
Nash Equilibria in the Two-Player Kidney Exchange Game

Margarida Carvalho · Andrea Lodi ·
João Pedro Pedroso · Ana Viana

Abstract Kidney exchange programs have been set in several countries within national, regional or hospital frameworks, to increase the possibility of kidney patients being transplanted. For the case of hospital programs, it has been claimed that hospitals would benefit if they collaborated with each other, sharing their internal pools and allowing transplants involving patients of different hospitals. This claim led to the study of multi-hospital exchange markets. We propose a novel direction in this setting by modeling the exchange market as an integer programming game. The analysis of the strategic behavior of the entities participating in the kidney exchange game allowed us to prove that the most rational game outcome maximizes the social welfare and that it can be computed in polynomial time.

Keywords Kidney Exchange · Nash Equilibrium · Social Welfare · Matching

1 Introduction

The Kidney Exchange Problem can be described as follows. A patient suffering from renal failure can see her life quality improved through the transplantation of a healthy kidney. Typically, a patient receives a kidney transplant from a deceased donor, or from a living donor that is a patient's relative or friend. Unfortunately, these two possibilities of transplantation can only satisfy a tiny

M. Carvalho · J. P. Pedroso
INESC TEC and Faculdade de Ciências da Universidade do Porto, Portugal
E-mail: margarida.carvalho@dcc.fc.up.pt, jpp@fc.up.pt

A. Lodi
University of Bologna, Italy and École Polytechnique de Montréal, Canada
E-mail: andrea.lodi@polymtl.ca

A. Viana
INESC TEC and Instituto Superior de Engenharia do Porto, Portugal
E-mail: aviana@inescporto.pt

fraction of the demand, since deceased donors are scarce and patient-donor incompatibilities may occur.

To potentially increase the number of kidney transplants, some countries' recent legislation (*e.g.*, United Kingdom [16], Netherlands [10]) allows a pairwise exchange: *e.g.*, for two patient-donor pairs P_1 and P_2 the patient of pair P_1 receives a kidney from the donor of pair P_2 and vice versa, forming a cycle of size 2. The idea can be extended to allow more than two pairs to be involved in an exchange (for L -pairs, P_2 receives a kidney from the donor P_1 , P_3 from the donor of P_2 , etc, and, finally, P_1 from the donor of P_L , closing a cycle), and to include undirected (altruistic) donors, as well as pairs with other characteristics [9]. The general aim is to define a match that maximizes the number of transplants in a pool. Because in most cases the operations must take place at the same time, for logistic reasons the number of pairs that can be involved in an exchange is limited to a maximum value, say L . Furthermore, because additional compatibility tests that must be performed prior to transplant may lead to new incompatibilities, resulting in the cancellation of all transplants involved in the cycle it is preferable for the cycles to be shorter.

Abraham et al. [1] formulated the kidney exchange problem (KEP) as an integer program with an exponential number of variables, which maximizes the number of vertices covered in a digraph by vertex-disjoint cycles of size at most L . In this model the vertices of the digraph represent patient-donor pairs and the arcs represent the compatibilities between pairs. A compact model, where the number of variables and constraints increases polynomially with the problem size, is proposed by Constantino et al. [9].

In the previous models, there is a centralized decision maker deciding the exchange program. However, there are other potential decision makers to be considered that can influence the exchange program. In Ceclárová et al. [8], patient-donor pairs are the players in a cooperative kidney exchange game that is structurally different from what is presented in this paper because the players, the set of actions and utilities interact differently, as will be clear after our game model description.

Multi-Agent Kidney Exchange. Although some countries have a national kidney exchange pool with the matches being done by a central authority, other countries have regional (or hospital) pools, where the matches are performed internally with no collaboration between the different entities. Since it is expected that as the size of a patient-donor pool increases more exchanges can take place, it became relevant to study kidney exchange programs involving several hospitals or even several countries. In such cases each entity is a self-interested agent that aims at maximizing the number of its patients receiving a kidney [3, 4].

To the extent of our knowledge, work in this area concentrates on the search of a strategyproof mechanism that decides all exchanges to be performed in a multi-hospital setting. A mechanism is *strategyproof* if the participating hospitals do not have incentive to hide information from a central authority that decides through that mechanism the exchanges that are to be executed. For the

2-hospital kidney exchange program with pairwise exchanges, the deterministic strategyproof mechanism in [2] provides a 2-approximation ratio on the maximum number of exchanges, while the randomized strategyproof mechanism in [7] guarantees a $\frac{3}{2}$ -approximation ratio. Additionally, Ashlagi et al. [2] built a randomized strategyproof mechanism for the multi-hospital case with approximation ratio 2, again only for pairwise exchanges. In these mechanisms, in order to encourage the hospitals to report all their incompatible pairs, the social welfare is sacrificed. In fact, the best lower bound for a strategyproof (randomized) mechanism is $2\left(\frac{8}{7}\right)$, which means that no mechanism returning the maximum number of exchanges is strategyproof [2]. In this context, the question is whether, analysing the hospitals interaction from a standpoint of a non-cooperative game, Nash equilibria would improve the program's social welfare.

We can formalize and generalize KEP to a competitive N -player kidney exchange game (N -KEG) with two sequential moves: first, simultaneously, each player n , for $n = 1, \dots, N$, decides the internal exchanges to be performed; second, an independent agent (IA) takes the first-stage unused pairs and decides the external exchanges to be done such that the number of pairs participating on it is maximized. Let us define V^n as the vertex set of player n , $V = \bigcup_{n=1}^N V^n$ and C as the set of cycles with size at most L . Let $C^n = \{c \in C : c \cap V^n = c\}$ be the subset of cycles involving only player n 's patient-donor pairs, and $I = C \setminus \bigcup_{n=1}^N C^n$ be the subset of cycles, involving at least two patient-donor pairs of distinct players. Each player solves the following bilevel programming:

$$\underset{x^n \in \{0,1\}^{|C^n|}}{\text{maximize}} \quad \sum_{c \in C^n} w_c^n x_c^n + \sum_{c \in I} w_c^n y_c \quad (1.1a)$$

$$\text{subject to} \quad \sum_{c \in C^n : i \in c} x_c^n \leq 1 \quad \forall i \in V^n \quad (1.1b)$$

where y solves the problem

$$\underset{y \in \{0,1\}^{|I|}}{\text{maximize}} \quad \sum_{c \in I} \sum_{n=1}^N w_c^n y_c \quad (1.1c)$$

$$\text{s.t.} \quad \sum_{c \in I : i \in c} y_c \leq 1 - \sum_{n=1}^N \sum_{c \in C^n : i \in c} x_c^n \quad \forall i \in V \quad (1.1d)$$

Player n controls a binary decision vector x^n with size equal to the cardinality of C^n . An element x_c^n of x^n is 1 if cycle $c \in C^n$ is selected, 0 otherwise. Similarly, the IA controls the binary decision vector y with size equal to the cardinality of I . The objective function (1.1a) translates on the maximization of player n 's patients receiving a kidney: w_c^n the number of player's n patient-donor pairs in cycle c (which is the size of c if it is an internal). Constraints (1.1b) ensure that every pair is in at most one exchange. The IA objective function (1.1c) represents the maximization of patient-donor pairs receiving a kidney in the second-stage. Constraints (1.1d) are analogous to (1.1b), but also

ensure that pairs participating in the first-stage exchanges are not selected by the IA.

In the way that we defined N -KEG, it is implicit that it is a complete information game. Initially, every player decides the pairs to reveal, and only revealed pairs will be considered in each player utility as well as in the second stage IA decision process. Note that there is no incentive for hiding information, as each player has complete control over her internal exchanges, and, therefore, can guarantee to be at least as good as if it was by herself. Moreover, if there are hidden pairs, they will not be considered in the IA decision, and thus, the players will not benefit from external exchanges including them. Consequently, this is intrinsically a complete information game.

The formulation above brings up the following research question: is the generalization of KEP to N -KEG relevant? In particular, it is worth noting that the special case of KEP with $L = 2$ can be formulated as a maximum matching problem and consequently, solved in polynomial time. Moreover, the multi-agent kidney exchange literature focuses mainly in exchanges with size 2. Thus, the most natural and relevant extension to look at is 2-KEG with pairwise exchanges.

Our Contributions. In this paper we concentrate on the non-cooperative 2-player kidney exchange game (2-KEG) with pairwise exchanges. A player can be a hospital, a region or even a country. Under this setting it is inefficient to follow the classical normal-form game approach [13] by specifying all the players' strategies. Note also that in our formulation of N -KEG, players' strategies are lattice points inside polytopes described by systems of linear inequalities. Thus, according to [15], N -KEG and, in particular, 2-KEG belongs to the class of integer programming games.

We show that 2-KEG has always a pure Nash equilibrium (NE) and that it can be computed in polynomial time. Furthermore, we prove the existence of a NE that is also a *social optimum*, *i.e.*, the existence of an equilibrium where the maximum number of exchanges is performed. Finally, we show how to determine a NE that is a social optimum, always the preferred outcome of both players, and can be computed in polynomial time.

Our work indicates that studying the players interaction through 2-KEG turns the exchange program efficient both from the social welfare and players' point of view. In contrast, as mentioned before, there is no centralized mechanism that is strategyproof and at the same time guarantees a social optimum. Although we provide strong evidence that under 2-KEG the players' most rational strategy is a social optimum, we note the possibility of multiple equilibria. We show that the worst case Nash equilibrium in terms of social welfare is at least $\frac{1}{2}$ of the social optimum. Thus, the worst case outcome for our game is comparable with the one for the best deterministic strategyproof mechanism (recall that it guarantees a 2-approximation of the social optimum). Therefore, the 2-KEG opens a new research direction in this field that is worth being explored.

Organization of the Paper. Section 2 formulates 2-KEG in mathematical terms. Section 3 proves the existence of a Nash equilibrium that maximizes the social welfare and measures the Nash equilibria quality enabling the comparison of our game with strategyproof mechanisms. Section 4 proves that the players have incentive to choose Nash equilibrium that are socially optimal. Section 5 refines the concept of social welfare equilibria motivating for a unique rational outcome for the game. Section 6 discusses extensions to our model and Section 7 draws some conclusions.

2 Definitions and preliminaries

Let the players of 2-KEG be labeled player A and player B . The operator $(\cdot)^{-n}$ for some $n \in \{A, B\}$ denotes the opponent of player n . For representing a 2-KEG as a graph, let V be a set of *vertices* representing the incompatible patient-donor pairs of players A and B , and E be the set of possible pairwise exchanges, *i.e.*, the set of *edges* (i, j) such that the patient of $i \in V$ is compatible with the donor of $j \in V$ and vice versa. For each player n , $V^n \subseteq V$ and $E^n \subseteq E$ are her patient-donor pairs and internal compatibilities, respectively. A subset M^n of E^n is called a *matching* of graph $G^n = (V^n, E^n)$ if no two edges of it share the same vertex. A player n 's *strategy set* is the set of matchings in graph $G^n = (V^n, E^n)$. A *profile of strategies* is the specification of a matching for all players. The independent agent controls the external exchanges $E^I \subseteq E$, *i.e.*, $(a, b) \in E^I$ if $a \in V^A$ and $b \in V^B$. Let $E^I(M^A, M^B)$ be a subset of E^I such that no edge is incident upon a vertex covered by M^A or M^B . For a player B 's matching M^B define the *player A 's reaction graph* $G^A(M^B) = (V, E^A \cup E^I(\emptyset, M^B))$ and for a player A 's matching M^A define the *player B 's reaction graph* $G^B(M^A) = (V, E^B \cup E^I(\emptyset, M^A))$. We will represent vertices that belong to V^A as gray circles and vertices that belong to V^B as white diamonds.

On the first stage of 2-KEG, each player n decides simultaneously a matching M^n of graph G^n to be executed. On the second stage of the game, given player A 's first-stage decision M^A and player B 's first-stage decision M^B , the IA decides the external exchanges to be performed such that the number of pairs covered by its decision is maximized. In other words, the IA finds a *maximum matching* $M^I(M^A, M^B)$ of $E^I(M^A, M^B)$, *i.e.*, a matching of maximum cardinality. In the end of the game, player A 's utility is $2|M^A| + |M^I(M^A, M^B)|$ and player B 's utility is $2|M^B| + |M^I(M^A, M^B)|$.

An important factor for a game is that its rules are executed efficiently. For 2-KEG this means that the IA optimization problem must be easy to solve. Edmonds [12] proved that the problem of computing a maximum matching can be solved in polynomial time for any graph. Therefore, given the players' decisions, the IA optimization problem is solved in polynomial time.

A legitimate question that must be answered is if the game is well defined in the sense that the rules are unambiguous. Note that the utility of each player depends on the IA decision rule. In the general N -KEG case, there might

be situations where there are multiple optimal IA's decisions that benefit the players differently. However, for 2-KEG that is not possible, because only pairwise exchanges are considered. That is, any IA matching leads to equal benefits for both players.

Proposition 1 *2-KEG is well defined.*

One apparent difficulty in the treatment of the game has to do with the bilevel optimization problem (1.1) of each player. However, computing a player's optimal strategy to a fixed matching of the other player can be simplified. From the standpoint of player A , the *best reaction* M^A to a player B 's fixed strategy M^B can be computed by dropping the IA objective function (1.1c) (game rule) and solving the single level matching problem in the reaction graph $G^A(M^B)$. Basically, we are claiming that player A best reaction predicts the appropriate IA decision given M^A and M^B . This holds because IA's edges have a positive impact on the utility of player A .

Lemma 1 *Let M^B be a matching of player B in 2-KEG. Player A 's best reaction to M^B can be achieved by solving a maximum weight matching problem on the graph $G^A(M^B)$, where the edges of G^A in E^A have weight 2 and those in $E^I(\emptyset, M^B)$ weight 1. The equivalent for player B also holds.*

3 Nash equilibria and social welfare

Normal form games are a class of finite games for which the players' strategies are explicitly specified. Unlike these games, the literature on integer programming games is almost nonexistent and the intuition is that they are more difficult to treat, since players' set of feasible strategies can have exponential size.

A Nash equilibrium is a widely accepted solution for a game. Nash [18] proved, in a non-constructive way, that any finite game has a NE. General algorithms to compute NE for normal form games were devised, but they fail to be polynomial [19]. In particular, these algorithms are inappropriate for integer programming games, as they require the explicit enumeration of all feasible strategies and here, the set of feasible matchings grows exponentially with the number of patient-donor pairs (vertices of the graph).

In this paper, we will concentrate on pure equilibria. A player A 's matching M^A of G^A and a player B 's matching M^B of G^B is a *pure Nash equilibrium* for 2-KEG if

$$2|M^A| + |M^I(M^A, M^B)| \geq 2|R^A| + |M^I(R^A, M^B)| \quad \forall \text{ matching } R^A \text{ of } G^A$$

$$2|M^B| + |M^I(M^A, M^B)| \geq 2|R^B| + |M^I(M^A, R^B)| \quad \forall \text{ matching } R^B \text{ of } G^B.$$

Along the paper we use NE to refer to pure Nash equilibria. Under 2-KEG, each player seeks to choose an internal matching that leads to the maximization of the number of its patients receiving a transplant in the end of the

game. Hence, a rational profile of strategies is one that simultaneously maximizes each players' utility. The NE satisfies this goal.

A mixed-strategy Nash equilibrium attributes a probability distribution over the players' feasible decisions; therefore, its description may involve an exponential number of players' strategies, which is computationally unsuitable.

In Section 3.1, we prove the existence of NE for 2-KEG and that it can be computed in polynomial time. Through these results, in Section 3.2 we prove the existence of a NE that maximizes the social welfare (sum of the players' utilities or, equivalently, number of vertices matched). In Section 3.3, we measure the quality of the NE in terms of social welfare. This analysis allow us to conclude that the worst case Nash equilibrium to 2-KEG and the best deterministic strategyproof mechanism guarantee that at least $\frac{1}{2}$ of the number of vertices matched in a social optimum is achieved.

3.1 Existence of a pure Nash equilibrium

In order to prove the existence of a pure NE we will use the concept of potential function to games, as defined in [17]. For 2-KEG, a potential function Φ is a real-valued function over the set of player A 's matchings in G^A and player B 's matchings in G^B such that the value of Φ increases strictly when a player switches to a new matching that improves its utility. In particular, a potential function is exact when this increase is equal to the player's utility increase when she unilaterally deviates to a new matching.

Observe that a player A 's decision does not interfere in the set of player B 's matchings in G^B . In particular, player A cannot influence the part of player B 's utility related with a matching in G^B . The symmetric observation holds for player B 's decision. With this in mind, it is not difficult to find an exact potential function to 2-KEG.

Proposition 2 *Function $\Phi(M^A, M^B) = 2|M^A| + 2|M^B| + |M^I(M^A, M^B)|$ is an exact potential function of 2-KEG.*

A profile of strategies for which the potential function maximum is attained is a NE (Lemma 2.1 of [17]). Otherwise, at least one of the players would have advantage in switching to a new strategy, which would imply that the potential function would strictly increase its value in this new profile. However, that contradicts the fact that the previous profile was a potential function optimum.

Theorem 1 *There exists at least one pure Nash equilibrium to 2-KEG and it can be computed in polynomial time.*

Proof A matching corresponding to the maximum of the function Φ of Proposition 2 is a NE of 2-KEG. Computing a maximum to Φ is equivalent to solving a maximum weight matching problem, where the edges in E^A and E^B weight 2 and the edges in E^I weight 1. This can be done in polynomial time (see,

e.g., [20]). Observe that the set of external matchings associated with the solution of this maximum weight matching problem is a maximum matching in E^I except edges incident with internal matchings of that solution. \square

Consider the 2-KEG instance represented in Figure 3.1. In this case, the NE achieved by computing the potential function maximum is $M^A = \{(4, 5)\}$, $M^B = \{(2, 3)\}$ (and thus, $M^I(M^A, M^B) = \emptyset$). There is another NE that does not correspond to a potential function maximum: $R^A = \emptyset$, $R^B = \emptyset$ and consequently $M^I(R^A, R^B) = \{(1, 2), (4, 3), (5, 6)\}$. The latter helps all the patient-donor pairs, and thus is more appealing to the players. This observation motivates the need of studying efficient Nash equilibria that are possibly not achieved through the potential function maximum.

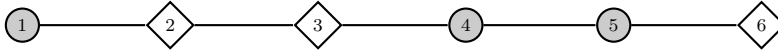


Fig. 3.1: Example of a 2-KEG instance with two distinct Nash equilibria.

3.2 Social welfare equilibrium

In what follows, we introduce a refinement of the NE concept in 2-KEG: the social welfare equilibrium.

A *social optimum* of 2-KEG is a maximum matching of the overall graph game $G = (V, E)$, corresponding to an exchange program that maximizes the number of patients receiving a kidney. A *social welfare equilibrium* (SWE) is a NE that is also a social optimum.

Observe that any NE, and thus any SWE, is a local maximum of Φ if the neighborhood of a strategy profile consists of a player's unilateral deviation. In what follows, we will use this fact to prove the existence and efficient computation of a SWE.

Let us define some concepts of graph theory in matching (see Chapter 5 of [6] for details). For a matching M in graph $G = (V, E)$, an *M -alternating path* is a path whose edges are alternately in $E \setminus M$ and M . An *M -augmenting path* is an M -alternating path whose origin and destination are M -unmatched vertices. The next property will be used often in what we will develop.

Property 1 Let M be a maximum matching of a graph $G = (V, E)$. Consider an arbitrary $R \subset M$ and the subgraph H of G induced by removing the R -matched vertices. The union of any maximum matching of H with R is a maximum matching of G .

Next, we recall Berge's theorem [5].

Theorem 2 (Berge [5]) *A matching M of a graph G is maximum if and only if it has no augmenting path.*

Berge's theorem is constructive, leading to an algorithm to find a maximum matching: start with an arbitrary matching M of G ; while there is an M -augmenting path p , switch the edges along the path p from in to out of M and vice versa: update M to $M \oplus p$, where \oplus represents the symmetric difference of two sets. The updated M is a matching with one more edge, where the previously matched vertices are maintained matched.

We have now the tools to prove the existence of a SWE.

Theorem 3 *There is always a social welfare equilibrium to 2-KEG.*

Proof Let M be a maximum matching (and thus, a social optimum) of the graph G representing a 2-KEG, where $E^A \cap M$ and $E^B \cap M$ are players' A and B strategies, respectively. If M is not a NE, let us assume, without loss of generality, that player A has incentive to deviate from $E^A \cap M$, given player B 's strategy $E^B \cap M$. Let M^A be player A 's best reaction to $E^B \cap M$. Observe that we can assume that $M^A \cup M^I(M^A, E^B \cap M)$ is a maximum matching of A in the reaction graph $G^A(E^B \cap M)$. If it is not, by Berge's theorem, there is a maximum matching such that it does not decrease the number of player A 's matched vertices. Therefore, by Property 1, $|M^A| + |M^I(M^A, E^B \cap M)| + |E^B \cap M| = |M|$.

Given that A has incentive to deviate, it holds by definition of potential function that $\Phi(E^A \cap M, E^B \cap M) < \Phi(M^A, E^B \cap M)$. If M^A together with $E^B \cap M$ is not a NE, then we can repeat the procedure above (alternating the player) until a NE is obtained. Note that the value of the potential function increases strictly, which means that no feasible profile of strategies is visited more than once. In addition, players have a finite number of feasible matchings, which implies that this process will terminate in an equilibrium. \square

Besides the fact that a SWE is an appealing NE to the players, it also has the advantage of being computable in polynomial time through the algorithm of the last proof (translated to pseudo-code in Algorithm 3.2.1). It is a well-known result that weighed matching problems can be solved in polynomial time (see, e.g., [20]). Therefore, it remains to prove that the number of iterations is polynomially bounded in the size of the instance. The next trivial result can be used to this end.

Lemma 2 *An upper bound to the maximum value of the 2-KEG potential function $\Phi(M^A, M^B) = 2|M^A| + 2|M^B| + |M^I(M^A, M^B)|$ is $|V^A| + |V^B|$.*

As noted before, the potential function Φ strictly increases whenever a player has incentive to unilaterally change her strategy. Therefore, our algorithm will in the worst case stop once the maximum value to Φ is reached, which is bounded by $|V^A| + |V^B|$. Taking into account that the value of Φ is always an integer number, the number of evaluations of Φ through the process is also bounded by $|V^A| + |V^B|$.

Theorem 4 *The computation of a social welfare equilibrium to 2-KEG can be done in polynomial time.*

Algorithm 3.2.1 Computation of a social welfare Nash equilibrium

Input: a 2-KEG instance G
Output: equilibrium matchings

```

1:  $M \leftarrow$  maximum matching of  $G$ 
2:  $M^A \leftarrow M \cap E^A, M^B \leftarrow M \cap E^B, M^I \leftarrow M \cap E^I$  /* initial matchings */
3: while  $\exists$  player  $n \in \{A, B\}$  with incentive to deviate from  $M^n$  do
4:    $R^n \leftarrow$  player  $n$ 's best reaction to  $M^{-n}$  /* maximum weight matching on  $G^n(M^{-n})$  */
5:   Apply (unweighted) augmenting paths to  $R^n$  until  $R^n$  is a maximum matching of  $G^n(M^{-n})$  /*  $R^n \cap E^A$  continues to be a best reaction to  $M^{-n}$  */
6:    $M^n \leftarrow R^n \cap E^A, M^I \leftarrow M^I(M^n, M^{-n})$  /* update solution */
7: end while
8: return  $M^A, M^B$ 

```

3.3 Price of stability and price of anarchy

In order to measure the quality of the Nash equilibria of a given game, we use the standard measures: price of stability and price of anarchy (see Chapter 17 of [19]). The *price of stability* (PoS) is the ratio between the highest total utilities value of one of its equilibria and that of a social optimum; the *price of anarchy* (PoA) is the ratio between the lowest total utilities value within its equilibria and that of a social optimum.

The following two results set PoS and PoA for 2-KEG.

Corollary 1 *The price of stability of the 2-KEG is 1.*

Proof Since we proved existence of a social welfare equilibrium:

$$\text{PoS} = \frac{\text{highest total utilities value among all Nash equilibria}}{\text{social optimum}} = 1.$$

□

Theorem 5 *The price of anarchy is $\frac{1}{2}$ for the 2-KEG.*

Proof By the definition of price of anarchy

$$\text{PoA} = \frac{\text{lowest total utilities value among all Nash equilibria}}{\text{social optimum}}.$$

Let M^A, M^B and $M^I(M^A, M^B)$ be the matchings of player A, B and the IA, respectively, that lead to the Nash equilibrium with lowest total utilities value, that is

$$z^* = 2|M^A| + 2|M^B| + 2|M^I(M^A, M^B)|.$$

Let M be a maximum matching of the game graph G . Therefore, the social optimum is equal to

$$\bar{z} = 2|M \cap E^A| + 2|M \cap E^B| + 2|M \cap E^I|.$$

By the definition of NE, we know that under M^A and M^B , none of the players has incentive to deviate, thus

$$\begin{aligned}
z^* &\geq 2|M \cap E^A| + |M^I(M \cap E^A, M^B)| + 2|M \cap E^B| + |M^I(M^A, M \cap E^B)| \\
\Leftrightarrow z^* &\geq 2|M \cap E^A| + 2|M \cap E^B| + 2|M \cap E^I| - 2|M \cap E^I| + |M^I(M \cap E^A, M^B)| + |M^I(M^A, M \cap E^B)| \\
\Leftrightarrow z^* &\geq \bar{z} - (2|M \cap E^I| - |M^I(M^A, M \cap E^B)| - |M^I(M \cap E^A, M^B)|).
\end{aligned} \tag{3.1a}$$

The set $M \cap E^I$ may include matchings of vertices also matched under M^A or M^B , therefore

$$2|M \cap E^I| \leq 2|M^A| + 2|M^B| + |R^A| + |R^B|$$

where R^n is a subset of E considering all the edges in $M \cap E^I$ but not in M^n and incident with a vertex of V^n , for $n = A, B$. See Figure 3.2. The number

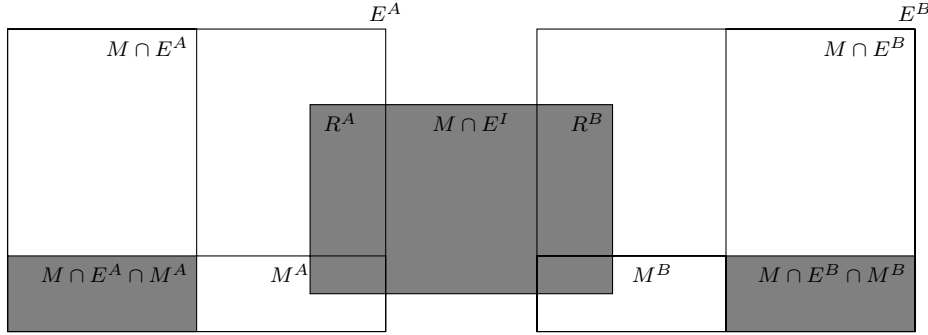


Fig. 3.2: Illustration of the solutions associated with the worst Nash equilibrium and the social optimum.

of player B 's vertices matched in $M^I(M \cap E^A, M^B)$ is equal or greater than R^B , because this external matching has available the vertices incident with the edges of R^B and can match them with any vertex not in $M \cap E^A$, thus

$$|R^B| - |M^I(M^A, M \cap E^B)| \leq 0.$$

In a completely analogous way, it can be shown that

$$|R^A| - |M^I(M \cap E^A, M^B)| \leq 0.$$

The inequalities above imply

$$2|M \cap E^I| - |M^I(M^A, M \cap E^B)| - |M^I(M \cap E^A, M^B)| \leq 2|M^A| + 2|M^B| \leq z^*,$$

which together with inequality (3.1a) results in

$$z^* \geq \bar{z} - z^* \Leftrightarrow \frac{z^*}{\bar{z}} \geq \frac{1}{2}.$$

Now, we will use an instance to prove that the bound $\frac{1}{2}$ is tight.

Consider a 2-KEG represented by the graph of Figure 3.3. It is easy to see that the worst Nash equilibrium in terms of total utilities is $M^A = \{(1, 2)\}$, $M^B = \emptyset$ and $M^I(M^A, M^B) = \emptyset$ with a total of $z^* = 2$. On the other hand, the social optimum is $M = \{(1, 3), (2, 4)\}$ with a value of $\bar{z} = 4$. In this instance the price of anarchy is $\frac{z^*}{\bar{z}} = \frac{2}{4} = \frac{1}{2}$. \square

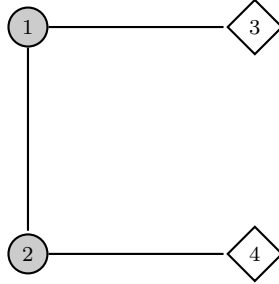


Fig. 3.3: 2-KEG with price of anarchy equal to $\frac{1}{2}$

4 Rational outcome: social welfare equilibrium

A profile of strategies is *dominated* if there is another profile in which all the players are equal or better, with at least one of them strictly better. A profile of strategies is said to be *Pareto efficient* if it is not dominated [21]. In this section, we will prove that the social welfare equilibria are Pareto efficient and any NE that is not social optimal is dominated by a SWE. Consequently, from both the social welfare and the players' point of view, these equilibria are the most desirable game outcomes. Moreover, recall that in Section 3.2, we presented an algorithm that computes a SWE in polynomial time emphasizing its practicality.

Below we show that no SWE is dominated, *i.e.*, all SWE are Pareto efficient.

Lemma 3 *In 2-KEG any social welfare equilibrium is Pareto efficient.*

Proof Let M^A and M^B be players' A and B strategies, respectively, in a SWE. Assume that this SWE is not Pareto efficient, that is, there is a player A 's feasible strategy R^A and a player B 's feasible strategy R^B that dominate this equilibrium. Without loss of generality, these assumptions translate into

$$2|M^A| + |M^I(M^A, M^B)| \leq 2|R^A| + |M^I(R^A, R^B)|$$

$$2|M^B| + |M^I(M^A, M^B)| < 2|R^B| + |M^I(R^A, R^B)|.$$

Summing the two inequalities above and simplifying, we obtain

$$|M^A| + |M^I(M^A, M^B)| + |M^B| < |R^A| + |M^I(R^A, R^B)| + |R^B|,$$

which contradicts the assumption that the equilibrium given by M^A and M^B is a social optimum (maximum matching). \square

Note that this result also holds for more than two players which reinforces the interest of studying SWE.

In the next section, we prove any NE that is not a social optimum is dominated by a SWE. In order to achieve this result we need the following theorem, which fully characterizes an optimal reaction of a player.

Theorem 6 *In 2-KEG, let M^B be a player B's fixed matching. A player A's matching M^A can be improved if and only if there is a $M^A \cup M^I(M^A, M^B)$ -alternating path in $G^A(M^B)$ whose origin is a vertex in V^A , unmatched in this path, and the destination is a*

- i. $M^A \cup M^I(M^A, M^B)$ -unmatched vertex belonging to V^A , or*
- ii. $M^I(M^A, M^B)$ -matched vertex in V^B , or*
- iii. $M^I(M^A, M^B)$ -unmatched vertex in V^B .*

The symmetric result for player B also holds.

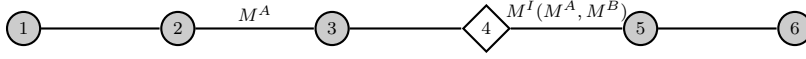
Proof Consider a fixed match M^B of G^B .

(Proof of “if”). Let M^A be a player A's strategy. Recall Lemma 1 in which we state that given M^B , we can assume that player A controls the IA decision. If there is a path p in $G^A(M^A)$ satisfying *i.*, *ii.* or *iii.*, then, $(M^A \cup M^I(M^A, M^B)) \oplus p$ improves player A's profit in comparison with $M^A \cup M^I(M^A, M^B)$; see Figure 4.1 for an illustration.

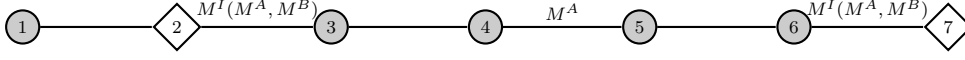
(Proof of “only if”). Let M^A be player A's best reaction to M^B and consider a feasible player A's strategy R^A that is not her best reaction to M^B . We will show that assuming that there is no $R^A \cup M^I(R^A, M^B)$ -alternating path of $G^A(M^B)$ as stated in the theorem leads to a contradiction.

Note that given any two matchings M^1 and M^2 of a graph, in the induced subgraph with edges $M^1 \oplus M^2$, each vertex can be incident to at most two edges; hence, any connected component of $M^1 \oplus M^2$ is either an even cycle with edges alternately in M^1 and M^2 , or a path with edges alternately in M^1 and M^2 . Let us define H^A as the subgraph of G^A that results from considering the edges in $M^A \oplus R^A$, and H as the subgraph of $G^A(M^B)$ that results from considering the edges in $(M^A \cup M^I(M^A, M^B)) \oplus (R^A \cup M^I(R^A, M^B))$. Connected components of H^A and of H are either even cycles or paths.

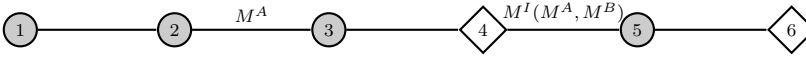
If $|M^A| > |R^A|$, H^A has more edges of M^A than of R^A , and therefore there exists a path p of H^A that starts and ends with edges of M^A . If the origin and destination of p are $M^I(R^A, M^B)$ -unmatched, then p is an $R^A \cup M^I(R^A, M^B)$ -alternating path as stated in *i.*, which contradicts our assumption. Thus, for all paths of H^A starting and ending with edges of M^A , it holds that all their



Case i. - The matching $\{(2, 3), (4, 5)\} \oplus \{(1, 2), (2, 3), (3, 4), (4, 5), (5, 6)\}$ increases player A 's utility by two units.



Case ii. - The matching $\{(2, 3), (4, 5), (6, 7)\} \oplus \{(1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7)\}$ increases player A 's utility by one unit.



Case iii. - The matching $\{(2, 3), (4, 5)\} \oplus \{(1, 2), (2, 3), (3, 4), (4, 5), (5, 6)\}$ increases player A 's utility by one unit.

Fig. 4.1: Possibilities for player A 's to have an incentive to deviate from strategy M^A , given the opponent strategy M^B .

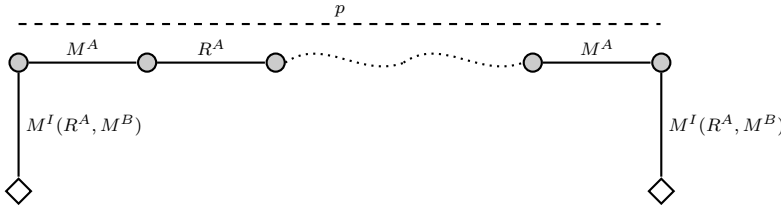


Fig. 4.2: The path p is not an $R^A \cup M^I(R^A, M^B)$ -alternating path of type i .

vertices are both M^A -matched and $R^A \cup M^I(R^A, M^B)$ -matched (see Figure 4.2). Therefore, the advantage of $M^A \cup M^I(M^A, M^B)$ over $R^A \cup M^I(R^A, M^B)$ must be outside H^A . Analogously, if $|M^A| \leq |R^A|$, we also conclude that the advantage of $M^A \cup M^I(M^A, M^B)$ over $R^A \cup M^I(R^A, M^B)$ must be outside H^A .

In this way, there is $a \in V^A$ and $b \in V^B$ such that $(a, b) \in M^I(M^A, M^B)$, but a is $R^A \cup M^I(R^A, M^B)$ -unmatched. Then, since we assumed that there is no $R^A \cup M^I(R^A, M^B)$ -alternating path as stated in the theorem (and the IA does not violate the game rules), the path of H starting in a must end in a vertex $a' \in V^A$ that is $R^A \cup M^I(R^A, M^B)$ -matched and $M^A \cup M^I(M^A, M^B)$ -unmatched. Therefore, the number of V^A vertices covered by $M^A \cup M^I(M^A, M^B)$ and $R^A \cup M^I(R^A, M^B)$ on this component is the same (see Figure 4.3). In conclusion, any path of H starting in a vertex of V^A that is $R^A \cup M^I(R^A, M^B)$ -unmatched and $M^I(M^A, M^B)$ -matched does not give

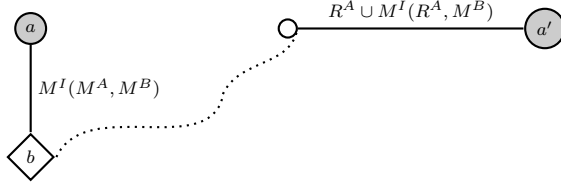


Fig. 4.3: Path component of H . The white circle is a vertex for which it is not important to specify the player to which it belongs.

advantage to $M^A \cup M^I(M^A, M^B)$ over $R^A \cup M^I(R^A, M^B)$. This contradicts the fact that strategy R^A is not a player A 's best reaction to M^B . \square

4.1 Computation of a dominant SWE

We present in Algorithm 4.1.1 a method that, given a 2-KEG graph and a socially suboptimal Nash equilibrium, computes a SWE that we claim dominates the given equilibrium.

Algorithm 4.1.1 Computation of a dominant SWE

Input: a 2-KEG instance G , a NE M of G

Output: M if it is a SWE, else a SWE dominating it

- 1: $S \leftarrow$ a maximum matching of G
 - 2: **if** $|M| = |S|$ **then**
 - 3: **return** M
 - 4: **end if**
 - 5: $t \leftarrow 1$
 - 6: $\mathcal{P}^t \leftarrow$ paths from $M \oplus S$ with both extreme edges in S /* M -augmenting paths */
 - 7: $M^t \leftarrow M \oplus p_1 \oplus \dots \oplus p_r$ where $\{p_1, p_2, \dots, p_r\} = \mathcal{P}^t$
 - 8: **while** there is an M^t -alternating path $x = (v_0, v_1, \dots, v_{2m})$ of type ii . in $G^n(M^t \cap E^{-n})$
for some $n \in \{A, B\}$ **do**
 - 9: Assume $(v_0, v_1) \in E^I \cap M^t$ with $v_0 \in V^{-n}$ and $v_1 \in V^n$.
 - 10: $j \leftarrow \max_{i=0, \dots, 2m-1} \{i : (v_i, v_{i+1}) \in q \text{ for some } q \in \mathcal{P}^t\}$
 - 11: $y \leftarrow (u_0, u_1, \dots, u_k, u_{k+1}, \dots, u_f) \in \mathcal{P}^t$ used to determine j with $(u_k, u_{k+1}) = (v_j, v_{j+1})$
 - 12: $z \leftarrow (v_{2m}, v_{2m-1}, \dots, v_{j+1}, u_{k+2}, \dots, u_f)$
 - 13: $M^{t+1} \leftarrow M^t \oplus y \oplus z$
 - 14: $\mathcal{P}^{t+1} \leftarrow (\mathcal{P}^t - \{y\}) \cup \{z\}$
 - 15: $t \leftarrow t + 1$
 - 16: $G' \leftarrow$ subgraph of $G^n(M^t \cap E^{-n})$ induced by considering only edges of x from v_0 to $v_j = u_k$ and of y from u_0 to $u_k = v_j$
 - 17: **if** there is a $x \leftarrow M^t$ -alternating path of type ii . in G' starting in (v_0, v_1) **go to** step 10
 - 18: **end while**
 - 19: **return** M^t .
-

In what follows we provide a proof of the correctness of this algorithm. For sake of clarity, first of all, we provide an illustration of how the algorithm works by applying it to a 2-KEG instance.

Example 1 Consider the 2-KEG instance represented in Figure 4.4.

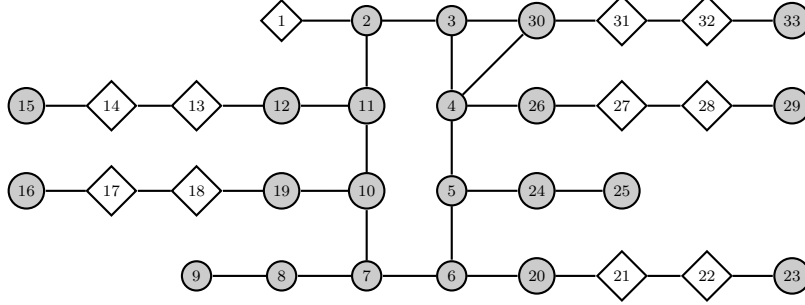


Fig. 4.4: A 2-KEG instance.

A Nash equilibrium M that is not a maximum matching is represented by bold edges in the top-left graph of Figure 4.5. The matching M is a Nash equilibrium, since there is no M -alternating path as stated in Theorem 6; and it is not a maximum matching because there are M -augmenting paths, *e.g.*, $(25, 24, 5, 6, 20, 21, 22, 23)$. We will apply Algorithm 4.1.1 to this NE in order to achieve one that is a SWE and dominates it.

The algorithm starts by computing an arbitrary maximum matching S , represented in the top-right graph of Figure 4.5; the symmetric difference between M and S is represented in the center-left graph of that figure. There are 6 connected components in $M \oplus S$, three of which include M -augmenting paths, namely

$$\mathcal{P}^1 = \{(33, 32, 31, 30, 3, 4, 26, 27, 28, 29), (25, 24, 5, 6, 20, 21, 22, 23), (15, 14, 13, 12, 11, 10, 19, 18, 17, 16)\}.$$

Therefore, at the end of step 7 we obtain a maximum matching M^1 , represented at the center-right of Figure 4.5.

The algorithm proceeds searching for an M^1 -alternating path of type ii. in $G^n(M^1 \cap E^{-n})$ for some $n \in \{A, B\}$, i.e., the algorithm will check if M^1 is a NE. In this step, path $x = (1, 2, 3, 4, 5, 6, 7, 8, 9)$ is found, which shows that M^1 is not an equilibrium. The M -augmenting path $y = (25, 24, 5, 6, 20, 21, 22, 23)$ is replaced by $z = (9, 8, 7, 6, 20, 21, 22, 23)$, leading to matching M^2 represented at the bottom-left graph of Figure 4.5. Next, step 17 is used to verify if there is an M^2 -alternating path of type ii. considering only the edges $(1, 2), (2, 3), (3, 4), (4, 5), (5, 24), (24, 25)$. There is one, namely path $(1, 2, 3, 4, 5, 24, 25)$. The M -augmenting path $(33, 32, 31, 30, 3, 4, 26, 27, 28, 29)$ is modified into $(25, 24, 5, 4, 26, 27, 28, 29)$,

obtaining M^3 represented in the lower-right graph of Figure 4.5. In the next iteration no M^3 -alternating path of type ii. can be found, and thus the algorithm terminates. M^3 is a SWE that dominates M .

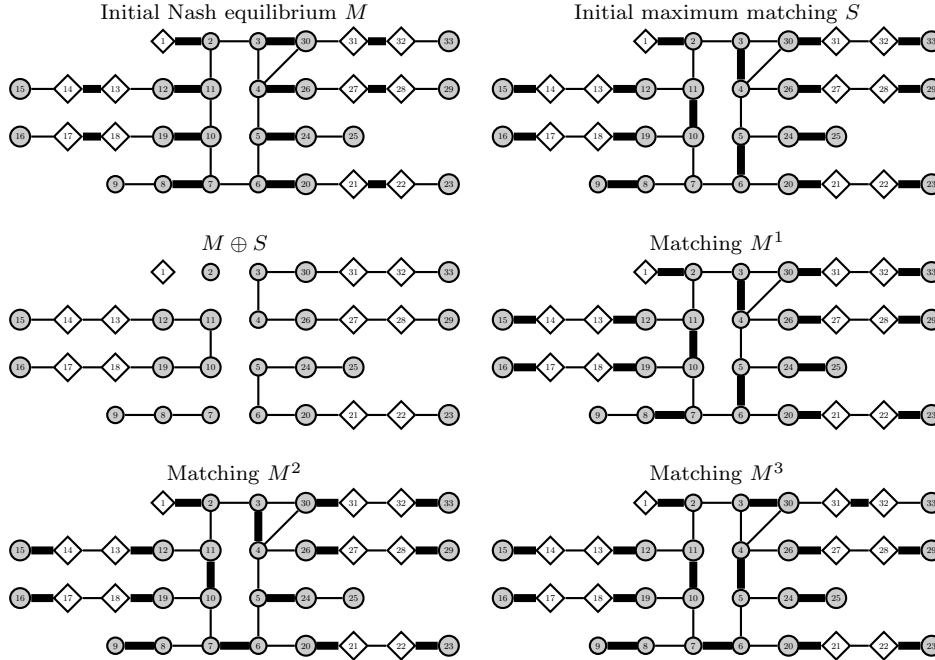


Fig. 4.5: Computation of a dominant SWE in the 2-KEG instance of Figure 4.4 starting from the initial equilibrium in the top-left graph, and the initial maximum matching of top-right graph.

Next we will prove that for any socially suboptimal NE, the Algorithm 4.1.1 returns a dominant SWE.

The algorithm starts by computing a maximum matching S . If the Nash equilibrium from the input is a maximum matching, the algorithm returns it and stops. Otherwise, it proceeds. At iteration t , \mathcal{P}^t is the set of M -augmenting paths used to compute the maximum matching M^t . In this way, step 7 augments M in order to obtain a maximum matching M^1 . Note that $|\mathcal{P}^1|$ augmenting paths of M are used in order to get M^1 and that the symmetric difference of a matching with an associated augmenting path only adds additional covered vertices. Therefore, none of the M -matched vertices is M^1 -unmatched, which shows that the players' utilities associated with M^1 are equal to or greater than the ones achieved through M .

Note that if there is an M^1 -alternating path of type *i.* or *iii.*, then it is also an augmenting path of M^1 contradicting the fact that M^1 is a maximum matching. Therefore, by Theorem 6, if M^1 is not a Nash equilibrium then there

is an M^1 -alternating path of type $ii.$ in $G^A(M^1 \cap E^B)$ or $G^B(M^1 \cap E^A)$. In this case, the algorithm will remove the M^1 -alternating path of type $ii.$ through steps 8 to 15. In these steps an M -augmenting path $y \in \mathcal{P}^1$ is replaced by a new M -augmenting path z . Thus, it is obvious that the new maximum matching M^2 dominates the utilities achieved through M .

Suppose that in step 8 an M^t -alternating path x of type $ii.$ is found. Since M is a NE, the path x cannot be M -alternating. Thus, x intersects at least one M^t -matched edge of a $y \in \mathcal{P}^t$. The algorithm picks such y accordingly with the one closest to v_{2m} , since this rule ensures that y never intersects x from $v_{j+1} = u_{k+1}$ to v_{2m} . Then, through step 13, v_{2m} is made M^{t+1} -matched, which eliminates the M^t -alternating path x of type $ii.$. See Figure 4.6 for illustration.

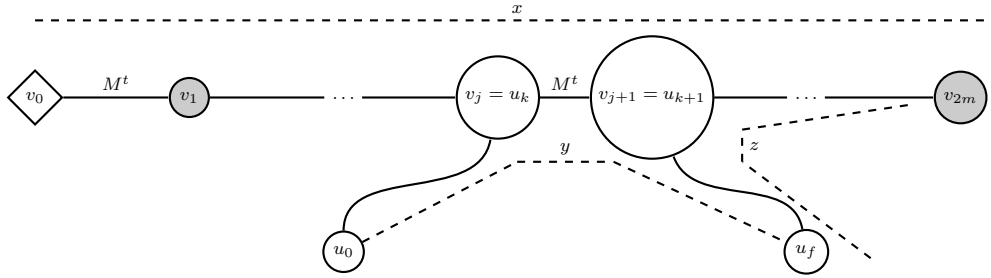


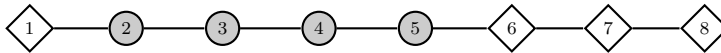
Fig. 4.6: Modification of y to z through x . White circle vertices mean that there is no need to specify the player to which the vertices belong.

So far, we proved that at any iteration t of Algorithm 4.1.1, the current maximum matching M^t dominates M and that if there is an M^t -alternating path of type $ii.$, we eliminate it in the next maximum matching M^{t+1} . It remains to show that the elimination of paths of type $ii.$ will stop, leading to a SWE.

By construction, the size of the augmenting path sets is maintained during the algorithm execution. Indeed, in each iteration, an M -augmenting path is replaced by a new one.

Lemma 4 $|\mathcal{P}^t| = |\mathcal{P}^k| \quad \forall t, k \geq 1.$

For an M -augmenting path $y = (u_0, u_1, \dots, u_f)$, define $\sigma(y)$ as the number of times that y switches the player's graph plus one unit if the first internal edge that follows the extreme $u_0 \in V^i$ is in E^{-i} , and plus one unit if the last internal edge that precedes the extreme $u_f \in V^j$ is in E^{-j} . For instance, the path



has σ -value equal to 3: count two unities because, the first extreme vertex, 1, is in V^B while the following internal edge, (2, 3), is in E^A and add 1 unit because the rest of the path is in E^B . Indeed, the σ -value of M -augmenting paths has to be greater or equal to two, otherwise it is not a Nash Equilibrium (*i.e.*, there is an M -alternating path as described in Theorem 6, or the independent agent is not choosing a maximum matching as obliged by the game rule). The following lemma states that the σ -value of the paths in \mathcal{P}^t is non-increasing.

Lemma 5 *In an iteration t of Algorithm 4.1.1 $\sigma(y) \geq \sigma(z)$.*

Proof Consider an arbitrary iteration t of Algorithm 4.1.1. Without loss of generality, assume that the M^t -alternating path x of type *ii*. found is in $G^A(M^t \cap E^B)$.

In step 11, $y = (u_0, u_1, \dots, u_f)$ is the selected augmenting path in \mathcal{P}^t . In order to get z , the part of y from u_0 to u_k is replaced by a path that has all the edges in $E^A \cup E^I$. Note that there must be an internal edge in y after u_{k+1} , otherwise M is not an equilibrium: the path $(u_f, u_{f-1}, \dots, u_{k+1}, v_{j+2}, v_{j+3}, \dots, v_{2m})$ would be an M -alternating path in $G^A(M \cap E^B)$ satisfying one of the conditions of Theorem 6. Thus, we continue the proof by distinguishing two possible cases: the first internal edge in y after u_{k+1} is in E^B or E^A .

Case 1: The first internal edge in y after u_{k+1} is in E^B . Then, $\sigma(z)$ is equal to one plus the number of times that the path y from u_{k+1} to u_f switches the player's graph plus one unit if the last internal edge before $u_f \in V^i$ is in E^{-i} . Observe that $\sigma(y)$ is greater or equal to the number of times that the path y from u_{k+1} to u_f switches the player's graph plus one unit if the last internal edge before $u_f \in V^i$ is in E^{-i} . In order to get equal, the part of y from u_0 to u_{k+1} must have the edges in $E^B \cup E^I$ and $u_0 \in E^B$. However, this contradicts the fact that M is a Nash equilibrium: one of the vertices u_k or u_{k+1} has to be in V^A , otherwise y is not in player A 's graph. If $u_{k+1} \in V^A$, then $u_{k+2} \in V^B$, which means that the part of x from v_{2m} to (u_{k+1}, u_{k+2}) is an M -alternating path of type *ii*. in $G^A(M \cap E^B)$. Otherwise, if $u_k \in V^A$, then $u_{k-1} \in V^B$ and the part of y from u_0 to u_k is an M -alternating path of type *ii*. in $G^B(M \cap E^A)$. In conclusion, $\sigma(y) \geq \sigma(z)$.

Case 2: The first internal edge in y after u_{k+1} is in E^A . Then, $\sigma(z)$ is equal to the number of times that the path y from u_{k+1} to u_f switches the player's graph plus one unit if the last internal edge before $u_f \in V^i$ is in E^{-i} . Note that $\sigma(y)$ is greater or equal to the number of times that the path y from u_{k+1} to u_f switches the player's graph plus one unit if the last internal edge before $u_f \in V^i$ is in E^{-i} . In conclusion, $\sigma(y) \geq \sigma(z)$.

□

An immediate consequence is the following corollary.

Corollary 2 *If $\sigma(y) > \sigma(z)$ holds in iteration t , then z will never evolve during the rest of the algorithm to be equal to y .*

Proof Assume that $\sigma(y) > \sigma(z)$ in iteration t . By Lemma 5, if z is selected in a forthcoming iteration then the resulting (modified) path has a σ -value less or equal to $\sigma(z)$ and, in particular, less than $\sigma(y)$. Therefore, it is impossible that from iteration z this path evolves to y , since that contradicts Lemma 5. \square

Whenever Algorithm 4.1.1 at iteration t modifies y such that $\sigma(y) > \sigma(z)$, we get that the maximum matching M^t will never be computed again in later iterations.

Corollary 3 *Algorithm 4.1.1 can only cycle after iteration t if $\sigma(y) = \sigma(z)$.*

Now, we will prove that when a modification of an augmenting path y to z has $\sigma(y) = \sigma(z)$, then the algorithm finds an M^{t+1} -alternating path of type *ii.* in step 17. This particular search for such a path is the important ingredient for the algorithm to stop after a finite number of iterations. If we remove this step from Algorithm 4.1.1 and we simply arbitrarily search for the elimination of paths of type *ii.* then the algorithm can cycle. For instance, in Example 1, when we are in iteration 2 and we do not perform the search as stated in step 17, then we can compute the M^2 -alternating path $(1, 2, 11, 10, 7, 6, 5, 24, 25)$ that would lead us to $M^3 = M^1$, making the algorithm to cycle.

Lemma 6 *If $\sigma(y) = \sigma(z)$ at the end of step 15 of Algorithm 4.1.1, then a path of type *ii.* is found in step 17.*

Proof Suppose that the algorithm is in the end of step 15. Without loss of generality, the proof concentrates only on the case for which x is in $G^A(M^{t-1} \cap E^B)$, since for x in $G^B(M^{t-1} \cap E^A)$ the proof is analogous.

We will make use of Lemma 5 proof in order to conclude that under the lemma hypothesis, $\sigma(y) = \sigma(z)$, the edges of y from u_0 to u_k are in $E^A \cup E^I$. Case 1 of that proof implies that in order to get $\sigma(y) = \sigma(z)$, the edges of the path y from u_0 to u_k should be in $E^A \cup E^I$ and $u_0 \in V^A$. In order to get $\sigma(y) = \sigma(z)$ in case 2, we also get that the edges of the path y from u_0 to u_k should be in $E^A \cup E^I$ and $u_0 \in V^A$.

Next, we will show that there is an M^t -alternating path of type *ii.* from (v_0, v_1) to u_0 that only uses the edges of x from v_0 to v_j and y from u_0 to u_k . Therefore, for sake of clarity, consider $y' = (u_0, u_1, \dots, u_k)$ and $x' = (v_0, v_1, v_2, \dots, v_j)$. Recall that $u_k = v_j$.

In step 17, the new M^t -alternating path of type *ii.* x can be built as follows. Start to follow x' from v_0 until it intersects a vertex u_{j_1} in y' (note that y' intersects x' at least in $u_k = v_j$). Consider the following possibilities.

Case 1: If $(u_{j_1}, u_{j_1-1}) \in M^t$, then $x = (v_0, v_1, \dots, u_{j_1}, u_{j_1-1}, \dots, u_0)$ is an M^t -alternating path of type *ii.*

Case 2: If $(u_{j_1}, u_{j_1+1}) \in M^t$, then $(u_{j_1}, u_{j_1-1}) \in M^{t-1}$ and $(u_{j_1}, u_{j_1-1}) \in x'$, which implies $u_{j_1+1} \notin x'$. Follow y' by index increasing order starting in u_{j_1+1} until it is reached a vertex $u_{j_2} = v_{i_1}$ of x' (note that such vertex exists since at least $u_k = v_j \in x'$, with $k > j_1 + 1$). The vertex

$u_{j_2-1} \notin x'$, otherwise, we would have stopped in u_{j_2-1} . Thus, $(u_{j_2}, u_{j_2-1}) \notin M^{t-1}$. Otherwise, x' would not be an M^{t-1} -alternating path. In conclusion, $(u_{j_2}, u_{j_2-1}) \in M^t$.

Next, we follow x' by index decreasing order starting in $u_{j_2} = v_{i_1}$ until we intersect a vertex u_{j_3} of y' (which has to occur, since we noted before that at least u_{j_1-1} is in x'). If $(u_{j_3}, u_{j_3-1}) \in M^t$, then the rest of the M^t -alternating is found as in case 1. Otherwise, $(u_{j_3}, u_{j_3+1}) \in M^t$ and we proceed as in the beginning of case 2. This process will terminate in u_0 since we are always adding new vertices to our M^t -alternating path and the number of vertices is finite.

□

Corollary 4 *The algorithm can only cycle if it remains in steps 15 to 17.*

Theorem 7 *After a finite number of executions of steps 15 to 17, the algorithm fails to find such a path in step 17.*

Proof The length of the path $(v_0, v_1, v_2, \dots, v_j)$ considered in step 17 strictly decreases in each consecutive execution of steps 15 to 17. □

As a corollary of the above Theorem we can now state the desired result.

Corollary 5 *After a finite number of iterations, the Algorithm 4.1.1 stops and finds a SWE that dominates the NE given in the input.*

5 Refinement of SWE

Although Algorithm 3.2.1 computes a SWE, the results obtained in Section 4 (see Theorem 6) allow the definition of a simpler polynomial time algorithm returning a SWE. Furthermore, the algorithm will solve another aspect left open in the previous sections where we discussed the advantage of SWE among the set of NE for 2-KEG. This refinement to select a NE is still not sufficient to get uniqueness, *i.e.*, there are 2-KEG instances for which there is more than one SWE. The algorithm presented in this section will solve this issue.

Example 2 Consider the 2-KEG instance represented in Figure 5.1. There are four maximum matchings M^1 to M^4 , of which matchings M^1 and M^2 are NE (SWE). Under M^1 player A has utility 4 and player B has utility 2; in contrast, under M^2 both players have utility 3.

This instance has two distinct SWE, and by repeating the relevant pattern we can create instances with multiple distinct SWE. For example, the game of Figure 5.2 has eight SWE.

In this context it seems rational to search for the social welfare equilibrium that minimizes the number of external exchanges, since that decreases the dependency of the players on each other; in practice, this seems to be a more

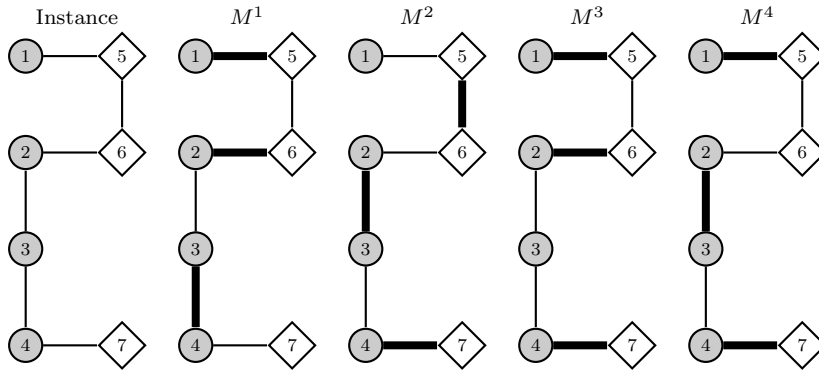


Fig. 5.1: Example of a 2-KEG instance with four maximum matchings, and two SWE M^1 and M^2 .

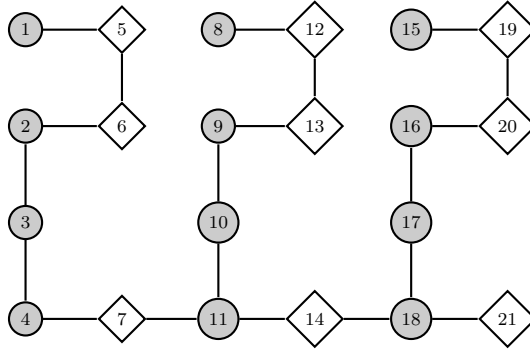


Fig. 5.2: Example of a 2-KEG instance with eight SWE.

desirable solution. Therefore, in what follows, we will show how to find such an equilibrium in polynomial time.

Consider Algorithm 5.0.1. This algorithm based on the number of vertices, $|V|$, it associates weight $2 + 2|V|$ for internal edges and weight $1 + 2|V|$ for external edges. Then, a maximum weight matching is returned. We will prove that this algorithm can be executed in polynomial time and that it computes a social welfare equilibrium that minimizes the number of external exchanges.

Lemma 7 *Algorithm 5.0.1 can be executed in polynomial time.*

Proof It is a well-known result that weighed matching problems can be solved in polynomial time (see, *e.g.*, [20]). Therefore, step 7 can be executed in polynomial time. Additionally, the attribution of weights for the graph edges is linear in the number of edges. Therefore, the algorithm can run in polynomial time. \square

Algorithm 5.0.1 Computation of the social welfare equilibrium that minimizes the number of external exchanges.

Input: a 2-KEG instance $G = (V, E)$

Output: a SWE that minimizes the number of external exchanges

```

1: for  $e$  in  $E^A \cup E^B$  do
2:    $w_e \leftarrow 2 + 2|V|$ 
3: end for
4: for  $e$  in  $E^I$  do
5:    $w_e \leftarrow 1 + 2|V|$ 
6: end for
7:  $M \leftarrow$  maximum weight matching in  $G$  given edge weights  $w_e, \forall e \in E$ 
8: return  $M$ 

```

In order to prove that Algorithm 5.0.1 outputs a SWE we need to prove that M is a maximum matching and a NE.

Lemma 8 *Algorithm 5.0.1 returns a maximum matching.*

Proof In step 7 of the algorithm, the maximum weight on an edge in the maximum weight matching problem considered is $2 + 2|V|$. Thus, any matching of size k has a total weight not greater than $k(2 + 2|V|)$. If that is not a maximum matching, *i.e.*, if $k < |S|$, where S is a maximum matching for G , the total weight is bounded above by

$$k(2+2|V|) = 2k(1+|V|) \leq 2(|S|-1)(1+|V|) = 2|S||V| + 2(|S|-|V|-1) < 2|S||V|,$$

where the last inequality comes from the fact that $|S| < |V|$.

A maximum matching on the graph game has a total weight at least equal to $|S|(1 + 2|V|) = |S| + 2|S||V|$. Therefore, a maximum matching has always a total weight greater than any non maximum matching. In conclusion, a maximum weight matching with the proposed edge weights is also a matching with maximum cardinality. \square

Lemma 9 *Algorithm 5.0.1 returns a NE.*

Proof Let M be the output of Algorithm 5.0.1.

By Lemma 8 we know that M is a maximum matching. If M is not a NE, then some player must have incentive to deviate; without loss of generality, assume that player A has incentive to deviate from $M \cap E^A$. Then, there must be an M -alternating path p of type *ii*. in $G^A(M \cap E^B)$ such that $M \oplus p$ increases player A 's utility

$$2|(M \oplus p) \cap E^A| + |(M \oplus p) \cap E^I| > 2|M \cap E^A| + |M \cap E^I|.$$

On the other hand, the matching $|M \oplus p|$ must have a total weight not greater than the one associated with M , *i.e.*,

$$\begin{aligned} (2 + 2|V|)|M \cap E^A| + (2 + 2|V|)|M \cap E^B| + (1 + 2|V|)|M \cap E^I| &\geq \\ (2 + 2|V|)|(M \oplus p) \cap E^A| + (2 + 2|V|)|(M \oplus p) \cap E^B| + (1 + 2|V|)|(M \oplus p) \cap E^I|. \end{aligned}$$

Since the path p only uses the edges in $E^A \cup E^I$, the set $M \cap E^B$ is equal to $(M \oplus p) \cap E^B$. Hence, in this inequality, we can remove the second term of both sides and rewrite as

$$\overbrace{2|M \cap E^A| + |M \cap E^I| - 2|(M \oplus p) \cap E^A| - |(M \oplus p) \cap E^I|}^{<0} + 2|V| (|M \cap E^A| + |M \cap E^I| - |(M \oplus p) \cap E^A| - |(M \oplus p) \cap E^I|) \geq 0.$$

Player A 's utility is bigger with $M \oplus p$ than with M . Thus, in this inequality the first four terms lead to a negative number. This implies that

$$|M \cap E^A| + |M \cap E^I| > |(M \oplus p) \cap E^A| + |(M \oplus p) \cap E^I| \geq 0,$$

which is impossible since, M and $M \oplus p$ have the same cardinality and, in particular, $|M \cap (E^A \cup E^I)| = |(M \oplus p) \cap (E^A \cup E^I)|$. \square

Finally, it remains to prove that Algorithm 5.0.1 returns a matching that minimizes the number of external edges on it among the set of SWE.

Lemma 10 *Algorithm 5.0.1 outputs a matching that minimizes the number of external edges among the set of social welfare equilibria.*

Proof Let M be the matching returned by Algorithm 5.0.1. We will prove by showing that assuming another SWE M' contains more internal exchanges than M leads to a contradiction. Since both M and M' are maximum matchings, M' has a total weight greater than M ; but this contradicts the fact that the algorithm returns a maximum weight matching (where the internal edges weight more than the external ones). \square

The next theorem concludes this section.

Theorem 8 *Algorithm 5.0.1 computes a SWE that minimizes the number of external exchanges in polynomial time.*

Unfortunately, for some 2-KEG instances this refinement of the SWE still does not lead to an unique solution.

Example 3 Consider the 2-KEG instance of Figure 5.3. There are two SWE that minimize the number of external exchanges, M^1 and M^2 . These matchings lead both players to an utility of 3.

However, the players utilities under social welfare equilibria that minimize the number of external exchanges are unique as we will prove next.

Lemma 11 *In any SWE that minimizes the number of external exchanges, for a fixed instance, the player's utilities are always the same.*

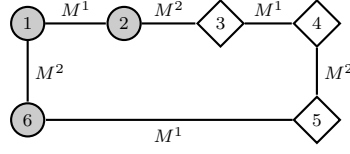


Fig. 5.3: Example of a 2-KEG instance with two distinct SWE that lead both players to same profit.

Proof Consider an instance of 2-KEG for which there are two different SWE minimizing the number of external exchanges, say M^1 and M^2 , of Algorithm 5.0.1. The proof is by contradiction, by assuming that player A 's utilities with M^1 and M^2 are different. Without loss of generality,

$$2|M^1 \cap E^A| + |M^1 \cap E^I| > 2|M^2 \cap E^A| + |M^2 \cap E^I|.$$

Build the subgraph H of G induced by the edges in the set $(M^1 \oplus M^2) \cap (E^A \cup E^I)$. As player A covers more of her vertices through M^1 than through M^2 , there must be at least one vertex $a \in V^A$ such that a is M^1 -matched and M^2 -unmatched. Consider each distinct component p of H ; p is a path starting in, say, vertex a . There are three possible cases. Namely,

- Case 1: path p terminates in an M^2 -matched vertex of V^A . Then, it is not this component that gives advantage to M^1 .
- Case 2: path p terminates in an M^2 -matched vertex of V^B . Then, p is an M^2 -alternating path of type $ii.$; by Lemma 9, this contradicts the fact that M^2 is a NE.
- Case 3: path p terminates in an M^1 -matched vertex. Then, p is an augmenting path to M^2 ; by Lemma 9, this contradicts the fact that M^2 is a maximum matching. \square

We finish this section by noting that another desirable SWE is that in which the difference of players' utilities is minimized, *i.e.*, the discrepancy of the players' utilities is minimized traducing in a more "fair" outcome. It is easy to show that the social welfare equilibrium introduced in this section, *i.e.*, that minimizing the number of external matchings achieves simultaneously the goal of minimizing the difference of players' utilities.

Theorem 9 *If M is the SWE with minimum number of external matchings then, it also the SWE that minimizes the difference of players' utilities.*

Proof Let M^A , M^B and $M^I(M^A, M^B)$ be the social welfare equilibrium that minimizes the number of external matchings. Let R^A , R^B and $M^I(R^A, R^B)$ be the social welfare equilibrium that minimizes the difference in the players utilities, *i.e.*, the value of $|2|R^A| + |M^I(R^A, R^B)| - 2|R^B| - |M^I(R^A, R^B)|| = ||R^A| - |R^B||$ is the minimum among all social welfare equilibria.

If $|M^I(M^A, M^B)| = |M^I(R^A, R^B)|$, then the matching $R^A \cup R^B \cup M^I(R^A, R^B)$ is also a SWE that minimizes the number of external matchings. Thus, by the uniqueness of the players' utilities under this refinement of the SWE, $M^A \cup M^B \cup M^I(M^A, M^B)$ also minimizes the difference of players' utilities.

If $|M^I(M^A, M^B)| \neq |M^I(R^A, R^B)|$ then, $|M^A| + |M^B| > |R^A| + |R^B|$ since, by hypothesis $|M^I(M^A, M^B)| < |M^I(R^A, R^B)|$ and both matchings have maximum cardinality. Without loss of generality, there must be a path p that starts and ends in M^A -matched vertices and alternates between edges in M^A and edges in R^A . Matching $R^A \cup R^B \cup M^I(R^A, R^B)$ is a NE which implies that p cannot be a path as described in Theorem 6. Therefore, the extreme vertices of p must be $M^I(R^A, R^B)$ -matched which does not show any advantage of $M^A \cup M^I(M^A, M^B)$ and $R^A \cup M^I(R^A, R^B)$ over each other in terms of player A 's utility. In this way, it follows that both matchings lead to the same profit for both players. \square

In conclusion, one may argue that the players will converge to social welfare equilibria since, given any Nash equilibrium, both players can improve their utilities through a SWE. Additionally, choosing a SWE that minimizes the number of external exchanges is a desirable propriety for both players, and we demonstrated that such equilibrium can be found in polynomial time. Moreover, players are indifferent among such equilibria, because utilities remain the same for any of them. Thus, it seems reasonable to consider that the players will agree in the SWE to be played.

6 Model Extensions

In what follows, we discuss extensions to the results when our assumptions (exchanges size players' objectives and number of players) are relaxed.

A common problem of these extensions is that the IA decision may become undefined (in contrast with Proposition 1), in the sense that there might exist more than one optimal solution maximizing the number of external exchanges that would benefit the players differently. In order to deal with this issue, we could, for example, impose a public preference on the external exchanges to the IA, associate a probability for each equivalent optimal solution of the IA or assume that the players are pessimistic/optimistic about the IA decision.

Relaxation of exchanges maximum size to $L > 2$. In the literature about kidney exchange programs, besides cycles of size two (matchings), typically cycles of size three (3-way exchanges) are allowed. In the latter case, we conjecture that (recall the notation introduced to N -KEG in Problem 1.1)

$$\Phi(x^A, x^B) = \sum_{c \in C^A} w_c x_c^A + \sum_{c \in C^B} w_c x_c^B + \sum_{c \in I: w_c^A = w_c^B = 1} y_c + \frac{3}{2} \sum_{c \in I: w_c^A = 2 \vee w_c^B = 2} y_c$$

is a (*non-exact*) potential function and thus, a maximum is a NE. However, for general values of L the game may fail to have a pure Nash equilibrium,

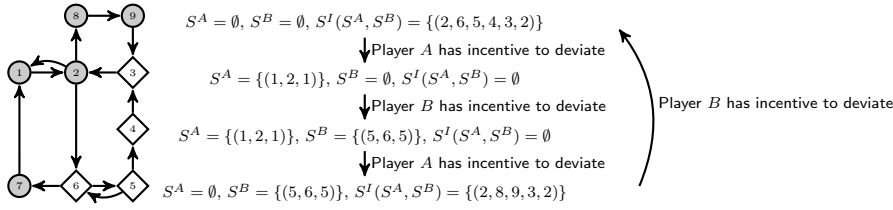


Fig. 6.1: A game instance with $L = 5$. Player A can select $(1, 2, 1)$ or \emptyset ; Player B can select $(5, 6, 5)$ or \emptyset . Let S^P be player P internal exchange program, for $P = A, B$, and $S^I(S^A, S^B)$ the IA external exchange program. The diagram on the right hand side of the graph shows that none of the (pure) game outcomes is a Nash equilibrium (implying that the game cannot be potential).

as shown in Figure 6.1. The main difference when $L > 3$ is that in this case external cycles may help strictly more patients of a same player than an internal exchange, while for $L = 3$ an external exchange helps at most as many patients as an internal one.

Besides cyclic exchanges, researchers have also included *chains*, where, there is an altruistic donor starting the exchange (see Figure 6.2).

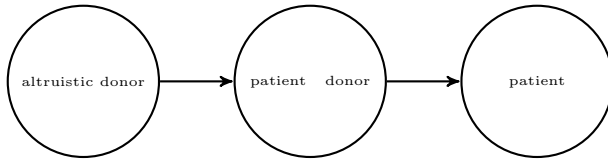


Fig. 6.2: Example of a chain of size 2.

Allowing exchanges beyond matchings ($L = 2$) is an extension with positive impact in the social optimum, and it calls for studying the existence of pure Nash equilibria with good social properties.

Change in Players' objective functions. Investigating different players' utilities is of crucial importance, the literature on the kidney exchange program being rich of examples analyzing different solution selection criteria (*e.g.*, see [11]).

A simple extension would be to assume that the players prioritize maximum matchings that maximize “hard-to-match” vertices. In this case, we could still have a SWE. We first compute a SWE for 2-KEG. If this SWE is not an equilibrium for this extension, then, w.l.o.g., there is a M -unmatched vertex $a \in V^A$ hard-to-match and a $M^A \cup M^I(M^A, M^B)$ -alternating path p that terminates in a player A M -matched vertex that is not hard-to-match. Because the maximum matching $M' = M \oplus p$ improves player A utility and does not

create alternating paths of type *ii*. (see Theorem 6), we just need to repeat this process until no player has incentive to deviate.

However, for more complicated players' utilities the game may fail to have pure Nash equilibria. For instance, consider the compatible graph of Figure 6.3. The IA behavior remains as before: maximize the number of external exchanges among the available vertices; be indifferent between the players' evaluation of the different matchings; have a deterministic decision, this is, for any combination of the players' strategies (internal matchings) the external exchange selected by the IA is known. In Figure 6.4, we have all the possible outcomes for the game. Observe that none of these 4 possible outcomes is a Nash equilibrium and thus, no pure equilibrium exists.

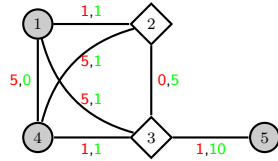


Fig. 6.3: The players' utility of each matching is given by the numbers in the edges: player *A* value is in red and player *B* value in green.

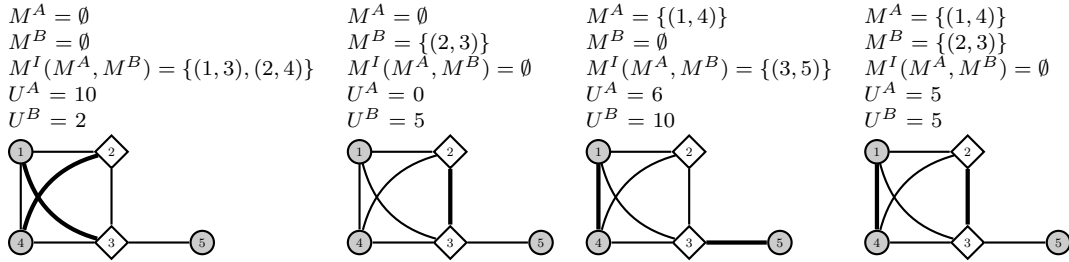


Fig. 6.4: U^A and U^B are player *A* and *B* utilities, respectively.

Another extension in this context is to Bayesian games. In this case, the players would not know their opponents evaluations/utilities for the exchanges. Under this incomplete information scenario, it would be interesting to explore how the players can build beliefs about the opponents' objectives by repeatedly observing the game outcomes and, thus, use them to compute (Bayesian) equilibria.

Increase number of players to $N > 2$. Extending our results about the existence of a NE and a SWE dominating it is immediate. Let $\{1, 2, \dots, N\}$ be

the set of players. Then, (by extending our notation in an obvious way)

$$\Phi(M^1, M^2, \dots, M^N) = \sum_{P=1}^N 2|M^P| + |M^I(M^1, M^2, \dots, M^N)|$$

is a (non-exact) potential function, and a optimum of it is a NE. The function is potential, since whenever a player increases her utility it is because she is increasing the number of internal exchanges. An increase in the number of internal exchanges has a greater impact in the value of Φ than external exchanges. The results in Section 4 remain valid in this setting. The ideas presented analyze each player's incentives for deviation, which hold for more than 2 players, because we can think of a player opponents' as a single one (reducing the study to 2-KEG).

It remains to investigate, if there is a NE in which the players would agree to choose.

7 Conclusions

In this paper, we have shown that 2-KEG has always a pure Nash equilibrium and that it can be computed in polynomial time. Furthermore, we have proven the existence of a NE that is also a social optimum. Finally, and more importantly, we have shown that for any NE there is always a social welfare Nash equilibrium that is a preferred outcome for both players.

There is no uniqueness result for social welfare equilibria. In order to find rational guidelines for the players' strategies, we add to the social welfare equilibrium the requirement that it must be the one that minimizes the number of external exchanges. For this type of solution, we were able to prove uniqueness in terms of the players' utilities and to show that it can be efficiently computed, thus strengthening the fact that this is a realistic outcome for the game.

Although we show that a social welfare equilibrium can be computed in polynomial time, a full characterization of the Pareto frontier of social welfare equilibria (with respect to pure Nash equilibria) remains to be done. This is an interesting subject for future research.

Our work also indicates that studying the players interaction through 2-KEG turns the exchange program efficient both from the social welfare and the players' point of view. These results motivate further research in the generalization of the game to more than two players, to exchanges including more than two patient-donor pairs and to different evaluation metrics of the exchanges. Some of these generalizations have been preliminarily discussed in Section 6.

Additional inspiration for future research is given by the recent paper [14], where a strategyproof mechanism for a multi-period dynamic model was shown to lead to a global maximum matching that cannot be guaranteed by a mechanism for the static case. Therefore, given that 2-KEG already provides such

solution as a rational outcome in the static case, investigating the 2-KEG by playing it repeatedly as the players' pools of patient-donor pairs change over time would be another line to explore in the future work.

Acknowledgments

We are indebted with Nicolás Stier-Moses for reading a preliminary version of the paper and providing useful feedbacks.

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) within projects PTDC/IIMGES/2830/2014 (mKEP - Models and optimisation algorithms for multi-country kidney exchange programs).

The first author acknowledges the support of Fundação para a Ciência e a Tecnologia (FCT) through a PhD grant number SFRH/BD/79201/2011 (POPH/FSE program). The second author acknowledges the support of MIUR under the PRIN2012 grant. We would like to thank the anonymous referees for their valuable comments that significantly contributed to the improvement of this work.

References

1. ABRAHAM, D. J., BLUM, A., AND SANDHOLM, T. Clearing algorithms for barter exchange markets: enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM conference on Electronic commerce* (New York, NY, USA, 2007), EC '07, ACM, pp. 295–304.
2. ASHLAGI, I., FISCHER, F., KASH, I. A., AND PROCACCIA, A. D. Mix and match: A strategyproof mechanism for multi-hospital kidney exchange. *Games and Economic Behavior*, 91 (2015), 284 – 296.
3. ASHLAGI, I., AND ROTH, A. Individual rationality and participation in large scale, multi-hospital kidney exchange. In *Proceedings of the 12th ACM conference on Electronic commerce* (New York, NY, USA, 2011), EC '11, ACM, pp. 321–322.
4. ASHLAGI, I., AND ROTH, A. Individual rationality and participation in large scale, multi-hospital kidney exchange. *Working paper*, http://web.mit.edu/iashlagi/www/papers/LargeScaleKidneyExchange_1_13.pdf (2011).
5. BERGE, C. Two theorems in graph theory. *Proceedings of the National Academy of Sciences* 43, 9 (1957), 842–844.
6. BONDY, J. A., AND MURTY, U. S. R. *Graph theory with applications*. Elsevier Science Publishing Co., Inc., 1976.
7. CARAGIANNIS, I., FILOS-RATSIKAS, A., AND PROCACCIA, A. D. An improved 2-agent kidney exchange mechanism. In *Internet and Network Economics*, N. Chen, E. Elkind, and E. Koutsoupias, Eds., vol. 7090 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011, pp. 37–48.
8. CECHLÁROVÁ, K. AND FLEINER, T. AND MANLOVE, D. The kidney exchange game. *Proc. of the 8th International Symposium on Operational Research SOR*, 5 (2005), 77 – 83.
9. CONSTANTINO, M., KLIMENTOVA, X., VIANA, A., AND RAIS, A. New insights on integer-programming models for the kidney exchange problem. *European Journal of Operational Research* 231, 1 (2013), 57 – 68.
10. DE KLERK, M., KEIZER, K. M., CLAAS, F. H., HAASE-KROMWIJK, B. J. J. M., AND WEIMAR, W. The Dutch national living donor kidney exchange program. *American Journal of Transplantation* 5 (2005), 2302–2305.

-
11. DICKERSON, J. P., PROCACCIA, A. D., SANDHOLM, T. Price of Fairness in Kidney Exchange. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems AAMAS '14* (2014), 1013–1020.
 12. EDMONDS, J. Paths, trees, and flowers. *Canad. J. Math.* 17 (1965), 449–467.
 13. FUDENBERG, D., AND TIROLE, J. *Game Theory*, 5th ed. MIT Press, Cambridge, MA, 1996.
 14. HAJAJ, C., DICKERSON, J. P., HASSIDIM, A., SANDHOLM, T., AND SARNE, D. Strategy-proof and efficient kidney exchange using a credit mechanism. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 921–928, 2015.
 15. KÖPPE, M., RYAN, C. T., AND QUEYRANNE, M. Rational generating functions and integer programming games. *Oper. Res.* 59, 6 (Nov. 2011), 1445–1460.
 16. MANLOVE, D. F., AND O'MALLEY, G. Paired and altruistic kidney donation in the UK: algorithms and experimentation. In *Experimental Algorithms*, R. Klasing, Ed., vol. 7276 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012, pp. 271–282.
 17. MONDERER, D. AND SHAPLEY, L. S. Potential Games. vol. 14 of *Games and Economic Behavior*. (1) 1996, pp. 124–143.
 18. NASH, J. Non-cooperative games. *Annals of Mathematics* 54, 2 (Sept. 1951), 286–295.
 19. NISAN, N., ROUGHGARDEN, T., TARDOS, E., AND VAZIRANI, V. V. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.
 20. PAPADIMITRIOU, C. H., AND STEIGLITZ, K. *Combinatorial optimization: algorithms and complexity*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1982.
 21. SCALZO, V. Pareto efficient Nash equilibria in discontinuous games. *Economics Letters* 107, 3 (2010), 364 – 365.