

*République Algérienne Démocratique et Populaire*  
*Ministère de l'enseignement supérieur et de la recherche scientifique*

**UNIVERSITE Dr. TAHAR MOULAY SAIDA**

**FACULTE : TECHNOLOGIE**

**DEPARTEMENT : INFORMATIQUE**



**MEMOIRE DE MASTER**

**OPTION : Réseaux Informatique Des Systèmes Répartis**

**Thème**

***La détection des communautés dans les réseaux sociaux par les méta heuristiques***

**Présenté par :**

**BENDOUMA CHAIMAA ZOUBEIDA**

**GOTT AICHA DJIHAD**

**Encadré par :**

**DR : BENYAHIA KADDA.**

**Co-Encadreur :**

**DR : GUENDOUZ MOHAMED.**

**Promotion : juin 2017/2018**

## DÉDICACE GOTT AICHA DJIHAD ET BENDOUMA CHAIMAA ZOUIBIDA

Je dédie humblement ce manuscrit à :

A celle qui s'est toujours dévouée et sacrifiée pour nous ,

celle qui nous a aidée du mieux qu'elle pouvait pour réussir; celle qui nous a accompagnée tout au long de ce parcours Périlleux, celle qui a toujours été là dans notre moment de détresse, **nos très chères Mères.**

A celui qui nous a toujours encouragée et soutenue moralement, **nos très chers pères.**

A celles qui nous ont toujours aidée, écoutée, soutenue et encouragée tout au long de notre parcours; celles qui ont toujours été présentes pour nous, nos très chères sœurs : **IMENE, AMINA, AMARIA.**

A nos très chers frères qui nous ont énormément aidés et à qui nous témoigne nous affection et nous profonde reconnaissance **AHMED, IBRAHIM**

A celles qui nous ont toujours aimées **nos familles.**

A nos très chères amies : **KOUIDRI CHAIMAA, MOKRFI SAMIA**

**Et tout nos amis** qui nous avons toujours encouragée et soutenue depuis le début de notre travail.

# REMERCIEMENTS

**N**ous exprimons notre gratitude et nos sincères appréciations : A nos Encadreurs Mr **BENYAHIA KADDA** et Mr **GUENDOZ MOHAMED** pour leur disponibilité, leurs critiques et leurs conseils toujours opportuns ainsi que pour leur grande gentillesse et de nous avoir donné toute cette confiance et d'avoir cru à nos compétences. Aux membres de jury pour leur disponibilité et acceptation d'examiner et de rapporter notre travail. A toutes les personnes qui nous ont aidées durant notre processus de recherche et d'écriture de ce mémoire.

## Résumé

Les réseaux sociaux à grande échelle ont émergé rapidement ces dernières années. Les réseaux sociaux sont devenus des réseaux complexes. Le domaine de recherche a attiré beaucoup d'intérêt scientifique. La communauté est une structure importante dans les réseaux sociaux.

Au cours de la dernière décennie, plusieurs algorithmes ont été proposés pour résoudre le problème de la détection de la communauté dans les réseaux sociaux. Beaucoup d'entre eux sont basés sur un essaim d'intelligence et algorithmes évolutifs.

La plupart de ces algorithmes utilisent la densité de modularité comme fonction de fitness pour maximiser. Cependant, ces algorithmes tentent de trouver la meilleure solution sans prendre en considération la structure du réseau. Dans cette mémoire nous proposons un algorithme génétique (GA) pour détecter les communautés. Nous appliquons notre algorithme sur quatre données classiques et on compare les résultats de notre approche avec les autres algorithmes.

Notre algorithme de détection de communauté s'est avéré efficace dans les expériences.

**Mots clés :** : Détection de communautaire , Algorithmes évolutifs , Algorithme génétique.

## Abstract

Social networks on a large scale have emerged rapidly in recent years. Social networks have become complex networks. The field of research has attracted a great deal of scientific interest. The community is an important structure in social networks. Over the last decade, several algorithms have been proposed to solve the problem of community detection in social networks. Many of them are based on a swarm of intelligence and evolutionary algorithms. Most of these algorithms use modularity density as a fitness function to maximize. However, these algorithms try to find a better solution without considering the structure of the network. In this memory we propose a genetic algorithm (GA) to detect

---

communities. We apply our algorithm on four classical data and we compare the results of our approach with the other algorithms. Our community detection algorithm proved effective in the experiments.

**Keywords :** Community detection , Evolutionary algorithms , genetic algorithm

## ملخص

الشبكات الاجتماعية على نطاق واسع ظهرت بسرعة في السنوات الأخيرة. أصبحت الشبكات الاجتماعية شبكات معقدة. وقد اجتذب مجال البحث قدرا كبيرا من الاهتمام العلمي. المجتمع هو هيكل مهم في الشبكات الاجتماعية. على مدى العقد الماضي، تم اقتراح العديد من الخوارزميات لحل مشكلة الكشف المجتمعي في الشبكات الاجتماعية. ويستند الكثير منها إلى سرب من الذكاء والخوارزميات التطورية. تستخدم معظم هذه الخوارزميات كثافة نمطية كوظيفة لياقة لتعظيم. ومع ذلك، تحاول هذه الخوارزميات للكشف عن (GA) إيجاد حل أفضل دون النظر إلى بنية الشبكة. في هذه المذكرة نقتراح خوارزمية جينية المجتمعات. نطبق خوارزمتنا على ثلاثة بيانات كلاسيكية ونقارن نتائج خوارزمتنا مع الخوارزميات الأخرى. أثبتت خوارزمية اكتشاف المجتمع لدينا فعاليتها في التجارب.

**الكلمات المفتاحية:** اكتشاف المجتمع • خوارزميات تطورية • خوارزمية جينية

# INTRODUCTION GÉNÉRALE

Avec le développement de la téléphones intelligents, de plus en plus de gens se connectent à leurs réseaux sociaux réseaux à travers leurs téléphones intelligents et partager du texte et d'informations multimédia avec leurs amis en ligne.

Les réseaux sociaux à grande échelle sont devenus des réseaux complexes, comme les réseaux métaboliques, réseaux épidémiologiques, réseaux écologiques, etc. Ils sont modélisés sous forme de graphe. Un réseau social se compose d'un ensemble de nœuds avec des bords reliant les nœuds. Les nœuds représentent les objets dans les réseaux sociaux, tels que les gens, les journaux, les marchandises, etc. Les communautés dans les réseaux du monde réel fournissent des informations très importantes pour détecter les relations structurelles et fonctionnelles entre les objets et dans la détermination de tout réseau d'identification. Par conséquent, la détection ou la découverte de communautés est un sujet très important dans l'analyse des réseaux complexes.

Une des recherches importantes dans l'analyse des réseaux sociaux est la détection de communautés (CD) pour sa large utilisation dans l'exploration de données. etc.

Il y a plusieurs algorithmes et techniques essayent pour résoudre le problème de détection de la communauté dans les réseaux sociaux comme , essaim de poisson artificiel, colonie artificielle d'abeilles et les algorithmes génétiques.

Dans ce, mémoire, nous avons traité principalement le problème de détection de communautés dans les réseaux sociaux que nous avons complété par une proposition d'une optimisation d'algorithmes génétiques pour la détection des communautés.

Le plan de ce mémoire s'organise autour d'une progression de la réflexion, partant de la problématique telle qu'elle est traitée dans l'état de l'art vers les

---

solutions opérationnelles qui ont été conçues, en passant par la description de notre contribution, les expérimentations et les résultats.

La première partie présente une vue d'ensemble des différents champs de recherche concernés par notre problématique, commençant par un état de l'art sur l'analyse des réseaux sociaux, ses concepts et ses représentations (cf. Chapitre 1).

Une présentation de la détection des communautés fait l'objet du deuxième chapitre (cf. Chapitre 2). Nous avons passé à une présentation guidée par la perspective de la détection de communauté, à savoir les différents travaux portant sur la détection (cf. Chapitre 2). Cette première étude nous permet d'exposer les notions essentielles pour la compréhension du domaine de recherche et de mettre l'accent sur les points clefs liées à notre problématique. Puis, nous abordons la partie contribution.

Dans la deuxième partie, nous présentons notre contribution de détection de communautés. Dans le chapitre 3, l'approche de détection de communautés en se basant sur les algorithmes génétiques. Le chapitre qui suit (cf. chapitre 4), est consacré à l'expérimentation de notre approche ainsi que l'évaluation des résultats. Afin de tester notre approche, nous avons implémenté un outil qui regroupe les différentes phases de notre approche. Nous présentons l'expérimentation effectuée pour évaluer l'approche proposée.

# Table des matières

<b>1</b>	<b>Etat de l'art sur l'analyse des réseaux sociaux</b>	<b>12</b>
1.1	Introduction	14
1.2	Définition	15
1.2.1	Réseau social	15
1.2.2	Réseaux sociaux en ligne	15
1.3	Origines des réseaux sociaux	17
1.4	Historique	18
1.4.1	Le classement des réseaux sociaux en Europe en 2016	19
1.4.2	Les Types des RS	20
1.4.2.1	RS de publication	21
1.4.2.2	RS de partage	21
1.4.2.3	RS de partage video	21
1.4.2.4	RS de partage de photos	21
1.4.2.5	RS de partage audio	22
1.4.2.6	RS de partage de présentation	22
1.4.2.7	RS de partage de discussion	22
1.5	Représentation d'un réseau social	22
1.5.1	Les acteurs sociaux	23
1.5.2	Les liens sociaux	23
1.5.3	Représentation graphique	24
1.5.4	Représentation matricielle	24
1.5.4.1	Théorème	25
1.5.5	Graphe orienté	25
1.5.6	Fonctionnement d'un Réseau Social	27
1.5.7	Domaines d'application réseaux sociaux en ligne	28
1.6	Analyse des réseaux sociaux en ligne	29
1.6.1	Origine de l'analyse des réseaux sociaux	29
1.7	La Représentation des RS dans l'ARS	30

## TABLE DES MATIÈRES

---

1.7.1	Notion de calcul en ARS	30
1.7.2	Centralité d'intermédiarité	30
1.7.3	Centralité de proximité	31
1.7.4	Centralisation de pouvoir	31
1.7.5	Prestige	31
1.7.6	Les branches de l'ARS	31
1.7.7	La détection des structures communautaires	31
1.7.8	L'étude de l'évolution dynamique des réseaux sociaux	32
1.8	Conclusion	32
<b>2</b>	<b>Détection de communauté</b>	<b>34</b>
2.1	Introduction	35
2.2	Communautés	35
2.2.1	Définitions	35
2.2.1.1	Les bases	35
2.2.1.2	Définitions locales	36
2.2.1.3	Définitions globales	37
2.2.2	Définitions basées sur la similarité des sommets	37
2.2.3	Les méthode pour calculer la similarité	37
2.2.4	Complexité informatique	39
2.2.5	Objective de détection de communautés dans les graphes	39
2.3	Les approches	40
2.3.1	Méthodes Traditionnelles	40
2.3.1.1	Partition de graphe	40
2.3.1.2	Partitions	41
2.3.1.3	Classification hiérarchique	43
2.3.1.4	Clustering partiel	44
2.3.1.5	Cluster spectral	44
2.3.1.6	Algorithmes proposés ( metaheuristiques )	45
2.3.1.7	Algorithme de chauve souris amélioré basé sur l'algorithme évolutionnaire différentiel	48
2.3.1.8	Algorithme Big Bang-Big Crunch modifié	51
2.4	Modularité	53
2.5	évaluation	55
2.6	Conclusion	56
<b>3</b>	<b>La detection de communautes par les algorithmes génétiques</b>	<b>58</b>
3.1	La detection de communautes par les algorithmes génétiques	61
3.1.1	Introduction	61
3.1.2	Définition	61
3.1.3	Calcul naturel	62

## TABLE DES MATIÈRES

---

3.1.4	Biologiquement Calcul Inspiré . . . . .	63
3.1.5	Biologie Motivé informatiquement . . . . .	63
3.1.6	Calcul avec biologie . . . . .	63
3.1.7	Informatique intelligente . . . . .	64
3.1.8	Processus de création d'un algorithme inspiré da La Nature . . . . .	64
3.1.9	Application de l'algorithme (méta heuristiques) . . . . .	65
3.2	Optimisation Combinatoire En mathématique . . . . .	66
3.2.1	Minimum Local . . . . .	66
3.2.2	Minimum Global . . . . .	66
3.2.3	Voisinage . . . . .	67
3.3	Classification des Méta heuristiques . . . . .	67
3.4	Notre approche Algorithme Génétique . . . . .	68
3.4.1	Introduction . . . . .	68
3.4.2	Définition . . . . .	69
3.4.3	Terminologie . . . . .	69
3.5	Fonctionnement générale . . . . .	70
3.5.1	Définition 01 (Gène) . . . . .	70
3.5.2	Définition 02 (Chromosome) . . . . .	71
3.5.3	Définition 03 (Individu) . . . . .	71
3.5.4	Définition 04 (population) . . . . .	71
3.5.5	Définition 05 (Génération) . . . . .	71
3.5.6	Dentition 06 (fonction de fitness ou fonction d'évaluation) . . . . .	71
3.5.7	Fonctionnement général . . . . .	72
3.6	Algorithme de base AG . . . . .	73
3.7	Principe de fonctionnement . . . . .	74
3.7.1	Codage . . . . .	74
3.7.1.1	Codage binaire . . . . .	74
3.7.1.2	Codage réel . . . . .	75
3.7.1.3	Codage alphanotique . . . . .	75
3.7.2	Population Initiale . . . . .	76
3.7.3	Adaptation . . . . .	76
3.8	Les opérateurs d'un algorithme génétique . . . . .	77
3.8.1	la sélection . . . . .	77
3.8.1.1	la sélection par roulette . . . . .	77
3.8.1.2	Sélection par rang . . . . .	78
3.8.1.3	Sélection par Elitisme . . . . .	79
3.8.2	Le croisement . . . . .	80
3.8.2.1	Le croisement binaire . . . . .	81
3.8.2.2	Croisement uniforme . . . . .	82
3.8.2.3	Croisement réel . . . . .	83
3.8.3	La Mutation . . . . .	83

## TABLE DES MATIÈRES

---

3.8.3.1	mutation binaire . . . . .	83
3.8.3.2	mutation réelle . . . . .	84
3.9	Critère d'arrêt . . . . .	84
3.10	Les algorithmes génétiques pour la détection de communauté dans les réseaux sociaux . . . . .	84
3.10.1	Codage . . . . .	84
3.10.2	Population Initiale . . . . .	85
3.10.3	Adaptation (fonction de fitness) . . . . .	85
3.10.4	la sélection . . . . .	85
3.10.5	Le croisement . . . . .	86
3.10.5.1	Le croisement utilisé c'est le croisement en un point . . . . .	86
3.10.6	La Mutation . . . . .	86
3.10.7	Critère d'arrêt . . . . .	87
3.11	Conclusion . . . . .	87
<b>4</b>	<b>Implémentation et Résultats</b>	<b>88</b>
4.1	Introduction . . . . .	90
4.2	Environnement de développement . . . . .	90
4.2.1	JAVA . . . . .	90
4.2.2	eclipse . . . . .	91
4.3	Présentation des réseaux de teste . . . . .	92
4.3.1	Club de karaté de Zachary . . . . .	92
4.3.2	Les dauphins de Lusseau . . . . .	92
4.3.3	Les livres politiques . . . . .	93
4.3.4	Football américain . . . . .	94
4.3.5	Présentation de l'application . . . . .	94
4.4	Discussion des résultats . . . . .	99
4.4.1	Evaluation . . . . .	99
4.5	Comparaison des résultats et discussion . . . . .	100
4.5.1	Club de Zachary . . . . .	100
4.5.2	dauphins de Lusseau . . . . .	101
4.5.3	Livres politiques . . . . .	101
4.5.4	Football américain . . . . .	101
4.6	Conclusion . . . . .	102
	<b>Bibliographie</b>	<b>105</b>

## TABLE DES MATIÈRES

---

<b>bibliographie</b>	<b>105</b>
<b>Table des figures</b>	<b>108</b>

# Chapitre **1**

## ÉTAT DE L'ART SUR L'ANALYSE DES RÉSEAUX SOCIAUX

### Sommaire

---

<b>1.1</b>	<b>Introduction</b>	14
<b>1.2</b>	<b>Définition</b>	15
1.2.1	Réseau social	15
1.2.2	Réseaux sociaux en ligne	15
<b>1.3</b>	<b>Origines des réseaux sociaux</b>	17
<b>1.4</b>	<b>Historique</b>	18
1.4.1	Le classement des réseaux sociaux en Europe en 2016	19
1.4.2	Les Types des RS	20
<b>1.5</b>	<b>Représentation d'un réseau social</b>	22
1.5.1	Les acteurs sociaux	23
1.5.2	Les liens sociaux	23
1.5.3	Représentation graphique	24
1.5.4	Représentation matricielle	24
1.5.5	Graphe orienté	25
1.5.6	Fonctionnement d'un Réseau Social	27
1.5.7	Domaines d'application réseaux sociaux en ligne	28
<b>1.6</b>	<b>Analyse des réseaux sociaux en ligne</b>	29
1.6.1	Origine de l'analyse des réseaux sociaux	29

---

<b>1.7 La Représentation des RS dans l'ARS</b> . . . . .	<b>30</b>
1.7.1 Notion de calcul en ARS . . . . .	30
1.7.2 Centralité d'intermédiarité . . . . .	30
1.7.3 Centralité de proximité . . . . .	31
1.7.4 Centralisation de pouvoir . . . . .	31
1.7.5 Prestige . . . . .	31
1.7.6 Les branches de l'ARS . . . . .	31
1.7.7 La détection des structures communautaires . . . . .	31
1.7.8 L'étude de l'évolution dynamique des réseaux sociaux . . . . .	32
<b>1.8 Conclusion</b> . . . . .	<b>32</b>

---

### 1.1. INTRODUCTION

**D**e nos jours, les réseaux sociaux ont pris une importante part dans la vie quotidienne des millions d'utilisateurs abonnés à ces derniers, en offrant une large gamme d'intérêts et de pratiques. La principale caractéristique des réseaux sociaux est la capacité de réunir différents individus autour d'un point de vue commun ou de les rassembler autour de croyances collectives.

Les réseaux sociaux permettent aux gens de se connecter avec des étrangers ou de maintenir leurs relations existantes.

Les gens se connectent aux réseaux sociaux pour des intérêts partagés tels que la profession, les études, le business ou pour des intérêts fondés sur la langue commune, la nationalité ou des partages raciaux, religieux, et sexuels.

Les réseaux sociaux intègrent plusieurs outils et de nouvelles technologies de communication comme la connectivité mobile ou le partage des photos et de vidéos ainsi que les techniques de géo localisation.

Le grand nombre d'abonnés et leur activités sur les RS a engendré une grande masse d'informations.

En effet l'exploitation de cette masse d'informations a conduit à la naissance de différentes études sur les RS , c'est pour ça il est important de bien cerner l'usage de ces réseaux sur le Web Pour cela il est nécessaire de définir les réseaux sociaux, leur typologie et de savoir comment ils peuvent se former et se construire.

Afin de mieux cerner le phénomène et d'en comprendre la portée, un cas particulier MySpace est le réseau social le plus utilisé au monde.

Dans ce chapitre, nous présentons l'état de l'art des réseaux sociaux, en couvrant la plupart de ses aspects, tout en nous consacrant plus particulièrement aux avancées les plus récentes du domaine, nous donnerons quelques définitions et la terminologie. Nous aborderons les réseaux sociaux et leurs représentations.

### 1.2. DÉFINITION

#### 1.2.1. RÉSEAU SOCIAL

Aujourd'hui, un réseau social est défini comme une structure définie par des relations entre des individus . Concrètement, c'est l'ensemble des individus avec qui une personne est en contact. Il s'agit également des liens entre des personnes : les habitants d'un quartier, une famille...

Un réseau social représente un système d'entités en interaction.

On le modélisera comme un graphe  $G=( D .S); A /$  où  $S$  est un ensemble d'entités (les sommets ou nœuds du graphe) et  $A$  est l'ensemble des arcs (ou connexions) représentant les interactions entre ces sommets .[1]

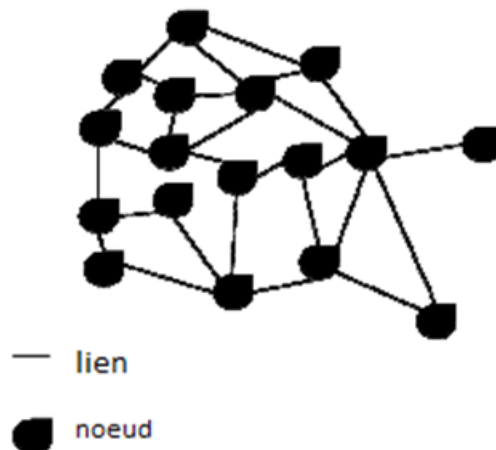


FIGURE 1.1 – Réseau social montrant les nœuds et les liens.[2]

#### 1.2.2. RÉSEAUX SOCIAUX EN LIGNE

La définition des réseaux sociaux en ligne englobe la mise en réseau pour les affaires, le plaisir, et tous les points entre les deux.

## 1.2. DÉFINITION

---

Les réseaux eux-mêmes ont des objectifs différents, et leurs homologues en ligne fonctionnent de diverses manières. Autrement dit, un réseau social permet aux gens de communiquer avec des amis et des connaissances à la fois anciennes et nouvelles.[15]

Définit un site de réseautage social comme un site Web qui fournit une communauté virtuelle pour les personnes intéressées par un sujet particulier ou simplement pour traîner ensemble .

Bien que ce soit une description précise, une définition plus détaillée du réseautage social en ligne engloberait également toutes les façons dont les gens peuvent se connecter.

Les réseaux en ligne tels que Facebook, Twitter, MySpace et LinkedIn offrent aux utilisateurs une variété de moyens pour augmenter leurs réseaux, partager des notes et divers types de médias, et se connecter à différents niveaux.

Certaines des façons dont les gens peuvent réseauter sur ces sites comprennent :

Avoir des amis ou des contacts actuels suggèrent d'autres personnes avec qui vous voudrez peut-être faire du réseautage Partage de photos, de fichiers audio et vidéo, d'œuvres écrites, de liens et d'autres médias

Afficher un curriculum vitae ou une expérience de travail pour la recherche d'emploi et le recrutement , promouvoir une marque ou un service auprès de ceux qui pourraient avoir besoin du produit ou du service.

### 1.3. ORIGINES DES RÉSEAUX SOCIAUX

---



FIGURE 1.2 – les différents types de RS avec des exemples de SRS réels.

### 1.3. ORIGINES DES RÉSEAUX SOCIAUX

La première personne à avoir représenté un réseau social est Jacob Levy Moreno au début des années 1930 [Moreno,1933]. [3]

Son objectif étant de visualiser graphiquement un réseau social, il a représenté les personnes par des points et une relation entre deux personnes par des flèches.

Au milieu du 20<sup>ème</sup> siècle, Cartwright et Harary sont les premiers à avoir appliqué la théorie des graphes dans l'analyse des réseaux sociaux.[4] 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.

### 1.4. HISTORIQUE

Depuis quelques années, les réseaux sociaux ou les sites de réseaux sociaux comme Facebook, Twitter, Google+ et LinkedIn ont fortement gagné en popularité sur Internet. Des millions de personnes se sont ainsi mises à créer leur profil public ou semi-public sur ces plateformes, à dresser la liste des utilisateurs avec lesquels elles sont en relation.

La notion de réseau social est née depuis que l'homme est constitué en société, en formant des groupes sociaux selon une croyance ou un thème fédérateur tel que la religion, la classe sociale, le niveau d'étude... etc.

Avec l'apparition d'internet cette notion a pris de nouvelles grandeurs, parmi les premiers réseaux sociaux apparus nous citons :

Le premier réseau social à avoir réuni toutes les catégories de base pour faire quelque chose de complet est Sivdegrees.com en 1997.

Cependant en 2000, le site dû fermer malgré des millions d'utilisateurs par manque d'argent.

Depuis Sivdegrees.com, la création de nouveaux réseaux sociaux ne cesse d'augmenter. De 1997 à 2001, beaucoup de plateformes comme communautés se sont créées comme par exemple :AsianAvenue (composé uniquement de la communauté asiatique),BlackPlanet (composé uniquement de la communauté noire),MiGente (composé uniquement de la communauté latine). Ces sites permettaient aux utilisateurs de créer des profils personnels ou professionnels mais aussi de faire des rencontres amoureuses entre communautés. Des réseaux sociaux réservés aux affaires professionnels .

En 2001, le site Ryze.com a vu le jour mais il ne connaît pas le succès,

C'est LinkedIn, créée en 2002 en Californie (Etats-Unis), qui devient un réseau professionnel très actif :En novembre 2015, le site comptait plus de 400 millions de membres issus de 170 secteurs différents d'activité dans plus de 200 pays et territoires.En France il y a environ 10 millions de membres actifs.C'est en 2004 que la création de Facebook a vu le jour par Marc Zuckerberg alors qu'il étudiait encore à Harvard.Pour la petite histoire, au lancement de Facebook, il fallait impérativement une adresse mail universitaire contenant Harvard.edu pour

## 1.4. HISTORIQUE

---

s'inscrire. Cela veut dire qu'à son lancement Facebook était un réseau social fermé.[11]

A partir de 2003, l'augmentation en volume de nouveaux réseaux sociaux a donné lieu à un terme anglais : YASN ( Yet Another Social Networking , soit encore un autre réseau social ).

En 2007 Facebook connaît une ascension phénoménale.

En France, c'est seulement lors de sa mise en ligne traduite en langue française le 10 mars 2008 que les inscriptions montèrent en flèches En juin 2009, Facebook comptait plus de 340 millions de visiteurs uniques selon la société de mesure d'audience ComScore.

### 1.4.1. LE CLASSEMENT DES RÉSEAUX SOCIAUX EN EUROPE EN 2016

Mesure en d'utilisateurs actifs par mois Facebook (1,79 milliards), YouTube (pas de résultats), WhatsApp (1 milliard), Facebook Messenger ( 1 milliard), Facebook Groups (1 milliard), Google Hangouts (1 milliard), Instagram (500 millions).[16]

## 1.4. HISTORIQUE

---

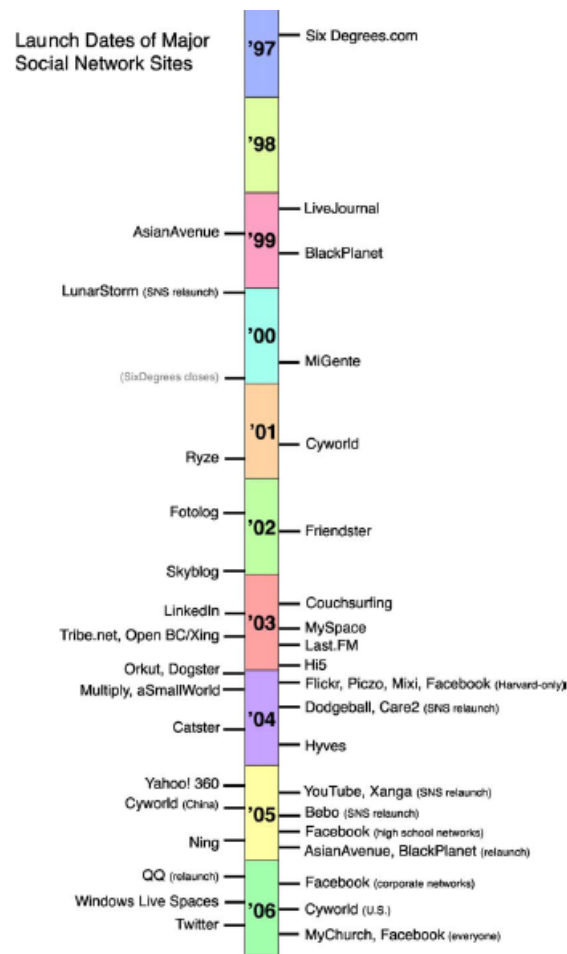


FIGURE 1.3 – Chronologie des dates des lancements des RS les plus importants.

### 1.4.2. LES TYPES DES RS

D'après la définition d'Esther Dyson des réseaux sociaux Les réseaux sociaux fournissent des outils qui facilitent le processus de mise en relation autour d'un centre d'intérêt commun et permettent la prise de contact en ligne qui amène une classification des réseaux sociaux selon différents critères :

Les services fournis aux acteurs sociaux.

D'autre part, si nous considérons un autre critère tel que le rôle, type de liens.

## 1.4. HISTORIQUE

---

### 1.4.2.1. RS DE PUBLICATION

Les RS de publication sont des RS de distribution d'informations fournies collectivement par un ensemble d'utilisateurs tels que les wikis ou individuellement dans le cas des blogs.

Exemple : Wikipedia.org, blogger.com , twitter.com.

### 1.4.2.2. RS DE PARTAGE

Ce type de RS permet à l'utilisateur de partager des documents avec les autres utilisateurs du RS.

En général, ils sont équipés par des moteurs de recherche pour rendre plus facile la fouille de documents ainsi qu'ils peuvent gérer la confidentialité du contenu (public ou privé).

On distingue quatre catégories selon les types des documents :

### 1.4.2.3. RS DE PARTAGE VIDEO

Cette catégorie de RS est la plus populaire parmi les autres, où l'utilisateur a la possibilité de visionner ou publier des vidéos, les commenter ou les télécharger si leurs contextes le permettent.[17]

Exemple : youtube.com, BlogTV.com, Buzdeo.com, GodTube.com, MetaCafe.com, Stickam.com .

### 1.4.2.4. RS DE PARTAGE DE PHOTOS

Les photos sont téléchargées vers les RS par l'utilisateur et elles sont partagées avec les autres membres souhaités du réseau.

Exemple : Clickfriends.com, Flickr.com, Fotki.com, Snappages.com, Picasa.com, Zoomr.com.

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

---

### 1.4.2.5. RS DE PARTAGE AUDIO

Dans ces communautés l'utilisateur peut développer ses propres musiques, les distribuer et au même temps jouer de la musique avec les autres membres en ligne.

Exemple : Last.FM, Pandora.com, iLike.com, Imeem.com, Radioblog.com.

### 1.4.2.6. RS DE PARTAGE DE PRÉSENTATION

Ces plateformes permettent aux utilisateurs de visualiser les fichiers de différents formats comme PDF (livre ou article scientifique...etc.) sous forme de présentation en ligne.

Exemple : Slideshare.net, Scribd, Issuu.

### 1.4.2.7. RS DE PARTAGE DE DISCUSSION

Les RS de discussions sont fondés et subsistent grâce aux utilisateurs qui se réunissent autour des sujets et des intérêts communs, en élaborant des conversations qui peuvent être instantanées ou hors ligne sur des thèmes spécifiques. Ces conversations amènent d'importantes masses d'informations actuelles, d'où nous pouvons dire que ces RS sont des moyens d'interaction sociale directe pour les utilisateurs.

Exemple :

Les forums, les RS de chat.

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

On peut définir la structure sociale par toutes les unités qui constituent une société ou un réseau social dans notre cas comme les individus, les groupes, les communautés... etc. Ainsi que les liens et les relations qui relient ces unités. On peut inclure aussi les réactions et le comportement des unités et le développement effectué sur les relations et les liens.

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

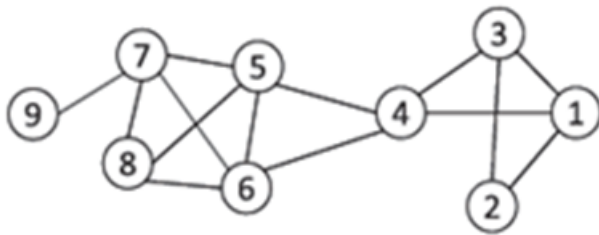
### 1.5.1. LES ACTEURS SOCIAUX

Les acteurs sociaux constituent l'unité la plus importante de la structure sociale, sans eux on ne peut pas parler de RS.

Les individus ou les utilisateurs d'un RS sont les acteurs évidents, néanmoins, les groupes et les communautés formés par les individus sont considérés comme acteurs sociaux.

### 1.5.2. LES LIENS SOCIAUX

Les liens sociaux sont les relations créées par les acteurs sociaux, les liens sociaux peuvent prendre plusieurs sens et types. Elles peuvent être amicales ou professionnelles ou familiales ...etc.



Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

- Représentation graphique

- Représentation matricielles

FIGURE 1.4 – Représentation graphique et Représentation matricielles.

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

---

### 1.5.3. REPRÉSENTATION GRAPHIQUE

Les graphes tirent leur nom du fait qu'on peut les représenter par des dessins.

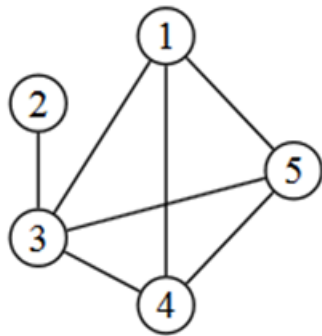
A chaque sommet de  $G$  on fait correspondre un point distinct du plan et on relie les points correspondant aux extrémités de chaque arête.

Il existe donc une infinité de représentations d'un graphe.

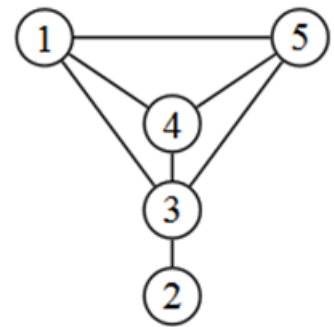
Les arêtes ne sont pas forcément rectilignes.

Si on peut dessiner un graphe  $G$  dans le plan sans qu'aucune arête n'en coupe une autre (les arêtes ne sont pas forcément rectilignes), on dit que  $G$  est planaire.

Le graphe  $G$  ci-dessus est planaire



Une représentation non planaire du graphe  $G$  (des arêtes se croisent)



Une représentation planaire de  $G$

FIGURE 1.5

### 1.5.4. REPRÉSENTATION MATRICIELLE

Graphe non-orienté :

Soit  $G$  un graphe non-orienté qui possède  $n$  sommets numérotés de 1 à  $n$ .

On appelle matrice d'adjacence du graphe la matrice  $A = (a_{i,j})$  où  $a_{i,j}$  est le nombre d'arêtes joignant le sommet  $i$  au sommet  $j$ . [6]

Exemple : Voici un graphe, et la matrice d'adjacence correspondante :

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

---

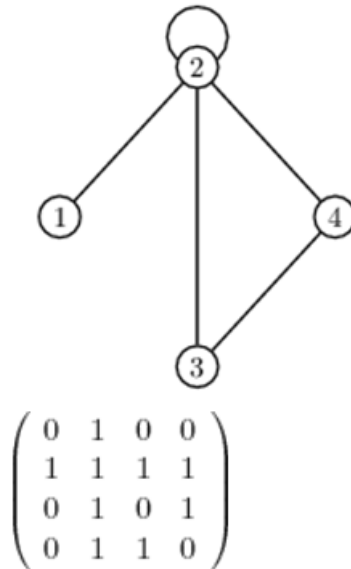


FIGURE 1.6

On peut remarque que cette matrice est symétrique.

C'est le cas pour toutes les matrices d'adjacence d'un graphe non-orienté.

Le résultat principal concernant les matrices d'adjacence est le théorème suivant :

### 1.5.4.1. THÉORÈME

Soit  $G$  un graphe non-orienté de matrice d'adjacence  $A$ . Le nombre de chaînes de longueur  $n$  joignant le sommet  $i$  au sommet  $j$  est donné par le terme d'indice  $i,j$  de la matrice  $A^n$ .

### 1.5.5. GRAPHE ORIENTÉ

on peut également définir la matrice d'adjacence d'un graphe orienté.

Cette fois, le coefficient  $a_{i,j}$  désigne le nombre d'arcs d'origine  $i$  et d'extrémité  $j$ .

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

---

Exemple : Pour le graphe suivant :

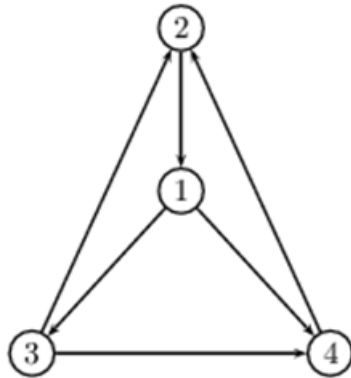


FIGURE 1.7

on trouve la matrice d'adjacence :

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

FIGURE 1.8

Cette matrice n'est plus symétrique. En revanche, le terme d'indice  $(i,j)$  de la matrice  $A^n$  compte toujours le nombre de chemins allant de  $i$  vers  $j$ . [12]

### 1.5.6. FONCTIONNEMENT D'UN RÉSEAU SOCIAL

Le fonctionnement d'un réseau social est souvent caractérisé par une même et unique procédure : création d'un profil, recherche d'autres profils en relation avec mes centres d'intérêt et enfin la mise en relation (directe ou non) Optionnellement, la prise de contact (commentaire, mail etc.) peut être intégrée. Bien entendu, en fonction de la typologie des réseaux certaines différences apparaissent. D'un côté les réseaux sociaux de type plate-forme, mettent à disposition des contenus sans création de profil, sans qu'il y ait besoin d'avoir une appartenance à la communauté. Cependant, si l'internaute souhaite faire partie de cette communauté, en diffusant ses propres vidéos par exemple, il doit obligatoirement se créer une fiche d'identité (profil). Ainsi, l'internaute est reconnu en tant que membre de ladite communauté . Une fois qu'il est membre, il est possible de rechercher des profils, de diffuser et de partager des contenus. Par contre, il n'y a pas de mise en relation à proprement parler. Les membres du réseau ne peuvent interagir entre eux qu'en réagissant, sous forme de commentaires, à des contenus publiés. Ainsi les réseaux de type plate-forme sont essentiellement orientés vers le partage et la diffusion de contenu (souvent des vidéos ou des photos) De l'autre côté, on a les réseaux sociaux personnels (généralistes ou thématiques) qui se rapprochent réellement plus de l'univers social et partage de contenu. La lecture du contenu est toujours libre d'accès, comme pour les réseaux de type plate-forme. Cependant, la création d'un profil est nettement plus avancée. Ce profil, véritable carte d'identité, est essentiel pour entrer dans la communauté mais surtout pour y participer et créer des liens avec les autres membres. En général, ce profil est présenté par un espace personnel (page web) visible par tous les internautes. Il s'agit d'un espace réservé aux membres où ils ont la possibilité d'y mettre tout ce qu'ils souhaitent : Textes, histoires, journal intime, photos de vacances ou encore vidéos. La mise en relation des membres se fait de manière très simple soit par un lien vers l'autre profil que l'internaute insère manuellement dans blogroll soit en invitant l'autre membre à se joindre à son cercle d'amis. Les réseaux sociaux personnels sont donc plus orientés vers la diffusion d'informations que vers la relation entre membres. Enfin, les réseaux sociaux professionnels sont les plus avancés d'un point de vue des

## 1.5. REPRÉSENTATION D'UN RÉSEAU SOCIAL

---

fonctionnalités proposées pour la gestion de sa communauté. La création de son profil est primordiale pour pouvoir profiter de tous les services associés aux réseaux professionnels. L'objectif de ce type de sites est clairement de se construire un réseau le plus grand et pertinent possible. Que ce soit dans le cadre d'une recherche d'emploi, pour trouver des capacités de financement ou encore des opportunités de partenariat, les réseaux sociaux peuvent s'avérer bien utiles pour ce genre de demande. De ce fait, la création de sa fiche personnelle (son curriculum vitae ici) est vitale. Si vous voulez être trouvé où trouver du monde, il faut que cette fiche soit la plus complète possible tout en définissant bien ses objectifs professionnels. En fait, l'utilisation d'un réseau social en ligne doit se faire à l'image d'un véritable réseau. D'un point de vue de la mise en relation entre les membres, les réseaux professionnels sont certainement les plus aboutis tout comme la recherche de profil[7]. Elles aboutissent souvent sur les profils recherchés et en quelques clics, avec un message de demande de mise en contact, l'invitation est envoyée par courrier électronique[?]. Par la suite, la personne invitée a le choix de refuser ou d'accepter la mise en relation. Si elle l'accepte, elle sera en contact direct avec la personne et aura accès à toutes ses informations professionnelles mais également, et surtout, verra le réseau du membre ainsi que son degré de proximité avec ses autres contacts. La création de son réseau n'est pas plus compliquée que cela et permet d'être rapidement en relation avec le monde entier .

### 1.5.7. DOMAINES D'APPLICATION RÉSEaux SOCIAUX EN LIGNE

Les domaines d'application des réseaux sociaux sont très divers : relation client, gestion des changements en interne, valorisation des expertises... Et toutes les fonctions de l'entreprise peuvent bénéficier de leur mise en œuvre. Comme pour le collaboratif où il existe autant d'outils que de visions de mode de collaboration, les éditeurs de RSE proposent une mise en scène du réseau social différente. Certaines solutions ont choisi de reprendre l'approche égocentrée en matière d'organisation de l'information des réseaux sociaux publics (Facebook, Viadeo, etc.) [17], mettant au 1er plan pour chaque membre ses informations, ses discussions et partages, ses contacts ou celles du membre auquel on s'intéresse.

### 1.6. ANALYSE DES RÉSEAUX SOCIAUX EN LIGNE

L'analyse des réseaux sociaux est une approche sociologique fondée sur la théorie des réseaux sociaux[18], elle est définie comme étant l'étude des entités sociales (les personnes dans les organisations qu'on appelle acteurs) ainsi que leurs interactions et leurs relations. Ces interactions et relations peuvent être représentées par un graphe ou un réseau, dans lequel chaque nœud représente un acteur et chaque lien est une relation. On peut étudier les propriétés de la structure et son rôle ainsi que la position et le prestige de chaque acteur social on peut rechercher aussi les différents types de sous-graphes comme par exemple les communautés formées par des groupes d'acteurs ayant des intérêts communs, en isolant le groupe d'individus ayant une densité élevée . Le réseau social peut être aussi une source permettant l'élaboration de recommandations : Trouver un expert dans un domaine donné, suggérer des produits à vendre, proposer un ami, etc. Cette élaboration peut être fondée sur des algorithmes d'exploration de chemins, d'analyse de degré.

#### 1.6.1. ORIGINE DE L'ANALYSE DES RÉSEAUX SOCIAUX

L'analyse des réseaux sociaux est apparue il y a environ 70 ans. Son cycle de vie peut être divisé en trois grandes périodes : Les fondements de l'approche entre les années 1940 et 1960, certaines études publiées parlent sur les systèmes de relations (1957, Elisabeth Bott, système de relations des familles)[8]. Articulation de la méthode : entre les années 1960 et 1970, caractérisée par le développement des recherches méthodologiques. Des recherches sur le nombre de relations reliant les acteurs des RS ont été publiées et mises en place (1967, Stanley Milgram). A période du développement actuel : de 1980 jusqu'au aujourd'hui. Les recherches actuelles se mènent à d'autres fins tels que la famille, les relations de travail, la camaraderie...etc. et surtout sur la notion du petit monde .

### 1.7. LA REPRÉSENTATION DES RS DANS L'ARS

Pour effectuer de L'ARS sur un Réseau social on devient avoir une structure analysable qui représente les RS. L'ARS utilise deux approches pour représenter un réseau, soit par matrices ou par graphes.[14] De nos jours, les graphes sont utilisés dans tous les domaines grâce à leur simplicité d'utilisation et leur capacité à résumer les informations. La représentation graphique des RS est la plus répandue et on peut dire que l'une des racines de l'ARS est la théorie des graphes. On va introduire dans ce qui suit certaines notions sur la théorie de graphes les plus utilisées dans l'ARS .

#### 1.7.1. NOTION DE CALCULE EN ARS

On présente dans ce qui suit quelque mesure utilisée dans l'ARS : La centralité : Les acteurs importants sont ceux qui sont liés et impliqués avec les autres acteurs D'une façon extensive[12]. Dans le cadre d'une organisation, une personne qui a beaucoup de contacts et qui communiquent bien avec les autres personnes est considérée plus importante qu'une autre personne ayant moins de contacts. Ces contacts sont modélisés par des liens. Un acteur central est un acteur qui est impliqué dans plusieurs liens. Chaque nœud dans le réseau social est un acteur et chaque lien indique que les deux acteurs aux extrémités communiquent ensemble. Intuitivement, on remarque que l'acteur 1 est l'acteur le plus central par ce qu'il/elle communique avec la majorité des autres acteurs. Il y a différents types de liens ou d'implications entre acteurs. Par conséquent, plusieurs types de centralités seront définis sur les graphes dirigés et non dirigés.

#### 1.7.2. CENTRALITÉ D'INTERMÉDIARITÉ

Cette notion exprime le degré de liaison d'un acteur social avec d'autres acteurs sociaux qui ne sont pas forcément interconnectés directement entre eux. C'est le nombre de gens auxquels une personne est connectée de façon indirecte, via des liens directs [13]. Si deux acteurs non adjacents  $j$  et  $k$  veulent communiquer et si l'acteur  $i$  se localise sur le chemin entre  $j$  et  $k$ , alors  $i$  a un certain contrôle sur leur interaction. L'intermédiation mesure ce contrôle de  $i$

## 1.7. LA REPRÉSENTATION DES RS DANS L'ARS

---

sur les deux autres acteurs. Par conséquent, si  $i$  se localise sur le chemin de plusieurs interactions alors  $i$  est un acteur important.

### 1.7.3. CENTRALIENNE DE PROXIMITÉ

La centralienne de proximité est un indice exprimant si un acteur social est près de tous les autres acteurs du même RS (d'une manière directe ou non). Soit  $d(i, j)$  la distance la plus courte à partir de l'acteur  $i$  vers l'acteur  $j$  (mesurée par le nombre de liens via le chemin le plus court).

### 1.7.4. CENTRALISATION DE POUVOIR

C'est la différence entre le nombre de liens pour chaque nœud divisée par le nombre total de liens. Plus cet indice est élevé plus le RS est centralisé au tour des acteurs importants de ce RS.

### 1.7.5. PRESTIGE

Le prestige est une mesure plus raffinée que la centralité. on distingue entre les liens sortants et ceux entrants. Un acteur prestigieux est un acteur ayant beaucoup de liens entrants. Le prestige d'un acteur est mesuré par le nombre des liens entrants, Ensuite présenter les deux approches de visualisation : par matrice et par graphe.

### 1.7.6. LES BRANCHES DE L'ARS

Les principales branches dans l'ARS sont la détection des structures communautaires ou les groupes d'acteurs sociaux fortement liés, et l'étude temporelle du développement de la structure sociale.

### 1.7.7. LA DÉTECTION DES STRUCTURES COMMUNAUTAIRES

Une des tâches principales de l'analyse des réseaux sociaux est de capturer les structures communautaires dans les RS. La notion des structures communautaires a été reconnue comme une caractéristique statistique importante des systèmes

## 1.8. CONCLUSION

---

en réseau cette dernière décennie. En effet, une communauté peut être définie dans un RS modélisé par un graphe comme un sous-graphe dont les nœuds sont fortement connectés entre eux mais moins connectés avec les autres nœuds du réseau. Cette branche de l'ARS s'intéresse à trouver de tels groupes de nœuds dans un réseau. On peut interpréter ce problème par un problème de classification, c'est à dire satisfaisant le critère qui dit que les nœuds dans chaque communauté (sous-graphe) ont des liens denses entre eux, mais moins chaînés avec les autres. C'est à dire, la recherche des communautés dans un RS revient à classer les acteurs sociaux par les méthodes statistiques de classification.

### 1.7.8. L'ÉTUDE DE L'ÉVOLUTION DYNAMIQUE DES RÉSEAUX SOCIAUX

L'un des aspects les plus importants des RS est l'évolution dynamique de leurs structures au fil du temps. Les RS sont fondés sur deux notions de base principales : Acteur social et lien social, sachant que le nombre des acteurs sociaux et leurs états (rôle, importance...etc.) changent avec le temps ainsi que les interactions et les relations qu'ils construisent sont aussi instables. [19] Par conséquent, il est nécessaire de trouver des techniques d'études des RS qui prennent en compte cet aspect. L'une des techniques qui existent est l'évaluation temporelle des RS . L'évaluation temporelle visualise l'état de la structure sociale à un instant donné en évaluant son comportement à cet instant.

## 1.8. CONCLUSION

Les réseaux sociaux ont toujours été présents dans la vie sociale de l'être humain. Ils n'attendaient plus qu'un moyen pour prendre de l'ampleur. Ce moyen est Internet et les nouvelles technologies. Ainsi il est possible de créer des réseaux de plusieurs millions de personnes. Tout individu peut entrer en relation par le biais de 5 contacts maximum, avec n'importe quelle personne se trouvant sur la surface de la terre. La puissance des réseaux sociaux est donc décuplée grâce au média Internet. Il s'agit là d'un phénomène que nul ne pouvait prévoir, il y a encore quelques années. La technologie a mis à la portée de tout à

## 1.8. CONCLUSION

---

chacun ou presque un outil possédant un très fort potentiel social. Enfin, il est intéressant de mettre en étroite relation entre les réseaux sociaux et le monde de l'entreprise. C'est en effet, un ensemble sous-estimé du domaine social sur Internet. Bien entendu, il existe des réseaux sociaux professionnels mais leur utilisation reste exclusivement externe à l'entreprise. Lorsque l'on parle de communauté, d'environnements collaboratifs, de dynamique de groupe, tout le monde pense à l'entreprise, lieu de prédilection des relations et de contacts humains.

Chapitre **2**

## DÉTECTION DE COMMUNAUTÉ

### Sommaire

---

2.1	Introduction	35
2.2	Communautés	35
2.2.1	Définitions	35
2.2.2	Définitions basées sur la similarité des sommets	37
2.2.3	Les méthode pour calculer la similarité	37
2.2.4	Complexité informatique	39
2.2.5	Objective de détection de communautés dans les graphes	39
2.3	Les approches	40
2.3.1	Méthodes Traditionnelles	40
2.4	Modularité	53
2.5	évaluation	55
2.6	Conclusion	56

---

### 2.1. INTRODUCTION

**L**A science moderne des réseaux a apporté des avancées significatives à notre compréhension des systèmes complexes.

Une des caractéristiques les plus pertinentes des graphes représentant des systèmes réels est la structure de communauté, ou regroupement, c'est-à-dire l'organisation de sommets dans des clusters, avec de nombreux arêtes reliant les sommets d'un même cluster et relativement peu.

De tels amas, ou communautés, peuvent être considérés comme des compartiments relativement indépendants d'un graphe, jouant un rôle similaire.

La détection des communautés revêt une grande importance en sociologie, en biologie et en informatique, disciplines où les systèmes sont souvent représentés sous forme de graphe.

Ce problème est très difficile et n'est pas encore résolu de manière satisfaisante, malgré les efforts considérables d'une grande communauté interdisciplinaire de scientifiques qui y travaillent depuis quelques années.

Dans cette section nous allons présenter les différentes méthodes de détection de communautés et quelques définitions sur les communautés .

### 2.2. COMMUNAUTÉS

#### 2.2.1. DÉFINITIONS

##### 2.2.1.1. LES BASES

Le premier problème dans le regroupement de graphe est de chercher une définition quantitative de la communauté. [19]

Aucune définition n'est universellement acceptée.

En fait la définition dépend souvent du système spécifique en question et / ou l'application que l'on a en tête.

Il nous obtenons l'idée qu'il doit y avoir plus d'arêtes "à l'intérieur" de la communauté que les sommets de la communauté avec le reste du graphe.

## 2.2. COMMUNAUTÉS

---

Une propriété requise d'une communauté est la connexité.

Nous nous attendons à ce que le sous graphe soit une communauté, il doit y avoir un chemin entre chaque paire de ses sommets, en cours d'exécution seulement à travers les sommets de sous graphe.

Cette fonctionnalité simplifie la tâche de détection de communauté sur des graphes déconnectés, comme dans ce cas on analyse juste chaque composant connecté séparément, sauf si des contraintes particulières sont imposées on clusters résultantes (Garey and Johnson,1990 ; Papadimitriou, 1994).

### 2.2.1.2. DÉFINITIONS LOCALES

Les communautés sont des parties du graphe avec quelques liens avec le reste du système.

Dans une certaine mesure, ils peuvent être considérés comme des entités séparées avec leur propre autonomie, Donc il est logique de les évaluer indépendamment.

Les communautés peuvent également être identifiées par une mesure de forme physique, exprimant dans quelle mesure un sous-graphe satisfait donné la propriété liée à sa cohésion.

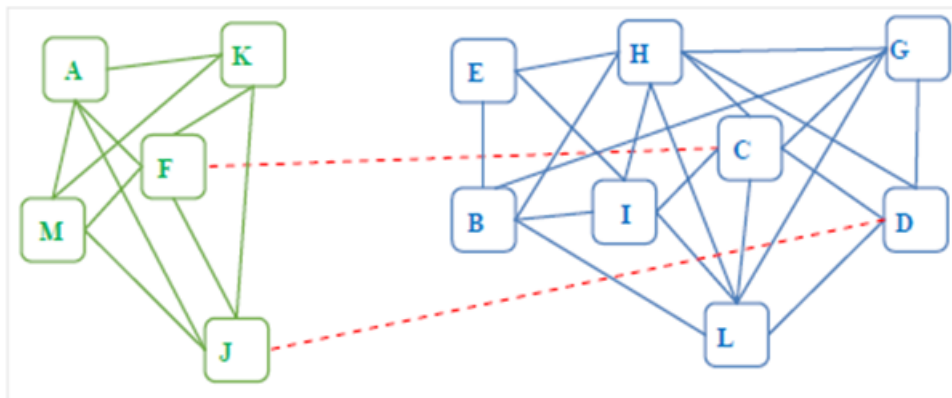


FIGURE 2.1 – *exemple simple communauté.*

## 2.2. COMMUNAUTÉS

---

### 2.2.1.3. DÉFINITIONS GLOBALES

Les communautés peuvent également être définies en graphe dans son ensemble.

Cela est raisonnable dans ces cas quels groupes sont des parties essentielles du graphe qui peuvent ne pas être démonté sans affecter sérieusement le fonctionnement du système.

La littérature offre de nombreux critères pour identifier les communautés.

Dans la plupart des cas, ils sont des définitions indirectes, dans lesquelles certaines propriétés globales du graphe sont utilisées dans un algorithme qui fournit des communautés à la fin.

Cependant, il existe une classe de définitions appropriées, basé sur l'idée qu'un graphe a une structure communautaire s'il est différent d'un graphe aléatoire.

### 2.2.2. DÉFINITIONS BASÉES SUR LA SIMILARITÉ DES SOMMETS

Il est naturel de supposer que les communautés sont des groupes de sommets similaires les uns aux autres.

On peut calculer la similarité entre chaque paire de sommets par rapport à certains propriété de référence, locale ou globale, peu importe si ils sont connectés par un bord ou non.

S'il est possible d'incorporer les sommets du graphe dans un espace euclidien n-dimensionnel, en assignant une position à eux, on pourrait utiliser la distance entre une paire de comme mesure de leur similitude (c'est en fait une mesure sûr de la dissimilitude parce que des sommets similaires sont attendus d'être proches les uns des autres).

Compte tenu des deux points de données :  $A = (a_1, a_2, \dots, a_n)$  et  $B = (b_1, b_2, \dots, b_n)$ .

### 2.2.3. LES MÉTHODE POUR CALCULER LA SIMILARITÉ

Euclidean distance(L2-norm)

## 2.2. COMMUNAUTÉS

$$d_{AB}^M = \sum_{k=1}^n |a_k - b_k|$$

Manhattan distance (L1-norm)

$$\rho_{AB} = \arccos \frac{a \cdot b}{\sqrt{\sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2}}$$

Et  $L_\infty$  - norm

$$\rho_{AB} = \arccos \frac{a \cdot b}{\sqrt{\sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2}}$$

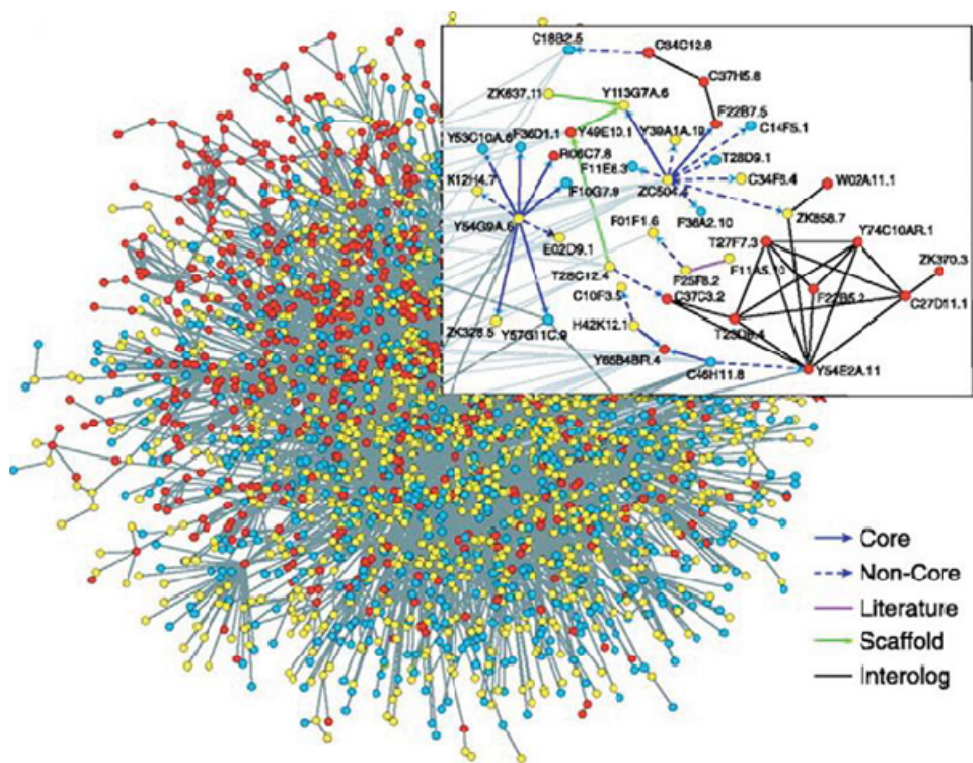


FIGURE 2.2 – Réseau d'interaction Elégans Protéine.

## 2.2. COMMUNAUTÉS

---

### 2.2.4. COMPLEXITÉ INFORMATIQUE

L'énorme quantité de données sur les réseaux réels actuellement disponible rend la question de l'efficacité du regroupement algorithmes essentiels.

La complexité de calcul d'un algorithme est l'estimation de la quantité de ressources requis par l'algorithme pour effectuer une tâche.

Cela implique à la fois le nombre d'étapes de calcul nécessaires et le nombre d'unités de mémoire qui doivent être simultanément alloué pour exécuter le calcul.

De telles demandes sont généralement exprimées par leurs évolutivités avec la taille de système.

Dans le cas d'un graphe, la taille est indiqué par le nombre de sommets  $n$  et  $m$  / ou nombre d'arêtes.

### 2.2.5. OBJECTIVE DE DÉTECTION DE COMMUNAUTÉS DANS LES GRAPHE

L'objectif de la détection de communautés dans les graphes, ou encore dans les réseaux sociaux est de créer une partition des sommets, en tenant compte des relations qui existent entre les sommets dans le graphe, de telle sorte que les communautés soient composées de sommets fortement connectés et qu'elles soient peu reliées entre eux (Brandes et al., 2003; Newman, 2004; Schaeffer, 2007; Lancichinetti et Fortunato, 2009).

Parmi les principales méthodes de détection de communautés proposées dans la littérature, on peut citer celles qui optimisent une fonction de qualité pour évaluer la qualité d'une partition donnée, comme la modularité, la coupe ratio, la coupe min-max ou la coupe normalisée (Kernighan et Lin, 1970; Chan et al., 1994; Shi et Malik, 2000; Ding et al., 2001; Newman et Girvan, 2004), les techniques hiérarchiques comme les algorithmes de division (Flake et al., 2003), les méthodes spectrales (Von Luxburg, 2007) ou l'algorithme de Markov et ses extensions (Satuluri et Parthasarathy, 2009).

Ces techniques de partitionnement de graphes sont très utiles pour détecter

## 2.3. LES APPROCHES

---

des composantes fortement connectées dans un graphe.

Cette section est consacrée à la détection de telles communautés qui repose donc uniquement sur les arêtes et leurs dispositions dans et entre les classes. Deux sommets classés un ensemble doivent être plus liés entre eux, directement ou par l'intermédiaire d'autres sommets, que vis-à-vis de sommets placés dans d'autres classes.

Deux sommets classés un ensemble doivent être plus liés entre eux, directement ou par l'intermédiaire d'autres sommets, que vis-à-vis de sommets placés dans d'autres classes.

## 2.3. LES APPROCHES

### 2.3.1. MÉTHODES TRADITIONNELLES

#### 2.3.1.1. PARTITION DE GRAPHE

Le problème du partitionnement de graphe consiste à diviser les sommets en groupes de taille prédéfinie .

La partition de graphe est un problème fondamental en parallèle en l'informatique, le partitionnement et la disposition des circuits, et dans conception de nombreux algorithmes en série, y compris les techniques résoudre des équations différentielles partielles et linéaires éparses de systèmes d'équations.

La plupart des variantes du partitionnement de graphe ont le problème de NP-difficile.

Il y a cependant plusieurs algorithmes qui peuvent faire du bon travail, même si leurs solutions ne sont pas nécessairement optimales (Pothen, 1997).

De nombreux algorithmes effectuer une bissection du graphe.

Partitions dans plus de deux groupes sont généralement atteintes par itération bissection De plus, dans la plupart des cas, on impose contrainte que les clusters ont une taille égale.

Ce problème est appelée bissection minimale et est NP-difficile.

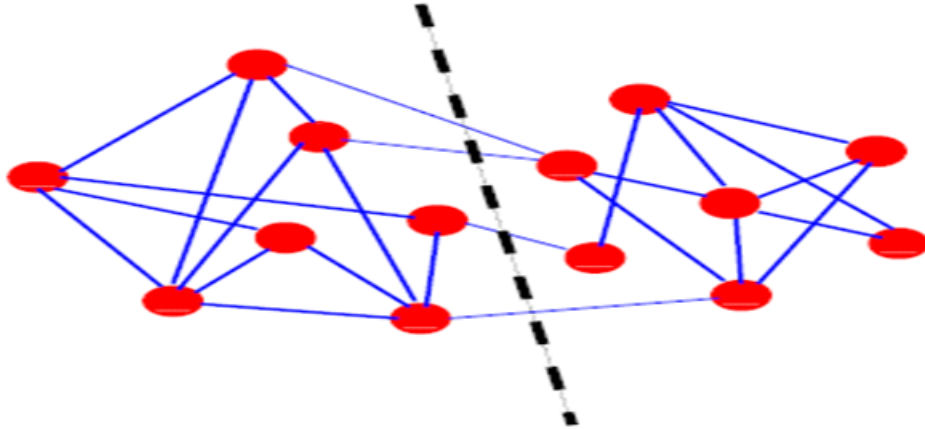


FIGURE 2.3

Partitionnement de graphe. La ligne pointillée montre la solution du problème de bissection minimum pour le graphe illustré, la partition en deux groupes de taille égale avec un minimum nombre d'arêtes entre les groupes.

### 2.3.1.2. PARTITIONS

Notions de base :

Une partition est une division d'un graphe en clusters, telle que chaque sommet appartient à un cluster, dans les sommets de systèmes réels peuvent être partagés entre différentes communautés. Une division d'un graphique en chevauchement (ou floue) communautés est appelée couverture.

Le nombre de partitions possibles dans  $k$  clusters d'un graphe avec  $n$  sommets est le nombre de Stirling de second.

Les partitions peuvent être hiérarchisées, lorsque le graphique a différents niveaux d'organisation / structure à différentes échelles.

Dans ce cas, les clusters s'affichent à leurs tours de structure communautaire, avec de plus petites communautés à l'intérieur, qui peut contenir de plus petites communautés.

### 2.3. LES APPROCHES

---

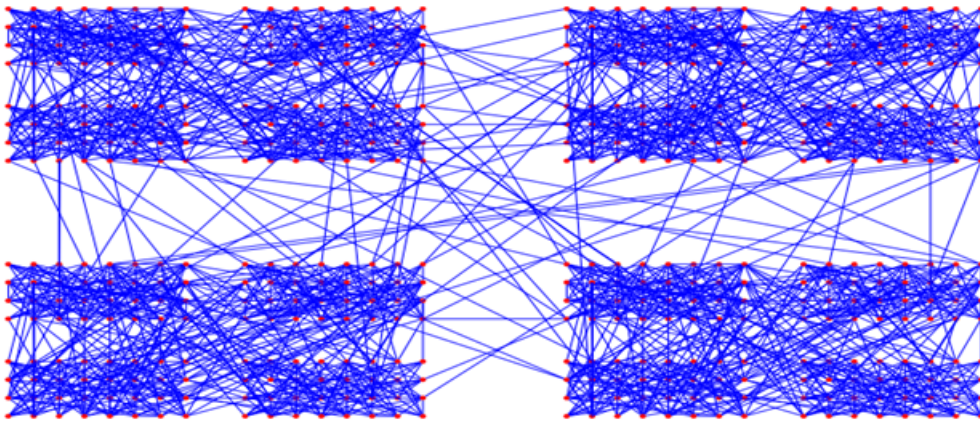


FIGURE 2.4 – Schématisation d'un graphe hiérarchique.[19]

Les fusions de communautés sont représentées par des lignes horizontales. Le niveau le plus élevé représente le graphique entier en tant que communauté unique.

Coupe le diagramme horizontalement à une certaine hauteur Le diagramme est hiérarchique par construction :

chaque communauté appartenir à un niveau est entièrement inclus dans une communauté à un niveau supérieur.

Les dendrogrammes sont régulièrement utilisés sociologie et biologie.

La technique de la classification hiérarchique.

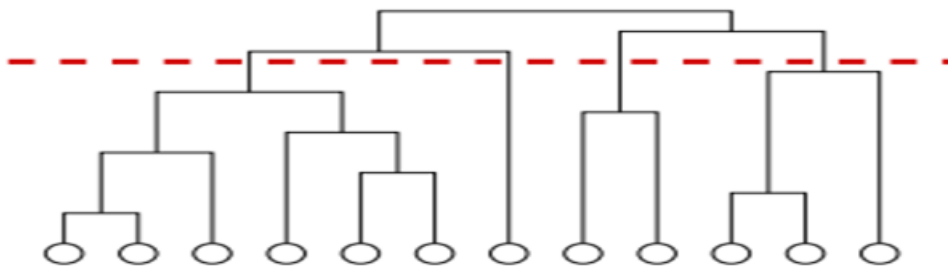


FIGURE 2.5 – Un dendrogramme, ou arbre hiérarchique ,Horizontal les coupes correspondent aux partitions du graphe dans les communautés.

## 2.3. LES APPROCHES

---

### 2.3.1.3. CLASSIFICATION HIÉRARCHIQUE

En général, on en sait très peu sur la communauté et la structure d'un graphe.

Il est rare de connaître le nombre des grappes dans lesquelles le graphe est divisé, ou d'autres indications à propos de l'adhésion des sommets.

Dans telle procédures de regroupement des cas comme les méthodes de partitionnement de graphe peut difficilement être utile, et on est obligé de faire hypothèses raisonnables quant au nombre et à la taille des clusters, qui sont souvent injustifiées.[20]

D'autre part, le graphe peut avoir une structure hiérarchique, i. e. afficher plusieurs niveaux de regroupement des sommets, avec petites clusters incluses dans de grands groupes, qui sont à son tours inclus dans les grands groupes, et ainsi de suite.

par exemple réseaux sociaux ont souvent une structure hiérarchique (Section III.C.1).

Dans de tels cas, on peut utiliser la hiérarchie algorithmes de regroupement (Hastie et al., 2001), i. e. regroupement techniques qui révèlent la structure à plusieurs niveaux de la graphe.

La classification hiérarchique est très fréquente dans les réseaux sociaux, analyse de réseau, biologie, ingénierie, marketing, etc.

Le point de départ de tout regroupement hiérarchique La méthode est la définition d'une mesure de similarité entre sommets.

Après avoir choisi une mesure, on calcule la similitude pour chaque paire de sommets, peu importe s'ils sont connecté ou non.

A la fin de ce processus, il en reste un avec une nouvelle matrice  $n \times n$ , la matrice de similarité.

Les techniques de classification hiérarchique visent à identifier les groupes de sommets avec une grande similitude, et peut être classé en deux catégories :

- Algorithmes d'agglomérations, dans lesquels les clusters sont itérativement fusionné si leur similarité est suffisante haute.
- Algorithmes de diviseurs, dans lesquels les clusters sont itérativement diviser en supprimant les arêtes reliant les sommets avec faible similitude[21].

## 2.3. LES APPROCHES

---

### 2.3.1.4. CLUSTERING PARTIEL

Le regroupement partitionnel indique une autre classe populaire des méthodes pour trouver des grappes dans un ensemble de points de données.

Ici, le nombre de groupes est pré assigné, disons  $k$ .

Les points sont intégrés dans un espace métrique, de sorte que chaque sommet est un point et une mesure de distance sont définis entre les paires de points dans l'espace.

La distance est une mesure de dissimilarité entre les sommets.

L'objectif est de séparer des points dans  $k$  groupes de manière à maximiser / minimiser un donné fonction de coût basée sur les distances entre les points et / ou des points aux centros, i. e. positions convenablement définies dans l'espace.

Certaines des fonctions les plus utilisées sont énumérées ci-dessous :

K-clustering minimum, a fonction de coût est ici le diamètre d'un cluster, qui est la plus grande distance entre deux points d'un cluster.

Les points sont classés de telle sorte que le plus grand du groupe  $k$  les diamètres sont les plus petits possibles .

L'idée est de Gardez les clusters très compact ".

La somme de k-clustering ,identique au k-clustering minimum, mais le diamètre est remplacé par la distance moyenne entre toutes les paires de points d'un cluster.

k-centre, pour chaque cluster, on définit une référence point  $x_i$ , le centro, et calcule le maximum  $d_i$  des distances de chaque point de cluster de le centro.

Les clusters et les centros sont auto-consciemment choisi de telle sorte à minimiser la plus grande valeur de  $d_i$  .

k-médiane,même que k-centre, mais la distance maximale du centro est remplacé par la moyenne distance .

### 2.3.1.5. CLUSTER SPECTRAL

Supposons avoir un ensemble de  $n$  objets  $x_1, x_2, \dots, x_n$  avec une fonction de similarité par paire  $S$  définie entre eux, qui est symétrique et non négatif (c'est-

## 2.3. LES APPROCHES

---

à-dire,

$$S(x_i; x_j) = S(x_j; x_i) \geq 0, \text{ pour tout } i, j = 1, \dots, n$$

. Cluster spectral ter inclut toutes les méthodes et techniques qui partitionnent l'ensemble en clusters en utilisant les vecteurs propres des matrices, comme S lui-même ou d'autres matrices qui en dérivent[22].

En particulier, les objets pourraient être des points dans un espace métrique, ou les sommets d'un graphe.

La classification spectrale consiste en une transformation de l'ensemble initial d'objets en un ensemble de points dans l'espace, dont les coordonnées sont des éléments de vecteurs propres :

l'ensemble des points est ensuite regroupé via la norme techniques, comme le clustering ,k-means.

Un peut se demander pourquoi il est nécessaire de regrouper les points obtenus à travers les vecteurs propres, quand on peut directement regrouper l'ensemble initial d'objets, en fonction de la matrice de similitude.

La raison en est que le changement de représentation induite par les vecteurs propres rend les propriétés du cluster de l'ensemble de données initial beaucoup plus évident.

De cette façon, clustering spectral est capable de séparer les points de données n'a pas pu être résolu en appliquant directement le clustering k-means, par exemple, comme ce dernier tend à livrer convexe ensembles de points[2].

### 2.3.1.6. ALGORITHMES PROPOSÉS ( METAHEURISTIQUES )

Dans cette section, six algorithmes adaptés au problème du CD sont décrits.

Les algorithmes originaux de recherche de chauves-souris et de gravitation ont été adaptés pour résoudre le problème d'optimisation de la modularité, qui est une optimisation discrète de problème[23].

Les autres algorithmes :

(big bang-big crunch modifié, chauve-souris améliorée :

algorithme basé sur différentiel évolutif, hyper-efficace, recherche et dispersion

### 2.3. LES APPROCHES

Dans cette étude, nous voulions analyser les résultats des algorithmes méta-heuristiques avec différents mécanismes les uns des autres dans la solution du problème CD.

En termes de l'original versions d'algorithmes, seulement l'algorithme de recherche de dispersion basé sur l'algorithme génétique est approprié pour résoudre des problèmes d'optimisation discrets. Les cinq autres algorithmes doivent être adaptés au problème d'optimisation discrète (le problème du CD).

Ces algorithmes sont conçus pour résoudre le problème de CD selon l'organigramme montré dans la figure .

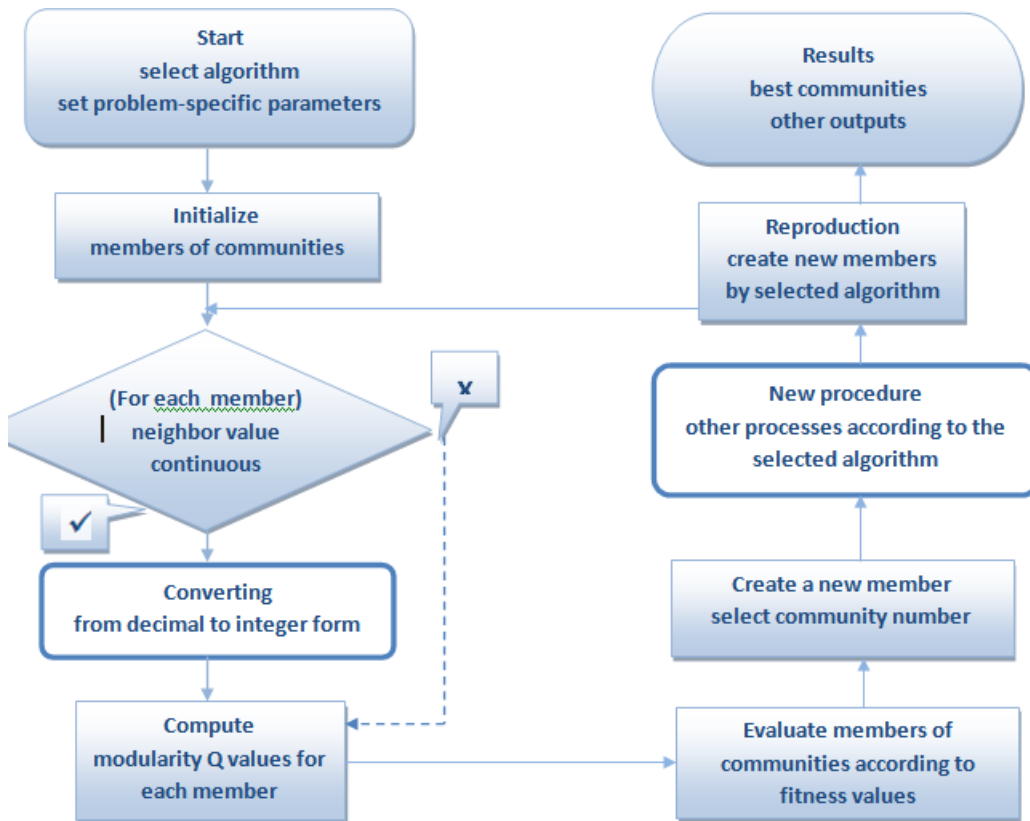


FIGURE 2.6 – Algorithme pour résoudre le problème de CD selon l'organigramme

**Algorithme de batte :** Bat Algorithm (BA), qui a été développé par Xin-She Yang est un méta heuristique algorithme d'optimisation basé sur le comportement

### 2.3. LES APPROCHES

---

d'écholocation des chauves-souris . L'écholocation est la détection d'emplacement par impulsions sonores [24].

Les chauves-souris génèrent une carte de direction en fonction de l'intensité des vibrations sonores réfléchies pendant l'exploration de la nourriture et du chemin, et ils convergent vers la direction appropriée.

Selon la règle mentionnée dans le algorithme, chaque chauve-souris a des paramètres tels que la vitesse, la position, la fréquence, longueur d'onde et volume.

Ces paramètres sont adaptés au problème du CD dans cette étude.

Chaque chauve-souris qui est générée aléatoirement à l'étape d'initialisation représente un ensemble de solutions en accord avec la matrice d'adjacence.

La longueur de cet ensemble est égale à nombre de sommets maximum, et les voisins appropriés sont choisis au hasard pour chaque bat en partant d'un (premier index).

Les valeurs de fitness des solutions candidates créées de cette fa sont calculées selon l'équation3.

La communauté la plus appropriée la détection vise à être faite en mettant à jour les vitesses et les positions des chauves-souris en fonction des valeurs Q obtenues.

Selon l'étape "Conversion" , la fonction de contrôle supplémentaire est définie pour la production de nouvelles chauves-souris conformément à la matrice d'adjacence.

## 2.3. LES APPROCHES

---

---

### Algorithme 1 : The pseudocode of the BA

---

```
1 Initialize the bat population  $x_i$  for  $i=1\dots n$ (in compliance with adjacency list)
2 Objective function  $f(x)$ ,  $x = (x_j, \dots, x_d)^T$  (according to modularity function)
3 Define pulse frequency  $Q_i \in [Q_{min}, Q_{max}]$ 
4 Initialize pulse rates  $r_i$  and the loudness  $A_i$ 
5 while  $t < \text{number of iterations}$  do
6   Generate new solutions by adjusting frequency according to step
   "Converting" in Fig.4
7   Updating velocities and locations/solutions
8   if  $\text{rand}(0, 1) > r_i$  then
9     Select a solution using  $\text{NewBat} = \text{BestBat} + \text{NewVelocity}$ 
10    Generate a locale solution around the best solution
11   Generate a new solution by flying randomly
12   if  $\text{rand}(0, 1) > A_i$  and  $f(x_i) < f(x)$  then
13     Accept the new solutions (if greater modularity values)
14     Increase  $r_i$  and reduce  $A_i$ 
15   Rank the bats and find the current best (bat with the best group)
16 end while.
```

---

#### 2.3.1.7. ALGORITHME DE CHAUVÉ SOURIS AMÉLIORÉ BASÉ SUR L'ALGORITHME ÉVOLUTIONNAIRE DIFFÉRENTIEL

Dans cette section, l'algorithme de chauve-souris amélioré basé sur l'évolution différentielle algorithm (BADE) est introduit.

Les principales étapes de cet algorithme ont été cons selon l'organigramme et la version originale de l'algorithme de la chauve-souris.

Le BADE implique deux étapes différentes, contrairement au BA précédent.

L'algorithme d'évolution différentielle (DEA) a été utilisé processus de régénération de la population de la BADE. En cours de production de solutions candidat , le DEA et la version originale de BA ont été utilisés l'ensemble pour la sélection de sommets adjacents .

Le DEA, qui est un algorithme d'optimisation métaheuristique basé sur la population, génère des résultats assez réussis en optimisant les paramètres dans un espace de solution adapté aux problèmes du monde réel .

Une nouvelle personne est sélectionnée en fonction de Equation 7 dans le

### 2.3. LES APPROCHES

---

processus de mutation de l'algorithme original DEA.

Dans le BADE, de nouveaux individus sont produits en ajoutant le troisième vecteur et la différence pondérée des deux premiers vecteurs, en utilisant la formule donnée dans l'équation 7, de l'original DEA à trois vecteurs qui ont été choisis au hasard parmi la population de chauves-souris généré par le BA original.

La phase de reproduction individuelle de la BADE est montré à la Fig. 2.7.

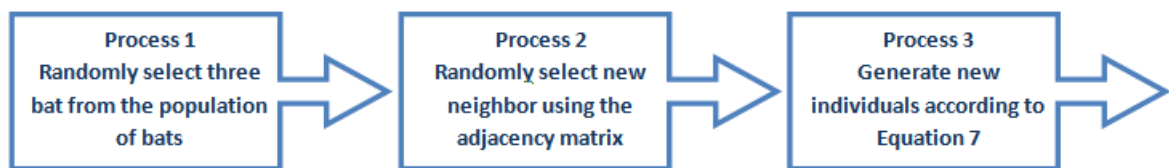


FIGURE 2.7 – Génération de nouvelle candidate solutions par BADE.

$$newBate = r^1 + fr^2 - r^3)$$

(7)

## 2.3. LES APPROCHES

---

---

### Algorithme 2 : The pseudocode of the BADE

---

```
1 Initialize the bat population  $x_i$  and  $v_i$  for  $i=1\dots p$ (in compliance with
   adjacency )
2 Objective function  $f(x)$ ,  $x = (x_j, \dots, x_d)^T$  (the modularity function)
3 Define pulse frequency  $Q_i \in [Q_{min}, Q_{max}]$ 
4 Initialize pulse rates  $r_i$  and the loudness  $A_i$ 
5 while  $t < \text{number of iterations}$  do
6   Generate new solutions by the DEA(by equation 7)
7   Updating new solutions according to step "converting" in fig.4
8   Updating velocities and locations/solutions
9   if  $\text{rand}(0, 1) > r_i$  then
10    modify the solution using
           
$$\text{NewBat} = \text{BestBat} + (1 - \frac{t}{\text{iter}}) + \text{NewVelocity}$$

11   Generate a new solution by flying randomly
12   if  $\text{rand}(0, 1) < A_i$  and  $f(x_i) < f(x)$  then
13     Accept the new solutions (if greater modularity values)
14     Increase  $r_i$  and reduce  $A_i$ 
15   Rank the bats and find the current best (bat with the best group)
16 end while.
```

---

## 2.3. LES APPROCHES

---

### 2.3.1.8. ALGORITHME BIG BANG-BIG CRUNCH MODIFIÉ

L'algorithme Big Bang-Big Crunch (BB-BC) a été développé en 2006, inspiré par une idée liée à la formation et la destruction de l'univers.

Selon cette approche, les étoiles sont initialement uniformément réparties dans l'espace de la solution. A savoir, les vecteurs des candidats sont formés.

Cette étape est similaire à la théorie du Big Bang concernant la formation de l'univers. Ensuite, les valeurs dans les vecteurs de la fonction à optimiser sont déterminées, et la moyenne pondérée de ces vecteurs est trouvées en donnant le plus grand poids.

Cette étape est comme la phase Big Crunch et le point obtenu est considéré comme le centre de l'univers.

à la prochaine étape, un nouveau Big Bang, qui est non uniforme distribué et concentré dans le centre de l'univers, se produit. Cette étape continue avec une réévaluation et un nouveau Big Crunch.

Ces étapes sont répétées tant que le centre de l'univers change. Sinon, le dernier vecteur obtenu au centre de l'univers est accepté comme solution acquise.

Selon l'approche BB-BC décrite ci-dessus, les valeurs de modularité peuvent être calculées sur la base de la relation d'adjacence. Par conséquent, chaque somme tout graphe donné est ajouté à la communauté la plus appropriée. Le BB-BC proposé utilisé dans les expériences a été exécuté sur un espace de solution qui est différent du vecteur d'espace de la version originale du BB-BC qui contient des valeurs continues. L'espace de solution de l'algorithme proposé est adapté au problème et consiste en 15 indices des sommets adjacents. De nouveaux individus sont produits autour du local meilleure solution dans chaque étape de l'algorithme original.

Selon l'approche de l'original BB-BC, chaque population nouvellement créée fonctionne indépendamment de l'espace de population précédent (espace vectoriel).

Selon le BB-BC proposé, les nouveaux individus produits dans chaque génération sont comparés aux meilleures solutions la population mondiale. Si les nouveaux individus ont obtenu à la fin de chaque génération qui a de meilleures solutions, ils sont inclus dans la population mondiale.

### 2.3. LES APPROCHES

---

Par conséquent, l'élite les individus des générations précédentes sont protégés et la population mondiale est également soutenue par les nouvelles solutions.

$$newCondidat = [selectedtemp\textit{solution}] + rand(1, n) * (2) - 1$$

(8)

Le pseudo code du BB-BC est donné dans le tableau 3. Les étapes entre 1 et 4 montrent la phase du Big Bang (construction de solutions).

D'autres étapes montrent la phase Big Crunch (mouvement de recherche locale) Aux étapes 7 et 8, on trouve le meilleur candidat et le meilleur centre de masse, respectivement.

Dans l'étape suivante, un candidat sur deux obtenu dans les étapes précédentes est choisi au hasard.

à l'étape 10, les nouveaux candidats sont obtenus par l'équation 8, où n indique le nombre de sommets dans un réseau donné.

---

**Algorithme 3** : The pseudocode du BB-BC

---

- 1 Initialize the population in a random manner(for only the first generation)
  - 2 Set the upper and lower limit ranges in accordance with the network
  - 3 Save candidates with the best communities
  - 4 Generate new population from the elite individuals (other generations)
  - 5 Generate some neighbors for all solutions in the population
  - 6 Replace the parent with its best young individuals for each
  - 7 temp solution 1 : find the center of mass
  - 8 temp solution 2 : find the best candidate of mass
  - 9 Randomly select a solution (temp solution 1 or 2)
  - 10 Perform local search and generate new population according to Equation 8
  - 11 Update the elite pool and the best found solution until now
  - 12 Return to step 1 if stopping criterion is not reached
  - 13 Return the best candidate (obtained best communities)
-

### 2.4. MODULARITÉ

L'une des fonctions de fitness les plus populaires et bien connues dans la littérature est la fonction de modularité.

Le concept de modularité a été introduit en 2004 .

Dans ces études, la fonction de modularité basées sur le principe selon lequel les relations entre les membres dans les communautés doivent être maximisés et que les relations de ces membres avec les membres des autres communautés doivent être minimisé.

La fonction d'évaluation ainsi obtenue est basée sur une approche non aléatoire et modèle mathématique.

Lorsque l'on considère également la difficulté de l'analyse des grandes et structures de graphe complexes, on peut comprendre que le problème de maximisation de la modularité est un problème de NP-difficile.

La fonction de modularité la plus basique  $Q$  qui représente la fonction de modularité du graphe est donnée dans l'équation suivante :

$$Q_{basic} = \sum_{i=1}^k (e_{ii} - a_i^2)$$

(1)

Où, indique les probabilités de bord des sommets binaires, montre le fraction d'arêtes ayant au moins une extrémité dans le groupe (pourcentage d'arêtes avec au moins 1 terminer dans le module  $i$ ).

Pour une fonction de modularité plus générale, lorsqu'un réseau d'échantillons donné est représenté avec toute structure de graphe, les structures communautaires acquises peuvent être considérées comme sous-graphes qui ont les qualités et les quantités telles que la propriété communes maximums , le nombre d'interaction, la similitude de position en soi.

Alors que les sommets (nœuds) qui sont les membres de ces structures doivent avoir le maximum relations ou propriétés communes maximums avec les nœuds dans sa propre communauté (à l'intérieur de la communauté), ils doivent avoir les relations minimales ou communes de propriétés avec les nœuds de l'autre

## 2.4. MODULARITÉ

---

communauté (en dehors de la communauté).

En domaines social , les personnes ayant de fortes relations se trouvent dans les communautés similaires, les colonies de créatures vivantes qui se nourrissent les unes les autres se trouvent dans la même communauté dans les domaines de l'environnement, et les clusters d'ordinateurs qui ont le maximum de données l'échange et la collaboration maximale qu'ils se forment sont situés dans la même communauté dans les réseaux informatiques peut être donné comme un exemple lié à la Problème de CD.

Supposons qu'une structure de graphe donnée  $G (V, E)$  représente un réseau d'échantillons, et  $G, V$  et  $E$  représente le graphe donné, l'ensemble des sommets et l'ensemble des arêtes, respectivement.

$$V = \{v_i | i = 1, 2, 3 \dots n\} \text{ et } E = \{e_j | j = 1, 2, 3 \dots m\}$$

où  $i, j, n$  et  $m$  représentent l'indice du sommet, l'indice du bord, le nombre des sommets totaux et le nombre des arêtes totales, respectivement. De plus, supposons qu'une matrice d'adjacence  $Adj$  qui démontre les relations entre les membres des groupes  $E$  et  $V$  avec des dimensions  $n \times n$  est défini.

$Adj$  est généré par l'équation 2

$$Adj_{(i,j)} = \begin{cases} 1 & \text{if node } i \text{ and node } j \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases}$$

(2)

$Q$  qui représente la fonction de modularité du graphe  $G$  est donnée dans l'équation 3.

$$Q = \frac{1}{2 * m} \sum_{ij} (Adj_{(i,j)} - \frac{k_i * k_j}{2 * m}) * \delta(c_i, c_j)$$

(3)

où  $Q$  se réfère à la fonction objective à maximiser. Alors que  $Adj_{(i,j)}$  représente la matrice d'adjacence du graphe  $G$ , le nombre de bords totaux dans le réseau est calculé par la formule donnée dans l'équation 4.  $k_i$  et  $k_j$  décrivent les degrés des  $i$ ème et  $j$ ème sommets respectivement, et  $k_i$  le degré du sommet est calculé

## 2.5. ÉVALUATION

---

selon l'équation 5.

$C_i$  et  $C_j$  indique la communauté dans laquelle appartiennent les  $i$ ème et  $j$ ème sommets, respectivement .

$(c_i, c_j)$  exprime également la fonction qui indique si les  $i$ ème et  $j$ ème sommets participent au même communauté ou non.  $(c_i, c_j)$  crée une sortie selon l'équation 6.

$$m = \frac{1}{2} \sum_{ij} Adj_{(i,j)}$$

(4)

$$k_i = \sum_j Adj_{(i,j)}$$

(5)

$$\delta = \begin{cases} 1 & \text{if } C_i = C_j \\ 0 & \text{if } C_i \neq C_j \end{cases}$$

(6)

Dans cette étude, la fonction donnée dans l'équation 3 a été utilisée comme une fonction de remise pour la détection des communautés les plus appropriées du monde réel de réseaux donné dans le processus de test.

Les paramètres de cette fonction et le complément , descriptions de l'autre problème CD ont été adaptées aux algorithmes donnés dans la section suivante et les résultats obtenus des études expérimentales ont été présenté dans les tableaux .

## 2.5. ÉVALUATION

Dans la littérature, on rencontre principalement deux modes d'évaluation des méthodes développées en classification non supervisée de sommets dans

## 2.6. CONCLUSION

---

les réseaux d'information. La première d'entre eux est l'évaluation à l'aide de critères internes. On cherche alors à obtenir une partition la meilleure possible au sens d'un critère particulier, comme la modularité. Or face à un problème réel, la question que l'on se pose est plutôt "La solution que m'apporte l'algorithme est-elle de qualité, a-t-elle de bonnes propriétés?" que "La solution que m'apporte l'algorithme est-elle la meilleure?" (Ou l'une des meilleures, la meilleure réponse face à un compromis temps / résultat). En particulier, la notion de "bonne solution" varie selon le contexte d'application. Le second mode d'évaluation est la confrontation des classes fournies par la méthode à des groupes "naturels". On dira alors que l'on évalue le résultat selon un critère externe. Le score se calcule alors en termes de précision et de rappel, ou encore de F-mesure. Si plusieurs mesures hybrides ont été proposées pour opérer la classification dans des réseaux d'information, elles ne peuvent pas réellement être employées en tant que critères d'évaluation, en tout cas si elles sont également utilisées pour construire les classes car l'évaluation sera biaisée. Dans l'état de l'art figure cependant le critère de Joint Silhouette Coefficient proposé par Moser et al. et qui permet de quantifier la qualité d'une partition construite sur un réseau d'information en tenant compte des deux types de données : attributs et relations (Moser et al., 2007).[19] Cet indice, inspiré de l'indice de Silhouette de Rousseeuw (Rousseeuw, 1987), est défini par :

$$JSC(p) = \frac{1}{|V|} \sum_{v \in V} \frac{b_E(v) - a(v)}{\max(b_E(v), a(v))}$$

## 2.6. CONCLUSION

Le fait que l'analyse de réseau social se situe entre plusieurs domaines (so-l'informatique, les mathématiques et la physique) a conduit à de nombreux approches thodologiques.

C'est pourquoi tant de programmes ont été créés dans l'ordre manipuler et étudier les réseaux sociaux. Alors qu'un outil autonome est très utile pour la

## 2.6. CONCLUSION

---

visualisation graphique (jusqu'à un maximum de quelques milliers des nœuds), format de données conversion ou calcul d'indicateurs, les bibliothèques sont plus adaptées aux tâches impliquant des dizaines de milliers de nœuds et pour des opérations telles que l'union et la différence entre les ensembles de nœuds ou pour le regroupement.

Une juste séparation des algorithmes, des interfaces d'utilisateurs (y compris les capacités de visualisation) et la manipulation des données apparaît comme une tâche importante afin de rassembler et promouvoir les meilleures fonctionnalités parmi les outils. Par exemple, récemment, Gephi a commencé une telle approche avec la publication récente de la Gephi boîte à outils , une bibliothèque créée à partir de la logique et des algorithmes de Gephi. Nous pouvons également dire qu'aujourd'hui, les outils disponibles gratuitement sont en mesure de fournir un très riche ensemble de fonctionnalités, mais si une analyse spécifique est requise, des développements de code spécifiques peut être nécessaire. Ce qui concerne la nécessité de stocker de plus grands réseaux, de nouvelles architectures proposé notamment les bases de données NoSQL, comme Neo4j 13 qui peut être utilisé pour stocker des données structuré en graphiques plutôt qu'en tableaux. Enfin à ce moment-là, les principaux défis concernant l'exploration graphique sont orientés vers une visualisation de haut niveau notamment pour les graphes hiérarchiques.

Parmi les améliorations possibles des outils d'analyse de réseaux sociaux, nous pouvons mentionner tout d'abord extraction sociale qui exploite simultanément les attributs de nœud et la structure graphique et d'autre part l'analyse temporelle qui devrait nous permettre d'étudier l'évolution des réseaux au fil du temps.

# Chapitre 3

## LA DETECTION DE COMMUNAUTES PAR LES ALGORITHMES GÉNÉTIQUES

### Sommaire

---

3.1	La detection de communautes par les algorithmes génétiques	61
3.1.1	Introduction	61
3.1.2	Définition	61
3.1.3	Calcul naturel	62
3.1.4	Biologiquement Calcul Inspiré	63
3.1.5	Biologie Motivé informatiquement	63
3.1.6	Calcul avec biologie	63
3.1.7	Informatique intelligente	64
3.1.8	Processus de création d'un algorithme inspiré da La Nature	64
3.1.9	Application de l'algorithme (méta heuristiques)	65
3.2	Optimisation Combinatoire En mathématique	66

---

3.2.1	Minimum Local	66
3.2.2	Minimum Global	66
3.2.3	Voisinage	67
<b>3.3</b>	<b>Classification des Méta heuristiques</b>	<b>67</b>
<b>3.4</b>	<b>Notre approche Algorithme Génétique</b>	<b>68</b>
3.4.1	Introduction	68
3.4.2	Définition	69
3.4.3	Terminologie	69
<b>3.5</b>	<b>Fonctionnement générale</b>	<b>70</b>
3.5.1	Définition 01 (Gène)	70
3.5.2	Définition 02 (Chromosome)	71
3.5.3	Définition 03 (Individu)	71
3.5.4	Définition 04 (population)	71
3.5.5	Définition 05 (Génération)	71
3.5.6	Définition 06 (fonction de fitness ou fonction d'évaluation)	71
3.5.7	Fonctionnement général	72
<b>3.6</b>	<b>Algorithme de base AG</b>	<b>73</b>
<b>3.7</b>	<b>Principe de fonctionnement</b>	<b>74</b>
3.7.1	Codage	74
3.7.2	Population Initiale	76
3.7.3	Adaptation	76
<b>3.8</b>	<b>Les opérateurs d'un algorithme génétique</b>	<b>77</b>
3.8.1	la sélection	77
3.8.2	Le croisement	80
3.8.3	La Mutation	83
<b>3.9</b>	<b>Critère d'arrêt</b>	<b>84</b>
<b>3.10</b>	<b>Les algorithmes génétique pour la détection de communauté dans les réseaux sociaux</b>	<b>84</b>

---

3.10.1 Codage . . . . .	84
3.10.2 Population Initiale . . . . .	85
3.10.3 Adaptation (fonction de fitness) . . . . .	85
3.10.4 la sélection . . . . .	85
3.10.5 Le croisement . . . . .	86
3.10.6 La Mutation . . . . .	86
3.10.7 Critère d'arrêt . . . . .	87
<b>3.11 Conclusion . . . . .</b>	<b>87</b>

---

## 3.1. LA DETECTION DE COMMUNAUTES PAR LES ALGORITHMES GÉNÉTIQUES

### 3.1.1. INTRODUCTION

La nature est une source puissante d'inspiration pour résoudre des problèmes complexes en informatique, puisqu'elle montre des phénomènes extrêmement divers, dynamiques, robustes, complexes et fascinants. Elle trouve toujours la solution optimale pour résoudre son problème, et maintient l'équilibre parfait entre ses composantes. Nouvelle ère est ouverte avec les algorithmes inspirés de la nature (Bio-inspiré) qui sont des méta heuristiques imitant la nature pour résoudre des problèmes d'optimisation. Dans les décennies passées, de nombreux efforts de recherches ont été concentrés dans ce secteur particulier. Ce chapitre présente une vue d'ensemble sur les algorithmes inspirés de la nature.

### 3.1.2. DÉFINITION

On peut définir l'informatique bio-inspirée comme un système ou une architecture dont l'organisation est issue de la connaissance que l'on a du fonctionnement des systèmes vivants[25].

Il s'agit donc de comprendre des phénomènes biologiques (propriétés d'adaptation, d'auto-organisation...), de les modéliser afin d'utiliser cette modélisation pour mettre au point de nouvelles méthodologies de traitement de l'information utilisant les propriétés essentielles du vivant. L'objectif est de reproduire et utiliser les mécanismes observés dans le vivant dans différents domaines (informatique, recherche, mathématiques, robotique...). Il est intéressant d'observer, par exemple, le fonctionnement d'une société d'insectes afin de comprendre comment est mise en œuvre leur intelligence collective.

En effet, des sociétés composées d'individus "simples" (termites, colonie de fourmis, abeilles par exemple), peuvent ensemble résoudre des problèmes complexes en suivant des règles élémentaires. Les défis visés sont par exemple de mieux formaliser de nouvelles méthodes de calculs cellulaires, d'élaborer de nouveaux algorithmes d'optimisation, de développer des architectures bio-inspirées

### 3.1. LA DETECTION DE COMMUNAUTES PAR LES ALGORITHMES GÉNÉTIQUES

---

pour des machines autonomes... Un algorithme est un ensemble d'instructions à suivre qui définissent "ce qu'il faut faire" et "dans quel ordre" pour résoudre un problème donné (ou un ensemble de problèmes). Les algorithmes vont permettre de résoudre des problèmes dits "combinatoires", caractérisés par leur complexité. En effet, leur résolution nécessite l'exploration d'un nombre important de combinaisons.

En pratique, on ne pourra pas résoudre toutes les instances d'un problème complexe, certaines pouvant être résolues plus rapidement que d'autres. Il s'agit donc de trouver une solution de qualité suffisante en un temps raisonnable, l'objectif étant d'avoir un comportement acceptable en pratique. Nous sommes fréquemment heurtés à des problèmes complexes : gérer un emploi du temps, trouver le chemin le plus court (optimisation des transports), trouver les max, min ou zéros de fonctions... Nous allons présenter des algorithmes dits méta heuristiques visant à résoudre des problèmes d'optimisation complexes par une méthode d'itération qui progresse vers un maximum ou un minimum local comme l'extrémum d'une fonction. Cette façon de faire ne garantit pas que la solution sera optimale, ni même l'écart avec la valeur optimale. Cependant, le temps nécessaire pour obtenir cette solution sera beaucoup plus faible.

#### 3.1.3. CALCUL NATUREL

Une perspective importante sur débraillés Intelligence artificielle est la motivation et d'inspiration pour la stratégie de traitement de l'information de base d'une donnée technique. Les ordinateurs ne peuvent faire ce qu'ils sont instruits, donc considération est de distiller le traitement des informations provenant d'autres domaines d'études, comme le monde physique et la biologie. L'étude des biologiquement motivés le calcul est appelé Biologiquement Informatique Inspiré, et l'un des trois domaines connexes de l'informatique naturelle. Informatique naturelle est un domaine interdisciplinaire concerné par la relation de calcul et la biologie, qui en plus de Biologiquement Inspiré Informatique est également composé de biologie et de l'informatique plan informatique Motivé avec Biologie.

### 3.1. LA DETECTION DE COMMUNAUTES PAR LES ALGORITHMES GÉNÉTIQUES

---

#### 3.1.4. **BIOLOGIQUEMENT CALCUL INSPIRÉ**

Biologiquement Calcul Inspiré est inspiré par calcul biologique métaphore, aussi appelé bio mimétisme et Biomimétiques dans d'autres systèmes de disciplines. Le but de ce champ est de concevoir mathématique et des outils d'ingénierie pour générer des solutions aux problèmes de calcul. Le domaine implique l'utilisation de procédures pour trouver des solutions abstraites du monde naturel pour traiter les problèmes de calcul phrasés.

#### 3.1.5. **BIOLOGIE MOTIVÉ INFORMATIQUEMENT**

Informatiquement Biologie motivé consiste à enquêter sur la biologie à l'aide des ordinateurs. Le but de ce domaine est d'utiliser les sciences de l'information et simulation pour modéliser les systèmes biologiques dans les ordinateurs numériques dans le but de reproduire et de mieux comprendre les comportements dans les systèmes biologiques. Le champ facilite la capacité de mieux comprendre la vie comme-il-est et d'enquêter comme pourrait-être la vie. En règle générale, le travail dans ce sous-domaine ne concerne pas la construction d'outils mathématiques et d'ingénierie, il est plutôt axé sur la simulation des phénomènes naturels. Des exemples courants comprennent artificielle la vie, la géométrie fractale (L-systèmes, systèmes de fonction itérative, Particule Systèmes, mouvement brownien), et Automates cellulaires. Un domaine connexe est que de biologie computationnelle généralement concernés par la modélisation biologique des systèmes et l'application des méthodes statistiques telles que dans le sous-domaine de Bioinformatique.

#### 3.1.6. **CALCUL AVEC BIOLOGIE**

Calcul avec la biologie est l'étude des substrats autres que dans lequel le silicium pour mettre en œuvre le calcul. Des exemples courants comprennent Informatique moléculaire ou de l'ADN et de l'informatique quantique.

## 3.1. LA DETECTION DE COMMUNAUTES PAR LES ALGORITHMES GÉNÉTIQUES

---

### 3.1.7. INFORMATIQUE INTELLIGENTE

L'intelligence informatique est un nom moderne pour le sous-domaine de l'IA concerné avec des techniques sous-symbolique (aussi appelé désordre, débrillé, et doux).

L'intelligence informatique décrit des techniques axées sur la stratégie et le résultat. Le champ couvre largement sous-disciplines axés sur adaptatif et les systèmes d'intelligence, sans pas limiter : calcul évolutionniste, Essaim Intelligence (Optimisation des particules Essaim et Fourmie Colony), Fuzzy System, Artificiel immunitaire Systèmes et Réseaux de neurones artificiels.

### 3.1.8. PROCESSUS DE CRÉATION D'UN ALGORITHME INSPIRÉ DA LA NATURE

Tout d'abord, la nature donne l'inspiration aux chercheurs de développer une observation D'un phénomène naturel particulier. Premièrement, Les développeurs créent et testent un modèle, Utilisent des simulations mathématiques qui aident à raffiner le modèle original. Ensuite, le modèle raffiné sera utilisé pour extraire une méta heuristique qui peut être utilisée comme une base pour finalement concevoir un algorithme inspiré de la nature.

La Figure illustre le cadre utilisé pour passer d'un phénomène naturel à un algorithme inspiré de la nature.

### 3.1. LA DETECTION DE COMMUNAUTES PAR LES ALGORITHMES GÉNÉTIQUES

---

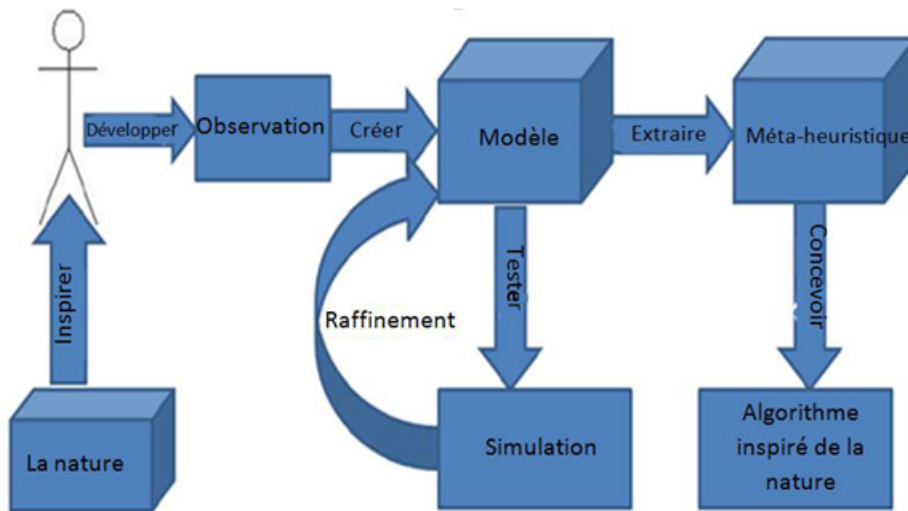


FIGURE 3.1 – Passage d'un phénomène naturel à un algorithme inspiré de la nature

#### 3.1.9. APPLICATION DE L'ALGORITHME (MÉTA HEURISTIQUES)

La recherche taboue est un méta heuristique de recherche locale très populaire, elle a connu un large succès pour résoudre divers problèmes d'optimisation combinatoires.

Le terme méta heuristique vient des mots grecs méta (au de la) et heuriskein (trouver)[26].

Il n'y a pas clairement de consensus sur la notion exacte des heuristiques et des métras heuristiques. Nous allons adopter Celles-ci :

Une heuristique est une technique de résolution spécialisée a un problème. Elle ne garantit pas la qualité de la solution obtenue.

Un méta heuristique est une heuristique générique qu'il faut adapter à chaque problème. Il ne faut pas confondre au méta heuristique et math euristique.

Une math euristique est une méthode d'optimisation combinant des techniques métras heuristiques a des algorithmes " de programmation mathématique (PM).

On peut ainsi avoir des métras heuristiques déniées ou améliorées par des techniques de PM, ou des métras heuristiques accélérant des algorithmes de PM.

Exemple : Utiliser un méta heuristique pour générer des colonnes en généra-

## 3.2. OPTIMISATION COMBINATOIRE EN MATHÉMATIQUE

---

tion de colonnes.

Un méta heuristique : est une méthode algorithmique capable de guider et d'orienter le processus de recherche dans un espace de solution, souvent très grand à des régions riches en solutions optimales.

Le fait de rendre cette méthode abstraite et plus générique conduite à une utilisation vaste pour des champs d'applications différents.

A ces applications, les méta heuristiques permettent, de trouver des solutions, peut être pas toujours optimales, en tout cas très proches de l'optimum et en un temps raisonnable.

### 3.2. OPTIMISATION COMBINATOIRE EN MATHÉMATIQUE

L'optimisation combinatoire recouvre toutes les méthodes qui permettent de déterminer l'optimum d'une fonction avec ou sans contraintes.

Soit  $S$  un ensemble de solutions à un problème d'optimisation et  $f$  une fonction objective qui mesure la valeur  $f(s)$  avec  $s \in S$ .

Pour un problème de minimisation on cherche à déterminer une solutions qui minimise la fonction objective.

Un minimum est une solution qui fait partie des solutions réalisable, dans le domaine d'optimisation on distingue deux types de minimums :

#### 3.2.1. MINIMUM LOCAL

une solutions est minimum local par apport à une structure de Voisinage  $N$  si pour tout  $s \in N(s)$ ,  $f(s) \leq f(s')$ .

#### 3.2.2. MINIMUM GLOBAL

une solutions est minimum global si pour tout  $s' \in S$ ,  $f(s) \leq f(s')$ .

### 3.3. CLASSIFICATION DES MÉTA HEURISTIQUES

---

#### 3.2.3. VOISINAGE

le voisinage est une fonction notée  $N$  qui associe un sous ensemble de  $S$  à toute solutions, les voisins des sont  $s \in N(s)$ .

### 3.3. CLASSIFICATION DES MÉTA HEURISTIQUES

Les problèmes d'optimisation combinatoire sont souvent des problèmes très difficiles dont la résolution par des méthodes exactes peut s'avérer très longue ou peu réaliste. L'utilisation de méthodes heuristiques permet d'obtenir des solutions de bonne qualité en un temps de résolution raisonnable.

Les Heuristiques sont aussi très utiles pour le développement de méthode des exactes fondées sur des techniques d'évaluation et de séparation (Branch And Bound).

Une Heuristique est un algorithme qui a pour but de trouver une solution réalisable, sans garantie d'optimalité, contrairement aux méthodes exactes qui garantissent des solutions exactes.

Comme Les algorithmes de résolution exacte sont de complexité exponentielle pour les problèmes difficiles, il peut être plus judicieux de faire appel aux heuristiques pour calculer une solution approchée d'un problème ou aussi pour accélérer le processus de résolution exacte. Généralement une heuristique est conçue pour un problème particulier, mais les approches peuvent contenir des principes plus généraux, On parle de méta heuristique.

## 3.4. NOTRE APPROCHE ALGORITHME GÉNÉTIQUE

---

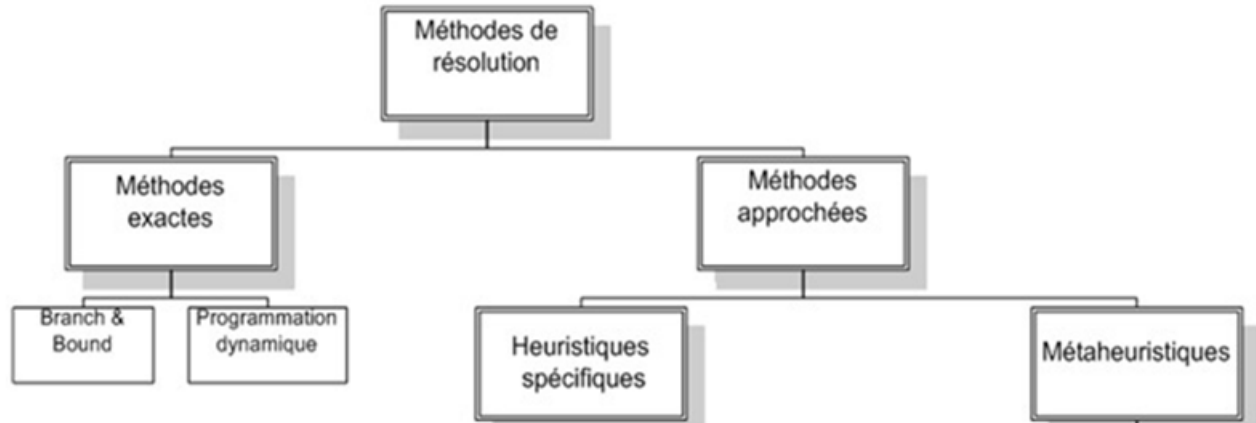


FIGURE 3.2 – les Méthode d'optimisation combinatoire

## 3.4. NOTRE APPROCHE ALGORITHME GÉNÉTIQUE

### 3.4.1. INTRODUCTION

Les algorithmes génétiques s'inspirent de l'évolution darwinienne des populations biologiques, pour définir leurs mécanismes. Les principes de cette évolution sont tels que, dans une population d'individus ce sont les plus forts, c'est à dire les mieux adaptés au milieu, qui survivront et pourront donner une descendance ou une bonne progéniture.

C'est en 1950 que les algorithmes génétiques ont vu le jour par des biologistes utilisant des ordinateurs pour simuler l'évolution des organismes. Vers 1960, John Holland et son équipe adaptaient ces algorithmes pour la recherche de solutions a des problèmes d'optimisation , en développant une analogie entre un individu dans une population et une solution d'un problème dans un ensemble de solutions.

## 3.4. NOTRE APPROCHE ALGORITHME GÉNÉTIQUE

---

### 3.4.2. DÉFINITION

Les algorithmes génétiques (AGs) sont des algorithmes d'optimisation stochastique fondés sur les mécanismes de la sélection naturelle et de la génétique [27].

On part avec une population de solutions potentielles (chromosomes) initiales arbitrairement choisies.

On évalue leur performance (fitness) relative.

Sur la base de ces performances on crée une nouvelle population de solutions potentielles en utilisant des opérateurs évolutionnaires : la sélection, le croisement et la mutation.

On recommence ce cycle jusqu'à ce que l'on trouve une solution satisfaisante.

### 3.4.3. TERMINOLOGIE

F Les chromosomes sont des chaînes d'ADN.

F L'élément de base des chromosomes est un gène.

F La position d'un gène sur le chromosome est son locus.

F L'ensemble des gènes d'un individu est son génotype.

F l'ensemble du patrimoine génétique d'une espèce est le génome. Les différentes versions d'un même gène sont appelées allèles.

### 3.5. FONCTIONNEMENT GÉNÉRALE

---

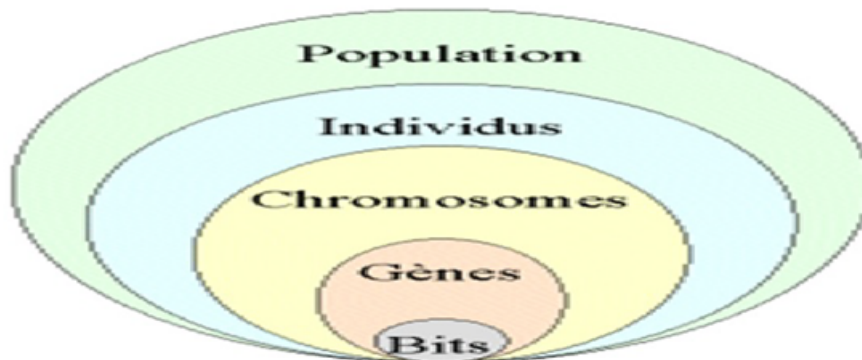


FIGURE 3.3 – les 5 niveaux d'organisation de notre algorithme Génétique

### 3.5. FONCTIONNEMENT GÉNÉRALE

Avant d'expliquer en détail le fonctionnement d'un algorithme génétique, nous allons présenter quelques mots relatifs à la génétique.

Le vocabulaire utilisé est le même que celui de la théorie de l'évolution et de la génétique, on emploie le terme individu (solution potentielle), population (ensemble de solutions), génotype (une représentation de la solution), gène (une partie du génotype), parent, enfant, reproduction, croisement, mutation, génération, etc.

#### 3.5.1. DÉFINITION 01 (GÈNE)

Un gène est une suite de bases azotées (adénine (A), cytosine (C), guanine (G), et la thymine (T)) qui contient le code d'une protéine.

On appellera gène la suite de symboles qui codent la valeur d'une variable.

Dans le cas général, un gène correspond à un seul symbole (0 ou 1 dans le cas binaire).

Une mutation changera donc systématiquement l'expression du gène muté.

### 3.5. FONCTIONNEMENT GÉNÉRALE

---

#### 3.5.2. DÉFINITION 02 (CHROMOSOME)

Un chromosome est constitué d'une séquence finie de gènes qui peuvent prendre des valeurs appelées allèles qui sont prise dans un alphabet qui doit être judicieusement choisi pour convenir de problème étudié.

#### 3.5.3. DÉFINITION 03 (INDIVIDU)

On appellera individu une des solutions potentielles. Dans la plupart des cas un individu sera représenté par un seul chromosome

#### 3.5.4. DÉFINITION 04 (POPULATION)

On appellera population l'ensemble des solutions potentielles qui 'utilise L'AG.

#### 3.5.5. DÉFINITION 05 (GÉNÉRATION)

On appellera génération l'ensemble des opérations qui permettent de passer d'une population  $P_i$  à une population  $P_j$ .

Ces opérations seront généralement : sélection des individus de la population courante, application des opérateurs génétiques, évaluation des individus de la nouvelle population.

#### 3.5.6. DÉFINITION 06 (FONCTION DE FITNESS OU FONCTION D'ÉVALUATION)

Est la pièce maitresse dans la processus de optimisation.

C'est l'élément qui permet aux algorithmes génétiques de prendre en compte un problème donné. Pour que le processus d'optimisation puisse donner de bons résultats, il faut concevoir une fonction de fitness permettent une évaluation pertinente des solutions d'un problème sous forme chiffrée. Cette fonction est déterminée en fonction du problème à résoudre et du codage choisi pour les chromosomes.

### 3.5. FONCTIONNEMENT GÉNÉRALE

---

Pour chaque chromosome, elle attribue une valeur numérique, qui est supposée proportionnelle à la qualité de l'individu en tant que solution.

Le résultat renvoyé par la fonction d'évaluation va permettre de sélectionner ou de refuser un individu selon une stratégie de sélection.

#### 3.5.7. FONCTIONNEMENT GÉNÉRAL

Leur fonctionnement est extrêmement simple, on part d'une population de solutions potentielles (chromosomes) initiales, arbitrairement choisies.

On évalue leur performance (Fitness) relative.

Sur la base de ces performances on crée une nouvelle population de solutions potentielles en utilisant des opérateurs évolutionnaires simples : la sélection, le croisement et la mutation.

Quelques individus se reproduisent, d'autres disparaissent et seuls les individus les mieux adaptés sont supposés survivre.

On recommence ce cycle jusqu'à ce qu'on trouve une solution satisfaisante.

En effet, l'héritage génétique à travers les générations permet à la population d'être adaptée et donc répondre au critère d'optimisation, la figure N3.4 illustre les principales étapes d'un algorithme génétique.

Un algorithme génétique recherche le ou les extrema d'une fonction définie sur un espace de données.

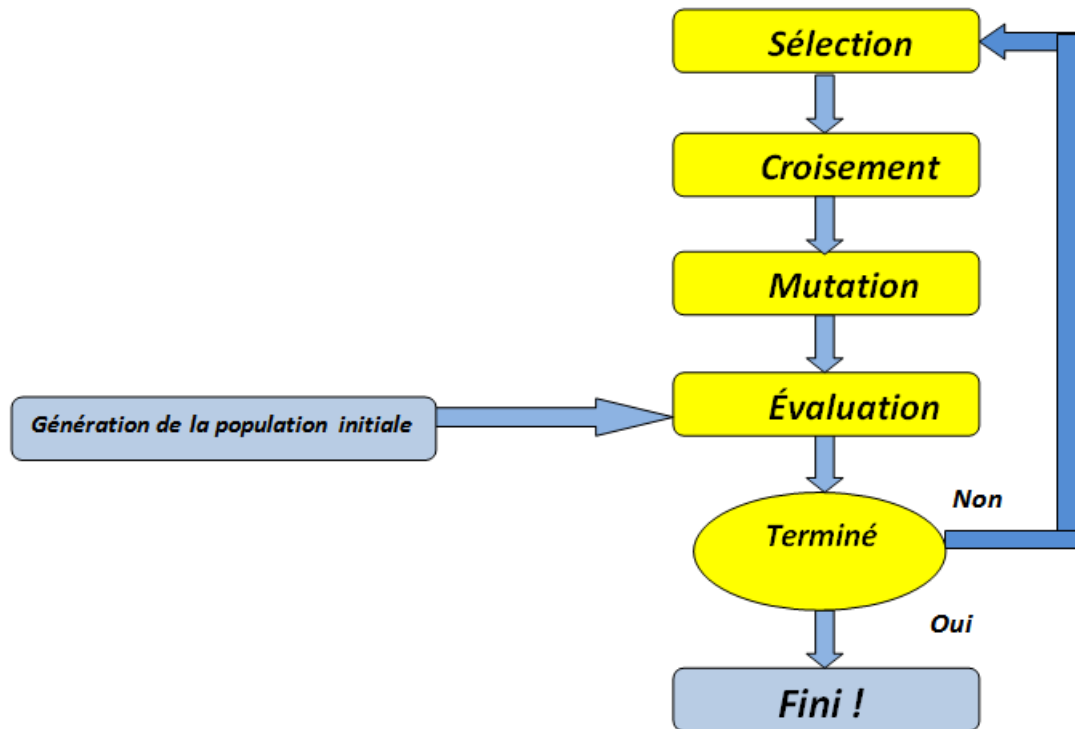


FIGURE 3.4 – fonctionnement de AG

### 3.6. ALGORITHME DE BASE AG

Début

- Générer une population initiale de n chromosome.
- Evaluer la fitness des chromosomes avec la fonction  $f(x)$
- Répéter
- Calculer la fonction fitness  $f(x)$ , pour tout chromosome  $x$
- Appliquer l'opération de sélection
- Appliquer l'opération de croisement avec une probabilité PC
- Appliquer l'opération de mutation avec une probabilité de PM

### 3.7. PRINCIPE DE FONCTIONNEMENT

---

- Ajouté les nouveaux chromosomes à la nouvelle population
- Calculer la fonction de fitness  $f(x)$  ,pour tous chromosomes  $x$
- Appliquer l'opération de remplacement
- Jusqu'à la satisfaction des conditions de terminaison
- Fin

## 3.7. PRINCIPE DE FONCTIONNEMENT

### 3.7.1. CODAGE

La première étape est de définir et coder convenablement le problème.

On retrouve différents techniques de codages

Le codage est un processus de représentation des gènes.

Le processus peut être effectué par utilisation des : bits, nombres, arbres, tableaux, listes ou tous autres objets.

La littérature définit deux types de codage : binaire et réel.

#### 3.7.1.1. CODAGE BINAIRE

C'est la représentation la plus fréquente. Chaque gène dispose du même alphabet binaire 0, 1 Un gène est alors représenté par Un entier long (32 bits), les chromosomes qui sont des suites des gènes sont représentés par des tableaux de gènes et les individus de notre espace de recherche sont représentés par des tableaux de chromosomes.

### 3.7. PRINCIPE DE FONCTIONNEMENT

---

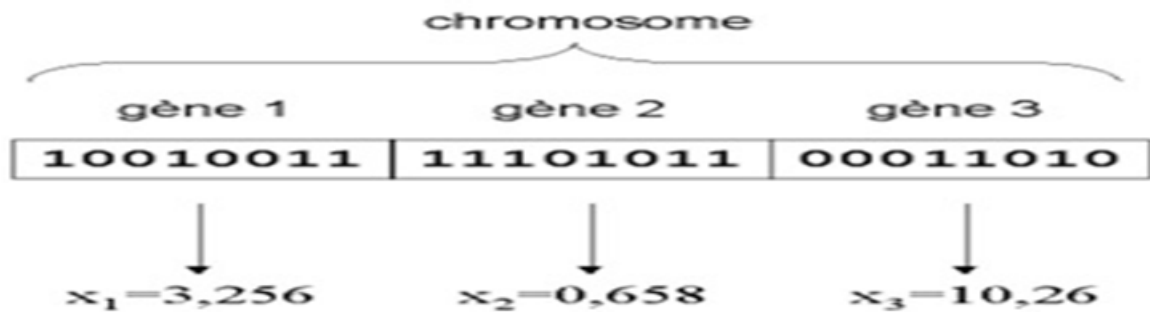


FIGURE 3.5

#### 3.7.1.2. CODAGE RÉEL

Une autre approche de codage pour coder les solutions en tant que une suite de nombres entiers ou de nombres réels, avec chaque position représentant encore un certain aspect particulier de la solution .

Ce type de codage présente certains avantages par apport au codage binaire :

Le codage réel est robuste pour les problèmes considérés comme un codage difficile pour le codage binaire.

Ce codage nécessite une adaptation des opérateurs de croisement et mutation.

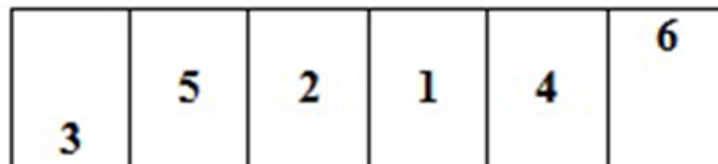


FIGURE 3.6

Tableau1 :

#### 3.7.1.3. CODAGE ALPHABOTIQUE

Une troisième approche est de représenter des individus dans AG comme une suite de caractère, ou chaque caractère représente un aspect spécifique de

### 3.7. PRINCIPE DE FONCTIONNEMENT

---

la solution.

Ce type de codage est utilisé dans de nombreux cas poussés d'algorithme génétique comme en bioinformatique.



FIGURE 3.7 – codage alphabétique d'un chromosome

#### 3.7.2. POPULATION INITIALE

Plusieurs mécanismes de génération de la population initiale sont utilisés dans la littérature.

Le choix de l'initialisation se fait en fonction des connaissances que l'utilisateur a sur le problème.

S'il n'a pas d'informations particulières, alors une initialisation aléatoire, la plus uniforme possible afin de favoriser une exploration de l'espace de recherche maximum, sera la plus adaptée.

Par ailleurs, cette étape présente un problème principal qui est celui de choix de la taille de la population.

En effet une taille de population trop grande augmente le temps de calcul et nécessite un espace mémoire considérable, alors qu'une taille de population trop petite conduit à l'obtention d'un optimum local.

#### 3.7.3. ADAPTATION

La fonction d'adaptation, ou fitness, associe une valeur pour chaque individu.

Cette valeur a pour but d'évaluer si un individu est mieux adapté qu'un autre à son environnement. Ce qui signifie qu'elle quantifie la réponse fournit au

## 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

problème pour une solution potentielle données.

Ainsi les individus peuvent être comparés entre eux.

Cette fonction, propre au problème, est souvent simple à formuler lorsqu'il ya peu de paramètres.

Au contraire, lorsqu'il ya beaucoup de paramètres ou lorsqu'ils sont corrèles, elle est plus difficile à définir.

Dans ce cas, la fonction devient une somme pondérée de plusieurs fonctions. Un ajustement des coefficients est nécessaire.

## 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

Quatre opérations caractérisent les algorithmes génétiques rappellent l'origine de ces méthodes.

Ils vont permettre à la population d'évoluer, par la création d'individus nouveaux construits à l'aide des individus anciens : la sélection et le croisement, la mutation et le remplacement.

### 3.8.1. LA SÉLECTION

La sélection permet d'identifier statistiquement les meilleurs individus de la population courante qui seront à se reproduire.

Cette opération est fondée la performance des individus, estimée à l'aide de la fonction d'adaptation. Il existe différents principes de sélection :

#### 3.8.1.1. LA SÉLECTION PAR ROULETTE

La sélection des individus par la méthode de la roulette s'inspire de la roue de loterie sur laquelle chaque individu est représenté par un secteur proportionnel à sa fitness.

On fait tourner la roue et on sélectionne un individu.

### 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

Les individus les mieux évalués ont statistiquement plus de chance d'être sélectionnés, mais donnent aussi une possibilité aux individus mal adaptés d'être choisis.

à chaque individu  $i$  une probabilité est associée, d'être choisi proportionnelle à son adaptation  $f_i$  :

$$prob(i) = \frac{f_i}{\sum f_j} \text{ où } \sum f_j \text{ désigne la somme des adaptations de la population.}$$

#### 3.8.1.2. SÉLECTION PAR RANG

La sélection par rang trie d'abord la population par fitness.

Chaque individu se voit associé un rang en fonction de sa position.

Rang est calculée selon le total des adaptations des individus divisé par la taille de la population, autrement  $\text{rang} = P \cdot \sum f_i/n$ , alors le plus mauvais individu aura le premier rang, le suivant le deuxième rang, ainsi de suite.

La sélection par rang d'un individu est la même que par roulette, mais les proportions sont en relation avec le rang plutôt qu'avec la valeur de l'évaluation.

Avec cette méthode de sélection, tous les individus ont une chance d'être sélectionnés. Cependant, elle conduit à une convergence plus lente vers la bonne solution.

### 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

Individu	Séquence	Adaptation	% du total
1	01011	11	18.6
2	10011	19	32.2
3	00101	5	8.5
4	11000	24	40.7
total		59	100

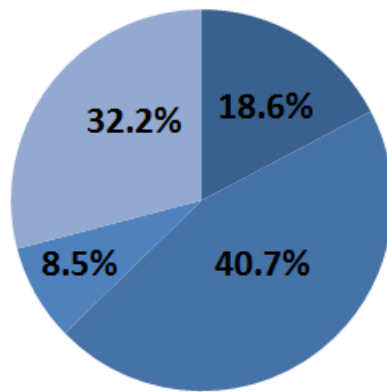


FIGURE 3.8

#### 3.8.1.3. SÉLECTION PAR ELITISME

Cette méthode consiste à sélectionner les  $n$  individus dont on a besoin ,pour la nouvelle génération  $P'$  en prenant les  $n$  meilleurs individus de la population  $P$  après l'avoir triée de manière décroissante selon la fonction d'adaptation (fitness) de ses individus.

Il est inutile de préciser que cette méthode est encore pire que celle de la loterie biaisée dans le sens où elle amènera à une convergence prématurée encore plus rapidement et surtout de manière encore plus sûre que la méthode de sélection de la loterie biaisée.

En effet, la pression de la sélection est trop forte, la variance nulle et la diversité inexistante.

Une sélection par élitisme après avoir trié la population selon la fonction d'adaptation permet de nous donner la population suivante :

### 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

Ind.(xj)	Code
1	11000
2	10011
3	11000
4	11000

FIGURE 3.9

#### 3.8.2. LE CROISEMENT

La naissance d'un nouvel individu, nécessite la prise aléatoire d'une partie des gènes de chacun des deux parents.

Ce phénomène, issu de la nature est appelé croisement (crossover).

Le croisement est le processus de prendre deux parents et de produire à partir d'elles des enfants.

Il s'agit d'un processus essentiel pour explorer l'espace des solutions possibles.

La littérature définit plusieurs opérateurs de croisement.

Ils diffèrent selon le type de codage adapté et la nature du problème traité.

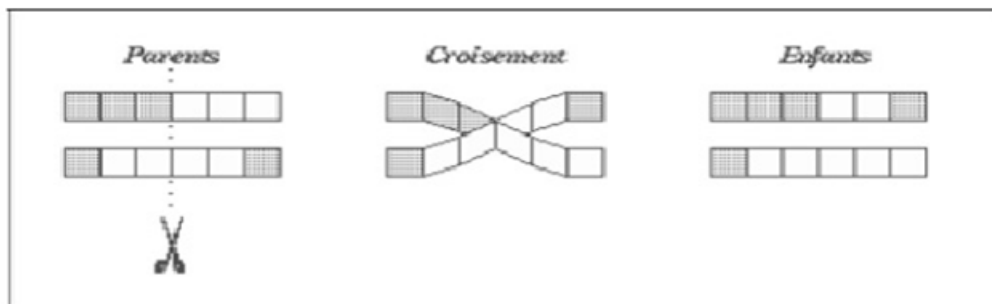


FIGURE 3.10

## 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

### 3.8.2.1. LE CROISEMENT BINAIRE

croisement en un point : on choisit au hasard un point de croisement, pour chaque couple (Figure 3.8).

Notons que le croisement s'effectue directement au niveau binaire, et non pas au niveau des gènes.

Un chromosome peut donc être coupé au milieu d'un gène.

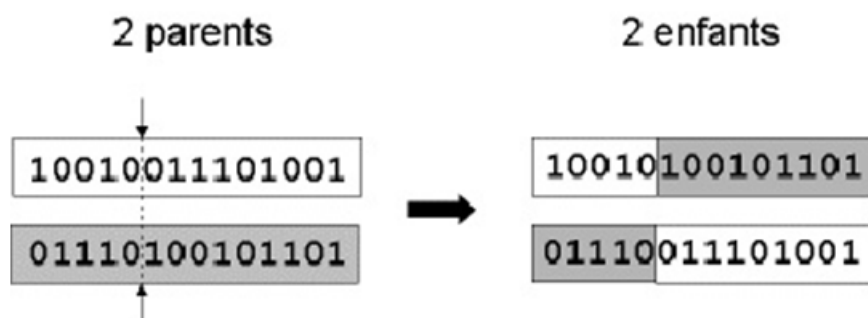


FIGURE 3.11

croisement en deux points : On choisit au hasard deux points de croisement (Figure 3.9).

Par la suite, nous avons utilisé cet opérateur car il est généralement considéré comme plus efficace que le précédent [Beasley, 1993b].

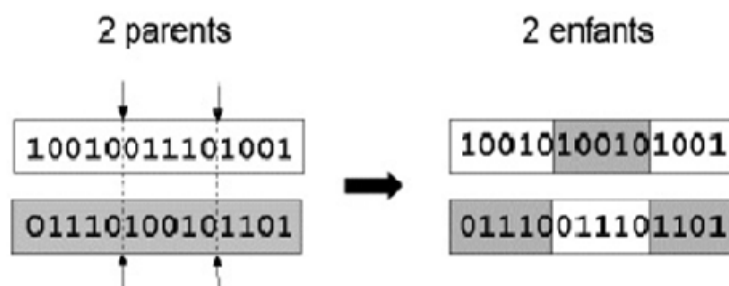


FIGURE 3.12

### 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

Croisement en n-points : Ce type de croisement s'énonce par un choix aléatoirement de n-points de coupure pour dissocier chaque parent en n+1 fragments.

Pour former un fils, il suffit de concaténer alternativement n+1 sous chaînes à partir des deux parents.

#### 3.8.2.2. CROISEMENT UNIFORME

Il existe des versions distinctes de ce croisement.

La plus connue consiste à définir de manière aléatoirement un masque, c'est-à-dire une chaîne de bits d'en même longueur que les chromosomes des parents sur lesquels il sera appliqué.

Ce masque est destiné pour savoir, pour chaque locus, quel est le parent de premier fils qui devra hériter du gène s'y trouvant ; si face à un locus le masque présente un 1, le fils héritera le gène s'y trouvant du parent1, s'il présente un 0 il en héritera du parent2.

La création du fils2 se fait de manière symétrique : si pour un gène donné le masque indique que le fils1 devra recevoir celui-ci du parent1 alors le fils2 le recevra du parent2, et si le fils1 le reçoit du parent2 alors le fils2 le recevra du parent1, (Figure3.13).

<b>Parent1</b>	<b>110110011001</b>
<b>Parent2</b>	<b>010011110011</b>
<b>masque</b>	<b>110011001000</b>

<b>Fils1</b>	<b>110010111011</b>
<b>Fils2</b>	<b>010111010010</b>

FIGURE 3.13

## 3.8. LES OPÉRATEURS D'UN ALGORITHME GÉNÉTIQUE

---

### 3.8.2.3. CROISEMENT RÉEL

Le croisement réel ne se différencie du croisement binaire que par la nature des éléments qu'il altère : ce ne sont plus des bits qui sont échangés à droite du point de croisement, mais des variables réelles.

### 3.8.3. LA MUTATION

C'est un processus où un changement mineur de code génétique est appliqué à un individu pour introduire de la diversité et ainsi d'éviter de tomber dans des optimums locaux. Différentes manières de mutation d'un chromosome sont aussi définies dans la littérature.

#### 3.8.3.1. MUTATION BINAIRE

Dans un algorithme génétique simple, la mutation en codage binaire est la modification aléatoire occasionnelle (de faible probabilité) de la valeur d'un caractère de la chaîne.

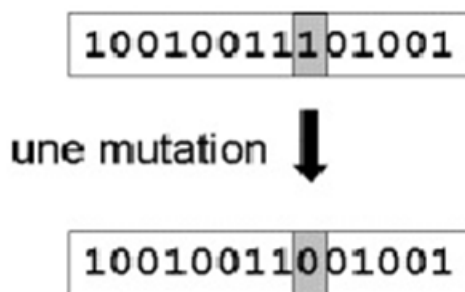


FIGURE 3.14 – représentation schématique d'une mutation dans un chromosome

## 3.9. CRITÈRE D'ARRÊT

---

### 3.8.3.2. MUTATION RÉELLE

La mutation réelle ne se différencie de la mutation binaire que par la nature de l'élément qu'elle altère : ce n'est plus un bit qui est inversé, mais une variable réelle qui est de nouveau tirée au hasard sur son intervalle de définition.

## 3.9. CRITÈRE D'ARRÊT

Le critère d'arrêt indique que la solution est suffisamment approchée de l'optimum.

Plusieurs critères d'arrêt de l'algorithme génétique sont possibles.

On peut arrêter l'algorithme après un nombre de générations suffisant pour que l'espace de recherche soit convenablement exploré.

Ce critère peut s'avérer coûteux en temps de calcul si le nombre d'individus à traiter dans chaque population est important.

On peut aussi envisager d'arrêter l'algorithme lorsque la fonction d'adaptation d'un individu dépasse un seuil fixé au départ.

## 3.10. LES ALGORITHMES GÉNÉTIQUE POUR LA DÉTECTION DE COMMUNAUTÉ DANS LES RÉSEAUX SOCIAUX

### 3.10.1. CODAGE

On a utilisé la méthode string-based

### 3.10. LES ALGORITHMES GÉNÉTIQUE POUR LA DÉTECTION DE COMMUNAUTÉ DANS LES RÉSEAUX SOCIAUX

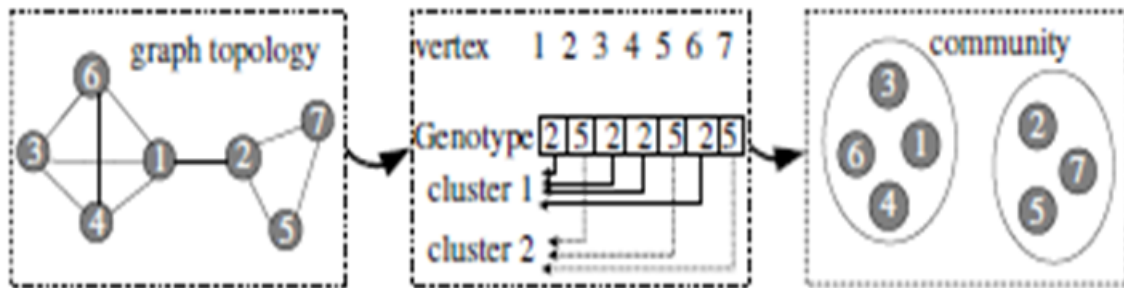


FIGURE 3.15

Le schéma basé sur une chaîne est plus facile à coder et décoder et il peut fournir une grande commodité à la diversité stratégies de promotion. Il convient de souligner que lorsque l'initialisation des individus, on peut utiliser pleinement le réseau connaissances préalables pour accélérer la convergence.

#### 3.10.2. POPULATION INITIALE

Dans notre application on a utilisé l'initialisation aléatoire qu'on a déjà défini dans ce chapitre .

#### 3.10.3. ADAPTATION (FONCTION DE FITNESS)

Dans notre application on a choisi comme fonction de fitness la fonction de la modularité (Q)

$$Q = \frac{1}{2 * m} \sum_{ij} (Adj_{(i,j)} - \frac{k_i * k_j}{2 * m}) * \delta(c_i, c_j)$$

Cette fonction est plus détaillée dans le deuxième chapitre .

#### 3.10.4. LA SÉLECTION

on a utilisés trois types de sélection :

Sélection par roulette, Sélection par rang et la Sélection par Elitisme Sachant que le type de selection peut changé selon notre choix .

### 3.10. LES ALGORITHMES GÉNÉTIQUE POUR LA DÉTECTION DE COMMUNAUTÉ DANS LES RÉSEAUX SOCIAUX

---

(qu'on a déjà défini ces types dans ce chapitre).

#### 3.10.5. LE CROISEMENT

##### 3.10.5.1. LE CROISEMENT UTILISÉ C'EST LE CROISEMENT EN UN POINT

on choisit au hasard un point de croisement, pour chaque couple (déjà défini).

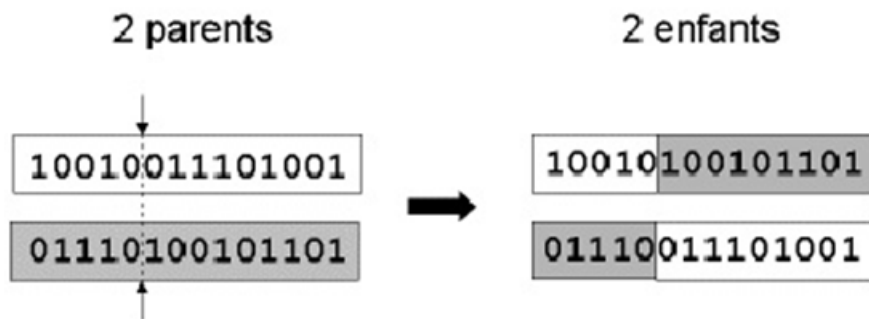


FIGURE 3.16

#### 3.10.6. LA MUTATION

La mutation qu'on a utilisé c'est la mutation binaire qui on a déjà la définie précédemment

### 3.11. CONCLUSION

---

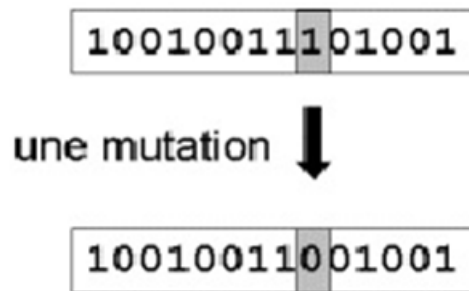


FIGURE 3.17 – représentation schématique d'une mutation dans un chromosome.

#### 3.10.7. CRITÈRE D'ARRÊT

L'algorithme termine son exécution selon le nombre d'itérations sélectionnée .

### 3.11. CONCLUSION

Dans ce chapitre, nous avons dans un premier temps défini des notions de base sur les méthodes bio inspirés et métaheuristiques, et aussi présenter l'approche utilisée pour l'application qui est l'approche algorithme génétique.

Finalement nous avons défini l'algorithme de notre approche algorithme génétique pour la détection de communauté et ces différentes étapes.

# Chapitre 4

## IMPLÉMENTATION ET RÉSULTATS

### Sommaire

---

4.1	Introduction	90
4.2	Environnement de développement	90
4.2.1	JAVA	90
4.2.2	eclipse	91
4.3	Présentation des réseaux de teste	92
4.3.1	Club de karaté de Zachary	92
4.3.2	Les dauphins de Lusseau	92
4.3.3	Les livres politiques	93
4.3.4	Football américain	94
4.3.5	Présentation de l'application	94
4.4	Discussion des résultats	99
4.4.1	Evaluation	99
4.5	Comparaison des résultats et discussion	100

---

4.5.1	Club de Zachary	100
4.5.2	dauphins de Lusseau	101
4.5.3	Livres politiques	101
4.5.4	Football américain	101
<b>4.6</b>	<b>Conclusion</b>	<b>102</b>

---

## 4.1. INTRODUCTION

Ce chapitre est essentiellement consacré aux grandes lignes qui visent à réaliser l'objectif de ce thème, et les outils exploités pour le développement du logiciel tel que le choix du langage de programmation, l'environnement de programmation, le matériel utilisé et les principales fonction de traitement au utiliser, et de présenter aussi une évaluation empirique de la méthode proposée de détection de communautés sur des réseaux réels. La performance de notre méthode est comparée à quatre algorithmes connus dans la littérature, soit Walktrap, Edge Betweenness, Fast Greedy et Label Propagation.

## 4.2. ENVIRONNEMENT DE DÉVELOPPEMENT

Notre application est développée en Java, avec outil de développement Eclipse oxygène.

### 4.2.1. JAVA

Le langage Java est un langage de programmation informatique Orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982), présenté officiellement le 23 mai 1995 au SunWorld.[28]

La société Sun a été ensuite rachetée en 2009 par la société Oracle qui détient et maintient désormais Java.

La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications. Pour cela, divers plateformes et frameworks associés visent à guider, sinon garantir, cette portabilité des applications développées en Java.



FIGURE 4.1 – *Logo de Java*

#### **4.2.2. ECLIPSE**

Eclipse est l'un des IDE (Integrated Development Environment) les plus utilisés pour le développement d'applications Java[29]. Integrated Development Environment ou IDE fournit de nombreuses fonctionnalités telles que la création, la modification, la compilation, le déploiement et le débogage de logiciels.

Eclipse n'est pas seulement utilisé pour développer une application Java. Il peut être utilisé pour développer des applications dans différents langages de programmation, notamment Ada, C, C ++, COBOL, Java, Perl, PHP, Python, R, Ruby (y compris le framework Ruby on Rails), Scala, Clojure, Groovy et Scheme. Il peut également être utilisé pour développer des packages pour les logiciels mathématiques.

Les environnements de développement incluent les outils de développement Java Eclipse (JDT) pour Java, Eclipse CDT pour C / C ++ et Eclipse PDT pour PHP, entre autres.

Beaucoup de plug-ins disponibles pour Eclipse afin que vous puissiez étendre ses capacités autant que vous le souhaitez.



FIGURE 4.2 – *eclipse oxygen*

## 4.3. PRÉSENTATION DES RÉSEAUX DE TESTE

### 4.3.1. CLUB DE KARATÉ DE ZACHARY

C'est un réseau construit à partir des relations entre 34 membres d'un club de karaté dans une université aux Etats- Unis. Il s'agit d'un réseau très populaire par une ligne verticale.

Nombre de nœuds : 34

Nombre de liens : 78

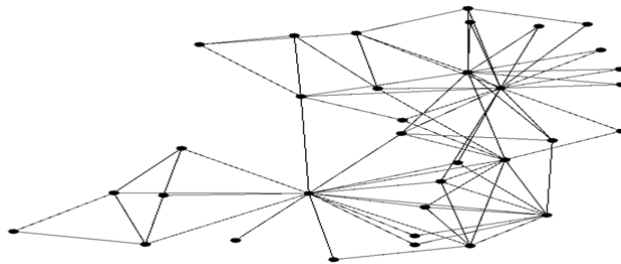


FIGURE 4.3 – *Visualisation de réseau Club de karaté de Zachary*

### 4.3.2. LES DAUPHINS DE LUSSEAU

Ce réseau contient 62 nœuds et 159 liens.

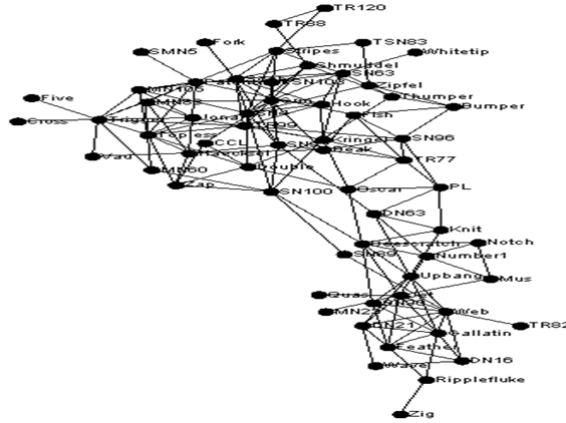


FIGURE 4.4 – Visualisation de reseau Les dauphins de Lusseau

### 4.3.3. LES LIVRES POLITIQUES

Ce réseau est constitué de 105 livres avec 441 liens se trouvant entre les livres achetés ensemble.

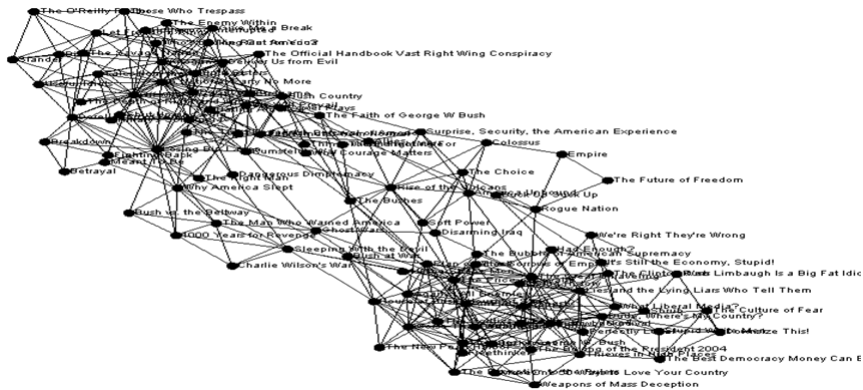


FIGURE 4.5 – Visualisation de reseau Les livres politiques

#### 4.3.4. FOOTBALL AMÉRICAIN

Le réseau de jeux de football américain (American football games) représente le calendrier des matchs entre des équipes américaines de football durant l'année 2 000. Ce réseau est constitué de, 115 nœuds et 613 liens.

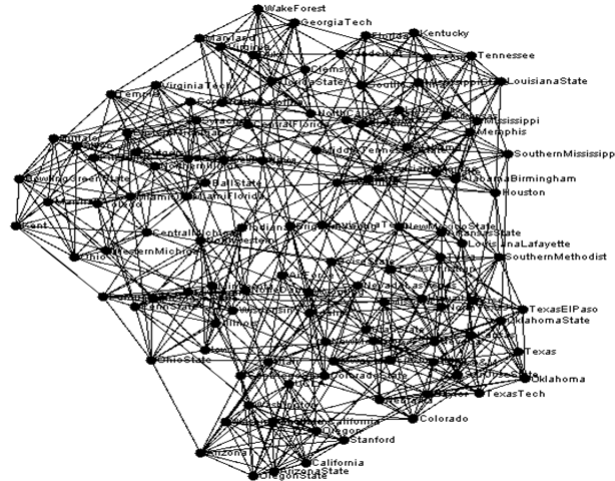


FIGURE 4.6 – Visualisation de réseau Football américain

#### 4.3.5. PRÉSENTATION DE L'APPLICATION



FIGURE 4.7 – capture d'écran <interface initiale>.

Cette image représente notre interface initiale, que nous la voir quand en clique sur run en eclipse.

- La touche num 1 : pour fermer l'interface.
- La touche num 2 : pour niveau inferieure.
- La touche num 3 : pour réduire.
- La touche num 4 : pour quitter l'application.
- La touche num 5 : pour entrer dans l'application.

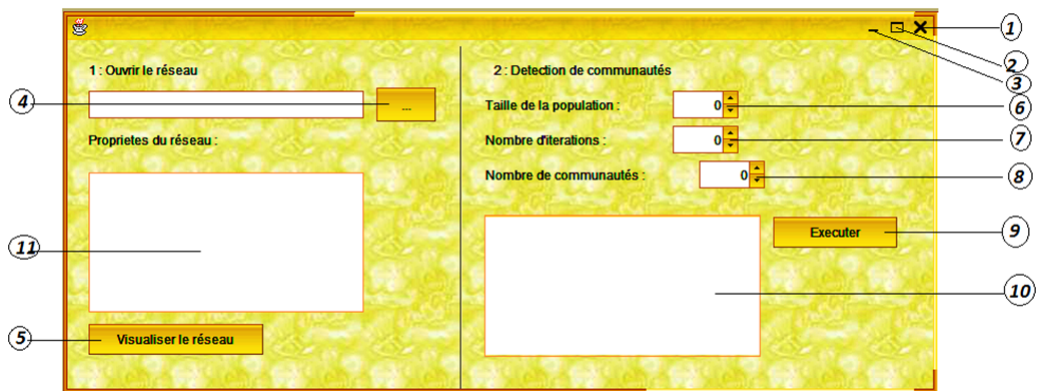


FIGURE 4.8 – capture d'écran<interface d'application>.

Cette capture d'image représente notre interface essentielle pour notre application :

- Les touches 1,2,3 : ont les mêmes fonctionnalités avec les touches de l'image précédente.
- La touche num 4 : pour choisir le réseau et le charger .
- La touche num 5 : pour visualiser le réseau choisi .
- La touche num 6 : pour définir la taille de population .
- La touche num 7 : pour définir le nombre d'itérations .
- La touche num 8 : pour définir le nombre des communautés .
- La touche num 9 : pour lancer l'exécution .
- L'espace num 10 : pour l'affichage des résultats .
- L'espace num 11 : pour l'affichage des propriétés de réseau choisi.

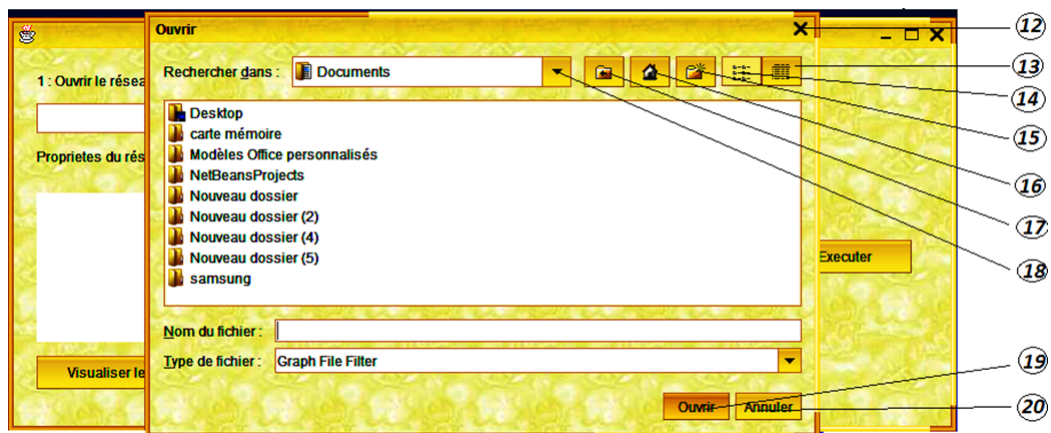


FIGURE 4.9 – capture d'écran <fenêtre de choix de réseau >.

#### Comparaison des résultats et discussion

- La touche num 12 : pour quitter la fenêtre .
- La touche num 13,14 : pour changer l'affichage de la liste des dossiers .
- La touche num 15 : pour créer un nouveau dossier .
- La touche num 16 : pour afficher les dossiers qui ont au bureau .
- La touche num 17 : pour afficher tout les dossiers.
- La touche num 18 : pour sélectionner le chemin de fichier de réseau quand veut choisir .
- La touche num 19 : pour ouvrir et charger le fichier choisi .
- La touche num 20 : pour annuler .



FIGURE 4.10 – capture d'écran <fenêtre de choix de réseau après la sélection de dossier qui contient la liste des réseaux >.

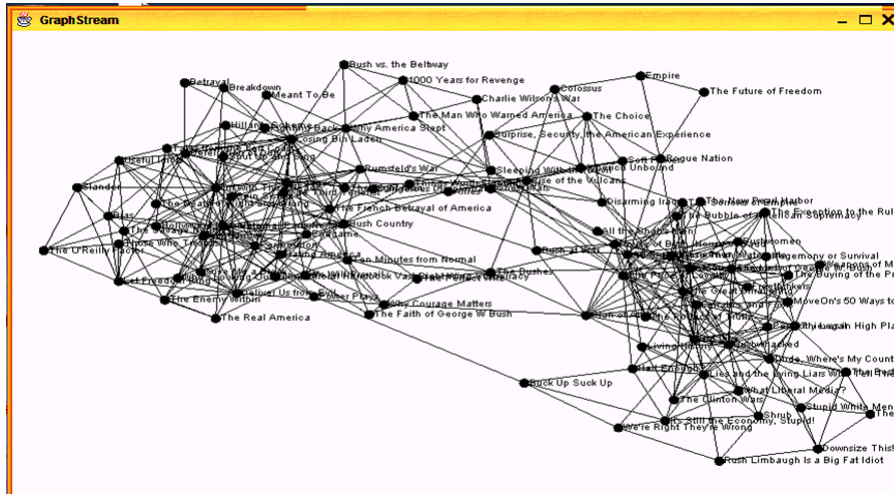


FIGURE 4.11 – capture d'écran < visualisation d'un réseau >.

Cette capture représente la structure de l'un des réseaux qu'on a choisi sous une forme graphique

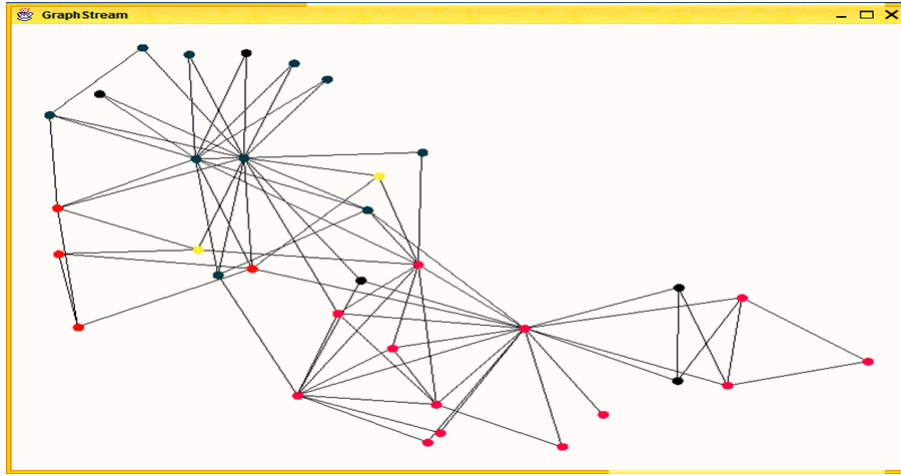


FIGURE 4.12 – capture d'écran < la visualisation des communautés détectées après l'exécution >.

## 4.4. DISCUSSION DES RÉSULTATS

### 4.4.1. ÉVALUATION

On a évalué avec deux mesures d'évaluations qui sont la modularité et NMI (Normalized Mutual Infomation).

Dans notre application on a choisi comme fonction d'évaluation la fonction de la modularité (Q)

$$Q = \frac{1}{2 * m} \sum_{ij} (Adj_{(i,j)} - \frac{k_i * k_j}{2 * m}) * \delta(c_i, c_j)$$

Cette fonction est plus détaillée dans le deuxième chapitre.

**NMI** Soit  $G = (V, E)$  le graphe modélisant un réseau, où  $V$  est l'ensemble des  $n$  nœuds constituant le réseau, et  $E$  les arêtes réglées reliant les couples d'éléments de  $V$ .

Une communauté est un groupe de sommets ayant une forte densité d'âges et une faible densité d'arêtes entre les groupes. Une structure de communauté sur  $V$  est un partitionnement  $A = \{A_1, \dots, A_R\}$

de  $V$  in un nombre  $R$  de sous-ensembles, tels que  $\bigcup_{i=1}^R A_i = V$  et  $A_i \cap A_j = \emptyset$  étant donné deux partitions et  $A = \{A_1, \dots, A_R\}$  et  $B = \{B_1, \dots, B_S\}$

de  $V$ , le chevauchement entre  $A$  et  $B$  peut être représenté par la table de contingence  $C$  également appelée matrice de confusion, de taille  $R * S$ , où  $C_{ij}$  désigne le nombre de nœuds que partagent les clusters  $A_i$  et  $B_j$ .

L'information mutuelle normalisée NMI ( $A, B$ ) de deux divisions  $A$  et  $B$  d'un réseau est définie comme suit.

$$NMI(A, B) = \frac{-2 \sum_{i=1}^R \sum_{j=1}^S C_{ij} \log(C_{ij}n / C_i C_j)}{\sum_{i=1}^R C_i \log(C_i / n) + \sum_{j=1}^S C_j \log(C_j / n)}$$

$NMI(A, B) = 1$ , si  $A$  et  $B$  sont complètement différents de  $NMI(A, B) = 0$ . [29]

## 4.5. COMPARAISON DES RÉSULTATS ET DISCUSSION

Nous rappelons que la méthode proposée est comparée avec Edge Betweenness, Label Propagation, Fast Greedy, Walktrap pour détecter les communautés, ils sont également très connus et considérés comme les plus performants et de bons étalons pour une comparaison de performance en tant que terme de précision du processus de détection des communautés. Pour chaque réseau généré, l'algorithme proposé dans ce mémoire ainsi que les quatre algorithmes choisis pour la comparaison sont exécutés. Pour Comparer ces résultats.

Plusieurs mesures peuvent être utilisées alors que Nous avons choisi la mesure NMI et la mesure de modularité  $Q$  car elles sont les plus utilisées.

Ces Mesures sont donc utilisées pour comparer les différents résultats trouvés par notre méthode.

On commence la comparaison par :

### 4.5.1. CLUB DE ZACHARY

La méthode proposée a donné la meilleure structure de communautés après Label Propagation avec une mesure de  $NMI=0.91$ . Les valeurs de la mesure NMI pour les autres algorithmes sont inscrites dans le tableau 4.1. Lorsque en

observant les valeurs de la modularité  $Q$  figurant dans le tableau 4.2, on trouve que la méthode proposée a la meilleure modularité avec  $Q=0.418$

#### 4.5.2. DAUPHINS DE LUSSEAU

La meilleure valeur de la mesure NMI (0.76) est donnée par la méthode proposée et Edge Betweenness et Fast Greedy. Edge Betweenness a la meilleure valeur de  $Q=0.519$  il est illustré dans le tableau 4.2.

#### 4.5.3. LIVRES POLITIQUES

La méthode proposée a une meilleure mesure de NMI égale à 0.89. Notre méthode a la meilleure valeur de  $Q=0.525$  comme il est illustré dans le tableau 4.2.

#### 4.5.4. FOOTBALL AMÉRICAIN

La meilleure valeur de la mesure NMI est donnée par walktrap (0.85) la méthode proposée a la meilleure valeur de  $Q=0.604$ .

NMI	Club de Zachary	dauphins de Lusseau	Livres politiques	Football américain
Méthode proposée	0.91	0.76	0.89	0.83
Edge Betweenness	0.76	0.76	0.81	0.83
Label Propagation	0.84	0.74	0.75	0.81
Fast Greedy	0.83	0.76	0.80	0.20
Walktrap	0.70	0.73	0.84	0.85

FIGURE 4.13 – Valeurs de la mesure NMI pour les différents réseaux.

Modularité (Q)	Club de Zachary	dauphins de Lusseau	Livres politiques	Football américain
Méthode proposée	0.418	0.501	0.525	0.604
Edge Betweenness	0.390	0.519	0.508	0.584
Label Propagation	0.336	0.409	0.489	0.597
Fast Greedy	0.374	0.485	0.502	0.576
Walktrap	0.351	0.478	0.499	0.587

FIGURE 4.14 – Valeurs de la modularité pour les différents réseaux.

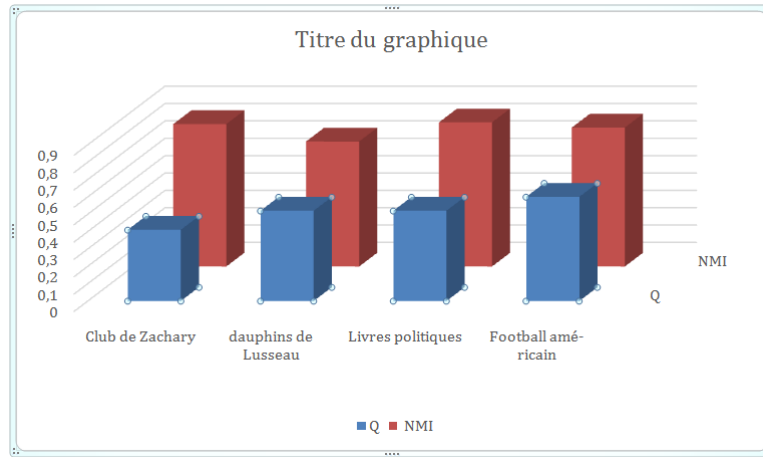


FIGURE 4.15 – Illustration graphique de comparaison des résultats pour NMI et Modularité.

## 4.6. CONCLUSION

Au cours de ce chapitre, nous avons présenté l’environnement de développement de notre application avec des petites définitions comme on a montré la structure de quelque réseaux, ainsi on a présenté notre application avec des captures d’écrans et on montré sa fonctionnalité. autrement nous avons comparé notre résultats avec des autres algorithmes, nous avons utilisé quatre réseaux réel sont Club de Zachary ,dauphins deLusseau ,Livres politiques,Football américain d’autre coté on a défini les mesures d’évaluation que notre comparaison basé sûr modularité et NMI . Finalement on a prouvé que dans la plupart de

temps la qualité de notre méthode de détection de communautés est la meilleure que celle des autres algorithmes.

# CONCLUSION

Les travaux suggérés dans notre mémoire rentrent dans le cadre de la détection des communautés sur les réseaux sociaux par les méta heuristiques qui est défini comme un domaine qui est encore dans une phase d'exploration, et pour lequel il faudra encore attendre quelques années avant d'arriver à un stade de maturation. Cette relative jeunesse du domaine, d'une part, a représenté un challenge et, d'autre part, a été un facteur de motivation important.

Dans le cadre de ce mémoire, nous avons proposé un algorithme génétique d'optimisation pour résoudre le problème de détection de communauté, et nous avons mené des comparaisons avec des algorithmes qui sont également très connus et considérés parmi les plus performants.

Les analyses comparatives de ces algorithmes ont été effectuées sur quatre réseaux sociaux. Après les tests, ces algorithmes ont été comparés statistiquement les uns avec les autres. Les résultats de notre méthode ont également été comparés aux résultats d'autres algorithmes. Comme critère de succès.

L'objectif principal est de montrer le succès de l'hyper-efficacité d'algorithme qui a offert les solutions contenant la meilleure modularité et NMI. Les résultats observés par les autres algorithmes étaient bien inférieurs.

Enfin, en fonction de la performance des algorithmes, et le critère de succès, notre méthode de détection de communautés a été observée pour être plus efficace et compétitive que les autres algorithmes.

## BIBLIOGRAPHIE

- [1] <https://www.researchgate.net/publication/314153472> L'analyse des reseaux sociaux .
- [2] Adedoyin-Olowe, M., Gaber, M. M., and Stahl, F. (2013). A survey of data mining techniques for social media analysis.arXivpreprintarXiv :1312.4617.
- [3] Erétéo, G., Gandon, F., Buffa, M., andGrohan, P. (2009). Analyse des réseaux sociaux et web sémantique : un état de l'art.Document ANR n ANR-08-cord-011-05.
- [4] Akrich, M. (1995). User representations : Practices, methods and sociology.
- [5] Mercanti-Guérin, M. (2010). Analyse des réseaux sociaux et communautés en ligne : quelles applications en marketing ?.Management and Avenir, (2), 132-153.
- [6] <http://www.cse.ust.hk/~qyang/621U/W1/621Intro.ppt>.
- [7] [https://recruteurs.apec.fr/files/live/mounts/media/mediasdelia/ documentsa-telecharger/etude\\_sapec/attitudeset\\_pratiques\\_desjeunesdiplomesconcernant\\_lusagedesreseauxsociaux\\_etlarecherchedemploi/e31ffd159a9853cc74c5657aa6e1c0aa.pdf](https://recruteurs.apec.fr/files/live/mounts/media/mediasdelia/documentsa-telecharger/etude_sapec/attitudeset_pratiques_desjeunesdiplomesconcernant_lusagedesreseauxsociaux_etlarecherchedemploi/e31ffd159a9853cc74c5657aa6e1c0aa.pdf).
- [8] Halloul, R. (2006). Réseaux d'alliances et asymétries du pouvoir : Microsoft et les navigateurs Internet. La Revue des Sciences de Gestion, 220-221,(4), 95-106. doi :10.3917/rsg.220.0095.
- [9] Gay, B., andDousset, B. (2006). Cartographie de réseaux d'alliances et analyse stratégique.Ingénierie des Systèmes d'Information,11(2), 37-51.

- [10] <https://www.theguardian.com/technology/2007/jul/25/media.newmedia21/03/2018> à 21h00.
- [11] Halloul, R. (2006). Réseaux d'alliances et asymétries du pouvoir : Microsoft et les navigateurs Internet. *La Revue des Sciences de Gestion*, 220-221,(4), 95-106. doi :10.3917/rsg.220.0095.
- [12] Gay, B., and Dousset, B. (2006). Cartographie de réseaux d'alliances et analyse stratégique. *Ingénierie des Systèmes d'Information*,11(2), 37-51.
- [13] <https://pdfs.semanticscholar.org/b47d/977827e74928db6f2e29aca2c46f16100f51.pdf>.
- [14] [https://hal.archives-ouvertes.fr/hal-00975998/file/Privacy\\_in\\_online\\_social\\_networks.pdf](https://hal.archives-ouvertes.fr/hal-00975998/file/Privacy_in_online_social_networks.pdf).
- [15] *Community Action and Social Media : A Review of the Literature*, Kevin Harris and Angus McCabe, January 2017.
- [16] <https://www.webmarketingconseil.fr/classement-reseaux-sociaux>.
- [17] Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5), 75-174.
- [18] [https://www.sagepub.com/sites/default/files/upm-binaries/35208\\_chapter1.pdf](https://www.sagepub.com/sites/default/files/upm-binaries/35208_chapter1.pdf) <https://tel.archives-ouvertes.fr/tel-00830882/document>
- [19] Combe, D. (2013). Détection de communautés dans les réseaux d'information utilisant liens et attributs. David Combe.
- [20] Atay, Y., Koc, I., Babaoglu, I., and Kodaz, H. (2017). Community detection from biological and social networks : A comparative analysis of metaheuristic algorithms. *Applied Soft Computing*,50, 194-211.
- [21] <https://tel.archives-ouvertes.fr/tel-01497593/document>
- [22] <https://ic.epfl.ch/biologie-computationnelle> , consulté le 15/03/2018 à 12h00
- [23] <https://www.sciencedirect.com/science/article/pii/S1568494616305944?via>
- [24] <https://dl.acm.org/citation.cfm?id=3030692>
- [25] <http://www.1-jour.fr/23-mai-1995-presentation-officielle-du-langage-deprogrammation-java/> consulté le 25/05/2018 à 19h00

- [26] <http://www.vogella.com/tutorials/Eclipse/article.html> , consulté le 25/05/2018 à 19h00
- [27] <https://hal.archives-ouvertes.fr/hal-00766241/document>
- [28] Loubier, E. (2009).Analyse et visualisation de données relationnelles par morphing de graphe prenant en compte la dimension temporelle(Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier).
- [29] <http://www.vogella.com/tutorials/Eclipse/article.html>
- [30] <https://www.researchgate.net/publication/281491352>*sNormalizedMutualInformation\_a Fair\_Measure\_for\_Comparing\_Community\_Detection\_Methods*
- [31] <https://pdfs.semanticscholar.org/b47d/977827e74928db6f2e29aca2c46f16100f51.pdf>.

## TABLE DES FIGURES

1.1	Réseau social montrant les nœuds et les liens.[2]	15
1.2	les différents types de RS avec des exemples de SRS réels.	17
1.3	Chronologie des dates des lancements des RS les plus importants.	20
1.4	Représentation graphique et Représentation matricielles.	23
1.5	.....	24
1.6	.....	25
1.7	.....	26
1.8	.....	26
2.1	exemple simple communauté.	36
2.2	Réseau d'interaction Elégants Protéine.	38
2.3	.....	41
2.4	Schématique d'un graphe hiérarchique.[19]	42
2.5	Un dendrogramme, ou arbre hiérarchique ,Horizontal les coupes correspondent aux partitions du graphe dans les communautés.	42
2.6	Algorithme pour résoudre le problème de CD selon l'organigramme	46
2.7	Génération de nouvelle candidate solutions par BADE.	49
3.1	Passage d'un phénomène naturel à un algorithme inspiré de la nature	65
3.2	les Méthode d'optimisation combinatoire	68
3.3	les 5 niveaux d'organisation de notre algorithme Génétique	70
3.4	fonctionnement de AG	73
3.5	.....	75
3.6	.....	75
3.7	codage alphabétique d'un chromosome	76
3.8	.....	79
3.9	.....	80
3.10	.....	80

3.11	.....	81
3.12	.....	81
3.13	.....	82
3.14	représentation schématique d'une mutation dans un chromosome	83
3.15	.....	85
3.16	.....	86
3.17	représentation schématique d'une mutation dans un chromosome.	87
4.1	Logo de Java	91
4.2	eclipse oxygen	92
4.3	Visualisation de réseau Club de karaté de Zachary	92
4.4	Visualisation de réseau Les dauphins de Lusseau	93
4.5	Visualisation de réseau Les livres politiques	93
4.6	Visualisation de réseau Football américain	94
4.7	capture d'écran <interface initiale>.	95
4.8	capture d'écran<interface d'application>.	96
4.9	capture d'écran <fenêtre de choix de réseau >.	97
4.10	capture d'écran <fenêtre de choix de réseau après la sélection de dossier qui contient la liste des réseaux >.....	98
4.11	capture d'écran < visualisation d'un réseau >.	98
4.12	capture d'écran < la visualisation des communautés détectées après l'exécution >.	99
4.13	Valeurs de la mesure NMI pour les différents réseaux.	101
4.14	Valeurs de la modularité pour les différents réseaux.	102
4.15	Illustration graphique de comparaison des résultats pour NMI et Modularité.	102

## LISTE DES ALGORITHMES

1	The pseudocode of the BA . . . . .	48
2	The pseudocode of the BADE . . . . .	50
3	The pseudocode du BB-BC . . . . .	52