

Année Universitaire 2020/2021

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

**UNIVERSITE M'HAMED BOUGARA-BOUMERDES**



**Faculté Des Sciences**

**Thèse de Doctorat**

Présentée par :

**DJERBI Rachid**

En vue de l'obtention du diplôme de **DOCTORAT** en :

**Filière : Informatique**

**Option : Informatique**

**TITRE**

**DÉTECTION DES COMMUNAUTÉS DANS LES RÉSEAUX SOCIAUX**

**Devant le jury composé de :**

Mr MEZGHICHE	Mohamed	Prof	UMBB	Président
Mr SEMCHEDINE	Fouzi	Prof	UFAS1	Examinateur
M NACER	Hassina	Prof	USTHB	Examinatrice
Mr BERRICHI	Ali	Prof	UMBB	Examinateur
Mr AMAD	Mourad	MCA	UAMOB	Directeur de thèse
Mr IMACHE	Rabah	MCA	UMBB	Co-directeur de thèse
Mr HADJ ALI	Allel	Prof	ENSMA/Fr	Invité

Année Universitaire 2020/2021

## ملخص

في السنوات الأخيرة ، تم اقتراح العديد من النماذج والنهج والخوارزميات لتحليل واستخراج المعرفة من الشبكات الاجتماعية.

أحد أكثر المعارف المطلوبة في هذا السياق هو العثور على تجميع المشتركين في "مجموعات" حول مجالات الاهتمام ومواضيع المناقشة. عندما نتحدث عن المفهوم الاجتماعي لـ "المجموعات" فإن المجتمع عبارة عن مجموعة من الأفراد (أو العقد في السياق الرسومي) مرتبطون بشدة ببعضهم البعض مع إتصال ضعيف بالآخرين.

أصبح لإكتشاف المجتمعات قيمة مهمة لفهم كيفية تغير بنية الشبكة الاجتماعية بمرور الوقت. كما أنها خطوة أساسية في تحليلها.

ومع ذلك ، لا يزال العثور على مجتمعات شبكة اجتماعية يمثل تحديًا ومجالًا للبحث الموضوعي يجذب العديد من الباحثين.

في هذه الأطروحة ، نقترح نهجًا جديدًا لاكتشاف التقسيم الأفضل للمجتمعات وفقًا لعدد العقد المشتركة بين كل زوج من الأفراد. بناءً على الحياة الاجتماعية للأفراد داخل مجتمعاتهم ، نبحث عن جميع الأزواج (الأباء) الذين لديهم الحد الأقصى من العقد (الأبناء) المشتركين لتشكيل مجتمع (عائلة) ، سينضم الأفراد الآخرون إلى العائلات التي تم العثور عليها وفقًا لبعض معايير التفضيل، المجتمعات الناتجة سيتم دمجها مع بعض في حال توفر بعض الشروط.

النموذج المقترح مستقر مما يعني أنه يعطي دائمًا نفس النتائج (أو متشابهة) لعدة عمليات تنفيذ على نفس الرسم البياني.

تظهر تجاربنا على الشبكات الاجتماعية الحقيقية أن النهج المقترح يمكن أن يحدد بدقة جميع المجتمعات. النموذج المقترح عام وتم اقتراح العديد من الإضافات مثل دعم الشبكات الموجهة / غير الموجهة ، الديناميكية / الثابتة ، الشبكات المرجحة أو غير المرجحة ، المجتمعات ذات التداخل أو بدونه.

في هذه الأطروحة نتحدث عن الشبكات الاجتماعية والكشف عن المجتمعات ، ونقدم أحدث ما توصل إليه محور البحث وتاريخه ، ثم نوضح إسهامنا وننتهي بخاتمة وبعض الأعمال المستقبلية.

### الكلمات المفتاحية

الشبكات الاجتماعية؛ كشف المجتمعات؛ تداخل المجتمعات؛ الحياة الاجتماعية للأفراد ؛ العائلات المتحكمة ؛ جودة الهياكل المجتمعية ؛ استقرار الخوارزميات.

# Résumé

Ces dernières années, plusieurs modèles, approches et algorithmes pour analyser et extraire les connaissances des réseaux sociaux (SN) ont été proposés. L'une des connaissances les plus recherchées dans ce contexte est de trouver le regroupement d'abonnés en "clusters" autour des centres d'intérêt et de sujets de discussion. On parle alors du concept social des «communautés». Une communauté est donc un groupe d'abonnés (ou de nœuds dans le contexte graphique) fortement connectés entre eux et faiblement connectés avec les autres.

La détection des communautés est devenue une tâche importante pour comprendre comment la structure du SN change avec le temps. C'est également une étape essentielle de l'analyse des SN.

Pendant, trouver les communautés d'un réseau social reste un défi et un domaine de recherche d'actualité qui attire de nombreux chercheurs.

Dans ce travail, nous proposons une nouvelle approche pour détecter la meilleure partition des communautés en fonction du nombre de nœuds en commun entre chaque paire d'entre eux. En se basant sur la vie sociale des individus au sein de leurs sociétés, nous cherchons l'ensemble des paires (parents) ayant le maximum de nœuds (fils) en commun pour en former une communauté (famille), les autres individus joindront les communautés adéquates selon quelques paramètres de préférences, les communautés trouvées se fusionnent selon quelques conditions.

Le modèle proposé est stable, veut dire qu'il donne toujours les mêmes résultats (ou similaires) pour plusieurs exécutions sur le même graphe.

Nos expériences sur des vrais SN montrent que l'approche proposée peut définir avec précision l'ensemble des communautés. Le modèle proposé est générique et plusieurs extensions ont été proposées comme la prise en charge des réseaux orientés/non orientés, dynamiques/statiques, pondérés ou non, communautés avec ou sans chevauchement.

Dans ce mémoire nous parlons des réseaux sociaux et la détection des communautés, donnons un état de l'art et historique de cet axe de recherche, puis détaillerons notre contribution et nous finissons avec une conclusion et quelques perspectives.

## **Mots clés:**

Réseaux sociaux; détection communautaire; les communautés chevauchantes; vie sociale des individus; familles larges; qualité des structures communautaires; stabilité des algorithmes.

# Abstract

In recent years, several models, approaches and algorithms for analyzing and extracting knowledge from social networks (SN) have been proposed. One of the most knowledge in this context is to find the grouping of subscribers into "clusters" around areas of interest and discussion topics. We then speak of the social concept of "communities". A community is therefore a group of subscribers (or nodes in the graphic context) strongly connected to each other and weakly connected to the others.

Detecting communities has become an important task to understand how the structure of the SN changes over time. It is also an essential step in the analysis of NS.

However, finding the communities of a social network remains a challenge and an area of topical research that attracts many researchers.

In this work, we propose a new approach to detect the best partition of communities according to the number of nodes in common between each pair of them. Based on the social life of individuals within their societies, we try to find all the pairs (parents) having the maximum number of nodes (children) in common to form a community (family), the other individuals will join the communities adequate according to a few preference parameters, the communities found merge under a few conditions.

The proposed model is stable means that it always gives the same results (or similar) for several executions on the same graph.

Our experiences with real SNs show that the proposed approach can precisely define all of the communities. The proposed model is generic and several extensions have been proposed such as the support of oriented / non-oriented networks, dynamic / static aspects, weighted or unweighted networks, communities with or without overlap.

In this thesis we talk about social networks and the detection of communities, give a state of the art and history of this research axis, then detail our contribution and we end with a conclusion and some perspectives.

## **Keywords:**

Dynamic social networks; community detection; communities overlap; social life of individuals ; large families; quality of community structures; stability of algorithms.

# Dédicace

Je dédie ce modeste travail à :  
L'âme de mon père  
Ma mère  
Ma femme et ma petite fille  
Mes frères et toute la famille  
Et tous mes amis.

# Remerciements

Tout d'abord, je remercie mon Dieu, qui m'a donné le tout pour pouvoir accomplir ce travail.

Je remercie tout particulièrement Dr. AMAD Mourad et Dr. IMACHE Rabah qui ont encadré ce mémoire. Ils ont toujours été disponibles et ont su me guider tout au long de mon mémoire tout en me laissant une grande liberté.

Je souhaite remercier vivement le Pr MEZGHICHE Mohamed qui m'a fait l'honneur de présider le jury de soutenance.

Je remercie de plus les membres de mon jury, Pr SEMCHEDINE Fouzi, Pr NACER Hassina et Pr BERRICHI Ali de m'avoir fait l'honneur d'être rapporteurs de mémoire et surtout d'avoir accepté de juger mon travail dans des délais très courts malgré leurs emplois de temps plus que saturés.

J'adresse mes remerciements à Monsieur Mohamed CHAABANI, Chef du département d'Informatique pour ses efforts et son aide.

Je remercie le Professeur HADJALI Allel de L'IAS, Poitiers (France) pour son aide et le temps qu'il m'avait accordé pendant mon stage PNE au niveau de son laboratoire.

Je tiens à adresser mes remerciements les plus sincères aux collègues enseignants, et personnel administratif du département d'Informatique à l'UMBB et du laboratoire LIMOSE.

Je remercie particulièrement mes collègues et amis: Chikh SALMI, Djamel GACEB, Abd Ellah REZZOUG, Ahmed AIT BOUZIAD, Lotfi ZERARGA, Karim AMIR, Samiya HAMADOUCHE, Abdelhak MESBAH, Youcef YAHYATEN et Selma DJEDDAI.

Je remercie également tous les membres du laboratoire LIMPAF de l'université du Bouira surtout Dr ABBAS, Dr CHOUIREF et Lyes BADIS de leurs soutiens, aides et conseils durant toute la durée de cette thèse.

Je remercie mes collègues et membres de mon équipe de recherche: Razika LOUNAS, Hocine MOKRANI, Dhia Eddine SALHI et Mohamed Taher BENNAI pour leur précieuse aide concernant la préparation de la soutenance.

J'exprime mes sincères remerciements à mes frères pour leur soutien et leurs encouragements.

Aucune page de remerciements ne suffira pour exprimer ma gratitude infinie et ma profonde reconnaissance à la personne la plus chère au monde: merci à ma mère à qui je dois tout et qui est un trésor dans ma vie.

# Table des matières

Introduction générale.....	12
<b>Introduction</b> .....	17
<b>I.1. Définitions</b> .....	17
<b>I.2. Caractéristiques des réseaux sociaux</b> .....	18
<b>I.3. Analyse des réseaux sociaux</b> .....	18
<b>I.4. Classification et exemples des réseaux sociaux</b> .....	19
<b>1.4.1. Réseaux sociaux généralistes (<i>grands publics</i>)</b> .....	19
<b>1.4.2. Réseaux sociaux spécialisés</b> .....	19
<b>a. Réseaux sociaux professionnels</b> .....	19
<b>b. Réseaux sociaux d'entreprise</b> .....	19
<b>I.5. Étude des cas</b> .....	20
<b>I.5.1. Facebook</b> .....	20
<b>I.5.2. Twitter</b> .....	21
<b>I.5.3. YouTube</b> .....	22
<b>I.5.4. Instagram</b> .....	23
<b>I.5.5. Google+</b> .....	24
<b>I.5.6. LinkedIn</b> .....	25
<b>I.5.7. Viadeo</b> .....	26
<b>I.5.8. Yammer</b> .....	27
<b>I.6. Étude comparative</b> .....	28
<b>I.7. Avantages et inconvénients des réseaux sociaux</b> .....	29
<b>1.7.1. Avantages des réseaux sociaux</b> .....	29
<b>1.7.2. Inconvénients des réseaux sociaux</b> .....	30
<b>I.8. Modélisation des Réseaux sociaux</b> .....	35
<b>Conclusion</b> .....	40
<b>Introduction</b> .....	42
<b>II.1. Communautés, définitions</b> .....	42
<b>II.2. Détection des communautés, objectifs</b> .....	43
<b>II.3. Chevauchement des communautés</b> .....	44
<b>A. Communautés chevauchantes</b> .....	44
<b>B. Communautés non chevauchantes</b> .....	45
<b>II.4. Types de communautés</b> .....	46

<b>A.</b>	<b>Communautés à sens forts</b> .....	46
<b>B.</b>	<b>Communautés à sens faible</b> .....	47
<b>II.5.</b>	<b>Qualité d'une distribution communautaire</b> .....	48
<b>II.6.</b>	<b>Stabilité des algorithmes de détection de communautés</b> .....	49
<b>II.7.</b>	<b>Historique et travaux connexes</b> .....	50
<b>II.8.</b>	<b>Principales approches de détections des communautés dans les SN</b> .....	52
<b>A.</b>	<b>Algorithme de Louvain</b> .....	54
<b>B.</b>	<b>Algorithme de Newman et Girvan</b> .....	56
<b>II.9.</b>	<b>Étude comparative entre quelques algorithmes de détection de communautés</b> .....	57
<b>A.</b>	<b>Comparaison qualitative</b> .....	57
<b>B.</b>	<b>Comparaison quantitative</b> .....	58
	<b>Conclusion</b> .....	58
	<b>Introduction</b> .....	61
<b>III.1.</b>	<b>Communautés moyennes</b> .....	61
<b>III.2.</b>	<b>Notre proposition</b> .....	63
<b>A.</b>	<b>LFM, concept</b> .....	63
<b>B.</b>	<b>Notations</b> .....	64
<b>C.</b>	<b>Principe de l'algorithme LFM</b> .....	65
<b>D.</b>	<b>Hypothèse</b> .....	66
<b>E.</b>	<b>Qualité de la distribution communautaire générée</b> .....	67
<b>F.</b>	<b>Stabilité de l'algorithme LFM</b> .....	68
<b>III.3.</b>	<b>Extensions du LFM</b> .....	68
<b>A.</b>	<b>Niveau d'extension 1: avec / sans chevauchement</b> .....	69
	<b>a. Amélioration 1: meilleure modularité globale</b> .....	71
	<b>b. Amélioration 2: paramètres de préférence</b> .....	71
<b>B.</b>	<b>Niveau d'extension 2: non orienté / orienté</b> .....	73
<b>C.</b>	<b>Niveau d'extension 3: dynamique / statique</b> .....	74
<b>III.4.</b>	<b>LFM et les autres</b> .....	75
	<b>Conclusion</b> .....	76
	<b>Introduction</b> .....	79
<b>IV.1.</b>	<b>Framework</b> .....	79
<b>IV.2.</b>	<b>Algorithme LFM</b> .....	79
<b>IV.3.</b>	<b>Étude de cas</b> .....	81
<b>A.</b>	<b>Les réseaux aléatoires</b> .....	82
<b>B.</b>	<b>Le réseau "Zachary's karate club"</b> .....	85
<b>C.</b>	<b>La multinationale «Enron»</b> .....	88



1. La base de données Enron .....	88
2. Construction du réseau social Enron.....	89
3. Application de l'algorithme de Louvain sur le réseau social Enron .....	91
4. Application de notre algorithme LFM sur le réseau social Enron.....	91
<b>IV.4. Performances et étude comparative.....</b>	<b>93</b>
A. Application sur un réseau aléatoire à grande échelle.....	93
B. Complexité .....	94
C. Temps d'exécution .....	94
<b>Conclusion .....</b>	<b>95</b>
<b>Introduction.....</b>	<b>97</b>
<b>V.1. Extension No_ELFM .....</b>	<b>97</b>
A. LFM de base, rappel .....	97
B. LFM sans chevauchement .....	98
1. Scénario 1: (P&P) .....	99
2. Scénario 2: (P&C).....	101
3. Scénario 3: (C&C) .....	103
<b>V.2. Vivacité des communautés dans No_ELFM.....</b>	<b>105</b>
<b>V.3. LFM léger.....</b>	<b>107</b>
<b>V.4. Mise en œuvre, expérimentation et discussion .....</b>	<b>107</b>
<b>Conclusion .....</b>	<b>113</b>
Conclusion générale .....	114
Bibliographie .....	117

# Table des figures

1- Figure I 1: Domaines de recherches des réseaux sociaux.....	17
2- Figure I 2: logos de Facebook [21].....	21
3- Figure I.3 : Logos de Twitter [21].....	22
4- Figure I.4 : Logos de Youtube [21].....	23
5- Figure I.5 : Logo d'Instagram [25]. ....	24
6- Figure I.6 : Logos de Google+ [21].....	25
7- Figure I.7 : Logos de LinkedIn [31].....	26
8- Figure I.8 : Logos de Viadeo [32].....	27
9- Figure I.9 : Logo de Yammer [21] .....	28
10- Figure I.10 : Représentation graphique et matricielle d'un graphe non orienté. ....	38
11- Figure I.11 : Représentation graphique et matricielle d'un graphe orienté. ....	38
12- Figure I.12 : Autres représentations mathématiques que la matrice d'adjacence. ....	39
13- Figure II.1 : Distribution communautaire, information cachée. ....	43
14- Figure II.2 : Le sens de communauté dans différents domaines.....	44
15- Figure II.3. Communautés chevauchantes.....	45
16- Figure II.4 : Communautés au sens fort.....	46
17- Figure II.5 : Communautés au sens faible.....	47
18- Figure II.6 : Stabilité des algorithmes de détection de communautés. ....	50
19- Figure II.7 : Approches et algorithmes de détection des communautés, Synthèse de [108]. ....	53
20- Figure II.8 : Approches hiérarchiques. ....	53
21- Figure II.9 : Dendrogramme d'une approche hiérarchique de détection de communautés.....	54
22- Figure II.10 : Les deux phases de l'algorithme de Louvain [4]. ....	55
23- Figure III.1 : Notre nouvelle notion de communautés à sens moyen. ....	62
24- Figure III.2 : Extension de la notion de communauté moyenne aux graphes pondérés. ....	63
25- Figure III.3 : MRC, communautés, parents, "In_node" et "Out_node" dans LFM. ....	64
26- Figure III.4 : Organigramme de l'algorithme LFM proposé.....	66
27- Figure III.5: Deux parents doivent avoir au moins deux enfants pour former une communauté. 67	
28- Figure III.6: Algorithme de qualité.....	68
29- Figure III.7: Les trois niveaux d'extension du LFM. ....	69
30- Figure III.8: Exemple de communautés dans le modèle LFM orienté. ....	74
31- Figure IV.1: La matrice de voisinage en commun MCadj.....	81
32- Figure IV.2: La matrice d'adjacence de l'exemple traité par Blondel et al. [4]. ....	82
33- Figure IV.3: Exemple traité par Blondel et al. [4] avant la détection des communautés.....	82
34- Figure IV.4: Communautés trouvée par Louvain sur l'exemple de Blondel et al. dans [4]. ....	83
35- Figure IV.5: Communautés trouvée par LFM sur l'exemple de Blondel et al. dans [4] (avant exécution de sa phase 3). ....	84
36- Figure IV.6: Communautés finales trouvées par LFM sur l'exemple de Blondel et al. dans [4]. ...	84
37- Figure.IV.7. Zachary Dataset: Communities number vs. CadjMax. ....	85
38- Figure.IV.8. Zachary Dataset: Nombre de nœuds "sortants" par rapport à CadjMax.....	86
39- Figure.IV.9. Zachary Dataset: modularité vs CadjMax.....	87
40- Figure.IV.10. Zachary Dataset: Temps d'exécution vs CadjMax.....	87
41- Figure IV.11: Les tables de la base de données Enron. ....	89

42- Figure.IV.12. Relations de contact entre les membres d'Enron.....	90
43- Figure.IV.13. Réseau d'ENRON avant l'application d'aucun algorithme de détection communautaire.....	90
44- Figure.IV.14. Réseau d'ENRON après application de l'algorithme de Louvain.....	91
45- Figure IV.15: Le réseau social d'ENRON: nombre de communautés vs CadjMax.....	91
46- Figure IV.16: Le réseau social d'ENRON: nombre de nœuds "Out" vs CadjMax.....	92
47- Figure IV.17 Ensemble de données ENRON: temps d'exécution vs CadjMax.....	92
48- Figure IV.18: Le réseau social d'ENRON: temps d'exécution vs nombre de communautés.....	93
49- Figure IV.19 Application sur un réseau aléatoire à grande échelle: nombre de nœuds en fonction du temps d'exécution.....	94
50- Figure.V.1. OOS, existence de relation sociale entre le parent et le conjoint choisis.....	99
51- Figure.V.2 : OOS, selon le nombre d'enfants.....	100
52- Figure.V.3. OOS, Communauté la plus dense.....	100
53- Figure.V.4. OOS, popularité des conjoints.....	101
54- Figure.V.5. Le scénario (P&C), le choix par défaut: parent.....	102
55- Figure.V.6 : Le scénario (P&C), relations dans chaque communauté.....	102
56- Figure.V.7. Le scénario (C&C), la communauté la plus dense (en termes de nœuds).....	103
57- Figure.V.8. Le scénario (C&C), la communauté la plus dense (en termes de relations).....	104
58- Figure.V.9. Scénario (C&C), destruction de la communauté (Nbr_children < CadjMax).....	105
59- Figure.V.10. Scénario (C&C), reconstruction communautaire (Nbr_children >= CadjMax).....	105
60- Figure.V.11. Vivacité des communautés dans LFM.....	107
61- Figure.V.12. 417 couples (parents) pour CadjMax=2 (meilleure valeur de Q)!!.....	108
62- Figure.V.13. Le nœud (0) a finalement choisi le nœud (8), (1) le nœud (14), etc.....	109
63- Figure.V.14. Élimination des chevauchements de type (P&P).....	110
64- Figure.V.15 : Élimination des chevauchements de type (P&C).....	110
65- Figure.V.16. Population avant élimination des chevauchements de type (C&C).....	111
66- Figure.V.17. Population après élimination des chevauchements de type (C&C).....	111
67- Figure.V.18 : Destruction des communautés avec nbr_chldren < CadjMax=2.....	112
68- Figure.V.19 : Liste des communautés après avoir rejoint les "Out_nodes".....	112
69- Figure.V.20: Illustration graphique des sept communautés trouvées.....	113

# Liste des tableaux

1- Tableau I.1: Comparative, selon quelques paramètres, entre les réseaux sociaux échantillon de notre étude .....	29
2- Tableau I.2: Avantages et inconvénients des réseaux sociaux échantillon de notre étude .....	35
3-Tableau II.1: Comparaison qualitative entre quelques algorithmes de détection communautaires	57
4- Tableau II.2 : Nombre de communautés détectées.....	58
5- Tableau II.3 : Modularité des distributions communautaires finales.....	58
6- Tableau III.1 : Notations utilisées dans l'approche LFM.....	65
7- Tableau III.2 : Niveau d'extension 1: avec / sans chevauchement.....	70
8- Tableau III.3 : LFM Vs Louvain.....	76
9- Tableau IV.1 Application sur un réseau aléatoire à grande échelle: statistiques .....	93
10- Tableau IV.2: LFM vs algorithmes d'Atay et al. : temps d'exécution sur Zachary .....	94

## **Introduction générale**

De nos jours, de nombreux systèmes réels sont représentés sous forme de réseaux qui sont composés d'entités interagissant entre elles. Ces réseaux sont modélisés par des graphes dont les nœuds sont les objets constituant le système et les liens représentent les interactions entre eux. Les graphes modélisant des objets réels qui sont appelés des graphes de terrain. En informatique, le web peut être vu comme un ensemble de pages web (*de simples fichiers textes*) liées par l'intermédiaire de liens hypertextes. En biologie, il existe des réseaux métaboliques dont les nœuds sont des protéines et les liens sont les interactions chimiques entre elles. En sociologie, les nœuds sont des individus ou entités sociales (*ex. personnes, organisations, entreprises, etc.*) qui sont liés entre eux par des relations de différentes natures. Dans le cadre de cette thèse, nous nous intéressons aux réseaux sociaux.

Les réseaux sociaux est un concept de psychologie organisationnelle définie sur l'établissement et le maintien de relations sociales qui pourraient donner accès à l'information et aux ressources [1].

Les recherches réalisées sur les réseaux sociaux se sont concentrées sur leur utilisation personnelle plutôt que professionnelle, alors que les auteurs de la presse populaire vantent régulièrement leurs avantages pour trouver des emplois et certains les considèrent comme la clé de communication pour les introvertis et renfermés [2].

À cet effet, encore plus de recherches fondées sur des preuves sont nécessaires pour connaître comment les individus utilisent réellement ce type de plateformes en ligne et quels sont leurs impacts sur le côté personnel ou professionnel des leurs usagers.

Les auteurs de [3] décrivent les réseaux sociaux comme «des services basés sur le Web qui permettent aux individus (1) de construire un profil (semi-) public au sein d'un système limité, (2) d'articuler une liste d'autres utilisateurs avec qu'ils partagent une connexion, et (3) afficher et parcourir leurs listes de connexions et celles faites par les autres». Également, ils permettent aux utilisateurs de parcourir les profils de leurs contacts pour en savoir plus sur les liens faibles des autres et éventuellement établir une connexion [2].

L'étude structurale des réseaux sociaux est une thématique spécialisée de la sociologie qui s'intéresse aux différentes informations, souvent cachées, des acteurs. L'analyse des réseaux, a développé une grande panoplie de métriques pour caractériser les réseaux sociaux.

Les réseaux sociaux, de par leur nature, peuvent facilement être représentés sous forme de graphes où chaque acteur du réseau social est représenté par un sommet (*ou nœud*) et chaque relation entre deux individus est représentée par une arête (*si la relation est symétrique*) ou un arc (*si la relation est asymétrique*). Par analogie, il est donc possible d'appliquer les concepts de la théorie des graphes aux réseaux sociaux.

L'analyse des réseaux sociaux permet d'extraire autant de propriétés, caractéristiques et beaucoup d'informations cachées. Parmi les problèmes actuels de l'analyse des réseaux sociaux, on trouve la détection des communautés ou le phénomène de regroupement des utilisateurs autour d'un même centre d'intérêt ou même profil.

Tout réseau social est caractérisé par l'existence de zones plus densément connectées que d'autres. Graphiquement, ces zones correspondent à des groupes de nœuds plus fortement connectés entre eux mais faiblement reliés aux autres. Dans le domaine des réseaux

sociaux, ce regroupement se fait en «communautés» qui représentent des groupes sociaux (ex. *familles, groupes d'amis, etc.*) avec des intérêts communs et des activités communes.

Dans la littérature, les chercheurs ont proposé de nombreux algorithmes pour détecter des communautés dans les réseaux sociaux comme celui de Louvain [4], Girvan [5] et bien d'autres.

La détection de ces communautés représente un axe de recherche très important dans l'analyse des réseaux sociaux qui a reçu beaucoup d'intérêt par des différents chercheurs, puisqu'elle permet d'obtenir une vision macroscopique de la structure des réseaux. Elle permet aussi de déterminer le rôle de différents acteurs au sein des communautés et dans la globalité du réseau. C'est dans cette thématique que s'inscrit cette thèse.

En général, la recherche d'une solution exacte à une tâche de partitionnement des graphes est considérée comme un problème NP-complet [5], ce qui rend sa résolution extrêmement difficile pour les grands graphes, mais une grande variété d'algorithmes heuristiques ont été développés qui donnent des solutions acceptables et bonnes dans de nombreux cas [5].

Pour détecter l'ensemble des communautés d'un réseau social, plusieurs approches et algorithmes de détection de communautés ont été proposés, ces derniers servent à décomposer le graphe et rechercher des "clusters" ayant une/ou plusieurs propriétés communes. Une bonne décomposition du graphe vérifie que la densité de chaque groupe soit élevée mais aussi celle entre les différents groupes soit faible.

La recherche de communautés est un problème important dans le domaine des réseaux complexes tels que les réseaux sociaux. La modélisation de ces derniers dans le domaine mathématique et graphique est une étape importante dans le cadre de notre travail. Étudier, analyser puis comprendre l'état de l'art des différentes approches proposées dans la littérature étaient un chemin incontournable dans la réalisation de notre travail. Finalement, proposer une nouvelle approche qui n'a jamais été discutée avant était notre problématique essentielle.

L'objectif principal de notre travail est d'étudier les algorithmes de détection des communautés les plus connus dans la littérature, recenser leurs limites et inconvénients pour ensuite proposer un nouveau qui serait stable, précis, facile à mettre en place et efficace pour les réseaux sociaux.

Dans cette thèse, nous avons proposé un nouveau modèle pour la détection des communautés dans les réseaux sociaux. Il se base sur la notion de familles larges au sein des sociétés où chaque communauté représente une famille formée de deux parents et un ensemble de fils, ces derniers représentent l'ensemble des nœuds en relation commune avec les deux parents, les autres nœuds joignent telle ou telle communauté sous certaines conditions. Avec ce système on a pu simuler le phénomène de migration des individus vers d'autres populations.

Dans la littérature il existe deux catégories de communautés, faibles et fortes. Dans notre travail nous avons proposé une nouvelle catégorie que nous appelons "Moyenne" qui est le concept de base de notre modèle proposé.

Nous avons détaillé les concepts de base de quelques algorithmes de détection de communautés, (1) la stabilité des algorithmes proposés où un algorithme est dit stable s'il donne les mêmes résultats (*ou similaires*) pour une multiple exécution et (2) la qualité de la distribution communautaire trouvée à la fin où une fonction objectif doit être améliorée afin

d'en qualifier telle ou telle solution. Notre travail consiste de proposer un algorithme stable mais aussi qui cherche la meilleure qualité possible parmi les solutions trouvées où on propose un compromis entre ces deux notions de base.

Notre modèle proposé se base sur la vie sociale des familles où chacune a deux parents et des enfants, puis d'autres nœuds "Out" qui vont joindre ces familles selon plusieurs paramètres de sélection. Les familles les plus larges qui vont maîtriser la société et qui prennent des décisions à la place des autres, pour cela, notre modèle a été appelé "LFM : Large Families Model".

LFM est générique où on a peu l'étendre aux réseaux orientés/non orientés, dynamiques/statiques, pondérés, simples et finalement avec sans chevauchement.

Une transcription sociologique de notre modèle a été proposée et bien discutée tout en montrant la faisabilité d'application sur les réseaux sociaux et bien d'autres types de réseaux.

L'organisation de cette thèse s'articule autour des cinq chapitres comme suit:

Dans le premier chapitre, nous allons présenter **un état de l'art des réseaux sociaux** en introduisant leur définition, intérêts et classification dont nous donnons un exemple de chaque catégorie afin de faire une étude comparative entre eux. Nous consacrons ce premier chapitre aux réseaux sociaux vu que ce type particulier de réseaux est un domaine de recherche en plein développement. Il est possible d'identifier plusieurs types de relations comme la collaboration, soutien et assistance, suivi et abonnement, contrôle, conseil, influence, parenté, amitié, etc. qu'entretiennent les acteurs les uns avec les autres au sein d'un réseau social.

Le deuxième chapitre est consacré à la présentation **d'un état de l'art sur les communautés** et les différentes approches de détection de communautés. Nous rapportons en premier lieu les diverses définitions relatives à la notion de communauté ainsi que sa définition formelle puis les différentes approches et méthodes proposées pour la détection des communautés pour illustrer leur diversité.

Dans le chapitre trois, nous proposons une modélisation et conception de notre **nouveau modèle** baptisé **LFM** (*ou Large Families Model*) en expliquant en détail son fonctionnement, le modèle générique, l'algorithme de base et ses différentes extensions.

Le chapitre quatre est consacré à **implémenter notre nouveau modèle**. Nous donnons tout d'abord l'algorithme LFM, présentons ensuite notre procédure d'implémentation et pour évaluer ses performances nous l'appliquons sur des jeux de données réelles et aléatoires.

Dans le cinquième chapitre, nous présentons et implémentons une extension très significative de notre modèle de base LFM qui concerne l'élimination des cas de chevauchements entre les communautés détectées ou le modèle **No\_ELFM** est proposé et discuté ainsi que les différents scénarios.

Une conclusion générale clôturera cette thèse dont laquelle nous avons rappelé les principales points abordés dans cette thèse en particulier notre proposition. Une liste des perspectives à court et à long termes sont données à la fin.

Cette thèse a fait l'objet de divers travaux:

Rachid DJERBI, Mourad AMAD, Rabah IMACHE, *A New Model for Communities' Detection in Dynamic Social Networks Inspired From Human Families*, *International Journal of Internet Technology and Secured Transactions (IJITST)*, Vol. 10, No. 1/2, 2020, *Inderscience*. <https://doi.org/10.1504/ijitst.2020.10026089>

DJERBI, Rachid, IMACHE, Rabah, et AMAD, Mourad. Communities' Detection in Social Networks: State of the art and perspectives. In: *2018 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, 2018. p. 1-6. <https://doi.org/10.1109/isncc.2018.8531055>

Rachid Djerbi, Mourad Amad, Rabah Imache, *A Novel Solution for Communities' Detection in Social Networks*, in The 3rd International Conference on Networking and Advanced Systems 2017, *ICNAS2017* (Annaba, Algeria), 2017.

Rachid Djerbi, Allel Hadjali, Mourad Amad, Rabah Imache, Mohamed T. Bennai, *Social context-based non-overlapping communities' detection model in social networks*. The International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD), Tangier, Marocco 2020.



## Chapitre 1 : État de l'art sur les Réseaux Sociaux

<b>Introduction</b> .....	17
<b>I.1. Définitions</b> .....	17
<b>I.2. Caractéristiques des réseaux sociaux</b> .....	18
<b>I.3. Analyse des réseaux sociaux</b> .....	18
<b>I.4. Classification et exemples des réseaux sociaux</b> .....	19
<b>1.4.1. Réseaux sociaux généralistes (<i>grands publics</i>)</b> .....	19
<b>1.4.2. Réseaux sociaux spécialisés</b> .....	19
<b>a. Réseaux sociaux professionnels</b> .....	19
<b>b. Réseaux sociaux d'entreprise</b> .....	19
<b>I.5. Étude des cas</b> .....	20
<b>I.5.1. Facebook</b> .....	20
<b>I.5.2. Twitter</b> .....	21
<b>I.5.3. YouTube</b> .....	22
<b>I.5.4. Instagram</b> .....	23
<b>I.5.5. Google+</b> .....	24
<b>I.5.6. LinkedIn</b> .....	25
<b>I.5.7. Viadeo</b> .....	26
<b>I.5.8. Yammer</b> .....	27
<b>I.6. Étude comparative</b> .....	28
<b>I.7. Avantages et inconvénients des réseaux sociaux</b> .....	29
<b>1.7.1. Avantages des réseaux sociaux</b> .....	29
<b>1.7.2. Inconvénients des réseaux sociaux</b> .....	30
<b>I.8. Modélisation des Réseaux sociaux</b> .....	35
<b>Conclusion</b> .....	40

Les réseaux sociaux (*SN* ou *Social Network*) sont devenus un outil incontournable dans le domaine de communication, partage et réception de l'information. Dans ce chapitre nous donnons un aperçu sur cette tendance technologique, ses catégories et quelques exemples avec une discussion de leurs avantages et inconvénients.

## Introduction

Aujourd'hui, les internautes sont attachés aux réseaux sociaux (RS), et presque tout le monde utilise cette ressource virtuelle qui relie les-uns aux autres directement ou indirectement. Depuis l'avancement de l'internet et l'évolution de la technologie web, les réseaux sociaux sont devenus des outils de communication incontournables, ils permettent aux différents utilisateurs de se regrouper et d'interagir en communautés, de partager leurs informations personnelles (*photos, musiques, vidéos, etc.*), réagir et commenter les informations des autres, créer leurs propres profils, etc. Ces réseaux sociaux sont de différents types, certains sont connus par tout le monde et comptent des millions de membres et d'autres sont relativement inaperçus ou restés confidentiels comme les réseaux sociaux des entreprises.

L'étude et l'analyse des interactions entre les nœuds sont très importantes pour comprendre l'évolution de ces types de réseaux dans le temps. Plusieurs chercheurs se sont intéressés à l'analyse de tels réseaux et ont révélé plusieurs caractéristiques structurelles (*figure 1.1*) telles que l'existence de motifs locaux répétés [6], la haute transitivity du réseau [7], la distribution des degrés de loi de puissance [8] et la détection de communautés [9].

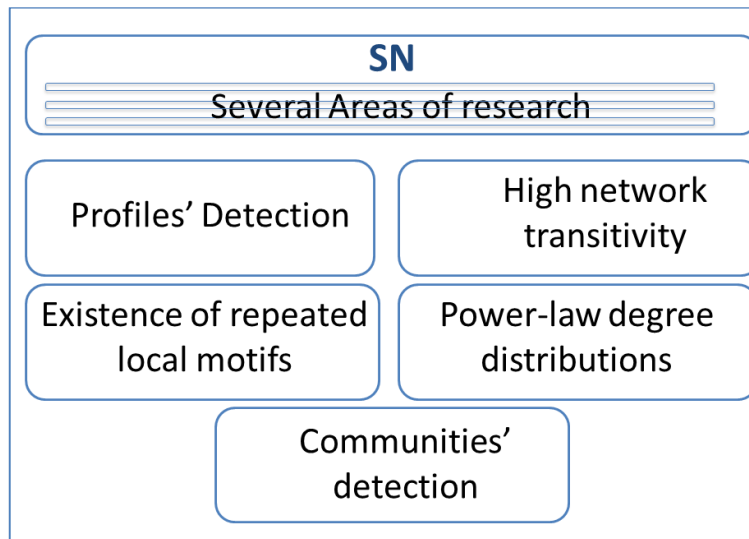


Figure 1.1: Domaines de recherches des réseaux sociaux.

Les SNs sont devenus un outil incontournable dans le domaine de communication, partage et réception de l'information. Dans ce chapitre nous donnons un aperçu sur cette tendance technologique, ses catégories et quelques exemples avec une discussion de leurs avantages et inconvénients.

### I.1. Définitions

Un réseau social est un système reliant des individus à travers des relations sociales, il leur permet plusieurs services comme la création de profils, la communication et le partage d'informations.

Autres définitions des réseaux sociaux ont été données dans la littérature, en citant quelques-unes comme suit :

Un réseau social représente toute plateforme en ligne dont la finalité est de mettre en relation des membres, et sur laquelle un individu peut s'inscrire librement, construire son propre réseau, produire du contenu, le partager et interagir avec les membres de son réseau. Un membre peut y créer un profil public visible par tous ou privé, visible par ses contacts uniquement. L'intérêt de telles plateformes est notamment de pouvoir suivre l'actualité des membres de son réseau et d'éventuellement la commenter. [10]

Un réseau social est un ensemble d'acteurs sociaux, tels que des individus ou des organisations, reliés entre eux par des connexions représentant des interactions sociales. Ces dernières peuvent avoir différentes natures : familiales, amicales, professionnelles, etc. [11]

Selon la définition proposée par Ellison et Boyd [12], les sites de réseaux sociaux sont des espaces d'échange sur internet qui permettent aux individus de construire des profils publics ou semi-publics dans le cadre d'un système délimité, puis d'articuler ces profils avec des listes d'utilisateurs interconnectés et de visualiser leurs listes de contacts et celles des autres.

Aussi, mustafa Hajeer et al. décrivent dans [13] un réseau social comme étant un ensemble d'acteurs qui interagissent fréquemment et participent à certaines discussions.

### I.2. Caractéristiques des réseaux sociaux

Le concept des réseaux sociaux est très large et plusieurs sites Web peuvent être classés dans la même famille. Alors qu'il existe quelques spécificités nous permettant de les différencier des autres. Chaque réseau social pourra être observé comme un site web caractérisé par les propriétés principales suivantes :

- La création d'un profil personnel représentant l'identité (*réelle ou virtuelle*) de l'utilisateur,
- L'utilisation obligatoire de cette notion de profil pour toute tentative de publication ou de mise à jour des informations,
- La recherche d'autres utilisateurs ou profils du système,
- La possibilité de commenter les publications des autres personnes de manière publique ou semi-publique,
- .. etc.

### I.3. Analyse des réseaux sociaux

L'analyse des réseaux sociaux (*ou Social Network Analysis SNA*) est une méthodologie pour décrire et étudier les relations entre les individus ou les nœuds d'un réseau. En tenant compte de la résilience de l'établissement de ces nœuds, le SNA peut être utilisé efficacement pour étudier les phénomènes organisationnels et inter-organisationnels.

De nombreux travaux ont étudié les différentes façons dont cette procédure d'analyse de réseaux sociaux est effectuée. L'analyse des réseaux sociaux a été appliquée sur plusieurs axes de recherche tels que la structure du réseau correspondant, le changement (*ou non*) des relations entre les différents utilisateurs, les citations dans le domaine de la rédaction d'articles scientifiques, dans le domaine commercial et les liens entre clients, produits et vendeurs, systèmes de recommandation de produits ou de films aux clients ... etc. En 2019, les auteurs de [14] ont étudié les différentes méthodes et applications du "Machine Learning" permettant d'analyser à grande échelle des données de mobilité humaine. De plus, la même année, les auteurs de [15] ont amélioré la sémantique de cette SNA afin d'identifier les différents influenceurs sur les réseaux sociaux en ligne. L'article [16] a proposé un modèle

d'aide à la décision dans des groupes à grande échelle où il a utilisé ce concept de SNA pour détecter puis éliminer les différents conflits basés sur les relations de confiance entre les utilisateurs des réseaux sociaux. Dans [17], les auteurs ont proposé une perspective très importante des réseaux sociaux, à savoir qu'il s'agit d'un moyen de soutien social afin d'assurer une transition de la formation au lieu de travail.

### 1.4. Classification et exemples des réseaux sociaux

Plusieurs études ont proposé des différentes classifications des réseaux sociaux, dans cette section nous donnons une vision plus générale en se basant sur l'échelle et la grandeur d'utilisation dans un côté et le domaine d'utilisation (*personnel ou professionnel*) d'un autre côté.

#### 1.4.1. Réseaux sociaux généralistes (*grands publics*)

Les réseaux sociaux généralistes sont des réseaux destinés à tout le monde où les relations entre individus peuvent être de différentes natures. Ils permettent aux utilisateurs de créer des profils et de faire des échanges d'information entre eux. Ils constituent donc des espaces personnalisables, simples et (*des fois*) gratuits à l'instar de **Facebook**<sup>1</sup>, **Twitter**<sup>2</sup>, **YouTube**<sup>3</sup>, **Instagram**<sup>4</sup> et **Google+**<sup>5</sup>, ..etc.

#### 1.4.2. Réseaux sociaux spécialisés

Les réseaux sociaux spécialisés sont des réseaux dédiés où les relations entre les utilisateurs sont d'un type spécifique. Le but de ce type de réseaux est de permettre aux utilisateurs d'enrichir leurs carnets professionnels, de trouver un emploi, de recruter des employés, gérer un curriculum vitae, etc. Cette catégorie pourra être divisée en deux classes à peu près similaires, réseaux sociaux professionnels et ceux d'entreprise.

##### a. Réseaux sociaux professionnels

Un réseau social est dit professionnel si la relation entre ses utilisateurs est purement professionnelle. Ces réseaux sociaux permettent à ses abonnés **professionnels** de se mettre en contact, d'échanger leurs idées, expériences et même des articles qu'ils ont rédigé. Les personnes sans emploi utilisent souvent ces réseaux sociaux professionnels afin d'en trouver un poste ou une fonction qui leur convient suivant leurs profils, compétences et valeurs. Comme exemple de ce type de réseaux on cite **LinkedIn**<sup>6</sup> et **Viadeo**<sup>7</sup>.

##### b. Réseaux sociaux d'entreprise

Toute application informatique installée et qui fonctionne sur le réseau local ou intranet d'une entreprise et permettant aux différents employés salariés de collaborer, partager et communiquer entre eux est dit «réseau social d'entreprise».

Un réseau social d'entreprise est, en général, une plateforme de réseautage collaborative et qui permet aux fonctionnaires de la même entreprise d'envoyer des informations

---

<sup>1</sup> <https://www.facebook.com/>

<sup>2</sup> <https://twitter.com/>

<sup>3</sup> <https://www.youtube.com/>

<sup>4</sup> <https://www.instagram.com/>

<sup>5</sup> <https://plus.google.com/>

<sup>6</sup> <https://www.linkedin.com/>

<sup>7</sup> <http://www.viadeo.com/>

professionnelles à leurs patrons, collaborateurs internes, ou même les membres de leurs équipes. Un exemple très pratique de ce genre de réseaux sociaux nous citons **Yammer**<sup>8</sup>.

### I.5. Étude des cas

Un réseau social est un outil collectif et collaboratif en ligne permettant à ses utilisateurs de d'interagir entre eux en partageant des images, des vidéos, des expériences et pratiques ou même des productions scientifiques.

Dans cette section, nous donnons une brève description de quelques exemples de réseaux sociaux parmi les plus utilisés dans la littérature à savoir **Facebook, Twitter, YouTube, Instagram, Google+, LinkedIn, Viadeo et Yammer**.

#### I.5.1. Facebook

Facebook [18] est un réseau social généraliste destiné aux grands publics. Il a été créé en février 2004 par Mark Zuckerberg (*étudiant en psychologie à l'université de Harvard*), il a commencé comme étant un service mis à la disposition à uniquement les membres de cette université. À l'époque et pour s'inscrire, l'utilisateur devait avoir une adresse e-mail universitaire «Harvard.edu».

Actuellement, Facebook offre plusieurs types d'interaction et de relations à ses membres entre autres l'amitié, commenter un statut ou une publication d'un autre membre, aimer...etc. L'inscription est obligatoire pour l'utiliser, sa fonction principale est de rassembler des individus proches ou inconnus. Facebook est actuellement (*Février 2020*) le réseau social le plus populaire des réseaux sociaux. Par défaut, les abonnés joignent progressivement des individus de leur milieu professionnel ou privé. Selon [19], en 2008 Facebook s'est composé de 47 000 réseaux collégiaux, secondaires, d'employés et régionaux, avec 600 millions de recherches et plus de 30 milliards de pages vues par mois.

Selon le site officiel du Facebook [18], consulté en Janvier 2020, plus de deux milliards d'utilisateurs inscrits sur la plateforme, 140 millions d'entreprises utilisent ses applications pour se connecter avec les clients et se développer, plus de 100 milliards de messages partagés chaque jour afin d'aider les gens à rester proches même lorsqu'ils sont éloignés. Ainsi, un milliard d'histoires partagées chaque jour juste pour permet aux utilisateurs à s'exprimer et à se connecter, déployé sur 15 centres de données (*Datacenter*) dans le monde et recrute 44,942 employés à temps plein en au 31 décembre 2019.

L'importance de Facebook provient de la diversité des langues, 96 langues sont disponibles, 2 milliards d'utilisateurs en mai 2017 et des revenus très élevés atteignant 17928 millions de dollars en 2015 [20]. Parmi ces technologies (*qu'il a développé ou racheté*) on trouve : Facebook app, Messenger, Instagram, WhatsApp, Oculus, Workplace, Portal et Calibra. Facebook a connu quatre changements de son logo depuis sa création (*figure 1.2*).

---

<sup>8</sup> <https://www.yammer.com/>

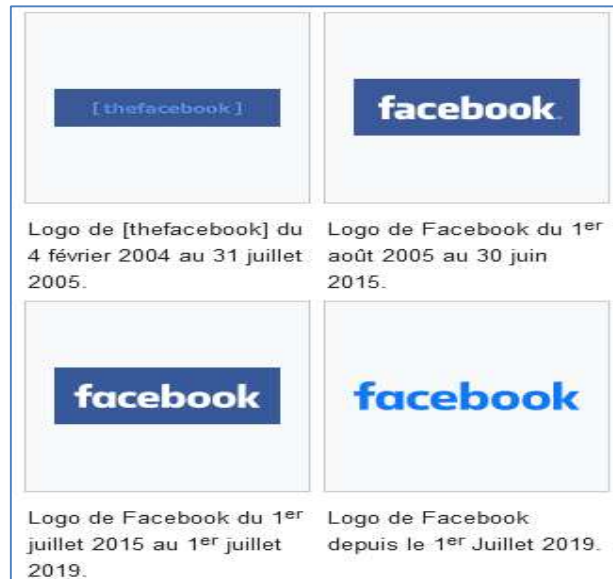


Figure I 2: logos de Facebook [21].

Facebook offre, à ses utilisateurs, plusieurs fonctions telles que :

- **Journal** : Une page personnalisée qui contient toutes les informations relatives à l'utilisateur, ses activités ordonnées chronologiquement et les contenus des différentes pages qu'il a aimé.
- **Amis** : Facebook est basé principalement sur la relation "amis" entre ses membres. Pour pouvoir consulter le profil d'une personne, il faut lui envoyer une invitation (*demande d'être amis*) et que cette personne l'accepte. Un cas exceptionnel où le profil est public et donc tout le monde pourra le consulter.
- **Fil d'actualité** : Page d'informations réservée au contenu des abonnés et leurs informations partagées, et toutes les nouveautés.
- **Statut** : C'est une case spécifique où l'utilisateur pourra partager, uniquement à ses amis, du texte, images, vidéos et liens.
- **Chat** : Facebook offre aussi un système de messagerie instantanée avec fonctions d'appels audio/vidéo et de partage privé des fichiers.
- **Bouton «j'aime»** : Ou l'émot-icône "cœur", est une caractéristique distinctive de Facebook est son bouton «j'aime» (*en anglais, like*). Quand un utilisateur actionne ce bouton sur une publication veut dire qu'il l'a apprécié.

## I.5.2. Twitter

Twitter [22] est un réseau social généraliste destiné aux grands publics. Il a été créé le 21 mars 2006 par Jack Dorsey, informaticien et entrepreneur américain. Ce réseau social nous permet de suivre nos centres d'intérêt, découvrir ce dont les gens parlent à ce moment, rejoindre des conversations même mondiales et regarder des événements en direct, collaborer avec des personnes créatives et curieuses dans le monde entier.

Twitter est un outil de communication permettant d'envoyer gratuitement des «tweets» aux différents utilisateurs qu'on appelle «twittos».

Un tweet est un message court de 280 caractères maximum que l'utilisateur poste sur son compte afin de le partager avec son groupe pour publier une expérience, idée ou tout sujet d'actualité et du quotidien. Ce réseau social permet à ses utilisateurs de cibler la recherche, par exemple, si on aime le "football", il suffit de taper «football» et on peut alors trouver toutes les personnes ayant ce centre d'intérêt et s'abonner à leurs «tweets». Twitter annonce les tendances du moment, ce qui permet aux entreprises de savoir de quoi les utilisateurs de Twitter parlent à un moment donné.

La spécificité la plus importante de Twitter par rapport à Facebook est que tous les tweets sont publics et peuvent être consultés par tout le monde.

Selon le site officiel de Twitter, ce réseau social a plus de 35 bureaux à travers le monde. Twitter a connu quatre changements de son logo depuis sa création (*figure 1.3*).



Figure 1.3 : Logos de Twitter [21].

Le premier tweet a été envoyé le 21 mars 2006 par Jack Dorsey. Le nombre d'utilisateurs de ce site est arrivé à 313 millions avec un nombre moyen de 500 tweets par jour, il est disponible en 40 langues. [20].

### 1.5.3. YouTube

YouTube [23] est un réseau social généraliste destiné aux grands publics. Il a été créé le 14 février 2005 puis racheté par Google en 2006. YouTube représente la plateforme de diffusion de vidéos la plus connue. Ainsi, il est le troisième site le plus visité après Google et Facebook [24]. Il permet à ses utilisateurs de diffuser, visionner, commenter et partager des vidéos en ligne. Il avait un effectif de 121 employés en 2014 [21]. Il compte actuellement plus de deux milliards d'utilisateurs dans le monde entier, plus de 100 pays, 80 langues et un milliard d'heures regardées par jour [23].

Youtube permet à ses abonnés d'en créer des chaînes (*play listes*) où ils peuvent mettre leurs vidéos d'une façon enchaînée. Aussi, il met à la disposition de ses utilisateurs plusieurs produits comme YouTube Go, YouTube Kids, YouTube Music, YouTube Originals, YouTube Premium, YouTube Studio et YouTube TV.

Plein d'autres fonctionnalités sont embarquées à youtube comme par exemple :

- La possibilité d'aimer ou de détester des vidéos via les deux fameux boutons de notes (*j'aime et je n'aime pas*).
- Permettre aux utilisateurs de s'abonner à nos chaînes YouTube en un seul clic (*via le bouton «S'abonner»*).
- Gagner de l'argent via plusieurs outils mis à la disposition de ses utilisateurs.
- Planifier, diffuser en direct et gérer nos flux vidéo diffusés.
- Avec son outil «Analytics & Reporting» on pourra mieux comprendre nos spectateurs et comment ils interagissent avec nos chaînes et vidéos.
- ..etc.

Youtube a connu cinq changements de son logo depuis sa création (*figure 1.4*).



Figure 1.4 : Logos de Youtube [21].

### I.5.4. Instagram

Instagram [25] est un réseau social généraliste destiné aux grands publics. Il est considéré comme une plateforme en ligne de capture et de partage de photos et vidéos. Il permet à ses utilisateurs de diffuser leurs moments de vie avec leurs amis, familles et collègues de travail. Il a été lancé en octobre 2010 et il a atteint un milliard utilisateurs en 2018 [25], avec une moyenne de 55 millions de photos téléchargées quotidiennement par les utilisateurs, et plus de 16 milliards de photos partagées à ce jour [26]. Avec 25 langues, il est clair qu'Instagram mérite l'attention de la communauté des chercheurs, ce qui est comparable à l'attention accordée à Twitter et à d'autres plateformes de réseaux sociaux [27]. Après son grand succès,



Instagram a été racheté par Facebook pour une somme d'un milliard de dollars américains [28]. La figure I.5 montre le logo d'Instagram.



Figure I.5 : Logo d'Instagram [25].

### I.5.5. Google+

Google+ [29] est un réseau social généraliste destiné aux grands publics. Il a été développé et lancé par Google le 28 juin 2011, il permet à ses utilisateurs de télécharger et partager des images et des vidéos, il a atteint, en 2019, plus de 90 millions d'utilisateurs [20] mais un taux très faible d'utilisation où 90% des sessions utilisateur durent moins de cinq secondes [30]. En octobre 2018 Google annonce dans son blog officiel [30] la fermeture progressive de la version grand public de son réseau social Google+ et définitivement fermée le 2 avril 2019.

Après un audit très approfondi en mars 2018, Google a détecté un bogue dans une des applications publiques (API) de Google+, les utilisateurs ont pu accéder à leurs données de profil et celles de leurs amis en exploitant cette API. Selon lui, ces données sont limitées aux champs statiques et facultatifs du profil Google+, notamment le nom, l'adresse e-mail, la profession, le sexe et l'âge. Elle n'inclut aucune autre donnée publiée ou connectée à Google+ ou à tout autre service, comme les publications, messages, données du compte et numéros de téléphone. Après une analyse détaillée de deux semaines, Google annonçait que les profils de jusqu'à 500 000 comptes Google+ étaient potentiellement affectés et aussi jusqu'à 438 applications ont, peut-être, utilisé cette API [30].

Actuellement, et selon le site officiel de Google, Google+ est uniquement disponible pour les comptes G Suite fournis par une entreprise ou un établissement scolaire. L'application n'est plus compatible avec les comptes personnels (*se terminant généralement par @gmail.com*) depuis le 2 avril 2019.

La figure I.6 montre le logo de Google+ et son alternatif utilisés depuis le 1<sup>er</sup> septembre 2015.

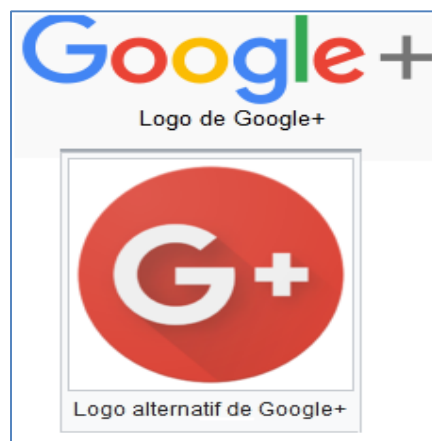


Figure I.6 : Logos de Google+ [21]

### I.5.6. LinkedIn

LinkedIn [31] est un réseau social spécialisé destiné aux professionnels. C'est une plateforme en ligne de communication professionnelle, lancée en 2003. En décembre 2012, Il a enregistré environ 200 millions d'utilisateurs, ce nombre s'était développé à 380 millions en 2015 [20] puis plus de 600 millions en 2019 [2]. Il compte actuellement 170 secteurs d'activités dans plus de 200 pays [21], pour la majorité des professionnels, LinkedIn est un outil très important pour le développement de carrière, en créant un profil nous permettra de construire un réseau de contacts et accéder à d'autres opportunités. Un bon profil augmente notre visibilité et crédibilité sur le web et permettra aux professionnels et entreprises de nous trouver facilement quand ils cherchent à recruter et renforcer leurs équipes. À cet effet, il est vivement recommandé de détailler le plus possible notre profil en tenant en compte nos expériences professionnelles, nos centres d'intérêt et compétences. Avec LinkedIn on pourra publier des offres d'emplois, faire la publicité à nos produits et services, publier des articles et montrer notre savoir et connaissance.

Selon [19], en 2008 LinkedIn été classé bien en dessous de Facebook, au 193<sup>ième</sup> rang des sites attirant le plus de trafic, avec une moyenne d'environ 500 millions de pages vues par mois.

En décembre 2016 et après cette preuve de succès dans le monde professionnel, Microsoft a finalisé son acquisition de LinkedIn, permettant ainsi la fusion entre le meilleur cloud professionnel au monde et le meilleur réseau professionnel au monde [31].

Selon le site officiel de LinkedIn [31] consulté en janvier 2020, il est actuellement disponible en 24 langues entre autres l'arabe, plus de 16 000 employés à temps plein ayant des bureaux dans plus de 30 pays à travers le monde, et encore plus de 675 millions utilisateurs dans le monde et les professionnels s'inscrivent pour rejoindre LinkedIn à un rythme de plus de deux nouveaux membres par seconde, plus de 30 millions d'entreprises représentées sur LinkedIn, plus de 20 millions d'emplois ouverts, 90 000 écoles répertoriées et 35 000 compétences répertoriées. La figure I.7 montre le logo de LinkedIn.



Figure I.7 : Logos de LinkedIn [31]

### I.5.7. Viadeo

Viadeo [32] est notre deuxième exemple des réseaux sociaux spécialisés destinés aux professionnels. Il a été créé en 2004 à Paris (*sous le nom Viaduc*) afin de faciliter la communication et l'interchange entre les professionnels. Pour ses membres, c'est un outil très puissant de gestion de réputation en ligne et de marketing personnel. Chaque utilisateur se crée un Curriculum Vitae en ligne tout en mettant son cursus professionnel, ses savoirs faire, ses diplômes et attestations, sa valeur à rajouter dans les entreprises et surtout ses centres d'intérêt et domaines de compétences. Aussi, un moteur de recherche est mis à la disposition de ses utilisateurs afin qu'ils puissent trouver des offres de recrutement, des formations ainsi que des propositions freelance des différentes entreprises.

Pour ses utilisateurs, Viadeo est l'un des plus importants outils en ligne permettant une comparaison et notation des entreprises afin d'en trouver l'entreprise la mieux adéquate grâce à un tableau qui compare vingt et un (21) critères différents fournissent une image exacte et détaillée de l'entreprise dans laquelle on travaille, ou dans laquelle on envisage de postuler [32]. Ces critères sont divisés en quatre volets: secteur d'activité, l'ampleur de l'entreprise, région et carrière.

Selon le site web officiel de Viadeo (*consulté le 17 février 2020*), plus de 30 miles offres d'emploi ont été déposés jusqu'au maintenant, 5.6 million de membres (*uniquement en France*), plus de 7 millions d'avis sur plus de 130 000 pages entreprises et 3 million visiteurs enregistrés par mois. Viadeo compte, actuellement, 145 employés et elle a été rachetée en 2016 par le groupe Figaro, journal quotidien Français [21].

En résumé, Viadeo nous accompagne tout au long de notre vie professionnelle et nous permet, autant que chercheurs d'emploi, de [32]:

- Postuler à des offres d'emploi;
- Consulter les avis sur notre future entreprise;
- Trouver l'entreprise qui nous correspond le mieux grâce au comparateur des entreprises;
- Partager des informations et suivre l'actualité de notre secteur.

Autant qu'entreprise, Viadeo nous permet de :

- Valorisation des atouts de notre entreprise grâce à des critères de notation détaillés;
- Preuve de la culture d'entreprise présentée sur notre page d'entreprise grâce aux avis rédigés;
- Candidatures ciblées et pertinentes.

Viadeo a connu quatre changements de son logo depuis sa création (figure 1.8).



Figure 1.8 : Logos de Viadeo [32]

### I.5.8. Yammer

Yammer [33] est un réseau social spécialisé destiné aux entreprises créé en septembre 2008. En 2012, il est utilisé par plus de 100 000 entreprises et revendique une base d'utilisateurs de plus de quatre millions de personnes [34] et vu son succès dans le marché des entreprises, il a été racheté par Microsoft en 2014 tout en lui permettant une facile intégration avec les produits d'Office 365.

Yammer est une copie (*clone*) de Twitter pour les entreprises. Il est censé être «un outil pour rendre les entreprises et les organisations plus productives grâce à l'échange de réponses courtes et fréquentes à une question simple: «Partager quelque chose avec mes collègues?»» [32]. Yammer diffère de Twitter à plusieurs égards: seuls les employés possédant une adresse e-mail d'entreprise valide peuvent rejoindre le réseau Yammer d'une entreprise; Yammer n'a pas la limite de 280 caractères sur les messages; Yammer prend en charge la pièce jointe aux messages; et Yammer permet aux utilisateurs de créer des groupes privés ou publics [35].

Yammer est un outil très puissant qui offre à ses utilisateurs la possibilité d'en créer et mettre en place gratuitement un réseau social interne spécifique à leurs entreprises. Avec Yammer, on prendra rapidement de meilleures décisions en communiquant avec les membres de notre équipe. Aussi, il permet aux dirigeants et aux employés de se connecter, de partager et de co-crée une culture avec des groupes ouverts et des événements en direct alimentés par de riches expériences vidéo. Encore, il favorise l'engagement et la communication dans les deux sens afin de créer une culture de transparence et de tenir au courant chaque partie de l'entreprise.

Autant qu'utilisateurs de Yammer, cet outil nous offre plusieurs autres fonctionnalités, entre autres il nous permet de:

- Rapprocher et impliquer les membres de notre organisation.
- Connecter les principaux acteurs ou responsables de haut niveau de notre organisation pour voir la progression de chacun vers un objectif commun.
- Produire des événements en direct et à la demande
- Créer et héberger facilement des réunions publiques, des réunions d'entreprise et des formations.
- Partager des communications enrichies avec des événements incluant jusqu'à 10 000 participants.
- Inviter des personnes à visionner des vidéos et à participer à des discussions sur le web et les applications mobiles.
- Identifier, capturer et présenter dynamiquement les connaissances et les meilleures pratiques de tous les coins de l'entreprise.
- Coopérer en toute sécurité avec nos sous-traitants, clients ou filiales.
- Créer des conversations ouvertes et dynamiques au sein de notre entreprise
- Conserver les documents, les photos et les vidéos en contexte en les joignant à des communications.
- Partager, créer et modifier des fichiers directement avec Office pour le web.
- Il nous permet une Intégration transparente avec Microsoft Teams, SharePoint et le reste de la suite Microsoft 365, y compris une sécurité, une conformité et une gestion de niveau professionnel auxquelles nous pouvons faire confiance.
- Autant qu'administrateur de l'entreprise et avec l'outil statistiques de Yammer, on pourra connaître l'activité de nos utilisateurs, connaître quels utilisateurs sont les plus actifs, analyser ce qu'ils partagent (*publications, likes, etc*) pour chaque utilisateur, etc.

La figure I.9 montre le logo de Yammer.



Figure I.9 : Logo de Yammer [21]

### I.6. Étude comparative

En se basant sur les sites officiels des réseaux sociaux suscités ainsi que d'autres sources, et pour faciliter la compréhension de leurs caractéristiques les-uns aux autres et avoir une vision plus claire sur leurs différences, nous avons préféré de tracer le tableau I.1 comparatif suivant :









Champs comparatifs	Réseaux sociaux personnels					Réseaux sociaux spécialisés		
	Facebook	Instagram	Twitter	Google+	Youtube	professionnels		entreprise
						LinkedIn	Viadeo	Yammer
1 Logos								
2 Date et lieu de création	2004/USA	2010/USA	2006/USA	2011/USA	2005/USA	2003 en USA	2004 en France	2008 en USA
3 Opérationnel?	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui
4 Nombre d'utilisateurs actifs (2016)	1.86 milliard	600 millions	317 millions	1,168 milliards	Un milliard	106 millions	65 millions	12 millions
5 Langues (2016)	96 langues	25 langues	35 langues	42 langues	76 langues	23 langues	Bilingue (Français, Anglais)	Bilingue (Anglais, Espagnol)
6 Contenu	Blogs, vidéos, photos, documents	Photos avec des filtres, vidéos de 60 seconds	Textes, blogs, vidéos, photos	Textes, blogs, vidéos, photos	Textes, vidéos, Liens documents	Contenu de nature professionnelle.	Contenu de nature professionnelle.	Textes, photos, documents, etc.
7 Type de relation	Symétrique /Asymétrique	Asymétrique	Asymétrique	Symétrique	Asymétrique	Symétrique	Symétrique	Symétrique

Tableau I.1: Comparative, selon quelques paramètres, entre les réseaux sociaux échantillon de notre étude

## I.7. Avantages et inconvénients des réseaux sociaux

### 1.7.1. Avantages des réseaux sociaux

Les réseaux sociaux ont plein d'avantage dans notre vie personnelle ou professionnelle, on cite à titre d'exemple les points suivants :

- Ils sont devenus un moyen gratuit et facile de communication avec les autres. On peut communiquer avec des amis ou des membres de la famille de partout dans le monde.
- Les réseaux sociaux peuvent être considérés comme un outil permettant à ses abonnés de maintenir leurs (*virtuelles*) rencontres avec leurs familles et amis, de communiquer avec eux des idées, des vidéos et photos de souvenir, etc.
- Ils sont vus, dans le domaine de l'économie, comme un intermédiaire entre les entreprises et leurs clients afin de les rapprocher les uns aux autres et les mettre en contact d'une façon synchrone et des fois asynchrone, ainsi pour les recommander des produits et des services.
- Ils donnent la possibilité aux gens qui travaillent en collaboration de mieux avancer dans leurs tâches quotidiennes, de discuter, de critiquer les uns et de lever les réserves des autres, etc.
- Les réseaux sociaux est un outil très rapide permettant un accès accru aux évènements ainsi que leur propagation.
- Les réseaux sociaux peuvent être vus comme un endroit de solidarité, comme la transmission d'informations sur la recherche des individus disparus, la donation (*monnaie, médicaments, sang, etc*) et les différents actes de soutien à ceux ayant besoin, etc.
- ..etc.

### 1.7.2. Inconvénients des réseaux sociaux

Depuis de nombreuses années, notre population est perpétuellement confrontée à divers types d'infections sociales entre les utilisateurs à travers les réseaux sociaux, ce qui cause des dangers et des risques imprévus, malgré le nombre important des avantages de cet outil de communication, partage et collaboration, on trouve aussi d'autres points négatifs comme par exemple :

- En mettant en ligne ses informations (*personnelles ou non*), l'utilisateur perdra tout droit de propriété sur elles.
- Les utilisateurs peuvent se retrouver surchargés et alourdis de plusieurs événements sans aucune importance.
- publier et divulguer trop d'informations personnelles (*la géolocalisation, la reconnaissance faciale, etc.*), volontairement ou non, implique la mise à la disposition du grand public son identité numérique, la chose qui peut se retourner contre soit.
- Les utilisateurs de ces réseaux sociaux peuvent devenir addictifs et avoir des mauvaises habitudes en restant de longues durées derrière leurs écrans, la chose qui pourra influencer négativement sur leurs relations familiales, études et devoirs envers les autres.
- Des fois, les utilisateurs des réseaux sociaux se permettent de dépasser leurs limites en parlant de la dignité et la vie privée des autres, la chose qui pourra avoir des conséquences assez graves sur les deux parties.
- L'inconvénient le plus flagrant quant à l'utilisation des réseaux sociaux est le risque de vol d'identité et la fraude, des virus, vers, hackings, violation de vie privée...etc. Ainsi, les renseignements personnels des utilisateurs peuvent être utilisés par des personnes douteuses pour des activités illégales.
- Plusieurs utilisateurs s'inscrivent via une fausse identité et de faux profils, la chose qui pourra tromper les autres en se mettant en relation avec eux. À cet effet, et pour mieux se protéger contre ce genre de «arnaqueurs», il faut être sûr de l'identité de la personne qu'on veut se mettre en relation avec elle.
- Avec nos «smartphones» mobiles, il est devenu très facile d'y accéder aux réseaux sociaux et à tout moment. À cet effet, dans le cas où les utilisateurs exagèrent, ils se trouvent vivres dans le virtuel en écartant tous les moments réels ainsi que la famille et amis.
- Trop de comptes dans différents réseaux sociaux tue son efficacité, il est presque impossible d'en créer autant de comptes dans plusieurs réseaux sociaux et qu'on reste connecté et répond aux messages et sollicitations de nos amis, une page qui ne sera pas alimentée régulièrement ne fera pas l'objectif de son existence : se communiquer.
- Les réseaux sociaux ne servent pas uniquement à mettre les gens en communication et les rapprocher les-uns aux autres, ils se cachent derrière une vérité très dangereuse où ils collectent toutes nos informations personnelles ou professionnelles, nos mouvements et positions géographiques, nos courses, achats et même nos commentaires, nos habitudes et tous ce qui concerne notre vie privée, toutes ces informations se finissent, un jour ou un autre, dans les mains de tierces personnes qui vont les utiliser d'une manière ou d'une autre.
- .. etc.

Ainsi, le tableau I.2 suivant donne une synthèse des avantages et inconvénients jugés les plus importants de chacun de ces réseaux sociaux.



	<b>Réseau social</b>	<b>Avantages</b>	<b>Inconvénients</b>
<b>1</b>	<b>Facebook</b>	<ul style="list-style-type: none"> <li>- Facebook est une plateforme rapide, pratique, facile, maîtrisable, accessible, interface adéquate et jamais plantée.</li> <li>- Un tableau de bord riche en fonctionnalités.</li> <li>- Possibilité d'en développer des applications spécifiques.</li> <li>- Leur système de «cuisson inversée» utilise la vision par ordinateur, une technologie qui extrait des informations à partir d'images et de vidéos numériques pour donner aux ordinateurs un niveau élevé de compréhension du monde visuel.</li> <li>- Amélioration des appels vidéo en ajoutant (<i>perspective</i>) de nouveaux contenus et fonctionnalités (<i>Portal Mini, Portal, Portal + et Portal TV</i>) plus amusants et utiles.</li> <li>- Agir contre la fraude publicitaire ou Facebook a déposé une plainte en Californie (<i>Décembre 2019</i>) contre une entité et deux personnes pour avoir enfreint ses conditions et politiques de publicité.</li> <li>- Les gens collectent plus de 2 milliards de dollars pour des causes sur Facebook (<i>Septembre 2019</i>).</li> <li>- Etc.</li> </ul>	<ul style="list-style-type: none"> <li>-Suite à l'enquête (<i>Facebook's Cambridge Analytica Scandal [36]</i>) en 2018, Facebook a monté un mur de confidentialité mais il a donné une ouverture pour les géants de la technologie comme Microsoft, Amazon et d'autres avec un accès beaucoup plus large aux données de ses utilisateurs. Des données personnelles étaient collectées à partir de profils de millions d'utilisateurs ont été exploitées pour microcibler des électeurs anglais (<i>Brexit ou British Exit</i>) et américains (<i>élections de Trump</i>).</li> <li>-La permission d'utilisation des faux comptes aux utilisateurs a permis d'en avoir plusieurs affaires d'arnaque sur facebook.</li> <li>-Sa large et massive utilisation dans le monde entier lui ont créé aux étudiants et élèves une forte dispersion de l'attention que peut amener cette plateforme dans le cadre de leurs études.</li> <li>-Même facebook est gratuit, mais il exagère dans son côté publicitaire afin d'en augmenter ses revenus. La chose qui va réduire son engagement envers la communauté.</li> </ul>
<b>2</b>	<b>Instagram</b>	<ul style="list-style-type: none"> <li>- Un réseau émotionnel, rapide, esthétique, facile à exploiter, interactif et amusant.</li> <li>- Discussions en vidéo via son application «Threads Instagram».</li> <li>- Programme de vérification des faits à l'échelle mondiale et combattre la désinformation.</li> <li>- Progrès en IA pour reconnaître différentes formes d'intimidation en ligne.</li> <li>- Empêcher les comportements inauthentiques où une action en</li> </ul>	<ul style="list-style-type: none"> <li>-Gratuit en général mais pour avoir un compte plus attrayant il faut dépenser davantage.</li> <li>-la perte de droits d'auteur.</li> <li>-Instagram est disponible seulement pour les systèmes d'exploitation iOS (<i>iPhone d'Apple</i>), Android (<i>de Google</i>) et Windows Mobile (<i>de Microsoft</i>), pour ceux travaillant sur BlackBerry, Linux ou tout autre OS doivent utiliser la version Web qui n'offre pas les mêmes fonctionnalités que l'application</li> </ul>



		<p>justice a été lancée en Avril 2019 contre une entreprise et un individu pour avoir fourni de faux likes, vues et followers aux utilisateurs.</p> <ul style="list-style-type: none"> <li>- La suppression de plusieurs pages, groupes et comptes pour un comportement inauthentique.</li> <li>- De nouveaux outils pour gérer son temps sur Instagram (<i>rappels quotidiens et nouvelles façons de limiter les notifications</i>).</li> <li>- Etc.</li> </ul>	<p>mobile.</p> <ul style="list-style-type: none"> <li>-Manque d'un service d'analyse où peu d'informations concernant nos publications, la chose qui rend difficile l'évaluation de nos audiences.</li> <li>-Pas de possibilité de programmer des évènements ou des publications.</li> </ul>
3	Twitter	<ul style="list-style-type: none"> <li>- L'avantage le plus important de Twitter est son immédiateté et l'instantanéité, toute personne désirant l'actualité n'a qu'à se connecté à son compte.</li> <li>- Twitter est une la vitrine de ce qui se passe dans le monde et des sujets de conversation du moment.</li> <li>- Vu que sur Twitter chaque mot, photo, vidéo et abonné peut avoir un impact, il est outillé d'un moyen de statistique assez performant afin d'en analyser nos Tweets et connaître nos abonnés.</li> <li>- Twitter est blindé de plein de mesures de sécurité entre autre les détecteurs de spam, des faux comptes et des liens vers des sites de hameçonnage (<i>phishing</i>), de préservation de la confidentialité, de signalement des comptes usurpant les identités et tous comportements inappropriés.</li> </ul>	<ul style="list-style-type: none"> <li>- Une étude effectuée par le centre de recherche américain Pew Research Center [37] a prouvé que les rebots (<i>ou les 'bots'</i>) ont une grande influence sur Twitter, ce sont des comptes automatisés programmés pour la diffusion des informations sans aucune intervention humaine. L'étude a découvert que plus de 66% des liens publiés sur Twitter parviennent des robots. Ce mécanisme est des fois ingérable et il pourra avoir des répercussions graves sur la société.</li> <li>- La limite de taille des messages qu'on pourra envoyer avec Twitter est 280 caractères, la chose qui pourra gêner ses utilisateurs.</li> <li>- Twitter affiche tous les tweets dans un ordre chronologique, la chose qui pourra les mélanger et donc l'utilisateur va rater quelques-uns.</li> <li>- ..etc.</li> </ul>
4	Google+	<ul style="list-style-type: none"> <li>- Ca ne nécessite pas la création d'un compte pour se connecter à Google+ pour tout individu possédant un compte Gmail.</li> <li>- Toute publication sur Google+ peut se classer dans les résultats de recherche de Google, la chose qui pourra influencer le classement sur ce moteur de recherche.</li> <li>- L'authoship est une fonctionnalité très intéressante de Google+ qui</li> </ul>	<p>La version grand public de ce réseau social n'est plus disponible actuellement, mais on pourra citer les points faibles suivants qui existaient pendant sa mise en service:</p> <ul style="list-style-type: none"> <li>- Google+ a utilisé autres produits de la même famille « GOOGLE » comme Youtube et Play Store. Si on veut se connecter à notre compte Youtube , par exemple, il nous demande d'utiliser celui de Google+,</li> </ul>

		<p>permet à ses utilisateurs de relier leur profil à leur site web ou blog, la chose qui pourra augmenter la visibilité de ces derniers (<i>sites</i>), ainsi que la photo de profil sera visible dans le moteur de recherche.</p> <ul style="list-style-type: none"> <li>- Possibilité d'interaction et d'intégration à d'autres plateformes comme Youtube.</li> <li>- On pourra segmenter nos publications grâce à l'outil des cercles où chacun à ses destinataires et types de profils qu'on choisisse.</li> <li>- Etc.</li> </ul>	<p>c'est comme s'il est imposé aux utilisateurs d'en avoir compte Google+. Puis toutes nos publications dans ces réseaux sociaux seront signées de notre compte Google+.</p> <ul style="list-style-type: none"> <li>- Google+ s'est permis d'en afficher nos commentaires (<i>avec notre nom et avatar</i>) sur les différents produits qu'on commente et annote, c'est comme s'il nous utilise pour promouvoir sa publicité sur ces produits. La chose que beaucoup de gens n'ont pas aimé et veulent rester discrets.</li> <li>- Google+ n'a pas de notion d'anonymat, ce dernier peut avoir, des fois, des effets négatifs comme l'agression verbale ou écrite, mais il pourra être vu comme une question de survie comme le cas des opposants politiques dans des régions où la liberté d'opinion politique est menacée ou contestée.</li> <li>- Etc.</li> </ul>
<p><b>5</b></p>	<p><b>Youtube</b></p>	<ul style="list-style-type: none"> <li>- Youtube est totalement gratuit, pas d'abonnements mensuels ni annuels.</li> <li>- YouTube est le leader de partage, diffusion et consultation de la vidéo dans le monde.</li> <li>- YouTube comme étant un moteur de recherche parmi les plus importants dans son domaine.</li> <li>- Si veut parler de nous, de notre entreprise, produit, service ou même passer une formation, il est plus facile de s'exprimer en vidéo qu'avec un texte de plusieurs pages, c'est le point fort le plus important de Youtube.</li> <li>- Youtube permet à ses utilisateurs d'en créer des chaînes qui regroupent des vidéos enchainées et qu'on peut les retrouver facilement.</li> <li>- Le gain de l'argent avec plusieurs moyens: devenir partenaire en mettant notre publicité en ligne et</li> </ul>	<ul style="list-style-type: none"> <li>- Dans Youtube, il est difficile d'établir une prise de contact.</li> <li>- Plusieurs fonctionnalités sont payantes.</li> <li>- L'inconvénient major de Youtube et qu'il n'a pas de contrôle parental et les enfants sont libres de choisir ce qu'ils veulent voir.</li> <li>- Avis de violation de droits d'auteur et de copyright limités.</li> <li>- Aucune restriction sur les commentaires (<i>des fois offensives et agressives</i>).</li> <li>- Pour ses utilisateurs, une vidéo à partager doit être bien fait, à cet effet, une certaine compétence quant au montage des vidéos est nécessaire.</li> <li>- La plus part des commentaires et avis ne sont pas toujours constructifs</li> <li>- Impossibilité de modifier ou corriger une vidéo sans perdre ses statistiques.</li> <li>- Publicité intégrée dans nos vidéos</li> </ul>

		<p>vendre nos produits et services plus facilement, utiliser le «financement participatif des fans» en demandant à notre public de nous soutenir via leur abonnement à nos chaînes et leurs mentions «j'aime» et partage, autoriser notre contenu aux médias ou travailler avec les marques comme influenceur..etc.</p> <ul style="list-style-type: none"> <li>- Etc.</li> </ul>	<p>d'une façon exagérée que ce soit au début, en cours de lecture ou à la fin et sans même pas demander notre autorisation,</p> <ul style="list-style-type: none"> <li>- On aurait aimé avoir un choix de protection par mot de passe de nos vidéos Youtube.</li> <li>- ..etc.</li> </ul>
<b>6</b>	<b>LinkedIn</b>	<ul style="list-style-type: none"> <li>- LinkedIn est devenu un outil incontournable dans le domaine des ressources humaines et un moyen très efficace de recrutement.</li> <li>- Possibilité de navigation en mode privé (<i>Ex. un employeur qui cherche l'élite à recruter..</i>) ou public.</li> <li>- Possibilité de connaître l'identité de ceux ayant consulté (<i>en mode public</i>) son profil.</li> <li>- Possibilité d'autoriser/interdire des utilisateurs à consulter son profil, activités et coordonnées personnelles.</li> <li>- Possibilité de création et programmation des événements.</li> <li>- Possibilité de consultation des offres d'emploi sans avoir un compte.</li> <li>- Les discussions ayant plus de succès sont affichées en premier.</li> <li>- ..etc</li> </ul>	<ul style="list-style-type: none"> <li>- Plusieurs fonctionnalités sont payantes, comme par exemple si on veut faire des appels audio ou vidéo, LinkedIn propose des formations payantes qui peut arriver jusqu'à 24,78 EUR/mois.</li> <li>- Pas possible de connaître l'identité de ceux ayant consulté (<i>en mode privé</i>) notre profil.</li> <li>- Pour contacter des personnes il faut en créer un compte et se connecter.</li> <li>- Plusieurs offres sont proposées pour avoir un compte premium : Career €24.78/mois, Business €37.18/mois, Sales €49.58/ mois et Hiring (<i>Recruiter</i>) €74.37/mois. Ces tarifs sont excessivement chers par rapport beaucoup d'autres plateformes de même genre.</li> <li>- ...etc</li> </ul>
<b>7</b>	<b>Viadeo</b>	<ul style="list-style-type: none"> <li>- Viadeo est un outil très efficace pour les consultants freelance, il leur permet de se faire connaître et de trouver des clients.</li> <li>- Beaucoup de chercheurs d'emploi le privilégient par rapport à d'autres plateformes.</li> <li>- Très facile dans son volet de gestion du profil, on pourra rajouter aisément nos expériences, compétences et langues maîtrisées, etc.</li> </ul>	<ul style="list-style-type: none"> <li>- Plusieurs options sont payantes et il faut en avoir un compte premium pour voir qui a visité notre profil, contacter tous les membres, utiliser tous les critères de recherche et personnaliser notre profil, etc.</li> <li>- Pour contacter des personnes il faut en créer un compte et se connecter.</li> <li>- L'abonnement premium est de 8,95 €/mois et 19,90€/mois pour un compte freelance (<i>avoir la priorité dans les résultats de recherche,</i></li> </ul>

	<ul style="list-style-type: none"> <li>- Possède un moteur de recherche et de filtrage très poussé, on pourra trouver des personnes, entreprises, communautés ou des annonces, etc.</li> <li>- La possibilité d'en créer des évènements et suivre leurs participants.</li> <li>- Possibilité de consultation des offres d'emploi sans avoir un compte.</li> <li>- Viadeo est doté de l'application "Let's Meet". qui permet à ses professionnels (<i>entreprises, clients, chercheurs d'emploi, partenaires..</i>) d'une même ville de se rencontrer physiquement grâce à la géolocalisation.</li> <li>- ..etc</li> </ul>	<p><i>proposer notre expertise en répondant directement aux missions et ne payer aucune commission sur les missions décrochées).</i></p> <ul style="list-style-type: none"> <li>- Affichage chronologique des messages sans aucun paramètre d'importance ou de succès.</li> <li>- Viadeo est utilisé beaucoup plus en France et il lui reste beaucoup d'effort et de temps pour atteindre un niveau international.</li> <li>- Etc.</li> </ul>
8	<p><b>Yammer</b></p> <ul style="list-style-type: none"> <li>- Yammer est un système collaboratif et d'intelligence collective permettant aux employés une large diffusion de leurs questions et en avoir le maximum de réponses.</li> <li>- Un outil important pour trouver de l'aide au sein de l'entreprise.</li> <li>- Avec Yammer, en évitant la circulation de l'information via la hiérarchie administrative, l'envoi et l'arrivée de l'information est tellement rapide.</li> <li>- Être au courant de ce qui se passe au niveau de l'entreprise et les sujets d'actualité.</li> <li>- Simple d'usage.</li> <li>- Discussion instantanée sur son mur.</li> <li>- Partage de tout type de fichiers: vidéo, texte, liens...etc.</li> <li>- Possibilité de donner accès aux personnes externes : clients, partenaires..etc.</li> <li>-</li> </ul>	<ul style="list-style-type: none"> <li>- Plusieurs options payantes.</li> <li>- Inciter les employés à s'inscrire et utiliser ce moyen technologique au sein de l'entreprise est un défi en soi pour les administrateurs et qui requiert la coopération tout le monde.</li> <li>- Beaucoup de responsables refusent de l'utiliser et circuler leurs documents avec la hiérarchie à cause de la confidentialité!</li> <li>- La maîtrise de son côté sécurité est primordiale pour les administrateurs afin d'en éviter tout genre d'attaques ou fuite des informations sensibles qui concernant l'entreprise.</li> </ul>

**Tableau I.2: Avantages et inconvénients des réseaux sociaux échantillon de notre étude**

### **I.8. Modélisation des Réseaux sociaux**

Pour étudier et analyser tout réseau social, nous devons le modéliser par un format que la machine peut comprendre et manipuler, il s'agit donc du format graphique.

La première fois que les réseaux sociaux sont présentés graphiquement date de l'année 1933 par Jacob Moreno [38], où les individus étaient représentés par des points et les relations entre eux représentées par des lignes. Cette visualisation graphique est l'un des processus de représentation de réseau structurel et d'analyse visuelle les plus utilisés. Différentes couleurs, nuances, symboles et tailles ont été utilisés pour spécifier les propriétés et les différents types d'acteurs et relations [39].

Dans cette section, nous présentons les notions de base sur la théorie des graphes afin de faciliter la compréhension des chapitres suivants.

### Définition 1 : Graphe

Pour chaque réseau social, on lui associe un graphe  $G$  défini par le couple  $(V, E)$  où  $V$  (*pour Vertices*) est un ensemble de points (*appelés sommets ou nœuds*) et  $E$  (*pour Edges*) l'ensemble des liens reliant certains sommets entre eux. Les sommets représentent les individus de notre réseau social, alors que les liens décrivent les relations entre eux.

### Définition 2 : Graphe non orienté

Dans ce type de graphes, la relation entre les sommets est symétrique, l'ensemble  $E$  représente l'ensemble de paires de sommets nommées **arêtes**. Un exemple concret de ce type de réseaux est Facebook où la relation d'amitié est symétrique.

### Définition 3 : Graphe orienté

Dans ce type de graphes, la relation entre les sommets est asymétrique, l'ensemble  $E$  représente l'ensemble de paires de sommets nommées **arcs**. Les extrémités de l'arc sont appelées extrémité initiale (*départ*) et extrémité finale (*arrivé*). Un exemple concret de ce type de réseaux est Twitter où la relation d'abonnement est asymétrique.

Un graphe orienté (resp. non orienté) est dit d'ordre  $n$  et d'une taille  $m$  si :

- $n$  : représente le nombre de ses sommets :  $n = |V|$ .
- $m$  : représente le nombre de ses arcs (resp. d'arêtes) :  $m = |E|$ .

Une boucle est une arête (*ou arc*) d'un graphe ayant pour extrémités le même sommet. Les boucles sont notamment interdites dans les graphes simples.

### Définition 4 : Adjacence

Deux sommets  $v$  et  $u$  sont adjacents s'ils sont des extrémités d'une même arête du graphe, c'est-à-dire :  $\forall u \in V, \forall v \in V, u \text{ est adjacent à } v \leftrightarrow \exists e \in E, e = (v, u)$

### Définition 5 : Voisinage

Le voisinage d'un nœud correspond à l'ensemble de tous ses nœuds adjacents. Autrement dit, l'ensemble des voisins d'un sommet  $v$ , notée **Vois(v)**, est défini comme :

$$\text{Vois}(v) = \{ \mu \in V : (v, \mu) \in E \}.$$

### Définition 6 : Demi-degré extérieur, Demi-degré intérieur

Dans le cas d'un graphe orienté, le demi-degré extérieur d'un sommet  $v$  (*noté  $d^+(v)$* ) est égal au nombre d'arcs ayant  $v$  comme nœud initial, et le demi-degré intérieur d'un sommet  $v$  (*noté  $d^-(v)$* ) est égal au nombre d'arcs ayant  $v$  comme nœud terminal.

**Définition 7 : Degré d'un sommet :**

Le degré d'un sommet  $v$  d'un graphe non orienté est le nombre d'arêtes incidentes à  $v$ , on note donc :  $d(v) = |\text{Vois}(v)|$ .

Le degré d'un sommet  $v$  d'un graphe orienté est la somme de ses deux demi-degrés extérieure et intérieure, c'est-à-dire :  $d(v) = d^+(v) + d^-(v)$ .

**Définition 8 : Incidence**

Une arête est incidente à un sommet si ce dernier est une extrémité pour cette arête.

C'est-à-dire :  $\forall e \in E, \exists v \in V, e \text{ est incident à } v \leftrightarrow \exists u \in V, e = (v, u)$

**Définition 9 : Chaîne**

On appelle chaîne, une suite de sommets adjacents, c'est à dire une suite de sommets reliés par des arêtes. La longueur de la chaîne est le nombre d'arêtes visitées pour passer du premier sommet de la chaîne au dernier.

**Définition 10 : Chemin**

Un chemin entre deux sommets  $v$  et  $\mu$ , est une séquence d'arcs à traverser pour aller de  $v$  à  $\mu$ . On dit, donc, que  $v$  est le début (*l'origine*) de ce chemin et  $\mu$  est sa fin (*l'extrémité finale*). La longueur de ce chemin est le nombre des arcs qu'il comporte.

**Définition 11 : Composante connexe**

Une composante connexe  $C$  d'un graphe  $G$  est un sous-ensemble **maximal** de sommets tel qu'entre tout couple de sommets  $(v, \mu)$  il existe une chaîne, c'est-à-dire :

- $\forall v \in C, \forall u \in C$ , il existe une chaîne reliant  $v$  à  $u$ .
- $\forall v \in C, \forall \gamma \in V \setminus C$ , il n'existe pas de chaîne reliant  $v$  à  $\gamma$ .

**Définition 12 : Graphe connexe**

Un graphe  $G = (V, E)$  est connexe s'il est possible, à partir de n'importe quel sommet, de rejoindre tous les autres. Un graphe non connexe se décompose en composantes connexes.

**Définition 13 : Graphe complet**

Un graphe  $G = (V, E)$  est dit complet, si tous ses sommets sont adjacents deux à deux, c'est-à-dire que :  $\forall (\mu, v) \in V^2, (\mu, v) \in E$ .

**Définition 14 : Matrice d'adjacence**

Pour effectuer des traitements d'analyse sur les graphes et comparer leurs structures, il est essentiel de les modéliser par un format accessible aux ordinateurs.

Une matrice d'adjacence  $A$  d'un graphe  $G$  d'ordre  $n$  est une matrice carrée dont les lignes et les colonnes représentent les sommets de  $G$ . Cette matrice  $A (n,n)$  est binaire telle que :

$$A[i,j] = \begin{cases} 1 & \text{s'il existe une arête entre le sommet } V_i \text{ et le sommet } V_j, \text{ donc, } (V_i, V_j) \in E \\ 0 & \text{sinon} \end{cases}$$

Dans le cas d'un graphe non orienté (figure I.10), la matrice d'adjacence est une matrice symétrique alors que la matrice d'adjacence d'un graphe orienté (figure I.11) n'est souvent pas symétrique. Dans une matrice d'adjacence d'un graphe orienté, les lignes représentent les nœuds de départ des arcs du graphe et les colonnes représentent les nœuds d'arrivée.



Figure I.10 : Représentation graphique et matricielle d'un graphe non orienté.



Figure I.11 : Représentation graphique et matricielle d'un graphe orienté.

### Définition 15 : Matrices d'incidence

La matrice d'incidence est une autre représentation matricielle des graphes et qui exploite la relation d'incidence entre arêtes et sommets. Soit  $G$  un graphe orienté sans boucle  $G=(V,E)$  comportant  $n$  sommets  $x_1, \dots, x_n$  et  $m$  arêtes  $e_1, \dots, e_m$ .

On appelle matrice d'incidence (aux arcs) de  $G$  la matrice  $M$  de dimension  $n \times m$  telle que :

$$M [i, j] = \begin{cases} +1 & \text{si } x_i \text{ est l'extrémité initiale (source) de } e_j. \\ -1 & \text{si } x_i \text{ est l'extrémité terminale (cible) de } e_j. \\ 0 & \text{sinon.} \end{cases}$$

Pour un graphe non orienté sans boucle, la matrice d'incidence (aux arêtes) est définie par:



$$M[i, j] = \begin{cases} 1 & \text{si } x_i \text{ est une extrémité de } e_j. \\ 0 & \text{sinon} \end{cases}$$

**Définition 16 : Listes d'adjacence**

Une liste d'adjacence est un tableau (ou une liste chaînée) contenant autant de cases que le nombre de sommets du graphe représenté. Chaque case de ce tableau pointe sur une liste de sommets qui représente les successeurs du nœud correspondant à la case considérée. Dans une liste d'adjacence les sommets ne sont pas forcément consécutifs.

**Définition 17 : Listes d'incidence**

Dans une liste d'incidence chaque case du tableau pointe sur la liste d'arêtes incidentes au sommet considéré.

La figure 1.12 illustre l'application de ces représentations mathématiques sur un même exemple.

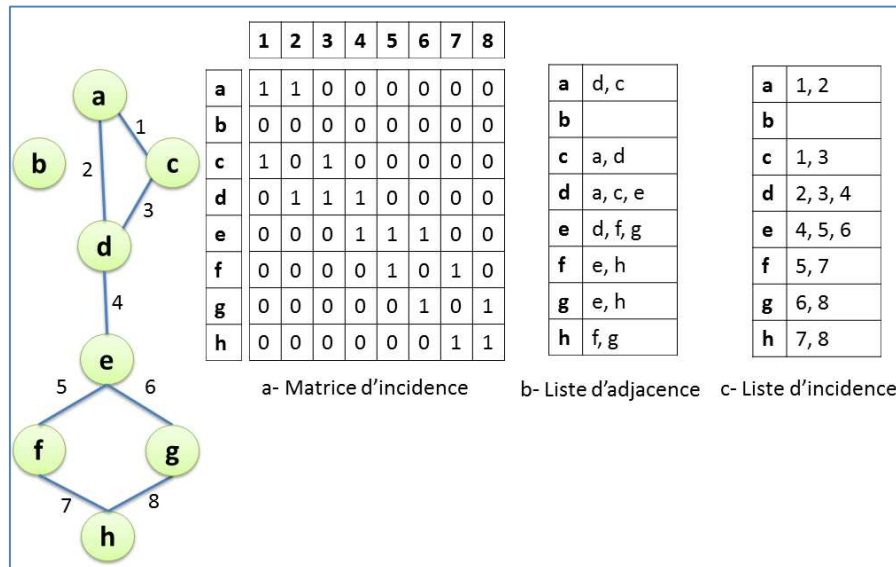


Figure 1.12 : Autres représentations mathématiques que la matrice d'adjacence.

Les représentations mathématiques des graphes sont différentes en terme d'espace mémoire et du temps de réponse. Dans les représentations matricielles nous sommes obligés de mémoriser non seulement l'existence de la relation entre les individus mais aussi son absence entre les autres individus, ce qui impose l'occupation inutile de plusieurs cases mémoires par des valeurs nulles. Pour cela, la représentation en liste chaînées est plus économique en termes d'espace mémoire, alors que le parcours de la liste nécessite un temps assez important de réponse.

Dans la représentation matricielle, la méthode de matrice d'adjacence est plus économique, en terme d'espace mémoire à allouer, que celle d'incidence. Elle a une taille carrée relative au nombre de nœuds alors que la matrice d'incidence a une taille très variée en



fonction de nombre de nœuds et d'arêtes. Ainsi, le changement fréquent de la taille, dû aux apparitions/disparitions d'arêtes, influence négativement sur le temps de réponse des algorithmes manipulant ces structures. Pour la suite et dans le cadre de notre travail, les concepts et les métriques seront calculés à partir de la matrice d'adjacence associée au réseau à étudier.

### Conclusion

Les réseaux sociaux ont toujours été présents dans la vie sociale de l'être humain. Ils sont devenus de plus en plus un endroit virtuel où les membres inscrits se communiquent entre eux d'une façon libre et non contrôlable. Dans ce chapitre nous avons présenté les réseaux sociaux en donnant quelques définitions, leurs caractéristiques et leurs intérêts, ensuite nous avons choisis et cité quelques exemples de réseaux sociaux les plus connus ainsi de tracer un tableau comparatif entre eux et nous avons terminé par présenter les avantages et les inconvénients de cette notion des réseaux sociaux d'une façon générale puis pour de chacun des réseaux cités.

Pour étudier, analyser et manipuler les réseaux sociaux, on doit passer par une modélisation ou représentation graphique, mathématique et informatique. À cet effet, nous avons détaillé, dans ce chapitre, cette notion de modélisation tout en donnant la définition de quelques concepts de base ayant relation avec notre travail.

Les individus d'un réseau social se regroupent, des fois, dans des groupes autour d'un thème fédérateur ou centre d'intérêt commun, ces groupes sont nommés «communautés». Ces dernières s'évaluent dans le temps à cause des changements (*des intégrations/ des disparitions*) effectués par les individus. Dans le chapitre suivant, nous allons entamer le noyau de notre thèse qui représente un état de l'art sur cette notion de communautés et les différentes approches et algorithmes de leurs détections dans les réseaux sociaux.

## Chapitre 2: Détection des communautés dans les Réseaux Sociaux

<b>Introduction</b> .....	42
<b>II.1. Communautés, définitions</b> .....	42
<b>II.2. Détection des communautés, objectifs</b> .....	43
<b>II.3. Chevauchement des communautés</b> .....	44
<b>A. Communautés chevauchantes</b> .....	44
<b>B. Communautés non chevauchantes</b> .....	45
<b>II.4. Types de communautés</b> .....	46
<b>A. Communautés à sens forts</b> .....	46
<b>B. Communautés à sens faible</b> .....	47
<b>II.5. Qualité d'une distribution communautaire</b> .....	48
<b>II.6. Stabilité des algorithmes de détection de communautés</b> .....	49
<b>II.7. Historique et travaux connexes</b> .....	50
<b>II.8. Principales approches de détections des communautés dans les SN</b> .....	52
<b>A. Algorithme de Louvain</b> .....	54
<b>B. Algorithme de Newman et Girvan</b> .....	56
<b>II.9. Étude comparative entre quelques algorithmes de détection de communautés</b> .....	57
<b>A. Comparaison qualitative</b> .....	57
<b>B. Comparaison quantitative</b> .....	58
<b>Conclusion</b> .....	58

La détection des communautés dans les réseaux sociaux est un axe très important dans le domaine de l'analyse de ces derniers. À cette fin, plusieurs approches et modèles ont été proposés dans la littérature et chacun avec ses spécificités, méthodes, complexités, avantage et inconvénients. Dans ce chapitre nous donnons un aperçu de l'état de l'art, les différentes approches proposées et quelques algorithmes les plus importants, ainsi que la notion de modularité et la stabilité des algorithmes de détection de communautés.

## Introduction

Le problème de la détection des distributions communautaires dans les réseaux sociaux (SN) a été étudié par un très grand nombre de chercheurs à travers le monde et de manière très approfondie ces dernières années. La chose qui prouve que ce sujet a attiré l'attention de nombreux chercheurs et qu'il se classe bien par rapport à de nombreux autres sujets dans le domaine d'analyse des réseaux sociaux (SNA).

Au cours de la dernière année de 2019, plusieurs modèles et approches ont été proposés pour détecter l'ensemble des communautés au sein des réseaux sociaux. Nous citons [40] qui a proposé deux nouveaux algorithmes, le premier basé sur la dynamique de consensus et le second son extension via les modèles de leader. Dans [41], les auteurs ont proposé un nouvel algorithme basé sur des attributs de nœuds pour détecter les communautés de réseau social, et dans [42] l'approche est basée sur la centralité de la périphérie locale. Dans [43], les auteurs utilisent un algorithme de détection de communauté d'évolution multi-objectif exploitant la centralité des nœuds. En outre, les auteurs de [44] ont présenté de nouvelles implémentations GPU (*Graphics Processing Units*), le parallèle dynamique et le parallèle imbriqué hybride pour améliorer la détection des communautés. Aussi, nous avons le travail de [45] qui propose une méthode incrémentale pour détecter l'ensemble des communautés dans les réseaux sociaux en évolution dynamique.

### II.1. Communautés, définitions

Dans son ouvrage [46], l'auteur a défini la notion de communauté par un ensemble d'individus connectés entre eux de manière à connaître l'information qui possède chacun. Aussi, l'auteur de [47] a expliqué comment l'appartenance d'un individu quelconque à une communauté et la redondance de ses contacts facilitent à la fois sa confiance mais aussi sa sanction. Au sein d'une communauté, les fautes et les mauvais comportements d'un individu se retransmettent plus vite à ses contacts directs, la chose qui pourrait amplifier la possibilité de sa sanction et isolement par la perte de confiance. En terme économique et au sein des entreprises, cette notion de communauté est équivalente à un ensemble d'employés ayant plus de communications en améliorant leur productivité.

Dans le domaine du SNA, la notion de communauté pourra être définie de deux manières :

**II.1.1. Définition sémantique:** ensemble de nœuds ayant le même centre d'intérêt ou le même profil.

**II.1.2. Définition structurelle:** ensemble de nœuds étroitement connectés mais ayant également des liens lâches avec le monde extérieur.

Dans la **figure 11.1**, les deux graphes sont identiques et la notion de distribution communautaire est une information cachée et avec la SNA qu'on découvre.

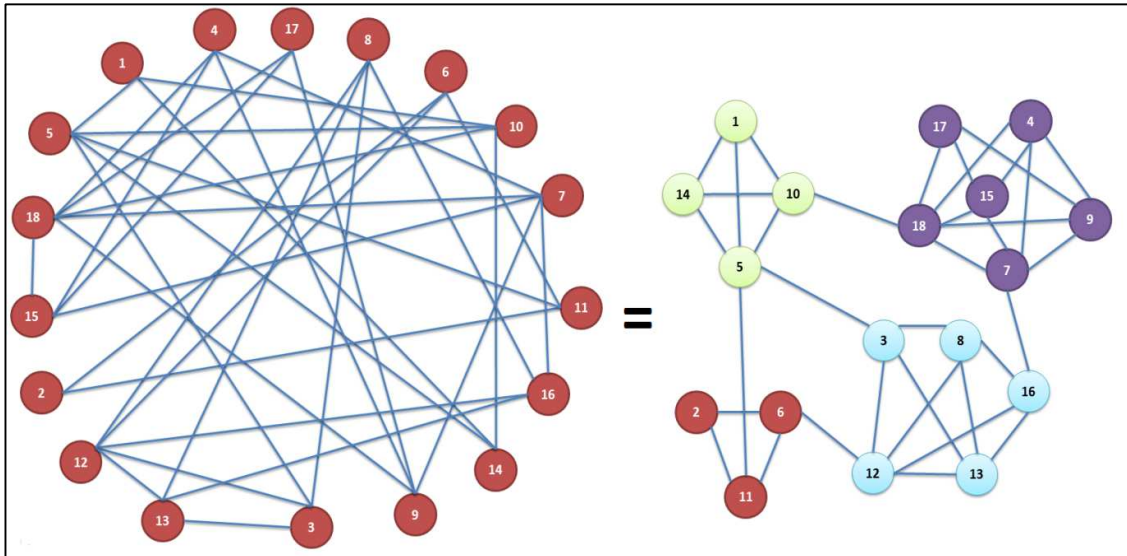


Figure II.1 : Distribution communautaire, information cachée.

## II.2. Détection des communautés, objectifs

Parmi les objectifs de cette notion de communautés, on peut citer ce qui suit:

- Mise en place d'une stratégie de marketing (*systèmes de recommandation*): à la base de la connaissance de l'ensemble des communautés, nous pouvons connaître les différents profils (*ou les intérêts communs regroupant les membres*), et ensuite nous pouvons faire une diffusion (*publicité, recommandation personnalisée, etc.*) d'informations précises à un ensemble bien connu d'utilisateurs.
- Identification des acteurs centraux d'un système: dans le domaine de la sécurité informatique, par exemple, il est important de savoir quels sont les nœuds (*leaders*) les plus importants qui doivent être attaqués (*ou mieux sécurisés si on est côté administrateur*) pour déstabiliser (*ou stabiliser*) un système ou un réseau.
- Solutions pour minimiser / maximiser la diffusion.
- Extraire les connaissances du réseau et mettre en évidence les principales propriétés du réseau.
- Fournir un résumé de la structure du réseau.
- Applicable à différents domaines tels que la biologie, la sociologie, le marketing...etc, pour permettre aux analystes d'extraire un aperçu utilisable et d'avoir un outil d'aide à la décision.
- Avoir de nombreuses propriétés sur le réseau comme l'importance d'un acteur donné (*son influence, sa popularité, etc.*) par rapport aux autres;
- Étudier la similarité entre les individus d'une même communauté et donc le degré d'interaction et mesurer ensuite la force de la relation entre eux
- Extraction des différents profils, centres d'intérêt, sujets d'actualité dont la population en parle;
- Connaître la tendance politique d'une population, connaître les goûts et les opinions des gens sur les produits proposés sur le marché;
- La connaissance d'un individu peut induire la connaissance des autres en relation avec lui ou avec le reste du groupe, le cas échéant.
- .. etc.

Plusieurs travaux de détection de communautés ont été proposés dans différents domaines (*figure II.2*) comme les protéines fonctionnelles [48], [49], les secteurs industriels [50], groupements des peuples [51], [52], etc.

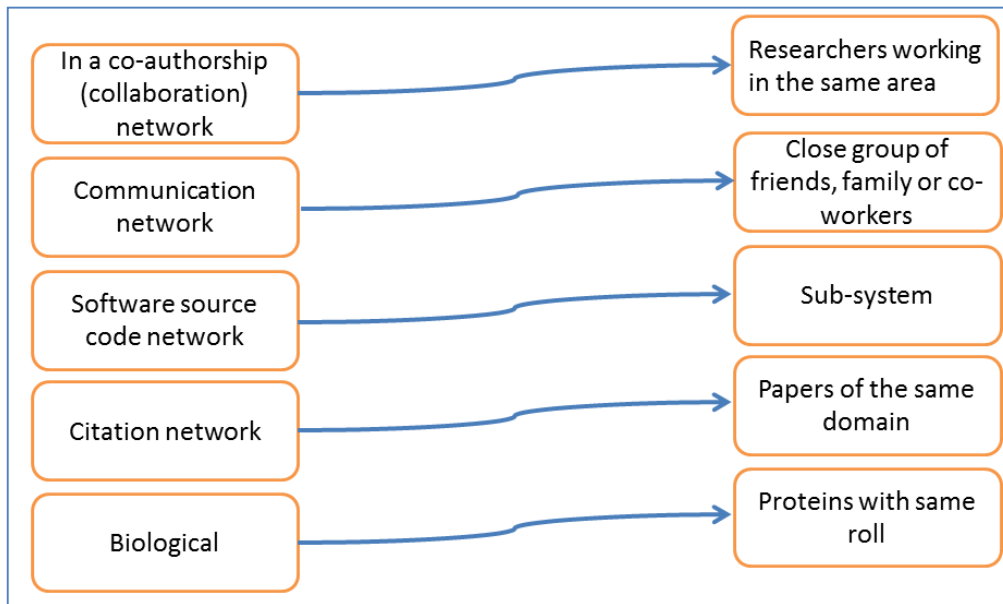


Figure II.2: Le sens de communauté dans différents domaines.

### II.3. Chevauchement des communautés

Plein de travaux de recherche se sont intéressés à la découverte des communautés, ces dernières peuvent se chevaucher (*des individus en commun*) ou non. Dans cette section nous donnons quelques algorithmes ayant traité ces deux cas.

#### A. Communautés chevauchantes

Selon les articles que nous avons étudié pendant la réalisation de notre projet, la plupart des recherches dans ce domaine de détection des communautés sur les réseaux sociaux ont abouti à des communautés non chevauchantes (*ou disjointes*). Cependant, dans la réalité, les abonnés de ces réseaux ont plusieurs centres d'intérêt et traitent plusieurs sujets à la fois, ce qui implique qu'ils peuvent appartenir, simultanément, à plusieurs communautés. Chacune d'elles peut être caractérisée par une relation très spécifique telle que l'amitié, la famille, le travail, etc.

Ainsi, quel que soit le domaine d'application, il sera parfois nécessaire de détecter toutes (*et donc plusieurs*) communautés d'un nœud spécifique. Comme illustré sur la figure II.3 où l'abonné "SARA" est un scientifique, sportif mais aime aussi les dessins animés.

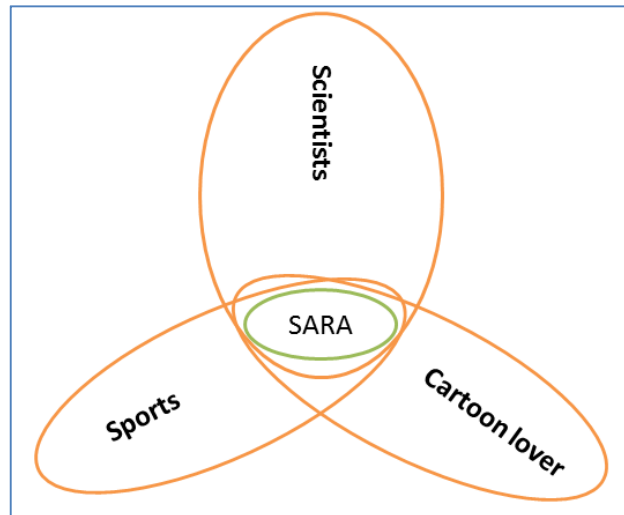


Figure II.3: Communautés chevauchantes

Certaines recherches se sont concentrées sur la détection de communautés se chevauchant dans les réseaux sociaux, comme [53] où les auteurs ont proposé une nouvelle approche pour découvrir les ensembles de communautés chevauchantes dans des systèmes complexes. De plus, dans [54], les auteurs ont proposé comment trouver les communautés qui se chevauchent dans les réseaux en appliquant un nouvel algorithme «propagation d'étiquette». Dans [55], les auteurs ont proposé une nouvelle approche pour trouver des communautés chevauchantes sur la base de partitions de nœuds et de liens associés et ils ont utilisé la méthode de factorisation matricielle non négative pour apprendre les paramètres de leur modèle. De plus, et récemment en 2018, les auteurs de [56] ont proposé un nouveau modèle pour détecter les communautés chevauchantes en se basant sur la stratégie d'expansion des grains s'appliquant aux réseaux non pondérés et pondérés, appelée OCSE. Et plus récemment, en 2019, les travaux de [57] étaient basés sur le modèle formel pour détecter les communautés qui se chevauchent dans les réseaux bipartites en affinant les groupes principaux à l'aide d'une analyse de silhouette. La même année, [43] a proposé un nouveau modèle pour couvrir les communautés qui se chevauchent dans les réseaux sociaux statiques et dynamiques.

### B. Communautés non chevauchantes

Selon les études de cas et le domaine d'application, nous devons parfois détecter des communautés qui ne se chevauchent pas, c'est-à-dire qu'un nœud appartient à une et une seule communauté. Comme si, par exemple, nous avons autant de rôles qu'ils peuvent exercer nos agents (*travailleurs, processus, protéines, etc.*) et que nous voulons connaître le rôle le plus approprié où un tel agent peut donner mieux et être plus rentable, efficace et ne lui pas surcharger avec d'autres tâches et se concentrer sur une seule (*avec les autres agents de sa communauté qui n'en font qu'une partie*), etc. Dans ce genre de cas, il faut "proposer" une distribution communautaire qui ne se chevauche pas afin de savoir puis affecter le rôle de chacun.

Dans ce contexte, nous pouvons citer le travail de [58] en 2016 qui a proposé une méthode unifiée pour détecter les communautés qui ne se chevauchent pas et une autre méthode pour celles qui se chevauchent. De plus, pour 2017, nous citons le travail de [59]

qui a proposé une approche intégrée pour la détection de communautés chevauchantes et celles non chevauchantes. En 2018, nous trouvons le travail de [60] qui a proposé une immunisation (ou protection) des réseaux avec une structure communautaire sans chevauchement et [61] qui a proposé un service Web pour l'évaluation de la qualité des communautés non chevauchantes.

## II.4. Types de communautés

Un SN peut contenir plusieurs et différentes compositions de communautés en fonction de leurs densités, d'une communauté unique contenant tous les nœuds aux plusieurs communautés contenant chacune un nœud.

Donc, tout dépend de la densité voulue, ces communautés sont principalement classées en deux grandes catégories [62].

### A. Communautés à sens forts

Chaque nœud a plus de connexions dans sa communauté qu'avec le reste du réseau (figure II.4). Mathématiquement, on dit que  $C$  est une communauté à sens forts si l'équation (eq.II.1) suivante est vérifiée:

$$K_i^{in}(C) > K_i^{out}(C), \forall i \in C \quad (eq.II.1)$$

Où  $K_i^{in}(C)$  signifié le degré (nombre de liens) du nœud  $i$  dans la communauté  $C$  (entre  $i$  et les différents membres de la communauté  $C$ ) et  $K_i^{out}(C)$  signifié le degré du  $i$  vers l'extérieur de la communauté  $C$ .

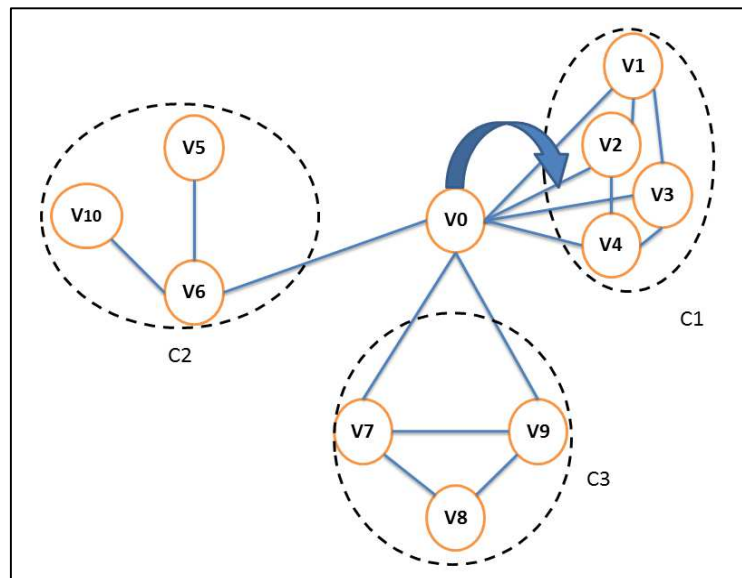


Figure II.4 : Communautés au sens fort.

Dans ce type de communautés, nous devrions avoir le degré interne de chaque nœud plus important (et même pas égal, et donc impossible d'avoir le cas de chevauchement) que l'externe. L'avantage de cette stratégie d'organisation en communautés est que la qualité de ces dernières est meilleure et que les relations entre leurs membres au sein de chacune sont



très fortes, l'inconvénient majeur est qu'elle n'est pas décidable pour un nœud ayant le même degré dans deux (ou plusieurs) communautés et il, donc, ne pourra pas décider laquelle choisir.

Ainsi, il est parfois impossible de diviser le graphe en communautés, comme par exemple, sur la figure II.4, si le nœud V0 crée une nouvelle relation avec V5, on aura  $K_{V0}^{in}(C1) = K_{V0}^{out}(C1) = 4$  et donc, V0 ne pourra pas joindre C1 (violation de l'équation eq.II.1).

**B. Communautés à sens faible**

Dans ce sens d'organisation, la somme de tous les degrés au sein d'une communauté doit être plus importante que la somme de tous les degrés vers le reste du réseau. Mathématiquement, on dit que C est une communauté à sens faible si l'équation 2 suivante est vérifiée:

$$\sum_{i \in C} K_i^{in}(C) > \sum_{i \in C} K_i^{out}(C) \quad (eq.II.2)$$

Dans la figure II.5, V0 pourra joindre C2 ( $\sum_{i \in C2} K_i^{in}(C2) = 5 > \sum_{i \in C2} K_i^{out}(C2) = 4$ ) et la condition est également vérifiée pour C1 (6, 5) et C3 (4, 3), comme il pourra joindre C1 ( $\sum_{i \in C1} K_i^{in}(C1) = 9 > \sum_{i \in C1} K_i^{out}(C1) = 5$ ) et la condition est également vérifiée pour C2 (3, 2) et C3 (4, 3).

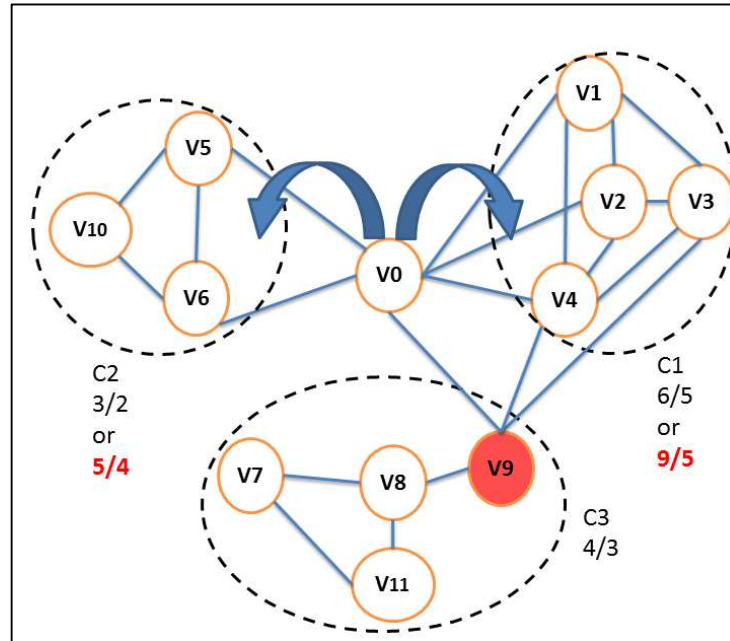


Figure II.5 : Communautés au sens faible.

L'avantage de cette organisation est qu'il est très facile de trouver une distribution communautaire qui vérifie la condition (eq.II.2) pour plusieurs cas de graphes, tandis que le principal inconvénient est l'indécidabilité dans la plupart des cas, la qualité des communautés trouvées est mauvaise, les relations entre les membres au sein de chaque communauté ne sont pas vraiment fortes et aussi, on peut trouver un nœud ayant plus de



relations à l'extérieur qu'à l'intérieur de sa communauté (comme  $V9$  vers l'extérieur de  $C3$ ). Mais, encore, envers une autre communauté que la sienne (comme  $V9$  qui a plus de relations avec  $C1$  que sa communauté  $C3$ ). Puis finalement, le cas de chevauchement est interdit.

Dans le cadre de cette thèse, et en plus du travail du [62] et pour construire un compromis entre ces deux sens des communautés, et afin de réduire leurs inconvénients et de combiner leurs avantages, nous définissons, dans le prochain chapitre, un nouveau sens de communautés que nous appellerons «communautés au sens moyen».

### II.5. Qualité d'une distribution communautaire

Le but de chaque algorithme de détection des communautés dans un réseau social est de le diviser en groupes (*chevauchés ou non*), chaque algorithme peut avoir autant de résultats et autant de distributions communautaires. À cet effet, chacun de ces algorithmes aura besoin d'une fonction objective qui précisera ses résultats et selon elle nous pourrions choisir (*et décider*) quelle distribution est la meilleure. Cette fonction objective témoigne du maximum possible de relations au sein des communautés et du minimum possible des relations intercommunautaires.

Plusieurs mesures et fonctions objectives ont été proposées pour quantifier la qualité des distributions communautaires des réseaux sociaux, comme l'approche spectrale augmentée proposée dans [63] et [64], dans ce travail, les auteurs ont développé un nouveau modèle basé sur l'optimisation itérative et ils l'ont utilisé pour étudier les communautés et l'équilibre structurel. Dans [65], les auteurs ont proposé un nouvel algorithme de détection communautaire qui vise à minimiser le critère de coupe standard et dispose d'une procédure de sélection de l'ordre du modèle, les performances de ce nouveau modèle sont comparables par rapport aux approches spectrales minimisant la coupe normalisée.

Dans ce domaine de qualification des résultats concernant les distributions communautaires, nous avons également les deux approches les plus utilisées dans la plupart de ce type de travaux: la «méthode de maximisation de la modularité» [66] et la «méthode d'inférence statistique par maximum de vraisemblance» [67], [68] et nous avons le travail de Newman [69] qui a examiné, étudié puis comparé les résultats de ces deux approches et leurs performances.

Dans la dernière année de 2019, nous trouvons d'autres travaux qui ont proposé de nouvelles approches pour qualifier les distributions communautaires comme [70] qui a combiné entre la maximisation de la modularité et certaines mesures de similarité (*la distance cosinus*) qui dépend de liens partagés. La même année, nous avons également le travail de [71] qui a utilisé la modularité avec une nouvelle approche qui supprime itérativement les bords avec des restrictions.

Dans le cadre de notre thèse, nous nous basons sur la notion de modularité ( $Q$ ) pour qualifier les distributions communautaires résultantes. Notre choix s'est porté sur cette notion car elle est parmi les plus utilisées dans la littérature et nous pouvons donc comparer nos résultats par rapport à d'autres algorithmes en fonction de la valeur de  $Q$  sur les mêmes benchmarks.

La recherche de la valeur de modularité optimale (*la plus grande*) est un problème NP-difficile [5]. Mathématiquement, plusieurs formules ont été proposées pour calculer la

valeur de  $Q$  de toute distribution communautaire et nous utiliserons celle donnée par [72] comme suit (équation II.3):

$$Q = 1/2m \sum_{ij} (AM_{ij} - P_{ij})\delta(C_i, C_j) \quad (eq. II.3)$$

Où  $m$  représente le nombre de liens dans le graphe,  $AM$  est la matrice d'adjacence du graphe,  $P_{ij} = K_i * K_j / 2m$  où  $K_i$  représente le degré de nœud  $i$ . La fonction  $\delta$  donne un si les sommets  $i$  et  $j$  sont dans la même communauté ( $C_i = C_j$ ), zéro sinon.

La modularité est comprise entre -1 et 1 mais la partition triviale regroupant tous les nœuds en une seule partie a pour modularité égale à 0 et donc les partitions intéressantes ont toutes une modularité positive et une modularité de 1 est inatteignable [73]. Des bornes plus précises sont données dans [74] où les auteurs ont prouvé, par exemple, que la limite supérieure de la valeur de la modularité pour n'importe quel réseau est +1.

Dans la plus part des cas, l'utilisation de cette fonction de modularité a un effet décisable sur la solution finale à prendre et donc un outil très important d'aide à la décision quant à la détection des communautés. En d'autre terme, si nous avons deux distributions communautaire  $DC1$  et  $DC2$ , pour les comparer et connaître la meilleur il suffit de calculer leurs valeurs correspondantes  $Q$ , on trouvera ainsi  $Q1$  et  $Q2$ , donc :

Si ( $Q1 > Q2$ ) alors *DC1 est meilleure que DC2 // retourne (DC1)*  
 sinon  
 Si ( $Q1 == Q2$ ) alors *DC1 est équivalente à DC2 // retourne (DC1 ou DC2)*  
 sinon *DC2 est meilleure que DC1 // retourne (DC2)*

## II.6. Stabilité des algorithmes de détection de communautés

Si nous ré-exécutons le même algorithme de détection de communautés (*CDA*) sur le même réseau ( $G$ ) et nous obtenons des résultats différents (communities distributions  $CD1$  et  $CD2$ ), on dits que cet algorithme est instable (figure II.6). Deux distributions communautaires sont égales si et seulement si pour tout nœud  $v$  du graphe ( $G$ ), sa communauté dans l'une est la même que celle dans l'autre.

Cette notion de stabilité a été introduite pour la première fois par Aynaud et Guillaume [73] pour l'optimisation du processus de détection de communautés dynamiques. Ils ont cherché à s'assurer que la partition trouvée au temps  $t + 1$  soit cohérente avec celle au temps  $t$ .

Par exemple, l'algorithme de Louvain [4] est instable car il fait des choix non déterministes en plusieurs points. Le premier choix est l'ordre d'évaluation des nœuds lors de l'itération. Louvain évalue les nœuds un par un et les place dans la communauté voisine en maximisant le gain de modularité, donc selon l'ordre d'évaluation, les choix et les fusions seront différents. Le deuxième choix a lieu lorsqu'il y a plusieurs communautés voisines permettant le même gain de modularité. Dans ce cas, un choix au début peut avoir beaucoup d'influence sur la suite. Aynaud [73] montre que bien que chacune de ces étapes soit déterminable, en ajoutant quelques règles pour que l'algorithme fasse toujours le même choix.

Ainsi, afin de rendre stable tout nouvel algorithme, nous devons d'abord le déterminer en suivant le même ordre sur les nœuds. De plus, nous parcourons les nœuds dans un ordre

croissant et déterministe et choisissons toujours la première communauté en cas de choix de jointure. De cette façon, les résultats obtenus par ce nouvel algorithme seront stables (*les mêmes*) à chaque ré-exécution sur le même réseau.

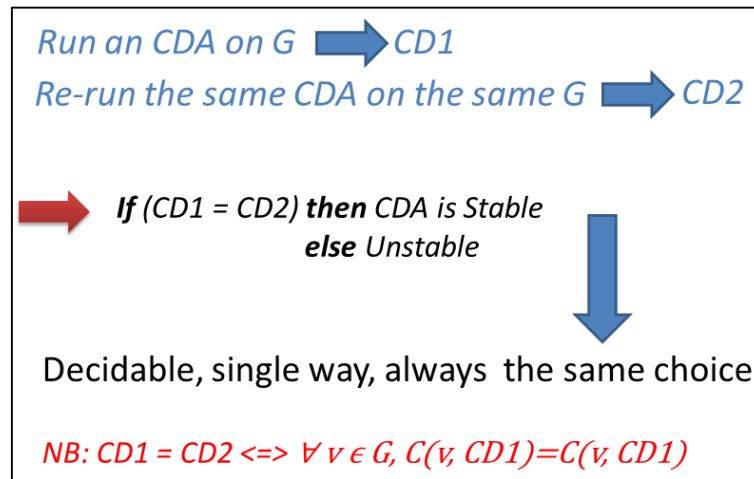


Figure II.6 : Stabilité des algorithmes de détection de communautés.

## II.7. Historique et travaux connexes

Plusieurs travaux de recherche se sont intéressés à l'état de l'art de cet axe de détection des communautés dans les réseaux sociaux, entre autre les nôtres [75] et [130]. Aussi, plusieurs algorithmes de détection communautaire ont été proposés dans plusieurs domaines (*réseaux de télécommunications, réseaux physiologiques et biologiques ainsi que les réseaux sociaux, etc.*). Dans cette section, nous donnons un aperçu CHRONOLOGIQUE sur l'évolution du centre d'intérêts des chercheurs, concernant cet axe de recherche, depuis 1970 jusqu'à ces jours-ci.

En 1970, nous remarquons l'apparition d'un des algorithmes les plus anciens est celui de Kernighan-Lin [77], qui utilise une méthode de division du graphe en deux sections pour trouver la coupe minimale des arêtes. En 2002, Girvan et Newman [78] ont présenté un algorithme basé sur le degré de centralité des bords et ont introduit pour la première fois la fonction de modularité pour détecter les communautés statiques non superposées. Cet algorithme ne traite pas les graphes pondérés et orientés. De plus, le temps d'exécution rend son utilisation impossible dans les très gros graphes. En 2004, Radicchi et al. [62] ont proposé un nouvel algorithme basé sur le degré de regroupement des bords, en supprimant ceux du coefficient minimal, à chaque itération. La principale différence entre cet algorithme et plusieurs autres est la condition d'arrêt de la procédure de détection communautaire. Cet algorithme est basé sur les propriétés des communautés détectées et non sur la valeur d'une fonction de qualité. Dans cet algorithme, l'existence de cycles courts pourrait être un problème. La même année, Newman et Girvan [5] ont proposé l'algorithme "Edge Betweenness" qui est basé sur une approche hiérarchique descendante. Malgré sa complexité, qui le rend pratiquement inutilisable pour des réseaux de plus de quelques milliers de nœuds, cet algorithme est utilisé comme la base de plusieurs études grâce à ses solides bases théoriques et est désormais implémenté dans de nombreuses bibliothèques d'analyse des réseaux complexes. Cependant, il n'aborde pas le concept dynamique des réseaux. Restant dans la même année, Fortunato et al. [79] ont proposé une amélioration de l'algorithme de Girvan et Newman, basant sur la centralisation de l'information. Toujours

dans la même année, Clauset et al. [80], ont proposé la méthode ‘‘Fast Greedy’’ qui est une extension de l’algorithme de Newman [81], et elle est basée sur l’optimisation gourmande de la modularité. En 2005, Pons et al. [82] ont utilisé une approche de regroupement hiérarchique pour détecter la similitude entre les nœuds. De plus, Palla et al. [83] ont proposé l’algorithme de ‘‘percolation de cliques’’ (CPM) qui détecte les communautés chevauchantes et statiques. L’avantage de cet algorithme est sa simplicité mais il souffre de certaines faiblesses. Dans les graphes très denses, il peut avoir une complexité et une consommation mémoire excessives ainsi que le réglage du paramètre  $k$  (*taille des cliques*) à priori est un inconvénient pour ce type de méthodes. En 2007, Palla et al. [84] ont adapté leur algorithme CPM pour prendre en charge l’aspect dynamique des réseaux sociaux. La même année, Raghavan et al. [85] ont proposé une méthode de détection des communautés statiques basée sur la propagation des étiquettes (LP). Ce type d’approches exploite l’appropriation de la densité des liens communautaires. Un signal émis par un nœud et retransmis par ses voisins est plus susceptible de rester dans la communauté source que de se propager à d’autres communautés en raison de la densité relative des communautés et de la faiblesse des inter-communautés. Cet algorithme permet la détection de communautés non chevauchantes. La même année, Tantipathananandh et al. [86] ont proposé une nouvelle approche permettant la détection de communautés non chevauchantes sur des réseaux temporels. Toujours la même année, Chevalyere et al. [87] ont proposé un algorithme dérivé de la théorie du choix social et de la notion de vote. En 2008, Falkowski et al. [88] ont proposé une nouvelle approche pour détecter les communautés qui se chevauchent sur les réseaux temporels. La même année, Rosvall et al. [89] ont proposé une nouvelle approche basée sur des relations aléatoires pour trouver la structure communautaire. De plus, Blondel et al. [4] ont proposé l’algorithme de Louvain, qui est une approche agglomérative (*ascendante*) et hiérarchique basée sur l’optimisation de la modularité. Initialement, ils considèrent chaque nœud comme une communauté; ils analysent aléatoirement les nœuds et les communautés voisines et calculent la nouvelle valeur de modularité en cas de jointure. En cas de gain en valeur de modularité, la jointure est exécutée, sinon elle est annulée. Cette opération est répétée jusqu’à ce qu’aucun gain ne soit possible. Cet algorithme n’est pas déterministe. En 2009, Kim et al. [90] ont traité la détection dynamique directe de communautés qui se chevauchent. Dans cette approche, le problème de la stabilité est étudié. Au lieu de modifier la fonction qualité, la stabilité est forcée en changeant les distances entre les objets à classer. Toujours en 2009, Xiang et al. [91] ont proposé une nouvelle métrique quantifiant la similarité entre les structures sous-graphiques. En 2010, une adaptation de l’algorithme Raghavan est proposée par Gregory et al. [92] pour la gestion des communautés qui se chevauchent (COPRA). Aussi, Aynaud et al. [73] ont traité la détection dynamique directe de communautés non chevauchantes où ils ont proposé une modification de la méthode de Louvain afin de réduire son problème de non-déterminisme. En 2011, Kanawati [93] a proposé l’algorithme LICOD (*Leaders Identification for Community Detection*) qui est une nouvelle approche de détection des communautés statiques, dans cette approche chaque communauté est composée de deux types de nœuds: les nœuds ‘‘Leader’’ et les nœuds ‘‘Follower’’. En 2013 et en 2016, nous avons remarqué l’apparition des algorithmes basés sur des méthodes génétiques, comme celui de Ronghua et al. [94], [95] qui optimise la fonction de modularité via un algorithme génétique. En 2015, Lakhlef [96] a proposé un algorithme pour un schéma de ‘‘clustering’’ à plusieurs niveaux basé sur des cliques et des ‘‘clusters’’ pour les réseaux de capteurs sans fil. En 2016, Saoud et al. [97] ont proposé une nouvelle méthode de fractionnement et de fusion pour la détection communautaire dans laquelle un arbre couvrant minimum (MST) de dissemblance entre les nœuds est utilisé. L’année suivante, les mêmes auteurs ont proposé dans [98] une nouvelle méthode pour les réseaux non pondérés et non orientés basée sur la similarité entre les nœuds en termes de

voisins partagés et l'optimisation de la modularité. De plus, De Salve et al. [99] ont proposé une évaluation des propriétés structurelles et temporelles d'Ego Networks pour la disponibilité des données dans les réseaux sociaux décentralisés (*Distributed Online Social Networks DOSN*), ils présentent une analyse de l'affinité temporelle et de la structure des communautés et de leur évolution dans le temps en utilisant une réelle base de données de Facebook. La même année, Fazlali et al. [100] ont proposé une adaptation parallèle de l'algorithme de Louvain sur une plateforme multicore. De plus, Atay et al. [101] ont effectué une analyse comparative des algorithmes méta-heuristiques. Aussi, Moosavi et al. [102], ont proposé une approche basée sur les intérêts, les schémas fréquents, les actions et enfin les activités des utilisateurs sur les réseaux. En 2018, d'autres travaux sont apparus comme celui de Huang et al. [103] qui ont proposé un nouveau modèle de détection des communautés qui se chevauchent dans les réseaux sociaux hétérogènes. De plus, Ahajjam et al. ont proposé dans [104] une nouvelle approche évolutive de détection de la "communauté leader". Toujours dans cette année, Al-Ayyoub et al. [105] ont proposé une amélioration de la détection de communautés en utilisant le parallélisme dynamique. Aussi, Kun He et al. [106] ont proposé une approche HICODE pour identifier les structures des communautés cachées ainsi que les structures des communautés dominantes. De même, Said et al. [107] ont proposé l'algorithme CC-GA pour détecter les communautés dans des réseaux complexes en utilisant des algorithmes génétiques et la notion de coefficient de "clustering".

### II.8. Principales approches de détections des communautés dans les SN

Un travail très intéressant a été effectué par NEDIOUI en 2015 dans [108] où il a classé les algorithmes de détection des communautés en deux grandes catégories, statiques où l'état du réseau, sujet d'étude, reste inchangé et dynamique où on pourra avoir, à tout moment, l'apparition de nouveaux nœuds ou nouvelles relations, ou la disparition de nœuds ou anciennes relations. Avec ces changements, on pourra avoir les six événements suivants : la naissance d'une nouvelle communauté, sa mort, la fusion de deux ou plusieurs communautés, la division d'une communauté en deux ou plusieurs autres, la croissance et la contraction des communautés.

Dans chacune de ces deux catégories, on trouve les deux segments: avec et sans chevauchement. Dans chacun, on trouve d'autres sous-approches. La figure II.7 illustre cette classification sous forme d'un arbre, avec quelques exemples d'implémentation et d'algorithmes au niveau des feuilles.

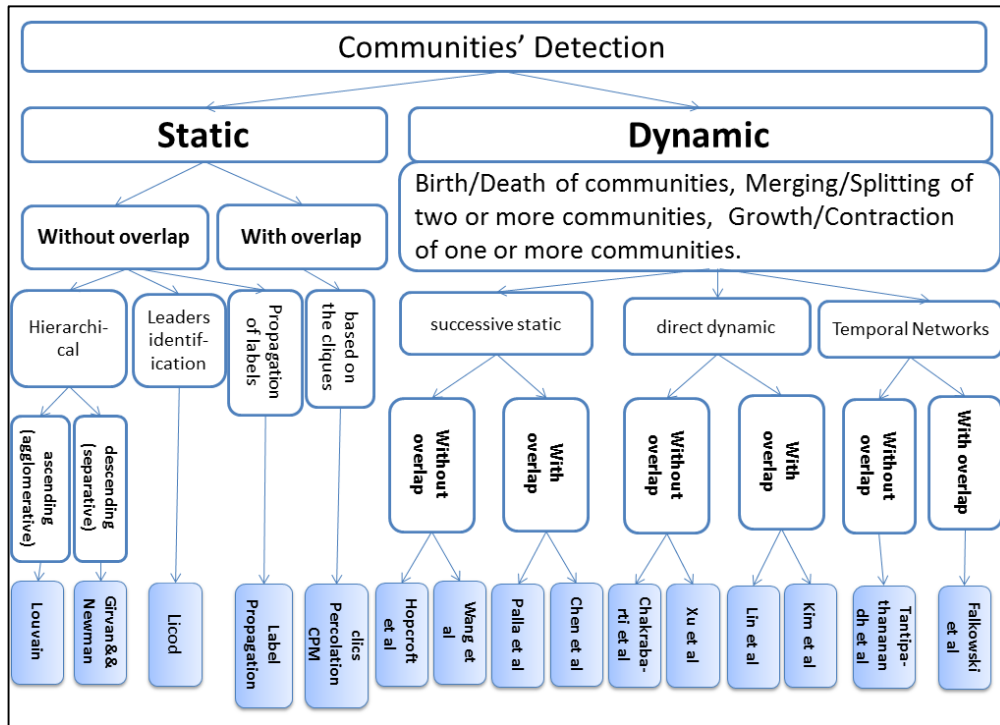


Figure II.7 : Approches et algorithmes de détection des communautés, Synthèse de [108].

Parmi les approches les plus basiques et que la majorité des travaux ultérieurs se sont basés sur leur concept on trouve celle hiérarchique où nous avons des approches descendantes (*séparatives*) et autres ascendantes (*agglomératives*), la figure II.8 illustre le principe de chacune de ces deux approches fondamentales.

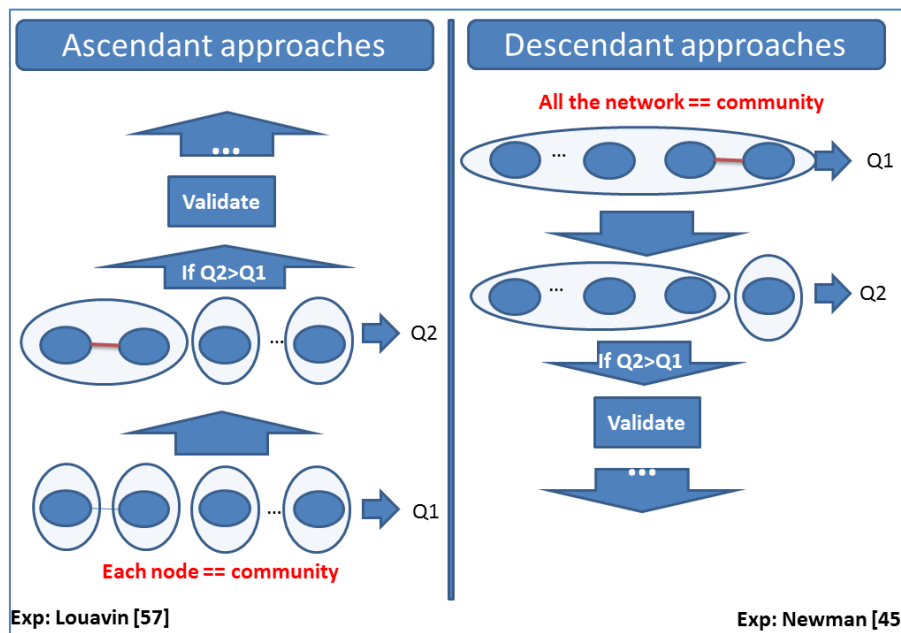


Figure II.8 : Approches hiérarchiques.



Toutes les approches hiérarchiques peuvent être représentées graphiquement comme étant un dendrogramme. La figure II.9 illustre un exemple de dendrogramme suivant l'application de ce type d'algorithmes de détection de communautés dans un réseau social.

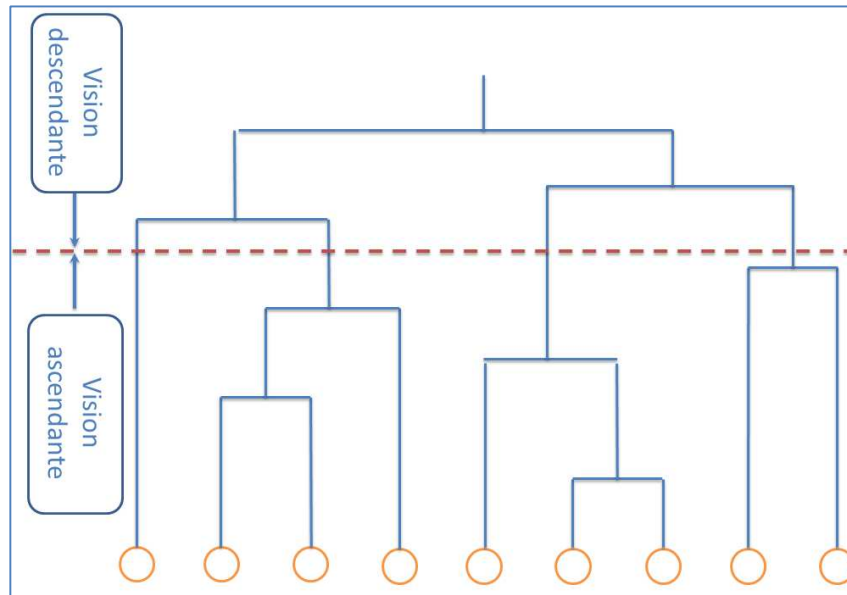


Figure II.9 : Dendrogramme d'une approche hiérarchique de détection de communautés.

La figure II.9 représente un dendrogramme sous forme d'un arbre hiérarchique. Les feuilles en bas de la figure représentent les sommets du réseau et en montant vers le haut, les sommets se rejoignent pour former des communautés de plus en plus grandes, comme indiqué par les lignes verticales, jusqu'à ce que nous atteignons la tête du dendrogramme, où tous les sommets sont réunis en une seule communauté. De même, en descendant du haut en bas, le graphe représente initialement un réseau connecté divisé en communautés de plus en plus petites au fur et à mesure. Une coupe transversale de l'arbre à n'importe quel niveau, comme indiqué par la ligne pointillée, donnera les communautés à ce niveau. La hauteur verticale des points de division dans l'arbre n'est indicative que de l'ordre dans lequel les divisions ou les fusions ont eu lieu.

Dans ce qui suit, et à titre illustratif, nous détaillons les deux approches hiérarchiques ascendantes et celle descendantes tout en donnant un algorithme, parmi les plus connus de la littérature, pour chacune. Louvain [4] proposé par Blondel et al. en 2008 et cité 10937 fois (*sur GoogleScholar le 09 Mars 2020*) pour les approches ascendantes ainsi que l'algorithme "Edge-Betweenness" proposé par Newman et Girvan en 2004 [5] qui est cité 12406 fois (*sur GoogleScholar le 09 Mars 2020*) pour celles descendantes. On a préféré de citer ces deux fameux exemples vu leur simplicité, facilité de mise en place et exploitabilité.

### A. Algorithme de Louvain

Cet algorithme a eu ce nom vu son originaire où il a été conçu au sein de l'université belge de Louvain. Louvain est un algorithme de détection de communautés hiérarchique ascendante, il démarre par l'hypothèse que tout nœud est une communauté, puis il regroupe chaque deux nœuds adjacents dans une même communauté en maximisant la modularité. Cet algorithme, contrairement à beaucoup d'autres, n'est pas supervisé c'est-à-dire que le nombre de communautés à former n'est pas demandé avant l'exécution.

Cet algorithme est très efficace en termes de temps d'exécution, il permet d'analyser des réseaux typiques de deux millions de nœuds en deux minutes [4]. Il comprend deux phases. Premièrement, il recherche les «petites» communautés en optimisant la modularité de manière locale (*séparément, au niveau de chaque communauté*). Deuxièmement, il regroupe les nœuds d'une même communauté et construit un nouveau réseau dont les nœuds sont les communautés. Ces deux phases sont répétées de manière itérative jusqu'à ce qu'un maximum (*local*) de modularité soit atteint. La figure II.10 que les auteurs ont utilisé dans leur papier illustre ces deux phases.

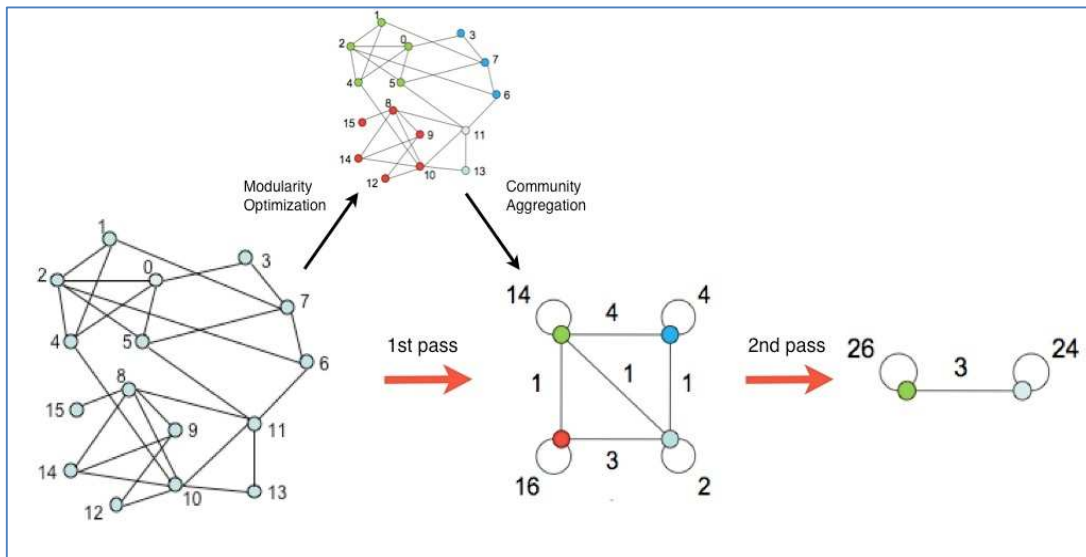


Figure II.10 : Les deux phases de l'algorithme de Louvain [4].

Le résultat (*Output*) de cet algorithme donne donc plusieurs partitions. La partition trouvée après la première phase se compose généralement de nombreuses communautés de petites tailles. Aux étapes suivantes, des communautés de plus en plus grandes se trouvent en raison du mécanisme d'agrégation. Ce processus conduit naturellement à une décomposition hiérarchique du réseau.

C'est évidemment une méthode approximative et rien ne garantit que le maximum global de modularité soit atteint, mais plusieurs tests ont confirmé que notre algorithme a une excellente précision et fournit souvent une décomposition en communautés qui a une modularité proche de l'optimalité.

Comme critique à cet algorithme, on pourra citer les points suivants :

- Il est conçu pour la détection de communautés non chevauchantes.
- Il est applicable uniquement sur des réseaux non orientés.
- Ce n'est pas un algorithme fournissant un maximum global, mais qu'il produit, en général, un optimum local.
- Il fait des choix non déterministes à plusieurs moments, pendant l'évaluation des nœuds durant une itération, quand plusieurs communautés permettant un même gain en modularité, et au moment du regroupement. Ce non déterminisme provoque une instabilité remarquable sur cet algorithme.
- En utilisant le dataset 'arXiv' qui est un réseau de scientifiques défini par Newman[109], les auteurs de l'algorithme de Louvain ont montré que le réseau initial



et après la suppression d'un seul nœud parmi 9377 du réseau, entre 1500 et 2500 changements de communautés ont eu lieu. Cela signifie que les communautés détectées entre chaque pas de temps sont très différentes (*plus d'un quart des nœuds se sont déplacés*) et que, par conséquent, cet algorithme est extrêmement instable. Ainsi, ils ont observé que cet algorithme est totalement inopérable pour détecter les communautés dynamiques sur plusieurs périodes en raison de cette instabilité.

- Etc.

Cet algorithme a été implémenté par les auteurs et le code source en C++ disponible sur <https://sites.google.com/site/findcommunities/newversion/community.tgz?attredirects=0>. Aussi, une mise à jour sur plusieurs aspects de cette version est disponible sur le lien <https://sites.google.com/site/findcommunities/newversion/community.tgz?attredirects=0>, ainsi qu'une implémentation avec le langage de programmation Matlab est disponible sur le lien [https://perso.uclouvain.be/vincent.blondel/research/Community\\_BGLL\\_Matlab.zip](https://perso.uclouvain.be/vincent.blondel/research/Community_BGLL_Matlab.zip), avec Python sur <https://bitbucket.org/taynaud/python-louvain/src/default/> et sur Github sur <https://github.com/taynaud/python-louvain/>. Ainsi, nous l'avons implémenté avec PHP et nous retournant vers lui dans le chapitre quatre de l'implémentation de notre algorithme, notre implémentation de Louvain nous a permis d'en comparer ses performance par rapport à notre modèle proposé.

### B. Algorithme de Newman et Girvan

Cet algorithme permet la découverte des communautés tout en éliminant d'une façon itérative les liens entre les sommets, qui vont devenir des liens "inter-communauté" par la suite. Pendant chaque itération (*suppression d'un line*), le processus calcul la nouvelle valeur de la fonction objective de modularité  $Q$  et valide la suppression si on aura un gain, la maximisation de cette valeur de modularité nous permettra d'en trouver la meilleure structure communautaire dans notre graphe. Cet algorithme est basé sur l'approche hiérarchique séparative (*ou descendante*), où nous commençons par le réseau en entier et essayons de trouver les paires de sommets connectés les moins similaires, puis supprimons les arêtes entre eux. La suppression itérative des liens entre les différents sommets implique un gain (*ou une perte*) dans la fonction de modularité, le lien donnant le maximum de gain après sa suppression est exactement celui reliant les deux sommets "les moins similaires" ou les plus faiblement connectés.

En faisant cela à plusieurs reprises, nous divisons le réseau en composants de plus en plus petits, et on pourra arrêter le processus à n'importe quelle étape et prendre les composants à ce stade pour qu'ils soient les communautés du réseau. Aussi, le processus peut être représenté comme un dendrogramme illustrant les divisions successives du réseau en groupes de plus en plus petits.

Plusieurs travaux ultérieurs ont appliqué avec succès l'algorithme de Newman et Girvan sur une variété de réseaux [80], y compris des réseaux de messages électroniques, des réseaux sociaux humains et animaux, des réseaux de collaborations entre scientifiques et musiciens, des réseaux métaboliques et des réseaux de gènes [43, 125, 126, 127, 128, 129, 88, 130, 131, 132].

Comme critique à cet algorithme, on pourra citer les points suivants :

- Cet algorithme, comme indiqué dans [80], impose de lourdes exigences sur les ressources de calcul, fonctionnant en complexité temporelle de  $O(n \times m^2)$  sur un

réseau aléatoire avec  $m$  arêtes et  $n$  sommets, c'est un algorithme qui doit être utilisé pour des réseaux d'au plus quelques milliers de sommets.

- Sa première version supporte uniquement les réseaux non pondérés et non orientés, puis des mises à jours se sont faites au fur et à mesure afin d'en supporter d'autres axes.
- Le nombre de communautés  $k$  est un input pour cet algorithme et doit être connu à l'avance.
- Pendant la suppression des liens, si l'algorithme trouve deux ou plusieurs ayant le même gain sur la modularité, le choix se fait d'une façon aléatoire, cet algorithme est donc instable.

Une implémentation de cet algorithme effectuée avec C++ est disponible sur le lien <https://github.com/SleepyBrain/GirvanNewmanAlgorithm> et pleins d'autres avec python sur <https://github.com/kjahan/community>, <https://github.com/ZwEin27/Community-Detection>, et <https://github.com/riteshkasat/Community-Detection-Algorithm>.

## II.9. Étude comparative entre quelques algorithmes de détection de communautés

Plusieurs études comparatives entre les différentes méthodes de détection de communautés ont été publiées telles que [64], [111], [72], [112]. De plus, plusieurs autres métriques ont été définies pour comparer les algorithmes (*structures des communautés détectées par chacun*), citant comme exemple: La densité interne [32], le rapport de coupe [72] et la coupe normalisée [113]. Pour voir plus de métriques (*mesures de similarité*), voir les deux travaux [114], [115].

Dans ce qui suit, nous donnons une étude comparative (*qualitative et quantitative*) entre quelques algorithmes de détection communautés les plus populaires sur certains benchmarks.

### A. Comparaison qualitative

Dans le tableau II.1 nous donnons une étude de comparaison qualitative entre les algorithmes les plus connus. Cette comparaison est basée sur les propriétés des communautés détectées (*statiques / dynamiques, avec ou sans chevauchement*).

Algorithm	With Overlap		Without Overlap	
	Static	Dynamic	Static	Dynamic
Louvain[4]			X	
Girvan et Newman[78]			X	
Licod[93]			X	
Percolation de cliques CPM[83]	X			
Label Propagation[85]			X	
COPRA[92]	X			
Hopcroft et al[116]				X
Palla et al[53]		X		
Kim et al[90]		X		
Tantipathananandh et al[86]				X
Falkowski et al[88]		X		

Tableau II.1: Comparaison qualitative entre quelques algorithmes de détection communautaires

**B. Comparaison quantitative**

Sur la base des simulations de leurs auteurs, dans les tableaux II.2 et II.3, nous donnons une étude de comparaisons quantitatives entre les algorithmes les plus connus. Le tableau II.2 (*N étant le nombre de nœuds du réseau*) donne une comparaison basée sur le nombre de communautés détectées sur les différents benchmarks les plus connus.

<b>Algorithms Benchmark</b>	<b>N</b>	<b>Louvain [4]</b>	<b>label propagation (LP) [85]</b>	<b>Licod [93]</b>	<b>Algo de Kuo et al. [117]</b>
Glass [118]	214	10	64	2	6
Iris [119]	150	11	40	5	3
Wine [120]	4898	12	59	2	3
Vehicle [121]	946	16	223	13	-

**Tableau II.2 : Nombre de communautés détectées.**

Le tableau II.3 donne une comparaison basée sur la valeur de la fonction qualité de la modularité.

<b>Algorithms Benchmark</b>	<b>N</b>	<b>Girvan &amp; Newman [78]</b>	<b>Newman04 [66]</b>	<b>Clauset et al. (CNM) [80]</b>	<b>Duch and Arenas (DA) [122]</b>	<b>Algo de KL et B Zalik [76]</b>
Zachary[123]	34	0.401	0.381	0.381	0.419	0.42
Jazz musicians[110]	1275	0.405	0.442	0.439	0.445	0.443
Metabolic[124]	453	0.403	0.435	0.402	0.434	0.443
Email ( <i>between students</i> ) [125]	1133	0.532	0.572	0.494	0.574	0.526
Key signing[126]	10 680	0.816	0.855	0.733	0.846	-
Physicists[109]	27 519	/	0.723	0.668	0.679	-

**Tableau II.3 : Modularité des distributions communautaires finales.**

**Conclusion**

La détection des communautés dans les réseaux sociaux est un sujet très posé dans le domaine de l'analyse des réseaux sociaux ces derniers temps. Plusieurs thèmes de recherches ont étudiés ce phénomène afin de faciliter l'analyse de ces réseaux.

Nous avons consacré ce chapitre à la définition de cette notion de communauté, ses types, la propriété de chevauchement, la qualité d'une distribution communautaire, la stabilité des algorithmes de détection des communautés, les différentes approches proposées dans la littérature ainsi qu'une étude comparative (*qualitative et quantitative*) entre quelques algorithmes les plus importants.

La plupart de ces algorithmes ne traitent pas en même temps les réseaux dynamiques, le chevauchement des communautés, les graphes pondérés et ceux orientés ainsi que la

propriété de stabilité. À cet effet, nous proposons, dans le prochain chapitre, une nouvelle approche générique et extensible qui pourra gérer tous ces aspects de manière flexible.

## Chapitre 3: Nouvelle approche de détection des communautés dans les RS

### Table des matières

Introduction.....	61
III.1. Communautés moyennes .....	61
III.2. Notre proposition .....	63
A. LFM, concept .....	63
B. Notations .....	64
C. Principe de l'algorithme LFM.....	65
D. Hypothèse.....	66
E. Qualité de la distribution communautaire générée.....	67
F. Stabilité de l'algorithme LFM .....	68
III.3. Extensions du LFM.....	68
A. Niveau d'extension 1: avec / sans chevauchement.....	69
a. Amélioration 1: meilleure modularité globale.....	71
b. Amélioration 2: paramètres de préférence.....	71
B. Niveau d'extension 2: non orienté / orienté.....	73
C. Niveau d'extension 3: dynamique / statique.....	74
III.4. LFM et les autres.....	75
Conclusion .....	76

Comme montré dans le chapitre précédent, dans la littérature on trouve beaucoup d'approches, modèles et d'algorithmes de détection des communautés dans les réseaux sociaux qui ont été proposés, la chose qui nous permet d'enrichir notre connaissance sur leurs aspects de base tout en étudiant leurs avantages et inconvénients. Cette variété nous a encouragés à proposer une nouvelle approche qui prendra en considération ces points faibles et forts mais aussi pour développer des algorithmes plus efficaces pour la détection communautaire et suggérer une discussion conceptuelle de la structure mathématique et sociologique des communautés dans les réseaux sociaux.

Dans ce chapitre nous détaillons notre nouvelle approche proposée tout en donnant ses principes, avantages par rapport à d'autres ainsi que les points faibles.

## **Introduction**

Ce troisième chapitre explique notre nouvelle approche de détection de communautés dans les réseaux sociaux. La méthode proposée est applicable sur des réseaux orientés et non orientés, pondérés et non pondérés, dynamiques et statiques et finalement pour détecter des communautés chevauchantes ou non. Il s'agit donc d'un modèle générique et généraliste qu'on pourra instancier sur plusieurs cas selon les domaines et cas d'étude ainsi que les champs d'application. Il s'agit d'une approche d'optimisation d'une fonction objective de qualité qui nous permet d'en détecter les meilleures structures communautaires du graphe objet de l'étude.

La détection de communautés dans notre approche se décompose en trois phases. La première vise à détecter les familles de base (*parents et enfants*) de notre société, la chose qui permettra de décomposer le réseau initial en petits groupes élémentaires dont les nœuds sont bien reliés entre eux. La deuxième phase consiste à joindre les autres membres n'ayant pas de parent (*Out-nodes*) de la première phase, plusieurs paramètres décident la jointure à telle ou telle communauté entre autres la maximisation de la modularité. Alors que la troisième phase nous permet d'en fusionner les communautés résultant de la deuxième phase sous certaines conditions.

Comme montré dans le deuxième chapitre, il existe deux types de communautés que tout algorithme détecte dans un réseau social, il s'agit des deux sens fort (*strong*) et faible (*weak*), nous et dans le cadre de ce projet avons proposé un troisième sens que nous appelons moyen (*middle*).

### **III.1. Communautés moyennes**

Dans le cas des communautés au sens fort, il faut vérifier, pour toutes les communautés, les degrés Out (*nombre de relations de la communauté vers son extérieur*) et In (*nombre de relations de la communauté vers elle-même*) de chaque nœud. La chose qui implique un temps d'exécution énorme. Par conséquent, proposer une nouvelle méthode pour réduire le temps d'exécution est important et nous avons besoin d'un autre type plus léger et plus rapide.

En ce qui concerne les communautés à sens faible, les relations entre les membres d'une même communauté sont parfois très faibles et ne reflètent pas la vérité d'une vraie communauté, et même, on peut trouver des individus avec plus de connexions à l'extérieur qu'à l'intérieur de leurs communautés, comme par exemple, le cas du nœud V9 dans la communauté C3 de l'exemple de la figure II.5 du chapitre précédent, la seule condition à vérifier est que le nombre de relations internes dans chaque communauté est plus important que celui externe. À cette fin, et aussi, nous avons besoin d'un autre type de communautés plus fort que le faible.

Dans cette section, nous présentons une nouvelle approche (*que nous appelons «communauté moyenne»*) comme intermédiaire entre les fortes et les faibles. Ainsi, le sous-graphe C est une communauté intermédiaire si l'équation (eq.III.1) est vérifiée:

$$K_i^{in}(C) > K_i^{in}(C'), \forall i \in C \text{ and } \forall C \neq C' \quad (\text{eq. III.1})$$

Pour chaque nœud, nous calculons ses degrés dans toutes les communautés actuelles, la communauté avec le degré maximal le prendra parmi ses nœuds. C'est exactement le même principe qu'on va suivre pendant la deuxième phase de notre approche où les nœuds vont joindre les différentes communautés adéquates.

Maintenant, pour un nœud donné, si nous avons deux ou plusieurs communautés ayant le même degré maximal, alors elles le partagent entre elles. Par conséquent, en appliquant l'équation (eq.III.1) les communautés seraient disjointes. À cette fin, et afin que les communautés partagent des individus, l'équation (eq.III.1) est adaptée comme suit (eq.III.2):

$$K_i^{in}(C) \geq K_i^{in}(C'), \forall i \in C \text{ and } \forall C \neq C' \quad (eq.III.2)$$

À cet effet et concernant le chevauchement des communautés, notre modèle est générique et c'est à l'utilisateur de le paramétrer selon son domaine d'utilisation et cas d'application.

Comme un exemple illustratif, dans la figure III.1 le degré du nœud V0 dans la communauté C2 est deux (2), en C3 il est égal à trois (3) et en C1 il est de quatre (4), donc V0 est pris par C1 malgré le degré général de V0 en dehors de C1 est de cinq (2 + 3).

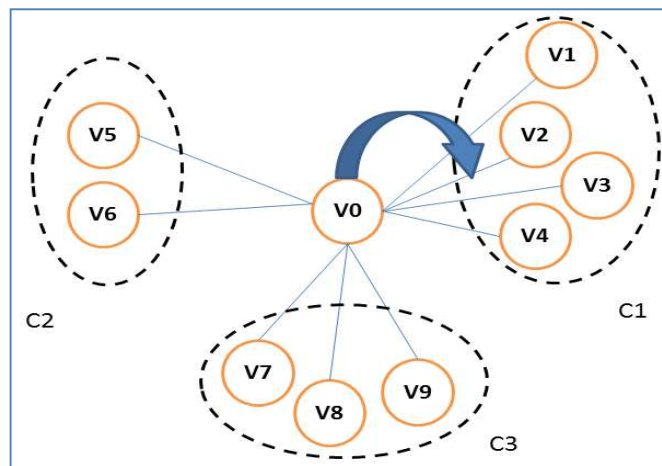


Figure III.1 : Notre nouvelle notion de communautés à sens moyen.

Dans cet exemple, si nous devons trouver un ensemble de communautés au sens fort, V0 ne pourra jamais rejoindre C1 car l'équation (eq.II.1) du chapitre précédant sera violée pour  $i=V0$  et  $C=C1$  (nous avons  $K_{V0}^{in}(C_1) = 4$  mais  $K_{V0}^{out}(C_1) = 5$ ).

La conception de cette idée pour cette nouvelle catégorie est venue d'avoir le maximum d'avantages des deux catégories (*forte et faible*) et le minimum d'inconvénients. Dans ce cas, le modèle sera décidable, un nœud n'a pas à choisir quelles communautés doivent rejoindre (*et donc il doit décider*), comme nous avons pris en charge le cas de chevauchement.

L'avantage de cette stratégie de regroupement est qu'il est facile de trouver une distribution communautaire qui vérifie la condition (eq.III.2) pour plusieurs cas de graphes (*avantage du «sens faible»*). De plus, nous ne pourrions jamais trouver un nœud avec plus de connexions dans d'autres communautés que la sienne (*avantage du sens fort*).

L'inconvénient majeur de ce sens des communautés est que la qualité des communautés trouvées n'est pas la meilleure mais aussi pas mauvaise. De plus, les relations entre les membres de chaque communauté sont moyennes. Une autre propriété importante est que nous pouvons trouver des nœuds avec plus de relations avec des nœuds externes (*avec plusieurs communautés unies mais jamais avec une seule*) qu'avec des nœuds internes à sa communauté (*soulagement du sens fort*).

### Extension aux graphes pondérés

Nous pouvons utiliser cette notion de «communauté moyenne» dans les graphes pondérés. Donc, pour un nœud quelconque, nous devons calculer la somme (pas le nombre) de ses poids dans toutes les communautés et celle(s) ayant la valeur maximale le prendra(ont).

La figure III.2 suivante illustre cette extension des communautés moyennes aux graphes pondérés comme suit :

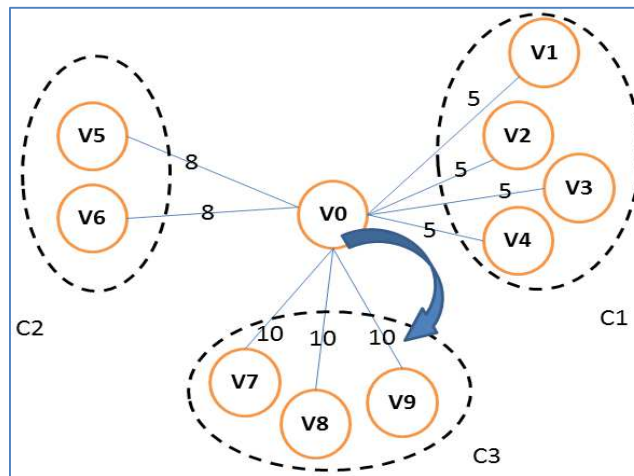


Figure III.2 : Extension de la notion de communauté moyenne aux graphes pondérés.

Dans la figure III.2, la somme des poids du nœud V0 dans la communauté C2 est de seize (16), dans C1 est de vingt (20) et dans C3 est de trente (30), donc V0 sera pris par C3 malgré que le poids global de V0 en dehors de C3 est de trente-six (16 + 20) et malgré qu'il a plus de relation (4) vers C1 que C3 (3).

### III.2. Notre proposition

Dans cette section, nous présentons notre nouvelle approche proposée, qui consiste à modéliser un concept sociologique et naturel de regroupement des individus dans leurs sociétés en familles, chaque famille est caractérisée par ses deux parents et l'ensemble des enfants, c'est l'ensemble des familles initiales qu'on doit découvrir en première étape, ces familles vont maîtriser la société et joindre, selon quelques paramètres de préférence, les autres individus, en parlant de la deuxième étape, puis ces nouvelles familles vont se fusionner, selon quelques conditions, on a donné un nom sociologique et significatif à notre modèle comme : le modèle des grandes familles (ou *Large Families Model: LFM*).

#### A. LFM, concept

Le réseau social ou son graphe associé est décomposé en un ensemble de composants connexes maximaux (*MRC pour Maximum Related Components*). Chacun d'eux a son ensemble de communautés. Une communauté dans LFM est initialement définie comme un ensemble de deux nœuds (*parents*) avec tous les nœuds communs (*enfants*) entre eux. Donc, dans notre approche, nous parcourons tous les nœuds deux à deux et calculons pour chaque pair (*parents*) le nombre de nœuds communs et nous formons ensuite les ensembles de familles. Chaque famille a un certain nombre d'enfants appelé "le degré de la famille" (*FD pour family degree*). Dans un premier temps, nous prenons les familles les plus larges c'est-à-



dire celles ayant le FD maximal dans la MRC actuelle, chacune de ces «grandes» familles forme une communauté initiale. La figure III.3 montre une composition initiale d'une communauté dans notre modèle LFM proposé.

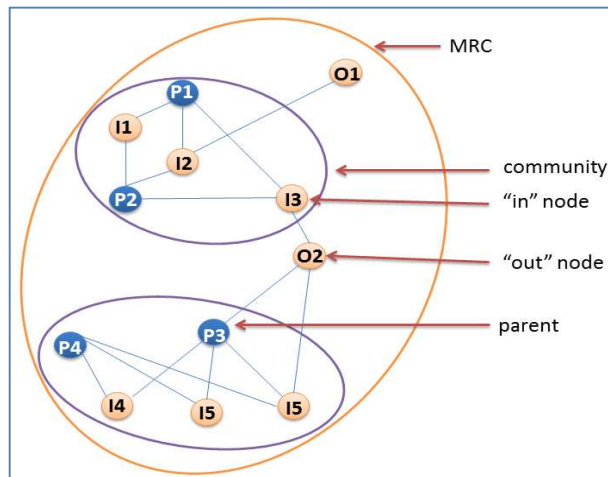


Figure III.3 : MRC, communautés, parents, "In\_node" et "Out\_node" dans LFM.

Comme illustré sur la figure III.3, le cercle extérieur représente le MRC. Les deux grands cercles représentent les deux communautés tandis que les petits cercles représentent les nœuds. Nous considérons trois types de nœuds comme le montre la figure II.3, les «parents» ( $P_i$ ) des communautés, les nœuds «in» ( $I_k$ ) et «Out» ( $O_j$ ) dans chaque MRC. Les nœuds «In» représentent les enfants et qui appartiennent à une (ou plusieurs) communautés, tandis que les nœuds «Out» sont ceux qui n'ont pas encore de communauté.

## B. Notations

Le tableau III.1 donne les notations utilisées dans notre approche.

Notation	Description
$V$	Ensemble de nœuds
$E$	Ensemble d'arêtes entre les nœuds
$G=(V, E)$	Le graphe associé du réseau social
$AM$	Matrice d'adjacence de $G$
$AM[i,j]$	Valeur booléenne représente la relation entre les nœuds $i$ et $j$ .
$MRC$	Ensemble de composants connexes maximaux.
$MRC(l)$	Le $l^{ième}$ MRC
$V(MRC(l))$	L'ensemble de nœuds du $l^{ième}$ MRC
$Adj(i)$	Ensemble de nœuds adjacents du nœud $i$ : $\forall i \in V, Adj(i) = \{j \in V, (i,j) \in E\}$ ou $\forall i \in V, Adj(i) = \{j \in V, AM[i,j]=1\}$
$Cadj(i,j)$	Nœuds adjacents communs entre les nœuds $i$ et $j$ : $\forall i, j \in V, Cadj(i,j) = \{h \in V, (i,h) \in E \text{ et } (j,h) \in E\}$ ou $\forall i, j \in V, Cadj(i,j) = Adj(i) \cap Adj(j)$

<b>MCadj[i][j]</b>	Nombre de voisins en commun entre les deux nœuds i et j : $\forall i, j \in V, MCadj[i][j] =  Cadj(i, j) $
<b>CadjMax(l)</b>	Le nombre maximal de nœuds adjacents communs entre tous les nœuds de MRC(l) : $CadjMax(l) = \text{MAX} ( Cadj(i, j) ), \forall i, j \in V(MRC(l))$
<b>Parents (l)</b>	Ensemble de parents (paire de nœuds avec le maximum de nœuds adjacents communs) de MRC(l), $\mathbb{N}$ : l'ensemble des entiers naturels, $\forall l \in \mathbb{N}, Parents(l) = \{k = (i, j) \in V(MRC(l)) * V(MRC(l)),  Cadj(i, j)  = CadjMax(l)\}$
<b>C(l)</b>	Ensemble de communautés initiales de MRC(l)
<b>Com(k)</b>	La k <sup>ième</sup> (ou (i, j) <sup>ième</sup> ) communauté initiale, k est unique pour tout le graphe: $Com(k) = \{i\} \cup \{j\} \cup Cadj(i, j)$ , tel que $k = (i, j) \in V^2$
<b>Out(l)</b>	L'ensemble de nœuds "out" du MRC <sub>l</sub> : $\forall l \in \mathbb{N}, out(l) = \{o \in V(MRC(l)), \nexists c \in C(l), o \in c\}$
<b>In(l)</b>	L'ensemble de nœuds "in" du MRC(l): $\forall l \in \mathbb{N}, in(l) = \{i \in V(MRC(l)), \exists c \in C(l), i \in c\}$
<b>Absb(o, k)</b>	L'ensemble d'absorbants (nœuds "in") du Com(k) ayant relation avec le nœud "out" o, $\forall l \in \mathbb{N}, \forall o \in out(l), \forall k \in Parents(l), Absb(o, k) = \{i \in Com(k), (i, o) \in E\}$
<b>D(o, k)</b>	Le degré (nombre de relations) du nœud o dans la Com(k) $\forall l \in \mathbb{N}, \forall o \in out(l), \forall k \in Parents(l), D(o, k) =  Absb(o, k) $
<b>Dmax(o)</b>	Le degré maximal du nœud "out" o dans toutes les communautés $\forall l \in \mathbb{N}, \forall o \in out(l), \forall k \in Parents(l), Dmax(o) = \text{Max}(D(o, k))$

**Tableau III.1 : Notations utilisées dans l'approche LFM.**

### C. Principe de l'algorithme LFM

Notre algorithme est basé sur la définition donnée dans la section (A). Il faut donc garder à l'esprit que le réseau social peut avoir un ou plusieurs composants connexes, chacun d'eux peut avoir ses communautés de tailles différentes d'un composant à un autre.

La forme générale de notre algorithme est divisée en trois phases comme suit:

**Étape 1 :** Rechercher et générer l'ensemble des parents et la valeur de CadjMax pour chaque MRC et donc l'ensemble des nœuds «Out».

**Étape 2 :** Pour chaque nœud «Out», trouver les bonnes communautés à rejoindre.

**Étape 3 :** Vérifier les communautés finales deux à deux pour les fusionner si possible (*sous certaines conditions*).

La figure III.4 montre l'organigramme de notre algorithme LFM proposé.

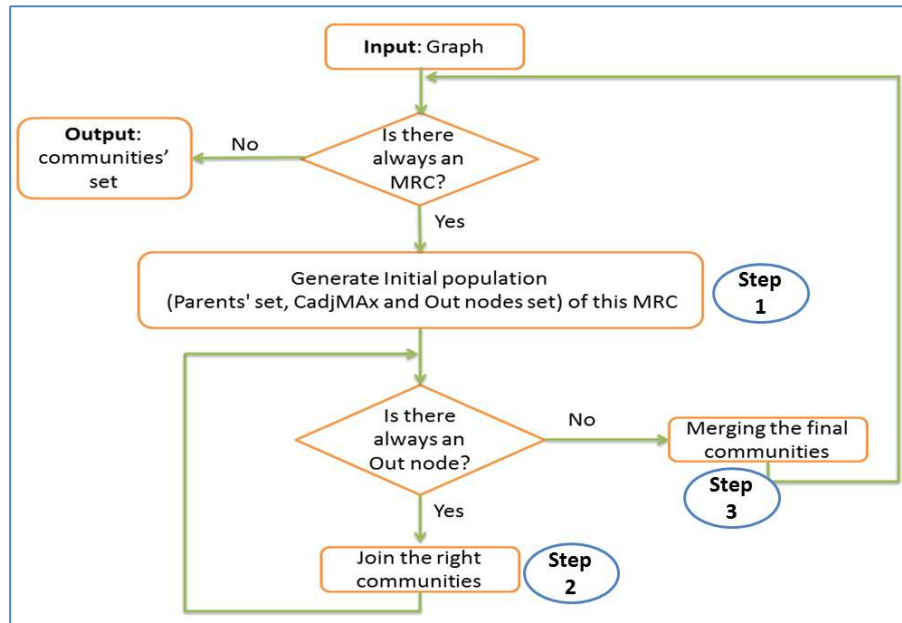


Figure III.4 : Organigramme de l'algorithme LFM proposé.

La spécificité de notre modèle proposé, par rapport aux autres, est qu'il est proche de la réalité et de la vie sociale des individus au sein de la population. Il commence par diviser l'ensemble de la population en sous-populations disjointes (*MRC*) qui n'interagissent pas entre elles. Puis, pour chaque sous-population, il cherche les relations (*en nombre*) entre les individus deux à deux et donc comme s'il cherchait les parents de chaque famille de la population. Ensuite, il prendra les familles les plus denses (*ayant le plus grand nombre d'enfants*), ces grandes familles qui maîtriseront et manipuleront la population, uniront autour d'elles les autres individus (nœuds Out) de la population. Chaque individu rejoindra une (*ou plusieurs*) de ces familles selon plusieurs critères (*paramètres de préférence*) expliqués dans la sous-section B de III.3.1 en bas. Par conséquent, la population est divisée en communautés. Une phase intéressante est que nous devons fusionner toutes les communautés ayant plus d'individus en commun que la différence.

Afin d'avoir des résultats plus pertinents et de respecter le sens sociologique d'une communauté, lors de la phase de détermination familiale (*phase 1*), on ne prendra que les couples de nœuds (*parents*) ayant des relations entre eux. De cette façon, nous allons réduire le nombre de communautés dans chaque MRC. Cependant, ces communautés seront plus justes et plus précises.

#### D. Hypothèse

Afin d'avoir une structure communautaire pour notre graphe, nous ne prenons en considération que les couples ayant plus que deux (2) enfants en commun comme parents, juste pour éviter le cas simple suivant (*figure III.5*):

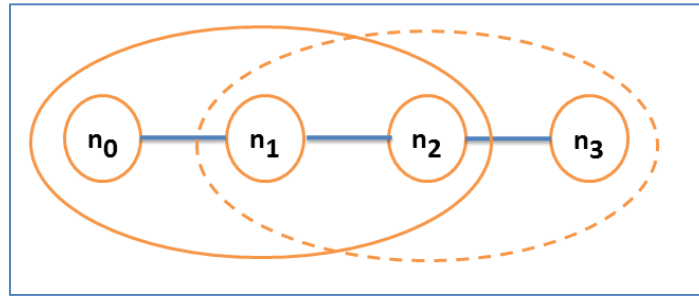


Figure III.5: Deux parents doivent avoir au moins deux enfants pour former une communauté.

### E. Qualité de la distribution communautaire générée

Comme expliqué dans la section II.4 du chapitre précédent, une bonne distribution communautaire signifie que le nombre de liens entre les communautés deux à deux est nettement inférieur à celui à l'intérieur de chacune. Notre objectif est de diviser le réseau en communautés où les membres communiquent entre eux plus qu'avec le monde extérieur et également d'arrêter la décomposition une fois qu'une bonne distribution est obtenue. Aussi, une distribution est dite "la meilleure" si sa fonction de qualité de modularité est maximisée.

Initialement, nous avons pris le maximum de nœuds communs entre chaque deux nœuds. Les paires ayant le nombre maximal de nœuds communs forment les communautés initiales. Maintenant et après avoir jointre les nœuds "Out" et fusionner les communautés finales, nous devons vérifier si cette distribution communautaire est la meilleure. Comme expliqué dans la section II.4 du chapitre précédent et pour mesurer la qualité de toute distribution communautaire, nous utilisons la modularité de Newman et Girvan.

En utilisant cette notion de modularité  $Q$ , nous pouvons améliorer notre algorithme avec l'extension suivante, nous calculons la valeur de  $Q$  correspondante au  $\text{CadjMax}(l)$  du  $l^{\text{ème}}$  MRC, nous obtenons la distribution communautaire initiale et sa valeur associée de  $Q$ . Nous répétons cette itération « $\text{CadjMax}(l)$ » fois, et pour chacun nous décrivons les valeurs de  $\text{CadjMax}(l)$  et  $Q$  puis nous prenons comme "meilleure distribution communautaire" celle équivalente à la plus grande valeur de  $Q$  comme illustré dans la figure III.6 et la portion du pseudo code suivantes.

**int GoodMaxVal (L)** // La meilleure distribution est différente d'une MRC à l'autre.

**Begin**

GMV= $\text{CadjMax}(L)$ ; // La valeur de  $\text{CadjMax}(L)$  est différente d'un MRC à l'autre.

Générer la première distribution communautaire pour GMV et sa modularité  $Q$

$Q_{\text{good}}=Q$ ;

$\alpha = \text{CadjMax}(L)-1$ ;

**For each**  $\alpha > 0$  **do**

Exécuter LFM\_Algorithm pour la nouvelle valeur de  $\alpha$ ;

Calculer la nouvelle valeur de la modularité  $Q_2$ ;

**if** ( $Q_2 > Q_{\text{good}}$ ) **then** GMV $\leftarrow \alpha$ ;  $Q_{\text{good}}\leftarrow Q_2$ ; **endif**

$\alpha - -$ ;

**endFor**

Return GMV;

**End**

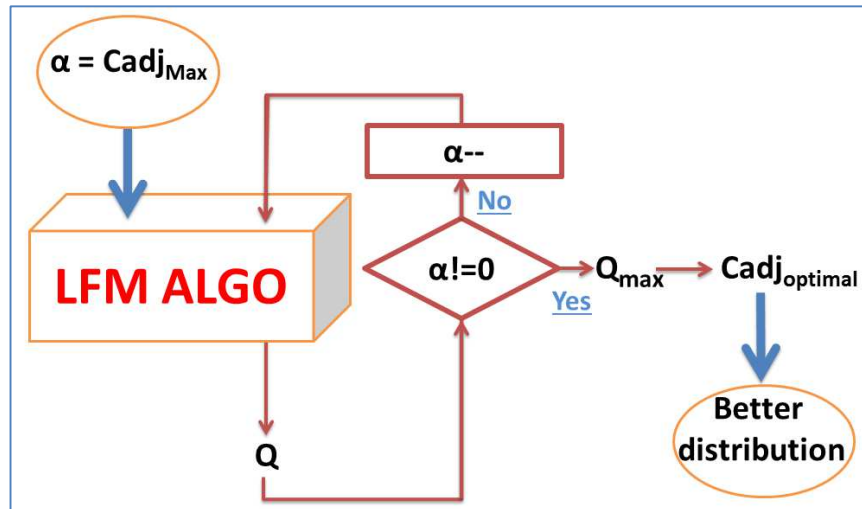


Figure III.6: Algorithme de qualité.

On ne s'appuiera pas uniquement sur cette notion de modularité pour assurer une bonne qualité, mais aussi sur le fait d'avoir, pour chaque nœud, le maximum de liens dans sa communauté que les autres (*communautés à sens moyen*).

Pendant la phase 3, et pour fusionner certaines communautés finales, il suffit de vérifier pour chaque deux d'entre elles si elles s'accordent (*nombre de nœuds en commun*) plus qu'elles se diffèrent (*nombre de nœuds en différence*). En d'autres termes, nous disons:

**Si**  $(|Com1 \cap Com2| \geq |Com1/Com2| + |Com2/Com1|)$  **alors**  
**Com1** ← **Com1 U Com2** and **Com2** ← **NULL** ;

### F. Stabilité de l'algorithme LFM

Afin de rendre notre algorithme stable, nous le déterminons d'abord en suivant le même ordre sur les nœuds. Nous parcourons les nœuds dans un ordre croissant et déterministe et choisissons toujours la première communauté en cas de choix de jointure. De cette manière, les résultats obtenus par notre algorithme sont stables à chaque réexécution sur le même réseau. Ainsi, l'algorithme LFM doit commencer par joindre les nœuds aux communautés ayant le maximum de relations avec eux. Dans ce cas, nous avons deux possibilités: si un nœud a plusieurs communautés avec le même degré maximum, elles le partageront entre elles. La deuxième possibilité est lorsque plusieurs nœuds ont le même degré (*maximum*) avec plusieurs communautés, il faut commencer donc par le nœud du plus petit indice (*identifiant*).

### III.3. Extensions du LFM

Notre modèle proposé est l'idée de base de nombreuses extensions, nous pouvons l'améliorer tout en rajoutant d'autres conditions comme la notion des communautés avec/sans chevauchement, la prise en charge de l'aspect statique/dynamique des réseaux sociaux et enfin les graphes orientés/non orientés. Pour rendre notre modèle plus efficace et plus général, nous proposons trois niveaux d'extensions, comme le montre la figure III.7.

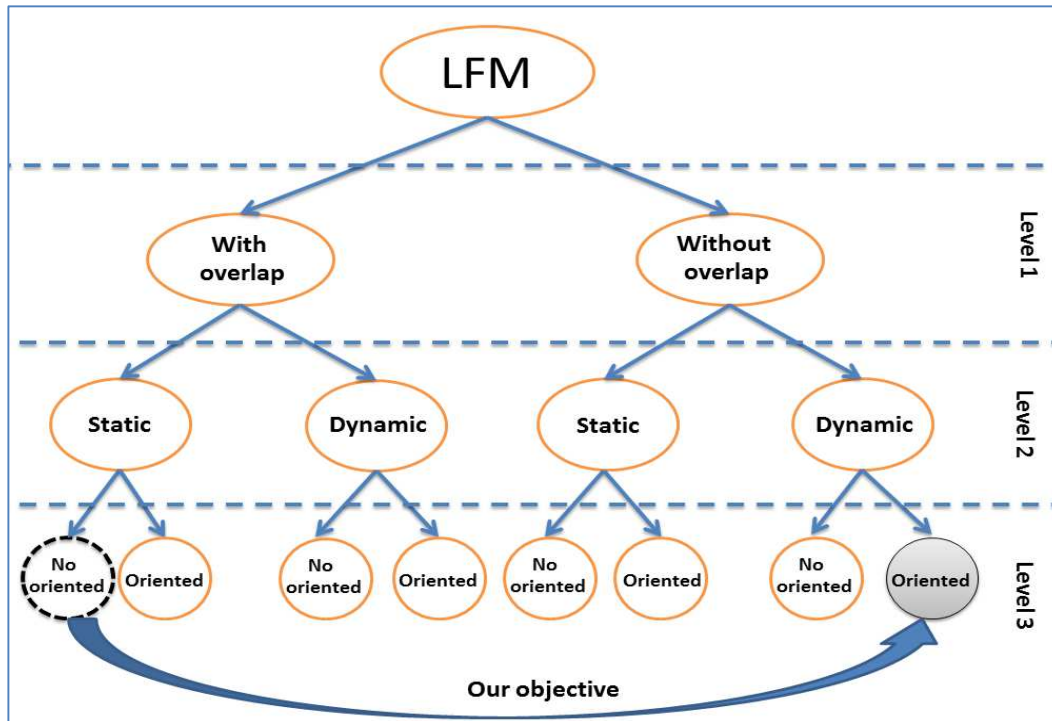


Figure III.7: Les trois niveaux d'extension du LFM.

Dans la figure III.7, le cercle discontinu représente le modèle de base tandis que le disque est notre objectif cible. Dans ce qui suit, nous détaillons la façon de prise en charge de ces trois niveaux d'extension afin de rendre notre modèle générique et extensible.

#### A. Niveau d'extension 1: avec / sans chevauchement

Y a-t-il une intersection entre les communautés? De plus, l'union des nœuds dans ces communautés est-elle égale à l'ensemble des nœuds du réseau? Et bien d'autres questions qu'on doit se poser lors de l'analyse de tout réseau social quand on veut détecter l'ensemble de ses communautés.

Dans de nombreux cas et selon le domaine d'application, nous sommes obligés d'éviter l'appartenance d'un individu à plusieurs groupes. Comme dans le cas du club de karaté de Zachary [123] qui montre un modèle d'amitié entre les membres d'un club de karaté à une université américaine dans les années 1970 où l'administrateur du club et son instructeur l'ont divisé en deux après un malentendu, dans ce cas, les membres ont été obligés de choisir un des deux (*sous*) clubs et ne pouvaient en aucun cas appartenir aux deux. L'application de notre algorithme sur cet exemple de données sera donnée dans le prochain chapitre.

Donc, il faut trouver un moyen d'adaptation de notre approche proposée afin de gérer ce cas de non-chevauchement, et voir, pour cet exemple de Zachary, quels individus rejoignent quel (*sous*)club.

Aussi, cette extension de prise en compte le non chevauchement élimine l'étape 3 de notre algorithme. Pour la première et la deuxième étape, cette extension peut être divisée en quatre cas:

- 1- **Parent / parent**: lors de la première étape, aucun parent d'une communauté ne doit être parent d'une autre.
- 2- **Parent / enfant**: lors de la première étape, aucun parent dans une communauté ne doit être enfant d'une autre.
- 3- **Enfant / parent**: lors de la première étape, aucun enfant d'une communauté ne doit être parent d'une autre.
- 4- **Enfant / enfant**: lors de la première et la deuxième étape, aucun enfant d'une communauté ne doit être enfant d'une autre.

La question qui doit être posée est de savoir comment un individu (*parent ou enfant*) doit choisir entre deux ou plusieurs communautés de manière **décidable et unique**, et si nous réexécutions le programme, ce choix est-il toujours le même? De cette façon, nous assurons la stabilité, mais nous devons également maximiser la qualité de la composition finale de la distribution communautaire.

Dans le cadre de notre thèse et à titre d'exemple, nous détaillons le seul cas traité dans la deuxième étape de notre algorithme, c'est le quatrième cas: enfant / enfant, comme suit:

Si nous avons un nœud "Out"  $o$  et deux communautés  $Com1$  et  $Com2$  où  $D(o,1)=D(o,2)=D_{max}(o)$ , alors  $Com1 \leftarrow Com1 \cup \{o\}$  et  $Com2 \leftarrow Com2 \cup \{o\}$ . Par conséquent, nous avons  $Com1 \cap Com2 \neq \{\}$ , donc  $Com1$  et  $Com2$  partagent  $o$ , nous sommes donc dans le cas de chevauchement.

Dans ce qui suit, nous observons (*le comportement de*) ces nœuds "Out" qu'ils souhaitent rejoindre l'une des communautés ayant la même valeur  $D_{max}$ , que nous appelons force de gravité car les individus (*personnes*) veulent joindre (*migrer vers*) une nouvelle communauté (*population*) et ils doivent choisir entre plusieurs et ces dernières vont exercer cette force de gravité pour le joindre parmi leurs membres.

Cela dépend du domaine d'étude et du but de l'application de notre algorithme, si on veut éviter ce cas de chevauchement la stratégie de mise à jour de l'algorithme LFM proposée est par défaut FIFO. Donc, pour un nœud "out"  $o$ , et parmi plusieurs communautés ayant la même force de gravité, la première de la liste (*générée par l'algorithme*) sera choisie comme le montre le tableau III.2 suivant :

Avec chevauchement	Sans chevauchement
<pre> ... <b>For each</b> <math>Com_k</math> in <math>C</math> <b>do</b>   <b>If</b> (<math>D(o,k) = D_{max}(o)</math>) <b>then</b>     <math>Com_k \leftarrow Com_k \cup \{o\}</math>;   <b>EndIf</b>; <b>EndFor</b>;  <math>Out(l) \leftarrow out(l) \cup \{o\}</math>; <math>In(l) \leftarrow In(l) \cup \{o\}</math>; //Changement de statut ... </pre>	<pre> ... <b>For each</b> <math>Com_k</math> in <math>C</math> <b>do</b>   <b>If</b> (<math>D(o,k) = D_{max}(o)</math>) <b>then</b>     <math>Com_k \leftarrow Com_k \cup \{o\}</math>;     <math>Out(l) \leftarrow out(l) \cup \{o\}</math>;     <math>In(l) \leftarrow In(l) \cup \{o\}</math>;   <b>EndIf</b>;   <b>If</b> (<math>o</math> <b>not in</b> <math>out(l)</math>) <b>then break</b>; <b>EndIf</b>; <b>EndFor</b>; ... </pre>

**Tableau III.2 : Niveau d'extension 1: avec / sans chevauchement.**



L'avantage de cette idée est qu'elle est simple et rapide à mettre en place. Cependant, l'inconvénient est que l'accès aléatoire d'un individu aux communautés peut induire une instabilité (*psychique, morale,...*) dans sa nouvelle vie. À cet effet, les améliorations suivantes sont proposées.

#### a. Amélioration 1: meilleure modularité globale

Pour toutes les communautés qui ont la même force de gravité, nous proposons que la migration soit effectuée et nous calculons la nouvelle valeur de modularité pour chacune de ces migrations. Nous prenons en compte la migration équivalente à la valeur maximale de la modularité comme suit:

$Q \leftarrow -1$ ; // La mauvaise valeur de la modularité

**For each**  $Com_k$  **in**  $C$  **do**

**If** ( $D(o,k) = Dmax(o)$ ) **then**

$Com_k \leftarrow Com_k \cup \{o\}$ ; // Intégrer  $o$  au  $com_k$

        Calculate NewQ; // Nouvelle modularité après fusionner  $o$  à  $com_k$

**If** (NewQ < Q) **then** // Pas de gain en modularité

$Com_k \leftarrow Com_k \setminus \{o\}$ ; // enlever  $o$  du  $com_k$

**else** Q  $\leftarrow$  NewQ; // MAJ de Q et maintenir  $o$  dans  $com_k$

**endIf**;

**endIf**;

**endFor**;

$out(l) \leftarrow out(l) \setminus \{o\}$ ; //  $in(l) \leftarrow in(l) \cup \{o\}$ ;

Ainsi, nous allons voir, dans le prochain chapitre et pendant la mise en place de cette amélioration, que la gestion de la qualité des distributions communautaires au cours de ce processus donne de meilleurs résultats par rapport à celui du FIFO et proches de ceux du cas de chevauchement.

Cette idée est plus efficace pour éviter l'accès aléatoire des individus aux communautés avec la même force de gravité. Cependant, l'inconvénient est que le choix de la meilleure distribution communautaire peut également induire une instabilité pour les individus.

#### b. Amélioration 2: paramètres de préférence

Lorsque nous avons plusieurs communautés qui génèrent la même valeur de modularité ou lorsque nous voulons améliorer la stabilité de l'individu migrant dans sa nouvelle vie, nous utilisons les paramètres de préférence suivants un par un et selon l'ordre ci-dessous:

- La densité de chaque communauté
- La «stabilité» élevée de chaque communauté
- Le pouvoir de mes voisins dans chaque communauté



- Le profil de chaque communauté.

Soit P le prédicat ou la fonction de préférence, définie comme suit:

```
If ( $D(o, k1) = D(o, k2) = Dmax(o)$ ) then  
  If ( $P(o, Com_{k1}) > P(o, Com_{k2})$ ) then  
     $Com_{k1} \leftarrow Com_{k1} \cup \{o\};$   
    else  $Com_{k2} \leftarrow Com_{k2} \cup \{o\};$   
  endIf;  
endIf;
```

Dans ce qui suit, nous décrivons le prédicat P pour chacun des paramètres de préférence mentionnés ci-dessus.

1. La densité de chaque communauté

Généralement, la nature humaine permet aux individus «migrants» de choisir les zones géographiques les plus populaires pour s'installer et se sentir plus à l'aise et plus heureux. Donc, si nous choisissons ce paramètre dans notre politique de migration, la fonction de préférence P pour un individu o et une communauté  $Com_k$  renvoie la taille (*nombre d'individus*) de la communauté  $Com_k$  comme suit:

$$P(o, Com_k) = |Com_k|.$$

Comme avantage de cette solution, nous citons l'évitement de l'accès aléatoire aux communautés. Comme inconvénients, nous pouvons dire que c'est plus difficile et plus lent, ainsi que les zones moins fréquentées et moins denses disparaîtront également.

2. La «stabilité» élevée de chaque communauté

Lorsque nous choisissons ce paramètre dans notre politique 'migratoire', les individus choisissent les zones les plus stables dans lesquelles ils s'installent. Une communauté  $Com_K$  est plus stable qu'une autre communauté  $Com_L$  si  $P(Com_K) > P(Com_L)$ , où

$$P(Com_i) = \text{Din}(Com_i) / \text{Dout}(Com_i), \text{ et}$$

$\text{Din}(Com_i)$ : le nombre de relations au sein de  $Com_i$

$\text{Dout}(Com_i)$ : le nombre de relations de  $Com_i$  avec l'extérieur.

Comme avantage de cette solution, nous pouvons sauver et donner des chances à des zones moins populaires d'exister et d'évoluer. Cependant, on peut dire que c'est plus difficile et plus lent, et après un certain temps, l'individu migrant peut se retrouver seul car les interactions entre les autres sont très importantes par rapport à lui.

3. Le pouvoir de mes voisins dans chaque communauté

Lorsque nous choisissons ce paramètre dans notre politique de migration, l'individu cherchera parmi ses voisins celui ayant le plus de relations dans sa communauté. Une fois trouvé, ledit individu choisira la communauté de ce voisin comme lieu d'installation et zone de migration. Un exemple très simple de notre vie quotidienne quand on veut migrer on choisira les zones où on connaît des personnes influenceuses et très bien classées comme le maire de la ville, l'imam ou tout autre responsable. Ce pouvoir de nos voisins correspond à leur degré interne dans leurs communautés. Le prédicat P se calcul ainsi comme suit:

$$P(o, com_i) = \sum_{v \in adj(o) \cap com_i} D(v, i) \quad (eq. III. 3)$$

Dans ce cas, l'individu migrant bénéficiera d'une protection maximale et évitera au maximum l'isolement et il peut être facilement intégré dans la nouvelle vie. Cependant, on peut dire que c'est plus difficile et plus lent et il est possible que ce migrant se retrouve loin de ses désirs, religion et traditions par rapport à la nouvelle population, y compris ses voisins.

#### 4. Profils de chaque communauté

Lorsque nous choisissons ce paramètre dans notre politique migratoire, les individus choisissent des populations d'un profil similaire à lui et surtout celui de leurs voisins pour s'installer. Le profil d'une communauté correspond à celui de ses individus.

Le prédicat P pourra être défini comme une variable booléenne pour dire que le profil d'un individu est similaire (*ou non*) à celui de telle communauté, comme suit :

**If (Profil(o)=Profil(com<sub>i</sub>)) then P(o,com<sub>i</sub>)=1 else P(o,com<sub>i</sub>)=0;endif.**

Comme il pourra être défini comme étant un pourcentage de similarité (*entre 0 et 1*) tout en calculant la distance (*Cosinus, Jaccard..etc*) entre les deux profils de l'individu et de la communauté.

Dans ce scénario, l'individu migrant aura le maximum de stabilité «psychique» et s'intégrera plus facilement dans la nouvelle vie. Cependant, on peut dire que c'est plus difficile et plus lent, mais aussi comment calculer le profil des individus et celui des communautés.

Si tous les paramètres de préférence sont égaux pour un individu dans deux ou plusieurs communautés, il en choisit une parmi elles car elles sont similaires pour lui.

#### **B. Niveau d'extension 2: non orienté / orienté**

Lorsque nous avons une relation asymétrique dans notre réseau social, comme l'abonnement dans «Twitter», le graphe associé sera orienté.

Nous construisons donc la matrice d'adjacence et prenons en compte les individus adjacents (*peu importe s'ils sont des successeurs ou prédécesseurs*) entre chaque deux individus et sans prise en considération le sens des arcs et, donc, le plus important est l'existence de la relation.

Ainsi, nous pouvons former l'ensemble des familles qui seront des futures communautés comme le montre la figure III.8. Pour l'intégration des nœuds «Out», la même logique et les mêmes paramètres de préférence sont suivis que dans le cas non orienté.

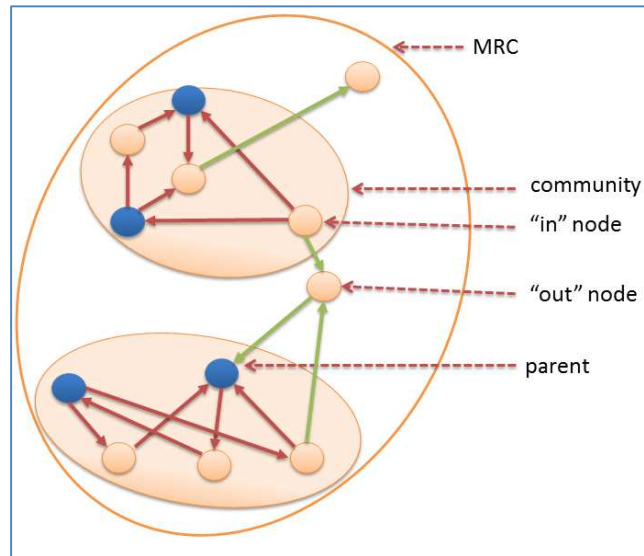


Figure III.8: Exemple de communautés dans le modèle LFM orienté.

La modularité a également une extension dans le cas des graphes orientés [135,136]. L'équation II.3 du chapitre précédent est modifiée comme suit [72]:

$$Q = 1/m \sum_{ij} (AM_{ij} - P_{ij}) \delta(C_i, C_j) \quad (\text{eq. III.4})$$

Où  $P_{ij} = K_i * K_j / m$ .

La même remarque peut être citée ici, c'est l'existence d'une relation entre les deux parents pour respecter le sens le plus juste de la communauté.

Pour ce cas d'extension et dans la partie d'implémentation du prochain chapitre, nous l'appliquerons sur un réseau orienté que nous allons créer nous même à partir d'une messagerie de la banque américaine ENRON[127] où la relation entre les individus est orienté (l'envoi d'un mail).

### C. Niveau d'extension 3: dynamique / statique

Étant donné que la plupart des réseaux réels changent au fil du temps et que la composition de ses communautés peut changer d'un moment à l'autre, dans cette sous-section, nous montrons comment notre algorithme est adapté pour prendre en charge ce nouvel aspect dynamique.

Dans l'aspect dynamique des réseaux sociaux, quatre types d'événements peuvent être déclenchés à tout moment. On peut avoir :

- 1- Abonnement d'une nouvelle personne.
- 2- Création de nouvelles relations entre individus.
- 3- Désinscription d'un ancien individu
- 4- Suppression d'une relation existante entre deux individus.

En plus de cela, nous pouvons capturer l'impact de l'exécution d'une combinaison de ces événements sur la qualité des structures communautaires au fil du temps et la stabilité de notre algorithme.

Ces quatre événements peuvent changer l'état de la composition communautaire du réseau comme suit:

- 1- Naissance d'une nouvelle communauté
- 2- Mort d'une ancienne communauté
- 3- Fusion de deux ou plusieurs communautés
- 4- Scission d'une ou plusieurs communautés en deux ou plusieurs
- 5- Croissance d'une ou plusieurs communautés
- 6- Contraction d'une ou plusieurs communautés.

Dans ce qui suit, nous détaillons comment mettre à jour notre algorithme de base pour prendre en charge un de ces quatre événements: nouvel individu, les autres peuvent être traités comme des travaux futurs.

La première idée est de mettre à jour la matrice d'adjacence après l'inscription de ce nouvel individu, puis relancer l'algorithme sur la nouvelle matrice d'adjacence.

L'avantage de cette solution est la simplicité de sa mise en place. Cependant, nous pouvons dire que la réexécution de l'algorithme entier à nouveau sur la nouvelle matrice d'adjacence et après chaque arrivée d'un nouvel individu impliquera un temps de réponse énorme.

**Amélioration:** au lieu de mettre à jour la matrice d'adjacence et de réexécuter tout l'algorithme, nous nous concentrons uniquement sur les communautés affectées par le ou les événements déclenchés. Les autres restent telles quelles-sont. À cet effet, nous détaillons cet événement (*nouvel individu*) et expliquons de quelle façon nous modifions notre algorithme LFM proposé.

**Nouvel individu:** l'ajout d'un nouveau nœud au graphe peut invoquer de nouveaux liens avec les autres. Nous avons des nœuds voisins de ce nouveau nœud mais aussi des communautés voisines. Nous ne consultons que ces communautés et nous vérifions la modularité du réseau pour chaque tentative de jointure de ce nœud à ces communautés. La modularité la plus élevée implique la communauté destinataire de ce nouveau nœud.

Également, nous pouvons commencer par l'état stable et considérer que ce nouveau nœud est un nœud "Out", puis nous effectuons les mêmes étapes que précédemment (*section des paramètres de préférence*).

Pour la partie d'implémentation du prochain chapitre, nous exécutons cette extension sur un réseau social que nous créons d'une façon aléatoire et nous rajoutons au fur et à mesure des nouveaux nœuds puis nous discutons les résultats dans les deux cas orienté et non orienté.

#### III.4. LFM et les autres

Le tableau III.3 donne une comparaison qualitative entre notre algorithme LFM proposé et celui du Louvain.

	<b>Louvain</b>	<b>LFM</b>
Prise en charge des boucles	✗	✓
Chevauchement des communautés	✗	✓
Graphes orientés	✗	✓
Graphes pondérés	✗	✓
Aspect dynamique	✗	✓
Décidabilité	✗	✓
Stabilité	✗	✓
Complexité temporelle	O(m), m est le nombre d'arêtes [128].	O(n <sup>2</sup> ), n est le nombre de nœuds.

**Tableau III.3 : LFM Vs Louvain.**

Dans ce tableau comparatif, l'algorithme de Louvain ne supporte pas la notion de boucles (*relations réflexives*) dans les graphes ainsi que les communautés qui se chevauchent; il est conçu pour les graphes statiques et non orientés, et aussi il n'est ni décidable ni stable. Cependant, notre modèle proposé a pris soin de tous ces axes.

Dans le prochain chapitre et après l'implémentation de notre algorithme nous donnons plus de comparaison à d'autres en termes de performance, temps d'exécution et autres.

## **Conclusion**

Dans ce chapitre, nous avons proposé une nouvelle approche de détection de communautés dans les réseaux sociaux basée sur le nombre de voisins communs entre les individus. Le concept de base est inspiré de la vie sociale de ces derniers et donne une vision sociologique sur la façon de leur regroupement en communautés.

Notre approche est basée sur un nouveau concept, que nous avons proposé, des communautés à sens moyen, où nous avons pris soin des inconvénients des sens faibles et forts et ainsi nous avons combiné leurs avantages.

Également, nous avons proposé plusieurs extensions pour prendre en charge autres cas comme le chevauchement entre les communautés, les graphes orientés, pondérés ainsi que l'aspect dynamique, l'objectif de ces différentes extensions est de rendre notre approche générique en donnant le choix à ses utilisateurs de le paramétrer selon le domaine d'application et le cas d'étude.

Aussi, pour qualifier notre algorithme, nous l'avons comparé à celui du Louvain tout en montrant les points positifs de notre démarche.

Dans le prochain chapitre, nous commencerons la phase de mise en œuvre et expérimentation, nous détaillerons l'implémentation de notre algorithme ainsi que son application sur plusieurs réseaux sociaux réels et aléatoires afin de montrer la puissance de notre approche proposée.

## Chapitre 4: Implémentation, Expérimentations et Discussion.

### Table des matières

Introduction.....	79
IV.1. Framework.....	79
IV.2. Algorithme LFM.....	79
IV.3. Étude de cas.....	81
A. Les réseaux aléatoires.....	82
B. Le réseau ‘‘Zachary’s karate club’’.....	85
C. La multinationale «Enron».....	88
1. La base de données Enron.....	88
2. Construction du réseau social Enron.....	89
3. Application de l’algorithme de Louvain sur le réseau social Enron.....	91
4. Application de notre algorithme LFM sur le réseau social Enron.....	91
IV.4. Performances et étude comparative.....	93
A. Application sur un réseau aléatoire à grande échelle.....	93
B. Complexité.....	94
C. Temps d’exécution.....	94
Conclusion.....	95

Dans ce chapitre, nous donnons la démarche suivie de l’implémentation et de mise en place de notre modèle proposé puis nous démontrons, à travers des expériences, que notre algorithme est très efficace pour découvrir la structure communautaire des réseaux générés aléatoirement par ordinateur ainsi que ceux du monde réel, qu’ils soient de structure simple voire complexe.

Nous appliquons notre modèle sur un véritable réseau social du club de karaté de Zachary. En outre, nous décrivons les expériences de notre modèle sur un réseau à grande échelle d’une base de données de messagerie électronique de la banque américaine ‘‘Enron’’ en tant que réseau de référence plus large. Les résultats des simulations montrent que notre modèle proposé est très satisfaisant.

## Introduction

Chaque nouvelle approche proposée doit avoir sa phase de mise en œuvre où nous parlons de son côté technique de programmation et de mise en œuvre. Ainsi, sans oublier l'aspect de l'expérimentation et son application sur des données réelles ou aléatoires afin de montrer la performance de cette nouvelle approche via des résultats qui doivent être discutés, validés ou même mettre en cause ladite approche.

Dans ce chapitre, nous démontrons, à travers des expériences, que notre algorithme est très efficace pour découvrir la structure communautaire des réseaux sociaux générés aléatoirement par ordinateur ainsi que ceux du monde réel.

Nous appliquons donc notre modèle sur un véritable réseau social du club de karaté de Zachary. En outre, nous décrivons les expériences de notre modèle sur un réseau à grande échelle de l'ensemble de données de la messagerie électronique de la banque américaine "Enron" en tant que réseau plus large. Les résultats des simulations montrent que notre modèle proposé est très satisfaisant.

### IV.1. Framework

Dans cette étape de notre travail, nous utilisons un ordinateur portable i5 avec 4 Go de RAM. Pour sauvegarder les différentes informations à traiter nous utilisons des fichiers textes ainsi qu'une base de données MySQL. En tant que langages de programmation, nous avons utilisé PHP (*Python et Java pour certaines extensions*).

### IV.2. Algorithme LFM

Notre algorithme est basé sur la définition donnée dans la section 2.1 du chapitre III. Il faut donc garder à l'esprit que le réseau social peut avoir un ou plusieurs composants connexes, chacun d'entre eux pouvant avoir des communautés de tailles différentes d'un autre composant connexe. Dans cette section, nous donnons un pseudo code de notre algorithme sans aller trop loin dans les détails d'implémentation.

**Input** : Graph;// Adjacency Matrix of G

**Output** : Initial and final communities' distributions.

**Begin**

Calculate MRC;// set of Maximum Related Components

**For** each MRC<sub>1</sub> in MRC **do**

    CadjMax(**l**)=0;     // maximum number of common adjacent nodes between all nodes of MRC<sub>1</sub>

**For** each i in MRC<sub>1</sub> **do** // node by node

**For** each j in MRC<sub>1</sub> **do** // node by node

            Calculate Cadj(i,j);     //set of common adjacent nodes between i and j

**If** (CadjMax(**l**) < |Cadj(i,j)|) **then**

                CadjMax(**l**) ← |Cadj(i,j)|; **EndIf**

**EndFor**;

**EndFor**;

// Calculer l'ensemble des parents de MRC<sub>1</sub>

Parents=array();// Dynamic array of two lines.

K=0;

**For** each i in MRC<sub>1</sub> **do** // node by node

**For** each j in MRC<sub>1</sub> **do** // node by node



```

    If (Cadj(i,j)==CadjMax(I)) then
        Parents[0][k] ← i;
        Parents[1][k] ← j;
        K++;
    EndIf
EndFor;
EndFor;
Nbr_parents=k-1;

// Étape 1: générer l'ensemble des communautés initiales de MRC1 et Out_nodes
out1 ← MRC1; // initially, all nodes of MRC1 are out
//in1 ← {};
For (k=0; k < Nbr_parents; k++) do
    Comk ← { Parents[0][k] } U { Parents[1][k] } U Cadj(Parents[0][k], Parents[1][k]);
    out1 ← out1 / Comk;
    // in1 ← in1 U Comk;
EndFor;

// Étape 2: joindre les nœuds ‘out’ aux communautés
For each o in out1 do
    Dmaxo = 0;
    For (k=0; k < Nbr_parents; k++) do
        Calculate Absbo,k;
        Do,k = | Absbo,k |;
        If (Dmaxo < Do,k) then
            Dmaxo ← Do,k;
        EndIf;
    EndFor;

    For (k=0; k < Nbr_parents; k++) do
        If (Do,k = Dmaxo) then
            Comk ← Comk U {o};
        EndIf;
    EndFor;
    out1 ← out1 / {o};
    // in1 ← in1 U {o};
EndFor;
// out1 == {};
Print the Set of Initial communities of MRC1 (without merging)

// Étape 3: fusion de ces communautés sous certaines conditions
For (k=0; k < Nbr_parents; k++) do // community by community
    For (r=0; r < Nbr_parents; r++) do // community by community
        Nbr_agree = |Comk ∩ Comr|;
        Nbr_disagree = |Comk U Comr| - Nbr_agree;
        If (Nbr_agree ≥ Nbr_disagree) then // |Comk ∩ Comr| ≥ (|Comk U Comr|)/2
            Comk ← Comk U Comr; // merging of Comk and Comr;
            Delete(Comr);
        EndIf;
    EndFor;
EndFor;

```

**EndFor;**

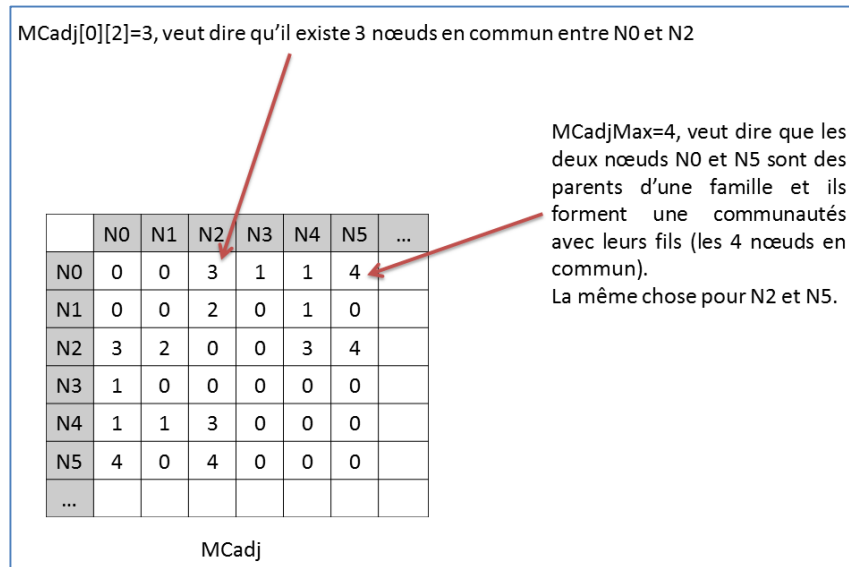
Print the Set of **Final** communities of  $MRC_L$  (with merging)

**EndFor;** //  $MRC_1$  by  $MRC_1$

**End.**

**NB:** À l'étape 3, nous fusionnons chaque deux communautés qui sont plus d'accord que désaccords (*l'intersection entre elles est plus importantes que la différence*).

Dans la phase de calcul des voisins communs (*files*) entre les nœuds (*parents*) nous avons utilisé une matrice  $MCadj$  de taille  $N \times N$  où  $MCadj[i][j]=nbr\_voisins\_en\_commun$  comme le montre la figure VI.1:



**Figure IV.1: La matrice de voisinage en commun  $MCadj$ .**

À partir de cette matrice nous pouvons calculer la valeur de  $CadjMax$  comme suit :  $CadjMax=Max_{i,j}(MCadj[i][j])$ . En trouvant la valeur de  $CadjMax$  on pourra déterminer l'ensemble des parents, leurs fils et donc l'ensemble des familles (*communautés*) et par la suite les nœuds "Out" qui vont choisir quelques communautés pour les joindre.

### IV.3. Étude de cas

Dans cette section nous envisageons deux cas, les réseaux sociaux aléatoires et ceux réels. Tout d'abord (*réseaux sociaux aléatoires*), nous avons implémenté et exécuté l'algorithme de Louvain sur un exemple du graphe traité par ses auteurs, et nous exécutons notre algorithme sur le même graphe pour comparer les résultats. Deuxièmement (*réseaux sociaux réels*), nous avons appliqué notre algorithme sur deux ensembles de données, l'un est le plus utilisé et créé en tant que réseau social, c'est le réseau de club de karaté de Zachary, le second est une base de données de mailing où nous avons créé un réseau social entre ses utilisateurs et la relation sociale est "envoyer à / recevoir à partir de" des courriers électroniques.

### A. Les réseaux aléatoires

Nous avons implémenté et exécuté l'algorithme de Louvain sur l'exemple du graphe traité par Blondel et al [4]. De plus, nous exécutons notre algorithme sur le même exemple. Dans cet exemple, le graphique se compose de 16 nœuds et de 28 arêtes. La matrice d'adjacence est illustrée dans la figure IV.2.

S\Sj	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
S0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
S1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0
S2	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0
S3	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
S4	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
S5	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
S6	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0
S7	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0
S8	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1
S9	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0
S10	0	0	0	0	1	0	0	0	1	0	0	1	1	1	1	0
S11	0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0
S12	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
S13	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
S14	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
S15	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Figure IV.2: La matrice d'adjacence de l'exemple traité par Blondel et al. [4].

Le graphe de cet exemple est donné dans la figure IV.3.

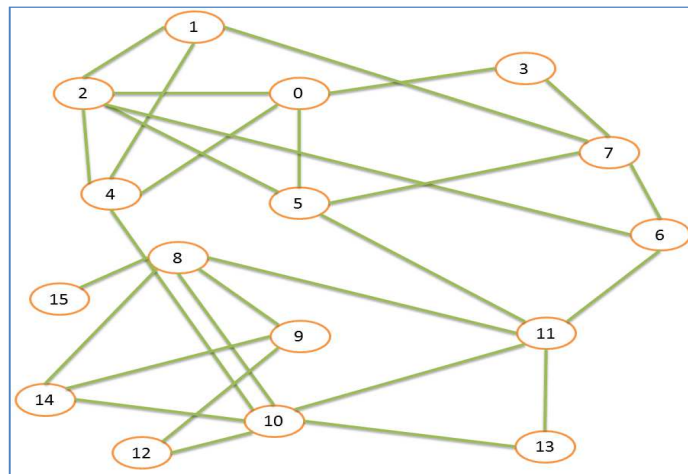


Figure IV.3: Exemple traité par Blondel et al. [4] avant la détection des communautés.

Après exécution de notre implémentation de l'algorithme de Louvain sur cet exemple, nous détectons quatre communautés (*les mêmes trouvées par Blondel*), comme le montre la figure IV.4. Il est à noter que la valeur de la fonction de modularité de cette distribution communautaire est  $Q=0.34630102040816$ .

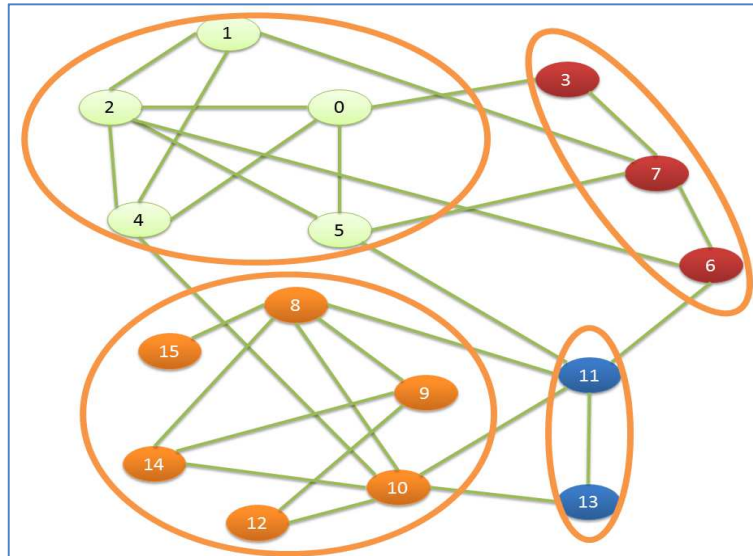


Figure IV.4: Communautés trouvée par Louvain sur l'exemple de Blondel et al. dans [4].

La première étape de l'algorithme LFM consiste à détecter, pour chaque MRC, la valeur de  $C_{adjMax}$  ainsi que l'ensemble des parents et donc l'ensemble des communautés initiales. La valeur maximale de  $C_{adjMax}$  détectée est égale à 3, donc l'ensemble des parents est  $(N_2, N_7)$ ,  $(N_5, N_6)$  et  $(N_9, N_{10})$  et donc, les trois communautés initiales suivantes:  $Com_0 = \{N_2, N_1, N_5, N_6, N_7\}$ ,  $Com_1 = \{N_5, N_2, N_7, N_{11}, N_6\}$  et  $Com_2 = \{N_9, N_8, N_{12}, N_{14}, N_{10}\}$ . De plus, nous avons détecté les cinq nœuds «out» suivants:  $\{N_0, N_3, N_4, N_{13}, N_{15}\}$ .

Comme nous l'avons cité dans la section 2.3 du chapitre III, nous avons rendu l'algorithme LFM déterministe et par conséquent il est devenu stable. Le premier nœud ayant un degré maximum dans ces communautés initiales est  $N_0$ , son degré est 2 dans  $Com_0$  et  $Com_1$ , et donc il rejoint ces deux communautés. L'étape suivante est pour  $N_3$  qui rejoint les nouvelles communautés  $Com_0$  et  $Com_1$ , et ainsi de suite, jusqu'à ce qu'il n'y ait plus de nœuds à traiter. Finalement, nous obtenons la distribution communautaire suivante,  $Com_0 = \{N_2, N_1, N_5, N_6, N_7, N_0, N_3, N_4\}$ ,  $Com_1 = \{N_5, N_2, N_7, N_{11}, N_6, N_0, N_3, N_{13}\}$  et  $Com_2 = \{N_9, N_8, N_{12}, N_{14}, N_{10}, N_{13}, N_{15}\}$ . La distribution communautaire trouvée est illustrée dans la figure IV.5:

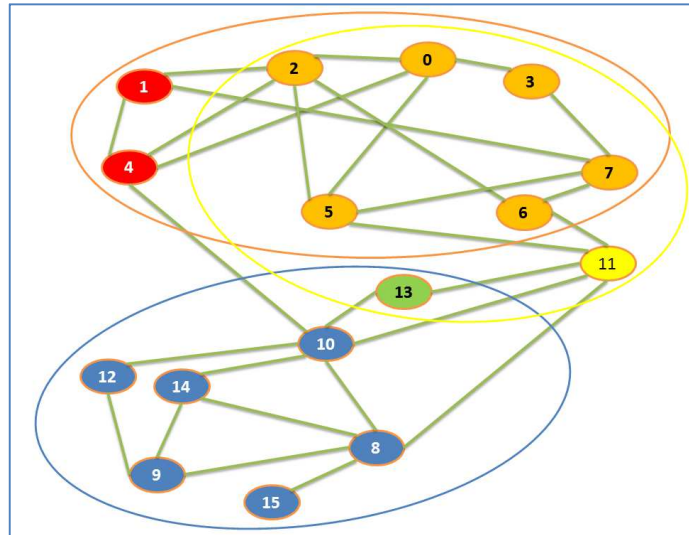


Figure IV.5: Communautés trouvée par LFM sur l'exemple de Blondel et al. dans [4] (avant exécution de sa phase 3).

Nous notons, et comme le montre la figure IV.5, que ces communautés finales se chevauchent **contrairement** à l'algorithme de Louvain qui ne supporte pas les communautés chevauchantes.

En appliquant la phase de fusion des communautés selon sa condition donnée à la section 2.2 du chapitre III, nous fusionnons les deux communautés Com0 et Com1 en une seule, de sorte que le résultat final est illustré dans la figure IV.6.

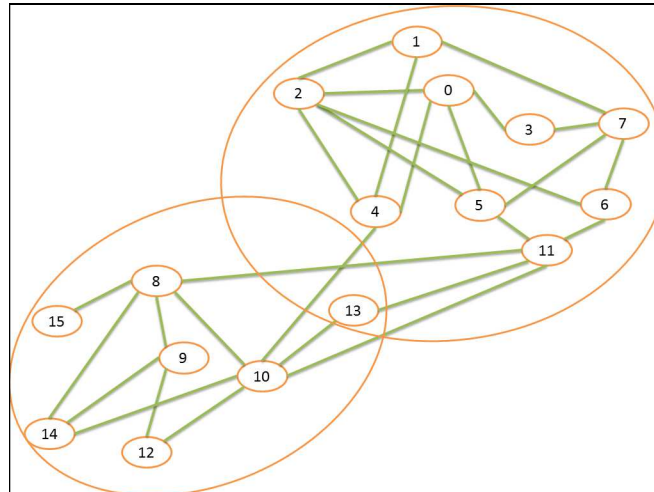


Figure IV.6: Communautés finales trouvées par LFM sur l'exemple de Blondel et al. dans [4].

La valeur de la fonction de modularité de cette distribution communautaire est  $Q=0.3265306122449$ , c'est moins que celle trouvée par l'algorithme du Louvain. Notre objectif est donc de trouver une distribution communautaire optimale mais pas forcément la meilleure. Le plus important pour LFM, par rapport à Louvain et beaucoup d'autres est sa stabilité (*donne toujours le même résultat avec plusieurs exécutions sur le même réseau*). À

cet effet, nous pouvons dire que LFM cherche un **compromis** entre les deux notions : la stabilité et la modularité.

### B. Le réseau ‘Zachary’s karate club’

Nous appliquons notre algorithme à un véritable réseau social, le réseau de club de karaté de Zachary [123]. Dans ce travail, les auteurs observent sur deux ans, 34 membres d'un club de karaté dans une université américaine. Au cours de cette étude, un désaccord s'est développé entre l'administrateur du club et son instructeur, qui a permis à l'instructeur de quitter et de créer un nouveau club, et il a pris près de la moitié du club. Zachary a construit un réseau d'amitié entre les membres du club d'origine (34 nœuds et 78 bords) afin d'en étudier la **conspiration** et les différents **complots** qui ont eu lieu dans le club d'origine.

Dans ce travail, nous utilisons une version simplifiée de ce réseau et nous appliquons notre algorithme LFM pour identifier les complots impliqués dans ce club.

Pour voir l'impact du nombre de relations communes entre les membres sur la structure communautaire finale, nous varions la valeur du CadjMax et voyons l'évolution de la composition de la communauté (figure IV.7) ainsi que des membres "out\_node" (figure IV.8) et donc le temps d'exécution (figure IV.10) qu'ils n'avaient rien à voir avec les complots du départ (communautés initiales).

L'exécution de l'algorithme de Louvain dans cet exemple a détecté six (06) communautés.

L'exécution de notre algorithme LFM sur cet exemple a trouvé un maximum de dix (10) membres communs entre seulement  $N_{32}$  et  $N_{33}$ . À cet effet, il n'aura qu'une seule communauté et les 22 autres membres seront considérés comme des "noeuds\_out" qui rejoindront cette communauté.

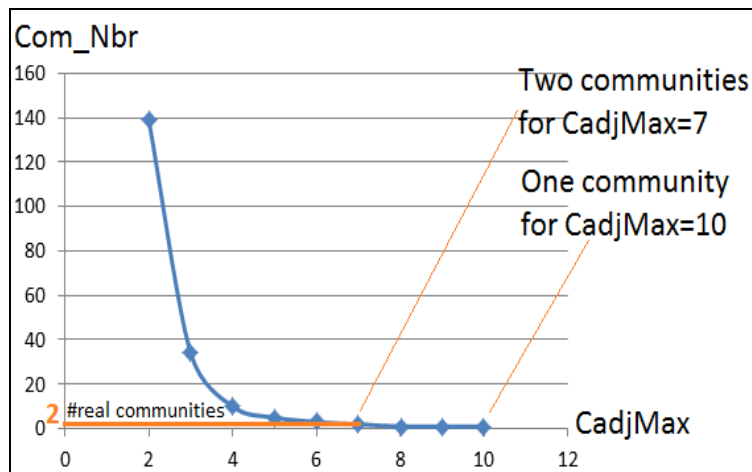


Figure.IV.7 : Zachary Dataset: Communities number vs. CadjMax.

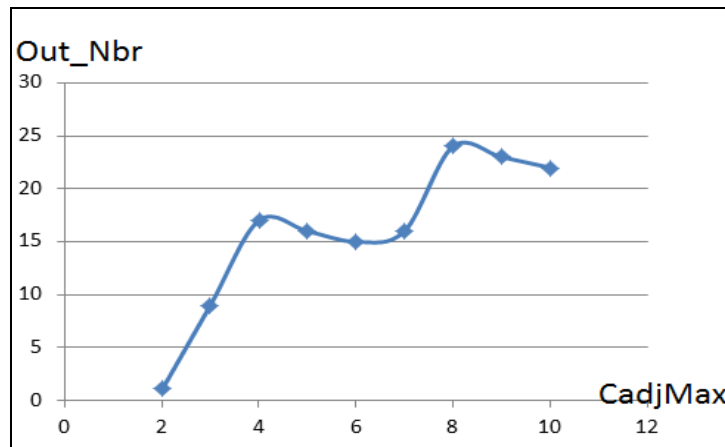


Figure.IV.8: Zachary Dataset: Nombre de nœuds "sortants" par rapport à CadjMax.

Le club de karaté de Zachary a été divisé réellement en deux, donc le nombre réel de communautés doit être égal à deux. D'après nos résultats de variation de la valeur de CadjMax, ce nombre réel correspond à la valeur de CadjMax qui est égale à sept ( $CadjMax=7$ ). Pour cet exemple de CadjMax, les deux parents sont  $(N_0, N_1)$  et  $(N_{32}, N_{33})$ . La distribution communautaire initiale est  $C_1=\{N_0; N_2; N_3; N_7; N_{13}; N_{17}; N_{19}; N_{21}; N_1\}$  et  $C_2=\{N_{32}; N_8; N_{14}; N_{15}; N_{18}; N_{20}; N_{22}; N_{23}; N_{33}\}$ .

Notant que les premiers parents de ces deux communautés sont  $N_0$  et  $N_{32}$  (détectés par notre algorithme proposé), ils représentent respectivement l'administrateur et l'instructeur qui ont divisé le club en deux parties. Par conséquent, et comme observation importante, l'algorithme LFM est capable de détecter les principaux acteurs (et leurs partenaires: le deuxième élément) qui ont initialement formé les familles puis les communautés, ainsi que (comme hypothèse à prouver) l'historique (ou bien l'ordre) de jointure de chaque individu à telle ou telle communauté.

L'ensemble des nœuds "out" est  $\{N_4; N_5; N_6; N_9; N_{10}; N_{11}; N_{12}; N_{16}; N_{24}; N_{25}; N_{26}; N_{27}; N_{28}; N_{29}; N_{30}; N_{31}\}$ .

La distribution communautaire finale est (comme hypothèse à prouver, nous remarquons l'ordre de la jointure de chaque nœud «out» à sa communauté):

$C_1=\{N_0; N_2; N_3; N_7; N_{13}; N_{17}; N_{19}; N_{21}; N_1; N_{12}; N_4; N_6; N_5; N_{10}; N_{16}; N_9; N_{11}\}$  et  $C_2=\{N_{32}; N_8; N_{14}; N_{15}; N_{18}; N_{20}; N_{22}; N_{23}; N_{33}; N_{29}; N_{30}; N_{26}; N_{27}; N_{31}; N_{24}; N_{25}; N_{28}; N_9\}$ . On remarque également le chevauchement entre ces deux communautés ( $N_9$  est partagé). Comme une autre remarque, nous observons l'équité (en termes de nombre d'individus) entre les deux communautés détectées par notre algorithme,  $C_1$  contient 17 membres (la moitié) et  $C_2$  contient 18 ( $N_9$  en plus) et c'est le cas réel où le club d'origine été divisé en deux.

Maintenant, nous devons vérifier la meilleure partition communautaire selon l'algorithme LFM qui correspond à la valeur maximale de CadjMax. La figure IV.9 donne la variation de la modularité par rapport à la valeur de CadjMax.



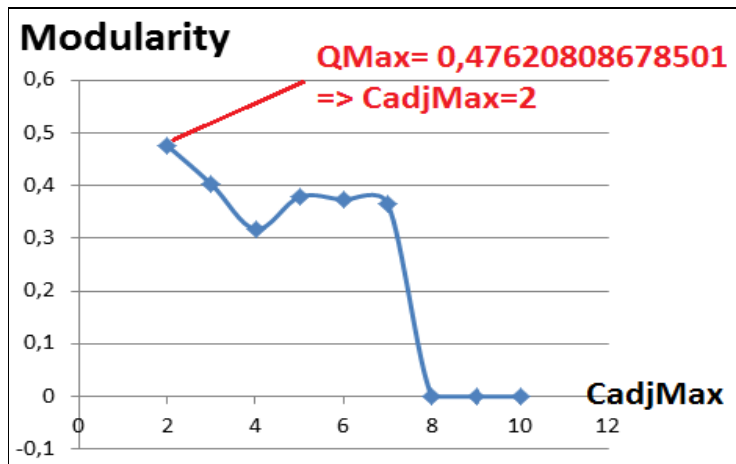


Figure.IV.9: Zachary Dataset: modularité vs CadjMax.

La valeur maximale de modularité correspond à la valeur de CadjMax=2. Pour avoir le nombre réel de communautés (*deux*), nous devons avoir une valeur de CadjMax=7, ce qui implique une valeur de modularité égale à  $Q=0,36534352399737$ .

Nous concluons que selon l'algorithme LFM, et pour en avoir la solution la plus optimale (*selon la fonction objectif de modularité*), le nombre adéquat de voisins en commun à prendre en considération est en effet deux et non pas sept.

Pour mesurer **le temps d'exécution**, nous modifions la valeur du CadjMax et voyons son impact sur le temps d'exécution et de détection des communautés finales et s'il existe une relation entre cette valeur, le nombre de communautés, le nombre de nœuds "out" et le temps d'exécution? (*figure IV.10*)

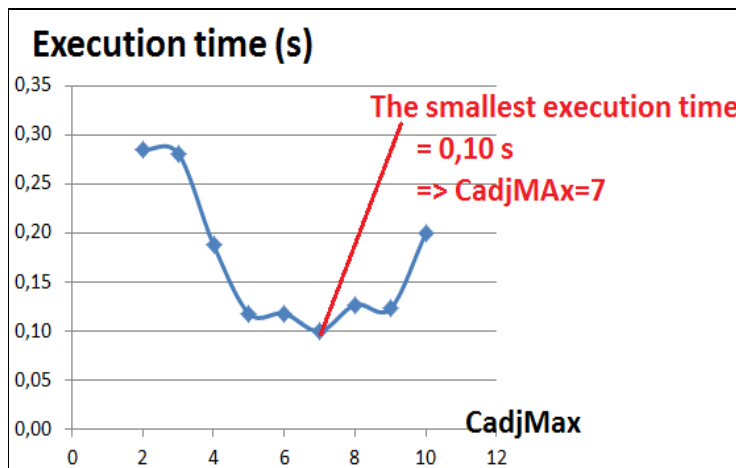


Figure.IV.10: Zachary Dataset: Temps d'exécution vs CadjMax.

Sur la figure IV.9, le plus petit temps d'exécution est celui correspondant à la valeur CadjMax=7, **c'est le cas** pour les communautés réelles (*deux communautés*).



### **C. La multinationale «Enron»**

Dans cette section, nous montrons l'évolutivité de notre modèle qui peut supporter des réseaux sociaux plus larges et orientés, nous appliquons notre algorithme, et aussi celui de Louvain, à un véritable réseau social que nous créons à partir de la base de données des emails de l'entreprise multinationale de domaine énergétique, pétrolier, financier et bancaire ENRON [127] qui a déclaré en faillite en décembre 2001.

Cette base de données est un ensemble important de courriels cités dans l'enquête de la «Federal Energy Regulatory Commission» des États Unis, qui a ensuite été mis à la disposition du grand public.

Le choix de cette base de données comme un échantillon d'étude et d'application de notre algorithme est justifié par les points suivants :

- 1- Cet ensemble de données est également similaire au type de données collectées pour la détection de la fraude ou le contre-terrorisme et est donc un banc d'essai parfait pour tester l'efficacité des techniques utilisées pour la détection de la fraude et le contre-terrorisme [127].
- 2- La faisabilité technique d'en créer un réseau social à partir de sa base de données des emails échangés entre ses employés.
- 3- La base de données contient des milliers de messages envoyés plus de deux millions fois vers des milliers d'adresses email différentes, la chose qui nous a permis de tester notre modèle sur un réseau social plus large afin d'en voir son côté extensibilité, flexibilité et souplesse.
- 4- La nature de la relation de contact (*mailing*) est une relation asymétrique, la chose qui nous permet d'en appliquer notre modèle sur un réseau social orienté.

Dans cette base de données (*disponible sur [http://www.ahschulz.de/pub/R/data/enron-mysqldump\\_v5.sql.gz](http://www.ahschulz.de/pub/R/data/enron-mysqldump_v5.sql.gz)*) mise en ligne par plusieurs ... en plusieurs versions dont celle qu'on a travaillé, et pour nos expériences, le nombre de messages est égal à 252.759 où certains ont été supprimés en raison des demandes des employés concernés. Ces messages ont été envoyés et reçus 2.064.442 fois vers et depuis 68.074 adresses e-mail distinctes. Le nombre d'utilisateurs (*employés*) qui échangent des e-mails entre eux est de 149.

#### **1. La base de données Enron**

Avec la base de données Enron, nous dérivons un réseau social de 149 employés. La base de données Enron contient les quatre tables illustrées sur la figure IV.11.

- Table "EmployeeList": représente la liste des employés, elle a neuf attributs: eid, firstName, lastName, Email\_id, Email2, Email3, Email4, dossier, statut.
- Table "Messages": contient les informations des messages électroniques de l'expéditeur, l'objet, le texte et d'autres informations. Elle a sept attributs: mid, sender, date, message\_id, Subject, body, folder.
- Table "RecipientInfo": est l'entité qui stocke les informations du destinataire. Elle contient l'adresse e-mail du destinataire, l'identifiant du message envoyé et le type d'envoi (*TO*, *CC*, *BCC*). Elle a cinq attributs: rid, mid, rtype, rvalue, date.
- Table "ReferenceInfo": stocke les références d'informations. Elle a trois attributs: rfid, mid, reference.

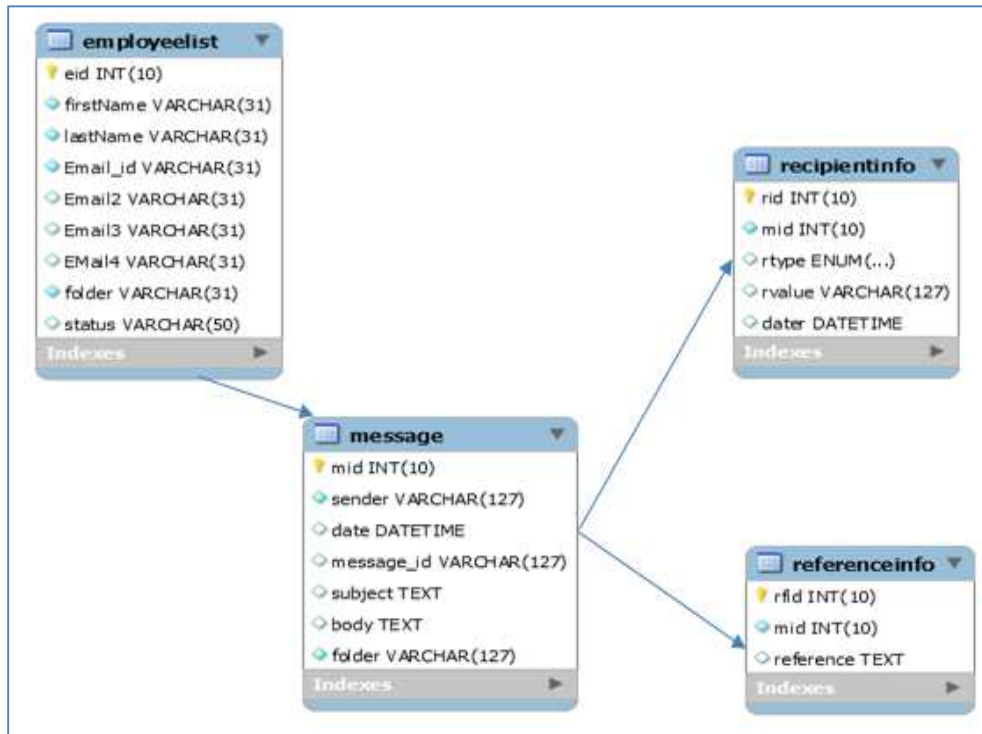


Figure IV.11: Les tables de la base de données Enron.

## 2. Construction du réseau social Enron

Afin d'en créer le réseau social d'Enron à l'aide de cette base de données, nous avons créé la requête SQL suivante pour obtenir les individus et les liens qui existent entre eux.

```

SELECT tmpoo.eid_sender, tmpoo.mailsender, tmpoo.mailrecipient, employeelist.eid eid_recipient,
tmpoo.nbr_mid_msg
FROM employeelist
JOIN (
SELECT employeelist.eid eid_sender, tmpo.mailsender, tmpo.mailrecipient, tmpo.nbr_mid_msg
FROM employeelist
JOIN (
SELECT message.sender mailsender, recipientinfo.rvalue mailrecipient, COUNT(message.mid)
nbr_mid_msg
FROM message
JOIN recipientinfo ON message.mid=recipientinfo.mid
WHERE recipientinfo.rvalue IN (
SELECT employeelist.Email_id
FROM employeelist) AND message.sender IN (
SELECT employeelist.Email_id
FROM employeelist)
GROUP BY message.sender, recipientinfo.rvalue) tmpo ON tmpo.mailsender = employeelist.Email_id)
tmpoo ON tmpoo.mailrecipient =employeelist.Email_id
ORDER BY mailsender
    
```

Le résultat de cette requête est illustré dans la figure IV.12.

eid_sender	mailsender	mailrecipient	eid_recipient	nbr_mid_msg
99	albert.meyers@enron.com	bill.williams@enron.com	135	5
99	albert.meyers@enron.com	ryan.slinger@enron.com	131	2
89	andrea.ring@enron.com	gerald.nemec@enron.com	75	1
89	andrea.ring@enron.com	scott.neal@enron.com	96	1
89	andrea.ring@enron.com	sandra.f.brawner@enron.com	120	16
89	andrea.ring@enron.com	judy.townsend@enron.com	81	1
89	andrea.ring@enron.com	peter.keavey@enron.com	62	1
89	andrea.ring@enron.com	chris.germany@enron.com	114	1
89	andrea.ring@enron.com	richard.ring@enron.com	5	22
89	andrea.ring@enron.com	brad.mckay@enron.com	47	1
89	andrea.ring@enron.com	scott.hendrickson@enron.com	63	1
78	andy.zipper@enron.com	phillip.k.ellen@enron.com	10	3
78	andy.zipper@enron.com	thomas.a.martin@enron.com	103	4
...	...	...	...	...

Figure.IV.12: Relations de contact entre les membres d'Enron.

Dans la figure IV.12, par exemple, la ligne numéro un (1) indique que M. Albert Meyers avec l'identifiant numéro 99 a envoyé exactement cinq (5) messages différents à M. Bill Williams.

Ainsi, la matrice d'adjacence à créer est de taille 149X149, chaque case  $[i, j]$  contient la valeur 0 s'il n'y a pas de messages électroniques échangés entre les deux individus  $i$  et  $j$ , et 1 si  $i$  a envoyé au moins un message à  $j$ . Le réseau construit correspond à la relation "a contacté", cette relation est asymétrique et donc le graphique correspondant est orienté

Le réseau avant l'application d'aucun algorithme de détection des communautés est comme illustré dans la figure IV.13.

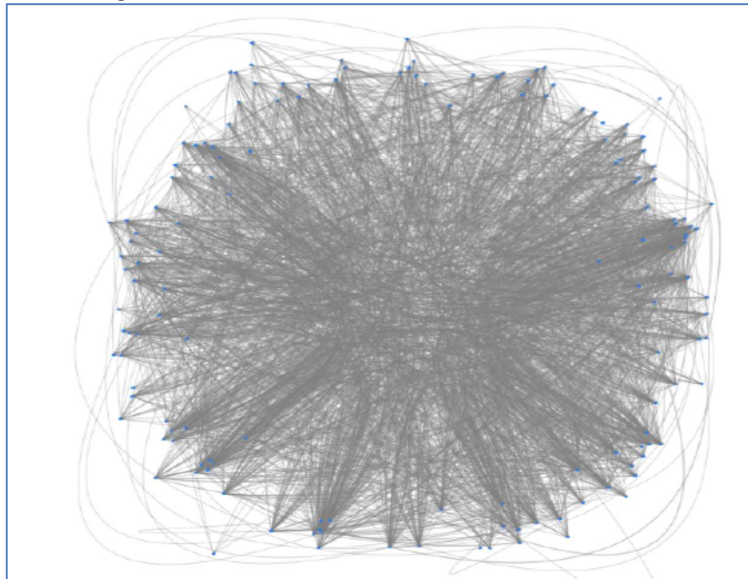


Figure.IV.13: Réseau d'ENRON avant l'application d'aucun algorithme de détection communautaire.

### 3. Application de l'algorithme de Louvain sur le réseau social Enron

En appliquant l'algorithme de Louvain, nous avons détecté 14 communautés avec une modularité égale à  $Q = 0,071883064734281$ . Le graphe après détection est illustré dans la figure IV.14.

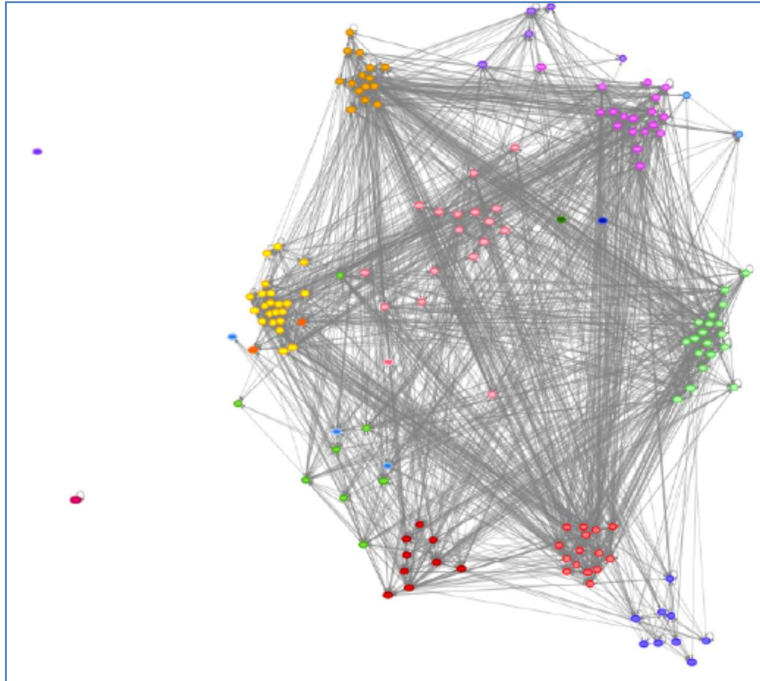


Figure.IV.14: Réseau d'ENRON après application de l'algorithme de Louvain.

### 4. Application de notre algorithme LFM sur le réseau social Enron

En appliquant notre algorithme LFM sur ce réseau social, nous avons trouvé une valeur de 52 voisins communs entre  $N_{53}$  (*M. John Lavorato*) et  $N_{122}$  (*M. Sally Beck*) et nous avons donc une seule communauté avec  $CadjMax = 52$  où  $N_{53}$  et  $N_{122}$  comme parents. Nous varions cette valeur de  $CadjMax$ , les statistiques suivantes (figures IV.15, IV.16, IV.17 et IV.18) sont trouvées.

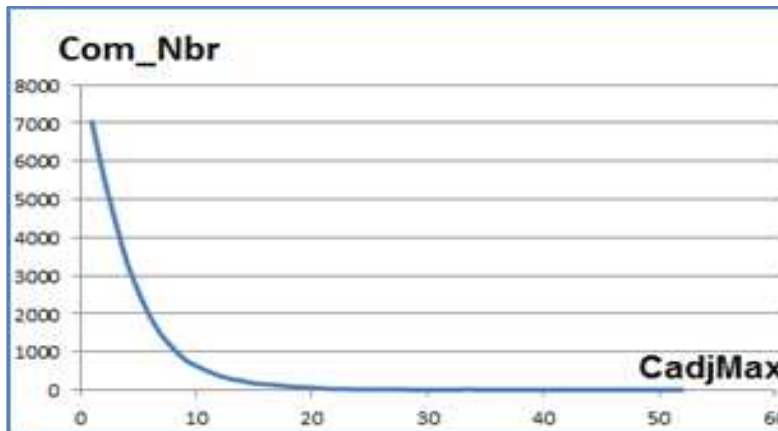


Figure IV.15: Le réseau social d'ENRON: nombre de communautés vs CadjMax

Nous remarquons dans la figure II.15 que la valeur CadjMax influence énormément nos résultats en termes de nombre de communautés et donc la distribution communautaire finale. Aussi, le chiffre **7000** communautés pour CadjMax=1 a eu lieu à cause de la commutativité des parents et les différentes combinaisons entre ces derniers et leurs fils. Exemple, pour CadjMax=1 avec uniquement quatre nœuds fortement connectés, on pourra trouver **les vingt (20)** communautés suivantes  $\{(1,3,2), (2,3,1), (1,4,2), (2,4,1), (1,2,3), (3,2,1), (1,4,3), (3,4,1), (1,2,4), (4,2,1), (1,3,4), (4,3,1), (2,1,3), (3,1,2), (2,4,3), (3,4,2), (3,1,4), (4,1,3), (3,2,4), (4,2,3)\}$ .

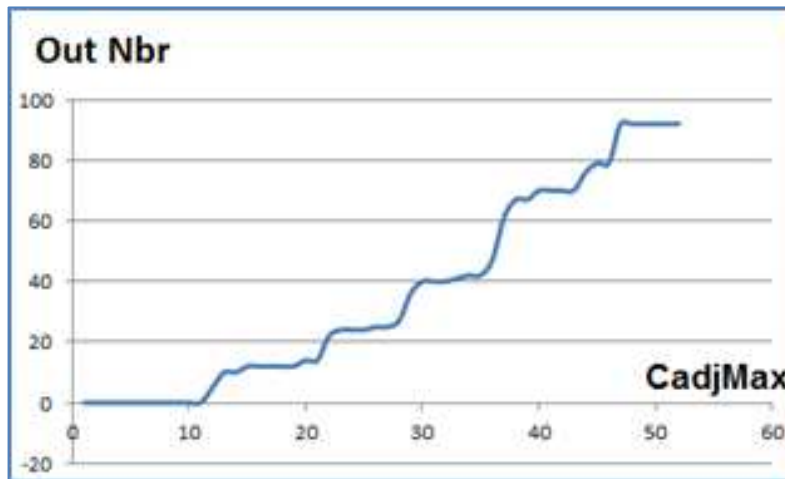


Figure IV.16: Le réseau social d'ENRON: nombre de nœuds "Out" vs CadjMax

Nous remarquons, dans la figure IV.16 que l'augmentation de la valeur du CadjMax implique forcément une augmentation de nombre de nœuds "Out". Augmentant la valeur de CadjMax veut dire une obligation d'en avoir un nombre élevé de fils entre les parents des communautés initiales, la chose qui impliquera un petit nombre de parents satisfaisant cette condition et donc un nombre petit de communautés initiales et aussi beaucoup de nœuds "Out".

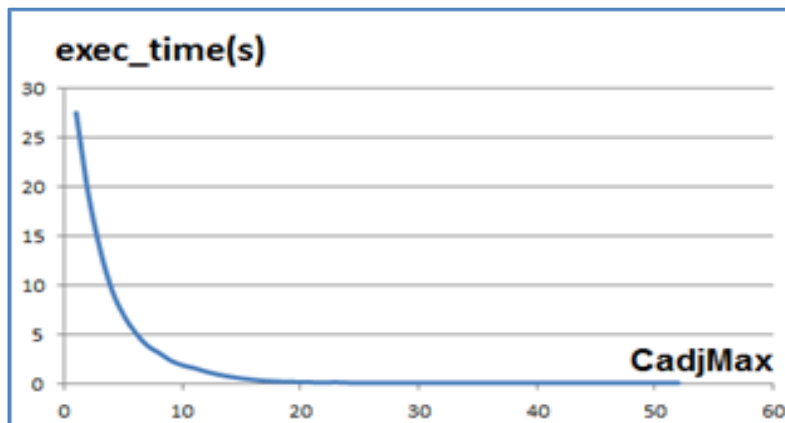


Figure IV.17: Ensemble de données ENRON: temps d'exécution vs CadjMax

Dans la figure IV.17 nous remarquons que l'augmentation du CadjMax implique une réduction du temps d'exécution. Ainsi, des petites valeurs pour CadjMax impliquent un temps d'exécution énorme à cause du nombre importants des parents satisfaisant la condition de "CadjMax" fils en commun. Un nombre important de communautés implique un

processus assez compliqué de jointure des nœuds ‘‘Out’’ à ces dernières (*chaque nœud ‘‘Out’’ aura autant de choix entre ce nombre important des communautés pour les joindre*), la chose qui augmentera le temps d’exécution.

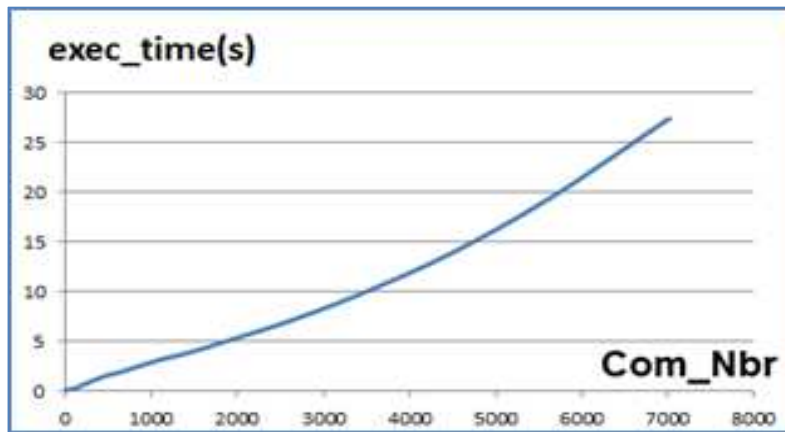


Figure IV.18: Le réseau social d'ENRON: temps d'exécution vs nombre de communautés

Dans la figure IV.18, l’augmentation de nombre de communautés impliquera une multitude de choix pour les nœuds ‘‘Out’’ d’en choisir telle(s) ou telle(s) communauté(s), la chose qui impliquera un temps d’exécution énorme par rapport à un nombre réduit de communautés.

#### IV.4. Performances et étude comparative

##### A. Application sur un réseau aléatoire à grande échelle

En utilisant une fonction aléatoire, nous créons un réseau social, avec notre ordinateur portable i5 de 4 Go de RAM, et nous simulons le temps d'exécution par rapport au nombre de nœuds et de liens (*avec leurs positions entre les nœuds*) générés aléatoirement. Dans les deux cas, orienté et non orienté, nous avons eu le tableau statistique IV.1 et la figure IV.19.

N	Orienté		Non-Orienté	
	M	Temps-Exec (Ms)	M	Temps-Exec (Ms)
100	5 027	64	5 021	72
200	19 761	243	20 127	207
300	44 729	680	44 942	653
400	80 272	1 558	80 426	1 633
500	125 016	3 059	125 053	3 011
600	179 926	5 355	179 982	5 128
700	245 289	8 705	244 246	8 306
800	320 042	17 924	320 736	16 980
900	404 953	27 346	405 428	23 829
1000	499 991	38 138	499 945	36 615
2 000	2 000 724	423 044	2 001 498	323 510

Tableau IV.1: Application sur un réseau aléatoire à grande échelle: statistiques

Nous remarquons que le temps d'exécution augmente linéairement avec le nombre de nœuds de notre réseau, et donc, nous pouvons dire que notre modèle est extensible et évolutif et sa complexité temporelle est d'ordre  $O(N^2)$  vu que le parcours le plus long se fait au niveau de la première phase où nous cherchons l'ensemble des parents en vérifiant les nœuds deux à deux.

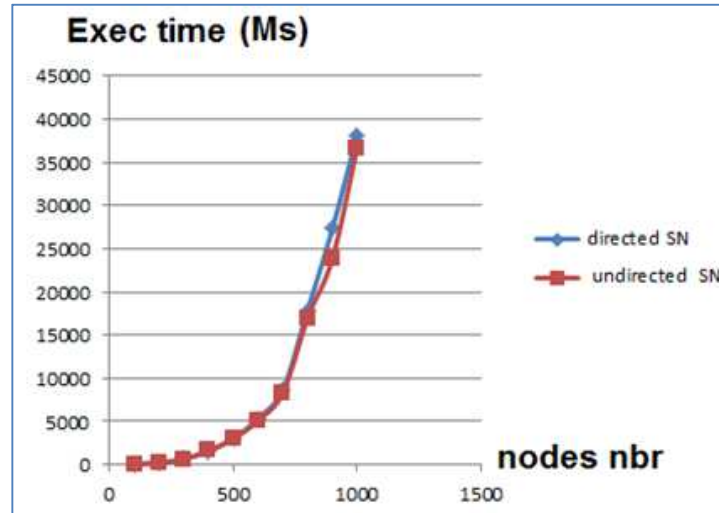


Figure IV.19: Application sur un réseau aléatoire à grande échelle: nombre de nœuds en fonction du temps d'exécution.

On note également que la détection des communautés avec le modèle LFM dans les réseaux sociaux non orientés est plus rapide que celle dans les réseaux orientés.

### B. Complexité

La complexité de l'algorithme de Louvain est  $O(m)$ ,  $m$  étant le nombre d'arêtes [128], tandis que notre algorithme LFM a une complexité égale à  $O(n^2)$ ,  $n$  est le nombre de nœuds. Cette complexité est obtenue vu que la recherche des nœuds conjoints qui ont formé des familles se fait en parcourant chaque nœud deux fois, le premier pour le «mari» et le second pour «sa femme».

### C. Temps d'exécution

Le tableau IV.2 donne une comparaison quantitative en fonction du temps d'exécution entre notre algorithme LFM proposé et les six algorithmes proposés par Atay et al. [101] sur le jeu de données Zachary [123].

Réseau / algo	Temps d'exécution sur chaque réseau (sec.)						
	HDSA	BADE	SSGA	BB-BC	BA	GSA	LFM ( <i>cadjMax=7</i> )
Zachary's	45.2	48	57.4	46.22	37.22	39.88	<b>0.10</b>

Tableau IV.2: LFM vs algorithmes d'Atay et al. [101]: temps d'exécution sur Zachary [123]

Dans ce tableau comparatif (tableau IV.2), nous constatons que notre algorithme est beaucoup plus rapide que les six algorithmes proposés par Atay et al. [101] malgré la configuration matérielle qu'ils ont utilisée (serveur avec 8 192 Mo de RAM, Intel (R) Xeon



(R) (12 CPU) E5-1650 3,20 GHz CPU et Windows 10 Pro 64 bits) est plus puissante que la nôtre (LapTop).

### **Conclusion**

Dans ce chapitre d'implémentation et d'expérimentation, nous avons appliqué notre algorithme sur le même exemple donné par Blondel et al afin d'implémenter son algorithme de Louvain et de le comparer avec notre algorithme proposé. De plus, nous avons appliqué notre algorithme sur un véritable réseau social d'amitiés entre les individus du club de karaté de Zachary en tant que petit réseau de référence non orienté, et également sur un autre réseau social, plus large et orienté, que nous avons construit à partir de la base de données de messagerie d'une banque américaine. De plus, nous avons discuté nos résultats en fonction de certains paramètres comme le nombre de nœuds, d'arêtes et donc la densité du réseau où nous avons pu exécuter notre algorithme sur un réseau de deux mille nœuds et plus de deux millions d'arêtes avec uniquement notre PC portable. Enfin, nous avons comparé notre algorithme LFM avec ceux proposés par Atay et al. [101] en fonction du temps d'exécution et nous avons constaté que le nôtre est beaucoup plus rapide qu'eux malgré la différence de performances matérielles entre nous.



## Chapitre 5: No-ELFM, Extension de LFM aux communautés non chevauchantes.

### Table des matières

<b>Introduction</b> .....	97
<b>V.1. Extension No_ELFM</b> .....	97
<b>A. LFM de base, rappel</b> .....	97
<b>B. LFM sans chevauchement</b> .....	98
1. Scénario 1: (P&P) .....	99
2. Scénario 2: (P&C).....	101
3. Scénario 3: (C&C) .....	103
<b>V.2. Vivacité des communautés dans No_ELFM</b> .....	105
<b>V.3. LFM léger</b> .....	107
<b>V.4. Mise en œuvre, expérimentation et discussion</b> .....	107
<b>Conclusion</b> .....	113

La détection des communautés dans les réseaux sociaux est une phase d'analyse très importante, mais comment trouver efficacement des communautés qui ne se chevauchent pas dans des réseaux sociaux réels reste un défi et un domaine de recherche d'actualité qui attire les chercheurs.

Dans ce chapitre, nous proposons l'extension No\_EFLM de notre modèle de base pour détecter la meilleure partition des communautés non chevauchantes dans les réseaux sociaux.

Nos expériences sur un vrai réseau social montrent que l'extension proposée peut définir avec précision des communautés qui ne se chevauchent pas, d'une part, et peut considérablement améliorer la qualité de la détection communautaire et obtenir une structure communautaire précise, d'autre part.

Dans ce chapitre nous fournissons le contexte nécessaire et certains travaux connexes. Donnons notre extension et ses différents scénarios. Deux autres nouvelles notions ont été données à savoir la vivacité des communautés et la version légère de No\_EFLM. Aussi, nous traitons la phase de l'implémentation et de mise en œuvre ainsi que de la discussion des différents résultats obtenus. Et finalement, nous concluons le chapitre.

## **Introduction**

Une communauté est un ensemble d'éléments (*abonnés ou utilisateurs d'un réseau social (SN), clients dans un site de vente en ligne, scientifiques dans un site publiant des articles et des livres ...etc.*) ayant des relations plus denses entre eux, mais moins de contacts avec le monde extérieur.

Dans ce chapitre, nous proposons une extension de notre modèle LFM pour prendre en charge la détection des communautés non chevauchantes dans les SN.

Cette extension se base sur la vie sociale des individus dans leurs sociétés. Premièrement, le LFM de base calcule l'ensemble des familles dans notre ensemble de données, tout en évitant le chevauchement entre elles (*en termes de parents et d'enfants*). On obtient alors un ensemble de "Out\_nodes" qui doivent joindre une et une seule famille parmi celles détectées dans la phase précédente. À cette fin, plusieurs paramètres et conditions peuvent jouer le rôle de décideurs pour rejoindre une famille particulière. Ces paramètres seront examinés plus en détail dans les sections suivantes.

Dans ce chapitre nous donnons notre extension et ses différents scénarios. Deux autres nouvelles notions ont été données à savoir la vivacité des communautés et la version légère de No\_EFLM. Aussi, nous traitons la phase de l'implémentation et de mise en œuvre ainsi que de la discussion des différents résultats obtenus. Et finalement, nous concluons le chapitre.

### **V.1. Extension No\_ELFM**

Dans cette section nous donnons un petit rappel du modèle de base LFM ainsi que son extension aux communautés non chevauchantes où nous proposons le nouveau modèle **No\_ELFM** (*ou Non-overlapping Extended LFM*) avec tous les différents scénarios.

#### **A. LFM de base, rappel**

Le principe de notre algorithme LFM consiste à rechercher deux à deux le nombre maximum de voisins communs entre les nœuds. Les couples (*parents*) ayant un nombre maximum (*CadjMax*) de voisins (*enfants*) en commun, forment les communautés initiales (*famille*). La deuxième étape consiste à permettre aux autres nœuds (*Out\_nodes*) de rejoindre ces communautés initiales.

Nous calculons la modularité  $Q$  de cette distribution, décrétons la valeur de *CadjMax* et re-exécutons l'algorithme (*la première génération de communautés et la jointure des "Out\_nodes"*) jusqu'à ce que *CadjMax* devienne nul.

La distribution communautaire correspondant à la meilleure valeur de  $Q$  est la plus qualifiée et la valeur correspondante de *CadjMax* doit être conservée.

Mathématiquement, nous calculons la valeur de ce paramètre *CadjMax* comme suit (formule V.1):

$$\mathbf{CadjMax} = \mathbf{Max}_{0 \leq k \leq K-1} (|\mathbf{com}_k|) - 2 \quad (\mathbf{V.1})$$

Où:

- $K$  est le nombre de communautés détectées à la première étape (*génération de familles*)
- $\mathbf{Com}_k$  est la  $k^{\text{ième}}$  communauté

- Nous mettons (-2) pour exclure les deux parents, chaque communauté n'a que deux parents.

### **B. LFM sans chevauchement**

Pour étendre le modèle LFM de base afin de prendre en charge le non-chevauchement entre les communautés, nous procédons comme suit:

Pour chacun des deux nœuds P1 et P2 ayant en commun "CadjMax" autres nœuds, une famille (*communauté*) est formée où P1 et P2 sont ses parents et les nœuds communs sont leurs enfants. À cette fin, et dans ce chapitre, P1 est considéré comme le conjoint de P2, et P2 est également le conjoint de P1.

Ainsi, et pour éviter le chevauchement des communautés, et lors de la génération de la distribution communautaire initiale (*avant l'étape de jointure des "Out\_nodes" aux différentes communautés initiales*), nous avons les scénarios possibles suivants:

**Scénario 1:** Un nœud se trouve comme un parent de deux familles (**P&P**): il ne doit choisir qu'un seul conjoint;

**Scénario 2:** Un parent se retrouve enfant dans une autre famille (**P&C**): il doit choisir son rôle, parent (*avec la première famille*) ou enfant (*avec la seconde*);

**Scénario 3:** Un nœud se trouve également un enfant dans une autre famille (**C&C**), ce phénomène n'existe pas dans la nature (*dans le domaine de la vie sociale des individus mais dans autres domaines comme les protéines et leurs rôles, les articles scientifiques et leurs domaines de recherches, ..., ce scénario peut être rencontré*) et vu que notre modèle de base pourra le générer on doit donc trouver un moyen comment le prendre en charge. Ces enfants doivent en choisir une parmi ces familles détectées dans la phase initiales.

### **Remarques**

1- Dans le cas où on élimine le chevauchement, la dernière phase de l'algorithme de base LFM, qui consiste à fusionner (*sous certaines conditions*) les communautés qui se chevauchent, sera supprimée car il n'y aura pas de chevauchement entre les communautés trouvées.

2- On peut se retrouver dans des cas plus complexes comme par exemple un nœud est parent dans deux familles mais aussi enfant dans une autre (**P&P&C**), dans ce cas, il suffit de décomposer le problème en «Atomic» sous-problèmes comme suit:

Le nœud concerné doit choisir dans quelle famille doit être parent (*donc choisir son conjoint et, de ce fait, "se divorcer" de l'autre et ainsi "dissocier" la deuxième communauté et ses membres seront considérés comme "Out\_nodes"*). Donc, nous avons le conflit entre le rôle parent et enfant et nous revenons au cas atomique (**P&C**), c'est-à-dire :  $P\&P\&C=(P\&P)\&C=(P\&C)$ .

Comme nous pouvons commencer par (**P&C**), où le nœud concerné doit choisir le rôle à jouer, parent dans une telle famille ou enfant dans l'autre, nous nous retrouvons dans l'un des deux cas atomiques (**P&P**) ou (**P&C**) comme suit:

$$P\&P\&C= P\&(P\&C)= \begin{cases} (P\&C) \text{ si le nœud concerné a choisi d'être enfant} \\ (P\&P) \text{ si le nœud concerné a choisi d'être parent} \end{cases}$$

Dans ce qui suit, nous détaillons les solutions envisagées pour résoudre ces trois scénarios.

### 1. Scénario 1: (P&P)

Si le parent P1 a deux conjoints ou plus, il doit en choisir un et donc dissocier les autres avec leurs enfants (*qui seront considérés comme des "Out\_nodes" qui rejoindront les différentes familles*).

Le paramètre de choix du conjoint, que ce parent doit choisir, est très important pour la distribution communautaire initiale et celle terminale et donc la qualité du résultat (modularité).

Dans cette section, nous proposons un modèle générique où l'utilisateur final peut le configurer en fonction de son domaine d'application et ses besoins. À cette fin, nous détaillons ces paramètres que le parent doit respecter pour choisir son seul conjoint. En parlant, donc, de **paramètres de préférence OOS (un seul conjoint ou Only One Spouse)**.

1- Selon la modularité: Nous pouvons proposer que le nœud concerné ait rejoint tous les conjoints un par un, et à chaque fois nous calculons la modularité puis prenons le conjoint qui augmente autant que possible cette fonction de qualité. Cette solution est assez lourde par rapport au temps d'exécution et la complexité temporelle, à cet effet nous pouvons envisager d'autres solutions plus applicables et plus faciles à implémenter.

2- Selon l'existence de la relation (*amitié, collaboration ...etc*) entre ce parent et les différents conjoints: Le parent peut choisir le conjoint avec lequel il a une relation sociale, comme illustré à la figure V.1.

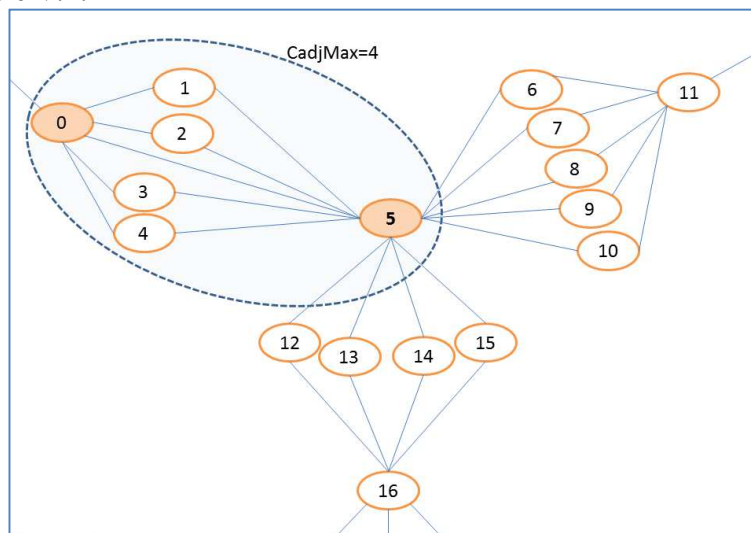
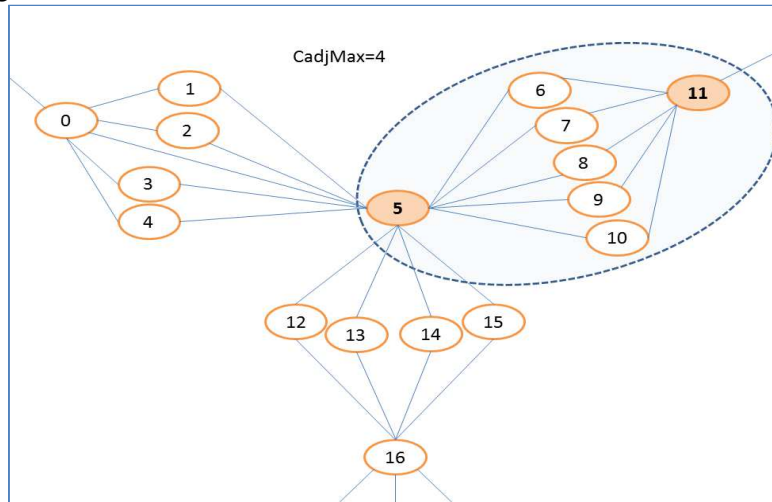


Figure.V.1: OOS, existence de relation sociale entre le parent et le conjoint choisis.

Comme nous pouvons le voir sur la figure V.1, le parent (5) choisira le conjoint (0) car c'est le seul nœud ayant une relation sociale avec lui. Maintenant, si ce parent aura deux conjoints (*ou plus*) avec lesquels il avait des relations sociales (*ou si l'utilisateur de notre modèle générique veut choisir un autre paramètre de choix*), nous exécutons le paramètre suivant: nombre d'enfants.

3- Selon le nombre d'enfants: notre approche consiste tout d'abord à construire toutes les familles, c'est-à-dire les parents avec le maximum (CadjMax) d'enfants, puis on décrémente cette valeur de CadjMax jusqu'à zéro (*inclut, car deux nœuds n'ayant pas d'enfants en commun peuvent être des parents*). Ainsi, par exemple, pendant la phase où CadjMax=5 enfants en commun, nous pouvons trouver un couple (*de parents*) ayant 6 (*ou plus*), car s'ils ont 6 enfants, ils ont également 5 et ils peuvent, dans les deux cas, créer leur propre communauté (*famille*), donc tous les couples avec CadjMax (*ou plus*) seront considérés comme parents.

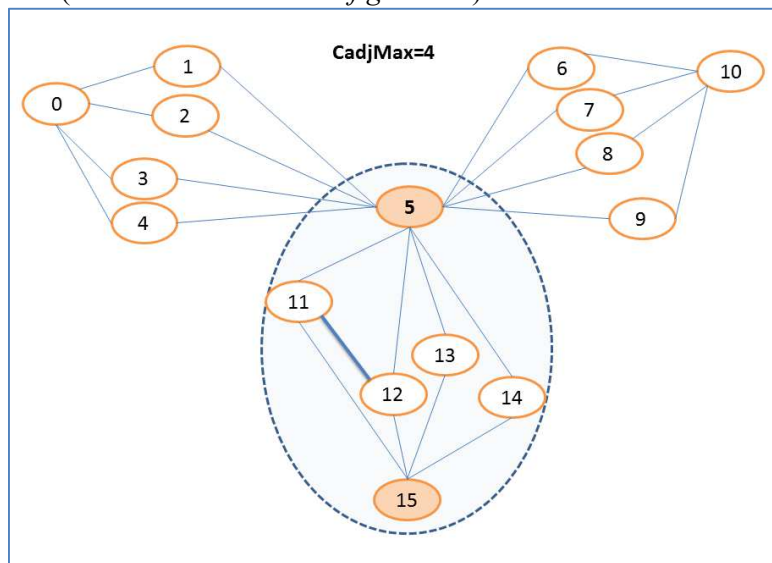
Dans ce cas, le nœud concerné doit choisir le conjoint ayant le nombre maximum d'enfants avec lui. Comme ça, c'est la plus grande famille qui doit maîtriser les autres comme le montre la figure V.2:



**Figure.V.2: OOS, selon le nombre d'enfants.**

Dans la figure V.2, nous sommes dans la phase (*de décrémentation*)  $CadjMax=4$ , donc tout couple avec 4 (*ou plus*) enfants en commun peut construire une communauté, et ainsi le parent (5) peut construire trois communautés "chevauchantes" avec les trois conjoints (0, 11 et 16). En appliquant ce nouveau paramètre "nombre maximum d'enfants", le parent (5) doit rejoindre le nœud (11) car il a cinq enfants en commun avec lui.

Dans le cas où on retrouve deux conjoints ayant le même nombre d'enfants en commun avec le parent en question, ce dernier choisira la communauté la plus dense en termes de relations (*comme illustré dans la figure V.3*).

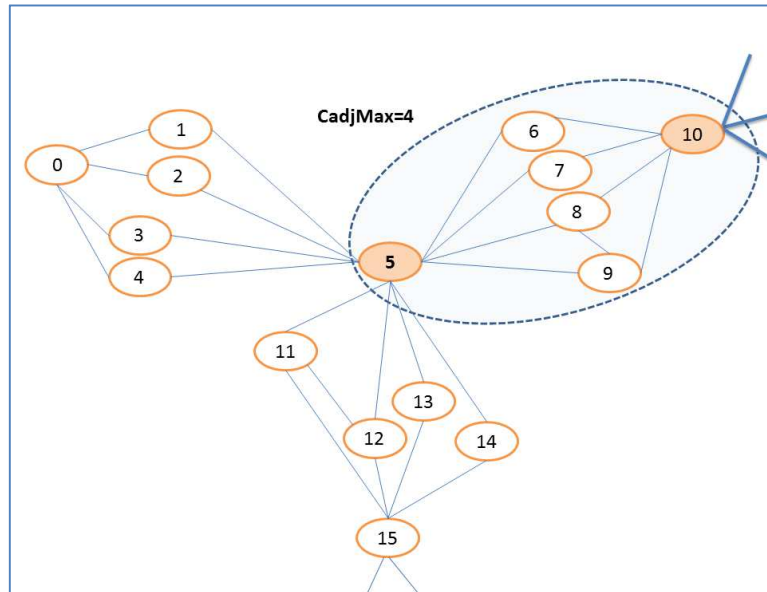


**Figure.V.3: OOS, Communauté la plus dense**

Dans la figure V.3, le parent (5) choisira le nœud (15) comme conjoint car ils forment la communauté la plus dense.

En cas d'égalité (entre deux conjoints ou plus), le parent en question fera le choix suivant (*point 4*).

4- Selon le nombre de relations de chaque conjoint: Le nœud concerné doit choisir le conjoint ayant le maximum de relations sociales avec le reste du monde (*réseau*). Comme ça, c'est la femme (*conjoint*) la plus populaire qui doit maîtriser les autres comme le montre la figure V.4.



**Figure.V.4: OOS, popularité des conjoints.**

Dans la figure V.4, le nœud (5) choisira entre les deux nœuds (10) et (15) car ils forment (*avec lui*) les deux communautés les plus denses en termes de relations sociales, il prendra le nœud (10) car il a plus de relations avec le monde extérieur.

## **2. Scénario 2: (P&C)**

Le nœud concerné doit choisir l'un des deux rôles, un parent dans une communauté ou un fils dans une autre. Par défaut, notre modèle permet à ce nœud de choisir d'être un parent juste pour créer une nouvelle famille (*communauté*). La figure V.5 illustre ce scénario.

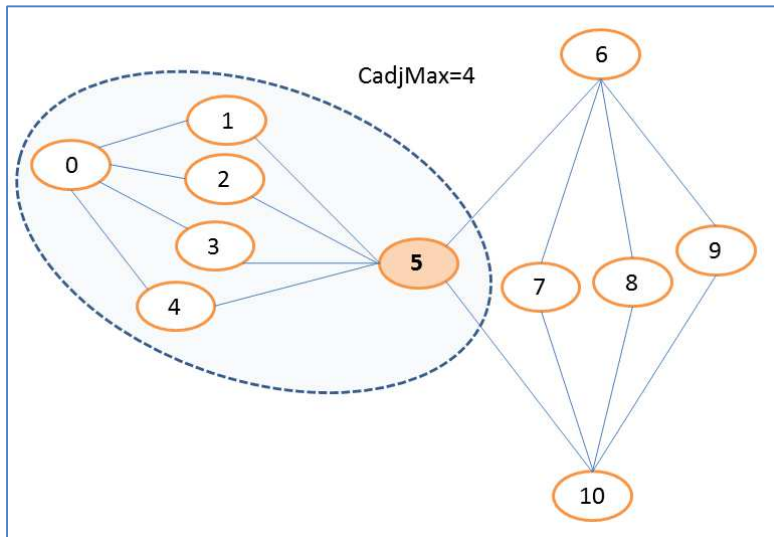


Figure.V.5: Le scénario (P&C), le choix par défaut: parent.

Dans la figure V.5, le nœud (5) a préféré rejoindre la communauté de gauche car il est par défaut un parent de famille contrairement à l'autre (*de droite*) où il est enfant.

Sinon, le nœud concerné pourra choisir le cas enfant (*ou parent*) dont d'autres paramètres de préférence qu'on pourrait imaginer comme les suivants:

1- Modularité: Nous calculons sa valeur pour tous cas et choisissons celui qui l'augmente. Cette solution reste toujours coûteuse en termes de complexité et de mise en œuvre.

2- Le nombre de ses relations dans chaque communauté: le nœud concerné choisira la communauté où il a plus de relations sociales. Ainsi, il se retrouvera plus à l'aise (*dans sa migration*) car il connaît plus de monde, comme illustré dans la figure V.6.

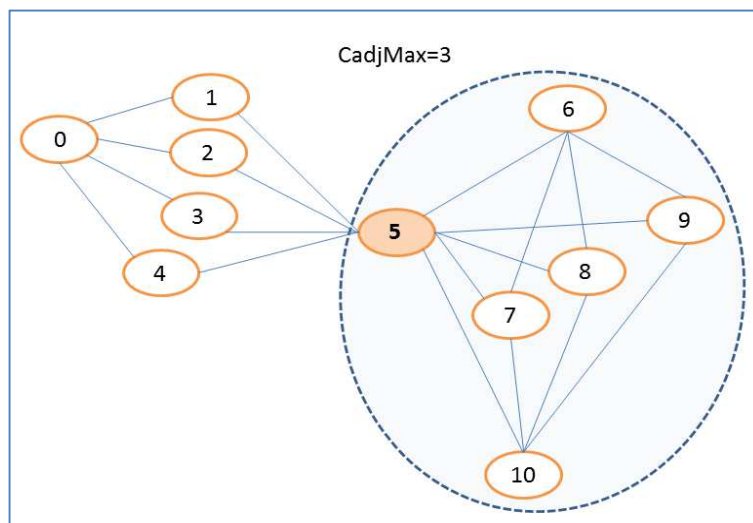


Figure.V.6: Le scénario (P&C), relations dans chaque communauté.



Dans la figure V.6, le nœud (5) a préféré de rejoindre la communauté de droite (comme un enfant des deux parents (6) et (10)) car il a cinq relations alors qu'il n'en a que quatre dans l'autre.

### 3. Scénario 3: (C&C)

Dans ce cas, le nœud concerné est un enfant dans deux ou plusieurs familles mais il doit choisir une et une seule afin d'en éviter le chevauchement, laquelle il doit rejoindre donc? On peut choisir la modularité comme paramètre de préférence qu'il faut augmenter. Mais par défaut et pour que notre modèle soit flexible, le nœud en question rejoindra la communauté (*famille*) la plus dense, comme illustré à la figure V.7.

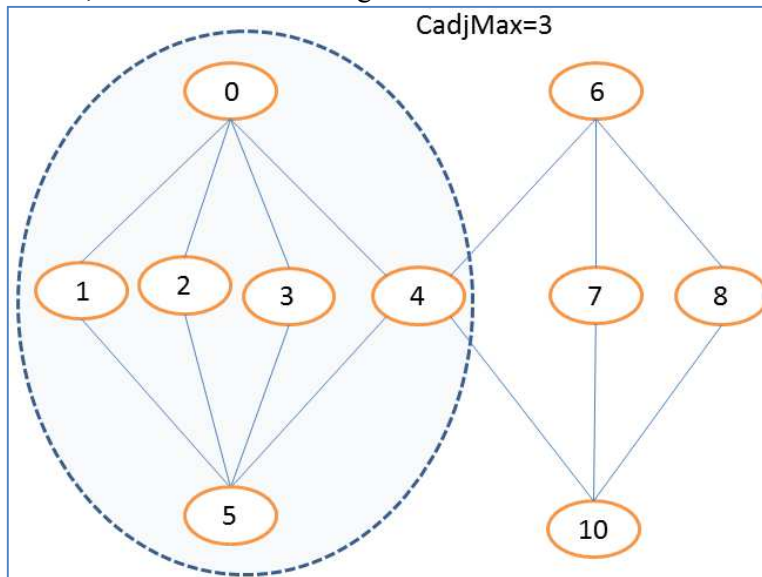


Figure.V.7: Le scénario (C&C), la communauté la plus dense (en termes de nœuds).

Dans la figure V.7, le nœud (4) choisira la communauté de gauche car elle est la plus dense.

En termes de conception, nous utilisons la matrice Com [K x N], où K est le nombre de communautés et N le nombre de nœuds, comme suit:

$$\text{Com}[k][j] = \begin{cases} 2 & \text{Si le nœud } j \text{ est un parent de la communauté } k \\ 1 & \text{Si le nœud } j \text{ est un fils de la communauté } k \\ 0 & \text{Sinon} \end{cases}$$

Afin d'en éliminer le cas de chevauchement, nous donnons, uniquement et à titre d'exemple d'implémentation pour ce scénario, la portion de code source suivante.



```

for (k=0; k<K; k++){ // all communities
  for (j=0; j<n; j++){ // all nodes
    if (com[k][j]==1) {
      for (i=k+1; i<K; i++) {
        if (com[i][j] ==1) { //the two communities k and i share node j, the larger one will take it
          if (nb_children(com[k])>= nb_children(com[i])) {com[i][j]=0;}
          else {com[k][j]=0;break;}
        }
      }
    }
  }
}

```

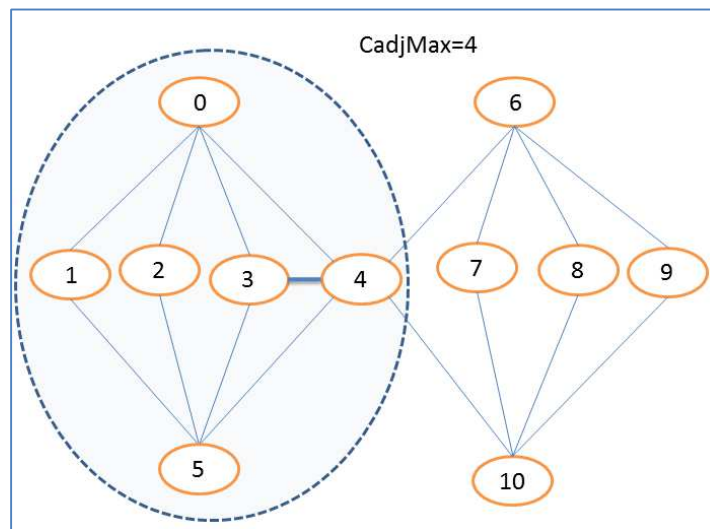


Figure.V.8: Le scénario (C&C), la communauté la plus dense (en termes de relations).

Dans ce cas, l'autre communauté doit décrémenter son nombre d'enfants en commun "Nbr\_children" et donc elle peut être considérée comme une autre communauté si  $Nbr\_children \geq CadjMax$ , sinon elle sera dissociée et tous ces membres seront considérés comme Out\_node comme illustré dans les deux figures V.9 et V.10.

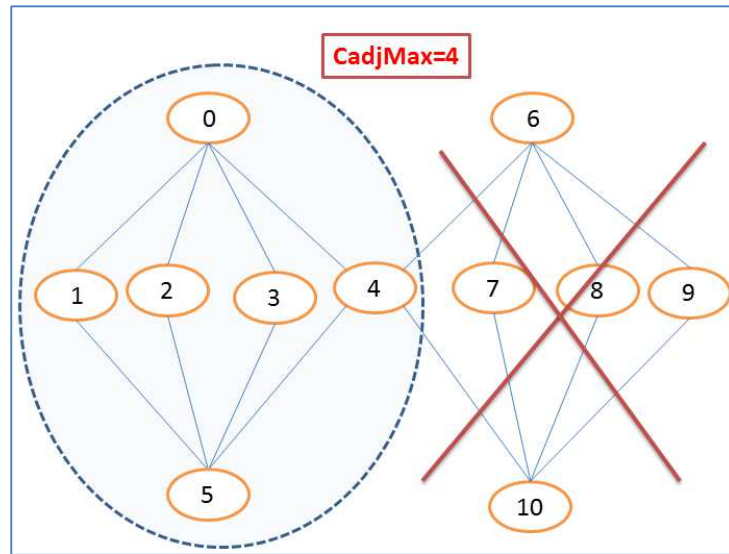


Figure.V.9: Scénario (C&C), destruction de la communauté ( $Nbr\_children < CadjMax$ ).

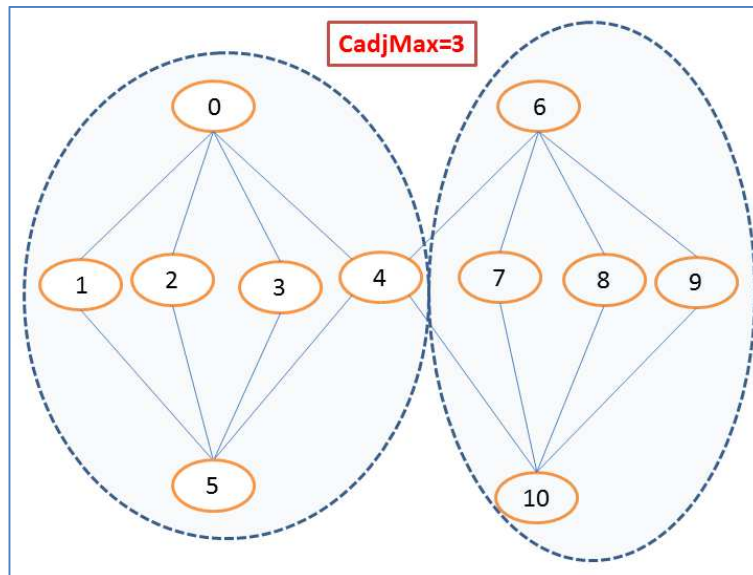


Figure.V.10: Scénario (C&C), reconstruction communautaire ( $Nbr\_children \geq CadjMax$ ).

Dans la figure V.9, nous avons  $CadjMax=4$  et le nombre d'enfants de la communauté de droite est  $Nbr\_children=3$ , donc cette dernière doit être dissociée et tous ses membres seront considérés comme *Out\_nodes*. Contrairement à la figure V.10 où  $CadjMax=3$  et donc ladite communauté restera et y joindra les nœuds externes qu'elle peut.

## V.2. Vivacité des communautés dans No\_ELFM

Jusqu'à présent, notre modèle de base ne s'intéresse qu'à l'existence de la relation sociale entre les membres de notre réseau. Quand il est vraiment insuffisant de décider si notre ami qui ne nous a contacté qu'une seule fois fera partie de notre communauté. Oui c'est vrai, il a une relation sociale avec nous mais il ne nous contacte que quelques fois. À cette

fin, nous étendrons notre modèle LFM pour soutenir une nouvelle notion que nous appelons "la vivacité des communautés" qu'il détecte.

Pour cela, on parle de "communautés pondérées" où il faut prendre en compte le nombre de fois où les nœuds communiquent et pas seulement l'existence de la relation de communication (*amitié ou autre*).

Dans ce cas, nous soutiendrons la notion de "Densité" des différentes interactions entre les membres du réseau social, un membre  $m_1$  ayant  $n_1$  interactions avec la communauté C est plus susceptible d'en faire partie qu'un autre membre  $m_2$  ayant  $n_2$  interactions avec C où  $n_1$  est supérieur à  $n_2$ .

De plus, le calcul de la valeur du paramètre  $CadjMax$  de la formule (V.1) devient comme suit (*formule V.2*):

$$CadjMax = \frac{1}{2} \text{Max}_{0 \leq k \leq K-1} \left( \sum_{i \in \text{com}_k / \text{Parents}(\text{com}_k)} [w(P_k^1, i) + w(P_k^2, i)] \right) \quad (V.2)$$

Où :

- K : Nombre de communautés détectées à la première étape (génération de familles),
- $\text{com}_k$  : La  $k^{\text{ième}}$  communauté
- Parents ( $\text{Com}_k$ ): Les deux parents de la  $k^{\text{ième}}$  communauté.
- $P_k^j$  : Le  $j^{\text{ième}}$  (premier ou deuxième) parent de la  $k^{\text{ième}}$  communauté.
- $w(P_k^j, i)$  : Poids de l'arête ( $P_k^j, i$ ).
- La division sur 2 est appliquée afin d'en forcer le modèle No\_ELFM à donner les mêmes résultats (*mêmes distributions communautaires*) dans le cas d'unicité des contacts entre les nœuds (*poids de toutes les arrêtes est égal à 1*). Dans ce cas, les deux formules (V.1) et (V.2) donnent la même valeur pour  $CadjMax$ .

Dans la figure V.11, pour une valeur de  $CadjMax=30$ , nous avons les nœuds (0) et (5) ont quatre enfants mais ils ne forment aucune communauté (*car la valeur de  $CadjMax$ , en appliquant la formule V.2, nous donne 17 qui est moins de 30*), contrairement aux nœuds (6) et (10) qui sont parents de la communauté {4,6,8,9,10}, simplement parce que le nombre (*somme*) de fois que les membres de cette communauté ont été en contact avec les deux parents (6,10) est beaucoup plus grand ou égal à la valeur de  $CadjMax=30$  et donc elle est plus importante que celle de {0,1,2,3,4,5} (*malgré que le nombre d'enfants est moins important  $3 < 4$* ). Ainsi, pour cette version d'extension, nous comptons non seulement le nombre de nœuds en commun mais plus tôt le nombre de fois que ces nœuds (*communs*) ont contacté leurs parents.

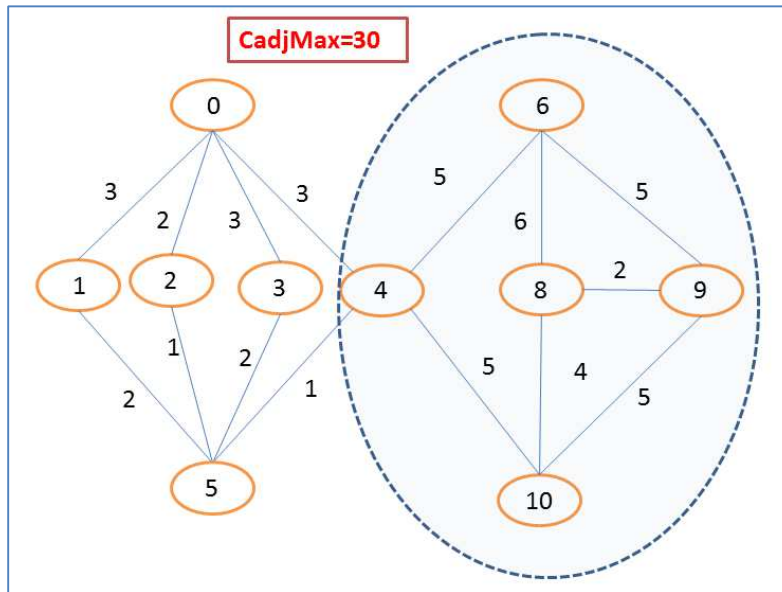


Figure.V.11: Vivacité des communautés dans LFM.

### V.3. LFM léger

Une remarque très importante concernant l'implémentation de cette extension No\_ELFM est que le modèle LFM de base décrémente la variable CadjMax jusqu'au zéro. Ainsi, pour chaque nouvelle valeur CadjMax, nous calculons la qualité  $Q$  (*modularité*) puis la distribution communautaire finale à retenir sera celle correspondant à la meilleure valeur de  $Q$ . Le problème qui se pose pour No\_ELFM est que les valeurs de CadjMax seront toujours très importantes (*parce que nous prenons le nombre d'interactions et pas seulement le nombre de nœuds en commun*). Par exemple, si nous avons 100 interactions entre deux parents et leurs enfants, nous devons produire 100 distributions communautaires (*et 100 calculs de  $Q$* ) et choisir la meilleure, mais 100 interactions prendront beaucoup de temps d'exécution.

La solution que nous proposons ici (*et que nous pouvons même appliquer au modèle de base*) est que nous devons produire une matrice  $N \times N$ , où l'intersection entre la ligne  $i$  et la colonne  $j$  donne le nombre d'interactions entre les deux parents  $i$  et  $j$  et leurs enfants, nous faisons un ordre décroissant de ces valeurs et nous les prenons pour la variable CadjMax. Par exemple, si nous trouvons la valeur maximale est 100, puis 90 puis 85..., nous ne commencerons pas par CadjMax=100 puis la diminuerons jusqu'au zéro mais juste CadjMax=100 puis directement CadjMax=90 puis 85.... À ce niveau, nous appellerons cette version **LFM léger** (*ou LFM\_Light*).

### V.4. Mise en œuvre, expérimentation et discussion

Dans cette section, nous effectuons nos différentes extensions sur un vrai jeu de données. Il concerne les interactions entre les abonnés du Facebook et contient une liste de tous les liens entre les utilisateurs du réseau Facebook de New Orleans. Nous utilisons le jeu de données "facebook-links" [129] disponible sur <http://socialnetworks.mpi-sws.mpg.de/data/facebook-links.txt.gz>.

Tous les liens sont traités comme orientés, même s'ils ne le sont pas sur Facebook vu que la relation d'amitié est symétrique. Son format est ASCII compressé. Chaque ligne

contient deux identifiants d'utilisateurs anonymes, ce qui signifie que le premier qui a demandé l'amitié du deuxième. Enfin, la troisième colonne est un horodatage UNIX avec l'heure d'établissement de la liaison (*si elle peut être déterminée, sinon c'est 'N'*).

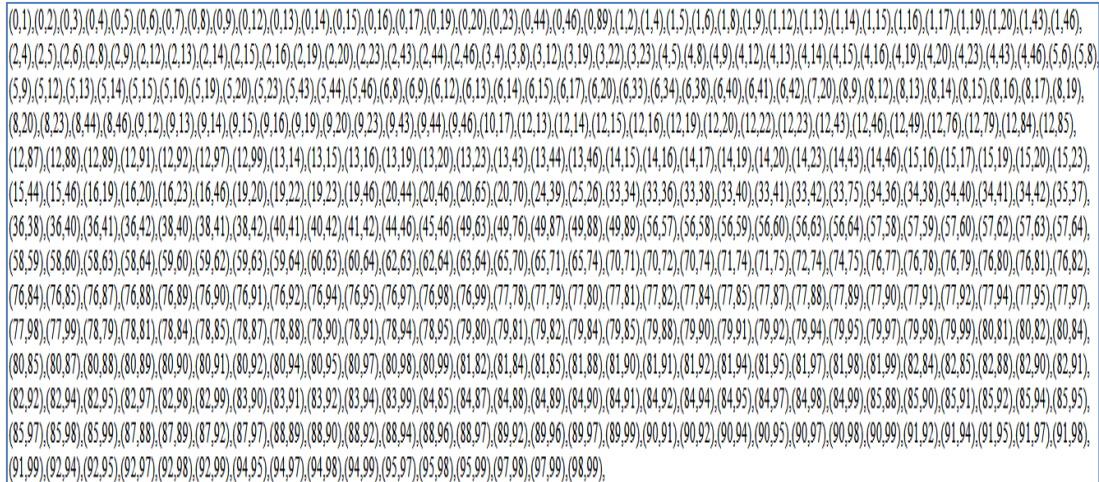
Dans ce chapitre, nous avons donné trois extensions pour le modèle de base LFM:

- 1- Communautés non chevauchantes (*No\_ELFM*) et les différents scénarios.
- 2- Vivacité des communautés
- 3- LFM-Light

En ce qui concerne cette section de mise en œuvre et d'expérimentation, nous nous concentrerons uniquement sur la première extension des communautés non chevauchantes. Aussi, et pour simplifier les calculs, nous avons pris, comme échantillon d'étude, les 100 premiers utilisateurs (identifiants) et toutes les relations entre eux.

Dans notre implémentation, nous avons trouvé  $CadjMax=13$ , où les parents étaient les deux nœuds avec identifiants (76) et (79), tandis que leurs enfants sont les suivants: {77,80,81,82,85,90,91,92, 94, 95,97,98,99}, c'est-à-dire que tous ces nœuds ont des relations avec les deux parents (76) et (79) et donc nous avons comme distribution communautaire initiale cette communauté ( $famille = \{parents\} \cup \{enfants\}$ ). Donc, nous avons ces 15 nœuds comme "In\_nodes" et aussi 85 "Out\_nodes" qui rejoindront cette communauté au fur et à mesure et nous avons eu comme distribution communautaire finale tous les nœuds du réseau (parce que le réseau est un composant connexe, tous les nœuds sont connectés directement ou indirectement).

La modularité trouvée est égale à la valeur de  $Q=5,9799773157978E-17$ , c'est-à-dire presque nulle et donc la distribution communautaire est mauvaise. Notre algorithme LFM de base consiste à décrémenter  $CadjMax$  et à déduire à chaque fois la distribution communautaire ainsi que la valeur correspondante de  $Q$ . Pour cela, nous avons fait 13 itérations (décrémentations) et la valeur maximale de  $Q$  trouvée est  $Q=0,70936414930556$  qui a donné la valeur correspondante de  $CadjMax=2$ . Une information importante sur ce résultat est que nous avons trouvé 417 (!) couples (*parents*) ayant 2 (*CadjMax*) enfants en commun comme par exemple (0,1), (0,2), (0,3), (0,4), (0,5),..., (60,63), (60,64), .. (95,97), (95,98), (95,99),... (*Voir la figure V.12*).



**Figure.V.12. 417 couples (parents) pour CadjMax=2 (meilleure valeur de Q)!!**

Comme critique de ce résultat, et donc de notre modèle de base, nous avons trouvé 417 communautés et la logique implique que nous devons trouver au maximum 50 communautés non chevauchantes (chacune ne contient que ses deux parents). Nous avons obtenu ce résultat car nous avons exécuté le modèle LFM de base sans prendre en compte le



non-chevauchement et nous pouvons voir que le nœud 0 est un parent dans 21 communautés avec les conjoints suivants  $\{(0,1), (0,2), (0,3), (0,4), (0,5), (0,6), (0,7), (0,8), (0,9), (0,12), (0,13), (0,14), (0,15), (0,16), (0,17), (0,19), (0,20), (0,23), (0,44), (0,46), (0,89)\}$ . Les mêmes cas sont produits pour plusieurs nœuds tels que le nœud 60  $\{(60.63), (60.64)\}$ , le nœud 95  $\{(95,97), (95,98), (95,99)\}$  ....etc.

Donc, nous sommes dans le scénario (P&P) où chaque parent doit choisir un et un seul conjoint. Lequel parmi tous ces 21 conjoints le nœud (0) doit alors choisir?

Les solutions à ce problème ont été bien expliquées dans la section "Only\_One\_Spouse", et nous avons choisi le point 3 comme exemple de mise en œuvre comme illustré sur la figure V.2 (en fonction du nombre d'enfants).

Dans notre exemple, et comme le montre la figure V.13, le nœud (0) a neuf enfants avec le nœud 1  $\{2,5,6,8,9,13,14,15,20\}$ , mais il a également 10 enfants avec le nœud (2)  $\{1, 4,5,8,9,12,13,14,15,16\}$  et donc il choisira (2) comme conjoint. De plus, avec le nœud (3), il n'a que (2)  $\{19,23\}$ , donc le nœud (0) conserve (2) comme conjoint. Nous exécutons cette itération avec les 21 nœuds et nous constatons à la fin qu'il a choisi le nœud (8) comme conjoint qui a 11 enfants en commun avec lui, comme le montre la figure V.13.

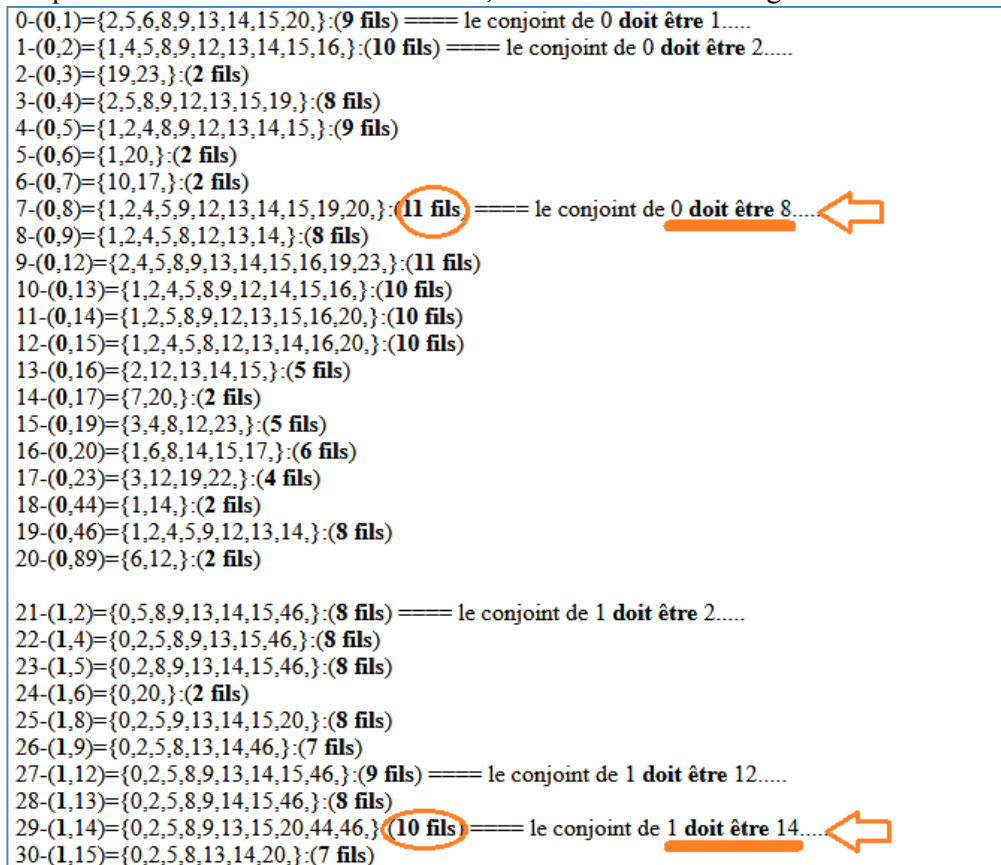


Figure.V.13. Le nœud (0) a finalement choisi le nœud (8), (1) le nœud (14), etc.

À la fin de ce processus d'élimination de chevauchement, nous trouvons la distribution communautaire initiale avec seulement les 25 communautés suivantes:  $\{(0,8); (1,14); (2,13); (4,12); (10,17); (15,46); (19,23); (20,65); (24,39); (25,26); (33,38); (34,41); (35,37); (56,59); (57,63); (58,60); (71,74); (72,74); (76,79); (81,82); (85,92); (87,89); (90,91); (94,97); (98,99)\}$  où nous pouvons voir qu'il n'y a pas de chevauchement de type P&P comme illustré dans la figure V.14.

Après suppression de chevauchement du scénario P&P, les 25 parents des familles les plus larges qui restent sont comme suit:

```
(0,8):{ 1,2,4,5,9,12,13,14,15,19,20,}:11 children
(1,14):{ 0,2,5,8,9,13,15,20,44,46,}:10 children
(2,13):{ 0,1,4,5,8,9,12,14,15,16,46,}:11 children
(4,12):{ 0,2,5,8,9,13,15,19,46,}:9 children
(10,17):{ 0,7,}:2 children
(15,46):{ 1,2,4,5,12,13,14,}:7 children
(19,23):{ 0,3,12,}:3 children
(20,65):{ 71,74,75,}:3 children
(24,39):{ 25,26,}:2 children
(25,26):{ 24,39,}:2 children
(33,38):{ 34,36,40,41,42,}:5 children
(34,41):{ 6,33,38,40,42,}:5 children
(35,37):{ 33,75,}:2 children
(56,59):{ 57,58,60,63,64,}:5 children
(57,63):{ 56,58,59,60,}:4 children
(58,60):{ 56,57,59,63,}:4 children
(71,74):{ 20,65,70,}:3 children
(72,74):{ 65,71,}:2 children
(76,79):{ 77,80,81,82,85,90,91,92,94,95,97,98,99,}:13 children
(81,82):{ 76,79,80,85,91,92,}:6 children
(85,92):{ 76,79,80,81,82,94,97,}:7 children
(87,89):{ 12,49,76,88,}:4 children
(90,91):{ 76,79,82,84,99,}:5 children
(94,97):{ 76,79,85,92,}:4 children
(98,99):{ 76,77,79,81,91,}:5 children
```

Figure.V.14. Élimination des chevauchements de type (P&P).

La phase suivante consiste à éliminer les chevauchements de type (P&C) dans notre résultat de la figure V.14. Par exemple, le nœud 0 est un parent dans la première communauté et enfant dans cinq autres. À cet effet, nous mettons en œuvre la solution par défaut, expliquée dans la section "Scénario2" où le nœud concerné choisira d'être parent plutôt qu'enfant. Cela implique que nous garderons la communauté ayant ce nœud comme parent et donc deux cas prévus pour les autres communautés. Dans ces dernières, en supprimant le nœud concerné, nous diminuons le nombre de fils en commun entre les parents "Nbr\_children":

- 1- Chaque communauté ayant  $Nbr\_children \geq CadjMax$  restera une communauté,
- 2- Sinon, elle sera dissociée et tous ces membres seront considérés comme 'Out\_nodes'.

La mise en œuvre et l'exécution de cette phase nous ont donné 15 communautés comme illustré dans la figure V.15.

La liste des 15 communautés (après suppression de chevauchement de type P&P puis P&C) est comme suit:

```
1-(0,8):{ 5,9,15,19,20,}:5 children
2-(1,14):{ 5,9,15,20,44,46,}:6 children
3-(2,13):{ 5,9,15,16,46,}:5 children
4-(4,12):{ 5,9,15,19,46,}:5 children
5-(25,26):{ 24,39,}:2 children
6-(33,38):{ 36,40,42,}:3 children
7-(34,41):{ 6,40,42,}:3 children
8-(57,63):{ 56,59,}:2 children
9-(58,60):{ 56,59,}:2 children
10-(71,74):{ 20,65,70,}:3 children
11-(76,79):{ 77,80,81,82,94,95,97,}:7 children
12-(85,92):{ 80,81,82,94,97,}:5 children
13-(87,89):{ 49,88,}:2 children
14-(90,91):{ 82,84,}:2 children
15-(98,99):{ 77,81,}:2 children
```

Figure.V.15 : Élimination des chevauchements de type (P&C).

L'étape suivante consiste à éliminer les chevauchements de type (C&C). Par exemple, les deux premières communautés partagent quatre enfants (5), (9), (15) et (20). Dans ce cas, chaque nœud (enfant) partagé entre deux (ou plusieurs) familles choisira (comme expliqué dans la section "Scénario C&C") la solution par défaut où il rejoindra la plus grande communauté (famille). Dans cet exemple, ces quatre nœuds rejoindront la deuxième communauté (1,14) et comme la première communauté (0,8) n'aura que le nœud (19) et donc  $Nbr\_children=1 < CadjMax=2$  elle sera dissociée (et donc les parents (0) et (8) deviendra *Out\_nodes*, le nœud 19 restera un *In\_node* car c'est aussi un enfant de la 4ème communauté de parents (4,12)) et ainsi de suite pour les autres.

La matrice "Com" définie dans la section «Scénario C&C» et avant l'élimination des chevauchements de type (C&C) est comme suivante (Figure V.16).

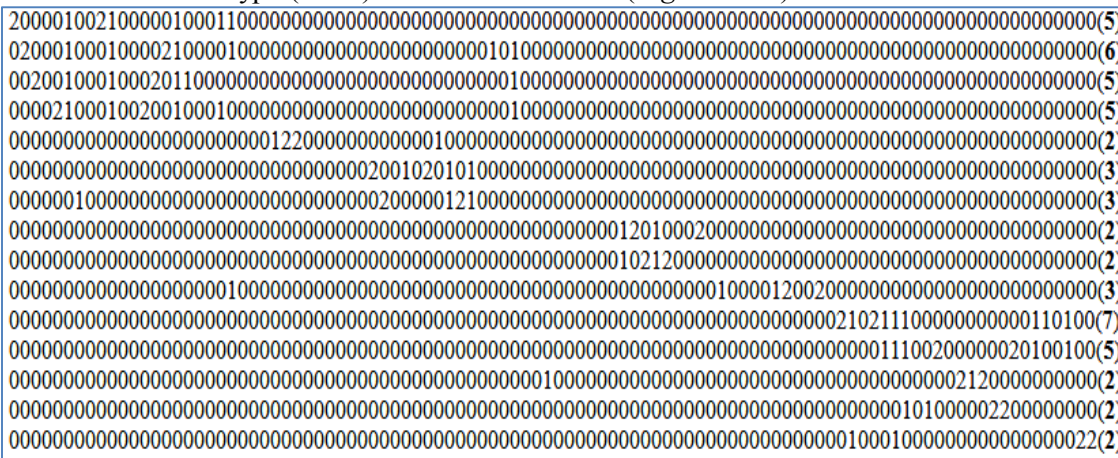


Figure.V.16. Population avant élimination des chevauchements de type (C&C).

Ainsi, chaque ligne étant une communauté et la dernière colonne (entre parenthèses) donne le nombre d'enfants de chacune.

Après avoir exécuté la partie du code donnée dans la section du scénario (C&C), la matrice «Com» est devenue comme illustré dans la figure V.17.

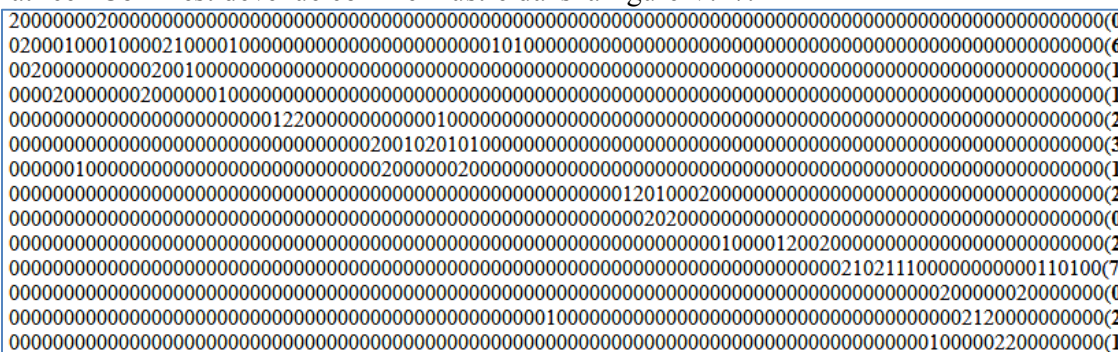
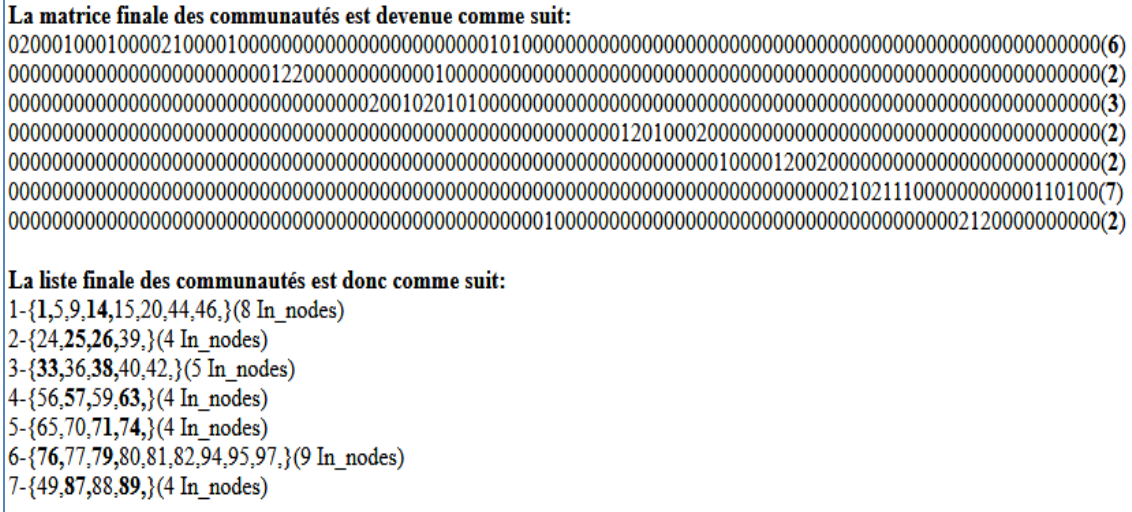


Figure.V.17. Population après élimination des chevauchements de type (C&C).

On voit qu'aucune colonne ne contient plus d'un nombre positif (1 ou 2), et donc pas de chevauchement entre les communautés. Il reste maintenant à dissocier les communautés ayant  $nbr\_children < CadjMax=2$ .

Après avoir effectué cette phase, nous trouvons les 7 communautés suivantes, comme illustré dans la figure V.18.



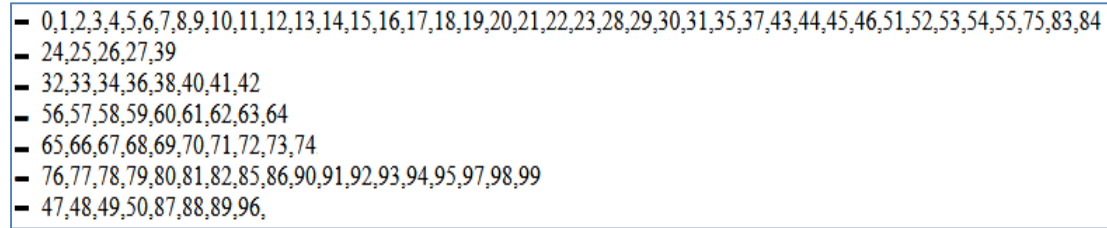


**Figure.V.18 : Destruction des communautés avec nbr\_children < CadjMax=2.**

Dans la liste des communautés de la figure V.18, les parents (*en gras*) sont fusionnés avec les enfants car la prochaine étape consiste à rejoindre les "Out\_nodes" et nous n'aurons pas besoin de cette notion de parent/enfants.

Pour la prochaine phase de jointure des nœuds "Out", nous avons 38 "In\_nodes" et 62 "Out\_nodes". Le processus est très simple, chaque "Out\_node" doit calculer ses poids (*nombre de relations*) aux sept communautés et celle avec la "force de gravité" la plus importante le rejoindra parmi ses "In\_nodes" et l'état de ce nœud changera de "Out\_node" à "In\_node". Pour tout "Out\_node", et en cas d'égalité de cette "force de gravité" entre deux ou plusieurs communautés, et juste pour simplifier notre implémentation, il choisira la première communauté parmi celles qui l'ont demandé (*pour assurer la décidabilité puis la stabilité de notre algorithme qui donnera toujours le mêmes résultat*). Il est préférable de se référer au chapitre III pour avoir plus d'options et paramètres de préférence concernant cette phase de jointure (*migration*) des nœuds (*individus*) vers les communautés (*populations*).

Après avoir effectué cette phase, nous avons trouvé les sept communautés suivantes (*Figure V.19*) où chaque ligne vaut une communauté.



**Figure.V.19 : Liste des communautés après avoir rejoint les "Out\_nodes"**

Et voici une illustration graphique des sept communautés trouvées (*Figure V.20*).

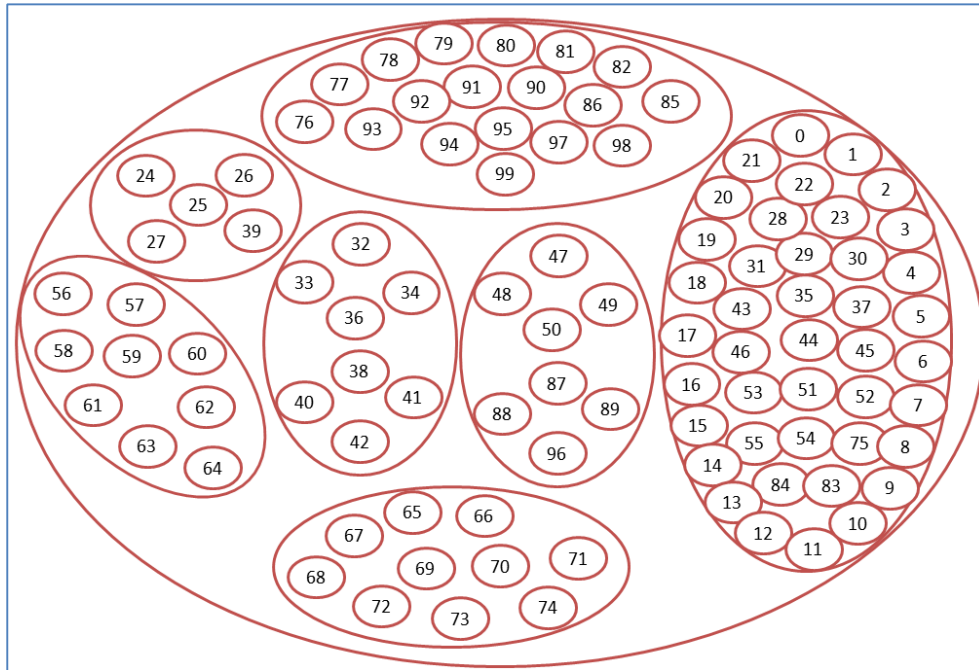


Figure.V.20: Illustration graphique des sept communautés trouvées.

La modularité de cette distribution communautaire est  $Q=0,57710624035494$ , elle est inférieure à celle du modèle de base mais elle est plus logique car le nombre de communautés est inférieur au nombre de nœuds (7 et non 417) et cette distribution communautaire est également sans aucun chevauchement.

## Conclusion

Les communautés détectées par les algorithmes peuvent ou non se chevaucher, mais il est parfois judicieux d'avoir des nœuds appartenant à plusieurs communautés (*comme dans le domaine religieux où chaque individu ne doit avoir qu'une seule religion, un autre exemple concernant la langue maternelle, etc.*). Pour cela, la proposition de nouvelles approches permettant la détection de communautés non chevauchantes a été un défi pour nous.

Dans ce chapitre, nous avons présenté une extension très significative de notre modèle de base LFM. Cette extension concerne l'élimination des cas de chevauchements entre les communautés détectées à la fin du processus.

Plusieurs scénarios ont été discutés afin d'éviter qu'un individu (parent ou enfant) appartienne à plusieurs communautés.

Nous avons démontré que notre modèle de base (LFM) est adaptable aux cas de communautés non chevauchantes avec une mise en œuvre et expérimentation sur une véritable base de données concernant la relation d'amitié entre les abonnés Facebook de la Nouvelle-Orléans.

## **Conclusion générale**

Chaque réseau social, du plus petit au plus grand, doit être analysé pour extraire autant de propriétés, caractéristiques et beaucoup d'informations cachées. Parmi ces informations cachées des réseaux sociaux figure le regroupement de ses utilisateurs en "clusters", on parle alors du concept social de "communautés". Une communauté est un ensemble d'éléments (*abonnés ou utilisateurs d'un réseau social, clients dans un site de vente en ligne, scientifiques dans un site publiant des articles et des livres ...*) ayant des relations plus denses entre eux mais aussi moins de contacts avec le monde extérieur de leur communauté. Par conséquent, parmi les problèmes actuels de l'analyse des réseaux sociaux, on trouve la détection des communautés.

Dans la littérature, les chercheurs ont proposé de nombreux algorithmes pour détecter des communautés comme celui de Louvain, Girvan et bien d'autres. Dans ce sens, nous avons effectué une étude qui a donné une image des approches et algorithmes les plus importants dans ce domaine.

Dans cette thèse, nous avons proposé un nouveau modèle de détection communautaire nommé LFM (*Large Families Model*) en prenant en compte le nombre de voisins communs aux individus. Notre modèle est basé sur la vie sociale des individus au sein de leurs sociétés. Il est divisé en trois principales phases.

Dans **la première** il détecte l'ensemble des parents de chaque famille (*communauté*), ces parents ce sont des paires de nœuds partageant le maximum de fils entre eux. Ils forment avec leurs fils (les individus en relation commune avec eux) cette notion de familles ou communautés initiales.

**La deuxième** phase permet d'en joindre les autres nœuds qui ne sont ni parents ni fils que nous avons appelé "Out\_nodes", ces derniers joindront les différentes communautés selon plusieurs paramètres de préférence que l'utilisateur pourra sélectionner selon son le domaine d'application. Ces paramètres de préférences ont été bien discutés dans notre mémoire.

**La troisième** phase de notre algorithme permet de vérifier les communautés trouvées deux à deux et étudier la faisabilité de les fusionner si elles ont un nombre de nœuds en commun plus que celui en différence.

Le modèle proposé est basique mais aussi générique et extensible. Basique car il détecte les communautés chevauchantes et statiques des réseaux non orientés et non pondérés. Générique car nous avons proposé comment il pourra supporter tous ces axes.

Plusieurs extensions ont été proposées pour prendre en charge le cas de chevauchement, les graphes orientés ainsi que l'aspect dynamique. Dans la section mise en œuvre et expérimentation, nous avons appliqué notre algorithme sur (1) le même exemple donné par Blondel [4] afin de mettre en œuvre son algorithme de Louvain et de le comparer avec notre algorithme proposé. De plus, des réseaux sociaux réels ont été utilisés comme cas d'étude et d'application de notre algorithme comme (2) le réseau d'amitiés entre les membres du club de karaté de Zachary, en tant que petit réseau de référence non orienté et (3) un autre plus large et orienté, que nous avons construit à partir d'une base de données d'une messagerie électronique d'une banque Américaine. Aussi, (4), des réseaux aléatoires créés par notre machine plus larges (*en termes de nombre de nœuds*) et plus denses (*en termes de nombre de relations*) ayant plus de deux mille nœuds et deux millions d'arêtes ont été utilisés afin d'en étudier les limites et les performances de notre algorithme dans les deux cas des réseaux orientés et non orientés. Enfin, nous avons discuté nos résultats en fonction de certains paramètres comme le temps d'exécution, le nombre de communautés détectées, la stabilité et la qualité des distributions communautaires trouvées.

L'objectif principal de la nouvelle approche présentée dans notre travail est de (1) contribuer à la littérature en développant des algorithmes plus efficaces pour la détection des communautés dans les réseaux sociaux, et (2) suggérer une discussion conceptuelle de la structure mathématique, graphique et sociologique des communautés.

Dans ce travail nous avons détaillé les deux catégories de communautés qui existent dans la littérature, faibles et fortes. En proposant une troisième catégorie que nous appelons "moyenne", nous avons pu proposer notre modèle LFM.

Notre modèle présente plusieurs avantages. Le plus importants est (1) ses étapes qui sont très simples à implémenter et mettre en place. Aussi, (2) ses résultats ne sont pas supervisés et il n'a besoin d'aucun paramètre en entrée (*à part le réseau social*). De plus, (3) son exécution est très rapide et son temps de réponse est très raisonnable par rapport à plusieurs autres algorithmes de la littérature comme ceux proposés par Atay [101]. En plus, notre modèle est (4) extensible, évolutif et il support des réseaux sociaux de grande échelle. Aussi, notre modèle est (5) générique et il pourra être extensible et étendu à plusieurs axes. Tout dépend du domaine d'application et du réseau en étude, on peut avoir des communautés chevauchantes ou non chevauchantes, on appliquera notre modèle sur des réseaux pondérés ou simples, orientés ou non, des réseaux statiques ou dynamiques où nous avons la notion de temps des différents événements déclenchés dans le réseau en étude comme l'inscription des nouveaux individus ou la désinscription des anciens, se mettre en contact avec d'autres personnes et créer des nouvelles relations ou bien se désabonner d'une relation sociale quelconque. Dans les réseaux dynamiques, nous prenons en compte le timing de tout événement déclenché afin d'en avoir des flashes instantanés sur les distributions communautaires. De plus, nous détectons les différents changements de ces dernières comme la naissance / mort de communautés, la fusion / division de deux communautés ou plus, la croissance / contraction d'une ou plusieurs communautés.

Un algorithme de détection des communautés dans les réseaux sociaux est dit «**stable**» s'il donne la même structure communautaire finale (*ou similaire*) s'il est ré-exécuté plusieurs fois sur le même réseau. Autrement dit, il fait des choix déterministes à chaque instant. Plusieurs algorithmes proposés dans la littérature ne sont pas stables comme Louvain, alors que notre modèle est stable et il donne toujours le même résultat sur le même exemple.

Le but de chaque algorithme de détection des communautés dans un réseau social est de le diviser en groupes, chaque algorithme peut avoir autant de résultats et autant de distributions communautaires. À cet effet, chacun de ces algorithmes aura besoin d'une fonction objectif qui précisera ses résultats et selon elle nous pourrions choisir (*et décider*) quelle distribution est la meilleure. Cette fonction objectif témoigne du maximum possible de relations au sein des communautés et du minimum possible des relations intercommunautaires. Parmi les fonctions objectifs les plus utilisées dans la littérature nous trouvons la «méthode de maximisation de la modularité» de Newman, c'est elle que nous avons utilisé dans notre travail pour quantifier la **qualité** de nos communautés trouvées.

Notre modèle est basé sur un compromis entre ces deux paramètres : la stabilité et la qualité des distributions communautaires finales où nous avons assuré sa stabilité avec la meilleure qualité possible.

Dans cette thèse, nous avons présenté et implémenté une extension très significative de notre modèle de base LFM. Cette extension concerne l'élimination des cas de chevauchements entre les communautés détectées à la fin du processus. Plusieurs scénarios ont été discutés afin d'éviter qu'un individu (parent ou enfant) appartienne à plusieurs

communautés. Nous avons démontré que notre modèle de base (LFM) est adaptable aux cas de communautés non chevauchantes avec une mise en œuvre et expérimentation sur une véritable base de données concernant la relation d'amitié entre les abonnés Facebook de la Nouvelle-Orléans.

### Perspectives

Comme perspectives et futures travaux nous pouvons proposer les axes suivants :

1. La prise en charge des différentes extensions proposées dans ce travail comme le cas de chevauchement, les graphes pondérés, orientés, complexes et dynamiques.
2. L'adaptation de l'approche aux communautés égo-centrées où au lieu de diviser tout le réseau en communautés, il suffit d'en rajouter un nœud comme "input" et trouver sa communauté.
3. Proposition d'un système de recommandation qui se base sur notre algorithme LFM.
4. Extension de l'approche vers les réseaux sociaux multiplexes où entre les nœuds on trouvera plusieurs relations et pas uniquement une seule. Un réseau multiplexe est un graphe multi-couches où chaque couche contient le même ensemble de nœuds reliés par une relation différente. Dans ce cas, on détectera plusieurs distributions communautaires chacune liée à une relation.
5. Combinaison entre la notion de l'égo-centré et les réseaux multiplexe.
6. Extension de l'approche vers une architecture décentralisée ou réseaux peer-to-peer (P2P) afin d'en garantir au maximum la confidentialité des données personnelles. Dans ce type de réseaux nous n'avons aucun serveur central qui définit les différentes relations entre les individus, chacun parmi ces derniers détient la liste de ses voisins et amis, et on n'a aucune matrice d'adjacence couvrant tout le réseau.
7. Extension de l'approche vers une architecture semi-centralisée où on choisira des super-nœuds (*leaders*) dans un environnement P2P qui vont jouer le rôle des serveurs (*secondaires*) hébergeant les relations entre leurs voisins. Dans ce contexte, l'élection de ces super-nœuds est un axe assez important.
8. Vérifions si notre modèle est applicable sur d'autres domaines que les réseaux sociaux, comme la détection des différents clusters dans les réseaux mobiles et Ad-Hoc, le regroupement des gènes et des protéines dans le domaine de la biologie, le regroupement des électeurs autour des candidats dans le domaine politique, dans le domaine sécuritaire et terrorisme, le domaine économique et systèmes de recommandation, etc.

---

## Bibliographie

- [1]: UTZ, Sonja et BREUER, Johannes. The relationship between networking, linkedin use, and retrieving informational benefits. *Cyberpsychology, Behavior, and Social Networking*, 2019, vol. 22, no 3, p. 180-185.
- [2]: DAVIS, Joanna, WOLFF, Hans-Georg, FORRET, Monica L., *et al.* Networking via LinkedIn: An examination of usage and career benefits. *Journal of Vocational Behavior*, 2020, p. 103396.
- [3]: BOYD, Danah M. et ELLISON, Nicole B. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication*, 2007, vol. 13, no 1, p. 210-230.
- [4]: BLONDEL, Vincent D., GUILLAUME, Jean-Loup, LAMBIOTTE, Renaud, *et al.* Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008, vol. 2008, no 10, p. P10008.
- [5]: NEWMAN, Mark EJ et GIRVAN, Michelle. Finding and evaluating community structure in networks. *Physical review E*, 2004, vol. 69, no 2, p. 026113.
- [6]: MILO, Ron, SHEN-ORR, Shai, ITZKOVITZ, Shalev, *et al.* Network motifs: simple building blocks of complex networks. *Science*, 2002, vol. 298, no 5594, p. 824-827.
- [7]: WATTS, Duncan J. et STROGATZ, Steven H. Collective dynamics of 'small-world' networks. *nature*, 1998, vol. 393, no 6684, p. 440-442.
- [8]: BARABÁSI, Albert-László et ALBERT, Réka. Emergence of scaling in random networks. *science*, 1999, vol. 286, no 5439, p. 509-512.
- [9]: DJERBI, Rachid, AMAD, Mourad, et IMACHE, Rabah. A new model for communities' detection in dynamic social networks inspired from human families. *International Journal of Internet Technology and Secured Transactions*, 2020, vol. 10, no 1-2, p. 24-60. <https://doi.org/10.1504/ijitst.2020.10026089>
- [10]: KOUAKOU, Kouassi Sylvestre. Réseaux sociaux numériques: quels enjeux pour les bibliothèques?. *Cahiers de la Documentation*, 2014, 2014/2. hal-01591750
- [11]: MLAIKI, Alya. *Compréhension de la continuité d'utilisation des réseaux sociaux numériques: Les apports de la théorie du don*. École doctorale de Dauphine-Paris, thèse de doctorat 2012 .
- [12]: BOYD, Danah M. et ELLISON, Nicole B. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication*, 2007, vol. 13, no 1, p. 210-230.
- [13]: HAJEER, Mustafa H., SINGH, Alka, DASGUPTA, Dipankar, *et al.* Clustering online social network communities using genetic algorithms. *arXiv preprint arXiv:1312.2237*, 2013.
- [14]: TOCH, Eran, LERNER, Boaz, BEN-ZION, Eyal, *et al.* Analyzing large-scale human mobility data: a survey of machine learning methods and applications. *Knowledge and Information Systems*, 2019, vol. 58, no 3, p. 501-523.



- [15]: RIOS, Sebastian A., AGUILERA, Felipe, NUÑEZ-GONZALEZ, J. David, *et al.* Semantically enhanced network analysis for influencer identification in online social networks. *Neurocomputing*, 2019, vol. 326, p. 71-81.
- [16]: LIU, Bingsheng, ZHOU, Qi, DING, Ru-Xi, *et al.* Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination. *European Journal of Operational Research*, 2019, vol. 275, no 2, p. 737-754.
- [17]: FROEHLICH, Dominik E. et GEGENFURTNER, Andreas. Social support in transitioning from training to the workplace: A social network perspective. *Beziehungen in pädagogischen Arbeitsfeldern und ihren Transitionen über die Lebensalter*, 2019, p. 208-222.
- [18]: Site officiel de Facebook. <https://www.facebook.com/>, 2020. consulté le 11/02/2020.
- [19]: PAPACHARISSI, Zizi. The virtual geographies of social networks: a comparative analysis of Facebook, LinkedIn and ASmallWorld. *New media & society*, 2009, vol. 11, no 1-2, p. 199-220.
- [20]: ZIDANE, Karima et ALIZIANE, Mohand Ouamer. E-CRM: une nouvelle révolution pour la relation client à l'ère de la digitalisation-Cas d'Optimum Télécom Algérie «Djezzy». *Dirassat Journal Economic Issue*, 2019, vol. 10, no 1, p. 381-393.
- [21]: Site officiel de wikipedia. <https://fr.wikipedia.org/>, 2020. consulté le 11/02/2020.
- [22]: Site officiel de Twitter. <https://twitter.com/>, 2020. consulté le 11/02/2020.
- [23]: Site officiel de YouTube. <https://www.youtube.com/>, 2020. consulté le 11/02/2020.
- [24]: MERRA, Lucile. *Pour une sociologie des médias sociaux. Internet et la révolution médiatique: nouveaux médias et interactions*. Thèse de doctorat Sociologie. Paris Sorbonne Cité - Paris Descartes, 2013. Français.
- [25]: Site officiel de Instagram. <https://www.instagram.com/>, 2020. consulté le 11/02/2020.
- [26]: HU, Yuheng, MANIKONDA, Lydia, et KAMBHAMPATI, Subbarao. What we instagram: A first analysis of instagram photo content and user types. In : *Eighth International AAI conference on weblogs and social media*, USA, 2014.
- [27]: NAAMAN, Mor, BOASE, Jeffrey, et LAI, Chih-Hui. Is it really about me? Message content in social awareness streams. In : *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. Georgia, USA, February 6–10. 2010. p. 189-192.
- [28]: A propos de facebook. <https://about.fb.com/news/2012/04/facebook-to-acquire-instagram/>, 2020. consulté le 12/02/2020.
- [29]: Site officiel de Google+. <https://plus.google.com/>, 2020. consulté le 28/02/2020.
- [30]: Blog Google. <https://www.blog.google/technology/safety-security/project-strobe/>, 2020. consulté le 12/02/2020.
- [31]: Site officiel de LinkedIn. <https://www.linkedin.com/>, 2020. consulté le 11/02/2020.
- [32]: Site officiel de Viadeo. <http://www.viadeo.com/>, 2020. consulté le 17/02/2020.
- [33]: Site officiel de Yammer. <https://www.yammer.com/>, 2020. consulté le 17/02/2020.

- [34]: RIEMER, Kai, SCIFLEET, Paul, et REDDIG, Ruwen. Powercrowd: Enterprise social networking in professional service work: A case study of Yammer at Deloitte Australia. 2012.
- [35]: ZHANG, Jun, QU, Yan, CODY, Jane, *et al.* A case study of micro-blogging in the enterprise: use, value, and related issues. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Georgia USA April, 2010. p. 123-132.
- [36]: T.B. Lee, "Facebook's Cambridge Analytica Scandal, Explained," *Ars Technica*, 20 Mar. 2018; vol. 20. <https://arstechnica.com/tech-policy/2018/03/facebooks-cambridge-analytica-scandal-explained>.
- [37]: WOJCIK, Stefan, MESSING, Solomon, SMITH, Aaron, *et al.* *Bots in the Twittersphere: An Estimated Two-thirds of Tweeted Links to Popular Websites are Posted by Automated Accounts-Not Human Beings*. Pew Research Center, USA, 2018.
- [38]: MORENO, J. L. Emotions mapped by new geography. *New York Times*, 1933, vol. 3, p. 17.
- [39]: DE, Sagar S. et DEHURI, Satchidananda. Machine learning for auspicious social network mining. In : *Social Networking*. Springer, Cham, 2014. p. 45-83.
- [40]: YANG, Bo, LI, Xu, LIU, Xiangwei, *et al.* Alternating between consensus and leader selection reveals community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 2019, vol. 515, p. 693-706.
- [41]: GMATI, Haifa, MOUAKHER, Amira, GONZALEZ-PARDO, Antonio, *et al.* A new algorithm for communities detection in social networks with node attributes. *Journal of Ambient Intelligence and Humanized Computing*, 2018, p. 1-13.
- [42]: LI, Xuequn, ZHOU, Shuming, LIU, Jiafei, *et al.* Communities detection in social network based on local edge centrality. *Physica A: Statistical Mechanics and its Applications*, 2019, vol. 531, p. 121552.
- [43]: ŽALIK, Krista Rizman. Evolution algorithm for community detection in social networks using node centrality. In : *Intelligent methods and big data in industrial applications*. Springer, Cham, 2019. p. 73-87.
- [44]: AL-AYYOUB, Mahmoud, AL-ANDOLI, Mohammed, JARARWEH, Yaser, *et al.* Improving fuzzy C-mean-based community detection in social networks using dynamic parallelism. *Computers & Electrical Engineering*, 2019, vol. 74, p. 533-546.
- [45]: ZHAO, Zhongying, LI, Chao, ZHANG, Xuejian, *et al.* An incremental method to detect communities in dynamic evolving social networks. *Knowledge-Based Systems*, 2019, vol. 163, p. 404-415.
- [46]: COLEMAN, James S. Social capital in the creation of human capital. *American journal of sociology*, 1988, vol. 94, p. S95-S120.
- [47]: BURT, Ronald S. Closure as social capital. *Social capital: Theory and research*, 2001, p. 31-55.



- [48]: RAVASZ, Erzsébet, SOMERA, Anna Lisa, MONGRU, Dale A., *et al.* Hierarchical organization of modularity in metabolic networks. *science*, 2002, vol. 297, no 5586, p. 1551-1555.
- [49]: SPIRIN, Victor et MIRNY, Leonid A. Protein complexes and functional modules in molecular networks. *Proceedings of the national Academy of sciences*, USA, 2003, vol. 100, no 21, p. 12123-12128.
- [50]: ONNELA, J.-P., CHAKRABORTI, Anirban, KASKI, Kimmo, *et al.* Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 2003, vol. 68, no 5, p. 056110.
- [51]: SCOTT, John. Social network analysis. *Sociology*, USA, 1988, vol. 22, no 1, p. 109-127.
- [52]: WATTS, Duncan J., DODDS, Peter Sheridan, et NEWMAN, Mark EJ. Identity and search in social networks. *science*, 2002, vol. 296, no 5571, p. 1302-1305.
- [53]: PALLA, Gergely, DERÉNYI, Imre, FARKAS, Illés, *et al.* Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 2005, vol. 435, no 7043, p. 814-818.
- [54]: GREGORY, Steve. Finding overlapping communities in networks by label propagation. *New journal of Physics*, 2010, vol. 12, no 10, p. 103018.
- [55]: JIN, Di, GABRYS, Bogdan, et DANG, Jianwu. Combined node and link partitions method for finding overlapping communities in complex networks. *Scientific reports*, 2015, vol. 5, p. 8600.
- [56]: LI, Qi, ZHONG, Jiang, LI, Qing, *et al.* A community merger of optimization algorithm to extract overlapping communities in networks. *IEEE Access*, 2018, vol. 7, p. 3994-4005.
- [57]: MESSAOUDI, Abir, MISSAOUI, Rokia, et IBRAHIM, Mohamed-Hamza. Detecting Overlapping Communities in Two-mode Data Networks using Formal Concept Analysis. In : *Extraction et Gestion des Connaissances: Actes de la conférence EGC'2019*. BoD-Books on Demand, 2019.
- [58]: CHAKRABORTY, Tanmoy, KUMAR, Suhansanu, GANGULY, Niloy, *et al.* GenPerm: A unified method for detecting non-overlapping and overlapping communities. *IEEE Transactions on knowledge and data engineering*, 2016, vol. 28, no 8, p. 2101-2114.
- [59]: HAJIABADI, Mahdi, ZARE, Hadi, et BOBARSHAD, Hossein. IEDC: An integrated approach for overlapping and non-overlapping community detection. *Knowledge-Based Systems*, 2017, vol. 123, p. 188-199.
- [60]: GHALMANE, Zakariya, EL HASSOUNI, Mohammed, et CHERIFI, Hocine. Immunization of networks with non-overlapping community structure. *Social Network Analysis and Mining*, 2019, vol. 9, no 1, p. 45.
- [61]: FENG, Yuhong, SHE, Song, WU, Yuanshi, *et al.* A RESTful Web Service for Non-overlapping Community Quality Assessment with MPI. In : *International Conference on Web Services*, USA. Springer, Cham, 2018. p. 280-294.

- [62]: RADICCHI, Filippo, CASTELLANO, Claudio, CECCONI, Federico, *et al.* Defining and identifying communities in networks. *Proceedings of the national academy of sciences*, 2004, vol. 101, no 9, p. 2658-2663.
- [63]: ANCHURI, Pranay et MAGDON-ISMAIL, Malik. Communities and balance in signed networks: A spectral approach. In : *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Turkey, IEEE, 2012. p. 235-242.
- [64]: CAPOCCI, Andrea, SERVEDIO, Vito DP, CALDARELLI, Guido, *et al.* Detecting communities in large networks. *Physica A: Statistical Mechanics and its Applications*, 2005, vol. 352, no 2-4, p. 669-676.
- [65]: TABATABAEI, Seyed Salim, COATES, Mark, et RABBAT, Michael. GANC: Greedy agglomerative normalized cut for graph clustering. *Pattern Recognition*, 2012, vol. 45, no 2, p. 831-843.
- [66]: NEWMAN, Mark EJ. Fast algorithm for detecting community structure in networks. *Physical review E*, 2004, vol. 69, no 6, p. 066133.
- [67]: SNIJDERS, Tom AB et NOWICKI, Krzysztof. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 1997, vol. 14, no 1, p. 75-100.
- [68]: AIROLDI, Edoardo M., BLEI, David M., FIENBERG, Stephen E., *et al.* Mixed membership stochastic blockmodels. *Journal of machine learning research*, 2008, vol. 9, no Sep, p. 1981-2014.
- [69]: NEWMAN, Mark EJ. Spectral methods for community detection and graph partitioning. *Physical Review E*, 2013, vol. 88, no 4, p. 042822.
- [70]: CHAUDHARY, Laxmi et SINGH, Buddha. Community detection using maximizing modularity and similarity measures in social networks. In : *Smart Systems and IoT: Innovations in Computing*. Springer, Singapore, 2020. p. 197-206.
- [71]: CHEN, Xiangtao et LI, Juan. Community detection in complex networks using edge-deleting with restrictions. *Physica A: Statistical Mechanics and its Applications*, 2019, vol. 519, p. 181-194.
- [72]: FORTUNATO, Santo. Community detection in graphs. *Physics reports*, 2010, vol. 486, no 3-5, p. 75-174.
- [73]: AYNAUD, Thomas et GUILLAUME, Jean-Loup. Static community detection algorithms for evolving networks. In : *8th International symposium on modeling and optimization in mobile, Ad Hoc, and wireless networks, USA*. IEEE, 2010. p. 513-519.
- [74]: FORTUNATO, Santo et BARTHELEMY, Marc. Resolution limit in community detection. *Proceedings of the national academy of sciences*, 2007, vol. 104, no 1, p. 36-41.
- [75]: DJERBI, Rachid, IMACHE, Rabah, et AMAD, Mourad. Communities' Detection in Social Networks: State of the art and perspectives. In : *2018 International Symposium on Networks, Computers and Communications, Italy (ISNCC)*. IEEE, 2018. p. 1-6. <https://doi.org/10.1109/isncc.2018.8531055>

- [76]: ZALIK, Krista Rizman et ZALIK, Borut. Memetic algorithm using node entropy and partition entropy for community detection in networks. *Information Sciences*, 2018, vol. 445, p. 38-49.
- [77]: KERNIGHAN, Brian W. et LIN, Shen. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, 1970, vol. 49, no 2, p. 291-307.
- [78]: GIRVAN, Michelle et NEWMAN, Mark EJ. Community structure in social and biological networks. *Proceedings of the national academy of sciences, USA*, 2002, vol. 99, no 12, p. 7821-7826.
- [79]: FORTUNATO, Santo, LATORA, Vito, et MARCHIORI, Massimo. Method to find community structures based on information centrality. *Physical review E*, 2004, vol. 70, no 5, p. 056104.
- [80]: CLAUSET, Aaron, NEWMAN, Mark EJ, et MOORE, Cristopher. Finding community structure in very large networks. *Physical review E*, 2004, vol. 70, no 6, p. 066111.
- [81]: NEWMAN, Mark EJ. Fast algorithm for detecting community structure in networks. *Physical review E*, 2004, vol. 69, no 6, p. 066133.
- [82]: PONS, Pascal et LATAPY, Matthieu. Computing communities in large networks using random walks. In : *International symposium on computer and information sciences*. Springer, Berlin, Heidelberg, 2005. p. 284-293.
- [83]: PALLA, Gergely, DERÉNYI, Imre, FARKAS, Illés, *et al.* Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 2005, vol. 435, no 7043, p. 814-818.
- [84]: PALLA, Gergely, BARABÁSI, Albert-László, et VICSEK, Tamás. Quantifying social group evolution. *Nature*, 2007, vol. 446, no 7136, p. 664-667.
- [85]: RAGHAVAN, Usha Nandini, ALBERT, Réka, et KUMARA, Soundar. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 2007, vol. 76, no 3, p. 036106.
- [86]: TANTIPATHANANANDH, Chayant, BERGER-WOLF, Tanya, et KEMPE, David. A framework for community identification in dynamic social networks. In : *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, USA*. 2007. p. 717-726.
- [87]: CHEVALEYRE, Yann, ENDRISS, Ulle, LANG, Jérôme, *et al.* A short introduction to computational social choice. In : *International Conference on Current Trends in Theory and Practice of Computer Science*. Springer, Berlin, Heidelberg, 2007. p. 51-69.
- [88]: FALKOWSKI, Tanja, BARTH, Anja, et SPILIOPOULOU, Myra. Studying community dynamics with an incremental graph mining algorithm. *AMCIS 2008 Proceedings*, Canada, 2008, p. 29.
- [89]: ROSVALL, Martin et BERGSTROM, Carl T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences, USA*, 2008, vol. 105, no 4, p. 1118-1123.

- [90]: KIM, Min-Soo et HAN, Jiawei. A particle-and-density based evolutionary clustering method for dynamic networks. *Proceedings of the VLDB Endowment*, France. 2009, vol. 2, no 1, p. 622-633.
- [91]: XIANG, Biao, CHEN, En-Hong, et ZHOU, Tao. Finding community structure based on subgraph similarity. In : *Complex networks*. Springer, Berlin, Heidelberg, 2009. p. 73-81.
- [92]: GREGORY, Steve. Finding overlapping communities in networks by label propagation. *New journal of Physics*, 2010, vol. 12, no 10, p. 103018.
- [93]: KANAWATI, Rushed. Licod: Leaders identification for community detection in complex networks. In : *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, USA. IEEE, 2011. p. 577-582.
- [94]: SHANG, Ronghua, BAI, Jing, JIAO, Licheng, *et al.* Community detection based on modularity and an improved genetic algorithm. *Physica A: Statistical Mechanics and its Applications*, 2013, vol. 392, no 5, p. 1215-1231.
- [95]: RAVASZ, Erzsébet, SOMERA, Anna Lisa, MONGRU, Dale A., *et al.* Hierarchical organization of modularity in metabolic networks. *science*, 2002, vol. 297, no 5586, p. 1551-1555.
- [96]: LAKHLEF, Hicham. A multi-level clustering scheme based on cliques and clusters for wireless sensor networks. *Computers & Electrical Engineering*, 2015, vol. 48, p. 436-450.
- [97]: SAOUD, Bilal et MOUSSAOUI, Abdelouahab. Community detection in networks based on minimum spanning tree and modularity. *Physica A: Statistical Mechanics and its Applications*, 2016, vol. 460, p. 230-234.
- [98]: BILAL, Saoud et ABDELOUAHAB, Moussaoui. Evolutionary algorithm and modularity for detecting communities in networks. *Physica A: Statistical Mechanics and its Applications*, 2017, vol. 473, p. 89-96.
- [99]: DE SALVE, Andrea, GUIDI, Barbara, et RICCI, Laura. Evaluation of structural and temporal properties of ego networks for data availability in DOSNs. *Mobile Networks and Applications*, 2018, vol. 23, no 1, p. 155-166.
- [100]: FAZLALI, Mahmood, MORADI, Ehsan, et MALAZI, Hadi Tabatabaee. Adaptive parallel Louvain community detection on a multicore platform. *Microprocessors and Microsystems*, 2017, vol. 54, p. 26-34.
- [101]: ATAY, Yilmaz, KOC, Ismail, BABA OGLU, Ismail, *et al.* Community detection from biological and social networks: A comparative analysis of metaheuristic algorithms. *Applied Soft Computing*, 2017, vol. 50, p. 194-211.
- [102]: MOOSAVI, Seyed Ahmad, JALALI, Mehrdad, MISAGHIAN, Negin, *et al.* Community detection in social networks using user frequent pattern mining. *Knowledge and Information Systems*, 2017, vol. 51, no 1, p. 159-186.
- [103]: HUANG, Mingqing, ZOU, Guobing, ZHANG, Bofeng, *et al.* Overlapping community detection in heterogeneous social networks via the user model. *Information Sciences*, 2018, vol. 432, p. 164-184.

- [104]: AHAJJAM, Sara, EL HADDAD, Mohamed, et BADIR, Hassan. A new scalable leader-community detection approach for community detection in social networks. *Social Networks*, 2018, vol. 54, p. 41-49.
- [105]: AL-AYYOUB, Mahmoud, AL-ANDOLI, Mohammed, JARARWEH, Yaser, *et al.* Improving fuzzy C-mean-based community detection in social networks using dynamic parallelism. *Computers & Electrical Engineering*, 2019, vol. 74, p. 533-546.
- [106]: HE, Kun, LI, Yingru, SOUNDARAJAN, Sucheta, *et al.* Hidden community detection in social networks. *Information Sciences*, 2018, vol. 425, p. 92-106.
- [107]: SAID, Anwar, ABBASI, Rabeeh Ayaz, MAQBOOL, Onaiza, *et al.* CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks. *Applied Soft Computing*, 2018, vol. 63, p. 59-70.
- [108]: NEDIOUI, MED ABDELHAMID. *Fouille et apprentissage automatique dans les réseaux sociaux dynamique*. 2015. Thèse de doctorat. Université Mohamed Khider-Biskra, Algérie.
- [109]: NEWMAN, Mark EJ. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, USA, 2001, vol. 98, no 2, p. 404-409.
- [110]: GLEISER, Pablo M. et DANON, Leon. Community structure in jazz. *Advances in complex systems*, 2003, vol. 6, no 04, p. 565-573.
- [111]: LANCICHINETTI, Andrea et FORTUNATO, Santo. Community detection algorithms: a comparative analysis. *Physical review E*, 2009, vol. 80, no 5, p. 056117.
- [112]: LESKOVEC, Jure, LANG, Kevin J., et MAHONEY, Michael. Empirical comparison of algorithms for network community detection. In : *Proceedings of the 19th international conference on World wide web*, USA. 2010. p. 631-640.
- [113]: SHI, Jianbo et MALIK, Jitendra. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 2000, vol. 22, no 8, p. 888-905.
- [114]: FORTUNATO, Santo et CASTELLANO, Claudio. Community structure in graphs. *arXiv preprint arXiv:0712.2716*, 2007.
- [115]: TRAUD, Amanda L., KELSIC, Eric D., MUCHA, Peter J., *et al.* Comparing community structure to characteristics in online collegiate social networks. *SIAM review*, 2011, vol. 53, no 3, p. 526-543.
- [116]: HOPCROFT, John, KHAN, Omar, KULIS, Brian, *et al.* Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences*, USA, 2004, vol. 101, no suppl 1, p. 5249-5253.
- [117]: KUO, R. J., LIN, T. C., ZULVIA, Ferani E., *et al.* A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. *Applied Soft Computing*, 2018, vol. 67, p. 299-308.
- [118]: REICHARDT, Jörg et BORNHOLDT, Stefan. Statistical mechanics of community detection. *Physical review E*, 2006, vol. 74, no 1, p. 016110.

- [119]: KANNAN, Kavitha. Data Mining Report on Iris and Australian Credit Card Dataset. *School of Computer Science and Information Technology, University Putra Malaysia, Serdang, Selangor, Malaysia*, 1988.
- [120]: CORTEZ, Paulo, CERDEIRA, António, ALMEIDA, Fernando, *et al.* Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 2009, vol. 47, no 4, p. 547-553.
- [121]: J. P. Siebert, “Vehicle Recognition Using Rule Based Methods”, Turing Institute Research Memorandum TIRM, United Kingdom, March 1987.
- [122]: DUCH, Jordi et ARENAS, Alex. Community detection in complex networks using extremal optimization. *Physical review E*, 2005, vol. 72, no 2, p. 027104.
- [123]: ZACHARY, Wayne W. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 1977, vol. 33, no 4, p. 452-473.
- [124]: JEONG, Hawoong, TOMBOR, Bálint, ALBERT, Réka, *et al.* The large-scale organization of metabolic networks. *Nature*, 2000, vol. 407, no 6804, p. 651-654.
- [125]: EBEL, Holger, MIELSCH, Lutz-Ingo, et BORNHOLDT, Stefan. Scale-free topology of e-mail networks. *Physical review E*, 2002, vol. 66, no 3, p. 035103.
- [126]: GUARDIOLA, X., GUIMERA, R., ARENAS, A., *et al.* Macro-and micro-structure of trust networks. *arXiv preprint cond-mat/0206240*, 2002.
- [127]: SHETTY, Jitesh et ADIBI, Jafar. Discovering important nodes through graph entropy the case of enron email database. In : *Proceedings of the 3rd international workshop on Link discovery*. 2005. p. 74-81.
- [128]: TRAAG, Vincent A. Faster unfolding of communities: Speeding up the Louvain algorithm. *Physical Review E*, 2015, vol. 92, no 3, p. 032801.
- [129]: VISWANATH, Bimal, MISLOVE, Alan, CHA, Meeyoung, *et al.* On the evolution of user interaction in facebook. In : *Proceedings of the 2nd ACM workshop on Online social networks*. 2009. p. 37-42.
- [130]: Rachid Djerbi, Mourad Amad, Rabah Imache, A Novel Solution for Communities' Detection in Social Networks, in *The 3rd International Conference on Networking and Advanced Systems*, 2017, ICNAS2017 (Annaba, Algeria), 2017.