Robust Mechanism Design on Networks with Externalities

Kohmei Makihara*

November 16, 2024

Abstract

Decision makers often seek to target the most deserving, or the most productive, people within a community, while lacking perfect information. This paper examines the allocation problem of a good with a positive externality without monetary payments. Agents, embedded in a network, know both their own and their neighbors' valuations. The principal's goal is to allocate the good to the agent with the highest valuation by proposing a mechanism which asks each agent to report their own and their neighbors' valuations and allocates the good based on these reports. This requires to correctly incentivize agents to report their truthful information. Due to the positive externalities, agents' incentives are partially aligned with the principal's objective—an agent not only wants to receive the good but also prefers that the agent with the highest valuation receives it if they do not. The paper identifies the network structures in which an efficient mechanism exists without assuming any common knowledge on the distribution from which the valuation is drawn, regardless of their beliefs about agents they are not directly connected to. We show that such mechanism exists if and only if at least two agents are connected to everyone. If agents do not use weakly dominated strategies, such mechanism exists if there is at least one agent connected to everyone. This insight guides decision makers in structuring agent networks when they have control over connections.

Keywords: network, full implementation, belief-free implementation, interdependent valuations, mechanism design without transfers

JEL classifications: D82, C72, D83

*Aix Marseille Univ, CNRS, AMSE, Marseille, France - kohmei.makihara@univ-amu.fr

1 Introduction

Many allocation problems involve distributing resources that benefit not only individual recipients but also the broader community. For example, consider a CEO deciding whom to appoint as the manager of a department within a company. The CEO aims to select the most productive individual for the role but has no information about each candidate's productivity. If a highly productive person is chosen, both the individual and the entire department benefit, as a capable leader can enhance overall performance. Another example is a government allocating a budget to subsidize research and development (R&D) for firms within a specific industry. While the subsidy directly benefits the recipient firm by fostering innovation, it also indirectly benefits the entire industry through spillover effects from new technologies developed by the subsidized firms. In both scenarios, because of the positive externalities, agents may have an incentive to truthfully report the most productive agent among those they know when asked by the CEO or the government whom aim to identify the most productive individual.

In such cases, a standard approach with monetary transfers is difficult to implement. For instance, a manager should not be promoted simply for transferring money to the CEO, and in the R&D context, the resource being allocated is the money itself, making direct monetary transfers impractical.

With these settings in mind, this paper examines an allocation problem without monetary transfers, faced by a principal such as a government or CEO, in situations where the characteristics of each agent (e.g., productivity, ability) are unknown to the principal. However, agents know their own characteristics and those of some other agents with whom they are connected in a network. For example, a department member may be more familiar with the work habits of colleagues who sit nearby or with whom they frequently collaborate. Additionally, the good being allocated generates a positive externality, which varies based on the characteristics of the agents who receive it. This means that each agent cares about who receives the good, even if they are not a recipient themselves.[1]

The principal's objective is to allocate the good to the most productive agent. To accomplish this, the principal designs a mechanism—an allocation rule based on information provided by agents—that incentivizes agents to reveal truthful information. We investigate mechanisms which asks each agent not only about themselves but also about their connected friends in the network. Agents, who have their own interest, might want to misreport information about himself or his neighbors if doing so increases their payoff. Moreover, whether such misreport can be prevented depends on the underlying network structure. For instance, if an agent i has a neighbor j who is only connected to i, both

[1] Thus, we can view our framework as a specific case of interdependent valuations, i.e. the utility of an agent depends on the valuations of the others which are not in general known by the agent. (Jehiel and Moldovanu (2001)).

agents know that the principal can only rely on the reports of i and j to identify the productivity of j. This might create a difficulty for the principal to detect the untruthful report of i and j, since the misreporting about this agent cannot be cross-checked by a third person. This underscores the importance of the network structure in helping the principal construct a mechanism that meets her objective. There are two natural questions in this setting. First, identifying which network structures allow for the existence of an efficient mechanism. Second, identifying such a mechanism when the existence conditions are met.

The answers to these questions depend heavily on each agent's beliefs about the information they lack. Since each department member may not know the characteristics of all other members but still cares about who receives the good, their decisions are influenced by their beliefs about unknown factors. For example, if an agent suspects that there may be a highly productive member they are unaware of, they might choose to underreport the productivity of a known colleague to increase the likelihood that this potentially more productive, unknown agent will be assigned the role.

The typical approach for modeling this situation assumes a common belief that each agent's productivity is drawn from a known distribution shared by all agents. This setting allows agents to compute the expected productivity of each agent, and to know that the others also have the same expectations as themselves. While this assumption simplifies the model considerably, it is restrictive and may not accurately reflect real-world scenarios, where agents often have different perceptions of how productive others are likely to be. Alternatively, a belief-free approach requires the existence of an efficient mechanism such that for any belief structure, all strategy profiles that can emerge as a Bayes-Nash equilibrium must be efficient. This approach necessitates not only to consider the setting that each agent might possess different expected productivity on the same agent, but also to model an extensive belief hierarchy for each agent—namely, beliefs about others' productivities, beliefs about others' beliefs about productivities, and so on.[2] This approach is much more demanding and, therefore, only applicable in limited contexts,[3] though it offers robustness to variations in belief structures.

Constructing a mechanism that identifies agents' productivities regardless of belief structure is challenging; however, the network helps to reduce the uncertainty agents face, making the setting less dependent on specific beliefs and thus opening up the possibility of belief-free efficiency. For instance, an agent connected to many others has access to more comprehensive information. If such agent exists, then this can help the principal in identifying agents' valuations, as the principal can rely more on the messages of well-connected agents, whose decisions depend less on belief compared to agents

[2]This is the characterization suggested by the seminal contribution of Harsanyi (1967) on the incomplete information games.

[3]See for instance Bergemann and Morris (2009) and Ollár and Penta (2017).

with limited information. However, a well-connected agent might also use this position strategically, providing misleading information to secure the good for himself, even if this contradicts the principal's objective. Thus, the interplay between network structure and the feasibility of accurately identifying agents' true productivities is complex and nuanced.

Our first result establishes the necessary and sufficient conditions for the existence of a mechanism that allocates the good to the most productive agent in any equilibrium, independent of each agent's belief hierarchy—a property we refer to as *robust efficiency.*[4] This concept is known to be stronger than several well-known solution concepts employed in the mechanism design literature.[5] We show that a robustly efficient mechanism exists *if and only if* the network is such that there are at least 2 agents who are connected to everybody in the community.

The intuition behind the necessary condition lies in the fact that each agent who is not connected to everyone may hold beliefs that an agent unknown to him has the highest productivity, and that others will send a message so that this agent receives the good. Since he believes that this unseen agent is the most productive, he has no incentive to change his message from the one which allows this unseen agent to obtain the good (even if it is untruthful), as he expects that allocating the good to this agent maximizes his utility. Consequently, even if he observes a more productive agent in his neighborhood, he might still send a message claiming that the central agent who is connected to everyone is more productive than this neighbor, assuming that the good will ultimately go to the most productive agent he cannot observe. This dynamic can incentivize the center to misrepresent his productivity in hopes of securing the good for himself, even if he is aware of a more productive agent.

We then prove sufficiency by exhibiting an efficient mechanism when the network has at least two centers. The efficient mechanism we construct to prove sufficiency operates within a simple framework that applies to a community with two agents who know each other. If both agents' messages contradict each other—each claiming to be more productive—the mechanism partially withholds the good, creating an allocation where the higher agent prefers partial allocation over the entire good going to the lower agent, but the lower agent prefers the entire good going to the higher agent over partial allo-

[4]The fact that *all equibria* must yield an efficient allocation is called *full implementation*, contrary to *partial implementation* which is satisfied when there is an equilibrium which is efficient but which does not exclude the existence of inefficient equilibria.

[5]This concept includes Bayesian implementability of Jackson (1991), since robust efficiency requires that for all belief structures, the mechanism is efficient in Bayes-Nash equilibrium. Moreover, Bergemann and Morris (2011) show that if a social choice function is robustly implementable, then it is *ex-post incentive compatible*. For ex-post implementation, see Crémer and McLean (1985), Bergemann and Morris (2008), and Feng et al. (2023).

cation. This balance depends on burning just the right amount of the good: too much, and both agents may collude untruthfully since the partial allocation is not attractive enough for the higher agent to deviate from the allocation giving the entire good to the lower; too little, and the contradiction might persist since the partial allocation is more attractive than the efficient allocation for the lower. Without monetary transfers, the principal cannot individually reward or punish agents, as the good benefits both through externalities. This framework ensures that when there are at least 2 agents connected to everyone, the less productive of the two has an incentive to report truthfully, given the positive externalities. This mechanism extends to a general network structure, with the two centers facing incentives similar as those just exposed.

Our necessity result restricts the set of networks to which our framework can be applied. This limitation arises from the demanding robustness requirement, which may lead agents to adopt untruthful strategies under the belief that their honesty will not impact the final allocation. Hence, we propose a weaker form of robustness to belief structures, assuming that agents do not use weakly dominated strategies—a concept we term *weakly robust efficiency*. Under this slightly relaxed criterion, we show that a weakly robustly efficient mechanism exists, provided there is at least one agent connected to everyone in the network. The importance of such network structures has been highlighted in multiple studies on social networks, and our work contributes another key implication of this structure.[6] Furthermore, we extend our concept of robustness to sequential mechanisms, showing that an efficient mechanism exists if the network can be divided into two sets of agents where each agent in one set is connected to every agent in the other set. This condition includes the networks for which a weakly robustly efficient mechanism exists.

This paper contributes to several areas of literature. First, it adds to recent work on mechanism design in information networks, where agents know their own types and those of their connected neighbors, and the principal designs a mechanism to gather this information to achieve her objective. In Bloch and Olckers (2022), the principal's goal is to create a complete ranking of agents without monetary transfer, with agents' utilities increasing based on their assigned ranks. They show that a necessary and sufficient condition for ex-post incentive compatible mechanism to exist is that every pair of agents shares a common neighbor. Besides the differing solution concept—ex-post incentive compatibility—which allows for inefficient equilibria, the key difference between our work and theirs is the role of externalities (hence, interdependent valuations): in their setting, agents only care about their own rank, fitting a private values framework.

Baumann (2023) examines full implementation in settings where goods are allocated

[6] For instance, in the network formation game of Bala and Goyal (2000), the star network is one of the 2 equilibria which might emerge in one of their models, as well as the unique efficient network for a wide range of parameters.

without monetary transfers. Here, the principal aims to allocate the good to the agent with the highest valuation, drawn from a distribution common to all agents. This model, based on private values, differs from ours with respect to externalities. Additionally, their framework assumes common knowledge of the distribution, allowing the use of standard Bayes-Nash implementation. Their study also relies on partially verifiable information, where agents have limited capacity to lie, and on lexicographic preferences that favor truth-telling when payoffs are equal—an aspect we do not incorporate in our model.

Secondly, this paper advances implementation theory and mechanism design in environments with interdependent valuations, particularly in contexts without monetary transfers. In these settings, agents' utility depends on the valuations of others, with each agent's valuation typically unknown to the others, creating an incomplete information game. Previous work on this topic has largely focused on allocations with monetary transfers (Crémer and McLean (1985), Jehiel and Moldovanu (2001)) and, more recently, on allocations without transfers (Bhaskar and Sadler (2019), Goldlücke and Tröger (2020), Feng et al. (2023)). Our study extends this literature by examining how specific network structures can ensure the existence of an efficient mechanism. In this context, outcomes rely heavily on assumptions about common knowledge, which shape the beliefs that agents hold. The standard approach assumes that agents have a common prior on the types of others, framing the situation as a Bayes-Nash game. Although this method provides key insights, the assumption of common knowledge does not take into account the complex belief architectures that agents might hold.

To address this point, Bergemann and Morris (2005) introduce the concept of robust implementation, which requires that all Bayes-Nash equilibria lead to efficient allocations across any belief space. This eliminates the need for common knowledge, as it guarantees efficient allocation regardless of the agents' beliefs. Bergemann and Morris (2009) establish necessary and sufficient conditions for the existence of such mechanisms; however, due to the allocative externalities that are present in our model, their framework does not cover our setting.[7] In a context where monetary transfers are allowed, Ollár and Penta (2017) address a related problem by providing conditions for an efficient transfer scheme to exist based on belief restrictions concerning the moments of the distributions from which each agent's valuation is drawn. Our contribution is to present a novel approach to achieving robust efficiency by leveraging network structures where agents have information about some of their peers, without imposing any belief restrictions and without relying on monetary transfers. Instead, we focus on the network structure and the positive externalities that align the incentives of the principal and the agents. It is important to note that we are not the first to utilize externalities as a means of incentive

[7]They assume that the utility of an agent depends only on the *aggregate* of the valuations, i.e. there exists an aggregator $h_i(\cdot)$ which gives a scalar depending on the realized valuation profiles.

alignment. For instance, Bhaskar and Sadler (2019) discuss the conditions for a second-best optimal mechanism and, using positive externalities, proposes an optimal mechanism in a Bayes-Nash equilibrium under a common prior, where agents' valuations are private information which, in our setting, amounts to considering that the network is empty. Our value-added regarding the use of the externalities is the following: by introducing the information network into such context, not only the first best is possible for some networks, but also with stronger solution concept which does not assume any common knowledge.

2 Model

A principal has 1 unit of divisible good to allocate among n agents. Let $N = \{1, \cdots, n\}$ with $n \geq 2$ be the set of all agents in the community. We denote the amount of good allocated to agent i as x_i, and the allocation profile as $\mathbf{x} = (x_1, \cdots, x_n) \in X$. We require that for all $i \in N$, $x_i \geq 0$, and $\sum_{i \in N} x_i \leq 1$. This implies that some of the good may not be allocated, allowing for money burning, which means the principal can choose not to allocate a portion of the good. The divisibility of the good can be interpreted in two ways: either as the inherent nature of the good itself (e.g., if the good to be allocated is a budget sum) or as the probability of allocating an indivisible good to an agent (e.g., the allocation of a single task or prize).

Each agent derives utility from the allocation profile. We take the formulation of Bhaskar and Sadler (2019), where the (ex-post) utility function for agent i is defined as follows:

$$u_i(\mathbf{x}; \mathbf{v}) = (1 - \alpha) v_i x_i + \alpha \sum_{j \in N} v_j x_j$$

where $v_i \in [0, 1]$ is the valuation of agent i, drawn from an unknown distribution, and $\alpha \in (0, 1)$ is the externality factor.[8] The term v_i represents the constant marginal utility of the good allocated to agent i, indicating either the agent's productivity or the degree to which each agent values the good. We denote the valuation profile as $\mathbf{v} = (v_1, \cdots, v_n)$. We assume that $v_i \neq v_j$ for any $i \neq j$.

The principal's objective is to allocate the entire good to the agent with the highest valuation. For instance, in the context of task allocation, this means assigning the task to the most productive agent. In the case of a simple allocation of a single good, the principal aims to allocate it to the agent who values it the most. We say that the allocation $\mathbf{x}$ is

[8] We assume that α depends neither on i nor on j. This represents a situation, for instance, where innovations are published through patents, and all firms can access them by paying a cost. Similarly, in an assignment setting, the benefit of a new manager is greater than the benefit that each individual department member receives which is considered as the same for everyone.

efficient for a given realization $\mathbf{v}$ if $x_i = 1$ for i such that $i = argmax_{i' \in N} v_{i'}$ and $x_j = 0$ otherwise.

Given the principal's objective, the externality factor α represents how closely aligned the incentives are between the principal and the agents.[9] When $\alpha = 1$, the utility function of an agent is simply $\sum_{j \in N} v_j x_j$ which is maximized when we allocate the entire good to the agent with the highest valuation. In this case, the efficient allocation maximizes the utility function of each agent. Conversely, as α decreases, the weight of the first term in the utility function increases, suggesting that the significance of each agent's own allocation becomes more pronounced. In the case, the efficient allocation may appear less appealing to agents whose valuations are not the highest, as they may prefer receiving the good for themselves.

The externality factor α can be interpreted in various ways depending on the context. For instance, consider a scenario where the good being allocated is the assignment of a project manager to a team member. Once the project manager is assigned and the project is completed, utility is realized based on the individual reward given only to the manager, alongside a group reward distributed to all team members. In this case, α represents the relative magnitude of the group reward compared to the individual reward. A larger α indicates that the group reward is more attractive than the individual reward, leading agents to prefer that another agent with a high valuation receives the good, thereby benefiting from the group reward, if such an agent exists. When $\alpha = 1$, it suggests that there is no individual reward, only the group reward, meaning maximizing the group reward is equivalent to maximizing the utility of each agent. Conversely, if $\alpha = 1$, it indicates that rewards are given solely to the individual who obtains the good, in this case, the project manager. In another context, when the principal aims to allocate the good to the individual who values it most, α can be seen as a measure of altruism, reflecting that the material benefit for agent i is $v_i x_i$. Furthermore, considering that the second term in the utility function represents total welfare, α can also be interpreted as a preference for efficiency, as discussed in various studies in experimental economics.[10]

We assume that the principal does not know the valuation of any agent, while each agent is aware of his own valuation as well as the valuations of his neighbors within a network. The network is represented by a matrix $\mathbf{G} = (g_{ij})_{i,j \in N}$, where $g_{ij} = 1$ indicates that agent i knows v_j. We further assume that the network is undirected, meaning $g_{ij} = g_{ji}$. Let N_i denote the set of neighbors of agent i.[11] The information possessed by agent i is described by $\theta_i(\mathbf{v}, \mathbf{G}) = (v_i, (v_j)_{j \in N_i})$. For sake of simplicity, we omit the notation for the dependence of the information type on the network and write $\theta_i = \theta_i(\mathbf{v}, \mathbf{G})$ when there is

[9] See also Bhaskar and Sadler (2019)

[10] See for instance Charness and Rabin (2002) or Engelmann and Strobel (2004).

[11] Crémer and McLean (1985) study the optimal auction design (i.e. with monetary transfer) where agents know the valuations of any other agent, i.e. where the network is complete.

no risk of confusion. The network $\mathbf{G}$ and the externality factor α are common knowledge among agents and the principal. The structure of the utility function, combined with the fact that an agent does not have complete knowledge of everyone's valuations, implies that an agent lacks full information about his utility function ex-ante.[12]

Mechanism and belief space

The principal's objective is to achieve an efficient allocation for any realization of valuation profile $\mathbf{v}$, which is unknown for her. To accomplish this, the principal proposes and commits to a mechanism, defined by the message space and the allocation rule. A mechanism is a pair $((M_i)_{i\in N}, \mathbf{x}(\cdot))$, where M_i is the message space for agent i, and $\mathbf{x} : \Pi_{i\in N} M_i \to X$ is the allocation rule. In essence, the principal defines the set of messages that each agent can send to her, and the allocation profile is determined according to the allocation rule $\mathbf{x}(\cdot)$ based on the message profile submitted by the agents.

From the agents' perspective, each mechanism induces a game form, in which the strategy set is the message space M_i, and the ex-post payoff function is defined as follows:

$$\pi_i(m_i, m_{-i}; \mathbf{v}) = u_i(\mathbf{x}(m_i, m_{-i}); \mathbf{v})$$

We consider a direct mechanism,[13] where the message space is defined as $M_i = \Theta_i(\mathbf{G})$ for all i. This means that each agent reports only the information they possess, specifically their own valuation and the valuations of their neighbors.

Since some parameters of the payoff function of agent i is unknown to himself in general, a mechanism constitutes a game of incomplete information. Consequently, the degree of common knowledge among agents and the principal becomes a crucial aspect of the model. The standard approach often assumes a Bayes-Nash equilibrium, where agents know that each valuation is drawn from a distribution common to all. However, this reliance on a common prior is often seen as overly restrictive, as it presumes uniform knowledge across agents. In many scenarios our model considers, it may be unrealistic to assume such common knowledge—for instance, it would be unlikely for all agents to have exact information about the distribution of each other's productivity. Avoiding this assumption of common knowledge, however, necessitates modeling the entire hierarchy of beliefs. This means incorporating not only each agent's beliefs about other agents'

[12] In implementation theory literature, this scenario is known as interdependent valuations, where an agent's utility depends on the types of other agents, which is generally assumed to be private information. See, for example, Jehiel and Moldovanu (2001).

[13] Some studies on implementation theory and mechanism design take the indirect mechanism approach to eliminate the inefficient allocation profile from the set of equilibria of the game induced by the mechanism (see for instance Palfrey and Srivastava (1989) or Bergemann and Morris (2008)). However, this often involves an augmented mechanism that uses variants of integer games as a device to eliminate inefficient allocations. Not only are integer games considered unrealistic to implement, but they might also lack any equilibrium due to the message space being infinite.

valuations but also beliefs about others' beliefs on those valuations, and so forth, as outlined by Harsanyi (1967). This recursive belief structure adds complexity but more accurately reflects situations with diverse information distributions.

3 Robustly efficient mechanism

In this paper, we take the approach of robust implementation proposed by Bergemann and Morris (2005) adapted to information networks. This robust concept ensures that the mechanism leads to an efficient allocation for any message profile that could be supported as a Bayes-Nash equilibrium under some belief hierarchy. In other words, it is robust against any possible configuration of beliefs and higher-order beliefs among agents. However, verifying Bayes-Nash equilibrium outcomes across all possible belief spaces is impractical. Instead, we use an alternative method that achieves the same result. This method involves iteratively eliminating strategies that are never a best response, which is effectively equivalent to assessing the Bayes-Nash equilibrium across all belief spaces.

To introduce this approach, we first define the expected payoff and the best-responses. Agent i's expected payoff depends on his conjecture conditional on his type, which is a joint probability distribution over the messages and valuations of the other agents. We require that the set of feasible conditional conjectures of an agent must be consistent with what the agent observes, namely, $\mu_i(\cdot \mid \theta_i) \in C_i(\theta_i)$, where

$$C_i(\theta_i) = \{\mu_i \in \Delta(M_{-i}, V_{-i}) \mid \mu_{-i}(\{(m_{-i}, v'_{-i}) \mid v'_j = v_j \text{ for } j \in N_i\}) = 1\}$$

where v_j is the true valuation of agent j. Given his conjecture $\mu_i(\cdot \mid \theta_i) \in C_i(\theta_i)$, the expected payoff when agent i adopts a message m_i is

$$E\pi_i(m_i; \theta_i, \mu_i) = \int_{M_{-i} \times V_{-i}} u_i(\mathbf{x}(m_i, m_{-i}); \mathbf{v}) d\mu_i(m_{-i}, v_{-i} | \theta_i)$$

and the best responses are

$$BR_i(\theta_i, \mu_i) = argmax_{m'_i \in M_i}\, E\pi_i(m'_i; \theta_i, \mu_i)$$

This formulation allows for uncertainty not only about other agents' valuations but also about the strategies they may choose. In the standard framework for games with incomplete information, strategies are functions of realized types, typically denoted by $s_i(\theta_i)$, making best responses functions of others' strategies. To see how this applies within our framework, suppose other agents' strategies $s_{-i}(\cdot)$ are fixed. Then, agent i's conjecture assigns positive probability only to pairs (m_{-i}, v_{-i}) such that $m_{-i} = s_{-i}(\theta_{-i}(\mathbf{v}))$, and such that the conjecture is consistent with the common prior. This approach thus models agent i's best response based on his conjecture over both messages and values of others, maintaining consistency with fixed strategies and prior beliefs.

The concept of robust efficiency is defined based on the iterative elimination of never-best responses for each agent given his information.[14] Let $S_i^0 = M_i$ and define inductively

$$S_i^k(\theta_i) = \Big\{ m_i \in M_i \,\Big|\, \exists \mu_i \in C_i(\theta_i) \text{ s.t. } m_i \in BR_i(\theta_i, \mu_i) \text{ and } \mu_i\big(\{(m_{-i}, v'_{-i}) \,|\, m_{-i} \in S_{-i}^{k-1}(\theta_{-i}(\mathbf{v}'))\}|\theta_i\big) = 1 \Big\}$$

and let $S_i(\theta_i) = \cap_{k\geq 0} S_i^k(\theta_i)$. Moreover, we say that a message profile $\mathbf{m}$ is rationalizable if $\mathbf{m} \in \Pi_{i\in N}\, S_i(\theta_i)$.

Definition 1 (Robust efficiency)**.** *Given the network* $\mathbf{G}$*, the allocation rule* $\mathbf{x}(\cdot)$ *is robustly efficient if for all* $\mathbf{v}$*,* $\theta \in \Pi_{i\in N}\, S_i(\theta_i)$ *and* $\mathbf{x}(\mathbf{m})$ *is efficient for all* $\mathbf{m} \in \Pi_{i\in N}\, S_i(\theta_i)$*.*

$S_i^k(\theta_i)$ is the set of all messages that survive the k-th round of elimination. In order for a strategy m_i to survive the k-th round of elimination given his observation θ_i, m_i must be a best response for some conjecture consistent with what he observes and which puts a positive probability only on the pairs (m_{-i}, v'_{-i}) such that the strategy profile m_{-i} survives the $(k-1)$-th round of elimination for all $j \neq i$, when he believes that the valuations of others are v'_{-i}. We can see that for all k, $S_i^k(\theta_i) \subseteq S_i^{k-1}(\theta_i)$ which guarantees that this process ends after certain number of rounds, i.e. there exists k^* such that $S_i^{k^*}(\theta_i) = S_i^{k^*+1}(\theta_i)$ for all i.[15] It is known that a message profile is in $\Pi_{i\in N} S_i(\theta_i)$ if and only if it can be played as an Bayes-Nash equilibrium in certain beliefs of agents.[16] Thus, robust efficiency requires that, regardless of the agents' beliefs—including all higher-order beliefs—the mechanism consistently results in an efficient allocation at the Bayes-Nash equilibrium. This invariance to belief variations is the essence of its "robust" nature.

To find a robustly efficient mechanism, it is necessary that all message profiles surviving the iterative elimination process yield an efficient allocation under the mechanism. This requires not only ensuring that among the surviving message profiles, there exists at least one that leads to an efficient allocation (partial implementation), but also that all inefficient message profiles are removed from the set of possible outcomes (full implementation).

Example 1. (Partial and full implementation) Assume that there are 2 agents, 1 and 2, connected to each other, hence they have complete information. The mechanism defines a message profile $\mathbf{m} = (m_1, m_2)$, where $m_i = (m_i^1, m_i^2)$ for $i \in \{1, 2\}$, in which

[14] This concept is a special case of Δ-*rationalizability* proposed by Battigalli and Siniscalchi (2003).

[15] We can see that by redefining this process in terms of an operator

$$b_i(S, \theta_i) = \{m_i \in M_i \,|\, \exists \mu_i \in C_i(\theta_i) \text{ s.t. } m_i \in BR_i(\theta_i, \mu_i) \text{ and } \mu_i\left(\{(m_{-i}, v_{-i}) \,|\, m_{-i} \in S_{-i}(\theta_{-i})\}\right) = 1\}$$

which is monotone in the set inclusion ordering, hence to which the Tarski's fixed point theorem is applicable.

[16] See Proposition 1 of Bergemann and Morris (2011).

m_i^j represents the message of agent i about the valuation of agent j. The allocation rule $\mathbf{x}(\mathbf{m}) = (x_1(\mathbf{m}), x_2(\mathbf{m}))$ is defined as follows:

- If $m_1^1 = m_2^1$ and $m_1^2 = m_2^2$, then $x_1(\mathbf{m}) = 1$ if $m_1^1 > m_1^2$, and $x_2(\mathbf{m}) = 1$ if $m_1^1 < m_1^2$.
- In any other case, $x_1(\mathbf{m}) = x_2(\mathbf{m}) = 0$.

This setup attempts to enforce efficiency by allocating the good only when the agents' messages are consistent, and it introduces a strong incentive for agents to be truthful. The mechanism includes an efficient profile where both agents report their true valuations, resulting in the allocation of the good to the agent with the higher valuation. However, the mechanism also allows for an inefficient outcome: both agents may send consistent but incorrect messages that still satisfy the consistency condition (i.e. $m_1 = m_2$). This possibility of a "consistent but wrong" message profile means that the good could be allocated based on incorrect information, failing to achieve efficiency.

For instance, let (v_1, v_2) with $v_1 > v_2$ be the true valuation profile, so that the efficient allocation is $x_1 = 1$. Consider a message profile $\mathbf{m}$ such that $m_1^j = m_2^j$ for $j \in \{1, 2\}$, and $m_i^1 < m_i^2$ for $i \in \{1, 2\}$. This message profile indeed survives the iterative elimination because for both agents, this (untruthful) message profile is the best response to the same (untruthful) message of the other agent.

As a result, while this mechanism includes an efficient allocation profile in the rationalizable set, it is not robustly efficient. Robust efficiency requires that all surviving message profiles (those that can be supported as rationalizable outcomes) yield an efficient allocation. The existence of inefficient outcomes—consistent yet incorrect message profiles—demonstrates that this mechanism does not meet the robust efficiency criterion.

This example illustrates a key point: achieving robust efficiency requires more than just penalizing agents for sending inconsistent messages. The mechanism must also provide incentives for agents to send truthful messages, even when doing so could result in an inconsistent message profile.

The primary challenge here stems from the lack of monetary transfers. In settings with monetary transfers, the principal could design a transfer scheme to reward agents individually, ensuring that one agent's reward does not impact the utility of others. However, without monetary transfers, the principal's only available punitive tool is to burn the good. This approach is inherently limited, as burning the good acts as a collective punishment. Because of the positive externalities in our setting, agents prefer that the good is allocated to someone—even if it is not themselves—rather than being destroyed.

Despite this difficulty, the following lemma shows that in 2 agents case there exists a simple allocation rule which rewards the agent with the higher valuation and punish the lower agent, so that the untruthful consistent message profiles are eliminated from the set of rationalizable message profiles.

		m_2	
		$m_2^1 > m_2^2$	$m_2^2 > m_2^1$
m_1	$m_1^1 > m_1^2$	$(1, 0)$	$\left(x_1^{(1,2)}, x_2^{(1,2)}\right)$
	$m_1^2 > m_1^1$	$\left(x_1^{(2,1)}, x_2^{(2,1)}\right)$	$(0, 1)$

Table 1: 2-agent mechanism

Lemma 1. *Assume that* $N = \{1, 2\}$ *and they are connected. The direct mechanism* $(M_1, M_2, \mathbf{x})$ *such that*

$$m_1^1 > m_1^2 \textit{ and } m_2^1 > m_2^2 \Rightarrow (x_1, x_2) = (1, 0)$$
$$m_1^1 < m_1^2 \textit{ and } m_2^1 < m_2^2 \Rightarrow (x_1, x_2) = (0, 1)$$
$$\textit{Otherwise} \Rightarrow (x_1, x_2) = \left(\frac{\alpha}{1+\alpha}, \frac{\alpha}{1+\alpha}\right)$$

is robustly efficient.

Proof. All proofs are relegated to the appendix. □

To illustrate the proof sketch of this lemma, we can examine the setup in Table 1. Inside the square is the allocation profile for each message profile. The payoff for each allocation profile is the following $\pi_1(x_1^{(1,2)}, x_2^{(1,2)}) = x_1^{(1,2)} v_1 + \alpha x_2^{(1,2)} v_2$ and $\pi_1(0, 1) = \alpha v_2$. Hence, by taking the difference, we have

$$\pi_1(x_1^{(1,2)}, x_2^{(1,2)}) - \pi_1(0, 1) = x_1^{(1,2)} v_1 - \alpha(1 - x_2^{(1,2)}) v_2$$

We need $\pi_1(x_1^{(1,2)}, x_2^{(1,2)}) > \pi_1(0, 1)$ if and only if $v_1 > v_2$. Hence

$$x_1^{(1,2)} = \alpha(1 - x_2^{(1,2)}) \tag{1}$$

By the same argument for agent 2, we obtain

$$\pi_2(x_1^{(1,2)}, x_2^{(1,2)}) - \pi_2(1, 0) = x_2^{(1,2)} v_2 - \alpha(1 - x_1^{(1,2)}) v_1$$

We need $\pi_2(x_1^{(1,2)}, x_2^{(1,2)}) > \pi_2(1, 0)$ if and only if $v_1 < v_2$. Hence

$$x_2^{(1,2)} = \alpha(1 - x_1^{(1,2)}) \tag{2}$$

We can see that the only allocation profile which satisfies (1) and (2) is indeed $x_1^{(1,2)} = x_2^{(1,2)} = \frac{\alpha}{1+\alpha}$. Besides, it is less restrictive for $(x_1^{(2,1)}, x_2^{(2,1)})$ can be anything low enough to have $\pi_1(x_1^{(2,1)}, x_2^{(2,1)}) < \pi_1(1, 0)$ if $v_1 > v_2$, and $\pi_2(x_1^{(2,1)}, x_2^{(2,1)}) < \pi_1(0, 1)$ if $v_1 < v_2$. These are also satisfied by the allocation $(\frac{\alpha}{1+\alpha}, \frac{\alpha}{1+\alpha})$ although it is not necessary.

The allocation of $\frac{\alpha}{1+\alpha}$ to both agents is more appealing for the higher-valued agent than allocating everything to the lower-valued agent, which occurs in the case of a consistent but untruthful message profile. Additionally, this allocation is less attractive for the lower-valued agent compared to the efficient allocation, where everything is given to the higher-valued agent. Note that as α increases, the amount of money burning decreases. This relationship illustrates that α serves as a measure of the degree of incentive alignment between the principal and the agents.

Using the same logic in larger networks with more than two agents, we can construct a robustly efficient mechanism even if not every agent is directly connected to every other agent. In fact, it suffices to have two agents within the network who are connected to all other agents. Furthermore, we can demonstrate that this particular network structure is also necessary for the existence of a robustly efficient mechanism.

Theorem 1. *A robustly efficient mechanism exists if and only if the network is such that there are at least 2 agents who are connected to everyone.*

We can understand the necessity of this network structure by considering the implications of having only one central agent connected to all other agents, and by assuming that a robustly efficient mechanism exists. In this scenario, since every agent except the central one has at least one other agent whom he does not observe, each agent might believe that the unseen agent has the highest valuation, and believe that all other agents send a message which results in the allocation of the entire good to this unseen agent. To simplify the argument, let us consider an agent i who is not the central agent and who holds such a belief. We will assume that agent i observes that a neighbor j has the highest valuation among the set of neighbors, which necessarily includes the central agent. However, as long as agent i believes that there is an unobserved agent with the highest valuation, and believes that everyone sends messages that result in this agent receiving the entire good, we can show that agent i has no incentive to report that agent j has a higher valuation than the central agent. This is because such a report should not change the allocation in favor of agent i, provided that the mechanism is robustly efficient. If this were the case, it would imply that the good should be allocated to agent i by some deviation of agent i, since he believes that the initial allocation is efficient.[17] This contradicts the concept of robust efficiency, as it suggests that agent i can change his report to benefit himself, even though the initial message profile allocating the good to the unseen agent was indeed the efficient allocation. Hence, we can always find a belief for agent i such that the untruthful message stating that the valuation of the central agent is higher than that of agent j is a best response for some strategy profile. Since this holds true for any non-central agent, the message profile indicating that the central

[17]For agent i, any allocation that does not allocate the good to him is worse than the efficient allocation.

agent has the highest valuation is always rationalizable, contradicting the assumption that the mechanism is robustly efficient.

We prove the sufficiency by constructing a specific mechanism that is robustly efficient. This mechanism employs the same logic as the mechanism in Lemma 1, where the two agents in question are those who are connected to everyone else in the network. The mechanism utilizes only the messages from these two agents. When there is an inconsistency in the claims made by both agents regarding who has the highest valuation, the mechanism allocates $\frac{\alpha}{1+\alpha}$ to both agents. This allocation structure makes the distribution of the entire good to the agent with the higher valuation more attractive for the agent with the lower valuation. The difference from Lemma 1 arises when there is another agent who has a higher valuation than both of them. In this case, the central agent with the lower valuation would prefer that this agent obtains the good rather than allowing the central agent who has the higher valuation among the two centers to obtain it.

Example 2. We illustrate the necessity of having a network with at least two agents connected to everyone for the existence of a robustly efficient mechanism. We do this by examining a network where there is only one agent in the center and demonstrating that no robustly efficient mechanism can exist in such a configuration.

Consider the network in Figure 1, where there is agent 1 who is connected to everyone, and take the valuation profile such that $v_2 > v_1 > v_3 > v_4$. We will see that for any agent

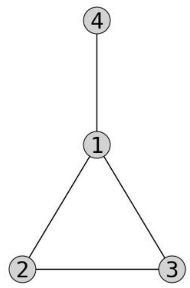

Figure 1: Only one agent is connected to everyone.

who observes agents 1 and 2, the message such that $m_i^1 > m_i^2$ cannot be eliminated, if we assume that there exists a robustly efficient mechanism. Let assume that there exists a robustly efficient mechanism, and let us take one of them.

Let us first take the agent 2's perspective. Since agent 2 does not observe agent 4, he can have a conjecture such that v_4 is the highest and $\alpha v_4 > v_2$,[18] and all other agents would send untruthful messages but consistent to each other, and tell that the valuation profile is $v_4 > v_1 > v_2 > v_3$. Since the mechanism is efficient, if everyone—including

[18]This implies that agent 2 prefers that agent 4 obtains the good rather than obtaining by himself.

agent 2—sends consistent messages telling that the valuation profile is $v_4 > v_1 > v_2 > v_3$, the mechanism must allocate the entire good to agent 4. Furthermore, given agent 2's beliefs, he has no incentive to deviate from this allocation, as he believes that it is more beneficial for him that agent 4 obtains the good. Therefore, for agent 2, sending the untruthful message telling that $v_1 > v_2$ can be a best response under certain conjectures.

Agent 3, who observes that agent 2 has indeed a higher valuation than agent 1, may also send an untruthful message in certain conjectures, specifically that agent 4 has the highest valuation and everyone sends consistent message telling that $v_4 > v_1 > v_2 > v_3$. In this case, exactly in the same way as in the case of agent 2, agent 3 does not have any incentive to deviate from this allocation since the good is allocated to agent 4, who is perceived as having the highest valuation according to agent 3's conjecture.

Given that agent 2 and 3 may send an untruthful message asserting that $v_1 > v_2$, and agent 4 may always send a truthful message indicating that $v_1 > v_4$, agent 1 might exploit these circumstances to send an untruthful message claiming that he has the highest valuation. If this occurs, the mechanism must allocate the entire good to agent 1, as everyone reports that v_1 is the highest. This situation implies that the mechanism is not robustly efficient, since in the true valuation profile, v_2 is actually the highest, and therefore the whole good should be allocated to agent 2.

4 Mechanism with weak robustness

The impossibility of a robustly efficient mechanism in networks without at least two agents who are connected to everyone arises from the fact that untruthful messages may constitute a best response in a weak sense, meaning they yield at most the same allocation as truthful messages for certain message profiles of other agents. This limitation significantly constrains the types of networks in which a robustly efficient mechanism exists. To expand the range of network structures for which we can construct an efficient mechanism, we slightly relax the definition of robustness. This new concept is still based on the iterative elimination of never-best responses; however, in each round of elimination, we also remove messages that are weakly dominated.

As in the previous section, let $T_i^0 = M_i$ and define inductively

$$
\begin{aligned}
T_i^k(\theta_i) = \Bigg\{ m_i \in M_i \,\Bigg|\, & \exists \mu_i \in C_i(\theta_i) \text{ s.t.} \\
& \mu_i(m_{-i}, v'_{-i} \,|\, \theta_i) > 0 \text{ for all } m_{-i} \in T_{-i}^{k-1}(\theta_{-i}(\mathbf{v}')), \\
& \mu_i(\{(m_{-i}, v'_{-i}) \,|\, m_{-i} \in T_{-i}^{k-1}(\theta_{-i}(\mathbf{v}'))\} \,|\, \theta_i) = 1, \text{ and } m_i \in BR_i(\theta_i, \mu_i) \Bigg\}
\end{aligned}
$$

and let $T_i(\theta_i) = \cap_{k \geq 0} T_i^k(\theta_i)$. We say that a message profile $\mathbf{m}$ is weakly rationalizable if $\mathbf{m} \in \Pi_{i \in N} T_i(\theta_i)$.

Definition 2 (Weakly robust efficiency). *Given the network* $\mathbf{G}$*, the allocation rule* $\mathbf{x}(\cdot)$ *is weakly robustly efficient if for all* $\mathbf{v}$, θ *is weakly rationalizable and* $\mathbf{x}(\mathbf{m})$ *is efficient for all* $\mathbf{m}$ *which are weakly rationalizable.*

In the iterative elimination process, we assume that each agent has a conjecture with full support for the messages of others; that is, each agent believes that if a message from another agent is not eliminated, then it might be used by that agent with strictly positive probability. This approach eliminates all messages that are weakly dominated in each round, given that the other agents only use the messages that survived in the previous round. Unlike the iterative elimination of never-best responses, it is important to note that convergence of the set is not guaranteed, as this process does not necessarily yield a monotonic operator within the lattice of set inclusion.

We propose a simple mechanism that is weakly robustly efficient when there is one agent who is connected to everyone. This mechanism applies to certain networks where it is known by Theorem 1 that a robustly efficient mechanism does not exist.

Mechanism with one agent in the center (WE mechanism)
Let agent 1 be the central agent who is connected to everyone. Let $i^* = argmax_{i \neq 1}\, m_1^i$.

- If $1 = argmax_{i \in N}\, m_1^i$, then
    - if $\forall i \neq 1$, $m_i^1 > m_i^i$, then $x_1 = 1$
    - otherwise, $x_1 = x_{i^*} = \frac{\alpha}{1+\alpha}$ and $x_i = 0$ for all $i \notin \{1, i^*\}$.
- Otherwise,
    - if $\forall i \neq 1$, $m_i^1 > m_i^i$, then $x_1 = 1$
    - otherwise, $x_{i^*} = 1$

Proposition 1. *The WE mechanism is weakly robustly efficient. Hence, there exists a weakly robustly efficient mechanism if there is one agent who is connected to everyone.*

For the central agent to obtain the entire good, all other agents must report that the central agent has a higher valuation than themselves. If there is another agent who claims to have a higher valuation than the center when the central agent claims to have the highest valuation, the mechanism allocates $\frac{\alpha}{1+\alpha}$ of the good to both the central agent and the non-central agent identified by the central agent as having the highest valuation among others. This mechanism incentivizes the central agent to report truthfully. Specifically, the central agent is discouraged from falsely claiming the highest valuation when he observes another agent with a higher valuation. This is because the central agent prefers the allocation where the highest-valuing agent receives the entire good over the alternative, in which $\frac{\alpha}{1+\alpha}$ is shared between the highest-valuing agent and the central agent, as shown in Lemma 1. Note that the central agent has an incentive to report truthfully only in a weak sense. Regardless of the message sent by the central agent, if all other agents report that the central agent has a higher valuation than themselves, the central agent

	m_3	
m_2	$m_3^1 > m_3^3$	$m_3^3 > m_3^1$
$m_2^1 > m_2^2$	$(1, 0, 0)$	$(\beta, \beta, 0)$
$m_2^2 > m_2^1$	$(\beta, \beta, 0)$	$(\beta, \beta, 0)$

(a) $m_1 : m_1^1 > m_1^2 > m_1^3$

	m_3	
m_2	$m_3^1 > m_3^3$	$m_3^3 > m_3^1$
$m_2^1 > m_2^2$	$(1, 0, 0)$	$(\beta, 0, \beta)$
$m_2^2 > m_2^1$	$(\beta, 0, \beta)$	$(\beta, 0, \beta)$

(b) $m_1 : m_1^1 > m_1^3 > m_1^2$

	m_3	
m_2	$m_3^1 > m_3^3$	$m_3^3 > m_3^1$
$m_2^1 > m_2^2$	$(1, 0, 0)$	$(0, 1, 0)$
$m_2^2 > m_2^1$	$(0, 1, 0)$	$(0, 1, 0)$

(c) $m_1 : m_1^2 = \max_{i \in N} m_1^i$

	m_3	
m_2	$m_3^1 > m_3^3$	$m_3^3 > m_3^1$
$m_2^1 > m_2^2$	$(1, 0, 0)$	$(0, 0, 1)$
$m_2^2 > m_2^1$	$(0, 0, 1)$	$(0, 0, 1)$

(d) $m_1 : m_1^3 = \max_{i \in N} m_1^i$

Table 2: 3 agents, agent 1 in the center

receives the entire good. Consequently, any untruthful message from the central agent is only weakly dominated.

We illustrate the mechanism and its incentive scheme in the example below by using the simple network of 3 agents.

Example 3. Let us consider a network with $N = \{1, 2, 3\}$, where $g_{12} = g_{13} = 1$ and $g_{23} = 0$. In this configuration, agent 1 is connected to both agents 2 and 3, while agents 2 and 3 are not connected to each other. Let $\beta := \frac{\alpha}{1+\alpha}$. Table 2 displays the allocations determined by the mechanism for each message profile. Each panel illustrates the resulting allocation profile for the messages of agents 2 and 3, corresponding to the message of agent 1 indicated in the caption. We will demonstrate that for any realization of the valuation profile, the message profiles that survive the iterative elimination yield an efficient allocation.

Assume first that the true valuation profile is $v_1 > v_2 > v_3$ or $v_1 > v_3 > v_2$. Then, we can see that for agent 1, $m_1^1 > m_1^2 > m_1^3$ or $m_1^1 > m_1^3 > m_1^2$ weakly dominate all other messages. We can see this by examining each allocation in the table. If both agent 2 and 3 chooses to send the truthful message, i.e. $m_2^1 > m_2^2$ and $m_3^1 > m_2^3$, then regardless of the message that agent 1 chooses, the allocation is $(1, 0, 0)$. However, if one of them sends a wrong message, then the resulting allocation is $(\beta, \beta, 0)$ if agent 1 chooses $m_1^1 > m_1^2 > m_1^3$ as shown in Panel (a), or $(0, 1, 0)$ if $m_1^2 = max_{i \in N}\, m_1^i$ as shown in Panel (c). According to Lemma 1, agent 1 prefers $(\beta, \beta, 0)$ to $(0, 1, 0)$ if and only if $v_1 > v_2$. On the other hand, if agent 1 observes that his valuation is not the highest, (w.l.o.g. let the realization be such that $v_2 > v_1 > v_3$), then, we find that for agent 1, the message $m_1^2 = max_{i \in N}\, m_1^i$ weakly dominates all other messages. This is because agent 1 prefers $(0, 1, 0)$ to $(\beta, \beta, 0)$ if and only if $v_2 > v_1$, as established in Lemma 1.

Furthermore, considering the perspective of agent 2, if the true valuation profile is such that $v_2 > v_1 > v_3$, he only observes that his valuation is higher than agent 1, but he does not observe agent 3. Consequently, agent 2 cannot determine whether for agent 1, this is $m_1^2 = max_{i \in N} m_1^i$ which is dominant (which occurs when $v_2 > v_3$) or $m_1^3 = max_{i \in N} m_1^i$ is dominant (which occurs when $v_3 > v_2$). However, in either case, the weakly dominant strategy for agent 2 is to report truthfully, that is, $m_2^2 > m_2^1$. This can be seen in Panels (c) and (d), where agent 2 weakly prefers the allocations corresponding to the message $m_2^2 > m_2^1$. It is important to note that if $m_1^3 = max_{i \in N} m_1^i$, it implies that $v_3 > v_2 > v_1$.[19] Therefore, in this scenario, agent 2 would prefer the allocation $(0, 0, 1)$ to $(1, 0, 0)$, which leads us to conclude that $m_2^2 > m_2^1$ is indeed weakly dominant.

Next, considering agent 3 who observes that $v_1 > v_3$, he knows that if $v_2 > v_1$, then agent 1 will choose $m_1^2 = max_{i \in N} m_1^i$, while agent 2 will choose $m_2^2 > m_2^1$. Consequently, agent 3 understands that his own message does not influence the allocation. On the other hand, if $v_1 > v_2$, it follows that agent 1 must choose either $m_1^1 > m_1^2 > m_1^3$ or $m_1^1 > m_1^3 > m_1^2$. In this scenario, the truthful message $m_3^1 > m_3^3$ for agent 3 remains dominant. This can be seen in Panel (a) which illustrates that agent 3 prefers $(1, 0, 0)$ to $(\beta, \beta, 0)$ when $v_1 > v_2$, and from Panel (b) which shows that agent 3 also prefers $(1, 0, 0)$ to $(\beta, 0, \beta)$ when $v_1 > v_3$.

As the above example illustrates, the process of iterative elimination begins with the central agent, who always has a weakly dominated message based on his observations. This fact allows us to eliminate the possibility of the central agent choosing an untruthful message to falsely claim that he has the highest valuation. In the context of (strictly) robust efficiency, the central agent may hold a conjecture that he is certain that others will send untruthful messages. By preventing this through the elimination of weakly never-best responses, we can construct a weakly robustly efficient mechanism.

5 Multistage mechanism

In this section, we extend the concept of robust efficiency to a multistage mechanism and examine whether the set of networks for which a robustly efficient mechanism exists can be broadened.

5.1 Multistage mechanism: model

We use the formulation of Battigalli and Siniscalchi (2003) to define the rationalizability on the multistage game, and apply to the mechanism design.

[19] Otherwise for agent 1, $m_1^3 = max_{i \in N} m_1^i$ would not be dominant.

We define the standard extensive-form games with observable actions and perfect recall. Unlike the simultaneous mechanism discussed in previous sections, here the principal designs an extensive-form mechanism, denoted by Γ, which includes a set of feasible histories $\mathcal{H}$ and a sequence of action profiles. We use $h' \subseteq h$ to indicate that h' is a subhistory of h, meaning h' is an initial subsequence of h. We denote the terminal history as $\overline{h}$ and the set of all terminal histories as Z. At each non-terminal history h, an agent is asked to choose an action from the set $A_i(h)$, with $|A_i(h)| > 1$ if agent i is active at that point.

Each terminal history is associated with an allocation defined by an allocation rule $\mathbf{x} : Z \to \mathcal{X}$. The (ex-post) utility function remains the same as in previous sections, which we present here:

$$u_i(\mathbf{x}; \mathbf{v}) = (1-\alpha)v_i x_i + \alpha \sum_{j \in N} v_j x_j$$

A message for agent i is a function $m_i : \mathcal{H} \to A_i$ such that $m_i(h) \in A_i(h)$. Each message profile $\mathbf{m} = (m_i)_{i \in N}$ corresponds to a terminal history through a function $O : M \to Z$, where $M = \Pi_{i \in N} M_i$ represents the set of all message profiles. Thus, we can express the payoff function of an agent as follows.

$$\pi_i(m_i, m_{-i}; \mathbf{v}) = u_i(\mathbf{x}(O(m_i, m_{-i})); \mathbf{v})$$

We omit the function $O(\cdot)$ and write $\mathbf{x}(m_i, m_{-i}) := \mathbf{x}(O(m_i, m_{-i}))$ when there is no confusion. As in the previous sections, agents are connected into a network $\mathbf{G}$ where $g_{ij} = 1$ implies that agent i knows the valuation agent j. Besides, we define the set of message profiles which are consistent with each history h, namely

$$M(h) = \{\mathbf{m} \in M \,|\, h \text{ is a subhistory of } O(\mathbf{m})\}$$

and let $M_i(h)$ be the set of m_i which does not prevent the game from reaching the history h. We also define a system of conditional conjectures $\mu_i(\cdot \,|\, \theta_i, h)$ which is a joint distribution over the messages and valuations of the other agents, conditional on each information type, and on each history. As in Section 3, the conditional conjecture must be consistent with what an agent observes, i.e. for all $\mu_i(\cdot \,|\, \theta_i, h) \in C_i(\theta_i)$. Moreover, the conjectures of each distinct history are connected via Bayes rule as follows. If h' is a subhistory of h'', then

$$\mu_i(m_{-i}, v_{-i}|\theta_i, h') = \mu_i(m_{-i}, v_{-i}|\theta_i, h'') \int_{V_{-i}} \int_{M_{-i}(h'')} \mu_i(m'_{-i}, v'_{-i}|\theta_i, h') dm'_{-i} dv'_{-i}$$

We denote a collection of all such conjectures as $D(\theta_i)$. Given a conjecture $\mu_i(\cdot \,|\, \theta_i, h) \in D(\theta_i)$, a non-terminal history h, the expected payoff of agent i with information θ_i when he plays message m_i is

$$E\pi_i(m_i; \theta_i, \mu_i, h) = \int_{M_{-i}(h) \times V_{-i}} u_i(\mathbf{x}(m_i, m_{-i}); \mathbf{v}) d\mu_i(m_{-i}, v_{-i}|\theta_i, h)$$

The rationalizability on the extensive form mechanism is defines by the sequential rationality. Given an information type θ_i and a conjecture μ_i, a message m_i^* is sequentially rational if for all $h \in \mathcal{H}$

$$E\pi_i(m_i^*; \theta_i, \mu_i, h) \geq E\pi_i(m_i; \theta_i, \mu_i, h)$$

for all m_i. Let $R_i(\theta_i, \mu_i)$ be the set of all sequentially rational messages given θ_i and μ_i.

The process of iterative elimination works similarly as the one in Section 4. Let $T_i^0 = M_i$, and define inductively

$$\begin{aligned} T_i^k(\theta_i) = \Bigg\{ m_i \in M_i \,\Bigg|\, & \exists \mu_i \in C_i(\theta_i) \text{ s.t. for all } h, \\ & \mu_i(m_{-i}, v'_{-i} \,|\, \theta_i, h) > 0 \text{ for all } m_{-i} \in T_{-i}^{k-1}(\theta_{-i}(\mathbf{v}')) \cap M_{-i}(h), \\ & \mu_i(\{(m_{-i}, v'_{-i}) \,|\, m_{-i} \in T_{-i}^{k-1}(\theta_{-i}(\mathbf{v}')) \cap M_{-i}(h)\} \,|\, \theta_i, h) = 1, \text{ and } m_i \in R_i(\theta_i, \mu_i) \Bigg\} \end{aligned}$$

and let $T_i(\theta_i) = \cap_{k \geq 0} T_i^k(\theta_i)$. We say that a message profile $\mathbf{m}$ is *sequentially weakly rationalizable* if $\mathbf{m} \in \Pi_{i \in N} T_i(\theta_i)$.

Definition 3 (Sequentially weakly robust efficiency)**.** *Given the network* $\mathbf{G}$*, the allocation rule* $\mathbf{x}(\cdot)$ *is sequentially weakly robustly efficient if for all* $\mathbf{v}$*,* θ *is sequentially weakly rationalizable and* $\mathbf{x}(\mathbf{m})$ *is efficient for all* $\mathbf{m}$ *which are sequentially weakly rationalizable.*

5.2 Multistage mechanism: implementation

Now, we propose a mechanism that satisfies sequentially weakly robust efficiency for a network in which the agents can be partitioned into two sets, with all agents in one set having knowledge of all agents in the other set.

Assume that a network $\mathbf{G}$ is such that the agents can be partitioned into two sets, L_1 and L_2, such that for any $i \in L_1$ and any $j \in L_2$, $g_{ij} = 1$. Without loss of generality, let $L_1 = \{1, \cdots, k\}$, and $L_2 = \{k+1, \cdots, n\}$. If there are several such partitions, take arbitrarily one of them. The numbering of agents can be arbitrary as well.

Multistage mechanism (MS mechanism)
The mechanism is composed by potentially n stages.

At each stage $i \in \{1, \cdots, k\}$, agent i sends a message on the valuations of himself and his neighbors belonging to the other set than his. Let $\overline{L}_l(\mathbf{m}) = \{i \in L_1 \,|\, m_i^i = argmax_{i' \in L_{l'} \cup \{i\}} m_i^{i'}\}$ for $l \in \{1, 2\}$ and $l' \neq l$.

- After k stages, if $\overline{L}_1(\mathbf{m}) = \emptyset$, then $x_{j^*} = 1$ for $j^* \in L_1$ such that $j^* = argmax_{j \in L_2 \cup \{1\}} m_1^j$, and stop the mechanism.
- Otherwise, go to the stage $k+1$.

- If $\overline{L}_2(\mathbf{m}) = \emptyset$, then $x_{i^*} = 1$ for $i^* \in L_2$ such that $i^* = argmax_{i \in L_1 \cup \{k+1\}} m^i_{k+1}$.
- After the stage n, define a bijection $\rho_{m_1} : \overline{L}_2(\mathbf{m}) \to \{1, \cdots, |\overline{L}_2(\mathbf{m})|\}$ which represents the ordering of agents in $\overline{L}_2(\mathbf{m})$ in the message of agent 1, that is, $\rho_{m_1}(j) < \rho_{m_1}(j')$ if and only if $m^j_1 > m^{j'}_1$.
- The allocation is $x_{i^*} = \left(\frac{\alpha}{1+\alpha}\right)^{|\overline{L}_2(\mathbf{m})|}$ for $i^* \in L_1$ such that $i^* = argmax_{i \in L_1} m^i_{k+1}$, $x_j = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(j)}$ for all $j \in \overline{L}_2(\mathbf{m})$, and $x_i = 0$ for all other agents.

Proposition 2. *The MS mechanism is sequentially weakly robustly efficient. Therefore, a sequentially weakly robustly efficient mechanism exists if the network is such that the agents can be partitioned into 2 sets where all agents in one set are connected to all agents in the other set.*

The MS mechanism involves selecting an agent from each set, L_1 and L_2, to assess the valuations of agents in the other set. This assessment functions similarly to the message of the center in the WE mechanism described in Section 4. In the WE mechanism, the center provides information about all other agents without incentive to misreport, as the agent ranked highest in the center's message receives a larger share of the good without affecting the center's own allocation. In the MS mechanism, an agent in L_1 similarly lacks incentive to misreport valuations of agents in L_2, and vice versa. This is because each agent's message only influences how the good is distributed among agents in the opposite set, leaving their own allocation unaffected. Through this setup, any misrepresentation of valuations across sets is effectively weakly eliminated.

Once all agents in L_2 observe the message from agent 1, they learn their own ranking as well as the ranking of others within L_2. They can trust that this ranking is truthful because any untruthful messages from agent 1 regarding the ranking of agents in L_2 were eliminated . Consequently, agents in L_2 are certain that anyone ranked lower genuinely has a lower valuation. Using this knowledge, we construct a mechanism such that if an agent in L_2 changes his message, it affects only the allocation of agents he observes and those ranked lower than himself. This setup incentivizes him to be truthful if all agents ranked below him are also truthful. Since the lowest-ranked agent in L_2 affects only the agents he observes,[20] we can, by applying the same method as in Lemma 1, ensure that this agent has an incentive to report truthfully. This reasoning then extends inductively: the agent ranked second-lowest knows that the lowest-ranked agent is truthful, which incentivizes him to be truthful as well, and so forth up the ranking. Therefore, this implies that if an agent in L_1 were to falsely claim that his valuation is higher than all agents in L_2 when it is not, at least one agent in L_2 would be motivated to contest this. As a result, all agents in L_1 are motivated to report truthfully. This is because, if an agent in L_2 has a higher valuation than his, then agents in L_1 would prefer that this high-value

[20]This is because there is no one who is ranked lower than him.

agent receives the entire good, rather than receiving an allocation which shares $\frac{\alpha}{1+\alpha}$ to L_1 and L_2 while the remainder is wasted. This final allocation, which results from any inconsistency, prevents any agent in L_1 from telling the wrong information.

Example 4. Let us take a network of 4 agents in Figure 2 which forms a rectangle. We apply the MS mechanism to this network to illustrate the incentive scheme of the mechanism. In this network, the partition is $L_1 = \{1, 2\}$ and $L_2 = \{3, 4\}$. Let us denote

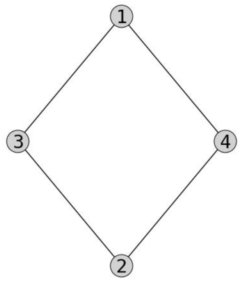

Figure 2: $L_1 = \{1, 2\}$, $L_2 = \{3, 4\}$

$\beta = \frac{\alpha}{1+\alpha}$ and let us assume first that the true valuation profile is such that $v_1 > v_2$ and $v_3 > v_4$. Each panel of Table 3 illustrates the allocations for each message of 3 and 4, when both agents 1 and 2 claim that they have a higher valuation than 3 and 4. Panel (a) shows the allocations when agent 1 announces that agent 3 has a higher valuation than agent 4, and Panel (b) the contrary. By comparing Panel (a) and (b), we can see that agent 1 is weakly better off by choosing the message of Panel (a) than choosing Panel (b) and is strictly better of when agent 3 chooses one of the 2 bottom rows and when agent 4 chooses the right column, i.e. when both agents claim that their valuation is higher than the one who is ranked the highest among $\{1, 2\}$ by agent 3. Therefore, the message of agent 1 in Panel (b) is eliminated.

Then, we can observe that the message of agent 4, who is ranked lower in agent 1's ranking, only impacts the allocation of himself and the agent in L_1 who is ranked the highest by agent 3, as agent 4 knows this due to the network structure. By Lemma 1, we already know that sharing the good at the rate of $\beta = \frac{\alpha}{1+\alpha}$ between two agents when there is a contradiction between two agents who know each other incentivizes the lower-ranked agent to tell the truth. This reasoning applies to this situation as well. In the valuation profile we assume, where $v_1 > v_4$, agent 4 has an incentive to tell the truth and thus chooses the message in the left column. On the other hand, if $v_4 > v_1$, agent 4 will choose the message in the right column.

Once the untruthful message of agent 4 is eliminated, we can also eliminate the untruthful message of agent 3. When agent 4 chooses the left message, this is a direct consequence of Lemma 1, since the message of agent 3 only alters the allocation of those whom agent 3 knows. When agent 4 chooses the right message, the situation is more

		m_4	
		$m_4^{i^*} > m_4^4$	$m_4^4 > m_4^{i^*}$
m_3	$m_3^1 > m_3^2 > m_3^3$	$(1,0,0,0)$	$(\beta,0,0,\beta)$
	$m_3^1 > m_3^3 > m_3^2$	$(1,0,0,0)$	$(\beta,0,0,\beta)$
	$m_3^2 > m_3^1 > m_3^3$	$(0,1,0,0)$	$(0,\beta,0,\beta)$
	$m_3^2 > m_3^3 > m_3^1$	$(0,1,0,0)$	$(0,\beta,0,\beta)$
	$m_3^3 > m_3^1 > m_3^2$	$(\beta,0,\beta,0)$	$(\beta^2,0,\beta,\beta^2)$
	$m_3^3 > m_3^2 > m_3^1$	$(0,\beta,\beta,0)$	$(0,\beta^2,\beta,\beta^2)$

(a) $m_1 : m_1^1 > m_1^3 > m_1^4$

		m_4	
		$m_4^{i^*} > m_4^4$	$m_4^4 > m_4^{i^*}$
m_3	$m_3^1 > m_3^2 > m_3^3$	$(1,0,0,0)$	$(\beta,0,0,\beta)$
	$m_3^1 > m_3^3 > m_3^2$	$(1,0,0,0)$	$(\beta,0,0,\beta)$
	$m_3^2 > m_3^1 > m_3^3$	$(0,1,0,0)$	$(0,\beta,0,\beta)$
	$m_3^2 > m_3^3 > m_3^1$	$(0,1,0,0)$	$(0,\beta,0,\beta)$
	$m_3^3 > m_3^1 > m_3^2$	$(\beta,0,\beta,0)$	$(\beta^2,0,\beta^2,\beta)$
	$m_3^3 > m_3^2 > m_3^1$	$(0,\beta,\beta,0)$	$(0,\beta^2,\beta^2,\beta)$

(b) $m_1 : m_1^1 > m_1^4 > m_1^3$

Table 3: Comparative tables for different agent preferences

complex because the message of agent 3 affects the allocation to agent 4, whom he is not directly connected to. This could lead agent 3 to believe that agent 4 has the highest valuation, which might cause agent 3 to choose a message different from the two bottom ones, even though agent 3 observes that his valuation is higher than those of agents 1 and 2, as this could reduce the allocation to agent 4. However, it is important to note that agent 3 knows that agent 4 has a lower valuation than he does, due to the message from agent 1. Therefore, if agent 4 chooses the right message, we can show that the allocation $(\beta^2, 0, \beta, \beta^2)$ is strictly preferred to $(\beta, 0, 0, \beta)$ by agent 3. This preference arises because agent 3 knows that $v_3 > v_1$ and can infer from agent 1's message that $v_3 > v_4$. Thus, agent 3 has an incentive to truthfully report, further reinforcing the mechanism's robustness.

Now, we know that if agent 1 claims that he is higher than both 3 and 4 although it is not true, then at least one of them would contest him so that he can at most get the allocation which shares β with one of them. Assume that $v_3 > v_1 > v_4$. If agent 1 chooses the message of Panel (a), then he gets $(\beta, 0, \beta, 0)$, since agent 3 will tell the truth. By Lemma 1, we know that agent 1 strictly prefers $(0, 0, 1, 0)$ to $(\beta, 0\beta, 0)$. Therefore, agent 1 would not claim that he is higher than both 3 and 4, but let agent 3 obtain the good by telling that agent 3 is the highest among everyone that he knows.

Now, we know that if agent 1 claims that he is ranked higher than both agents 3 and 4, although it is not true, then at least one of them will contest this claim, meaning agent 1 can, at most, get the allocation that shares β with one of them. Assume that $v_3 > v_1 > v_4$. If agent 1 chooses the message in Panel (a), then he will receive the allocation $(\beta, 0, \beta, 0)$, because agent 3 will truthfully report. By Lemma 1, we know that agent 1 strictly prefers the allocation $(0, 0, 1, 0)$ to $(\beta, 0, \beta, 0)$. Therefore, agent 1 would not claim that he is ranked higher than both agents 3 and 4. Instead, he would let agent 3 receive the good by truthfully reporting that agent 3 is the highest among the agents he knows. This ensures that agent 1 receives a better outcome than if he falsely claims to be ranked higher than agents 3 and 4, further incentivizing truthful behavior.

6 Conclusion

This paper analyzes robust mechanism design with positive externalities in networks where agents know the valuations of their neighbors. Instead of following the standard approach, which restricts agents' beliefs by assuming common knowledge, we investigate network structures that allow the principal to identify the agent with the highest valuation through a mechanism, regardless of the beliefs agents may hold. This approach offers insights into how the network of an organization should be structured. For example, if a team manager has the ability to create links between team members, such as through a buddy program or by controlling desk placements, the network structures proposed in this paper are the ones the manager should aim for in order to identify the most productive member when needed.

More broadly, our framework highlights the importance of considering the asymmetry of information that each agent possesses about others, in order to prevent dishonesty when agents are asked to report their information. The framework shows that high concentration of information among a few agents may actually facilitate the elicitation of truthful information.

References

Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000.

Pierpaolo Battigalli and Marciano Siniscalchi. Rationalization and incomplete information. *Advances in Theoretical Economics*, 3(1), 2003.

Leonie Baumann. Robust implementation with peer mechanisms and evidence. *Available at SSRN 2873888*, 2023.

Dirk Bergemann and Stephen Morris. Robust mechanism design. *Econometrica*, pages 1771–1813, 2005.

Dirk Bergemann and Stephen Morris. Ex post implementation. *Games and Economic Behavior*, 63(2):527–566, 2008.

Dirk Bergemann and Stephen Morris. Robust implementation in direct mechanisms. *The Review of Economic Studies*, 76(4):1175–1204, 2009.

Dirk Bergemann and Stephen Morris. Robust implementation in general mechanisms. *Games and Economic Behavior*, 71(2):261–281, 2011.

Dhruva Bhaskar and Evan Sadler. Resource allocation with positive externalities. *Available at SSRN 2853085*, 2019.

Francis Bloch and Matthew Olckers. Friend-based ranking. *American Economic Journal: Microeconomics*, 14(2):176–214, 2022.

Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.

Jacques Crémer and Richard P McLean. Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53(2): 345–361, 1985.

Dirk Engelmann and Martin Strobel. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, 94(4):857–869, 2004.

Tangren Feng, Axel Niemeyer, and Qinggong Wu. The limits of ex post implementation without transfers. *Theoretical Economics*, 18(2):463–479, 2023.

Susanne Goldlücke and Thomas Tröger. The multiple-volunteers principle. *Available at SSRN: https://ssrn.com/abstract=3753985*, 2020.

John C Harsanyi. Games with incomplete information played by "bayesian" players, i–iii. *Management science*, 14(3):159–182, 1967.

Mathew O Jackson. Bayesian implementation. *Econometrica*, pages 461–477, 1991.

Philippe Jehiel and Benny Moldovanu. Efficient design with interdependent valuations. *Econometrica*, 69(5):1237–1259, 2001.

Mariann Ollár and Antonio Penta. Full implementation and belief restrictions. *American Economic Review*, 107(8):2243–2277, 2017.

Thomas R Palfrey and Sanjay Srivastava. Mechanism design with incomplete information: A solution to the implementation problem. *Journal of Political Economy*, 97(3):668–691, 1989.

Appendix: Proofs

Proof of Lemma 1: w.l.o.g, let $v_1 > v_2$. Let m_i be such that $m_i^1 > m_i^2$ and m_i' be such that $m_i^{1\prime} < m_i^{2\prime}$ for $i \in \{1, 2\}$. We will check that for agent 1, m_1' is strictly dominated. We have

$$\pi_1(m_1, m_2) = v_1$$
$$\pi_1(m_1', m_2) = \frac{\alpha}{1+\alpha}v_1 + \frac{\alpha^2}{1+\alpha}v_2 < v_1 \text{ if } v_1 > v_2$$

Moreover,

$$\pi_1(m_1, m_2') = \frac{\alpha}{1+\alpha}v_1 + \frac{\alpha^2}{1+\alpha}v_2$$
$$\pi_1(m_1', m_2') = \alpha v_2 < \frac{\alpha}{1+\alpha}v_1 + \frac{\alpha^2}{1+\alpha}v_2 \text{ if and only if } v_1 > v_2$$

Therefore, m_1' is strictly dominated.
Once m_1' is eliminated, we can check that for agent 2 m_2' is strictly dominated. We have

$$\pi_1(m_1, m_2) = \alpha v_1$$
$$\pi_1(m_1, m_2') = \frac{\alpha}{1+\alpha}v_1 + \frac{\alpha^2}{1+\alpha}v_2 < \alpha v_1 \text{ if and only if } v_1 > v_2$$

Hence, m_2' is strictly dominated. $\square$

Proof of Theorem 1: *Necessity*
We prove it by contradiction. Assume that $\mathbf{x}(\cdot)$ is robustly efficient, and assume that there is only one agent who is connected to everyone. Without loss of generality, let agent 1 be such agent.
Consider two valuation profiles $\mathbf{v}^*$ and $\mathbf{v}^{**}$, with corresponding type profiles θ^* and θ^{**}, such that for some agent i, $v_i^* > v_1^*$, $v_i^{**} < v_1^{**}$, and $v_{i'}^{**} = v_{i'}^*$ for all $i' \neq i$.
We prove that for all j, there exists a valuation profile $\hat{\mathbf{v}}^j$ such that $\hat{\theta}_j^j = \theta_j^{**}$, and $\hat{\theta}_j^j \in S_j(\theta_j^*)$, hence $\theta_j^{**} \in S_j(\theta_j^*)$.

Assume by contradiction that there exists an agent $j^* \neq 1$ such that for all $\hat{\mathbf{v}}^{j^*}$ such that $\hat{\theta}_{j^*}^{j^*} = \theta_{j^*}^{**}$ and $\hat{\theta}_{j^*}^{j^*} \notin S_{j^*}(\theta_{j^*}^*)$. It implies that there exists a largest k such that $\hat{\theta}_l^{j^*} \in S_l^k(\theta_l^*)$ for all $l \in N$, and let $\hat{k}$ be such k. Thus, there exists an agent j such that for any $\mu_j \in \Delta(M_{-j} \times V_{-j}) \cap C_j(B_j(\theta_j^*))$ such that

$$\mu_j(m_{-j}, v_{-j}) > 0 \Rightarrow m_l \in S_l^{\hat{k}}(\theta_l) \text{ for all } l \neq j \text{ and for } \theta_l \text{ corresponding to } \mathbf{v}$$

there exists m_j^* such that

$$\int_{M_{-j}\times V_{-j}} u_j(\mathbf{x}(m_j^*, m_{-j}), (v_j^*, v_{-j}))d\mu_j > \int_{M_{-j}\times V_{-j}} u_j(\mathbf{x}(\hat{\theta}_j^{j^*}, m_{-j}), (v_j^*, v_{-j}))d\mu_j$$

This remains true by taking μ_j which puts probability 1 to $m_{-j} = \hat{\theta}^{j^*}_{-j}$. Hence, for any $\psi_j \in \Delta(V_{-j}) \cap B_j(\theta^*_j)$ such that

$$\psi_j(v_{-j}) > 0 \Rightarrow \hat{\theta}^{j^*}_l \in S^{\hat{k}}_l(\theta_l) \text{ for all } l \neq j \text{ and for } \theta_l \text{ corresponding to } \mathbf{v} \tag{3}$$

there exists m^*_j such that

$$\int_{V_{-j}} u_j(\mathbf{x}(m^*_j, \hat{\theta}^{j^*}_{-j}), (v^*_j, v_{-j}))d\psi_j > \int_{V_{-j}} u_j(\mathbf{x}(\hat{\theta}^{j^*}_j, \hat{\theta}^{j^*}_{-j}), (v^*_j, v_{-j}))d\psi_j$$

Take $\psi^*_j \in \Delta(V_{-j}) \cap B_j(\theta^*_j)$ such that $\psi^*_j(v^*_{-j}) = 1$. Then, we have

$$\int_{V_{-j}} u_j(\mathbf{x}(m^*_j, \hat{\theta}^{j^*}_{-j}), (v^*_j, v_{-j}))d\psi^*_j > \int_{V_{-j}} u_j(\mathbf{x}(\hat{\theta}^{j^*}_j, \hat{\theta}^{j^*}_{-j}), (v^*_j, v_{-j}))d\psi^*_j$$

$$u_j(\mathbf{x}(m^*_j, \hat{\theta}^{j^*}_{-j}), (v^*_j, v^*_{-j})) > u_j(\mathbf{x}(\hat{\theta}^{j^*}_j, \hat{\theta}^{j^*}_{-j}), (v^*_j, v^*_{-j})) \tag{4}$$

Let us denote $\hat{\mathbf{x}} = \mathbf{x}(m^*_j, \hat{\theta}^{j^*}_{-j})$ and $\hat{\mathbf{x}}^* = \mathbf{x}(\hat{\theta}^{j^*}_j, \hat{\theta}^{j^*}_{-j})$. Let $\hat{\mathbf{v}}^{j^*}$ be such that $\hat{v}^{j^*}_{j'} = max_{i' \in N}\, \hat{v}^{j^*}_{i'}$ and $v^*_{j'} - \hat{v}^{j^*}_{j'} \geq v^*_i - \hat{v}^{j^*}_i$ for some agent j' with $g_{j'j^*} = 0$, and $\hat{v}^{j^*}_l = v^{**}_l$ for all $l \neq j'$. Since the mechanism is robustly efficient, we have

$$\hat{x}^*_l = \begin{cases} 1, & \text{for } l = j' \\ 0, & \text{otherwise} \end{cases}$$

Therefore, we have

$$u_j(\hat{\mathbf{x}}, (v^*_j, v^*_{-j})) = v^*_j \hat{x}_j + \alpha \sum_{l \neq j} v^*_l \hat{x}_l$$

$$u_j(\hat{\mathbf{x}}^*, (v^*_j, v^*_{-j})) = \alpha v^*_{j'}$$

and then from (4), we have

$$v^*_j \hat{x}_j + \alpha \sum_{l \neq i,j,j'} v^*_l \hat{x}_l + \alpha v^*_i \hat{x}_i + \alpha v^*_{j'} \hat{x}_{j'} - \alpha v^*_{j'} > 0 \tag{5}$$

Besides, since the mechanism is robustly efficient, we have

$$u_j(\mathbf{x}(\hat{\theta}^{j^*}_j, \hat{\theta}^{j^*}_{-j}), (\hat{v}^{j^*}_j, \hat{v}^{j^*}_{-j})) \geq u_j(\mathbf{x}(m^*_j, \hat{\theta}^{j^*}_{-j}), (\hat{v}^{j^*}_j, \hat{v}^{j^*}_{-j}))$$

$$\hat{v}^{j^*}_j \hat{x}_j + \alpha \sum_{l \neq j,j'} \hat{v}^{j^*}_l \hat{x}_l + \alpha \hat{v}^{j^*}_{j'} \hat{x}_{j'} - \alpha \hat{v}^{j^*}_{j'} \leq 0$$

Since $\hat{v}^{j^*}_l = v^{**}_l$ for all $l \neq j'$, and $v^{**}_l = v^*_l$ for all $l \neq i$, we have

$$v^{**}_j \hat{x}_j + \alpha \sum_{l \neq j,j'} v^{**}_l \hat{x}_l + \alpha \hat{v}^{j^*}_{j'} \hat{x}_{j'} - \alpha \hat{v}^{j^*}_{j'} \leq 0$$

$$v^*_j \hat{x}_j + \alpha \sum_{l \neq i,j,j'} v^*_l \hat{x}_l + \alpha v^{**}_i \hat{x}_i + \alpha \hat{v}^{j^*}_{j'} \hat{x}_{j'} - \alpha \hat{v}^{j^*}_{j'} \leq 0 \tag{6}$$

Therefore, by (5) and (6), we have

$$\alpha[(v_i^* - v_i^{**})\hat{x}_i - (v_{j'}^* - \hat{v}_{j'}^{j^*})(1 - \hat{x}_{j'})] > 0$$

This cannot hold if $\hat{x}_{j'} = 1$. Since $\sum_{l \in N} \hat{x}_l \leq 1$, if $\hat{x}_{j'} = 1$, then $\hat{x}_i = 0$, so that

$$\alpha[(v_i^* - v_i^{**})\hat{x}_i - (v_{j'}^* - \hat{v}_{j'}^{j^*})(1 - \hat{x}_{j'})] = 0$$

Hence, assume $\hat{x}_{j'} < 1$. Then, since $v_{j'}^* - \hat{v}_{j'}^{j^*} \geq v_i^* - v_i^{**} > 0$, we have

$$\alpha\left[(v_i^* - v_i^{**})\hat{x}_i - (v_i^* - v_i^{**})(1 - \hat{x}_{j'})\right] \geq \alpha[(v_i^* - v_i^{**})\hat{x}_i - (v_{j'}^* - \hat{v}_{j'}^{j^*})(1 - \hat{x}_{j'})] > 0$$
$$\alpha(v_i^* - v_i^{**})(\hat{x}_i + \hat{x}_{j'} - 1) > 0$$

This is a contradiction since $\hat{x}_i + \hat{x}_{j'} \leq 0$ by the model.
If such j is i, then let $\hat{\mathbf{v}}^{j^*}$ be such that $\hat{v}_{i'}^{j^*} = max_{l \in N}\, \hat{v}_l^{j^*}$ and $\alpha(v_{i'}^* - \hat{v}_{i'}^{j^*}) > v_i^* - v_i^{**}$ for an agent i' with $g_{i'j^*} = 0$, and $\hat{v}_l^{j^*} = v_l^{**}$ for all $l \neq i'$. By the same argument, we obtain

$$(v_i^* - v_i^{**})\hat{x}_i - \alpha(v_{i'}^* - \hat{v}_{i'}^{j^*})(1 - \hat{x}_{i'}) > 0$$

Since $\alpha(v_{i'}^* - \hat{v}_{i'}^{j^*}) > v_i^* - v_i^{**} > 0$ by assumption, we have

$$\hat{x}_i(v_i^* - v_i^{**}) - (1 - \hat{x}_{i'})(v_i^* - v_i^{**}) > \hat{x}_i(v_i^* - v_i^{**}) - \alpha(1 - \hat{x}_{i'})(v_{i'}^* - \hat{v}_{i'}^{j^*}) > 0$$
$$(v_i^* - v_i^{**})(\hat{x}_{i'} + \hat{x}_i - 1) > 0$$

which is a contradiction since $\sum_{l \in N} \hat{x}_l \leq 1$.
If such j is 1, then let $\mathbf{v}^*$ be such that $v_1^* \geq \alpha v_i^*$. Moreover, let $\mathbf{v}^{**}$ be such that $v_1^{**} = max_{l \in N}\, v_l^{**}$ and $v_i^{**} = max_{l \neq 1}\, v_l^{**}$. Then, with the same argument, there exists m_1^* such that

$$u_1(\mathbf{x}(m_1, \theta_{-1}^{**}), (v_1^*, v_{-1}^*)) > u_1(\mathbf{x}(\theta_1^{**}, \theta_{-1}^{**}), (v_1^*, v_{-1}^*))$$
$$v_1^*\hat{x}_1 + \alpha \sum_{l \neq 1} v_l^* \hat{x}_l > v_1^*$$

Since $v_i^* = max_{l \neq 1}\, v_l^*$, we have

$$v_1^*\hat{x}_1 + \alpha v_i^*(1 - \hat{x}_1) > v_1^*\hat{x}_1 + \alpha \sum_{l \neq 1} v_l^* \hat{x}_l > v_1^*$$

$$\alpha v_i^* > v_1^*$$

which contradicts the assumption that $v_1^* \geq \alpha v_i^*$.
Hence, we proved that for all $j \in N$, there exists $\hat{\mathbf{v}}^j$ such that $\hat{\theta}_j^j = \theta_j^{**}$, and $\hat{\theta}_j^j \in S_j(\theta_j^*)$, thus $\theta_j^{**} \in S_j(\theta_j^*)$. This implies that for all $j \neq 1$, if we assume that $\theta_j^{**} \in S_j(\theta_j^{**})$, then it must be that $\theta_j^{**} \in S_j(\theta_j^*)$.

Now, suppose that the true valuation profile is $\mathbf{v}^*$ such that $v_i^* = max_{j\in N}\, v_j^*$, $v_1^* = max_{j\neq i}\, v_j^*$, and $v_1^* > \alpha v_j^*$ for any j. We know that there exists $\mathbf{v}^{**}$ such that $v_1^{**} = max_{j\in N}\, v_j^{**}$, and $\theta_j^{**} \in S_j(\theta_j^*)$ if $\theta_j^{**} \in S(\theta_j^{**})$ for all $j \neq 1$. Since the mechanism is robustly efficient, we have $\theta_j^{**} \in S_j(\theta_j^{**})$, and hence $\theta_j^{**} \in S_j(\theta_j^*)$. If this is the case, then $\theta_1^{**} \in S_1(\theta_1^*)$ since $\mathbf{x}(\theta^{**})$ is such that $x_1 = 1$ and $x_j = 0$ for any $j \neq 1$, and thus $\theta^{**} \in S(\theta^*)$, and then $\mathbf{x}(\theta^{**})$ must be efficient in $\mathbf{v}^*$. However, $\mathbf{x}(\theta^{**})$ is such that $x_1 = 1$ and $x_j = 0$ for any $j \neq 1$ if the mechanism is robustly efficient, but this allocation is not efficient in $\mathbf{v}^*$ since $v_i^* = max_{j\in N}\, v_j^*$.

Sufficiency

Without loss of generality, let $C \subseteq N$ be the set of agents who are connected to everyone. Choose randomly 2 agents from C (w.l.o.g let agents 1 and 2 be such agents). Let the mechanism $\mathbf{x}(\cdot)$ be the following:

- If $max_{j\in N}\, m_1^j = m_1^{j_1} \neq m_1^1$ and $max_{j\in N}\, m_2^j = m_2^{j_2} \neq m_2^2$, then $x_{j_1} = x_{j_2} = \frac{1}{2}$. If $j_1 = j_2 = j^*$, then $x_{j^*} = 1$.
- If $max_{j\in N}\, m_1^j = m_1^1$ and $max_{j\in N}\, m_2^j = m_2^2$, then $x_1 = x_2 = \frac{\alpha}{1+\alpha}$ and $x_i = 0$ for all other i.
- Otherwise, w.l.o.g. assume that $max_{j\in N}\, m_1^j = m_1^{j_1} \neq m_1^1$ and $max_{j\in N}\, m_2^j = m_2^2$. Then, $x_{j_1} = m_1^2/m_1^{j_1} + \epsilon$ and $x_i = 0$ for all other i.

We prove that this mechanism is robustly efficient. Without loss of generality, assume that $v_1 > v_2$, and $i^* = argmax_{i\in N}\, v_i$. First, assume that $i^* \neq 1$.
We prove first that m_1 such that $m_1^2 = max_{j\in N}\, m_1^j$ is never a best response. We prove that by considering the best responses of agent 1 to each strategy of agent 2.
First, let m_2 be such that $m_2^2 = max_{j\in n} m_2^j$. In this case, the best response of agent 1 is either $m_1^1 = max_{j\in N} m_1^i$ and obtain the allocation $x_1 = x_2 = \frac{\alpha}{1+\alpha}$, or $m_1^{i^*} = max_{j\in N}\, m_1^j$ and obtain the allocation profile $x_{i^*} = m_1^2/m_1^{i^*} + \epsilon$ and $x_i = 0$ for all other $i \neq j_1$. The first case is the best response if $v_{i^*} \leq \frac{v_1+\alpha v_2}{1+\alpha}$, and this can be proved by Lemma 1. The second case is the best response if $v_{i^*} > \frac{v_1+\alpha v_2}{1+\alpha}$. We can see it by checking the resulting payoff of agent 1 by taking this strategy, which is

$$u_1(\mathbf{x}(\mathbf{m})) = \alpha x_{i^*} v_{j_1}$$

and there exists $x_{i^*} = m_1^2/m_1^{i^*} + \epsilon < 1$ such that

$$\alpha x_{i^*} v_{i^*} > \frac{\alpha}{1+\alpha}\,(v_1 + \alpha v_2)$$

where the RHS is the payoff when agent 1 plays the first case.

Next, let m_2 be such that $m_2^1 = max_{j\in N} m_2^j$. Then, the best responses of agent 1 is either $m_1^1 = max_{j\in N}\, m_1^i$ and obtain the allocation $x_1 = 1$, or $m_1^{i^*} = max_{j\in n}\, m_1^j$ and obtain the allocation profile $x_1 = x_{i^*} = \frac{1}{2}$ if $\alpha v_{i^*} > v_1$.

Then, let m_2 be such that $m_2^i = max_{j \in n} m_2^j$ for some $i \notin \{1, 2\}$. Then, the best response of agent 1 is $m_1^{i^*} = max_{j \in n}\, m_1^j$ and obtain the allocation profile $x_i = x_{i^*} = \frac{1}{2}$ or $x_{i^*} = 1$ if $i = i^*$.

Once m_1 such that $m_1^2 = max_{j \in N}\, m_1^j$ is eliminated, the strategy for agent 2 such that $m_2^2 = max_{j \in N}\, m_2^j$ is never a best response and $m_2^{i^*} = max_{j \in N}\, m_2^j$ strictly dominates any other strategy.
If $m_1^1 = max_{j \in N}\, m_1^j$, then there exists a strategy m_2 such that $x_{i^*} = m_2^1 / m_2^{i^*} + \epsilon$ and

$$\alpha x_{i^*} v_{i^*} > \alpha v_1$$

where the RHS is the payoff of agent 2 when he plays m_2 such that $m_2^1 = max_{j \in N}\, m_2^j$.
Note that by Lemma 1, we have

$$\alpha v_1 > \frac{\alpha}{1 + \alpha} \left(\alpha v_1 + v_2 \right)$$

where the RHS is the payoff when agent 2 plays $m_2^2 = max_{j \in N}\, m_2^j$.
The fact that $m_2^{i^*} = max_{j \in N}\, m_2^j$ strictly dominates any other strategy for agent 2 yields that for agent 1 $m_1^{i^*} = max_{j \in N}\, m_1^j$ strictly dominates any other strategy.

If $1 = i^*$, same argument holds. $\square$

Proof of Proposition 1: We prove the statement by a series of lemmata.

Lemma 2. *Assume that $v_1 \neq max_{i \in N}\, v_i$. Then, m_1 such that $m_1^1 = max_{i \in N} m_1^i$ is not in $S_1^1(\theta_1)$. Moreover, if $v_1 = max_{i \in N}\, v_i$, then m_1 such that $m_1^1 = max_{i \in N} m_1^i$ and $m_1^{i^*} = max_{i \neq 1} m_1^i$ are the only messages in $S_1^1(\theta_1)$.*

Proof. We first prove the first part. It is sufficient to prove that such m_1 is weakly dominated by some other message if $v_1 \neq max_{i \in N}\, v_i$. Assume that $v_{i_{max}} = max_{i \in N} v_i$, and let us take a message $\hat{m}_1$ such that $max_{i \in N}\, \hat{m}_1^i = m_1^{i_{max}}$. We prove that $\hat{m}_1$ weakly dominates m_1.
Assume first that m_{-1} is such that $m_i^1 > m_i^i$ for all i. Then, $\mathbf{x}(m_1, m_{-1}) = \mathbf{x}(\hat{m}_1, m_{-1})$, the allocation is the same irrespective of the message of agent 1.
Otherwise, $\mathbf{x}(m_1, m_{-1})$ is

$$x_i(m_1, m_{-1}) = \begin{cases} \frac{\alpha}{1+\alpha}, & \text{for } i \in \{1, i^*\} \\ 0, & \text{otherwise} \end{cases}$$

and $\mathbf{x}(\hat{m}_1, m_{-1})$ is

$$x_i(\hat{m}_1, m_{-1}) = \begin{cases} 1, & \text{for } i = i_{max} \\ 0, & \text{otherwise} \end{cases}$$

By Lemma 1, $\pi_1(\mathbf{x}(\hat{m}_1, m_{-1})) > \pi_1(\mathbf{x}(m_1, m_{-1}))$.

We prove the second part. Let m_1 be a message as in the statement, and $\hat{m}_1$ be a message such that $\hat{m}_1^1 \neq max_{i \in N}\, \hat{m}_1^i$. We prove that m_1 weakly dominates $\hat{m}_1$.
If m_{-1} is such that $m_i^1 > m_i^i$ for all i, then with the same argument with the previous case, the allocation is the same irrespective of the message of agent 1. So assume otherwise. Then, $\mathbf{x}(m_1, m_{-1})$ is

$$x_i(m_1, m_{-1}) = \begin{cases} \frac{\alpha}{1+\alpha}, & \text{for } i \in \{1, i^*\} \\ 0, & \text{otherwise} \end{cases}$$

and $\mathbf{x}(\hat{m}_1, m_{-1})$ is

$$x_i(\hat{m}_1, m_{-1}) = \begin{cases} 1, & \text{for } i = i^* \\ 0, & \text{otherwise} \end{cases}$$

and by Lemma 1, $\pi_1(\mathbf{x}(m_1, m_{-1})) > \pi_1(\mathbf{x}(\hat{m}_1, m_{-1}))$ if $v_1 = max_{i \in N}\, v_i$. □

Lemma 3. *For all i such that $v_i > v_1$, m_i such that $m_i^1 > m_i^i$ is not in $S_i^2(\theta_i)$.*

Proof. Assume that there is an agent i with $v_i > v_1$. Then, by Lemma 2, he knows that m_1 such that $m_1^1 = max_{i \in N} m_1^i$ is not in $S_1^1(\theta_1)$, and only messages m_1 such that $m_1^{i^*} = max_{i \in N} m_1^i$ are in $S_1^1(\theta_1)$.
Let m_i and $\hat{m}_i$ be such that $m_i^1 > m_i^i$ and $\hat{m}_i^i > \hat{m}_i^i$. For m_i, we have

$$x_i(m_i, m_{-i}) = \begin{cases} 1, & \text{for } i = 1 \\ 0, & \text{otherwise} \end{cases}$$

if for all $j \notin \{1, i\}$, $m_j^1 > m_j^j$, and

$$x_i(m_i, m_{-i}) = \begin{cases} 1, & \text{for } i = i_{max} \\ 0, & \text{otherwise} \end{cases}$$

otherwise.
For $\hat{m}_i$, we always have

$$x_i(\hat{m}_i, m_{-i}) = \begin{cases} 1, & \text{for } i = i_{max} \\ 0, & \text{otherwise} \end{cases}$$

which weakly dominates m_i. □

Lemma 4. *For all i such that $v_1 > v_i$, the messages m_i such that $m_i^1 > m_i^i$ are the only messages in $S_i^3(\theta_i)$.*

Proof. Let m_i be as in the statement, and let $\hat{m}_i$ be a message such that $\hat{m}_i^1 > \hat{m}_i^i$. We prove that m_i weakly dominates $\hat{m}_i$ for any belief.
Assume first that agent i believes that there is an agent j such that $v_j > v_1$. Then, by Lemma 2, all $m_1 \in S_1^1(\theta_1)$ are such that $m_1^{i_{max}} = max_{i' \in N}\, m_1^{i'}$, and by Lemma 3 all $m_j \in S_j^2(\theta_j)$ are such that $m_j^j > m_j^1$. Hence, $\mathbf{x}(m_i, m_{-i})$ is such that $x_{i^{max}}(m_i, m_{-i}) = 1$ for any message of i.
Assume next that i believes that $v_1 = max_{i' \in N}\, v_{i'}$. Then, all messages $m_1 \in S_1^1(\theta_1)$ are such that $m_1^1 = max_{i' \in N} m_1^{i'}$. Hence, $\mathbf{x}(m_i, m_{-i})$ is such that

$$x_k(m_i, m_{-i}) = \begin{cases} 1, & \text{for } i = 1 \\ 0, & \text{otherwise} \end{cases}$$

if $m_j^1 > m_j^j$ for all $j \notin \{1, i\}$, and

$$x_k(m_i, m_{-i}) = \begin{cases} \frac{\alpha}{1+\alpha}, & \text{for } i \in \{1, i^*\} \\ 0, & \text{otherwise} \end{cases}$$

otherwise.
On the other hand, $\mathbf{x}(\hat{m}_i, m_{-i})$ is

$$x_i(\hat{m}_i, m_{-i}) = \begin{cases} \frac{\alpha}{1+\alpha}, & \text{for } i \in \{1, i^*\} \\ 0, & \text{otherwise} \end{cases}$$

We can see by Lemma 1 that $\mathbf{x}(m_i, m_{-i})$ yields a weakly higher payoff for i than $\hat{m}_i$ for any i^*, even $i^* = i$. □

By Lemmata 2, 3, and 4, for any $\mathbf{v}$, $\mathbf{x}(\mathbf{m})$ is efficient for any $\mathbf{m} \in T^3(\theta)$. □

Proof of Proposition 2: We prove with a series of lemmata.

Lemma 5. *For agent 1, a message such that $1 \in \overline{L}_1$ and $m_1^i < m_1^j$ whereas $v_i > v_j$ for some $i, j \in L_2$ is eliminated in the first round of weak elimination.*

Proof. We prove it for agent 1 first.
Take two messages m_1 and m_1^* with $1 = argmax_{i' \in L_2 \cup \{1\}}\, m_1^{i'} = argmax_{i' \in L_2 \cup \{1\}}\, m_1^{i'*}$ such that $m_1^i < m_1^j$ and $m_1^{i*} > m_1^{j*}$ for some $i, j \in L_2$ with $v_i > v_j$, and there is no $l \neq i, j$ and $l' \neq i, j$ such that $m_1^i < m_1^l < m_1^j$ and $m_1^{i*} > m_1^{l'*} > m_1^{j*}$. Moreover, assume that $m_1^{l_1} > m_1^{l_2}$ if and only if $m_1^{l_1\prime} > m_1^{l_2\prime}$ for any other pair $(l_1, l_2) \neq (i, j)$. We prove that m_1 is weakly dominated by m_1^*.
Assume first that either $i \notin \overline{L}_2$ or $j \notin \overline{L}_2$, then $\pi_1(m_1, m_{-1}) = \pi_1(m_1^*, m_{-1})$ for any m_{-1}.

So, assume $i, j \in \overline{L}_2$. Then, since $\rho_{m_1}(j) = \rho_{m_1}(i) - 1$, we have

$$x_i(m_1, m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)}$$
$$x_j(m_1, m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(j)} = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)-1}$$

Conversely, for m_1^*, we have $\rho_{m_1^*}(i) = \rho_{m_1}(j)$ and $\rho_{m_1^*}(j) = \rho_{m_1}(i)$, and thus

$$x_i(m_1^*, m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1^*}(i)} = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(j)} = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)-1}$$
$$x_j(m_1^*, m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1^*}(j)} = \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)}$$

and $x_l(m_1, m_{-1}) = x_l(m_1^*, m_{-1})$ for all other l. Hence, $\pi_1(m_1^*, m_{-1}) > \pi_1(m_1, m_{-1})$ since $v_i > v_j$. This implies that agent 1 is weakly better off by choosing m_1^* than choosing m_1, no matter what m_{-1}. Hence, m_1 is weakly eliminated in the first round of weak iterative eliminations.

Assume now that m_1 is such that there exists a sequence of agents $(l_1, \cdots, l_k)$ with $l_1 = i$ and $l_k = j$ such that $m_1^{l_1} < \cdots < m_1^{l_k}$ and no such agent l' as $m_1^{l_p} < m_1^{l'} < m_1^{l_{p+1}}$ for any $p \in \{1, \cdots, k-1\}$. If $v_i > v_j$, then necessarily there exists $q \in \{1, \cdots, k\}$ such that $v_{l_q} > v_{l_{q+1}}$. Therefore, we can apply the same argument to this pair. □

Lemma 6. *For agent $k+1$, a message such that $i_{max} \neq argmax_{i \in L_1} m_{k+1}^i$ where $i_{max} = argmax_{i \in L_1} v_i$ is eliminated in the first round of weak elimination.*

Proof. Obviously true. □

Lemma 7. *Let i^* be such that $i^* = argmax_{i \in L_1} m_{k+1}^i$. For all $j \in L_2$, $m_j \in T_j(\theta_j)$ for $m_j^{i^*} > m_j^j$ if and only if $v_{i^*} > v_j$.*

Proof. Let $\sigma_{m_1} : L_2 \to \{1, \cdots, |L_2|\}$ such that $\sigma_{m_1}(j) < \sigma_{m_1}(j')$ if and only if $m_1^j > m_1^{j'}$, i.e. a bijection which represents the ranking of all agents in L_2 according to m_1. We prove the lemma by induction, starting from agent j such that $\sigma_{m_1}(j) = 1$. For sake of simplicity, let us renumber all agents in L_2 according to $\sigma_{m_1}(j)$, i.e. we rename agent $j \in L_2$ as agent j' if $\sigma_{m_1}(j) = j'$.

(Agent 1) : Take a given message profile (m_1, m_{-1}) where m_1 is such that $m_1^1 > m_1^{i^*}$ and $m_{-1} \in T_{-1}^1(\theta_{-1})$. Let m_1' be such that $m_1^{1\prime} < m_1^{i^*\prime}$. We can easily compute that we have

$$x_1(m_1, m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|}$$
$$x_{i^*}(m_1, m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|}$$

and

$$x_1(m_1', m_{-1}) = 0$$
$$x_{i^*}(m_1', m_{-1}) = \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|-1}$$

and $x_i(m_1', m_{-1}) = x_i(m_1, m_{-1})$ for any other i. Hence, by Lemma 1, $\pi_1(m_1, m_{-1}) > \pi_1(m_1', m_{-1})$ if and only if $v_1 > v_{i^*}$.

(Agent j') : Assume that for all agents $j'' < j'$, the statement is true.
We consider first the case where $v_{j'} < v_{i^*}$. In this case, by Lemma 5, it must be that for all $j'' < j'$, we have $v_{j''} < v_{j'}$. Therefore, it must be that $m_{j''}^{j''} < m_{j''}^{i^*}$ for any $m_{j''} \in T_{j''}^k(\theta_{j''})$ after some rounds of eliminations, and let k be such smallest number of rounds. Let $(m_{j'}, m_{-j'})$ be a message profile with $m_{j'}^{j'} > m_{j'}^{i^*}$ and $m_{-j'} \in T_{-j'}^k(\theta_{-j'})$. Moreover, let $m_{j'}'$ be such that $m_{j'}^{j'} < m_{j'}^{i^*}$. By the mechanism, we have

$$x_{j'}(m_{j'}, m_{-j'}) = \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|}$$
$$x_{i^*}(m_{j'}, m_{-j'}) = \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|}$$

and

$$x_{j'}(m_{j'}, m_{-j'}) = 0$$
$$x_{i^*}(m_{j'}, m_{-j'}) = \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})-1|}$$

and $x_i(m_{j'}', m_{-j'}) = x_i(m_{j'}, m_{-j'})$ for any other i. Hence, by Lemma 1, $\pi_{j'}(m_{j'}, m_{-j'}) < \pi_{j'}(m_{j'}', m_{-j'})$ since $v_{j'} < v_{i^*}$.
Now assume that $v_{j'} > v_{i^*}$. Then, if j' believes that for all $j'' < j'$, $v_{j''} < v_{i^*}$, then by the same argument, we can prove that $\pi_{j'}(m_{j'}, m_{-j'}) > \pi_{j'}(m_{j'}', m_{-j'})$ if and only if $v_{j'} > v_{i^*}$, since agent j' believes that no $j'' < j'$ chooses $m_{j''}^{j''} > m_{j''}^{i^*}$. So, let j' believes otherwise, i.e. there is some $j'' < j'$ with $v_{j''} > v_{i^*}$. We define $J = \{j'' \in \overline{L}_2(\mathbf{m}) \,|\, j'' < j'\}$. By assumption, we know that $v_{j''} > v_{i^*}$. Let us take $\mathbf{m} = (m_{j'}, m_{-j'})$ and $\mathbf{m}' = (m_{j'}', m_{-j'})$ such that $m_{j'}^{j'} > m_{j'}^{i^*}$ and $m_{j'}^{j'}{}' < m_{j'}^{i^*}{}'$. Then, we have

$$x_i(\mathbf{m}) = \begin{cases} \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)}, & \text{for } i \in \overline{L}_2 \\ \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|}, & \text{for } i = i^* \\ 0, & \text{otherwise} \end{cases}$$

and

$$x_i(\mathbf{m}') = \begin{cases} \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)}, & \text{for } i \in \overline{L}_2 \setminus J \\ \left(\frac{\alpha}{1+\alpha}\right)^{\rho_{m_1}(i)-1}, & \text{for } i \in J \\ \left(\frac{\alpha}{1+\alpha}\right)^{|L_2(\mathbf{m})|-1}, & \text{for } i = i^* \\ 0, & \text{otherwise (including } i = j) \end{cases}$$

We can compute that $\pi_{j'}(\mathbf{m}) > \pi_{j'}(\mathbf{m}')$, since for all $j'' < j'$, $v_{j''} < v_{j'}$ and $v_{i^*} < v_{j'}$. Hence, the statement is proved for j' if for all $j'' < j'$ the statement is true. Together with the case for agent 1, the statement is proved. □