

Interactions in Information Spread

Under the supervision of J. Velcin and S. Loudcher

Gaël Poux-Médard

Université de Lyon, France
Lyon 2, ERIC UR 3083

April 21, 2022



Introduction

- Every minute:

 400h of video
 350 000 tweets

 500 000 comments
 4 200 000 searches

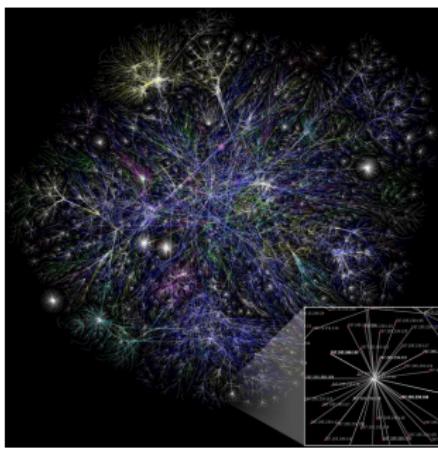


Figure 1: Snapshot of the internet (Wikipedia)

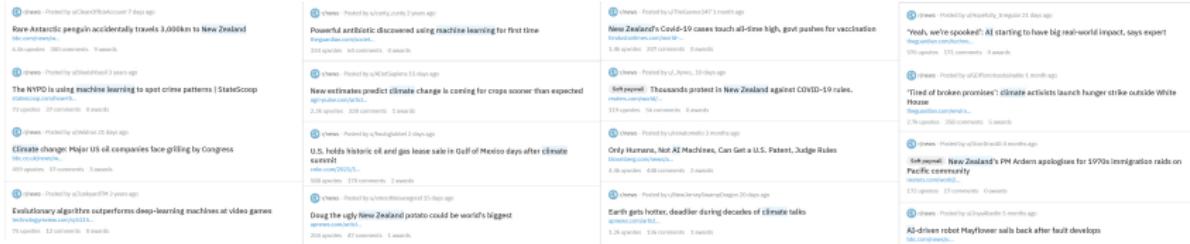
Motivation

- Every minute:

 400h of video
 350 000 tweets

 500 000 comments
 4 200 000 searches

- How does this data get generated?



The screenshot shows a horizontal scroll of a news feed from the subreddit r/news. Each post includes a thumbnail, the title, the number of upvotes, the number of comments, and the number of awards.

- Post 1:** More Antarctic penguin accidentally travels 3,000km to New Zealand | National Geographic. 1.6k upvotes, 380 comments, 9 awards.
- Post 2:** We've discovered a machine learning algorithm that can predict climate change coming for crops sooner than expected | Ars Technica. 2.1k upvotes, 228 comments, 3 awards.
- Post 3:** The NYPD is using machine learning to spot crime patterns | StateScoop. 1.9k upvotes, 37 comments, 4 awards.
- Post 4:** Climate change: Major US oil companies face grilling by Congress | Ars Technica. 200 upvotes, 53 comments, 3 awards.
- Post 5:** Evolutionary algorithm outperforms deep-learning machines at video games | Ars Technica. 1.2k upvotes, 12 comments, 1 award.
- Post 6:** Powerful antibiotic discovered using machine learning for first time | Ars Technica. 1.8k upvotes, 165 comments, 4 awards.
- Post 7:** New Zealand's Covid-19 cases touch all-time high, govt pushes for vaccination | The New Zealand Herald. 3.1k upvotes, 207 comments, 6 awards.
- Post 8:** Cheats - Printed by u/DeezNuts 13 days ago. 1.5k upvotes, 104 comments, 10 awards.
- Post 9:** Cheats - Printed by u/DeezNuts 13 days ago. 2.1k upvotes, 218 comments, 3 awards.
- Post 10:** Thousands protest in New Zealand against COVID-19 rules. | cnet.com/pic/... 3.1k upvotes, 211 comments, 4 awards.
- Post 11:** Cheats - Printed by u/DeezNuts 2 months ago. 1.5k upvotes, 179 comments, 2 awards.
- Post 12:** U.S. holds historic oil and gas lease sale in Gulf of Mexico days after climate summit | Ars Technica. 1.6k upvotes, 178 comments, 2 awards.
- Post 13:** Cheats - Printed by u/DeezNuts 2 months ago. 2.1k upvotes, 218 comments, 3 awards.
- Post 14:** Only Humans, Not AI Machines, Can Get a U.S. Patent, Judge Rules | Ars Technica. 4.3k upvotes, 410 comments, 2 awards.
- Post 15:** Cheats - Printed by u/DeezNuts 20 days ago. 2.1k upvotes, 47 comments, 1 award.
- Post 16:** Earth gets hotter, deadlier during decades of climate talks | Ars Technica. 1.2k upvotes, 136 comments, 1 award.
- Post 17:** Cheats - Printed by u/DeezNuts 2 months ago. 1.5k upvotes, 177 comments, 2 awards.
- Post 18:** Cheats - Printed by u/DeezNuts 2 months ago. 1.5k upvotes, 177 comments, 2 awards.
- Post 19:** New Zealand's PM Ardern apologizes for 1970s immigration raids on Pacific community | news.com.au... 1.5k upvotes, 57 comments, 3 awards.
- Post 20:** Cheats - Printed by u/DeezNuts 2 months ago. 1.5k upvotes, 177 comments, 2 awards.
- Post 21:** Ai-driven robot Mayflower sails back after fault develops | Ars Technica. 1.2k upvotes, 136 comments, 1 award.

Figure 2: A typical stream from r/news

Motivation

- Every minute:

400h of video
 350 000 tweets

500 000 comments
 4 200 000 searches

- How does this data get generated?
 → Hidden interactions?

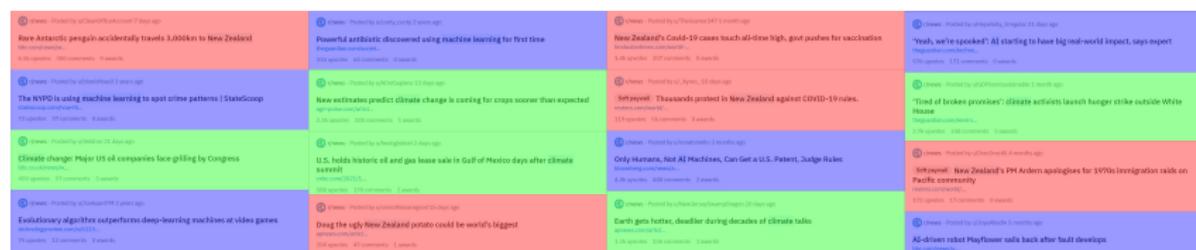


Figure 2: A typical stream from r/news – with topics

Context

Modelling the interaction between pieces of information and characterizing their role in information spreading processes

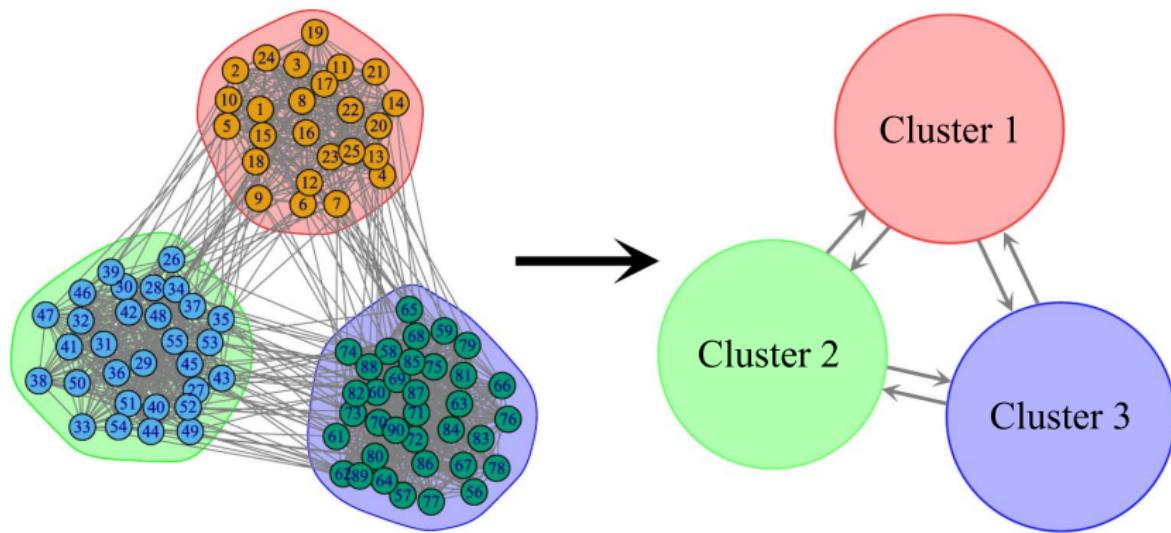
- Information: any item susceptible to spread (tweets, news, ...)
- Action: reaction to a piece of information (retweet, share, ...)
- Interaction: when the joint effect of several pieces of information is different than the product of their independent effects.
 - $P(x|A, B) \neq P(x|A)P(x|B)$

State of the art

- Few works considered the topic from a machine learning perspective
 - Clash of the Contagions (S.A.Myers et J.Leskovec, ICDM 2012)
 - Correlated cascades (Zarezade et al., AAAI 2017)
- Several theoretical works on interacting processes [Prakash et al., 2012, Wang et al., 2019, Zhu et al., 2020]
 - Define micro-rules first
 - Compare to global statistics then
 - No learning from the data

Stochastic Block Models

- Dimension reduction via clustering
 - Nodes: pieces of information (e.g. tweets)
 - Links: outcome given a pair of nodes (e.g. retweet or not)



Results

- Four datasets:
 - PubMed: how symptoms interact to refine a diagnosis
 - Reddit: how words interact to trigger an answer
 - Spotify: how hearing some songs influence the next one listened to
 - Twitter: how exposure to tweets trigger retweeting ulterior ones

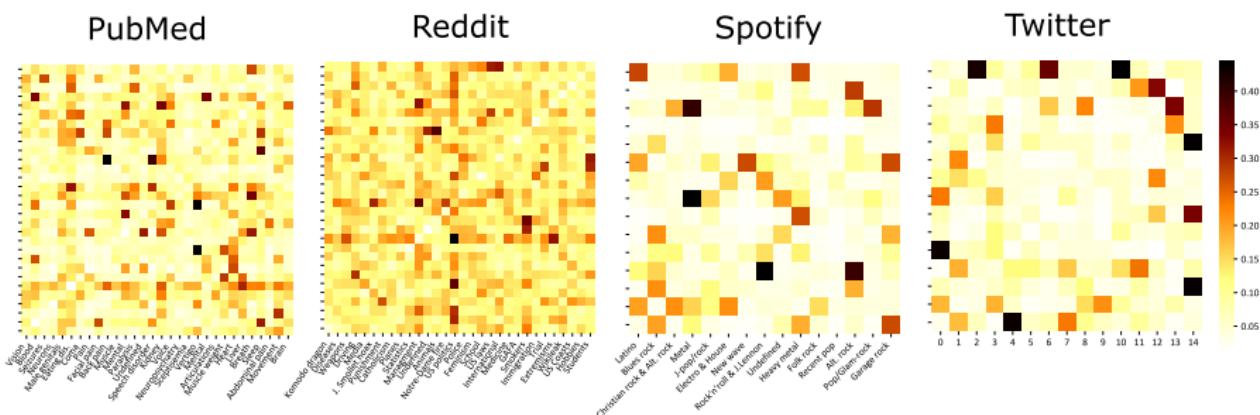
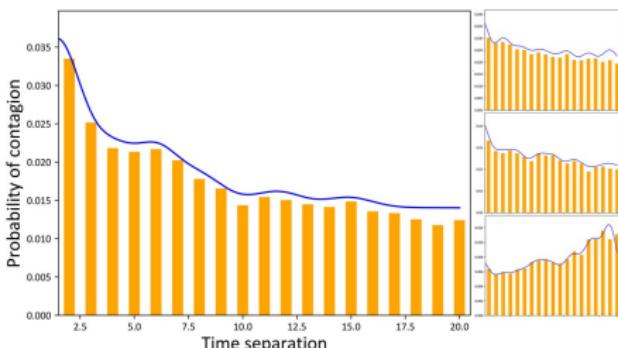


Figure 3: Conclusion: significant interactions between pieces of information are rare (Poux-Médard et al., RecSys 2021)

InterRate

- How long do pair interactions last in time?
 - Piece of information A at time t_A and B at time $t_B > t_A$: how A relates to B after a time $\Delta t = t_B - t_A$?

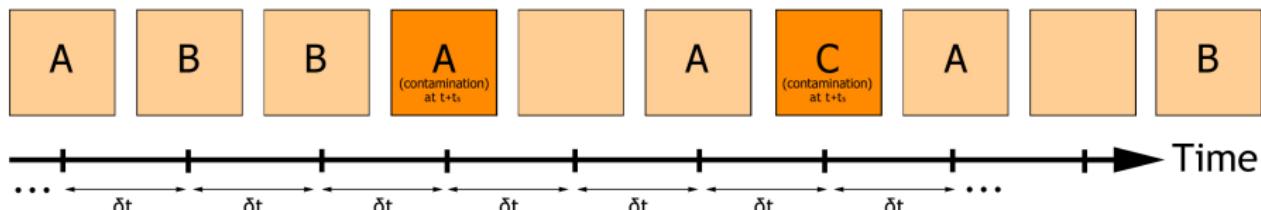


A : exposure to A

A_{contaminated} : exposure to A at t and contamination by A at $t+t_s$

t_s : time between exposures and contaminations

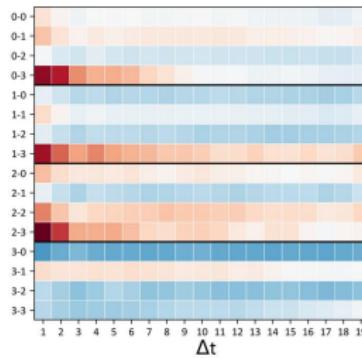
δt : time between exposures



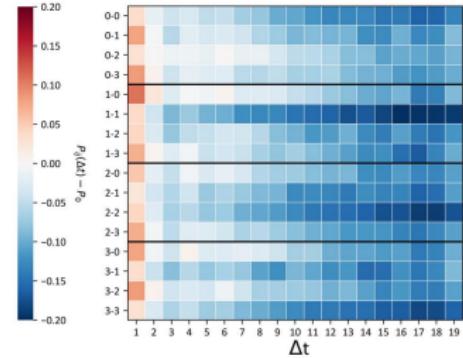
Results

- 3 datasets:
 - Twitter: how old tweets influence the probability of retweets
 - Ads: how exposure to ads influence the probability of clicking one
 - Prisoner's dilemma: how past cooperation or betrayal influence one's decisions

Twitter



Ads



PD

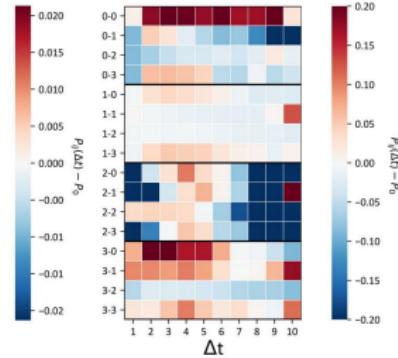


Figure 4: Conclusion: interactions between pieces of information do not last long (Poux-Médard *et al.*, ECML-PKDD 2021)

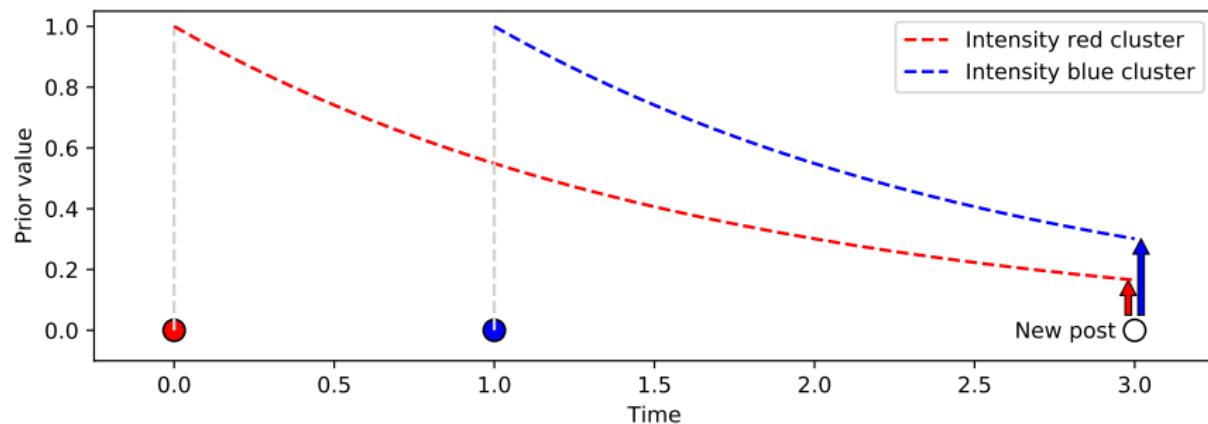
Synthesis

- Interactions only take place between a restrictive set of entities
 - Clustering is needed
- Interactions do not last in time
 - Modeling time is necessary
- Interactions improve datasets description
- Solution: jointly model clusters and their dynamic interactions
 - Promising lead: Dirichlet-Hawkes processes (Du *et al.*, KDD 2015)

Dirichlet-Hawkes process

- (Du et al., KDD 2015): Dirichlet-Hawkes prior (Bayesian inference)
- Clusters can self-replicate (self-interaction)

$$P(\text{cluster}|\text{text}, \text{time}, H) \propto \underbrace{P(\text{text}|\text{cluster})}_{\text{Textual likelihood} \\ (\text{Dirichlet-Multinomial})} \times \underbrace{P(\text{cluster}|\text{time}, H)}_{\text{Temporal prior} \\ (\text{Dirichlet-Hawkes})}$$



Powered Dirichlet-Hawkes process

- (Poux-Médard *et al.*, ICDM 2021): Powered Dirichlet-Hawkes prior

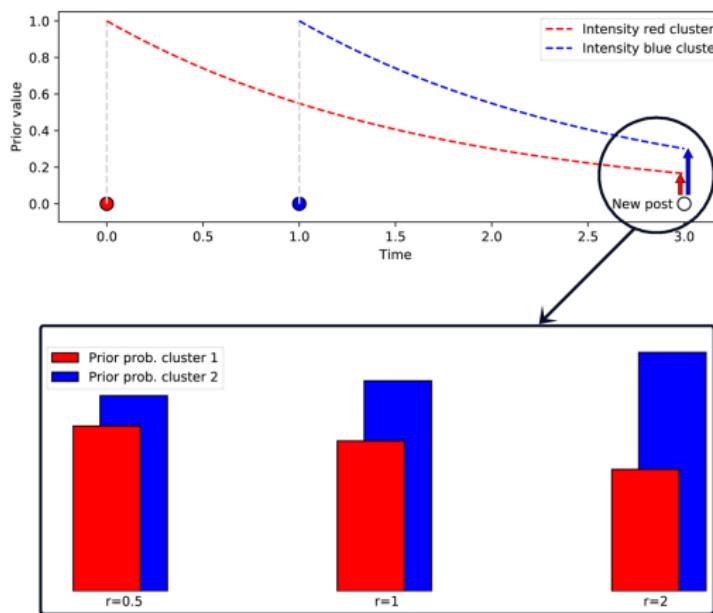


Figure 6: Powered Dirichlet-Hawkes prior

Results for various overlaps

- Powered Dirichlet-Hawkes process: works better in challenging situations
 - Scarce textual information (short texts, overlapping voc.)
 - Scarce temporal information (entangled dynamics)

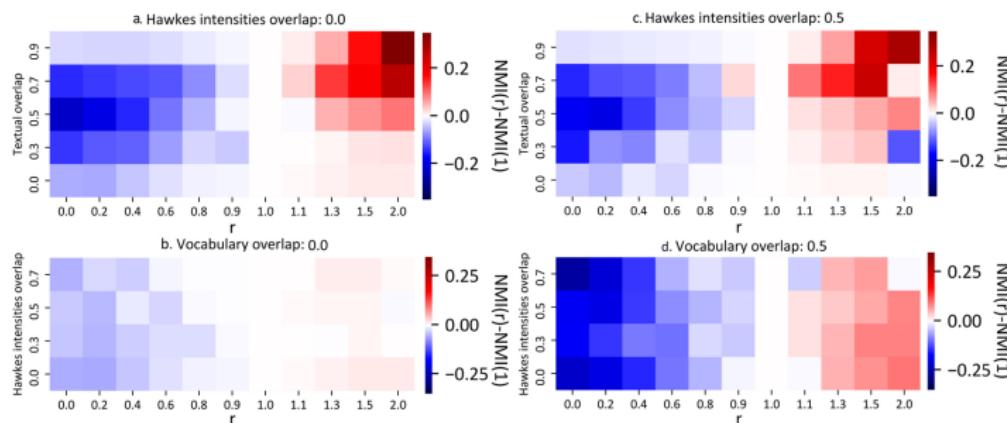
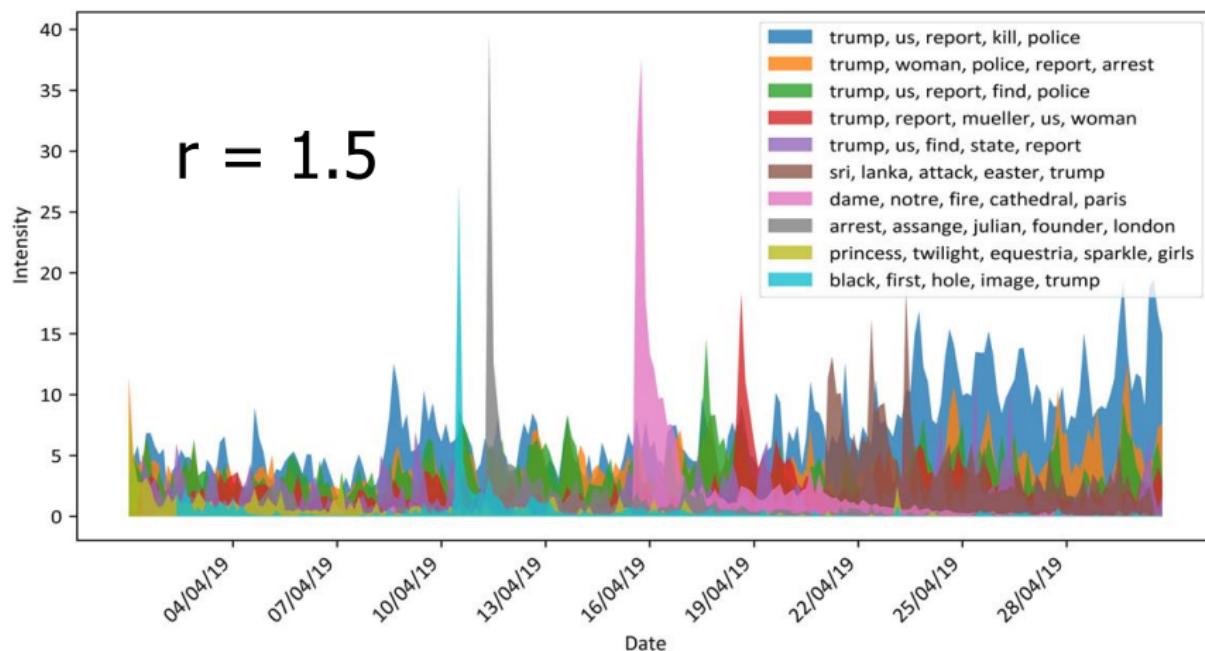


Figure 7: Red is better — (Poux-Médard et al., ICDM 2021)

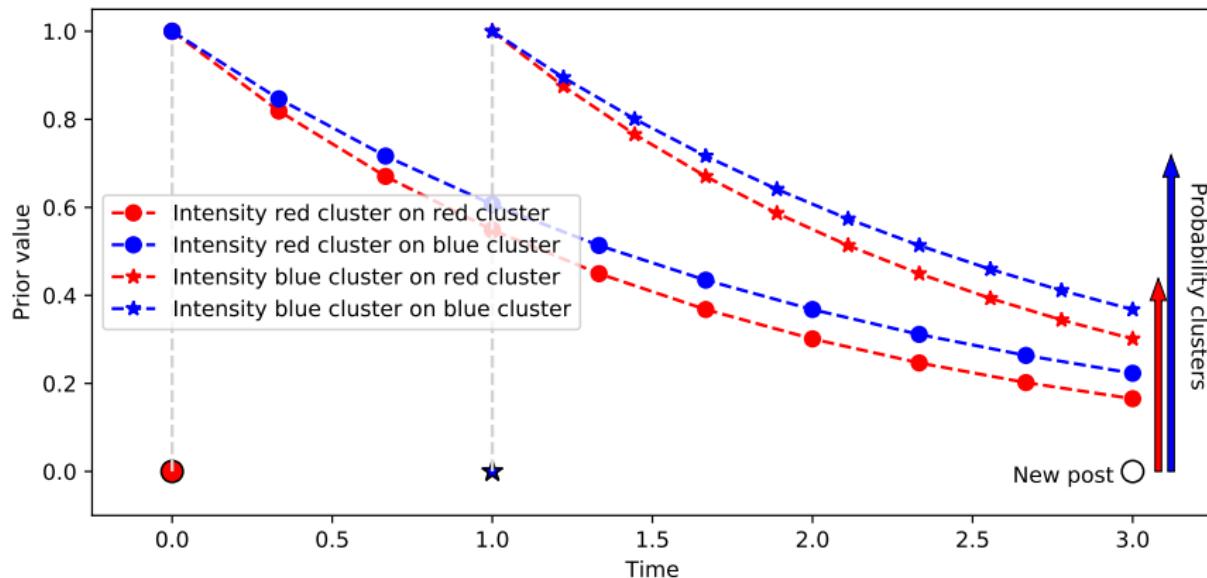
Powered Dirichlet-Hawkes process

- Powered Dirichlet-Hawkes prior: summary from data flows using temporal interactions



Perspective: Multivariate Powered Dirichlet-Hawkes process

- Perspective: Multivariate Powered Dirichlet-Hawkes prior
 - How clusters influence each other



Perspective MPDHP – Cluster interaction network

- Multivariate Powered Dirichlet-Hawkes prior: Cluster interaction network

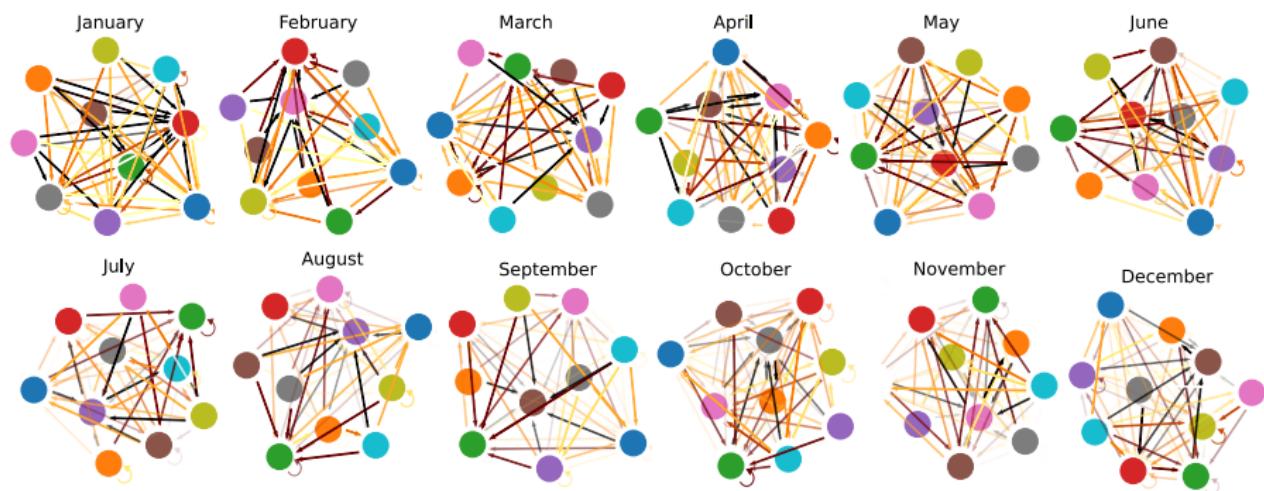


Figure 8: Topical interaction network

Perspective MPDHP – Summary generation

- Multivariate Powered Dirichlet-Hawkes prior: Cluster interaction network

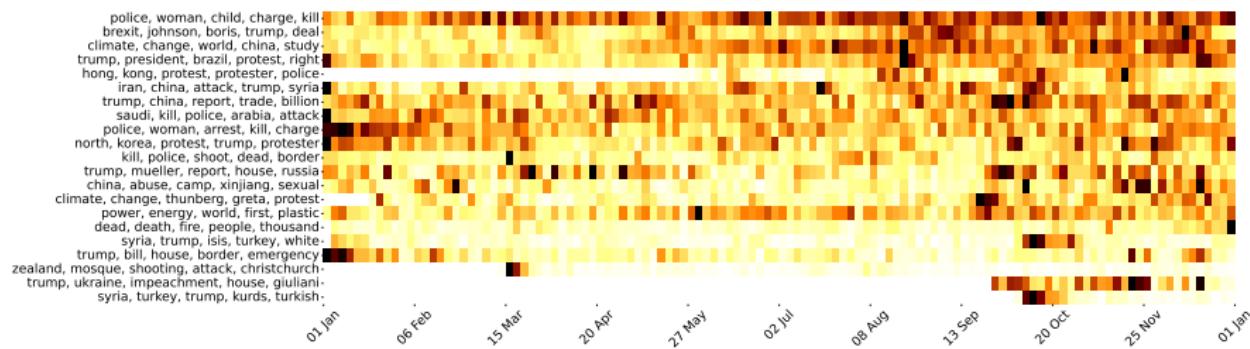
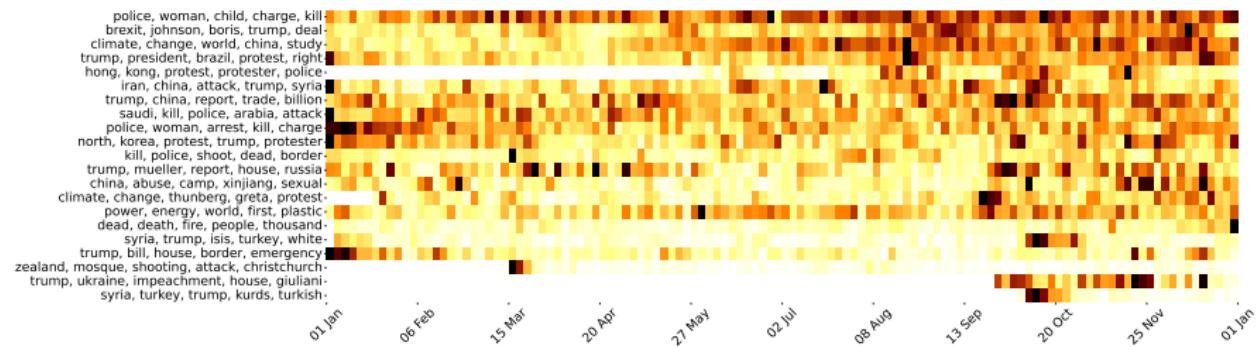


Figure 9: Inferred topics timeline

Thanks for your attention!



Bibliography I

[Prakash et al., 2012] Prakash, B., Beutel, A., Rosenfeld, R., and Faloutsos, C. (2012).

Winner takes all: Competing viruses or ideas on fair-play networks.

WWW.

[Wang et al., 2019] Wang, W., Liu, Q.-H., Liang, J., Hu, Y., and Zhou, T. (2019).

Coevolution spreading in complex networks.

Physics Reports, 820:1–51.

Coevolution spreading in complex networks.

[Zhu et al., 2020] Zhu, Z., Gao, C., and Zhang, Y. (2020).

Cooperation and competition among information on social networks.

Nature Sci Rep, 3103:12160.