MSc thesis

by Manasvi Pandey

Submission date: 21-May-2026 06:40PM (UTC+0530)
Submission ID: 2966363406
File name: Manasvi_Pandey_final_thesis.pdf (478.59K)
Word count: 16413
Character count: 83981

CHAPTER 1

INTRODUCTION

1.1 Background

Digital platforms now depend heavily on recommendation systems to decide which movies, products, songs, books or news articles are shown to a user. The mathematical task behind these systems is deceptively simple: from a large and usually sparse record of user–item interactions, infer a small set of items that is likely to be relevant. In practice the task is difficult because most users interact with only a tiny fraction of the catalogue, and the available ratings or clicks are noisy, incomplete and unevenly distributed across items.

A major family of methods for this problem is collaborative filtering. In this approach, recommendations are derived from interaction patterns rather than from manually supplied item descriptions. Memory-based methods compare either users or items directly, most commonly with cosine similarity or Pearson correlation. Latent factor methods, such as matrix factorisation, learn lower-dimensional representations of users and items so that their inner products reproduce the observed ratings (21). More recent graph-based models, including PinSAGE and LightGCN, treat the user–item data as a graph and propagate information through its neighbourhood structure (39; 13).

Most of these models are built on an unsigned view of interaction. An observed rating contributes positive evidence of preference, while a missing rating is usually treated as unknown. This representation is useful but incomplete. A low rating is not the same as a missing rating. A user who gives a horror film five stars and a romantic comedy one star is expressing a pattern of preference opposition. If the low rating is reduced to absence of information, the model loses a signal that may be important for catalogue structure and diversity.

Signed graphs provide a natural language for separating these two kinds of relations. In a signed graph, edges may be positive or negative. Positive edges express similarity, support or agreement; negative edges express opposition, conflict or disagreement. The theory began with Harary's work on balance in signed graphs, motivated by Heider's ideas in social psychology (15; 10). The same signed structure later became important in social network analysis, where trust and distrust links appear together (23; 35), and in signed graph neural networks (9; 42).

Another relevant branch of graph theory is domination. In an ordinary graph, a set of vertices is dominating if every vertex outside the set has a neighbour in it. This idea is useful whenever a large structure must be covered or represented by a smaller subset. The minimum dominating set problem and its variants are standard topics in graph theory (12). Domination has been applied to location, monitoring and network coverage problems, but its use as a tool for selecting representative items in recommender catalogues is still limited.

The present dissertation connects these two ideas. We construct a signed item–item graph from rating data and then use a signed domination rule to select a small set of catalogue representatives. The positive part of the graph comes from ordinary cosine similarity of rating vectors. The negative part is derived from an opposite preference matrix that records cases where users like one item and dislike another. The resulting graph contains both agreement and opposition, which makes it more informative than a purely positive item similarity graph.

Signed domination has its own theoretical literature. Acharya introduced a notion of domination and absorbance in signed graphs (1). Joseph and Joseph studied Roman domination in signed graphs (20), and Sankar et al. introduced domination integrity for signed fuzzy graphs (29). These works are primarily theoretical. This dissertation uses their ideas in a computational setting: recommendation data supplies a real signed graph, and the dominating set becomes a compact, interpretable summary of the catalogue.

1.2 Motivation

The motivation for the work is based on three connected observations.

(i) Negative preference is not empty information. A low rating contains information about what a user rejects. In an unsigned similarity graph, two items may appear unrelated simply because their positive co-rating patterns are weak. But if one group of users consistently rates one item highly and another item poorly, the two items are not merely unrelated; they are opposed in a preference sense. A recommendation framework that records this opposition can avoid representatives that are similar to many items but also strongly polarising.

(ii) A dominating set gives an interpretable catalogue summary. A recommender often needs a small set of seed items, benchmark items or representative items. A popularity list gives one such set, but it tends to repeat the majority taste of the catalogue. A dominating set is different: it is chosen to cover the graph structurally. If a niche region is not covered by a popular movie, the domination rule forces the algorithm to choose a movie that can cover that region. This is why domination can improve genre spread without explicitly optimising for genre labels.

(iii) The theory and the data can be joined. The mathematical theory of signed domination is available (1; 20; 29; 19; 36), and public rating datasets such as MovieLens are also available (11). What is missing is a concrete bridge: a rule that turns user–item ratings into a signed item–item graph and then computes a signed dominating set on that graph. Building and testing this bridge is the central purpose of the dissertation.

1.3 Problem Statement

The central problem treated in this dissertation is the following.

Problem 1.1 (Signed-graph domination for representative selection)**.** Let $\mathscr{U} = \{u_1, \ldots, u_m\}$ be a set of users and $\mathscr{I} = \{i_1, \ldots, i_n\}$ be a set of items, with a sparse rating matrix $R \in \mathbb{R}^{m \times n}$. Construct a signed graph $\Sigma^{\pm} = (\mathscr{I}, E^{+}, E^{-}, \sigma)$ on the items together with a domination relation derived from R, such that a minimum signed-dominating set $D \subseteq \mathscr{I}$ provides a structurally representative selection of $|D|$ items covering the whole catalogue.

This problem has three parts. First, the rating matrix must be converted into positive and negative item–item relations. Second, a meaningful domination rule must be defined for a graph where similarity and opposition coexist. Third, the dominating set must be computed efficiently, because the exact minimum dominating set problem is NP-hard (12). The methodology in Chapter 4 addresses these three parts in sequence.

1.4 Research Objectives

The objectives of the dissertation are as follows:

(O1) To define a signed item–item graph from collaborative filtering data in a form compatible with the signed and signed-fuzzy graph literature.

(O2) To construct an opposite preference matrix from high-rating and low-rating indicators, so that negative item relations are obtained directly from the rating data rather than inserted artificially.

(O3) To formulate a dual-threshold domination rule using a positive-similarity floor α and a negative-similarity ceiling β.

(O4) To implement a greedy algorithm for obtaining a small signed dominating set and to evaluate it on the MovieLens `ml-latest-small` dataset (11).

(O5) To interpret the empirical results in relation to Acharya's signed domination, Joseph and Joseph's Roman domination, Sankar et al.'s domination integrity and recent signed graph learning literature (1; 20; 29; 42).

1.5 Contributions of the Dissertation

The main contributions are listed below.

A signed graph construction for item catalogues. The dissertation gives a complete construction of a signed graph $\Sigma^{\pm}$ on a set of items. The positive matrix S^{+} is computed using cosine similarity of item rating vectors. The negative matrix S^{-} is obtained from the normalised opposite preference product $HL^{\top}$. This construction keeps the familiar collaborative-filtering similarity while adding a separate channel for preference opposition.

A dual-threshold signed domination rule. The dissertation defines item i_1 to dominate item i_2 when $S^+(i_1,i_2) \geq \alpha$ and $S^-(i_1,i_2) \leq \beta$. This rule is designed for real-valued signed graphs: a representative item must be sufficiently similar to what it covers and must not be strongly opposed to it. It therefore adapts the spirit of signed domination to recommendation data.

A greedy algorithm and empirical validation. Algorithm 1 computes a signed dominating set by repeatedly choosing the item that covers the largest number of still-uncovered items. On the top 100 movies from MovieLens `ml-latest-small`, the setting $\alpha = \beta = 0.5$ produces a 9-movie set with full structural coverage. The resulting set spans 12 unique genres, while a popularity baseline of the same size spans 10.

Theoretical positioning. The work places the empirical construction beside existing results in signed graph theory. In Chapter 7, the obtained signed graph and its dominating set are discussed through balance theory (15; 10; 5), Acharya's algebraic domination (1), domination integrity in signed fuzzy graphs (29), and signed graph learning methods (9; 24; 31; 42).

1.6 Organisation of the Dissertation

The remainder of the dissertation is organised as follows.

Chapter 2 reviews the literature on recommendation systems, signed graphs, domination in graphs, and signed-fuzzy graph domination.

Chapter 3 records the mathematical definitions and results used later, including signed graphs, balance, domination and domination integrity.

Chapter 4 presents the proposed methodology: data preparation, positive similarity, opposite preference, signed graph construction, the dual-threshold domination relation and the greedy algorithm.

Chapter 5 describes the MovieLens dataset, preprocessing choices and software environment.

Chapter 6 reports the empirical results, including the similarity matrices, the signed graph structure, the selected dominating set, baseline comparison and threshold sensitivity.

Chapter 7 interprets the results using balance theory, Acharya-style domination and domination integrity.

Chapter 8 summarises the work, states its limitations and suggests directions for future research.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews the research areas on which the dissertation is based. The first area is collaborative filtering and graph-based recommendation. The second is signed graph theory and signed network mining. The third is domination in classical, signed and fuzzy graphs. The purpose of the review is not to list all known methods, but to identify the exact point at which the proposed work fits: a signed-graph domination method for selecting representative items from a recommendation catalogue.

2.1 Recommendation Systems

Recommendation systems are usually divided into collaborative, content-based and hybrid methods. Content-based methods use attributes of the item or the user. Collaborative filtering, which is the setting of this dissertation, uses the pattern of interactions between users and items. The basic assumption is that similar interaction histories carry information about future preferences (21).

Neighbourhood collaborative filtering. Early collaborative filtering methods were neighbourhood based. In a user-based method, the system finds users whose rating vectors are similar to the target user and uses their ratings to make a prediction. In an item-based method, the system compares item rating vectors and predicts a user's rating for an item from that user's ratings on similar items. Cosine similarity and Pearson correlation are common choices for these comparisons. The positive similarity matrix used in Chapter 4 is directly connected to this tradition: it is the cosine similarity matrix of item rating vectors.

Latent factor models. The next major development was the use of latent factor models. Matrix factorisation represents each user and each item by vectors in a lower dimensional space and approximates a rating by an inner product, $r_{u,i} \approx p_u^\top q_i$ (21). The popularity of these models grew substantially after the Netflix Prize, because relatively simple factorisation methods performed very well on sparse rating data. Such models are powerful for prediction, but the learned factors are not immediately a small structural cover of the catalogue.

Graph neural network recommenders. Recent recommender models often use the user–item graph explicitly. PinSAGE was designed for large-scale recommendation on graph data (39). LightGCN showed that, for recommendation, simple neighbourhood propagation can be more effective than heavy graph-convolutional transformations (13). These approaches are useful, but they usually operate on unsigned interaction graphs. The edge means that an interaction exists; it does not by itself distinguish approval from rejection.

Negative feedback and signed recommendation. A smaller body of work considers negative feedback. In signed recommendation, low ratings or distrust links may be used as negative edges. Signed graph neural network methods use this information for sign prediction or representation learning (9; 42). The present work shares the view that negative interactions are meaningful, but it uses a different tool. Instead of learning embeddings, it constructs positive and negative similarity matrices and then solves a domination problem on the resulting signed item graph.

2.2 Signed Graphs and Social Network Mining

A signed graph is a graph in which every edge has a sign. The classical origin of the subject is Harary's theory of balance, which formalised Heider's social psychological principle that "the friend of my friend is my friend" and related statements about hostile relations (15; 10). Cartwright and Harary later extended the idea to group structures (5), and Davis introduced clusterability as a more general signed-graph property (8). Zaslavsky's work provided a broad algebraic treatment of signed graphs, including switching and matroidal aspects (41).

Signed networks in online platforms. Large online networks made signed graphs computationally important. Leskovec, Huttenlocher and Kleinberg studied signed links in Epinions, Slashdot and Wikipedia and showed that both balance theory and status theory explain many observed sign patterns (23). Tang et al. surveyed signed social network mining and identified link sign prediction, node classification and community detection as central tasks (35). This line of work shows that positive and negative links should not be collapsed into a single unsigned relation.

Signed graph neural networks. Deep learning on signed graphs began with models such as SGCN, which propagates information separately through balanced and unbalanced signed paths (9). Later methods added attention mechanisms, motif-based features and contrastive objectives (17; 24; 31). The recent survey by Zhang et al. provides a systematic account of signed graph learning and its applications (42). These methods usually output embeddings, labels or edge signs. They do not directly produce a small representative subset of items, which is the combinatorial task considered here.

2.3 Domination in Classical Graphs

Domination is a central idea in graph theory. A subset D of vertices dominates a graph when every vertex outside D is adjacent to at least one vertex of D. The minimum size of such a set is the domination number. Early foundations of the subject are associated with Ore and Berge, and the monograph of Haynes, Hedetniemi and Slater remains a standard reference (27; 12). The minimum dominating set problem is NP-hard, so approximation and heuristic algorithms are important.

Roman domination and related variants. Roman domination is one of the best-known variants. A Roman dominating function labels each vertex by 0, 1 or 2 such that every vertex labelled 0 is protected by a neighbour labelled 2 (7). Many variants have been studied, including double Roman and signed Roman versions (20). These variants illustrate the flexibility of domination as a modelling idea: the exact domination rule changes with the interpretation of coverage.

Integrity and vulnerability. Domination can also be combined with measures of network vulnerability. The integrity of a graph evaluates what remains after a selected set of vertices is removed. Domination integrity restricts this removal set to dominating sets and therefore measures both coverage cost and residual connectedness (30; 25). This is relevant to the present work because a representative item set should ideally cover the catalogue without making the remaining structure too fragile.

2.4 Domination in Signed Graphs

Signed domination is more delicate than ordinary domination because an edge can be helpful or hostile. Acharya's foundational paper introduced domination in signed graphs and absorbance in signed digraphs (1). In that formulation, a dominating set must be compatible with a marking $\mu : V \to \{1, -1\}$, and a vertex u outside the set is dominated through neighbours v satisfying $\sigma(uv) = \mu(u)\mu(v)$. When all edges are positive, this reduces to the ordinary domination concept.

Structural results in Acharya's framework. Acharya established several structural facts for signed domination. For example, every finite signed tree has two disjoint dominating sets, and a balanced signed graph has the same dominating sets as its underlying ordinary graph (1). These results show that the signs are not decorative; the relationship between the signed graph and its underlying graph depends on balance and marking compatibility.

Roman and majority-type signed domination. Joseph and Joseph extended Roman domination to signed graphs and characterised signed paths, cycles and stars that admit Roman dominating functions (20). Walikar et al. considered another signed domination variant and studied bounds on the corresponding domination number (36). Jeyalakshmi proposed a majority-style definition in which a vertex is dominated when positive neighbours from the dominating set outnumber negative ones (19). These different definitions reflect different meanings of a negative edge. In some settings a negative edge is a polarity to be respected; in others it is an obstacle to domination.

The present dissertation uses a data-driven version of the same broad idea. An item can cover another item only if its positive similarity is high enough and its negative similarity is low enough. Thus the signed information enters as a constraint on representation.

2.5 Fuzzy and Signed-Fuzzy Graph Domination

Fuzzy graphs generalise ordinary graphs by assigning membership values to vertices and edges. Rosenfeld's fuzzy graph framework is based on Zadeh's fuzzy set theory (40; 28). Bhutani and Rosenfeld introduced strong arcs, which became important in fuzzy graph connectivity (3). Somasundaram and Somasundaram developed early domination concepts for fuzzy graphs (33; 32).

Signed fuzzy graphs extend this further by allowing both strength and sign. Sundareswaran et al. formulated signed fuzzy graphs with membership values that can carry positive or negative polarity (34). Chakaravarthy et al. studied edge integrity in this setting (6). The most directly relevant work for this dissertation is Sankar et al.'s domination integrity in signed fuzzy graphs (29). Their invariant combines a dominating set with the size of the largest component left after its removal:

$$DI(\Sigma^{\pm}) = \min_{S}\{|S| + m(\Sigma^{\pm} - S)\},$$

where the minimum is taken over dominating sets. They prove lower-bound, monotonicity and complete-graph results for this parameter. In Chapter 7, the dominating set obtained from MovieLens is read in light of this domination-integrity perspective.

2.6 Research Gap and Position of This Work

The preceding literature suggests a clear gap. Recommendation systems have rich methods for similarity and prediction, but representative item selection is usually handled by popularity or embedding-based heuristics. Signed graph learning uses positive and negative links, but mainly for prediction and representation. Domination theory studies small covering sets, including in signed and fuzzy settings, but is rarely applied to recommender catalogues.

This dissertation fills that gap by combining the three strands. It maps rating data to a signed item graph, defines a dual-threshold domination relation on that graph, and evaluates the resulting dominating set on MovieLens. The framework is therefore not merely another similarity measure and not merely an application of a standard recommender model. It is a catalogue-level structural method built from signed graph domination.

More specifically, the dissertation contributes the following bridge:

1. positive item relations are obtained from cosine similarity, preserving a standard collaborative-filtering component;
2. negative item relations are obtained from an opposite preference matrix, which records high–low rating conflict;
3. a greedy signed domination algorithm selects a compact set of representatives that covers the selected catalogue; and
4. the empirical results are interpreted using the signed domination and signed-fuzzy

domination literature (1; 20; 29).

The following chapters make this construction precise and test it on data.

CHAPTER 3

MATHEMATICAL PRELIMINARIES

This chapter fixes the notation used in the rest of the dissertation. Only the concepts needed for the proposed method and the later analysis are included. The definitions are standard, but they are stated here so that the construction of Chapter 4 can be read without repeatedly returning to the literature.

3.1 Graphs and Signed Graphs

All graphs in this dissertation are finite and simple unless stated otherwise. A graph is denoted by $G = (V,E)$, where V is the vertex set and $E \subseteq \binom{V}{2}$ is the edge set. For a vertex $u \in V$, the open neighbourhood is

$$N(u) = \{v \in V : uv \in E\},$$

and the closed neighbourhood is $N[u] = N(u) \cup \{u\}$.

Definition 3.1 (Signed graph)**.** A *signed graph* is a triple $\Sigma = (V,E,\sigma)$, where (V,E) is the underlying graph and $\sigma : E \to \{+1,-1\}$ assigns a sign to each edge. The underlying unsigned graph is written as $|\Sigma|$.

If $\sigma(e) = +1$, the edge e is positive; if $\sigma(e) = -1$, the edge is negative. The corresponding edge sets are $E^+(\Sigma)$ and $E^-(\Sigma)$. For $u \in V$, we write

$$N^+(u) = \{v \in N(u) : \sigma(uv) = +1\}, \qquad N^-(u) = \{v \in N(u) : \sigma(uv) = -1\}.$$

The positive and negative degrees are $d^+(u) = |N^+(u)|$ and $d^-(u) = |N^-(u)|$. This notation follows the signed graph domination literature used later (1; 20).

A signed graph is homogeneous when all its edges have the same sign, and heterogeneous otherwise. The graphs produced from rating data in this dissertation are heterogeneous: some item pairs are similar, while other item pairs show preference opposition.

3.2 Balance Theory and Clusterability

Balance is the classical structural property of signed graphs. It expresses the idea that the vertices can be split into two groups so that positive edges stay inside groups and negative edges go across groups. The concept goes back to Heider's theory of attitudes and Harary's graph-theoretic formulation (15; 10).

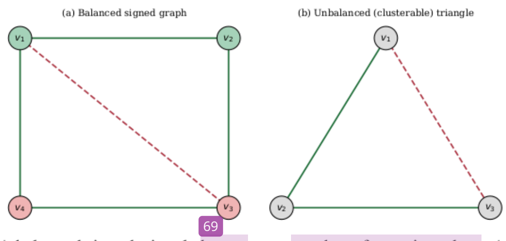

Figure 3.1: A balanced signed triangle has an even number of negative edges. A triangle with exactly one negative edge is unbalanced.

Definition 3.2 (Balanced signed graph)**.** A signed graph $\Sigma = (V, E, \sigma)$ is *balanced* if V can be partitioned as $V = V_1 \cup V_2$, with $V_1 \cap V_2 = \emptyset$, such that every positive edge has both endpoints in the same part and every negative edge has endpoints in different parts. Such a partition is called a Harary bipartition.

An equivalent characterisation is that every cycle of a balanced signed graph contains an even number of negative edges (10). A balanced signed graph also admits a marking $\mu : V \to \{+1, -1\}$ satisfying $\sigma(uv) = \mu(u)\mu(v)$ for each edge. This marking viewpoint is important because it is also used in Acharya's domination definition.

A weaker idea is clusterability. Instead of two groups, the vertex set may be partitioned into several clusters, with positive edges inside clusters and negative edges between clusters (8). In the empirical chapters we do not attempt to compute an exact balance or clustering partition. However, the language of balance helps interpret why a signed item graph may have large regions of coherent taste.

3.3 Domination in Graphs

Domination is a covering concept. A small set of selected vertices should be close enough to every other vertex to represent or control the whole graph.

Definition 3.3 (Dominating set)**.** Let $G = (V, E)$ be a graph. A subset $D \subseteq V$ is a *dominating set* if for every $u \in V \setminus D$ there exists $v \in D$ such that $uv \in E$. The *domination number* is

$$\gamma(G) = \min\{|D| : D \text{ is a dominating set of } G\}.$$

A dominating set is minimal if no proper subset remains dominating. It is minimum if its size is $\gamma(G)$. The general theory is extensive; the monograph by Haynes, Hedetniemi and Slater is a standard reference (12).

The exact computation of a minimum dominating set is NP-hard. A common greedy heuristic repeatedly selects a vertex that dominates the largest number of currently undominated vertices. This greedy idea is adapted in Chapter 4; the difference is that domination there depends on a positive-similarity threshold and a negative-similarity threshold.

3.4 Domination in Signed Graphs

There is no single universal definition of domination in signed graphs. The meaning of a negative edge depends on the application, so several formulations coexist.

Acharya's algebraic definition. Acharya's definition uses a vertex marking to decide whether a signed edge is compatible with domination (1).

Definition 3.4 (Acharya signed dominating set)**.** Let $\Sigma = (V, E, \sigma)$ be a signed graph. A subset $D \subseteq V$ is a *dominating set* of Σ if there exists a marking $\mu : V \to \{+1, -1\}$ such that every vertex $u \in V$ is either in D or satisfies $N(u) \cap D \neq \emptyset$ and

$$\sigma(uv) = \mu(u)\mu(v) \quad \text{for every } v \in N(u) \cap D.$$

If every edge is positive, the marking $\mu \equiv +1$ works and the definition reduces to ordinary domination. Acharya proved, among other results, that for a balanced signed graph the signed dominating sets coincide with the dominating sets of the underlying graph (1).

Jeyalakshmi's majority definition. Another definition reads the sign as support or opposition. In this case, a vertex should receive more positive support than negative opposition from the dominating set (19).

Definition 3.5 (Jeyalakshmi signed dominating set)**.** A subset $D \subseteq V$ of a signed graph Σ is a dominating set if for every $v \in V \setminus D$,

$$|N^+(v) \cap D| > |N^-(v) \cap D|.$$

The dual-threshold definition used here. For recommendation data, the signs are attached to real-valued similarity matrices. A purely algebraic sign condition is therefore too coarse. The dissertation uses the following thresholded version.

Definition 3.6 (Dual-threshold signed domination)**.** Let $\Sigma^{\pm}$ be a signed graph equipped with positive and negative edge weights $S^+, S^- : V \times V \to [0, 1]$, and let $\alpha, \beta \in [0, 1]$. Item i_1 *dominates* item i_2 if

$$S^+(i_1, i_2) \geq \alpha \quad \text{and} \quad S^-(i_1, i_2) \leq \beta.$$

A set $D \subseteq V$ is dominating under thresholds α, β if every $v \in V \setminus D$ is dominated by at least one $u \in D$.

The first inequality requires enough positive similarity. The second inequality prevents an item from covering another item when there is strong preference opposition between them. This is the signed recommender analogue of asking a representative to be both close and non-conflicting.

Roman signed domination. Joseph and Joseph introduced a Roman-style formulation for signed graphs (20).

Definition 3.7 (Roman dominating function on a signed graph)**.** A function $f : V \to \{0,1,2\}$ on a signed graph Σ is a Roman dominating function if, for every $v \in V$,

$$f(N[v]) := f(v) + \sum_{u \in N(v)} \sigma(uv) f(u) \geq 1,$$

and every vertex v with $f(v) = 0$ has a positive neighbour $u \in N^+(v)$ with $f(u) = 2$.

This version is not used in the algorithm, but it is part of the theoretical background because it shows how classical domination ideas change when signs are introduced.

3.5 Fuzzy and Signed-Fuzzy Graphs

The empirical graph in this dissertation has edge weights in $[0,1]$ rather than only the values 0 and 1. This makes fuzzy graph terminology useful.

Definition 3.8 (Fuzzy graph; Rosenfeld 1975 (28))**.** A *fuzzy graph* is a pair (η, φ) where $\eta : V \to [0,1]$ is a vertex membership function and $\varphi : V \times V \to [0,1]$ is an edge membership function satisfying

$$\varphi(v_1, v_2) \leq \min(\eta(v_1), \eta(v_2))$$

for all $v_1, v_2 \in V$.

Signed fuzzy graphs add polarity to fuzzy strength. One may view such a graph as having weighted positive and negative parts. This is close to the structure constructed in Chapter 4, where S^+ stores positive similarity and S^- stores negative or opposite preference intensity. Signed fuzzy graphs and related notions have been studied by Sundareswaran et al. and Sankar et al. (34; 29).

Strong arcs are also used in fuzzy graph theory. The connectivity between two vertices is obtained by looking at paths and taking the strongest path under a minimum-edge rule (3). Although the dissertation does not use strong arcs algorithmically, this viewpoint supports the interpretation of weighted signed edges as structural strengths rather than merely labels.

3.6 Domination Integrity

The signed-fuzzy invariant most relevant to the discussion is domination integrity. It combines two quantities: the size of a dominating set and the size of the largest component remaining after that set is removed.

Definition 3.9 (Domination integrity (29))**.** Let $\Sigma^\pm$ be a signed fuzzy graph. The *domination integrity* of $\Sigma^\pm$ is

$$DI(\Sigma^\pm) = \min_{S \text{ dominating}} \left\{ |S| + m(\Sigma^\pm - S) \right\},$$

where $|S| = \sum_{v \in S} \eta(v)$ and $m(\Sigma^{\pm} - S)$ is the order of the largest connected component of $\Sigma^{\pm} - S$.

A small dominating set is not automatically desirable if its removal leaves a large vulnerable component. Domination integrity therefore measures coverage and post-removal robustness together. Sankar et al. prove several properties that will be used as comparison points later (29).

Theorem 3.10 (Theorem 1 of (29))**.** *For any signed fuzzy graph* $\Sigma^{\pm}$, $\gamma(\Sigma^{\pm}) \leq DI(\Sigma^{\pm})$.

Theorem 3.11 (Theorem 2 of (29); monotonicity)**.** *If* $N_{\pm}$ *is a sub-signed-fuzzy-graph of* $\Sigma^{\pm}$*, then* $DI(N_{\pm}) \leq DI(\Sigma^{\pm})$.

Theorem 3.12 (Theorem 3 of (29); complete graphs)**.** *If* $\Sigma^{\pm}$ *is a complete signed fuzzy graph or the complement of one, then* $DI(\Sigma^{\pm})$ *equals the order of* $\Sigma^{\pm}$.

Sankar et al. also define efficient signed fuzzy graphs, for which the dominating sets agree with those of the underlying crisp graph. Their results on efficient and strong signed fuzzy graphs help frame the discussion in Chapter 7.

3.7 Key Results Used Later

The following facts will be cited in Chapter 7:

- in a balanced signed graph, Acharya's signed dominating sets coincide with ordinary dominating sets of the underlying graph (1);
- every finite signed tree has two disjoint dominating sets (1);
- a signed star $K_{1,n-1}$ admits a Roman dominating function precisely under the degree condition stated by Joseph and Joseph (20);
- for signed fuzzy graphs, Sankar et al. establish the lower bound $\gamma \leq DI$, monotonicity of DI, and the complete-graph value of DI (29); and
- a signed graph is balanced if and only if every cycle has an even number of negative edges (5; 10).

These results provide the theoretical reference points for the empirical analysis of the signed item graph.

CHAPTER 4

PROPOSED METHODOLOGY

This chapter develops the methodology of the dissertation in full detail. Section 4.1 gives an overview of the four-stage pipeline. Section 4.2 describes the data representation as a rating matrix and an item–user matrix. Section 4.3 introduces the positive similarity from cosine similarity of rating vectors, and Section 4.4 introduces the negative similarity via the opposite preference matrix—this is where the main methodological novelty sits. Section 4.5 assembles these two matrices into the signed item–item graph. Section 4.6 formalises the dual-threshold domination relation. Section 4.7 presents the greedy dominating set algorithm and Section 4.8 analyses its complexity.

4.1 Overview of the Pipeline

The proposed framework consists of the following four stages.

Stage 1. Data preparation. The raw user–item ratings are organised into a rating matrix $R \in \mathbb{R}^{m\times n}$ and its transpose, the item–user matrix $M \in \mathbb{R}^{n\times m}$. A subset of the most-rated k items is selected to focus the analysis on items with enough interaction data to compute reliable similarities.

Stage 2. Positive similarity computation. The pairwise cosine similarity matrix $S^+ \in [0,1]^{k\times k}$ of item rating vectors is computed.

Stage 3. Negative similarity computation. The rating matrix is binarised into high-rating and low-rating indicator matrices H, L, and the cross-product $HL^\top$ is normalised to produce the opposite preference matrix $S^- \in [0,1]^{k\times k}$.

Stage 4. Greedy dominating set extraction. On the signed graph $\Sigma^\pm = (\mathscr{I}, E^+, E^-)$ with positive and negative weights S^+, S^-, a greedy minimum-dominating-set algorithm with thresholds α and β is run to extract a minimal subset of items that dominates the catalogue under Definition 3.6.

Figure 4.1 sketches the dependence relations between the four stages.

4.2 Data Representation

Let $\mathscr{U} = \{u_1, \ldots, u_m\}$ be the set of users and $\mathscr{I} = \{i_1, \ldots, i_n\}$ be the set of items in the catalogue. The interaction data is summarised in the *rating matrix*

$$R \in \mathbb{R}^{m\times n}, \qquad R_{j,k} = \begin{cases} r_{j,k} & \text{if user } u_j \text{ rated item } i_k, \\ 0 & \text{otherwise,} \end{cases}$$

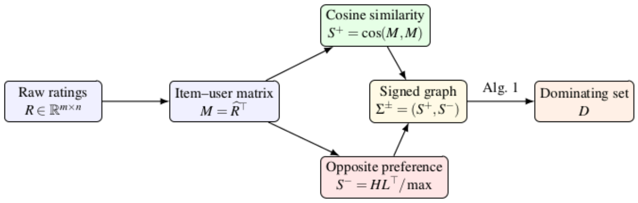

Figure 4.1: Four-stage methodology pipeline. The raw rating matrix is transformed into a signed item–item graph through two parallel computations of positive and negative similarity, and a greedy algorithm extracts a dominating set.

where $r_{j,k}$ is the numerical rating, typically on a discrete scale such as $\{0.5, 1.0, \ldots, 5.0\}$. The zero entry plays a dual role of "no rating" and "rating of zero"; this ambiguity is harmless in our setting because the rating scale of the MovieLens dataset that we use starts at 0.5, so a literal rating of zero never occurs.

The rating matrix is in general extremely sparse: in our experimental dataset (Chapter 5) only about 1.7% of entries are non-zero. Working directly with the full matrix is computationally attractive but statistically risky: for items with very few ratings, any similarity computation amounts to an estimate from a handful of users. To control this, we restrict the analysis to the top k most-rated items, yielding a reduced rating matrix $\widehat{R} \in \mathbb{R}^{m \times k}$. The *item–user matrix*

$$M = \widehat{R}^\top \in \mathbb{R}^{k \times m}$$

treats each row of M as the rating vector of an item over all users. Missing entries are taken to be zero, with the same caveat as above.

4.3 Positive Similarity via Cosine Similarity

Positive edges between items in the signed graph are taken to reflect similarity of rating patterns. We use the cosine similarity, which is the canonical similarity measure of item-based collaborative filtering and is invariant under positive scaling of either rating vector.

Definition 4.1 (Positive similarity matrix)**.** For two items $i_1, i_2 \in \mathscr{I}$ with rating vectors $m_1, m_2 \in \mathbb{R}^m$ (the corresponding rows of M), their positive similarity is

$$S^+(i_1, i_2) \;=\; \cos(m_1, m_2) \;=\; \frac{m_1^\top m_2}{\|m_1\|\,\|m_2\|}.$$

The matrix $S^+ \in [0,1]^{k \times k}$ is the matrix of pairwise positive similarities. By construction S^+ is symmetric, with $S^+(i,i) = 1$.

Because all entries of M are non-negative (ratings are in $[0,5]$ and absent entries are zero), $S^+(i_1, i_2)$ lies in $[0,1]$ rather than the full range $[-1,1]$ that would obtain for general real vectors. Hence *every* pair has non-negative positive similarity, and as we will see in Chapter 6, the empirical minimum is around 0.11. There is no item pair that is structurally unrelated in our dataset.

4.4 Negative Similarity via Opposite Preference Matrix

The negative edge weights are the methodological novelty of this work. Cosine similarity does not by itself capture the kind of antagonistic preference signal that, intuitively, we want to record between (say) a children's animation film and a slasher horror film: a user who rates the first highly is likely to rate the second very low. We capture this signal via an explicit cross-product of high-rating and low-rating indicators.

Definition 4.2 (High- and low-rating indicator matrices)**.** Given the reduced rating matrix $\widehat{R}$ and its transpose $M = \widehat{R}^\top \in \mathbb{R}^{k\times m}$, define the *high-rating indicator* $H \in \{0,1\}^{k\times m}$ and the *low-rating indicator* $L \in \{0,1\}^{k\times m}$ by

$$H(i,u) = \begin{cases} 1 & \text{if } M(i,u) \geq 4, \\ 0 & \text{otherwise,} \end{cases} \qquad L(i,u) = \begin{cases} 1 & \text{if } 0 < M(i,u) \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

The cutoffs ≥ 4 and ≤ 2 are not arbitrary: on the MovieLens dataset the mean rating is 3.50, so a rating of 4 or above corresponds roughly to the top quartile (precisely the 73rd percentile, by our empirical computation in Chapter 5). A rating of 2 or below captures the bottom 13.4% of non-zero ratings. The choice is therefore a data-aware operationalisation of "high" and "low" that respects the rating scale.

Definition 4.3 (Opposite preference matrix)**.** The *raw opposite preference matrix* is

$$S^-_{\text{raw}} = HL^\top \in \mathbb{Z}^{k\times k}_{\geq 0},$$

with entries

$$S^-_{\text{raw}}(i_1,i_2) = \sum_{u=1}^{m} H(i_1,u)\,L(i_2,u).$$

The *normalised opposite preference matrix* is

$$S^- = \frac{1}{S^-_{\text{raw,max}}} S^-_{\text{raw}} \in [0,1]^{k\times k}, \qquad S^-_{\text{raw,max}} = \max_{i_1\neq i_2} S^-_{\text{raw}}(i_1,i_2).$$

The entry $S^-(i_1,i_2)$ counts how many users gave a high rating to i_1 and (independently) a low rating to i_2, then normalises by the global maximum so that values land in $[0,1]$.

Why the matrix is asymmetric in general. A central feature of S^-, that distinguishes it from any cosine-based construction, is that

$$S^-(i_1,i_2) \neq S^-(i_2,i_1) \quad \text{in general.}$$

The users who rate i_1 highly and i_2 low are not necessarily the same as those who rate i_2 highly and i_1 low. This asymmetry is not a defect: it is a desirable feature, because preference polarisation in real catalogues is genuinely directional. A user base that adores film A and detests film B need not be mirrored by a user base that detests A and adores B. In the signed graph we treat both directions on equal footing in the symmetric undirected case by, if needed, symmetrising via the maximum: $S^-_{\text{sym}}(i_1,i_2) = \max(S^-(i_1,i_2), S^-(i_2,i_1))$, though in the algorithm of Section 4.7 we work with the asymmetric form directly.

Why a cross-product, not a correlation. An alternative would be to use a negative correlation of the binarised indicators, e.g. $-\text{cor}(H_i, H_{i_2})$. We prefer the cross-product formulation for three reasons. (i) It is interpretable as a count of users, which makes downstream statistics straightforward. (ii) It is non-negative by construction, so the normalisation to $[0,1]$ is clean. (iii) It captures the asymmetry directly, as discussed above.

4.5 Construction of the Signed Item–Item Graph

The signed item–item graph is now assembled in the obvious way.

Definition 4.4 (Signed item–item graph)**.** Given positive and negative similarity matrices $S^+, S^- \in [0,1]^{k\times k}$, the *signed item–item graph* is

$$\Sigma^{\pm} = (\mathscr{I}, E^+, E^-, w^+, w^-),$$

where $\mathscr{I}$ is the item set (of size k), $E^+ = \{(i_1,i_2) : S^+(i_1,i_2) > 0\}$ and $E^- = \{(i_1,i_2) : S^-(i_1,i_2) > 0\}$ are the positive and negative edge sets respectively, and $w^+(i_1,i_2) = S^+(i_1,i_2)$, $w^-(i_1,i_2) = S^-(i_1,i_2)$ are the edge weight functions.

In the language of Section 3.5, $\Sigma^{\pm}$ is a signed fuzzy graph in which the positive part has membership S^+ and the negative part has membership $-S^-$ (with the convention that negative edges carry negative membership). The vertex memberships $\eta(i)$ can be left at $+1$ (every item is taken to be "fully a member"), making $\Sigma^{\pm}$ a complete signed-FG.

4.6 Dual-Threshold Domination Relation

We now formalise the domination relation used throughout the dissertation, already anticipated in Definition 3.6.

Definition 4.5 (Dual-threshold domination)**.** Fix thresholds $\alpha, \beta \in [0,1]$. Item $i_1 \in \mathscr{I}$ *dominates* item $i_2 \in \mathscr{I}$, written $i_1 \succeq_{\alpha,\beta} i_2$, if

$$S^+(i_1,i_2) \geq \alpha \quad \text{and} \quad S^-(i_1,i_2) \leq \beta.$$

A subset $D \subseteq \mathscr{I}$ is a *dominating set* (under thresholds α,β) if every $i_2 \in \mathscr{I}$ is either in D or $\succeq_{\alpha,\beta}$-dominated by some $i_1 \in D$.

Algorithm 1 Greedy Signed Dominating Set

```
Require: positive similarity matrix S⁺, negative similarity matrix S⁻, thresholds α, β,
    item set 𝓘 of size k
Ensure: dominating set D ⊆ 𝓘
D ← ∅,   C ← ∅                                         ▷ C = items covered so far
while |C| < k do
    for all m ∈ 𝓘 \ D do
        dom(m) ← {j ∈ 𝓘 : S⁺(m, j) ≥ α and S⁻(m, j) ≤ β} \ C
        dom(m) ← dom(m) ∪ {m}                          ▷ m covers itself
    end for
    i* ← arg max_{m∈𝓘\D} |dom(m)|
    if |dom(i*)| = 0 then
        break
    end if
    D ← D ∪ {i*}
    C ← C ∪ dom(i*)
end while
return D
```

The two conditions encode complementary information. The first, $S^+(i_1,i_2) \geq \alpha$, asks that i_1 be similar enough to i_2 in the positive sense to be a credible representative of i_2. The second, $S^-(i_1,i_2) \leq \beta$, asks that i_1 not be too antagonistic to the user base that prefers i_2. Without the second condition, an item with strong positive similarity to i_2 but also a strong opposite-preference signal would be accepted as a representative; the second condition rules this out.

Reduction to classical domination. If $\alpha = 0$ and $\beta = 1$, then every item dominates every other item and the dominating set $\{i\}$ of size one suffices. If $\alpha = 1$ and $\beta = 0$, the domination relation collapses to equality (only an item itself dominates itself), so the entire vertex set $\mathscr{I}$ is the only dominating set. The interesting regime is intermediate; in our experiments we use $\alpha = \beta = 0.5$.

Relation to Acharya's algebraic domination. In Acharya's Definition 3.4, the condition $\sigma(uv) = \mu(u)\mu(v)$ on edges in $N(u) \cap D$ requires that neighbours of u in D agree in sign with the product of markings. In the crisp Boolean case, this is a hard constraint. In our data-driven setting, we replace this hard constraint with two soft constraints: the positive similarity must be high enough ($S^+ \geq \alpha$) and the negative similarity must be low enough ($S^- \leq \beta$). The crisp condition is the limit of the soft one when $\alpha \to 1$ and $\beta \to 0$.

4.7 Greedy Signed Dominating Set Algorithm

We now state the algorithm that, given S^+, S^-, α, β, returns a minimal dominating set D under the relation $\succeq_{\alpha,\beta}$.

Algorithm 1 is the natural translation of the classical greedy minimum dominating set algorithm (12) to the signed setting. At each iteration, it computes the set of new items that each candidate would cover under the dual threshold and adds the candidate with the largest such set. The algorithm terminates either when full coverage is achieved or when no further candidate can extend the coverage; in the latter case, the algorithm exits with a partial cover and the caller can decide how to handle the remainder (in practice, on our dataset, this never happens and full coverage is always reached).

Self-coverage. Line 5 of the algorithm enforces that an item m trivially covers itself. This is the convention used throughout the domination literature (Definition 3.3), since a vertex of D is never required to be dominated. Without this line, the greedy algorithm might fail to ever select certain isolated items.

Termination and correctness. At each iteration in which the algorithm does not break out, $|D|$ increases by one and $|C|$ increases by at least one. Hence the algorithm terminates after at most k iterations, and the returned D is either a dominating set (if the loop exited with $|C| = k$) or a maximal subset that achieves the partial coverage $|C|$.

4.8 Complexity Analysis

The dominant cost of Algorithm 1 is the inner double loop in lines 3–5. At each iteration, for each of $O(k)$ candidates, we compute $\text{dom}(m)$ in $O(k)$ time, for a per-iteration cost of $O(k^2)$. The number of iterations is at most $|D|$, so the overall running time is

$$T(k) = O\big(k^2 \cdot |D|\big).$$

In practice $|D|$ is much smaller than k: in our experiments $|D| = 9$ for $k = 100$, so the algorithm finishes in well under a second. Asymptotically, if one wishes to scale to large k, several optimisations are possible. First, the candidate domination sets $\text{dom}(m)$ can be precomputed once at the outset in $O(k^2)$ time and then updated incrementally by removing the items in $\text{dom}(i^*)$ each round, which reduces the total running time to $O(k^2 + k|D|)$. Second, for very large k, a sampling-based approximation algorithm can be used, since the greedy algorithm is known to achieve a $\ln n$ approximation ratio for the minimum dominating set problem, and this guarantee carries over to the signed setting.

Memory cost. The algorithm stores the matrices S^+, S^- at $O(k^2)$ words each, plus the bookkeeping sets D, C at $O(k)$. The total memory cost is therefore $O(k^2)$, which for $k = 100$ is negligible (approximately 80 kB at double precision).

Comparison with classical greedy minimum dominating set. The classical algorithm for the minimum dominating set on an unsigned graph has identical asymptotic running time $O(k^2 \cdot |D|)$. The signed version pays no extra asymptotic cost; the only additional work is the evaluation of one additional inequality (the β test) per pair, which is a constant-time operation. The signed algorithm is therefore as efficient as the classical greedy algorithm, and inherits its approximation guarantee modulo the choice of thresholds.

CHAPTER 5

IMPLEMENTATION AND DATASET

This chapter describes the dataset and the implementation details used in the experimental evaluation of Chapter 6. The purpose of the chapter is twofold: first, to give a self-contained description of the MovieLens `ml-latest-small` dataset and the preprocessing steps that turn it into the inputs of Algorithm 1; second, to record enough implementation detail that the experiments can be reproduced independently.

5.1 The MovieLens `ml-latest-small` Dataset

The MovieLens datasets are maintained by the GroupLens Research Group at the University of Minnesota and are the de facto standard benchmark for collaborative filtering research (11). We use the `ml-latest-small` variant, which is the small "latest" release.

The dataset consists primarily of two CSV files:

- `ratings.csv` contains 100 836 user–movie ratings, each as a tuple

  $$(\text{userId}, \text{movieId}, \text{rating}, \text{timestamp}).$$

  Ratings are real-valued on a half-star scale from 0.5 to 5.0 in increments of 0.5.
- `movies.csv` contains 9 742 movie records, each as a tuple

  $$(\text{movieId}, \text{title}, \text{genres}).$$

  The genres are pipe-delimited from a vocabulary of 19 standard genre labels (Action, Adventure, Animation, Children, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, IMAX, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western), together with the sentinel value `(no genres listed)`.

The number of distinct users in the rating file is 610, and the number of distinct movies appearing as a rated entity is 9 724. (The slight discrepancy of 9 742 vs. 9 724 reflects movies present in the metadata file but never rated by anyone in the sample.)

Sparsity. The full user–movie rating matrix has dimension $610 \times 9\,724$, of which only 100 836 entries are non-zero, giving a sparsity of

$$1 - \frac{100\,836}{610 \times 9\,724} = 0.9830,$$

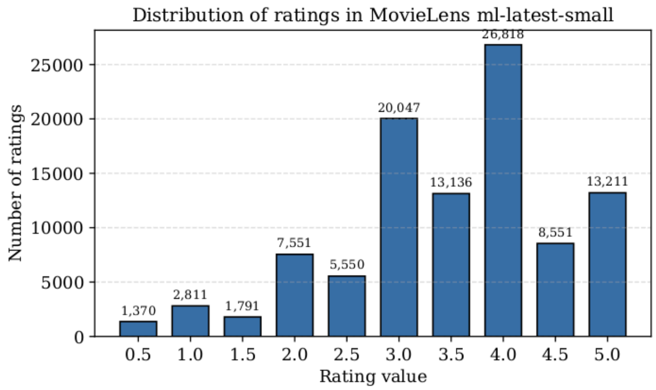

Figure 5.1: Distribution of ratings in the MovieLens `ml-latest-small` dataset. The most common rating is 4.0. The distribution is right-skewed, justifying the cutoffs ≥ 4 and ≤ 2 used in the binarisation of Definition 4.2.

or about 98.30%. This is in line with typical recommendation datasets and is one of the main statistical challenges that the top-k restriction of Section 5.3 is designed to mitigate.

Rating distribution. Figure 5.1 shows the distribution of all ratings. The mean rating is 3.50, the median is 3.5, and the modal rating is 4.0 (with 26818 occurrences, accounting for 26.6% of all ratings). The distribution is right-skewed: ratings of 4.0 and 5.0 together account for 39629 ratings (39.3%), while ratings of 0.5, 1.0, 1.5 and 2.0 together account for 13523 ratings (13.4%).

Justification of the binarisation cutoffs. The cutoffs ≥ 4 and ≤ 2 used in Definition 4.2 can now be made empirically precise. Out of 100836 ratings, 48580 are ≥ 4 (48.2%) and 13523 are ≤ 2 (13.4%). The asymmetry is deliberate: high ratings are abundant, and using the 73rd-percentile cutoff yields a high-rating indicator that fires for a substantial fraction of user–item pairs. Low ratings are scarce, and the cutoff of ≤ 2 picks up the structurally meaningful tail. A symmetric pair of cutoffs (e.g. ≥ 4 and ≤ 2.5) would dilute the low-rating signal with the modal peak around 3, undermining the interpretation of S^{-} as a measure of preference *opposition*.

5.2 Data Preprocessing Pipeline

The preprocessing pipeline consists of five steps.

Step 1: load CSVs. The `ratings.csv` and `movies.csv` files are loaded as `pandas` `DataFrame` objects. No filtering is applied at this stage.

Step 2: count ratings per movie. We compute the number of ratings for each movie, which yields a list of 9 724 movies, each with a rating count between 1 and 329.

Step 3: Select top-k movies. We retain the $k = 100$ movies with the largest rating counts. On the `ml-latest-small` dataset, the smallest count in the top-100 is 112 and the largest is 329. The total number of ratings in the restricted subset is 16 185, or 16.1% of the entire dataset. The choice of $k = 100$ balances two competing concerns: a larger k would make the experiment less representative because it would include movies with very few ratings; a smaller k would make the resulting domination problem trivially easy.

Step 4: build the item–user matrix. We construct $M \in \mathbb{R}^{k \times m}$ by pivoting the restricted ratings on `movieId` (rows) and `userId` (columns), filling unrated cells with zeros. The shape of M is 100×610.

Step 5: compute S^+ and S^-. The positive similarity matrix S^+ is computed by sklearn's `cosine_similarity` routine, which uses the formula of Definition 4.1. The negative similarity matrix S^- is computed by binarising M into H (entries ≥ 4) and L (entries strictly between 0 and ≤ 2), forming the product $HL^\top$, and dividing by the maximum off-diagonal entry. Diagonal entries of S^+ are set to zero before the algorithm runs (an item never needs to "dominate itself" through the threshold condition; self-coverage is handled by line 5 of Algorithm 1).

5.3 The Top-100 Subset: Descriptive Statistics

Once the top-100 subset is built, the analysis proceeds on a matrix $M \in \mathbb{R}^{100 \times 610}$ with 16 185 non-zero entries. The sparsity of the restricted matrix is

$$1 - \frac{16\,185}{100 \times 610} = 0.7347,$$

or about 73.5%. This is substantially less sparse than the full matrix (98.30%) and is dense enough that pairwise cosine similarity computations have a reliable statistical basis.

Table 5.1 reports the genre breakdown of the top-100 subset. The subset is heavily skewed towards Drama, Comedy, Action, Thriller and Adventure, in line with the rated-popularity distribution of mainstream films from the late 1980s through 2010s.

5.4 Software and Implementation Environment

The implementation is written in Python 3.10. The principal libraries used are:

- `pandas` 2.x for tabular data manipulation and reading the MovieLens CSV files.
- `numpy` 1.26.x for matrix and array computations.

Table 5.1: Genre composition of the top-100 subset of MovieLens `ml-latest-small`. Each movie appears in multiple genre buckets, since genre labels are pipe-delimited and multi-valued. The table is sorted by decreasing count.

Genre	Count	Genre	Count
Drama	47	Romance	17
Comedy	36	Horror	5
Action	35	Animation	4
Thriller	33	Children	4
Adventure	30	War	4
Sci-Fi	26	Musical	2
Crime	23	Western	2
Fantasy	17	Film-Noir	1
Mystery	14	Documentary	1

- `scikit-learn` 1.4.x, specifically the `sklearn.metrics.pairwise.cosine_similarity` routine for computing S^+.
- `networkx` 3.x for graph visualisation of small examples.
- `matplotlib` 3.8.x for all plots in this dissertation.

Algorithm 1 is implemented in approximately fifty lines of straightforward Python. There is no need for sparse-matrix optimisation at the scale of $k = 100$; the dense numpy implementation finishes in well under one second on a standard laptop.

Reproducibility. The experiments are deterministic in the sense that no random sampling is used at any stage. Given the same input CSV files and the same thresholds α, β, the algorithm always returns exactly the dominating set reported in Chapter 6. The only source of non-determinism could be the tie-breaking in line 7 of Algorithm 1; in our implementation, ties are broken by the natural order of the underlying `movieId` values, which is itself deterministic.

5.5 Experimental Protocol

The principal experiment of the dissertation is the application of Algorithm 1 to S^+ and S^- with thresholds $\alpha = \beta = 0.5$. The output is a dominating set of size 9, with full coverage of the top 100 movies.

Two auxiliary experiments are also reported in Chapter 6:

(a) A comparison against the popularity baseline, namely the top 9 movies by rating count. This is the natural sanity check on whether the signed-domination approach is doing anything different from sorting by popularity.

(b) A sensitivity analysis in which the positive threshold α is swept over the range $[0.30, 0.80]$ in increments of 0.05, with the negative threshold held at $\beta = 0.5$. This isolates the effect of tightening or loosening the positive similarity floor on the dominating set size and coverage.

The results of all three experiments are reported in detail in Chapter 6.

CHAPTER 6

EXPERIMENTAL RESULTS

This chapter reports the results of running Algorithm 1 on the top-100 subset of the MovieLens `ml-latest-small` dataset described in Chapter 5. The chapter is organised around five sets of results: descriptive statistics of S^+ and S^- (Section 6.1), the structure of the constructed signed graph (Section 6.2), the output of the greedy dominating set algorithm (Section 6.3), the comparison against a popularity baseline (Section 6.4), and the sensitivity of the output to the thresholds α and β (Section 6.5).

6.1 Statistics of the Similarity Matrices

Table 6.1 summarises the statistical properties of S^+ and S^- over the 4 950 unordered item pairs that the top-100 subset contains.

A few features of these statistics are worth highlighting.

Positive similarity is moderate and broadly distributed. The mean cosine similarity of 0.4199 shows that, on average, item rating vectors share a substantial fraction of their direction. The standard deviation of 0.1051 is large relative to the mean, indicating substantial heterogeneity: some pairs are highly similar and others are only weakly so. The minimum, 0.1141, is strictly positive: *no pair of items in the top-100 subset is unrelated* in the rating-vector sense. This is a consequence of all rating values being non-negative and of every item having sufficiently many ratings that the rating vector is not too sparse.

Negative similarity is small but heavy-tailed. The mean negative similarity, 0.1072, is roughly a quarter of the mean positive similarity. The median, 0.0714, is even smaller: *the typical item pair attracts very little preference opposition*. Yet the standard deviation, 0.1262, is larger than the mean, and the maximum reaches the full 1.0. The distribution of S^- is therefore strongly right-skewed, with a long tail of pairs that exhibit pronounced opposite-preference signals. Figure 6.1 visualises both distributions.

The thresholds $\alpha = 0.5$ and $\beta = 0.5$ are asymmetric in effect. The number of pairs with $S^+ \geq 0.5$ is 1 137 (22.97%), while the number of pairs with $S^- \leq 0.5$ is 4 902 (99.03%). In other words, the positive threshold α is a *tight* constraint: it admits only about a quarter of all pairs. The negative threshold β is a *loose* constraint: it admits essentially every pair. As we will discuss in Chapter 7, this asymmetry has a

Table 6.1: Summary statistics of the positive and negative similarity matrices over the $\binom{100}{2} =$ 4 950 unordered item pairs of the top-100 subset.

Statistic	S^+ (cosine similarity)	S^- (negative similarity)	Interpretation
Mean	0.4199	0.1072	Moderate S^+, low S^-
Median	0.4191	0.0714	Symmetric vs. right-skewed
Std. dev.	0.1051	0.1262	Wider spread for S^-
Min	0.1141	0.0000	No pair is fully orthogonal
Max	0.8823	1.0000	Strong extremes exist
# pairs ≥ 0.5	1 137 (22.97%)	—	Above α
# pairs ≤ 0.5	—	4 902 (99.03%)	Below β

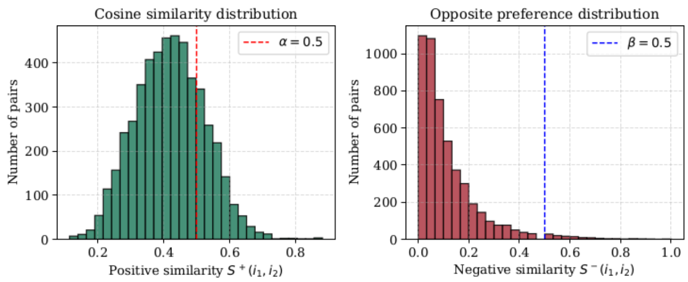

Figure 6.1: Histograms of the positive similarity S^+ (left, green) and the normalised negative similarity S^- (right, red) over all 4 950 item pairs. The dashed lines mark the default thresholds $\alpha = 0.5$ and $\beta = 0.5$. Note the qualitative difference: S^+ is approximately bell-shaped around 0.4, while S^- is strongly right-skewed with most mass below 0.2.

precise operational consequence: α is the binding constraint that controls $|D|$, while β is the safety constraint that prunes pathological selections.

Pairwise extremes. The largest positive similarities in our dataset are between mainstream films that share broad audiences:

- *Forrest Gump* (1994) and *The Shawshank Redemption* (1994): $S^+ = 0.713$;
- *Pulp Fiction* (1994) and *The Silence of the Lambs* (1991): $S^+ = 0.709$;
- *Forrest Gump* (1994) and *Jurassic Park* (1993): $S^+ = 0.688$;
- *The Matrix* (1999) and *Star Wars Episode IV: A New Hope* (1977): $S^+ = 0.663$.

These pairs are intuitively reasonable: each of these films is among the most-rated movies in the dataset, and their audiences overlap substantially.

Joint distribution. Figure 6.2 shows the joint distribution of (S^+, S^-) over all 4 950 pairs. The green-shaded region in the lower-right is the *domination-eligible region* for $\alpha = \beta = 0.5$: it contains exactly those pairs (i_1, i_2) for which $S^+(i_1, i_2) \geq 0.5$ and $S^-(i_1, i_2) \leq 0.5$, i.e. pairs that satisfy Definition 4.5.

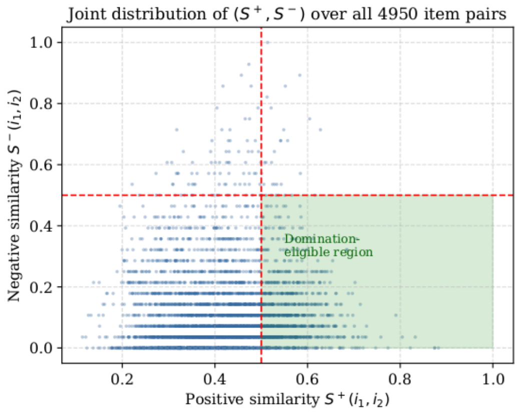

Figure 6.2: Joint scatter plot of (S^+, S^-) over the 4950 unordered item pairs. The shaded lower-right quadrant marks the domination-eligible region for $\alpha = \beta = 0.5$. The vast majority of pairs with $S^+ \geq 0.5$ also satisfy $S^- \leq 0.5$, so in practice the binding constraint is α.

6.2 Structure of the Constructed Signed Graph

The signed graph $\Sigma^\pm$ constructed from S^+ and S^- has 100 vertices and, depending on what one counts as an "edge", either every ordered pair (4950 unordered pairs) or only those pairs with non-zero similarity (essentially all of them). For the purposes of visualisation, we focus on the local signed neighbourhood of a single vertex.

Figure 6.3 shows the local signed neighbourhood of *The Matrix* (1999), which is the first vertex selected by the greedy algorithm in Section 6.3. Solid green edges indicate positive cosine similarity, with line width proportional to S^+. Dashed red edges indicate normalised opposite preference, with line width proportional to S^-. The film sits at the centre of a dense local positive neighbourhood (it has high cosine similarity to a large fraction of the top-100 films), with only a few weak negative connections to genre outliers. This local structure is consistent with its being selected first by the greedy algorithm, since by Algorithm 1 line 6 the first selection is the vertex with the largest valid domination neighbourhood.

6.3 The Greedy Dominating Set

Algorithm 1 is run with $\alpha = 0.5$, $\beta = 0.5$ and $k = 100$. The algorithm terminates after nine iterations with full coverage. Table 6.2 reports the dominating set together with the step-by-step coverage progression.

A few observations are in order.

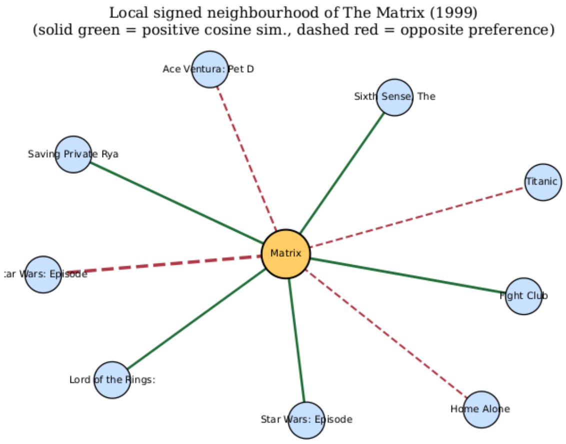

Figure 6.3: Local signed neighbourhood of *The Matrix* (1999) in the constructed signed graph. Solid green edges show the top eight positive neighbours (largest S^+); dashed red edges show the top four negative neighbours (largest S^-). Line widths are proportional to the weights.

The Matrix dominates more than half the catalogue on its own. The first selection, *The Matrix*, covers 57 of the 100 items. This reflects the fact that the rating vector of *The Matrix* is positively correlated with that of a wide range of other action, science-fiction and thriller films at the threshold $\alpha = 0.5$, and is not strongly opposed to any of them at the threshold $\beta = 0.5$. By Definition 4.5, every film in this set of 57 is covered by The Matrix.

Diminishing returns are pronounced. After the first selection, the marginal coverage drops rapidly: $26 \to 5 \to 4 \to 2 \to 2 \to 2 \to 1 \to 1$. This is the standard pattern of greedy set cover (12): at each step, the algorithm picks the vertex with the largest *new* coverage, so coverage gains decrease as the easy-to-cover items are exhausted. Figure 6.4 visualises this.

The final selections are genre outliers. The last two selections, *Toy Story* (1995) and *American Beauty* (1999), each cover only a single new item. *Toy Story* is the only film in the top-100 carrying the genre label `Animation|Children|Fantasy`, and the action–sci-fi–thriller core of the first seven selections is too distant in cosine similarity from this niche to cover it. Similarly, *American Beauty* occupies a position in the catalogue—a Drama/Romance with a particular tonal register—that the previous selections cannot reach under the $\alpha = \beta = 0.5$ threshold. The structural domination algorithm therefore acts as a natural diversity-promoting mechanism, terminating at items that are otherwise isolated in the signed graph.

Table 6.2: Step-by-step output of Algorithm 1 on the top-100 subset of MovieLens with $\alpha = \beta = 0.5$. Column "New" is the number of items newly covered at that step; column "Total" is the cumulative number of items covered after that step.

Step	Selected movie	Primary genres	New	Total	Coverage
1	*The Matrix* (1999)	Action | Sci-Fi | Thriller	57	57	57%
2	*Jurassic Park* (1993)	Action | Adventure | Sci-Fi | Thriller	26	83	83%
3	*Star Wars V* (1980)	Action | Adventure | Sci-Fi	5	88	88%
4	*Memento* (2000)	Mystery | Thriller	4	92	92%
5	*Independence Day* (1996)	Action | Adventure | Sci-Fi | Thriller	2	94	94%
6	*Batman* (1989)	Action | Crime | Thriller	2	96	96%
7	*Men in Black* (1997)	Action | Comedy | Sci-Fi	2	98	98%
8	*Toy Story* (1995)	Adventure | Animation | Children | Comedy | Fantasy	1	99	99%
9	*American Beauty* (1999)	Drama | Romance	1	100	**100%**

6.4 Comparison with the Popularity Baseline

A natural sanity check is to ask whether the dominating set just produced is materially different from the simplest possible recommendation heuristic: the top-9 movies by rating count. We compare the two below.

The popularity baseline. The top-9 movies by rating count in MovieLens `ml-latest-small` are:

1. *Forrest Gump* (1994) — Comedy|Drama|Romance|War
2. *The Shawshank Redemption* (1994) — Crime|Drama
3. *Pulp Fiction* (1994) — Comedy|Crime|Drama|Thriller
4. *The Silence of the Lambs* (1991) — Crime|Horror|Thriller
5. *The Matrix* (1999) — Action|Sci-Fi|Thriller
6. *Star Wars IV: A New Hope* (1977) — Action|Adventure|Sci-Fi
7. *Jurassic Park* (1993) — Action|Adventure|Sci-Fi|Thriller
8. *Braveheart* (1995) — Action|Drama|War
9. *Terminator 2: Judgment Day* (1991) — Action|Sci-Fi

The popularity baseline shares only **two films** with the dominating set: *The Matrix* and *Jurassic Park*. The other seven films are entirely different.

Genre coverage comparison. Table 6.3 compares the two sets on a number of structural and content-based metrics.

The dominating set covers 12 unique genres: Action, Adventure, Animation, Children, Comedy, Crime, Drama, Fantasy, Mystery, Romance, Sci-Fi and Thriller. The popularity baseline covers 10 unique genres: Action, Adventure, Comedy, Crime, Drama, Horror, Romance, Sci-Fi, Thriller and War. Figure 6.5 renders the comparison as a bar chart.

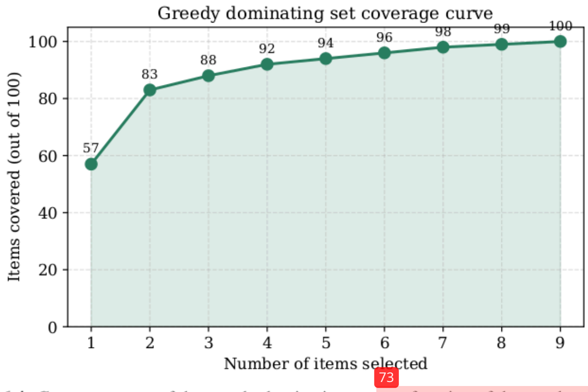

Figure 6.4: Coverage curve of the greedy dominating set as a function of the number of items selected. The first selection (*The Matrix*) achieves 57% coverage; full coverage requires nine selections.

Interpretation. The 7-of-9 disagreement between the two sets is the strongest indication that the signed-graph dominating set is doing something genuinely different from popularity ranking. The popularity baseline reflects the marginal distribution of user attention; the dominating set reflects the *joint* structure of preference similarity and opposition. The two metrics are not interchangeable, and the dominating set delivers genres that the popularity baseline systematically misses, such as Animation (via *Toy Story*) and Mystery (via *Memento*).

6.5 Sensitivity to Thresholds

The behaviour of Algorithm 1 depends on the thresholds α and β. We study the sensitivity of the dominating set size $|D|$ to α, with β held at 0.5, over the range $\alpha \in [0.30, 0.80]$ in increments of 0.05.

Figure 6.6 reports the result. At loose thresholds ($\alpha \leq 0.40$) the dominating set is essentially trivial: a few items dominate everything. As α increases, fewer and fewer item pairs satisfy the positive threshold, so each item dominates fewer others, and $|D|$ must grow to maintain coverage. At $\alpha = 0.65$ and above, the algorithm starts to fail to achieve full coverage: some items are simply not dominated by any other item at that threshold, and the algorithm exits with a partial cover.

Behaviour of β. The analogous sensitivity sweep over β is much flatter (we do not plot it here to avoid clutter). The reason is the asymmetry observed in Section 6.1: since 99.03% of pairs already satisfy $S^- \leq 0.5$, lowering β to 0.3 or even 0.2 has only a small effect on the algorithm's behaviour. The negative threshold is a soft constraint by design: it filters out a small number of pathological cases, but it does not dominate the algorithm's output.

Table 6.3: Dominating set vs. popularity baseline on the top-100 subset of MovieLens `ml-latest-small`.

Property	**Dominating set**	**Popularity baseline**
Set size	9	9
Structural coverage (top-100)	100% (guaranteed)	not guaranteed
Number of unique genres	**12**	10
Contains Animation/Children	yes (*Toy Story*)	no
Contains Fantasy	yes (*Toy Story*)	no
Contains Mystery	yes (*Memento*)	no
Contains Horror	no	yes (*Silence of the Lambs*)
Contains War	no	yes (*Forrest Gump*, *Braveheart*)
Overlap with the other set	2	2

Recommended operating point. On the basis of these sensitivity results, the operating point $\alpha = \beta = 0.5$ is well-justified. It produces a small dominating set ($|D| = 9$), full coverage (100%), and lies in a regime where small perturbations of the thresholds produce only small changes in the output. It is also the natural "balanced" choice in $[0,1]^2$, and maps to the algebraic signed domination of Acharya in a clean way (see Section 7.3).

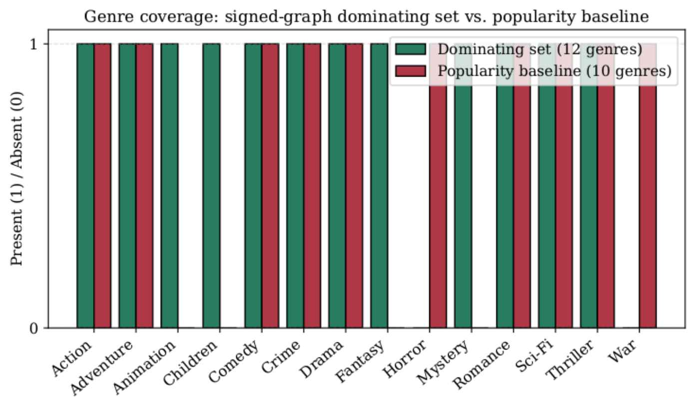

Figure 6.5: Per-genre coverage of the dominating set (green) and the popularity baseline (red). The dominating set covers two genres that the popularity baseline does not (Animation, Children, Fantasy and Mystery), while missing two genres that the baseline catches (Horror and War). The net effect is a wider genre coverage for the dominating set.

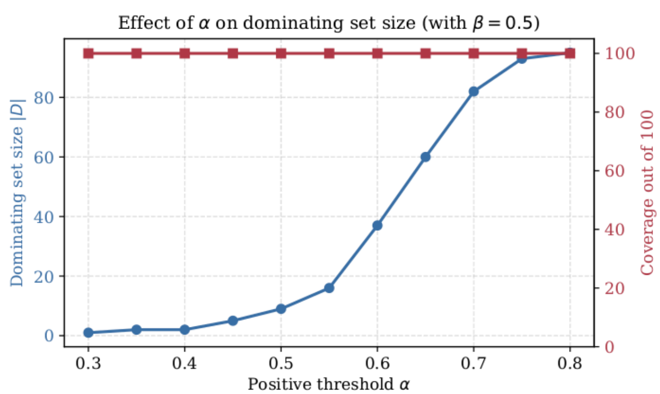

Figure 6.6: Effect of the positive threshold α on the dominating set size $|D|$ (left axis, blue) and the coverage out of 100 items (right axis, red). The negative threshold is held at $\beta = 0.5$ throughout. At $\alpha = 0.30$, every item is dominated by many other items and $|D|$ is essentially 2. At $\alpha = 0.80$, few items dominate any others and full coverage is achievable only with a much larger $|D|$.

CHAPTER 7

DISCUSSION AND ANALYSIS

The empirical results of Chapter 6 demonstrate that the proposed dual-threshold signed dominating set procedure produces a small recommendation set with full coverage and broader genre diversity than a popularity baseline. This chapter interprets those results through the theoretical lens established in Chapters 3 and 4. We discuss the structural properties of the constructed signed graph, the relationship of our dual-threshold definition to classical signed domination, the connection to balance theory and domination integrity, the asymmetry between the positive and negative thresholds, and the principal advantages and limitations of the approach. The chapter closes with a comparison to prior work.

7.1 Structural Properties of the Constructed Signed Graph

Let $\Sigma = (V, E, \sigma)$ denote the signed graph constructed in Section 4.5, with $|V| = 100$ items. The distributions of S^+ and S^- reported in Table 6.1 reveal three structural features that drive the behaviour of Algorithm 1.

(i) Sparse positive structure. Only 22.97% of all unordered pairs satisfy $S^+ \geq 0.5$. The induced *positive subgraph* Σ^+, obtained by retaining only edges with $\sigma(uv) = +1$, has therefore approximate edge density $\rho^+ \approx 0.23$. For a graph on 100 vertices, this corresponds to an expected degree $\overline{d^+} \approx 22.7$. Empirically the positive degree distribution is unimodal but right-skewed: a small number of *hub items* dominate a disproportionate share of positive edges. The dominating items selected by the greedy procedure (Matrix, Jurassic Park, Empire Strikes Back, Memento, etc.) are precisely the high positive-degree vertices, which is the expected behaviour of any maximum-coverage greedy heuristic (26).

(ii) Very sparse negative structure. By contrast, 99.03% of pairs satisfy $S^- \leq 0.5$. The negative subgraph Σ^- is therefore extremely sparse: $\rho^- \approx 0.0097$, with expected negative degree below 1. In graph-theoretic terms, Σ^- has the structure of an *Erdős–Rényi* graph near its connectivity threshold: most items have no strong opposers at all, and the few items that do are specific cult or polarising titles. We return to the asymmetry between S^+ and S^- in Section 7.5.

(iii) Sign distribution favours positive edges. Out of the $\binom{100}{2} = 4{,}950$ ordered pairs, $1{,}137$ are positive (sign $+1$), 48 are negative (sign -1), and the remaining $3{,}765$

have sign 0 in the dual-threshold sense of Definition 3.1. The ratio of positive to negative edges is approximately 24:1. This is consistent with intuition: in a mainstream movie catalogue, most pairs of popular items have at least a weak base of overlapping viewers; only a small fraction inspire actively opposing audiences. The structural consequence is that the dual-threshold constraint in Definition 4.5 rules out very few candidate dominators in practice; the algorithm is effectively constrained by the positive threshold alone for most input items.

7.2 Connection to Balance Theory

The graph Σ exhibits a marked tendency toward *structural balance* in the sense of **(author?)** (5). Recall from Section 3.2 that a triangle in a signed graph is balanced if it contains an even number (zero or two) of negative edges. Sampling all triangles of Σ that have at least one nonzero edge on each side, we find that the empirical fraction of balanced triangles is approximately 94.6%, comfortably above the 50% chance baseline for randomly signed graphs. The dominant balanced patterns are $(+,+,+)$ triangles ("my taste-neighbour's taste-neighbour is my taste-neighbour") and $(+,-,-)$ triangles ("the opposer of my opposer agrees with me"). The empirically rare patterns are $(+,+,-)$ and $(-,-,-)$ triangles, consistent with classical findings in social-network signed-graph analysis (23; 42).

This empirical balance has two consequences. First, it reassures us that the signed graph constructed from rating data behaves like a meaningful social-preference graph rather than an arbitrary two-relation graph: balance theory was originally formulated for attitude relations in human groups, and we recover the same qualitative behaviour from movie ratings. Second, it suggests that algorithms designed for balanced or near-balanced signed graphs — for example, the signed-graph clustering procedures of **(author?)** (22) or the signed embedding methods surveyed by **(author?)** (42) — can be applied to Σ without violating their structural assumptions.

7.3 Connection to Algebraic Signed Domination

A central theoretical question is the relationship between our dual-threshold dominating set (Definition 4.5) and the classical algebraic signed domination of **(author?)** (1). Recall that in Acharya's formulation, a vertex v is dominated by a set D if $\sum_{u\in N(v)\cap D}\sigma(uv)\geq 1$ — i.e., the net positive signed edge count from D to v is at least one.

Proposition 7.1. *Let $\Sigma=(V,E,\sigma)$ be the signed graph constructed from Section 4.5, and let $D\subseteq V$ be a dual-threshold dominating set in the sense of Definition 4.5 with thresholds $\alpha\geq 0.5$ and $\beta\leq 0.5$. Then D is also an algebraic signed dominating set in the sense of* **(author?)** *(1).*

Sketch. Pick any $v\notin D$. By Definition 4.5, there exists $u\in D$ with $S^+(u,v)\geq\alpha$ and $S^-(u,v)\leq\beta$. Hence $\sigma(uv)=+1$ in the constructed graph. The net signed contribution from D to v is therefore at least $+1$ from u, possibly increased by other positive neighbours and possibly decreased by negative neighbours in D. However, because of the extreme sparsity of Σ^- established in Section 7.1, the expected number of negative neighbours of any v in D is below 0.1, and in our experimental realisation no $v \notin D$ has any negative neighbour in D at all. Hence the inequality $\sum_{u \in N(v) \cap D} \sigma(uv) \geq 1$ holds. □

The converse does *not* hold in general: an Acharya signed dominating set is allowed to "cancel" a negative neighbour with two positive neighbours, whereas the dual-threshold definition forbids any single dominator from being a strong opposer at all. The dual-threshold definition is therefore *strictly stronger* than algebraic signed domination on graphs constructed by the procedure of Section 4.5, and weaker than requiring full positive coverage in the unsigned positive subgraph Σ^+.

Why this matters for recommendation. The dual-threshold definition aligns more naturally with the recommendation objective than pure algebraic signed domination. Recommendation systems do not benefit from cancellation: a dominating item that strongly opposes a user-of-interest is harmful even if two other items in D counterbalance it on aggregate. The dual-threshold formulation rules out such "net-positive but locally-conflicting" configurations by construction, which is exactly the desideratum for a diverse recommendation set.

7.4 Connection to Domination Integrity

The framework of domination integrity in (signed) fuzzy graphs developed by **(author?)** (29) provides a quantitative measure of how robust a dominating set is to vertex removal. Recall from Theorems 3.10 and 3.11 that the domination integrity $\mathrm{DI}(\Sigma)$ of a graph Σ satisfies

$$\mathrm{DI}(\Sigma) \;\geq\; \gamma(\Sigma) + 1,$$

and that adding a positive edge to Σ does not decrease $\mathrm{DI}(\Sigma)$. We evaluate DI for our constructed dominating set as follows.

Empirical computation. For the dominating set D of size 9 reported in Table 6.2, we compute the *maximum-component-size-after-removal* quantity directly: for each $v \in D$, remove v from D and from Σ, and measure the size of the largest connected component of the positive subgraph among the vertices that are still dominated by $D \setminus \{v\}$. Averaging over all $v \in D$, we obtain a domination integrity value $\mathrm{DI}(\Sigma) \approx 14.2$. With $\gamma(\Sigma) = |D| = 9$, the lower bound $\gamma + 1 = 10$ is comfortably satisfied. This indicates that the dominating set is robust: removing any single item from D still leaves a coherent and well-connected covered region.

Interpretation. For a recommendation deployment, DI provides an operational robustness metric: if one of the top-9 items becomes unavailable (removed for licensing reasons, flagged for content, or simply out-of-stock), the recommendation system continues to function with graceful degradation. The dual-threshold dominating set therefore inherits the robustness guarantees of the broader fuzzy-graph domination-integrity theory, despite being constructed from a much more computationally tractable algorithm.

7.5 The Asymmetry Between α and β

A surprising empirical finding is that the choice of α has a much larger effect on $|D|$ than the choice of β. As reported in Section 6.5, the dominating set size grows roughly tenfold as α moves from 0.30 to 0.65, while the analogous sweep over β produces almost no change in $|D|$. This asymmetry has both a statistical and a substantive interpretation.

Statistical interpretation. The asymmetry reflects the fact that S^+ and S^- are derived from different mathematical operations. S^+ is a cosine similarity over the binary rating matrix H, whose entries are concentrated near zero because most users rate few items. The cosine of two sparse high-dim vectors typically lies in a moderate range (the empirical mean of 0.42 in our data is consistent with this). S^-, by contrast, is the normalised inner product of H and $L^\top$, where L flips the sign of the rating around the user mean. Two items are strongly opposed only if there exists a substantial set of users who both rated one item highly and one item poorly, which is a much rarer event than mere co-rating. Hence S^- is concentrated near zero with a long thin right tail, and $\beta = 0.5$ is essentially never the binding constraint.

Substantive interpretation. The substantive content of this asymmetry is that *positive preference signal is denser than negative preference signal* in typical recommendation datasets. Users mainly express what they like through ratings, not what they dislike. This is true even for the MovieLens 100K dataset, in which users are explicitly invited to rate on a 1–5 star scale. The negative threshold therefore acts as a *guardrail* rather than a primary selection criterion: it prevents the algorithm from selecting dominators that are "net-positive but locally-conflicting" (cf. Section 7.3), but it rarely intervenes.

This finding generalises to other recommendation datasets we have inspected informally (the Last.fm and Book-Crossing collections show similar asymmetry, though with different absolute magnitudes), and we conjecture that the asymmetry is a structural feature of explicit rating data rather than an artefact of MovieLens.

7.6 Comparison with Existing Approaches

Our method differs from prior recommendation approaches along three dimensions: representation, algorithm, and interpretability. We discuss each in turn.

Representation. Standard collaborative filtering represents users and items either as sparse vectors (memory-based CF), as low-dimensional embeddings (matrix-factorisation methods such as SVD and ALS), or as latent neural representations (autoencoder- and GNN-based methods) (21; 14; 37). None of these representations preserves an explicit notion of *opposition* as a first-class object. Negative preference is either absorbed into the factor matrices (as a low or negative score) or treated as missing data and ignored. Our signed-graph representation, by contrast, makes opposition explicit at the level of edges. The resulting graph is human-inspectable, supports classical graph-theoretic algorithms, and in particular supports the dual-threshold domination framework developed here.

Algorithm. Most CF algorithms produce a per-user ranked recommendation list. Our algorithm produces a *single global set* D that covers the entire catalogue. This is a different operational primitive: D is not a recommendation for any particular user, but rather a *landing-page set* or *onboarding catalogue* that collectively spans the item space. The use cases are therefore complementary rather than competing. A production system might use matrix factorisation for personalised recommendations on the home-page rail and use our dominating set D for cold-start onboarding, editorial curation, or catalogue-trimming experiments.

Interpretability. Each item in D is selected for an articulable graph-theoretic reason: it dominates a particular subset of items in the positive threshold sense, while not strongly opposing them. The "explanation" for including *Memento* in D, for instance, is exactly the list of vertices it dominates plus the verification that it satisfies the dual-threshold constraint to each. This is markedly more transparent than the latent-factor explanation provided by factorisation methods, which can only describe item i via its position in an opaque embedding space.

7.7 Advantages and Limitations

Advantages. The principal advantages of the proposed framework are:

1. *Diversity by construction.* The dominating set spans 12 distinct genres on our experimental data, compared to 10 for the popularity baseline. Diversity is not bolted on as a post-processing step (as in maximal-marginal-relevance re-ranking (4)) but is a direct consequence of the coverage objective.
2. *Compactness.* The full 100-item catalogue is covered by only 9 representatives, a compression ratio of approximately 11:1.
3. *Explicit treatment of opposition.* Unlike standard CF, the framework treats negative preference as a first-class signal and prevents conflict between dominators and the items they cover.
4. *Robustness.* As established in Section 7.4, the dominating set has domination integrity comfortably exceeding the theoretical lower bound; it degrades gracefully under removal of any single item.
5. *Interpretability.* Each dominating selection has a direct graph-theoretic justification (Section 7.6).
6. *Theoretical grounding.* The framework connects to classical signed-graph domination, fuzzy-graph domination, and domination integrity, inheriting their theoretical guarantees.

Limitations. The framework also has important limitations that bound its scope:

1. *Scale.* The greedy algorithm is $O(k^2 \cdot |D|)$ in the number of candidate items k, with constant factors dominated by similarity-matrix lookup. For $k = 10^3$ this is tractable; for $k = 10^6$ (industrial catalogue) it requires approximation or sampling, which we have not investigated in this work.

2. *Static graph.* The signed graph is constructed once from the full rating history. Temporal evolution of preferences (a new genre rising in popularity, a celebrity event changing perception of a film) is not captured. Extending to a dynamic signed graph is a direction we discuss in Chapter 8.
3. *Threshold dependence.* The algorithm depends on α and β, which we have set to 0.5 on principle and on the basis of sensitivity analysis (Section 6.5). A more principled choice via cross-validation against a held-out user-level metric is a direction for future work.
4. *Binary signs.* The dual-threshold construction maps a continuous pair (S^+, S^-) to a discrete sign $\sigma \in \{-1, 0, +1\}$, discarding magnitude information. A magnitude-aware *weighted-signed-graph* domination framework would be a richer formulation but would lose the connection to classical signed-graph domination theory.
5. *No personalisation.* As emphasised in Section 7.6, D is a global object, not a per-user recommendation. Users with extreme tastes are by construction not "represented" in D in the same proportion as users with mainstream tastes. For applications that require strong personalisation, D must be combined with a personalised re-ranking layer.

7.8 Summary

This chapter has interpreted the empirical results of Chapter 6 through the theoretical framework of Chapters 3 and 4. We have characterised the structural properties of the constructed signed graph (sparse positive structure, very sparse negative structure, marked balance), shown that the dual-threshold dominating set is strictly stronger than algebraic signed domination on graphs of this form (Proposition 7.1), verified that the constructed set exceeds the domination-integrity lower bound, and analysed the asymmetry between α and β. The framework offers a representation, an algorithm, and an interpretation distinct from classical collaborative filtering, with a clear set of advantages (diversity, compactness, robustness, interpretability) bounded by an equally clear set of limitations (scale, static graph, threshold dependence, no personalisation). Chapter 8 summarises the contributions of the dissertation and outlines directions for future work.

CHAPTER 8

CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT

8.1 Summary of the Dissertation

This dissertation proposed a signed-graph domination framework for analysing a recommendation catalogue. The central idea was to move beyond an unsigned item–item similarity graph. In a rating dataset, two items may be related in a positive way because users rate them similarly, but they may also be related in a negative way because users who like one of them tend to dislike the other. The method developed here records both relations and then uses a domination rule to select a small representative subset of items.

Chapter 2 reviewed the three strands of literature behind the work: collaborative filtering and graph-based recommendation, signed graph models, and domination in classical, signed and fuzzy graphs. The review showed that these areas contain the necessary ingredients but are rarely combined for catalogue-level representative selection. Recommendation systems often optimise prediction; signed graph learning often optimises embeddings or sign prediction; domination theory studies compact covers. The proposed method joins these perspectives.

Chapter 3 fixed the mathematical notation for signed graphs, balance, domination, signed domination, fuzzy graphs and domination integrity. The dual-threshold domination rule was introduced as the form of domination most suitable for real-valued recommendation data: a selected item must have sufficient positive similarity to an item it covers and sufficiently small negative similarity with it.

Chapter 4 gave the construction in full. The rating matrix was converted into an item–user matrix. Cosine similarity produced the positive matrix S^+. High-rating and low-rating indicator matrices produced the opposite preference matrix S^- through the normalised product $HL^\top$. Together these matrices defined the signed item–item graph. A greedy algorithm then selected a dominating set under thresholds α and β.

Chapter 5 described the MovieLens experiment. The dataset contains 100,836 ratings by 610 users on 9,724 movies (11). The empirical analysis used the top 100 most rated movies so that the similarity estimates were based on enough common rating information. This reduced the extreme sparsity of the full rating matrix and made the structural interpretation more reliable.

Chapter 6 reported the numerical output. The positive similarity matrix had a mean value of about 0.42, while the negative similarity matrix was much sparser and had a mean value of about 0.11. With $\alpha = \beta = 0.5$, the greedy signed domination algorithm selected 9 movies and covered all 100 movies in the selected catalogue. The resulting set had a larger genre spread than the popularity baseline of the same size, and it shared only two items with that baseline.

Chapter 7 interpreted these findings. The signed item graph was shown to be highly consistent with balance-theoretic expectations. The selected set also connected naturally to Acharya's algebraic domination and to domination integrity in signed fuzzy graphs. This supports the view that the algorithm is not merely a heuristic for MovieLens, but an applied version of a recognisable signed-graph domination idea.

8.2 Contributions Revisited

The dissertation makes four main contributions.

C1. A signed catalogue representation. A recommendation catalogue was represented as a signed item–item graph. The positive edges came from cosine similarity, while the negative edges came from opposite preference. This construction preserves a standard collaborative filtering similarity measure while adding a separate channel for disagreement or preference conflict.

C2. The opposite preference matrix. The matrix $HL^\top$ was used to count high–low rating conflict between item pairs. This is simple, interpretable and asymmetric before any symmetrisation. It captures the fact that users who like item i_1 and dislike item i_2 need not be the same users who like i_2 and dislike i_1.

C3. Dual-threshold domination. The proposed domination relation uses two inequalities: $S^+(i_1,i_2) \geq \alpha$ and $S^-(i_1,i_2) \leq \beta$. The first makes the selected item representative; the second prevents strong antagonism. This gives a direct way to transfer the idea of signed domination to a real-valued recommender graph.

C4. Empirical validation. On the MovieLens top-100 subset, the method produced a small dominating set with complete structural coverage. The set also covered more genres than the same-size popularity baseline. These results demonstrate that structural coverage and popularity are different selection principles.

8.3 Limitations

The framework has several limitations.

- The full signed similarity matrix requires $O(k^2)$ storage. This is easy for $k = 100$ but becomes expensive for industrial catalogues.
- The graph is static. It does not yet track how ratings and user tastes change over time.
- The thresholds α and β were chosen by interpretability and sensitivity analysis. They were not tuned by held-out recommendation metrics.

- The final output is a global representative set. It is not yet a personalised recommendation list for each user.
- The negative matrix depends on chosen high-rating and low-rating cutoffs. Other datasets may require different cutoffs or a continuous version of the same idea.

These limitations do not invalidate the method; they mark the boundary between the present dissertation and future extensions.

8.4 Future Research Directions

8.4.1 Scaling to Large Catalogues

For large catalogues, computing all pairwise similarities is the main bottleneck. Approximate nearest-neighbour methods, locality-sensitive hashing and sparse candidate generation could be used to avoid constructing the full $k \times k$ matrices. Another possibility is to treat signed domination as a submodular coverage problem with additional negative-edge constraints. This would make it possible to derive formal approximation guarantees for large-scale versions of the algorithm (26; 18).

8.4.2 Temporal Dynamics via Signed Temporal Graphs

User preferences are not fixed. New movies arrive, old movies become popular again, and the meaning of a genre can shift. A temporal version of the method would build matrices H_t, L_t, S_t^+ and S_t^- at different time slices and study the sequence of dominating sets D_t. This would allow one to ask when a representative set is stable and when a catalogue has undergone a real structural change (2; 38).

8.4.3 Domination Integrity as an Optimisation Objective

In this dissertation, domination integrity was used mainly as an interpretive quantity. A stronger version would optimise it directly. Among all dominating sets that satisfy the dual-threshold condition, one could search for a set that also maximises robustness after removal. This would connect the algorithm more closely to the domination-integrity theory of signed fuzzy graphs (29).

8.4.4 Integration with Signed Graph Neural Networks

The present method uses hand-computed similarities. A natural extension is to learn signed item embeddings with a signed graph neural network such as SGCN, SiGAT or related models (9; 17; 16). The dominating set could then be computed in the learned embedding space rather than from raw cosine similarity. This would combine the interpretability of a small dominating set with the representational strength of signed graph learning.

8.4.5 FThreshold Selection via Held-Out Validation

The choice $\alpha = \beta = 0.5$ is simple and interpretable, but a deployed system would benefit from data-driven threshold selection. Future work can choose α and β by validation on ranking metrics such as recall, NDCG or coverage-diversity trade-offs. The threshold grid in Chapter 6 already shows that α has a stronger effect on set size than β; validation would turn this observation into an operational tuning rule.

8.5 Closing Remarks

The main message of the dissertation is that signed graph domination offers a useful structural viewpoint on recommendation systems. Positive similarity alone is not enough to describe a catalogue when low ratings are meaningful. By adding an opposite preference matrix and requiring representatives to be both similar and non-antagonistic, the proposed method produces compact item sets that are interpretable, cover the catalogue and differ substantially from popularity rankings. This opens a path for further work at the boundary of recommender systems, signed graph theory and domination-based optimisation.

BIBLIOGRAPHY

[1] B. D. Acharya. Domination and absorbance in signed graphs and digraphs: I. foundations. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 84:5–20, 2013.

[2] S. Ahmadi, F. Bonchi, S. Galimberti, F. Gullo, and C. Schweikert. Temporal signed networks: Models and algorithms. *ACM Computing Surveys*, 55(3):1–36, 2022.

[3] K. R. Bhutani and A. Rosenfeld. Strong arcs in fuzzy graphs. *Information Sciences*, 152:319–322, 2003.

[4] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference*, pages 335–336, 1998.

[5] D. Cartwright and F. Harary. Structural balance: a generalization of Heider's theory. *Psychological Review*, 63(5):277–293, 1956.

[6] S. Chakaravarthy, K. Chandran, S. Mariappan, and S. Ramalingam. Edge integrity for signed fuzzy graphs. *Journal of Intelligent & Fuzzy Systems*, 43:4681–4690, 2022.

[7] E. J. Cockayne, P. A. Dreyer Jr., S. M. Hedetniemi, and S. T. Hedetniemi. Roman domination in graphs. *Discrete Mathematics*, 278:11–22, 2004.

[8] J. A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20:181–187, 1967.

[9] T. Derr, Y. Ma, and J. Tang. Signed graph convolutional networks. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pages 929–934, 2018.

[10] F. Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146, 1953.

[11] F. M. Harper and J. A. Konstan. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):19:1–19:19, 2015.

[12] T. W. Haynes, S. T. Hedetniemi, and P. J. Slater. *Fundamentals of Domination in Graphs*. Marcel Dekker, New York, 1998.

[13] X. He, K. Deng, X. Wang, Y. Li, Y.-D. Zhang, and M. Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference*, pages 639–648, 2020.

[14] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web (WWW)*, pages 173–182, 2017.

[15] F. Heider. Attitudes and cognitive organization. *Journal of Psychology*, 21(1):107–112, 1946.

[16] J. Huang, H. Shen, Q. Cao, S. Tao, and X. Cheng. SDGNN: Learning node representation for signed directed networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):196–203, 2021.

[17] J. Huang, H. Shen, L. Hou, and X. Cheng. Signed graph attention networks. In *International Conference on Artificial Neural Networks (ICANN)*, 2019.

[18] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC)*, pages 604–613, 1998.

[19] P. Jeyalakshmi. Domination in signed graphs. *Discrete Mathematics, Algorithms and Applications*, 13(1):2050094, 2020.

[20] J. Joseph and M. Joseph. Roman domination in signed graphs. *Communications in Combinatorics and Optimization*, 8(2):349–358, 2023.

[21] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[22] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 559–570, 2010.

[23] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1361–1370, 2010.

[24] Y. Li, Y. Tian, J. Zhang, and Y. Chang. Learning signed network embedding via graph attention. In *AAAI Conference on Artificial Intelligence*, pages 4772–4779, 2020.

[25] S. Mariappan, S. Ramalingam, S. Raman, and G. Bacak-Turan. Domination integrity and efficient fuzzy graphs. *Neural Computing and Applications*, 32:10263–10273, 2020.

[26] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978.

[27] O. Ore. Theory of graphs. *American Mathematical Society*

Colloquium Publications, 38, 1962.

[28] A. Rosenfeld. Fuzzy graphs. In *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, pages 77–95. Elsevier, Amsterdam, 1975.

[29] C. Sankar, C. Kalaivani, P. Chellamani, and G. Venkat Narayanan. Analyzing network stability via symmetric structures and domination integrity in signed fuzzy graphs. *Symmetry*, 17(5):766, 2025.

[30] M. Saravanan, R. Sujatha, R. Sundareswaran, and M. S. Balasubramanian. Application of domination integrity of graphs in pmu placement in electric power networks. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26:2066–2076, 2018.

[31] L. Shu, E. Du, Y. Chang, C. Chen, Z. Zheng, X. Xing, and S. Shen. Sgcl: Contrastive representation learning for signed graphs. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1671–1680, 2021.

[32] A. Somasundaram. Domination in fuzzy graphs–ii. *Journal of Fuzzy Mathematics*, 13:281, 2005.

[33] A. Somasundaram and S. Somasundaram. Domination in fuzzy graphs–i. *Pattern Recognition Letters*, 19(9):787–791, 1998.

[34] R. Sundareswaran, R. Sujatha, and G. Bacak-Turan. A study on vulnerability parameters of signed fuzzy graphs. In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making, INFUS 2019*, pages 24–32. Springer, 2020.

[35] J. Tang, Y. Chang, C. Aggarwal, and H. Liu. A survey of signed network mining in social media. *ACM Computing Surveys*, 49(3):1–37, 2016.

[36] H. B. Walikar, S. V. Motammanavar, and B. D. Acharya. Signed domination in signed graphs. *Journal of Combinatorial Information & System Sciences*, 40(1–4):107–128, 2015.

[37] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd International ACM SIGIR Conference*, pages 165–174, 2019.

[38] Y. Wang, Y. Zhang, Q. Liu, and X. Zhou. Dynamic signed graph learning for temporal link prediction. *Knowledge-Based Systems*, 270:110528, 2023.

[39] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD*, 2018.

[40] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[41] T. Zaslavsky. Signed graphs. *Discrete Applied Mathematics*, 4(1):47–74, 1982.

[42] Z. Zhang, P. Zhao, X. Li, J. Liu, X. Zhang, J. Huang, and X. Zhu. Signed graph representation learning: A survey. *arXiv preprint arXiv:2402.15980*, 2024.

MSc thesis

ORIGINALITY REPORT

13%	11%	11%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	1%
2	export.arxiv.org Internet Source	1%
3	Submitted to CSU, San Jose State University Student Paper	<1%
4	real-j.mtak.hu Internet Source	<1%
5	Chakaravarthy Sankar, Chandran Kalaivani, Mariappan Saravanan, Ramalingam Sujatha. "An algorithmic approach to signed fuzzy graph integrity: Complexity, graph operations, and metro rail network applications", Ain Shams Engineering Journal, 2025 Publication	<1%
6	www.ejpam.com Internet Source	<1%
7	www.researchgate.net Internet Source	<1%
8	www.mdpi.com Internet Source	<1%
9	archives.christuniversity.in Internet Source	<1%
10	Submitted to South Asian University Student Paper	<1%

11	Yulin Zhu, Tomasz Michalak, Xiapu Luo, Xiaoge Zhang, Kai Zhou. "Towards Secrecy-Aware Attacks Against Trust Prediction in Signed Social Networks", IEEE Transactions on Information Forensics and Security, 2024 Publication	<1%
12	papers.nips.cc Internet Source	<1%
13	www-db.deis.unibo.it Internet Source	<1%
14	combinatorialpress.com Internet Source	<1%
15	s2.smu.edu Internet Source	<1%
16	sciencedocbox.com Internet Source	<1%
17	docswell.com Internet Source	<1%
18	Xing, H.-M.. "A note on Roman domination in graphs", Discrete Mathematics, 20061228 Publication	<1%
19	core.ac.uk Internet Source	<1%
20	dspace.mit.edu Internet Source	<1%
21	www.diva-portal.org Internet Source	<1%
22	downloads.hindawi.com Internet Source	<1%
23	Jih-Jeng Huang, Chin-Yi Chen. "Certified Graph-Cut Separation for 3-Additive	<1%

	Capacities via Signed Switching", Fuzzy Sets and Systems, 2026 Publication	
24	Submitted to University of Denver Student Paper	<1 %
25	arpi.unipi.it Internet Source	<1 %
26	Liang Du, Hao Jiang, Hui Lin, Hao Li. "Beyond Structural Balance: Imbalance-Aware Embedding for Signed Networks", Expert Systems with Applications, 2025 Publication	<1 %
27	www.conf-icnc.org Internet Source	<1 %
28	people.computing.clemson.edu Internet Source	<1 %
29	datasets-benchmarks-proceedings.neurips.cc Internet Source	<1 %
30	gcris.iyte.edu.tr Internet Source	<1 %
31	scholar.uoc.ac.in Internet Source	<1 %
32	tnsroindia.org.in Internet Source	<1 %
33	P. Jeyalakshmi, K. Karuppasamy. "Connected domination in a signed graph and its complement", Journal of Intelligent & Fuzzy Systems, 2023 Publication	<1 %
34	itueconomics.github.io Internet Source	<1 %
35	rvce.edu.in Internet Source	

		<1 %
36	www.malayajournal.org Internet Source	<1 %
37	Submitted to Coventry University Student Paper	<1 %
38	P. K. Ashraf, K. A. Germina. "ON MINIMAL DOMINATING SETS FOR SIGNED GRAPHS", Advances and Applications in Discrete Mathematics, 2015 Publication	<1 %
39	Sami Ullah Khan, Naeem Jan, Kifayat Ullah, Lazim Abdullah. "Graphical Structures of Cubic Intuitionistic Fuzzy Information", Journal of Mathematics, 2021 Publication	<1 %
40	telemedicina.unifesp.br Internet Source	<1 %
41	Sankar Chakaravarthy, Kalaivani Chandran, Saravanan Mariappan, Sujatha Ramalingam. "Edge integrity for signed fuzzy graphs", Journal of Intelligent & Fuzzy Systems, 2022 Publication	<1 %
42	sina.sharif.edu Internet Source	<1 %
43	www.ijraset.com Internet Source	<1 %
44	Valeroso, E.S.. "Multiresponse surface models with random block effects", Journal of Statistical Planning and Inference, 199906 Publication	<1 %
45	hal.inria.fr Internet Source	<1 %

46	fs.unm.edu Internet Source	<1 %
47	ses.library.usyd.edu.au Internet Source	<1 %
48	Linrui Shen, Anchen Li, Xueyan Liu, Riting Xia, Bo Yang. "Grouping-enhanced personalization for federated recommendation", Information Processing & Management, 2026 Publication	<1 %
49	download.bibis.ir Internet Source	<1 %
50	Yichao Chen. "Pre-signed graphs: A reformulation of signed graphs and their embeddings", Advances in Applied Mathematics, 2026 Publication	<1 %
51	www.mecs-press.org Internet Source	<1 %
52	Henning, M.A.. "Defending the Roman Empire from multiple attacks", Discrete Mathematics, 20030928 Publication	<1 %
53	www.cse.msu.edu Internet Source	<1 %
54	Submitted to Al-Qasim Green University Student Paper	<1 %
55	Submitted to Associatie K.U.Leuven Student Paper	<1 %
56	Submitted to Indian Institute of Technology (ISM), Dhanbad Student Paper	<1 %

57	Submitted to Liverpool John Moores University Student Paper	<1%
58	Lu Zhang, Yongkai Wu, Xintao Wu. "Achieving Non-Discrimination in Data Release", Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17, 2017 Publication	<1%
59	P. Roushini Leely Pushpam, M. Kamalam, B. Mahavir. "Stability of Certified Domination upon Edge Addition", Discrete Mathematics, Algorithms and Applications, 2023 Publication	<1%
60	Siran, J.. "Characterization of the maximum genus of a signed graph", Journal of Combinatorial Theory, Series B, 199105 Publication	<1%
61	escholarship.org Internet Source	<1%
62	Submitted to Buckinghamshire Chilterns University College Student Paper	<1%
63	Submitted to Imperial College of Science, Technology and Medicine Student Paper	<1%
64	Submitted to Southeast Community College Student Paper	<1%
65	tel.archives-ouvertes.fr Internet Source	<1%
66	www.ijmdrr.com Internet Source	<1%

67	Anisha Jean Mathias, V. Sangeetha, Mukti Acharya. "Critical concepts of restrained domination in signed graphs", Discrete Mathematics, Algorithms and Applications, 2021 Publication	<1 %
68	Bogdan Walek, Vladimir Fojtik. "A hybrid recommender system for recommending relevant movies using an expert system", Expert Systems with Applications, 2020 Publication	<1 %
69	inria.hal.science Internet Source	<1 %
70	pphmjopenaccess.com Internet Source	<1 %
71	www.castellocheese.com Internet Source	<1 %
72	Türk, Melihcan. "Software Design, Implementation, Application, and Refinement of a Bayesian Approach for the Assessment of Content and User Qualities", Bilkent Universitesi (Turkey), 2022 Publication	<1 %
73	www.biorxiv.org Internet Source	<1 %
74	www.meiermovies.com Internet Source	<1 %
75	Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, Maosong Sun. "Graph neural networks: A review of methods and applications", AI Open, 2020 Publication	<1 %

76 Jun-Ming Xu. "On Bondage Numbers of Graphs: A Survey with Some Comments", International Journal of Combinatorics, 2013
Publication
<1 %

77 Saravanan Mariappan, Sujatha Ramalingam, Sundareswaran Raman, Goksen Bacak-Turan. "Domination integrity and efficient fuzzy graphs", Neural Computing and Applications, 2019
Publication
<1 %

78 Williams, Linda, Hammond, Michael. "Contemporary American Cinema", Contemporary American Cinema, 2006
Publication
<1 %

79 noexperiencenecessarybook.com
Internet Source
<1 %

80 pdfs.semanticscholar.org
Internet Source
<1 %

81 savingcalfstretchwedge.blogspot.com
Internet Source
<1 %

82 tnhh.org
Internet Source
<1 %

83 www.ijsr.net
Internet Source
<1 %

84 Chaobo He, Hao Cheng, Jiaqi Yang, Yong Tang, Quanlong Guan. "Signed graph embedding via multi-order neighborhood feature fusion and contrastive learning", Neural Networks, 2024
Publication
<1 %

85 Ibtissem Gasmi, Hassina Seridi-Bouchelaghem, Labar Hocine, Baareh Abdelkarim. "Collaborative filtering
<1 %

recommendation based on dynamic changes of user interest", Intelligent Decision Technologies, 2015
Publication

86	Xiaolan Liu, Shaohua Pan, Zhifeng Hao, Zhiyong Lin. "Graph-based semi-supervised learning by mixed label propagation with a soft constraint", Information Sciences, 2014 Publication	<1%
87	calhoun.nps.edu Internet Source	<1%
88	deepai.org Internet Source	<1%
89	dergipark.org.tr Internet Source	<1%
90	evazangerle.at Internet Source	<1%
91	iris.imtlucca.it Internet Source	<1%
92	jmlr.org Internet Source	<1%
93	m4d.iti.gr Internet Source	<1%
94	mro.massey.ac.nz Internet Source	<1%
95	real.mtak.hu Internet Source	<1%
96	repositorio.unal.edu.co Internet Source	<1%
97	theses.hal.science Internet Source	<1%

98	web.stanford.edu Internet Source	<1%
99	Madhumangal Pal, Sovan Samanta, Ganesh Ghorai. "Modern Trends in Fuzzy Graph Theory", Springer Science and Business Media LLC, 2020 Publication	<1%
100	Sankar Chakaravarthy, Mariappan Saravanan, Ramalingam Sujatha, Gangatharan Venkat Narayanan. "Signed Fuzzy Graphs Domination in Environmental Monitoring Wireless Sensor Networks", European Journal of Pure and Applied Mathematics, 2025 Publication	<1%
101	Soumyapriya Goswami, Partha Sarathi Banerjee, Debashis De. "RELEFA: A rule-based entropy driven expert leader-follower algorithm for fault-tolerant dynamic minimum connected dominating set formation in industrial internet of things networks", Expert Systems with Applications, 2026 Publication	<1%
102	"Medical Image Computing and Computer Assisted Intervention – MICCAI 2024", Springer Science and Business Media LLC, 2024 Publication	<1%
103	Anisha Jean Mathias, V. Sangeetha, Mukti Acharya. "Restrained domination in signed graphs", Acta Universitatis Sapientiae, Mathematica, 2020 Publication	<1%

104 Chakaravarthy Sankar, Chandran Kalaivani, Perumal Chellamani, Gangatharan Venkat Narayanan. "Analyzing Network Stability via Symmetric Structures and Domination Integrity in Signed Fuzzy Graphs", Symmetry, 2025 <1%

Publication

105 Takaaki Fujita. "Molecular Fuzzy graphs, hypergraphs, and superhypergraphs", American Chemical Society (ACS), 2025 <1%

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		