



HAL
open science

Modélisation du contexte social : application aux réseaux opportunistes

Yaofu Cao

► **To cite this version:**

Yaofu Cao. Modélisation du contexte social : application aux réseaux opportunistes. Sociologie. Université de Technologie de Troyes, 2017. Français. NNT : 2017TROY0002 . tel-02410058

HAL Id: tel-02410058

<https://theses.hal.science/tel-02410058>

Submitted on 13 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse
de doctorat
de l'UTT

Yaofu CAO

Modélisation du contexte social : application aux réseaux opportunistes

Spécialité :

**Ingénierie Sociotechnique des Connaissances, des Réseaux
et du Développement Durable**

2017TROY0002

Année 2017



THESE

pour l'obtention du grade de

DOCTEUR de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

**Spécialité : INGENIERIE SOCIOTECHNIQUE DES CONNAISSANCES,
DES RESEAUX ET DU DEVELOPPEMENT DURABLE**

présentée et soutenue par

Yaofu CAO

le 24 janvier 2017

**Modélisation du contexte social :
application aux réseaux opportunistes**

JURY

M. P. LORENZ	PROFESSEUR DES UNIVERSITES	Président (Rapporteur)
Mme I. FAJJARI	DOCTEURE	Examineur
M. A. GIBOIN	CHARGE DE RECHERCHE INRIA	Examineur
M. G. PUJOLLE	PROFESSEUR DES UNIVERSITES	Rapporteur

Personnalités invitées

Mme D. GAÏTI	PROFESSEUR DES UNIVERSITES	Directrice de thèse
M. P. SALEMBIER	PROFESSEUR UTT – HDR	Directeur de thèse

Résumé

Cette thèse traite de l'adaptation automatique des applications sensibles au contexte par l'utilisation d'informations liées à l'environnement social des utilisateurs afin d'enrichir le service rendu par les applications. Pour cela, notre contribution s'articule autour de la modélisation multidimensionnelle des différents niveaux de contextes sociaux, notamment le poids de la relation entre les acteurs. Plus spécifiquement, nous synthétisons des contextes sociaux non seulement liés à la familiarité mais aussi liés à la similitude des communautés statiques et dynamiques. Deux modèles basés respectivement sur les graphes et les ontologies sont proposés afin de satisfaire l'hétérogénéité des réseaux sociaux de la vie réelle. Nous utilisons les données réelles recueillies sur les réseautages sociaux en ligne pour conduire nos expérimentations et analysons les résultats en vérifiant l'efficacité de ces modèles. En parallèle nous traitons le point de vue de l'application, et nous présentons deux algorithmes utilisant des contextes sociaux pour améliorer la stratégie de transmission des données dans le réseau opportuniste, et particulièrement la contre-mesure aux nœuds égoïstes. Les simulations des scénarios réels confirment les avantages liés à l'introduction des contextes sociaux, en termes de taux de succès et de délais de transmission. Nous effectuons une comparaison avec d'autres algorithmes de transmission traditionnellement décrits dans la littérature pour compléter notre démonstration.

Mots Clés : Contexte (linguistique), Prise de conscience, Contexte (linguistique) -- Aspect social, Réseaux sociaux, Systèmes de communication sans fil, Données – Transmission

Abstract

This thesis deals with the dynamic adaptation of context-aware applications using information related to the social environment of users to enrich the service rendered by the applications. To achieve this goal our contribution mobilizes multidimensional modeling of the different levels of social contexts, especially the weight of the relationship between the actors. Particularly, we synthesize not only social contexts related to familiarity but also social contexts reasoned from the similarity of static and dynamic communities. Two models based on respectively graphs and ontologies are proposed in order to satisfy the heterogeneity of the social networks in real life. We use the actual data gathered on online social networking services for conducting experiments and the results are analyzed by checking the effectiveness of the models. In parallel we consider the point of view of the application, and we present two algorithms using social contexts to improve the strategy of transmission of data in the opportunistic network, particularly countermeasure against selfish nodes. The simulations of real scenarios confirm the advantages of introducing social contexts in terms of success rate and delay of transmission. We carry out a comparison with other popular transmission algorithms in the literature.

Keywords: Context (linguistics), Awareness, Context (linguistics) -- Social aspects, Social Networks, Wireless communication systems, Data transmission systems

Remerciements

Je voudrais exprimer ma plus profonde gratitude à mes superviseurs Dominique GAITI et Pascal SALEMBIER. Merci à eux de m'avoir donné l'occasion de démarrer et de profiter de ce voyage de recherche. Ils m'ont aidé à préciser le sujet de la thèse et m'ont prodigué soutien continu, conseils et encouragements. Ils ont toujours été si généreux et patients, c'était vraiment agréable de travailler avec eux. Je souhaite également les remercier pour le temps qu'ils ont su me consacrer pour proposer corrections, remarques et conseils avisés sur la rédaction de la thèse et la présentation de mes travaux.

Merci au China Scholarship Council (CSC), pour leur soutien financier. Merci aux laboratoires ERA et Tech-CICO pour l'aide financière qu'ils m'ont consentie durant les derniers mois avant la soutenance, et ce grâce à mes superviseurs.

J'ai également eu une chance de bénéficier d'échanges avec d'excellents collègues des laboratoires ERA et Tech-CICO. En particulier, je tiens à remercier Jason, Hocine et Malte, qui ont toujours été gentils en m'aidant à améliorer ma compréhension de la culture internationale.

En outre, je voudrais remercier Fei et Yanyan pour les discussions, source d'inspiration et les suggestions constructives pendant le cours de mon travail de thèse.

Beaucoup d'amis chinois m'ont apporté beaucoup de joie durant les trois années et demie de cette thèse. Il est impossible d'énumérer tous leurs noms et ce que je peux dire est: merci à toutes et tous pour votre accompagnement.

Enfin, je voudrais exprimer ma plus profonde gratitude à ma femme Heping mais il y aurait trop de choses à écrire ici.

Table des Matières

RESUME	II
ABSTRACT	IV
REMERCIEMENTS	VI
CHAPITRE 1. INTRODUCTION	1
1.1 CHALLENGE	2
1.2 MOTIVATION ET CONTRIBUTION	3
1.3 ORGANISATION	6
CHAPITRE 2. 'CONTEXT AWARENESS'	7
2.1 LE COURANT DU 'CONTEXT AWARENESS'	7
2.2 DEFINITIONS ET CATEGORIES DE CONTEXTE	10
2.2.1 <i>Classification par mode d'acquisition</i>	12
2.2.2 <i>Classification par niveau</i>	12
2.2.3 <i>Classification par type</i>	14
2.3 UTILISATION DU CONTEXTE AVEC L'EVOLUTION DES RESEAUX	15
2.3.1 <i>Applications desktop</i>	15
2.3.2 <i>Applications web</i>	15
2.3.3 <i>Informatique mobile (Mobile computing)</i>	16
2.3.4 <i>Informatique ubiquitaire (Ubiquitous computing)</i>	16
2.3.5 <i>L'internet des objets (Internet of Things : IoT)</i>	17
2.4 EXIGENCES POUR UN SYSTEME CONTEXTUEL	18
2.4.1 <i>Acquisition de contexte</i>	18
2.4.2 <i>Modélisation du contexte</i>	19
2.4.2.1 Méthode de paires attributs-valeurs (Key-value pair Model)	20
2.4.2.2 Modèle de balisage (Make-up Scheme Model)	21
2.4.2.3 Modèle graphique	21
2.4.2.4 Modèle orienté objet (Object Oriented Model)	22
2.4.2.5 Modèle basé sur la logique (Logic Based Model)	23
2.4.2.6 Modèle basé sur les ontologies	24
2.4.3 <i>Inférence contextuelle</i>	25
2.5 LES CARACTERISTIQUES ET LIMITATIONS DE CONTEXTE DE LA NATURE 'PHYSIQUE'	26
CHAPITRE 3. CONTEXTE SOCIAL	27
3.1 INTRODUCTION: OPPORTUNITE ET CHALLENGE POUR UNE VISION SOCIALE	27
3.2 TROIS DEFIS PRINCIPAUX POUR L'UTILISATION DU CONTEXTE SOCIAL	30
3.2.1 <i>Standards de représentation unifiée</i>	30
3.2.2 <i>Capacité d'inférence</i>	30
3.2.3 <i>Vie privée - intimité</i>	31
3.3 PRISE EN COMPTE DES DIFFERENTS NIVEAUX DE CONTEXTES SOCIAUX	31
3.4 QUELQUES EXEMPLES DE REALISATION UTILISANT LE CONTEXTE SOCIAL	32
CHAPITRE 4. EVOLUTION DES RESEAUX VERS L'UTILISATION DU CONTEXTE SOCIAL	35
4.1 CONTEXTE MULTIDIMENSIONNEL ET RESEAUX SOCIAUX	35
4.2 UTILISATION DU CONTEXTE SOCIAL DANS UN RESEAU OPPORTUNISTE	37
4.2.1 <i>Le réseau opportuniste</i>	37

4.2.2	<i>Protocole de transmission</i>	39
4.3	SECURITE DE LA VIE PRIVEE DU RESEAU OPPORTUNISTE ET CONTEXTE SOCIAL.....	42
4.3.1	<i>Menace de sécurité du réseau opportuniste</i>	42
4.3.1.1	Menace de transmission	42
4.3.1.2	Menace de comportement de nœud	43
4.3.1.3	Menace de la vie privée.....	44
4.3.2	<i>Mécanisme de la confiance sur la base du contexte social</i>	45
CHAPITRE 5.	MODELISATION DE CONTEXTE SOCIAL	48
5.1	POIDS DE LA RELATION.....	49
5.2	MODELISATION DU POIDS DE LA RELATION DANS LA PRATIQUE DES RESEaux SOCIAUX	50
5.2.1	<i>Poids de la relation amenée par la familiarité</i>	53
5.2.2	<i>Poids de la relation lié à la similitude</i>	59
5.2.3	<i>Processus d’algorithme et analyse de performance</i>	62
5.2.4	<i>Acquisition de données sur le réseautage social</i>	63
5.2.5	<i>Un exemple concret de calcul de poids de relation</i>	67
5.2.6	<i>Vérification de l’efficacité du modèle</i>	69
5.3	MODELISATION DU POIDS DE LA RELATION POUR LE RESEAU SOCIAL REEL	72
5.3.1	<i>Modélisation du réseau social sémantique</i>	74
5.3.1.1	Couche de réseau de concepts de l'ontologie.....	76
5.3.1.2	Couche de réseau social.	78
5.3.2	<i>Poids de la relation entre « chercheurs »</i>	78
5.3.2.1	Dimension « citer » et évaluation de son poids de relation	78
5.3.2.2	Relation de coopération et évaluation de son poids de relation.....	79
5.3.2.3	Relation d’événement en commun et évaluation de son poids de relation.....	80
5.3.2.4	Relation de chercheur et évaluation de son poids de relation	80
5.3.3	<i>Expérimentation et analyse</i>	80
CHAPITRE 6.	APPLICATION BASEE SUR LE CONTEXTE SOCIAL POUR LES RESEaux OPPORTUNISTES	86
6.1	UTILISATION DE LA CONFIANCE DANS LE DOMAINE DES RESEaux OPPORTUNISTES	86
6.1.1	<i>Définition de la confiance pour les réseaux opportunistes</i>	86
6.1.2	<i>Gestion de confiance et modèle de confiance</i>	88
6.2	ALGORITHME DE TRANSMISSION BASE SUR LA CONFIANCE SOCIALE BINAIRE	91
6.2.1	<i>Degré de familiarité</i>	94
6.2.2	<i>Niveau d’intimité</i>	94
6.2.3	<i>Qualité de service</i>	96
6.2.4	<i>Calcul de la confiance sociale binaire</i>	97
6.2.5	<i>Le processus de transmission basé sur la confiance sociale binaire</i>	98
6.3	ALGORITHME DE TRANSMISSION DE CONFIANCE BASE SUR LA SIMILITUDE DE COMMUNAUTE DYNAMIQUE DES NŒUDS.	101
6.3.1	<i>Similitude de trajectoire</i>	103
6.3.2	<i>Similitude de relation sociale</i>	105
6.3.3	<i>Niveau de confiance de nœud</i>	107
6.3.4	<i>Processus de transmission basé sur la confiance</i>	108
CHAPITRE 7.	SIMULATION ET EXPERIMENTATION DE L’APPLICATION	111
7.1	PLATE-FORME DE SIMULATION	111
7.2	PARAMETRES D’EVALUATION	115
7.2.1	<i>Taux de succès de transmission</i>	115
7.2.2	<i>Temps de délai moyen</i>	115

7.3	SIMULATION D'ALGORITHME DE TRANSMISSION DE DONNEES BASE SUR LA CONFIANCE SOCIALE BINAIRE	116
7.3.1	<i>Environnement d'expérimentation</i>	116
7.3.2	<i>Résultat d'expérimentation et analyse</i>	117
7.4	SIMULATION DE L'ALGORITHME DE TRANSMISSION DE CONFIANCE BASE SUR LA SIMILITUDE DES COMMUNAUTES DYNAMIQUES DES NŒUDS	122
7.4.1	<i>Environnement d'expérimentation</i>	122
7.4.2	<i>Résultat d'expérimentation et analyse</i>	123
CHAPITRE 8.	CONCLUSION ET PERCEPTIVES	129
REFERENCES	133

Chapitre 1. Introduction

En tant que technologie clé dans le domaine étendu de l'intelligence artificielle, un champ de recherche en pleine croissance appelé *context awareness* [1] a pour but de fournir de l'information pour construire des services plus intelligents et accessibles. L'objectif des services envisagés est de favoriser l'émergence des environnements de *Pervasive computing* [2], où les individus sont aidés dans leurs tâches quotidiennes par des technologies qui s'exécutent en tâches de fond, de manière invisible et/ou implicites mais significativement utile. L'information fournie ici, que nous appellerons contexte, est largement valorisée aujourd'hui, et se réfère généralement à des informations sur les systèmes, les entités et leurs environnements. Les applications qui adaptent leurs comportements en fonction des changements d'informations de contexte sont appelées des applications sensibles au contexte [3].

De nombreux efforts de recherche dans le domaine des applications sensibles au contexte sont axés sur la modélisation, le raisonnement et la gestion d'informations contextuelles, afin de réduire la complexité de l'ingénierie de tels systèmes. Ces efforts ont permis d'accomplir des progrès considérables dans la formulation de modèles de contexte qui sont une condition préalable importante pour la construction de systèmes capables de répondre aux changements dans leur environnement.

Les contextes utilisés par des modèles, des architectures logicielles et plateformes regroupant des applications sensibles au contexte sont essentiellement de nature physique (emplacement, temps, activité, etc.). Mais les utilisateurs présentent également des caractéristiques sociales. De ce fait, l'introduction des contextes sociaux permet d'étendre le service rendu par les applications d'une manière plus riche et plus intelligente. Comparativement, il n'y a eu que peu de travaux qui étudient les contextes de nature sociale. Même si récemment certains travaux ont tenté de modéliser le contexte social, ils considèrent certains types de contextes sociaux limités à une source unique [66,76,81]. En outre, il existe un manque de support pour aider à la modélisation et au raisonnement sur ces différents types de contextes sociaux multi dimensions.

1.1 Challenge

Notre recherche explore le concept de contexte social en l'examinant sous différents angles afin de représenter et de raisonner sur les différents contextes sociaux, notamment le poids de la relation entre les acteurs. Dans le processus de compréhension du contexte social et la conception d'applications et de services sociaux connexes, nous devons donc faire face à divers défis sources de questions théoriques et technologiques intéressantes.

Premièrement, il est nécessaire de définir et de collecter des informations sur le contexte social de diverses sources. La représentation et la modélisation du contexte social nécessite l'accès à des relations à la fois liées à l'interaction de la vie réelle et à la connexion virtuelle. La modélisation des relations orientées vers l'interaction de la vie réelle entre les personnes dans un environnement ou domaine (par exemple, famille, bureau, domaine de recherche etc.) d'une part, et d'autre part, les relations orientées connexion des utilisateurs peuvent être acquises à partir des différentes applications des réseaux sociaux en ligne, autrement dit l'activité de réseautage social, sur des plateformes telles que Facebook, LinkedIn, Twitter, etc., car les utilisateurs utilisent ces applications pour spécifier leurs relations, mettre à jour leur statut et noter des contenus. Le défi consiste donc à modéliser et à représenter de la manière la plus complète possible tous ces différents types de contextes sociaux afin de faciliter leur utilisation par des applications.

Dans un deuxième temps, une application de contexte social nécessite des informations contextuelles de haut niveau plus ou moins abstraites (par exemple, la situation sociale, le degré de relation entre utilisateurs, etc.) qui ne sont pas directement disponibles mais peuvent être inférées à partir d'informations acquises ou prédéfinies. Le défi consiste donc à soutenir l'inférence la plus significative à partir d'un ensemble d'informations contextuelles de base. Il est alors nécessaire de travailler à partir de raisonnements basés sur les situations en analysant les activités et événements d'interaction actuels et passés en s'appuyant éventuellement sur certaines théories et connaissances issues des sciences humaines et sociales.

Troisièmement, la collecte et le traitement de données à grande échelle est nécessaire. Pour une plate-forme de réseau social donnée, le nombre d'utilisateurs existants et le volume des diverses données numériques que les utilisateurs enrichissent peuvent être très importants. La collecte de données à grande échelle est un problème pratique. Il est presque impossible de capturer un graphe complet de réseautage social avec toutes les informations téléchargées, en particulier pour une plateforme de réseau social populaire comportant un grand nombre d'utilisateurs actifs tels que Facebook ou Twitter. De plus, les techniques d'échantillonnage sont généralement utilisées pour recueillir une petite partie représentative des données à des fins de recherche. Le défi ici est de savoir comment assurer la représentativité des données échantillonnées. De plus, certaines informations des utilisateurs ne sont pas accessibles au public en raison des problèmes de confidentialité personnelle ou professionnelle.

Quatrièmement, il faut faire progresser les applications et les services en utilisant des informations sociales. Par rapport aux contextes sociaux capturés sur les sources en ligne et qui ont une structure stable, la situation sociale d'application change en temps réel. Afin de développer des applications sociales ou de faire progresser les services existants, il faut prendre en considération la caractéristique d'évolution des aspects des contextes sociaux dans l'environnement réel. En particulier pour le type d'application qui va nous servir de démonstrateur à savoir les applications de stratégie de transmission pour les réseaux opportunistes. Les réseaux opportunistes constituent un nouveau type de réseau qui peut établir une communication entre différentes personnes par rencontre opportuniste ; les nœuds du réseau opportuniste sont influencés par des comportements sociaux humains, et ils intègrent métaphoriquement certains attributs des sociétés humaines. La règle de raisonnement est de fait différente de celle des réseaux sociaux normaux.

1.2 Motivation et Contribution

Réseautage social ou réseaux sociaux en ligne ont acquis une grande popularité et ont accumulé des données numériques massives sur la société humaine. Ces données massives, représentant les informations personnelles et sociales des

individus, nous fournissent des opportunités d'étudier, d'analyser et de modéliser la structure du réseau, les connexions humaines, la similitude des personnes, etc. La façon d'extraire et d'exploiter des contextes sociaux efficaces à partir des données massives disponibles pour améliorer les applications et les services a reçu beaucoup d'attention [29,64].

Il existe actuellement une multiplicité de réseaux sociaux liés à des finalités différentes mais complémentaires et qui conduisent à des volumes d'échange d'informations importants entre personnes ou organisations : les réseaux sociaux en ligne, les réseaux de recherche, les réseaux d'entreprise et d'autres types de réseaux communautaires. Avec l'augmentation de l'activité sociale, la taille des réseaux augmente, et leur structure devient de plus en plus complexe. Les réseaux sociaux sont hétérogènes, et regroupent des relations variés et des nœuds différents et caractéristique propre à la vie proches des caractéristiques des sociétés humaines.

Nous proposons deux modèles pour modéliser respectivement le poids de relation du réseau social homogène et du réseau social hétérogène de la vie réelle. Avec ces deux modèles, nous définissons un poids de relation de réseau social multidimensionnel, c'est-à-dire que le contexte social est à la source du monde cyber mais aussi du monde physique.

Inspirée de certaines théories développées en sciences humaines et sociales (sociologie et psychologie sociale notamment), notre approche de modélisation du poids de la relation du réseautage social le distingue en poids relatif au caractère de familiarité et poids relatif au caractère de similitude, le caractère de similitude comprend la similitude de la communauté statique et la similitude de la communauté dynamique. L'ensemble des contextes sociaux de différents niveaux de raisonnement sont considérés pour mettre en œuvre l'inférence. Nous proposons une méthode spécifique de calcul basée sur les graphes. Une solution adaptée pour recueillir les données du contexte social sur les réseaux est introduite. Nous utilisons des données réelles pour conduire des expérimentations et nous analysons les résultats en les comparants avec l'approche existante.

Pour les réseaux sociaux de la vie réelle, nous proposons un modèle basé sur les ontologies pour construire le réseau social sémantique et pour évaluer le poids de la relation entre les membres. Le modèle est réalisé d'une manière spécifique sur le réseau social de recherche. La familiarité et la similitude sont aussi utilisées pour la modélisation. Nous utilisons également des données en ligne d'un réseau social de recherche pour vérifier l'usage et analyser l'effet de notre modèle.

Afin d'éviter la transmission des données aux nœuds égoïstes dans le réseau opportuniste, et dans le même temps surmonter des problèmes d'algorithmes de transmission de confiance lorsqu'il existe de nombreux nœuds égoïstes dans le réseau, tel qu'un délai de transmission élevé et une évaluation inexacte de la confiance etc. Une application pour améliorer la performance et la sécurité de stratégie de transmission dans le réseau opportuniste basée sur la confiance est définie dans la dernière partie de notre recherche. Nous utilisons les contextes sociaux de familiarité et de similitude de manière similaire aux deux modèles de calcul de poids de relation évoqués ci-dessus. Pour adapter la caractéristique d'évolution des aspects des contextes sociaux des nœuds dans le réseau opportuniste, et pour tenir compte de la pauvreté des ressources d'information contextuelle, et pour inférer plus exactement la relation des utilisateurs selon les activités des nœuds, nous proposons deux algorithmes distincts : un algorithme de transmission de confiance basé sur la confiance sociale binaire et un algorithme de transmission de confiance basé sur la similitude de communauté dynamique des nœuds.

Pour le premier algorithme, basé sur l'algorithme de transmission d'infection *Epidemic*, nous décrivons trois attributs sociaux : degré de familiarité, niveau d'intimité et qualité de service, basé sur des contextes sociaux pour établir une relation de confiance entre les nœuds, et quantifier les valeurs spécifiques de la confiance ; le nœud de confiance peut être sélectionné comme nœud de transmission, pour éviter la transmission aux nœuds égoïstes et assurer le fonctionnement normal du réseau. Le deuxième algorithme introduit la théorie de cercle de connaissance et analyse la similitude de la trajectoire de mouvement et la situation d'interaction des nœuds pour évaluer la confiance entre deux nœuds. Les messages sont transmis au nœud qui a la valeur de

confiance la plus stable, ce qui permet d'assurer la fiabilité de la transmission d'une manière plus poussée.

Nous simulons enfin le réseau opportuniste avec des scénarios réels pour évaluer la pertinence de nos approches sur différents indices, et en comparant avec les résultats obtenus avec d'autres algorithmes.

1.3 Organisation

La suite de la thèse est organisée comme suit: le chapitre 2 présente les notions de base relatives au concept de *context awareness*, la définition et les catégories de contexte, notamment l'importance du raisonnement et l'introduction du contexte social, l'utilisation du contexte avec l'évolution des réseaux, les étapes de traitement du système contextuel et les caractéristiques et limitations du contexte de nature physique pour introduire le contexte social. Le chapitre 3 détaille l'opportunité et les défis inhérents à la prise en compte d'une vision basée sur la prise en compte des caractéristiques sociales dans le cadre du *context awareness*. Nous présentons ensuite la prise en compte des différents niveaux de contextes sociaux et des essais déjà réalisés en utilisant le contexte social. Le chapitre 4 détaille premièrement la notion de contexte multidimensionnel, et deuxièmement le concept et les caractéristiques de réseau opportuniste, ainsi que les stratégies de transmission et de gestion de la sécurité que notre application de contexte social est susceptible d'améliorer. Les chapitres 5, 6, 7 correspondent respectivement aux trois contributions principales précisées ci-dessus. Le chapitre 8 conclue et donne quelques perspectives de recherche potentielles à venir.

Chapitre 2. 'Context awareness'

Dans ce chapitre, nous présentons l'origine et le contexte d'émergence de la notion de « conscience contextuelle » (*context awareness*) ainsi que d'autres concepts de recherche relatifs à cette notion. Un ensemble de définitions et de catégories de contexte sont listées pour mieux circonscrire et comprendre le contexte dans le cadre du développement d'applications et de services, notamment l'importance du raisonnement inférentiel et l'introduction du contexte social. En combinant avec l'application du domaine de réseau que nous avons sélectionnée, la question de l'utilisation du contexte liée à l'évolution des réseaux est détaillée dans la dernière partie. Les processus nécessaires (acquisition de contexte, modélisation de contexte et inférence contextuelle), et les techniques existantes pour l'utilisation de données contextuelles sont ensuite présentés. Nous détaillons plus particulièrement une solution basée sur les ontologies que nous utilisons pour modéliser le réseau social. Finalement nous indiquons les caractéristiques et limitations de la modélisation du contexte basée uniquement sur des données de nature physique pour introduire le contexte social.

2.1 Le courant du 'context awareness'

Avec le développement des technologies de l'information et de la communication, une large gamme de produits électroniques ont pénétré tous les aspects de la vie humaine. L'émergence de chaque nouvelle technologie et nouvelle plate-forme va profondément affecter la façon dont nous vivons au quotidien. De plus, les cycles d'innovation se raccourcissent sans arrêt, et à chaque fois qu'une nouvelle technologie qui anticipe une demande réelle apparaît, elle trouve plus ou moins rapidement son domaine d'application, de manière parfois non prévue par les concepteurs. Les innovations technologiques se traduisent souvent en une augmentation de la productivité et une réduction des coûts de production.

Nous sommes en train de changer progressivement le mode interactif classique au fur et à mesure de la vulgarisation des terminaux intelligents puissants.

Pendant ces décennies marquant l'apparition de l'ordinateur, l'humain est souvent dans une position « dominante » dans l'interaction homme-machine; il prend l'initiative de fournir des données et des instructions à l'ordinateur, puis demande les résultats de l'exécution du programme à partir d'un ordinateur. Ce modèle d'interaction renvoie aux limitations de la technologie et/ou à des contraintes de type sociocognitives : nécessité de garder systématiquement l'humain dans la boucle de contrôle des automatismes afin de garantir un fonctionnement adéquat du système cognitif joint (agent humain + système technique) notamment en situation incidentelle; acceptabilité sociale limitée vis-à-vis des automatismes intrusifs et invisibles.

Avec la progression des technologies et l'évolution des besoins humains, les limitations de ce modèle d'interaction « centré sur les personnes » émergent en permanence : l'utilisateur doit toujours observer et surveiller l'état de son environnement en termes d'informations de contexte pertinent, comme des changements dans l'environnement ou des transformations dans l'état du comportement des gens. Il doit ensuite explicitement l'indiquer à l'ordinateur pour qu'il puisse produire la réponse appropriée.

Dans le domaine de l'interaction et de la communication entre humains, grâce à la richesse de la langue que nous partageons, à la compréhension commune de la façon dont le monde fonctionne, et à la compréhension implicite des situations quotidiennes, nous pouvons facilement transmettre des idées à l'autre et réagir de manière appropriée. Nous pouvons également utiliser les informations implicites d'une situation, ou des informations du contexte, afin de renforcer l'efficacité de la coordination entre agents humains.

Malheureusement, la capacité à transmettre des idées entre les personnes n'est pas directement reproductible dans le domaine de l'interaction entre l'humain et l'ordinateur. En informatique interactive traditionnelle, l'utilisateur dispose de modalités encore limitées pour fournir des données en entrée à un ordinateur. Par ailleurs les capacités de traitement contextuel par une machine restent encore limitées. Par conséquent, en raison de l'absence d'une perception efficace, appropriée et même correcte du contexte, la machine n'est souvent pas en capacité de fournir les bons services aux bons utilisateurs au bon moment. Cette question est d'autant plus critique que l'évolution de

l'interaction Homme-ordinateur va vers de plus en plus de mobilité. La vulgarisation des terminaux intelligents et le déploiement à grande échelle des dispositifs embarqués constituent une étape importante dans la mise en œuvre d'applications sensibles au contexte [3].

Tout d'abord, les terminaux intelligents et les appareils embarqués s'équipent d'une variété de capteurs, tels que des gyroscopes, capteurs de gravité, capteurs de lumière et capteurs de distance etc., qui peuvent obtenir et recueillir des éléments divers pour la caractérisation du contexte. D'autre part, les équipements sont de plus en plus petits, et la puissance de calcul augmente. Enfin, en termes de transmission dans le réseau, transmission filaire ou sans fil, la capacité et la vitesse ont été largement augmentées.

A partir de ce constat, les recherches menées dans le domaine de l'interaction Homme-machine ont commencé à réfléchir à la manière de profiter du confort apporté par l'informatique ubiquitaire [2] (accessibilité des ressources en tout lieu et à tout moment) en utilisant des techniques et des méthodes existantes, et en donnant une importance particulière à la notion de sensibilité au contexte (*context awareness* [1]).

La sensibilité au contexte est en effet à l'origine de l'informatique ubiquitaire. En 1994, Schilit et Theimer [4] ont ainsi proposé le concept de la sensibilité au contexte, en se référant à des applications logicielles qui opèrent adaptativement sur la base d'informations sur l'emplacement d'utilisation, la collecte des personnes proches et les dispositifs accessibles. Ces informations changent tout le temps et ils sont étendus continuellement grâce à la sensibilité au contexte. Dey et al. [5][6] ont introduit l'option d'auto-adaptativité dans les services sensibles au contexte, mais limitée à l'utilisation des informations de contexte de l'utilisateur pour obtenir des services logiciels auto-adaptifs. Salber et al. [7] ont ensuite introduit la notion de service sensible au contexte basé sur temps réel, en fournissant des services de flexibilité maximale. Dey et Abowd [8] ont étendu le champ des applications sensibles au contexte une fois encore, avec notamment l'idée que le système peut fournir des informations ou des services pertinents pour l'utilisateur en profitant du contexte, où la pertinence dépend de la tâche de l'utilisateur.

Cette définition est non seulement assez aisée à appliquer en pratique, mais aussi assez générale et adaptée à la plupart de cas.

Lorsque les applications sensibles au contexte sont apparues, le passage de la surveillance systématique et répétitive et des possibilités très limitées des entrées machine en mode explicite, à des systèmes riches et performants, utilisant la notion de contexte a été évidente (au moins théoriquement). Le système utilise des informations contextuelles acquises pour faire de la déduction, enrichir le comportement actuel de l'utilisateur par des caractéristiques de contexte, puis fournit spontanément et automatiquement aux utilisateurs une assistance ou un service. Un système sensible au contexte brise les contraintes existantes du « mode de service passif », et permet à l'individu de se concentrer sur la réalisation de ses tâches (qu'elles soient professionnelles, culturelles, ludiques, sociales).

L'informatique contextuelle peut ainsi contribuer à augmenter le niveau de couplage implicite entre l'utilisateur et le dispositif et réduire la charge de travail de l'utilisateur.

Un des signes de l'importance des travaux menés sur *context-aware computing* est que la sensibilité au contexte peut être vue comme indépendante de la recherche sur l'informatique ubiquitaire et ouvrir sur un nouveau domaine scientifique à part entière [9]: les chercheurs et les experts dans ce domaine sont unanimes pour penser que le contexte est le noyau de leur recherche, et que les autres aspects se développent autour de ce point.

2.2 Définitions et catégories de contexte

Le terme contexte est largement utilisé dans plusieurs disciplines. Une part, les études menées sur la cognition naturelle mettent en évidence que le contexte est constitutif de la manière dont les humains donnent du sens au monde qui les entoure, interprètent des événements et transmettent des idées à d'autres personnes [10]. Il est largement admis que les humains utilisent le traitement contextuel dans la compréhension linguistique, qu'elle soit écrite ou parlée [11]. Par exemple, le contexte déclenche la réactivation des souvenirs à long terme, puis l'enrichit avec de nouvelles informations et le redéfinit [12]. C'est donc une notion fondamentale en psychologie, notamment dans les modèles «

situés » de la cognition [13]. D'autre part, les travaux menés en intelligence artificielle se donnent pour objectif de développer des capacités de traitement (discernement et raisonnement) inspirés de la cognition humaines. A côté des espoirs mis dans les capacités brutes de traitement informatique des machines, et illustrés par le nouveau domaine des *Big Data analytics*, le traitement et le raisonnement contextuel apparaît pertinent pour créer des systèmes dont le fonctionnement reste intelligible par l'homme [14]. En général, ce terme appartient aux deux mondes de la cognition et du calcul. La littérature [15] a souligné la difficulté de modélisation en informatique parce que le contexte possède une dimension infinie.

En particulier, dans le domaine de l'information de base, le contexte décrit un ensemble d'entités conceptuelles à partir d'un point de vue spécifique. Dans le domaine de l'intelligence artificielle, le contexte se réfère à la division de base de données d'information en un ensemble manipulable ou structure logique pour soutenir les activités de raisonnement correspondantes [16,17]. Et dans le cadre du réseau sémantique, les chercheurs ont montré que des objets tels que l'emplacement ou l'identité et leur changement est un élément de contexte ; ce type de définition du contexte sous méthode d'énumération ou par l'exemple est utilisé souvent dans la recherche primitive d'application sensible au contexte [4, 19].

Brown et Peter [18] ont défini le contexte comme toutes les informations d'environnement perçues par l'ordinateur. Ryan [19] a défini que le contexte est constitué par l'emplacement, l'environnement, l'identité, l'heure et d'autres informations sur l'utilisateur. Hull et al. [20] ont décrit le contexte comme un ensemble d'aspects de l'environnement. Dey et Abowd [8] pensent que toutes les informations sur l'utilisateur et l'environnement où l'application se situe pendant l'interaction entre l'utilisateur et l'application peuvent être considérées comme le contexte.

Même si le contexte est un concept difficile à saisir et à définir, une définition générale largement acceptée est la suivante :

“Context is any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the

interaction between a user and an application, including the user and applications themselves” [21].

Cependant, Zimmerman [22] indique que cette définition est trop formelle, universelle et manque d'une partie opérationnelle importante. En effet, le type d'information nécessaire pour définir la relation entité-situation n'est pas définie dans la définition ci-dessus, mais est nécessaire pour toute approche pragmatique à la modélisation du contexte.

Sur la base des diverses tentatives de définition et de conceptualisation du contexte que l'on peut trouver dans la littérature, nous considérerons que la question peut être traitée sous trois angles (ou modes de classification) principaux : mode d'acquisition, niveau et type.

2.2.1 Classification par mode d'acquisition

Compte tenu de l'origine de l'information de contexte, certains auteurs ont classé le contexte en fonction de la façon dont les données sont capturées : contexte physique, contexte virtuel et contexte logique. Le contexte physique est représenté avec des données à partir des capteurs physiques tels que les appareils GPS, les capteurs de lumière etc. Le contexte virtuel est basé sur des informations qui proviennent des applications ou des services logiciels ; par exemple, la position de l'utilisateur peut être déterminée en parcourant son agenda électronique, e-mails, etc. Enfin, le contexte logique est tiré d'informations à partir de diverses sources. Il combine les informations à la fois des capteurs physiques et virtuels (application qui parcourt l'agenda électronique et l'e-mail d'utilisateur est un exemple de capteur virtuel) [23].

2.2.2 Classification par niveau

Selon la hiérarchie du contexte, de manière générale, les informations contextuelles peuvent être divisées en contexte primaire, contexte moyen et contexte supérieur, comme illustré dans la figure 2-1 :

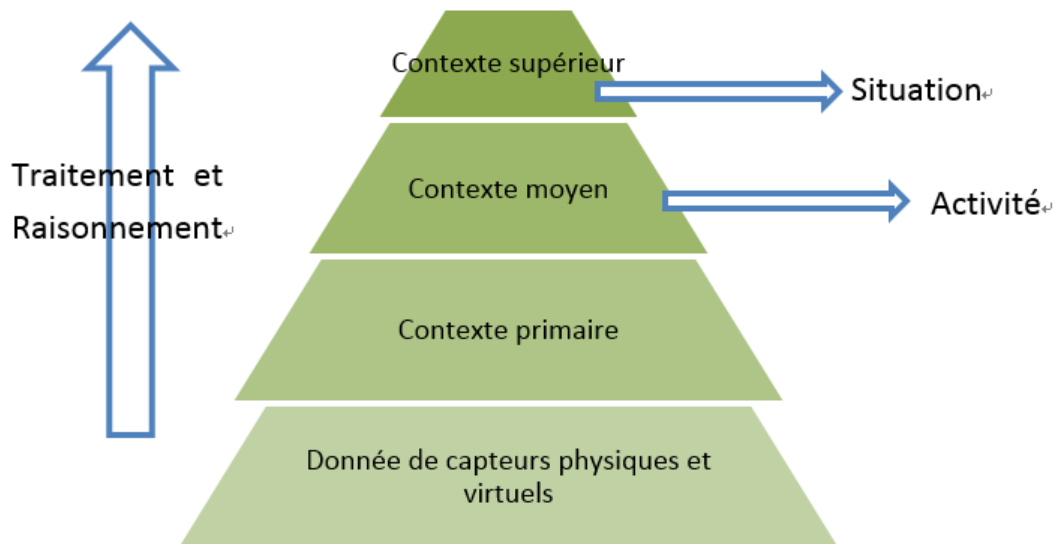


Figure 2-1 : Classification des contextes par niveau

Le contexte primaire est la valeur des caractéristiques désirées et obtenues de manière directe ou indirecte des capteurs bruts. Le contexte moyen est les données sous forme compréhensible après le traitement des données brutes, cette approche est généralement une cartographie simple. Le contexte supérieur est produit d'une manière plus poussée à partir du contexte moyen, lui donnant un sens sémantique qui est facile à comprendre par un humain et qui peut être utilisé directement par le système d'application.

Une forme importante de contexte moyen renvoie aux activités, qui sont représentées après une inférence simple au sujet des informations contextuelles, comme l'activité de « hacher » dans la cuisine inférée en observant le mouvement au fil du temps [24]. Une étape de traitement supplémentaire, permet de faire émerger une nouvelle notion importante : la situation, qui est une abstraction des événements qui se produisent dans le monde réel provenant de contexte et des hypothèses sur la façon dont le contexte observé rapporte des éléments d'intérêt pour les concepteurs et les applications. Les situations fusionnent généralement plusieurs sources de contexte, ainsi que des connaissances du domaine et des modèles spatiaux et temporels du comportement attendu des phénomènes observés [25].

2.2.3 Classification par type

Du point de vue du type de contexte, nous pouvons trouver différentes versions de classification basées sur la méthode de l'épuisement [26]. Par exemple, l'énumération suivante : contexte informatique, contexte d'utilisateur, contexte physique, contexte de temps et contexte d'histoire, ou bien contexte de dispositif, contexte d'espace-temps, contexte d'utilisateur et contexte de tâche.

Dans cette recherche, nous divisons le contexte d'une manière différente en distinguant entre « contexte social » et « contexte physique » que nous introduirons plus loin.

Bien que les définitions pour la classification de contexte mentionnées ci-dessus correspondent plutôt bien pour les applications omniprésentes plus petite dans les environnements avec des frontières bien définies, elles ne suffisent pas pour décrire la multitude de chevauchements des contextes de situations quotidiennes, en particulier si nous voulons envisager une multitude d'utilisateurs. Dans cette recherche, nous le divisons d'une manière totalement différente en distinguant le « contexte social » du « contexte physique » qui est la notion pour les contextes utilisés dans des recherches antérieures. Le <contexte physique> ici se réfère à des informations contextuelles sur les systèmes, les entités et leurs environnements. Le terme <contexte social> est utilisé pour une vision sur la communauté des personnes, groupes et organisations avec lesquels l'individu interagit.

La plupart des applications sensibles au contexte existantes traitent des « contextes physiques » tels que l'emplacement, le temps, l'activité, etc. En comparaison, il y a peu d'études qui traitent des contextes de nature sociale. Cependant, nous, humains, autour desquels s'organisent des systèmes d'informatique ubiquitaire sommes des êtres sociaux. Par conséquent, le « contexte social » devrait théoriquement permettre d'aller plus loin dans la réalisation de l'informatique ubiquitaire. Les applications sensibles au contexte social des utilisateurs sont en mesure de les aider dans leurs activités quotidiennes par de nouveaux services et, finalement, d'enrichir leur bien-être social. Par exemple, une application d'appel téléphonique sensible au contexte

social peut réduire les interruptions en permettant à l'utilisateur de filtrer les appels entrants en fonction de son information sur le contexte social (les relations avec l'appelant, l'état courant de l'appelé, etc.) [27,28,29]

2.3 Utilisation du contexte avec L'évolution des réseaux

Dans cette partie, nous essayons d'introduire les différents rôles et les diverses utilisations du contexte en lien avec les avancées dans les domaines des technologies de l'information et de la communication et de l'évolution des réseaux.

2.3.1 Applications desktop

Malgré la proposition universelle et l'accord sur l'importance de la recherche en informatique sur le contexte, ces recherches passent après la prévision <troisième vague de l'informatique> [30] — le paradigme *ubiquitous computing* était proposé et la notion *context awareness* était incluse. En fait, l'utilisation du contexte a été enrichi accompagnant en même temps le développement et l'innovation dans le domaine des technologies de l'information et de la communication. Même à une époque où les réseaux étaient balbutiants, l'information contextuelle était déjà utilisée pour mieux maîtriser l'ordinateur. Bien qu'il faille importer les données manuellement par l'interface graphique par exemple, le contexte comme l'identité et la référence historique étaient présentes dans l'application *Desktop* ; par exemple, la personnalisation de logiciels de traitement de texte pour faciliter les pratiques de travail des participants [31]. Les aspects les plus communément pris en compte sont l'accès plus facile à la routine ou la fonction utilisée le plus souvent.

2.3.2 Applications web

La généralisation de l'usage des réseaux, impliquant un large accès aux e-mails, une connexion au web à l'échelle mondiale, et de nombreuses autres sources d'information, a conduit les utilisateurs d'ordinateurs à mobiliser différents moyens et stratégies de gestion de l'information. Pour aider les individus à organiser cette information, les systèmes informatiques doivent être en mesure de refléter les catégories, les relations et les indices dont les usagers ont besoin et qu'ils vont pouvoir mémoriser.

À cette fin, un outil [32] est proposé, qui se situe entre les applications de l'utilisateur et les sources d'information et avec lequel l'utilisateur interagit (par exemple, système de fichiers, bases de données, e-mail, web, etc.). Il collecte les propriétés de l'application et indexe les documents sur différentes dimensions de ces propriétés pour créer un dossier personnel de l'utilisation de l'information, et fournir des moyens pratiques pour l'accès à cette information. Il inclut une tentative d'utilisation de la fonction de capteur virtuel qui peut être utilisé comme une extension d'application desktop.

De plus, bien que la condition n'était pas entièrement satisfaite pour la réalisation d'applications sensibles au contexte telles que nous les définissons aujourd'hui, il était déjà possible d'obtenir certains des contextes partagés sur l'internet, par exemple le profil personnel, la préférence, l'activité, etc., ainsi que des éléments sur le contexte physique comme l'emplacement ou la température à l'aide d'analyse sémantique [33].

2.3.3 Informatique mobile (*Mobile computing*)

Au cours de ces dernières années, avec le développement rapide de l'industrie des terminaux portables et de la mobilité, téléphones intelligents, tablettes et autres appareils mobiles sont devenus l'un des plus importants outils de divertissement sociaux et l'un des principaux médias d'accès à l'information. Ces nouveaux appareils mobiles sont équipés de nombreux capteurs de haute performance, tels que les capteurs de géolocalisation, les accéléromètres 3D et les capteurs optiques, etc. Ils peuvent capturer des données complexes et contextuelles de l'utilisateur mobile qui permettent d'aider à atteindre une série de service mobile basé sur la sensibilité au contexte, comme les guides de voyage sensible au contexte [34,35], les rappels basés sur la localisation [35,36], et le système de recommandation sensibles au contexte[37,38], etc.

2.3.4 Informatique ubiquitaire (*Ubiquitous computing*)

L'*ubiquitous computing* (*Ubicomp*) [2,30,39] se réfère à une société dans laquelle l'interaction homme-machine est intégrée parfaitement et imperceptiblement dans la vie quotidienne. Il est suggéré que les espaces de travail peuvent contenir de nombreux ordinateurs qui aident les gens à améliorer leur performance. *Ubicomp* renverse la situation où les humains

adaptent l'activité sociale à la technologie, ce sont les machines qui s'adaptent à l'environnement de l'homme [40]. Au cœur d'*Ubicomp* se trouve une idéologie dans laquelle la technologie informatique est envisagée pour réellement nous servir discrètement en arrière-plan de nos activités, non pas comme des assistants personnels, mais plutôt comme une extension de nos propres capacités.

En dehors des appareils mobiles, *Ubicomp* met en évidence l'utilisation des capteurs du contexte intégrés dans l'environnement, portés par et même incorporés dans l'utilisateur, dans le but de construire un espace intelligent [41].

2.3.5 L'internet des objets (*Internet of Things : IoT*)

L'internet des objets était initialement similaire à l'idée d'*Ubicomp* appliquée à la technologie de l'Internet. Avec le temps, cette notion a été considérablement élargie. En termes simples, l'*IoT* envisage une société où tous les membres disposent d'un plein accès à un véritable environnement peuplé par des technologies intelligentes à tout moment et n'importe où :

"Internet of things (IoT) is a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual 'things' have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network." [42]

Dans *IoT*, des objets « intelligents » sont appelés à devenir des participants actifs dans le commerce, l'information et les processus sociaux. Ils sont activés pour interagir et communiquer entre eux et avec l'environnement en échangeant des données et de l'information sensible à propos de l'environnement. Ils peuvent réagir de manière autonome aux événements du monde réel / physique et l'influencer par des processus qui déclenchent des actions et créent des services avec ou sans intervention humaine directe.

Un développement technologique avec des conséquences évidentes pour l'*IoT* est la mise à jour du protocole Internet (d'IPv4 à IPv6 en augmentant la taille de l'adresse IP de 32 bits à 128 bits). Cette transformation permettra à

chaque personne sur terre à avoir 4.86×10^{28} d'adresses IP [43] à sa disposition. Tous les objets quotidiens peuvent avoir accès à l'Internet individuel, non seulement être identifiés dans les catalogues de l'Internet, mais avoir aussi leurs propres adresses IP.

Les réseaux de capteurs sont les composants les plus essentiels de l'*IoT*. L'*IoT* ne peut exister sans réseaux de capteurs, parce qu'ils fournissent un support majeur à l'infrastructure matérielle (par exemple la détection et la communication), en offrant un accès à des capteurs et des actionneurs.

Perera et al. [44] ont listé l'utilisation du contexte et des applications sensibles au contexte pour l'*IoT*.

2.4 Exigences pour un système contextuel

Le développement d'applications sensibles au contexte implique de relever de nombreux défis sans précédent, y compris la perception et le recueil des informations contextuelles, la manière de l'identifier et de le modéliser, ainsi que l'interprétation et l'inférence du contexte. Nous présentons brièvement ces points dans la partie qui suit.

2.4.1 Acquisition de contexte

La perception et le recueil des informations contextuelles est le point de départ pour les applications sensibles au contexte. Il faut être capable de choisir d'utiliser une variété d'informations contextuelles selon la demande. Ces contextes peuvent être des données qui proviennent des capteurs en temps réel, tels que des informations de localisation GPS, la température et l'humidité, des informations de concentration de polluants, des données obtenues par l'interface avec la base de connaissance, comme les informations personnelles de l'utilisateur et les données historiques, etc. En profitant de la technique du Cloud, des données acquises par des stockages en ligne et par l'interrogation de terminaux connectés au Cloud, il devient possible de disposer d'informations mises à jour sur par exemple, les conditions météorologiques, l'état du trafic aérien en temps réel, les manifestations de sociabilité sur les réseaux sociaux.

Les problèmes principaux pour l'acquisition de données contextuelles sont les suivantes : 1) les informations proviennent de différentes sources et manquent de modèle uniforme des données et de manière uniforme pour les recevoir et les traiter efficacement 2) en raison de la différence entre les capteurs, des pertes et des incohérences des données peuvent se produire lors de la transmission et 3), les données acquises ne sont pas totalement fiables, et il faut généralement un processus de prétraitement avant que les données puissent être utilisées.

On peut distinguer deux mécanismes pour l'acquisition du contexte: un premier basé sur les événements et un second basé sur la requête. La méthode basée sur les événements définit d'abord l'événement qui implique certains changements d'état des données. Après la détection de l'événement, la couche d'acquisition de contexte envoie une notification à la couche d'application intéressée par cet événement. La méthode basée sur la requête permet aux applications d'envoyer des requêtes pour obtenir des contextes intéressants à partir de la base de données ou des plates-formes de détection sous-jacents. La méthode basée sur la requête peut fournir l'interface déclarative comme le langage de requête structuré pour les opérations de la requête.

La méthode basée sur les événements utilise la communication asynchrone et elle est adaptée aux plates-formes de ressources limitées telles que les réseaux de capteurs. La méthode basée sur la requête est plus performante pour la récupération d'informations issues de base de données telle que les serveurs de réseau social en ligne.

2.4.2 Modélisation du contexte

Dans des conditions normales, les sources d'informations contextuelles fournissent les données brutes, mais la présentation et la modélisation de contexte sont la base pour obtenir le contexte de niveau supérieur. Actuellement, les diverses applications sensibles au contexte contiennent leur propre modèle du contexte. Selon les caractéristiques des informations contextuelles utilisées, chaque méthode de modélisation a des problèmes particuliers. L'objectif d'une bonne modélisation devrait être de construire un modèle unifié qui peut être utilisé généralement, de mettre en œuvre

facilement des mécanismes de représentation du contexte et de promouvoir l'interopérabilité du niveau sémantique avec d'autres systèmes sensibles au contexte.

Un modèle général du contexte a une signification pratique importante pour le développement de systèmes sensibles au contexte. Il peut réduire la charge de travail des développeurs et augmenter l'efficacité du développement.

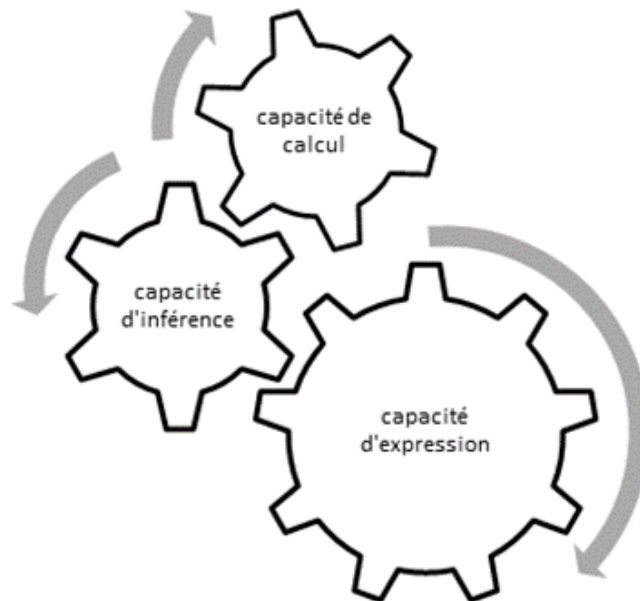


Figure 2-2 : Trois grands besoins de Modélisation de contexte

L'intention initiale du modèle des informations contextuelles est le transfert et le partage des connaissances entre l'humain et la machine, la machine et la machine, et donc n'importe quelle méthode de modélisation doit résoudre trois grands besoins que sont la capacité d'expression, la capacité de calcul et la capacité d'inférence (figure 2-2). Dans la partie suivante, les méthodes principales de modélisation du contexte vont être introduites.

2.4.2.1 Méthode de paires attributs-valeurs (Key-value pair Model)

La méthode de paires attributs-valeurs a été proposée par [26] pour établir les modèles des informations de localisation, et, par la suite, elle a été généralisée dans les autres contextes en la développant et l'étendant. La méthode de paires attributs-valeurs est la méthode la plus simple et la plus ancienne pour la modélisation de contexte ; son principe est de considérer l'information contextuelle comme un attribut qui peut être représenté par une

valeur ; ensuite les valeurs correspondantes peuvent construire une paire pour décrire les informations contextuelles.

Cette méthode est bien adaptée aux capteurs physiques déployés dans l'environnement, par exemple, le capteur de température qui mesure la température ambiante de la chambre 211 trouve la valeur 26°C ; il peut être représenté par la forme :

<Room 211 Temperature, 26>

Le modèle de paires attributs-valeur est souvent utilisé dans les services distribués. Grâce à sa simplicité et son accessibilité directe, chaque genre de service peut être représenté par une paire attribut—valeur. Pendant le processus de découverte du service, le principe simple d'appairage est utilisé, si tous les attributs et les valeurs sont bien correspondants, il est considéré comme le service cible.

Les inconvénients de la méthode de paires attributs-valeurs sont qu'elle ne peut pas représenter directement des structures d'informations complexes.

2.4.2.2 Modèle de balisage (Make-up Scheme Model)

Le modèle de balisage est un modèle d'utilisation de langage de balisage pour décrire les informations contextuelles formalisées, plus précisément, en utilisant le langage de balisage sous la forme d'une structure de données hiérarchiques. Chaque couche est construite par l'attribut et le contenu sous un signe de balisage.

Le concept typique de modèle de balisage est le profil, normalement en utilisant SGML (*Standard Generic Markup Language*) pour faire le développement. Les méthodes de modèle de balisage sont UAProf et CC/PP [45].

2.4.2.3 Modèle graphique

Le modèle graphique utilise les outils orientés vers les graphiques pour modéliser les informations contextuelles, et se concentre sur les descriptions des structures des informations contextuelles.

UML (*Unified Modeling Language*) peut être utilisé comme un langage de modélisation graphique grâce à sa forte capacité de représentation graphique. Le modèle typique est ORM (*Object-Role Modeling*) proposé par Henrichen et al. [46] ; ORM est basé sur des faits, selon la longueur de la durée du fait, les faits peuvent être divisés en deux types, un type de fait stable et un type de fait dynamique.

Bien que le modèle graphique puisse décrire les relations entre les entités contextuelles de manière graphique, il présente des limitations dans l'application pratique. Tout d'abord, il n'y a pas un logiciel universel ni un outil qui peut supporter toutes les représentations du contexte, et ensuite il manque une norme uniforme, qui résoud le problème des différences importantes entre les notions contextuelles et les expressions. Donc, ce modèle ne favorise pas l'inférence des informations contextuelles d'une manière plus poussée.

2.4.2.4 *Modèle orienté objet (Object Oriented Model)*

La conception orientée objet peut être introduite pour la modélisation des informations contextuelles, c'est-à-dire, les informations peuvent être encapsulées et réutilisées suivant leur classe, les informations qui ont des attributs et les méthodes similaires peuvent être encapsulés comme une classe. Comme la classe a la capacité d'encapsulation et d'héritage, non seulement les duplications des définitions peuvent être évitées facilement, mais aussi les descriptions de la qualité des attributs des informations contextuelles et les descriptions de complétude peuvent être envisagées. De plus, la hiérarchie des informations contextuelles devient plus structurée, et conduit à un système complexe des informations contextuelles.

Comme l'héritage de la classe et la contenance de la classe sont utilisés, le modèle orienté objet est plus naturel et plus efficace que le modèle graphique pour décrire le monde réel.

Un modèle typique est le CML (*Context Modeling Language*) proposé dans la littérature [47]. CML est développé à partir du modèle de l'ORM, et supporte donc les représentations graphiques des informations contextuelles pour satisfaire des besoins de l'analyse et l'expression formelle des développeurs du logiciel.

L'avantage de l'utilisation du CML est de faciliter la mise en correspondance avec une base de données relationnelle. En même temps, les expressions et les recherches des informations contextuelles de haut niveau peuvent être réalisés pendant le calcul. L'inconvénient de CML est qu'il est un modèle <en plan>, il ne supporte pas le modèle hiérarchique des informations contextuelles, en conséquence, la bibliothèque de modèle serait trop lourd et compliquée à manipuler.

2.4.2.5 Modèle basé sur la logique (Logic Based Model)

Le modèle basé sur la logique est un modèle qui décrit les informations contextuelles en utilisant le système logique et permet le raisonnement déductif des informations contextuelles. Il définit les contextes comme une série de fait, d'expression ou de rôle. La logique est de définir les conditions contraintes de développements du fait ou de l'expression connue (existe, brut, de niveau bas) à du fait et de l'expression inconnue (dérivatif, précis, de haut niveau).

Par conséquent, dans un modèle contextuel basé sur la logique, les opérations comme ajouter, supprimer, mettre à jour etc. de contexte sont sous la forme de fait ou d'expression.

Il a deux avantages principaux : la forme facile à comprendre ; la possibilité de montrer la précarité de contexte par la probabilité et la capacité d'inférence mathématique sur cette précarité. Mais le modèle basé sur la logique ne peut pas exprimer les relations entre les entités dans différents contextes, pour la validation partitive et la convivialité, il a des limitations importantes.

En effet, un bon modèle contextuel doit répondre aux exigences autant que possible. Mais ce n'est pas possible de trouver un modèle universel qui peut satisfaire tous les besoins, en plus de la précision de l'expression et de la richesse sémantique en même temps. Il faut donc trouver les équilibres entre différents besoins. Les méthodes de modélisation des informations contextuelles sont résumées dans [48,49], et une méthode d'évaluation approfondie pour évaluer les caractéristiques de chaque modèle est proposée.

Le modèle a besoin d'une manière d'exprimer les choses qui peut être comprise par l'humain et aussi par la machine, afin d'atteindre l'interaction entre les deux ; Il faut équilibrer le coût des entrées et le coût des sorties pour évaluer la capacité de calcul. Un excellent modèle doit permettre aux ordinateurs de sortir rapidement et précisément les résultats. Afin d'obtenir la bonne sortie correspondante à l'entrée, il faut suivre la bonne manière d'inférer.

2.4.2.6 Modèle basé sur les ontologies

Le terme ontologie est dérivé du domaine de la philosophie qui était destinée à caractériser la nature de « l'être ».

Studer et al. définissent l'ontologie comme une spécification explicite et formelle d'une conceptualisation partagée :

An ontology is a formal, explicit specification of a shared conceptualisation [50]

Dans le domaine de la gestion des connaissances, l'ontologie est une compréhension commune des domaines spécifiques, une description conceptuelle des notions et relations existantes. En exprimant le modèle conceptuel et les relations logiques précisément, elle forme une compréhension commune d'agrégation des notions et réduit le malentendu du domaine du problème par les autres.

Les modèles d'ontologie sont connus pour les avantages suivants dans la modélisation de contexte:

- ✓ En raison du fait qu'ils sont basés sur un langage de description simple et flexible, ils ont une grande puissance expressive [51]. Ceci est important pour la modélisation de contexte, car les données de contexte peuvent évoluer et être hétérogènes.
- ✓ Ils possèdent un haut degré de formalisation, facilitant le partage de la connaissance des données de contexte hétérogènes [48]. En effet, ils fournissent des classes, propriétés et règles prédéfinies, qui fournissent une base formelle sur laquelle on peut créer de nouvelles classes et propriétés pour les domaines de contexte spécifique.

- ✓ Ils sont réutilisables. C'est également un avantage important des modèles d'ontologie. [52] démontre que l'utilisation d'ontologies supérieures génériques et réutilisables tels que CONON ou CoDAMoS [53] permet de créer des ontologies uniques et spécifiques à un domaine.
- ✓ Enfin, les règles formelles fournissent des mécanismes de raisonnement logique. Cela permet, comme on l'a déjà dit, que l'inférence de nouvelles informations sur la base des données brutes de niveau inférieur soit possible, et aussi d'évaluer le contexte en vérifiant la cohérence, la compatibilité, l'incomplétude et de l'ambiguïté.

Cependant, le modèle ne peut pas modéliser l'incertitude du contexte et il est complexe d'étendre l'ontologie.

En conclusion, l'approche de la modélisation basée sur les ontologies pour le contexte offre un bon équilibre entre la capacité d'expression et la capacité d'inférence, et donc cette approche reçoit de plus en plus d'attention dans le cadre de recherches académiques et industrielles. Elle peut soutenir aussi la relation dynamique entre les entités, et par conséquent, pour toutes les méthodes, le modèle basée sur les ontologies est adapté pour la tâche de modélisation du contexte social comme un type de contexte spécial dans notre recherche. Des recherches significatives ont été réalisées dans le domaine du développement d'ontologies spécifiquement conçues pour les descriptions de contexte, généralement par l'extension de langages d'ontologie de base existants, tels qu'OWL [54]. Ces ontologies comprennent CONON [52], CoDAMoS [53] et SOUPA [55] ontologies qui définissent de nouvelles classes liées à des applications qui interagissent avec les humains.

2.4.3 Inférence contextuelle

L'inférence contextuelle se réfère au processus d'inférer des informations de contexte implicites de niveau élevé basées sur l'information existante de contexte explicite de niveau bas. Les techniques existantes peuvent être classées en deux types: inférence exacte et inférence inexact ou floue. L'approche exacte comprend la méthode basée sur les réseaux bayésiens [56], méthode basée sur la logique [57], basée sur les cas [58], basée sur les ontologies [55], basée sur les règles [59] techniques. Tandis que des

techniques inexactes comprennent la méthode basée sur la preuve [60] et la méthode basée sur la logique floue [61]. En héritant de la modélisation du contexte basée sur les ontologies, nous utilisons aussi l'approche d'inférence basée sur les ontologies. Cette partie n'est pas le point important pour notre recherche, donc nous ne la présenterons pas en détail.

2.5 Les caractéristiques et limitations de contexte de la nature 'physique'

Bien que les définitions pour la classification de contexte mentionnées ci-dessus fonctionnent plutôt bien pour les applications omniprésentes plus petites dans les environnements avec des frontières bien définies, elles ne suffisent pas pour décrire la multitude de chevauchement des contextes de situations quotidiennes, en particulier si nous voulons envisager une multitude d'utilisateurs. Dans cette recherche, nous classons les choses d'une manière différente en distinguant le « contexte social » du « contexte physique » qui était le cœur des recherches antérieures. Le « contexte physique » se réfère ici à des informations contextuelles sur les systèmes, les entités et leurs environnements. Le terme « contexte social » est utilisé pour une vision sur les communautés de personnes, groupes et organisations avec lesquels l'individu interagit.

Chapitre 3. Contexte social

Dans le chapitre 3, nous détaillons l'opportunité et les défis pour une vision du *context awareness* intégrant la composante sociale. Trois défis principaux pour l'utilisation du contexte social sont introduits : standards de représentation unifiée, capacité d'inférence et protection de la vie privée. Nous présentons ensuite une partie importante pour le raisonnement et l'évaluation d'indices de contexte social dans notre proposition de modélisation : la prise en compte des différents niveaux de contextes sociaux. Des essais antérieurs utilisant le contexte social sont décrits et discutés dans deux types d'application : applications de présentation et applications de présentation d'exécution

3.1 Introduction: Opportunité et challenge pour une vision sociale

Grâce à l'expansion rapide des techniques de *context-awareness*, les applications de *pervasive computing* deviennent plus intelligentes et accessibles dans le but de réaliser des environnements où les gens sont aidés dans leurs tâches quotidiennes par des technologies qui fonctionnent en arrière-plan, et demeurent invisibles bien que significativement utiles. Mais les utilisateurs sont des êtres sociaux. Dans les cas normaux, il est inévitable que nous rencontrions des gens avec qui nous avons des interactions, puis avec qui nous construisons une relation, pour éventuellement devenir un membre d'une certaine communauté. Poser la question de l'utilisation par des applications de ces informations relatives aux activités sociales constitue donc une question potentiellement pertinente.

Il existe une tendance récente qui vise à tirer parti des téléphones mobiles des usagers sur une grande échelle pour effectuer de la détection mobile. Les exemples incluent la surveillance du trafic routier, les patterns de circulation, ou même la présence de nids de poule sur les routes [62]. Ce genre de détection est décrit comme la détection centrée sur les personnes [63]. Malgré l'échelle de détection assez simple, cela constitue déjà un bon essai pour obtenir des informations sur la dimension sociale. Dans une seconde étape, grâce à une extension de la notion initiale de « sensibilité au contexte » à celle de « sensibilité sociale », il permet de penser de nouveaux types d'applications,

par la reconnaissance des contextes sociaux à partir des patterns de communication, des situations d'interaction au niveau communautaire, et des comportements humains collectifs [64]. Par conséquent, la notion de « sensibilité sociale » peut étendre le champ du *context-aware* plus loin dans la réalisation de la vision *pervasive computing*, où les applications seront informées des contextes sociaux des utilisateurs et seront en mesure de les aider dans leurs activités quotidiennes et de contribuer à enrichir leur bien-être social.

Récemment, nous avons assisté à un nombre croissant d'efforts visant à fournir des applications sensible au contexte social dans des domaines tels que l'informatique ubiquitaire, le web sémantique et la recherche d'information. Ces applications exploitent les informations de contexte social des utilisateurs tels que les relations, l'historique de l'interaction, etc. pour fournir des services et des capacités « intelligentes ». Par exemple, SmartObject [65] considère les informations de relation de l'utilisateur pour activer un lecteur audio lorsque les amis de l'utilisateur sont présents. CAMEO [66] supporte le service de réseau social mobile en temps réel (MSN) basée sur non seulement les relations sociales existantes mais aussi l'interaction sociale entre utilisateurs inconnus qui partagent des intérêts et des besoins. Il tient compte non seulement des interactions sociales humaines sur le réseau physique de dispositifs électroniques, mais aussi des intérêts des utilisateurs et de leurs conditions actuelles.

D'autre part, la popularité des réseaux sociaux tels que Facebook, LinkedIn et Twitter [67] favorise la production d'une quantité sans précédent d'informations relatives au contexte social (*Social Context Informations (SCIs)*) dans la mesure où les usagers précisent leurs relations, renouvèlent leurs statuts et partagent des contenus.

Du fait de l'insertion souvent très forte des réseaux sociaux dans la vie quotidienne de la plupart des gens, les données générées par ces échanges ont donné naissance à une sorte de *big data* dynamique, et la question de l'analyse de ces est devenue un enjeu de recherche, mais aussi commercial, prometteur. Le terme de *Social Mining*, en lien avec les *big data analytics*, a ainsi émergé [68]. Cette approche peut être utilisée pour prédire le comportement d'un

utilisateur, pour ensuite définir des stratégies d'adaptation et de personnalisation applicables à cet utilisateur. Par exemple, un système de *business intelligence* peut utiliser des résultats d'analyse pour encourager une cible de clients particuliers à acheter les produits qui les intéressent. Plus généralement, il devient possible d'exploiter toute source d'information, qui peut être identifiée dans les services de réseaux sociaux, et qui exprime les intérêts individuels, les préférences et les activités dans la sphère sociale, bien que les contextes sociaux soient répartis sur tout le web, fragmentés à travers des nombreuses applications sur des plates-formes différentes. Par conséquent, une occasion unique est offerte pour la création d'applications de service sensibles au contexte social qui permettrait d'enrichir considérablement la portée et l'intérêt de ces applications.

On peut classifier le contexte social selon les modalités d'interaction sociale, et selon que celle-ci se déroule dans le monde réel ou virtuel.

La construction d'un contexte social peut ainsi provenir de l'interaction directe ou indirecte entre les personnes à la fois dans le monde virtuel et physique. L'interaction directe comprend la conversation face à face, la visioconférence, etc., tandis que l'interaction indirecte comprend la co-localisation pendant une période de temps, la présence d'un même événement, etc. (tableau 3-1).

	<i>Monde physique</i>	<i>Monde virtuel (Réseau social)</i>
<i>Interaction directe :</i>	Conversation face à face	Message lass é
<i>Pour inférer le contexte social dyadique</i>	Conversation sur dispositif t él écommunication ou r éseau	Communication en temps r éel, litt érale, vocale ou visio
	Lettre ou SMS	D'autre comportement virtuel, partager, aimer, @, etc.
<i>Interaction indirecte :</i>	Co-localisation pendant une p ériode de temps	Suivre m ême gens, dans m ême groupe, choisir m ême tags
<i>Pour raisonner sur les communautés statiques et dynamiques</i>	Présence d'un même événement	Co ñcidence de profil
	Pattern similaire d'activité ou de mouvement	

Tableau 3-1 : Situations d'interaction

Du fait de l'hétérogénéité des conceptions de la notion de contexte et des données qui le constituent, les définitions du contexte social dans la littérature

sont diverses et souvent exclusives. Liang et al. [29] par exemple ont listé chaque définition du contexte social dans le cas de 12 middlewares sensibles au contexte social. Nous avons retenu une définition assez satisfaisante et complète qui nous paraît convenir aux objectifs de notre travail: « le contexte social se réfère à un ensemble d'informations qui caractérisent plusieurs utilisateurs tels que le lien social, le groupe social ou la dynamique de groupe » [69].

3.2 Trois défis principaux pour l'utilisation du contexte social

Le contexte social, pour être utilisé efficacement, doit répondre à trois principaux défis qui sont la représentation unifiée, la capacité d'inférence et la garantie de la vie privée.

3.2.1 Standards de représentation unifiée

Il existe différents types de contexte social englobant à la fois une information sur les réseaux sociaux et une information physique qui doivent être représentées de façon cohérente pour faciliter le fonctionnement et l'utilisation de l'application.

3.2.2 Capacité d'inférence

Une application peut avoir besoin des contextes sociaux qui ne sont pas directement disponibles à partir des sources de contexte mais qui peut être dérivées d'autres informations disponibles. Par exemple, pour l'inférence du lien social, nous avons besoin des informations «ami commun » [70], « intérêt commun » [71] et « co-localisation » [72] qui caractérisent la relation. La manière dont nous pouvons détecter et inférer l'existence de groupes est aussi un enjeu important. Ainsi, il y a ici nécessité de déduire le contexte social (à différents niveaux d'abstraction) à partir des relations de base (lien familial et ami proche par exemple) ainsi que les attributs (par exemple, la force et la confiance).

3.2.3 Vie privée - intimité

L'information du contexte social de l'utilisateur est intrinsèquement sensible et peut être utilisée pour déduire d'autres informations sensibles.

C'est toujours un problème d'actualité car l'application qui utilise et profite des contextes sociaux oblige les utilisateurs à partager leurs informations pour le bien théorique du plus grand nombre, mais cela peut aussi compromettre l'intimité de l'utilisateur, le contrôle d'accès sur son statut et autres données personnelles. Chen et Rahman [73] ont ainsi étudié les réseaux sociaux mobiles actuels et identifié leurs faiblesses et leurs points forts. Comme procédés visant à protéger l'intimité des individus, l'agrégation et l'anonymisation des données peuvent fournir une solution grossière utile [74], mais une analyse plus approfondie est nécessaire pour identifier les effets de futures applications omniprésentes sur la protection de la vie privée.

3.3 Prise en compte des différents niveaux de contextes sociaux

Selon l'analyse de la notion de réseau social [75], le contexte social peut être caractérisé à trois niveaux : niveau de l'acteur, niveau de la dyade et niveau des sous-groupes. Dans notre recherche, nous développons cette manière de classer en l'étendant à 4 catégories: contexte social individuel, contexte social binaire, contexte social des communautés statiques et contexte social des communautés dynamiques.

- Le contexte social individuel réfère au profil social de l'individu, y compris le contexte, l'intérêt, les amis, etc.
- Le contexte social dyadique caractérise la relation entre deux utilisateurs en se référant notamment à la force du lien social.
- La communauté statique concerne les individus reliés par une relation directe identifiée par une définition formelle et déclarative. Par exemple, une personne identifie d'autres personnes comme son père, son superviseur, un ami proche, etc. Ce type de relation peut être utilisé dans une application d'appel téléphonique socialement sensible.
- La communauté dynamique est définie par un ensemble d'individus qui ont manifesté des intérêts communs (via des *like* ou *tag* dans Facebook), qui

participent à des activités communes ou deviennent membres de groupes similaires. Ce type de relation a été utilisé dans des applications qui visent à inférer des préférences [76] et des mesures incitatives de partage des ressources [77].

3.4 Quelques exemples de réalisation utilisant le contexte social

Chen et al. [78] définissent trois niveaux d'interactivité pour les applications sensibles au contexte : personnalisation, sensibilité passive au contexte et sensibilité active au contexte.

- ✓ La « personnalisation » permet à l'utilisateur de spécifier ses propres paramètres et la façon dont l'application doit se comporter dans une situation donnée;
- ✓ La « sensibilité passive au contexte » met à jour le contexte ou les informations du capteur de l'utilisateur, mais permet à l'utilisateur de décider comment changer le comportement de l'application;
- ✓ La « sensibilité active au contexte » change automatiquement le comportement de l'application en fonction de l'information détectée.

Dans cette étude, nous nous appuyons sur la définition ci-dessus sur la sensibilité active et passive au contexte. La sensibilité active au contexte décrit les applications qui, sur la base des données des capteurs mesurées, modifient leur contenu de façon autonome. Et la sensibilité passive au contexte présente simplement le contexte mis à jour pour l'utilisateur et permet à l'utilisateur de spécifier comment l'application doit changer. Une autre distinction est fournie par Brown et Jones, qui définissent les niveaux d' « interactivité » et de « proactivité » [79]. Les applications interactives incluent des caractéristiques des deux premiers niveaux de Chen et al., et la proactivité est presque identique à la sensibilité active au contexte.

Après analyse et comparaison des deux précédents travaux menés par Schilit et al. [26] et Pascoe [80], Abowd et al. [6] ont identifié trois caractéristiques qu'une application sensible au contexte peut prendre en charge : présentation, exécution et marquage. On peut appliquer ces caractéristiques au paradigme de l'application sensible au contexte social.

Nous précisons les deux catégories présentation et exécution dans une perspective des applications de sensibilité sociale.

- L'application de présentation présente le contexte nouveau ou mis à jour à l'utilisateur intéressé (ou rend le contexte persistant pour le récupérer plus tard) et, permet à l'utilisateur de décider comment changer le comportement de l'application.
- L'application d'exécution change automatiquement le comportement de l'application en fonction de l'information détectée.

En fait, tous les services d'application peuvent être construits comme ces deux sortes d'application mais ces usages doivent être distingués pour les différents utilisateurs. Un exemple simple d'une application d'exécution est le téléphone mobile qui change son horodatage automatiquement lorsque le téléphone entre dans un nouveau fuseau horaire. L'application de présentation correspondante est le téléphone mobile qui demande à l'utilisateur des informations sur le changement de fuseau horaire et permet à l'utilisateur de choisir si le temps doit être mis à jour ou non.

Par exemple, dans un premier scénario de tourisme de middleware sensible au contexte social CAMEO [66], des informations de contexte sont représentées par les préférences de l'utilisateur, son profil, l'emplacement, les attractions à visiter dans la ville, et le contenu généré par l'utilisateur, mais aussi par des informations relatives aux utilisateurs environnants et leurs périphériques (les types de connectivité disponibles, leurs intérêts, contenu, etc.). En recueillant et en gérant toutes ces informations, CAMEO optimise des protocoles de diffusion du contenu des gens qui viennent visiter des attractions intéressantes et qui peuvent fournir des informations utiles, non disponible sur Internet et récemment mises à jour. Grâce à ce type des communications opportunistes, l'application est capable d'exécuter automatiquement des routines pour aider les utilisateurs intéressés à réécherlonner leurs visites, optimiser leur temps et éviter les expériences désagréables.

Un autre exemple de l'application d'exécution est CallCensor [81] qui exploite des informations sociales reçues pour décider d'autoriser ou non les appels entrants à être présentés à l'appelé. En plus de l'information sociale, cette

application utilise l'emplacement du téléphone pour en déduire si l'utilisateur est en réunion (par exemple, dans le bureau). Pour chaque appel entrant, l'application réalise un traitement basé sur l'intensité sociale ou l'inférence de voisinage pour évaluer le type de lien social entre l'appelant et le propriétaire du téléphone. Sur la base des paramètres du propriétaire (par exemple, autoriser les appels du conjoint à tout moment), l'application décide si le téléphone doit sonner, vibrer ou demeurer silencieux lors de la réception de l'appel.

Nous pouvons également penser l'application de présentation dans le cas où un appel de numéro étranger arrive et que le propriétaire est libre ; l'application peut alors l'informer, présenter les contextes sociaux inférés et laisser le propriétaire décider du comportement à suivre. Cela peut être utile pour connaître les pratiques d'utilisateur (sous forme de patterns) et améliorer ultérieurement le service automatique.

Chapitre 4. Evolution des réseaux vers l'utilisation du contexte social

Nous présentons premièrement la notion de contexte multidimensionnel pour illustrer la diversité de sources de contexte social et la différence entre les différentes couches du réseau social.

Deuxièmement, pour une préparation d'application de contexte social dans le chapitre 6, nous développons le concept et les caractéristiques de réseau opportuniste, ainsi que l'état de l'art sur les stratégies de transmission et de gestion de la sécurité que notre application de contexte social est susceptible d'améliorer.

4.1 Contexte multidimensionnel et réseaux sociaux

Notre environnement physique devient de plus en plus saturé d'entités informatiques et communicantes qui interagissent entre elles, et avec les utilisateurs. Des composants du monde réel interagissent avec le cyberspace via des éléments de détection, des équipements informatique et de la communication, qui amènent donc la convergence du *Cyber-Physical World (CPW)* [82].

Les structures sociales humaines jouent un rôle crucial pour la diffusion d'information dans le monde cyber, et réciproquement, le cyberspace adapte le monde convergent au comportement humain et à la dynamique sociale. L'information sociale est extraite des rencontres entre les appareils mobiles détectés par des interfaces sans fil telles que Bluetooth, ZigBee et Wi-Fi [83,84] ; d'autres techniques d'interactions sociales qui ne sont pas fondées sur des rencontres physiques nous donnent également un aperçu du comportement de l'utilisateur, tel que le courrier électronique, les chats et les services en ligne de réseaux sociaux (par exemple Facebook, Twitter et LinkedIn). Bien que ces deux types de contexte social appartiennent toutes deux à la catégorie de l'interaction du monde cyber, les premières traitent des rencontres physiques tandis que les secondes sont liées à un contact virtuel entre les nœuds.

A part la couche physique qui correspond à des relations entre individus dans le monde réel, si nous étendons le nombre de contextes sociaux et considérons

plusieurs réseaux sociaux pour un ensemble particulier d'utilisateurs, nous obtenons une structure multi couches, qui représente les connexions d'un seul utilisateur à d'autres utilisateurs sur plusieurs couches autonomes. Deux utilisateurs peuvent ainsi être connectés par de nombreuses couches en même temps.

Le résultat est un système complexe dans lequel plusieurs couches du monde réel et cyber cohabitent, et où les utilisateurs exploitent différents types de relations, comme le montre la figure 4-1.

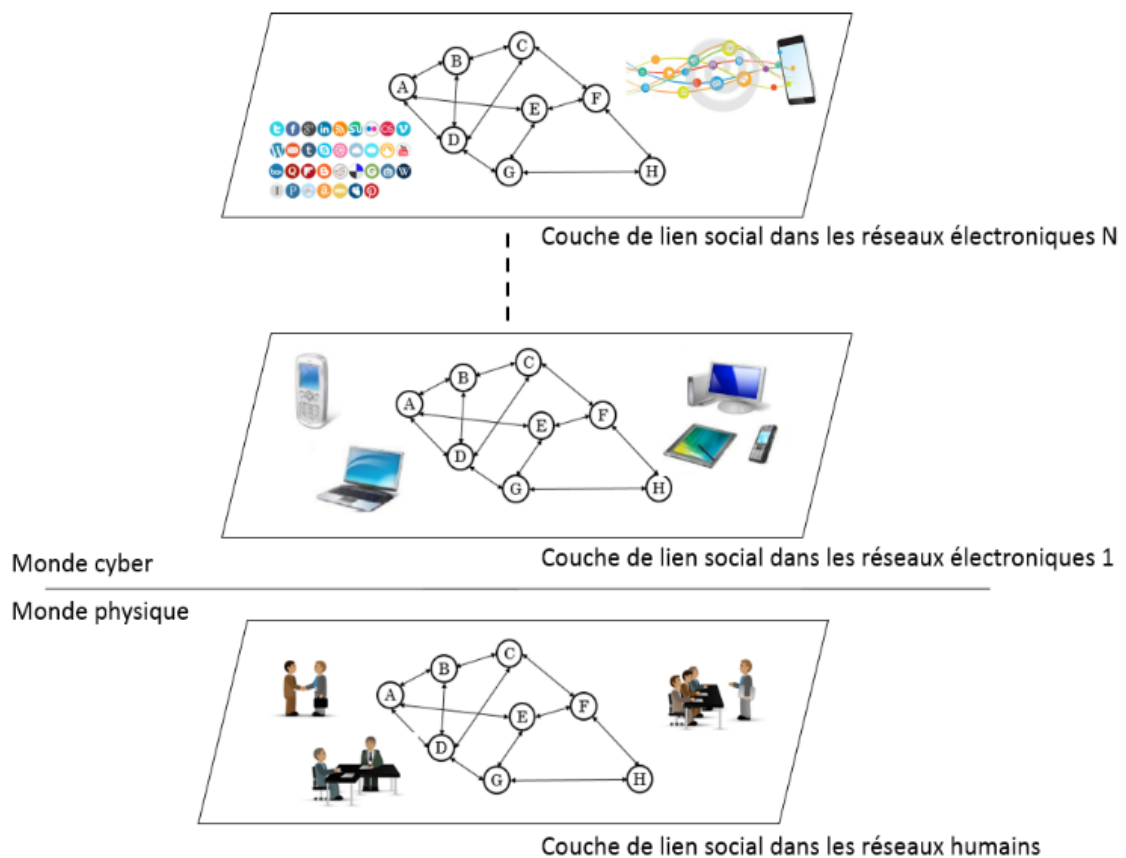


Figure 4-1 : Contexte multidimensionnel et réseaux sociaux

Différents types de contextes sociaux et réseaux sociaux de différents niveaux seront générés par plusieurs sources de ces couches qui doivent être gérés dans un traitement futur. Pour utiliser cette énorme quantité de données hétérogènes provenant d'une variété de sources, la qualité de ces données doit être évaluée et prise en considération avant l'utilisation de ces données. Pour cette raison, notre but est de fournir un ensemble de solutions d'extraction, de modélisation et de raisonnement pour ces divers contextes sociaux grâce à les

propositions de raisonnement dans les chapitres suivantes et le poids de relation entre les utilisateurs.

L'exploration des relations humaines dans le monde virtuel ajoute des nouvelles dimensions au *Pervasive Computing* et aux paradigmes de communication. Plus précisément, cela peut donner lieu à la mise en œuvre de la notion de réseaux opportunistes, *Opportunistic Networking and Computing (ONC)* [85].

4.2 Utilisation du contexte social dans un réseau opportuniste

4.2.1 Le réseau opportuniste

Le réseau opportuniste est un réseau particulier capable de choisir les bons algorithmes réseaux au bon moment et que l'on pourrait voir comme un mélange de *delay-tolerant networks (DTNs)* et de réseau ad hoc mobile. Il ne s'agit pas de concevoir un chemin complet entre les nœuds source et destination, mais de réaliser la communication inter-nœuds de manière à réaliser la suite «stocker-porter-transférer» en fonction des nœuds rencontrés dépendante de la mobilité des nœuds [86]. Comme une branche de DTNs, le réseau opportuniste partage toutes ses caractéristiques, telles qu'une connectivité intermittente élevée, une grande latence des données, une mauvaise stabilité de la topologie du réseau. Les DTNs sont présents pour fournir des services de communication à d'autres réseaux, en utilisant ses points forts pour compenser de nombreux problèmes de communication, mais ils se composent généralement d'un certain nombre de différents sous-réseaux, et ne peuvent pas créer des connexions à long terme entre les sous-réseaux, de sorte que le protocole TCP / IP ne peut s'appliquer à ce réseau [87]. En outre, le réseau opportuniste a également des caractéristiques du réseau mobile ad hoc, telles que le caractère d'auto-organisation forte, de distribution, de capacité très limitée, etc. Le réseau ad-hoc mobile est établi par les comités normatifs IEEE 802.11 après les années 90, et utilisé à l'origine dans le domaine militaire, pour empêcher l'effondrement de la communication des réseaux causés par la guerre, et est aujourd'hui largement appliqué dans le domaine des communications mobiles [88]. Par conséquent, le réseau opportuniste peut

être considéré comme un réseau mobile ad hoc qui possède des caractéristiques de DTNs, qui combine les avantages des deux, et qui possède également son propre mode de composition de réseau et une forme de communication unique.

Maintenant, avec la large utilisation des différents types de dispositifs portables sans fil dans les sociétés humaines, le réseau opportuniste peut être largement utilisé, de sorte que la communication entre les personnes peut également être réalisée par rencontre opportuniste. Comme le montre la figure 4-2, la manière de transmettre la séquence «stocker - porter - transférer» dans le réseau opportuniste social par Bob pour envoyer un message à Alice d'une autre société est décrite. Tout d'abord, Bob envoie le message à Lucy qui est autour de lui, Lucy transmet le message à un bus quand elle passe par un arrêt, ensuite, le bus le passe à Jack qui est collègue de Alice se déplaçant, et Jack finalement envoie le message à Alice en retournant au bureau.

Pendant la totalité du processus de transmission, les dispositifs de chacun d'entre eux doivent être à portée de communication de l'objet afin d'envoyer les messages les uns aux autres. Bien que la chaîne de transmission puisse être multiple, le processus de transmission de message reste le même.

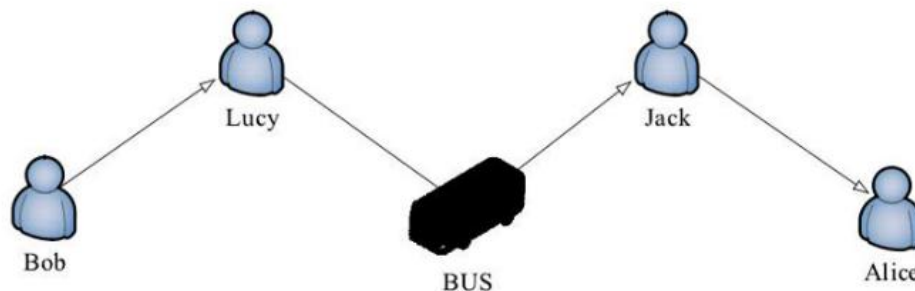


Figure 4-2 : Exemple de transmission de réseau opportuniste

Par rapport aux réseaux sans fil traditionnels, la topologie du réseau opportuniste est tout d'abord instable. Parce que des nœuds du réseau opportuniste sont toujours en mouvement, formant parfois des sous-réseaux de tailles différentes, ils peuvent entrer et sortir des sous-réseaux à tout moment, ce qui provoque une topologie réseau toujours en état instable. Comme il n'existe essentiellement pas de structure centrale et des

infrastructures tierces, tous les nœuds du réseau peuvent être à la fois une source ou une destination du message, et à la fois nœud de transmission de cœur. Tous les nœuds se déplacent toujours dans un certain espace, la vitesse et la direction du mouvement sont contrôlées par eux-mêmes, mais influencés par ses propres équipements. La portée de communication est très limitée, et donc la communication entre les nœuds ne peut pas toujours maintenir la connexion et le routage du message n'est pas en conformité avec les chemins établis au préalable. En plus, les nœuds de réseau opportuniste sont autonomes, à cause de la transmission sans fil, ses ressources comme l'énergie, la mémoire et la bande passante sont très limitées. Si les ressources sont gaspillées, le nœud va rapidement perdre sa capacité à fonctionner et réduire ainsi les performances du réseau. Enfin, le réseau opportuniste peut être composé d'une variété de différents réseaux sans fil ; un tel réseau sans fil doit être flexible et compatible avec différentes interfaces réseau pour améliorer efficacement le débit du réseau et le taux de transmission du message. En conséquence, comparés à d'autres technologies réseau, les problèmes posés par la transmission et le routage du réseau opportuniste sont plus importants.

4.2.2 Protocole de transmission

Comme décrit ci-dessus, pour concevoir le routage efficace dans le réseau opportuniste, le problème clé est devenu la manière de déterminer le meilleur nœud suivant et l'opportunité adaptée de transmission pour chaque message. Par conséquent, l'établissement et le maintien de la fonction de routage se transforment en une décision stratégique du nœud de transmission suivant. C'est pourquoi, nous utilisons le terme « transmission » au lieu de « routage » dans ce texte.

Sur la base des stratégies de transfert différentes, les mécanismes de transmission actuels peuvent être divisés en quatre catégories : le mécanisme basé sur la redondance, le mécanisme fondé sur l'utilité, l'hybride des deux premiers et le mécanisme de transmission basé sur le mouvement du nœud (figure 4-3).

Chaque message dans le mécanisme basé sur la redondance génère une pluralité de messages redondants qui se diffusent dans le réseau.

Donc la performance de transmission est améliorée grâce à une transmission parallèle multi-chemins.

Le mécanisme fondé sur l'utilité est similaire au réseau traditionnel, il utilise un seul chemin et une seule copie du message. Nous prenons la décision du prochain saut par les informations d'état du réseau. Ce mécanisme assigne à chaque nœud une valeur utilitaire, en utilisant différents paramètres comme la prévision de rencontre, l'état de la liaison et le contexte pour estimer l'utilité de transmission des nœuds. Lorsque deux nœuds se rencontrent, le message est transmis à partir du nœud qui a une valeur d'utilité faible vers le nœud d'utilité plus élevée jusqu'au nœud de destination.

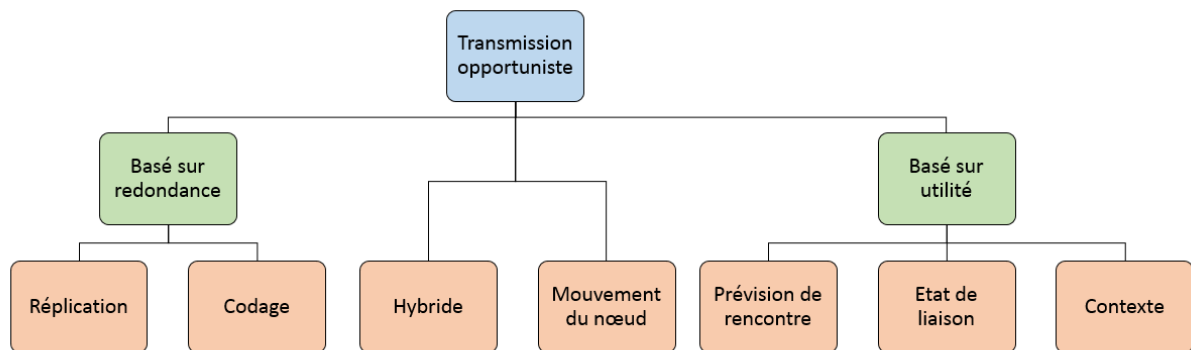


Figure 4-3 : Stratégies de transmission opportuniste

En plus de l'état des paramètres de réseau, le mécanisme basé sur l'utilité utilise aussi des informations contextuelles plus étendues comme l'énergie du nœud, la vitesse de transmission, et la position, etc. Le protocole *Context-Aware Routing (CAR)* [89] profite des attributs contextuels que sont l'énergie résiduelle, le taux de variation de la topologie, la probabilité et la vitesse pour parvenir au domaine cible, etc. L'utilité des nœuds est calculée synthétiquement selon la prévision des valeurs d'attribut en utilisant la théorie d'analyse de série temporelle.

Le CAR étend l'entrée de la table de routage du protocole *Destination-Sequenced Distance-Vector routing (DSDV)* au nœud cible, au saut suivant, au coût et à la probabilité de transmission, grâce à l'échange périodique des informations de table de routage dans le domaine connecté, chaque nœud

conservant la probabilité du nœud de destination. Si le nœud courant et le nœud de destination sont dans le même domaine connecté, il transmet directement ; par ailleurs, le message est transmis au nœud qui a la probabilité la plus élevée pour la destination dans le domaine connecté actuel.

Il est complexe de calculer la valeur d'utilité en considérant les contextes, car le poids des différents paramètres influence fortement la performance de transmission.

En fait, il existe déjà un certain nombre de protocoles de transmission basés sur des informations qui ont des caractéristiques de contexte social. La plupart de ces protocoles utilise la structure communautaire comme des métriques de routage, et la calcule sur le graphe social détecté grâce à des contacts du monde réel entre les appareils mobiles. Comme la couche intermédiaire le montre dans la figure 4-3 Bubble Rap [90] est un protocole basé sur le contexte social en utilisant deux valeurs de centralité qui sont associés à chaque nœud sur la base de la popularité mondiale du nœud dans l'ensemble du réseau et de la popularité locale au sein des communautés. Le système de transmission utilise ces valeurs de centralité de sorte qu'un message est transféré à des nœuds avec des valeurs de centralité globales plus élevées jusqu'à ce que le nœud porteur rencontre un nœud avec la même étiquette de communauté en tant que nœud de destination. Un message est transmis aux nœuds avec des classements locaux plus élevés jusqu'à la livraison réussie.

Seuls quelques travaux prennent des décisions en utilisant également l'information sociale extraite de contacts virtuels ou auto-déclarés, qui correspondent à la couche haute de la figure 4-3 Dans [91], un protocole de transmission opportuniste, appelé *Social Role Routing (SRR)*, utilise ce type d'information pour un *bootstrap* du réseau opportuniste. SRR catégorise les nœuds dans des rôles en appliquant la technique d'analyse de réseau social d'équivalence régulière qui partitionne les nœuds en classes où tous les nœuds d'une classe sont connectés aux mêmes classes de nœuds. Avant que le réseau ne démarre, chaque nœud stocke une copie d'un graphique de connectivité de rôle, qui a été préalablement calculé à partir de ces contextes sociaux des nœuds participants, en leur permettant de calculer la distance géodésique entre les rôles. Les messages sont ensuite transmis uniquement aux nœuds

intermédiaires qui se trouvent dans le même rôle, ou dans le rôle adjacent au rôle de la destination.

En plus, la stratégie de dissémination de donnée est aussi importante pour les autres réseaux sans fil de nouvelle génération comme réseaux ad hoc. VANETs ont une nature relativement plus dynamique par rapport aux MANETs en raison des changements rapides dans la topologie du réseau. La conception et la mise en œuvre d'un algorithme efficace et évolutif pour la dissémination d'informations dans les VANETs sont des enjeux majeurs. De manière similaire au réseau opportuniste, dans cet environnement dynamique, un nombre croissant de messages de diffusion redondant augmentera l'utilisation des ressources, ce qui aurait une incidence indirecte sur la performance du réseau. [92] propose une approche adaptative pour la dissémination d'informations dans les VANETs. Chaque nœud prend en compte les informations locales recueillies auprès des nœuds voisins. Cette information comprend le nombre de voisins et la distance entre les nœuds. Mais la relation sociale entre les nœuds n'est pas considérée.

4.3 Sécurité de la vie privée du réseau opportuniste et contexte social

4.3.1 Menace de sécurité du réseau opportuniste

Des applications de réseau opportuniste peuvent être mieux adaptées aux environnements sans infrastructure de réseau. Cependant, face à des caractéristiques telles que l'absence de chemin de bout en bout entre le nœud source et le nœud destination, l'interruption des liens de communication et le retard élevé de transmission de données, les réseaux opportunistes souffrent souvent d'un certain nombre de menaces potentielles de sécurité.

4.3.1.1 Menace de transmission

La transmission ou bien le routage est l'enjeu clé, et il est nécessaire d'assurer la confidentialité, la fiabilité et l'intégrité de message. Boudec et al. [93] ont étudiés le mécanisme basé sur la redondance et deux types de menace ont été identifiées. Pour la première menace, du fait d'une forte densité de paquets de données fabriqués et transmis dans le réseau, les services de transmission des

nœuds normaux sont contraints et les ressources réseau diminuent, ou alors le cycle de vie des paquets est modifié de sorte qu'ils vont être éliminés prématurément. La deuxième menace est de faire prendre l'identité de l'autre nœud, comme nœud *Sybil* [94], de sorte que le nœud de destination correcte ne peut pas recevoir des données en temps opportun. En général, dans le processus de transmission de paquets, certains mécanismes d'authentification sont utilisés pour authentifier la source et l'intégrité des données par les nœuds intermédiaires en raison de la demande de sécurité. Mais le mécanisme d'authentification de certificat n'est pas approprié dans une situation intermittente comme le réseau opportuniste. Behrouz et al. [95] propose un système *Identity-Based Cryptography (IBC)* qui peut périodiquement mettre à jour la clé privée et l'identifiant dans le réseau ce qui est plus adapté à la transmission dans les réseaux opportunistes.

4.3.1.2 Menace de comportement de nœud

Les nœuds du réseau opportuniste sont fortement autonomes, et leurs comportements sont entièrement décidés par eux-mêmes, sans contrôle ou gestion par d'autres installations ou nœuds. En règle générale, le nœud normal se déplace dans le réseau et aide d'autres nœuds à réaliser la transmission de message, et il profite bien évidemment du même service des autres nœuds au même moment. Cependant, en fonctionnement réel, le comportement du nœud devient inhabituel en raison, par exemple, du comportement égoïste et malveillant des nœuds. Si ces comportements sont laissés sans contrôle, leur conséquence peut provoquer l'effondrement du réseau.

Les nœuds qui ont un comportement égoïste s'appellent des nœuds égoïstes [96]. Dans le réseau opportuniste, des ressources comme l'énergie du nœud, la mémoire et la bande passante sont limitées, les nœuds les économisent souvent pour leurs propres intérêts, en refusant de fournir des services de transmission pour d'autres nœuds. Selon ses comportements particuliers, les nœuds égoïstes peuvent être divisés en trois types : nœud de trou noir, nœud de trou de ver et nœud auto-stop.

Un nœud de trou noir est le nœud égoïste le plus courant, et les caractéristiques du nœud de trou noir sont de recevoir des paquets à partir

d'autres nœuds, mais de les rejeter de sorte que des paquets ne peuvent plus être transmis à d'autres nœuds [97]. Ce comportement égoïste réduit considérablement le taux de succès de transmission.

Des nœuds de trou de ver partagent une partie des caractéristiques du nœud de trou noir, et en rajoute d'autres. Tout d'abord, il reçoit tous les paquets transmis par les nœuds rencontrés, ensuite il ne rejette pas les paquets, mais les transmet en entier à tous les nœuds rencontrés. Bien que ce comportement égoïste a également achevé la transmission des données, le nœud de trou de ver consomme trop de ressources de bande passante et peut facilement conduire à une congestion du réseau partiel.

Comme son nom l'indique, le nœud auto-stop bénéficie de la transmission de ses propres messages à l'aide d'autres nœuds, mais refuse de fournir des services à d'autres sauf quand ses paquets sont passés par des nœuds de destination. Ce comportement égoïste économise la consommation de ressources, mais réduit sérieusement le taux de réussite de transmission et affecte la performance globale du réseau.

Le nœud malveillant est un terme générique qui caractérise des nœuds qui écoutent en cachette, interceptent ou altèrent des messages en provenance d'autres nœuds. L'impossibilité d'assurer la sécurité des communications est la plus grande menace du nœud malveillant pour le réseau opportuniste, non seulement préjudiciable au processus de communication, mais aussi aux contenus des messages.

4.3.1.3 Menace de la vie privée

Concernant la question de vie privée, les contenus des messages transitent dans le réseau opportuniste, mais les informations de contexte sont aussi impliquées. Dans la mesure où le réseau opportuniste peut faire transiter l'information dans une variété de réseaux hétérogènes, les protocoles de sécurité traditionnels ne peuvent pas être appliqués et il est inévitable d'accéder aux données privées, surtout les contextes. Par conséquent, une fois qu'un nœud malveillant a une possibilité d'intrusion dans le réseau opportuniste, les données privées vont être exposées à l'enlèvement et à la falsification. Pour lutter contre les problèmes d'atteinte à la vie privée dans les

réseaux opportunistes, certains chercheurs ont proposé des stratégies de protection connexes.

Dans le processus de transmission utilisant le contexte comme la position ou les ressources de réseau, et afin d'éviter aux données privées d'être accessible directement, Parris et al. [98] ont proposé une stratégie de protection pour le mécanisme de transmission basée sur la redondance qui remplace l'identité du nœud de destination par les données de contexte, et chiffre des données de contexte par un algorithme de chiffrement basé sur l'identité. Les nœuds intermédiaires comparent des données de contexte chiffrées pour trouver le nœud de destination.

En fait, de manière similaire dans l'aspect de contexte social de communautés partageant des caractéristiques communes, le contenu de message est souvent prélevé pour améliorer le mécanisme de transmission pour le réseau opportuniste. Le routage du message est basé sur l'intérêt des nœuds intermédiaire. Shikfa et al [99] ont proposé une solution en terme de confidentialité de transmission sur la base du contenu. Contrairement à des contextes «physiques», l'intérêt des nœuds n'est pas fixe, et il varie selon le changement de l'environnement de réseau et de l'état du nœud lui-même. Il existe aussi des recherches plus relatives aux contextes sociaux comme la relation sociale sur la stratégie de la protection de la vie privée au niveau de la couche application [100].

4.3.2 Mécanisme de la confiance sur la base du contexte social

Aujourd'hui, l'assurance de la sécurité et la fiabilité du réseau opportuniste deviennent une des recherches clé. Le but ultime du réseau opportuniste est l'aide à la transmission des messages par les nœuds. Dans un environnement de réseau réel, des comportements égoïstes et malveillants des nœuds affectent gravement le fonctionnement normal du réseau, et les mécanismes d'authentification et de protection traditionnels sont assez restreints pour résister efficacement à ces comportements. Néanmoins, dans le cadre d'une stratégie de protection élémentaire [101], la confiance, il est possible de prévoir le comportement suivant grâce aux enregistrements des comportements historiques des nœuds. De plus, la valeur concrète quantifiée

qui représente la fiabilité entre les nœuds est présentée pour éviter l'attaque des nœuds malveillants.

Dans l'étude et la recherche sur la confiance sociale dans les réseaux opportunistes, [66] présente la formation de la relation sociale par les attributs sociaux différents, et la force de la relation reflète la force de la confiance entre des nœuds. Sur la base de ce lien, [102] construit un modèle de la confiance pour le réseau opportuniste basé sur la confiance sociale. Le modèle décrit le degré de confiance des nœuds sur deux dimensions ; la première partie est mesurée par la topologie du réseau et la distance en nombre de sauts, et la deuxième partie de la confiance utilise des données historiques. La confiance ultime de nœud synthétise les deux confiances ensemble. D'ailleurs, l'auteur décrit bien le lien d'amitié entre des nœuds par la structure de leur distribution, en confirmant l'identité véridique des nœuds, résistant efficacement de cette façon à l'attaque *sybil* [103]. Dans [104], on construit un modèle de confiance du point de vue de la coopération des nœuds. Sur la base des données de la relation sociale collectionnée sur Facebook et Mobi Clique app, et des chemins de chaque nœud, les nœuds populaires sont d'abord sélectionnés par niveau de coopération des nœuds. Ensuite, d'une part la confiance est mesurée par l'intérêt commun et la relation d'ami entre des nœuds intermédiaires, et d'autre part le meilleur chemin de confiance est déterminé selon l'évaluation entre eux. Et finalement le message est transmis sur ce chemin jusqu'au nœud de destination. Ce modèle utilise des contextes sociaux sur les réseaux sociaux pour aider efficacement la sécurité de la transmission, mais il ne convient qu'au réseau opportuniste à petite échelle en raison de la contrainte de délai.

Bien que certaines des informations de contexte social soient déjà utilisées dans la littérature relative à la sécurité des réseaux opportunistes, elles mesurent la confiance unilatéralement sur des données de la relation des nœuds. C'est-à-dire que quelque parties seulement de contexte sur plusieurs couches de contexte social multidimensionnel sont prises en compte. En plus, la gestion de la confiance du réseau opportuniste a ses propres particularités. A cause d'une intermittence forte de la communication, l'évaluation et le stockage de la confiance vont être réalisés seulement entre les nœuds eux même sans l'aide des diapositives de surveillance tiers. Et comme les nœuds de

réseau opportuniste sont toujours en mouvement, le processus d'évaluation commence en premier lieu lorsque des nœuds se rencontrent. L'algorithme doit donc être conçu de manière simple et efficace pour veiller à ce que le processus soit terminé dans un très court laps de temps, d'une part afin de fournir suffisamment de temps pour la transmission de données, et d'autre part afin d'économiser la consommation de l'énergie. Donc en utilisant notre modèle de contexte social, notre ambition est de fournir un mécanisme plus efficace pour assurer la sécurité dans le réseau opportuniste.

Chapitre 5. Modélisation de contexte social

Dans un premier temps, nous présentons le contexte social auquel nous nous intéressons dans notre recherche.

Ensuite, nous proposons deux modèles pour modéliser respectivement le poids de la relation du réseau social homogène et du réseau social hétérogène correspondant à la vie réelle. Avec ces deux modèles, nous définissons le poids de la relation dans le réseau social multidimensionnel, c'est-à-dire le contexte social considéré à la fois dans le domaine des réseaux sociaux et dans le monde physique.

Basé sur certaines théories des sciences humaines et sociales (sociologie et psychologie sociale), notre approche de modélisation du poids de la relation dans le contexte de réseautage social divise le poids de la relation en deux parties : le poids relatif au caractère de familiarité et le poids relatif au caractère de similitude ; la similitude est elle-même divisée en une similitude dans les communautés statiques et dans les communautés dynamiques selon leur importance. Les contextes sociaux à différents niveaux de raisonnement sont tous introduit pour permettre l'inférence. Nous proposons également une méthode spécifique de calcul basée sur les graphes. Une solution adaptée au recueil des données du contexte social sur le réseautage est introduit. Finalement nous utilisons les données réelles pour conduire notre expérimentation et analysons les résultats en comparant avec les approches existantes.

Pour les réseaux sociaux à large échelle de la vie réelle, nous proposons un modèle, pour construire le réseau social sémantique et évaluer le poids de la relation entre les membres, basé sur les ontologies. Ce modèle est réalisé d'une manière spécifique au réseau social de recherche. La familiarité et la similitude sont aussi utilisées pour la modélisation. Nous utilisons également des données disponibles en ligne sur un réseau social de recherche pour vérifier l'usage et analyser l'effet de notre modèle.

5.1 Poids de la relation

Présenté dans la figure 2-1, selon la hiérarchie du contexte, les informations de contexte peuvent être divisées en contexte primaire, contexte moyen et contexte supérieur. Nous nous intéressons au contexte social supérieur qui renvoie à la mesure du niveau d'intimité entre les utilisateurs, c'est-à-dire à leur degré de proximité sociale. Nous l'appelons ici le poids de la relation.

La littérature [29] indique une notion similaire à notre poids de relation. L'inférence des liens sociaux déduit le lien social entre une paire d'utilisateurs. L'indicateur de lien social contient les informations qui indiquent certaines relations telles que l'ami commun, l'intérêt commun ou la localisation commune. Basées sur l'indicateur de lien social, des techniques d'inférence sont utilisées pour extraire le lien entre une paire d'utilisateurs. L'inférence du lien social vise à extraire la relation entre les utilisateurs de l'information cyber et physique. Cependant, la façon de fusionner et d'intégrer à la fois l'information cyber et physique pour obtenir des relations sociales reste une question ouverte.

Il faut d'abord étudier des contextes sociaux primaires et moyens, que nous présentons en détail dans la partie suivante, pour inférer ce contexte de niveau supérieur. Il peut être utilisé par des applications pour améliorer la performance de traitement et fournir une stratégie de protection de la vie privée. Par exemple nous devons connaître le niveau de confiance si nous devons avoir des contacts avec un autre utilisateur, et ainsi pouvoir décider de la stratégie d'interaction à l'avance et établir les limites de l'autorité pour éviter les problèmes sur les données privés. Dans un autre exemple, certaines des applications d'exécution, peuvent automatiquement décider de leur comportement selon le poids de la relation avec les autres personnes. Comme dans l'application *Do-Not-Disturb* [105], par exemple, où la décision de déranger ou non le propriétaire d'un *Smart Phone* peut être contextuelle et pas seulement basée sur l'état tel que « en train de dormir », au travail ou en réunion, mais aussi sur le poids de la relation entre le propriétaire du smartphone et la personne qui appelle.

La pratique des réseaux sociaux en ligne s'est considérablement développée et a accumulé des données numériques massives sur la société humaine. Ces données massives, représentant les informations personnelles et sociales des individus, nous fournissent des opportunités d'étudier, d'analyser et de modéliser la structure du réseau, les connexions humaines, la similitude entre les personnes, etc. La façon d'extraire et d'exploiter des contextes sociaux efficaces à partir des données massives disponibles pour améliorer les applications et les services a retenu l'intérêt de la communauté de recherche.

Dans la vie réelle, il existe une grande variété de réseaux sociaux du fait de l'échange d'informations entre les personnes ou les organisations : les réseaux sociaux en ligne, les réseaux de recherche, les réseaux d'entreprise et d'autres communautés. Avec une activité sociale plus riche, la taille du réseau social croît, et la structure devient de plus en plus complexe. Ces réseaux sont hétérogènes, contiennent des relations variées, des nœuds différents et des attributs attachés.

5.2 Modélisation du poids de la relation dans la pratique des réseaux sociaux

Pour évaluer le poids de la relation entre personnes, notre approche divise le poids de la relation en caractère de familiarité et caractère de similitude ; en général, dans le réseau de relations sociales, les gens sont toujours plus proches avec quelqu'un qui leur est familier ; nous pouvons donc dire que le poids de relation entre des utilisateurs est fort quand les utilisateurs sont proches ; en outre, parmi les amis avec le même niveau de familiarité, les individus ont souvent tendance à préférer les utilisateurs qui ont des attitudes, des intérêts, des valeurs et une personnalité similaires aux leurs [106].

Comme nous l'avons présenté dans la section 3.3, le contexte social peut être classé en quatre niveaux : contexte social individuel, contexte social binaire, contexte social des communautés statiques (comme l'âge, l'emplacement de la famille, profession, etc.) et contexte social des communautés dynamiques (la personnalité, l'intérêt, etc.).

La littérature [107] souligne que les individus sont plus susceptibles de communiquer avec d'autres individus d'âge similaire, de langue commune, et

de localisation rapprochée. Le degré de relation représente la probabilité de communication entre les utilisateurs. Il a également été mis en évidence que les utilisateurs qui utilisent les mêmes tags ou qui notent les mêmes éléments peuvent partager plus d'intérêts [108].

Le poids relatif au caractère de familiarité correspondant au contexte social dyadique est principalement évalué par l'activité d'interaction entre les deux utilisateurs. Le poids relatif au caractère de similitude correspondant au contexte social des communautés et nous devons inférer par le contexte social individuel. La similitude comprend également la similitude de la communauté statique et la similitude de la communauté dynamique. Dans la situation d'un même niveau de familiarité bas (les personnes ne se connaissent pas), la similitude de communauté statique influence le poids de la relation entre les utilisateurs de manière plus profonde ; par exemple, pour un groupe d'amis qui viennent de faire connaissance, ils préféreront toujours communiquer avec les gens qui leur paraissent similaires (originaires de la même ville, camarades d'écoles), et il est donc plus facile pour eux d'établir une relation plus forte. Dans la situation d'un niveau de familiarité déjà relativement élevé, la similitude de la communauté dynamique influence le poids de la relation entre des utilisateurs de manière plus forte ; par exemple, parmi les amis (en dehors de la famille), les gens vont généralement maintenir des contacts avec les amis qui ont les mêmes intérêts, et dans ce cas, le poids de relation entre ces amis est plus grand. En résumé, le calcul du poids de la relation entre utilisateurs par la familiarité et la similitude peut mieux refléter la réalité du processus de génération de la relation, et être aussi plus proche de la relation entre les utilisateurs dans le monde réel.

Comme nous le présentons dans le chapitre 3, avec l'augmentation des performances des terminaux intelligents mobiles et l'accroissement continu des réseaux sociaux, le contexte de l'utilisateur des réseaux sociaux mobile, qui peut être obtenu, est très riche : localisation de l'utilisateur (via GPS ou localisé par le réseau de téléphonie mobile), état de l'utilisateur (présence, en ligne, hors ligne, en réunions, etc.), agenda, préférences de l'utilisateur, réseaux de relations sociales de l'utilisateur et informations d'évaluation d'amis etc.

Pour calculer le contexte social de niveau supérieur, et le poids de la relation, notre proposition utilise les contextes sociaux de niveau bas tels que : l'échange de messages entre les utilisateurs, l'information des attributs et les préférences de l'utilisateur, les réseaux de relations sociales de l'utilisateur et les informations d'évaluation d'amis etc. La clé de cette proposition est de mesurer le niveau d'intimité entre les utilisateurs, de la même manière qu'un contexte social est utilisé directement par les applications intelligentes, à partir des contextes sociaux disponibles sur les réseaux sociaux.

Ce poids de la relation est évalué à partir de la familiarité et de la similitude des relations humaines entre les utilisateurs. Nous pouvons prendre la formule suivante pour présenter les poids de la relation entre l'utilisateur A (l'utilisateur initial) et l'utilisateur N (l'utilisateur cible) :

$$W(A, N) = FW(A, N) + SW(A, N) \quad (5-1)$$

Dans cette formule, $W(A, N)$ représente le poids de la relation de l'utilisateur A à l'utilisateur N, $FW(A, N)$ représente le poids de la relation amené par la familiarité. Dans l'étude de la familiarité, les gens utilisent souvent la fréquence de communication ou l'effet de rencontre pour mesurer la familiarité, c'est-à-dire que, généralement, plus il y a de communications ou de rencontres, plus les utilisateurs sont familiers. Donc si la fréquence d'interaction entre les utilisateurs est plus grande, la familiarité entre eux sera plus élevée, et plus le poids de la relation entre eux sera plus grand. Selon certaines données [106] issues de la psychologie sociale, nous sélectionnons le nombre d'interactions entre les utilisateurs comme le contexte social pour calculer le degré de familiarité entre les utilisateurs de confiance ; $SW(A, N)$ signifie le poids de la relation engendré par la similitude ; cette partie est divisée entre la similitude de communauté statique et la similitude de communauté dynamique : nous utilisons les attributs des utilisateurs, comme l'âge, l'emplacement de la famille, etc., pour calculer la similitude de communauté statique entre des utilisateurs. De plus, sur la base des informations de préférences des utilisateurs, la similitude de communauté dynamique peut être calculée. Finalement, selon la différence obtenue dans le paramètre de la familiarité, nous rajustons les

valeurs des paramètres entre la similitude de communauté statique et la similitude de communauté dynamique pour finaliser le poids de la relation lié à la similitude.

5.2.1 Poids de la relation amenée par la familiarité

Un graphe dans la théorie des graphes [109] est un ensemble de points nommés nœuds (parfois sommets ou cellules) reliés par des traits (segments) ou flèches nommées arêtes (ou liens ou arcs). L'ensemble des arêtes entre nœuds forme une figure similaire à un réseau, qui est habituellement utilisé pour décrire une relation particulière entre certaines choses. Des points représentent les choses en connectant par les lignes qui représentent la relation respectivement entre les choses. Grâce à une démonstration claire de la topologie entre les choses, ceci a provoqué une large attention de la recherche scientifique et est de plus en plus utilisé, notamment dans le domaine de la stratégie de recommandations sur les réseaux sociaux, [110] par exemple.

Si nous considérons que le réseau social peut être représenté comme un graphe orienté avec une fonction de poids $G(N, E, W)$, illustré par la figure 5-1, où N est l'ensemble des nœuds dans le réseau qui représente tous les utilisateurs ; E les bords orientés entre les nœuds qui signifient la relation de communication entre des utilisateurs, c'est-à-dire qu'il existe une relation de communication si il y a un bord entre deux nœuds, le nœud orienté est le côté de réception. W sont les poids des bords qui représentent le nombre de message de communications unidirectionnelles entre les utilisateurs.

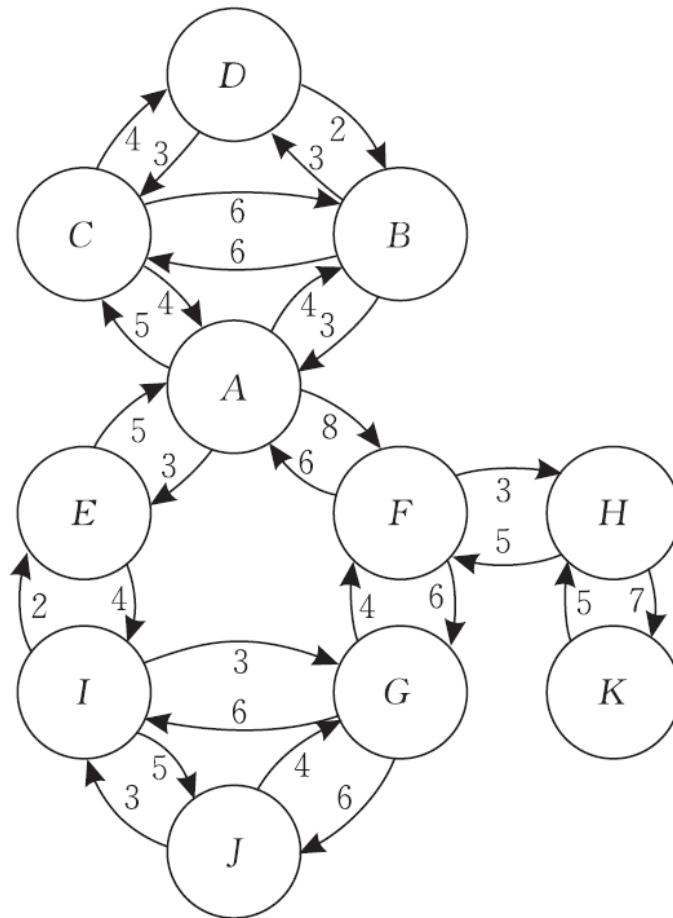


Figure 5-1 : Un graphe d'échange de messages pour le réseautage social

En général, plus il y a d'échanges entre les utilisateurs, plus ils sont familiers. D'autre part, étant donné que la communication est souvent bidirectionnelle, si un seul partenaire envoie des messages mais que l'autre ne répond jamais, nous ne pouvons pas déclarer ces deux utilisateurs comme étant familiers. Nous sélectionnons donc le nombre minimum de messages dans l'interaction comme une mesure de familiarité. Dans la figure 5-1, si nous enlevons le bord de poids le plus grand des deux bords entre deux nœuds et conservons le petit, nous pouvons obtenir le réseau de message minimum entre les utilisateurs, comme illustré dans la figure 5-2.

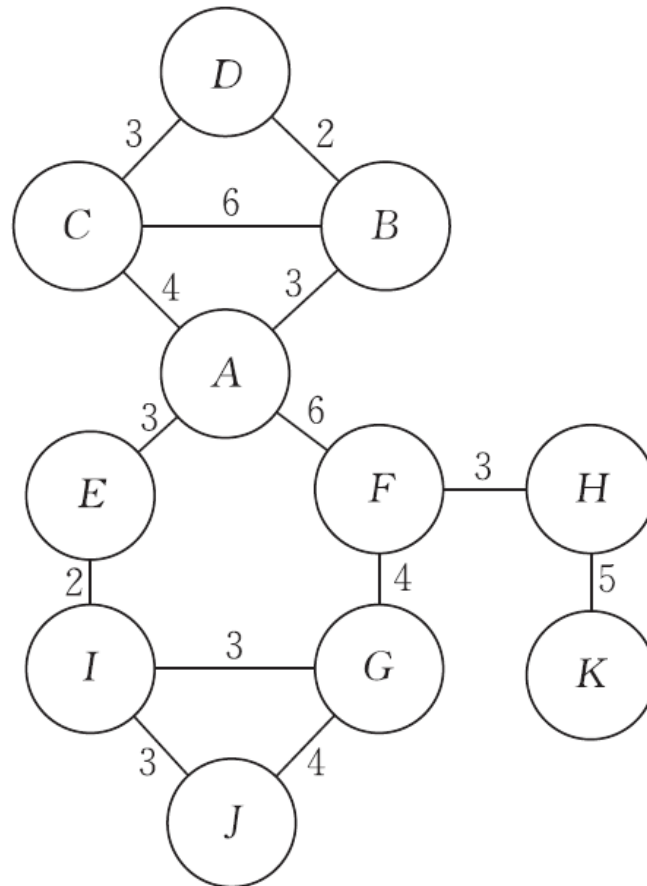


Figure 5-2 : Réseau de messages minimum entre les utilisateurs

Dans la figure 5-2, en prenant l'utilisateur A comme nœud racine par exemple, nous réorganisons tous les utilisateurs qui ont un lien (direct et indirect) avec l'utilisateur A pour former un cercle concentrique centré sur A, comme cela est illustré par la figure 5-3 : les nœuds dans la couche 1 sont les amis directs de l'utilisateur A, la couche 2 sont les amis des amis directs de l'utilisateur A, et ainsi de suite. Selon la théorie des six degrés de séparation [111] (aussi appelée théorie des 6 poignées de main ou théorie de petit monde, etc.), toute personne sur le globe peut être reliée à n'importe quelle autre, au travers une chaîne de relations individuelles comprenant au plus six maillons. Autrement dit, nous pouvons connaître un étranger par au maximum 6 personnes, et par conséquent, tout le monde est dans une des couches du cercle concentrique à 7 couches centré sur A.

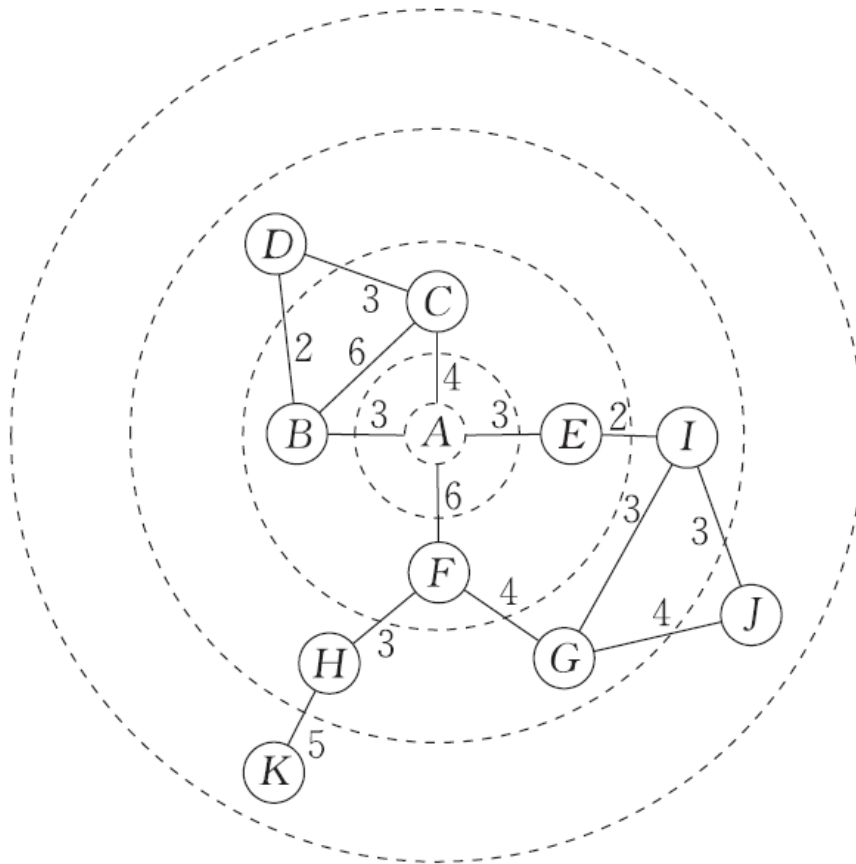


Figure 5-3 : graphe du réseau de nœud cible

Cependant, en raison de la réalité du système de réseau social qui ne peut pas couvrir l'ensemble du monde, ou d'un manque de nœuds correspondant à des utilisateurs clé obligeant le passage par certains utilisateurs etc., il existe des utilisateurs dans des couches en dehors des 7 couches du cercle concentrique, ou qui deviennent des nœuds isolés. Nous conservons donc dans notre étude les utilisateurs connectés par plus de 6 utilisateurs et calculons le poids de la relation liée à la familiarité du nœud racine au nœud cible par la méthode ci-dessous. Particulièrement pour les nœuds isolés, ses poids de relation liés à la familiarité sont mis à 0. Dans la figure 5-3, B et C sont amis de A, D est ami commun de B et C, et pendant le calcul du poids de la relation lié à la familiarité, puisque B et C sont aussi amis, le chemin de A à D est composé de quatre segments $\{(A, B, D), (A, C, D), (A, B, C, D), (A, C, B, D)\}$. Pour les amis indirectement connus, moins il y a de nœuds traversés, plus le poids de relation

est grand ; nous supprimons les liens des nœuds dans la même couche de la figure 4-3, comme illustré par la figure 4-4.

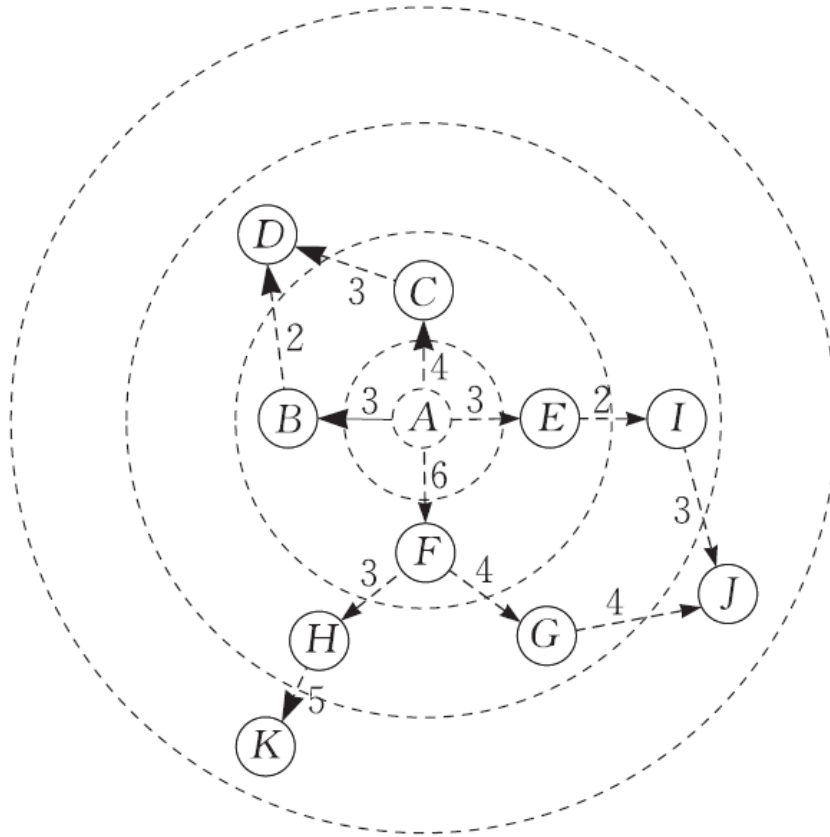


Figure 5-4: graphe du réseau de nœud cible sans lien dans une même couche

Avec l'aide de la figure 5-4, nous pouvons calculer le poids de la relation par la formule suivante :

$$FW(A, N) = w_n \cdot \sum_{i=1}^n \left[\prod_{j=1}^m \frac{N(S_{j-1}, S_j)}{L_j} \right] \quad (5-2)$$

Dans laquelle A est le nœud racine, N est l'autre nœud cible sauf le nœud racine. Les 4 nœuds B, C, E, F sont dans la première couche, et ainsi de suite. i signifie le i ème chemin dans les n chemins de nœud A à nœud N, j signifie la j ème couche de ce chemin, m est la couche où le nœud cible est situé, L_j représente la somme des messages entre tous les nœuds associés dans les couches j et $j - 1$, par exemple, $L_1 = 16, L_2 = 14$. $N(S_{j-1}, S_j)$ est le nombre de messages entre nœud S_{j-1} et nœud S_j , S_j est le nœud de la j ème couche

sur le chemin. En plus, pour représenter la relation loin ou proche de couche du nœud racine A au nœud cible N, nous apportons w_n comme le poids de la couche de nœud A au nœud N, plus il est proche de A, le plus le poids est grand. w_n peut être calculé par :

$$1 - \frac{layer_N}{layer_{SUM}} \quad (5-3)$$

$layer_N$ représente le numéro de couche où le nœud N est situé, $layer_{SUM}$ signifie la somme des couches.

Pour le nœud B par exemple, le poids de relation lié à la familiarité de A à B est $[3/(3 + 4 + 3 + 6)] \times (1 - 1/6)$. Mais pour le nœud D, il y a deux chemins $\{(A, B, D), (A, C, D)\}$ de A à D, et donc le poids de relation amené par la familiarité de A à D est $\{[3/(3 + 4 + 3 + 6)] \times [2/(2 + 3 + 2 + 4 + 3)] + [4/(3 + 4 + 3 + 6)] \times [3/(2 + 3 + 2 + 4 + 3)]\} \times (1 - 2/6)$, le processus de calcul du poids de la relation des autres nœuds est similaire.

Les résultats du poids de la relation lié à la familiarité de l'utilisateur A aux autres nœuds sont indiqués dans le tableau 5-1. Pour tous les nœuds cibles, nous choisissons en priorité celui de poids de la relation le plus grand, et s'il existe des nœuds qui ont la même valeur de poids, nous choisissons le chemin traversant le moins de nœuds, et si le nombre de nœuds traversés est identique, nous choisissons le nœud cible qui a la somme maximum du poids des bords sur le chemin complet, et ainsi de suite. Finalement, nous obtenons la liste avec l'ordre du poids de la relation lié à la familiarité du nœud racine aux autres : F, C, B, E, G, D, H, J, K, I. Le résultat de mise en ordre est conforme à l'analyse ci-dessus.

Nœud	Poids de relation ω
<i>B</i>	0.1563
<i>C</i>	0.2083
<i>D</i>	0.0557
<i>E</i>	0.1563
<i>F</i>	0.3125
<i>G</i>	0.0714
<i>H</i>	0.0536
<i>I</i>	0.0119
<i>J</i>	0.0212
<i>K</i>	0.0167

Tableau 5-1 : exemple du résultat du poids de la relation lié à la familiarité

5.2.2 Poids de la relation lié à la similitude

Sur la base du poids de la relation lié à la similitude, les utilisateurs cibles sont divisés en deux types : familier et non-familier. Nous évaluons le poids de la relation lié à la similitude séparément pour ces deux types d'utilisateurs. Comme nous l'avons déjà présenté, cette partie de poids inclut la similitude de communauté statique et la similitude de communauté dynamique. La méthode de calcul est la suivante :

$$SW(A, N) = \alpha S_{cs}(A, N) + (1 - \alpha) S_{cd}(A, N) \quad (5-4)$$

$SW(A, N)$ représente le poids de la relation lié à la similitude entre l'utilisateur A et l'utilisateur N, $S_{cs}(A, N)$ signifie la similitude de communauté statique de l'utilisateur A et de l'utilisateur N, et $S_{cd}(A, N)$ signifie la similitude de communauté dynamique de A et N. α est le coefficient d'ajustement, lorsque les utilisateurs A et N sont étrangers, la similitude de communauté statique est le rôle le plus important pour déterminer la relation entre eux, et dans ce cas-là nous prendrons $\alpha = 0.6107$ sur la base des résultats d'expérimentation de la littérature [112]. Par contre, si l'utilisateur N est familier avec l'utilisateur A, la

similitude de communauté dynamique joue un rôle plus important ; $\alpha = 0.3983$ alors.

Les contextes sociaux de l'âge et de la localisation de la famille sont sélectionnés pour calculer la similitude de communauté statique en raison du fait que la parenté et l'amitié concerne seulement une petite partie des utilisateurs dans le réseau social et ces caractéristiques sont parfois absentes ou difficiles à obtenir. En plus, du fait qu'ils peuvent être mesurés par la méthode similaire à la localisation de la famille, nous retiendrons l'âge et l'emplacement de la famille du fait de son caractère universel et quantifiable. Puisque nous ne disposons pas de résultats de recherche spécifiques sur le rôle de ces contextes sociaux pour évaluer la relation, mais qu'ils sont tout de même importants pour notre évaluation, nous leur attribuons un poids uniforme. Voici la formule pour calculer la similitude de communauté statique :

$$S_{cs}(A, N) = S_a(A, N) + S_h(A, N) \quad (5-5)$$

$S_a(A, N)$ signifie la similitude de l'âge, et $S_a(A, N) \in [0,1]$, plus cette valeur est grande, plus leur âge est similaire. La méthode de calcul est comme suit :

$$\begin{cases} 1 - \frac{|Age(A) - Age(N)|}{10}, & |Age(A) - Age(N)| \leq 10 \\ 0, & |Age(A) - Age(N)| > 10 \end{cases} \quad (5-6)$$

$S_h(A, N)$ signifie la similitude de l'emplacement de la famille. En général, dans le contexte de la même familiarité d'un utilisateur vers d'autres utilisateurs, avec la même localisation de la famille, moins il y a d'unité régionale administrative, plus le poids de relation est grand ; la méthode de calcul est la suivante: si l'emplacement de la famille des utilisateurs A et N se trouve dans la même rue, $S_h(A, N)$ est égal à 1, et égal à 0.9 si c'est dans le même quartier résidentiel, à 0.8 si c'est la même commune, à 0.5 si c'est dans la même ville, à 0.3 si c'est dans le même département, à 0.1 dans le même pays et à 0 pour le reste.

Nous utilisons la préférence de l'utilisateur comme critère d'évaluation pour la similitude de communauté dynamique ; les utilisateurs qui montrent la même

préférence sont aussi similaires dans leurs intérêts. Dans de nombreuses applications impliquant un grand nombre d'utilisateurs des réseaux sociaux, il n'est pas difficile de recueillir les intérêts des utilisateurs, et nous pouvons également utiliser ces données pour calculer le degré de similitude entre les préférences des utilisateurs. Nous prendrons pour référence ici la méthode de filtrage collaboratif pour calculer la similitude des utilisateurs, et sélectionnons la formule pertinente, à savoir le coefficient de corrélation de Pearson [112] et faisons les ajustements appropriés en tant que mesure de la similitude de communauté dynamique. Supposons que l'ensemble des articles notés en commun par l'utilisateur A et l'utilisateur N est $I_{A,N}$, la similitude de communauté dynamique de l'utilisateur A à l'utilisateur N est :

$$S_{cd}(A, N) = 1 + \frac{\sum_{C \in I_{A,N}} (R_{A,c} - \bar{R}_A)(R_{N,c} - \bar{R}_N)}{\sqrt{\sum_{C \in I_{A,N}} (R_{A,c} - \bar{R}_A)^2} \sqrt{\sum_{C \in I_{A,N}} (R_{N,c} - \bar{R}_N)^2}} \quad (5-7)$$

Dans la formule, $R_{A,c}$ signifie la note de l'utilisateur A donnée à l'article c, \bar{R}_A et \bar{R}_N signifie la note moyenne de l'utilisateur A et l'utilisateur N donnée à l'article. Notez que la note moyenne se réfère à l'article noté en commun par les deux utilisateurs. Le coefficient de corrélation de la similitude obtenue est compris dans l'intervalle [0,2], plus le coefficient de corrélation est grand, plus les intérêts des deux utilisateurs sont similaires. L'ensemble des articles utilisés pour calculer les notes peut être constitué des articles dans les systèmes d'évaluation par exemple de films ou de livres populaires dans les réseaux sociaux.

5.2.3 Processus d'algorithme et analyse de performance

Selon l'algorithme ci-dessus, en supposant A l'utilisateur racine A , N le groupe d'utilisateurs cibles, l'ensemble du processus de calcul de l'algorithme est le suivant:

```

1.  ORD  $\emptyset$ 
2.  for all target user  $n$  in  $N$  do
3.    calculate the weight generated by the familiarity of
       $A$  and  $n$ 
4.  end for
5.  for all target user  $n$  in  $N$  do
6.    if  $n$  is familiar to  $A$  then
7.      calculate the weight generated by the similarity
        of  $A$  and  $n$  with a community static's higher
        weight
8.    else { / * if  $n$  is not familiar to  $A$  * / }
9.      calculate the weight generated by the similarity
        of  $A$  and  $n$  with a community dynamic's higher
        weight
10.   end if
11. end for
12. add the weight generated by familiarity and similarity
13. sort all the users of  $N$  by their weight
14. return ORD

```

La fonction principale de cet algorithme consiste à calculer le poids de la relation lié à la familiarité de l'utilisateur racine A et l'utilisateur cible n (étapes 2 à 4) ; puis, sur cette base, à calculer le poids de la relation lié à la similitude selon la familiarité des utilisateurs (étapes 5 à 11), puis additionner le poids des deux relations (lié à la familiarité et à la similitude) (étapes 12), et enfin

ordonner par le poids de la relation des utilisateurs (étape 13). Lors du calcul du poids de la relation lié à la familiarité, il faut parcourir toutes les utilisateurs cibles relatifs à l'utilisateur racine A, et utiliser la formule 5-2 pour calculer les chemins multiples. En théorie, dans le graphe de N nœuds que la figure 5-4 montre, la complexité la pire en temps de calcul des chemins entre deux nœuds est $O(3^{N/3})$, et par conséquent, la complexité plus grande en temps du poids de la relation lié à la familiarité est $O(N \cdot 3^{N/3})$. Mais dans le réseau social réel, le réseau de relations sociales des utilisateurs est statiquement stocké dans la base de données du système, et la connection entre les nœuds n'est pas si dense, et par conséquent, le nombre de chemins entre deux utilisateurs est constant, de sorte que la complexité en temps de calcul basé sur la familiarité est $O(N)$. Dans les étapes de calcul du poids de la relation lié à la similitude, la similitude de communauté statique est calculée en se basant sur les attributs de l'utilisateur, et les attributs sont généralement des quantités constantes, et la complexité en temps de cette étape est donc $O(N)$. Mais pour la similitude de communauté dynamique, il faut parcourir l'ensemble Pour des articles notés en commun par les deux utilisateurs, supposant M, la complexité en temps est $O(N \cdot M)$. Dans le système réel, M est aussi une quantité constante, et la complexité en temps de la similitude de communauté dynamique est donc $O(N)$. En raison de la nécessité de parcourir l'ensemble des utilisateurs cibles N, la complexité en temps de calcul du poids de la relation lié à la similitude est $O(N^2)$. En outre, la complexité en temps des algorithmes de mise en ordre en général est également $O(N^2)$. Ainsi, la complexité en temps global de l'algorithme est $O(N^2)$.

5.2.4 Acquisition de données sur le réseautage social

Facebook est le réseau social le plus populaire dans le monde. Il est parvenu à retenir l'attention de tous dans le monde, y compris les célébrités, les commerçants, les politiciens, les artistes et les chercheurs en démographie. Facebook est également un référentiel complet qui intègre diverses informations utilisateur, des applications sociales, des communautés d'utilisateurs, des événements, etc. En outre, Facebook se préoccupe (tout du moins officiellement) de la confidentialité des utilisateurs et fournit aux utilisateurs des fonctions personnalisées de confidentialité. Par conséquent,

Facebook est un outil de réseau social idéal et représentatif pour recueillir des données et mener des études.

Dans la mesure où les données sont la base pour l'étude des comportements des utilisateurs et l'analyse des caractéristiques du réseau, de nombreuses approches ont été proposées afin de les recueillir. Dans cette section, nous allons brièvement présenter quatre méthodes typiques de collecte de données: la réalisation d'enquête ou d'entretiens, la capture du trafic réseau, le robot d'indexation de la page en ligne et l'accès à l'interface API.

La conduite d'un sondage ou d'un entretien est le moyen le plus traditionnel de recueillir des données dans le cadre d'une recherche visant à caractériser des comportements d'usagers. De cette façon, les chercheurs recueillent des données en demandant à un certain nombre de personnes de remplir un ensemble de questions personnalisées. Bien que nous soyons capables de poser volontairement des questions pour la recherche, les échantillons sont généralement réduits à une petite échelle et le coût induit peut s'avérer important (recrutement des sujets, réalisation des entretiens et analyse des réponses).

L'évolution du réseau laisse une énorme quantité de traces des comportements en ligne des utilisateurs. Capturer les trafics de réseau générés par les personnes accédant à des sites en ligne ou communiquant par le réseau est l'une des méthodes largement utilisées. De nombreux outils (par exemple, Wireshark, Fiddler et NetworkMiner) peuvent nous aider à surveiller et à capturer librement ces trafics réseau. Néanmoins, ces outils peuvent capturer seulement le trafic réseau passant par le même routeur. Et les données d'émission sont habituellement chiffrées, de sorte que nous ne pouvons pas acquérir les informations détaillées telles que le contenu d'un site en ligne ou les informations de base d'un utilisateur.

Au lieu de capturer le trafic réseau, l'exploration des pages en ligne est une méthode alternative largement appliquée. Un robot d'indexation (*crawler*), en tant qu'agent demandant le serveur, peut extraire des informations contextuelles de la page en ligne accédée. Le robot continue à « ramper » et à sauter à une autre page liée. Un avantage concurrentiel du *crawler* est qu'il

peut recueillir toutes les informations détaillées que nous pouvons afficher sur un navigateur. En outre, les données analysées conservent généralement une topologie des nœuds analysés dans le graphe réseau.

Presque toutes les plates-formes de réseau social, telles que Facebook et Twitter, fournissent des API publiques pour les tiers ou l'individu afin de recueillir des données. La demande API est relativement simple et les données proposées sont structurées. Malheureusement, les données disponibles via l'API sont limitées et une autorisation (*access token*) est nécessaire, ce qui peut ne pas satisfaire les exigences des chercheurs pour mener des études approfondies et des analyses.

Nous utilisons à la fois les méthodes de *crawler* et les API pour collecter notre ensemble de données de Facebook. En particulier, nous mettons en œuvre le *crawler* pour saisir les profils des utilisateurs, y compris l'information démographique, les relations sociales et les intérêts des utilisateurs, et nous utilisons l'API pour recueillir l'information d'interaction en ligne entre eux. Nous avons demandé l'autorisation des étudiants dans le Groupe UTT qui a les échantillons suffisants, mais après une période de recueil de données, nous avons estimé que la communauté n'était malheureusement pas assez dynamique pour fournir les informations de contextes sociaux intéressantes.

Pour simuler une situation plus réelle et satisfaire les besoins d'un grand nombre d'échantillons expérimentaux, nous avons demandé l'aide d'autres institutions de recherches scientifiques. Finalement, l'institut d'application informatique de l'Université de Jilin, où j'ai obtenu ma licence, a accédé à notre requête et a accepté que nous utilisions leurs données de laboratoire. Vingt-deux étudiants de master et doctorants de l'équipe ont participé au recueil de données de leurs amis basés sur leurs propres cercles ; leurs amis qui sont actifs sur le réseau social ont également été mobilisés pour collecter plus de données. La source de données est un célèbre site de réseau social chinois *Renren* [113], qui fournit les mêmes services et applications que Facebook. Globalement, les données dynamiques de 1286 utilisateurs ont été recueillies pendant un mois. Selon la relation d'amis entre ces utilisateurs, un petit réseau social complexe a pu être formé, qui contient 1286 nœuds et 21612 bords. Afin de mieux comprendre ce réseau social, il est divisé en 19 sous-communautés

par algorithme de division de communauté *Label Propagation*. Comme illustré par la figure 5-5, les clusters de nœuds en différentes couleurs représentent différentes sous-communautés.

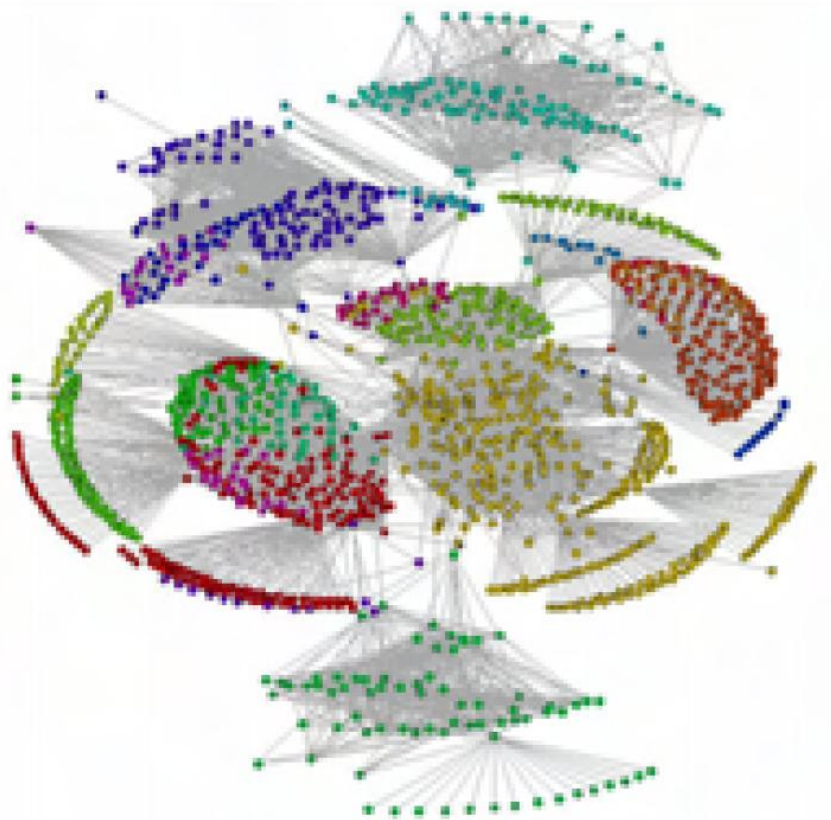


Figure 5-5: communauté de réseau de relation recueillie sur Renren

Pour les expériences suivantes, nous choisissons un utilisateur aléatoire comme utilisateur racine et parcourons aléatoirement une certaine communauté de l'application. Nous faisons l'acquisition du nombre d'interactions (message laissé, messagerie instantanée, téléphone en ligne, etc.) entre l'utilisateur racine et ces utilisateurs correspondant au contexte social moyen pour calculer le contexte supérieur : le poids de la relation lié à la familiarité. Le contexte primaire, ses propriétés de base (âge et emplacement de la famille sélectionnée pour l'expérimentation) permet d'inférer le poids de la relation lié à la similitude de communauté statique. Nous avons recueilli les notes d'articles (communauté géographique par exemple) qui sont notés en commun par l'utilisateur racine et l'utilisateur cible pour calculer le poids de la relation lié à la similitude de communauté dynamique.

d'attribut comme le nom et les façons de contact de l'utilisateur. Sous la condition que la plupart des utilisateurs remplissent seulement la ville pour l'emplacement de la famille en ligne, nous ajustons S_h pour calculer le poids de la relation lié à la similitude de communauté statique : si l'emplacement de la famille des utilisateurs est dans la même commune, $S_h = 1$, la même ville 0.9, le même département 0.8, située dans la même région (nord-est, nord, sud, etc.) 0.7, au nord ou au sud en commune 0.6, le reste 0.

ID	Age	Emplacement de famille	Lycée
193	25	Ville Weihai Département Shandong	Lycée 2 Commune Rongcheng
065	25	Ville Weihai Département Shandong	Lycée 2 Commune Rongcheng
318	24	Commune Haidian Ville Beijing	Lycée 11 Ville Beijing
357	24	Ville Yantai Département Shandong	Lycée 1 Commune Longkou
437	25	Ville Heze Département Shandong	L-expérimental Commune Dong
455	25	Ville Jinyang Département Anhui	L-banlieue Ville Jinyang
531	35	Ville Yicheng Département Anhui	Lycée 2 Ville Yicheng
588	24	Ville Jmen Département Guangdong	Lycée 1 Ville Jmen
605	24	Ville Lyang Département Liaoning	Lycée supérieur 1 Ville Lyang
928	24	Ville Nanping Département Fujian	Lycée 1 Commune Guangze

Tableau 5-2: Données d'attribut d'une partie des utilisateurs

Les résultats partiels de calcul sont listés dans le tableau 5-3, et nous sélectionnons les 30% plus grands poids de la relation liés à la familiarité comme les utilisateurs familiers pour évaluer les deux parties du poids de la relation lié à la similitude. Les résultats finaux de l'expérimentation sont mis en ordre sur la valeur du poids de relation entre l'utilisateur ID193 et les utilisateurs cibles pour réduire le temps et l'effort de filtrage et de sélection de l'utilisateur.

ID	Familiarité	Similitude-communauté-statique	S-c-dynamique
065	0.0632478632476832	0.840	2.0
318	0.126495726495726	1.750	1.6
357	0.00769230769230769	1.482	1.7
437	0.00728194399742516	1.959	1.8
455	0.0905982905982905	1.945	1.0
531	0.0000000000000000	1.088	0
588	0.00197288279222498	1.781	0.9
605	0.02655294496298680	0.985	1.5
357	0.0000000000000000	0.978	0.9

Tableau 5-3: une partie des résultats d'expérimentation

5.2.6 Vérification de l'efficacité du modèle

Afin de vérifier l'efficacité de notre modèle du contexte social concernant le poids de la relation entre les utilisateurs du réseau social, la méthode *TidalTrust* d'inférence de confiance dans un réseau social [114] est analysé et comparée avec la nôtre. La valeur du niveau de la confiance de cette approche peut représenter et posséder le même effet que le poids de la relation de notre proposition. La logique de l'algorithme *TidalTrust* est que tous les utilisateurs du réseau social donnent des notes de confiance aux amis qui lui sont directement liés, puis sur la base de ce réseau de niveau de confiance, la confiance entre les amis indirectement liés peut être aussi calculée. Basé sur le jugement subjectif de l'utilisateur, il est plus facile de faire une évaluation de la relation relativement précise pour ses amis, mais pour les amis des amis et les utilisateurs de la relation plus lointaine, l'exactitude de l'évaluation sera considérablement réduite (parce que la note d'utilisateur devient de plus en plus arbitraire). Par conséquent, afin de réduire la subjectivité et l'arbitraire de la note de l'utilisateur à d'autres utilisateurs qui ne lui sont pas directement reliés, nous utilisons l'algorithme *TidalTrust* pour calculer le niveau de relation entre les amis indirectement liés, et puis nous analysons et comparons ses résultats avec notre modèle. En fait, dans la plupart des situations, la valeur

absolue de la relation n'est pas d'une grande importance pour l'utilisateur, le degré relatif de la relation est plus intéressant, tant qu'il est suffisant que l'utilisateur puisse distinguer l'ordre de la relation avec d'autres utilisateurs. Nous comparons alors l'ordre de confiance de l'algorithme *TidalTrust* à la mise en ordre du poids de la relation par notre méthode de calcul.

Afin d'évaluer la précision de la méthode de calcul du poids de la relation proposée, nous utilisons l'indice d'évaluation *NDCG* (*Normalized Discounted Cumulative Gain*) pour mesurer la fidélité de l'ordre de poids de la relation. *NDCG* est un des outils d'évaluation de mise en ordre largement utilisé dans le monde de la recherche ; *NDCG* reflète des résultats de recherche de forte corrélation très utiles et des résultats de recherche de faible corrélation et de non corrélation. Le résultat de forte corrélation se situe en haut de la liste (*classement* le plus élevé), il est le plus utile ; on doit s'interroger et le modifier si un résultat de forte corrélation obtient un mauvais classement. La définition du *NDCG* est :

$$NDCG_p = \frac{DCG_p}{IDCG_p} \quad (5-8)$$

La formule pour calculer DCG_p pour notre expérimentation est la suivante :

$$\sum_{i=1}^p \frac{2^{w_i-1}}{\log_2(1+i)} \quad (5-9)$$

w_i représente le poids de la relation de l'utilisateur (le nœud) i , p représente le nombre d'utilisateurs cibles mis en ordre. $IDCG_p$ signifie le résultat de DCG_p avec la condition parfaite de mise en ordre, l'algorithme parfaite de mise en ordre conduite $IDCG_p$ et DCG_p en même valeur, donc $NDCG_p$ sera 1. En réel, la valeur de *NDCG* est entre 0 et 1 ; plus elle est proche de 1, plus l'algorithme est exact.

Nous utilisons les données concernant les notes de niveau de confiance données aux amis de 50 utilisateurs qui ont des liens d'amitié relativement denses ; ce sous-réseau de relation sociale comprend 476 nœuds d'utilisateur. Puis sur la base de ce sous-réseau, nous avons fait trois analyses

expérimentales de trois différents nœuds d'utilisateur racine. Trente nœuds d'utilisateurs cibles ont été sélectionnés respectivement pour faire la comparaison selon la règle suivante. Avec le changement du nœud racine et des nœuds cibles, chaque expérimentation a été répétée sept fois pour établir des statistiques. Les expérimentations 1 et 2 sont respectivement dans le cas du rapport d'amis de 60% et 20% de l'utilisateur cible, les utilisateurs cibles de l'expérimentation 3 sont choisis au hasard.

Comme nous pouvons le voir à partir des résultats expérimentaux indiqués dans la figure 5-7, si les amis de l'utilisateur racine constituent la majorité des utilisateurs cible, la performance de fidélité de l'algorithme *TidalTrust* est meilleure que notre approche. C'est principalement parce que le poids de la relation est donné directement par l'utilisateur, alors que notre approche le calcule de manière synthétique en se basant sur la familiarité et la similitude entre les utilisateurs. Certains des amis sont assez familiers mais il est probable que la quantité d'interaction et la similitude entre eux soient faibles. Avec la réduction du nombre d'amis de l'utilisateur cible, notre méthode est meilleure que l'algorithme *TidalTrust* en terme de précision, car notre approche envisage complètement la familiarité et la similitude entre les utilisateurs, ce qui peut creuser profondément la relation entre les utilisateurs, de sorte qu'il correspond mieux avec la situation réelle de l'utilisateur. En outre, puisque la condition préalable de l'algorithme *TidalTrust* est l'existence des notes de relation de l'utilisateur à tous ses amis dans le réseau social, et que le poids de la relation des autres individus qui lui sont indirectement liés peut être calculé, la qualification de cette condition est difficile dans les réseaux sociaux de grande taille. Notre méthode correspond plus à la demande d'évaluation du poids des relations entre les utilisateurs dans des applications réelles, et a un domaine d'utilisation plus large.

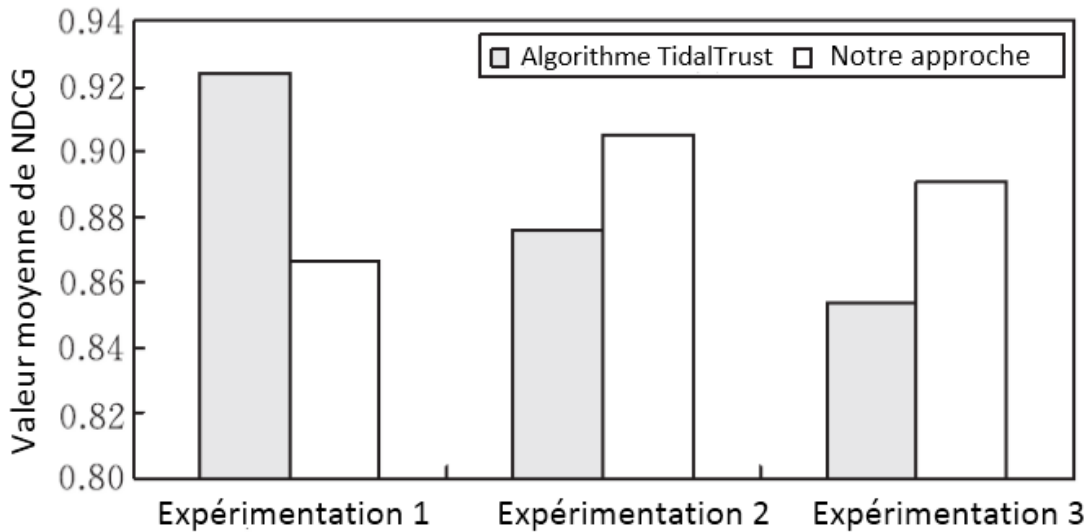


Figure 5-7 Résultat d'Analyse

5.3 Modélisation du poids de la relation pour le réseau social réel

Notre approche pour modéliser le poids de la relation de réseautage social est basée sur la théorie des graphes. Mais généralement, le réseau social est une structure réticulaire de relations à la suite de l'échange d'informations entre personnes et/ou organisations. Au sens large, il existe beaucoup d'exemples de réseaux sociaux dans le domaine des sciences et technologies, du commerce, de l'économie et de la biologie etc. Un exemple de réseau social du commerce, [115] utilise la représentation basée sur l'ontologie pour intégrer les données commerciales aux informations sociales. Il exploite les données hétérogènes extraites des e-mails, des forums et des réseaux sociaux pour soutenir des activités de marketing telles que la réputation de la marque, la découverte de thèmes et l'analyse des sentiments. La société peut exploiter les opinions des clients pour avoir une vision complète de ses propres affaires, et aussi pour mieux comprendre la position sur le marché des concurrents directs.

Traditionnellement le réseau social peut être représenté à l'aide de graphes ou de matrices. Mais dans la mesure où ce moyen ne peut pas exprimer de manière uniforme la variété des relations et des nœuds des réseaux hétérogènes, et du fait de mauvaises performances dans la recherche

d'information, de compréhension et de facteur d'échelle, il est plus adapté à un certain type de réseau social simple des nœuds et des relations unitaires, et avec une information de propriété simple. Grâce aux caractéristiques des contextes sociaux, qui satisfont ces conditions, les résultats du calcul du poids de la relation de réseautage social sélectionné pour notre approche ci-dessus sont déjà de niveau assez élevé, ce qui montre que la méthode basée sur les graphes peut très bien fonctionner. Pour les autres cas plus complexes, nous proposons un nouveau modèle basé sur les ontologies pour les réseaux sociaux hétérogènes qui contiennent des relations variées et des nœuds différents, caractéristiques propre à la vie réelle.

La première étape de ce modèle a pour but de construire le réseau social sémantique dérivé du réseau social réel. La deuxième étape est également l'évaluation du poids de la relation du réseau social sémantique.

Pour la construction du réseau social sémantique, nous extrayons d'abord les concepts et les relations des ontologies à partir du réseau social réel et ses données de description originales, puis l'annotation sémantique des données originales peuvent être effectuées en utilisant des ontologies ; de cette manière, le réseau social exprimé par le langage des métadonnées peut être élaboré. La définition des ontologies et la relation entre les ontologies sont des abstractions et de la modélisation du réseau social, et le réseau social produit des nouvelles relations selon la règle d'inférence des ontologies. Ce processus construit le réseau social sémantique à deux couches : la couche de réseau de concepts de l'ontologie et la couche de réseau social.

Pour l'évaluation du poids de la relation dans le réseau social sémantique, il faut d'abord obtenir l'attribut de la relation des classes depuis la couche de réseau de concepts de l'ontologie. Et puis nous évaluons la relation de familiarité par l'attribut de relation qui satisfait la condition selon laquelle l'ensemble de définition (*Domain of a function*) et l'image d'une application (*Range*) sont tous dans la classe cible, nous évaluons la relation de la similitude en utilisant l'inférence d'ontologies par d'autres attributs de relation. Puis nous mesurons le poids de la relation relatif à la familiarité et à la similitude séparément sur le nombre de relations entre les entités et sur le degré de correspondance des entités voisines, et le poids de la relation globale par la

fusion des relations multidimensionnelles. Enfin, les résultats sont analysés et évalués.

Nous prenons l'exemple du réseau de recherche scientifique pour expliquer le modèle ci-dessus. Le réseau de la recherche est un réseau social hétérogène qui contient des relations variées et des nœuds différents et l'information d'attribut attaché. Il existe des types différents de nœuds dans le réseau, par exemple « le chercheur », « l'article », « la conférence », « le journal » et « l'institution » etc. Des relations peuvent émerger entre les mêmes types et également entre des nœuds de différents types. Par exemple, entre les chercheurs il existe la relation de « coopération » et de « collègue » ; entre les « chercheurs » et les « articles » il existe la relation de « citation » ; entre les « chercheurs » et « l'institution » il existe la relation « appartient à ». En plus de l'information des nœuds et des relations, il existe également des informations d'attribut, par exemple, l'intérêt de recherche du chercheur, l'année de publication du papier, etc.

5.3.1 Modélisation du réseau social sémantique

La technologie de réseau sémantique est la base du réseau social sémantique. Selon la déclaration de Mika [116], l'idée de réseau sémantique transfère les informations non-structurées à une description exploitable par la machine, avec pour but de fournir les informations manquantes nécessaires. Le cadre du réseau sémantique comprend un modèle de graphe (RDF), une sorte de langage de requête (SPARQL) et les cadres de définition de modèle (OWL et RDFS). Grâce à la technologie de réseau sémantique, nous pouvons définir le modèle de métadonnées universelles: ontologie, et établir un cadre universel pour l'organisation des connaissances et le partage de la sémantique. L'ontologie est un outil de modélisation conceptuelle qui peut décrire l'information au niveau sémantique et au niveau des connaissances. OWL et RDF sont des langages d'ontologie recommandés par le W3C (*World Wide Web Consortium*).

Nous considérons que le réseau social sémantique se réfère à un réseau social qui utilise l'ontologie pour sa description. Le modèle se compose de deux couches, la couche de réseau de concept d'ontologie et la couche de réseau

social. La couche de réseau de concepts de l'ontologie est la couche du modèle, y compris les concepts, les relations entre les concepts et les contraintes. La couche de réseau social est la couche de données, contenant les instances de l'ontologie. Le modèle de réseau social sémantique peut être exprimé comme $SSN = \{O, SN, M\}$, où SSN représente le réseau social sémantique, O représente la couche de réseau de concept d'ontologie, qui est l'abstraction et la modélisation du réseau, SN représente la couche de réseau social, avec la fonction de stockage des informations spécifiques du réseau social, M représente la relation d'application (*map* en anglais) entre les deux couches.

La figure 5-8 est un exemple simple de réseau social sémantique. La couche de réseau de concepts de l'ontologie comprend une partie des concepts et les relations entre eux pour le réseau de recherche. S , P , E , J , C signifient respectivement la classe de chercheur, la classe de papier, la classe d'événement [117], la classe de journal et la classe de conférence. La couche de réseau est le réseau social mémorisé dans le fichier de métadonnées, après le traitement d'annotation par ontologie sur les données originales. La ligne en pointillé indique la relation d'application entre les deux couches. Nous montrons par exemple l'application de classe de chercher à l'instance de chercheur et l'application d'attribut de rédiger à leur instance d'attribut de rédiger.

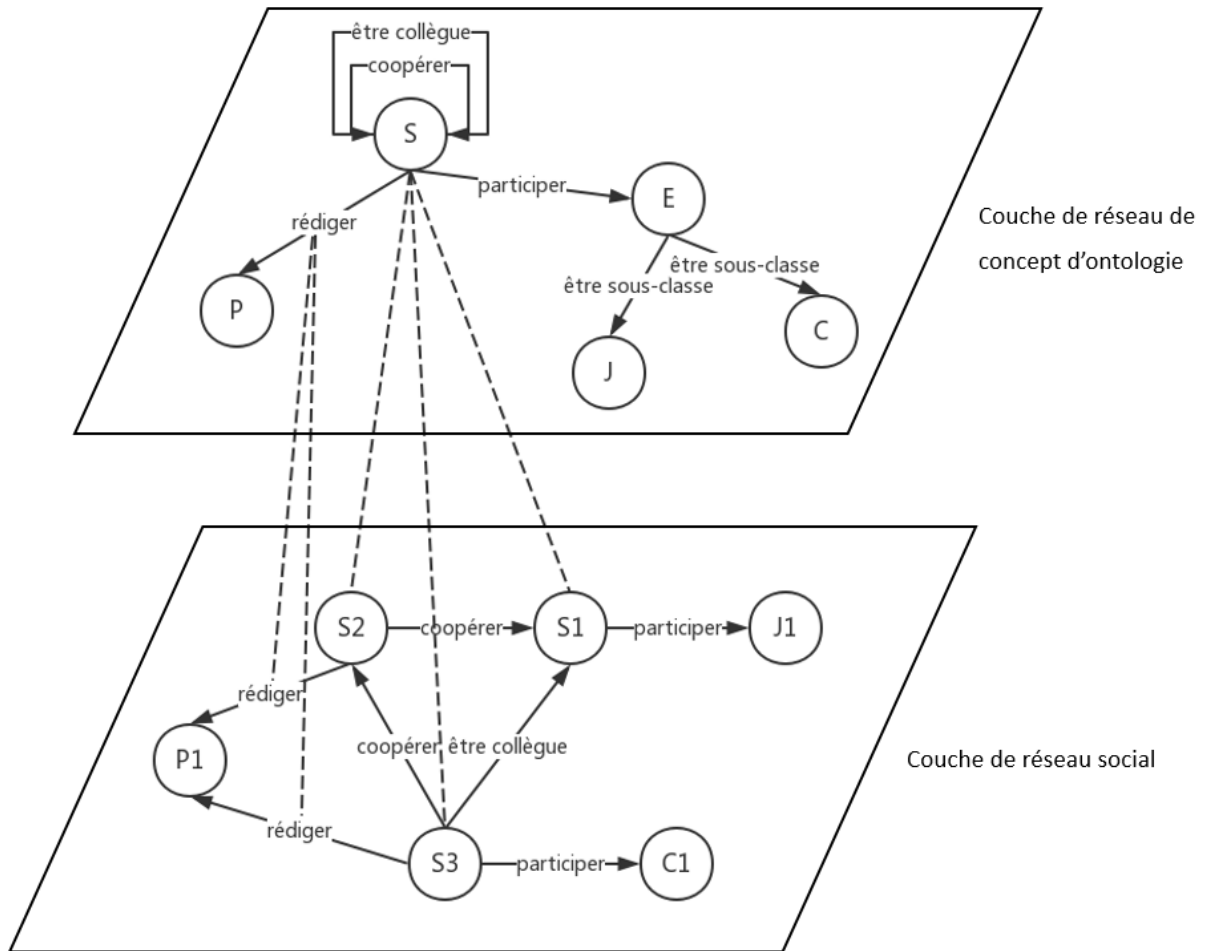


Figure 5-8 : exemple de réseau social sémantique

5.3.1.1 Couche de réseau de concepts de l'ontologie.

La couche de réseau de concepts de l'ontologie contient des descriptions de niveau concept pour la structure du réseau social par abstraction et modélisation du réseau social. Elle peut être exprimée comme $O = \{C, R, P\}$, où O signifie le réseau de concept d'ontologie, C l'ensemble des concepts d'ontologie, R l'ensemble de relation entre les concepts d'ontologie, P la contrainte des concepts d'ontologie, c'est à dire que l'ensemble des axiomes.

Nous avons extrait les concepts et les relations entre eux depuis les données originales de description de réseau de recherche réel. La figure 5-9 représente les relations des concepts d'ontologie. Elle montre les classes et les attributs d'ontologie. Les attributs d'ontologie comprennent deux types : l'attribut d'objet et l'attribut de relation. L'attribut d'objet indique des attributs, comme le nom d'auteur, le numéro et l'année du papier par exemple, possédés par la classe individuel. L'attribut de relation se réfère à l'attribut de relation entre les classes, qui peut être de même type comme les relations de coopération et de collègue, mais aussi entre les différentes classes comme l'attribut de rédiger et de citer entre auteur et papier.

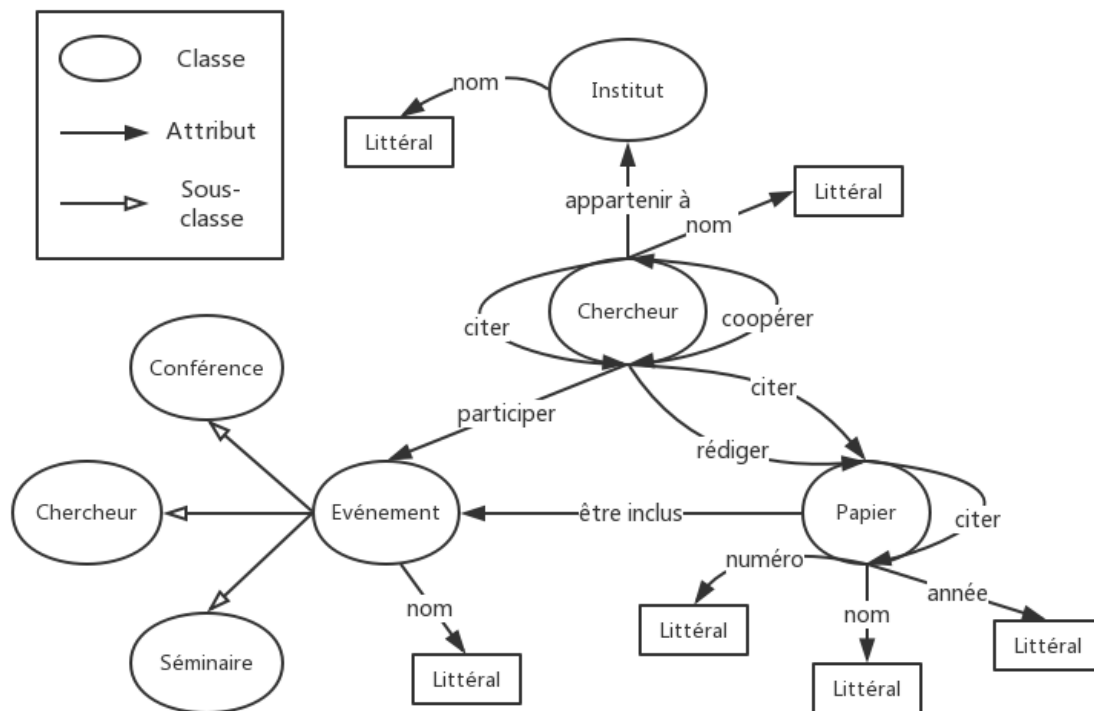


Figure 4-8 : Relations des concepts de l'ontologie

La capacité d'inférence est un point fort de l'ontologie. Pour notre recherche, la caractéristique de raisonnement sur les ontologies nous fournit des conditions pratiques pour inférer le poids de la relation entre les membres du réseau. Par exemple une définition de la règle de raisonnement, lorsque le chercheur A et le chercheur B sont tous l'auteur d'un papier, alors A et B ont la relation de coopération.

5.3.1.2 Couche de réseau social.

La couche de réseau social est la couche de données du réseau social sémantique. Elle contient l'information d'instance d'ontologie du réseau social. De manière similaire à la couche de réseau de concepts de l'ontologie, la couche de réseau social est aussi exprimée par un langage de métadonnées et peut être formalisée comme $SN = \{V, A, E\}$, où SN représente la couche de réseau social, V l'ensemble des nœuds du réseau, A un ensemble d'attributs et E l'ensemble des relations entre les nœuds.

L'ensemble des relations entre les nœuds E de la couche de réseau social permet l'existence d'une certaine redondance, c'est-à-dire que certaines relations peuvent être inférées par raisonnement d'autres relations selon P de la couche de réseau de concepts de l'ontologie. Par exemple nous pouvons obtenir la relation de coopération et la relation d'être cité en commun, etc. pour établir des règles de raisonnement.

5.3.2 Poids de la relation entre « chercheurs »

Le poids de la relation de familiarité entre deux chercheurs est calculé par le nombre des relations entre deux entités, et le poids de la relation de similitude est évalué par le niveau de correspondance des autres entités connectées aux deux chercheurs. Basé sur la sémantique, le niveau de correspondance peut être divisé en trois dimensions : la dimension « citer », la dimension « coopération » et la dimension « événement en commun » (participer conjointement à une conférence ou publier l'article dans le même journal par exemple).

5.3.2.1 Dimension « citer » et évaluation de son poids de relation

La dimension de la relation de citer est une dimension formée par le fait que l'information est citée par l'auteur, y compris la relation d'être cité en commun, la relation de citer en commun et la relation de citer en direct.

En supposant que le chercheur A a un papier C, le chercheur B a un papier D, les papiers C et D sont cités en commun par le papier E, nous pouvons inférer qu'il existe la relation d'être cité en commun entre les chercheurs A et B. Nous utilisons u_A et u_B pour représenter les chercheurs A et B, cd_A et cd_B pour

représenter respectivement le papier qui cite A et le papier qui cite B. Le poids de relation d'être cité en commun $W_{c1}(u_A, u_B)$ peut être calculé comme suit :

$$W_{c1}(u_A, u_B) = \frac{|cd_A \cap cd_B|}{\min\{|cd_A|, |cd_B|\}} \quad (5-10)$$

Dans le calcul de la relation de citer en direct, nous considérons les deux directions de citer entre le chercheur A et le chercheur B. c_{AB} et c_{BA} signifie respectivement le papier du chercheur B cité par le chercheur A et le papier de A cité par B. Le poids de relation de citer en direct mutuellement $W_{c2}(u_A, u_B)$ peut être calculé comme :

$$W_{c2}(u_A, u_B) = \frac{c_{AB}}{cd_B} + \frac{c_{BA}}{cd_A} \quad (5-11)$$

Si le chercheur A a un papier E et le chercheur B a un papier F, les papiers E et F citent un papier G en commun, nous disons des auteurs A et B qu'ils ont la relation de citer en commun. c_A et c_B signifie respectivement le papier cité par A et le papier cité par B. Le poids de relation de citer en commun peut être calculé comme :

$$W_{c3}(u_A, u_B) = \frac{|c_A \cap c_B|}{|c_A \cup c_B|} \quad (5-12)$$

Et finalement le poids de relation de citer entre le chercheur A et le chercheur B est calculé comme :

$$W_c(u_A, u_B) = W_{c1}(u_A, u_B) + W_{c2}(u_A, u_B) + W_{c3}(u_A, u_B) \quad (5-13)$$

5.3.2.2 Relation de coopération et évaluation de son poids de relation

La relation de coopération est formée grâce à l'article co-rédigé par des chercheurs. p_{AB} signifie l'article co-rédigé par les chercheurs A et B, p_A et p_B représentent respectivement l'article du chercheur A et l'article de B, le poids de la relation de coopération $W_{co}(u_A, u_B)$ entre A et B est calculé comme :

$$W_{co}(u_A, u_B) = \frac{|p_{AB}|}{\min\{|p_A|, |p_B|\}} \quad (5-14)$$

5.3.2.3 Relation d'événement en commun et évaluation de son poids de relation

La relation d'événement en commun indique la relation d'activité en commun comme quand les chercheurs participent conjointement à une conférence ou un séminaire, publient l'article dans un même journal. e_A et e_B représente respectivement l'événement où A et B participent, le poids de la relation d'événement en commun $W_a(u_A, u_B)$ entre A et B est calculé comme :

$$W_a(u_A, u_B) = \frac{|e_A \cap e_B|}{|e_A \cup e_B|} \quad (5-15)$$

5.3.2.4 Relation de chercheur et évaluation de son poids de relation

Selon les trois dimensions ci-dessus, le poids de la relation entre le chercheur A et le chercheur B peut être calculé comme :

$$W_a(u_A, u_B) = \beta_1 \times W_c(u_A, u_B) + \beta_2 \times W_{co}(u_A, u_B) + \beta_3 \times W_a(u_A, u_B) \quad (5-16)$$

$\beta_1, \beta_2, \beta_3$ sont respectivement les fonction poids des trois dimensions. Nous leur attribuons 0.4, 0.4, 0.2 selon le degré de leur importance.

5.3.3 Expérimentation et analyse

Les membres d'ACL (*Association for Computational Linguistics*) sont des chercheurs dans le domaine de la linguistique informatique, qui est un champ interdisciplinaire permettant de simuler une partie ou toutes des fonctions du langage par la machine. Les événements ACL sont notamment les conférences et les séminaires, et y compris également un petit nombre de journaux. Grâce à une source de données en ligne [118], nous avons sélectionné les données des papiers publiés sur les événements ACL entre 2008 et 2012, y compris les références bibliographiques (le nom, l'auteur, la source (événement) et l'année), les relations de citer et l'institution de l'auteur etc. L'ensemble de données statistiques comprend 19,424 articles, 11,524 auteurs, 13 conférences, 326 séminaires et trois journaux.

Les données ci-dessus sont la description d'un réseau de recherche dans le domaine de la linguistique informatique. Elles peuvent être utilisées pour extraire des relations de coopération entre auteurs, des relations de citer entre

auteurs et des relations de citer entre papiers. Et puis nous pouvons établir respectivement le réseau de coopération et le réseau de citer. Leurs caractéristiques sont indiquées au tableau 5-4. Nous pouvons trouver que c'est un réseau social hétérogène qui contient les relations variées, les nœuds différents et les informations d'attributs.

Nom de Classe	Attribut
Chercheur	nom, rédiger, coopérer, citer, collègue, appartenir à, participer à
Papier	numéro, nom, citer, être cité, auteur, année
Événement	nom, publier, chercheur

Tableau 5-4 : Concepts et relations d'ontologie de réseau de recherche

Nous utilisons la méthode de la section 5.3.1 pour construire le réseau social sémantique. Tout d'abord, nous extrayons les concepts et les relations comme indiqué dans le tableau 5-5 à partir des données ci-dessus. Et puis nous utilisons le logiciel Protégé pour construire l'ontologie du réseau de recherche ; cette ontologie se compose de trois classes principales : « auteur », « article » et « événement », où « conférence », « séminaire » et « journal » sont des sous-classes d'événement. Ensuite, nous utilisons une API d'ontologie Jena pour marquer la sémantique d'ontologie pour les données originales, ce qui nous permet d'obtenir les données de réseau social exprimées par le langage OWL. Et finalement, un réseau social sémantique correspondant au réseau de recherche est établi.

De même, nous utilisons la méthode de la section 5.3.2 pour calculer le poids de relation entre chercheurs, c'est-à-dire la découverte des relations de la communauté statique et dynamique et le calcul du poids de la relation entre les auteurs basé sur l'ontologie. Les relations sont divisés en trois dimensions : dimension de relation de citer, dimension de relation de coopération et dimension de relation d'événement en commun. Et le poids de la relation est calculé par la fusion des dimensions multiples. Nous analysons ensuite les résultats.

No.	Chercheur A	Chercheur B	Etre cité en commun	Citer mutuel	Citer en commun	Citer	Coopération	Participer en commun	Poids de relation
01	V. W. Zue	L. Hirschman	0.419	0.271	0.145	1.000	0.875	0.238	0.798
02	E. Reiter	A. Belz	0.417	0.206.	0.201	0.986	0.656	0.512	0.759
03	D. D. McDonald	J. Pustejovsky	0.327	0.165	0.056	0.637	1.000	0.109	0.677
04	B. M. Sundheim	R. Grishman	0.557	0.085	0.025	0.789	0.716	0.238	0.650
05	R. Johansson	A. Moschitti	0.125	0.130	0.223	0.549	0.829	0.296	0.610
06	M. A. Walker	D. J. Litman	0.297	0.123	0.183	0.708	0.583	0.352	0.587
07	S. Bangalore	A. K. Joshi	0.412	0.070	0.143	0.735	0.539	0.238	0.557
08	L. Hirschman	R. Grishman	0.268	0.092	0.077	0.497	0.677	0.389	0.547
09	B. Roark	M. Johnson	0.291	0.067	0.152	0.590	0.389	0.379	0.467
10	H. Alshawi	M. Rayner	0.211	0.066	0.075	0.390	0.404	0.247	0.367
11	T. Pedersen	R. Mihalcea	0.381	0.018	0.113	0.593	0.175	0.273	0.362
12	I. Mani	J. Pustejovsky	0.065	0.044	0.077	0.181	0.606	0.199	0.354
13	P. S. Jacobs	J. R. Hobbs	0.157	0.024	0.056	0.246	0.000	0.887	0.276
14	Nederhof, Mark-Jan	M. Johnson	0.359	0.023	0.082	0.531	0.000	0.278	0.268
15	P. S. Jacobs	R. M. Weischedel	0.157	0.046	0.040	0.252	0.000	0.819	0.265
16	J. Nivre	R. Johansson	0.302	0.081	0.085	0.537	0.000	0.223	0.259
17	E. Bill	R. C. Moore	0.174	0.012	0.064	0.262	0.210	0.346	0.258
18	R. M. Weischedel	J. R. Hobbs	0.058	0.016	0.078	0.137	0.000	1.000	0.255
19	P. Koehn	Wu Dekai	0.428	0.022	0.077	0.611	0.000	0.000	0.244
20	E. Brill	K. W. Church	0.190	0.013	0.126	0.361	0.000	0.469	0.238
21	E. Brill	R. M. Weischedel	0.208	0.031	0.056	0.317	0.000	0.473	0.221
22	A. Koller	C. Gardent	0.164	0.140	0.075	0.425	0.000	0.249	0.220
23	P. Koehn	R. C. Moore	0.296	0.011	0.072	0.424	0.000	0.247	0.219
24	L. Hirschman	P. S. Jacobs	0.100	0.006	0.000	0.080	0.000	0.920	0.216

Tableau 5-5 : Information des chercheurs qui ont le poids de relation le plus grand

Les 50 premiers auteurs qui ont publié le plus des papiers ont été sélectionnés et nous calculons respectivement le poids de la relation entre eux ; 1200 paires des valeurs du poids de la relation entre les chercheurs ont été obtenues. Nous normalisons toutes les dimensions et le tableau 5-5 montre la partie des valeurs supérieures des mesures des 5 types de relation des 3 dimensions et les poids des relations globales.

Les valeurs du poids de la relation dans chaque dimension de certains chercheurs sont tous relativement élevées, et donc les poids des relations après fusion sont également élevés. Par exemple, pour les deux chercheurs V. W. Zue et L. Hirschman qui ont le poids de relation le plus grand, ils ont cité en commun 12 papiers, 13 papiers cités en commun, 10 « citer mutuellement », 3 coopérations et 2 « participer à un événement » conjointement. Le poids de la relation de citer entre ces deux chercheurs est le plus fort, le poids de relation de coopération est le 2^{ième}, le poids de relation d'événement en commun est le 23^{ième} ; donc le poids de relation global est le premier.

Certains chercheurs ont des relations assez faibles, et d'autres relations plus fortes, et ils possèdent donc également un grand poids de relation. Cela aide à ne pas manquer des relations qui peuvent être ignorées lors de la mesure ordinaire de la relation unique. Par exemple, P. S. Jacobs et R. J. Hobbs, bien qu'il n'existe pas de relation de coopération, mais grâce à la forte relation de participation en commun, leurs classement du poids de relation est le 13.

Chercheurs centraux	Chercheurs en connexion
E. Brill	R. C. Moore, K. W. Church
R. M. Weischedel	E. Brill, K. W. Church, P. S. Jacobs, J. R. Hobbs
L. Hirschman	V. W. Zue, R. Grishman, P. S. Jacobs
P. S. Jacobs	J. R. Hobbs, R. M. Weischedel, L. Hirschman

Tableau 5-6 : Chercheurs centraux

Nous pouvons observer par le tableau 5-5 que, bien que les chercheurs dans le domaine de la linguistique informatique viennent de différentes régions du monde, il existe une forte corrélation entre eux et une structure de réseau est formée. Le tableau 5-6 montre les chercheurs centraux qui possèdent au moins

3 chercheurs importants avec un poids de relation élevé. Les 50 chercheurs que nous choisissons représentent les études les plus élevés de la linguistique informatique pour une large part, et les chercheurs centraux promeuvent le partage et l'échange de connaissances pour ce domaine grâce à la connexion avec d'autres chercheurs.

De plus, d'un point de vue de communauté dynamique, la linguistique informatique est un domaine énorme, et une grande différence existe entre les différents intérêts de recherche des chercheurs. Nous récupérons les intérêts de recherche des chercheurs depuis leurs sites web personnels et complétons par le service d'*Arnetminer* [119]. Le tableau 5-7 montre les intérêts de recherche de 5 paires de chercheurs qui ont le poids de relation le plus élevé. Les systèmes de recommandation de filtrage collaboratif sont fondés sur le principe selon lequel deux utilisateurs auront plus d'intérêts en communs s'ils partagent déjà certains intérêts [120]. Nous pouvons trouver une familiarité élevée entre les intérêts des chercheurs qui ont le poids de relation élevé. Cela confirme que notre modèle du poids de la relation basé sur les ontologies pour le réseau social de recherche peut révéler efficacement le degré de familiarité entre les utilisateurs, et relève le poids de la relation entre eux.

No.	Chercheur	Intérêt de recherche
01	V. W. Zue	<i>Spoken Language Recognition, medical informatics, education informatics</i>
01	L. Hirschman	Language Recognition, Natural Language Processing, biomedical informatics
02	E. Reiter	natural-language generation, data-to-text, multimodal systems, Automatic Generation
02	A. Belz	Natural Language Generation, data-to-text generation, Automatic generation, Formal models of natural language
03	D. D. McDonald	natural language generation, lexical inference, spatial description

03	J. Pustejovsky	natural language processing, Lexical Semantics, Generative Lexicon Theory, Qualitative Temporal and Spatial Reasoning
04	B. M. Sundheim	Natural Language & Speech Processing, Human-Computer Interaction, Performance Evaluation (information extraction system evaluation)
04	R. Grishman	Natural Language Processing, Information Extraction
05	R. Johansson	Semantic Role Labeling, Online Learning, Automatic Annotation
05	A. Moschitti	Semantic Role Labeling, Online Learning, Text Categorization, Semantic Modeling

Tableau 5-7 : Intérêt de recherche des chercheurs

Dans ce chapitre, afin de prendre en compte l'hétérogénéité des réseaux sociaux de la vie réelle, deux modèles de contexte social basés respectivement sur les graphes et les ontologies sont proposés. Nous utilisons les données réelles recueillies sur les réseautages sociaux en ligne pour conduire nos expérimentations et analysons les résultats en vérifiant l'efficacité de ces modèles.

Chapitre 6. Application basée sur le contexte social pour les réseaux opportunistes

Une application [121] basée sur la confiance pour améliorer la performance et la sécurité dans la stratégie de transmission dans les réseaux opportunistes est proposée dans ce chapitre. Nous présentons tout d'abord l'utilisation de la confiance dans le domaine du réseau opportuniste. Ensuite, nous exploitons des contextes sociaux basés sur la familiarité et la similitude en utilisant les deux modèles sur le poids des relations, tel que décrit auparavant, mais d'une part pour adapter la caractéristique d'évolution des aspects de contextes sociaux des nœuds dans le réseau opportuniste, et d'autre part en raison du choix restreint en informations contextuelles, et pour inférer plus exactement la relation des utilisateurs selon les activités des nœuds, nous proposons séparément deux algorithmes : un algorithme de transmission de confiance basé sur la confiance sociale binaire et un algorithme de transmission de confiance basé sur la similitude de communauté dynamique des nœuds. La description spécifique de chaque algorithme est développée dans la suite de ce chapitre.

6.1 Utilisation de la confiance dans le domaine des réseaux opportunistes

6.1.1 Définition de la confiance pour les réseaux opportunistes

La confiance est un concept multidimensionnel complexe difficile à définir. Comme une tendance ou une volonté subjective humaine, avec le développement de la société, la confiance a évolué et est devenue un attribut naturel de notre société. Dans le domaine de la recherche informatique, l'étude de la confiance est dans sa phase initiale ; de nombreux chercheurs l'utilisent en introduisant des conceptions de la confiance issues des domaines de la sociologie, de la psychologie, et des sciences politiques, combinée à l'analyse quantitative de modèles mathématiques, sous des angles et des aspects différents.

En sociologie, O Donovan a défini la confiance de la manière suivante : Des phénomènes d'identité mutuelle et d'interaction sincère pendant la mise en œuvre des comportements sociaux humains [122].

Hoffman, qui est dans le domaine de la psychologie a décrit comme : Une bonne prévision certaine des comportements futurs incertains d'autres [123].

Et plus, un chercheur en économie, Sable est défini comme : Une personne est convaincue que l'autre ne sera pas tiré parti de ses faiblesses pour obtenir le bénéfice [124].

Dans le domaine du réseau opportuniste, des nœuds sont influencés par des comportements sociaux humains, et intègrent certains attributs des sociétés humaines. Pour un nœud dans le réseau opportuniste, la confiance est le niveau de crédibilité quantifiée du nœud, évalué par son comportement futur, sa performance et sa fiabilité, etc. La confiance peut être aussi comprise comme une relation de dépendance entre les nœuds, une anticipation subjective des nœuds pour la tendance des comportements à venir des autres nœuds sur la base de leurs propres observations et informations fournies par les nœuds voisins.

En général, les notions de confiance et de réputation sont deux concepts susceptibles d'être confondus. Il existe déjà beaucoup de travaux sur la réputation dans le domaine des réseaux opportunistes. Afin de permettre au lecteur de comprendre la différence essentielle entre les deux, nous la présenterons dans cette partie. La confiance et la réputation sont deux états différents ; la confiance est un processus d'évaluation de la crédibilité d'un nœud par un autre nœud; la réputation est le résultat de cette évaluation. La première notion est active, par contre la deuxième est passive et il faut un processus d'accumulation. La relation entre les deux est illustrée dans le Figure 6-1.

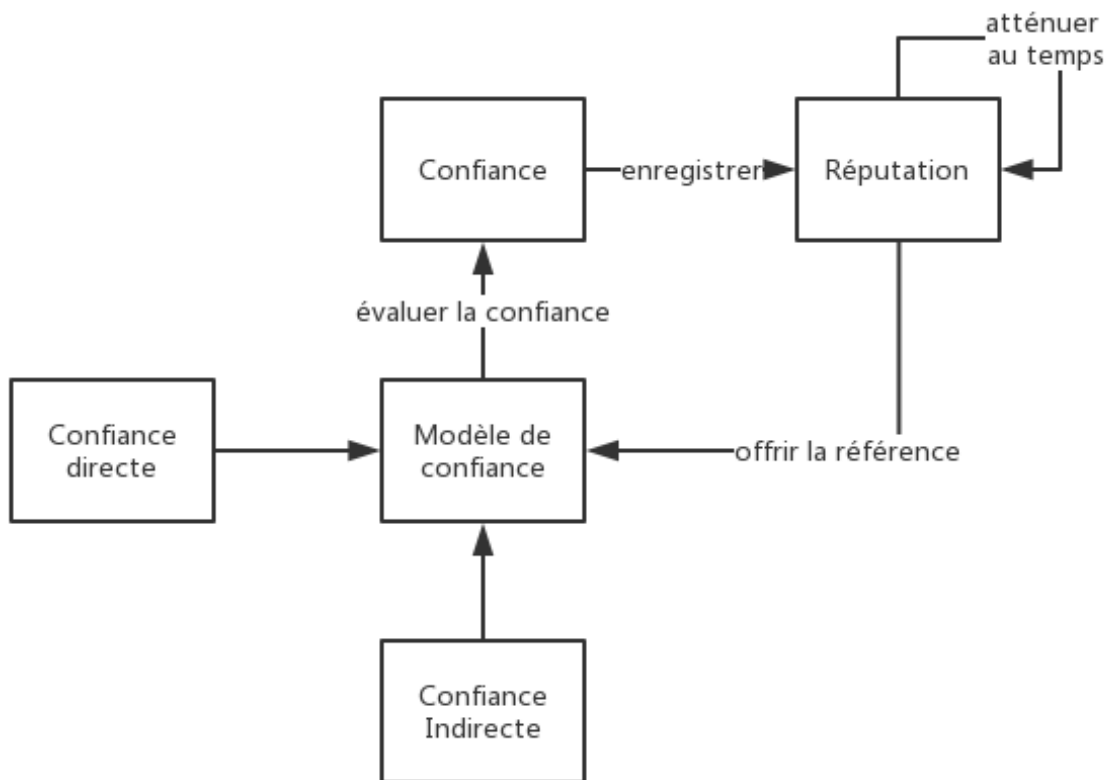


Figure 6-1 Relation entre confiance et réputation

6.1.2 Gestion de confiance et modèle de confiance

Le modèle de confiance donne la méthode pour calculer la valeur de la confiance de l'entité par mesure de la relation de confiance entre les entités. La gestion de confiance est principalement présente pour compléter la collecte de la confiance, la classification de la confiance, la décision stratégique et le stockage de la confiance, et la maintenance de la valeur de confiance de l'entité. Il s'agit de fournir des services de gestion de l'information dans le processus d'évaluation du modèle de confiance.

Le concept initial de la gestion de confiance est proposé par Blaze et al. [125] en 1996, et se réfère à "l'adoption d'une approche unifiée pour décrire et expliquer la stratégie de sécurité, les informations d'identification de sécurité et la relation de confiance critique pour des opérations d'autorisation directe." Avec la recherche et le développement de la gestion de confiance, le concept

s'enrichit progressivement, et le concept de gestion de confiance est maintenant défini comme suit: "Des comportements de collection, de codage, d'analyse et de représentation de la relation de confiance dans l'application du réseau avec pour but l'évaluation et la prise de décision [126]".

La gestion de confiance dans le réseau opportuniste intègre de nombreux paramètres du concept ci-dessus, mais le réseau opportuniste se différencie des réseaux traditionnels passés car la topologie du réseau est modifiée régulièrement, et la stratégie de la maintenance de la confiance des nœuds par des équipements tiers ne peut pas s'adapter à la gestion de confiance du réseau opportuniste.

Selon les caractéristiques du réseau opportuniste, la gestion de confiance du réseau opportuniste est construite sur la stratégie de gestion de confiance dynamique. Le concept de la gestion dynamique de confiance est résumé ainsi: Lorsque la gestion de confiance pour l'entité d'application d'internet, considérant pleinement la sécurité, l'intégrité et la fiabilité, et d'autres raisons qui influencent la relation de confiance, en se fondant sur la confiance concrète métrique par représentation numérisée, collectionnant dynamiquement les facteurs subjectifs et objectifs pour l'établissement de la relation de confiance entre l'entité, le processus des ajustements en même temps en réalisant l'évaluation, la prise de décision et le stockage de confiance [127]. Dans ce concept, un certain nombre de facteurs influencent l'établissement de la confiance. Dans le processus de collection de confiance, la collection d'information d'évaluation de confiance non seulement prend en compte l'information subjective, mais aussi l'information d'évaluation objective. Dans le même temps, en insistant sur la décision et le stockage de la confiance, des ajustements appropriés doivent être fait fondée sur l'environnement d'évaluation. Par conséquent, par rapport à la gestion traditionnelle de confiance, la gestion dynamique peut être mieux adaptée au réseau opportuniste, et peut résoudre efficacement les différents problèmes dans le processus de gestion de confiance, afin de fournir la décision raisonnable pour construire le modèle de confiance.

La gestion dynamique de confiance explique précisément la raison pour laquelle l'information contextuelle est importante pour la gestion de confiance

dans le réseau opportuniste, et notamment l'information liée aux contextes sociaux, en tenant compte des caractéristiques du réseau opportuniste. Nous présentons cet aspect dans la partie suivante de ce chapitre.

La stratégie de gestion dynamique de confiance exploite le réseau opportuniste car elle peut soutenir les caractéristiques de dynamique et de multi-sources pour l'évaluation de la confiance. Bien que différente de la gestion traditionnelle de la confiance il n'existe pas de meilleures stratégies, mais le modèle de confiance dynamique renforce la caractéristique dynamique d'établissement de la confiance, qui peut être mieux appliquée au réseau opportuniste.

La confiance peut aider à mesurer la fiabilité et la sécurité qui influencent les paramètres de qualité de service des autres services de réseau. Par exemple, [128] traite de l'algorithme d'allocation de ressources dynamique optimisé pour le réseau backbone de Cloud. En raison des exigences imprévisibles et variables des utilisateurs, un schéma d'allocation de ressources flexible et intelligent est nécessaire. L'article aborde le défi fondamental posé par la question de l'allocation efficace des ressources dans le réseau backbone de cloud. L'objectif ultime est de satisfaire les besoins des utilisateurs du nuage tout en maximisant les revenus du fournisseur de services cloud. La proposition coordonne les nœuds virtuels et les étapes de mappage de liens virtuels pour optimiser l'utilisation des ressources. Si la confiance entre les utilisateurs et fournisseurs peut être considéré, nous pouvons avoir l'effet plus intéressant.

Selon les différentes représentations de la relation de confiance, celle-ci peut être distinguée en confiance rationnelle et confiance émotionnelle. La confiance rationnelle signifie que l'entité croit qu'une autre entité va accomplir certains actes dans certaines circonstances ; la relation de confiance de la confiance rationnelle ne possède que deux types de relation, la confiance ou la méfiance, qui peuvent être représentés par le nombre 0 ou 1, elle est couramment utilisée dans l'e-mail, l'e-commerce et l'opération en ligne etc.

Nous utilisons la confiance émotionnelle pour mesurer la relation sociale des nœuds du réseau opportuniste en profitant des 3 catégories de contexte social: contexte social binaire, contexte social des communautés statiques et contexte

social des communautés dynamiques présentées dans la section 3.3. La confiance émotionnelle est plus semblable à la notion de confiance présentée ci-dessus, et qui se réfère à l'analyse du niveau de confiance de l'autre par l'interaction historique entre les entités, les informations d'évaluation et l'état de fonctionnement etc., et l'amélioration de la fidélité qui découle de l'interaction entre les entités. Souvent il y a des entités qui mènent leur propre évaluation de la confiance, avec une forte subjectivité. La plus grande différence avec la confiance rationnelle est que la confiance émotionnelle n'est plus absolue, grâce à l'algorithme d'évaluation spécifique, la confiance est représentée par un certain continuum de valeur, puis divisé en différents niveaux de confiance.

6.2 Algorithme de transmission basé sur la confiance sociale binaire

Nous présentons d'abord les algorithmes de transmission souvent utilisés dans les réseaux opportunistes : *Epidemic* [129], *Spray and Wait* [130], *PROPHET* [131].

Epidemic et *Spray and Wait* sont des algorithmes de transmission basés sur la reproduction. *Epidemic* stocke toutes les données reçues par tous les nœuds dans la mémoire cache, et si des nœuds se rencontrent, ils vérifient mutuellement l'information manquante dans leurs propre mémoire cache et envoient les données manquantes à l'autre [132]. En adoptant un tel mécanisme de transmission de *flooding*, si la mémoire cache du nœud est assez grande en théorie, après une certaine période de temps, chaque nœud du réseau va recevoir toutes les données dans le réseau, ce qui améliore considérablement le taux de succès de transmission de données et réduit le temps de délai de transmission. Mais le défaut d'*Epidemic* est qu'un grand nombre de copies des données sont générées dans le réseau, ce qui consomme beaucoup de ressources et affecte les performances.

Spray and Wait est un algorithme de *flooding* limité pour résoudre le défaut d'*Epidemic*. Sur le modèle d'*Epidemic*, le processus de transmission de données est divisé en phase de *Spray* et phase de *Wait*, pour contrôler le nombre de copies de données dans le réseau. Dans la phase du *Spray*, le nœud source

envoi des copies données à un nombre limité de nœuds rencontrés. Dans la phase de *Wait*, le nœud de transmission ne copie plus les données pour envoyer à l'autre nœud et attend la rencontre avec le nœud de destination pour les transférer directement. Cet algorithme non seulement surmonte le problème de nombre excessif des copies de données d'*Epidemic*, et améliore efficacement les performances du réseau, mais présente également une excellente efficacité de transmission et permet une grande évolutivité.

Prophet est un algorithme de transmission basé sur la prévision probabiliste. L'algorithme prévoit la probabilité de transmission au nœud de destination grâce à l'historique des interactions enregistré par les nœuds. Lorsqu'il rencontre d'autres nœuds, il décide si le nœud peut être un nœud de transmission en utilisant la probabilité prédictive, et réalise sélectivement la reproduction de message ; de cette manière, il contrôle le numéro des copies et améliore les performances du réseau [133]. D'une manière générale, si le nœud i , qui transporte un message, rencontre le nœud j , la formule de prédiction de probabilité entre les nœuds est donnée par les formules (6-1) et (6-2). Plus il y a de rencontres entre les nœuds, plus grande est la valeur de la probabilité.

$$P_{(i,j)} = P_{(i,j)old} + (1 - P_{(i,j)old}) \cdot P_{init} \quad (6-1)$$

$$P_{(i,j)} = P_{(i,j)old} \cdot \gamma^k \quad (6-2)$$

Etant donné que des nœuds du réseau opportuniste sont influencés par des comportements sociaux humains, on peut dire que, d'une certaine façon, ils intègrent certains attributs des sociétés humaines. Mais le mode de fonctionnement de chaque nœud reste conforme à celui du réseau opportuniste, donc ces nœuds possèdent également toutes les caractéristiques du nœud de réseau opportuniste. Les contextes sociaux des personnes de différents niveaux et types peuvent donc également être utilisés dans des applications du réseau opportuniste.

Dans le processus d'interaction sociale humaine, quand un membre doit communiquer, partager ou se confier à un autre membre étranger, il ou elle va

d'abord déterminer si la personne est digne de confiance, et alors seulement l'interaction et l'échange d'information peuvent être réalisées. Généralement la personne est jugée avec nos propres critères d'évaluation au début, et puis confirmée synthétiquement grâce aux recommandations des membres également crédibles.

Une bonne relation de confiance est importante pour l'échange d'information d'autres aspects. La base de l'établissement de la relation interpersonnelle est la relation de confiance mutuelle, la confiance sociale binaire est une sorte de phénomène de reconnaissance mutuelle et d'interaction sincère entre les gens généré par des comportements sociaux [134]. Par conséquent, compte tenu de leurs caractéristiques, la confiance du réseau social peut être introduite dans le réseau opportuniste, afin d'établir une relation de confiance entre les nœuds du réseau opportuniste, pour améliorer l'influence de sa performance en dépit des nœuds égoïstes.

Afin d'éviter la transmission des données aux nœuds égoïstes dans le réseau opportuniste, et en même temps de surmonter des problèmes d'algorithme de transmission de confiance lorsqu'il existe nœuds égoïstes nombreux dans le réseau, tel que le délai de transmission élevé et l'évaluation inexacte de la confiance, etc. Basé sur l'algorithme de transmission d'infection *Epidemic*, nous proposons un algorithme de confiance sociale : *ST-Epidemic (Social Trust Epidemic)*. En utilisant l'efficacité de la transmission d'*Epidemic*, nous décrivons trois attributs sociaux : degré de familiarité, niveau d'intimité et qualité de service, basé sur des contextes sociaux. Par l'établissement d'une relation de confiance entre les nœuds, et la quantification par des valeurs spécifiques de la confiance, le nœud de confiance peut être finalement sélectionné comme le nœud de transmission, pour éviter la transmission aux nœuds égoïstes et assurer le fonctionnement normal du réseau.

Dans cette partie nous décrirons d'abord la base théorique et l'idée de construction de l'algorithme de confiance sociale *ST-Epidemic*. En introduisant la confiance sociale binaire du réseau social, notre algorithme décrit les attributs des contextes sociaux des nœuds en utilisant les paramètres de base d'interaction des nœuds, et évalue le niveau de confiance entre les nœuds par la construction de la relation sociale des nœuds en profitant des contextes

sociaux. Puis nous présenterons la manière d'éviter les nœuds égoïstes dans le processus de transmission de données, illustré par un diagramme de flux. Enfin, une composante clé du code de conception de l'algorithme est donnée, et la complexité de l'algorithme et la complexité de l'espace sont analysées.

6.2.1 Degré de familiarité

Dans la société humaine, on a tendance à associer fréquence des rencontres avec les mêmes personnes et degré de familiarité élevé. Nous pouvons donc considérer qu'un certain niveau de la relation de confiance est établi en même temps. La stratégie de gestion des nœuds du réseau opportuniste est adoptée sous le principe de l'autogestion, bien que le statut du nœud soit le même dans le réseau, mais les comportements sociaux de chaque nœud sont variés, de sorte que la fréquence de contact des nœuds varie aussi. Ceci constitue une utilisation du contexte social binaire présenté dans la section 3.3.

Ainsi, dans le processus de transmission des données du nœud, nous pouvons choisir le nœud le plus familier comme nœud de transmission, le degré de familiarité du nœud étant défini de la manière suivante:

$$fam_{i,j} = \frac{conn_{i,j}}{\sum_{k=0}^n conn_{i,k}} \quad (6-3)$$

$fam_{i,j}$ signifie le degré de familiarité du nœud j à nœud i , $fam_{i,j} \in [0,1]$, $conn_{i,j}$ signifie le nombre de fois où il y a eu échange entre le nœud i et le nœud j , $\sum_{k=0}^n conn_{i,k}$ signifiant la somme des contacts du nœud i avec tous les nœuds. Comme les nœuds contactés sont différents pour chaque nœud, le nombre de contacts est différent également, de sorte que le degré de familiarité du nœud j à nœud i $fam_{i,j}$ et le degré de familiarité du nœud i à nœud j $fam_{j,i}$ ne sont pas nécessairement les mêmes. Quand nous évaluons le degré de familiarité des nœuds, ce processus d'évaluation est donc unidirectionnel.

6.2.2 Niveau d'intimité

Que ce soit du fait des nécessités inhérentes aux activités professionnelles, scolaires, au déroulement de la vie privée les individus sont amenés à partager

des intérêts communs sans être forcément rassemblés dans les mêmes lieux. Dans ce contexte, plus la proximité (qu'elle soit physique ou virtuelle) est importante plus l'intimité entre les individus est supposée grande. Nous pouvons même les considérer comme faisant partie de la même communauté. De plus, dans ce cas-là, nous pouvons les considérer comme les contextes sociaux des communautés statiques présentées dans la section 3.3. Dans le processus de mouvement des nœuds de réseau opportuniste, un phénomène similaire aux cercles d'amis se produit également, si différents nœuds à portée de communication l'un et l'autre maintiennent l'état de la communication pendant une longue période, ils sont alors considérés comme partageant un haut niveau d'intimité.

En outre, du point de vue du réseau opportuniste, il est possible qu'il existe plusieurs nœuds qui ont le même niveau d'intimité. Lorsque cette situation apparaît, le niveau d'intimité ne peut pas mesurer complètement la confiance sociale binaire des nœuds. Il y a deux raisons principales à cela : d'une part, dans le scénario réel, des nœuds égoïstes se déplacent constamment, et peuvent avoir une fréquence de contact élevée avec un nœud normal, c'est-à-dire que des nœuds égoïstes peuvent donc abuser de la confiance du nœud normal. D'autre part, dans le processus de transmission de données, il faut assurer un temps suffisamment long pour transmettre les données au nœud suivant. Seulement quand le temps de connexion entre les nœuds est plus long, l'intégrité de la transmission de données ne peut être garantie. Ainsi, selon la théorie du cercle des connaissances du réseau social [135], sur la base de la fréquence des contacts des nœuds, en tenant compte en plus du temps de connexion entre les nœuds, nous pouvons décrire plus précisément la confiance sociale binaire du nœud. Donc le niveau d'intimité est présenté dans la formule :

$$inti_{i,j} = \frac{connecttime_{i,j}}{\sum_{k=0}^n connecttime_{i,k}} \quad (6-4)$$

$inti_{i,j}$ signifie le niveau d'intimité de nœud j à nœud i , et $inti_{i,j} \in [0,1]$, $connecttime_{i,j}$ signifie le temps de connexion du nœud i et nœud j , $\sum_{k=0}^n connecttime_{i,k}$ signifie le somme du temps de connexion du nœud i avec

tous les nœuds. En raison du fait que les états de fonctionnement de chaque nœud sont différents, le temps de connexion marque la différence, de sorte que le niveau d'intimité de nœud j à nœud i $inti_{i,j}$ et le degré de familiarité de nœud i à nœud j $inti_{j,i}$ ne sont pas nécessairement les mêmes. Quand nous évaluons le niveau d'intimité des nœuds, ce processus d'évaluation est également unidirectionnel.

6.2.3 Qualité de service

D'après notre expérience de la vie quotidienne [124], la fréquence d'interaction entre les gens et la longueur de temps d'activité conjointe n'est pas suffisant pour une description complète d'une bonne relation de confiance. Seulement en fournissant des services entre eux, et en estimant les services apportés, nous pouvons bien discerner l'intention et la nature de chacun.

Étant donné que le nœud normal du réseau peut fournir un service de transmission en temps voulu pour d'autres nœuds, les nœuds égoïstes profitent seulement des services fournis par les autres sans rien donner en échange. Tirant profit de ce fait, nous intégrons la qualité de service dans le chaînon d'évaluation de la confiance. Grâce à l'échange du nombre de fois où il y a eu réception et transmission de données enregistré par chaque nœud, la qualité de service peut être mesurée. La fiabilité du service de transmission du nœud est reflétée par la qualité de service, et en même temps pour distinguer plus efficacement les nœuds égoïstes et nœuds normaux. La qualité de service du nœud est définie comme suit :

$$serv_{i,j} = \frac{tran_{i,j}}{rece_{i,j}} \quad (6-5)$$

$serv_{i,j}$ signifie la qualité de service du nœud j à nœud i, et $serv_{i,j} \in [0,1]$, $tran_{i,j}$ signifie le nombre de fois où il y a eu transmission de données du nœud i aidé par le nœud j, $rece_{i,j}$ signifie le nombre de fois où il y a eu réception de données du nœud j par le nœud i ; ce nombre de $rece_{i,j}$ ici est inévitable supérieur ou égal à $tran_{i,j}$.

6.2.4 Calcul de la confiance sociale binaire

La constitution de la confiance sociale binaire basée sur les contextes sociaux est décrite sous trois angles : le degré de familiarité, le niveau d'intimité et la qualité de service. La base théorique de chaque propriété est également présentée. Comme la confiance sociale binaire est fonction des contextes sociaux d'attributs différents, le calcul de la confiance sociale binaire est une somme pondérée, telle que représentée par la formule (6-6)

$$T_{i,j} = \alpha \cdot fam_{i,j} + \beta \cdot inti_{i,j} + \gamma \cdot serv_{i,j} \quad (6-6)$$

$T_{i,j}$ signifie la confiance sociale binaire du nœuds j à nœud i , α, β, γ signifie respectivement les poids du degré de familiarité, du niveau d'intimité et de la qualité de service ci-dessus dans la confiance sociale binaire, $\alpha, \beta, \gamma \in (0,1]$ et $\alpha + \beta + \gamma = 1$.

En tenant compte que la confiance sociale binaire du nœud est décrite du point de vue des trois attributs de contextes sociaux différents, qui construit la base de la confiance, la distribution des poids est comme (6-7) et (6-8),

$$\alpha = \frac{fam_{i,j}}{fam_{i,j} + inti_{i,j}} \cdot (1 - \gamma) \quad (6-7)$$

$$\beta = \frac{inti_{i,j}}{fam_{i,j} + inti_{i,j}} \cdot (1 - \gamma) \quad (6-8)$$

Pour refléter la subjectivité d'évaluation de la confiance, dans le processus de détermination des trois poids, γ est établi dès le début, dans la mesure où dans ces trois attributs de contextes sociaux, la qualité de service constitue la plus grande part de la confiance, qui non seulement décrit l'attribut social du nœud, mais aussi distingue progressivement les nœuds égoïstes des nœuds normaux. Ensuite, la fréquence des contacts et le temps de connexion entre les nœuds varient, les nœuds dans les différents environnements de réseau mettent l'accent sur différents propriétés des contextes sociaux et calculent la confiance des autres nœuds. Par conséquent, nous mettons l'accent sur le degré de familiarité et le niveau d'intimité pour déterminer α et β .

6.2.5 Le processus de transmission basé sur la confiance sociale binaire

Le processus de routage des données des réseaux opportunistes est la réalisation de chaque étape du processus de transmission coordonné par des nœuds. Nous proposons le calcul de la confiance sociale binaire comme ci-dessus. Nous calculons ici la confiance du nœud j par le nœud i $T_{i,j}$, et si $T_{i,j} \geq \theta$, le nœud est considéré comme nœud de confiance et la transmission de données vers ce nœud peut commencer; si $T_{i,j} < \theta$, le nœud est considéré comme nœud égoïste et la transmission de données est refusée. θ est donc la valeur de seuil de la confiance.

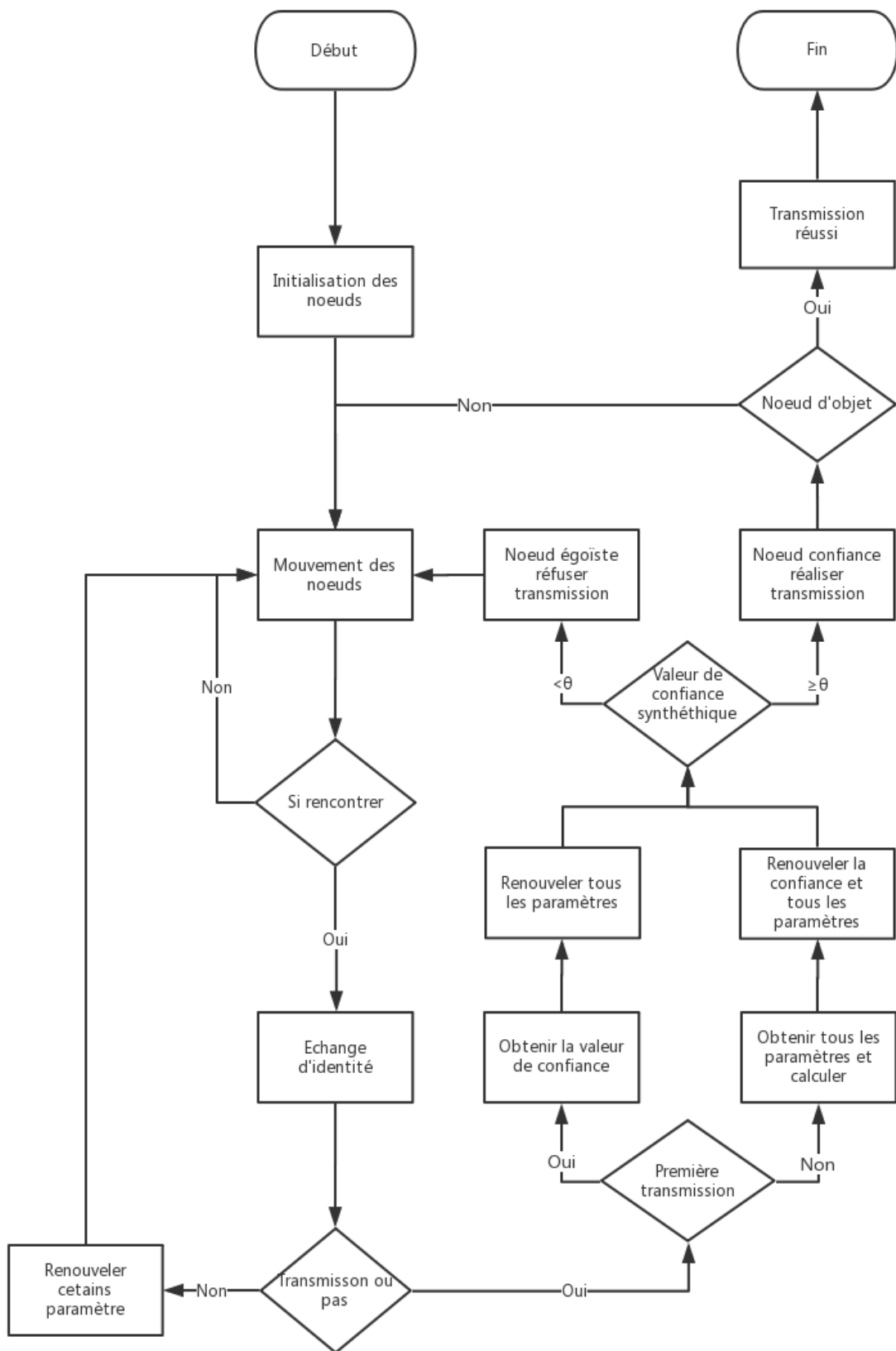


Figure 6-2 : Processus de transmission de ST-Epidemic

Selon les règles de transmission ci-dessus, *ST-Epidemic* réalise la transmission de données pour juger de la valeur de la confiance sociale binaire du nœud. Lorsque le nœud i transmet les données au nœud j pour la première fois, la transmission est déterminée selon la valeur de confiance initiale du nœud. Si la transmission n'est pas la première, la confiance sociale binaire du nœud j à nœud i $T_{i,j}$ doit être calculée par $fam_{i,j}$, $inti_{i,j}$, $serv_{i,j}$, si $T_{i,j}$ atteint la valeur de seuil de transmission de confiance, il transmet les données au nœud j , sinon, il refuse la transmission. Le processus est illustré à la figure 6-2.

Selon les règles et la procédure de transmission des données de *ST-Epidemic*, l'algorithme de transmission basée sur la confiance sociale est représenté sur l'algorithme.

```

Input :  $fam_{i,j}$   $inti_{i,j}$   $serv_{i,j}$   $T_{i,j} \leftarrow \theta$   $tran_{i,j} \leftarrow 0$ 
Output : ensemble des nœuds de confiance TR[j]

1:  While 1 do
2:    while  $j$  is in contact with  $i$  do
3:      If  $tran_{i,j} = 1$   $j = destination(m)$  then
4:         $Forward(j, m)$ 
5:         $update(i)$ 
6:      else
7:         $T_{i,j} = \alpha \cdot fam_{i,j} + \beta \cdot inti_{i,j} + \gamma \cdot serv_{i,j}$ 
8:        if  $T_{i,j} \geq \theta$  then
9:          while  $\exists m \in buffer(i)$  do
10:             $Forward(j, m)$ 
11:             $update(i)$ 
12:          else
13:             $update(i)$ 
14:          end if

```

L'initialisation de la valeur de confiance θ du nœud dans l'algorithme au-dessus a pour but d'encourager l'établissement plus rapide de la relation de confiance entre les nœuds, et de fournir une meilleure base de confiance pour la transmission des données par la suite. Ainsi, la première fois que les données sont transmises à un nœud, si le nœud est un nœud normal, il continuera à transmettre des données à un autre nœud ; si le nœud est un nœud égoïste, il ne transmettra aucune des données vers d'autres nœuds. Par cette approche exploratoire, une petite partie de l'ensemble des paquets de données peut être perdue, mais elle aide à distinguer plus rapidement les nœuds normaux et les nœuds égoïstes, et donc l'établissement de la valeur de confiance initiale θ est nécessaire.

Nous analysons la complexité de l'algorithme, qui dépend du nombre de nœuds se rencontrant, du nombre de paquets à transmettre et du seuil de confiance de transmission θ . Dans le scénario, le numéro des nœuds est n , le nombre maximum de paquets à transmettre est m , et le seuil de confiance de transmission $\theta \in [0,1]$. Si θ est plus élevé, la demande de confiance des nœuds sera augmentée, de sorte qu'il y aura une baisse du nombre de transmission de données, une augmentation du délai de transmission et une réduction de l'efficacité de l'algorithme. Par conséquent, le cas le plus extrême de l'algorithme considère un nœud qui rencontre $n-1$ nœuds et la confiance de tous ces nœuds rencontrés satisfait le seuil de confiance θ . Dans ce cas la complexité de calcul de l'algorithme est $O(nm)$. A l'opposé, le cas le plus favorable est que tous les nœuds rencontrés ne satisfassent pas le seuil de confiance θ , la complexité de calcul d'algorithme serait alors $O(n)$. En tenant compte de la situation réelle dans le réseau, la complexité de calcul d'algorithme est probablement comprise entre $O(nm)$ et $O(n)$.

6.3 Algorithme de transmission de confiance basé sur la similitude de communauté dynamique des nœuds

Dans les réseaux opportunistes sociaux, la mobilité et la trajectoire du dispositif du nœud sont complètement contrôlées par la personne. L'interaction sociale quotidienne des membres de la société est parfois simultanée, comme par exemple des étudiants rassemblés dans une même salle de cours, les gens qui

vont au cinéma pendant leurs loisirs, les consommateurs qui vont au marché pendant la semaine, etc. Selon la nature et le type d'occasions et/ou d'activité, les différents membres de la société ont convergé et intégré différents groupes ou communautés, ce qui fait émerger un phénomène de "petit monde" [136] des nœuds du réseau opportuniste social, et réduit la distance entre les membres de la société dans une certaine mesure. Donc dans cette partie nous profitons du contexte social des communautés dynamiques présenté dans la section 3.3.

En combinant la réalité de l'environnement social et les expériences de vie des membres de la société, les membres sociaux qui regroupent leurs activités souvent dans la même région, physique ou virtuelle, peuvent avoir plus de possibilités d'interaction, et peuvent établir progressivement une relation plus familière. Dans ce processus, grâce à la trajectoire de déplacement et à la situation sociale similaire des membres d'une « communauté », la définition de certains niveaux de confiance entre les uns et les autres prend naissance. Ainsi, les nœuds du réseau opportuniste social, en imitant le processus d'établissement des relations décrit ci-dessus, peuvent aussi réaliser la construction de relation de confiance selon la similitude de trajectoire des nœuds et leur similitude de relation sociale.

En analysant la relation entre les membres de la société qui ont la même trajectoire dans la même zone au sein d'un réseau social, ainsi que les relations des membres dans différentes zones, on remarque que la relation entre les membres de la société est étroitement liée à leurs activités et la diversité de leur choix. Souvent les membres peuvent avoir de nombreuses possibilités d'interagir sur différentes activités sociales dans une même zone, et donc établissent progressivement une relation plus intime. Dans ce processus, principalement en raison du fait que la portée des activités entre les membres se chevauche souvent, et que les différents membres de la société dans une même zone d'activité connaissent beaucoup de membres qui leur sont similaires, l'établissement de la confiance entre eux est plus facile. Par conséquent, nous proposons un algorithme de transmission de la confiance basé sur la similitude de communauté dynamique des nœuds pour les réseaux opportunistes. Selon la méthode de division des mailles du réseau que l'on

peut trouver dans la littérature [137], notre algorithme divise tout d’abord le scénario du réseau à des mailles de même surface, et tous étiquetés de manière unique. Dans la mémoire de tous les nœuds dans le réseau, une même table de division du scénario est posée. Nous marquons la trajectoire des nœuds par 0 et 1, et nous pouvons ainsi comparer la similitude de trajectoire entre les nœuds. Dans le même temps, les nœuds d’interaction dans chaque zone pour un nœud sont enregistrés en utilisant la liste d’adjacence, pour analyser la situation sociale des nœuds. Ensuite, nous déterminons le niveau de confiance en synthétisant la similitude de trajectoire des nœuds et leur similitude de relation sociale. Enfin, lors du choix du nœud de confiance, et en exploitant la connaissance que l’on a de la façon dont le consommateur conduit ses achats et de sa psychologie, nous choisissons autant que possible un nœud idéal de confiance qui a une crédibilité plus stable, afin d’assurer d’une manière plus poussée la fiabilité des données du réseau.

6.3.1 Similitude de trajectoire

Dans le réseau opportuniste social, la vitesse, la direction et la trajectoire etc. des nœuds sont tous contrôlés par les membres de la société. En même temps que ces membres construisent la relation sociale par leurs activités sociales, la possibilité d’établissement de la relation des nœuds est également apportée. Par exemple, les gens qui se rencontrent souvent sur un même chemin peuvent facilement se connaître et construire la relation sociale. Donc, nous pouvons déterminer la relation de confiance entre les nœuds en comparant la similitude de trajectoire.

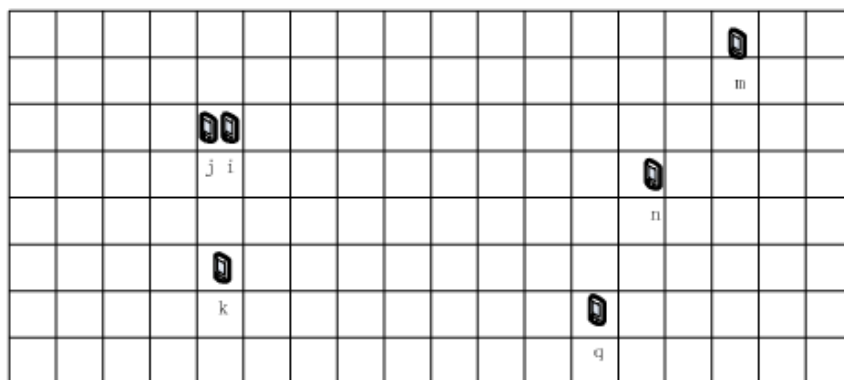


Figure 6-3 : Exemple de division de zone de réseau

Tout d'abord, afin de décrire précisément la trajectoire mobile du nœud, il est nécessaire de diviser en différentes zones la totalité du réseau. En prenant en compte la règle de l'activité humaine décrite dans la littérature [138], nous divisons le réseau en sous-zones carrées de $200m \times 200m$, et étiquetons toutes les zones carrées. Comme illustré dans la Figure 5-3, les nœuds i, j, k, m, n, q sont distribués dans différentes sous-zones du réseau.

De plus, dans le mémoire de chaque nœud, une table matricielle à deux dimensions est configurée avec le même nombre de division que celles données ci-dessus. C'est une table de position du nœud L_i qui est utilisée pour enregistrer le chemin du mouvement du nœud. Nous supposons qu'initialement tous les termes de la table de position de nœud L_i sont 0, le nœud peut détecter de lui-même la sous-zone où il est situé pendant son mouvement. Pour les sous-zones actuelles et passagères, nous les marquons avec 1, et les zones non fréquentées conservent leur 0. En observant la position de la table finale de distribution 1, nous pouvons avoir la trajectoire de déplacement du nœud.

Pour faciliter le calcul de similitude de la trajectoire, après la rencontre des nœuds, la table de position L_i est transformée en matrice de position L_i , les nœuds échangent leurs matrice de position VL_i et VL_j , et puis calculent respectivement la similitude de trajectoire entre les deux ; la formule de calcul est donnée ci-dessous.

$$LSim(L_i, L_j) = \frac{VL_i \cdot VL_j}{\|VL_i\| \cdot \|VL_j\|} \quad (6-9)$$

La formule (6-9) utilise la formule d'angle de vecteur pour calculer la similitude de la trajectoire des matrices de position VL_i et VL_j , elle est montrée par la valeur de cosinus d'angle compris. $\|VL_i\|$ signifie le nombre des sous-zones de passage dans le réseau du nœud i , $LSim(L_i, L_j)$ signifie la similitude de la trajectoire de nœud i et du nœud j .

En outre, avec le mouvement incessant des nœuds, le chemin de chaque nœud mobile change constamment. Pour les zones qui n'ont pas été accédées depuis une certaine période, il faut renouveler le calcul de la trajectoire du mouvement

pour assurer que la trajectoire actuelle représente bien les sous-zones de passage récentes. Par conséquent, chaque nœud enregistre l'heure d'entrée précédente t_{in} d'une sous-zone, l'heure de sortie précédente t_{out} d'une sous-zone et l'heure actuelle t_{now} . Selon le temps de séjour du nœud dans une certaine sous-zone, la vitesse de mise à jour d'étiquette est proportionnelle à la longueur du temps de séjour, la méthode de calcul de mise à jour est :

$$LUpdate(t_{in}, t_{out}, t_{now}) = e^{-\left\{\frac{1}{\text{round}[t_{now}-(t_{out}-t_{in})]}\right\}} \quad (6-10)$$

La fonction $\text{round}()$ est une fonction arrondie, $LUpdate(t_{in}, t_{out}, t_{now})$ donnant la valeur de mise à jour d'une certaine sous-zone, seulement si la valeur de mise à jour est en dessous du seuil φ , l'étiquette de cette sous-zone dans la table de position va être remis à 0, et sinon restera à l'état de l'étiquette d'origine.

6.3.2 Similitude de relation sociale

Dans des circonstances normales, le nombre d'amis communs des différents membres de la société peut illustrer dans un certain sens l'intensité de la relation entre les membres [139] ; plus ils ont d'amis en commun, et plus leur relation est forte. Par exemple, les élèves d'une même classe ont tendance à connaître tous les autres élèves de la classe, une bonne relation de confiance entre eux peut être établie. Par conséquent, nous pouvons comparer cela à la relation sociale entre les nœuds de réseau opportuniste qui reflète la relation de confiance.

Tout d'abord, le nœud a besoin de la liste d'adjacence pour enregistrer le nombre de nœuds d'interaction dans les sous-zones ; le nœud de tête de la liste d'adjacence est composé par le champ de données des sous-zones et le pointeur qui pointe sur le premier nœud de la liste d'adjacence. Le nœud de la liste est composé par le champ des nœuds voisins, le champ de données comprenant la notification du nom des nœuds d'interaction et le pointeur qui pointe sur le nœud suivant, comme la figure 6-4 le montre. Si le nœud i a une interaction respectivement avec le nœud b et le nœud u dans la zone 1, les

nœuds b et u vont être ajoutés après la zone 1 suivant l'ordre de rencontre, ce qui signifie que le nœud i a deux notes d'interaction dans la zone 1.

Dans le même temps, afin d'enregistrer la situation de relation sociale des nœuds, une table de relation sociale L_i est installée dans le mémoire du nœud, identique à la table de position du nœud, chaque terme de la table de relation sociale correspondant aussi aux sous-zones du réseau. Le nombre des nœuds rencontré dans chaque sous-zone comptabilisé par la liste d'adjacence est également ajouté à la table de relation sociale. Tous les paramètres initiaux des termes de la table de relation sociale sont à 0, ce qui représente que le nombre de nœud rencontré par le nœud i dans les sous-zones est à 0. Suivant l'interaction des nœuds dans les différentes zones, les paramètres des termes de la table se modifient avec l'augmentation de nombre d'interaction.

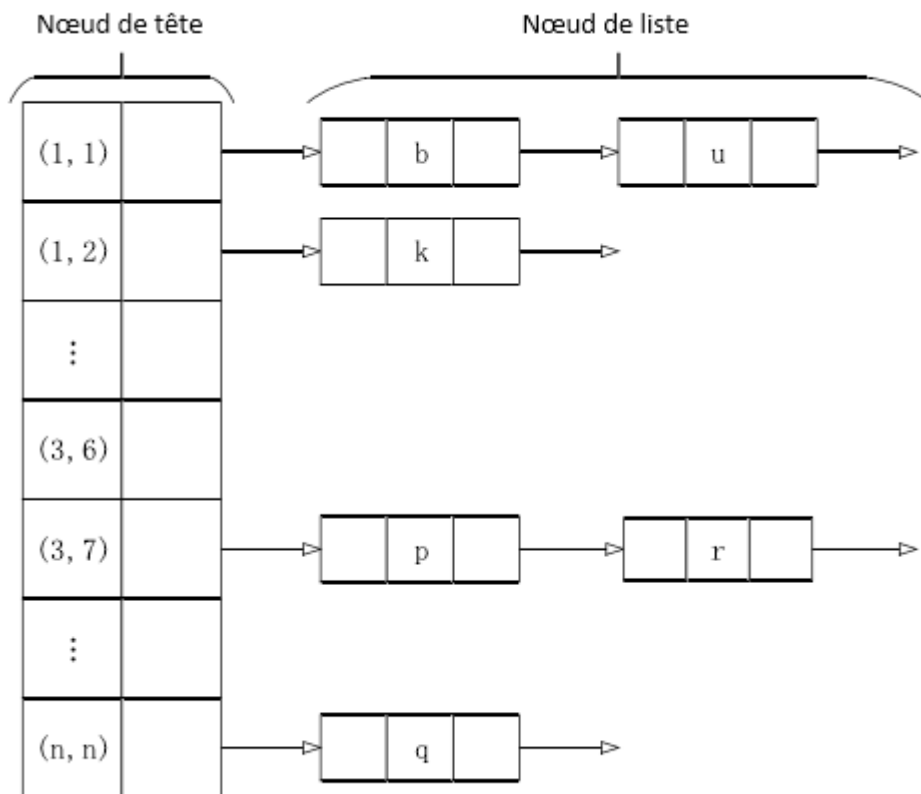


Figure 6-4 : Situation d'interaction de nœud i

Si le nœud i et le nœud j se rencontrent, ils échangent premièrement leurs matrices de relation sociale $V F_i$, et la similitude de relation sociale des nœuds peut ensuite être calculée par la formule :

$$FSim(F_i, F_j) = \frac{VF_i \cdot VF_j}{\|VF_i\| \cdot \|VF_j\|} \cdot Dec_{i,j}(t_{last} t_{now}) \quad (6-11)$$

VF_i correspond au nombre de nœuds de confiance de la table de confiance, $FSim(F_i, F_j)$ représente la similitude dans la relation sociale entre les nœuds i et j , $Dec_{i,j}(t_{last} t_{now})$ correspond au degré d'affaiblissement de la similitude de relation sociale entre les nœuds i et j . Pour des nœuds sans interaction pendant une longue période, la similitude au niveau de la relation sociale va certainement baisser avec le temps. Ainsi, la similitude de la relation sociale entre les nœuds doit avoir une caractéristique d'affaiblissement. Nous supposons que le temps d'interaction depuis la dernière interaction entre les nœuds est t_{last} , le degré d'affaiblissement de relation sociale des nœuds est donc :

$$Dec_{i,j}(t_{last}, t_{now}) = e^{-round(t_{now} - t_{last})} \quad (6-12)$$

$Dec_{i,j}(t_{last}, t_{now})$ dans la formule signifie le degré d'affaiblissement de relation sociale entre les nœuds i et j ; il est proportionnel à l'intervalle de temps d'interaction : plus le temps sans interaction est long, plus le degré d'affaiblissement est grand.

6.3.3 Niveau de confiance de nœud

D'un point de vue sociologique, la similitude de trajectoire et la similitude de la relation sociale décrivent l'établissement de la relation en profitant de la mobilité et la région active des nœuds. Dans le réseau opportuniste social, nous pouvons donc déterminer la relation de confiance entre les nœuds en comparant la similitude de trajectoire et la similitude de relation sociale des différents nœuds. Nous utilisons une somme pondérée pour intégrer les deux, et le niveau de confiance est $T_{i,j}$:

$$T_{i,j} = Sim_{i,j} \quad (6-13)$$

$T_{i,j}$ signifie le niveau de confiance du nœud j pour le nœud i , $Sim_{i,j}$ représente la similitude globale des nœuds j et i , et la somme pondérée de la similitude de trajectoire et la similitude de relation sociale. La formule est comme suit :

$$Sim_{i,j} = \alpha \cdot LSim(L_i, L_j) + \beta \cdot FSim(F_i, F_j) \quad (6-14)$$

α et β signifie respectivement le coefficient de poids de la similitude de trajectoire et de la similitude de relation sociale dans la similitude globale, $\alpha, \beta \in (0,1)$ et $\alpha + \beta = 1$. La contrôle de poids de α et β ici doit tenir compte de la subjectivité et de la caractéristique d'évolution d'évaluation de la confiance, c'est-à-dire que la détermination de la valeur de poids doit être basée sur l'environnement d'évaluation de confiance des nœuds. La méthode de distribution de poids est calculée par les formules (6-15) et (6-16).

$$\alpha = \frac{FSim(F_i, F_j)}{FSim(F_i, F_j) + LSim(L_i, L_j)} \quad (6-15)$$

$$\beta = 1 - \frac{FSim(F_i, F_j)}{FSim(F_i, F_j) + LSim(L_i, L_j)} \quad (6-16)$$

6.3.4 Processus de transmission basé sur la confiance

D'une manière générale, après l'achèvement de l'évaluation de la confiance du nœud, il faut sélectionner le nœud de confiance pour le prochain saut dans la transmission de données. Si l'on fait un parallèle avec la psychologie du consommateur, lorsque les membres d'une société achètent des biens sur le marché, ils préfèrent des produits qui ont une garantie de stabilité de qualité. Notre algorithme examine la stabilité de confiance selon la valeur d'évaluation de la confiance historique des nœuds pour choisir le nœud de confiance comme nœud de transmission, le nœud avec la meilleure stabilité de confiance sera sélectionné comme le nœud de prochain saut de transmission de données parmi des nœuds de confiance. La précision de la confiance est décrite d'une manière plus poussée, afin d'assurer la fiabilité de la transmission de données.

Nous supposons que la séquence de la valeur de confiance historique du nœud i sur le nœud j est $T_{1i,j}, T_{2i,j}, T_{3i,j}, T_{hi,j}$, h signifiant le nombre de fois où il y a eu une évaluation de confiance du nœud i par le nœud j , la méthode de calcul de sa stabilité est exprimée par les deux formules suivantes :

$$\overline{T_{i,j}} = \frac{1}{p} \cdot \sum_{h=1}^p T_{hi,j} \quad (6-17)$$

$$Sta_{i,j} = \frac{1}{p} \cdot \sum_{h=1}^p (T_{h_{i,j}} - \overline{T_{i,j}})^2 \quad (6-18)$$

Le formule (6-17) utilise d'abord la valeur de confiance historique de nœud i sur nœud j pour calculer la valeur moyenne d'évaluation de confiance de p fois $\overline{T_{i,j}}$. Et puis $\overline{T_{i,j}}$ est substitué dans la formule (6-18) pour calculer la variance $Sta_{i,j}$ d'évaluation de confiance de p fois précédent, la variance $Sta_{i,j}$ est donc la stabilité de confiance du nœud. Enfin, pour comparer la variance $Sta_{i,j}$ entre les différents nœuds, nous pouvons dire que si la variance est plus petite, la stabilité et la fiabilité de la confiance entre des nœuds est meilleure, au contraire, si la variance est plus grande, cela indique une mauvaise stabilité et fiabilité dans la relation de confiance.

Le processus spécifique de la transmission basé sur la confiance peut être divisé en plusieurs étapes:

1. Calculer la valeur de confiance de chaque nœud avec les autres nœuds dans son champs de communication $T_{i,j}$, et puis calculer la valeur moyenne d'évaluation de confiance $\overline{T_{i,j}}$ et la variance $Sta_{i,j}$ des autres nœuds grace à la valeur historique de confiance de ce nœud.
2. Selon la valeur moyenne de confiance $\overline{T_{i,j}}$, sélectionner des nœuds qui ont une valeur moyenne de confiance $\overline{T_{i,j}}$ plus grande que la valeur du seuil de confiance T_θ , soit $T_{i,j} \geq T_\theta$, et les mettre dans l'ensemble des nœuds $TempN_h = \{N_h\}$.
3. Observer la variance $Sta_{i,j}$ du nœud dans l'ensemble des nœuds $TempN_h = \{N_h\}$, et les classer de la plus petite à la plus grande, et choisir le nœud de la plus petite variance comme un nœud de transmission de confiance.

L'analyse de la complexité du processus d'évolution de confiance ci-dessus montre que la confiance dépend du nombre de nœuds se rencontrant et de la dimension de la division du réseau. En supposant que le nombre des nœuds est m pour le scénario et la dimension de la division du réseau est $n \times n$, le cas le plus défavorable de l'algorithme est un nœud rencontrant n-1 autres nœuds ; dans ce cas la complexité de calcul d'algorithme est $O(mn^2)$. La meilleure situation est lorsqu'un nœud rencontre seulement un autre nœud ; la

complexité de calcul d'algorithme est alors de $O(n^2)$. En outre, puisque les données de stockage et données d'entrée et de sortie de l'ensemble de l'algorithme d'évaluation comprennent une table à deux dimensions et un tableau à une dimension etc., la complexité spatiale de l'algorithme est alors de $O(n)$.

En résumé de ce chapitre, nous avons présenté dans un premier temps les raisons pour lesquelles nous avons fait la proposition de l'algorithme de transmission de confiance basé sur la similitude de mouvement des nœuds et la base théorique de l'algorithme. Ce modèle analyse la trajectoire et la zone d'activité des membres sociaux dans le réseau opportuniste social, divise le réseau opportuniste en de nombreuses sous-zones qui sont installées dans le mémoire du nœud, marque la trajectoire de mouvement des nœuds par des étiquettes différentes, et mesure le niveau de confiance en comparant la trajectoire des différents nœuds. Puis, sur la base des nœuds d'interaction enregistrés dans les différentes zones, nous avons comparé la situation de relation sociale des nœuds, et déterminé ainsi le niveau de confiance entre les nœuds. Enfin, le niveau de confiance globale des nœuds est obtenu en synthétisant la similitude de trajectoire et la similitude de la relation sociale. En outre, en référence aux études sur la psychologie du consommateur, nous avons calculé la stabilité de confiance pour sélectionner le nœud qui a une meilleure stabilité comme prochain saut.

Chapitre 7. Simulation et expérimentation de l'application

7.1 Plate-forme de simulation

Afin de vérifier l'algorithme de transmission de confiance proposé, nous utilisons l'«*Opportunistic Network Environment (ONE)*» [140] qui est conjointement développé par l'entreprise finlandaise Nokia et l'Université de technologie d'Helsinki comme logiciel de simulation. ONE fournit le développement en langage java et possède les caractéristiques de la programmation orientée objet, basée sur les événements discrets et la simulation de scénarios réels du réseau. Ce logiciel intègre la simulation de différents aspects du réseau opportuniste, par exemple le modèle de mobilité, la stratégie de routage et la disposition d'environnement etc. Grâce à l'interface visuelle, tous les détails du processus de simulation peuvent être montrés et il fournit des rapports abondants sur les résultats et les outils d'analyse en fonction des besoins de l'utilisateur. En outre, ONE est un logiciel open source qui peut être modifié et configuré selon la demande des utilisateurs, avec des extensions possibles.

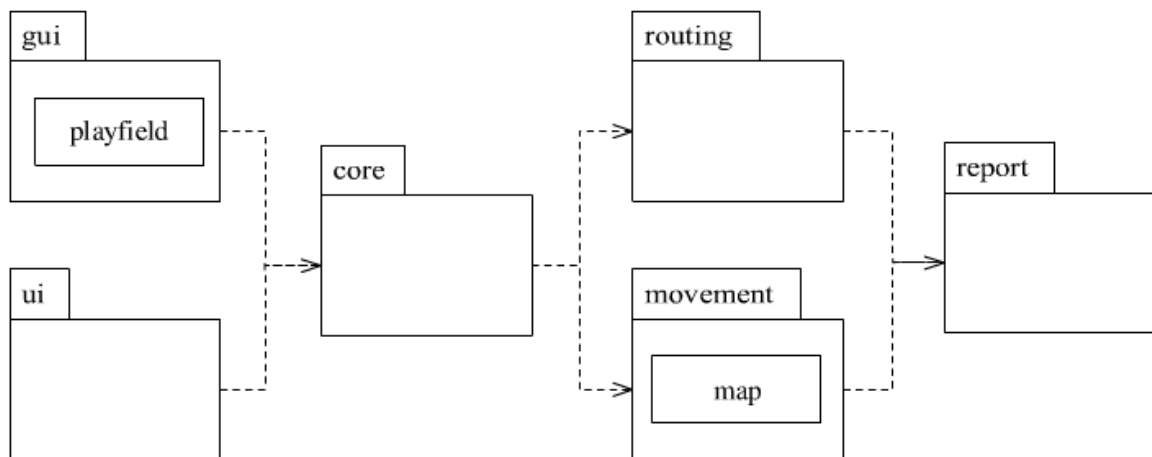


Figure 7-1 : Relation des paquetages de simulateur ONE

Présenté dans la figure 7-1, le simulateur ONE principalement fonctionne en coordonnant les différents modules *gui*, *ui*, *core*, *mouvement*, *routage*, *rapport* et d'autres modules. Le module *gui* complète la conception de l'interface graphique d'ONE, le module *ui* complète la partie interface graphique de la

conception des caractères, et ils réalisent ensemble une fenêtre de visualisation. Le module *core* est le module de base des fonctions d'ONE, qui réalise l'initialisation des fichiers de configuration et plusieurs fonctions comme l'installation des nœuds, des paquets et l'interface de communication etc. Le module *movement* implémente le mode de mouvement des nœuds dans le réseau en fonction des besoins d'utilisateur, comme le mode de mouvement basé sur la carte, le mode de mouvement des bus, et le mode aléatoire etc. Le module *routing* réalise principalement la fonction de routage du réseau opportuniste, ce module comprend des classes de toutes les méthodes de routage du réseau, qui peuvent être grossièrement divisées en types actifs et passifs. Le module *report* réalise la génération des données d'expérimentation, ce module contient une foule des classes de génération de données, et fournit l'utilisateur une base importante pour l'analyse expérimentale.

La figure XXXXXX montre la relation d'appel des modules du simulateur ONE. Dans le processus de fonction, d'abord l'instance de scénario du module *core* est 'instancé' par les modules *gui* et *ui*, ensuite le module *core* crée l'instance de mode de mouvement du module *routing* et du module *movement* du scénario. Après la fin de l'expérimentation, le module *routing* et le module *movement* vont appeler des classes relatives au module *report*, et exporter des données expérimentales au format txt.

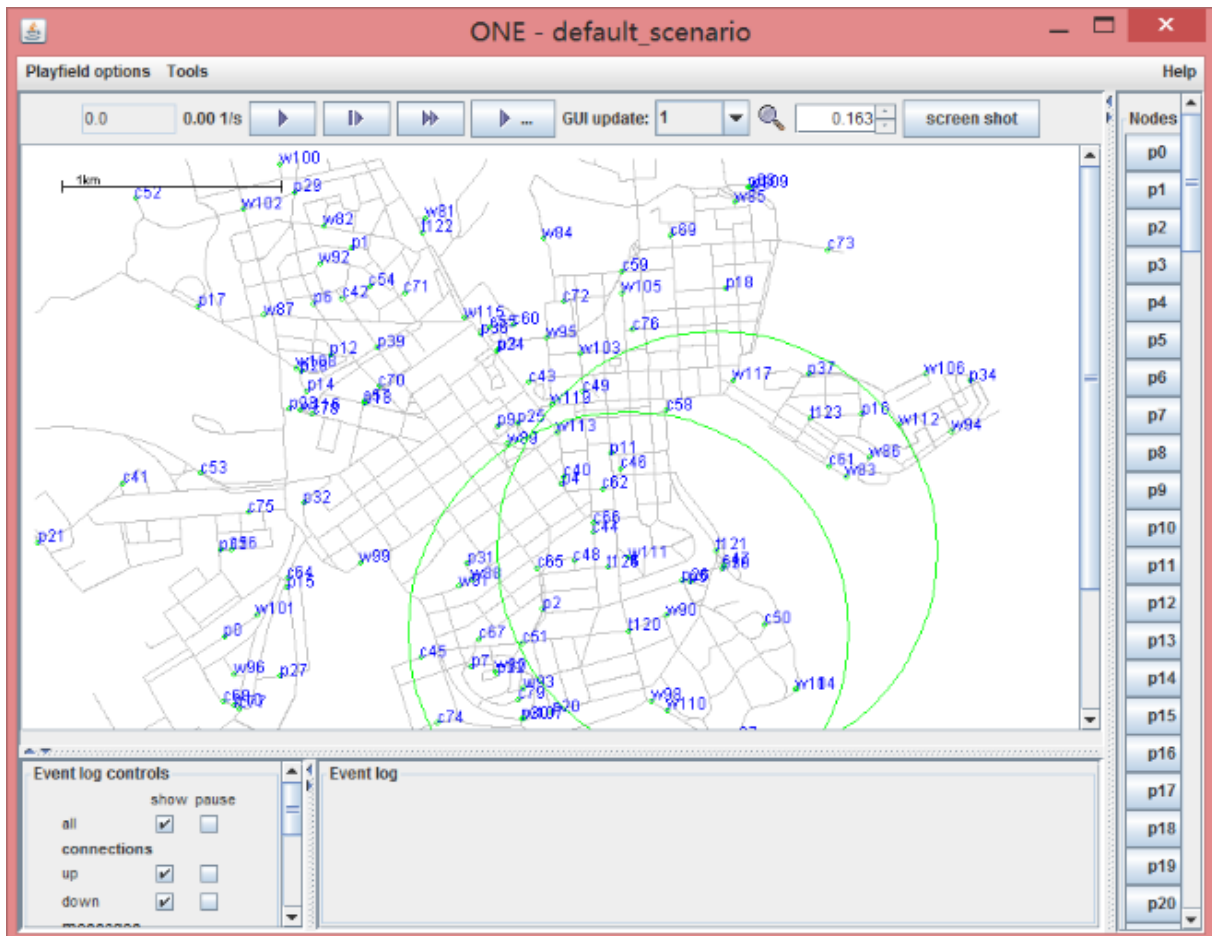


Figure 7-2 Fenêtre graphique visuelle du Simulateur ONE

ONE dispose de deux modes de fonctionnement, respectivement le mode graphique visuel de l'utilisateur (GUI) et le mode de batch. Le mode visuel affiche un certain processus de simulation par une interface graphique visuelle en appelant le module *gui* du simulateur. Décrit dans la figure 7-2 nous pouvons clairement voir la distribution de l'espace, les types de nœuds, toutes sortes de boutons de contrôle de scénario et l'enregistreur d'événements etc. pour le scénario fonctionnel. Le mode batch permet de procéder à de multiples

expériences de simulation simultanées, mais ne fournit pas une interface graphique visuelle dans ce mode, donc nous ne pouvons pas connaître la situation fonctionnelle de chaque nœud dans le processus de simulation, mais seulement par le rapport sur les données expérimentales après la fin de la simulation.

Avant le démarrage des deux modes de fonctionnement ci-dessus, le simulateur lit d'abord le fichier de configuration de scénario *default_settings.txt* déjà configuré par l'utilisateur, et puis initialise les paramètres de fonctionnement du scénario d'ONE. Les réglages des paramètres généraux peuvent être divisés en: réglages des paramètres de scénario, des paramètres de l'interface de transmission, des paramètres des paquets de données, des paramètres d'informations de nœud, des résultats d'expérimentation. Les réglages des paramètres de scénario comprennent le nom du scénario, le temps de fonctionnement du scénario, la vitesse de pas de temps, ainsi que l'intervalle de renouvellement d'interface graphique etc. Les réglages des paramètres de l'interface de transmission règlent principalement le type et le mode de communication des nœuds utilisés par les nœuds dans le scénario, y compris le mode de transmission de communication, la vitesse de transmission de communication et la sphère de transmission de communication. Les réglages des paramètres des paquets de données définissent des aspects relatifs aux paquets de données du réseau, y compris la méthode de génération des paquets, l'intervalle de génération des paquets, le nom du paquet préfixe, la taille des paquets, le temps de survie des paquets, le cadre d'adresses des paquets etc. Les réglages des paramètres d'informations de nœud comprennent l'ID des nœuds, l'étiquette de classification des nœuds, la vitesse de mouvement des nœuds, la taille du cache, le mode de mobilité et méthodes de routage etc. ONE offre une variété de rapports de résultat de données d'expérimentation ; nous pouvons régler les paramètres des résultats selon les besoins en utilisant de multiples rapports de données, et les placer dans le même répertoire. En outre, dans le fichier de configuration, nous pouvons également régler l'interface graphique, le temps de mise en route du scénario et les informations de regroupement.

7.2 Paramètres d'évaluation

Le but ultime du mécanisme de confiance de réseau opportuniste concerne le service de routage et la fiabilité des données sur le réseau. Afin de mesurer l'efficacité de l'algorithme proposé, nous sélectionnons le taux de succès de transmission et le temps de délai moyen en tant que paramètres d'évaluation ; ces deux paramètres sont des indicateurs clés pour vérifier l'algorithme de routage, que nous introduirons brièvement ci-dessous.

7.2.1 Taux de succès de transmission

Le taux de succès de transmission indique, pour une certaine période de temps, le ratio entre le nombre de paquets de données reçus par le nœud de destination finale et le nombre de paquets envoyés par le nœud source dans le réseau. Il est calculé par la formule (7-1) ci-dessous.

$$DeliverySR(i, j, data_k) = \frac{NumReceive_j}{NumSend_i} \quad (7-1)$$

Le $DeliverySR(i, j, data_k)$ signifie le taux de succès des paquets k du nœud de source i au nœud de destination k , $NumSend_i$ signifie le nombre de paquets envoyés par le nœud source, $NumReceive_j$ signifie le nombre de paquets reçus par le nœud de destination.

7.2.2 Temps de délai moyen

Le temps de délai moyen est le temps passé au total depuis l'envoi du paquet par le nœud source jusqu'à son arrivée dans le nœud de destination. En raison des caractéristiques du réseau opportuniste, et le mode de transmission de données de « porter - stocker – transmettre », pour de nombreux paquets de même type, le temps de délai du nœud source au nœud de destination varie entre l'un et l'autre, généralement le temps moyen va être calculé comme le critère d'évaluation.

7.3 Simulation d'algorithme de transmission de données basé sur la confiance sociale binaire

7.3.1 Environnement d'expérimentation

L'expérimentation 1 réalise la simulation de l'algorithme de transmission *ST-Epidemic* proposé dans le chapitre 6. Nous mesurons les deux indicateurs, le taux de succès de transmission et le délai moyen pour examiner sa performance. Notre expérimentation ici est divisée en deux groupes, en comparant respectivement le taux de succès de transmission et le temps de délai moyen de quatre modes de transmission *ST-Epidemic*, *Epidemic*, *Prophet*, *Spray-and-Waits* sous différents ratio de nœuds égoïstes. Les résultats d'expérimentation de simulation vont être analysés après.

Pour correspondre à la situation réelle, nous prenons le scénario d'un quartier de la capitale de Finlande : la ville Helsinki présente dans le simulateur ONE. Il y a 200 nœuds au total dans la zone de simulation, et en utilisant la communication Bluetooth entre eux, le rayon de communication est de 10m, la vitesse de mouvement du nœud est compris entre [1-3] km / h, et le débit de transmission des données est 200Kbps. Le réglage des paramètres environnementaux spécifiques sont présentés dans le tableau 6-1.

Paramètres environnementaux	Valeur
Surface de couverture du réseau	4500m*3500m
Nombre total de nœuds	200
Nombre des nœuds égoïstes	20, 40,, 160
Mode de mouvement	SPMBM
Cadre de la vitesse de mouvement	1-3m/s
Canal de communication	Bluetooth
Rayon de communication	10m
Débit de transmission	200Kbps
Taille des paquets de données	50Kb

Cycle de vie des paquets	3600s
Fréquence de génération des paquets	30s
Capacité de la mémoire des nœuds	30Mb
Temps de simulation	21600s

Tableau 7-1: Paramètres environnementaux de simulation

7.3.2 Résultat d'expérimentation et analyse

Les Figures 7-3 et 7-4 présentent respectivement le taux de succès de transmission des nœuds égoïstes présents à 30% et 70% dans le réseau sous quatre modes de transmission.

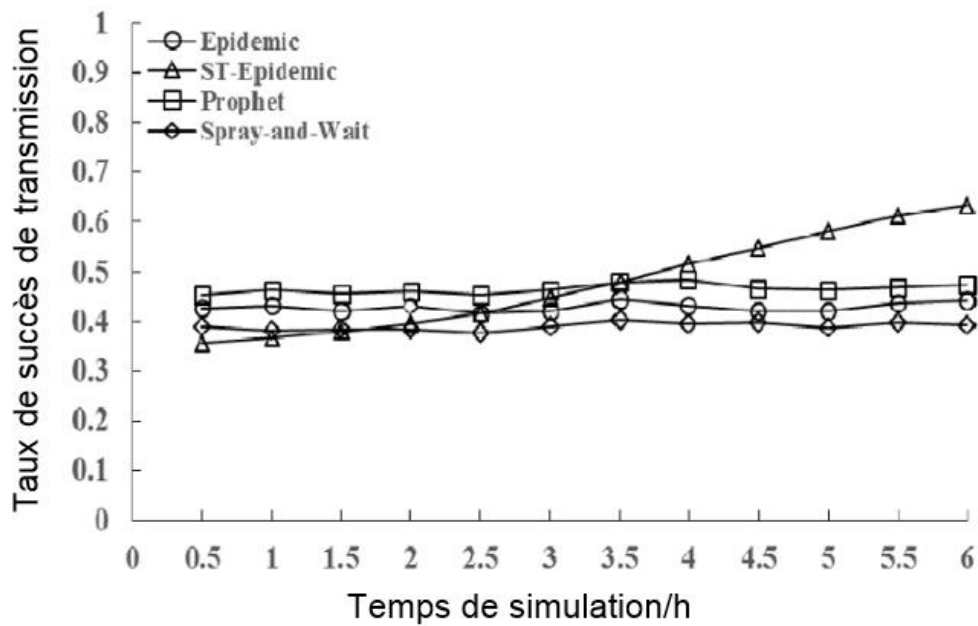


Figure 7-3 : Taux de succès de transmission avec un ratio de nœuds égoïstes de 30%

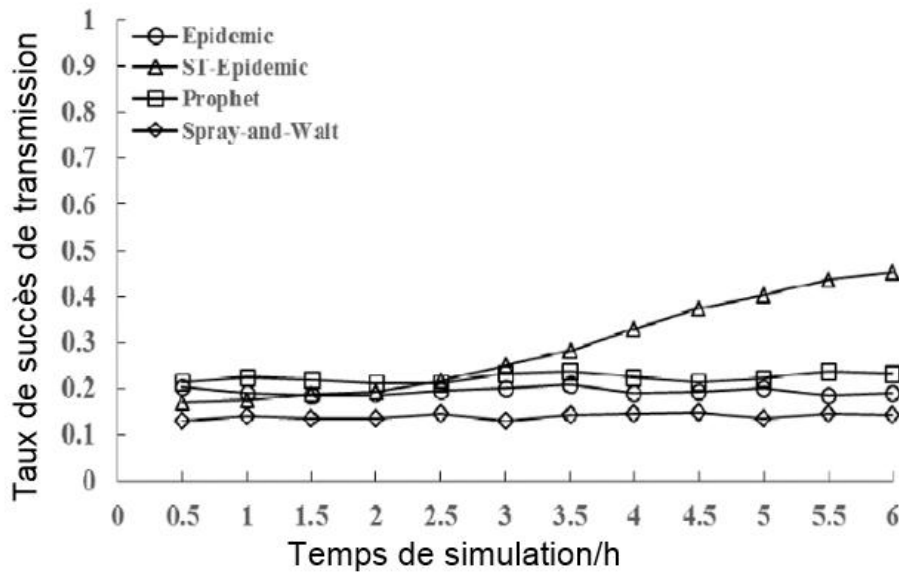


Figure 7-4 : Taux de succès de transmission avec un ratio de nœuds égoïstes de 70%

Dans la première partie de l'expérience, le taux de succès de transmission de *ST-Epidemic* est similaire au mode de transmission *Epidemic*, *Prophet*, *Spray-and-Waits*, mais avec l'augmentation du nombre d'interactions entre les nœuds, le taux de succès de transmission de *ST-Epidemic* devient progressivement supérieur aux trois autres modes de transmission. Parce qu'en début d'expérimentation, la relation de confiance entre les nœuds est encore dans le processus d'établissement, il peut y avoir le cas où certains nœuds normaux sont temporairement considérés comme des nœuds égoïstes, l'augmentation du taux de succès de transmission de *ST-Epidemic* est alors lent. Dans la dernière partie de l'expérience, grâce à l'accumulation progressive de bonnes relations de confiance entre des nœuds normaux, les données sont transférées en sécurité entre les nœuds de confiance; d'autre part, en raison du fait que les nœuds égoïstes refusent d'aider les autres nœuds pour la transmission de données, la proportion de service des nœuds égoïstes est basse, ce qui entraîne également une valeur de la confiance sociale binaire des nœuds égoïstes aussi basse; les nœuds normaux évitent progressivement les nœuds égoïstes par la détermination de la confiance pour la transmission de

données, de sorte que le taux de succès de transmission de *ST-Epidemic* augmente rapidement.

La figure 7-5 montre que tous les taux de succès de transmission de *ST-Epidemic* , *Epidemic* , *Prophet* et *Spray-and-Waits* baissent avec l'augmentation de la présence de nœuds égoïstes dans le scénario, alors que la tendance à la baisse de *ST-Epidemic* est modérée. Bien que quand la présence de nœuds égoïstes est inférieure à 30%, les taux de succès de transmission des modes *Epidemic* , *Prophet* et *Spray-and-Waits* dépassent celui de *ST-Epidemic*, mais si le ratio des nœuds égoïstes est supérieur à 30%, le taux de succès de transmission de *ST-Epidemic* sera plus élevé que les autres modes. La raison principale est que les nœuds égoïstes ont un impact sérieux sur le processus de transmission de *Epidemic* , *Prophet* et *Spray-and-Waits*, et un grand nombre de données transférées est intercepté et supprimé par les nœuds égoïstes, engendrant un taux de succès de transmission de données diminuant rapidement. En revanche, pour le mode de *ST-Epidemic*, sous différents ratio de nœuds égoïstes, la transmission de données est efficacement assurée par la relation de confiance déjà établie, donc le déclin du taux de succès de transmission est atténué.

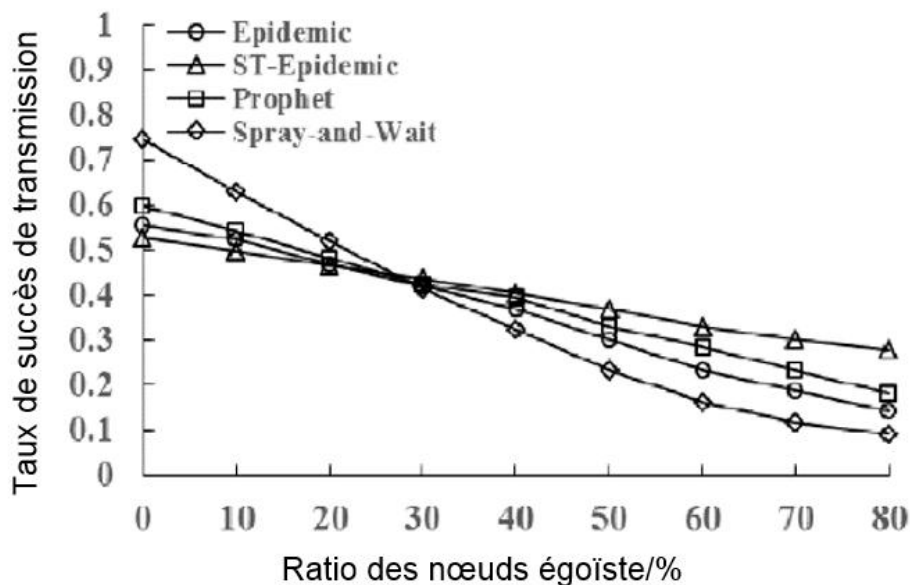


Figure 7-5 : Taux de succès de transmission à différent ratio de nœuds égoïstes

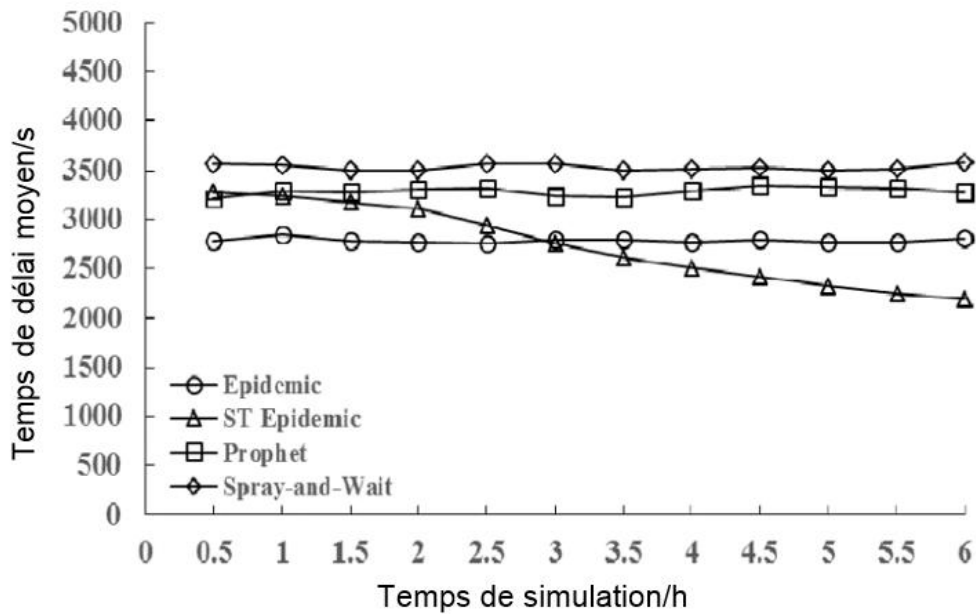


Figure 7-6 : temps de délai moyen avec un ratio de nœuds égoïstes de 30%

Les figures 7-6 et 7-7 montrent le temps de délai moyen des quatre modes de transmission avec un ratio de nœuds égoïstes de 30% et 70%

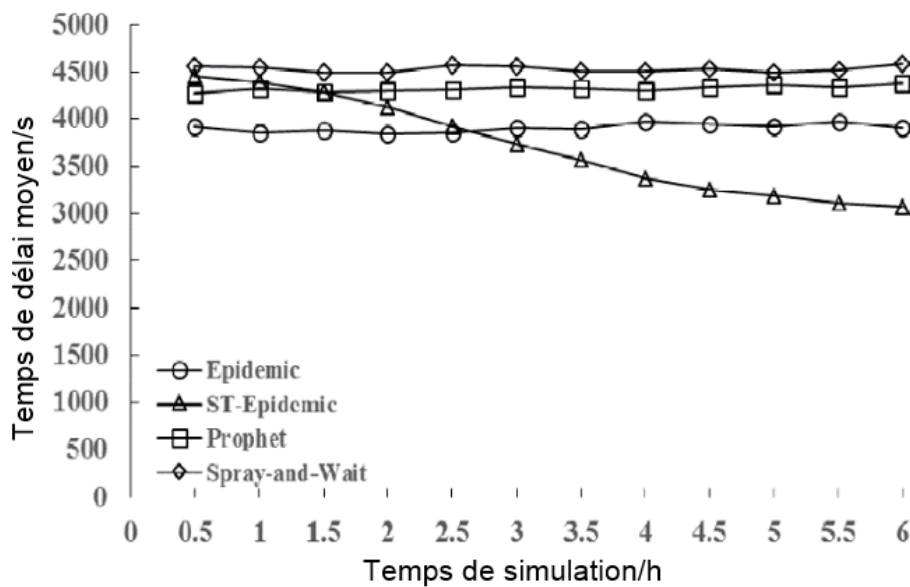


Figure 7-7 : temps de délai moyen avec un ratio des nœuds égoïstes de 70%

Dans la première partie de l'expérience, le temps de délai moyen de ST-Epidemic est principalement entre celui d'Epidemic et celui de Spray-and-Waits. Mais pour la dernière partie de l'expérience, le temps de délai moyen de ST-

Epidemic est plus court que les trois autres. C'est parce qu'en début de l'expérimentation, il existe la possibilité que certains des nœuds normaux soient considérés comme des nœuds égoïstes, ce qui cause un temps de délai moyen de *ST-Epidemic* similaire aux autres modes de transmission. Dans la dernière partie de l'expérience, grâce à l'établissement de bonnes relations de confiance entre des nœuds normaux, qui peuvent transférer des données en évitant des nœuds égoïstes peu par peu, la situation où les données transmises soient récupérées par des nœuds égoïstes est moins fréquente, et donc le temps de délai moyen est bien réduit.

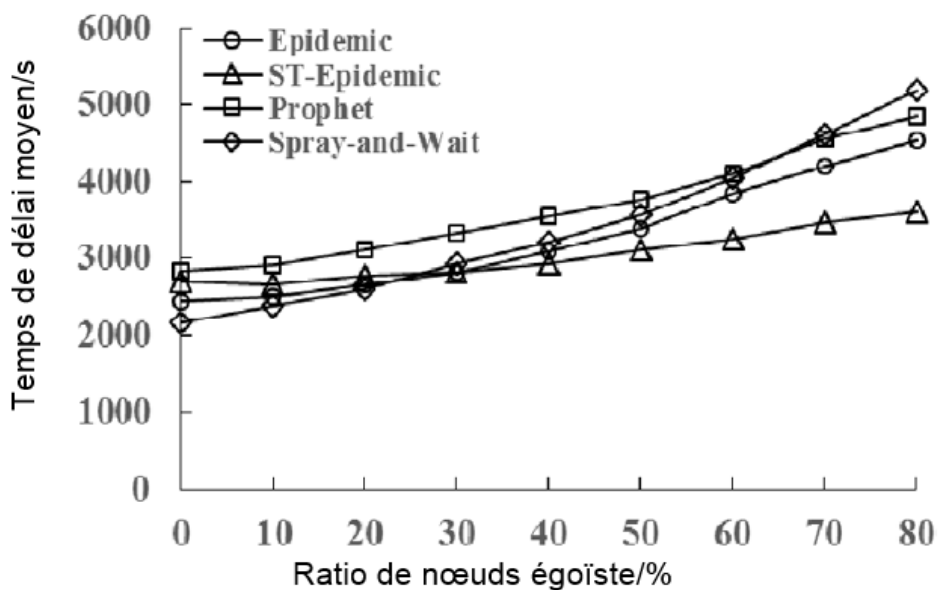


Figure 7-8 : temps de délai moyen avec différents ratio de nœuds égoïstes

La figure 7-8 montre le temps de délai moyen des modes *ST-Epidemic*, *Epidemic*, *Prophet* et *Spray-and-Waits* avec différents ratio de nœuds égoïstes. Avec l'augmentation du ratio de nœuds égoïstes, le temps de délai moyen des quatre algorithmes est en hausse, mais la tendance à la hausse de *ST-Epidemic* est plus faible. Lorsque le ratio des nœuds égoïstes est inférieur à 30%, le temps de délai moyen de *ST-Epidemic* est légèrement plus long que *Epidemic* et *Spray-and-Waits*. Mais quand le ratio est supérieur à 30%, le temps de délai moyen de *ST-Epidemic* est plus court que des trois autres modes de transmission. C'est parce que la relation de confiance établie empêche effectivement l'interférence des nœuds égoïstes, et ralentit la hausse du temps de délai moyen. Mais *Epidemic*, *Prophet* et *Spray-and-Waits* sont gravement

affectées par des nœuds égoïstes, ce qui réduit la probabilité des données envoyées au nœud de destination, de sorte que la moyenne des temps de délai augmente rapidement.

7.4 Simulation de l'algorithme de transmission de confiance basé sur la similitude des communautés dynamiques des nœuds

7.4.1 Environnement d'expérimentation

Nous utilisons également le simulateur ONE pour évaluer cet algorithme proposé. Parce qu'il faut cette fois diviser le zone d'expérimentation en sous-zones, nous profitons des paramètres similaire à la situation réelle du scénario de la ville d'Helsinki et simulons des zones carrées qui peuvent être divisées en sous-zones de 200m*200m. 200 nœuds sont installés pour la simulation, en utilisant la communication Bluetooth entre eux, un rayon de communication de 10m, une vitesse de mouvement du nœud de [1-3] km/h, un débit de transmission des données de 200Kbps, et également un réglage des paramètres environnementaux spécifiques présenté dans le tableau 7-2.

Paramètres environnementaux	Valeur
Surface de couverture du réseau	1000m*1000m, 2000m*2000m, 3000m*3000m
Surface des sous-zones du réseau	200m*200m
Nombre total de nœuds	200
Nombre des nœuds égoïstes	20, 40,, 160
Mode de mouvement	SPMBM
Cadre de la vitesse de mouvement	1-3m/s
Canal de communication	Bluetooth
Rayon de communication	10m
Débit de transmission	200Kbps
Taille des paquets de données	50Kb

Cycle de vie des paquets	3600s
Fréquence de génération des paquets	30s
Capacité de la mémoire des nœuds	30Mb
Temps de simulation	21600s

Tableau 7-2 : Paramètres environnementaux de simulation

7.4.2 Résultat d'expérimentation et analyse

Les figure 7-9, 7-10, et 7-11 comparent respectivement dans les couvertures réseaux de 1000m*1000m, 2000m*2000m, 3000m*3000m et avec un ratio de nœuds égoïstes différent, le taux de succès des trois modes de transmission *Epidemic*, *Prophet* et *Spray-and-Waits* en l'absence et avec notre algorithme de confiance.

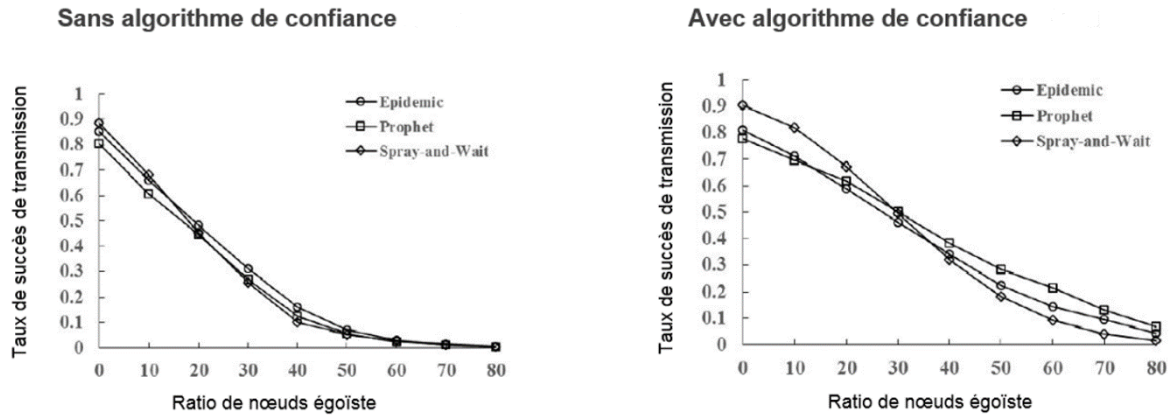


Figure 7-9 : Taux de succès de transmission dans la zone 1000m*1000m

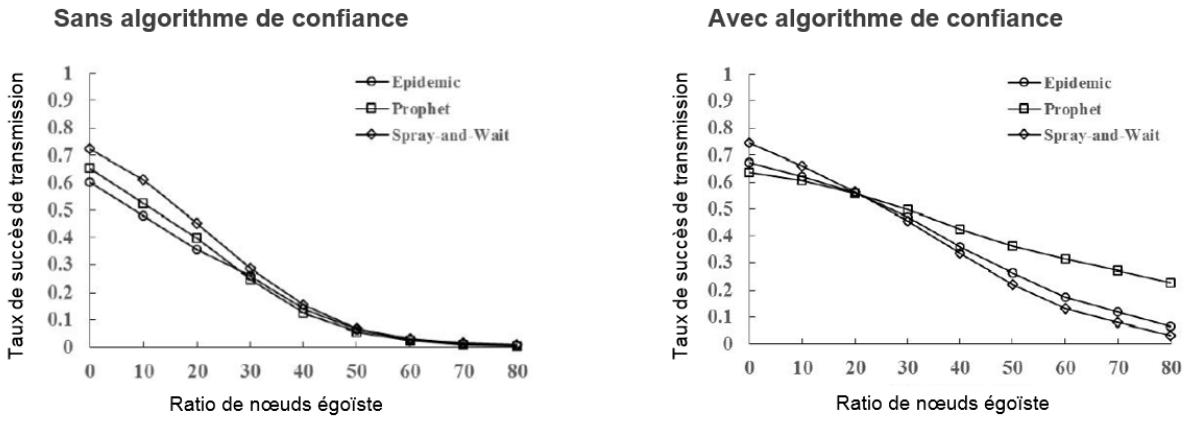


Figure 7-10 : Taux de succès de transmission dans la zone 2000m*2000m

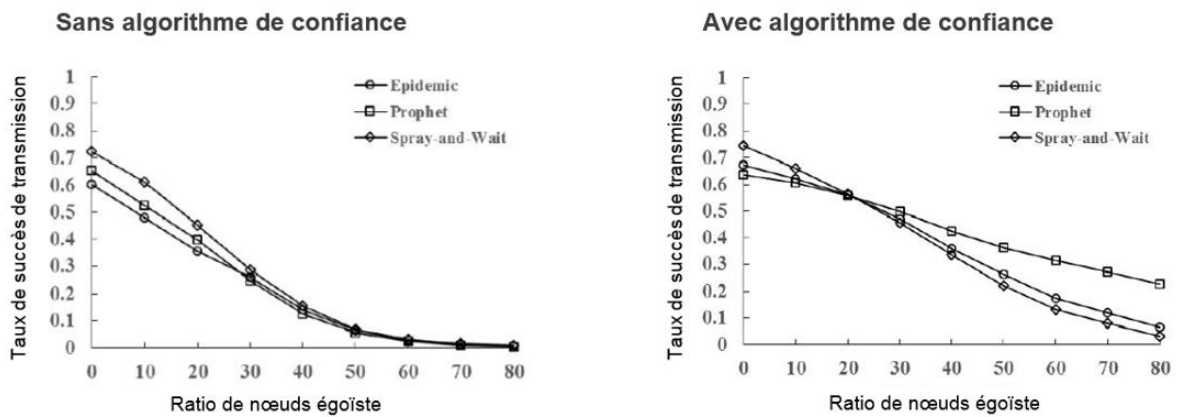


Figure 7-11 : Taux de succès de transmission dans la zone 3000m*3000m

Nous pouvons voir sur les trois figures ci-dessus qu'en l'absence d'algorithme de confiance, et avec l'augmentation de la couverture du réseau, le taux de succès de transmission diminue. La tendance générale est également à la baisse si on augmente le ratio de nœuds égoïstes. Lorsque le ratio de nœuds égoïstes augmente à 50% ou 60%, le taux de succès de trois types de mode de transmission est pratiquement à 0, et seulement une très petite quantité de données peut être transférée avec succès, le réseau est essentiellement dans un état d'effondrement. Dans le cas de l'ajout d'algorithme de confiance, bien que la tendance du taux de succès diminue en fonction de l'augmentation du ratio de nœuds égoïstes, cependant, grâce à la relation de confiance établie par la stratégie basée sur le contexte social, la fiabilité de la transmission de données est assurée efficacement, ce qui réduit l'influence des nœuds égoïstes dans le processus de transmission de données.

En outre, en comparant le taux de succès des trois modes de transmission sur la même surface de réseau, le taux de succès de transmission d'*Epidemic* est plus grand que les deux autres modes si le ratio de nœuds égoïstes est bas ; le taux de succès de transmission de *Prophet* est relativement grand quand le ratio de nœuds égoïstes est élevé ; la performance du taux de succès de transmission de *Spray-and-Waits* est globale sur l'ensemble. Cela peut être interprété par le fait que le taux de succès de transmission est élevé grâce au mécanisme de *flooding* d'*Epidemic* quand le ratio de nœuds égoïstes est petit, mais par contre si le ratio est grand, un grand nombre de paquets d'*Epidemic* sont détournés par des nœuds égoïstes et seulement une petite partie des données sont transférées entre une partie des nœuds avec un haut niveau de confiance, et cela diminue rapidement le taux de succès de transmission. En raison de leur capacité à prédire le nœud de destination de *Prophet*, associée à la mise en place de l'établissement de la relation de confiance basée sur le contexte social, la vitesse du déclin du taux de succès de transmission devient plus lente avec l'augmentation du ratio de nœuds égoïstes. *Spray-and-Waits* est l'algorithme amélioré basé sur *Epidemic*, et donc sa performance se situe entre *Prophet* et *Epidemic*.

Les figures 7-12, 7-13, 7-14 comparent respectivement dans les couvertures réseaux de 1000m*1000m, 2000m*2000m, 3000m*3000m et avec des ratio

différents des nœuds égoïstes, le temps de délai des trois modes de transmission *Epidemic*, *Prophet* et *Spray-and-Waits* en l'absence et avec notre algorithme de confiance.

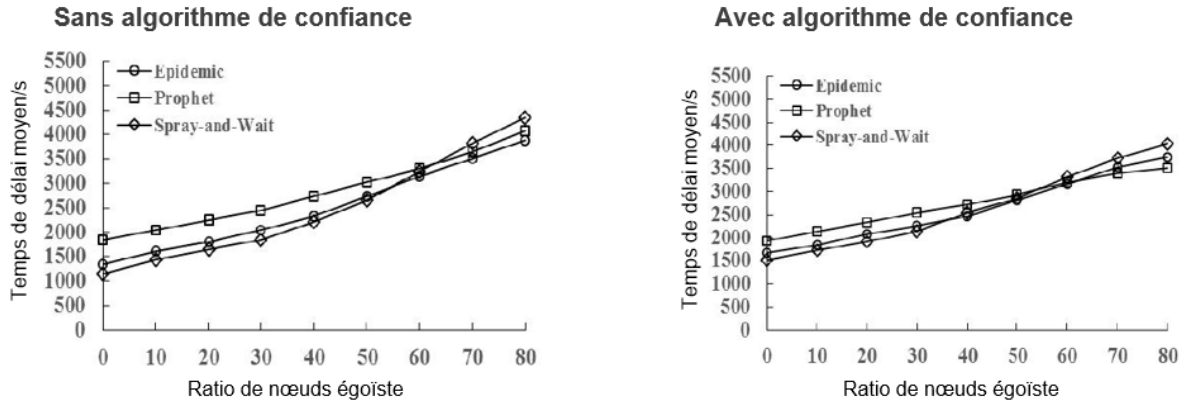


Figure 7-12 : Temps de délai de transmission dans la zone 1000m*1000m

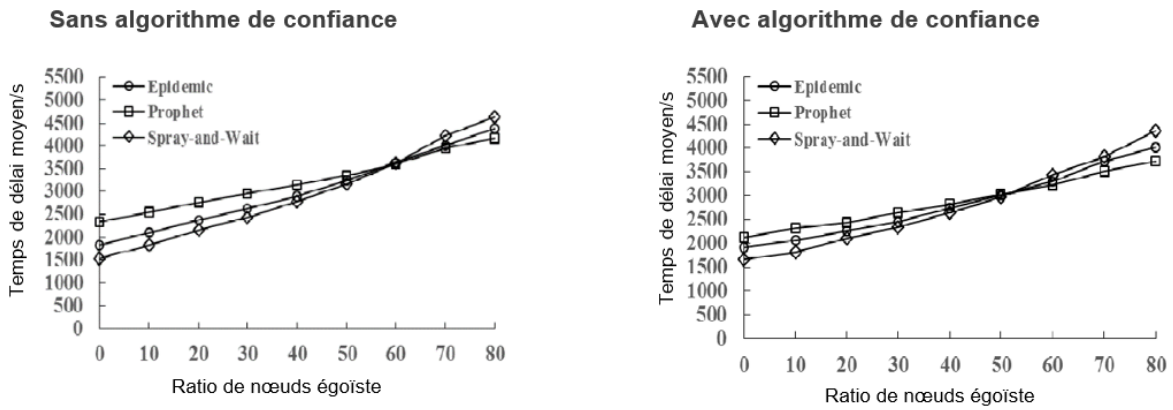


Figure 7-13 : Taux de succès de transmission dans la zone 2000m*2000m

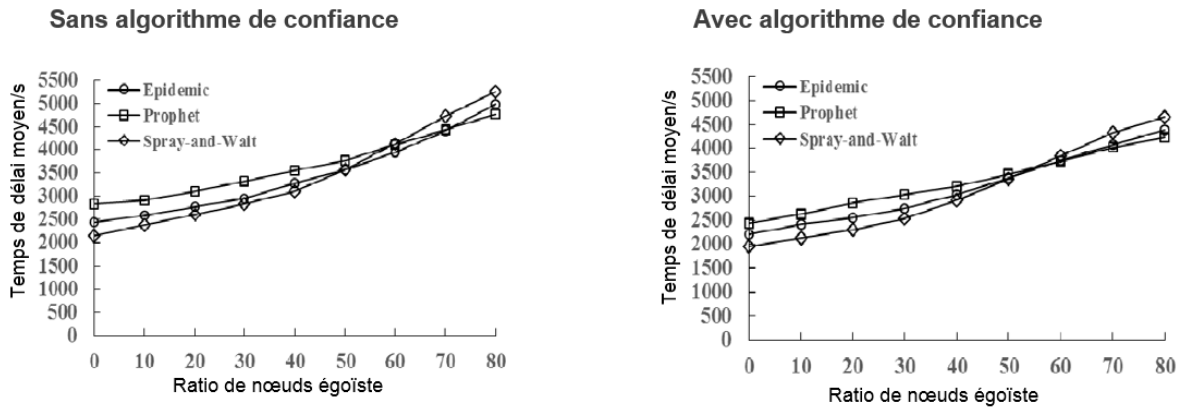


Figure 7-14 : Taux de succès de transmission dans la zone 3000m*3000m

Le temps de délai moyen augmente progressivement avec l'augmentation de la couverture de réseau selon les trois figures ci-dessus. En l'absence d'algorithme de confiance, dans les trois modes de transmission, le temps de délai moyen augmente suivant l'augmentation du ratio de nœuds égoïste. En ajoutant notre algorithme de confiance en revanche, le temps de délai moyen de transmission est réduit effectivement, c'est-à-dire que la tendance à l'augmentation du temps de délai moyen est bien atténuée.

En comparant les temps de délai moyen des trois modes de transmission sur la même surface de réseau, si le ratio de nœuds égoïstes est plus bas que 50% ou 60%, le temps de délai moyen de *Spray-and-Waits* est plus court que celui d'*Epidemic* et de *Prophet*. Pour *Prophet*, parce que le nombre de paquets transféré est relativement faible, sous l'influence des nœuds égoïstes, son temps de délai moyen est plus long qu'*Epidemic* et *Spray-and-Waits*. Si le ratio de nœuds égoïstes est plus haut que 50% ou 60%, *Prophet* réalise la transmission en profitant de sa propre prévision et l'algorithme de transmission basé sur la confiance sociale binaire, dans la situation où le ratio de nœuds égoïstes est élevé, le temps de délai moyen est plus court que les deux autres.

Ce chapitre a présenté dans un premier temps la structure et le principe de fonctionnement de l'outil de simulation ONE. Et puis, en fonction des besoins des deux expériences, différents paramètres des scénarios ont été mis en place

pour effectuer nos simulations. Les résultats expérimentaux ont été respectivement décrits et analysés, et la cause des phénomènes a été résumée. Enfin, les résultats ont montré que l'algorithme de transmission de confiance proposé peut améliorer le taux de succès de transmission et réduire les temps de délai des données. En utilisant le contexte social dans le réseau opportuniste, la sécurité des données du réseau est bien assurée en tenant compte de la performance du réseau.

Chapitre 8. Conclusion et perceptives

Dans le but de favoriser l'émergence des environnements de *Pervasive computing*, où les individus sont aidés dans leurs tâches quotidiennes par des technologies qui s'exécutent en tâches de fond, de manière invisible mais significativement utile, le domaine de recherche *Context awareness* joue un rôle important pour permettre de fournir des connaissances sur l'environnement et les utilisateurs que les applications intelligentes peuvent utiliser directement pour offrir des informations ou des services selon la situation constatée. En complément des contextes physiques les plus utilisés tels que l'emplacement, le temps et l'activité, cette thèse propose d'utiliser les contextes sociaux pour étendre le service rendu par les applications sensible au contexte.

Dans la partie état de l'art, nous avons d'abord présenté les connaissances de base sur le contexte et la conscience du contexte (*context awareness*). Nous avons ensuite pointé les limitations du contexte de nature physique et introduit le contexte social, y compris l'opportunité et les challenges de la vision sociale, la définition, la classification et différentes utilisations des contextes sociaux. Nous avons identifié la notion de contexte multidimensionnel pour expliquer un des défis importants de notre recherche. Le domaine d'application, que nous avons choisi d'enrichir en utilisant les contextes sociaux, est le réseau opportuniste. Nous en présentons le concept et les caractéristiques, les stratégies de transmission, ainsi que les problèmes de sécurité.

Ensuite, Nous avons proposé deux modèles pour modéliser respectivement le poids de la relation du réseautage social homogène et du réseau social hétérogène correspondant à la vie réelle. Nous avons synthétisé des contextes sociaux non seulement liés à la familiarité mais aussi lié à la similitude des communautés statiques et dynamiques. Avec ces deux modèles, nous pouvons modéliser les contextes sociaux des utilisateurs dans le réseau social multidimensionnel. Le premier modèle s'adapte à un réseautage social ; les informations telles que le nombre d'interactions, les propriétés de base et les notes d'articles des utilisateurs sont prises en compte. Après les avoir synthétisés, nous avons détaillé une méthode pour calculer le poids de la

relation basée sur les graphes. Pour les réseaux sociaux à large échelle de la vie réelle, nous proposons un deuxième modèle basé sur les ontologies qui est réalisé d'une manière similaire au réseau social de recherche. Nous utilisons les données réelles recueillies sur les réseautages sociaux en ligne pour conduire nos expérimentations et analysons les résultats en les comparant avec l'approche existante.

Enfin, nous avons présenté deux algorithmes utilisant des contextes sociaux pour améliorer la stratégie de transmission des données dans le réseau opportuniste, et particulièrement en contre-mesure aux nœuds égoïstes. Nous avons présenté l'utilisation de la confiance dans le domaine du réseau opportuniste. Nous avons également utilisé des contextes sociaux basés sur la familiarité et la similitude : un algorithme de transmission de confiance basé sur la confiance sociale binaire et un algorithme de transmission de confiance basé sur la similitude de communauté dynamique des nœuds. Les simulations de scénarios réels confirment les avantages de l'introduction des contextes sociaux par comparaison également avec d'autres algorithmes de transmission bien connus dans la littérature sur certains paramètres tels que le taux de succès et le délai de transmission des données.

En résumé, cette thèse étudie de manière approfondie les effets des contextes sociaux des utilisateurs sur notre ensemble de données en ligne et introduit de plus ces effets pour améliorer efficacement une application de réseaux opportuniste. Mais notre travail comporte néanmoins encore plusieurs limitations. Dans ce qui suit, certaines orientations de recherche qui nous paraissent prometteuses sont discutées.

En premier lieu, cette thèse n'a utilisé qu'une partie des informations qui peut être consacrée à inférer le contexte social sur le réseautage social ou toute autre source en ligne disponible. Nous avons tiré parti des propriétés de base statiques et du nombre d'interactions (messages laissés, messagerie instantanée, téléphone en ligne, etc.) sur des réseaux sociaux tels que Facebook et Renren. Mais en dehors de l'aspect numérique, l'activité d'interaction en ligne contient beaucoup de connaissances, et des actions de l'utilisateur telles que le post, le commentaire et *like*, qui peuvent être utilisées pour mesurer plus précisément la relation entre les utilisateurs et enrichir plus

efficacement les applications et les services intelligents. De plus, les contenus inscrits dans les comportements des utilisateurs sont aussi intéressants pour produire un raisonnement plus poussé. En outre, de nombreuses autres plateformes florissantes accumulent des données dans leurs propres spécialités. Par exemple, Twitter diffuse des informations dans un message court de 140 caractères sur l'emplacement où il a été envoyé, Foursquare partage et rassemble un grand nombre de check-ins. Ces informations massives qui se développent et laissent des traces numériques peuvent représenter différents patterns de comportement humains, voire des marqueurs de personnalité. Une variété de règles d'inférence concernant la familiarité et la similitude peut être saisie ou la connaissance du contexte social peut être réinterprétée ou enrichie sous des angles différents. Par conséquent, l'intégration des données étendues à étudier, à analyser et à modéliser serait plus intéressante et permettrait d'enrichir les applications et les services. Bien que la fusion de données puisse enrichir nos sources de données, elle présente également des défis : la manière de collecter et d'unifier les informations de l'utilisateur à partir de plateformes multiples, la quantité et le coût des données à stocker, et la préservation de la vie privée des utilisateurs lors de l'intégration des données.

En second lieu, le nombre total d'utilisateurs dans les réseaux sociaux est dans les faits immenses, sans parler de la quantité de connexions, d'interactions et de contenu générés par l'utilisateur. Cependant, cette thèse n'a pas pris en compte le problème de passage à l'échelle et n'a évalué nos algorithmes et nos approches proposés que sur des ensembles de données d'échelle limitée. Comment appliquer notre approche à des milliards d'utilisateurs et le contexte social qui leur est associé ? Cela pourrait engendrer de nombreux problèmes difficiles à traiter, et il faudra probablement des techniques issues d'autres domaines de recherche tels que les systèmes distribués et le calcul parallèle. Par conséquent, une autre direction pour nos futures recherches est d'adapter l'approche que nous proposons à des approches évolutives qui peuvent être exploitées dans des plateformes réelles constituées d'un nombre très conséquent d'utilisateurs. Il reste néanmoins à déterminer s'il pourrait y avoir un intérêt à étudier l'influence du contexte social dans un environnement d'interactions envisagé à l'échelle planétaire.

Troisièmement, il existe encore un large spectre d'autres applications qui peuvent être enrichies par l'utilisation des contextes sociaux, depuis l'amélioration des services d'ICT et les applications en ligne jusqu'à la réalisation d'environnements dit « intelligents » tels que le véhicule intelligent, la maison intelligente, les transports intelligents et la ville intelligente.

En outre, nous nous sommes intéressés à l'interaction liée au contexte social dans le cas d'interactions dyadiques (deux utilisateurs) ; mais il pourrait être intéressant, pour certaines applications, d'étudier le cas d'interactions entre plusieurs utilisateurs (>à 2). Par exemple, on pourrait étudier le poids de la relation entre utilisateurs et villes ou pays, ce qui permettrait d'exploiter de grands volumes de données issues des réseaux sociaux afin d'esquisser automatiquement une carte culturelle mondiale (qui est traditionnellement créée par des enquêtes), dans le but de refléter les spécificités communautaires et de distinguer entre différentes régions du monde.

Références

- [1] Dourish, Paul. "Seeking a foundation for context-aware computing." *Human-Computer Interaction* 16.2-4 (2001): 229-241.
- [2] Abowd, Gregory D., and Elizabeth D. Mynatt. "Charting past, present, and future research in ubiquitous computing." *ACM Transactions on Computer-Human Interaction (TOCHI)* 7.1 (2000): 29-58.
- [3] Dix, Alan, et al. "Exploiting space and location as a design framework for interactive mobile systems." *ACM Transactions on Computer-Human Interaction (TOCHI)* 7.3 (2000): 285-321.
- [4] Schilit, Bill N., and Marvin M. Theimer. "Disseminating active map information to mobile hosts." *Network, IEEE* 8.5 (1994): 22-32.
- [5] Dey, Anind K. "Context-aware computing: The CyberDesk project." *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments*. 1998.
- [6] Abowd, Gregory D., et al. "Towards a better understanding of context and context-awareness." *Handheld and ubiquitous computing*. Springer Berlin Heidelberg, 1999.
- [7] Salber, Daniel, Anind K. Dey, and Gregory D. Abowd. "Ubiquitous computing: Defining an hci research agenda for an emerging interaction paradigm." (1998).
- [8] Dey, Anind K., Gregory D. Abowd, and Andrew Wood. "CyberDesk: A framework for providing self-integrating context-aware services." *Knowledge-Based Systems* 11.1 (1998): 3-13.
- [9] Dourish, Paul. "What we talk about when we talk about context." *Personal and ubiquitous computing* 8.1 (2004): 19-30.
- [10] *Cognition, Everyday*. "Its Development in Social Context." Cambridge, MA: Harvard (1984).
- [11] Glucksberg, Sam, Roger J. Kreuz, and Susan H. Rho. "Context can constrain lexical access: Implications for models of language comprehension." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12.3 (1986): 323.
- [12] Hupbach, Almut, Rebecca Gomez, and Lynn Nadel. "Episodic memory updating: The role of context familiarity." *Psychonomic bulletin & review* 18.4 (2011): 787-797.
- [13] Seifert, C. "Situated cognition and learning." *The MIT encyclopedia of the cognitive sciences*. The MIT Press, Cambridge (1999).
- [14] Dey, Anind K. "Understanding and using context." *Personal and ubiquitous computing* 5.1 (2001): 4-7.
- [15] McCarthy, John. "Notes on formalizing context." (1993).
- [16] Bazire, Mary, and Patrick Brézillon. "Understanding context before using it." *International and Interdisciplinary Conference on Modeling and Using Context*. Springer Berlin Heidelberg, 2005.
- [17] Brézillon, Patrick. "Representation of procedures and practices in contextual graphs." *The Knowledge Engineering Review* 18.02 (2003): 147-174.

- [18] Brown, Peter J. "The stick-e document: a framework for creating context-aware applications." *ELECTRONIC PUBLISHING-CHICHESTER*- 8 (1995): 259-272.
- [19] Ryan, Nick, Jason Pascoe, and David Morse. "Enhanced reality fieldwork: the context aware archaeological assistant." *Bar International Series* 750 (1999): 269-274.
- [20] Hull, Richard, Philip Neaves, and James Bedford-Roberts. "Towards situated computing." *Wearable Computers, 1997. Digest of Papers., First International Symposium on. IEEE, 1997.*
- [21] Drost, Chiel. "Privacy in context-aware systems." University of Twente, Enschede, Netherlands, Department of Informatics, Federal University of Espirito Santo, Vitoria, Brazil. Retrieved November 5 (2004): 2011.
- [22] Zimmermann, Andreas, Andreas Lorenz, and Reinhard Oppermann. "An operational definition of context." *International and Interdisciplinary Conference on Modeling and Using Context. Springer Berlin Heidelberg, 2007.*
- [23] Nadoveza, Drazen, and Dimitris Kiritsis. "Concept for Context-Aware Manufacturing Dashboard Applications." *Manufacturing Modelling, Management, and Control. Vol. 7. No. 1. 2013.*
- [24] Olivier, Patrick, et al. "Ambient kitchen: designing situated services using a high fidelity prototyping environment." *Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments. ACM, 2009.*
- [25] Ye, Juan, Simon Dobson, and Susan McKeever. "Situation identification techniques in pervasive computing: A review." *Pervasive and mobile computing* 8.1 (2012): 36-66.
- [26] Schilit, Bill, Norman Adams, and Roy Want. "Context-aware computing applications." *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on. IEEE, 1994.*
- [27] Nagel, Kristine Susanne. "Using availability indicators to enhance context-aware family communication applications." (2006).
- [28] Zulkernain, Sina, Praveen Madiraju, and Sheikh Iqbal Ahamed. "A context-aware cost of interruption model for mobile devices." *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on. IEEE, 2011.*
- [29] Liang, Guanqing, and Jiannong Cao. "Social context-aware middleware: A survey." *Pervasive and Mobile Computing* 17 (2015): 207-219.
- [30] Weiser, Mark. "Ubiquitous computing." *Computer* 10 (1993): 71-72.
- [31] Page, Stanley R., et al. "User customization of a word processor." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 1996.*
- [32] Farrell, Stephen, et al. "Information programming for personal user interfaces." *Proceedings of the 7th international conference on Intelligent user interfaces. ACM, 2002.*
- [33] Delphi Group, " Perspectives on Information Retrieval. " Boston, Mass., Delphi Group (2002)

- [34] Emrich A, Chapko A, and Werth D. "Context-aware recommendations on mobile services: the m;Ciudad approach." Proceedings of the 4th European conference on Smart sensing and context, Berlin, Heidelberg, Springer-Verlag, 2009. 107-120.
- [35] Gulliver, Stephen R., et al. "A context-aware Tour Guide: User implications." Mobile Information Systems 3.2 (2007): 71. Ludford, Pamela J., et al. "Because I carry my cell phone anyway: functional location-based reminder applications." Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM, 2006. 88.
- [36] Lin, Chi-Yi, and Ming-Tze Hung. "A location-based personal task reminder for mobile users." Personal and ubiquitous computing 18.2 (2014): 303-314.
- [37] Bader, Roland, et al. "Context-aware POI recommendations in an automotive scenario using multi-criteria decision making methods." Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation. ACM, 2011.
- [38] Liu, Qi, et al. "A survey of context-aware mobile recommendations." International Journal of Information Technology & Decision Making 12.01 (2013): 139-172.
- [39] Weiser, Mark. "The computer for the 21st century." Scientific american 265.3 (1991): 94-104.
- [40] York, Judy, and Parag C. Pendharkar. "Human-computer interaction issues for mobile computing in a variable work context." International Journal of Human-Computer Studies 60.5 (2004): 771-797.
- [41] Augusto, Juan Carlos, Hideyuki Nakashima, and Hamid Aghajan. "Ambient intelligence and smart environments: A state of the art." Handbook of ambient intelligence and smart environments. Springer US, 2010. 3-31.
- [42] Sundmaeker, Harald, et al. "Vision and challenges for realising the Internet of Things." Cluster of European Research Projects on the Internet of Things, European Commission (2010).
- [43] Chen, Edwin, et al. "Analysis of IPv6 Network Communication Using Simulation." Research and Development, 2006. SCORed 2006. 4th Student Conference on. IEEE, 2006.
- [44] Perera, Charith, et al. "Context aware computing for the internet of things: A survey." IEEE Communications Surveys & Tutorials 16.1 (2014): 414-454.
- [45] Held, Albert, Sven Buchholz, and Alexander Schill. "Modeling of context information for pervasive computing applications." Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics. 2002.
- [46] Henriksen, Karen, Jadwiga Indulska, and Andry Rakotonirainy. "Modeling context information in pervasive computing systems." Pervasive Computing. Springer Berlin Heidelberg, 2002. 167-180.
- [47] Henriksen, Karen, and Jadwiga Indulska. "Developing context-aware pervasive computing applications: Models and approach." Pervasive and mobile computing 2.1 (2006): 37-64.

- [48] Bettini, Claudio, et al. "A survey of context modelling and reasoning techniques." *Pervasive and Mobile Computing* 6.2 (2010): 161-180.
- [49] Hoareau, Christian, and Ichiro Satoh. "Modeling and processing information for context-aware computing: A survey." *New Generation Computing* 27.3 (2009): 177-196.
- [50] Studer, Rudi, V. Richard Benjamins, and Dieter Fensel. "Knowledge engineering: principles and methods." *Data & knowledge engineering* 25.1 (1998): 161-197.
- [51] Ejigu, Dejene, Marian Scuturici, and Lionel Brunie. "An ontology-based approach to context modeling and reasoning in pervasive computing." *Pervasive Computing and Communications Workshops, 2007. PerCom Workshops' 07. Fifth Annual IEEE International Conference on. IEEE, 2007.*
- [52] Wang, Xiao Hang, et al. "Ontology based context modeling and reasoning using OWL." *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on. IEEE, 2004.*
- [53] Preuveneers, Davy, et al. "Towards an extensible context ontology for ambient intelligence." *European Symposium on Ambient Intelligence. Springer Berlin Heidelberg, 2004.*
- [54] McGuinness, Deborah L., and Frank Van Harmelen. "OWL web ontology language overview." *W3C recommendation 10.10 (2004): 2004.*
- [55] Chen, Harry, et al. "Soupa: Standard ontology for ubiquitous and pervasive applications." *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on. IEEE, 2004.*
- [56] Mamei, Marco, and Radhika Nagpal. "Macro programming through Bayesian Networks: Distributed inference and anomaly detection." *Pervasive Computing and Communications, 2007. PerCom'07. Fifth Annual IEEE International Conference on. IEEE, 2007.*
- [57] Ranganathan, Anand, and Roy H. Campbell. "An infrastructure for context-awareness based on first order logic." *Personal and Ubiquitous Computing* 7.6 (2003): 353-364.
- [58] Zhang, Daqiang, et al. "Extended dempster-shafer theory in context reasoning for ubiquitous computing environments." *Computational Science and Engineering, 2009. CSE'09. International Conference on. Vol. 2. IEEE, 2009.*
- [59] Nieto, JA Fraile, ME Beato Gutiérrez, and B. Pérez Lancho. "Developing home care intelligent environments: from theory to practice." *7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS 2009). Springer Berlin Heidelberg, 2009.*
- [60] Zhang, Daqiang, et al. "Context reasoning using extended evidence theory in pervasive computing environments." *Future Generation Computer Systems* 26.2 (2010): 207-216.
- [61] Haghighi, Pari Delir, et al. "Reasoning about context in uncertain pervasive computing environments." *Smart Sensing and Context. Springer Berlin Heidelberg, 2008. 112-125.*

- [62] Eriksson, Jakob, et al. "The pothole patrol: using a mobile sensor network for road surface monitoring." Proceedings of the 6th international conference on Mobile systems, applications, and services. ACM, 2008.
- [63] Campbell, Andrew T., et al. "The rise of people-centric sensing." IEEE Internet Computing 12.4 (2008): 12-21.
- [64] Modes, Beyond. "From context awareness to socially aware computing." (2012).
- [65] Biamino, Giulia. "Modeling social contexts for pervasive computing environments." Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on. IEEE, 2011.
- [66] Arnaboldi, Valerio, Marco Conti, and Franca Delmastro. "CAMEO: A novel context-aware middleware for opportunistic mobile social networks." *Pervasive and Mobile Computing* 11 (2014): 148-167.
- [67] Quan, wu et al. "Online Social Networks and Social Network Services: A Technical Survey." Pervasive Communications Handbook. Chapter 18, 2011.
- [68] Tan, Wei, et al. "Social-Network-Sourced Big Data Analytics." IEEE Internet Computing 17.5 (2013): 62-69.
- [69] Schuster, Daniel, et al. "Pervasive social context: Taxonomy and survey." ACM Transactions on Intelligent Systems and Technology (TIST) 4.3 (2013): 46.
- [70] Yu, Xiao, et al. "Geo-friends recommendation in gps-based cyber-physical social network." Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, 2011.
- [71] Zhuang, Jinfeng, et al. "Modeling social strength in social media community via kernel-based learning." Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011.
- [72] Wirz, Martin, Daniel Roggen, and Gerhard Tröster. "Decentralized detection of group formations from wearable acceleration sensors." Computational Science and Engineering, 2009. CSE'09. International Conference on. Vol. 4. IEEE, 2009.
- [73] Chen, Guanling, and Faruq Rahman. "Analyzing privacy designs of mobile social networking applications." Embedded and Ubiquitous Computing, 2008. EUC'08. IEEE/IFIP International Conference on. Vol. 2. IEEE, 2008.
- [74] Shi, Jing, et al. "Prisense: privacy-preserving data aggregation in people-centric urban sensing systems." INFOCOM, 2010 Proceedings IEEE. IEEE, 2010.
- [75] Prell, Christina. Social network analysis: History, theory and methodology. Sage, 2011.
- [76] Mislove, Alan, Krishna P. Gummadi, and Peter Druschel. "Exploiting social networks for internet search." 5th Workshop on Hot Topics in Networks (HotNets06). Citeseer. 2006.
- [77] Li, Jinyang, and Frank Dabek. "F2F: Reliable Storage in Open Networks." IPTPS. 2006.
- [78] Chen, Guanling, and David Kotz. A survey of context-aware mobile computing research. Vol. 1. No. 2.1. Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, 2000.

- [79] Brown, Peter J., and Gareth JF Jones. "Context-aware retrieval: Exploring a new environment for information retrieval and information filtering." *Personal and Ubiquitous Computing* 5.4 (2001): 253-263.
- [80] Pascoe, Jason. "Adding generic contextual capabilities to wearable computers." *Wearable Computers, 1998. Digest of Papers. Second International Symposium on. IEEE, 1998.*
- [81] Kourtellis, Nicolas, et al. "Prometheus: User-controlled p2p social data management for socially-aware applications." *Proceedings of the ACM/IFIP/USENIX 11th International Conference on Middleware. Springer-Verlag, 2010.*
- [82] Conti, Marco, et al. "Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber-physical convergence." *Pervasive and Mobile Computing* 8.1 (2012): 2-21.
- [83] Hui, Pan, Jon Crowcroft, and Eiko Yoneki. "Bubble rap: Social-based forwarding in delay-tolerant networks." *Mobile Computing, IEEE Transactions on* 10.11 (2011): 1576-1589.
- [84] De Rango, Floriano, and Filippo Monteverdi. "Social and dynamic graph-based scalable routing protocol in a DTN network." *Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2012 International Symposium on. IEEE, 2012.*
- [85] Pelusi, Luciana, Andrea Passarella, and Marco Conti. "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks." *IEEE Communications Magazine* 44.11 (2006): 134-141.
- [86] Distl, Bernhard, and Theus Hossmann. "Privacy in opportunistic network contact graphs." *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2014 IEEE 15th International Symposium on a. IEEE, 2014.*
- [87] Galati, Adriano. "Delay tolerant network." (2010).
- [88] Chen, Ling-Jyh, et al. "A hybrid routing approach for opportunistic networks." *Proceedings of the 2006 SIGCOMM workshop on Challenged networks. ACM, 2006.*
- [89] Musolesi, Mirco, Stephen Hailes, and Cecilia Mascolo. "Adaptive routing for intermittently connected mobile ad hoc networks." *World of wireless mobile and multimedia networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a. IEEE, 2005.*
- [90] Hui, Pan, Jon Crowcroft, and Eiko Yoneki. "Bubble rap: Social-based forwarding in delay-tolerant networks." *Mobile Computing, IEEE Transactions on* 10.11 (2011): 1576-1589.
- [91] Bigwood, Greg, and Tristan Henderson. "Bootstrapping opportunistic networks using social roles." *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a. IEEE, 2011.*
- [92] Bakhouya, Mohamed, Jaafar Gaber, and Pascal Lorenz. "An adaptive approach for information dissemination in vehicular ad hoc networks." *Journal of Network and Computer Applications* 34.6 (2011): 1971-1978.

- [93] El Fawal, Alaeddine, Jean-Yves Le Boudec, and Kave Salamatian. "Vulnerabilities in epidemic forwarding." *World of Wireless, Mobile and Multimedia Networks*, 2007. WoWMoM 2007. IEEE International Symposium on a. IEEE, 2007.
- [94] Tan, Yi, et al. "Spectrum stealing via Sybil attacks in DSA networks: Implementation and defense." *Communications (ICC), 2011 IEEE International Conference on*. IEEE, 2011.
- [95] Jedari, Behrouz, and Feng Xia. "A survey on routing and data dissemination in opportunistic mobile social networks." *arXiv preprint arXiv:1311.0347* (2013).
- [96] Zhang, Sihai, et al. "Evolutionary reputation model for node selfishness resistance in opportunistic networks." *Concurrency and computation: Practice and experience* 24.11 (2012): 1261-1269.
- [97] Nelson, Samuel C., et al. "Encounter: based routing in DTNs." *ACM SIGMOBILE Mobile Computing and Communications Review* 13.1 (2009): 56-59.
- [98] Parris, Iain, and Tristan Henderson. "The impact of location privacy on opportunistic networks." *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a. IEEE, 2011*.
- [99] Shikfa, Abdullatif, Melek Onen, and Refik Molva. "Privacy in content-based opportunistic networks." *2009 International Conference on Advanced Information Networking and Applications Workshops*. IEEE, 2009.
- [100] Daly, Elizabeth M., and M. Haahr. "Social Network Analysis for information flow in disconnected Delay-Tolerant Manets." *IEEE Transactions on Mobile Computing* 8.5 (2009): 606.
- [101] Cho, Jin-Hee, Ananthram Swami, and Ing-Ray Chen. "A survey on trust management for mobile ad hoc networks." *Communications Surveys & Tutorials, IEEE* 13.4 (2011): 562-583.
- [102] Trifunovic, Sacha, Franck Legendre, and Carlos Anastasiades. "Social trust in opportunistic networks." *INFOCOM IEEE Conference on Computer Communications Workshops, 2010. IEEE, 2010*.
- [103] Douceur, John R. "The sybil attack." *International Workshop on Peer-to-Peer Systems*. Springer Berlin Heidelberg, 2002.
- [104] Mtibaa, Abderrahmen, and Khaled A. Harras. "Social-Based Trust Mechanisms in Mobile Opportunistic Networks." *Proc. IEEE SIMNA* (2011).
- [105] Qin, Yujue, et al. "A context-aware do-not-disturb service for mobile devices." *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2014.
- [106] Simpson, Jeffry A., and William Steven Ed Rholes. *Attachment theory and close relationships*. Guilford Press, 1998.
- [107] Adamic, Lada A., and Eytan Adar. "Friends and neighbors on the web." *Social networks* 25.3 (2003): 211-230.
- [108] Symeonidis, Panagiotis, Alexandros Nanopoulos, and Yannis Manolopoulos. "A unified framework for providing recommendations in social tagging systems

- based on ternary semantic analysis." IEEE Transactions on Knowledge and Data Engineering 22.2 (2010): 179-192.
- [109] https://fr.wikipedia.org/wiki/Th%C3%A9orie_des_graphes
- [110] Yao, Weilong, et al. "A Graph-based model for context-aware recommendation using implicit feedback data." World Wide Web 18.5 (2015): 1351-1371.
- [111] https://fr.wikipedia.org/wiki/Six_degr%C3%A9s_de_s%C3%A9paration
- [112] Ahlgren, Per, Bo Jarneving, and Ronald Rousseau. "Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient." Journal of the American Society for Information Science and Technology 54.6 (2003): 550-560.
- [113] <http://www.renren.com/>
- [114] Golbeck, Jennifer Ann. "Computing and applying trust in web-based social networks." (2005).
- [115] Buffa, Michel, et al. "ISICIL: Semantics and Social Networks for Business Intelligence." International Conference on Current Trends in Theory and Practice of Computer Science. Springer Berlin Heidelberg, 2013.
- [116] Mika, Peter. "Social networks and the semantic web: The next challenge." (2005): 82-85.
- [117] Li, Jianguo, et al. "Semantic description of scholar-oriented social network cloud." The Journal of Supercomputing 65.1 (2013): 410-425.
- [118] <http://aclanthology.info/>
- [119] <https://aminer.org/>
- [120] Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." IEEE transactions on knowledge and data engineering 17.6 (2005): 734-749
- [121] Yaofu Cao, Dominique Gaïti, Pascal Salembier, Prise en compte du contexte pour la fiabilité des réseaux opportunistes et la gestion de la privacy, la 30ème édition du congrès DNAC, Paris, France, Jan 2017
- [122] O'Donovan, John, and Barry Smyth. "Trust in recommender systems." Proceedings of the 10th international conference on Intelligent user interfaces. ACM, 2005.
- [123] Hoffman, Robert R. "Whom (or what) do you (mis) trust?: Historical reflections on the psychology and sociology of information technology." Human Interaction with Complex Systems, 1998. Proceedings., Fourth Annual Symposium on. IEEE, 1998.
- [124] Hall, Simin, and William McQuay. "Review of trust research from an interdisciplinary perspective-psychology, sociology, economics, and cyberspace." Proceedings of the IEEE 2010 National Aerospace & Electronics Conference. IEEE, 2010.
- [125] Blaze, Matt, Joan Feigenbaum, and Jack Lacy. "Decentralized trust management." Security and Privacy, 1996. Proceedings., 1996 IEEE Symposium on. IEEE, 1996.

- [126] Cho, Jin-Hee, Ananthram Swami, and Ray Chen. "A survey on trust management for mobile ad hoc networks." *IEEE Communications Surveys & Tutorials* 13.4 (2011): 562-583.
- [127] Blaze, Matt, et al. "Dynamic trust management." *Computer* 42.2 (2009): 44-52.
- [128] Fajjari, Ilhem, et al. "An optimised dynamic resource allocation algorithm for Cloud's backbone network." *Local Computer Networks (LCN), 2012 IEEE 37th Conference on. IEEE, 2012.*
- [129] Pajevic, Ljubica, Gunnar Karlsson, and Ólafur Helgason. "Performance of Epidemic Content Distribution in Opportunistic Networks." *IFIP WG 7.3 Performance 2013 31 st International Symposium on Computer Performance, Modeling, Measurements and Evaluation 2013 Student Poster Abstracts September 24-26, Vienna, Austria. 2013.*
- [130] Spyropoulos, Thrasyvoulos, Konstantinos Psounis, and Cauligi S. Raghavendra. "Spray and wait: an efficient routing scheme for intermittently connected mobile networks." *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking. ACM, 2005.*
- [131] Huang, Ting-Kai, Chia-Keng Lee, and Ling-Jyh Chen. "Prophet+: An adaptive prophet-based routing protocol for opportunistic network." *2010 24th IEEE International Conference on Advanced Information Networking and Applications. IEEE, 2010.*
- [132] Niu, Jianwei, et al. "Predict and spread: An efficient routing algorithm for opportunistic networking." *2011 IEEE Wireless Communications and Networking Conference. IEEE, 2011.*
- [133] Verma, Anshul, and Dr Srivastava. "Integrated routing protocol for opportunistic networks." *arXiv preprint arXiv:1204.1658 (2012).*
- [134] Rothstein, Bo, and Eric M. Uslaner. "All for all: Equality, corruption, and social trust." *World politics* 58.01 (2005): 41-72.
- [135] Zhou, W-X., et al. "Discrete hierarchical organization of social group sizes." *Proceedings of the Royal Society of London B: Biological Sciences* 272.1561 (2005): 439-444.
- [136] Gallos, Lazaros K., Hernán A. Makse, and Mariano Sigman. "A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks." *Proceedings of the National Academy of Sciences* 109.8 (2012): 2825-2830.
- [137] Zong Z. Study of grid trust model nased on virtual domain. *Applied Mechanics and Materials*, 2013, 385: 1059-1065
- [138] Ahmed, Ahmedin Mohammed, et al. "Social community-partition aware replica allocation in ad-hoc social networks." *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCoM), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, 2013.*

- [139] Golbeck, Jennifer, and James Hendler. "Inferring binary trust relationships in web-based social networks." *ACM Transactions on Internet Technology (TOIT)* 6.4 (2006): 497-529.
- [140] Keranen, Ari. "Opportunistic network environment simulator." Special Assignment report, Helsinki University of Technology, Department of Communications and Networking (2008).

Yaofu CAO

Doctorat : Ingénierie Sociotechnique des Connaissances, des Réseaux et du Développement Durable

Année 2017

Modélisation du contexte social : application aux réseaux opportunistes

Cette thèse traite de l'adaptation automatique des applications sensibles au contexte par l'utilisation d'informations liées à l'environnement social des utilisateurs afin d'enrichir le service rendu par les applications. Pour cela, notre contribution s'articule autour de la modélisation multidimensionnelle des différents niveaux de contextes sociaux, notamment le poids de la relation entre les acteurs. Plus spécifiquement, nous synthétisons des contextes sociaux non seulement liés à la familiarité mais aussi liés à la similitude des communautés statiques et dynamiques. Deux modèles basés respectivement sur les graphes et les ontologies sont proposés afin de satisfaire l'hétérogénéité des réseaux sociaux de la vie réelle. Nous utilisons les données réelles recueillies sur les réseautages sociaux en ligne pour conduire nos expérimentations et analysons les résultats en vérifiant l'efficacité de ces modèles. En parallèle nous traitons le point de vue de l'application, et nous présentons deux algorithmes utilisant des contextes sociaux pour améliorer la stratégie de transmission des données dans le réseau opportuniste, et particulièrement la contre-mesure aux nœuds égoïstes. Les simulations des scénarios réels confirment les avantages liés à l'introduction des contextes sociaux, en termes de taux de succès et de délais de transmission. Nous effectuons une comparaison avec d'autres algorithmes de transmission traditionnellement décrits dans la littérature pour compléter notre démonstration.

Mots clés : contexte (linguistique) - prise de conscience - contexte (linguistique), aspect social - réseaux sociaux - systèmes de communication sans fil - données, transmission.

Social Context Modeling: Application to Opportunistic Networks

This thesis deals with the dynamic adaptation of context-aware applications using information related to the social environment of users to enrich the service rendered by the applications. To achieve this goal our contribution mobilizes multi-dimensional modeling of the different levels of social contexts, especially the weight of the relationship between the actors. Particularly, we synthesize not only social contexts related to familiarity but also social contexts reasoned from the similarity of static and dynamic communities. Two models based on respectively graphs and ontologies are proposed in order to satisfy the heterogeneity of the social networks in real life. We use the actual data gathered on online social networking services for conducting experiments and the results are analyzed by checking the effectiveness of the models. In parallel we consider the point of view of the application, and we present two algorithms using social contexts to improve the strategy of transmission of data in the opportunistic network, particularly countermeasure against selfish nodes. The simulations of real scenarios confirm the advantages of introducing social contexts in terms of success rate and delay of transmission. We carry out a comparison with other popular transmission algorithms in the literature.

Keywords: context (linguistics) - awareness - context (linguistics), social aspects - social networks - wireless communication systems - data transmission systems.

Thèse réalisée en partenariat entre :



Ecole Doctorale "Sciences et Technologies"