



Acfas

# 89<sup>E</sup> CONGRÈS



SCIENCES —  
INNOVATIONS  
— SOCIÉTÉS

9 — 13 MAI 2022  
UNIVERSITÉ LAVAL





# *Anomalies dans les réseaux*

Benoît Corsini

En collaboration avec Pierre-André Noël, David Vázquez, et Perouz Taslakian.



- **Introduction**
- **Méthode**
- **Résultats**



- **Introduction**
- Méthode
- Résultats



## ***Le contexte***

- Nous avons un ensemble d'éléments connectés les uns aux autres :
  - Utilisateurs d'un réseau social.
  - Emails entre collègues.
  - Collaborations entre chercheurs.
- Ces éléments possèdent des attributs :
  - Information sur l'utilisateur.
  - Contenu de l'email.
  - Sujet de recherche et titre de l'article.
- Peut-on trouver les anomalies qui habitent cet ensemble :
  - Bot ou influenceurs.
  - Spam ou tentative de phishing.
  - Articles influents ou plagiat.

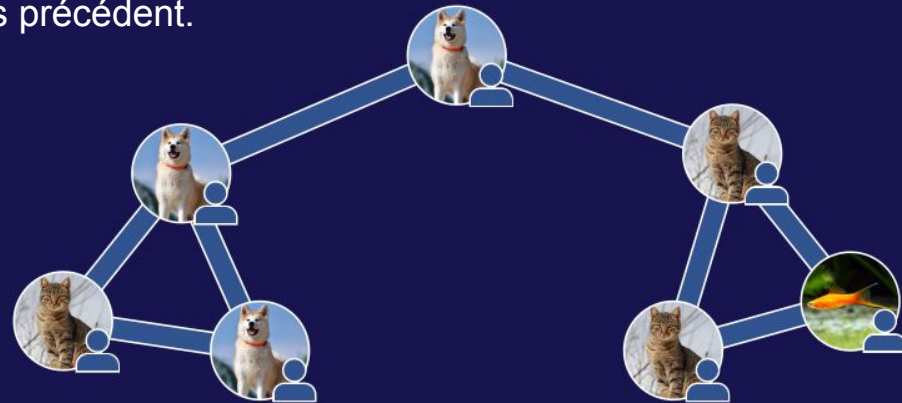


## *L'objectif*

- Créer un algorithme qui prend en entrée
  - un ensemble d'éléments formant un réseau,
  - des attributs pour chaque élément,
  - aucune information préalable concernant ce qui est défini comme une anomalie;
- et qui retourne
  - Un ensemble d'éléments identifiés comme anomalies, ou
  - Une valeur de "normalité" pour chaque élément, ou encore
  - Un classement de la "normalité" de chaque élément.

## Quelques problèmes :

- Il n'existe pas de définition uniforme de ce qu'est une "anomalie".
- Les anomalies sont par définition rares et difficiles à observer.
- Les anomalies peuvent provenir de multiples facteurs :
  - Les attributs individuels des éléments;
  - Les manières dont les éléments sont connectés les uns aux autres;
  - Une combinaison des deux phénomènes précédent.





- Introduction
- **Méthode**
- Résultats



## L'idée

Utiliser ce qui était un problème comme avantage :

*“Les anomalies sont par définition rares et difficiles à observer.” - BC*

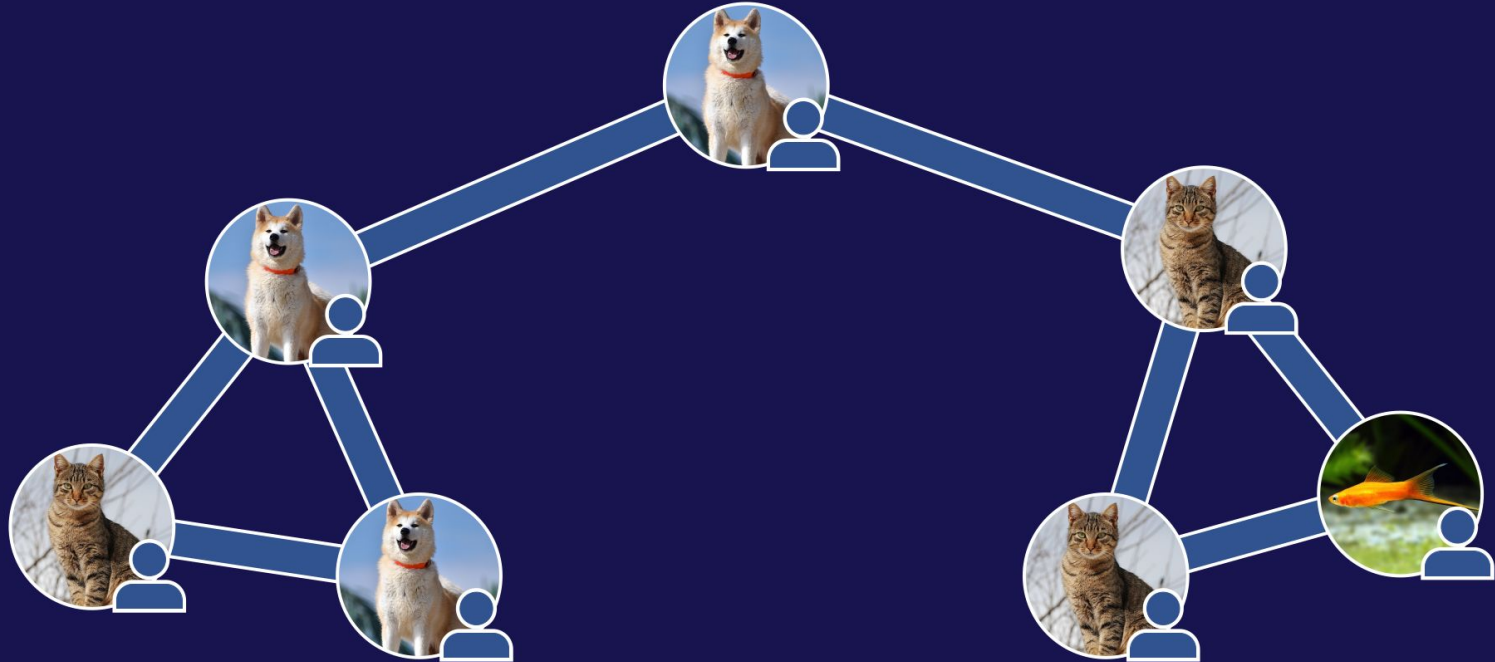
- Si l'on observe un élément au hasard, il sera probablement normal.
- On a donc accès à de très nombreux exemples “normaux”.
- On peut utiliser cela pour définir ce qui est normal et définir “anormal” comme étant “non normal”.



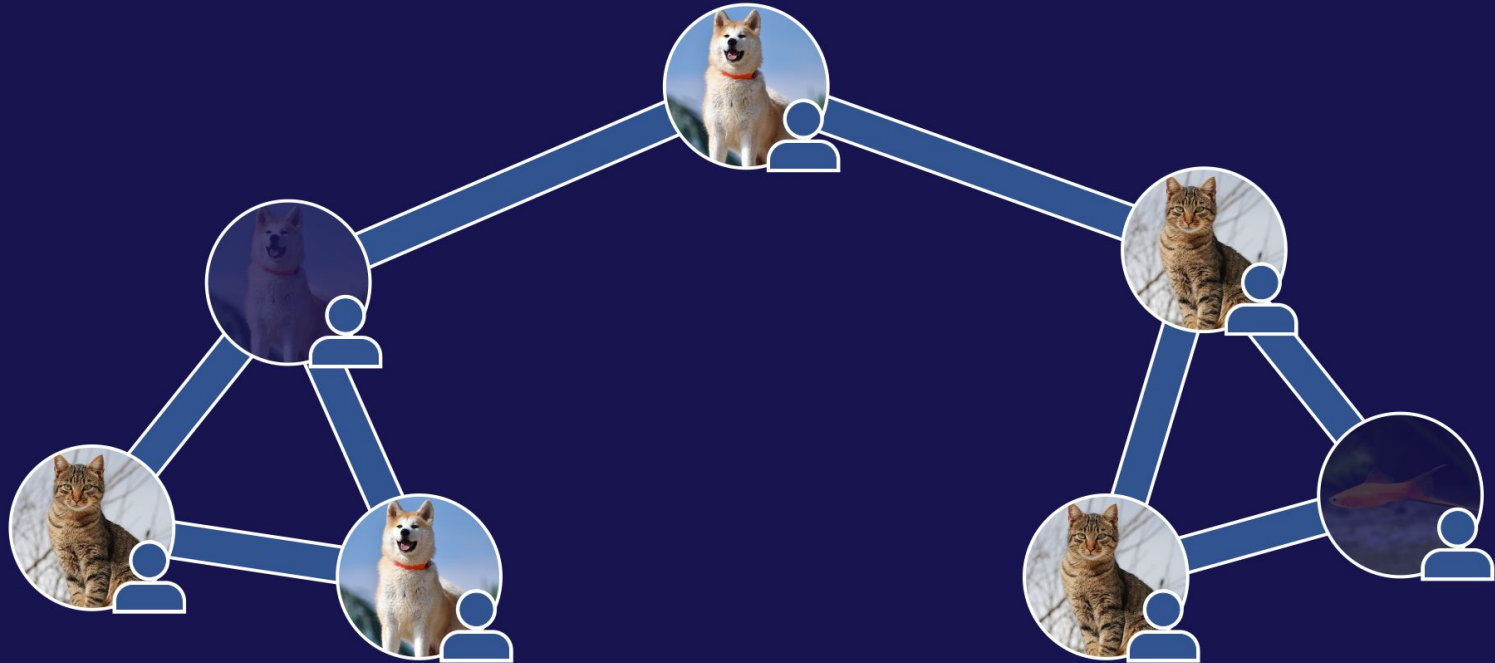
## *La méthode*

- On commence par “cacher” certains attributs.
- En utilisant l’information non cachée (réseaux et autres attributs), on devine ce qu’on vient de cacher.
- En répétant cela de nombreuses fois, on est capable de reproduire les attributs.
- On compare maintenant ce qu’on a reproduit avec la valeur originale pour identifier les anomalies.

## La méthode



## La méthode





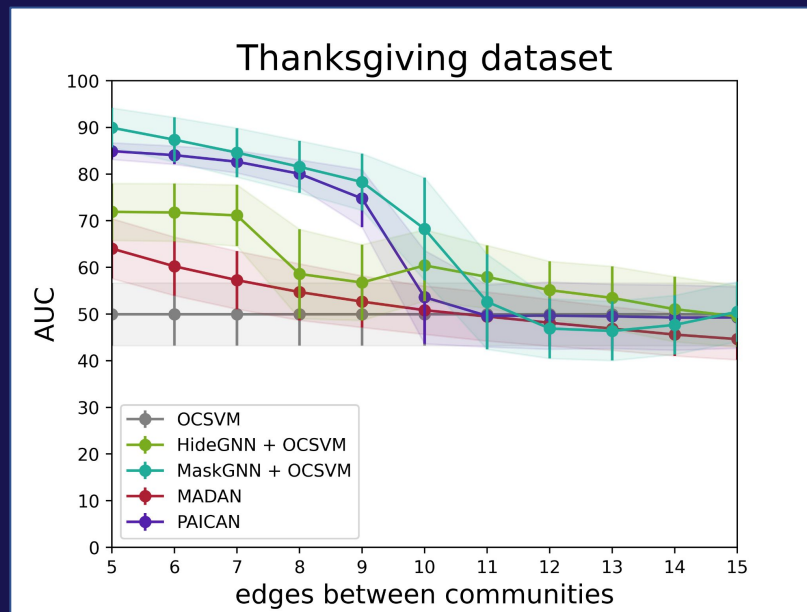
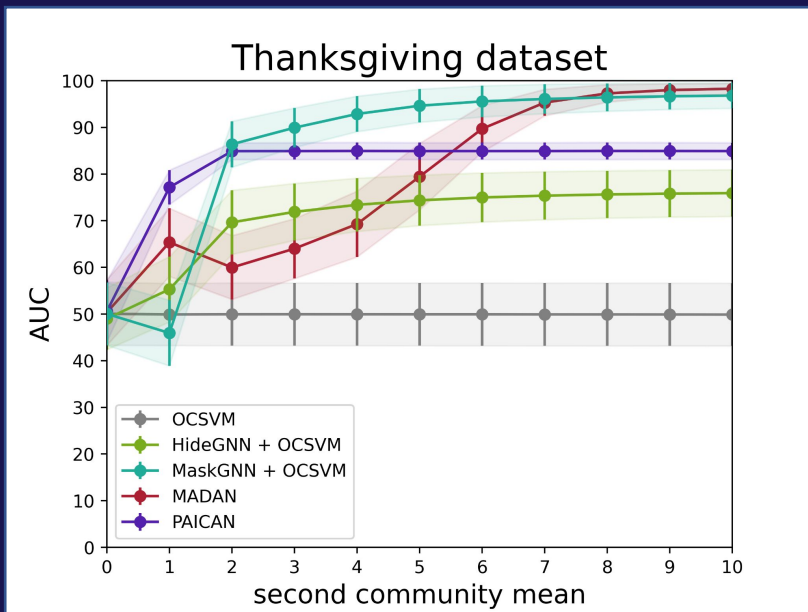
- Introduction
- Méthode
- Résultats



## ***Résultats***

- Méthode prometteuse surpassant les précédents résultats.
- Nos modèles sont capables de détecter différents types d'anomalies.
- Ils permettent d'adapter des algorithmes de détection d'anomalies à des réseaux.
- Ils apportent une nouvelle approche pour les algorithmes d'apprentissage sur les réseaux.

# Résultats



## Résultats

	Books		Disney*		Enron*	
	Original	Reduced	Original	Reduced	Original	Reduced
OCSVM <sup>†</sup>	35.90 <sup>‡</sup>	37.33 <sup>‡</sup>	36.44 <sup>‡</sup>	<b>85.88<sup>‡</sup></b>	55.90 <sup>‡</sup>	43.22 <sup>‡</sup>
Hide + OCSVM	49.58 (±1.47)	<b>52.21</b> (±2.41)	46.97 (±2.26)	43.16 (±1.03)	32.76 (±3.79)	<b>51.84</b> (±3.35)
Mask + OCSVM	<b>55.06</b> (±1.64)	48.49 (±2.07)	<b>47.63</b> (±1.73)	43.07 (±6.30)	<b>56.28</b> (±9.73)	50.46 (±6.55)
MEG <sup>†</sup> [24]	<b>63.12<sup>‡</sup></b>	56.60 <sup>‡</sup>	50.28 <sup>‡</sup>	39.55 <sup>‡</sup>	26.47 <sup>‡</sup>	<b>63.08<sup>‡</sup></b>
Hide + MEG	60.42 (±7.03)	<b>65.36</b> (±3.26)	46.51 (±6.51)	<b>69.44</b> (±3.47)	<b>44.22</b> (±14.05)	40.12 (±6.54)
Mask + MEG	57.06 (±6.55)	62.17 (±4.67)	<b>51.25</b> (±3.46)	62.44 (±13.38)	42.36 (±16.82)	40.63 (±12.21)
MADAN [13]	48.80 <sup>‡</sup>	45.54 <sup>‡</sup>	<b>67.51<sup>‡</sup></b>	19.49 <sup>‡</sup>	<b>61.81<sup>‡</sup></b>	<b>73.64<sup>‡</sup></b>
Hide + MADAN	<b>57.21</b> (±9.26)	52.08 (±7.64)	59.09 (±15.03)	<b>49.82</b> (±15.02)	54.24 (±14.41)	47.47 (±13.33)
Mask + MADAN	58.43 (±8.87)	<b>52.34</b> (±7.89)	55.98 (±14.07)	48.87 (±12.54)	60.55 (±14.58)	50.20 (±14.13)
PAICAN [2]	67.76 <sup>‡</sup>	55.38 <sup>‡</sup>	<b>75.14<sup>‡</sup></b>	<b>73.45<sup>‡</sup></b>	26.94 <sup>‡</sup>	57.77 <sup>‡</sup>
Hide + PAICAN	68.23 (±1.45)	<b>72.82</b> (±3.38)	62.73 (±8.14)	71.92 (±1.15)	<b>54.59</b> (±3.31)	58.65 (±10.07)
Mask + PAICAN	<b>68.90</b> (±1.15)	61.25 (±3.43)	67.31 (±6.87)	64.76 (±10.82)	45.92 (±8.08)	<b>65.40</b> (±5.46)





***Merci !***

# 89<sup>E</sup> CONGRÈS

## ORGANISATRICES



## GRANDS PARTENAIRES



## PARTENAIRES DES ACTIVITÉS SCIENTIFIQUES

