

Motivation  
oooo

DP  
oooo

HP  
oooooo

DHP  
oooooooo

PDP  
oooo

PDHP  
oooooooo

Houston  
oooooooooooo

Conclusion  
oo

# Dirichlet-Point processes

Gaël Poux-Médard

Université de Lyon, France  
Lyon 2, ERIC UR 3083

November 2021



# Introduction

- Every minute:

 400h of video  
 350 000 tweets

 500 000 comments  
 4 200 000 searches

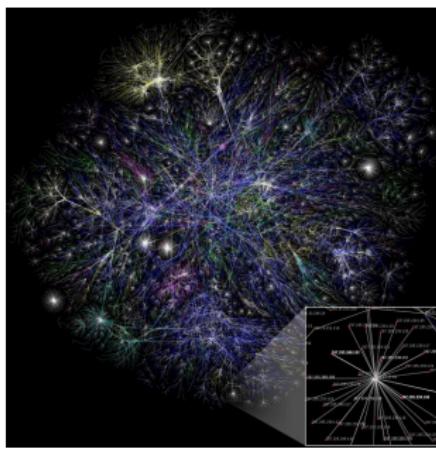


Figure 1: Snapshot of the internet (Wikipedia)

## Motivation

- Every minute:



400h of video



350 000 tweets



500 000 comments



4 200 000 searches

- How to make sense out of *that*?

- 1 news - Probability of being infected 7 days ago.  
Rare Antarctic penguin accidentally travels 3,000km to New Zealand  
1,630 updates, 380 comments, 3 tweets.
- 1 news - Probability of being infected 1 years ago.  
The NYPD is using machine learning to spot crime patterns | Statista  
1,000 updates, 29 comments, 4 tweets.
- 1 news - Probability of being infected 25 days ago.  
Climate change: Major US oil companies face grilling by Congress  
490 updates, 97 comments, 3 tweets.
- 1 news - Probability of being infected 1 years ago.  
Evolutionary algorithm outperforms deep-learning machine at video game  
<https://arxiv.org/abs/1906.04238>  
71 update, 12 comments, 3 tweets.

- [1]  [1] **Health**** - Posted by [schwartz.com](#), 2 years ago  
**Powerful antibiotic resistance discovered using machine learning for first-time sequencing**  
1,000 reads, 40 comments, 10 weeks ago
- [1]  [1] **Health**** - Posted by [u/NetGardens](#), 13 days ago  
**New estimates predict climate change is costing us crops across the globe**  
2,114 upvotes, 238 comments, 5 weeks ago
- [1]  [1] **Health**** - Posted by [u/RedGiantGiant](#), 3 days ago  
**U.S. holds historic oil and gas lease sale in Gulf of Mexico days after Biden's executive order**  
1,000 upvotes, 379 comments, 2 weeks ago
- [1]  [1] **Health**** - Posted by [u/Chromophore25](#), 15 days ago  
**Using the ugly New Zealand potato could be world's biggest breakthrough**  
420 upvotes, 47 comments, 1 week ago

1. **Levi** - Printed by iFabrics.com (24 hours ago)  
New Zealand's Covid-19 cases touch all-time high, govt pushes for vaccination  
1.6k upvotes, 207 comments, 3 awards

2. **Levi** - Printed by iFabs, 15 days ago  
SoftBank: Thousands protest in New Zealand against COVID-19 rules.  
[readers.comment](https://www.pcmag.com/coronavirus/softbank-thousands-protest-in-new-zealand-against-covid-19-rules)  
151 upvotes, 14 comments, 8 awards

3. **Levi** - Printed by iFabs (1 month ago)  
Only Humans, Not AI Machines, Can Get a U.S. Patent, Judge Rules  
[readers.comment](https://www.pcmag.com/coronavirus/only-humans-not-ai-machines-can-get-a-u-s-patent-judge-rules)  
4.4k upvotes, 481 comments, 2 awards

4. **Levi** - Printed by iFabs (20 days ago)  
Early geno-hacker, during decades of climate talk  
[readers.comment](https://www.pcmag.com/coronavirus/early-geno-hacker-during-decades-of-climate-talk)  
1.2k upvotes, 139 comments, 3 awards

 [@news](#) [Facebook](#) [Instagram](#) | 2 hours ago  
"Yeah, we're spooked": At starting to have big real-world impact, says expert  
[They're not here to harm us...](#)

 [@news](#) [Facebook](#) [Instagram](#) | 1 month ago  
'Tired of broken promises': climate activists launch hunger strike outside White House  
[They've been here for months...](#)  
2.7K upvotes | 113 comments | 5 search

 [@news](#) [Facebook](#) [Instagram](#) | 4 months ago  
New Zealand's PM Ardern apologises for 1970s immigration raids on Pacific community  
[They're not here to harm us...](#)  
1.7K upvotes | 13 comments | 5 search

 [@news](#) [Facebook](#) [Instagram](#) | 6 months ago  
AI-driven robot Mayflower sails back after fault develops

**Figure 2:** A typical stream from r/news

## Motivation

- Every minute:



400h of video



350 000 tweets

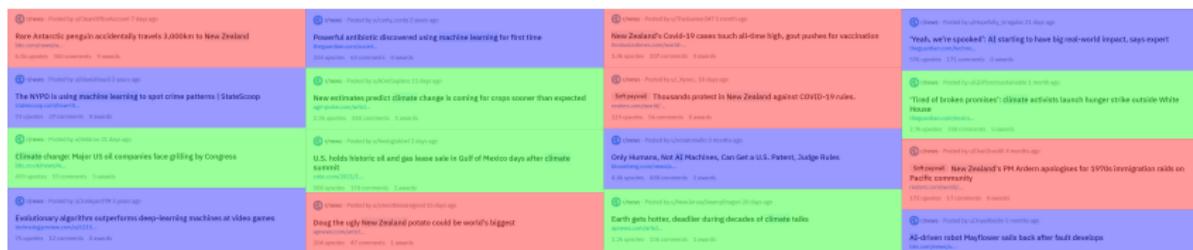


500 000 comments



4 200 000 searches

- How to make sense out of *that*?  
→ Hidden semantic links



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

# Available information

- Main clues:
  - Textual information

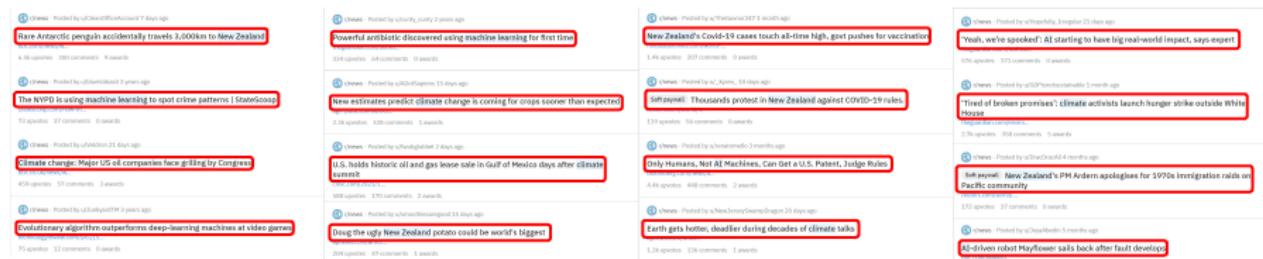


Figure 3: We can use textual information

# Available information

- Main clues:
  - Textual information
  - Temporal information

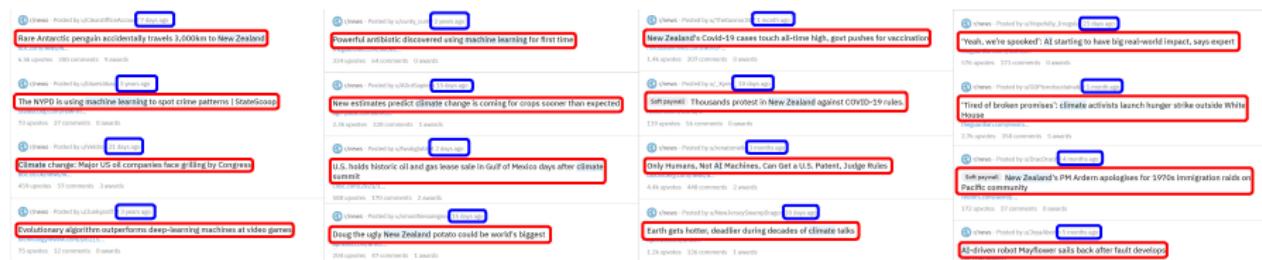


Figure 3: We can use textual information and temporal information

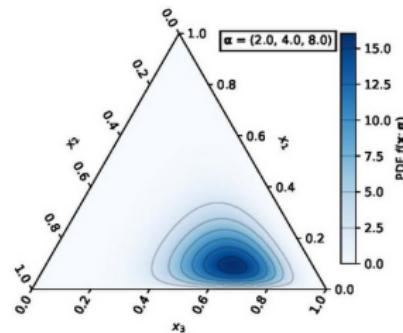
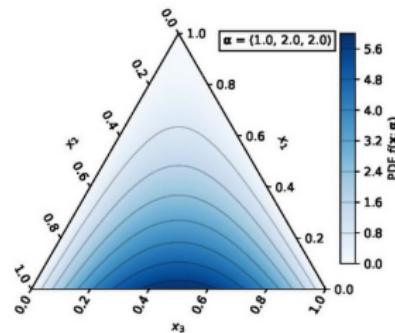
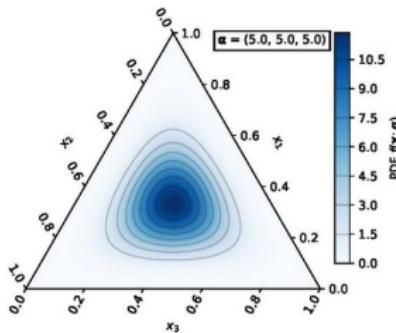
# Documents stream

- The data is therefore a documents stream



# Dirichlet process

- Dirichlet processes fit to consider streams as inputs
- Dirichlet distribution:  $\vec{X} \sim Dir(\alpha)$  s.t.  $\sum_k X_k = 1$
- Often used as a prior distribution in Bayesian clustering
  - ◊ Typically  $X_k$  is the probability to belong to cluster  $k$
- Can be represented in several ways:
  - ◊ Stick-breaking process
  - ◊ Polya-Urn process
  - ◊ Chinese restaurant process



# Chinese restaurant process

- Chinese Restaurant Process:

$$CRP(C_i = c | C_1, C_2, \dots, C_{i-1}, \alpha) = \begin{cases} \frac{N_c}{\alpha + N} & \text{if } c = 1, \dots, K \\ \frac{\alpha}{\alpha + N} & \text{if } c = K+1 \end{cases}$$



## Handling a stream of documents

- Chinese Restaurant Process:

$$CRP(C_i = c | C_1, C_2, \dots, C_{i-1}, \alpha) = \begin{cases} \frac{N_c}{\alpha + N} & \text{if } c = 1, \dots, K \\ \frac{\alpha}{\alpha + N} & \text{if } c = K+1 \end{cases}$$

- Useful for sequential modeling (explicit posterior at each step, allows Gibbs sampling)

$$\underbrace{P(n^{th} obs = c | D, history)}_{Posterior} \propto \underbrace{P(D | n^{th} obs = c)}_{Likelihood} \times \underbrace{P(n^{th} obs = c | history)}_{CRP \ prior}$$

- Hypothesis: “rich-get-richer”

## Variants

- Variants of DP exist:
  - ◊ Uniform process [Wallach et al., 2010]
  - ◊ Pitman-Yor process [Pitman and Yor, 1997]
  - ◊ Hierarchical Dirichlet process [Teh et al., 2006]
  - ◊ Nested Dirichlet process [Rodríguez et al., 2008]
- Most exhibit “rich-get-richer” property
- All consider counts, none consider temporal dimension

## Modeling time as a continuous variable

- Time often “modeled” by sampling observations (DTM [Blei and Lafferty, 2006], TOT [Wang and McCallum, 2006], RCRP [Ahmed and Xing, 2008], DDCRP [Blei and Frazier, 2010], etc.)
  - ◊ Problems: how to slice data, which sampling function use, how to weight observations, etc.
- Whole literature modeling time explicitly: point processes

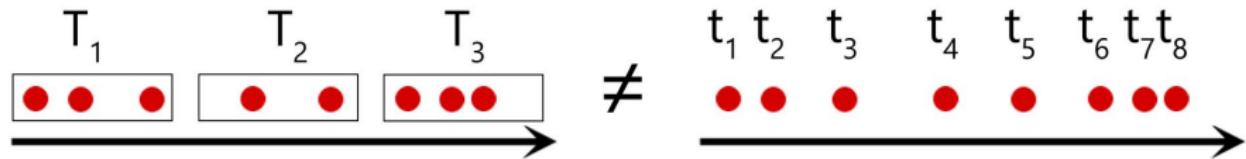


Figure 4: Data sampling/slicing is an approximation

## Poisson process

- Poisson processes are characterized by an **intensity**  $\lambda$ .
  - ◊  $P(\mathbb{N}(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$  = probability for  $n$  events to happen within a time  $t$
- Instantaneous PDF of **one** event (or inter-arrival time PDF):

$$f(t) = \frac{P(\mathbb{N}(t) = 1)}{t} = \lambda e^{-\lambda t}$$

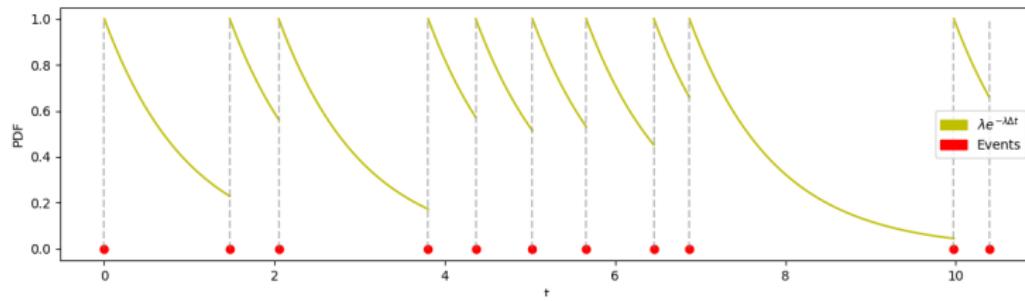
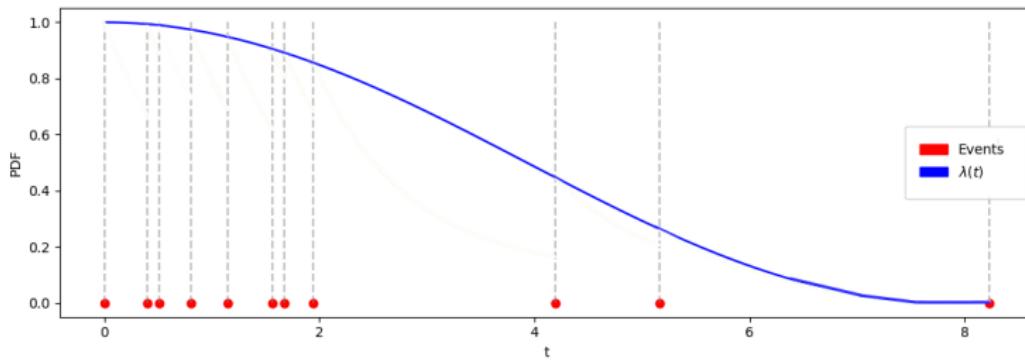


Figure 5: Could model radioactive decay events of atoms whose half-life is 1

## Non-homogeneous Poisson process

- $\lambda(t)$  is a function
  - $\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(\mathbb{N}(t+\Delta t) - \mathbb{N}(t) = 1)}{\Delta t}$



**Figure 6:** Could model cars arrival at gas station throughout a day

## Hawkes process

- Hawkes processes:  $\lambda(t)$  depends on past events  $\mathcal{H}_t = \{t_i | t_i < t\}$   
→ “Self-exciting process”
- Typically:  $\lambda(t) = \lambda_0 + \sum_{t_i \in \mathcal{H}_t} \phi(t - t_i)$

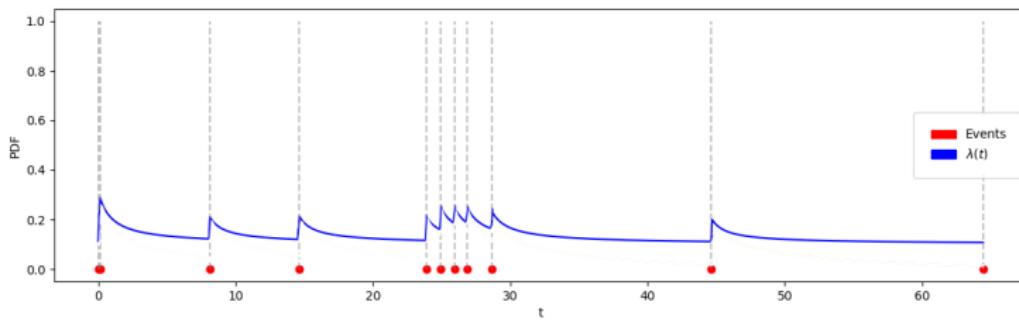


Figure 7: Could model online posting dynamics

## Inference

- Log-likelihood  $\ell(\lambda)$  fit for data streams:

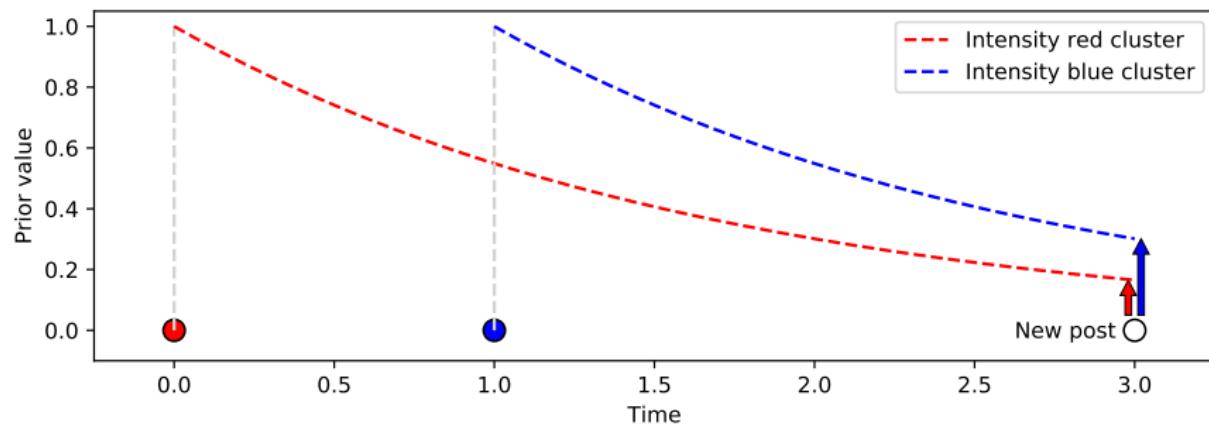
$$\begin{aligned}\ell(\lambda) = & - \int_{t_0}^{t_N} \lambda(t) dt + \sum_{t_i < t_N} \log \lambda(t_i) = \log \lambda(t_1) - \int_{t_0}^{t_1} \lambda(t) dt \\ & + \log \lambda(t_2) - \int_{t_1}^{t_2} \lambda(t) dt \\ & + \dots \\ & + \log \lambda(t_N) - \int_{t_{N-1}}^{t_N} \lambda(t) dt\end{aligned}$$

- Convex for certain shapes of  $\lambda(t)$  (exp, ray, PL, Gaussian, ...).

## Dirichlet-Hawkes process

- [Du et al., 2015]: Dirichlet-Hawkes prior (Bayesian inference)
- Merges Dirichlet priors and Hawkes processes

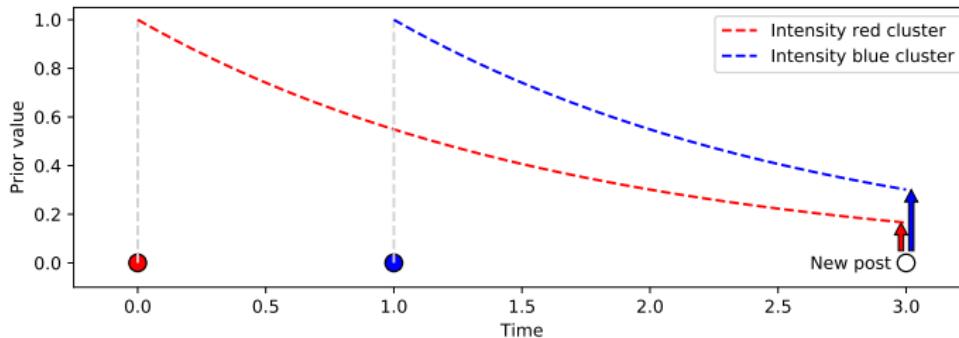
$$P(\text{cluster}|\text{text}, \text{time}, H) \propto \underbrace{P(\text{text}|\text{cluster})}_{\text{Textual likelihood} \\ (\text{Dirichlet-Multinomial})} \times \underbrace{P(\text{cluster}|\text{time}, H)}_{\text{Temporal prior} \\ (\text{Dirichlet-Hawkes})}$$



# Dirichlet-Hawkes process – Explicit

- $P(c|t, \mathcal{H})$ : prior probability of cluster  $c$  at time  $t$  given history  $\mathcal{H}$
- $\lambda_c(t)$ : intensity of cluster  $c$  at time  $t$
- Dirichlet process with counts  $N_c$  replaced by  $\lambda_c(t)$

$$\underbrace{P(c|t, \mathcal{H})}_{\substack{\text{Temporal prior} \\ (\text{Dirichlet-Hawkes})}} = \begin{cases} \frac{\lambda_c(t)}{\alpha_0 + \sum_k \lambda_k(t)} & \text{if } c = 1, \dots, K \\ \frac{\alpha_0}{\alpha_0 + \sum_k \lambda_k(t)} & \text{if } c = K+1 \end{cases}$$



## Motivation

oooo

HP

○○○○

DHP

○○●○○○○○○

PDP

0000

PDHP

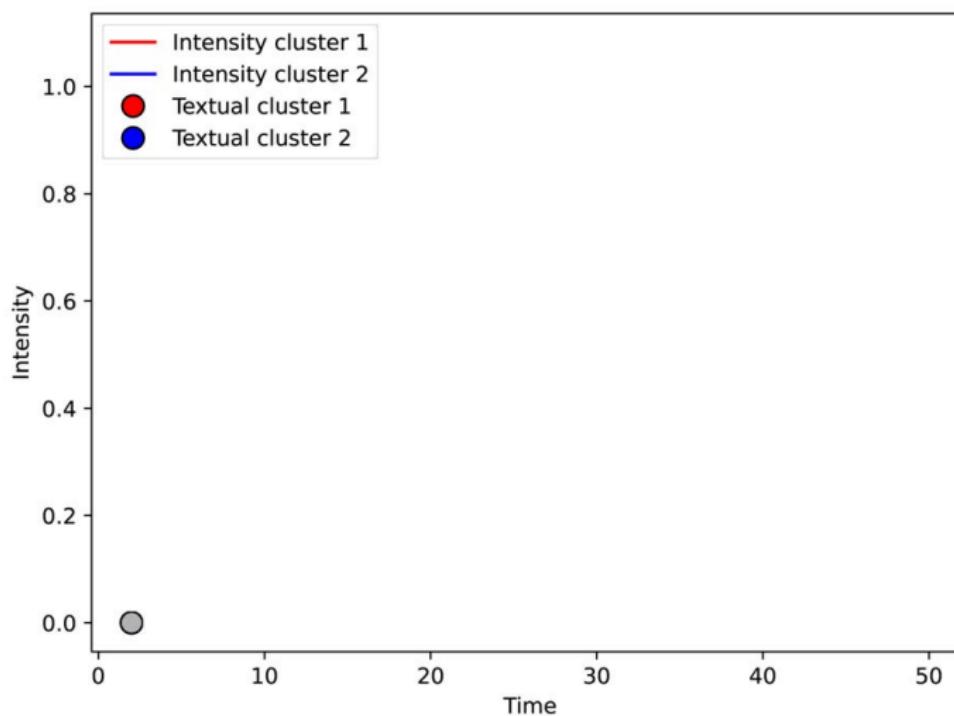
○○○○○○○○○○

Houston

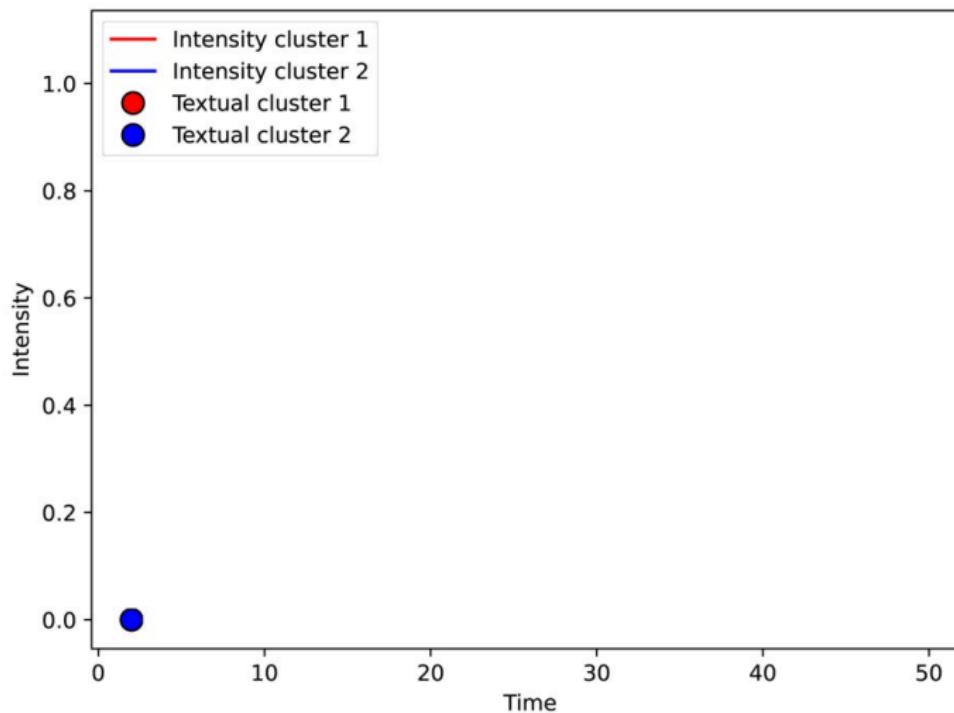
○○○○○○○○○○

## Conclusion

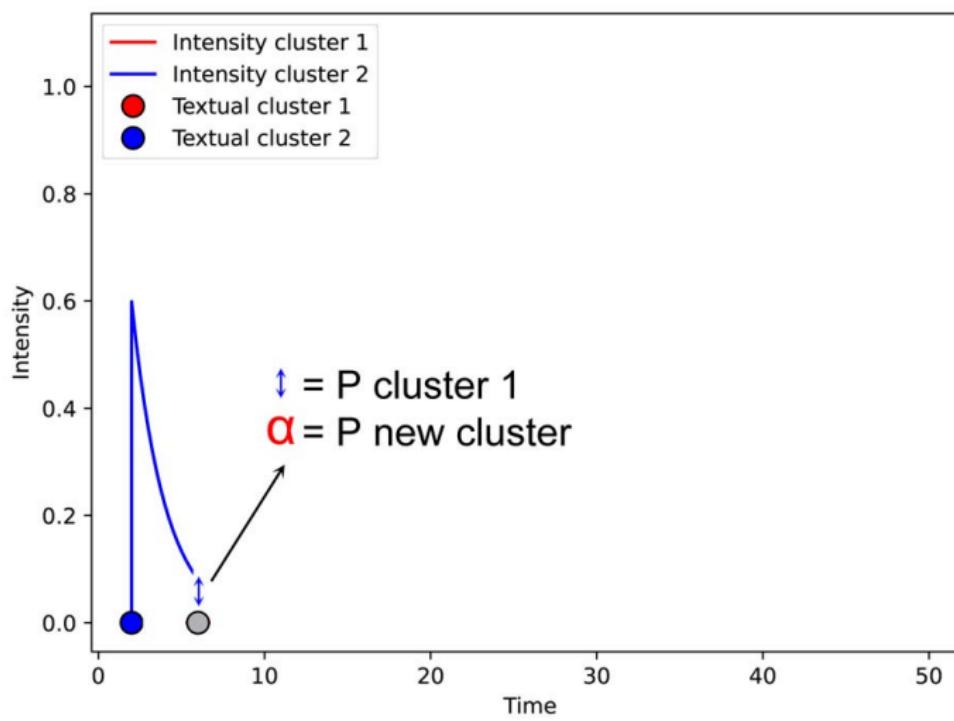
## Inference (1 particle)



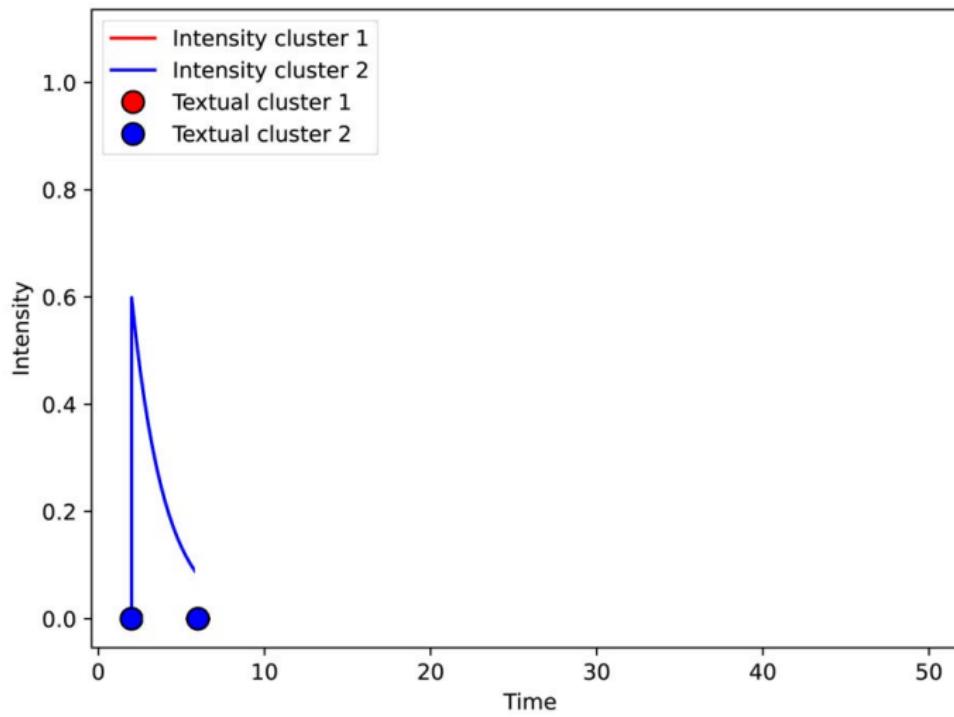
## Inference (1 particle)



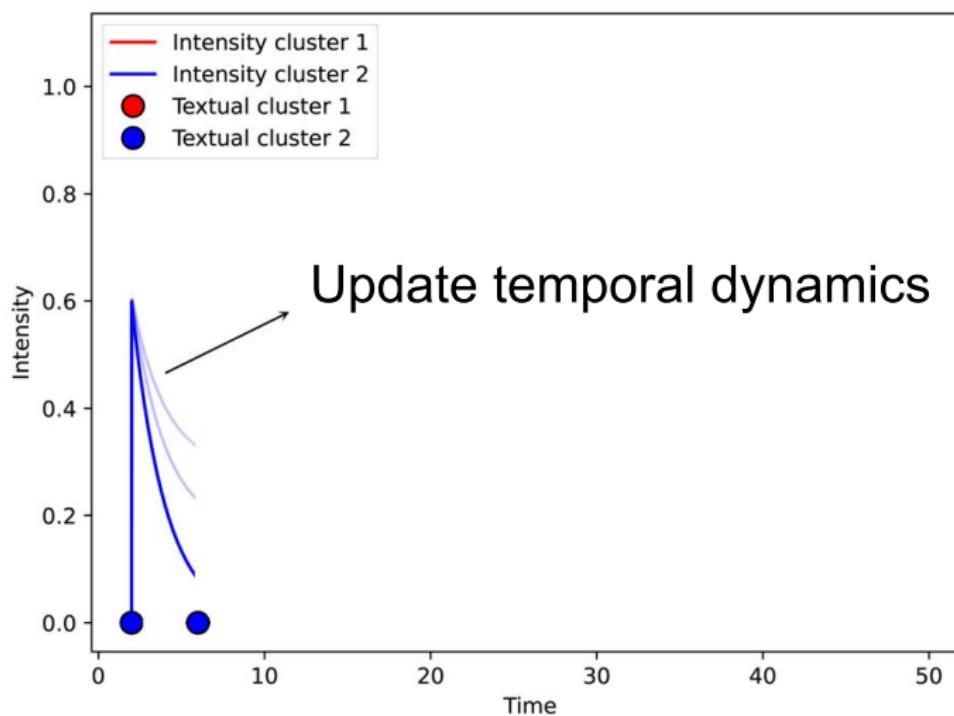
## Inference (1 particle)



## Inference (1 particle)

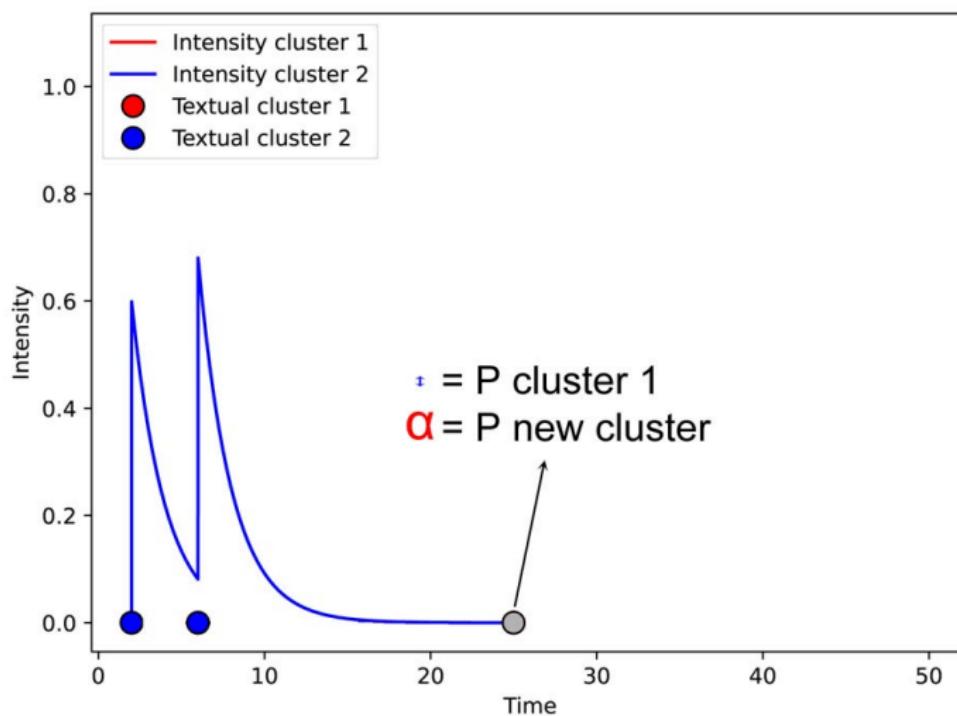


## Inference (1 particle)

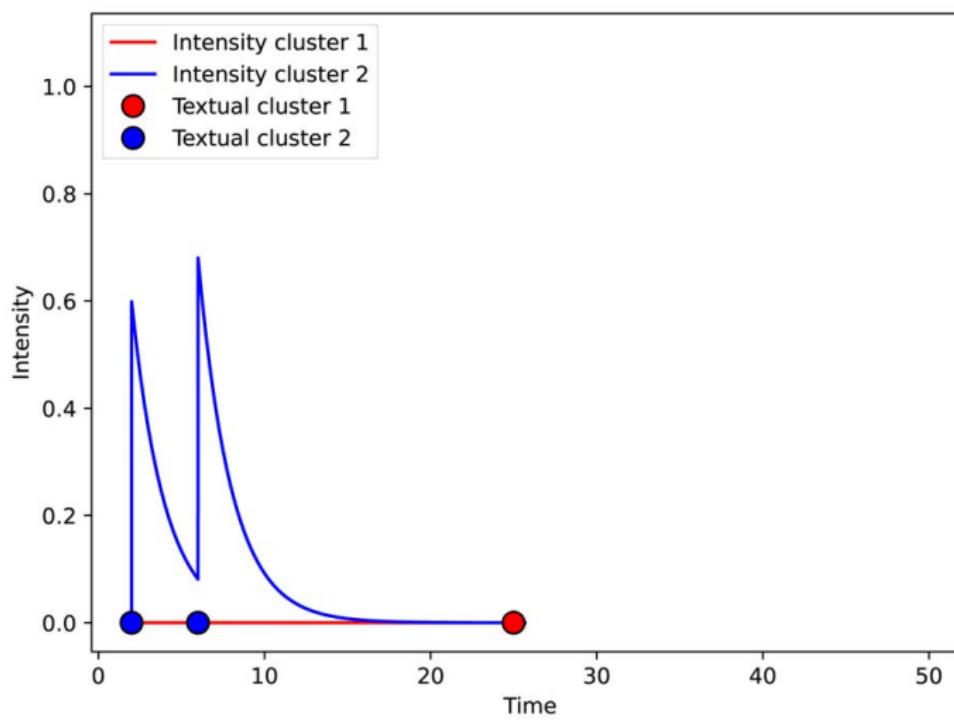


Motivation  
ooooDP  
ooooHP  
ooooooDHP  
○○●○○○○○○PDP  
ooooPDHP  
ooooooooooHouston  
ooooooooooooConclusion  
oo

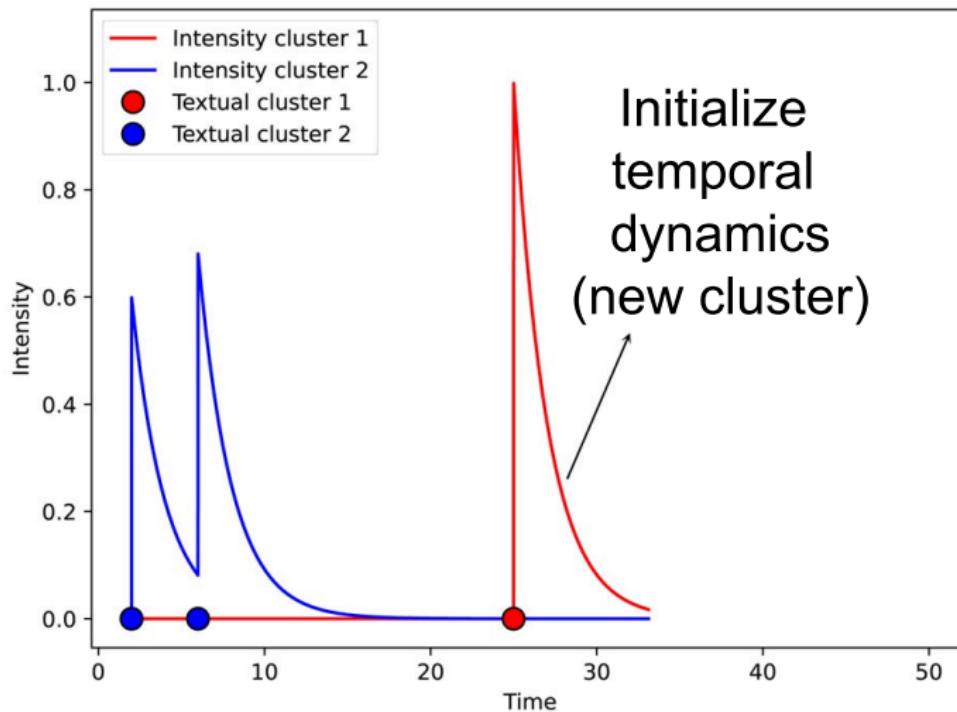
## Inference (1 particle)



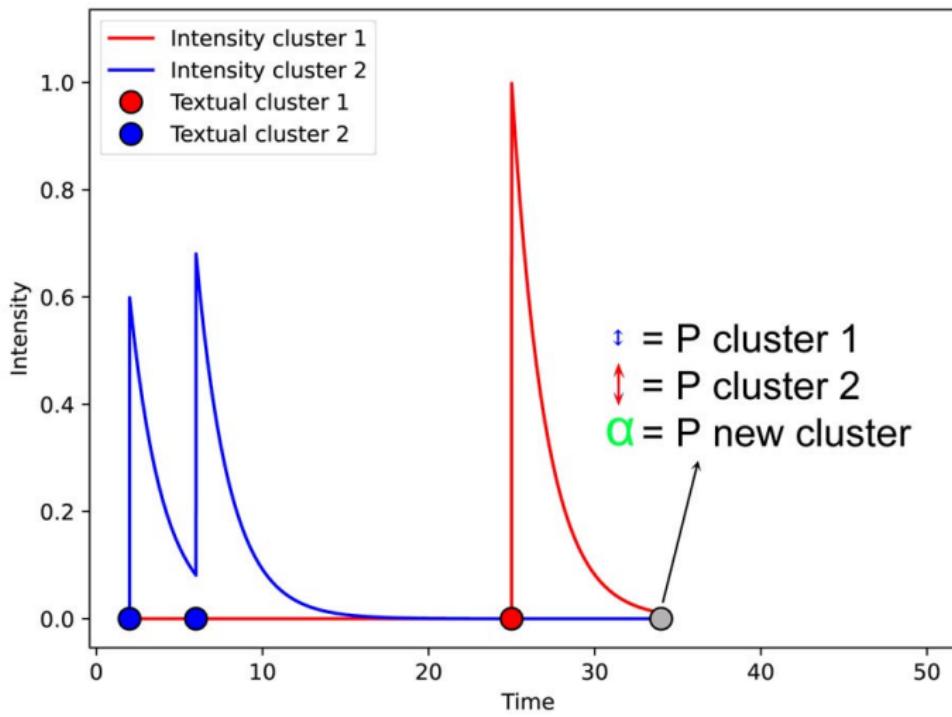
## Inference (1 particle)



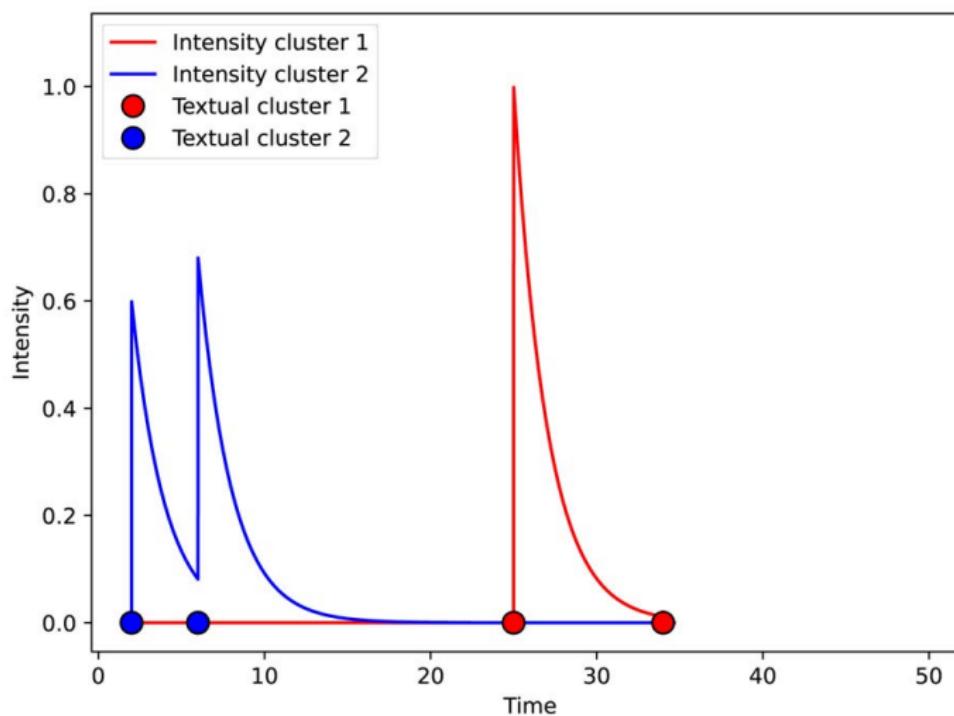
## Inference (1 particle)



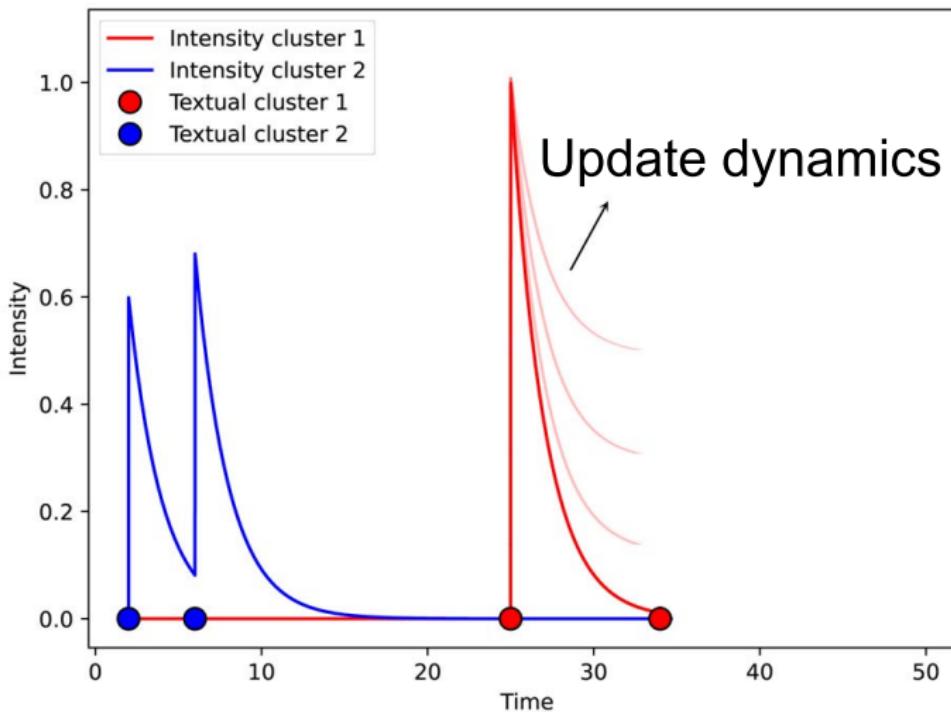
## Inference (1 particle)



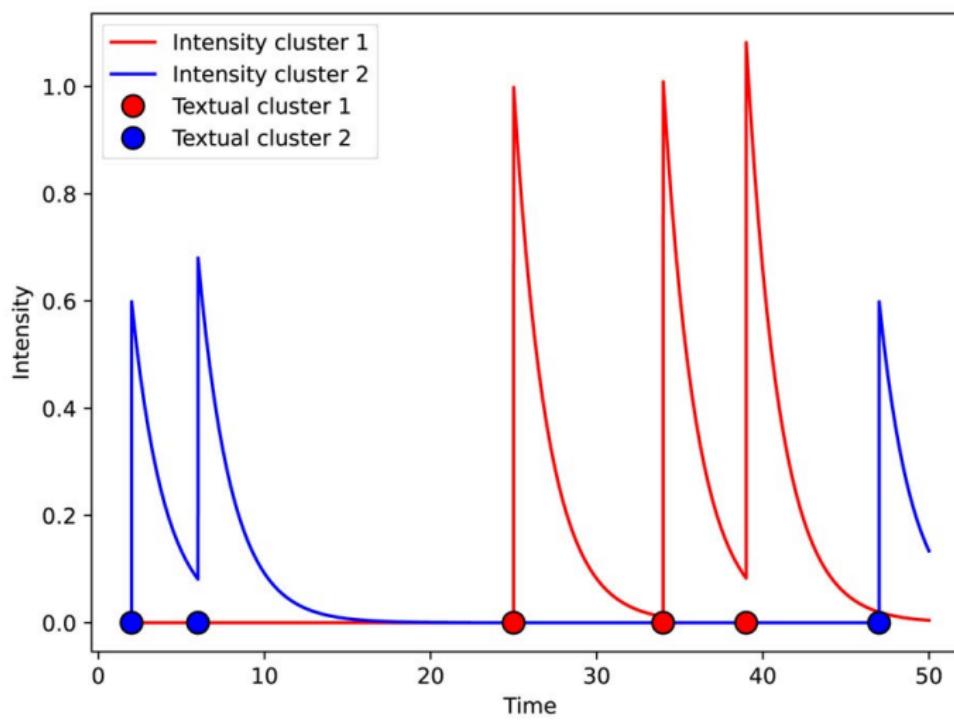
## Inference (1 particle)



## Inference (1 particle)

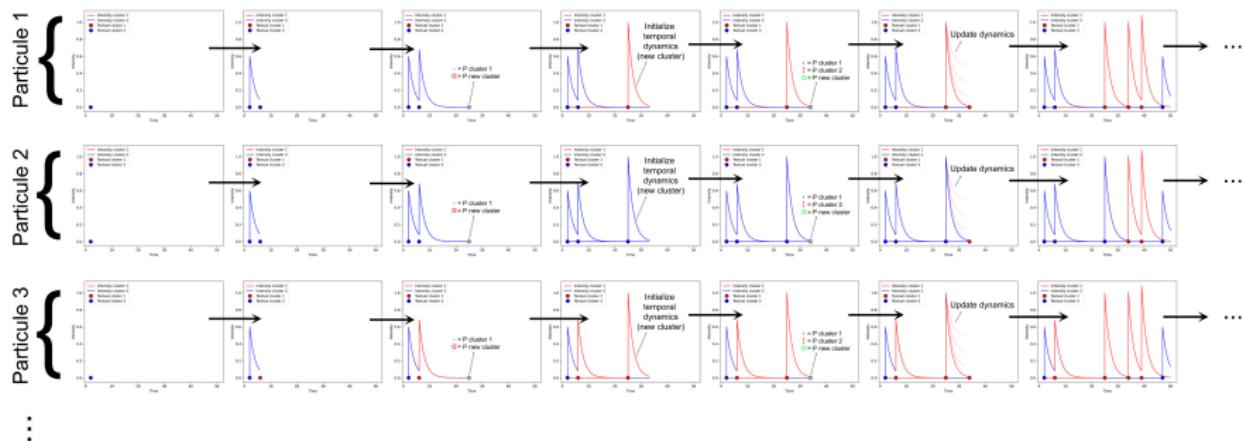


## Inference (1 particle)



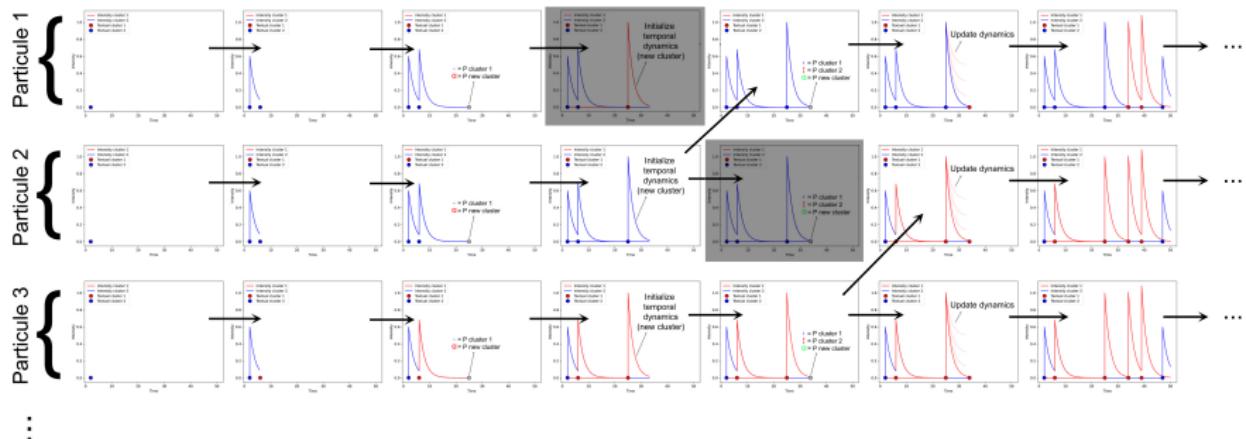
# Inference (all particles)

- Run simultaneously on several *particles*



# Inference (all particles)

- Discard unlikely particles and replace them by more likely ones

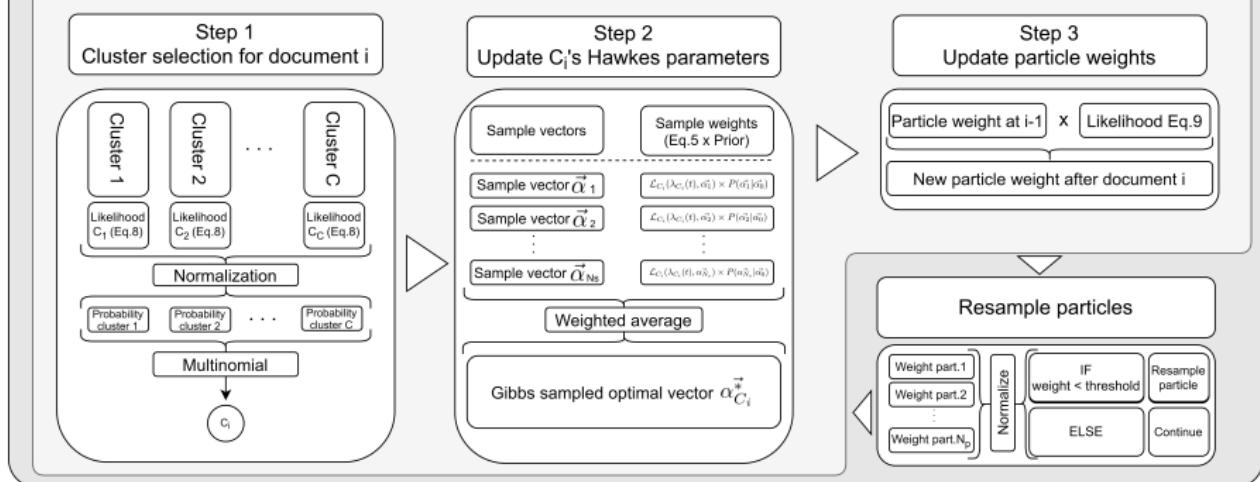


Motivation  
ooooDP  
ooooHP  
ooooooDHP  
oooo●ooooPDP  
ooooPDHP  
ooooooooHouston  
ooooooooooooConclusion  
oo

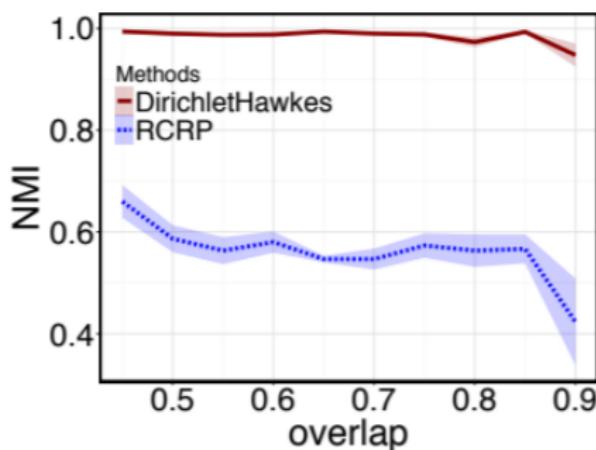
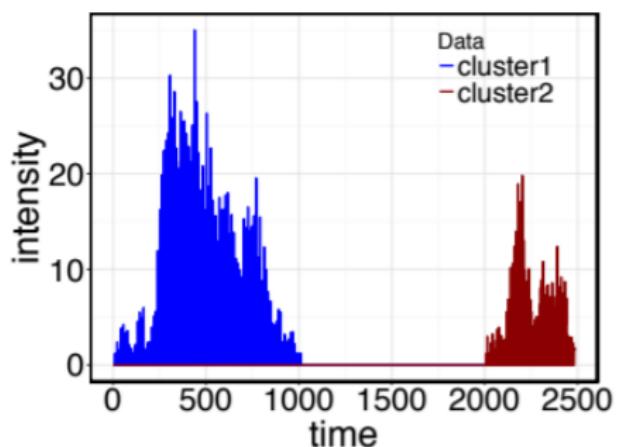
# Inference (summarized)

For each new document

For each particle



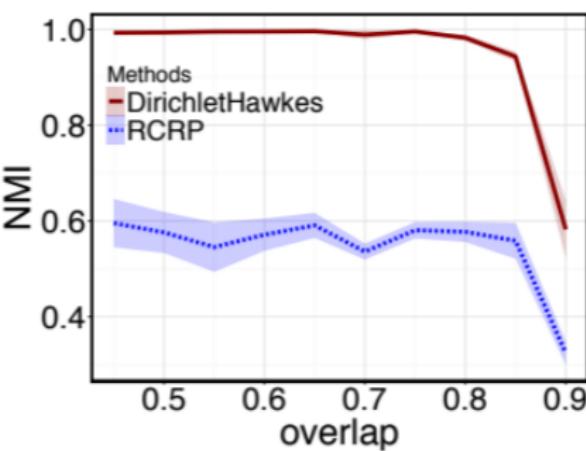
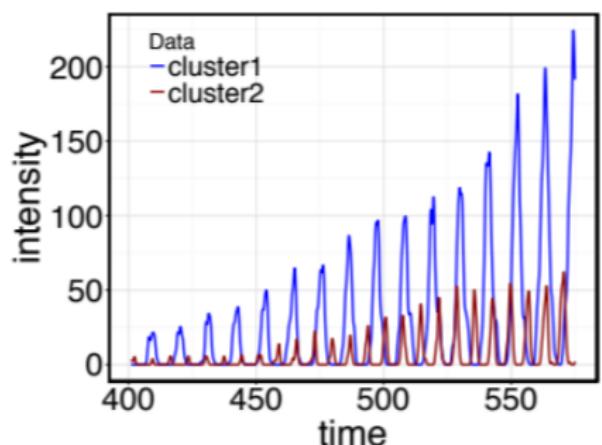
## Performances (well-separated)



(a) Temporally well-separated clusters.

Figure 10: [Du et al., 2015]

## Performances (“not” well-separated)



(b) Temporally interleaved clusters.

Figure 11: [Du et al., 2015]

## Output

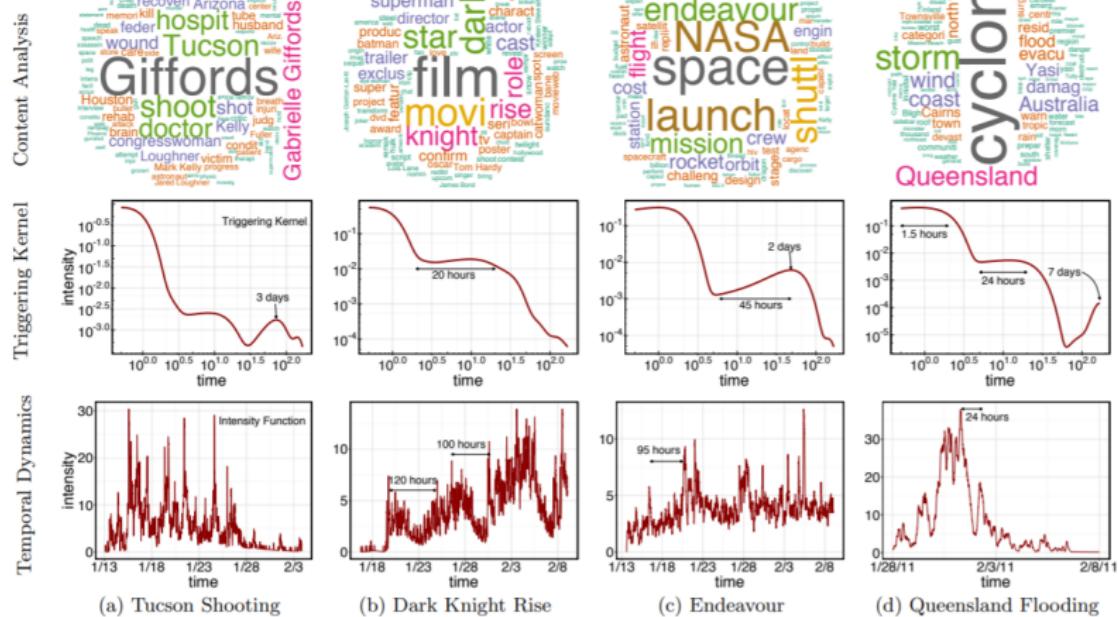


Figure 12: [Du et al., 2015]

## Variants

- Numerous variants based on Dirichlet-Hawkes process
  - Hierarchical (CRF) and Nested (nCRP) extensions of DHP
  - Multivariate DHP [Zheng et al., 2021]
  - Not-vanishing DHP prior [Kapoor et al., 2018]

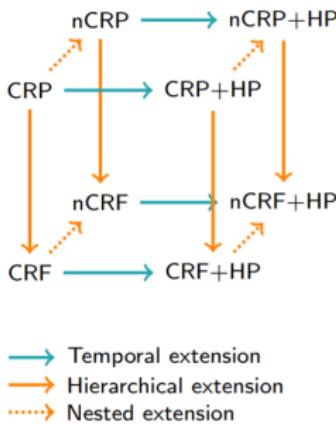


Figure 13: [Kapoor et al., 2018]

BUT!

# Dirichlet prior is a choice

- Dirichlet-based priors are an arbitrary choice
  - ◊ Other priors are as fit [Welling, 2006]
  - ◊ The choice of the prior matters [Wallach et al., 2009]
  - ◊ Few variations proposed [Wallach et al., 2010, Pitman and Yor, 1997]
- DP exhibits “rich-get-richer” property
  - ◊ Why linear dependence?
  - ◊ Why this assumption at all? [Wallach et al., 2010]

Motivation  
ooooDP  
ooooHP  
ooooooDHP  
ooooooooPDP  
○●○○PDHP  
ooooooooHouston  
ooooooooooooConclusion  
oo

# Powered Dirichlet process

- Powered Chinese Restaurant Process:

$$PCRP(C_i = c | C_1, \dots, C_{i-1}, \alpha, r) = \begin{cases} \frac{N_c^r}{\alpha + \sum_k N_k^r} & \text{if } c = 1, \dots, K \\ \frac{\alpha}{\alpha + \sum_k N_k^r} & \text{if } c = K+1 \end{cases}$$

- ◊  $r < 0$ : “rich-get-poorer”
- ◊  $r = 0$ : “rich-get-no-richer” (Uniform Process)
- ◊  $0 < r < 1$ : “rich-get-less-richer”
- ◊  $r = 1$ : “rich-get-richer” (Dirichlet Process)
- ◊  $r = \frac{\log(N_k - \beta)}{\log(N_k)}$ : “rich-get-richer” (Pitman-Yor Process)
- ◊  $r > 1$ : “rich-get-more-richer”

## PDP impact

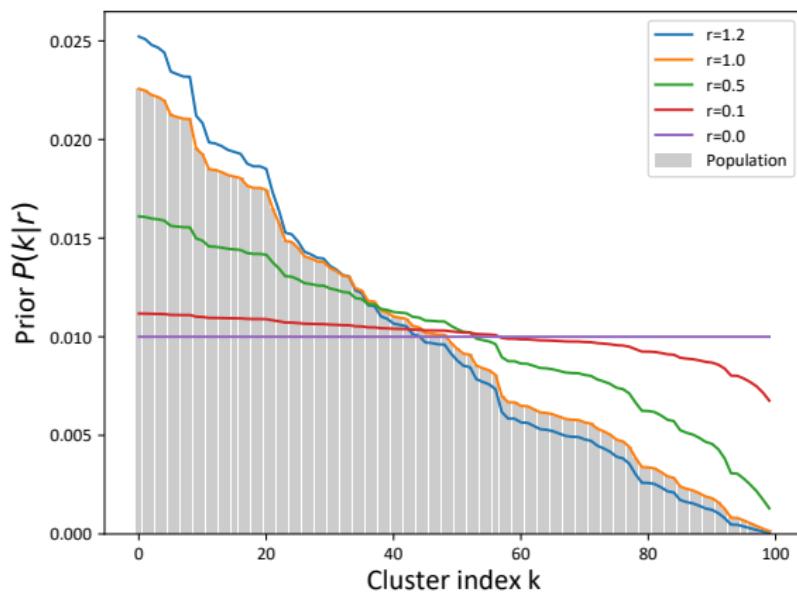
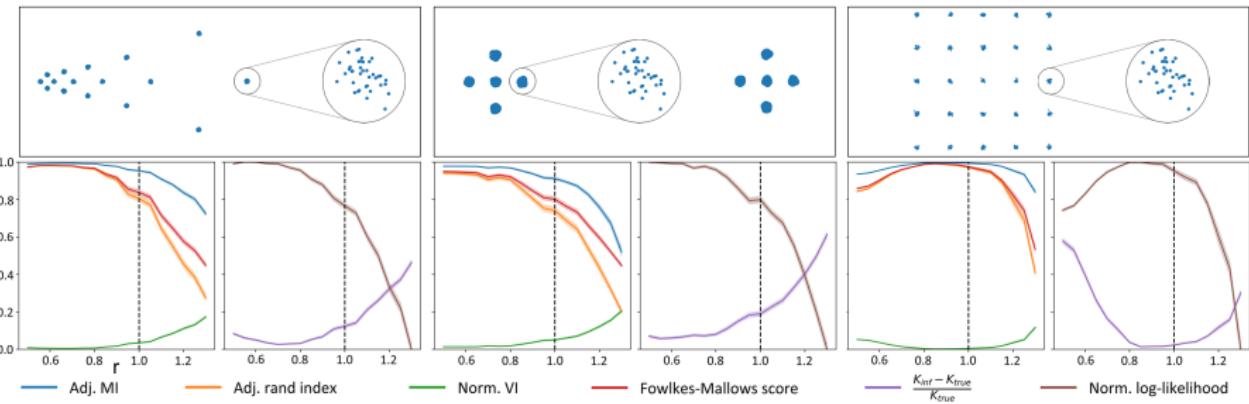


Figure 14: Prior probability for each of 100 clusters whose population is known (grey bars) w.r.t.  $r$

# Results

- Use as prior for IGMM
- DP not always the best prior



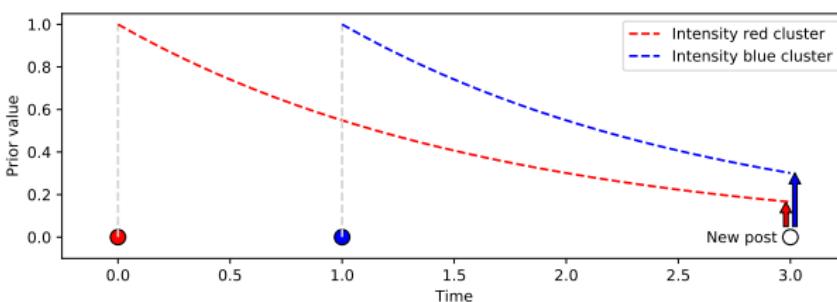
## PDP into DHP

- Powered priors: controlling the informativeness of the prior
  - ◊ PDP: strength of the “rich-get-richer” hypothesis
  - ◊ PDHP: strength of the temporal dependence hypothesis
- PDHP [Poux-Médard et al., 2021]:

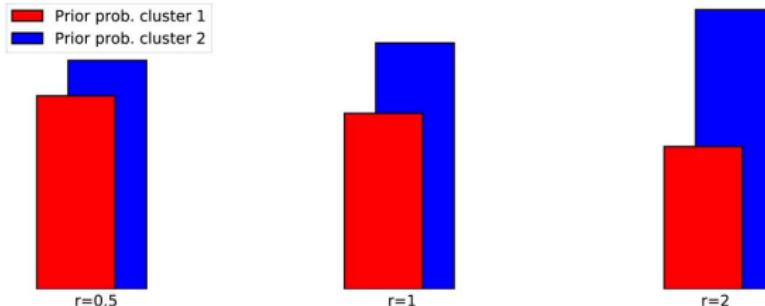
$$\underbrace{P(c|t, \mathcal{H}, \textcolor{red}{r})}_{\text{PDHP prior}} = \begin{cases} \frac{\lambda_c(t)^{\textcolor{red}{r}}}{\alpha_0 + \sum_k \lambda_k(t)^{\textcolor{red}{r}}} & \text{if } c = 1, \dots, K \\ \frac{\alpha_0}{\alpha_0 + \sum_k \lambda_k(t)^{\textcolor{red}{r}}} & \text{if } c = K+1 \end{cases}$$

- Generalization:
  - ◊ Uniform process:  $r = 0$  (only textual information)
  - ◊ Dirichlet-Hawkes process:  $r = 1$  (temporal and textual information)
  - ◊ Deterministic Hawkes process:  $r \rightarrow \infty$  (only temporal information)

## Effect of $r$



↓



Motivation  
ooooDP  
ooooHP  
ooooooDHP  
ooooooooPDP  
ooooPDHP  
○○●○○○○○○Houston  
ooooooooooooConclusion  
oo

## Changes induced by PDHP

$$P(\text{cluster}|\text{text}, \text{time}) \propto \underbrace{P(\text{text}|\text{cluster})}_{\text{Textual likelihood}} \times \underbrace{P(\text{cluster}|\text{time}, r, \text{history})}_{\text{PDHP temporal prior}}$$

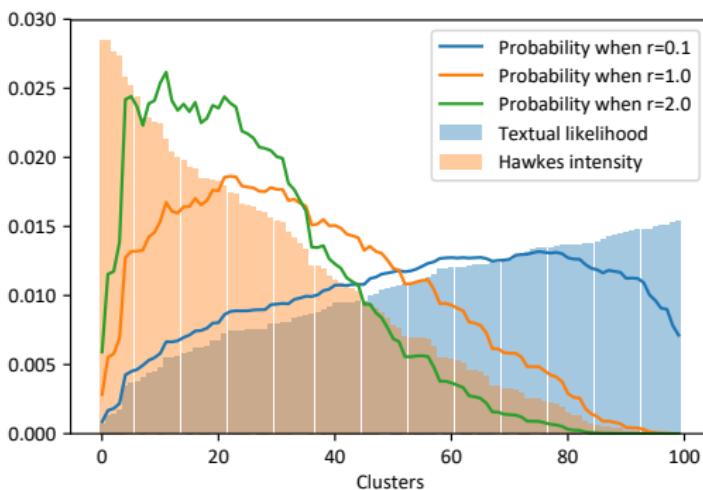


Figure 15: [Poux-Médard et al., 2021]

## Why is it relevant - Overlaps

- Often, a piece of information is more informative than the other:
  - Twitter: short texts (few textual information) but informative cascade dynamics (helpful temporal information)
- Happens often because of overlaps:

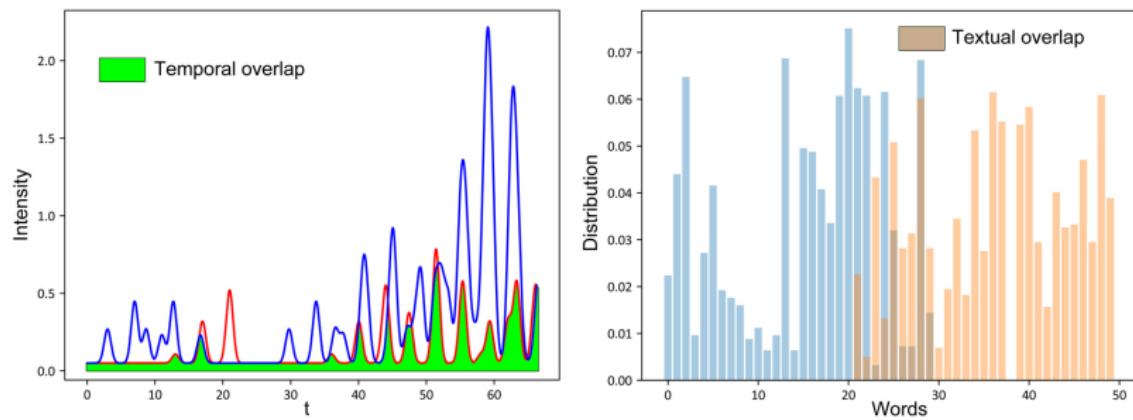
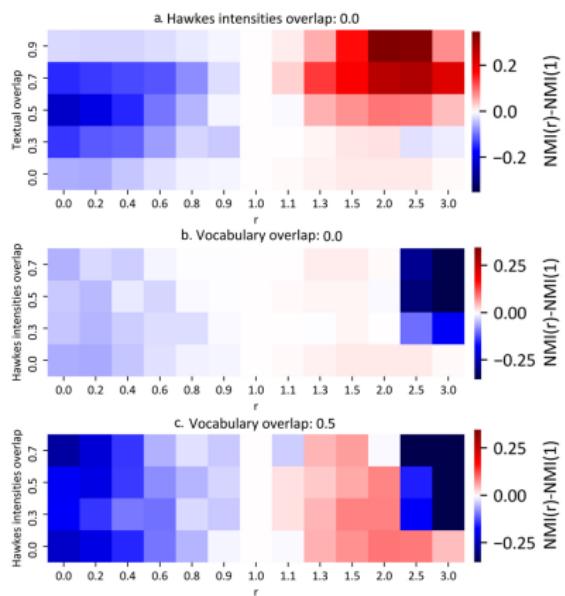


Figure 16: [Poux-Médard et al., 2021]

## Results for various overlaps



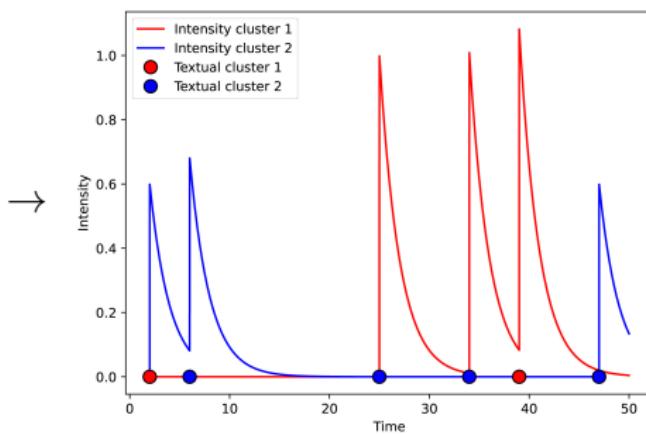
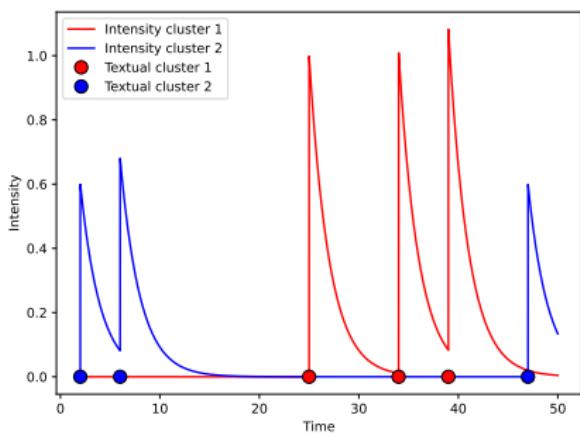
- PDHP adapts to various situations better than DHP:
  - ◊ Large textual overlap
  - ◊ Large temporal overlap
  - ◊ No overlap
- Up to +0.3 NMI in our case

Figure 17: [Poux-Médard et al., 2021]

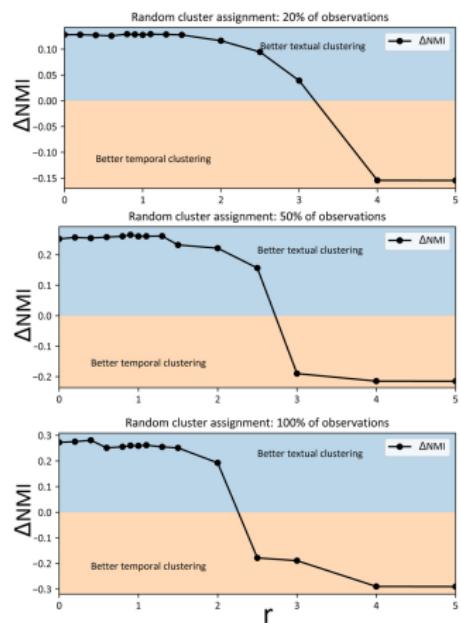
## Why is it relevant - Decorrelations

- Decorrelations:

- ◊ Ex: influent journal publishing on a topic does not have same dynamics as less influent one on the same topic



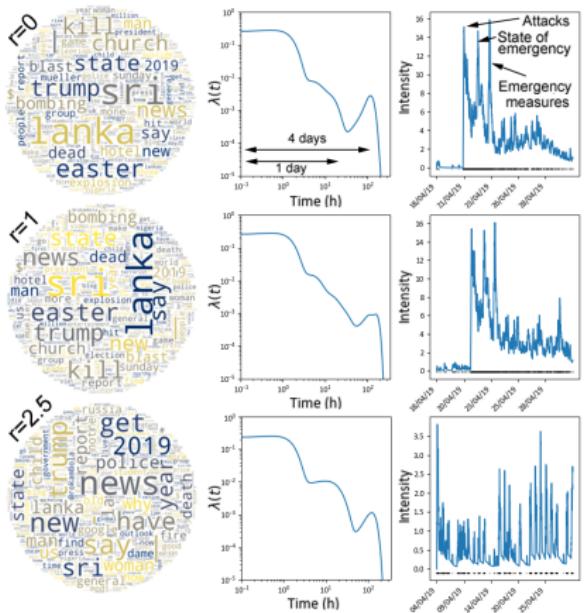
## Results for various decorrelations



- PDHP retrieves either temporal or textual clusters
  - ◊ Small  $r$ : good textual clusters
  - ◊ Large  $r$ : good temporal clusters

Figure 18: [Poux-Médard et al., 2021]

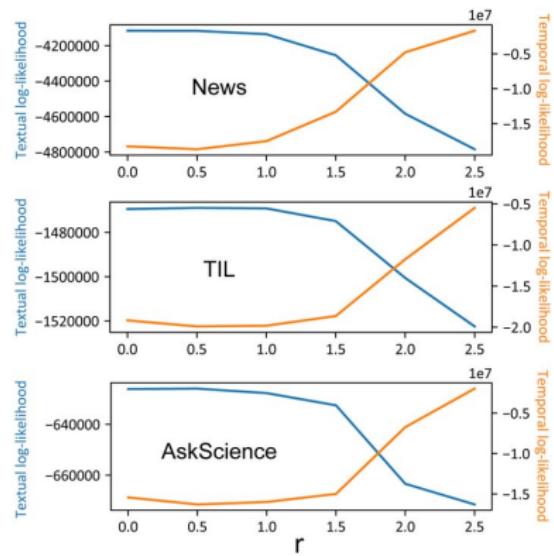
## Reddit r/news - Typical output



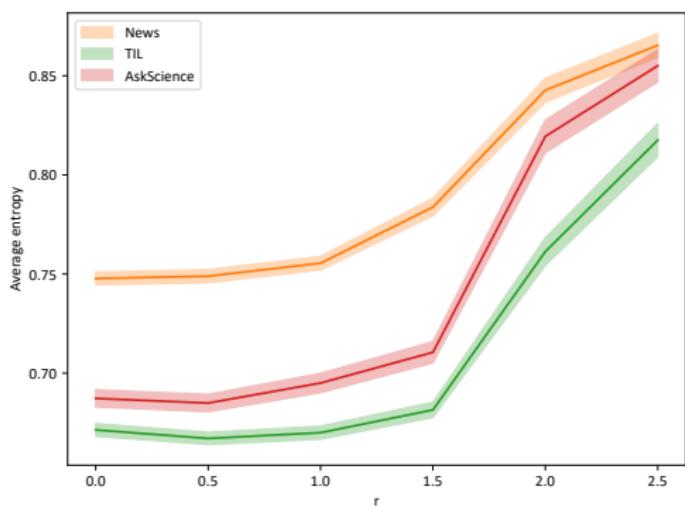
- Real world data: r/news
  - Different clusters and dynamics for different  $r$ 
    - ◊ Small  $r$ : similar vocabulary
    - ◊ Large  $r$ : specific dynamics

**Figure 19:** [Poux-Médard et al., 2021]

# Reddit r/news, r/TodayILearned, r/AskScience - Some metrics

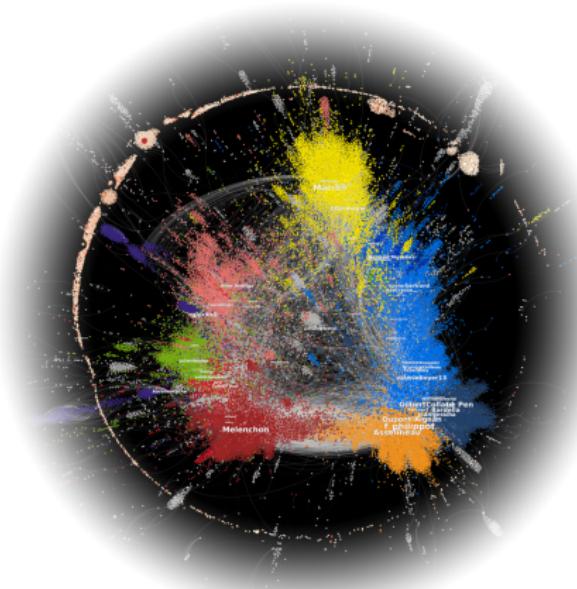


**Figure 20:** Textual and temporal likelihood vs  $r$   
[Poux-Médard et al., 2021]



**Figure 21:** Entropy of textual clusters:  
sharper textual clusters for low  $r$   
[Poux-Médard et al., 2021]

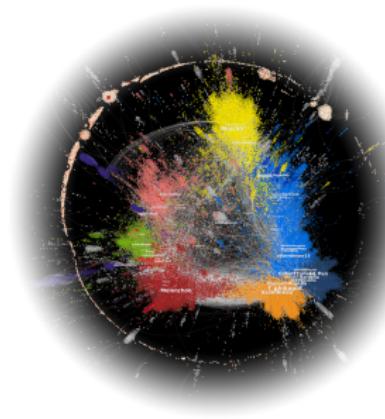
# Structure matters!



**Figure 22:** A sample from the Twitter structure (Politoscope [Gaumont et al., 2018])

## Why (P)DHP is incomplete

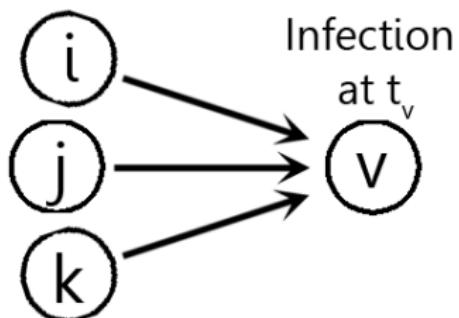
- DHP prior accounts for time but not structure
  - ◊ Infers aggregated dynamics
  - ◊ Misses the structural aspect: discussions are not the same among different groups



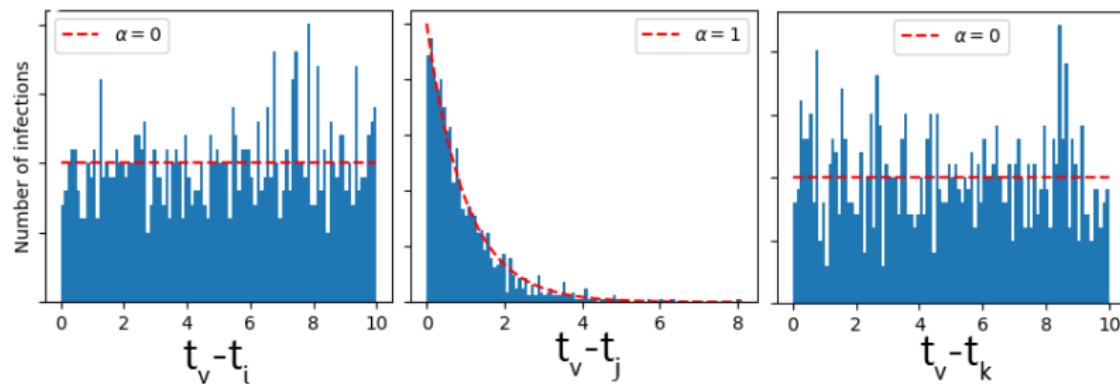
**Figure 23:** A sample from the Twitter structure (Politoscope [Gaumont et al., 2018])

Motivation  
ooooDP  
ooooHP  
ooooooDHP  
ooooooooPDP  
ooooPDHP  
ooooooooHouston  
ooo●ooooooConclusion  
oo

## Network inference



Exponential model  $P(t) = a \cdot e^{-\alpha t}$



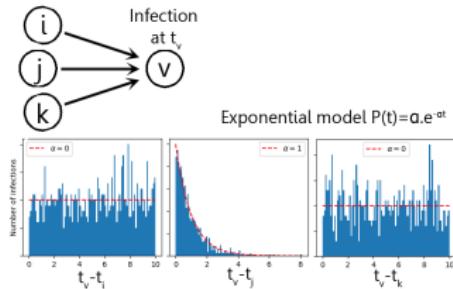
## Network inference – Literature

- Several works on network inference using survival analysis:
  - ◊ NetRate [Gomez-Rodriguez et al., 2011]
  - ◊ InfoPath [Gomez-Rodriguez et al., 2013a]
  - ◊ KernelCascade [Du et al., 2012]
  - ◊ MoNet [Wang et al., 2012]
  - ◊ TopicCascade [Du et al., 2013]
- They are all special cases of [Gomez-Rodriguez et al., 2013b]
  - ◊ Bridges the gap between survival analysis and point processes
  - ◊ Formulates each of previous models as a counting point process



# Point process

- Network inference naturally embeds into point processes literature
  - We can derive a temporal *and* structural Bayesian prior



Both are  
point  
processes  
 $\langle \approx \rangle$

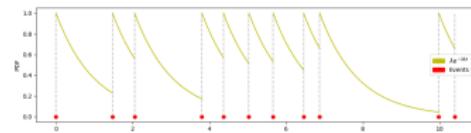


Figure 25: Hawkes process

Figure 24: Survival process

# Temporal and structural prior

- Houston: **Heterogeneous Online User-Topic Network** inference
- Prior on cluster membership  $C_i$  of observation  $i$  observed on node  $u$  at time  $t$  given history  $\mathcal{H}$  and cluster-dependent networks  $A$ :

$$P(C_i = k | u, t, \mathcal{H}, A)$$

$$= \begin{cases} \frac{\lambda_0^{(k)} + \sum_{\mathcal{H}_{i,c}^{(k)}} H(t_i^c | t_j^c, \alpha_{u_j^c, u_i^c}^{(k)})}{\lambda_0^{(K+1)} + \sum_k^K \lambda_0^{(k)} + \sum_{\mathcal{H}_{i,c}^{(k)}} H(t_i^c | t_j^c, \alpha_{u_j^c, u_i^c}^{(k)})} & \text{if } k = 1, \dots, K \\ \frac{\lambda_0^{(K+1)}}{\lambda_0^{(K+1)} + \sum_k^K \lambda_0^{(k)} + \sum_{\mathcal{H}_{i,c}^{(k)}} H(t_i^c | t_j^c, \alpha_{u_j^c, u_i^c}^{(k)})} & \text{if } k = K+1 \end{cases}$$

$$= \begin{cases} \frac{\text{Strength of incoming edges of cluster/subnetwork } k \text{ at time } t}{\text{Normalizing term}} & \text{if } k = 1, \dots, K \\ \frac{\text{Probability of a new cluster/subnetwork } k+1 \text{ at time } t}{\text{Normalizing term}} & \text{if } k = K+1 \end{cases}$$

Motivation  
oooo

DP  
ooo

HP  
oooo

DHP  
oooooooo

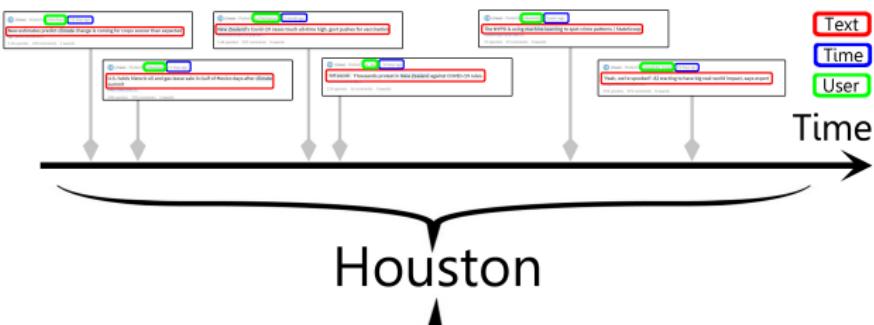
PDP  
ooo

PDHP  
oooooooo

Houston  
oooooo•ooo

Conclusion  
oo

# Task



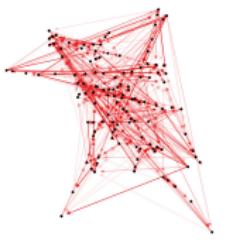
Cluster 1

• New COVID-19 cases change & testing to map areas that exceed social distancing levels.

• New students need to download older high school papers to access them.

• Millions of people are still working from home.

• Thousands arrived in São Paulo against COVID-19 rules.

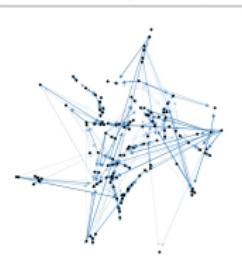


Cluster 2

• The WHO's using mobile learning to help children in Ethiopia learn English.

• New software allows users to track their progress.

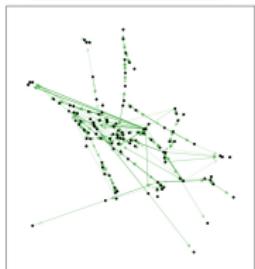
• India's new app is helping people learn English.



Cluster 3

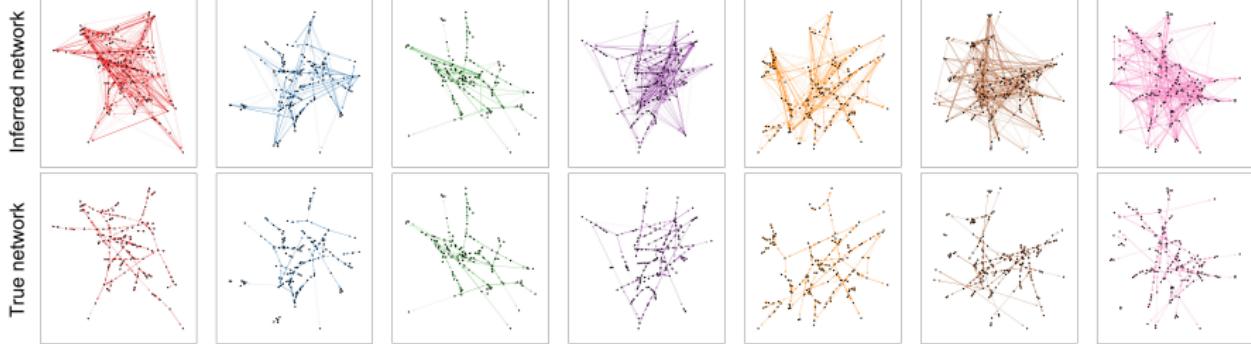
• U.S. adds millions of gas lease rate to Gulf of Mexico days after climate action.

• New software predicts climate change by looking at tree cover, then matched with satellite imagery.



## Results – Synthetic

- We simulate the spread of documents drawn from 5 topics, each with its own vocabulary and subnetwork



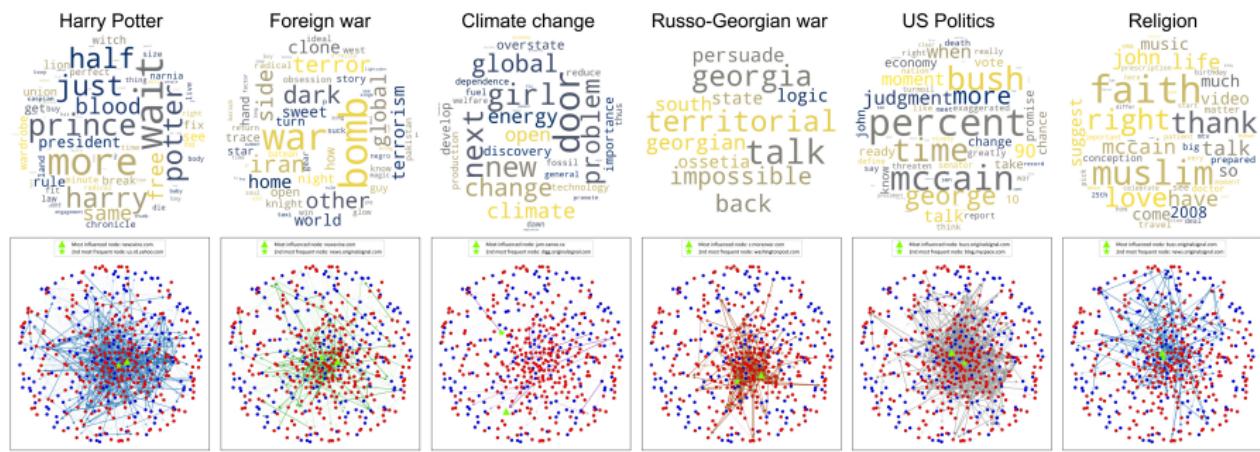
Motivation  
ooooDP  
ooooHP  
ooooooDHP  
ooooooooPDP  
ooooPDHP  
ooooooooHouston  
ooooooooooooConclusion  
oo

## Numerical results

		Houston	TC	DHP	NetRate
PL	NMI	<b>0.809</b>	0.669	0.449	-
	ARI	<b>0.688</b>	0.330	0.063	-
	AUC	<b>0.807</b>	0.719	-	0.731
	MAE	<b>0.267</b>	0.338	-	0.460
ER	NMI	<b>0.787</b>	0.711	0.638	-
	ARI	<b>0.631</b>	0.488	0.411	-
	AUC	<b>0.849</b>	0.800	-	0.659
	MAE	<b>0.229</b>	0.278	-	0.481
Blogs	NMI	<b>0.750</b>	0.668	0.372	-
	ARI	<b>0.609</b>	0.365	0.023	-
	AUC	<b>0.701</b>	0.613	-	0.710
	MAE	<b>0.374</b>	0.444	-	0.499

# Results – Real world

- Memetracker data (2009)



# Conclusion

- Dirichlet and Hawkes process have an old and separate history
  - ◊ Only recently (2015) they have been brought together
  - ◊ Their reunion launched a new branch of inductive machine learning
- The number of extensions based on Dirichlet-Point-Processes might be enormous, because we touched core concepts of machine learning
  - ◊ Dirichlet processes (PDP): could be used to redefine hierarchical DP, nested DP, or any models built on them (LDA, SBMs, among others)
  - ◊ Point processes (Poisson, Hawkes, Survival/Counting, etc.): the new possibility to merge them with DP could lead to a potentially infinite number of different Dirichlet-Point-Process priors.
- We presented 2 of such extensions:
  - ◊ PDP+HP → PDHP (flexible temporal prior)
  - ◊ DP+Survival → Houston (temporal+structural prior)

# Thanks for your attention!

(DP, HDP, nHDP, **PDP**, IBP, PIBP, PnHDP, PPY, PnPY, PHPY, ...)

×

(Hawkes, Survival, Cox, Poisson, Determinantal, Geometric, ...)

=

(DHP, HDHP, IBHP, **PDHP**, **Houston**, ...?)



## Bibliography I

[Ahmed and Xing, 2008] Ahmed, A. and Xing, E. (2008).

Dynamic non-parametric mixture models and the recurrent chinese restaurant process: with applications to evolutionary clustering.

In *SIAM International Conference on Data Mining*, pages 219–230.

[Blei and Frazier, 2010] Blei, D. M. and Frazier, P. (2010).

Distance dependent chinese restaurant processes.

In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 87–94, Madison, WI, USA. Omnipress.

[Blei and Lafferty, 2006] Blei, D. M. and Lafferty, J. D. (2006).

Dynamic topic models.

In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 113–120, New York, NY, USA. Association for Computing Machinery.



## Bibliography II

[Du et al., 2015] Du, N., Farajtabar, M., Ahmed, A., Smola, A., and Song, L. (2015).

Dirichlet-hawkes processes with applications to clustering continuous-time document streams.

*21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

[Du et al., 2012] Du, N., Song, L., Smola, A., and Yuan, M. (2012).

Learning networks of heterogeneous influence.

*NIPS*, 4:2780–2788.

[Du et al., 2013] Du, N., Song, L., Woo, H., and Zha, H. (2013).

Uncover topic-sensitive information diffusion networks.

In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS*, volume 31 of *JMLR Workshop and Conference Proceedings*, pages 229–237. JMLR.org.



## Bibliography III

[Gaumont et al., 2018] Gaumont, N., Panahi, M., and Chavalarias, D. (2018).

Reconstruction of the socio-semantic dynamics of political activist twitter networks—method and application to the 2017 french presidential election.

*PLOS ONE*, 13(9):1–38.

[Gomez-Rodriguez et al., 2011] Gomez-Rodriguez, M., Balduzzi, D., and Schölkopf, B. (2011).

Uncovering the temporal dynamics of diffusion networks.

In *ICML*, pages 561–568.

[Gomez-Rodriguez et al., 2013a] Gomez-Rodriguez, M., Leskovec, J., and Schoelkopf, B. (2013a).

Structure and dynamics of information pathways in online media.

*WSDM*.



## Bibliography IV

[Gomez-Rodriguez et al., 2013b] Gomez-Rodriguez, M., Leskovec, J., and Schölkopf, B. (2013b).

Modeling information propagation with survival theory.

In *ICML*, volume 28, pages III–666–III–674.

[Kapoor et al., 2018] Kapoor, J., Vergari, A., Valera, I., and Gomez-Rodriguez, M. (2018).

Bayesian nonparametric hawkes processes.

In *Proceedings of the Bayesian Nonparametrics workshop at the 32nd Conference on Neural Information Processing Systems (NIPS)*, NIPS workshops.

[Pitman and Yor, 1997] Pitman, J. and Yor, M. (1997).

The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator.

*The Annals of Probability*, 25(2):855 – 900.



## Bibliography V

[Poux-Médard et al., 2021] Poux-Médard, G., Velcin, J., and Loudcher, S. (2021).

Powered hawkes-dirichlet process: Challenging textual clustering using a flexible temporal prior.

*ICDM*.

[Rodríguez et al., 2008] Rodríguez, A., Dunson, D. B., and Gelfand, A. E. (2008).

The nested dirichlet process.

*Journal of the American Statistical Association*, 103(483):1131–1154.

[Teh et al., 2006] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006).

Hierarchical dirichlet processes.

*Journal of the American Statistical Association*, 101(476):1566–1581.

[Wallach et al., 2010] Wallach, H., Jensen, S., Dicker, L., and Heller, K. (2010).

An alternative prior process for nonparametric bayesian clustering.

In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 892–899. JMLR.



## Bibliography VI

- [Wallach et al., 2009] Wallach, H., Mimno, D., and McCallum, A. (2009).  
Rethinking Ida: Why priors matter.  
In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc.
- [Wang et al., 2012] Wang, L., Ermon, S., and Hopcroft, J. E. (2012).  
Feature-enhanced probabilistic models for diffusion network inference.  
In *Machine Learning and Knowledge Discovery in Databases*, pages 499–514, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Wang and McCallum, 2006] Wang, X. and McCallum, A. (2006).  
Topics over time: A non-markov continuous-time model of topical trends.  
In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 424–433, New York, NY, USA. Association for Computing Machinery.



## Bibliography VII

[Welling, 2006] Welling, M. (2006).

Flexible priors for infinite mixture models.

In *Workshop on learning with non-parametric Bayesian methods*.

[Zheng et al., 2021] Zheng, P., Yuan, S., and Wu, X. (2021).

Using dirichlet marked hawkes processes for insider threat detection.

*Digital Threats: Research and Practice*, 3(1).