

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

Ferhat Abbas University – Setif 1
Faculty of Sciences
Department of Computer Science

Community Detection in Signed Social Networks

MASTER THESIS

Specialty : Networks and Distributed Systems

Presented by :

Kerrouche Hamza
Tarchoune Akram

Supervisor :

Dr. Ikram Bouras

Academic Year : 2025 – 2026

Abstract

Social networks often encode not only positive relationships — such as friendship, trust, and collaboration — but also negative ones, including rivalry, distrust, and opposition. Networks in which edges carry an explicit positive or negative sign are known as **signed social networks**. Community detection in such networks aims to partition the node set into cohesive groups that simultaneously maximise positive intra-community links and negative inter-community links, in accordance with **structural balance theory**.

This thesis addresses the problem of multi-community detection in signed social networks through a **hybrid exact–heuristic** (matheuristic) framework named **HybridVNS-ILP**. The proposed algorithm combines three complementary components : (i) a balance-informed greedy construction that exploits the local sign structure to build a high-quality initial partition ; (ii) a **Variable Neighborhood Search** (VNS) core that explores four neighbourhood structures of increasing perturbation strength to escape local minima ; and (iii) a **conflict-driven Integer Linear Programming** (ILP) refinement phase that identifies the most problematic nodes and optimally reassigns them on a restricted subgraph.

For the exact restricted-refinement variant, the method converges to a local optimum of the disagreement function $f(\mathcal{C})$ in a finite number of iterations, with time complexity $O(m \log n)$ and space complexity $O(n + m)$. Experimental evaluation on four benchmark signed network datasets ($n \in \{8, 12, 20, 100\}$, $K \in \{2, 3, 4\}$) demonstrates that HybridVNS-ILP outperforms Random and Greedy, and matches SLPA on the two largest datasets — Random, Signed Label Propagation, and Greedy — achieving an average improvement rate of 36.1% over the greedy baseline and **exactly recovering the planted community structure** on the two largest datasets. A fully documented Python implementation and an interactive graphical interface are also contributed.

Keywords : signed social networks, community detection, Variable Neighborhood Search, Integer Linear Programming, matheuristic.

Résumé

Les réseaux sociaux encodent non seulement des relations positives — telles que l'amitié, la confiance et la collaboration — mais aussi des relations négatives, incluant la rivalité, la méfiance et l'opposition. Les réseaux dans lesquels chaque arête porte un signe explicite positif ou négatif sont appelés **réseaux sociaux signés**. La détection de communautés dans de tels réseaux vise à partitionner l'ensemble des nœuds en groupes cohésifs maximisant simultanément les liens positifs intra-communauté et les liens négatifs inter-communauté, conformément à la **théorie de l'équilibre structurel**.

Ce mémoire traite le problème de la détection multi-communautaire dans les réseaux sociaux signés à travers un cadre **hybride exact–heuristique** (matheuristique) dénommé **HybridVNS-ILP**. L'algorithme proposé combine trois composantes complémentaires : (i) une construction gloutonne informée par l'équilibre, qui exploite la structure locale des signes pour produire une partition initiale de qualité ; (ii) un cœur de **Recherche à Voisinage Variable** (VNS) qui explore quatre structures de voisinage de force perturbatrice croissante afin d'échapper aux minima locaux ; et (iii) une phase de **raffinement par Programmation Linéaire en Nombres Entiers** (ILP) pilotée par les conflits, qui identifie les nœuds les plus problématiques et les réaffecte de manière optimale sur un sous-graphe restreint.

Dans sa version à raffinement exact restreint, le cadre est conçu pour converger vers un optimum local en un nombre fini d'itérations local de la fonction de désaccord $f(\mathcal{C})$ en un nombre fini d'itérations est démontrée, avec une complexité temporelle $O(m \log n)$ et spatiale $O(n + m)$. Les résultats montrent que HybridVNS-ILP surpasse nettement les approches Aléatoire et Gloutonne, et obtient des performances comparables à Signed Label Propagation (SLPA) sur les deux plus grands jeux de données, Propagation d'Étiquettes Signée et Glouton — avec un taux d'amélioration moyen de 36,1 % par rapport à l'initialisation gloutonne et une **récupération exacte de la structure communautaire plantée** sur les deux jeux de données les plus grands. Une implémentation Python entièrement documentée et une interface graphique interactive sont également fournies.

ملخص

تُرمِّز الشبكات الاجتماعية ليس فقط العلاقات الإيجابية — كالصداقة والثقة والتعاون — بل أيضاً العلاقات السلبية، كالتنافس وانعدام الثقة والمعارضة. تُعرَّف الشبكات التي تحمل أضلاعها إشارةً صريحة موجبة أو سالبة بـ **الشبكات الاجتماعية الموقَّعة.**

يهدف اكتشاف المجتمعات في هذه الشبكات إلى تقسيم مجموعة العقد إلى مجموعات متماسكة تُعظِّم في آنٍ واحد الروابط الموجبة داخل المجتمعات والروابط السالبة بين المجتمعات، وفقاً لـ**نظرية التوازن الهيكلي.**

تعالج هذه الأطروحة مشكلة اكتشاف المجتمعات المتعددة في الشبكات الاجتماعية الموقَّعة من خلال إطار **هجين دقيق-استكشافي** يُسمَّى **HybridVNS-ILP.** يجمع الخوارزم المقترَح بين ثلاثة مكوِّنات متكاملة: (i) بناء جشِع مُستنير بالتوازن يستغل البنية المحلية للإشارات لإنتاج تقسيم أولي عالي الجودة؛ (ii) نواة **البحث في الجوار المتغير** (VNS) التي تستكشف أربع هياكل جوارية بقوة اضطراب متصاعدة للإفلات من المينيما المحلية؛ (iii) مرحلة **تحسين دقيقة مدفوعة بالصراع** باستخدام **البرمجة الخطية الصحيحة** (ILP)، تُحدِّد العقد الأكثر إشكاليةً وتُعيد تعيينها بصورة مثلى على مصفوفة فرعية مقيَّدة.

يُثبَت أن الخوارزم يتقارب نحو ذروة محلية لدالة الاختلاف $f(\mathcal{C})$ في عدد منتهٍ من التكرارات، بتعقيد زمني $O(m \log n)$ ومكاني $O(n+m)$.

كشفت التجارب التي أُجريت على أربع مجموعات بيانات مرجعية للشبكات الموقَّعة ($n \in \{8, 12, 20, 100\}$، $K \in \{2, 3, 4\}$) وتُظهر النتائج أن HybridVNS-ILP يتفوق بوضوح على الطريقتين العشوائية والجشعة ، كما يحقق أداءً مقاربًا لطريقة Propagation Label Signed (SLPA) على أكبر مجموعتي بيانات مُحقّقاً معدَّل تحسين متوسطاً يبلغ 36.1 % مقارنةً بالتهيئة الجشعة، **واسترجاعاً دقيقاً للبنية المجتمعية الحقيقية المزروعة** في أكبر مجموعتَي بيانات.

كما تُقدَّم تنفيذاً برمجياً كاملاً بلغة Python وواجهة رسومية تفاعلية.

Acknowledgement

First and we would like to express my sincere gratitude to Allah for giving me the strength, patience, and guidance to complete this work.

we would like to thank our supervisor for her valuable guidance, continuous support, and insightful feedback throughout this project.

we are also deeply grateful to all the teachers of the Department of Computer Science for their dedication and for sharing their knowledge during our academic journey.

Special thanks go our families for their unconditional love, encouragement, and support during our studies.

Finally, we like to thank our friends and colleagues who helped us directly or indirectly in the completion of this work.

Table des matières

Abstract **i**

Résumé **ii**

ملخص **iii**

Acknowledgement **iv**

List of Figures **x**

List of Tables **xi**

1 Introduction **1**
1.1 Context and Motivation 1
1.2 Problem Statement 3
1.3 Research Objectives 4
1.4 Thesis Contribution 4
1.5 Organization of the Thesis 5

2 Background and Preliminaries **7**
2.1 Basic Concepts of Graph Theory 7
2.1.1 Graphs, Nodes, and Edges 7
2.1.2 Key Graph Matrices 8
2.1.3 Paths, Connectivity, and Subgraphs 11
2.2 Signed Social Networks 11
2.2.1 Definition and Formal Model 12
2.2.2 Signed Degrees and Signed Laplacian 12
2.2.3 Real-World Signed Network Datasets 13
2.2.4 Challenges Specific to Signed Networks 13
2.3 Community Detection in Networks 14
2.3.1 Definition and Problem Formulation 14

2.3.2 Unsigned Modularity 14
2.3.3 Why Unsigned Methods Fail on Signed Networks 15
2.4 Structural Balance Theory 15
2.4.1 Origins and Intuition 15
2.4.2 Formal Definition of Structural Balance 16
2.4.3 Weak Balance and k-Balance 16
2.5 Quality Measures for Signed Communities 17
2.5.1 Disagreement Function 17
2.5.2 Balance Ratio 18
2.5.3 Internal and External Quality 18
2.5.4 Signed Normalised Cut 19
2.5.5 Summary 19
2.6 Review of Existing Methods 20
2.6.1 Exact Algorithms 20
2.6.2 Heuristic Algorithms 23
2.6.3 Comparative Summary and Discussion 28

3 Proposed Algorithm **30**
3.1 Problem Formulation 30
3.1.1 Formal Statement 30
3.1.2 Relationship to Signed Modularity 31
3.1.3 Decomposition of the Objective 31
3.2 Design Principles of the Heuristic 32
3.3 Description of the Proposed Algorithm 33
3.3.1 Initialisation Strategy 34
3.3.2 Community Assignment and Refinement 35
3.3.3 Stopping Criterion 38
3.4 Pseudocode of the Algorithm 39
3.5 Complexity Analysis 40
3.5.1 Time Complexity per Iteration 40
3.5.2 Space Complexity 41
3.5.3 Parameter Sensitivity 41
3.5.4 Comparison with Baseline Methods 42

4 Numerical Experiments **44**
4.1 Experimental Setup 44
4.1.1 Objectives 44
4.1.2 Hardware and Software Environment 45
4.1.3 Parameter Configuration 45
4.2 Datasets 45

4.2.1 Overview . . . 45
4.2.2 Dataset 1 — Chapter 3 Example ($n = 8$, $K = 2$) . . . 46
4.2.3 Dataset 2 — 3-Community Graph ($n = 12$, $K = 3$) . . . 46
4.2.4 Dataset 3 — Random Signed Graph ($n = 20$, $K = 3$) . . . 46
4.2.5 Dataset 4 — Medium Signed Graph ($n = 100$, $K = 4$) . . . 47
4.2.6 Visual Representation . . . 47
4.3 Baseline Methods . . . 48
4.3.1 Baseline 1 — Random Assignment . . . 48
4.3.2 Baseline 2 — Signed Label Propagation (SLPA) . . . 49
4.3.3 Baseline 3 — Greedy Construction Only . . . 49
4.4 Evaluation Metrics . . . 50
4.4.1 Total Disagreement (f) . . . 50
4.4.2 Balance Ratio (BR) . . . 50
4.4.3 Improvement Rate (IR) . . . 50
4.4.4 Running Time . . . 50
4.5 Implementation Details . . . 51
4.5.1 Code Architecture . . . 51
4.5.2 Greedy Initialisation Fix . . . 51
4.5.3 ILP Approximation . . . 51
4.5.4 Stopping Criterion . . . 51
4.6 Results and Comparison . . . 52
4.6.1 Main Results . . . 52
4.6.2 Robustness over Multiple Runs . . . 53
4.6.3 Detailed Analysis per Dataset . . . 53
4.6.4 Convergence Behaviour . . . 55
4.6.5 Improvement Rate Analysis . . . 56
4.6.6 Balance Ratio Summary . . . 56
4.6.7 Runtime Analysis . . . 57
4.7 Discussion . . . 57
4.7.1 Strengths of HybridVNS-ILP . . . 57
4.7.2 Limitations . . . 58
4.7.3 Expected Benefit of Full ILP . . . 58
4.7.4 Parameter Sensitivity . . . 58
4.7.5 Positioning Relative to the Literature . . . 59
Chapter Summary . . . 59

5 Conclusion **60**
5.1 Summary of Contributions . . . 60
5.1.1 Comprehensive Literature Review . . . 60

5.1.2 Formal Problem Formulation 61
5.1.3 The HybridVNS-ILP Algorithm 61
5.1.4 Python Implementation and Interactive GUI 61
5.1.5 Experimental Evaluation 62
5.2 Main Findings 62
5.3 Limitations 63
5.3.1 Approximate ILP Refinement 63
5.3.2 Small-Scale Evaluation 63
5.3.3 Fixed Number of Communities K 64
5.3.4 Undirected and Unweighted Edges Only 64
5.3.5 No Theoretical Approximation Guarantee 64
5.4 Future Work 64
5.4.1 Evaluation on Real-World Signed Benchmarks 64
5.4.2 Full ILP Integration 65
5.4.3 Automatic Community Number Selection 65
5.4.4 Extension to Directed and Weighted Signed Networks 65
5.4.5 Parallel and Distributed VNS 65
5.4.6 Graph Neural Network Integration 65
5.4.7 Online and Dynamic Signed Networks 66
Closing Remarks 66