Likelihood-Based Methods Improve
Parameter Estimation in Opinion Dynamics Models

Jacopo Lenti
jcp.lenti@gmail.com
CENTAI, Turin, Italy
Sapienza University, Rome, Italy

Corrado Monti
corrado.monti@centai.eu
CENTAI, Turin, Italy

Gianmarco De Francisci Morales
gdfm@acm.org
CENTAI, Turin, Italy

ABSTRACT

We show that a maximum likelihood approach for parameter estimation in agent-based models (ABMs) of opinion dynamics outperforms the typical simulation-based approach. Simulation-based approaches simulate the model repeatedly in search of a set of parameters that generates data similar enough to the observed one. In contrast, likelihood-based approaches derive a likelihood function that connects the unknown parameters to the observed data in a statistically principled way. We compare these two approaches on the well-known bounded-confidence model of opinion dynamics.

We do so on three realistic scenarios of increasing complexity depending on data availability: (*i*) fully observed opinions and interactions, (*ii*) partially observed interactions, (*iii*) observed interactions with noisy proxies of the opinions. To realize the likelihood-based approach, we first cast the model into a probabilistic generative guise that supports a proper data likelihood. Then, we describe the three scenarios via probabilistic graphical models and show the nuances that go into translating the model. Finally, we implement such models in an automatic differentiation framework, thus enabling easy and efficient maximum likelihood estimation via gradient descent. These likelihood-based estimates are up to 4× more accurate and require up to 200× less computational time.

CCS CONCEPTS

• **Mathematics of computing** → **Probabilistic inference problems**; • **Human-centered computing** → *Collaborative and social computing*; • **Theory of computation** → Social networks.

KEYWORDS

Agent-based models, opinions dynamics, social media, probabilistic modeling, maximum likelihood

Figure 1: Estimation of the 'open-mindedness' parameter ϵ for a Bounded-Confidence Model with a baseline simulation-based approach (Method of Simulated Moments, left) vs a likelihood-based approach (Maximum Likelihood, right). Perfect estimates lie on the diagonal. The likelihood-based approach produces much more accurate estimates.

1 INTRODUCTION

The challenges currently facing humanity require global coordination. The Covid-19 pandemic and climate change are two clear examples of such global problems. However, people are struggling to find a consensus to tackle these problems, and even simple and effective interventions such as face masks become controversial [31]. Why does this happen?

This question is the main focus of opinion dynamics, the research field that studies the evolution of people's opinions. And while the literature is replete with possible theoretical models [34, 51], few of these have been empirically validated [47, 58]. The rise of online social media has provided an unprecedented amount of data about their user's opinions and interactions. And yet, we are no closer to understanding opinion formation today than 20 years ago.

One of the main obstacles to validating models of opinion formation lies in their nature: agent-based models (ABMs). A staple of complex systems methodology, ABMs are mechanistic models where *agents* interact among themselves and with an environment according to predefined local rules. The interactions change the state of the system which in turn affects the interactions in a possibly complex feedback loop, and drive the evolution of the system (e.g., users' opinions). The system is then simulated to derive outcomes and implications, and to explore what-if scenarios. The rules are typically *parameterized*, and allow for a variety of agent interactions and system steady states. The flexible and bottom-up nature of ABMs represents a double-edged sword: on the one hand, these models are simple to build and understand; on the other hand, they lack a principled way to use data to estimate their parameters.

Parameter estimation in ABMs, also called *'calibration'* in the literature, is the focus of the current paper. The vast majority of approaches to the task are *simulation-based*: they explore the parameter space and run a set of simulations with the parameter values under consideration. The set of parameters able to produce the data traces "closest" to the real data is taken as the result of this procedure. The definition of closest is arbitrary, and usually consists in comparing summary statistics of the data according to some chosen metric (e.g., differences in mean and variance). Unfortunately, this procedure can generate quite inaccurate estimates for the parameters of even very simple models.

Figure 1 shows two scatterplots of the true and estimated parameter values for the well-known bounded-confidence model of opinion dynamics [17, 27]. This model, detailed in Section 3, is the simplest non-linear opinion dynamics ABM, and has a single continuous parameter $\epsilon \in [0, 0.5]$ that determines the "open-mindedness" of agents. Even in this basic example, a simulation-based calibration approach produces relative errors of up to 89% (90th percentile), as shown in the left plot. Conversely, the right plot shows the maximum likelihood approach (22% relative error at 90th percentile). It is evident that the likelihood-based approach significantly improves on the simulation-based one. Experiments reported in Section 5 demonstrate that the likelihood-based method offers more accurate estimates on a variety of different opinion dynamics scenarios. In addition, this increased accuracy does not come at the cost of increased computational time. On the contrary, the likelihood-based approach is markedly faster than the simulation-based one.

In general, the goal of this paper is to compare the likelihood-based parameter estimation to the standard simulation-based solution on a set of scenarios with increasing complexity. We increase the complexity of the task by taking inspiration from the real challenges that empirical researchers face every day. We first consider partial observability of the data, e.g., on social media we can only observe 'positive' interactions (retweets and likes) by design. Then, we consider that estimating the exact opinion of users on social media is often impossible, and we have to rely on noisy proxies such as the use of informative hashtags or participation in partisan communities. These scenarios offer a much more realistic benchmark of the accuracy of parameter estimation for opinion dynamics models. Our experimental results show that even in the simplest scenario the gap in accuracy between the two methods is surpisingly large.

This paper represents a first step towards a more structured evaluation of the methodological practices in opinion dynamics research. It emphasizes the importance of aligning our methodologies with the models we use, especially considering the abundant data accessible via social media. Our aspiration to enhance the understanding of opinion formation hinges on the reliability of our findings. Given the complexity of the issue, it is evident that building more solid foundations than the current field norms is an essential step forward.

2 RELATED WORK

The literature on opinion dynamics is incredibly vast and it would be impossible to review it in this paper, so we limit ourselves to giving a broad overview. Typically, in opinion dynamics ABMs individuals are represented by agents who share their opinion with their neighbors on an underlying social network, which possibly evolves over time. The opinion state variable can be discrete [53] or continuous [17], and represent the stance of a user towards a specific topic, such as politics [16], vaccines [3], or civil rights [8]. The interactions between agents cause the opinion variables to change, thus leading to different scenarios depending on the model in consideration and the specific set of parameters [45].

The advantage of using ABMs for opinion dynamics lies in being able to encode directly individual behaviors and consider only the local properties of the system. ABMs are also incredibly flexible, as they can easily incorporate complex strategies or different types of interactions, both with other peers or with the environment. Thanks to their flexibility, ABMs are used also in several other fields such as economics, ecology, and human mobility [4].

The interest in opinion dynamics in particular has been recently renewed due to the emergence of phenomena such as polarization, filter bubbles, and echo chambers on social media [6, 14, 22, 43]. For instance, the data mining community has produced several interesting models [2, 10, 12, 42]. The availability of data from social media has surely been an incredible catalyst for this type of research. However, this line of research has not been accompanied by appropriate methodological progress in calibration and validation of the models [45].

Calibration is the process of determining the parameters that align a model with an observed dataset [47]. Given the generative essence of Agent-Based Models (ABMs), the research community in this realm has commonly embraced simulation-based methodologies as the conventional calibration approach [15]. These techniques require the execution of numerous simulations of the model, each with a different parameter configuration. The procedure then selects the parameter configuration that most accurately emulates the data trajectories or system property. A wide range of approaches have been developed in this direction, which all share three fundamental requisites: (*i*) the identification of summary statistics to quantitatively characterize the model output, (*ii*) the choice of a distance metric to measure the similarity between the summary statistics of the simulated and observed data, and (*iii*) a parameter sampling procedure aimed at exploring the parameter space [37, 47]. In addition, these methods require choosing an acceptance threshold to assess whether the obtained simulations have replicated the dataset well enough, and a stopping criterion for the exploration of the parameter space [24].

The predominant and de-facto standard simulation-based method is the Method of Simulated Moments (MSM) [20, 37]. In each iteration, this method simulates the ABM to get a time series of generated data. Then it compares specific moments of the real data and the simulated time series. After a sufficiently long exploration of the parameter space, the parameters that minimize the discrepancy between the selected moments are chosen as the best estimates [20]. This approach is also described as reproducing "stylized facts", i.e., broad tendencies, in economics.

Researchers have explored different metrics for quantifying the distance between simulated and observed data [7, 29, 55]. The quest for a more efficient parameter space exploration stands as another challenge within simulation-based methods for calibration. To minimize the number of simulations required, the methods look for the

most promising regions of the parameter space. Several Bayesian strategies integrate a prior knowledge of the distribution of the parameters of interest have been developed. Grazzini et al. [23] offer a comprehensive spectrum of approaches, encompassing parametric, non-parametric, and likelihood-free techniques. Specifically, Approximate Bayesian Computation (ABC) is an established likelihood-free Bayesian technique in common practice. ABC seeks to maximize the posterior probability of the target quantity, by relying on specific summary statistics, when a likelihood function is not available [56]. Additionally, several techniques employ deep learning to reduce the number of simulations needed to determine the desired parameters [19, 50]. In this case, the deep learning model is used as a surrogate for the ABM, so that new data traces can be generated quickly after a more expensive training phase. Within the stream of Bayesian methods, particle filtering and Sequential Monte Carlo techniques calibrate the parameters online, by simulating the ABM step by step [25, 35, 36, 54]. Other strategies use simulations to approximate the likelihood [30, 48], or a simulated maximum likelihood estimator [28]. Particularly notable are the advances in differentiable ABMs, which facilitate the calibration via auto-differentiation techniques and fast tensorized simulations [13, 49].

However, all the aforementioned approaches share some evident limitations. First, the arbitrary choices of the summary statistics and of the distance metric heavily affect outcomes of calibration [20]. Second, the computational cost associated with simulating ABMs presents a substantial hurdle, making iterative runs computationally taxing and challenging to alleviate [47]. Consequently, models are often validated solely in simplified versions featuring a limited number of agents [23]. Third, these methods suffer from the curse of dimensionality. As the dimensionality of the parameter space increases, an exhaustive exploration of such space becomes practically unfeasible [29]. Moreover, when increasing the complexity of the ABMs, unexpected or unclear behaviors can emerge, which makes purely simulative approaches less reliable [32, 52]. Furthermore, comparing only some summary statistics leads to an inevitable waste of valuable information from the data, thus forcing one to concentrate on a few specific macroscopic properties of the system [23]. For these reasons, research in ABMs has recently highlighted the urgency to find new robust methods of calibration for validating existing models, instead of focusing on the definition of new models able to capture different behaviors [4, 45].

In contrast to simulation-based approaches, only a handful of methodologies employ data to maximize the likelihood of generating them [18, 39, 41, 59]. The goal is always to maximize the likelihood associated with the observed data: Dong [18] and Xu et al. [59] derive the likelihood through stochastic kinetic models, and use Variational Inference to optimize it and make predictions. However, to the best of our knowledge, a standardized protocol for implementing these approaches is currently missing, and the solutions presented address specific problems in epidemic models and traffic congestion scenarios. Conversely, we employ Probabilistic Graphical Models (PGMs) to represent the ABMs. This novel approach enables the translation of the ABM dynamics into conditional probabilities, by encapsulating agent states and interactions as random variables, whether latent or observed. Within this framework, parameter estimation is not attained in the simulations aimed at replicating the desired properties of the data. Rather, the focus

shifts to the maximization of the likelihood of generating the observed variables. Consequently, parameter calibration takes the form of an optimization task, which can be tackled via automatic differentiation (e.g., PyTorch [44], JAX [11], and TensorFlow [1]). However, this promising direction has solely delved into specific scenarios so far [39, 41]. Beyond the theoretical robustness of maximum likelihood estimation, the questions pertaining to its broader feasibility and empirical outcomes are still open. In particular, a head-to-head comparison with the simulation-based method serves to explain the real utility of the Maximum Likelihood strategy and its practical viability.

3 MODEL AND SCENARIOS

We consider one of the most well-known opinion dynamics model, i.e., the bounded-confidence model (BCM) [17, 27], in its stochastic version [39]. Our choice is motivated by the popularity of the model, which has been studied extensively. Despite being easy to understand and describe, this model is the first non-linear model of opinion dynamics ever proposed.

The aim of the model is to represent a situation where the opinions of agents are influenced only by interactions with neighbors that have similar opinions. More specifically, an agent is influenced by an interaction with another agent only if their opinions are within their *confidence interval*, encoded in a parameter ϵ. These interactions drive the evolution of their opinions by mutual influence and convergence. This model is an ABM: each agent behaves according to independent, local rules.

We consider a population V of N agents. At time t, each agent u has an opinion $x_t^u \in [0, 1]$, with initial opinion x_0^u drawn uniformly at random. At each time $t \in \{1, \ldots, T\}$, m pairs of agents are extracted uniformly at random from $V \times V$. We indicate these extractions with e_t. An interaction occurs for each pair of agents in e_t, for a total of $M = m \cdot T$ interactions.

In the original formulation of BCM, when u and v interact, if $|x_t^u - x_t^v| < \epsilon$ their opinions converge towards their average at a rate μ (convergence parameter). Otherwise, their opinions remain unchanged. In the stochastic version we consider, each outcome happens with some probability. For each interaction (u, v) at time t, a sigmoid function defines the probability of that interaction being positive. Formally, we extract a sign $s_t^{u,v} \in \{0, 1\}$ as a Bernoulli trial

$$P(s_t^{u,v} = 1) = \sigma\Big(\rho \cdot (\epsilon - |x_t^u - x_t^v|)\Big), \quad (1)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the logistic function and ρ is a steepness parameter that tunes the amount of stochastic behavior. Then, the agents update their opinions as follows:

$$\begin{cases} x_{t+1}^u = x_t^u + s_t^{u,v} \cdot \mu(x_t^v - x_t^u) \\ x_{t+1}^v = x_t^v + s_t^{u,v} \cdot \mu(x_t^u - x_t^v), \end{cases} \quad (2)$$

that is, after an interaction agents are likely to get closer if they were within their *confidence bound*. Note that for $\rho \to \infty$ the model is equivalent to the original, deterministic BCM model [39].

Scenarios. The goal of this work is to compare two different paradigms for the estimation of parameters: likelihood-based methods based on probabilistic generative models and simulation-based

methods. We focus on the estimation of ϵ, since it is the most characteristic parameter of the model. Recapping, the variables involved are thus the following:

- ϵ is the bounded-confidence parameter of the BCM model, it is the target of the estimation.
- μ is the convergence parameter, it is known.
- $X_t = (x_t^1, \ldots, x_t^N) \in [0, 1]^N$ the opinions of the agents at time t. It depends on the opinions and interactions at previous time steps. The initial state X_0 is extracted from a uniform distribution in $[0, 1]^N$.
- $e_t = (e_t^1, \ldots, e_t^m) \in (V^2)^m$ are the pairs of agents extracted at time t. They are extracted uniformly at random (with replacement) from V^2. A single pair of agents is an interaction, and it is defined as $e_t^i = (u, v)$.
- $s_t \in \{0, 1\}^m$ are the sign of interactions at time t, that is, whether each interaction has resulted in opinion change. Sign s_t^i is the outcome of the interaction e_t^i. If $s_t^i = 1$ the interacting agents update their opinions according to the BCM rules, otherwise $s_t^i = 0$. The outcomes s_t of the interactions depend on the agents involved e_t, their opinions X_t, and the parameter ϵ.

To comprehensively evaluate the performances of the two estimation methods, we run experiments in three scenarios with varying degrees of knowledge about the opinion formation process: (i) fully-observed interactions (BCM-F); (ii) partially-observed interactions (BCM-P); (iii) noisy proxies for opinions(BCM-N). Let us now give the details of these three scenarios.

Fully-observed Bounded Confidence (BCM-F). In the first scenario, we assume complete knowledge of the other variables in the process, which means that we observe the initial condition X_0, the identities of the pairs of interacting agents e_t, and the resulting outcome of the interaction s_t. Given the observations of all of these variables, the task is always to estimate the unknown parameter ϵ. The results in Figure 1 are obtained from this scenario.

Partially-observed Bounded Confidence (BCM-P). In our second scenario, we reduce the available knowledge: we can only observe interactions with a positive outcome. This scenario is realistic in many cases where data is available: for instance, retweets on Twitter represent support, but obtaining data about disagreeing users is much harder. Hence, we assume to observe the opinions of the agents X_t, and the pairs of interacting agents having a positive interaction: e_t is observable only when $s_t = 1$.

Bounded Confidence with Noisy Opinions (BCM-N). The third scenario assumes having only partial and noisy information on the opinions while observing all the interactions e and their signs s. We introduce an observed variable $Y_t \in \{0, 1\}^k$ which is a proxy of the opinions of a subset k of the agents. To obtain this proxy, at each time t a set of k agents is selected uniformly at random to share some content with label $y_t^u \sim \text{Bernoulli}(x_t^u)$. In total, we have $K = k \cdot T$ binary proxies. The opinion vector X_0 is instead latent, as well as its evolution X_t. These conditions fit computational social science studies, where the exact stance of social media users is hard to know, but we have some proxies of their opinions in the form of the hashtags they use or the communities they participate in.

Similarly, binary labels of the stance of a user are easier to elicit from human labelers than their continuous hidden representation.

4 LIKELIHOOD-BASED ESTIMATION

The goal of this work is to compare a likelihood-based method, and in particular, one based on maximum likelihood (ML) estimation via gradient-based optimization, against a simulation-based one (method of simulated moments, MSM). As explained in Section 2, MSM simulates the ABM while exploring the space of the unknown parameters. The parameter that generated the simulated trajectory closest to the observed one is the final parameter estimate $\hat{\epsilon}$.

Conversely, for ML, we maximize the likelihood function for the parameter estimate $\hat{\epsilon}$ after having cast each ABM into an equivalent probabilistic generative ABM [39, 41].

To maximize the likelihood function we specify the conditional probabilities connecting the parameters, the latent variables, and the observed data. Often, the locality of ABMs allows to capture conditional (in-)dependencies between random variables, thus making it easier to treat the likelihood. Then, we estimate the target parameter by maximizing the marginal likelihood of the parameters of interest from the observed data z, i.e., $\hat{\epsilon} = \arg\max_\epsilon \mathcal{L}(\epsilon \mid z)$.

In the remainder of this section, we analyze each of the scenarios considered, show how it gets represented as a probabilistic graphical model (PGM), and derive its marginal likelihood w.r.t. the parameter of interest ϵ. Figure 2 depicts the PGMs of the three scenarios. Circles represent stochastic variables, diamonds deterministic variables, and the known parameters have no shape. (μ and ρ are omitted for simplicity). Grey-colored shapes are observed variables, and white ones are latent. X_t is known in BCM-F and BCM-P, because it is a deterministic function of observed variables.

4.1 Fully-observed Scenario (BCM-F)

The PGM induced by the model in this scenario is represented in Figure 2(a). Our goal is to optimize the marginal likelihood of ϵ given the observed data. Since the only observed variable that depends on the latent parameter is the sequence of interaction outcomes s, we write the likelihood function as

$$\mathcal{L}(\epsilon) = P(s \mid \epsilon). \tag{3}$$

Note that X_t is known for any t, as it is a deterministic function of the interaction signs s and of X_0, which are both observed in this scenario. Thanks to the conditional independencies found in the PGM, Equation (3) factorizes over the sequence of interactions as

$$P(s \mid \epsilon) = \prod_{j=1,\ldots,M} P(s^j \mid X_t, \epsilon).$$
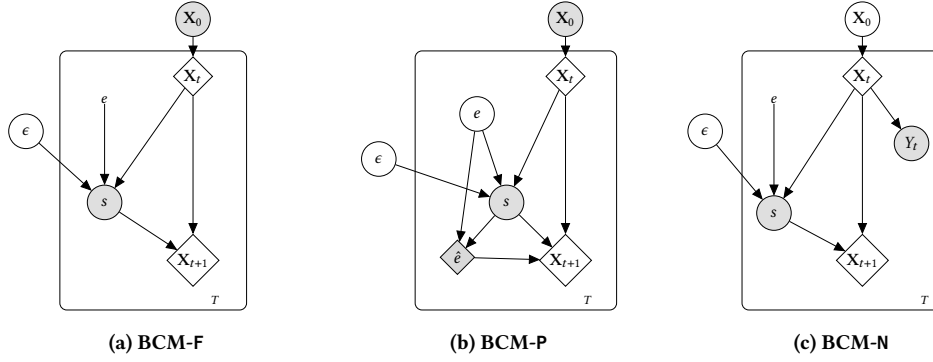
Finally, we simply derive the likelihood of a Bernoulli distribution from Equation (1),

$$P(s^j \mid X_t, \epsilon) = s^j \kappa_\epsilon(e^j) + (1 - s^j)(1 - \kappa_\epsilon(e^j)), \tag{4}$$

where for brevity we indicate with $\kappa_\epsilon(e^j)$ the sigmoid as a function of the pair of agents $e^j = (u, v)$, that is,

$$\kappa_\epsilon(e^j) = \sigma\Big(\rho \cdot (\epsilon - |x_t^u - x_t^v|)\Big).$$

By switching to the log-likelihood for optimization purposes (with abuse of notation), we obtain the following objective function

(a) BCM-F   **(b) BCM-P**   **(c) BCM-N**

Figure 2: Representation of the Bounded Confidence Model as a probabilistic graphical model in three different scenarios. In the first scenario (a) BCM-F, both the initial state X_0 and the signs s are fully-observed. In BCM-P (b), X_0 is still known but only the positive interactions are observed, represented by the fact that only some of the edges $\hat{e}$ are observed (chosen deterministically). In the final one BCM-N (c), the initial state X_0 is unknown, and at each timestamp we get a random noisy proxy Y_t of the state of some of the agents.

for the PGABM

$$\mathcal{L}(\epsilon) = \sum_{j=1,\dots,M} \log\left(s^j \kappa_\epsilon(e^j) + (1-s^j)(1-\kappa_\epsilon(e^j))\right). \quad (5)$$

We get an estimate $\hat{\epsilon}$ simply by maximizing this function with gradient descent.

4.2 Partially-observed Scenario (BCM-P)

In this scenario, we only have knowledge of which agents interacted if their interaction caused opinion change. Therefore, the pair of interacting agents e is, in the general case, latent. We introduce $\hat{e}^j$ as an observed deterministic variable, that is equal to e^j if $s^j = 1$, and is equal to $\varnothing$ if $s^j = 0$ (where the symbol $\varnothing$ denotes a null observation). The induced PGM is represented in Figure 2(b). Again, our goal is to estimate the bounded-confidence parameter ϵ from the observable data. That is, we want to maximize the marginal likelihood w.r.t. the parameter ϵ given s and $\hat{e}$. Similarly to the previous scenario, we have knowledge of X_0 and of the interactions that cause opinion change (which are only the positive ones according to the BCM), and therefore we fully observe X_t. When considering the likelihood, we can again factorize each element of the sequence of interactions

$$\mathcal{L}(\epsilon) = \log P(\hat{e}, s \mid \epsilon) = \sum_{j=1,\dots,M} \log P(\hat{e}^j, s^j \mid \epsilon)$$

In this scenario, we do not know which agents interacted when the interaction does not change opinions. Thus, we have to integrate over all the possible extractions of e, which is drawn uniformly from V^2. The probability above further factorizes as

$$P(\hat{e}^j, s^j \mid \epsilon) = \sum_{e^j \in V^2} P(\hat{e}^j, s^j, e^j \mid \epsilon)$$
$$= \sum_{e^j \in V^2} P(\hat{e}^j \mid s^j, e^j, \epsilon) P(s^j \mid e^j, \epsilon) P(e^j \mid \epsilon).$$

Let us now look at these three probabilities. Since $\hat{e}^j$ is deterministic given s^j, e^j, the first one is equal to 1 when $s^j = 0 \wedge \hat{e}^j = \varnothing$, or when $s^j = 1 \wedge \hat{e}^j = e^j$, and is equal to 0 otherwise. The second one, $P(s^j \mid e^j, \epsilon)$ is given by the sigmoid, and we write it as $\kappa_\epsilon(e^j)$ when

$s^j = 1$ and $(1 - \kappa_\epsilon(e^j))$ otherwise. Finally, $P(e^j \mid \epsilon)$ is constant with respect to ϵ and therefore we can ignore it for our estimation task. Therefore, we can divide interactions according to their sign s^j and rewrite the log-likelihood as

$$\mathcal{L}(\epsilon) = \sum_{j:s^j=1} \log\left(\kappa_\epsilon(\hat{e}^j)P(e^j)\right) + \sum_{j:s^j=0} \log\left(\sum_{e^j \in V^2}(1-\kappa_\epsilon(e^j))P(e^j)\right),$$

which is our objective function in this scenario. In practice we do not need to sample all the possible pairs of agents, as we can approximate $\sum_{e^j \in V^2}(1 - \kappa_\epsilon(e^j))P(e^j)$ with a sample mean.

4.3 Noisy Opinions Scenario (BCM-N)

Figure 2(c) shows the corresponding PGM. Differently from the previous scenarios, here X_0 is latent, and therefore we do not know X_t. Even though we wish to estimate only ϵ, we need to also estimate X_0 (and thus X_t) as the two variables are not independent given the signs s. Since gradient descent relies on backpropagation to update the estimate of X_0, we write the function $X_t = f_t(X_0)$ as the following efficient matrix operation. Let A_t be the adjacency matrix of positive interactions at time t, that is $A_{i,j} = 1$ iff $\exists j : e^j = (i, j) \wedge s^j = 1$. Then, we can write Equation (2) as

$$X_{t+1} = X_t + \mu\left(\left(X_t \mathbf{1}^T - \mathbf{1}X_t^T\right) \odot A_t\right)\mathbf{1}^T,$$

where $\odot$ indicates the Hadamard product.

The log-likelihood function of this model can be divided into two factors, one for the interaction signs s and one for the opinion proxies Y, since they are independent given X_0, ϵ

$$\mathcal{L}(\epsilon) = \log P(s, Y \mid \epsilon, X_0) = \log P(s \mid \epsilon, X_0) + \log P(Y \mid \epsilon, X_0)$$
$$= \sum_{j=1,\dots,M} \log\left(s^j \kappa_\epsilon(e^j) + (1-s^j)(1-\kappa_\epsilon(e_t))\right)$$
$$+ \sum_{j=1,\dots,K} \log\left(y_j f_t(X_0)_j + (1-y_j)(1-f_t(X_0)_j)\right).$$

In the optimization procedure, we initialize X_0 randomly and use the update rule f to maximize this objective function.

5 EXPERIMENTS

We compare the accuracy in parameter estimation of the simulation- and likelihood-based methods with a set of experiments with equal data availability. For all our experiments, the ground-truth model used to generate the data traces and the generative model assumed for both methods are the same, i.e., there is no misspecification error. For each experimental configuration, we generate 120 data traces by varying the seed of the pseudorandom generator. In all the experiments we fix $N = 100$, $\rho = 16$, $\mu = 0.1$. We vary $T \in [16, 32, 64, 128, 256, 512]$ and $m \in [1, 4, 8, 16]$. In BCM-N we also vary $k \in [4, 8, 16]$. These parameters are available to both methods and the only unknown one is ϵ. The code to reproduce these results is open-source and available at https://github.com/jaclenti/ABMcalibration-MLvsMSM.

MSM parameter settings. For the implementation of the MSM we rely on Black-it [9], a recent, open-source, simulation-based calibration framework. It encapsulates several widely-used approaches in the field, thus representing a crystallization of the state-of-the-art. As pointed out in Section 2, MSM requires some arbitrary choices:

- *Summary statistics.* We compute the time series w of the number of interactions with positive outcomes ($w_t = \sum_j s_t^j$). Following previous literature on MSM [9], we define a vector of 18 moments of w (first 4 moments and the first 5 auto-correlations of w, first 4 moments and first 5 auto-correlations of the absolute differences at lag one of w).
- *Loss function.* It incorporates the notion of distance between the simulated and observed time series of the ABM. We use the Mean Squared Error (MSE) between the simulated and observed summary statistics.
- *Exploration strategy.* We use the Halton sampler [26] to explore the parameter space, as it achieves the best results in tuning.
- *Number of simulations.* This one is the main tuning hyperparameter for MSM: since the estimation time increase linearly with the number of simulations, there is a time/accuracy trade-off. When tuning this parameter, we observe an increase in accuracy until 200 simulations, and a knee corresponding to such value. Hence, in all the experiments we run 200 simulations (details on the tuning are shown in Supplementary Material).

In addition, some small customization to each scenario is required to run MSM.

- BCM-F: at each timestep, the signs s_t are simulated from the known agents e_t and opinions X_t.
- BCM-P: in this case, the agents e_t are latent. However, simulating s_t from the observed $\hat{e}_t$ would end up in an upward-biased estimate of $\hat{\epsilon}$, since the sample $\hat{e}_t$ represent pairs of agents that have had a positive interaction, and are thus likely to be close each other in opinion space. Hence, at each step, we sample $e_t \in V^2$ and simulate s_t by knowing X_t.
- BCM-N: since X_0 is latent, a sample from a uniform distribution is used as its estimate $\hat{X}_0$. At each timestep, the signs s_t are simulated from e_t and from the current estimate of $\hat{X}_t$ (as a function of $\hat{X}_0$, $e_{\tau<t}$, $s_{\tau<t}$). Note that MSM does not provide a way to estimate latent variables X_t, but only parameters such as ϵ. Moreover, to consider the information from Y_t, we sample

a set of k agents $\tilde{V}_t$, and sample $\hat{Y}_t \sim \text{Bernoulli}(X_t(\tilde{V}_t))$. In this scenario, we compute the moments of three different time series: $w_t = \sum_j s_t^j$, $\gamma_t^0 = \mathbb{E}(\hat{Y}_t)$ and $\gamma_t^1 = \mathbb{V}(\hat{Y}_t)$. The distance between the observed and simulated time series is the sum of the MSEs between the 18 moments of w, γ^0, and γ^1.

ML parameter settings. We maximize the objective functions resulting from each scenario by using gradient descent on PyTorch, a Python package that enables automatic differentiation. Following hyperparameter tuning, we choose distinct optimizers for the three scenarios. Notably, and differently from MSM, in BCM-N the optimization process involves the estimation both of the target parameter ϵ and the latent variable X_0.

The specific optimizers and their respective learning rates for each model are as follows:

- BCM-F: Nadam with a learning rate of 0.1.
- BCM-P: Adam with a learning rate of 0.05.
- BCM-N: RMSprop with learning rates for ϵ set to 0.05 and for X_0 set to 0.5.

All optimizers are stopped after 200 epochs if they have not converged yet. For further details on the tuning of these parameters refer to the Supplementary Material.

Implementation details. Some implementation tricks are used in the optimization pipeline of our likelihood functions:

- To constrain the parameter space ($[0, 0.5]$ for ϵ and $[0, 1]$ X_0), we represent the target parameters as the logit function of an unbounded optimization variable $\theta \in \mathbb{R}$. For example, $\hat{\epsilon} = \sigma(\theta)/2$. This choice avoids the need for clipping and possible problems with vanishing gradient.
- To estimate the parameters of BCM-P we need to compute the population mean, which involves averaging $\kappa_\epsilon(e)$ at each step for all possible agent pairs e. To reduce the computational cost of this process, we employ a sample mean of $\kappa_\epsilon(e)$ instead.
- In BCM-N X_0 is latent, and the likelihood function is symmetric around the midpoint of the opinion space (0.5), i.e., X_0 and $1-X_0$ are both equally good estimates for the latent opinions. This behavior arises from the arbitrariness of the two endpoints of the opinion scale, in addition to the fact that the likelihood computed from the interactions dependent only on the distance between agent opinions, not their absolute values. We implement an approach to circumvent encountering a local optimum when $\hat{X}_0 \approx 1-X_0$. During each epoch, we calculate the objective function as $\tilde{\mathcal{L}}(\epsilon, X_0 \mid s, Y) = \max(\mathcal{L}(\epsilon, X_0 \mid s, Y), \mathcal{L}(\epsilon, 1-X_0 \mid s, Y))$. Consequently, we select the value of $\hat{X}_0$ that optimizes this modified objective function.

Results. Figure 3 shows the average absolute errors $\delta_{ML} = |\hat{\epsilon}_{ML}-\epsilon|$ and $\delta_{MSM} = |\hat{\epsilon}_{MSM} - \epsilon|$ in the three scenarios, as a function of T (and relative 95% confidence intervals). The difference in the performances of the two methods is evident. Aggregating over the three scenarios, we obtain a median decrease in the absolute error of 3.1× (median δ_{ML} 0.013 vs δ_{MSM} 0.040). Considering that ϵ has maximum value 0.5, an average error of 0.1 as obtained by MSM in low-data configurations represents a minimum relative error of 20%. The error of MSM increases slightly on more complex scenarios,
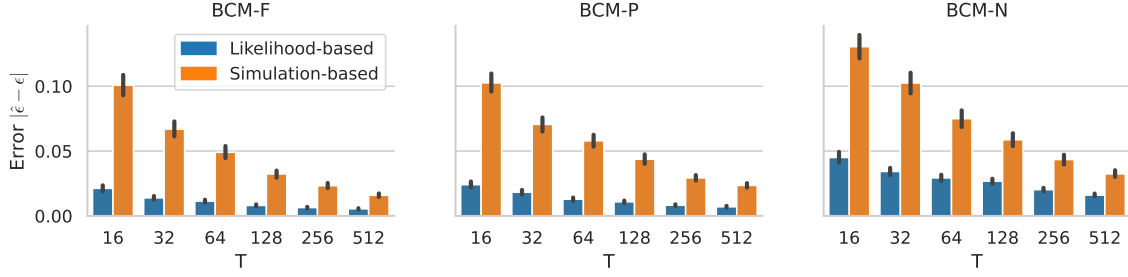
Figure 3: Average absolute error of estimates of ϵ with Maximum Likelihood and Method of Simulated Moments as a function of T. Error bars represent the 95% confidence intervals. The likelihood-based approach produces estimates with several times smaller errors.
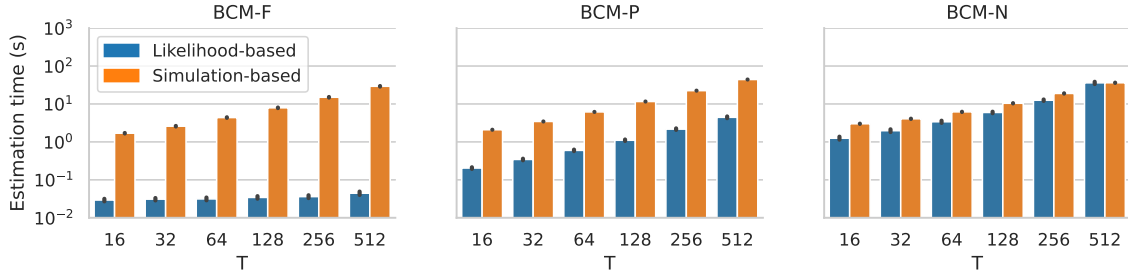

Figure 4: Average time in seconds required to estimate ϵ with Maximum Likelihood and Method of Simulated Moments as a function of T. Error bars represent the 95% confidence intervals. The likelihood-based approach improves the computational cost by orders of magnitude when it does not need to reproduce the entire opinion trajectory (left and center plots). In the last scenario (right), at each iteration the entire opinion trajectory X needs to be re-computed from the new estimate of X_0.

although not in a significant way. Instead, the ML approach achieves very low errors for the fully-observed scenario, and only marginally higher errors for the more complex scenarios. Reassuringly, both methods are able to improve their estimates when using more data.

Figure 4 compares the computational costs of the estimation (note the logarithmic axis). In BCM-F and BCM-P the comparison is strikingly one-sided. MSM requires repeating the whole simulation of the ABM hundreds of times, thus incurring a significant computational burden. However, in BCM-N the Maximum Likelihood method shows similar estimation time. This increase in computational cost is due to the need to re-compute X from the current estimate of X_0 at each epoch. The median ratios between the estimation times, T_{MSM}/T_{ML}, are, respectively, 207.0, 11.8, and 2.2, for the three scenarios respectively. Aggregating the three scenarios, we have a median reduction of estimation time by 3.9× (median estimation time of 1.7s for ML vs 7.5s for MSM).

ML outperforms MSM in all the configurations of parameters tested. In particular, ML always reduces the average error as the dataset size increases, both in terms of T, m, and k. Conversely, MSM shows an unclear trend when varying these parameters: the error decreases as T increases, is not affected by k, and has a bell shape when varying m. Due to space constraints, we leave the exploration of this particular behavior of MSM to future research and show these results in the Supplementary Material.

One important advantage of ML is that it can also provide an estimate of the latent variable X_0 in BCM-N. Figure 5 displays the distribution of the coefficients of determination $R^2(\hat{X}_0, X_0)$ as a function

of T and k. Besides the experiments with very small datasets, ML is able to estimate X_0 with a high level of accuracy. Looking at experiments with $T \geq 64$, we obtain a median value of $R^2(X_0) = 0.90$. This result is particularly interesting since $X_0 \in [0, 1]^{100}$ is high-dimensional, which indicates that this method could be employed for more complicated models with larger parameter spaces. A comparison of the true values of X_0 and estimated ones $\hat{X}_0$ for an experiment with a median R^2 can be found in Supplementary Material.

6 DISCUSSION

Our paper presents a comprehensive comparison between two distinct approaches to calibrating Agent-Based Models (ABMs) for opinion dynamics. We highlight several compelling factors that advocate for the adoption of a maximum likelihood (ML) approach over a simulation-based method (MSM) when calibrating an opinion dynamics model. While opinion dynamics models are notoriously rudimentary and unrealistic, we see our endeavor to improve their connection to real data as a crucial step to make them useful as statistical models in empirical work [38, 40, 46].

Experimental evidence. Our experiments span a broad range of scenarios w.r.t. data availability. The results clearly show the superiority of ML in terms of computational cost and estimation accuracy in all the contexts we considered. The differences in the performances reach orders of magnitude. The gap in performance persists as the dataset size increases.
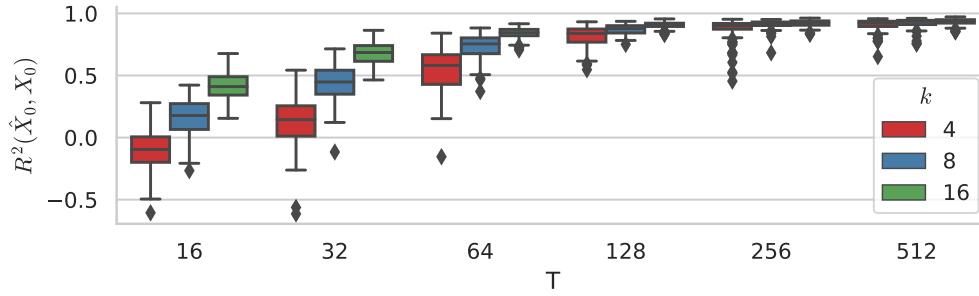
Figure 5: Distribution of R^2 scores of the estimate of the initial condition $\hat{X}_0$ with Maximum Likelihood as a function of the length of the data trace T and the number of noisy observations of the opinions k. The number of edges per timestep is $m = 4$.

We compared the ML approach with a representative simulation-based method. We opted for MSM as it is the most popular and consolidated technique, and it captures most of the typical behaviors of its class of methods. While more complex methods exist (see Section 2), there is no extensive assessment of their comparative performance, and thus it is hard to prefer one approach over another. Additionally, several methods are restricted to a specific framework, and their adaptation to different models is non-trivial.

Theoretical robustness. The ML approach is grounded in a robust statistical framework that aligns with our specific problem. Rather than resorting to a trial-and-error methodology, we directly identify the parameters that most likely generated a given dataset. In contrast, the MSM method inherently relies on selected summary statistics and a chosen distance metric to characterize and compare simulated and observed time series. These selections wield significant influence over the calibration results, potentially discarding valuable information not encapsulated within these summaries.

Although there is potential to improve the choices of the summary statistics, distance measures, and sampling strategies, conducting an extensive optimization of the MSM for the presented scenarios is out of the scope of this paper. Indeed, our point is that these choices will always be somewhat arbitrary, while using the data likelihood represents a principled approach to the issue. Conversely, the ML approach casts the inference problem into an optimization task, which nevertheless needs careful hyperparameter tuning. Fortunately, this process can leverage the extensive body of knowledge built by the machine learning community, which offers valuable insights and established methodologies.

In real applications, the true data-generating process is unknown, which can cause a model mismatch. In such contexts, simulation-based methods can attempt to replicate some macroscopic properties of the system. However, likelihood-based methods can compare the likelihood of different models to unveil the most realistic scenario on the basis of the microscopic behavior of the agents.

Appropriatedness of the probabilistic framework. We used Probabilistic Graphical Models (PGMs) to formalize the rules of an Agent-Based Model (ABM) in a mathematical framework. In itself, this formalization is already a noteworthy contribution, given the reproducibility issues that affect the ABM literature [21, 57]. Several factors support this choice:

- We have complete knowledge of the conditional probability distributions implied by the rules of the ABM. Therefore, it is natural to represent the causal connections via a graph of dependencies.

- The resultant PGMs often exhibit remarkable sparsity, mirroring the tendency of ABMs to involve only a handful of elements in causal relationships. This characteristic greatly simplifies the inference process for PGMs.

- PGMs can leverage the entire dataset to infer latent variables from the observed ones. This feature is missing in simulation-based approaches, which seed simulations with random initial conditions in an attempt to align with observed variables, essentially by chance.

Estimated variables. In this work, we focused on the estimation of a specific parameter, ϵ. Notably, the limited dimensionality of the parameter space and the straightforward nature of the examined ABM should provide an optimal environment for the MSM approach. Despite these conditions, ML consistently outperforms MSM across all scenarios, even in the simplest, fully-observed one. Additionally, the experiments in BCM-N display that ML is able to effectively recover the initial conditions of the ABM in a 100-dimensional space. These findings open the door to broader investigations involving higher-dimensional parameter spaces and more intricate ABMs.

Generalization and future work. A limitation of our work is the use of a single opinion dynamics model. While our approach can be extended to more complex models, the process is not straightforward. Indeed, one of the root causes of the difficulty in calibrating and validating ABMs is that their rules can quickly become very complex. The generalizability of our results deserves further study.

However, there are reasons to be optimistic. For instance, even though optimizing discrete probabilistic programs is still challenging, recent advances have been made in the area [5, 33]. These capabilities would allow including a larger class of models with discrete parameters and variables, common in ABMs, within the maximum likelihood framework. In addition, even when the likelihood function is not tractable exactly, approximate techniques such as continuous approximations or variational inference can be used to tackle complex models. By connecting the world of ABM to probabilistic models and machine learning, we can hope to leverage the rapid advancements that these fields are experiencing.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467* (2016). (Cited on 3)

[2] Rediet Abebe, Jon Kleinberg, David Parkes, and Charalampos E Tsourakakis. 2018. Opinion dynamics with varying susceptibility to persuasion. In *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1089–1098. (Cited on 2)

[3] Lucila G Alvarez-Zuzek, Cristian E La Rocca, José R Iglesias, and Lidia A Braunstein. 2017. Epidemic spreading in multiplex networks influenced by opinion exchanges on vaccination. *PloS one* 12, 11 (2017), e0186492. (Cited on 2)

[4] Li An, Volker Grimm, Abigail Sullivan, BL Turner Ii, Nicolas Malleson, Alison Heppenstall, Christian Vincenot, Derek Robinson, Xinyue Ye, Jianguo Liu, et al. 2021. Challenges, tasks, and opportunities in modeling agent-based complex systems. *Ecological Modelling* 457 (2021), 109685. (Cited on 2, 3)

[5] Gaurav Arya, Moritz Schauer, Frank Schäfer, and Christopher Rackauckas. 2022. Automatic differentiation of programs with discrete randomness. *Advances in Neural Information Processing Systems* 35 (2022), 10435–10447. (Cited on 8)

[6] Chris Bail. 2021. *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton University Press. (Cited on 2)

[7] Sylvain Barde. 2016. Direct comparison of agent-based models of herding in financial markets. *Journal of Economic Dynamics and Control* 73 (2016), 329–353. (Cited on 2)

[8] Fabian Baumann, Philipp Lorenz-Spreen, Igor M Sokolov, and Michele Starnini. 2021. Emergence of polarized ideological opinions in multidimensional topic spaces. *Physical Review X* 11, 1 (2021), 011012. (Cited on 2)

[9] Marco Benedetti, Gennaro Catapano, Francesco De Sclavis, Marco Favorito, Aldo Glielmo, Davide Magnanimi, and Antonio Muci. 2022. Black-it: A Ready-to-Use and Easy-to-Extend CalibrationKit for Agent-based Models. *Journal of Open Source Software* 7, 79 (Nov. 2022), 4622. (Cited on 6)

[10] Nikita Bhalla, Adam Lechowicz, and Cameron Musco. 2023. Local edge dynamics and opinion polarization. In *16th ACM International Conference on Web Search and Data Mining*. 6–14. (Cited on 2)

[11] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. 2018. *JAX: composable transformations of Python+NumPy programs*. http://github.com/google/jax (Cited on 3)

[12] Uthsav Chitra and Christopher Musco. 2020. Analyzing the impact of filter bubbles on social network polarization. In *13th International Conference on Web Search and Data Mining*. 115–123. (Cited on 2)

[13] Ayush Chopra, Alexander Rodríguez, Jayakumar Subramanian, Balaji Krishnamurthy, B Aditya Prakash, and Ramesh Raskar. 2022. Differentiable Agent-based Epidemiology. *arXiv:2207.09714* (2022). (Cited on 3)

[14] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The Echo Chamber Effect on Social Media. *PNAS* 118, 9 (2021), e2023301118. (Cited on 2)

[15] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. 2020. The frontier of simulation-based inference. *PNAS* 117, 48 (2020), 30055–30062. (Cited on 2)

[16] Henrique Ferraz de Arruda, Felipe Maciel Cardoso, Guilherme Ferraz de Arruda, Alexis R Hernández, Luciano da Fontoura Costa, and Yamir Moreno. 2022. Modelling how social network algorithms can influence opinion polarization. *Information Sciences* 588 (2022), 265–278. (Cited on 2)

[17] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems* 3, 01n04 (2000), 87–98. (Cited on 2, 3)

[18] Wen Dong. 2016. Variational inference with agent-based models. *arXiv:1605.04360* (2016). (Cited on 3)

[19] Joel Dyer, Patrick Cannon, J Doyne Farmer, and Sebastian Schmon. 2022. Blackbox Bayesian inference for economic agent-based models. *arXiv:2202.00625* (2022). (Cited on 3)

[20] Giorgio Fagiolo, Mattia Guerini, Francesco Lamperti, Alessio Moneta, and Andrea Roventini. 2019. Validation of agent-based models in economics and finance. In *Computer simulation validation*. Springer, 763–787. (Cited on 2, 3)

[21] Ben G Fitzpatrick. 2019. Issues in reproducible simulation research. *Bulletin of mathematical biology* 81 (2019), 1–6. (Cited on 3)

[22] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying Controversy on Social Media. *ACM Transactions on Social Computing* 1, 1 (2018), 3. (Cited on 2)

[23] Jakob Grazzini, Matteo G Richiardi, and Mike Tsionas. 2017. Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control* 77 (2017), 26–47. (Cited on 3)

[24] C Marijn Hazelbag, Jonathan Dushoff, Emanuel M Dominic, Zinhle E Mthombothi, and Wim Delva. 2020. Calibration of individual-based models to epidemiological data: A systematic review. *PLoS computational biology* 16, 5 (2020), e1007893. (Cited on 2)

[25] Mira Kattwinkel and Peter Reichert. 2017. Bayesian parameter inference for individual-based models using a Particle Markov Chain Monte Carlo method. *Environmental modelling & software* 87 (2017), 110–119. (Cited on 3)

[26] Ladislav Kocis and William J Whiten. 1997. Computational investigations of low-discrepancy sequences. *ACM Transactions on Mathematical Software (TOMS)* 23, 2 (1997), 266–294. (Cited on 6)

[27] Ulrich Krause. 2000. A discrete nonlinear and non-autonomous model of consensus formation. *Communications in difference equations* 2000 (2000), 227–236. (Cited on 2, 3)

[28] Jiri Kukacka and Jozef Barunik. 2017. Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control* 85 (2017), 21–45. (Cited on 3)

[29] Francesco Lamperti. 2018. An information theoretic criterion for empirical validation of simulation models. *Econometrics and Statistics* 5 (2018), 83–106. (Cited on 2, 3)

[30] Francesco Lamperti, Andrea Roventini, and Amir Sani. 2018. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control* 90 (2018), 366–389. (Cited on 3)

[31] Jun Lang, Wesley W Erickson, and Zhuo Jing-Schmidt. 2021. # MaskOn!# MaskOff! Digital polarization of mask-wearing in the United States during COVID-19. *PloS one* 16, 4 (2021), e0250817. (Cited on 1)

[32] Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska, Behroexoz Hassani-Mahmooei, Forrest Stonedahl, Iris Lorscheid, Alexey Voinov, J Gareth Polhill, Zhanli Sun, and Dawn C Parker. 2015. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation* 18, 4 (2015). (Cited on 3)

[33] Alexander K Lew, Mathieu Huot, Sam Staton, and Vikash K Mansinghka. 2023. ADEV: Sound automatic differentiation of expected values of probabilistic programs. *ACM on Programming Languages* 7, POPL (2023), 121–153. (Cited on 8)

[34] Jan Lorenz. 2007. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C* 18, 12 (2007), 1819–1838. (Cited on 1)

[35] Thomas Lux. 2021. Bayesian estimation of agent-based models via adaptive particle Markov chain Monte Carlo. *Computational Economics* (2021), 1–27. (Cited on 3)

[36] Thomas Lux. 2023. Sequential Monte Carlo Squared for Agent-Based Models. In *Artificial Intelligence, Learning and Computation in Economics and Finance*. Springer, 59–69. (Cited on 3)

[37] Thomas Lux and Remco CJ Zwinkels. 2018. Empirical validation of agent-based models. In *Handbook of computational economics*. Vol. 4. Elsevier, 437–488. (Cited on 2)

[38] Corrado Monti, Luca Maria Aiello, Gianmarco De Francisci Morales, and Francesco Bonchi. 2022. The language of opinion change on social media under the lens of communicative action. *Scientific Reports* 12, 1 (Oct. 2022), 17920. (Cited on 7)

[39] Corrado Monti, Gianmarco De Francisci Morales, and Francesco Bonchi. 2020. Learning opinion dynamics from social traces. In *26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 764–773. (Cited on 3, 4)

[40] Corrado Monti, Jacopo D'Ignazi, Michele Starnini, and Gianmarco De Francisci Morales. 2023. Evidence of Demographic rather than Ideological Segregation in News Discussion on Reddit. In *ACM Web Conference 2023*. ACM, Austin TX USA, 2777–2786. (Cited on 7)

[41] Corrado Monti, Marco Pangallo, Gianmarco De Francisci Morales, and Francesco Bonchi. 2023. On learning agent-based models from data. *Scientific Reports* 13, 1 (2023), 9268. (Cited on 3, 4)

[42] Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. 2018. Minimizing polarization and disagreement in social networks. In *2018 world wide web conference*. 369–378. (Cited on 2)

[43] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. penguin UK. (Cited on 2)

[44] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019). (Cited on 3)

[45] Antonio F Peralta, János Kertész, and Gerardo Iñiguez. 2022. Opinion dynamics in social networks: From models to data. *arXiv:2201.01322* (2022). (Cited on 2, 3)

[46] Flavio Petruzzellis, Francesco Bonchi, Gianmarco De Francisci Morales, and Corrado Monti. 2023. On the Relation between Opinion Change and Information Consumption on Reddit. *International AAAI Conference on Web and Social Media* 17 (June 2023), 710–719. (Cited on 7)

[47] Donovan Platt. 2020. A comparison of economic agent-based model calibration methods. *Journal of Economic Dynamics and Control* 113 (2020), 103859. (Cited on 1, 2, 3)

[48] Donovan Platt. 2022. Bayesian estimation of economic simulation models using neural networks. *Computational Economics* 59, 2 (2022), 599–650. (Cited on 3)

[49] Arnau Quera-Bofarull, Ayush Chopra, Anisoara Calinescu, Michael Wooldridge, and Joel Dyer. 2023. Bayesian calibration of differentiable agent-based models.

arXiv:2305.15340 (2023). (Cited on 3)

[50] Takashi Shiono. 2021. Estimation of agent-based models using Bayesian deep learning approach of BayesFlow. *Journal of Economic Dynamics and Control* 125 (2021), 104082. (Cited on 3)

[51] Alina Sîrbu, Vittorio Loreto, Vito DP Servedio, and Francesca Tria. 2017. Opinion dynamics: models, extensions and external effects. *Participatory sensing, opinions and collective awareness* (2017), 363–401. (Cited on 1)

[52] Vivek Srikrishnan and Klaus Keller. 2021. Small increases in agent-based model complexity can result in large increases in required calibration data. *Environmental Modelling & Software* 138 (2021), 104978. (Cited on 3)

[53] Katarzyna Sznajd-Weron and Jozef Sznajd. 2000. Opinion evolution in closed community. *International Journal of Modern Physics C* 11, 06 (2000), 1157–1165. (Cited on 2)

[54] Patricia Ternes, Jonathan A Ward, Alison Heppenstall, Vijay Kumar, Le-Minh Kieu, and Nick Malleson. 2022. Data assimilation and agent-based modelling: towards the incorporation of categorical agent parameters. *Open Research Europe*

1 (2022), 131. (Cited on 3)

[55] Carlo M Valensise, Matteo Cinelli, and Walter Quattrociocchi. 2023. The drivers of online polarization: fitting models to data. *Information Sciences* (2023), 119152. (Cited on 2)

[56] Elske van der Vaart, Mark A Beaumont, Alice SA Johnston, and Richard M Sibly. 2015. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling* 312 (2015), 182–190. (Cited on 3)

[57] Uri Wilensky and William Rand. 2007. Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation* 10, 4 (2007), 2. (Cited on 8)

[58] Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. 2007. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 10, 2 (2007), 8. (Cited on 1)

[59] Zhen Xu, Wen Dong, and Sargur N Srihari. 2016. Using social dynamics to make individual predictions: variational inference with a stochastic kinetic model. *Advances in Neural Information Processing Systems* 29 (2016). (Cited on 3)