Point processes

ft. Processus de Hawkes-Dirichlet

Gaël Poux-Médard

20/10/2020

Sommaire

Introduction

Motivations

- Modéliser des processus aléatoires dans des espaces continus : temps, positions.
- Exemples : probabilité d'un retweet à un temps donné, popularité d'une app, mouvements du marché financier, influence d'événements externes sur un comportement, etc.

Données disponibles

- Un historique des événements survenus à un temps t.
  - Un article sur Trump a été publié au temps t_1, un autre sur Biden à t_2, etc.
  - Noël est arrivé à t_1, une Wii a été achetée à t_2 et t_3, une PS5 à t_4 et t_5, etc.

Idée

- La présence d'un événement au temps t correspond à la réalisation d'un processus stochastique défini par une distribution de probabilités.

Sommaire

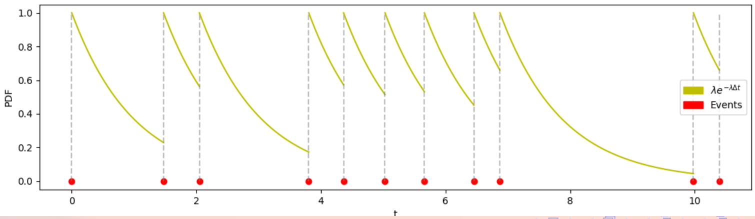
Poisson process

Définition

- Un processus de Poisson est un counting process paramétré par une **intensité** λ.
  - $P(\mathbb{N}(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$ est la probabilité que n événements arrivent en un temps t.
- PDF de la date d'arrivée d'un seul événement :

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

- Propriétés : $\mathbb{E}(f(t)) = \frac{1}{\lambda}, \sigma^2(f(t)) = \frac{1}{\lambda^2}$



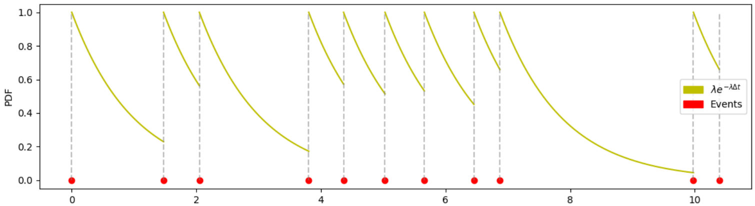
Processus sans mémoire

Memorylessness

- On reprend la probabilité qu'un événement se produise :

$$f(t) = \lambda e^{-\lambda t}$$

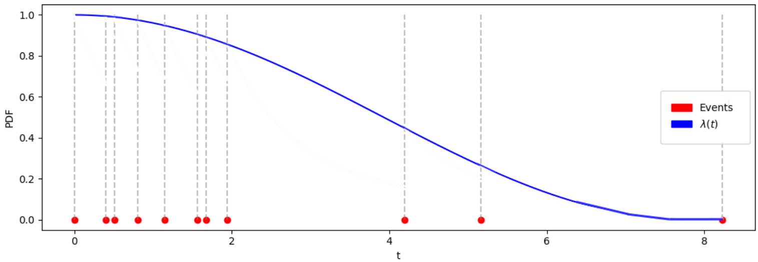
- $f(t > \tau) = \int_{\tau}^{\infty} \lambda e^{-\lambda t} = e^{-\lambda \tau}$
- Si on a déjà attendu un temps $\Delta\tau$, on calcule :
  - $f(t > \tau + \Delta\tau | t > \Delta\tau) = \dfrac{f(t > \tau + \Delta\tau, t > \Delta\tau)}{f(t > \Delta\tau)} \overset{t \geq 0}{\equiv} \dfrac{f(t > \tau + \Delta\tau)}{f(t > \Delta\tau)} = e^{-\lambda\tau}$ $\square$
- Implique aussi $P(N(t + \Delta t) - N(t) = n) = P(N(\Delta t) = n)$



Processus de Poisson non-homogène

Intensité

- La paramètre λ est appelé l'intensité du processus.
- $\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t+\Delta t) - N(t) = 1)}{\Delta t}$
- Dans les exemples précédents $\lambda(t) = \lambda$.
- $\lambda(t)$ permet une modélisation plus fine (ex: traffic de voitures)

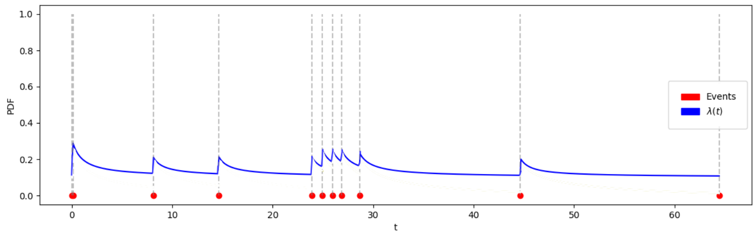


Processus de Hawkes

Intensité

- On peut définir un processus stochastique pour n'importe quelle intensité $\lambda(t)$
- Les processus de Hawkes incorporent une mémoire à l'intensité. On définit l'ensemble $\mathcal{H}_t$ contenant tous les événements qui sont arrivés avant t.
- Typiquement : $\lambda(t) = \lambda_0 + \sum_{t_i \in \mathcal{H}_t} \phi(t - t_i)$

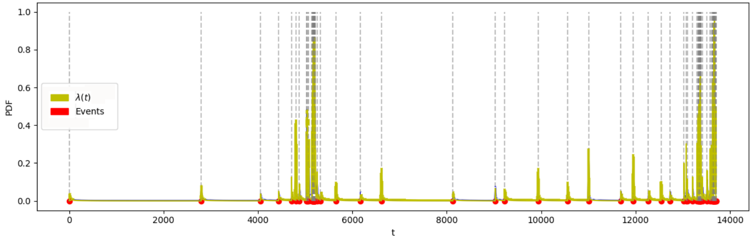
Figure 1: Ici $\lambda(t) = \lambda_0 + \sum_{t_i \in max(\mathcal{H}_t, 2)} \frac{1}{t - t_i}$



Processus de Hawkes

Bursts

- Choix de modélisation intéressant pour les processus réels:
  - Intuitivement correct
  - Reproduit des phénomènes de burst



Processus de Hawkes multivariés

Intensité

- Processus de Hawkes : $\lambda(t) = \lambda_0 + \sum_{t_i \in \mathcal{H}_t} \phi(t - t_i)$
- Processus de Hawkes multivarié : $\lambda(t) = \lambda_0 + \sum_{t_i^{(m)} \in \mathcal{H}_t} \phi^{(m \to n)}(t - t_i^{(m)})$

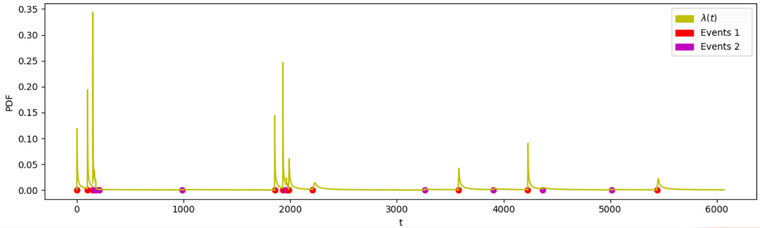


Figure 2: Realisation avec 2 types d'événements différents ; les événements rouges ont plus d'influence que les violets ($\phi^{rouge}(t) > \phi^{violet}(t)$).

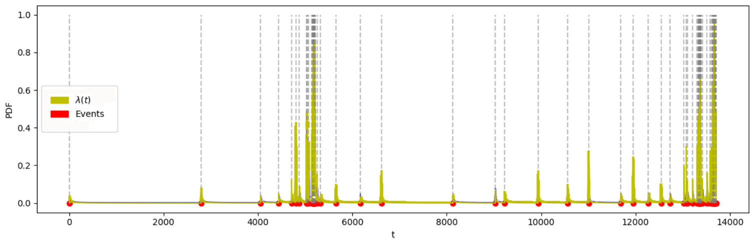
Optimisation

Likelihood

- Les processus stochastiques sont bien connus dans la littérature.
- Likelihood d'observations entre t=0 et t=T :

$$\mathcal{L}(\lambda) = e^{-\int_0^T \lambda(t)dt} \prod_{t_i<T} \lambda(t_i)$$

- Complexité $|T|^2$, peut être réduite à $|T|$ pour certaines formes de $\lambda(t)$ (combinaisons linéaires d'exponentielles, gaussiennes, power-law, ...).



Sommaire

Memetracker

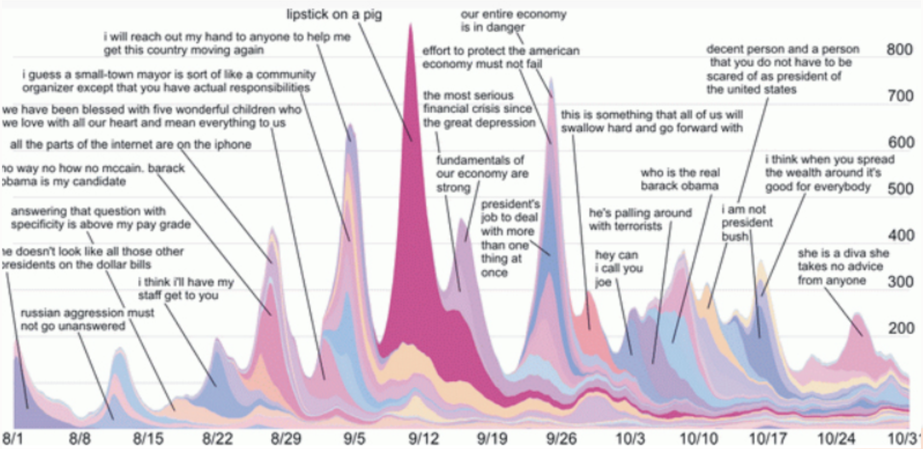
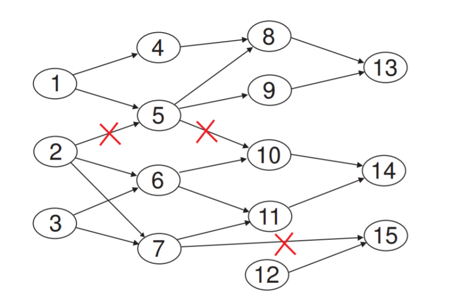


Figure 3: Histogramme de la publication de memes dans le temps par des sites de News. Taille : 1 million de sites d'information considérés.

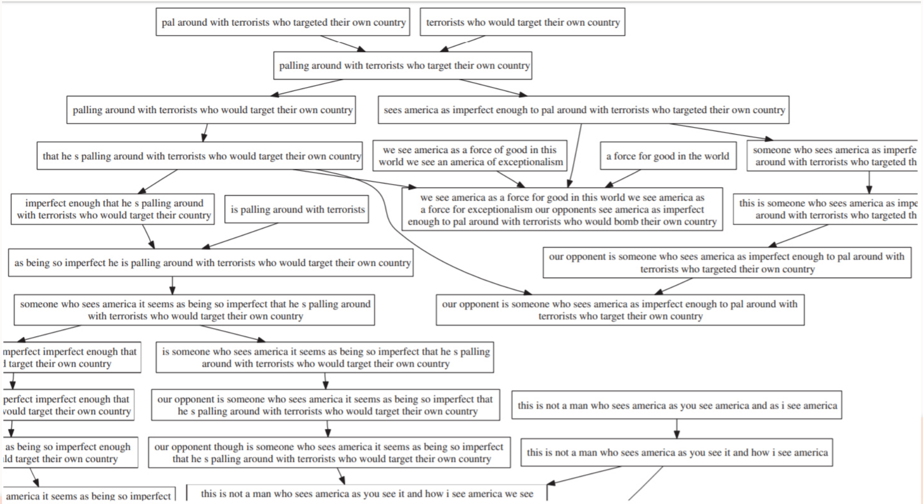
Memetracker

Création du dataset

- Graph de phrases où les liens sont dirigés des phrases les plus courtes (parent) aux plus longues (enfant).
- Pondéré positivement par la fréquence du noeud enfant, et negativement par le temps séparant les publications.
- Partition DAG : supprimer le poids minimal en liens pour que le graphe n'ait qu'une feuille.



Memetracker



Interaction information-information -très- sparse

Les interactions entre informations est sparse

- Peu de clusters d'informations intéragissent les uns avec les autres.

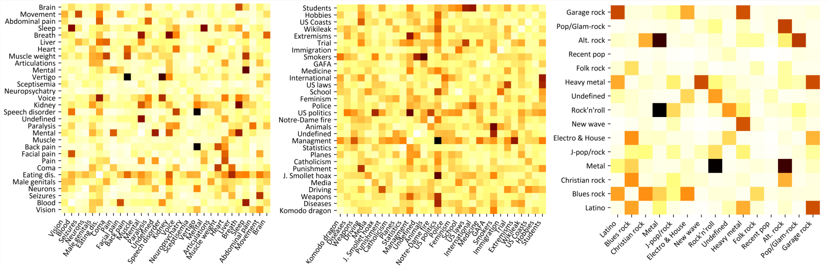


Figure 4: IMMSBM, soumis à ~~ECML-PKDD~~ ~~WSDM~~ ~~WWW~~ RecSys

Interaction information-information -très- sparse

Les interactions entre informations est courte

- Lorsque deux informations intéragissent, elles le font sur une fenêtre de temps très courte.
- De plus, l'information elle-même a une durée de vie très limitée :
  - Twitter : 18 minutes
  - Facebook : 5 heures
  - Instagram : 21 heures
  - YouTube : 20 jours

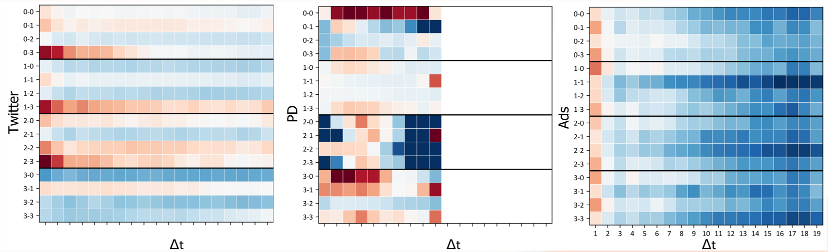


Figure 5: Papier ~~passé en première phase à AAAI-21~~ soumis à ~~KDD~~ desk reject, mauvais template ECML-PKDD

Sommaire

Processus de Hawkes-Dirichlet (HDP)

Solution

- On a donc impérativement besoin de clustering pour espérer capturer des interactions entre informations sur des temps longs.
- Processus de Hawkes-Dirichlet : processus de Hawkes avec clustering :
  - Si deux informations intéragissent de manière similaire avec une autre, elles feront partie d'un cluster.
  - On aboutit à un processus de Hawkes avec des clusters d'information en déclenchant d'autres.

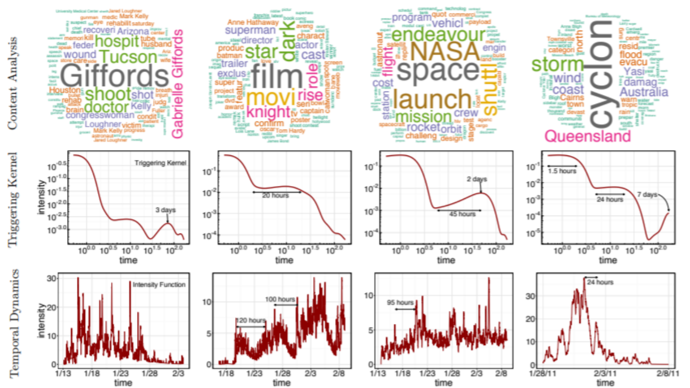
Petit état de l'art

- Les processus de Hawkes-Dirichlet sont récents et peu explorés dans la littérature.
- Liste exhaustive (d'après les trois premières pages Google en tapant "Hawkes-Dirichlet process") :
  - **KDD 2015**, *Hawkes-Dirichlet Processes with Applications to Clustering Continuous-Time Document Streams*, Nan Du, Mehrdad Farajtabar & al.
  - **WWW 2017**, *Modeling the Dynamics of Online Learning Activity*, Charalampos Mavroforakis, Isabel Valera, Manuel Gomez-Rodriguez
  - **NIPS 2017**, *A Dirichlet Mixture Model of Hawkes Processes for Event Sequence Clustering*, Hongteng Xu, Hongyuan Zha

Processus de Hawkes-Dirichlet (HDP)

Solution

- **KDD 2015** : adaptative hard clustering + inference kernel, Monte-Carlo
- **WWW 2017** : fixed mixed clustering, Monte-Carlo (variation de KDD 2015)
- **NIPS 2017** : adaptative mixed clustering + inference kernel, EM



(a) Tucson Shooting (b) Dark Knight Rise (c) Endeavour (d) Queensland Flooding

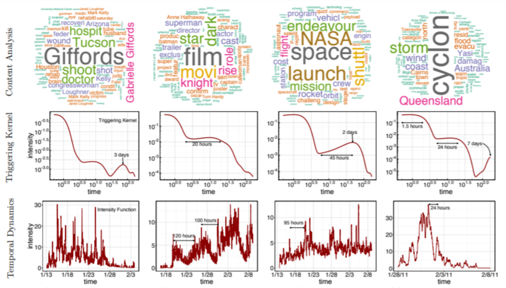
Processus de Hawkes-Dirichlet (HDP)

Du 2015

- **KDD 2015**, *Hawkes-Dirichlet Processes with Applications to Clustering Continuous-Time Document Streams*, Nan Du, Mehrdad Farajtabar & al.
- Likelihood définie comme :

$$L(cluster|contenu\ txt, date) \propto \underbrace{L(cluster|contenu\ txt)}_{\text{Modèle de langue}} \times \underbrace{L(cluster|date)}_{\text{Prior temporel (HDP)}}$$

- Optimisé via Sequential Monte Carlo



Processus de Hawkes-Dirichlet (HDP)

Dirichlet prior

- Un prior de Dirichlet donne la likelihood d'un ensemble de comptes s'ils émergent d'un processus "rich-get-richer" :
- $P(c|\vec{N}) \approx \frac{N_c}{\sum_k N_k}$

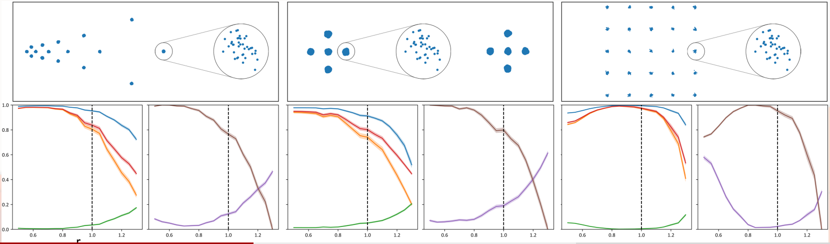
Processus HPD : parlons-en

- Remplacer les N_k par un "compte" qui dépend du temps et des clusters : la fonction d'intensité $\lambda(t)$
- $P(c|\vec{\lambda}(t)) \approx \frac{\lambda_c(t)}{\sum_k \lambda_k(t)}$
- Ca donne un prior temporel sur les clusters.

Processus de Hawkes-Dirichlet (HDP)

Dirichlet prior : pourquoi

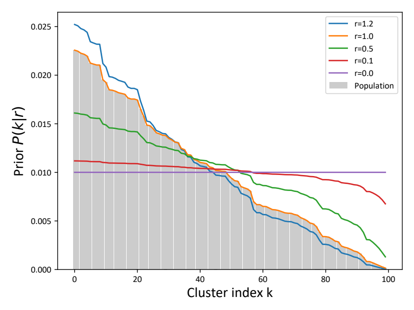
- Le prior de Dirichlet n'est pas toujours un très bon prior ($P(c|\vec{N}) \approx \frac{N_c}{\sum_k N_k}$):
  - Pourquoi dépendance linéaire ?
  - Pourquoi dépendance tout court ?
- Variations :
  - **Uniform process** : supprime la dépendance vàv des populations $P(c|\vec{N}) \approx \frac{1}{\sum_k 1}$
  - **Powered Chinese Restaurant process** (ça c'est moi, soumis à ECML-PKDD) : généralise Uniform process, Dirichlet-Process $P(c|\vec{N}, r) \approx \frac{N_c^r}{\sum_k N_k^r}$



Processus de Hawkes-Dirichlet Puissancé (PHDP)

Powered Hawkes-Dirichlet prior

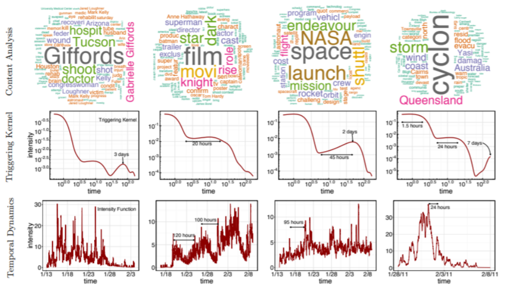
- $$L(cluster\ c|contenu\ txt, date) \propto \underbrace{L(c|contenu\ txt)}_{\text{Modèle de langue}} \times \underbrace{L(c|date)}_{\text{Powered HDP prior}}$$
  $$\propto L(c|contenu\ txt) \times \frac{\lambda_c^r(date)}{\sum_k \lambda_k^r(date)}$$



Processus de Hawkes-Dirichlet Puissancé (PHDP)

Powered Hawkes-Dirichlet prior : pourquoi ?

- Si les clusters textuels sont différents des clusters temporels ?
  - Un article de journal aura une influence temporelle sur des publications ultérieures due à la notoriété du journal (par exemple) et à son contenu
  - Un journal hebdomadaire publiera toutes les semaines quel que soit son contenu $\rightarrow$ Journaux hebdomadaires : cluster, pas cluster ?
  - Quel clustering choisir ? Inférer l'influence temporelle en fonction du journal éditeur ou autres informations jouant un rôle temporel (limite r $\gg$ 0) ou du contenu (limite r = 0) ?
- Dans Nan Du : $r = 1$, un choix parmi tant d'autres.



(a) Tucson Shooting (b) Dark Knight Rise (c) Endeavour (d) Queensland Flooding

Extensions de HDP

Powered Hawkes-Dirichlet prior

- $$L(cluster\ c|contenu\ txt, date) \propto \underbrace{L(c|contenu\ txt)} \times \frac{\lambda_c^r(date)}{\sum_k \lambda_k^r(date)}$$

- Dans N.Du 2015, ils utilisent un modèle de langue simple (bag of words / recurrent chinese restaurant process), je suis preneur d'un modèle plus abouti qui fonctionne dans un SMC (nouveau document : calcul de la likelihood pour chaque cluster + mise à jour après allocation)

Multivariate Powered Hawkes-Dirichlet prior

- Plutôt que de juste inférer $\lambda_i(t)$, inférer $\lambda_{i,j}(t)$
- Les différents clusters ont une influence l'un sur l'autre ; on ne décrit plus uniquement la dynamique d'un cluster isolé.

Application : Memetracker

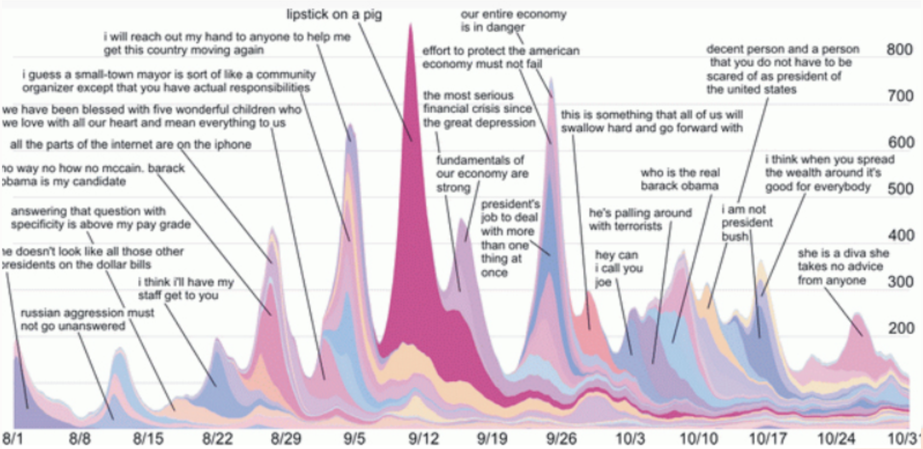


Figure 6: Histogramme de la publication de memes dans le temps par des sites de News. Taille : 1 million de sites d’information considérés.

Sommaire

Conclusion

Remarques

- Dans le cas de la modélisation de l'interaction entre information, le clustering est nécessaire.
- Les processus de Hawkes-Dirichlet permettent de créer des clusters textuo-temporels en temps linéaire (via SMC)
- Sujet récent dans la littérature et relativement peu exploré
- Résultats potentiellement intéressants à analyser

Merci

Merci

- Merci de votre attention !