A $\frac{21}{16}$-Approximation for the Minimum 3-Path Partition Problem

Yong Chen  
Department of Mathematics, Hangzhou Dianzi University, Hangzhou, Zhejiang, China  
chenyong@hdu.edu.cn

Randy Goebel  
Department of Computing Science, University of Alberta, Edmonton, Alberta T6G 2E8, Canada  
grgoebel@ualberta.ca

Bing Su  
School of Economics and Management, Xi’an Technological University, Xi’an, Shaanxi, China  
su_bing684@sohu.com

Weitian Tong\footnote{Corresponding author}  
Department of Computer Science, Eastern Michigan University, Ypsilanti, Michigan 48197, USA  
wttong.research@gmail.com

Yao Xu  
Department of Computer Science, Kettering University, Flint, Michigan 48504, USA  
skyaoxu@outlook.com

An Zhang  
Department of Mathematics, Hangzhou Dianzi University, Hangzhou, Zhejiang, China  
anzhang@hdu.edu.cn

---

Abstract

The minimum $k$-path partition (Min-$k$-PP for short) problem targets to partition an input graph into the smallest number of paths, each of which has order at most $k$. We focus on the special case when $k = 3$. Existing literature mainly concentrates on the exact algorithms for special graphs, such as trees. Because of the challenge of NP-hardness on general graphs, the approximability of the Min-3-PP problem attracts researchers’ attention. The first approximation algorithm dates back about 10 years and achieves an approximation ratio of $\frac{3}{2}$, which was recently improved to $\frac{13}{9}$ and further to $\frac{4}{3}$. We investigate the $\frac{3}{2}$-approximation algorithm for the Min-3-PP problem and discover several interesting structural properties. Instead of studying the unweighted Min-3-PP problem directly, we design a novel weight schema for $\ell$-paths, $\ell \in \{1, 2, 3\}$, and investigate the weighted version. A greedy local search algorithm is proposed to generate a heavy path partition. We show the achieved path partition has the least 1-paths, which is also the key ingredient for the algorithms with ratios $\frac{13}{9}$ and $\frac{4}{3}$. When switching back to the unweighted objective function, we prove the approximation ratio $\frac{21}{16}$ via amortized analysis.

2012 ACM Subject Classification  
Mathematics of computing → Graph algorithms; Mathematics of computing → Approximation algorithms; Theory of computation → Graph algorithms analysis; Theory of computation → Approximation algorithms analysis

Keywords and phrases  
3-path partition, exact set cover, approximation algorithm, local search, amortized analysis

Digital Object Identifier  
10.4230/LIPIcs.ISAAC.2019.46

Funding  
Yong Chen: NSFC Grants 11971139 and 11571252; CSC Grant 201508330054  
Randy Goebel: NSERC Canada  
An Zhang: NSFC Grants 11771114 and 11571252; CSC Grant 201908330090

© Yong Chen, Randy Goebel, Bing Su, Weitian Tong, Yao Xu, and An Zhang; licensed under Creative Commons License CC-BY

Editors: Pinyan Lu and Guochuan Zhang; Article No. 46; pp. 46:1–46:20  
Leibniz International Proceedings in Informatics  
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
1 Introduction

In the minimum $k$-path partition (abbreviated as Min-$k$-PP) problem, a simple graph $G = (V, E)$ is partitioned into the smallest number of paths such that each path has at most $k$ vertices for a given positive integer $k$. We can observe that when $k = 2$ the Min-$k$-PP problem is closely related to the maximum matching problem in an unweighted graph and that when $k = n$ the Min-$k$-PP problem has an optimal value 1 if and only if the input graph contains a Hamiltonian path. Besides, the Min-$k$-PP problem can be treated as a special case of the minimum $k$-set exact cover problem by associating each path (of order at most $k$) with a set of size at most $k$. The decision version of the minimum exact cover problem is one of Karp’s 21 NP-complete problems [5]. To the best of our knowledge, there are no non-trivial approximation algorithms for the minimum $k$-set exact cover problem. The Min-$k$-PP problem also has a close relation to the classic minimum $k$-set cover problem, which does not require the mutual disjointness for the resultant cover. Though the Min-$k$-PP and minimum $k$-set cover problems share some similarities, none contains the other as a special case. A detailed discussion of the relationship between them can be found in [1].

It is not hard to see the Min-$k$-PP problem is NP-hard on general graphs [4]. It remains intractable on cographs [8] and chordal bipartite graphs [9] when $k$ is an input. Moreover, the Min-$k$-PP problem remains to be NP-hard in comparability graphs even for $k = 3$ [9]. Recently, Korpelainen [6] further investigated and depicted the NP-hardness of the Min-$k$-PP problem in some special graph classes.

On the positive side, the Min-$k$-PP problem is polynomial-time solvable in several special cases. Motivated by the application in network broadcasting, which finds the minimum number of message originators necessary to broadcast a message to all vertices in a tree network in one or two time units, Yan et al. [10] presented a linear-time algorithm for the Min-$k$-PP problem on trees. A polynomial-time algorithm on cographs when $k$ is fixed was designed by Steiner [8], who later proposed a polynomial-time solution for the Min-$k$-PP problem, with any $k$, on bipartite permutation graphs [9].

Monnot and Toulouse [7] are the pioneer to investigate the approximability for the Min-$k$-PP problem. In particular, they studied the special case, Min-3-PP, and designed a neat $3/2$-approximation algorithm with a running time $O(nm + n^2 \log n)$ for general graphs, where $n$ and $m$ are the numbers of vertices and edges in the graph. Recently, Chen et al. [2] presented an improved approximation algorithm with a ratio $13/9$ by first computing a $k$-path partition with the least 1-paths for any $k \geq 3$ and then greedily merging three 2-paths into two 3-paths whenever possible. Their greedy algorithm takes $O(nm)$ and $O(n^3)$ time respectively in these two steps. Based on the first step of the $13/9$-approximation algorithm, Chen et al. [1] designed a novel local search scheme to improve the approximation ratio to $4/3$. Specifically, their local search algorithm repeatedly searches for an expected collection of 2- and 3-paths and replaces it by a strictly smaller replacement collection of new 2- and 3-paths. It is worth noting that the ratio $4/3$ matches the best approximation ratio for the minimum 3-set cover problem [3]. Due to the similarity between the Min-3-PP problem and the minimum 3-set cover problem, it seems difficult to improve this ratio a step further and Chen et al. [1] left an open question for a better approximation algorithm for the Min-3-PP problem.

Our paper addresses this open question proposed by Chen et al. [1]. Our main contributions are as follows. 1) We propose a novel weight function for $\ell$-paths, $\ell \in \{1, 2, 3\}$, which forces any heavy path partition prefer specific combinations of $\ell$-paths. In particular, the number of 1-paths in a heavy path partition cannot be too large. 2) We design a
greedy local search strategy (named as Greedy) to generate a path partition that contains 2-paths “sparsely” compared with the 3-paths. Moreover, we are able to show there exits an optimal solution whose number of 1-paths is upper bounded by our greedy solution. 3) We design another well-designed wide-range tree-like search strategy (named as TreeSearch) to further reduce the number of 1-paths. More specifically, the resultant path partition contains the least amount of 1-paths. 4) We conduct an delicate amortized analysis to show the “sparseness” of 2-paths quantitatively, leading to a better approximation ratio \( \frac{21}{16} \) (= 1.3125 < 1.3333 ≈ \( \frac{4}{3} \)) for the Min-3-PP problem.

In the following, Section 2 introduces basic concepts and notations; Section 3 restates the classic \( \frac{4}{3} \)-algorithm by Monnot and Toulouse [7]; Section 4 shows our \( \frac{21}{16} \)-approximation algorithm based on non-trivial discoveries of the structural properties of the Min-3-PP problem; Section 5 concludes the paper and proposes open questions.

2 Preliminaries

We begin with definitions and notations which hold throughout this paper. Let \( G = (V,E) \) be a simple undirected graph, defined by a set of vertices \( V = \{v_1, v_2, \ldots, v_n\} \) and a set of undirected edges \( E = \{e_1, e_2, \ldots, e_m\} \), where each edge \( e = \{u,v\} \) connects two vertices \( u,v \in V \). Let \( U \subset V \) be any subset of vertices of \( G \). Then the (vertex) induced subgraph \( G[U] \) is the subgraph whose vertex set is \( U \) and whose edge set consists of edges in \( E \) with both endpoints in \( U \). For any subgraph \( S \) of \( G \), let \( V_G(S) \) and \( E_G(S) \) denote the vertex set and edge set of \( S \), respectively. The order of \( S \) is defined as the cardinality of \( V_G(S) \). For each vertex \( v \in V \), define its neighbor set as \( N_G(v) = \{u \mid \{u,v\} \in E(G)\} \) and its degree as \( d_G(v) = |N_G(v)| \). If the underlying graph \( G \) is clear, we may omit the subscript \( G \) in all notations for the sake of simplicity. In sequel, we use \( \cup \) and \( \uplus \) to denote the set union and multiset union respectively. For any (even multigraph) \( S \) with \( V(S) \subseteq V_G \), we abuse the notation \( G[S] \) to denote the induced graph \( (V_G(S), E(S)) \).

A path \( P \) in \( G \) is a sequence of distinct vertices \( \langle v_1, v_2, \ldots, v_\ell \rangle \), \( \ell \geq 1 \), such that \( \{v_i, v_{i+1}\} \in E \), for \( i = 1, 2, \ldots, \ell - 1 \). We say a path is an \( \ell \)-path if its order is \( \ell \), i.e., \( |V(P)| = \ell \). A path partition of \( G \) is a collection of vertex disjoint paths \( \mathcal{P} \) such that \( V(G) = \bigcup_{P \in \mathcal{P}} V(P) \). A \( k \)-path partition is a path partition \( \mathcal{P} \) with each path having at most \( k \) vertices for a given positive integer \( k \), and the minimum \( k \)-path partition problem aims at minimizing the cardinality of \( \mathcal{P} \). Our paper considers the special case when \( k = 3 \). In the following context, a path partition is a 3-path partition by default.

Consider an optimal path partition \( \mathcal{P}^* = \{P_1^*, P_2^*, P_3^*\} \), where \( P_\ell^*, \ell \in \{1,2,3\} \) denotes the set of \( \ell \)-paths. Suppose \( \mathcal{P} = \{P_1, P_2, P_3\} \) is any feasible path partition on \( G \). Let \( \text{OPT} = |\mathcal{P}^*| \) and \( \text{SOL} = |\mathcal{P}| \). Denote the cardinality of \( P_\ell^* \) (\( P_\ell \), respectively) as \( p_\ell^* \) (\( p_\ell \), respectively), \( \ell \in \{1,2,3\} \). Then we have

\[
\begin{align*}
n &= p_1^* + 2p_2^* + 3p_3^* = p_1 + 2p_2 + 3p_3, \\
\text{OPT} &= p_1^* + p_2^* + p_3^* \\
\text{SOL} &= p_1 + p_2 + p_3,
\end{align*}
\]

which implies

\[
\text{SOL} \geq \text{OPT} \geq \frac{n}{3}.
\]

Let \( \mathcal{Q}_\ell, \ell \in \{1,2,3\} \) denote the collection of all possible \( \ell \)-paths in the given graph \( G \) and define \( \mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2 \cup \mathcal{Q}_3 \). We introduce two important concept: conflict graph and intersection graph.
Definition 1 (Conflict graph). In a conflict graph, denoted by CG, each vertex represents an element of Q, i.e., some ℓ-path in G, ℓ ∈ {1, 2, 3}; and each edge {P, Q} with P, Q ∈ Q exists only if V_G(P) ∩ V_G(Q) ≠ ∅. In addition, we add i parallel edges between two vertices in CG if and only if two corresponding paths in G have i vertices in common.

Definition 2 (Intersection graph). Considering the optimal partition P* and any feasible partition P, the intersection graph induced by P and P*, denoted by IG, is a bipartite graph, where each vertex represents an element of P and P*, and two vertices in IG are adjacent if and only if the vertex sets of the corresponding paths in G intersect. Similar to the conflict graph, parallel edges are allowed in IG.

According to the above definitions, parallel edges are allowed in both CG and IG. Next, we partition the edges of IG into sets Eij, i, j ∈ {1, 2, 3} such that Eij = {P, P*} | P ∩ P* ≠ ∅, P ∈ P, P* ∈ P*. Let m_{ij} denote the cardinality of Eij. We have

\[ p_i = \frac{1}{ℓ} \sum_{j=1}^{3} m_{ij}, \quad p_j^* = \frac{1}{ℓ} \sum_{j=1}^{3} m_{ij}, \quad n = \sum_{i,j} m_{ij}. \]  

Then the relation between any feasible solution and the optimal solution can be represented as follows.

\[
\text{SOL} = p_1 + p_2 + p_3 = \sum_{i=1}^{3} \left( \frac{1}{ℓ} \sum_{j=1}^{3} m_{ij} \right) \\
= (m_{11}/1 + m_{12}/1 + m_{13}/1) + (m_{21}/2 + m_{22}/2 + m_{23}/2) + (m_{31}/3 + m_{32}/3 + m_{33}/3) \\
= (m_{31}/3 + m_{32}/2 + m_{33}/1) + (m_{31}/3 + m_{32}/2 + m_{12}/1) + (m_{31}/3 + m_{23}/2 + m_{13}/1) \\
= \left( \sum_{i=1}^{3} m_{11} - \frac{2}{3} m_{31} - \frac{1}{2} m_{21} \right) + \left( \frac{1}{2} \sum_{i=1}^{3} m_{i2} - \frac{1}{6} m_{32} + \frac{1}{2} m_{12} \right) + \left( \frac{1}{3} \sum_{i=1}^{3} m_{i3} + \frac{1}{6} m_{23} + \frac{2}{3} m_{33} \right) \\
= p_1^* + p_2^* + p_3^* + \left( \frac{2}{3} m_{13} + \frac{1}{2} m_{12} - \frac{1}{6} m_{23} - \frac{2}{3} m_{31} - \frac{1}{2} m_{21} - \frac{1}{6} m_{32} \right) \\
= \text{OPT} + \left( \frac{2}{3} m_{13} + \frac{1}{2} m_{12} - \frac{1}{6} m_{23} - \frac{2}{3} m_{31} - \frac{1}{2} m_{21} - \frac{1}{6} m_{32} \right). 
\]

3 A brief review of the classic 3/2-approximation algorithm

For the algorithm designed by Monnot and Toulouse [7], the main idea is to first compute a maximum matching M_1 in the input graph G and then find another maximum matching M_2 to connect M_1 with the vertices left from the calculation of M_1. Since each time an edge is added to the solution, the number of connected components decreases by 1 and therefore SOL = n − |M_1| − |M_2|. Let’s consider the vertex set left over after the first maximum matching M_1* is found, i.e., V \ V(M_1*). For any vertex \( v \in (V \setminus V(M_1^*)) \setminus \bigcup_{P \in P^*} V(P) \), \( v \) must be contained in some ℓ-path with ℓ ≥ 2 in the optimal partition, which implies \( v \) is adjacent to some edge in M_1*. Each vertex in V(M_1*) can be adjacent to at most two such vertices and from the maximality each edge \{u, v\} in M_1* can be adjacent to at most two such vertices. Also from the maximality, we have \|M_2\| ≥ \frac{1}{4}(n − 2|M_1| − p_1^*), which implies SOL ≤ \frac{1}{2}(n + p_1^*). Therefore, OPT = p_1^* + p_2^* + p_3^* = \frac{3}{4}(n + p_1^* + 2p_1^*) ≥ \frac{3}{4}(n + p_1^*) ≥ \frac{3}{4}\text{SOL}, which shows an approximation ratio of 3/2.
4 The 21/16-Approximation Algorithm

Based on the observation and analysis of the equation (6), our main idea is to find a partition $P$ to minimize $m_{13}$, $m_{12}$ and $m_{23}$. Note that the analysis of our algorithm only depends on the existence of an optimal partition and we do not need to find an optimal partition.

Due to the definitions of the conflict graph and intersection graph, we may abuse an $\ell$-path to denote a vertex in $CG$ and $IG$, vice versa. Recall that $Q$ is the collection of all $\ell$-paths, $\ell \in \{1, 2, 3\}$, in $G$. We define a weight function $w(\cdot)$ mapping each path to a real value. Specifically, let

$$w(P) = \begin{cases} 1, & \text{if } P \in Q_1; \\ 5, & \text{if } P \in Q_2; \\ 8, & \text{if } P \in Q_3. \end{cases} \quad (7)$$

For any subset $Q' \subset Q$, let $w(Q')$ denote the total weight of paths in $Q'$.

▶ Definition 3 (Improving set). For any $I \subset Q\setminus P$, $N_{CG}(I, P) = N_{CG}(I) \cap P$ is the set of neighbors of $I$ restricted in $P$. $I$ is an improving set if $w(I) > w(N_{CG}(I, P))$ and $(P \setminus N_{CG}(I, P)) \cup I$ is a partition of $G$.

Our algorithm, named as Greedy-TreeSearch, is a local search algorithm, which invokes two local search strategies, Greedy and TreeSearch, as subroutines iteratively until the total weight of the partition cannot be increased.

The intuition behind the design of our weight function is that we prefer partitioning a 4-path into two 2-paths over partitioning into one 1-path and one 3-path and partitioning a 6-path into two 3-paths over three 2-paths. When analyzing the algorithm Greedy-TreeSearch, an optimal solution is compared. Without loss of generality, we choose the optimal path partition $P^*$ with the maximum weight with respect to our weight function.

Considering two vertex disjoint paths $X, Y \in P$, we say $X$ and $Y$ are friends if there is an edge $\{u, v\}$ incident to both $X$ and $Y$ in the original graph $G$. Refer to Figure 1 for more details. If $u$ ($v$, respectively) is the ending vertex of $X$ ($Y$, respectively), connecting $X$ and $Y$ in $G$ via $\{u, v\}$ forms a path in $G$ and we say $X$ and $Y$ are close friends with respect to $\{u, v\}$. We also say $X$ is a close friend to $Y$ via $\{u, v\}$ or via $u$ or via $v$. Otherwise, assuming $u$ is the middle vertex of $X$, we say $X$ is an ordinary friend to $Y$ via $\{u, v\}$. Or we just say $X$ and $Y$ are ordinary friends via $\{u, v\}$. According to the definitions of different types of friends, we have the following observation.

![Figure 1](image-url) An illustration of friends. The solid and dashed edges denote the edges in $E(P)$ and $E(G) \setminus E(P)$, respectively. The paths in $P$ are indicated in ellipses. The $\{(z_1, z_2), (y_1, y_2, y_3)\}$ and $\{(y_1, y_2, y_3), (x_1, x_2, x_3)\}$ are two pairs of friends. The first pair are close friends while the second pair are ordinary friends.

▶ Observation 4. For any two paths $X$ and $Y$ in a path partition $P$, $X$ and $Y$ are ordinary friends only if at least one of them is a 3-path and the friendship is built through the middle vertex of the 3-path.
4.1 The Greedy Algorithm

The algorithm GREEDY starts with an arbitrary path partition for $G$ and then increases the partition’s weight by iteratively updating the current partition with an improving set of size at most 6 until there are no improving sets. The value 6 is determined in the proof for Lemma 14, where at most 6 paths are involved for the detection of an improving set. We abuse the notation $P$ to denote the resultant partition when GREEDY terminates.

Lemma 5. For any two friends $X$ and $Y$ in $P$, if $X$ is a 1-path, $Y$ can only be a 3-path and the friendship is built through the middle vertex on $Y$.

Proof. Recall that the weight of a 1-, 2-, 3-path is 1, 5, 8, respectively. We prove the lemma by contradiction. If $Y$ is an $\ell$-path with $\ell \leq 2$, connecting $X$ and $Y$ forms an $(\ell + 1)$-path, which is an improving set. Refer to the left two subfigures of Figure 2. If $Y$ is a 3-path and $Y$ is a close friend of $X$, connecting $X$ and $Y$ produces a 4-path, which can be partitioned into two 2-paths $\langle x_1, y_1 \rangle$ and $\langle y_2, y_3 \rangle$. It is easy to check these two paths form an improving set. Refer to the third subfigure of Figure 2. The GREEDY algorithm will not terminate if there is an improving set. This implies the correctness of the lemma.

Figure 2 Four subgraphs of $G$ show the different cases of the friendship involving a 1-path. The solid and dashed edges denote the edges in $E(P)$ and $E(G) \setminus E(P)$, respectively. The paths in $P$ are indicated in ellipses.

Figure 3 The left and right subgraphs show the friendship of a 2-path and 3-path, respectively. The solid and dashed edges denote the edges in $E(P)$ and $E(G) \setminus E(P)$, respectively. The paths in $P$ are indicated in ellipses.

Lemma 6. Consider an $\ell$-path $X = \langle x_1, \ldots, x_\ell \rangle$ in $P$. Let $F_i$ denote the set of 2-path friends in $P$ via $x_i$, $i \in \{1, \ldots, \ell\}$ in the multigraph induced by the union of paths in $P$ and $P^*$, that is, $G[E(P^*) \cup E(P)]$.

- In $G[E(P^*) \cup E(P)]$, $X$ has at most $2\ell$ distinct friends in $P$.
- There are at most $\ell - 1$ distinct 2-paths $Y_i$ such that $Y_i \in F_i$.

Proof. By Lemma 5, we only need to consider $\ell \in \{2, 3\}$. Since each path in $P^*$ has an order at most 3, $X$ can be adjacent to at most 2 other paths in $P$ via one vertex and therefore $X$ has at most $2\ell$ distinct friends in $P$. Suppose there are $\ell$ distinct 2-paths $Y_i$ such that $Y_i \in F_i$. Assume $Y_i = \langle y_{i1}, y_{i2} \rangle$ are the 2-path friends via the ending vertex $x_i$. Refer to Figure 3. Then the $\ell$ 3-paths $\langle x_i, y_{i1}, y_{i2} \rangle$, $i \leq \ell$, form an improving set, which indicates a contradiction.
Recall that the chosen optimal path partition $\mathcal{P}^*$ has the maximum weight with respect to our weight function. In addition, we require that $|E_G(\mathcal{P}^*) \cap E_G(\mathcal{P})|$ is maximized over all the heaviest optimal path partitions. In other words, we consider the heaviest optimal path partition that overlaps our solution as many edges as possible. Then we have the following lemma.

Lemma 7. Let $\mathcal{P}^*$ be the heaviest optimal path partition that maximizes $|E_G(\mathcal{P}^*) \cap E_G(\mathcal{P})|$. We have $\mathcal{P}_1^* \subseteq \mathcal{P}_1$ and thus $p_1^* \leq p_1$.

Proof. Suppose $X = \{x\} \in \mathcal{P}_1^* \setminus \mathcal{P}_1$. $x$ must be on some $\ell$-path $P$ in $\mathcal{P}_2 \cup \mathcal{P}_3$. Since $\mathcal{P}^*$ has the maximum weight, any 1-path in $\mathcal{P}^*$ can only have an ordinary friend in $\mathcal{P}^*$ by a similar argument in the proof for Lemma 5. Without loss of generality, we assume $y$ is a neighbor of $x$ on $P$. Then $y$ has to be the middle vertex of a 3-path in $\mathcal{P}^*$, denoted by $Y = \langle y_1, y, y_3 \rangle$. Since $\{x, y\} \in E(\mathcal{P})$, at least one of $\{y, y_1\}$ and $\{y, y_3\}$ is not in $E(\mathcal{P})$. Assume $\{y, y_1\} \notin E(\mathcal{P})$, we modify the paths $X$ and $Y$ to $\{x, y, y_3\}$ and $\{y_1\}$. This modification does not change the weight of $\mathcal{P}^*$ and $|E_G(\mathcal{P}^*) \cap E_G(\mathcal{P})|$ is increased by 1, which contradicts to the maximality. ▶

4.2 The TreeSearch Algorithm

The algorithm TreeSearch aims at reducing the number of 1-paths in the partition $\mathcal{P}$ returned by the Greedy algorithm. Though the TreeSearch algorithm also targets to find an improving set, the size of an improving set may be fairly large.

Fix any 1-path $\langle r \rangle$ in $\mathcal{P}$. We modify the depth first search (DFS for short) to explore $G$ with the root $r$. Recall that the DFS grows a tree node by node as deep as possible to expand the whole connected graph. We modify the DFS as follows: if the currently exploring vertex $u$ is connected with its parent via an edge in $E(G \setminus E(\mathcal{P}))$ in the current tree, instead of expanding the vertex as the normal DFS, we expand the vertex simply with the $\ell$-path in $\mathcal{P}$ containing $u$, and then the normal DFS is applied on the ending vertex (or vertices) of this $\ell$-path. Abbreviate the modified DFS as MDFS. Note that the MDFS produces a forest instead of a spanning tree as the MDFS does not fully expand every node in the graph.

We say an $\ell$-path is involved in a tree path $P_T$ starting at $r$, if the intersection of $P_T$ and this $\ell$-path is not empty. In Figure 4, the tree path $\langle r, x_{22}, x_{23}, x_{32}, x_{42}, x_{43} \rangle$ involves paths $X_1$, $X_2$, $X_3$, $X_4$. Once the MDFS finds an improving set for the $\ell$-paths involved along a path starting at $r$, the TreeSearch algorithm refines the current partition. Our TreeSearch algorithm invokes the MDFS on the 1-paths iteratively until no more improving sets can be identified.

Let’s consider the multigraph induced by the union of paths in $\mathcal{P}$ and $\mathcal{P}^*$, i.e., $G[E(\mathcal{P}^*) \cup E(\mathcal{P})]$. Let $F$ be the resultant forest returned by the TreeSearch algorithm on this multigraph. For any 3-path in $\mathcal{P}$ involved in $F$, its endpoints are expanded by the MDFS but its middle vertex is ignored during the MDFS search. It is possible that this middle vertex is connected to at most two friends in $\mathcal{P}$. In particular, if the friends is a 1-path, we call it as a free 1-path, which forms a tree of size 1 in the forest $F$. Refer to Figure 4 for an example of a free 1-path.

Lemma 8. Consider any tree $T$ in the forest $F$. For any path connecting the root $r$ and a leaf in $T$, suppose the $\ell$-paths in $\mathcal{P}$ involved in the path are $\{X_1, X_2, \ldots, X_t\}$ in order with $X_1 = \{r\}$. We claim that

- $X_2, \ldots, X_t$ are all 3-paths and moreover $X_i$ and $X_{i+1}$ are ordinary friends, $i \in \{1, \ldots, t - 1\}$;
- any two involved 3-paths say, $X$ and $Y$, cannot be connected via the ending vertices in $G[E(\mathcal{P}^*) \cup E(\mathcal{P})]$;
- the two ending vertices of $X_i$ are not connected in $G[E(\mathcal{P}^*) \cup E(\mathcal{P})]$.
Proof. We prove the first claim in the lemma by contradiction. Assume $X_j, j \geq 2$ is the first \ell-path such that \ell < 3 or it is a close friend to $X_{j-1}$. Suppose $X_i = (x_{i,1}, x_{i,2}, x_{i,3}), i \in \{2, \ldots, j-1\}$. If $j = 2$, the proof reduces to the proof of Lemma 5. We assume $j \geq 3$ without loss of generality.

- $X_j$ is an \ell-path with \ell \leq 2. The 3-paths $\langle r, x_{2,2}, x_{2,1} \rangle, \langle x_{2,3}, x_{3,2}, x_{3,1} \rangle, \ldots, \langle x_{j-2,3}, x_{j-1,2}, x_{j-1,1} \rangle$ together with the (\ell + 1)-path connecting $x_{j-1,3}$ and $X_j$ form an improving set.

- $X_j$ is a 3-path and $X_j$ is a close friend of $X_{j-1}$. Connecting $X_j$ and $X_{j-1,3}$ produces a 4-path, which can be partitioned into two 2-paths. Together with the 3-paths $\langle r, x_{2,2}, x_{2,1} \rangle, \langle x_{2,3}, x_{3,2}, x_{3,1} \rangle, \ldots, \langle x_{j-2,3}, x_{j-1,2}, x_{j-1,1} \rangle$, we find an improving set.

The TreeSearch algorithm will not terminate if there is an improving set, which implies the correctness of the claim.

By the first claim, each tree in the forest produced by the TreeSearch algorithm on $G[E(P^*) \cup E(P)]$ can only involve 3-paths if its root is a 1-path in $P$. Suppose there are two involved 3-paths $X$ and $Y$ such that they are connected via the ending vertices. Consider $X$ if 1) $X$ and $Y$ are involved in the same root-to-leaf path and $X$ is the ancestor; 2) $X$ and $Y$ are involved in different root-to-leaf paths. An improving set can be found similarly following the argument for the first claim.

The third claim states the edge $\{x_{i,3}, x_{i,1}\}$ does not exist in $G[E(P^*) \cup E(P)]$. Otherwise, we can find a 4-path $\langle x_{i-1,3}, x_{i,2}, x_{i,3}, x_{i,1} \rangle$ if $i \geq 3$ or $\langle r, x_{i,2}, x_{i,3}, x_{i,1} \rangle$ if $i = 2$. Then an improving set can be identified similar to the proof for the first claim. \hfill ◼

Let’s consider any tree $T$ in the forest $F$ returned by the TreeSearch algorithm. If we contract the edges in $P^*$, $T$ become a tree where each internal node has a degree 2 or 4. Denote the contracted tree as $\tilde{T}$. Refer to Figure 4 for details.

Lemma 9. $m_{13} + m_{12} \leq m_{31}$.

Proof. Since the $P^*$ contains paths of order at most 3, each edge in $E_G(T) \cap E_G(P^*)$ can be connected at most once to some leaf in $T$. In the left subfigue in Figure 4, the edge $\{x_{23}, x_{32}\}$ is connected with the leaf $x_{43}$ in a 3-path in $P^*$. Each leaf in $T$ is also a vertex in $G$ thus must be included in some \ell-path in $P^*$. According to Lemma 8, different leaves can only be connected to different edges in $E(T) \cap E(P^*)$, i.e., different internal nodes in $\tilde{T}$. If a leaf is not connected to the other vertex, it must be a 1-path in $P^*$, which contributes 1 to the value of $m_{31}$. Suppose the number of leaves is $L$. Also assume the number of leaves that are connected to edges in $E(T) \cap E(P^*)$ is $L_1$. Denote the number of remaining leaves as $L_2 = L - L_1$.

In the contracted tree $\tilde{T}$, each internal node representing edge(s) in $E(P^*)$ has at least 2 children. Thus, the number of leaves in $\tilde{T}$ is at least 1 plus the number of the internal vertices. Since the $P^*$ contains paths of order at most 3, each internal node in $\tilde{T}$ can be
connected to at most one leaf or one free path. Assume without loss of generality that the number of these two types of internal nodes are $I_1$ and $I_2$ respectively. Each node of the second type contributes 1 to $m_{13}$ if the free path is a 1-path. Denote the number of internal nodes to be $I$. Then $I_1 + I_2 \leq I$.

According to the definitions of $I_1$, $I_2$, $L_1$, $L_2$, and $L$, we have

$$I_1 + I_2 + 1 \leq I + 1 \leq L = L_1 + L_2; L_1 = I_1; m_{13} + m_{12} = I_2 + 1; m_{31} = L_2; m_{13} \leq I_2 + 1; m_{12} \leq 1,$$

where the last inequality is because of the root contributing 1 to $m_{12}$ or $m_{13}$. Using the above equations and summarizing over all trees in the forest $F$, we have

$$m_{13} + m_{12} = I_2 + 1 \leq L - L_1 = L_2 = m_{31}.$$

\[\square\]

**Corollary 10.** Comparing the resultant partition $\mathcal{P}$ and any optimal partition $\mathcal{P}^*$, $\mathcal{P}$ has less amount of 1-paths than $\mathcal{P}^*$.

**Proof.** $p_1 = m_{13} + m_{12} + m_{11} \leq m_{31} + m_{11} \leq m_{31} + m_{21} + m_{11} = p_1^*$. \[\square\]

Corollary 10 is coincident with an intermediate result in [2, 1]. Combining with Lemma 7, we have the following theorem.

\[\square\]

**Theorem 11.** For the path partition obtained by our Greedy-TreeSearch algorithm, there exists an optimal path partition $\mathcal{P}^*$ such that $p_1^* = p_1$ and $m_{13} = m_{12} = m_{31} = m_{21} = 0$.

### 4.3 Algorithm Analysis

Recall that

$$\text{SOL} = \text{OPT} + \left( \frac{2}{3} m_{13} + \frac{1}{2} m_{12} + \frac{1}{6} m_{23} - \frac{2}{3} m_{31} - \frac{1}{2} m_{21} - \frac{1}{6} m_{32} \right).$$

By Theorem 11, we have $m_{13} = m_{12} = m_{31} = m_{21} = 0$ and thus

$$\text{SOL} \leq \text{OPT} + \frac{1}{6} (m_{23} - m_{32}). \quad (8)$$

$m_{23} \leq n$ holds trivially as there are at most $\frac{n}{2}$ 2-paths in $\mathcal{P}$ and each 2-path contributes at most 2 to $m_{23}$. Thus, we have

$$\text{SOL} \leq \text{OPT} + \frac{n}{6} \leq \text{OPT} + \frac{1}{2} \text{OPT} = \frac{3}{2} \text{OPT},$$

which matches the ratio obtained by Monnot and Toulouse [7]. Next, we present an amortized analysis to show the value of the second term in (8). $\frac{1}{6} m_{23}$ in particular, is actually much less than $\frac{n}{6}$. The idea behind is to lower bound the number of 3-paths in $\mathcal{P}$ by the number of effective 2-paths in $\mathcal{P}$, where an effective 2-path means a 2-path that contributes to the value of the second term in (8). The relation between 2-paths and 3-paths is in turn used to upper bound the number of 2-paths. Assume $\alpha \cdot p_2 \leq p_3$ for some $\alpha > 0$. Since $2p_2 + 3p_3 \leq n$, we have $p_2 \leq \frac{n}{2 + 3\alpha}$. Each 2-path contributes at most 2 to the value of $m_{23}$. Then we have

$$\text{SOL} \leq \text{OPT} + \frac{n}{6} \cdot m_{23} \leq \text{OPT} + \frac{1}{6} \cdot 2p_2 \cdot \frac{2n}{2 + 3\alpha} \leq \left( \frac{1}{2} \cdot \frac{2n}{2 + 3\alpha} + 1 \right) \text{OPT}, \quad (9)$$

where the last inequality follows from the fact $\text{OPT} \geq \frac{n}{3}$.\[\square\]
In the following, we consider a connect component in \(G[E(P^*) \cup E(P)]\), which is induced by the 2-paths in \(P\) and the neighbors in IG from \(P^*\) and thus may contain parallel edges. Besides, we only consider the 2-path that shares at least one vertex with some 3-path in \(P^*\), as otherwise a 2-path contributes 0 to \(m_{23}\) and thus a non-positive value to the second term in (8).

\[\textbf{Lemma 12.} \quad \text{For any 2-path } X = \langle x_1, x_2 \rangle \text{ in } P, \text{ either } X \text{ has at least one 3-path friend(s), or } X \text{ has only one 2-path friend in } P \text{ which has at least one 3-path friend in } P.\]

\[\textbf{Proof.} \quad \text{Let } F_i \text{ denote the set of 2-path friends via } x_i, i \in \{1, 2\}. \text{ Define the cardinality of } F_i \text{ as } f_i. \text{ Let } \bar{f} = (f_1, f_2). \text{ It is possible that } F_1 \cap F_2 \neq \emptyset. \text{ We introduce } F = \bigcup_i F_i \text{ to denote the set of distinct 2-path friends of } X. \text{ Let } f = |F|. \text{ Suppose } X \text{ does not have a 3-path friend in } P.\]

According to Theorem 11, \(X\) does not share a vertex with a 1-path in \(P^*\). If \(X\) is contained in a 3-path in \(P^*\), \(X\) has only one friend via one ending vertex; otherwise, \(X\) has friend(s) via both ending vertices. By Lemma 6, \(f \leq 2\). When \(f = 2\), either \(\bar{f} = (2, 0)\) or \(\bar{f} = (1, 1)\), both of which are impossible due to Theorem 11 and Lemma 6, respectively. Thus, we have \(f = 1\). Let \(F = \{Y\} \text{ and } Y = \langle y_1, y_2 \rangle.\)

1. \(\bar{f} = (1, 0)\): The symmetric case \(\bar{f} = (0, 1)\) can be discussed similarly. \(X\) must be contained in a 3-path in \(P^*\), say \(\langle x_1, x_2, y_1 \rangle. y_2\) cannot form a 1-path in \(P^*\) by Theorem 11, which implies \(Y\) has another friend via \(y_2\), denoted by \(Z\). If \(Z\) is a 2-path as shown in the first subfigure of Figure 5, an improving set \(\langle x_1, x_2, y_1 \rangle\) and \(\langle y_2, z_1, z_2 \rangle\) can be identified. This is a contradiction

2. \(\bar{f} = (1, 1)\): We have \(F_1 = F_2 = \{Y\}\). At least one of the \(\langle x_1, y_1 \rangle\) and \(\langle x_2, y_2 \rangle\) is a part of a 3-path in \(P^*\) as otherwise \(X\) contributes 0 to \(m_{23}\) and thus can be ignored. Without loss of generality, assume \(Y\) has another friend via \(y_2\), denoted by \(Z\). If \(Z\) is a 2-path as shown in the second subfigure of Figure 5, an improving set \(\langle x_2, x_1, y_1 \rangle\) and \(\langle y_2, z_1, z_2 \rangle\) can be identified. This is a contradiction.

\[\text{Figure 5} \quad \text{The solid and dashed edges denote the edges in } E(P) \text{ and } E(G) \setminus E(P), \text{ respectively. The paths in } P \text{ are indicated in ellipses. The thick paths form an improving set.}\]

\[\textbf{Definition 13.} \quad \text{We say a 2-path } X \text{ in } P \text{ is the special 1-hop-away friend of a 3-path } Z \text{ in } P, \text{ if they satisfy the relation in Figure 5.}\]

In the resultant partition \(P\), suppose each 3-path owns one token. We distribute each token to 2-paths in \(P\) evenly if these 2-paths are the friends or special 1-hop-away friends of this 3-path in the induced graph \(G[E(P^*) \cup E(P)]\). We say such 2-paths are \textit{associated with} this 3-path. Assume each 2-path can receive at least \(\gamma\) token in average and the value of \(\gamma\) will be estimated in Lemma 14.

\[\textbf{Lemma 14.} \quad \gamma = \min \left\{ \frac{2}{5}, \frac{2+\gamma}{4}, \frac{2+3\gamma}{7} \right\}.\]

\[\textbf{Proof.} \quad \text{Consider any 3-path } X = \langle x_1, x_2, x_3 \rangle \text{ in } P. \text{ Let } F_i \text{ denote the set of 2-path friends via } x_i, i \in \{1, 2, 3\}. \text{ Define the cardinality of } F_i \text{ as } f_i. \text{ Let } \bar{f} = (f_1, f_2, f_3). \text{ Suppose } F = \bigcup_i F_i \text{ and } f = |F|. \text{ Assume } F = \{Y_1, Y_2, \ldots, Y_f\} \text{ and } Y_i = \{y_{i1}, y_{i2}\}.\]
We claim $f \leq 4$. Otherwise, there are five different 2-path friends and we are able to find three distinct 2-paths $Y_i$ such that $Y_i \in \mathcal{F}_i$, $i \in \{1, 2, 3\}$, which is contradictory to Lemma 6. Next, we discuss how to distribute the token case by case with respect to the value of $f \leq 4$ and $\bar{f}$. Due the page limit, we only discuss Case 1: $f \leq 1$ and Case 2: $f = 2$. The discussions for Case 3: $f = 3$ and Case 4: $f = 4$ will delayed to Appendix A.

**Case 1:** $f \leq 1$. There is at most one 2-path friend and possibly one special 1-hop-away 2-path friend. Thus, each 2-path associated with $X$ receives at least $1/2$ token.

**Case 2:** $f = 2$. There are two distinct 2-path friends. If $f_1 + f_2 + f_3 \geq 4$, there exist two pairs of $i$ and $j$ with $i \neq j \in \{1, 2, 3\}$ such that $\mathcal{F}_i \cap \mathcal{F}_j \neq \emptyset$, which indicates that $X$ has no special 1-hop-away 2-path friends. If $X$ has no special 1-hop-away 2-path friends, only two 2-paths are associated with $X$ and each receives $\frac{1}{2}$ token. In the following discussion under the Case 2, we assume $X$ has at least one special 1-hop-away 2-path friend(s) and thus we suppose $f_1 + f_2 + f_3 \leq 3$.

**Case 2.1:** $f = 2$ and $f_1 + f_2 + f_3 = 2$, i.e., $\mathcal{F}_i \cap \mathcal{F}_j = \emptyset, \forall i \neq j \in \{1, 2, 3\}$.

**Case 2.1.1:** $\bar{f} = (2, 0, 0)$. The symmetric case $\bar{f} = (0, 0, 2)$ can be discussed similarly. Suppose $\mathcal{F}_1 = \{Y_1, Y_2\}$. By Theorem 11, $Y_i$ must have a friend via $y_{i2}$, $i \in \{1, 2\}$. We claim $X$ has at most one special 1-hop 2-path friend either via $Y_1$ or $Y_2$. Suppose $Y_i$ has a 2-path friend $Z_i$ via $y_{i2}$, $i \in \{1, 2\}$. As shown in Figure 6, there exists an improving set $\{\langle z_{12}, z_{11}, y_{11} \rangle, \langle z_{22}, z_{21}, y_{22} \rangle, \langle y_{11}, x_1, y_{21} \rangle, \langle x_2, x_3 \rangle\}$. Moreover, at least one of $Y_1$ and $Y_2$ has a 3-path friend in $\mathcal{P}$, which cannot be $X$ as $\bar{f} = (2, 0, 0)$. There are at most three 2-paths associated with $X$ and each receives at least $\frac{1+\gamma}{4}$ token in average.

**Figure 6** Case 2.1.1 $f = 2$ and $\bar{f} = (2, 0, 0)$. The solid and dashed edges denote the edges in $E(\mathcal{P})$ and $E(\mathcal{P}^*)$, respectively. The thick paths form an improving set.

**Case 2.1.2:** $\bar{f} = (0, 2, 0)$. Suppose $\mathcal{F}_2 = \{Y_1, Y_2\}$. By Theorem 11, $Y_i$ has another friend via $y_{i2}$, denoted as $Z_i$, $i \in \{1, 2\}$. If at least one of the $Z_1$ and $Z_2$ is a 3-path, each 2-path (special 1-hop-away) friend receives at least $\frac{1+\gamma}{3}$ token in average. Suppose both $Z_1$ and $Z_2$ are the special 1-hop-away 2-path friends of $X$. In the following discussion, we focus on $Z_1$. Without loss of generality. By Theorem 11, $X$ have friends in $\mathcal{P}$ via both $x_1$ and $x_3$. More specifically, these friends are 3-paths as we are discussing the case $\bar{f} = (0, 2, 0)$. Let’s focus on the 3-path friend via $x_3$, denoted as $Y_3 = \{y_{31}, y_{32}, y_{33}\}$. For $Y_3$, we define $\mathcal{F}_1', f_1', f_2$, $f_1'$, and $\bar{f}$ similarly.

**Case 2.1.2.1:** $Y_3$ is connected with $X$ via $y_{32}$. We claim $f_1' = f_2' = 0$ and $f_2' \leq 1$. Suppose $Y_3$ has a 2-path friend via $y_{33}$, denoted as $Z_3 = \{z_{31}, z_{32}\}$. It is possible that $Z_3 = Y_2$. As shown in the first subfigure in Figure 7, there is an improving set $\{\langle z_{12}, z_{11}, y_{11} \rangle, \langle y_{11}, x_2, x_1 \rangle, \langle x_3, y_{32}, y_{31} \rangle, \langle y_{33}, z_{31}, z_{32} \rangle\}$, which is a contradiction. $f_2' \leq 1$ because $X$ is a friend of $Y_3$ via $y_{32}$ and $Y_3$ has at most one more friend via $y_{32}$. Thus, there are at most five 2-paths associated with $X$ and $Y_3$ and each receives at least $\frac{2}{5}$ token in average.
Case 2.1.2.2: $Y_3$ is connected with $X$ via $y_{31}$. We claim $f'_3 = 0$ and $f'_1 \leq 1$. The proof is similar to the Case 2.1.2.1. An analogous example is shown in the second subfigure of Figure 7.

If $f'_1 = 1$, denote this friend via $y_{31}$ as $Z_3 = \langle z_{31}, z_{32} \rangle$. We claim $Z_3$ has another 3-path in $\mathcal{P}$, except for $X$ and $Y_3$. It is possible that $Z_3$ is coincident with $Y_2$. By Theorem 11, $Z_3$ has another friend via $z_{32}$. $W_3$ cannot be $X$ as we are discussing under the case $f = (0, 2, 0)$. If $W_3 = Y_3$, depending on whether the friendship is via $y_{32}$ or $y_{33}$ there exits an improving set $\{ \langle z_{12}, z_{31}, y_{12} \rangle, \langle y_{11}, x_{2}, x_{1} \rangle, \langle x_{3}, y_{31}, z_{31} \rangle \}$ or $\{ \langle z_{12}, z_{31}, y_{12} \rangle, \langle y_{11}, x_{2}, x_{1} \rangle, \langle x_{3}, y_{31}, z_{31} \rangle \}$, respectively, as shown in the fourth subfigure of Figure 7. This is a contradiction. We claim $W_3$ is a 3-path except for $X$ and $Y_3$. Otherwise, suppose $W_3$ is a 2-path. As shown in the fifth subfigure of Figure 7, there is an improving set $\{ \langle z_{12}, z_{31}, y_{12} \rangle, \langle y_{11}, x_{2}, x_{1} \rangle, \langle x_{3}, y_{31}, z_{31} \rangle \}$, which is a contradiction.

If $Y_3$ has a 2-path friend $Z_3$ via $y_{32}$ and $Z_3$ has another friend in $\mathcal{P}$, except for $X$ and $Y_3$, we claim $Z_3$ is a 3-path. Otherwise, suppose $W_3 = (z_{31}, z_{32})$ is a 2-path. It is possible that $Z_3$ is coincident with $Y_2$. As shown in the sixth subfigure of Figure 7, there is an improving set $\{ \langle z_{12}, z_{31}, y_{12} \rangle, \langle y_{11}, x_{2}, x_{1} \rangle, \langle x_{3}, y_{31}, z_{31} \rangle \}$, which is a contradiction.

Due to the previous discussion, we have $f' \leq 3$ and $Y_3$ has no special 1-hop-away 2-path friends.

1. $f' = 0$: there are at most four 2-paths associated with $X$ and $Y_3$, and each receives at least $\frac{1}{2}$ token in average.

2. $f' = 1$: There is only one 2-path friend associated with $Y_3$. $\vec{f} = (1, 1, 0)$ cannot happen as we discussed above. If $\vec{f} = (1, 0, 0)$, $Y_3$' 2-path friend receives extra $\gamma$ token from another 3-path. If $\vec{f} = (0, 1, 0)$, $Y_3$' 2-path friend may not receive extra tokens from other 3-paths. There are at most five 2-paths associated with $X$ and $Y_3$. Thus each 2-path receives at least $\min\{\frac{2+\gamma}{3}, \frac{\gamma}{2}\}$ token in average.

3. $f' = 2$: There are two 2-paths associated with $Y_3$. $\vec{f} = (1, 2, 0)$ cannot happen as we discussed above (also refer to the fourth subfigure of Figure 7). If $\vec{f} = (1, 1, 0)$, $Y_3$' 2-path friend via $y_{31}$ receives extra $\gamma$ token from another 3-path, but $Y_3$' 2-path friend via $y_{32}$ may not receive extra token from other 3-paths. If $\vec{f} = (0, 2, 0)$, both $Y_3$' 2-path friends via $y_{32}$ have another friends
except for \(Y_3\) and \(X\), and each receives extra \(\gamma\) token from other 3-paths. There are at most six 2-paths associated with \(X\) and \(Y_3\). Thus each 2-path receives at least \(\min\{\frac{2+\gamma}{6}, \frac{2+3\gamma}{6}\}\) token in average.

4. \(f^* = 3\): There are three 2-paths associated with \(Y_3\). If \(\vec{f} = (1, 2, 0)\), every \(Y_3\) 2-path friend receives extra \(\gamma\) token from another 3-path except for \(Y_3\) and \(X\). There are at most seven 2-paths associated with \(X\) and \(Y_3\). Thus each 2-path receives at least \(\frac{2+3\gamma}{6}\) token in average.

To summarize, each 2-path receives at least \(\min\{\frac{2}{6}, \frac{2+\gamma}{6}, \frac{2+3\gamma}{6}\}\) token in average.

Case 2.1.3: \(\vec{f} = (1, 1, 0)\). The symmetric case \(\vec{f} = (0, 1, 1)\) can be discussed similarly. Suppose \(\mathcal{F}_1 = \{X_1\}\) and \(\mathcal{F}_2 = \{Y_2\}\). We claim that \(X\) cannot have a special 1-hop-away 2-path friend via \(Y_2\). Otherwise, denote this friend as \(Z_2 = (z_{21}, z_{22})\). As shown in the first subfigure of Figure 8, there is an improving set \(\{x_1, y_{11}, y_{12}\}, (y_{21}, x_2, x_3), (y_{22}, z_{21}, z_{22})\}\), which is a contradiction.

Suppose \(X\) has a special 1-hop-away 2-path friend via \(Y_1\), denoted as \(Z_1\).

1. If \(Y_2\) has a friend via \(y_{22}\), denoted as \(Z_2\), \(Z_2\) is another 3-path except for \(X\), following the previous argument. Each of three associated 2-paths receives \(\frac{1+3\gamma}{4}\) token from \(X\).

2. If \(Y_2\) has no other friends via \(y_{22}\), it is contained in a 3-path in \(\mathcal{P}^*\) as shown in the second subfigure of Figure 8, \(X\) have a 3-path friend in \(\mathcal{P}\) via \(x_3\), denoted as \(Y_3 = \{y_{31}, y_{32}, y_{33}\}\). There are three (special 1-hop-away) 2-path friends associated with \(X\). By a similar discussion in the Case 2.1.2 “\(f = 2\) and \(\vec{f} = (0, 2, 0)^t\)”, each 2-path receives at least \(\min\{\frac{1}{2}, \frac{2+\gamma}{6}, \frac{2+3\gamma}{6}\}\) token from \(X\) and \(Y_3\) token in average.

![Figure 8](image_url) Case 2.1.3 \(f = 2\) and \(\vec{f} = (1, 1, 0)\) (left two subfigures) and Case 2.1.4 \(f = 2\) and \(\vec{f} = (1, 0, 1)\) (rightmost subfigure). The solid and dashed edges denote the edges in \(E(\mathcal{P})\) and \(E(\mathcal{P}^*)\), respectively. The thick paths form an improving set.

Case 2.1.4: \(\vec{f} = (1, 0, 1)\). Suppose \(\mathcal{F}_1 = \{X_1\}\) and \(\mathcal{F}_3 = \{Y_2\}\). Without loss of generality, assume \(X\) has a special 1-hop-away 2-path friend via \(Y_1\), denoted as \(Z_1\).

By the definition of the special 1-hop-away 2-path friend, \(Z_1 \neq Y_2\). As shown in the third subfigure of Figure 8, there is an improving set \(\{y_{12}, z_{11}, z_{12}\}, (y_{11}, x_1, x_2), (x_3, y_{21}, y_{22})\}\), which implies this case is impossible. That is, \(X\) has no a special 1-hop-away 2-path friends via \(Y_i, i \in \{1, 2\}\).

Case 2.2: \(f = 2\) and \(f_1 + f_2 + f_3 = 3\). There exist some \(i \neq j \in \{1, 2, 3\}\) such that \(\mathcal{F}_i \cap \mathcal{F}_j \neq \emptyset\).

Case 2.2.1: \(f = 2\) and \(\vec{f} = (1, 1, 1)\).

Suppose \(\mathcal{F}_1 = \{X_1\}, \mathcal{F}_2 = \{Y_1\}, \) and \(\mathcal{F}_3 = \{Y_2\}\). The symmetric case can be discussed similarly. We claim \(X\) has no a special 1-hop-away 2-path friends via \(Y_i, i \in \{1, 2\}\). Otherwise, if \(X\) has a special 1-hop-away 2-path friend via \(Y_1\), this special 1-hop-away 2-path friend can only in the format shown in the first subfigure of Figure 9, where an improving set \(\{(z_{11}, y_{11}, x_1), (z_{12}, y_{12}, x_2), (x_3, y_{21}, y_{22})\}\) can be identified; if \(X\) has a special 1-hop-away 2-path friend via \(Y_2\), as shown in the
The minimum average number of tokens a 2-path can receive is \( \gamma = \frac{2}{5} \) by solving the equation in Lemma 14. The total number of tokens is \( p_3 \) and thus we have

\[
\frac{2}{5} \cdot p_2 \leq p_3.
\]
Combining inequalities (10) and (9), we have

\[ \text{SOL} \leq \frac{21}{16} \cdot \text{OPT}. \]  

(11)

**Theorem 15.** Our algorithm Greedy-TreeSearch is a $\frac{21}{16}$-approximation algorithm.

*Proof.* The approximation ratio is shown in (11). Next we argue the running time of the Greedy-TreeSearch algorithm is polynomial.

According to the definition of our weight function, the upper bound for the weight of a path partition is $5 \times \lceil \frac{n}{2} \rceil$. Each iteration of Greedy-TreeSearch identifies an improving set and the weight of the partition increases by at least 1. Therefore, our local search algorithm terminates within $O(n)$ iterations. In the worst case, finding an improving set needs to invoke both Greedy and TreeSearch. The subroutine Greedy searches for the improving set of size at most 6 by exhausting all possible path set of size 6. Note that Greedy does not need to recheck all examined path sets. Since the collection of all $\ell$-path, $\ell \in \{1, 2, 3\}$, in $G$ has a size $O(n^3)$, the total time of invoking Greedy is $O(n^{18})$. The subroutine TreeSearch applies the modified depth first search algorithm to $G$ and has a time complexity $O(n + m)$. To summarize, the time complexity for Greedy-TreeSearch is $O(n^{18} + n \cdot (n + m)) = O(n^{18})$, which is a polynomial.

\[ \Box \]

## 5 Conclusion

We study the approximability of the minimum 3-path partition (Min-3-PP) problem, which has wide applications in the communication network. Several intrinsic structural properties on the feasible and optimal solutions are discovered. In particular, a quantitative relation between any feasible solution and the optimal solution to an arbitrary Min-3-PP instance is described. A further exploration of the optimal solution’s structure distills the quantitative relation to estimate the number of a special type of 2-paths, named as effective 2-paths. Then we show that the number of effective 2-paths is upper bounded by a ratio of the 3-paths, which implies the number of effective 2-paths cannot be too large. Inspired by the discovered properties, a novel weighted local search algorithm is designed to obtain a better approximation ratio $\frac{21}{16}$ for the Min-3-PP problem.

As we discussed in the introduction section, the Min-3-PP problem is closely related to the minimum 3-set cover problem, for which it is widely believed difficult to break the approximation barrier of $4/3$. However, we break this barrier for the Min-3-PP problem. It will be interesting to further investigate the differences and similarities between these two problems. Since the inapproximability of the Min-3-PP problem is still open, it is interesting to investigate whether there exists a better approximation algorithm or there is an approximation barrier.

For the general minimum $k$-path partition problem, its approximability is open in the literature. We think it should be also interesting to design non-trivial approximation algorithms even for some fixed $k \geq 4$.

**References**


Approximation for Min-3-PP

Recall that Lemma 14 states each 2-path receives at least \( \gamma = \min \left\{ \frac{2}{3}, \frac{2+\sqrt{3}}{6}, \frac{2+3\sqrt{3}}{9} \right\} \) token in average. We have already discussed Case 1: \( f \leq 1 \) and Case 2: \( f = 2 \). Now we continue to discuss Case 3: \( f = 3 \) and Case 4: \( f = 4 \).

Case 3: \( f = 3 \). There are three distinct 2-path friends. If \( f_1 + f_2 + f_3 \geq 5 \), we have \( \vec{f} = (1, 2, 2) \) or \( (2, 2, 1) \) or \( (2, 1, 2) \) or \( (2, 2, 2) \), which is impossible by Lemma 6. Thus, \( f_1 + f_2 + f_3 \leq 4 \).

Case 3.1: \( f = 3 \) and \( f_1 + f_2 + f_3 = 3 \), i.e., \( \mathcal{F}_i \cap \mathcal{F}_j = \emptyset \), \( \forall i \neq j \in \{1, 2, 3\} \).

Case 3.1.1: \( f = 3 \) and \( \vec{f} = (0, 2, 1) \). The symmetric case \( \vec{f} = (1, 2, 0) \) can be discussed similarly. Suppose \( \mathcal{F}_2 = \{Y_1, Y_2\} \) and \( \mathcal{F}_3 = \{Y_3\} \). Using a similar argument in the Case 2.1.3 “\( f = 2 \) and \( \vec{f} = (1, 1, 0) \)”, \( X \) cannot have a special 1-hop-away 2-path friend via \( Y_1 \) and \( Y_2 \), which can also be observed from the first subfigure of Figure 11. By Theorem 11, \( Y_i \) has friends via \( y_{i2} \), \( i \in \{1, 2\} \).

![Figure 11](image)

**Figure 11** Case 3.1.1 \( f = 3 \) and \( \vec{f} = (0, 2, 1) \). The solid and dashed edges denote the edges in \( E(P) \) and \( E(P^*) \), respectively. The thick paths form an improving set.

We discuss the following cases.

Case 3.1.1.1: \( X \) has a special 1-hop-away 2-path friend via \( Y_3 \). We claim both \( Y_1 \) and \( Y_2 \) have 3-path friends except for \( X \), denoted as \( Z_1 \) and \( Z_2 \). It is possible that \( Z_1 = Z_2 \). Assume without loss of generality, let’s focus on \( Z_1 \). Suppose \( Z_1 = \{z_{11}, z_{12}\} \). \( Z_1 = Y_3 \) is impossible in this case.
To summarize, each 2-path receives at least \(\frac{1+\gamma}{3}\) token in average.

**Case 3.1.2:** \(f = 3\) and \(\tilde{f} = (0, 1, 2)\). The symmetric case \(\tilde{f} = (2, 1, 0)\), can be discussed similarly. Suppose \(F_2 = \{Y_1\}\) and \(F_3 = \{Y_2, Y_3\}\). By Theorem 11, \(Y_1\) has another friend via \(y_{12}\), denoted as \(Z_1\). We claim \(Z_1\) is a 3-path in \(P\), which implies \(X\) has no special 1-hop-away 2-path friend via \(Y_1\). Otherwise, let \(Z = \langle z_1, z_2 \rangle\). If \(Z \notin \{Y_2, Y_3\}\), \(Z\) must be a 3-path with a similar argument in Case 3.1.1.1; if \(Z \in \{Y_2, Y_3\}\), say \(Z = Y_2\), the friendship between \(Y_1\) and \(Y_2\) can only be built via the edge \(\langle y_{12}, y_{22} \rangle\) and we can find an improving set \(\{\langle x_1, x_2, y_{11} \rangle, \langle y_{12}, y_{22}, y_{21} \rangle, \langle x_3, y_{31}, y_{32} \rangle\}\) for \(P\).

To summarize, there are at most one special 1-hop away 2-path friend associated with \(X\) and each 2-path receives at least \(\frac{1+\gamma}{3}\) token from other 3-paths in \(P\). Thus each 2-path receives at least \(\frac{1+\gamma}{3}\) token in average.

**Case 3.1.3:** \(f = 3\) and \(\tilde{f} = (2, 0, 1)\). The symmetric case \(\tilde{f} = (1, 0, 2)\) can be discussed similarly. Suppose \(F_1 = \{Y_1, Y_2\}\) and \(F_3 = \{Y_3\}\). Using a similar argument in the Case 2.1.4 “\(f = 2\) and \(\tilde{f} = (1, 0, 1)\)”, \(X\) cannot have special 1-hop-away 2-path
friends via $Y_1$ and $Y_2$, which can also be observed from the first subfigure of Figure 13. The Case 3.1.1 "$f = 3$ and $\vec{f} = (0, 2, 1)$" is a "semi-symmetric" to this case. Since the edge \{\(x_2, x_3\)\} is not used to construct an improving set during the discussion for the Case 3.1.1, the same argument still holds correctly. Subfigures for corresponding different subcases are shown in Figure 13. To summarize, each 2-path receives at least $\min\{\frac{1+f_2}{4}, \frac{1+f_3}{4}\}$ token in average.

![Figure 12](image_url) Case 3.1.2 $f = 3$ and $\vec{f} = (0, 1, 2)$. The solid and dashed edges denote the edges in $E(\mathcal{P})$ and $E(\mathcal{P}^*)$, respectively. The thick paths form an improving set.

![Figure 13](image_url) Case 3.1.3 $f = 3$ and $\vec{f} = (2, 0, 1)$. The solid and dashed edges denote the edges in $E(\mathcal{P})$ and $E(\mathcal{P}^*)$, respectively. The thick paths form an improving set.

Case 3.1.4: $f = 3$ and $\vec{f} = (1, 1, 1)$. That is, $\mathcal{F}_1 = Y_1$, $\mathcal{F}_2 = Y_2$, $\mathcal{F}_3 = Y_3$, which is impossible by Lemma 6.

Case 3.2: $f = 3$ and $f_1 + f_2 + f_3 = 4$. There exist some $i \neq j \in \{1, 2, 3\}$ such that $\mathcal{F}_i \cap \mathcal{F}_j \neq \emptyset$.

Case 3.2.1: $\vec{f} = (0, 2, 2)$. The symmetric case $\vec{f} = (2, 2, 0)$ can be discussed similarly. Without loss of generality, suppose $\mathcal{F}_2 = \{Y_1, Y_2\}$ and $\mathcal{F}_3 = \{Y_1, Y_3\}$. By Theorem 11, $Y_1$ has another friend via $y_{22}$, denoted as $Z_i$, $i \in \{2, 3\}$. By a similar argument in Case 2.2.2 "$f = 2$ and $\vec{f} = (1, 2, 0)$" and Case 2.2.3 "$f = 2$ and $\vec{f} = (2, 1, 0)$", we claim $X$ has no special 1-hop-away 2-path friends via $Y_i$, $i \in \{1, 2, 3\}$.

We claim $Z_2$ is a 3-path. Otherwise, let $Z_2 = \{z_{21}, z_{22}\}$. If $Z_2 = Y_3$, the friendship between $Y_2$ and $Y_3$ can only be built via the edge $\{y_{22}, y_{22}\}$, we can find an improving set $\{(x_1, x_2, y_{21}), (y_{22}, y_{22}, y_{31}), (y_{11}, y_{12}, x_3)\}$ for $\mathcal{P}$, as shown in the first subfigure of Figure 14. If $Z_2 \neq Y_3$, there is an improving set $\{(x_1, x_2, y_{21}), (y_{22}, z_1, z_2), (x_3, y_{31}, y_{32})\}$ for $\mathcal{P}$, as shown in the second subfigure of Figure 14.

Similarly, we can also prove $Z_3$ is a 3-path. To summarize, there are three 2-paths associated with $X$. Each of $Y_2$ and $Y_3$ receives $\gamma$ token from other 3-path friends in $\mathcal{P}$. Thus each 2-path receives at least $\frac{1+f_2}{3}$ token in average.

Case 3.2.2: $\vec{f} = (2, 0, 2)$. Without loss of generality, suppose $\mathcal{F}_1 = \{Y_1, Y_2\}$ and $\mathcal{F}_3 = \{Y_1, Y_3\}$. By a similar argument in Case 2.2.4 "$f = 2$ and $\vec{f} = (1, 0, 2)$", we claim $X$ has no special 1-hop-away 2-path friends via $Y_i$, $i \in \{1, 2, 3\}$. The Case 3.2.1 "$f = 3$ and $\vec{f} = (2, 0, 2)$" is a "semi-symmetric" to this case. Since the edge \{\(x_2, x_3\)\} is not used to construct an improving set during the discussion for the Case 3.2.1, the same argument still holds correctly. A set of subfigures for different subcases is shown in Figure 14. Each associated 2-path receives at least $\frac{1+f_2}{3}$ tokens in average.
Case 3.2.3: \( \bar{f} = (2, 1, 1) \). The symmetric case \( \bar{f} = (1, 1, 2) \) can be discussed similarly.

Suppose \( F_1 = \{Y_1, Y_2\} \), \( F_2 = F_3 = \{Y_3\} \). We claim \( X \) cannot have a special 1-hop-away 2-path friend via \( Y_i \), \( i \in \{1, 2, 3\} \). The correctness proof follows from a similar argument in Case 2.2.1 “\( f = 2 \) and \( \bar{f} = (1, 1, 1) \)”. By Theorem 11, \( Y_i \) has another friend via \( y_{i2} \), denoted as \( Z_i \), \( i \in \{1, 2\} \). We claim \( Z_i \) is a 3-path in \( P \), \( i \in \{1, 2\} \). Otherwise, let \( Z_1 = \{z_{11}, z_{12}\} \).

- If \( Z \notin \{Y_2, Y_3\} \), we can find an improving set \( \{y_{11}, x_1, x_2\}, \{y_{12}, z_{11}, z_{12}\}, \{x_3, y_{32}, y_{31}\}\} \) for \( P \). Refer to the first subfigure in Figure 15.
- If \( Z = Y_2 \), the friendship between \( Y_1 \) and \( Z \) can only be built via the edge \( \{y_{12}, y_{22}\} \) and we can find an improving set \( \{y_{11}, x_1, x_2\}, \{y_{12}, y_{22}, y_{21}\}, \{x_3, y_{32}, y_{31}\}\} \) for \( P \). Refer to the second subfigure in Figure 15.
- If \( Z = Y_3 \) and the friendship between \( Y_1 \) and \( Z \) is built via the edge \( \{y_{12}, y_{31}\} \), we can find an improving set \( \{x_1, y_{21}, y_{22}\}, \{x_3, y_{32}, y_{31}\}, \{y_{11}, y_{12}, y_{32}\}\} \) for \( P \). Refer to the third subfigure in Figure 15.
- If \( Z = Y_3 \) and the friendship between \( Y_1 \) and \( Z \) is built via the edge \( \{y_{12}, y_{32}\} \), we can find an improving set \( \{x_1, y_{21}, y_{22}\}, \{x_3, y_{32}, y_{31}\}, \{y_{11}, y_{12}, y_{32}\}\} \) for \( P \). Refer to the fourth subfigure in Figure 15.

To summarize, each 2-path receives at least \( \frac{1 + 2\gamma}{3} \) token in average.

Case 3.2.4: \( \bar{f} = (1, 2, 1) \). Suppose \( F_1 = \{Y_1\} \), \( F_2 = \{Y_2\} \), \( F_3 = \{Y_3\} \). We claim \( X \) cannot have a special 1-hop-away 2-path friend via \( Y_i \), \( i \in \{1, 2, 3\} \). The correctness proof follows from a similar argument in Case 2.2.1 “\( f = 2 \) and \( \bar{f} = (1, 1, 1) \)”. By Theorem 11, \( Y_i \) has another friend via \( y_{i2} \), denoted as \( Z_i \), \( i \in \{1, 2\} \). We claim \( Z_i \) is a 3-path in \( P \), \( i \in \{1, 2\} \). Otherwise, let \( Z_1 = \{z_{11}, z_{12}\} \).

- If \( Z \notin \{Y_2, Y_3\} \), we can find an improving set \( \{x_1, x_2, y_{11}\}, \{y_{12}, z_{11}, z_{12}\}, \{x_3, y_{32}, y_{31}\}\} \) for \( P \). Refer to the first subfigure in Figure 16.
- If \( Z = Y_2 \), the friendship between \( Y_1 \) and \( Z \) can only be built via the edge \( \{y_{12}, y_{22}\} \) and we can find an improving set \( \{x_1, x_2, y_{11}\}, \{y_{12}, y_{22}, y_{21}\}, \{x_3, y_{32}, y_{31}\}\} \) for \( P \). Refer to the second subfigure in Figure 16.
If \( Z = Y_3 \), the friendship between \( Y_1 \) and \( Z \) is built via the edge \{\( y_{12}, y_{31} \)\} or \{\( y_{12}, y_{31} \)\}. We consider the first case without loss of generality. By Theorem 11, \( Y_2 \) has another friend via \( y_{22} \), denoted as \( Z' \). We claim \( Z' \neq Z \). Otherwise, the friendship between \( Y_2 \) and \( Z \) can only be built via the edge \{\( y_{22}, y_{31} \)\}. Refer to the third subfigure in Figure 16. We can find an improving set \{\( \langle y_{21}, y_{22}, y_{31} \rangle \), \( \langle y_{11}, y_{12}, y_{32} \rangle \)\} for \( \mathcal{P} \).

To summarize, each 2-path receives at least \( \frac{1 + 2\gamma}{4} \) token in average.

![Figure 16](image-url) Case 3.2.4 \( f = 3 \) and \( \bar{f} = (1, 2, 1) \). The solid and dashed edges denote the edges in \( E(\mathcal{P}) \) and \( E(\mathcal{P}^*) \), respectively. The thick paths form an improving set.

Case 4: \( f = 4 \).

By Lemma 6, we cannot find three distinct 2-paths \( Y_i \) such that \( Y_i \in \mathcal{F}_i \), \( i \in \{1, 2, 3\} \), which implies \( f_1 + f_2 + f_3 < 5 \), that is \( f_1 + f_2 + f_3 = 4 \). Moreover, there exits one \( i \in \{1, 2, 3\} \) such that \( f_i = 0 \). Otherwise, each \( \mathcal{F}_i \) contains distinct 2-path friends, which is a contradiction.

Case 4.1: \( f = 4 \) and \( \bar{f} = (0, 2, 2) \). The symmetric case \( \bar{f} = (2, 2, 0) \) can be discussed similarly. Suppose \( \mathcal{F}_2 = \{Y_1, Y_2\} \) and \( \mathcal{F}_3 = \{Y_3, Y_4\} \). Using a similar argument in the Case 2.1.3 \( f = 2 \) and \( \bar{f} = (1, 1, 0) \), \( X \) cannot have a special 1-hop-away 2-path friend via \( Y_1 \) and \( Y_2 \). Besides, using a similar argument in the Case 2.1.1 \( f = 2 \) and \( \bar{f} = (2, 0, 0) \), \( X \) has at most one special 1-hop-away 2-path friend either via \( Y_3 \) or \( Y_4 \), and at least one of \( Y_3 \) and \( Y_4 \) has a another 3-path friend in \( \mathcal{P} \). We discuss the following cases.

Case 4.1.1: \( X \) has a special 1-hop-away 2-path friend via \( Y_4 \) without loss of generality. Followin in the argument in Case 3.1.1.1, both \( Y_1 \) and \( Y_2 \) have 3-path friends except for \( X \). There are five 2-paths associated with \( X \). \( Y_i \), \( i \in \{1, 2, 3\} \), receives \( \gamma \) token from other 2-paths in \( \mathcal{P} \). Therefore, each 2-path receives at least \( \frac{1 + 2\gamma}{5} \) token in average.

Case 4.1.2: \( X \) has no special 1-hop-away 2-path friends. There are four 2-paths associated with \( X \). For at least two paths in \{\( Y_i \), \( i \in \{1, 2, 3, 4\} \)\}, each receives \( \gamma \) token from other 2-paths in \( \mathcal{P} \). Therefore, each 2-path receives at least \( \frac{1 + 2\gamma}{4} \) token in average.

To summarize, each 2-path associated with \( X \) receives at least min\{\( \frac{1 + 3\gamma}{5} \), \( \frac{1 + 2\gamma}{4} \)\} token in average.

Case 4.2: \( f = 4 \) and \( \bar{f} = (2, 0, 2) \). Suppose \( \mathcal{F}_1 = \{Y_1, Y_2\} \) and \( \mathcal{F}_3 = \{Y_3, Y_4\} \). Using a similar argument in the Case 2.1.4 \( f = 2 \) and \( \bar{f} = (1, 0, 1) \), \( X \) cannot have a special 1-hop-away 2-path friend via \( Y_1 \), \( i \in \{1, 2, 3, 4\} \). Besides, using a similar argument in the Case 2.1.1 \( f = 2 \) and \( \bar{f} = (2, 0, 0) \), at least one of \( Y_1 \) and \( Y_2 \) \( (Y_3 \) and \( Y_4 \) \) has a another 3-path friend in \( \mathcal{P} \). There are four 2-paths associated with \( X \) For at least two paths in \{\( Y_i \), \( i \in \{1, 2, 3, 4\} \)\}, each receives \( \gamma \) token from other 2-paths in \( \mathcal{P} \). Therefore, each 2-path associated with \( X \) receives at least \( \frac{1 + 2\gamma}{4} \) token in average.