greenrgb0.1,0.1,0.1

# Information Interaction Profile of Choice Adoption

Gael Poux-Medard[1] ✉[0000−0002−0103−8778], Julien Velcin[1][0000−0002−2262−045X], and Sabine Loudcher[1][0000−0002−0494−0169]

[1] Université de Lyon, Lyon 2, ERIC UR 3083, 5 avenue Pierre Mendès France, F69676 Bron Cedex, France
gael.poux-medard@univ-lyon2.fr
julien.velcin@univ-lyon2.fr
sabine.loudcher@univ-lyon2.fr

**Abstract.** Interactions between pieces of information (*entities*) play a substantial role in the way an individual acts on them: adoption of a product, the spread of news, strategy choice, etc. However, the underlying interaction mechanisms are often unknown and have been little explored in the literature. We introduce an efficient method to infer both the entities interaction network and its evolution according to the temporal distance separating interacting entities; together, they form the *interaction profile*. The interaction profile allows characterizing the mechanisms of the interaction processes. We approach this problem *via* a convex model based on recent advances in multi-kernel inference. We consider an ordered sequence of exposures to entities (URL, ads, situations) and the actions the user exerts on them (share, click, decision). We study how users exhibit different behaviors according to *combinations* of exposures they have been exposed to. We show that the effect of a combination of exposures on a user is more than the sum of each exposure's independent effect–there is an interaction. We reduce this modeling to a non-parametric convex optimization problem that can be solved in parallel. Our method recovers state-of-the-art results on interaction processes on three real-world datasets and outperforms baselines in the inference of the underlying data generation mechanisms. Finally, we show that interaction profiles can be visualized intuitively, easing the interpretation of the model.

## 1 Introduction

When told in the year 2000 that the XX$^{th}$ century was the century of physics and asked whether he agrees that the next one would be the century of biology, Stephen Hawkins answered that he believed the XXI$^{th}$ century would be the century of complexity. Be it a reasoned forecast or a tackle to promote scientific multidisciplinarity, there has been an undeniable growing interest for complex systems in research over the past decades. A complex system can be defined as a system composed of many components that interact with each other. Their

study often involves network theory, a branch of mathematics that aims at modeling those interactions –that can be physical, biological, social, etc. A significant point of interest is understanding how information spreads along the edges of a network–with a particular interest in social networks. If the social network skeleton (edges, nodes) plays a significant role in such processes, recent studies pointed out that the interaction between spreading entities might also play a non-trivial role in it [23, 2]. The histograms presented Fig.1 illustrate this finding: the probability for a piece of information to be adopted (or spread) varies according to the exposure to another one at a previous time. We refer to this figure as the *interaction profile*. The study of this quantity is a novel perspective: the interaction between pieces of information has been little explored in the literature, and no previous work aims at unveiling trends in the information interaction mechanisms.

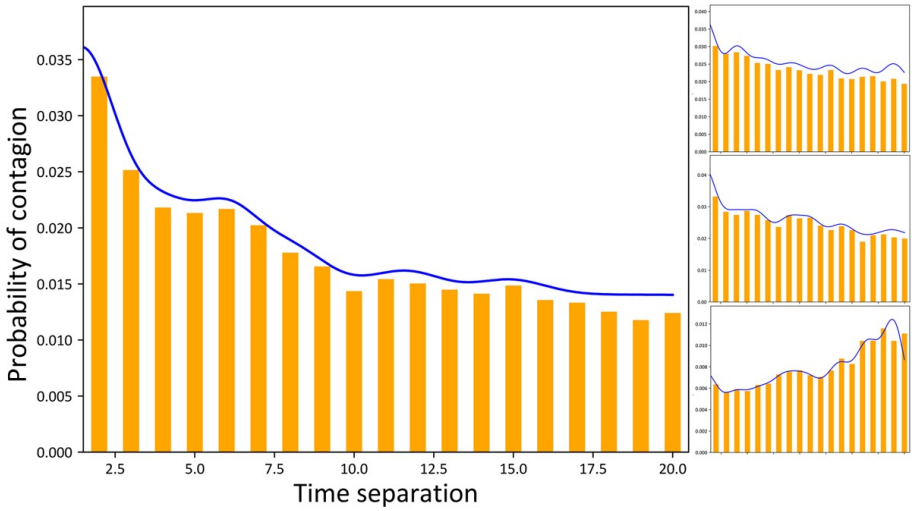


**Fig. 1. Interaction profiles between pairs of entities** — Examples of interaction profiles on Twitter; here is shown the effect of URL shortening services migre.me (left), bit.ly (right-top), tinyurl (right-middle) and t.co (right-bottom) on the probability of tweeting a t.co URL and its evolution in time. This interaction profile shows, for instance, that there is an increased probability of retweet for a t.co URL when it appears shortly after a migre.me one (interaction). This increase fades when the time separation grows (no more interaction). In blue, the interaction profile inferred by our model.

The study of interactions between pieces of information (or entities) has several applications in real-world systems. We can mention the fields of recommender systems (the probability of adoption is influenced by what a user saw shortly prior to it), news propagation and control (when to expose users to an entity in order to maximize its spreading probability [20]), advertising (same

reasons as before [3]), choice behavior (what influences one's choice and how [4]).

In the present work, we propose to go one step further and to unveil the mechanisms at stake within those interacting processes by inferring information interaction profiles. Let us imagine, for instance, that an internet user is exposed to an ad at time $t_1$ and to another ad for a similar product at time $t_2 > t_1$. Due to the semantic similarity between the two exposures, we suppose that the exposure to the first one influences the user's sensitivity (likeliness of a click) to the second one a time $t_2 - t_1$ later. Modeling this process involves quantifying the influence an ad exerts on the other and how it varies with the time separation between the exposures. The reunion of those quantities form what we call the *interaction profile* –illustrated Fig.1), that is, the influence an exposure exerts on the adoption (click, buy, choice, etc.) of another one over time.

Following this idea, we introduce an efficient method to infer both the entities (ad, tweets, products, etc.) interaction network and its evolution according to the temporal distance separating the interacting entities (the influence of an entity A on entity B will not be the same depending on whether A appeared 10 minutes or 10 hours before B). Together they form the interaction profile.

First, we develop a method for inferring this interaction profile in a continuous-time setup using multi-kernel inference methods [5]. Then we show that the inference of the parameters boils down to a convex optimization problem for kernel families obeying specific properties. Moreover, the problem can be subdivided into as many subproblems as entities, which can be solved in parallel. The convexity of the problem guarantees convergence to the likelihood's global optimum for each subproblem and, therefore, to the problem's optimal likelihood. We apply the model to investigate the role of interaction profiles on synthetic data and in various corpora from different fields of research: advertisement (the exposure to an ad influences the adoption of other ads [3]), social dilemmas (the previous actions of one influences another's actions [4]) and information spread on Twitter (the last tweets read influence what a user retweets [12])[1]. Finally, we provide analysis leads and show that our method recovers state-of-the-art results on interaction processes on each of the three datasets considered.

### 1.1 Contributions

The main contributions of this paper are the following:

- We introduce the interaction profile, which is the combination of both the interaction network between entities and its evolution with the interaction time distance, according to the inferred kernels. The interaction profile is a powerful tool to understand how interactions take place in a given corpus (see Fig.4) and has not been developed in the literature. Its introduction in research is the main contribution of the present work.

[1] Supplementary materials, codes and datasets can be found at https://github.com/GaelPouxMedard/InterRate

- We design a convex non-parametric algorithm that can be solved in parallel, baptized InterRate. InterRate automatically infers the kernels that account the best for information interactions in time within a given kernel family. Its output is the aforementioned interaction profile.
- We show that InterRate yields better results than non-interacting or non-temporal baseline models on several real-world datasets. Furthermore, our model can recover several conclusions about the datasets from state-of-the-art works.

## 2 Related work

Previous efforts in investigating the role of interactions in information diffusion have shown their importance in the underlying spreading processes. Several works study the interaction of information with users' attention [11], closely linked to information overload concepts [18], but not the interaction between the pieces of information themselves. On the other hand, whereas most of the modeling of spreading processes are based on either no competition [19, 15] or perfect competition [17] assumption, it has been shown that relaxing this hypothesis leads to a better description of competitive spread [2] –with the example of Firefox and Chrome web browsers, whose respective popularities are correlated. According to this finding, a significant effort has been done in elaborating complex processes to *simulate* interaction [22, 17] on real-world networks.

However, fewer works have been developed to tackle interaction in information spread from a machine learning point of view. The correlated cascade model [23] aims to infer an interacting spreading process's latent diffusion network. In this work, the interaction is modeled by a hyper-parameter $\beta$ tuning the intensity of interactions according to an exponentially decaying kernel. In their conclusion, the authors formulate the open problem of learning several kernels and the interaction intensity parameter $\beta$, which we address in the present work.

To our knowledge, the attempt the closest to our task to model the interaction intensity parameter $\beta$ is Clash of the contagions [12]; this aims to predict retweets on Twitter based on tweets seen by a user. This model estimates the probability of retweet for a piece of information, given the last tweets a user has been exposed to, according to their relative position in the Twitter feed. The method suffers various flaws (scalability, non-convexity). It also defines interactions based on an arguable hypothesis made on the prior probability of a retweet (in the absence of interactions) that makes its conclusions about interactions sloppy. It is worth noting that in [12], the authors outline the problem of the inference of the interaction profile but do so without searching for global trends such as the one shown in Fig.1. Recent works address the various flaws observed in [12] and suggest a more general approach to the estimation of the interaction intensity parameters [16]. The latter model develops a scalable algorithm that correctly accounts for interacting processes but neglects the interactions' temporal aspect. To take back the Twitter case study, it implies that in the case of a retweet at time $t$, a tweet appearing at $t_1 \ll t$ in the news feed has the same influence on

the retweet as a tweet that appeared at $t_2 \approx t$. A way to relax this assumption is to integrate a temporal setting in the interaction network inference problem.

In recent years, temporal networks inference has been a subject of interest. Significant advances have been made using survival theory modeling applied to partial observations of independent cascades of contagions [6, 7]. In this context, an infected node tries to contaminate every other node at a rate that is tuned by $\beta$. While this work is not directly linked to ours, it has been a strong influence on the interaction profile inference problem we develop here; the problems are different, but the methodology they introduce deserves to be mentioned for its help in the building of our model (development and convexity of the problem, analogy between interaction profile and hazard rate). Moreover, advances in network inference based on the same works propose a multi-kernel network inference method that we adapted to the problem we tackle here [5]. Inspired by these works, we develop a flexible approach that allows for the inference of the best interaction profile from several candidate kernels.

## 3 InterRate

### 3.1 Problem definition

We illustrate the process to model in Figure 2. It runs as follows: a user is first exposed to a piece of information at time $t_0$. The user then chooses whether to act on it at time $t_0 + t_s$ (an act can be a retweet, a buy, a booking, etc.); $t_s$ can be interpreted as the "reaction time" of the user to the exposure, assumed constant. The user is then exposed to the next piece of information a time $\delta t$ later, at $t_1 = t_0 + \delta t$ and decides whether to act on it a time $t_s$ later, at $t_1 + t_s$, and so on. Here, $\delta t$ is the time separating two consecutive exposures, and $t_s$ is the reaction time, separating the exposure from the possible contagion. In the remaining of the paper, we refer to the user's action on an exposure (tweet appearing in the feed, exposure to an ad, etc.) as a contagion (retweet or the tweet, click on an ad, etc.).

This choice of modeling comes with several hypotheses. **First**, the pieces of information a user is exposed to appear independently from each other. It is the main difference between our work and survival analysis literature: the pseudo-survival of an entity is conditioned by the random arrival of pieces of information. Therefore, it cannot be modeled as a point process. This assumption holds in our experiments on real-world datasets, where users have no influence on what information they are exposed to. **Second** hypothesis, the user is contaminated solely on the basis of the previous exposures in the feed [12, 23]. **Third**, the reaction time separating the exposure to a piece of information from its possible contagion, $t_s$, is constant (i.e. the time between a read and a retweet in the case of Twitter). Importantly, this hypothesis is a deliberate simplification of the model for clarity purposes; relaxing this hypothesis is straightforward by extending the kernel family, which preserves convexity and time complexity. Note that this simplification does not always hold, as shown in recent works concluding that response time can have complex time-dependent mechanisms [21].

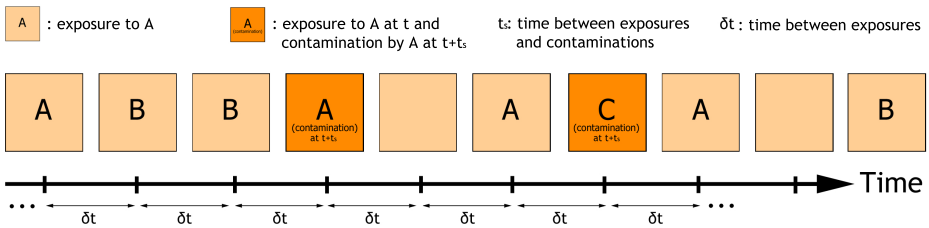


**Fig. 2. Illustration of the interacting process** — Light orange squares represent the exposures, dark orange squares represent the exposures that are followed by contagions and empty squares represent the exposures to the information we do not consider in the datasets (they only play a role in the distance between exposures when one considers the order of appearance as a time feature). A contagion occurs at a time $t_s$ after the corresponding exposure. Each new exposure arrives at a time $\delta$t after the previous one. Contagion takes place with a probability conditioned by all previous exposures. In the example, the contagion by A at time $t+t_s$ depends on the effect of the exposure to A at times $t$ and $t-3\delta t$, and to B at times $t-\delta t$ and $t-2\delta t$.

### 3.2 Likelihood

We define the likelihood of the model of the process described in Figure 2. Let $t_i^{(x)}$ be the exposure to $x$ at time $t_i$, and $t_i^{(x)}+t_s$ the time of its possible contagion. Consider now the instantaneous probability of contagion (*hazard function*) $H(t_i^{(x)}+t_s|t_j^{(y)},\beta_{xy})$, that is the probability that a user exposed to the piece of information $x$ at time $t_i$ is contaminated by $x$ at $t_i+t_s$ given an exposure to $y$ at time $t_j \leq t_i$. The matrix of parameters $\beta_{ij}$ is what the model aims to infer. $\beta_{ij}$ is used to characterize the interaction profile between entities. We define the set of exposures preceding the exposure to $x$ at time $t_i$ (or history of $t_i^{(x)}$) as $\mathcal{H}_i^{(x)} \equiv \{t_j^{(y)} \leq t_i^{(x)}\}_{j,y}$. Let $\mathcal{D}$ be the whole dataset such as $\mathcal{D} \equiv \{(\mathcal{H}_i^{(x)}, t_i^{(x)}, c_{t_i}^{(x)})\}_{i,x}$. Here, c is a binary variable that account for the contagion ($c_{t_i}^{(x)}=1$) or non-contagion ($c_{t_i}^{(x)}=0$) of $x$ at time $t_i+t_s$. The likelihood for one exposure in the sequence given $t_j^{(y)}$ is:

$$L(\beta_{xy}|\mathcal{D},t_s) = P(\mathcal{D}|\beta_{xy},t_s) = \underbrace{H(t_i^{(x)}+t_s|t_j^{(y)},\beta_{xy})^{c_{t_i}^{(x)}}}_{\text{contagion at } t_i^{(x)}+t_s \text{ due to } t_j^{(y)}} \cdot \underbrace{(1-H(t_i^{(x)}+t_s|t_j^{(y)},\beta_{xy}))^{(1-c_{t_i}^{(x)})}}_{\text{Survival at } t_i^{(x)}+t_s \text{ due to } t_j^{(y)}}$$

The likelihood of a sequence (as defined Fig.2) is then the product of the previous expression over all the exposures that happened before the contagion event $t_i^{(x)}+t_s$ e.g. for all $t_j^{(y)} \in \mathcal{H}_i^{(x)}$. Finally, the likelihood of the whole dataset $\mathcal{D}$ is the product of $L(\beta_x|\mathcal{D},t_s)$ over all the observed exposures $t_i^{(x)}$. Taking the logarithm

of the resulting likelihood, we get the final log-likelihood to maximize:

$$\begin{aligned}\ell(\beta|\mathcal{D}, t_s) = \\ \sum_{\mathcal{D}} \sum_{t_j^{(y)} \in \mathcal{H}_i^{(x)}} c_{t_i}^{(x)} \log\Big(H(t_i^{(x)} + t_s | t_j^{(y)}, \beta_{xy})\Big) \\ + (1 - c_{t_j}^{(y)}) \log\Big(1 - H(t_i^{(x)} + t_s | t_j^{(y)}, \beta_{xy})\Big)\end{aligned} \tag{1}$$

### 3.3 Proof of convexity

The convexity of a problem guarantees to retrieve its optimal solution and allows using dedicated fast optimization algorithms.

**Proposition 1** *The inference problem* $\min_\beta -\ell(\beta|\mathcal{D}, t_s)\ \forall \beta \geq 0$*, is convex in all of the entries of* $\beta$ *for any hazard function that obeys the following conditions:*

$$\begin{cases} H'^2 \geq H''H \\ H'^2 \geq -H''(1-H) \\ H \in ]0;1[ \end{cases} \tag{2}$$

*where* $'$ *and* $''$ *denote the first and second derivative with respect to* $\beta$*, and* $H$ *is the shorthand notation for* $H(t_i^{(x)} + t_s | t_j^{(y)}, \beta_{xy})\ \forall i, j, x, y$*.*

*Proof.* The negative log-likelihood as defined in Eq.1 is a summation of $-\log H$ and $-\log(1-H)$; therefore $H \in ]0;1[$. The second derivative of these expressions according to any entry $\beta_{mn}$ (noted $''$) reads:

$$\begin{cases} (-\log H)'' = \left(\frac{-H'}{H}\right)' = \frac{H'^2 - H''H}{H^2} \\ (-\log(1-H))'' = \left(\frac{H'}{1-H}\right)' = \frac{H'^2 + H''(1-H)}{(1-H)^2} \end{cases} \tag{3}$$

The convexity according to a single variable holds when the second derivative positive, which leads to Eq.2. The convexity of the problem then follows from composition rules of convexity. □

A number of functions obey the conditions of Eq.2, such as the exponential ($e^{-\beta t}$), Rayleigh ($e^{-\frac{\beta}{2}t^2}$), power-law ($e^{-\beta \log t}$) functions, and any log-linear combination of those [5]. These functions are standard in survival theory literature [8].

The final convex problem can then be written $\min_{\beta \geq 0} -\ell(\beta|\mathcal{D}, t_s)$. An interesting feature of the proposed method is that the problem can be subdivided into N convex subproblems that can be solved independently (one for each piece of information). To solve the subproblem of the piece of information $x$, that is to find the vector $\beta_x$, one needs to consider only the subset of $\mathcal{D}$ where $x$ appears. Explicitly, each subproblem consists in maximizing Eq.1 over the set of observations $\mathcal{D}^{(x)} \equiv \{(\mathcal{H}_i^{(x)}, t_i^{(x)}, c_{t_i}^{(x)})\}_i$.

# 4 Experimental setup

## 4.1 Kernel choice

**Gaussian RBF kernel family (IR-RBF)** Based on [5], we consider a log-linear combination of Gaussian radial basis function (RBF) kernels as hazard function. We also consider the time-independent kernel needed to infer the base probability of contagion discussed in the section "Background noise in the data" below. The resulting hazard function is then:

$$\log H(t_i^{(x)} + t_s | t_j^{(y)}, \beta_{ij}) = -\beta_{ij}^{(bg)} - \sum_{s=0}^{S} \frac{\beta_{ij}^{(s)}}{2} (t_i + t_s - t_j - s)^2$$

The parameters $\beta^{(s)}$ of Rayleigh kernels are the amplitude of a Gaussian distribution centered on time $s$. The parameter S represents the maximum time shift we consider. In our setup, we set S=20. We think it is reasonable to assume that an exposition does not significantly affect a possible contagion 20 steps later. The parameter $\beta_{ij}^{(bg)}$ corresponds to the time-independent kernel –base probability of contagion by i. The formulation allows the model to infer complex distributions from a reduced set of parameters whose interpretation is straightforward.

**Exponentially decaying kernel (IR-EXP)** We also consider an exponentially decaying kernel that can be interpreted as a modified version of a multivariate Hawkes point process (or self-exciting point process). A multivariate Hawkes process is a point process where different classes of objects exert an influence on the point process of the others [9]. The most common (and historical) form of such a process considers a time-decaying exponential function to model a class's influence on others. However, in our setup, we cannot consider a Hawkes process: we infer the variation of contagion probabilities in time conditionally on the earlier random exposure to a piece of information. The studied process is therefore not rigorously self-exciting. We consider the following form for the hazard function and refer to this modeling as IR-EXP instead of a Hawkes process:

$$\log H(t_i^{(x)} + t_s | t_j^{(y)}, \beta_{xy}) = -\beta_{ij}^{(bg)} - \beta_{ij}(t_i + t_s - t_j)$$

Where $\beta_{ij}^{(bg)}$ once again accounts for the background noise in the data discussed further in this section.

## 4.2 Parameters learning

Datasets are made of sequences of exposures and contagions, as shown in Fig.2. To assess the robustness of the proposed model, we apply a 5-folds cross-validation method. After shuffling the dataset, we use 80% of the sequences as a training set and the 20% left as a test set. We repeat this slicing five times, taking care that an interval cannot be part of the test set more than once. The optimization is made in parallel for each piece of information via the convex optimization module for Python CVXPY.

We also set the time separating two exposures $\delta t$ as constant. It means that we consider only the order of arrival of exposures instead of their absolute arrival time. The hypothesis that the order of exposures matters more than the absolute exposure times has already been used with success in the literature [12]. Besides, in some situations, the exact exposure time cannot be collected, while the exposures' order is known. For instance, in a Twitter corpus, we only know in what order a user read her feed, unlike the exact time she read each of the posts. However, from its definition, our model works the same with non-integer and non-constant $\delta t$ in datasets where absolute time matters more than the order of appearance.

### 4.3 Background noise in the data

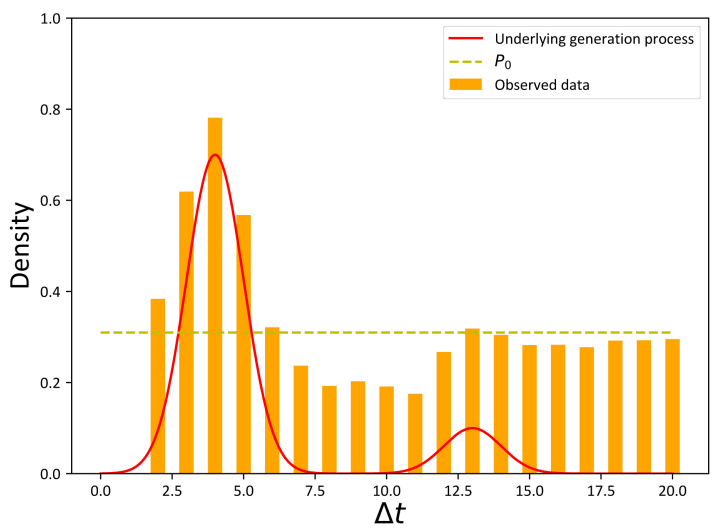


**Fig. 3. Underlying generation process vs observed data** — The red curve represents the underlying probability of contagion by C given an exposure observed $\Delta t$ steps before C. The orange bars represent the observed probability of such events. We see that there is a noise $P_0(C)$ in the observed data. The underlying generation process can then only be observed in the dataset when its effect is larger than some threshold $P_0(C)$.

Because the dataset is built looking at all exposure-contagion correlations in a sequence, there is inherent noise in the resulting data. To illustrate this, we look at the illustrated example Fig.2 and consider the exposure to C leading to a contagion happening at time $t$. We assume that in the underlying interaction process, the contagion by C at time $t+t_s$ took place only because C appeared at

time $t$. However, when building the dataset, the contagion by C is also attributed to A appearing at times $t - \delta t$, $t - 3\delta t$ and $t - 6\delta t$, and to B appearing at times $t - 4\delta t$ and $t - 5\delta t$. It induces a noise in the data. In general, for any contagion in the dataset, several observations (pair exposure-contagion) come from the random presence of entities unrelated to this contagion.

We now illustrate how this problem introduces noise in the data. In Fig.3, we see that the actual underlying data generation process (probability of a contagion by C given an exposure present $\Delta t$ step earlier) does not exactly fit the collected resulting data: the data gathering process induces a constant noise whose value is noted $P_0(C)$ –that is the average probability of contagion by C. Thus the interaction effect can only be observed when its associated probability of contagion is larger than $P_0(C)$. Consequently, the performance improvement of a model that accounts for interactions may seem small compared with a baseline that only infers $P_0(C)$. That is we observe in the experimental section. However, in this context, a small improvement in performance shows an extended comprehension of the underlying interacting processes at stake (see Fig.3, where the red line obviously explains the data better than a constant baseline). Our method efficiently infers $P_0(C)$ *via* a time-independent kernel function $\beta_{i,j}^{P_0(i)}$.

### 4.4 Evaluation criteria

The main difficulty in evaluating these models is that interactions might occur between a small number of entities only. It is the case here, where many pairs of entities have little to no interaction (see the Discussion section). This makes it difficult to evaluate how good a model is at capturing them. To this end, our principal metric is the residual sum of squares (**RSS**). The RSS is the sum of the squared difference between the observed and the expected frequency of an outcome. This metric is particularly relevant in our case, where interactions may occur between a small number of entities: any deviation from the observed frequency of contagion is accounted for, which is what we aim at predicting here. We also consider the Jensen-Shannon (**JS**) divergence; the JS divergence is a symmetric version of the Kullback–Leibler divergence, which makes it usable as a metric [14].

We finally consider the best-case F1-score (**BCF1**) of the models, that is, the F1-score of the best scenario of evaluation. It is not the standard F1 metric (that poorly distinguishes the models since few interactions occur), although its computation is similar. Explicitly, it generalizes F1-score for comparing probabilities instead of comparing classifications; the closer to 1, the closer the inferred and observed probabilities. It is derived from the best-case confusion matrix, whose building process is as follows: we consider the set of every information that appeared before information i at time $t_i$ in the interval, that we denote $\mathcal{H}_i$. We then compute the contagion probability of i at time $t_i + t_s$ to every exposure event $t_j^{(y)} \in \mathcal{H}_i$. Confronting this probability with the observed frequency f of contagions of i at time $t_i + t_s$ given $t_j^{(y)}$ among N observations, we can build the best-case confusion matrix. In the best case scenario, if out of N observations

the observed frequency is f and the predicted frequency is p, the number of True Positives is $N \times \min\{p, f\}$, the number of False Positives is $N \times \min\{p - f, 0\}$, the number of True Negatives is $N \times \min\{1 - p, 1 - f\}$, the number of False Positives is $N \times \min\{f - p, 0\}$.

Finally, when synthetic data is considered, we also compute the mean squared error of the $\beta$ matrix inferred according to the $\beta$ matrix used to generate the observations, that we note MSE $\beta$.

We purposely ignore evaluation in prediction because, as we show later, interactions influence quickly fades over time: probabilities of contagion at large times are mainly governed by the background noise discussed in previous sections. Therefore, it would be irrelevant to evaluate our approach's predictive power on the whole range of times where it does not bring any improvement over a naive baseline (see Fig.1). A way to alleviate this problem would be to make predictions only when interactions effects are above/below a certain threshold (at short times, for instance). However, such an evaluation process would be debatable. Here, we choose to focus on the descriptive aspect of InterRate.

### 4.5 Baselines

**Naive baseline** For a given piece of information i, the contagion probability is defined as the number of times this information is contaminated divided by its number of occurrences.

**Clash of the contagions** We use the work presented in [12] as a baseline. In this work, the authors model the probability of a retweet given the presence of a tweet in a user's feed. This model does not look for trends in the way interactions take place (it does not infer an interaction profile), considers discrete time steps (while our model works in a continuous-time framework), and is optimized via a non-convex SGD algorithm (which does not guarantee convergence towards the optimal model). More details on implementation are provided in SI file.

**IMMSBM** The Interactive Mixed-Membership Stochastic Block Model is a model that takes interactions between pieces of information into account to compute the probability of a (non-)contagion [16]. Note that this baseline does not take the position of the interacting pieces of information into account (time-independent) and assumes that interactions are symmetric (the effect of A on B is the same as B on A).

**ICIR** The Independent Cascade InterRate (ICIR) is a reduction of our main IR-RBF model to the case where interactions are not considered. We consider the same dataset, enforcing the constraint that off-diagonal terms of $\beta$ are null. The (non-)contagion of a piece of information i is then determined solely by the previous exposures to i itself.

# 5 Results

## 5.1 Synthetic data

**Table 1. Experimental results** — Darker is better (linear scale). Our model outperforms all of the baselines in almost every dataset for every evaluation metric. The standard deviations of the 5 folds cross-validation are negligible and reported in SI file.

| | | RSS | JS div. | BCF1 | MSE $\beta$ |
|---|---|---|---|---|---|
| **Synth-20** | IR-RBF | 18.4151 | 0.0023 | 0.9188 | 0.0005 |
| | ICIR | 139.5926 | 0.0100 | 0.8270 | 0.0159 |
| | Naive | 145.5132 | 0.0104 | 0.8221 | |
| | CoC | 123.0583 | 0.0094 | 0.8220 | |
| | IMMSBM | 222.0555 | 0.0173 | 0.7265 | |
| **Synth-5** | IR-RBF | 0.1169 | 0.0002 | 0.9742 | 0.0053 |
| | ICIR | 8.2661 | 0.0081 | 0.8499 | 0.0192 |
| | Naive | 10.0264 | 0.0100 | 0.8214 | |
| | CoC | 0.1154 | 0.0002 | 0.9763 | |
| | IMMSBM | 11.6936 | 0.0136 | 0.7693 | |

We generate synthetic data according to the process described in Fig.2 for a given $\beta$ matrix using the RBF kernel family. First, we generate a random matrix $\beta$, whose entries are between 0 and 1. A piece of information is then drawn with uniform probability and can result in a contagion according to $\beta$, the RBF kernel family and its history. We simulate the outcome by drawing a random number and finally increment the clock. The process then keeps on by randomly drawing a new exposure and adding it to the sequence. We set the maximum length of intervals to 50 steps and generate datasets of 20,000 sequences.

We present in Tab. 1 the results of the various models with generated interactions between 20 (Synth-20) and 5 (Synth-5) entities. The interactions are generated using the RBF kernel, hence the fact we are not evaluating the IR-EXP model –its use would be irrelevant. The InterRate model outperforms the proposed baselines for every metric considered. It is worth noting that performances of non-interacting and/or non-temporal baselines are good on the JS divergence and F1-score metrics due to the constant background noise $P_0$. For cases where interactions do not play a significant role, IMMSBM and Naive models perform well by fitting only the background noise. By contrast, the RSS metric distinguishes very well the models that are better at modeling interactions.

Note that while the baseline [12] yields good results when few interactions are simulated (Synth-5), it performs as bad as the naive baseline when this number increases (Synth-20). This is due to the non-convexity of the proposed model, which struggles to reach a global maximum of the likelihood even after 100 runs (see supplementary materials for implementation details).

### 5.2 Real data

**Table 2. Experimental results** — Darker is better (linear scale). Our model outperforms all of the baselines in almost every dataset for every evaluation metric. The standard deviations of the 5 folds cross-validation are negligible and reported in SI file.

| | | RSS | JS div. | BCF1 |
|---|---|---|---|---|
| **Twitter** | IR-RBF | 0.0015 | 0.0001 | 0.9832 |
| | IR-EXP | 0.0011 | 0.0000 | 0.9862 |
| | ICIR | 0.0137 | 0.0006 | 0.9614 |
| | Naive | 0.0161 | 0.0007 | 0.9379 |
| | CoC | 0.0017 | 0.0001 | 0.9572 |
| | IMMSBM | 0.0147 | 0.0007 | 0.9543 |
| **PD** | IR-RBF | 1.1268 | 0.0076 | 0.9789 |
| | IR-EXP | 1.5526 | 0.0087 | 0.9661 |
| | ICIR | 3.5359 | 0.0182 | 0.9381 |
| | Naive | 3.6527 | 0.0191 | 0.9455 |
| | CoC | 1.2409 | 0.0081 | 0.9736 |
| | IMMSBM | 20.3773 | 0.0870 | 0.7672 |
| **Ads** | IR-RBF | 0.0043 | 0.0000 | 0.9814 |
| | IR-EXP | 0.0030 | 0.0000 | 0.9852 |
| | ICIR | 0.0983 | 0.0008 | 0.9659 |
| | Naive | 0.1453 | 0.0013 | 0.9126 |
| | CoC | 0.0045 | 0.0000 | 0.9741 |
| | IMMSBM | 0.0155 | 0.0002 | 0.9543 |

We consider 3 real-world datasets. For each dataset, we select a subset of entities that are likely to interact with each other. For instance, it has been shown that the interaction between the various URL shortening services on Twitter is non-trivial [23]. The datasets are a **Twitter** dataset (104,349 sequences, exposure are tweets and contagions are retweets) [10], a Prisoner's dilemma dataset (**PD**) (2,337 sequences, exposures are situations, contagions are players choices) [13, 1] and an **Ads** dataset (87,500 sequences, exposures are ads, contagions are clicks on ads) [3]. A detailed description of the datasets is presented in SI file, section Datasets.

The results on real-world datasets are presented in Tab.2. We see that the IMMSBM baseline performs poorly on the PD dataset: either considering the time plays a consequent role in the probability of contagion, or interactions are not symmetric. Indeed, the core hypothesis of the IMMSBM is that the effect of exposition A on B is the same as B on A, whichever is the time separation between them. In a prisoner's dilemma game setting, for instance, we expect that a player does not react in the same way to defection followed by cooperation as to cooperation followed by defection, a situation for which the IMMSBM does not account. When there are few entities, the CoC baseline performs as good as IR, but fails when this number increases; this is due mainly to the non-convexity of

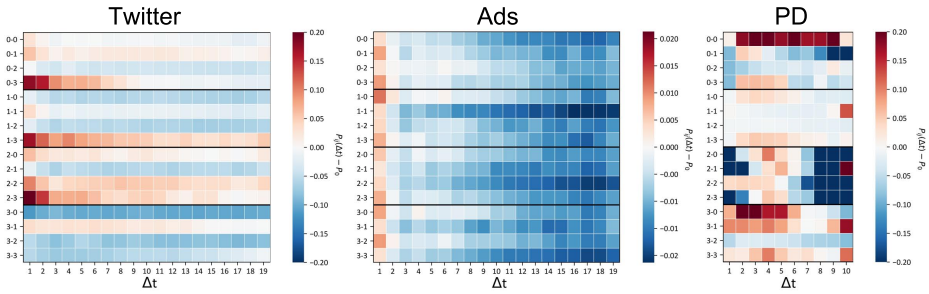


**Fig. 4. Visualization of the interaction profiles** — Intensity of the interactions between every pair of entities according to their time separation (one line is one pair's interaction profile, similar to Fig.1 seen "from the top"). A positive intensity means that the interaction helps the contagion, while a negative intensity means it blocks it. The key linking numbers on the y-axis to names for each dataset is provided in the Datasets section.

the problem that does not guarantee convergence towards the optimal solution. Overall, the InterRate models yield the best results on every dataset.

## 6 Discussion

In Fig.4, we represent the interaction intensity over time for every pair of information considered in every corpus fitted with the RBF kernel model. The intensity of the interactions is the inferred probability of contagion minus the base contagion probability in any context: $P_{ij}(t) - P_0(i)$. Therefore, we can determine the characteristic range of interactions, investigate recurrent patterns in interactions, whether the interaction effect is positive or negative, etc. Overall, we understand why the EXP kernel performs as good as the RBF on the Twitter and Ads datasets: interactions tend to have an exponentially decaying influence over time. However, this is not the case on the PD dataset: the effect of a given interaction is very dependent on its position in the history (pike on influence at $\Delta t = 3$, shift from positive to negative influence, etc.).

In the Twitter dataset, the most substantial positive interactions occur before $\Delta$t=3. This finding agrees with previous works, which stated that the most informative interactions within Twitter URL dataset occur within the 3 time steps before the possible retweet [12]. We also find that the vast majority of interactions are weak, matching with previous study's findings [12, 16]. However, it seems that tweets still exert influence even a long time after being seen, but with lesser intensity.

In the Prisoner's Dilemma dataset, players' behaviors are heavily influenced by the previous situations they have been exposed to. For instance, in the situation where both players cooperated in the previous round (pairs 2-x, $3^{rd}$ section in Fig.4-PD). The probability that the player defects is then significantly in-

creased if both players cooperated or if one betrayed the other exactly two rounds before but decreased if it has been two rounds that players both cooperate.

Finally, we find that the interactions play a lesser role in the clicks on ads. We observe a slightly increased probability of click on every ad after direct exposure to another one. We also observe a globally decreasing probability of click when two exposures distant in time, which agrees with previous work's findings [3]. Finally, the interaction profile is very similar for every pair of ads; we interpret this as a similarity in users' ads perception.

We showed that for each of the considered corpus, considering the interaction profile provides an extended comprehension of choice adoption mechanisms and retrieves several state-of-the-art conclusions. The proposed graphical visualization also provides an intuitive view of how the interaction occurs between entities and the associated trends, hence supporting its relevance as a new tool for researchers in a broad meaning.

## 7 Conclusion

We introduced an efficient convex model to investigate the way interactions happen within various datasets. Interactions modeling has been little explored in data science, despite recent clues pointing to their importance in modeling real-world processes [23, 12]. Unlike previous models, our method accounts for both the interaction effects and their influence over time (the interaction profile). We showed that this improvement leads to better results on synthetic and various real-world datasets that can be used in different research fields, such as recommender systems, spreading processes, and human choice behavior. We also discussed the difficulty of observing significant interaction profiles due to the data-gathering process's inherent noise and solved the problem by introducing a time-independent kernel. We finally proposed a way to easily explore the results yielded by our model, allowing one to read the interaction profiles of any couple of entities quickly.

In future works, we will explore the way interactions vary over time and work on identifying recurrent patterns in interaction profiles. It would be the next step for an extended understanding of the role and nature of interacting processes in real-world applications.

## References

1. Bereby-Meyer, Y., Roth, A.E.: The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. American Economic Review **96**(4), 1029–1042 (2006)
2. Beutel, A., Prakash, B.A., Rosenfeld, R., Faloutsos, C.: Interacting viruses in networks: Can both survive? SIGKDD p. 426–434 (2012)
3. Cao, J., Sun, W.: Sequential choice bandits: learning with marketing fatigue. AAAI-19 (2019)

4. Cobo-López, S., Godoy-Lorite, A., Duch, J.: Optimal prediction of decisions and model selection in social dilemmas using block models. EPJ Data Sci **7(48)** (2018). https://doi.org/10.1140/epjds/s13688-018-0175-3
5. Du, N., Song, L., Smola, A., Yuan, M.: Learning networks of heterogeneous influence. NIPS (2012)
6. Gomez-Rodriguez, M., Balduzzi, D., Schölkopf, B.: Uncovering the temporal dynamics of diffusion networks. In: ICML. p. 561–568 (2011)
7. Gomez-Rodriguez, M., Leskovec, J., Schoelkopf, B.: Structure and dynamics of information pathways in online media. WSDM (2013)
8. Gomez-Rodriguez, M., Leskovec, J., Schölkopf, B.: Modeling information propagation with survival theory. In: ICML. vol. 28, p. III–666–III–674 (2013)
9. Hawkes, A.: Point spectra of some mutually exciting point processes. JRSS. Series B **33** (07 1971). https://doi.org/10.1111/j.2517-6161.1971.tb01530.x
10. Hodas, N.O., Lerman, K.: The simple rules of social contagion. Scientific Reports **4**(4343) (2014)
11. L., W., A., F., al.: Competition among memes in a world with limited attention. Nature Sci Rep 2 **335** (2012). https://doi.org/10.1038/srep00335
12. Myers, S., Leskovec, J.: Clash of the contagions: Cooperation and competition in information diffusion. IEEE pp. 539–548 (2012)
13. Nay, J.J., Vorobeychik, Y.: Predicting human cooperation. PLoS One **11**(5) (2016)
14. Nielsen, F.: On a generalization of the jensen–shannon divergence and the jensen–shannon centroid. Entropy **22**, 221 (02 2020). https://doi.org/10.3390/e22020221
15. Poux-Médard, G., Pastor-Satorras, R., Castellano, C.: Influential spreaders for recurrent epidemics on networks. Phys. Rev. Research **2**, 023332 (Jun 2020). https://doi.org/10.1103/PhysRevResearch.2.023332
16. Poux-Médard, G., Velcin, J., Loudcher, S.: Interactions in information spread: quantification and interpretation using stochastic block models. RecSys (2021)
17. Prakash, B., Beutel, A., Rosenfeld, R., Faloutsos, C.: Winner takes all: Competing viruses or ideas on fair-play networks. WWW (04 2012). https://doi.org/10.1145/2187836.2187975
18. Rodriguez, M., Gummadi, K.P., Schölkopf, B.: Quantifying information overload in social media and its impact on social contagions. ICWSM (03 2014)
19. Senanayake, R., O'Callaghan, S., Ramos, F.: Predicting spatio–temporal propagation of seasonal influenza using variational gaussian process regression. AAAI p. 3901–3907 (2016)
20. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**, 1146–1151 (03 2018). https://doi.org/10.1126/science.aap9559
21. Yu, L., Cui, P., Song, C., Zhang, T., Yang, S.: A temporally heterogeneous survival framework with application to social behavior dynamics. KDD'17 p. 1295–1304 (2017). https://doi.org/10.1145/3097983.3098189
22. Z., Z., C., G., Y, Z.: Cooperation and competition among information on social networks. Nature Sci Rep **3103**, 12160 (07 2020). https://doi.org/10.1038/s41598-020-69098-5
23. Zarezade, A., Khodadadi, A., Farajtabar, M., Rabiee, H.R., Zha, H.: Correlated cascades: Compete or cooperate. AAAI (2017)