Donation Center Location Problem

Chien-Chung Huang[*] Zoya Svitkina[†]

October 11, 2009

Abstract

We introduce and study the *donation center location* problem, which has an additional application in network testing and may also be of independent interest as a general graph-theoretic problem. Given a set of agents and a set of centers, where agents have preferences over centers and centers have capacities, the goal is to open a subset of centers and to assign a maximum-sized subset of agents to their most-preferred open centers, while respecting the capacity constraints.

We prove that in general, the problem is hard to approximate within $n^{1/2-\epsilon}$ for any $\epsilon > 0$. In view of this, we investigate two special cases. In one, every agent has a bounded number of centers on her preference list, and in the other, all preferences are induced by a line-metric. We present constant-factor approximation algorithms for the former and exact polynomial-time algorithms for the latter. Of particular interest among our techniques are an analysis of the greedy algorithm for a variant of the maximum coverage problem called *frugal coverage*, the use of maximum matching subroutine with subsequent modification, analyzed using a counting argument, and a reduction to the independent set problem on *terminal intersection graphs*, which we show to be a subclass of trapezoid graphs.

1 Introduction

Suppose that a charitable organization wishes to open a number of locations where people can make donations (e.g. donate blood). There is no cost for opening these centers, but they do have capacities for the number of donors that they can accommodate. We model the potential donors, whom we call agents, as each having a list of locations where she would be willing to go to make a donation. Once some of the centers are opened, each agent goes to the most convenient open one from her list. However, if that center is full (i.e. has exceeded its capacity), then the agent gives up and decides not to donate at all. Our goal is to choose a set of centers to open to maximize the number of collected donations.

Formally, we define the DONATION CENTER LOCATION (DCL) problem as follows. Let $G = (A \cup L, E)$ be a directed bipartite graph, with edges directed from the set A of agents to the set L of donation centers. Every center $l \in L$ has a capacity $c_l \in \mathbb{Z}^+$, and every vertex $a \in A$ has a strictly-ordered preference ranking of its neighbors in L (or, equivalently, of its outgoing edges). These preferences model either distance or some other measure of convenience for the agents over the locations. We have to choose a subset $L' \subseteq L$ of centers to open, and to assign a subset $A' \subseteq A$

[*]Max-Planck-Institut für Informatik, Saarbrücken, Germany
[†]Department of Computing Science, University of Alberta, Canada

of agents to centers in L', in such a way that the number of agents assigned to any center $l \in L'$ is at most c_l, and each $a \in A'$ is assigned to its highest-ranked neighbor in L'. The goal is to maximize $|A'|$, the number of assigned agents. Note that once a set L' of locations is selected, it is very easy to find an optimal assignment of agents: if some open center $l \in L'$ is the first choice of more than c_l agents, then an arbitrary subset of c_l of them can be assigned to it (and others will remain unassigned). Thus, our problem statement requiring an explicit assignment from A' to L' is equivalent to one motivated above, which just asks to find L' and assumes that each center $l \in L$ will accommodate an unspecified subset of at most c_l agents who prefer it.

We use notation $l \succ_a l'$ to indicate that agent $a \in A$ prefers center l to center l', where both (a, l) and (a, l') are edges in E. If a solution assigns agent $a \in A'$ to center $l \in L'$, then we write $\mu(a) = l$. We also define $\mu^{-1}(l) = \{a \in A' : \mu(a) = l\}$ to be the set of agents assigned to l. If an assignment μ satisfies the constraints of the DCL problem, then we call it *valid*. Formally, a valid assignment $\mu : A' \to L'$ meets the following conditions:

1. if $a \in A'$, then $(a, \mu(a)) \in E$
2. if $a \in A'$, then there is no $l \in L'$ such that $(a, l) \in E$ and $l \succ_a \mu(a)$
3. if $l \in L'$, then $|\mu^{-1}(l)| \leq c_l$

One special case of DCL that we focus on is the *unit-capacity* case, where $c_l = 1$ for all centers. In that case the assignment $\mu : A' \to L'$ is a matching. This special case establishes a connection between DCL and various matching problems under preferences that have been extensively studied in both computer science and economics literature. It also has an application in network testing [8, 24], which is as follows. In a wireless network consisting of transmitters and receivers, the transmitter nodes have to be tested. For one round of testing, a maximum-cardinality set A' of transmitters has to be matched to a set L' of receivers. The power setting of a transmitter is adjusted based on the distance to its intended receiver, and the signal reaches this receiver as well as all receivers that are closer to the transmitter than it is. The preference lists of transmitters over receivers are complete and are induced by the distance, with closer ones ranked higher. Then Condition 2 for a valid matching requires that a matched receiver not simultaneously be in the range of two active (matched) transmitters, thus preventing interference.

1.1 Related work

Matching entities with preferences is an extensively studied topic in the literature. The most representative is the stable matching (also known as stable marriage) problem [11, 13, 16], where both sides have preferences and a matching is considered stable if there are no two elements that both prefer each other to their assigned matches. Recently, the matching problems in the context of one-sided preferences have also been studied. Examples include popular matching [2, 17, 19], rank-maximal matching [15, 20], and pareto-optimal matching [1]. A major distinction in our model is that an unopened center does not influence the feasibility of a given solution, even if some agent prefers this center to his assigned open one. However, for instance, in the stable marriage problem, a bachelor and a married woman can disturb the stability of a matching.

Our model also resembles the well-studied facility location problems [7, 23] and their capacitated versions [21, 26]. However, in most facility location problems, the algorithm is allowed to assign clients to arbitrary opened facilities, whereas in our case, each client has to go to its nearest one. Also, DCL is a maximization problem and does not have a requirement of assigning all clients, whereas facility location is usually formulated as assigning all clients while minimizing cost. Thus, there is no direct way to apply known algorithmic techniques for it to our setting.

Network testing is a possible application of DCL. Maximum-cardinality matching between transmitters and receivers has been studied in [8, 24], where the interference between transmitters is modeled in a more crude way: just the presence of an edge between a transmitter and a receiver in the connection graph represents a possible source of interference. In contrast, we use the notion of preferences (relative physical distance) to give a more fine-grained model of interference.

1.2 Our results and techniques

We study the general DCL problem as well as several special cases of it. Most of the versions that we consider here are NP-complete, in which case we study their approximability, either by finding good approximation algorithms, or by proving hardness of approximation results. However, some of the special cases are solvable in polynomial time, and for these we present exact algorithms. Our results are summarized in Table 1.

	unit capacity	general capacity
complete preferences	$n^{1/2-\varepsilon}$-hard to approximate (§2)	
bounded degree	APX-hard (§2)	
out-degree bound d	$1/d$ (§3.1)	$1/2d$ (§3.3), $1/\phi d$ (§A.3)
out-degree bound $d = 2$	$\frac{e}{e+1}$ (§3.2)	$1/2$ (§A.4)
line metric	polynomial-time (§4)	

Table 1: Summary of results

For the general case of DCL, we show that it is hard to approximate to a factor of $n^{1/2-\varepsilon}$, for any $\varepsilon > 0$. This result also holds for the special case of complete preferences (when G is a complete bipartite graph). In view of this, we focus on two types of special cases, one of which is the bounded degree case. Here the degree in G of any vertex $a \in A$ is upper-bounded by a constant d. We show that the problem remains APX-hard, even in the unit-capacity case with degree bound of 2. For any degree bound d, we give a $1/d$-approximation algorithm for the unit-capacity case. For the special case of degree bound $d = 2$, we improve this ratio to $\frac{e}{e+1} \approx 0.731$. To do this, we introduce a new variant of the maximum coverage problem, called frugal coverage, and analyze the performance of the greedy algorithm on it. For the problem with general capacities and degree bound d, we present a $1/2d$ approximation algorithm that makes use of a maximum matching subroutine. We then improve the analysis to give a $1/\phi d$ approximation, for $\phi \approx 1.618$, and also improve the ratio to $1/2$ for the special case of $d = 2$.

The second special case that we consider is one in which the preferences are induced by a line metric. In particular, all nodes of $A \cup L$ are located on a single line, and each agent ranks the centers in the order of proximity. For this case, we give an exact linear-time algorithm for the unit-capacity setting. Then we extend it to obtain an exact polynomial-time algorithm for general capacities. To design these algorithms, we reduce the problems to maximum independent set on a special kind of graphs that we call terminal intersection graphs. We then show that these graphs form a subclass of trapezoid graphs [5, 6], for which there are known polynomial-time algorithms that solve maximum independent set [10, 18].

2 Hardness results

We prove that DCL is hard to approximate to a factor of $n^{1/2-\epsilon}$, and that it remains APX-hard even in the bounded-degree case. The proof of the first result uses a non-trivial reduction from the maximum independent set problem, which increases the size of the instance while approximately preserving the value of the optimal solution. The proof of APX-hardness uses a different reduction from independent set on 3-regular graphs. Both proofs appear in the Appendix.

Theorem 2.1 *DCL problem is hard to approximate within* $O\left(|A \cup L|^{1/2-\epsilon}\right)$ *for any* $\epsilon > 0$*, unless NP=ZPP. This is true even in the case of unit capacities and complete preferences.*

Theorem 2.2 *DCL problem is APX-hard, even with unit capacities, out-degree bound of 2 on A and in-degree bound of 3 on L.*

We also show that the special case of DCL in which preferences are induced by a metric is no easier than the general problem with complete preferences[1]. In fact, arbitrary complete preferences of A over L can be represented by embedding all points of $A \cup L$ into a metric space. To do this, we use the ℓ_∞ metric over an $|L|$-dimensional space. Each element $l_i \in L$ (for $i = 1$ to $|L|$) is mapped to a location x^i, with coordinates $x^i_j = 0$ for $j \neq i$, and $x^i_i = 1$. Each element $a \in A$ is mapped to a location x^a, with $x^a_i = \frac{1}{2} - \frac{rank(a,l_i)}{2|L|}$. Here $rank(a, l_i)$ is the rank that agent a assigns to center l_i, ranging from 1 for the most-preferred center up to $|L|$. With this embedding, the ℓ_∞ distance from a to l_i, for each $1 \leq i \leq |L|$, becomes $\frac{1}{2} + \frac{rank(a,l_i)}{2|L|}$ (with $|x^i_i - x^a_i|$ being the largest coordinate difference). This ensures that for each $a \in A$, the ordering of elements of L by distance is the same as it is by preference.

3 Algorithms for bounded-degree DCL

In view of the hardness results for the general problem, in this section we focus on special cases in which the lengths of the agents' preference lists are bounded by a constant d.

3.1 A linear-time $1/d$ approximation for the unit-capacity case

We partition L into d subsets $L_1, L_2, \cdots, L_d$. A center l is in group L_k if, among all edges from agents to l, the highest rank of these edges is k. We now consider each set L_k separately, and let μ_k denote an arbitrary matching in which each center in L_k is matched to an agent that ranks it k. Note that at least one such agent for each center must exist by definition of L_k, and no agent will be matched twice as it can't have the same rank for two different centers. We claim that μ_k is a valid matching in the original problem. If not, suppose that both (a, l) and (a', l') are part of μ_k and $l' \succ_a l$. Then a ranks l' higher than k, contradicting the assumption that $l' \in L_k$. We output the largest μ_k, which satisfies $|\mu_k| = |L_k| \geq \frac{1}{d}|L| \geq \frac{1}{d}\textsc{Opt}$, and note that the algorithm can be implemented in linear time.

[1]This reduction was suggested to us by Uri Feige

3.2 A $e/(e+1)$ approximation for unit-capacity DCL with degree bound $d = 2$

Here we consider the unit-capacity case in which the out-degree of each agent is at most 2. Our algorithm in the preceding section gives a ratio of 1/2 in this case, but here we improve it to $\frac{e}{e+1} \geq 0.731$. We first give an approximation-preserving reduction to a problem that we call frugal coverage, and then give a $\frac{e}{e+1}$-factor approximation for frugal coverage. The input to this problem is the same as for set cover, but the objective function is different. We wish to maximize the number of elements covered by the chosen sets *plus* the number of sets that are not chosen.

Definition 3.1 *In the* frugal coverage problem*, the input is a universe U of elements and a collection $\mathcal{C}$ of subsets of U. The goal is to select a subset $\mathcal{C}' \subseteq \mathcal{C}$ that maximizes $|\bigcup_{S \in \mathcal{C}'} S| + |\mathcal{C} \setminus \mathcal{C}'|$.*

Lemma 3.2 *If there is an α-approximation for the frugal coverage problem, then there is an α-approximation for unit-capacity DCL with degree bound 2.*

Proof. To obtain a reduction, we first do one step of pre-processing on the given DCL instance. If any center $l \in L$ has incoming edges of both rank 1 and rank 2, then we remove all its incoming edges of rank 2. We claim that the value of the optimum is maintained, because any feasible solution that uses the edges that were removed can be transformed into one of the same size which does not use these edges. Suppose $a_1 \in A$ ranks l first, and $a_2 \in A$ ranks l second. Now, if a_2 is matched to l in the optimal solution, then a_1 is unmatched, as otherwise it would prefer l to its match. So we can replace the matched pair (a_2, l) with (a_1, l), preserving the size and feasibility of the solution.

Now we give a reduction from the DCL instance to frugal coverage, assuming that no node $l \in L$ has both rank-1 and rank-2 incoming edges. We also assume without loss of generality that there are no nodes in $A \cup L$ with degree zero. We partition the set L into subsets X and Y, where X contains all the nodes with incoming rank-1 edges, and Y contains all the nodes with incoming rank-2 edges. By our assumptions, these sets are disjoint and cover L. For each $l_1 \in X$, we create a set $S(l_1) \in \mathcal{C}$. For each $l_2 \in Y$, we create an element $e(l_2) \in U$. For each agent $a \in A$ whose preference list is of length two, with $l_1 \succ_a l_2$, we include the element $e(l_2)$ into the set $S(l_1)$.

Given a valid matching μ for the DCL instance, we create a solution to the derived frugal coverage instance with value at least $|\mu|$. In particular, this solution $\mathcal{C}'$ consists of all sets $S(l)$ that correspond to unmatched nodes $l \in X$. If we let $|\mu| = x + y$, where x is the number of matches on rank-1 edges, and y is the number of matches on rank-2 edges, then $|\mathcal{C} \setminus \mathcal{C}'| = x$ and $|\bigcup_{S \in \mathcal{C}'} S| \geq y$. The equality follows because $\mathcal{C}$ corresponds to all nodes of X, and $\mathcal{C}'$ corresponds to the unmatched ones. For the inequality, suppose that a rank-2 match (a, l_2) is part of μ, and consider the center l_1 such that $l_1 \succ_a l_2$. Then l_1 is unmatched, as otherwise the feasibility of μ is violated, and therefore $S(l_1) \in \mathcal{C}'$. Also, by construction, $e(l_2) \in S(l_1)$. So we have that for each $l_2 \in Y$ matched on rank-2 edge, there exists $l_1 \in X$ such that $e(l_2) \in S(l_1) \in \mathcal{C}'$, and therefore $|\bigcup_{S \in \mathcal{C}'} S| \geq y$.

Conversely, given a solution $\mathcal{C}'$ to the constructed frugal coverage instance, a feasible solution μ to the original DCL instance, with at least as big a value, can be produced. For each $l \in X$ whose corresponding set is not chosen ($S(l) \notin \mathcal{C}'$), choose an arbitrary node $a \in A$ such that (a, l) is an edge, and include (a, l) in μ. For each $l_2 \in Y$ whose corresponding element is covered by the frugal coverage solution ($e(l_2) \in \bigcup_{S \in \mathcal{C}'} S$), find a node $l_1 \in X$ whose corresponding set covers $e(l_2)$ (i.e. with $e(l_2) \in S(l_1)$ and $S(l_1) \in \mathcal{C}'$), choose a node $a \in A$ such that $l_1 \succ_a l_2$ (which enabled us to include $e(l_2)$ in $S(l_1)$ when constructing the instance), and match a to l_2. To ensure that no $a \in A$ is matched twice, and that μ is a valid matching, suppose that there is a node $a \in A$

with $\mu(a) = l_2$, $l_1 \succ_a l_2$, and l_1 also matched. But this is a contradiction because we only matched (a, l_2) if $S(l_1) \in \mathcal{C}'$, and only matched l_1 if $S(l_1) \notin \mathcal{C}'$. Since for each covered element and for each unchosen set we have included one pair into the matching, the size of μ is at least the objective function value of the frugal coverage solution.

To obtain an α-approximation for DCL, perform the above construction, producing an instance of frugal coverage whose optimum is at least $|\mu^*|$, where μ^* is an optimal valid matching. Find an α-approximation to frugal coverage of value at least $\alpha \cdot |\mu^*|$, and transform it back to a DCL solution with at least as big a value. □

3.2.1 Algorithm for frugal coverage

We analyze the performance of the greedy algorithm for the frugal coverage problem. This is the same algorithm as is used for set cover [25]: while there is a set that covers at least one new element, choose the one that covers maximum number of new elements and include it in the solution. We note that our approximation guarantee for frugal coverage is better than the best possible factor of $\frac{e-1}{e} \approx 0.632$ for the maximum coverage problem [9].

Lemma 3.3 *The greedy algorithm is a $\frac{e}{e+1}$ approximation for the frugal coverage problem.*

Proof. Let $m = |\mathcal{C}|$ be the number of sets in the instance, $n = |U|$ be the total number of elements, and $n' = |\bigcup_{S \in \mathcal{C}} S|$ be the number of elements that are contained in at least one set. Suppose that the greedy algorithm completes after taking l sets. Then its objective function value is equal to $\textsc{Alg} = n' + (m - l)$. Let $\mathcal{C}_k$ denote the intermediate solution obtained by the greedy algorithm after including $0 \leq k \leq l$ sets. We observe that the solution $\mathcal{C}_l$ is at least as good as any $\mathcal{C}_k$, because with each step of the algorithm, the number of unused sets $|\mathcal{C} \setminus \mathcal{C}_k|$ decreases by one, and the number of covered elements $|\bigcup_{S \in \mathcal{C}_k} S|$ increases by at least one. By the same reasoning, we know that there is an optimal solution $\mathcal{C}^* \subseteq \mathcal{C}$ to the frugal coverage problem that covers all elements that are contained in at least one set. Let $k^* = |\mathcal{C}^*|$ be the number of sets chosen by this optimal solution. Then its objective function value is $\textsc{Opt} = n' + (m - k^*)$.

We first give an easy proof to show that the greedy algorithm is at least a $\frac{e-1}{e}$ approximation, and then improve the guarantee. Consider the intermediate greedy solution $\mathcal{C}_{k^*}$ (note that $l \geq k^*$, as k^* is the minimum number of sets that can cover all n' elements). By the guarantee of the greedy algorithm for the maximum coverage problem [9], $\mathcal{C}_{k^*}$ covers at least $\frac{e-1}{e} \cdot n'$ elements. So the value of the solution is $\textsc{Alg} \geq \frac{e-1}{e} \cdot n' + (m - k^*) \geq \frac{e-1}{e} \cdot \textsc{Opt}$.

To improve the guarantee, we observe that $l \leq m$, and therefore $\textsc{Alg} \geq n'$. Combining with the previous result, we get $\textsc{Alg} \geq \max(\frac{e-1}{e} n' + m - k^*, n')$. We now consider two cases. The first case is that $n' \geq \frac{e-1}{e} n' + (m - k^*)$, and therefore $n' \geq e(m - k^*)$. Then

$$\textsc{Alg} \;\geq\; n' \;=\; \frac{en'}{e+1} + \frac{n'}{e+1} \;\geq\; \frac{en'}{e+1} + \frac{e(m-k^*)}{e+1} \;=\; \frac{e}{e+1} \cdot \textsc{Opt}.$$

In the second case, $n' < \frac{e-1}{e} n' + (m - k^*)$, and therefore $m - k^* > n'/e$. Then

$$\begin{aligned} \textsc{Alg} \;&\geq\; \frac{e-1}{e} n' + m - k^* \;=\; \frac{e-1}{e} n' + \frac{m-k^*}{e+1} + \frac{e(m-k^*)}{e+1} \\ &>\; \frac{e-1}{e} n' + \frac{n'}{e(e+1)} + \frac{e(m-k^*)}{e+1} \;=\; \frac{en'}{e+1} + \frac{e(m-k^*)}{e+1} \;=\; \frac{e}{e+1} \cdot \textsc{Opt}, \end{aligned}$$

so in either case we get the desired approximation. □

Combining Lemmas 3.2 and 3.3, we arrive at the following result.

Theorem 3.4 *There is an $\frac{e}{e+1} \geq 0.731$ approximation for unit-capacity DCL with degree bound 2.*

We make two remarks before we close this section. First, by the APX-hardness result of Theorem 2.2, the reduction in Lemma 3.2, and the constant approximation in Lemma 3.3, it follows that the frugal coverage problem is APX-complete. Second, the following special case of DCL is solvable in polynomial time: every agent in A has out-degree at most 2 and every center in L has at most two incoming rank-1 edges. To see this, observe that in this setting, under the reduction of Lemma 3.2, we derive a frugal coverage instance with every set in $\mathcal{C}$ of size at most 2. By the same reasoning as in Lemma 3.3, there is an optimal solution that covers all elements in $\bigcup_{S\in\mathcal{C}} S$. Thus, the problem is equivalent to finding an optimal set cover where every set is of size at most 2 and can be easily shown to be equivalent to the edge cover problem, which is known to be in P [12].

3.3 A $1/2d$ approximation for DCL with general capacities

As the hardness results of Section 2 still apply to the problem with general capacities, we consider the special case in which each agent has at most d outgoing edges in G. Our algorithm consists of the following steps.

1. Using flow techniques, find a maximum-size assignment μ (not necessarily valid) between A and L on the edges of G, where each agent is assigned to at most one center, and each center l gets at most c_l agents. This assignment disregards the preferences of the agents, and serves as an upper bound on the optimum.
2. Create a directed graph on the set of centers $H = (L, F)$ based on μ. An arc $(l, l') \in F$ is drawn if there is some agent that is assigned to center l by μ, but prefers l' to l. If H contains a directed cycle, then update μ by transferring one agent along each arc of this cycle so as to improve the transferred agents' assignments. Update H, and repeat until H is acyclic. Note that this process terminates in polynomial time.
3. Discard all unassigned agents and unused centers from the graph G to produce a subgraph G'. Furthermore, remove from G' edges from each agent a to centers which a ranks lower than $\mu(a)$. Also remove unused centers from H.
4. Define a topological order over H so that all directed arcs of H go "from left to right".
5. Consider each center node l in H, scanning from left to right, and delete it from G' if the degree of l in G' is greater than $\xi \cdot c_l$, where $\xi > 1$ is a parameter to be optimized later. To delete l, update G' by removing l and the agents assigned to it by μ, along with the incident edges.
6. Return the set U of centers that are still part of G'.

Note that the final solution is not μ, as μ is not necessarily a valid assignment. Instead, it is the set $U \subseteq L$ of open centers, with the best valid assignment of agents to them, which is easy to find as mentioned in the introduction. The possible loss in value of this solution compared to the size of μ is analyzed below.

Theorem 3.5 *The above algorithm is a $1/2d$-approximation for DCL with degree bound d.*

Proof. As mentioned, the number of agents assigned by μ serves as an upper bound on the optimum. Moreover, step 2 does not alter the size of μ. There are two ways in which the algorithm can lose agents that are matched in step 1. The first is the deletion of centers in step 5, as agents assigned to them may not have any edges to the remaining centers, and thus be lost to the solution. The second reason is that even from centers in U, the contribution to the objective function may be smaller than the number of agents assigned to them by μ. This is because the agents 'switch' from their assigned centers to their best centers in U. As a simple example, consider an instance with two agents and two centers, where both agents prefer l_1 to l_2, and $c_{l_1} = c_{l_2} = 1$. Then μ assigns one agent to each center, and has size two. But opening both centers produces a solution with objective function of 1.

We let $|\mu| = n_u + n_r$, where n_u is the number of agents that are assigned by μ (after step 2) to centers in U, and n_r is the number of agents assigned by μ to other centers, i.e. ones removed by the algorithm in step 5. We first lower-bound n_u, and then lower-bound the size of the solution in terms of n_u. Observe that for every center l deleted in step 5, its degree in G' (at the time of deletion) is greater than ξc_l. At most c_l of these incoming edges come from agents assigned to it by μ, and the rest come from agents that are assigned elsewhere by μ, but prefer l to their current centers (this is because in step 3, we removed edges from each agent a to centers that rank lower than $\mu(a)$). Let us say that one such agent, a, is assigned to a center l' but prefers l to l'. In this case the graph H would contain an edge from l' to l, which means that l' occurs before l in the topological ordering. Furthermore, when l' was considered by step 5 of the algorithm (which happened before l was considered), it was not deleted, since otherwise we would have also deleted all its agents, including a. So any such center l' must be part of U. Now, each agent has at most $d-1$ edges in G' to centers other than its assigned one, so the number of agents assigned to U by μ that contribute the extra $\xi c_l - c_l$ edges to centers $l \notin U$ can be bounded as

$$n_u \;\geq\; \frac{\sum_{l\notin U}(\xi c_l - c_l)}{d-1} \;=\; \frac{\xi-1}{d-1}\cdot\sum_{l\notin U} c_l \;\geq\; \frac{\xi-1}{d-1}\cdot n_r. \tag{1}$$

The value of the final solution that assigns agents from A to centers in U in an optimal way can only be higher than if we restrict the assignment to only use agents from some subset $\tilde{A} \subseteq A$. In particular, let $\tilde{A}$ be the set of n_u agents that are assigned to U by μ. For a center $l \in U$, consider the set of agents $\tilde{A}_l \subseteq \tilde{A}$ that rank l highest among centers in U. For any such agent $a \in \tilde{A}_l$, there is an edge in G' from a to l. But since the degree of l in G' is at most ξc_l, the size of $\tilde{A}_l$ is at most ξc_l. Thus, at least a $1/\xi$ fraction of agents in $\tilde{A}_l$ can be assigned to l by a valid assignment. As the sets $\tilde{A}_l$ partition $\tilde{A}$, overall $\textsc{Alg} \geq n_u/\xi$. Using (1), the approximation ratio becomes

$$\frac{\textsc{Alg}}{\textsc{Opt}} \;\geq\; \frac{n_u/\xi}{|\mu|} \;=\; \frac{n_u/\xi}{n_u+n_r} \;\geq\; \frac{n_u/\xi}{n_u + n_u\frac{d-1}{\xi-1}} \;=\; \frac{1/\xi}{1+\frac{d-1}{\xi-1}} \;\equiv\; f(\xi).$$

Calculus shows that $f(\xi)$ is maximized if we set $\xi = 1+\sqrt{d-1}$, and the approximation guarantee becomes $1/(d+2\sqrt{d-1}) \geq 1/2d$. In fact, for large d, it approaches $1/d$. □

With a more detailed analysis (see Section A.3), the above algorithm can be shown to deliver a $1/\phi d$ approximation, for $\phi \approx 1.618$. In addition, for the special case of $d = 2$, another algorithm with an improved guarantee of $1/2$ appears in Section A.4.

Theorem 3.6 *There is a $\frac{1}{2}$-approximation for DCL with degree bound 2.*

4 DCL on a line

In this section we show how to find optimal solutions to unit-capacity and general DCL, in the case that preferences are complete and defined according to distances on a line (with closer points ranked higher). Our algorithms work through a reduction to the independent set problem on a special class of graphs, which we call terminal intersection graphs. As we show, terminal intersection graphs are a subclass of trapezoid graphs [5, 6], for which polynomial-time algorithms for independent set are known. We assume that no two nodes are co-located on the line, and no two distances are equal. Distance between two points on the line is denoted by $d(x, y)$.

Definition 4.1 *A graph $H = (W, F)$ is a* terminal intersection graph *if there exists a set of intervals $\mathcal{I} = \{I_w = [a_w, b_w] \ : \ w \in W\}$ on a line, each with a terminal $c_w \in I_w$, such that there is an edge $(w, w') \in F$ if and only if either $c_w \in I_{w'}$ or $c_{w'} \in I_w$.*

Definition 4.2 *A graph $H = (W, F)$ is a* trapezoid graph *if there exist two parallel lines such that each vertex $w \in W$ corresponds to a trapezoid T_w defined by the convex hull of two points on the top line and two points on the bottom line, and $(w, w') \in F$ if and only if T_w and $T_{w'}$ intersect.*

Lemma 4.3 *Every terminal intersection graph is a trapezoid graph, and the trapezoid model can be found in linear time when a terminal intersection model is given.*

Proof of Lemma 4.3 appears in the Appendix. We now give the main results of the section.

Theorem 4.4 *Unit-capacity DCL on a line can be solved to optimality in linear time.*

Proof. We reduce to the independent set problem on terminal intersection graphs, which can be solved in linear time using Lemma 4.3 and the algorithm for independent set on trapezoid graphs ([18] and the fact that trapezoid graphs are a subclass of co-comparability graphs [10]). Our reduction is also linear-time, so this gives an overall $O(n)$-time algorithm.

We start with a useful observation about the structure of an optimal valid matching on a line. We claim that for any instance, there exists an optimal solution in which every matched center is matched either to its closest agent to the right of it on the line, or to its closest agent to the left. To see this, consider a center l, its closest agent to the right a_1, and an agent a_2 farther to the right. Suppose that l and a_2 are matched, and the distance between them is $d(l, a_2)$. Then there is no other matched center between l and a_2, and there is no matched center within distance $d(l, a_2)$ to the right of a_2 (otherwise a_2 would prefer those centers to l). This implies that a_1 is not matched, as otherwise it would prefer l to its match. So if, instead of (a_2, l), we match (a_1, l), this would also be a feasible solution, since l would be the closest matched center to a_1.

Given the above observation, it suffices to only consider two possible matches for each center l: (a_l, l) and (a_r, l), where a_l and a_r are the nearest agents to the left and to the right, respectively. So there are at most $2|L|$ possible matches to consider, and we have to choose the maximum subset of them which matches each node at most once and fulfills the condition of a valid matching. We do this by creating a terminal intersection graph $H = (W, F)$ in which the nodes correspond to possible matches, and edges correspond to pairs of matches that interfere with each other. Then the maximum independent set in H corresponds to the maximum valid matching in our instance.

For each center $l \in L$ and its potential match a_r on the right we create a vertex w in H specified by the interval $I_w = [l, \, l + 2 \cdot d(l, a_r)]$ and the terminal $c_w = l$ (we identify the nodes in $A \cup L$ with their coordinates on the line, and hence treat them as numbers). Similarly, for l and a_l we create a node with interval $[l - 2 \cdot d(a_l, l),\, l]$ and terminal l. This is an interval that is twice as long as the distance between the agent and the center, centered at the agent, and with a terminal at the endpoint corresponding to the center.

We now verify that two vertices have an edge in H if and only if both of their corresponding pairs cannot be included in the matching. Suppose that H contains an edge (w, w'). Then either $c_w \in I_{w'}$ or $c_{w'} \in I_w$, so assume without loss of generality that $c_w \in I_{w'}$. Let (a_w, l_w) be the potential match corresponding to the vertex w, and $(a_{w'}, l_{w'})$ be one corresponding to w'. Then, by geometry, $d(a_{w'}, l_w) \leq d(a_{w'}, l_{w'})$. So either $l_w = l_{w'}$, or $a_{w'}$ is closer to l_w than to $l_{w'}$, and both pairs cannot be matched simultaneously. Conversely, suppose that two pairs (a_w, l_w) and $(a_{w'}, l_{w'})$ cannot both participate in the matching. This could be because $l_w = l_{w'}$, or $a_w = a_{w'}$, or because they violate the preference condition. In the first case, immediately $c_w \in I_{w'}$, so the edge (w, w') is in H; in the second case, assume l_w is closer than $l_{w'}$ to a_w, but then $c_w = l_w \in I_{w'}$; in the last case, assume that $d(a_{w'}, l_w) < d(a_{w'}, l_{w'})$. But since $I_{w'}$ is an interval of length $2d(a_{w'}, l_{w'})$ centered at $a_{w'}$, it includes $c_w = l_w$, so again (w, w') is an edge in H. □

Theorem 4.5 *DCL on a line can be solved to optimality in polynomial time.*

We sketch the ideas for extending the unit-capacity algorithm to general capacities. The running time is no longer linear, but it remains polynomial. We again construct a terminal intersection graph H, but this time we reduce to the maximum *weighted* IS problem on it, which is still solvable in polynomial time for trapezoid graphs [10]. Consider a center u with $c_u \leq n$. Any feasible solution assigns k_l agents to it that are to the left of u on the line, and k_r agents that are on the right, for some k_l and k_r with $k_l + k_r \leq c_u$. Analogously to the proof of Theorem 4.4, we can assume that these agents are the closest k_l ones on the left, and the closest k_r ones on the right. So for each center u, and for each possible k_l and k_r, we create an interval I_u with a terminal located at u. To specify the endpoints of I_u, we let a_l be the k_l-th farthest agent to the left of u, and a_r be the k_r-th agent to the right of u. Then $I_u = [u - 2 \cdot d(a_l, u),\, u + 2 \cdot d(u, a_r)]$. Finally, we set the weight of the corresponding vertex in H to $k_l + k_r$. As before, it can be verified that a set of vertices in H is independent if and only if the corresponding assignment in the original problem is valid. Moreover, the weight of this set is equal to the number of agents assigned in the corresponding solution.

5 Conclusion

We have introduced a new combinatorial problem with a number of applications and made significant progress toward characterizing its complexity and approximability. In doing so, we used a variety of techniques, including a non-trivial hardness proof, an analysis of the greedy algorithm for a new variant of set cover, a counting argument for establishing the approximation ratio in the general capacity case, and a reduction to geometric graphs. Our definitions of the frugal coverage problem and the terminal intersection graphs, as well as our algorithms, may be of more general interest and find applications in other contexts.

One extension of the DCL problem is the non-bipartite version, where G is a general directed graph, and all vertices have preferences over their outgoing edges. A solution consists of sets A', L', and a valid assignment from A' to L' as before, but now A' and L' can be arbitrary disjoint subsets of vertices of G. In the network testing application, the bipartite problem corresponds to the case that transmitters and receivers are two different types of devices, whereas the non-bipartite version models a more general setting in which some or all of devices are capable of performing either function. Our hardness of approximation results extend to the non-bipartite version, and in the unit-capacity setting it admits a $1/d$ approximation for the bounded-degree case as well as a polynomial-time exact algorithm on a line metric. Full description of these results is omitted from this version of the paper.

Our results highlight a number of questions and related problems that remain open. For example, the weighted version of the problem is a possible extension. For network testing, a natural problem is to minimize the number of rounds required to test all transmitters. We note that an algorithm for DCL can be used as part of a greedy *set cover*-type procedure for this, but a more direct approach may produce better results. Also, an extension of our algorithm for DCL on a line to the case of Euclidean plane may be relevant for this setting. Given our hardness results for DCL, it may be worthwhile to consider alternative formulations. For example, it would be interesting to study the approximability of the version that seeks to minimize the number of unassigned nodes, instead of maximizing the number of assigned ones. By analogy with minimum vertex cover and maximum independent set, the minimization formulation may be more amenable to approximation. Finally, we leave for future work the investigation of similar problems with two-sided preferences.

Acknowledgements

We thank Peter Winkler and Kurt Mehlhorn for useful discussions, and Uri Feige for suggesting the reduction to the metric case.

References

[1] D. Abraham, K. Cechlárová, D. Manlove, and K. Mehlhorn. Pareto optimality in house allocation problems. In *Proc. of the 15th ISAAC*, pages 3–15, 2004.

[2] D. Abraham, R. Irving, K. Mehlhorn, and K. Telikepalli. Popular matchings. In *Proc. 16th ACM Symp. on Discrete Algorithms*, pages 424–432, 2005.

[3] P. Alimonti and V. Kann. Some APX-completeness results for cubic graphs. *Theoretical Computer Science*, 237:123–134, 2000.

[4] P. Berman and T. Fujito. On approximation properties of the independent set problem for low degree graphs. *Theory of Computing Systems*, 32:115–132, 1999.

[5] D.G. Corneil and P.A. Kamula. Extensions of permutation and interval graphs. Proc. 18th Southeast. Conf. on Combinatorics, Graph Theory, and Computing, 1987.

[6] I. Dagan, M. C. Golumbic, and R. Y. Pinter. Trapezoid graphs and their coloring. *Discrete Appl. Math.*, 21(1):35–46, 1988.

[7] Z. Drezner. *Facility Location: A Survey of Applications and Methods*. Springer, 1995.

[8] S. Even, O. Goldreich, S. Moran, and P. Tong. On the NP-completeness of certain network testing problems. *Networks*, 14, 1984.

[9] U. Feige. A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4):634–652, 1998.

[10] S. Felsner, R. Muller, and L. Wernisch. Trapezoid graphs and generalizations, geometry and algorithms. *Discrete Appl. Math.*, 74(1):13–32, 1997.

[11] D. Gale and L. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.

[12] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1979.

[13] D. Gusfield and R. Irving. *The Stable Marriage Problem*. The MIT Press, 1989.

[14] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[15] R. Irving, T. Kavitha, K. Mehlhorn, D. Michail, and K. Paluch. Rank-maximal matchings. In *Proc. 15th ACM Symp. on Discrete Algorithms*, pages 68–75, 2004.

[16] D. Knuth. *Mariages stables et leurs relations avec d'autre problèmes combinatoires*. Les Presses de l'université de Montréal, 1976.

[17] D. Manlove and C. Sng. Popular matchings in the capacitated house allocation problem. In *Proc. 14th European Symposium on Algorithms*, pages 492–503, 2006.

[18] R. McConnell and J. Spinrad. Modular decomposition and transitive orientation. *Discrete Math.*, 201(1-3):189–241, 1999.

[19] J. Mestre. Weighted popular matchings. In *Proc. of the 33rd ICALP*, pages 715–726, 2006.

[20] D. Michai. Reducing rank-maximal to maximum weight matching. *Theoretical Computer Science*, 389(1-2):125–132, 2007.

[21] M. Pal, E. Tardos, and T. Wexler. Facility location with nonuniform hard capacities. In *Proc. 42nd IEEE Symp. on Foundations of Computer Science*, page 329, 2001.

[22] C. H. Papadimitriou and M. A. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, 43:425–440, 1991.

[23] D. Shmoys. Approximation algorithms for facility location problems. In K. Jansen and S. Khuller, editors, *Approximation Algorithms for Combinatorial Optimization*, pages 27–33. Springer, Berlin, 2000.

[24] L. Stockmeyer and V. Vazirani. NP-completeness of some generalizations of the maximum matching problem. *Information Processing Letters*, 15(1), 1982.

[25] V. V. Vazirani. *Approximation algorithms*. Springer-Verlag New York, Inc., 2001.

[26] J. Zhang, B. Chen, and Y. Ye. A multiexchange local search algorithm for the capacitated facility location problem. *Math. Oper. Res.*, 30(2):389–403, 2005.

A Appendix

A.1 Proof of Theorem 2.1

Proof. We present a gap-preserving reduction from MAXIMUM INDEPENDENT SET (IS), which is known to be hard to approximate within $n^{1-\epsilon}$ for any $\epsilon > 0$ unless NP=ZPP [14].

Let $H = (W, F)$ be the given instance of IS, and assume that $W = \{0, ..., n-1\}$. For each vertex $w \in W$, we create n agents and n unit-capacity centers and denote them as $\{a_j^w\}$ and $\{l_j^w\}$, respectively, with j ranging from 0 to $n-1$. We say that these nodes *correspond* to the vertex w. Let A and L be the sets of all such agents and centers, respectively. We define the preferences below. The intuition behind our construction is that if a vertex w is part of an independent set, then the pairs (a_j^w, l_j^w), for all j, can be part of a valid matching. Conversely, if some pairs (a_j^w, l_j^w) are in the valid matching, then we can add w to an independent set of the original instance.

Define the *index* of a node a_j^w (and also of l_j^w) as $w \cdot n + j$. Then the indices in each of A and L are unique and range from 0 to $n^2 - 1$. The preference list of an agent a_j^w consists of four parts:

1. Highest are the centers $l_k^{w'}$ with $(w, w') \in F$ and $w' < w$, ranked in order of increasing index.
2. Next on the list is this agent's intended match, l_j^w.
3. Following are the centers $\{l_k^w : k \neq j\}$, again ordered by increasing index. These are all the centers that correspond to vertex w, except for l_j^w.
4. Finally, the list contains all other centers, also in the order of increasing index.

Now we argue that: (1) if there is an independent set $W' \subseteq W$ in H, then there exists a valid matching of size at least $n \cdot |W'|$; and (2) if there is a valid matching μ, then H contains an IS of size at least $\frac{|\mu|-n}{n}$. This implies that the size of the maximum IS on H, denoted $\alpha(H)$, and the size of the maximum valid matching in the constructed instance, $|\mu^*|$, satisfy

$$n \cdot \alpha(H) \ \leq \ |\mu^*| \ \leq \ n \cdot \alpha(H) + n. \tag{2}$$

(1) Given W', if $w \in W'$, then we match a_j^w to l_j^w for all j. We claim this yields a valid matching of size $n|W'|$. To see this, suppose that some a_j^w prefers another matched center $l_k^{w'}$ to its current one l_j^w. By the structure of the preferences, this means that $w' \neq w$ and $(w, w') \in F$. But if $l_k^{w'}$ is matched, then both w and w' are in W', contradicting the assumption that W' is independent.

(2) We define S^w to be the set of nodes (both agents and centers) that correspond to a given vertex $w \in W$, and show that any valid matching can contain at most n matched pairs in which the matched nodes belong to different sets S^w. We prove this by arguing that at most one element of each S^w can be matched to a node that is outside of S^w and has a higher index. We consider three cases. First, suppose that two centers, l_j^w and l_k^w, with $j < k$, are matched to higher-index agents outside S^w. But this contradicts the definition of a valid matching, because the agent $\mu^{-1}(l_k^w)$ prefers l_j^w to its own match. As a second case, consider an agent a and a center l, both in S^w, that are matched to higher-index nodes outside S^w. But then a would prefer l to its own match. The last case is if two agents in S^w, a and a', are matched outside S^w. Assume, without loss of generality, that $\mu(a)$ has a lower index than $\mu(a')$. But then a' prefers $\mu(a)$ to $\mu(a')$.

We remove from μ all pairs whose elements correspond to different vertices of H. As we just showed, there can be at most n of them. Let the remaining matching be μ'. Then μ' is still a valid matching and has size at least $|\mu| - n$. Create a set $W' \subseteq W$ by including all vertices w such that $(a_h^w, l_j^w) \in \mu'$ for some h and j. We claim that W' is an independent set in H. If not, and two adjacent vertices w and w' are in W', then μ' must contain two pairs (a_h^w, l_j^w) and $(a_k^{w'}, l_l^{w'})$. Assume that $w > w'$. Then by the construction of preferences, a_h^w prefers $l_l^{w'}$ to its own match, contradicting the assumption that μ' is a valid matching. Finally, observe that μ' can contain at most n pairs consisting of agents and centers from any particular S^w. This implies that $|W'| \geq |\mu'|/n \geq (|\mu| - n)/n$.

To show that our reduction is gap-preserving, we phrase the known results about the hardness of IS (or, equivalently, maximum clique) [14] in the following way. For some $\beta(n) \geq 1$ and for $\gamma(n) = n^{1-\epsilon}$ with any $\epsilon > 0$, there are graphs H on n vertices for which the following two cases are NP-hard to distinguish, unless NP=ZPP:

1. $\alpha(H) \leq \beta(n)$
2. $\alpha(H) \geq \gamma(n) \cdot \beta(n)$.

We use such a graph H to construct an instance of DCL as described above, and let μ^* be its optimum solution. Then inequality (2) implies that the following two cases are hard to distinguish:

1. $\alpha(H) \leq \beta(n)$, and therefore $|\mu^*| \leq n \cdot \alpha(H) + n \leq n \cdot \beta(n) + n$
2. $\alpha(H) \geq \gamma(n) \cdot \beta(n)$, and therefore $|\mu^*| \geq n \cdot \alpha(H) \geq n \cdot \gamma(n) \cdot \beta(n)$.

So DCL is hard to approximate within a factor of $\frac{n\gamma(n)\beta(n)}{n\beta(n)+n} = \frac{\gamma(n)\beta(n)}{\beta(n)+1} \geq \frac{1}{2}\gamma(n)$, using the fact that $\beta(n) \geq 1$. Now, since the size of the derived DCL instance is $|A \cup L| = 2n^2$, we get a hardness of approximation of $\frac{1}{2}\gamma(n) = \frac{1}{2}\gamma(\sqrt{|A \cup L|/2}) = O\left(|A \cup L|^{1/2-\epsilon}\right)$. □

We note that to obtain an equivalent result for the case with incomplete preferences, it suffices to include only preferences listed as items 1 and 2 in the proof. Then the constructed instance satisfies an additional interesting property of *consistency*: there is a global ordering of the set of centers L such that each agent $a \in A$, among centers for which it has preferences, prefers the earlier ones in the ordering to the later ones.

A.2 Proof of Theorem 2.2

Proof. We present an *L-reduction* [22] from MAXIMUM INDEPENDENT SET, which is known to be APX-complete even for 3-regular graphs [3, 4]. Let $H = (W, F)$ be the original instance of IS on a 3-regular graph (so $|F| = 3|W|/2$). We transform it into an instance of DCL so that H has an independent set of size $|W'|$ if and only if the derived instance allows a valid matching μ of size $|W'| + |F|$.

For each vertex $w \in W$, we create a center l_w, and for each edge $e \in F$, we create a center l_e. For each edge $e = (w, w')$, we create two agents $a_{e,w}$ and $a_{e,w'}$ with preference lists (l_w, l_e) and $(l_{w'}, l_e)$ respectively. We now argue that this is a valid L-reduction.

($\rightarrow$) Let W' be an independent set in the original instance H. Suppose $w \in W'$ and it is incident to edges $(w, w_1), (w, w_2), (w, w_3)$. Choose any incident edge, say (w, w_1), and let $\mu(a_{(w,w_1),w}) = l_w$. To match the centers l_e that correspond to edges, observe that for every edge $e = (w, w')$, at most one of $\{w, w'\}$ can be in the independent set W', and thus at most one of the agents $\{a_{e,w}, a_{e,w'}\}$ can be matched already. So we match the other one (or an arbitrary one, if neither is already matched) to l_e. The resulting matching μ is valid and has size $|W'| + |F|$. Thus $|\mu^*| \geq \alpha(H) + |F|$, where μ^* is the optimum valid matching in the derived instance, and $\alpha(H)$ is the size of the maximum independent set in H.

($\leftarrow$) We pre-process a valid matching μ to make sure that vertices $w \in W$ that correspond to matched centers l_w form an independent set in H. If for some edge $e = (w, w')$, both l_w and $l_{w'}$ are matched, then it must be the case that l_e is not matched (if it were matched to, say, $a_{e,w}$, then $l_w \succ_{a_{e,w}} l_e$, contradicting the feasibility of μ). So we can delete $(\mu^{-1}(l_w), l_w)$ from μ, and add $(a_{e,w}, l_e)$ instead, preserving the size and validity of the matching. This can be repeated for other violated edges. This implies that $|\mu| \leq \alpha(H) + |F|$, because μ includes at most $\alpha(H)$ centers l_w that correspond to vertices $w \in W$ and at most $|F|$ centers l_e that correspond to edges $e \in F$.

Finally, we remark that in our construction, every agent has a preference list of size 2 and every center appears in at most 3 agents' preference lists (as H is 3-regular). □

A.3 Improving the ratio from $1/2d$ to $1/\phi d$ for DCL with general capacities

We give a finer analysis of the algorithm presented in Section 3.3. Recall that when a center $u \notin U$ is deleted in step 5, its in-degree in G', at the time of deletion, is greater than ξc_u. Let us divide the sources of these incoming edges, for all $u \notin U$, into three disjoint groups:

1. From agents a with $\mu(a) \notin U$
2. From agents a with $\mu(a) \in U$, and such that a ranks $\mu(a)$ highest among all $l \in U$
3. From agents a with $\mu(a) \in U$, and such that a prefers some $l \in U$ to $\mu(a)$

We upper-bound the total number of such edges. Note that for any a with $\mu(a) \notin U$, there is only one edge, namely the one toward $\mu(a)$, that is counted toward this sum. This is because from the topological ordering of H, we know that a does not have any edges to centers occurring before $\mu(a)$; and when $\mu(a)$ is deleted in step 5, all edges from a to other centers are deleted as well, so they are not counted when those later centers are processed. Thus, there is a total of n_r edges of the first type. Each agent of the second type can have up to $d-1$ edges toward deleted centers, but each agent of the third type can have at most $d-2$ edges toward deleted centers, as it has at least two directed edges toward centers in U. Let s be the fraction of agents assigned by μ to centers in U that rank their assigned center highest among ones in U. The above discussion implies that

$$n_r + n_u\, s\,(d-1) + n_u\,(1-s)\,(d-2) \;\geq\; \sum_{u \notin U} \xi c_u \;\geq\; \xi n_r,$$

so $n_r \leq \frac{s+d-2}{\xi-1} n_u$. As in Section 3.3, $\textsc{Alg} \geq n_u/\xi$. But also, $\textsc{Alg} \geq s n_u$, since the final solution is at least as big as one that only assigns the $s n_u$ agents that are already matched to their best centers in U by μ. Thus, the approximation ratio can be expressed as

$$\frac{\textsc{Alg}}{\textsc{Opt}} \;\geq\; \frac{\max(s, 1/\xi)\cdot n_u}{n_u + n_r} \;\geq\; \frac{\max(s, 1/\xi)\cdot n_u}{n_u + \frac{s+d-2}{\xi-1} n_u} \;=\; \frac{\max(s, 1/\xi)\cdot(\xi-1)}{\xi + s + d - 3}.$$

Calculus shows that the last expression is minimized with respect to s at $s = 1/\xi$, so

$$\frac{\textsc{Alg}}{\textsc{Opt}} \;\geq\; \frac{\xi - 1}{\xi^2 + (d-3)\xi + 1},$$

which is maximized at $\xi = 1 + \sqrt{d-1}$. Thus, the approximation ratio becomes $\frac{1}{d-1+2\sqrt{d-1}} \geq 1/\phi d$, for $\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$, the golden ratio.

A.4 Proof of Theorem 3.6

Proof. We divide all centers into two groups, L_1 and L_2. A center u belongs to L_1 if at least c_u agents rank u as their top choice; otherwise, u belongs to L_2. Let $\textsc{Alg}_1$ be the solution in which each center $u \in L_1$ is assigned c_u agents who rank it the highest. Let $\textsc{Alg}_2$ be the solution in which each center $u \in L_2$ is first assigned the set of agents ranking it the highest, followed by the set of agents ranking it second and who do not rank another center $u' \in L_2$ first, until either all such agents are assigned or the capacity c_u is reached. Finally, let $\textsc{Opt}_1$ and $\textsc{Opt}_2$ be the assignment of the optimal solution $\textsc{Opt}$ restricted to centers in L_1 and L_2 respectively. We choose the larger-size solution between $\textsc{Alg}_1$ and $\textsc{Alg}_2$ and claim that $\textsc{Alg}_1 \geq \textsc{Opt}_1$ and $\textsc{Alg}_2 \geq \textsc{Opt}_2$. Then $\max\{\textsc{Alg}_1, \textsc{Alg}_2\} \geq \frac{1}{2}(\textsc{Opt}_1 + \textsc{Opt}_2) = \frac{1}{2}\textsc{Opt}$, yielding the desired approximation ratio.

The fact that $\textsc{Alg}_1 \geq \textsc{Opt}_1$ follows because $\textsc{Alg}_1$ fills all centers in L_1 to capacity. To argue that $\textsc{Alg}_2 \geq \textsc{Opt}_2$, we use the following claim, where we let $L_2^* \subseteq L_2$ denote the subset of centers that are opened in $\textsc{Opt}_2$.

Claim A.1 *For every $u \in L_2^*$, the number of agents assigned to center u in* Opt *is at most the number of agents assigned to u in* Alg_2 *plus the number of agents for which u is the second choice and which are assigned to another center $u' \in L_2 \setminus L_2^*$ in* Alg_2.

If u has c_u agents in Alg_2, or if every agent with an edge to u is assigned to u in Alg_2, the claim is trivial. Otherwise assume that there is some agent a who is assigned to u in Opt but not in Alg_2. Since all agents for which u is the first choice must be assigned to u based on the algorithm, u must be the second choice for a. Moreover, a must be assigned to its first-choice center $u' \in L_2 \setminus \{u\}$ by Alg_2. This is because if a's first choice were in L_1, then either a would be assigned to u in Alg_2, or u would already have c_u agents, both of which violate our earlier assumptions. This means that a's first choice is in $L_2 \setminus \{u\}$, in which case Alg_2 must match a to its first choice. Observe that $u' \notin L_2^*$, as otherwise Opt is not a valid matching. So $u' \in L_2 \setminus L_2^*$, proving the claim.

To see that the claim implies $\text{Alg}_2 \geq \text{Opt}_2$, notice that the sets of agents assigned in Alg_2 to centers in $L_2 \setminus L_2^*$ and having u as their second choice are disjoint for different u. □

A.5 Proof of Lemma 4.3

Proof. Given a terminal intersection graph $H = (W, F)$, for each vertex $w \in W$ that corresponds to an interval $[a_w, b_w]$ and a terminal c_w, we define a trapezoid T_w by the points a_w and c_w on the top line, and points c_w and b_w on the bottom line. Thus, the construction takes linear time.

To prove correctness, we first show that if $(w, w') \in F$, then trapezoids T_w and $T_{w'}$ intersect. Suppose without loss of generality that $c_w \in I_{w'}$. The line segment between c_w on the top line and c_w on the bottom line is contained in T_w. And since $c_w \in I_{w'}$, this segment must intersect the segment from $a_{w'}$ on the top line to $b_{w'}$ of the bottom line, which is part of the trapezoid $T_{w'}$. Thus, T_w and $T_{w'}$ intersect. Conversely, suppose that $(w, w') \notin F$. Assume without loss of generality that $c_w < c_{w'}$. Now, since $c_w \notin I_{w'}$, it must be that $a_{w'} > c_w$. Similarly, $c_{w'} \notin I_w$, so $b_w < c_{w'}$. So the right side of T_w is strictly to the left of the left side of $T_{w'}$, and the two trapezoids are disjoint. □