INFERENCE OF INTERACTION KERNELS IN MEAN-FIELD MODELS OF OPINION DYNAMICS

WEIQI CHU*, QIN LI†, AND MASON A. PORTER‡

Abstract. In models of opinion dynamics, many parameters—either in the form of constants or in the form of functions—play a critical role in describing, calibrating, and forecasting how opinions change with time. When examining a model of opinion dynamics, it is beneficial to infer its parameters using empirical data. In this paper, we study an example of such an inference problem. We consider a mean-field bounded-confidence model with an unknown interaction kernel between individuals. This interaction kernel encodes how individuals with different opinions interact and affect each other's opinions. Because it is often difficult to quantitatively measure opinions as empirical data from observations or experiments, we assume that the available data takes the form of partial observations of a cumulative distribution function of opinions. We prove that certain measurements guarantee a precise and unique inference of the interaction kernel and propose a numerical method to reconstruct an interaction kernel from a limited number of data points. Our numerical results suggest that the error of the inferred interaction kernel decays exponentially as we strategically enlarge the data set.

Key words. opinion dynamics, inverse problems, kinetic equations

MSC codes. 91D30, 35R30, 45Q05, 65K10

1. Introduction.

The opinions and associated actions of the individuals in a population have major effects on financial markets [46], pandemic responses [7], climate change [54], and many other phenomena [5]. There are many investigations of opinion dynamics in both applied and theoretical contexts [28], and the mathematical modeling of opinions is prevalent in many disciplines, including sociology, economics, political science, mathematics, and physics [11, 16, 52].

Models of opinion dynamics, the spread of information, and other social phenomena can take the form of either deterministic or stochastic processes on networks [55]. In such models, the nodes of a network encode social entities (such as individual people) and the edges of a network encode interactions between entities [51]. In a model of opinion dynamics, each entity of a social network holds a time-dependent opinion. Such opinions are continuous-valued in some models and discrete-valued in others [52]. When entities interact with each other, they may adjust their opinions according to some update rule. The update rule and network structure jointly affect the steady-states and transient dynamics of opinion models. Relevant phenomena include how long it takes a system to converge to a steady state (and whether or not it does so) [48], whether or not the entities of a system eventually reach a consensus state [52], and how extremist opinions can take root in a system and affect the overall dynamics [2, 53].

Many studies of opinion models take a "forward" perspective and explore how model parameters and network structure affect their dynamics [43, 47, 56]. Although it is valuable to study mathematical models of opinion dynamics, one cannot rely on models alone. To make viable forecasts and sufficiently match empirical data, it is desirable to use parameter values that one obtains from real-life measurements. Unfortunately, it is difficult to directly measure parameters of opinion models from empirical observations in a trustworthy way [10, 27]. It is also difficult to justify

*Department of Mathematics, University of California, Los Angeles

†Department of Mathematics, University of Wisconsin, Madison

‡Department of Mathematics, University of California, Los Angeles; Department of Sociology, University of California, Los Angeles; and Santa Fe Institute

precise choices of the functional forms of the mathematical terms in such models [25]. Individuals in a population interact with each other in complex and time-dependent ways, and the opinions of individuals and groups can change drastically from both endogenous evolution and exogenous events. Such multifaceted complexities make it difficult to quantify opinions [37] and measure parameter values in a scientifically rigorous way. Consequently, it is useful to develop approaches from the perspective of inverse problems to infer the parameters of opinion models from data observations. Such efforts can help advance methods of validating opinion models and forecasting future dynamics from past observations.

Prior research on models of interacting agents has employed parameter inference in a variety of settings, including disease spread [22, 60], election forecasts [64], opinion dynamics [25,49], and dynamics that are governed by distance-based interactions [42,45]. Such works often formulate an inference problem using a regression framework, in which one seeks to determine optimal parameter values for a model to produce output data that is as close as possible to observed data [1]. One can also view this process as a numerical execution of "inversion" to tune parameters to characterize the true properties of a dynamical process [23, 33]. A fundamental topic that ties closely with such numerical-inversion procedures is whether or not parameters are identifiable from the provided data. To do this, it is necessary to relate empirical data to model parameters. In the present paper, we examine the following two questions:

(1) What type of data uniquely reconstructs parameters in a model of opinion dynamics?

(2) How much data does one need for such a reconstruction?

The answers to these questions depend both on the opinion model itself and on the employed numerical-inversion procedure [36, 62]. Researchers have studied these questions in the context of optical tomography [3], seismology [15], geophysics [59], and many other applications. In our paper, we examine the above questions for a *bounded-confidence models* (BCM) of opinion dynamics [9, 43].

In a BCM, the opinions of the entities in a population take continuous values, which perhaps represent a range of opinions from very liberal to very conservative on a one-dimensional political spectrum. The interactions between these entities encapsulate the idea of "selective exposure" from psychology [26, 57]. When two or more entities interact, they compromise their opinions by some amount if and only if their current opinions are sufficiently close to each other. One can measure such closeness with a scalar "confidence bound". Entities that interact compromise their opinions to some extent if and only if the difference between their current opinions is smaller than the confidence bound; otherwise, in most BCMs, their opinions remain unchanged after an interaction [9, 43].

The idea of a confidence bound appears both in agent-based BCMs and in density-based BCMs [43]. Agent-based models and density-based models describe opinion dynamics at different scales, and they serve complementary purposes. Agent-based BCMs characterize fine-grained interactions and are useful for examining the opinion trajectories of a discrete number of agents. They are helpful for obtaining insights into the influence of network architecture on dynamics, such as the qualitative characteristics of steady states (e.g., consensus versus polarization versus fragmentation) and the convergence time to attain a steady state [48]. There is a large body of research on elaborating agent-based BCMs with various features, such as by incorporating heterogeneous confidence bounds [17], media nodes [13], and other extensions. Density-based BCMs take a macroscale perspective and examine the evolution of the opinions of infinitely many agents using population densities [8]. Density-based BCMs describe the

macroscale collective behavior of agents, so they are useful for studying ensemble and average effects. In some situations, one can derive density-based BCMs as mean-field limits of associated agent-based BCMs [20]. Such descriptions arise both in opinion dynamics [4] and in many other models of the collective behavior of a large number of agents [66], such as in bird flocking and fish schooling [29].

In the present paper, we examine a density-based model of opinion dynamics. Such a model describes the time evolution of an opinion density using a kinetic equation. We seek to develop a mathematically rigorous approach to infer the interaction kernel θ of our model. During the past decade, there has been considerable theoretical and computational progress in the study of inverse kinetic theory [63]. Key prior work in this area has considered parameter inference in the classical Boltzmann equation [38,40] and the radiative-transfer equation [6,19,31,41]. Inference results in inverse kinetic theory depend heavily on the particular form of a kinetic equation, and it is thus important to examine a variety of models.

In prior investigations, researchers have examined the inference of interaction kernels in agent-based models using nonparametric methods [44,45] and maximum-likelihood methods [39]. Such models have finitely many agents and often take the form of a stochastic process or a system of ordinary differential equations. To obtain a specified accuracy, both the number of data measurements and the computational cost typically increase with the number of agents [14,45,65]. In the present paper, we consider a density-based model, which gives a mean-field description when there are infinitely many agents. Therefore, our inference procedure is in a regime that was not considered in [39,44,45].

Our paper proceeds as follows. In section 2, we present the mean-field opinion model and set up an inverse problem to infer the interaction kernel between the agents in a population. In section 3, we state and prove two theorems that give theoretical guarantees in the form of sufficient conditions for the data to uniquely and precisely reconstruct the interaction kernel. In section 4, we propose a numerical method to infer the interaction kernel using a differential-equation-constrained optimization framework. In section 5, we develop an adaptive optimization algorithm to accelerate the convergence of our numerical method. In section 6, we conclude and discuss future work. In Appendix A, we derive an adjoint problem, which we use to obtain an explicit formula for an associated Fréchet derivative. In Appendix B, we prove Theorem 4.1.

2. Inverse-problem setup. In this section, we present a mean-field model of opinion dynamics [8] and propose a sensible way to measure data for opinion dynamics.

2.1. A mean-field model of opinion dynamics. We consider a density-based BCM [8,20] whose governing equation is the Kac-type integro-differential equation

$$\partial_t f_\theta(x,t) = \int_{\Omega\times\Omega} \theta(x_1-x_2) f_\theta(x_1,t) f_\theta(x_2,t) \big[2\delta\big(x-\tfrac{x_1+x_2}{2}\big) - \delta(x-x_1) - \delta(x-x_2)\big]\, dx_1\, dx_2\,, \tag{2.1}$$

where $f_\theta(x,t)$ is the probability density of agents with continuous-valued opinion $x \in \Omega$, the set $\Omega \subset \mathbb{R}$ is the space of possible opinions of an agent, $\theta(\cdot)$ is an interaction kernel, and $\delta(\cdot)$ is the Dirac delta function. We use the subscript in f_θ to indicate explicitly that the solution depends on the interaction kernel θ. In (2.1), two agents with opinions x_1 and x_2 interact with each other with a probability that is proportional to $\theta(x_1 - x_2) f(x_1,t) f(x_2,t)$. After this interaction, the two agents compromise with each other and change their opinions to their mean opinion $\frac{x_1+x_2}{2}$. This process leads to a "gain term" in (2.1) at $x = \frac{x_1+x_2}{2}$ from the post-interaction opinions and "loss terms" at $x = x_1$ and $x = x_2$ from the pre-interaction opinions. One can extend the model (2.1)

to situations in which the post-interaction opinions have more complicated forms [61], instead of only considering compromises to the mean opinion of two interacting agents. One can also examine generalizations of (2.1) that incorporate interactions between three or more entities [20].

The interaction kernel $\theta(x_1 - x_2)$ encodes the probability that two agents, with opinions x_1 and x_2, interact with each other. It can take various forms, depending on the specific BCM. For example, in the classic Deffuant–Weisbuch BCM [24], interacting agents exchange their opinions if their opinion difference is less than a constant c. The interaction kernel thus takes the form of the indicator function [8] $\theta(r) = \mathbb{1}_{(-c,c)}(r)$, which is parameterized by a constant confidence bound c. By contrast, Sîrbu et al. [58] examined a BCM that favors edges between nodes whose opinions are very close to each other. They implemented this favoritism using an interaction kernel with a power-law decay. The interaction kernel θ plays a critical role in determining the dynamics of a BCM or other model of collective behavior. For example, Motsch and Tadmor showed that heterophilous interactions promote convergence to consensus in models of interacting agents that adjust to environmental averages [50]. To ensure the well-posedness of (2.1), we assume that the interaction kernel is nonnegative [20]. For simplicity, we also assume that $\theta(r)$ is symmetric around $r = 0$.

A natural question is the following: Given the time series of opinions of all agents, can we infer the opinion-update rule that governs how agents interact and compromise with each other? In the context of the BCM (2.1), this question amounts to inferring the interaction kernel θ from observations of the probability density $f(x, t)$. This problem thus falls into the framework of inverse problems of kinetic equations.

2.2. Accessibility of data. It is difficult to directly measure people's opinions as continuous values [37]. Instead of assuming that we have a direct measurement of opinions as a function of time, we suppose that the available data takes the form of aggregated values from a distribution of opinions. Each data point is the cumulative probability density up to a value a at time t. That is, each data point takes the form

$$M_\theta(a, t; f_0) = \int_{-\infty}^{a} f_\theta(x, t)\, dx\,, \tag{2.2}$$

where f_θ is the opinion-distribution density, which is governed by the mean-field opinion model (2.1), and f_0 is its associated initial density. We use the subscript θ to emphasize the dependence on the interaction kernel θ.

We refer to a as the *measurement threshold*, which we assume is a known quantity that we are given along with the data. Consider a situation in which people vote in a binary way, such as by choosing "No" or "Yes". We assume that a vote for "No" results from an underlying continuous opinion value that lies in the interval $(-\infty, a]$ and that a vote for "Yes" results from a continuous opinion value in the interval $(a, +\infty)$. In this example, M_θ is the fraction of a population that votes "No". The quantity $M_\theta(a, t; f_0)$ is a function of the measurement threshold a, the time t, and the initial opinion density f_0. It is impractical to assume that M_θ is accessible on the entire domain for all a, t, and f_0. To cope with this reality, we suppose that we only have access to data at certain values of a, t, and f_0.

3. Theoretical guarantees in two scenarios. We consider two specific scenarios: either M_θ is measured at a fixed measurement threshold a or it is measured for a fixed initial opinion distribution f_0. Both scenarios correspond to certain simplistic real-life situations. For each scenario, we prove that the associated inverse problem is well-defined in the sense that the data uniquely determines the interaction kernel θ.

3.1. Data measured at a fixed measurement threshold a. Suppose that we fix the measurement threshold $a = a_0$ and allow the initial opinion distribution f_0 to vary. We define

$$m_\theta(t; f_0) = M_\theta(a_0, t; f_0)\,, \tag{3.1}$$

which is the fraction of opinions that are less than or equal to the threshold a_0. This scenario models a situation in which the collected data has a binary nature. This situation occurs frequently in political systems, such as whether the United Kingdom should remain part of the European Union or leave the European Union in the 2016 "Brexit" referendum. Another example is a presidential election with two candidates.

We define a functional $\mu_\theta : \mathcal{L}^1(\Omega) \to \mathbb{R}$ that maps the initial opinion distribution f_0 to a real number. This functional, which is given by

$$\mu_\theta[f_0] = \partial_t m_\theta(t = 0; f_0)\,, \tag{3.2}$$

evaluates the rate of change of the "left-wing" opinion fraction m_θ at the initial time $t = 0$, given the initial opinion distribution f_0. In this scenario, we always fix the measurement threshold a to the value a_0. One key result of our paper is that the measurement μ_θ is sufficient to infer the interaction kernel θ in Equation (2.1). In particular, if we know the functional values for the entire function space $\mathcal{L}^1(\Omega)$ (i.e., $\mu_\theta[f_0]$ is available for all $f_0 \in \mathcal{L}^1(\Omega)$), then we have sufficient data to uniquely determine θ. We state the result in Theorem 3.1, which we prove in section 3.3.

THEOREM 3.1. *Let θ be an interaction kernel, and let μ_θ be the functional* (3.2)*. There is a one-to-one correspondence between θ and μ_θ.*

Theorem 3.1 implies that different interaction kernels θ_1 and θ_2 (with $\theta_1 \neq \theta_2$) induce different functionals μ_{θ_1} and μ_{θ_2}. Namely, there exists at least one initial opinion distribution f_0 such that values of $\mu_{\theta_1}[f_0]$ and $\mu_{\theta_2}[f_0]$ are not equal. This implies that a measurement that includes $\mu_\theta[f_0]$ for all $f_0 \in \mathcal{L}^1(\Omega)$ yields a data set that is sufficient to uniquely reconstruct an interaction kernel.

At first sight, the required data seems to be rather substantial because we need to evaluate the functional $\mu_\theta[f_0]$ for all initial opinion distributions f_0. However, this strict data requirement is unavoidable to ensure the unique reconstruction of θ. The interaction kernel $\theta(r)$ is itself a function, so inferring it requires one to deal with an infinite-dimensional function space. The observed data needs to be sufficiently abundant to do this. Because the measurement threshold a is fixed at a single value a_0, we require flexibility in the initial opinion distribution f_0. For example, if the initial opinion distribution $f_0(x)$ is a Dirac delta function, then all nonnegative symmetric kernels yield the same solution $f(x, t) = f_0(x)$. In this case, no measurement is able to identify the kernel θ with this initial opinion distribution. Another reconstruction failure occurs when the initial opinion distribution $f_0(x)$ is symmetric about $x = a_0$. By symmetry, $m(t; f_0) = 0.5$ for all interaction kernels (i.e., all nonnegative and symmetric functions), so it is not possible to identify the true interaction kernel. In the proof of Theorem 3.1, we show that one can slightly relax the stated conditions. Instead of requiring $\mu_\theta[f_0]$ for all f_0, we only need $\mu_\theta[f_0]$ for a basis of the $\mathcal{L}^1(\Omega)$ function space.

In practice — both in reality and in our computations — we can work with data that consists of a finite number of data points, rather than infinitely many of them. Even with only a limited number of data points, we can still numerically reconstruct the interaction kernel θ. For example, in section 5, we consider a data set that includes

data from 4 different initial conditions. In this example, our numerical inference method is able to reconstruct the interaction kernel up to an accuracy of 10^{-4} within 1000 iterations.

3.2. Data measured for a fixed initial opinion distribution f_0. In our second scenario, we fix the initial opinion distribution f_0 but have data at multiple measurement thresholds. That is, we fix f_0 in (2.2) and measure M_θ for different values of a. In such a scenario, individuals' opinions about a specified topic take discrete values, which may correspond to satisfaction levels or happiness levels [30]. Some surveys also ask participants to rate their views on some scale (e.g., using a Likert scale [34], which is a common psychometric scale), such as by choosing an integer between 1 and 10. If one obtains the initial opinion distributions f_0 by shifting the same opinion distribution, the second scenario becomes equivalent to the first scenario (see section 3.1).

Recall that $M_\theta(a, t; f_0)$ [see (2.2)] indicates the fraction of a population whose opinion has a value of at most a. We define a function $\nu_\theta : \Omega \to \mathbb{R}$ that depends on the measurement threshold a and ν_θ. This function takes the form

$$\nu_\theta(a) = \partial_t M_\theta(a, t = 0; f_0)\,. \tag{3.3}$$

We now present our second main result, which states that the measurement ν_θ is sufficient to uniquely reconstruct θ. In particular, if we know the function values $\nu_\theta(a)$ for all $a \in \Omega$, we can precisely reconstruct the interaction kernel θ. We state the result in Theorem 3.2, which we prove in section 3.3.

THEOREM 3.2. *Suppose that the interaction kernel $\theta(r)$ is compactly supported on the interval $(-B, B)$ and that the initial opinion distribution f_0 that we use to define ν_θ in (3.3) is uniform (i.e., $f_0(x) = \frac{1}{2B}$ if $x \in (-B, B)$ and $f_0(x) = 0$ otherwise). There is then a one-to-one correspondence between θ and ν_θ.*

Theorem 3.2 implies that for any interaction kernel θ_1 that differs from the true kernel θ, its induced functional ν_{θ_1} is also different from ν_θ at least at one point (i.e., $\nu_{\theta_1}(a_0) \neq \nu_\theta(a_0)$ for some a_0). Therefore, the data set is sufficient to uniquely identify the true interaction kernel if the data set includes a measurement for all values of a.

Theorem 3.2 states that if we measure opinions at sufficiently finely-grained opinion levels, then a single initial opinion distribution yields enough data to uniquely reconstruct the interaction kernel. Because the interaction kernel θ is a function, inferring it precisely requires measurements of different a in a continuous manner, which yields infinitely many data points. As in the scenario in section 3.1, we only possess a discrete version of the data in practice. In other words, instead of knowing $M_\theta(a, t; f_0)$ for all a in a continuous interval, we only have a data set $\{M(a, t; f_0) :\ t \in \mathcal{T},\ a \in \mathcal{A}\}$, where $\mathcal{T}$ and $\mathcal{A}$ are finite sets (see section 4). As we enlarge the measurement-threshold set $\mathcal{A}$, we can reconstruct the interaction kernel with better accuracy.

3.3. Proofs of Theorems 3.1 and 3.2. We now prove Theorems 3.1 and 3.2. To ensure mathematical rigor, we first need to justify that the functional μ_θ in Equation (3.2) and the function ν_θ in Equation (3.3) are well-defined. From results in [20], we know that Equation (2.1) is well-posed. In particular, for any nonnegative and symmetric $\theta(r)$ and any density function $f_0(x) \in \mathcal{L}^1(\Omega)$, there exists a unique solution $f_\theta(x, t)$ of Equation (2.1), with initial opinion distribution $f(x, 0) = f_0(x)$, such that $f_\theta(x, t)$ is differentiable with respect to time t and $\int_\Omega f_\theta(x, t)\, dx = 1$ for all $t \geq 0$. This implies that (1) the value of the functional $\mu_\theta[f_0]$ is well-defined for all

interaction kernels $\theta(r)$ and all initial opinion distributions $f_0 \in \mathcal{L}^1(\Omega)$ and that (2) the value $\nu_\theta(a)$ is well-defined for all interaction kernels $\theta(r)$ and all $a \in \Omega$.

We now prove Theorem 3.1, which guarantees that we can uniquely reconstruct an interaction kernel from data that we measure at a single measurement threshold.

Proof of Theorem 3.1. A direct computation yields

$$\mu_\theta[f_0] = \partial_t m_\theta(0; f_0) = 2\int_0^\infty \theta(y) k(y; f_0)\, dy\,, \tag{3.4}$$

where

$$k(y; f_0) = \int_0^{y/2} \left[f_0(a-y+x) f_0(a+x) - f_0(a-y-x) f_0(a-x)\right]\, dx \tag{3.5}$$

is the integral kernel. We choose the initial opinion distribution f_0 to be a sum of indicator functions. For any c and $w > 0$, we construct the initial opinion distribution

$$f_0(x) = \frac{1}{2w}\left[\mathbb{1}_{(a-c-w,a-c)}(x) + \mathbb{1}_{(a+c-w,a+c)}(x)\right]\,, \tag{3.6}$$

where $\mathbb{1}_{(l,r)}$ is the indicator function on the interval (l, r). The amplitude $1/(2w)$ ensures that f_0 is normalized. For this specific type of f_0, we obtain

$$k(y; f_0) = \frac{4}{w} h_{2c,w}(y)\,, \tag{3.7}$$

where $h_{2c,w}(y)$ is the triangular hat function that is centered at $y = 2c$ and has a base of width $2w$. That is,

$$h_{2c,w}(y) = \max\left\{1 - |(y-2c)/w|\,, 0\right\}\,, \tag{3.8}$$

where $\max\{\cdot,\cdot\}$ denotes the larger number of its two arguments.

By considering sufficiently many values of c and w, the set of triangular hat functions $h_{2c,w}(y)$ gives a basis of the $\mathcal{L}^1(\mathbb{R}_+)$ function space. Therefore, given two functions θ_1 and θ_2, there must exist a pair c, w such that

$$\int_0^\infty \left[\theta_1(y) - \theta_2(y)\right] h_{2c,w}(y)\, dy \neq 0\,. \tag{3.9}$$

Using Equation (3.4), we know that $\partial_t m_{\theta_1}(0; f_0) \neq \partial_t m_{\theta_2}(0; f_0)$ when $f_0(x)$ takes the form in Equation (3.6). This implies that $\mu_{\theta_1}[f_0] \neq \mu_{\theta_2}[f_0]$ and thus that $\mu_{\theta_1} \neq \mu_{\theta_2}$. Because we can choose θ_1 and θ_2 arbitrarily, we know that the correspondence between θ and μ_θ is one-to-one. ∎

We now prove Theorem 3.2, which guarantees that we can uniquely reconstruct an interaction kernel from data that we measure for a fixed initial opinion distribution.

Proof of Theorem 3.2. A direct computation yields the Fredholm integral

$$\partial_a \nu_\theta(a) = \frac{1}{2B^2}\int_0^B \theta(y)\, G(a, y)\, dy\,, \tag{3.10}$$

where

$$G(a, y) = 4B^2\left[2 f_0\Big(a - \frac{y}{2}\Big) f_0\Big(a + \frac{y}{2}\Big) - f_0(a)\, f_0(a+y) - f_0(a)\, f_0(a-y)\right] \tag{3.11}$$

is the integral kernel. Recall that f_0 is a uniform function on $(-B, B)$. A direct computation yields

$$G(a,\cdot) = \begin{cases} \mathbb{1}_{(B-a,B)}(\cdot)\,, & a \in (0, B/2] \\ \mathbb{1}_{(B-a,2(B-a))}(\cdot) - \mathbb{1}_{(2(B-a),B)}(\cdot)\,, & a \in (B/2, B)\,. \end{cases} \tag{3.12}$$

We claim that $\mathbb{1}_{(c,B)}(\cdot)$ is a linear combination of $G(a,\cdot)$ with different values of a. We prove this claim for the two possible cases: (1) $c \in [B/2, B)$ and (2) $c \in (0, B/2)$.

For case (1), we have

$$\mathbb{1}_{(c,B)}(\cdot) = G(B-c,\cdot) \text{ for all } c \in [B/2, B) \tag{3.13}$$

directly from (3.12). We prove case (2) by induction. For any $c \in (0, B/2)$, we have

$$\mathbb{1}_{(c,B)}(\cdot) = \mathbb{1}_{(c,2c]}(\cdot) + \mathbb{1}_{(2c,B)}(\cdot) = G(B-c,\cdot) + 2\mathbb{1}_{(2c,B)}(\cdot)\,. \tag{3.14}$$

When $c \in [B/4, B/2)$, a direct computation from (3.14) yields

$$\mathbb{1}_{(c,B)}(\cdot) = G(B-c,\cdot) + 2G(B-2c,\cdot)\,, \tag{3.15}$$

which implies that $\mathbb{1}_{(c,B)}(\cdot)$ is a linear combination of $G(a,\cdot)$ when $c \in [B/4, B/2)$. Suppose that $\mathbb{1}_{(c,B)}(\cdot)$ is a linear combination of $G(a,\cdot)$ for $c \in [B/2^k, B/2^{k-1})$. From (3.14), we know that $\mathbb{1}_{(c,B)}(\cdot)$ is a linear combination of $G(a,\cdot)$ for $c \in [B/2^{k+1}, B/2^k)$. By induction, $\mathbb{1}_{(c,B)}(\cdot)$ is a linear combination of $G(a,\cdot)$ for $c \in (0, B/2)$.

The above two cases imply that the set $\{G(a,\cdot)\}_a$ is a basis of the function space $\mathcal{L}^1(0,B)$. For any two distinct interaction kernels θ_1 and θ_2, there exists at least one basis function of $\mathcal{L}^1(0,B)$ (e.g., $G(a^*,\cdot)$) such that

$$\int_0^B [\theta_1(y) - \theta_2(y)]\, G(a^*, y)\, dy \neq 0\,. \tag{3.16}$$

Equation (3.10) implies that $\partial_a \nu_{\theta_1}(a^*) \neq \partial_a \nu_{\theta_2}(a^*)$, which in turn implies that $\nu_{\theta_1} \neq \nu_{\theta_2}$. This concludes the proof. ∎

We used the same strategy to prove Theorems 3.1 and 3.2. In this approach, one rewrites the proposed measurement as a Fredholm integral (see Equations (3.4) and (3.10)) and proves that the Fredholm kernel spans the function space as one varies one of the variables. When it does, the reconstruction is unique in the dual space. Researchers use this type of strategy for many inverse problems [36, 62], such as linearized Calderón problems and linearized inverse-scattering problems.

In the present paper, we only consider strategies to select either the initial opinion distribution f_0 or the measurement threshold a for which the generated data yields a unique inferred interaction kernel θ in (2.1). However, there are other ways that one can design data measurement to ensure the uniqueness of the inferred interaction kernel θ. For example, one can consider different strategies to generate data at specific times. The uniqueness of the inferred interaction kernel is necessary for the well-posedness of the associated inverse problem (which we introduced in section 2) and is also fundamental to the development of numerical inference methods.

3.4. Inference stability. Theorem 3.1 guarantees that we can uniquely reconstruct an interaction kernel θ from data that is measured at a fixed measurement threshold. We now discuss the stability of this inference problem. We assume that the

derivatives of θ and f_0 are sufficiently smooth (i.e., their derivatives of all orders are continuous and bounded).

To examine the stability of $\mu_\theta[f_0]$ with respect to θ, we need to find a function F such that

$$\|\theta_1 - \theta_2\| \leq F(\|\mu_{\theta_1} - \mu_{\theta_2}\|) \tag{3.17}$$

when θ and μ_θ are equipped with proper norms. If $F(\cdot)$ is linear and $F(0) = 0$, then the reconstruction of θ is Lipschitz stable for the selected norms.

We define the infinity norm of μ by

$$\|\mu_{\theta_1} - \mu_{\theta_2}\| = \max_{c,w} \int (\theta_1 - \theta_2)(y) h_{2c,w}(y) dy \tag{3.18}$$

and equip θ with the standard Banach-space L_∞ norm. Suppose that $\|\theta_1 - \theta_2\|_\infty = \tau > 0$. Because θ is continuous, we know that there is a point y_0 such that $|\theta_1(y_0) - \theta_2(y_0)| = \tau$. Additionally, if θ is differentiable (or smoother), there exists a small neighborhood $(y_0 - \epsilon, y_0 + \epsilon)$ in which $|\theta_1(y) - \theta_2(y)| > \tau/2$ for all $y \in (y_0 - \epsilon, y_0 + \epsilon)$. The size of ϵ depends on the smoothness of θ; a smoother θ yields a larger ϵ. Setting $2c_0 = y_0$ and $w_0 = \epsilon$, we define a function $h_{2c_0,w_0}(y)$ using (3.8). We insert $h_{2c_0,w_0}(y)$ into (3.18) to obtain

$$\|\mu_{\theta-1} - \mu_{\theta_2}\| \geq \int (\theta_1 - \theta_2)(y) h_{2c_0,w_0}(y) dy > \tau \cdot \epsilon/2 = \frac{\epsilon}{2} \|\theta_1 - \theta_2\|_\infty . \tag{3.19}$$

Accordingly, we choose $F(\cdot) = 2/\epsilon$ in (3.17) and obtain the Lipschitz constant $2/\epsilon$ in (3.17).

4. A numerical inference method. Our theoretical results in section 3 identify forms of data that guarantee a unique reconstruction of the interaction kernel. Theorems 3.1 and 3.2 require knowledge of the time derivatives $\partial_t M_\theta(a, t = 0; f_0)$ for either all measurement threshold values a or all initial opinion distributions f_0. Additionally, the data and the to-be-reconstructed interaction kernel both take a continuous form, so they live in infinite-dimensional spaces. In practice, however, the time derivatives $\partial_t M_\theta$ are often inaccessible and both the data sets and the to-be-reconstructed interaction kernel are finite-dimensional. In this section, we develop a numerical method to infer the interaction kernel of the mean-field opinion model (2.1) in this discrete setting.

Suppose that a data set takes the form

$$\mathcal{D} = \left\{ M^*(a, t; f_0) : a \in \mathcal{A}\,,\ t \in \mathcal{T}\,,\ f_0 \in \mathcal{F} \right\}, \tag{4.1}$$

where $M^*(a, t; f_0)$ is the measurement (2.2) that is generated by the true interaction kernel θ^* (which we specify in our numerical computations) and $\mathcal{A}$, $\mathcal{T}$, and $\mathcal{F}$ are finite collections of discrete values of a, t, and f_0, respectively. We use the notation $|\cdot|$ to represent the cardinality of a set.

4.1. Loss-function formulation. We consider the ℓ_2 error between simulated data and measured data. This yields a loss function

$$\mathrm{L}(\theta) = \frac{1}{|\mathcal{A}||\mathcal{T}||\mathcal{F}|} \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} \sum_{f_0 \in \mathcal{F}} \frac{1}{2} \left[M_\theta(a, t; f_0) - M^*(a, t; f_0) \right]^2 . \tag{4.2}$$

In our numerical inference, we seek an interaction kernel $\hat{\theta}$ (i.e., a nonnegative and symmetric function) that minimizes the loss function $\mathrm{L}(\theta)$. That is, we seek to

determine

$$\hat{\theta} = \arg\min_{\theta\in\Phi} \mathrm{L}(\theta)\,, \tag{4.3}$$

where $\Phi = \{\theta(r) \in \mathcal{L}^\infty :\ \theta(r) \geq 0\,,\ \theta(-r) = \theta(r)\}$ is the admissible set. If there are any other restrictions on θ (such as requiring that θ is compactly supported on an interval), we place the minimization over an admissible set as constraints to account for them. To reduce overfitting, one can also add a regularization term of θ to the loss function in (4.2).

There are many strategies to minimize the loss function (4.2). Roughly speaking, existing methods are either gradient-based or Hessian-based. In theory, Hessian-based methods have higher-order convergence rates (so one may expect them to be faster, in principle) than gradient-based methods, but they require the computation of second-order functional derivatives and are thus computationally prohibitive. Therefore, we use a gradient-based minimization algorithm; see section 4.3 for details. Such an algorithm typically involves an iterative update

$$\theta_{n+1} = \theta_n - \alpha_n r_n\,, \quad r_n \approx \partial_\theta \mathrm{L}(\theta)|_{\theta=\theta_n}\,, \tag{4.4}$$

where $\partial_\theta \mathrm{L}(\theta)$ is the Fréchet derivative with respect to θ. We adjust the step size α_n at each iteration to expedite convergence and ensure that θ_{n+1} is in the admissible set. The descent direction r_n is a modification of $\partial_\theta \mathrm{L}(\theta)$ that ensures that the update does not leave the admissible set. In other words, it does not violate the nonnegativity constraint and remains symmetric. We compute

$$\partial_\theta \mathrm{L}(\theta) = \frac{1}{|\mathcal{A}||\mathcal{T}||\mathcal{F}|} \sum_{a\in\mathcal{A}} \sum_{t\in\mathcal{T}} \sum_{f_0\in\mathcal{F}} \left(M_\theta(a,t;f_0) - M^*(a,t;f_0)\right) \partial_\theta M_\theta(a,t;f_0)\,. \tag{4.5}$$

The core of the computation lies in evaluating the Fréchet derivative $\partial_\theta M_\theta$. This gives a function in the updating direction that lies in the same function space as $\theta(r)$. When the context is clear, we omit the dependence on a, t, and f_0 in our notation and write $\partial_\theta M_\theta(r)$ as a function of only $r \in \Omega_\theta$. In section 4.2, we derive the formula for $\partial_\theta M_\theta$.

4.2. Computation of the Fréchet derivative. To compute the Fréchet derivative in Equation (4.5), we view M_θ as a functional that maps the function θ to a real number. The Fréchet derivative measures the rate of change of M_θ when one perturbs the function θ. By using variational calculus, we show that evaluating $\partial_\theta M_\theta$ amounts to solving two integro-differential equations with specifically designed initial and final conditions. The former is the forward problem, and the latter is the adjoint problem. The forward problem is (2.1), which we rewrite as

$$\begin{cases} \partial_t f_\theta(x,t) = \displaystyle\int_{\Omega\times\Omega} \theta(x_1 - x_2) f_\theta(x_1,t) f_\theta(x_2,t) F(x,x_1,x_2)\ dx_1\, dx_2 \\ f_\theta(x,0) = f_0(x)\,, \end{cases} \tag{4.6}$$

where $F(x,x_1,x_2) = 2\delta\left(x - \frac{x_1+x_2}{2}\right) - \delta(x-x_1) - \delta(x-x_2)$ and $f_0(x) \in \mathcal{L}^1(\Omega)$ is the initial opinion distribution. The adjoint problem is

$$\partial_\tau g_\theta(x,\tau) = -\mathcal{L}^*_\theta[g_\theta](x,\tau)\,, \quad g_\theta(x,t) = \mathbb{1}_{(-\infty,a]}(x)\,, \tag{4.7}$$

where the operator $\mathcal{L}^*_\theta$ is the adjoint operator of the integral term in (4.6). That is,

$$\mathcal{L}^*_\theta[g_\theta](x,\tau) = \int_\Omega 2 f_\theta(y,\tau)\theta(x-y)\left[2g_\theta\left(\tfrac{x+y}{2},\tau\right) - g_\theta(x,\tau) - g_\theta(y,\tau)\right] dy\,. \tag{4.8}$$

In Appendix A, we give a detailed derivation of $\mathcal{L}_\theta^*$. We state the formula for the Fréchet derivative (4.5) in Theorem 4.1, which we prove in Appendix B.

THEOREM 4.1. *Let f_θ and g_θ be solutions of Equations* (4.6) *and* (4.7)*, respectively. The Fréchet derivative in* (4.5) *satisfies*

$$\partial_\theta M_\theta(r) = \int_0^t \int_\Omega 2 f_\theta(r+y,\tau) f_\theta(y,\tau) \left[2 g_\theta\left(\frac{r}{2}+y\right) - g_\theta(r+y) - g_\theta(y) \right] \, dy \, d\tau \,, \tag{4.9}$$

with $r \in \Omega_\theta$.

Evaluating $\partial_\theta M_\theta(a,t;f_0)$ involves solving the forward problem (4.6) for f_θ and the adjoint problem (4.7) for g_θ, where a, t, and f_0 enter the equations as parameters. The initial opinion distribution f_0 is the initial condition of the forward problem, and the interval $(-\infty, a]$ is the final condition of the adjoint problem. We solve both problems on the time interval $[0,t]$. The minimization of (4.2) is an example of *differential-equation-constrained optimization* because of its relationship (see (4.9)) between the loss function (4.2) and the forward (4.6) and adjoint (4.7) problems.

4.3. An optimization algorithm for problems with constraints. In section 4.1, we formulate the numerical inference of the interaction kernel θ as a minimization problem of the loss function $\mathrm{L}(\theta)$ [see (4.2)]. To minimize the loss function, we adopt a gradient-based method and evaluate the Fréchet derivative $\partial_\theta \mathrm{L}(\theta)$ [see (4.9)]. Evaluating $\partial_\theta \mathrm{L}(\theta)$ amounts to repeatedly solving the forward (4.6) and adjoint (4.7) problems. In particular, for each iteration in (4.4), evaluating $\partial_\theta \mathrm{L}(\theta)$ requires solving $|\mathcal{F}|$ forward problems and $|\mathcal{F}||\mathcal{A}|$ adjoint problems. For each (forward or adjoint) problem, we solve the associated integro-differential equation ((4.6) or (4.7)) for $|\mathcal{T}|$ time steps.

To reduce computational cost, it is important to decrease the number of iterations that we need. We design an algorithm (see Algorithm 4.1) with an adaptive step size to expedite the convergence of the minimization problem while accommodating the nonnegativity constraint of θ. We discretize $\theta(r)$ with a uniform grid and store the discretized kernel as a vector Θ. We use a subscript to denote the solution at a particular iteration. For example, Θ_r is the solution at the rth iteration.

Algorithm 4.1 is a gradient-based minimization algorithm that uses an adaptive step size both to maintain the nonnegativity constraint and to expedite convergence. We normalize the step size by the norm of the gradient vector (see lines 7 and 13) to try to avoid getting stuck in a local minimum. As Θ_n approaches a local minimum, the gradient-vector norm $\|r_n\|$ becomes closer to 0. This leads to a larger step size than what one would obtain without normalization and helps the algorithm jump over the local minimum. When a forward step satisfies the nonnegativity constraint, we double the step size (see line 9) to speed up the convergence; when a forward step breaks the constraint, we halve the step size (see line 11) and introduce a mechanism to adjust the descent direction to ensure that $\Theta_n \geq 0$ (see line 12). After fixing the descent direction, we adjust the step size again by using the new gradient norm (see line 13). The new gradient vector r_n may have a fairly small norm, which results in an excessive step size. To avoid abrupt changes in the step size, we impose a lower bound $\alpha_{\min}$ and an upper bound $\alpha_{\max}$ of the step size (see lines 11 and 13). Based on our numerical observations, the two parameters ($\alpha_{\min}$ and $\alpha_{\max}$) are necessary for our algorithm to succeed.

There are other available optimization methods to minimize the loss function (4.2). One appealing choice is to use a stochastic-gradient-descent (SGD) method [12, 18, 35]

Algorithm 4.1 An adaptive algorithm to minimize (4.2) with a nonnegativity constraint on the interaction kernel θ

Input: $\mathcal{T}$, $\mathcal{A}$, $\mathcal{F}$, Θ_0, α, $\alpha_{\min}$, $\alpha_{\max}$, $n_{\max}$
Output: $\Theta_{n_{\max}}$

1: **for** $n = 0, 1, 2, \ldots, n_{\max}$ **do**
2: **for** f_0 in $\mathcal{F}$ **do**
3: Solve the forward problem (4.6) for $f(x, t; f_0)$
4: Solve the adjoint problem (4.7) with $f(x, t; f_0)$ for all $a \in \mathcal{A}$
5: **end for**
6: Compute $r_n = \partial_\theta \mathrm{L}(\Theta_n)$ using Equations (4.5) and (4.9)
7: $\alpha_n = \max\{\alpha, \alpha/\|r_n\|\}$; $\widetilde{\Theta}_{n+1} = \Theta_n - \alpha_n r_n$
8: **if** $\min\{\widetilde{\Theta}_{n+1}\} \geq 0$ **then**
9: $\alpha = \alpha \times 2$; $\Theta_{n+1} = \widetilde{\Theta}_{n+1}$
10: **else**
11: $\alpha = \max\{\alpha/2, \alpha_{\min}\}$
12: $r_n(\widetilde{\Theta}_{n+1} < 0) = 0$; $\Theta_{n+1} = \max\{\widetilde{\Theta}_{n+1}, 0\}$
13: $\alpha_n^* = \min\{\alpha_{\max}, \max\{\alpha, \alpha/\|r_n\|\}\}$
14: $\widetilde{\Theta}_{n+1} = \Theta_n - \alpha_n^* r_n$
15: **if** $\min\{\widetilde{\Theta}_{n+1}\} \geq 0$ **then**
16: $\Theta_{n+1} = \widetilde{\Theta}_{n+1}$
17: **end if**
18: **end if**
19: **end for**

▷ In line 12, the first assignment $r_n(\widetilde{\Theta}_{n+1} < 0) = 0$ is entrywise, so we assign the ith entry of r_n to 0 if the ith entry of $\widetilde{\Theta}_{n+1}$ is negative.

with nonnegativity constraints. Because all of the sub-loss functions — which are given by $[M_\theta(a, t; f_0) - M^*(a, t; f_0)]^2$ for some combination of a, t, and f_0 — in (4.2) have the same minimizer, we expect to obtain a faster convergence rate with our approach than is typically the case for SGD methods. We do not compare our approach to an SGD approach in this paper, as such a comparison is tangential to our main goals, which are to formulate inverse problems for parameter inference in a BCM and to provide theoretical guarantees for our approach.

5. Numerical computations. We now do some computations to demonstrate the numerical inference method that we proposed in section 4. We consider data sets (with, necessarily, finitely many data points) that follow the two scenarios in section 3. With our computations, we demonstrate that our approach from section 4 is able to reconstruct the interaction kernel θ with good accuracy. We also demonstrate that the reconstruction accuracy increases exponentially as we increase the number of data points.

We consider an opinion space $\Omega = [-1, 1]$ and generate data with the true interaction kernel $\theta^*(r) = \mathbb{1}_{|r|<0.36}$ of the mean-field opinion model (2.1). We discretize Ω and Ω_θ with uniform square grids; each square is of size $d_x \times d_r$, with $d_x = 0.02$ and $d_r = 0.19$. We deliberately choose d_r to be much larger than d_x. Inferring the interaction kernel using a coarser grid than the one that generates the data helps mitigate overfitting problems. All of our computations use the same initial point Θ_0; we draw each element of Θ_0 independently from the uniform distribution on $(0, 1)$. The

iteration parameters in Algorithm 4.1 are $\alpha = 0.01$, $\alpha_{\min} = 0.003$, and $\alpha_{\max} = 0.05$. The set of time steps is $\mathcal{T} = 0.2 \times \{0, 1, \dots, 100\}$.

5.1. Reconstruction using data from a single measurement threshold. We fix the measurement-threshold set $\mathcal{A} = \{-1/3\}$ and let the initial-condition set $\mathcal{F}$ (i.e., the set of initial opinion distributions) be collections of uniform probability densities on intervals $(-B, B)$ for several values of B. We consider $B = 1$, $B = 0.9$, $B = 0.8$, and $B = 0.7$. These values yield the corresponding uniform probability density functions

$$f_0^1(x) = \frac{1}{2}\mathbb{1}_{(-1,1)}(x)\,, \quad f_0^2(x) = \frac{1}{2\times 0.9}\mathbb{1}_{(-0.9,0.9)}(x)\,, \tag{5.1}$$
$$f_0^3(x) = \frac{1}{2\times 0.8}\mathbb{1}_{(-0.8,0.8)}(x)\,, \quad f_0^4(x) = \frac{1}{2\times 0.7}\mathbb{1}_{(-0.7,0.7)}(x)\,.$$

In our numerical computations, $\mathcal{F}$ is one of the following four sets:

$$\mathcal{F} = \{f_0^1\}\,, \quad \mathcal{F} = \{f_0^1, f_0^2\}\,, \quad \mathcal{F} = \{f_0^1, f_0^2, f_0^3\}\,, \quad \mathcal{F} = \{f_0^1, f_0^2, f_0^3, f_0^4\}\,. \tag{5.2}$$

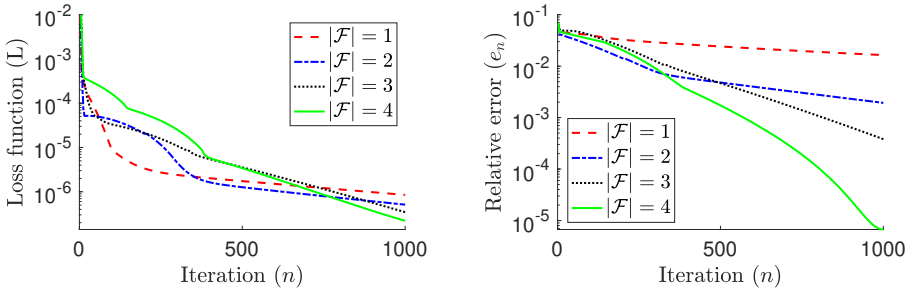

FIGURE 1. *The (left) loss-function value* $\mathrm{L}(\Theta_n)$ *and (right) relative error* $e_n = \|\Theta_n - \Theta^*\|/\|\Theta^*\|$ *a function of the iteration* n *of our optimization algorithm (see Algorithm 4.1). We fix the measurement threshold to* $a = -1/3$*. We perform* $n = 1000$ *iterations of the algorithm. When the algorithm finishes (i.e., when* $n = 1000$*), we obtain a relative error of* $e_n \approx 1.6 \times 10^{-2}$ *for* $|\mathcal{F}| = 1$*, a relative error of* $e_n \approx 1.9 \times 10^{-3}$ *for* $|\mathcal{F}| = 2$*, a relative error of* $e_n \approx 3.8 \times 10^{-4}$ *for* $|\mathcal{F}| = 3$*, and a relative error of* $e_n \approx 6.7 \times 10^{-6}$ *for* $|\mathcal{F}| = 4$*.*

In Figure 1, we show the decrease of the loss function (4.2) and the relative error $e_n = \|\Theta_n - \Theta^*\|/\|\Theta^*\|$ as a function of the iteration n of our optimization algorithm. Early in the optimization process (until about iteration $n = 400$), we observe that the loss function (4.2) decays irregularly. This arises from the adaptive nature of our optimization algorithm. Each iteration uses a new step size and adjusts the gradient-descent direction to preserve nonnegativity. Consequently, the algorithm is unlikely to have a constant convergence rate. In our numerical experiments, the optimization algorithm mostly selects $\alpha_{\min}$ as the step size after about 400 iterations. During this phase of the iterative process, the loss function decays exponentially at an approximately constant rate.

We observe that our optimization algorithm (see Algorithm 4.1) reduces the loss function (4.2) to a value of roughly the same order of magnitude (about 10^{-6}–10^{-7}) after $n = 1000$ iterations for data sets of different sizes ($|\mathcal{F}| = 1, \dots, 4$). In the right panel of Figure 1, we observe that the error drops dramatically as we enlarge the data set. After $n = 1000$ iterations, even though the loss function attains a similar value (of about 10^{-6}) for all $|\mathcal{F}| = 1, \dots, 4$, the associated inference of the interaction

kernel θ has different accuracies. With more data points, the loss function (4.2) becomes more effective at measuring the deviation from the true interaction kernel. The computational cost of minimizing the loss function increases as we enlarge the data set. At each iteration, we need to solve $|\mathcal{F}|$ forward problems (4.6) and $|\mathcal{A}| \times |\mathcal{F}|$ (which is equal to $|\mathcal{F}|$ in this example) adjoint problems (4.7).

5.2. Reconstruction using data from a single initial opinion distribution. We now fix the initial opinion distribution $f_0(x) = \mathbb{1}_{(-1,1)}(x)$ and vary the measurement-threshold set $\mathcal{A}$. We take $\mathcal{A} = \{a_0 - 1, 2a_0 - 1, \ldots, |\mathcal{A}|a_0 - 1\}$, with $a_0 = 1/(3|\mathcal{A}|)$, and we consider $|\mathcal{A}| = 1, \ldots, 4$. In Figure 2, we show the loss-function values and the relative errors as functions of the number of iterations of our optimization algorithm.

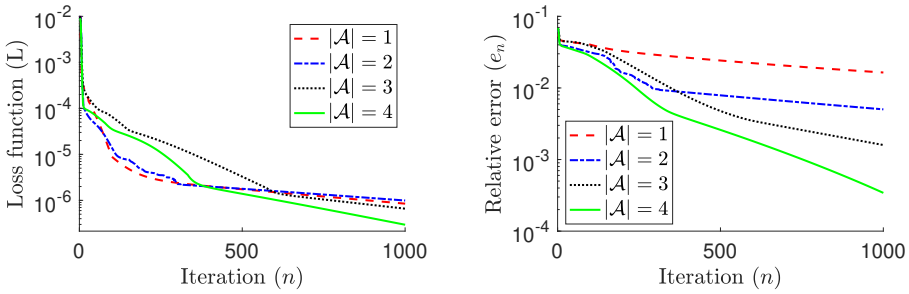

FIGURE 2. *The (left) loss-function value* $\mathrm{L}(\Theta_n)$ *and (right) relative error* $e_n = \|\Theta_n - \Theta^*\|/\|\Theta^*\|$ *as a function of the iteration* n *of our optimization algorithm (see Algorithm 4.1) for a data set that we generate from a single initial opinion distribution. After* $n = 1000$ *iterations, we obtain a relative error of* $e_n \approx 1.6 \times 10^{-2}$ *for* $|\mathcal{A}| = 1$*, a relative error of* $e_n \approx 5.0 \times 10^{-3}$ *for* $|\mathcal{A}| = 2$*, a relative error of* $e_n \approx 1.6 \times 10^{-3}$ *for* $|\mathcal{A}| = 3$*, and a relative error of* $e_n \approx 3.4 \times 10^{-4}$ *for* $|\mathcal{A}| = 4$*.*

In the left panel of Figure 2, the value of the loss function (4.2) decreases to about 10^{-6} after $n = 1000$ iterations for all data sets (i.e., for all $|\mathcal{A}| = 1, \ldots, 4$). Additionally, as in Figure 1, we again observe that the loss function decays irregularly for an initial set of iterations (until about $n = 550$) before decaying exponentially (for $n \geq 550$). After $n = 1000$ iterations, the loss function decreases to a value of the same order of magnitude for all 4 data sets. However, the relative error between the inferred and true interaction kernels differs across the 4 data sets. As we enlarge the size of the measurement-threshold set $\mathcal{A}$, the relative error decreases dramatically and the loss function becomes more effective at measuring the deviation from the true interaction kernel. At each iteration, we need to solve 1 forward problem (4.6) (because we fix the initial opinion distribution f_0) and $|\mathcal{A}|$ adjoint problems (4.7).

5.3. Comparing the two scenarios. By comparing Figures 1 and 2, we observe that when we terminate the optimization process, the loss-function values in both scenarios are about 10^{-6} but that the relative error between the inferred interaction kernel and the true interaction kernel is much smaller for $|\mathcal{F}| = 4$ than for $|\mathcal{A}| = 4$. In Figure 3, we plot the relative error after $n = 1000$ iterations for different values of $|\mathcal{F}|$ and $|\mathcal{A}|$. We observe that the relative error decays faster as we increase $|\mathcal{F}|$ than it does as we increase $|\mathcal{A}|$.

Importantly, our comparison between the two kernel-reconstruction scenarios does not imply that including data from multiple initial opinion distributions is more efficient than including data from multiple measurement thresholds. The performance of our optimization algorithm also depends on the initial opinion distributions and the measurement points. When two initial opinion distributions are too similar, one

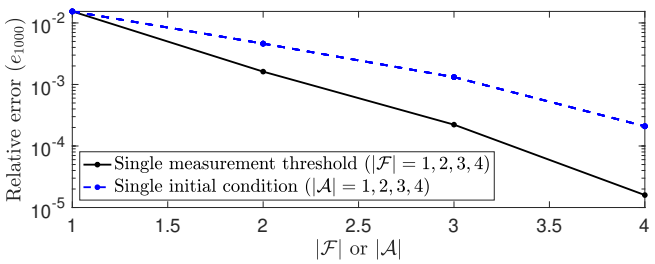

FIGURE 3. *A comparison of the sizes* $\|\hat{\Theta} - \Theta\|/\|\Theta\|$ *of the relative error between the inferred interaction kernel and the true interaction kernel for our two scenarios. We plot the relative error from data that we measure at a fixed measurement threshold but with different numbers of initial opinion distributions* ($|\mathcal{F}| = 1, \ldots, 4$) *as a solid black curve. We plot the relative error from data that we measure for a fixed initial opinion distribution but with different numbers of cumulative thresholds* ($|\mathcal{A}| = 1, \ldots, 4$) *as a dashed blue curve.*

typically expects them to yield similar dynamics and in turn to yield similar output data (although this need not be the case for chaotic dynamics), which thus may do little or nothing to improve the performance of our optimization algorithm and parameter inference. An analogous situation occurs when two measurement points are too close to each other. It is an important goal for future work to develop techniques to assess choices of the initial opinion distribution f_0 and measurement threshold a before applying an optimization algorithm to infer an interaction kernel.

6. Conclusions and discussion. In the mathematical modeling of opinion dynamics, it is typically difficult (or even impossible) to directly observe or measure parameter values or the functional forms of interactions. Therefore, it is important to develop methods to infer unknown parameters (either constants or functions) from empirical opinion data. To explicitly perform such a procedure, we formulated and examined an inverse problem using a mean-field bounded-confidence model (BCM) of opinion dynamics. We inferred the mean-field BCM's interaction kernel, which is a function that encodes how two agents interact and compromise their opinions. We examined this procedure both theoretically and numerically.

In our inverse problems, we considered two types of data sets: one with a fixed initial opinion distribution and the other with a fixed measurement threshold. For both scenarios, we proved that the given data has sufficient information to uniquely identify the interaction kernel θ (i.e., that the associated inverse problem is well-posed). We then developed a numerical inference strategy that employs a differential-equation-constrained optimization framework and seeks parameter values that produce simulated data that best matches the given empirical data. We formulated a gradient-based algorithm to execute the optimization. In this algorithm, one computes a Fréchet derivative with respect to the interaction kernel θ and repeatedly solves a set of forward and adjoint problems. Our numerical results showcased our well-posedness results for both scenarios.

The perspective of inverse problems is promising for the study of opinion dynamics. In the present paper, we examined inverse problems that are associated with a density-based BCM. It is also worthwhile to formulate and analyze inverse problems for agent-based models of opinion dynamics, such as for inferring the discordance function in agent-based BCMs on hypergraphs [32] or inferring waiting-time distributions in a non-Markovian opinion model on temporal networks [21]. We expect that it will

be valuable to study such inverse problems to validate opinion models and concretely connect them to real-life phenomena. It is also desirable to investigate a variety of approaches for the numerical inference strategy. We employed a simple gradient-based method, and approaches such as Hessian-based methods and interior-point methods may yield computational benefits.

Appendix A. Perturbed equations and adjoint operators. In this appendix, we examine the perturbed dynamics of the forward problem (4.6) and derive the adjoint operator $\mathcal{L}_\theta^*$ [see (4.8)].

Let $f_{\theta+\widetilde{\theta}}(x,t)$ be the solution of the forward problem (4.6) with the interaction kernel $\theta+\widetilde{\theta}$, and let $\widetilde{f} = f_{\theta+\widetilde{\theta}} - f_\theta$. Differentiating $\widetilde{f}$ with respect to time yields

$$\partial_t \widetilde{f}(x,t) = \mathcal{L}_\theta[\widetilde{f}] + \mathcal{S}_\theta[\widetilde{\theta}]\,, \quad \widetilde{f}(x,0) = 0\,, \tag{A.1}$$

where

$$\begin{aligned} \mathcal{L}_\theta[\widetilde{f}](x,t) &= \int_{\Omega\times\Omega} 2\widetilde{f}(x_1,t) f_\theta(x_2,t)\theta(x_1-x_2)F(x,x_1,x_2)\ dx_1\,dx_2\,, \\ \mathcal{S}_\theta[\widetilde{\theta}](x,t) &= \int_{\Omega\times\Omega} \widetilde{\theta}(x_1-x_2) f_\theta(x_1,t) f_\theta(x_2,t)F(x,x_1,x_2)\ dx_1\,dx_2\,. \end{aligned} \tag{A.2}$$

With a direct computation, we see that the adjoint operator $\mathcal{L}_\theta^*$ satisfies

$$\begin{aligned} \mathcal{L}_\theta^*[g](x,t) &= \int_{\Omega\times\Omega} 2g(x_1,t)\theta(x-x_2) f_\theta(x_2,t)F(x_1,x,x_2)\ dx_1\,dx_2 \\ &= \int_\Omega 2f_\theta(y,t)\theta(x-y)\left[2g\left(\frac{x+y}{2},t\right) - g(x,t) - g(y,t)\right]\ dy\,. \end{aligned} \tag{A.3}$$

Let g_θ be the solution of the adjoint problem

$$\partial_t g_\theta(x,t) = -\mathcal{L}_\theta^*[g_\theta](x,t)\,, \quad g_\theta(x,T) = \mathbb{1}_{(-\infty,a]}(x)\,. \tag{A.4}$$

A direct computation from (A.1) and (A.4) yields

$$\partial_t(\widetilde{f} g_\theta) = \mathcal{L}_\theta[\widetilde{f}]g_\theta + \mathcal{S}_\theta[\widetilde{\theta}]g_\theta - \mathcal{L}_\theta^*[g_\theta]\widetilde{f}\,. \tag{A.5}$$

We integrate both sides of (A.5) in both time and space. Using the initial opinion distribution (i.e., initial condition) of (A.1) and the final condition of (A.4), the left-hand side of (A.5) becomes

$$\int_0^T \int_\Omega \partial_t(\widetilde{f} g_\theta)(x,t)\ dx\,dt = \int_{-\infty}^a \widetilde{f}(x,T)\ dx = \widetilde{M}(T)\,, \tag{A.6}$$

where $\widetilde{M} = M_{\theta+\widetilde{\theta}} - M_\theta$ and M_θ is defined in (2.2). After integration, the first and last terms of (A.5) on the right-hand side cancel each other. Consequently,

$$\widetilde{M}(T) = \int_0^T \int_\Omega \mathcal{S}_\theta[\widetilde{\theta}](x,t) g_\theta(x,t)\ dx\,dt\,. \tag{A.7}$$

We conclude the derivations in this appendix with the following lemma.

LEMMA A.1. *Let f_θ and $f_{\theta+\tilde\theta}$ be solutions of Equations* (4.6) *with interaction kernels θ and $\theta+\tilde\theta$, respectively. Let g_θ be the solution of* (4.7). *We then have*

$$M_{\theta+\widetilde{\theta}}(a,t;f_0) - M_\theta(a,t;f_0) = \int_0^t\int_\Omega \mathcal{S}_\theta[\widetilde{\theta}](x,\tau) g_\theta(x,\tau)\, dx\, d\tau\,, \tag{A.8}$$

where $M_{\theta+\widetilde{\theta}}$ and M_θ are defined in (2.2) *using $f_{\theta+\widetilde{\theta}}$ and f_θ, respectively, and $\mathcal{S}_\theta$ is defined in* (A.2).

Appendix B. Proof of Theorem 4.1. In this appendix, we prove Theorem 4.1.

Proof. Recall Lemma A.1 and the definition of $\mathcal{S}_\theta$ in (A.2). We compute

(B.1)

$$\begin{aligned} M_{\theta+\widetilde{\theta}} - M_\theta &= \int_0^t\int_\Omega \mathcal{S}_\theta[\widetilde{\theta}](x,\tau) g_\theta(x,\tau)\, dx\, d\tau \\ &= \int_0^t\int_{\Omega_\theta\times\Omega\times\Omega} \widetilde{\theta}(r) g_\theta(x,\tau) f_\theta(r+y,\tau) f_\theta(y,\tau) F(x,r+y,y)\, dx\, dy\, dr\, d\tau \\ &= \int_0^t\int_{\Omega_\theta\times\Omega} 2\widetilde{\theta}(r) f_\theta(r+y,\tau) f_\theta(y,\tau) \\ &\qquad \times \left[2g_\theta\left(\frac{r}{2}+y,\tau\right) - g_\theta\left(r+y,\tau\right) - g_\theta(y,\tau)\right]\, dy\, dr\, d\tau\,, \end{aligned}$$

where we have omitted writing the dependence on a, t, and f_0. Equation (B.1) yields the Fréchet derivative

(B.2)

$$\partial_\theta M_\theta(r) = \int_0^t\int_\Omega 2 f_\theta(r+y,\tau) f_\theta(y,\tau)\left[2g_\theta\left(\frac{r}{2}+y,\tau\right) - g_\theta\left(r+y,\tau\right) - g_\theta(y,\tau)\right]\, dy\, d\tau\,.$$

∎

Acknowledgements. QL was supported in part by the National Science Foundation (through the grants NSF-CAREER-1750488 and NSF-DMS-2023239). MAP was funded by the National Science Foundation (grant 1922952) through the Algorithms for Threat Detection (ATD) program. WC was supported by the Wisconsin Alumni Research Foundation for her visit to QL at UW-Madison, where the research was initiated.

REFERENCES

[1] G. ALBI, E. CALZOLA, AND G. DIMARCO, *A data-driven kinetic model for opinion dynamics with social network contacts*, arXiv preprint arXiv:2307.00906, (2023).

[2] F. AMBLARD AND G. DEFFUANT, *The role of network topology on extremism propagation with the relative agreement opinion dynamics*, Physica A: Statistical Mechanics and its Applications, 343 (2004), pp. 725–738.

[3] S. R. ARRIDGE, *Optical tomography in medical imaging*, Inverse Problems, 15 (1999), pp. R41–R93.

[4] N. AYI AND N. P. DUTEIL, *Mean-field and graph limits for collective dynamics models with time-varying weights*, Journal of Differential Equations, 299 (2021), pp. 65–110.

[5] J. B. BAK-COLEMAN, M. ALFANO, W. BARFUSS, C. T. BERGSTROM, M. A. CENTENO, I. D. COUZIN, J. F. DONGES, M. GALESIC, A. S. GERSICK, J. JACQUET, A. B. KAO, R. E. MORAN, P. ROMANCZUK, D. I. RUBENSTEIN, K. J. TOMBAK, J. J. VAN BAVEL, AND E. U. WEBER, *Stewardship of global collective behavior*, Proceedings of the National Academy of Sciences of the United States of America, 118 (2021), e2025764118.

[6] G. Bal and F. Monard, *Inverse transport with isotropic time-harmonic sources*, SIAM Journal on Mathematical Analysis, 44 (2012), pp. 134–161.
[7] J. J. V. Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman, et al., *Using social and behavioural science to support COVID-19 pandemic response*, Nature Human Behaviour, 4 (2020), pp. 460–471.
[8] E. Ben-Naim, P. L. Krapivsky, and S. Redner, *Bifurcations and patterns in compromise processes*, Physica D: Nonlinear Phenomena, 183 (2003), pp. 190–204.
[9] C. Bernardo, C. Altafini, A. Proskurnikov, and F. Vasca, *Bounded confidence opinion dynamics: A survey*, Automatica, 159 (2024), 111302.
[10] G. Bohner and N. Dickel, *Attitudes and attitude change*, Annual Review of Psychology, 62 (2011), pp. 391–417.
[11] P. Bonacich and P. Lu, *Introduction to Mathematical Sociology*, Princeton University Press, Princeton, NJ, USA, 2012.
[12] L. Bottou, *Stochastic gradient descent tricks*, in Neural Networks: Tricks of the Trade, Springer-Verlag, Heidelberg, Germany, 2012, pp. 421–436.
[13] H. Z. Brooks and M. A. Porter, *A model for the influence of media on the ideology of content in online social networks*, Physical Review Research, 2 (2020), 023041.
[14] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences of the United States of America, 113 (2016), pp. 3932–3937.
[15] K. P. Bube and R. Burridge, *The one-dimensional inverse problem of reflection seismology*, SIAM Review, 25 (1983), pp. 497–559.
[16] C. Castellano, S. Fortunato, and V. Loreto, *Statistical physics of social dynamics*, Reviews of Modern Physics, 81 (2009), pp. 591–646.
[17] G. Chen, W. Su, W. Mei, and F. Bullo, *Convergence properties of the heterogeneous Deffuant–Weisbuch model*, Automatica, 114 (2020), 108825.
[18] K. Chen, Q. Li, and J.-G. Liu, *Online learning in optical tomography: A stochastic approach*, Inverse Problems, 34 (2018), 075010.
[19] M. Choulli and P. Stefanov, *Inverse scattering and inverse boundary value problems for the linear Boltzmann equation*, Communications in Partial Differential Equations, 21 (1996), pp. 763–785.
[20] W. Chu and M. A. Porter, *A density description of a bounded-confidence model of opinion dynamics on hypergraphs*, arXiv preprint arXiv:2203.12189 (SIAM Journal on Applied Mathematics, in press), (2023).
[21] W. Chu and M. A. Porter, *Non-Markovian models of opinion dynamics on temporal networks*, SIAM Journal on Applied Dynamical Systems, 22 (2023), pp. 2624–2647.
[22] S. Clémençon, V. Chi Tran, and H. De Arazoza, *A stochastic SIR model with contact-tracing: Large population limits and statistical inference*, Journal of Biological Dynamics, 2 (2008), pp. 392–414.
[23] M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, in Handbook of Uncertainty Quantification, Springer-Verlag, Heidelberg, Germany, 2017, pp. 311–428.
[24] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, *Mixing beliefs among interacting agents*, Advances in Complex Systems, 3 (2000), pp. 87–98.
[25] A. Franci, M. Golubitsky, A. Bizyaeva, and N. E. Leonard, *A model-independent theory of consensus and dissensus decision making*, arXiv preprint arXiv:1909.05765, (2019).
[26] D. Frey, *Recent research on selective exposure to information*, Advances in Experimental Social Psychology, 19 (1986), pp. 41–80.
[27] S. Fricker, M. Galesic, R. Tourangeau, and T. Yan, *An experimental comparison of web and telephone surveys*, Public Opinion Quarterly, 69 (2005), pp. 370–392.
[28] M. Galesic, H. Olsson, J. Dalege, T. van der Does, and D. L. Stein, *Integrating social and cognitive aspects of belief dynamics: Towards a unifying framework*, Journal of The Royal Society Interface, 18 (2021), 20200857.
[29] S.-Y. Ha and E. Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinetic & Related Models, 1 (2008), pp. 415–435.
[30] J. Helliwell, R. Layard, and J. Sachs, *World Happiness Report*, The Earth Institute, Columbia University, New York City, NY, USA, 2012.
[31] K. Hellmuth, C. Klingenberg, Q. Li, and M. Tang, *Kinetic chemotaxis tumbling kernel determined from macroscopic quantities*, 2022, https://arxiv.org/abs/2206.01629.
[32] A. Hickok, Y. Kureh, H. Z. Brooks, M. Feng, and M. A. Porter, *A bounded-confidence model of opinion dynamics on hypergraphs*, SIAM Journal on Applied Dynamical Systems, 21 (2022), pp. 1–32.

[33] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE constraints*, vol. 23, Springer-Verlag, Heidelberg, Germany, 2008.
[34] A. T. Jebb, V. Ng, and L. Tay, *A review of key Likert scale development advances: 1995–2019*, Frontiers in Psychology, 12 (2021), 637547.
[35] B. Jin and X. Lu, *On the regularizing property of stochastic gradient descent*, Inverse Problems, 35 (2018), 015004.
[36] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag, Heidelberg, Germany, 2011.
[37] I. V. Kozitsin, *Opinion dynamics of online social network users: A micro-level analysis*, The Journal of Mathematical Sociology, 47 (2023), pp. 1–41.
[38] R.-Y. Lai, G. Uhlmann, and Y. Yang, *Reconstruction of the collision kernel in the nonlinear Boltzmann equation*, SIAM Journal on Mathematical Analysis, 53 (2021), pp. 1049–1069.
[39] J. Lenti, C. Monti, and G. De Francisci Morales, *Likelihood-based methods improve parameter estimation in opinion dynamics models*, 2023, https://arxiv.org/abs/2310.02766.
[40] L. Li and Z. Ouyang, *Determining the collision kernel in the Boltzmann equation near the equilibrium*, Proceedings of the American Mathematical Society, 151 (2023), pp. 4855–4865.
[41] Q. Li and W. Sun, *Applications of kinetic tools to inverse transport problems*, Inverse Problems, 36 (2020), 035011.
[42] Y. Liu, S. G. McCalla, and H. Schaeffer, *Random feature models for learning interacting dynamical systems*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 479 (2023), 20220835.
[43] J. Lorenz, *Continuous opinion dynamics under bounded confidence: A survey*, International Journal of Modern Physics C, 18 (2007), pp. 1819–1838.
[44] F. Lu, M. Maggioni, and S. Tang, *Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories*, Foundations of Computational Mathematics, 22 (2022), pp. 1013–1067.
[45] F. Lu, M. Zhong, S. Tang, and M. Maggioni, *Nonparametric inference of interaction laws in systems of agents from trajectory data*, Proceedings of the National Academy of Sciences of the United States of America, 116 (2019), pp. 14424–14433.
[46] R. M. May, S. A. Levin, and G. Sugihara, *Ecology for bankers*, Nature, 451 (2008), pp. 893–894.
[47] S. McQuade, B. Piccoli, and N. Pouradier Duteil, *Social dynamics models with time-varying influence*, Mathematical Models and Methods in Applied Sciences, 29 (2019), pp. 681–716.
[48] X. F. Meng, R. A. Van Gorder, and M. A. Porter, *Opinion formation and distribution in a bounded-confidence model on various networks*, Physical Review E, 97 (2018), 022312.
[49] C. Monti, G. De Francisci Morales, and F. Bonchi, *Learning opinion dynamics from social traces*, in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, New York, NY, USA, 2020, Association for Computing Machinery, pp. 764–773.
[50] S. Motsch and E. Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Review, 56 (2014), pp. 577–621.
[51] M. E. J. Newman, *Networks*, Oxford University Press, Oxford, UK, second ed., 2018.
[52] H. Noorazar, K. R. Vixie, A. Talebanpour, and Y. Hu, *From classical to modern opinion dynamics*, International Journal of Modern Physics C, 31 (2020), 2050101.
[53] J. Ojer, M. Starnini, and R. Pastor-Satorras, *Modeling explosive opinion depolarization in interdependent topics*, Physical Review Letters, 130 (2023), 207401.
[54] I. M. Otto, J. F. Donges, R. Cremades, A. Bhowmik, R. J. Hewitt, W. Lucht, J. Rockström, F. Allerberger, M. McCaffrey, S. S. P. Doe, A. Lenferna, N. Morán, D. P. v. Vuuren, and H. J. Schellnhuber, *Social tipping dynamics for stabilizing Earth's climate by 2050*, Proceedings of the National Academy of Sciences of the United States of America, 117 (2020), pp. 2354–2365.
[55] M. A. Porter and J. P. Gleeson, *Dynamical Systems on Networks: A Tutorial*, vol. 4 of Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Springer International Publishing, Cham, Switzerland, 2016.
[56] S. Redner, *Reality inspired voter models: A mini-review*, Comptes Rendus Physique, 20 (2019), pp. 275–292.
[57] D. O. Sears and J. L. Freedman, *Selective exposure to information: A critical review*, Public Opinion Quarterly, 31 (1967), pp. 194–213.
[58] A. Sîrbu, D. Pedreschi, F. Giannotti, and J. Kertész, *Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model*, PloS ONE, 14 (2019), e0213246.

[59] R. Snieder and J. Trampert, *Inverse problems in geophysics*, in Wavefield Inversion, Springer-Verlag, Heidelberg, Germany, 1999, pp. 119–190.

[60] B. State and L. Adamic, *The diffusion of support in an online social movement: Evidence from the adoption of equal-sign profile pictures*, in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15, New York, NY, USA, 2015, Association for Computing Machinery, pp. 1741–1750.

[61] G. Toscani, *Kinetic models of opinion formation*, Communications in Mathematical Sciences, 4 (2006), pp. 481–496.

[62] G. Uhlmann, *Inverse Problems and Applications: Inside Out II*, Cambridge University Press, Cambridge, UK, 2013.

[63] C. Villani, *A review of mathematical topics in collisional kinetic theory*, Handbook of Mathematical Fluid Dynamics, 1 (2002), pp. 3–8.

[64] A. Volkening, D. F. Linder, M. A. Porter, and G. A. Rempala, *Forecasting elections using compartmental models of infection*, SIAM Review, 62 (2020), pp. 837–865.

[65] S. Wang, E. D. Herzog, I. Z. Kiss, W. J. Schwartz, G. Bloch, M. Sebek, D. Granados-Fuentes, L. Wang, and J.-S. Li, *Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks*, Proceedings of the National Academy of Sciences of the United States of America, 115 (2018), pp. 9300–9305.

[66] W. H. Warren, J. B. Falandays, K. Yoshida, T. D. Wirth, and B. A. Free, *Human crowds as social networks: Collective dynamics of consensus and polarization*, Perspectives on Psychological Science, (2023), 17456916231186406.