Causal ABMs: Learning Plausible Causal Models using Agent-based Modeling

Konstantina Valogianni KONSTANTINA.VALOGIANNI@IE.EDU
Department of Information Systems
IE Business School, IE University
Madrid, Spain

Balaji Padmanabhan BP@USF.EDU
Muma College of Business
University of South Florida
Tampa, FL, USA

Thuc Le, Lin Liu, Emre Kıcıman, Sofia Triantafyllou and Huan Liu

Abstract

We present *Causal ABM*, a methodology to derive causal structures describing complex underlying behavioral phenomena. Agent-based models (ABMs) have powerful advantages for causal modeling that have not been explored sufficiently. Unlike traditional causal estimation approaches which often result in "one best" causal structure that is learned, two properties of ABMs - *equifinality* (the ability of different sets of conditions or model representations to yield the same outcome) and *mutlifinality* (the same ABM might yield different outcomes) - can be exploited to learn multiple diverse "plausible causal models" from data. Using an illustrative example of news sharing on social networks we show how this idea can be applied to learn such causal sets. We also show how genetic algorithms can be used as a estimation technique to learn multiple plausible causal models from data due to their parallel search structure. However, significant computational challenges remain before this can be generally applied, and we, therefore, highlight specific key issues that need to be addressed in future work.

Keywords: causal inference, agent-based models, equifinality, mutlifinality

1. Introduction

Explanatory statistical models, and in particular *causal frameworks*, are powerful methodological tools (Pearl, 2019). Besides offering much needed explanatory insights using tested theoretical frameworks, they allow for exploiting the abundance of data toward a deeper understanding of diverse underlying phenomena. In some phenomena, however, such models often do not reach their full potential, mainly because of the rigidness imposed by the frameworks used to estimate them. For example, typical econometric estimation approaches offer possibilities for linear, polynomial or exponential model estimations, but they do not offer the same possibilities when it comes to estimating models of other – possibly ex-ante unknown – forms. In a large majority of phenomena, linear, polynomial or exponential models suffice to describe the underlying relationships with high precision (Chaibub Neto, 2020; Chan et al., 2010; Kiritoshi et al., 2021; Malinsky and Spirtes, 2018). Yet, there are phenomena that could benefit from more flexible explanatory models.

This need for higher flexibility has been recently identified in domains such as a infectious epidemiology (Germann et al., 2006; Hernán, 2015; Marshall and Galea, 2015). For example, epidemiologists have been trying to model the exact causal outcomes of pandemic mitigation strategies (e.g., social distancing measures) (Germann et al., 2006) which requires the ability to model complex relationships such as network effects within patient communities. Agent-based models allow for building very granular representations of the world (e.g., infected and non-infected individuals), with distinct heterogeneous behaviors (e.g., daily interaction of individuals with heterogeneous behavior). They can also be used to derive counterfactual scenarios using these realistic representations (e.g., the probability of an certain individual getting infected). If such models are (i) driven by theory and (ii) built to find the most accurate relationship between input and output data sets, they can serve as *causal* agent-based frameworks. However, such models include many non-linear relationships with feedback loops; therefore, estimating them with high precision requires advanced computational methods (Lamperti et al., 2018; Zhang et al., 2020).

ABM has traditionally been used to help build bottom-up models of complex phenomena (Grignard et al., 2013; Peters et al., 2018; Rand and Stummer, 2021; Wellman, 2016). However, the use of ABM as a means to derive causal frameworks is scarce (Istrate, 2021). Agent-based model estimation techniques known formally as *validation* or *calibration* (Oliva, 2003; Stonedahl and Rand, 2014; Stonedahl et al., 2011) to some extent resemble the causal framework estimation methods, as they aim at finding the most realistic calibration of a model, matching inputs and outputs. Yet, these approaches do not necessarily require the *adherence to a theoretical model*, which is one of the major elements of establishing causality (Shmueli and Koppius, 2011). As a result, frequently such approaches sacrifice model realism for computational ease. As Rand and Stummer (2021) point out, the *lack of causality* is a fair criticism against ABM.

We show how ABM can be used as a framework for building causal models by leveraging its modeling flexibility and ability to capture complex relationships. Especially given the advancement of ML, computational researchers can use their know-how toward building powerful *causal ABMs* that do not sacrifice realism for computational ease. Such a new pathway could allow researchers to explore possibly more complex underlying causal relationships in real-world phenomena leveraging large data sets. At the same time, estimating the parameters and structural forms of causal agent-based models requires the design of appropriate ML algorithms. Hence, the contribution of such work can be enriched by new ML algorithms proposed to estimate these models.

ABMs offer two unique ideas for causal modeling - *equifinality* (the ability of different sets of initial conditions or model representations to yield the same outcomes/data) and *mutlifinality* (the same ABM might create different outcomes/data in different runs). Given the difficulty of learning causal models from observational data, we suggest that it may be useful to develop approaches that offer the ability to learn *multiple plausible solutions* in a causal inference setting. Causal ABMs, as we show in this paper, present one approach to doing this.

This paper presents an approach for designing Causal ABMs and using genetic algorithms for its estimation - i.e. to learn multiple plausible causal sets from data. The benefits of a Causal ABM framework are (i) its expressive power (ii) its ability to present a realistic approach to learning causal models from data; specifically, a recognition that in many cases it might be possible to have not one, but multiple plausible causal explanations from observed data. As far as we know, this paper is the first to introduce the notion of using ABM-based modeling to learn multiple plausible causal models given data. The rest of the paper presents the approach using a specific example of learning causal models in the context of news sharing behavior on social networks. We use this example to illustrate the potential of our approach while identifying challenges that need to be addressed before this idea can be practically applied in larger settings.

2. ABM estimation and causal inference in the literature

Our work builds on many related ideas in the literature. Below, we summarize the key works, and note the terminologies used in these.

2.1 ABM and model abduction

Abduction is rooted in the theory of logic (Mayer and Pirri, 1996) and refers to the process of deriving a *reasonable* explanatory connection between inputs and outputs (Glass, 2019), as opposed to induction and deduction which are based on *logical* inferences to get from inputs to outputs. Based on deductive reasoning, if Y is derived by X, then Y is a "formal logical consequence" of X. In contrast, in inductive reasoning, if we can infer Y from X, it is quite probable that all instances that resemble Y are inferred by X, without this being the absolute norm. In this sense, deduction is often considered as going logically from general to specific, while induction is going from specific to general. Abductive reasoning, on the other hand, allows us to infer Y as a reasonable explanation for X. This indicates that there might be unobserved relationships connecting X and Y, and we can only infer some of those by observing the outcome of this connection $X \rightarrow Y$. In effect, abduction is, therefore, a process of considering alternative explanations given data and then choosing what seems to be "the best one". It has been noted that human decision making is often abductive in nature, and a statistical analogy is maximum likelihood estimation, except that this process of identifying the most likely explanation usually happens inside a human brain. Abductive reasoning has been used extensively in the fields of AI and knowledge representation (Boutilier and Beche, 1995) to provide plausible explanations of various phenomena.

Especially in ABM, abduction is being used to explain the outcomes of counterfactual scenarios, using logic programming (Pereira and Saptawijaya, 2016). Alberti et al. (2005) present an abductive logic programming framework, which can update the derived model dynamically based on upcoming facts and as a second step allows for hypotheses confirmation/rejection. For example, a realistic agent-based model of disease spread may have been calibrated on real data. This model can then be used to examine counterfactuals such as 'what would have happened in social distancing was in place earlier.' Along similar lines, Gavanelli et al. (2004) propose an abductive reasoning logic programming framework that provides a group of agents the same abductive semantics, and the agents using their own knowledge base are expected to make abductive inferences about a common goal.

Building on the abductive logic programming literature, Satoh et al. (2000) propose *speculative computation* based on abductive reasoning. They show that the agents in the multi-agent system can infer certain literals even using incomplete information sets, applying abductive reasoning. Similar abductive frameworks have been used in cyber-security (Karafili et al., 2018), defense-system algorithmic design (Das et al., 2011), or robotics and automation (Dennis et al., 2016). Karafili et al. (2018) propose the use of abductive reasoning on observed cyber-attack technical data and social evidence to infer the origin of a cyber-attack. In this work, whereas no claims about causality are made, by observing the argumentation (logical) rules that are used in their system it becomes clear that causality is the underlying assumption. Specifically, attributing an attack to an attacker entails a strong causal element. Das et al. (2011) use abduction to identify and reason about agent actions in an integrated air-defense system. In every simulation counterfactual scenario, the agent behavior and their environments are observed, and abductive claims are being made. With a slightly different objective, Dennis et al. (2016) use abductive reasoning to infer the rationale of an autonomous vehicle choice. Their goal is to show that when an autonomous vehicle makes a choice, there is reason to believe that this is the choice that causes the least ethical harm and, thus, it is chosen by the vehicle.

The most representative work of using abductive modeling to derive an agent-based model that is built on causal theories and matches reality is presented by Cedeno-Mieles et al. (2020). The authors build an agent-based model based on causal assumptions derived from theory, and in parallel run behavioral experiments. They combine the outputs of the experiments to adjust the model so that it matches the experimental output, and for this they use iterations of abduction. This has some flavors of *causal ABM*, which we define as ABM with strong theoretical framework and model estimation based on real-world data.

2.2 ABM and model learning

Model learning in ABM refers to observing inputs and outputs of an agent-based model to derive the best model that connects the two. Grimm et al. (2005), borrowing concepts from ecology, derive the underlying agent-based model not only by finding the best model that fits inputs and outputs, but also by falsifying alternative theories, attempting to establish causality in their resulting model. The work of Grimm et al. (2005) belongs to a broader family of solutions that is known as *inverse simulation*, based on which multiple “simulated worlds” are generated, and via comparing the outcomes of these worlds with real data, the researchers select the most realistic modeling representation (Kurahashi, 2018).

A different stream of literature (Chen and Liao, 2005; Janssen et al., 2019; Kvassay et al., 2017; Maes et al., 2003, 2007; Mao and Gratch, 2005, 2006, 2012; Nagoev et al., 2020; Wan and Singh, 2003; Wurzer and Lorenz, 2014) uses ABM to infer explanations for emergent phenomena arising from simulations. Often these explanations have a causal flavor. Chen and Liao (2005) design an agent-based model to replicate the stock market functionality, and by observing the output of the simulation model infer explanations about macro- or micro-phenomena that emerge. Janssen et al. (2019) present an agent-based methodology to conduct causal discovery of emergent phenomena via simulating and analyzing different scenarios. Kvassay et al. (2017) create an agent-based model to simulate a set of scenarios, and using causal inferences they try to derive the most plausible explanation to the outcomes of these scenarios. Similarly to Kvassay et al. (2017), Nagoev et al. (2020) infer reasoning from the agents actions and emergent interactions. Mao and Gratch (2005, 2006, 2012) present a method that infers causal beliefs from social interactions, and Wan and Singh (2003) infer causal commitments from agent interactions without proposing an explicit model learning.

In the articles belonging in the latter category, there is no explicit matching of the output of the model with real-world outputs. Therefore, this model learning could be considered implicit. A more explicit approach is followed by the works dealing with model estimation via calibration and validation, as presented below.

2.3 ABM and model calibration and validation

Model calibration in ABM, typically, refers to calibrating the parameters of a model based on some realistic conditions (Oliva, 2003). However, if this calibration is not conducted in a way that establishes a causal (and best-fit) relationship between inputs and outputs, the ABM model cannot be considered as causal. Such a non-causal approach is presented by Gilli and Winker (2003), who use an "indirect model estimation" method to calibrate their agent-based model so that it reflects the conditions of financial markets realistically. Similarly, Stonedahl et al. (2011), Stonedahl and Rand (2014) and Nguyen et al. (2019) propose genetic algorithms (GAs) to find the best calibration values for their agent-based model, so that it matches real output data. Oliva (2003) posits that such a model calibration can be seen as a hypotheses testing methodology, provided that the calibration is using a solid theoretical basis.

Bianchi et al. (2007) and Zhang and Vorobeychik (2019) connect the calibration of an agent-based model with the term *validation*. Model validation reflects this model calibration that makes the model "correct", in the sense of reflecting reality Bianchi et al. (2007). To this end, they find the best calibration of their model (validation) by matching the model-generated output with real-world output data. Similar validation processes are being presented by Rand and Rust (2011), who place particular emphasis on validation as a dimension of rigor in ABM. Windrum et al. (2007) present a critical review of ABM validation approaches. Naturally, the agent-based model calibration and validation process can be very computationally complex and sometimes intractable, as it has been pointed out by Gilli and Winker (2003) and Oliva (2003), among others. Thus, to address this limitation, Lamperti et al. (2018); Zhang et al. (2020) propose a set of ML surrogates that can explore the calibration space more efficiently and derive more accurately the set of parameters that establish causal relationships.

Summarizing the current literature, there is a need for building causal models, and existing ideas offer some directions in this context. While models consistent with data have been derived (best-fit function that connects inputs and outputs), we do not see derivations or learning of causal models purposefully.

2.4 Causal inference in the literature

Causal inference is a rich field with contributions stemming from disciplines spanning from computer science to economics. One of the most established causal inference methodologies deals with directed acyclic graphs (DAGs) (Dawid, 2010; Elwert, 2013; Knight and Winship, 2013; Pearl, 1998; Williams et al., 2018), which represent causal relationships between variables (nodes). DAGs most of the times assume specific types of relationships among the variables, which poses some expressive limitations in the kinds of causal structures that might be represented. Partially in response to this limitation, non-linear models have been proposed to model the relationships among variables in graphical models such as DAGs (Glymour et al., 2019) or new fuzzy directed graphs with feedback have been proposed (Osoba and Kosko, 2019). This stream of literature despite its longstanding presence is still facing challenges such as identifying the strength of a causal relationship (Janzing et al., 2013), identifying possibly hidden causal factors that are not properly modeled in a DAG (Dablander, 2020) or even computationally searching for the best DAG representation given raw data (Viinikka et al., 2020; Vowels et al., 2021). Recent advancements in causal modeling using DAGs include increase in efficiency by using "recursive Markov boundary-based causal structure learning" (Mokhtarian et al., 2021), interactive causal structure learning (Melkas et al., 2021), and latent causal structure learning (Young et al., 2020), among others.

The focus of this work is to provide a different approach to modeling causal relationships. Specifically, the objective of this work is to offer more "specific and informative formalism than its simpler (yet intuitive) graphical counterpart" (Vowels et al., 2021). As highlighted by Vowels et al. (2021) graphical models are intuitive and simple, but they might sacrifice expressive power (e.g., modeling the relationship between each variable with another in a more detail) for simplicity and macro-focus. With the proposed causal ABM, we aim to provide a framework for modeling complex dynamics and feedback loops among agents while non-linear learning causal relationships. Feedback loops are harder to model using DAGs (Strobl, 2019), and the ABM approach can address this limitation. However, the expressive power advantages offered by this approach do come with greater computational complexity in estimation, which we acknowledge here to be an issue that needs significant work in order for Causal ABM methodologies to become practical.

3. An Illustrative Framework

To motivate how we can approach causality in a richer sense with ABMs, here we present a framework that can learn *plausible causal models* in a setting where agents interact in a network and receive time- and agent-dependent feedback from other agents or themselves. Notationally, we assume agents $i \in [1, M]$ that have a set of attribute vectors $\mathbf{X^i} = \{x_1^i, ..., x_N^i\}$, where $x_1^i, ..., x_N^i$ are temporal vectors. Agents generate an output $\mathbf{Y}$ (agent observed behavior) while receiving environment-induced signals $\mathbf{E}$ and affected by time- or agent-dependent feedback.

For example, the network can be a social network where agents are connected, and the outcome of interest could be news sharing behavior. In this domain, prior knowledge from theory (from areas such as psychology, network science, and consumer behavior) can provide insights into sharing behavior of individuals; such theory needs to be explicitly modeled in causal frameworks as opposed to purely predictive ones. The agents interact in an environment from which they receive signals, such as a major news event outbreak. Agents can have a combination of static and dynamic attributes, such as gender, income, and propensity to share. As agents interact with each other and the environment and share information, data is constantly generated as a stream; such temporal data could even reflect endogenous interactions between the outcome of interest (news sharing) and the agent attributes or the presence of confounding factors leading to a causal outcome. In such a context, the problem of learning causal models is essentially being able to learn the underlying, theory-driven agent interactions that lead to the emergent data stream that is observed - i.e., learning the true data generating process. Interestingly, the *same* emergent data stream can be generated from different starting points; this is referred to as equifinality. A critical observation, therefore, is that multiple patterns of interactions - all consistent with theory - could generate the same observed data. This is our motivation for learning *plausible causal sets*, rather than seeking a *single* causal model as commonly done in the literature. As we show later in the paper, the learning such of such sets, or the estimation problem, is particularly interesting given one important observation. If we "seed" our ABM with *one* causal model driving the interactions, it is in fact possible that the same causal model leads to multiple datasets in different runs of the real world, due to stochasticity. Hence, estimation methods need to be robust in the sense of not "requiring" learned models to be fine-tuned only to the single observed temporal data instance (more on this in Section 5).

Formally, in such a model, the relationships between the input $\mathbf{X^i} = \{x_1^i, ..., x_N^i\}$, and output $\mathbf{Y}$ are captured by a function in the form of $Y = f(g^{(1)}(\mathbf{X^1}, .., \mathbf{X^M}), ..., g^{(z)}(\mathbf{X^1}, .., \mathbf{X^M})$, $\mathbf{E}$), where z is the number of sub-functions that capture inner-relationships among input attributes $\mathbf{X^i} = \{x_1^i, ..., x_N^i\}$ over time, as well as among input attributes and output over time (Figure 1). Each sub-function $g^{(1)}(\mathbf{X^1}$,.., $\mathbf{X^M}), ..., g^{(z)}(\mathbf{X^1}, .., \mathbf{X^M})$ captures certain relationships among any of the inputs $\mathbf{X^1}, .., \mathbf{X^M}$ or output $\mathbf{Y}$, driven by theory. Theory in psychology or network science can offer more than one reason $g(\mathbf{X^1}$,.., $\mathbf{X^M})$, e.g., how homophily influences sharing behavior, captured by $g^{(1)}(\mathbf{X^1}$,.., $\mathbf{X^M})$; and how behavior of influencers in the network affect sharing behavior, captured by $g^{(2)}(\mathbf{X^1}$,.., $\mathbf{X^M})$, etc. The function $f(\cdot)$ models all the interactions that subsequently take place for an individual to share a piece of news. The functions $g^{(1)}(\mathbf{X^1}$,.., $\mathbf{X^M}), ..., g^{(z)}(\mathbf{X^1}, .., \mathbf{X^M})$ can be parametrized in order to model functional, theory-driven rules, but the function $f(\cdot)$ can be as complex as needed to model all interactions that subsequently occur among agents and the environment that finally ends in some sharing behavior. In our case, the function $f(\cdot)$ is the ABM itself (i.e. the computational mechanism that creates the emergent outcome). It is tempting to learn the relationship, $Y = f(\mathbf{X^1}, .., \mathbf{X^M})$, $\mathbf{E}$), between inputs and outputs directly from data in a continuous manner using deep learning frameworks; however, such a model, while likely accurate and predictive, would not be causal. For causal inference, it is important to learn the actual data generating process that is playing out behind the scenes; which are the actual relationships $f(\cdot)$ and $g(\cdot)$ captured in a causal ABM framework.

Before providing more details, it is useful to ask how conventional causal modeling frameworks would consider such a setting. Traditional frameworks estimate a given causal model directly from the data using different approaches (e.g. panel estimation models, bayesian networks, etc.), which are also theory-driven. What do ABMs add in this context? We believe ABMs offer something unique which existing approaches do not - the ability to model highly complex and flexible interactions among agents and the environment, which capture not just the effects of the theoretical rules, but the effect of subsequently generated data on the behavior of agents in the next time period. At the same time, ABMs through their theory-driven rules can model endogenous relationships among variables and the effect of confounding factors on the final outcome (all this can be captured by the functions $f(\cdot)$

and $g(\cdot)$ but needs to be modeled explicitly). In addition, ABMs can include counterfactual analysis via simulating the ABM world without the presence of the causal rules, and comparing the outcome.

We define *Causal ABMs* as ABMs that are informed by theory and are consistent with observed data. For example, theory informs the way that inputs $\mathbf{X^1}, .., \mathbf{X^M}$ connect with the output $\mathbf{Y}$. Also, theory determines the way that environment-induced signals $\mathbf{E}$ affect $\mathbf{Y}$ (the signals $\mathbf{E}$ could capture the presence of confounding factors on the final outcome $\mathbf{Y}$). In addition, theory can specify the way that different attributes $x_1^i, ..., x_N^i$ of each agent can interact with one another or the way that past outputs (Y at $t-\kappa$) affect current outputs (Y at t). Therefore, a Causal ABM is a particularly good framework to model data generating processes exactly as they occur in practice. But, for an ABM to be *causal*, not only should it reflect theory, it should also be consistent with observed data, i.e. the theory-driven rules embedded in the ABM should actually generate the data that is observed in the real world. Learning such Causal ABMs from data is indeed challenging; this paper provides one approach and highlights the opportunities ahead for researchers.

Estimating causal ABMs shows a lot of parallels with either cross-sectional or panel data model estimations. In both ABM and econometric causal model estimations, a causal model that best expresses the relationship $\mathbf{Y} = f(\mathbf{X^1}, .., \mathbf{X^M}, \mathbf{E})$ is estimated. The main difference is found in the form of the function $f(\cdot)$, which in the econometric model estimations has a pre-specified structure, potentially less complex and flexible than the function $f(g^{(1)}(\mathbf{X^1}, .., \mathbf{X^M}), ..., g^{(z)}(\mathbf{X^1},..,\mathbf{X^M}), \mathbf{E})$. In causal ABM learning, the function $f(g^{(1)}(\mathbf{X^1}, .., \mathbf{X^M}),...,g^{(z)}(\mathbf{X^1}, .., \mathbf{X^M}), \mathbf{E})$ might not even have a closed-form representation; instead, it might be a sequence of if-then rules, ensembles of networks or other ML algorithms such as the approaches described by Cui et al. (2020). Hence, ensembles of ML can be used to learn the most *realistic and causal* ABM. Naturally, this flexibility and ability to model complex relationships comes at a estimation complexity cost. Therefore, each of these groups of methods are suitable for solutions with different complexity requirements. Particularly powerful in this framework is the ability of ABM approaches, due to their equifinality and multifinality properties, to learn multiple plausible sub-functions consistent with theory, that can lead to a much richer understanding of the causal phenomena.

4. Case Study: Sharing Political News

Estimating *causal ABMs* can be challenging depending on the examined scenario, and it requires specific estimation algorithm design. Through a case study example about an agent's decision to share political news depending on the sharing behavior in her social network, we explore estimation challenges and complexities.

4.1 Agent-based simulation preliminaries

We create a simulation in which agents $i \in [1, M]$ connected in a network decide to share or not a piece of political news in their social media. The decision to share political news is denoted as $Y_t^i \in \mathbf{Y}$ and is a binary decision that varies across agents i and across time t. In addition, our simulation randomly decides to connect two agents by creating a bidirectional edge between them. Furthermore, in our simulation there are some agents that are considered *influential* (e.g., influencers). These agents have the power to influence

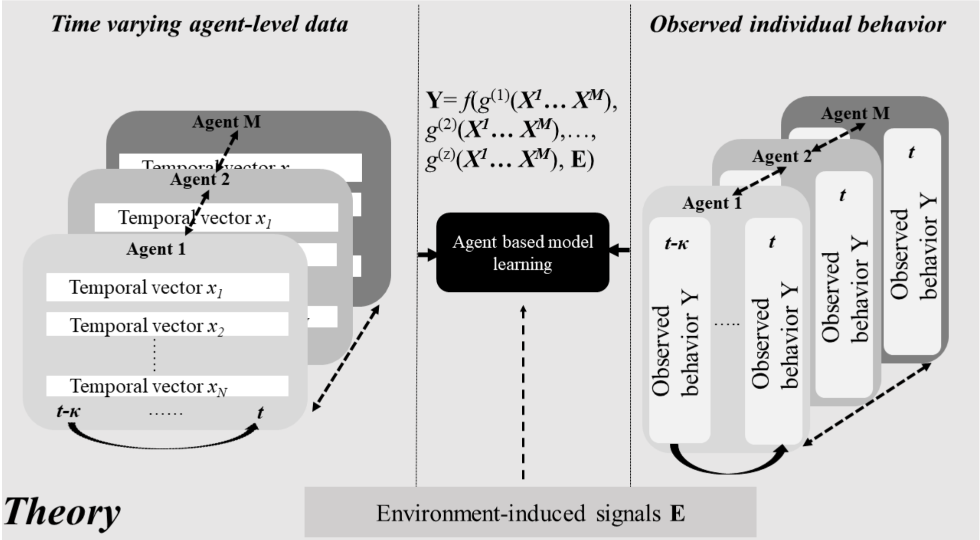
Time varying agent-level data
Observed individual behavior
Agent M
Agent 2
Agent 1
Temporal vector x_1
Temporal vector x_2
Temporal vector x_N
t-κ
t
Y= f(g^{(1)}(X^1… X^M), g^{(2)}(X^1… X^M),…, g^{(z)}(X^1… X^M), E)
Agent based model learning
Observed behavior Y
Theory
Environment-induced signals E

Figure 1: Causal ABM: inference of causal relationships from agent-level data with agent and time-dependent feedback

other agents connected with them in their decision to share political news in their social media. Finally, we assume that at random points in time, some exogenous events take place that are very popular on aggregate and this increases the probability of agents to share political news about these events. Such events represent aggregate signals induced by the environment surrounding the simulation (these signals are denoted by $\mathbf{E}$ in Figure 1).

4.2 Agent characteristics

The agents have certain attributes ($\mathbf{X^1}...\mathbf{X^M}$ in Figure 1). For expositional simplicity, we assume that the agents have as differentiating attribute their *endowment* $s_t^i \in \mathbf{S}$. Each agent's endowment s_t^i could change over time t or remain constant. This attribute could establish "similarity" (or homophily) between agents. Agents with homophily along the endowment dimension have a higher probability to influence one another in their decision to share political news in their social media if they are connected. Other characteristics that could establish homophily among two agents could be age, gender, geographical proximity, educational background, etc.

Second, we assume that past sharing decisions $Y_{t-\kappa}^i$ affect current decisions to share or not Y_t^i. To express agent heterogeneity not only in endowment and ability to be influential, but also in terms of decision making, we model agents that belong in two categories, A and B, with respect to the propensity to be affected by their past preferences in sharing. Agents belonging in category A have higher propensity to be affected by their past preferences as opposed to agents belonging in category B. The assignment of agents in the categories A and B takes place randomly once the simulation gets initialized.

Next, we present the theory-driven decision rules of agents that lead to their decision to share or not political news.

4.3 Theory-driven decision rules

For demonstration purposes, we assume that there are six rules that influence each agent's decision to share or not a piece of political news in their social media. These rules serve as an example; depending on the problem, domain experts must decided which rules are necessary or sufficient to model causal relationships. Ideally, such decisions can be accompanied with the use of real data. In the absence of real data coupled with knowledge about the exact causal underlying phenomena, rules such as the presented ones can be used to create synthetic data. In this example, the combination of the rules Eq. (1)-(6) are comprising the function $f(\cdot)$, whereas each of the Eq. (1)-(6) correspond to $g^{(1)}(\cdot)$-$g^{(6)}(\cdot)$ (following the notation of the presented causal ABM framework).

The first rule is called "homophily 1". Specifically, following the theory about homophily (Aral et al., 2009; Fang and Hu, 2018; Kossinets and Watts, 2009; Shalizi and Thomas, 2011), agents are influenced in their decision to share political news by agents that are homophilus to them. Based on the *threshold* model by Granovetter (1978), if the percentage of the connected neighbors of an agent i that have the same endowment and have shared a piece of political news during time t is surpassing a certain threshold $H1$ (e.g., 50%), then the agent i will also share this piece of political news during time $t+1$. We denote the set of connected neighbors of each agent i as $\mathbf{C^i}$, and the set of neighbors that have same homophily trait (endowment) with agent i and have shared the piece of news during time t as P_s^i, where s denotes the homophily trait "endowment". Since in this example endowment is the only homophily trait, for presentation clarity we omit the index s from P^i. Then, this rule is:

$$Y_{t+1}^i = \begin{cases} 1, & \text{if } \frac{|P^i|}{|\mathbf{C^i}|} > H1 \\ 0, & \text{else} \end{cases}, \quad \forall\ i \in [1, M] \tag{1}$$

In addition, we implement the decision rule "homophily 2". Based again on the theory about homophily (Aral et al., 2009; Fang and Hu, 2018; Kossinets and Watts, 2009; Shalizi and Thomas, 2011) and the threshold model Granovetter (1978), if the percentage of an agent's i connected neighbors that have shared this piece of news during time t surpasses a certain threshold $H2$, then agent i will also share this piece of news during time $t+1$. Mathematically, this rule is expressed as:

$$Y_{t+1}^i = \begin{cases} 1, & \text{if } \frac{\sum_j Y_t^j}{|\mathbf{C^i}|} > H2 \\ 0, & \text{else} \end{cases}, \quad \forall\ j \in \mathbf{C^i},\ \forall\ i \in [1, M] \tag{2}$$

The sharing decision is also known to be affected by social influence (Shalizi and Thomas, 2011). Hence, for every time t some agents are randomly chosen to be *influential.* These influential agents, in case they have decided to share a piece of political news, are likely to influence their connected agents with a certain probability p_t^{infl} to also share this piece of news. In the social influence literature (Fang and Hu, 2018; Rice et al., 1990; Shalizi and Thomas, 2011), the probability of influence can be measured by the frequency of social interactions or any other interaction proxy. If $Ber(p_t^{\text{infl}})$ denotes a Bernoulli random variable with probability p_t^{infl} to be 1, the above influence rule is expressed as:

$$Y_{t+1}^i = \begin{cases} \sim Ber(p_t^{\text{infl}}), & \text{if } j = influential \\ 0, & \text{else} \end{cases}, \forall\, j \in \mathbf{C^i},\ \forall\, i \in [1, M] \tag{3}$$

Furthermore, agent past preferences or decisions about sharing influence their current decision to share Y_t^i. According to preference theory (Hanley et al., 2006; Rafailidis and Nanopoulos, 2014), user preferences are relatively stable over time and change as a result of exogenous events. Thus, in our simulation, we assume that agents i, depending on whether they belong in category A or B, are influenced by their past preferences with a probability p^A or p^B. However, after a critical point in time t_c, the probability of sharing the piece of news is not anymore determined by the category, but becomes proportional to the times they have shared a piece of news before. The above rule is summarized as:

$$Y_{t+1}^i \sim \begin{cases} Ber(p^A), & \text{if } \ i \in \mathbf{A}, \text{and } \ t \le t_c \\ Ber(p^B), & \text{if } \ i \in \mathbf{B}, \text{and } \ t \le t_c \ , \ \forall \ \in [1, M] \\ Ber(\frac{\sum_t Y_t^i}{t}), & \text{if } \ t > t_c \end{cases} \tag{4}$$

In terms of exogenously-induced signals $\mathbf{E}$, some events are "viral" leading to sharing by more agents in the network. According to Salganik et al. (2006), some events are much more influential than others leading to a cascade of influence in a network. Such a cascade follows the "rich-get-richer" propagation theory (Easley et al., 2010) also known as the "Matthew Effect" (Rigney, 2010), based on which if the percentage of agents that has shared the news during time t surpasses a threshold V in the whole network, then this piece of news is considered viral and all agents in the simulation will share it with a probability p^v during $t+1$. This rule is described by Eq. (5).

$$Y_{t+1}^i = \begin{cases} \sim Ber(p^v), & \text{if } \ \frac{\sum_t Y_t^i}{M} > V \\ 0, & \text{else} \end{cases} , \ \forall \ i \in [1, M] \tag{5}$$

Finally, we model sporadic exogenous events that we call "rare events". When these take place, then the whole agent network is likely to share the piece of news about such events with a probability p^r. Similarly to viral events, these rare events also follow the the "rich-get-richer" propagation theory (Easley et al., 2010; Rigney, 2010; Salganik et al., 2006). We denote the presence of rare events as $t = t_{\text{rare}}$ and this rule is:

$$Y_{t+1}^i = \begin{cases} \sim Ber(p^r), & \text{if } \ t = t_{\text{rare}} \\ 0, & \text{else} \end{cases} , \ \forall \ i \in [1, M] \tag{6}$$

The last two rules express the effect of exogenous environment-induced signals $\mathbf{E}$ to the final outcome $\mathbf{Y}$. For example, the probability of a rare event (for example a rare shooting incident) is directly influenced by the environment, and as such these rules indirectly express the effect of $\mathbf{E}$ on the $\mathbf{Y}$.

5. Estimation Method

The estimation problem in a Causal ABM framework can be stated informally as follows. Given (i) temporal data, and (ii) some knowledge about the underlying dynamics within the ABM, identify the exact data generating process inside the ABM that generated the observed data. Of course, with only observed data and no knowledge or assumptions the estimation problem is intractable. Based on problem settings, specific versions of the estimation problem may be proposed, that differ in what kind of prior knowledge is assumed. Below we discuss this in the context of the case study.

5.1 Preliminaries

In order to estimate the causal agent-based model described in section 4, first, a set of environment inputs should be provided. These inputs are known beforehand and describe the agent-based environment within which agents make decisions. In our case study, these environment inputs are the number of agents M, the number of connected neighbors $\mathbf{C}^i$ of each agent i, the endowment s^i of each agent i, the preference category A or B of each agent i and the critical time threshold t_c. We denote this input set as $\mathbf{I} = \{M, \mathbf{C}^i, s^i, \mathbf{A}, \mathbf{B}, t_c\}$.

In combination with the above inputs, the output $\mathbf{Y} = \{\sum_{i=1}^{M} Y_t^i\}$, expressed as the aggregate decisions of all agents M to share or not the piece of news over time, is known. This output $\mathbf{Y}$ can be the result of any of the decision rules described by Eq. (1) - (6); however, the rule that caused sharing or not sharing is not known.

Finally, the set of causal rules described by Eq. (1) - (6) are known but without their exact parameters. That is, using theory, a researcher can derive the general structure of these rules; however, the exact parameters that connect the known inputs and outputs need to be discovered. For example, in Eq. (1) the threshold $H1$ is the unknown parameter. Similarly, in Eq. (2) the threshold $H2$, in Eq. (3) the probability p^{infl}, in Eq. (4) the probabilities p^A and p^B, and in Eq. (5) and (6) the probabilities p^v and p^r are the unknown parameters that need to be estimated. Denoting these unknown parameters as $\mathbf{K} = \{H1, H2, p^{\text{infl}}, p^A, p^B, p^v, p^r\}$, the set of causal rules can be expressed as: $R(\mathbf{K})$. And, therefore, assuming that the agent-based model is $f(\cdot)$, we have that $\mathbf{Y} = f(\mathbf{I}, R(\mathbf{K}))$. The objective of this agent-based model estimation is to find the parameter set $\mathbf{K}$ that satisfies $\mathbf{Y} = f(\mathbf{I}, R(\mathbf{K}))$.

We refer to each realization of $\mathbf{Y}$, as a result of causal rules $R(\mathbf{K})$, as a 'world'. It should be noted here that the agent-based model $f(\mathbf{I}, R(\mathbf{K}))$ can generate more than one "worlds", described by $\mathbf{Y}$. In other words, for a set of inputs $\mathbf{I}$ and a set of rules $R(\mathbf{K})$ more than one output vectors $\mathbf{Y}$ are possible. The latter is coined in the literature as *multifinality* (Chaturvedi et al., 2011) and has been elaborated in section 3. Equivalently, more than a set of inputs $\mathbf{I}$ and a set of rules $R(\mathbf{K})$ can yield the same world $\mathbf{Y}$, also known as *equifinality*.

Furthermore, we denote as $f(\mathbf{I}, R(\mathbf{K}))$ one "run" of the ABM, whereas as $f_W(\mathbf{I}, R(\mathbf{K}))$ W "runs" of the ABM. The latter means that for the same inputs $\mathbf{I}$, same rules $R(\mathbf{K})$, the whole agent-based model is run W times. Similarly to one ABM run which is characterized by *multifinality* and *equifinality*, the W ABM runs have the same characteristics.

5.2 Evaluation Metric

As a direct consequence of the above-mentioned *multifinality* and *equifinality*, an evaluation metric that is able to measure the "goodness" of each "world" $\mathbf{Y}$ should be established. For this purpose, we set as $D(\mathbf{Y^1}, \mathbf{Y^2})$ a distance function that can measure the distance between two "worlds" $\mathbf{Y^1}$ and $\mathbf{Y^2}$. The distance metric $D(\cdot)$ could be comparing the worlds in as granular or coarse a manner as acceptable for the problem under consideration (e.g., the worlds could be compared based on distribution of total user shares of news across time, or the actual daily shares). The choice of the distance metric $D(\cdot)$ could be another possible contribution, as the right choice of $D(\cdot)$ allows for faster convergence of the estimation to the desired parameter set $\mathbf{K}$ that satisfies $\mathbf{Y} = f(\mathbf{I}, R(\mathbf{K}))$.

5.3 Estimation Algorithm

In order to estimate the parameter set $\mathbf{K}$ that satisfies the causal agent-based model $\mathbf{Y} = f(\mathbf{I}, R(\mathbf{K}))$ researchers can propose a variety of methods depending on the nature and the complexity of the problem under consideration. In this paper, we offer an estimation algorithm example based on Genetic Algorithms (GAs) due to its parallel search structure that naturally lends itself to learning multiple solutions (plausible causal sets). It should be noted that the estimation algorithm serves as an example of how causal structures can be learned in the proposed causal ABM framework. Researchers can expand on this direction and propose novel contributions with regards to learning causal structures in agent-based models. A summary of the proposed estimation algorithm is presented in Algorithm 1.

Step 1 receives inputs $\mathbf{I}$ which are known ex-ante, as well as the output $\mathbf{Y^1}$ which serves as *ground truth*. In other words, the objective of the GA is to find the parameter rules that generate a world as close as possible to $\mathbf{Y^1}$. Step 2 finds an initial set of rule parameters $\mathbf{K}$ as a result of either one model run $f(\mathbf{I}, R(\mathbf{K}))$ or W ABM model runs $f_W(\mathbf{I}, R(\mathbf{K}))$ ($f_W(\mathbf{I}, R(\mathbf{K}))$ produces a matrix containing the outcomes of all W runs). Step 3 includes a set of steps that are repeated until the GA produces a satisfactory solution (expressed in the termination condition). In this step the GA continuously evaluates the generated solutions against the ground truth $\mathbf{Y^1}$, and through crossover operations and a population maintenance policy creates a new population of solutions $\mathbf{Q}$. In the end, in Step 4, the algorithm returns the set of solutions that have an acceptable distance from the ground truth $\mathbf{Y^1}$. To establish the acceptable distance from the ground truth we set a $fitness\_threshold$, which can be determined exogenously by the algorithm designer.

As mentioned previously, the choice of the distance function is crucial for the discovery of the underlying causal ABM. In the presented case study, there is substantial stochasticity in the environment of the agents which poses significant challenges to the estimation of the true causal ABM. This stochasticity stems from many behavioral attributes of the agents, for example, the initialization of the input parameters $\mathbf{I} = \{M, \mathbf{C}^i, s^i, \mathbf{A}, \mathbf{B}, t_c\}$ is done stochastically, and in addition, in every time t different agents are defined as influential, increasing the stochastic nature of the case study. Therefore, we set our fitness function as the L2 norm of the difference between ground truth $\mathbf{Y^1}$ and W ABM runs, $D(\mathbf{Y^1}, f(\mathbf{I}, R(\mathbf{K}))) = ||\mathbf{Y^1} - \frac{\sum_{(w=1)}^{W} f_W(\mathbf{I}, R(\mathbf{K}))}{W}||_2$. L2 norms are commonly used in GA algorithms to guide the algorithm to convergence (Stonedahl et al., 2011). However, in our case study we are benefiting from multiple ABM runs, as opposed to one. Such a function, using the average output of W ABM runs is more accurate in discovering the actual causal ABM structure, as opposed to a single ABM run, because it can eliminate part of the stochastic noise that is injected in the GA algorithm via the stochastic initialization and behavior of the agent population. It is important to mention that while W increases, the convergence of the GA takes place faster, as well. However, this comes at a computational cost, therefore, researchers need to identify the right balance between accuracy and computational complexity while estimating the causal ABM structure, and as a result the sets $\mathbf{K}$. Future work can produce meaningful contributions in this direction, because for example, there are rules that might influence the outcome more than others; hence, more elaborate distance functions could achieve faster convergence.

Algorithm 1 Summary of the Proposed Estimation Algorithm

```
Step Description
1 Receive the input parameter set $\mathbf{I}$ and the world $\mathbf{Y^1}$
2 Initialize a population $\mathbf{Q}$ of candidate solutions
  $\mathbf{K}$, where each candidate solution is a specific
  binding for the parameters of the causal rules.
  $\mathbf{Q}$ can contain the result of $W$ ABM model runs $f_W(\mathbf{I}, R(\mathbf{K}))$
3 While termination condition is met:
    a. For each candidate solution $\mathbf{K}$ in the population $\mathbf{Q}$
       compute fitness as $fitness(\mathbf{K}) = D(\mathbf{Y^1}, f_W(\mathbf{I}, R(\mathbf{K}))=$
       $||\mathbf{Y^1} - \frac{\sum_{(w=1)}^{W} f_W(\mathbf{I},R(\mathbf{K}))}{W}||_2$
    b. Use the fitness values $fitness(\mathbf{K}) \ \forall \ \mathbf{K} \in \mathbf{Q}$ to
       probabilistically select parents $\mathbf{K^1}$ and $\mathbf{K^2}$ for
       a crossover operation
    c. Create new candidate solutions $\mathbf{L^1}$, $\mathbf{L^2}$ based on a
       crossover and compute $fitness(\mathbf{L^1})$ and $fitness(\mathbf{L^2})$
    d. Insert children into the population and update the
       population by removing the worst solutions
    e. With some random probability mutate a randomly
       chosen solution
    f. If the average fitness function is not decreasing,
       perform diversity boosting by replacing a
       share of the population with new solutions
    g. Update $\mathbf{Q}$ with the set of new solutions
4Return plausible causal sets $\mathbf{R} = \{\mathbf{K} | fitness(\mathbf{K})$
  $<< fitness\_threshold\}$
```

GAs have been used in the ABM literature to calibrate or validate models in order to match real-data (Calvez and Hutzler, 2005; Nguyen et al., 2019; Stonedahl and Rand, 2014; Stonedahl et al., 2011), without, however, the constraint of causality in mind. The latter adds extra complexity in the model calibration process but it allows for making causal inferences using the proposed ABM. Furthermore, in our approach, we are interested in finding sets of plausible causal parameters so that the mechanism behind the ABM is revealed, as opposed to finding the "best-fit" parameters that optimize a model, which is the objective of model validation methods. As Rand and Stummer (2021) highlight, causality is missing from the current model validation or calibration methods in ABM, and is the natural next step to enhance our understanding behind emergent phenomena.

Note, here, that in the presented example, for expositional simplicity, we have not included a counterfactual analysis (which is common in causal modeling). Incorporating a counterfactual analysis in causal ABMs can be done by simulating the ABM world without the presence of the causal rules, represented by $f(\cdot)$ and $g(\cdot)$ or in Eq. (1)-(6) in our specific example, and adjust the distance function to measure the proximity or (lack thereof) of the counterfactual outcome. More in-depth counterfactual analyses can examine the impact of specific rules (Eq. (1)-(6)) on the outcome.

5.4 Baseline

As commonly done with evolutionary algorithms, to evaluate the presented approach, we implemented a random generator benchmark. This benchmark serves as our baseline and generates equal number of candidate solutions $\mathbf{K}$ as the ones generated by the proposed approach. If τ the number of solutions generated by the proposed GA approach, the benchmark algorithm behaves as shown in Algorithm 2.

Algorithm 2 Summary of the Baseline Algorithm

```
Step Description
1 Receive the number of solutions, $\tau$, generated by the
proposed GA approach
2 For $\tau$:
    initialize random solution vector $\mathbf{K}$, where $\mathbf{K}$ is a
    specific binding for the parameters of the causal rules.
3 Return sets $\mathbf{R} = \{\mathbf{K} | fitness(\mathbf{K}) << fitness\_threshold\}$
```

6. Estimation Results

6.1 Preliminaries and Simulation Environment

Next, we present the results of our proposed estimation approach and we compare them to the results of the baseline. The objective of the proposed estimation algorithm and the baseline is to estimate the true causal rule parameters denoted as $\mathbf{A} = \{H1, H2, p^{\text{infl}}, p^{A}, p^{B}, p^{v}, p^{r}\}$. We denote as $\hat{\mathbf{A}} = \{\hat{H}1, \hat{H}2, \widehat{p^{\text{infl}}}, \widehat{p^{A}}, \widehat{p^{B}}, \widehat{p^{v}}, \widehat{p^{r}}\}$ the parameters estimated either by the proposed GA-based approach or the baseline. The sets $\hat{\mathbf{A}}$ are the sets in $\mathbf{R} = \{\mathbf{K}\}$ with the lowest fitness. To quantify the comparison between the actual and the estimated parameter set, we use the euclidean distance: $\{(H1 - \hat{H}1)^2 + (H2 - \hat{H}2)^2 + (p^{\text{infl}} - \widehat{p^{\text{infl}}})^2 + (p^{\text{A}} - \widehat{p^{\text{A}}})^2 + (p^{\text{B}} - \widehat{p^{\text{B}}})^2 + (p^{v} - \widehat{p^{v}})^2 + (p^{r} - \widehat{p^{r}})^2\}^{\frac{1}{2}}$.

In this simulation run, we have $M = 23$ agents (randomly generated number in each simulation). These agents interact with each other for 23 days (randomly generated number). The number of agents and the number of days of the simulation are randomly drawn. The agent characteristics are also randomly initialized. Figure 2 displays the sharing behavior of some agents of the simulation over the 23 day horizon. We observe a diverse agent behavior: agent 8 during the first day does not share the piece of news, but after day 1 she keeps on sharing until the end of the simulation; agent 21 starts with sharing and later this behavior changes, finishing the simulation without sharing the piece of news. Agent 3 also changes behavior between sharing and not sharing, possibly influenced by more than one causal rules. We should note here that in our simulation, more than one rules might cause sharing, making the estimation challenge more complex.

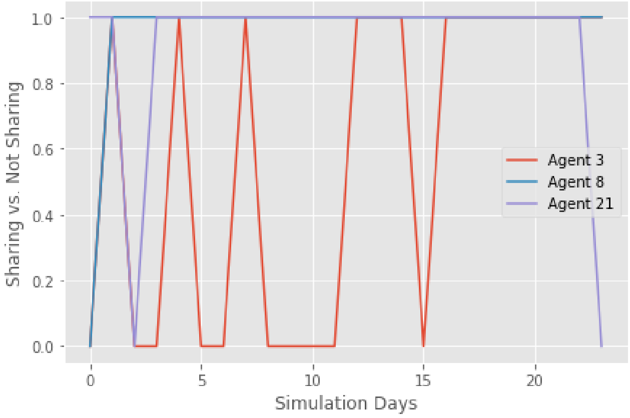

Figure 2: Sharing behavior of some randomly selected agents.

6.2 Learning Causal Sets

The proposed estimation's objective is to learn the causal rule parameters of the agents that lead to sharing or not a piece of news ($\mathbf{Y}$). The actual rules that drive the agent behavior as well as the best learned rules by the proposed algorithm are presented in Table 1. The rules in Table 1 have been shortened due to space limitations, and their full expressions can be found in Eq. (1) - (6). The parameters of interest are presented in bold.

In addition, we present the best 3 results generated by the proposed algorithm and the baseline in Tables 2 and 3. The baseline algorithm generated $\tau = 735$ causal sets $\mathbf{K}$, equal to the number of causal sets generated by the GA-based approach. The solutions in both Tables 2 and 3 are sorted based on their ability to generate an output closer to the ground truth $\mathbf{Y^1}$. We should note that the Euclidean Distance is comparing the causal parameters and not their generated output.

The proposed approach outperforms the baseline in learning the causal sets. Specifically, the best estimated causal set deviates only by 1.25 units of Eucl. Dist. from the true set, whereas the best estimated causal set by the baseline deviates by much more from the actual parameters, i.e., 8.3 units of Eucl. Dist. Also, on average the best 3 solutions estimated by the proposed method deviate by 1.42 units from the actual parameters, whereas the respective average deviation of the baseline is 6.38 units.

6.3 Multifinality Observations

Because of the many stochastic factors in our simulation, which are meant to mimic the real world, we notice substantial presence of multifinality. As a proxy for multifinality, we use the fitness function defined to ensure convergence of the GA-based algorithm. Specifically, we use the actual causal parameter set $\mathbf{A} = \{0.88, 22, 0.70, 0.90, 0.10, 0.90, 0.80\}$ to run $W = 50$ "worlds" using our simulation. The initialization of the simulation remains the same (agent network, neighbors, attributes, etc), and only the stochastic factors change.

Table 1: Sharing Political News: Decision Rules

Actual Rule	Estimated Rule
$\mathrm{Y}^i_{t+1} = \begin{cases} 1, & \text{if } \frac{\lvert P^i \rvert}{\lvert \mathbf{C^i} \rvert} > \mathbf{0.88} \\ 0, & \text{else} \end{cases}$	$\mathrm{Y}^i_{t+1} = \begin{cases} 1, & \text{if } \frac{\lvert P^i \rvert}{\lvert \mathbf{C^i} \rvert} > \mathbf{0.61} \\ 0, & \text{else} \end{cases}$
$\mathrm{Y}^i_{t+1} = \begin{cases} 1, & \text{if } \frac{\sum_j Y^j_t}{\lvert \mathbf{C^i} \rvert} > \mathbf{22} \\ 0, & \text{else} \end{cases}$	$\mathrm{Y}^i_{t+1} = \begin{cases} 1, & \text{if } \frac{\sum_j Y^j_t}{\lvert \mathbf{C^i} \rvert} > \mathbf{21} \\ 0, & \text{else} \end{cases}$
$\mathrm{Y}^i_{t+1} = \begin{cases} \sim Ber(\mathbf{0.70}) \\ 0, \end{cases}$	$\mathrm{Y}^i_{t+1} = \begin{cases} \sim Ber(\mathbf{0.90}), & \text{if } j = infl. \\ 0, & \text{else} \end{cases}$
$\mathrm{Y}^i_{t+1} \sim \begin{cases} Ber(\mathbf{0.90}) \\ Ber(\mathbf{0.10}) \\ Ber(\frac{\sum_t Y^i_t}{t}) \end{cases}$	$\mathrm{Y}^i_{t+1} \sim \begin{cases} Ber(\mathbf{0.90}), \text{ if } i \in \mathbf{A}, \& \ t \leq t_c \\ Ber(\mathbf{0.29}), \text{ if } i \in \mathbf{B}, \& \ t \leq t_c \\ Ber(\frac{\sum_t Y^i_t}{t}), \text{ if } t > t_c \end{cases}$
$\mathrm{Y}^i_{t+1} = \begin{cases} \sim Ber(\mathbf{0.90}), \\ 0, \end{cases}$	$\mathrm{Y}^i_{t+1} = \begin{cases} \sim Ber(\mathbf{0.27}), & \text{if } \frac{\sum_t Y^i_t}{M} > V \\ 0, & \text{else} \end{cases}$
$\mathrm{Y}^i_{t+1} = \begin{cases} \sim Ber(\mathbf{0.80}) \\ 0, \end{cases}$	$\mathrm{Y}^i_{t+1} = \begin{cases} \sim Ber(\mathbf{0.93}), & \text{if } t = t_{\text{rare}} \\ 0, & \text{else} \end{cases}$

Table 2: Best 3 GA Solutions

	$H1$	$H2$	p^{infl}	p^A	p^B	p^v	p^r	Eucl. Dist.
Actual Causal Parameters	0.88	22	0.70	0.90	0.10	0.90	0.80	
Best 3 GA Solutions	0.61	21	0.90	0.90	0.29	0.27	0.93	1.25
	0.22	21	0.90	0.90	0.29	0.27	0.14	1.53
	0.22	21	0.90	0.90	0.29	0.07	0.93	1.49
Mean Eucl. Dist.								**1.42**

Table 3: Best 3 Baseline Solutions

	$H1$	$H2$	p^{infl}	p^A	p^B	p^v	p^r	Eucl. Dist.
Actual Causal Parameters	0.88	22	0.70	0.90	0.10	0.90	0.80	
Best 3 Baseline Solutions	0.26	14	0.63	0.85	0.03	0.79	0.43	8.03
	0.74	15	0.56	0.96	0.07	0.73	0.45	7.05
	0.26	18	0.74	0.91	0.12	0.48	0.69	4.07
Mean Eucl. Dist.								**6.38**

Calculating the fitness function of these runs compared with the ground truth $\mathbf{Y^1}$, we get: $fitness(\mathbf{A}) = D(\mathbf{Y^1}, f(\mathbf{I}, R(\mathbf{A}))) = ||\mathbf{Y^1} - \frac{\sum_{(w=1)}^{50} f_W(\mathbf{I}, R(\mathbf{A}))}{50}||_2 = 1.80$. This shows, that even with the actual causal rules known, the "simulation world" is not identical, if it is run many times.

Similarly, we evaluate the learned sets along this dimension. Specifically, we use the best 3 sets learned by the proposed GA-based approach and the best 3 solutions estimated by the baseline as parameters in the described simulation to simulate $W = 50$ worlds. Next, we compare their output (sharing behavior) in terms of fitness. The results are shown in Tables 4 and 5. We observe that the parameters learned by our proposed solution have the same fitness as the actual causal set, indicating an excellent performance of the proposed GA-based approach. Furthermore, these results show the estimation challenges in deriving causal sets (and a as result causal rules) from real-world data; even when the actual causal sets are known, the realization of the "world" is not identical.

Table 4: Fitness of the Best 3 GA Solutions

	$H1$	$H2$	p^{infl}	p^A	p^B	p^v	p^r	Fitness
Actual Causal Parameters	0.88	22	0.70	0.90	0.10	0.90	0.80	**1.80**
Best 3 GA Solutions	0.61	21	0.90	0.90	0.29	0.27	0.93	1.80
	0.22	21	0.90	0.90	0.29	0.27	0.14	1.82
	0.22	21	0.90	0.90	0.29	0.07	0.93	1.82
Mean Fitness								**1.81**

Table 5: Fitness of the Best 3 Baseline Solutions

	$H1$	$H2$	p^{infl}	p^A	p^B	p^v	p^r	Fitness
Actual Causal Parameters	0.88	22	0.70	0.90	0.10	0.90	0.80	**1.80**
Best 3 Baseline Solutions	0.26	14	0.63	0.85	0.03	0.79	0.43	1.99
	0.74	15	0.56	0.96	0.07	0.73	0.45	2.02
	0.26	18	0.74	0.91	0.12	0.48	0.69	2.05
Mean Fitness								**2.02**

This presence of stochasticity is also captured in the convergence of the algorithm. As shown in Figure 3, the fitness of the best solution in the GA-based approach does not improve below 1.80. Specifically, the algorithm reaches solutions with $fitness = 1.80$ around the 300^{th} iteration, and afterwards starts performing diversity boosts (spikes in the average fitness graph) in order to possibly discover solutions with lower fitness. However, till the end of the estimation procedure, the fitness does not improve below 1.80. As explained previously, even multiple runs with the actual causal rules give a fitness of 1.80 because of the stochasticity present in the simulation.

7. Discussion

This paper introduced the idea of *Causal ABMs* and presented one approach to learning these in the context of a case study. In the case study, we showed how genetic algorithms

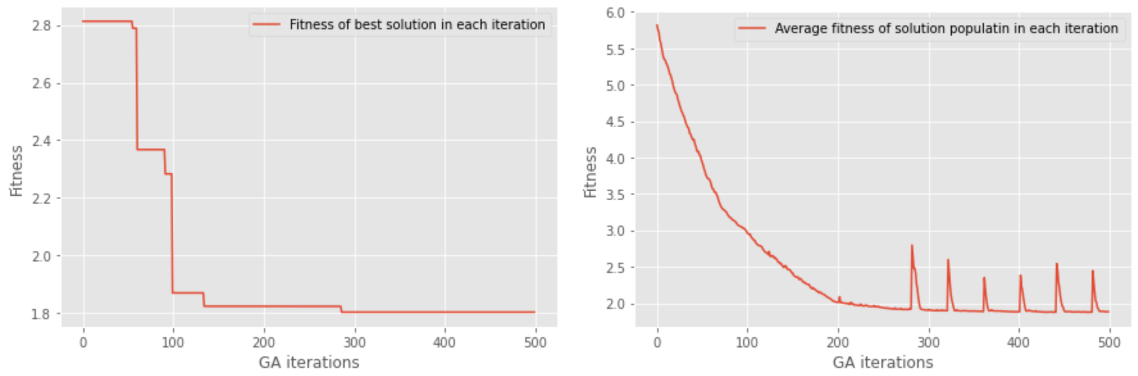

Figure 3: Fitness of best solution and average fitness of the solution population in each GA iteration

can be used to learn multiple plausible causal sets based on data generated from the ABM that embodied specific rules. More generally, causality is a key area for research, and there are many open questions and challenges regarding learning causal ABMs given data. The flexibility of ABMs to model real-world data generating processes is particularly appealing from the perspective of expressive power of causal models. This, combined with effective computational techniques for estimation, can generate fundamentally new ideas for learning causal models from data.

Yet, as our paper highlighted throughout, there are significant challenges and hurdles that need to be overcome. We conclude, here, with three important ones to be addressed in future work. First, in contrast to standard econometric model estimation, the computational complexity of the estimation method poses unique challenges. For instance, one approach to learn the plausible Causal ABMs, here, would be to launch multiple ABMs with varying sets of causal rules with specific parameters, and comparing the world(s) generated in each of those cases with the actual outcomes observed in data. Hence, this opens fruitful pathways for researchers to devise ML algorithms that are able to estimate complex *causal ABMs* without suffering from high computational complexity. Second, because of the stochasticity present in realistic agent-based simulations, there might be more than one possible causal ABM that generates the same output $\mathbf{Y}$. As noted earlier, this is known as *equifinality*, or the ability of different set of conditions or model representations to yield the same outcome (Chaturvedi et al., 2011). Here, researchers can devise new metrics that assess the "goodness" of a causal ABM, as well as algorithms that use such metrics to learn not one, but multiple plausible causal ABMs. Philosophically this is consistent with the notion that the real world is just one realization of what could have happened due to a combination of causal factors with random components; in such an interpretation it is often possible to have multiple possible causes that could have resulted in the same outcome. Third, the same causal ABM might yield different results (multifinality) (Chaturvedi et al., 2011) in different runs, where some realizations are closer to the "actual" observed data while others are less so. This does not mean that the estimations are wrong; it instead, means that we need new ways of thinking about such occurrences. This phenomenon is also related to what has recently been coined as *the red ribbon* phenomenon.[1] The *the red ribbon* schematically captures the set of possible outcomes in data - even if the observed data only reflects some of these. One implication is that we might need to re-think the use of optimizing specific distance functions in order to learn causal ABMs, and instead design new distance functions that have the flexibility and capture the semantic complexities of what should be considered "good match" with the true underlying data generating process.

In addition, the use of GAs, in a way, demonstrates the complexity of estimating suitable parameters for ABM calibration or validation, and the need for designing appropriate meta-heuristics. In our case, where causality is imposing an additional, possibly hard to satisfy constraint, such complexity issues can become more central. However, such a challenge opens important pathways for research contributions, as computational researchers have the expertise to design novel meta-heuristics, possibly leveraging ML (Cui et al., 2020), to make such causal estimation processes more efficient. Another angle that provides interesting paths for future contributions is choice of fitness functions, something that has also been explored in model validation approaches (Calvez and Hutzler, 2005; Stonedahl and Rand, 2014; Stonedahl et al., 2011). Finally, while this approach was motivated here by the need for more flexibility and expressive power, and therefore presented as an alternate framework, future work is needed to explore stronger connections between *causal ABMs* and other causal frameworks (Pearl, 2010; Sekhon, 2008; VanderWeele et al., 2016).

References

Marco Alberti, Marco Gavanelli, Evelina Lamma, Paola Mello, Paolo Torroni, et al. Abduction with hypotheses confirmation. In *International Joint Conferences on Artificial Intelligence(IJCAI)*, volume 2005, pages 1545–1546. Citeseer, 2005.

Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009.

Carlo Bianchi, Pasquale Cirillo, Mauro Gallegati, and Pietro A Vagliasindi. Validating and calibrating agent-based models: a case study. *Computational Economics*, 30(3):245–264, 2007.

Craig Boutilier and Veronica Beche. Abduction as belief revision. *Artificial intelligence*, 77 (1):43–94, 1995.

Benoît Calvez and Guillaume Hutzler. Automatic tuning of agent-based models using genetic algorithms. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 41–57. Springer, 2005.

Vanessa Cedeno-Mieles, Zhihao Hu, Yihui Ren, Xinwei Deng, Abhijin Adiga, Christopher Barrett, Noshir Contractor, Saliya Ekanayake, Joshua M Epstein, Brian J Goode, et al.

[1] Yann LeCun, Director of AI Research at Facebook, and Professor at New York University, source: https://bit.ly/2NxUOCF

Networked experiments and modeling for producing collective identity in a group of human subjects using an iterative abduction framework. *Social Network Analysis and Mining*, 10(1):1–43, 2020.

Elias Chaibub Neto. A causal look at statistical definitions of discrimination. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 873–881, 2020.

David Chan, Rong Ge, Ori Gershony, Tim Hesterberg, and Diane Lambert. Evaluating online ad campaigns in a pipeline: causal models at scale. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–16, 2010.

Alok R Chaturvedi, Daniel R Dolk, and Paul Louis Drnevich. Design principles for virtual worlds. *MIS Quarterly*, pages 673–684, 2011.

Shu-Heng Chen and Chung-Chih Liao. Agent-based computational modeling of the stock price–volume relation. *Information Sciences*, 170(1):75–100, 2005.

Peng Cui, Zheyan Shen, Sheng Li, Liuyi Yao, Yaliang Li, Zhixuan Chu, and Jing Gao. Causal inference meets machine learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3527–3528, 2020.

Fabian Dablander. An introduction to causal inference. 2020.

Sumanta K Das, Sumant Mukherjee, et al. Agent-based decision making for integrated air defence systems. *Journal of Battlefield Technology*, 14(1):25, 2011.

A Philip Dawid. Beware of the dag! In *Causality: objectives and assessment*, pages 59–86. PMLR, 2010.

Louise Dennis, Michael Fisher, Marija Slavkovik, and Matt Webster. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, 77:1–14, 2016.

David Easley, Jon Kleinberg, et al. Power laws and rich-get-richer phenomena. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press*, 2010.

Felix Elwert. Graphical causal models. In *Handbook of causal analysis for social research*, pages 245–273. Springer, 2013.

Xiao Fang and Paul J Hu. Top persuader prediction for social networks. *MIS Quarterly*, 42(1):63–82, 2018.

Marco Gavanelli, Evelina Lamma, Paola Mello, and Paolo Torroni. An abductive framework for information exchange in multi-agent systems. In *International Workshop on Computational Logic in Multi-Agent Systems*, pages 34–52. Springer, 2004.

Timothy C Germann, Kai Kadau, Ira M Longini, and Catherine A Macken. Mitigation strategies for pandemic influenza in the united states. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.

Manfred Gilli and Peter Winker. A global optimization heuristic for estimating agent based models. *Computational Statistics & Data Analysis*, 42(3):299–312, 2003.

David H Glass. Competing hypotheses and abductive inference. *Annals of Mathematics and Artificial Intelligence*, pages 1–18, 2019.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524, 2019.

Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

Arnaud Grignard, Patrick Taillandier, Benoit Gaudou, Duc An Vo, Nghi Quang Huynh, and Alexis Drogoul. Gama 1.6: Advancing the art of complex agent-based modeling and simulation. In *International Conference on Principles and Practice of Multi-agent Systems*, pages 117–131. Springer, 2013.

Volker Grimm, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M Mooij, Steven F Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L DeAngelis. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310(5750):987–991, 2005.

Gregory P Hanley, Brian A Iwata, and Eileen M Roscoe. Some determinants of changes in preference over time. *Journal of Applied Behavior Analysis*, 39(2):189–202, 2006.

Miguel A Hernán. Invited commentary: agent-based models for causal inference—reweighting data and theory in epidemiology. *American Journal of Epidemiology*, 181(2):103–105, 2015.

Gabriel Istrate. Models we can trust: Toward a systematic discipline of (agent-based) model interpretation and validation: Blue sky track. *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, 2021.

Stef Janssen, Alexei Sharpanskykh, Richard Curran, and Koen Langendoen. Using causal discovery to analyze emergence in agent-based models. *Simulation Modelling Practice and Theory*, 96:101940, 2019.

Dominik Janzing, David Balduzzi, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Quantifying causal influences. *The Annals of Statistics*, 41(5):2324–2358, 2013.

Erisa Karafili, Linna Wang, Antonis C Kakas, and Emil Lupu. Helping forensic analysts to attribute cyber-attacks: An argumentation-based reasoner. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 510–518. Springer, 2018.

Keisuke Kiritoshi, Tomonori Izumitani, Kazuki Koyama, Tomomi Okawachi, Keisuke Asahara, and Shohei Shimizu. Estimating individual-level optimal causal interventions combining causal models and machine learning models. In *The KDD'21 Workshop on Causal Discovery*, pages 55–77. PMLR, 2021.

Carly R Knight and Christopher Winship. The causal implications of mechanistic thinking: Identification using directed acyclic graphs (dags). In *Handbook of causal analysis for social research*, pages 275–299. Springer, 2013.

Gueorgi Kossinets and Duncan J Watts. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–450, 2009.

Setsuya Kurahashi. Model prediction and inverse simulation. In *Innovative Approaches in Agent-Based Modelling and Business Intelligence*, pages 139–156. Springer, 2018.

Marcel Kvassay, Peter Krammer, Ladislav Hluchỳ, and Bernhard Schneider. Causal analysis of an agent-based model of human behaviour. *Complexity*, 2017, 2017.

Francesco Lamperti, Andrea Roventini, and Amir Sani. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389, 2018.

Sam Maes, Joke Reumers, and Bernard Manderick. Identifiability of causal effects in a multi-agent causal model. In *IEEE/WIC International Conference on Intelligent Agent Technology, 2003. IAT 2003.*, pages 605–608. IEEE, 2003.

Sam Maes, Stijn Meganck, and Bernard Manderick. Inference in multi-agent causal models. *International Journal of Approximate Reasoning*, 46(2):274–299, 2007.

Daniel Malinsky and Peter Spirtes. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD workshop on causal discovery*, pages 23–47. PMLR, 2018.

Wenji Mao and Jonathan Gratch. Social causality and responsibility: Modeling and evaluation. In *International Workshop on Intelligent Virtual Agents*, pages 191–204. Springer, 2005.

Wenji Mao and Jonathan Gratch. Evaluating a computational model of social causality and responsibility. In *Proceedings of the fifth International Joint conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 985–992. International Foundation for Autonomous Agents and Multiagent Systems, 2006.

Wenji Mao and Jonathan Gratch. Modeling social causality and responsibility judgment in multi-agent interactions. *Journal of Artificial Intelligence Research*, 44:223–273, 2012.

Brandon DL Marshall and Sandro Galea. Formalizing the role of agent-based modeling in causal inference and epidemiology. *American Journal of Epidemiology*, 181(2):92–99, 2015.

Marta Cialdea Mayer and Fiora Pirri. Abduction is not deduction-in-reverse. *Logic Journal of the IGPL*, 4(1):95–108, 1996.

Laila Melkas, Rafael Savvides, Suyog H Chandramouli, Jarmo Mäkelä, Tuomo Nieminen, Ivan Mammarella, and Kai Puolamäki. Interactive causal structure discovery in earth system sciences. In *The KDD'21 Workshop on Causal Discovery*, pages 3–25. PMLR, 2021.

Ehsan Mokhtarian, Sina Akbari, AmirEmad Ghassami, and Negar Kiyavash. A recursive markov boundary-based approach to causal structure learning. In *The KDD'21 Workshop on Causal Discovery*, pages 26–54. PMLR, 2021.

Zalimhan Nagoev, Inna Pshenokova, and Murat Anchekov. Model of the reasoning process in a multiagent cognitive system. *Procedia Computer Science*, 169:615–619, 2020.

Hung Khanh Nguyen, Raymond Chiong, Manuel Chica, Richard H Middleton, and Sandeep Dhakal. Agent-based modeling of migration dynamics in the mekong delta, vietnam: Automated calibration using a genetic algorithm. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pages 3372–3379. IEEE, 2019.

Rogelio Oliva. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, 151(3):552–568, 2003.

Osonde Osoba and Bart Kosko. Beyond dags: modeling causal feedback with fuzzy cognitive maps. *arXiv preprint arXiv:1906.11247*, 2019.

Judea Pearl. Graphical models for probabilistic and causal reasoning. *Quantified representation of uncertainty and imprecision*, pages 367–389, 1998.

Judea Pearl. Causal inference. *Causality: objectives and assessment*, pages 39–58, 2010.

Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3):54–60, 2019.

Luís Moniz Pereira and Ari Saptawijaya. Abduction and beyond in logic programming with application to morality. *IFCoLog Journal of Logic and its Applications*, 3(1):37–72, 2016.

Markus Peters, Maytal Saar-Tsechansky, Wolfgang Ketter, Sinead A Williamson, Perry Groot, and Tom Heskes. A scalable preference model for autonomous decision-making. *Machine Learning*, 107(6):1039–1068, 2018.

Dimitrios Rafailidis and Alexandros Nanopoulos. Modeling the dynamics of user preferences in coupled tensor factorization. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 321–324, 2014.

William Rand and Roland T Rust. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3):181–193, 2011.

William Rand and Christian Stummer. Agent-based modeling of new product market diffusion: an overview of strengths and criticisms. *Annals of Operations Research*, pages 1–23, 2021.

Ronald E Rice, August E Grant, Joseph Schmitz, and Jack Torobin. Individual and network influences on the adoption and perceived outcomes of electronic messaging. *Social Networks*, 12(1):27–55, 1990.

Daniel Rigney. *The Matthew effect: How advantage begets further advantage.* Columbia University Press, 2010.

Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762): 854–856, 2006.

Ken Satoh, Katsumi Inoue, Koji Iwanuma, and Chiaki Sakama. Speculative computation by abduction under incomplete communication environments. In *Proceedings Fourth International Conference on MultiAgent Systems*, pages 263–270. IEEE, 2000.

Jasjeet S Sekhon. The neyman-rubin model of causal inference and estimation via matching methods. *The Oxford handbook of political methodology*, 2:1–32, 2008.

Cosma Rohilla Shalizi and Andrew C Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40 (2):211–239, 2011.

Galit Shmueli and Otto R Koppius. Predictive analytics in information systems research. *MIS Quarterly*, pages 553–572, 2011.

Forrest Stonedahl and William Rand. When does simulated data match real data? In *Advances in Computational Social Science*, pages 297–313. Springer, 2014.

Forrest Stonedahl, David Anderson, and William Rand. When does simulated data match real data? In *Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation*, pages 231–232, 2011.

Eric V Strobl. Improved causal discovery from longitudinal data using a mixture of dags. In *The 2019 ACM SIGKDD Workshop on Causal Discovery*, pages 100–133. PMLR, 2019.

Tyler J VanderWeele, John W Jackson, and Shanshan Li. Causal inference and longitudinal data: a case study of religion and mental health. *Social psychiatry and psychiatric epidemiology*, 51(11):1457–1466, 2016.

Jussi Viinikka, Antti Hyttinen, Johan Pensar, and Mikko Koivisto. Towards scalable bayesian learning of causal dags. *Advances in Neural Information Processing Systems*, 33:6584–6594, 2020.

Matthew J Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery. *arXiv preprint arXiv:2103.02582*, 2021.

Feng Wan and Munindar P Singh. Commitments and causality for multiagent design. In *Proceedings of the second International Joint conference on Autonomous agents and Multiagent Systems (AAMAS)*, pages 749–756. International Foundation for Autonomous Agents and Multiagent Systems, 2003.

Michael P Wellman. Putting the agent in agent-based modeling. *Autonomous Agents and Multi-Agent Systems*, 30(6):1175–1189, 2016.

Thomas C Williams, Cathrine C Bach, Niels B Matthiesen, Tine B Henriksen, and Luigi Gagliardi. Directed acyclic graphs: a tool for causal studies in paediatrics. *Pediatric research*, 84(4):487–493, 2018.

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

Gabriel Wurzer and Wolfgang E Lorenz. Causality in hospital simulation based on utilization chains. In *Proceedings of the Symposium on Simulation for Architecture and Urban Design*, 2014.

Jonathan D Young, Bryan Andrews, Gregory F Cooper, and Xinghua Lu. Learning latent causal structures with a redundant input neural network. In *Proceedings of the 2020 KDD Workshop on Causal Discovery*, pages 62–91. PMLR, 2020.

Haifeng Zhang and Yevgeniy Vorobeychik. Empirically grounded agent-based models of innovation diffusion: a critical review. *Artificial Intelligence Review*, pages 1–35, 2019.

Yi Zhang, Zhe Li, and Yongchao Zhang. Validation and calibration of an agent-based model: A surrogate approach. *Discrete Dynamics in Nature and Society*, 2020, 2020.