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

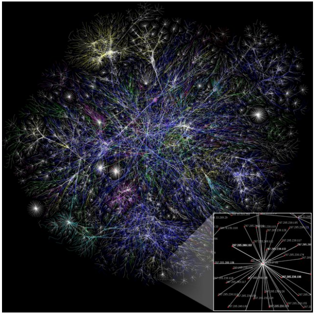


Figure 1: Snapshot of the internet (Wikipedia)

Motivation

- Every minute:

  - 400h of video
  - 350 000 tweets
  - 500 000 comments
  - 4 200 000 searches

- How does this data get generated?

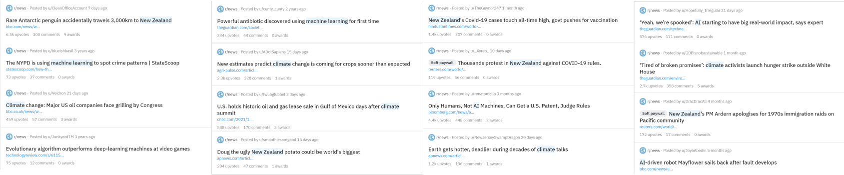


Figure 2: A typical stream from r/news

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.
  - $P(x|A,B) \neq P(x|A)P(x|B)$

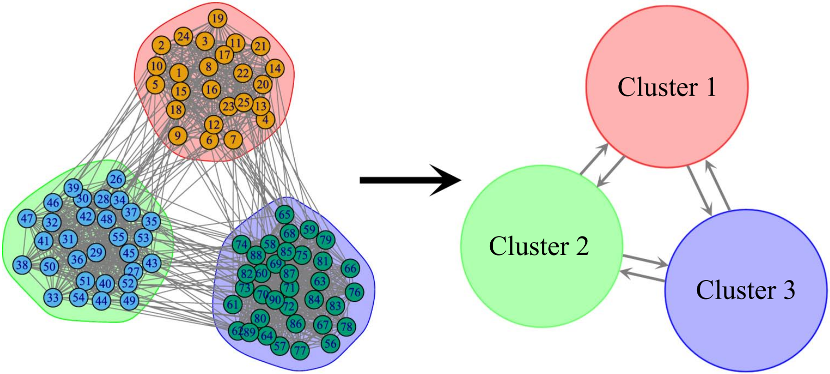
State of the art

- 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

  $\rightarrow$ 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

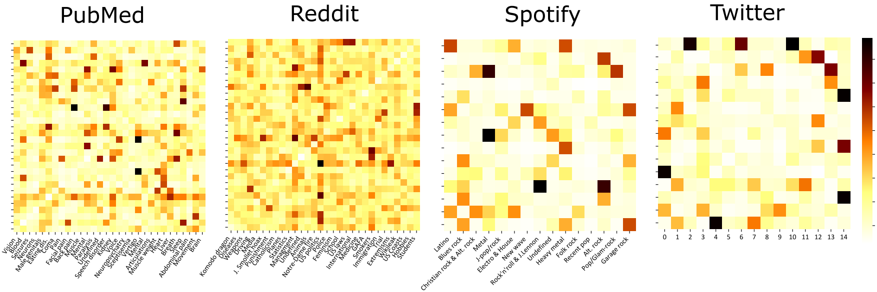
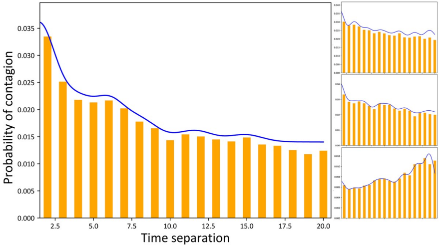
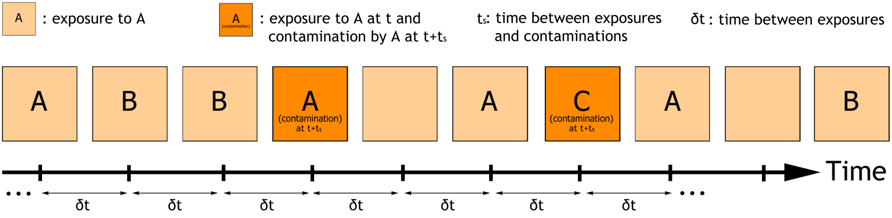


Figure 3: Conclusion: significant interactions between pieces of information are rare (Poux-Médard *et al.*, RecSys 2021)

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 $\Delta t = t_B - t_A$?





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

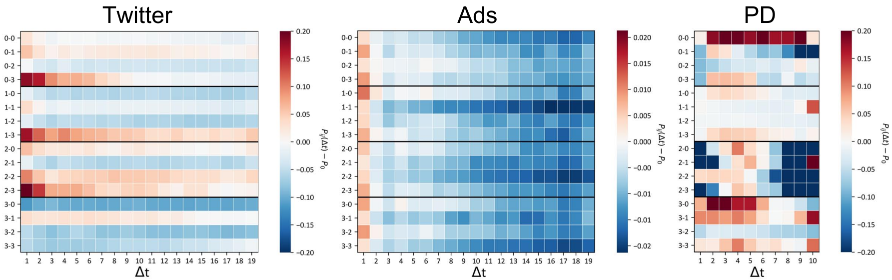


Figure 4: Conclusion: interactions between pieces of information do not last long (Poux-Médard *et al.*, ECML-PKDD 2021)

Synthesis

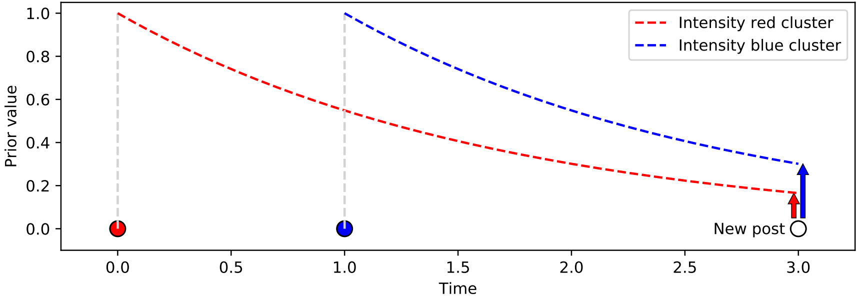
- Interactions only take place between a restrictive set of entities
  $\rightarrow$ Clustering is needed
- Interactions do not last in time
  $\rightarrow$ Modeling time is necessary
- Interactions improve datasets description

- 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)

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



Powered Dirichlet-Hawkes process

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

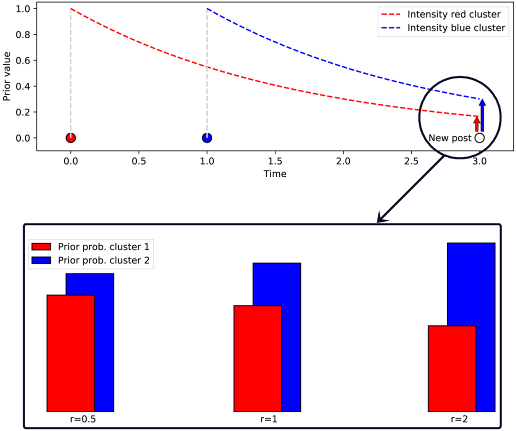


Figure 6: Powered Dirichlet-Hawkes prior

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)

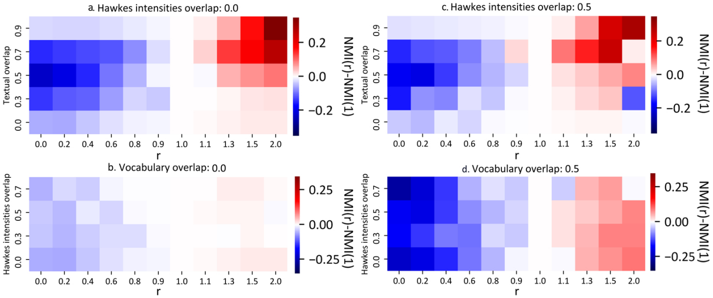
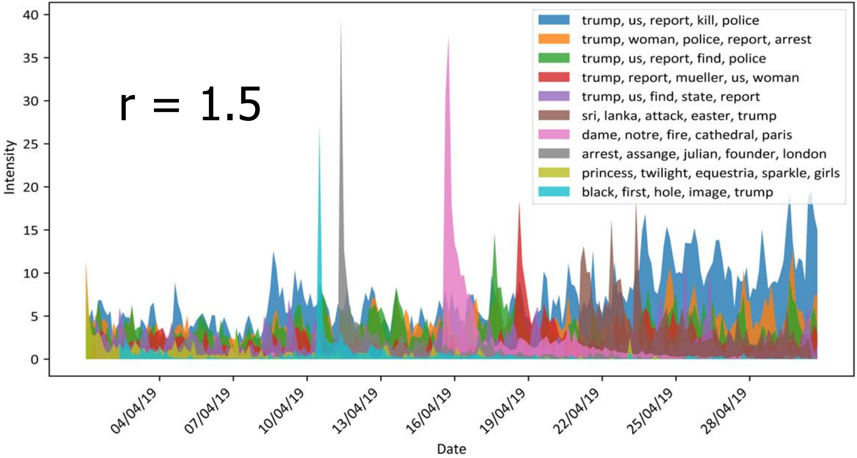


Figure 7: Red is better — (Poux-Médard *et al.*, ICDM 2021)

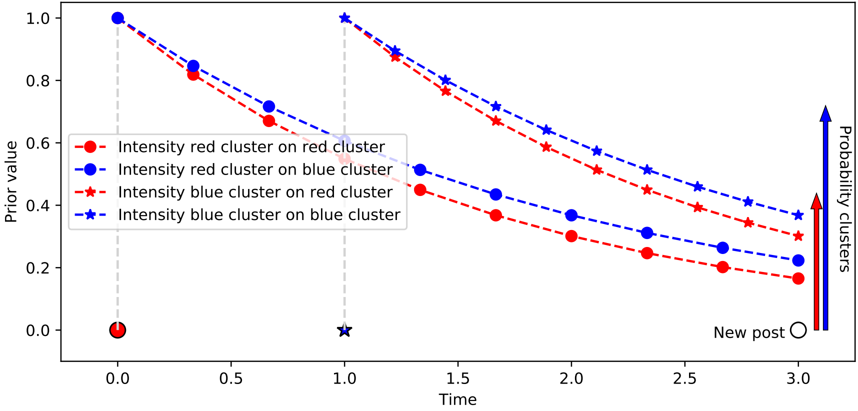
Powered Dirichlet-Hawkes process

- 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

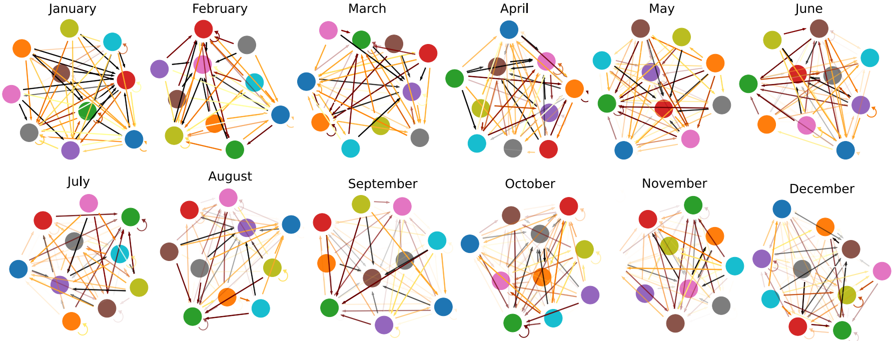


Figure 8: Topical interaction network

Perspective MPDHP – Summary generation

- Multivariate Powered Dirichlet-Hawkes prior: Cluster interaction network

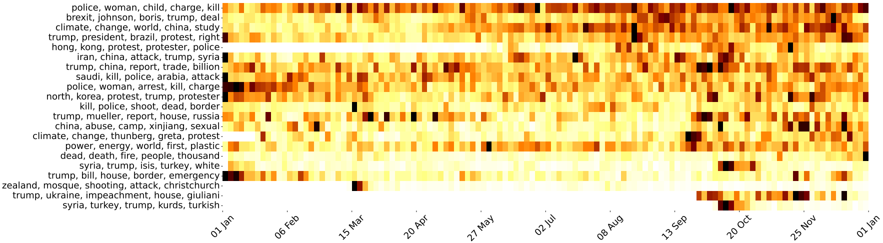
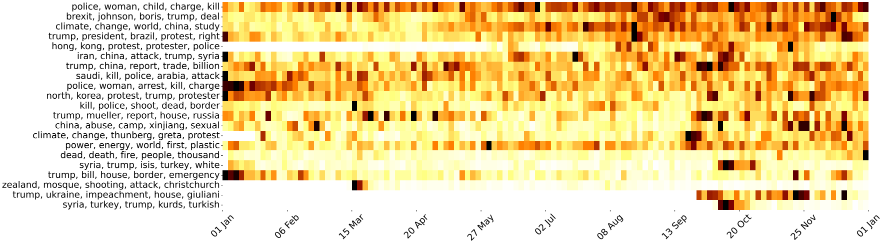


Figure 9: Inferred topics timeline

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.