An agent-based model with social interactions for scalable probabilistic prediction of performance of a new product

Abstract

Understanding the spreading process of new products provides valuable knowledge that can be used for effective marketing. The ability to make early prediction of success or failure is a great advantage in innovation processes. Extending current literature in a novel way, we propose a data-driven agent-based methodology that accomplishes this task. Inference and predictions are based on short-time observations of the product adoption history and knowledge of the social network of consumers. We model and predict adoptions at the agent level as driven by unobserved peer-to-peer influence and external factors such as marketing. The method compares interaction between consumers and general campaigns, and quantifies the importance of characteristics of customers and their social relations. Our computationally efficient method is demonstrated by analyzing real data, predicting the process far into the future using data from a short period after launch, and validated by simulation experiments on a true full-scale communication network.

Keywords: Networks, stochastic agent-based modeling, viral marketing, predicting product performance, customer behavior modeling.

1 Introduction

Imagine that you have recently launched a new product in the market, and that you have collected data on all the adoptions so far; there have been some general advertising campaigns, and there are signs, or at least hopes, that adoptions start to get "viral" among your customers, who are interacting in a social network. You have a good picture of your customer base and would like to understand how adoptions will develop in the next months, and ideally in the next year, under the assumption that the adoption processes which are in action now, will persist. You would also like to ask what-if questions, related to personalized marketing of your product. In this paper we present a new data-driven agent-based model (ABM) [**?**], which allows performing these analyses in a real market and in useful computational time. Studying and understanding the spreading of innovations can point the way to more effective marketing. This work presents methodology that gives unique insights into the spreading process. Our methodology enable fast learning of what is going on with the adoption of a particular innovation. The results and methods presented represent a rare combination of insights, both for understanding and for practical applicability. A recent analysis of the field of digital marketing [**?**] suggests six important themes for future research, including marketing strategy, social media, customer behavior and location data, which are all topics related to the work in this paper.

The methodology we propose is based on (i) an available social network, here captured by the complete agent-to-agent telecommunications over a mobile phone network observed over one month, (ii) a set of covariates related to the agent-to-agent communication, such as the intensity of the communication (for example total number of minutes spoken) and the nature of the connection (for example similarity in age), (iii) a set of agent related covariates, e.g. demography and agent location, (iv) the time points, targets and volumes of specific advertising campaigns and (v) the list of adopters and adoption times during a short period, often during product launch time. In our model, time is continuous. By modeling the interactions between customers by means of an observed social network, which is assumed to represent interactions among all pairs of agents, ABMs capture the effect of social influences. Data-driven ABMs differ from non-empirically constructed ABMs, which are informed by general knowledge and hypotheses, and which are useful tools in understanding and comparing policies and marketing strategies, less in predicting the future of actual products in specific markets. We refer to the excellent reviews [**?**] and [**?**] for discussions of advantages and challenges of data-driven ABMs. Next we highlight some of the important criteria set up by these seminal papers for successful ABMs, including current challenges, pinpointing how the present paper contributes advancements of method and practice.

The modeling process for data-driven ABMs has three fundamental stages: initialization, calibration and validation [**?** **?**]. Initialization comprises the model formulation of the agent behavior, the social interaction and the exogenous influences, its parametrization, and the data-sources needed and available to inform models from data. Calibration consists of the method and algorithms used to learn parameters from data. Validation is the process of determining if the ABM is able to describe the market appropriately. [**?**] finds that very few papers present all three phases in a satisfactory way, and concludes that the current literature on data-based ABMs "commonly exhibits several major shortcomings in model calibration and validation".

Often, calibration relies on the matching of aggregated statistics (called macro-calibration), for example comparing actual weekly total sales in the training data with ABM-based simulated weekly total sales. This approach can suffer from non-identifiability of parameter estimates, leading to potentially wrong conclusions. Instead, proper statistical inference, based on cross-validation and maximum-likelihood of all data (micro-calibration) is recommended. In the present paper, we develop a full maximum-likelihood based calibration, which uses each single agent's behavior in the training period. An additional important challenge of calibration is computational complexity. [**?**] explains that it is sometimes necessary to sub-sample the set of agents, or the set of adoption times, in order to be able to estimate parameters in useful computational time. However, reducing sample size can lead to bad parameter estimates. For our proposed generic ABM we develop a computational approach which reduces computational complexity markedly, so that we can scale our algorithm to full societal level. Our telecommunication case has approximately 2 million agents, the social network has approximately 20 million edges, and there are $36,448$ adoptions in the training data and $245,728$ in the validation data. A final pitfall of calibration is related to stochastic ABMs, for which uncertainty in parameter estimation must be properly quantified. As we will see

later, this is particularly important in prediction, as the uncertainty of the calibration must be propagated into the validation, for this to be valid. Using penalized regression (like lasso [**?**]) and cross-validation does not easily lead to valid confidence intervals for the parameters [**?**]. Instead, the maximum likelihood approach used in this paper naturally leads to confidence intervals, whose coverage appears to be appropriate.

Three more pitfalls found by [**?**] are related to the validation step. First, citing from this paper, "rigorous quantitative validation on independent data is uncommon"–that is, testing of the model is seldom performed on data which have not been used for training. [**?**] writes "Few conduct validation of forecasting effectiveness on independent future data". As recommended in [**?**], we perform validation by prediction of future adoptions, comparing how the true adoptions are captured within the uncertainty ranges of the ABM-based prediction. In this paper, we do both macro- and micro-validation. The first verifies how the ABM predictions fit aggregated statistics of the future data, for example total adoptions. Micro-validation on the other hand focuses on how the predicted agent-level behavior fits observed individual behavior. For a stochastic ABM, one cannot expect exact correspondence between predicted future individual events and true ones [**?**], but a good correspondence of individual characteristics, for example in terms of shares of adopters belonging to one gender or to a given age group, can be expected.

The present paper makes the following contributions:

1. We propose a continuous-time, intensity-based modeling approach for ABMs, which include an agent-level model, an external-level model and an interaction-level model, based on an existing social network representing social interactions, i.e. agent-to-agent influence.

2. We perform maximum-likelihood inference for the model parameters, based on each agent's properties and behavior in the training period. We present computationally efficient algorithms for calibration and prediction, allowing to run our ABM in useful time at a full societal level.

3. We are able to distinguish the exogenous, out-of-network sources of influence from the viral, word-of-mouth spread, which is important and generally difficult [**?**].

4. We perform both macro- and micro-validation on a real data set by prediction of total sales two years into the future, based on a short training period of three months, and by predicting useful properties of the agents who will adopt in the future.

5. We illustrate how to use the ABM in what-if scenario simulations, which can be for example changing the general advertising campaigns or simulating that the product will become more or less viral. This opens important possibilities for personalized marketing, for example by targeting the most difficult group of customers, or the strongest influencers.

The paper is organized in six sections as follows: in Section **??** we present previous research on the central concepts of this paper. Then in Section **??**, we present the methodology and data.

Results are presented in Section **??**, then Section **??** discusses the contributions of work as well as limitations and future work. Conclusions are drawn in Section **??**.

2 Background

Artificial intelligence and statistical modeling is increasingly more important in marketing and information sciences. For a general background on the use of artificial intelligence in these areas see [**?** **?** **?** **?**]. The use of ABMs [**?** **?** **?** **?**] is by now a classical approach to build digital twins [**?**] of complex systems, which are composed of many simple units (agents, here the customers) who can act autonomously, but are influenced by the interactions with other units. Because each single agent is modeled using its specific features (which we call covariates in this paper), ABMs allow to represent a heterogeneous population. ABMs are constructed by defining simple behavioral rules which the agents follow; these rules can be deterministic or stochastic (as in this paper) and depend on the interaction patterns between subsets of agents. In contrast to ABMs, compartmental or aggregated models represent groups of agents together (the compartment), often by means of ordinary differential equations [**?**]. By construction, these models can capture heterogeneity of the population only by introducing multiple compartments. However, their number must remain small for practical and computational purposes, which severely limits their capacity to describe complex markets [**?**]. ABMs have been proposed very successfully in many different areas, from sociology [**?**] to psychology [**?**], from economy [**?**] to information science [**?**], from infectious diseases [**?** **?**] to marketing ([**?**] and references their-in), to mention a few.

Social relations are well known to affect the adoption of products [**?**] and lay the foundation of viral marketing [**?** **?** **?** **?**]. They have also been used in other contexts, e.g. for detecting fake news campaigns [**?**]. Many possible models for social relations have been proposed in the literature, ranging from theoretical assumptions about who influences who and how much, to fully observed or surveyed recommendation messaging between customers. Networks are an intuitive framework for social relations, where customers are nodes, and edges between nodes represent existing social relations between customers. Among the theoretical models which have been assumed in the literature, an important role is taken by regular stochastic networks, for example based on preferential attachment and scale-free growth [**?**]. In many applications, one can assume that the influence is in the form of local interaction: for example, all customers resident in a certain geographic area influence each other when adopting a product; or customers belonging to specific groups (age-based, income-based etc) influence each other. Within each local group the network is complete (with an edge between each pair of customers), with further links between groups originating from multiple group membership of customers. In more recent times, it is becoming possible to observe the social network between customers directly, for example, as we do in the present paper, through the telecommunication between customers, or by explicitly designed recommendation referral programs [**?**], or on social media websites [**?** **?** **?** **?**]. Customer behavior does of course not only depend on social influence. The adoption of products is affected by external interventions,

such as general or targeted marketing campaigns, and by overall trends and fashions [**?**]. These exogenous, out-of-network sources of influence can be easily modeled within an ABM, for example as additional synthetic nodes which are connected to all customers [**?**].

Some other papers have considered aspects of what we present in this work, but not together. [**?**] combines a non-parametric model of the external influence with word-of-mouth influence, but assumes this latter contribution to be fully known, and hence estimates only the parameters for the external influence. [**?**] proposes a similar non-parametric model, doing inference on both viral and external forces, with an algorithm that alternates between the two components, but without agent-level heterogeneity. [**?**] proposes a framework which similarly to our work is based on an epidemic SIS model. The framework combines word-of-mouth with external factors, but differs from ours in that it is based on the structure of the social network, and disregards agent-level information on individual nodes and edges. Alternative approaches to predicting future adoptions include the self-exciting Hawkes point process [**?** **?** **?**]. This process also mimics the effect of the viral spread in a network, but does not utilize the available network structure nor does it model the individual actions; it is therefore quite different from ABMs.

The previous work which to our knowledge is closest to our paper is [**?**]. Like us, the authors present a data-driven ABM and perform calibration and validation in the recommended way. Still, there are several differences between this and our paper: first, we use an existing and real agent-to-agent social network, as a proxy for social relations, while [**?**] builds a network based on the geographic location of the agents. This interesting construction depends on the definition of region, and the authors work with three different geographical units (entire area, ZIP code and a circle with a given radius around each agent). During calibration they select the unit which allows best cross-validation fit. However, there is a certain level of arbitrariness in the definition of the geographical units, which can have a large impact on predictions. Because in real-life situations validation by prediction is not possible (the future is still unknown), and because we believe that geographical closeness might not be enough in many situations, and oftentimes not even correct to describe social relations, we prefer to rely on a given network, which captures in an unequivocal way the interactions between agents. Telecommunication operators have the opportunity to observe a proxy of the social network of their customers [**?**], via communication links (voice and text), which is what we use in this paper. We believe that such data describing networks, defined by telecommunication, financial transactions, or ownership (when agents are companies), etc. are becoming more common. The second important difference relates to the agent-level model. We assume continuous time and work with adoption intensities, in the spirit of ABMs for infectious diseases [**?**], rather than using adoption probabilities and discrete time. Because intensities are not normalized, the modeling is less constrained. Our approach is generic and we indicate how other modelers can build their own ABMs using our mathematical approach. The third important advantage of our approach is scalability. "Solving the resulting maximum likelihood estimation can be computationally intractable", states [**?**]. We are able to transform several computational components of the maximum likelihood computations into operations on sparse matrices, which

are computationally very efficient. Instead of building our algorithms upon an existing simulation software (like the Repast ABM simulation toolkit [**?**] used in [**?**]), we have optimized our code for fast calibration, and in this way we are able to use all the data available in the training set. During prediction, we are also able to update all states at each single adoption, instead of accumulating several adoptions before updating, as done in [**?**].

3 Material and methods

3.1 Model

We let $i = 1, 2, ..., n$ identify each agent in a given market. The n agents are considered as nodes in a network, with edges representing direct social contact between agents. For instance in a telecom context, an edge could be present between nodes i and j if they phone or text message above a given threshold. The network of nodes does not need to be connected and can consist of separate components. We denote the existence of an edge from i to j by $e_{ij} = 1$, while $e_{ij} = 0$ indicates absence of an edge. We assume that agent i may influence agent j only if $e_{ij} = 1$. The edges can vary in time if the network is time-dynamic, but for simplicity here we assume that the network is static. Depending on the context, edges may be directed or undirected, with $e_{ij} = e_{ji}$ in the latter case. The ABM is inspired by epidemic models for infectious diseases [**?** **?** **?**].

Both nodes and edges carry covariate information, and we denote the vector of p covariates related to the node i as x_i. Demographic covariates can for instance include age, gender and geographic location, while behavioral covariates can include information on how nodes have used similar products or services in the past and their past propensities to adopt new versions of products. In general, covariates should be chosen so that they potentially contribute information to the propensity of a node to adopt the new product and to influence others to do so. Covariates connected to an edge between nodes i and j carry information about the propensity for i to specifically influence j, and may be directed or undirected. An example of a directed edge covariate is an indicator variable of whether i is younger than j or not, while an indicator of whether i and j live in the same municipality would be an undirected edge covariate. As these covariates may be thought of as a weight (or strength) of the connection between nodes i and j, we let w_{ij} denote the vector of covariates related to the edge between nodes i and j.

For the rest of the paper, we assume that a product or service becomes available on the market from time $t = 0$, and we observe the adoption process in the time period $[0, T^{\text{obs}}]$. We assume that all nodes can adopt, i.e. are susceptible in the epidemic metaphor, until adoption or T^{obs}. Let B denote the set of agents or nodes adopting the product during the time period until T^{obs}. For a specific node $i \in B$ we denote by $t_i \in [0, T^{\text{obs}}]$ the adoption time of agent i.

Figure **??** sketches the essential components of our approach.

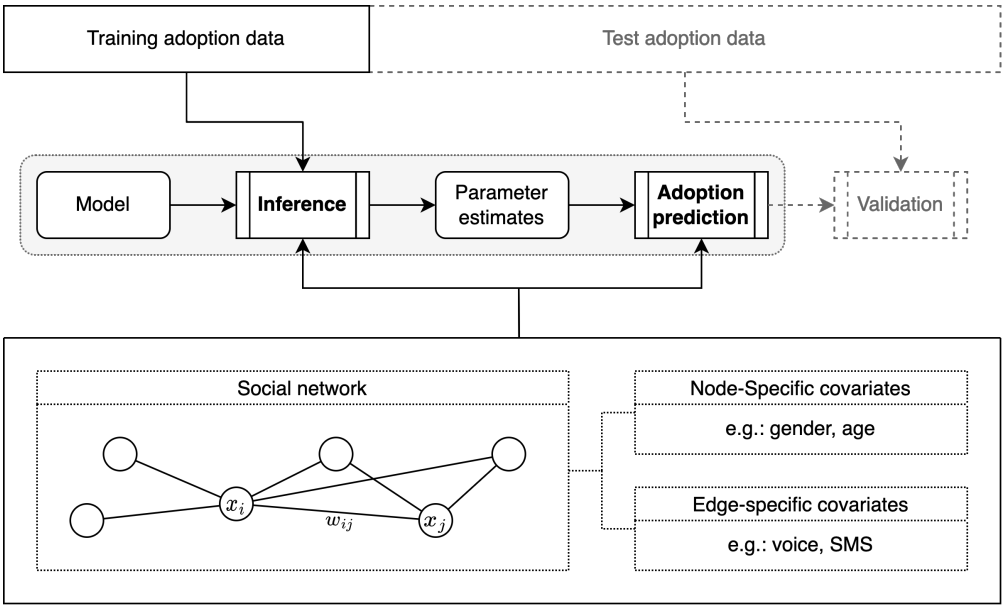


Figure 1: The flow of our methodology is indicated by arrows. Inference is based on the model, training adoption data and social network with covariate information. The estimated parameters and uncertainty are then used in the prediction of future adoptions. For validating predictions, test adoption data are used. Both the estimation and the prediction phases rely on information from the same social network.

3.1.1 Intensity based model of viral influence

We assume that node i can influence its neighbors in the network from the adoption time t_i, until time $t_i^A = \min(t_i + A, T^{\text{obs}})$. For simplicity, and following [**?**], the duration of the influence window A is here assumed constant and equal for all nodes.

Consider two nodes i and j. Assume that $i \in B$ and that j has not adopted the product by time t_i. If (1) $j \in B$ and adopts the product at time t_j, (2) $t_j \in [t_i, t_i^A]$, and (3) $e_{ij} = 1$, then i may have influenced j to adopt. If on the other hand node j does not adopt the product during a period $[t_i, t_i^A]$ even if $e_{ij} = 1$, then we assume that the influence effect from i to j has failed.

Let $\lambda_{ij}(t)$ denote the transmission intensity of the influence from node i to node j at time t, which is different from 0 only if node i can influence node j at time t. This is ensured by an influence indicator function $I_{ij}(t)$, which equals 1 only if node i can influence node j at time t. There are three conditions that must be fulfilled for this: (1) $t \in [t_i, t_i^A]$ such that node i has adopted the product and may still influence its neighbors at time t, (2) node j has not adopted the product before time t and hence is susceptible, and (3) $e_{ij} = 1$ such that there is contact from i to j. If these three conditions are not fulfilled at time t, i cannot influence j at that time and $I_{ij}(t) = 0$, which implies that also $\lambda_{ij}(t) = 0$. Furthermore, we assume a common baseline intensity for the influence strength between all pairs of nodes, denoted by $\lambda_{\text{vir}} = \exp(\alpha_0)$. This could be extended to be time-varying, including for instance seasonality. Finally, the effect of the agent covariates is modeled by three different groups of parameters, denoted by the vectors α_1, α_2 and

α_3. The first group α_1 models the effect of the influence strength of the infectious node i, using the covariate vector x_i of node i. The second group α_2 models how difficult or easy it is to influence the susceptible node j to adopt the product, using the covariate vector x_j of node j. The term α_3 allows the intensity of influence to depend on covariates of the edge between i and j. All together, the overall intensity of influence from node i to node j at time t is modeled as

$$\lambda_{ij}(t) = \lambda_{\text{vir}} \exp\left(x_i^T \alpha_1\right) \exp\left(x_j^T \alpha_2\right) \exp\left(w_{ij}^T \alpha_3\right) I_{ij}(t). \tag{1}$$

This form of dependency on the covariates is based on the model in [**?**], which was inspired from [**? ?**]. Generally, it may be the case that the covariates that are relevant for influence are different from the ones that are relevant for susceptibility. This can be handled technically by including all covariates in the covariate vector x_i and fixing the appropriate elements of α_1, α_2 to zero (and similarly for β_2 below).

3.1.2 Intensity based model of external influence

Each node j can be influenced to adopt the product not only by its neighbors, but also by external factors such as marketing campaigns or an underlying propensity towards the product. External sources or background influence have been included in several other modeling approaches [e.g. **? ?**]. For consistent notation, we add an external node to the network, indexed by $i = 0$, connected to all nodes, such that $e_{0j} = 1, \forall j$. Node 0 represents the source of any external convincement. This node is always infectious and can thus influence all susceptible nodes in the network at all times. Hence $I_{0j}(t) = 1$ if node j is susceptible at time t, otherwise $I_{0j}(t) = 0$.

We start by assuming a common time-varying baseline intensity $\lambda_{\text{ext}}(t)$, expressing the external influence, such as specific marketing campaigns. Marketing covariates or information describing the external influence over time are denoted by $m(t)$, for instance the intensity of marketing campaigns at time t. We model the time-varying baseline as a function of marketing covariates with parameters β_1, so that $\lambda_{\text{ext}}(t) = \exp\left(\beta_0 + m(t)^T \beta_1\right)$. The value $m(t) = 0$ then corresponds to a constant baseline, so that $\lambda_{\text{ext}}(t) = \exp\left(\beta_0\right)$. In addition we model the effect of agent covariates of j by parameters β_2, expressing how easy or difficult it is for the external node to influence the node j. It is reasonable to assume that the set of covariates connected to β_2 is often the same as for α_2. Overall, the external intensity of influence from the external 0 node to node j at time t is modeled as

$$\lambda_{0j}(t) = \lambda_{\text{ext}}(t) \exp\left(x_j^T \beta_2\right) I_{0j}(t). \tag{2}$$

3.1.3 Combining all sources of influence

The total intensity of influence to a node j at time t is given by combining Equations (**??**) and (**??**)

$$\lambda_j(t) = \sum_{i=0}^{n} \lambda_{ij}(t) = \lambda_{0j}(t) + \sum_{i=1}^{n} \lambda_{ij}(t). \tag{3}$$

Here, influence intensities are summed over all sources which may influence node j. Define $C_j(t)$ to be the set of influential neighbors of node j at time t. Then Equation (**??**) simplifies to

$$\lambda_j(t) = \lambda_{0j}(t) + \sum_{i \in C_j(t)} \lambda_{ij}(t), \tag{4}$$

as $\lambda_{ij}(t)$ is non-zero only for influential neighbors of j at time t. At time t, the time to adoption for the nodes that are susceptible are assumed to be independent and exponentially distributed with intensities given in (4).

3.2 Data analysis

3.2.1 Maximum likelihood estimation

We fit the model to data by maximum likelihood, to provide estimates of the parameters of the model together with their uncertainty. The necessary data consist of all individual adoptions with the corresponding adopting times, for the initial training period of time $[0, T^{obs}]$. In addition, the covariates of all nodes and edges are needed. Missing values need to be imputed, for example as in [**?**]. To identify relevant covariates for nodes and edges, expert information and/or a preliminary data analysis can be useful. It is recommended to use an alternative data set to avoid overfitting, for instance data from previous similar products.

We derive the likelihood for the parameters given the observed data [see e.g. **?** , equation (5.2)]. Defining $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2)$, the likelihood resulting from Equation (**??**) is

$$\begin{aligned} \mathcal{L}(\theta) &= \left\{ \prod_{k \in B} \lambda_k(t_k) \right\} \exp\left\{ - \int_0^{T^{\text{obs}}} \sum_{j=1}^{n} \lambda_j(t) dt \right\} \\ &= \left\{ \prod_{k \in B} \left[\lambda_{0k}(t_k) + \sum_{i \in C_k(t_k)} \lambda_{ik}(t_k) \right] \right\} \exp\left\{ - \sum_{j=1}^{n} \int_0^{T^{\text{obs}}} \left[\lambda_{0j}(t) + \sum_{i \in C_j(t)} \lambda_{ij}(t) \right] dt \right\}. \end{aligned} \tag{5}$$

For $j = 1, \ldots, n$, define T_j^{s} as the time period in which node j is susceptible, i.e. $T_j^{\text{s}} = [0, t_j]$ if j adopts in the observation period, or $T_j^{\text{s}} = [0, T^{\text{obs}}]$ otherwise, and let N_j be the set of all nodes that have been influential neighbors of j at some point in time period T_j^{s}. Furthermore, for $i \in N_j$, let T_{ij}^{s} be the time interval in which i is infectious and j is susceptible, i.e.

$$T_{ij}^{\text{s}} = \begin{cases} \left[t_i, \min(t_i^A, t_j)\right] & \text{if } j \in B, t_i < t_j \text{ and } i \in N_j \\ \left[t_i, t_i^A\right] & \text{if } j \notin B \text{ and } i \in N_j \\ \emptyset & \text{otherwise.} \end{cases}$$

Now, to simplify notation, introduce $\gamma_{ij} = \exp\left(x_i^T \alpha_1 + x_j^T \alpha_2 + w_{ij}^T \alpha_3\right)$ as the total viral effect of agent covariates and let $\gamma_{0j} = \exp\left(x_j^T \beta_2\right)$ be the total external effect of agent covariates. Then we

can write Equation (**??**) in the following way

$$\mathcal{L}(\theta) = \left\{ \prod_{k \in B} \left[\lambda_{\text{ext}}(t_k)\gamma_{0k} + \sum_{i \in C_k(t_k)} \lambda_{\text{vir}}\gamma_{ik} \right] \right\} \cdot \\ \cdot \exp\left\{ -\sum_{j=1}^{n} \left[\gamma_{0j} \int_{T_j^{\text{s}}} \lambda_{\text{ext}}(t)dt + \sum_{i \in N_j} \left(\gamma_{ij} \int_{T_{ij}^{\text{s}}} \lambda_{\text{vir}} dt \right) \right] \right\}. \tag{6}$$

Note that the likelihood comprises two parts. The first includes contributions from nodes that eventually adopt and consists of multiplying the total intensity of influence for each $k \in B$ at the time of adoption t_k. The second part summarizes the influence exerted onto all nodes $j = 1, \ldots, n$ during the period that they are susceptible. The two parts collect the effects of external pressure alongside two different sets of influential neighbors. The set N_j contains all neighbors that have exerted influence on j in its susceptibility period T_j^s, and appears in the second part. The first part makes use of $C_k(t_k)$, which only contains the nodes that are influential neighbors of k at the specific time point t_k. For all adopting nodes $k \in B$, the set $C_k(t_k)$ is a subset of N_k.

An essential contribution of our approach is its computational efficiency, which allows to run both inference (maximum likelihood) and predictions (see Section **??**) in useful operational time with a real-life number of agents and network connections. Details can be found in Appendix **??**. An investigation of the time complexity can be seen in Appendix **??**.

3.2.2 Estimating the influence period A

An adoption due to external pressure may happen at any time, while a viral adoption may only have occurred if an adopter has at least one influential neighbor who adopted during the A previous time units. The length of the influence period A has an important role in describing the viral influence. In our model, every adopting neighbor of a given agent j contributes to $\lambda_j(t)$ immediately after the adoption time and for an amount of time A. An estimate of the value A can be found by analyzing the difference in the adoption times of all dyads where both nodes adopted in the observed period. More precisely, for all pairs of observed adoption times t_i and t_j for $\{i, j\} \in B \times B$ such that $t_i < t_j$ and $e_{ij} = 1$, we study the empirical distribution of $\Delta t_{ij} = t_j - t_i$. We expect to see a distribution with a clear change-point at some value Δt_{ij}^*, which we then select as the estimate of A. This method is used in the real-life application (Section **??**), and then successfully validated in the simulation study (Section **??**).

This method, however, requires some refinement when used in real-life, because several marketing campaigns are launched during the training period, creating time segments with a higher intensity of non-viral adoptions. This produces noise in the empirical distribution of Δt_{ij}. The effect of marketing campaigns must therefore be removed from the data by omitting dyads for which the adoption of node j happens at a time with an active marketing campaign.

An alternative approach is to treat A as a tuning parameter, selecting the value giving the best accuracy or performance according to some error measure. For example, we may evaluate

the out-of-sample prediction error over a suitable range of possible A values (based on expert knowledge). If $[0, T^{\text{obs}}]$ is the time interval in which the adoptions are observed, we use $[0, T^*]$ as the training period (with $T^* < T^{\text{obs}}$), and then validate the model's prediction error on the period $[T^*, T^{\text{obs}}]$. The procedure is repeated for each value of A on a grid, and the A-value yielding the best predictions is then selected.

3.2.3 Prediction

Based on the results from the maximum likelihood estimation, we can predict the market dynamics after T^{obs} by simulating the model forward in time. One predicted trajectory for a given set of parameter values is obtained in the following way. Let $S(t)$ denote the set containing the nodes that are susceptible at time t. At time t, the time to adoption $\tau_j(t)$ for a susceptible node $j \in S(t)$ is exponentially distributed with intensity $\lambda_j(t)$. We assume that $\tau_j(t)$ and $\tau_{j'}(t)$ are independent for all pairs of susceptible nodes $j \neq j'$, conditional on all $\lambda_j(t), j = 1, \ldots, n$. If we assume that $\lambda_{\text{ext}}(t)$ is constant in time, then the individual intensities $\lambda_j(t)$ are also constant in time, between consecutive adoptions. Hence, at a given time t, the time to the next adoption in the market, $\min\{\tau_j(t), j \in S(t)\}$, is exponentially distributed with intensity $\sum_{j \in S(t)} \lambda_j(t)$. The probability that this adoption is done by node k is then $\frac{\lambda_k(t)}{\sum_{S(t)} \lambda_j(t)}$. Using these results, we can simulate a trajectory in the future by repeatedly sampling the next time-to-adoption and then the actual node which adopts at this time, until we have predicted as far into the future as needed.

Prediction uncertainty is handled using the fact that the maximum likelihood estimate $\hat{\theta}$ is approximately multinormally distributed. The predictions are simulated by repeating the following procedure M times, with M sufficiently large: each time, one parameter vector $\hat{\theta}_m$ is sampled from the multinormal distribution and using this, one future trajectory is simulated. This results in a collection of M predicted trajectories of the future dynamics, from which a point-wise prediction band (say 95%) can be constructed. This procedure allows to take into account both sources of uncertainty of predictions, namely the natural randomness of the model and estimation uncertainty.

3.3 Data collection

3.3.1 Social network

We illustrate our proposed methodology on a customer telecommunication network using simulated and real-life adoption data. The nodes and edges of the network are based on the calling history of the customers of a European telecom company ('the Company') during September 2016. This includes aggregate information about contacts between a sender and a receiver in that period, including both total duration of voice calls (measured in seconds) and number of direct text messages (SMS). When defining the edges in the network, we include only edges between two nodes that have a minimum amount of contact: we take a threshold of at least one sent or received SMS or a call duration of at least 1 minute in total during the whole period. The edges are undirected and constant over time. In addition, we only retained the largest connected component of the network.

Table 1: The percentage of customers in the different age groups in the data set.

Age group	1	2	3	4	5
Ages	0-18	19-30	31-50	51-65	66-100
Total share of age group	7.8 %	12.9 %	28.3 %	24.6 %	26.5%

The final network consists of two millions of nodes, with almost ten times as many connections.

3.3.2 Covariates for nodes and edges

Node and edge covariates are based on information on gender, age and ZIP code for the nodes in the network as in September 2016. Gender is missing for 13.7% of nodes, age is missing for 13.6% of nodes, and both variables are typically missing for the same nodes. In addition, there are 0.9% missing ZIP codes. All missing values are here simply imputed by sampling from the observed population distribution. The covariates used for the nodes are gender (male = 0, female = 1) and age discretized into 5 age groups (see Table **??**). The ZIP code is converted to an indicator variable for whether two connected nodes are in the same municipality, and used as an edge covariate. The additional edge covariates are: 1) an indicator variable for equal gender of the nodes of the edge, 2) a categorical variable indicating either equal age group of the nodes of the edge, one level difference in age group, or more than one level difference in the age group. We also use the amount of voice minutes and the number of SMSs exchanged between the two nodes as edge covariates. Due to extreme communication behavior of some nodes, we impose an upper threshold on voice duration (30 minutes) and the number of SMSs (50). The covariates for voice minutes and number of SMSs are normalized to assume values in $[0, 1]$. The edge covariates are all symmetric.

3.3.3 Marketing covariates

We specifically consider targeted marketing, e.g. SMS or e-mails with specific offers and advertisement sent to selected customers, referred to as Below-the-Line (BTL) campaigns. Such targeted campaigns stand in contrast to mass audience advertising targeted at high volumes of consumers, e.g. TV commercials and posters in the subway, called Above-the-Line (ATL) campaigns [**?**]. For the given product, we have access to all recorded BTL marketing via SMS for the observation period, consisting of the date of the campaign and the total number of dispatched SMSs, $O(t)$. The specific identities of the customers receiving an SMS is not available, but the number of customers targeted by a BTL campaign is incorporated at an aggregate level. We incorporate the marketing covariates $m(t)$ by grouping the BTL campaigns in three different categories of low, medium and

high intensity depending on the number of SMSs that were sent on a given date:

$$\mathrm{m}(t) = \begin{cases} \text{Reference}, & \text{if } O(t) = 0, \\ \text{Low intensity}, & \text{if } 0 < O(t) \leq 10000, \\ \text{Medium intensity}, & \text{if } 10000 < O(t) \leq 50000, \\ \text{High intensity}, & \text{if } O(t) > 50000. \end{cases} \tag{7}$$

In our data there is no information available on the ATL campaigns related to the product. With the categorization of the BTL campaigns described in Equation **??**, the baseline of the external influence, $\lambda_{\text{ext}}(t)$, is piece-wise constant with four different levels.

3.3.4 Real life adoption data

We demonstrate our method on a data set for the adoption of an unspecified service offered, supplied by the Company. The data set, which is on a daily time scale, cover 144 weeks from the launch date, during which over 250,000 agents adopted the product on the network described in Section **??**. To demonstrate the early inference and prediction of the method, we use the adoptions from the 82 first days after the launch of the service as training data.

3.4 Robustness checks

3.4.1 Simulation study

For investigating how the quality of the maximum likelihood estimates and the prediction of future adoptions depend on the quantity of training data, we perform a simulation study of the adoption processes based on the real-life network, with associated covariates and on the same (daily) time scale, described in Section **??**. It is important to be able to make rapid predictions, based on a short initial observation of a market.

The simulation consists of three steps which are repeated $S = 100$ times: (1) we simulate an adoption process for a specific set of parameters θ, (2) we estimate the parameter set $\widehat{\theta}$ for this adoption process using training periods of different length, and (3) we simulate $M = 100$ prediction trajectories of the future. For speed of computation, the network is a subset of the one described in Section **??**, obtained as the largest connected component of the sub-network containing all agents belonging to a specific city (which remains unnamed for privacy reasons) and neighboring municipalities. The network consists of around 57,000 nodes and 440,000 edges. Although the size is significantly smaller compared to the full network, the topological properties are preserved: the degree distribution, the average clustering coefficient for nodes having same degree and the ratio of edges to nodes are very similar (see Section 1.1 of the Supplementary Material). We use all the node covariates described in Section **??** for the adopter node (as x_i) as well as for the susceptible node (as x_j), for both the viral and the external parts. We also include all the edge covariates described in Section **??** for the edges in the viral part (w_{ij}), except for the unnecessary

municipality indicator. The parameter A is set to 5 days. The parameter values (see Table 1 in the Supplementary Material) used for simulating the $S = 100$ adoption processes are chosen to produce 'S'-shaped cumulative adoption curves. The 'S'-shape is natural when the adoption process includes viral effects and the network has a dense core [**?**]: the slope of the adoption curve increases rapidly as the sale starts, while it decreases as the number of adopters saturates the dense parts of the network. After these phases, the slope eventually becomes constant, as a further viral outbreak cannot take place anymore, and new adoptions rely solely on the external terms.

For validating the model estimates during the different phases of the adoption process; before, during and after the viral outbreak, we investigate inference performance for a varying number of adopters in the training period, namely 250, 500, 1000, 1500, 2000, 2500, 3000, and 4000.

To represent the variability of predictions into the future, for each of the $S = 100$ simulated adoption processes, and for each training set size, we simulate $M = 100$ independent trajectories into the future as described in Section **??**.

3.4.2 Micro-validation on the real data

To micro-validate our model, we compare the characteristics of the predicted adopters with the characteristics of the observed adopters in the two weeks following the end of the training period. Given the covariate information at our disposal, the comparison is carried out by comparing the relative frequencies of agents belonging to either gender and to every age group in the two populations. For each of these groups, the predicted relative frequency of the agents belonging to the group is computed by averaging over the results obtained by the $M = 100$ simulations.

4 Results

4.1 Results for analysis of real data

4.1.1 Inference

Following the procedure in Section **??**, we use the histogram of Δt_{ij} obtained from the training data to estimate A, after removing the marketing campaign effects. The histogram is shown in Figure **??**, and we infer that A should be 4 days.

The parameter estimates are given in Table **??**. For the covariates related to the infectious nodes, the effect of gender is estimated to be negative, meaning that male adopters will influence others more than female adopters. All the age groups 2 to 5 are estimated to have a negative effect relative to the first group (0–18 years), suggesting that the youngest adopters will influence others at a higher rate. All parameters for the adopter are statistically significantly different from zero.

Among the covariates related to the susceptible nodes in the viral part, the effect of gender is estimated to be negative but not significant. The effects for the age groups 2, 3 and 4 are estimated to be positive and significant relative to age group 1, meaning that susceptible nodes of age 18–66

Table 2: Estimated parameters for the telecom service example using an observation period of 82 days and an inferred influence period of $A = 4$ days.

Viral	Parameter	CI
α_0	-6.616	[-7.003, -6.230]
α_1		
Gender	-0.305	[-0.403, -0.208]
Age group 2	-0.831	[-1.176, -0.487]
Age group 3	-1.040	[-1.357, -0.723]
Age group 4	-1.021	[-1.347, -0.696]
Age group 5	-1.008	[-1.368, -0.649]
α_2		
Gender	-0.022	[-0.114, 0.070]
Age group 2	0.450	[0.117, 0.783]
Age group 3	0.846	[0.540, 1.152]
Age group 4	0.731	[0.418, 1.044]
Age group 5	0.022	[-0.321, 0.365]
α_3		
Voice call minutes	1.140	[1.016, 1.263]
Number of SMS	0.723	[0.584, 0.863]
Same gender	-0.446	[-0.543, -0.349]
$= 1$ age group diff.	-0.376	[-0.486, -0.267]
> 1 age group diff.	-0.727	[-0.895, -0.560]
Same municipality	0.258	[0.161, 0.355]
External		
β_0	-10.95	[-11.04, -10.85]
β_1		
Low intensity BTL	1.213	[1.175, 1.251]
Medium intensity BTL	2.176	[2.139, 2.213]
High intensity BTL	2.563	[2.530, 2.596]
β_2		
Gender	0.162	[0.136, 0.188]
Age group 2	1.365	[1.270, 1.460]
Age group 3	1.657	[1.566, 1.748]
Age group 4	1.531	[1.440, 1.623]
Age group 5	0.510	[0.414, 0.606]

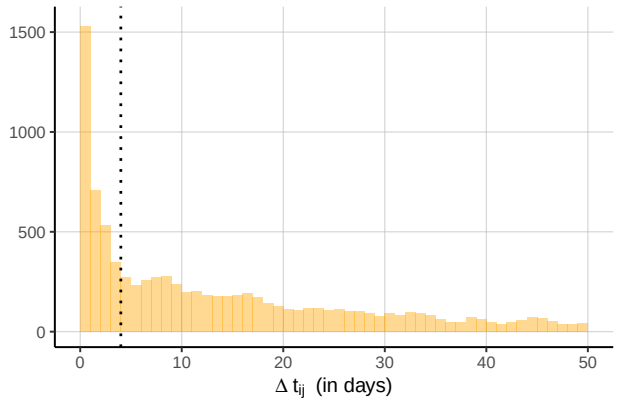


Figure 2: The plot shows the distribution of Δt_{ij} after having removed the marketing campaign effects: we conclude that A should be 4.

years will have a higher propensity to be influenced. The effect for the oldest age group ($>$ 66 years) is not significant.

For covariates related to the pairing between the adopter and the susceptible, the estimated parameters for voice call duration and number of SMSs were both positive. This indicates that more telephone contact, as a proxy for social contact, between the nodes gave a higher probability for the susceptible to adopt the product. The same holds true for the covariate measuring geographic co-location of the node pair, where co-location leads to a stronger propensity for the susceptible to adopt the product. Further, we see a negative effect for pairs of the same gender, which indicates a larger influence rate when the genders of the nodes are different. The indicators of an age difference also has a negative effect, with larger age differences yielding a larger negative effect. The pairing with the highest influence is therefore co-located agents of different gender, but equal age, with high telephone contact. All effects of the pair covariates were significant.

Contrary to the viral component, the results related to the external marketing influence indicate that females will be influenced by the external sources at a higher rate than males. The estimates related to the remaining covariates for the susceptible in the external part are comparable in sign and ordering to those given for the viral part. The estimated effects of the BTL campaigns are estimated to be positive and increasing for increasing campaign intensity, as expected. All BTL effects are significant.

4.1.2 Prediction

Using the estimated probabilistic model, we construct a prediction band for the future adoptions by running the adoption process forward in time, as explained in Section **??**. We predict the process using $M = 100$ independent simulations for the next 951 days from the end of the training period at day 82, covering the full adoption history. The mean cumulative number of adoptions (solid blue) and the 90% band (dashed blue) are shown in Figure **??**, together with the actual observed cumulative adoptions (solid black). The predicted adoptions, approximately 2.6 years into the

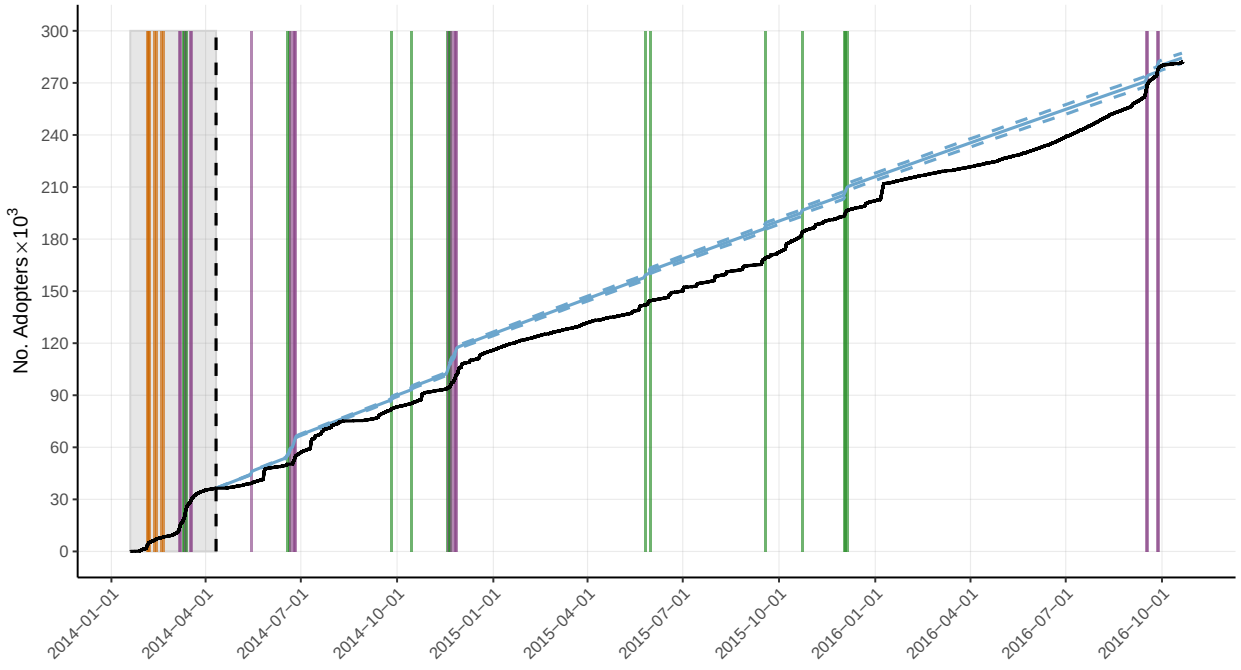


Figure 3: Predicted cumulative adoptions 138 weeks (2.6 years) after the training period of 82 days. The observed cumulative adoption curve is the solid black line; our point prediction is the solid blue line. The dashed blue lines show the uncertainty, as a 90% point-wise prediction band. Colored vertical bars show the day in which BTL campaigns were held. The intensity is color-coded, with dark orange, green, and purple signaling increasing BTL intensity.

future, fit very well with the real adoptions.

4.1.3 Results for micro-validation and what-if-scenarios

The results reported in Table **??** show that the predicted population has very similar characteristics to the true one.

As an example of what-if scenario simulations, Figure **??** shows the effect of adding hypothetical BTL campaigns. In addition to the BTL campaigns which actually took place (columns in blue), for the first what-if scenario we add the hypothetical BTLs indicated by the columns in orange. The predicted cumulative adoption curve for this what-if scenario is shown as the orange curve. For a second what-if scenario, in addition to the BTLs added for the first what-if scenario, we add even

Table 3: True and predicted shares of agents belonging to several groups.

	Female	Male	Age 1	Age 2	Age 3	Age 4	Age 5
True share	56.4%	43.6%	3.9%	14.3%	42.5%	26.4%	12.9%
Predicted share	52.3%	47.7%	3.1%	13.5%	40.5%	30.5%	12.4%

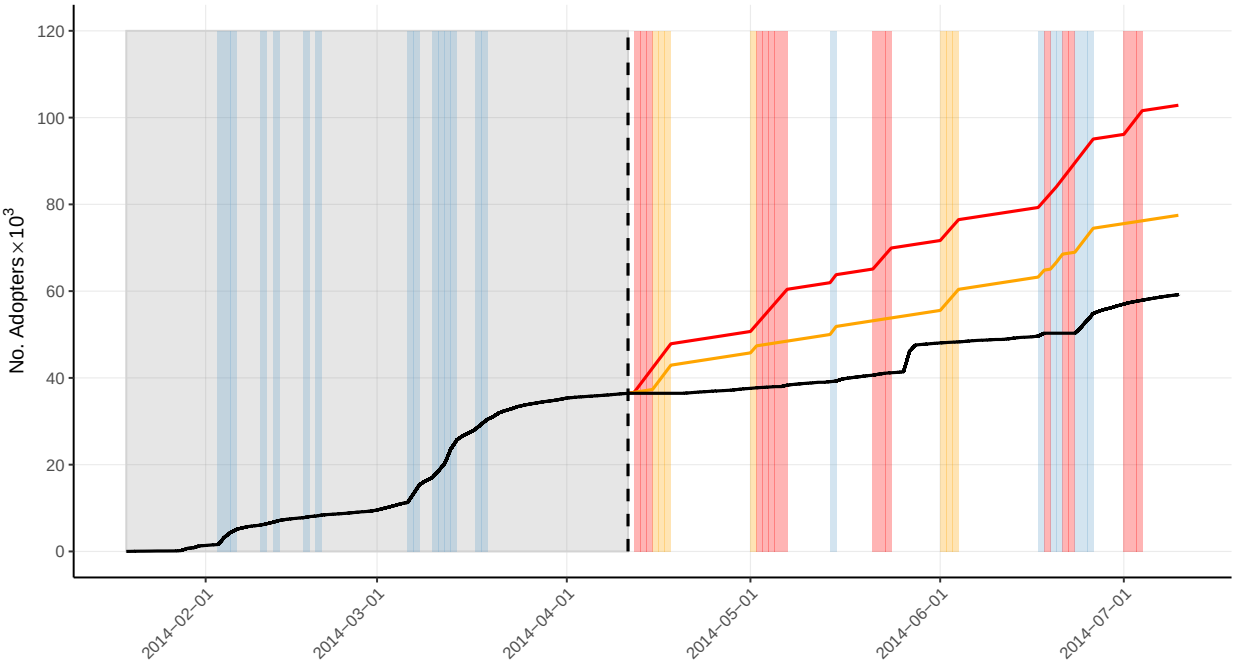


Figure 4: What-if simulations. Expected cumulative adoptions 90 days after the training period of 82 days, for two hypothetical scenarios of BTL campaigns. The observed cumulative adoption curve is the solid black line; the solid orange and red lines represent the point predictions for the two what-if scenarios. In order to produce the first scenario, the BTL indicated by the orange columns are added to the original ones (blue columns). In the second scenario we add even more BTL campaigns (red columns). Here, the color-coding distinguishes the real and hypothetical BTLs, not intensities as in Figure **??**. The hypothetical BTLs are of similar intensity as the real BTLs shown by the blue columns.

more hypothetical BTLs (red columns). The predicted cumulative adoption curve of this second what-if scenario is shown as the red curve.

4.2 Results for simulation study

4.2.1 Inference

Figure **??** shows box-plots of some estimated parameters, for varying size of the training set (complete results are in Section 2 of the Supplementary Material). When the training set contains 1000 adopters or more, all parameters are well estimated around the true value (black horizontal line). In our simulation, the viral outbreak just started around 1000 adoptions. This means that our method is capable of performing early assessment of the virality of the adoption process. When using very small training sets, estimates show larger variability around the true value, and some also exhibit additional bias. This behavior is due to lack of information about the effect of some covariates in the very small data sets. For instance, the estimates of age group 5 parameter for the susceptible nodes (Figure **??**) are not precise for small training data sets, because in our simulation

only very few customers in this specific age segment adopt early in the process. The estimates of the parameters of the edge related covariates are more precise than for those for the node-wise covariates for small training sets, as illustrated by the results for the voice call minutes parameter in Figure **??**. The reason is that for a given number of adopters, the information available about edge-wise covariates is richer than the information available about node-wise covariates.

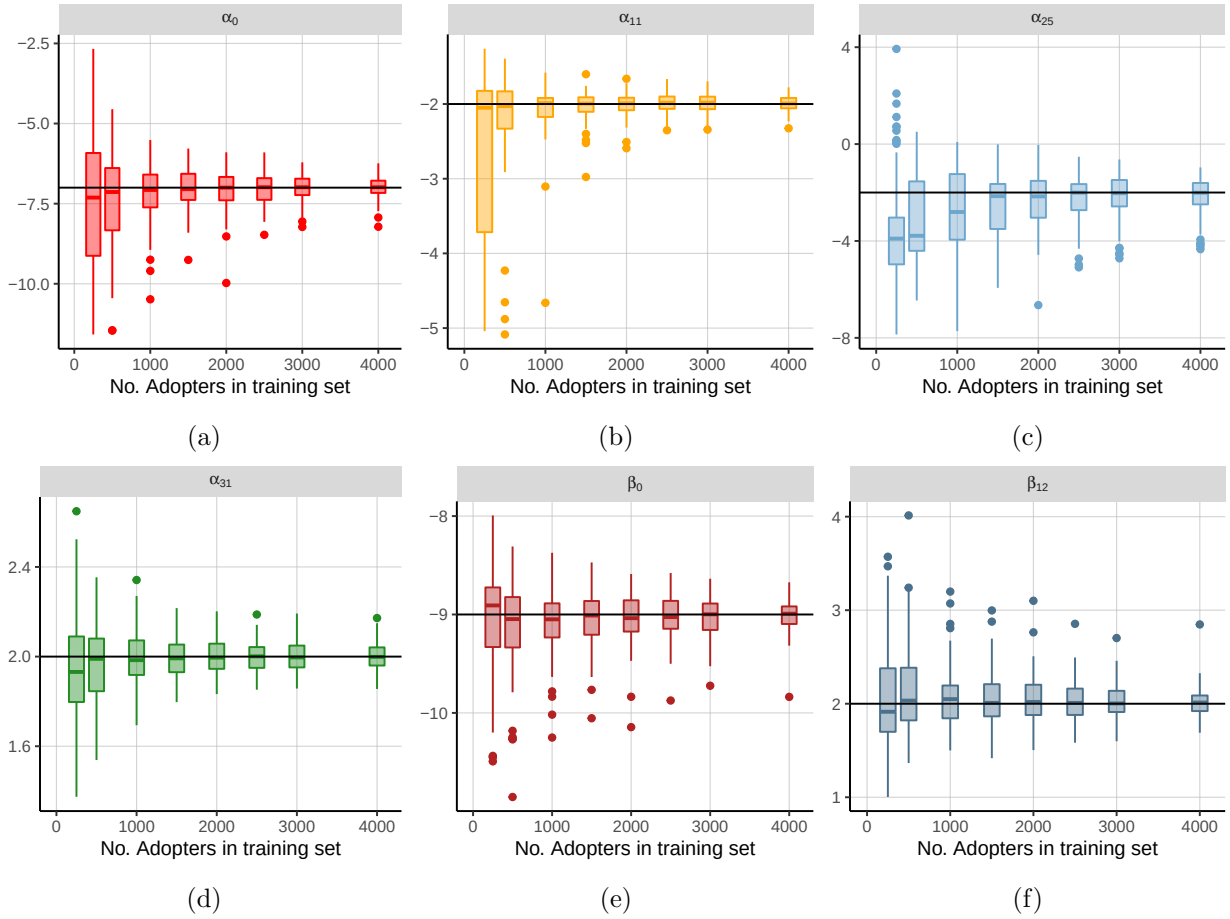


Figure 5: Boxplots for the point estimates of the parameters for (a) the baseline for the viral part, (b) gender for the infectious nodes, (c) age group 5 for the susceptible nodes, (d) voice call minutes, (e) the baseline for the external part and (f) low intensity BTL in the simulation study, for training set size 250, 500, 1000, 1500, 2000, 2500, 3000, and 4000. The solid line represents the true value of the parameter.

We also tested the method for estimating the A parameter described in Section **??**. The four panels in Figure **??** show histograms of the distribution of the time differences Δt_{ij} for four of the simulated adoption processes from Section **??**, using the first 1000 adoptions as the training set. The bin size of the histogram is set to one day. Visual inspection of these histograms shows a clear change-point at the true value of A, which was 5 for all the simulation setups.

4.2.2 Prediction

The prediction results for a single simulated adoption process is shown in Figure **??**, for training sets containing 500, 1000, and 1500 adopters. We see that the true trajectory is contained in the prediction variability band for all three training set sizes. In fact, for all the $S = 100$ adoption processes, and for all eight training set sizes, the true trajectory is always contained in the prediction variability band (complete results for four selected simulated adoption processes are shown in Section 2 of the Supplementary Material). The important changes in the viral outbreak are well captured, demonstrated from all trajectories exhibiting the 'S'-shaped curve. Hence, even though the prediction bands exhibit high variability when training data are scarce, the prediction offers a valid qualitative insight into the potential of a viral adoption. Also, as expected, the variances of the prediction bands decrease as the training set size increases.

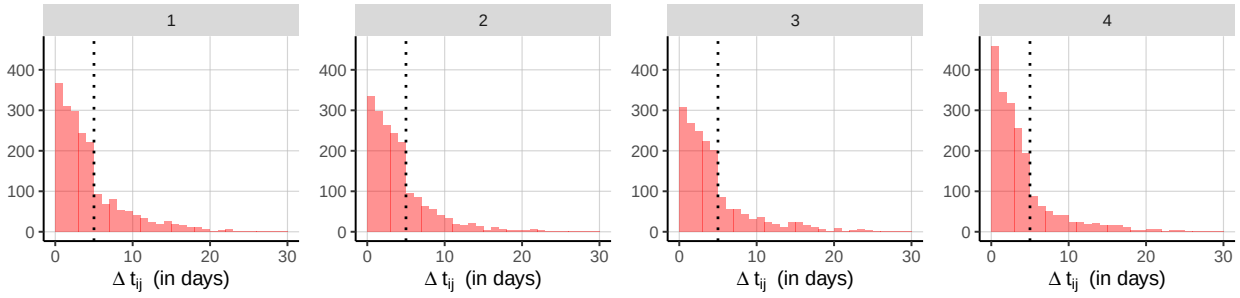
5 Discussion

5.1 Contributions to literature and implications for practice

The work presented in this paper is an important contribution to the literature on the use of data-driven ABMs for marketing innovations on social networks. It adds to previous work in different directions: We use a real agent-to-agent network, we model adoption intensities assuming a continuous time process, and we have made a scalable methodology, enabling full use of training data.

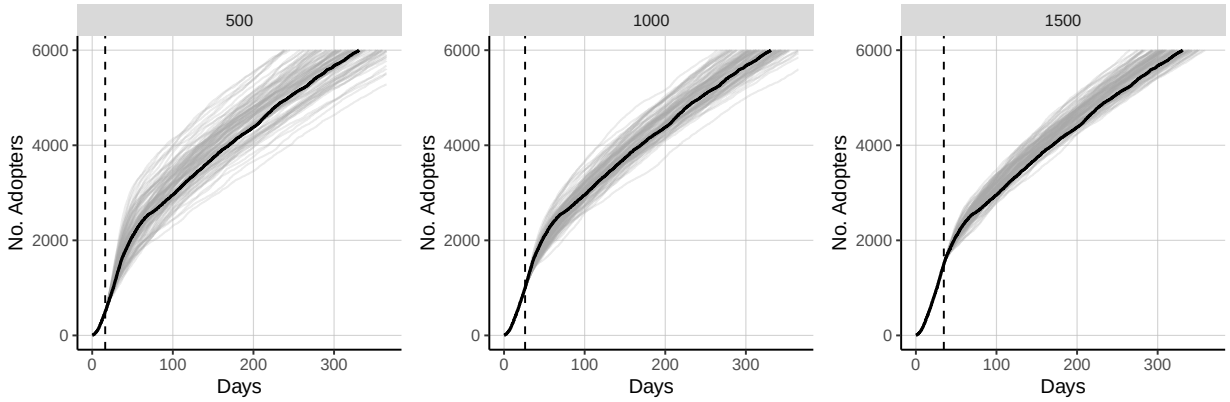
We have shown that our intensity-based, continuous time stochastic ABM is able to learn the true model parameters after a short training period, so that we can predict the future dynamics already in "early days" after launch. This has the potential to drastically improve the efficiency of the innovation process, by capturing early signals that correctly predict success or failure of the innovation, also revealing the potential benefit of different targeted marketing interventions.

Our model successfully handles all important mechanisms in the real case study. Viral effects



(a) (b) (c) (d)

Figure 6: Adoption time differences distributions: panels (a), (b), (c), and (d) show the distribution of Δt_{ij} for four simulated adoption processes, after 1,000 adoptions. The dashed line represents the true value of the parameter A, which was 5.



(a) (b) (c)

Figure 7: Prediction results of a single simulated adoption process, are shown i for three different sizes of training sets: (a) 500, (b) 1000, and (c) 1500 adoptions. The black curve is the true adoption process, the gray lines represent $M = 100$ prediction trajectories, each simulated with a parameter set drawn from the estimated multinormal approximation as described in Section **??**. The vertical dashed line indicates the number of days in the training set.

are detected in the training period — during which there seems to be some viral activity — and are small but statistically significant. The external terms are also detected and significant. Estimation results are verified by the prediction of the following two years of adoption—for which the dense core of the network can no longer sustain viral spreading, as we will argue below. The prediction results for our simulation study with a stronger viral component (Figure **??**) show that our approach can handle stronger virality also well — even when the training period includes only the very early days of the viral takeoff (leftmost panel of Figure **??**). This supports the claim that our approach can predict accurately the viral activity in the future.

Earlier studies [**?**] identified a dense core in large social networks, where the edge density is high and social spreading can be very strong. The strong rise in the classical 'S'-shaped curve of adoption is associated with strong spreading in the dense core [**?**]. The simulation results reported in Section **??**—with finite A—reveal that spreading in the dense core typically dies off without fully saturating it. Once this happens, the dense core becomes impervious to further viral spreading, while the peripheral parts of the network also do not support significant viral spreading. The result is that the adoption curve becomes dominated by the external term. This occurs both in our simulations (Figure **??**) and in the real-data (Figure **??**). Our viral simulations give new understanding of the evolution in time of viral spreading in the case of a finite influence time A, which is often a very realistic assumption.

5.2 Limitations and future work

In the following, we describe how our method can be extended in multiple directions to overcome limitations. We assume that a node $i \in B$ is infectious, i.e. can convince neighbors to adopt, for a period A after adoption. Instead we could assume a random length of this period, or a decay of

influence during a time interval $[t_i, t_i^A]$. Further, this decay could depend on agent covariates, or have different parametric forms for different subgroups. The influence could also start before t_i, to allow for node i to start influencing its neighbors already before adoption time t_i. All these options can be very important in specific situations.

It is also possible to assume that the network, covariates or viral baseline are time-varying instead than constant. The effect of the external 0–node, representing e.g. advertising campaigns, could be modeled in a more complex way—for example, incorporating a decay in time, calendar time effects, and other features of advertising campaigns.

We also believe that including the viral and exogenous effects of competing products could be important, but some data on the market penetration of such alternative products would be needed. If the competing products were offered by the same company, then an extension is relatively straightforward—because the company would have the needed information (adoptions, customer network, covariates, and campaigns) for each of the products. If the company would market both competing products to its customers, our method could predict the likely winner early on. If on the other hand product P from company Y faces competition from product Q from company Z, it is unlikely that any analyst would get complete information for both products. Instead, we imagine having only company Y's detailed information. The effect of Q on company Y's customers—in particular, on their adoption of P—could then only be informed by publicly available aggregated data, for example on adoption of Q for the customers of Y, and on marketing campaigns for Q. In short, the effect of product Q could be modeled as a negative external influence over the market.

Each agent i in the network may also be characterized by its network properties, which can be used as covariates. For example, a network measure of centrality for node i could be associated with its strength of influence. Here we note that we are allowing for a kind of 'double counting' in our approach to centrality. If we think of network centrality as a measure of well-connectedness, then it follows that well-connected nodes are well placed to influence many others — even if their ability to influence their contacts is only average. One might however also allow for the possibility that the *perception* (among the contacts) of a node's centrality (or importance) makes the central node more influencing. Two good candidate measures of centrality are the degree centrality (number of neighbors) and the eigenvector centrality [**? ? ?**]. Either of these measures is likely to contribute to the perception of a node's importance among that node's contacts.

Currently, the model allows multiple neighbors of a node j to influence j simultaneously, and we treat the effect of multiple sources of influence as being additive. However, one could also incorporate synergistic effects, such that the influence received by j from two or more influenced neighbors is different from the simple addition of incoming influence. One such model is complex contagion [**?**], defined as a threshold phenomenon, where more than one adopting friend (influencer) is needed before the node, exposed to this influence, will adopt. Complex contagion is expected to occur when what is being spread is behavioral change, rather than, say, the spread of information, or an infectious disease. Since adoption is a behavior, the possibility of complex contagion can be introduced in our approach.

Our method has shown ability to perform reliable inference early in the adoption process. Sometimes, however, it is hard to obtain good estimates of parameters related to covariates that are not well represented in the training data. A Bayesian framework can help obtain better inference faster, and also solve the aforementioned problem. Information from the history of similar past products can be incorporated nicely in informative priors for the study of a new product, and can thus help in reaching reliable inference in a situation in which data are particularly scarce. This will help make early predictions. A Bayesian framework also allows for sequential updating of the parameter inference when new data arrive. This enables capturing potential dynamic behavior of the covariate effects, which may change the adoption probabilities of certain customers during the product lifetime.

6 Conclusions

This paper extends in several directions the work on agent-based models to predict the results of a new product or service in a market. Methodologically, we develop a new intensity-based, continuous time stochastic ABM, whose parameters can be estimated by maximum likelihood in useful time. Useful time means that a real-life market simulation can be run in a few computing hours, so that its results can be interpreted and made operational before the data become obsolete. Maximum likelihood allows a coherent quantification of the uncertainty. We investigate, by realistic simulation examples, how much training data are needed to obtain good enough estimates and predictions, and conclude that the full history of adoptions in the very early phase can be sufficient to distinguish the viral part from the exogenous drift, and provide reliable and useful predictions. We apply our data-driven approach to a real data case from the telecommunication sector, showing that prediction in this case is computationally feasible and precise.

Appendix

A Computational efficiency

We achieve computational efficiency by exploiting fast sparse matrix computations, by reformulating the two parts of the likelihood in Equation (**??**). The second factor of Equation (**??**) can be written by introducing the notation $L_{0j} = \int_{T_j^{\mathrm{s}}} \lambda_{\mathrm{ext}}(t)dt$ and $L_{ij} = \int_{T_{ij}^{\mathrm{s}}} \lambda_{\mathrm{vir}} dt$ as

$$\exp\left\{-\sum_{j=1}^{n}\left[\gamma_{0j}L_{0j} + \sum_{i\in N_j}\gamma_{ij}L_{ij}\right]\right\}. \tag{A.1}$$

The L_{0j} and L_{ij} terms represent the baseline influence exerted onto node j during its period of susceptibility. These terms can be collected in matrix form as $\boldsymbol{L} = \{L_{ij}\}_{i,j=1}^{n}$ with the external terms in the diagonal, $L_{jj} = L_{0j}$, thus obtaining a compact representation of both the viral and

external influence.

The elements of $\boldsymbol{L}$ are defined for all nodes irrespective of their final adoption status. For a specific node j, however, the only non-zero elements of column j of $\boldsymbol{L}$ are those for which $i \in N_j$, as the integrals for which $i \notin N_j$ in Equation (**??**) are computed over the empty set. Because N_j is a subset of node j's relatively small set of neighbors, as compared to the whole network, $\boldsymbol{L}$ is a sparse matrix.

By collecting all covariate effects in an $n \times n$ matrix $\boldsymbol{\Gamma}$, so that $\Gamma = \{\gamma_{ij}\}_{i,j=1}^{n}$ with $\Gamma_{ii} = \gamma_{0i}$, we can express the j-th term of the outer sum in Equation (**??**) as the j-th element of the diagonal vector of the matrix product $\boldsymbol{\Gamma}^T \boldsymbol{L}$, for $j = 1, \ldots, n$. Hence, the contribution of Equation (**??**) to the log-likelihood is given by

$$\log\left(\exp\left\{-\sum_{j=1}^{n}\left[\gamma_{0j}L_{0j} + \sum_{i\in N_j}\gamma_{ij}L_{ij}\right]\right\}\right) = -\sum_{j=1}^{n}\operatorname{diag}(\boldsymbol{\Gamma}^T\,\boldsymbol{L})_j. \tag{A.2}$$

Note that by definition, $\boldsymbol{\Gamma}$ is potentially a full matrix. However, as highlighted in Equation (**??**), the only elements of $\boldsymbol{\Gamma}$ that are needed to carry out the computation are those corresponding to the non-zero elements of $\boldsymbol{L}$.

Now focusing on the first part of Equation (**??**),

$$\prod_{k\in B}\left[\lambda_{\text{ext}}(t_k)\gamma_{0k} + \sum_{i\in C_k(t_k)}\lambda_{\text{vir}}\gamma_{ik}\right], \tag{A.3}$$

let $\boldsymbol{\Lambda}$ be the $n \times n$ matrix defined for all node pairs in the network, with elements

$$\Lambda_{ik} = \begin{cases} \lambda_{\text{vir}}, & k \in B, i \in C_k(t_k), \\ \lambda_{\text{ext}}(t_k), & k \in B, i = k, \\ 0, & \text{otherwise.} \end{cases}$$

Although the non-zero elements of $\boldsymbol{\Lambda}$ could be represented by a $b \times b$ matrix, where $b = |B|$ is the size of the set B, utilizing a sparse $n \times n$ matrix simplifies the indexing of the adopter nodes. In this way each node is associated with the same row and column in $\boldsymbol{\Lambda}$, $\boldsymbol{L}$, and $\boldsymbol{\Gamma}$. Due to the relationship between the sets $C_k(t_k)$ and N_k, all non-zero elements of $\boldsymbol{\Lambda}$ appear in positions in which also $\boldsymbol{L}$ has a non-zero element, but not vice versa.

Now, organizing Equation (**??**) in a similar fashion as Equation (**??**), the contribution to the log-likelihood of Equation (**??**) can be written in compact form as

$$\log\left(\prod_{k\in B}\lambda_k(t_k)\right) = \sum_{k\in B}\log\left(\operatorname{diag}(\boldsymbol{\Gamma}^T\boldsymbol{\Lambda})_k\right).$$

B Computational time

Social networks may be very large, and we therefore investigate the time complexity of our algorithms. The estimation procedure requires the numerical optimization of the likelihood in Equation (**??**). Section **??** provides a computationally efficient formulation of the log-likelihood

$$\log \mathcal{L}(\theta) = \sum_{k \in B} \log \left(\text{diag}(\mathbf{\Gamma}^T \mathbf{\Lambda})_k \right) - \sum_{j=1}^{n} \text{diag}(\mathbf{\Gamma}^T \boldsymbol{L})_j.$$

Albeit $\mathbf{\Lambda}$ and $\boldsymbol{L}$ have off-diagonal non-zero elements, each node generally has few neighbors such that both matrices are sparse. The numbers of off-diagonal non-zero elements of $\mathbf{\Lambda}$ and $\boldsymbol{L}$ scale as a function of the number of rows, which is b, in $\mathbf{\Lambda}$, and n, in $\boldsymbol{L}$. Furthermore, the matrix $\mathbf{\Gamma}$ does not need to be filled in every position, as only its elements corresponding to the non-zero elements of $\mathbf{\Lambda}$ and $\boldsymbol{L}$ are required, hence the time needed for the numerical optimization of $\log \mathcal{L}$ for a given network is $\mathcal{O}(b + n)$.

Prediction is based on simulating new adopters using the estimated model, as in Section **??**. The time-consuming operation is the computation of the probability $\frac{\lambda_k(t)}{\sum_{S(t)} \lambda_j(t)}$ at time t, when node k adopts. This probability needs to be computed for all nodes present in $S(t)$, hence it scales with the number of susceptible nodes n_s, at time t. This operation always requires the same amount of time, as the location of the susceptible nodes in the matrices can be pre-stored to target them directly. Furthermore, the computation of the adoption probabilities for the susceptible nodes has to be repeated r times, where r denotes the number of adoptions to be simulated. It is worth to notice that when using the method for performing early prediction, i.e. with a rather small training sets, the number of susceptible nodes in a network is very close to its natural upper-bound n, the size of the network itself. This fact holds through in this setting because the number of non-susceptible nodes represent a very small share of the total number of nodes contained in the network. The time complexity of the whole prediction algorithm is then approximately $\mathcal{O}(rn)$. A further analysis of the computational times is given in Section 1.2 of the Supplementary Material.