An Analysis of Top Trading Cycles in Two-Sided Matching Markets

Yeon-Koo Che* Olivier Tercieux†

July 30, 2015

Preliminary and incomplete.

Abstract

We study top trading cycles in a two-sided matching environment (Abdulkadiroglu and Sonmez (2003)) under the assumption that individuals' preferences and objects' priorities are drawn *iid* uniformly. The distributions of agents' preferences and objects' priorities remaining after a given round of TTC depend nontrivially on the exact history of the algorithm up to that round (and so need not be uniform *iid*). Despite the nontrivial history-dependence of evolving economies, we show that the *number of individuals/objects assigned* at each round follows a simple Markov chain and we explicitly derive the transition probabilities.

JEL Classification Numbers: C70, D47, D61, D63.
Keywords: Random matching markets, Markov property.

1 Introduction

Top Trading Cycles (TTC) algorithm, introduced by Abdulkadiroglu and Sonmez (2003) in a two-sided matching environment has been an influential method for achieving efficient outcomes in particular in school choice environments. For instance, TTC was used until recently in New Orleans school systems for assigning students to public high schools and

*Department of Economics, Columbia University, USA. Email: yeonkooche@gmail.com.
†Department of Economics, Paris School of Economics, France. Email: tercieux@pse.ens.fr.

recently San Francisco school system announced plans to implement a top trading cycles mechanism. A generalized version of TTC is also used for kidney exchange among donor-patient pairs with incompatible donor kidneys (see Sonmez and Unver (2011)).

The TTC algorithm proceeds in multiple rounds as follows: In Round $t = 1, ...,$ each individual $i \in I$ points to his most preferred object (if any). Each object $o \in O$ points to the individual who has the highest priority at that object. Since the number of individuals and objects are finite, the directed graph thus obtained has at least one cycle. Every individual who belongs to a cycle is assigned the object he is pointing at. Any individuals and objects that are assigned are then removed. The algorithm terminates when all individuals have been assigned; otherwise, it proceeds to Round $t + 1$.

In the sequel, we shall simply consider an **random market** consisting of a set I of agents and a set O of objects such that *the preferences of each side with respect to the other side are drawn iid uniformly.* The difficulty with the analysis of TTC in this random environment stems from the fact that the preferences of the agents and objects remaining after the first round of TTC need not be uniform, with their distributions affected nontrivially by the realized event of the first round of TTC, and the nature of the conditioning is difficult to analyze.[1] The current paper shows that, even though the exact composition of cycles are subject to the conditioning issue, the number of agents assigned in each round follows a Markov chain, and is thus free from the conditioning issue. Indeed, our main result stated below states that the numbers of agents and objects that are assigned in each round of TTC follow a simple Markov chain depending *only* on the numbers of agents and objects at the beginning of that round. It also characterizes the probability structure of the Markov chain. This implies that there are no conditioning issues at least with respect to the *total* numbers of agents and objects that are assigned in each round of TTC. Namely, one does not need to keep track of the precise history leading up to a particular economy at the beginning of a round, as far as the numbers of objects assigned in that round is concerned.

[1] To see this, assume that the set of agents and objects have the same size n and that they are indexed from $1, \ldots, n$. Observe first that in Round 1 of TTC, each pair of an individual and an object has probability $1/n^2$ to form a cycle of order 2. Since there are n^2 such pairs, at Round 1, the expected number of cycles of order 2 is 1. Now, to see where the conditioning issue comes from, consider the event that at Round 1 of TTC, each object points to the individual with the same index while each individual with index $k \leq n-1$ points to the object with index $k+1$. Finally assume that individual n points to object n. Given this event, observe that at Round 1 a single cycle clears and it only involves the individual and the object with index n. Thus, conditionally on this event, the expected number of cycles of order 2 in Round 2 is much smaller than 1. Indeed, in Round 2, only individual $n-1$ can be part of a cycle of order 2 and the only way for this to happen is for individual $n-1$ to point to object $n-1$. This occurs with probability $1/(n-1)$ and so the expected number of cycles of order 2 goes to 0 as n grows.

The formal statement is as follows:

Theorem 1. *Suppose any round of TTC begins with n agents and o objects remaining in the market. Then, the probability that there are $m \le \min\{o, n\}$ agents assigned at the end of that round is*

$$p_{n,o;m} = \left(\frac{m}{(on)^{m+1}}\right)\left(\frac{n!}{(n-m)!}\right)\left(\frac{o!}{(o-m)!}\right)(o+n-m).$$

Thus, denoting n_i and o_i the number of individuals and objects remaining in the market at any round i, the random sequence (n_i, o_i) is a Markov chain.

Beyond the technical contribution, this result can be useful for several purposes. First, we believe the result can be useful in order to analyze the distribution of the number of steps needed for TTC to converge. Because of this, one may expect that this result will help to get a better understanding of the distribution of ranks of individuals/objects under TTC which is directly welfare relevant. Second, even though the random environment we consider does not allow for correlation in agents' preferences, as we show in Che and Tercieux (2015), this result can actually be used in order to make interesting predictions in richer environments where agents' preferences are positively correlated. Indeed, appealing to Theorem 1, Che and Tercieux (2015) show that the fraction of agents/objects assigned via cycles of order strictly greater than 2 does not converge in probability to 0. In the environment where agents have positively correlated preferences, they show that this can be used in order to prove that with probability bounded away from 0 there will be a significant number of pairs of agents and objects who would significantly benefit from being matched together rather than with their partners given by TTC.

Finally, we note that Theorem 1 parallels the corresponding result by Frieze and Pittel (1995) on the Shapley-Scarf version of TTC. The difference between the two versions of TTC is not trivial, so their proofs do not carry over. Section 2 below is devoted to the proof of this result while Section 3 discusses some of implication of the main result.

2 Proof of Theorem 1

Consider any two finite sets I and O, with cardinalities $|I| = n, |O| = o$. A **bipartite digragh** $G = (I \times O, E)$ consists of vertices I and O on two separate sides and directed edges $E \subset (I \times O) \cup (O \times I)$, comprising ordered pairs of the form (i, o) or (o, i) (corresponding to edge originating from i and pointing to o and an edge from o to i, respectively). A **rooted**

tree is a bipartite digraph where all vertices have out-degree 1 except the root which has out-degree 0.[2] A **rooted forest** is a bipartite graph which consists of a collection of disjoint rooted trees. A **spanning rooted forest over** $I \cup O$ is a forest comprising vertices $I \cup O$. From now on, a spanning forest will be understood as being over $I \cup O$.

We begin by noting that TTC induces a random sequence of spanning rooted forests. Indeed, one could see the beginning of the first round of TTC as a situation where we have the trivial forest consisting of $|I| + |O|$ trees with isolated vertices. Within this step each vertex in I will randomly point to a vertex in O and each vertex in O will randomly point to a vertex in I. Note that once we delete the realized cycles, we again get a spanning rooted forest. So we can think again of the beginning of the second round of TTC as a situation where we start with a spanning rooted forest where the agents and objects remaining from the first round form this spanning rooted forest, where the roots consist of those agents and objects that had pointed to the entities that were cleared via cycles. Here again objects that are roots randomly point to a remaining individual and individuals that are roots randomly point to a remaining object. Once cycles are cleared we again obtain a forest and the process goes on like this.

Formally, the random sequence of forests, $F_1, F_2, ...$ is defined as follows. First, we let F_1 be a trivial unique forest consisting of $|I|+|O|$ trees with isolated vertices, forming their own roots. For any $i = 2, ...$, we first create a random directed edge from each root of F_{i-1} to a vertex on the other side, and then delete the resulting cycles (these are the agents and objects assigned in round $i-1$) and F_i is defined to be the resulting rooted forest.

2.1 Preliminaries

We shall later use the following lemma, which characterizes the number of spanning rooted forests.

LEMMA 1 (Jin and Liu (2004)). *Let $V_1 \subset I$ and $V_2 \subset O$ where $|V_1| = \ell$ and $|V_2| = k$. The number of spanning rooted forests having k roots in V_1 and ℓ roots in V_2 is $f(n, o, k, \ell) := o^{n-k-1} n^{o-\ell-1}(\ell n + ko - k\ell)$.*

For the next result, consider agents I' and objects O' such that $|I'| = |O'| = m > 0$. We say a mapping $f = h \circ g$ is a **bipartite bijection**, if $g : I' \to O'$ and $h : O' \to I'$ are

[2]Sometimes, a tree is defined as an acyclic undirected connected graph. In such a case, a tree is rooted when we name one of its vertex a "root." Starting from such a rooted tree, if all edges now have a direction leading toward the root, then the out-degree of any vertex (except the root) is 1. So the two definitions are actually equivalent.

both bijections. A **cycle** of a bipartite bijection is a cycle of the induced digraph. Note that a bipartite bijection consists of disjoint cycles. A **random bipartite bijection** is a (uniform) random selection of a bipartite bijection from the set of all bipartite bijections. The following result will prove useful for a later analysis.

LEMMA 2. *Fix sets I' and O' with $|I'| = |O'| = m > 0$, and a subset $K \subset I' \cup O'$, containing $a \geq 0$ vertices in I' and $b \geq 0$ vertices in O'. The probability that each cycle in a random bipartite bijection contains at least one vertex from K is*

$$\frac{a+b}{m} - \frac{ab}{m^2}.$$

PROOF. We begin with a few definitions. A **permutation** of X is a bijection $f : X \to X$. A **cycle** of a permutation is a cycle of the digraph induced by the permutation. A permutation consists of disjoint cycles. A **random permutation** chooses uniform randomly a permutation f from the set of all possible permutations. Our proof will invoke following result:

FACT 1 (Lovasz (1979) Exercise 3.6). *The probability that each cycle of a random permutation of a finite set X contains at least one element of a set $Y \subset X$ is $|Y|/|X|$.*

To begin, observe first that a bipartite bijection $h \circ g$ induces a permutation of set I'. Thus, a random bipartite bijection defined over $I' \times O'$ induces a random permutation of I'. To compute the probability that each cycle of a random bipartite bijection $h \circ g$ contains at least one vertex in $K \subset I' \times O'$, we shall apply Fact 1 to this induced random permutation of I'.

Indeed, each cycle of a random bipartite bijection contains at least one vertex in $K \subset I' \times O'$ if and only if each cycle of the induced random permutation of I' contains either a vertex in $K \cap I'$ or a vertex in $I' \setminus K$ that points to a vertex in $K \cap O'$ in the original random bipartite bijection. Hence, the relevant set $Y \subset I'$ for the purpose of applying Fact 1 is a random set that contains $|K \cap I'| = a$ vertices of the former kind and Z vertices of the latter kind.

The number Z is random and takes a value z, $\max\{b-a, 0\} \leq z \leq \min\{m-a, b\}$, with probability:

$$\Pr\{Z = z\} = \frac{\binom{m-a}{z}\binom{a}{b-z}}{\binom{m}{b}}.$$

This formula is explained as follows. $\Pr\{Z = z\}$ is the ratio of the number of bipartite bijections having exactly z vertices in $I' \setminus K$ pointing toward $K \cap O'$ to the total number of bipartite bijections.

Note that since we consider bipartite bijections, the number of vertices in I' pointing to the vertices in $K \cap O'$ must be equal to b. Focusing first on the numerator, we have to compute the number of bipartite bijections having exactly z vertices in $I' \setminus K$ pointing toward $K \cap O'$ and the remaining $b - z$ vertices pointing to the remaining $K \cap O'$. There are $\binom{m-a}{z}\binom{a}{b-z}$ ways one can choose z vertices from $I' \setminus K$ and $b - z$ vertices from $K \cap I'$. Thus, the total number of bipartite bijections having exactly z vertices in $I' \setminus K$ that point to $K \cap O'$ is $\binom{m-a}{z}\binom{a}{b-z}\upsilon$, where υ is the total number of bipartite bijections in which the b vertices thus chosen point to the vertices in $K \cap O'$. This gives us the numerator. As for the denominator, the total number of bipartite bijections having b vertices in I' pointing to $K \cap O'$ is $\binom{m}{b}$ (the number of ways b vertices are chosen from I'), multiplied by υ (the number of bijections in which the b vertices thus chosen point to the vertices in $K \cap O'$). Hence, the denominator is $\binom{m}{b}\upsilon$. Thus, we get the above formula.

Recall our goal is to compute the probability that each cycle of the random permutation induced by the random bipartite bijection contains at least one vertex in the random set Y, with $|Y| = a + Z$, where $\Pr\{Z = z\} = \frac{\binom{m-a}{z}\binom{a}{b-z}}{\binom{m}{b}}$. Applying Fact 1, then the desired probability is

$$\begin{aligned}
\mathbb{E}\left[\frac{|Y|}{|I'|}\right] &= \sum_{z=\max\{b-a,0\}}^{\min\{m-a,b\}} \Pr\{Z = z\}\frac{a+z}{m} \\
&= \frac{a}{m} + \sum_{z=\max\{b-a,0\}}^{\min\{m-a,b\}} \Pr\{Z = z\}\frac{z}{m} \\
&= \frac{a}{m} + \sum_{z=\max\{b-a,0\}}^{\min\{m-a,b\}} \frac{\binom{m-a}{z}\binom{a}{b-z}}{\binom{m}{b}}\left(\frac{z}{m}\right) \\
&= \frac{a}{m} + \left(\frac{m-a}{m\binom{m}{b}}\right) \sum_{z=\max\{b-a,1\}}^{\min\{m-a,b\}} \binom{a}{b-z}\binom{m-a-1}{z-1} \\
&= \frac{a}{m} + \left(\frac{m-a}{m\binom{m}{b}}\right)\binom{m-1}{b-1} \\
&= \frac{a}{m} + \frac{b(m-a)}{m^2} \\
&= \frac{a+b}{m} - \frac{ab}{m^2},
\end{aligned}$$

where the fifth equality follows from Vandermonde's identity. □

2.2 Markov Chain Property of TTC

For any rooted forest F_i, let $N_i = I_i \cup O_i$ be its vertex set and $k_i = (k_i^I, k_i^O)$ be the vector denoting the numbers of roots on both sides, and use (N_i, k_i) to summarize this information. And let $\mathcal{F}_{N_i,k_i}$ denote the set of all rooted forests having N_i as the vertex set and k_i as the vector of its root numbers.

LEMMA 3. *Given $(N_j, k_j), j = 1, ..., i$, every (rooted) forest of $\mathcal{F}_{N_i,k_i}$ is equally likely.*

PROOF. We prove this result by induction on i. Since for $i = 1$, by construction, the trivial forest is the unique forest which can occur, this is trivially true for $i = 1$. Fix $i \geq 2$, and assume our statement is true for $i - 1$.

Fix $N_i = I_i \cup O_i \subset N_{i+1} = I_{i+1} \cup O_{i+1}$, and k_i and k_{i+1}. For each forest $F \in \mathcal{F}_{N_{i+1},k_{i+1}}$, we consider a possible pair (F', ϕ) that could have given rise to F, where $F' \in \mathcal{F}_{N_i,k_i}$ and ϕ maps the roots of F' in I_i to its vertices in O_i as well as the roots of F' in O_i to its vertices in I_i. In words, such a pair (F', ϕ) corresponds to a set N_i of agents and objects remaining at the beginning of round i of TTC, of which k_i^I agents of I_i and k_i^O objects have lost their favorite parties (and thus they must repoint to new partners in N_i under TTC in round i), and the way in which they repoint to the new partners under TTC in round i causes a new forest F to emerge at the beginning of round $i + 1$ of TTC. There are typically multiple such pairs that could give rise to F.

We start by showing that each forest $F \in \mathcal{F}_{N_{i+1},k_{i+1}}$ arises from the same number of such pairs—i.e., that the number of pairs (F', ϕ), $F' \in \mathcal{F}_{N_i,k_i}$, causing F to arise does not depend on the particular $F \in \mathcal{F}_{N_{i+1},k_{i+1}}$. To this end, for any given $F \in \mathcal{F}_{N_{i+1},k_{i+1}}$, we construct all such pairs by choosing a quadruplet (a, b, c, d) of four non-negative integers with $a + c = k_i^I$ and $b + d = k_i^O$,

(i) choosing c old roots from I_{i+1}, and similarly, d old roots from O_{i+1},

(ii) choosing a old roots from $I_i \backslash I_{i+1}$ and similarly, b old roots from $O_i \backslash O_{i+1}$,

(iii) choosing a partition into cycles of $N_i \backslash N_{i+1}$, each cycle of which contains at least one old root from (ii),[3]

[3] Within round i of TTC, one cannot have a cycle creating only with nodes that are not roots in the forest obtained at the beginning of round i. This is due to the simple fact that a forest is an acyclic graph. Thus, each cycle creating must contain at least one old root. Given that, by definition, these roots are eliminated from the set of available nodes in round $i+1$, these old roots that each cycle must contain must be from (ii).

(iv) choosing a mapping of the $k^I_{i+1}+k^O_{i+1}$ new roots to $N_i \backslash N_{i+1}$.[4]

Clearly, the number of pairs (F', ϕ), $F' \in \mathcal{F}_{N_i,k_i}$, satisfying the above restrictions depends only on $|I_i|$, $|O_i|$, k_i, k_{i+1}, and $|N_{i+1}|-|N_i|$.[5] We denote the number of such pairs by $\beta(|I_i|,|O_i|,k_i;|N_{i+1}|-|N_i|,k_{i+1})$. Let $\phi_i=(\phi^I_i,\phi^O_i)$ where ϕ^I_i is the random mapping from the roots of F_i in I_i to O_i and ϕ^O_i is the random mapping from the roots of F_i in O_i to I_i. Let $\phi=(\phi^I,\phi^O)$ be a generic mapping of that sort. Since, conditional on $F_i=F'$, the mappings ϕ^I_i and ϕ^O_i are uniform, we get

$$\Pr(F_{i+1}=F|F_i=F') = \frac{1}{|O_i|^{k^I_i}}\frac{1}{|I_i|^{k^O_i}}\sum_{\phi}\Pr(F_{i+1}=F|F_i=F',\phi_i=\phi). \tag{1}$$

Therefore, we obtain

$$\begin{aligned}
&\Pr(F_{i+1}=F|(N_1,k_1),...,(N_i,k_i))\\
=&\sum_{F'\in\mathcal{F}_{N_i,k_i}}\Pr(F_{i+1}=F,F_i=F'|(N_1,k_1),...,(N_i,k_i))\\
=&\sum_{F'\in\mathcal{F}_{N_i,k_i}}\Pr(F_{i+1}=F|(N_1,k_1),...,(N_i,k_i),F_i=F')\Pr(F_i=F'|(N_1,k_1),....,(N_i,k_i))\\
=&\frac{1}{|\mathcal{F}_{N_i,k_i}|}\sum_{F'\in\mathcal{F}_{N_i,k_i}}\Pr(F_{i+1}=F|F_i=F')\\
=&\frac{1}{|\mathcal{F}_{N_i,k_i}|}\sum_{F'\in\mathcal{F}_{N_i,k_i}}\frac{1}{|O_i|^{k^I_i}}\frac{1}{|I_i|^{k^O_i}}\sum_{\phi}\Pr(F_{i+1}=F|F_i=F',\phi_i=\phi)\\
=&\frac{1}{|\mathcal{F}_{N_i,k_i}|}\frac{1}{|O_i|^{k^I_i}}\frac{1}{|I_i|^{k^O_i}}\sum_{F'\in\mathcal{F}_{N_i,k_i}}\sum_{\phi}\Pr(F_{i+1}=F|F_i=F',\phi_i=\phi)\\
=&\frac{1}{|\mathcal{F}_{N_i,k_i}|}\frac{1}{|O_i|^{k^I_i}}\frac{1}{|I_i|^{k^O_i}}\beta(|I_i|,|O_i|,k_i;|N_{i+1}|-|N_i|,k_{i+1}),
\end{aligned} \tag{2}$$

where the third equality follows from the induction hypothesis and the Markov property of $\{F_j\}$, the fourth follows from (1), and the last follows from the definition of β and from the fact that the conditional probability in the sum of the penultimate line is 1 or 0, depending upon whether the forest F arises from the pair $(F^{'},\phi)$ or not. Note that this probability is

[4] Since, by definition, any root in $F\in\mathcal{F}_{N_{i+1},k_{i+1}}$ does not point, this means that, in the previous round, this node was pointing to another node which was eliminated at the end of that round.

[5] Recall that by definition of TTC, whenever a cycle creates, the same number of individuals and objects must be eliminated in this cycle. Hence, $|O_{i+1}|-|O_i|=|I_{i+1}|-|I_i|$ and $|N_{i+1}|-|N_i|=2|I_{i+1}|-|I_i|$.

independent of $F \in \mathcal{F}_{N_{i+1},k_{i+1}}$. Hence,

$$\begin{aligned}&\Pr(F_{i+1}=F|(N_1,k_1),...,(N_i,k_i),(N_{i+1},k_{i+1}))\\=&\frac{\Pr(F_{i+1}=F|(N_1,k_1),...,(N_i,k_i))}{\Pr(F_{i+1}\in\mathcal{F}_{N_{i+1},k_{i+1}}|(N_1,k_1),...,(N_i,k_i))}\\=&\frac{\Pr(F_{i+1}=F|(N_1,k_1),...,(N_i,k_i))}{\sum_{\tilde{F}\in\mathcal{F}_{N_{i+1},k_{i+1}}}\Pr(F_{i+1}=\tilde{F}|(N_1,k_1),...,(N_i,k_i))}=\frac{1}{|\mathcal{F}_{N_{i+1},k_{i+1}}|},\end{aligned}\tag{3}$$

which proves that, given $(N_j,k_j), j=1,...,i$, every rooted forest of $\mathcal{F}_{N_i,k_i}$ is equally likely. □

The next lemma then follows easily.

LEMMA 4. *Random sequence (N_i,k_i) forms a Markov chain.*

PROOF. By (2) we must have

$$\begin{aligned}\Pr((N_{i+1},k_{i+1})|(N_1,k_1),....,(N_i,k_i))&=\sum_{F\in\mathcal{F}_{N_{i+1},k_{i+1}}}\Pr(F_{i+1}=F|(N_1,k_1),....,(N_i,k_i))\\&=\sum_{F\in\mathcal{F}_{N_{i+1},k_{i+1}}}\frac{1}{|\mathcal{F}_{N_i,k_i}|}\frac{1}{|O_i|^{k_i^I}}\frac{1}{|I_i|^{k_i^O}}\beta(|I_i|,|O_i|,k_i;|N_{i+1}|-|N_i|,k_{i+1}).\end{aligned}$$

Observing that the conditional probability depends only on (N_{i+1},k_{i+1}) and (N_i,k_i), the Markov chain property is established. □

The proof of Lemma 4 reveals in fact that the conditional probability of (N_{i+1},k_{i+1}) depends on N_i) only through its cardinalities $(|I_i|,|O_i|)$, leading to the following conclusion. Let $n_i := |I_i|$ and $o_i := |O_i|$.

COROLLARY 1. *Random sequence $\{(n_i,o_i,k_i^I,k_i^O)\}$ forms a Markov chain.*

PROOF. By symmetry, given $(n_1,o_1,k_1^I,k_1^O),...,(n_i,o_i,k_i^I,k_i^O)$, the set (I_i,O_i) is chosen uniformly at random among all the $\binom{n}{n_i}\binom{o}{o_i}$ possible sets. Hence,

$$
\begin{aligned}
&\Pr((n_{i+1}, o_{i+1}, k^I_{i+1}, k^O_{i+1})|(n_1, o_1, k^I_1, k^O_1), ..., (n_i, o_i, k^I_i, k^O_i))\\
=&\sum_{(I_i,O_i):|I_i|=n_i,|O_i|=o_i} \Pr\{(n_{i+1}, o_{i+1}, k^I_{i+1}, k^O_{i+1})|(n_1, o_1, k^I_1, k^O_1), ..., (n_i, o_i, k^I_i, k^O_i), (I_i, O_i)\}\\
&\qquad \times \Pr\left\{(I_i, O_i) \mid (n_1, o_1, k^I_1, k^O_1), ..., (n_i, o_i, k^I_i, k^O_i)\right\}\\
=&\left(\sum_{(I_i,O_i):|I_i|=n_i,|O_i|=o_i} \Pr\{(n_{i+1}, o_{i+1}, k^I_{i+1}, k^O_{i+1})|(n_1, o_1, k^I_1, k^O_1), ..., (I_i, O_i, k^I_i, k^O_i)\}\right) \frac{1}{\binom{n}{n_i}\binom{o}{o_i}}\\
=&\left(\sum_{\substack{(I_i,O_i):|I_i|=n_i,|O_i|=o_i\\(I_{i+1},O_{i+1}):|I_{i+1}|=n_{i+1},|O_{i+1}|=o_{i+1}}} \Pr\{(I_{i+1}, O_{i+1}, k^I_{i+1}, k^O_{i+1})|(n_1, o_1, k^I_1, k^O_1), ..., (I_i, O_i, k^I_i, k^O_i)\}\right)\\
&\quad \times \frac{1}{\binom{n}{n_i}\binom{o}{o_i}}\\
=&\frac{1}{\binom{n}{n_i}\binom{o}{o_i}} \sum_{\substack{(I_i,O_i):|I_i|=n_i,|O_i|=o_i\\(I_{i+1},O_{i+1}):|I_{i+1}|=n_{i+1},|O_{i+1}|=o_{i+1}}} \Pr\{(I_{i+1}, O_{i+1}, k^I_{i+1}, k^O_{i+1})|(I_i, O_i, k^I_i, k^O_i)\},
\end{aligned}
$$

where the second equality follows from the above reasoning and the last equality follows from the Markov property of $\{(I_i, O_i, k^I_i, k^O_i)\}$. The proof is complete by the fact that the last line, as shown in the proof of Lemma 4, depends only on $(n_{i+1}, o_{i+1}, k^I_{i+1}, k^O_{i+1}), (n_i, o_i, k^I_i, k^O_i))$.
□

We are now in a position to obtain our main result:

LEMMA 5. *The random sequence* (n_i, o_i) *is a Markov chain, with transition probability given by*

$$
\begin{aligned}
p_{n,o;m} :=& \Pr\{n_i - n_{i+1} = o_i - o_{i+1} = m | n_i = n, o_i = o\}\\
=& \left(\frac{m}{(on)^{m+1}}\right)\left(\frac{n!}{(n-m)!}\right)\left(\frac{o!}{(o-m)!}\right)(o + n - m).
\end{aligned}
$$

PROOF. We first compute the probability of transition from (n_i, o_i, k^I_i, k^O_i) such that $k^I_i + k^O_i = \kappa$ to $(n_{i+1}, o_{i+1}, k^I_{i+1}, k^O_{i+1})$ such that $k^I_{i+1} = \lambda^I$ and $k^O_{i+1} = \lambda^O$:

$$
\begin{aligned}
&\mathbf{P}(n, o, \kappa; m, \lambda^I, \lambda^O)\\
:=& \Pr\left\{n_i - n_{i+1} = o_i - o_{i+1} = m, k^I_{i+1} = \lambda^I, k^O_{i+1} = \lambda^O \mid n_i = n, o_i = o, k^I_i + k^O_i = \kappa\right\}.
\end{aligned}
$$

This will be computed as a fraction $\frac{\Theta}{\Upsilon}$. The denominator Υ counts the number of rooted forests in the bipartite digraph with k^I_i roots in I_i and k^O_i roots in O_i where $k^I_i + k^O_i = \kappa$, multiplied by the ways in which k_i^I roots of I_i could point to O_i and k_i^O roots of O_i could point to I_i.[6] Hence, letting $f(n, o, k^I, k^O)$ denote the number of rooted forests in a bipartite digraph (with n and o vertices on both sides) containing k^I and k^O roots on both sides.

$$\begin{aligned}\Upsilon &= \sum_{(k^I,k^O):k^I+k^O=\kappa} o^{k^I} n^{k^O} f(n, o, k^I, k^O) \\ &= \sum_{k^I+k^O=\kappa} o^{k^I} n^{k^O} \binom{n}{k^I} \binom{o}{k^O} o^{n-k^I-1} n^{o-k^O-1} (nk^O + ok^I - k^I k^O) \\ &= \sum_{k^I+k^O=\kappa} \binom{n}{k^I} \binom{o}{k^O} o^{n-1} n^{o-1} (nk^O + ok^I - k^I k^O) \\ &= o^n n^o \left(2\binom{n+o-1}{\kappa-1} - \binom{n+o-2}{\kappa-2} \right).\end{aligned}$$

The first equality follows from the fact that there are $o^{k^I} n^{k^O}$ ways in which k^I roots in I_i point to O_i and k^O roots in O_i could point to I_i. The second equality follows from Lemma 1. The last uses Vandermonde's identity.

The numerator Θ counts the number of ways in which m agents are chosen from I_i and m objects are chosen from O_i to form a bipartite bijection each cycle of which contains at least one of κ old roots, and for each such choice, the number of ways in which the remaining vertices form a spanning rooted forest and the λ^I roots in I_{i+1} point to objects in $O_i \setminus O_{i+1}$ and λ^O roots in O_{i+1} point to agents in $O_i \setminus O_{i+1}$. To compute this, we first compute

$$\alpha(n, o, \kappa; m, \lambda^I, \lambda^O) = \sum_{(k^I,k^O):k^I+k^O=\kappa} \beta(n, o, k^I, k^O; m, \lambda^I, \lambda^O),$$

where β is defined in the proof of Lemma 3. In words, α counts, for any F with $n-m$ agents and $o-m$ objects and roots λ^I and λ^O on both sides, the total number of pairs (F', ϕ) that could have given rise to F, where F' has n agents and o objects with κ roots and ϕ maps the roots to the remaining vertices. Following the construction in the beginning of

[6] Given that we have $n_i = n$ individuals, $o_i = o$ objects and $k_i^I + k_i^O = \kappa$ roots at the beginning of step i under TTC, one may think of this as the total number of possible bipartite digraph one may obtain via TTC at the end of step i when we let k_i^I roots in I_i point to their remaining most favorite object and k_i^O roots in O_i point to their remaining most favorite individual.

the proof of Lemma 3, the number of such pairs is computed as

$$
\begin{aligned}
&\alpha(n,o,\kappa;m,\lambda^I,\lambda^O)\\
:=&\sum_{a+b+c+d=\kappa}\binom{n-m}{c}\binom{o-m}{d}\binom{m}{a}\binom{m}{b}\left(\frac{a+b}{m}-\frac{ab}{m^2}\right)(m!)^2m^{\lambda^I+\lambda^O}\\
=&(m!)^2m^{\lambda^I+\lambda^O}\times\left(\sum_{a+b+c+d=\kappa}\binom{n-m}{c}\binom{o-m}{d}\binom{m-1}{a-1}\binom{m}{b}\right.\\
&\left.+\sum_{a+b+c+d=\kappa}\binom{n-m}{c}\binom{o-m}{d}\binom{m}{a}\binom{m-1}{b-1}-\sum_{a+b+c+d=\kappa}\binom{n-m}{c}\binom{o-m}{d}\binom{m-1}{a-1}\binom{m-1}{b-1}\right)\\
=&(m!)^2m^{\lambda^I+\lambda^O}\left(2\binom{n+o-1}{\kappa-1}-\binom{n+o-2}{\kappa-2}\right).
\end{aligned}
$$

The first equality follows from Lemma 2, along with the fact that there are $(m!)^2$ possible bipartite bijections between $n-m$ agents and $o-m$ objects, and the fact that there are $m^{\lambda^I}m^{\lambda^O}$ ways in which new roots λ^I agents and λ^O objects) could have pointed to $2m$ cyclic vertices (m on the individuals' side and m on the objects' side), and the last equality follows from Vandermonde's identity.

The numerator Θ is now computed as:

$$
\begin{aligned}
\Theta &= \binom{n}{m}\binom{o}{m}f(n-m,o-m,\lambda^I,\lambda^O)\alpha(n,o,\kappa;m,\lambda^I,\lambda^O)\\
&= \binom{n}{m}\binom{o}{m}f(n-m,o-m,\lambda^I,\lambda^O)(m!)^2m^{\lambda^I+\lambda^O}\left(2\binom{n+o-1}{\kappa-1}-\binom{n+o-2}{\kappa-2}\right).\\
&= \left(\frac{n!}{(n-m)!}\right)\left(\frac{o!}{(o-m)!}\right)m^{\lambda^I+\lambda^O}f(n-m,o-m,\lambda^I,\lambda^O)\left(2\binom{n+o-1}{\kappa-1}-\binom{n+o-2}{\kappa-2}\right).
\end{aligned}
$$

Collecting terms, let us compute

$$
\mathbf{P}(n,o,\kappa;m,\lambda^I,\lambda^O)=\frac{1}{o^n n^o}\left(\frac{n!}{(n-m)!}\right)\left(\frac{o!}{(o-m)!}\right)m^{\lambda^I+\lambda^O}f(n-m,o-m,\lambda^I,\lambda^O).
$$

A key observation is that this expression does not depend on κ, which implies that (n_i,o_i) forms a Markov chain.

Its transition probability can be derived by summing the expression over all possible (λ^I,λ^O)'s:

$$
p_{n,o;m}:=\sum_{0\le\lambda^I\le n-m,0\le\lambda^O\le o-m}P(n,o,\kappa;m,\lambda^I,\lambda^O).
$$

To this end, we obtain:

$$\begin{aligned}
&\sum_{0\le\lambda^I\le n-m}\sum_{0\le\lambda^O\le o-m} m^{\lambda^I} m^{\lambda^O} f(n-m, o-m, \lambda^I, \lambda^O)\\
=&\sum_{0\le\lambda^I\le n-m}\sum_{0\le\lambda^O\le o-m} m^{\lambda^I} m^{\lambda^O} \binom{n-m}{\lambda^I}\binom{o-m}{\lambda^O}\times\\
&\quad (o-m)^{n-m-\lambda^I-1}(n-m)^{o-m-\lambda^O-1}((n-m)\lambda^O+(o-m)\lambda^I-\lambda^I\lambda^O)\\
=&m\left(\sum_{0\le\lambda^I\le n-m}\binom{n-m}{\lambda^I}m^{\lambda^I}(o-m)^{n-m-\lambda^I}\right)\left(\sum_{1\le\lambda^O\le o-m}\binom{o-m-1}{\lambda^O-1}m^{\lambda^O-1}(n-m)^{o-m-\lambda^O}\right)\\
&+m\left(\sum_{1\le\lambda^I\le n-m}\binom{n-m-1}{\lambda^I-1}m^{\lambda^I-1}(o-m)^{n-m-\lambda^I}\right)\left(\sum_{0\le\lambda^O\le o-m}\binom{o-m}{\lambda^O}m^{\lambda^O}(n-m)^{o-m-\lambda^O}\right)\\
&-m^2\left(\sum_{1\le\lambda^I\le n-m}\binom{n-m-1}{\lambda^I-1}m^{\lambda^I-1}(o-m)^{n-m-\lambda^I}\right)\left(\sum_{1\le\lambda^O\le o-m}\binom{o-m-1}{\lambda^O-1}m^{\lambda^O-1}(n-m)^{o-m-\lambda^O}\right)\\
=&mo^{n-m}n^{o-m-1}+mo^{n-m-1}n^{o-m}-m^2o^{n-m-1}n^{o-m-1}\\
=&mo^{n-m-1}n^{o-m-1}(n+o-m),
\end{aligned}$$

where the first equality follows from Lemma 1, and the third follows from the Binomial Theorem.

Multiplying the term $\frac{1}{o^n n^o}\left(\frac{n!}{(n-m)!}\right)\left(\frac{o!}{(o-m)!}\right)$, we get the formula stated in the Lemma. □

This last lemma concludes the proof of Theorem 1.

3 Discussion

We show how our result can be exploited in order to compute the expected number of agents matched at a given stage of TTC given the the remaining number of individuals and objects at the beginning of that round.

Consider an arbitrary mapping, $g : I \to O$ and $h : O \to I$, defined over our finite sets I and O. Note that such a mapping naturally induces a bipartite digraph with vertices $I \cup O$ and directed edges with the number of outgoing edges equal to the number of vertices, one for each vertex. In this digraph, $i \in I$ points to $g(i) \in O$ while $o \in O$ points to $h(o) \in I$. Such a mapping will be called a bipartite mapping. A **cycle** of a bipartite mapping is a cycle in the induced bipartite digraph, namely, distinct vertices $(i_1, o_1, ..., i_{k-1}, o_{k-1}, i_k)$ such that $g(i_j) = o_j, h(o_j) = i_{j+1}, j = 1, ..., k-1, i_k = i_1$. A **random bipartite mapping** selects a composite map $h \circ g$ uniformly from a set $\mathcal{H} \times \mathcal{G} = I^O \times O^I$ of all bipartite mappings. Note that a random bipartite mapping induces a random bipartite digraph consisting of vertices $I \cup O$ and directed edges emanating from vertices, one for each vertex. We say that a vertex in a digraph is **cyclic** if it is in a cycle of the digraph.

The following lemma states the number of cyclic vertices in a random bipartite digraph induced by a random bipartite mapping.

LEMMA 6 (Jaworski (1985), Corollary 3). *The number q of the cyclic vertices in a random bipartite digraph induced by a random bipartite mapping $g : I \to O$ and $h : O \to I$ has an expected value of*

$$\mathbb{E}[q] := 2 \sum_{i=1}^{\min\{o,n\}} \frac{(o)_i (n)_i}{o^i n^i},$$

and a variance of

$$8 \sum_{i=1}^{\min\{o,n\}} \frac{(o)_i (n)_i}{o^i n^i} i - \mathbb{E}[q] - \mathbb{E}^2[q],$$

where $(x)_j := x(x-1) \cdots (x-j-1)$.

It is clear that at the beginning of the first round of TTC, if there are n agents and o objects in the economy, the distribution of the number of individuals and objects assigned is the same as that of q. Appealing to Theorem 1 we can further obtain that *for any* round of TTC which begins with n agents and o objects remaining in the market, the number of individuals and objects assigned has the same distribution as q. Hence, the first and second moments of the number of individuals/objects matched at that round corresponds exactly to those in the above lemma. Jaworski (1985) also shows that asymptotically (as o and n grow) the expectation of q is $\sqrt{2\pi \frac{no}{n+o}}(1 + o(1))$ while its variance is $(4 - \pi)\frac{2no}{n+o}(1 + o(1))$. Given the number n of individuals and o of objects available at the beginning of Stage t of TTC, if we denote X_t the number of agents and objects matched at that stage, we have that $\mathbb{E}[\frac{X_t}{\sqrt{2\pi \frac{no}{n+o}}}]$ converges to 1 as n grows while the variance of $\frac{X_t}{\sqrt{2\pi \frac{no}{n+o}}}$ converges to the constant $\frac{4-\pi}{\pi}$.

Finally, Frieze and Pittel (1995) get a similar markov chain result for TTC but in a Shapley-Scarf economy. Our result allows to compare the two Markov chains. Interestingly, we can order the two chains in terms of likelihood ratio order To see this, let us recall the transition probabilities of the Markov chain obtained by Frieze and Pittel (1995):

$$\hat{p}_{n;m} = \frac{n!}{n^m(n-m)!}\frac{m}{n}$$

By Theorem 1,we obtain (assuming $n = o$):

$$\begin{aligned} p_{n;m} \quad : &= p_{n,n;m} = \left(\frac{m}{(n)^{2(m+1)}}\right)\left(\frac{n!}{(n-m)!}\right)^2 (2n-m) \\ &= \left(\frac{n!}{n^m(n-m)!}\right)^2 \left(\frac{m(2n-m)}{n^2}\right). \end{aligned}$$

Let us compare the two distributions in terms of likelihood ratio order. Fix $n \geq 1$ and any $m' \geq m$. It is easy to check that

$$\frac{\hat{p}_{n,m'}}{\hat{p}_{n,m}} = \frac{n^m(n-m)!}{n^{m'}(n-m')!}\frac{m'}{m}$$

while

$$\frac{p_{n,m'}}{p_{n,m}} = \left(\frac{n^m(n-m)!}{n^{m'}(n-m')!}\right)^2 \frac{m'}{m}\frac{2n-m'}{2n-m}.$$

Now, observe that

$$\begin{aligned} \left(\frac{\hat{p}_{n,m'}}{\hat{p}_{n,m}}\right)^{-1} \frac{p_{n,m'}}{p_{n,m}} &= \left(\frac{1}{n^{m'-m}}\right)\left(\frac{(n-m)!}{(n-m')!}\right)\frac{(2n-m')}{(2n-m)} \\ &= \frac{(n-m)(n-m-1)...(n-m'+1)}{n^{m'-m}}\frac{2n-m'}{2n-m} \leq 1. \end{aligned}$$

This proves that for any n, the distribution $\hat{p}_{n,\cdot}$ dominates $p_{n,\cdot}$ in terms of likelihood ratio order. One can prove prove an interesting implication of this result: for each $t \geq 1$, the probability that TTC stops before Round t is smaller than the probability that Shapley-Scarf TTC stops before Round t. Put in another way, the random round at which TTC stops is (first order) stochastically dominated by that at which the Shapley-Scarf TTC stops.

References

Abdulkadiroglu, A., and T. Sonmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Che, Y.-K., and O. Tercieux (2015): "Efficiency and Stability in Large Matching Markets," Columbia University and PSE, Unpublished mimeo.

Frieze, A., and B. Pittel (1995): "Probabilistic Analysis of an Algorithm in the Theory of Markets in Indivisible Goods," *The Annals of Applied Probability*, 5, 768–808.

Jaworski, J. (1985): "A Random Bipartite Mapping," *The Annals of Discrete Mathematics*, 28, 137–158.

Jin, Y., and C. Liu (2004): "Enumeration for spanning forests of complete bipartite graphs," *Ars Combinatoria - ARSCOM*, 70, 135–138.

Lovasz, L. (1979): *Combinatorial Problems and Exercises*. North Holland, Amsterdam.

Sonmez, T., and U. Unver (2011): "Market Design for Kidney Exchange," in *Oxford Handbook of Market Design*, ed. by M. N. Z. Neeman, and N. Vulkan. forthcoming.