Working Paper

The Role of Communication and Network Technologies in the Dynamics of Social Movements

Krystyna Marcinek, Rushil Zutshi, Omair Khan, Justin Grana, Marek N. Posard, Todd C. Helmus, Aaron B. Frank

RAND National Security Research Division

WR-A1646-2
June 2022

RAND
CORPORATION

For more information on this publication, visit www.rand.org/t/WRA1646-2.

Published by the RAND Corporation, Santa Monica, Calif.

The Role of Communication and Network Technologies in the Dynamics of Social Movements

Krystyna Marcinek Rushil Zutshi Omair Khan Justin Grana
Marek Posard Todd Helmus Aaron Frank

May 13, 2022

Abstract

We investigate the multi-faceted role of information technologies in the spread and dynamics of social movements. Specifically, we ask two main questions: 1) how do communication *and* network technologies impact the number and connectivity of movement participants, and 2) how does more efficient and more accurate surveillance technology impact an authority's ability to learn about the movement. Importantly, our simulation model includes both homophily and social influence, two established tenants of social movements and social relationships more broadly. Our results show that communication technology that increases spontaneous interaction helps to ignite social movements, while improvements in networking technology are more effective at accelerating the growth of social movements in their intermediate stages. However, when agents are allowed to join the movement, outreach is more effective at accelerating the growth of the number of participants. Our results also show that authority can gain highly accurate beliefs simply by observing network links (instead of individual actors) in all but the smallest social movements.

The ability to communicate with peers, both proximate and distant, is a central component of social movements. Historically, American slaves invented codes in the form of songs [1] and quilts [2] to facilitate communication and escape along the underground railroad. The advent of coffee shops in 17th and 18th century England is often acknowledged as an expansion of communication channels and a key contributor to the Age of Enlightenment [3]. Online communication platforms allowed Hong Kong protesters "to form small groups to initiate and coordinate their actions [4]." And in early 2021, political protesters used social media and other internet forums to share information, which ultimately resulted in violent protests at the U.S. Capitol.

As protesters develop and leverage advances in communication technologies, the empowered authority also leverages advances in data collection, storage, and processing technologies to better monitor and subdue the movement. For example, in the 2011 Syrian Civil Uprising, the Syrian government engaged in "monitoring and controlling a user's dynamic web-based activities," and implemented systems that were "capable of capturing webcam activity, logging keystrokes, [and] stealing passwords [5]." This suggests that both protesters and authority figures are leveraging cutting-edge technologies to achieve their objectives, but it is still unclear how this technological co-evolution impacts the overall dynamics of social movements.

This work develops and analyzes an agent-based model (ABM) of network formation to investigate such dynamics. Specifically, we first investigate how an increased ability to form communication channels impacts the overall growth (both nodes and links) of social movements. Then, we investigate how improved surveillance technologies—motivated by technological advances in data collection, storage, and processing power—impact an authority's ability to monitor and potentially disrupt a social movement. Ultimately, we characterize the interaction between the participants and the authority in the context of their technological endowments.

In order to remain as general as possible, the model examines how changes in communication technologies impact the number of participants and the connectivity among them without any reference to the ultimate goal of the social movement. Alternatively put, we abstract away from how—once connected—participants mobilize and coordinate resources in pursuit of an objective. Instead, we focus on the preliminary step of how the participants coalesce in the first place. The main benefit of this approach is that the model is agnostic to the participants' goals and thus provides general insights into *any* social movement. Of course, the main drawback is that the mechanisms through which a group achieves its goal are relevant for understanding (and predicting) any specific social movement. However, by establishing a framework that explains how social movements grow and connect, the model can serve as a building block onto larger models that consider both communication and mobilization. We also recognize that the surveilling body might not always be a government or law enforcement agency but may include private corporations or special interest groups. However, we use the term "authority" loosely to mean an agent with the desire and means to monitor a social movement

We follow the same logic for the authority. That is, we abstract away from any particular goal of the authority and focus on the authority's need to decide which of the agents are participants. The value of this approach rests on the assumption that, except in the most extreme cases, regardless of the authority's goal, acquiring information about who is a participant is a necessary pre-requisite for action. For example, if an authority wanted to disrupt a social movement by severing communication links between participant factions, it would need to know which of the citizens are participants that serve as links between disparate groups. Again, while this means that our model does not apply to *particular* social movements, it serves as a framework for building models that include the authority's disruptive action. We also recognize that the surveilling body might not always be a government or law enforcement agency but may include private corporations or special interest groups. However, we use the term "authority" loosely to mean an agent with the desire and means to monitor a social movement.

In order to ensure that the results are rooted in established principles, we include both homophily and social influence in the model. Homophily is the notion that similar individuals are more likely to have contact than two individuals that are different [6]. Such phenomena are well documented in domains such as adolescent friendship formation [7], entrepreneurial relationships [8], and, perhaps unsurprisingly, political opinion networks [9]. Social influence is the notion that an individual's opinions and beliefs are shaped by their peers [10]. Like homophily, social influence has been observed in a variety of contexts, including exercise habits [11] and product purchase decisions [12]. Both homophily and social influence have been observed in social movements [13, 14, 15].

Our model yields several interesting results. It illustrates how a change in communication technologies that increase participants' ability to contact agents outside of their immediate network (a term we call *outreach*) is most effective in igniting a social movement, both in terms of increasing the number of its supporters, as well as the accelerating connectivity among them. On the other hand, a change in networking technologies that increases participants' ability to contact agents within close network proximity (i.e., friends of friends) is most effective at accelerating the growth of the social movement through its intermediate stages. The final number of social movement participants in our model is determined by the underlying characteristics of the population—social influence, homophily, and the initial seed of the movement. However, for some sets of population characteristics, the final number of participants is very difficult to predict—it is possible that no new agent becomes a participant (that is, the social movement does not grow beyond the initial seed), as well as that all of them do, turning the entire population into supporters of the social movement.

On the surveillance side, we find that the authority figure is able to form accurate beliefs about the participation status of agents even when the node observations may be noisy so long as it can reliably observe the links between agents. Expectedly, we also find that the accuracy of these beliefs improves when agents are sampled more frequently. When we then compare improvements in the accuracy of these beliefs across these three parameters over different kinds of social movements (based on size and connectivity), we find that the accuracy of the authority's beliefs is most improved when the node observation process is less noisy, but this is less the case as the social movement becomes larger and more connected. Most of the same results hold even with social influence, but the dynamic nature of the social movement size results in less monotonic relationships between the accuracy of the authority's beliefs and the surveillance parameters.

Communication, Social Movements, and Agent-Based Models

Our work builds on—and incorporates—insights from several traditionally disparate disciplines. As mentioned above, our model incorporates homophily and social influence, two well-established tenants in the sociology of network formation. We quantitatively define[1] degrees of homophily and social influence and

[1]We do not claim that our quantitative definition of homophily and social influence is the unique definition, we only claim that it is reasonable in the context of social movements.

contribute to the literature by evaluating how changes in such degrees impact the dynamics of a social movement.

Our work is also related to the vast literature on behavior spread [16] and dynamic networks [17]. Using both analytical (examples can be found in [18], [19]) and computational (examples can be found in [20], [21]) methods, the goal is to derive properties of networks that form according to some underlying stochastic process. Our work is related to this line of research in that we also propose an underlying stochastic process for the evolution of the network. However, our work is different in several ways. First, we are interested in the state of the network as a function of time and nodes. This contrasts with the dynamic network literature that typically (though not always) focuses on the limits as time and the number of nodes gets large. Furthermore, we incorporate an authority figure that does not participate in the network but *monitors* the network. This is an element that is absent from the study of dynamic networks.

Our work is most related to other ABMs of communication and social movements. This includes models of political insurgency [22], revolution and technology [23], civil violence [24, 25, 26], radicalization [27] and more [28]. We also build on game theoretic models of mass action [29]. Our model adds to this literature in several ways.

First, our model trades intention for generality. That is, we focus specifically on how well participants can form a network and how well an authority can monitor the network without any regard for either party's objectives. This means that our model is not constrained to a specific social movement but can be used to explain network formation more generally. Of course, the trade-off is that our work stops short of expressing how changes in communication technologies affect how well the participants achieve their goals, but our model is meant to be modular and can be included in models that specifically incorporate participants' goals.

Secondly, we include technological change from both the participants' and the authority's perspectives. Many recent ABMs examine the impact of more accessible social media on the evolution of a social movement [30]. However, these models typically ignore that as participants are exposed to more technologically sophisticated tools, the authority's technological endowments are also increasing. So, for example, while participants are more able to connect, the authority's increased processing power and algorithmic advances may also increase its ability to monitor a social movement and eventually disrupt it. Our model explicitly takes these dynamics into account.

Third, our model is not only concerned with "steady-state" distributions but instead asks "how fast" do the participant's network and the authority's knowledge evolve? This contrasts with many ABMs of social movements that focus on the distribution of end-states [31]. That is not to say that our model is better or more relevant than those focusing on end-states. Instead, it only illuminates a different aspect of complex social movements.

Model

In this section, we present a narrative overview of the model followed by the full technical specification. The model overview suffices for understanding the key components of the model and the results. The full technical specification provides the necessary detail to replicate the model.

Model Overview

Our model is a dynamic model of network formation with both *citizens* and an *authority*. A fixed number of citizens are deemed social movement *participants* while the rest of the citizens are non-participants. Each citizen, either participant or non-participant, has an additional attribute called its *identity*. The identity is independent of whether the citizen is a social movement participant and represents characteristics of the citizen other than its proclivity to join the movement. For example, it might capture demographic information, tastes, interests, or education levels. At the start of the model, all citizens are disconnected, and there are no communication channels between any two citizens.

In each time step, an agent is chosen at random to potentially form links with other agents under two mechanisms, which we label *outreach* and *networking*. Outreach is the process where the agent can survey a fixed number of random agents. If both agents are participants or both agents are similar enough in identity, they form a link. This process represents a simple notion of homophily and can be interpreted as serendipitous networking such as meeting a person at a public park or cafe. As a digital example, an agent might 'stumble' upon another agent's blog while on the internet and subsequently connect to that person through a social media platform.

Networking refers to the process where the agent can survey a fixed number of agents who have connections with its current connections. For example, if agent Alex is connected to agent Bob and agent Bob is connected to agent Chris, then Alex can survey Chris because it is connected to Chris through Bob. Again, homophily dictates that agents under this process form a link if they are either both participants or if their identities are sufficiently similar. A digital example of the networking mechanism would be two agents that connect via a closed online group they were both invited to by a common friend.

The parameters that govern how many agents a given agent can survey in one time step can be interpreted as the inverse of the communication costs to forming links. For example, if an agent has a fixed amount of time it can dedicate to communicating and meeting new people, raising the number of queries an agent can make to other agents is equivalent to lowering the communication cost. A key feature of our model is that there are separate parameters that govern outreach and networking, and thus it is possible to isolate the impacts of advances in networking and advances in outreach technology. In real-world communication systems, advances in social media platforms allow providers to suggest connections based on friends of friends and thus reduce the cost of networking. On the other hand, websites centered around filling job vacancies facilitate communication between parties that are not otherwise linked and thus reduce outreach costs. Our model is able to disentangle the impacts of these similar but subtly different notions of advances in communication technologies. The results ultimately show that the relative impact of increasing networking versus increasing outreach depends on how much the movement has already progressed and the initial prevalence of the social movement.

Finally, we allow agents that were not originally participants to become participants if they are connected to enough participants. There is a parameter that specifies the agent's propensity to become a participant, and that parameter captures the degree of social influence in the model. While the parameter is the same for every agent, we run several experiments—including a base case with no social influence—to explore the impacts of various degrees of social influence.

We are interested in two main quantities that govern the dynamics of the social movement. First, we characterize how changes in communication technologies (the increased ability for outreach and networking) impact the growth of the number of participants. This insight tells us how different communication technologies can influence whether a social movement is a slow boil, explosive, or some combination of the two. We also analyze the connectivity (number of links) between participants. Connectivity is important under the assumption that regardless of the participants' goals, having more communication channels facilitates the mobilization of resources towards the participants' ultimate goal.

The final piece of the model is an authority agent. While the authority ultimately desires to alter the course of the social movement, this work specifically focuses on how the authority collects information about the social movement. The key question we address is how changes in surveillance and data processing technologies impact how an authority determines which citizens are participants.

To form beliefs about which agents are participants, the authority observes a random subset of agents at regular time intervals. The number of agents per time step (which can be less than 1) represents the processing power and storage capacity available to the authority. Specifically, the interpretation is that as the authority's processing power and data storage capacities increase, it is able to collect and analyze data on a wider set of individuals in a given time period. The authority's observation of a given agent consists of two components. First, there is a *node observation process* where the authority observes a noisy signal of whether a given agent is a participant. The amount of noise captures the quality of the authority's surveillance technology, where a low-noise signal represents a more sophisticated data collection and analysis technology.

Second, for each agent the authority observes, there is an *edge observation process* in which the authority observes a subset of the agent's personal network. Specifically, the authority has a fixed probability of observing each of the agent's existing edges. This probability represents the quality of the authority's technology allowing network structure surveillance. A low probability may mean that the authority can observe connections only under very strict conditions, for example, when two agents are physically co-located. A high probability could, for example, mean that the authority can observe connections despite the mode of communication chosen by the participants, such as social network data mining. Notably, the node and edge observation technologies are governed by separate parameters. This allows us to disentangle the impacts of increased individual surveillance versus increased network surveillance, our key comparative statics when analyzing the authority.

A fully rational authority would have a prior belief and use the data from its node observation process and its edge observation process to compute a Bayesian probability of each agent being a participant. However, this exact probability is analytically intractable due to the underlying model of network formation and would require significant Monte Carlo experiments to estimate. This is exacerbated in models of social influence where agents can switch from being non-participants to participants. Since it is unlikely that a real-world authority would be able to compute these probabilities accurately, our implementation of an authority adopts a heuristic that, with infinite samples, can separate participants from non-participants. Intuitively, the heuristic uses the fact that, on average, participants have more connections than non-participants (because participants are connected to one another as well as with those with a similar identity) and thus combines the degree of a node with the node observation data to assign a numerical value (not a proper probability) to a node being a participant. With infinite samples, the score of all participants would be strictly positive, while the score of all non-participants would approach 0. We show how well this classification process works using a receiver operator curve (ROC) in the results section of the paper.

Model Specification

The main object of the model is a time-indexed graph $\mathcal{G}_t = (V, \mathcal{E}_t)$ where V is the set of N vertices or nodes and $\mathcal{E}_t$ is the set of undirected edges at time t, and time is discrete.[2] Each vertex represents an agent and edge $e^t_{ij} \in \mathcal{E}_t$ represents a communication channel between agent i and agent j at time t.

Each agent i, (equivalently, a vertex) has two attributes, $\beta^t_i \in \{0, 1\}$ and $\theta_i \in (0, 1)$. The binary attribute β^t_i indicates whether the agent is a participant and equals 1 if the agent is a participant and 0 otherwise. The uniformly distributed attribute θ_i represents an agent's identity and captures characteristics other than its proclivity to protest. Again, this may represent the agent's socioeconomic status or a collection of interests. While β^t_i can change over time, θ_i is fixed.

At each time step t, an agent is selected at uniform random. Denote the agent selected at time t as a_t. Agent a_t can form links in two ways:

- ***Outreach*** First, agent a_t randomly queries L_1 other agents from the set $V^t_{a_t} = \{v \in V | v \neq a_t, (a_t, v) \notin \mathcal{E}_t\}$, i.e., any agent that is not already connected to a_t. Let $v \in V^t_{a_t}$ be any of the L_1 agents that agent a_t queries. Then, the edge (a_t, v) is added to $\mathcal{E}_t$ if *either* of the two conditions is met:

  1. $\beta^t_{a_t} = \beta^t_v = 1$
  2. $|\theta_{a_t} - \theta_v| \leqslant c$

  Intuitively, an edge is formed if either both agents are movement participants or their value of θ is sufficiently close.

- ***Networking*** Second, agent a_t randomly queries L_2 agents from the set $\widetilde{V}^t_{a_t} = \{v \in V | v \neq a_t, (a_t, v) \notin \mathcal{E}_t, (a_t, j) \wedge (j, v) \in \mathcal{E}$ for some $j\}$ and connects according to the same conditions as when it conducts outreach. Intuitively, the set $\widetilde{V}^t_{a_t}$ is all nodes where the shortest path between a_t and a node in $\widetilde{V}^t_{a_t}$ is 2.

These two processes define the evolution of $\mathcal{E}$. Specifically, $\mathcal{E}_t$ is the union of $\mathcal{E}_{t-1}$ and the links that are formed during outreach and networking at time t.

After all of the links at time t are formed, agents update their value of β. Specifically, if agent i is connected to at least w participants and is itself not a participant, β^t_i switches from 0 to 1. Note that this is an iterative process since the effects of an individual agent turning into a participant can spread throughout the network if that agent is connected to several other agents that were previously connected to $w - 1$ participants.

[2] Since edges are undirected, as a shorthand, we use notation such that the edge $(x, y) = (y, x)$.

The key parameters are L_1, L_2, c, p, and w. Specifically, L_1 controls an agent's ability to leverage communication technology for outreach, where L_2 represents the agent's ability to leverage communication technology for networking. The parameter w represents how susceptible agents are to social influence. The parameter p defines the number of participants at $t = 0$, i.e., the initial seed of the movement. The parameter c represents *inclusivity* since it represents how much agents are willing to include other agents with different identities into their network. Furthermore, c is the inverse of homophily. The lower the value of c, the more agents need to have similar identities for them to connect. Since we are interested in communication technologies, our main computational experiments explore how changing L_1 and L_2 impact the growth of the social movement and how that growth rate depends on c, w, and p, which are measures of homophily, social influence, and the initial movement size, respectively.

The authority agent dynamically observes the network as it forms, but not necessarily at every time step. Specifically, every k time-step, the authority observes m citizens. The parameter k represents how fast an authority can process data obtained through citizens and draw insights, while m represents how much data an authority can collect simultaneously. Of course, the actual data collection and processing procedures are likely asynchronous, but we assume the processes are synchronous to maintain simplicity.

For each citizen i of the m citizens sampled at time t, the authority has a **node observation process** where it observes y_{it}, which equals β_i^t with probability $1 - \delta$ and $1 - \beta_i^t$ with probability δ. Intuitively, the authority observes the correct value of whether or not the citizen is a participant with probability $1 - \delta$ but makes an error with probability δ.[3] The parameter δ captures yet another notion of surveillance technology; the lower δ, the more accurate the surveillance. The authority keeps a record of these observations $y_i^t = \{y_{i1}, y_{i2}...y_{ik_t}\}$ where y_{ij} represents the authority's jth observation of citizen i and k_t represents the total number of times the authority observed agent i up to and including time t.

In addition to the value of β_i^t, for each agent i of the m sampled agents, the authority also completes an **edge observation process** where it observes $e_{ij}^t \in \mathcal{E}_t$, $j \in \{1, 2...N\}$ with probability γ. This process represents the authority's ability to conduct surveillance on a citizen's network. If $\gamma = 0$, then the authority cannot observe network links. If $\gamma = 1$, the authority perfectly observes a citizen's connections. The authority also keeps track of the degree of each agent i, given by

$$d_i^t = |\{e_{ij}^t | \exists t \text{ such that the authority observes } e_{ij}^t\}|$$

To quantify the authority's belief on whether a citizen is a participant, it must combine observations y_i and d_i. It does this through the function

$$s_i^t = (r_i^t)^\alpha (d_i^t)^{1-\alpha} \tag{1}$$

where

$$r_i^t = \frac{\prod_{t=1}^{k_t} (1-\delta)^{y_{it}} \delta^{1-y_{it}} \frac{p}{N}}{\prod_{t=1}^{k_t} (1-\delta)^{y_{it}} \delta^{1-y_{it}} \frac{p}{N} + \prod_{t=1}^{k_t} \delta^{y_{it}} (1-\delta)^{1-y_{it}} (1 - \frac{p}{N})} \tag{2}$$

Intuitively, r_i^t is the authority's Bayesian posterior belief of citizen i being a participant *if it only completes a node observation process*. However, to leverage the data collected from the edge observation process, it weights this probability by the node's degree to obtain s_i^t. Again, the motivation is that a participant will, on average, have a higher degree than a non-participant, so raising an agent's degree would increase an agent's score. In the case of social influence, one could expect that if the participant status of each citizen can change over time, the authority should limit their "memory" to some number of last observations. However, the experiments conducted for the model with social influence have shown that constraining memory negatively affects the classification. Consequently, we allow the authority to form their belief based on all previous observations of the citizen node.

We are interested in an aggregate measure that captures the accuracy of the authority's belief across all agents. For this, we will use ROC curves and compute the area under the curve (AUC) [32] as an aggregate

[3] Without loss of generality, we assume $\delta \leqslant 0.5$.

measure of the authority's beliefs. A ROC graph is a two-dimensional graph with the true positive rate plotted on the y-axis and the false positive rate plotted on the x-axis, and each point on the graph represents a different classification threshold. Specifically, the bottom left of the graph (0,0) represents a strategy of never issuing any positive classifications—so there are no false positives, but the true positive rate is also zero. Conversely, the upper right (1,1) represents a strategy of always issuing positives. The point (0,1) represents a perfect classifier. The diagonal represents the strategy of randomly guessing a class. As a rule of thumb, one classification strategy is better than another if it yields a point that's located more northwest than the other in the ROC space.

While a ROC graph represents the performance of a classifier in two dimensions by varying the classification threshold, it may often be helpful to reduce this performance down to a single scalar value to compare the performance of multiple classifiers. One such method is to calculate the area under the curve (AUC) in a ROC graph. Since the AUC is some portion of a unit square, its value is always between 0 and 1. However, as mentioned previously, the diagonal line corresponds to a random guessing strategy, which corresponds to an AUC of 0.5. Consequently, any realistic classifier should have a value between 0.5 and 1. The AUC has an important interpretation which makes it a reasonable proxy for classifier performance. It is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. Tracking the AUC over time allows examining the changes in classifier quality throughout a simulation.

Results

This section details the main results. We begin with the simplest scenario and examine the dynamics of the participant network formation without social influence. In these initial simulations, we focus on the *connectivity* among participants since the number of participants is fixed. We then introduce the authority and show how the accuracy of its beliefs depends on the parameters governing its observation process. We then repeat the exercise with social influence. Since the number of participants changes over time, these results focus on the evolution of the *number* of participants. Finally, we examine how the social movement growth influences the authority's ability to track participants accurately.

Parameters

We vary both the quality of technologies themselves and the underlying characteristics of the population, according to the parameter values in Table 1. For the authority agent, we collapsed sample size and sampling frequency into one parameter reflecting the frequency of sampling one agent. Parameters defining surveillance technologies were varied separately for node and edge observation processes, as shown in Table 2. While the exact values of the parameters are not necessarily of interest, we performed several robustness checks to ensure our experiments captured the full spectrum of the parameter space. To ease presentation, in the subsequent analysis, we refer to the parameter values as "low", "medium", and "high" instead of their precise numerical values as specified in tables 1 and 2.

Characteristics	**Parameter**	**Values**
Underlying characteristics of the population	Population size N	100
	Initial seed of movement p	5, 10, 15
	Inclusivity c	0.05, 0.075, 0.10
	Social influence w	∞ (no influence), 5, 4, 3
Technological capabilities	Outreach technology L_1	1, 2, 3
	Networking technology L_2	1, 2, 3

Table 1: Participants' parameters

Characteristics	Parameter	Values
Survailance capacity	Sampling frequency k	20, 10, 5
Surveillance accuracy	Noise in node observation δ	0.1, 0.3, 0.5
	Probability of link observation γ	0.1, 0.5, 0.9

Table 2: Authority's parameters.

Participant Dynamics Without Social Influence

To understand how technology impacts participant connectivity, we show how the degree of participants evolves over time when there is no social influence, and thus the number of participants is fixed (i.e., $w = \infty$). Figure 1 shows the average degree of a participant and a non-participant for 40 simulations at the same parameter values. The key insight is that the average degree of participants 1) converges, 2) is stochastic and 3) is higher than non-participants. The fact that the quantity is convergent and higher for participants is not surprising. With infinite samples and without social influence, each participant will be connected to every other participant *and* every non-participant with a similar identity, while non-participants connect exclusively on the identity. The stochastic nature is also not surprising. Since each agent's identity is drawn from a uniform distribution at the start of the simulation, the number of agents another participant connects to based on identity is random.

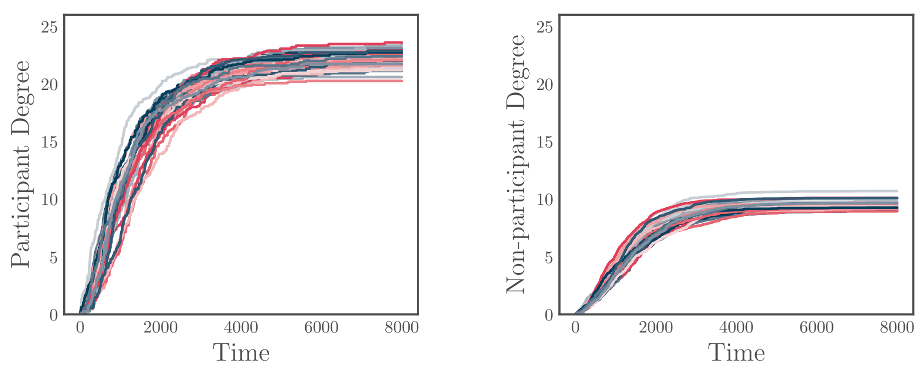

Figure 1: Convergence of average participant and non-participant degree for fixed parameters ($L_1 = L_2$: low, p: high, c: low).

Due to the stochastic nature of the participant's degree, we use Monte Carlo simulations to compute the average degree when holding the parameters fixed. Therefore, when we use the term "average degree," we refer to the average over *all participants* and *40 Monte Carlo simulations* for a fixed set of parameters. Specifically, we repeat each computational experiment with the same parameter values 40 times, compute the average participant degree and then compute the average over each of the 40 Monte Carlo simulations. Specifically

$$\text{Average participant degree at time t, } \bar{d}_t = \frac{1}{40} \sum_{m=1}^{40} \frac{1}{p} \sum_{i=1}^{N} \beta_i^{t,m} d_i^{t,m} \tag{3}$$

where the superscript m corresponds to simulation number m.

Another baseline result is that the final average degree of a participant depends on two key parameters, the initial number of participants, p and inclusivity, c. Figure 2 plots the average participant degree for each of the different parameter sets at the end of the simulation. For each parameter value, the results group around three values corresponding to three levels of the other parameter. For example, results for the parameter sets with low inclusivity—depicted in the most left stack—group around 13, 18, 22.5 for the sets with low, medium, and high initial seed, respectively. The plots show that the parameter values are such that increasing the initial number of participants by one category (low to medium, for example) has roughly the same effect as increasing inclusivity by the same category (also low to medium, for example). This is important because it establishes that c and p are on a similar scale, and differences in scale are not driving the results.

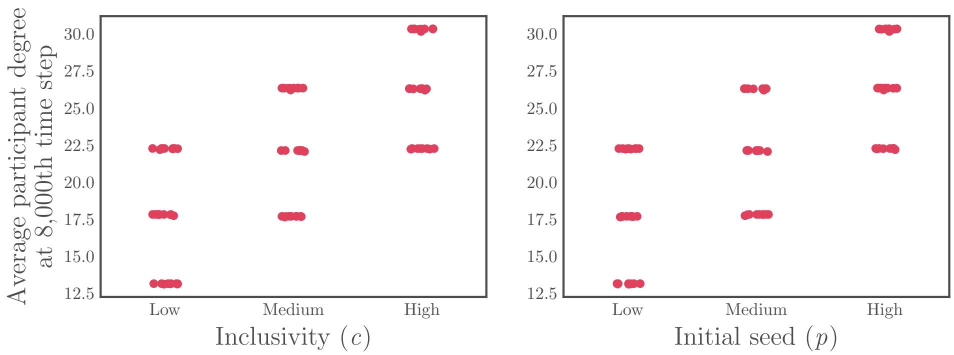

Figure 2: Distribution of average participants degree at the limit, grouped by c and p.

In this baseline model, we were particularly interested in the speed of social movement network formation, i.e., we only investigate connections between participants. To understand the relative impacts of increasing network and communication technologies on participant connectivity, we investigate how much faster the participants' network converges to its limits as we increase one technology endowment versus another.

This result is presented in Figure 3. The horizontal axis represents the proportion of links among participants relative to the network's limit (all participants connected to one another). The vertical axis represents how much faster that level of convergence is reached due to an increase in outreach versus an increase in networking. So, for example, in the left plot, when p is low and c in medium (blue annotated line), increasing outreach (L_1) from low to medium means that the proportion of connections reaches 50% almost 500 time steps sooner than if networking (L_2) were to increase from low to medium. These graphs provide two high-level insights.

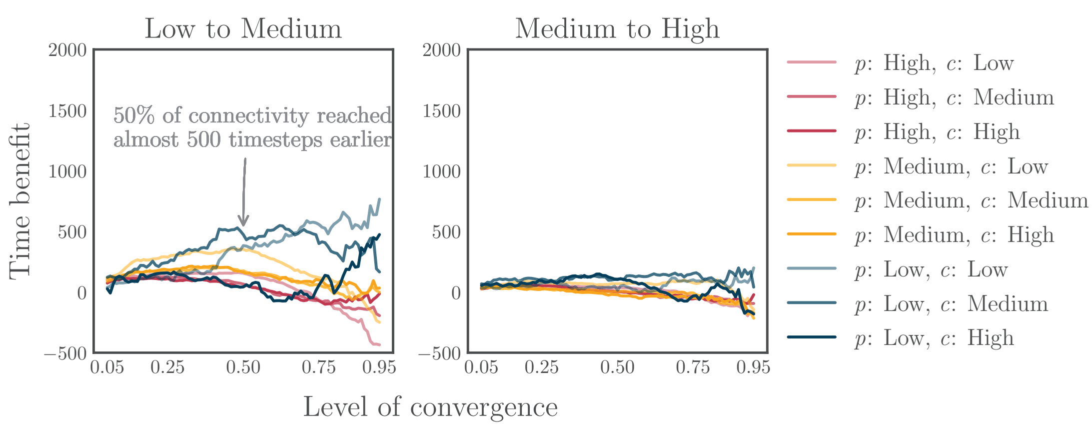

Figure 3: Time benefits of increasing outreach versus networking (as a function of the level of convergence).

First, the relative benefits of outreach and networking depend on the initial size of the social movement. With a small number of participants, increasing outreach leads to faster link formation for any given level of connectivity. However, with a larger number of participants, increasing networking can lead to faster convergence to the ultimate network state. Intuitively, if there are very few participants, most of the connections are formed based on identity. Therefore, it is unlikely that a participant would find another participant through networking. However, when there are more participants in the population, more links are initially formed among participants, and the higher the chance of meeting other participants via one's existing connections. In total, this means that in all but the smallest movement size, increasing outreach is more apt to "ignite" connectivity among participants, whereas increasing networking is more effective at quickly bringing the participant network to its ultimate potential.

Secondly, the plots show that the relative impact of increasing outreach versus networking is nonlinear, which is evident from the difference in the height of the lines in the left and right plots. Specifically, the relative difference in convergence time from increasing outreach versus networking from low to medium is much more than an increase from medium to high. This suggests that only in technologically limited cases do increases in networking and outreach technologies have asymmetric impacts on network formation. However, as both technologies reach an advanced state, their impact on participant connectivity becomes indistinguishable.

Surveillance Without Social Influence

This section investigates how improvements in different surveillance technologies impact an authority's ability to discern participants from non-participants. Since each participant receives a score from the authority, a key quantity is what participant score cutoff the authority uses to classify citizens as participants. This score would ultimately depend on the authority's cost and benefits of correct and incorrect classifications. In order to gauge the efficacy of the authority's classification process in the general sense without reference to the authority's payoff, we use a Receiver Operator Curve (ROC) to understand the trade-off between the true positives (benefits) and false positives (costs).

An example of the ROC graph from our analysis is shown in figure 4. The black line represents a ROC at the beginning of a simulation ($t = 50$)—it almost overlaps with the diagonal (gray dashed line), representing a random guess. As the simulation progresses, the curves shift towards the top left corner, which means that the authority's ability to classify and detect participants improves.

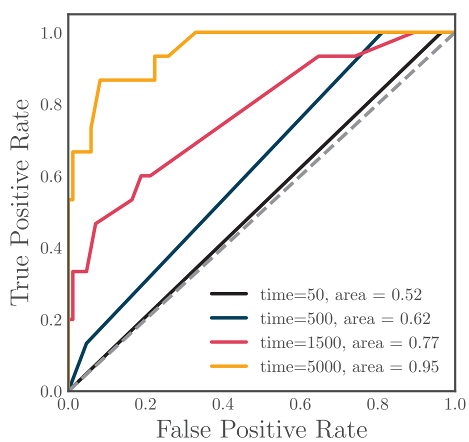

Figure 4: Receiver Operating Characteristic (ROC) curves at various time steps for fixed parameters ($L_1 = L_2$: low, p: high, c: low, k: low, δ: medium, γ: low).

Figure 5 shows how the surveillance ability of the authority changes over time (here for 10 Monte Carlo simulations). Specifically, the figure plots the area under the curve (AUC) at various time steps. Just like in the participant case, the AUC at any timestep is stochastic and thus necessitates Monte Carlo simulations. Therefore, when we use the term "mean AUC" we refer to the average over *40 Monte Carlo simulations* for a fixed set of parameters.

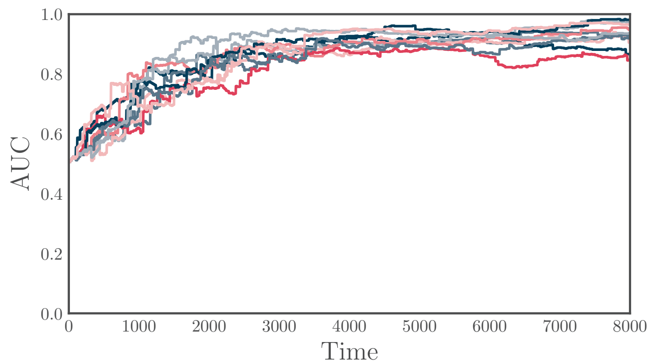

Figure 5: AUC at various time steps for fixed parameters ($L_1 = L_2$: low, p: high, c: low, k: low, δ: medium, γ: low) in 10 Monte Carlo simulations.

Noise in Node Observation (δ) and Probability of Link Observation (γ)

Figure 6 presents results of the AUC over time for different configurations of the participant parameters and link and node observation parameters while sampling frequency is held constant at 1 agent observation per 10 time steps. The figure illustrates that even when the authority's node observation process is very noisy, it still performs relatively well as long as it observes links with a high probability. This is best illustrated in the right column where γ is high. In all but the first row, the red line—representing high noise in the node observation—is relatively close to the blue line—representing low noise in the node observation process. However, this is *not* the case with a low number of initial participants and low communication technologies, which is the scenario in the first row of plots in figure 6. In this case, participants connect so slowly that network information provides very little information to the authority, and the authority's beliefs are driven primarily by its observation of nodes. Thus when the node observation process is noisy, the authority cannot "fall back" on the edge observation process to form its beliefs.

Another curious result arising from these plots is that the authority may sometimes have a higher AUC with *higher* noise in the node observation process, which is the case when the red line is higher than the yellow line. This could be caused by the node observation process being effectively ignored when it is too noisy. While this is undoubtedly an artifact of the heuristic score function, it shows the importance of correctly aggregating participant data. Of course, this is true from a modeling perspective but is also true in practice. Data come from disparate sources during real protests, and with constraints on data sharing, modeling, and processing, authorities are often required to use heuristics. This result shows how seemingly benign and innocuous heuristics can lead to counter-intuitive results and reinforces the importance of an authority taking a principled approach to data analysis.

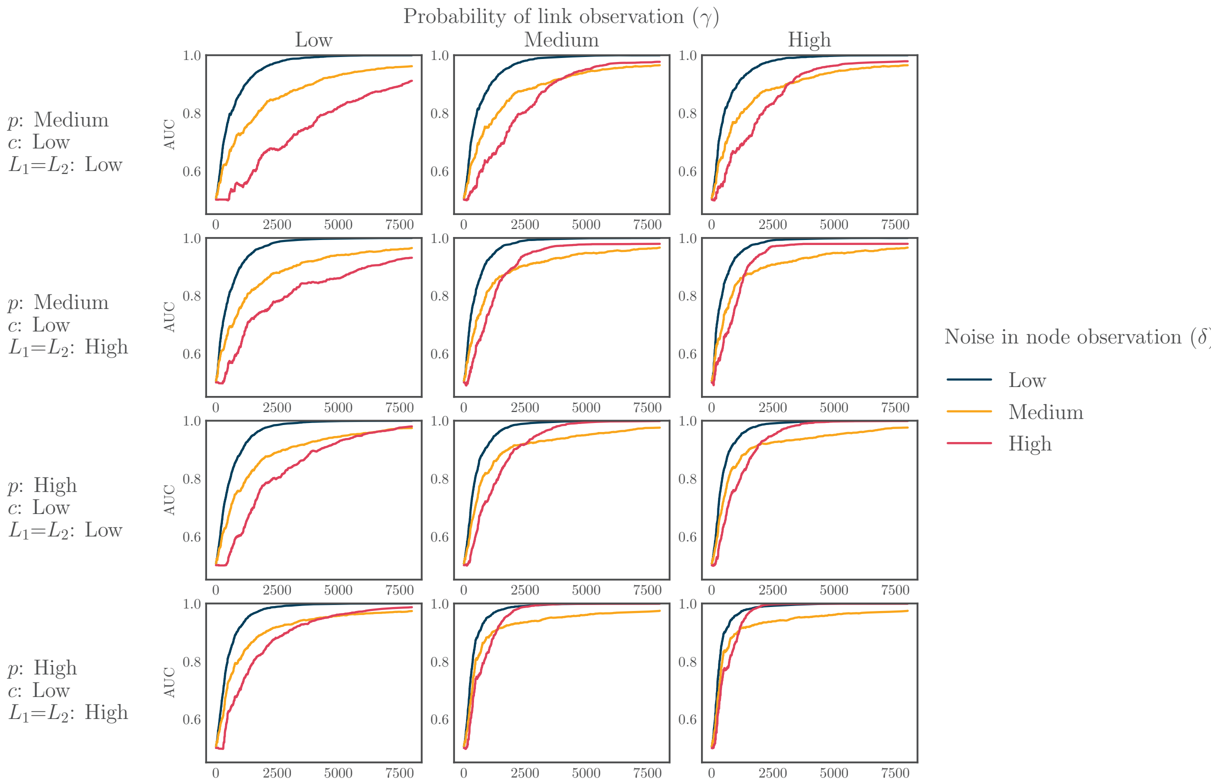

Figure 6: AUC at various time steps for δ (low, medium, high) over different γ and 40 Monte Carlo simulation (k: medium)

Sampling Frequency (k)

We can compare how changes in the sampling frequency k impact the authority's beliefs at different probabilities of link observation and participant parameters. These results are summarized succinctly in figure 7. This depicts an expected trend of a quicker convergence to AUC=1 when the sampling frequency increases from sampling a node every 20 time steps (low sampling frequency) to every 5 time steps (high sampling frequency).

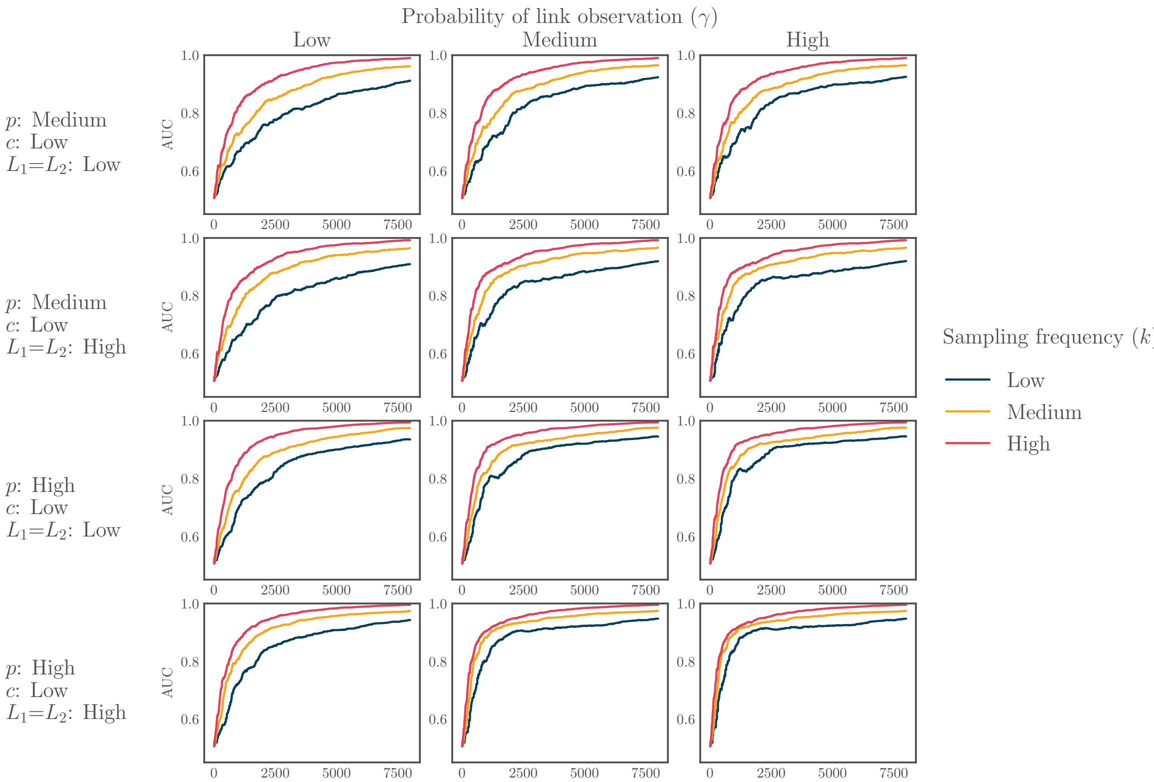

Figure 7: AUC at various time steps for k (low, medium, high) over different γ and 40 Monte Carlo simulation (δ: medium)

Finally, we can compare how changes in the sampling frequency k impact the authority's beliefs relative to changes in the noise in its node and link observation process. These results are summarized succinctly in figure 8. This figure shows gains to AUC over time (averaged for each parameter over all other parameter combinations) for changes in each of the three authority parameters. The plot shows that the impact of changing the sampling frequency is independent of the underlying participant dynamics. This is demonstrated by the relatively small changes in the height of the lines in the first column of the figure. This is in contrast to noise in the link observation process, where improvements have the largest impact in large social movements with fast connectivity (middle column plots are highest in the bottom row) and noise in the node observation process, where improvements have the largest impact in small social movements with slow connectivity (third column plots are highest in the top row).

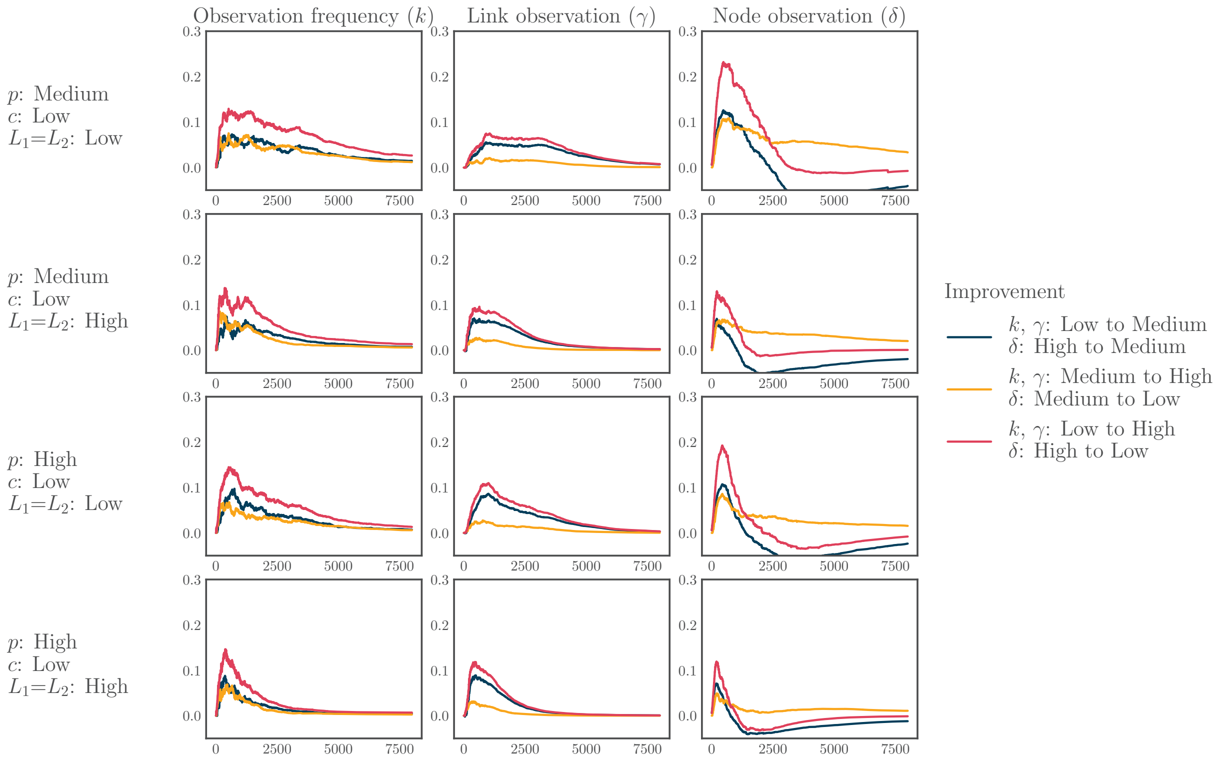

Figure 8: Average AUC improvement over time for improvements in k, γ, and δ.

Participant Dynamics With Social Influence

In this section, we re-investigate the participant dynamics, allowing non-participants to become participants via social influence. Specifically, if a non-participant is connected to at least w participants, it becomes a participant. Importantly, social influence is a one direction mechanism—participants cannot become non-participants. In the previous sections without social influence, the number of participants was fixed, so we focused on the connectivity among participants. With social influence, the number of participants varies over time, and therefore, we focus our analysis on the size of the social movement.

Once again, the number of participants is stochastic and convergent. This is illustrated in Figure 9, which shows the change in the number of participants over time for 40 simulations at the same parameter values. In this particular case, all 40 simulations started with 10 participants; however, the final number of participants varies between 40 and 100. Similar to the term "average degree", the term "average number of participants" refers to the number of participants averaged over 40 Monte Carlo simulations:

$$\text{Average number of participants at time t, } \bar{p}_t = \frac{1}{40} \sum_{m=1}^{40} \sum_{i=1}^{N} \beta_i^{t,m} \tag{4}$$

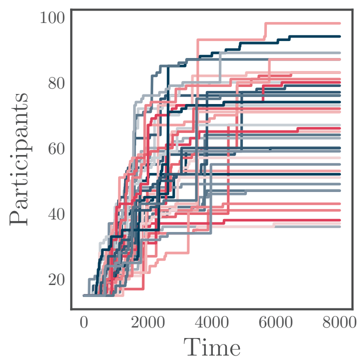

Figure 9: Convergence of the number of participant for fixed parameters ($L_1 = L_2$: low, p: high, $c$low, w: high).

Figure 10 shows the distribution of the average number of participants at the 8,000th time step grouped by the inclusivity and the initial number of participants parameters. For example, the most left stack presents results for the parameter sets with low inclusivity, where colors signify sets with different levels of social influence; dots in each color cluster around three different values corresponding to three levels of the initial number of participants. The plots show that even with a high degree of social influence (red dots), it is not guaranteed that everyone will become a participant. For instance, when both the inclusivity and the initial number of participants are low, the final number of participants remains low, which is represented on the plot by the red dots clustered around 5 in the most left stack. This is because even if social influence is high if c and p are low, non-participants do not have the opportunity to connect to participants and subsequently become participants themselves.

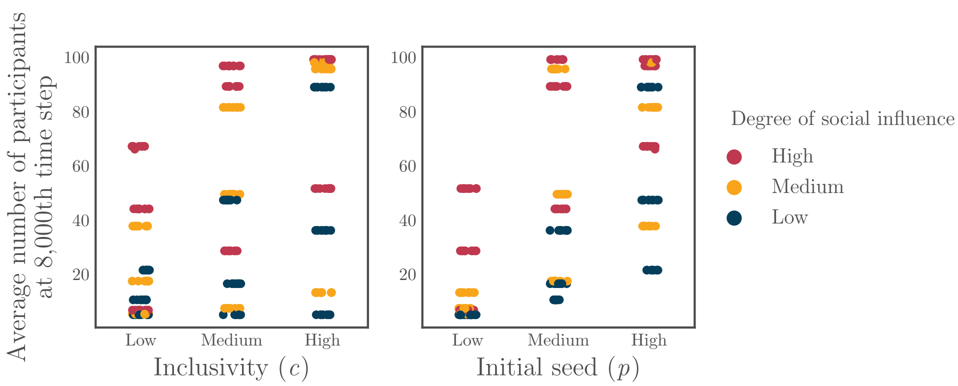

Figure 10: Distribution of the average number of participants at the limit, grouped by c and p.

In addition to the number of participants, another relevant quantity is the variance in the number of participants, as this captures the uncertainty associated with how large a social movement might ultimately get. To gain insight into this feature, figure 11 presents the standard deviation of the number of participants at the limit for three different levels of social influence as a function of the inclusivity and the initial number of participants. As we can see at each level of social influence in this population of 100 agents, the standard deviation of the ultimate number of participants ranges from 0 to 40.

The plot shows that the standard deviation is smallest either when almost every agent becomes a participant or when almost none does. The former is possible with high social influence in a population of high inclusivity and considerable size of initial movement seed (left graph, top row). The latter happens when both social influence and initial movement size are low (right graph, left column). High uncertainty is in turn associated with an overlap of these conditions. For example, in a society that encourages connections (high threshold of c), even a small number of participants can spark a social movement growth since the society is well-connected. We can see this effect in the left graph with high social influence, where we observe the highest standard deviation when the inclusivity is high, but the initial number of participants is low.

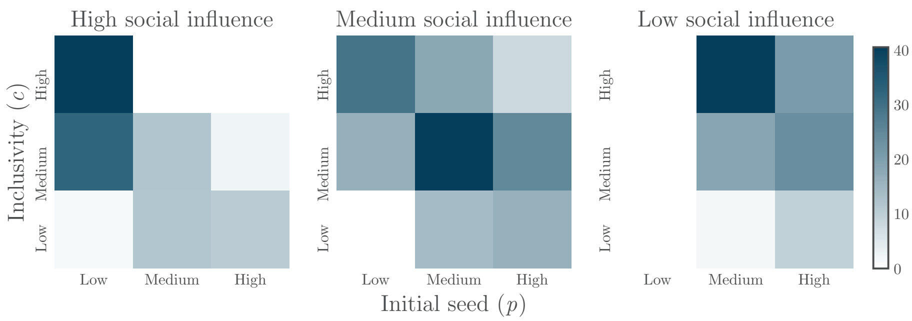

Figure 11: Standard deviation of the number of participants at the limit

Interestingly, the cases with the highest variance are, in fact, cases of multimodal distribution, where either no one becomes a participant or everyone. This result is presented in figure 12. It is especially true for the highest variance parameter sets with high and low social influence (blue and yellow histograms). With medium influence (red histogram), the steady-state number of participants can be anywhere between the initial number of participants (10) and the maximum number of participants (100). This speaks to the difficulty of predicting the pace of propagation of a social movement. The same characteristics of a movement at the outset can lead to vastly different popularity over time.

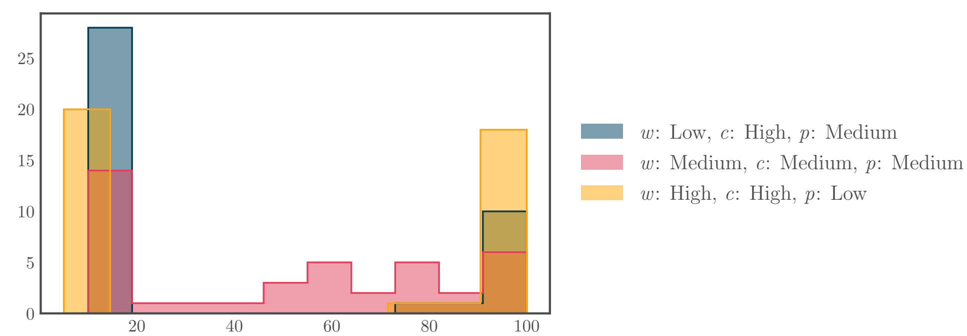

Figure 12: Distribution of participants number at the limit for the highest variance parameter sets.

Another interesting question is how fast the social movement grows as a function of communication and network technologies. As increasing either networking or outreach technologies will unambiguously speed up the growth of the movement, the key question is how much does an increase in networking technology speed up the growth of the movement relative to an increase in outreach. Figure 13 demonstrates this by showing at each time step how many more participants there are (on average) by increasing outreach instead of networking when the degree of social influence is high. For example, the red annotated line in the left plot of figure 13 says that when p and c are high, increasing outreach ability from low to medium will result in an average of 27 more participants at time step 700 than if networking were to increase from low to medium.

Since most of the lines are weakly greater than 0, the plot shows that increasing outreach appears to almost always lead to a higher number of participants at any given timestep. That pace of movement growth depends on the underlying characteristics of the population—the higher the inclusivity, the higher the pace. Increasing networking leads to faster growth in the number of participants only when the initial number of participants is low (dark blue line on the left graph). This is due to the fact that when the number of participants is low, it is very unlikely to randomly sample a non-participant that is already connected to w participants. The advantage of increasing outreach instead of networking in all other cases indicates that a social movement would likely grow faster in societies endowed with better outreach technologies. The networking, on the other hand, becomes important later in the process, when members of the movement connect, and thus, the movement recognizes its strength, as the model without social influence indicated.

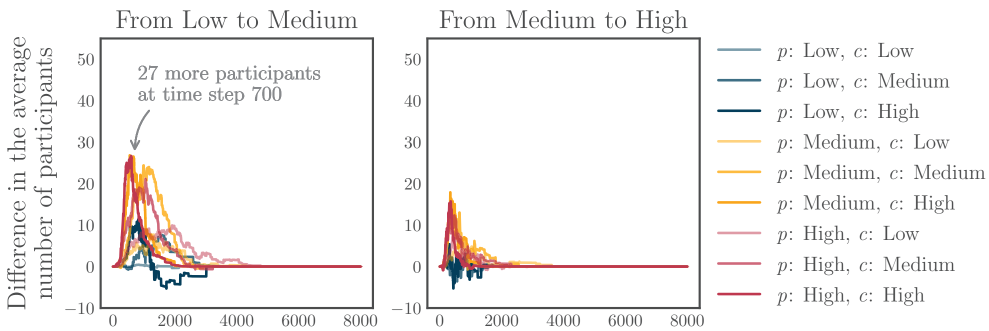

Figure 13: The difference in the average number of participants per timestep resulting from increasing outreach versus networking (high social influence)

Surveillance with Social Influence

In this section, we return to the analysis of authority's surveillance capabilities allowing for social influence. As described in the model section, the computation of the authority's belief of citizen i is based on all previous observations of the citizen node, even though now each citizen's participant status can change over time. While results are similar to the case of surveillance without social influence, there are some important distinctions highlighted below.

Noise in Node Observation (δ) and Probability of Link Observation (γ)

Figure 14 presents results of the AUC over time for different configurations of the participant parameters and probability of link observation parameter while sampling frequency is held constant at the medium level. As was the case when there was no social influence, even when the authority's node observation process is very noisy, it still performs well as long as the link is observed with high probability. In fact, in most cases, it performs better than the configuration with the least noise in the node observation process. Moreover, an important distinction is that in this case, this result holds even with a relatively low probability link observation process. This is likely because the noise in the node observation process is exacerbated when agents can turn into participants. For similar reasons as those outlined in the case without social influence, counterintuitively, the authority may perform better when the noise in node observation is higher in this case as well.

The convergence of the AUC is also not as monotonic as is in the case without social influence due to agents turning into participants. This is especially true when it is easier for "neutral" agents to become participants (rows 1 and 3). Figure 15 presents results of the AUC over time for different configurations of the participant parameters and noise in node observation parameter while sampling frequency is held constant at 1 agent observation per 10 time steps. We find that the main result from the no social influence case holds here—there are still diminishing returns to the improvements made to the AUC by improving the link observation process. However, what is different is that the biggest improvements in AUC are made at the beginning before new agents start becoming participants. Further, if the noise in the node observation process is low and the networking and outreach parameters are low, then the improvement in AUC over time is non-monotonic. The AUC starts to dip in the period of dynamic movement growth and then improves again once the number of participants stabilizes.

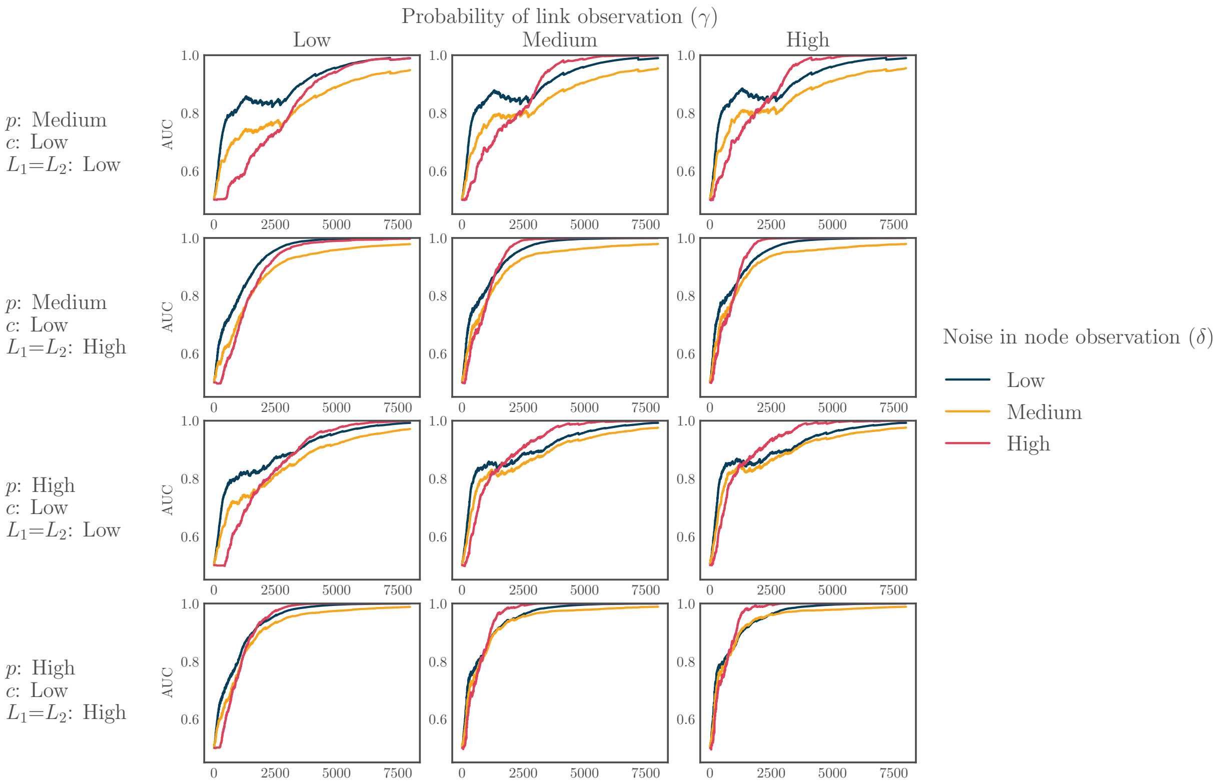

Figure 14: AUC at various time steps for δ (low, medium, high) over different γ and 40 Monte Carlo simulations (k: medium)

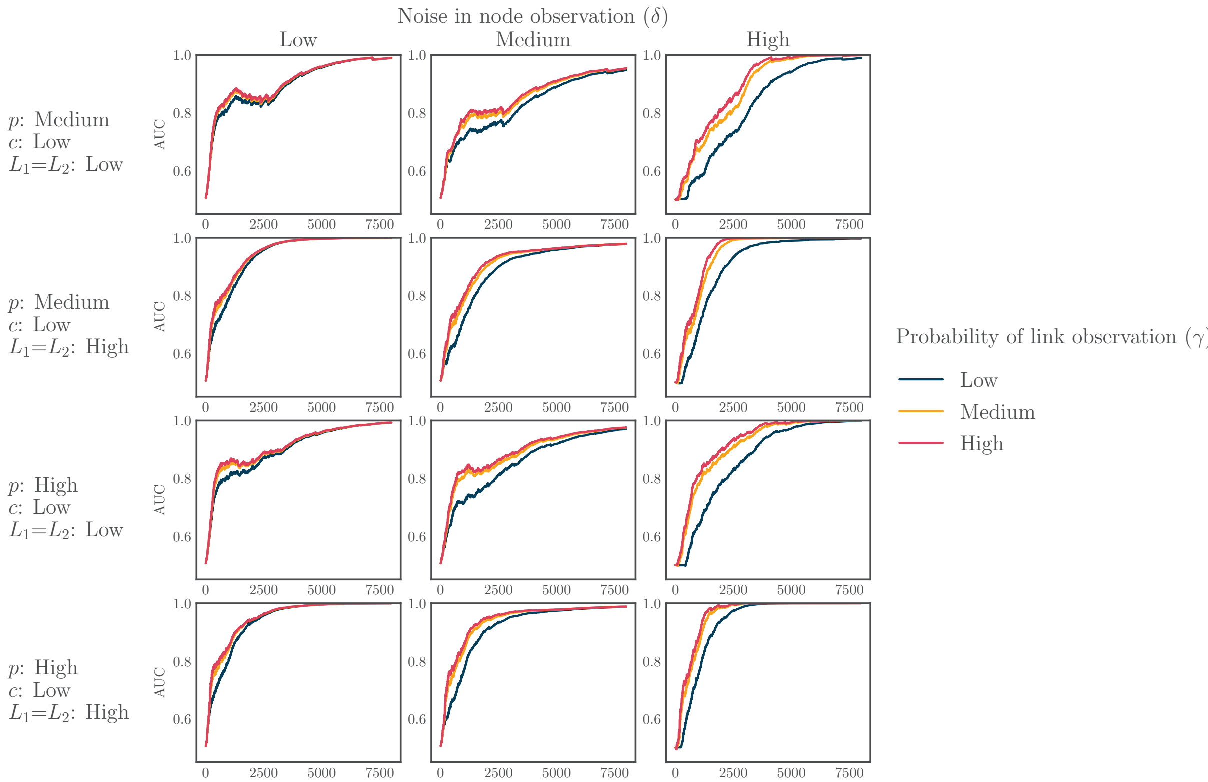

Figure 15: AUC at various time steps for γ (low, medium, high) over different δ and 40 Monte Carlo simulations (k: medium)

Sampling Frequency (k)

Changing sampling frequency from every 5 to every 10 time steps (which corresponds to an average of 16 and 8 observations per node) seems to have a similar effect on increasing time to convergence as changing it from every 10 to every 20 time steps (which corresponds to an average of 8 and 4 observations per node), just like in the case of no social influence. The only difference in the trend is due to the dynamic movement growth, which makes the convergence less monotonic than the no social influence case.

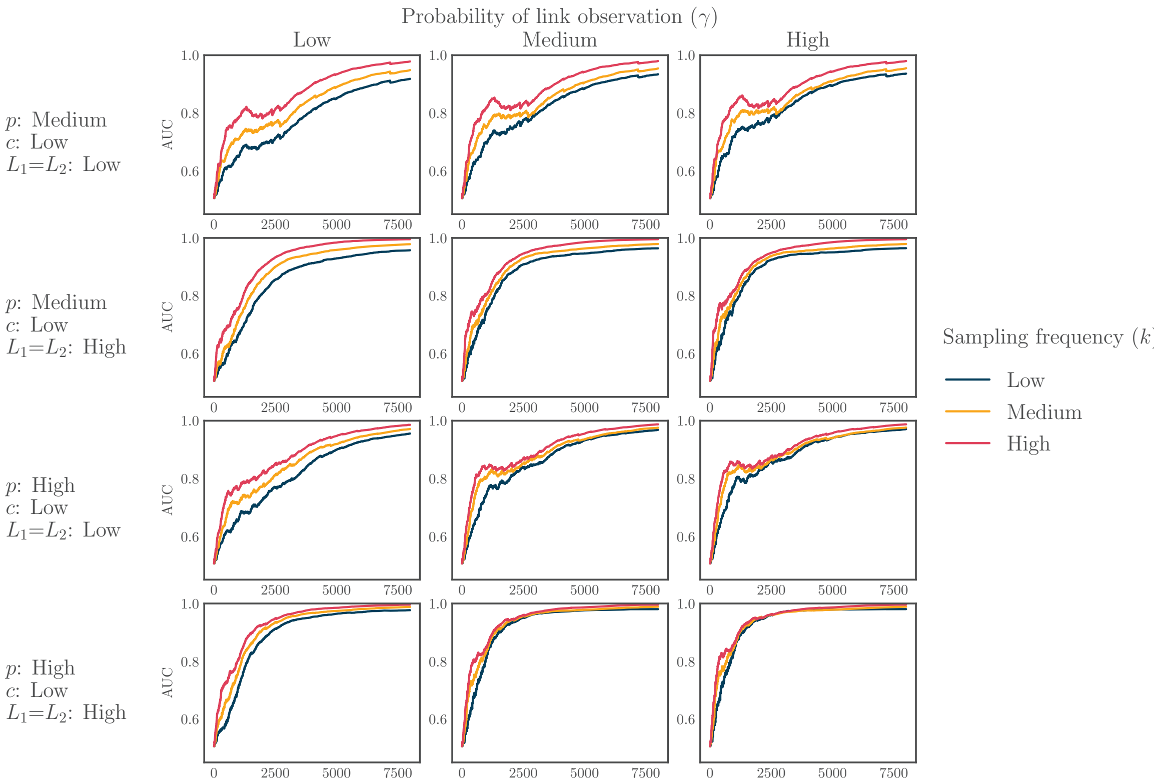

Figure 16: AUC at various time steps for k (low, medium, high) over different γ and 40 Monte Carlo simulations (δ: medium)

The results are summarized succinctly in figure 17. This figure shows gains to AUC over time for various changes in each of the three authority parameters. The results are similar to the case without social influence. The gains from improving technology consistently show diminishing returns (especially over longer time horizons) except in the case of reducing noise in the node observation where we see increasing returns (i.e., reducing the node observation noise from medium to low shows more improvement in AUC than reducing it from high to medium).

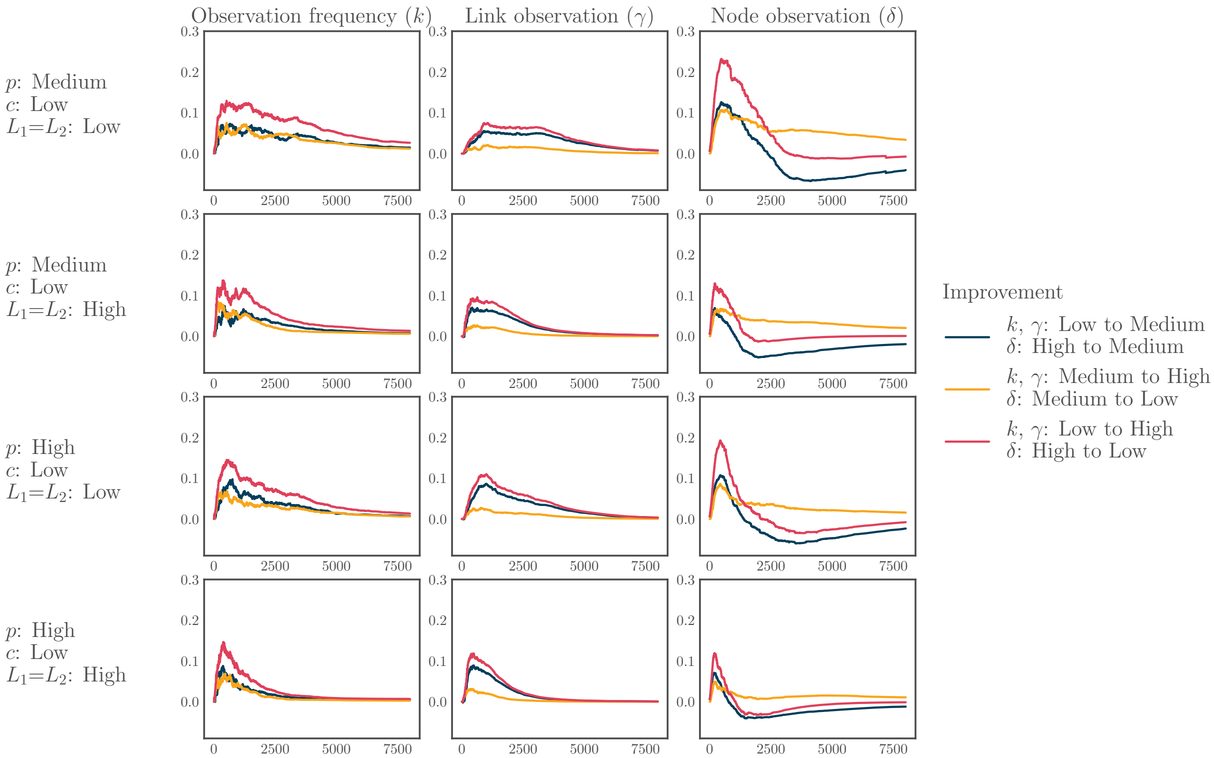

Figure 17: AUC improvement over time for improvements in k, γ, and δ

Summary and Future Work

Communication technologies are not intrinsically good or bad for civil society. They can be used by violent extremists to coordinate terrorist attacks or by advocates of just social causes to organize legitimate protests. Governments can use communication technologies to track and suppress activists or to track and neutralize terrorists. Rather than focusing on participants' or authorities' objectives, this work aims to understand better the dynamics of technological facilitation of both sides' goal, whatever they may be. One of the assumptions of our model is that participants need to recognize themselves as a group, build a network to support their actions, and recruit new supporters. We also assume that effective and efficient actions on the authority side require first recognizing who is a member of the protest and who is only a bystander. Consequently, we study the impact of communication technologies on the ability to perform those functions.

Our model yields some interesting results. First, the model reflects a common observation of sociologists and political scientists who try to predict how much support a given social movement might receive. Namely, depending on the underlying characteristics of the population, the final number of participants can significantly vary. Secondly, the model indicates that in the initial phase of social movement formation, technologies that enable casting a wide net accelerate both the social movement growth and the network formation more than those allowing elite, invitation-only membership. That is, open forums can lead to mobilization faster than closed chat rooms, even if members of those closed groups might be more direct about their goals. However, due to the snowball effect, the bigger the network, the bigger the utility of friend-of-friends networking technologies.

On the other hand, faster network growth enables better classification of agents by the authority. This is an artifact of the adopted link formation rule allowing participant to have more links than non-participants, making them stand out in the population. In real life, this property might be more difficult to observe as participants might deliberately limit their communication with others, precisely to stay below law enforcement radar as long as possible.

On the authority side, our model suggests that ensuring a reliable link observation process can result in reasonable accuracy even when the node observation process is noisy. The fact that our model allows the authority to perform better when the node observation process is noisier signals the importance of having a score function that correctly aggregates participant data. Further, we find that depending on the size of the social movement, the authority's belief formation is made more accurate by improving a different aspect of the surveillance process. Reducing the noise in node observation improves accuracy more effectively when there are fewer participants and lower outreach and networking. On the other hand, the improvements in accuracy made from increasing the sampling frequency are fairly independent of social movement characteristics.

As mentioned several times, one way to view our model is as a building block to more complex models of social movements that integrate participant and authority objectives. For that reason, an obvious extension is to integrate our network formation and surveillance model into a larger model of social movement. This would include adding participant and authority objectives and actions. For example, one research question might be the optimal time for which the participants decide to mobilize and act. From the authority's perspective, it might choose when to disrupt communication channels by either severing links or removing individuals they believe to be key participants (such as through incarceration).

It is also possible to build on our baseline model of participant dynamics. For example, in our model, citizens sample other citizens at random. In the real world, citizens will likely sample other citizens strategically in an effort to actively seek connections. Therefore, one extension to our model would be to incorporate more sophisticated networking and outreach behavior. Additionally, our model only allowed agents to switch from non-participants to participants and not switch in the other direction, which is, of course, possible in real social movements. This is also true of the authority; the authority may not randomly sample citizens to monitor but might strategically dedicate resources to monitoring key citizens. Furthermore, in the current model, the authority monitors links regardless of how they were formed. Strategic allocation of resources could also account for a difference in costs of monitoring open forums (outreach technologies) versus closed forums (networking technologies).

One hurdle in extending the model to include more complex and strategic behavior is determining how to modify the authority's belief process. While in this work we proposed a simple score that uses partial Bayesian updating, this score will likely be too imprecise to be relevant with more complex stochastic processes. However, the burgeoning fields of graph embeddings and graph neural networks [33] provide a promising path forward. Specifically, graph neural networks take as an input observed graph features (node features, links, etc.) and attempt to predict other features such as unobserved node characteristics. This is *precisely* what the authority wants to accomplish. Specifically, the authority observes noisy links and noisy node features and attempts to predict whether the node is a participant. Consequently, graph neural networks provide a promising direction to generalize the process in which the authority assigns scores as a function of its observations and thus allows the model to incorporate much more complex network formation dynamics.

References

[1] Charles J Stewart, Craig Allen Smith, and Robert E Denton Jr. *Persuasion and social movements*. Waveland Press, 2012.

[2] Jacqueline Tobin and Jacqueline L Tobin. *Hidden in plain view: The secret story of quilts and the Underground Railroad*. Anchor, 2000.

[3] Bonnie Calhoun. Shaping the public sphere: English coffeehouses and french salons and the age of the enlightenment. *Colgate Academic Review*, 3(1):7, 2012.

[4] Sum Lok-kei. 'hong kong reddit': how leaderless extradition protests took a lead from social media. *South China Morning Post*.

[5] Olesya Tkacheva. *Internet freedom and political space*. Rand Corporation, 2013.

[6] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[7] Wesley Shrum, Neil H Cheek Jr, and Saundra MacD. Friendship in school: Gender and racial homophily. *Sociology of Education*, pages 227–239, 1988.

[8] Martin Ruef, Howard E Aldrich, and Nancy M Carter. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American sociological review*, pages 195–222, 2003.

[9] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. *Journal of communication*, 64(2):317–332, 2014.

[10] Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[11] Juho Hamari and Jonna Koivisto. "working out for likes": An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, 50:333–347, 2015.

[12] Sinan Aral and Dylan Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science*, 57(9):1623–1639, 2011.

[13] Monica Di Gregorio. Networking in environmental movement organisation coalitions: interest, values or discourse? *Environmental Politics*, 21(1):1–25, 2012.

[14] Ersin Dincelli, Yuan Hong, and Nic DePaula. Information diffusion and opinion change during the gezi park protests: Homophily or social influence? *Proceedings of the Association for Information Science and Technology*, 53(1):1–5, 2016.

[15] Paul C Stern, Thomas Dietz, Troy Abel, Gregory A Guagnano, and Linda Kalof. A value-belief-norm theory of support for social movements: The case of environmentalism. *Human ecology review*, pages 81–97, 1999.

[16] Damon Centola. *How Behavior Spreads: The Science of Complex Contagions (Introduction)*. Princeton University Press, 2018.

[17] Charu Aggarwal and Karthik Subbian. Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1):1–36, 2014.

[18] Béla Bollobás and Oliver M Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, pages 1–34, 2003.

[19] Enghin Atalay, Ali Hortacsu, James Roberts, and Chad Syverson. Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202, 2011.

[20] Roger Guimera, Marta Sales-Pardo, and Luís A Nunes Amaral. Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2):025101, 2004.

[21] Kibae Kim and Jörn Altmann. Effect of homophily on network formation. *Communications in Nonlinear Science and Numerical Simulation*, 44:482–494, 2017.

[22] Claudio Cioffi-Revilla and Mark Rouleau. Mason rebeland: An agent-based model of politics, environment, and insurgency. *International Studies Review*, 12(1):31–52, 2010.

[23] Michael D Makowsky and Jared Rubin. An agent-based model of centralized institutions, social network technology, and revolution. *PloS one*, 8(11):e80380, 2013.

[24] Joshua M Epstein. Modeling civil violence: An agent-based computational approach. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7243–7250, 2002.

[25] Alessandro Moro. Understanding the dynamics of violent political revolutions in an agent-based framework. *PLOS one*, 11(4):e0154175, 2016.

[26] Carlos M Lemos. *Agent-based modeling of social conflict: From mechanisms to complex behavior*. Springer, 2017.

[27] Dana Downey. Convergence versus emergence of youth extremism: An agent-based model of the arab spring. In Paul A. Youngman and Mirsad Hadzikadic, editors, *Complexity and the Human Experience*, pages 183–202. Pan Stanford, Boca Raton, 2014.

[28] Carlos Lemos, Helder Coelho, Rui J Lopes, et al. Agent-based modeling of social conflict, civil violence and revolution: State-of-the-art-review and further prospects. In *EUMAS*, pages 124–138. Toulouse, 2013.

[29] Susanne Lohmann. The dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989–91. *World politics*, 47(1):42–101, 1994.

[30] Annie Waldherr and Nanda Wijermans. Modelling the role of social media at street protests. In *Advances in Social Simulation 2015*, pages 445–449. Springer, 2017.

[31] Hai-hua Hu, Wen-tian Cui, Jun Lin, and Yan-jun Qian. Icts, social connectivity, and collective action: A cultural-political perspective. *Journal of Artificial Societies and Social Simulation*, 17(2):7, 2014.

[32] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[33] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*, 2018.