

The Roommates Problem Revisited

Thayer Morrill*

North Carolina State University

May 2009

Abstract

One of the oldest but least understood matching problems is Gale and Shapley's (1962) "roommates problem": is there a stable way to assign $2N$ students into N roommate pairs? Unlike the classic marriage problem or college admissions problem, there need not exist a stable solution to the roommates problem. However, the traditional notion of stability ignores the key physical constraint that roommates require a room, and it is therefore too restrictive. Recognizing that many assignments, such as a roommates assignment, require bilateral approval to dissolve motivates a new matching problem: matching agents subject to an initial assignment. A particularly important example of this is kidney exchange where after an assignment has been made, subsequent tests may determine that a patient and donor are incompatible. This paper introduces an efficient algorithm for finding a Pareto improvement starting from any *status quo* roommates assignment.

*Email address: thayermorrill@gmail.com. I would like to thank Larry Ausubel, Peter Cramton, Melinda Morrill, and Daniel Vincent for their comments throughout this project. I would also like to thank Daniel Aromi, Emel Filiz Ozbay, Jonah Gelbach, and Erkut Ozbay. Two anonymous referees provided a number of thoughtful comments that greatly benefited this paper.

1 Introduction

Two-sided matching theory has been well studied by economists, mathematicians, and computer scientists who have created an elegant and applicable theory. Unfortunately, one-sided matching theory has been comparatively neglected.¹ This neglect is likely due to the very paper that introduced it. In their classic 1962 article *College Admissions and the Stability of Marriage*, Gale and Shapley introduce both the marriage problem and the roommates problem. While Gale and Shapley prove a stable match always exists in a two-sided market, they introduce the roommates problem to demonstrate that a stable pairing need not exist in a one-sided market. Since a stable match need not exist, economists have been stymied in their attempts to find and analyze solutions to this important assignment problem. Unfortunately, this has led many economists to turn their attention elsewhere, and as a result, the economics literature on this classic problem is sparse.

This paper introduces Pareto optimality as an alternative solution concept to stability for a wide class of matching models including the roommates problem. Gale and Shapley define a set of marriages as unstable if either there exist a man and woman who are not married but prefer each other to their current spouse or there exists someone who would prefer to be single than married to her current partner. Stability in the roommates problem is borrowed from the marriage model. A pairing is unstable if two students prefer to live with each other rather than their current assignment.² This paper is concerned with the following question: *ex-post*, what types of coalitions

¹Roth and Sotomayor (1990) is an excellent introduction to the two-sided matching literature. Gusfield and Irving (1989) is also a nice introduction. Interestingly, although the economics literature on the roommates problem is very small, there is a comparatively large computer science literature on it. Roth and Sotomayor, two economists, mention the roommates problem only as an example. In contrast, Gusfield and Irving, two computer scientists, devote nearly a quarter of the book to the roommates problem. Finding a traditionally-stable roommate pairing (if one exists) is considered a “hard” algorithmic question. The bulk of their presentation is a polynomial-time algorithm for finding a traditionally-stable pairing when one exists. Tan (1991) establishes a necessary and sufficient condition for the existence of a stable pairing. Chung (2000) extends Tan’s result to a sufficient condition for the existence of a stable pairing when preferences are weak.

²I am interested in the case where each student is required to have a roommate. Consequently, I do not include in my definition of stability the additional requirement that each student prefers her assignment to being unassigned.

will be able to block a given assignment? A key consideration is whether bilateral approval is required to dissolve the original partnership. For example, roommates face an additional constraint that married couples do not; roommates must have a room in which to live. A student may prefer another to her assigned roommate; however, she needs a room in which to live and presumably does not have the right to evict her current roommate. Therefore, after an assignment has already been made, the traditional notion of stability is too restrictive.

I will present Gale and Shapley's original example to highlight this point.

Example (Gale and Shapley, 1962): *A Stable Assignment Need Not Exist*

Suppose there are four students: α, β, γ and δ . α 's top choice is β , β 's top choice is γ , γ 's top choice is α , and all three rank δ last. Gale and Shapley define an assignment to be unstable if two students are not currently roommates but prefer each other to their current assignment. Under this definition, there does not exist a stable assignment since whoever is assigned to δ prefers the other two students to δ and is the top choice of one of these students. In the words of Gale and Shapley:

“...whoever has to room with δ will want to move out, and one of the other two will be willing to take him in.”

While one of the other two may be willing to take him in, it is quite a different matter whether this student is able to take him in. In order to take him in, either his current roommate must voluntarily leave, be evicted, or an additional room must be available. With a scarcity of rooms and with no student willing to change his assignment to δ , the original assignment will not be disturbed after all.

This does not contradict Gale and Shapley's original analysis but rather suggests a new class of assignment problems. Gale and Shapley looked at agents and preferences and asked if there exists a stable assignment. This paper looks at agents, preferences, and a *status-quo* assignment and asks the following question: if each agent has the option to maintain her *status-quo*, what is the core of this coalitional game?

If the *status-quo* assignment can be dissolved unilaterally, then stability is the natural solution concept. If an agent finds someone she prefers who also prefers her, then both parties will dissolve their current partnership and pair together. As a result, the non-transferable utility core will consist only of stable assignments. However, if a partnership requires bilateral agreement to dissolve, then two agents wanting to change their assignment is not enough to block the original pairing. If bilateral agreement is required, an assignment will only be changed if all reassigned parties agree. Since an agent will only agree if the new assignment makes her better off, any deviation from the original set of assignments must be a Pareto improvement. Therefore, when bilateral agreement is required to dissolve a partnership, the core consists of Pareto-optimal, Pareto-improvements of the initial assignment.

The theoretical contribution of this paper is to introduce a new class of matching problems: matching agents subject to an initial assignment. These problems are important to market design as they extend several classic matching theory applications to a dynamic environment. Typically, a match is made based on a fixed population and set of preferences. However, in real-world applications invariably both preferences and the population being matched change. This is precisely the scope of this paper: a *status-quo* assignment exists and must be honored, but the primitives of the problem have changed so that the match may no longer be in the core.

The practical contribution is to present a fast algorithm for Pareto-improving any inefficient roommates assignment. This algorithm will be important for extending any application of the roommates problem to a dynamic environment. Section 3.1 discusses a particular important application of this algorithm to kidney exchange.

The paper is organized as follows. Section 2 formally introduces the problem. Section 3 details the Roommate Swap algorithm and discusses applications. Section 4 examines the strategic implications of mechanisms for selecting a core assignment, and Section 5 concludes. The appendix provides several technical proofs and a discussion of the computational complexity of the Roommate Swap algorithm.

2 The Model and Applications

2.1 Matching Subject to an Initial Assignment

This section formally introduces the assignment game.³ Each game consists of the triple (S, \succ, μ) . $S = \{s_1, \dots, s_{2N}\}$ is a set of students. The second element, $\succ = (\succ_1, \dots, \succ_{2N})$ is the list of preferences of each student where \succ_i denotes the strict preferences of student i over the $2N - 1$ other students. The last element, μ is an initial matching of students. A **matching** is an assignment of students such that every student is assigned to exactly one other student and assignments are symmetric. More formally, it is a function from S to S such that for every student s , $\mu(s) \neq s$ and $\mu(\mu(s)) = s$. Let \mathcal{M} denote the set of matchings. For convenience of notation, I fix S and \succ and denote the assignment game by μ .

A matching μ is **Pareto efficient** if there does not exist a $\nu \in \mathcal{M}$ such that $\nu(s) \succeq \mu(s)$ for all $s \in S$ and $\nu(s) \succ \mu(s)$ for some $s \in S$.

Definition 1. A matching η is in the **core** of the assignment game μ , denoted $core(\mu)$, if

1. $\nu(s) \succeq \eta(s)$ for every $s \in S$.

and there is no matching ν such that

2. $\nu(b) \succeq \eta(b)$ for every $s \in S$.
3. $\nu(b) \succ \eta(b)$ for some $s \in S$.

ν is said to **block** matching η . μ is a **core assignment** if $\mu \in core(\mu)$.

In this context, the only coalition we need to consider is the grand coalition. As each agent has the right to maintain her initial assignment, a reassignment will be vetoed if it is not a Pareto improvement. Therefore, $core(\mu) = \{\mu\}$ if

³The framework presented here is modeled after Abdulkadiroglu and Sonmez (1998).

μ is Pareto-efficient. Otherwise, $core(\mu)$ consists of Pareto-efficient, Pareto-improvements of μ .

One can make a Pareto-optimal assignment a number of ways, but one mechanism with particular desirable properties is the *Random Serial Dictatorship*.⁴ Assign every student a priority (randomly or otherwise). Assign the student with highest priority her most preferred roommate and remove them both from consideration. From students who remain, assign the student with highest priority her most preferred roommate among those students that are unassigned. Remove these two from consideration and repeat until no students remain. This assignment is Pareto efficient.⁵

In contrast to the ease of making a core assignment, it is difficult to determine if a given roommates assignment is efficient. Preferences between students need not interact when assigning students, but they interact directly when determining if one assignment Pareto improves another. At this point the reader may object as there is an obvious algorithm to determine if an assignment is Pareto-efficient: look at each possible reassignment and determine if it Pareto improves the original. If no assignment Pareto improves the original, then the original is efficient. Unfortunately, this algorithm is of no practical use as the growth of the number of assignments relative to students being assigned is factorial. Specifically, given $2N$ students there exists $\frac{(2N)!}{2^N(N!)}$ = $(2N - 1)(2N - 3)(2N - 5) \cdots (3)(1)$ many ways of assigning them to be roommates.⁶ Even for small N , this is prohibitively large. For example, there exists on order of 6 quadrillion (6×10^{15}) many ways to assign 30 students to be roommates. Therefore, a more sophisticated process is required.

⁴See Abdulkadiroglu and Sonmez (1998).

⁵To see this, note that if a student is involved in a Pareto improvement, then necessarily her roommate must be involved as well. The student with highest priority, s_1 , receives her top choice, s_2 , so neither she nor her choice can be involved in a Pareto improvement. Let s_3 be the student who chooses second. Since neither s_1 or s_2 are involved in any Pareto improvements, if s_3 is part of a Pareto improvement she must be reassigned to a student among $S \setminus \{s_1, s_2\}$. However, s_3 already receives her top choice among this set. Therefore, s_3 (and consequently the student she chooses) is not part of any Pareto improvement. Similarly, the student who chooses third is not part of any Pareto improvement, and so on.

⁶A short proof appears in the Appendix.

3 The Roommate Swap Algorithm

This section demonstrates an $O(n^2)$ algorithm for determining if a roommate assignment is efficient. Moreover, when an assignment is inefficient I provide an $O(n^3)$ algorithm, The Roommate Swap, for finding a Pareto improvement.⁷

The extra structure inherent in a two-sided match makes it simple to find a Pareto improvement for an inefficient assignment. I present one algorithm here as it provides intuition for the more complicated one-sided case.⁸ For a given man m , define a woman w to be **achievable** for m if w weakly prefers m to her current husband. Have each male point to his most-preferred, achievable woman. Note that every man has a woman to point to as his current assignment is achievable. Have each woman point to her current assignment. There must exist a cycle since there are only a finite number of agents and each agent is pointing to someone. If the cycle is trivial (the man is pointing to the woman he is currently married to), then neither the man nor the woman can be involved in a Pareto improvement and we can remove them from consideration. If the cycle is non-trivial, then reassigning each man in the cycle to the woman he is pointing to is a Pareto-improvement.

The arrow from a man to an achievable woman ensures that both prefer the new match. The arrow from a woman ensures her old match will be re-assigned. Therefore, a non-trivial cycle that alternates between “male” arrows and “female” arrows will be a Pareto improvement. Moreover, the structure inherent in a two-sided match guarantees that any cycle will alternate between these two types of arrows. This need not be the case for a one-sided match, but the intuition still holds: a cycle that alternates between a preferred, achievable agent and her old assignment is a Pareto-improvement.

I will define a graph to aid our search for a Pareto-improvement.⁹ Given a set of preferences \succ and assignment μ , induce a graph, G_{\succ}^{μ} , as follows:

⁷A discussion on the computational complexity of the algorithm appears in the appendix.

⁸This algorithm is a slight variation of the Top-Trading Cycles algorithm (Shapley and Scarf, 1974) and is very similar to the algorithm presented in Erdil and Ergin (2008).

⁹I refer the reader to *Introduction to Graph Theory*, second edition, by Douglas West for a detailed introduction to graph theory.

- Each vertex corresponds to a student. Label the vertices s_1 through s_n . When referring to the graph, I will use the term vertex and student interchangeably.
- A *solid* edge is drawn between roommates. By definition, each vertex is incident to exactly one solid edge.
- Draw a *dashed* edge between any two students that form a blocking pair to μ . That is to say, if s_i prefers s_j to her current roommate and vice versa.

When the preferences and assignment are clear from the context, I will just refer to the graph as G . I will call a path that alternates between dashed and solid edges (or vice versa) an **alternating path**. Similarly, a cycle that alternates between dashed and solid edges is an **alternating cycle**.

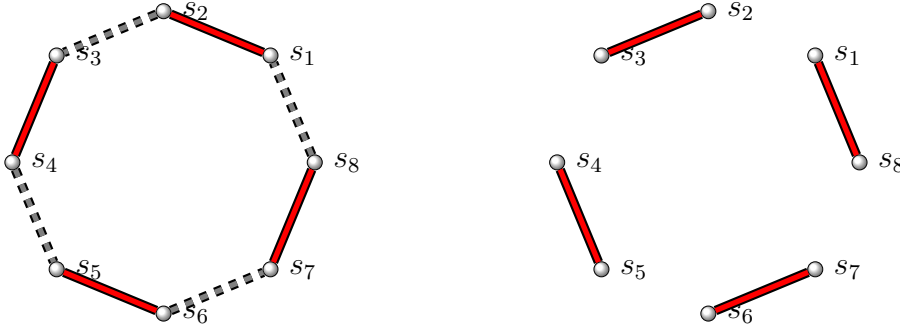
Lemma 1. *An assignment μ is efficient under preferences \succ if and only if G_\succ^μ contains no alternating cycle. Moreover, if μ' Pareto improves μ and s is a student such that $\mu(s) \neq \mu'(s)$, then s is contained in an alternating cycle in G_\succ^μ .¹⁰*

The intuition for sufficiency is captured in Figure 1. In an alternating cycle, we can simply “swap” roommates. We eliminate the solid edges, make the dashed edges in the cycle solid, and leave everyone outside the cycle unchanged. This is a well-defined reassignment that Pareto improves the initial assignment.

Proof. Suppose G_\succ^μ contains an alternating cycle C . An alternating cycle is closed under roommates as each vertex is incident to a solid edge in the cycle. This implies $V(G) \setminus C$ is closed under roommates as well ($V(G)$ means the vertex set of G). We will construct a Pareto improvement μ' . For every $v \in V(G) \setminus C$ let $\mu'(v) = \mu(v)$. This is well defined since $V(G) \setminus C$ is closed under roommates. For every $v \in C$, let $\mu'(v)$ be the vertex it shares a dashed edge with in the cycle C . This is well defined as each vertex is incident to exactly one dashed edge in the cycle and sharing a dashed edge is a reciprocal

¹⁰This Lemma was discovered independently by Abraham and Manlove (2004).

Figure 1: An alternating cycle with its corresponding Pareto improvement.



relationship. A dashed edge indicates that both vertices prefer each other to their original assignment. Therefore, μ' Pareto improves μ .

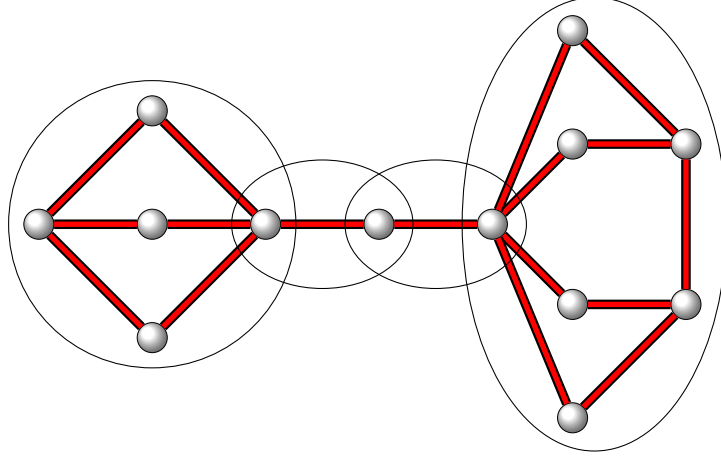
Suppose that μ' is a Pareto improvement of μ . Let G' be the subgraph consisting of all solid edges in G_\succ^μ and only the dashed edges between vertices not paired by μ that are paired by μ' (since μ' is a Pareto improvement, there must be a dashed edge between such vertices). Note that any vertex v in G' either has degree¹¹ 1 (if $\mu(v) = \mu'(v)$) or degree 2 (if $\mu(v) \neq \mu'(v)$). Moreover, for any vertex v , if $d(v) = 2$, then $d(\mu(v)) = d(\mu'(v)) = 2$. Choose any vertex t such that $d(t) = 2$. t is connected via a solid edge to $\mu(t)$. Since $d(t) = 2$, $d(\mu(t)) = 2$ and so $\mu(t)$ must be connected via a dashed edge to $\mu'(\mu(t))$. $\mu'(\mu(t))$ must be connected via a solid edge to $\mu(\mu'(\mu(t)))$ which must be connected to a dashed edge via $\mu'(\mu(\mu'(\mu(t))))$, and so on. Eventually this process must cycle as there are a finite number of vertices. However, a cycle to any vertex $s \neq t$ would mean the degree of s is at least three which is not possible. Therefore, the process must cycle back to our first vertex t . Moreover, it must cycle via a dashed edge as we have already exhausted t 's solid edge. By construction, this is an alternating cycle. \square

For a graph G , a vertex v is a **cut-vertex** if G is connected but $G \setminus \{v\}$ is not. A **block** is a maximal subgraph containing no cut vertex. Note that the subgraph consisting of two vertices and an edge between them contains no cut-vertex, so any edge is either a block or a subset of a block. I will refer

¹¹The degree of a vertex v , denoted $d(v)$, is the number of edges v is incident to.

to any block containing only two vertices as a **trivial block**. Since every vertex in our graph has at least one edge, this is the smallest block possible. Figure 2 shows an example where the blocks have been circled.

Figure 2: An example of a graph with four blocks.



Lemma 2. *Let t be any student.*

1. t and $\mu(t)$ are contained in a unique block, B_t .
2. If t is part of an alternating-cycle C , then $C \subseteq B_t$.
3. If t is involved in a Pareto improvement, then B_t is non-trivial. That is to say if there exists an assignment μ' such that μ' Pareto improves μ and $\mu'(t) \neq \mu(t)$, then $|B_t| > 2$.

Proof.

1. Since there is an edge between t and $\mu(t)$, they are in at least one block together. Since the intersection of two blocks contains at most one student,¹² t and her roommate must be in exactly one block together.

¹²See West pg. 156. The intuition is that if two blocks B_1 and B_2 share two vertices, then after cutting a vertex, at least one of the two must remain. Call this vertex v . v is connected to all remaining vertices as it is in a block with each of them. But if every vertex has a path to v , then all vertices are connected. Therefore $B_1 \cup B_2$ has no cut-vertex contradicting the maximality of a block.

Call this block B_t .

2. A cycle contains no cut-vertex, so it must be a subset of a block. An alternating-cycle containing t must contain $\mu(t)$ since t lies on a solid edge in the alternating-cycle. Since B_t is the unique block containing t and $\mu(t)$, the cycle must be contained in B_t .
3. If t is involved in a Pareto improvement, then by Lemma 1 t is contained in an alternating-cycle. By (2) this alternating-cycle is contained in B_t , so B_t must contain more than just t and $\mu(t)$.

□

Lemma 2, part (2) says that if a student t is part of a Pareto improvement (and consequently an alternating-cycle), then she must be reassigned to a member of B_t . Therefore, no edge between t and a vertex outside of B_t can be part of an alternating-cycle. Let G' be the graph obtained by deleting all edges between t and any vertex not in B_t . Then G contains an alternating-cycle if and only if G' contains an alternating-cycle. This motivates the following procedure.

Pruning a Graph

1. Start with a graph G .
2. Determine the set of blocks B_1, B_2, \dots, B_m .
3. For each student-roommate pair s and $\mu(s)$, locate the unique block that both are in. Remove *all* edges from either s or $\mu(s)$ to any student outside this block.

A key point is that if a student s was in a block B with $\mu(s) \notin B$, then after pruning the graph, s is no longer in B . By iterating the pruning process we end up with a graph in which all blocks are **closed under roommates**¹³. Note that these blocks may be trivial, but by Lemma 2, the students in such a block are not involved in any Pareto improvements.

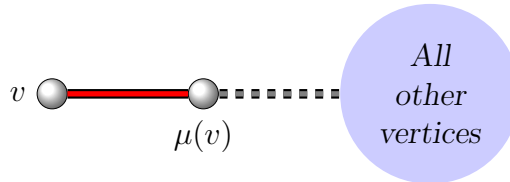
¹³Given an assignment μ , define a set S of students to be closed under roommates if $s \in S$ implies $\mu(s) \in S$.

Proposition 1. *Any non-trivial block closed under roommates contains an alternating cycle.*

The algorithm in this proof was inspired by Edmonds' Blossom Algorithm from graph theory¹⁴ and Gale's Top-Trading Cycles Algorithm.¹⁵

Proof. Suppose we have a non-trivial block G closed under roommates with $|V(G)| = 2N$. First note that in any non-trivial block closed under roommates, every vertex is incident to exactly one solid edge and at least one dashed edge. Each vertex v has a solid edge to $\mu(v)$ as the block is closed under roommates. There must be additional vertices as the block is non-trivial. If v had no dashed edge, then $\mu(v)$ would be a cut-vertex as removing $\mu(v)$ would disconnect v from these other vertices.

Figure 3: In a non-trivial block closed under roommates, any vertex v must be incident to at least one dashed edge or else $\mu(v)$ would be a cut-vertex.



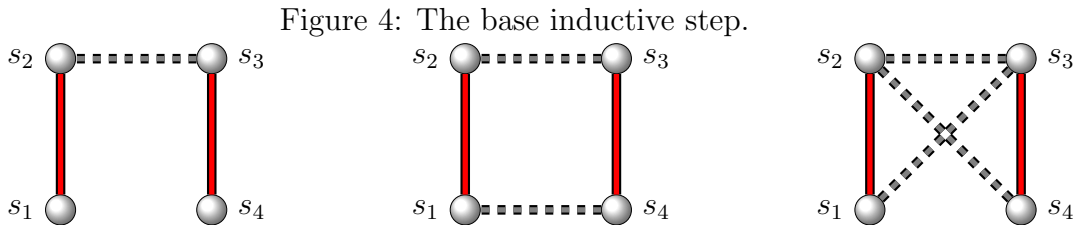
The proof proceeds by induction on N .

Base Step: $N=2$. Figure 4 illustrates the argument that follows. The smallest non-trivial block closed under roommates contains four vertices, $\{s_1, \dots, s_4\}$. Without loss of generality, $\mu(s_1) = s_2$, $\mu(s_3) = s_4$, and there is a dashed edge between s_2 and s_3 . If there is a dashed edge between s_1 and s_4 then we are done. If not, then the only remaining possibility is s_1 is adjacent to s_3 and s_4 is adjacent to s_2 , but in this case, $\{s_1, s_3, s_4, s_2\}$ is an alternating-cycle.

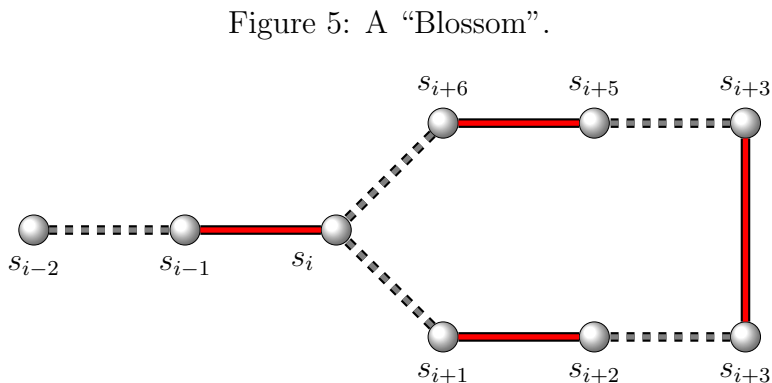
Inductive Step. Assume $N > 2$. Our inductive hypothesis is that any graph H that is closed under roommates, has no cut-vertex, and has strictly fewer than $2N$ vertices must contain an alternating cycle.

¹⁴Edmonds (1965). A discussion of the Blossom algorithm appears in West, page 142.

¹⁵Shapley and Scarf (1974).



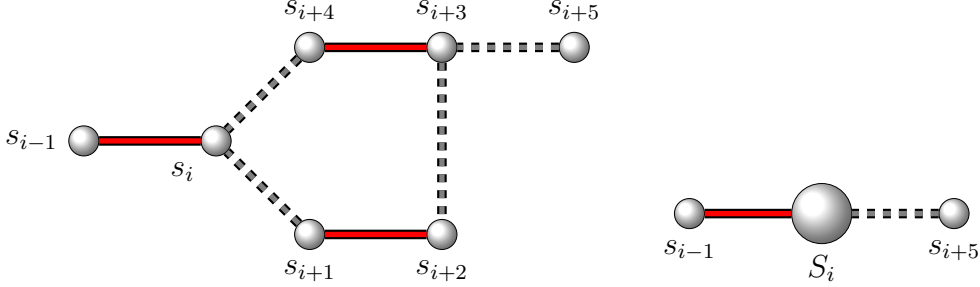
Start with any vertex s_1 , take its solid edge to $s_2 = \mu(s_1)$, and continue on a dashed edge to some s_3 . Continue alternating between dashed and solid edges until we cycle. We must eventually cycle since there are a finite number of vertices.



If our cycle is even (a cycle is even if it contains an even number of vertices), then we are done. By construction, an even cycle alternates between dashed and solid edges and is therefore an alternating cycle. Therefore, assume our cycle is odd, $\{s_i, s_{i+1}, s_{i+2}, \dots, s_{i+2m}\}$. By construction, any odd cycle looks like Figure 5, except possibly of different length. Edmonds refers to this as a blossom. The vertices $\{s_1, s_2, \dots, s_i\}$ are the stem, s_i is the base of the blossom, and s_i must connect to s_{i+1} and s_{i+2m} via dashed edges. It is important to note that for any odd cycle, the blossom minus the base, $\{s_{i+1}, s_{i+2}, \dots, s_{i+2m}\}$, is closed under roommates.

Intuitively, the way we would like to proceed is to treat the odd cycle $\{s_i, \dots, s_{i+2m}\}$ as a single vertex, I call it a super-vertex, and then continue to look for an alternating cycle. This will work for two reasons. First, if we

Figure 6: A blossom before and after contraction.

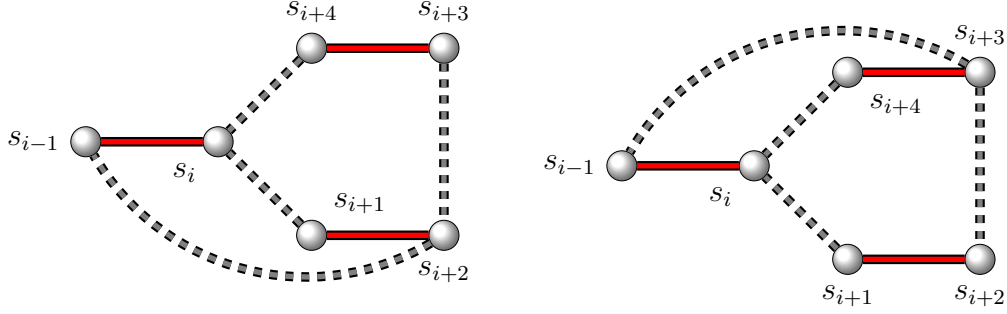


find an alternating cycle which contains the super-vertex in the “contracted” graph, we can always expand it to an alternating cycle in the original graph. Figures 8 - 10 on Pages 15 and 16 give the intuition behind this, but a formal proof will appear later. Second, by contracting the odd cycle, our new graph has strictly fewer vertices than the previous graph did. Therefore, as long as we can show it is closed under roommates and contains no cut-vertex, we know there will be an alternating-cycle by our inductive hypothesis.

Contract the odd cycle as follows: replace the set of vertices $\{s_i, \dots, s_{i+2m}\}$ with the single vertex S_i . Define a new graph G' with $V(G') = V(G) \setminus \{s_i, \dots, s_{i+2m}\} \cup S_i$. For all $u, v \in V(G) \setminus \{s_i, \dots, s_{i+2m}\}$, draw a dashed (respectively solid) edge between u and v if and only if there was a dashed (solid) edge between u and v in G . Add a solid edge between s_{i-1} and S_i , and add a dashed edge between S_i and any other vertex v if and only if there was a dashed edge between v and any vertex in $\{s_i, \dots, s_{i+2m}\}$ ¹⁶. s_{i-1} may now have both a solid and a dashed edge with S_i , but if this is the case then we have found our alternating cycle. See Figure 7 for an example, but more formally, if s_{i-1} has a dashed edge to $s_{i+2j} \in \{s_i, \dots, s_{i+2m}\}$, then $\{s_{i-1}, s_i, s_{i+1}, s_{i+2}, \dots, s_{i+2j}\}$ is an alternating cycle. If s_{i-1} is incident to $s_{i+2j+1} \in \{s_i, \dots, s_{i+2m}\}$, then $\{s_{i+2j+1}, s_{i+2j+2}, \dots, s_{i+2m}, s_i, s_{i-1}\}$ is an alternating cycle. In either case, we have found an alternating cycle, so we will assume there is no dashed edge between s_{i-1} and any vertex in $\{s_i, \dots, s_{i+2m}\}$. Therefore, there are no multiple edges in our new graph and it is unambiguous to call s_{i-1} and S_i roommates as they share a solid edge.

¹⁶In the graph G , the only solid edge between our odd cycle and any vertex outside of the odd cycle is between s_{i-1} and s_i

Figure 7: If there is a dashed edge between s_{i-1} and a vertex in $\{s_i, \dots, s_{i+2m}\}$, there must be an alternating cycle. By definition, there can not be a dashed edge between s_{i-1} and s_i .



It is important to note that the graph $G' \setminus \{S_i\}$ is identical to the graph $G \setminus \{s_i, \dots, s_{i+2m}\}$. G' may contain a cut-vertex¹⁷, but we can find a subgraph of G' which contains no cut-vertex and is closed under roommates. Let $\mathcal{C}(s_{i-1})$ be s_{i-1} 's component in the graph $G' \setminus \{S_i\}$. s_{i-1} 's component is the set of all vertices s_{i-1} is connected to.

Claim: $\mathcal{C}(s_{i-1}) \cup \{S_i\}$ is closed under roommates and contains no cut-vertex.

Closed under Roommates:

s_{i-1} and S_i are roommates and both are in $\mathcal{C}(s_{i-1}) \cup \{S_i\}$. Any other vertex $v \in \mathcal{C}(s_{i-1}) \cup \{S_i\}$ is connected to s_{i-1} by definition of component and connected to $\mu(v)$ by its solid edge. Therefore $\mu(v)$ and s_{i-1} are connected as connected is a transitive relationship. Therefore, $\mu(v) \in \mathcal{C}(s_{i-1}) \cup \{S_i\}$.

No Cut-Vertex:

S_i can not be a cut-vertex as $\mathcal{C}(s_{i-1})$ is connected by the definition of component. Look at any $u \in \mathcal{C}(s_{i-1})$. We will show that u is not a cut-vertex by showing that all vertices in $\mathcal{C}(s_{i-1}) \cup \{S_i\} \setminus \{u\}$ are connected to S_i and therefore connected to each other. Look at any $v \in \mathcal{C}(s_{i-1}) \setminus \{u\}$. G contains no cut-vertex, so $G \setminus \{u\}$ is connected. Therefore, there is a path in $G \setminus \{u\}$ from v to every vertex in $\{s_i, \dots, s_{i+2m}\}$. Among all paths in $G \setminus \{u\}$ from v to any vertex in $\{s_i, \dots, s_{i+2m}\}$, let $\{v, p_1, \dots, p_n, t\}$ be one of minimal length. This implies that $\{v, p_1, \dots, p_n\} \cap \{s_i, \dots, s_{i+2m}\} = \emptyset$, since otherwise there would

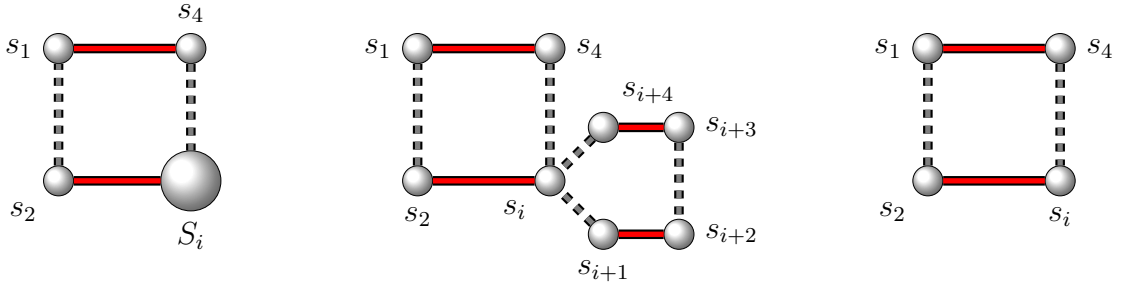
¹⁷In particular, S_i may be a cut-vertex. I would like to thank an anonymous referee for pointing this out.

exist a shorter path to a vertex in $\{s_i, \dots, s_{i+2m}\}$. Since $\{v, p_1, \dots, p_n\} \cap \{s_i, \dots, s_{i+2m}\} = \emptyset$, $\{v, p_1, \dots, p_n\}$ is a path in $G \setminus \{s_i, \dots, s_{i+2m}\} = G' \setminus \{S_i\}$. Therefore, $\{v, p_1, \dots, p_n\} \subseteq \mathcal{C}(s_{i-1})$ as they are all connected to $v \in \mathcal{C}(s_{i-1})$. Moreover, since p_n is adjacent to $t \in \{s_i, \dots, s_{i+2m}\}$ in G , p_n has a dashed edge with S_i in G' . So indeed, we can remove any u and still find a path that lies entirely in $\mathcal{C}(s_{i-1}) \cup \{S_i\} \setminus \{u\}$ from any v to S_i . Since all vertices in $\mathcal{C}(s_{i-1}) \cup \{S_i\} \setminus \{u\}$ are connected to S_i , all vertices are connected and u is not a cut-vertex. Therefore $\mathcal{C}(s_{i-1}) \cup \{S_i\}$ contains no cut-vertices.

$\mathcal{C}(s_{i-1}) \cup \{S_i\}$ must be non-trivial since s_{i-1} has a dashed edge to some vertex not in $\{s_i, \dots, s_{i+2m}\}$ ¹⁸. Moreover, it contains strictly fewer vertices than G as it is a subset of G' which is the result of contracting three or more vertices into one vertex. Therefore, by the inductive hypothesis $\mathcal{C}(s_{i-1}) \cup \{S_i\}$ contains an alternating cycle¹⁹.

Any alternating cycle that does not contain the contraction S_i is an alternating cycle in G . Any alternating cycle containing the contracted vertex S_i can be expanded to an alternating cycle in G . If we enter the blossom through the base vertex s_i then we can just replace S_i with s_i and we maintain an alternating cycle. See Figure 8 for an example.

Figure 8: Expanding a contraction if we enter the odd cycle through the base.



¹⁸Remember, we are assuming s_{i-1} does not have a dashed edge with any vertex in $\{s_i, \dots, s_{i+2m}\}$ or else we would have already found our alternating cycle.

¹⁹Proving existence by induction makes the proof cleaner, but it obscures the underlying algorithm for finding the alternating cycle. Now that we have contracted the odd cycle, we can simply continue alternating between solid and dashed edges until we cycle. If it is an odd cycle, we contract and continue. If it is an even cycle, we stop. Every time we contract we have strictly fewer vertices, so eventually the process must conclude with us finding an alternating-cycle

Suppose we enter the blossom through a vertex other than the base. Either the edge to the left or to the right is solid. We can follow the cycle in the direction of the solid edge all the way to base vertex. This is an alternating path to the base, the cycle connects to the base with a dashed edge, and then continues along the stem starting with a solid edge. So indeed, this expands an alternating cycle through S_i to an alternating cycle in G . See Figures 9 and 10 for examples.

Figure 9: Expanding a contraction if we enter the odd cycle on an odd vertex. Continue around the odd cycle counterclockwise.

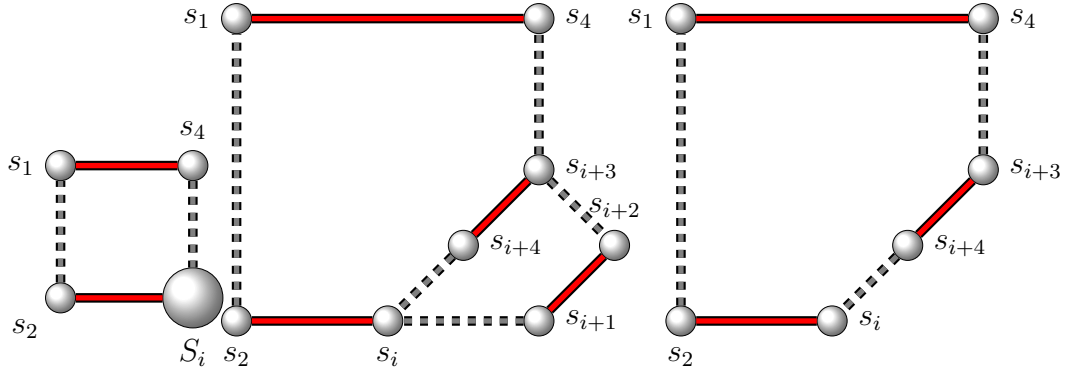
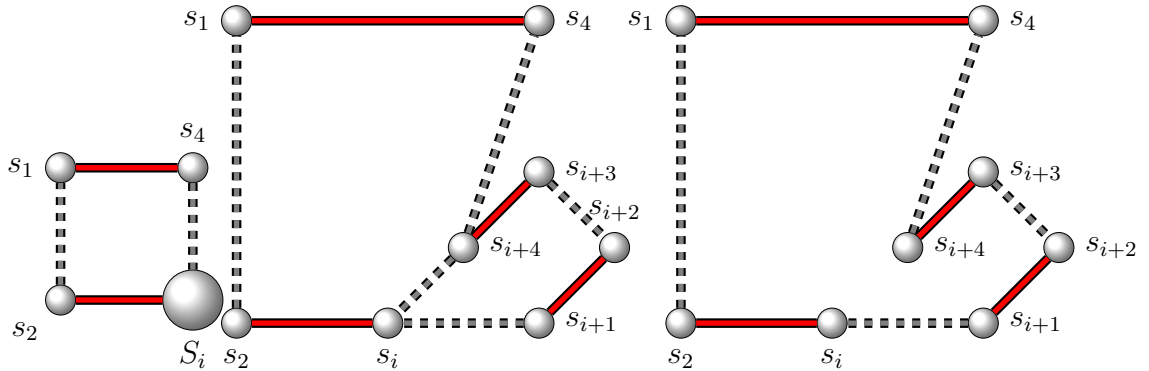


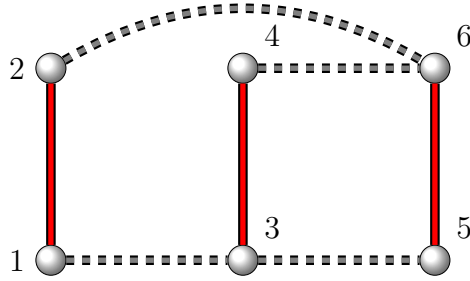
Figure 10: Expanding a contraction if we enter the odd cycle on an even vertex. Continue around the cycle clockwise.



□

Proposition 1 implies a simple procedure for determining whether an assignment is efficient: (1) Induce graph G_{ζ}^{μ} , (2) Iteratively prune G_{ζ}^{μ} until all

Figure 11: A non-trivial block closed under roommates, but s_1 and s_2 are not contained in any alternating-cycle.



blocks are closed under roommates, (3) If all blocks are trivial, then our assignment is efficient. If there exists a non-trivial block, then by Proposition 1 and Lemma 1, our assignment is inefficient. Moreover, the proof of Proposition 1 provides an algorithm for finding an alternating cycle when one exists.

The Roommate Swap determines if a given assignment is efficient. However, a particular student likely does not care whether the assignment can be Pareto improved. Rather, she would like to know if *she* can be part of a Pareto improvement. Unfortunately, Proposition 1 does not generalize to the statement if a student t is contained in a non-trivial block closed under roommates, then t is involved in a Pareto improvement. Figure 11 is a non-trivial block that is closed under roommates, but s_1 and s_2 are not part of any Pareto improvements.

The Roommate Swap does not determine if a particular student can improve her assignment. However, it is not biased. In particular, starting from any assignment, we can reach all assignments which Pareto dominate it. If we randomly choose the vertex we start with, and when we have a choice, we randomly choose which edge to continue on, then the Roommate Swap will find any Pareto improvement with probability that is uniformly bounded away from zero. Therefore, if the Roommate Swap is run repeatedly, it will determine if an individual student is involved in a Pareto improvement with probability one.

3.1 Applications

The practical importance of matching theory with an initial assignment is that it enables the market designer to extend any matching theory application to a dynamic environment. For example, in a kidney exchange, preference over a kidney change if tests determine the patient and donor are incompatible. The population being matched changes if a patient passes away. In school choice, a student's preferences over schools change if she moves within the school district. The population being matched changes if a student elects to attend a private school or moves to a different school district. In either instance, the initial assignment must be honored; however, after the primitives of the matching problem have changed, the initial assignment may no longer be Pareto optimal.

Roth, Sonmez, and Unver (2005) propose a mechanism for kidney exchange. In this setting an agent consists of a patient in need of a kidney along with her (incompatible) donor. For institutional reasons, exchanges are currently limited to two donor-recipient pairs. Since there is only one type of agent being matched, this is a Roommates Problem.

A patient needing a kidney may be incompatible with a donor if their blood-types are incompatible or if the patient has an antibody against one or more of the donor's proteins (this is called a positive crossmatch).²⁰ Blood type incompatibility follows a very organized structure: type O patients can only receive a kidney from another type O donors, type A patients can only receive a kidney from type O or type A donors, type B patients are compatible with type O and type B, and type AB patients can receive a kidney from all types. However, positive crossmatches are much less predictable as they result from individual donor proteins.

Positive crossmatches are a significant obstacle for any kidney exchange mechanism. A positive crossmatch cannot be determined from a database of donor-patient characteristics but rather must be tested for on an individual basis. After agents have been matched for an exchange, a test determines if there is a positive donor-patient crossmatch. Unfortunately, positive crossmatches are relatively common. For example, Zenios et al. (2001) estimate

²⁰Roth, Sonmez, and Unver 2007 includes a nice description of the factors behind kidney-exchange compatibility.

the probability two strangers will have a positive crossmatch to be 11 percent.²¹ Since each assignment requires two donor-patient pairs to be compatible, roughly 20 percent of the pairings in any initial assignment will be incompatible.

After an assignment is made, compatible matches cannot be forced to give up their initial assignment. Therefore, this is exactly the domain of the Roommate Swap. There exists a *status-quo* assignment which must be honored; however, after preferences over kidneys have been updated to reflect positive crossmatches, the initial assignment may not be Pareto-efficient. The Roommate Swap determines whether or not the initial kidney exchange assignment is still Pareto-optimal and Pareto-improves it if the assignment is not.

Matching subject to an initial assignment is relevant to two-sided applications as well. The Boston public school system uses the deferred acceptance algorithm to assign students to schools.²² In 2009 the assignment was made during the middle of March.²³ In between March and the start of the school year both the preferences of the students and the student body itself are subject to change. In either case, all students have the right to keep their initial assignment, but the initial assignment may no longer be Pareto optimal. Similarly, a college student may change her preferences over universities and wish to transfer. A doctor might change her opinion of a hospital after working there. In each instance, the initial assignment either must be honored or is not easily severed but may be Pareto improved.

²¹The probability of a crossmatch is difficult to measure precisely. Roth et al. (2005) used the characteristics of the current waiting list to indicate the scope of the problem. The OPTN/SRTR database categorizes participants as either low, medium, or high PRA (Percent Reactive Antibody). Low PRA patients have a positive crossmatch with less than 10 percent of the population, medium have a positive crossmatch with 10 to 80 percent, and high percent have a positive crossmatch with more than 80 percent of the population. Roth et al (2005) reports that the distribution of patients on the OPTN/SRTR registry over the years from 1993 to 2002 was 70.19 percent low, 20 percent medium PRA, and 9.81 percent high PRA.

²²See Abdulkadiroglu and Sonmez (2003), Abdulkadiroglu, Pathak, Roth and Sonmez (2005), and Pathak and Sonmez (2006)

²³From <http://Bostonpublicschool.org/register> accessed May 8, 2009.

4 Strategic Implications

In this section I examine the strategic implications of any mechanism for selecting a core assignment of the coalitional game μ . Unfortunately, although not surprisingly, I obtain a negative result: there does not exist a mechanism for selecting a core assignment where truthfully reporting one's preferences is always a dominant strategy. This result follows very closely the results for two-sided matching theory presented in Roth and Sotomayor (1990).

Following the matching literature, I will use dominant strategy as my solution concept.

Definition 2. *A **dominant strategy** is a strategy that is a best response to all possible strategies of the other agents. An assignment mechanism is **strategy proof** if it is a dominant strategy for each agent to reveal her preferences truthfully.*

Lemma 3. *There does not exist a strategy-proof mechanism for selecting a Pareto improvement of an inefficient assignment.*

Lemma 3 is proved in the appendix. This is quite a general result, but it is rather easy to prove. A strategy-proof mechanism must be able to handle any initial assignment and any profile of preferences. Following the path of Roth (1982), I demonstrate a case that no mechanism is able to handle.

5 Conclusion

This paper studies a new class of matching problems: matchings subject to an initial assignment. When bilateral approval is required to dissolve a partnership, the core consists of Pareto optimal matches. A particularly appealing example, and the focus of this paper, is the roommates problem. After an assignment has been made, roommates have common rights to the room. When the assignment cannot be changed without mutual consent, Pareto optimality and not stability defines the core.

The key practical contribution of this paper is to introduce a fast algorithm that determines whether or not a particular roommates assignment is Pareto

efficient. If a given assignment is not in the core, the algorithm finds a core assignment. This has an important application to kidney exchange as it enables any assignment mechanism to respond to positive crossmatches.

6 Appendix

Lemma 4. *There are $\frac{(2N)!}{2^N(N!)}$ = $(2N - 1)(2N - 3)(2N - 5) \cdots (3)(1)$ many ways to assign $2N$ students to be roommates.*

Proof. The proof is by induction. When $N = 1$, the result is trivial as there is only one way to assign two students to be roommates. Assume $N > 1$ and by induction there are $(2N - 3)(2N - 5) \cdots (3)(1)$ many ways to assign $2(N-1)$ many students to be roommates. Select a student s . There are $2N-1$ possible roommates for s , and by assumption, for any roommate we pick, there are $(2N - 3)(2N - 5) \cdots (3)(1)$ many ways to assign the remaining $2N-2$ many students. Therefore, there is a total of $[2N - 1] \times [(2N - 3)(2N - 5) \cdots (3)(1)]$ many ways of assigning roommates. \square

Lemma 3 *There does not exist a strategy-proof mechanism for selecting a Pareto improvement of an inefficient assignment.*

Proof. Suppose there are four students, a , b , c , and d , and an initial assignment, μ_1 pairing a with b and c with d . Moreover, suppose the student's preferences are as follows.

$$\begin{aligned} a : c \succ d \succ b \\ b : c \succ d \succ a \\ c : b \succ a \succ d \\ d : b \succ a \succ c \end{aligned}$$

With four students, there are three possible assignments. Note that an assignment is completely determined by who a (or any other student) is assigned to. Let μ_2 denote the assignment where a is paired with c and μ_3 denote the assignment where a is paired with d . In our original assignment

μ_1 , each person is paired with their least preferred roommate, so μ_1 is Pareto dominated by both of the other assignments. Suppose for contradiction that there exists a strategy-proof mechanism M for selecting an efficient, Pareto improving assignment. Note that if a submits the preferences $c \succ b \succ d$ and all other students submit true preferences, then μ_2 is the only assignment that Pareto improves μ_1 (relative to the submitted preferences). In such a case, M must select μ_2 . Similarly, if b submits the preferences $c \succ a \succ d$ and all other students submit true preferences, then M must select μ_3 as it is now the only Pareto improving assignment. When all students submit true preferences, M must select either μ_2 or μ_3 . If M selects μ_2 , then b can do better by deviating and submitting the preferences $c \succ a \succ d$. If M selects μ_3 , then a can do better by submitting preferences $c \succ b \succ d$. Either way, M is not strategy proof which is a contradiction. \square

6.1 Computational Complexity

The purpose of this section is to demonstrate that the Roommate Swap is a polynomial time algorithm. I demonstrate that it is at worst an $O(N^3)$ algorithm where N is the number of students.

Each iteration of the algorithm involves the following steps, performed in sequence:

1. *Induce the graph.* This is at worst $O(N^2)$ as a graph is defined by its edges and there are at most $\frac{N(N-1)}{2}$ many edges.
2. *Iteratively prune the graph until all blocks are closed under roommates.* West (2001), pg. 157, details an $O(N)$ algorithm for determining blocks. We need to iterate at most N times as each iteration eliminates at least one student from each block or stops looking at a block if it is already closed under roommates. Therefore iteratively pruning the graph is at worst $O(N^2)$.
3. *Find an alternating-cycle.* This process is $O(N)$. At each step we either travel to a previously unvisited vertex, which we can do at most N times, or contract a minimum of two vertices, which we can do at most $\frac{N}{2}$ times. So the algorithm must conclude in at most $N + \frac{N}{2}$ steps.

As it takes at most N steps to expand a cycle containing super-vertices to a proper cycle, finding an alternating-cycle concludes in $O(N)$ time.

Therefore each iteration is $O(N^2)$.

Observation 1. *In each iteration of the Roommate Swap, at least one student can be reassigned her top achievable student.*

Proof. The search process ends with an alternating-cycle that may or may not contain super-vertices. Choose the dashed edges from standard vertices to be the vertex's most preferred student among those who prefer her to their current assignment. Therefore, if the alternating cycle contains no super-vertices, then half the students receive their top achievable match. A grey edge from a super-vertex is not necessarily the student's most preferred achievable student. However, if the alternating-cycle contains a super-vertex and we need to expand our contractions, then there must be a *last* odd-cycle that needs to be expanded.

Figure 12: An odd cycle.

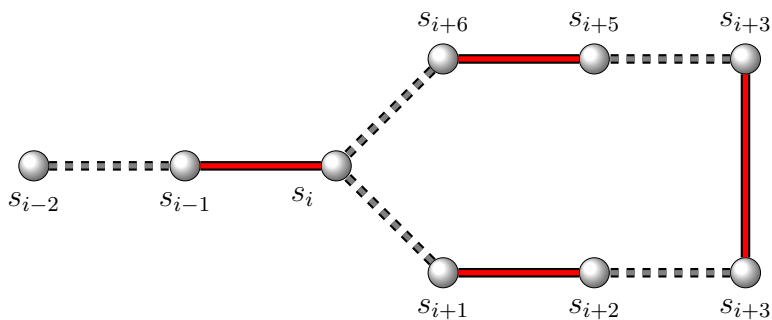


Figure 12 shows a last cycle with six vertices, but the analysis is the same for fewer or greater vertices. Our alternating path must go through s_i and either s_{i+1} or s_{i+6} . None of these edges involve super-vertices (this is our last expansion) so by construction, s_{i+1} is s_i 's top achievable student and s_i is s_{i+6} 's top achievable choice. Either way, at least one student receives her top achievable choice.

□

The significance of this is that once a student has been assigned her top achievable choice, neither she nor her roommate can ever be involved in another Pareto improvement. Therefore we can eliminate them both from consideration. Since we eliminate at least two students after every iteration, there can be at most $\frac{N}{2}$ iterations.

The algorithm performs $O(N)$ many iterations of an $O(N^2)$ process. Therefore it is, at worst, $O(N^3)$.

References

- [1] Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., and Sonmez, T. (2005), “The Boston Public School Match,” *American Economic Review Papers & Proceedings* 95-2: 368-371.
- [2] Abdulkadiroglu, A. and Sonmez, T. (1998), “Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems,” *Econometrica* 66: 689-701.
- [3] Abdulkadiroglu, A. and Sonmez, T. (2003), “School Choice: A Mechanism Design Approach,” *American Economic Review*, 93: 729-747.
- [4] Abraham, D. J. and Manlove, D. F. (2004), “Pareto Optimality in the Roommates Problem,” mimeo, University of Glasgow.
- [5] Chung, K. (2000), “On the Existence of Stable Roommate Matchings,” *Games and Economic Behavior* 33: 206-230.
- [6] Edmonds, J. (1965), “Paths, Trees, and Flowers,” *Canad. J. Math.* 17: 449-467.
- [7] Erdil, A. and Ergin, H. (2008), “What’s the Matter with Tie-Breaking? Improving Efficiency in School Choice?” *American Economic Review* 98-3, 669-689.
- [8] Gale, D. and Shapley, L. (1962), “College Admissions and the Stability of Marriage,” *Amer. Math. Monthly* 69: 9-15.

- [9] Gusfield, D. and Irving, R. (1989), “The Stable Marriage Problem: Structure and Algorithms,” MIT Press, Boston, MA.
- [10] Pathak, P. A. and Sonmez, T. (2006) “Leveling the Playing Field: Sincere and Strategic Players in the Boston Mechanism,” *American Economic Review* forthcoming.
- [11] Roth, A. E., Sonmez, T., and Unver, U. (2005). “Pairwise Kidney Exchange,” *Journal of Economic Theory* 125: 151-188.
- [12] Roth, A. E., and Sotomayor, M. (1990). “Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis,” Econometric Society Monograph 18, Cambridge Univ. Press, Cambridge.
- [13] Segev, D. L., Sommer, E. G., Warren, D. S., et al. (2005). “Kidney Paired Donation and Optimizing the Use of Live Donor Organs,” *JAMA* 293(15): 1883-1890.
- [14] Shapley, L. and Scarf, H. (1974), “On Cores and Indivisibility,” *Journal of Mathematical Economics* 1: 23-28.
- [15] Tan, J. J. M. (1991). A Necessary and Sufficient Condition for the Existence of a Complete Stable Matching, *J. Algorithms* 12: 154-178.
- [16] West, D. B. (2001). “Introduction to Graph Theory,” Prentice Hall, Englewood Cliffs.
- [17] Zenios, S. A., Woodle, E. S., Ross, L. F., (2001) “Primum Non Nocere: Avoiding Harm to Vulnerable Wait List Candidates in an Indirect Kidney Exchange.” *Transplantation* 72:648-654.