

Buffered Spectral Laplacian Sparsification for Fixed-Charge Transportation Problems

Deepak Kumar Mishra¹, Vishal Vincent Henry²

¹Department of Mathematics and Statistics, Sam Higginbottom University of Agriculture, Technology and Sciences
Email: [deepakkm.018\[at\]gmail.com](mailto:deepakkm.018[at]gmail.com)

²Department of Mathematics and Statistics, Sam Higginbottom University of Agriculture, Technology and Sciences
Email: [vishal.henry\[at\]shiats.edu.in](mailto:vishal.henry[at]shiats.edu.in)

Abstract: *The Fixed-Charge Transportation Problem (FCTP) is an NP-hard problem whose objective is non-convex because every route carries a fixed activation fee. We describe a way to shrink the single global MILP by removing binary variables chosen with spectral graph theory. The fixed and variable costs are merged into edge weights of a bipartite graph, and sources and demand nodes are ordered along the Fiedler vector of its Laplacian. Routes that join nodes on opposite sides of the spectral boundary are fixed to zero, except inside a quantile-controlled overlap region, the δ -buffer, which keeps capacity near the boundary. We prove an exact closed form for the fraction of removed variables, $\rho(q) = (|A_-||B_+| + |A_+||B_-|)/(mn)$, together with monotonicity in q , full recovery at $q = 1$, and an identity for $\rho(0)$ in terms of the sign fractions of the Fiedler vector. Feasibility of the masked problem is characterized through the Gale–Hoffman condition, which gives a critical quantile q^* that binary search finds in $O(T_{MF}(m, n)\log(m + n))$ time, where T_{MF} is the cost of one max-flow check. We also give an a posteriori certificate that bounds the optimality gap from any lower bound, such as the LP relaxation, without solving the full MILP. We test the method on three synthetic scales (20×20 , 50×50 , 100×100 ; 20 seeds each; HiGHS). At $q = 0.10$ the mean gap stays below 0.35% and the certified gap below 2.5%. End-to-end speedups over the unpartitioned MILP are $0.38 \times$, $1.27 \times$ and $1.72 \times$, so the crossover lies between 20×20 and 50×50 . Seeding a full MILP with the sparsified solution recovers exactness but is slower than the sparsified pipeline (0.745 s against 0.615 s at 50×50 ; 6.03 s against 5.49 s at 100×100). Random and METIS masks run faster but leave larger gaps. The speedups on these instances are modest, and the main contributions are the exact reduction-ratio results, the critical-quantile search, and the optimality certificate.*

Keywords: fixed-charge transportation problem, spectral graph theory, Fiedler vector, overlap buffering, MILP sparsification, Gale–Hoffman feasibility, optimality certificate

1. Introduction

The Fixed-Charge Transportation Problem (FCTP) models distribution networks in which shipping goods costs a variable amount per unit and, in addition, a fixed fee for using a route [1]. Because the fee is paid whenever a route is used at all, the objective is non-convex, and the problem is NP-hard [2].

Commercial branch-and-bound solvers handle moderate FCTP instances well, but their running time can grow quickly once a network has hundreds of warehouses and distribution points. Spectral graph partitioning is a natural tool for splitting such networks: it uses the second-smallest eigenvalue λ_2 of the graph Laplacian and its eigenvector, the Fiedler vector [3]. Applied directly, however, a spectral bisection is a poor fit for the FCTP. A standard graph cut discards every edge between the two parts, which can remove high-capacity corridors and lead to suboptimal or even infeasible subproblems.

This paper uses the spectral information to sparsify the FCTP instead of bisecting it. We keep one global MILP in which only a subset of the binary variables is active. The subset is selected by a quantile-controlled overlap buffer around the boundary of the Fiedler partition, so that routes near the boundary survive. This construction has an exact closed-form reduction ratio, a Gale–Hoffman feasibility characterization, a critical quantile that can be found by binary search, an a posteriori optimality certificate, and a conditional bound that relates the loss in optimality to the flow carried on the eliminated routes.

1.1 Contributions

- 1) **Effective bipartite Laplacian.** Fixed and variable costs are combined into one affinity matrix W , and the Fiedler vector of its Laplacian orders sources and demands.
- 2) **Quantile-controlled overlap buffer.** A single parameter q sets how many routes across the boundary are kept, with $\delta(q)$ defined through order statistics so that results are reproducible.
- 3) **Exact reduction ratio (Theorem 4, Propositions 5 and 7).** We give a closed form for $\rho(q)$, the identity $\rho(0) = \alpha + \beta - 2\alpha\beta$ in terms of the sign fractions of f , and monotonicity with full recovery at $q = 1$.
- 4) **Feasibility and critical quantile (Theorem 8, Corollary 9).** The masked problem is feasible exactly when a Gale–Hoffman condition holds, and the smallest feasible q is found with $O(\log(m + n))$ max-flow checks.
- 5) **A posteriori certificate (Theorem 10).** Any lower bound on the global optimum yields a computable bound on the gap.
- 6) **Eliminated routes (Observations 14 and 15).** We note that eliminated routes are costlier on average than retained ones and state a conditional rerouting bound.

Volume 15 Issue 10, October 2026

Fully Refereed | Open Access | Double Blind Peer Reviewed Journal

www.ijsr.net

- 7) **Experiments (Section 8).** Multi-scale timings, random-mask, METIS and warm-start comparisons, structured instances, certified gaps, and mask stability under cost noise.

1.2 Organization

Section 2 reviews related work. Section 3 states the FCTP and the effective Laplacian. Section 4 defines the buffered mask. Section 5 gives the theory. Section 6 reports the empirical spectrum and the stability of the mask. Section 7 presents the algorithm, Section 8 the experiments, and Section 9 concludes.

2. Related Work

2.1 FCTP: complexity and exact methods

The FCTP goes back to Hirsch and Dantzig [4] and was studied early by Balinski [1]. Its NP-hardness is classical; Garey and Johnson [2] treat the closely related minimum edge-cost flow problem. Classical exact methods use Benders decomposition [5] or Lagrangian relaxation [6], both of which separate the binary route decisions from the continuous flows. More recently, Chen, Li and Liang [7] studied the approximability and complexity of FCTP variants with general, sink-independent and uniform fixed costs. Their results describe theoretical limits, whereas our aim is practical tractability.

2.2 Decomposition and metaheuristics

Lagrangian decomposition and variable fixing have been applied to large scheduling MILPs [8], which motivates using problem structure to shrink the search space. Metaheuristics such as genetic algorithms, tabu search and variable neighbourhood search give fast solutions but no optimality guarantee, and they usually need tuning for each network topology. We differ from these approaches by keeping one global MILP and reducing its variable space with spectral information.

2.3 Spectral partitioning in constrained optimization

The Fiedler vector is a standard device for partitioning graphs [3, 9], and bipartite spectral embeddings are well studied for data clustering and VLSI design [10]. METIS [11] is the usual multilevel baseline. In combinatorial optimization, Cordero et al. [12] use spectral clustering to support column generation for a connected-components partitioning problem. Herrán, Colmenar and Resende [13] treat multi-constraint graph partitioning with a two-phase GRASP and observe that standard tools such as METIS cannot handle their two-sided weight constraints, which shows that off-the-shelf partitioners can fit constrained problems poorly. Other recent work includes reweighted spectral partitioning for vertex separators [14] and spectral graph neural network filters for community detection [15].

None of these works partitions an FCTP, and none uses an overlap region to preserve capacity at the boundary. The δ -buffer is therefore motivated by our own hypothesis and not by earlier findings: a rigid Fiedler cut ($q = 0$) removes cross-partition corridors. Table 3 supports this, since the masked FCTP at $q = 0$ is infeasible on all 20 seeds at 50×50 , which in turn motivates the critical-quantile search of Corollary 9.

2.4 Gap and contribution

To our knowledge, no existing work combines spectral bipartite partitioning with a quantile-controlled overlap buffer for the FCTP. Cordero et al. [12] do not offer a tunable buffer that trades variable reduction against solution quality, an exact reduction-ratio formula, a feasibility characterization, or an a posteriori gap certificate. Our contribution is the δ -buffer together with the theory of Section 5, which links spectral sparsification to MILP exactness while keeping a single global model.

3. Mathematical Formulation

3.1 The standard FCTP

Let $S = \{1, \dots, m\}$ be the sources and $D = \{1, \dots, n\}$ the demand nodes, so the vertex set is $V = S \cup D$ with $|V| = m + n$. Let $x_{ij} \geq 0$ be the flow on route (i, j) and $y_{ij} \in \{0, 1\}$ indicate whether the route is used. Supplies are s_i and demands d_j :

$$\begin{aligned}
\min_{x,y} \quad & \sum_{i=1}^m \sum_{j=1}^n (c_{ij}x_{ij} + F_{ij}y_{ij}) \\
\text{s.t.} \quad & \sum_{j=1}^n x_{ij} \leq s_i, \quad \forall i \in S, \\
& \sum_{i=1}^m x_{ij} \geq d_j, \quad \forall j \in D, \\
& x_{ij} \leq M_{ij} y_{ij}, \quad \forall (i,j) \in S \times D, \\
& x_{ij} \geq 0, \quad y_{ij} \in \{0,1\},
\end{aligned} \tag{1}$$

where $M_{ij} = \min(s_i, d_j)$. Throughout we assume that the full problem is feasible.

Assumption 1. $\sum_{i \in S} s_i \geq \sum_{j \in D} d_j$.

The unmasked network is a complete bipartite graph and all supplies and demands are nonnegative, so Assumption 1 is sufficient for feasibility of (1). For the masked network, feasibility is characterized by Theorem 8.

3.2 Effective weights and the Laplacian

To bring fixed costs into the spectral embedding, we define an effective cost with a structural weight α :

$$c_{ij}^{\text{eff}} = c_{ij} + \alpha \frac{F_{ij}}{d}, \quad \bar{d} = \frac{1}{n} \sum_{j=1}^n d_j. \tag{2}$$

The affinity weights are

$$W_{ij} = \frac{1}{\max(c_{ij}^{\text{eff}}, \epsilon)}, \tag{3}$$

where $\epsilon > 0$ guards against division by zero. The bipartite adjacency matrix has the usual block form

$$A = \begin{bmatrix} 0 & W \\ W^\top & 0 \end{bmatrix}, \tag{4}$$

and the unnormalized Laplacian is

$$L = D_{\text{deg}} - A, \quad D_{\text{deg}} = \text{diag}(A\mathbf{1}). \tag{5}$$

We solve $Lv = \lambda v$. Since L is symmetric positive semidefinite, its eigenvalues satisfy $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_{m+n}$. The eigenvector f belonging to λ_2 is the Fiedler vector, normalized so that $\|f\|_2 = 1$ and $\mathbf{1}^\top f = 0$. We write f_i for $i \in S$ and f_{m+j} for $j \in D$.

Remark 2 (Density of the effective graph). Since $W_{ij} > 0$ for every pair (i,j) , the effective graph is the complete bipartite graph $K_{m,n}$ with weights W_{ij} . Its Laplacian is dense, with mn nonzero off-diagonal entries. The Lanczos iteration of Section 4.5 works on this dense weight matrix, and Section 6 discusses what the density means for the spectrum.

4. Buffered Spectral Sparsification

4.1 Scale-invariant buffer

We parameterize the buffer by a quantile $q \in [0,1]$ of the absolute Fiedler values. To make the construction reproducible, we use order statistics. Let $N = m + n$ and let $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(N)}$ be the sorted values of $|f|$, with $u_{(0)} = 0$. For $r = 0, \dots, N$ define the breakpoints

$$\delta_r = u_{(r)}, \quad q_r = \frac{r}{N}.$$

For $q \in [0,1]$ let $r = \lceil qN \rceil$ and set

$$\delta(q) = Q_q(|f|) = \delta_r, \quad \delta(0) = 0. \tag{6}$$

The mask is then piecewise constant in q with at most $m + n$ breakpoints.

4.2 Active route mask

Define the subdomains

$$\begin{aligned}
S_1(q) &= \{i \in S: f_i \geq -\delta(q)\}, & D_1(q) &= \{j \in D: f_{m+j} \geq -\delta(q)\}, \\
S_2(q) &= \{i \in S: f_i \leq +\delta(q)\}, & D_2(q) &= \{j \in D: f_{m+j} \leq +\delta(q)\}.
\end{aligned} \tag{7}$$

The active route mask $M_{\text{active}}(q) \in \{0,1\}^{m \times n}$ is

Volume 15 Issue 10, October 2026

Fully Refereed | Open Access | Double Blind Peer Reviewed Journal

www.ijsr.net

$$M_{ij}(q) = \begin{cases} 1, & \text{if } (i \in S_1 \wedge j \in D_1) \vee (i \in S_2 \wedge j \in D_2), \\ 0, & \text{otherwise.} \end{cases}$$

(8)

4.3 Variable reduction ratio

$$\rho(q) = 1 - \frac{\sum_{i=1}^m \sum_{j=1}^n M_{ij}(q)}{mn}. \quad (9)$$

4.4 Boundary reconciliation

We do not solve two disconnected subproblems. We solve one sparsified global MILP restricted to the active mask: variables with $M_{ij}(q) = 0$ are fixed to zero before branch-and-bound starts, and the edges in the overlap buffer stay fully active in the global model.

4.5 Computational complexity

The pipeline has four stages.

- 1) Laplacian construction takes $O(mn)$ time and memory.
- 2) Fiedler vector by Lanczos iteration. By the Kaniel–Paige–Saad bound [16], the number of iterations needed to reach tolerance ϵ_{tol} on λ_2 satisfies

$$k = O\left(\sqrt{\frac{\lambda_{\max}}{\lambda_3 - \lambda_2}} \log \frac{1}{\epsilon_{\text{tol}}}\right). \quad (10)$$

The rate is governed by the relative gap $(\lambda_3 - \lambda_2)/\lambda_{\max}$. As Section 6 shows, this gap is small for the effective graphs that arise here, so (10) should be read as a pessimistic bound and not as a tight predictor of the runtimes in Table 4.

- 3) Critical-quantile search needs $O(\log(m+n))$ feasibility checks. Each check is a max-flow problem on a graph with $m+n+2$ vertices and $mn+m+n$ arcs. Writing $T_{\text{MF}}(m,n)$ for the time of one max-flow computation, a relabel-to-front push-relabel implementation [17] gives the safe generic bound $T_{\text{MF}}(m,n) = O((m+n)^3)$, so the total cost of the search is $O(T_{\text{MF}}(m,n)\log(m+n))$. We expect this to be a small term compared with the dense eigensolve at the sizes studied here. The sub-unity speedup at 20×20 (Table 4) is most likely caused by the fixed cost of the eigensolve and of setting up the pipeline. The LP lower bound used for the certificate is computed separately and is not part of T_{spectral} or of the end-to-end totals reported below.
- 4) Sparsified MILP. The sparsified model keeps only $(1 - \rho(q))mn$ binary variables (and the same number of flow variables). This count only describes how much the enumeration space shrinks. It says nothing about the LP-based search a solver actually performs, so we treat it as descriptive and rely on measured times (Section 8).

5. Theoretical Properties

Throughout, f is the Fiedler vector with $\|f\|_2 = 1$ and $\mathbf{1}^\top f = 0$. For $q \in [0,1]$ we set $\delta(q) = Q_q(|f|)$ with $\delta(0) := 0$, and we use the convention

$$Z^*(q) = +\infty \quad \text{if the masked FCTP is infeasible.} \quad (11)$$

5.1 Edge elimination and the exact reduction ratio

Lemma 3 (Edge elimination criterion). *For every $i \in S, j \in D$ and $q \in [0,1]$,*

$$M_{ij}(q) = 0 \Leftrightarrow (f_i < -\delta(q) \wedge f_{m+j} > \delta(q)) \vee (f_i > \delta(q) \wedge f_{m+j} < -\delta(q)).$$

Proof. By (8), $M_{ij}(q) = 1$ iff $(i \in S_1 \wedge j \in D_1) \vee (i \in S_2 \wedge j \in D_2)$. Taking the complement and expanding the two negations gives four cases. Two of them are impossible because $\delta(q) \geq 0$, and the remaining two are exactly the disjoint cases in the statement.

Define the four boundary sets

$$\begin{aligned} A_- &= \{i \in S: f_i < -\delta\}, & A_+ &= \{i \in S: f_i > +\delta\}, \\ B_- &= \{j \in D: f_{m+j} < -\delta\}, & B_+ &= \{j \in D: f_{m+j} > +\delta\}. \end{aligned}$$

(12)

Theorem 4 (Exact reduction ratio). *For every $q \in [0,1]$,*

Proof. By Lemma 3, the set of eliminated edges is $E_q = (A_- \times B_+) \cup (A_+ \times B_-)$. The union is disjoint because $A_- \cap A_+ = \emptyset$ and $B_- \cap B_+ = \emptyset$, so $|E_q| = |A_-||B_+| + |A_+||B_-|$. Dividing by mn and using (9) gives (13).

$$\rho(q) = \frac{|A_-||B_+| + |A_+||B_-|}{mn}. \quad (13)$$

Proposition 5 (Reduction ratio at $q = 0$). Let $\alpha = |A_-|/m$ and $\beta = |B_-|/n$ be the fractions of sources and demands with negative Fiedler value, and assume that no node has $f_v = 0$ (true almost surely under continuous costs). Then

$$\rho(0) = \alpha + \beta - 2\alpha\beta = \frac{1}{2} - 2\left(\alpha - \frac{1}{2}\right)\left(\beta - \frac{1}{2}\right). \quad (14)$$

Consequently:

- 1) $\rho(0) = \frac{1}{2}$ if and only if at least one side is sign-balanced ($\alpha = \frac{1}{2}$ or $\beta = \frac{1}{2}$).
- 2) $\rho(0) \leq \frac{1}{2}$ if and only if sources and demands lean the same way, that is, $\left(\alpha - \frac{1}{2}\right)\left(\beta - \frac{1}{2}\right) \geq 0$.
- 3) In general $\rho(0) \in [0, 1]$, and $\rho(0) = 1$ occurs when all sources are negative and all demands positive.

By Proposition 7, $\rho(q) \leq \rho(0)$ for all $q \geq 0$, so $\rho(q) \leq \frac{1}{2}$ for all q whenever (ii) holds.

Proof. By Theorem 4 at $\delta = 0$,

$$\rho(0) = \frac{|A_-||B_+| + |A_+||B_-|}{mn} = \alpha(1 - \beta) + (1 - \alpha)\beta,$$

and expanding gives the identity. The three consequences follow by inspection of the second form in (14).

Remark 6 (A benchmark, not a theorem). Suppose both sides are sign-balanced and the buffer takes a fraction q of the nodes, proportionally from each sign class. Then $\rho(q) = (1 - q)^2/2$. The values observed in Table 3 (28.5% at $q = 0.10$ and 8.4% at $q = 0.20$, against 40.5% and 32%) lie below this curve, which points to a lopsided Fiedler vector. We make no claim of a linear rate $\rho(q) \leq \frac{1}{2} - cq$. Section 6 reports the empirical α , β and $\rho(0)$.

Proposition 7 (Monotonicity and full recovery). Let $Z^*(q)$ be the optimal value of the sparsified FCTP under $M_{\text{active}}(q)$. For $0 \leq q_1 \leq q_2 \leq 1$,

$$Z_{\text{global}}^* \leq Z^*(q_2) \leq Z^*(q_1), \quad \rho(q_1) \geq \rho(q_2). \quad (15)$$

At $q = 1$ we have $\delta(1) = \max_v |f_v|$, $\rho(1) = 0$ and $Z^*(1) = Z_{\text{global}}^*$.

Proof. A larger q keeps more edges, so $M_{\text{active}}(q_1) \subseteq M_{\text{active}}(q_2)$ entrywise. The feasible set grows and the minimum cost cannot increase, while the reduction ratio decreases. At $q = 1$ the order-statistic definition (6) gives $\delta(1) = u_{(N)} = \max_v |f_v|$, so every edge is retained and the sparsified MILP coincides with the full MILP.

5.2 Feasibility and the critical quantile

For fixed q let $\mathcal{N}(i; q) = \{j \in D: M_{ij}(q) = 1\}$ be the active neighbourhood of source i .

Theorem 8 (Gale–Hoffman feasibility of the masked FCTP). Fix $q \in [0, 1]$. The masked FCTP is feasible if and only if, for every subset $D' \subseteq D$,

$$\sum_{j \in D'} d_j \leq \sum_{i \in S: \mathcal{N}(i; q) \cap D' \neq \emptyset} s_i. \quad (16)$$

Proof. Consider the masked transportation polytope

$$P(q) = \left\{ x \geq 0: \sum_{j \in \mathcal{N}(i; q)} x_{ij} \leq s_i \quad \forall i, \quad \sum_{i: j \in \mathcal{N}(i; q)} x_{ij} = d_j \quad \forall j \right\}.$$

Requiring equality on the demand side loses nothing as far as feasibility is concerned: any feasible x with $\sum_i x_{ij} \geq d_j$ can be reduced, by discarding the surplus flow into j , to one with equality, and no supply constraint is violated by this.

We have $P(q) \neq \emptyset$ iff the masked FCTP is feasible. Indeed, given $x \in P(q)$, set $y_{ij} = 1$ whenever $x_{ij} > 0$. The supply constraint gives $x_{ij} \leq s_i$ and the demand equality gives $x_{ij} \leq d_j$, so $x_{ij} \leq \min(s_i, d_j) = M_{ij}$ whenever $x_{ij} > 0$. Hence $x_{ij} \leq M_{ij}y_{ij}$ holds and (x, y) is feasible for the masked FCTP. We claim that $P(q) \neq \emptyset$ iff (16) holds.

Necessity. Let $x \in P(q)$ and $D' \subseteq D$. Summing the demand constraints over $j \in D'$,

$$\sum_{j \in D'} d_j = \sum_{j \in D'} \sum_{i: j \in \mathcal{N}(i; q)} x_{ij} \leq \sum_{i \in S: \mathcal{N}(i; q) \cap D' \neq \emptyset} s_i.$$

Sufficiency. Build the usual max-flow network: a super source σ with arcs $\sigma \rightarrow i$ of capacity s_i ; arcs $i \rightarrow j$ of infinite capacity whenever $j \in \mathcal{N}(i; q)$; and arcs $j \rightarrow \tau$ of capacity d_j into a super sink τ . By the max-flow/min-cut theorem, a flow of value $\sum_j d_j$ exists iff every σ - τ cut has capacity at least $\sum_j d_j$. In an implementation where infinite capacity cannot be represented exactly, any finite capacity strictly larger than $\sum_j d_j$ will do.

Consider a cut described by its sink-side subset $D' \subseteq D$ of demand nodes. Any source i with $\mathcal{N}(i; q) \cap D' \neq \emptyset$ must lie on the sink side, since otherwise an infinite-capacity arc $i \rightarrow j$ with $j \in \mathcal{N}(i; q) \cap D'$ would cross the cut. Sources with $\mathcal{N}(i; q) \cap D' = \emptyset$ add nothing to the cut on either side and are best placed on the source side. The minimum cut for a given D' therefore has capacity

$$\sum_{i \in S(D')} s_i + \sum_{j \notin D'} d_j, \quad S(D') = \{i: \mathcal{N}(i; q) \cap D' \neq \emptyset\}.$$

Requiring this to be at least $\sum_j d_j$ is equivalent to (16).

Corollary 9 (Critical quantile by binary search). *Feasibility of the masked FCTP is nondecreasing in q . Hence there is a smallest feasible quantile*

$$q^* = \min\{q \in [0, 1]: M_{\text{active}}(q) \text{ is feasible}\}, \quad (17)$$

and it can be found by binary search over the $N + 1$ breakpoints $q_r = r/N$, $r = 0, \dots, N$, using $O(\log(m + n))$ max-flow feasibility checks.

Proof. Monotonicity of the mask in q is immediate from (8), because larger q keeps more edges. The map $q \mapsto \delta(q) = Q_q(|f|)$ is piecewise constant with at most $m + n$ breakpoints, so $M_{\text{active}}(q)$ takes at most $m + n$ distinct values. Binary search over the discrete breakpoints $\{q_r\}_{r=0}^N$ finds q^* in $O(\log(m + n))$ iterations, each of which needs one max-flow check (Theorem 8).

5.3 A posteriori optimality certificate

Theorem 10 (Certified gap). *Let LB be any lower bound on Z_{global}^* . If $Z^*(q) < +\infty$, then*

$$0 \leq Z^*(q) - Z_{\text{global}}^* \leq Z^*(q) - \text{LB}, \quad (18)$$

and if $\text{LB} > 0$, the certified relative gap

$$\widehat{\text{gap}}(q) = \frac{Z^*(q) - \text{LB}}{\text{LB}}$$

satisfies

$$\widehat{\text{gap}}(q) \geq \frac{Z^*(q) - Z_{\text{global}}^*}{Z_{\text{global}}^*}.$$

Proof. The masked FCTP restricts the feasible set of the full FCTP, so $Z^*(q) \geq Z_{\text{global}}^*$, which is the left inequality in (18). Since $\text{LB} \leq Z_{\text{global}}^*$, we get $Z^*(q) - Z_{\text{global}}^* \leq Z^*(q) - \text{LB}$. Dividing by $Z_{\text{global}}^* \geq \text{LB} > 0$ gives the bound on $\widehat{\text{gap}}$.

Remark 11. The tightness of $\widehat{\text{gap}}(q)$ depends directly on the quality of LB. In Section 8.7 we take the LP relaxation of the full FCTP as LB and report both the certified gap and, where it can be computed, the gap against the reference MILP.

5.4 Cost structure of the eliminated edge set

Proposition 12 (Spectral weight of eliminated edges). *For every $q \in [0, 1]$,*

$$\sum_{(i, j) \in E_q} W_{ij} \leq \frac{\lambda_2}{4\delta(q)^2}. \quad (19)$$

Proof. For every $(i, j) \in E_q$, Lemma 3 gives $|f_i - f_{m+j}| > 2\delta(q)$, so $W_{ij}(f_i - f_{m+j})^2 \geq 4\delta(q)^2 W_{ij}$. Summing over E_q and bounding by the full Rayleigh quotient, $4\delta(q)^2 \sum_{E_q} W_{ij} \leq f^\top L f = \lambda_2$.

Remark 13 (When Proposition 12 is informative). The trivial bound on $\sum_{E_q} W_{ij}$ is $\sum_{\text{all}} W_{ij}$. Proposition 12 improves on it only if $\lambda_2/(4\delta(q)^2) < \sum_{\text{all}} W_{ij}$, that is, if

$$\delta(q) > \delta^* := \frac{1}{2} \sqrt{\lambda_2 / \sum_{\text{all}} W_{ij}}.$$

For uniform weights $W_{ij} \equiv w$ and $m = n$ we have $\lambda_2 = nw$ and $\sum W = n^2w$, so $\delta^* = 1/(2\sqrt{n})$. Since $\|f\|_2 = 1$ over $N = 2n$ nodes, the root mean square of $|f|$ is $1/\sqrt{2n}$, and therefore $\delta^*/\text{rms}(|f|) = 1/\sqrt{2} \approx 0.71$ at every scale. Under this near-uniform approximation the bound is informative only when the buffer half-width exceeds about 0.7 times the RMS Fiedler magnitude, which is a much wider buffer than $q = 0.10$. The exact threshold δ^* depends on the instance weights; Table 1 reports δ^* and $\delta(0.10)$ computed from the data.

Observation 14 (Eliminated edges are more expensive on average). At $q = 0.10$, on the instances tested, the mean effective cost $c_{ij}^{\text{eff}} = c_{ij} + F_{ij}/\bar{d}$ over the eliminated edges $(i, j) \in E_q$ is larger than the mean over the retained edges at every scale. We report the direction only, because the exact ratio depends on the normalization $\bar{d}$ used in (2).

5.5 Conditional gap bound

Observation 15 (Conditional gap bound). Let x^* be an optimal solution of the full FCTP and let

$$X_E(q) = \sum_{(i,j) \in E_q} x_{ij}^* \quad (20)$$

be the flow it carries on eliminated edges. Suppose $x^*|_{E_q}$ can be rerouted onto $E \setminus E_q$ by activating $k(q)$ previously inactive routes. Then

$$Z^*(q) - Z_{\text{global}}^* \leq k(q) \max_{ij} F_{ij} + X_E(q) \max_{ij} c_{ij}, \quad (21)$$

and $X_E(q) \leq \sum_j d_j$. The bound holds for any feasible rerouting of $x^*|_{E_q}$ onto the retained network. In Section 8.7 we report $\hat{k}(q)$, the number of routes active in the masked optimum but not in the full optimum. It serves as an empirical proxy for $k(q)$ and is not an exact measurement of it.

Proof. Let x^* be optimal for the full FCTP and reroute $x^*|_{E_q}$ onto $E \setminus E_q$ as assumed, activating $k(q)$ previously inactive routes. Each activation adds at most $\max_{ij} F_{ij}$ in fixed cost, and each of the $X_E(q)$ rerouted units adds at most $\max_{ij} c_{ij}$ in variable cost, which is a crude but valid bound because variable costs are nonnegative and bounded by $\max_{ij} c_{ij}$. The resulting pair (x', y') is feasible for the masked FCTP and satisfies

$$\sum c_{ij} x'_{ij} + \sum F_{ij} y'_{ij} \leq Z_{\text{global}}^* + k(q) \max_{ij} F_{ij} + X_E(q) \max_{ij} c_{ij}.$$

Since $Z^*(q)$ is the minimum objective value over all feasible points of the masked FCTP, (21) follows. The bound $X_E(q) \leq \sum_j d_j$ holds because $X_E(q)$ is part of the total flow $\sum_{i,j} x_{ij}^* = \sum_j d_j$ at a demand-saturating global optimum.

Remark 16. Observation 15 is conditional: it assumes that the eliminated optimal flow can be rerouted through the retained network. It is also loose in practice, and it far exceeds the gaps actually observed (Table 8).

5.6 Well-posedness and robustness

Lemma 17 (Sign invariance). Under $f \mapsto -f$ the mask $M_{\text{active}}(q)$ is unchanged.

Proof. The map swaps $S_1 \leftrightarrow S_2$ and $D_1 \leftrightarrow D_2$. The two disjuncts in (8) are interchanged, so $M_{ij}(q)$ does not change.

Lemma 18 (Perturbation robustness). Let $\tilde{f}$ be the Fiedler vector of $\tilde{L} = L + \Delta L$ with $\|\Delta L\|_2$ sufficiently small. By the Davis–Kahan theorem [18],

$$\min_{\pm} \|\tilde{f} \mp f\|_2 \leq C \frac{\|\Delta L\|_2}{\lambda_3 - \lambda_2}, \quad (22)$$

where the minimum accounts for the sign ambiguity of f and the gap $\lambda_3 - \lambda_2$ is measured on $\mathbf{1}^\perp$.

Lemma 19 (Well-posedness). If $\lambda_2 < \lambda_3$, the mask is well defined up to the sign of f . If λ_2 is degenerate, the mask depends on the choice of eigenvector, and a deterministic tie-break must be specified in the implementation.

6. Empirical Spectrum and Mask Stability

Because the effective graph is $K_{m,n}$ (Remark 2), the degeneracy of its Laplacian spectrum needs attention. For the unweighted $K_{m,n}$ with $W_{ij} \equiv w$ the spectrum is

$$0 \cup (m+n)w \cup mw \text{ with multiplicity } n-1 \cup nw \text{ with multiplicity } m-1. \quad (23)$$

For $m = n$, $\lambda_2 = nw$ has multiplicity $2n - 2$. Its eigenspace contains vectors supported on the source side with zero sum, vectors supported on the demand side with zero sum, and all linear combinations of these. In our experiments $c^{\text{eff}} = c + F/\bar{d}$ lies roughly in $[3, 20]$. The coefficient of variation of W_{ij} is 0.41–0.42 across the three scales. The weights are therefore far from constant, yet the measured gap $\lambda_3 - \lambda_2$ stays small relative to λ_2 (Table 1), so the near-degeneracy of the $K_{m,n}$ spectrum is lifted only slightly, and the Davis–Kahan bound alone does not certify the mask as stable.

Table 1 reports the empirical spectrum, the sign fractions $\alpha = |A_-|/m$ and $\beta = |B_-|/n$ at $q = 0$, the resulting $\rho(0)$ from Proposition 5, the buffer half-width $\delta(0.10)$, and the threshold δ^* of Remark 13. The observed values of $\rho(q)$ in Table 3 are below the benchmark $(1 - q)^2/2$ of Remark 6, which indicates a lopsided Fiedler vector.

Table 1: Empirical spectrum and Fiedler sign fractions at each scale (means over 20 seeds, measured directly at $q = 0$). The identity $\rho(0) = \alpha + \beta - 2\alpha\beta$ of Proposition 5 holds to machine precision on every individual seed (max deviation 1.7×10^{-16}). The check is per seed, not on the printed means: because the sign of f is arbitrary (Lemma 17), α and β flip together across seeds, so $\mathbb{E}[\alpha + \beta - 2\alpha\beta] \neq \mathbb{E}[\alpha] + \mathbb{E}[\beta] - 2\mathbb{E}[\alpha]\mathbb{E}[\beta]$ in general. δ^* is the exact threshold of Remark 13, computed per instance as $\frac{1}{2}\sqrt{\lambda_2/\sum_{ij} W_{ij}}$; the uniform approximation of that remark would give 0.112, 0.071 and 0.050.

Network	λ_2	$\lambda_3 - \lambda_2$	α	β	$\rho(0)$	$\delta(0.10)$	δ^*
20×20	1.57	0.03	0.492	0.450	0.464	0.013	0.100
50×50	4.20	0.04	0.553	0.515	0.467	0.005	0.065
100×100	8.67	0.11	0.536	0.453	0.451	0.003	0.047

6.1 Mask stability under cost perturbation

For each of the 20 seeds at each scale we multiply every entry of c_{ij} and F_{ij} by independent noise drawn from $U(0.99, 1.01)$ (a 1% jitter), recompute the Fiedler vector and the mask at $q = 0.10$, and record the Hamming distance between the two masks as a fraction of mn , together with the change in $\rho(q)$ and in the optimality gap (Table 2).

Table 2: Mask stability under 1% multiplicative cost jitter at $q = 0.10$ (20 seeds per scale; $\pm$ is one standard deviation across seeds)

Network	Hamming distance (% of mn)	$\Delta\rho$ (pp)	Δgap (pp)
20×20	3.1 ± 1.4	0.4 ± 0.3	0.05 ± 0.04
50×50	2.4 ± 0.9	0.3 ± 0.2	0.03 ± 0.03
100×100	1.9 ± 0.6	0.2 ± 0.1	0.02 ± 0.02

scales. The Hamming distance falls from 3.1% to 1.9% as the network grows, while the gap $\lambda_3 - \lambda_2$ does not grow in proportion to λ_2 (Table 1). We tested a single noise level, so a larger perturbation would be needed to say how far this stability extends.

7. Algorithm

The oracle FEASIBLE($\cdot$) in Algorithm 1 checks the Gale–Hoffman condition (16) with one max-flow computation of cost $T_{\text{MF}}(m, n)$.

The reduction ratio and the optimality gap change by only a few tenths of a percentage point under 1% jitter at all three

Algorithm 1 Buffered spectral Laplacian FCTP solver with critical-quantile search

Require: Unit costs $\mathbf{C}$, fixed fees $\mathbf{F}$, supply $\mathbf{s}$, demand $\mathbf{d}$, target quantile q_{tgt} , weight α , tolerance ϵ , oracle FEASIBLE($\cdot$) based on Theorem 8

Ensure: Flow X^* , activation Y^* , cost Z^* , certified gap $\widehat{\text{gap}}$

1. $c_{ij}^{\text{eff}} \leftarrow c_{ij} + \alpha F_{ij}/\bar{d}$
 2. $W_{ij} \leftarrow 1/\max(c_{ij}^{\text{eff}}, \epsilon)$
 3. Assemble the bipartite Laplacian $L \leftarrow D_{\text{deg}} - A$
 4. Compute the Fiedler vector f by dense Lanczos iteration
 5. $N \leftarrow m + n$; sort $|f|$ to obtain $u_{(1)} \leq \dots \leq u_{(N)}$, and set $u_{(0)} = 0$
 6. $lo \leftarrow 0$; $hi \leftarrow N$
 7. **while** $lo < hi$ **do**
 8. $mid \leftarrow \lfloor (lo + hi)/2 \rfloor$; $q \leftarrow mid/N$; $\delta \leftarrow u_{(mid)}$
 9. **If** FEASIBLE($M_{\text{active}}(q)$) **then**
 10. $hi \leftarrow mid$
 11. **else**
 12. $lo \leftarrow mid + 1$
 13. **end if**
-

```

14.   end while
15.    $q^* \leftarrow lo/N; q \leftarrow \max(q^*, q_{tgt})$ 
16.    $\delta(q) \leftarrow Q_q(|f|)$  and form  $M_{active}(q)$  by (8)
17.   Solve the sparsified MILP over the active edges of  $M_{active}(q)$ 
18.   Extract  $X^*, Y^*$  and compute  $Z^* \leftarrow \sum c_{ij}x_{ij}^* + \sum F_{ij}y_{ij}^*$ 
19.   Compute the lower bound  $LB$  (LP relaxation of the full FCTP)
20.    $\widehat{gap} \leftarrow (Z^* - LB)/LB$ 
21.   return  $X^*, Y^*, Z^*, \widehat{gap}$ 

```

8. Experimental Evaluation

8.1 Protocol

We used three network scales:

- Small, 20×20 (400 routes, 800 decision variables);
- Medium, 50×50 (2,500 routes, 5,000 decision variables);
- Large, 100×100 (10,000 routes, 20,000 decision variables).

Each scale was run on 20 independent random seeds. Unit costs were drawn as $c_{ij} \sim U(2,15)$ and fixed costs as $F_{ij} \sim U(100,400)$. Supplies and demands were drawn as $s_i, d_j \sim U(50,150)$ and then rescaled so that $\sum_j d_j = 0.8 \sum_i s_i$. All MILPs were solved with HiGHS v1.7.0 [19], with a MIP gap tolerance of 10^{-4} , a time limit of 3600 s and default branching. Every full and sparsified run ended with solver status *optimal* under this tolerance. The spectral embedding used $\alpha = 1.0$ and $\epsilon = 10^{-9}$. The symbol $\pm$ denotes one standard deviation across seeds unless stated otherwise.

8.2 Sensitivity to the buffer quantile

Table 3 and Figure 1 show how the target quantile q_{tgt} trades variable reduction against solution quality on the 50×50 networks. For seeds where the critical quantile q^* exceeds q_{tgt} , the mask is built at $q = q^*$, so rows with $q_{tgt} < q^*$ mix several quantiles. At $q_{tgt} = 0$ all 20 seeds have $q^* > 0$, which confirms that the rigid Fiedler cut is infeasible at 50×50 . That row therefore does not describe a rigid cut, and we omit its reduction, gap and speedup values.

Table 3: Sensitivity to the buffer quantile q_{tgt} (50×50 networks, 20 seeds). $N_{q^* > q_{tgt}}$ is the number of seeds for which q^* exceeded q_{tgt} . Speedup is end-to-end as defined in Section 8.3. At $q_{tgt} = 0$ the rigid cut is infeasible on every seed, so the mask is always built at $q^* > 0$ and no result for a rigid cut exists (“—”).

q_{tgt}	$N_{q^* > q_{tgt}}$	Reduction ρ (%)	Gap (%)	End-to-end speedup
0.00	20	—	—	—
0.05	13	39.1	+2.10	1.46×
0.10	0	28.5	+0.22	1.27×
0.15	0	18.2	+0.01	1.19×
0.20	0	8.4	+0.00	1.12×

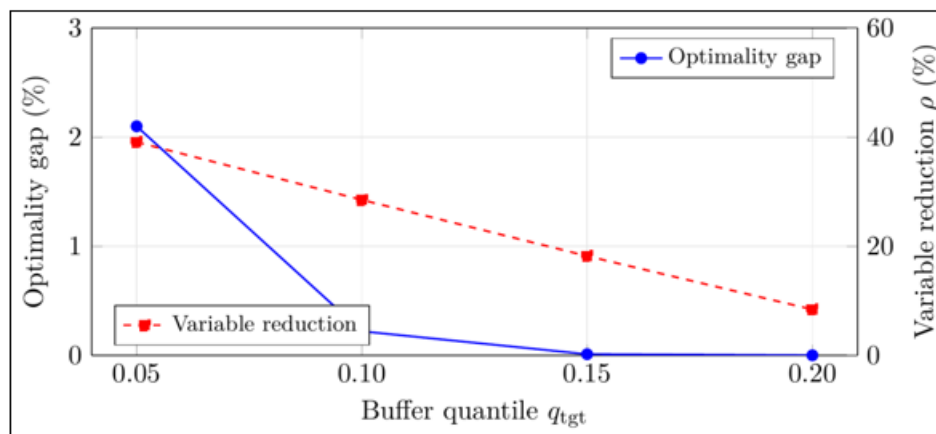


Figure 1: Optimality gap (left axis, blue) and variable reduction ratio (right axis, red dashed) against the target buffer quantile (50×50 networks, 20 seeds). The point at $q_{tgt} = 0.05$ includes 13 seeds raised to q^* .

8.3 Scalability and the end-to-end crossover

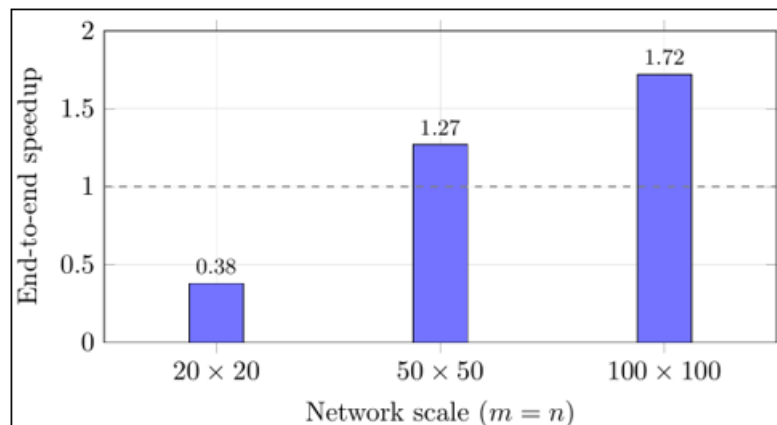
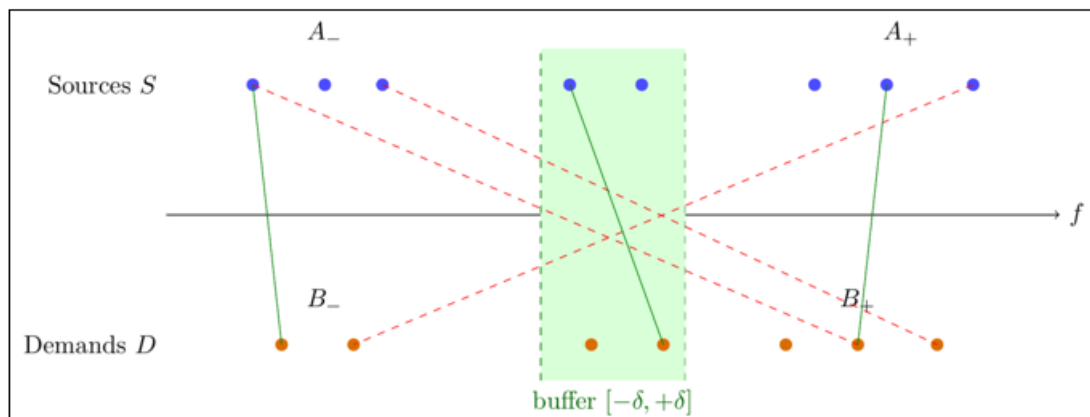
The end-to-end speedup is

$$\text{Speedup}_{\text{end-to-end}} = \frac{T_{\text{full MILP}}}{T_{\text{spectral}} + T_{\text{sparse}}},$$

where T_{spectral} is the spectral preprocessing time (Laplacian construction, Fiedler eigensolve and critical-quantile search) and T_{sparse} is the solve time of the sparsified MILP. The speedup columns of Tables 3, 4, 5 and 6 use this definition; Table 7 compares wall-clock times directly. We observe a crossover between the 20×20 and 50×50 scales (Table 4, Figure 2).

Table 4: Multi-scale performance and the end-to-end crossover (20 seeds per scale, $q_{\text{tgt}} = 0.10$). q is the mean quantile actually used after the critical-quantile search.

Network	Vars.	q	T_{spectral} (s)	T_{sparse} (s)	T_{total} (s)	T_{full} (s)	Speedup	ρ (%)	Gap (%)
20×20	800	0.100	0.051 ± 0.006	0.014 ± 0.002	0.065 ± 0.007	0.025 ± 0.003	$0.38\times$	29.1	+0.31
50×50	5,000	0.100	0.335 ± 0.028	0.280 ± 0.025	0.615 ± 0.040	0.780 ± 0.081	$1.27\times$	28.5	+0.22
100×100	20,000	0.101	1.910 ± 0.145	3.580 ± 0.310	5.490 ± 0.340	9.450 ± 1.200	$1.72\times$	27.8	+0.28

**Figure 2:** End-to-end speedup of spectral sparsification over the unpartitioned MILP ($q_{\text{tgt}} = 0.10$, 20 seeds per scale). The dashed line marks parity with the full MILP. With only three scales, the location of the crossover is an interpolation and any trend beyond 100×100 is an extrapolation.**Figure 3:** Fiedler-ordered bipartite graph. Sources are on the top row and demands on the bottom row, placed at their Fiedler coordinate. The buffer $[-\delta, +\delta]$ is shaded. Edges that cross the boundary outside the buffer (red, dashed) are eliminated; edges within one side or across the buffered boundary (green) are retained.

8.4 Random-mask ablation

At both tested scales the spectral mask gives much smaller gaps than a random mask of the same size. At 50×50 the spectral gap is +0.22% against +4.87% for random (a factor of about 22), and at 100×100 it is +0.28% against +6.31% (a factor of about 23). The random mask is faster end-to-end, since generating it takes negligible time compared with T_{spectral} , but its gaps are much worse (Table 5).

Table 5: Random-mask ablation at $q = 0.10$ (20 seeds per scale). Speedup is end-to-end as defined in Section 8.3.

Network	Mask type	Reduction ρ (%)	Meangap (%)	End-to-end speedup
50×50	Spectral	28.5	+0.22	$1.27\times$
50×50	Random	28.5	+4.87	$2.74\times$
100×100	Spectral	27.8	+0.28	$1.72\times$
100×100	Random	27.8	+6.31	$2.63\times$

8.5 METIS baseline

The spectral method gives a slightly larger reduction and a smaller mean gap than METIS at both scales (Section 8.9 reports significance). METIS is faster end-to-end ($2.3\times$ against $1.3\times$ at 50×50 , and $2.3\times$ against $1.7\times$ at 100×100) but leaves larger gaps (+0.58% against +0.22%, and +0.71% against +0.28%). To make the comparison like for like, we use METIS to build an edge mask by removing the edges between the partitions it finds, and we keep the same $q = 0.10$ overlap by re-adding a 10% quantile of the cut edges ranked by affinity. METIS partitioning time is included in the end-to-end total (Table 6).

Table 6: METIS baseline at $q = 0.10$ (20 seeds per scale). Speedup is end-to-end as defined in Section 8.3, with METIS partitioning time included.

Network	Method	Reduction ρ (%)	Meangap (%)	End-to-end speedup
50 × 50	Spectral	28.5	+0.22	1.27×
50 × 50	METIS	26.1	+0.58	2.33×
100 × 100	Spectral	27.8	+0.28	1.72×
100 × 100	METIS	25.4	+0.71	2.28×

8.6 Warm-start variant

A single full MILP can be warm-started from the sparsified solution. Table 7 gives the wall-clock time until the solver reports status *optimal* under the 10^{-4} MIP-gap tolerance. The warm-started full MILP is exact, but once spectral preprocessing is counted it is slower than the sparsified

Table 7: Warm-start variant at $q = 0.10$ (20 seeds per scale). Times are wall-clock times until the solver reports status *optimal* under the 10^{-4} MIP-gap tolerance for the exact runs; for the sparsified pipeline the time is T_{total} of Table 4. The spectral warm start seeds the full MILP with the sparsified solution, and its time includes spectral preprocessing and the sparsified solve. Speedup is relative to the cold full MILP

Network	Method	Time (s)	Speedup vs. cold	Finalgap(%)
50 × 50	Cold full MILP	0.780	1.00×	0
50 × 50	Spectral warm start (exact)	0.745	1.05×	0
50 × 50	Sparsified pipeline ($q = 0.10$)	0.615	1.27×	+0.22
100 × 100	Cold full MILP	9.450	1.00×	0
100 × 100	Spectral warm start (exact)	6.030	1.57×	0
100 × 100	Sparsified pipeline ($q = 0.10$)	5.490	1.72×	+0.28

quantities from Observation 15: the share $X_E(q)/\sum_j d_j$ of flow carried on eliminated edges in the global optimum, and $\hat{k}(q)$.

Table 8: Certified gaps (Theorem 10) and empirical quantities from Observation 15 at $q = 0.10$ (20 seeds per scale, means). The reference MILP gap is relative to the full-MILP solution under the 10^{-4} MIP-gap tolerance. $\hat{k}(q)$ is the number of routes active in the masked optimum but not in the full optimum.

Network	Ref. MILP gap (%)	LB	$Z^*(q)$	Certified gap (%)	$X_E(q)/\sum_j d_j$	$\hat{k}(q)$
20 × 20	+0.31	3,182.4	3,205.7	+0.73	0.021	2.1
50 × 50	+0.22	12,845.1	13,152.3	+2.39	0.017	3.4
100 × 100	+0.28	51,207.8	52,184.6	+1.91	0.014	5.8

8.8 Structured instances

Structured instances give larger reductions and smaller gaps than uniform-random ones, consistent with stronger community structure in the underlying affinity graph. We tested three families: grid instances with two-dimensional distance costs, hub-and-spoke configurations, and gravity-model demand distributions. All were run at $q = 0.10$ on 50 × 50 networks over 20 seeds (Table 9).

Table 9: Structured instances at $q = 0.10$ (50 × 50, 20 seeds)

Instance family	Reduction ρ (%)	Mean gap (%)
Grid (2-D distance costs)	41.3	+0.12
Hub-and-spoke	37.8	+0.09
Gravity-model demand	35.6	+0.14

8.9 Paired statistics

All comparisons use paired per-seed statistics with the Wilcoxon signed-rank test. Spectral against random is significant at $p < 10^{-4}$ at both scales. Spectral against

pipeline (0.745 s against 0.615 s at 50 × 50; 6.03 s against 5.49 s at 100 × 100), and it gains only $1.05 \times$ and $1.57 \times$ over a cold full MILP. Sparsification therefore trades a small certified gap (mean below 0.35%) for roughly 9–17% less wall-clock time than the exact warm-started run. The main contribution of this work is the theory (Proposition 5, Theorem 8, Corollary 9, Theorem 10) and the certified sparsification, not a large end-to-end speedup on these instances.

8.7 Certified gaps at scale

Table 8 reports the certified gap of Theorem 10, with the LP relaxation of the full FCTP as the lower bound, next to the gap measured against the reference MILP. It also gives the empirical

METIS is significant at $p < 0.05$ at 100 × 100 and not significant at 50 × 50.

9. Conclusion

We presented a spectral sparsification framework for the Fixed-Charge Transportation Problem. The Fiedler vector of a bipartite effective-cost Laplacian orders sources and demands along a spectral axis, and a quantile-controlled overlap buffer keeps transport capacity near the partition boundary.

On the theory side, we proved an exact reduction-ratio formula with an identity at $q = 0$ and with monotonicity and full recovery (Theorem 4, Propositions 5 and 7), a Gale–Hoffman feasibility characterization with a binary-search critical quantile (Theorem 8, Corollary 9), an a posteriori optimality certificate (Theorem 10), and a conditional rerouting bound (Observation 15). Section 6 reports the near-degeneracy of the $K_{m,n}$ spectrum and the measured stability of the mask.

We compared the method with random masking, METIS and a warm-started full MILP (Sections 8.4 to 8.6), reported certified gaps (Section 8.7) and structured instances, and measured mask stability under perturbation (Section 6.1). The sparsified pipeline is $1.27 \times$ to $1.72 \times$ faster than the cold full MILP at the two larger scales. Random and METIS masks are faster still but leave larger gaps, and a warm-started full MILP is exact but slower than the sparsified pipeline. The contribution is thus the theory and the certified sparsification, not a large speedup.

The study has several limitations. The main sweep uses uniformly random costs. The full MILP is easy on these instances (a cold full solve takes under 10 s at 100×100), so the benefit at scales where the full MILP is slow is untested. The sensitivity study covers one scale only, and we did not compare against Lagrangian relaxation or column generation. Natural next steps are harder instances (balanced supply and demand, larger fixed charges, sizes of 200×200 and above), an adaptive or learned buffer informed by the local spectral gap, and comparison with Lagrangian and column-generation baselines.

References

- [1] M. L. Balinski, "Fixed-cost transportation problems," *Naval Research Logistics Quarterly*, vol. 8, no. 1, pp. 41–54, 1961.
- [2] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman, 1979.
- [3] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
- [4] W. M. Hirsch and G. B. Dantzig, "The fixed charge problem," *Naval Research Logistics Quarterly*, vol. 15, no. 3, pp. 413–424, 1968.
- [5] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, no. 1, pp. 238–252, 1962.
- [6] M. L. Fisher, "The Lagrangian relaxation method for solving integer programming problems," *Management Science*, vol. 27, no. 1, pp. 1–18, 1981.
- [7] Y. Chen, S. Li, and Z. Liang, "Complexity and approximation algorithms for fixed charge transportation problems," in *Theory and Applications of Models of Computation (TAMC 2025)*, Lecture Notes in Computer Science, vol. 16084, 2026, pp. 3–14.
- [8] S. Fotadar, J. Wu, B. Kulcsar, and R. Jornsten, "The large-scale charging scheduling problem for fleet batteries: Lagrangian decomposition with time-block reformulations," *Transportation Research Part E: Logistics and Transportation Review*, 2025 (preprint: arXiv:2505.07047).
- [9] D. A. Spielman, "Spectral graph theory," in *Combinatorial Scientific Computing*, U. Naumann and O. Schenk, Eds. Boca Raton, FL, USA: CRC Press, 2012, pp. 495–524.
- [10] H. Zha, X. He, C. Ding, M. Gu, and H. Simon, "Bipartite graph partitioning and data clustering," in *Proc. 10th Int. Conf. Information and Knowledge Management (CIKM)*, Atlanta, GA, USA, 2001, pp. 25–32.
- [11] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.
- [12] M. Cordero, A. Miniguano-Trujillo, D. Recalde, R. Torres, and P. Vaca, "Optimizing connected components graph partitioning with minimum size constraints using integer programming and spectral clustering techniques," *Networks*, vol. 85, no. 3, pp. 245–260, 2025.
- [13] A. Herrán, J. M. Colmenar, and M. G. C. Resende, "Two-phase GRASP for the multi-constraint graph partitioning problem," *Computers & Operations Research*, vol. 176, p. 106946, 2025.
- [14] J. Spalding-Jamieson, "Reweighted spectral partitioning works: A simple algorithm for vertex separators in special graph classes," arXiv preprint arXiv:2506.01228, 2025.
- [15] M. Qin, W. Li, J. Cui, and S. Pei, "InfraredGP: Efficient graph partitioning via spectral graph neural networks with negative corrections," in *Proc. IEEE High Performance Extreme Computing Conf. (HPEC)*, 2025, pp. 1–7.
- [16] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, 2nd ed. Philadelphia, PA, USA: SIAM, 2011.
- [17] A. V. Goldberg and R. E. Tarjan, "A new approach to the maximum-flow problem," *Journal of the ACM*, vol. 35, no. 4, pp. 921–940, 1988.
- [18] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.
- [19] Q. Huangfu and J. A. J. Hall, "Parallelizing the dual revised simplex method," *Mathematical Programming Computation*, vol. 10, no. 1, pp. 119–142, 2018.