On (Coalitional) Exchange-Stable Matching

Jiehua Chen(✉), Adrian Chmurovic, Fabian Jogl, and Manuel Sorge(✉)

TU Wien, Vienna, Austria
jiehua.chen@tuwien.ac.at, manuel.sorge@ac.tuwien.ac.at

Abstract. We study *(coalitional) exchange stability*, which Alcalde [Economic Design, 1995] introduced as an alternative solution concept for matching markets involving property rights, such as assigning persons to two-bed rooms. Here, a matching of a given Stable Marriage or Stable Roommates instance is called *coalitional exchange-stable* if it does not admit any *exchange-blocking coalition*, that is, a subset S of agents in which everyone prefers the partner of some other agent in S. The matching is *exchange-stable* if it does not admit any *exchange-blocking pair*, that is, an exchange-blocking coalition of size two.

We investigate the computational and parameterized complexity of the Coalitional Exchange-Stable Marriage (resp. Coalitional Exchange Roommates) problem, which is to decide whether a Stable Marriage (resp. Stable Roommates) instance admits a coalitional exchange-stable matching. Our findings resolve an open question and confirm the conjecture of Cechlárová and Manlove [Discrete Applied Mathematics, 2005] that Coalitional Exchange-Stable Marriage is NP-hard even for complete preferences without ties. We also study bounded-length preference lists and a local-search variant of deciding whether a given matching can reach an exchange-stable one after at most k *swaps*, where a swap is defined as exchanging the partners of the two agents in an exchange-blocking pair.

1 Introduction

An instance in a matching market consists of a set of agents that each have preferences over other agents with whom they want to be matched with. The goal is to find a matching, i.e., a subset of disjoint pairs of agents, which is *fair*. A classical notion of fairness is *stability* [14], meaning that no two agents can form a *blocking pair*, i.e., they would prefer to be matched with each other rather than with the partner assigned by the matching. This means that a matching is fair if the agents cannot take local action to improve their outcome. If we assign property rights via the matching, however, then the notion of blocking pairs may not be actionable, as Alcalde [3] observed: For example, if the matching

JC was supported by the WWTF research project (VRG18-012). MS was supported by the Alexander von Humboldt Foundation.

I. Caragiannis and K. A. Hansen (Eds.): SAGT 2021, LNCS 12885, pp. 205–220, 2021.
https://doi.org/10.1007/978-3-030-85947-3_14

represents an assignment of persons to two-bed rooms, then two persons in a blocking pair may not be able to deviate from the assignment because they may not find a new room that they could share. Instead, we may consider the matching to be *fair* if no two agents form an *exchange-blocking pair*, i.e., they would prefer to have each other's partner rather than to have the partner given by the matching [3]. In other words, they would like to *exchange* their partners. Note that such an exchange would be straightforward in the room-assignment problem mentioned before. We refer to the work of Alcalde [3], Cechlárová [9], and Cechlárová and Manlove [10] for more discussion and examples of markets involving property rights.

If a matching does not admit an exchange-blocking pair, then the matching is *exchange-stable*. If we also want to exclude the possibility that several agents may collude to favorably exchange partners, then we arrive at *coalitional exchange-stability* [3]. In contrast to classical stability and exchange-stability for perfect matchings (i.e., everyone is matched), it is not hard to observe that coalitional exchange-stability implies *Pareto-optimality*, another fairness concept which asserts that no other matching can make at least one agent better-off without making some other agent worse-off (see also Abraham and Manlove [2]). Cechlárová and Manlove [10] showed that the problem of deciding whether an exchange-stable matching exists is NP-hard, even for the marriage case (where the agents are partitioned into two subsets of equal size such that each agent of either subset has preferences over the agents of the other subset) with complete preferences but without ties. They left open whether the NP-hardness transfers to coalitional exchange-stability, but observed NP-containment.

In this paper, we study the algorithmic complexity of problems revolving around (coalitional) exchange-stability. In particular, we establish a first NP-hardness result for deciding coalitional exchange-stability, confirming a conjecture of Cechlárová and Manlove [10]. The NP-hardness reduction is based on a novel *switch-gadget* wherein each preference list contains at most three agents. Utilizing this, we can carefully complete the preferences so as to prove the desired NP-hardness. We then investigate the impact of the maximum length d of a preference list. We find that NP-hardness for both exchange-stability and coalitional exchange-stability starts already when $d = 3$, while it is fairly easy to see that the problem becomes polynomial-time solvable for $d = 2$. For $d = 3$, we obtain a fixed-parameter algorithm for exchange-stability regarding a parameter which is related to the number of switch-gadgets.

Finally, we look at a problem variant, called Path to Exchange-Stable Marriage (P-ESM), for uncoordinated (or decentralized) matching markets. Starting from an initial matching, in each iteration the two agents in an exchange-blocking pair may swap their partners. An interesting question regarding the behavior of the agents in uncoordinated markets is whether such iterative swap actions can reach a stable state, i.e., exchange-stability, and how hard is it to decide. It is fairly straight-forward to verify that if the number k of swaps is bounded by a constant, then P-ESM is polynomial-time solvable since there are only polynomially many possible sequences of exchanges to be checked. From the parameterized complexity point of view, we obtain an XP algorithm for k, i.e., the exponent in the polynomial running time depends on k. We further show that the dependency of the exponent on k is unlikely to be removed by showing W[1]-hardness with respect to k.

Related Work. Alcalde [3] introduced (coalitional) exchange stability and discussed restricted preference domains where (coalitional) exchange stability is guaranteed to exist. Abizada [1] showed a weaker condition (on the preference domain) to guarantee the existence of exchange stability. Cechlárová and Manlove [10] proved that it is NP-complete to decide whether an exchange-stable matching exists, even for the marriage case with complete preferences without ties. Aziz and Goldwasser [4] introduced several relaxed notions of coalitional exchange-stability and discussed their relations.

The P-ESM problem is inspired by the Path-to-Stability via Divorces (PSD) problem, originally introduced by Knuth [16], see also Biró and Norman [5] for more background. Very recently, Chen [11] showed that PSD is NP-hard and W[1]-hard when parameterized by the number of divorces. P-ESM can also be considered as a local search problem and is a special case of the Local Search Exchange-Stable Seat Arrangement (Local-STA) problem, introduced by Bodlaender et al. [6]: Given a a set of agents, each having cardinal preferences (i.e., real values) over the other agents, an undirected graph G with the same number of vertices as agents, and an initial assignment (bijection) of the agents to the vertices in G, is it possible to swap two agents' assignments iteratively so as to reach an exchange-stable assignment? Herein an assignment is called *exchange-stable* if no two agents can each have a higher sum of cardinal preferences over the other's neighboring agents. P-ESM is a restricted variant of Local-STA, where G consists of disjoint edges and the agents have ordinal preferences. Bodlaender et al. [7] showed that Local-STA is W[1]-hard wrt. the number k of swaps. Their reduction relies on the fact that the given graph contains cliques and stars, and the preferences of the agents may contain ties. Our results for P-ESM that Local-STA is W[1]-hard even if the given graph consists of disjoint edges and the preferences do not have ties. Finally, we mention that Irving [15] and McDermid et al. [17] studied the complexity of computing stable matchings in the marriage setting with preference lists, requiring additionally that the matching should be man-exchange stable, i.e., no two men form an exchange-blocking pair, obtaining hardness and tractability results.

Organization. In Sect. 2, we introduce relevant concepts and notation, and define our central problems. In Sect. 3, we investigate the complexity of deciding (coalitional) exchange-stability, both when the preferences are complete and when the preferences length are bounded. In Sect. 4, we provide algorithms for profiles with preference length bounded by three. In Sect. 5, we turn to the local search variant of reaching exchange-stability. Section 6 concludes with open questions. Due to space constraints, results marked by $\star$ are deferred to [12].

2 Basic Definitions and Observations

For each natural number t, we denote the set $\{1, 2, \ldots, t\}$ by $[t]$.

Let $V = \{1, 2, \ldots, 2n\}$ be a set of $2n$ agents. Each agent $i \in V$ has a nonempty subset of agents $V_i \subseteq V$ which he finds *acceptable* as a partner and has a *strict preference list* $\succ_i$ on V_i (i.e., a linear order on V_i). The *length* of preference list $\succ_i$ is defined as the number of acceptable agents of i, i.e., $|V_i|$. Here, $x \succ_i y$ means that i *prefers* x to y.

We assume that the acceptability relation among the agents is *symmetric*, i.e., for each two agents x and y it holds that x is acceptable to y if and only if y is acceptable to x. For two agents x and y, we call x *most acceptable* to y if x is a maximal element in the preference list of y. For notational convenience, we write $X \succ_i Y$ to indicate that for each pair of agents $x \in X$ and $y \in Y$ it holds that $x \succ_i y$.

A *preference profile* $\mathcal{P}$ is a tuple $(V, (\succ_i)_{i \in V})$ consisting of an agent set V and a collection $(\succ_i)_{i \in V}$ of preference lists for all agents $i \in V$. For a graph G, by $V(G)$ and $E(G)$ we refer to its vertex set and edge set, respectively. Given a vertex $v \in V(G)$, by $N_G(v)$ and $d_G(v)$ we refer to the neighborhood and degree of v in G, respectively. To a preference profile $\mathcal{P}$ with agent set V we assign an *acceptability graph* $G(\mathcal{P})$ which has V as its vertex set and two agents are connected by an edge if they find each other acceptable. A preference profile $\mathcal{P}$ may have the following properties: Profile $\mathcal{P}$ is *bipartite*, if the agent set V can be partitioned into two agent sets U and W of size n each, such that each agent from one set has a preference list over a subset of the agents from the other set. Profile $\mathcal{P}$ has *complete* preferences if the underlying acceptability graph $G(\mathcal{P})$ is a complete graph or a complete bipartite graph on two disjoint sets of vertices of equal size; otherwise it has *incomplete* preferences. Profile $\mathcal{P}$ has *bounded length d* if each preference list in $\mathcal{P}$ has length at most d.

(Coalitional) Exchange-stable Matchings. A *matching M* for a given profile $\mathcal{P}$ is a subset of disjoint edges from the underlying acceptability graph $G(\mathcal{P})$. Given a matching M for $\mathcal{P}$, and two agents x and y, if it holds that $\{x, y\} \in M$, then we use $M(x)$ (resp. $M(y)$) to refer to y (resp. x), and we say that x and y are their respective assigned *partners* under M and that they are *matched* to each other; otherwise we say that $\{x, y\}$ is an *unmatched pair* under M. If an agent x is *not* assigned any partner by M, then we say that x is *unmatched by M* and we put $M(x) = x$. We assume that each agent x prefers to be matched than remaining unmatched. To formalize this, we will always say that x prefers all acceptable agents from V_x to himself x. A matching M is *perfect* if every agent is assigned a partner. It is *maximal* if for each unmatched pair $\{x, y\} \in E(G(\mathcal{P})) \setminus M$ it holds that x or y is matched under M. For two agents x, y, we say that *x envies y under M* if x prefers the partner of y, i.e., $M(y)$, to his partner $M(x)$. We omit the "under M" if it is clear from the context.

Matching M admits an *exchange-blocking coalition* (in short *ebc*) if there exists a sequence $\rho = (x_0, x_1, \ldots, x_{r-1})$ of r agents, $r \geq 2$, such that each agent x_i envies her successor x_{i+1} in ρ (index $i + 1$ taken modulo r). The *size* of an ebc is defined as the number of agents in the sequence. An *exchange-blocking pair* (in short *ebp*) is an ebc of size two. A matching M of $\mathcal{P}$ is *exchange-stable* (resp. *coalitional exchange-stable*) if it does not admit any ebp (resp. ebc). Note that a coalitional exchange-stable matching is exchange-stable. For an illustration, let us consider the following example.

Example 1. The following bipartite preference profile $\mathcal{P}$ with agent sets $U = \{x, y, z\}$ and $W = \{a, b, c\}$ admits 2 (coalitional) exchange-stable matchings M_1 and M_2 with $M_1 = \{\{x, c\}, \{y, b\}, \{z, a\}\}$ (marked in red boxes) and $M_2 = \{\{x, b\}, \{y, c\}, \{z, a\}\}$ (marked in blue boxes).

$$x\colon a \succ \boxed{b} \succ \boxed{c},\quad a\colon y \succ x \succ \boxed{\boxed{z}},$$
$$y\colon \boxed{b} \succ a \succ \boxed{c},\quad b\colon \boxed{x} \succ \boxed{y} \succ z,$$
$$z\colon \boxed{\boxed{a}} \succ c \succ b,\quad c\colon \boxed{x} \succ \boxed{y} \succ z.$$

Matching M_3 with $M_3 = \{\{x, c\}, \{y, a\}, \{z, b\}\}$ is not exchange-stable and hence not coalitional exchange-stable since for instance (y, z) is an exchange-blocking pair of M_3.

As already observed by Cechlárová and Manlove [10], exchange-stable (or coalitional exchange-stable) matchings may not exist, even for bipartite profiles with complete preferences. Every coalitional exchange-stable matching is maximal ($\star$).

We are interested in the computational complexity of deciding whether a given profile admits a coalitional exchange-stable matching.

Coalitional Exchange-Stable Roommates (CESR)
Input: A preference profile $\mathcal{P}$.
Question: Does $\mathcal{P}$ admit a coalitional exchange-stable matching?

The bipartite restriction of CESR, called Coalitional Exchange-Stable Marriage (CESM), has as input a *bipartite* preference profile. Exchange-Stable Roommates (ESR) and Exchange-Stable Marriage (ESM) are defined analogously.

We are also interested in the case when the preferences have bounded length. In this case, not every coalitional exchange-stable (or exchange-stable) matching is perfect. In keeping with the literature [9, 10], we focus on the perfect case.

d-Coalitional Exchange-Stable Roommates (d-CESR)
Input: A preference profile $\mathcal{P}$ with preferences of bounded length d.
Question: Does $\mathcal{P}$ admit a coalitional exchange-stable and *perfect* matching?

We analogously define the bipartite restriction d-Coalitional Exchange-Stable Marriage *(d-CESM)*, and the exchange-stable variants d-Exchange-Stable Roommates (d-ESR) and d-Exchange-Stable Marriage (d-ESM). Note that the above problems are contained in NP [10].

Finally, we investigate a local search variant regarding exchange-stability. To this end, given two matchings M and N of the same profile $\mathcal{P}$, we say that M is *one-swap reachable* from N if there exists an exchange-blocking pair (x, y) of N such that $M = (N \setminus \{\{x, N(x)\}, \{y, N(y)\}\}) \cup \{\{x, y\}, \{N(x), N(y)\}\}$. Accordingly, we say that M is *k-swaps reachable* from N if there exists a sequence $(M_0, M_1, \cdots, M_{k'})$ of k' *matchings* of profile $\mathcal{P}$ such that (a) $k' \leq k$, $M_0 = N$, $M_{k'} = M$, and (b) for each $i \in [k']$, M_i is one-swap reachable from M_{i-1}.

The local search problem variant is defined as follows:

Path to Exchange-Stable Marriage (P-ESM)
Input: A bip. preference profile $\mathcal{P}$, a matching M_0 of $\mathcal{P}$, and an integer k.
Question: Is there an exchange-stable matching M for $\mathcal{P}$ that is k-swap reachable from M_0?

3 Deciding (Coalitional) Exchange-Stability is NP-complete

Cechlárová and Manlove [10] proved NP-completeness for ESM. It is, however, not immediate how to adapt Cechlárová and Manlove's proof to show hardness for coalitional exchange-stability since their constructed exchange-stable matching is not always coalitional exchange-stable. To obtain a hardness reduction for CESM, we first study the case when the preferences have length bounded by three, and show that 3-CESM is NP-hard, even for strict preferences. We reduce from an NP-complete ($\star$) variant of 3SAT, called (2,2)-3SAT: Is there a satisfying truth assignment for a given Boolean formula $\phi(X)$ with variable set X in 3CNF (i.e., a set of clauses each containing at most 3 literals) where no clause contains both the positive and the negated literal of the same variable, and each literal appears *exactly* two times?

A crucial ingredient for our reduction is the following *switch-gadget* which enforces that each exchange-stable matching results in a valid truth assignment. The gadget and its properties are summarized in the following lemma.

Lemma 1 ($\star$). *Let $\mathcal{P}$ be a bipartite preference profile on agent sets U and W. Let $A = \{a^z \mid z \in \{0, 1, \ldots, 6\}\}$ and $B = \{b^z \mid z \in \{0, 1, \ldots, 6\}\}$ be two disjoint sets of agents, and let $Q = \{\alpha, \beta, \gamma, \delta\}$ be four further distinct agents with $A \cup \{\alpha, \gamma\} \subseteq U$ and $B \cup \{\beta, \delta\} \subseteq W$. The preferences of the agents from A and B are as follows; the preferences of the other agents are arbitrary but fixed.*

- a^0: $b^1 \succ \beta$,
- a^1: $b^0 \succ b^2 \succ b^1$,
- a^2: $b^3 \succ b^1 \succ b^2$,
- a^3: $b^2 \succ b^3 \succ b^4$,
- a^4: $b^4 \succ b^3 \succ b^5$,
- a^5: $b^6 \succ b^4 \succ b^5$,
- a^6: $b^5 \succ \delta$,

- b^0: $a^1 \succ \alpha$,
- b^1: $a^0 \succ a^2 \succ a^1$,
- b^2: $a^2 \succ a^3 \succ a^1$,
- b^3: $a^4 \succ a^3 \succ a^2$,
- b^4: $a^3 \succ a^5 \succ a^4$,
- b^5: $a^6 \succ a^4 \succ a^5$,
- b^6: $a^5 \succ \gamma$.

Every perfect *matching M of $\mathcal{P}$ satisfies the following, where*

$$N^1 := \{\{\alpha, b^0\}, \{a^6, \delta\}\} \cup \{\{a^{z-1}, b^z\} \mid z \in [6]\},$$
$$N^2 := \{\{a^0, \beta\}, \{\gamma, b^6\}\} \cup \{\{a^z, b^{z-1}\} \mid z \in [6]\}, \text{ and}$$
$$N^D := \{\{\alpha, b^0\}, \{a^0, \beta\}, \{a^6, \delta\}, \{\gamma, b^6\}, \{a^1, b^2\}, \{a^2, b^1\}, \{a^3, b^3\}, \{a^4, b^5\}, \{a^5, b^4\}\}.$$

(1) If M is exchange-stable, then either (i) $N^1 \subseteq M$, or (ii) $N^2 \subseteq M$, or (iii) $N^D \subseteq M$.
(2) If $N^1 \subseteq M$, then every ebc of M which involves an agent from A (resp. B) also involves α (resp. δ).
(3) If $N^2 \subseteq M$, then every ebc of M which involves an agent from A (resp. B) also involves γ (resp. β).
(4) If $N^D \subseteq M$, then every ebc of M which involves an agent from A (resp. B) also involves an agent from $\{\alpha, \gamma\}$ (resp. $\{\beta, \delta\}$).

Using Lemma 1, we can show NP-hardness for bounded preference length.

Theorem 1. 3-CESM*,* 3-ESM*,* 3-CESR*, and* 3-ESR *are NP-complete.*

Proof. As already mentioned [10], by checking for cycles in the envy graph all discussed problems are in NP ($\star$). For the NP-hardness, it suffices to show that 3-CESM and 3-ESM are NP-hard. We use the same reduction from (2,2)-3SAT for both. Let (X, C) be an instance of (2,2)-3SAT where $X = \{x_1, x_2, \cdots, x_{\hat{n}}\}$ is the set of variables and $\phi = \{C_1, C_2, \cdots, C_{\hat{m}}\}$ the set of clauses.

We construct a bipartite preference profile on two disjoint agent sets U and W. The set U (resp. W) will be partitioned into three different agent-groups: the variable-agents, the switch-agents, and the clause-agents. The general idea is to use the variable-agents and the clause-agents to determine a truth assignment and satisfying literals, respectively. Then, we use the switch-agents from Lemma 1 to make sure that the selected truth assignment is consistent with the selected satisfying literals. For each literal $\mathsf{lit}_i \in X \cup \overline{X}$ that appears in two different clauses C_j and C_k with $j < k$, we use $\mathsf{o}_1(\mathsf{lit}_i)$ and $\mathsf{o}_2(\mathsf{lit}_i)$ to refer to the indices j and k; recall that in ϕ each literal appears exactly two times.

The Variable-agents. For each variable $x_i \in X$, introduce 6 *variable-agents* v_i, w_i, x_i, $\overline{x}_i$, y_i, $\overline{y}_i$. Add $v_i, x_i, \overline{x}_i$ to U, and $w_i, y_i, \overline{y}_i$ to W. For each literal $\mathsf{lit}_i \in X \cup \overline{X}$ let $y(\mathsf{lit}_i)$ denote the corresponding Y-variable-agent, that is, $y(x_i) = y_i$ and $y(\overline{x}_i) = \overline{y}_i$. Define $\overline{X} := \{\overline{x}_i \mid i \in [\hat{n}]\}$, and $\overline{Y} := \{\overline{y}_i \mid i \in [\hat{n}]\}$.

The Clause-agents. For each clause $C_j \in C$, introduce two *clause-agents* c_j, d_j. Further, for each literal $\mathsf{lit}_i \in C_j$ with $\mathsf{lit} \in \{x, \overline{x}\}$, introduce two more *clause-agents* e^i_j, f^i_j. Add c_j, f^i_j to U, and d_j, e^i_j to W. For each clause $C_j \in \phi$, define $E_j := \{e^i_j \mid \mathsf{lit}_i \in C_j\}$, and $F_j := \{f^i_j \mid \mathsf{lit}_i \in C_j\}$. Moreover, define $E := \bigcup_{C_j \in \phi} E_j$ and $F := \bigcup_{C_j \in \phi} F_j$.

The Switch-agents. For each clause $C_j \in C$, and each literal $\mathsf{lit}_i \in C_j$ introduce fourteen *switch-agents* $a^z_{i,j}, b^z_{i,j}$, $z \in \{0, 1, \cdots, 6\}$. Define $A_{i,j} = \{a^z_{i,j} \mid z \in \{0, 1, \ldots, 6\}\}$ and $B_{i,j} = \{b^z_{i,j} \mid z \in \{0, 1, \ldots, 6\}\}$. Add $A_{i,j}$ to U and $B_{i,j}$ to W.

In total, we have the following agent sets:

$$U := \{v_i \mid i \in [\hat{n}]\} \cup X \cup \overline{X} \cup \{c_j \mid j \in [\hat{m}]\} \cup F \cup \bigcup_{C_j \in \phi \wedge \mathsf{lit}_i \in C_j} A_{i,j}, \text{ and}$$
$$W := \{w_i \mid i \in [\hat{n}]\} \cup Y \cup \overline{Y} \cup \{d_j \mid j \in [\hat{m}]\} \cup E \cup \bigcup_{C_j \in \phi \wedge \mathsf{lit}_i \in C_j} B_{i,j}.$$

The Preference Lists. The preference lists of the agents are shown in Fig. 1. Herein, the preferences of the switch-agents of each occurrence of the literal correspond to those given in Lemma 1. Note that all preferences are specified except those of $\alpha_{i,j}$ and $\delta_{i,j}$, which we do now. Defining them in an appropriate way will connect the two groups of switch-agents that correspond to the same literal as well as literals to clauses. For each literal $\mathsf{lit}_i \in X \cup \overline{X}$, recall that $\mathsf{o}_1(\mathsf{lit}_i)$ and $\mathsf{o}_2(\mathsf{lit}_i)$ are the indices of the clauses which contain lit_i with $\mathsf{o}_1(\mathsf{lit}_i) < \mathsf{o}_2(\mathsf{lit}_i)$. Let

$$\alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)} := \mathsf{lit}_i,\ \delta_{i,\mathsf{o}_1(\mathsf{lit}_i)} := b^0_{i,\mathsf{o}_2(\mathsf{lit}_i)}, \alpha_{i,\mathsf{o}_2(\mathsf{lit}_i)} := a^6_{i,\mathsf{o}_1(\mathsf{lit}_i)},\ \delta_{i,\mathsf{o}_2(\mathsf{lit}_i)} := y(\mathsf{lit}_i). \quad (1)$$

$\forall i \in [\hat{n}]$:

- $v_i : y_i \succ \overline{y}_i$, ■ $w_i : x_i \succ \overline{x}_i$,
- $x_i : w_i \succ b^0_{i,\mathsf{o}_1(x_i)}$, ■ $y_i : v_i \succ a^6_{i,\mathsf{o}_2(x_i)}$,
- $\overline{x}_i : w_i \succ b^0_{i,\mathsf{o}_1(\overline{x}_i)}$, ■ $\overline{y}_i : v_i \succ a^0_{i,\mathsf{o}_2(\overline{x}_i)}$,

$\forall j \in [\hat{m}]$:

- $c_j : [E_j]$, ■ $d_j : [F_j]$,

$\forall i \in [\hat{n}], \forall j \in [\hat{m}]$ with $\mathsf{lit}_i \in C_j$:

- $f^i_j : d_j \succ b^6_{i,j}$ ■ $e^i_j : c_j \succ a^0_{i,j}$
- $a^0_{i,j} : b^1_{i,j} \succ e^i_j$, ■ $b^0_{i,j} : a^1_{i,j} \succ \alpha_{i,j}$,
- $a^1_{i,j} : b^0_{i,j} \succ b^2_{i,j} \succ b^1_{i,j}$, ■ $b^1_{i,j} : a^0_{i,j} \succ a^2_{i,j} \succ a^1_{i,j}$,
- $a^2_{i,j} : b^3_{i,j} \succ b^1_{i,j} \succ b^2_{i,j}$, ■ $b^2_{i,j} : a^2_{i,j} \succ a^3_{i,j} \succ a^1_{i,j}$,
- $a^3_{i,j} : b^2_{i,j} \succ b^3_{i,j} \succ b^4_{i,j}$, ■ $b^3_{i,j} : a^4_{i,j} \succ a^3_{i,j} \succ a^2_{i,j}$,
- $a^4_{i,j} : b^4_{i,j} \succ b^3_{i,j} \succ b^5_{i,j}$, ■ $b^4_{i,j} : a^3_{i,j} \succ a^5_{i,j} \succ a^4_{i,j}$,
- $a^5_{i,j} : b^6_{i,j} \succ b^4_{i,j} \succ b^5_{i,j}$, ■ $b^5_{i,j} : a^6_{i,j} \succ a^4_{i,j} \succ a^5_{i,j}$,
- $a^6_{i,j} : b^5_{i,j} \succ \delta_{i,j}$, ■ $b^6_{i,j} : a^5_{i,j} \succ f^i_j$.

Fig. 1. The preferences constructed in the proof for Theorem 1. Recall that for each literal $\mathsf{lit}_i \in X \cup \overline{X}$, expressions $\mathsf{o}_1(\mathsf{lit}_i)$ and $\mathsf{o}_2(\mathsf{lit}_i)$ denote the two indices $j < j'$ of the clauses that contain lit_i. For each clause $C_j \in \phi$, the expression $[E_j]$ (resp. $[F_j]$) denotes an arbitrary but fixed order of the agents in E_j (resp. F_j).

This completes the construction of the instance for 3-CESM, which can clearly be done in polynomial-time. Let $\mathcal{P}$ denote the constructed instance with $\mathcal{P} = (U \uplus W, (\succ_x)_{x \in U \cup W})$. It is straight-forward to verify that $\mathcal{P}$ is bipartite and contains no ties and each preference list $\succ_x$ has length bounded by three. Before we give the correctness proof, for each literal $\mathsf{lit}_i \in X \cup \overline{X}$ and each clause C_j with $\mathsf{lit}_i \in C_j$ we define the following three matchings:

$$\begin{aligned}
N^1_{i,j} &\coloneqq \{\{\alpha_{i,j}, b^0_{i,j}\}, \{a^6_{i,j}, \delta_{i,j}\}\} \cup \{\{a^{z-1}_{i,j}, b^z_{i,j}\} \mid z \in [6]\},\\
N^2_{i,j} &\coloneqq \{\{a^0_{i,j}, e^i_j\}, \{b^6_{i,j}, f^i_j\}\} \cup \{\{a^z_{i,j}, b^{z-1}_{i,j}\} \mid z \in [6]\}, \text{ and}\\
N^D_{i,j} &\coloneqq \{\{\alpha_{i,j}, b^0_{i,j}\}, \{a^0_{i,j}, e^i_j\}, \{a^6_{i,j}, \delta_{i,j}\}, \{f^i_j, b^6_{i,j}\},\\
&\qquad \{a^1_{i,j}, b^2_{i,j}\}, \{a^2_{i,j}, b^1_{i,j}\}, \{a^3_{i,j}, b^3_{i,j}\}, \{a^4_{i,j}, b^5_{i,j}\}, \{a^5_{i,j}, b^4_{i,j}\}\}.
\end{aligned} \tag{2}$$

Now we show the correctness, i.e., ϕ admits a satisfying assignment if and only if $\mathcal{P}$ admits a perfect and coalitional exchange-stable (resp. exchange-stable) matching. For the "only if" direction, assume that $\sigma\colon X \to \{\mathsf{true}, \mathsf{false}\}$ is a satisfying assignment for ϕ. Then, we define a perfect matching M as follows.

- For each variable $x_i \in X$, let $M(\overline{x}_i) \coloneqq w_i$ and $M(v_i) \coloneqq \overline{y}_i$ if $\sigma(x_i) = \mathsf{true}$; otherwise, let $M(x_i) \coloneqq w_i$ and $M(v_i) \coloneqq y_i$.
- For each clause $C_j \in \phi$, fix an arbitrary literal whose truth value satisfies C_j and denote the index of this literal as $\mathsf{s}(j)$. Then, let $M(c_j) \coloneqq e^{\mathsf{s}(j)}_j$ and $M(f^{\mathsf{s}(j)}_j) \coloneqq d_j$.
- Further, for each literal $\mathsf{lit}_i \in X \cup \overline{X}$ and each clause C_j with $\mathsf{lit}_i \in C_j$, do:
  (a) If $\mathsf{s}(j) = i$, then add to M all pairs from $N^1_{i,j}$.
  (b) If $\mathsf{s}(j) \neq i$ and lit_i is set true under σ (i.e., $\sigma(x_i) = \mathsf{true}$ iff. $\mathsf{lit}_i = x_i$), then add to M all pairs from $N^D_{i,j}$.
  (c) If $\mathsf{s}(j) \neq i$ and lit_i is set to false under σ (i.e., $\sigma(x_i) = \mathsf{true}$ iff. $\mathsf{lit}_i = \overline{x}_i$), then add to M all pairs from $N^2_{i,j}$.

One can verify that M is perfect. Hence, it remains to show that M is coalitional exchange-stable. Note that this would also imply that M is exchange-stable.

Suppose, for the sake of contradiction, that M admits an ebc ρ. First, observe that for each variable-agent $z \in X \cup \overline{X} \cup Y \cup \overline{Y}$ it holds that $M(z)$ either is matched with his most-preferred partner (i.e., either v_i or w_i) or only envies someone who is matched with his most-preferred partner. Hence, no agent from $X \cup \overline{X} \cup Y \cup \overline{Y}$ is involved in ρ. Analogously, no agent from $E \cup F$ is involved in ρ. Next, we claim the following.

Claim 1 ($\star$). *For each literal $\mathsf{lit}_i \in X \cup \overline{X}$ and each clause C_j with $\mathsf{lit}_i \in C_j$, it holds that neither $\alpha_{i,j}$ nor $\delta_{i,j}$ is involved in ρ.*

Using the above observations and claim, we continue with the proof. We successively prove that no agent is involved in ρ, starting with the agents in U.

- If v_i is involved in ρ for some $i \in [\hat{n}]$, then he only envies someone who is matched with y_i. By the preferences of y_i, this means that $M(y_i) = a^6_{i,\mathsf{o}_2(x_i)}$

and v_i envies $a^6_{i,\mathsf{o}_2(x_i)}$. Hence, $a^6_{i,\mathsf{o}_2(x_i)}$ is also involved in ρ. Moreover, since $M(a^6_{i,\mathsf{o}_2(x_i)}) = y_i$, we have $N^1_{i,\mathsf{o}_2(x_i)} \subseteq M$ or $N^D_{i,\mathsf{o}_2(x_i)} \subseteq M$. By Lemma 1(2) and Lemma 1(4) (setting $\alpha = \alpha_{i,\mathsf{o}_2(x_i)}$, $\beta = e^i_{\mathsf{o}_2(x_i)}$, $\gamma = f^i_{\mathsf{o}_2(x_i)}$, and $\delta = \delta_{i,\mathsf{o}_2(x_i)}$), ρ involves an agent from $\{\alpha_{i,\mathsf{o}_2(x_i)}, f^i_{\mathsf{o}_2(x_i)}\}$. Since no agent from F is involved in ρ, it follows that ρ involves $\alpha_{i,\mathsf{o}_2(x_i)}$, a contradiction to Claim 1.

– Analogously, if $c_j \in \rho$ for some $j \in [\hat{m}]$, then this means that E_j contains two agents e^i_j and e^t_j such that $M(c_j) = e^t_j$ but c_j prefers e^i_j to e^t_j, and $M(e^i_j) \in \rho$. Since M is perfect and c_j is not available, it follows that $M(e^i_j) = a^0_{i,j}$, implying that $a^0_{i,j} \in \rho$. Moreover, by the definition of M we have that $N^2_{i,j} \subseteq M$ or $N^D_{i,j} \subseteq M$. By Lemmas 1(3)–(4) (setting $\alpha = \alpha_{i,j}$, $\beta = e^i_j$, $\gamma = f^i_j$, and $\delta = \delta_{i,j}$), ρ involves an agent from $\{\alpha_{i,j}, f^i_j\}$, a contradiction since no agent from F_j is involved in ρ and by Claim 1 $\alpha_{i,j}$ is not in ρ.
– Analogously, we can obtain a contradiction if w_i with $i \in [\hat{n}]$ is in ρ: By the definition of M, if $w_i \in \rho$, then $M(x_i) = b^0_{i,\mathsf{o}_1(x_i)}$ and w_i envies $b^0_{i,\mathsf{o}_1(x_i)}$. Hence, $b^0_{i,\mathsf{o}_1(x_i)}$ is also involved in ρ. Moreover, since $M(b^0_{i,\mathsf{o}_1(x_i)}) = x_i$, it follows that $N^1_{i,\mathsf{o}_1(x_i)} \subseteq M$ or $N^D_{i,\mathsf{o}_1(x_i)} \subseteq M$. By Lemmas 1(2) and (4) (setting $\alpha = \alpha_{i,\mathsf{o}_1(x_i)}$, $\beta = e^i_{\mathsf{o}_1(x_i)}$, $\gamma = f^i_{\mathsf{o}_1(x_i)}$, and $\delta = \delta_{i,\mathsf{o}_1(x_i)}$), ρ involves an agent from $\{e^i_{\mathsf{o}_1(x_i)}, \delta_{i,\mathsf{o}_1(x_i)}\}$. Since no agent from E is involved in ρ, it follows that ρ involves $\delta_{i,\mathsf{o}_1(x_i)}$, a contradiction to Claim 1.
– Again, analogously, if $d_j \in \rho$ for some $j \in [\hat{m}]$, then we obtain that $\delta_{i,j}$ is involved in ρ, which is a contradiction to Claim 1.
– Finally, if ρ involves an agent from $A_{i,j}$ (resp. $B_{i,j}$), then by Lemma 1(2) and (4) (setting $\alpha = \alpha_{i,j}$, $\beta = e^i_j$, $\gamma = f^i_j$, and $\delta = \delta_{i,j}$), it follows that ρ involves an agent from $\{\alpha_{i,j}, f^i_j\}$ (resp. $\{\beta_{i,j}, e^i_j\}$), a contradiction to our observation and to Claim 1.

Summarizing, M is coalitional exchange-stable and exchange-stable.

For the "if" direction, assume that M is a perfect and exchange-stable matching for $\mathcal{P}$. We show that there is a satisfying assignment for ϕ. Note that this then also implies that, if M is perfect and coalitional exchange-stable, then there is a satisfying assignment for ϕ.

We claim that the selection of the partner of w_i defines a satisfying truth assignment for ϕ. More specifically, define a truth assignment $\sigma\colon X \to \{\mathsf{true}, \mathsf{false}\}$ with $\sigma(x_i) = \mathsf{true}$ if $M(w_i) = \overline{x}_i$, and $\sigma(x_i) = \mathsf{false}$ otherwise. We claim that σ satisfies ϕ. To this end, consider an arbitrary clause C_j and the corresponding clause-agent. Since M is perfect, it follows that $M(c_j) = e^i_j$ for some $\mathsf{lit}_i \in C_j$. Since e^i_j is not available, it also follows that $M(a^0_{i,j}) = b^1_{i,j}$. By Lemma 1(1) (setting $\alpha = \alpha_{i,j}$, $\beta = e^i_j$, $\gamma = f^i_j$, and $\delta = \delta_{i,j}$), it follows that $N^1_{i,j} \subseteq M$. In particular, $M(\alpha_{i,j}) = b^0_{i,j}$ so that $\alpha_{i,j}$ is not available to other agents anymore.

Now, if we can show that $\mathsf{lit}_i = \alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ is matched to $b^0_{i,\mathsf{o}_1(\mathsf{lit}_i)}$, then since M is perfect, we have $M(w_i) = \overline{x}_i$ if $\mathsf{lit}_i = x_i$, and $M(w_i) = \overline{x}_i$ otherwise By definition, we have $\sigma(x_i) = \mathsf{true}$ if $\mathsf{lit}_i = x_i$ and $\sigma(x_i) = \mathsf{false}$ otherwise. Thus,

C_j is satisfied under σ, implying that σ is a satisfying assignment. It remains to show that lit_i is matched to $b^0_{i,\mathsf{o}_1(\mathsf{lit}_i)}$. We distinguish between two cases;

- If $j = \mathsf{o}_1(\mathsf{lit}_i)$, then $\mathsf{lit}_i = \alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ is matched to $b^0_{i,\mathsf{o}_1(\mathsf{lit}_i)}$, as required.
- If $j = \mathsf{o}_2(\mathsf{lit}_i)$, then by definition, it holds that $\alpha_{i,j} = a^6_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ and $\delta_{i,\mathsf{o}_1(\mathsf{lit}_i)} = b^0_{i,j}$. In other words, $M(a^6_{i,\mathsf{o}_1(\mathsf{lit}_i)}) = \delta_{i,\mathsf{o}_1(\mathsf{lit}_1)}$. By Lemma 1(1) (setting $\alpha = \alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$, $\beta = e^i_{\mathsf{o}_1(\mathsf{lit}_i)}$, $\gamma = f^i_{\mathsf{o}_1(\mathsf{lit}_i)}$, and $\delta = \delta_{i,\mathsf{o}_1(\mathsf{lit}_i)}$), it follows that $N^1_{i,j} \subseteq M$ or $N^D_{i,j} \subseteq M$. In both cases, it follows that $\alpha_{i,\mathsf{o}_1(i)}$ is matched to $b^0_{i,\mathsf{o}_1(i)}$. □

Next, we show how to complete the preferences of the agents constructed in the proof of Theorem 1 to show hardness for complete and strict preferences.

Theorem 2. CESM *and* CESR *are NP-complete even for complete and strict preferences.*

Proof. We only show NP-hardness for CESM as the hardness for CESR will follow immediately by using the same approach as [10, Lemma 3.1]. To show hardness for CESM, we adapt the proof of Theorem 1. In that proof, given (2,2)-3SAT instance (X, ϕ) with $X = \{x_1, x_2, \cdots, x_{\hat{n}}\}$ and $\phi = \{C_1, C_2, \cdots, C_{\hat{m}}\}$, we constructed two disjoint agent sets U and W with $U \coloneqq \{v_i \mid i \in [\hat{n}]\} \cup X \cup \overline{X} \cup \{c_j \mid j \in [\hat{m}]\} \cup F \cup \bigcup_{C_j \in \phi \wedge \mathsf{lit}_i \in C_j} A_{i,j}$ and $W \coloneqq \{w_i \mid i \in [\hat{n}]\} \cup Y \cup \overline{Y} \cup \{d_j \mid j \in [\hat{m}]\} \cup E \cup \bigcup_{C_j \in \phi \wedge \mathsf{lit}_i \in C_j} B_{i,j}$. For each agent $z \in U \cup W$ let L_z denote the preference list of z constructed in the proof. The basic idea is to extend the preference list L_z by appending to it the remaining agents appropriately.

We introduce some more notations. Let $\rhd_U$ and $\rhd_W$ denote two arbitrary but fixed linear orders of the agents in U and W, respectively. Now, for each subset of agents $S \subseteq U$ (resp. $S \subseteq W$), let $[S]_\rhd$ denote the fixed order of the agents in S induced by $\rhd_U$ (resp. $\rhd_W$), and let $S \setminus \mathsf{L}_z$ denote the subset $\{t \in S \mid t \notin \mathsf{L}_z\}$, where $z \in W$ (resp. $z \in U$). Finally, for each agent $z \in U$ (resp. $z \in W$), let R_z denote the subset of agents which do not appear in L_z or in $Y \cup \overline{Y} \cup E$ (resp. $X \cup \overline{X} \cup F$). That is, $\mathsf{R}_z \coloneqq \big(W \setminus (Y \cup \overline{Y} \cup F)\big) \setminus \mathsf{L}_z$ (resp. $\mathsf{R}_z \coloneqq \big(U \setminus (X \cup \overline{Y} \cup F)\big) \setminus \mathsf{L}_z$).
Now, we define the preferences of the agents as follows.

$$\forall z \in U,\ z\colon \mathsf{L}_z \succ [Y \cup \overline{Y} \cup E \setminus \mathsf{L}_z]_\rhd \succ [\mathsf{R}_z]_\rhd, \quad \text{and}$$
$$\forall z \in W,\ z\colon \mathsf{L}_z \succ [X \cup \overline{X} \cup F \setminus \mathsf{L}_z]_\rhd \succ [\mathsf{R}_z]_\rhd.$$

Let $\mathcal{P}'$ denote the newly constructed preference profile. Clearly, the constructed preferences are complete and strict. Before we show the correctness, we claim the following for each coalitional exchange-stable matching of $\mathcal{P}'$.

Claim 2 **(⋆).** *If M is a coalitional exchange-stable matching for $\mathcal{P}'$, then*

(i) for each agent $z \in U \cup W$ it holds that $M(z) \notin \mathsf{R}_z$, and
(ii) for each agent $z \in U \cup W \setminus (X \cup \overline{X} \cup F \cup Y \cup \overline{Y} \cup E)$ it holds that $M(z) \in \mathsf{L}_z$.

Now we are ready to show the correctness, i.e., ϕ admits a satisfying assignment if and only if $\mathcal{P}'$ admits a coalitional exchange-stable matching.

For the "only if" direction, assume that ϕ admits a satisfying assignment, say $\sigma\colon X \to \{\mathsf{true}, \mathsf{false}\}$. We claim that the coalitional exchange-stable matching M for $\mathcal{P}$ that we defined in the "only if" direction of the proof for Theorem 1 is a coalitional exchange-stable matching for $\mathcal{P}'$. Clearly, M is a perfect matching for $\mathcal{P}'$ since $G(\mathcal{P}')$ is a supergraph of $G(\mathcal{P})$. Since each agent $z \in U \cup W$ has $M(z) \in \mathsf{L}_z$, for every two agents $z, z' \in U$ (resp. W), it holds that z envies z' only if $M(z') \in \mathsf{L}_z$. In other words, if M would admit an ebc $\rho = (z_0, z_1, \cdots, z_{r-1})$ $(r \geq 2)$ for $\mathcal{P}'$, then for each $i \in \{0, 1, \ldots, r-1\}$ it must hold that $M(z_i) \in \mathsf{L}_{z-1}$ ($z-1$ taken modulo r). But then, ρ is also an ebc for $\mathcal{P}$, a contradiction to our "only if" part of the proof for Theorem 1.

For the "if" direction, let M be a coalitional exchange-stable matching for $\mathcal{P}'$. Note that in the "if" part of the proof of Theorem 1 we heavily utilize the properties given in Lemma 1(1). Now, to construct a satisfying assignment for ϕ from M, we will prove that the lemma also holds for profile $\mathcal{P}'$. To this end, for each literal $\mathsf{lit}_i \in X \cup \overline{X}$ and each clause C_j with $\mathsf{lit}_i \in C_j$, recall the three matchings $N^1_{i,j}$, $N^2_{i,j}$, $N^D_{i,j}$ and the agents $\alpha_{i,j}$ and $\delta_{i,j}$ that we have defined in Eqs. (2) and (1) .

Claim 3 ($\star$). *Matching M satisfies for each literal $\mathsf{lit}_i \in X \cup \overline{X}$ and each clause $C_j \in \phi$ with $\mathsf{lit}_i \in C_j$, either (i) $N^1_{i,j} \subseteq M$, or (ii) $N^2_{i,j} \subseteq M$, or (iii) $N^D_{i,j} \subseteq M$.*

Now we show that the function $\sigma\colon X \to \{\mathsf{true}, \mathsf{false}\}$ with $\sigma(x_i) = \mathsf{true}$ if $M(w_i) = \overline{x}_i$, and $\sigma(x_i) = \mathsf{false}$ otherwise is a satisfying truth assignment for ϕ. Clearly, ϕ is a valid truth assignment since by Claim 2(ii) every variable agent w_i is matched to either x_i or $\overline{x}_i$. We claim that σ satisfies ϕ. Consider an arbitrary clause C_j and the corresponding clause-agent c_j. By Claim 2(ii), we know that $M(c_j) = e^i_j$ for some $\mathsf{lit}_i \in C_j$. Since e^i_j is not available, by Claim 2(ii), it also follows that $M(a^0_{i,j}) = b^1_{i,j}$. By Claim 3, it follows that $N^1_{i,j} \subseteq M$. In particular, $M(\alpha_{i,j}) = b^0_{i,j}$ so that $\alpha_{i,j}$ is not available to other agents anymore.

We aim to show that $\alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ is matched to $b^0_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ by M, which implies that lit_i is not available to w_i since $\alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)} = \mathsf{lit}_i$ by the definition of $\alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$. We distinguish two cases: If $j = \mathsf{o}_1(\mathsf{lit}_i)$, then by the definition of $\alpha_{i,j}$, it follows that $\alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ is matched to $b^0_{i,\mathsf{o}_1(\mathsf{lit}_i)}$. If $j = \mathsf{o}_2(\mathsf{lit}_i)$, then by the definition of $\alpha_{i,j}$, we have $\alpha_{i,j} = a^6_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ and by the definition of $\delta_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ we have $\delta_{i,\mathsf{o}_1(\mathsf{lit}_i)} = b^0_{i,\mathsf{o}_2(\mathsf{lit}_i)} = b^0_{i,j}$. In particular, since $M(\alpha_{i,j}) = b^0_{i,j}$ we have $M(a^6_{i,\mathsf{o}_1(\mathsf{lit}_i)}) = \delta_{i,\mathsf{o}_1(\mathsf{lit}_1)}$. By Claim 3, it follows that $N^1_{i,\mathsf{o}_1(\mathsf{lit}_i)} \subseteq M$ or $N^D_{i,\mathsf{o}_1(\mathsf{lit}_i)} \subseteq M$. In both cases, it follows that $\alpha_{i,\mathsf{o}_1(\mathsf{lit}_i)}$ is matched to $b^0_{i,\mathsf{o}_1(\mathsf{lit}_i)}$. We have just shown that lit_i is *not* available to w_i. Hence, by Claim 2(ii), $M(w_i) = \overline{x}_i$ if $\mathsf{lit}_i = x_i$, and $M(w_i) = \overline{x}_i$ otherwise. By definition, we have that $\sigma(x_i) = \mathsf{true}$ if $\mathsf{lit}_i = x_i$ and $\sigma(x_i) = \mathsf{false}$ otherwise. Thus, C_j is satisfied under σ, implying that σ is a satisfying assignment. $\square$

4 Algorithms for Bounded Preferences Length

When bounding the preference length by two it is not hard to show that (coalitional) exchange-stability can be decided in linear time.

Theorem 3 (⋆). 2-ESM*,* 2-ESR*,* 2-CESM*, and* 2-CESR *can be solved in linear time.*

Fixed-parameter Algorithm for 3-ESR. We now turn to preference length at most three . In Theorem 1 we have seen that even this case remains NP-hard, even for bipartite preference profiles. Moreover, the proof suggests that a main obstacle that one has to deal with when solving 3-ESM (and hence 3-ESR) are the switch gadgets. Here we essentially show that they are indeed the *only* obstacles, that is, if there are few of them present in the input, then we can solve the problem efficiently. We capture the essence of the switch gadgets with the following structure that we call hourglasses.

Definition 1. *Let $\mathcal{P}$ be a preference profile and $V_H \subseteq V$ a subset of $2h$ agents with $V_H = \{u_i, w_i \mid 0 \leq i \leq h-1\}$. We call the subgraph $G(\mathcal{P})[V_H]$ induced by V_H an hourglass of height h if it satisfies the following:*

- *For each $i \in \{0, h-1\}$ the degrees of u_i and w_i are both at least two in $G(\mathcal{P})[V_H]$;*
- *For each $i \in [h-2]$, the degrees of u_i and w_i are exactly three in $G(\mathcal{P})[V_H]$;*
- *For each $i \in \{0, 1, \ldots, h-1\}$ we have $\{u_i, w_i\} \in E(G(\mathcal{P})[V_H])$;*
- *For each $i \in \{0, 1, \ldots, h-2\}$ we have $\{u_i, w_{i+1}\}, \{u_{i+1}, w_i\} \in E(G(\mathcal{P})[V_H])$.*

We refer to the agents u_i and w_i from V_H as layer-i agents. We call an hourglass H maximal if no larger agent subset $V' \supsetneq V(H)$ exists that induces an hourglass.

Given an hourglass H in $G(\mathcal{P})$, we call a matching M for $\mathcal{P}$ perfect for H if for each agent $v \in V(H)$ we have $M(v) \in V(H) \setminus \{v\}$. Further, M is exchange-stable for H if no two agents from $V(H)$ can form an exchange-blocking pair.

Notice that the smallest hourglass has height two and is a cycle with four vertices. We are ready to show the following fixed-parameter tractability result.

Theorem 4 (⋆). *An instance of* 3-ESR *with $2n$ agents and ℓ maximal hourglasses can be solved in $O(6^\ell \cdot n\sqrt{n})$ time.*

The main ideas are as follows. The first observation is that a matching for a maximal hourglass can interact with the rest of the graph in only six different ways: The only agents in an hourglass H of height h that may have neighbors outside H are the layer-0 and layer-$(h-1)$ agents; let us call them *connecting agents* of H. A matching M may match these agents either to agents inside or outside H. Requiring M to be perfect means that an even number of the connecting agents has to be matched inside H. This then gives a bound of at most six different possibilities of the matching M with respect to whether the connecting agents are matched inside or outside H. Let us call this the *signature* of M with respect to H. Hence, we may try all 6^ℓ possible combinations of signatures for all hourglasses and check whether one of them leads to a solution (i.e., an exchange-stable matching).

The second crucial observation is that each exchange-blocking pair of a perfect matching yields a four-cycle and hence, is contained in some maximal hourglass. Thus, the task of checking whether a combination of signatures leads to a solution decomposes into (a) checking whether each maximal hourglass H allows for an exchange-stable matching adhering to the signature we have chosen for H and (b) checking whether the remaining acceptability graph after deleting all agents that are in hourglasses or matched by the chosen signatures admits a perfect matching.

Task (b) can clearly be done in $O(n \cdot \sqrt{n})$ time by performing any maximum-cardinality matching algorithm (note that the graph $G(\mathcal{P})$ has $O(n)$ edges). We then prove that task (a) for all six signatures can be reduced to checking whether a given hourglass admits a perfect and exchange-stable matching. This, in turn, we show to be linear-time solvable by giving a dynamic program that fills a table, maintaining some limited but crucial facts about the structure of partial matchings for the hourglass.

5 Paths to Exchange-Stability

We now study the parameterized complexity of P-ESM with respect to the number of swaps. Observe that it is straightforward to decide an instance of P-ESM with $2n$ agents in $O((2n)^{2k+2})$ time by trying k times all of the $O(n^2)$ possibilities for the next swap and then checking whether the resulting matching is exchange-stable. The next theorem shows that the dependency of the exponent on k in the running time cannot be removed unless FPT = W[1].

Theorem 5 ($\star$). Path to Exchange-Stable Marriage *is W[1]-hard with respect to the number k of swaps.*

Proof (Sketch). We provide a parameterized reduction from the W[1]-complete Independent Set problem, parameterized by the size of the independent set [13]: Therein, given a graph H and an integer h, we want to decide whether G admits an h-vertex *independent* set, i.e., a subset of h pairwise nonadjacent vertices.

Let $I = (H, h)$ be an instance of Independent Set with vertex set $V(H) = \{v_1, v_2, \ldots, v_n\}$ and edge set $E(H)$. We construct an instance $I' = (\mathcal{P}, M_0, 2h)$ of P-ESM where $\mathcal{P}$ has two disjoint agent sets U and W, each of size $2n + h$. Both U and W consist of h *selector-agents* and $2n$ *vertex-agents* with preferences which encode the adjacency of the vertices in $V(H)$. More precisely, for each $j \in [h]$, we create two selector-agents, called s_j and t_j, and add them to U and W, respectively. For each $i \in [n]$, we create four vertex-agents, called x_i, u_i, y_i, w_i, add x_i and u_i to U, and add y_i and w_i to W. Altogether, we have $U = \{s_j \mid j \in [h]\} \cup \{u_i, x_i \mid i \in [n]\}$ and $W = \{t_j \mid j \in [h]\} \cup \{w_i, y_i \mid i \in [n]\}$.

Now we define the preferences of the agents from $U \cup W$. For notational convenience, we define two subsets of agents which shall encode the neighborhood of a vertex: For each vertex $v_i \in V(H)$, define $Y(v_i) \coloneqq \{y_z \mid \{v_i, v_z\} \in E(H)\}$ and $U(v_i) \coloneqq \{u_z \mid \{v_i, v_z\} \in E(H)\}$.

$$\begin{aligned}
&\forall j \in [h]\colon\ s_j : w_1 \succ \cdots \succ w_n \succ t_j, && t_j : u_1 \succ \cdots \succ u_n \succ x_1 \succ \cdots \succ x_n \succ s_j,\\
&\forall i \in [n]\colon\ x_i : t_1 \succ \cdots \succ t_h \succ y_i, && y_i : u_i \succ x_i \succ [U(v_i)],\\
&\forall i \in [n]\colon\ u_i : w_i \succ [Y(v_i)] \succ y_i \succ t_1 \succ \cdots \succ t_h, && w_i : s_1 \succ \cdots \succ s_h \succ u_i\ .
\end{aligned}$$

Herein, $[Y(v_i)]$ (resp. $[U(v_i)]$) denotes the unique preference list where the agents in $Y(v_i)$ (resp. $U(v_i)$) are ordered ascendingly according to their indices. Observe that the acceptability graph $G(\mathcal{P})$ includes the following edges:

- For all $i \in [h]$ and $j \in [n]$, the edges $\{s_i, t_i\}$, $\{s_i, w_j\}$, $\{t_i, x_j\}$, $\{t_i, u_j\}$, $\{w_j, u_j\}$, $\{y_j, x_j\}$, $\{y_j, u_j\}$ are in $E(G(\mathcal{P}))$.
- For all edges $\{v_i, v_{i'}\} \in E(G)$, the edges $\{u_i, y_{i'}\}$ and $\{u_{i'}, y_i\}$ are in $E(G(\mathcal{P}))$.

We define an initial matching M_0 on $G(\mathcal{P})$ as $M_0 = \{\{s_j, t_j\} \mid j \in [h]\} \cup \{\{w_i, u_i\}, \{y_i, x_i\} \mid i \in [n]\}$. This completes the construction of I', which can clearly be done in polynomial time. It is straight-forward to check that that $\mathcal{P}$ is bipartite and the construction can be done in linear time. The correctness proof is given in the full version [12]. □

6 Conclusion

Regarding preference restrictions [8], it would be interesting to know whether deciding (coalitional) exchange-stability for complete preferences would be become tractable for restricted preferences domains, such as single-peakedness or single-crossingness. Further, the NP-containment of the problem of checking whether a given matching may reach an exchange-stable matching is open.

References

1. Abizada, A.: Exchange-stability in roommate problems. Rev. Econ. Des. **23**, 3–12 (2019). https://doi.org/10.1007/s10058-018-0217-0
2. Abraham, D.J., Manlove, D.F.: Pareto optimality in the roommates problem. Technical report, University of Glasgow, Department of Computing Science (2004). tR-2004-182
3. Alcalde, J.: Exchange-proofness or divorce-proofness? Stability in one-sided matching markets. Econ. Des. **1**, 275–287 (1995)
4. Aziz, H., Goldwaser, A.: Coalitional exchange stable matchings in marriage and roommate market. In: Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), pp. 1475–1477 (2017). extended Abstract
5. Biró, P., Norman, G.: Analysis of stochastic matching markets. Int. J. Game Theory **42**(4), 1021–1040 (2012). https://doi.org/10.1007/s00182-012-0352-8

6. Bodlaender, H.L., Hanaka, T., Jaffke, L., Ono, H., Otachi, Y., van der Zanden, T.C.: Hedonic seat arrangement problems. In: Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), pp. 1777–1779 (2020). extended Abstract
7. Bodlaender, H.L., Hanaka, T., Jaffke, L., Ono, H., Otachi, Y., van der Zanden, T.C.: Hedonic seat arrangement problems. Technical report, arXiv:2002.10898 (cs.GT) (2020)
8. Bredereck, R., Chen, J., Finnendahl, U.P., Niedermeier, R.: Stable roommate with narcissistic, single-peaked, and single-crossing preferences. Auton. Agent. Multi-Agent Syst. **34**(53), 1–29 (2020)
9. Cechlárová, K.: On the complexity of exchange-stable roommates. Discret. Appl. Math. **116**(3), 279–287 (2002)
10. Cechlárová, K., Manlove, D.F.: The exchange-stable marriage problem. Discret. Appl. Math. **152**(1–3), 109–122 (2005)
11. Chen, J.: Reaching stable marriage via divorces is hard. Technical report, arXiv:1906.12274(cs.GT) (2020)
12. Chen, J., Chmurovic, A., Jogl, F., Sorge, M.: On (coalitional) exchange-stable matching. Technical report, arXiv:2105.05725(cs.GT) (2021)
13. Cygan, M., et al.: Lower bounds for kernelization. In: Parameterized Algorithms, pp. 523–555. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21275-3_15
14. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. Am. Math. Mon. **120**(5), 386–391 (1962)
15. Irving, R.W.: Stable matching problems with exchange restrictions. J. Comb. Optim. **16**(4), 344–360 (2008)
16. Knuth, D.: Mariages Stables. Les Presses de L'Université de Montréal (1976)
17. McDermid, E., Cheng, C.T., Suzuki, I.: Hardness results on the man-exchange stable marriage problem with short preference lists. Inf. Process. Lett. **101**(1), 13–19 (2007)