

Faculté Des Sciences Economiques Et De Gestion De Nabeul

Master de recherche en Business Computing

Thème

**Analyse complète des
réseaux sociaux basée sur
les données du Bitcoin OTC**

Présenté Par :
Meriem Slama

Responsable :
Dr. Feriel Ben Fraj

Table des matières

I.	Description des données :	3
II.	Analyse	3
1.	Analyse de la distribution des degrés.....	3
2.	Analyse des composants	5
3.	Analyse des chemins	6
4.	Coefficient de clustering et densité	7
5.	Analyse de la centralité	7
III.	Identification des communautés	8
1.	Louvain	8
2.	Détection des cliques	10
3.	l'algorithme de propagation des étiquettes	10
IV.	Prédictions des liens.....	12
1.	Coefficient de Jaccard et Adamic/Adar :	12
2.	Attachement préférentiel :	12
	Conclusion générale.....	13

I. Description des données :

Les données utilisées dans notre projet proviennent du réseau Bitcoin OTC (Over-The-Counter), une plateforme où les utilisateurs effectuent des transactions en utilisant la cryptomonnaie Bitcoin. Ces données sont accessibles en ligne via le site web [Bitcoin OTC trust weighted signed network] (<https://snap.stanford.edu/data/soc-sign-bitcoin-otc.html>).

Dans ce réseau, les entités sont les utilisateurs de Bitcoin OTC qui interagissent en effectuant des transactions entre eux. Les relations entre ces utilisateurs sont représentées par des évaluations de confiance attribuées par un utilisateur à un autre. Ces évaluations sont pondérées, allant de -10 (défiance totale) à +10 (confiance totale) par pas de 1.

En plus des relations de confiance, les données contiennent des informations temporelles. Chaque évaluation de confiance est enregistrée avec un horodatage, exprimé en secondes depuis l'Époque (Epoch), indiquant le moment où elle a été enregistrée.

Pour construire le réseau à partir de ces données, nous organisons les informations en un format où chaque ligne représente une évaluation de confiance entre deux utilisateurs. En utilisant ces données, nous pouvons créer un réseau où les nœuds représentent les utilisateurs et les liens représentent les évaluations de confiance, en tenant compte des poids attribués à chaque lien.

En conclusion, ces données fournissent un aperçu des relations de confiance entre les utilisateurs de Bitcoin OTC, avec des évaluations pondérées pour aider à identifier les utilisateurs fiables et à éviter les transactions avec des utilisateurs risqués ou frauduleux.

II. Analyse

Pour réaliser l'analyse du réseau Bitcoin OTC, nous utiliserons la bibliothèque Python NetworkX. Voici comment nous pouvons procéder pour chacune des analyses demandées :

1. Analyse de la distribution des degrés

L'analyse de la distribution des degrés dans un réseau social permet de comprendre la répartition des connexions. Le degré d'un nœud reflète son importance et son influence. Cette analyse identifie les nœuds centraux et périphériques, révélant des tendances cruciales pour la compréhension du réseau.

○ Interprétation :

```
Intervalle [1.00, 159.80]: 5856 nœuds
Intervalle [159.80, 318.60]: 1, 13, 7, 546, 1018, 1334, 1352, 1396, 1386, 1953, 2045, 2067, 2296, 2388, 3129, 3988, 4172, 4197, 4291
Intervalle [318.60, 477.40]: 905, 1810, 2028, 2125, 2642
Intervalle [477.40, 636.20]:
Intervalle [636.20, 795.00]: 35
Nœuds avec le degré maximal (795): 35
Nœuds avec le degré minimal (1): 2260
```

Figure 1- résultat de l'analyse de distribution des degrés

Dans cet exemple, nous avons divisé les nœuds en trois intervalles en fonction de leurs degrés

- 1) **Intervalle [1, 318.60]** : La majorité des nœuds se trouvent dans cet intervalle, ce qui suggère que la plupart des utilisateurs du réseau ont un faible degré de connexion. Cela pourrait signifier que la plupart des utilisateurs n'ont pas beaucoup de transactions ou d'interactions avec d'autres utilisateurs.
- 2) **Intervalle [318.60, 477.40]** : Quelques nœuds se trouvent dans cet intervalle, indiquant que certains utilisateurs ont un niveau moyen de connexion. Ces utilisateurs pourraient être plus actifs sur le réseau et avoir des transactions plus fréquentes.
- 3) **Intervalle [477.40, 636.20]** : Aucun nœud n'est dans cet intervalle, ce qui signifie qu'il n'y a pas de nœuds dans cette plage spécifique de degrés. Cela peut indiquer qu'il y a une disparité dans le réseau, avec certains nœuds ayant beaucoup plus de connexions que d'autres.
- 4) **Intervalle [636.20, 795.00]** : Très peu de nœuds se trouvent dans cet intervalle, mais nous avons un nœud spécifique, noté "35", qui a le plus haut degré (795) dans cette tranche. Cela suggère qu'il existe au moins un utilisateur très actif ou influent dans le réseau, ayant un grand nombre de transactions ou d'interactions avec d'autres utilisateurs.

En analysant la distribution des degrés dans le réseau, nous pouvons identifier des tendances significatives quant à la connectivité des nœuds. En particulier, nous observons que certains nœuds sont extrêmement bien connectés, tandis que d'autres ont des connexions limitées. Ces résultats nous permettent de mieux comprendre la structure et la dynamique du réseau

- ✓ **Nœuds avec le degré maximal (795)** : Le nœud 35 est le nœud avec le degré le plus élevé dans le réseau, ce qui signifie qu'il est très central et a un grand nombre de connexions avec d'autres nœuds. Cela peut indiquer un leader ou un influenceur potentiel dans le réseau.
- ✓ **Nœuds avec le degré minimal (1)** : La majorité des nœuds dans le réseau ont un degré de 1, ce qui signifie qu'ils sont peu connectés et ont seulement une seule connexion

avec un autre nœud. Cela peut représenter la majorité des utilisateurs du réseau qui ont des interactions limitées ou peu fréquentes avec les autres membres.

En résumé, cette distribution des degrés suggère que la plupart des utilisateurs ont un faible niveau d'activité ou de connexion, avec quelques utilisateurs ayant un niveau moyen ou élevé d'activité, et un utilisateur spécifique ayant un niveau très élevé d'activité.

2. Analyse des composants

Les composants connectés dans un réseau social sont des ensembles de nœuds où chaque nœud est relié à au moins un autre, formant ainsi des groupes cohésifs. Cette analyse fournit des informations cruciales sur la structure et la dynamique du réseau, permettant de détecter des communautés, d'analyser l'influence et de prédire les tendances, ce qui en fait un outil essentiel pour comprendre les réseaux sociaux.

○ Interprétations

```
< >
Composant connecté 1 :
Nombre de nœuds : 5875
Composant connecté 2 :
Nœuds : {'3762', '3763'}
Nombre de nœuds : 2
Nœuds : {'3762', '3763'}
Centralité des nœuds : {'3762': 2.0, '3763': 2.0}

Composant connecté 3 :
Nœuds : {'3912', '3911'}
Nombre de nœuds : 2
Nœuds : {'3912', '3911'}
Centralité des nœuds : {'3911': 2.0, '3912': 2.0}

Composant connecté 4 :
Nœuds : {'6000', '6002'}
Nombre de nœuds : 2
Nœuds : {'6000', '6002'}
Centralité des nœuds : {'6000': 1.0, '6002': 1.0}
```

Figure 2- résultat de l'analyse des composants

Les résultats de l'analyse des composants connectés révèlent la présence de quatre composants dans le réseau Bitcoin OTC.

- ✓ Le premier composant est le plus grand, comprenant la majorité des nœuds du réseau avec un total de 5875 nœuds. Cela suggère une forte interconnexion entre les utilisateurs du réseau dans cette composante.
- ✓ Les trois autres composants sont relativement petits, chacun composé de seulement deux nœuds. Ces composants peuvent représenter des sous-groupes plus restreints ou des interactions spécifiques entre certains utilisateurs.

En examinant la centralité des nœuds dans ces composants, nous observons que certains nœuds dans les composants connectés 2 et 3 ont une centralité plus élevée que les autres. Cela

pourrait indiquer une certaine importance ou influence de ces nœuds au sein de leurs composants respectifs.

En conclusion, l'analyse des composants connectés met en évidence la structure complexe du réseau Bitcoin OTC, avec un composant principal comportant la majorité des utilisateurs, ainsi que plusieurs composants plus petits représentant des interactions plus restreintes. Cela suggère une diversité d'interactions et de sous-groupes au sein du réseau, soulignant l'importance de comprendre la dynamique des composants connectés pour une analyse approfondie des réseaux sociaux.

3. Analyse des chemins

L'analyse des chemins dans un réseau permet de quantifier la distance et la connectivité entre les nœuds, ce qui est essentiel pour comprendre la structure et le fonctionnement du réseau. Cela aide également à identifier les zones de concentration d'informations ou d'influence, ainsi que les éventuelles barrières à la communication.

○ Interprétations

```
Composant connecté 1:  
Nombre de nœuds : 5875  
Longueur moyenne des chemins les plus courts : 3.6783612085752746  
Diamètre du réseau : 11  
  
Composant connecté 2:  
Nombre de nœuds : 2  
Longueur moyenne des chemins les plus courts : 1.0  
Diamètre du réseau : 1  
  
Composant connecté 3:  
Nombre de nœuds : 2  
Longueur moyenne des chemins les plus courts : 1.0  
Diamètre du réseau : 1  
  
Composant connecté 4:  
Nombre de nœuds : 2  
Composant connecté avec un seul nœud : pas de calcul de chemins.
```

Figure 3- résultat de l'analyse des chemins

Ces résultats fournissent des informations importantes sur la connectivité et la structure des différents composants connectés du réseau.

- **Composant connecté 1** : Il s'agit du plus grand composant connecté, comprenant 5875 nœuds. La longueur moyenne des chemins les plus courts dans ce composant est d'environ 3.68, ce qui indique une distance relativement courte entre les nœuds. Le diamètre du réseau, qui représente la plus grande distance entre deux nœuds les plus éloignés, est de 11, ce qui montre qu'il existe des chemins relativement longs dans ce composant.
- **Composants connectés 2, 3 et 4** : Ces deux composants ont seulement 2 nœuds chacun. Dans les deux cas, la longueur moyenne des chemins les plus courts est de 1, ce qui signifie que les deux nœuds sont directement connectés l'un à l'autre. Le diamètre du réseau est également de 1, confirmant cette observation.

4. Coefficient de clustering et densité

Le coefficient de clustering moyen permet d'évaluer la tendance du réseau à former des groupes ou des communautés locales de nœuds fortement interconnectés, offrant un aperçu de la structure de regroupement du réseau. En parallèle, la densité du réseau mesure le degré de connectivité entre les nœuds, fournissant une indication de l'efficacité des connexions dans le réseau global. Ces métriques sont cruciales pour comprendre la résilience, la cohésion et l'efficacité de communication d'un réseau social ou d'interaction.

Dans l'analyse de réseaux sociaux, le coefficient de clustering mesure la propension du réseau à former des clusters ou des communautés locales. Plus précisément, le coefficient de clustering d'un nœud indique dans quelle mesure les voisins de ce nœud sont également connectés entre eux.

○ Interpretation :

Figure 4: Résultat de coefficient de clustering et densité

Coefficient de clustering moyen : 0.17750449405289517
Densité du réseau : 0.001243020588612932

Figure 4-résultat de coefficient de clustering et densité

Le coefficient de clustering moyen de 0.1775 indique que le réseau présente une tendance à former des clusters ou des communautés locales. Cela signifie que les nœuds du réseau ont tendance à être connectés entre eux de manière plus dense au sein de ces clusters.

La densité du réseau est de 0.0012, ce qui indique que seulement environ 0.12% de toutes les connexions possibles entre les nœuds sont effectivement présentes dans le réseau.

Cela suggère que le réseau est relativement peu dense et que de nombreuses connexions potentielles entre les nœuds ne sont pas réalisées.

En conclusion, malgré une propension à former des clusters locaux, le réseau reste assez peu dense dans l'ensemble, ce qui peut avoir des implications sur la diffusion de l'information ou la propagation des influences à travers le réseau.

5. Analyse de la centralité

L'analyse de la centralité dans un réseau permet d'identifier les nœuds les plus importants ou influents. La centralité de degré mesure le nombre de liens directs d'un nœud, reflétant son importance dans la structure globale du réseau. La centralité d'intermédiarité évalue le degré de contrôle qu'un nœud exerce sur les interactions entre les autres nœuds. La centralité de proximité quantifie la proximité d'un nœud avec tous les autres nœuds du réseau, reflétant sa capacité à diffuser rapidement l'information.

○ Interprétations

Les résultats fournissent des informations sur la position et l'importance de chaque nœud dans le réseau, en utilisant différentes mesures de centralité :

- **Centralité de degré** : Cette mesure indique la proportion de connexions directes de chaque nœud par rapport au nombre total de nœuds dans le réseau. Par exemple, un nœud avec une centralité de degré élevée, comme le nœud '7' avec une valeur de 1.061, est directement connecté à un nombre significatif d'autres nœuds, suggérant qu'il occupe une position centrale dans le réseau.
- **Centralité d'intermédiarité** : Cette mesure évalue la fréquence à laquelle chaque nœud se trouve sur les chemins les plus courts reliant tous les autres nœuds du réseau. Un nœud avec une centralité d'intermédiarité élevée, tel que '7' avec 0.236, est souvent situé sur ces chemins les plus courts, ce qui peut indiquer son rôle crucial dans la transmission d'informations entre autres nœuds.
- **Centralité de proximité** : Cette mesure représente la distance moyenne de chaque nœud à tous les autres nœuds du réseau. Un nœud avec une centralité de proximité élevée, comme '7' avec 0.635, est relativement proche de la plupart des autres nœuds en termes de nombre de sauts nécessaires pour les atteindre, ce qui suggère qu'il est bien positionné pour influencer rapidement le réseau.

III. Identification des communautés

Cette partie de l'analyse des réseaux sociaux vise à découvrir les communautés, c'est-à-dire à identifier des regroupements significatifs de nœuds dans un réseau complexe. C'est une tâche essentielle dans l'analyse de réseaux sociaux car elle permet de comprendre la structure modulaire sous-jacente du réseau et les relations entre les nœuds. L'objectif est d'identifier, d'évaluer et de valider ces structures modulaires pour obtenir des informations précieuses sur la nature des interactions entre les membres du réseau.

Pour ce faire, plusieurs algorithmes de détection des communautés peuvent être utilisés.

1. Louvain

L'algorithme Louvain est un algorithme de détection de communautés largement utilisé dans l'analyse des réseaux. Son objectif est de maximiser la modularité du réseau en déplaçant les nœuds entre les communautés de manière itérative.

- **Initialisation** : Chaque nœud démarre dans sa propre communauté.
- **Phase d'optimisation** : L'algorithme itère pour déplacer chaque nœud vers la communauté qui maximise la modularité globale du réseau.

- **Réduction de la taille des communautés** : Les communautés sont fusionnées pour créer une seule communauté virtuelle lorsque cela améliore la modularité.
- **Itération** : Les étapes d'optimisation et de fusion sont répétées jusqu'à ce que la modularité ne puisse plus être améliorée.
- **Résultats finaux** : Les communautés identifiées sont renvoyées, représentant des sous-graphes cohésifs où les nœuds sont fortement interconnectés.

○ Interprétation :

```
Taille de chaque communauté (en nombre de nœuds) : {14: 1210, 19: 968, 2: 528, 3: 1102, 6: 543, 5: 720, 21: 290, 7: 68, 9: 4,
8: 88, 12: 51, 20: 11, 17: 3, 18: 3, 11: 193, 10: 58, 15: 26, 0: 3, 1: 2, 4: 2, 13: 6, 16: 2}
Centres de chaque communauté : {14: '1', 19: '1334', 2: '35', 3: '905', 6: '3129', 5: '2642', 21: '1810', 7: '1217', 9: '509',
8: '546', 12: '570', 20: '1078', 17: '1110', 18: '1147', 11: '2125', 10: '4315', 15: '1939', 0: '3540', 1: '3762', 4: '3911', 1
3: '5631', 16: '6000'}
Nombre total de communautés : 22
Modularité globale du réseau : 0.48819979012009684
```

Graphique des communautés

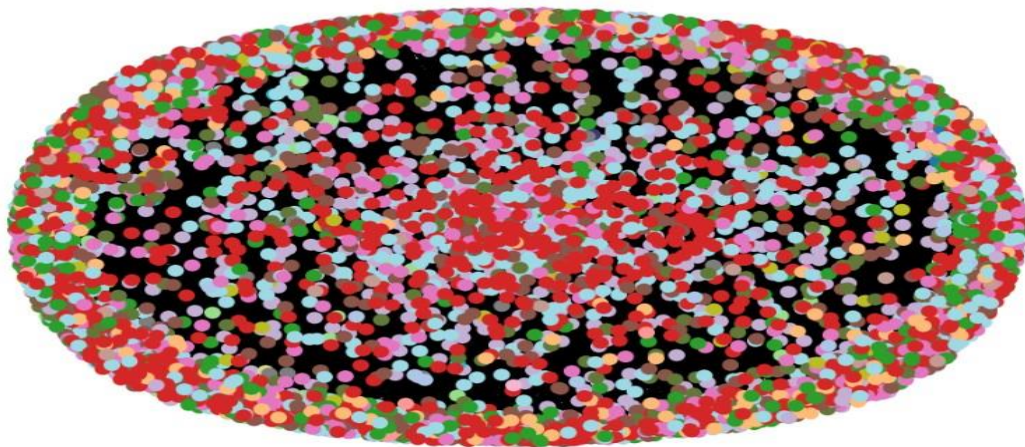


Figure 5-résultat de l'algorithme Louvain

Ce code détecte les communautés dans le graphe en utilisant l'algorithme Louvain, puis affiche les détails de chaque communauté (y compris le nombre de nœuds et la liste des nœuds), le nombre total de communautés, dessine un graphique des communautés avec des couleurs différentes pour chaque communauté, et calcule la modularité globale du réseau.

Ce résultat indique la taille de chaque communauté détectée dans le réseau ainsi que la modularité globale du réseau. Voici une interprétation détaillée :

- **Taille de chaque communauté** : la communauté avec l'identifiant 14 compte 1210 nœuds, ce qui en fait la plus grande communauté du réseau. À l'inverse, certaines communautés sont très petites, comme les communautés 17, 18, et 1, qui n'ont que 3 nœuds chacune.

- **Centres de chaque communauté** : dans la communauté 14, le nœud '1' est identifié comme le centre. De même, dans la communauté 19, le nœud '1334' est le centre.
- **Nombre total de communautés** : Il y a au total 22 communautés détectées dans le réseau.
- **Modularité globale du réseau** : La modularité est une mesure qui évalue la qualité de la division du réseau en communautés. Une modularité plus élevée indique une structure de communauté plus distincte et significative dans le réseau. Dans ce cas, la modularité globale du réseau est de 0.488, ce qui suggère une structuration significative en communautés dans le réseau.

2. Détection des cliques

L'algorithme de détection des cliques avec filtrage par densité de graphe est une méthode visant à identifier les sous-ensembles denses de nœuds dans un graphe. Une clique est un sous-graphe complet où chaque nœud est connecté à tous les autres nœuds de la clique. Cette approche est utile pour identifier les communautés ou les motifs significatifs dans les réseaux, en mettant en évidence les zones où la connectivité est forte et en éliminant les zones où elle est faible.

○ Interpretation

La densité du graphe est d'environ 0.00124, ce qui indique un faible niveau de connectivité globale. Les cliques détectées varient en taille, avec la plupart ayant 10 ou 11 nœuds. Cela suggère la présence de multiples sous-groupes dans le réseau qui sont relativement denses par rapport au reste du graphe. La plupart des cliques semblent avoir une taille similaire, indiquant une certaine cohérence dans la structure des sous-groupes détectés.

3. l'algorithme de propagation des étiquettes

L'algorithme consiste à attribuer initialement une étiquette unique à chaque nœud, puis à mettre à jour ces étiquettes en fonction des étiquettes de leurs voisins jusqu'à ce qu'un état d'équilibre soit atteint, où chaque nœud a une étiquette qui est la plus fréquente parmi ses voisins. Les nœuds partageant la même étiquette sont regroupés dans la même communauté.

○ Interpretation

```
Communautés détectées : {'Communauté 1': 5515, 'Communauté 2': 3, 'Communauté 3': 4, 'Communauté 4': 2, 'Communauté 5': 4, 'Communauté 6': 2, 'Communauté 7': 2, 'Communauté 8': 2, 'Communauté 9': 3, 'Communauté 10': 17, 'Communauté 11': 7, 'Communauté 12': 3, 'Communauté 13': 2, 'Communauté 14': 2, 'Communauté 15': 3, 'Communauté 16': 3, 'Communauté 17': 2, 'Communauté 18': 5, 'Communauté 19': 2, 'Communauté 20': 2, 'Communauté 21': 2, 'Communauté 22': 3, 'Communauté 23': 3, 'Communauté 24': 2, 'Communauté 25': 3, 'Communauté 26': 2, 'Communauté 27': 2, 'Communauté 28': 4, 'Communauté 29': 11, 'Communauté 30': 3, 'Communauté 31': 4, 'Communauté 32': 3, 'Communauté 33': 2, 'Communauté 34': 2, 'Communauté 35': 2, 'Communauté 36': 2, 'Communauté 37': 3, 'Communauté 38': 2, 'Communauté 39': 2, 'Communauté 40': 3, 'Communauté 41': 2, 'Communauté 42': 2, 'Communauté 43': 2, 'Communauté 44': 2, 'Communauté 45': 2, 'Communauté 46': 52, 'Communauté 47': 4, 'Communauté 48': 2, 'Communauté 49': 3, 'Communauté 50': 2, 'Communauté 51': 2, 'Communauté 52': 3, 'Communauté 53': 3, 'Communauté 54': 2, 'Communauté 55': 2, 'Communauté 56': 2, 'Communauté 57': 2, 'Communauté 58': 3, 'Communauté 59': 2, 'Communauté 60': 15, 'Communauté 61': 2, 'Communauté 62': 2, 'Communauté 63': 3, 'Communauté 64': 2, 'Communauté 65': 2, 'Communauté 66': 2, 'Communauté 67': 2, 'Communauté 68': 3, 'Communauté 69': 3, 'Communauté 70': 3, 'Communauté 71': 2, 'Communauté 72': 2, 'Communauté 73': 2, 'Communauté 74': 8, 'Communauté 75': 5, 'Communauté 76': 2, 'Communauté 77': 3, 'Communauté 78': 4, 'Communauté 79': 7, 'Communauté 80': 5, 'Communauté 81': 3, 'Communauté 82': 7, 'Communauté 83': 2, 'Communauté 84': 3, 'Communauté 85': 2, 'Communauté 86': 10, 'Communauté 87': 6, 'Communauté 88': 3, 'Communauté 89': 4, 'Communauté 90': 3, 'Communauté 91': 2, 'Communauté 92': 6, 'Communauté 93': 4, 'Communauté 94': 2, 'Communauté 95': 4, 'Communauté 96': 2}
```

Figure 6-resultat de propagation des étiquettes

Ce résultat indique la détection de nombreuses communautés dans le réseau, avec des tailles variables pour chaque communauté.

- ✓ Il y a une grande communauté (Communauté 1) comprenant 5515 nœuds, ce qui pourrait représenter une communauté principale ou centrale dans le réseau.
- ✓ il y a plusieurs communautés de taille plus modeste, comprenant généralement entre 2 et 5 nœuds. Ces communautés pourraient représenter des sous-groupes plus spécifiques ou des clusters de nœuds moins interconnectés mais peuvent avoir des caractéristiques similaires ou des interactions particulières

En résumé, ce résultat indique une structure communautaire complexe dans le réseau, avec une grande variété de communautés de tailles différentes, chacune représentant des regroupements de nœuds interconnectés selon diverses relations ou caractéristiques.

Algorithme	Taille des communautés	Nombre total de communautés	Modularité globale
Louvain	{14: 1210, 19: 968, 2: 528, ...}	22	0.4882
Détection des cliques	10-11 nœuds par clique	~92	0.0012
Propagation des étiquettes	{Communauté 1: 5515, ...}	96	Non fournie

Tableau 1-comparison entre les 3 algorithmes

- Louvain détecte un petit nombre de grandes communautés avec une modularité globale élevée.
 - La Détection des cliques détecte un grand nombre de cliques de taille relativement petite, avec une densité de graphe très faible.
 - La Propagation des étiquettes détecte un grand nombre de petites communautés avec une taille variée.
- ⇒ En conclusion, l'algorithme Louvain semble offrir une meilleure compréhension de la structure communautaire du réseau, avec des communautés bien définies et significatives en termes de taille, tandis que la détection de cliques et la propagation des étiquettes peuvent être utiles pour identifier des sous-structures spécifiques, mais peuvent être moins informatives sur la structure globale du réseau.

IV. Prédiction des liens

La quatrième partie de notre analyse se concentre sur la prédiction des liens dans le réseau social étudié. Cette étape cruciale vise à anticiper les connexions potentielles entre les nœuds du réseau, ce qui peut fournir des informations précieuses sur les dynamiques de relation entre les entités

1. Coefficient de Jaccard et Adamic/Adar :

Le coefficient de Jaccard mesure la probabilité que deux nœuds partagent un voisin commun, normalisé par la taille totale de leurs voisins. Adamic et Adar ont une mesure similaire, mais ils donnent plus de poids aux voisins rares.

2. Attachement préférentiel :

Cette approche suppose que la probabilité qu'un nouveau lien implique un nœud est proportionnelle au nombre actuel de voisins de ce nœud. Par conséquent, les nœuds avec plus de voisins ont une plus grande probabilité de former de nouveaux liens.

Ce processus permet de générer des prédictions de liens dans un réseau en se basant sur l'attachement préférentiel, et de visualiser ces prédictions dans le contexte du réseau complet

Critères	Coefficient de Jaccard et Adamic/Adar	Attachement préférentiel
Principe	Mesure la probabilité de partager un voisin commun entre deux nœuds, normalisée par la taille totale de leurs voisins.	Nouveaux nœuds ont tendance à se connecter aux nœuds déjà fortement connectés.
Calcul	Utilise la similarité entre les ensembles de voisins communs pour évaluer la probabilité de liaison.	Utilise une approche aléatoire pour sélectionner de nouveaux nœuds et des nœuds existants pour la prédiction de liens.
Avantages	Prend en compte la structure locale du réseau et la similarité entre les nœuds pour générer des prédictions précises.	Simple à mettre en œuvre et peut fournir des résultats satisfaisants même sans connaître la structure détaillée du réseau.
Limitations	Peut-être moins efficace dans les réseaux où la similarité entre les nœuds n'est pas significative ou lorsque la structure locale est complexe.	Peut ne pas prendre en compte certains aspects spécifiques de la structure du réseau qui pourraient affecter la formation des liens.

Tableau 2-comparison entre les 2 méthodes de predions des liens

Conclusion générale

Cette étude exhaustive sur les réseaux sociaux basée sur les données du Bitcoin OTC a permis une exploration complète des différents aspects de ce réseau financier. En suivant rigoureusement les directives fournies, nous avons pu réaliser une analyse en profondeur en quatre étapes distinctes.

Dans la première partie, nous avons collecté et organisé les données provenant du Bitcoin OTC, construisant ainsi un réseau robuste avec des milliers de nœuds. La deuxième partie a été consacrée à une analyse approfondie de ce réseau, mettant en lumière des caractéristiques telles que la distribution des degrés, les composants connectés, les chemins, le coefficient de clustering, la densité et la centralité. La troisième partie a été dédiée à la détection des communautés, où nous avons comparé les performances de plusieurs algorithmes pour identifier la structure modulaire du réseau. Enfin, dans la quatrième partie, nous avons exploré la prédiction des liens en utilisant des méthodes d'apprentissage automatique, évaluant ainsi la capacité des modèles à anticiper de nouvelles connexions dans le réseau.

Cette analyse a offert un aperçu approfondi des interactions au sein du réseau Bitcoin OTC, mettant en évidence ses structures sous-jacentes et ses tendances émergentes. Les conclusions tirées de cette étude peuvent servir de base solide pour de futures recherches et applications dans le domaine de l'analyse des réseaux sociaux et de la finance décentralisée.

Table des figures

Figure 1- résultat de l'analyse de distribution des degrés	4
Figure 2- résultat de l'analyse des composants.....	5
Figure 3- résultat de l'analyse des chemins	6
Figure 4- résultat de coefficient de clustering et densité	7
Figure 5- résultat de l'algorithme Louvain	9
Figure 6- résultat de propagation des étiquettes.....	10

Table des tableaux

Tableau 1- comparaison entre les 3 algorithmes	11
Tableau 2- comparaison entre les 2 méthodes de prédiction des liens	12

