



Properties of Reddit News Topical Interactions

Gaël Poux-Médard^(✉), Julien Velcin, and Sabine Loudcher

Université de Lyon, Lyon 2, ERIC UR 3083, Bron 69676, France

gael.poux-medard@univ-lyon2.fr

<https://gaelpouxmedard.github.io>

Abstract. Most models of information diffusion online rely on the assumption that pieces of information spread independently from each other. However, several works pointed out the necessity of investigating the role of interactions in real-world processes, and highlighted possible difficulties in doing so: interactions are sparse and brief. As an answer, recent advances developed models to account for interactions in underlying publication dynamics. In this article, we propose to extend and apply one such model to determine whether interactions between news headlines on Reddit play a significant role in their underlying publication mechanisms. After conducting an in-depth case study on 100,000 news headline from 2019, we retrieve state-of-the-art conclusions about interactions and conclude that they play a minor role in this dataset.

Keywords: Dirichlet Hawkes process · Topic modeling · Interaction · Information spread

1 Introduction

As the volume of data available on the Internet keeps on growing exponentially, so does the need for efficient data-processing tools and methods. In particular, the user-generated content produced on online social platforms provides a detailed snapshot of the world population's thoughts and interests. This kind of data can be used for many different applications (e.g. in marketing, opinion mining, fake news control, summary generation, etc.). However, we need a fine understanding of the underlying data-generation mechanisms at stake to refine the results of these possible applications.

Early works on information spread consider that a user publishing a piece of information (or meme) did so due to an earlier exposure to this same meme [11]. Users are represented as a network on which edges pieces of information flow independently from each other. This model has seen several refinements, that consider additional information about nodes and edges to model the way information spreads. Nodes participate the spread when exposed a certain number of times [15, 17], have a different influence on their neighbours depending on their position in the network [12], can publish exogenous memes (e.g. that were not

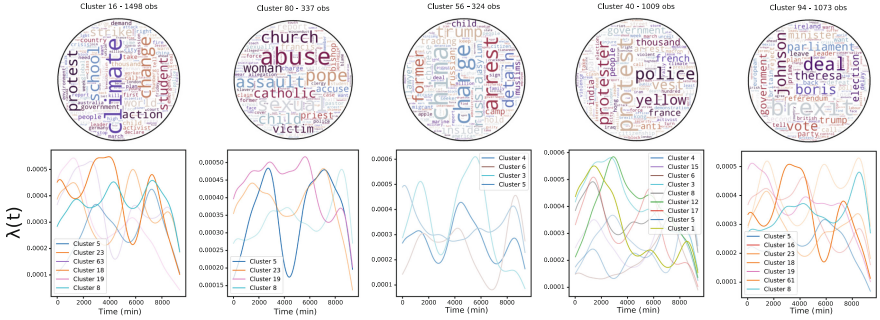


Fig. 1. An output of the proposed approach—(Top) A set of inferred topics along with the vocabulary of their documents. (Bottom) Instantaneous probability for one observation of a topic to trigger other observations in other topics.

flowing on the network beforehand) [10, 16], etc. Edges between nodes represent the likeliness of a meme flowing from one node to the other; later works consider edges whose intensity depends on the meme considered [1, 6], on the temporal distance to the last exposure [8, 9], etc.

However, a core assumption of most of these models and extensions is that memes spread independently from each other. It has been highlighted in some occasions that modeling the interaction between pieces of information might bring interesting insights in the way data gets generated online [3, 7, 15]. For instance, we expect memes about two politically opposed candidates to have some interaction (the publication of one would inhibit the publication chances of the other one); memes raising climate change awareness might be more likely to spread when coexisting with memes about ecologic disasters, etc.

In this work, we conduct an in-depth study of such interaction mechanisms on a large-scale Reddit dataset—gathered from 9 news subreddits over 2019. We propose to use a model that answers the challenges of interaction modeling raised in recent works on interactions—interactions between memes are likely sparse and brief. This model, the Multivariate Powered Dirichlet-Hawkes Process (MPDHP), groups textual memes into topics. A possible output is represented in Fig. 1. Topics interact in pairs to yield a probability for a new meme about a given topic to get published, in a continuous-time setting. The output of MPDHP is a time-dependent topical interaction network. Once this network has been inferred, we can analyse the properties of topical interactions. We conduct experiments using several iterations of MPDHP (accounting for different timescales and topical modeling), so that we get an exhaustive panorama of interactions at stake. Overall, our conclusions hint that interactions play a minor role on news subreddits; most information indeed spreads independently in this context. From a broader perspective, we introduce a methodology to assess the role of temporal topical interactions textual datasets.

2 Background

Modeling interactions In this section, we briefly review previous works on interaction modeling in spreading processes. To the best of our knowledge, the first work tackling the problem proposes a SIR-based model that comprises an interaction term β between two co-existing viruses—in their case, the adoption of either Firefox or Chrome web browser [3]. Interaction is modeled as a hyper-parameter, that has to be tuned manually to retrieve global shares of each virus over time. Building on this work, [15] later proposed to model interaction between memes at the agent level: given a user is exposed to tweet A at time t_A , what is the probability to retweet tweet B at a time t_B later. This model attempts to learn the interaction parameter β from the data instead of fine-tuning it. Their conclusion is that interactions can have a large overall effect of the data-generation processes (here retweeting mechanisms), but that most interactions have little influence. Later works tackling the problem from a similar perspective on several real-world datasets find similar results: significant interactions between clusters of memes are sparse [19,21]. This highlight the need to cluster memes so that it becomes possible to retrieve meaningful interaction terms.

Some other works tackled the problem of pair-interaction modeling from a temporal perspective. Simple considerations show that interactions cannot remain constant over time; a user that read a meme 5 min earlier and another meme 5 days earlier is likely to be much more influenced the first one than the second one. Interactions are likely to be brief. This temporal dependence is explicitly modeled in [18], where the interaction strength is shown to typically decay exponentially, which correlated the findings of [4], and of [15] to a certain extent. It highlights the need to consider time to relevantly model interactions. Thus the need for a model able to handle large piles of data, performing topic inference, allowing interactions between these topics, and accounting for time.

Dirichlet-Hawkes processes The Dirichlet-Hawkes process [5] seem to qualify for these requirements. In particular, it has been extended as the Powered Dirichlet-Hawkes process [20] to allows for extended modeling flexibility—to which extend should we favor the textual information over the temporal information. However, both these models do not account for pair-interaction between the inferred topics. Instead, topics can only trigger new observations from themselves. In this work, we consider a Multivariate extension of the Powered Dirichlet-Hawkes process (MPDHP). This extension boils down to substituting the Hawkes process described in [5,20] by a multivariate Hawkes process [13], so that topics can influence the probability of new observations to belong to either other topics.

The principle of the Dirichlet-Hawkes approaches relies on Bayesian inference. An online language model accounts for the textual content of documents (Fig. 1-top). This model, typically a Dirichlet-Multinomial bag of words as we will use later, expresses the likelihood for a new document to belong to any existing

topic. This language model is coupled to a Dirichlet-Hawkes prior, that can be expressed as a sequential process. The prior assigns to a new observation a prior probability to belong to either topic based on its temporal interaction with earlier publications (Fig. 1-middle). In the remaining of this section, we present the modified formulation of the base models [5, 20] so that it accounts for topical pair-interactions.

We first rewrite and detail the expression of the Dirichlet-Hawkes process as introduced in [5, 20]:

$$\begin{aligned}
 & P(C_i = c | n_i, N_{<i,c}, t_i, \lambda(t_i), \mathcal{H}, \theta_0, \lambda_0, r) \\
 & \propto \underbrace{P(n_i | C_i = c, N_{<i,c}, \theta_0)}_{\substack{\text{Textual likelihood} \\ \text{(Dirichlet-Multinomial)}}} \underbrace{P(C_i = c | t_i, \lambda(t_i), \mathcal{H}, \lambda_0, r)}_{\substack{\text{Temporal prior} \\ \text{(MPDHP)}}} \\
 & = \frac{\Gamma(N_c + \theta_0)}{\Gamma(N_c + n_i + \theta_0)} \prod_v \frac{\Gamma(N_{c,v} + n_{i,v} + \theta_{0,v})}{\Gamma(N_{c,v} + \theta_{0,v})} \begin{cases} \frac{\lambda_c^r(t_i)}{\lambda_0 + \sum_{c'} \lambda_{c'}^r(t_i)} \text{ if } c \leq K \\ \frac{\lambda_0}{\lambda_0 + \sum_{c'} \lambda_{c'}^r(t_i)} \text{ if } c = K + 1 \end{cases} \quad (1)
 \end{aligned}$$

where C_i represents the cluster chosen by the i th document, c is the random variable accounting for this allocation, n_i is a vector that whose v th entry represents the count of word v in document i , $N_{<i,c}$ is the vector whose entry v represents the total count of word v in all documents up to i that belong to cluster c , t_i the arrival time of document i , $\lambda(t_i)$ the vector of intensity functions at time t_i whose c th entry corresponds to cluster c , and \mathcal{H} the history of all previous documents that appeared before time t_i . The three last symbols are hyperparameters: θ_0 is the concentration parameter of the Dirichlet-Multinomial language model [5, 22], λ_0 is the concentration parameter of the MPDHP temporal prior [5], and r controls the extent to which we rely of either textual or temporal information [20]. Data is processed sequentially using a Sequential Monte Carlo (SMC) algorithm similar to [5, 14, 20]. For each new observation, we get from Eq. 1 the posterior probability that it belongs to either of K existing clusters, or to start a cluster of its own ($K + 1$). The SMC algorithm accounts for several allocation hypotheses at once, and discard the most unlikely ones every other iteration.

Now, the extension of the Dirichlet-Hawkes process to the multivariate case boils down to giving a new definition for the vector $\lambda(t)$. We express it as:

$$\lambda_c(t) = \sum_{t_i^{c'} < t} \alpha_{c,c'}^T \cdot \kappa(t - t_i^{c'}) \quad \text{where} \quad \kappa_l(\Delta t) = \frac{1}{\sqrt{2\pi\sigma_l^2}} e^{-\frac{(\Delta t - \mu_l)^2}{2\sigma_l^2}} \quad (2)$$

where $\alpha_{c,c'}$ is a vector of L parameters to infer, and $\kappa(t - t_i^{c'})$ is a vector of L given temporal kernel functions depending only on the time difference between two events. We consider $\kappa(\Delta t)$ to be a Gaussian RBF kernel with fixed mean μ and deviation σ , which allows us to model a range of different intensity functions. Each parameter $\alpha_{c,c',l}$ accounts for the influence of c' on c according to the l th

entry of the temporal kernel. The dot product of α with κ yields the intensity function vector λ , which represents the topical interactions’ temporal adjacency matrix (represented in an alternative way in Fig. 1-middle).

3 Experimental Setup

Dataset The dataset used in this study has been gathered from the Pushshift Reddit repository [2], which contains archives of the entirety of Reddit posts and comments up to June 2021. For each Reddit post, we can retrieve the subreddit it came from, the title of the publication, its publication date and its score (number of upvotes minus number of downvotes).

We choose to only consider popular English news subreddits. Namely, we select only posts from 2019 published on the following subreddits: inthenews, neutralnews, news, nottheonion, offbeat, open news, qualitynews, truenews, worldnews. This leaves us with a corpus of 867,328 headlines, which makes a total of 1,111,955 words drawn from a vocabulary of size 36,284.

Finally, we discard uninformative words and documents from the dataset. Explicitly, we remove the stopwords, punctuation signs, web addresses, words whose length is lesser than 4 characters, and words that appear less than 3 times in the whole dataset. Then, we remove publications that carry lesser textual or temporal information. Firstly, we choose not to consider the publications that have a popularity lesser than 20—meaning that they received less than 20 positive votes more than negative votes. We make this choice so that we consider publications that are visible enough to have any influence on the data generation

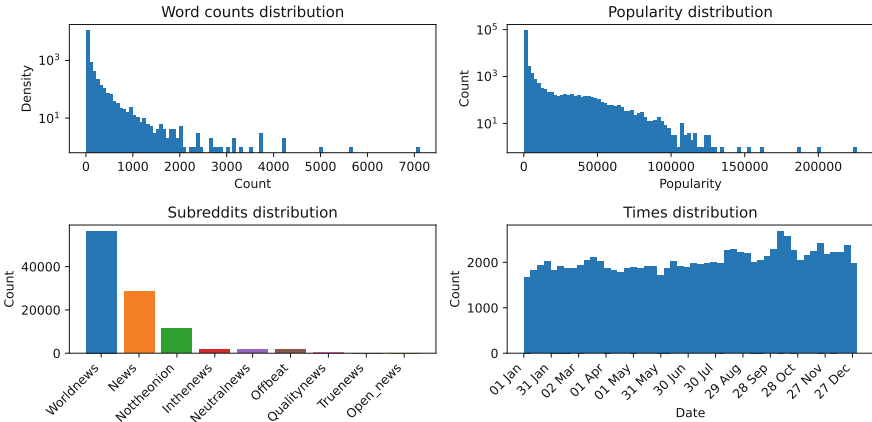


Fig. 2. Characteristics of the News dataset—For $\sim 100,000$ headlines and $\sim 13,000$ different words: (Top-Left) Distribution of the words count. (Top-Right) Distribution of headlines popularity. (Bottom-Left) Distribution of headlines over subreddits. (Bottom-Right) Distribution of headlines over time.

process. Secondly, we remove publications that comprise less than 3 words. The semantic information so-carried is expected to be poor and is not considered in our analysis.

After curating the dataset in the way described above, we are left with 102,045 news headlines (one-eighth of the original data), which makes a total of 875,334 tokens (named entities, verbs, numbers, etc.) drawn from a vocabulary of size 13,241 (one-third of the original vocabulary). The characteristics of this dataset are shown in Fig. 2.

Temporal kernel We run our experiments using three different RBF kernels, which account for publication dynamics at three different timescales: minute, hour, and day. The “**Minute**” RBF kernel is made of Gaussian functions centered at the following times: [0, 10, 20, 30, 40, 05, 60, 70, 80] minutes; each entry shares a same standard deviation σ of 5 minutes, and $\lambda_0 = 0.01$. The “**Hour**” RBF kernel has Gaussians centered around [0, 2, 4, 6, 8] hours, with a standard deviation σ of 1 hour, and $\lambda_0 = 0.001$. The “**Day**” RBF kernel is centered around [0, 1, 2, 3, 4, 5, 6] days, with a standard deviation σ of 0.5 days, and $\lambda_0 = 0.0001$. For each of these kernels, we set the concentration parameter λ_0 so that it reaches roughly the value of one Gaussian function evaluated at 2σ . It means that an event which is 2σ away from the center of the Gaussian kernel of a single observation has 50% chances of getting associated with this Gaussian kernel entry, and 50% chances of opening a new cluster.

Hyper-parameters We consider two values for the concentration parameter of the language model: $\theta_0 = 0.001$ and $\theta_0 = 0.01$. The choice of this range is standard in the literature [5] and supported by our own observations. A larger value of θ_0 makes the inferred clusters cover a broader range of document types, whereas a small value makes the inferred clusters more specific to a topic.

The value of r is chosen to be either 0 (no use of the temporal information), 0.5, 1 and 1.5. The larger r , the more the inference relies of the temporal dynamics instead of the textual content of the documents.

The SMC algorithm described in [5, 20] is run using 8 particles and 100,000 samples used to estimate the matrix of parameters α .

4 Results

4.1 Overview of the Experiments

In Table 1, we represent the main characteristics of each run in terms of number of inferred clusters K , the average cluster population $\langle N \rangle$ (where $\langle \cdot \rangle$ denotes the average), the average normalized entropy of the vocabulary of the top 20 clusters $S_{text}^{(20)}$, the average normalized entropy of the subreddits partition of the most populated 20 clusters $S_{sub}^{(20)}$. The normalized entropy is bounded between 0 and 1. It is defined so that a low entropy $S_{text}^{(20)}$ (resp. $S_{sub}^{(20)}$) means that

Table 1. Results for each experiment—We characterize each run in terms of clusters (number, size, textual entropy, subreddits entropy). The standard deviation on the last digits is given between parenthesis.

$\kappa(t)$	θ_0	r	K	$\langle N \rangle$	$S_{text}^{(20)}$	$S_{sub}^{(20)}$
Minute	0.01	0.0	16150	6	0.744 (68)	0.400 (36)
		0.5	8498	12	0.796 (55)	0.441 (24)
		1.0	5069	20	0.790 (49)	0.476 (53)
		1.5	2730	37	0.808 (43)	0.490 (52)
	0.001	0.0	46304	2	0.475 (48)	0.239 (56)
		0.5	37277	3	0.485 (40)	0.256 (60)
		1.0	26858	4	0.501 (37)	0.266 (72)
		1.5	19275	5	0.506 (39)	0.280 (58)
Hour	0.01	0.0	3792	27	0.798 (52)	0.474 (106)
		0.5	1735	59	0.791 (45)	0.469 (77)
		1.0	825	124	0.803 (47)	0.484 (75)
		1.5	426	240	0.795 (37)	0.489 (69)
	0.001	0.0	18012	6	0.760 (86)	0.397 (63)
		0.5	11923	9	0.784 (72)	0.425 (38)
		1.0	4837	21	0.821 (50)	0.497 (30)
		1.5	2368	43	0.814 (42)	0.481 (91)
Day	0.01	0.0	609	168	0.713 (34)	0.413 (103)
		0.5	326	313	0.728 (33)	0.429 (98)
		1.0	172	593	0.743 (31)	0.461 (94)
		1.5	96	1063	0.755 (36)	0.464 (86)
	0.001	0.0	4349	23	0.705 (49)	0.396 (103)
		0.5	2654	38	0.721 (59)	0.404 (104)
		1.0	1399	73	0.734 (57)	0.431 (106)
		1.5	764	134	0.741 (54)	0.442 (98)

the top 20 clusters contain documents that are concentrated around a reduced set of words (resp. of subreddits); conversely, a large entropy means that these top 20 clusters do not account for documents concentrated around a specific vocabulary (resp. set of subreddits). We can make several observations from Table 1:

- The number of inferred clusters decreases with r , and their average population increases.
- The number of clusters grows large for the “Minute” kernel. This is because the short time range considered does not allow for clusters to last in time. A cluster that does not replicate within 1h30 is forgotten.

- We recover the fact that textual clusters have a lower entropy for small values of r [20]; this is because their creation is based more on textual coherence than on temporal coherence.
- The subreddit entropy seems to increase as r grows. A possible interpretation is that favouring the temporal information for cluster creation results in larger clusters (see $\langle N \rangle$). They would be too large to account for subreddit-specific dynamics. However, the entropy remains lower than the entropy of the distribution Fig. 2-top-left, equal to 0.51.

4.2 Quantifying Interactions

Effective interaction We introduce the parameters we are going to use in follow-up analyses. The output of MPDHP consists of a list of clusters comprising timestamped bags of words—news headlines. Between each pair of clusters, MPDHP inferred a temporal influence function $\lambda(t)$, that represents the probability for one cluster to trigger publications from another. Therefore, our model yields an adjacency matrix $A \in \mathbb{R}^{K \times K \times L}$, where K is the number of clusters and L the size of the RBF kernel $\kappa(t)$. One entry $a_{i,j,l}$ represents the strength of the influence of j in i due to the l th entry of $\kappa(t)$.

However, we must consider the effective number of interactions, that is to which extent the intensity function has effectively had a role in the inference. To do so, we simply consider a weight matrix $W \in \mathbb{R}^{K \times K \times L}$, whose entries $w_{i,j,l}$ are the average of the intensity of i above λ_0 due to j from the kernel entry l for all observations. Explicitly:

$$w_{i,j,l} = \frac{1}{|\mathcal{H}_i|} \sum_{t_i \in \mathcal{H}_i} \sum_{t_j < t_i} \max(a_{i,j,l} \kappa_l(t_i - t_j) - \lambda_0, 0) \quad (3)$$

The notations are the same as in Eqs. 1 and 2. Note that we retract λ_0 from the intensity term, because it is considered as a background probability for a publication to happen. Therefore, W can also be interpreted as the instantaneous increase in probability due to interactions.

Interactions strength In Table 2, we investigate the effective impact of interactions in the dataset. We consider the following metrics:

- $\langle A \rangle$: the average value of the whole adjacency matrix; to which extent topics interact with each other according to MPDHP.
- $\langle W \rangle$: the average value of the effective interactions; the extent to which the interactions (encoded in A) effectively happen in the dataset.
- $\langle A \rangle_W$: the average of the inferred interaction matrix A weighted by the effective interactions W . In this case, W can be interpreted as our confidence in the corresponding entries of A .
- $\frac{\langle W^{intra} \rangle}{\langle W^{extra} \rangle}$: ratio of the intra-cluster effective interactions with the extra-cluster effective interactions; how much clusters self-interact versus how much they interact with other ones.

Table 2. Interaction strength—Overall, interaction between clusters is weak. The standard deviation on the last digits are given between parenthesis. The large standard deviations suggest that there is a large variety of interacting behaviours. Interactions tend to happen within a cluster (self-interactions).

$\kappa(t)$	θ_0	r	$\langle A \rangle (10^{-3})$	$\langle W \rangle (10^{-5})$	$\langle A \rangle_W (10^{-3})$	$\frac{\langle W^{intra} \rangle}{\langle W^{extra} \rangle}$
Minute	0.01	0.5	49 (21)	342 (889)	66 (17)	1.8 (62)
		1.0	48 (20)	478 (1124)	60 (17)	1.4 (43)
		1.5	48 (20)	746 (1901)	60 (17)	1.0 (33)
	0.001	0.5	50 (22)	316 (882)	66 (17)	3.1 (138)
		1.0	50 (21)	279 (752)	67 (16)	2.6 (105)
		1.5	50 (22)	268 (665)	67 (16)	2.3 (84)
Hour	0.01	0.5	49 (18)	389 (843)	56 (17)	0.5 (13)
		1.0	49 (18)	478 (1187)	56 (17)	0.6 (15)
		1.5	48 (17)	471 (789)	52 (15)	0.7 (13)
	0.001	0.5	50 (21)	110 (398)	61 (17)	1.7 (67)
		1.0	50 (18)	133 (506)	57 (17)	1.4 (60)
		1.5	49 (17)	183 (554)	55 (17)	1.1 (37)
Day	0.01	0.5	49 (18)	41 (97)	55 (17)	1.2 (34)
		1.0	49 (19)	63 (131)	54 (17)	1.2 (31)
		1.5	49 (19)	91 (187)	53 (18)	1.2 (31)
	0.001	0.5	50 (20)	18 (90)	60 (19)	1.1 (59)
		1.0	50 (19)	23 (101)	58 (17)	1.0 (50)
		1.5	50 (19)	37 (111)	56 (18)	1.0 (36)

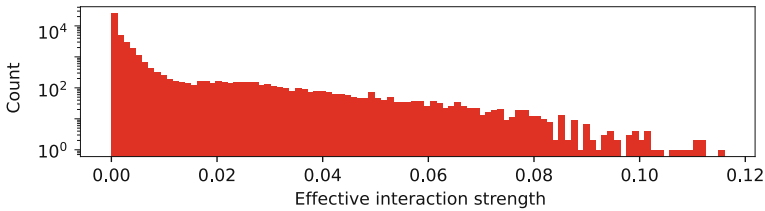


Fig. 3. Distribution of interaction strength—A large effective interaction strength means that the publication of a new document can be significantly explained by the presence of previously published news. We see that most interactions are weak.

The main conclusion of the results Table 2 is that most interactions are weak. The average value of A tells us that the average value of the inferred parameters is around 0.05, which is few given entries of A are bounded between 0 and 1. The metric $\langle W \rangle$ tells us that on all events, the interaction between clusters rose the probability of publication by 0.1–1% on average. We can also note that the values of $\langle W \rangle$ are of the same order of magnitude as λ_0 (0.01 for the “Minute” kernel, 0.001 for “Hour”, and 0.0001 for “Day”). We can interpret this as the probability for a new document belonging to a cluster or being from a new cluster is roughly the same from a temporal perspective. The metric $\langle A \rangle_W$ tells us that when weighting the average of A with the effective interaction, the values of A are slightly higher than 0.05; we can now be confident in this value, given it has been inferred on a statistically significant number of observations. Still, only some interactions seem to be significant, which correlates with [15, 19]. Finally, the last metric $\frac{\langle W^{intra} \rangle}{\langle W^{extra} \rangle}$ finds that most effective interactions take place more within the same cluster; clusters tend to self-replicate. Further studies of this phenomena would involve an extension of the MPDHP that considers Nested Dirichlet Processes instead of Powered Dirichlet Processes. Clusters would then be broken down into smaller ones, whose interactions could be analyzed.

Another major observation from Table 2 is that standard deviations of effective interactions are large: it hints that some of interactions may play a more significant role in the dataset. In Fig. 3, we plot the distribution of effective interactions for one specific run (“Hour” kernel, $\theta_0 = 0.01$, $r = 1$; we recover the same trend in all other experiments). The results of this figure are similar to the ones of previous studies [15, 19].

Interactions range Finally, in Table 3, we investigate the range of effective interactions. We compute the effective interaction for each entry of the temporal kernel individually and average it over all existing clusters. Importantly, the effective interaction of κ_1 is smaller than others. This is induced by our kernel choice: because κ_1 is centered around $t = 0$, half of the associated Gaussian function accounts negative time differences, which never happens by design. Therefore, only other kernels entries can contribute on both sides of their mean.

We see in Table 2 that influence tends to decrease over time for all the kernels considered. Overall, the interaction between documents still plays a marginal role. We did not plot the standard deviation for visualization purposes, but they are similar as in Table 2; *most* interactions do not play a significant role in the publication of subsequent documents over time. Overall, the increase in probability for a new document to belong to a cluster due to interactions is within 0.1–1%.

5 Conclusion

In this work, we conducted extensive experiments on a real-world large-scale dataset from Reddit. We conducted 24 different experiments, each accounting

Table 3. Interaction range—All the values for effective interaction are given in ten-thousandth (10^{-5}). Influence tends to decrease over time for all kernels.

$\kappa(t)$	θ_0	r	κ_1	κ_2	κ_3	κ_4	κ_5	κ_6	κ_7	κ_8	κ_9
Minute (m)	0.01	0.5	133	421	407	451	428	403	395	345	121
		1	198	591	580	607	532	575	521	507	224
		1.5	308	937	893	914	955	840	808	810	304
	0.001	0.5	218	509	457	424	371	340	313	178	52
		1	142	435	396	388	343	327	272	187	45
		1.5	104	388	366	353	326	333	290	215	61
Hour (h)	0.01	0.5	247	430	502	456	324				
		1	329	538	549	542	451				
		1.5	229	615	532	526	411				
	0.001	0.5	62	149	119	137	92				
		1	77	164	172	149	111				
		1.5	104	244	223	197	156				
Day (d)	0.01	0.5	22	45	47	46	47	47	37		
		1	35	71	72	68	68	70	61		
		1.5	51	100	101	105	98	105	82		
	0.001	0.5	9	20	21	21	21	21	21	17	
		1	12	26	24	27	28	26	22		
		1.5	11	41	42	41	41	41	35		

for a given combination of parameters, that determine the timescale considered ($\kappa(\Delta t)$), the sparsity of the language modeling (θ_0), and the extent on which we rely on text or time during the inference (r).

Our experiments hint that interactions do not play a significant role in this dataset. We proposed several ways to assess the role of interactions in the dataset. In particular, we introduced the notion of *effective interaction* as a way to evaluate how confident we are in MPDHP’s output. On this basis, we analysed the importance of interactions in general, as well as from a temporal perspective. We recovered the conclusions of prior works: interactions are sparse and decay over time. By looking at the global effective interaction average, we conclude that interactions play a minor role this dataset. Overall, they only increase the instantaneous probability for a new observation to appear by 1%. Even the most extreme interactions seem to only increase this probability by 12% top.

However, despite intending our study as exhaustive, there is room for improvement in interaction modelling using MPDHP. In particular, there are two biases that we could not explore in this work. Firstly, the parameter λ_0 has been set according to a heuristic (so that a new cluster is opened with fifty percent chances when we are 95% sure that it does not match the existing one). Its direct inference would robustify the approach. Nevertheless, this inclusion sounds challenging: λ_0 does not account for individual events realizations, but for Hawkes processes starts, whose inference is not a trivial extension. Another improvement would be to allow clusters to passively replicate—without the need for an interaction. We expect that this would boil down to adding a time-independent kernel entry to κ . However, other questions may arise from such modification: when to consider a cluster as extinct given a non-fading kernel? How should this kernel relate to the temporal concentration parameter λ_0 ? We believe such improvements would make MPDHP more robust and interpretable, and find applications beyond interaction modelling.

References

1. Barbieri, N., Manco, G., Ritacco, E.: Survival factorization on diffusion networks. In: Machine Learning and Knowledge Discovery in Databases, pp. 684–700 (2017)
2. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J.: The pushshift reddit dataset. Proc. Int. AAAI Conf. Web Social Media **14**(1), 830–839 (2020)
3. Beutel, A., Prakash, B., Rosenfeld, R., Faloutsos, C.: Interacting viruses in networks: can both survive? Proc. ACM SIGKDD (2012)
4. Cao, J., Sun, W.: Sequential choice bandits: learning with marketing fatigue. AAAI-19 (2019)
5. Du, N., Farajtabar, M., Ahmed, A., Smola, A., Song, L.: Dirichlet-Hawkes processes with applications to clustering continuous-time document streams. In: 21th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2015)
6. Du, N., Song, L., Woo, H., Zha, H.: Uncover topic-sensitive information diffusion networks. Proc. Sixteenth Int. Conf. Artif. Intell. Stat. **31**, 229–237 (2013)
7. Gomez-Rodriguez, M.: Structure and Dynamics of Diffusion Networks. Ph.D. thesis (2013)

8. Gomez-Rodriguez, M., Balduzzi, D., Schölkopf, B.: Uncovering the temporal dynamics of diffusion networks. *ICML*, pp. 561–568 (2011)
9. Gomez-Rodriguez, M., Leskovec, J., Schölkopf, B.: Structure and dynamics of information pathways in online media. *WSDM* (2013)
10. He, X., Rekatsinas, T., Foulds, J.R., Getoor, L., Liu, Y.: Hawkestopic: A joint model for network inference and topic modeling from text-based cascades. *ICML* (2015)
11. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146 (2003)
12. Larremore, D., Carpenter, M., Ott, E., Restrepo, J.: Statistical properties of avalanches in networks. *Phys. Rev. E* **85** (2012)
13. Liniger, T.J.: Multivariate Hawkes processes. Ph.D. thesis (ETH Zurich) (2009)
14. Mavroforakis, C., Valera, I., Gomez-Rodriguez, M.: Modeling the dynamics of learning activity on the web. In: *Proceedings of the 26th International Conference on World Wide Web*, pp. 1421–1430 (2017)
15. Myers, S.A., Leskovec, J.: Clash of the contagions: Cooperation and competition in information diffusion. In: *2012 IEEE 12th International Conference on Data Mining*, pp. 539–548 (2012)
16. Myers, S.A., Zhu, C., Leskovec, J.: Information diffusion and external influence in networks. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 33–41 (2012)
17. Niemczura, B., Gliwa, B., Zygmunt, A.: Linear Threshold Behavioral Model for the Spread of Influence in Recommendation Services, pp. 98–105 (2015)
18. Poux-Médard, G., Velcin, J., Loudcher, S.: Information interaction profile of choice adoption. In: *Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*. Research Track, pp. 103–118 (2021)
19. Poux-Médard, G., Velcin, J., Loudcher, S.: Information interactions in outcome prediction: quantification and interpretation using stochastic block models. In: *Fifteenth ACM Conference on Recommender Systems (RecSys)*, pp. 199–208 (2021)
20. Poux-Médard, G., Velcin, J., Loudcher, S.: Powered hawkes-dirichlet process: challenging textual clustering using a flexible temporal prior. In: *2021 IEEE International Conference on Data Mining (ICDM)*, pp. 509–518 (2021)
21. Poux-Médard, G.: Interactions in information spread. In: *WWW 2022: Companion Proceedings of the Web Conference 2022*, pp 313–317 (2022). <https://doi.org/10.1145/3487553.3524190>
22. Yin, J., Chao, D., Liu, Z., Zhang, W., Yu, X., Wang, J.: Model-based clustering of short text streams. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2634–2642 (2018)