Strategy-Proofness and Non-Obvious Manipulability of Top-Trading-Cycles with Strategic Invitations

Shinnosuke Hamasaki, Taiki Todo and Makoto Yokoo
Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, 819-0395, Japan
hamasaki@agent.inf.kyushu-u.ac.jp, {todo, yokoo}@inf.kyushu-u.ac.jp

Keywords: Two-Sided Matching, Top-Trading-Cycles Algorithm, Non-Obvious Manipulability.

Abstract: Diffusion mechanism design is one of the recent trends in the literature of mechanism design. Its purpose is to incentivize agents to diffuse the information about the mechanism to as many followers as possible, as well as reporting their preferences. This paper is the first attempt to consider diffusion mechanism design for two-sided matching from the perspective of non-obvious manipulability. We focus on the top-trading-cycles (TTC) mechanism for the many-to-one two-sided matching problem. We clarify the necessary and sufficient condition for the mechanism to satisfy strategy-proofness and non-obvious manipulability, respectively. We also propose a new TTC-based matching mechanism that violates strategy-proofness but satisfies non-obvious manipulability, which illustrates how we can handle strategic information diffusion in two-sided matching.

1 INTRODUCTION

As one of the active fields in the area of artificial intelligence, multi-agent systems have been attracting considerable attention of researchers and practitioners. In a multi-agent system, multiple agents interact with each other and the society containing these agents makes a joint decision. Such a process is called a multi-agent decision-making.

For the research of multi-agent decision-making, game theory has played an important role. More specifically, mechanism design is considered as a mathematical foundation of multi-agent decision-making, especially when agents are self-interested and not cooperative with each other. The main purpose of mechanism design is to develop decision-making rules, also known as *mechanisms*, which incentivize selfish agents to take desirable actions.

Strategy-proofness is a well-known incentive property in the literature. It requires that for each agent, reporting a true private information, which is in many cases referred to a true *type* of an agent, to a mechanism is a dominant strategy. While it is quite appealing for achieving a socially-acceptable outcome, there are a lot of negative results regarding strategy-proofness, because requiring the existence of dominant strategy equilibria is too demanding. Weakening strategy-proofness and/or choosing other incentive properties is then a natural direction.

In this paper we consider *diffusion mechanism design* (Li et al., 2017), in which participation to a decision-making is invitation-based. An agent can participate in a decision-making only when other participating agents invite her. In diffusion mechanism design, strategy-proofness is rather demanding, since it requires that telling a true preference and inviting as many colleagues as possible is a dominant strategy for every agent. Cho et al. (2022) showed several impossibility theorems on strategy-proofness in diffusion mechanism design for two-sided matching.

Given above impossibility theorems, in this paper we consider *non-obvious manipulability* as a weaker notion of incentive property, instead of strategy-proofness. Non-obvious manipulability intuitively requires that, for each agent, truth-telling is weakly better than any manipulation in both the best- and worst-cases. While several existing works have investigated non-obvious manipulability in the literature of mechanism design for two-sided matching (please refer to the next section for a survey), analysis of non-obvious manipulability in diffusion mechanism design does not exist, as far as the authors know.

To sum up, this paper is the first attempt to consider non-obvious manipulability in the diffusion mechanism design for two-sided matching. We first show that, even if the incentive property is weakened to non-obvious manipulability, known impossibility results by Cho et al. (2022) come over. We

Hamasaki, S., Todo, T. and Yokoo, M.
Strategy-Proofness and Non-Obvious Manipulability of Top-Trading-Cycles with Strategic Invitations.
DOI: 10.5220/0013319100003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 1*, pages 616-623
ISBN: 978-989-758-737-5; ISSN: 2184-433X

then focus on the *Top-Trading-Cycles* (TTC) mechanism (Abdulkadiroğlu and Sönmez, 2003), which is known not to be manipulable in the standard settings without strategic invitation, and provide two if-and-only-if conditions for it to satisfy strategy-proofness and non-obvious manipulability, respectively, in two-sided matching with strategic invitation. We further propose a new mechanism that violates SP but satisfies NOM.

1.1 Related Works

Gale and Shapley (1962) initiated the research of two-sided matching and proposed the seminal Deferred-Acceptance mechanism. Crawford (1991) studied the effect of having more students/colleges in the two-sided matching. In their model the set of students varies as a result of exogenous events, while in our model it is due to the strategic actions of students. Various extensions of two-sided matching have also been studied, including school choice with diversity constraints (Kurata et al., 2017), matchings with budget constraints (Aziz et al., 2020), and uncertain preferences (Rastegari et al., 2013; Todo et al., 2021).

Several works investigated strategy-proof mechanisms with monetary compensations from the perspective of diffusion mechanism design (Kawasaki et al., 2020; Li et al., 2024). On the other hand, there is limited research on diffusion mechanism design without money. Recently, Kawasaki et al. (2021) and You et al. (2022) considered house allocation over social networks, without monetary compensations. Another recent work by Ando et al. (2025) studied strategy-proof social choice over social networks. Cho et al. (2022) dealt with two-sided matching with strategic information diffusion. However, all these works focused on strategy-proofness, and never considered non-obvious manipulability.

Given the difficulties of designing strategy-proof mechanisms, analysis based on the non-obvious manipulability (Troyan and Morrill, 2020) is one of the recent trends. Ortega and Klein (2023) proposed a two-sided matching mechanism that violates strategy-proofness but satisfies the non-obvious manipulability, although their model is quite standard and not dealing with strategic invitation.

2 MODEL

Our model of two-sided matching over social networks is basically identical with Cho et al. (2022), while we assume that every student is acceptable by any college and the underlying social network among students are restricted to be a tree. We will further define some additional notations to formalize the property of non-obvious manipulability.

In our model, there are two sets of agents, *students* and *colleges*. Let $C = \{c_1, c_2, \ldots, c_{|C|}\}$ be the set of colleges, and let $S = \{s_1, s_2, \ldots, s_{|S|}\}$ be the set of students. We usually use $c \in C$ and $s \in S$ to represent a college and a student without specifying their identifiers. The symbol $\varnothing$ denotes an "unmatched" status for students. Furthermore, special agent o, called *moderator*, corresponds to a trusted third party.

Each college c has a *priority* $\succ_c$, given as a strict ordering over S, specifying its one-to-one comparison over students. As Cho et al. (2022) assumed, we do not care about how colleges compare two subsets of students. Let $\succ_C = (\succ_c)_{c \in C}$ represent a profile of the priorities of colleges C. Each college c has its *maximum quota* $q_c \in \mathbb{Z}_{>0}$, indicating the number of students that college c can accept. Let $q_C := (q_c)_{c \in C}$.

Each student s has a *preference* $\succ_s$, given as a strict ordering over $C \cup \{\varnothing\}$. A notation $c \succ_s c'$ indicates that s strictly prefers being assigned to c instead of c'. Analogously, $c \succ_s \varnothing$ indicates that s strictly prefers being assigned to c to being unmatched. Symbol $\succsim_s$ indicates the "weak preference" associated with $\succ_s$; since we focus on strict preferences, $c \succsim_s c'$ indicates either $c \succ_s c'$ or $c = c'$. Let $\succ_S = (\succ_s)_{s \in S}$ represent a preference profile of students S.

Students are distributed over a social network. Let $r_o \subseteq S$ be the set of the neighbors of o, which are also called the *direct children* of o. For each s, let $r_s \subseteq S \setminus \{s\}$ denote s's neighbors. The neighborhood relation is asymmetric, i.e., $s' \in r_s$ does not imply $s \in r_{s'}$. Given $r_S := (r_s)_{s \in S}$ and r_o, all the neighborhood relations are defined, specifying *social network* $G(r_S, r_o)$ among students and the moderator. In this paper we assume that social networks are tree-shaped, where the moderator is located at the root.

Matching m specifies to which college each student is assigned. Given matching m, let $m(s) \in C \cup \{\varnothing\}$ denote the college (if any) to which student s is assigned, and $m(c) \subseteq S$ denote the set of students (if any) with which college c is matched. We abuse $\succ_s$ and write $m \succ_s m'$ (or $m \succsim_s m'$) instead of $m(s) \succ_s m'(s)$ (or $m(s) \succsim_s m'(s)$).

Let $\theta_s = (\succ_s, r_s)$ denote the *true type* of student s, and let $\theta = (\theta_s)_{s \in S}$ denote a profile of the students' true types. Let θ_{-s} denote a profile of the types owned by the students except s. Analogously, given subset (also called as a *coalition*) $T \subseteq S$, let θ_T denote a profile of the types owned by T, and let θ_{-T} denote a profile of the types owned by the students except T. Let $R(\theta_s) = \{\theta'_s = (\succ'_s, r'_s) \mid r'_s \subseteq r_s\}$ denote the set of *reportable types* by s with true type

θ_s, assuming that each s cannot pretend to be connected to any student to whom s is not really connected. When s reports r'_s as her neighbors, we say s *diffuses the information toward* r'_s, or s *invites* r'_s. Let $\theta' = (\theta'_s)_{s\in S} \in \times_{s\in S} R(\theta_s) = R(\theta)$ denote a reportable type profile. Analogously, given subset $T \subseteq S$, let θ'_T denote a profile of the types reported by T, and θ'_{-T} a profile of the types reported by the students except T.

Given type profile θ', let $\hat{S}(\theta') \subseteq S$ denote the set of *connected students* to whom a path exists from o in $G(r'_S, r_o)$. Given θ', let $M(\theta')$ denote a set of *feasible* matchings m satisfying the followings:

1. consistency; for any $s \in \hat{S}(\theta')$ and any $c \in C$, $m(s) = c \Leftrightarrow s \in m(c)$
2. max. quota constraint; for any $c \in C$, $|m(c)| \leq q_c$
3. connectivity; for any $s \in S$, $s \notin \hat{S}(\theta') \Rightarrow m(s) = \varnothing$

Given true type profile θ (which is not observable), *mechanism* μ maps a reported profile $\theta' \in R(\theta)$ into feasible matching $m \in M(\theta')$, while μ can use $\succ_C$, q_C, and r_o as parameters.

We further define some terms related to social networks. Given θ' and $s \in \hat{S}(\theta')$, let $d_s(\theta') \in \mathbb{Z}$ be the distance (the number of edges in the shortest path) from o to s in $G(r'_S, r_o)$. For any $s \notin \hat{S}(\theta')$, let $d_s(\theta') = \infty$. Also, for any $s \in S \cup \{o\}$, let $\delta_s(\theta')$ be the set of descendants of s in $G(r'_S, r_o)$.

Definition 1 (Strategy-Proofness (SP)). *Given mechanism μ, arbitrarily chosen student $s \in S$, profile $\theta'_{-s} \in R(\theta_{-s})$, true type θ_s, and misreport $\theta'_s \in R(\theta_s)$, let $m := \mu(\theta_s, \theta'_{-s})$ and $m' := \mu(\theta'_s, \theta'_{-s})$. Then, μ satisfies* strategy-proofness (SP) *if $m(s) \succsim_s m'(s)$ holds.*

Strategy-proofness requires that, for each student s, telling a true type θ_s is a dominant strategy, i.e., the outcome under the truth-telling is weakly better than the outcome under reporting a fake type θ'_s.

Definition 2 (Non-Obvious Manipulability (NOM)). *Given mechanism μ, arbitrarily chosen student $s \in S$, a true profile $\theta'_{-s} \in R(\theta_{-s})$, given preference $\succ_s$, and a type θ'_s of student s,*

- $B_{\succ_s}(\theta'_s) := c \in C \cup \{\varnothing\}$ *s.t.* $\exists \theta'_{-s} \in R(\theta_{-s})$:
  - $\mu_i(\theta'_s, \theta'_{-s}) = c$, *and*
  - $\forall \theta''_{-s} \in R(\theta_{-s}),\ c \succsim_s \mu_s(\theta_s, \theta''_{-s})$.
- $W_{\succ_s}(\theta_s) := c \in C \cup \{\varnothing\}$ *s.t.* $\exists \theta'_{-s} \in R(\theta_{-s})$:
  - $\mu_s(\theta_s, \theta'_{-s}) = c$, *and*
  - $\forall \theta''_{-s} \in R(\theta_{-s}),\ \mu_s(\theta_s, \theta''_{-s}) \succsim_s c$.

Then, μ satisfies non-obvious manipulability *if both*

$$B_{\succ_s}(\theta_s) \succsim_s B_{\succ_s}(\theta'_s) \quad \text{and} \quad W_{\succ_s}(\theta_s) \succsim_s W_{\succ_s}(\theta'_s)$$

holds for any $\theta_s := (\succ_s, r_s)$ and any $\theta'_s \in R(\theta_s)$.

The condition on $B_{\succ_s}(\cdot)$ denotes the best-case incentive constraint, which requires that telling a true type θ_s is better than telling any other type θ'_s, under the true preference $\succ_s$, in the best-case. Here the best-case is calculated by changing the reports of other students $S \setminus \{s\}$. Analogously, the condition on $W_{\succ_s}(\cdot)$ denotes the worst-case incentive constraint.

We also define stability, efficiency, and fairness properties, which are used in Cho et al. (2022), for obtaining our impossibility theorems.

The mutually-best property intuitively require that, if there is a pair of a student and a college who ranks them at the top with each other, such a pair is matched. Some weaker variants of the mutually-best properties are also defined.

Definition 3 (Mutually-Best (MB)). *Given θ', we say a pair of student s and college c is a* mutually-best pair (MB-pair) *if $c \succ_s c'$ for any $c' \neq c$ and $s \succ_c s'$ for any $s' \in \hat{S}(\theta')$. Matching m is* mutually-best *if any MB-pair (s, c) is matched as long as $s \in \hat{S}(\theta')$. A mechanism is said to satisfy MB if it always returns a mutually-best matching. Matching m is* mutually-best for direct children *if any MB-pair (s, c) is matched as long as $s \in \hat{S}(\theta')$ and $s \in r_o$. A mechanism is said to satisfy MB-D if it always returns a matching that is mutually-best for direct children.*

Non-wastefulness is a well-known notion of efficiency, which intuitively requires that, if a student cannot enter a college that she strictly prefer to her current assignment, it must be the case that the college is full. One of its weaker notions, introduced here as weak non-wastefulness, only requires that, the existence of such a student implies that either she is already assigned to some college or the college she prefers already have some student assigned.

Definition 4 (Weak Non-Wastefulness (WNW)). *A matching m is* weakly non-wasteful (WNW) *for given θ' if, for any $s \in \hat{S}(\theta')$ and any $c \in C$, $c \succ_s m(s)$ implies either $m(s) \neq \varnothing$ or $|m(c)| > 0$.*

Fairness is a condition related to students' envies toward other students. A pair of a college c and a student s forms a blocking coalition for a given matching m if both $c \succ_s m(s)$ and $s \succ_c s'$ for some $s' \in m(c)$ hold. In words, the blocking pair like them with each other. Intuitively, a matching is fair if such a blocking pair does not exist. In the definition below, a weaker notion of fairness is also defined.

Definition 5 (Fairness). *For student s and college c, assume $c \succ_s m(s)$. Then, s has (i)* justified envy with respect to priority *toward student $s' \in m(c)$ if $s \succ_c s'$, and (ii)* justified envy with respect to network *toward $s' \in m(c)$ if both $s \succ_c s'$ and $s' \notin \pi_s(\theta')$. Matching m is* fair *(FR) for given θ' if, there is no student with jus-*

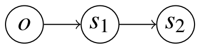

Figure 1: An example showing the incompatibility between NOM and MB, and the incompatibility among NOM, FR, and MB-D. Each circle indicates students (or the moderator), and each arrow indicates invitation.

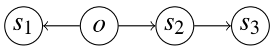

Figure 2: An example showing the incompatibility among NOM, FRN, and WNW.

tified envy with respect to priority. Matching m is fair with respect to network *(FRN) for given* θ' *if there is no student with justified envy with respect to network.*

Cho et al. (2022) showed the following theorems, all of which requires SP as an incentive property.

Theorem 1 (Cho et al. (2022))**.** *There exists no mechanism that simultaneously satisfy SP and MB.*

Theorem 2 (Cho et al. (2022))**.** *There exists no mechanism that simultaneously satisfy SP, FR, and MB-D.*

Theorem 3 (Cho et al. (2022))**.** *There exists no mechanism that simultaneously satisfy SP, FRN, and WNW.*

Finally, we define the *Top-Trading-Cycles* (TTC) mechanism for the classical two-sided matching problem (Abdulkadiroğlu and Sönmez, 2003).

Definition 6. *In Round $t = 1, 2, \ldots$, each student $s \in S$ points to her most preferred college, if any, among those who still have a non-zero capacity. Each college $c \in C$ points to the student who has the highest priority at that college. For each cycle, each belonging student is assigned to the college that she is pointing to. Remove that student, and reduce the quota of the college by one. The algorithm terminates when all the students are assigned or all the colleges have zero capacity; otherwise, it proceeds to Round $t+1$.*

3 IMPOSSIBILITIES

We first show three impossibility theorems, each of which strengthens an existing result (Theorems 1, 2, and 3) by replacing SP with NOM. The examples in their proofs are almost identical with the original ones in Cho et al. (2022), while our proofs are a bit more complicated due to the weakened incentive property.

Theorem 4. *There exists no mechanism that simultaneously satisfy NOM and MB.*

Proof. Consider the social network shown in Fig. 1, with two students, s_1 and s_2, both of whom have preference $c \succ \varnothing$. Note that θ_{s_1} is defined here as tuple $\theta_{s_1} = (c_1 \succ_{s_1} \varnothing, \{s_2\})$. There is one college c_1 with priority $s_2 \succ_{c_1} s_1$ and quota $q_{c_1} = 1$.

When s_1 sincerely reports θ_{s_1}, implying that student s_2 is invited, there is at least a case where s_2 reports a preference $c_1 \succ_{s_2} \varnothing$. In such a case, any mechanism satisfying MB must match s_2 to c_1 and leave s_1 unmatched. This is clearly the worst possible case for s_1. Thus, $W_{\succ_{s_1}}(\theta_{s_1}) = \varnothing$ holds.

On the other hand, when s_1 reports $\theta'_{s_1} = (c_1 \succ_{s_1} \varnothing, \emptyset)$, i.e., decides not to invite student s_2, MB implies that s_1 is matched to c_1. Therefore, $W_{\succ_{s_1}}(\theta'_{s_1}) = c_1$ holds. Since her true preference assumes $c_1 \succ_{s_1} \varnothing$, $W_{\succ_{s_1}}(\theta'_{s_1}) \succ_{s_1} W_{\succ_{s_1}}(\theta_{s_1})$ holds, violating NOM. □

Theorem 5. *There exists no mechanism that simultaneously satisfy NOM, FR, and MB-D.*

Proof. Consider the social network shown in Fig. 1, with two students, s_1 and s_2, both of whom have preference $c \succ \varnothing$. There is one college c_1 with priority $s_2 \succ_{c_1} s_1$ and quota $q_{c_1} = 1$.

When student s_1 does not invite student s_2, the MB-D condition requires that s_1 is matched to c_1. Thus, the worst-case outcome when s_1 does not invite s_2 is c_1. To guarantee NOM, it must be the case that the worst-case outcome when s_1 sincerely reports her true type, i.e., invites s_1, must be also c_1. However, if s_1 is assigned to college c_1, student s_2 has a justified envy with respect to priority toward s_1, since we have $s_2 \succ_{c_1} s_1$. This is a violation to FR. □

Theorem 6. *There exists no mechanism that simultaneously satisfy NOM, FRN, and WNW.*

Proof. Consider three students s_1, s_2, s_3 and two colleges c_1 and c_2. The social network among students are given as Fig. 2, and their preferences are given as $c_1 \succ c_2 \succ \varnothing$. Colleges have capacity $q_{c_1} = q_{c_2} = 1$, and their priorities are given as $s_3 \succ s_1 \succ s_2$. Assume that all the three students report their types truthfully. To guarantee WNW, two students must be assigned to two colleges, one-by-one. In other words, exactly one student is unmatched.

If student s_1 is unmatched, she has a justified envy with respect to network toward s_2, since both colleges prefer s_1 to s_2. Thus it violates FRN. If student s_3 is unmatched, s_3 has a justified envy with respect to network toward s_1, which violates FRN.

Finally, if student s_2 is unmatched, consider the worst-case outcomes. When s_2 invites s_3, s_2 is unmatched in the worst-case outcome. On the other hand, if s_2 does not invite s_3, s_2 is matched to c_2 in the worst-case outcome, to guarantee FRN. Thus, the worst-case condition of NOM is violated for s_2. □

These impossibility results have a quite negative implication. Given original impossibility theorems presented by Cho et al. (2022), it is quite natural to consider designing new two-sided matching mechanisms satisfying all the requirements in the theorems except for SP, and guaranteeing NOM instead. Our impossibility results therefore show that such a natural direction never provides any positive finding.

4 ACHIEVING STRATEGY-PROOFNESS

We now consider restricting the target instances. More precisely, we will give a necessary and sufficient conditions on parameters (e.g., colleges' priorities, quotas, and the underlying social network) for TTC to satisfy SP and/or NOM. This section focuses on SP, and the next section focuses on NOM.

Theorem 7. *TTC satisfies SP if and only if*

$$\forall s \in S,\ \ \forall c \text{ s.t. } q_c \leq \#\nu_{c,s},\ \ \forall s' \in \delta_s(\theta'),\quad s \succ_c s'$$

holds, where $\nu_{c,s} := \{s'' \in S \mid s'' \neq s, s'' \succ_c s\}$.

The notation $\nu_{c,s}$ denotes the set of students s'' that is more prioritized than student s at college c. The condition intuitively requires that any student s has a higher priority than any of her descendants at any college c. Only the exception is where college c has an enough capacity so that c can still accept student s after accepting some of s's descendants.

Proof. We first show the sufficiency. When the condition holds, we can guarantee that s is matched to such a college c before any of her descendants $s' \in \delta_s$. Thus, s's invitation strategy r'_s never affects the assignment of s. Also, if c has enough capacity, s can be assigned to c under truth-telling. Thus, strategy-proofness is guaranteed.

We then show the necessity. Assume that $\exists s \in S$, $\exists c$ s.t. $q_c \leq \#\nu_{c,s}$, and $\exists s' \in \delta_s(\theta')$, it holds that $s' \succ_c s$. Now consider the case where student s' is a descendant of student s, both s and s' have preference $c \succ \varnothing \succ \cdots$, arbitrarily chosen $q_c - 1$ students among $\nu_{c,s} \setminus \{s'\}$ have preference $c \succ \cdots$, and all the other students have preference $\varnothing \succ \cdots$. In such a case, student s is assigned to college c when she choose not to invite s' (more precisely, the unique child of s who is an ancestor of s'), but left unmatched when she invites s'. This is a violation to SP. $\square$

5 ACHIEVING NON-OBVIOUS MANIPULABILITY

We then consider achieving NOM by the TTC mechanism. The following theorem provides the necessary and sufficient condition for TTC to satisfy NOM.

Theorem 8. *TTC satisfies NOM if and only if either*

$$\forall s \in S,\ \ \forall c \text{ s.t. } q_c \leq \#\nu_{c,s},\ \ \forall s' \in \delta_s(\theta'),\quad s \succ_c s'$$

or

$$\forall s \in S,\ \forall \beta_s \subseteq \alpha_s := \{c \in C \mid c \succ_s \varnothing\},\ \ \sum_{c \in \beta_s} q_c \leq \# \bigcup \chi_{c,s}$$

holds, where $\nu_{c,s} := \{s'' \in S \mid s'' \neq s, s'' \succ_c s\}$ *and* $\chi_{c,s} := \{s'' \in S \setminus \delta_s(\theta') \mid s'' \neq s, s'' \succ_c s\}$.

The first condition is exactly identical with Theorem 7, since SP implies NOM. On the other hand, the second condition requires that, for any subset β_s of colleges that student s is willing to go, the sum of their quotas must not exceed the number of students that are more prioritized than s in at lease one of these colleges. This condition resembles the well-known Hall's marriage theorem (Hall, 1935). Indeed, our proof strategy for sufficiency essentially searches a perfect matching between students and colleges.

Proof Sketch. First of all, it is obvious that the best-case condition of NOM holds for every student $s \in S$; consider the case where all the students except s reports $\varnothing \succ c$ for every c, and invites all their neighbors. Then, every student except s points to herself, and thus s is matched to the college that she ranks the best, by definition of TTC. In other words, the LHS of the best-case condition of NOM is the most preferred college of s. Therefore, the best-case condition holds regardless of what the RHS is. Furthermore, from Proposition 1 and the fact that TTC is resilient to preference misreport, we can restrict our attention to showing NOM only for the diffusion strategy.

Therefore, it suffices to show that the worst-case condition of NOM, restricted only for the diffusion strategy, is satisfied if and only if the given conditions are satisfied. From now on we will show both the sufficiency and the necessity.

About the sufficiency, it is obvious that TTC satisfies NOM, and even SP, if the first condition is satisfied, as we have already shown in Theorem 7. Here we show that, the second condition is also sufficient to guarantee the worst-case condition of NOM, restricted only for the diffusion strategy. As lemma 9 shows, not inviting any student achieves the best-possible worst case. However, if the second condition holds, we can find a profile of reports θ'_{-s} under which all the remaining sheets in α_s are filled. Thus, the RHS of the worst-case condition of NOM is $\varnothing$. Since TTC never assigns students to any college that they are not willing to go, the RHS of the worst-case condition of NOM is weakly better than $\varnothing$, which complete the proof for the sufficiency.

We then show the necessity. Due to its complexity, we will show an example and explain the intuition.

See Fig. 3, where there are four students $s_1,\dots,s_4$. We assume that there are two colleges c_1 and c_2, whose quotas are set as $q_{c_1}=2$ and $q_{c_2}=1$. Now consider student s_1's strategic invitation.

Assume that student s_1's preference is given as $c_1 \succ c_2 \succ \varnothing$, and colleges' priorities are given as

$$c_1 : s_4 \succ s_2 \succ s_1 \succ s_3 \qquad c_2 : s_4 \succ s_1 \succ s_2 \succ s_3$$

Here, $s_2 \succ_{c_1} s_1$ violates the first condition of NOM, and the second condition for NOM is also violated for s_1 and $\beta_{s_1} := \{c_1,c_2\}$; $\sum_{c\in\beta_{s_1}} q_c = 3$ and $\#\bigcup_{c\in\beta_{s_1}} \chi_{c,s_1} = 2$.

In this example, the worst-case outcome when s_1 sincerely invites s_2 is that both s_4 and s_2 are assigned to college c_1 and s_1 is assigned to c_2. On the other hand, the worst-case outcome when s_1 does not invite s_2 is that both s_4 and s_1 are assigned to c_1 and s_3 is assigned to c_2. In other words, the LHS of the worst-case condition of NOM is c_2, and the RHS is c_1. Since $c_1 \succ_{s_1} c_2$ holds, NOM is violated, which concludes the sketch of the proof. □

The following proposition and lemma are used in the proof of Theorem 8.

Proposition 1. *Assume a mechanism μ is resistant to preference misreport for every student s, i.e., telling a true preference dominates telling a false preference, when s's invitation strategy and all the other students' reports are fixed. Then, μ satisfies NOM if and only if μ satisfies NOM only for the diffusion strategies.*

Proof. The only-if direction is obvious, since we restrict possible manipulations by student s. We then show the if direction. For the sake of contradiction, assume that, under such a mechanism μ, there exists student s, s's true type $\theta_s := (\succ_s, r_s)$, and s's misreport $\theta'_s := (\succ'_s, r'_s) \in R(\theta_s)$ such that either

$$B_{\succ_s}(\theta'_s) \succ_s B_{\succ_s}(\theta_s) \quad \text{or} \quad W_{\succ_s}(\theta'_s) \succ_s W_{\succ_s}(\theta_s).$$

Note that μ is resistant to preference misreport. Therefore, for a type $\theta''_s := (\succ_s, r'_s)$, both

$$B_{\succ_s}(\theta''_s) \succsim_s B_{\succ_s}(\theta'_s) \quad \text{and} \quad W_{\succ_s}(\theta''_s) \succsim_s W_{\succ_s}(\theta'_s)$$

holds; even for any fixed θ'_{-s}, it holds that $\mu_s(\theta''_s,\theta'_{-s}) \succsim_s \mu_s(\theta'_s,\theta'_{-s})$. Thus, either

$$B_{\succ_s}(\theta''_s) \succsim_s B_{\succ_s}(\theta'_s) \succ_s B_{\succ_s}(\theta_s)$$

or

$$W_{\succ_s}(\theta''_s) \succsim_s W_{\succ_s}(\theta'_s) \succ_s W_{\succ_s}(\theta_s)$$

holds, which contradicts the assumption that μ satisfies NOM for the diffusion strategies. □

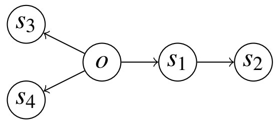

Figure 3: Example with Four Students.

Lemma 9. *Given type $\theta_s := (\succ_s, r_s)$ of student $s \in S$, consider a type misreport $\theta'_s := (\succ_s, r'_s) \in R(\theta_s)$ in which student s is not misreporting her preference and just consider changing her invitation strategy r'_s. Then, under the TTC mechanism, the worst-case outcome $W_{\succ_s}(\theta'_s)$ becomes the most preferred by student s when s does not invite any student.*

Proof Sketch. To consider the worst-case for student s, we can arbitrarily change the preference of all the other invited students. Therefore, from the definition of TTC, when student s invites some students s', we can imitate the case where s does not invite s' by just setting the preference $\succ_{s'}$ such that $\varnothing \succ \cdots$. In other words, there are weakly more possible outcomes when s invites s', implying that the worst-case outcome is weakly worse when s invites s'. □

The following example explains what the necessary and sufficient condition requires, and shows that how TTC violates SP and satisfies NOM.

Example 1. *Assume there are four students, $s_1,\dots,s_4$, and two colleges, c_1 and c_2. The social network among students are given as Fig. 3, and student s_1 has a preference $c_1 \succ c_2 \succ \varnothing$. Colleges have capacity $q_{c_1} = q_{c_2} = 1$, and their priorities are:*

$$c_1 : s_4 \succ s_2 \succ s_3 \succ s_1 \qquad c_2 : s_4 \succ s_1 \succ s_2 \succ s_3$$

Note that $s_2 \succ_{c_1} s_1$ violates the condition in Theorem 7, but satisfies the other condition in Theorem 8 for every $s \in S$. For example, if we choose $\beta_{s_1} := \{c_1,c_2\}$, both $\sum_{c\in\beta_{s_1}} q_c = 2$ and $\#\bigcup_{c\in\beta_{s_1}} \chi_{c,s_1} = 2$ hold, which do not violate the other condition.

Here, a profile of preferences of $s_2,\dots,s_4$ exists, under which s_1 has an incentive not to invite s_2, say,

$$s_2 : c_1 \succ c_2 \succ \varnothing \quad s_3 : c_2 \succ \varnothing \succ c_1 \quad s_4 : c_2 \succ c_1 \succ \varnothing$$

Student s_1 is unmatched under truth-telling, but would be matched to c_1 if she does not invite s_2.

However, student s_1 cannot get a better outcome in both of the best- and worst-cases; s_1 is assigned to c_1 in the best-case under her sincere preference report, and s_1 is left unmatched in the worst-case even if she does not invite s_2, e.g., consider the case where s_3 is still willing to go to college c_1.

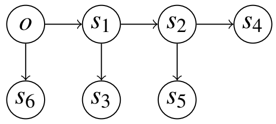

Figure 4: An Example Showing that IBSI-TTC Violates SP.

6 NEW MECHANISM SATISFYING NOM

As we have already shown in Section 3, just achieving NOM with other desirable properties is difficult. Therefore, in this section we ignore other properties and focus on achieving NOM.

Definition 7 (Invitation-Based Stepwise-Improving TTC (IBSI-TTC))**.** *Let τ be an empty matching under which no student is assigned to any college. Then, in each Phase $p = 1, 2, \ldots$, apply TTC for all the students with distance less than or equal to p, except for those who are at distance one and do not invite any student. Let τ_p be the outcome of Phase p. If all the students who invite at least one student weakly prefers τ_p to τ, then let $\tau := \tau_p$; otherwise keep the current τ. If all the students are at distance less than or equal to p, the algorithm terminates and returns τ; otherwise go to Phase $p+1$.*

The following example demonstrates how the new algorithm works, and shows that it violates SP.

Example 2. *There are six students, $s_1, \ldots, s_6$, and three colleges, $c_1, \ldots, c_3$. The social network are given as Fig. 4, and preferences are:*

$$
\begin{array}{ll}
s_1 : c_3 \succ \varnothing \succ c_1 \succ c_2 & s_2 : c_1 \succ c_2 \succ \varnothing \succ c_3 \\
s_3 : c_1 \succ c_2 \succ \varnothing \succ c_3 & s_4 : c_2 \succ c_1 \succ \varnothing \succ c_3 \\
s_5 : c_1 \succ c_2 \succ \varnothing \succ c_3 & s_6 : \cdots
\end{array}
$$

Colleges have capacity $q_{c_1} = q_{c_2} = q_{c_3} = 1$, and their priorities are given as follows:

$$
\begin{array}{ll}
c_1 : & s_5 \succ s_4 \succ s_3 \succ s_2 \succ s_1 \succ s_6 \\
c_2 : & s_2 \succ s_3 \succ s_5 \succ s_4 \succ s_1 \succ s_6 \\
c_3 : & s_1 \succ s_2 \succ s_3 \succ s_4 \succ s_5 \succ s_6
\end{array}
$$

Note that s_6 is a direct child of o and does not invite any student. Such a student is ignored at all in IBSI-TTC, and thus left unmatched in any case.

In Phase 1, IBSI-TTC applies TTC for s_1. In the tentative matching τ_1, s_1 is matched to c_1. Since this is better than $\varnothing$ for s_1, let $\tau := \tau_1$.

In Phase 2, IBSI-TTC applies TTC for s_1, s_2, and s_3. Under the tentative outcome τ_2, s_1 is matched to c_3, s_3 is matched to c_1, and s_2 is matched to c_2. Since τ_2 is weakly better than τ for both s_1 and s_2, who invited at least one student, let $\tau := \tau_2$.

In Phase 3, IBSI-TTC applies TTC for all the students. Under the tentative outcome τ_3, s_1 is matched to c_3, s_5 is matched to c_1 and s_2 is matched to c_2. Since this outcome τ_3 is weakly better than (more precisely, identical to) the current τ for s_1 and s_2, let $\tau := \tau_3$. We then get the final outcome τ which assigns s_1 to c_3, s_2 to c_2, and s_5 to c_1.

Now consider the case where s_2 decides not to invite s_5. Then, Phase 3 differs from the above; under the tentative outcome τ_3 in this case, s_1 is matched to c_3, s_4 is matched to c_2 and s_2 is matched to c_1. This is weakly better than the current τ for both s_1 and s_2. Thus, the final outcome assigns s_2 to c_1, which is strictly better for s_2 than the above case. violating SP.

Intuitively, this mechanism keeps updating the outcome by increasing the number of participating students, based on the agreement of those who invited at least one student. Thus, the candidate of final outcome, which is represented as τ in the description, is weakly monotonically increasing for them. As a result, the following lemma holds. Due to the space limitations, we omit the proof.

Lemma 10. *For any direct child $s \in r_o$ who invites at least one student, the final outcome of IBSI-TTC is weakly better than the first tentative outcome τ_1.*

Indeed, the following theorem shows that IBSI-TTC satisfies NOM, while it violates SP. As far as the authors know, this is the first example of mechanisms that satisfy NOM but violate SP under the two-sided matching model with information diffusion; the three mechanisms proposed by Cho et al. (2022) satisfies SP (or even stronger incentive properties), which implies that they also satisfy NOM.

Theorem 11. *IBSI-TTC satisfies NOM.*

Proof. There are three categories of students; (i) students who are not direct children of o (corresponding to students $s_2, \ldots, s_5$ in Example 2), (ii) students who are direct children of o and have at least one child (s_1 in Example 2), and (iii) students who are direct children of o but do not have children (s_6 in Example 2).

First of all, for those students in the category (iii), both the best- and worst-case conditions clearly hold, since such a student is always unmatched, regardless of what the other students report. In other words, both LHS and RHS are $\varnothing$, in both of the conditions.

It is also obvious that the best-case condition of NOM holds for every student $s \in S$ in the categories (i) and (ii); consider the case where all the students except s reports $\varnothing \succ_s c$ for every college c, and invites all their neighbors. In this case, s is matched to the college that she most prefers. In other words, the LHS of the best-case condition of NOM is the most preferred college of s. Therefore, the best-case condition holds regardless of what the RHS is.

We now show that the worst-case condition holds for every $s \in S$ in the categories (i) and (ii). For (i) students $s \notin r_o$, when the parent of s does not invite s, s is unmatched. Since IBSI-TTC never matches any student to a college that she is not willing to go, this is a worst-case, regardless of what s reports. Thus, both LHS and RHS of the worst-case condition is $\varnothing$, implying that the worst-case condition holds.

For (ii) students $s \in r_o$ who have at least one child, let $\theta_s := (\succ_s, r_s)$ be the true preference of s. Phase 1 of IBSI-TTC is resistant to preference misreport, from the known property of TTC in the traditional setting. Thus, when s arbitrarily misreports her preference, τ_1 gets weakly worse. Also, the tentative outcome at Phase 1 is identical, regardless to s's invitation strategy r'_s. Let τ'_1 be the tentative outcome at Phase 1 under s's arbitrary misreport θ'_s. We then have $\tau_1 \succsim_s \tau'_1$.

Furthermore, there is a possible profile of reports $\theta'_{-s} \in R(\theta_{-s})$ by the other students so that students in distance larger than one prefer to be unmatched. Under such a profile, the final outcome coincides with the Phase 1 outcome. That is, τ_1 is an upper bound of the worst-case outcome, even from the viewpoint of true preference $\succ_s$ of s, implying both

$$\tau_1 \succsim_s W_{\succ_s}(\theta_s) \quad \text{and} \quad \tau'_1 \succsim_s W_{\succ_s}(\theta'_s).$$

Also, Lemma 10 implies $W_{\succ_s}(\theta_s) \succsim_s \tau_1$. Thus, $W_{\succ_s}(\theta_s) \sim_s \tau_1 \succsim_s \tau'_1 \succsim_s W_{\succ_s}(\theta'_s)$ holds, which guarantees that the worst-case condition of NOM holds. $\square$

IBSI-TTC violates FR, FRN, WNW, MB, and MB-D. This is mainly due to two facts. First, it is based on TTC, which violates FRN (and thus FR). Second, it ignores direct children of o if they do not invite any students, which is totally wasteful and results in an unstable outcome. However, we believe that the idea behind the IBSI-TTC mechanism, guaranteeing NOM only and not paying too much attention to achieve SP, will be a useful building block for designing better NOM mechanisms in the future.

7 CONCLUDING REMARKS

Our model of two-sided matching with strategic invitation is limited in the sense that the social network among students are restricted as a tree-shaped. Handling more general structures would be a promising future direction. For each impossibility theorem, it would also be required to show the independence of the properties by providing mechanisms satisfying all except one properties, though we strongly believe that they are independent. Diffusion mechanism design is still a new and developing model of mechanism design. We believe there are further various extensions to achieve relatively positive results, including restricting preferences and allowing randomization.

REFERENCES

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747.

Ando, R., Kimura, K., Todo, T., and Yokoo, M. (2025). Parameterized voter relevance in facility location games with tree-shaped invitation graphs. To appear in *Proc. WALCOM '25*.

Aziz, H., Baychkov, A., and Biró, P. (2020). Summer internship matching with funding constraints. In *Proc. AAMAS '20*, pages 97–104.

Cho, S.-H., Todo, T., and Yokoo, M. (2022). Two-sided matching over social networks. In *Proc. IJCAI-ECAI '22*, pages 186–193.

Crawford, V. P. (1991). Comparative statics in matching markets. *J. Econ. Theory*, 54(2):389–400.

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

Hall, P. (1935). On representatives of subsets. *J. the London Mathematical Society*, s1-10(1):26–30.

Kawasaki, T., Barrot, N., Takanashi, S., Todo, T., and Yokoo, M. (2020). Strategy-proof and non-wasteful multi-unit auction via social network. In *Proc. AAAI '20*, pages 2062–2069.

Kawasaki, T., Wada, R., Todo, T., and Yokoo, M. (2021). Mechanism design for housing markets over social networks. In *Proc. AAMAS '21*, pages 692–700.

Kurata, R., Hamada, N., Iwasaki, A., and Yokoo, M. (2017). Controlled school choice with soft bounds and overlapping types. *J. Artif. Intell. Res.*, 58:153–184.

Li, B., Hao, D., Zhao, D., and Zhou, T. (2017). Mechanism design in social networks. In *Proc. AAAI '17*, pages 586–592.

Li, M., Cao, Y., and Zhao, D. (2024). Double auction on diffusion network. In *Proc. AAAI '24*, pages 9848–9855.

Ortega, J. and Klein, T. (2023). The cost of strategy-proofness in school choice. *Games and Economic Behavior*, 141:515–528.

Rastegari, B., Condon, A., Immorlica, N., and Leyton-Brown, K. (2013). Two-sided matching with partial information. In *Proc. ACM-EC '13*, pages 733–750.

Todo, T., Wada, R., Yahiro, K., and Yokoo, M. (2021). Lazy gale-shapley for many-to-one matching with partial information. In *Proc. ADT '21*, pages 390–405.

Troyan, P. and Morrill, T. (2020). Obvious manipulations. *J. Econ. Theory*, 185:104970.

You, B., Dierks, L., Todo, T., Li, M., and Yokoo, M. (2022). Strategy-proof house allocation with existing tenants over social networks. In *Proc. AAMAS '22*, pages 1446–1454.