

UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉTECTION DE NŒUDS CENTRAUX ET DE COMMUNAUTÉS DANS LES  
RÉSEAUX SOCIAUX À DEUX MODES

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN SCIENCES ET TECHNOLOGIES DE L'INFORMATION

PAR

AYAO BOBI

et évalué par le jury composé de :

Prof. Ana-Maria Cretu ..... Présidente du jury  
Prof. Alan Davoust ..... Membre du jury  
Prof. Rokia Missaoui ..... Directrice de recherche  
Dr. Mohamed Hamza Ibrahim ..... Codirecteur de recherche

DÉCEMBRE 2023

# Résumé

Dans le domaine de l'analyse des réseaux sociaux, l'identification des nœuds centraux et la détection des communautés sont des thèmes de recherche essentiels qui permettent de mieux comprendre la dynamique des réseaux, le flux d'informations et le comportement collectif. Ce mémoire étudie ces domaines critiques en exploitant la puissance de l'Analyse Formelle des Concepts (AFC).

Les objectifs principaux de cette recherche sont d'améliorer, de valider et d'appliquer un indice de pertinence développé par l'équipe du Laboratoire de Recherche sur l'Information Multimédia (LARIM) de l'UQO. L'application de la mesure s'étend à la détection de communautés cohésives et chevauchantes dans les réseaux à deux modes, ainsi qu'à l'identification des nœuds utiles au sein du réseau et potentiellement au sein de chaque communauté identifiée. En outre, l'étude exploite l'AFC non seulement pour atteindre ces objectifs essentiels, mais aussi pour découvrir des concepts formels sémantiquement pertinents dans les treillis de concepts, révélant ainsi des grappes homogènes d'objets caractérisés par leurs attributs et favorisant la génération de nouvelles idées.

L'intérêt de cette recherche réside dans ses implications profondes pour la compréhension de la nature complexe des réseaux sociaux et de leurs structures sous-jacentes. En identifiant efficacement les nœuds centraux et les communautés au sein du réseau, les chercheurs et les praticiens acquièrent des connaissances inestimables sur la propagation de l'influence, la diffusion de l'information et les comportements émergents. Alors que les mesures de centralité traditionnelles ont apporté des contributions substantielles, cette recherche introduit une nouvelle perspective en incorporant l'indice de pertinence. Cette méthode a le potentiel d'améliorer la précision et l'utilité de l'évaluation de la centralité, en particulier dans les réseaux à deux modes où les nœuds appartiennent à deux types distincts. Cela donnerait une image plus complète du fonctionnement des réseaux.

Cette recherche se distingue par l'intégration de l'AFC dans le domaine de l'analyse des réseaux sociaux. L'AFC, développée à l'origine pour formaliser les relations entre

les objets et les attributs et identifier des groupes conceptuels , a trouvé une autre application dans cette étude. Cette nouvelle intégration met en évidence la polyvalence de l'AFC en tant qu'outil permettant non seulement d'améliorer les techniques d'analyse des réseaux, mais aussi de repousser les frontières de l'extraction de connaissances dans le domaine des réseaux sociaux.

Les expériences sur des réseaux et des contextes réels et générés aléatoirement montrent l'efficacité de notre amélioration de l'indice de pertinence des concepts formels ainsi que la capacité à détecter des communautés chevauchantes basées sur des concepts et des nœuds centraux par rapport à d'autres méthodes de pointe de l'état de l'art.

**Mots-clés :** Analyse formelle de concepts - pertinence de concepts, d'objets et d'attributs - Détection de communautés - Réseaux à deux modes de données - Graphes bipartis.

# Abstract

In the realm of social network analysis, the identification of central nodes and the detection of communities stand as pivotal key research topics that enable a deeper understanding of network dynamics, information flow, and collective behaviour. This thesis embarks on a comprehensive exploration of these critical areas by harnessing the power of Formal Concept Analysis (FCA). The primary objectives of this thesis are to enhance, validate, and apply a relevance score developed by the *LABoratoire de Recherche sur l'InforMation Multimédia* (LARIM) team. The score's applicability extends to identifying cohesive and overlapping communities within two-mode networks, as well as unveiling actionable nodes dispersed throughout the network and potentially within each revealed community. Furthermore, the study exploits FCA not only to achieve these pivotal goals but also to uncover semantically relevant formal concepts within concept lattices, thereby revealing homogeneous clusters of objects characterized by their attributes and fostering the generation of novel insights.

The relevance of this research lies in its profound implications for understanding the intricate nature of social networks and their underlying structures. By effectively identifying central nodes and communities, researchers and practitioners gain invaluable insights into influence propagation, information dissemination, and emerging behaviours. While traditional centrality measures have provided substantial contributions, this thesis introduces a novel perspective through the incorporation of the relevance score. This method has the potential to improve the accuracy and usefulness of centrality assessment, especially in two-mode networks where nodes belong to two different types. This would give a more complete picture of how networks work.

A distinctive feature of this research is its integration of FCA into the realm of social network analysis. FCA, originally developed for formalizing the relationships between objects and attributes and generating conceptual clusters, has found another application in this study.

This novel integration showcases the versatility of FCA as a tool not only for enhancing network analysis techniques but also for expanding the frontiers of knowledge extraction within the social network domain.

The experiments on real-world and randomly generated two-mode networks and contexts show the efficiency of our improvement in the relevance score as well as the ability to detect conceptual-based overlapping communities and central nodes compared to other state-of-the art methods.

**Keywords :** Formal Concept Analysis - Concept, object and attribute relevancy - Community detection - Two-mode data networks - Bipartite graphs.

# Remerciements

Avant de présenter ce travail, je tiens à exprimer ma gratitude aux personnes qui m'ont aidé, de près ou de loin avec leur soutien. Qu'ils trouvent ici l'expression de toute ma gratitude, collectivement et individuellement.

Je tiens à exprimer ma gratitude à tous les membres du jury pour m'avoir fait l'honneur de participer à l'évaluation de ce travail.

Je voudrais exprimer toute ma gratitude à la professeure Rokia, ma directrice de recherche et au docteur Mohamed, pour leur disponibilité, leur patience et surtout pour leurs conseils, suggestions et recommandations judicieuses et utiles tout au long de ces travaux de recherche.

Je tiens aussi à saluer chaleureusement tous mes professeurs de l'UQO.

Enfin, je veux exprimer ma gratitude sincère à ma famille, qui reste un soutien moral indéfectible, et à tous mes amis de promotion pour leurs conseils.

# Table des matières

Résumé	i
Remerciements	v
Liste des figures	viii
Liste des tableaux	x
Liste des abréviations, sigles et acronymes	xii
<b>1 Introduction</b>	<b>1</b>
1.1 Contexte . . . . .	1
1.2 Problématique . . . . .	2
1.3 Objectifs de la recherche . . . . .	2
<b>2 Rappels et état de l’art</b>	<b>4</b>
2.1 Définition et types de réseaux . . . . .	4
2.1.1 Définition . . . . .	4
2.1.2 Types des réseaux . . . . .	5
2.2 Analyse formelle de concepts . . . . .	6
2.3 Les mesures des concepts formels . . . . .	11
2.3.1 Stabilité . . . . .	11
2.3.2 Séparation . . . . .	13
2.4 Détection de nœuds centraux et de communautés . . . . .	14
2.4.1 Méthodes d’analyse des réseaux sociaux . . . . .	17
2.4.2 Détection des communautés . . . . .	19
2.4.3 Algorithmes de détection des communautés . . . . .	20

<b>3</b>	<b>Approche proposée</b>	<b>23</b>
3.1	Calcul de la pertinence des concepts . . . . .	24
3.2	Détection des nœuds importants . . . . .	28
3.3	Détection des communautés fondées sur les concepts . . . . .	32
3.3.1	Extraction des concepts pertinents . . . . .	36
3.3.2	Détection d'objets pertinents . . . . .	43
<b>4</b>	<b>Analyse empirique</b>	<b>46</b>
4.1	Expérimentation . . . . .	46
4.1.1	Méthodologie . . . . .	46
4.1.2	Résultats . . . . .	48
4.1.3	Discussions . . . . .	54
<b>5</b>	<b>Conclusion</b>	<b>57</b>
<b>A</b>		<b>59</b>
	<b>Bibliographie</b>	<b>61</b>



# Liste des figures

2.1	Exemple de treillis de concepts pour le contexte du tableau 2.1 . . . . .	8
2.2	Exemple de treillis de concepts en étiquetage réduit pour le contexte du tableau 2.1 . . . . .	9
2.3	Exemple de treillis de concepts avec les générateurs minimaux pour le contexte du tableau 2.1 . . . . .	10
2.4	Visualisation du graphe dans le logiciel UCINET . . . . .	15
2.5	Exemple de distance géodésique . . . . .	16
2.6	Exemple de trois communautés délimitées par des cercles en pointillé . .	20
3.1	Diagramme des étapes de notre approche. Les losanges représentent les différentes opérations à réaliser sur les données alors que les ovales montrent les entrées ou sorties des opérations. Le sens des flèches indique le sens du déroulement des opérations. . . . .	34
3.2	Treillis issu du contexte <i>Woman-Southern-Davis</i> . . . . .	36
3.3	Figure des communautés de base du contexte <i>Woman-Southern-Davis</i> identifiées par notre approche . . . . .	40
3.4	Résultat de la méta-analyse de Freeman [17] . . . . .	42
4.1	Diagramme de dispersion entre l'indice de pertinence et l'indice de stabilité pour le jeu de données <i>Coin-Toss</i> . . . . .	49
4.2	Diagramme de dispersion entre l'indice de pertinence et l'indice de stabilité pour le jeu de données <i>Woman-Southern-Davis</i> . . . . .	49
4.3	Diagramme de dispersion entre l'indice de pertinence et l'indice de stabilité pour le jeu de données " <i>Phytothérapie</i> " . . . . .	50
4.4	Temps moyen d'exécution de l'indice de pertinence et l'indice de stabilité pour le jeu de données <i>Coin-Toss</i> . . . . .	51

4.5	Temps moyen d'exécution de l'indice de pertinence et l'indice de stabilité pour le jeu de données <i>Woman-Southern-Davis</i> . . . . .	51
4.6	Temps moyen d'exécution de l'indice de pertinence et l'indice de stabilité pour le jeu de données " <i>Phytothérapie</i> ". . . . .	52
A.1	Exemple de code Python pour le calcul des générateurs minimaux . . . .	59
A.2	Exemple de code Python pour l'affichage de $\alpha_{ex}(c)$ et $\beta_{ex}(c)$ . . . . .	60

# Liste des tableaux

2.1	Exemple de contexte formel tiré de [12]	6
2.2	Les concepts formels avec leurs extensions et intentions	7
2.3	Les résultats du calcul des indices de séparation $\gamma(c)$	14
3.1	Les résultats des calculs de $\alpha_{ex}(c)$ et $\beta_{ex}(c)$	28
3.2	Tableau des concepts extensionnellement pertinents.	30
3.3	Tableau des objets pertinents parmi les concepts du tableau 3.2.	30
3.4	Les résultats des calculs de $\alpha_{in}(c)$ , $\beta_{in}(c)$ et du $\mathcal{CR}_{in}(c)$	31
3.5	Tableau des concepts intentionnellement pertinents.	31
3.6	Tableau des attributs conceptuellement pertinents parmi les concepts du tableau 3.5.	32
3.7	Contexte formel de Woman-Southern-Davis	35
3.8	Tableau des concepts de base	37
3.9	Tableau des concepts avec les composantes $\alpha_{ex}$ , $\beta_{ex}$ , la séparation ( $\gamma(c)$ ) et l'autonomie ( $\zeta(c)$ )	39
3.10	Tableau des communautés de base avec les liens partagés.	40
3.11	Tableau des communautés finales.	41
3.12	Tableau des concepts extensionnellement pertinents.	43
3.13	Tableau des objets pertinents parmi les concepts du tableau 3.12.	44
3.14	Tableau des concepts pertinents selon l'indice de stabilité.	45
4.1	Une description des ensembles de données testés, où $ \mathcal{G} $ est le nombre d'objets, $ \mathcal{M} $ est le nombre d'attributs, $n$ est le nombre de liens entre les ensembles $\mathcal{G}$ et $\mathcal{M}$ , $k$ est le nombre de concepts produits par le treillis et pour lesquels l'indice de pertinence et de stabilité a été calculé et $\delta$ est la densité du contexte avec $\delta = \frac{n}{ \mathcal{G}  \times  \mathcal{M} }$	47

4.2	Tableau des calculs des mesures de centralité de degré ( $D_c$ ), de la centralité de proximité ( $C_c$ ) et de la centralité d'intermédiation ( $B_c$ ) . . . . .	53
4.3	Coefficient de Kendall entre les mesures de centralité de degré ( $D_c$ ), de la centralité de proximité ( $C_c$ ), la centralité d'intermédiation ( $B_c$ ) et la centralité basée sur la pertinence des concepts $Freq_{CR}$ . . . . .	54

# Liste des abréviations, sigles et acronymes

**AFC** : Analyse formelle de concepts

**ARS** : Analyse des réseaux sociaux

**CR Index** : *Conceptual Relevancy Index*

**DC** : Détection de communautés

**FCA** : *Formal Concept Analysis*

**FD** : Fouille des données

**LARIM** : LAboratoire de Recherche sur l'Information Multimédia

# Chapitre 1

## Introduction

### 1.1 Contexte

L'Internet et les progrès technologiques des dernières décennies ont largement changé notre façon de communiquer et de partager les informations avec nos interlocuteurs pouvant se trouver dans des zones géographiques très éloignées. Cette révolution a été possible notamment grâce à Internet qui a facilité la naissance des réseaux sociaux.

Par leur mode de fonctionnement qui permet de mettre en contact des individus, familles ou organisations liés d'une certaine manière, les réseaux sociaux sont aujourd'hui la base d'une grande partie des échanges de données et d'informations effectués dans le monde. Ils permettent d'acheminer des informations presque instantanément. Plusieurs acteurs y sont présents presque tous les jours et contribuent à son évolution. Ce domaine a connu une forte accélération ces dernières années avec l'apparition et le développement des médias sociaux comme Meta (*Facebook*), *LinkedIn* ou encore *Twitter*.

Par ce fait même, nous assistons à un développement de plusieurs groupes ou communautés qui se sont formés de manière explicite et constitués par un ensemble d'individus qui interagissent plus souvent entre eux qu'avec les autres et ayant une affinité commune ou un certain nombre d'intérêts. La détection des communautés prend alors une dimension importante et permet notamment l'identification des profils types, la mise en œuvre des actions ciblées, un meilleur ajustement des recommandations ainsi que la réorganisation et l'identification des acteurs centraux ou influents.

---

## 1.2 Problématique

Les réseaux sociaux sont souvent modélisés par des graphes, une structure permettant l’encodage d’instances (objets, individus, événements, etc.) interreliées où les nœuds correspondent aux différentes instances et les liens représentent les relations entre elles. Quel que soit le domaine applicatif, cette modélisation sert à l’étude de la structure émergeant des entrelacements entre individus. Généralement, la densité des liens entre les nœuds du réseau varie d’une zone à une autre, ce qui implique l’existence de groupes de nœuds fortement connectés entre eux mais faiblement reliés aux autres nœuds du réseau. Ces zones, appelées communautés, peuvent être définies comme des ensembles de nœuds fortement liés entre eux, et plus faiblement liés avec le reste du réseau. De ce fait, l’identification de communautés est un sujet d’actualité puisqu’il peut être rencontré dans plusieurs domaines d’application et des situations du monde réel. En plus, il est également possible de déterminer le rôle de différents acteurs au sein des communautés et dans le réseau dans sa globalité.

Par ailleurs, la fouille de données (FD) [23] est un domaine de recherche très actif qui permet de découvrir des connaissances souvent appelés motifs (*patterns*) et dont le nombre peut être très grand et dont une partie s’avère souvent peu pertinente à exploiter pour la prise de décision. Il est donc important de filtrer les résultats de la FD en ne considérant que les motifs jugés pertinents selon des critères établis.

## 1.3 Objectifs de la recherche

Dans le cadre de ce mémoire de recherche, nous allons exploiter, améliorer et valider une nouvelle mesure de pertinence développée par l’équipe du Laboratoire de Recherche sur l’Information Multimédia (LARIM) en vue d’une part, d’identifier les communautés dans les réseaux sociaux à deux modes de données (deux types de noeuds), et d’autre part, de cibler les noeuds les plus importants au sein du réseau et éventuellement dans chacune des communautés identifiées. Pour cela, le mémoire exploite l’analyse formelle de concepts non seulement pour atteindre les objectifs ci-haut mentionnés, mais également pour identifier les concepts formels d’un treillis de concepts les plus sémantiquement pertinents, c.-à-d. les groupes homogènes d’objets (ex. individus) décrits par des attributs et qui génèrent de nouvelles connaissances.

Notre travail est focalisé sur les réseaux sociaux à deux modes et comporte plusieurs

---

chapitres. Le premier chapitre est consacré à la présentation du contexte général de notre étude, de la problématique et des objectifs visés. Dans le chapitre 2, nous présentons d'abord des rappels sur les notions de base, et plus précisément sur l'analyse formelle de concepts (AFC) comme cadre de référence pour le regroupement en lien avec l'analyse des réseaux sociaux (ARS). Ensuite, l'état de l'art sur les réseaux sociaux et la détection des communautés sont fournis. Dans le chapitre 3, on présente la formule révisée de l'indice de pertinence pour ensuite présenter la démarche utilisée pour identifier des communautés et des objets/attributs conceptuellement pertinents au sein de chacune des communautés. Le chapitre 4 présente la partie expérimentale alors que le dernier chapitre est consacré à la conclusion et aux travaux futurs.



# Chapitre 2

## Rappels et état de l'art

Dans ce chapitre, nous allons présenter quelques notions relatives à l'analyse des réseaux sociaux et à l'analyse formelle de concepts (AFC) comme étant une technique de fouille de données [23] par la production de groupes chevauchants et de règles d'association.

### 2.1 Définition et types de réseaux

Le réseau se définit comme une “trame ou une structure composée d'éléments ou de points, souvent qualifiés de nœuds ou de sommets, reliés entre eux par des liens ou liaisons, assurant leur interconnexion ou leur interaction et dont les variations obéissent à certaines règles de fonctionnement [16].”

Les réseaux sont principalement divisés en quatre catégories : les réseaux *sociaux*, comme Twitter, Facebook et Google, les réseaux *d'information* comme les réseaux sémantiques et WordNet, les réseaux de *citations* et les réseaux *technologiques* comme Internet, les circuits électriques et électroniques connectés, les réseaux de télécommunication, les réseaux routiers et ferroviaires, et finalement les réseaux *biologiques* comme les chaînes alimentaires, les espèces animales et les réseaux de neurones [1].

#### 2.1.1 Définition

Dans le domaine de l'analyse des réseaux sociaux, nous distinguons plusieurs types de réseaux, principalement le réseau à un seul mode de données et celui à deux modes de données.

**Définition 1.** *Un réseau à un seul mode comporte alors un seul type de nœuds et un seul type de liens.*

**Définition 2.** Les réseaux de données à deux modes, également appelés réseaux d'affiliation, sont largement utilisés pour représenter deux ensemble de nœuds  $V_1$  et  $V_2$  (ex. personnes et événements) de types distincts et des liens dans l'ensemble  $E$  qui relient un nœud de  $V_1$  avec un autre nœud de  $V_2$  (ex. participation de chercheurs à des événements scientifiques). Ils sont représentés par des graphes bipartis de la forme  $G = (V_1 \cup V_2, E \subseteq V_1 \times V_2)$ .

Tel qu'indiqué dans le chapitre 1, l'accent sera mis sur les réseaux à deux modes.

### 2.1.2 Types des réseaux

Parmi les familles de réseaux, nous pouvons distinguer : les réseaux *homogènes* et les réseaux *hétérogènes* [39].

**Définition 3.** *Réseau hétérogène. Un réseau est appelé hétérogène si le nombre des types nœuds est supérieur à 1 ou celui des types de relations est supérieur à 1.*

Notons aussi que ces réseaux peuvent être orientés ou non. En présence d'une orientation, un lien  $e_1 = (v_i, v_j)$  entre  $v_i$  et  $v_j$  n'implique pas nécessairement l'existence d'un lien  $e_2 = (v_j, v_i)$  et la représentation des liens est exprimée par une flèche. Pour les réseaux *non orientés*, si les nœuds  $v_i$  et  $v_j$  sont connectés, il existe alors un lien entre  $v_i$  et  $v_j$ .

**Définition 4.** *Réseau homogène*

*C'est un réseau qui ne contient que des liens connectant des nœuds ayant le même type. Il s'agit alors d'un réseau à un seul mode de données.*

Un exemple de ces réseaux peut être le réseau médical formé par des médecins reliés entre eux par des liens de collaboration. Bien que ce type de structure est simple et facile à manipuler, son analyse peut prendre différentes formes et présenter des défis. Ainsi, soient  $v_i$  et  $v_j$  deux nœuds du réseau, s'il existe un lien direct entre ces nœuds, alors les nœuds  $v_i$  et  $v_j$  sont dits adjacents, ou encore connectés ou voisins.

## 2.2 Analyse formelle de concepts

L'analyse formelle de concepts [20] est un formalisme de représentation et de découverte de la connaissance, basé sur la formalisation des concepts et la hiérarchie de concepts. Elle est considérée comme une méthode de regroupement conceptuel (*conceptual clustering*) puisqu'elle permet de produire des groupes homogènes à la fois chevauchants (*overlapping clusters*) et conceptuels du fait que chaque groupe représenté par un concept formel est décrit non seulement par son extension, mais également par son intention (sa description). Elle permet également la production de règles d'association.

**Définition 5.** Soit  $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$  un contexte formel où  $\mathcal{G}$ ,  $\mathcal{M}$  et  $\mathcal{I}$  sont respectivement un ensemble d'objets, une collection d'attributs et une relation binaire entre  $\mathcal{G}$  et  $\mathcal{M}$ . L'expression  $(g, m) \in \mathcal{I}$  ou encore  $gIm$  signifie que l'objet  $g$  possède l'attribut  $m$ .

Le tableau 2.1 représente un contexte formel basé sur l'exemple de Dong et *al.* [12], extrait d'une base de données des transactions pour illustrer les réseaux de données à deux modes. Il comporte cinq objets formant l'ensemble  $\mathcal{G} = \{1, 2, 3, 4, 5\}$  et huit attributs différents formant l'ensemble  $\mathcal{M} = \{a, b, c, d, e, g, h, i\}$ . Les objets font référence aux identifiants des transactions alors que les attributs représentent l'ensemble de produits contenus dans la transaction. Ce tableau peut être perçu comme un réseau décrivant par exemple des chercheurs et les conférences qu'ils ont fréquentées au cours de l'année 2010.

Id & Produit	a	b	c	d	e	g	h	i
1	X	X	X	X	X	X	X	X
2	X		X	X		X		
3		X	X	X		X	X	X
4	X	X		X			X	X
5		X	X		X	X	X	X

TABLEAU 2.1 – Exemple de contexte formel tiré de [12]

Dans le tableau 2.1, le contexte formel est représenté sous forme d'un tableau où les objets sont en lignes et les attributs en colonnes. L'intersection des lignes et des colonnes représente la relation binaire  $\mathcal{I}$  entre les objets et les attributs.

**Définition 6.** Un concept formel  $c$  est une paire d'ensembles  $c := (A, B)$  avec  $A \subseteq \mathcal{G}$ ,  $B \subseteq \mathcal{M}$ ,  $A = B'$  et  $B = A'$ , où  $A'$  est l'ensemble des attributs partagés par les objets dans

$A$  et  $B'$  est l'ensemble des objets ayant tous leurs attributs dans  $B$ . Les sous-ensembles  $A$  et  $B$  sont appelés respectivement l'extension et l'intention du concept  $c$ . Les valeurs de  $A'$  et  $B'$  sont obtenues comme suit :  $A' := \{m \in M \mid gIm \forall g \in A\}$  et  $B' := \{g \in G \mid gIm \forall m \in B\}$ .

À partir du contexte de Dong, nous remarquons que le couple  $(\{1, 3, 5\}, \{b, c, g, h, i\})$  forme un concept, où  $(\{1, 3, 5\})$  est son extension et  $(\{b, c, g, h, i\})$ , son intention. En effet,  $\{1, 3, 5\}' = \{b, c, g, h, i\}$  et  $\{b, c, g, h, i\}' = \{1, 3, 5\}$ . Cela signifie que l'ensemble des attributs partagés par les objets dans  $A$ , c'est-à-dire par les objets 1, 3 et 5, correspond exactement à  $\{b, c, g, h, i\}$ , donc  $B = A'$ . De même, l'ensemble des objets possédant les attributs dans  $B$ , c'est-à-dire les attributs  $b, c, g, h$  et  $i$  coïncide avec  $\{1, 3, 5\}$ , et donc  $A = B'$ . Tous les concepts formels de cet exemple sont listés dans le tableau 3.1

No	Concept	Extension	Intention
1	$(\{1\}, \{a, b, c, d, e, g, h, i\})$	$\{1\}$	$\{a, b, c, d, e, g, h, i\}$
2	$(\{1, 2\}, \{a, c, d, g\})$	$\{1, 2\}$	$\{a, c, d, g\}$
3	$(\{1, 3\}, \{b, c, d, g, h, i\})$	$\{1, 3\}$	$\{b, c, d, g, h, i\}$
4	$(\{1, 4\}, \{a, b, d, h, i\})$	$\{1, 4\}$	$\{a, b, d, h, i\}$
5	$(\{1, 5\}, \{b, c, e, g, h, i\})$	$\{1, 5\}$	$\{b, c, e, g, h, i\}$
6	$(\{1, 2, 3\}, \{c, d, g\})$	$\{1, 2, 3\}$	$\{c, d, g\}$
7	$(\{1, 2, 4\}, \{a, d\})$	$\{1, 2, 4\}$	$\{a, d\}$
8	$(\{1, 3, 4\}, \{b, d, h, i\})$	$\{1, 3, 4\}$	$\{b, d, h, i\}$
9	$(\{1, 3, 5\}, \{b, c, g, h, i\})$	$\{1, 3, 5\}$	$\{b, c, g, h, i\}$
10	$(\{1, 2, 3, 4\}, \{d\})$	$\{1, 2, 3, 4\}$	$\{d\}$
11	$(\{1, 2, 3, 5\}, \{c, g\})$	$\{1, 2, 3, 5\}$	$\{c, g\}$
12	$(\{1, 3, 4, 5\}, \{b, h, i\})$	$\{1, 3, 4, 5\}$	$\{b, h, i\}$
13	$(\{1, 2, 3, 4, 5\}, \{\})$	$\{1, 2, 3, 4, 5\}$	$\{\}$

TABLEAU 2.2 – Les concepts formels avec leurs extensions et intentions

Un ordre partiel existe entre deux concepts  $c_1 = (A_1, B_1)$  et  $c_2 = (A_2, B_2)$  de la forme  $c_1 \preceq c_2$  si  $A_1 \subseteq A_2 \iff B_1 \supseteq B_2$ . Dans un tel cas,  $c_1$  est appelé surconcept ou successeur de  $c_2$  alors que  $c_2$  est appelé sous-concept ou prédécesseur de  $c_1$ . Lorsque la relation d'ordre est stricte entre les deux concepts,  $c_1$  est appelé successeur immédiat (*upper cover*) de  $c_2$  alors que ce dernier est prédécesseur immédiat (*lower cover*) de  $c_1$ .

L'ensemble  $\mathcal{C}$  de tous les concepts ainsi que l'ordre partiel forment un treillis de concepts [20] noté  $\mathfrak{B}(G, M, I)$ .

Plusieurs algorithmes et outils permettent de générer les concepts ainsi que leur treillis à partir d'un contexte, tels que [32] : Bordat, Next Closure, Close by One, Lindig, Chein, Nourine, Norris, Godin, et Dowling. Dans notre projet, nous avons utilisé Lattice Miner [34], un outil développé par le Laboratoire de Recherche sur l'Information Multimédia (LARIM) de l'UQO. La figure suivante montre le treillis basé sur l'exemple de Dong et *al.* [12]

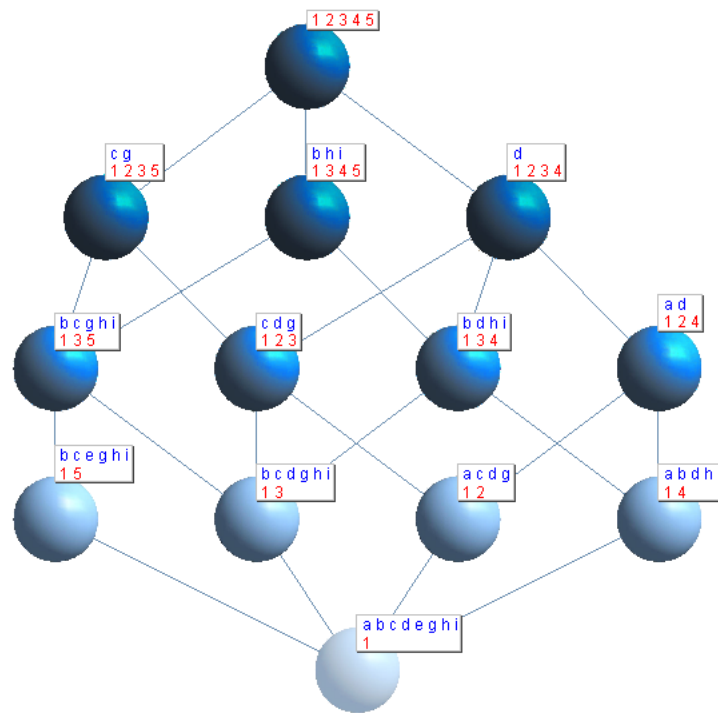


FIGURE 2.1 – Exemple de treillis de concepts pour le contexte du tableau 2.1

Les treillis de concepts formels (ou treillis de Galois) sont une structure mathématique permettant de représenter les classes non disjointes sous-jacentes à un ensemble d'objets (exemples, instances, tuples ou observations) décrits à partir d'un ensemble d'attributs (propriétés, descripteurs ou items). Ces classes non disjointes sont aussi appelées concepts formels, hyper-rectangles ou ensembles fermés. Une classe matérialise un concept (à savoir une idée générale que l'on a d'un objet). Ce concept peut être défini formellement par une extension (exemples du concept) ou par une intension (abstraction du concept) [37].

Dans un treillis de concepts, les concepts sont représentés par les nœuds alors que les arcs représentent une relation d'ordre, la longueur des arcs n'ayant aucune signification particulière. La figure 2.1 présente un treillis à étiquetage complet (*full labeling*), où les concepts sont affichés avec leur extension et intention. D'autres outils permettent de faire un étiquetage réduit (*reduced labeling*) où un objet est indiqué uniquement la première fois qu'il est rencontré à partir du bas du treillis, et dualement, un attribut est indiqué uniquement la première fois qu'il est rencontré à partir du haut du treillis [1].

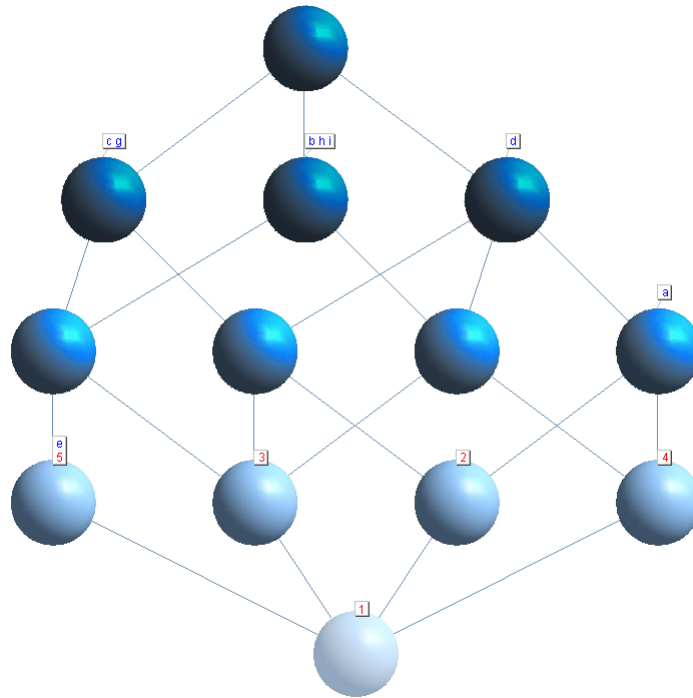


FIGURE 2.2 – Exemple de treillis de concepts en étiquetage réduit pour le contexte du tableau 2.1

À titre d'exemple, le noeud du treillis de la figure 2.1 dont l'étiquette est  $(5, e)$  représente en fait le concept formel  $(\{1, 5\}, \{b, c, e, g, h, i\})$  indiquant que les objets 1 et 5 ont en commun les attributs b, c, e, g, h et i. En outre, il s'agit d'un concept dit à la fois objet et attribut pour l'objet 5 et l'attribut e indiquant qu'il est le plus petit concept contenant dans son extension l'objet 5 et le plus grand concept contenant dans son intention l'attribut e.

**Définition 7.** *Un générateur minimal  $h$  est un sous-ensemble minimal de l'intention  $B$  tel que  $h'' = B$ .*

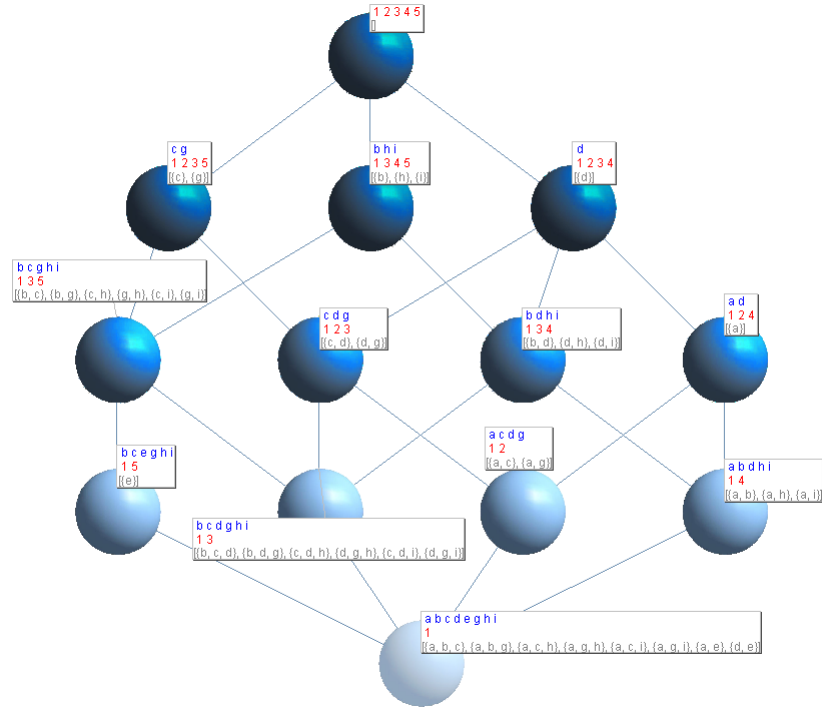


FIGURE 2.3 – Exemple de treillis de concepts avec les générateurs minimaux pour le contexte du tableau 2.1

Sur cette figure générée à partir de l'outil Lattice Miner, nous remarquons la présence de tous les générateurs minimaux des concepts formels. Comme exemple, le concept formel  $(\{1, 3, 4, 5\}, \{b, h, i\})$  a trois générateurs minimaux dans l'ensemble  $\{\{b\}, \{h\}, \{i\}\}$ .

**Définition 8.** Les attributs  $m_1$  et  $m_2$  sont dits équivalents dans l'intention  $B$  du concept  $c := (A, B)$  si  $|m_1''| > 1$  et  $|m_2''| > 1$  et  $m_1' = m_2' = A$ . Autrement dit,  $m_1$  et  $m_2$  sont équivalents si le concept attribut pour ces deux attributs est le même, c.-à-d.  $c = \mu(m_1) = \mu(m_2) = (m_1', m_1'')$ . Les attributs équivalents sont donc au moins deux attributs qui apparaissent pour la première fois dans un concept en partant du supremum du treillis et dont la présence de l'un implique nécessairement la présence de l'autre.

## 2.3 Les mesures des concepts formels

Les concepts formels jouent un rôle important dans la découverte des connaissances. Toutefois, ils peuvent être démesurément nombreux. Récemment, plusieurs approches ont été introduites pour traiter le problème : soit par la sélection des concepts avec des extensions dépassant un certain seuil [28] ou par la construction d'une sous structure dite *Iceberg lattice* [40] ou encore par la détection des concepts plus pertinents.

Une idée qui devient de plus en plus omniprésente est d'évaluer la pertinence des concepts au moyen d'indices d'intérêt comme : la stabilité (*Stability*) [29], la robustesse (*Robustness*) [41], la probabilité du concept (*Concept Probability*) [27], la séparation (*Separation*) [27], la fréquence (*Frequency*) [33], etc. Dans ce que suit, nous rappelons la notion de l'indice de stabilité et de séparation qui nous intéressent en particulier.

### 2.3.1 Stabilité

Bien que plusieurs mesures d'intérêt aient été introduites pour sélectionner les concepts pertinents, l'indice de stabilité s'est avéré être le plus approprié [31].

La stabilité intentionnelle  $\sigma(c)$  du concept  $c = (A, B)$  mesure la force de dépendance entre l'intention  $B$  et les objets de l'extension  $A$ . Plus précisément, elle exprime la probabilité de maintenir  $B$  fermé lorsqu'un sous-ensemble d'objets dans  $A$  est supprimé avec une même probabilité. Aussi, un concept stable résiste au bruit et ne s'effondre pas lorsque certains objets sont supprimés de son extension.

$\sigma(c)$  est définie comme suit :

$$\sigma(c) = \frac{|\{e \in \mathcal{P}(A) | e' = B\}|}{2^{|A|}} \quad (2.1)$$

En considérant les communautés, la stabilité aide à identifier les concepts pertinents qui représentent des groupes cohésifs.

Pour le concept formel  $(\{1, 5\}, \{b, c, e, g, h, i\})$ , l'indice de stabilité équivaut à  $\sigma(c) = \frac{1}{2^2} = 0.25$

Dans l'article [24], les auteurs ont proposé une mesure qui raffine l'index de stabilité. Il s'agit de l'indice de pertinence des concepts (*Conceptual Relevance Index*).

L'indice de pertinence des concepts basé sur l'intention d'un concept  $c = (A, B)$  dans un contexte formel  $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$  peut être calculé comme suit :



$$\mathcal{CR}_{in}(c) = \mathcal{F}(\alpha_{in}(c), \beta_{in}(c)) \quad (2.2)$$

avec

$$\alpha_{in}(c) = \begin{cases} \frac{|\{m \in B \mid (B \setminus \{m\})' \neq A\}|}{|B|} & \text{si } B \neq \emptyset \\ 0 & \text{sinon.} \end{cases} \quad (2.3)$$

et

$$\beta_{in}(c) = \begin{cases} \frac{|\mathcal{H}_c|}{2^{|B|-2}} & \text{si } |\mathcal{H}_c| > 1 \text{ et } |B| > 1 \\ 0 & \text{sinon.} \end{cases} \quad (2.4)$$

La fonction  $\mathcal{F}$  peut être n'importe quelle fonction d'activation (minimum, maximum ou moyenne...) ou une relation non linéaire entre les termes  $\alpha_{in}(c)$  et  $\beta_{in}(c)$ . Elle offre une flexibilité pour adapter la formule de l'indice de pertinence des concepts à trouver la variante appropriée qui correspond à chaque problème spécifique.

Pour le calcul de  $\alpha_{in}(c)$ , nous faisons une itération à travers les attributs de l'intention  $B$  pour compter le nombre de ceux qui sont conceptuellement pertinents et qui satisfont la condition de dérivation  $(B \setminus \{m\})' \neq A$ ). La taille de l'intention  $|B|$  est utilisée pour normaliser le terme.  $\alpha_{in}(c)$  quantifie le nombre d'attributs conceptuellement pertinents dans l'intention  $B$  du concept.

Pour le concept formel  $(\{1, 5\}, \{b, c, e, g, h, i\})$ , nous avons  $\alpha_{in}(c) = \frac{0}{6} = 0$ , tous les attributs n'étant pas pertinents.

Dans le terme  $\beta_{in}(c)$ , nous comptons le nombre de générateurs minimaux  $\mathcal{H}_c$  parmi tous générateurs possibles de l'intention  $B$ . Le fait de soustraire 2 au dénominateur correspond à l'exclusion de l'ensemble vide et l'intention elle-même. Il quantifie le ratio d'informations locales pertinentes au sein d'un concept pouvant aboutir à des schémas potentiellement pertinents. Il peut également être considéré comme une version améliorée de la stabilité car il utilise les générateurs minimaux. Pour calculer l'ensemble des générateurs minimaux, nous appliquons la procédure *Mingen()* qui est donnée par l'algorithme 1.

Pour le concept formel  $(\{1, 5\}, \{b, c, e, g, h, i\})$ ,  $\beta_{in}(c) = 0$  puisque  $|\mathcal{H}_c| = 1$ . Dans le chapitre 3, nous allons apporter une amélioration à ces équations.

---

**Algorithme 1** *Minigen()* : procédure de calcul des générateurs minimaux

---

**Input:** Concept  $c = (A, B)$ , ensemble des successeurs  $\mathcal{U}(c)$  du concept  $c$ .

**Output:** Ensemble des générateurs minimaux  $\mathcal{H}_i$

```

1:  $\mathcal{H}_i \leftarrow \emptyset$ ;
2: pour chaque  $c_u = (A_u, B_u)$  dans  $\mathcal{U}(c)$  faire
3:    $f_u \leftarrow B \setminus B_u$ ;
4:   si  $\mathcal{H}_i == \emptyset$  alors
5:      $\mathcal{H}_i \leftarrow \{a \mid \forall a \in f_u\}$ ;
6:   sinon
7:      $\text{Gen} \leftarrow \emptyset$ ;
8:     pour chaque  $h_i$  dans  $\mathcal{H}_i$  faire
9:       si  $h_i \cap f_u == \emptyset$  alors
10:         $\text{Gen} \leftarrow (\text{Gen} \cup \{h_i \cup a \mid \forall a \in f_u\})$ ;
11:       sinon
12:         $\text{Gen} \leftarrow (\text{Gen} \cup \{h_i\})$ ;
13:       fin si
14:     fin pour
15:      $\mathcal{H}_i \leftarrow \text{minimal}(\text{Gen})$ ;
16:   fin si
17: fin pour
18: Retourner  $\mathcal{H}_i$ ;

```

---

### 2.3.2 Séparation

La séparation  $\gamma(c)$  d'un concept formel  $c = (A, B)$  [27] est calculée comme suit :

$$\gamma(c) = \frac{|A| \times |B|}{\sum_{g \in A} |g'| + \sum_{m \in B} |m'| - |A| \times |B|} \quad (2.5)$$

où  $g'$  est l'ensemble des attributs de l'objet  $g$  de  $A$  et  $m'$  est l'ensemble des objets associés à l'attribut  $m$  dans  $B$ . L'indice de séparation est utilisé pour évaluer la qualité des attributs ou des caractéristiques d'un concept ou d'un groupe d'objets, en mettant l'accent sur leur spécificité par rapport au contexte plus large dans lequel ils sont considérés. Il peut être utilisé en combinaison avec d'autres mesures ou techniques d'analyse

de données pour prendre des décisions ou obtenir des informations supplémentaires sur les relations et les caractéristiques des objets étudiés.

Un indice de séparation élevé indique que les attributs associés à un concept sont relativement exclusifs et rares, ce qui signifie qu'ils ne sont pas souvent partagés avec d'autres concepts ou groupes d'objets. En revanche, un indice de séparation faible suggère que les attributs associés au concept sont plus courants et peuvent être partagés avec d'autres concepts ou groupes d'objets.

Pour le concept formel  $(\{1, 5\}, \{b, c, e, g, h, i\})$ , la séparation équivaut à :

$$\gamma(c) = \frac{12}{14+22-12} = 0.50$$

Pour notre exemple illustratif, le calcul des indices de séparation se résume dans le tableau suivant.

No	Concept(c)	$\gamma(c)$
1	$(\{1\}, \{a, b, c, d, e, g, h, i\})$	0.28
2	$(\{1, 2\}, \{a, c, d, g\})$	0.42
3	$(\{1, 3\}, \{b, c, d, g, h, i\})$	0.46
4	$(\{1, 4\}, \{a, b, d, h, i\})$	0.45
5	$(\{1, 5\}, \{b, c, e, g, h, i\})$	0.50
6	$(\{1, 2, 3\}, \{c, d, g\})$	0.43
7	$(\{1, 2, 4\}, \{a, d\})$	0.33
8	$(\{1, 3, 4\}, \{b, d, h, i\})$	0.52
9	$(\{1, 3, 5\}, \{b, c, g, h, i\})$	0.60
10	$(\{1, 2, 3, 4\}, \{d\})$	0.17
11	$(\{1, 2, 3, 5\}, \{c, g\})$	0.33
12	$(\{1, 3, 4, 5\}, \{b, h, i\})$	0.48
13	$(\{1, 2, 3, 4, 5\}, \{\})$	0.00

TABLEAU 2.3 – Les résultats du calcul des indices de séparation  $\gamma(c)$

## 2.4 Détection de nœuds centraux et de communautés

Dans cette partie du document, nous allons exposer quelques méthodes de détection des communautés dans les réseaux à deux modes.

Avant d'entamer la détection des communautés, rappelons d'abord quelques notions de base sur la théorie des graphes.

L'analyse des réseaux sociaux est une confluence de plusieurs théories et dont la théorie des graphes. Un graphe  $G$  est défini comme un couple  $G = (V, E)$  composé d'un ensemble de sommets  $V$  (*vertices*) et d'un ensemble d'arêtes  $E$  (*edges*) reliant certaines paires de sommets de  $V$ . Formellement,  $E \subseteq V \times V$ . Nous considérons le cas de graphes finis, c'est-à-dire ayant un nombre fini de sommets et d'arêtes et notons  $n = |V|$  et  $p = |E|$ . Les sommets sont les objets en relation dans le graphe. On les désigne aussi par les termes nœuds et acteurs. Les arêtes décrivent les relations entre les sommets. Notons aussi que dans un graphe, les arêtes peuvent être valuées. Plus cette valeur est élevée, plus le lien entre les sommets est intense. Il est alors question de graphe valué. Dans le présent document, nous considérons uniquement les graphes non valués.

La figure ci-dessous donne une visualisation de notre exemple de réseau dans le logiciel UCINET [6]. Il s'agit d'un réseau de données à deux modes représenté sous forme de graphe biparti.

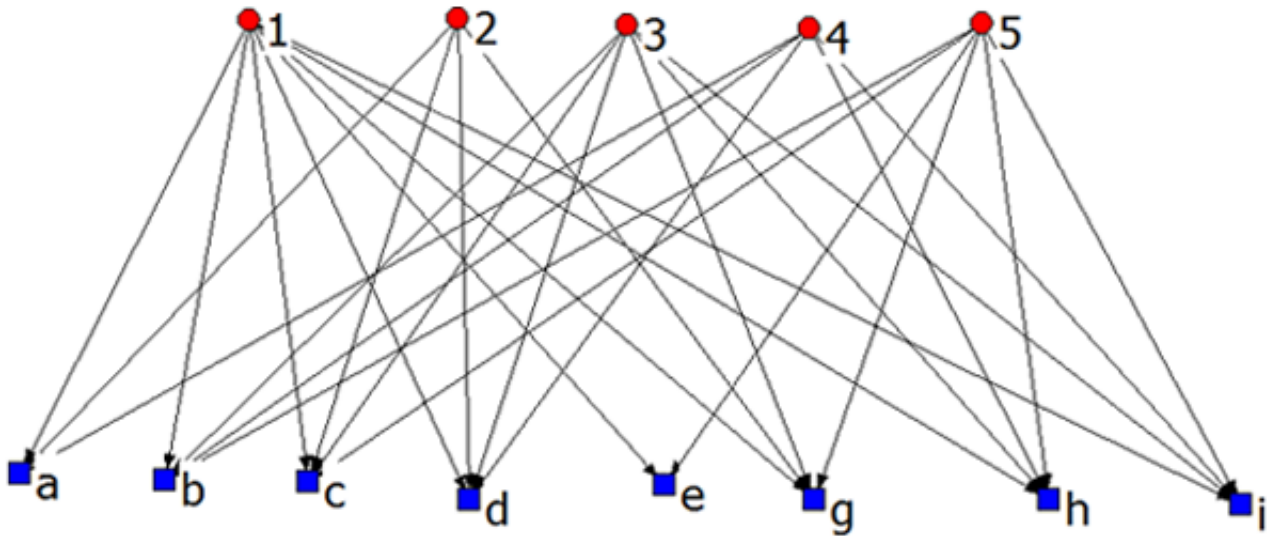


FIGURE 2.4 – Visualisation du graphe dans le logiciel UCINET

En théorie des graphes, un graphe est dit biparti si son ensemble de sommets peut être divisé en deux sous-ensembles disjoints  $U$  et  $V$  tels que chaque arête a une extrémité dans  $U$  et l'autre dans  $V$ . Un graphe biparti permet notamment de représenter une

relation binaire. Notons qu'il ne peut y avoir de relation entre les éléments d'un même ensemble.

Dans le cas de notre exemple [12], l'ensemble  $U$  représente les transactions alors que l'ensemble  $V$  est celui des produits. Le lien entre ses deux ensembles est décrit comme "un produit est contenu dans une transaction".

Une arête est incidente à un sommet si ce sommet constitue au moins une de ses extrémités et un graphe est dit complet si tous ses sommets sont adjacents deux à deux.

**Définition 9.** *Partant d'un nœud, la distance géodésique désigne le plus petit nombre de liens à traverser pour arriver à un second nœud. Dans la figure ci-dessous, la distance géodésique entre les nœuds 4 et 5 est égale à 2.*

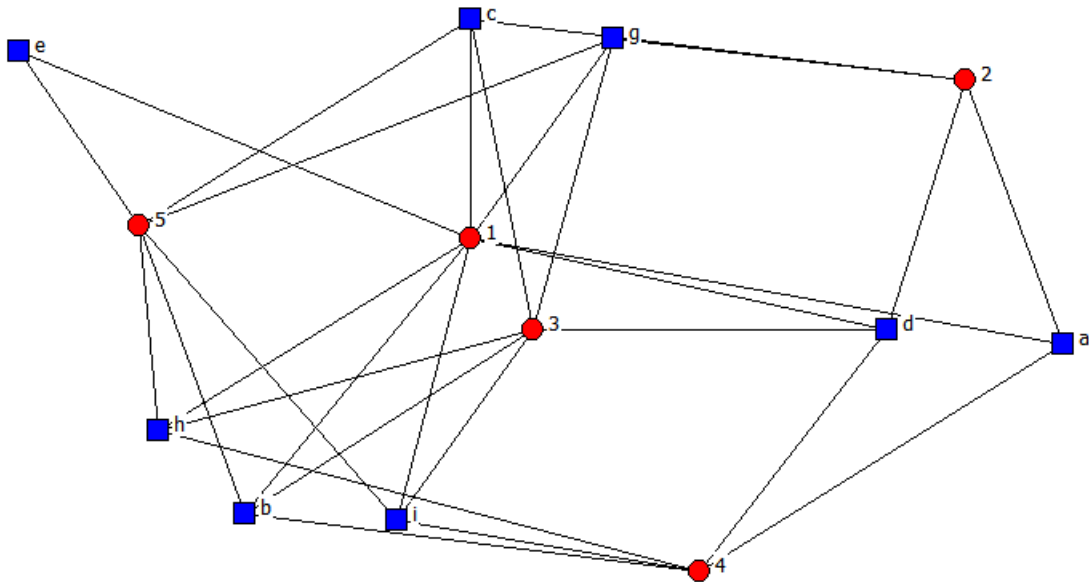


FIGURE 2.5 – Exemple de distance géodésique

**Définition 10.** *Le degré d'un sommet correspond au nombre de sommets auquel il est relié. Dans un graphe simple d'ordre  $n$ , le degré d'un sommet est compris entre 0 et  $n - 1$ .*

**Définition 11.** *La densité d'un réseau est une mesure de sa connectivité (le nombre de liens ou d'arêtes). Plus un graphe est dense, plus il est connecté. La densité indique la quantité de liens au sein d'un réseau et permet de définir la cohésion d'un réseau social.*

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

$$\Delta = \frac{2p}{n(n-1)} \quad (2.6)$$

où  $p$  est le nombre de liens et  $n(n-1)/2$  est le nombre maximal de connexions possibles pour un réseau de  $n$  nœuds.

### 2.4.1 Méthodes d'analyse des réseaux sociaux

De nos jours, les réseaux sociaux stockent de grands volumes d'information grâce aux échanges quotidiens de données entre les acteurs. De telles données peuvent nous fournir de la connaissance. Pour y arriver, une analyse de ces données s'avère nécessaire en passant par une modélisation sous forme de graphe.

**Définition 12.** *Une mesure de centralité d'un nœud décrit son niveau d'importance au sein d'un réseau social. Les acteurs centraux sont alors ceux qui sont liés et impliqués dans plusieurs relations avec les autres acteurs d'une façon extensive. Bien que de nombreuses métriques de centralité aient été proposées, le degré, la proximité et l'intermédiarité se sont avérés les plus remarquables dans une variété d'applications, et sont donc largement utilisés.*

La centralité de degré d'un nœud dans un réseau de graphes à deux modes est définie comme suit :

$$D_c(u_i) = \sum_{v_j \in \mathcal{M}} I_{ij}, \forall u_i \in \mathcal{G}, \quad (2.7)$$

$$D_c(v_j) = \sum_{u_i \in \mathcal{G}} I_{ij}, \forall v_j \in \mathcal{M} \quad (2.8)$$

où  $I_{ij}$  vaut 1 lorsqu'il existe un lien entre  $u_i$  et  $v_j$ , et 0 sinon. Ainsi, la sommation dans l'équation (2.7) représente le nombre d'arêtes (ou de liens avec d'autres types de nœuds voisins) impliquant le nœud.

**Définition 13.** *Centralité de proximité*

C'est une catégorie d'algorithmes qui évaluent la centralité d'un nœud par sa proximité (distance) par rapport aux autres nœuds. Elle mesure le nombre d'étapes ou de liens nécessaires à un acteur particulier pour accéder à tous les autres acteurs du réseau.

Dans un réseau de graphes à deux modes, la centralité de proximité normalisée d'un nœud ( $u_i$  ou  $v_j$ ) de l'un des deux types est définie par :

$$C_c(u_i) = \frac{|\mathcal{M}| + 2(|\mathcal{G}| - 1)}{\sum_{v_j \in \mathcal{M}} d(u_i, v_j)}, \forall u_i \in \mathcal{G}, \quad (2.9)$$

$$C_c(v_j) = \frac{|\mathcal{G}| + 2(|\mathcal{M}| - 1)}{\sum_{u_i \in \mathcal{G}} d(u_i, v_j)}, \forall v_j \in \mathcal{M} \quad (2.10)$$

où  $d(u_i, v_j)$  est la distance géodésique (le plus court chemin) entre les nœuds  $u_i$  et  $v_j$ .

**Définition 14.** *Centralité d'intermédiarité [9]*

Elle mesure l'utilité du sommet dans la transmission de l'information au sein du réseau. Le sommet joue un rôle central si beaucoup de plus courts chemins entre deux sommets doivent emprunter ce sommet.

Dans les graphes bipartis, la centralité d'intermédiarité normalisée d'un nœud est définie comme dans [7] :

$$B_c(u_i) = \sum_{u_j \neq u_k \neq u_i, u_j, u_k, u_i \in \mathcal{G}} \frac{\sigma_{u_j u_k}(u_i)}{\sigma_{u_j u_k}}, \forall u_i \in \mathcal{G}, \quad (2.11)$$

$$B_c(v_j) = \sum_{v_j \neq v_k \neq v_i, v_j, v_k, v_i \in \mathcal{M}} \frac{\sigma_{v_i v_k}(v_j)}{\sigma_{v_i v_k}}, \forall v_j \in \mathcal{M}, \quad (2.12)$$

où  $\sigma_{u_j u_k}(u_i)$  désigne le nombre total de chemins les plus courts entre les nœuds  $u_j$  et  $u_k$  et qui traversent  $u_i$ . Pour normaliser l'intermédiarité, nous divisons simplement  $B_c(u_i)$  (respectivement  $B_c(v_j)$ ) par le terme correspondant à son ensemble de nœuds  $B_c(\mathcal{G})$  (respectivement  $B_c(\mathcal{M})$ ) [7] :

$$B_c(\mathcal{G}) = \frac{1}{2} [|\mathcal{M}|^2(s+1)^2 + |\mathcal{M}|(s+1)(2t-s-1) - t(2s-t+3)], \forall u_i \in \mathcal{G}, \quad (2.13)$$

où  $s = (|\mathcal{G}| - 1 \mid \text{div } |\mathcal{M}|)$  et  $t = (|\mathcal{G}| - 1 \mid \text{mod } |\mathcal{M}|)$ ,

$$B_c(\mathcal{M}) = \frac{1}{2} [|\mathcal{G}|^2(p+1)^2 + |\mathcal{G}|(p+1)(2r-p-1) - r(2p-p+3)], \forall v_j \in \mathcal{M}, \quad (2.14)$$

où  $p = (|\mathcal{M} - 1| \operatorname{div} |\mathcal{G}|)$  et  $r = (|\mathcal{M} - 1| \operatorname{mod} |\mathcal{G}|)$

Dans [25], les auteurs ont défini une nouvelle mesure de centralité appelée *Bi-face* pour identifier les nœuds importants dans les réseaux sociaux à deux modes en utilisant l'analyse formelle de concepts. Cette mesure exploite les faces du concept (c.-à-d. la différence entre l'intention du concept et celle de ses successeurs) pour détecter les nœuds qui ont une connectivité biclique influente<sup>1</sup> et ne sont pas situés dans des ponts<sup>2</sup>. Contrairement aux indices de centralité connus, elle quantifie la façon dont un nœud a une influence sur ses nœuds voisins via les cliques. Des expériences sur une variété de réseaux réels et synthétiques montrent que cette mesure est plus performante que plusieurs mesures de centralité. Les résultats montrent également que la présence de nœuds terminaux, de ponts influents, et le chevauchement des cliques clés a un impact à la fois sur les performances et le comportement de la mesure ainsi que sur sa relation avec d'autres mesures de centralité traditionnelles. Sur les jeux de données testés, le calcul de *Bi-face* est d'au moins vingt-trois fois plus rapide que l'intermédiarité, onze fois plus rapide que la percolation, neuf fois plus rapide que le vecteur propre, et dix fois plus rapide que la proximité.

### 2.4.2 Détection des communautés

L'analyse des réseaux sociaux (ARS) avec l'approche basée sur la théorie des graphes a beaucoup contribué à la détection des communautés à l'intérieur d'un réseau. Cependant, la notion de communautés dans les graphes n'a pas de définition formelle. Sémantiquement, une communauté est un ensemble de nœuds qui partagent les mêmes centres d'intérêt ou qui ont des profils similaires. Une communauté est donc un ensemble de nœuds fortement liés entre eux et faiblement liés avec les autres nœuds du graphe [22, 15].

---

1. La connectivité biclique influente permet de quantifier à quel point certains nœuds sont importants dans un réseau en termes de leur capacité à influencer les autres nœuds. Cela peut être utilisé pour identifier les nœuds clés dans un réseau, qui, s'ils sont supprimés, peuvent perturber la propagation de l'information ou l'efficacité du réseau

2. Un pont est une arête ou un lien dont la suppression entraîne la division du réseau en deux composantes connexes distinctes.



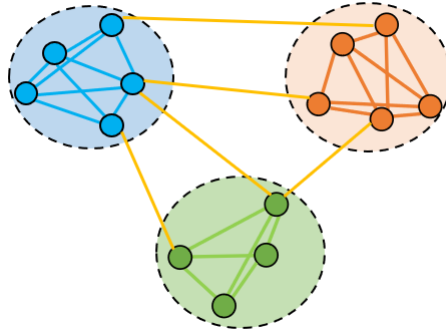


FIGURE 2.6 – Exemple de trois communautés délimitées par des cercles en pointillé

La détection des communautés peut être perçue comme une opération de regroupement (*clustering*) des nœuds en plusieurs entités tout en respectant les relations existant entre eux dans le graphe. Nous pouvons trouver des exemples de communautés dans la vie réelle telles que : la communauté des chercheurs d'université, celle des lecteurs de revues particulières ou encore celles des médecins ayant de l'expertise dans le traitement de certaines maladies. Dans notre exemple, les communautés sont des transactions regroupant en leur sein des produits similaires ou qui peuvent être décrites par les caractéristiques des produits qu'elles contiennent.

### 2.4.3 Algorithmes de détection des communautés

De nos jours, plusieurs chercheurs ont travaillé sur la catégorisation des méthodes proposées pour la détection des communautés dans les réseaux sociaux. La plupart de ces catégorisations sont basées sur le type des algorithmes de détection des communautés et leurs principes méthodologiques. Fortunato [14] a mené une étude exhaustive et classé les méthodes en huit catégories suivantes :

#### Méthodes traditionnelles

Elles se basent sur le partitionnement optimal des graphes représentant les réseaux sociaux en  $k$  groupes ou "clusters",  $k$  étant connu. Dans cette catégorie de méthodes, la méthode la plus connue est l'algorithme de Kernighan-Lin [26]. Cette méthode se base sur la recherche des communautés ou des partitions souvent de même taille, ce qui s'avère difficile dans le cas des réseaux complexes. Cette contrainte étant trop stricte et difficile à atteindre dans des cas réels, elle a été relâchée au profit de la méthode de recherche (le

---

partitionnement hiérarchique) des communautés sans toutefois à avoir à préciser la taille exacte [19]. Le partitionnement hiérarchique se base alors sur une fonction de similarité de telle sorte que les partitions contiennent des nœuds avec des similarités fortes.

### **Algorithmes décisifs**

Ils sont basés sur la recherche d'une propriété des liens intercommunautaires pour pouvoir les identifier afin de les éliminer. Ces éliminations déconnectent le graphe pour former des composantes connexes qui représentent les communautés. L'algorithme le plus populaire est celui proposé par Girvan et Newman [36].

### **Méthodes basées sur la modularité**

La modularité, proposée par Girvan et Newman, est utilisée par plusieurs algorithmes comme fonction de qualité en l'optimisant ou la modifiant telles que les techniques gloutonnes et d'autres fonctions d'optimisation [2]. Cependant, il a été démontré que trouver la partition optimale en modularité est un problème NP-complet. D'autres algorithmes, comme celui très connu de Louvain, sont basés sur des techniques gloutonnes pour apporter des temps de calcul acceptables.

### **Algorithmes de partitionnement spectral**

Ces algorithmes sont basés sur la notion du spectre définissant la proximité entre les nœuds. Les vecteurs propres, agissant comme des propagateurs de temps pour le processus de marche aléatoire dans le graphe du réseau social, associés aux valeurs propres les plus faibles décrivent les groupes à similarité interne forte [4]. Généralement la matrice Laplacienne est utilisée comme matrice de similarité.

### **Algorithmes dynamiques**

Ils simulent des processus dynamiques où les particules s'influencent entre elles. Ainsi, les particules proches les unes des autres ont tendance à partager le même état. Parmi les processus appliqués aux graphes des réseaux sociaux, on peut citer Spin-Spin et la synchronisation où le système unifie progressivement tous ses éléments au même état [5].

### **Méthodes basées sur l'inférence statistique**

Ces méthodes supposent que le graphe a été généré suivant un modèle admettant l'appartenance des nœuds aux communautés comme des paramètres. Le but est alors d'inférer les paramètres qui auraient généré les observations trouvées avec la probabilité la plus élevée.

### **Méthodes d'extraction des communautés chevauchantes**

Il est évident qu'un nœud peut appartenir à plusieurs groupes ou communautés. La première méthode prenant efficacement le chevauchement entre groupes a été proposée en 2005 par Palla [38]. Par la suite, d'autres approches ont été proposées.

### **Méthodes multi-résolution**

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

Dans ce chapitre, nous avons présenté les notions de base de l'analyse des réseaux sociaux, de l'analyse formelle de concepts (AFC) et quelques méthodes de détermination des communautés dans les réseaux sociaux. Ces notions nous ont permis de comprendre l'état de l'art de notre sujet.

Dans ce chapitre, nous avons rappelé l'état de l'art du domaine d'étude, surtout les définitions relatives aux réseaux, à l'analyse formelle de concepts, aux mesures des concepts formels et aux algorithmes de détection des communautés dans les réseaux sociaux. Dans le chapitre suivant, nous allons présenter les améliorations apportées à l'indice de pertinence des concepts qui est le point de départ de nos travaux et notre approche fondée sur les concepts et qui permet la détection des objets (individus) centraux et les communautés.

# Chapitre 3

## Approche proposée

Dans le chapitre précédent, nous avons exposé les éléments de base de l'analyse formelle de concepts, l'état de l'art sur la détection des communautés dans les réseaux sociaux et les mesures de pertinence des concepts formels. Et comme nous l'avions annoncé dans l'introduction de ce travail, notre objectif est d'une part, d'exploiter, d'améliorer et de valider une nouvelle mesure de pertinence de concepts formels décrite dans [24], en vue de mettre en évidence des communautés dans des réseaux sociaux à deux modes, et d'autre part, de cibler les nœuds les plus importants au sein du réseau dans sa globalité et plus tard, à l'intérieur de chaque communauté générée. Pour la détection des communautés dans les réseaux sociaux à deux modes, nous avons couplé la mesure améliorée de l'indice de pertinence avec l'indice de séparation.

Notre démarche dans la détection des nœuds centraux et des communautés dans des réseaux sociaux à deux modes se compose des éléments suivants. Dans un premier temps, nous construisons le contexte formel ou la matrice d'adjacence du réseau social pour produire le treillis de concepts. Ensuite, nous calculons pour chaque concept du treillis (sauf le supremum et l'infimum), l'indice modifié de pertinence des concepts en faisant usage des équations 3.2 et 3.3.

Pour les nœuds centraux du réseau (et non pas du treillis), nous cherchons d'abord les concepts pertinents en parcourant tous les concepts, classés par ordre décroissant de l'indice de pertinence, jusqu'à ce que les objets apparaissant dans ces concepts couvrent l'ensemble des objets du contexte. En fonction de leur fréquence d'apparition dans l'extension des concepts retenus, nous déterminons les objets centraux qui sont en réalité les acteurs (membres d'une population ou d'une communauté) et plus précisément les objets du contexte formel (*i.e.*, un des deux types du réseau à deux modes).

Quant à la détection des communautés, nous faisons appel à la moyenne harmonique de l'indice de séparation et de l'indice de pertinence pour obtenir l'autonomie des concepts. Cette autonomie est classée par ordre décroissant et un second parcours est fait pour recenser les concepts pertinents pouvant former de potentielles communautés. Nous détaillerons plus tard cette procédure. Rappelons que toute notre démarche concerne les réseaux sociaux à deux types de nœuds où le type-1 représente l'ensemble des objets et le type-2, l'ensemble des attributs.

### 3.1 Calcul de la pertinence des concepts

Nous rappelons que la formule améliorée de l'indice de pertinence des concepts se base sur les deux composantes  $\alpha$  et  $\beta$  de l'indice de pertinence des concepts ( $\mathcal{CR}$ ). Cette mesure peut être calculée en termes d'intention (description) ou d'extension (instances) des concepts. Nous présentons ci-après la formulation de la partie extension qui met l'accent sur les attributs des objets et les associations entre ces attributs.

Nous avons amélioré le numérateur du terme  $\alpha_{ex}(c)$  pour y inclure, entre autres, les attributs équivalents (c'est-à-dire au moins deux attributs qui apparaissent pour la première fois dans un concept en partant du supremum du treillis et donc dont la présence de l'un implique nécessairement la présence de l'autre) car ils sont tout aussi pertinents que ceux définis dans la formule initiale (cf. partie en bleu dans le numérateur du 2ème cas). Quant au terme  $\beta_{ex}(c)$ , il est une adaptation de l'indice de stabilité, en tenant compte au numérateur de l'ensemble des générateurs minimaux au lieu de l'ensemble (plus grand) des sous-ensembles de l'intention  $B$ , et au dénominateur de la taille de l'intention au lieu de  $2^{|B|}$ .

$$\mathcal{CR}_{ex}(c) = (\alpha_{ex}(c) + \beta_{ex}(c))/2 \quad (3.1)$$

Avec

$$\alpha_{ex}(c) = \begin{cases} N_1 = \frac{|\{m \in B \text{ tel que } (B \setminus \{m\})' \neq A\}|}{|B|} & \text{si } B \neq \emptyset \\ \frac{|\{m \in B \text{ tel que } |m''| > 1 \text{ et } m' = A\}|}{|B|} & \text{si } N_1 = 0 \text{ et } B \neq \emptyset \\ 0 & \text{sinon} \end{cases} \quad (3.2)$$

Et

$$\beta_{ex}(c) = \begin{cases} \text{Min}(\frac{|\mathcal{H}_i|}{|B|}, 1) & \text{si } |\mathcal{H}_i| > 1 \\ \frac{1}{|B|} & \text{si } |\mathcal{H}_i| = 1 \text{ et } |gen| < |B| \text{ avec } gen \in \mathcal{H}_i \\ 0 & \text{sinon} \end{cases} \quad (3.3)$$

Il est important de noter que si l'intention  $B$  est telle que  $B_1 \cup B_2 \subseteq B$  où  $B_1$  correspond aux attributs conceptuellement pertinents vérifiant la condition  $B \setminus \{m\}' \neq A$  et  $B_2$  représente des attributs équivalents (cf. la définition 8), alors la valeur de  $N_1$  dans le calcul de  $\alpha_{ex}(c)$  va être nécessairement nulle puisque tout attribut  $m \in B_1$  ne vérifie pas la condition  $B \setminus \{m\}' \neq A$  en raison de la présence d'au moins deux attributs équivalents. Une valeur nulle de  $N_1$  nécessite l'utilisation du 2ème cas pour vérifier la présence d'attributs équivalents sûrement non détectés par la condition ci-haut mentionnée. Si par contre, il existe des attributs conceptuellement pertinents, et donc une valeur non nulle de  $N_1$ , le premier cas ne se produit qu'en l'absence d'attributs équivalents, et donc le deuxième cas ne s'applique pas.

Pour le concept  $(\{1, 3, 4\}, \{b, d, h, i\})$ , nous avons  $N_1 = (\{d'\})$ ,  $|N_1| = 1$ ,  $|B| = 4$  et  $\mathcal{H}_i = \{bd', dh', di'\}$  (voir la figure 2.3). Ce qui nous donne  $\alpha_{ex}(c) = \frac{1}{4} = 0.25$  et  $\beta_{ex}(c) = 0.75$ , d'où une valeur de l'indice de pertinence égale à 0.50.

L'algorithme du calcul de l'indice de pertinence des concepts est fourni ci-après.

---

**Algorithme 2** Indice extensionnel de pertinence du concept  $c$ .

---

**Input:** Concept  $c = (A, B)$ ,  $\mathcal{H}_i$  : ensemble des générateurs minimaux de l'intention  $B$

**Output:** Indice de pertinence des concepts  $\mathcal{CR}(c)$ .

```

1:  $\alpha \leftarrow \beta \leftarrow 0$ ; Count  $\leftarrow 0$ ;
   // Calcul du terme  $\alpha$ .
2: si  $B \neq \emptyset$  alors
3:   pour chaque attribut  $m$  dans  $B$  faire
4:     si  $(B \setminus \{m\})' \neq A$  alors
5:       Count  $\leftarrow$  Count + 1
6:     fin si
7:   fin pour
8:   si Count  $\neq 0$  alors
9:      $\alpha \leftarrow \frac{\text{Count}}{|B|}$ ;
10:  sinon
11:    pour chaque attribut  $m$  dans  $B$  faire
12:      si  $|m''| > 1$  &  $m' = A$  alors
13:        Count  $\leftarrow$  Count + 1;
14:      fin si
15:    fin pour
16:     $\alpha \leftarrow \frac{\text{Count}}{|B|}$ 
17:  fin si
   // Calcul du terme  $\beta$ .
18:  si  $|\mathcal{H}_i| > 1$  alors
19:     $\beta \leftarrow \text{Min}(\frac{|\mathcal{H}_i|}{|B|}, 1)$ ;
20:  sinon
21:    si  $|\mathcal{H}_i| = 1$  &  $|gn \in \mathcal{H}_i| < |B|$  alors
22:       $\beta \leftarrow \frac{1}{|B|}$ ;
23:    fin si
24:  fin si
25: fin si
26:  $\mathcal{CR}(c) \leftarrow (\alpha + \beta)/2$ ;
27: Retourner  $\mathcal{CR}(c)$ ;

```

---

Cet algorithme se compose de deux parties essentielles permettant de calculer la valeur de  $\alpha$  (ligne 2 à 17) et celle de  $\beta$  (ligne 18 à 24). Dans le calcul de  $\alpha$ , et pour un concept  $c = (A, B)$  du treillis, nous évaluons d'abord la présence d'au moins un attribut dans l'intention  $B$ , ensemble des attributs du concept sélectionné. Si cette condition est vérifiée, nous procédons à la recherche des attributs conceptuellement pertinents, et dans le cas échéant, des attributs équivalents. Pour illustrer le fonctionnement de cet algorithme, nous prenons le concept  $(\{ '1', '2', '3' \}, \{ 'c', 'd', 'g' \})$  avec  $A = \{ '1', '2', '3' \}$  et  $B = \{ 'c', 'd', 'g' \}$ . Pour tous les attributs de  $B$ , nous avons vérifié la condition de dérivation  $(B \setminus \{m\})' \neq A$  et obtenu les résultats suivants :  $(\{ 'd', 'g' \})' = \{ '1', '2', '3' \}$ ,  $(\{ 'c', 'g' \})' = \{ '1', '2', '3', '5' \}$  et  $(\{ 'c', 'd' \})' = \{ '1', '2', '3' \}$ . La partie bleue de la formule 2.3 exprimée par les lignes 11 à 15 de l'algorithme ne s'exécute que lorsqu'aucun attribut pertinent n'est trouvé. Dans cet exemple, seul l'attribut  $\{ 'd' \}$  respecte la condition de la ligne 4 de cet algorithme, ce qui permet d'avoir  $\frac{1}{3}$  comme valeur de  $\alpha_{ex}$ . La valeur de  $\mathcal{H}_i$  étant égale à  $(\{ 'd', 'c' \}, \{ 'd', 'g' \})$ , la formule à la ligne 19 s'applique alors pour le calcul de la valeur de  $\beta_{ex}$  qui est  $\frac{2}{3}$ .

D'une manière duale, nous proposons une formule similaire à  $\mathcal{CR}_{ex}(c)$  pour la version intentionnelle de l'indice de pertinence d'un concept, notée  $\mathcal{CR}_{in}(c)$  et définie comme suit :

$$\mathcal{CR}_{in}(c) = (\alpha_{in}(c) + \beta_{in}(c))/2 \quad (3.4)$$

avec

$$\alpha_{in}(c) = \begin{cases} N_1 = \frac{|\{g \in A \text{ tel que } (A \setminus \{g\})' \neq B\}|}{|A|} & \text{si } A \neq \emptyset \\ \frac{|\{g \in A \text{ tel que } |g'| > 1 \text{ et } g' = B\}|}{|A|} & \text{si } |N_1| = 0 \text{ et } A \neq \emptyset \\ 0 & \text{sinon} \end{cases} \quad (3.5)$$

et

$$\beta_{in}(c) = \begin{cases} \text{Min}(\frac{|\mathcal{H}_e|}{|A|}, 1) & \text{si } |\mathcal{H}_e| > 1 \\ \frac{1}{|A|} & \text{si } |\mathcal{H}_e| = 1 \text{ et } |gen| < |A| \text{ avec } gen \in \mathcal{H}_e \\ 0 & \text{sinon} \end{cases} \quad (3.6)$$

Où  $H_e$  est l'ensemble des générateurs minimaux de l'extension  $A$ .



Le tableau ci-après montre les différents résultats obtenus avec la formule révisée de  $\mathcal{CR}_{ex}(c)$ .

No	Concept(c)	$\alpha_{ex}(c)$	$\beta_{ex}(c)$	$\mathcal{CR}_{ex}(c)$
1	({'1'}, {'a','b','c','d','e','g','h','i'})	0.00	1.00	0.50
2	({'1','2'}, {'a','c','d','g'})	0.25	0.50	0.38
3	({'1','3'}, {'b','c','d','g','h','i'})	0.17	1.00	0.58
4	({'1','4'}, {'a','b','d','h','i'})	0.20	0.60	0.40
5	({'1','5'}, {'b','c','e','g','h','i'})	0.17	0.17	0.17
6	({'1','2','3'}, {'c','d','g'})	0.33	0.67	0.50
7	({'1','2','4'}, {'a','d'})	0.50	0.50	0.50
8	({'1','3','4'}, {'b','d','h','i'})	0.25	0.75	0.50
9	({'1','3','5'}, {'b','c','g','h','i'})	0.00	1.00	0.50
10	({'1','2','3','4'}, {'d'})	1.00	0.00	0.50
11	({'1','2','3','5'}, {'c','g'})	1.00	1.00	1.00
12	({'1','3','4','5'}, {'b','h','i'})	1.00	1.00	1.00

TABLEAU 3.1 – Les résultats des calculs de  $\alpha_{ex}(c)$  et  $\beta_{ex}(c)$

Si l'on considère ces concepts par ordre décroissant d'indice de pertinence  $\mathcal{CR}_{ex}(c)$ , selon l'ordre d'apparition des concepts, nous obtenons les concepts suivants :  $n^{\circ}11$ ,  $n^{\circ}12$ ,  $n^{\circ}3$ ,  $n^{\circ}1$ ,  $n^{\circ}6$ ,  $n^{\circ}7$ ,  $n^{\circ}8$ ,  $n^{\circ}9$ ,  $n^{\circ}10$ ,  $n^{\circ}4$ ,  $n^{\circ}2$  et  $n^{\circ}5$ . Les plus pertinents sont les concepts  $n^{\circ}11$ ,  $n^{\circ}12$  puisqu'ils ont leur valeur de  $\alpha_{ex}$  et  $\beta_{ex}$  égale à 1. En plus, ils couvrent tous les objets du treillis et partagent les objets 1, 3 et 5. Cela nous amène à noter que la détection de communautés va se faire sans aucune prédéfinition de seuil de sélection des concepts.

## 3.2 Détection des nœuds importants

L'identification des individus centraux ou importants dans un réseau social constitue une étape importante de notre recherche. Les individus centraux du réseau ou les objets pertinents du treillis sont généralement les individus par qui l'information circule le plus souvent. Pour les détecter, nous recherchons d'abord tous les concepts pertinents. Ce sont tous les concepts pris par ordre décroissant de l'indice de pertinence, jusqu'à ce que l'union de leur extension couvre l'ensemble des objets du contexte et en faisant appel à

l'algorithme 3 suivant.

De la même façon, nous pouvons proposer un algorithme similaire pour la détection des attributs intentionnellement pertinents.

---

**Algorithme 3** Recherche des objets extensionnellement pertinents

---

**Input:**  $C_p$  : ensemble des concepts extensionnellement pertinents,  $T_o$  : ensemble de tous les objets du contexte.

**Output:** Ensemble des objets pertinents  $O_p$ .

```

1:  $n \leftarrow \text{taille}(T_o)$ ;
   // Vecteur de comptage du nombre de fois qu'un objet apparaît.
2:  $O_p : \text{Tableau}(n)$ ;
3: pour  $i$  allant de 0 à  $n - 1$  faire
    $O_p(i) \leftarrow 0$ ;
4: fin pour
   // Listage des objets pertinents dans les concepts pertinents.
5: pour chaque concept  $c = (A, B)$  dans  $C_p$  faire
6:   pour  $k$  allant de 0 à  $\text{taille}(A) - 1$  faire
7:     pour  $v$  allant de 0 à  $\text{taille}(T_o) - 1$  faire
8:       si  $T_o(v) = A(k)$  alors
9:          $O_p(v) \leftarrow O_p(v) + 1$ 
10:      fin si
11:   fin pour
12: fin pour
13: Retourner  $O_p$ ;

```

---

L'algorithme 3 a été établi pour faire le comptage des objets pertinents dans un ensemble de concepts pertinents passé en paramètre. Il prend aussi en entrée tous les objets du contexte étudié. Une première boucle permet d'initialiser le tableau du résultat final noté  $O_p$ . La seconde boucle incluant une autre boucle imbriquée s'occupe de parcourir l'extension des concepts pertinents et de vérifier si un objet a été déjà identifié afin d'incrémenter sa fréquence d'apparition et de la mettre à jour dans la structure  $O_p$  à la fin de l'exécution de cet algorithme. À titre d'exemple, nous fournissons à l'algorithme 3 les concepts du tableau 3.2 et le résultat est donné dans le tableau 3.3 avec les objets

pertinents suivants  $\{‘1’\}$ ,  $\{‘2’\}$ ,  $\{‘3’\}$ ,  $\{‘4’\}$ ,  $\{‘5’\}$  et leur fréquence respective de 2, 1, 2, 1 et 2.

Dans le tableau 3.2, nous montrons les concepts pertinents issus de notre exemple en application de l’algorithme 3.

No	Concept(c)	$\alpha_{ex}(c)$	$\beta_{ex}(c)$	$\mathcal{CR}_{ex}(c)$
11	( $\{‘1’$ , $\{‘2’$ , $\{‘3’$ , $\{‘5’\}$ , $\{‘c’$ , $\{‘g’\}$ )	1.00	1.00	1.00
12	( $\{‘1’$ , $\{‘3’$ , $\{‘4’$ , $\{‘5’\}$ , $\{‘b’$ , $\{‘h’$ , $\{‘i’\}$ )	1.00	1.00	1.00

TABLEAU 3.2 – Tableau des concepts extensionnellement pertinents.

Parmi ces concepts pertinents, nous avons extrait les objets pertinents avec leur fréquence d’apparition, toujours selon l’algorithme 3. Les résultats sont consignés dans le tableau suivant.

No	Objet	Fréquence	Score
1	( $\{‘1’\}$ )	2	1.00
2	( $\{‘2’\}$ )	1	0.50
3	( $\{‘3’\}$ )	2	1.00
4	( $\{‘4’\}$ )	1	0.50
5	( $\{‘5’\}$ )	2	1.00

TABLEAU 3.3 – Tableau des objets pertinents parmi les concepts du tableau 3.2.

À partir de ces résultats, nous pouvons conclure que les objets les plus pertinents du treillis sont les objets  $\{‘1’\}$ ,  $\{‘3’\}$  et  $\{‘5’\}$ , suivis des objets  $\{‘2’\}$  et  $\{‘4’\}$ .

Dans ce qui suit et d’une manière duale, nous calculons l’indice intentionnel de pertinence des concepts pour déterminer les concepts importants ainsi que les attributs conceptuellement pertinents en suivant la même démarche que dans le cas des concepts extensionnellement pertinents. Les différents résultats sont répertoriés dans les tableaux ci-après.

No	Concept(c)	$\alpha_{in}(c)$	$\beta_{in}(c)$	$\mathcal{CR}_{in}(c)$
1	({'1'}, {'a','b','c','d','e','g','h','i'})	0.00	0.00	0.00
2	({'1','2'}, {'a','c','d','g'})	0.50	0.50	0.50
3	({'1','3'}, {'b','c','d','g','h','i'})	0.50	0.50	0.50
4	({'1','4'}, {'a','b','d','h','i'})	0.50	0.50	0.50
5	({'1','5'}, {'b','c','e','g','h','i'})	0.50	0.50	0.50
6	({'1','2','3'}, {'c','d','g'})	0.67	0.33	0.50
7	({'1','2','4'}, {'a','d'})	0.67	0.33	0.50
8	({'1','3','4'}, {'b','d','h','i'})	0.67	0.33	0.50
9	({'1','3','5'}, {'b','c','g','h','i'})	0.67	0.33	0.50
10	({'1','2','3','4'}, {'d'})	0.75	0.25	0.50
11	({'1','2','3','5'}, {'c','g'})	0.50	0.25	0.38
12	({'1','3','4','5'}, {'b','h','i'})	0.50	0.25	0.38

TABLEAU 3.4 – Les résultats des calculs de  $\alpha_{in}(c)$ ,  $\beta_{in}(c)$  et du  $\mathcal{CR}_{in}(c)$ 

Le tableau 3.4 montre le résultat de l'application de la nouvelle formule de calcul de l'indice intentionnel de pertinence des concepts. Nous avons identifié les concepts pertinents en fonction de la valeur de l'indice de pertinence  $\mathcal{CR}_{in}(c)$  classé par ordre décroissant, de telle sorte que l'union de leur intention couvre l'ensemble des attributs du contexte. Le tableau 3.5 suivant recense ces concepts intentionnellement pertinents.

No	Concept(c)	$\alpha_{in}(c)$	$\beta_{in}(c)$	$\mathcal{CR}_{in}(c)$
10	({'1','2','3','4'}, {'d'})	0.75	0.25	0.50
7	({'1','2','4'}, {'a','d'})	0.67	0.33	0.50
6	({'1','2','3'}, {'c','d','g'})	0.67	0.33	0.50
8	({'1','3','4'}, {'b','d','h','i'})	0.67	0.33	0.50
5	({'1','5'}, {'b','c','e','g','h','i'})	0.50	0.50	0.50

TABLEAU 3.5 – Tableau des concepts intentionnellement pertinents.

Pour retrouver les attributs conceptuellement pertinents décrivant les objets du contexte, nous avons utilisé un algorithme similaire à l'algorithme 3 qui prend en entrée, la liste de tous les concepts intentionnellement pertinents et l'ensemble des attributs du contexte et qui retourne la fréquence d'apparition de chaque attribut dans l'intention des

concepts. Selon le tableau suivant, nous remarquons que, sur les six concepts pertinents, l'attribut  $\{d\}$  est le plus pertinent avec un score de 0.67, suivi des attributs  $\{b\}$ ,  $\{c\}$ ,  $\{g\}$ ,  $\{h\}$  et  $\{i\}$  avec un score de 0.33 pour chacun et enfin, les attributs  $\{a\}$  et  $\{e\}$  avec un score de 0.17.

No	Attribut	Fréquence	Score
1	( $\{d\}$ )	4	0.67
2	( $\{b\}$ )	2	0.33
3	( $\{c\}$ )	2	0.33
4	( $\{g\}$ )	2	0.33
5	( $\{h\}$ )	2	0.33
5	( $\{i\}$ )	2	0.33
5	( $\{a\}$ )	1	0.17
5	( $\{e\}$ )	1	0.17

TABLEAU 3.6 – Tableau des attributs conceptuellement pertinents parmi les concepts du tableau 3.5.

Dans cette partie de notre travail, nous avons détecté et identifié les individus centraux du réseau social ou objets pertinents du treillis. Nous avons d'abord recherché les concepts pertinents en fonction de l'indice de pertinence des concepts classés par ordre décroissant et en retenant les concepts jusqu'à ce que l'union des extensions couvre l'ensemble des objets du contexte. Enfin, nous avons fait appel à l'algorithme 3 pour récupérer tous les objets pertinents en fonction de leur fréquence d'apparition.

Dans la sous-section suivante, nous montrons comment détecter les communautés sur la base des concepts pertinents.

### 3.3 Détection des communautés fondées sur les concepts

Pour la détection des communautés, nous adoptons une approche qui se repose sur une adaptation de l'algorithme CoDeBi décrit dans [35, 1]. Cette adaptation nous permet de détecter des communautés qui se chevauchent, voire qui sont imbriquées, dans des graphes bipartis. Pour ce faire, nous avons utilisé l'indice de pertinence des concepts  $\mathcal{CR}_{ex}(c)$  au lieu de l'indice de stabilité, ainsi que l'indice de séparation mentionné précédemment dans le chapitre 2. Ainsi, une communauté est retenue si ses objets présentent

---

une forte cohésion, identifiée par un indice de pertinence élevé, et une faible séparation entre ces objets. Les deux indices sont pris en compte en calculant leur moyenne harmonique.

La procédure comporte trois principales étapes :

1. Calcul de l'ensemble des concepts à partir du contexte formel décrivant les données du réseau à deux modes.
2. Calcul de l'autonomie du concept comme étant la moyenne harmonique des indices de pertinence et de séparation, puis classement de cette moyenne par ordre décroissant.
3. Sélection des concepts en suivant l'ordre décroissant jusqu'à ce que l'union de tous les objets de ces concepts couvre l'ensemble des objets du contexte. Ces concepts constituent alors des communautés potentielles.

Si le nombre de communautés finales reste élevé, une étape de percolation optionnelle peut être envisagée afin de fusionner les groupes qui partagent des éléments communs. Le diagramme suivant montre les différentes étapes de notre démarche en partant du contexte formel jusqu'à la détection des individus pertinents du réseau.

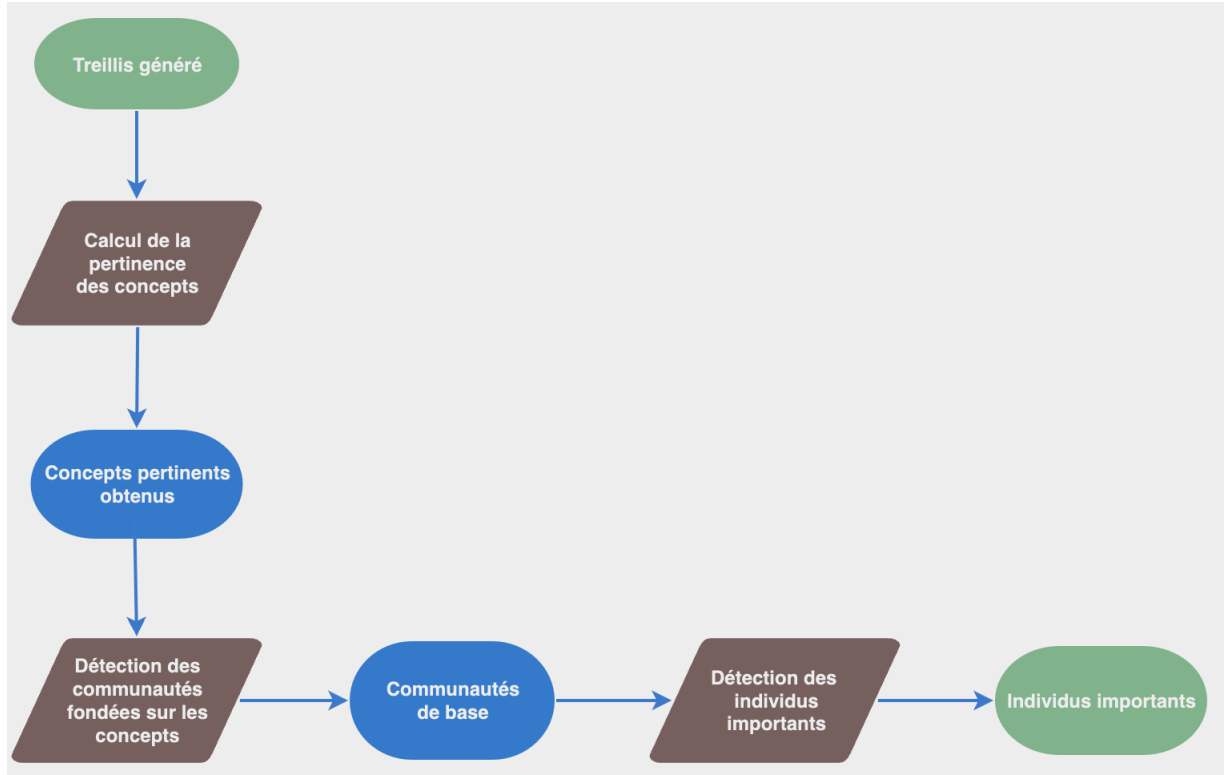


FIGURE 3.1 – Diagramme des étapes de notre approche. Les losanges représentent les différentes opérations à réaliser sur les données alors que les ovales montrent les entrées ou sorties des opérations. Le sens des flèches indique le sens du déroulement des opérations.

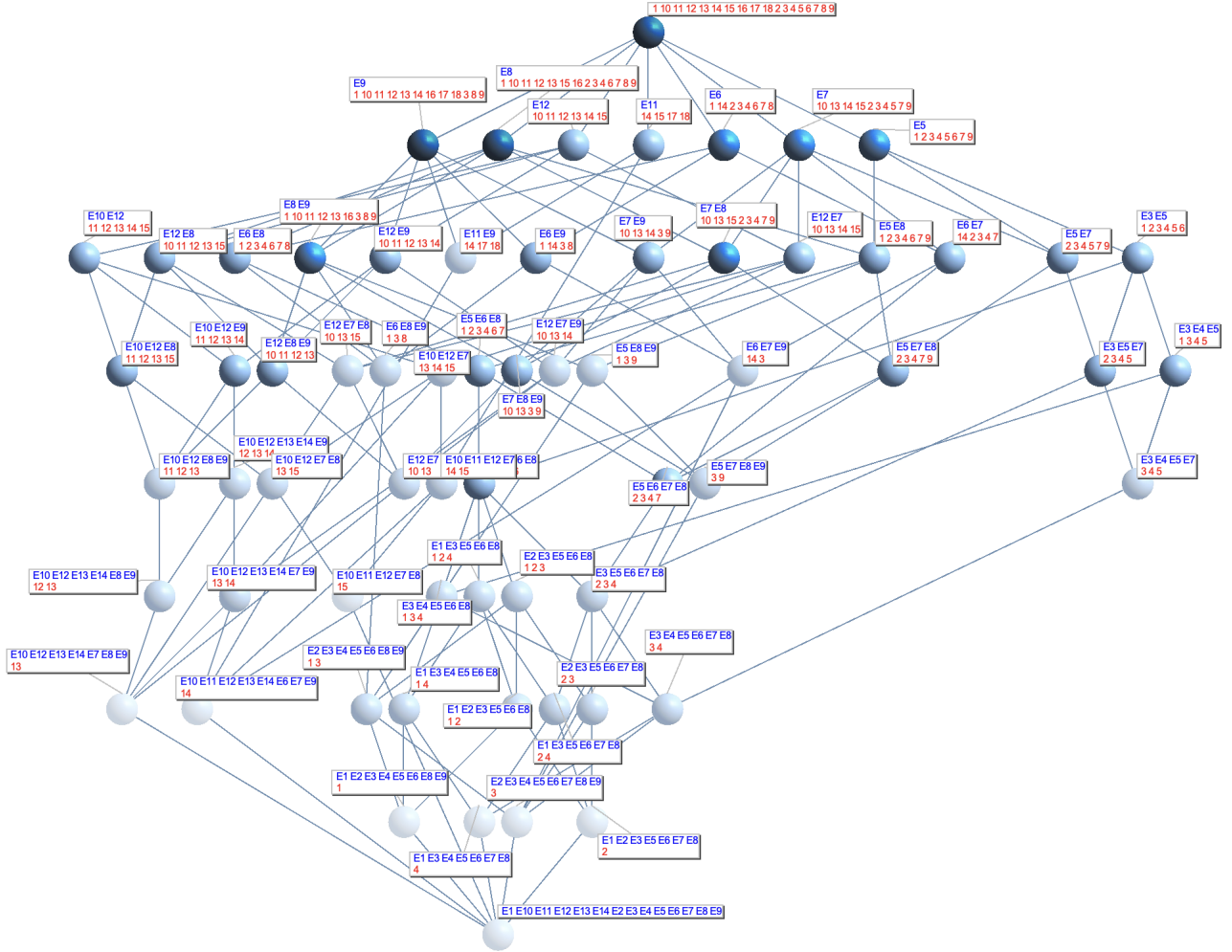
Pour illustrer notre démarche, nous utilisons l'exemple connu et exploité fréquemment en analyse de réseaux sociaux. Il s'agit de l'étude menée par Davis [11] sur la participation de dix-huit femmes du sud des États-Unis à quatorze événements sociaux. Freeman [18] en a fait une méta-analyse de plusieurs travaux d'identification de groupes sociaux et de femmes importantes (i.e., ayant une forte importance ou centralité au sein du réseau). Le contexte formel correspondant comporte donc dix-huit lignes/objets (nœuds de type 1) notées de 1 à 18 et 14 attributs/événements (nœuds de type 2) identifiées par E1 à E14, et 89 liens qui indiquent la participation de ces femmes aux événements et qui est symbolisée par un "X" dans le tableau 3.7. Notons que le jeu de données à l'origine comporte plutôt le nom des femmes que nous avons remplacé par des numéros pour des raisons de simplicité d'affichage et d'interprétation.

$\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
1	X	X	X	X	X	X		X	X					
2	X	X	X		X	X	X	X						
3		X	X	X	X	X	X	X	X					
4	X		X	X	X	X	X	X						
5			X	X	X		X							
6			X		X	X		X						
7					X	X	X	X						
8						X		X	X					
9					X		X	X	X					
10							X	X	X			X		
11								X	X	X		X		
12								X	X	X		X	X	X
13							X	X	X	X		X	X	X
14						X	X		X	X	X	X	X	X
15							X	X		X	X	X		
16								X	X					
17									X		X			
18									X		X			

TABLEAU 3.7 – Contexte formel de Woman-Southern-Davis

Le treillis correspondant au contexte *Woman-Southern-Davis* est donné par la figure 3.2 où les nœuds représentent les concepts alors que les arcs sont synonymes de la relation d'ordre partiel. Dans ce treillis, et pour un concept donné, nous affichons tous les attributs (en rouge) et les objets (en bleu). Le supremum c'est-à-dire le concept le plus haut du treillis est un concept tout-objet (qui possède tous les objets) alors que l'infimum est le concept le plus bas du treillis et qui possède tous les attributs.



FIGURE 3.2 – Treillis issu du contexte *Woman-Southern-Davis*

### 3.3.1 Extraction des concepts pertinents

L'objectif de cette partie est de déterminer les concepts pertinents pouvant former une communauté, constituée des objets ayant une cohésion interne entre eux et une séparation avec le reste des objets dans le contexte.

La première étape de notre démarche nous a permis d'extraire au total 63 concepts de base que comportait le jeu de données *Woman-Southern-Davis*, comme le montre le tableau 3.8 Nous avons ensuite calculé l'indice de pertinence des concepts ( $\mathcal{CR}_{ex}(c)$ ). Ce calcul n'est pas fait pour le supremum du treillis mais est effectué sur l'infimum uniquement lorsque l'extension n'est pas nulle. Notons que cette extraction ne tient compte

Concepts	A	B
C1	{1}	{'E1', 'E2', 'E3', 'E4', 'E5', 'E6', 'E8', 'E9'}
C2	{2}	{'E1', 'E2', 'E3', 'E5', 'E6', 'E7', 'E8'}
C3	{3}	{'E2', 'E3', 'E4', 'E5', 'E6', 'E7', 'E8', 'E9'}
C4	{4}	{'E1', 'E3', 'E4', 'E5', 'E6', 'E7', 'E8'}
C5	{13}	{'E7', 'E8', 'E9', 'E10', 'E12', 'E13', 'E14'}
C6	{14}	{'E6', 'E7', 'E9', 'E10', 'E11', 'E12', 'E13', 'E14'}
C7	{15}	{'E7', 'E8', 'E10', 'E11', 'E12'}
C8	{1, 2}	{'E1', 'E2', 'E3', 'E5', 'E6', 'E8'}
C9	{1, 3}	{'E2', 'E3', 'E4', 'E5', 'E6', 'E8', 'E9'}
C10	{1, 4}	{'E1', 'E3', 'E4', 'E5', 'E6', 'E8'}
C11	{2, 3}	{'E2', 'E3', 'E5', 'E6', 'E7', 'E8'}
C12	{2, 4}	{'E1', 'E3', 'E5', 'E6', 'E7', 'E8'}
C13	{3, 4}	{'E3', 'E4', 'E5', 'E6', 'E7', 'E8'}
C14	{3, 9}	{'E5', 'E7', 'E8', 'E9'}
C15	{3, 14}	{'E6', 'E7', 'E9'}
C16	{10, 13}	{'E7', 'E8', 'E9', 'E12'}
C17	{12, 13}	{'E8', 'E9', 'E10', 'E12', 'E13', 'E14'}
C18	{13, 14}	{'E7', 'E9', 'E10', 'E12', 'E13', 'E14'}
C19	{13, 15}	{'E7', 'E8', 'E10', 'E12'}
C20	{14, 15}	{'E7', 'E10', 'E11', 'E12'}
C21	{1, 2, 3}	{'E2', 'E3', 'E5', 'E6', 'E8'}
C22	{1, 2, 4}	{'E1', 'E3', 'E5', 'E6', 'E8'}
C23	{1, 3, 4}	{'E3', 'E4', 'E5', 'E6', 'E8'}
C24	{1, 3, 8}	{'E6', 'E8', 'E9'}
C25	{1, 3, 9}	{'E5', 'E8', 'E9'}
C26	{2, 3, 4}	{'E3', 'E5', 'E6', 'E7', 'E8'}
C27	{3, 4, 5}	{'E3', 'E4', 'E5', 'E7'}
C28	{10, 13, 14}	{'E7', 'E9', 'E12'}
C29	{10, 13, 15}	{'E7', 'E8', 'E12'}
C30	{11, 12, 13}	{'E8', 'E9', 'E10', 'E12'}
C31	{12, 13, 14}	{'E9', 'E10', 'E12', 'E13', 'E14'}
C32	{13, 14, 15}	{'E7', 'E10', 'E12'}
C33	{14, 17, 18}	{'E9', 'E11'}
C34	{1, 3, 4, 5}	{'E3', 'E4', 'E5'}
C35	{1, 3, 8, 14}	{'E6', 'E9'}
C36	{2, 3, 4, 5}	{'E3', 'E5', 'E7'}
C37	{2, 3, 4, 7}	{'E5', 'E6', 'E7', 'E8'}
C38	{3, 9, 10, 13}	{'E7', 'E8', 'E9'}
C39	{10, 11, 12, 13}	{'E8', 'E9', 'E12'}
C40	{10, 13, 14, 15}	{'E7', 'E12'}
C41	{11, 12, 13, 14}	{'E9', 'E10', 'E12'}
C42	{11, 12, 13, 15}	{'E8', 'E10', 'E12'}
C43	{14, 15, 17, 18}	{'E11'}
C44	{1, 2, 3, 4, 6}	{'E3', 'E5', 'E6', 'E8'}
C45	{2, 3, 4, 7, 9}	{'E5', 'E7', 'E8'}
C46	{2, 3, 4, 7, 14}	{'E6', 'E7'}
C47	{3, 9, 10, 13, 14}	{'E7', 'E9'}
C48	{10, 11, 12, 13, 14}	{'E9', 'E12'}
C49	{10, 11, 12, 13, 15}	{'E8', 'E12'}
C50	{11, 12, 13, 14, 15}	{'E10', 'E12'}
C51	{1, 2, 3, 4, 5, 6}	{'E3', 'E5'}
C52	{1, 2, 3, 4, 6, 7}	{'E5', 'E6', 'E8'}
C53	{2, 3, 4, 5, 7, 9}	{'E5', 'E7'}
C54	{10, 11, 12, 13, 14, 15}	{'E12'}
C55	{1, 2, 3, 4, 6, 7, 8}	{'E6', 'E8'}
C56	{1, 2, 3, 4, 6, 7, 9}	{'E5', 'E8'}
C57	{1, 2, 3, 4, 5, 6, 7, 9}	{'E5'}
C58	{1, 2, 3, 4, 6, 7, 8, 14}	{'E6'}
C59	{2, 3, 4, 7, 9, 10, 13, 15}	{'E7', 'E8'}
C60	{1, 3, 8, 9, 10, 11, 12, 13, 16}	{'E8', 'E9'}
C61	{2, 3, 4, 5, 7, 9, 10, 13, 14, 15}	{'E7'}
C62	{1, 3, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18}	{'E9'}
C63	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16}	{'E8'}

TABLEAU 3.8 – Tableau des concepts de base

---

d'aucun critère particulier de sélection. Une fois que les concepts sont sélectionnés, nous avons calculé la séparation selon l'équation 2.5 et l'autonomie  $\zeta(c)$  des concepts qui est la moyenne harmonique de la séparation et l'indice de pertinence des concepts, pour ne garder que ceux ayant une plus grande autonomie de telle sorte que l'union de leur extension couvre l'ensemble des objets du contexte. Plus la valeur de l'autonomie est élevée, plus les concepts retenus sont pertinents et forment ainsi de potentielles communautés de base.

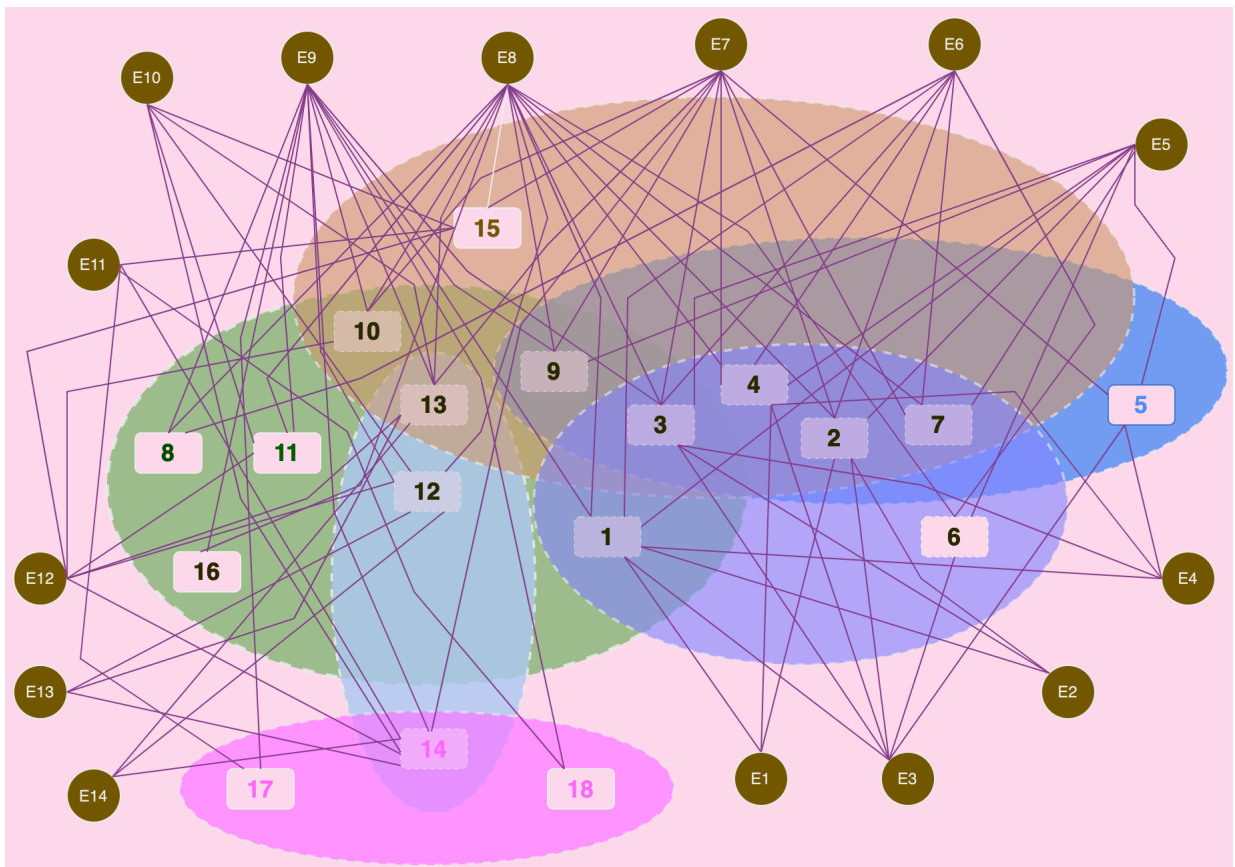
Le tableau 3.9 montre les différents résultats avec le calcul de l'autonomie des concepts. Ces résultats sont classés par ordre décroissant de l'autonomie. En parcourant ce tableau et en récupérant les objets de la colonne A de telle sorte que leur union forme l'ensemble de tous les objets du contexte, nous avons identifié six concepts pertinents formant ainsi les six communautés de base du tableau 3.10. Un schéma de ces communautés de base est donné par la figure 3.3.

Concepts	A	B	$\alpha_{ex}$	$\beta_{ex}$	$\mathcal{CR}_{ex}(c)$	$\gamma(c)$	$\zeta(c)$
C60	[1', 3', 8', 9', 10', 11', 12', 13', 16']	[E8', E9']	1	0.5	0.75	0.33	0.46
C53	[2', 3', 4', 5', 7', 9']	[E5', E7']	1	0.5	0.75	0.33	0.43
C45	[2', 3', 4', 7', 9']	[E5', E7', E8']	1	0.33	0.67	0.32	0.43
C52	[1', 2', 3', 4', 6', 7']	[E5', E6', E8']	0.67	0.33	0.50	0.36	0.42
C59	[2', 3', 4', 7', 9', 10', 13', 15']	[E7', E8']	1.00	0.50	0.75	0.30	0.42
C42	[11', 12', 13', 15']	[E8', E10', E12']	0.67	0.33	0.5	0.34	0.41
C49	[10', 11', 12', 13', 15']	[E8', E12']	1	0.5	0.75	0.28	0.41
C31	[12', 13', 14']	[E9', E10', E12', E13', E14']	0.40	0.40	0.40	0.43	0.41
C55	[1', 2', 3', 4', 6', 7', 8']	[E6', E8']	1	0.5	0.75	0.29	0.41
C56	[1', 2', 3', 4', 6', 7', 9']	[E5', E8']	1	0.5	0.75	0.28	0.41
C39	[10', 11', 12', 13']	[E8', E9', E12']	1	0.33	0.67	0.29	0.41
C41	[11', 12', 13', 14']	[E9', E10', E12']	0.67	0.33	0.5	0.33	0.4
C48	[10', 11', 12', 13', 14']	[E9', E12']	1	0.5	0.75	0.27	0.4
C33	[14', 17', 18']	[E9', E11']	1.00	0.50	0.75	0.27	0.40
C37	[2', 3', 4', 7']	[E5', E6', E7', E8']	0.5	0.5	0.5	0.32	0.39
C36	[2', 3', 4', 5']	[E3', E5', E7']	0.67	0.33	0.5	0.32	0.39
C50	[11', 12', 13', 14', 15']	[E10', E12']	0.5	0.5	0.5	0.32	0.39
C44	[1', 2', 3', 4', 6']	[E3', E5', E6', E8']	0.25	0.5	0.38	0.4	0.39
C40	[10', 13', 14', 15']	[E7', E12']	1	0.5	0.75	0.25	0.38
C38	[3', 9', 10', 13']	[E7', E8', E9']	1	0.33	0.67	0.26	0.37
C51	[1', 2', 3', 4', 5', 6']	[E3', E5']	0.5	0.5	0.5	0.3	0.37
C30	[11', 12', 13']	[E8', E9', E10', E12']	0.75	0.25	0.5	0.29	0.36
C27	[3', 4', 5']	[E3', E4', E5', E7']	0.5	0.25	0.38	0.34	0.36
C47	[3', 9', 10', 13', 14']	[E7', E9']	1	0.5	0.75	0.23	0.36
C46	[2', 3', 4', 7', 14']	[E6', E7']	1	0.5	0.75	0.24	0.36
C29	[10', 13', 15']	[E7', E8', E12']	1	0.33	0.67	0.24	0.36
C32	[13', 14', 15']	[E7', E10', E12']	0.67	0.33	0.5	0.28	0.36
C28	[10', 13', 14']	[E7', E9', E12']	1	0.33	0.67	0.24	0.35
C34	[1', 3', 4', 5']	[E3', E4', E5']	0.33	0.33	0.33	0.36	0.35
C20	[14', 15']	[E7', E10', E11', E12']	0.25	0.75	0.5	0.27	0.35
C26	[2', 3', 4']	[E3', E5', E6', E7', E8']	0.4	0.4	0.4	0.28	0.33
C35	[1', 3', 8', 14']	[E6', E9']	1	0.5	0.75	0.21	0.32
C43	[14', 15', 17', 18']	[E11']	1	0	0.5	0.24	0.32
C18	[13', 14']	[E7', E9', E10', E12', E13', E14']	0.17	0.5	0.34	0.29	0.31
C24	[1', 3', 8']	[E6', E8', E9']	1	0.33	0.67	0.2	0.31
C23	[1', 3', 4']	[E3', E4', E5', E6', E8']	0.2	0.4	0.3	0.31	0.31
C19	[13', 15']	[E7', E8', E10', E12']	0.75	0.25	0.5	0.21	0.29
C25	[1', 3', 9']	[E5', E8', E9']	0.67	0.33	0.5	0.2	0.29
C9	[1', 3']	[E2', E3', E4', E5', E6', E8', E9']	0	0.71	0.35	0.25	0.29
C62	[1', 3', 8', 9', 10', 11', 12', 13', 14', 16', 17', 18']	[E9']	1	0	0.5	0.21	0.29
C63	[1', 2', 3', 4', 6', 7', 8', 9', 10', 11', 12', 13', 15', 16']	[E8']	1	0	0.5	0.19	0.28
C16	[10', 13']	[E7', E8', E9', E12']	1	0.25	0.62	0.18	0.28
C13	[3', 4']	[E3', E4', E5', E6', E7', E8']	0.33	0.33	0.33	0.23	0.27
C8	[1', 2']	[E1', E2', E3', E5', E6', E8']	0.33	0.17	0.25	0.27	0.26
C17	[12', 13']	[E8', E9', E10', E12', E13', E14']	0.17	0.33	0.25	0.27	0.26
C61	[2', 3', 4', 5', 7', 9', 10', 13', 14', 15']	[E7']	1	0	0.5	0.17	0.26
C10	[1', 4']	[E1', E3', E4', E5', E6', E8']	0.33	0.17	0.25	0.26	0.26
C54	[10', 11', 12', 13', 14', 15']	[E12']	1	0	0.5	0.18	0.26
C57	[1', 2', 3', 4', 5', 6', 7', 9']	[E5']	1	0	0.5	0.17	0.26
C21	[1', 2', 3']	[E2', E3', E5', E6', E8']	0.2	0.2	0.2	0.32	0.25
C15	[3', 14']	[E6', E7', E9']	1	0.33	0.67	0.15	0.25
C14	[3', 9']	[E5', E7', E8', E9']	0.75	0.25	0.5	0.17	0.25
C58	[1', 2', 3', 4', 6', 7', 8', 14']	[E6']	1	0	0.5	0.16	0.25
C22	[1', 2', 4']	[E1', E3', E5', E6', E8']	0.2	0.2	0.2	0.33	0.25
C6	[14']	[E6', E7', E9', E10', E11', E12', E13', E14']	0	1	0.5	0.16	0.24
C11	[2', 3']	[E2', E3', E5', E6', E7', E8']	0.33	0.17	0.25	0.23	0.24
C12	[2', 4']	[E1', E3', E5', E6', E7', E8']	0.33	0.17	0.25	0.24	0.24
C3	[3']	[E2', E3', E4', E5', E6', E7', E8', E9']	0.12	0.75	0.43	0.12	0.19
C5	[13']	[E7', E8', E9', E10', E12', E13', E14']	0.29	0.43	0.36	0.13	0.19
C2	[2']	[E1', E2', E3', E5', E6', E7', E8']	0.43	0.14	0.29	0.13	0.18
C4	[4']	[E1', E3', E4', E5', E6', E7', E8']	0.43	0.14	0.29	0.13	0.18
C7	[15']	[E7', E8', E10', E11', E12']	0.4	0.2	0.3	0.13	0.18
C1	[1']	[E1', E2', E3', E4', E5', E6', E8', E9']	0.12	0.25	0.18	0.14	0.16

TABLEAU 3.9 – Tableau des concepts avec les composantes  $\alpha_{ex}$ ,  $\beta_{ex}$ , la séparation ( $\gamma(c)$ ) et l'autonomie ( $\zeta(c)$ )

Concepts	Communautés	Événements partagés
C60	{'1', '3', '8', '9', '10', '11', '12', '13', '16'}	{'E8', 'E9'}
C53	{'2', '3', '4', '5', '7', '9'}	{'E5', 'E7'}
C52	{'1', '2', '3', '4', '6', '7'}	{'E5', 'E6', 'E8'}
C59	{'2', '3', '4', '7', '9', '10', '13', '15'}	{'E7', 'E8'}
C31	{'12', '13', '14'}	{'E9', 'E10', 'E12', 'E13', 'E14'}
C33	{'14', '17', '18'}	{'E9', 'E11'}

TABLEAU 3.10 – Tableau des communautés de base avec les liens partagés.

FIGURE 3.3 – Figure des communautés de base du contexte *Woman-Southern-Davis* identifiées par notre approche

Sur cette figure, les rectangles comportant les chiffres de 1 à 18, représentent les individus du réseau alors que les petits cercles avec les valeurs E1 à E14 à l'intérieur,

sont les événements auxquels participent les individus. Les arcs entre les individus et les événements symbolisent la relation existant entre eux.

Selon notre démarche, nous avons trouvé des communautés cohésives qui ont la particularité suivante, ce qui est rare dans les autres travaux de recherches : du recouvrement entre les communautés et même des communautés imbriquées. Toutefois, il est possible, selon le besoin de l'utilisateur, d'appliquer une technique de fusion des communautés sur ces résultats.

Dans notre cas, nous remarquons que la communauté C31 partage avec C60 les objets {'12', '13'} et avec C33, l'objet {'14'} et que C52, C53 et C59 partagent entre eux les objets {'2', '3', '4', '7'}. En plus, les communautés C53 et C59 partagent l'objet {'9'}. Ces observations nous mènent à penser que ces communautés peuvent être regroupées pour en former une plus grande. En procédant ainsi, ce regroupement supplémentaire produit les communautés finales suivantes :

Communautés	Objets
1	{'1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '13', '15'}
2	{'11', '12', '14', '16', '17', '18'}

TABLEAU 3.11 – Tableau des communautés finales.

Tel qu'indiqué auparavant, Freeman [17] a effectué une méta-analyse dans le domaine de la détection des communautés dans les réseaux sociaux à deux modes en prenant l'exemple des femmes et des événements. La plupart des méthodes utilisées par les différents chercheurs ont révélé la présence de deux communautés distinctes avec parfois, un certain nombre d'individus qui se chevauchent entre les deux communautés, comme le montre la figure ci-dessous.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	DGG41	W	W	W	W	W	W	W	W	WW	W	W	W	W	W	W	W	W	W
2	HOM50	W	W	W	W	W	W	W	WW			W	W	W	W	W	W	W	W
3	P&C72	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
4	BGR74	W	W	W	W	W	W	W		W	W	W	W	W	WW	WW		W	W
5	BBA75	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
6	BCH78	W	W	W	W	W	W				W	W	W	W	W	W			
7	DOR79	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
8	BCH91	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
9	FRE92	W	W	W	W	W	W	W		W	W	W	W	W	W	W	W		
10	E&B93	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
11	FR193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
12	FR293	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
13	FW193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	WW	W	W
14	FW293	W	W	W	W	W	W	W		W	W	W	W	W	W	W		W	W
15	BE197	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
16	BE297	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
17	BE397	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
18	S&F99	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W		W	W
19	ROB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
20	OSB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
21	NEW01	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W

FIGURE 3.4 – Résultat de la méta-analyse de Freeman [17]

Dans la figure 3.4, les 18 colonnes représentent les 18 femmes recensées dans le jeu de données *Woman-Southern-Davis*. Les 21 lignes renvoient aux 21 procédures analytiques utilisées par les différents auteurs. Une femme dans chaque cellule est désignée par un "W" et les groupes sont désignés par des couleurs (rouge, verte ou bleue). Tous les "W" rouges d'une ligne donnée ont été affectés au même groupe par la procédure désignée dans cette ligne. Tous les "W" bleus ont été affectés à un deuxième groupe. Enfin, dans la quatrième rangée, les "W" verts ont été affectés à un troisième groupe. Toute femme affectée à deux groupes par la procédure en question a reçu une paire de codes de couleur.

D'après les communautés obtenues et recensées dans le tableau 3.11, nos résultats montrent que notre méthode d'identification des communautés se rapproche de la méthode DGG41, en identifiant aussi deux communautés de base après l'étape de regroupement supplémentaire. Cette méthode a été celle qui a identifié les individus 1 à 9 comme faisant partie d'un premier groupe et les individus 10 à 18 comme membres du second groupe. L'individu 9 étant partagé entre les deux groupes détectés.

### 3.3.2 Détection d'objets pertinents

La détection des nœuds ou individus centraux dans les réseaux sociaux revêt une importance capitale aujourd'hui puisque ces acteurs centraux ont la capacité d'influencer les autres acteurs de réseau lorsqu'ils interagissent avec eux. Détecter ces individus revient à les identifier pour toucher largement le réseau dans le but de mener une action concrète.

Dans le cadre de ce mémoire, et en partant des concepts pertinents déjà obtenus, nous avons pu détecter les objets pertinents (algorithme 3) qui sont les acteurs de type 1 du réseau. L'application de cette démarche sur le jeu de données *Woman-Southern-Davis*, et en se basant sur les résultats du tableau 3.9, nous avons obtenu les objets pertinents dans le tableau 3.13.

No	Concept(c)	Attributs	$\alpha_{ex}(c)$	$\beta_{ex}(c)$	$\mathcal{CR}_{ex}(c)$
C49	{'10', '11', '12', '13', '15'}	{'E8', 'E12'}	1.00	0.50	0.75
C56	{'1', '2', '3', '4', '6', '7', '9'}	{'E5', 'E8'}	1.00	0.50	0.75
C46	{'2', '3', '4', '7', '14'}	{'E6', 'E7'}	1.00	0.50	0.75
C35	{'1', '3', '8', '14'}	{'E6', 'E9'}	1.00	0.50	0.75
C53	{'2', '3', '4', '5', '7', '9'}	{'E5', 'E7'}	1.00	0.50	0.75
C60	{'1', '3', '8', '9', '10', '11', '12', '13', '16'}	{'E8', 'E9'}	1.00	0.50	0.75
C33	{'14', '17', '18'}	{'E9', 'E11'}	1.00	0.50	0.75

TABLEAU 3.12 – Tableau des concepts extensionnellement pertinents.

En établissant la fréquence d'apparition de chaque objet dans l'ensemble des extensions de ces concepts pertinents, nous retrouvons le tableau des objets pertinents suivant :



No	Objet	Fréquence	Score
1	({'1'})	3	0.43
2	({'2'})	3	0.43
3	({'3'})	5	0.71
4	({'4'})	3	0.43
5	({'5'})	1	0.14
6	({'6'})	1	0.14
7	({'7'})	3	0.43
8	({'8'})	2	0.28
9	({'9'})	3	0.43
10	({'10'})	2	0.28
11	({'11'})	2	0.28
12	({'12'})	2	0.28
13	({'13'})	2	0.28
14	({'14'})	3	0.43
15	({'15'})	1	0.14
16	({'16'})	1	0.14
17	({'17'})	1	0.14
18	({'18'})	1	0.14

TABLEAU 3.13 – Tableau des objets pertinents parmi les concepts du tableau 3.12.

Nous remarquons que l'objet le plus pertinent est le {'3'} avec une fréquence de 5 ou un score de 0.71, suivi des objets {'1'}, {'2'}, {'4'}, {'7'}, {'9'} et {'14'} pour une fréquence de 3 ou une score de 0.43.

À titre d'exemple, nous avons utilisé l'indice de stabilité  $\sigma(c)$  (cf. équation 2.1) pour déterminer les concepts pertinents sur le jeu de données *Woman-Southern-Davis*, dont les résultats sont consignés dans le tableau 3.14.

No	Concept(c)	Attributs	$\sigma(c)$
C31	{'12', '13', '14'}	{'E9', 'E10', 'E12', 'E3', 'E4'}	0.75
C9	{'1', '3'}	{'E2', 'E3', 'E4', 'E5', 'E6', 'E8', 'E9', }	0.58
C63	{'1', '2', '3', '4', '6', '7', '8', '9', '10', '11', '12', '13', '15', '16'}	{'E8'}	0.50
C51	{'1', '2', '3', '4', '5', '6', }	{'E3', 'E5'}	0.50
C43	{'14', '15', '17', '18'}	{'E11'}	0.50

TABLEAU 3.14 – Tableau des concepts pertinents selon l'indice de stabilité.

Ce tableau montre les cinq concepts pertinents avec l'utilisation de l'indice de stabilité contre sept avec l'approche basée sur l'indice de pertinence.

Dans ce chapitre, nous avons apporté une amélioration de la formule initiale de l'indice de pertinence en tenant compte des attributs équivalents pour la composante  $\alpha_{ex}$  et l'utilisation des générateurs minimaux dans la composante  $\beta_{ex}$ . Une approche de détection des concepts formels sémantiquement pertinents a été élaborée et basée sur les nouvelles améliorations de l'indice de pertinence et a été appliquée à la détection des communautés et des objets (individus) importants dans le contexte. Dans le chapitre suivant, nous allons procéder à quelques expérimentations afin de valider la nouvelle formule de l'indice de pertinence.

# Chapitre 4

## Analyse empirique

### 4.1 Expérimentation

Afin de valider la nouvelle formule de l'indice de pertinence des concepts, nous avons mené une série de tests sur des jeux des données de réseaux synthétiques et réels.

L'objectif principal de l'évaluation expérimentale est de répondre aux trois questions suivantes :

- (Q1.) L'indice  $\mathcal{CR}$  est-il empiriquement précis par rapport à d'autres indices pour évaluer la pertinence des concepts formels ? Nous allons étudier empiriquement la relation entre l'indice de pertinence conceptuelle et les indices qui y sont reliés tels que la stabilité.
- (Q2.) L'indice  $\mathcal{CR}$  est-il plus rapide que les indices d'intérêt actuels ? Nous voulons valider l'efficacité de l'indice de pertinence conceptuelle en termes de temps d'exécution.
- (Q3.) Existe-t-il une corrélation entre notre approche de détection des nœuds de type-1 et les autres mesures de centralité dans l'état de l'art, comme la centralité de degré ( $D_c$ ), la centralité de proximité ( $C_c$ ) et la centralité d'intermédiarité ( $B_c$ ) ?

#### 4.1.1 Méthodologie

Pour obtenir des réponses fiables aux questions Q1 et Q2, nous choisissons d'abord les ensembles de données synthétiques et réels suivants.

### Données

- **Coin-Toss** [13] qui est un ensemble de données produit par le générateur de modèle *Coin-Toss*
- **Woman-Southern-Davis**<sup>1</sup> [10] qui décrit un ensemble de dix-huit femmes du Sud des États-Unis et les quatorze événements auxquels elles ont assisté.
- **Phytothérapie** [3] qui décrit un ensemble de maladies et de plantes médicinales qui les traitent.

Pour l'ensemble de données appelé "Phytothérapie" [3], nous avons produit une matrice de 142 objets (plantes) et 108 attributs (maladies). L'intersection d'une ligne et d'une colonne de cette matrice contient la valeur 1 ou la valeur 0, selon que la plante indiquée en ligne soigne la maladie présentée en colonne. La description des trois ensembles est fournie dans le tableau 4.1.

Données	$ \mathcal{G} $	$ \mathcal{M} $	$n$	$k$	$\delta$
<i>Coin-Toss</i>	793	10	3310	911	0.41
<i>Woman-Southern-Davis</i>	18	14	89	63	0.35
Phytothérapie	142	108	511	302	0.03

TABLEAU 4.1 – Une description des ensembles de données testés, où  $|\mathcal{G}|$  est le nombre d'objets,  $|\mathcal{M}|$  est le nombre d'attributs,  $n$  est le nombre de liens entre les ensembles  $\mathcal{G}$  et  $\mathcal{M}$ ,  $k$  est le nombre de concepts produits par le treillis et pour lesquels l'indice de pertinence et de stabilité a été calculé et  $\delta$  est la densité du contexte avec  $\delta = \frac{n}{|\mathcal{G}| \times |\mathcal{M}|}$

Les résultats du calcul de l'indice de pertinence ont été ensuite comparés à l'indice de stabilité qui est actuellement le plus connu pour le calcul de l'intérêt des concepts formels. Lors de cette comparaison des résultats, nous enregistrons une liste de scores  $\{(x_i, y_i)\}_{i=1}^n$ , où  $x_i$  et  $y_i$  sont respectivement, l'indice de pertinence et celui de la stabilité mesurés du concept  $i$ .

1. Source : <http://www.upriss.org.uk/fca/examples.html>

### Métriques d'évaluation

Les deux mesures suivantes ont été prises en considération pour évaluer la performance (en temps d'exécution) et la précision du  $\mathcal{CR}$  (c'est-à-dire sa capacité à évaluer la pertinence des concepts) :

- Le coefficient de corrélation de Pearson  $\xi$  (et son diagramme de dispersion) :

$$\xi = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)} \sqrt{(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} \quad (4.1)$$

Où  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  et  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  sont les valeurs moyennes des mesures de pertinence obtenues à partir des deux indices respectivement.

- Le temps CPU moyen ( $\tau$ ) :

$$\tau = \frac{\sum_{i=1}^n t_i}{n} \quad (4.2)$$

Où  $t_i$  est le temps CPU pour calculer l'indice de pertinence du concept  $c_i$ .

### 4.1.2 Résultats

Pour automatiser le calcul de l'indice de pertinence des concepts, la détection des nœuds centraux (objet de type-1) et les communautés de base, nous avons développé un programme en code Python dont quelques extraits sont présentés dans les annexes A.1 et A.2. Les résultats obtenus lors de ces expérimentations sont issus de l'exécution de notre algorithme sur un *MacBook Pro* à 16 Go de mémoire vive avec un processeur *Apple M1 Pro* à dix coeurs CPU.

### Expérimentation I

Pour le jeu de données *Coin-Toss*, nous avons obtenu les résultats suivants :

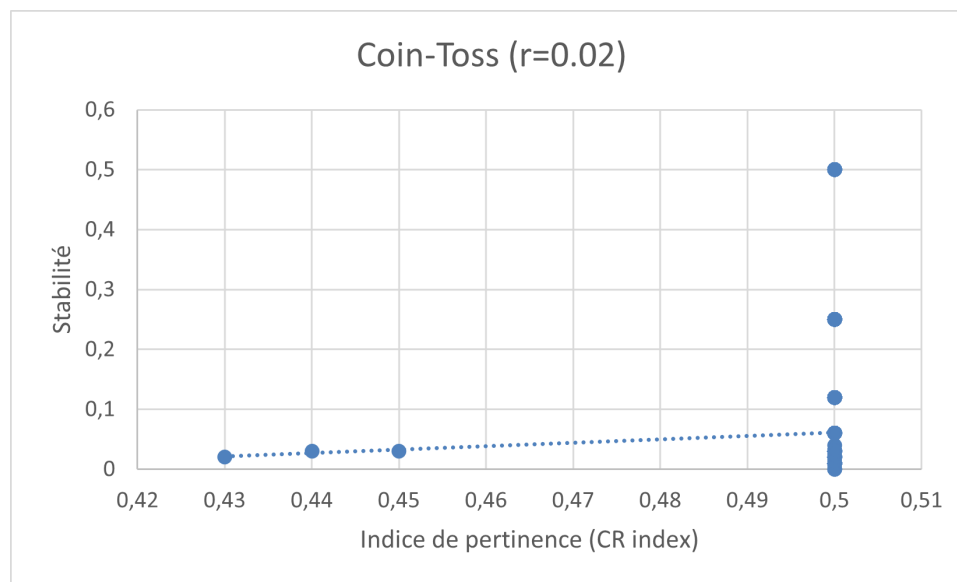


FIGURE 4.1 – Diagramme de dispersion entre l’indice de pertinence et l’indice de stabilité pour le jeu de données *Coin-Toss*.

Les résultats sont les suivants pour le jeu de données *Woman-Southern-Davis* :

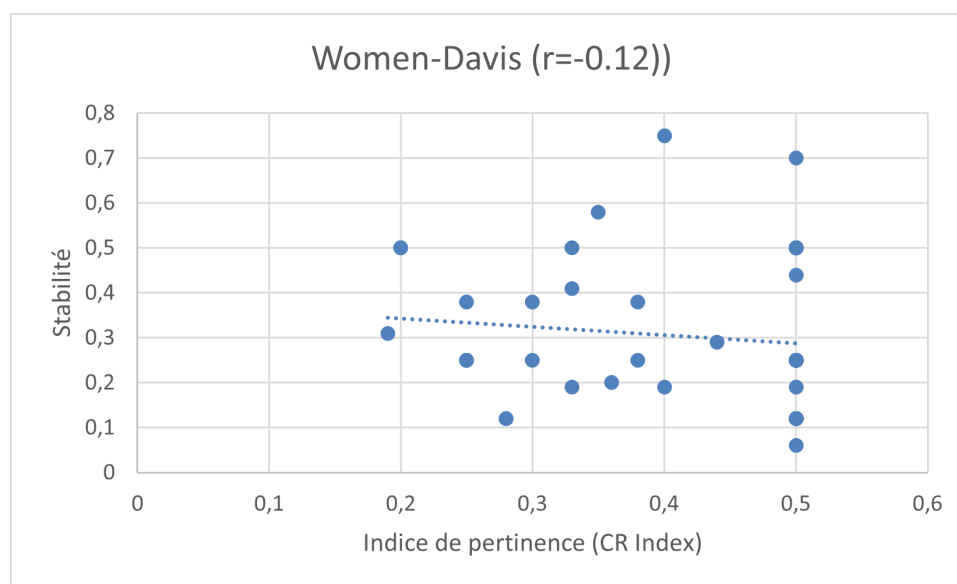


FIGURE 4.2 – Diagramme de dispersion entre l’indice de pertinence et l’indice de stabilité pour le jeu de données *Woman-Southern-Davis*

Pour le jeu de données “Phytothérapie”, nous avons obtenu les résultats suivants :

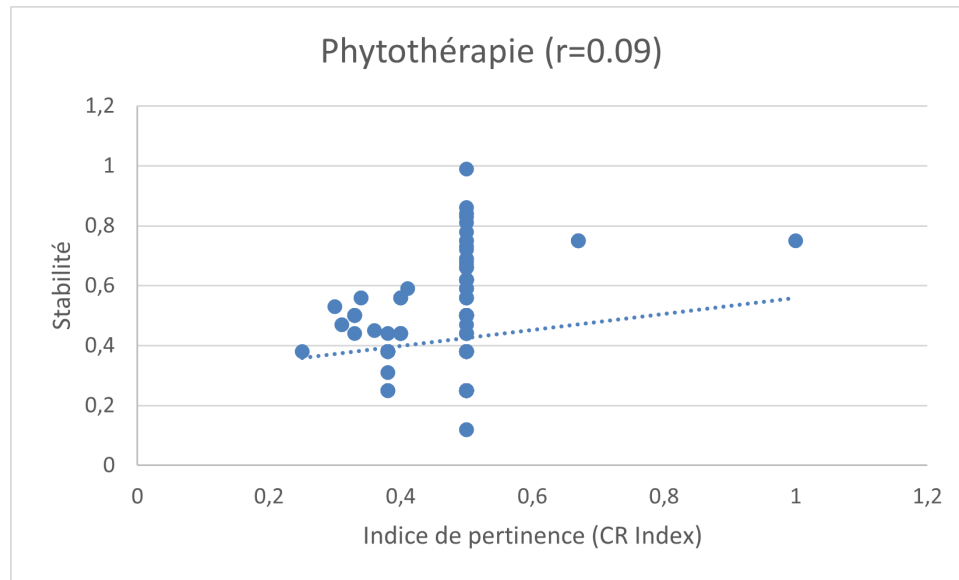


FIGURE 4.3 – Diagramme de dispersion entre l’indice de pertinence et l’indice de stabilité pour le jeu de données “*Phytothérapie*”

L’objectif premier de la corrélation de Pearson est d’évaluer la relation linéaire entre l’indice  $\mathcal{CR}$  et celui de la stabilité. Contrairement à la stabilité,  $\mathcal{CR}$  prend en compte deux types d’information distinctes : les attributs conceptuellement pertinents et les générateurs minimaux. L’étude de la relation linéaire entre ces deux indices donne une indication du degré de similitude ces deux types d’information du  $\mathcal{CR}$  et les générateurs de la stabilité. Ainsi, le fait qu’il existe des corrélations faibles et négatives entre les indices montre clairement que  $\mathcal{CR}$  capture certaines caractéristiques importantes ou pertinentes pour évaluer la qualité du concept, ce que la stabilité ne fait pas. Cela laisse croire que l’indice  $\mathcal{CR}$  renforce la pertinence conceptuelle des concepts.

## Expérimentation II

Notre deuxième expérimentation consiste à mesurer le temps d’exécution global de notre algorithme, comparé à celui de l’indice de stabilité. Les résultats sont présentés sur les graphiques suivants :

- Pour le jeu de données *Coin-Toss*

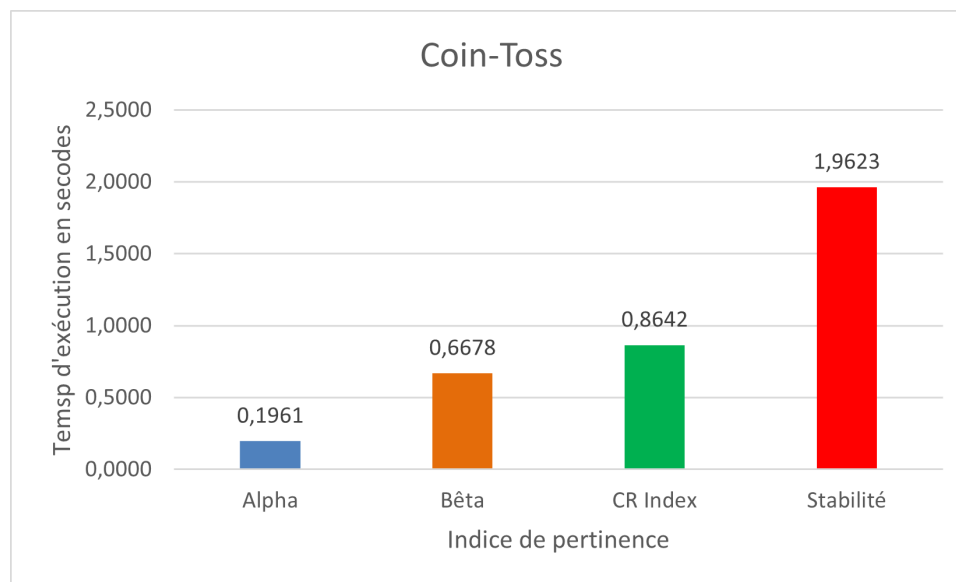


FIGURE 4.4 – Temps moyen d'exécution de l'indice de pertinence et l'indice de stabilité pour le jeu de données *Coin-Toss*.

- Pour le jeu de données *Woman-Davis*

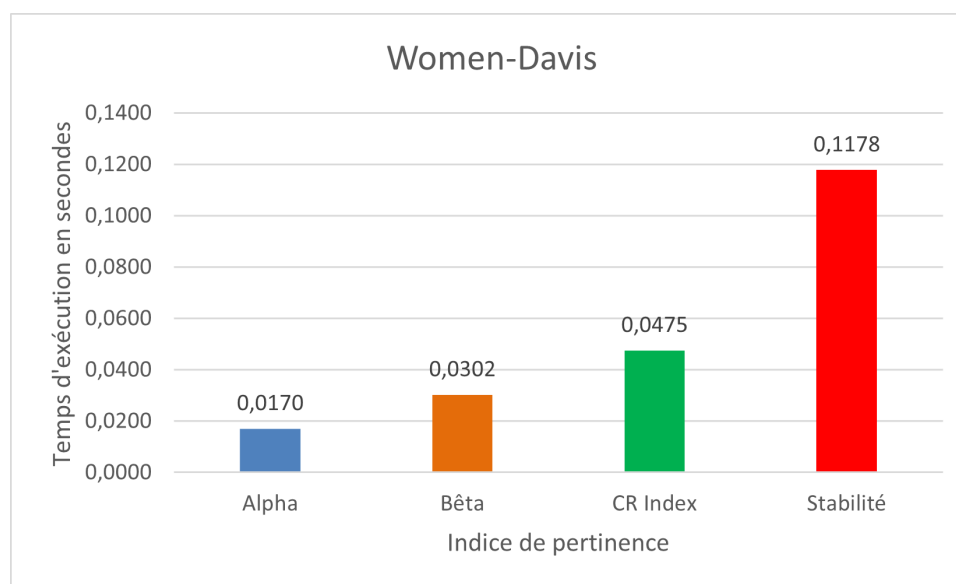


FIGURE 4.5 – Temps moyen d'exécution de l'indice de pertinence et l'indice de stabilité pour le jeu de données *Woman-Southern-Davis*.

- Pour le jeu de données "*Phytothérapie*"



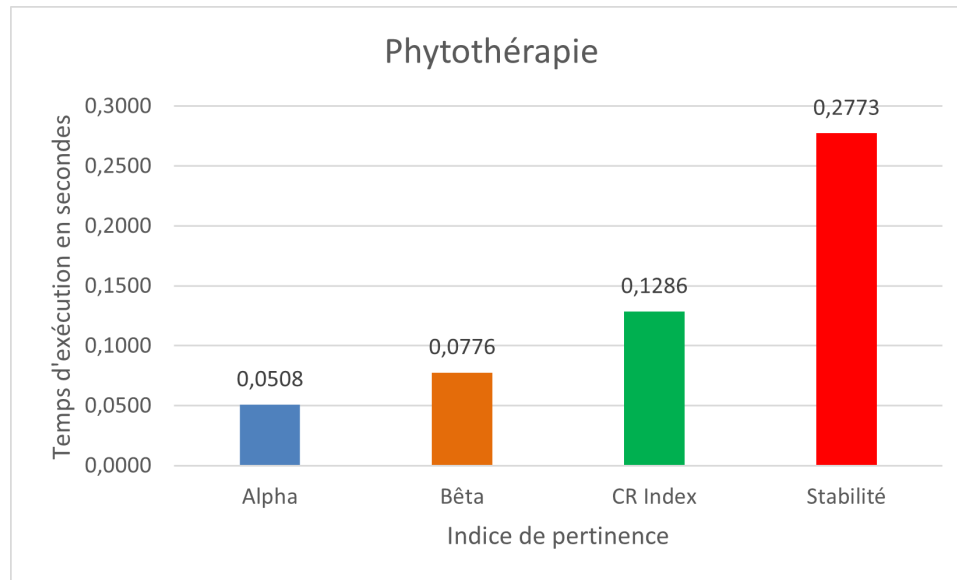


FIGURE 4.6 – Temps moyen d’exécution de l’indice de pertinence et l’indice de stabilité pour le jeu de données “*Phytothérapie*”.

### Expérimentation III

Dans cette expérimentation, nous nous penchons sur la question **Q3**. À cet effet, nous avons élaboré un algorithme permettant de détecter les nœuds de type-1 issus des concepts pertinents, identifiés à partir de l’indice de pertinence ( $\mathcal{CR}_{ex}$ ). Nos résultats sont alors comparés à l’état de l’art, notamment les mesures de centralité comme la centralité de degré ( $D_c$ ), la centralité de proximité ( $C_c$ ) et la centralité d’intermédiarité ( $B_c$ ) qui sont les mesures de centralité connues. Dans un premier temps, les différentes mesures ont été calculées et renseignées dans le tableau 4.2. Pour la centralité de degré et la centralité de proximité, les objets les plus pertinents sont  $\{1\}$ ,  $\{3\}$ ,  $\{14\}$ ,  $\{2\}$ ,  $\{4\}$ ,  $\{13\}$  et  $\{12\}$  alors que pour la centralité d’intermédiarité, ces objets sont  $\{14\}$ ,  $\{1\}$ ,  $\{3\}$ ,  $\{13\}$ ,  $\{2\}$ ,  $\{4\}$ ,  $\{12\}$  contre  $\{3\}$ ,  $\{1\}$ ,  $\{2\}$ ,  $\{4\}$ ,  $\{7\}$ ,  $\{9\}$  et  $\{14\}$  pour la centralité basée sur la pertinence des concepts. Sans toutefois tenir compte de l’ordre de ces nœuds, nous nous apercevons qu’ils sont les mêmes pour les trois premières mesures alors que deux nouveaux nœuds ( $\{7\}$  et  $\{9\}$ ) ont été identifiés par notre approche, en remplacement des nœuds  $\{12\}$  et  $\{13\}$  des autres mesures. Mais plus les nœuds à identifier sont nombreux, plus les mesures traditionnelles et notre approche basée sur la pertinence des concepts se rapprochent considérablement. Ce qui nous mène à utiliser le

coefficient de Kendall pour s'assurer de cette probable corrélation. Le tableau 4.3 nous montre les résultats obtenus.

Objets	$D_c$	$C_c$	$B_c$	$Freq_{CR}$
1	0.0909	0.0210	0.0924	0.4286
2	0.0795	0.0194	0.0491	0.4286
3	0.0909	0.0210	0.0838	0.7143
4	0.0795	0.0194	0.0474	0.4286
5	0.0455	0.0160	0.0101	0.1429
6	0.0455	0.0160	0.0103	0.1429
7	0.0455	0.0160	0.009	0.4286
8	0.0341	0.0151	0.0065	0.2857
9	0.0455	0.0160	0.0161	0.4286
10	0.0455	0.0160	0.0151	0.2857
11	0.0455	0.0160	0.0156	0.2857
12	0.0682	0.0182	0.0453	0.2857
13	0.0795	0.0194	0.0686	0.2857
14	0.0909	0.0210	0.1086	0.4286
15	0.0568	0.0170	0.0406	0.1429
16	0.0227	0.0143	0.0019	0.1429
17	0.0227	0.0143	0.0048	0.1429
18	0.0227	0.0143	0.0048	0.1429

TABLEAU 4.2 – Tableau des calculs des mesures de centralité de degré ( $D_c$ ), de la centralité de proximité ( $C_c$ ) et de la centralité d'intermédiarité ( $B_c$ )

Ce tableau nous recense les valeurs normalisées des différentes mesures de centralité y compris celle issue de la pertinence des concepts (colonne  $Freq_{CR}$ ), en application des équations 2.7, 2.10 et 2.13 pour les colonnes  $D_c$ ,  $C_c$  et  $B_c$ .

	$Freq_{CR}$
$D_c$	0.62
$C_c$	0.62
$B_c$	0.57

TABLEAU 4.3 – Coefficient de Kendall entre les mesures de centralité de degré ( $D_c$ ), de la centralité de proximité ( $C_c$ ), la centralité d’intermédiarité ( $B_c$ ) et la centralité basée sur la pertinence des concepts  $Freq_{CR}$

### 4.1.3 Discussions

Lors de l’expérimentation I, nous avons essayé d’obtenir une relation entre l’indice de pertinence et l’indice de stabilité, pour les jeux de données sélectionnés, en utilisant le coefficient de corrélation de Pearson  $\xi$ , selon la formule 4.1. La corrélation mesure une relation linéaire entre les deux scores (l’indice de pertinence et l’indice de stabilité) et n’implique pas que l’un soit supérieur à l’autre. Sa valeur est comprise entre -1 et 1 et il s’agit d’un moyen simple d’explorer cette relation, qui peut être, d’ailleurs, non linéaire en fonction du contexte des ensembles de données. Pour une corrélation linéaire, une valeur positive indique que des scores plus élevés pour un indice tendent à être associés à des scores plus élevés pour l’autre, tandis qu’une corrélation négative indique le contraire (c’est-à-dire que des scores plus élevés pour l’un tendent à être associés à des scores moins élevés pour l’autre). Si le coefficient de corrélation est proche de zéro, cela signifie qu’il y a peu ou pas de corrélation entre les deux indices, ce qui implique que l’importance du concept mesurée par un score n’a pas de relation linéaire cohérente avec son importance mesurée par l’autre score.

Avec le jeu de données *Coin-Toss* nous avons obtenu une corrélation positive de 0.02 entre l’indice de pertinence et l’indice de stabilité. Les deux indices évoluent alors dans le même sens c’est-à-dire, l’augmentation du score de l’un entraîne aussi l’augmentation du score de l’autre. Cette observation est aussi valable pour le jeu de données *Phytothérapie* où la valeur du coefficient de Pearson est de 0.09. Cependant, avec le jeu de données *Woman-Southern-Davis*, nous observons un coefficient négatif de -0.12, ce qui nous amène à dire que les deux indices évolueraient dans le sens contraire.

À la lumière des ces résultats, nous remarquons que les coefficients de corrélation entre l’indice de pertinence et l’indice de stabilité sont très faibles et proches de zéro pour

les trois jeux de données étudiés. Nous pouvons conclure que les deux ensembles de concepts pertinents obtenus à partir de ces indices n'ont que peu ou pas de relation linéaire cohérente.

Dans l'expérimentation II, nous mesurons la performance de l'indice de pertinence par rapport à celui de la stabilité en termes de temps d'exécution. En effet, nous avons exécuté notre algorithme une dizaine de fois sur chacun des jeux de données et avons pris la moyenne du temps nécessaire pour effectuer le calcul des termes  $\alpha$ ,  $\beta$ , l'indice de pertinence  $\mathcal{CR}_{ex}$  et l'indice de stabilité. Les différents résultats de notre expérimentation peuvent être observés sur les figures 4.4, 4.5 et 4.6 à partir desquelles nous faisons les observations suivantes :

- Pour le jeu de données *Coin-Toss* : notre algorithme s'exécute en 0.19 secondes pour la valeur du terme  $\alpha_{ex}$ , 0.66 secondes pour le terme  $\beta_{ex}$ , 0.86 pour l'indice de pertinence  $\mathcal{CR}_{ex}$  et 1.96 secondes pour la stabilité.
- Pour le jeu de données *Woman-Southern-Davis* : notre algorithme prend 0.01 secondes pour calculer la valeur du terme  $\alpha_{ex}$ , 0.03 secondes pour celle du terme  $\beta_{ex}$ , 0.04 pour l'indice de pertinence  $\mathcal{CR}_{ex}$  et 0.11 secondes pour la valeur de la stabilité.
- Pour le jeu de données *Phytothérapie* : notre algorithme s'exécute en 0.05 secondes pour la valeur du terme  $\alpha_{ex}$ , 0.07 secondes pour le terme  $\beta_{ex}$ , 0.12 pour l'indice de pertinence  $\mathcal{CR}_{ex}$  et 0.27 secondes pour l'indice de stabilité.

Si nous prenons la valeur du terme  $\beta_{ex}$  qui est une amélioration de la stabilité, nous remarquons une réduction du temps d'exécution de 66% par rapport à la stabilité dans le cas du jeu de données *Coin-Toss*. Cette même remarque est valable pour les jeux de données *Woman-Southern-Davis* et *Phytothérapie* qui ont une réduction du temps d'exécution respectivement de 74% et 70%.

Tous ces résultats montrent que l'indice de pertinence est plus rapide que la stabilité en termes du temps d'exécution. Cette performance s'explique par le fait que, dans l'indice de pertinence, le calcul de la composante  $\alpha$  est peu coûteux. Dans la formule de  $\beta$ , nous avons fait plutôt usage des générateurs minimaux de l'intention  $B$  d'un concept, contrairement à la production des sous-ensembles de l'intention  $B$  dans le cas de l'indice de stabilité. Cela est dû au fait que l'exploration des parties de  $2^{|B|}$  parties consomme plus de temps et est pénalisant lorsque l'intention  $B$  est de grande taille.

Dans l'expérience III, l'accent est mis sur l'identification des nœuds de type 1 dans le réseau de *Woman-Southern-Davis* en se fondant sur la pertinence des concepts et en

---

comparant notre approche avec les mesures de centralité les plus connues. Les résultats issus de cette expérimentation montrent que, sur les 7 nœuds les plus pertinents identifiés par les mesures traditionnelles, 5 font partie de la liste des nœuds trouvés par notre approche, soit un taux de détection de 71%. Cela peut s'expliquer aussi bien par le jeu de données utilisé que par sa taille ou la densité du réseau. Quant au coefficient de Kendall, nous retrouvons une corrélation de 0.62 pour la centralité de degré et la centralité de proximité et 0.57 pour la centralité d'intermédiarité. Ces valeurs positives de la corrélation nous permettent de dire que l'utilisation de la pertinence des concepts pour ensuite identifier les objets (les nœuds de type 1) centraux est prometteuse.

D'autres tests seront envisagés dans le futur pour mieux valider les résultats obtenus.

# Chapitre 5

## Conclusion

Dans le cadre de ce mémoire, nous avons exploité et amélioré une nouvelle mesure de pertinence de groupes (concepts formels) développée par l'équipe du LARIM [24] en vue d'une part, d'identifier les communautés dans les réseaux à deux modes de données (deux types de nœuds), et d'autre part, de cibler les nœuds les plus importants au sein de chacune des communautés identifiées, en utilisant l'analyse formelle de concepts (AFC). Cette recherche se distingue par l'intégration de l'AFC dans le domaine de l'analyse des réseaux sociaux. Cette nouvelle intégration met en évidence la polyvalence de l'AFC en tant qu'outil permettant non seulement d'améliorer les techniques d'analyse des réseaux, mais aussi de repousser les frontières de l'extraction de connaissances dans le domaine des réseaux sociaux.

Pour identifier les communautés, et en considérant la version extensionnelle des formules, nous avons d'abord apporté une amélioration aux équations 2.3 et 2.4 en ajoutant une composante nous permettant de tenir compte de tous les attributs conceptuellement pertinents et en modifiant la composante  $\beta$ . Un raisonnement similaire a aussi été appliqué à la version intentionnelle des formules.

Bien que Kuznetsov [30] affirme que l'index de stabilité aboutit à des groupes cohésifs, nous pensons que notre mesure de pertinence, combinée avec d'autres mesures tel que l'indice de séparation a donné des résultats plus concluants permettant de vérifier les deux propriétés du regroupement, soit la cohésion et la distinction. Le manque de corrélation entre l'indice de pertinence des concepts et la stabilité de Kuznetsov implique que notre indice renforce la pertinence sémantique des concepts.

Quant à l'identification des nœuds centraux du réseau et correspondant en fait aux objets importants dans le treillis de concepts, nous avons d'abord détecté les concepts

pertinents constituant les communautés de base en calculant l'autonomie des concepts qui est la moyenne harmonique de l'indice de pertinence et de l'indice de séparation. Les résultats sont ensuite triés par ordre décroissant de l'autonomie de telle sorte que l'union des concepts identifiés couvre l'ensemble des objets du treillis. Avec ces concepts pertinents, nous avons élaboré un algorithme qui fait le comptage des objets se retrouvant dans l'extension des concepts pertinents.

Rappelons que toutes les expérimentations ont été menées avec l'exemple du contexte formel représenté dans le tableau 2.1 et ensuite, avec le contexte de *Woman-Southern-Davis* (tableau 3.7) afin de comparer nos résultats par rapport à quelques méthodes déjà existantes dans la détection des communautés. L'algorithme proposé a été aussi capable d'identifier des communautés de petite taille comme le montre le tableau 3.10. Les résultats obtenus nous permettent d'affirmer que les améliorations apportées à la formule de base sont importantes et que ces résultats dépendent non seulement de la densité du treillis mais aussi du nombre d'attributs. Cependant, nous estimons que d'autres tests sur des réseaux sociaux réels et synthétiques permettront plus tard de valider et d'affiner la démarche.

Parmi les travaux que nous comptons mener dans le futur, nous énumérons les suivants :

- Adaptation de la notion de “extremal point” définie dans [21] à l'identification des attributs conceptuellement pertinents de l'intention d'un concept  $c = (A, B)$  en vue de comparer, au niveau de la précision et de la performance en temps de calcul, notre indice extensionnel de pertinence du concept  $c$  avec sa contrepartie dans laquelle le numérateur de  $\alpha_{ex}$  est un nombre de points extrêmes dans  $B$
- Élaboration d'une méthode de percolation pour les réseaux à deux modes de données en vue de fusionner des communautés lorsque leur nombre est grand
- Adaptation de la mesure Bi-Face [25] en vue d'identifier des objets centraux d'une manière locale (et non globale) au sein d'une communauté identifiée.

# Annexe A

```
def get_upper_neighbors(Extent, Intent, conte):
    Upper_neighbors_concept = []
    LL = conte.lattice[Convert(Intent)].upper_neighbors
    for j in range(len(LL)):
        lc = []
        lc = find_between(str(LL[j]).replace(" ", ""), "{", "}").split(",")
        lc = list(filter(lambda a: a != ';', lc))
        lc = list(filter(lambda a: a != "'", lc))
        if lc == []:
            lc = []
        lc_o = list(conte.intension(lc))
        Upper_neighbors_concept.append(lc_o)
    print('Upper Covers :', Upper_neighbors_concept)
    return Upper_neighbors_concept

#-----

def minimal_generators(gen):
    minimal_list = []
    if len(gen) > 1:
        for e1 in range(len(gen)):
            count = 0
            for e2 in range(len(gen)):
                if gen[e1] != gen[e2] and set(gen[e1]).issuperset(set(gen[e2])):
                    count = count + 1
            if count == 0:
                minimal_list.append(gen[e1])
    else:
        minimal_list.append(gen)
    return minimal_list
```

FIGURE A.1 – Exemple de code Python pour le calcul des générateurs minimaux



```

print("\nThe Conceptual Relevance (intensional) for the context: ', filename)
print('=====')
print('=====')
i=1
for ext,inte in c.lattice:
    Intent = list(inte)
    Extent = list(ext)
    if len(Extent) > 0 :
        print('C'+str(i), '(,Extent,',',Intent,)', "\n")
        print ('Intent = B: ', Intent)
        print ('Extent = A: ', Extent, "\n")
        generateur = Generators_(Intent, c)
        print('Minimal Generator : ', minimal_generators(generateur), "\n")
        alpha_in = round(CR_intention(Intent,c), 2)
        beta_in = round(St_intention(Extent,Intent,c), 2)
        print('Alpha(in) =',alpha_in, 'Beta(in) =',beta_in)
        print('-----')
        i=i+1
#|

```

FIGURE A.2 – Exemple de code Python pour l’affichage de  $\alpha_{ex}(c)$  et  $\beta_{ex}(c)$

## Bibliographie

- [1] ABIR, M. *Détection des communautés dans les réseaux complexes*. PhD thesis, Université du Québec en Outaouais, 2020.
- [2] AGARWAL, G., AND KEMPE, D. Modularity-maximizing graph communities via mathematical programming. *The European Physical Journal B* 66, 3 (2008), 409–418.
- [3] ALPEN, É. *Précis de Phytothérapie*. Édition Alpen, 2010.
- [4] ALVES, N. A. Unveiling community structures in weighted networks. *Physical Review E* 76, 3 (2007), 036101.
- [5] BOCCALETTI, S., IVANCHENKO, M., LATORA, V., PLUCHINO, A., AND RAPSARDA, A. Detecting complex network modularity by dynamical clustering. *Physical Review E* 75, 4 (2007), 045102.
- [6] BORGATTI, S. P., EVERETT, M. G., AND FREEMAN, L. C. Ucinet for windows : Software for social network analysis. *Harvard, MA : analytic technologies* 6 (2002), 12–15.
- [7] BORGATTI, S. P., AND HALGIN, D. S. Analyzing affiliation networks. *The Sage handbook of social network analysis* 1 (2011), 417–433.
- [8] BOTHOREL, C., CRUZ, J. D., MAGNANI, M., AND MICENKOVA, B. Clustering attributed graphs : models, measures and methods. *Network Science* 3, 3 (2015), 408–444.
- [9] BRANDES, U. A faster algorithm for betweenness centrality. *Journal of mathematical sociology* 25, 2 (2001), 163–177.
- [10] DAVIS, A., GARDNER, B., AND GARDNER, M. A. davis and b.b. gardner and m.r. gardner. Deep South, 1941.
- [11] DAVIS, A., GARDNER, B., AND GARDNER, M. *Deep South*. The University of Chicago Press, 1941.

- [12] DONG, G., JIANG, C., PEI, J., LI, J., AND WONG, L. Mining succinct systems of minimal generators of formal concepts. In *International Conference on Database Systems for Advanced Applications* (2005), Springer, pp. 175–187.
- [13] FELDE, M., HANIKA, T., AND STUMME, G. Null models for formal contexts. *Information 11*, 3 (2020), 135.
- [14] FORTUNATO, S. Community detection in graphs. *Physics reports 486*, 3-5 (2010), 75–174.
- [15] FORTUNATO, S., LATORA, V., AND MARCHIORI, M. Method to find community structures based on information centrality. *Physical review E 70*, 5 (2004), 056104.
- [16] FRANÇOIS, O., AND MICHEL, V. D. K. De la pyramide au réseau? pour une théorie dialectique du droit. In *Publication des Facultés universitaires Saint Louis*, no. 14, Bruxelles, 2002, p. 24.
- [17] FREEMAN, L. Finding groups : A meta-analysis of the southern women data. *Dynamic Social Network Modeling and Analysis* (01 2003).
- [18] FREEMAN, L. C. Finding social groups : A meta-analysis of the southern women data. In *Dynamic Social Network Modelling and Analysis : Workshop Summary and Papers*, K. C. Ronald Breiger and P. Pattison, Eds. The National Academies Press, 2003, pp. 1–60.
- [19] FRIEDMAN, J., HASTIE, T., AND TIBSHIRANI, R. The elements of statistical learning.(2001). *NY : Springer Series in Statistics.[Google Scholar]*.
- [20] GANTER, B., AND WILLE, R. *Formal Concept Analysis : Mathematical Foundations*. Springer-Verlag New York, Inc., 1999.
- [21] GANTER, B., AND WILLE, R. *Formal Concept Analysis : Mathematical Foundations*. Springer-Verlag New York, Inc., 1999. Translator-C. Franzke.
- [22] GIRVAN, M., AND NEWMAN, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences 99*, 12 (2002), 7821–7826.
- [23] HAN, J., PEI, J., AND TONG, H. *Data Mining : Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2022.

- [24] IBRAHIM, M.-H., MISSAOUI, R., AND VAILLANCOURT, J. Conceptual relevance index for identifying actionable formal concepts. In *International Conference on Conceptual Structures* (2021), Springer, pp. 119–126.
- [25] IBRAHIM, M. H., MISSAOUI, R., AND VAILLANCOURT, J. Identifying influential nodes in two-mode data networks using formal concept analysis. *IEEE Access* 9 (2021), 159549–159565.
- [26] KERNIGHAN, B. W., AND LIN, S. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal* 49, 2 (1970), 291–307.
- [27] KLIMUSHKIN, M., OBIEDKOV, S. A., AND ROTH, C. Approaches to the selection of relevant concepts in the case of noisy data. In *ICFCA* (2010), vol. 20, Springer, pp. 255–266.
- [28] KUZNETSOV, S. Interpretation on graphs and complexity characteristics of the problems of finding regularities of a certain type. *Nauchn.-Tekhn. Inform., Ser 2* (1989), 23–28.
- [29] KUZNETSOV, S., OBIEDKOV, S., AND ROTH, C. Reducing the representation complexity of lattice-based taxonomies. In *International Conference on Conceptual Structures* (2007), Springer, pp. 241–254.
- [30] KUZNETSOV, S. O., AND MAKHALOVA, T. On interestingness measures of formal concepts. *Information Sciences* 442 (2018), 202–219.
- [31] KUZNETSOV, S. O., AND MAKHALOVA, T. P. Concept interestingness measures : a comparative study. In *CLA* (2015), vol. 1466, pp. 59–72.
- [32] KUZNETSOV, S. O., AND OBIEDKOV, S. A. Comparing performance of algorithms for generating concept lattices. *Journal of Experimental & Theoretical Artificial Intelligence* 14, 2-3 (2002), 189–216.
- [33] MANNILA, H., TOIVONEN, H., AND VERKAMO, A. I. Efficient algorithms for discovering association rules. In *KDD-94 : AAAI workshop on Knowledge Discovery in Databases* (1994), pp. 181–192.
- [34] MISSAOUI, R., AND EMAMIRAD, K. Lattice miner-a formal concept analysis tool. In *14th International Conference on Formal Concept Analysis* (2017), p. 91.

- [35] MISSAOUI, R., MESSAOUDI, A., AND IBRAHIM, MOHAMED HAMZA ABDESSALEM, T. *Advances in Knowledge Discovery and Management : Volume 9*, vol. 9. Springer, March 2022, ch. Detecting Communities in Multilayer Networks using Formal Concept Analysis, pp. 77–105.
- [36] NEWMAN, M. E., AND GIRVAN, M. Finding and evaluating community structure in networks. *Physical review E* 69, 2 (2004), 026113.
- [37] NJIWOUA, P., AND NGUIFO, E. M. Treillis de concepts et classification supervisée. *Technique et Science Informatiques* 24, 4 (2005), 449–488.
- [38] PALLA, G., DERÉNYI, I., FARKAS, I., AND VICSEK, T. Uncovering the overlapping community structure of complex networks in nature and society. *nature* 435, 7043 (2005), 814–818.
- [39] SHI, C., LI, Y., ZHANG, J., SUN, Y., AND PHILIP, S. Y. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2017), 17–37.
- [40] STUMME, G., TAOUIL, R., BASTIDE, Y., PASQUIER, N., AND LAKHAL, L. Computing iceberg concept lattices with titanic. *Data & knowledge engineering* 42, 2 (2002), 189–222.
- [41] TATTI, N., MOERCHEN, F., AND CALDERS, T. Finding robust itemsets under subsampling. *ACM Transactions on Database Systems (TODS)* 39, 3 (2014), 20.