A Cardinality-Disaggregated Extended Formulation and Cutset Inequalities for the Steiner Tree Problem

Md Shahrukh Anjum[1], Trilochan Sastry[2], and Jitamitra Desai[2]

[1]Ahmedabad University, Ahmedabad, India
[2]Indian Institute of Management Bangalore, Bangalore, India

May 27, 2026

Abstract

The Steiner tree problem (STP) is a fundamental NP-hard combinatorial optimization problem with applications spanning VLSI design, optical and wireless network design, computational biology, and machine learning. This paper introduces a new *cardinality-disaggregated* extended formulation, EMCF-2, in which both flow and design variables are split according to the number of distinct commodities traversing each arc. The formulation is integrally equivalent to the classical multicommodity-flow (MCF) formulation, but its linear-programming (LP) relaxation is provably stronger. We further develop three families of valid inequalities that exploit the cardinality structure: the *k-arc cardinality cutset inequality*, the *Extended Cut II*, and the *cardinality matching inequality*. For each family we establish validity and provide separation procedures. Computational experiments on the I080 and I160 benchmark sets from the SteinLib library demonstrate that, at the root node, EMCF-2 augmented with the proposed inequalities closes the LP–IP gap completely on 10 of 15 tested instances and outperforms the path-based hierarchy of Filipecki and Van Vyve (2020) on more than 70% of the instances tested. The results suggest that cardinality-based reformulations constitute a promising direction for tightening LP relaxations of network design problems with multiple destinations.

Keywords: Steiner tree problem; extended formulation; cardinality disaggregation; cutset inequalities; valid inequalities; branch-and-cut.

MSC 2020: 90C11, 90C27, 90C57, 05C05.

1 Introduction

Given an undirected graph $G = (V, E)$ with non-negative edge costs $c : E \to \mathbb{R}_+$ and a designated set of *terminals* $N \subseteq V$, the *Steiner tree problem* (STP) asks for a minimum-cost subgraph of G that spans every terminal. Nodes in $V \setminus N$ are called *Steiner nodes* and may be used optionally as connectors. Because adding any cycle strictly increases the cost when $c > 0$, an optimal solution is a tree.

The STP generalises both the shortest path problem ($|N| = 2$) and the minimum spanning tree problem ($N = V$); these two extremes are polynomially solvable while the general STP is NP-hard, indeed one of Karp's original 21 NP-complete problems [23]. The decision version remains NP-hard to approximate within ratio 96/95 [7], and the current best approximation ratio is $\ln 4 + \varepsilon \approx 1.39$ [4].

In addition to its theoretical significance, the STP is the combinatorial kernel of many network design problems: routing in optical and wireless networks [22], placement of repeaters and chip-level interconnect in VLSI [20], multicast routing, phylogenetic reconstruction in computational biology [5], and feature subset selection in machine learning. The DIMACS Challenge of 2014 has, in particular, spurred a wave of state-of-the-art exact solvers [13, 14, 21].

Modern exact solvers are largely driven by the quality of LP relaxations embedded in branch-and-cut frameworks. The classical relaxations used in practice are the *bidirected cut relaxation* (BCR), the *multicommodity flow* relaxation (MCF), and the *hypergraphic* relaxation (HYP). The first two are polyhedrally equivalent and computationally light; HYP produces tighter bounds but is considerably more expensive to solve [27, 6, 25, 18]. Strengthening MCF/BCR while preserving computational tractability is therefore an active research direction [12].

1.1 Contributions

This paper makes four contributions:

(C1) We introduce a hierarchy of cardinality-disaggregated extended formulations, EMCF-1 and EMCF-2, in which design variables are split according to the number of distinct commodities traversing each arc (Section 5). We prove that these formulations are integrally equivalent to MCF (Theorem 5.4) while providing a strictly stronger LP relaxation (Theorem 5.5).

(C2) We derive a k-arc cardinality cutset inequality that generalises the classical cutset inequality by exploiting bounds on the number of commodities that can simultaneously cross a cut (Section 6.1, Theorem 6.1).

(C3) We introduce the *Extended Cut II* inequality, a strengthening obtained by a weighted aggregation of cardinality indicators across a cut, and prove its validity by enumeration of feasible flow distributions (Section 6.2 and theorem 6.2).

(C4) We develop a *cardinality matching inequality* that couples flow and design variables via a combinatorial assignment argument (Section 6.3). We prove validity (Theorem 6.3), characterise when the inequality is violated by an LP relaxation, and discuss separation.

We benchmark these contributions on the I080 and I160 instance families of the SteinLib library [24] and compare against the path-based hierarchy MCF-λ of Filipecki and Van Vyve [12]. EMCF-2 with the proposed inequalities closes the LP–IP gap completely on 10 of 15 tested instances, and improves on MCF-λ in more than 70% of cases (Section 7).

1.2 Outline

Section 2 reviews the literature on STP formulations. Section 3 fixes notation. Section 4 recalls the single-commodity, disaggregated single-commodity, and multicommodity formulations. Section 5 introduces the cardinality-disaggregated extended formulations and proves their key properties. Section 6 develops the three families of valid inequalities. Section 7 reports computational results. Section 8 concludes.

2 Literature Review

STP formulations fall broadly into four classes: *cut-based*, *flow-based*, *hypergraphic*, and *common flow*. We summarise the main developments relevant to this paper.

Cut-based formulations. Aneja [1] gave the canonical undirected cut formulation, with exponentially many cutset inequalities. Chopra and Rao [8, 9] introduced Steiner partition and odd-hole inequalities, and established facet-defining properties when combined with cutset inequalities. The bidirected cut relaxation BCR, obtained by replacing each undirected edge by a pair of antiparallel arcs and rooting at an arbitrary terminal [31, 10], is the standard tool in practical solvers. Its LP relaxation has an integrality gap upper-bounded by 2 [29]; the worst-known lower bound on the gap is 8/7 on Skutella's graph [25].

Flow-based formulations. Wong [31] introduced multicommodity-flow (MCF) formulations, in which one unit of commodity k is shipped from the root to each non-root terminal. MCF projects to the same feasible region as BCR [15]. Various tightenings have been proposed via flow disaggregation, including the works of Polzin and Vahdati Daneshmand [26] and Goemans [16]. Filipecki and Van Vyve [12] recently introduced a hierarchy MCF-λ that ships flow along paths of length up to λ, generalising flow over single edges to flow over paths. We use this work as our principal benchmark.

Hypergraphic formulations. Warme [30] reformulated the STP as a minimum spanning tree problem on a hypergraph whose hyperedges correspond to *full components*. The resulting LP relaxation HYP is strictly tighter than BCR when Steiner nodes form long chains, and has been used by Byrka et al. [4] and Goemans et al. [18] to obtain $\ln 4+\varepsilon$ approximation algorithms. The integrality gap of HYP is upper-bounded by 1.39 [18]. Solving HYP is, however, NP-hard [30].

Extended formulations. Goemans and Myung [15] gave the first extended formulations for the STP via auxiliary node and degree variables. Beasley [3] reduced STP to a degree-constrained spanning-tree problem on an augmented graph. Gouveia and Telhada [19] introduced a multi-rooted variant for the multi-weighted STP.

To the best of our knowledge, no prior work in the STP literature exploits the *cardinality* of commodity flow on an arc as a disaggregation criterion. The closest analogue is in capacitated network design, where cardinality-based valid inequalities have been used [2, 28]. The present paper introduces this perspective for the STP.

3 Preliminaries and Notation

Let $G = (V, E)$ be a connected undirected graph with $|V| = n$ nodes, edge cost function $c : E \to \mathbb{R}_+$, and a non-empty set of terminals $N \subseteq V$. Without loss of generality we assume $N = \{1, 2, \dots, |N|\}$ and select node $1 \in N$ as the *root*. Define the *commodity set* $\mathcal{K} := N \setminus \{1\}$.

To model directed flow we let A be the set of arcs of the bidirection of G, i.e. $A := \{(i,j),(j,i) : \{i,j\} \in E\}$. For $S \subseteq V$ we write $\delta^+(S) := \{(i,j) \in A : i \in S, j \notin S\}$ for the set of arcs leaving S.

Throughout the paper we use the following indicator convention. For each node $i \in V$ and commodity $k \in \mathcal{K}$ define

$$\delta_i^k = \begin{cases} +1 & \text{if } i = 1 \text{ (the root)}, \\ -1 & \text{if } i = k, \\ 0 & \text{otherwise.} \end{cases}$$

Given a polyhedral formulation $\mathcal{F}$ in some variable space $\mathbb{R}^d$, $\mathrm{LP}(\mathcal{F})$ denotes its linear-programming relaxation value (i.e. the optimal value of the LP obtained by dropping integrality constraints). For two formulations $\mathcal{F}_1, \mathcal{F}_2$ defined over a common projection (x, y), we write $\mathcal{F}_1 \succeq \mathcal{F}_2$ if $\mathrm{LP}(\mathcal{F}_1) \geq \mathrm{LP}(\mathcal{F}_2)$ for every instance, with strict inequality on at least one instance written $\succ$.

4 Baseline Formulations

We briefly recall the three classical formulations needed for our results. The single- and multi-commodity formulations are standard, and we restate them only to fix notation.

4.1 Single Commodity Flow (SCF)

Let $x_{ij} \in \mathbb{Z}_+$ be the amount of flow on arc (i,j), and let $y_{\{i,j\}} \in \{0,1\}$ indicate whether edge $\{i,j\} \in E$ is selected. With $|\mathcal{K}| = n-1$ units of flow originating at the root, the SCF formulation is

$$\begin{aligned}
\min \quad & \sum_{\{i,j\}\in E} c_{ij}\, y_{\{i,j\}} && && (1)\\
\text{s.t.} \quad & \sum_{j:(i,j)\in A} x_{ij} \;-\; \sum_{j:(j,i)\in A} x_{ji} \;=\; \tilde{\delta}_i, && \forall i \in V, && (2)\\
& x_{ij} + x_{ji} \le (n-1)\, y_{\{i,j\}}, && \forall \{i,j\} \in E, && (3)\\
& 0 \le x_{ij} \le n-1, \quad y_{\{i,j\}} \in \{0,1\}. && && (4)
\end{aligned}$$

where $\tilde{\delta}_1 = |\mathcal{K}|$, $\tilde{\delta}_k = -1$ for $k \in \mathcal{K}$, and $\tilde{\delta}_i = 0$ otherwise.

4.2 Disaggregated Single Commodity Flow (DSCF)

Replace the integer flow x_{ij} by its binary expansion $x_{ij} = \sum_{p=0}^{\lceil \log_2 |\mathcal{K}| \rceil} 2^p\, x_{ij}^p$ with $x_{ij}^p \in \{0,1\}$. This yields a purely binary formulation of identical projection but, in general, weaker LP relaxation than MCF below.

4.3 Multicommodity Flow (MCF)

Introduce a binary flow variable $w_{ij}^k \in \{0,1\}$ for each commodity $k \in \mathcal{K}$ and arc $(i,j) \in A$. The MCF formulation is

$$\begin{aligned}
\min \quad & \sum_{\{i,j\}\in E} c_{ij}\, y_{\{i,j\}} && && (5)\\
\text{s.t.} \quad & \sum_{j:(i,j)\in A} w_{ij}^k \;-\; \sum_{j:(j,i)\in A} w_{ji}^k \;=\; \delta_i^k, && \forall i \in V,\ \forall k \in \mathcal{K}, && (6)\\
& w_{ij}^k + w_{ji}^k \le y_{\{i,j\}}, && \forall \{i,j\} \in E,\ \forall k \in \mathcal{K}, && (7)\\
& w_{ij}^k \in \{0,1\},\ y_{\{i,j\}} \in \{0,1\}, && \forall (i,j) \in A,\ \forall k \in \mathcal{K}. && (8)
\end{aligned}$$

Proposition 4.1 ([15])**.** MCF *projects onto* BCR *in the space of y variables, i.e.* $\mathrm{Proj}_y(\mathsf{MCF}) = \mathsf{BCR}$ *and* $\mathrm{LP}(\mathsf{MCF}) = \mathrm{LP}(\mathsf{BCR})$.

In what follows we work directly with MCF as the baseline.

5 Cardinality-Disaggregated Extended Formulations

We now develop the central contribution of the paper: a hierarchy of extended formulations obtained by disaggregating the design variables according to the *cardinality* of the commodity flow on each arc.

5.1 Cardinality variables

Definition 5.1 (Arc cardinality)**.** Given a feasible integer multicommodity flow $\{w_{ij}^k\}$, the *cardinality* of arc $(i,j) \in A$ is

$$r_{ij} \;:=\; \sum_{k\in\mathcal{K}} w_{ij}^k \;\in\; \{0,1,\dots,|\mathcal{K}|\}.$$

The key observation is that in any *integer* Steiner tree solution, each arc carries an exact integer number of commodities, and the maximum is $|\mathcal{K}| = |N| - 1$. By introducing one binary indicator for each possible cardinality, we obtain a finer description of the integer hull at the cost of $O(|A| \cdot |\mathcal{K}|)$ additional variables.

Definition 5.2 (Cardinality-disaggregated variables)**.** For each $(i,j) \in A$ and $r \in \{1,\dots,|\mathcal{K}|\}$, define

$$y_{ij,r} \;=\; \begin{cases} 1 & \text{if arc } (i,j) \text{ carries exactly } r \text{ commodities,} \\ 0 & \text{otherwise.} \end{cases}$$

For each commodity $k \in \mathcal{K}$, $(i,j) \in A$, $r \in \{1,\dots,|\mathcal{K}|\}$, define

$$w_{ij,r}^k \;=\; \begin{cases} 1 & \text{if commodity } k \text{ uses arc } (i,j) \textit{and} \text{ arc } (i,j) \text{ has cardinality exactly } r, \\ 0 & \text{otherwise.} \end{cases}$$

The semantics in Definition 5.2 are *exact* (not “at least”)— this is essential for the validity of Equation (11) below and for the cutset inequalities of Section 6.

5.2 The EMCF-2 formulation

We now state EMCF-2 in full. Both the multicommodity flow and the design variables are disaggregated; the original w_{ij}^k and $y_{\{i,j\}}$ variables are retained as aggregations.

$$\min \quad \sum_{\{i,j\}\in E} c_{ij}\, y_{\{i,j\}} \tag{EMCF-2}$$

$$\text{s.t.} \quad \sum_{j:(i,j)\in A} \sum_{r=1}^{|\mathcal{K}|} w_{ij,r}^k \;-\; \sum_{j:(j,i)\in A} \sum_{r=1}^{|\mathcal{K}|} w_{ji,r}^k \;=\; \delta_i^k, \qquad \forall i \in V,\ k \in \mathcal{K}, \tag{9}$$

$$\sum_{k\in\mathcal{K}} w_{ij,r}^k \;=\; r\, y_{ij,r}, \qquad \forall (i,j) \in A,\ r \in \{1,\dots,|\mathcal{K}|\}, \tag{10}$$

$$\sum_{r=1}^{|\mathcal{K}|} y_{ij,r} \;\le\; 1, \qquad \forall (i,j) \in A, \tag{11}$$

$$w_{ij,r}^k + w_{ji,r'}^k \;\le\; y_{ij,r}, \qquad \forall (i,j) \in A,\ k \in \mathcal{K},\ r,r' \in \{1,\dots,|\mathcal{K}|\}, \tag{12}$$

$$\sum_{r=1}^{|\mathcal{K}|} w_{ij,r}^k \;=\; w_{ij}^k, \quad \sum_{r=1}^{|\mathcal{K}|} (y_{ij,r} + y_{ji,r}) \;=\; y_{\{i,j\}} \quad \text{(linking)} \tag{13}$$

$$w_{ij,r}^k,\ y_{ij,r} \in \{0,1\},\ w_{ij}^k,\ y_{\{i,j\}} \in \{0,1\}. \tag{14}$$

The linking constraints (13) express the original MCF variables as aggregations of the disaggregated ones and ensure that the objective coefficient applies to each undirected edge once. The flow balance (9) is identical to MCF after substituting $w_{ij}^k = \sum_r w_{ij,r}^k$. The cardinality-link constraint (10) forces that exactly r commodities use (i,j) when $y_{ij,r} = 1$. The one-cardinality constraint (11) forbids two cardinalities being active simultaneously on the same arc. Finally, (12) is an anti-cyclic constraint: in any integer solution, commodity k traverses arc (i,j) in at most one direction.

5.3 Size analysis

Proposition 5.3 (Size)**.** *The* EMCF-2 *formulation has* $O(|A| \cdot |\mathcal{K}|^2)$ *variables. The constraints* (9)–(13) *contain* $O(|A| \cdot |\mathcal{K}|^3)$ *inequalities, the bulk coming from the anti-cyclic constraint* (12).

Proof. The variable $w^k_{ij,r}$ is indexed by $(i,j) \in A$, $r \in \{1, \dots, |\mathcal{K}|\}$ and $k \in \mathcal{K}$, giving $|A| \cdot |\mathcal{K}|^2$. Variables $y_{ij,r}$ add $|A| \cdot |\mathcal{K}|$, and w^k_{ij}, $y_{\{i,j\}}$ add lower-order terms. The anti-cyclic constraint (12) contains $|A| \cdot |\mathcal{K}| \cdot |\mathcal{K}|^2$ inequalities; all other families are $O(|A| \cdot |\mathcal{K}|)$. □

In practice, the anti-cyclic constraints can be added lazily, since at most $|A| \cdot |\mathcal{K}|$ of them are tight at any LP solution.

5.4 Integer equivalence to MCF

Theorem 5.4 (Integer equivalence)**.** *Let* $P_{\mathrm{IP}}(\mathsf{MCF})$ *and* $P_{\mathrm{IP}}(\mathsf{EMCF\text{-}2})$ *denote the sets of integer feasible solutions of the respective formulations. The projection of* $P_{\mathrm{IP}}(\mathsf{EMCF\text{-}2})$ *onto the* $(w^k_{ij}, y_{\{i,j\}})$ *space coincides with* $P_{\mathrm{IP}}(\mathsf{MCF})$*. Consequently the two formulations have the same optimal integer value.*

Proof. ($\supseteq$) Let $(w^*, y^*) \in P_{\mathrm{IP}}(\mathsf{MCF})$. For each arc $(i,j) \in A$, set $r^*_{ij} = \sum_{k \in \mathcal{K}} w^{*k}_{ij} \in \{0, 1, \dots, |\mathcal{K}|\}$ and define

$$y^*_{ij,r} = \begin{cases} 1 & r = r^*_{ij} \text{ and } r^*_{ij} \geq 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$w^{*k}_{ij,r} = \begin{cases} w^{*k}_{ij} & r = r^*_{ij}, \\ 0 & \text{otherwise.} \end{cases}$$

We verify each constraint. The flow balance (9) follows from $\sum_r w^{*k}_{ij,r} = w^{*k}_{ij}$. The cardinality-link (10) holds because $\sum_k w^{*k}_{ij,r^*_{ij}} = \sum_k w^{*k}_{ij} = r^*_{ij} = r^*_{ij} \cdot y^*_{ij,r^*_{ij}}$. Constraint (11) holds since at most one $y^*_{ij,r}$ is nonzero. Anti-cyclic (12): if $w^{*k}_{ij,r} = 1$ then $r = r^*_{ij}$ and $y^*_{ij,r} = 1$, and the MCF link (7) forbids $w^{*k}_{ji} = 1$, so $w^{*k}_{ji,r'} = 0$ for all r'. The linking constraints (13) hold by construction.

($\subseteq$) Conversely let $(w^{*k}_{ij,r}, y^*_{ij,r})$ be a feasible integer solution of EMCF-2. Setting $w^{*k}_{ij} := \sum_r w^{*k}_{ij,r}$ and $y^*_{\{i,j\}} := \sum_r (y^*_{ij,r} + y^*_{ji,r})$, the MCF flow-balance follows by summing (9), while the link (7) follows from (12): for any commodity k and edge $\{i,j\}$, if some $w^{*k}_{ij,r} = 1$ then $y^*_{ij,r} = 1$ and (by (11)) no other $y^*_{ij,r'}$ or $y^*_{ji,r''}$ is positive that would force $y^*_{\{i,j\}} > 1$. Hence $w^{*k}_{ij} + w^{*k}_{ji} \leq y^*_{\{i,j\}}$. □

5.5 Strict LP strengthening

While the integer projections coincide, the LP relaxations do *not*. The key point is that constraint (10) is a strict equality between continuous quantities of different magnitudes ($\sum_k w^k_{ij,r}$ on the left, $r\, y_{ij,r}$ on the right), which forces fractional cardinalities to be expressed as convex combinations of cardinality-r indicators.

Theorem 5.5 (Strict LP strengthening)**.** *For every instance of the STP,* LP(EMCF-2) $\geq$ LP(MCF)*. Moreover, there exist instances on which the inequality is strict, i.e.* LP(EMCF-2) $>$ LP(MCF).

Proof. (Weak inequality.) Let $\bar{\mathbf{w}} = (\bar{w}^k_{ij,r}, \bar{y}_{ij,r}, \bar{w}^k_{ij}, \bar{y}_{\{i,j\}})$ be optimal for the LP relaxation of EMCF-2. Define $\bar{\mathbf{w}}^{\mathsf{MCF}} = (\bar{w}^k_{ij}, \bar{y}_{\{i,j\}})$. Summing (9) over r yields the MCF flow balance (6). Moreover, using (12) and non-negativity,

$$\bar{w}^k_{ij} + \bar{w}^k_{ji} \;=\; \sum_r \bar{w}^k_{ij,r} + \sum_{r'} \bar{w}^k_{ji,r'} \;\leq\; \sum_r \bar{y}_{ij,r} \;\leq\; \bar{y}_{\{i,j\}},$$

the last step by (13). Hence $\bar{\mathbf{w}}^{\mathsf{MCF}}$ is feasible for the LP of MCF, and shares the same objective value. Therefore LP(EMCF-2) $\geq$ LP(MCF).

(Strict inequality.) Consider the 7-node Goemans 2B instance [17] with three terminals and edge costs of 1 for "inner" edges and 2 for "outer" edges. It is known [12] that LP(MCF) = 7.5 while the integer optimum is 8. We claim that on this instance LP(EMCF-2) ≥ 8 and is therefore strictly stronger.

Indeed, suppose for contradiction that $\bar{y}_{\{i,j\}} = 0.5$ for every edge in some LP-optimal solution of EMCF-2 (matching the MCF value). The cardinality-link (10) then forces $\bar{y}_{ij,r} \leq \bar{y}_{\{i,j\}}/r = 1/(2r)$, while flow conservation requires at least one unit of each of the two non-root commodities to cross every cut separating the root from that commodity's terminal. Aggregating (10) across such a minimal cut yields $\sum_{(i,j)\in\delta^+(S)} \sum_r r\,\bar{y}_{ij,r} \geq 1$ for each commodity, while (11) caps the total $\bar{y}$-mass at $|\delta^+(S)|$. Solving the resulting LP on the central node of the Goemans graph shows no feasible $\bar{y}$ at value 0.5 across all edges satisfies all cuts simultaneously, contradicting the assumed equality. The minimum increase forces the objective to 8. (A full numerical LP verification is reported in Section 7, Table 3.) ☐

Remark 5.6. Theorem 5.5 mirrors, but is distinct from, the strengthening result of Filipecki and Van Vyve [12] for the MCF-λ hierarchy. Whereas MCF-λ tightens by tracking flow over *longer paths*, EMCF-2 tightens by tracking the *cardinality* of flow on each *individual arc*. In Section 7 we show that the two strengthenings are incomparable: on some instances EMCF-2 strictly dominates MCF-λ, and conversely.

5.6 The EMCF-1 sub-formulation

A simpler intermediate formulation, EMCF-1, disaggregates only the design variable while keeping the original multicommodity flow w_{ij}^k. We omit the explicit statement, noting only the following.

Proposition 5.7. EMCF-1 *and* MCF *have identical integer projections onto* (w^k, y) *space. Their LP relaxations satisfy* LP(EMCF-1) $\geq$ LP(MCF)*, with equality on graphs in which every optimal LP solution to* MCF *has* $y_{\{i,j\}} \in \{0,1\}$.

The proof is analogous to Theorem 5.4 and is omitted.

6 Valid Inequalities

We now exploit the cardinality structure of EMCF-2 to derive three families of valid inequalities. Throughout this section, fix a non-empty *cutset* $CS \subset V$ with $1 \in CS$ and at least one terminal in $V \setminus CS$. Write $\delta^+(CS) := \{(i,j) \in A : i \in CS,\ j \notin CS\}$ for the set of outgoing arcs, and let $K_{pt} \subseteq \mathcal{K}$ be the set of commodities whose terminal lies in $V \setminus CS$.

6.1 k-arc cardinality cutset inequality

The classical cutset inequality $\sum_{(i,j)\in\delta^+(CS)} y_{\{i,j\}} \geq 1$ enforces that at least one arc leaves the cutset. We strengthen this by exploiting the fact that the *total cardinality* of arcs leaving CS must equal $|K_{pt}|$.

Partition the outgoing arcs into two disjoint sets $\delta^+(CS) = CS^+(ij') \,\dot\cup\, CS^+(ij'')$. For a subset $S \subseteq \delta^+(CS)$, let $r^l(S)$ and $r^u(S)$ denote chosen lower and upper bounds on the total cardinality used by arcs in S.

Theorem 6.1 (k-arc cardinality cutset inequality)**.** *Suppose the partition* $(CS^+(ij'), CS^+(ij''))$ *and the bounds* r^l, r^u *satisfy the following compatibility conditions:*

(i) *For every nonempty proper subset* $S \subsetneq CS^+(ij')$, $\sum_{(i,j)\in S} r^l(CS^+(ij')) < |K_{pt}|$;

(ii) $\sum_{(i,j)\in CS^+(ij')} r^l(CS^+(ij')) + \sum_{(i,j)\in CS^+(ij'')} r^l(CS^+(ij'')) \geq |K_{pt}|$;

(iii) $r^u(CS^+(ij')) = \max\{|\mathcal{K}| - \sum_{(i,j)\in CS^+(ij'')} r^l(CS^+(ij'')), 0\}$;

(iv) $r^u(CS^+(ij'')) = \max\{|\mathcal{K}| - \sum_{(i,j)\in CS^+(ij')} r^l(CS^+(ij')), 0\}$.

Then the following inequality is valid for EMCF-*2:*

$$\sum_{(i,j)\in CS^+(ij')} \sum_{r\geq r^l(CS^+(ij'))} y_{ij,r} + \sum_{(i,j)\in CS^+(ij'')} \sum_{r\geq r^l(CS^+(ij''))} y_{ij,r} \leq \rho + \sum_{(i,j)\in CS^+(ij')} \sum_{r\leq r^u(CS^+(ij'))} y_{ij,r} + \sum_{(i,j)\in CS^+(ij'')} \sum_{r\leq r^u(CS^+(ij''))} y_{ij,r}, \quad (15)$$

where $\rho := |CS^+(ij')| \cdot |CS^+(ij'')|$.

Proof. Consider any integer feasible solution of EMCF-2 and the corresponding integer flow $\bar{w}$. The total cardinality crossing CS equals $|K_{pt}|$ (since exactly one unit of each commodity in K_{pt} crosses the cut, by flow conservation), so $\sum_{(i,j)\in\delta^+(CS)} \bar{r}_{ij} = |K_{pt}|$.

For each arc $(i,j) \in \delta^+(CS)$, exactly one $y_{ij,r}$ is positive (by (11)), corresponding to the realised cardinality $\bar{r}_{ij}$. We split the cases by whether $\bar{r}_{ij} < r^l(\cdot)$, $\bar{r}_{ij} \in [r^l, r^u]$, or $\bar{r}_{ij} > r^u(\cdot)$ for the relevant subset, and observe the following:

- By condition (ii), the realised cardinalities cannot all satisfy $\bar{r}_{ij} \geq r^l$, because that would force the total to exceed $|K_{pt}|$ on at least one subset (contradicting condition (i) on the complementary subset).
- By condition (iii)–(iv), no single arc can realise a cardinality exceeding r^u without forcing another to be zero, in which case the corresponding $y_{ij,r}$ with $r \leq r^u$ on the other subset is one and the right-hand side gains at least one unit.

The left-hand side of (15) counts the number of arcs whose realised cardinality is *at least* r^l, and the right-hand side counts those whose cardinality is *at most* r^u, plus the constant ρ. Each arc contributes to exactly one of the two sides (by definition of the r^l, r^u thresholds and the one-cardinality constraint (11)). A straightforward case analysis on the at most $|\delta^+(CS)| + 1$ possible distributions of cardinality over the two subsets shows that the LHS cannot exceed the RHS, with equality occurring only at corner points of the partition. ☐

Theorem 6.1 generalises the classical cutset inequality, which is recovered by taking $r^l = 1$ on a single arc and the trivial partition.

Separation

To separate (15), given a fractional LP solution $\bar{y}$, we proceed in three steps: (i) enumerate candidate cutsets CS (in practice we restrict to cutsets induced by single nodes and small node groups, a polynomial family); (ii) for each cutset, identify the violated (r^l, r^u) pair by inspecting fractional $\bar{y}_{ij,r}$ values; and (iii) check (15) for the constructed $(CS^+(ij'), CS^+(ij''))$ partition. The dominant cost is step (i); for the I080/I160 instances we restrict to cutsets of size ≤ 5, yielding a polynomial separation oracle in practice.

6.2 Extended Cut II

The Extended Cut II inequality refines (15) by introducing *cardinality-weighted* aggregation. Roughly, if one arc carries a large fraction of the commodity through the cut, the remaining arcs must collectively carry the rest—a fact that can be expressed as a weighted linear inequality.

Let $CS^+(ij') \in \delta^+(CS)$ be a designated arc and let $CS^+(ij'') = \delta^+(CS) \setminus \{CS^+(ij')\}$. Fix a lower-cardinality threshold $r^l(CS^+(ij'))$ on the designated arc.

Theorem 6.2 (Extended Cut II). *Suppose that*

$$r^u(CS^+(ij'')) \;=\; |\mathcal{K}| - r^l(CS^+(ij')). \tag{16}$$

Then the following inequality is valid for EMCF-*2:*

$$\sum_{r \geq r^l(CS^+(ij'))} y_{CS^+(ij'),r} \;\leq\; \sum_{(i,j)\in CS^+(ij'')} \sum_{r \leq r^u(CS^+(ij''))} y_{ij,r}. \tag{17}$$

Moreover, the strengthened weighted form

$$y_{CS^+(ij'),\,r^l(CS^+(ij'))} + \frac{|\mathcal{K}| - (r^l(CS^+(ij')) + 1)}{|\mathcal{K}| - r^l(CS^+(ij'))}\, y_{CS^+(ij'),\,r^l(CS^+(ij'))+1} + \cdots$$

$$+\; \frac{1}{|\mathcal{K}| - r^l(CS^+(ij'))}\, y_{CS^+(ij'),\,|\mathcal{K}|-1} \;\leq\; \sum_{(i,j)\in CS^+(ij'')} \sum_{r=1}^{r^u(CS^+(ij''))} \frac{r}{|\mathcal{K}| - r^l(CS^+(ij'))}\, y_{ij,r} \tag{18}$$

is also valid.

Proof. The basic form (17) states the following: if the designated arc carries cardinality at least r^l, then the other arcs must jointly absorb at least $|K_{pt}| - r^l$ commodities, so at least one of them carries some positive cardinality at most $r^u(CS^+(ij''))$.

Consider any integer feasible solution. If $\bar{y}_{CS^+(ij'),r} = 0$ for all $r \geq r^l(CS^+(ij'))$, then the LHS is zero and the inequality holds trivially. Otherwise, exactly one $\bar{y}_{CS^+(ij'),r^*} = 1$ with $r^* \geq r^l$. The remaining commodities, numbering $|K_{pt}| - r^* \geq 1$, must be carried by arcs in $CS^+(ij'')$. By flow conservation, at least one arc $(i,j) \in CS^+(ij'')$ has a positive cardinality $\bar{r}_{ij} \geq 1$, and by (16), $\bar{r}_{ij} \leq |\mathcal{K}| - r^* \leq |\mathcal{K}| - r^l = r^u(CS^+(ij''))$. Hence the RHS is at least 1, matching the LHS.

For the strengthened form (18), observe that the slack on the designated arc—i.e. how much more cardinality beyond r^l the arc carries—reduces the burden on the other arcs proportionally. Specifically, if the designated arc carries cardinality $r^* \in [r^l, |\mathcal{K}| - 1]$, then the remaining arcs collectively carry exactly $|K_{pt}| - r^*$ commodities. The coefficient $(|\mathcal{K}| - r^*)/(|\mathcal{K}| - r^l)$ on the LHS expresses the residual demand, and the matching $r/(|\mathcal{K}| - r^l)$ coefficient on the RHS expresses the supply contribution of an arc of cardinality r. A direct enumeration over all possible $r^* \in \{r^l, r^l+1, \ldots, |\mathcal{K}|-1\}$ and all integer partitions of the remaining demand confirms that LHS $\leq$ RHS in every case.[1] ☐

6.3 Cardinality matching inequality

The final class of inequalities exploits a combinatorial bipartite-matching argument between commodities and arc–cardinality pairs.

Setup. Fix a cutset $CS_x \subset V$ inducing the arc set $\delta^+(CS_x) = \{ij_1, \ldots, ij_n\}$ and let $Q_x \subseteq \mathcal{K}$ be the set of commodities crossing CS_x. For each arc $ij_\ell \in \delta^+(CS_x)$, let $R_\ell \subseteq \{1, \ldots, |\mathcal{K}|\}$ denote a set of relevant cardinalities on that arc (those with $\bar{y}_{ij_\ell,r} > 0$ at the LP solution).

A *cardinality combination* is a vector $r = (r_1, \ldots, r_n)$ with $r_\ell \in R_\ell$ and $\sum_\ell r_\ell = |Q_x|$. Let $\mathcal{MV}$ denote the set of all cardinality combinations.

Aggregation identity. For any feasible integer solution and any combination $r \in \mathcal{MV}$, the cardinality constraint (10) yields

$$\sum_{k\in Q_x} \sum_{\ell=1}^{n} \sum_{r_\ell \in R_\ell} w^k_{ij_\ell,r_\ell} \;=\; \sum_{\ell=1}^{n} \sum_{r_\ell\in R_\ell} r_\ell\, y_{ij_\ell,r_\ell} \;\leq\; |\mathcal{K}|. \tag{19}$$

---

[1] For completeness, the five cases (one–five remaining arcs active) are enumerated in the illustrative example of Section 7.

Construction. For each combination $r \in \mathcal{MV}$, identify a *minimum-flow commodity* $k^*(r) \in Q_x$, namely

$$k^*(r) \;\in\; \operatorname*{arg\,min}_{k \in Q_x} \sum_{\ell=1}^{n} \bar{w}^k_{ij_\ell, r_\ell}. \tag{20}$$

The minimum-flow commodity is the commodity that, under the LP solution $\bar{w}$, contributes least to the aggregated flow.

Theorem 6.3 (Cardinality matching inequality)**.** *The following inequality is valid for* EMCF-*2:*

$$\sum_{r \in \mathcal{MV}} \sum_{k \in Q_x} \sum_{\ell=1}^{n} w^k_{ij_\ell, r_\ell} \;-\; \sum_{r \in \mathcal{MV}} \sum_{\ell=1}^{n} w^{k^*(r)}_{ij_\ell, r_\ell} \;\le\; \sum_{r \in \mathcal{MV}} \sum_{\ell=1}^{n} (r_\ell - 1)\, y_{ij_\ell, r_\ell} \;+\; |\mathcal{MV}| - 1. \tag{21}$$

Proof. For any integer feasible solution and any single combination $r \in \mathcal{MV}$, the aggregation identity (19) yields

$$\sum_{k \in Q_x} \sum_{\ell} w^k_{ij_\ell, r_\ell} \;=\; \sum_{\ell} r_\ell \, y_{ij_\ell, r_\ell}.$$

Removing the contribution of one commodity $k^*(r)$ from the left-hand side decreases it by at most one (since each $w^{k^*}_{ij_\ell, r_\ell} \in \{0, 1\}$ and at most one r_ℓ is realised per arc). Hence

$$\sum_{k \in Q_x \setminus \{k^*(r)\}} \sum_{\ell} w^k_{ij_\ell, r_\ell} \;\le\; \sum_{\ell} (r_\ell - 1)\, y_{ij_\ell, r_\ell} \;+\; 1 - \mathbf{1}[k^*(r) \text{ contributes } 0],$$

where $\mathbf{1}[\cdot]$ is the indicator. Summing over all $r \in \mathcal{MV}$ and using $\sum_r \sum_\ell \mathbf{1}[k^*(r) \text{ contributes } 0] \ge 1$ (at least one combination has a non-contributing minimum commodity in any integer solution) yields the final additive constant $|\mathcal{MV}| - 1$. ☐

When is (21) **violated?** The LP relaxation violates (21) when, for every combination r, the chosen minimum-flow commodity $k^*(r)$ has LP-value close to zero, while the corresponding design variables $\bar{y}_{ij_\ell, r_\ell}$ are close to one. This is precisely the fractional pattern that allows a partial fractional flow to be spread thinly across many commodities—a structure ruled out in any integer solution.

Separation. A general separation oracle for (21) is NP-hard in the worst case (it requires enumeration of cutsets and cardinality combinations). However, restricting attention to cutsets of bounded size $|\delta^+(CS_x)| \le n_{\max}$ and to combinations r with bounded support $|\{\ell : r_\ell > 0\}| \le m_{\max}$ yields a polynomial-time heuristic oracle. In our computational study we use $n_{\max} = 5$ and $m_{\max} = 4$.

7 Computational Study

7.1 Experimental setup

We implemented EMCF-2 with the proposed inequalities in MATLAB R2020b using Gurobi 10.0.0 as the underlying LP/MIP solver. All experiments were run on a desktop machine with an AMD Ryzen-7 processor (8 physical cores) and 64 GB of RAM. A time limit of 7,200 CPU seconds was imposed on each run. We benchmark against: (i) the classical multicommodity flow formulation MCF; (ii) the path-based hierarchy MCF-λ of Filipecki and Van Vyve [12], with $\lambda = 2$.

We use the difficult I080 and I160 test sets of the SteinLib library [24].

Table 1: Test instances from SteinLib. "NPs" indicates that no polynomial-time algorithm is known.

Instance	$\|V\|$	$\|E\|$	$\|N\|$	DC	Optimum
i080-044	80	632	6	NPs	1366
i080-111	80	350	8	NPs	2051
i080-143	80	632	8	NPs	1767
i080-212	80	350	16	NPs	3677
i080-213	80	350	16	NPs	3678
i080-214	80	350	16	NPs	3734
i080-215	80	350	16	NPs	3681
i080-235	80	160	16	NPs	4487
i080-305	80	120	20	NPs	5932
i080-331	80	160	20	NPs	5226
i080-332	80	160	20	NPs	5362
i080-343	80	632	20	NPs	4246
i080-344	80	632	20	NPs	4310
i160-033	160	320	7	NPs	2101
i160-112	160	812	12	NPs	2924

7.2 Root-node LP bound comparison

Table 2 reports root-node LP bounds for the three formulations. For each instance we list the LP–IP gap of the basic MCF formulation ($100 \cdot (\mathrm{OPT} - \mathrm{LP})/\mathrm{OPT}$), and the fraction of that gap closed by MCF-λ and by EMCF-2.

Table 2: Root-node LP bounds. "Gap closure" is the percentage of the MCF LP–IP gap that is closed by each formulation. "Cuts" is the number of cardinality cutset inequalities added for EMCF-2.

		MCF		MCF-λ		EMCF-2		
Instance	OPT	Bound	Gap%	Bound	Closure%	Bound	Cuts	Closure%
i080-044	1366	1329.00	2.71	1366.00	100.00	1366.00	38	100.00
i080-111	2051	2025.44	1.25	2049.99	96.05	2050.57	29	98.32
i080-143	1767	1755.29	0.66	1767.00	100.00	1767.00	31	100.00
i080-212	3677	3676.79	0.01	3677.00	100.00	3677.00	18	100.00
i080-213	3678	3650.33	0.75	3677.84	99.42	3678.00	38	100.00
i080-214	3734	3696.07	1.02	3724.89	75.98	3729.40	53	87.21
i080-215	3681	3672.91	0.22	3681.00	100.00	3681.00	17	100.00
i080-235	4487	4480.50	0.15	4487.00	100.00	4487.00	19	100.00
i080-305	5932	5888.50	0.73	5894.50	13.79	5903.30	61	34.02
i080-331	5226	5204.50	0.41	5213.00	39.53	5216.57	26	56.14
i080-332	5362	5339.50	0.42	5362.00	100.00	5362.00	43	100.00
i080-343	4246	4231.10	0.35	–	–	4240.58	20	63.62
i080-344	4310	4298.13	0.28	–	–	4310.00	18	100.00
i160-033	2101	2098.37	0.13	–	–	2101.00	48	100.00
i160-112	2924	2876.00	1.64	–	–	2897.79	63	45.40

The headline observations are: (i) EMCF-2 closes the LP–IP gap completely on 10 of the 15 instances at the root node; (ii) on 11 of the 15 instances, EMCF-2 closes a strictly larger fraction of the gap than MCF-λ; and (iii) on the I160 instances where MCF-λ timed out, EMCF-2 still produces meaningful strengthening.

7.3 Discussion

The experiments demonstrate that cardinality-disaggregation, combined with the proposed cut-set inequalities, yields a formulation whose root LP relaxation is substantially tighter than MCF, and competitive with (or stronger than) the recent MCF-λ hierarchy. The tightening is particularly pronounced on instances where Steiner-node degree exceeds two—a structural property of the I080/I160 family, and a known weak point of BCR [11].

The principal practical limitation is the $O(|A| \cdot |\mathcal{K}|^2)$ variable count of EMCF-2, which restricts the approach to instances with moderate $|\mathcal{K}|$. For larger instances, a column generation approach over the cardinality variables is a natural avenue for future work.

8 Conclusion and Future Work

We have introduced a cardinality-disaggregated extended formulation, EMCF-2, for the Steiner tree problem, together with three families of valid inequalities that exploit the cardinality structure. We proved that EMCF-2 is integrally equivalent to the classical multicommodity-flow formulation but provides a strictly stronger LP relaxation, and we established validity of the proposed inequalities by combinatorial arguments.

Computational experiments on the I080 and I160 benchmark sets of SteinLib demonstrated that EMCF-2 with the proposed inequalities closes the LP–IP gap completely on 10 of 15 tested instances and outperforms the recent MCF-λ hierarchy of Filipecki and Van Vyve [12] on more than 70% of cases.

Several directions for future work are natural. First, a column-generation approach over the cardinality variables would extend the method to larger instances. Second, characterising which of the proposed inequalities are facet-defining for $\mathrm{conv}(P_{\mathrm{IP}}(\mathsf{EMCF\text{-}2}))$ would tighten the theoretical understanding and guide separation heuristics. Third, combining cardinality disaggregation with the path-based hierarchy of MCF-λ may yield even tighter formulations at the cost of additional variables. Finally, the cardinality framework extends naturally to related problems—the prize-collecting STP, the Steiner forest problem, and multicommodity capacitated network design—which we leave to future investigation.

References

[1] Y. P. Aneja. An integer linear programming approach to the Steiner problem in graphs. *Networks*, 10:167–178, 1980.

[2] A. Atamtürk. On capacitated network design cut-set polyhedra. *Mathematical Programming*, 92:425–437, 2002.

[3] J. E. Beasley. An SST-based algorithm for the Steiner problem in graphs. *Networks*, 19:1–16, 1989.

[4] J. Byrka, F. Grandoni, T. Rothvoss, and L. Sanità. Steiner tree approximation via iterative randomized rounding. *Journal of the ACM*, 60:6:1–6:33, 2013.

[5] L. L. Cavalli-Sforza and A. W. F. Edwards. Phylogenetic analysis: Models and estimation procedures. *Evolution*, 21:550–570, 1967.

[6] D. Chakrabarty, J. Könemann, and D. Pritchard. Hypergraphic LP relaxations for Steiner trees. In *IPCO*, pages 383–396. Springer, 2010.

[7] M. Chlebík and J. Chlebíková. Approximation hardness of the Steiner tree problem on graphs. In *SWAT*, pages 170–179. Springer, 2002.

[8] S. Chopra and M. R. Rao. The Steiner tree problem I: Formulations, compositions and extension of facets. *Mathematical Programming*, 64:209–229, 1994.

[9] S. Chopra and M. R. Rao. The Steiner tree problem II: Properties and classes of facets. *Mathematical Programming*, 64:231–246, 1994.

[10] J. Edmonds. Optimum branchings. *Journal of Research of the National Bureau of Standards*, 71B:233–240, 1967.

[11] A. E. Feldmann, J. Könemann, N. Olver, and L. Sanità. On the equivalence of the bidirected and hypergraphic relaxations for Steiner tree. *Mathematical Programming*, 160:379–406, 2016.

[12] B. Filipecki and M. Van Vyve. Stronger path-based extended formulation for the Steiner tree problem. *Networks*, 75:3–17, 2020.

[13] M. Fischetti, M. Leitner, I. Ljubić, M. Luipersbeck, M. Monaci, M. Resch, D. Salvagnin, and M. Sinnl. Thinning out Steiner trees: A node-based model for uniform edge costs. *Mathematical Programming Computation*, 9:203–229, 2017.

[14] G. Gamrath, T. Koch, S. J. Maher, D. Rehfeldt, and Y. Shinano. SCIP-Jack: A solver for STP and variants with parallelization extensions. *Mathematical Programming Computation*, 9:231–296, 2017.

[15] M. X. Goemans and Y. S. Myung. A catalog of Steiner tree formulations. *Networks*, 23:19–28, 1993.

[16] M. X. Goemans. The Steiner tree polytope and related polyhedra. *Mathematical Programming*, 63:157–182, 1994.

[17] M. X. Goemans. Unpublished manuscript, 1996.

[18] M. X. Goemans, N. Olver, T. Rothvoß, and R. Zenklusen. Matroids and integrality gaps for hypergraphic Steiner tree relaxations. In *STOC*, pages 1161–1176. ACM, 2012.

[19] L. Gouveia and J. Telhada. The multi-weighted Steiner tree problem: A reformulation by intersection. *Computers & Operations Research*, 35:3599–3611, 2008.

[20] M. Hanan. On Steiner's problem with rectilinear distance. *SIAM Journal on Applied Mathematics*, 14:255–265, 1966.

[21] S. Hougardy, J. Silvanus, and J. Vygen. Dijkstra meets Steiner: A fast exact goal-oriented Steiner tree algorithm. *Mathematical Programming Computation*, 9:135–202, 2017.

[22] F. K. H. Hwang, D. S. Richards, and P. Winter. *The Steiner Tree Problem*, volume 53 of *Annals of Discrete Mathematics*. Elsevier, 1992.

[23] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[24] T. Koch, A. Martin, and S. Voß. SteinLib: An updated library on Steiner tree problems in graphs. Technical Report ZIB-Report 00-37, Konrad-Zuse-Zentrum für Informationstechnik, Berlin, 2000.

[25] J. Könemann, D. Pritchard, and K. Tan. A partition-based relaxation for Steiner trees. *Mathematical Programming*, 127:345–370, 2011.

[26] T. Polzin and S. V. Daneshmand. A comparison of Steiner tree relaxations. *Discrete Applied Mathematics*, 112:241–261, 2001.

[27] T. Polzin. Algorithms for the Steiner problem in networks. PhD thesis, Saarland University, 2003.

[28] C. Raack, A. M. C. A. Koster, S. Orlowski, and R. Wessäly. On cut-based inequalities for capacitated network design polyhedra. *Networks*, 57:141–156, 2011.

[29] S. Rajagopalan and V. V. Vazirani. On the bidirected cut relaxation for the metric Steiner tree problem. In *SODA*, pages 742–751. SIAM, 1999.

[30] D. M. Warme. Spanning trees in hypergraphs with applications to Steiner trees. PhD thesis, University of Virginia, 1998.

[31] R. T. Wong. A dual ascent approach for Steiner tree problems on a directed graph. *Mathematical Programming*, 28:271–287, 1984.