

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université A.MIRA-BEJAIA



جامعة بجاية  
Tasdawit n Bgayet  
Université de Béjaïa

Faculté des Sciences Exactes  
Département d'Informatique

# Mémoire

Présenté par

**BENSIMESSAOUD Sihem**

Pour l'obtention du diplôme de Magister

Filière : Informatique

Option : Cloud Computing

Thème

**Préservation de la confidentialité des informations  
personnelles dans la publication des réseaux sociaux**

Soutenu le : 29/09/2016

Devant le Jury composé de :

Nom et Prénom	Grade		
M. TARI Abdelkamel	Professeur	Université de Bejaïa	Président
M. BADACHE Nadjib	Professeur	CERIST, Alger	Rapporteur
M. BOUKERRAM Abdallah	Professeur	Université de Bejaïa	Examineur
Mme BENZAID Chafika	MCA	USTHB, Alger	Examinatrice
Mme BENMEZIANE Souad	CR	CERIST	Invitée

Année Universitaire : 2015/2016

## *Résumé*

Aujourd'hui, de plus en plus les données du réseau social sont rendues publiquement disponibles à des fins d'analyse des données. Bien que cette analyse soit importante pour les chercheurs, il peut y avoir un risque de violation de la vie privée des utilisateurs constituant ce réseau social. Avec peu de connaissances locales sur les individus dans un réseau social, un adversaire peut réaliser différents types d'attaques. Parmi les informations que peut collecter un adversaire sur une victime cible, nous trouvons « le voisinage », c'est à dire, quels sont les voisins de la victime et comment ces voisins sont connectés. Cette information pourra ainsi aider l'adversaire à identifier la personne même si d'autres informations d'identification sont supprimées.

Dans notre travail, nous avons fait une étude détaillée de l'approche d'anonymisation proposée par Zhou et Pei [54] contre les attaques de voisinage et nous avons pu identifier quelques faiblesses concernant l'altération des propriétés structurelles du graphe anonymisé résultant. Ceci nous a conduit à proposer une nouvelle approche d'anonymisation des graphes sociaux à publier tout en maintenant l'utilité des données du réseau qui reflète les propriétés structurelles du graphe original notamment APL. La solution consiste à implémenter le modèle « k-voisinage » pour garantir que tout individu ne peut pas être identifié correctement dans le graphe social anonymisé avec une probabilité supérieure à  $1/k$ , tel que pour chaque sommet appartenant au graphe, il existe au moins  $(k-1)$  autres sommets ayant des voisinages isomorphes.

Le but de l'approche proposée est d'une part de protéger les données publiées contre les attaques de voisinages et de préserver l'utilité du graphe social anonymisé d'autre part.

<b>Introduction générale</b>	<b>1</b>
<b>Chapitre 1 : Les réseaux sociaux et la sécurité</b>	<b>4</b>
1 Introduction.....	4
2 Définition des réseaux sociaux .....	5
3 Les réseaux sociaux en ligne.....	5
4 Les médias sociaux .....	6
5 Les sites de réseautage social.....	7
5.1 Définition d'un site de réseautage social SNS « Social Networking Site » .....	8
5.2 Classification des sites de réseautage social.....	8
5.2.1 SNSs personnels.....	8
5.2.2 SNSs professionnels.....	9
5.2.3 Loisirs et Intérêts.....	10
5.2.4 SNS fonctionnels.....	10
6 Quelques axes de recherche dans le domaine des réseaux sociaux .....	11
7 Modélisation des réseaux sociaux et la théorie des graphes .....	11
8 Les caractéristiques communes des graphes sociaux.....	14
8.1 La distribution des degrés suit une loi de puissance.....	14
8.2 Réseaux petit monde.....	15
8.3 Structure communautaire.....	16
9 L'Analyse des réseaux sociaux.....	17
9.1 Historique et Définition.....	17
9.2 Types d'analyse des réseaux sociaux .....	18
9.3 Applications de l'analyse des réseaux sociaux.....	19
10 La sécurité dans les réseaux sociaux en ligne .....	19
10.1 Menaces de sécurité dans les réseaux sociaux.....	20
10.1.1 L'ingénierie sociale « Social engineering » .....	20
10.1.2 URLs raccourcies malveillants “Malicious Shortened URLs” .....	21
10.1.3 Les logiciels malveillants « Malwares » .....	22
10.1.4 Applications tierces malicieuse « Malicious Third Party Applications ».....	22
10.1.5 Usurpation ou vol d'identité « Identity Theft ».....	23
10.1.6 Les Spams .....	24

## Table des matières

10.1.7	Les faux utilisateurs “Fake Users” .....	24
10.1.8	Redirection d’apparence légitime « Legitimate Look Redirect » .....	24
10.2	Quelques solutions existantes.....	25
10.2.1	Outils de sécurité d’URL.....	25
10.2.2	Ajuster les paramètres de confidentialité .....	25
10.2.3	Ajuster le niveau d'accès des applications .....	25
10.2.4	Partage limité.....	26
10.2.5	Réfléchir à deux fois .....	26
11	Conclusion.....	27
<b>Chapitre 2 : L’anonymisation et les attaques contre les réseaux sociaux anonymisés</b>		<b>28</b>
1	Introduction.....	28
2	La publication des données de réseaux sociaux.....	29
3	La « vie privée » ou la « privacy » dans les réseaux sociaux .....	30
4	Modélisation de la préservation de la vie privée dans les réseaux sociaux .....	31
4.1	Les informations personnelles ou privées dans les réseaux sociaux .....	32
4.2	Les connaissances de base de l’adversaire .....	33
4.3	L’utilité dans les réseaux sociaux.....	36
5	Technique de base de protection de la privacy : l’ <i>anonymisation naïve</i> .....	38
6	Les attaques sur les réseaux sociaux naïvement anonymisés .....	39
6.1	La divulgation d’identité .....	39
6.1.1	Attaques de ré-identification de sommet.....	41
6.1.2	Attaques de réassociation d’informations .....	42
6.2	La divulgation de lien ou la ré-identification de lien.....	43
6.3	La divulgation de contenu .....	43
7	État de l’art sur les approches d’anonymisation proposées dans la littérature .....	44
7.1	Techniques de préservation de privacy dans les bases de données .....	44
7.1.1	Le modèle k-anonymat.....	45
7.1.2	Le modèle l-diversité.....	46
7.2	Challenges (Anonymisation des réseaux sociaux VS anonymisation des bases de données).....	47
7.3	Techniques de préservation de privacy dans les réseaux sociaux .....	48
7.3.1	Le modèle k-candidat .....	48
7.3.2	Le modèle « k-degee » et « KDLD » .....	49

## Table des matières

7.3.3	Approches d'anonymisation des voisinages .....	50
7.3.4	Le modèle k-automorphisme.....	51
7.3.5	Approches d'anonymisation de liens .....	51
7.3.6	Approches d'anonymisation de graphes dynamiques .....	53
8	Conclusion .....	53

### **Chapitre 3 : Proposition d'une nouvelle approche d'anonymisation d'un réseau social 55**

1	Introduction.....	55
2	Illustration des attaques de voisinage .....	56
3	Contribution .....	58
4	Formulation du problème traité .....	59
5	Modélisation d'un réseau social.....	60
6	Quelques concepts de graphes sociaux .....	61
6.1	Voisinage et d-voisinage d'un sommet .....	61
6.2	Composante de voisinage .....	61
6.3	L'isomorphisme de graphe .....	62
6.4	L'isomorphisme de sous-graphe.....	62
6.5	Le k-anonymat et l'anonymat de k-voisinage ou «k-neighborhood ».....	63
7	Utilité du graphe anonymisé et propriétés structurelles.....	63
8	Proposition d'une nouvelle approche d'anonymisation.....	64
8.1	Extraction et représentation des voisinages et des composantes de voisinages .....	67
8.2	Comparaison des voisinages et anonymisation .....	69
8.2.1	Mesure de la qualité de l'anonymisation « Coût d'anonymisation ».....	71
8.2.2	L'Anonymisation de deux voisinages .....	72
8.2.3	Méthode de vérification de la similarité et d'anonymisation de deux composantes .....	74
8.2.4	Méthode d'ajout de nœuds .....	76
9	Conclusion .....	80

### **Chapitre 4: Tests et Expérimentations 81**

1	Introduction.....	81
2	Les outils utilisés.....	81
2.1	Le langage de programmation « JAVA ».....	81

## Table des matières

2.2	Outils de représentation et d'analyse «Gephi ».....	82
2.3	Outil de génération de données synthétiques « Pajek ».....	84
3	La représentation d'un graphe en mémoire .....	84
4	Interfaces de l'outil développé « AnonSN ».....	85
5	Expérimentation.....	89
5.1	Le modèle de référence de Zhou et Pei .....	89
5.2	Environnement expérimental.....	89
5.3	Le jeu de données utilisé .....	89
5.3.1	Ensemble de données synthétiques .....	89
5.3.2	Ensemble de données réelles.....	90
5.4	Exemple d'anonymisation .....	91
5.5	Mesures d'évaluation .....	92
5.5.1	Première expérience : comparaison avec le modèle de référence selon les propriétés structurelles .....	92
5.5.2	Deuxième expérience : Comparaison avec le modèle de référence selon les valeurs de k .....	96
6	Discussion et conclusion.....	100
	<b>Conclusion générale</b>	<b>101</b>
	<b>Bibliographie</b>	<b>103</b>