



Interactions in Information Spread

Gaël Poux-Médard

► To cite this version:

Gaël Poux-Médard. Interactions in Information Spread. WWW '22: The ACM Web Conference 2022, 2022, Virtual Event, France. pp.313-317, 10.1145/3487553.3524190 . hal-03778886

HAL Id: hal-03778886

<https://hal.science/hal-03778886>

Submitted on 16 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interactions in Information Spread

Gaël Poux-Médard

Under the supervision of Julien Velcin and Sabine Loudcher

gael.poux-medard@univ-lyon2.fr

Université de Lyon, Lyon 2, ERIC UR 3083

Bron, France

ABSTRACT

Large quantities of data flow on the internet. When a user decides to help the spread of a piece of information (by retweeting, liking, posting content), most research works assumes she does so according to information's content, publication date, the user's position in the network, the platform used, etc. However, there is another aspect that has received little attention in the literature: the information interaction. The idea is that a user's choice is partly conditioned by the previous pieces of information she has been exposed to. In this document, we review the works done on interaction modeling and underline several aspects of interactions that complicate their study. Then, we present an approach seemingly fit to answer those challenges and detail a dedicated interaction model based on it. We show our approach fits the problem better than existing methods, and present leads for future works. Throughout the text, we show that taking interactions into account improves our comprehension of information interaction processes in real-world datasets, and argue that this aspect of information spread is should not be neglected when modeling spreading processes.

CCS CONCEPTS

• General and reference → Surveys and overviews; • Information systems → Clustering; Social recommendation.

KEYWORDS

Information, Interaction, Diffusion, MMSBM, Dirichlet-Hawkes Process

ACM Reference Format:

Gaël Poux-Médard, Under the supervision of Julien Velcin and Sabine Loudcher. 2022. Interactions in Information Spread. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3487553.3524190>

1 MOTIVATION

As of today, every minute, approximately 400h of Youtube content, 350.000 tweets and 500.000 Facebook comments are uploaded to the internet. Such amounts of data do not appear randomly. They are generated by users given a context: publication platform, user

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9130-6/22/04...\$15.00

<https://doi.org/10.1145/3487553.3524190>

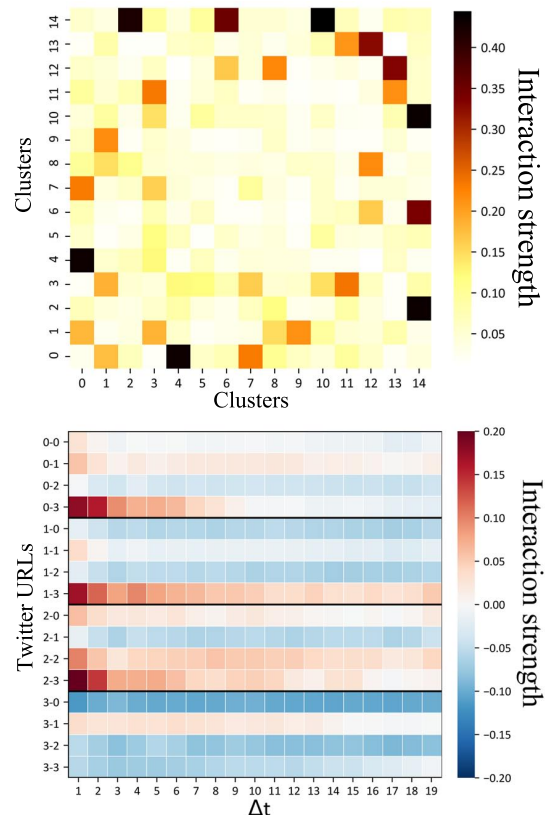


Figure 1: State of the art results on information interaction — (top) Interactions take place between very specific pairs of clusters; interactions are sparse [15]. (bottom) Interactions between specific pairs of entities tend to follow an exponential decay in time; interactions do not last [14]

interests, date of publication, position of the user in a communication network, etc. Many of these aspects have been widely explored in the literature; it has been shown that users interests can be efficiently modeled [6, 11], dynamics of publication can be inferred in efficient ways [5], and methods have been proposed to model the underlying diffusion networks of information diffusion [7, 19]. However, an open problem has been independently raised on several occasions, and received little interest over the last decade: how to efficiently model interactions between the spreading entities?

Intuitively, we say there is an interaction anytime the combination of entities yields more than their independent sum. Imagine answering a simple question, whose interacting entities are unique

words: to the question "Capital France?", the answer is obvious. However, to the bit "Capital?", answers can be very broad; same for the bit "France?". The interaction between those words narrowed the field of possible answers; they are interacting. This idea can easily be generalized to retweets (does a user retweet A only because of A, or also because of tweet B she has seen before?), online buying (does the user buy A because she likes it, or because she also bought B before?), and in general to online publication of documents (what are the bits of information making a user publish this post in particular on this particular platform?).

The understanding of interactions in information spread should provide a more exhaustive description of the publication process on the internet, as well as provide leads on how users react in a given context. At a time where three quarters of the world population has an online identity¹, the understanding of these mechanisms could find applications anytime an internet user has to make a choice: click, buy, like, comment, etc. One could think of nudging users towards healthier internet behaviours on forum or better consumption habits on online retail websites for instance.

The goal of this document is to briefly review the challenges that arise when it comes to modeling interactions in information spread –interactions are sparse and short. We then present a candidate methodology to deal with them, based on both Bayesian clustering and temporal point-processes. We show this method has flaws and propose a way to correct them. Our method allows to get better results on several datasets. Finally, we present leads to extend the existing framework and discuss research perspectives on interactions modeling.

2 PROBLEM

2.1 Interactions are sparse

In a first attempt to explicitly model interactions in a real-world dataset, the authors in [12] proposed to apply a classical block model to Twitter data. The goal is to predict the probability that a link will be retweeted, given a history of links that appeared before it. The method makes a central assumption: the interaction between tweets is a small modulation of a base retweet probability, which is precomputed as the total frequency of this tweet being retweeted. On this basis, the authors show that interaction modeling indeed helps to improve the predictive performances of the model. The analysis of the inferred interaction terms has led them to the following conclusions: most interactions are small, and interactions are slightly negative overall (meaning the lower the base probability of a retweet). By running their model for different sizes of prior exposures (or history), they also conclude that users rely mostly on a short-term memory, which correlates previous findings on users modeling [9, 13].

However, more recent works proved the core-hypothesis of this modeling false [15]. By precomputing the base probability of a retweet as the total number of times a tweet has been retweeted divided by the number of times it appeared, the authors in [12] already account for interactions. Imagine that interactions lead to a shift of Δp on the true base probability of a retweet p , and that this interaction happens in a fraction f of all observations of a given

tweet being retweeted. The base probability as defined in [12] then equals $p(1 - f) + (p + \Delta p)f = p + f\Delta p$, which is likely not to equal the actual probability of a retweet in the absence of interaction p . Therefore, defining interaction according to p is wrong. The base probability for a retweet needs to be inferred by the model at the same time as the contribution of interactions to be properly defined.

The authors in [15] then propose a mixed membership stochastic block model (MMSBM) in order to jointly infer the base probability of an entity spreading, as well as the interaction terms that modulate it. The idea is to group entities into clusters according to the way those clusters interact together. The authors in [15] study the importance of interactions in 4 different datasets. The conclusions are as follow: **modeling interactions improve predictive results, most interactions are small, and significant interactions are sparse**, meaning that they take place only between rare specific pairs of clusters. This result for the Twitter dataset is reported on Fig. 1-top.

2.2 Interactions are short

An important aspect of interactions that [15] does not consider is time. The position of entities in the considered history is not considered. In [12], the order of appearance is taken into account, but each temporal slice is modeled as independent from the others.

In [14], the authors propose to study the evolution of the interaction term in continuous time, using a convex model based on survival theory. This framework allows to infer both a time-independent probability of a retweet *and* a linear combination of time-dependent kernel functions that account for temporal evolution of the interaction terms. The data considered is similar as in [12, 15] as it consists of pairs of entities (typically tweets) associated to a given action (typically retweet), separated by a time Δt in a user's history. Note that this model does not have any clustering component; the interaction is studied only between the most interacting pairs of entities for every of the 3 considered datasets.

The results of [14] are as follow: **interactions magnitudes follow an exponential decay in time, and therefore do not last in time**. Besides, the model recovers previous findings on interactions: interactions are sparse, and their maximum magnitude is small. This result is reported Fig. 1-bottom.

3 STATE OF THE ART

3.1 Historical approach to dynamic clustering

We reviewed the main challenges associated with interactions modeling. First, interactions take place between specific pairs of clusters, and even more specific pairs of entities; to be able to spot them using non-convex models requires a **clustering aspect**. Second, interactions quickly fade in time, meaning that the information clusters must also take into account a **temporal aspect**.

The use of temporal dimension in documents clustering has been studied several occasions; a notable spike of interest happened in 2006. Many authors tackled the problem of inferring time-dependent clusters using models based on LDA [4, 23]. The idea is to sample the data provided to each model with a temporal function; the data can be selected using a temporal sliding window, or a temporal data sampling function. Time is not explicitly modeled. Most of these models are parametric, meaning the number of clusters

¹https://en.wikipedia.org/wiki/Global_Internet_usage

is fixed at the beginning of the algorithm. In 2008, A. Ahmed & al proposed the Recurrent Chinese Restaurant Process (RCRP) as an answer to this problem [1]. Instead of considering a fixed-size dataset, this model can handle a stream of documents arriving in chronological order, and the number of clusters is automatically updated. In this model, time is still not modeled explicitly (data is split into episodes to capture the temporal aspect of cluster formation). A later version of the model from 2010, the Distance-Dependent Chinese Restaurant Process (DD-CRP), tries to alleviate this approximation by replacing fixed-time episodes with a continuous-time sampling function [3]. However, the model is still not designed to explicitly model temporal information in a continuous-time setting. The slicing of the considered data introduces a consequent bias in the modeling.

3.2 Dirichlet-Hawkes process (DHP)

In 2015, N. Du & al [5] answered this problem by developing the Dirichlet-Hawkes Bayesian prior, that can be used to model the appearance of events in a continuous-time setting. The key idea is to replace the counts in a Dirichlet process prior by the intensity function of a Hawkes process.

Explicitly, given a standard Bayesian textual clustering model:

$$\underbrace{P(\text{cluster}|\text{text}, \text{time})}_{\text{Posterior probability}} \propto \underbrace{P(\text{text}|\text{cluster})}_{\text{Textual likelihood}} \times \underbrace{P(\text{cluster}|\text{time})}_{\text{Temporal prior}} \quad (1)$$

the authors propose the following form for the temporal prior:

$$P(C_i = c|t_i, \lambda_0, \mathcal{H}_{<t_i,c}) = \begin{cases} \frac{\lambda_c(t_i)}{\lambda_0 + \sum_{c'} \lambda_{c'}(t_i)} & \text{if } c \leq C \\ \frac{\lambda_0}{\lambda_0 + \sum_{c'} \lambda_{c'}(t_i)} & \text{if } c = C+1 \end{cases} \quad (2)$$

where C_i is the cluster allocation among C clusters of the i^{th} document appearing at time t_i , c a random variable, $\mathcal{H}_{<t_i,c}$ the date of all documents within cluster c published before t_i , $\lambda_c(t)$ the intensity of cluster c at time t , and λ_0 a parameter.

The functions $\lambda_c(t)$ need to be inferred as the intensities of a Hawkes process [18], and represent the instantaneous probability that a document appears at time t . Note that in Eq.2, the prior probability that a new document belongs to a given cluster depends on the intensity of this cluster at a time t : **clusters are self-stimulated**. In the problematic of information interaction, it means that pieces of information can only interact with other pieces of information from the same cluster. An example of a situation where a new observation gets allocated to any of the existing clusters is illustrated Fig. 2.

The resulting Dirichlet-Hawkes process (DHP) is then used as a prior for clustering documents appearing in a continuous-time stream. The inference is realized with a Sequential Monte-Carlo (SCM) algorithm. Following DHP, two articles have been published extending the idea: the Hierarchical Dirichlet Hawkes process (HDHP) [11] in 2016 and Indian Buffet Hawkes process in 2018 [20]. Recently, [10] underlined that this formulation suffers from a vanishing prior problem ($\lambda_c(t)$ can go to 0), and proposes a small procedure in order to avoid it, without modifying the core idea of the prior.

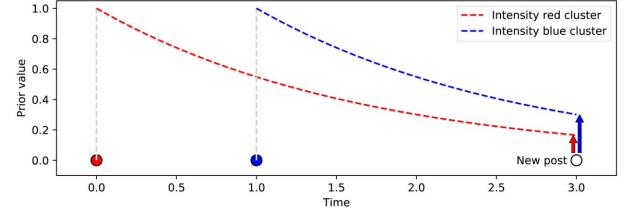


Figure 2: Illustration of the Dirichlet-Hawkes prior — Each cluster is associated to a temporal intensity function. The prior probability (Eq.2) of a new observation belonging to a cluster is linearly proportional to this cluster’s intensity.

4 PROPOSED APPROACH

4.1 Improvements over DHP

A common feature of all the models we mentioned built on DHP is that they use a non-parametric Dirichlet process (DP) prior or variations built on it, such as DHP and HDHP. Yet, on several occasions, it has been pointed out that there are no specific reasons to use this process in particular and that alternative forms might work better depending on the dataset [21, 24]. In [24], the author relaxes several properties associated with DP and shows that alternative priors are an equally valid choice in Bayesian modeling. In [21], the authors derive the Uniform process (UP) and show that it performs better on a document clustering task. In [16], the authors generalize UP and DP within a more general framework, the Powered Dirichlet process (PDP), and show it performs better than DP on several datasets.

Moreover, it has recently been highlighted that DHP does not work well when the textual information within documents conveys little information, that is when the text is short [25] or when vocabularies overlap significantly. To answer this problem, the authors develop an approach based on Dirichlet process mixtures, which is not designed for continuous-time document streams – the temporal aspect comes from a sampling function as in [1, 3]. There are other limiting cases for DHP, for instance when temporal information is conveys little information (few observations, entangled dynamics) or when documents within textual clusters do not follow a unique temporal dynamic. To overcome those limitations, we develop the Powered Dirichlet-Hawkes process in the next section.

4.2 The model

Taking back Eq.2, we note that the linear dependence of the prior on $\lambda_c(t)$ is justified only by analogy with the standard Dirichlet process. At this stage, we believe that a small modification of Eq.2 can lead to great variations in the comprehension of a dataset. In particular, making the prior more or less dependent on the temporal dimension (in the same way that [17, 21] makes the DP more or less dependent on the “rich-get-richer” hypothesis) could lead to clusters that are more text or time orientated. By replacing the standard Dirichlet process in [5] by the Powered Dirichlet process from [16], we derive the Powered Dirichlet-Hawkes process[17]:

$$P(C_i = c|r, t_i, \lambda_0, \mathcal{H}_{<t_i,c}) = \begin{cases} \frac{\lambda_c^r(t_i)}{\lambda_0 + \sum_{c'} \lambda_{c'}^r(t_i)} & \text{if } c \leq C \\ \frac{\lambda_0}{\lambda_0 + \sum_{c'} \lambda_{c'}^r(t_i)} & \text{if } c = C+1 \end{cases} \quad (3)$$

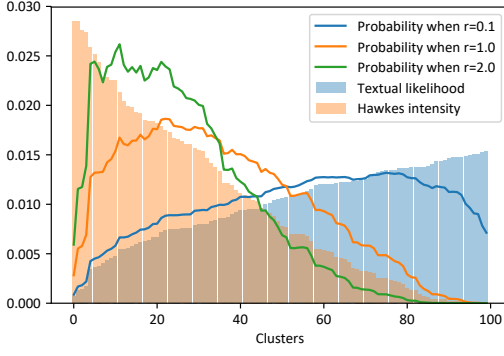


Figure 3: Effect of the parameter r in PDHP on the posterior probability (Eq.1 using the prior in Eq.3) for a new observation to belong to each cluster.

We show in Fig. 3 what this change implies for a standard Bayesian textual clustering model. The value of r tunes how much we rely on either the textual or the temporal information every time the algorithm chooses a cluster for a new document.

The algorithm used for inference is a sequential Monte-Carlo; it is the same as described in [5, 11, 17]. It is especially fit for modeling data streams, as it considers data sequentially. The underlying idea is that for each new observation, a number of *particles* N_{part} will make a cluster allocation hypothesis according to Eq.1, update the cluster's intensity function $\lambda_c(t)$, and pass on to the next document. When the likelihood of a particle L_p goes under a given threshold ω , the particle is discarded and replaced by a more likely one. In our case, we consider 8 particles and a threshold of $\omega = \frac{\sum_p L_p}{2N_{part}}$.

5 METHODOLOGY AND RESULTS

5.1 Data generation

We simulate cases where only two clusters are considered. Each cluster has its own vocabulary distribution over 1 000 words and its associated intensity function $\lambda_c(t)$. We first simulate one independent Hawkes process per cluster using the Tick Python library [2]. The processes are stopped at time $t = 1500$, which makes a rough average of 7 000 events per run. Then we associate each simulated observation with a sample of 20 words drawn from the corresponding cluster's word distribution.

We evaluate our results according to the Normalized Mutual Information score (NMI), which is a standard metric to compare the results of different clustering algorithms when the number of clusters can vary.

5.2 Intra-cluster interactions...

5.2.1 ...with unreliable information. A first lead to explore how information interact in time is to consider the case where clusters are self-stimulated. However, [25] showed that the approach in [5] is not fit for cases where textual information is weak or unreliable (e.g. Twitter) or when temporal dynamics are highly entangled or when textual information is sufficient (e.g. long documents or real-world process at large scales). In order to confirm that our modeling

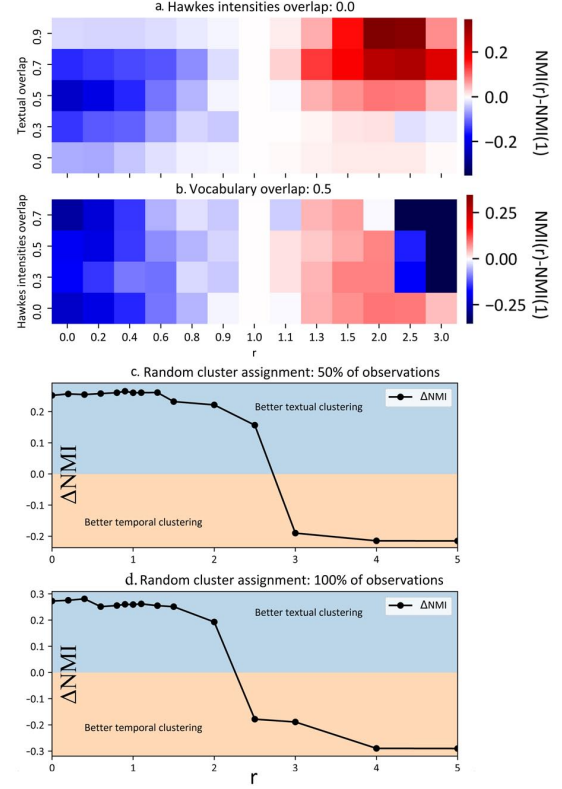


Figure 4: Results of PDHP when textual and temporal information are not fully reliable (a,b) and when textual content is decorrelated from temporal dynamics (c,d).

in [17] answers those limitations, we run experiments for various values of informativeness of either textual or temporal content.

To do so, we will make the clusters' vocabulary distributions and temporal intensity functions overlap. Given two distributions $\mathcal{A}(x)$ and $\mathcal{B}(x)$, the overlap is defined as $\frac{\int_x \min(\mathcal{A}(x), \mathcal{B}(x)) dx}{\int_x \mathcal{A}(x) + \mathcal{B}(x) dx}$.

In the case of vocabulary overlap, we shift the words distribution so that they have a certain fraction of shared words. In the case of temporal overlap, we first simulate each cluster's Hawkes process, and compute the intensity function of the whole resulting dataset. Then, we shift each process realization on the temporal axis so that the overlap reaches the desired value. We evaluate the models according to the documents clusters allocations compared to the true clusters from the data generation process. We report the average improvement for every of the 20 datasets considered in each overlaps configuration, for various values of r .

Our results are shown in Figs. 4a-b. PDHP allows better performances comparing to DHP ($r=1$) and UP ($r=0$) when textual information is unreliable (high textual overlap), up to a gain of +0.3 NMI w.r.t. DHP (Fig. 4a). In more realistic situations (average textual and temporal overlaps), PDHP allows better results (up to +0.2 NMI, Fig. 4b). Our method is therefore able to accurately uncover which cluster triggered a new document, that is to which interaction chain

this document belongs, in the case of weak temporal and textual information.

5.2.2 ... with decorrelated content and dynamics. Another limiting situation of [5] is when textual documents about a same topic do not follow similar dynamics. Picture a famous and an infamous newspapers publishing the exact same article; these articles would not follow the same dynamics despite an identical textual content.

In order to explore the case where identical textual contents do not follow identical temporal dynamics, we first generate the whole dataset with null temporal and textual overlaps. Then, we randomly select a given fraction of the generated observations, and resample their associated vocabulary according to a randomly selected cluster. We have two different possible clustering: textual (how well the model recovers the vocabulary a document's content has been drawn) and temporal (which intensity function made this document appear where it is). In this case, we report the difference between the textual clustering NMI and the temporal clustering NMI: $\Delta NMI = NMI_{text} - NMI_{time}$.

Our results are shown Figs. 4c-d. When textual content is decorrelated from publication dynamics (which is more in line with what happens in real-world processes), tuning r allows to consistently recover one clustering or the other. That means we can successfully recover clusters based on either textual similarity or on the temporal interaction between documents, or a mixture of both.

5.3 Ongoing: Inter-cluster interactions

As for now, we presented introductory results about self-interacting clusters. However, in order to get the full picture, we cannot only consider the case where only one given cluster interacts with incoming documents. Instead, we need to consider the interaction between all existing clusters. This embeds naturally in the proposed approach [5, 17] by considering a multivariate Hawkes process instead of a uni-variate one [18]. In implies redefining the numerator in Eq.3 from $\lambda_c(t)$ to $\sum_{c'} \lambda_{c' \rightarrow c}(t)$, where $\lambda_{c' \rightarrow c}(t)$ is the temporal influence of cluster c' on cluster c at a time t . The challenge here is to preserve the efficiency of the proposed algorithm while the model's complexity increases in a polynomial fashion with the number of inferred clusters.

5.4 Future: User-level inter-cluster interactions

As for future works, we believe it is possible to go further by combining temporal point processes to Dirichlet processes. While [5] opened the door to this union, it only explored a small part or possible applications. In particular, we think of the works of Gomez-Rodriguez on underlying network inference [7, 19] and extensions [6, 22]. In [8], he showed that all these models can be formulated in terms of a counting temporal point process, which naturally allows to substitute any such modeling to the Hawkes process in Eq.2 and Eq.3. In this case, we would be able to study the interaction between pieces of information at the user-level, providing a deeper understanding of information interaction in real-world processes.

6 CONCLUSIONS

In this document, we first motivated the need to model the interaction between pieces of information in order to get a better

understanding of real-world spreading processes. We then highlighted the main challenges that arise in this perspective using recent published works. We proposed to address the problem by considering a recent approach that seems to fit to answer the challenges of interaction modeling, despite having its own flaws. We proposed a simple method to overcome these, and conducted systematic experiments to demonstrate how our methodology works at doing so. Finally, we proposed leads for further improving our understanding of real-world information interaction modeling.

REFERENCES

- [1] Amr Ahmed and E. Xing. 2008. Dynamic Non-Parametric Mixture Models and the Recurrent Chinese Restaurant Process: with Applications to Evolutionary Clustering. In *SIAM International Conference on Data Mining*. 219–230.
- [2] Emmanuel Bacry, Martin Bompairé, Philip Deegan, Stéphane Gaïffas, and Søren V. Poulsen. 2017. Tick: A Python Library for Statistical Learning, with an Emphasis on Hawkes Processes and Time-Dependent Models. *J. Mach. Learn. Res.* 18, 1 (2017), 5 pages.
- [3] David M. Blei and Peter Frazier. 2010. Distance Dependent Chinese Restaurant Processes. In *ICML '10 (ICML '10)*. 87–94.
- [4] David M. Blei and John D. Lafferty. 2006. Dynamic Topic Models. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*. 113–120.
- [5] Nan Du, Mehrdad Farajtabar, Amr Ahmed, Alex Smola, and L. Song. 2015. Dirichlet-Hawkes Processes with Applications to Clustering Continuous-Time Document Streams. *21th ACM SIGKDD (2015)*.
- [6] Nan Du, Le Song, Hyenkyun Woo, and Hongyuan Zha. 2013. Uncover Topic-Sensitive Information Diffusion Networks. In *AISTATS (JMLR, Vol. 31)*. 229–237.
- [7] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Schölkopf. 2011. Uncovering the Temporal Dynamics of Diffusion Networks. In *ICML*. 561–568.
- [8] Manuel Gomez-Rodriguez, Jure Leskovec, and Bernhard Schölkopf. 2013. Modeling Information Propagation with Survival Theory. In *ICML*. Vol. 28. 9 pages.
- [9] Mario Gutiérrez-Roig, Carlota Segura, Jordi Dutch, and Josep Perello. 2016. Market Imitation and Win-Stay Lose-Shift Strategies Emerge as Unintended Patterns in Market Direction Guesses. *PLOS One* (2016).
- [10] J. Kapoor, A. Vergari, I. Valera, and M. Gomez-Rodriguez. 2018. Bayesian non-parametric Hawkes processes. In *Bayesian Nonparametrics workshop NIPS'18*.
- [11] Charalampos Mavroforakis, Isabel Valera, and Manuel Gomez-Rodriguez. 2017. Modeling the Dynamics of Learning Activity on the Web. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. 1421–1430.
- [12] Seth A. Myers and J. Leskovec. 2012. Clash of the Contagions: Cooperation and Competition in Information Diffusion. *2012 IEEE 12th International Conference on Data Mining (2012)*, 539–548.
- [13] Gaël Poux-Médard, Sergio Cobo-Lopez, Jordi Duch, Roger Guimerà, and Marta Sales-Pardo. 2021. Complex decision-making strategies in a stock market experiment explained as the combination of few simple strategies. *EPJ Data Science* 10 (2021), Issue 26.
- [14] Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. 2021. Information Interaction Profile of Choice Adoption. *ECML-PKDD '21 (2021)*.
- [15] Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. 2021. Interactions in information spread: quantification and interpretation using stochastic block models. *RecSys '21 (2021)*.
- [16] Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. 2021. Powered Dirichlet Process for Controlling the Importance of "Rich-Get-Richer" Prior Assumptions in Bayesian Clustering. *ArXiv* (2021). arXiv:2104.12485 [cs.LG]
- [17] Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. 2021. Powered Hawkes-Dirichlet Process: Challenging Textual Clustering using a Flexible Temporal Prior. *ICDM (2021)*.
- [18] Marian-Andrei Rizoio, Young Lee, Swapnil Mishra, and Lexing Xie. 2018. A Tutorial on Hawkes Processes for Events in Social Media. *FMR* (2018), 191–218.
- [19] Manuel Gomez Rodriguez, Jure Leskovec, and Bernhard Schoelkopf. 2013. Structure and Dynamics of Information Pathways in Online Media. *WSDM* (2013).
- [20] X. Tan, Vinayak A. Rao, and Jennifer Neville. 2018. The Indian Buffet Hawkes Process to Model Evolving Latent Influences. In *UAI*.
- [21] Hanna Wallach, Shane Jensen, Lee Dicker, and Katherine Heller. 2010. An alternative prior process for nonparametric Bayesian clustering. In *AIS '13*. 892–899.
- [22] Liaoruo Wang, Stefano Ermon, and John E. Hopcroft. 2012. Feature-Enhanced Probabilistic Models for Diffusion Network Inference. In *Machine Learning and Knowledge Discovery in Databases*. 499–514.
- [23] Xuerui Wang and Andrew McCallum. 2006. Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends. *KDD '06*, 424–433.
- [24] Max Welling. 2006. Flexible priors for infinite mixture models. In *Workshop on learning with non-parametric Bayesian methods*.
- [25] Jianhua Yin, Daren Chao, Zhongkun Liu, Wei Zhang, Xiaohui Yu, and Jianyong Wang. 2018. Model-Based Clustering of Short Text Streams (*KDD '18*). 2634–2642.