République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Larbi Tébessi -Tébessa -
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département : mathématiques et informatique

MEMOIRE DE MASTER

Domaine : Math et Informatique

Filière : informatique

Option : réseau et sécurité informatique

Thème :

Détection des communautés dans les réseaux sociaux basée sur l'analyse formelle de concepts

Présenté par :

Madjid Boughanem

Radia Slama

Devant le jury :

S.Hadjej	M.A.B	Université de Larbi Tébessi	Président
E.Bradji	M.C.B	Université de Larbi Tébessi	Rapporteur
H.Ghougha	M.A.A	Université de Larbi Tébessi	Examinateur

Date de soutenance : 30/05/2016

Note :……. Mention :……..……

ملخص

لقد أصبحت دراسة بنية المجتمع في الشبكات ذات أهمية بالغة، كما أن معرفة الوحدات الأساسية للمجتمعات في الشبكات يسهل فهمنا لعملها وتصرفاتها، وتساعدنا على فهم أداء هذه الأنظمة.،كما أنها لأنها تتيح لناالحصول على وجهة نظر دقيقة للنظام المعقد وهي أداة قيمة لفهم وتحليل هذه الأنظمة.

المجتمع في الشبكة هو عبارة عن مجموعة من العقد التي ترتبط بقوة فيما بينها و لكن ارتباطها يكون ضعيفا مع بقية العقد في الشبكة (الرسم البياني).

يهدف هذا العمل المتواضع لإثبات أنه من المناسب استخدام تحليل مفهوم الرسمي للكشف عن المجتمع، على عكس المناهج التقليدية التي تعتمد فقط على استخدام الرسوم البيانية.

للقيام بذلك، بدأنا موضوعنا بالتعرض لدراسات سابقة، الشيء الذي سمح لنا فيما بعد بتقديم تصنيف حول طرق كشف المجتمع سواء على أساس الرسوم البيانية أو على أساس تحليل مفهوم رسمي.

في الجزء الثاني، اهتممنا بتطوير نهج للكشف عن المجتمعات المفككة والمتداخلة في شبكات التواصل الاجتماعي و التي تأخذ بعين الاعتبار جميع العناصر الفاعلة في الشبكة الاجتماعية و التي تربط كافة العقد المعزولة بالمجتمعات.

الفكرة الرئيسية تكمن في الاعتماد على معيارين و المتمثلين في تعظيم النمطية المقترحة من طرف نيومان، و حساب التصرف.

Abstract

The study of community structure in networks becomes an important question. Know the basic modules (communities) networks make our understanding of their workings and behaviors, and help us to understand the performance of these systems, it allows us to obtain a macroscopic view of the complex system is a valuable tool for understanding and analyzing these systems.

A community in a network is a set of nodes that are strongly linked but weakly bound with the rest of the graph.

This modest work aims to demonstrate that it is appropriate to use formal concept analysis for the detection of communities unlike conventional approaches that use only graphs. To do this, we started our theme with a study of the state of the art which allowed us then to present a classification of community detection methods based on the graphs or formal concept analysis.

In the second part, we are interested in developing a detection approach disjointed and overlapping communities in social networks which takes into account all the actors of a social network.

The main idea lies in fact in mining communities based on two parameters: maximization of modularity introduced by Newman, and conductance.

Résumé

L'étude des structures de communautés dans les réseaux devient, à profusion, une question préoccupante. Connaitre les modules de base (communautés) des réseaux nous facilite la compréhension de leurs fonctionnements et comportements, et nous aide à appréhender les performances de ces systèmes, elle nous permet d'obtenir une vue macroscopique des système complexes et constitue une aide précieuse pour comprendre et analyser ces systèmes.

Une communauté dans un réseau correspond à un ensemble de nœuds qui sont fortement liés entre eux, mais faiblement liés avec le reste du graphe.

Ce modeste travail vise à démontrer qu'il est pertinent d'utiliser l'analyse formelle de concepts pour la détection de communautés, contrairement aux approches classiques qui n'utilisaient que des graphes.

Pour ce faire, nous avons entamé notre thème par une étude de l'état de l'art ce qui nous a permis, ensuite, de présenter une classification des méthodes de détection de communautés à savoir celle basée sur les graphes et celle portant sur l'analyse de concept formelle.

Dans le second volet, nous nous sommes intéressés à l'élaboration d'une approche de détection des communautés disjointes et chevauchantes dans des réseaux sociaux qui prend en considération tous les acteurs d'un réseau social ; et affecte tous les nœuds isolés.

L'idée principale réside en fait, dans l'extraction des communautés en se basant sur deux paramètres ; la maximisation de modularité introduite par Newman, et le calcul de la conductance.

Dédicace

Je dédie ce modeste travail à :

Mon adorable famille :

Ma mère source de tendresse…

Mon père source de courage…

Mes chères sœurs (Ismahene, Soraya, Nadia et Mounia)

Mon cher frère Yacine

A mon neveux Moeze et mes nièces Djihene et Ghizlene

Mon beau-frère Yacine

Ma propre famille :

Mon mari Karim

Mes petits anges

Anas Abd El Moumen et Farés Salah Eddine

A tous mes amis

A tous ceux qui m'aiment.
A tous ceux que j'aime.

Slama Radhia

Dédicace

Je dédie ce modeste travail à :

Mon adorable famille :

Ma mère source de tendresse…

Mon père source de courage…

Ma femme

Mes gosses : Mohammed Ali et Maha

Mes propres sœurs

Mes propres frères

A mes neveux et mes nièces

Et mes amis

Boughanem Madjid

Remerciements

Nous remercions en priorité ALLAH tout puissant de nous avoir donné le courage, la force, la patience et la volonté d'achever ce mémoire.

Nous remercions en premier lieu notre directeur de recherche Monsieur le Docteur Lwardi Bradji pour son encouragement et ses conseils qui nous ont permis de mener à bien ce mémoire.

Nous remercions, aussi notre directeur de travail monsieur Abdelmalek Metrouh pour son encouragement et toute l'équipe du centre des réseaux et système, de téléenseignement et d'enseignement à distance.

Nos remerciements vont également à tous nos collègues qui ont soutenus et à tous ceux qui ont contribué de près ou de loin à la réalisation de notre travail en particulier Boualeg Yakoub, Necib Farouk, Slama Mounia.

Table des matières

Contents

Liste des tableaux

Listes des figures

26	Nombre de communautés trouvés pour différents réseaux par différents Algorithmes	74

Introduction générale

La compréhension d'un phénomène de groupe, qu'il soit sociologique, économique, biologique ou informatique, passe par l'observation et l'analyse individuelle des membres du groupe, mais aussi par celle de leurs interactions.

Les réseaux sociaux constituent un exemple typique de réseau d'interactions. Un tel réseau est défini comme l'ensemble des liens sociaux établis entre les individus d'un groupe : liens d'amitié, politiques ou professionnels par exemple.

Actuellement, La perception actuelle de la notion de réseau a permis de réaliser des progrès significatifs pour la compréhension des systèmes complexes.

Un réseau est formalisé en mathématiques par un graphe, les membres du réseau sont appelés nœuds, et les interactions sont représentées par des paires de nœuds, appelées arêtes.

L'existence des zones est l'une des caractéristiques communes dans de nombreux réseaux ces zones sont plus densément connectées que d'autres. Celles-ci sont appelés communautés et correspondent intuitivement à des groupes de nœuds plus fortement connectés entre eux qu'avec les autres nœuds du réseau.

La détection de ces zones dites communautaires est un outil important pour la compréhension des structures et des fonctionnements des grands réseaux.

Dans les grands réseaux, la détection de sous-ensembles de sommets plus densément connectés que d'autres, est un problème que l'on retrouve dans plusieurs disciplines - Biologie (réseaux d'interactions entre protéines), Informatique (recherche d'informations sur le Web), mais aussi, Recherche Opérationnelle (détermination d´équipes), Sociologie (groupes dans des réseaux sociaux). Ces communautés jouent un rôle important dans l'organisation ou la structuration des réseaux. Ce problème est donc fortement lié à celui du partitionnement.

En Informatique, les réseaux sociaux ont été principalement étudiés par deux familles d'approches, celles qui se basent sur la théorie des graphes et celles basées sur l'AFC (Analyse Formelle des Concepts).

Les méthodes de détection de communautés ont fait l'objet de nombreux travaux, depuis l'article fondateur de Girvan et Newman. La plupart d'entre elles consistent à déterminer une partition des sommets du graphe optimisant un certain critère de qualité d'un partitionnement.

Il existe plusieurs indicateurs pour la détection des communautés à savoir La modularité ; la densité, la conductance, le temps d'exécution, le coefficient de clustring.

Malgré les nombreuses approches recensées dans le domaine de détection de communautés dans les réseaux, la détection est encore un problème complexe pour lequel aucun algorithme n'est le meilleur dans tous les cas.

Cette problématique nous a motivé à répondre à la question suivante :

Comment détecter des communautés dans des réseaux sociaux en combinant l'analyse formelle de concept et les graphes ?

Pour répondre à cette problématique nous nous intéresserons dans ce travail à :

- La proposition d'une nouvelle approche de détection de communautés dans les réseaux sociaux.
- La proposition d'un algorithme pour détection de communautés dans les réseaux sociaux.
- La comparaison des résultats de l'algorithme proposé avec d'autres algorithmes.

Nous introduisons ce mémoire par un premier chapitre qui porte sur les notions des réseaux sociaux et les communautés ; chapitre 2 s'intéresse à quelques notions de l'analyse formelle de concept et aux différentes approches de détections de communautés à savoir celle basés sur les graphes et celles portant sur l'analyse formelle de concept.

Chapitre 3 est consacré pour la contribution, l'expérimentation et la discussion des différents résultats.

Enfin, ce mémoire s'achève par une conclusion et les perspectives de recherches.

Chapitre01 :
Les réseaux sociaux

Chapitre01 : les réseaux sociaux

I. Introduction

Depuis la nuit des temps, les hommes communiquent sur les quatre coins de la terre, des réseaux sociaux y ont été présents, et des groupes sociaux se sont formés pour échanger des idées relevant de divers thèmes à savoir la religion, les arts et la vie de tous les jours.

On peut parler de l'existence de ces réseaux dès que deux personnes ou plus partagent leur avis.

A travers l'histoire, les hommes ont toujours cherché à échanger leurs opinions, la raison pour laquelle ces réseaux n'ont pas cessé de se diversifier, se multiplier voire se développer.

Avec l'avènement d'internet, une nouvelle catégorie de réseaux sociaux plus importants apparait, classmates (crée en 1995 par Randy Conrads) en fut le premier ayant pour objectif de retrouver des anciens camarades de classe, collègues ou soldats de l'armée.

La véritable apparition des réseaux en ligne remonte à l'an 2001 et ce n'est qu'en 2002 qu'ils deviennent de plus en plus populaire, et c'est grâce à Friendster, crée en Mars de la même année par Jonathan Abrams, mettant en exergue la notion de cercles d'amis.

Avec le Web 2.0, on assiste à une nouvelle ère de réseaux sociaux en ligne, qui valorisent l'activité des utilisateurs des sites, il s'agit d'ailleurs de participants actifs des réseaux.

II. Origines des réseaux sociaux

Au début des années 1930 Jacob Levy Moreno fut la première personne à avoir représenté un réseau social. [1]

Il s'agissait d'une visualisation graphique d'un réseau social, ou les personnes ont été représentées par des points et la relation qui unit deux personnes par des flèches.

Le milieu du 20ième siècle, fut marqué par les travaux de Cartwright et Harary qui étaient les premiers mis en œuvre la théorie des graphes dans l'analyse des réseaux sociaux. Cette théorie, est devenue, par la suite la représentation adoptée par toutes les disciplines manipulant l'analyse des réseaux sociaux à savoir la sociologie, les mathématiques et l'informatique.

II.1 Définition d'un réseau social

II.1.1. Définition d'un réseau

On désigne par réseau tous ensemble d'éléments reliés entre eux et réglés de manière qu'ils puissent communiquer ; ou « c'est un ensemble de nœuds reliés entre eux par des liens ».

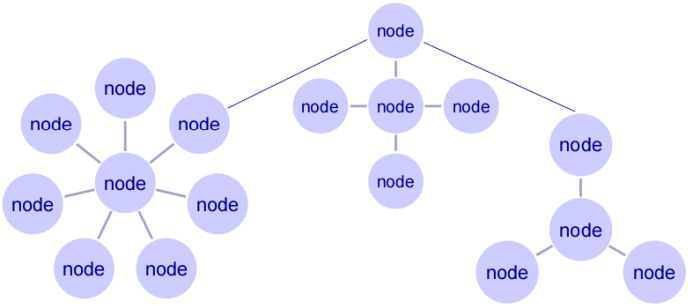


Figure01.structure d'un réseau

II.1.2 Exemples de réseau

Le terme utilisé pour identifier un réseau diffère selon le type d'entité concernée, on distingue :

Réseau de transport : ensemble d'infrastructures et de disposition permettant de transporter des personnes et des biens entre plusieurs zones géographiques

Réseau téléphonique : infrastructure permettant de faire circuler la voix entre plusieurs postes téléphoniques

Réseau de neurones : ensemble de cellules interconnectées entre-elles

Réseau de malfaiteurs : ensemble d'escrocs qui sont en contact les uns avec les autres (un escroc en cache généralement un autre !)

Réseau informatique : ensemble d'ordinateurs reliés entre eux grâce à des lignes physiques et échangeant des informations sous forme de données numériques (valeurs binaires, c'est-à-dire codées sous forme de signaux pouvant prendre deux valeurs : 0 et 1)

II.1.3 Définition d'un réseau social

Définition1 : un réseau social, tel qu'il est défini par Pierre Mercklé dans Sociologie des Réseaux Sociaux, est : « un ensemble d'unités sociales et des relations que ces unités sociales entretiennent les unes avec les autres directement, ou indirectement à travers des chaînes de longueurs variables ».

La présence de trois personnes au minimum est indispensable pour que l'on puisse parler de réseau social, car, dans un réseau, il existe ce que nous appelons relations fortes et d'autres dites faibles (relations moins fortes).

Deux individus peuvent former une relation personnelle, elle est donc forte, il y a un de l'intimité.

La disparition de l'un des deux individus entraine la disparition totale de la relation. Lorsque l'on rassemble trois internautes, la relation devient interpersonnelle et donc plus impersonnelle. Des stratégies se développent. Un réseau social, n'a pas de frontières délimitées, un réseau peut-être potentiellement infini . [2]

Définition2 :

Un réseau social est un ensemble d'acteurs (individus, groupes ou organisations) reliés par des interactions sociales. [3]

•Ces interactions sociales peuvent être de différentes natures : familiales, sentimentales (liens forts) ou plus distantes : affinité, relation d'affaire, de travail (liens faibles) …

•Elles peuvent être reliées à travers des contacts directs ou médiés technologiquement : échange de lettres, de méls, chat, réseaux sociaux, mondes virtuels…

III. Analyse des réseaux sociaux

L'analyse de réseaux est au centre d'intérêts de divers travaux ces dernières années. En effet, dans de nombreuses disciplines, il est possible de représenter des systèmes réels sous forme de réseaux. Nous pouvons ainsi, citer les réseaux de neurones en biologie, les réseaux sociaux dans les sciences humaines, voire les réseaux interbancaires en économie.

Récemment, la popularité croissante d'internet et du web a conduit à la naissance d'immenses réseaux, pouvant contenir des milliards de nœuds et de liens, et pour lesquels il est possible d'obtenir de grandes quantités d'information.

Etant donné la complexité des réseaux et la diversité de leur application, la détection des communautés est le problème majeur des réseaux.

Les communautés peuvent être définies comme étant des ensembles de nœuds fortement liés entre eux, et plus faiblement liés avec le reste du réseau [4]**.** Ces structures sont signifiantes dans les réseaux : par exemple, dans un réseau social, ces communautés pourront correspondre à des groupes sociaux (famille, groupe d'amis, etc.). Elles pourront correspondre à des champs lexicaux dans des réseaux de synonymie, ou encore à des blogs traitant de sujets similaires dans une analyse de pages web, pour donner quelques exemples.

V*I.* Propriétés des réseaux sociaux

VI .1. Types des réseaux sociaux

Dans ce contexte, on distingue différents types de réseaux sociaux :

- **Généralistes :** ces sites offrent la possibilité de créer et d'élargir son cercle d'amis, le plus connu étant Facebook.

- **Professionnels :** tel que LinkedIn ou Viadeo qui sont devenus des outils primordiaux pour les ressources humaines.

On outre, il existe aussi des réseaux sociaux professionnels spécialisés par corps de métiers (avocats, marketing, finance…) ;

- **Focalisés sur les centres d'intérêts :**

Comme la musique avec MySpace, Deezer, Spotify ou LastFM, la littérature (Babelio), le cinéma, la religion.

- **Centrés sur les services et la vie pratique**

Pratique :tel que Yahoo! Questions Réponses, Peuplade sur sa vie de quartier, les réseaux de jeunes mamans …

Par ailleurs, si des réseaux sociaux préexistent et ne nécessitent qu'une inscription ou la création d'un compte, il est également possible pour un individu, une marque, une institution, …, de créer son propre réseau social grâce à des plateformes de créations de réseaux comme Affinitiz ou Ning.

VI .2. Caractéristiques des réseaux sociaux

VI.2.1 Le caractère communautaire

Depuis l'apparition du Net Le regroupement des individus par centres d'intérêts ou origines pour interagir entre eux, existe :

- Newsgroups/usenet (forums de discussion par mails), listes de diffusion, chats / IRC.

-Regroupement autour d'un site web ou des personnes partagent un même centre d'intérêt, avec un forum éventuellement.

- Commentaires sur sites / blogs.

VI.2.2 Loi de distribution des degrés

La loi de distribution des degrés des graphes est par excellence l'une des caractéristiques principales des réseaux sociaux.

Commençant par un constat [5] puis une expérimentation, il a été démontré que dans un graphe, les degrés des nœuds suivent une distribution de type loi de puissance, de type $P(k) \ \alpha \ k^{-\alpha}$ avec k les degrés d'un nœud.

VI.2.3 L'effet petit-monde

Outre la loi de distribution des degrés, il a été vérifié [6] que les graphes possèdent généralement, un faible diamètre, qui représente la distance la plus longue de toutes les plus courtes distances d'un graphe.

Cette théorie Caractérise un réseau dans lequel il est possible de trouver un chemin très court entre toute paire de points, chemin trouvable sans même connaitre le réseau dans sa totalité. En

général, le diamètre d'un graphe représentant un « petit monde » est de l'ordre de ***O(log n)*** avec n le nombre de nœuds du graphe.

VI.2.4 Clusterisation

Un réseau social se distingue par son taux de clusterisation, qui peut se mesurer par la proportion (pourcentage) de **« mes »**, amis qui sont amis entre eux. Ce taux, pour un nœud ***i*** est calculé de la manière suivante : $C_i = \frac{2e_i}{k_i(k_i-1)}, C_i \in [0,1]$, e_i présente le nombre de liens entre les voisins de i et avec k_i le degré de i. Le taux de clusterisation peut également être calculé pour un graphe en effectuant la moyenne de tous les coefficients de chaque nœud :$C = \frac{1}{N}\sum_i^N C_i$ Cette propriété de clustering peut être expliquée par la propriété de similarité (les amis de x lui sont similaires) et par une transitivité de la similarité ou si x et y sont amis avec z alors x et z sont similaires, de même entre y et z d'où par transitivité x et y sont similaires ou encore x et y sont amis. [7]

VI.2.5 Un coefficient de clustering local élevé

Les individus dans un tel réseau optent pour le regroupent en communautés pour se socialiser.

Si un nœud S1 est connecté à un autre nœud S2 qui est lui-même connecté à un nœud S3, alors il y a une forte probabilité pour que S1 soit aussi connecté à S3. En d'autres termes, la façon avec laquelle les nœuds sont dans un réseau social favorise l'émergence de structures de graphe de type Triangle. Ces communautés possèdent une forte densité locale et une faible densité globale.

Le coefficient de clustering est donné par la formule suivant :

$$\sum \frac{3 \times \#\Delta}{\#\Lambda}$$

Où #Δ est le nombre de triangles dans le graphe et #Λ est le nombre de triades. Noter que dans un graphe aléatoire le coefficient de clustering sera de l'ordre de la probabilité de l'existence d'un lien [8].

VI.3 Représentation d'un réseau social

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. Les définitions suivantes listent quelques notions manipulées par la théorie des graphes pour les réseaux sociaux [1] :

- Un **sommet** est l'unité de base d'un réseau, il en représente une ressource, dans un réseau social on parle d'acteur.
- Le terme **nœud** est également utilisé pour désigner un sommet.
- Une **arête** est une connexion entre deux sommets. On parle également d'arc ou de lien.
- Une arête est **orientée** si elle ne s'utilise que dans une seule direction. Inversement, on parle d'arête **non orientée** pour une arête qui s'utilise dans les deux directions.
- Une arête est **pondérée** lorsqu'on lui attribue un poids.
- Une **arête est étiquetée** lorsqu'on lui attribue un label.
- Un **graphe** est défini par un ensemble de sommets et un ensemble d'arêtes.
- Un **graphe orienté** désigne un graphe avec des arêtes orientées.
- Un **graphe pondéré** désigne un graphe avec des arêtes pondérées.
- Un **graphe étiqueté** désigne un graphe avec des arêtes étiquetées.
- Le **degré** d'un sommet est le nombre de ses arêtes adjacentes.
- Un **chemin** est une séquence d'arêtes qui relie deux sommets.
- Un **chemin orienté** est une séquence d'arêtes qui relie deux sommets en respectant l'orientation du parcours à chaque arrête.
- Un graphe est **complet** lorsqu'il existe une arête entre toute paire de sommets.
- Un graphe est dit **connexe** lorsqu'il existe un chemin entre toute paire de sommets.

VI.4 Le réseautage social

Le réseautage social correspond à l'ensemble des moyens utilisés afin de relier des personnes physiques ou personnes morales entre elles. Avec l'apparition d'Internet, il recouvre les applications Web connues sous le nom de « service de réseautage social en ligne ». Ces applications ont de nombreuses finalités et divers objectifs. Elles servent en effet à constituer un réseau social en reliant des amis, des associés, et plus généralement des individus employant ensemble une variété d'outils dans le but de faciliter, par exemple, la gestion des carrières professionnelles, la distribution et la visibilité artistique ou les rencontres privées.

V. Utilité des réseaux sociaux-avantages et inconvénients

La croissance spectaculaire des abonnées aux réseaux sociaux est préoccupante voire alarmante. Facebook, en est le meilleur exemple ; à la fin du premier trimestre 2016 Facebook revendique 1,65 milliards d'utilisateurs actifs chaque mois et 1,51 milliards d'utilisateurs actifs sur mobile chaque mois. Ces chiffres attestent de la place qu'occupent les réseaux sociaux dans notre vie. C'est pourquoi, il faut être en mesure d'évaluer les répercussions. Commençons par

les répercussions positives de ces technologies. Évidemment, l'utilisation des réseaux sociaux englobe plusieurs avantages qui sont largement connus. En effet, sa facilité d'utilisation, peu coûteux, généralement gratuit, ça permet d'échanger rapidement, mais surtout, avec ces technologies, il est maintenant possible de rejoindre des milliers de personnes quasiment instantanément. Parmi les avantages des réseaux sociaux on peut citer :

1- Se connecter avec des amis que vous n'avez pas encore rencontrés :

En psychologie, des recherches expliquent le rôle que joue la discussion avec les amis virtuels dans l'amélioration de leur mode de vie ; en effet les individus se sentent heureux et soutenus par leurs amis virtuels.

2- Améliorer la communication avec votre famille :

Il n'est pas toujours évident de rester en contact avec notre famille, soit à cause du travail, la distance, le temps, l'argent, etc. Par contre, avec les réseaux sociaux, il est facile d'échanger quelques petites phrases pour prendre des nouvelles de la famille ou donner de nos nouvelles à ceux-ci.

3- Se renouer avec les anciens amis ou membres de la famille :

Grace à la popularité du Facebook, il devient facile de retrouver des anciens amis en faisant une recherche avec leur nom. Il est dit que le fait de communiquer avec votre passé peut être émotionnellement sain.

4- Clarifier ce qui est important :

C'est un lien de partage avec l'autre. Il nous donne l'occasion de parler de nos préoccupations, partager nos opinions et même nos sentiments.

Donnant aussi une image sur nos centres d'intérêts.

5- Partager avec le monde :

Sur ces sites, il est facile de partager avec les autres de vos ressources, connaissances et renseignements précieux. Vous pouvez donc aider et influencer les autres et faire une différence dans leur vie.

6- Aider d'autres gens :

Il peut être simple de faire une différence dans la vie des gens qui ne vont pas bien. Seulement en leur écrivant un petit message de sympathie ou d'empathie, vous pourriez contribuer à aider cette personne.

Cependant les réseaux sociaux ont des inconvénients. Ils peuvent être très nuisibles. A force d'être connectés sur ces sites tout le temps ; les gens en deviennent accros ! Cela peut engendre de graves problèmes pour eux. La majorité des étudiants se plaignent de la difficulté de réviser car ils sont trop distraits par leur Facebook, ce qui influence très négativement leurs résultats.

En outre, ces sites peuvent causer des problèmes de santé à savoir, l'obésité, le manque de sommeil ; et les maladies oculaires.

De plus, les gens inconscients des paramètres de confidentialités de ces sites peuvent affronter des problèmes sérieux touchant leurs vies privées (par exemple le partage de photos). C'est pourquoi, il est hyper important d'être conscient de tous les paramètres avant d'oser utiliser ces sites.

En somme, les réseaux sociaux ont des avantages et des inconvénients, on doit les utiliser avec modération.

IV. Notion de communautés

De nombreux systèmes complexes dans divers domaines comme la biologie, l'informatique, la linguistique, le commerce, etc., peuvent être représentés de manière abstraite par des réseaux.

L'existence des zones est l'une des caractéristiques communes dans de nombreux réseaux. Celles-ci sont appelés communautés et correspondent intuitivement à des groupes de nœuds plus fortement connectés entre eux qu'avec les autres nœuds du réseau.

La détection de ces zones dites communautaires est un outil important pour la compréhension des structures et des fonctionnements des grands réseaux. De tels réseaux peuvent être modélisés en termes de graphes, où un nœud représente un membre individuel du système, et une arête représente un lien entre les nœuds selon une relation bien déterminée du système.

Actuellement, La perception actuelle de la notion de réseau a permis de réaliser des progrès significatifs pour la compréhension des systèmes complexes.

IV.1. Définitions d'une communauté

Définition1 : Une « communauté » est un terme ambigu ; Selon l'article Wikipédia sur les communautés, « Une communauté est une interaction d'organismes partageant un environnement commun. Dans les communautés humaines, l'intention, la croyance, les ressources, les besoins ou les risques sont des conditions communes affectant l'identité des participants et le degré de leur cohésion. ».

Définition2 : Une communauté, que nous appelons aussi groupe ou module, est définie de façon la plus universelle et minimaliste comme un sous-ensemble de nœuds possédant plus d'arêtes internes, c'est-à-dire entre nœuds du groupe que d'arêtes externes, c'est-à-dire avec une extrémité dans le groupe et l'autre en dehors.

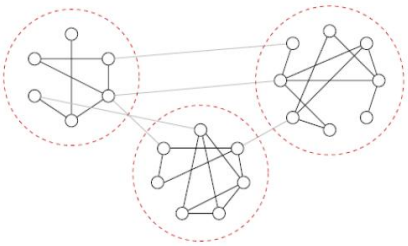


Figure2.Un réseau constitué de 3 communautés.

Communautés, Clusters et Groupes

Une communauté correspond à un ensemble de nœuds ayant plusieurs connexions entre eux mais par contre, un très faible nombre de connexions vers l'extérieur.

Il ne faut pas confondre les communautés, dites modules avec les groupes, ces derniers étant une caractéristique des nœuds. En effet, chaque nœud peut déclarer qu'il appartient `a un ou plusieurs groupes (comme c'est le cas avec les groupes Facebook par exemple). Un groupe est donc constitué de membres adhérents et ne doit donc pas être confondu avec les communautés.

Enfin, la différence entre la détection de communautés et le clustering consiste en ce que le premier cherche `à diviser le graphe en des structures selon leurs connexions (c'est uniquement de la topologie) alors que le deuxième consiste en un regroupement de structures selon une mesure de similarité, mesure qui est à définir au préalable. Cependant, ces deux notions sont similaires et sont souvent confondues.

Il existe deux définitions de communautés, l'une sémantique et l'autre structurelle :

Définitions d'une communauté de point de vue sémantique :

Ensemble des nœuds ayant des **centres d'intérêts communs** ou des profils similaires

Définition d'une communauté de point de vue structurelle :

Ensemble des nœuds **fortement liés entre eux** et faiblement liés avec les autres nœuds du graphe.

IV.2. Les différents types de communautés

Il existe diffèrent types de communautés [9] :

IV.2.1. Communautés d'intérêt

Il s'agit d'un type de communauté qui rassemble des individus qui partagent des idées, des croyances, une cause commune ou simplement une proximité propice à l'échange. Ces communautés sont parfois implicites. Elles constituent de temps à autre des réseaux souterrains de pouvoir. Deux lois expliquent ce "pouvoir" : la loi de Metcalfe, « l'influence d'un groupe augmente au carré du nombre de participants » ou la loi de Reed (1999) partant du principe que les réseaux encourageant la construction de groupes qui communiquent créent une valeur qui croît de façon exponentielle avec la taille du réseau. Ces lois de croissance indiquent comment la connectivité potentielle crée la valeur d'un réseau pour ses usagers. Les communautés d'intérêt sont ouvertes, elles jouent un rôle dans la dissémination d'informations. Par ailleurs, appartenir à plusieurs communautés d'intérêt permet d'être plus réceptif aux signaux faibles annonciateurs d'innovations.

IV.2.2. Communautés de pratiques

Dans ce type de communauté, les membres s'identifient par des pratiques communes. Ils s'engagent à s'entraider, échanger de l'information, apprendre les uns des autres, construire des relations, partager leurs savoir-faire. La communauté de pratique est informelle et spontanée, mais moins ouverte qu'une communauté d'intérêt. Souvent, les individus doivent être cooptés pour en devenir membre. Ce sont essentiellement les flux de connaissances qui caractérisent les communautés de pratiques**.**

IV.2.3. Communautés de projet

Pour ce qui est des communautés de projet est mettent la tâche au centre d'intérêt. Le flux d'information et de connaissance y est important, mais complètement dédié au projet. Il s'agit de délivrer un rendu, un produit ou une prestation, dans un délai alloué. Les acteurs ont un rôle donné. Pour être efficace, une communauté de projet ne peut compter trop de membres. Au-delà d'une dizaine, il est généralement conseillé de créer des échelons intermédiaires. Le nombre 13 est superstitieusement souvent évoqué comme une limite à ne pas dépasser…

IV.2.4. Communautés épistémiques

Une communauté épistémique quant à elle, est centrée sur la connaissance. Elle correspond à un nombre réduit de membres reconnus et acceptés, le plus souvent selon un

principe de cooptation. Ces derniers travaillent sur un sous-ensemble conjointement défini de questions en lien direct avec la création de nouvelles connaissances. Les membres d'une communauté épistémique acceptent de contribuer ensemble selon une autorité procédurale. Une telle autorité peut se définir comme, « un ensemble de règles ou de codes de conduite définissant les objectifs de la communauté et les moyens à mettre en œuvre pour les atteindre et régissant les comportements collectifs au sein de la communauté ». Cette structuration autour d'une autorité procédurale est acceptée car elle est essentielle à la création de nouvelles connaissances.

Conclusion

Dans ce chapitre nous avons essayé de présenter un éventail de définitions afin de mettre au clair les différentes notions relevant du thème des réseaux sociaux. Pour ce faire, nous avons commencé par les origines des réseaux sociaux pour passer ensuite à la définition de la notion du réseau et celle de communauté tout en démontrant leurs types. Ce petit rappel notionnel va nous permettre de passer au deuxième chapitre qui portera sur les différentes approches de détection de communauté dans les réseaux sociaux

Chapitre 02 : Différentes approches de détection de communauté dans les réseaux sociaux

Chapitre 02 : différentes approches de détection de communauté dans les réseaux sociaux

I. Introduction

Dans toutes les disciplines d'origine des réseaux complexes, on y trouve une étude de la structure communautaire qui possède des applications. Elle permet de comprendre les interactions entre la structure topologique et la dynamique d'un réseau, en démontrant, par exemple, pourquoi les liens se forment et comment ils évoluent et permet, en outre, de faciliter la compréhension des structures et leur fonctionnement.

La détection de communautés réside dans un procédé commun s'appliquant à tout type de réseau et permettant le partitionnement des nœuds en communautés. L'ensemble de ces communautés doit répondre à certains critères de qualité, notamment la modularité. De nouvelles approches sont de plus en plus proposées afin de minimiser les coûts de détection en termes de temps et d'espace, tout en optimisant la qualité des partitions trouvées. Selon les cas, différents compromis (sur le temps, l'espace, la définition adoptée, etc.) sont autorisés ; ce qui aboutit à la multiplication de méthodes de détection proposées.

Dans ce chapitre, nous décrirons les principales approches proposées dans la détection des communautés, et ce sans se limiter uniquement aux approches qui ont reçu le plus d'attention de la part des chercheurs du domaine. De ce fait, la liste des travaux cités, reste non exhaustive. Pour mettre en exergue les caractéristiques des méthodes, nous avons estimé important de regrouper ces dernières sous deux catégories : les approches basées sur les graphes et les approches basées sur l'analyse formelle des concepts sans pour autant négliger le type de communautés découvertes (disjointes ou chevauchantes).

II. Détection de communautés : définitions et intérêts

Nous essayerons, ici, de rapporter les diverses définitions relatives à la notion de communautés. Ainsi, nous nous parlerons des différents intérêts de la détection des communautés dans les réseaux. Nous considérons ce dernier point comme très important car il constitue, en quelques sortes, une partie fondamentale de notre travail.

II.1. Qu'est-ce qu'une communauté ?

Le terme communauté désigne, comme son radical l'indique, un ensemble d'individus ayant une ou plusieurs caractéristiques en commun. Cette notion de communauté a occupé, depuis plusieurs décennies, le centre d'intérêt de nombreuses études de la part des sociologues, notamment dans le cadre de l'analyse des réseaux sociaux [10]. Les réseaux sociaux sont des structures modélisant les relations sociales (par exemple l'amitié, la collaboration, la parenté, etc.) qui existent entre un ensemble d'individus appelés aussi acteurs. Ceci nous donne nécessairement une idée intuitive de ce qu'est une communauté.

Depuis quelques années, l'utilisation du mot communauté s'est généralisée à d'autres types de réseaux et n'est désormais plus réservée aux réseaux sociaux. Nous retrouvons cette extension d'usage dans plusieurs disciplines telles que l'informatique la physique ou encore la biologie. En informatique par exemple, la notion de communauté est devenue dès la fin des années 90 très populaire avec le développement du Web et de la recherche d'information basée sur les liens hypertextes. Des chercheurs comme Kleinberg [11] ou Flake [12]ont introduit la notion de communauté web qui désigne un ensemble de pages web traitant d'une même thématique ou d'un même sujet (ou topic en anglais). Ainsi, des structures communautaires sont définies et observées dans de nombreux réseaux, et jouent un rôle important dans leur organisation ou leur structuration. De ce fait, il est fondamental de les définir nettement et de les détecter automatiquement.

Bien qu'aucune définition formelle de ce qu'est une communauté ne soit actuellement reconnue, on désigne intuitivement par ce terme un groupe de nœuds du réseau plus fortement connectés entre eux qu'avec les autres nœuds du réseau.

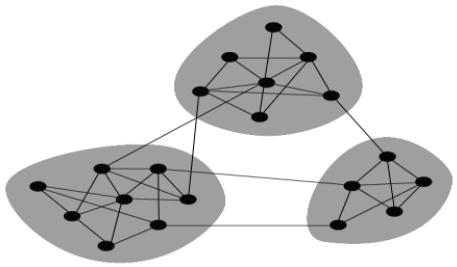


Figure 03. Structure de trois communauté dans un réseau

Diverses définitions adoptées

La première question qu'on se pose inévitablement est : comment formuler mathématiquement l'appartenance d'un nœud à une communauté plutôt qu'à une autre ? Si son sens sociologique est assez clair pour un être humain, il devient tout à fait inexploitable lorsqu'il s'agit d'un calcul exact.

Dans la théorie des graphes, il existe déjà quelques définitions qui répondent à ces contraintes. Nous citons :

- **Un clique** d'un graphe est un sous ensemble de sommets tous en relation deux à deux. Elle définit un sous graphe à n sommets et n (n − 1) /2 arêtes.

– Une **k-clique** est un sous-graphe complet maximal de k sommet, où chaque nœud est lié avec tous les autres. Il est certain qu'une k-clique possède les propriétés requises pour une communauté.

La définition forte de la communauté tient compte de chaque nœud du sous ensemble.

Ainsi, un sous-graphe V est considéré comme une communauté si :

$$K_i^{in}(V) > K_i^{out}(V), \forall\, i \in V$$

où $K_i^{in}(V)$ représente le nombre de liens du nœud i avec les nœuds de l'ensemble V et

$K_i^{out}(V)$ Est le nombre de liens du noeud i vers l'extérieur de la communauté formée par les nœuds de l'ensemble V.

Cette définition exige une contrainte sur chaque nœud, ce qui n'est pas le cas pour une communauté au sens faible dont la contrainte est donnée par :

$$\sum i \in V\, K_i^{in}(V) > \sum i \in V\, K_i^{out}(V).$$

Pour ainsi dire, dans une communauté forte chaque nœud a plus de liens vers sa communauté que vers le reste du graphe et dans une communauté faible le nombre de liens internes est supérieur au nombre de liens externes.

Une nouvelle définition, encore moins contraignante, d'une communauté a été proposé récemment par Hu et al.

Dans leur définition, un nœud i fait partie de la communauté V_a s'il a au moins autant de liens vers cette communauté que vers n'importe quelle autre communauté, dans d'autres termes, ils vérifient la relation suivante :

$$K_i^{in}(V_a) \geq K_i^{out}(V_b), \forall\, b \neq a\,, \forall\, i \in V$$

IL n'existe pas donc une définition usuellement notoire pour la notion de communauté, ce qui rend plus difficile de construire un algorithme accepté par tous, néanmoins cela offre la

possibilité de diversifier et de particulariser les approches proposées pour détecter ce type de structures. [13]

II.2 Définitions fondées sur une fonction de qualité du partitionnement

Une définition plus moderne de la notion de communauté s'appuie sur l'utilisation d'une fonction de qualité. En effet, un nombre important de chercheurs se préoccupent de cet ensemble de définitions et ce depuis la publication en 2004 d'un article par Newman et Girvan [14] . Une fonction de qualité est une fonction qui associe à un sous-graphe S une valeur quantifiant le fait que S correspond à une communauté. Suite à l'apparition de cet article de référence, plusieurs fonctions de qualité ont été proposées par d'autres chercheurs. Nous présentons ici deux de ces fonctions. La première est largement utilisée, il s'agit de la modularité de Newman, la deuxième est celle de Mancoridis [15].

II.2.1. Modularité de Newman

Connue sous le nom de modularité réseau Q et largement utilisée dans les travaux sur les réseaux, cette fonction de qualité a été introduite par Newman et Girvan [16] ; il s'agit d'une métrique de qualité pour l'évaluation du partitionnement d'un réseau en communautés. Elle est basée sur l'idée intuitive que les réseaux aléatoires ne renferment pas de structure communautaire. Par conséquent un sous-graphe forme une communauté si la distribution des liens entre ses nœuds n'est pas due au hasard.

Elle s'appuie donc sur la différence entre le nombre de liens à l'intérieur d'une communauté et le nombre de liens attendus à l'intérieur de cette communauté si les liens apparaissent aléatoirement dans le graphe tout en respectant la distribution des degrés des nœuds.

Une communauté est d'autant mieux appréciée que sa proportion d'arêtes internes sera supérieure à sa proportion attendue d'arêtes. Ceci est synthétisé cette équation donnant la modularité, et que l'on cherchera toujours à maximiser :

$$Q = \sum_{C_i} e(C_i) - a(C_i)^2$$

Où e (C_i) représente la proportion d'arêtes ayant les deux extrémités dans C_i et, a(C_i) est la probabilité pour qu'une arête ait une extrémité dans C_i.

L'expression détaillée de la modularité de Newman est donnée par l'équation suivante :

$$Q = \frac{1}{2m} \sum_{i\,j} \left(A[i,j] - \frac{k_i\, k_j}{2m} \right) \partial(C_i, C_J)$$

Où $\partial(C_i, C_J) = \{^{1\ si\ C_i=C_j}_{0\ sinon}$

La modularité Q peut avoir des valeurs entre -1 et 1 : $-1 \leq Q \leq 1$. Une partition contenant une communauté unique regroupant tous les sommets, possède une modularité de valeur nulle. Cette définition a été introduite pour des graphes non pondérés, mais nous pouvons l'adapter aux graphes pondérés en remplaçant le nombre d'arêtes internes par le poids total des arêtes internes, et en remplaçant le nombre total d'arêtes par le poids total du graphe.

II.2.2. Modularité de Mancoridis

La modularité *MQ* (Modularity Quality) a été introduite par Mancoridis [17] et a pour fin de mesurer la qualité d'une partition de nœuds du réseau. Cette métrique repose sur la différence entre les ratios de connectivité interne et externe des communautés. Elle est définie comme suit par l'équation :

$$MQ = \frac{1}{K}\sum_{i=1}^{K}[\,S(C_i, C_j) - \frac{1}{(K-1)}\sum_{i,j=1\ i\neq j}^{K} S(C_i, C_j)]$$

Avec $S(C_i, C_j) = \frac{E(C_i,C_j)}{|C_i|+|C_j|}$

E (C_i , C_j) représente le nombre de liens entre les nœuds des classes C_i et C_j, et le produit $|C_i|$ x $|C_j|$ représente le maximum de liens qu'on peut avoir entre les nœuds des classes C_i et C_j.

II.2.3. Intérêt de la détection de communauté

La détection de communauté permet de faciliter la compréhension des structures et leur fonctionnement. L'analyse de la structure réseaux permet de mettre en évidence des communautés qui sont des sous-ensembles de nœuds très fortement liés par rapport au reste des nœuds du graphe.

L'identification de ce type de structure est intéressante à plus d'un titre. En effet, cette structure existe dans de nombreux réseaux réels, et possède, dans la plupart du temps, une signification concrète en termes d'organisation. De ce fait, ces communautés permettent de donner un point de vue macroscopique sur la structure des réseaux. Elles sont sujet à différentes interprétations suivant le type de réseau considéré. Ainsi, elles correspondent, dans les réseaux sociaux à des ensembles d'individus ayant des points communs et dont les liens sociaux sont certainement plus forts, par exemple, des groupes d'individus avec des intérêts communs, des activités communes, etc. Contrairement à ce qui précède, dans les réseaux biologiques d'interactions protéine-protéine, les communautés correspondent communément à des ensembles de protéines qui collaborent à une même fonction cellulaire. Les propriétés

topologiques des réseaux d'interactions protéine-protéine sont les clés de la compréhension des maladies et devraient permettre aux biologistes de trouver des objectifs thérapeutiques précis.

III. Approches de détection des communautés

Nous allons lister ici les principales approches qui ont été proposées à ce jour. Quoiqu'elle soit importante, cette liste est non exhaustive : en vue d'en limiter la longueur, nous n'avons retenu que les approches qui ont reçu le plus d'attention de la part de la communauté scientifique. Notre but est de proposer une vue d'ensemble des méthodes proposées, et d'en illustrer la diversité. Afin d'organiser la présentation, nous avons essayé de les regrouper en fonction du type d'approche utilisée. Nous nous sommes tout d'abord focalisés sur les méthodes basées sur les graphes, pour passer par la suite aux méthodes utilisant l'analyse formelle de concept.

III.1. Approches basés sur les graphes

III.1.1 Approche statique sans chevauchement

La première méthode moderne pour la détection de communautés, encore utilisée dans plusieurs domaines et sur laquelle de nombreuses méthodes postérieures sont basées, est celle proposée en 2002 par Girvan et Newman. Nous avons jugé important de présenter cette méthode, car, bien que certaines méthodes proposées ultérieurement donnent de meilleurs résultats, l'approche statique sans chevauchement reste à leur origine et représente la source de leurs principes.

III.1.1.1 Approches hiérarchiques

A. Approches hiérarchiques ascendantes (Agglomérative)

L'idée commune de toutes ces méthodes est bien d'utiliser une approche ressemblant à celle du clustering hiérarchique, dans laquelle les sommets sont regroupés itérativement en communautés en partant d'une partition de *n* communautés composées d'un seul sommet. Les regroupements de communautés sont poursuivis jusqu'à obtenir une seule communauté regroupant tous les sommets, et une structure hiérarchique de communautés (dendrogramme) est ainsi construite. Ces approches sont similaires aux approches séparatives à la différence qu'elles travaillent de bas en haut dans la hiérarchie des communautés au lieu de travailler de haut en bas.

-L'algorithme d'optimisation de la modularité proposé par Newman [18]**et amélioré par Clauset et al** [19]

Une notion de modularité a été introduite par Newman ; il s'agit d'une fonction Q mesurant la qualité d'une partition du graphe en communautés. (Elle se base sur les proportions d'arêtes internes aux communautés et les proportions d'arêtes liées à chaque communauté). Pour mieux maximiser cette quantité ; l'algorithme glouton proposé fusionne à chaque étape les communautés permettant d'avoir la plus grande augmentation de la modularité. Afin de perfectionner les performances de l'algorithme seules les communautés ayant une arête entre elles peuvent être fusionnées à chaque étape. Chaque fusion se fait en O(n) et la mise à jour des valeurs des variations de Q (pour chaque nouvelle fusion possible) peut être effectuée facilement en O(m). La complexité globale est donc O ((m + n) n) = O(mn). Cette méthode considérée comme très rapide, permet de traiter de très grands graphes. En revanche, la qualité des partitions obtenues est moins bonne qu'avec des méthodes plus coûteuses.
En utilisant une structure de données adaptée, Clauset a pu améliorer la complexité de cette méthode. Une autre optimisation de cette méthode a été dernièrement proposée par Wakita et Tsurumi [80] afin de traiter des graphes de taille encore plus importante.

-Méthode de Louvain : [20] **Méthode d'optimisation de la modularité.**

Introduite initialement par Girvan et Newman, la modularité est une fonction de mesure de la qualité d'un partitionnement ; elle a pour finalité de choisir une coupe privilégiée dans un dendrogramme issu d'un clustering hiérarchique. La modularité mesure le nombre d'arêtes à l'intérieur des communautés moins ce même nombre obtenu sur un graphe aléatoire (c.à.d. sans structure) de même taille mais gardant exactement la même distribution de degrés.

L'idée est que dans un graphe aléatoire, certains liens ont naturellement une forte chance d'exister (notamment entre 2 nœuds de très fort degrés). L'existence de ce lien dans le graphe réel ne sera donc pas un argument pertinent pour considérer que ces 2 nœuds sont effectivement dans la même communauté. Au contraire, l'existence d'un lien entre 2 nœuds ayant peu de chances d'être liés dans le graphe aléatoire sera un argument fort pour les regrouper. Louvain utilise une méthode d'optimisation gloutonne de la modularité. Au départ chaque sommet est dans sa propre communauté, puis chaque sommet prend la communauté d'un de ces voisins de tel sorte que le gain de modularité soit maximal. Cette opération est répété plusieurs fois sur l'ensemble des sommets jusqu'à ce qu'aucun gain ne soit possible. Toute l'opération est ensuite répétée sur le graphe des communautés.

B. Approches hiérarchiques descendantes (Séparatives)

L'idée commune à toutes ces méthodes est d'essayer de scinder le graphe en plusieurs communautés en retirant progressivement les arêtes reliant des communautés distinctes. Les arêtes sont retirées une à une, et à chaque étape les composantes connexes du graphe obtenu sont identifiées à des communautés. Le processus est répété jusqu'au retrait de toutes les arêtes. On obtient alors une structure hiérarchique de communautés (dendrogramme), comme pour les méthodes de clustering hiérarchique. Les méthodes existantes diffèrent par la façon de choisir les arêtes à retirer.

-L'algorithme de Girvan et Newman basé sur la centralité d'intermédiarité [21]

Cette approche retire les arêtes de plus forte centralité d'intermédiarité. Cette centralité est définie pour une arête comme le nombre de plus courts chemins passant par cette arête. Il existe en effet peu d'arêtes reliant les différentes communautés et les plus courts chemins entre deux sommets de deux communautés différentes ont de grandes chances de passer par ces arêtes. Un algorithme calculant la centralité de toutes les arêtes en O(mn) est proposé. Ce calcul est effectué à chaque étape sur le graphe obtenu après retrait des arêtes. La complexité de l'algorithme est donc $O(m^2n)$. Une variante considérant des marches aléatoires à la place des plus courts chemins est aussi introduite. Elle donne des résultats légèrement meilleurs mais demande plus de calculs.

Nous devons rappeler que le même algorithme de calcul de la centralité d'intermédiarité des arêtes a aussi été introduit en même temps par Brandes [22].

- l'algorithme Edge Betweenness

Cette approche s'appuie sur le de chemins géodésiques entre les différents nœuds. Par conséquent, Il a une grande valeur de centralité d'intermédiarité. Cependant, quelques liens se trouvant entre les communautés et qui sont plus longs que les chemins géodésiques, ont des EBC (centralité d'intermédiarité d'un lien) faibles. D'où l'idée d'éliminer, dans chaque itération, le lien le plus intermédiaire (i.e., ayant la plus grande valeur EBC), puis recalculer à nouveau les EBC de tous les liens restants. [23]

L'algorithme est constitué de quatre étapes principales :

1- Calculer les EBC pour tous les liens du graphe.

2- Éliminer le lien le plus intermédiaire.

3- Recalculer les EBC de tous les liens restants.

4- Répéter les étapes 2 et 3 jusqu'à l'élimination de tous les liens.

La méthode Edge Betweenness commence par une seule communauté C contenant tous les nœuds du graphe, après chaque itération, un nouveau sous-graphe apparait, de nouvelles communautés C1 et C2 se génèrent avec C = C1 ∪ C2

III.1.1.2. Les approches utilisant des marches aléatoires

Les marches aléatoires dans les graphes représentent des processus aléatoires dans lesquels un marcheur est positionné sur un sommet de graphe et peut à chaque étape se déplacer vers un des sommets voisins. Le comportement des marches aléatoires est étroitement lié à la structure du graphe, ainsi plusieurs approches de détection de communautés se basent sur ces comportements.

-Algorithme Walktrap: [24] C'est un algorithme qui utilise une distance entre sommets basée sur des marches aléatoires. Une marche aléatoire courte, partant d'un sommet donné, tend à rester dans la (les) communauté(s) de ce sommet. Ainsi la distance entre les résultats de deux marches aléatoires partant de deux sommets distincts, révèle efficacement l'appartenance commune ou non de ces sommets à une même communauté. Cette distance permet à cette méthode de partitionner le graphe par l'intermédiaire d'un algorithme de clustering hiérarchique

-Infomap [25]: De même que Walktrap cette méthode exploite le fait qu'un marcheur suivant aléatoirement les arêtes du graphe a tendance à rester bloqué dans les communautés. Si on décrit un parcours aléatoire sur le graphe comme une séquence de numéros indiquant soit le code de la communauté dans laquelle entre le marcheur, soit le code du sommet sur lequel il arrive (code propre à la communauté courante), alors un bon partitionnement doit permettre de compresser au mieux cette séquence, les changements de communautés devant être rares. Le nombre de bits utiles, en moyenne, pour coder un pas du marcheur est défini théoriquement (pour un partitionnement donné) permet de mesurer la qualité d'un partitionnement. La détection de communautés en elle-même est ensuite réalisée par une méthode d'optimisation similaire à Louvain.

-FCD : fast community detection algorithm [26]

C'est un algorithme de détection de communautés dans les réseaux. Il initie chaque sommet à la rechercher pour indépendamment une communauté voisine.

Chaque sommet choisit sa communauté sur la base des connaissances des degrés et des coefficients de clustring des voisins et le nombre de communes voisines.il est à noter que cet algorithme travail aussi sur des communautés statiques avec chevauchement.

III.1.2. Approche statique avec chevauchement

Si l'idée de la hiérarchie de communautés était déjà présente implicitement dans les premières méthodes de détection de communautés telles que celle de Newman, il a fallu attendre plus longtemps pour qu'une autre caractéristique des réseaux soit prise en compte : le recouvrement de communautés. Cette propriété, observée depuis longtemps dans les réseaux sociaux.

Si l'on considère, à titre d'exemple un réseau d'individus dans lesquels les liens ne sont pas limités à un type précis d'interaction, mais peuvent correspondre à tout type de relation (comme on peut le trouver dans les réseaux sociaux Web 2.0 tels que Facebook), il est évident que chaque personne appartient à plusieurs groupes, ou communautés : les personnes de sa famille, ses collègues de travail, ses amis de lycée et de l'université, etc. Même si l'on ne considère que le réseau égocentré de cette personne, il est probable qu'au moins l'une des personnes de sa famille fasse aussi partie de son cercle d'amis du lycée, ou autre appartenance multiple.

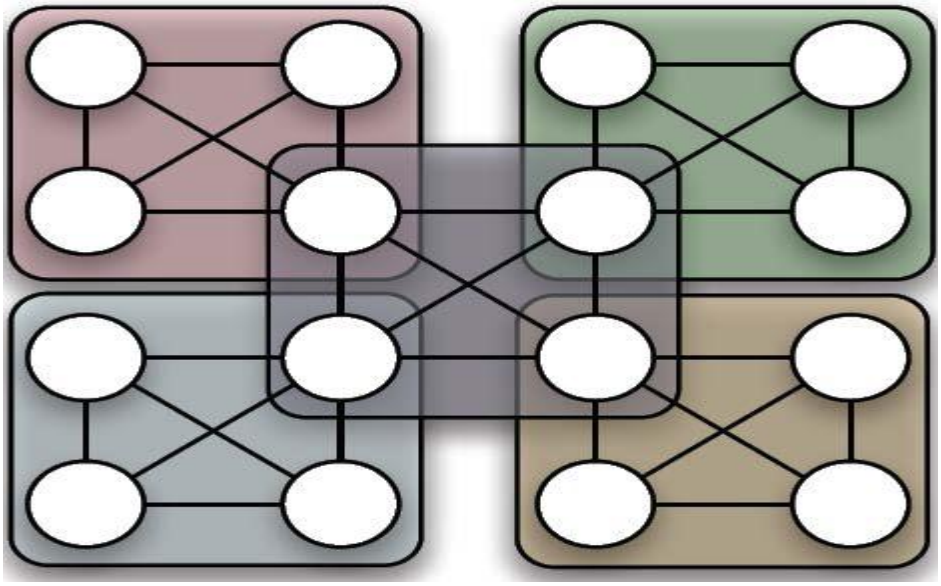


Figure 04 : Exemple d'un graphe avec communauté recouvrante [45]

A. Approches basées sur des cliques

Afin de trouver des communautés chevauchantes, une première solution proposée et consiste d'utiliser les cliques du réseau. Les cliques sont des ensembles de nœuds dont tous les membres sont liés à tous les autres. On sait que dans de nombreux réseaux, les relations ont tendance à être transitives : si les nœuds i et j sont connectés à un même nœud k, alors la probabilité qu'un lien existe entre i et j est augmentée, ce qui créé naturellement des cliques.

Les communautés étant composées de nœuds fortement connectées entre eux, elles sont riches en cliques, dans les réseaux pour lesquels la transitivité est forte. Une définition caricaturale des communautés utilisée dans certains cas consiste d'ailleurs à considérer que les communautés correspondent aux cliques maximales du réseau, ce qui fournit des communautés recouvrantes. Certaines approches ont donc essayé d'utiliser les cliques comme élément de base de leurs communautés.

La première de ces méthodes est celle de Palla et al. [27], CPM, qui consiste d'abord à détecter toutes les cliques d'une taille donnée, puis à agglomérer celles ayant suffisamment de nœuds en commun, ce qui est moins restrictif que de simplement considérer les cliques maximales.

Kumpala et al. [28] quant à eux, ont proposé une optimisation de CPM, plus performante mais reprenant le même mécanisme.

Shen et al. [29] ont proposé un algorithme appelé EAGLE, qui utilise un dendrogramme de cliques. L'algorithme commence par identifier toutes les cliques maximales. Ce sont les communautés initiales. Ensuite, les communautés ayant le plus fort taux de similarité sont fusionnées, formant de nouvelles communautés, qui, à leur tour, pourront être fusionnées avec des communautés semblables.

B. Méthodes basées sur la propagation de labels

Le premier algorithme implanta l'idée de propagation de labels est proposé dans [30]. C'est un algorithme itératif où à chaque itération un nœud envoi son label à ses voisins directs, et reçoit ceux de ses voisins. Chaque nœud détermine le label majoritaire qu'il adopte pour l'itération suivante. Ces processus itératifs mènent à un consensus sur un label précis pour chaque groupe de nœuds. La propagation de labels peut se faire en mode synchrone ou asynchrone. L'avantage du premier mode est qu'il peut facilement passer à l''échelle vu le parallélisme du calcul. Par contre, il peut y avoir un problème de convergence lié à un échange infini de label entre deux nœuds. Ce problème a été évité dans le mode asynchrone. Une version semi-synchrone qui tente d'avoir les avantages de ces deux versions a été proposée dans [31].

III.1.3 Les approches dynamiques

Les algorithmes que nous avons présentées supra s'appliquent à des réseaux statiques. Or, la majorité des réseaux sont dynamiques.

Les réseaux dynamiques sont des réseaux évoluant dans le temps. Dès lors, une nouvelle définition de communautés dynamiques s'impose : c'est une succession de communautés statiques. L'évolution des communautés dynamiques peut se faire de diverses manières :

- ***La croissance et la contraction*** : correspondant à l'ajout et au retrait de nœuds d'une communauté existante.

-***La naissance et la mort de communautés*** : des nouvelles communautés peuvent apparaître, et d'anciennes communautés peuvent disparaître avec l'évolution de réseau.

- ***La fusion et la division de communautés*** : Deux communautés - ou plus - peuvent en effet, se fusionner en une seule au cours du temps. De manière semblable, une communauté peut se diviser en deux ou plus en communautés, plus petites que celle dont elles sont issues.

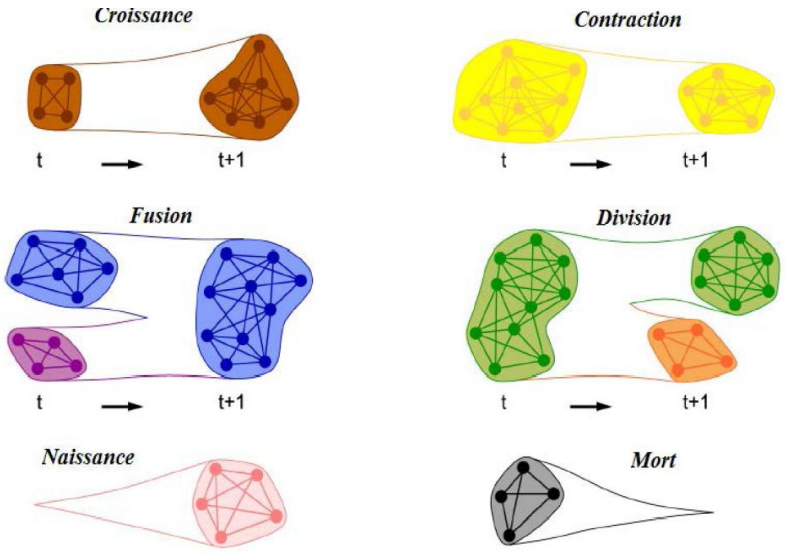


Figure05 Les différentes opérations possibles sur les communautés dynamiques [32]

III 1.3.1.Approches par détections statiques successives

Etant donné qu'elles ne font intervenir que des détections de communautés statiques suivies d'un post-traitement, ces méthodes sont considérées comme les plus simples. Cependant, toutes ces méthodes soufrent du problème de l'instabilité de la détection, Le principe général de ces approches est illustré dans la figure 06.

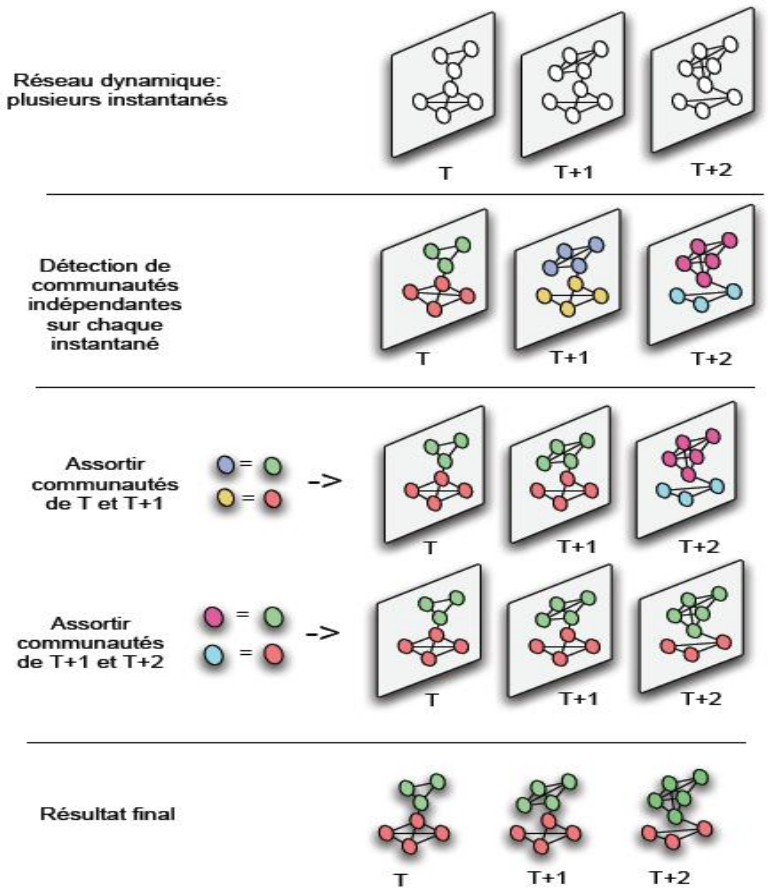


Figure 06 Illustration de l'approche par détections statiques successives. [33]

III.1.3.2 Approches par détections statiques informées successives

Il s'agit ici des approches qui utilisent toujours des instantanés, et effectuent une détection pour chacun d'entre eux. Cependant, pour résoudre le problème de l'instabilité des algorithmes, ces méthodes proposent de prendre en compte les résultats obtenus à l'étape t lors de la détection des communautés à l'étape t + 1. Ceci réduit l'instabilité, car, au cas où l'algorithme ne saurait lequel choisir entre deux découpages différents, il pourrait par exemple prendre le plus semblable au découpage précédent. Le principe général de cette approche est présenté dans la figure 07.

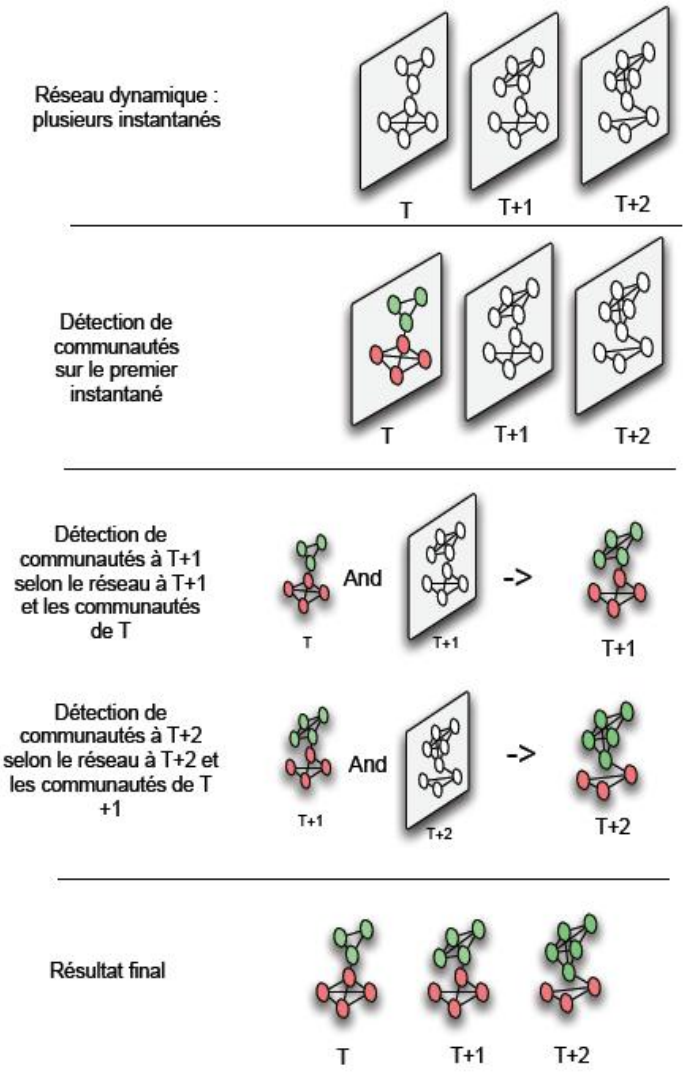


Figure 07 Illustration de l'approche par détections statiques informées successives [33].

III.1.3.3. Approches travaillant sur des réseaux temporels

La dernière section est réservée aux algorithmes travaillant directement sur le réseau temporel.

Ici, l'évolution du réseau n'est plus considérée comme une succession d'instantanés, mais comme une succession de modifications sur le réseau. Il s'agit donc de prendre en compte la ou les dernières modifications effectues sur le réseau, et de modifier les communautés existantes en conséquence. Il n'y a plus ici de problème d'instabilité, les communautés perdurant naturellement.

Toutefois, d'autres problèmes se posent, tel que l'aspect très local des modifications, qui peut entrainer une dérive vers des communautés qui ne sont plus valables à un moment donné,

par rapport à l'état du réseau à cet instant (la détection de communauté n'étant jamais faite sur l'ensemble du réseau, mais uniquement par petites modifications locales successives). Le principe général de cette approche est présenté sur la figure 08

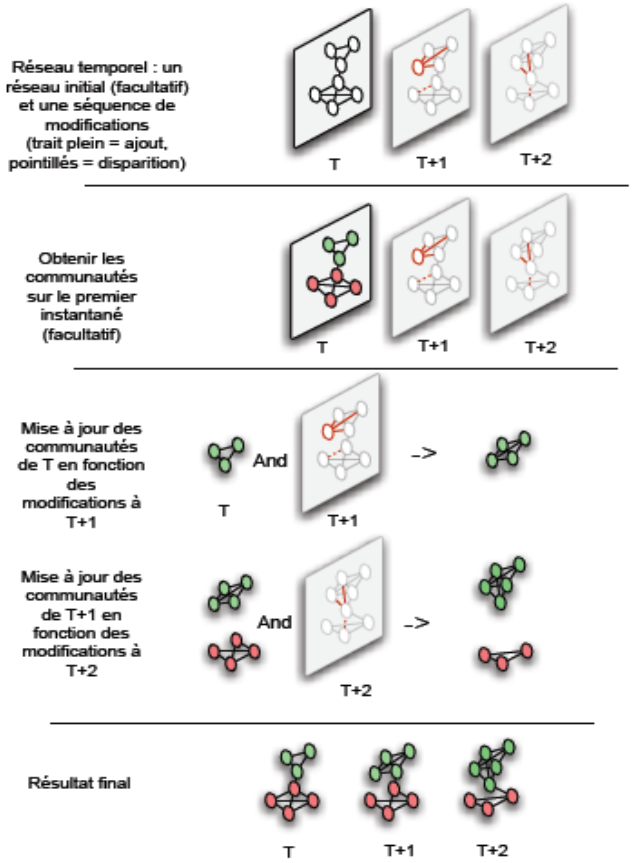


Figure 08 Illustration de l'approche par détection de communautés sur des réseaux temporels. [33]

III.2. Les approches basées sur l'analyse de concept formel

III.2.1 Quelques notions sur l'AFC

L'Analyse de Concepts Formels (ACF), appelée aussi Analyse Formelle de Concepts (AFC), est un formalisme qui constitue un pont entre les mathématiques, en particulier la théorie des ensembles ordonnés, et les applications d'analyse de données.

III.2.1.1 Théorie des treillis : Notions de base

Ensemble ordonné

Définition 1 (Relation binaire) Une *relation binaire R* entre deux ensembles *M* et *N* est un ensemble de couples d'éléments tels que m ϵ M et n ϵ N, i.e. un sous ensemble de M x N. (m, n) ϵ R (aussi noté par *m R n*) signifie que l'élément m est en relation R avec l'élément n. Si M = N, on parle de relation binaire sur M. R^{-1} est la relation inverse de R,

i.e. la relation entre N et M telle que $nR^{-1}m \Leftrightarrow mRn$.

Définition 2 (Relation d'ordre (partiel)) Une relation binaire R sur un ensemble E est dite *relation d'ordre partiel* (ou simplement relation d'ordre) sur E si elle vérifie les conditions suivantes pour tous $x, y, z \in E$:

1. $(x, x) \in R$ (R est réflexive)
2. si $(x ; y) \in R$ et $x \neq y$ alors $(y, x) /\in R$ (R est antisymétrique)
3. si $(x ; y) \in R$ et $(y, z) \in R$ alors $(x, z) \in R$ (R est transitive)

Une relation d'ordre R est souvent notée par $\leq$ (R^{-1} est notée par "$\leq$") et on dit que "*x* est plus petit que *y* " lorsque $x \leq y$.

Définition 3 (Ensemble ordonné) Un *ensemble partiellement ordonné* (ou simplement ensemble ordonné) est un couple $(E, \leq)$ où E est un ensemble et "$\leq$" est une relation d'ordre sur E.

Dans un ensemble ordonné $(E, \leq)$, deux éléments *x* et *y* de E sont dits comparables lorsque $x \leq y$ ou $y \leq x$, autrement ils sont dits incomparables. Pour deux éléments comparables et différents, $x \leq y$ et $x \neq y$, on note $x < y$. Un sous ensemble de $(E, \leq)$ dans lequel tous les éléments sont comparables est appelé ***chaîne.*** Un sous ensemble de $(E, \leq)$ dans lequel tous les éléments sont incomparables est appelé ***anti-chaîne.***

Définition 4 (Successeur, prédécesseur, couverture) Soient $(E, \leq)$ un ensemble ordonné et $x, y \in E$. Y est dit successeur de x lorsque $x < y$ et il n'existe aucun élément $z \in E$ tel que $x < z < y$. Dans ces cas, x est dit ***prédécesseur*** de y et on note $x \prec y$. Lorsque x est un prédécesseur de y on dit que x couvre y (et que y est couvert par x). La couverture de x est formée par tous ses successeurs.

Tous ensemble ordonné, $(E, \leq)$, peut être représenté graphiquement par un diagramme appelé ***"diagramme de Hasse"*** (ou diagramme de couverture) et obtenu comme suit :

1. Tout élément de E est représenté par une petit cercle dans le plan
2. Si $x, y \in E$ et $x \prec y$ alors le cercle correspondant à y doit être au-dessus de celui correspondant à x et les deux cercles sont reliés par un segment.

À partir d'un tel diagramme on peut lire la relation d'ordre comme suit : $x < y$ si et seulement s'il existe un chemin ascendant qui relie le cercle correspondant à x à celui correspondant à y.

Définition 5 (Principe de dualité des ensembles ordonnés) Soit $(E, \leq)$ un ensemble ordonné. La relation inverse "$\geq$" de "$\leq$" est aussi une relation d'ordre sur E. "$\geq$" est appelée ***duale*** de "$\leq$" et $(E, \geq)$ est appelé le dual de l'ensemble ordonné $(E, \leq)$.

Le diagramme de Hasse de $(E, \geq)$ peut être obtenu à partir de celui de $(E, \leq)$ par une simple réflexion horizontale. De plus, il possible de dériver les propriétés duales de $(E, \geq)$ à partir des propriétés de $(E, \leq)$.

III.2.1.2 Treillis

Définition 6 ***(Majorant, minorant, supremum, infimum)*** Soient (E, ≤) un ensemble ordonné et S un sous ensemble de E. Un élément a ϵ E est dit ***majorant*** de S lorsque a ≥ s ∀ s ϵ S.

De façon duale, a ϵ E est dit ***minorant*** de S lorsque a ≤ s ∀ s ϵ S.

Le plus petit majorant (respectivement minorant) de S, s'il existe, est appelé ***supremum*** ou borne supérieure (respectivement ***infimum*** ou borne inférieure) de S et noté V S (respectivement Λ S). Dans le cas ou S ={x,y}, V S et Λ S sont aussi notés par x V y et x Λ y respectivement.

Dans tout ensemble ordonné, lorsque le supremum (respectivement l'infimum) existe, il est unique.

Définition 7 ***(Treillis, treillis complet)*** Un treillis est un ensemble partiellement ordonné (E, ≤) tel que x V y et x Λ y existent pout tout couple d'éléments x, y ϵ E. Un treillis est dit ***complet si*** V S et Λ S existent pour tout sous ensemble S de E. En particulier, un treillis complet admet un élément maximal (top) noté par T et un élément minimal (bottom) noté par ⊥ Tout treillis fini est un treillis complet.

Définition 8 ***(Demi-treillis)*** Un ensemble ordonné (E, ≤) est un ***sup-demi-treillis*** (respectivement ***inf-demi-treillis)*** si tout couple d'éléments x, y ϵ E admet un supremum x V y (respectivement un infimum x Λ y).

III.2.1.3Connexion de Galois

Définition 11 (Connexion de Galois) Soient ϕ : P → Q et ψ : Q → P deux applications entre deux ensembles ordonnés (P, ≤P) et (Q, ≤Q). ϕ et ψ forment ***une connexion de Galois*** entre (P, ≤P) et (Q, ≤Q) si elles vérifient les conditions suivantes pour tous p, p_1, p_2 ϵ P et q ; q_1 ; q_2 ϵ Q :

1. si p_1 ≤P p_2 alors ϕ (p_2) ≤ Q ϕ (p_1),
2. si q_1 ≤Q q_2 alors ψ (q_2) ≤P ψ (q_1),
3. p ≤P ψ (ϕ (p)) et q ≤Q ϕ (ψ (q)).

Les conditions données dans la définition précédente sont équivalentes à la formule suivante :

p ≤P ψ (q) $\Longleftrightarrow$ q ≤Q ϕ (p)

III.2.2 Les approches existantes

III.2.2.1. Méthodes de Freeman pour la détection de communautés

Afin de déterminer les communautés Freeman [34] a associé le concept formel de la CM au concept formel du treillis de Galois. En outre, pour détecter des communautés, Freeman [34] s'appuie notamment sur la notion de chevauchement de CM dans le treillis de Galois. Freeman détermine à partir du treillis de Galois T(K), un ensemble de CM dont au moins deux chemins allant de sa position courante dans le treillis vers l'infimum ne sont pas de même longueur qu'il appelle cliques intermédiaires. Ensuite il procède à l'élimination des arêtes

partant de ces nœuds pour obtenir des groupes disjoints. Dans les grands treillis plusieurs nœuds appartiennent à des cliques intermédiaires et sont éliminés lors du processus de détection proposé par Freeman.

III.2.2.2 Méthode de Falzon pour la détection de communautés

Falzon [35] a proposé une nouvelle approche qui n'élimine pas les nœuds intermédiaires et la justifie à travers l'observation d'une propriété du treillis de CM. Falzon [35] propose une méthode de construction qui ne calcule qu'en partie ce dernier de manière polynomiale et qui classe tous les acteurs des CM. Pour la construction du treillis, il prend en entrée l'ensemble des CM L1 puis construit pour chaque niveau (k) un ensemble Lk créé à partir des intersections des paires de nœuds du niveau k−1. Ensuite, une liste LS[k] d'ensembles de nœuds pour chaque couche de niveau (k) est créée. Cette liste contient les nœuds dont les acteurs n'apparaissent pas aux niveaux supérieurs à k. Enfin, une comparaison de chaque paire de couches adjacentes du treillis est établie pour déterminer si elles ont des ensembles de nœuds communs, les nœuds communs de la couche supérieure sont éliminés. L'algorithme 1 de la structure de groupes proposé par Falzon appelle la fonction chevauchement nœud o plusieurs fois.
Cette fonction est une extension aux nœuds de la définition de chevauchement de CM, son appel s'effectue une fois pour chaque couche dans le réseau lors du processus de détection de groupes.

La création des groupes G_k se fait d'abord par la génération des groupes au niveau 1. Ces groupes sont formés en appliquant la fonction de chevauchement sur les nœuds du niveau 1, puis les groupes suivants (avec k > 1) sont formés en appelant la fonction de chevauchement sur les nœuds du niveau k et les nœuds contenus dans les listes LS [1] jusqu'à LS [k − 1]. Falzon [35] affirme que la méthode de détection de groupes de Freeman est bonne pour les contextes de petite taille dont le treillis de CM associé n'excède pas trois niveaux. Au-delà de trois niveaux le nombre d'acteurs appartenant aux cliques intermédiaires devient important et par conséquent ils sont éliminés de l'ensemble des groupes détectés. Falzon [35] répond à cette perte d'information pertinente en proposant une méthode qui n'élimine pas les acteurs appartenant aux cliques intermédiaires et qui détermine des groupes à chaque niveau du treillis, mais toujours à partir du même treillis de CM que Freeman propose de construire en premier. Or dans les graphes de RS un grand nombre d'acteurs n'appartient pas à des CM. L'exemple de RS construit par Falzon pour illustrer ces algorithmes comprend au départ 97 acteurs, ensuite 18 sont éliminés car ils n'appartiennent à aucune CM, ce qui représente 19 % du RS.

Algorithme 1 : L'algorithme de la structure de groupe selon la méthode de *Falzon*

```
1:  LS[0] = ∅
2:  for i := 0 to maxplayer − 1 do
3:     L := L_{i+1} ∪ LS[1] ∪ LS[2] ∪ ... ∪ LS[i] ;
4:     k := 1 ;
5:     while L non vide ; do
6:        Soit n le premier ensemble nœud de L ;
7:        GS := n ;
8:        Déterminer tous les ensembles nœud n_j tel que (n, n_j) ∈ o (fonction de
          chevauchement) ;
9:        G_k := ∪n_j ; GS := n_j ;
10:       L := L − GS ;
11:       Ajouter G_k à la liste de groupes pour la couche i ;
12:       k + +
13:    i := i + 1 ;
```

III.2.2.3 Méthode de Sid Ali Selmane pour la détection de communautés

Une nouvelle approche basée sur l'Analyse Formelle des Concepts (AFC) pour la détection de communautés dans un réseau social a été proposée par Sid Ali Selmane [36]. Ce dernier a proposé une fonction basée sur une modularité adaptée, appelée GroupNode modularity, une extension de la méthode de Falzon qui améliore une méthode de détection partielle proposée par Falzon en prenant en considération tous les acteurs du réseau social. Il a appelé son approche GNOM-FCA [37] (Group Node Modularity combined with Formal Concept Analysis approach). D'autre part, il a adapté une fonction issue du domaine de la recherche d'information, à savoir la F-mesure, dans le cas de classes multiples pour évaluer et comparer la qualité des communautés détectées. Enfin, il a validé son approche par des expérimentations sur des réseaux sociaux issus du monde réel connus dans le domaine.

Une telle idée consistant à trouver une approche qui considère l'ensemble des acteurs du RS est très intéressante. Effectivement, dans le domaine de détection de communautés, certains acteurs peuvent jouer des rôles importants malgré qu'ils n'entretiennent pas beaucoup de liens au sein du RS lui-même (en d'autres termes, ils n'appartiennent pas à des CM). Par exemple, dans une organisation terroriste, les personnes qui exécutent des attentats entretiennent peu de contacts avec les autres terroristes d'un même groupe, ils attendent juste un ordre d'une seule personne qui leur assigne une mission. Ainsi un seul contact permet au terroriste en question de commettre son acte. La modélisation de cette organisation en RS représentera ce contact par un seul lien et donc cet acteur n'appartiendrait à aucune CM. Dans les approches basées sur l'AFC décrites précédemment, ces acteurs n'appartiendront à aucune CM et de ce fait ne seront pas considérés lors du processus de détection. Par contre les éléments chargés de la logistique, de la préparation de cette organisation auront beaucoup de contacts entre eux et appartiendront à

des CM alors que ces éléments ne vont pas agir concrètement. D'où l'importance de considérer des éléments isolés lors du processus d'identification de communautés car non seulement un réseau de terroristes peut être démantelé mais aussi un attentat peut être évité en reliant ces éléments entretenant peu de contacts avec les groupes détectés.

La démarche que ***Sid Ali Selmane*** a proposé permet d'apporter une solution qui combine l'approche de Falzon pour la détection de communautés avec une notion de la théorie des graphes qui consiste en l'adaptation de la fonction de modularité de Newman [16]. Ce processus d'identification de communautés qu'il a proposé s'effectue en deux étapes. Dans la première étape il déterminera les groupes de niveaux G_k en appliquant l'algorithme de Falzon sur une partie du RS (les acteurs appartenant à des CM), puis dans la deuxième étape il construit son contexte formel K_ = (G (V, E), Gk, L1) formé du graphe du RS G (V, E), des groupes de niveaux Gk et l'ensemble des nœuds L1 représentant les CM au niveau 1 du treillis.
Ensuite, il a déterminé l'ensemble des nœuds Na du supremum du treillis, c.à.d. les acteurs qui appartiennent au graphe G (V, E) et n'appartenant pas à l'ensemble des CM L1. Il s'est basé sur l'hypothèse selon laquelle un acteur appartient à un groupe avec lequel il existe au moins un lien avec l'un des acteurs de ce groupe, pour chaque nœud ni appartenant à l'ensemble Na, il a généré des conteneurs $B(n_i)$ à partir de l'ensemble des arêtes E et l'ensemble des groupes G_k qui contiennent l'ensemble des paires (n_j, G_k) représentant les nœuds et les groupes avec lesquels le nœud ni a un lien. Une fois ce conteneur construit, il vérifie si le nœud ni a des liens uniquement avec des nœuds appartenant au même groupe ou avec des nœuds qui appartiennent à différents groupes. Dans le premier cas le nœud ni sera assigné au groupe auquel il a un ou plusieurs liens. Dans le second cas il a calculé la fonction de modularité adaptée par rapport aux différents groupes auxquels il a des liens, c.à.d. en considérant le nœud n_i appartenant à chacun des groupes G_k avec lesquels il partage un lien il a calculé et comparé les différentes valeurs de $Q\ (G_k, n_i)$ pour la maximiser, elle est donnée par :

$$Q(G_k, n_i) = \sum_j (e_{jj}(G_k) - a_j^2\ (G_k^*))$$

Où $e(G_k)_{jj}$ est la proportion d'arêtes à l'intérieur des groupes (nombre d'arêtes dans le groupe G_k en prenant en considération le nœud n_i divisé par le nombre total d'arêtes dans le graphe), $a_j\ (G_k^*) = \sum_j (e_{jj}\ (G_k^*))$ est la proportion d'arêtes attendue dans le graphe de G en assignant le nœud n_i au groupe G_k^*. Le nœud ni sera assigné au groupe dont $Q\ (G_k, n_i)$ est maximale.

Pour évaluer la précision des partitionnements qu'il a obtenus à travers chaque niveau il a adapté les mesures de Rappel et de Précision dans le cas des multiclasses à son approche

d'identification de communautés, soit P = (G_1,G_2, ...,G_k),$U_{(i=1..k)}$ Gi = V et $\forall\, i \neq j, G_i \cap G_j = \emptyset$; un partitionnement de V en k communautés. Initialement chaque nœud est attribué à une seule communauté G_j , cette attribution est déterminée soit par des études (ethnographiques, sociologiques, politiques, etc.) pour le cas des RS issus du monde réel, ou bien obtenu par le modèle de génération pour les RS synthétiques. A travers sa méthode pour chaque niveau du treillis il obtient un partitionnement P0 que nous comparons avec le partitionnement initial P, P' = (G'$_1$,G'$_2$, ...,G'$_l$), $U_{(i=1..l)}$ G'$_i$ = V et $\forall\, i \neq j, G_i \cap G_j = \emptyset$;La précision **P**(P, P') et le rappel **R**(P, P') du partitionnement sont donnés par les formules suivantes :

$$\mathbf{P}(\mathrm{P},\mathrm{P'}) = \frac{\sum_{i=1}^{l} \mathbf{P}\{G'_i\}}{l} \qquad \mathbf{R}(\mathrm{P},\mathrm{P'}) = \frac{\sum_{i=1}^{l} \mathrm{R}\{G'_i\}}{l}$$

La précision ***P{G'$_i$}*** représente le nombre d'individus correctement regroupés dans une communauté G0i par rapport au nombre d'individus initialement dans la communauté G_i et le rappel ***R{G'$_i$}*** le nombre d'individus correctement regroupés dans une communauté G'_i par rapport au nombre total d'individus dans la communauté G'_i.Ces mesures sont données par les formules suivantes :

$$\boldsymbol{P}\{G'_i\} = \frac{|G'_i \cap G_i|}{|G'_i|} \qquad \boldsymbol{R}\{G'_i\} = \frac{|G'_i \cap G_i|}{|G'_i|}$$

VI. Faiblesses des méthodes existantes

VI.1 concernant les approches basées sur des graphes

VI.1.1 Instabilité des algorithmes statiques

Les algorithmes statiques de détection de communautés dans les réseaux peuvent donner des résultats assez éloignés pour des graphes très proches. Ceci est lié à leur caractère heuristique, à leur non déterminisme ou au fait qu'il y a souvent plusieurs bonnes décompositions. Si les résultats varient trop, les événements suivis sont plus liés à l'algorithme qu'à un changement structurel et n'apportent donc pratiquement pas d'informations sur l'évolution du réseau.

VI.1.2 Problème de recouvrement

Le recouvrement : c'est un comportement important connu depuis longtemps dans le domaine des réseaux sociaux. Il désigne l'existence d'un mécanisme qui permet de classer les nœuds tout en permettant à certains d'appartenir à plusieurs communautés.

Dans le cas d'un graphe dynamique, une grande partie des transformations se fait continuellement : un ensemble de nœuds d'une communauté la quitte pour en rejoindre une autre par exemple, il est impossible de représenter ceci sans chevauchement et cela impose des évolutions faites d'importantes actions comme la division ou la fusion d'une communauté. Ainsi, une compréhension profonde de la dynamique nous semble difficile à représenter un tel cas de figure sans recouvrement.

VI.1.3 Problème lié au maximisation de modularité

Les méthodes de détection de communautés et en particulier celles basés sur la modularité souffrent de plusieurs problèmes, notamment ; la limite de résolution (elles sont incapables de détecter des petites communautés) ; la possibilité d'identifier des communautés dans des graphes n'ayant aucune structure communautaire et le non-déterminisme et l'absence de stabilité qui en découle. [38]

Fortunato et al [39] ont montré que tous les algorithmes basés sur la maximisation de modularité souffrent d'une limite de résolution qui empêche les petites communautés d'être détectées dans les grands réseaux. Ils ont plus précisément prouvé que les algorithmes de maximisation de la modularité ont du mal à trouver des communautés ayant moins de $\sqrt{\frac{m}{2}}$ liens, où m est le nombre de liens dans l'ensemble du réseau.

VI.2 Concernant les approches basées sur l'analyse de concept formelle

- Il est à noter que la méthode de Freeman pour la détection de communautés ; ne prend pas en considération l'ensemble des acteurs du RS.

Dans les grands treillis plusieurs nœuds appartiennent à des cliques intermédiaires et sont éliminés lors du processus de détection proposé par Freeman,Falzon [35] a affirmé que la méthode de détection de groupes de Freeman est bonne pour les contextes de petite taille dont le treillis de CM associé n'excède pas trois niveaux.

- La méthode de Falzon ne considère pas les individus qui appartiennent au graphe du réseau social et n'appartenant pas à l'ensemble des cliques maximales du réseau, Or dans les réseaux sociaux un grand nombre d'individus n'appartient pas à des cliques maximales.

Comme dans l'exemple de Sid Ali Selmane, dans une organisation terroriste, les personnes qui exécutent des attentats entretiennent peu de contacts avec les autres terroristes d'un même groupe, ils attendent juste un ordre d'une seule personne qui leur assigne une mission. Ainsi un seul contact permet au terroriste en question de commettre son acte. La modélisation de cette

organisation en RS représentera ce contact par un seul lien et donc cet acteur n'appartiendrait à aucune CM.

-La méthode de détection de communautés dans les réseaux sociaux appelée GNOM-FCA qui améliore la méthode proposée par Falzon, est une approche qui permet de considérer l'ensemble des acteurs d'un réseau social lors du processus d'identification des communautés tout en se basant sur un calcul assez compliqué, et en cas d'égalité il affecte aux communauté du dernier calcul de la modularité ; d'où l'inconvénient d'élargir une communauté par rapport aux autres.

- Les deux méthodes soit celle de Falzon ou de Sid Ali Selmane se basent sur une fonction de modularité et cette dernière malgré qu'elle présente une fonction de qualité importante mais elle n'est pas forcément le signe d'une structure modulaire, vu la limite de résolution qui empêche les petites communautés d'être détectées dans les grands réseaux.

Conclusion

Comme nous l'avons bien explicité dans ce chapitre, il existe moult méthodes de détection de communautés dans les réseaux sociaux ayant objectif de faciliter la compréhension des structures des communautés et leur fonctionnement.

Nous avons décrit les principales approches proposées dans la détection des communautés dans les réseaux, à savoir celles basées sur les graphes et celles portant sur l'analyse formelle de concepts.

Chapitre3 : Contribution

Chapitre3 : Contribution

I Introduction

Dans ce chapitre nous présentons notre approche de détection de communautés dans les réseaux sociaux basée sur l'analyse formelle de concept (AFC) et qui prend en considération tous les acteurs d'un réseau social ; car certains acteurs peuvent jouer des rôles importants malgré qu'ils n'entretiennent pas beaucoup de liens au sein du réseau social lui-même (autrement dit, même s'ils n'appartiennent pas à des cliques).

L'algorithme proposé s'inscrit dans la catégorie des approches statiques sans chevauchement et avec chevauchement, il affecte chaque acteur du réseau social à une communauté en basant sur le nombres de liens avec les cliques du graphe, et s'appuie sur deux paramètres : la maximisation de modularité et le calcul de la conductance.

On a choisi le paramètre maximiser la modularité car la modularité est un critère très naturel, et très proche de la définition des communautés. Et comme la modularité souffre de la limite de résolution qui empêche les petites communautés d'être détectées dans les grands réseaux on a préféré de travailler avec un autre critère qui est la conductance, elle aussi une fonction de qualité qui se base sur la différence entre le nombre de liens à l'intérieur d'une communauté et le nombre de liens attendus à l'intérieur de cette communauté.

Pour prouver l'efficacité de l'approche proposée, l'expérimentation est réalisée sur des réseaux sociaux issus du monde réel car l'identification des communautés est une tâche difficile à accomplir vu le nombre de communautés et leurs tailles qui ne sont pas connues à priori.

II. Approche proposée pour la détection de communautés

II.1 Principe de l'approche proposée

L'approche de détection de communautés, que nous allons présenter est basée sur l'analyse formelle de concept, en utilisant aussi les graphes.

Il s'agit d'une approche qui considère tous les acteurs du réseau social. Cette dernière est applicable sur des réseaux non orientés et non pondérés.

C'est une approche de détection de communautés basée sur la maximisation de modularité en utilisant une fonction de qualité qui permet de trouver des structures meilleures de communautés de graphe, et comme la modularité souffre d'un problème de limite de résolution dans le sens qu'ils ne peuvent pas distinguer des communautés plus petites d'une certaine taille limite ; nous avons estimé intéressant de travailler sur un autre paramètre ; outre la maximisation de la modularité la conductance.

-La fonction de qualité introduite par Newman et Girvan [16] connue sous le nom de la **modularité réseau Q**, est une métrique de qualité pour l'évaluation du partitionnement d'un réseau en communautés.

$$Q = \frac{1}{2m} \sum_{i\,j} \left(A[i,j] - \frac{k_i\, k_j}{2m} \right) \partial(C_i, C_J)$$

Ou $\partial(C_i, C_J) = \{^{1\ si\ C_i = C_j}_{0\ sinon}$

A_{ij} : matrice adjacence,

k_i: degré du sommet i (d(i))

m : le nombre de total d'arêtes dans le graphe

La fonction $\partial(C_i, C_J)$est égale à 1 si i et j appartiennent à la même communauté et 0 sinon.

-La conductance est une fonction de qualité basée sur le nombre de liens internes et externes, [40] elle compare le nombre de liens internes et externes de la communauté et plus la conductance d'un ensemble de nœuds est faible, plus cet ensemble est censé être une bonne communauté, s'appuie sur la différence entre le nombre de liens à l'intérieur d'une communauté et le nombre de liens attendus à l'intérieur de cette communauté.

Une communauté est d'autant mieux appréciée lorsque sa proportion d'arêtes internes sera supérieure à sa proportion attendue d'arêtes.

La conductance a été et demeure une fonction de qualité de choix pour évaluer la pertinence d'un ensemble de nœuds en tant que communauté.

La conductance pour un ensemble de sommets S est définie comme [41]:

$$\text{conductance}(S) = \frac{C_s}{2m_s + C_s}$$

Cs est le nombre des arêtes inter-communauté.

m_s est le nombre des nœuds dans la communauté.

Définitions :

Graphe. Soit G = (V, E) un graphe représentant un réseau social où :

– V est l'ensemble des acteurs (nœuds) $\{x_i\}n_{i=1}$ du réseau social et n = |V | est le nombre de nœuds dans G.

– E est l'ensemble des liens sociaux entre les acteurs et m = |E| est le nombre d'arêtes dans G.

- **Un clique** d'un graphe est un sous ensemble de sommets tous en relation deux à deux. Elle définit un sous graphe à n sommets et n (n − 1) /2 arêtes.

-Une **k-clique** est un sous-graphe complet maximal de k sommet, où chaque nœud est lié avec tous les autres.

- **Contexte formel**. Soit F = (V, C, I) le contexte formel associant les cliques d'un graphe à L'ensemble de ses acteurs V où :

- C est l'ensembles des cliques $\{Cj\}^{k}_{j=1}$ extraites du graphe G.

- I est la relation binaire qui lie les ensembles V et C. (si un acteur x_i appartient à une clique Cj $I(x_i,Cj) = 1$ sinon Cj , $I(x_i,Cj) = 0$.

Le principe sur lequel est fondu notre travail est similaire à celui de représentation du réseau sociaux basés sur l'AFC qui fait correspondre le concept formel de la clique au concept de l'ensemble dans la théorie des ensembles en se basant sur des opérateurs de base qui sont :

L'intersection : L'ensemble intersection de deux cliques A et de B, noté « A ∩ B » est l'ensemble des éléments de clique A qui sont également éléments de clique B, soit :

$A \cap B = \{ x \in U \,|(x \in A) \wedge (x \in B)\}$ / *x* est un nœud du réseau

C'est-à-dire que :

$x \in A \cap B$ si et seulement si $x \in A$ *et* $x \in B$.

Deux cliques qui n'ont aucun élément en commun, c'est-à-dire que leur intersection est vide, sont dites disjointes.

L'union : L'union de deux cliques A et B est l'ensemble qui contient tous les éléments qui appartiennent à A ou appartiennent à B. On la note A ∪ B.

Formellement :

$$\forall x, x \in A \cup B \Leftrightarrow ((x \in A) \vee (x \in B))$$

- Par exemple l'union des cliques A = {1, 2, 3} et B = {2, 3, 4} est l'ensemble {1, 2, 3, 4}.

Ces deux opérateurs avec l'ensemble des communautés permettre mathématiquement d'obtenir un treillis de Galois.

-Les étapes de l'approche :

L'approche proposée pour la détection de communautés dans les réseaux sociaux contient plusieurs étapes comme l'illustre la figure09.

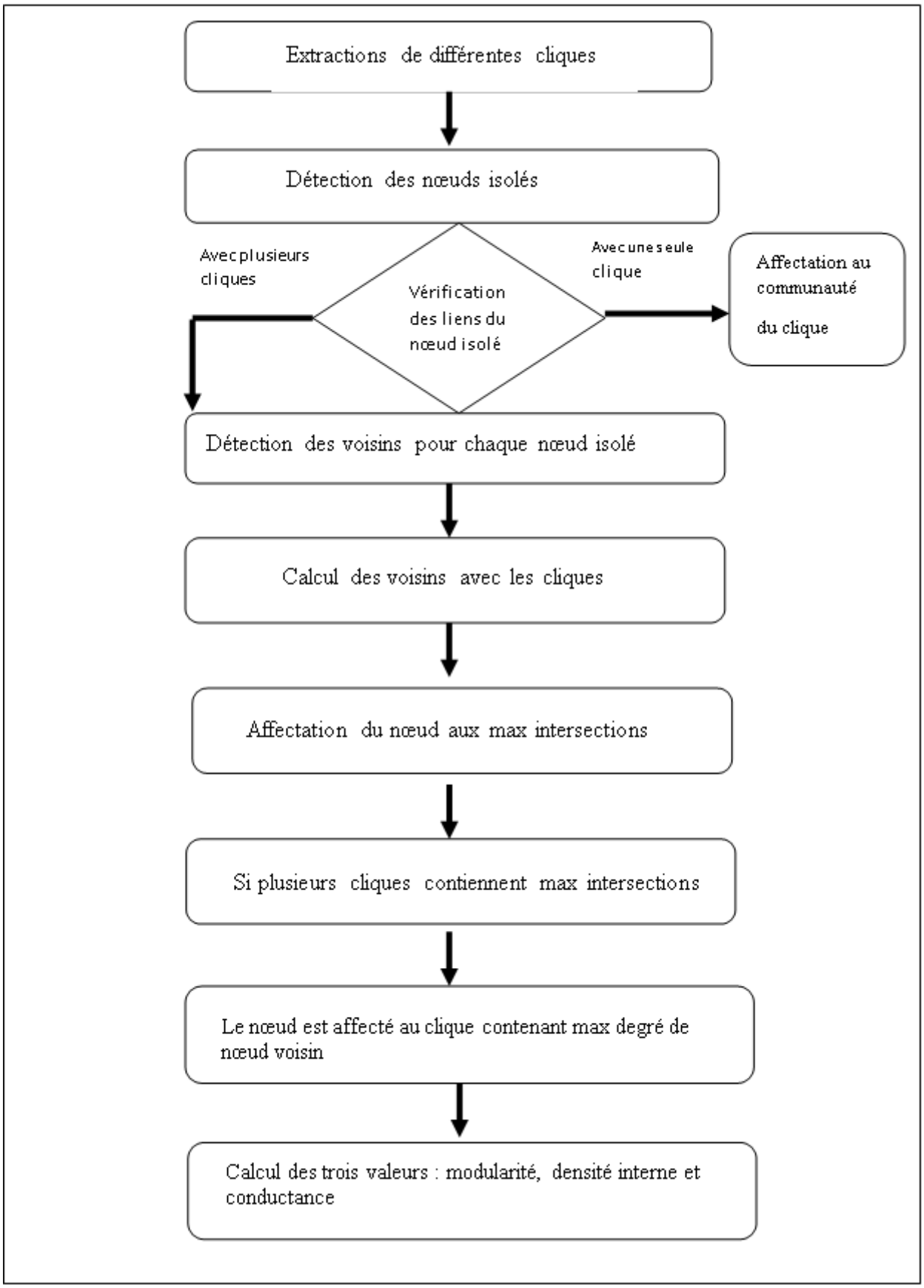


Figure09. Les étapes de l'approche

II.2 Algorithme

Algorithme : *Approche-AFC*

Entrées : $G = (V, E)$ *: le graphe initial.*

Extraction les cliques : $k_i \in \{k1,k2,\ldots\ldots,kn\}$ *en utilisant la fonction heuristique C-Finder.*

Sorties : $G_0 = (V_0, E_0)$: *le graphe après l'affectation des nœuds isolés*.

Variables : N : *nœud dans graphe G*

C : les communautés dans G ;

Ni : ensemble des nœuds isolés.

NVC : nombre de max voisin ;

MaxDR : *le maximum degré du clique à identifier* ;

Début

Pour *chaque nœud* $N \in G$ ***Faire***

Si $N \in K_i$ // $k_i \in \{k_1,k_2,\ldots\ldots,k_n\}$

Communauté(N) $\leftarrow C_i$ // $C_i \in \{C_1,C_2,\ldots\ldots,C_n\}$

// affecter le nœud dans la communauté C_i

Sinon

Ajoute (N, Ni)

// ajouter le nœud dans la liste des nœuds isolé Ni

Finsi

FinPour

Pour *chaque nœud* $n \in Ni$ ***Faire*** // $k_i \in \{k_1,k_2,\ldots\ldots,k_n\}$

NVC $\leftarrow$*nombre max de nœud (* $voisin(n) \cap k_i$ *)*

// voisin(n) est les voisins de nœud isolé

// NVC est le max intersection

Si *plusieurs cliques* k_i , k_j *contiennent NVC*

$MaxDR_i \leftarrow$ *Max degré (NVC* $\cap k_i$ *)* // $k_i \in \{k_1,k_2,\ldots\ldots,k_n\}$

$MaxDR_j \leftarrow$ *Max degré (NVC* $\cap k_j$ *)* // $k_J \in \{k_1,k_2,\ldots\ldots,k_n\}$

Si Ci $< C_j$

Communauté(N) $\leftarrow C_j$

Sinon

Communauté(N) ← C_i

Finsi

Sinon

Communauté(N) ← C_i // k_i *contient le Max NVC*

Finsi

Finpour

Fin

II.3 Détail des différentes étapes de l'approche

La détection de communautés se fait en quatre phases, les détails de chaque phase sont présentés dans les sections suivantes.

-La première phase vise à diviser le graphe à des cliques :

- Dans notre étude nous allons opter pour une étude ou K=4 ; (4-clique) généralement acceptée comme la plus efficace.

-Pour extraire les différentes cliques du graphe ; nous allons utiliser un outil gratuit appelé CFinder.

- En entrée nous fournissons au logiciel le réseau représenté par un graphe G (V, E).

-En sortie CFinder retourne l'ensemble des cliques C= {K_1, K_2, …K_n} extraites du graphe G.

-Chaque clique Kn contient un ensemble de nœuds : K_n= {n_1,n_2, …….n_n}

- En cas d'intersection entre deux cliques ou plus les cliques seront fusionnées en une seule.

Si $K_i \cap K_j \neq \phi$ alors $K_{ij}=K_i \cup K_j$

-La deuxième phase vise à identifier les nœuds isolés ; **un nœud isolé** est un nœud qui appartient au réseau (graphe) et n'appartenant à aucune clique.

-La troisième phase consiste à Détecter l'ensemble des voisins pour chaque nœud isolé du graphe, c.à.d. les nœuds avec les quelle un nœud isolé à des arrêtes.

-La quatrième phase porte sur l'affectation des nœuds isolés aux max intersections, En cas de conflit ; ou l'intersection d'un nœud isolé avec deux cliques ou plus est la même, l'affectation est basée sur le calcul du maximum degré des ensembles des nœuds de cliques ayant un lien

avec le nœud isolé puis la comparaison entre les résultats obtenus, le nœud sera donc affecté au clique qui contient un degré maximum.

Au cas où un nœud aurait un lien un lien avec une seule clique, il sera automatiquement affecté à ce dernier.

ET enfin le calcul des trois valeurs : modularité, densité et conductance et la visualisation des réseaux.

III. Expérimentation

Il est nécessaire voire indispensable d'effectuer des tests afin d'évaluer l'efficacité de notre approche. Pour ce faire, des expérimentations ont été menées, ces derniers consistent à affecter tous les nœuds du réseau, calculer la modularité, et la conductance.

A cet effet, un groupe de réseaux réels dont la structure est connue a été également utilisé.

III.1 Outil d'extraction des cliques

CFinder est un logiciel gratuit pour trouver et visualiser des groupes denses qui se chevauchent de nœuds dans les réseaux basés sur la méthode de percolation de clique ;

Une communauté est définie comme une chaîne de k-cliques adjacentes. Une k-clique est un sous-ensemble de k sommets tous adjacents les uns aux autres, et deux k-cliques sont adjacentes si elles partagent k - 1 sommets. L'avantage immédiat d'une telle approche réside dans la détection de communautés avec recouvrement, un sommet pouvant appartenir à plusieurs k-cliques non forcément adjacentes.

CFinder a été récemment appliqué à la description quantitative de l'évolution des groupes sociaux. Un cluster - appelé aussi une communauté - dans un réseau est un groupe de nœuds plus densément connectés les uns aux autres que de nœuds en dehors du groupe. Dans les réseaux réels des clusters se chevauchent souvent.

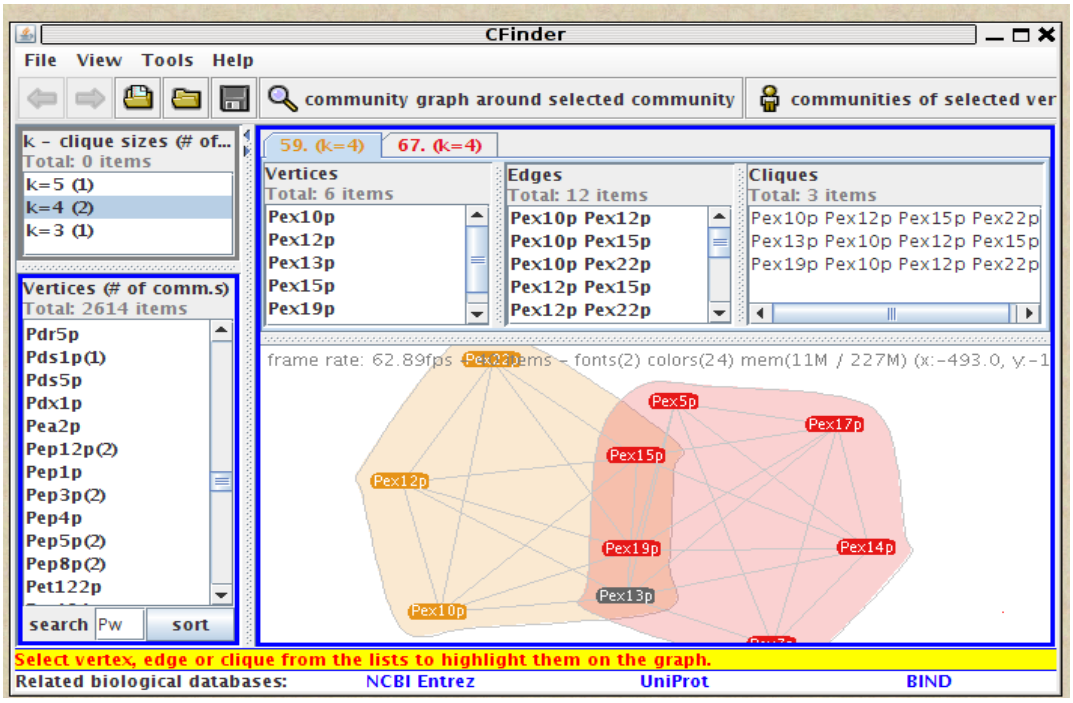


Figure10. Outil d'extraction de cliques CFinder

III.2 Outil de visualisation de réseaux

Pour la visualisation des réseaux nous avons opté pour le logiciel **Gephi.**

Gephi est un logiciel pour visualiser, analyser et explorer en temps réel les graphes (aussi appelés réseaux ou données relationnelles) de tout type. Sorte de Photoshop pour les réseaux, l'utilisateur interagit avec la représentation graphique, manipule les structures, les formes et les couleurs pour en révéler les propriétés cachées via des saillances visuelles.

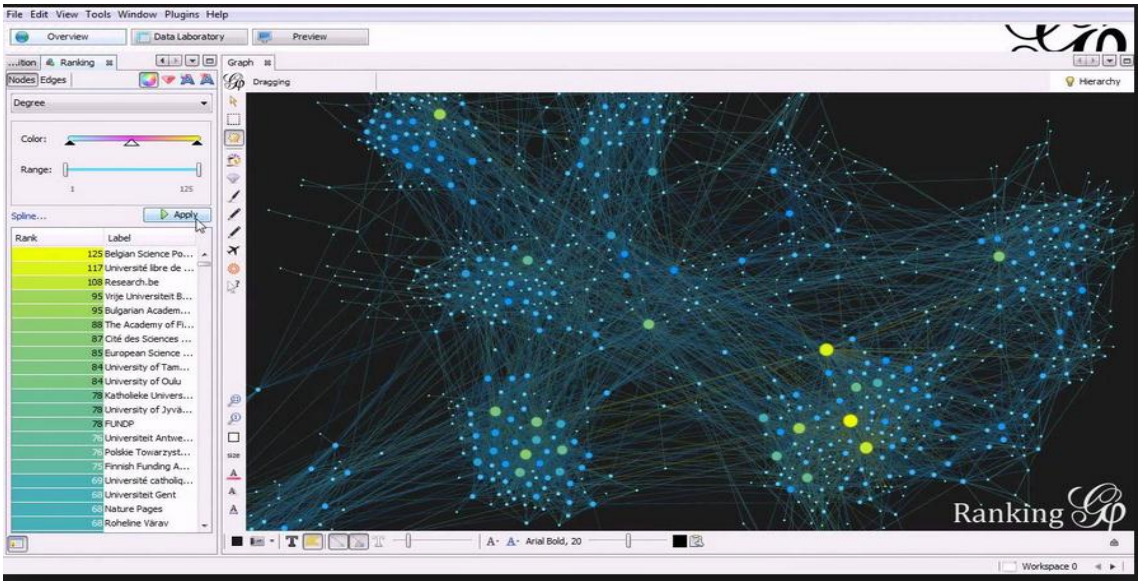


Figure 11. Outil de visualisation de réseaux Gephi

III.3 Choix du langage et de l'environnement d'implémentation

Microsoft Visual Studio est une suite de logiciels de développement pour Windows conçue par Microsoft. La dernière version s'appelle Visual Studio 2015.

Visual Studio est un ensemble complet d'outils de développement permettant de générer des applications web ASP.NET, des services web XML, des applications bureautiques et des applications mobiles. Visual Basic, Visual C++, Visual C# utilisent tous le même environnement de développement intégré (IDE), qui leur permet de partager des outils et facilite la création de solutions faisant appel à plusieurs langages. Par ailleurs, ces langages permettent de mieux tirer parti des fonctionnalités du Framework .NET, qui fournit un accès à des technologies clés simplifiant le développement d'applications web ASP et de services web XML grâce à Visual Web Développer.

III.4 les réseaux choisis pour l'expérimentation

L'approche proposée est testés sur des réseaux issus du monde réel ; des réseaux dont la structure de communautés est connue à l'avance.

L'algorithme est exécuté sur tous les réseaux utilisés, puis on calcul la fonction de modularité de Newman et la conductance pour chacun des réseaux.

Les réseaux utilisés dans l'expérimentation sont :

-Le club de karaté de Zachary

-Les dauphins de Lusseau

-Le football américain

III.4.1 Club de karaté de Zachary *[42]*

Le premier exemple est le club de karaté de Zachary. Il s'agit d'un réseau construit à partir des relations entre 34 membres d'un club de karaté dans une université aux États-Unis. C'est un réseau répandu et très utilisé par plusieurs algorithmes pour vérifier leurs performances puisque sa structure de communautés est connue à l'avance.

Comme l'illustre la figure qui suit, ce réseau comporte deux groupes, une ligne verticale sépare les deux communautés.

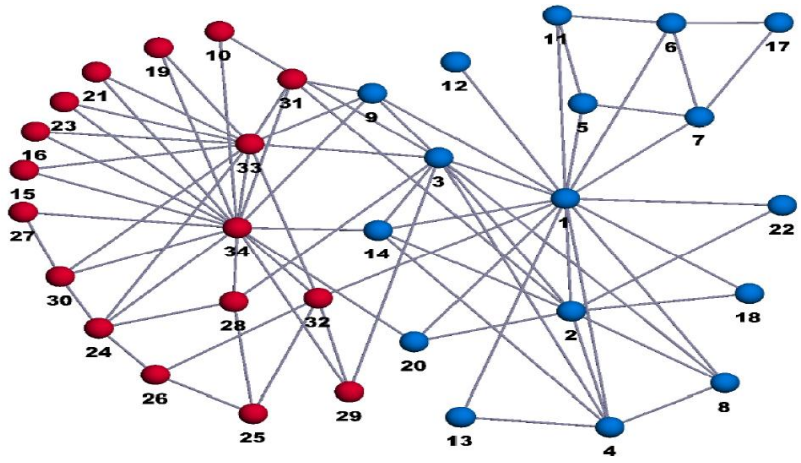


Figure 12. Club de karaté du Zachary [42]

III.4.2. Les dauphins de Lusseau

Le deuxième exemple est le réseau de dauphins de Lusseau [43]**.** Deux communautés constituent essentiellement ce réseau, elles sont séparées par une ligne verticale. Il contient 62 nœuds et 159 liens.

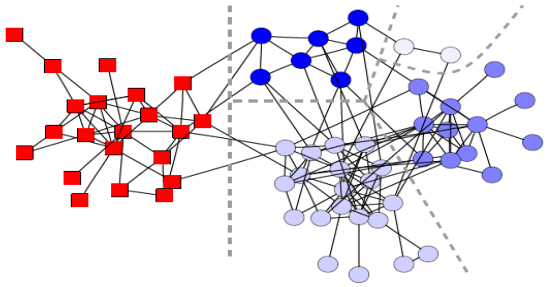


Figure 13. Les dauphins de Lusseau

III.4.3 Le graphe du Football Américain [44]

C'est un réseau de relations entre les équipes du football américain de première division entre des universités aux USA. Les nœuds du réseau représentent les équipes (étiquetés par les noms des collèges) et les liens représentent les matchs entre les équipes (un lien est mis entre deux équipes si elles se rencontrent durant la saison). Il représente le calendrier des matchs entre des équipes américaines de football durant l'année 2 000. Ce réseau est constitué de douze communautés, 115 nœuds et 613 liens.

VI. Expérimentation sur les différents réseaux

A. le réseau Club de karaté du Zachary

- Ce réseau contient 34 nœuds et 78 liens.

La base de données représentant les liens entre les différents nœuds est représentée dans le tableau suivant :

nœud	**relations**
1	2,3,4,5,6,7,8,9,11,12,13,14,18,20,22,32
2	1, 3, 4, 8, 14, 18, 20, 22,31
3	1, 2, 4, 8, 9, 10, 14, 28, 29,33
4	1, 2, 3, 8, 13,14,
5	1, 7,11
6	1, 7, 11,17
7	1, 5, 6,17
8	1, 2, 3,4
9	31, 33,34
10	3,34
11	1, 5,6
12	1
13	1,4
14	1, 2, 3, 4,34
15	33,34
16	33,34
17	6,7
18	1,2
19	33,34
20	1,2,34
21	33 ,34
22	1,2
23	33,34
24	34, 26, 28, 30, 33,34
25	26, 28,32
26	24, 25,32
27	30,34
28	3, 24, 25,34
29	3,32, 34
30	24, 27, 33,34
31	2, 9, 33,34
32	1, 25, 26, 29, 33,34
33	3, 9, 15, 16, 19, 21, 23, 24, 30,31
34	9,10,14,15,16,19,20,21,23,24,27,28,29,30,31,32,33

Tableau01 : Base de donnée du réseau club de Zachary

-L'étape d'extraction de cliques :

- Diviser le réseau en cliques en utilisant le logiciel CFinder : k = 4

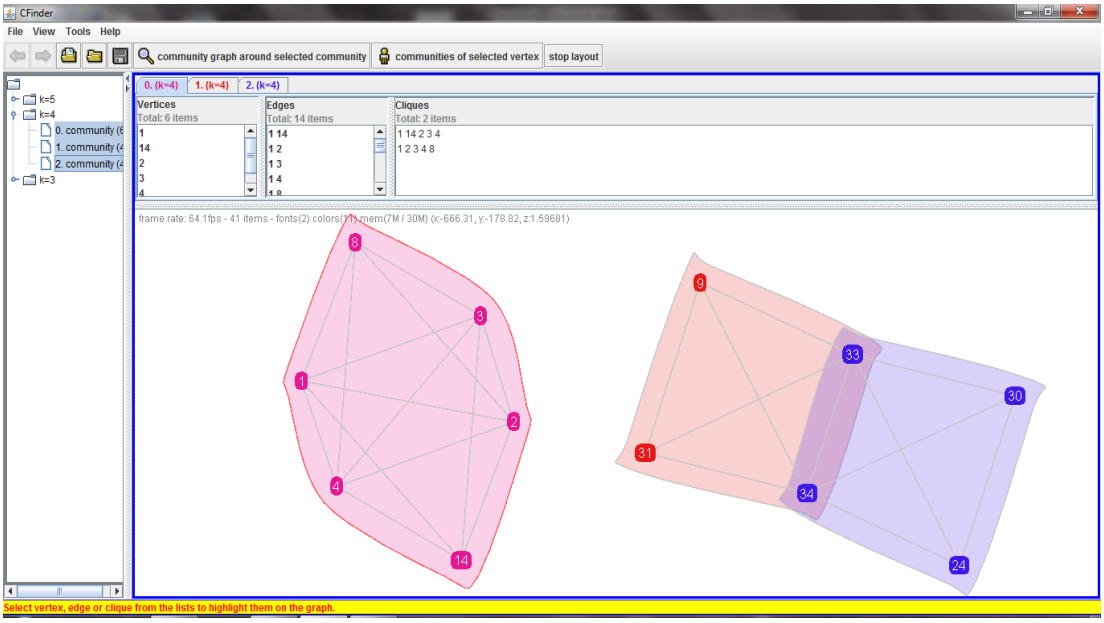


Figure14. extraction des cliques du réseau club de Zachary avec CFinder

- Les différents cliques du réseau :

Id-clique	**nœuds**
1	1 2 3 4 8 14
2	9 31 33 34
3	24 30 33 34

Tableau02 : Les cliques du réseau Zachary avant la fusion

-Nous remarquons qu'il y a une intersection entre le clique k2 et k3 :

$K_2 \cap K_3 = \{33, 34\}$ alors les deux cliques seront fusionnées :

$K_2 \cup K_3 = \{9, 24, 30, 31, 33, 34\}$

Id-clique	**nœuds**
1	1 2 3 4 8 14
2	9 24 30 31 33 34

Tableau03 : Les cliques du réseau Zachary après la fusion

-L'étape d'identification des nœuds isolés

-Les nœuds isolés sont des nœuds qui appartiennent au base de donnée et n'appartenant pas aux cliques 1 et 2.

-Ni est l'ensemble des nœuds isolés, dans ce cas :

Ni= {5,6,7,10,11 ,12,13,15,16,17,18,19,20,21,22,23,25 ,26,27,28,29,32}

-L'étape de détection des voisins pour chaque nœud isolé :

-l'ensemble des nœuds voisins du nœud isolé n_i ; représente t des nœuds qui ont une relation avec le nœud isolé n_i.

-L'étape d'affectation des nœuds isolés

- Pour chaque nœud isolé ; si le nœud n_i a des liens uniquement avec des nœuds appartenant au même groupe, il sera assigné au groupe auquel il a un ou plusieurs liens.

-Dans le cas où le nœud n_i a des liens avec différents groupes ; on calcule le max intersections avec tous les voisins de n_i.

- S'il existe plusieurs cliques qui possèdent le max intersections on calcule le max degré des nœuds voisins de n_i, n_i sera affecté à la communauté qui contient le max degré ; dans le cas contraire le nœud isolé n_i sera affecté à la communauté de max intersection.

Affichage des résultats

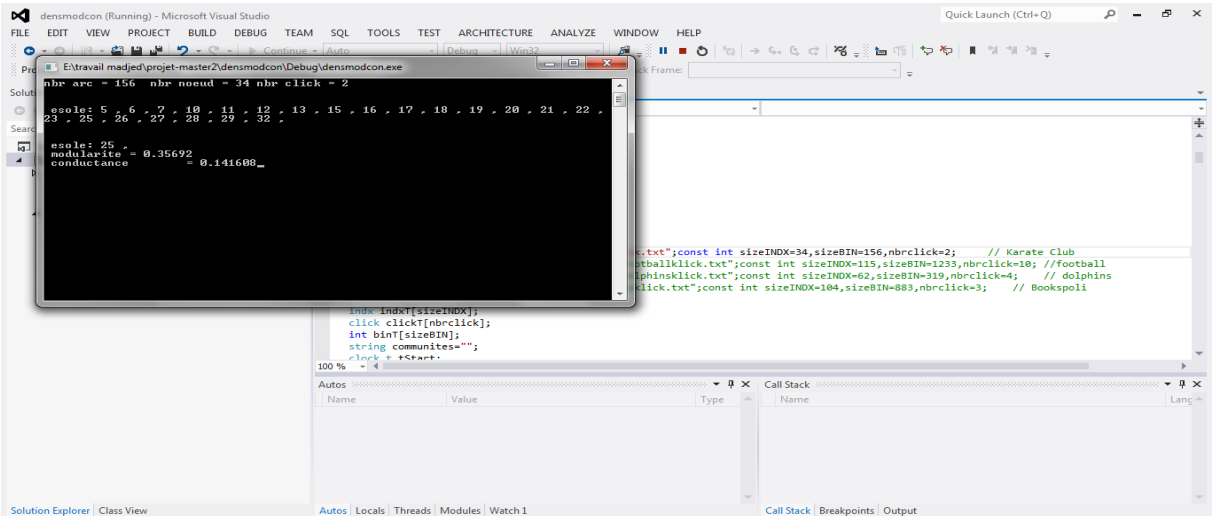


Figure15. Résultat d'exécution d'algorithme sur le réseau club de Zachary

La modularité	**0.3569**
La conductance	**0.1416**

Tableau04 : Résultat d'exécution d'algorithme sur le réseau Zachary

-Visualisation du réseau club de karaté :

-Après toutes les étapes de l'algorithme de détection de communauté proposé, le réseau est visualisé avec l'outil GEPHI 0.9.1 comme l'illustre la figure16

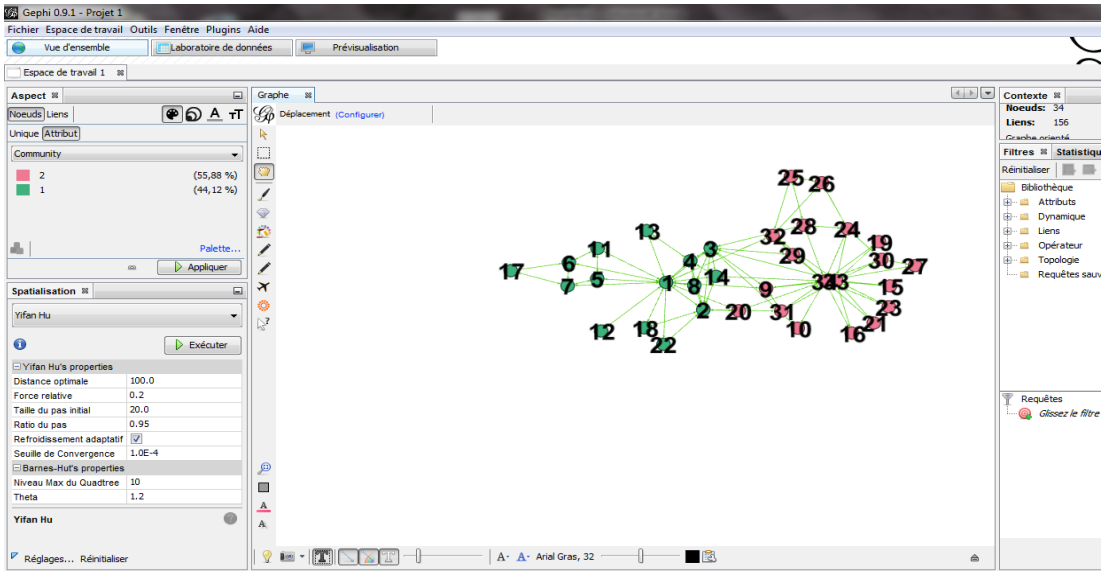


Figure16. Visualisation du réseau club de karaté avec Gephi

B. Les dauphins de lusseau

- Ce réseau contient 62 nœuds et 159 liens.

- La base de données représentant les liens entre les différents nœuds est représentée dans l'annexe.

-L'étape d'extraction de cliques :

- Diviser le réseau en cliques en utilisant le logiciel CFinder :

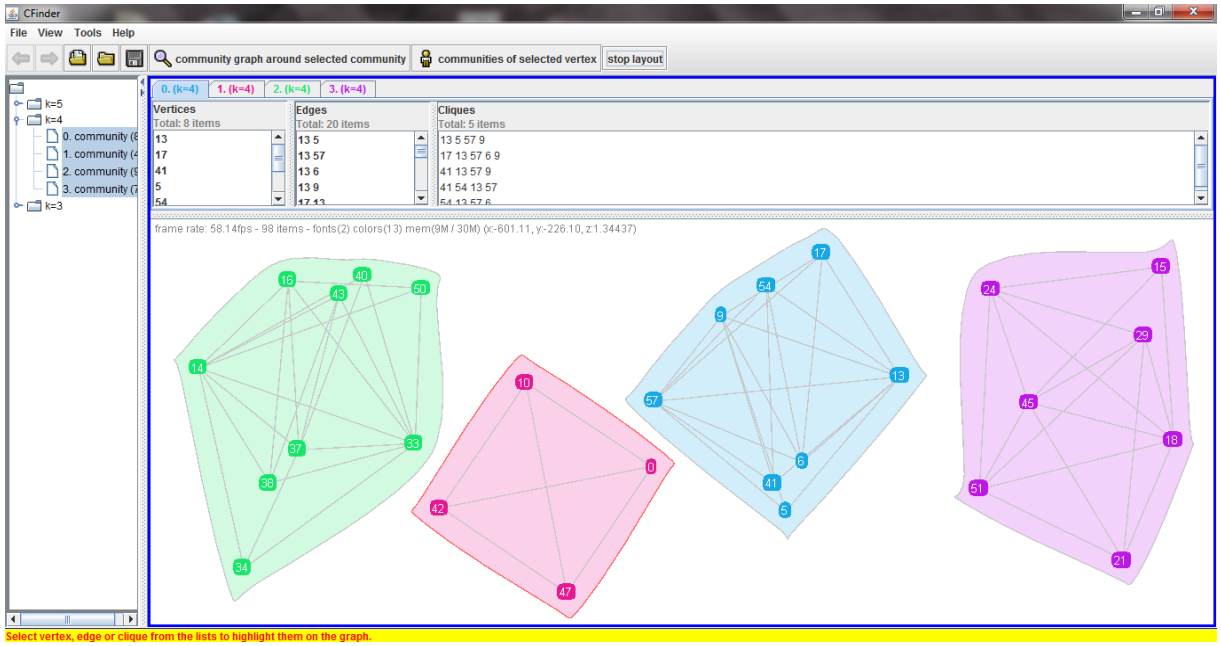


Figure17. Extraction des cliques du réseau Les dauphins de lusseau avec CFinder

-Différentes cliques de réseau :

Id-clique	nœuds
1	17 41 54 13 5 57 6 9
2	0 10 42 47
3	14 40 33 16 34 37 38 43 50
4	15 29 51 24 18 45 21

Tableau05 : Les cliques du réseau Dauphin de lusseau

-L'étape d'identification des nœuds isolés

-Ni est l'ensemble des nœuds isolés, dans ce cas :

Ni={1,2,3,4,7,8,11,12,19,20,22,23,25,26,27,28,30,31,32,35,36,39,44,46,48,49,52,53,55,56,58,59,60,61}

-L'étape suivante consiste à détecter l'ensemble des voisins pour chaque nœud isolé puis l'affectation de ces derniers au différentes cliques selon l'algorithme présenté.

-Affichage des résultats

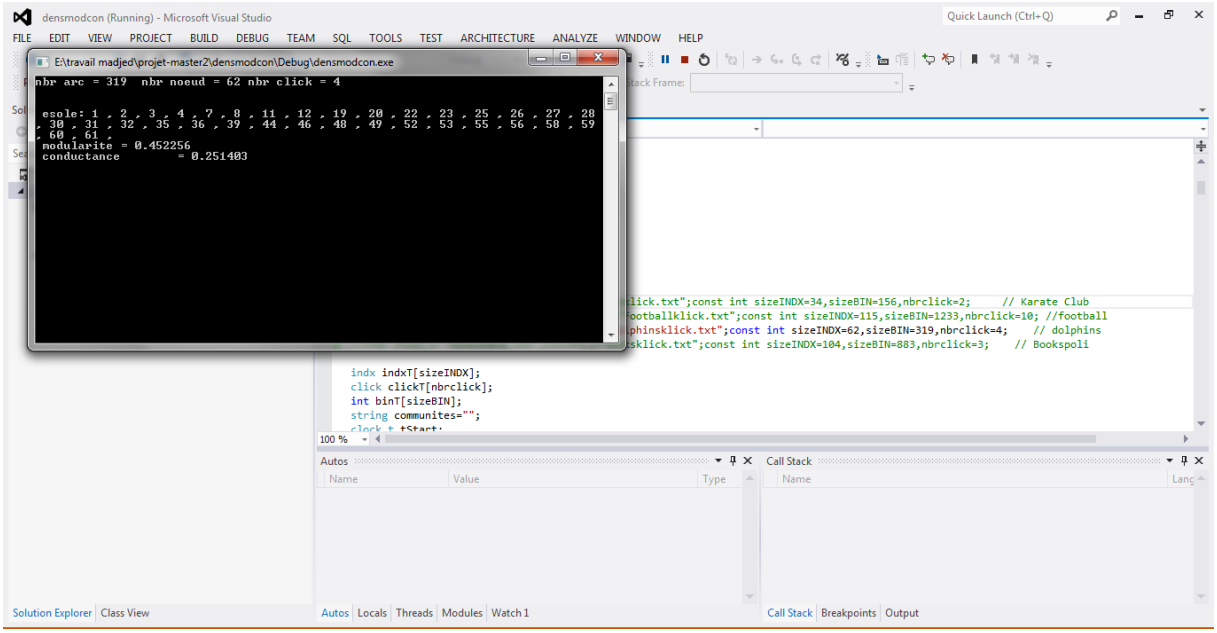


Figure18. Résultat d'exécution d'algorithme sur les dauphins du lusseau

La modularité	**0.4522**
La conductance	**0.2514**

Tableau06 : Résultat d'exécution d'algorithme sur les dauphins du lusseau

Visualisation du réseau Dauphin du lusseau :

-Après toutes les étapes de l'algorithme de détection de communauté proposé, le réseau est visualisé avec l'outil GEPHI 0.9.1 comme la montre la figure19.

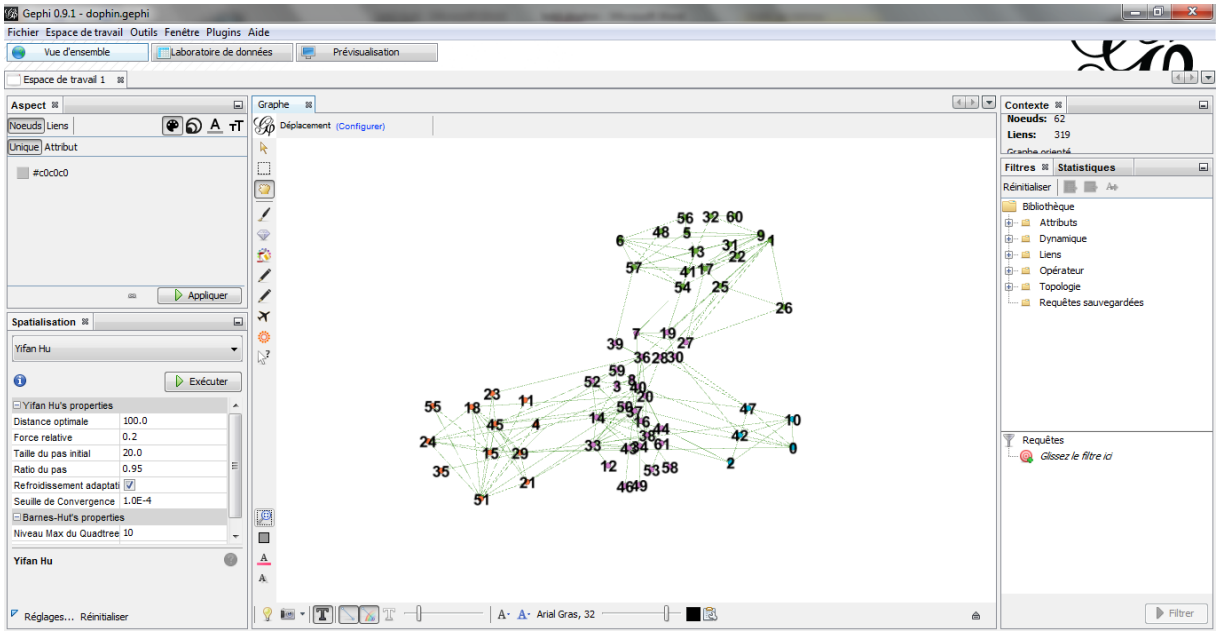


Figure19. Visualisation du réseau Dauphin du lusseau avec Gephi

C. le réseau Football Américain

-Ce réseau contient 115 nœuds et 613 liens.

- La base de données représentant les liens entre les différents nœuds est représentée dans l'annexe.

-L'étape d'extraction de cliques :

- Diviser le réseau en cliques en utilisant le logiciel CFinder :

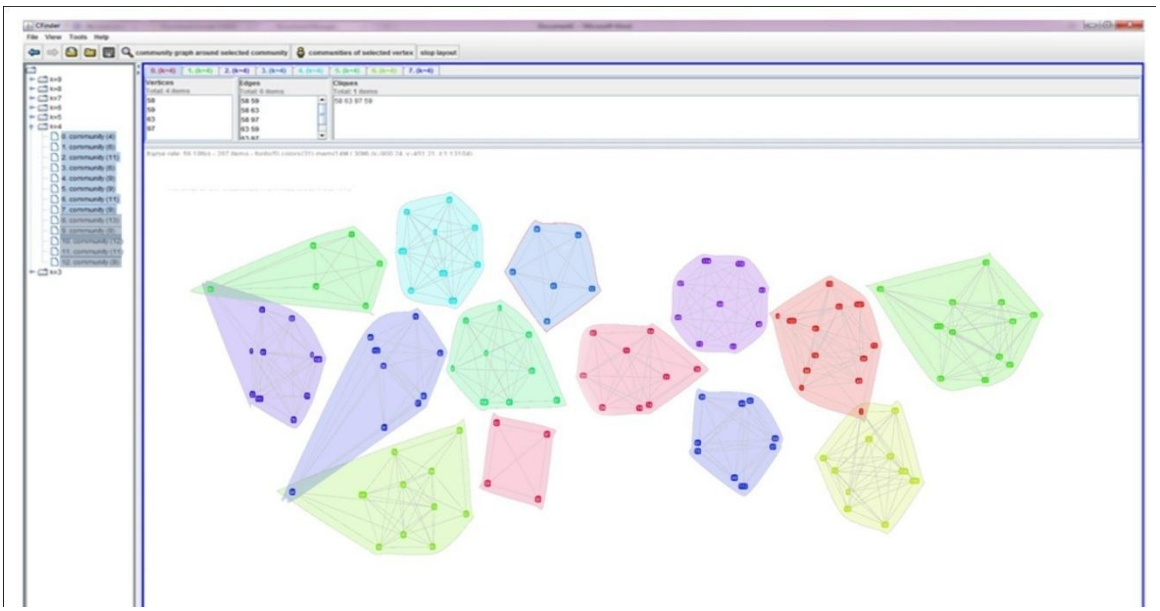


Figure20. Extraction des cliques du le réseau Football Américain avec CFinder

-Différentes cliques du réseau :

Id-clique	nœuds
1	58 63 97 59
2	90 11 24 28 50 69
3	51 108 111 21 22 68 7 77 78 50 8
4	12 14 26 38 43 85
5	1 33 103 105 109 25 37 45 89
6	0 104 16 23 4 41 9 93 67
7	35 101 82 19 29 30 55 79 94 80 44
8	57 92 75 112 44 66 91 48 86
9	10 102 107 5 72 74 81 84 98 2 3 40 52
10	99 14 18 34 38 54 71 31 61
11	65 27 70 87 62 56 96 113 17 20 76 95
12	60 100 13 15 2 32 39 47 6 64 106
13	114 67 88 110 46 49 53 73 83

Tableau07 : les cliques du réseau Football Américain avant la fusion

-Nous remarquons qu'il y a une intersection entre les cliques K_2 *et* K_3

$K_2 \cap K_3$= {50} alors les deux cliques seront fusionnées :

$K_2 \cup K_3$= {7,8,11,21,22,24,28,50,51,69,68,77,78,90,108,111}

-Nous remarquons aussi qu'il y a une intersection entre les cliques K_7 *et* K_8

$K_7 \cap K_8$= {44} alors les deux cliques seront fusionnées :

$K_7 \cup K_8$= {19,29,30,35,44,48,55,57,66,75,79,80,82,86,91,92,94,101,112}

-La troisième intersection est entre les 02 cliques K_9 et K_{12}

$K_9 \cap K_{12}$= {2} alors les deux cliques seront fusionnées :

$K_9 \cup K_{12}$ = {2 ,3 ,5 ,6 ,10 ,13 ,15 ,32 ,39 ,40 ,47 ,52 ,60 ,64 ,72 ,74 ,81, 84 ,98 ,100 ,102 ,106 ,107}

-Les cliques après la fusion :

Id-clique	nœuds
1	58 59 63 97
2	7 8 11 21 22 24 28 50 51 69 68 77 78 90 108 111
3	12 14 26 38 43 85
4	1 25 33 37 45 89 103 105 109
5	0 4 9 16 23 41 67 93 104
6	19 29 30 35 44 48 55 57 66 75 79 80 82 86 91 92 94 101 112
7	2 3 5 6 10 13 15 32 39 40 47 52 60 64 72 74 81 84 98 100 102 106 107
8	14 18 31 34 38 54 61 71 99
9	17 20 27 56 62 65 70 76 87 95 96 113
10	46 49 53 67 73 83 88 110 114

Tableau08 : les cliques du réseau Football Américain après la fusion

-L'étape d'identification des nœuds isolés

-Les nœuds isolés sont des nœuds qui appartiennent à la base de donnée et n'appartiennent pas à aucune clique

-Ni est l'ensemble des nœuds isolés, dans ce cas :

Ni = {36, 42}

L'étape suivante consiste à détecter l'ensemble des voisins pour chaque nœud isolé puis l'affectation de ces derniers aux différentes cliques selon l'algorithme présenté.

-Affichage des résultats

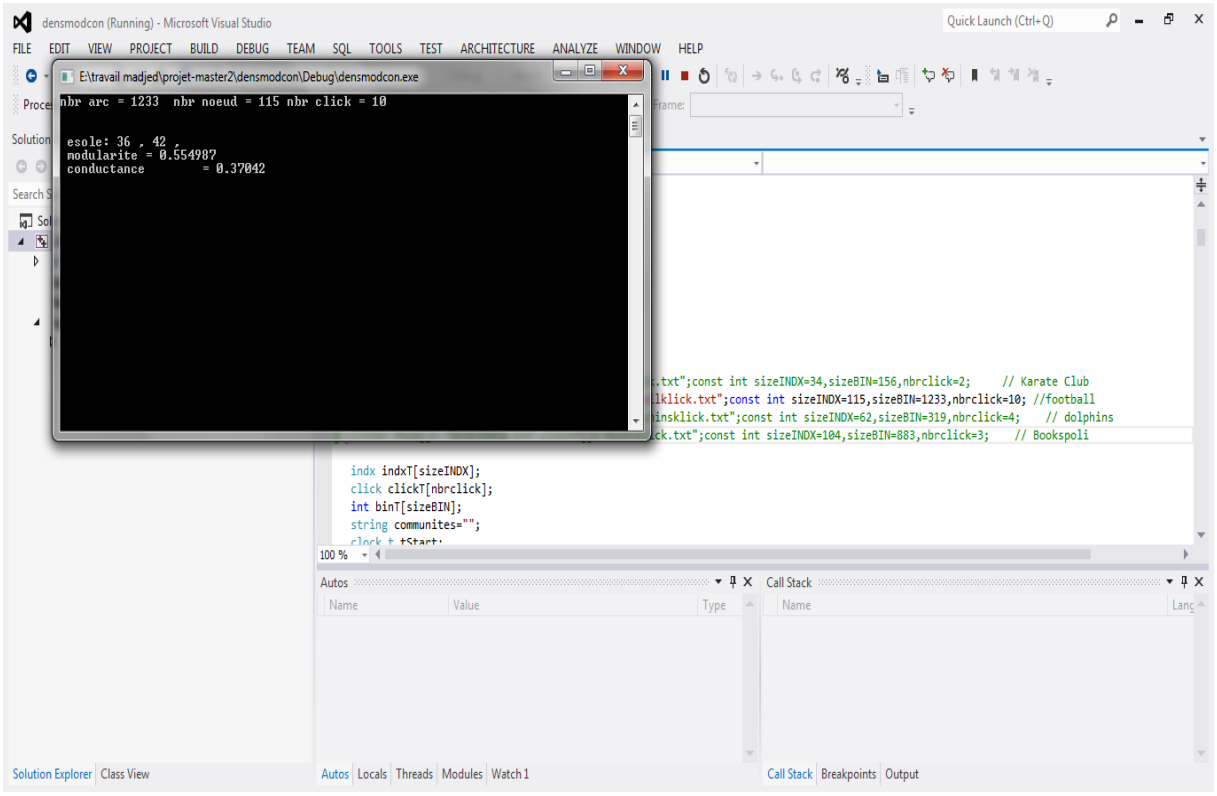


Figure21. Résultat d'exécution de l'algorithme sur le réseau Football Américain avec Gephi

La modularité	**0.5549**
La conductance	**0.3704**

Tableau09 :Résultat d'exécution de l'algorithme sur le réseau Football Américain

Visualisation du réseau Football américain :

-Après toutes les étapes de l'algorithme de détection de communauté proposé, le réseau est visualisé avec l'outil GEPHI 0.9.1 comme l'illustre la figure 22.

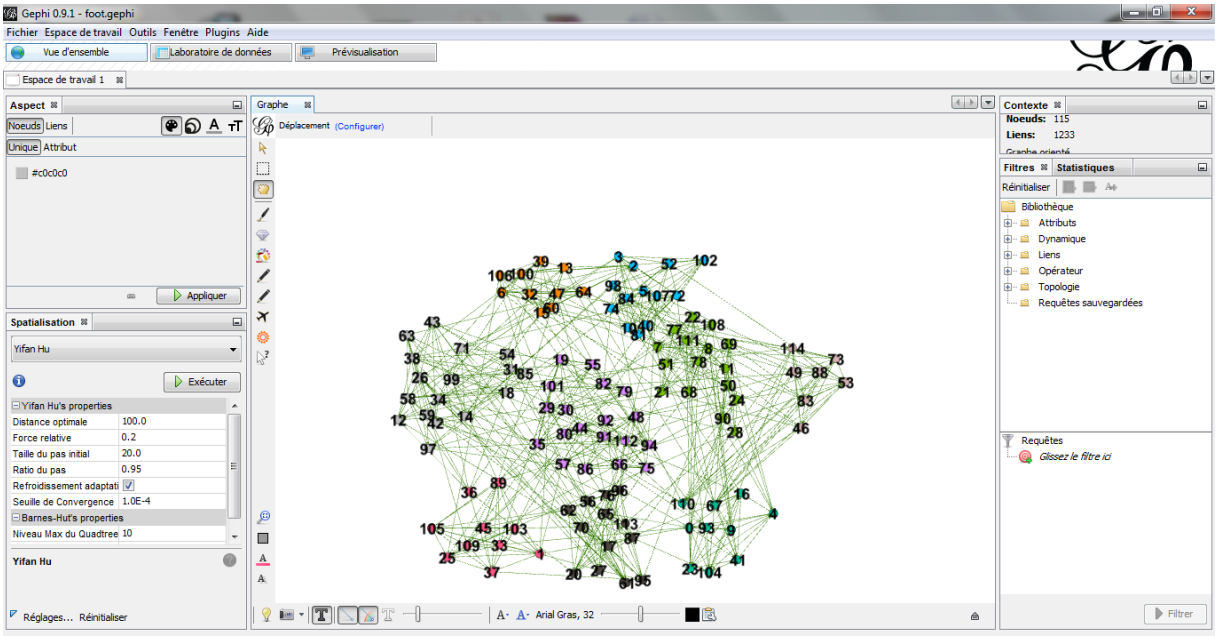
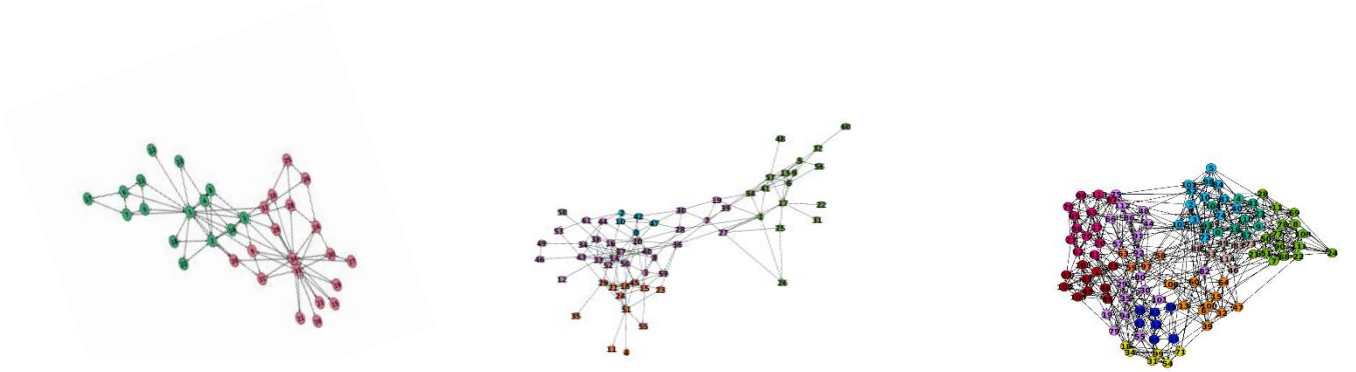


Figure22. Visualisation du réseau Football américain avec Gephi



(a) Communautés détectées -club Zachary-

(b) Communautés détectées -Dauphin de lusseau-

(c) Communautés détectées -Football Américain -

Figure23. Communautés détectées sur différents réseaux

	Club Zachary	Dauphin de lusseau	Football Américain
Modularité	0.3569	0.4522	0.5549
La conductance	0.1416	0.2514	0.3704

Tableau10 : Résultat sur les différents réseaux

V. Discussion

L'approche proposée est comparée avec quelques algorithmes présentés dans le deuxième chapitre, soit Infomap, Walktrap, FCD et Label Propagation. Ces algorithmes ont été choisis pour le fait qu'ils représentent d'une part les différents types d'algorithmes (agglomératif, divisif, à base du modèle et heuristique) de détection de communautés, et d'autre part, ils sont également très connus et considérés comme les plus performants pour une comparaison.

Comme l'illustre La Figure 24 qui représente les différentes valeurs de la conductance trouvée par différents algorithmes et la figure 25 qui représente les différentes valeurs de la modularité trouvée par différents algorithmes, les résultats des expérimentations que nous avons réalisées ont montré que les résultats sont satisfaisants et que l'algorithme est efficace.

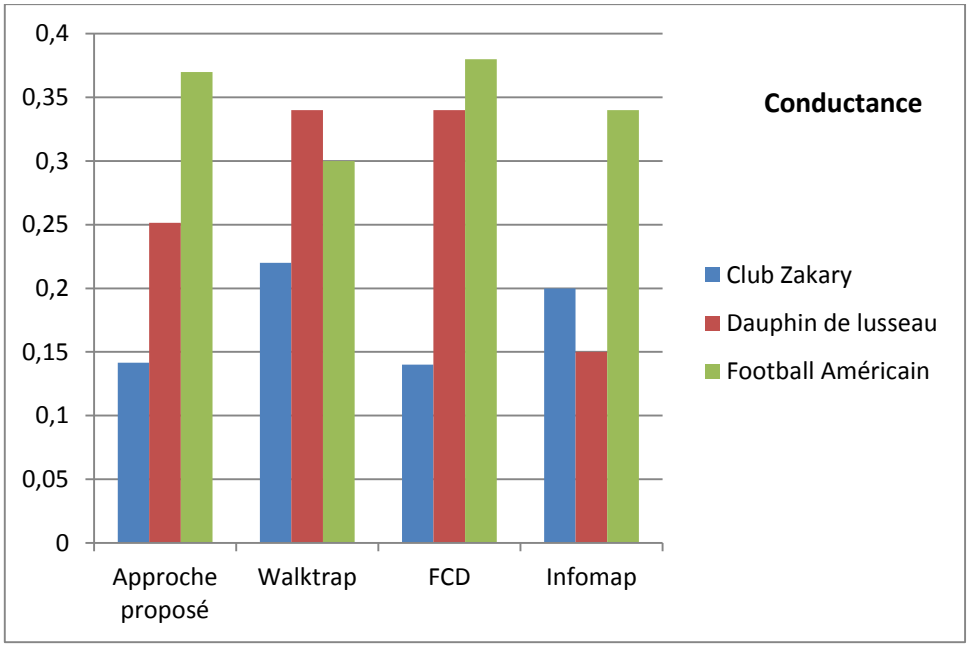


Figure 24. Valeurs de la conductance sur différents réseaux par différents Algorithmes

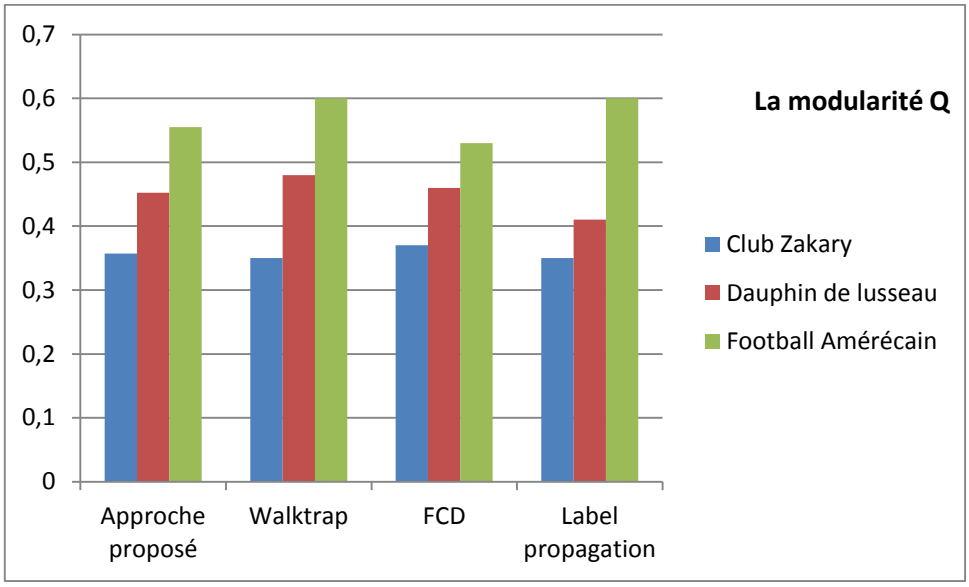


Figure 25. Valeurs de la modularité sur différents réseaux par différents Algorithmes

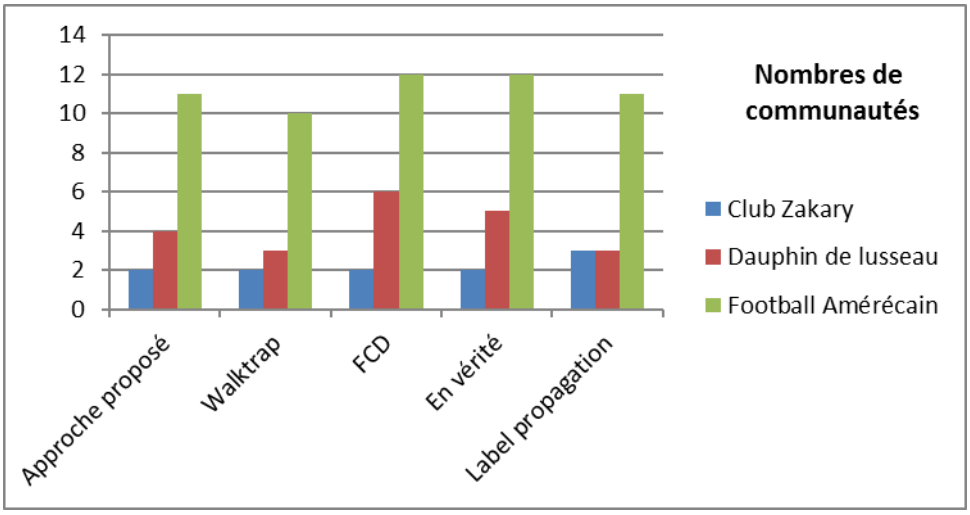


Figure 26. Nombre de communautés trouvées pour différents réseaux par différents Algorithmes

Conductance	Club Zachary	Dauphin de lusseau	Football Américain
Approche proposé	0.1416	0.2514	0.37
Walktrap	0.22	0.34	0.30
FCD	0.14	0.34	0.38
Infomap	0.20	0.15	0.34

Tableau11 : Valeurs de la conductance sur différents réseaux par différents Algorithmes

La modularité Q	Club Zachary	Dauphin de lusseau	Football Américain
Approche proposée	0.3569	0.4522	0.5549
Walktrap	0.35	0.48	0.60
FCD	0.37	0.46	0.53
Label propagation	0.35	0.41	0.60

Tableau12 : Valeurs de la modularité sur différents réseaux par différents Algorithmes

Nombres de communautés	Club Zachary	Dauphin de lusseau	Football américain
Approche proposée	02	04	11
En vérité	02	05	12
Walktrap	02	03	10
FCD	02	06	12
Label propagation	03	03	11

Tableau13 : Nombre de communautés trouvées pour différents réseaux par différents Algorithmes

Analyse des résultats

- La comparaison des résultats calculés par notre algorithme avec les autres sur chacun des réseaux choisi montre les différences suivantes :

- La modularité et la conductance

1- le réseau club de Zachary

- En comparant les valeurs calculées avec l'algorithme Walktrap nous remarquons que la conductance et la modularité calculées par notre approche sont mieux que celles de walktrap (tableau 11,12).

- La comparaison entre l'algorithme FCD et l'algorithme proposé concernant la valeur de la modularité montre que les résultats sont très proches par contre la valeur de la conductance trouvée par notre algorithme est mieux que FCD. (Tableau11,12)

- La comparaison de la valeur de la conductance calculée par l'algorithme Infomap, et celle trouvée par l'algorithme proposé démontre que le résultat de ce dernier est mieux. (Tableau11).

- La modularité comparée entre l'algorithme Label propagation et notre algorithme montre l'efficacité de ce dernier selon les résultats trouvés.

2-Le réseau Dauphin de lusseau

- En comparant les valeurs calculées avec l'algorithme Walktrap et FCD nous remarquons que la conductance trouvée par notre algorithme est mieux par contre la modularité calculée par Walktrap est mieux et pour l'algorithme FCD les deux résultats sont presque égaux. (Tableau 11,12).

- La comparaison de la valeur de la conductance calculée par l'algorithme Infomap, et celle trouvée par l'algorithme proposé prouve que le résultat de ce dernier est mieux. (Tableau11).

- La modularité comparé entre l'algorithme Label propagation et notre algorithme montre l'efficacité de ce dernier selon les résultats trouvés.

3-Le réseau Football Américain

Nous remarquons que les valeurs de la conductance et la modularité montrées par les algorithme Walktrap, Infomap, Label propagation sont mieux que celles calculés par l'algorithme proposé. (Tableau11,12).

Par contre les valeurs calculées par notre algorithme soit la conductance ou la modularité sont mieux que les résultats de l'algorithme FCD. (Tableau11,12).

-Le nombre de communautés trouvées

Les résultats des expérimentations que nous avons réalisées sur les différents réseaux choisis ont montré, que les communautés trouvées par notre approche étaient comparables à celles trouvées par d'autres algorithmes sur des réseaux sociaux, et surtout très proches du nombre de communautés originales (Tableau 13), (figure25).

-Avantages et limites de l'approche proposée

L'approche proposée présente certains avantages et limites :

A- Les avantages

La stabilité

Les communautés trouvées par notre approche restent stables dans plusieurs exécutions sur le même réseau.

La rapidité de l'algorithme

Pendant l'affectation des différents nœuds isolés ; l'algorithme parcourt la base de donnée une seule fois.

La capacité de recouvrement

La phase de détection des cliques aide à la détection s des différents chevauchements entes les nœuds, ce qui est hyper important dans les réseaux sociaux.

B- Les limites

-Cette Approche est basée d'un côté sur le principe de maximiser la modularité, qui est l'un des paramètres les plus classiques de détection de communautés, alors elle hérite la faiblesse de difficulté de détecter les communautés de petites taille dans les grands réseaux.

Conclusion

Dans ce chapitre, nous avons exposé notre approche de détection de communauté dans les réseaux sociaux basée sur l'analyse formelle de concept, et les différents résultats des expérimentations que nous avons effectuées sur des réseaux réels.

L'expérimentation a montré la capacité et la fiabilité de notre algorithme à détecter les Communautés.

Les résultats sont acceptables et montrent que les valeurs de modularité sont comparables avec plusieurs algorithmes, tandis que les valeurs de la conductance étaient convaincantes.

Conclusion et perspectives

Le travail réalisé dans le cadre de ce mémoire s'inscrit dans le domaine de la détection de communautés dans les réseaux sociaux. Nous nous sommes particulièrement intéressés à la détection de communautés en se basant sur l'AFC et en utilisant aussi la notion des graphes.
Notre approche est caractérisée par l'utilisation de deux paramètres : la maximisation de modularité qui est une fonction de qualité qui permet de trouver de meilleures structures de communautés de graphe et le calcul de la conductance qui compare le nombre de liens internes et externes de la communauté et plus la conductance d'un ensemble de nœuds est faible, plus cet ensemble est censé être une bonne communauté

La contribution de ce mémoire réside dans la proposition d'un algorithme qui prend en considération tous les acteurs du réseau social, et affecte tous les nœuds isolés aux différentes communautés en s'appuyant sur l'extraction des cliques du graphe, et le calcul des intersections avec les différentes communautés pour l'affectation des nœuds isolés. Cet algorithme appartient à la catégorie des approches statiques sans chevauchement et avec chevauchement.

Etant donné que le nombre de communautés et leurs tailles ne sont pas connus à priori ; l'algorithme a été expérimenté sur un ensemble de données de réseaux issu du monde réel.
Les expérimentations effectuées ont montré l'efficacité et la capacité de cet algorithme à découvrir les communautés et à affecter tous les acteurs du réseau social.

Pour valider notre travail, nous avons établi des comparaisons des déférentes valeurs de modularité et de la conductance ainsi que le nombre de communautés avec des algorithmes qui sont également connus et considérés parmi les plus performants, et les résultats étaient acceptables.

Nous avons en particulier remarqué que les valeurs de la conductance et le nombre de communautés trouvés par l'algorithme proposé sont tous les deux convaincants voire satisfaisants.

Les travaux et les résultats présentés dans ce travail offrent naturellement de nombreuses perspectives. La première découle du fait que nous n'avons travaillé que sur deux paramètres ; la maximisation de la modularité et le calcul de la conductance quoiqu'il existe d'autres paramètres à savoir la densité interne, le coefficient de clustring et le temps d'exécution.

La prise en compte des réseaux dynamiques est aussi une perspective importante de la détection de communautés.

Bibliographie

[1] G. E. (, Analyse des réseaux sociaux et web sémantique: un état de l'art, France, 2008.

[2] R. Bachelet, «Réseaux sociaux,» Lille-France, 2014.

[3] P. torloting, «Enjeux et perspectives des reseaux sociaux,» Paris, 20006.

[4] M. a. M.e.J.Newman, «Community structure in social and biological networks,» *Proceedings of the National Academy of Sciences,* 2002.

[5] Price, «Networks of scienti,» 1965.

[6] R. Kanawati, «Détection de communautés dans les grands graphes d'interactions (multiplexes) : état de l'art,» p. 3, 2014.

[7] Newman, «Fast algorithm for detecting community structure in networks,» 2003.

[8] R. Kanawati, «Détection de communautés dans les grands graphes d'interactions (multiplexes) : état de l'art,» p. 4, 2014.

[9] P. Diane-Gabrielle Tremblay, «LES COMMUNAUTÉSVIRTUELLES DE PRATICIENS VERS DE NOUVEAUX MODES D'APPRENTISSAGE ET DE CRÉATION DE CONNAISSANCES,» 2003.

[10] J. S. &. P. J. Carrington, «Social Network Analysis: A Handbook. Sage Publications Ltd,» 2011.

[11] K. R. R. P. R. S. a. T. A. Kleinberg J, «The Web as a Graph: Measurements, Models, and Methods Computing and Combinatorics,» pp. 1627: 1-17, 1999.

[12] L. S. a. G. C. Flake G.W, «Efficient identification of Web communities,» pp. 150-160.

[13] C. C. C. F. L. V. a. P. D. Radicchi F, «Defining and identifying communities in networks,» pp. 2658-2663, 2004.

[14] G. M. Newman M.E.J, «Finding and evaluating community structure in networks,» *Physical Review E,,* p. 69:026113, 2004.

[15] M. B. R. C. C. Y. a. G. E. Mancoridis S, «Using Automatic Clustering to Produce High-Level System Organizations of Source Code,» 1998.

[16] G. M. Newman M.E.J, «Finding and evaluating community structure in networks,» *Physical Review E,* 2004.

[17] M. B. R. C. C. Y. a. G. E. Mancoridis S, «Using Automatic Clustering to Produce High-Level System Organizations of Source Code,» pp. 45-53, 1998.

[18] M. E. J. Newman, «Fast algorithm for detecting community structure in networks,» *Physical Review E,,* p. 69(6) :066133, 2004.

[19] M. E. J. N. a. C. M. Aaron Clauset, «Finding community structure in very large networks,» *Physical Review E,* p. 70(6) :066111, 2004.

[20] G. J. L. R. L. E. BLONDEL V. D., «Fast unfolding of communities in large networks,» *Journal of Statistical Mechanics Theory and Experiment,* 2008.

[21] M. G. a. M. E. J. Newman, «Community structure in social and biological networks,» *PNAS,* 2002.

[22] U. Brandes, «A faster algorithm for betweenness centrality,» *Journal of Mathematical,* 2001.

[23] Mohamed Talbi, «Une nouvelle approche de détection de communautés dans les réseaux sociaux,» OUTAOUAIS, 2013.

[24] P. a. L. Pons, «Computing communities in large networks using random walks,» p. 10(2) :191–218, 2006.

[25] B. C. T. ROSVALL M., «Maps of random walks on complex networks reveal community structure,» *Proceedings of the National Academy of Sciences vol. 105, n°04,* p. p. 1118–1123, 2008.

[26] Y. S. a. S. Bressan, «Fast Community Detection-TRC/13,» pp. 1-2, 2013.

[27] «Uncovering the overlapping community structure of complex networks in nature and society. Nature,» p. 435(7043) :814{818, 2005.

[28] M. K. K. K. a. J. S. J.M. Kumpula, «Sequential algorithm for fast clique percolation,» *Physical Review E,,* p. 78(2) :026109, 2008.

[29] X. C. K. C. a. M. H. H. Shen, «Detect overlapping and hierarchical community structure in networks.,» *Physica A : Statistical Mechanics and its Applications,* p. 388(8) :1706{1712, 2009.

[30] U. N. A. R. a. K. S. Raghavan, «Near linear time algorithm to detect community structures in large-scale networks,» *Physical Review E,* p. 76 :1–12, 2007.

[31] M. Bajec, «Robust network community detection using balanced propagation. Nature,» 2011.

[32] A. B. a. T. G. Palla, «Vicsek. Quantifying social group evolution. Nature,» 2007.

[33] R.cazabet, «R. CAZABDétection des communautés dynamiques dans des réseaux temporels,,» Toulouse-France, 2013.

[34] L. C. Freeman, «Cliques, Galois lattices, and the structure of human social groups,» p. 159–172.

[35] L. Falzon, «Determining groups from the clique structure in large social networks,» p. 159–172.

[36] F. B. R. M. O. B. Sid Ali Selmane, «Identification des communautés au sein des réseaux sociaux par l'Analyse Formelle de Concept».

[37] S. A. Selmane, «Identification des communautés au sein des réseaux sociaux par l'Analyse Formelle de Concept,» pp. 25-26.

[38] M. SEIFI, «Coeurs stables de communautés dans les graphes de terrain,» 2012.

[39] S. F. e. M. Barthélemy, «Resolution limit in community detection,» *Proceedings of the National Academy of Sciences, vol. 104, no. 1,,* p. 36, 2007.

[40] M. DANISCH, «Mesures de proximité appliquées à la détection de communautés dans les grands graphes de terrain,» Paris.

[41] Y. S. a. S. Bressan, «Fast Community Detection,» p. 8, May 2013.

[42] W. W. Zachary, «An information flow model for conflict and fission in small groups,» *Journal of anthropological research,* pp. 452-473, 1977.

[43] K. S. O. J. B. P. H. E. E. S. a. S. M. D. David Lusseau, «The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations,» p. 54(4) :396–405, 2003.

[44] M. G. a. M. E. Newman, «Community structure in social and biological networks.,» *Proceedings of the National Academy of Sciences,,* p. 99(12) :7821–7826, 2002.

[45] R. cazabet, «Détéction des communautés dynamiques dans des réseaux temporels,» France, 2013.

[46] O. K. B. K. a. B. S. J. Hopcroft, «Tracking evolving communities in large linked networks,» *Proceedings of the national academy of sciences of the United States of America,* pp. 5249-5253, 2004.

[47] B. W. a. N. D. Y. Wang, «Community evolution of social network : feature algorithm and model,» p. Science And Technology, 2008.

[48] A. B. a. T. V. G. Palla, «Quantifying social group evolution,» p. 446(7136) :664{667, 2007.

[49] K. W. Y. J. W. H. a. N. S. Z. Chen, «Detecting and tracking community dynamics in evolutionary networks.,» *In Data Mining Workshops (ICDMW),* p. pages 318{327, 2010.

Annexe

I. Les bases de données des réseaux utilisés dans l'expérimentation

A-Base de donnée du réseau dauphin de lusseau

nœud	relations
0	10,14,15,40,42,47
1	17,19,26,27,28,36,41,54
2	10,42,4461
3	8,14,59
4	51
5	9,13,56,57,
6	9,13,17,54,56,57
7	19,27,30,40,54
8	3,20,28,37,45,59
9	5,6,13,17,32,41,57
10	0,2,29,42,47
11	51
12	33
13	5,6,9,17,32,41,54,57
14	0,3,16,24,33,34,37,38,40,43,50,52
15	0,18,24,40,45,55,59
16	14,20,33,37,38,50
17	1,6,9,13,22,25,27,31,57
18	15,20,21,24,29,45,51
19	1,7,30,34
20	20,16,18,28,36,38,44,47,50
21	18,29,33,37,45,51
22	17
23	36,45
24	14,15,18,29,45,51
25	17,26,27
26	1,25,27
27	1,7,17,25,26
28	1,8,20,30,47
29	10,18,21,24,35,43,45,51,52
30	7,19,28,42,47
31	17
32	9,13,60
33	12,14,16,21,34,37,38,40,43,50
34	14,33,37,44,49
35	29
36	1,20,23,37,39,40,59
37	8,14,16,21,33,34,36,40,43,45,61
38	14,16,20,33,43,44,52,58
39	36,57

40	0,7,14,15,33,36,37,52
41	1,9,13,54,57
42	0,2,10,30,47,50
43	14,29,33,37,38,46,53
44	2,20,34,38
45	8,15,18,21,23,24,29,37,50,51,59
46	43,49
47	0,10,20,28,30,42
48	57
49	34,46
50	14,16,20,33,42,45,51
51	4,11,18,21,23,24,29,45,50,55
52	14,29,38,40
53	43,61,53
54	1,6,7,13,19,41,57
55	15,51
56	5,6
57	5,6,13,17,39,41,48,54
58	38
59	3,8,15,36,45
60	32
61	2,37,53

B-Base de donnée du réseau football Américain

nœud	relations
0	1,4,9,16,23,33,35,41,65,90,93,104
1	0,25,27,33,37,45,57,89,101,103,105,109,
2	3,6,13,14,15,47,60,64,72,74,100,106
3	2,5,11,26,40,52,58,72,74,81,84,102
4	0,5,9,16,23,28,41,69,93,104,108
5	3,4,10,11,52,74,81,84,90,97,98,107
6	2,7,32,39,47,55,58,60,64,85,100,106
7	6,8,21,22,40,68,73,77,78,82,108,111
8	7,9,21,22,41,51,68,77,78,90,111
9	0,4,8,16,22,23,41,64,93,104,108
10	5,11,60,72,74,81,84,98,102,107
11	3,5,10,24,28,50,69,90,97,104
12	13,14,17,18,26,34,36,38,43,85
13	2,12,15,32,39,45,60,64,100,106,110
14	2,12,15,26,38,43,54,71,85,99,
15	2,13,14,32,39,47,60,68,92,100,106,114
16	0,4,9,17,23,38,41,67,81,93,104
17	12,16,20,27,58,62,65,87,95,96,113
18	12,19,31,34,36,38,42,54,61,71,99
19	18,29,30,33,35,36,44,55,79,94,101

20	17,21,36,62,65,70,75,76,87,96,113
21	7,8,20,22,32,46,51,68,77,108,111
22	7,8,9,21,23,47,51,68,77,78,108
23	0,4,9,16,22,41,78,90,93,104,111
24	11,25,28,50,66,69,84,87,90,110
25	1,24,33,37,45,53,89,103,105,106,109
26	3,12,14,27,34,38,42,43,61,85
27	1,17,26,56,62,63,65,70,76,95,96
28	4,11,24,38,50,69,78,90,113
29	19,30,35,42,55,79,80,82,9194,101
30	19,29,35,44,50,55,79,82,94,101,109
31	18,32,34,43,54,55,61,71,79,85,99
32	6,13,15,21,31,39,47,49,64,100,106
33	0,1,19,25,37,45,89,103,105,109
34	12,18,26,31,35,42,54,61,71,94,99
35	0,19,29,30,34,44,55,79,92,94,101
36	12,18,19,20,37,43,58,59
37	1,25,33,36,45,80,89,95,103,105,109
38	12,14,16,18,26,28,39,43,54,71,85
39	6,13,15,32,38,47,54,60,82 ,100,106
40	3,7,41,51,52,72,74,81,98,102,107
41	0,4,8,9,16,23,40,67,93,104
42	18,26,29,34,43,57,63
43	12,14,26,31,36,38,42,61,70,79,85
44	19,30,35,45,48,57,66,75,86,91,112
45	1,13,25,33,37,44,62,89,103,105,109
46	21,47,49,53,67,73,83,88,110,111,114
47	2,6,15,22,32,39,46,60,61,64,100
48	44,49,53,57,66,75,86,91,92,96,98
49	48,32,46,53,67,73,83,88,110,114
50	11,24,28,30,51,68,69,78,90
51	8,21,22,40,50,68,77,778,101,108,111
52	3,5,40,53,72,74,84,98,102,112
53	25,46,48,49,52,67,73,83,86,88,110,114
54	14,18,31,34,38,39,55,61,71,99
55	6,19,29,30,31,35,54,79,89,94,101
56	27,57,62,65,70,76,87,95,96,106
57	1,42,44,48,56,75,86,91,92,112
58	3,6,17,36,59,63,88,97,101,114
59	36,58,60,63,66,76,97,11
60	2,6,10,13,15,39,47,59,64,71,106
61	18,26,31,34,49,47,54,62,71,92,99
62	17,20,27,45,56,61,70,76,87,95,105
63	27,42,58,59,64,65,97,109,112
64	2,6,9,13,32,47,60,63,100,106,111
65	0,17,20,27,56,63,66,70,87,96,113
66	24,44,48,59,65,75,76,86,91,92,112
67	16,41,46,49,53,68,73,83,88,104,114
68	7,8,15,21,22,50,51,67,78,108,111

69	4,11,24,28,50,70,83,88,90,91,95
70	20,27,43,56,62,65,69,76,95,103,113
71	14,18,31,34,38,54,60,61,72,99
72	2,3,10,40,52,71,74,81,102,104,107
73	7,46,49,53,67,74,77,83,88,110,114
74	2,3,5,10,40,52,72,73,82,84,102
75	20,44,48,57,66,76,86,92,107,112
76	20,27,56,59,62,66,70,75,95,96,113
77	7,8,21,22,51,73,78,82,98,108,111
78	7,8,22,23,28,50,51,68,108,111
79	19,29,30,31,35,55,80,94,101,109
80	29,37,79,82,85,86,91,93,94,105,110
81	3,5,10,16,40,72,82,83,84,98,107
82	7,29,30,39,74,77,80,81,93,94,100
83	46,49,53,67,69,73,84,88,110,114
84	3,5,10,24,49,52,74,81,83,98,107
85	6,12,14,26,31,38,43,80,99
86	44,48,53,57,66,75,80,87,91,92,97
87	17,20,24,56,62,65,86,95,96,104,113
88	46,49,53,58,67,69,73,83,89,107,110,114
89	1,25,33,37,45,55,88,99,103,105,109
90	0,5,8,11,23,24,28,50,69
91	29,44,48,57,66,69,80,86,92,93,112
92	15,35,48,57,61,66,75,91,106,112
93	6,4,9,16,23,41,80,82,91,104
94	19,29,30,34,35,55,79,80,101
95	17,27,37,56,62,69,70,76,87,113
96	17,20,27,48,56,65,76,87,112,113
97	5,11,58,59,63,86,98,112
98	5,10,40,48,52,77,81,84,97,102,107
99	14,18,31,34,54,61,71,89,100
100	2,6,15,32,39,47,64,82,99,102
101	1,19,29,30,35,51,55,58,79,94
102	3,10,40,52,72,74,98,100,103,107
103	1,25,33,37,45,70,89,105,109
104	0,4,9,11,16,23,41,67,72,87,93,104,114
105	1,25,33,37,45,62,80,89,103,109
106	2,6,13,15,25,32,39,56,60,64,92
107	5,10,40,72,75,81,84,88,98,102
108	4,7,9,21,22,51,68,77,78,111
109	1,25,30,33,37,45,63,79,89,103,105
110	13,24,46,49,53,67,73,80,83,88,114
111	7,8,21,23,46,51,64,68,77,78,108
112	44,52,57,63,66,75,91,92,96,97
113	17,20,28,59,65,70,76,87,95,96
114	15,46,49,53,58,67,73,83,88,104