Edge Disjoint Paths Revisited *

Chandra Chekuri[†] Sanjeev Khanna[‡]

Corresponding Author: Sanjeev Khanna
574 Levine Hall
Dept. of CIS
3330 Walnut Street
University of Pennsylvania
Philadelphia, PA 19104
sanjeev@cis.upenn.edu
(215)898-0375 (tel)
(215)573-8190 (fax)

*A preliminary version of this paper appeared in Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 628–637, 2003.

[†]Bell Labs, 600 Mountain Ave, Murray Hill, NJ 07974. E-mail: chekuri@research.bell-labs.com.

[‡]Dept. of CIS, University of Pennsylvania, Philadelphia, PA 19104. E-mail: sanjeev@cis.upenn.edu.
Supported in part by an Alfred P. Sloan Research Fellowship and by an NSF Career Award CCR-0093117.

Abstract

The approximability of the maximum edge disjoint paths problem (EDP) in directed graphs was seemingly settled by an $\Omega(m^{1/2-\varepsilon})$-hardness result of Guruswami et al. and an $O(\sqrt{m})$ approximation achievable via a natural multicommodity flow based LP relaxation and also a greedy algorithm. Here m is the number of edges in the graph. However, we observe that the $\Omega(m^{1/2-\varepsilon})$-hardness of approximation applies to *sparse* graphs and hence when expressed as a function of n, the number of vertices, only an $\Omega(n^{1/2-\varepsilon})$-hardness follows. On the other hand, the $O(\sqrt{m})$-approximation algorithms do not guarantee a *sub-linear* (in terms of n) approximation algorithm for *dense* graphs. We note that a similar gap exists in the known results on the integrality gap of the flow based LP relaxation: an $\Omega(\sqrt{n})$ lower bound and an $O(\sqrt{m})$ upper bound. Motivated by this discrepancy in the upper and lower bounds we study algorithms for EDP in directed and undirected graphs obtaining improved approximation ratios. We show that the greedy algorithm has an approximation ratio of $O(\min(n^{2/3}, \sqrt{m}))$ in undirected graphs and a ratio of $O(\min(n^{4/5}, \sqrt{m}))$ in directed graphs. For acyclic graphs we give an $O(\sqrt{n \ln n})$ approximation via LP rounding. These are the first sub-linear approximation ratios for EDP. Our results also extend to EDP with weights and to the uniform capacity unsplittable flow problem (UCUFP).

1 Introduction

In this paper we consider the well-studied *edge-disjoint paths* problem (EDP). An instance of EDP consists of a (directed or undirected) graph G and a set $\mathcal{T} = \{(s_i, t_i) : 1 \leq i \leq k\}$ of k source-sink pairs, and the objective is to connect a maximum number of these pairs via edge-disjoint paths. EDP is one of the classical NP-hard problems [8] and in addition to having several applications, most notably in VLSI routing, it is a problem of central importance in combinatorial optimization and graph theory. Consequently substantial effort has gone into understanding the approximability of this problem, several of its special cases, and generalizations. See [12, 14, 2, 11, 7, 1] and the references therein for the extensive literature.

Throughout the paper, we use the term general graph to refer to both directed and undirected graphs, unless we explicitly indicate otherwise. The best known approximation ratio for general graphs is $O(\sqrt{m})$ [23, 14] where m is the number of edges in the graph. Essentially there have been two approaches (algorithms) to obtain upper bounds for general graphs: (randomized) rounding of the fractional solution to the natural multicommodity flow based relaxation [23, 2, 14, 19], and the greedy algorithm [12, 14, 11, 1]. We refer to the former as the LP approach and the latter as the greedy approach. Let $\mathcal{O}^*$ denote an optimal solution to the LP relaxation for a given instance and let OPT denote the integral optimum value for the same instance. It is clear that $\mathcal{O}^* \geq \text{OPT}$. The best known results for EDP are summarized in [14]. For the LP approach, Srinivasan [23] shows that randomized rounding obtains $\Omega(\max\{\frac{\mathcal{O}^*}{\sqrt{m}}, \frac{(\mathcal{O}^*)^2}{m}, \frac{\mathcal{O}^*}{d}\})$ paths where d is average length of flow paths in the LP solution. Kolliopoulos and Stein [14] show similar bounds for the greedy algorithm obtaining $\Omega(\max\{\frac{\text{OPT}}{\sqrt{m_0}}, \frac{\text{OPT}^2}{m_0}, \frac{\text{OPT}}{d_0}\})$ paths where m_0 and d_0 are the number of edges and average length of paths respectively, in some optimum solution.

EDP in directed graphs appears to be fundamentally harder than in undirected graphs. For example, given a directed graph G, the problem of deciding if two given pairs, (s_1, t_1) and (s_2, t_2), can be connected via disjoint paths is already NP-hard [6]. In undirected graphs, for any fixed k, Robertson and Seymour [21] gave a polynomial time algorithm to decide if a given set of k pairs can be connected via disjoint paths. However, their algorithm is highly non-trivial and requires extensive machinery from the work on graph minors. Surprisingly, there is a fairly simple algorithm for the same problem in directed *acyclic* graphs [6]. Until recently, it was only known that EDP is APX-hard [10]. Guruswami et al. [11], via a simple reduction to the two pair decision problem mentioned above, showed that it is NP-hard to approximate EDP in directed graphs to within a $\Omega(m^{1/2-\varepsilon})$ factor for every fixed $\varepsilon > 0$. Independently, Ma and Wang [17] showed, via a more complex and indirect reduction from the maximization version of the label cover problem, that EDP in directed graphs is $\Omega(2^{\log^{1-\varepsilon} n})$-hard. Although the hardness result in [11] is stronger, we observe that the result in [17] applies even to *acyclic* graphs; the reduction in [11] uses graphs with cycles in an essential way.

At first glance it would appear that the complexity of the directed EDP has been resolved.

However, this is only true if we use m as the parameter for expressing the approximability of EDP. The hardness of approximation shown in [11] applies only to *sparse* graphs. When expressed as a function of n, the number of vertices in the graph, the hardness of approximation obtained is $\Omega(n^{1/2-\varepsilon})$. An obvious question to ask is if one can obtain an $o(n)$-approximation for EDP in directed graphs. In particular, is there an $O(\sqrt{n})$ approximation algorithm? A positive answer to this question would settle the approximability of the problem, and also improve the current ratio of $O(\sqrt{m})$ for all graphs with super-linear number of edges (by a substantial factor for *dense* graphs). Furthermore, answering this question would help us understand whether m or n is the right parameter to express the approximability results for EDP. We note that the *vertex-disjoint paths* problem is a special case of EDP in directed graphs for which an essentially tight $\Theta(\sqrt{n})$ approximability threshold has been established [11].

In this paper, we make progress towards answering the above question in the positive by obtaining an $O(n^{4/5})$ approximation for EDP in directed graphs. For directed *acyclic* graphs we obtain an $O(\sqrt{n \ln n})$-approximation. For undirected graphs we obtain an $O(n^{2/3})$-approximation. These are the first sub-linear (in n) approximation guarantees for the problem. In the process we obtain a *tight* analysis for the greedy algorithm and show improved bounds on the integrality gap of the natural LP relaxation. We remark that Kolliopoulos [13] has obtained an $O(n^{3/4})$ bound in acyclic graph instances in which no vertex in the graph is an end point of more than one of the pairs to be routed.

Overview of Results and Techniques: Our first set of results are based on the analysis of the greedy algorithm for EDP [14] that iteratively picks a shortest path among unrouted pairs[1]. We henceforth refer to this algorithm as Greedy. We show that Greedy is an $O(\min\{n^{2/3}, \sqrt{m}\})$-approximation algorithm for undirected graphs and an $O(\min\{n^{4/5}, \sqrt{m}\})$-approximation algorithm for directed graphs. An upper bound of $O(\sqrt{m})$ on the approximation ratio of Greedy is known [14], and is based on two observations: if a greedily chosen path is "short" (has $O(\sqrt{m})$ edges) then it can result in the removal of only $O(\sqrt{m})$ paths in some fixed optimum solution; if it is "long" (has $\Omega(\sqrt{m})$ edges), then the graph has capacity to route only $O(\sqrt{m})$ pairs in the residual graph. The key observation underlying our analysis of the greedy algorithm is that when there are no short paths, the capacity of the graph to route additional pairs is more strongly diminished than suggested by the above counting argument. Even and Tarjan [5] showed that in a simple unit capacity directed graph, for any two vertices s and t, the maximum flow from s to t is $O(n^2/\ell^2)$ where ℓ is the shortest path distance from s to t. Our analysis of the greedy algorithm is based on the following question. Given a unit capacity graph G, let S_ℓ be the set of *all* ordered pairs (s_i, t_i) such that $\text{dist}_G(s_i, t_i) \geq \ell$. What is the maximum *multi-commodity*

[1] Kleinberg [12] defined an online *bounded* greedy algorithm that picks a pair as long as the shortest path between them is at most a parameter L. Our analysis is in fact for the bounded greedy algorithm, however we use the greedy algorithm as defined in [14] since it is a more natural offline algorithm.

flow[2] between pairs in S_ℓ? We show via a simple counting argument that in undirected graphs the maximum flow is $O(n^2/\ell^2)$. However, our argument does not generalize to directed graphs. We use a different idea based on the work of Cheriyan, Karloff, and Rabani [4] on directed multicuts to show that there exists a cut of value $O(n^4/\ell^4)$ that separates all pairs in S_ℓ. This of course also implies that the maximum flow is $O(n^4/\ell^4)$.

The bounds we establish on the approximation ratio of Greedy can be translated into similar upper bounds on the integrality gap of the flow based LP relaxation. We are thus able to improve the gap between the known upper bound of $O(\sqrt{m})$ [23] and lower bound of $\Omega(\sqrt{n})$ [10]. We show that Greedy has an improved approximation in uniformly dense undirected graphs. In graphs with minimum degree δ, Greedy yields an $O(n/\delta)$ approximation. For example, if $\delta = \Omega(n^{1-\varepsilon})$, we obtain an $O(n^\varepsilon)$ approximation. Our analysis for this case uses a charging scheme that is different from the one for the general case. We also show that there is an infinite family of undirected (and directed) graphs for which the approximation ratio obtained by Greedy is $\Omega(n^{2/3})$. Thus our analysis of Greedy is tight for undirected graphs.

Our second result is for EDP in directed *acyclic* graphs. For these graphs we obtain an $O(\sqrt{n \ln n})$ approximation. We also establish the same upper bound on the LP integrality gap which nearly matches the $\Omega(\sqrt{n})$ lower bound [10]. The underlying idea behind our algorithm is different from the earlier approaches and we describe it next.

The difficult case for all existing approaches to EDP is when the paths chosen by the optimal algorithm are $\Omega(\sqrt{n})$ long. We address this case as follows: a simple counting argument shows that if there are many such paths, then there must a vertex v with high congestion (that is, many paths go through it). In order to take advantage of this observation we introduce a restricted class of instances that we call v-*separable* instances. Informally speaking, these instances have the property whereby the vertex set can be partitioned into two sets, one containing only sources and other only sinks, such that v is the only common vertex in these two sets. Hence all paths between the pairs share the bottleneck vertex v. Such instances naturally define two single source disjoint paths problems. We use an LP-based approach to obtain an $O(\ln n)$-approximation for v-*separable* instances. In acyclic graphs, every vertex v naturally defines a v-separable instance, and this is the key to obtaining our $O(\sqrt{n \ln n})$ approximation.

Based on our results we conjecture the following.

Conjecture 1.1 *There is an $O(\sqrt{n})$-approximation for EDP in directed graphs. Moreover, the LP integrality gap is $\Theta(\sqrt{n})$.*

In this paper we concentrate on the basic edge disjoint paths problem, however, our ideas extend to some generalizations of the problem. In particular, with a loss of an additional logarithmic factor in the approximation ratio, it is straightforward to extend our results to the *weighted*

[2]There are two variants of multi-commodity flow: maximum concurrent flow, and maximum sum multi-commodity flow. In this paper we are concerned only with sum multi-commodity flow.

EDP: each pair (s_i, t_i) has a non-negative weight associated with it and the goal is to find a maximum weight set of pairs that can be routed. We also show that our results extend to the *uniform capacity unsplittable flow problem* (UCUFP). In a UCUFP instace all edges of the given graph have the same capacity C. Each pair (s_i, t_i) has a demand ρ_i associated with it and the goal is to maximize the number of pairs that can be feasibly routed where a pair is routed if all of its demand is satisfied on a single path.

Organization: We start by describing the multicommodity flow based LP relaxation for EDP in Section 2. In Section 3, we present a new analysis of Greedy as well as our upper bound on the LP integrality gap. We also show here that Greedy can be used to get stronger approximation guarantees when the underlying graph has a large minimum degree. In Section 4 we describe our $O(\sqrt{n \ln n})$ approximation for directed acyclic graphs. To obtain this result we formalize the notion of a v-separable instance and show that a good approximation for v-separable instances translates into a good approximation for EDP in directed acyclic graphs. In Section 5 we give an LP-based approximation algorithm for v-separable instances. Section 6 describes extension to the uniform capacity unsplittable flow problem.

2 Preliminaries

EDP has a natural IP formulation based on multicommodity flow. Here we give the path formulation for convenience, there is an equivalent and compact flow formulation. Let $\mathcal{P}_i$ denote the set of all paths in G from s_i to t_i. For each path π in $\cup_i \mathcal{P}_i$ we have a binary variable f_π that indicates if π is chosen in the solution or not. For each pair (s_i, t_i) we have a variable x_i indicating whether the pair is chosen to be connected or not.

$$\max \sum_{i=1}^{k} x_i \text{ s.t.}$$

$$\begin{aligned}
\sum_{\pi \in \mathcal{P}_i} f_\pi &= x_i & i = 1, 2, \ldots, k \\
\sum_{i=1}^{k} \sum_{\pi \in \mathcal{P}_i : \pi \ni e} f_\pi &\leq 1 & e \in E(G) \\
x_i &\leq 1 & i = 1, \ldots, k \\
f_\pi &\in \{0, 1\} & \pi \in \bigcup_{i=1}^{k} \mathcal{P}_i
\end{aligned}$$

By relaxing the final set of constraints to $f_\pi \in [0, 1]$ for $\pi \in \bigcup_{i=1}^{k} \mathcal{P}_i$, we obtain an LP relaxation that we shall refer to as LP-MCF. Given a fractional solution $(x_1, \ldots, x_k, f_{\pi_1}, f_{\pi_2}, \ldots)$, we shall refer to $\sum_{i=1}^{k} x_i$ as the value of the solution. We shall say that the solution uses a flow path π if $f_\pi > 0$. An optimal solution to LP-MCF can be obtained in polynomial time using standard techniques. We will assume that the number of flow paths with non-zero flow is $O(m)$; from standard polyhedral theory, a basic solution to the linear program satisfies this property. Simple and efficient combinatorial methods that obtain a $(1 + \varepsilon)$ approximation are also known [9].

A cut in a graph, undirected or directed, is a subset of edges $E' \subseteq E$. A cut E' separates an ordered pair of vertices (u, v) if there is no path from u to v (a directed path if G is directed) in the graph obtained by removing the edges in E'. In undirected graphs we restrict our attention to cuts induced by a subset of vertices $S \subset V$; the cut induced by S is the set of all edges that have one end point in S and the other in $V \setminus S$. In a directed graph, a cut induced by a vertex subset S consists of all edges (u, v) with u in S and $v \in V \setminus S$.

We need the following large deviation bounds.

Proposition 2.1 *Let $X_1, X_2, \ldots, X_n$ be independent binary random variables and let $X = \sum_i X_i$ with $\mu = \mathrm{E}[X]$. Then the following hold for sufficiently large n.*

- *If $\mu \geq \gamma \ln n$ then for $\delta > 0$,* $\Pr[X < (1 - \delta)\mu] \leq 1/n^{\gamma\delta^2/2}$.
- *If $\mu \leq \ln n$, for $\delta > e^2 - 1$,* $\Pr[X > (1 + \delta)\ln n] < 1/n^{1+\delta}$.

Proof: Let $\delta > 0$. Standard Chernoff-Hoeffding bounds [18] show that $\Pr[X < (1 - \delta)\mathrm{E}[X]] < \exp(-\mu\delta^2/2)$ and $\Pr[X > (1 + \delta)\mathrm{E}[X]] < (e^\delta/(1+\delta)^{(1+\delta)})^\mu$. Substituting for μ and simplifying yields the desired bounds. ■

3 The Greedy Algorithm

We show that Greedy gives an approximation ratio of $O(n^{2/3})$ for the undirected disjoint paths problem and an approximation ratio of $O(n^{4/5})$ for the directed case. The algorithm simply considers the shortest path length for each unrouted (s_i, t_i) pair and connects a pair with minimum shortest path length via its shortest path. The earlier analysis of Greedy [12, 14] used the fact that when the minimum path length for unrouted pairs exceeds a certain bound, say ℓ, then OPT can route only $O(m/\ell)$ pairs. We show that when m is large, the capacity of the network to route pairs in this case is more restricted than suggested by this bound. In particular, we show that in the undirected case, an optimal solution can route at most $O(n^2/\ell^2)$ pairs while in the directed case, it can route at most $O(n^4/\ell^4)$ pairs. These bounds apply even to an optimum solution to LP-MCF and we thus obtain the first sub-linear bounds on its integrality gap. We start by establishing these bounds.

Theorem 3.1 *Let $G(V, E)$ be a simple unit capacity undirected graph and let $\mathcal{T}$ be the collection of all source-sink pairs such that the shortest distance in G between source and sink is at least ℓ. Then the value of the maximum (fractional) multicommodity flow for the pairs in $\mathcal{T}$ is bounded by $O(n^2/\ell^2)$.*

Proof: In the following, for convenience of notation, we assume that ℓ is a multiple of 6. We refer to a vertex in G as a high-degree vertex if its degree is strictly greater than $6n/\ell$. A vertex that is not high-degree is a low-degree vertex. We claim that for every pair (s, t) in $\mathcal{T}$, any path from s to t must use at least $\ell/6$ low-degree vertices. It then follows that a fractional flow path carrying a flow of $\varepsilon > 0$ consumes $\Omega(\varepsilon \cdot \ell)$ capacity incident on the low-degree vertices. Thus the total fractional flow that can be supported by G is $O(n^2/\ell^2)$ since the total capacity incident to the low-degree vertices is $O(n^2/\ell)$.

We prove the above claim by contradiction. Suppose there exists a path P that connects a pair $(s, t) \in \mathcal{T}$ using less than $\ell/6$ low-degree vertices. Consider a breadth-first search tree rooted at s, and let *layer* L_i denote the set of vertices at distance i from s. It is easy to see that every edge connects vertices that are either in the same layer or in two adjacent layers. We claim that at least $\ell/6$ layers among the first ℓ layers must consist only of low-degree vertices. Suppose not. Partition the layers into blocks of 3 contiguous layers each such that the block B_i consists of vertices in the layers L_{3i+1}, L_{3i+2}, and L_{3i+3}. Discard any blocks that contain a layer consisting only of low-degree vertices. By our assumption on P, at most $\ell/6$ blocks are discarded and thus we remain with at least $\ell/6$ blocks. Since the blocks are disjoint, one of the remaining blocks contains at most $6n/\ell$ vertices altogether. Consider a high-degree vertex in the middle layer of such a block — all its edges go to other vertices within the block. A contradiction! Thus there must be at least $\ell/6$ layers among $L_1, L_2, ..., L_\ell$ that consist entirely of low-degree vertices. Now any s-t path must cross these layers and thus will be incident to at least one low-degree vertex in each such layer. This proves the claim and the theorem. ■

The preceding argument does not extend to directed graphs. The technical difficulty is that in a breadth first tree from a given vertex in a directed graph, even though there cannot be forward arcs between non-adjacent layers, there could be many back arcs between non-adjacent layers. It is easy to construct graphs where the in-degree and the out-degree of every vertex is $\Omega(n)$ but there still exists an s-t pair such that the shortest path distance from s to t is $\Omega(n)$.

For directed graphs we use a different argument that is in part inspired by the algorithm and analysis of Cheriyan, Karloff and Rabani [4] for the directed multicut problem. They find a cut based on an LP relaxation while we find a cut using the fact that pairs are far apart in the graph. We recall that in the context of directed graphs, a cut is simply a subset of edges. We say that a pair (u, v) is separated by a cut $E' \subset E$ if there is no directed path from u to v in the graph obtained by removing the edges in the cut E'.

Theorem 3.2 *Let $G(V, E)$ be a simple unit capacity directed graph and let $\mathcal{T}$ be the collection of all source-sink pairs such that the shortest distance in G between source and sink is at least ℓ. Then there exists a cut of size $O(n^4/\ell^4)$ that separates all pairs in $\mathcal{T}$, and hence the value of the maximum (fractional) multicommodity flow for the pairs in $\mathcal{T}$ is bounded by $O(n^4/\ell^4)$.*

Proof: In the following, pairs are ordered and when we say an s-t path we mean a directed path from s to t. Separating s from t does not imply that t is separated from s.

We show that there exists a set $E' \subseteq E$ of size $O(n^4/\ell^4)$ such that by deleting edges in E', we can separate every source-sink pair in $\mathcal{T}$. We construct the set E' iteratively as follows. Initialize E' to $\emptyset$. We denote the residual graph $G(V, E \setminus E')$ by G'. At the beginning of the iteration, we pick any pair $(s, t) \in \mathcal{T}$ such that G' contains an s-t path. We do a breadth-first search from s in the graph G'. Let layer L_j denote the vertices at distance j from s. We partition these layers into blocks of 2 adjacent layers each such that the ith block B_i consists of vertices in the layers L_{2i+1} and L_{2i+2}. By our assumption, t occurs in some block B_r where $r \geq \ell/2$. Let $Y = B_1 \cup B_2 ... \cup B_{\ell/6}$ and let $Z = B_{\ell/3+1} \cup B_{\ell/3+2} ... \cup B_r$. Consider the $\ell/6$ blocks not contained in Y or Z, namely, the blocks $B_{\ell/6+1}$ through blocks $B_{\ell/3}$. Since the blocks are disjoint, there exists a block B_k among these blocks such that the total number of vertices in B_k does not exceed $6n/\ell$. We include in E' all edges within this block, that is, edges that go from layer L_{2k+1} to L_{2k+2} — the total number of these edges is $O(n^2/\ell^2)$. After we remove these edges it is clear that there is no s-t path in G'. We iterate until there are no pairs in $\mathcal{T}$ left to separate. Now we argue that the total number of iterations cannot exceed $O(n^2/\ell^2)$. This, combined with the fact that we remove $O(n^2/\ell^2)$ edges in each iteration, gives us the desired bound on $|E'|$.

Consider an iteration above that separates the pair s and t and let P be a shortest path from s to t in G'. From the above arguments it is clear that not only do we separate s from t, we also separate Y from Z. In particular we separate $P \cap Y$ from $P \cap Z$. Note that since P is a path, every vertex in $P \cap Y$ could reach every vertex in $P \cap Z$ before this iteration. The sets $P \cap Y$ and $P \cap Z$ contain at least $\ell/3$ vertices each, thus each iteration separates $\Omega(\ell^2)$ pairs of vertices that were connected before the iteration. Hence the total number of iterations is $O(n^2/\ell^2)$. ■

We now show how Theorems 3.1 and 3.2 can be used to obtain tighter bounds on the performance of Greedy. For the purpose of our analysis, we will find it more convenient to work with the variant of Greedy that terminates as soon as the minimum shortest path length among unrouted pairs exceeds a certain bound, say ℓ. We refer to this algorithm as bounded Greedy. We will assume for convenience that at least one pair is routed by the algorithm. Clearly, Greedy routes more pairs than bounded Greedy. The analysis below applies to undirected as well as directed graphs. Unless necessary, we will not make any explicit reference to the directed or the undirected nature of the underlying graph.

Let S be the set of paths chosen by Greedy. Fix an optimum solution, say OPT. Let O_u denote the set of pairs that are routed by OPT and not by Greedy. If a path in O_u shares an edge with a path $P \in S$, we charge it to the path P in the solution output by Greedy. Let O_1 denote the set of paths in O_u that can be accounted for in this manner. Clearly, $|O_1| \leq |S| * \ell$. Let $O_2 \stackrel{\text{def}}{=} O_u \setminus O_1$. Also, let $G_r(V, E_r)$ be the subgraph of G that remains at the termination of Greedy. Applying Theorems 3.1 and 3.2 to the graph G_r, it follows that $|O_2|$ is $O(n^2/\ell^2)$ in the undirected case and $O(n^4/\ell^4)$ in the directed case. Thus choosing a threshold of $\ell = n^{2/3}$ for the undirected case, and a threshold of $\ell = n^{4/5}$ in the directed case gives us the following result.

Theorem 3.3 *For EDP, Greedy gives an $O(n^{2/3})$-approximation in undirected graphs and an $O(n^{4/5})$-approximation in directed graphs.*

3.1 Lower Bound for Greedy

We next show that the performance ratio of Greedy is $\Omega(n^{2/3})$ for both undirected and directed EDP. Thus our analysis is asymptotically tight in the undirected setting.

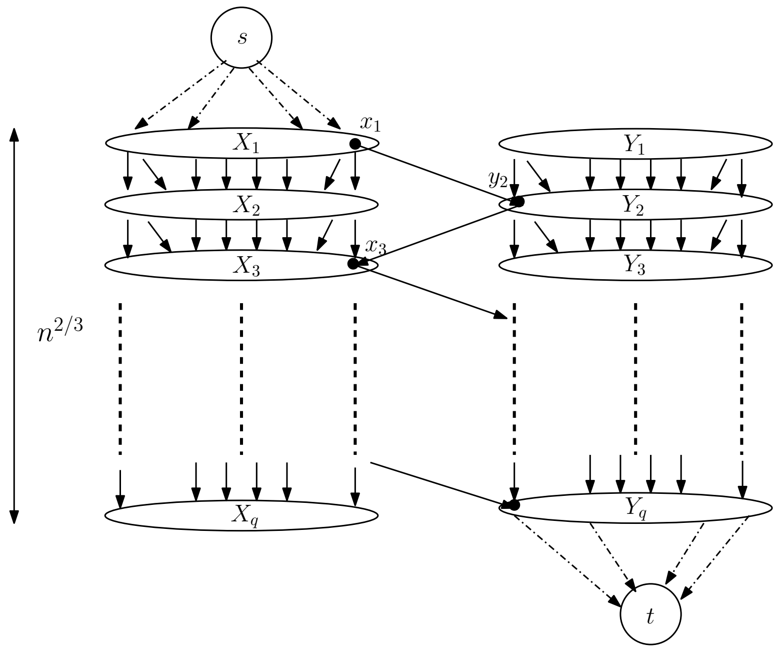

Figure 1: A lower bound instance for Greedy. s and t are only for the proof.

Theorem 3.4 *There exists an infinite family of directed and undirected instances on which the approximation ratio achieved by Greedy is $\Omega(n^{2/3})$. Moreover, the directed instances in the family are acyclic.*

Proof: We now describe a family of instances on which Greedy routes only one path while there exists a solution of size $\Omega(n^{2/3})$. We focus on the directed case here; however, the underlying undirected graph serves as a lower bound example for the undirected case. The instance consists of two layered graphs, say G_1 and G_2, where each G_i contains $n^{2/3}$ layers of $n^{1/3}$ vertices each. For notational ease, let $p = n^{1/3}$ and let $q = n^{2/3}$. Let $X_1, ..., X_q$ be the layers of G_1, and let $Y_1, ..., Y_q$ be the layers of G_2. Then for $1 \le i < q$, every vertex in X_i is connected to every vertex in X_{i+1}. Y_i is connected to Y_{i+1} in a similar way. The two graphs are connected as follows. For each odd i, we pick a representative vertex in X_i, say x_i, and for each even i, we pick a representative vertex in Y_i, say y_i. We then add edges (x_1, y_2), (y_2, x_3),...,(x_{q-1}, y_q) — assume without any loss of generality that q is even. The source sink pairs are given by the set $\{(x, y) \mid x \in X_1, y \in Y_q\}$. This completes the description of our instance. See Figure 1.

Observe that every source-sink path is of length at least $q - 1$. Suppose Greedy chooses to connect the pair (x_1, y_q) along the path $x_1 \to y_2 \to ...x_{q-1} \to y_q$. Note that this path is a shortest x_1-y_q path. (By a simple modification of adding an extra vertex on each edge between X_1 and X_2, we can ensure that this path is the unique shortest path between unrouted pairs.) Then all remaining source-sink pairs are disconnected in the residual graph and thus no more pairs can be routed by Greedy. On the other hand, we claim that an optimal algorithm can route $\Omega(n^{2/3})$ pairs. Notice that by our choice of the source-sink pairs, it suffices to show that any cut that separates X_1 from Y_q must have capacity at least $\Omega(n^{2/3})$. In order to show this, it is convenient to imagine a super-source vertex s connected to X_1 via edges of infinite capacity, and a super-sink vertex t to which all vertices in Y_q are connected by edges of infinite capacity. We will argue that any cut that separates s and t has capacity at least $n^{2/3}/4$. Consider any such cut (S, T). Suppose there exists a layer X_i such that $|S \cap X_i| \le n^{1/3}/2$. Consider the smallest index i for which this occurs. Since $|S \cap X_{i-1}| \ge n^{1/3}/2$, it is easy to see that the capacity of the (S, T) cut is at least $n^{2/3}/4$ and we are done. Thus we may assume that $|S \cap X_i| \ge n^{1/3}/2$ for each X_i. A similar argument implies that $|T \cap Y_j| \ge n^{1/3}/2$ for each Y_j. Now suppose S contains more than $n^{1/3}/2$ representative vertices from G_2. Then by the preceding condition, each such vertex has at least $n^{1/3}/2$ neighbors in T and thus the capacity of the cut is at least $n^{2/3}/4$. So it must be that at least $(n^{2/3} - n^{1/3})/2$ of the representative vertices in G_2 are in T. A similar argument shows that at least $(n^{2/3} - n^{1/3})/2$ of the representative vertices in G_1 are in S. This once again implies that the capacity of the cut is at least $n^{2/3}/4$. ■

3.2 Integrality Gap of the LP Formulation

The preceding analysis of Greedy can be used to obtain improved upper bounds on the integrality gap of LP-MCF in directed and undirected graphs. The best previous bound is $O(\sqrt{m})$. A lower bound of $\Omega(\sqrt{n})$ is known even in directed acyclic graphs and it remains open if a matching upper bound exists. As we will shortly see in Section 4, this question can be essentially answered in the affirmative for directed acyclic graphs.

Theorem 3.5 *The integrality gap of* LP-MCF *in undirected graphs is $O(n^{2/3})$ and in directed graphs is $O(n^{4/5})$.*

Proof: Let $\ell = n^{2/3}$ in the undirected case and $\ell = n^{4/5}$ in the directed case. Given a feasible fractional solution, let us consider the subgraph G' of G that contains edges with positive fractional flow. We now run the bounded Greedy on G' (i.e. terminate once the shortest path length for each unrouted pair exceeds ℓ). Let S be the set of paths chosen by the algorithm. We will focus only on pairs (fractionally) routed by the fractional solution but unrouted by Greedy. Fix a decomposition of the fractional solution into fractional flow paths, and discard any flow paths for pairs that are routed by Greedy. Any fractional flow path that intersects with a path in S is charged to S. Clearly, the total flow on the flow paths charged in this manner is upper bounded by $|S| \cdot \ell$.

Now consider the residual graph G_r that consists of edges in G' that do not intersect with S. Any unaccounted fractional flow path is strictly contained in G_r. As in the proof of Theorem 3.5, the total fractional flow that can be routed in G_r can be bounded using Theorems 3.1 and 3.2. Putting together, we get the claimed bounds. ∎

We note here that the algorithm used in Theorem 3.5 ignores the flow paths of the fractional solution. This is in contrast to randomized rounding techniques that route each pair on some flow path for that pair provided by the fractional solution.

3.3 Undirected Dense Graphs

We show that in dense graphs (i.e. graphs where the minimum degree is large), Greedy gives improved approximation ratios. We use a slightly different analysis here. Specifically, we show that any pair that is routed by an optimal solution and is not routed by Greedy, must pass through a "heavily-used" vertex in the solution output by Greedy. As a result, we can account for these unrouted pairs by charging to the paths in Greedy that pass through these heavily-used vertices.

Theorem 3.6 *The approximation ratio of Greedy is $O(n/\delta)$ in undirected graphs with minimum degree at least δ. In particular, for any $0 \leq \varepsilon < 1$, in graphs with minimum degree $\Omega(n^{1-\varepsilon})$, the approximation ratio of Greedy is $O(n^{\varepsilon})$.*

Proof: As before, we work with the bounded Greedy algorithm – we use a bound of $6n/\delta$. Upon termination of the Greedy, let G_r denote the residual graph. We partition the vertices into two sets, namely, *saturated* and *unsaturated*, based on the ratio of their degrees in G and G_r. A vertex is called saturated if its degree in G_r is less than half of its initial degree in G and unsaturated otherwise.

Let S be the set of paths chosen by Greedy. Fix an optimal solution, say OPT, and let O_u denote the set of pairs in OPT not connected by the solution of Greedy. For each saturated vertex v, define S_v to be the subset of paths in S that pass through vertex v. If a path in O_u passes through a saturated vertex v, we simply charge this path to the set S_v. Clearly, since v is saturated, at most $2|S_v|$ paths in O_u are charged to a set S_v in this manner. On the other hand, $\sum_v |S_v| \leq (6n/\delta)|S|$. Thus we account for at most $12(n/\delta)|S|$ paths in O_u in this manner; let O_1 denote this set. We now consider $O_2 \stackrel{\text{def}}{=} O_u \setminus O_1$. Every path in O_2 passes through only vertices that are unsaturated and hence have degree at least $\delta/2$ in G_r. An elementary result in graph theory shows that any connected graph with minimum degree δ_0 has diameter at most $3n/\delta_0$ [25]. Thus any pair in O_2 can be connected by a path of length at most $6n/\delta$. This contradicts the fact that Greedy terminated. Thus $O_2 = \emptyset$. Putting together, $|\text{OPT}| \leq |S| + 12(n/\delta)|S|$ and thus the theorem follows. ■

The above theorem does not hold in directed graphs for reasons similar to those mentioned before Theorem 3.2.

4 An $O(\sqrt{n \ln n})$ approximation for acyclic graphs

In this section we obtain an $O(\sqrt{n \ln n})$-approximation algorithm for EDP in acyclic graphs, based on the LP relaxation. This ratio nearly matches the known lower bound of $\Omega(\sqrt{n})$ on the integrality gap of the relaxation. For a large class of general graphs our result is the first to provably demonstrate that the LP provides asymptotically better approximation ratios than Greedy and its variants. We also remark that in contrast to typical randomized rounding approaches, our algorithm routes pairs on paths that are not necessarily provided by the LP solution.

The Main Idea: We now describe the key high level idea behind our algorithm. Let α be a parameter to be chosen later. To obtain an $O(\alpha)$ approximation algorithm it is sufficient to restrict ourselves to instances in which the shortest path between s_i and t_i for every i is at least α long; we can greedily route pairs with paths of length at most α and remove them. Let OPT be the number of paths in an optimal solution to an instance with only long (of length at least α) paths. Since each of the optimal paths is constrained to be long, they share many vertices. In particular, a straightforward averaging argument shows that there must be a vertex u such that $\Omega(\text{OPT} \cdot \alpha/n)$ paths go through u. We can assume knowledge of u by enumerating over all vertices. How does this information help us? Consider the following variant of EDP: find the maximum number of pairs that can be routed via disjoint paths all of which are constrained to go through a specific vertex v. We show that this variant can be well approximated in acyclic graphs because of the following fact: given any vertex v, the set of vertices which can reach v via a directed path, and the set of vertices which can be reached by v via a directed path, are *disjoint*. In particular the variant of EDP mentioned above induces what we call a v-*separable* instance (to be formally

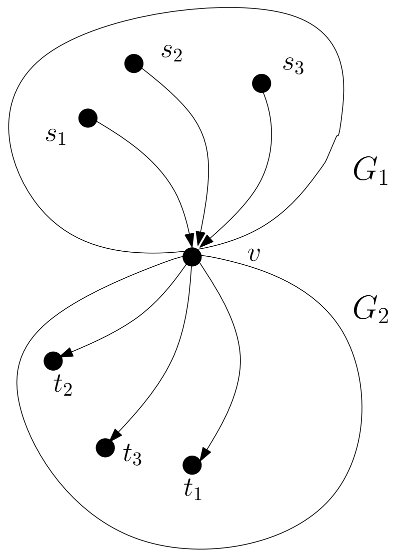

Figure 2: A v-separable instance.

defined shortly). In Section 5 we show that an $O(\ln n)$ approximation can be obtained for a v-separable instance via LP rounding. Therefore by guessing u, we can route $\Omega(\text{OPT} \cdot \alpha/(n \ln n))$ pairs which leads to the promised $O(\sqrt{n \ln n})$ approximation if α is chosen to be $\sqrt{n \ln n}$. We now define a v-separable instance.

Definition 4.1 *Let I be an instance of EDP on $G = (V, E)$ with k pairs (s_i, t_i), $1 \leq i \leq k$. An instance I is v-separable if we can partition G into two induced graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ such that:*

- *$V = V_1 \cup V_2$ and $V_1 \cap V_2 = \{v\}$, and*
- *$E = E_1 \cup E_2$, and $E_1 \cap E_2 = \emptyset$, and*
- *for $1 \leq i \leq k$, $s_i \in V_1$ and $t_i \in V_2$.*

Note that in a v-separable instance all paths from s_i to t_i have to go through v. See Figure 2. Let I be an instance of EDP on an acyclic graph G. For a vertex $v \in V(G)$, let G_v^{in} (for G into v) be the subgraph of G induced by the vertices of $V(G)$ that can reach v in G. Similarly let G_v^{out} (for G out of v) be the subgraph induced by vertices that can be reached by v in G. Since G is acyclic, G_v^{in} and G_v^{out} share only v. Given a set of pairs (s_i, t_i) to be routed in G, we restrict the pairs to those such that $s_i \in V(G_v^{in})$ and $t_i \in V(G_v^{out})$. By doing this we naturally obtain a v-separable instance of the edge disjoint paths problem with $G_1 = G_v^{in}$ and $G_2 = G_v^{out}$. Further we note that any path in G that goes through v is preserved in the new v-separable instance. In Section 5 we show the following result.

Theorem 4.2 *Given a v-separable instance I of EDP along with the partition of G into G_1 and G_2, there is a polynomial time algorithm that routes $\Omega(\mathcal{O}/\ln n)$ pairs where $\mathcal{O}$ is the value of an optimum solution of* LP-MCF *on I.*

We now formally describe our approximation algorithm for acyclic graphs based on the above theorem. In the algorithm we use a threshold ℓ that will be specified later in the analysis.

Algorithm Acyclic-DP:

1. Let $S = \emptyset$. $X = \{1, 2, \ldots, k\}$.
2. While there exists an i in X and an $s_i \rightsquigarrow t_i$ path P_i such that $|P_i| \leq \ell$ do **[Greedy Phase]**
   (a) Add i to S and remove i from X.
   (b) Remove the edges of P_i from the graph.
3. Let G' be the residual graph. Initialize MAX $= 0$, $S' = \emptyset$.
4. For each vertex $v \in G'$ do **[v-separable Phase]**
   (a) Approximately solve the v-separable instance induced by v and X in G'. Let S_v be the set of indices of the paths routed through v.
   (b) If (MAX $< |S_v|$) set MAX $= |S_v|$ and set $S' = S_v$.
5. Output $S \cup S'$.

There are two phases in the algorithm. Step 2 implements a *greedy* phase in which we remove paths as long as they are *short*, that is, of length at most ℓ. S is the set of pairs routed in this phase. Step 4 implements the second phase that solves, for each vertex $v \in V$, a v-separable instance induced by v in the residual graph G'. We pick the vertex that results in the most number of paths routed. S' is the set of pairs routed in this phase. We now analyze the performance of the above algorithm.

Let O be the set of pairs routed in some fixed optimal solution. Let O_u be the set of pairs in O but not in $S \cup S'$. We partition O_u into two sets O_1 and O_2 as follows. A pair from O_u is in O_1 if and only if its path in the optimal solution intersects the path of some pair in S. It is easily seen that $|O_1| \leq \ell \cdot |S|$ since the path for a pair in S is at most ℓ edges long. Recall from the algorithm description that G' is the residual graph after edges in the paths used to route S are removed. The following claim is easy to see.

Claim 4.3 *All the pairs in O_2 are routed using only the edges in $E(G')$ and the path assigned to each pair is at least ℓ long.*

Using the above claim and a simple averaging argument, we get the following: there exists a vertex u such that at least $\ell|O_2|/n$ paths from O_2 go through u in G'. Note that $O_2 \subset X$ where X is the set of unrouted pairs at the beginning of Step 4. Hence, in the u-separable instance induced by u and X in G', at least $\ell|O_2|/n$ paths from O_2 can be routed. Therefore, from Theorem 4.2, the algorithm can route $\Omega(\ell|O_2|/(n \ln n))$ pairs.

Theorem 4.4 *There is an $O(\sqrt{n \ln n})$ approximation for the edge disjoint path problem in acyclic graphs. Furthermore, the LP integrality gap is $O(\sqrt{n \ln n})$.*

Proof: From the arguments above, $|O_1| \leq \ell \cdot |S|$ and $|O_2| \leq cn \ln n |S'|/\ell$ for some sufficiently large constant $c \geq 1$. Choosing ℓ to be $\sqrt{n \ln n}$, we obtain that $|O_u| = |O_1| + |O_2| \leq c\sqrt{n \ln n}(|S| + |S'|)$. Hence we obtain the desired approximation. In the above algorithm the greedy phase is not based on an LP solution. However, it is possible to argue along the same lines as in the proof of Theorem 3.5 to establish a bound of $O(\sqrt{n \ln n})$ on the LP integrality gap. We omit the details. ■

5 Edge disjoint paths in a v-separable instance

In this section we prove Theorem 4.2. We use properties of the *single* source flow problem in an important way. We start with a preliminary lemma that guarantees a near optimal feasible solution to LP-MCF with some additional properties.

Lemma 5.1 *For a given instance I of EDP let $(x_1, \ldots, x_k, f_{\pi_1}, f_{\pi_2}, \ldots)$ be a feasible fractional solution to* LP-MCF *of value $\mathcal{O} = \sum_i x_i$. If $\mathcal{O} \geq \gamma$ for a sufficiently large constant γ, then for some absolute positive constants α, β, there is a feasible solution $(y_1, \ldots, y_k, g_{\pi_1}, g_{\pi_2}, \ldots)$ to* LP-MCF *such that:*

- *for $1 \leq i \leq k$, $y_i = 0$ or $y_i \geq \alpha/\ln n$, and*
- *$\mathcal{O}' = \sum_i y_i \geq \beta\mathcal{O}$.*

Moreover, such a solution can be computed in polynomial time.

Proof: We alter $(\bar{x}, \bar{f})$ to obtain $(\bar{y}, \bar{g})$. We first decompose $(\bar{x}, \bar{f})$ into two feasible solutions $(\bar{x}^1, \bar{f}^1)$ and $(\bar{x}^2, \bar{f}^2)$ as follows. For $1 \leq i \leq k$ let $x_i^1 = \lfloor x_i \ln n \rfloor / \ln n$ and $x_i^2 = x_i - x_i^1$. For each $\pi \in \mathcal{P}_i$ such that $f_\pi > 0$ we set $f_\pi^1 = (x_i^1/x_i)f_\pi$ and $f_\pi^2 = f_\pi - f_\pi^1$. We consider two cases. If $\sum_i x_i^1 \geq \sum_i x_i/2$ then we observe that $(\bar{x}^1, \bar{f}^1)$ is the desired solution $(\bar{y}, \bar{g})$ and satisfies the properties of the lemma with $\alpha = \beta = 1/2$.

Otherwise, we have that $\sum_i x_i^2 \geq \sum_i x_i/2$. We now obtain $(\bar{y}, \bar{g})$ from $(\bar{x}^2, \bar{f}^2)$ using randomized rounding with a scaling factor. For each i we independently pick at most one path from

$\mathcal{P}_i$ such that probability of picking a path $\pi \in \mathcal{P}_i$ is exactly $\ln n \cdot f_\pi^2$. Note that the probability of picking a path for i is exactly $\ln n \cdot x_i^2$. If we do not pick a path for i, we set $y_i = 0$. Otherwise, letting π_i denote the path picked for i, we set $y_i = g_{\pi_i} = 1/\ln n$. We observe that y_i is a random variable and $\mathrm{E}[y_i] = x_i$. Let ρ_1, ρ_2 be absolute constants to be specified later. We claim that the random solution $(\bar{y}, \bar{g})$ satisfies the following two properties with high probability: (i) $\sum_i y_i \geq \rho_1 \sum_i x_i^2$ and (ii) for each edge e, $\sum_{\pi \ni e} g_\pi \leq \rho_2$. Although the above procedure is randomized, it can be derandomized using pessimistic estimators [20] to obtain a deterministic solution that satisfies the two properties.

We now prove that the random solution satisfies the two properties claimed above. Let A_i be a binary random variable that is 1 if a path is picked for pair i and 0 otherwise. From the rounding procedure, $\mathrm{E}[A_i] = \Pr[A_i] = \ln n \cdot x_i^2$. Letting $A = \sum_i A_i$ we have that $\mathrm{E}[A] = \ln n \sum_i x_i^2 \geq \ln n \cdot \mathcal{O}/2 \geq \ln n \cdot \gamma/2$. From Proposition 2.1, $\Pr[A < (1-\delta)\mathrm{E}[A]] < 1/n^{\gamma\delta^2/4}$. Choosing $\delta = 1/2$, $\sum_i y_i = A/\ln n \geq (\sum_i x_i^2)/2 \geq \mathcal{O}/4$ with probability at least $1 - 1/n^{\gamma/8}$. Assuming $\gamma \geq 8$, we can conclude that property (i) above holds with probability at least $1 - 1/n$ for $\rho_1 = 1/2$. For property (ii) above, fix any edge e. Let $x_i^2(e)$ denote the total flow on e for pair i. We let $A_i(e)$ be a binary random variable that is 1 if a path is chosen for i and it contains e. From the rounding procedure it is clear that $\mathrm{E}[A_i(e)] = \Pr[A_i(e)] = \ln n \cdot x_i^2(e)$. Letting $A(e) = \sum_i A_i(e)$, observe that $\mathrm{E}[A(e)] = \ln n \sum_i x_i^2(e) \leq \ln n$, the inequality follows from the feasibility of the solution $(\bar{x}^2, \bar{f}^2)$. Again using Proposition 2.1, for $\delta \geq e^2 - 1$, $A(e) < (1+\delta)\ln n$ with probability at least $1 - 1/n^{1+\delta}$. Note that the total flow on e in the random solution is $A(e)/\ln n$, and hence with high probability, the flow on e is at most $1+\delta$. By choosing $\delta = 8$, we ensure that this holds true for all edges e with probability at least $1 - 1/n$. Thus property (ii) above holds with high probability with $\rho_2 = 9$.

Assuming that properties (i) and (ii) are satisfied, we can scale down $(\bar{y}, \bar{g})$ by ρ_2 to obtain a feasible solution that satisfies the requirements of the lemma with $\alpha = 1/\rho_2$ and $\beta = 1/(4\rho_2)$. ∎

Single source flows: Given a directed (undirected) graph with a single source s and terminals $u_1, \ldots, u_k$, finding the maximum number of disjoint paths from v to the terminals is easy. We can reduce it to the standard source-sink flow by connecting each terminal u_i to a new sink vertex t by a directed edge of capacity 1 and then computing the maximum flow from s to t. An integral flow is guaranteed. Not only is the flow integral, the matrix describing the s-t flow problem is totally unimodular (see [16, 22]), hence we also obtain decomposition properties; we state a lemma in a form that will be useful to us.

Lemma 5.2 *Let G be a unit-capacity directed graph with a source vertex v and distinct terminals $u_1, u_2, \ldots, u_k$. Let $\bar{f}$ be a feasible flow vector in which each terminal u_i, $1 \leq j \leq k$ gets a flow of y_j, $0 \leq y_j \leq 1$ from v such that $\rho = \sum_j y_j$ is an integer. Then there exists an $\ell \leq m$ and for $1 \leq i \leq \ell$, a positive real α_i and an edge disjoint path collection F_i from v to a subset of terminals, with the following properties.*

- $\sum_{1\le i\le \ell} \alpha_i = 1$, *and*

- *for each edge* e, f_e *the total flow on* e *in* $\bar{f}$ *satisfies* $f_e = \sum_{i:F_i \ni e} \alpha_i$, *and*

- *for each terminal* u_j, $\sum_{i:u_j\in F_i} \alpha_i = y_j$.

Such a decomposition can be computed in polynomial time.

Proof: We connect the terminals to a sink vertex t by directed edges of capacity 1 and then there is an v-t flow of value ρ. Let e_j be the edge connecting u_j to t. We can send flow of value ρ from v to t where the flow on each e_j is exactly y_j. Since the flow value is integral and all capacities are 1, this fractional v-t flow can be decomposed into a convex combination of *integral flows*, each of value ρ, such that each edge e that has flow f_e in $\bar{f}$ is used in the convex combination to f_e. Each integral flow from v to t of value ρ defines a disjoint path collection from v to t and hence from v to the terminals. We put u_j in the path collection corresponding to an integral flow if e_j is used in that flow. The lemma follows. ■

Our algorithm for a v-separable instance is based on the following observation that establishes the connection to single source flows.

Proposition 5.3 *Let* $(\bar{x}, \bar{f})$ *be a feasible solution to* LP-MCF *for a* v*-separable instance* I. *Then it defines a feasible solution to two separate single source flow problems:* I_1 *with* v *and* $s_1, s_2, \ldots, s_k$ *in* G_1, *and* I_2 *with* v *and* $t_1, t_2, \ldots, t_k$ *in* G_2. *The flow value in* I_1 *from* s_i *to* v *and in* I_2 *from* v *to* t_i *is* x_i.

Algorithm v-sep:

1. Solve LP-MCF on I to obtain a fractional solution $(\bar{x}, \bar{f})$ with value $\mathcal{O}$.

2. Convert $(\bar{x}, \bar{f})$ to a solution $(\bar{y}, \bar{g})$ with value $\mathcal{O}'$ and the properties described in Lemma 5.1.

3. Let $\bar{z}$ and and $\bar{w}$ be the feasible fractional solutions to the single source flow problems induced by $(\bar{y}, \bar{g})$ in G_1 and G_2 respectively. Decompose the single source flow in each instance separately to obtain convex combinations $(\bar{\alpha}, \bar{F})$ and $(\bar{\beta}, \bar{H})$ of disjoint path collections, as guaranteed by Lemma 5.2.

4. From $\bar{F}$ pick a random path collection F_a according to the distribution $\bar{\alpha}$. Independently, from $\bar{H}$, pick a random path collection H_b according to the distribution $\bar{\beta}$.

5. If s_i is connected to v in F_a and v is connected to t_i in H_b, connect s_i and t_i by joining the the two paths into a single path. Otherwise the pair (s_i, t_i) is not connected.

Since $F_a \subseteq E_1$ and $H_b \subseteq E_2$ and $E_1 \cap E_2 = \emptyset$, it is easy to see that a feasible solution is generated for the original instance I.

Lemma 5.4 *The probability that a pair (s_i, t_i) with positive flow y_i in the fractional solution $(\bar{y}, \bar{g})$ is connected by a disjoint path is y_i^2.*

Proof: Consider the convex decomposition $(\bar{\alpha}, \bar{F})$ in the single source instance in G_1. We pick a random disjoint path collection from the decomposition. Lemma 5.2 guarantees that s_i will be in the disjoint path collection with probability y_i, the value of the flow to v. By exactly the same reasoning, v will be connected to t_i with probability y_i. Since the two events are independent, with probability y_i^2, s_i and t_i will be connected by a disjoint path via v. ■

Hence the expected number of pairs that will be connected is $\sum_i y_i^2$. However, $\mathcal{O}' = \sum_i y_i$, and from Lemma 5.1, $\mathcal{O}' = \Omega(\mathcal{O})$ and if $y_i > 0$ then $y = \Omega(1/\ln n)$. Therefore the expected number of pairs that are connected is $\Omega(\mathcal{O}/\ln n)$. $\mathcal{O}$ is an upper bound on the integer optimal solution and this completes the proof of Theorem 4.2. We note that the algorithm can be easily derandomized by trying all possible tuples (F_a, H_b) with F_a from $\bar{F}$ and H_b from $\bar{H}$ and picking the tuple that maximizes the number of pairs that are connected. There are only a polynomial number of such pairs.

6 Unsplittable Flow

The *unsplittable flow problem* (UFP) is a generalization of the maximum edge disjoint path problem. We are given an edge-capacitated graph $G = (V, E)$ with c_e denoting the capacity of edge $e \in E$. Each pair (s_i, t_i) also specifies a demand ρ_i which needs to be routed on a single path. The objective is to find the maximum number of pairs that can be routed without violating edge capacities. Azar and Regev [1] showed that UFP in directed graphs is hard to approximate to within an $\Omega(m^{1-\varepsilon})$ factor. When expressed as a function of n, their result shows an $\Omega(n^{1-\varepsilon})$-hardness. We obtain interesting and useful special cases of UFP by imposing additional restrictions. An instance is said to satisfy the *no-bottleneck* assumption if $\rho_{\max} = \max_i \rho_i \leq c_{\min} = \min_e c_e$; that is, the maximum demand is at most the minimum capacity. The uniform capacity UFP (UCUFP for short) imposes a stronger constraint that all edges have the same capacity, say C. Observe that both these special cases generalize EDP and hence are at least as hard. The best known approximation ratio for UFP with the no-bottleneck assumption is $O(\sqrt{m})$ [23, 2]. This is also the best known ratio for UCUFP.

We now observe that the approximability bounds for EDP obtained in Theorems 3.3 and 4.4 apply to UCUFP as well. Given an instances, we partition the pairs into two sets S_1 and S_2 where S_1 contains a pair i iff $\rho_i \leq C/2$, and S_2 contains the remaining pairs. Baveja and Srinivasan [2], and Azar and Regev [1] give an $O(\sqrt{n})$-approximation for instances of UFP that satisfy the property that $\rho_{\max} \leq c_{\min}/2$. Thus we can obtain an $O(\sqrt{n})$-approximation for the instance induced by pairs in S_1. Note that any two pairs in S_2 that are routed cannot share an edge in their paths. Thus pairs in S_2 induces an instance of EDP for which we can apply the bounds in Theorems 3.3 and 4.4. We thus obtain the desired algorithm for UCUFP by taking the better of the solutions for S_1 and S_2.

7 Conclusions

We obtained improved approximation ratios for EDP and UFP in both undirected and directed graphs. In particular we have obtained a near optimal bound for EDP in acyclic graphs. The complexity of EDP for general graphs remains wide open for undirected graphs. For directed graphs there is still a gap between the $O(\min(n^{4/5}, \sqrt{m}))$ bound that we prove and the hardness factor of $\Omega(n^{1/2-\varepsilon})$. Subsequent to our work, Varadarajan and Venkataraman [24] have improved the upper bound for directed graphs to $O(\min(n^{2/3}\ln^{1/3} n, \sqrt{m}))$; they establish this bound for Greedy by improving Theorem 3.2. Building on our work, Kolman [15] analyzed a variant of Greedy and obtained bounds for UFP that match our bounds for UCUFP. The LP relaxation is a potential way to resolve Conjecture 1.1. In some recent work [3], an $O(\sqrt{n})$ approximation and integrality gap for EDP and UFP have been obtained for undirected graphs and directed acyclic graphs.

Acknowledgments: We thank Anupam Gupta and Bruce Shepherd for useful discussions and clarifications. We are grateful to Sanjeev Arora for pointing out an idea in the proof of Lemma 5.1 that improved the ratio from $O(\sqrt{n}\log n)$ to $O(\sqrt{n \log n})$.

References

[1] Y. AZAR AND O. REGEV. Combinatorial algorithms for the unsplittable flow problem. *Algorithmica*, 44(1):49–66, 2006. Preliminary version in *Proc. of 8th IPCO*, 2001.

[2] A. BAVEJA AND A. SRINIVASAN. Approximation algorithms for disjoint paths and related routing and packing problems. *Mathematics of Operations Research*, Vol. 25, pp. 255–280, 2000.

[3] C. CHEKURI, S. KHANNA, AND F. B. SHEPHERD. An $O(\sqrt{n})$ Approximation and Integrality Gap for Disjoint Paths and Unsplittable Flow. *Theory of Computing*, Vol 2, 137–146, 2006.

[4] J. CHERIYAN, H. KARLOFF, AND Y. RABANI. Approximating directed multicuts. *Combinatorica*, 25(3):251–269, 2005. Preliminary version in *Proc. of IEEE FOCS*, 2001.

[5] S. EVEN AND R. E. TARJAN. Network flow and testing graph connectivity. *SIAM Journal on Computing*, Vol. 4, 1975, pp. 507-518.

[6] S. FORTUNE, J. HOPCROFT AND J. WYLLIE. The directed subgraph homeomorphism problem. *Theoretical Computer Science*, 10(2):111–121, 1980.

[7] A. FRANK. Packing paths, cuts, and circuits - a survey. In B. Korte, L. Lovász, H. J. Prömel, and A. Schrijver, eds., *Paths, Flows and VLSI-Layout*, 49–100. Springer Verlag, Berlin, 1990.

[8] M. R. GAREY AND D. S. JOHNSON. *Computers and Intractability: A Guide to the Theory of NP-completeness*. Freeman, 1979.

[9] N. GARG AND J. KONEMANN. Faster and Simpler Algorithms for Multicommodity Flow and other Fractional Packing Problems. In *Proc. of IEEE FOCS*, 300–309, 1998.

[10] N. GARG, V. VAZIRANI AND M. YANNAKAKIS. Primal-dual approximation algorithms for integral flow and multicut in trees. *Algorithmica*, 18, pp. 3–20, 1997. Preliminary version in *Proc. of ICALP*, 1993.

[11] V. GURUSWAMI, S. KHANNA, B. SHEPHERD, R. RAJARAMAN, AND M. YANNAKAKIS. Near-Optimal Hardness Results and Approximation Algorithms for Edge-Disjoint Paths and Related Problems. *Journal of Computer System Sciences*, 67(3): 473–496, 2003. Preliminary version in *Proc. of ACM STOC*, 1999.

[12] J. M. KLEINBERG. Approximation algorithms for disjoint paths problems. PhD thesis, MIT, May 1996.

[13] S. G. KOLLIOPOULOS. Exact and Approximation Algorithms for Disjoint-Path Problems. PhD Thesis, Dept. of Computer Science, Dartmouth College, 1998.

[14] S. G. KOLLIOPOULOS AND C. STEIN. Approximating disjoint-path problems using greedy algorithms and Packing Integer Programs. *Math. Programming A*, (99):63–87, 2004. Preliminary version in *Proc. of IPCO*, 1998.

[15] P. KOLMAN. A Note on the Greedy Algorithm for the Unsplittable Flow Problem. *Information Processing Letters*, 88(3):101–105, 2003.

[16] E. L. LAWLER. Combinatorial Optimization: Networks and Matroids. Holt, Rinehart and Winston, 1976.

[17] B. Ma and L. Wang. On the inapproximability of disjoint paths and minimum steiner forest with bandwidth constraints. *Journal of Computer and Systems Sciences*, vol 60, pages 1–12, 2000.

[18] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, New York, 1995.

[19] P. Raghavan and C. D. Thompson. Randomized rounding: A technique for provably good algorithms and algorithmic proofs. *Combinatorica*, Vol. 7, pp. 365–374, 1987.

[20] P. Raghavan. Probabilistic Construction of Deterministic Algorithms: Approximating Packing Integer Programs. *JCSS*, 37(2): 130-143, 1988.

[21] N. Robertson and P. D. Seymour. Outline of a disjoint paths algorithm. In B. Korte, L. Lovász, H. J. Prömel, and A. Schrijver, Eds., *Paths, Flows and VLSI-Layout*. Springer-Verlag, Berlin, 1990.

[22] A. Schrijver. Theory of Linear and Integer Programming. John Wiley and Sons, 1986.

[23] A. Srinivasan. Improved approximations for edge-disjoint paths, unsplittable flow, and related routing problems. *Proc. of the IEEE FOCS*, pp. 416–425, 1997.

[24] K. Varadarajan and G. Venkataraman. Graph Decomposition and a Greedy Algorithm for Edge-Disjoint Paths. *Proc. of ACM-SIAM SODA*, 379–380, 2004.

[25] D. B. West. Introduction to Graph Theory. Prentice-Hall, Inc., 1996.