

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITE MOHAMED BOUDIAF - M'SILA
FACULTE DES MATHÉMATIQUES ET
DE L'INFORMATIQUE

DEPARTEMENT D'INFORMATIQUE



MEMOIRE de fin d'étude

Présenté pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Informatique Décisionnel et Optimisation

Par : LOGRADA Nassira

SUJET

**La détection de communautés dans les réseaux
sociaux**

Soutenu publiquement le : / /2019 devant le jury composé de :

AKROUF Samir

Université de M'sila

Rapporteur

KHETTAF Abdelouahab

Université de M'sila

Président

AMRI Said

Université de M'sila

Examineur

Promotion : 2018 /2019

Dédicaces

A mes très chers parents.

Remerciements

Je remercie Dieu le tout puissant qui m'a donné le courage et la force pour aboutir à L'accomplissement de ce travail.

Je voudrais exprimer toute ma reconnaissance et mon respect à Monsieur Samir Akrouf pour son orientation.

Je remercie les membres du jury pour l'honneur qu'ils m'ont attribué en acceptant d'évaluer et de juger ce modeste travail.

Je remercie mes amis pour leur aide appréciable, leurs encouragements continus et leur soutien moral ininterrompu.

Ce travail n'aurait jamais été possible sans le soutien et l'appui moral des membres de ma famille. Je les remercie tous.

Table des Matières

Introduction générale.....	1
-----------------------------------	----------

CHAPITRE 1 LES RESEAUX SOCIAUX

Introduction.....	4
--------------------------	----------

1. Les médias sociaux.....	4
-----------------------------------	----------

1.1 Définition.....	4
---------------------	---

1.2 Évolution des medias sociaux.....	5
---------------------------------------	---

1.3 Les cinq piliers.....	5
---------------------------	---

1.4 Les différents types de médias sociaux.....	6
---	---

2. Les réseaux sociaux	7
-------------------------------------	----------

2.1 Définition.....	7
---------------------	---

2.2 Historique des réseaux sociaux	8
--	---

2.3 Les types de réseaux étudiés.....	9
---------------------------------------	---

2.4 Les principaux réseaux sociaux classés par catégorie.....	10
---	----

2.5 Le top des plateformes en 2019.....	16
---	----

2.6 L'utilité des réseaux sociaux	17
---	----

2.7 Les dangers des réseaux sociaux	18
---	----

Conclusion.....	19
------------------------	-----------

CHAPITRE 2 L'ANALYSE DES RESEAUX SOCIAUX

Introduction.....	21
--------------------------	-----------

1. Analyse des réseaux sociaux.....	21
--	-----------

1.1 Définition.....	21
---------------------	----

1.2 Les concepts d'analyse des réseaux sociaux.....	21
---	----

1.3 Les représentations des réseaux sociaux.....	22
--	----

1.4 Quelques définitions et indicateurs.....	23
--	----

2. Propriétés des réseaux sociaux	26
--	-----------

2.1 Effet de petits mondes	26
2.2 Distribution des degrés en loi de puissance	27
2.3 Structure communautaire	27
3. Méthodes d'analyse des réseaux sociaux.....	27
3.1. Méthodes traditionnelles.....	27
3.2. Fouille de données dans les réseaux sociaux	29
4. Utilisation de l'analyse des réseaux sociaux.....	30
Conclusion.....	31

CHAPITRE 3 LA DETECTION DE COMMUNAUTE : ETAT DE L'ART

Introduction.....	33
1. communauté	33
2. Détection de communauté	34
3. Classification des méthodes de détection de communauté.....	34
4. Mesure de qualité d'une partition	36
5. Les algorithmes de détection de communauté.....	38
5.1 Girvan et Newman.....	38
5.2 Louvain.....	39
5.3 Label propagation	41
6. caractéristiques des algorithmes.....	43
Conclusion.....	44

CHAPITRE 4 IMPLEMENTATION

Introduction.....	46
--------------------------	-----------

1. Outils et environnement de développement.....	45
2. Les bibliothèques	47
3. Le travail	47
Conclusion.....	52
Conclusion générale.....	53

Liste des Figures

Figure 1.1 Tous les médias sociaux	7
Figure 1.2 Historique des réseaux sociaux	9
Figure 1.3 Les Différents réseaux	14
Figure 1.4 Logo de Facebook	11
Figure 1.5 Logo de Twitter.....	11
Figure 1.6 Logo de Google+.....	11
Figure 1.7 Logo de Snapchat.....	12
Figure 1.8 Logo de LinkedIn	12
Figure 1.9 Logo de viadeo.....	12
Figure 1.10 Logo de Medium.....	13
Figure 1.11 Logo Tumblr	13
Figure 1.12 Logo YouTube	14
Figure 1.13 Logo Periscope	14
Figure 1.14 Logo d'Instagram.....	15
Figure 1.15 Logo de Pinterest.....	15
Figure 1.16 Logo de Flickr.....	15
Figure 1.17 Logo de Reddit	16
Figure 1.18 Le top des plateformes en 2019.....	16
Figure 2.1 Un exemple d'un graphe.....	22
Figure 2.2 Matrice d'adjacence.....	23
Figure 2.3(a) Illustre un graphe orienté, (b) montre un graphe non orienté.....	24
Figure 2.3 Réseau sociaux centre.....	28
Figure 3.1 Communautés dans un graphe	33
Figure 3.2 Pseudo-code de l'algorithme de Girvan-Newman.....	38
Figure 3.3 Graphe pour la présentation de l'algorithme de Girvan-Newman.....	39
Figure 3.4 Étapes de l'algorithme de Girvan-Newman.....	39
Figure 3.5 Pseudo-code de l'algorithme Louvain.....	40
Figure 3.6 Exemple d'exécution algorithme Louvain.....	40

Figure 3.7 Exemple d'exécution de l'algorithme Label Propagation.....	42
Figure 4.1 Plateforme ANACONDA	46
Figure 4.2 IDE spyder	47
Figure 4.3 la création et l'installation	48
Figure 4.4 interface de l'exécution	48
Figure 4.5 extraction d'un fichier.....	49
Figure 4.6 le résultat de la détection de communautés dans l'interface.....	49
Figure 4.7 le résultat dans la console.....	50
Figure 4.8 quelque ligne de code (1).....	50
Figure 4.9 quelque ligne de code (2).....	51
Figure 4.10 quelque ligne de code (3).....	51

Liste des Tables

Table 2.1 Type d'analyse du datamining.....	30
Table 3.1 Caractéristique des algorithmes.....	42

INTRODUCTION GENERALE

De nos jours, les réseaux sociaux sur internet apparaissent comme un nouveau moyen important de communication. Tout d'abord, il faut savoir qu'un réseau social est un ensemble d'identités sociales où des individus sont reliés entre eux par des liens d'interactions sociales, il regroupe de nombreux individus, nommés des utilisateurs.

Les réseaux sont souvent modélisés par des graphes, une structure permettant l'encodage de données relationnelles. Quel que soit le domaine applicatif, cette modélisation sert à l'étude de la structure émergeant des entrelacements entre individus. Ainsi, les utilisateurs d'un réseau social auront tendance à former des groupes plus fortement connectés entre eux qu'avec le reste du réseau.

Généralement, la densité des liens entre les nœuds du réseau varie d'une zone à une autre, ce qui implique l'existence de groupes de nœuds fortement connectés entre eux mais faiblement reliés aux autres nœuds du réseau. Ces zones, appelées communautés, on peut les définir comme des ensembles de nœuds fortement liés entre eux, et plus faiblement liés avec le reste du réseau .De ce fait , l'identification de communautés est un sujet important, puisqu'il peut être rencontré dans plusieurs domaines d'application et des situations du monde réel. En plus ils nous permet également de déterminer le rôle de différents acteurs au sein des communautés et dans le réseau dans sa globalité.

Par ailleurs, beaucoup de travaux portant sur la définition puis la détection de ces communautés ont été effectués durant ces dernières années, Et nous en décrirons certaines dans notre mémoire.

L'objectif principal de cet mémoire est la détection de communautés dans les réseaux sociaux.

Dans le premier chapitre, on commence par une vue d'ensemble sur les médias sociaux, ainsi que les réseaux sociaux.

Dans Le deuxième chapitre, nous allons aborder l'analyse des réseaux sociaux, ainsi que les propriétés des réseaux sociaux ajouter à cela les méthodes d'analyse des réseaux sociaux, l'utilisation de l'analyse des réseaux sociaux.

Le chapitre trois, est consacré aux communautés et leur détection ainsi que la classification des méthodes de détection des communautés, Ensuite les mesures de qualité d'une partition. Et nous définissons en suite les différents algorithmes de détection des communautés.

Dans le quatrième et le dernier chapitre, nous présentons l'implémentation de l'algorithme de Louvain utilisé afin de détecter nous clôturons ce document par une conclusion générale.

CHAPITRE 1
LES RESEAUX SOCIAUX

Introduction

De nos jours, les réseaux sociaux sur internet apparaissent comme un nouveau moyen important de communication. Tout d'abord, il faut savoir qu'un réseau social est un ensemble d'identités sociales où des individus sont reliés entre eux par des liens d'interactions sociales, il regroupe de nombreux individus, nommés des utilisateurs.

Dans ce chapitre introductif on présente les principaux traits des réseaux sociaux, on utilisera deux sections : la première vise à proposer une définition aussi complète que possible de ce qui sont les médias sociaux en précisant ces différents types et ces évolutions. Enfin dans la dernière section on va clore ce chapitre en présentant les réseaux sociaux en évoquant notamment ces principaux traits.

1. Les médias sociaux

1.1 Définition

Fred Cavazza, blogueur reconnu dans le domaine des médias sociaux, a publié une définition des médias sociaux très précise : « Les médias sociaux désignent l'ensemble des services permettant de développer des conversations et des interactions sociales sur internet ou en situation de mobilité ».

Selon sa définition des médias sociaux, ils présentent donc trois grands intérêts :

- Les médias sociaux permettent d'instaurer un dialogue avec sa communauté, et donc une relation concrète.
- Les médias sociaux permettent de développer des interactions sociales (« like », « retweet », « partage »...) révélant un engagement de la part d'une communauté et entraînant de la viralité.
- Les médias sociaux permettent de communiquer auprès de sa communauté à tout moment, même en situation de mobilité. A l'heure des smartphones et tablettes, votre cible est perpétuellement connectée et donc toujours susceptible de recevoir des informations. [1]

1.2 ÉVOLUTION DES MÉDIAS SOCIAUX

À la fin des années 1990, au fur et à mesure que s'est accrue la popularité de l'Internet à large bande, des sites Web permettant de créer et de télécharger du contenu ont commencé à apparaître. Le premier site de réseautage social (SixDegrees.com) date de 1997. À partir de 2002, les sites de réseautage social se sont multipliés. Certains – comme Friendster – ont créé un véritable engouement initial, qui s'est cependant vite dissipé. D'autres se sont trouvés des créneaux, comme MySpace, qui a séduit les adolescents friands de musique.

À la fin des années 2000, les médias sociaux étaient largement entrés dans les mœurs, et certains services ont vu le nombre de leurs adeptes grossir de façon exponentielle. Ainsi, en novembre 2012, Facebook annonçait qu'il comptait un milliard d'utilisateurs dans le monde. En juillet 2012, on estimait à 517 millions le nombre d'utilisateurs Twitter.

Plusieurs facteurs ont contribué à la croissance rapide de l'utilisation des médias sociaux : des facteurs technologiques, comme l'accès croissant à la large bande, l'amélioration des outils logiciels et la puissance toujours plus grande des ordinateurs et des appareils mobiles, des facteurs sociaux, comme l'engouement instantané de la part des groupes d'âge plus jeunes, des facteurs économiques, comme le prix de plus en plus abordable des ordinateurs et des logiciels et l'intérêt commercial croissant que suscitent les sites de médias sociaux [2].

1.3 Les cinq piliers des medias sociaux

Nous pouvons définir les médias sociaux selon les 5 piliers suivants :

- **Participation** : Tous les efforts sont faits pour encourager les utilisateurs à contribuer à leur opinion, éliminant ainsi la barrière entre le public et les médias.
- **Ouverture** : Les médias sociaux dépendent des principes de coopération et d'échange d'informations. Tout le monde peut participer, il n'y a pas d'obstacle à l'entrée.
- **Conversation** : Alors que les médias traditionnels ont tendance à "raconter" ou à transmettre un message, le dialogue social est plus qu'une écoute attentive.
- **Communauté** : Les médias sociaux permettent de constituer rapidement des communautés de personnes partageant les mêmes intérêts et les mêmes idées.
- **Interconnexion** : La plupart des médias sociaux se développent par interconnexion en tirant partie des liens avec les autres sites, ressources ou personnes.

1.4 Les différents types de médias sociaux

Forum : est un espace de discussion asynchrone sur un site Web ou un service en ligne qui permet aux internautes d'échanger, de poser des questions ou de poster des réponses aux sujets proposés.

Blog : est une page personnelle ou d'entreprise comportant des avis, des liens ou chroniques périodiquement créés par son ou ses auteurs sous forme d'articles. Un flux RSS (Un flux RSS est une ressource Web dont le contenu est généralement produit automatiquement, en fonction des mises à jour d'un site Web) permet en général aux internautes intéressés par un blog d'être avertis des nouveaux articles.

Wiki : Une base de connaissance en ligne où les internautes rédigent et corrigent eux-mêmes le contenu. Les wikis sont constitués d'un ensemble de pages. Chaque page dispose d'un historique des modifications et peut être commentée.

Service de partage : Service en ligne où les internautes peuvent publier des photos, vidéos, liens... Chaque élément publié est rattaché à un membre et peut être commenté et noté.

Réseau social: désigne généralement l'ensemble des sites internet permettant de se constituer un réseau d'amis ou de connaissances professionnelles et fournissant à leurs membres des outils et interfaces d'interactions, de présentation et de communication

Microblog : Service de publication, de partage et de discussion reposant sur des messages très courts. La consultation des messages et profils ne requiert pas d'inscription et peut se faire sur le web.

Agrégateur : Service en ligne permettant de regrouper l'ensemble des publications d'un utilisateur des médias sociaux, de très nombreuses formes de contributions sont acceptées (RSS, photos, vidéos, liens, email...). Les utilisateurs peuvent s'abonner aux flux des autres membres.

FAQ collaborative : Le terme FAQ se traduit en français par Foire Aux Questions , qui signifie questions fréquemment posées, Les FAQ sont largement utilisés sur l'internet. Les sites internet publient une liste de questions fréquemment posées avec les réponses associées.

Jeux sociaux : Jeux en ligne reposant sur une plateforme sociale exploitant les profils des membres pour proposer différentes interactions sociales entre les joueurs.

Service de géolocalisation : Applications permettant de publier, partager et discuter sur des terminaux mobiles. Les articles ou photos publiés sont rattachés à un lieu afin de leur donner un contexte géographique. Chaque membre dispose d'un profil où sont listées ses dernières publications ainsi que les lieux qu'il a visités. Chaque lieu dispose également d'une page où sont listés les membres qui s'y sont signalés. [3]



Figure 1.1 Tous les médias sociaux [1]

2. Les réseaux sociaux

2.1 Définition

Le réseau social est un ensemble d'identités sociales telles que des individus ou encore des organisations reliées entre elles par des liens créés lors des interactions sociales. Ils se présentent par une structure ou une forme dynamique d'un groupement social.

Les réseaux sociaux sont des sites sur lesquels les membres d'un même réseau peuvent rester en contact. Ils doivent avoir un « *profil d'utilisation* » sur lequel ils peuvent partager différentes informations avec les autres. Ces informations sont accessibles à tous membres du réseau.

2.2 Historique des réseaux sociaux

- Le premier email a été envoyé en 1971 !
- En 1991, le World Wide Web fait ses premiers pas lorsque Tim Berners-Lee du CERN propose un nouveau mode de distribution de l'information.
- En 1994, un étudiant, Justin Hall lance son site Justin's Links from the Underground pour se connecter au monde extérieur. Hall a publié sur son blog pendant 11 ans et est reconnu comme le père fondateur des blogs personnels.
- En 1995, le réseau Classmates permettait aux américains de retrouver leurs anciens camarades d'école, l'ancêtre de Copains d'Avant en quelque sorte. Le site utilise les technologies pour créer des relations humaines dans le monde réel et pas uniquement sur la toile.
- En 1998, Open Diary permet à ses utilisateurs de publier leurs journaux en ligne de manière publique ou privée sans avoir de connaissances HTML. Pour la première fois, les utilisateurs peuvent aussi ajouter leurs commentaires.
- En 2000, Wikipédia fait son apparition et publie 20 000 articles dès la première année! Ce site a révolutionné notre accès à la connaissance et à l'information. Les adolescents en ont d'ailleurs fait leur meilleur ami pour leurs devoirs...
- En 2001, Meetup est un site de partage d'intérêts et de passion.
- En 2002, le réseau social Friendster est le premier site à atteindre un million d'utilisateurs
- En 2003, ce sont trois grands sites qui font leur apparition et qui vont révolutionner notre façon d'utiliser le web, que ce soit dans la sphère privée et dans la sphère professionnelle : MySpace, WordPress et bien sûr LinkedIn. En un mois, MySpace réunit 1 million d'utilisateurs. Pour WordPress, ce sont des centaines de personnes dans le monde qui travaillent ensemble pour créer un système gratuit de management de contenus. LinkedIn est un réseau social professionnel. Aujourd'hui, ce sont 200 millions de personnes qui ont un profil sur le réseau social.
- En 2004, Mark Zuckerberg lance TheFacebook.com depuis sa chambre d'étudiant d'Harvard La même année, l'équipe de Flickr lance son site de partage de photos.
- En 2005, ce sont les vidéos qui sont à l'honneur avec le lancement de YouTube. On peut désormais télécharger et partager des vidéos gratuitement avec sa famille et ses amis.
- Depuis 2006, nous avons pris l'habitude d'écrire en 140 caractères avec Twitter. La même année, Spotify permet à ses utilisateurs de partager leurs playlists musicales.

- En 2007, Tumblr permet à ses souscripteurs de partager rapidement et facilement des photos, des textes, des notes et des liens avec leurs communautés.
- Avec Foursquare, qui est apparu en 2009, vous pouvez suivre vos amis à la trace et savoir où ils se trouvent et ce qu'ils font en temps réel. Plus récemment encore, Google a voulu concurrencer les réseaux sociaux comme Facebook et Twitter et a créé Google + en 2011.
- 2011: Pinterest (Babillard des objets de ses rêves)
- 2013: SnapChat (Journal personnel éphémère en photos et vidéos)

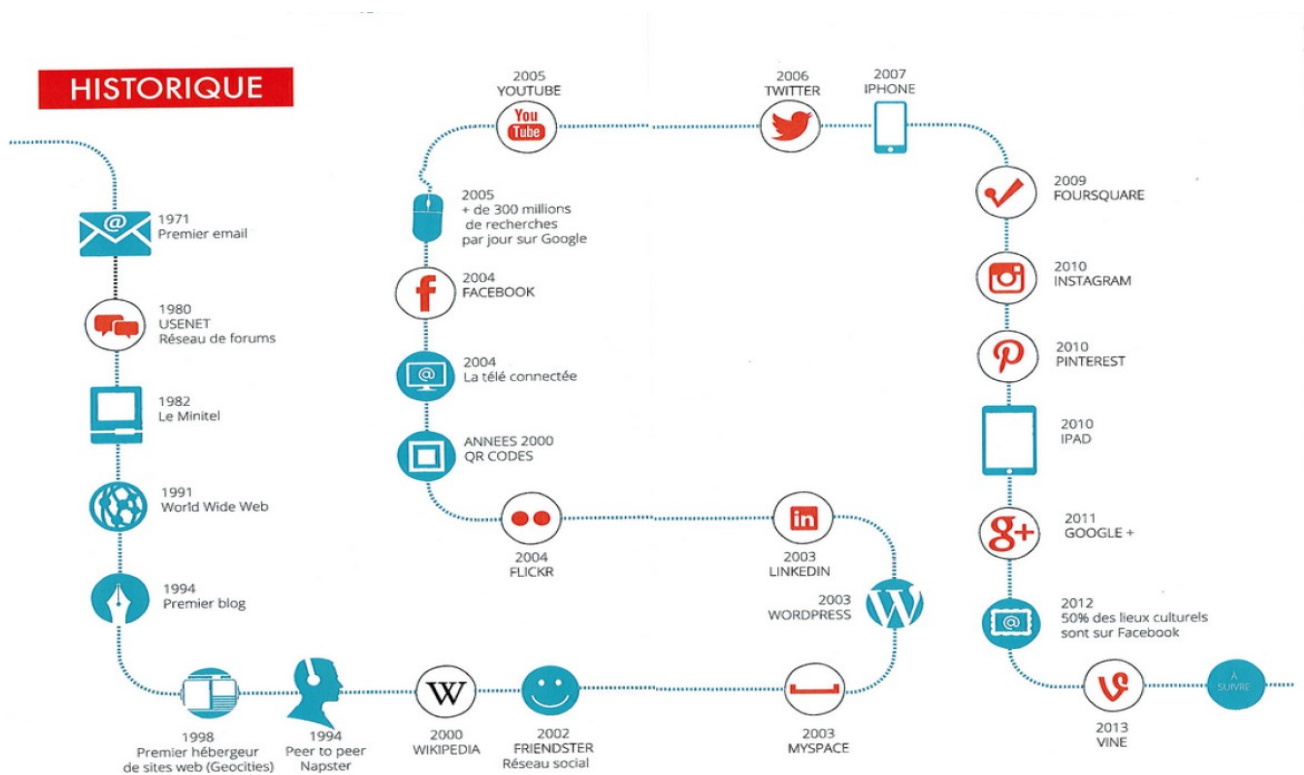


Figure 1.2 Historique des réseaux sociaux [4]

2.3 Les types de réseaux étudiés

Informatique : pages Web, routeurs, P2P, etc.

Réseaux : routage, protocoles, sécurité P2P, conception de systèmes, déviances Web , indexation, moteurs de recherche.

Biologie : protéines, neurones cérébraux, etc.

Sciences sociales : amitiés, collaboration, contacts sexuels, etc.

Économie : échanges financiers.

Histoire : mariages.

Linguistique : synonymie, co-occurrence.

Transports : réseau aérien, électrique. [5]

Exemple de réseaux :

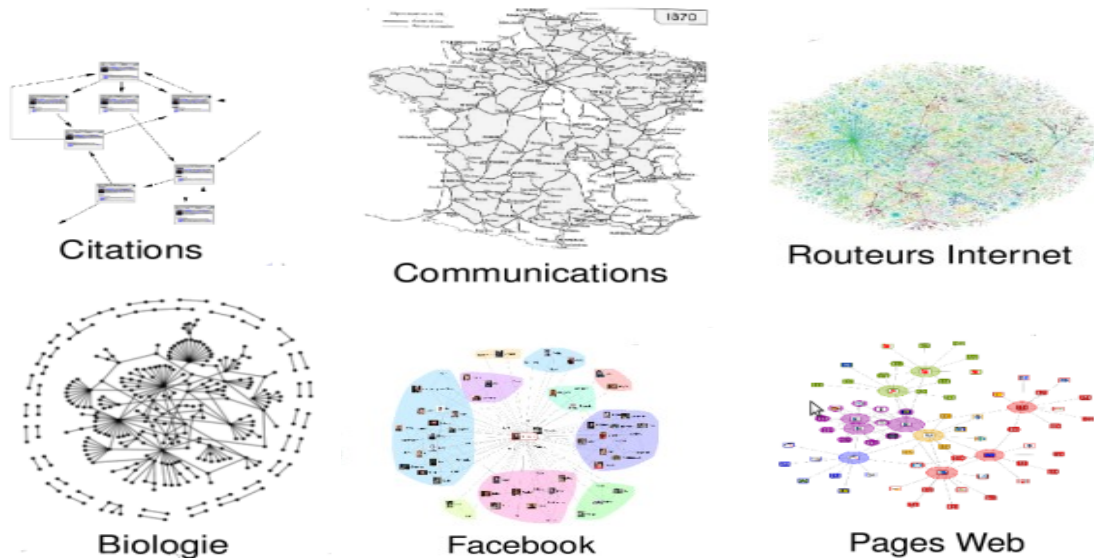


Figure 1.3 les Différents réseaux [5]

2.4 Les principaux réseaux sociaux classés par catégorie

2.4.1 Les réseaux sociaux généralistes

Facebook : Considéré comme synonyme de «réseaux sociaux» par certains, Facebook est le seul site où vous êtes susceptible de trouver des amis, des collègues et des parents autour de vous tellement la plateforme est devenu populaire. Bien que Facebook soit principalement axé sur le partage de photos, de liens et de la vie quotidienne, les utilisateurs peuvent également montrer leur soutien aux marques ou aux organisations en devenant des fans.[6]



Figure 1.4 logo de Facebook [6]

Twitter : Peut-être la plus simple de toutes les plateformes de médias sociaux, Twitter se justifie être l'un des plus fun et intéressants à utiliser. Les messages sont limités à 140 caractères ou moins, mais c'est plus que suffisant pour afficher un lien, partager une image ou même échanger des idées avec votre célébrité préférée ou un influenceur. La création d'un nouveau profil ne prend que quelques minutes et l'interface de Twitter est facile à prendre en main.[6]



Figure 1.5 logo de Twitter[6]

Google+ : En combinant le meilleur de Facebook et Twitter dans une seule plateforme, Google a donné aux utilisateurs un réseau social qui a un petit quelque chose de tout le monde. Vous pouvez ajouter du nouveau contenu, mettre en surbrillance des sujets avec des hashtags et même séparer des contacts en cercles. Même si un profil G+ ne prend que quelques minutes pour être configuré, le succès de Google+ n'est sûrement pas ce qui était espéré par la firme de MountainView. [6]



Figure 1.6 logo de Google+ [6]

Snapchat : Il s'agit d'une plate-forme sociale de partage photos et de vidéos disponible sur smartphone et qui permet de discuter avec des amis en utilisant des images. La particularité de ce réseau social est que chaque image ou vidéo envoyée ne peut être visible que durant une période de temps par son destinataire. Il vous permet d'explorer les nouvelles et même de consulter des histoires en direct qui se passent dans le monde.[6]



Figure 1.7 logo de Snapchat [6]

2.4.2 Les réseaux sociaux professionnels :

LinkedIn : Un des seuls réseaux sociaux qui est réellement orienté vers les entreprises. LinkedIn est au cyberspace ce que les groupes de réseautage ont été pour les communautés d'aires locales. Il est idéal pour rester en contact avec ses anciens collègues, rencontrer des clients, des fournisseurs, recruter de nouveaux employés et suivre les dernières nouvelles de votre secteur d'activité.[6]



Figure 1.8 logo de LinkedIn[6]

Viadeo : Viadeo est un site de réseautage social en ligne Français. Concurrent de LinkedIn, il permet de construire son réseau professionnel et facilite le dialogue entre les membres. L'une des particularités de Viadeo est de réunir des professionnels issus des petites et moyennes entreprises puisque les profils présents dans des entreprises de moins de 50 employés représentent presque la moitié des inscrits. On peut cependant déplorer que la plateforme est nettement en perte de vitesse.[8]



Figure 1.9 logo de Viadeo [6]

2.4.3 Les réseaux sociaux de blogging :

Medium : Medium est une plateforme internet de blogging minimaliste conçue pour des textes longs. L'objectif du site est que la plateforme offre une visibilité maximale aux textes indépendamment de leur auteur, pour dépasser la difficulté de faire connaître son blog. Résultats, on trouve des milliers d'articles très utiles et très intéressants tout en restant agréables à lire.[6]

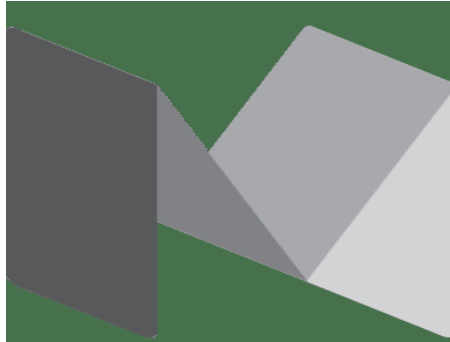


Figure 1.10 logo de Medium [6]

Tumblr : Cette plateforme est différente de beaucoup d'autres puisqu'elle héberge essentiellement des microblogs pour ses utilisateurs. Les individus et les entreprises, peuvent construire de manière très simple et intuitive leurs blogs avec du multimédia (comme des images et des clips vidéo courts). La force de Tumblr est de s'appuyer principalement sur le reblogage et donc de démultiplier la viralité des contenus.[6]



Figure 1.11 logo de Tumblr [6]

2.4.4 Les réseaux sociaux de vidéo :

YouTube : En tant que service de partage de vidéos, YouTube est devenu si populaire que son catalogue de milliards et de milliards de vidéos est devenu le «deuxième moteur de recherche mondial». Les utilisateurs peuvent partager, évaluer et commenter toutes les vidéos qu'ils visionnent. La plateforme qui a été rachetée par Google et faisant parti à part entière de son écosystème est le leader incontestable des réseaux sociaux de vidéo.[6]



Figure 1.12 logo de YouTube [6]

Periscope : Periscope est à la fois une application pour appareils mobiles et une plateforme pour appareils mobiles qui permet aux utilisateurs de retransmettre en direct ce qu'il est en train de filmer. Appartenant désormais à Twitter, les utilisateurs doivent se connecter à ce dernier puis peuvent diffuser en streaming des vidéos filmées n'importe où dans le monde, et de permettre à d'autres utilisateurs de les visualiser durant les 24 heures suivant la fin de l'enregistrement.[6]

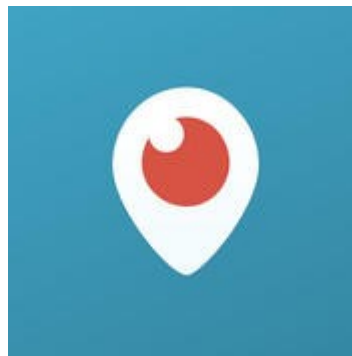


Figure 1.13 logo de Periscope [6]

2.4.5 Les réseaux sociaux “visuels”

Instagram : Instagram permet de partager ses photographies ou vidéos prises avec son smartphone avec ses amis sur la plateforme ou via les autres réseaux sociaux notamment Facebook qui l'a racheté. Vous pouvez aussi choisir parmi une variété de filtres photo et inviter des amis à les commenter.[6]



Figure 1.14 logo d'Instagram[6]

Pinterest : Servant de gigantesque idée virtuelle d'inspiration, Pinterest a joué très grand rôle dans les médias sociaux au cours des dernières années. Particulièrement populaire auprès des femmes et de la communauté du "Do It Yourself", la plateforme vous permet de partager des photos, et toute sorte de visuels créatifs que d'autres peuvent épingler, enregistrer ou dupliquer.[6]



Figure 1.15 logo de Pinterest [6]

Flickr : Flickr est un site web de partage de photographies et de vidéos très utilisé par des photographes professionnels ou amateurs. Les utilisateurs peuvent éditer leurs photos en ligne, les gérer, les partager et obtenir des licences Créative Commons.[6]



Figure 1.16 logo de Flickr[6]

2.4.6 Les réseaux sociaux communautaires

Reddit : Reddit est un site communautaire de partage de signets permettant aux membres de soumettre leurs liens et de voter pour les contenus proposés par les autres utilisateurs. Fortement ancrée dans la culture web, la plateforme est un des réseaux sociaux les plus influents.[6]



Figure 1.17 logo de Reddit[6]

2.5 Le top des plateformes en 2019

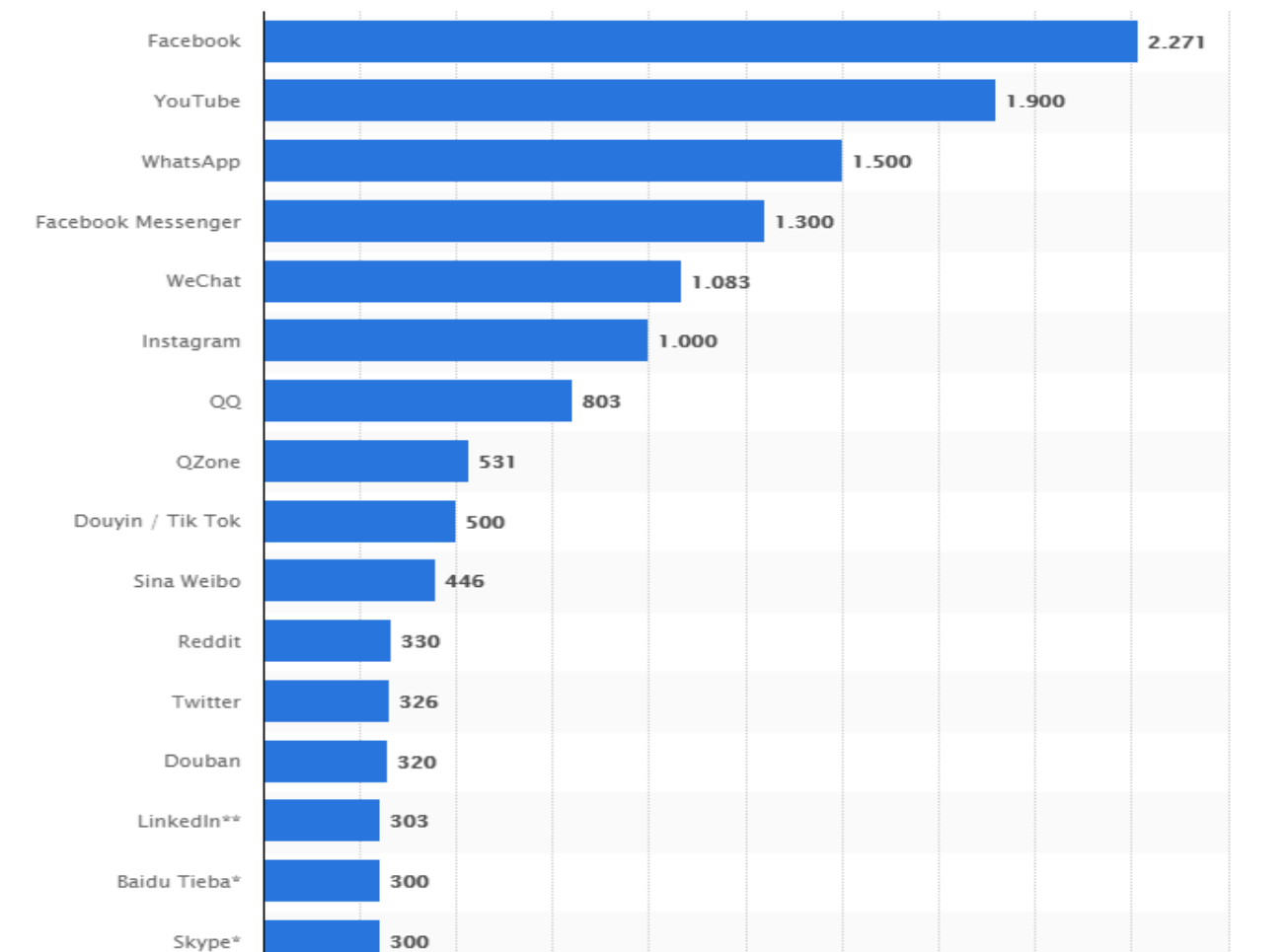


Figure 1.18 Le top des plateformes en 2019 [7]

2.6 L'utilité des réseaux sociaux

Les réseaux sociaux sont utilisés par deux catégories d'individus dans des buts différents : Tout d'abord par les particuliers, ils utilisent les réseaux sociaux pour partager des informations, des liens, des centres d'intérêts, rechercher des personnes, publier des photos, des vidéos, etc...

Puis par les professionnels qui eux les utilisent pour générer du business, faire connaître leur entreprise, leurs services, suivre les tendances actuelles, optimiser leurs recrutements, etc...

- Pour le contact et la communication

Un réseau social est un espace permettant aux personnes inscrites de communiquer entre elles. C'est une plateforme de discussion, de publications personnelles, de vidéos de tout type qui permettent de se divertir, de s'informer, de commenter. C'est un avantage pour la communication, de part la gratuité de son utilisation et de sa facilité d'accès. On peut communiquer avec des amis ou des membres de la famille de partout dans le monde de manière instantanée. En outre, les réseaux sociaux, comme Facebook, sont un bon moyen de rester en contact avec vos amis anciens et nouveaux. On peut partager ses pensées, photos, vidéos, etc....., avec tous ses amis en un seul clic.

- Pour développer une entreprise :
 - Améliorer la visibilité de l'entreprise

La présence des entreprises et des marques sur les réseaux sociaux est de plus en plus généralisée. Ces sites représentent un vecteur de communication au même titre que les médias classiques, mais présentent un coût moindre et un retour sur investissement beaucoup plus intéressant.

- Augmenter le nombre de prospects, fidéliser les clients

Les réseaux sociaux représentent un outil marketing à bas coût permettant de dynamiser les ventes et d'élargir la clientèle. Facebook et Twitter sont très pratiques pour annoncer la sortie d'un nouveau produit ou pour lancer une opération événementielle (par exemple, des réductions accordées uniquement aux personnes connectées aux réseaux sociaux).

L'information sera rapidement relayée entre les membres du réseau si bien que le phénomène généré permettra d'attirer de nouveaux prospects. Par ailleurs, l'utilisation du Web se voulant de plus en plus interactive, les internautes (surtout les particuliers) n'hésitent pas à donner leur avis sur tel produit ou telle société. Cela permet donc d'avoir accès à des retours d'expérience utilisateurs intéressants sans avoir à faire appel à des enquêtes ou des panels plus compliqués à mettre en place.

- Optimiser les recrutements

De plus en plus de recruteurs utilisent les réseaux sociaux pour embaucher leurs futurs collaborateurs :

Il est possible de publier une offre d'emploi sur certains réseaux sociaux pour toucher soit un lectorat très large, soit, au contraire, une cible très spécifique.

On peut adopter une démarche plus proactive et scruter les différents profils sur Internet pour dénicher le candidat idéal. [8]

2.7 Les dangers des réseaux sociaux

Internet est un avantage pour ceux qui souhaitent apprendre, améliorer leur culture, découvrir de nouveaux domaines ou communiquer. Cependant, Internet représente une multitude de dangers envers ses usagers et notamment pour les jeunes qui sont les plus vulnérables car ils n'ont pas le recul ni l'expérience leur permettant de discerner une situation à risque ou un contenu potentiellement nuisible, il peut s'agir de sites au contenu explicite ou de la rencontre de personnes malintentionnées.

- Le harcèlement en ligne ou cyber-harcèlement

Une nouvelle forme de harcèlement se répand : le harcèlement sur Internet ou cyber-harcèlement. Ce problème est apparu lors de la survenue des réseaux sociaux sur Internet, nouvel outil utilisé pour faire de l'intimidation : diffusion de textes, d'images ou de films malveillants ou diffamatoire. C'est un acte agressif, intentionnel commis par un individu ou par un groupe d'individus par le biais des communications électroniques.

- L'usurpation d'identité

Les réseaux sociaux sont aussi théâtre d'usurpation d'identité. L'usurpation d'identité consiste à prendre délibérément l'identité virtuelle d'une personne existante, en soutirant son mot de passe ainsi que son identifiant et de se faire pour elle en utilisant son compte. L'usurpateur peut également créer un nouveau compte au nom de sa victime et y dresser un profil avec des photos détournées, il peut ainsi se faire passer pour elle.

- Risque de cyberdépendance

La dépendance à Internet (ou cyberdépendance) est un trouble qui se traduit généralement par un usage pathologique et abusif d'Internet ainsi que des réseaux sociaux. Cette dépendance se caractérise par des comportements incontrôlables, les personnes atteintes expriment un besoin compulsif de se connecter à Internet pour maintenir une activité passive ou active en ligne. La personne atteinte montre des signes d'anxiété et d'agressivité et souffre de l'impact de son addiction sur sa vie sociale, professionnelle et privée.[8]

Conclusion

Les médias sociaux permettent aux entreprises, aux organisations, aux gouvernements de communiquer avec un grand nombre de personnes. En même temps, ils changent les interactions humaines. Par ailleurs, il existe des conséquences du partage des renseignements personnels. Dans le chapitre suivant, nous présentons la notion d'analyse des réseaux sociaux en évoquant ses principaux traits.

CHAPITRE 2

L'ANALYSE DES RESEAUX SOCIAUX

Introduction

La méthode structurale appelée analyse des réseaux sociaux est une méthode de description et de modélisation de la structure relationnelle d'un ensemble d'acteurs, sans hypothèse a priori .Elle part de l'observation de la présence ou de l'absence de relations entre les membres d'un groupe pour reconstituer un système de relations.

Dans ce chapitre Nous présentons ce qu'est l'analyse des réseaux sociaux en évoquant notamment les concepts d'analyses qui le déterminent .Ensuite nous allons représenter les différents propriétés des réseaux sociaux, ainsi que les méthodes d'analyse de ces dernier.

Enfin dans la dernière section on va clôture se chapitre en présentons utilisation de l'analyse des réseaux sociaux.

1. Analyse des réseaux sociaux

1.1 Définition

L'analyse de réseaux sociaux s'appuie sur les acquis de la théorie des graphes pour formaliser le réseau social comme un ensemble de nœuds et de liens où chaque nœud modélise un acteur et chaque lien une relation entre deux acteurs. Une valeur peut être affectée à un lien et représentera alors la force de celui-ci. Elle peut servir à représenter l'importance d'une relation, que ce soit en comptant simplement le nombre d'occurrences de cette relation, ou en prenant en compte d'autres processus de pondération (qualité de l'interaction, système d'évaluation, ...).

1.2 Les concepts d'analyse des réseaux sociaux

Selon Wassermann et Faust, trois concepts sont également retenus dans cette analyse des réseaux sociaux :

1. Les acteurs et leurs actions sont considérés comme des entités indépendantes.
2. L'environnement des acteurs procure des opportunités et exerce des contraintes sur leurs actions individuelles.
3. Les structures sociales, politiques, économiques, ...etc. ont une influence sur les formes de relations entre les acteurs. [9]

1.3 Les représentations des réseaux sociaux

Les concepts clés de l'analyse de réseaux sociaux se base sur deux représentations : la représentation nœuds-liens et la représentation matricielle.

- La représentation nœuds-liens

La première personne à avoir représenté un réseau social est Jacob Levy Moreno au début des années 1930 [10]. Son objectif étant de visualiser graphiquement un réseau social, il a représenté les personnes par des points et une relation entre deux personnes par des flèches. Moreno a ainsi introduit le concept d'étoile pour désigner les personnes ayant le plus de relations dans un réseau social, en référence à l'étoile formée par un point et ses connections.

Les mathématiciens ont rapidement fait le rapprochement entre les représentations sociogrammes (est un schéma qui illustre un réseau social. Il peut aussi être connu sous le nom de diagramme de réseau social ou graphe) et la théorie des graphes au sens mathématique. Au milieu du vingtième siècle, Cartwright et Harary sont les premiers à avoir appliqué la théorie des graphes à l'analyse des réseaux sociaux. Le graphe est devenu par la suite la représentation adoptée par toutes les sciences manipulant l'analyse des réseaux sociaux, dont la sociologie, les mathématiques et l'informatique.

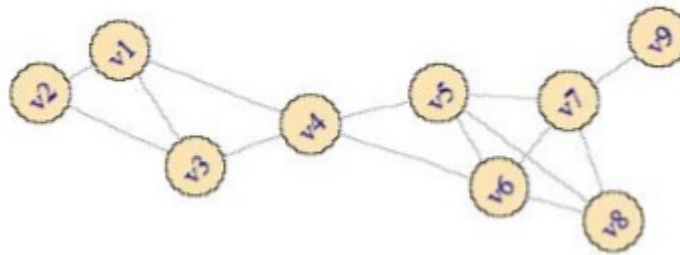


Figure 2.1 un exemple d'un graphe [11]

- La représentation matricielle

La représentation matricielle des graphes est une représentation beaucoup moins utilisée que les diagrammes nœuds-liens. Un graphe peut être représenté sous la forme d'une matrice A à n lignes et n colonnes représentant un tableau. Chaque case de ce tableau est notée a_{ij} avec i et j les numéros respectifs de ligne et de colonne de la case.

La valeur contenue dans la case a_{ij} est le poids de la relation entre les ressources v_i et v_j (égal à 1 dans le cas d'un graphe non pondéré), 0 correspond à une absence de relation. On distingue deux types de matrices dans un réseau social, les matrices d'incidence et les matrices d'adjacence.

On parle de matrice d'adjacence lorsqu'on a les mêmes ressources en ligne et en colonne, on obtient ainsi une matrice carrée avec la ligne i et la colonne i représentant la même ressource.

S	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	0	1	1	1	0	0	0	0	0
v2	1	0	1	0	0	0	0	0	0
v3	1	1	0	1	0	0	0	0	0
v4	1	0	1	0	1	1	0	0	0
v5	0	0	0	1	0	1	1	1	0
v6	0	0	0	1	1	0	1	1	0
v7	0	0	0	0	1	1	0	1	1
v8	0	0	0	0	1	1	1	0	0
v9	0	0	0	0	0	0	1	0	0

Figure 2.2 matrice d'adjacence [11]

1.4 Quelques définitions et indicateurs

- Le graphe est un ensemble $G=\{V,E\}$ où V désigne les sommets et E les arêtes.
- La taille du graphe est le nombre de liens.
- L'ordre du graphe est le nombre de sommets.
- Voisinage N_i d'un sommet = l'ensemble des nœuds qui lui sont directement connectés.
- Une géodésique est l'un des plus courts chemins entre deux sommets donnés.
- Le diamètre d'un graphe est le plus long chemin géodésique de ce graphe.

- Un graphe est dit vide lorsqu'il ne contient aucun lien.
- Un graphe est dit complet lorsque tous les sommets sont connectés à tous les autres. Il y a alors $\frac{n(n-1)}{2}$ liens dans un graphe complet d'ordre n .
- Un graphe orienté est un ensemble de sommets et d'arêtes, chaque arête étant un couple de sommets ordonnés. Ainsi, la relation entre les sommets x et y est différente de celle entre y et x .

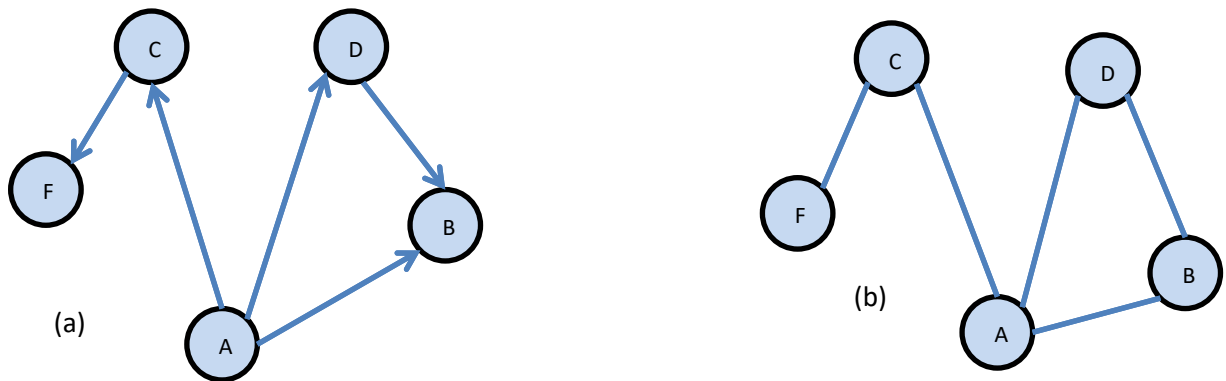


Figure 2.3 (a) illustre un graphe orienté, (b) montre un graphe non orienté.

- Un graphe valué, par opposition à un graphe non valué, comporte des liens multiples (deux sommets sont liés plusieurs fois). Dans cette communication, on se limitera à des graphes non orientés, dans les quelles les relations entre sommets sont de fait symétriques. Un graphe simple est un graphe non valué et sans boucle (sans arête d'un sommet vers lui-même).
- Graphe biparti Un graphe $G = (V,E)$ est dit biparti si et seulement si il existe une partition (V_1,V_2) de V telle que toute arête $e \in E$ a une extrémité dans V_1 et l'autre dans V_2 .
- Le degré d'un sommet est le nombre de sommets auquel il est relié. Dans un graphe simple d'ordre n , le degré d'un sommet est compris entre 0 et $n-1$.
- La densité La densité d'un réseau est une mesure de sa connectivité. Plus un graphe est dense, plus il est connecté. La densité indique la quantité de liens au sein d'un réseau et permet de définir la cohésion d'un réseau social.

Pour mesurer la densité dans un réseau, on divise le nombre total de connexions par le nombre de connexions possibles.

$$\Delta = \frac{L}{M(M-1)/2} \dots\dots\dots 2.1$$

Où $M(M-1)/2$ est le nombre maximal de connexions possibles pour un réseau de M nœuds.

Un réseau où la densité vaut un est dit complet, et L le nombre des liens .

- Graphe connexe Il existe au moins un chemin entre chaque paire de nœuds. Il n'y a pas de sommets isolés.
- Un composant (dans un graphe déconnecté) est un ensemble de groupes connectés de nœuds. Il peut être considéré comme un graphe connexe isolément. La notion de composants est utile pour un raisonnement sur les réseaux et sur une échelle beaucoup plus petite.
- Graphe trivial c'est un graphe qui contient un seul sommet.
- Un graphe aléatoire de taille n est un graphe à n sommets dont on a choisi aléatoirement les arêtes.
- Graphe pondéré désigne un graphe avec des arêtes pondérées, si un certain nombre (appelée poids) est affecté à chaque bord d'un graphe.
- Ces poids peuvent représenter des coûts, des longueurs ou des capacités, en fonction du problème étudié. La somme de tous les poids est le poids du graphe. Ce graphe est adopté aux réseaux sociaux qui contiennent différents niveaux d'intensité dans les relations
- Chemin Un chemin permettant d'aller de v_i à v_j est la succession de sommets et arêtes qui les deux sommets. De v_i à v_j , plusieurs chemins sont possibles, il existe un plus court chemin minimisant le nombre d'arêtes parcourus.
- Chaîne suite finie de sommets reliés entre eux par une arête.
- Chaîne eulérienne chaîne simple passant par toutes les arêtes d'un graphe.
- Chaîne hamiltonienne chaîne simple passant par tous les sommets d'un graphe une et une seule fois.
- Cycle chaîne qui revient à son point de départ.

- Matrice d'incidence. La matrice d'incidence M est une matrice binaire $N \times P$ où l'élément $M_{i\alpha}$ de M est égal à 1 si l'arc α admet le sommet i comme origine, et 0 autrement.
- Matrice d'adjacence : La matrice d'adjacence A est une matrice $N \times N$ dont les colonnes et les lignes sont associées aux nœuds du graphe $G = (V,E)$. L'élément a_{ij} de cette matrice représente le poids de l'interaction entre les nœuds v_i et v_j .
- Matrice des degrés La matrice des degrés est une matrice diagonale avec le degré des nœuds du graphe sur les diagonales.
- La matrice Laplacienne qui se définit telle que : $L = D - A$ où D est la matrice des degrés, et A la matrice d'adjacence.
- Force d'un lien entre deux sommets peut être évalué à partir de l'importance du recouvrement (overlap) entre leurs voisinages. Plus il (le recouvrement) est élevé, plus le lien est fort.

$$\text{overlap}(v_i, v_j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j| - 2} \dots \dots \dots 2.2$$

$|N_i \cap N_j|$: Voisinage commun.

$|N_i \cup N_j|$: Voisinage total.

-2 : Ne pas prendre en compte les 2 sommets.

2. Propriétés des réseaux sociaux

L'étude des Réseaux Sociaux a révélé que ceux-ci partagent des propriétés formelles communes telles que l'effet de petits mondes, la distribution des degrés en loi de puissance et la structure communautaire

2.1 Effet de petits mondes

La principale caractéristique d'un réseau social est «l'effet petit monde», mise en évidence historiquement par la fameuse expérience de Milgram [12], et qui exprime le fait que les graphes ont souvent des diamètres très faibles. Le plus court chemin entre deux nœuds dans un réseau social de taille n est de l'ordre de $\log(n)$. Newman a montré que les individus d'un même réseau social possèdent la faculté de trouver facilement ces plus courts chemins.

2.2 Distribution des degrés en loi de puissance

Dans les Réseaux Sociaux, ils existent beaucoup de sommets de faibles degrés et très peu de sommets de forts degrés. En effet, la distribution des sommets suit une loi de puissance $p(k) \sim k^{-\gamma}$, où k est le degré et γ l'exposant de la loi de puissance. Cet exposant est en pratique entre 2 et 3 être présente la vitesse de décroissance de la courbe des degrés. Plus γ est grand, plus la probabilité d'obtenir des sommets de forts degrés est petite.[15]

2.3 Structure communautaire

Pour une meilleure compréhension des réseaux sociaux, les chercheurs sont essayés de trouver plus de caractéristiques structurelles de ces réseaux. Et ont montré quels réseaux réels ne sont pas des graphes aléatoires car ils présentent de grandes hétérogénéités, révélant un haut niveau d'ordre et d'organisation. En outre, la répartition des arêtes n'est pas seulement globale, mais aussi localement non homogène, avec des concentrations élevées dans des groupes spéciaux de sommets et de faibles concentrations entre ces groupes, Cette caractéristique des réseaux réels s'appelle la structure communautaire.

En effet, les graphes représentant les réseaux sociaux du monde réel présentent une structure modulaire avec des nœuds formant des groupes et éventuellement des groupes au sein de groupes. [13]

3. Méthodes d'analyse des réseaux sociaux

Tous ces réseaux sociaux amassent de très nombreuses données les amis, les messages, les images, la fréquence d'utilisation, etc. Tous ces échanges et informations sont soigneusement enregistrés. Dès lors se pose le problème de l'exploitation de cette masse d'informations.

Il faut tout d'abord modéliser le réseau sous forme mathématique. La structure de base est bien entendu le graphe : l'analyse des figures produites permet de tirer un grand nombre d'informations, et aussi de prédire en partie l'évolution future du réseau.

Des mises à jour sont effectuées, pour offrir des méthodes d'analyse flexibles qui tiennent compte de toutes les informations en termes de structure et de contenu. On distingue alors deux grandes familles : traditionnelle et fouille de données. [14]

3.1 Méthodes traditionnelles

- **La centralité** : Les acteurs importants sont ceux qui sont liés et impliqués avec les autres acteurs d'une façon extensive. Après avoir modélisé les contacts par des liens

dans un réseau, nous pouvons définir un acteur central comme étant celui qui est impliqué dans plusieurs liens.

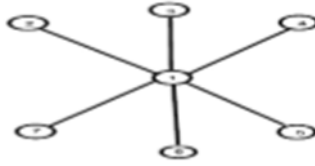


Figure 2.4 réseau sociaux centre [15]

Il existe différents types de liens entre les nœuds (acteur), ce qui a permis de générer plusieurs types de centralités dont :

- **Degré de Centralité** : Soit n le nombre total de nœud (acteurs) dans un réseau (graphe). Le degré de centralité d'un acteur i noté $CD(i)$ est le degré du nœud acteur (le nombre d'arêtes) noté $d(i)$ normalisé par le degré maximal $n-1$. [15]

$$CD(i) = \frac{d(i)}{(n-1)} \dots \dots \dots 2.3$$

- **La centralité de proximité** (closeness centrality) indique si le sommet est situé à proximité de l'ensemble des sommets du graphe et s'il peut rapidement interagir avec

$$c_c(v) = \frac{1}{\sum_{u \in V \setminus \{v\}} d_G(u,v)} \dots \dots \dots 2.4$$

Ces sommets. Il s'écrit formellement :

avec $d_G(u,v)$ la distance entre les sommets u et v . [15]

- **La centralité d'intermédiarité** (betweenness centrality) est un des concepts les plus importants. Il mesure l'utilité du sommet dans la transmission de l'information au sein du réseau. Le sommet joue un rôle central si beaucoup de plus courts chemins entre deux sommets doivent emprunter ce sommet. Elle s'écrit :

$$C_B(v) = \sum_{i \neq j \neq v} \frac{\sigma_{ij}(v)}{\sigma_{ij}} \dots \dots \dots 2.5$$

Avec $\sigma_{ij}(v)$ le nombre de chemins entre i et j qui passent par v . [15]

- **Le prestige**

La notion de prestige est une autre manière de mesurer l'importance d'un nœud.

Un nœud prestigieux est un nœud à lequel un grand nombre d'autres nœuds se lient- il reçoit un grand nombre de liens entrants. Nous distinguons les mesures suivantes :

- **Le degré de prestige** : avec $dI(i)$ le degré entrant du nœud i , n le nombre nœuds dans le réseau, le degré de prestige est donné par la relation :

$$PD(i) = \frac{dI(i)}{(n-1)} \dots \dots \dots 2.6$$

- **Le prestige de tri** : Les mesures proposées jusqu'au là sont fondées sur les liens entrants et sortants d'un acteur donné. La mesure du prestige du tri considère la réputation et l'importance des acteurs choisissant l'acteur i . Ce type d'algorithme est utilisé pour trier les résultats de recherche comme Google (algorithme Rank Page).[15]

3.2 Fouille de données dans les réseaux sociaux

Les réseaux sociaux amassent de très nombreuses données : les amis, les messages, les images, la fréquence d'utilisation...etc. Dès lors se pose le problème de l'exploitation de cette masse d'informations. La fouille de données s'avère être un outil incroyablement riche et puissant lorsqu'il est appliqué aux réseaux sociaux, puisque cette méthode nous permet de modéliser le réseau sous forme mathématique, l'analyser pour tirer le maximum d'information mais aussi prédire en partie l'évolution future du réseau.

La fouille de données ou Datamining est l'ensemble des méthodes scientifiques destinées à l'exploration et l'analyse de grande quantité de données informatiques en vue de détecter des profils-type, des comportements récurrents, des règles, des liens, des tendances inconnues, des structures particulières restituant de façon concise l'essentiel de l'information utile pour l'aide à la décision. [16]

Appliquer cette technique dans notre cas, cette technique va nous permettre d'extraire depuis les réseaux sociaux, les données relatives à un client puis les analyser et classer. Le datamining permet d'accomplir les quatre types d'analyse suivant : Classification, Estimation, Segmentation, Préviation. Ces types d'analyse se répartissent dans deux catégories descriptives et prédictives comme illustré dans le tableau suivant :

Techniques descriptives	Techniques prédictives
Classification	- Estimation. - Segmentation. - Préviation.

Table 2.1 Type d'analyse du Datamining

- **Les techniques descriptives :** (ex : la classification) permettent de décrire, résumer, synthétiser et classer les ces techniques essaye de mettre en évidence des informations présentes mais cachées par le volume des données. Par contre, les techniques prédictives essayent d'extrapoler des nouvelles informations à partir des données présentées.
- **La prévision (technique prédictive) :** Consiste à estimer une valeur future d'un champ à partir des données réelles possédées. Les méthodes de classification et d'estimation peuvent être utilisées en prédiction. Dans notre cas, la prévision va nous servir de prévenir le comportement d'une personne dans le futur, les produits qui vont lui intéresser dans le futur. Cette technique est très utilisée dans le domaine d'intelligence artificielle, spécialement dans les algorithmes machine Learning.[15]

4. Utilisations de l'analyse des réseaux sociaux

La grande masse d'informations conservées dans les réseaux sociaux peuvent être très utiles dans de nombreux domaines. Elles peuvent servir à surveiller la marque de son entreprise, en lui offrant des informations sur: l'opinion du public sur ses produits, avoir une idée sur l'état actuel du marché, les entreprises concurrentielles, ainsi acquérir une nouvelle stratégie de marketing car grâce à ses données elle pourra enfin établir une comparaison des produits existants dans le marché, assurer une meilleure gestion de gamme de produits, permettre un meilleur soutien à sa clientèle, pouvoir suivre les influenceurs.

Ces données peuvent aussi être utilisées lors des élections pour voir l'opinion du public.[15]

Conclusion

Dans ce chapitre nous avons défini le concept d'analyse des réseaux sociaux ,ainsi que ces propriétés .Par la suite nous avons introduit les méthodes qui permettent l'analyse des réseaux sociaux, puis les utilisation de l'analyse des réseaux sociaux.

Dans le chapitre suivant nous introduisons la notion de détection de communauté dans les réseaux sociaux, nous citons par la suite la classification des algorithmes, puis proposer les différents approches utiliser.

CHAPITRE 3

LA DETECTION DE COMMUNAUTES

ETAT DE L'ART

Introduction

Nous présentons dans ce chapitre le principe fondamental des communautés qui constituent l'objet de ce mémoire. Ensuite, nous évoquons dans les sections qui suivent, la notion de détection des communautés, ainsi que la classification des méthodes de détection des communautés, ensuite les mesures de qualité d'une partition . Dans la dernière section on a présenté les différents algorithmes de détection des communautés.

1. Communauté

La notion de communautés dans les graphes n'a pas de définition formelle. Cependant, l'existence de zones plus densément connectées que d'autres est le résultat d'une présence de structures de graphes dont les nœuds se sont regroupés en communautés du fait de leur ressemblance ou de leurs intérêts communs. Cette ressemblance ou ce partage d'intérêt peut avoir des interprétations différentes selon la nature et le type du réseau d'interaction considéré (réseaux sociaux, réseaux biologiques, etc).

Nous allons donner ici deux définitions des communautés, l'une sémantique et l'autre structurelle.

- **Définition sémantique** : Une communauté est un ensemble de nœuds qui partagent les mêmes centres d'intérêt ou ayant le même profil.
- **Définition structurelle**: Une communauté est un ensemble de nœuds fortement liés entre eux et faiblement liés avec les autres nœuds du graphe. [17]

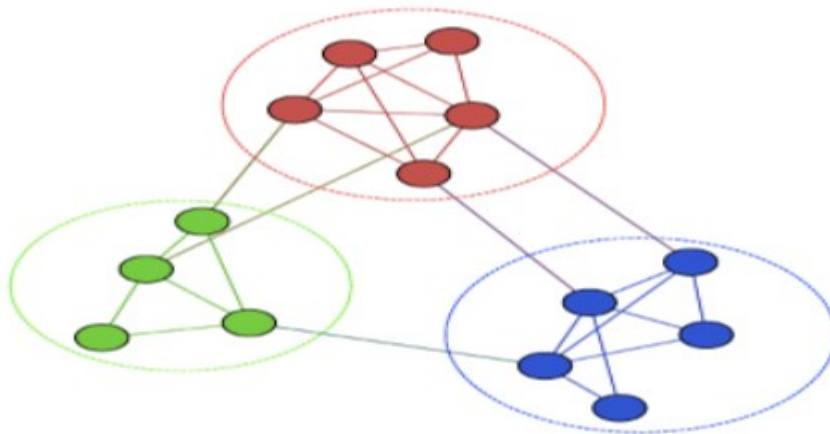


Figure 3.1 Communautés dans un graphe [17]

2. Détection des communautés

La recherche des partitions en communautés est l'un des principaux problèmes liés à l'analyse des réseaux sociaux. Les méthodes de sa résolution ont fait l'objet de nombreux travaux, depuis l'article fondateur de Girvan et Newman .

Cependant, il reste encore plusieurs questions en suspens. En effet, les réseaux sociaux, qui sont des réseaux complexes, présentent en général des propriétés topologiques non triviales, contrairement aux graphes aléatoires.

Ces propriétés caractérisent la connectivité du réseau et impactent la dynamique des processus qui y sont appliqués.

De plus, les Réseaux Sociaux sont parmi les plus gros producteurs de données. Les utilisateurs y postent, consultent et échangent des données de toutes sortes. Donc, l'identification des communautés dans les réseaux sociaux s'avère un problème complexe qui a été abordé sous différentes perspectives. En général, les techniques de détection de communautés d'intérêt se réfèrent à une classification des nœuds du réseau plus densément connectés que d'autres, pour construire des classes connexes d'utilisateurs ayant les mêmes caractéristiques au regard d'une mesure de similarité se référant à des intérêts communs.[17]

3. Classifications des méthodes de détection des communautés

Plusieurs classifications des méthodes proposées pour la détection des communautés d'intérêt dans les Réseaux Sociaux ont été publiées. La plupart de ces classifications sont basées sur les types des algorithmes de détection des communautés et leurs principes méthodologiques. Notons que, aujourd'hui, détection de communautés, partitionnement de graphe et clustering en anglais sont souvent utilisés indifféremment. Fortunato [18] a mené une étude exhaustive et classé les méthodes en huit catégories, qui sont les suivantes :

- 1 **Méthodes traditionnelles** : Consistent au partitionnement optimal en k partitions ou "clusters", k étant donné, des graphes représentant les Réseaux Sociaux. La méthode la plus connue est l'algorithme de Kernighan-Lin[19]. La solution proposée est de chercher des partitions souvent de même taille. Cette contrainte étant trop stricte et difficile à atteindre dans des cas réels, elle a été relâchée de manière à chercher des communautés mais sans avoir à préciser la taille exacte [20]. Le partitionnement hiérarchique se base

alors sur une fonction de similarité de telle sorte que les “clusters” contiennent des nœuds avec des similarités fortes.

- 2 **Algorithmes divisifs** : Ils sont basés sur la recherche d'une propriété des liens intercommunautaires pour pouvoir les identifier afin de les éliminer. Ces éliminations déconnectent le graphe pour former des composantes connexes qui représentent les communautés. L'algorithme le plus populaire est celui proposé par Girvan et Newman [21].
- 3 **Méthodes basées sur la modularité**: La modularité, proposée par Girvan et Newman, est utilisée par plusieurs algorithmes comme fonction de qualité en l'optimisant ou la modifiant telles que les techniques gloutonnes et d'autres fonctions d'optimisation.[22] Cependant, ont démontré que trouver la partition optimale en modularité est un problème NP-complet. Donc d'autres alternatives, comme l'algorithme de Louvain, sont basées sur des techniques gloutonnes pour apporter des solutions acceptables en temps de calcul.
- 4 **Algorithmes de partitionnement spectral** : Ces algorithmes sont basés sur la notion du spectre définissant la proximité entre les nœuds. Les vecteurs propres, agissant comme des propagateurs de temps pour le processus de marche aléatoire dans le graphe du Réseau Social, associés aux valeurs propres les plus faibles décrivent les groupes à similarité interne forte.[23] Généralement la matrice Laplacienne est utilisée comme matrice de similarité.
- 5 **Algorithmes dynamiques** : Simulent des processus dynamiques où les particules s'influencent entre elles. Ainsi, les particules proches les unes des autres ont tendance à partager le même état. Parmi les processus appliqués aux graphes des Réseaux Sociaux nous citons Spin-Spin et la synchronisation où le système unifie progressivement tous ses éléments au même état [24].
- 6 **Méthodes basées sur l'inférence statistique** : Comme l'inférence Bayésienne, y compris les modèles génératifs et la modélisation de blocs. Ces méthodes supposent que le graphe a été généré suivant un modèle admettant l'appartenance des nœuds aux communautés comme des paramètres. Le but est alors d'inférer les paramètres qui auraient généré les observations trouvées avec la probabilité la plus élevée.
- 7 **Méthodes pour extraire des communautés recouvrantes** : Il est évident qu'un nœud peut appartenir à plusieurs groupes ou communautés c'est la caractéristique

de recouvrement de communautés. La première méthode prenant efficacement le recouvrement a été proposée en 2005 par Palla.[25] Par la suite, d'autres approches ont été proposées.

- 8 **Méthodes multi-résolution** : L'application du paradigme multi-résolution à la détection de communautés cherche à intégrer un facteur d'échelle permettant de déterminer directement l'échelle de détection et indirectement la taille caractéristique des communautés. [26]

Bothorel ont mené une étude sur les méthodes de clustering dans les graphes attribués et les ont classé en trois familles:

- 1 **Les méthodes exploitant les attributs puis les relations** : ces méthodes se basent sur l'enrichissement du graphe du réseau social soit en ajoutant des sommets et des arêtes basés sur les attributs soit en valant directement les arêtes à l'aide des attributs .
- 2 **Les méthodes exploitant les relations puis les attributs**: dans ce contexte, un regroupement des communautés est effectué en se basant sur les valeurs des attributs .
- 3 **Les méthodes exploitant les attributs et les relations** conjointement : parmi les méthodes proposées dans cette perspective, des extensions de la méthode de Louvain sont identifiées.

Cependant, la richesse de l'information disponible n'est pas limitée aux liens existant entre les entités. Dans un Réseau Social, un individu interagit avec d'autres, et ces interactions obéissent aux différents phénomènes sociaux réels comme l'influence et la confiance. Le large éventail des données sur ces phénomènes suscitent un intérêt certain dans l'analyse des communautés. [13]

4. Mesure de qualité d'une partition

Une mesure de qualité peut être utilisée pour comparer différentes partitions et choisir la meilleure. En effet, les mesures de qualité associées à un objet mathématique a priori dépourvu d'une relation d'ordre, une valeur numérique standardisée. Ainsi, ces mesures sont utilisées comme fonction objective à maximiser dans le cadre d'algorithmes de partitionnement.

La modularité. C'est une mesure de la qualité d'un partitionnement de graphe. Si l'on considère P la partition en p clusters du graphe $G = \{V,E\}$, alors : $P = \{c_1, \dots, c_n, \dots, c_p\}$ La

$$Q(P) = \sum_i (e_{ci} - a_{ci}^2) \dots \dots \dots 3.1$$

modularité peut être introduite de façon assez simple de la façon suivante, en se référant à l'idée de Newman.

avec e_{c_i} la part des liens d'un cluster c_i sur le total, a_{c_i} la probabilité qu'un sommet se trouve dans le cluster c_i et donc $a_{c_i}^2$ la probabilité que les deux sommets d'un lien se trouvent dans le même cluster c_i . Cette expression générale est transformée dans la première forme usuelle de présentation de la modularité. On montre que la modularité peut s'écrire sous la forme :

$$Q(P) = \frac{1}{2m} \sum_{i,j \in V} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_i, c_j) \dots \dots \dots 3.2$$

avec

- m le nombre d'arêtes du graphe.
- A la matrice d'adjacence du graphe.
- A_{ij} le poids des liens entre les sommets i et j .
- d_i la somme des degrés de i avec $d_i = \sum_j A_{ij}$.
- $a_{c_i}^2 = \sum_j \frac{d_i d_j}{4m^2}$.
- $\delta(c_i, c_j)$ une fonction de Kronecker qui vaut 1 si les deux sommets appartiennent à la même communauté et 0 sinon.
- $A_{ij} - \frac{d_i d_j}{2m}$ correspond à la différence de liens entre notre graphe et un graphe aléatoire dont la contrainte est la conservation des degrés de sommets.[27]

Elle est un indice de la qualité du partitionnement d'un graphe donné. Cet indice est important si le nombre de liens entre les communautés est faible et le nombre de liens intra-communautés est élevé. La meilleure structure de communautés est celle qui maximise la modularité. Une étude comparative a d'ailleurs montré l'efficacité de la fonction de modularité de Newman par rapport à d'autres fonctions de qualité de partitionnement comme Modularity ration. [28]

5. Les algorithmes de détection de communauté

5.1 Girvan et Newman

C'est la méthode séparative la plus classique qui introduit une mesure de centralité appelée centralité d'intermédiarité des liens (Edge-BetweennessCentrality)[17].

Elle repose sur l'idée suivante : si un lien se trouve fréquemment sur les plus courts chemins entre les nœuds du graphe, alors il est naturel de penser qu'il ne se trouve pas « au sein d'une communauté donnée », mais plutôt qu'il relie des portions distantes du graphe, des communautés distinctes. En retirant progressivement le lien qui a la plus forte « centralité », on obtient un découpage des blocs de notre réseau.

Voici le détail de l'algorithme :

- on calcule la centralité (*betweenness*) de chaque lien .
- on enlève le lien de plus forte centralité.
- on recommence jusqu'à ce qu'il n'y ait plus de lien.
- les composantes connexes restantes sont les communautés.
- on obtient une décomposition hiérarchique du réseau.[5]

```

RevisedNewmanGirvan(G)
Data: a graph G
Result: the hierarchy of nested partitions of  $V(G)$ 
 $G_0 := G$ ;
 $P_0 := \{V(G)\}$ ;
 $i := 0$ ;
 $c := 0$ ;
while  $E(G_i) \neq \emptyset$  do
   $S_i := \{B(e) : e \in E(G_i)\}$ ;
   $mx := \max S_i$ ;
   $M_i := \{e \in E(G_i) : B(e) = mx\}$ ;
   $\hat{E}_i := M_i$ ;
   $\ell_i := 1$ ;
  while  $|\hat{E}_i| > 1$  do
     $\ell_i := \ell_i + 1$ ;
     $U_{\ell_i} := \{B(A) : A \subset M_i \wedge |A| = \ell_i\}$ ;
     $gmx := \max U_{\ell_i}$ ;
     $\hat{E}_i := \{A \subset M_i : |A| = \ell_i \wedge B(A) = gmx\}$ ;
  end
   $G_{i+1} := G_i \setminus \hat{E}_i$ ;
  if  $b_0(G_{i+1}) > b_0(G_i)$  then
     $c := c + 1$ ;
     $P_c := \{V_1, \dots, V_{r_c} :$ 
       $\langle V_1 \rangle, \dots, \langle V_{r_c} \rangle$  are connected components of  $G_{i+1}$ 
     $\}$ ;
  end
   $i := i + 1$ ;
end
return  $\{P_i : i = 0, \dots, c\}$ 

```

Figure 3.2 Pseudo-code de l'algorithme Girvan et Newman [5]

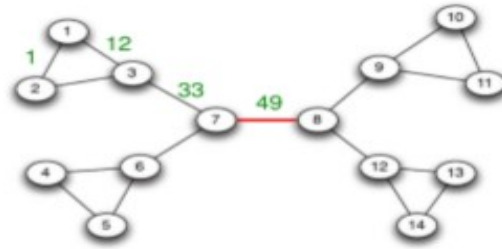
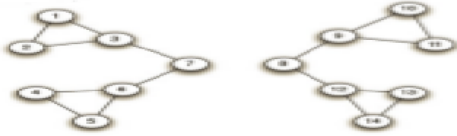


Figure 3.3 graphe pour la présentation de l'algorithme de Girvan-Newman [5]

Étape 1



Étape 2



Étape 3



Figure 3.4 Étapes de l'algorithme de Girvan-Newman [5]

5.2 Louvain

En 2008, trois chercheurs de l'université de Louvain ont proposé une autre méthode "gloutonne", plus rapide que la majorité des autres approches. Sa particularité est de se fonder sur une approche locale de la modularité

Situation initiale : Chaque nœud est une petite communauté à lui tout seul.

Ensuite la passe 1 est constituée de deux phases :

Migrations

- Une migration consiste à considérer successivement chaque nœud et à le faire migrer vers une autre communauté voisine si ce déplacement améliore la modularité.
- Les migrations sont itérées tant qu'elles apportent un gain de modularité.

Réduction

à la fin des migrations on dispose d'une partition du graphe et l'on construit alors le graphe réduit selon cette partition.

Ensuite la passe 2 consiste à réappliquer les deux phases décrites dans la passe 1 à ce graphe réduit. Et les passes sont répétées tant qu'elles mènent à un gain de modularité.[29]

```

1:  $G$  the initial network
2: repeat
3:   Put each node of  $G$  in its own community
4:   while some nodes are moved do
5:     for all node  $n$  of  $G$  do
6:       place  $n$  in its neighboring community including
         its own which maximizes the modularity gain
7:     end for
8:   end while
9:   if the new modularity is higher than the initial
     then
10:     $G =$  the network between communities of  $G$ 
11:  else
12:    Terminate
13:  end if
14: until

```

Figure 3.5 Pseudo-code de méthode Louvain [29]

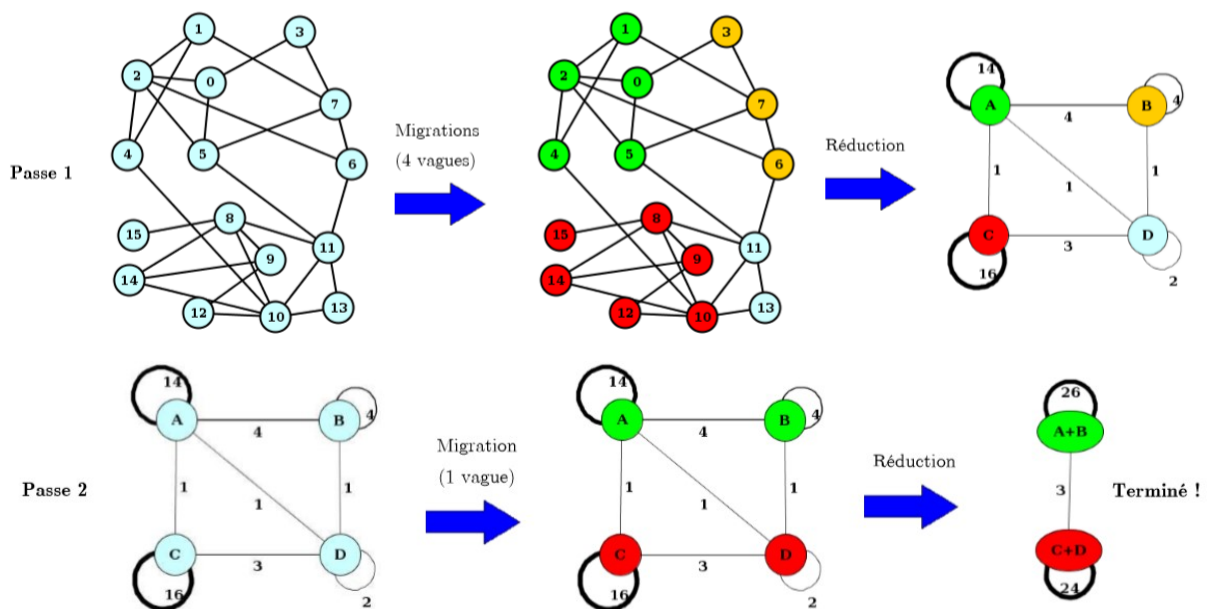


Figure 3.6 exemple d'exécution algorithme Louvain[29]

5.3 Label propagation

Cet algorithme se base sur le principe que chaque nœud change de communauté en fonction de la communauté à laquelle appartiennent ses voisins. Un nœud fait partie de la communauté qui contient le plus grand nombre de nœuds voisins. Ce processus est le modèle d'apprentissage pour Label propagation, qui s'exécute sur tous les nœuds dans chaque itération. Au début de l'algorithme, chaque nœud se trouve dans une seule communauté. Puis, les nœuds changent de communautés tout en respectant le modèle d'apprentissage. Avec cette méthode, un groupe de nœuds, fortement liés entre eux, finit par se trouver dans la même communauté. L'exemple de la figure 3.5 illustre ceci avec son réseau à 7 nœuds. Au début de l'algorithme, chaque nœud se trouve dans une communauté (a,b,c,d,e,f,g) comme illustré dans la figure 3.5 (1). L'algorithme choisit d'une façon aléatoire le traitement des nœuds. Supposons qu'il a choisi le nœud 3 de la communauté c.

Comme chacun de ses voisins appartient à une seule communauté, ce nœud peut appartenir à n'importe quelle communauté (a,b, ou d) qui lui est voisine. Dans notre cas, l'algorithme a choisi aléatoirement d'associer le nœud 3 à la communauté b (figure 3.5(2)). Il s'en suit que l'algorithme a choisi aléatoirement de traiter le nœud 7 se trouvant dans la communauté g. Ce dernier a un voisin dans la communauté e et un autre voisin dans la communauté f.

Le deuxième nœud à traiter passe donc de la communauté g à la communauté e (figure 3.5 (3)). Le nouveau nœud à traiter est le nœud 4 qui a deux voisins de la communauté b et un voisin de la communauté a. Puisque le nombre de voisin dans la communauté b est plus grand que le nombre de voisin dans la communauté a, le nœud 4 fait partie de la communauté b. En appliquant ce processus avec tous les autres nœuds, nous avons eu deux communautés b et e pour le graphe entier (figure 3.5 (6)).

La communauté d'un nœud v_i à la $t^{\text{ème}}$ itération dépend des communautés des voisins de v_i à la $t-1^{\text{ème}}$ itération. Il existe aussi une autre façon asynchrone pour affecter les communautés. La communauté d'un nœud v_i à la $t^{\text{ème}}$ itération dépend des communautés des voisins de v_i à la $t^{\text{ème}}$ itération. Idéalement, l'algorithme s'arrête lorsque les nœuds ne changent plus de communautés même si ce n'est pas toujours le cas. En effet, lorsqu'un nœud a un nombre maximal de voisins qui appartiennent à deux communautés ou plus, il change de communauté à chaque itération.

L'algorithme Label Propagation est décrit principalement par les étapes suivantes :

1-Initialement chaque nœud v_i , à l'itération $t = 0$, appartient à une seule communauté i : $C_{v_i}(0) = i, i = 1 \dots n$.

2-Définir $t = 1$.

3-Pour chaque nœud v_i , trouver la communauté du nœud v_i à la $t^{\text{ème}}$ itération $C_{v_i}(t)$ en fonction des communautés de voisins de v_i à l'itération t ou $t-1$.

4-Si tous les nœuds ne changent plus de communautés ($C_{v_i}(t) = C_{v_i}(t-1)$ pour $i = 1 \dots n$), l'algorithme s'arrête. Sinon, $t = t+1$ et aller à l'étape (3).[28]

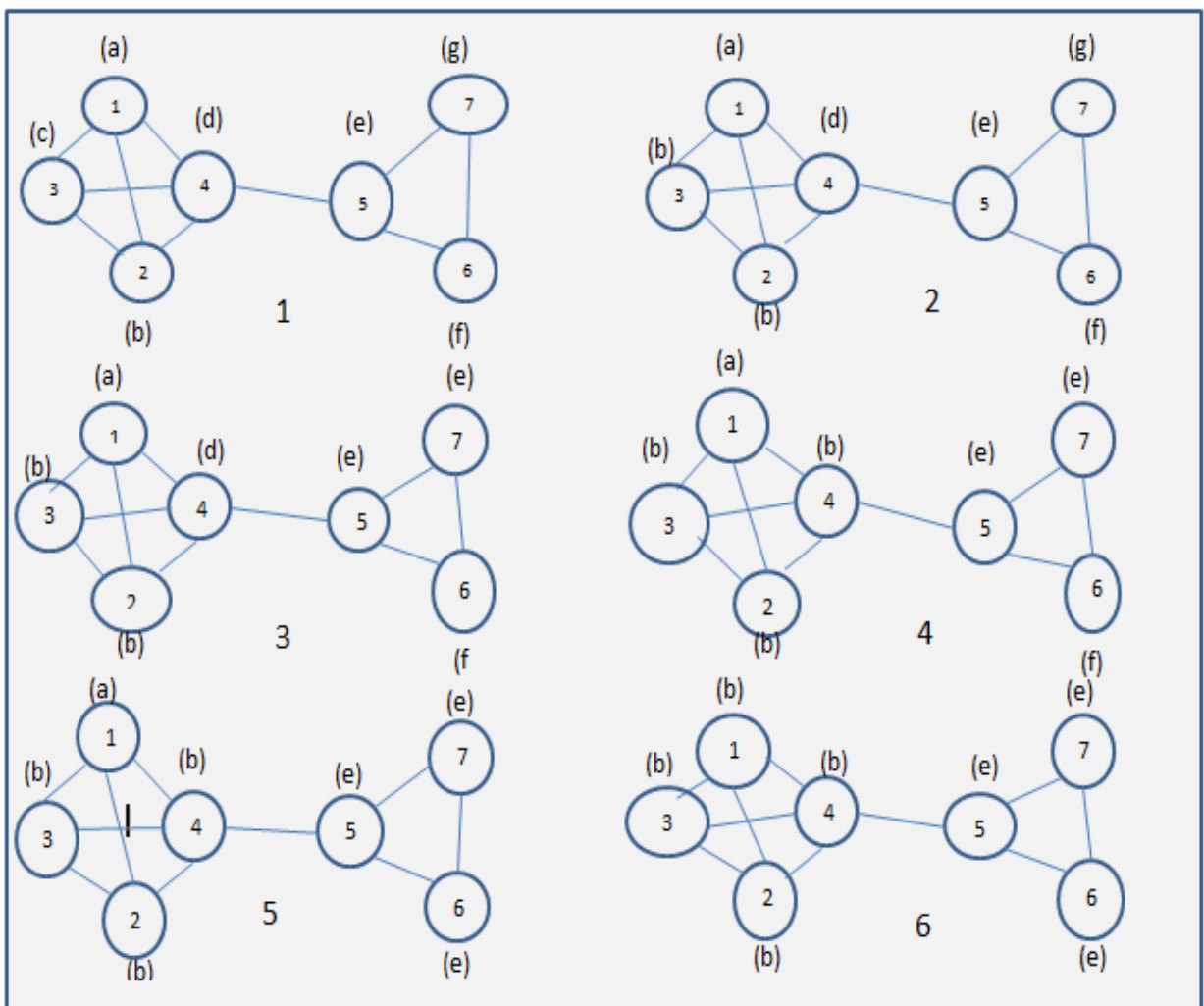


Figure 3.7 Exemple d'exécution de l'algorithme Label Propagation.[28]

6. caractéristiques des algorithmes

Algorithmes	Avantages	Inconvénients
Girvan et Newman	-résultats de qualité raisonnable [17]	-une forte complexité algorithmique ($O(m^2n)$) -usage à de petits graphes (500 nœuds maximum) [17]
Louvain	-Simplicité de Parallélisations -La rapidité ($O(n \log n)$) -traiter des graphes ayant jusqu'à plusieurs milliards de liens [30]	-la sensibilité du résultat à l'ordre de traitement des sommets.[30]
Label propagation	-Simplicité de Parallélisations -Complexité : $O(m)$ - permet de travailler sur de grands graphes [28]	-non déterministe -mauvaise propagation qui peut mener à de trop grandes communautés [28]

Table 3.1 caractéristiques des algorithmes

Conclusion

La présentation générale menée au cours de ce chapitre sur les communautés portant sur leur définition et détection ainsi que la classification des méthodes de leurs détections. Nous avons présenté aussi la mesure de qualité d'une partition «la modularité» et les différents algorithmes de détection des communautés.

CHAPITRE 4
IMPLEMENTATION

Introduction

Dans ce chapitre, nous avons commencé par présenter les outils et environnement de développement que nous avons utilisé. Ensuite, on a présenté l'implémentation de notre travail. Après Nous avons procédé un exemple pour tester, Le tests est réalisés avec la méthode déjà vu dans le chapitre 3. Enfin on a présenté les résultats obtenus.

1. Outils et environnement de développement

Avant de commencer l'implémentation de notre travail, nous allons tout d'abord spécifier les outils utilisés qui nous ont semblé être un bon choix vu les avantages qu'ils offrent.

Langage Python

Notre choix du langage Python2.7. Python est un langage facile à apprendre et son code est plus lisible, il est donc plus facile à maintenir. Il est parfois jusqu'à 5 fois plus concis que le langage Java par exemple, ce qui augmente la productivité du développeur et réduit mécaniquement le nombre de bugs. L'environnement python est riche en bibliothèques. [31]

ANACONDA

Anaconda est un utilitaire pour Python offrant de nombreuses fonctionnalités. Il offre par exemple la possibilité d'installer des bibliothèques et de les utiliser dans ses programmes.

Outre cette console, ANACONDA propose un logiciel servant à divers usages. Celui-ci permet principalement d'accéder aux autres logiciels installés dans la suite ANACONDA. Il est possible grâce à ce logiciel de contrôler les bibliothèques utilisées par le programme, ainsi que de vérifier si celles-ci sont bien à jour. De plus, il offre à ses utilisateurs une grande documentation et un espace de partage communautaire.

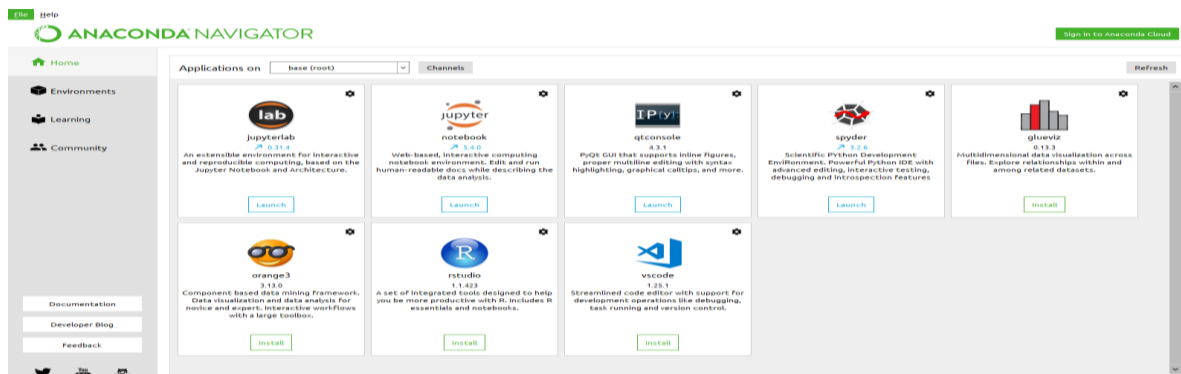


Figure 4.1 Plateforme ANACONDA

Spyder

Spyder est un logiciel de développement simplifiant la vie de ses utilisateurs. On peut accéder à ce logiciel en passant par l'ANACONDA Navigator.

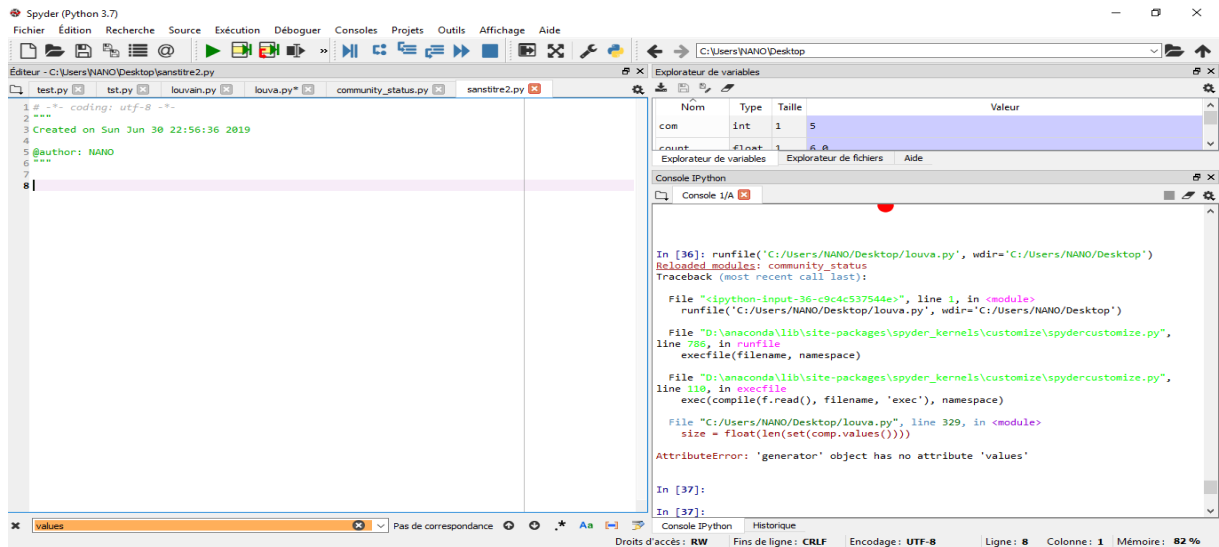


Figure 4.2 IDE Spyder

2. Les bibliothèques

Une bibliothèque est un ensemble de fonctions. Celles-ci sont regroupées et mises à disposition afin de pouvoir être utilisées sans avoir à les réécrire.

Celles-ci permettent de faire : du calcul numérique, du graphisme, de la programmation internet ou réseau, du formatage de texte, de la génération de documents...

On à utiliser plusieurs paquets qui sont : [32]

Matplotlib est une bibliothèque python qui dessine des graphiques . Nul besoin de connaissances en interfaces graphiques pour créer un graphique dynamique avec possibilité de zoom et de sauvegarde par l'utilisateur. Il est d'ailleurs possible de sauvegarder les graphiques en format matriciels comme le PNG, JPEG, etc. et vectoriels comme le PDF et le SVG.

Pillow est une bibliothèque de traitement d'image, Elle est conçue de manière à offrir un accès rapide aux données contenues dans une image, et offre un support pour différents formats de fichiers tels que PPM, PNG, JPEG, GIF, TIFF et BMP.

NetworkX est une bibliothèque Python qui permet d'instancier des graphes composés de nœuds et de ponts, ou liens. Grâce à cette bibliothèque, la manipulation de ces graphes est simplifiée.

Installation et organisation

Avant commencer l'installation des dépendances et Framework :

- Tout d'abord créer un environnement pour le projet figure 4.3 (1)
- Installer les bibliothèques figure 4.3 (2)

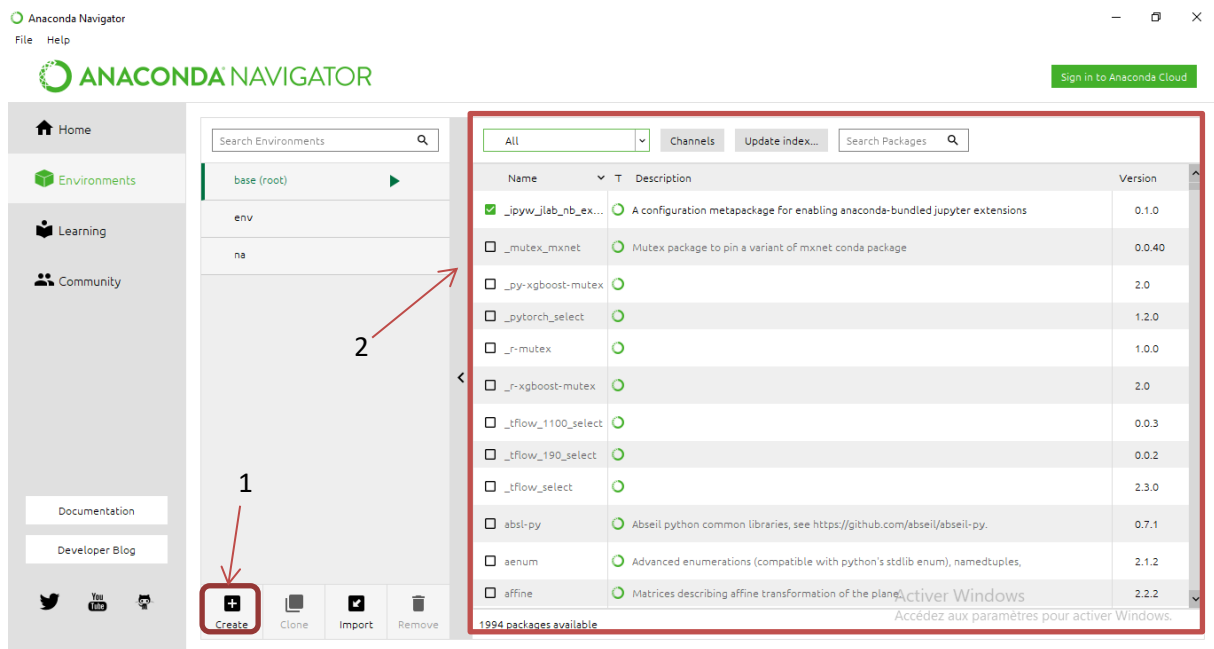
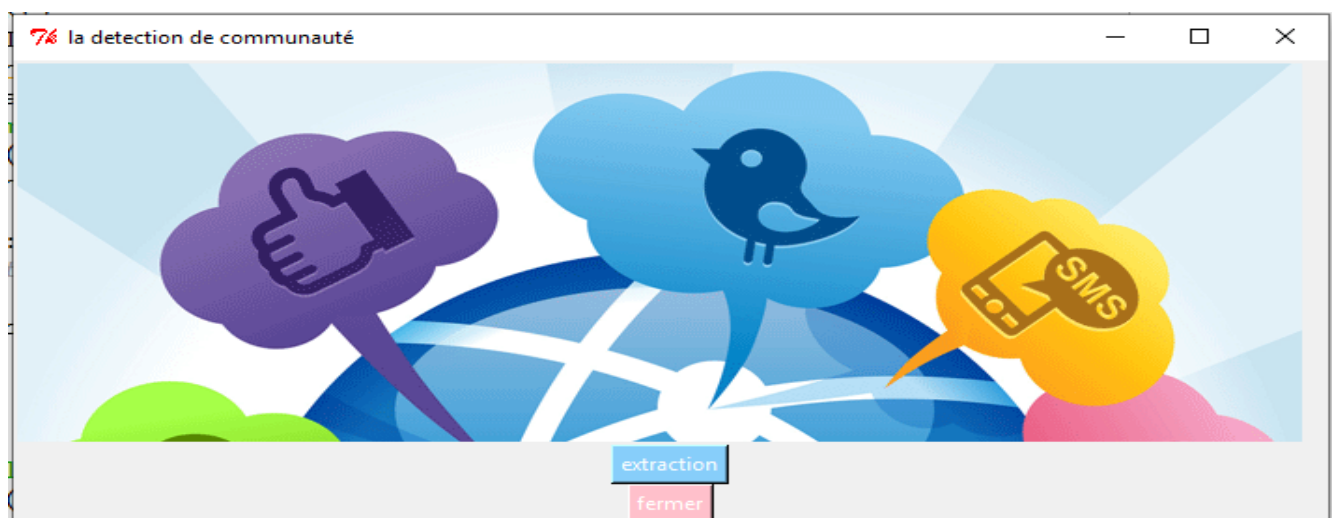


Figure 4.3 la création et l'installation

3. Le travail



3.1 Présentation de l'interface

Figure 4.4 interface de l'exécution

On réalise une nouvelle fenêtre en appuyant sur le bouton (extraction), cette fenêtre permet de sélectionner un fichier de format .txt de Facebook.

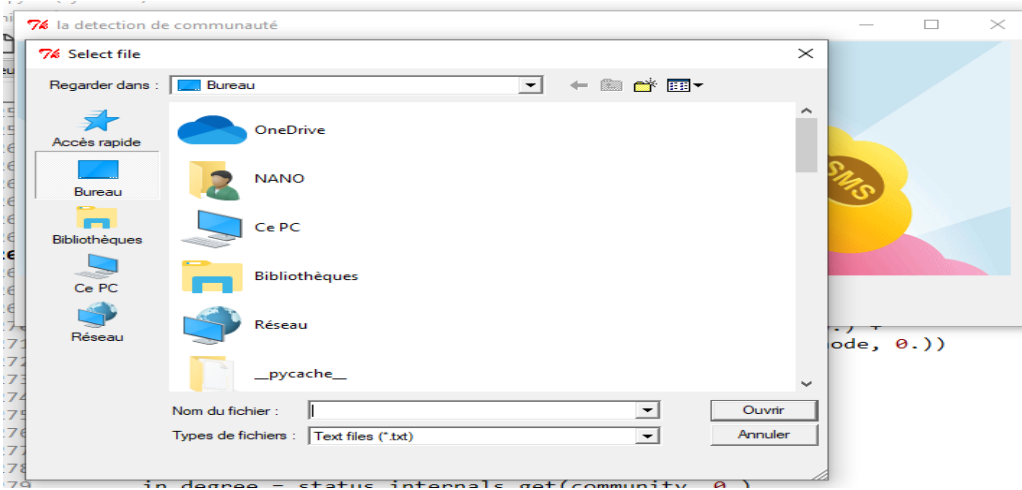


Figure 4.5 extraction d'un fichier

3.2 Le résultat obtenu



Figure 4.6 le résultat du détection de communautés dans l'interface

```

Console 1/A
IPython 5.8.0 -- An enhanced Interactive
Python.
?          -> Introduction and overview of
IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use
'object??' for extra details.

In [1]: runfile('C:/Users/NANO/Desktop/
louva.py', wdir='C:/Users/NANO/Desktop')
Image de fond (item 1 )
Name:
Type: Graph
Number of nodes: 4039
Number of edges: 88234
Average degree: 43.6910
number of community : 16
    
```

Figure 4.7 le résultat dans la console

```

def resul():
    fen2 = Tk()
    fen2.title("le resultat du detection de communauté")
    tex2 = Label(fen1, text='le resultat du detection de communauté ', fg='red')
    tex2.pack()
    photo = ImageTk.PhotoImage(Image.open("path1.png"))
    can = Canvas(fen1,height=500,width=500)
    item = can.create_image(0,0,anchor=NW,image=photo)

    can.pack()
    fen2.mainloop()
def fermer():
    p=fich()
    graph(p)
    resul()

fen1 = Tk()
#fen2 = Tk()
fen1.title("la detection de communauté")
tex1 = Label(fen1, text='la detection de communauté ', fg='red')
photo = PhotoImage(file="n.gif")
can1 = Canvas(fen1,height=250,width=700)
item = can1.create_image(0,0,anchor=NW,image=photo)

can1.pack()
Button(fen1,text='extraction',bg='lightskyblue', fg='white',command=fermer).pack(anchor=N)
Button(fen1,text='fermer', bg='pink', fg='white',command=fen1.quit).pack(anchor=N)

fen1.mainloop()
    
```

Figure 4.8 quelque ligne de code (1)

```

1 def save():
2
3     plt.savefig ('path1.png')
4
5 def fich():
6     filename = tkFileDialog.askopenfilename(initialdir = "/",title = "Select file",
7                                             filetypes = (("Text files","*.txt"),("all files","*.*")))
8
9     return filename
10
11 def graph(filename):
12     G=nx.read_edgelist(filename)
13     print (nx.info(G))
14     partition = louvain(G)
15     print ("modularity is :",modularity(partition, G))
16     size = len(set(partition.values()))
17     print("number of community :",len(set(partition.values())))
18     pos = nx.spring_layout(G)
19     count = 0
20     for com in set(partition.values()) :
21         count = count + 1.
22         list_nodes = [nodes for nodes in partition.keys()
23                       if partition[nodes] == com]
24
25     nx.draw_networkx_nodes(G, pos, list_nodes, with_labels=True,node_size = 150,
26                           node_color = str(count/ size ))
27     nx.draw_networkx_edges(G, pos, alpha=0.5)
28     save()
29     # ...

```

Figure 4.9 quelque ligne de code(2)

```

1 def louvain(graph,
2             partition=None,
3             weight='weight',
4             resolution=1.,
5             randomize=None,
6             random_state=None):
7
8     dendo = generate_dendrogram(graph,
9                                 partition,
10                                weight,
11                                resolution,
12                                randomize,
13                                random_state)
14
15     return partition_at_level(dendo, len(dendo) - 1)
16
17
18 def generate_dendrogram(graph,
19                         part_init=None,
20                         weight='weight',
21                         resolution=1.,
22                         randomize=None,
23                         random_state=None):
24
25     if graph.is_directed(): #graph non orienté
26         raise TypeError("Bad graph type, use only non directed graph")

```

Figure 4.10 quelque ligne de code (3)

Conclusion

Dans ce chapitre nous avons présenté l'implémentation de notre travail, On a commencé par présenter l'environnement matériel et logiciel d travail réaliser, ensuite nous avons présenté notre implémentation.

Enfin nous avons discuté sur les résultats obtenus.

CONCLUSION GENERALE

La détection de communautés est un domaine qui est encore dans une phase d'exploration, et pour lequel il faudra encore attendre quelques années avant d'arriver à un stade de maturation. Cette relative jeunesse du domaine a, d'une part, représenté un challenge et, d'autre part, a été un facteur de motivation important.

Le travail exposé dans ce mémoire s'intéresse à la détection de communauté dans un réseau social « Facebook ».

Dans le premier chapitre, on a présenté les médias sociaux ainsi que les réseaux sociaux.

Dans le deuxième chapitre, on a étudié les différents notions liés l'analyse des réseaux sociaux.

Après avoir introduit les notions principales de l'étude, Dans le troisième chapitre, on a présenté la notion des communautés ainsi que la détection de ces derniers. Après on a décrit d'une manière détaillé l'algorithme implémenté « Louvain ».

Dans le dernier chapitre, on a implémenter l'algorithme de Louvain sur le réseau social Facebook pour détecter ces différentes communautés. Le résultat obtenu c'est été 16 différentes communautés. Alors on peut déduire que se algorithme est capable de détecter les communautés d'un réseau.

REFERENCES BIBLIOGRAPHIQUES

- [1] <https://www.ludosln.net/medias-sociaux-definition-et-utilites/> consulté le 16/6/2019
- [2] M. DEWING , Les médias sociaux – Introduction Publication no 2010-03-F , Le 3 février 2013
- [3] <https://fredcavazza.net/2011/02/06/description-des-differents-types-de-medias-sociaux/> consulté le 16/6/2019
- [4] B.BENITA.et al. Lumière sur les réseaux sociaux (p14/15)
- [5] R.FOURNIER, M. CRUCIANU, M.FERECATU cour RCP216 Fouille de graphes et réseaux sociaux
- [6] <https://digitiz.fr/blog/reseaux-sociaux-definition-liste/> ,consulté le 16/06/19
- [7] <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> ,consulté le 28/06/2019
- [8] <http://socialonline.over-blog.com/les-reseaux-sociaux-en-general.html>, consulté le 16/06/19.
- [9] P. TORLOTING, enjeux et perspectives des réseaux sociaux paris ,2006 .
- [10] J. MORENO, Emotions mapped by new geography, <http://diana-jones.com/wp-content/uploads/Emotions-Mapped-by-New-Geography.pdf>, 1933.
- [11] L.TANG, H.LIU, Livre Community Detection and Mining in Social Media, 2010.
- [12] T ,M.,An experimental study of the small world problem, 1969.
- [13] N.CHOUCHANI ,Thèse de doctorat Une approche de détection des communautés d'intérêt dans les réseaux sociaux : application à la génération d'IHM personnalisées , 2018.
- [14] M.MALEK. Introduction à l'analyse des réseaux sociaux, 2009.
- [16] M.GAËTAN, B.STEPLIANE, Fouille de réseaux sociaux en ligne. Synthèse bibliographique sur le Data Mining Sociale, 2012.
- [15] M.DJELLOULI,A.TABTI, Mémoire Analyse de Contexte des Personnes Sur Les Réseaux Sociaux : Analyse des sentiments sur Twitter 2017/2018.
- [17] M. NEDIOUI, Mémoire fouille de donnée et apprentissage automatique dans les réseaux sociaux dynamiques,2015.

- [18] S. FORTUNATO, Community detection in graphs, 2010.
- [19] B. KERNIGHAN , S. LIN, An efficient heuristic procedure for partitioning graphs, 1970.
- [20] T. HASTIE, R. TIBSHIRANI, J. H. FRIEDMAN, The Elements of Statistical Learning , 2001.
- [21] J. NEWMAN and M. GIRVAN, Finding and evaluating community structure in networks, 2004.
- [22] G. AGARWAL, D. KEMPE, Modularity-maximizing graph communities via mathematical programming, 2008.
- [23] N. ALVES. Unveiling community structures in weighted networks, 2007.
- [24] S. BOCCALETTI, et al , complex network modularity by dynamical clustering, 2007.
- [25] G. PALLA, et al , The overlapping community structure of complex networks in nature and society, 2005.
- [26] C. BOTHOREL, et al , Clustering attributed graphs : models, measures and methods, 2015.
- [27] P. EUSEBIO, J. FLOCH, D. LEVY. Article chapitre 13 partitionnement et analyse de graphes. 2015.
- [28] M. TALBI. Mémoire une nouvelle approche de détection de communautés dans les réseaux sociaux, 2013.
- [29] J. CELLIER , Cour en quête de communauté, 2012.
- [30] <https://groupefmr.hypotheses.org/544> consulté le 07/07/19
- [31] <https://python-django.dev/> consulté le 07/07/19
- [32] https://fr.wikibooks.org/wiki/Programmation_Python/Biblioth%C3%A8ques_pour_Python consulté le 07/07/19

ملخص

في هذا العمل ، أجرينا دراسة متعلقة بمشكلة تحليل الشبكات الاجتماعية وطرق اكتشاف المجتمعات. هذا سمح لنا بتبني نموذج تمثيلي عام ومرن للشبكات الاجتماعية لتكون قادرة على دراسة أي نوع من العلاقات والتعامل مع شبكة كبيرة ، وبعد تحديد مفهوم المجتمع بطريقة واضحة ، درسنا عدة حلول للكشف عن المجتمعات مع مراعاة تعقيد الوقت والهيكل الديناميكي للرسوم البيانية. تمت مواجهة الحلول المقترحة مع الأساليب الحالية باستخدام مجموعات البيانات الحقيقية. هذه الخطوة الأخيرة للتحقق من مساهمة عملنا وإبرازها.

الكلمات المفتاحية: الشبكات الاجتماعية , تحليل الشبكة الاجتماعية , طرق الكشف عن المجتمع .

Abstract

In this work, we carried out a study about the state of art which is related with the problem of social networks analysis and methods of communities detection.

This allowed us to adopt a general and flexible representation model of social networks to study any type of relationship and to manipulate a network with huge dimension, and after having defined the community notion in a clear way, we studied a several solutions of communities detection taking into account complexity time and the dynamic structure of graphs. The proposed solutions have been confronted with existing approaches using real datasets. its the last step to validate and highlight the contribution of our work.

Keywords: Social networks , social network analysis , community detection methods.

Résumé

Dans ce travail, nous avons effectué une étude de l'état de l'art liée à la problématique d'analyse de réseaux sociaux et aux méthodes de détection de communautés.

Ceci nous a permis d'adopter un modèle de représentation des réseaux sociaux général et flexible pour pouvoir étudier tout type de relation et manipuler un réseau de grande dimension, et après avoir défini la notion de communauté de façon claire, nous avons étudié plusieurs solutions de détection de communautés en prenant en compte la complexité en temps et la structure dynamique des graphes. Les solutions proposées ont été confrontées aux approches existantes en utilisant des jeux de données réels. Ceci est l'ultime étape pour valider et mettre en valeur l'apport de notre travail.

Les mots clés : Réseaux sociaux, l'analyse des réseaux sociaux, méthodes de détection de communautés.