arXiv:2402.15169v2 [cs.GT] 20 Nov 2024

Platforms for Efficient and Incentive-Aware Collaboration

Nika Haghtalab, Mingda Qiao, and Kunhe Yang

University of California, Berkeley
{nika,mingda.qiao,kunheyang}@berkeley.edu

Abstract

Collaboration is crucial for reaching collective goals. However, its potential for effectiveness is often undermined by the strategic behavior of individual agents — a fact that is captured by a high Price of Stability (PoS) in recent literature [Blum et al., 2021a]. Implicit in the traditional PoS analysis is the assumption that agents have full knowledge of how their tasks relate to one another. We offer a new perspective on bringing about efficient collaboration across strategic agents using information design. Inspired by the increasingly important role collaboration plays in machine learning (such as platforms for collaborative federated learning and data cooperatives), we propose a framework in which the platform possesses more information about how the agents' tasks relate to each other than the agents themselves. Our results characterize how and to what degree such platforms can leverage their information advantage and steer strategic agents towards efficient collaboration.

Concretely, we consider collaboration networks in which each node represents a task type held by one agent, and each task benefits from contributions made in their inclusive neighborhood of tasks. This network structure is known to the agents and the platform. On the other hand, the real location of each agent in the network is known to the platform only — from the perspective of the agents, their location is determined by a uniformly random permutation. We employ the framework of private Bayesian persuasion and design two families of persuasive signaling schemes that the platform can use to guarantee a small total workload when agents follow the signal. The first family aims to achieve the minmax optimal approximation ratio compared to the total workload in the optimal collaboration, which is shown to be $\Theta(\sqrt{n})$ for unit-weight graphs, $\Theta(n^{\frac{2}{3}})$ for graphs with edge weights lower bounded by $\Omega(1)$, and $O(n^{\frac{3}{4}})$ for general weighted graphs. The second family ensures per-instance strict improvement in the total workload compared to scenarios with full information disclosure.

Contents

1 Introduction **4**
- 1.1 Our Contribution . . . 6
- 1.2 Related Works . . . 7
- 1.3 Structure of the paper . . . 8

2 Model **8**
- 2.1 Collaboration system . . . 8
- 2.2 Benchmarks and the price of stability . . . 10
- 2.3 Information design and signaling schemes . . . 10

3 Technical Overview and Structural Results **11**
- 3.1 Structural results . . . 12
- 3.2 Overview of upper bounds in unit-weight graphs . . . 13
- 3.3 Overview of upper bounds in weighted graphs . . . 16
- 3.4 Overview of lower bounds . . . 18

4 Discussion and Future Directions **19**

5 Upper Bounds for Unit-Weight Graphs **19**
- 5.1 Weights of cuts and induced sub-graphs . . . 20
- 5.2 General unit-weight graphs . . . 20
- 5.3 Matching the cost of full revelation . . . 22
- 5.4 Strict improvement upon full revelation . . . 24

6 Lower Bounds for Unit-Weight Graphs **27**
- 6.1 Lower bound for the double-star graph . . . 27
- 6.2 A more general lower bound via k-stars . . . 29

7 Upper Bounds for Weighted Graphs **30**
- 7.1 Upper bound for approximating $\mathsf{OPT}^{\mathsf{IR}}$ . . . 30
- 7.2 Strict improvement upon full revelation . . . 31

8 Lower Bounds for Weighted Graphs **32**

A Omitted Proofs from Section 2 **38**
- A.1 Alternative Benchmarks . . . 38
- A.2 Proof of Lemma 2.1 . . . 38

B Omitted proofs from Section 3.1 **39**
- B.1 Proof of Lemma 3.2 . . . 39
- B.2 Proof of Lemma 3.4 . . . 40

C Details for Section 4 **40**
- C.1 Conjectured impossibility of matching $\mathsf{OPT}^{\mathsf{stable}}$ using $O(1)$ signals. . . . 40
- C.2 Conjectured $\Omega(n^{3/4})$ lower bound . . . 42

D Omitted Proofs from Section 5 **42**
- D.1 Technical Lemmas in Section 5.1 . . . 42

D.2 Proof of Lemma 5.8 43

E Omitted Proofs from Section 6 **44**

F Omitted Proofs from Section 7 **47**
F.1 Proof of Lemma 7.1 47
F.2 Proof of Theorem 7.2 50
F.3 Proof of Theorem 7.3 51
F.4 Proof of Lemma 7.5 52
F.5 Proof of Theorem 7.4 52

G Omitted Proofs from Section 8 **54**
G.1 Proof of Theorem 8.1 54
G.2 Projection to a low-dimensional space 56
G.3 Choice of the test function 58
G.4 Verify the dual feasibility 60

1 Introduction

Collaboration is the cornerstone of modern achievements across various disciplines. Effective implementation of collaborative systems has substantially increased what can be accomplished by the limited capabilities of individual agents. For instance, the global collaboration between agencies and institutions under the Genome-Wide Association Studies (GWAS) has enabled the decoding of genetic foundations of diseases [Bergen and Petryshen, 2012]; collaborations across hundreds of mathematicians have led to resolving longstanding open problems [Steingart, 2012]; collaboratively maintained virus signature databases [AlienVault, 2012] have made technology safer; and collaborative federated learning [McMahan et al., 2017] has led to the training of models with superior performance on a large range of applications.

However, for every highly successful collaboration, there are many others that never came to fruition. This is in part due to the inherently strategic nature of participants in a collaboration that can be a significant barrier to the realization of optimal collaboration. On the one hand, achieving optimal collaboration has been shown to typically require levels of effort from some participants that surpass what is individually rational (the amount of work required when working alone) or what is considered stable (the amount of work deemed reasonable given others' contributions). On the other hand, stable collaboration systems where everyone is satisfied with their assigned effort, if exist, often suffer from significant inefficiency.

The above issues are evident even in a simple double-star network shown in Figure 1 where each node represents an agent and their task, and each agent's task benefits from both their own and neighboring agents' contributions[1]. For collaboration to be feasible, each agent's task requires at least one unit of contribution. Ideally, in the optimal collaboration that minimizes total workload, the two central nodes take on the entire workload by each contributing one unit of effort to support all their leaf neighbors. However, when incentives are factored in, both center nodes have the incentive to unilaterally reduce their contribution, as they have already received enough support from each other. In fact, in any stable collaboration in this network, the total workload inevitably scales linearly with the number of agents, which implies that the benefits gained from collaboration are merely marginal.

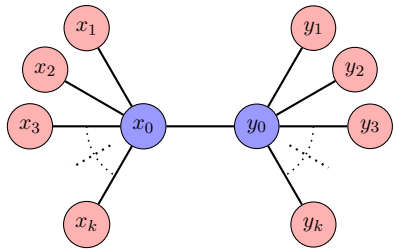

Figure 1: Double Star Graph. Formally, let θ_v denote the contribution of node v. The social optimum is achieved when $\theta_{x_0} = \theta_{y_0} = 1$ with all other $\theta_v = 0$. *Feasibility* requires that each agent v receives a total contribution of $\theta_v + \sum_{u:\{u,v\}\in E} \theta_u \geq 1$. *Stability* requires no node can unilaterally reduce their contribution without hurting their own feasibility. Following are typical stable solutions: (1) One-sided leaf contribution, where $\theta_v = 1$ for $v \in \{x_1, x_2, \ldots, x_k, y_0\}$ and $\theta_v = 0$ for the rest, and (2) Full leaf contribution, with $\theta_v = 1$ for all leaf nodes $v \in \{x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k\}$, while $\theta_{x_0} = \theta_{y_0} = 0$. Both solutions suffer from a total workload of $\Omega(k)$.

This example highlights the inefficiency of collaboration in the presence of strategic agents, which is quantitatively reflected by a high Price of Stability (PoS) [Anshelevich et al., 2008, Blum et al.,

[1] A variant of this structure was used by Blum et al. [2021a] for establishing a lower bound on the PoS in several abstractions of collaborative federated learning.

2021a], a well-established concept in game theory measuring the ratio between the total workload in the best stable solution and the socially optimal solution[2].

However, the traditional notion of PoS may be overly pessimistic for analyzing collaboration networks for two primary reasons. First, it assumes that agents have full knowledge of their positions and roles within the collaborative system. This overlooks real-life and practical uncertainties that often exist in collaborations centered on data gathering and usage; for instance, in data cooperatives and federated learning, participants are often uncertain of how their data contribute to the bigger picture, including their specific uses, their broader implications and value, and how they can help solve tasks for other participants. Collaborations are often hosted on platforms, yet, platforms lack the leverage to steer the agents toward more efficient collaborative outcomes in the traditional notion of PoS. As a result, the system is severely affected by the misalignment between the strategic agents' incentives and the platform's goal of improving the overall efficiency.

Inspired by the increasingly important role collaboration plays in machine learning, such as platforms for collaborative federated learning and data cooperatives, we revisit the problem of efficient and stable collaboration under a new framework. To address the aforementioned issues, it is crucial to recognize that while agents typically have some understanding of the types of tasks in the system and the general structure relating different task types, they often lack precise knowledge about how their specific tasks relate to others. This precise knowledge is usually exclusive to the platform, which highlights the platform's crucial role in communicating this knowledge to the agents. This level of uncertainty challenges the traditional PoS analysis that assumes complete knowledge of one's position and role. Yet, it also opens up new possibilities: by strategically distributing information about agents' task types, not only platforms can address these challenges but also they can create opportunities for efficient collaboration that was previously thought to be impossible.

Reflecting this into the aforementioned double-star structure in Figure 1, we now consider a scenario where agents are assigned to nodes of this network by a uniformly random permutation. That is, while the network structure is known to the agents, they are unaware of their exact positions within the graph. Not knowing whether one's task is represented by a central node or a leaf node significantly influences agents' decisions and collaboration outcomes. This is where the platform's role of strategically sharing information becomes clear. If the platform withholds all information, every agent will contribute constant effort because they face maximum and independent uncertainty about whether their task is sufficiently covered by contributions of others. This high level of uncertainty will lead agents to contribute an unnecessary amount of effort. On the other extreme, full information disclosure removes all uncertainty, leading central nodes to reduce their contributions (as discussed above) and pushing leaf nodes into a less efficient pattern of contribution, thus significantly increasing the PoS.

The crux of steering strategic agents into efficient collaboration lies in *creating the right level of correlated uncertainty*. By controlling the flow of information, the platform can negatively correlate the beliefs of central and leaf nodes, while maintaining a desirable level of uncertainty. The negative correlation between agents' beliefs ensures that agents would not perform unnecessary amount of work simultaneously. Moreover, the remaining uncertainty (as well as negative correlation) leverages the central agents' fear that their tasks might have not received sufficient contributions from others, thereby encouraging them to increase their effort.

This leads us to the main question of our study: Is it possible for the platform to always ensure

[2]In the original definition, PoS is the ratio between the cost of the best Nash equilibrium and that of the social optimum. In our collaborative context, the Nash equilibrium analog is a stable solution conditioned on being feasible.

efficient, stable collaboration among strategic agents by controlling the information flow about their types? Specifically, we seek to determine whether a general information structure can be designed to:

(1) Guarantee a sublinear[3] upper bound on the PoS for all collaboration instances;

(2) Consistently improve the PoS compared to a scenario of full information revelation.

To address these questions, our approach builds on the framework of the private Bayesian persuasion model in information design [Kamenica and Gentzkow, 2011, Arieli and Babichenko, 2019]. We allow the platform to credibly commit to a signaling scheme, which is a probabilistic mapping from the realization of joint types to private signals that are sent to each agent in the system. These private signals distill the information about joint types into direct recommendations on how much each agent should contribute. Key to this framework is the signaling scheme's *persuasiveness* – it guarantees that agents are incentivized to follow the recommendations. The platform facilitates efficient collaboration by committing to persuasive signaling schemes with low total workload when agents follow the recommendations.

1.1 Our Contribution

We initiate the study of improving the efficiency of stable collaboration through the lens of information design. We formalize the private Bayesian persuasion problem in a linear collaborative network and characterize the structural properties of persuasive signaling schemes. For a special class of *binary* signaling schemes, we identify how its persuasiveness relates to graph characteristics such as cuts and induced subgraphs.

Based on these structural insights, we design two families of signaling schemes that achieve the aforementioned goals (1) and (2), respectively. We will use n to denote the number of agents (nodes) in the network, and use benchmarks OPT, $\mathsf{OPT}^{\mathsf{IR}}$ and $\mathsf{OPT}^{\mathsf{stable}}$ to respectively quantify the total workload under the optimal feasible collaboration as well as the optimal feasible collaboration which also satisfies individual rationality (IR) and stability constraints.

Toward our first goal of achieving a nontrivial approximation ratio to OPT in the worst case, we design a binary signaling scheme that achieves the approximation ratio of $O(\sqrt{n})$ in any unit-weight graphs. In general weighted graphs, we argue that $\mathsf{OPT}^{\mathsf{IR}}$ serves as a more appropriate benchmark than OPT, because OPT might require some agents to contribute more than their individually rational amount, which rational agents cannot be persuaded to do. However, we show that binary signaling schemes fail to achieve this goal. To address this, we introduce a third signal to achieve the optimal approximation ratios of $O(n^{\frac{2}{3}})$ for graphs with lower bounded edge weights, and $O(n^{\frac{3}{4}})$ for general weighted graphs. Our results are summarized in Theorems 1.1 and 1.2.

Theorem 1.1. *In any unit-weight graph, there exists a binary signaling scheme that is persuasive and has cost $O(\sqrt{n} \cdot \mathsf{OPT})$. Moreover, the $O(\sqrt{n})$ approximation ratio is tight for certain graphs.*

Theorem 1.2. *In any weighted graph, there exists a ternary signaling scheme that is persuasive and has cost $O(n^{3/4} \cdot (\mathsf{OPT}^{\mathsf{IR}})^{1/2})$. In addition, for graphs with all edge weights bounded below by δ, the cost is improved to $O((n \cdot \mathsf{OPT}^{\mathsf{IR}})^{2/3} \cdot \delta^{-1/3})$, with the $O(n^{2/3})$ approximation ratio being tight on a family of graphs with $\delta = 1/2$.*

We introduce a different family of signaling schemes that achieve our second goal of strictly improving $\mathsf{OPT}^{\mathsf{stable}}$. This approach uses only two signals in unit-weight graphs and at most $n + 1$ signals

[3]Recall that worst-case PoS can be linear in the number of agents, as demonstrated by Figure 1

in general weighted graphs. We summarize the results in Theorems 1.3 and 1.4. Despite the substantial increase in the signal space required for weighted graphs, we demonstrate the infeasibility of achieving this goal through binary or ternary signaling schemes, and leave the identification of the minimum number of signals needed as an open problem discussed in Section 4.

Theorem 1.3. *In any unit-weight graph, whenever* $\mathsf{OPT} < \mathsf{OPT}^{\mathsf{stable}}$*, there exists a binary signaling scheme that is persuasive and has a cost strictly lower than* $\mathsf{OPT}^{\mathsf{stable}}$*.*

Theorem 1.4. *In any general weighted graph, whenever* $\mathsf{OPT}^{\mathsf{IR}} < \mathsf{OPT}^{\mathsf{stable}}$*, there exists a signaling scheme with at most* $n+1$ *signals that is persuasive and has a cost strictly lower than* $\mathsf{OPT}^{\mathsf{stable}}$*.*

1.2 Related Works

Collaborative federated learning. Our work connects to the broader topic of collaborative and federated learning among strategic agents, where incentives and fairness have been extensively explored [Donahue and Kleinberg, 2021a,b,c, Blum et al., 2021a]. Most related to us is the work of Blum et al. [2021a] that introduces a formalism for collaborative learning among network of agents and highlights the inefficiency in collaboration among strategic agents as captured by a high PoS. Our work introduces a novel approach to reducing the PoS under a more realistic scenario where agents face uncertainty regarding the types of their tasks.

Recent studies have adopted a *mechanism design* approach in federated learning, targeting objectives such as incentivizing data contribution and preventing free-riding [Huang et al., 2023, Karimireddy et al., 2022, Chen et al., 2023b, Xu et al., 2021], encouraging participation [Hu et al., 2023, Xu et al., 2023, Cohen and Shao, 2023, Han et al., 2023], and promoting fairness [Lin et al., 2023, Lyu et al., 2020]. See recent surveys [Zhan et al., 2021, Tu et al., 2022] for a more comprehensive overview. Additionally, research on federated bandits [Wei et al., 2023] explores strategies to motivate agents to share their data. However, these papers primarily rely on either monetary incentives or penalization through reduced model accuracy and usually operate under the condition that agents know how their data or tasks relate to those of others. In contrast, we assume that the platform possesses more information about the agents' types but without a direct control over the distribution of data, gradients, or models.

In addition, there is research focusing on incentives and mechanism design in *data markets* where consumers purchase data from the market [Agarwal et al., 2019, 2020, Jia et al., 2019], but they do not focus on incentivizing collaboration among participants.

Another line of literature focuses on the communication cost [Blum et al., 2021b] or sample efficiency of collaborative learning without incentives [Blum et al., 2017, Nguyen and Zakynthinou, 2018, Chen et al., 2018, Haghtalab et al., 2022, Awasthi et al., 2023, Zhang et al., 2023, Peng, 2023]. In contrast, we achieve *both efficiency and incentive-awareness* through the lens of information design.

Information design. Bayesian persuasion, a seminal work of Kamenica and Gentzkow [2011] rooted in the context of incomplete information games [Aumann et al., 1995], is a canonical model for information design that studies how an informed sender can strategically share information to influence the actions of a single uninformed receiver. This framework has been expanded to multi-receiver settings [Bergemann and Morris, 2016] and explored through algorithmic perspectives [Dughmi and Xu, 2016, Dughmi, 2017], supporting its applications in diverse problems. These applications include but are not limited to routing games [Das et al., 2017, Wu and Amin,

2019], coordination games [Wu and Amin, 2019], voting [Schnakenberg, 2015, Alonso and Câmara, 2016, Bardhi and Guo, 2018, Wang, 2013, Arieli and Babichenko, 2019], and incentivizing exploration [Mansour et al., 2020, 2022, Kremer et al., 2014]. We refer the readers to survey papers [Candogan, 2020, Kamenica, 2019] for a comprehensive overview.

The problem we study involves designing signaling schemes in a specific multi-receiver setting, where the literature offers two general approaches: characterizing and selecting *Bayes Correlated Equilibria (BCE)* based on sender utility [Bergemann and Morris, 2013, 2016, 2019, Taneva, 2019], and belief-based optimization [Mathevet et al., 2020]. However, applying these approaches directly in our context is computationally intractable due to the exponentially large state space. Indeed, finding optimal signal schemes is proved hard even in some simple multi-receiver settings [Bhaskar et al., 2016, Rubinstein, 2017, Dughmi, 2017], emphasizing the importance of leveraging our collaborative network's structural properties.

Our work also contributes to the recent line of work on *information design in networks and platforms*. In the context of social networks, Candogan and Drakopoulos [2017], Candogan [2019], Egorov and Sonin [2019] design optimal signaling schemes to maximize overall engagement, while Arieli and Babichenko [2019] studies the setting when the sender's utility is more general and non-additive. Mathevet and Taneva [2022] studies vertical and horizontal information transmission, identifying scenarios where *single-meeting schemes* — signals shared with and observed by a specific subset of agents — are optimal. Our analysis of signaling schemes in the collaboration network also focuses on optimality but diverges in its definition: we assess optimality not by the ability to implement all BCE, as previous studies do, but by achieving an optimal asymptotic approximation ratio compared to the socially optimal solution when agents are non-strategic. To the best of our knowledge, we are the first work that focuses on the perspective of multi-receiver Bayesian persuasion to enhance and encourage collaboration between strategic agents.

In the context of online platforms with strategic agents, Papanastasiou et al. [2018] studies information provision policy for sequentially arriving consumers who choose among alternative products or services, Bergemann et al. [2022], Chen et al. [2023a] focuses on maximizing revenues in click-through auctions when the platform has exclusive knowledge about the true click-through rates, among many other studies in this field. Our work complements the literature by studying the efficiency of collaboration platforms.

1.3 Structure of the paper

We formally introduce our model in Section 2. In Section 3, we discuss structural results that characterize the persuasiveness of signaling schemes, and provide a technical overview of our main results. In Section 4, we discuss a few open problems and future research directions. In Section 5, we focus on unit-weight graphs and prove the upper bounds in Theorems 1.1 and 1.3. We prove the lower bound side of Theorem 1.1 in Section 6. For weighted graphs, we provide the upper bounds (Theorems 1.2 and 1.4) in Section 7, and the lower bound (second half of Theorem 1.2) in Section 8.

2 Model

2.1 Collaboration system

Task-based contribution. Let V be the set of task types in the collaboration system, with $|V| = n$. Let $\theta_v \in \mathbb{R}_{\geq 0}$ be the effort contributed directly towards solving task v, and let $\boldsymbol{\theta} = (\theta_v)_{v \in V}$

be its vector form. We assume that the total contribution that each type receives is a linear combination of the direct contributions that are put into all tasks in the system, with coefficients $W_{u,v}$ that quantifies the extent to which efforts dedicated to a type-u task help complete a type-v task. We assume that $W_{u,v} = W_{v,u} \in [0,1]$ and $W_{u,u} = 1$ for all pairs $(u,v) \in V^2$, i.e., W is an $n \times n$ symmetric matrix with coefficients $W_{u,v}$ and has a diagonal of all ones. For $v \in V$, we use u_v to denote the total contribution entering a type-v task, and let $\boldsymbol{u} = (u_v)_{v \in V}$ be its vector form. We then have

$$u_v(\boldsymbol{\theta}) = \sum_{v' \in V} W_{v,v'} \theta_{v'}, \quad \text{and} \quad \boldsymbol{u}(\boldsymbol{\theta}) = W\boldsymbol{\theta}.$$

Graph representation. The task structure can be equivalently represented as a weighted undirected graph $G = (V, E, w)$, where each vertex is a task type in V. Each edge $\{v_1, v_2\} \in E$ represents that efforts towards task v_2 positively impact task v_1. In other words, $E = \{\{v_1, v_2\} : v_1, v_2 \in V, v_1 \neq v_2, W_{v_1,v_2} > 0\}$. The weight function $w : E \to [0,1]$ maps each edge $\{v_1, v_2\} \in E$ to W_{v_1,v_2} and reflects the coefficients between tasks. We also use $N(v)$ to denote the open neighborhood of vertex $v \in V$.

Strategic agents. Let there be n agents in the collaboration system. Each agent i is associated with a unique task type $t_i \in V$ in the network. Task types for all agents form a type profile $\boldsymbol{t} = (t_1, \ldots, t_n)$, which is a permutation of the types in V, i.e., $\boldsymbol{t} \in \mathsf{Sym}(V)$ with $\mathsf{Sym}(V)$ being the symmetric group on V. Additionally, each agent i can take an action $a_i \in \mathbb{R}_{\geq 0}$ that is the effort they put into solving their own task. The action profile $\boldsymbol{a}$ and task type profile $\boldsymbol{t} \in \mathsf{Sym}(V)$ naturally induce a type-based direct contribution profile $\boldsymbol{\theta}(\boldsymbol{a}; \boldsymbol{t}) = (\theta_v)_{v \in V}$, where $a_i = \theta_{t_i}$, or more succinctly $\boldsymbol{\theta}(\boldsymbol{a}; \boldsymbol{t}) = \Pi^{-1}\boldsymbol{a}$. In the above equation, $\Pi = (\pi_{i,j})$ is the $n \times n$ permutation matrix corresponding to permutation $\boldsymbol{t}$, where $\pi_{i,j} = 1$ if $j = t_i$ and 0 otherwise. We abbreviate $\boldsymbol{\theta}(\boldsymbol{a}; \boldsymbol{t})$ to $\boldsymbol{\theta}$ when it is clear from context.

For each agent i, the quality of a collaborative solution, denoted with $\mathsf{quality}_i$, is a function of both $\boldsymbol{a}$ and $\boldsymbol{t}$ and equals the total contribution entering the task of type t_i. We also use $\mathsf{quality} = (\mathsf{quality}_1, \cdots, \mathsf{quality}_n)$ to denote the agent-based utility vector. Formally, we have

$$\mathsf{quality}_i(\boldsymbol{a}; \boldsymbol{t}) = u_{t_i}(\boldsymbol{\theta}(\boldsymbol{a}; \boldsymbol{t})), \quad \text{and} \quad \mathsf{quality}(\boldsymbol{a}; \boldsymbol{t}) = \Pi \cdot \boldsymbol{u}(\boldsymbol{\theta}(\boldsymbol{a}; \boldsymbol{t})).$$

Strategic agents have three main concerns about the collaboration outcome: *feasibility, individual rationality (IR), and stability.* Given a type profile $\boldsymbol{t}$, an action profile $\boldsymbol{a}$, and the type-based action profile $\boldsymbol{\theta}$ they induce, we say that a solution is feasible if all agents secure a quality goal of 1, i.e., $\mathsf{quality}(\boldsymbol{a}; \boldsymbol{t}) \geq \mathbf{1}$ or equivalently $\boldsymbol{u}(\boldsymbol{\theta}) \geq \mathbf{1}$ coordinate-wise — For example, when specializing our model to the context of collaborative learning, $\mathsf{quality}_i$ represents the accuracy of the final model on agent i's learning task, and feasibility requires the model to achieve sufficient accuracy across all agents' tasks. IR is satisfied if participating in the collaboration system is more beneficial for each agent than completing the task independently. Note that a solution satisfies IR iff $\boldsymbol{a} \leq \mathbf{1}$ (or equivalently $\boldsymbol{\theta} \leq \mathbf{1}$) coordinate-wise. This is because the matrix W has an all-one diagonal, implying that completing a task independently would require a unit of effort. In addition, a feasible solution is stable if, given the contribution of others, no agent can unilaterally reduce their contribution without compromising their own feasibility. Formally, this requires

$$\forall i,\ a_i = \min\{x \geq 0 \mid \mathsf{quality}_i(x, \boldsymbol{a}_{-i}; \boldsymbol{t}) \geq 1\} \iff \forall v,\ \theta_v = \min\{x \geq 0 \mid u_v(x, \boldsymbol{\theta}_{-v}) \geq 1\}.$$

2.2 Benchmarks and the price of stability

Our benchmarks are the minima of the total workload (i.e., $\|\boldsymbol{a}\|_1 = \|\boldsymbol{\theta}\|_1$) among solutions that satisfy different combinations of the feasibility, IR, and stability requirements. We introduce three benchmarks: OPT as the optimal workload in feasible collaborations, and $\mathsf{OPT}^{\mathsf{IR}}$, $\mathsf{OPT}^{\mathsf{stable}}$ as the optimal workload subject to additional IR and stability constraints. Formally, they are solutions to the following programs:

$$\begin{aligned}
\mathsf{OPT} &= \min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \boldsymbol{u}(\boldsymbol{\theta}) \geq \mathbf{1},\ \boldsymbol{\theta} \geq \mathbf{0};\\
\mathsf{OPT}^{\mathsf{IR}} &= \min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \boldsymbol{u}(\boldsymbol{\theta}) \geq \mathbf{1},\ \boldsymbol{\theta} \geq \mathbf{0},\ \boldsymbol{\theta} \leq \mathbf{1};\\
\mathsf{OPT}^{\mathsf{stable}} &= \min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \boldsymbol{u}(\boldsymbol{\theta}) \geq \mathbf{1},\ \boldsymbol{\theta} \geq \mathbf{0},\ \theta_v = \min\{x \geq 0 \mid u_v(x, \boldsymbol{\theta}_{-v}) \geq 1\}\ \ (\forall v \in V).
\end{aligned}$$

The *price of stability (PoS)* is formally defined as the gap between the socially optimal solution and the optimal stable solution, i.e., $\mathsf{PoS} \triangleq \mathsf{OPT}^{\mathsf{stable}}/\mathsf{OPT}$.

2.3 Information design and signaling schemes

Prior distribution and information asymmetry. We now introduce the information design perspective of our model. Let τ be the prior distribution from which the type profile $\boldsymbol{t}$ is drawn. We model τ as the uniform distribution over the symmetric group, i.e., $\tau = \mathsf{Unif}(\mathsf{Sym}(V))$. As a result, for each fixed agent i, the marginal distribution on its type t_i is uniform over V. We assume that the graph structure G and prior distribution τ are common knowledge, but the realization of the true types $\boldsymbol{t}$ is exclusive knowledge held by the platform. The platform uses the private Bayesian persuasion protocol to strategically communicate this exclusive knowledge to the agents.

Private Bayesian persuasion. A *signaling scheme*, denoted with $\varphi : \mathsf{Sym}(V) \to \Delta(\mathcal{S}^V)$, is a mapping from the task type profile $\boldsymbol{t} \in \mathsf{Sym}(V)$ to a correlated distribution over private signals $\boldsymbol{s} = (s_v)_{v \in V} \in \mathcal{S}^V$, where $\mathcal{S} \subset [0, 1]$ is a finite space of signals, and each s_v is the signal value sent to the agent with type v.[4] The interaction protocol between the platform and the agents is as follows: initially, both the platform and the agents start with a common prior τ, and the platform credibly commits to a signaling scheme φ, which also becomes common knowledge upon commitment. Then, the true types $\boldsymbol{t} \sim \tau$ are realized from the prior distribution and observed by the platform. With the knowledge of $\boldsymbol{t}$, the platform generates a set of signals $\boldsymbol{s} = (s_v)_{v \in V} \sim \varphi(\boldsymbol{t})$, and sends s_{t_i} privately to agent i for all $i \in [n]$, which represents the recommended action to be taken. When an agent i receives the private signal s_{t_i}, they form a posterior belief about the type profile $\boldsymbol{t}$ and the signals sent to other types $\boldsymbol{s}_{-t_i}$, which we denote with μ_i:

$$\mu_i(\boldsymbol{t}, \boldsymbol{s}_{-t_i} \mid s_{t_i}) = \frac{\tau(\boldsymbol{t}) \cdot \Pr_{\varphi(\boldsymbol{t})}\left[s_{t_i}, \boldsymbol{s}_{-t_i}\right]}{\sum_{\boldsymbol{t}', \boldsymbol{s}'_{-t_i}} \tau(\boldsymbol{t}') \cdot \Pr_{\varphi(\boldsymbol{t}')}\left[s_{t_i}, \boldsymbol{s}'_{-t_i}\right]}. \tag{1}$$

Persuasiveness. We say that a signaling scheme is *persuasive* if no agent has the incentive to unilaterally deviate from the signal, assuming that all the other agents follow their signals. Formally, this requires that for every agent i, the following two conditions hold:

[4] In the original definition, the signals should be sent to each *agent* rather than each *type*, but the two definitions are equivalent as the platform can reassign the signals to agents according to $\Pi \cdot \boldsymbol{s}$.

- (Feasibility) For every $\theta \in \mathcal{S}$, conditioning on receiving a private signal $s_{t_i} = \theta$, taking action θ is feasible in expectation:

$$\mathop{\mathbb{E}}_{(\boldsymbol{t},\boldsymbol{s}_{-t_i})\sim\mu_i(\cdot|\boldsymbol{s}_{t_i}=\theta)}\left[u_{t_i}(\theta,\boldsymbol{s}_{-t_i})\right] \geq 1.$$

- (Stability) For every $\theta \in \mathcal{S}$, conditioning on receiving a private signal of value θ, the agent cannot contribute strictly less than θ effort while still meeting the aforementioned feasibility condition:

$$\theta = \min\left\{ x \geq 0 \;\middle|\; \mathop{\mathbb{E}}_{(\boldsymbol{t},\boldsymbol{s}_{-t_i})\sim\mu_i(\cdot|\boldsymbol{s}_{t_i}=\theta)}\left[u_{t_i}(x,\boldsymbol{s}_{-t_i})\right] \geq 1 \right\}.$$

  Equivalently, the feasibility condition must be tight whenever $\theta > 0$.

The *cost* of a persuasive signaling scheme φ is the expected total contribution assuming that all agents follow the signals. Formally,

$$\mathsf{Cost}(\varphi) \triangleq \mathop{\mathbb{E}}_{\boldsymbol{t}\sim\tau,\boldsymbol{s}\sim\varphi(\boldsymbol{t})}\left[\|\boldsymbol{s}\|_1\right].$$

Identity-independent signaling schemes. Note that in the above formulation, since φ is a function of permutations of V, the signal sent to each task type could depend on the identity of the agent performing the task. Nevertheless, we claim that it is sufficient to consider signaling schemes that are *identity-independent*: one such that the signal distribution is independent of the realized permutation, i.e., there is a distribution $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ such that $\varphi(\boldsymbol{t}) = \mathcal{D}_\varphi$ holds for all $\boldsymbol{t} \in \mathsf{Sym}(V)$. We prove the following lemma in Appendix A.2.

Lemma 2.1 (Identity-independent signaling schemes)**.** *For any persuasive signaling scheme φ, there exists an identity-independent signaling scheme $\widetilde{\varphi}$ that is persuasive and has $\mathsf{Cost}(\widetilde{\varphi}) = \mathsf{Cost}(\varphi)$.*

Since identity-independent signaling schemes are equivalent to joint distributions on $\mathcal{S}^V$, this lemma sets the groundwork for interpreting a signaling scheme as a "random labeling" or "random assignment" that directly assigns values in $\mathcal{S}$ to types in V. In the rest of the paper, we will adopt this terminology and refer to the signaling process in task structure graphs as "labeling" or "assigning" randomized values to its vertices. Furthermore, Lemma 2.1 shows that, to lower bound the cost of all persuasive signaling schemes on a particular instance, it suffices to prove such a lower bound against identity-independent schemes.

3 Technical Overview and Structural Results

This section outlines our approach for proving all of our main results. In Section 3.1, we give a sufficient and necessary condition for a signaling scheme to be persuasive and show that for *binary* signaling schemes this condition reduces to a simple inequality involving cuts and induced subgraphs in the graph representation of the collaboration system. This characterization is useful both for verifying the persuasiveness of the signaling schemes that we design, and for proving lower bounds against persuasive schemes. With these tools in hand, we give an overview of our upper bounds for unit-weight graphs (Section 3.2) and general weighted graphs (Section 3.3) starting from concrete illustrative examples. We end the section by explaining how we use the LP duality framework to prove our lower bounds.

3.1 Structural results

Consider an identity-independent signaling scheme $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ on a finite signal space $\mathcal{S} \subset [0,1]$. For each $\theta \in \mathcal{S}$, we introduce the concept of *slack* for receiving signal θ as follows:

Definition 3.1 (Slack)**.** *For a signaling scheme $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ and any $\theta \in \mathcal{S}$, we define $\mathsf{Contrib}_\theta$ as the expected total contribution entering vertices receiving θ, and Num_θ as the expected number of vertices receiving θ:*

$$\mathsf{Contrib}_\theta \triangleq \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v=\theta\right] \sum_{v'\in N(v)} W_{v,v'} s_{v'}\right]; \quad \mathsf{Num}_\theta \triangleq \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v=\theta\right]\right],$$

where $N(v)$ stands for the open neighborhood of v. The slack *of receiving signal θ, denoted with Δ_θ, is the difference between $\mathsf{Contrib}_\theta$ and Num_θ scaled by the factor $(1-\theta)$:*

$$\Delta_\theta \triangleq \mathsf{Contrib}_\theta - (1-\theta)\cdot\mathsf{Num}_\theta.$$

Lemma 3.2 (Persuasiveness of general signaling schemes)**.** *An identity-independent signaling scheme $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ with finite $\mathcal{S}$ is persuasive if and only if the following conditions are met:*

- *For every $\theta \in \mathcal{S}$ where $\theta > 0$, the slack Δ_θ must equal zero ($\Delta_\theta = 0$).*
- *If the signal space includes zero ($0 \in \mathcal{S}$), the slack Δ_0 must be non-negative ($\Delta_0 \geq 0$).*

Proof sketch. Recall from the definition of persuasiveness, $\mathcal{D}_\varphi$ is persuasive if and only if $\forall\theta \in \mathcal{S}$, $\mathbb{E}_{(\boldsymbol{t},\boldsymbol{s}_{-t_i})\sim\mu_i(\cdot|\boldsymbol{s}_{t_i}=\theta)}\left[u_{t_i}(\theta,\boldsymbol{s}_{-t_i})\right] \geq 1$, with this inequality being tight for $\theta > 0$. Therefore, to prove Lemma 3.2, it suffices to show that

$$\mathop{\mathbb{E}}_{(\boldsymbol{t},\boldsymbol{s}_{-t_i})\sim\mu_i(\cdot|\boldsymbol{s}_{t_i}=\theta)}\left[u_{t_i}(\theta,\boldsymbol{s}_{-t_i})\right] = \theta + \frac{\mathsf{Contrib}_\theta}{\mathsf{Num}_\theta}. \tag{2}$$

This is because $\theta + \frac{\mathsf{Contrib}_\theta}{\mathsf{Num}_\theta} \geq 1 \Leftrightarrow \mathsf{Contrib}_\theta - (1-\theta)\cdot\mathsf{Num}_\theta \geq 0$ when $\mathsf{Num}_\theta > 0$, which is true because it is without loss of generality to assume that every $\theta \in \mathcal{S}$ is realized with a non-zero probability. We establish Equation (2) by straightforward application of Bayes' rule in Appendix B.1. □

A special class of binary signals Consider a binary scheme φ with signal space $\mathcal{S} = \{0,\alpha\}$ for $\alpha \in (0,1]$. Recall from Lemma 2.1 that $\mathcal{D}_\varphi$ can be interpreted as a random labeling of V using either 0 or α. It is useful to view the random labeling $\boldsymbol{s}$ as first selecting a random subset $S \subseteq V$ from a distribution $\mathcal{D}$ over all subsets of V (namely 2^V), and then assigning $s_v = \alpha \cdot \mathbb{1}\left[v \in S\right]$ for each $v \in V$. Using this view, we can rewrite $\mathsf{Contrib}_0$ and $\mathsf{Contrib}_\alpha$ using the random subset $S \sim \mathcal{D}$:

$$\mathsf{Contrib}_\alpha = \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v=\alpha\right] \sum_{v'\in N(v)} W_{v,v'} s_{v'}\right] = \alpha \cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[\sum_{v\in S}\sum_{v'\in S\setminus\{v\}} W_{v,v'}\right];$$

$$\mathsf{Contrib}_0 = \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v=0\right] \sum_{v'\in N(v)} W_{v,v'} s_{v'}\right] = \alpha \cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[\sum_{v\notin S}\sum_{v'\in S} W_{v,v'}\right].$$

The above two quantities have graph-theoretical interpretation using the cut $(S, V\setminus S)$ and the subgraph induced by S. We define their total weight respectively as follows.

Definition 3.3 (Weights of Cut and Induced Subgraph)**.** *The weight of the cut between S and its complement $V \setminus S$, denoted $\mathsf{Cut}(S, V \setminus S)$, and the weight of the subgraph induced by S, denoted $\mathsf{Induced}(S)$, are given by*

$$\mathsf{Cut}(S, V \setminus S) \triangleq \sum_{u \in S} \sum_{v \in (V \setminus S)} W_{u,v};$$
$$\mathsf{Induced}(S) \triangleq \sum_{u \in S} \sum_{v \in S} W_{u,v} = |S| + 2 \sum_{\{u,v\} \subseteq S} W_{u,v}.$$

Here, $\mathsf{Cut}(S, V \setminus S)$ measures the total weight of edges crossing the cut from S to $V \setminus S$, and $\mathsf{Induced}(S)$ measures the total weight of all edges within S, including self-loops with $W_{v,v} = 1$, and counts each undirected internal edge twice.

Leveraging the characteristics of S defined in Definition 3.3, we provide a graph-theoretical characterization of the persuasiveness of a binary signaling scheme in the next lemma. The detailed proof mirrors the arguments used in Lemma 3.2 and is deferred to Appendix B.2.

Lemma 3.4 (Persuasiveness of binary signaling schemes)**.** *Let $\mathcal{D} \in \Delta(2^V)$ be a distribution over subsets of V. If $\mathcal{D}$ satisfies the following inequality,*[5]

$$\frac{\mathbb{E}_{S \sim \mathcal{D}}\left[\mathsf{Cut}(S, V \setminus S)\right]}{\mathbb{E}_{S \sim \mathcal{D}}\left[|V \setminus S|\right]} \geq \frac{\mathbb{E}_{S \sim \mathcal{D}}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{S \sim \mathcal{D}}\left[|S|\right]},$$

for $\alpha = \frac{\mathbb{E}_{S \sim \mathcal{D}}[|S|]}{\mathbb{E}_{S \sim \mathcal{D}}[\mathsf{Induced}(S)]}$, there exists an identity-independent persuasive signaling scheme $\mathcal{D}_\varphi \in \Delta\left(\{0, \alpha\}^V\right)$ that assigns $s_v = \alpha \cdot \mathbb{1}\left[v \in S\right]$ according to a random subset $S \sim \mathcal{D}$. The cost of this signaling scheme is $\mathsf{Cost}(\varphi) = \frac{(\mathbb{E}_{S \sim \mathcal{D}}[|S|])^2}{\mathbb{E}_{S \sim \mathcal{D}}[\mathsf{Induced}(S)]} \leq \mathbb{E}_{S \sim \mathcal{D}}\left[|S|\right]$.

Corollary 3.5 (Cost of the no-information scheme)**.** *When $\mathcal{D}$ in Lemma 3.4 is the degenerate distribution at V, the resulting signaling scheme is persuasive for $\alpha = \frac{n}{\mathsf{Induced}(V)}$ and incurs a cost of $\frac{n^2}{\mathsf{Induced}(V)}$. Using $m = \sum_{\{v_1, v_2\} \in E} W_{v_1, v_2}$ to denote the total edge weights in the graph, the cost simplifies to $\frac{n^2}{n+2m} = O(\frac{n^2}{m})$. This scheme, by not differentiating between vertices, essentially conveys no information. Therefore, its cost is equal to the total contribution of agents when no signal is sent.*

3.2 Overview of upper bounds in unit-weight graphs

We start by sketching the proof of Theorem 1.1, which gives a binary signaling scheme with an $O(\sqrt{n} \cdot \mathsf{OPT})$ cost. Recall that by Lemma 2.1, a signaling scheme is equivalent to a randomized labeling of vertices with values between 0 and 1.

A signaling scheme for the double-star graph. Our proof is best illustrated by the double-star example from Figure 1, on which the signaling scheme simply randomizes between the following two plans for some small $\epsilon > 0$:

- Plan A: Label the two centers with $1 - \epsilon$. Label the $2k$ leaves with 0.
- Plan B: Label the two centers with 0. Label the $2k$ leaves with $1 - \epsilon$.

Plan A is almost the same as the socially optimal solution (without the stability constraint), in which each center contributes a unit amount. While Plan A has a low cost, we should not always follow it,

[5] When $\mathcal{D}$ is the degenerate distribution at V, the left-hand side is treated as $+\infty$.

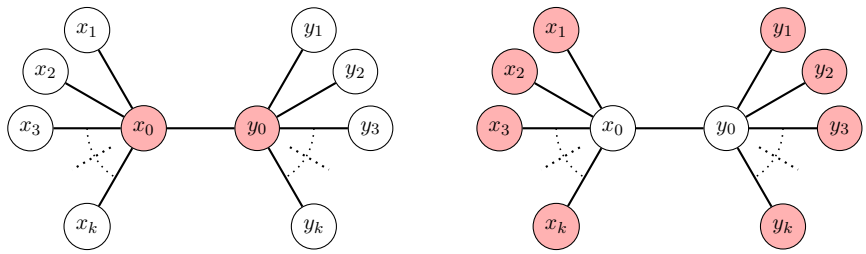

Figure 2: Plan A (left) and Plan B (right) for the double-star graph. The shaded vertices are labeled with $1-\epsilon$. The empty vertices are labeled with 0.

for two different reasons. First, it violates the stability condition: conditioning on receiving signal $1-\epsilon$, the agent knows for sure that one of their neighbors will play $1-\epsilon$, and thus has an incentive to deviate and play a much lower value. Second, it violates feasibility: conditioning on receiving 0, the agent expects a total contribution of $1-\epsilon$ from their neighbors, so following the signal would not satisfy their demand. On the other hand, while Plan B appears extremely inefficient at first glance, it does remedy both issues of Plan A by reducing the slack when signal $1-\epsilon$ is received, and increasing the total contributions from neighbors when the signal is 0.

With the notation from Definition 3.1 (i.e., Δ_θ denotes the amount of slack corresponding to signal θ), always following Plan A gives

$$\Delta_{1-\epsilon} = 2\cdot(1-\epsilon) - \epsilon\cdot 2 = +\Theta(1), \quad \Delta_0 = 2k(1-\epsilon) - 1\cdot 2k = -\Theta(k\epsilon),$$

whereas following Plan B gives

$$\Delta_{1-\epsilon} = 0 - \epsilon\cdot 2k = -\Theta(k\epsilon), \quad \Delta_0 = 2k(1-\epsilon) - 1\cdot 2 = +\Theta(k).$$

A simple calculation shows that, for some $\epsilon, p = \Theta(1/\sqrt{k})$, we can ensure $\Delta_{1-\epsilon} = 0$ and $\Delta_0 \geq 0$ by following Plan A with probability $1-p$ and following Plan B with probability p. The resulting cost would then be $(1-p)\cdot 2\cdot(1-\epsilon) + p\cdot 2k\cdot(1-\epsilon) = O(\sqrt{k}) = O(\sqrt{n})$ as desired.

A binary signal perspective. To generalize this result to all unit-weight graphs, it is helpful to view the signaling scheme above through the lens of Lemma 3.4, our characterization of persuasive binary signaling schemes. Recall that the lemma states that a distribution $\mathcal{D} \in \Delta(2^V)$ gives a persuasive binary signaling scheme if the following inequality holds:

$$\frac{\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Cut}(S, V\setminus S)\right]}{\mathbb{E}_{S\sim\mathcal{D}}\left[|V\setminus S|\right]} \geq \frac{\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]}. \tag{3}$$

Furthermore, the cost of the scheme is at most $\mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]$.

In the double-star example, let DS be the set of the two centers, and IS be the set of the $2k$ leaves. (The names are justified since DS is a dominating set of the graph, and IS is an independent set.) When $\mathcal{D}$ is the degenerate distribution at DS, Equation (3) reduces to $\frac{2k}{2k} \geq \frac{4}{2}$, which does not hold. In contrast, when $\mathcal{D}$ is the degenerate distribution at IS, Equation (3) gives $\frac{2k}{2} \geq \frac{2k}{2k}$, which not only holds, but holds with a large margin! Again, it follows from an elementary calculation that, Equation (3) can be satisfied by setting $\mathcal{D}(\mathsf{DS}) = 1-p$ and $\mathcal{D}(\mathsf{IS}) = p$ for some $p = \Theta(1/\sqrt{k})$, and the resulting cost is, as expected, bounded by $\mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right] = (1-p)\cdot|\mathsf{DS}| + p\cdot|\mathsf{IS}| = O(\sqrt{n})$.

Signaling scheme for general unit-weight graphs. In general, we choose DS as a minimum dominating set of G, and choose IS as a maximal independent set of the sub-graph induced by $V \setminus \mathsf{DS}$. As in the double-star graph, DS gives a good approximation of the socially optimal cost OPT, but does not guarantee persuasiveness. Concretely, $|\mathsf{DS}|/\mathsf{OPT}$ is upper bounded by the integrality gap of the LP relaxation of the dominating set problem, which is $O(\log n)$ (e.g., [Williamson and Shmoys, 2011, Section 1.7]). To see that DS might not be persuasive, note that the left-hand side of Equation (3), $\frac{\mathsf{Cut}(\mathsf{DS}, V \setminus \mathsf{DS})}{|V \setminus \mathsf{DS}|}$, can be as small as 1, while the right-hand side $\frac{\mathsf{Induced}(\mathsf{DS})}{|\mathsf{DS}|}$ can be as large as $|\mathsf{DS}|$, when the induced sub-graph of DS is a clique.

Once again, the other set IS comes to the rescue of persuasiveness. By the independence of IS, the right-hand side of Equation (3) gives $\frac{\mathsf{Induced}(\mathsf{IS})}{|\mathsf{IS}|} = 1$. For the left-hand side, we note that

$$\mathsf{Cut}(\mathsf{IS}, V \setminus \mathsf{IS}) = \mathsf{Cut}(\mathsf{IS}, \mathsf{DS}) + \mathsf{Cut}(\mathsf{IS}, V \setminus (\mathsf{IS} \cup \mathsf{DS})) \geq |\mathsf{IS}| + |V \setminus (\mathsf{IS} \cup \mathsf{DS})| = n - |\mathsf{DS}|,$$

The first step holds since DS and $V \setminus (\mathsf{IS} \cup \mathsf{DS})$ partitions $V \setminus \mathsf{IS}$ (see Figure 3 for a pictorial illustration). In the second step, $\mathsf{Cut}(\mathsf{IS}, \mathsf{DS}) \geq |\mathsf{IS}|$ since DS is a dominating set of G, which guarantees that every vertex in $\mathsf{IS} \subseteq V \setminus \mathsf{DS}$ has a neighbor in DS. Meanwhile, the maximality of IS implies that every vertex in $V \setminus (\mathsf{DS} \cup \mathsf{IS})$ has a neighbor in IS. This gives $\mathsf{Cut}(\mathsf{IS}, V \setminus (\mathsf{IS} \cup \mathsf{DS})) \geq |V \setminus (\mathsf{IS} \cup \mathsf{DS})|$. Therefore, the left-hand side of Equation (3), $\frac{\mathsf{Cut}(\mathsf{IS}, V \setminus \mathsf{IS})}{|V \setminus \mathsf{IS}|}$, is at least $\frac{n - |\mathsf{DS}|}{n - |\mathsf{IS}|}$, which is typically strictly larger than 1. This shows that Equation (3) holds with a margin for IS.

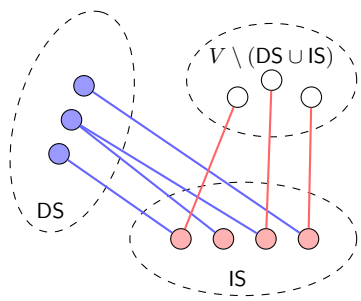

Figure 3: $\mathsf{Cut}(\mathsf{IS}, V \setminus \mathsf{IS})$. Every vertex in IS has a neighbor in DS; every vertex in $V \setminus (\mathsf{DS} \cup \mathsf{IS})$ has a neighbor in IS.

By carefully randomizing between DS and IS, we obtain a persuasive binary signaling scheme with cost $O(\sqrt{n}) \cdot |\mathsf{DS}|$, which is an $O(\sqrt{n} \cdot \log n)$-approximation of OPT. To shave this extra $\log n$ factor, our actual proof replaces DS with a distribution over sets obtained from an independent rounding of the optimal solution. This essentially preserves all the desirable properties of DS, while reducing the cost from $|\mathsf{DS}|$ to OPT.

Strict improvement upon $\mathsf{OPT}^{\mathsf{stable}}$. We sketch the proof of Theorem 1.3, which states that there is a binary signaling scheme with a strictly lower cost than $\mathsf{OPT}^{\mathsf{stable}}$, whenever $\mathsf{OPT} < \mathsf{OPT}^{\mathsf{stable}}$. A key ingredient of the proof is the construction of a binary scheme with cost exactly $\mathsf{OPT}^{\mathsf{stable}}$, which is based on the following structural result regarding stable solutions in unit-weight graphs.

Lemma 3.6. *Let $\boldsymbol{\theta} \in [0,1]^V$ be an arbitrary stable solution on a unit-weight graph $G = (V, E)$. There exists a distribution $\mathcal{D} \in \Delta(2^V)$ supported over the independent sets of G, such that for every $v \in V$, $\Pr_{S \sim \mathcal{D}}[v \in S] = \theta_v$.*

We prove Lemma 3.6 in Section 5 by rounding the fractional solution $\boldsymbol{\theta}$ (which satisfies stability) into an integral solution (i.e., a subset of V), via a rounding scheme known as *competing exponential clocks*. Given the lemma, it then follows quite easily that such a distribution $\mathcal{D}$ satisfies Equation (3), and gives a binary signaling scheme of cost exactly $\mathsf{OPT}^{\mathsf{stable}}$. To see this, note that since every set in the support of $\mathcal{D}$ is an independent set, we have $\mathbb{E}_{S \sim \mathcal{D}}[\mathsf{Induced}(S)] = \mathbb{E}_{S \sim \mathcal{D}}[|S|]$, i.e., the right-hand side of Equation (3) is 1. Then, using the assumption that $\boldsymbol{\theta}$ is a stable solution, along

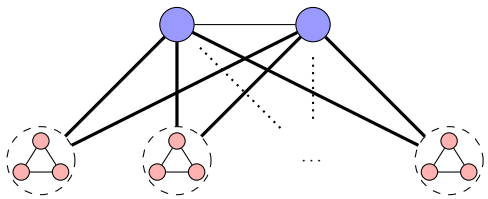

Figure 4: A weighted graph on which all binary schemes fail. There are k triangles in total. Each thick line indicates the edges between a center and all the three vertices in a triangle. Every edge is of weight 1/2.

with the fact that the marginal of $\mathcal{D}$ exactly matches $\boldsymbol{\theta}$, we can prove that $\mathbb{E}_{S\sim\mathcal{D}}[\mathsf{Cut}(S, V \setminus S)] \geq n - \|\boldsymbol{\theta}\|_1 = \mathbb{E}_{S\sim\mathcal{D}}[|V \setminus S|]$. This lower bounds the left-hand side of Equation (3) by 1.

The other ingredient of our proof is an argument based on LP duality, which shows that, whenever $\mathsf{OPT} < \mathsf{OPT}^{\mathsf{stable}}$, there exists an optimal stable solution $\boldsymbol{\theta} \in [0,1]^V$ such that the rounding of $\boldsymbol{\theta}$ actually satisfies Equation (3) with a positive margin. This allows us to randomize between this scheme (with cost $\mathsf{OPT}^{\mathsf{stable}}$) and an independent rounding of the optimal solution (with cost $\mathsf{OPT} < \mathsf{OPT}^{\mathsf{stable}}$), and achieve a cost strictly below $\mathsf{OPT}^{\mathsf{stable}}$.

3.3 Overview of upper bounds in weighted graphs

Recall that for unit-weight graphs, our binary signaling scheme is based on sending a positive signal to either DS (a minimum dominating set) or IS (a maximal independent set of the sub-graph induced by $V \setminus \mathsf{DS}$), each with a carefully chosen probability. Sending signal 1 to the vertices in DS gives a low cost, guarantees feasibility, but might not be stable. The purpose of the set IS is to reduce the slack introduced by DS.

What goes wrong when we repeat this strategy on weighted graphs? The argument about DS generalizes easily. On unit-weight graphs, a dominating set is equivalent to an integral solution, in which the contribution from each vertex is either 0 or 1. Naturally, for weighted graphs, we choose DS as the set corresponding to the optimal integral solution. Towards satisfying Equation (3), the feasibility of the integral solution guarantees $\mathsf{Cut}(\mathsf{DS}, V \setminus \mathsf{DS}) \geq |V \setminus \mathsf{DS}|$, i.e., the left-hand side of Equation (3) is lower bounded by 1. Under mild assumptions on the edge weights, we can still show that $|\mathsf{DS}|$ is not much larger than $\mathsf{OPT}^{\mathsf{IR}}$, the optimal fractional solution subject to IR.

Unfortunately, as we demonstrate in the example below, it might be impossible to find an analogue of IS for our purpose.

A concrete example where binary schemes fail. The graph is shown in Figure 4. Every edge in the graph has a weight of 1/2. The graph contains two "centers" and k triangles (consisting of $3k$ "leaves"). Each center is connected to all other vertices in the graph.

On this graph, the optimum subject to IR, $\mathsf{OPT}^{\mathsf{IR}} = 2$, is achieved when each center plays 1. As discussed earlier, we pick DS as the set of the two centers. If we, as in unit-weight graphs, pick IS as a maximal independent set of the induced sub-graph of $V \setminus \mathsf{DS}$, IS would contain exactly one vertex in each triangle. However, it can be verified that no distribution $\mathcal{D}$ supported over $\{\mathsf{DS}, \mathsf{IS}\}$ satisfies Equation (3). Therefore, we cannot derive a persuasive binary signaling scheme following the previous approach.

In fact, this is not even a problem with this specific choice of DS and IS — in Theorem 8.1, we show that *every* persuasive binary signaling scheme has a cost of $\Omega(n)$ on the instance above!

Solution: a third signal. While binary schemes are insufficient for weighted graphs, a minimal fix — adding a third signal — would solve the problem. Concretely, for the graph in Figure 4, there is a non-trivial signaling scheme via randomizing among the three plans below:

- Plan A: Label the centers with $1-\epsilon$. Label the leaves with 0.
- Plan B: Label one vertex from each triangle with $1-\epsilon$. Label all other vertices with 0.
- Plan C: Label the leaves with α. Label the centers with 0.

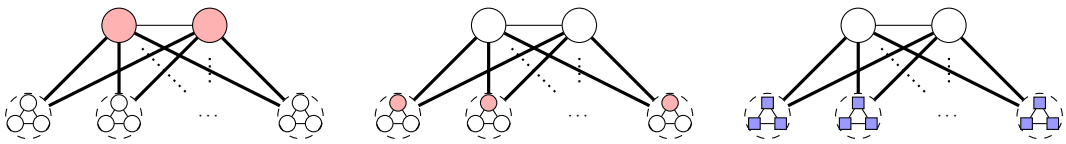

Figure 5: The three plans (left: Plan A; middle: Plan B; right: Plan C) of signaling for the graph from Figure 4. Shaded circle vertices are labeled with $1-\epsilon$. Empty circle vertices are labeled with 0. Shaded square vertices are labeled with α.

To ensure persuasiveness, we go back to the characterization in Lemma 3.2. The three plans, when we follow each of them alone, give

- Plan A: $\Delta_{1-\epsilon} = (1-\epsilon) - \epsilon \cdot 2 = +\Theta(1)$ and $\Delta_0 = 3k(1-\epsilon) - 1 \cdot 3k = -\Theta(k\epsilon)$.
- Plan B: $\Delta_{1-\epsilon} = 0 - \epsilon \cdot k = -\Theta(k\epsilon)$ and $\Delta_0 = 2k(1-\epsilon) - 1 \cdot (2k+2) = -\Theta(1+k\epsilon)$.
- Plan C: $\Delta_0 = 3k\alpha - 1 \cdot 2 = 3k\alpha - 2$ and $\Delta_\alpha = 3k\alpha - (1-\alpha) \cdot 3k = 6k\alpha - 3k$.

To satisfy the constraint $\Delta_\alpha = 0$, we set $\alpha = 1/2$. The effect of Plan C is then simplified to

$$\Delta_0 = +\Theta(k) \quad \text{and} \quad \Delta_{1/2} = 0.$$

In hindsight, the introduction of Plan C and the third signal α is natural. Our previous strategy, which corresponds to randomizing between only Plans A and B, cannot guarantee both $\Delta_{1-\epsilon} = 0$ and $\Delta_0 \geq 0$ — enforcing $\Delta_{1-\epsilon} = 0$ inevitably results in $\Delta_0 < 0$. The purpose of Plan C is then to bring Δ_0 back to 0, without introducing additional slacks.

Formally, it can be verified that, for some $\epsilon, p, q = \Theta(1/\sqrt{k})$, following the three plans with probability $1-p-q$, p, and q respectively gives a persuasive scheme. The resulting cost is given by

$$(1-p-q) \cdot 2(1-\epsilon) + p \cdot k(1-\epsilon) + q \cdot 3k\alpha = O(\sqrt{k}) = O(\sqrt{n}).$$

Upper bound for general graphs. For general weighted graphs, generalizing the ternary signaling scheme above gives a cost of $n|\mathsf{DS}|/\sqrt{|\mathsf{IS}|}$, where IS is an independent set of the graph. In particular, the first two plans still correspond to assigning signal $1-\epsilon$ to DS and IS, respectively, while Plan C assigns a different non-zero value to $V \setminus \mathsf{DS}$, in the hope of making the slack at 0 non-negative. When we can find a large IS, this would give a non-trivial approximation of $|\mathsf{DS}|$, which in turn approximates $\mathsf{OPT}^{\mathsf{IR}}$. What if there are no large independent sets, e.g., when the graph is dense?

Assuming a lower bound on the non-zero edge weights, this issue can be resolved via a win-win argument. Suppose that the edge weights are at least $\delta > 0$. Let m be the total edge weight in the graph. If m is large, by Corollary 3.5, we have a persuasive signaling scheme with cost $O(n^2/m)$. If m is small, by our assumption on the edge weights, there are at most m/δ edges in the graph.

Intuitively, this sparse graph must have a large independent set. Indeed, we can lower bound $|\mathsf{IS}|$ by $\Omega(\delta n^2/m)$. Then, regardless of the value of m, the better between the two schemes gives an $O((n \cdot \mathsf{OPT}^{\mathsf{IR}})^{2/3} \cdot \delta^{-1/3})$ cost.

Without such a lower bound δ, we still have a non-trivial approximation of $\mathsf{OPT}^{\mathsf{IR}}$. The key observation is that IS does not need to be an independent set — the same proof strategy goes through as long as, in the sub-graph induced by IS, every vertex has a (weighted) degree of $O(\epsilon)$. Going one step further, we do not need IS to be a deterministic set at all! In our actual proof, we replace IS with a random set $\widetilde{\mathsf{IS}} \subseteq V$, obtained from including each vertex in the graph with a fixed, carefully chosen probability. This gives the $O(n^{3/4} \cdot \left(\mathsf{OPT}^{\mathsf{IR}}\right)^{1/2})$ bound for the general case.

3.4 Overview of lower bounds

To prove the tightness of our approximation guarantees, we start by reverse-engineering the proofs of the upper bounds, and identifying graphs on which the analyses are tight. For unit-weight graphs, the hard instance is exactly the double-star graph in Figure 1. For graphs with edge weights bounded by $\Omega(1)$, we modify the construction in Figure 4 by replacing the $\Theta(n)$ triangles with $\Theta(n^{2/3})$ copies of a clique of size $\Theta(n^{1/3})$.

The more difficult part is, of course, to lower bound the cost of *all* persuasive signaling schemes on these graphs. To this end, we revisit the characterization of persuasiveness in Lemma 3.2: $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ is persuasive if and only if the induced slacks satisfy $\Delta_0 \geq 0$ and $\Delta_\theta = 0$ for all $\theta \in \mathcal{S} \setminus \{0\}$. By Definition 3.1, each Δ_θ can be expressed as an expectation over the distribution $\mathcal{D}_\varphi$. Therefore, the family of persuasive signaling schemes with signal space $\mathcal{S}$ is exactly a subset of $\Delta(\mathcal{S}^V)$ defined by finitely many linear constraints. Consequently, the minimum-cost persuasive scheme is characterized by a linear program.

This simple observation leads us to lower bound the cost of persuasive signaling schemes by constructing a feasible solution of the dual LP. Concretely, suppose that we could find a function $f : [0,1] \to \mathbb{R}$ such that $f(0) \geq 0$ and, for any $\mathcal{S}$ and degenerate distribution at $s \in \mathcal{S}^V$ (i.e., a deterministic labeling of V using values in $\mathcal{S}$), the resulting slacks satisfy

$$\|s\|_1 \geq \sum_{\theta \in \mathcal{S}} f(\theta) \cdot \Delta_\theta + C. \tag{4}$$

Then, for any persuasive signaling scheme $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$, the linearity of expectation gives

$$\mathop{\mathbb{E}}_{s \sim \mathcal{D}_\varphi} [\|s\|_1] \geq \sum_{\theta \in \mathcal{S}} f(\theta) \cdot \Delta_\theta + C \geq C.$$

The last step above holds since when $\theta = 0$, $f(0) \cdot \Delta_0 \geq 0$ holds as both $f(0)$ and Δ_0 are non-negative. When $\theta \neq 0$, $\Delta_\theta = 0$ implies $f(\theta) \cdot \Delta_\theta = 0$.

Our proof of the $\Omega(n^{2/3})$ lower bound (for graphs with edge weights lower bounded by $\Omega(1)$) proceeds by carefully choosing the function $f(\theta)$, and proving Equation (4) for a sufficiently large C via an involved case analysis. For unit-weight graphs, while the $\Omega(\sqrt{n})$ lower bound admits a simpler proof, in the proof we implicitly consider a dual solution, in which f is a piece-wise constant function f that takes a negative value with a large magnitude when θ is close to 1.

4 Discussion and Future Directions

In this section, we highlight a few concrete open problems and discuss a few natural extensions of our model.

The power of simple signaling schemes. A recurring theme in our positive results is the surprising effectiveness of *simple* signaling schemes that only use a few different signal values. The only exception is Theorem 1.4: we use $\Omega(n)$ different signals to achieve a cost below $\mathsf{OPT}^{\mathsf{stable}}$. Can we reduce the number of signals to $O(1)$? In Appendix C.1, we explain the technical difficulties in proving such a result, and give instances which suggest that the binary and ternary schemes that we consider are insufficient in general.

Tight approximation ratio for weighted graphs. In terms of worst-case approximation guarantees, the only result that is not shown to be optimal is the $O(n^{3/4})$ approximation on general weighted graphs. In Appendix C.2, we give a concrete instance on which the $n^{3/4}$ ratio is conjectured to be tight, and discuss why the conjecture does not following easily from our current proof strategy.

Alternative collaboration systems. Our work focuses on collaboration systems in which the quality of each agent's task is a linear combination of the agent's own contribution and the amounts contributed by the neighbors. In the context of collaborative federated learning, this corresponds to the *random discovery* model proposed by Blum et al. [2021a]. A non-linear version of the collaboration system is another model of Blum et al. [2021a], termed *random coverage*, in which each agent is associated with a (discrete) data distribution, and the accuracy of each agent's task is linear in the total probability mass of the elements sampled by the agent themself and their neighbors. This formulation brings more structure to the relation between each pair of agents' tasks and might allow us to circumvent some hard instances in the linear model. For instance, agent i's data are maximally effective for another agent j only if they share the same distribution, and this property would be transitive. In contrast, in our current model, there might exist task types i, j, and k such that $W_{i,j} = W_{j,k} = 1$ yet $W_{i,k} = 0$.

Alternative models of incentives. Another modeling assumption that we made is regarding the agents' incentives. Implicit in the definition of persuasiveness (Section 2.3) is that each agent prioritizes satisfying feasibility, and exactly minimizes their own effort subject to the feasibility constraint. A natural relaxation is to allow *approximate stability*, i.e., a signaling scheme is considered stable as long as no agent has the incentive to decrease their action by some $\epsilon > 0$. We may also consider alternative models in which each agent maximizes the expected quality of their task minus a penalization term that depends on their own workload.

5 Upper Bounds for Unit-Weight Graphs

In this section, we prove the upper bounds in Theorems 1.1 and 1.3. To this end, we first state a few bounds on the cuts and induced sub-graphs in the unit-weight graph representation of the collaboration system. We then present a binary signaling scheme that is shown to be persuasive and achieve an $O(\sqrt{n})$ approximation of OPT. Towards proving Theorem 1.3, we give a binary scheme with a cost of exactly $\mathsf{OPT}^{\mathsf{stable}}$, which, under the additional assumption that $\mathsf{OPT} < \mathsf{OPT}^{\mathsf{stable}}$, easily implies a strict improvement upon $\mathsf{OPT}^{\mathsf{stable}}$.

5.1 Weights of cuts and induced sub-graphs

We start by stating several elementary bounds on the cuts and induced edge weights (from Definition 3.3) for dominating sets and maximal independent sets in a graph. These bounds will be used towards constructing binary signaling schemes via Lemma 3.4. We defer the proofs to Appendix D.

The first bound is regarding maximal independent sets in an induced sub-graph obtained from removing a dominating set.

Lemma 5.1. *Let* DS *denote a dominating set of a unit-weight graph* $G = (V, E)$*, and* IS *be an arbitrary maximal independent set of the sub-graph induced by* $V \setminus \mathsf{DS}$*. Then,* $\mathsf{Cut}(\mathsf{IS}, V \setminus \mathsf{IS}) \geq |V| - |\mathsf{DS}|$.

The second lemma considers a random set $\widetilde{\mathsf{DS}}$ obtained from the independent rounding of a feasible solution that satisfies the IR requirement as defined in Section 2.2. This lemma holds for weighted graphs as well.

Lemma 5.2. *Let* $G = (V, E, w)$ *be a weighted graph, and* W *be the corresponding matrix. Let* $\boldsymbol{\theta} \in [0,1]^V$ *denote a feasible solution subject to IR, i.e.,* $W\boldsymbol{\theta} \geq 1$ *and* $\boldsymbol{\theta} \leq 1$ *hold coordinate-wise. Let* $\mathcal{D} \in \Delta(2^V)$ *denote the distribution of a random subset of* V *that includes each* $v \in V$ *independently with probability* θ_v*. Then, the following bounds hold:*

- $\mathbb{E}_{S \sim \mathcal{D}}[|S|] = \|\boldsymbol{\theta}\|_1$.
- $\mathbb{E}_{S \sim \mathcal{D}}[\mathsf{Induced}(S)] \leq \|\boldsymbol{\theta}\|_1^2 + \|\boldsymbol{\theta}\|_1$.
- $\mathbb{E}_{S \sim \mathcal{D}}[\mathsf{Cut}(S, V \setminus S)] \geq |V| - 2\|\boldsymbol{\theta}\|_1$.

5.2 General unit-weight graphs

In this section, we prove the $O(\sqrt{n} \cdot \mathsf{OPT})$ upper bound in Theorem 1.1 by designing a distribution $\mathcal{D} \in \Delta(2^V)$ that satisfies Lemma 3.4. Our proof closely relies on the properties developed in Section 5.1. Before the proof, we first present a lemma on the integrality gap of OPT. Note that the program for computing OPT is essentially the relaxed program determining the minimum dominating set of a graph. Therefore, our lemma follows from the integrality gap upper bound for the dominating set problem, e.g., see [Williamson and Shmoys, 2011, Section 1.7].

Lemma 5.3 (Integrality gap)**.** *There exists a dominating set* $\mathsf{DS} \subseteq V$ *such that* $|\mathsf{DS}| = O(\log n \cdot \mathsf{OPT})$*. Moreover, the set* DS *can be computed efficiently through randomized rounding of* OPT*.*

We are now ready to prove the upper bound theorem.

Theorem 5.4 (Upper bound part of Theorem 1.1)**.** *In any unit-weight graph, there is a persuasive signaling scheme with a cost of* $O(\sqrt{n} \cdot \mathsf{OPT})$*.*

Proof of Theorem 5.4. We start with constructing a family of distributions $\mathcal{D}_p \in \Delta(2^V)$ parametrized by $p \in [0,1]$, and then show that there exists a choice of p such that the binary signaling scheme induced by $\mathcal{D}_p$ is persuasive and has a small cost.

Let $\boldsymbol{\theta}^\star \in [0,1]^V$ be a socially optimal solution that achieves $\|\boldsymbol{\theta}^\star\|_1 = \mathsf{OPT}$. We define $\mathcal{D}_A \in \Delta(2^V)$ to be the distribution of a random subset $\widetilde{\mathsf{DS}} \subseteq V$ that includes each $v \in V$ independently with probability $\theta_v^\star$ — we call $\widetilde{\mathsf{DS}} \sim \mathcal{D}_A$ an *independent rounding* of $\theta^\star$. In addition, let DS be a dominating set satisfying Lemma 5.3. Then, $\mathcal{D}_B \in \Delta(2^V)$ is defined as the degenerate distribution on IS where IS is a maximal independent set of $V \setminus \mathsf{DS}$.

We set the probability mass function of the distribution $\mathcal{D}_p$ to be

$$\forall S \subseteq V, \qquad \mathcal{D}_p(S) = (1-p) \cdot \mathcal{D}_A(S) + p \cdot \mathcal{D}_B(S).$$

Recall from Lemma 3.4 that $\mathcal{D}_p$ induces a persuasive binary signaling scheme if it satisfies

$$\frac{\mathbb{E}_{S\sim\mathcal{D}_p}\left[\mathsf{Cut}(S, V\setminus S)\right]}{\mathbb{E}_{S\sim\mathcal{D}_p}\left[|V\setminus S|\right]} \geq \frac{\mathbb{E}_{S\sim\mathcal{D}_p}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{S\sim\mathcal{D}_p}\left[|S|\right]}. \tag{5}$$

We analyze each term separately to obtain a sufficient condition for eq. (5) to hold. For brevity, we denote $\beta \triangleq |\mathsf{IS}|$.

- Applying the first property in Lemma 5.2 to $\mathcal{D}_A$ gives

  $$\mathop{\mathbb{E}}_{S\sim\mathcal{D}_p}\left[|S|\right] = (1-p) \cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}\left[|S|\right] + p \cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}\left[|S|\right] = (1-p)\cdot\mathsf{OPT} + p\cdot\beta,$$

  and thus, $\mathbb{E}_{S\sim\mathcal{D}_p}\left[|V\setminus S|\right] = n - \mathbb{E}_{S\sim\mathcal{D}_p}\left[|S|\right] = (1-p)\cdot(n-\mathsf{OPT}) + p\cdot(n-\beta)$.

- For the expected size of the cut, we can lower bound it as below.

  $$\begin{aligned}\mathop{\mathbb{E}}_{S\sim\mathcal{D}_p}\left[\mathsf{Cut}(S, V\setminus S)\right] =& (1-p)\cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}\left[\mathsf{Cut}(S, V\setminus S)\right] + p\cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}\left[\mathsf{Cut}(S, V\setminus S)\right]\\ \geq& (1-p)\cdot(n-2\mathsf{OPT}) + p\cdot(n-|\mathsf{DS}|)\\ \geq& (1-p)\cdot(n-2\mathsf{OPT}) + p\cdot(n-\iota\cdot\mathsf{OPT}).\end{aligned}$$

  In the above equations, the second step uses the third property of Lemma 5.2 and Lemma 5.1. In the last step, we denote $|\mathsf{DS}|$ by $\iota\cdot\mathsf{OPT}$ with $\iota = O(\log n)$ according to Lemma 5.3.

- For the expected size of induced subgraphs, we upper bound it using the second property of Lemma 5.2 and the independence of IS,

  $$\begin{aligned}\mathop{\mathbb{E}}_{S\sim\mathcal{D}_p}\left[\mathsf{Induced}(S)\right] =& (1-p)\cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}\left[\mathsf{Induced}(S)\right] + p\cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}\left[\mathsf{Induced}(S)\right]\\ \leq& (1-p)\left(\mathsf{OPT}^2 + \mathsf{OPT}\right) + p\beta.\end{aligned}$$

Plugging the above bounds into eq. (5), we obtain a sufficient condition for persuasiveness:

$$\frac{(1-p)\,(n-2\mathsf{OPT}) + p\,(n-\iota\cdot\mathsf{OPT})}{(1-p)(n-\mathsf{OPT}) + p\cdot(n-\beta)} \geq \frac{(1-p)(\mathsf{OPT}^2+\mathsf{OPT}) + p\beta}{(1-p)\mathsf{OPT} + p\beta}. \tag{6}$$

For the remaining proof, we assume $\mathsf{OPT} \leq \sqrt{n}/C$ and $\beta \geq C\sqrt{n}\cdot\mathsf{OPT}$ for a constant $C = 10$. The first assumption is WLOG because if $\mathsf{OPT} > \sqrt{n}/C$, Corollary 3.5 guarantees a cost of $O(\frac{n^2}{n+2m}) = O(n)$ by not sending any signals, which is already an $O(\sqrt{n})$ approximation to OPT. The second assumption is WLOG because if $\beta = |\mathsf{IS}| < C\sqrt{n}\cdot\mathsf{OPT}$, we can expand IS into a maximal independent set of V by including some additional vertices from DS. The size of the resulting maximal independent set is at most $|\mathsf{IS}| + |\mathsf{DS}| = \beta + \iota\cdot\mathsf{OPT} = O(\sqrt{n}+\log n)\cdot\mathsf{OPT}$. Therefore, sending signal 1 to the resulting maximal independent set gives a persuasive binary signaling scheme with cost $O(\sqrt{n}\cdot\mathsf{OPT})$.

Under these two additional assumptions, there always exists $p \in [0,1]$ that satisfies eq. (6) — for example, taking $p = 1$ satisfies eq. (6) by making the left-hand side > 1 and the right-hand side $= 1$. However, since

$$\mathsf{Cost}(\mathcal{D}_p) \leq \mathop{\mathbb{E}}_{S\sim\mathcal{D}_p}\left[|S|\right] = (1-p)\cdot\mathsf{OPT} + p\cdot\beta,$$

the larger p is, the more costly the signaling scheme would be. Therefore, we seek the smallest p that satisfies eq. (6). To do so, we continue simplifying eq. (6) by subtracting 1 from both sides:

$$(6) \iff \frac{(1-p)\left(-\mathsf{OPT}\right) + p\left(\beta - \iota \cdot \mathsf{OPT}\right)}{(1-p)(n-\mathsf{OPT}) + p \cdot (n-\beta)} \geq \frac{(1-p)\mathsf{OPT}^2}{(1-p)\mathsf{OPT} + p\beta}.$$

Dividing $(1-p)$ on both denominators and enumerators and setting $\kappa = p/(1-p)$ gives us:

$$\begin{aligned}
(6) \iff & \frac{-\mathsf{OPT} + \kappa\left(\beta - \iota \cdot \mathsf{OPT}\right)}{(n-\mathsf{OPT}) + \kappa(n-\beta)} \geq \frac{\mathsf{OPT}^2}{\mathsf{OPT} + \kappa\beta} \\
\iff & \left(-\mathsf{OPT} + \kappa(\beta - \iota \cdot \mathsf{OPT})\right)\left(\mathsf{OPT} + \kappa\beta\right) \geq \left((n-\mathsf{OPT}) + \kappa(n-\beta)\right)\mathsf{OPT}^2 \\
\iff & \kappa^2\beta(\beta - \iota \cdot \mathsf{OPT}) \geq \mathsf{OPT}^2(n - \mathsf{OPT} + \kappa(n-\beta)) + \mathsf{OPT}^2 - \kappa((\beta - \iota \cdot \mathsf{OPT})\mathsf{OPT} - \beta\mathsf{OPT}) \\
\iff & \kappa^2\beta(\beta - \iota \cdot \mathsf{OPT}) \geq \mathsf{OPT}^2(n - \mathsf{OPT} + \kappa(n-\beta+\iota) + 1).
\end{aligned}$$

In the last equation above, we have $\beta - \iota \cdot \mathsf{OPT} \geq \frac{\beta}{2}$ for all sufficiently large n, by the assumption that $\beta \geq C\sqrt{n} \cdot \mathsf{OPT}$ and the fact that $\iota = O(\log n)$. In addition, we upper bound the right-hand side by dropping the negative terms. This gives the following sufficient condition:

$$(6) \Longleftarrow \frac{\kappa^2\beta^2}{2} \geq (2 + 2\kappa)n\mathsf{OPT}^2,$$

which can be satisfied by choosing $\kappa = \Theta(\sqrt{n}\mathsf{OPT}/\beta)$. In particular, under this choice, we have $\kappa = O(1)$ due to the assumption that $\beta \geq C\sqrt{n}\mathsf{OPT}$, so both the left- and right-hand sides are of the order $\Theta(n\mathsf{OPT}^2)$. This gives us the choice of p via $\kappa = \frac{p}{1-p}$.

Finally, by Lemma 3.4, the cost of the resulting signaling scheme satisfies

$$\mathsf{Cost}(\mathcal{D}_p) \leq \mathop{\mathbb{E}}_{S \sim \mathcal{D}_p}[|S|] = (1-p) \cdot \mathsf{OPT} + p \cdot \beta \leq \mathsf{OPT} + \kappa\beta \leq O(\sqrt{n} \cdot \mathsf{OPT}).$$

□

5.3 Matching the cost of full revelation

Towards proving Theorem 1.3, in this section, we give a binary signaling scheme with a cost of exactly $\mathsf{OPT}^{\mathsf{stable}}$.

Theorem 5.5. *For any unit-weight graph G, there is a binary signaling scheme with cost $\mathsf{OPT}^{\mathsf{stable}}$.*

Note that Theorem 5.5 would be trivial if we allow signaling schemes with a large signal space. This is because, for any optimal stable solution $\boldsymbol{\theta}^\star \in [0,1]^V$, labeling each vertex $v \in V$ with $\theta_v^\star$ gives a persuasive signaling scheme with cost $\mathsf{OPT}^{\mathsf{stable}}$. Indeed, the interesting aspect of Theorem 5.5 is that, while $\boldsymbol{\theta}^\star$ might contain many different entries, it is possible to achieve the same cost, via signaling, using only two different values.

Recall that Lemma 3.6 states that any stable solution in a unit-weight graph can be written as the marginal of a distribution over independent sets. We first show that the lemma immediately implies Theorem 5.5.

Proof of Theorem 5.5. Let $\boldsymbol{\theta}^\star \in [0,1]^V$ be an optimal stable solution for graph $G = (V, E)$ with $\|\boldsymbol{\theta}^\star\|_1 = \mathsf{OPT}^{\mathsf{stable}}$. We consider the binary signaling scheme defined by distribution $\mathcal{D} \in \Delta(2^V)$ obtained from $\boldsymbol{\theta}^\star$ via Lemma 3.6.

By Lemma 3.4, to verify the persuasiveness of the binary signaling scheme, it suffices to show that

$$\frac{\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Cut}(S, V\setminus S)\right]}{\mathbb{E}_{S\sim\mathcal{D}}\left[|V\setminus S|\right]} \geq \frac{\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]}.$$

Since every S in the support of $\mathcal{D}$ is an independent set, we have $\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right] = \mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]$. It remains to show that $\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Cut}(S, V\setminus S)\right] \geq \mathbb{E}_{S\sim\mathcal{D}}\left[|V\setminus S|\right]$.

Again, since S is always an independent set, the cut size $|\mathsf{Cut}(S, V\setminus S)|$ is equal to $\sum_{v\in S}\deg(v)$. Then, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[|\mathsf{Cut}(S, V\setminus S)|\right] &= \sum_{v\in V}\Pr_{S\sim\mathcal{D}}\left[v\in S\right]\cdot\deg(v) && (S \text{ is an independent set})\\
&= \sum_{v\in V}\theta^\star_v\cdot\deg(v) && (\text{property of } \mathcal{D} \text{ from Lemma 3.6})\\
&= \sum_{v\in V}\sum_{u\in N(v)}\theta^\star_v = \sum_{v\in V}\sum_{u\in N(v)}\theta^\star_u\\
&\geq \sum_{v\in V}(1-\theta^\star_v) && (\boldsymbol{\theta}^\star \text{ is a feasible solution})\\
&= n - \|\boldsymbol{\theta}^\star\|_1 = \mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[|V\setminus S|\right].
\end{aligned}$$

The fourth step holds since both the third line and the fourth line are equal to $\sum_{\{u,v\}\in E}(\theta^\star_u+\theta^\star_v)$. This proves persuasiveness.

By Lemma 3.4, the cost of this scheme is exactly

$$\frac{\left(\mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]\right)^2}{\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right]} = \mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[|S|\right] = \sum_{v\in V}\Pr_{S\sim\mathcal{D}}\left[v\in S\right] = \sum_{v\in V}\theta^\star_v = \|\boldsymbol{\theta}^\star\|_1 = \mathsf{OPT}^{\mathsf{stable}}.$$

The third step above applies the property of $\mathcal{D}$ guaranteed by Lemma 3.6. □

Now we prove Lemma 3.6 via rounding the stable solution through a correlated rounding method. The method is referred to as the *competing exponential clocks* in the context of approximation algorithms (e.g., [Buchbinder et al., 2013]).

Proof of Lemma 3.6. Let $V' \coloneqq \{v\in V : \theta_v > 0\}$ be the support of the stable solution $\boldsymbol{\theta}\in[0,1]^V$. We define $\mathcal{D}\in\Delta(2^V)$ as the distribution of the set $S\subseteq V$ generated by the following process:

- For each $v\in V'$, independently draw $X_v\sim\mathsf{Exponential}(\theta_v)$.
- Choose the set S as $\left\{v\in V' : X_v < \min_{u\in N(v)\cap V'} X_v\right\}$.

It is clear that S must be an independent set: if S contains two neighbouring vertices u and v, we must have both $X_u < X_v$ and $X_v < X_u$, a contradiction.

For every $v\in V\setminus V'$, we clearly have $\Pr_{S\sim\mathcal{D}}\left[v\in S\right] = 0 = \theta_v$. To see that $\Pr_S\left[v\in S\right] = \theta_v$ holds for every $v\in V'$, we use the condition $\theta_v + \sum_{u\in N(v)\cap V'}\theta_u = 1$, which follows from stability. Then, the property of the exponential distribution shows that $\min_{u\in N(v)\cap V'} X_u$ follows the exponential distribution $\mathsf{Exponential}(1-\theta_v)$ and is independent of X_u. Therefore, we have

$$\Pr\left[v\in S\right] = \Pr\left[X_v < \min_{u\in N(v)\cap V'} X_u\right]$$

$$= \mathop{\mathbb{E}}_{X_v \sim \mathsf{Exponential}(\theta_v)} [\exp(-(1-\theta_v)X_v)]$$
$$= \int_0^{+\infty} \theta_v \exp(-\theta_v x) \cdot \exp(-(1-\theta_v)x) \, \mathrm{d}x = \theta_v,$$

as desired. □

5.4 Strict improvement upon full revelation

In this section, we build on Theorem 5.5 and characterize conditions under which signaling results in a *strict* improvement compared to $\mathsf{OPT}^{\mathsf{stable}}$ and the amount of improvement achieved.

Theorem 5.6 (Formal version of Theorem 1.3)**.** *On any unit-weight graph that satisfies* $\mathsf{OPT}^{\mathsf{stable}} > \mathsf{OPT}$*, there exists a binary signaling scheme* $\mathcal{D}_\varphi$ *with* $\mathsf{Cost}(\varphi) = \mathsf{OPT}^{\mathsf{stable}} - \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT})$*, where*

$$\epsilon = \max\left\{\min\left\{\frac{1}{2}, \frac{\mathsf{PoS}-1}{\mathsf{PoS} + \frac{2n}{\mathsf{PoS}+1}}\right\}, \frac{\mathsf{PoS}-1}{\mathsf{PoS} + (n - \mathsf{OPT})}\right\} > 0, \tag{7}$$

and $\mathsf{PoS} = \mathsf{OPT}^{\mathsf{stable}}/\mathsf{OPT}$ *is the price of stability.*

To prove Theorem 5.6, we first present two technical lemmas that connect the "wastefulness" of the optimal stable solution to the gap $\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}$. Formally, we call a feasible solution $\boldsymbol{\theta} \in [0,1]^V$ *wasteful* if it satisfies $W\boldsymbol{\theta} \neq \mathbf{1}$. Recall that feasibility requires $W\boldsymbol{\theta} \geq \mathbf{1}$ coordinate-wise. Therefore, a wasteful solution is one in which at least one of the tasks receives strictly more contribution than what is required. Our first lemma states that, even on general weighted graphs, every non-wasteful feasible solution must be optimal. In particular, assuming $\mathsf{OPT}^{\mathsf{stable}} > \mathsf{OPT}$, the lemma below implies that the optimal stable solution must be wasteful, and this wastefulness allows us to design a signaling scheme with a cost strictly below $\mathsf{OPT}^{\mathsf{stable}}$.

Lemma 5.7. *In a general weighted graph, if there exists a non-wasteful feasible solution* $\boldsymbol{\theta} \in [0,1]^V$*, it holds that* $\mathsf{OPT} = \|\boldsymbol{\theta}\|_1$*.*

Proof. Recall from the definition that OPT can be computed by the following linear program:

$$\begin{aligned} \min_{\boldsymbol{\theta} \in \mathbb{R}^n} \quad & \mathbf{1}^\top \boldsymbol{\theta} \\ \text{s.t.} \quad & W\boldsymbol{\theta} \geq \mathbf{1}; \\ & \boldsymbol{\theta} \geq \mathbf{0}. \end{aligned} \tag{primal LP}$$

Since $W = W^\top$, the dual linear program of (primal LP) is given by:

$$\begin{aligned} \max_{\boldsymbol{\phi} \in \mathbb{R}^n} \quad & \mathbf{1}^\top \boldsymbol{\phi} \\ \text{s.t.} \quad & W\boldsymbol{\phi} \leq \mathbf{1}; \\ & \boldsymbol{\phi} \geq \mathbf{0}. \end{aligned} \tag{dual LP}$$

Since $\boldsymbol{\theta}$ is feasible and non-wasteful, we have $W\boldsymbol{\theta} = \mathbf{1}$, which implies that $\boldsymbol{\theta}$ is feasible solution for both (primal LP) and (dual LP), and achieves the same objective value of $\mathbf{1}^\top\boldsymbol{\theta} = \|\boldsymbol{\theta}\|_1$. Therefore, $\boldsymbol{\theta}$ must be an optimal solution for (primal LP), and we have $\mathsf{OPT} = \|\boldsymbol{\theta}\|_1$. □

The second lemma quantitatively captures the gap $\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}$ in terms of the "wastefulness" of the optimal stable solution $\boldsymbol{\theta}^\star$, i.e., the gap $\|W\boldsymbol{\theta}^\star\|_1 - n$, in the case of unit-weight graphs. Again, we state a more general result, and will apply it specifically to the optimal stable solution.

Lemma 5.8. *Let $\boldsymbol{\theta} \in [0,1]^V$ be a feasible solution on a unit-weight graph. Then,*

$$\|W\boldsymbol{\theta}\|_1 \geq n + \|\boldsymbol{\theta}\|_1 - \mathsf{OPT}.$$

We prove Lemma 5.8 in Appendix D.2. The high-level idea of the proof is to convert $\boldsymbol{\theta}$, which is a feasible solution of (primal LP), into a feasible solution $\boldsymbol{\phi}^\star$ of (dual LP) by lowering the contributions that enter wasteful coordinates. Through our careful construction, we ensure the resulting decrease in the objective value (i.e., $\mathbf{1}^\top\boldsymbol{\theta} - \mathbf{1}^\top\boldsymbol{\phi}^\star$) is upper bounded by the wastefulness of $\boldsymbol{\theta}$ (i.e., $\|W\boldsymbol{\theta}\|_1 - n$). Finally, we establish the lemma by invoking weak duality to show that $\mathbf{1}^\top\boldsymbol{\phi}^\star \leq \mathsf{OPT}$.

Now we prove Theorem 5.6 by constructing a signaling scheme that randomizes between an independent rounding of OPT and the signaling scheme in Theorem 5.5.

Proof of Theorem 5.6. Let $\boldsymbol{\theta}^\star \in [0,1]^V$ be an optimal solution with cost OPT, and $\boldsymbol{\theta} \in [0,1]^V$ be an optimal stable solution with cost $\mathsf{OPT}^{\mathsf{stable}}$. Then, let $\mathcal{D}_A$ be the distribution defined by rounding $\boldsymbol{\theta}^\star$ independently, and $\mathcal{D}_B$ be the distribution obtained from rounding $\boldsymbol{\theta}$ according to Lemma 3.6. We define a parametrized family of distributions $\mathcal{D}_\epsilon \in \Delta(2^V)$ for $\epsilon \in [0,1]$ as

$$\forall S \subseteq V, \qquad \mathcal{D}_\epsilon(S) = \epsilon \cdot \mathcal{D}_A(S) + (1-\epsilon) \cdot \mathcal{D}_B(S).$$

Recall from Lemma 3.4 that the distribution $\mathcal{D}_\epsilon$ induces a persuasive binary signaling scheme if the following inequality holds:

$$\frac{\mathbb{E}_{S\sim\mathcal{D}_\epsilon}\left[\mathsf{Cut}(S, V\setminus S)\right]}{\mathbb{E}_{S\sim\mathcal{D}_\epsilon}\left[|V\setminus S|\right]} \geq \frac{\mathbb{E}_{S\sim\mathcal{D}_\epsilon}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{S\sim\mathcal{D}_\epsilon}\left[|S|\right]}. \tag{8}$$

We analyze each term separately to obtain a sufficient condition for eq. (8) to hold.

- For the expected size of $|S|$ and $|V \setminus S|$, we combine Lemma 3.6 with the first property of Lemma 5.2 to get

$$\begin{aligned}
\mathop{\mathbb{E}}_{S\sim\mathcal{D}_\epsilon}\left[|S|\right] &= \epsilon \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}\left[|S|\right] + (1-\epsilon) \mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}\left[|S|\right] = \epsilon\mathsf{OPT} + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}};\\
\mathop{\mathbb{E}}_{S\sim\mathcal{D}_\epsilon}\left[|V\setminus S|\right] &= n - \mathop{\mathbb{E}}_{S\sim\mathcal{D}_\epsilon}\left[|S|\right] = n - \mathsf{OPT}^{\mathsf{stable}} + \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}).
\end{aligned}$$

- For the expected size of the cut, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}\left[\mathsf{Cut}(S, V\setminus S)\right] &= \mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}\left[\sum_{v\in S}\deg(v)\right] && (S\sim\mathcal{D}_B \text{ is an independent set})\\
&= \sum_{v\in V}\Pr_{S\sim\mathcal{D}_B}\left[v\in S\right]\cdot\deg(v) && (\text{linearity of expectation})\\
&= \sum_{v\in V}\theta_v\cdot\deg(v) && (\text{property of } \mathcal{D}_B \text{ from Lemma 3.6})\\
&= \boldsymbol{\theta}^\top(W\mathbf{1} - \mathbf{1}) && (W \text{ is the adjacency matrix of } G)\\
&= \|W\boldsymbol{\theta}\|_1 - \|\boldsymbol{\theta}\|_1 && (\boldsymbol{\theta}^\top W\mathbf{1} = (W\boldsymbol{\theta})^\top\mathbf{1} = \|W\boldsymbol{\theta}\|_1)
\end{aligned}$$

$$\geq (n + \mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}) - \mathsf{OPT}^{\mathsf{stable}} \qquad \text{(Lemmas 3.6 and 5.8)}$$
$$= n - \mathsf{OPT}.$$

Combining the above inequalities with the third property of Lemma 5.2 gives us

$$\begin{aligned}\mathop{\mathbb{E}}_{S\sim\mathcal{D}_\epsilon}[\mathsf{Cut}(S, V\setminus S)] &= \epsilon \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}[\mathsf{Cut}(S, V\setminus S)] + (1-\epsilon)\mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}[\mathsf{Cut}(S, V\setminus S)]\\ &\geq \epsilon\,(n - 2\mathsf{OPT}) + (1-\epsilon)(n - \mathsf{OPT})\\ &= n - (1+\epsilon)\mathsf{OPT}.\end{aligned}$$

- For the expected size of induced subgraphs, we upper bound it using the second property of Lemma 5.2 and the independence of subsets drawn from $\mathcal{D}_B$ (Lemma 3.6).

$$\begin{aligned}\mathop{\mathbb{E}}_{S\sim\mathcal{D}_\epsilon}[\mathsf{Induced}(S)] &= \epsilon \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}[\mathsf{Induced}(S)] + (1-\epsilon)\mathop{\mathbb{E}}_{S\sim\mathcal{D}_B}[\mathsf{Induced}(S)]\\ &\leq \epsilon\left(\mathsf{OPT}^2 + \mathsf{OPT}\right) + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}}.\end{aligned}$$

Plugging the above bounds into Equation (8), we obtain its sufficient condition:

$$\frac{n - (1+\epsilon)\mathsf{OPT}}{n - \mathsf{OPT}^{\mathsf{stable}} + \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT})} \geq \frac{\epsilon\left(\mathsf{OPT}^2 + \mathsf{OPT}\right) + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}}}{\epsilon\mathsf{OPT} + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}}} \tag{9}$$

We first show the existence of $\epsilon \in (0, 1]$ to make eq. (9) hold. Note that when $\epsilon = 0$, it reduces to

$$\frac{n - \mathsf{OPT}}{n - \mathsf{OPT}^{\mathsf{stable}}} \geq \frac{\mathsf{OPT}^{\mathsf{stable}}}{\mathsf{OPT}^{\mathsf{stable}}} = 1,$$

which holds with a strict inequality because $\mathsf{OPT}^{\mathsf{stable}} > \mathsf{OPT}$. Since both sides are continuous in ϵ, there should exist a small neighborhood of ϵ around 0 that all satisfies eq. (9).

Moreover, according to Lemma 3.4, the cost of the signaling scheme induced by $\mathcal{D}_\epsilon$ is

$$\mathsf{Cost}(\mathcal{D}_\epsilon) \leq \mathop{\mathbb{E}}_{S\sim\mathcal{D}_\epsilon}[|S|] = \epsilon\mathsf{OPT} + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}} = \mathsf{OPT}^{\mathsf{stable}} - \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}).$$

Therefore, to prove Theorem 5.6, it remains to show that the choice of ϵ in eq. (7) satisfies eq. (9). We start with simplifying the condition by subtracting 1 on both sides:

$$(9) \iff \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT} - \epsilon\mathsf{OPT}^{\mathsf{stable}}}{n - \mathsf{OPT}^{\mathsf{stable}} + \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT})} \geq \frac{\epsilon\mathsf{OPT}^2}{\epsilon\mathsf{OPT} + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}}}.$$

On the one hand, since lower bounding the left-hand side and upper bounding the right-hand side results in a sufficient condition, we use $\epsilon\mathsf{OPT} + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}} \geq \mathsf{OPT}$ and obtain

$$\begin{aligned}(9) &\Longleftarrow \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT} - \epsilon\mathsf{OPT}^{\mathsf{stable}}}{n - \mathsf{OPT}} \geq \frac{\epsilon\mathsf{OPT}^2}{\mathsf{OPT}}\\ &\iff \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{n - \mathsf{OPT}} \geq \epsilon\left(\mathsf{OPT} + \frac{\mathsf{OPT}^{\mathsf{stable}}}{n - \mathsf{OPT}}\right)\\ &\iff \epsilon \leq \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{\mathsf{OPT}(n - \mathsf{OPT}) + \mathsf{OPT}^{\mathsf{stable}}} = \frac{\mathsf{PoS} - 1}{\mathsf{PoS} + (n - \mathsf{OPT})},\end{aligned}$$

where $\mathsf{PoS} = \mathsf{OPT}^{\mathsf{stable}}/\mathsf{OPT}$ is the price of stability. This justifies the second choice of ϵ in Equation (7).

On the other hand, under the condition of $\epsilon \le \frac{1}{2}$, we have $\epsilon\mathsf{OPT} + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}} \ge \frac{\mathsf{OPT}+\mathsf{OPT}^{\mathsf{stable}}}{2}$. We use this to provide an alternative sufficient condition to eq. (9).

$$
\begin{aligned}
(9) \Longleftarrow\ & \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT} - \epsilon\mathsf{OPT}^{\mathsf{stable}}}{n} \ge \frac{\epsilon\mathsf{OPT}^2}{\frac{\mathsf{OPT}+\mathsf{OPT}^{\mathsf{stable}}}{2}} \\
\Longleftrightarrow\ & \mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT} \ge \epsilon\left(\frac{2n\mathsf{OPT}^2}{\mathsf{OPT}+\mathsf{OPT}^{\mathsf{stable}}} + \mathsf{OPT}^{\mathsf{stable}}\right) \\
\Longleftrightarrow\ & \epsilon \le \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{\frac{2n\mathsf{OPT}^2}{\mathsf{OPT}+\mathsf{OPT}^{\mathsf{stable}}} + \mathsf{OPT}^{\mathsf{stable}}} = \frac{\mathsf{PoS}-1}{\mathsf{PoS} + \frac{2n}{\mathsf{PoS}+1}},
\end{aligned}
$$

which justifies the first option of ϵ when $\epsilon \le \frac{1}{2}$. □

6 Lower Bounds for Unit-Weight Graphs

In this section, we prove the lower bound part of Theorem 1.1 by showing that on the double-star graph in Figure 1, every persuasive signaling scheme must have an $\Omega(\sqrt{n})$ cost, while the optimal total workload is $O(1)$. Therefore, the $O(\sqrt{n})$ approximation guarantee in Theorem 1.1 cannot be significantly improved, even when non-binary schemes are allowed.

Then, we prove a more general lower bound, which states that for any n and $k \in [2, n]$, there is an instance on which $\mathsf{OPT} = \Theta(k)$, $\mathsf{OPT}^{\mathsf{stable}} = \Theta(n)$, and no persuasive signaling scheme can achieve a cost that is much better than $\min\{k\sqrt{n}, n\}$. Recall that by Theorems 1.1 and 1.3, on this instance there is a persuasive binary signaling scheme with cost $O(\min\{\mathsf{OPT}\cdot\sqrt{n}, \mathsf{OPT}^{\mathsf{stable}}\}) = O(\min\{k\sqrt{n}, n\})$. Therefore, for a wide range of OPT, between the two signaling schemes that we develop for unit-weight graphs, the better one is essentially optimal.

6.1 Lower bound for the double-star graph

Theorem 6.1 (Lower bound part of Theorem 1.1)**.** *On the double-star graph with $n = 2k+2$ vertices, every persuasive signaling scheme has a cost of $\Omega(\sqrt{n})$.*

Proof. For clarity, we rename the vertices in the graph with $V = \{(c,1),(c,2)\} \cup \{(i,j) : i \in [2], j \in [k]\}$, where (c,i) is the center of the i-th star, and (i,j) is the j-th leaf in the i-th star.

Fix a persuasive signaling scheme for the graph. By Lemma 2.1, it is without loss of generality to assume that the scheme is identity-independent and is specified by $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ on signal space $\mathcal{S}$. In other words, the signaling scheme samples $(s_{c,1}, s_{c,2}, s_{1,1}, \ldots, s_{1,k}, s_{2,1}, \ldots, s_{2,k}) \in \mathcal{S}^V$ from $\mathcal{D}_\varphi$, labels the two centers with $s_{c,1}$ and $s_{c,2}$, and labels the j-th leaf in the i-th star with $s_{i,j}$. For brevity, we use x and y as shorthands for $s_{c,1}$ and $s_{c,2}$, and let random variable sum denote $x + y + \sum_{i=1}^{2}\sum_{j=1}^{k} s_{i,j}$. Then, the cost of the scheme is given by $\mathbb{E}_{\mathcal{D}_\varphi}[\mathsf{sum}]$.

Let $C \coloneqq 3$. We will show that, under the assumption that $\mathbb{E}_{\mathcal{D}_\varphi}[\mathsf{sum}] \le \sqrt{k}/C$, the persuasiveness of $\mathcal{D}_\varphi$ implies $\mathbb{E}_{\mathcal{D}_\varphi}[\mathsf{sum}] = \Omega(\sqrt{k})$. Therefore, $\mathbb{E}_{\mathcal{D}_\varphi}[\mathsf{sum}] = \Omega(\sqrt{k}) = \Omega(\sqrt{n})$ always holds.

By Lemma 3.2, the persuasiveness of $\mathcal{D}_\varphi$ implies that for every $\theta \in \mathcal{S}$,

$$0 \le \Delta_\theta = \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v = \theta\right]\left(\sum_{u\in N(v)} s_u - (1-s_v)\right)\right].$$

Summing over all $\theta \in \mathcal{S}$ gives

$$0 \le \sum_{\theta\in\mathcal{S}} \Delta_\theta = \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v\in V}\left(s_v + \sum_{u\in N(v)} s_u - 1\right)\right] = \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{u\in V} s_u\left(1+|N(u)|\right)\right] - n.$$

Since $|N(u)| = 1$ when u is a leaf and $|N(u)| = k+1$ when u is a center, we have

$$n \le \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{u\in V} s_u\left(1+|N(u)|\right)\right] = 2\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v\in V} s_v\right] + k\cdot\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[s_{c,1}+s_{c,2}\right] = 2\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\mathsf{sum}\right] + k\cdot\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[x+y\right],$$

which further implies $\mathbb{E}_{\mathcal{D}_\varphi}\left[x+y\right] \ge \frac{n-2\,\mathbb{E}_{\mathcal{D}_\varphi}[\mathsf{sum}]}{k}$. By the assumption that $\mathbb{E}_{\mathcal{D}_\varphi}\left[\mathsf{sum}\right] \le \sqrt{k}/C$, we have

$$\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[x+y\right] \ge \frac{2k+2-2\sqrt{k}/C}{k} \ge 2 - \frac{2}{C\sqrt{k}}.$$

Let $\epsilon \coloneqq 1/\sqrt{k}$. Since $2-(x+y)$ is always non-negative, by Markov's inequality, we have

$$\begin{aligned}\Pr_{\mathcal{D}_\varphi}\left[x+y < 2-\epsilon\right] = \Pr_{\mathcal{D}_\varphi}\left[2-(x+y) > \epsilon\right] &\le \frac{\mathbb{E}_{\mathcal{D}_\varphi}\left[2-(x+y)\right]}{\epsilon} \\ &\le \frac{2-\left(2-\frac{2}{C\sqrt{k}}\right)}{1/\sqrt{k}} = \frac{2}{C}. \qquad (\mathbb{E}_{\mathcal{D}_\varphi}\left[x+y\right] \ge 2-2/(C\sqrt{k}))\end{aligned}$$

Equivalently, we have $\Pr_{\mathcal{D}_\varphi}\left[x+y \ge 2-\epsilon\right] \ge 1-2/C$.

As long as $k \ge 2$, we have $1-\epsilon = 1-1/\sqrt{k} > 0$, so Lemma 3.2 implies $\sum_{\theta\in\mathcal{S}} \Delta_\theta \cdot \mathbb{1}\left[\theta \ge 1-\epsilon\right] = 0$, which is equivalent to

$$\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v \ge 1-\epsilon\right]\left(s_v + \sum_{u\in N(v)} s_u - 1\right)\right] = 0.$$

The left-hand side above can be written as the sum of the following four terms:

- $T_1 \coloneqq \mathbb{E}_{\mathcal{D}_\varphi}\left[\mathbb{1}\left[x+y \ge 2-\epsilon\right]\cdot\left(2(x+y-1) + \sum_{i=1}^{2}\sum_{j=1}^{k} s_{i,j}\right)\right]$, the contribution from the two centers when $x+y \ge 2-\epsilon$. Note that this condition implies both $x \ge 1-\epsilon$ and $y \ge 1-\epsilon$.
- $T_2 \coloneqq \mathbb{E}_{\mathcal{D}_\varphi}\left[\mathbb{1}\left[x+y < 2-\epsilon \wedge x \ge 1-\epsilon\right]\left(x+y-1+\sum_{i=1}^{k} s_{1,i}\right)\right]$, the contribution from the center of the first star, when $x+y < 2-\epsilon$.
- $T_3 \coloneqq \mathbb{E}_{\mathcal{D}_\varphi}\left[\mathbb{1}\left[x+y < 2-\epsilon \wedge y \ge 1-\epsilon\right]\left(x+y-1+\sum_{i=1}^{k} s_{2,i}\right)\right]$, the contribution from the center of the second star, when $x+y < 2-\epsilon$.
- $T_4 \coloneqq \mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{v\in V\setminus\{(c,1),(c,2)\}} \mathbb{1}\left[s_v \ge 1-\epsilon\right]\left(s_v + \sum_{u\in N(v)} s_u - 1\right)\right]$, the contribution from the leaves.

Recall that we showed $\Pr_{\mathcal{D}_\varphi}[x+y \geq 2-\epsilon] \geq 1-2/C$ earlier, so the first term T_1 satisfies

$$\begin{aligned} T_1 &\geq \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\mathbb{1}[x+y \geq 2-\epsilon] \cdot 2(x+y-1)\right] \geq \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\mathbb{1}[x+y \geq 2-\epsilon] \cdot 2(2-\epsilon-1)\right] \\ &= \Pr_{\mathcal{D}_\varphi}[x+y \geq 2-\epsilon] \cdot (2-2\epsilon) \geq (1-2/C) \cdot 2 \cdot (1-\epsilon) \geq 2 \cdot (1-2/C-\epsilon). \end{aligned}$$

For the second term, note that when $x \geq 1-\epsilon$ holds, $x+y-1+\sum_{i=1}^{k} s_{1,i} \geq x-1 \geq -\epsilon$. Thus,

$$T_2 \geq -\epsilon \cdot \Pr_{\mathcal{D}_\varphi}[x+y < 2-\epsilon \wedge x \geq 1-\epsilon] \geq -\epsilon \cdot \Pr_{\mathcal{D}_\varphi}[x+y < 2-\epsilon] \geq -2\epsilon/C.$$

By symmetry, we also have $T_3 \geq -2\epsilon/C$. For the same reason, the last term can be lower bounded by $-\epsilon \cdot \mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{v \in V \setminus \{(c,1),(c,2)\}} \mathbb{1}[s_v \geq 1-\epsilon]\right]$. Note that

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}[\mathsf{sum}] &\geq \sum_{v \in V \setminus \{(c,1),(c,2)\}} \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}[s_v] \geq \sum_{v \in V \setminus \{(c,1),(c,2)\}} (1-\epsilon) \cdot \Pr_{\mathcal{D}_\varphi}[s_v \geq 1-\epsilon] \\ &= (1-\epsilon) \cdot \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v \in V \setminus \{(c,1),(c,2)\}} \mathbb{1}[s_v \geq 1-\epsilon]\right], \end{aligned}$$

so we have $T_4 \geq -\frac{\epsilon}{1-\epsilon}\mathbb{E}_{\mathcal{D}_\varphi}[\mathsf{sum}]$.

Recall that $T_1+T_2+T_3+T_4 = 0$. Therefore, combining the four lower bounds gives

$$0 \geq 2 \cdot (1-2/C-\epsilon) - 2\epsilon/C - 2\epsilon/C - \frac{\epsilon}{1-\epsilon}\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}[\mathsf{sum}],$$

which is equivalent to

$$\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}[\mathsf{sum}] \geq \frac{1-\epsilon}{\epsilon} \cdot [2 \cdot (1-2/C-\epsilon) - 4\epsilon/C].$$

By our choice of $C = 3$ and $\epsilon = 1/\sqrt{k}$, the right-hand side above is $\Omega(\sqrt{k}) = \Omega(\sqrt{n})$ for all sufficiently large k. This establishes the $\Omega(\sqrt{n})$ lower bound on the cost. □

6.2 A more general lower bound via k-stars

The previous lower bound gives an instance where $\mathsf{OPT} = O(1)$ and the optimal persuasive signaling scheme has an $\Omega(\sqrt{n})$ cost. Now, we present a more general lower bound in Theorem 6.2 that holds for a wider range of OPT. The proof of Theorem 6.2 is deferred to Appendix E.

Theorem 6.2. *For any $n \geq k \geq 2$, there is a unit-weight graph with $\leq n$ vertices, on which the following hold simultaneously:*

- $\mathsf{OPT} = \Theta(k)$.
- $\mathsf{OPT}^{\mathsf{stable}} = \Omega(n)$.
- *Every persuasive signaling scheme has a cost of $\Omega(\min\{k\sqrt{n}, n\})$.*

When n is sufficiently large, the $\Theta(\cdot)$ and $\Omega(\cdot)$ notations above hide universal constants that do not depend on n and k.

When $k = O(\sqrt{n})$, we consider a generalized version of the double-star construction, in which there are k disjoint stars, each with $\Theta(n/k)$ leaves. The centers of the k stars form a clique. This gives an instance on which $\mathsf{OPT} = k$ and $\mathsf{OPT}^{\mathsf{stable}} = \Omega(n)$. By a similar analysis to that of the double-star graph, we show that every persuasive signaling scheme has an $\Omega(k\sqrt{n})$ cost on this k-star graph.

When $k \gg \sqrt{n}$, the above construction cannot be directly used for establishing an $\Omega(n)$ lower bound. This is because the k-clique formed by the centers already contains $\Omega(k^2) \gg n$ edges; by Corollary 3.5, an $O(n^2/m) = o(n)$ cost can be achieved. Instead, we apply this construction on $n' = (n/k)^2$ and $k' = n/k$ (so that $k' = O(\sqrt{n'})$ still holds), and construct $n/n' = k^2/n$ disjoint copies of such a graph. Intuitively, the optimal solutions in the large graph must be n/n' times those for a single copy. This gives $\mathsf{OPT} = (n/n') \cdot k' = k$, $\mathsf{OPT}^{\mathsf{stable}} = (n/n') \cdot n' = n$, and the cost of every persuasive scheme is lower bounded by $(n/n') \cdot \Omega(n') = \Omega(n)$.

7 Upper Bounds for Weighted Graphs

In this section, we prove upper bounds in weighted graphs using the ternary signaling approach outlined in Section 3.3. Together, they give the upper bound part of Theorem 1.2. We also prove Theorem 1.4, which states that we can achieve a cost strictly lower than $\mathsf{OPT}^{\mathsf{stable}}$ whenever $\mathsf{OPT}^{\mathsf{IR}} < \mathsf{OPT}^{\mathsf{stable}}$.

7.1 Upper bound for approximating $\mathsf{OPT}^{\mathsf{IR}}$

Before presenting the formal theorem statements, we first establish the following technical lemma, which shows that if the graph contains a large set that is "almost independent", we have a good approximation of $\mathsf{OPT}^{\mathsf{IR}}$.

Lemma 7.1. *Suppose that distribution $\mathcal{D} \in \Delta(2^V)$ satisfies*

$$\mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right] \le (1+\gamma) \cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}}\left[|S|\right].$$

Then, there is a persuasive signaling scheme with a cost of

$$O\left(\mathsf{OPT}^{\mathsf{IR}} + \gamma n + \frac{n \cdot \mathsf{OPT}^{\mathsf{IR}}}{\sqrt{\mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]}}\right).$$

Note that if $\gamma = 0$, the condition $\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right] \le (1+\gamma) \cdot \mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right]$ is equivalent to that $\mathcal{D}$ is supported over independent sets of V. When $\gamma > 0$, we allow some edges (with a small total weight) in the sub-graph induced by a random $S \sim \mathcal{D}$, at the cost of an additional γn term in the cost of the resulting signaling scheme.

We include a proof of Lemma 7.1 in Appendix F.1. The proof is based on optimizing the probabilities of sending signals to $\widetilde{\mathsf{DS}}, \widetilde{\mathsf{IS}} \sim \mathcal{D}$, and $V \setminus \widetilde{\mathsf{DS}}$ as well as the values of the three signals, as described in Section 3.3. In particular, let $\boldsymbol{\theta}^\star \in [0,1]^V$ be an optimal solution subject to IR with cost $\|\boldsymbol{\theta}^\star\|_1 = \mathsf{OPT}^{\mathsf{IR}}$. We consider the following family of ternary signaling schemes with parameters $p, q \ge 0$ and $\epsilon, \alpha \in [0,1]$:

- With probability $\frac{1}{1+p+q}$, draw a random set $\widetilde{\mathsf{DS}} \subseteq V$ that includes each $v \in V$ independently with probability $\theta_v^\star$. Label (the vertices in) $\widetilde{\mathsf{DS}}$ with $1-\epsilon$, and label $V \setminus \widetilde{\mathsf{DS}}$ with 0.

- With probability $\frac{p}{1+p+q}$, draw a random set $\widetilde{\mathsf{IS}} \sim \mathcal{D}$. Label $\widetilde{\mathsf{IS}}$ with $1-\epsilon$, and label $V \setminus \widetilde{\mathsf{IS}}$ with 0.
- With probability $\frac{q}{1+p+q}$, draw a random set $\widetilde{\mathsf{DS}} \subseteq V$ that includes each $v \in V$ independently with probability $\theta_v^\star$. Label $V \setminus \widetilde{\mathsf{DS}}$ with α, and label $\widetilde{\mathsf{DS}}$ with 0.

In the next section, we prove the upper bound side of Theorem 1.2 by instantiating Lemma 7.1 with specific choices of $\mathcal{D}$.

In Theorem 7.2, we provide the $O(n^{3/4})$ approximation guarantee for general weighted graphs. The distribution $\mathcal{D}$ is chosen as the distribution of a random subset of the vertices sampled with a carefully chosen probability. The proof of Theorem 7.2 is in Appendix F.2.

Theorem 7.2 (The first upper bound in Theorem 1.2)**.** *In any weighted graph, there exists a ternary signaling scheme that is persuasive and has cost* $O\left(n^{3/4} \cdot \left(\mathsf{OPT}^{\mathsf{IR}}\right)^{1/2}\right)$.

When the edges in the graph have weights lower bounded by $\delta > 0$, we apply Lemma 7.1 with $\gamma = 0$ and choose $\mathcal{D}$ as a degenerate distribution at an independent set of the graph. The proof of Theorem 7.3 is in Appendix F.3.

Theorem 7.3 (The second upper bound in Theorem 1.2)**.** *In any weighted graph, if every edge has a weight of at least* δ*, there is a persuasive ternary signaling scheme with a cost of* $O\left(\left(n \cdot \mathsf{OPT}^{\mathsf{IR}}\right)^{2/3} \cdot \delta^{-1/3}\right)$.

7.2 Strict improvement upon full revelation

In this section, we state and prove a more formal version of Theorem 1.4 regarding strict improvement upon $\mathsf{OPT}^{\mathsf{stable}}$.

Theorem 7.4 (Formal version of Theorem 1.4)**.** *In any weighted graphs, if* $\mathsf{OPT}^{\mathsf{IR}} < \mathsf{OPT}^{\mathsf{stable}}$*, there exists a persuasive signaling scheme that uses at most* $(n+1)$ *signals and achieves a cost of*

$$\mathsf{Cost} = \mathsf{OPT}^{\mathsf{stable}} - \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}^{\mathsf{IR}}), \quad \textit{where } \epsilon = \frac{\mathsf{PoS} - 1}{\mathsf{PoS} - 1 + \frac{\mathsf{OPT}^{\mathsf{IR}}(n - \mathsf{OPT}^{\mathsf{IR}} + 1)}{\mathsf{OPT}^{\mathsf{IR}} + 1}} > 0, \tag{10}$$

and $\mathsf{PoS} = \mathsf{OPT}^{\mathsf{stable}}/\mathsf{OPT}$ *is the price of stability.*

Similar to Theorem 5.6, our proof relies on quantifying the *wastefulness* of any feasible solution. We first present a quantitative version of Lemma 5.7 in the case of weighted graphs.

Lemma 7.5. *Let* $\boldsymbol{\theta}$ *be a feasible solution in a general weighted graph. Then, we have*

$$\|W\boldsymbol{\theta}\|_1 \geq n + \frac{\|\boldsymbol{\theta}\|_1 - \mathsf{OPT}}{\mathsf{OPT}}.$$

The proof of Lemma 7.5 (deferred to Appendix F.4) mirrors that of Lemma 5.8 and follows from a different construction of the feasible dual variables. Leveraging this lemma, Theorem 7.4 is achieved by randomizing between the independent rounding of the optimal IR solution and the degenerate scheme of deterministically sending the optimal stable solution. We prove this theorem in Appendix F.5.

8 Lower Bounds for Weighted Graphs

In this section, we prove the lower bound part of Theorem 1.2. Even when all the edges are of weight 1/2, there are graphs on which $\mathsf{OPT}^{\mathsf{IR}}$ is a constant, yet no signaling scheme achieves a cost lower than $n^{2/3}$. In addition, when restricted to binary signaling schemes, the gap becomes $\Omega(n)$.

We start with the lower bound against binary signaling schemes. In contrast to the unit-weight case, we cannot achieve any non-trivial approximation via a binary signaling scheme as before. This shows that the introduction of a third signal in Theorem 1.2 is necessary. The proof of the following theorem is in Appendix G.1.

Theorem 8.1. *There exists a family of weighted graphs on which (1) all edge weights are* $1/2$*; (2)* $\mathsf{OPT}^{\mathsf{IR}} = O(1)$*; (3) every persuasive binary signaling scheme has a cost of* $\Omega(n)$*.*

The rest of this section is devoted to proving the lower bound part of Theorem 1.2.

Theorem 8.2 (Lower bound part of Theorem 1.2)**.** *There exists a family of weighted graphs on which (1) all edge weights are* $1/2$*; (2)* $\mathsf{OPT}^{\mathsf{IR}} = O(1)$*; (3) every persuasive signaling scheme has a cost of* $\Omega(n^{2/3})$*.*

Recall that when the edge weights are 1/2, our proof of Theorem 7.3 implies a signaling scheme with a cost of

$$O\left(\min\left\{\mathsf{OPT}^{\mathsf{IR}}\sqrt{n+m}, \frac{n^2}{m}\right\}\right),$$

where m is the total edge weight in the graph. In order to establish an $\Omega(n^{2/3})$ gap, we need $\mathsf{OPT}^{\mathsf{IR}} = \Theta(1)$ and $m = \Theta(n^{4/3})$. This naturally leads to the following construction, which is a modified version of the graph in Figure 4:

- The graph contains $n = k^3 + 2$ vertices: two centers and k^3 leaves.
- The k^3 leaves form k^2 disjoint cliques, each of size k.
- Each of the two centers is adjacent to every other vertex in the graph.

We indeed have $\mathsf{OPT}^{\mathsf{IR}} = 2 = O(1)$ (achieved by letting both centers play 1) and $m = \Theta(k^4) = \Theta(n^{4/3})$. In the rest of this section, we show that any persuasive signaling scheme for this graph must have an $\Omega(k^2) = \Omega(n^{2/3})$ cost.

Our proof of Theorem 8.2 can be divided into the following three steps. We provide a sketch in the remainder of this section and offer more details in Appendices G.2 to G.4.

Step 1. Dimensionality reduction. By Lemma 2.1, it suffices to consider an identity-independent persuasive signaling scheme specified by $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ with signal space $\mathcal{S}$. According to Lemma 3.2, the persuasiveness of such a signaling scheme is characterized by $|\mathcal{S}|$ constraints on the slack for each signal, which are linear in the n-dimensional probability simplex. As a result, the optimal signaling scheme with minimum cost can be characterized by a linear program. In Appendix G.2, we show that this is equivalent to considering the projection of $\mathcal{D}_\varphi$ onto a lower-dimensional probability simplex supported on $k+2$ vertices — two centers and a random clique of size k. As a result, we obtain a $k+2$-dimensional distribution $\mathcal{D}'$ and significantly simplify the constraints for persuasiveness.

Step 2. Choosing test functions. In the second step, we aim to use the duality of linear programs to lower bound the cost of $\mathcal{D}'$. We consider two choices of test functions $f : \mathcal{S} \to \mathbb{R}_{\geq 0}$: a constant function $f_1(\theta) = 1$, and a function that puts more weight on larger signal values $f_2(\theta) = \frac{\theta}{1-\theta}$. Since the second test function is ill-defined at $\theta = 1$, we additionally show that it is without loss of generality to consider signaling schemes without the signal 1, as the restriction of distribution $\mathcal{D}'$ to $\mathcal{S} \setminus \{1\}$ incurs at most a constant blow-up in the cost. With this, the outputs of the two test functions essentially serve as two sets of dual variables — they give rise to $\sum_{\theta \in \mathcal{S}} f_1(\theta)\Delta_\theta \geq 0$ and $\sum_{\theta \in \mathcal{S}} f_2(\theta)\Delta_\theta = 0$. See more details in Appendix G.3.

Step 3. Verify the dual feasibility In the final step, we consider the dual variables $\beta_1 f_1(\theta) - \beta_2 f_2(\theta)$ with carefully chosen constants β_1, β_2, and verify the feasibility by establishing the inequality

$$\mathop{\mathbb{E}}_{\mathcal{D}'}[\mathsf{sum}] \geq \sum_{\theta \in \mathcal{S}} (\beta_1 f_1(\theta) - \beta_2 f_2(\theta))\Delta_\theta + \Omega(1) \geq \Omega(1),$$

where sum is the sum of the signals sent to the randomly selected clique. We prove the above inequality by an involved case analysis in Appendix G.4. As a result, this implies that $\mathsf{Cost}(\mathcal{D}_\varphi) \geq k^2 \mathbb{E}_{\mathcal{D}'}[\mathsf{sum}] \geq \Omega(k^2) = \Omega(n^{\frac{2}{3}})$ and completes the proof of Theorem 8.2.

Acknowledgments

This work was supported in part by the National Science Foundation under grant CCF-2145898, by the Office of Naval Research under grant N00014-24-1-2159, a C3.AI Digital Transformation Institute grant, and Alfred P. Sloan fellowship, and a Schmidt Science AI2050 fellowship.

References

Anish Agarwal, Munther Dahleh, and Tuhin Sarkar. A marketplace for data: An algorithmic solution. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC)*, pages 701–726, 2019.

Anish Agarwal, Munther Dahleh, Thibaut Horel, and Maryann Rui. Towards data auctions with externalities. *arXiv preprint arXiv:2003.08345*, 2020.

AlienVault. Alienvault - open threat exchange. https://en.wikipedia.org/wiki/Open_Threat_Exchange, 2012. Accessed: 2024-02-12.

Ricardo Alonso and Odilon Câmara. Persuading voters. *American Economic Review*, 106(11): 3590–3605, 2016.

Elliot Anshelevich, Anirban Dasgupta, Jon Kleinberg, Éva Tardos, Tom Wexler, and Tim Roughgarden. The price of stability for network design with fair cost allocation. *SIAM Journal on Computing*, 38(4):1602–1623, 2008.

Itai Arieli and Yakov Babichenko. Private Bayesian persuasion. *Journal of Economic Theory*, 182: 185–217, 2019.

Robert J Aumann, Michael Maschler, and Richard E Stearns. *Repeated games with incomplete information*. MIT press, 1995.

Pranjal Awasthi, Nika Haghtalab, and Eric Zhao. Open problem: The sample complexity of multi-distribution learning for vc classes. In *The Thirty Sixth Annual Conference on Learning Theory (COLT)*, pages 5943–5949. PMLR, 2023.

Arjada Bardhi and Yingni Guo. Modes of persuasion toward unanimous consent. *Theoretical Economics*, 13(3):1111–1149, 2018.

Dirk Bergemann and Stephen Morris. Robust predictions in games with incomplete information. *Econometrica*, 81(4):1251–1308, 2013.

Dirk Bergemann and Stephen Morris. Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, 11(2):487–522, 2016.

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

Dirk Bergemann, Paul Duetting, Renato Paes Leme, and Song Zuo. Calibrated click-through auctions. In *Proceedings of the ACM Web Conference 2022*, pages 47–57, 2022.

Sarah E Bergen and Tracey L Petryshen. Genome-wide association studies (GWAS) of schizophrenia: does bigger lead to better results? *Current opinion in psychiatry*, 25(2):76, 2012.

Umang Bhaskar, Yu Cheng, Young Kun Ko, and Chaitanya Swamy. Hardness results for signaling in Bayesian zero-sum and network routing games. In *Proceedings of the 17th ACM Conference on Economics and Computation (EC)*, pages 479–496, 2016.

Avrim Blum, Nika Haghtalab, Ariel D Procaccia, and Mingda Qiao. Collaborative PAC learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2389–2398, 2017.

Avrim Blum, Nika Haghtalab, Richard Lanas Phillips, and Han Shao. One for one, or all for all: Equilibria and optimality of collaboration in federated learning. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 1005–1014. PMLR, July 2021a.

Avrim Blum, Shelby Heinecke, and Lev Reyzin. Communication-aware collaborative learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 6786–6793, 2021b.

Niv Buchbinder, Joseph (Seffi) Naor, and Roy Schwartz. Simplex partitioning via exponential clocks and the multiway cut problem. In *Symposium on Theory of Computing (STOC)*, pages 535–544, 2013.

Ozan Candogan. Persuasion in networks: Public signals and k-cores. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 133–134, 2019.

Ozan Candogan. Information design in operations. In *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, pages 176–201. INFORMS, 2020.

Ozan Candogan and Kimon Drakopoulos. Optimal signaling of content accuracy: Engagement vs. misinformation.(2017). *URL https://ssrn. com/abstract*, 3051275, 2017.

Jiecao Chen, Qin Zhang, and Yuan Zhou. Tight bounds for collaborative PAC learning via multiplicative weights. volume 31, 2018.

Junjie Chen, Minming Li, Haifeng Xu, and Song Zuo. Bayesian calibrated click-through auction. *arXiv preprint arXiv:2306.06554*, 2023a.

Yiding Chen, Xiaojin Zhu, and Kirthevasan Kandasamy. Mechanism design for collaborative normal mean estimation. *arXiv preprint arXiv:2306.06351*, 2023b.

Lee Cohen and Han Shao. Incentivized collaboration in active learning. *arXiv preprint arXiv:2311.00260*, 2023.

Sanmay Das, Emir Kamenica, and Renee Mirka. Reducing congestion through information design. In *2017 55th annual allerton conference on communication, control, and computing (allerton)*, pages 1279–1284. IEEE, 2017.

Kate Donahue and Jon Kleinberg. Model-sharing games: Analyzing federated learning under voluntary participation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 5303–5311, 2021a.

Kate Donahue and Jon Kleinberg. Optimality and stability in federated learning: A game-theoretic approach. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:1287–1298, 2021b.

Kate Donahue and Jon Kleinberg. Models of fairness in federated learning. *arXiv preprint arXiv:2112.00818*, 2021c.

Shaddin Dughmi. Algorithmic information structure design: a survey. *ACM SIGecom Exchanges*, 15(2):2–24, 2017.

Shaddin Dughmi and Haifeng Xu. Algorithmic Bayesian persuasion. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 412–425, 2016.

Georgy Egorov and Konstantin Sonin. Persuasion on networks.(2019). *URL https://ssrn.com/abstract*, 3375521, 2019.

Nika Haghtalab, Michael Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 406–419, 2022.

Minbiao Han, Kumar Kshitij Patel, Han Shao, and Lingxiao Wang. On the effect of defections in federated learning and how to prevent them. *arXiv preprint arXiv:2311.16459*, 2023.

Shengyuan Hu, Dung Daniel Ngo, Shuran Zheng, Virginia Smith, and Zhiwei Steven Wu. Federated learning as a network effects game. *arXiv preprint arXiv:2302.08533*, 2023.

Baihe Huang, Sai Praneeth Karimireddy, and Michael I Jordan. Evaluating and incentivizing diverse data contributions in collaborative learning. *arXiv preprint arXiv:2306.05592*, 2023.

Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, Dawn Song, and Costas J Spanos. Towards efficient data valuation based on the Shapley value. In *Artificial Intelligence and Statistics (AISTATS)*, pages 1167–1176. PMLR, 2019.

Emir Kamenica. Bayesian persuasion and information design. *Annual Review of Economics*, 11: 249–272, 2019.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6): 2590–2615, 2011.

Sai Praneeth Karimireddy, Wenshuo Guo, and Michael Jordan. Mechanisms that incentivize data sharing in federated learning. In *Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022)*, 2022.

Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

Xiaoqiang Lin, Xinyi Xu, See-Kiong Ng, Chuan-Sheng Foo, and Bryan Kian Hsiang Low. Fair yet asymptotically equal collaborative learning. *arXiv preprint arXiv:2306.05764*, 2023.

Lingjuan Lyu, Xinyi Xu, Qian Wang, and Han Yu. Collaborative fairness in federated learning. *Federated Learning: Privacy and Incentive*, pages 189–204, 2020.

Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. *Operations Research*, 68(4):1132–1161, 2020.

Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in Bayesian games. *Operations Research*, 70(2):1105–1127, 2022.

Laurent Mathevet and Ina Taneva. Organized information transmission. 2022.

Laurent Mathevet, Jacopo Perego, and Ina Taneva. On information design in games. *Journal of Political Economy*, 128(4):1370–1404, 2020.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282. PMLR, 2017.

Huy Nguyen and Lydia Zakynthinou. Improved algorithms for collaborative PAC learning. volume 31, 2018.

Yiangos Papanastasiou, Kostas Bimpikis, and Nicos Savva. Crowdsourcing exploration. *Management Science*, 64(4):1727–1746, 2018.

Binghui Peng. The sample complexity of multi-distribution learning. *arXiv preprint arXiv:2312.04027*, 2023.

Aviad Rubinstein. Honest signaling in zero-sum games is hard, and lying is even harder. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

Keith E Schnakenberg. Expert advice to a voting body. *Journal of Economic Theory*, 160:102–113, 2015.

Alma Steingart. A group theory of group theory: Collaborative mathematics and the 'uninvention' of a 1000-page proof. *Social Studies of Science*, 42(2):185–213, 2012.

Ina Taneva. Information design. *American Economic Journal: Microeconomics*, 11(4):151–185, 2019.

Xuezhen Tu, Kun Zhu, Nguyen Cong Luong, Dusit Niyato, Yang Zhang, and Juan Li. Incentive mechanisms for federated learning: From economic and game theoretic perspective. *IEEE transactions on cognitive communications and networking*, 8(3):1566–1593, 2022.

Yun Wang. Bayesian persuasion with multiple receivers. *Available at SSRN 2625399*, 2013.

Zhepei Wei, Chuanhao Li, Haifeng Xu, and Hongning Wang. Incentivized communication for federated bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011.

Manxi Wu and Saurabh Amin. Information design for regulating traffic flows under uncertain network state. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 671–678. IEEE, 2019.

Xinyi Xu, Lingjuan Lyu, Xingjun Ma, Chenglin Miao, Chuan Sheng Foo, and Bryan Kian Hsiang Low. Gradient driven rewards to guarantee fairness in collaborative machine learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:16104–16117, 2021.

Xinyi Xu, Zhaoxuan Wu, Arun Verma, Chuan Sheng Foo, and Bryan Kian Hsiang Low. FAIR: Fair collaborative active learning with individual rationality for scientific discovery. In *Artificial Intelligence and Statistics (AISTATS)*, pages 4033–4057. PMLR, 2023.

Yufeng Zhan, Jie Zhang, Zicong Hong, Leijie Wu, Peng Li, and Song Guo. A survey of incentive mechanism design for federated learning. *IEEE Transactions on Emerging Topics in Computing*, 10(2):1035–1044, 2021.

Zihan Zhang, Wenhao Zhan, Yuxin Chen, Simon S Du, and Jason D Lee. Optimal multi-distribution learning. *arXiv preprint arXiv:2312.05134*, 2023.

A Omitted Proofs from Section 2

A.1 Alternative Benchmarks

Recall that in Section 2.2, we defined the benchmarks assuming agents all know the type profile $\boldsymbol{t} = (t_1, \ldots, t_n)$. In this section, we discuss alternative benchmarks for the Bayesian setting where the agents only know the prior distribution τ of $\boldsymbol{t}$ but not its actual realization. In this case, agents evaluate the quality of a collaborative solution in expectation under the type distribution τ. We start by defining the alternative benchmarks $\overline{\mathsf{OPT}}, \overline{\mathsf{OPT}}^{\mathsf{IR}}, \overline{\mathsf{OPT}}^{\mathsf{stable}}$, then show how they can be easily achieved by a no-information signaling scheme and discuss the comparison with the benchmarks defined in Section 2.2.

We define benchmarks $\overline{\mathsf{OPT}}$ as the optimal total workload of a feasible action profile $\boldsymbol{a}$, and $\overline{\mathsf{OPT}}^{\mathsf{IR}}, \overline{\mathsf{OPT}}^{\mathsf{stable}}$ as the optimal workload under additional IR and stability constraints.

$$\begin{aligned}
\overline{\mathsf{OPT}} &= \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_1 \quad \text{s.t.} \quad \mathop{\mathbb{E}}_{\boldsymbol{t}\sim\tau}\left[\mathsf{quality}(\boldsymbol{a};\boldsymbol{t})\right] \geq 1,\ \boldsymbol{a} \geq 0;\\
\overline{\mathsf{OPT}}^{\mathsf{IR}} &= \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_1 \quad \text{s.t.} \quad \mathop{\mathbb{E}}_{\boldsymbol{t}\sim\tau}\left[\mathsf{quality}(\boldsymbol{a};\boldsymbol{t})\right] \geq 1,\ \boldsymbol{a} \geq 0,\ \boldsymbol{a} \leq 1;\\
\overline{\mathsf{OPT}}^{\mathsf{stable}} &= \min_{\boldsymbol{a}} \|\boldsymbol{a}\|_1 \quad \text{s.t.} \quad \mathop{\mathbb{E}}_{\boldsymbol{t}\sim\tau}\left[\mathsf{quality}(\boldsymbol{a};\boldsymbol{t})\right] \geq 1,\ \boldsymbol{a} \geq 0;\\
&\qquad\qquad \forall i \in [n],\ a_i = \min\left\{x \geq 0 \;\middle|\; \mathop{\mathbb{E}}_{\boldsymbol{t}\sim\tau}\left[\mathsf{quality}_i(x, \boldsymbol{a}_{-i};\boldsymbol{t})\right] \geq 1\right\}.
\end{aligned}$$

In is not hard to see that because $\mathsf{quality}_i(\boldsymbol{t}, \boldsymbol{a}) = \Pi W \Pi^{-1}\boldsymbol{a}$ for the permutation matrix that corresponds to $\boldsymbol{t}$, the benchmarks $\overline{\mathsf{OPT}}, \overline{\mathsf{OPT}}^{\mathsf{IR}}, \overline{\mathsf{OPT}}^{\mathsf{stable}}$ for W equals the corresponding benchmarks under the original definition for a different problem instance $\overline{W}$ with entries

$$\overline{W}_{ij} = \mathop{\mathbb{E}}_{\boldsymbol{t}\sim\tau}\left[W_{t_i,t_j}\right] = \begin{cases} 1, & i = j,\\ \frac{\mathsf{Induced}(V)-n}{n(n-1)}, & i \neq j, \end{cases}$$

in which the unique optimal solution for all three benchmarks is $\frac{n^2}{\mathsf{Induced}(V)}$. According to Corollary 3.5, one can achieve this cost by either incorporating a degenerate signaling scheme or not sending any signals at all. Therefore, we do not use these alternative benchmarks to evaluate our signaling schemes.

Next, we use the comparison of OPT and $\overline{\mathsf{OPT}}$ in unit-weight graphs as an example to show that the two sets of benchmarks are incomparable in general.

- Consider a unit-weight graph that consists of a complete subgraph $K_{\frac{n}{2}}$ and $\frac{n}{2}$ isolated vertices. In this graph, we have $\mathsf{OPT} = \Theta(n)$ as all isolated vertices have to contribute unit effort. On the other hand, $\overline{\mathsf{OPT}} = \Theta(1)$ because $\mathsf{Induced}(V) = \Theta(n^2)$. This shows the possibility of having $\overline{\mathsf{OPT}} \ll \mathsf{OPT}$.
- Consider a star graph with $n-1$ leaves. We have $\mathsf{OPT} = 1$, achieved by having the center node contribute unit effort. However, $\overline{\mathsf{OPT}} = \Theta(n)$ as $\mathsf{Induced}(V) = \Theta(n)$. Therefore, it is also possible to have $\overline{\mathsf{OPT}} \gg \mathsf{OPT}$.

A.2 Proof of Lemma 2.1

Let $\varphi : \mathsf{Sym}(V) \to \Delta(\mathcal{S}^V)$ be a persuasive signaling scheme that is not necessarily identity-independent. Consider the following identity-independent signaling scheme $\widetilde{\varphi}$ that always chooses

$\widetilde{\varphi}(\boldsymbol{t}) = \mathcal{D}_{\widetilde{\varphi}}$ to be the mixture of $\{\varphi(\boldsymbol{t}) \mid \boldsymbol{t} \in \mathsf{Sym}(V)\}$, i.e.,

$$\mathcal{D}_{\widetilde{\varphi}}(\boldsymbol{s}) = \frac{1}{|\mathsf{Sym}(V)|} \sum_{\boldsymbol{t} \in \mathsf{Sym}(V)} \varphi(\boldsymbol{t})(\boldsymbol{s}).$$

We show that $\widetilde{\varphi}$ is persuasive. Let $\mu_i, \widetilde{\mu}_i$ be the posterior distribution of agent i under signaling schemes φ and $\widetilde{\varphi}$, respectively. Since $\widetilde{\varphi}(\boldsymbol{t})$ is defined to be independent of $\boldsymbol{t}$, we have

$$\widetilde{\mu}_i(\boldsymbol{t}, \boldsymbol{s}_{-t_i} \mid s_{t_i}) = \tau(\boldsymbol{t}) \cdot \frac{\sum_{\boldsymbol{t}' \in \mathsf{Sym}(V)} \Pr_{\varphi(\boldsymbol{t}')}\left[s_{t'_i}, \boldsymbol{s}_{-t'_i}\right]}{\sum_{\boldsymbol{s}'_{-t'_i}} \sum_{\boldsymbol{t}' \in \mathsf{Sym}(V)} \Pr_{\varphi(\boldsymbol{t}')}\left[s_{t'_i}, \boldsymbol{s}'_{-t'_i}\right]} = \frac{1}{|\mathsf{Sym}(V)|} \sum_{\boldsymbol{t}' \in \mathsf{Sym}(V)} \mu_i(\boldsymbol{t}', \boldsymbol{s}_{-t'_i} \mid s_{t'_i}).$$

where the last step follows from $\tau(\boldsymbol{t}) = \frac{1}{|\mathsf{Sym}(V)|}$ and the definition of μ_i in Equation (1). Therefore, for the expected quality under the posterior distribution $\widetilde{\mu}_i$, we have

$$\mathop{\mathbb{E}}_{(\boldsymbol{t}, \boldsymbol{s}_{-t_i}) \sim \widetilde{\mu}_i(\cdot \mid \boldsymbol{s}_{t_i} = \theta)} [u_{t_i}(\theta, \boldsymbol{s}_{-t_i})] = \frac{1}{|\mathsf{Sym}(V)|} \mathop{\mathbb{E}}_{\boldsymbol{t} \sim \tau} \left[\sum_{\boldsymbol{t}' \in \mathsf{Sym}(V)} \mathop{\mathbb{E}}_{(\boldsymbol{t}', \boldsymbol{s}_{-t'_i}) \sim \mu_i(\cdot \mid \boldsymbol{s}_{t'_i} = \theta)} \left[u_{t'_i}(\theta, \boldsymbol{s}_{-t'_i}) \right] \right] \geq 1,$$

where the inequality is because the persuasiveness of φ implies that

$$\mathop{\mathbb{E}}_{(\boldsymbol{t}, \boldsymbol{s}_{-t_i}) \sim \mu_i(\cdot \mid \boldsymbol{s}_{t_i} = \theta)} [u_{t_i}(\theta, \boldsymbol{s}_{-t_i})] \geq 1$$

for any agent $i \in [n]$ and any $\theta \in \mathcal{S}$, which holds with equality when $\theta > 0$. We have thus established the persuasiveness of $\widetilde{\varphi}$.

Finally, as for the cost, we have

$$\mathsf{Cost}(\widetilde{\varphi}) = \mathop{\mathbb{E}}_{\boldsymbol{s} \sim \mathcal{D}_{\widetilde{\varphi}}} [\|\boldsymbol{s}\|_1] = \frac{1}{|\mathsf{Sym}(V)|} \sum_{\boldsymbol{t} \in \mathsf{Sym}(V)} \mathop{\mathbb{E}}_{\boldsymbol{s} \sim \varphi(\boldsymbol{t})} [\|\boldsymbol{s}\|_1] = \mathop{\mathbb{E}}_{\boldsymbol{t} \sim \tau, \boldsymbol{s} \sim \varphi(\boldsymbol{t})} [\|\boldsymbol{s}\|_1] = \mathsf{Cost}(\varphi).$$

B Omitted proofs from Section 3.1

B.1 Proof of Lemma 3.2

In this subsection, we finish the proof of Lemma 3.2 by establishing eq. (2), which is restated as follows:

$$\mathop{\mathbb{E}}_{(\boldsymbol{t}, \boldsymbol{s}_{-t_i}) \sim \mu_i(\cdot \mid \boldsymbol{s}_{t_i} = \theta)} [u_{t_i}(\theta, \boldsymbol{s}_{-t_i})] = \theta + \frac{\mathsf{Contrib}_\theta}{\mathsf{Num}_\theta}.$$

We begin by substituting the expression for the posterior $\mu_i(\cdot \mid \boldsymbol{s}_{t_i} = \theta)$ from eq. (1):

$$\begin{aligned}
\sum_{\boldsymbol{t}, \boldsymbol{s}_{-t_i}} \mu_i(\boldsymbol{t}, \boldsymbol{s}_{-t_i} \mid s_{t_i = \theta}) \cdot u_{t_i}(\theta, \boldsymbol{s}_{-t_i}) &\overset{\text{(a)}}{=} \sum_{\boldsymbol{t}, \boldsymbol{s}_{-t_i}} \frac{\tau(\boldsymbol{t}) \mathcal{D}_\varphi(\theta, \boldsymbol{s}_{-t_i})}{\sum_{\boldsymbol{t}', \boldsymbol{s}'_{-t_i}} \tau(\boldsymbol{t}') \mathcal{D}_\varphi(\theta, \boldsymbol{s}'_{-t_i})} \cdot u_{t_i}(\theta, \boldsymbol{s}_{-t_i}) \\
&\overset{\text{(b)}}{=} \frac{\sum_{\boldsymbol{t}, \boldsymbol{s}} \tau(\boldsymbol{t}) \mathcal{D}_\varphi(\boldsymbol{s}) \cdot \mathbb{1}[s_{t_i} = \theta] \cdot u_{t_i}(\boldsymbol{s})}{\sum_{\boldsymbol{t}', \boldsymbol{s}'} \tau(\boldsymbol{t}') \mathcal{D}_\varphi(\boldsymbol{s}') \cdot \mathbb{1}[s'_{t_i} = \theta]} = \frac{\mathbb{E}_{\boldsymbol{s} \sim \mathcal{D}_\varphi}\left[\sum_{\boldsymbol{t}} \tau(\boldsymbol{t}) \cdot \mathbb{1}[s_{t_i} = \theta] \cdot u_{t_i}(\boldsymbol{s})\right]}{\mathbb{E}_{\boldsymbol{s}' \sim \mathcal{D}_\varphi}\left[\sum_{\boldsymbol{t}'} \tau(\boldsymbol{t}') \cdot \mathbb{1}[s'_{t_i} = \theta]\right]} \\
&\overset{\text{(c)}}{=} \frac{\mathbb{E}_{\boldsymbol{s} \sim \mathcal{D}_\varphi}\left[\sum_{v \in V} \frac{1}{n} \cdot \mathbb{1}[s_v = \theta] \cdot u_v(\boldsymbol{s})\right]}{\mathbb{E}_{\boldsymbol{s}' \sim \mathcal{D}_\varphi}\left[\sum_{v \in V} \frac{1}{n} \cdot \mathbb{1}[s'_v = \theta]\right]} \overset{\text{(d)}}{=} \frac{\mathbb{E}_{\boldsymbol{s} \sim \mathcal{D}_\varphi}\left[\sum_{v \in V} \mathbb{1}[s_v = \theta] \cdot \left(\theta + \sum_{v' \in N(v)} W_{v,v'} s_{v'}\right)\right]}{\mathbb{E}_{\boldsymbol{s}' \sim \mathcal{D}_\varphi}\left[\sum_{v \in V} \mathbb{1}[s'_v = \theta]\right]}
\end{aligned}$$

$$= \theta + \frac{\mathsf{Contrib}_\theta}{\mathsf{Num}_\theta}.$$

In the above equations, step (a) uses the expression of posterior μ_i in eq. (1) together with the identity-independent property of scheme $\mathcal{D}_\varphi$. Step (b) follows from $\mathcal{D}_\varphi(\theta, \boldsymbol{s}_{-t_i}) = \sum_{s_{t_i}} \mathcal{D}_\varphi(\boldsymbol{s}) \cdot \mathbb{1}[s_{t_i} = \theta]$, and that $u_{t_i}(\theta, \boldsymbol{s}_{-t_i}) = u_{t_i}(\boldsymbol{s})$ when $s_{t_i} = \theta$. Step (c) is because $\sum_{\boldsymbol{t}:t_i=v} \tau(\boldsymbol{t}) = \frac{1}{n}$ since the marginal distribution of τ on t_i is uniform over V. Step (d) follows from $u_v(\boldsymbol{s}) = \theta + \sum_{v' \in N(v)} W_{v,v'} s_{v'}$ when $s_v = \theta$. The final step uses the linearity of expectations.

B.2 Proof of Lemma 3.4

We prove Lemma 3.4 using the characterizations in Lemma 3.2. If the signaling scheme $\mathcal{D}_\varphi$ induced by sending α to a random subset of vertices $S \sim \mathcal{D}$ is persuasive, its slack (defined in Definition 3.1) must satisfy $\Delta_0 \geq 0$ and $\Delta_\alpha = 0$ for all $\alpha \in (0, 1]$. We have

$$\begin{aligned}
\Delta_\alpha &= \mathsf{Contrib}_\alpha - (1-\alpha)\mathsf{Num}_\alpha \\
&= \mathop{\mathbb{E}}_{\boldsymbol{s} \sim \mathcal{D}_\varphi} \left[\sum_{v \in V} \mathbb{1}[s_v = \alpha] \left(\sum_{v' \in N(v)} W_{v,v'} s_{v'} - (1-\alpha) \right) \right] \\
&= \mathop{\mathbb{E}}_{S \sim \mathcal{D}} \left[\sum_{v \in S} \left(\sum_{v' \in S \setminus \{v\}} W_{v,v'} \cdot \alpha - 1 + \alpha \right) \right] && (s_{v'} = \alpha \cdot \mathbb{1}[v' \in S]) \\
&= \mathop{\mathbb{E}}_{S \sim \mathcal{D}} \left[\alpha \cdot \sum_{v \in S} \sum_{v' \in S} W_{v,v'} - |S| \right] && (W_{v,v} = 1) \\
&= \alpha \cdot \mathop{\mathbb{E}}_{S \sim \mathcal{D}} [\mathsf{Induced}(S)] - \mathop{\mathbb{E}}_{S \sim \mathcal{D}} [|S|].
\end{aligned}$$

Therefore, to make $\mathcal{D}_\varphi$ persuasive, we should set $\alpha = \frac{\mathbb{E}_{S \sim \mathcal{D}}[|S|]}{\mathbb{E}_{S \sim \mathcal{D}}[\mathsf{Induced}(S)]}$.

Similarly, for Δ_0, we have

$$\Delta_0 = \mathsf{Contrib}_0 - \mathsf{Num}_0 = \mathop{\mathbb{E}}_{S \sim \mathcal{D}} \left[\sum_{v \notin S} \left(\sum_{v' \in S} \alpha \cdot W_{v,v'} - 1 \right) \right] = \alpha \cdot \mathop{\mathbb{E}}_{S \sim \mathcal{D}} [\mathsf{Cut}(S, V \setminus S)] - \mathop{\mathbb{E}}_{S \sim \mathcal{D}} [|V \setminus S|].$$

Plugging in the choice of α, to make $\Delta_0 \geq 0$, the distribution $\mathcal{D}$ should satisfy

$$\frac{\mathbb{E}_{S \sim \mathcal{D}} [\mathsf{Cut}(S, V \setminus S)]}{\mathbb{E}_{S \sim \mathcal{D}} [|V \setminus S|]} \geq \frac{\mathbb{E}_{S \sim \mathcal{D}} [\mathsf{Induced}(S)]}{\mathbb{E}_{S \sim \mathcal{D}} [|S|]},$$

which is exactly the inequality in Lemma 3.4. Finally, for the cost of $\mathcal{D}_\varphi$, we have

$$\mathsf{Cost}(\mathcal{D}_\varphi) = \alpha \cdot \mathop{\mathbb{E}}_{S \sim \mathcal{D}} [|S|] = \frac{(\mathbb{E}_{S \sim \mathcal{D}} [|S|])^2}{\mathbb{E}_{S \sim \mathcal{D}} [\mathsf{Induced}(S)]}.$$

C Details for Section 4

C.1 Conjectured impossibility of matching $\mathsf{OPT}^{\mathsf{stable}}$ using $O(1)$ signals.

We conjecture that, in general, $\omega(1)$ different signals are needed to achieve a cost below $\mathsf{OPT}^{\mathsf{stable}}$. On a technical level, even if it is possible to achieve a cost of $\mathsf{OPT}^{\mathsf{stable}}$ using $O(1)$ different signals,

such a result is unlikely to follow from a "natural" proof strategy, because signaling schemes with small signal spaces do not *compose* well.

On a graph G that consists of several disjoint components $G_1, G_2, \ldots$, a natural proof strategy would first derive a signaling scheme φ_i with a constant-size signal space for each G_i (such that $\mathsf{Cost}(\varphi_i) = \mathsf{OPT}^{\mathsf{stable}}(G_i)$), and then appropriately combine them into a scheme for G. While the direct product of $\varphi_1, \varphi_2, \ldots$ does give a persuasive scheme for the larger graph G, the number of different signals also gets added up, and may scale with the size of G.

Another natural idea is to mimic our approach to designing binary schemes — we view the signaling scheme as a randomized partitioning of G_i into $O(1)$ parts[6], and take the direct product of each $\mathcal{D}_i$ instead. Unfortunately, this breaks even for the case of binary schemes: Recall from Lemma 3.4 that $\mathcal{D}_i \in \Delta(2^{V_i})$ gives a persuasive binary scheme for G_i if and only if

$$\frac{\mathbb{E}_{S\sim\mathcal{D}_i}\left[\mathsf{Cut}(S, V_i \setminus S)\right]}{\mathbb{E}_{S\sim\mathcal{D}_i}\left[|V_i \setminus S|\right]} \geq \frac{\mathbb{E}_{S\sim\mathcal{D}_i}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{S\sim\mathcal{D}_i}\left[|S|\right]}. \tag{11}$$

If we choose $\mathcal{D}$ as the direct product of $\mathcal{D}_1, \mathcal{D}_2, \ldots$, the corresponding condition reduces to

$$\frac{\sum_i \mathbb{E}_{S\sim\mathcal{D}_i}\left[\mathsf{Cut}(S, V_i \setminus S)\right]}{\sum_i \mathbb{E}_{S\sim\mathcal{D}_i}\left[|V_i \setminus S|\right]} \geq \frac{\sum_i \mathbb{E}_{S\sim\mathcal{D}_i}\left[\mathsf{Induced}(S)\right]}{\sum_i \mathbb{E}_{S\sim\mathcal{D}_i}\left[|S|\right]}. \tag{12}$$

Perhaps counter-intuitively, even if Equation (11) holds for every i, the combined inequality in Equation (12) might *not* hold![7]

Based on this technical insight, we found a simple instance on which a numerical optimization over ternary signaling schemes suggests that three different signals are not enough to achieve a cost of $\mathsf{OPT}^{\mathsf{stable}}$. We consider a graph that consists of multiple connected components $G_1, G_2, \ldots$ with vertex sets $V_1, V_2, \ldots$. Each graph G_i admits a persuasive binary signaling scheme with cost $\mathsf{OPT}^{\mathsf{stable}}(G_i)$ induced by a distribution $\mathcal{D}_i \in \Delta(2^{V_i})$ that satisfies Equation (11). Towards ensuring that the direct product of $\mathcal{D}_1, \mathcal{D}_2, \ldots$ does not give a persuasive signaling scheme, we choose these components such that: (1) Equation (11) is tight for each i; (2) the two sides of Equation (11) take different values for different values of i.

Concretely, we consider the graph that consists of:

- **Component 1:** Three vertices that form a path with two unit-weight edges.
- **Component 2:** Two additional vertices connected by an edge of weight 1/2.

In the optimal stable solution, the vertex in the middle of the 3-path plays 1, while both vertices in the other component play 2/3. Recall that a binary signaling scheme is specified by a distribution over 2^V, which is a 31-dimensional simplex on this specific graph. Therefore, the optimal binary signaling scheme is characterized by a constrained continuous optimization problem on this simplex, in which the constraint is given by Lemma 3.4. We relax this optimization problem to multiple linear programs, and numerically verified that no binary signaling scheme achieves a cost of $\mathsf{OPT}^{\mathsf{stable}}$ or lower.

The same idea gives another graph that consists of:

[6]For instance, the distribution $\mathcal{D} \in \Delta(2^V)$ that specifies a binary scheme is essentially a randomized 2-partition of V.

[7]This is similar to *Simpson's paradox* in statistics. For unit-weight graphs, we avoided this issue by showing that the left-hand side of Equation (11) is always lower bounded by 1, which in turn upper bounds the right-hand side. This stronger condition *does* imply Equation (12) when multiple components are combined.

- **Component 1:** Three vertices that form a path with two unit-weight edges.
- **Component 2:** Two vertices connected by an edge of weight 2/3.
- **Component 3:** Three vertices that form a triangle with three edges of weights 1/2, 3/4, and 3/4.

A similar but more complicated numerical optimization on ternary signaling schemes indicates that no ternary scheme achieves a cost of $\mathsf{OPT}^{\mathsf{stable}}$ or lower.

Despite this negative evidence, it *might* still be possible to achieve an $O(\mathsf{OPT}^{\mathsf{stable}})$ cost using $O(1)$ signals. However, such a constant factor approximation, in conjunction with our proof of Theorem 1.4, would not guarantee a strict improvement upon $\mathsf{OPT}^{\mathsf{stable}}$.

C.2 Conjectured $\Omega(n^{3/4})$ lower bound

We conjecture that the $n^{3/4}$ ratio is tight on the following instance: The graph consists of two centers connected by a unit-weight edge, along with $n-2$ leaves that form a clique with edge weight $n^{-3/4}$. In addition, each center is connected to each leaf by an edge of weight 1/2.

When only the centers play 1 each, we obtain a feasible IR solution with cost $2 = O(1)$, whereas our approach at best gives an $O(n^{3/4})$ cost on this graph. For a sanity check, note that the total edge weight in the graph is $m = \Theta(n^2 \cdot n^{-3/4} + n) = \Theta(n^{5/4})$. Our proof of Theorem 1.2 (more specifically, Theorem 7.2) gives a persuasive signaling scheme with cost upper bounded by a constant factor times

$$\min\left\{(\mathsf{OPT}^{\mathsf{IR}})^{2/3} m^{1/3} n^{1/3} + \mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n}, \frac{n^2}{m}\right\},$$

yet both terms reduce to $\Theta(n^{3/4})$ when we plug in $\mathsf{OPT}^{\mathsf{IR}} = \Theta(1)$ and $m = \Theta(n^{5/4})$.

The graph above resembles the hard instances for the other cases on which we manage to prove tight lower bounds, so why is it harder to prove an $\Omega(n^{3/4})$ bound? Recall from Section 3.4 that we prove lower bounds by carefully choosing a dual solution $f : [0,1] \to \mathbb{R}$ to the LP that characterizes the optimal scheme. On the lower bounds that we managed to prove, we could choose f as a linear combination of the constant function and the function $f(\theta) = \theta/(1-\theta)$. This choice gives a simple closed-form expression, and satisfies concavity and Lipschitz continuity (except when θ is close to 1). These make the choice of f relatively amenable to formal proofs. However, in the hard instance defined above, a numerical computation suggests that the optimal dual solution is much more ill-behaved — it is non-convex, non-concave, and has a large derivative around 1/2. Therefore, we expect that, even if the instance above does witness an $\Omega(n^{3/4})$ bound, the proof would involve a much more complicated choice of f, and a more involved analysis.

D Omitted Proofs from Section 5

D.1 Technical Lemmas in Section 5.1

We prove Lemmas 5.1 and 5.2 stated in Section 5.1.

Proof of Lemma 5.1. Since DS and $V \setminus (\mathsf{IS} \cup \mathsf{DS})$ partitions $V \setminus \mathsf{IS}$, we have

$$\mathsf{Cut}(\mathsf{IS}, V \setminus \mathsf{IS}) = \mathsf{Cut}(\mathsf{IS}, \mathsf{DS}) + \mathsf{Cut}(\mathsf{IS}, V \setminus (\mathsf{IS} \cup \mathsf{DS})).$$

Since DS is a dominating set of G, every vertex in $\mathsf{IS} \subseteq V \setminus \mathsf{DS}$ has a neighbour in DS, so we have $\mathsf{Cut}(\mathsf{IS}, \mathsf{DS}) \geq |\mathsf{IS}|$. Since IS is maximal with respect to the induced sub-graph of $V \setminus \mathsf{DS}$, every vertex $v \in V \setminus (\mathsf{DS} \cup \mathsf{IS})$ has a neighbour in IS; otherwise, v can be added to IS to give a larger independent set. This gives $\mathsf{Cut}(\mathsf{IS}, V \setminus (\mathsf{IS} \cup \mathsf{DS})) \geq |V \setminus (\mathsf{IS} \cup \mathsf{DS})|$. Therefore, we conclude that

$$\mathsf{Cut}(\mathsf{IS}, V \setminus \mathsf{IS}) \geq |\mathsf{IS}| + |V \setminus (\mathsf{IS} \cup \mathsf{DS})| = |V| - |\mathsf{DS}|. \qquad \square$$

Proof of Lemma 5.2. The first bound follows easily from the fact that each vertex is included independently with probability $\theta_v^\star$. For the second bound, note that

$$\mathop{\mathbb{E}}_{S \sim \mathcal{D}} [\mathsf{Induced}(S)] \leq \sum_{u,v \in V} \Pr_{S \sim \mathcal{D}} [u, v \in S] .$$

By definition of $\mathcal{D}$, $\Pr_{S \sim \mathcal{D}} [u, v \in S]$ is given by $\theta_u \theta_v$ if $u \neq v$, and θ_u if $u = v$. It follows that

$$\mathop{\mathbb{E}}_{S \sim \mathcal{D}} [\mathsf{Induced}(S)] \leq \sum_{u,v \in V : u \neq v} \theta_u \theta_v + \sum_{u \in V} \theta_u \leq \left(\sum_{u \in V} \theta_u \right)^2 + \left(\sum_{u \in V} \theta_u \right) = \|\boldsymbol{\theta}\|_1^2 + \|\boldsymbol{\theta}\|_1.$$

For the third bound, note that

$$\mathop{\mathbb{E}}_{S \sim \mathcal{D}} [\mathsf{Cut}(S, V \setminus S)] = \sum_{u,v \in V} W_{u,v} \cdot \Pr_{S \sim \mathcal{D}} [u \notin S \wedge v \in S] = \sum_{u \in V} \sum_{v \in N(u)} W_{u,v} \cdot (1 - \theta_u) \cdot \theta_v.$$

For each $u \in V$, the condition $W\boldsymbol{\theta} \geq \mathbf{1}$ implies $\sum_{v \in N(u)} W_{u,v} \theta_v \geq 1 - \theta_u$, and it follows that

$$\mathop{\mathbb{E}}_{S \sim \mathcal{D}} [\mathsf{Cut}(S, V \setminus S)] = \sum_{u \in V} (1 - \theta_u) \sum_{v \in N(u)} W_{u,v} \theta_v \geq \sum_{u \in V} (1 - \theta_u)^2 \geq \sum_{u \in V} (1 - 2\theta_u) = |V| - 2\|\boldsymbol{\theta}\|_1.$$

$\square$

D.2 Proof of Lemma 5.8

Proof of Lemma 5.8. Recall from the proof of Lemma 5.7 that OPT is characterized by the linear program in (primal LP), and its dual is given by (dual LP). In the following, we construct a feasible solution $\boldsymbol{\phi}^\star$ to the dual LP, such that $\mathbf{1}^\top(\boldsymbol{\theta} - \boldsymbol{\phi}^\star) \leq \|W\boldsymbol{\theta}\|_1 - n$. By the weak duality of LP, we have $\mathbf{1}^\top \boldsymbol{\phi}^\star \leq \mathsf{OPT}$. It follows that

$$\|\boldsymbol{\theta}\|_1 - \mathsf{OPT} \leq \mathbf{1}^\top \boldsymbol{\theta} - \mathbf{1}^\top \boldsymbol{\phi}^\star \leq \|W\boldsymbol{\theta}\|_1 - n,$$

which is equivalent to the desired bound.

Now we show how to construct $\boldsymbol{\phi}^\star$. Without loss of generality, the vertices of the graph are labeled with $V = [n]$. Let $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}$. For $i = 1, 2, \ldots, n$, we examine the i-th coordinate of $W\boldsymbol{\theta}^{(i-1)}$. If $(W\boldsymbol{\theta}^{(i-1)})_i \leq 1$, we set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$ and move on to the next coordinate. If $(W\boldsymbol{\theta}^{(i-1)})_i > 1$, we choose vector $\boldsymbol{\theta}^{(i)}$ such that:

- $\theta_i^{(i)} + \sum_{j \in N(i)} \theta_j^{(i)} = 1$, and $\theta_j^{(i)} \in \left[0, \theta_j^{(i-1)}\right]$ holds for every $j \in N(i) \cup \{i\}$. This is possible since $\theta_i^{(i-1)} + \sum_{j \in N(i)} \theta_j^{(i-1)} = (W\boldsymbol{\theta}^{(i-1)})_i > 1$.
- $\theta_j^{(i)} = \theta_j^{(i-1)}$ for each $j \notin N(i) \cup \{i\}$.

If $(W\boldsymbol{\theta}^{(i-1)})_i \le 1$, we immediately obtain $(W\boldsymbol{\theta}^{(i)})_i = (W\boldsymbol{\theta}^{(i-1)})_i \le 1$; otherwise, our choice of $\boldsymbol{\theta}^{(i)}$ guarantees

$$(W\boldsymbol{\theta}^{(i)})_i = \theta^{(i)} + \sum_{j \in N(i)} \theta_j^{(i)} = 1,$$

Thus, in both cases, the dual constraint $W\boldsymbol{\phi} \le \mathbf{1}$ is satisfied by $\boldsymbol{\phi} = \boldsymbol{\theta}^{(i)}$ on the i-th coordinate. Furthermore, it is clear from our procedure that $\mathbf{0} \le \boldsymbol{\theta}^{(n)} \le \boldsymbol{\theta}^{(n-1)} \le \cdots \le \boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}$ holds coordinate-wise. Since all entries in W are non-negative, $W\boldsymbol{\theta}^{(n)} \le W\boldsymbol{\theta}^{(n-1)} \le \cdots \le W\boldsymbol{\theta}^{(0)} = W\boldsymbol{\theta}$ also holds coordinate-wise.

Therefore, if we let $\boldsymbol{\phi}^\star = \boldsymbol{\theta}^{(n)}$, it holds for every $i \in [n]$ that

$$(W\boldsymbol{\phi}^\star)_i = (W\boldsymbol{\theta}^{(n)})_i \le (W\boldsymbol{\theta}^{(i)})_i = 1.$$

In other words, $\boldsymbol{\phi}^\star$ is a feasible solution to (dual LP). Moreover, for each $i \in [n]$, if $\boldsymbol{\theta}^{(i)}$ is different from $\boldsymbol{\theta}^{(i-1)}$, we have

$$\mathbf{1}^\top \left(\boldsymbol{\theta}^{(i-1)} - \boldsymbol{\theta}^{(i)}\right) = \left(\theta_i^{(i-1)} + \sum_{j \in N(i)} \theta_j^{(i-1)}\right) - \left(\theta_i^{(i)} + \sum_{j \in N(i)} \theta_j^{(i)}\right) = (W\boldsymbol{\theta}^{(i-1)})_i - 1 \le (W\boldsymbol{\theta})_i - 1.$$

If $\boldsymbol{\theta}^{(i)}$ remains the same as $\boldsymbol{\theta}^{(i-1)}$, we trivially have

$$\mathbf{1}^\top \left(\boldsymbol{\theta}^{(i-1)} - \boldsymbol{\theta}^{(i)}\right) = 0 \le (W\boldsymbol{\theta})_i - 1,$$

where the last step follows from the feasibility of $\boldsymbol{\theta}$. Therefore, we have $\mathbf{1}^\top \left(\boldsymbol{\theta}^{(i-1)} - \boldsymbol{\theta}^{(i)}\right) \le (W\boldsymbol{\theta})_i - 1$ in both cases, and the desired bound follows from

$$\mathbf{1}^\top (\boldsymbol{\theta} - \boldsymbol{\phi}^\star) = \sum_{i=1}^{n} \mathbf{1}^\top \left(\boldsymbol{\theta}^{(i-1)} - \boldsymbol{\theta}^{(i)}\right) \le \sum_{i=1}^{n} [(W\boldsymbol{\theta})_i - 1] = \|W\boldsymbol{\theta}\|_1 - n.$$

□

E Omitted Proofs from Section 6

Proof of Theorem 6.2. We start with the case that $k = O(\sqrt{n})$. The $k \gg \sqrt{n}$ case would easily follow from the lower bound for the first case.

The $k = O(\sqrt{n})$ case. Concretely, we assume that $k \le \sqrt{n}/2$. Consider the following graph: There are k star graphs, each with a center connected to $l := \lfloor \frac{n}{k} \rfloor - 1$ leaves. The k centers of the stars form a clique. For this graph, we have $\mathsf{OPT} = k$ (achieved when each of the k centers contributes 1) and $\mathsf{OPT}^{\mathsf{stable}} = \Omega(n)$ (e.g., achieved when the center in one of the stars and the leaves in all the other stars contribute 1 each). We will show that every persuasive signaling scheme must have a cost of $\Omega(k\sqrt{n}) = \Omega(\min\{k\sqrt{n}, n\})$.

Again, we assume that there is a signaling scheme with cost $\le k\sqrt{n}/C$ where $C := 20$, and we will still derive a lower bound of $\Omega(k\sqrt{n})$ on the cost under this assumption. By Lemma 2.1, without loss of generality, the signaling scheme is identity-independent and is specified by a distribution $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$. In other words, the scheme samples $((s_{c,1}, \ldots, s_{c,k}), (s_{i,j})_{i \in [k], j \in [l]}) \in \mathcal{S}^V$ from $\mathcal{D}_\varphi$,

labels the center of the i-th star with $s_{c,i}$ and labels the j-th leaf in the i-th star with $s_{i,j}$. For brevity, we shorthand x_i for $s_{c,i}$, and let $\mathsf{sum} := \sum_{i=1}^{k} x_i + \sum_{i=1}^{k}\sum_{j=1}^{l} s_{i,j}$ denote the sum of the signals.

We first use the persuasiveness of $\mathcal{D}_\varphi$ to derive a lower bound on $\mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^{k} x_i\right]$, the expected total signal received by the k centers. By Lemma 3.2, it holds for every $\theta \in \mathcal{S}$ that

$$0 \le \Delta_\theta = \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v = \theta\right]\left(\sum_{u\in N(v)} s_u - (1 - s_v)\right)\right].$$

Summing over $\theta \in \mathcal{S}$ gives

$$0 \le \sum_{\theta\in\mathcal{S}} \Delta_\theta = \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v\in V}\left(s_v + \sum_{u\in N(v)} s_u - 1\right)\right] = \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{u\in V} s_u\left(1 + |N(u)|\right)\right] - k(l+1).$$

Since $|N(u)| = 1$ when u is a leaf and $|N(u)| = l + k - 1$ when u is a center, we have

$$k(l+1) \le \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{u\in V} s_u\left(1 + |N(u)|\right)\right] = 2\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\mathsf{sum}\right] + (l + k - 2)\cdot \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^{k} x_i\right].$$

Applying the assumption that $\mathbb{E}_{\mathcal{D}_\varphi}\left[\mathsf{sum}\right] \le k\sqrt{n}/C$ gives the desired lower bound on $\mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^{k} x_i\right]$:

$$\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^{k} x_i\right] \ge \frac{k(l+1) - 2\,\mathbb{E}_{\mathcal{D}_\varphi}\left[\mathsf{sum}\right]}{l + k - 2} \ge \frac{k(l+1) - 2k\sqrt{n}/C}{l + k}.$$

Let $y := \frac{\sqrt{n}}{\sqrt{n}+2k}$ and define the random variable $X := \sum_{i=1}^{k} \mathbb{1}\left[x_i \ge y\right]$. Next, we use the lower bound on $\mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^{k} x_i\right]$ to lower bound the expectation of X by $\Omega(k)$. Applying the inequality $\mathbb{1}\left[x_i \ge y\right] \ge \frac{x_i - y}{1-y}$ gives

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[X\right] &\ge \frac{\mathbb{E}_{\mathcal{D}}\left[\sum_{i=1}^{k} x_i\right] - ky}{1-y} \ge \frac{\frac{k(l+1) - 2k\sqrt{n}/C}{l+k} - ky}{1-y} \\
&= \frac{kl(1-y) + k - 2k\sqrt{n}/C - yk^2}{(1-y)(l+k)} \\
&\ge \frac{kl}{l+k} - \frac{2k\sqrt{n}}{C(1-y)(l+k)} - \frac{yk^2}{(1-y)(l+k)}.
\end{aligned}$$

Plugging $y = \frac{\sqrt{n}}{\sqrt{n}+2k}$ into the last term above gives

$$\frac{kl}{l+k} - \frac{n + 2k\sqrt{n}}{C(l+k)} - \frac{k\sqrt{n}}{2(l+k)}.$$

Recall that $k \le \sqrt{n}/2$, $l = \lfloor n/k \rfloor - 1 = (n/k)\cdot(1 + o_n(1))$ and $C = 20$. For all sufficiently large n, the three terms above can be bounded as follows:

$$\frac{kl}{l+k} = \frac{k}{1 + k/l} = \frac{k}{1 + k^2/n\cdot(1 + o_n(1))} \ge \frac{k}{1 + 1/4 + o_n(1)} \ge \frac{k}{2},$$

$$\frac{n+2k\sqrt{n}}{C(l+k)} \le \frac{2n}{Cl} = \frac{k}{10}(1+o_n(1)) \le \frac{k}{8},$$

and

$$\frac{k\sqrt{n}}{2(l+k)} = \frac{k}{2(\sqrt{n}/k + k/\sqrt{n})} \cdot (1+o_n(1)) \le \frac{k}{2\cdot(2+1/2)} \cdot (1+o_n(1)) \le \frac{k}{4}.$$

Therefore, we conclude that $\mathbb{E}_{\mathcal{D}_\varphi}[X] \ge k/2 - k/8 - k/4 = k/8$.

Since $k - X$ is always non-negative, by Markov's inequality, we have

$$\Pr_{\mathcal{D}_\varphi}\left[X \le \frac{k}{10}\right] = \Pr_{\mathcal{D}_\varphi}\left[k - X \ge \frac{9k}{10}\right] \le \frac{\mathbb{E}_{\mathcal{D}_\varphi}[k-X]}{9k/10} \le \frac{7k/8}{9k/10} = \frac{35}{36},$$

which implies $\Pr_{\mathcal{D}_\varphi}\left[X \ge \frac{k}{10}\right] \ge \frac{1}{36}$.

Applying Lemma 3.2 again gives

$$\begin{aligned}
0 = \sum_{\theta\in\mathcal{S}} \Delta_\theta \cdot \mathbb{1}[\theta \ge y] &= \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_v \mathbb{1}[s_v \ge y]\left(\sum_{u\in N(v)} s_u - (1-s_v)\right)\right] \\
&\ge \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^k \mathbb{1}[x_i \ge y]\cdot\left(\sum_{i=1}^k x_i - 1\right) + \sum_{i=1}^k\sum_{j=1}^l \mathbb{1}[s_{i,j} \ge y](y-1)\right] \qquad (13)\\
&= \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[X\left(\sum_{i=1}^k x_i - 1)\right)\right] - (1-y)\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^k\sum_{j=1}^l \mathbb{1}[s_{i,j} \ge y]\right].
\end{aligned}$$

In the second line above, we drop the contributions from the leaves when $x_i \ge y$, and drop the contribution from the incident center when a leaf satisfies $s_{i,j} \ge y$.

In the rest of the proof, we will first lower bound the $\mathbb{E}_{\mathcal{D}_\varphi}\left[X\left(\sum_{i=1}^k x_i - 1)\right)\right]$ term. By Equation (13), this gives a lower bound on $(1-y)\,\mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^k\sum_{j=1}^l \mathbb{1}[s_{i,j} \ge y]\right]$, which, in turn, lower bounds the cost.

Note that

$$X\cdot\left(\sum_{i=1}^k x_i - 1\right) \ge X\cdot\left(\sum_{i=1}^k y\cdot\mathbb{1}[x_i \ge y] - 1\right) = X\cdot(yX-1),$$

and the minimum of the quadratic function $x \mapsto yx^2 - x$ is $-1/(4y)$. By the assumption that $k \le \sqrt{n}/2$, we have $y = \frac{\sqrt{n}}{\sqrt{n}+2k} \in [1/2, 1]$, which implies $X\cdot\left(\sum_{i=1}^k x_i - 1\right) \ge -1/(4y) \ge -1/2$. When $X \ge k/10$, we have a stronger lower bound of $X\cdot\left(\sum_{i=1}^k x_i - 1\right) \ge (k/10)\cdot[y\cdot(k/10) - 1]$. It follows that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[X\left(\sum_{i=1}^k s_i - 1)\right)\right] &\ge \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\mathbb{1}\left[X \ge \frac{k}{10}\right]\cdot\frac{k}{10}\cdot\left(\frac{ky}{10} - 1\right) + \mathbb{1}\left[X < \frac{k}{10}\right]\cdot(-1/2)\right] \\
&\ge \Omega(k^2)\cdot\Pr_{\mathcal{D}_\varphi}\left[X \ge \frac{k}{10}\right] - \frac{1}{2} \ge \Omega(k^2).
\end{aligned}$$

The last step follows from the inequality $\Pr_{\mathcal{D}_\varphi}[X \ge k/10] \ge 1/36$ that we proved earlier. Plugging the above into Equation (13) gives $(1-y)\,\mathbb{E}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^k\sum_{j=1}^l \mathbb{1}[s_{i,j} \ge y]\right] \ge \Omega(k^2)$.

Finally, we conclude that

$$\mathop{\mathbb{E}}_{\mathcal{D}_\varphi}[\mathsf{sum}] \geq y \cdot \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{i=1}^{k}\sum_{j=1}^{l} \mathbb{1}\left[s_{i,j} \geq y\right]\right] \geq \Omega(k^2) \cdot \frac{y}{1-y} \geq \Omega(k\sqrt{n}),$$

where the last step follows from $y = \frac{\sqrt{n}}{\sqrt{n}+2k}$. This completes the proof for the $k \leq \sqrt{n}/2$ case.

The $k \gg \sqrt{n}$ case. We handle the $k > \sqrt{n}/2$ case by a reduction to the first case. Let $k' \coloneqq \lfloor n/(4k) \rfloor$ and $n' = 4(k')^2$. Since $k' \leq \sqrt{n'}/2$, our proof for the first case gives a graph G' with n' vertices, on which $\mathsf{OPT} = k'$, $\mathsf{OPT}^{\mathsf{stable}} = \Omega(n')$, and every persuasive signaling scheme has an $\Omega(k'\sqrt{n'}) = \Omega(n')$ cost.

Let $m \coloneqq \lfloor n/n' \rfloor = \Theta(k^2/n)$. Consider the graph G that consists of m disjoint copies of G', denoted by G'_1 through G'_m. It is clear that the benchmarks OPT and $\mathsf{OPT}^{\mathsf{stable}}$ are additive on a graph consisting of multiple connected components. Thus, on the graph G, we have $\mathsf{OPT} = m \cdot k' = \Theta(k)$ and $\mathsf{OPT}^{\mathsf{stable}} = m \cdot \Omega(n') = \Omega(n)$.

It remains to show that the lower bound on the cost of persuasive signaling schemes also composes. Suppose that $\mathcal{D}_\varphi$ is a persuasive signaling scheme for graph G, with a cost of C. Consider the following induced signaling scheme $\mathcal{D}'_\varphi$ for graph G': Pick $i \in [m]$ uniformly at random, draw $s \sim \mathcal{D}_\varphi$, and choose the signal as the restriction of s to G'_i. Let Δ and Δ' denote the slacks induced by the signaling schemes $\mathcal{D}_\varphi$ and $\mathcal{D}'_\varphi$, respectively. Note that $\mathcal{D}'_\varphi$ is still persuasive, since for every $\theta \in [0,1]$, we have

$$\Delta'_\theta = \frac{1}{m}\sum_{i=1}^{m} \mathop{\mathbb{E}}_{\mathcal{D}_\varphi}\left[\sum_{v \in G'_i} \mathbb{1}\left[s_v = \theta\right]\left(\sum_{u \in N(v)} s_u - (1 - s_v)\right)\right] = \frac{\Delta_\theta}{m} \geq 0,$$

and the inequality is tight for all $\theta > 0$. Moreover, the cost of the signaling scheme $\mathcal{D}'_\varphi$ is given by C/m.

Now, applying the lower bound for graph G' shows that $C/m \geq \Omega(k'\sqrt{n'}) = \Omega(n')$. Therefore, we have $C \geq \Omega(mn') = \Omega(n)$. In other words, every persuasive signaling scheme must have an $\Omega(n) = \Omega(\min\{k\sqrt{n}, n\})$ cost on graph G. This completes the proof. □

F Omitted Proofs from Section 7

F.1 Proof of Lemma 7.1

Proof of Lemma 7.1. Let $\boldsymbol{\theta}^\star \in [0,1]^V$ be an optimal solution subject to IR with cost $\|\boldsymbol{\theta}^\star\|_1 = \mathsf{OPT}^{\mathsf{IR}}$. Consider the following signaling scheme with parameters $p, q \geq 0$ and $\epsilon, \alpha \in [0,1]$.

- With probability $\frac{1}{1+p+q}$, draw a random set $\widetilde{\mathsf{DS}} \subseteq V$ that includes each $v \in V$ independently with probability $\theta^\star_v$. Label (the vertices in) $\widetilde{\mathsf{DS}}$ with $1-\epsilon$, and label $V \setminus \widetilde{\mathsf{DS}}$ with 0.
- With probability $\frac{p}{1+p+q}$, draw a random set $\widetilde{\mathsf{IS}} \sim \mathcal{D}$. Label $\widetilde{\mathsf{IS}}$ with $1-\epsilon$, and label $V \setminus \widetilde{\mathsf{IS}}$ with 0.
- With probability $\frac{q}{1+p+q}$, draw a random set $\widetilde{\mathsf{DS}} \subseteq V$ that includes each $v \in V$ independently with probability $\theta^\star_v$. Label $V \setminus \widetilde{\mathsf{DS}}$ with α, and label $\widetilde{\mathsf{DS}}$ with 0.

Properties of $\widetilde{\mathsf{DS}}$ and $\widetilde{\mathsf{IS}}$. We will use the following properties of $\widetilde{\mathsf{DS}}$ and $\widetilde{\mathsf{IS}}$:

- $\mathbb{E}\left[|\widetilde{\mathsf{DS}}|\right] = \|\boldsymbol{\theta}^\star\|_1 = \mathsf{OPT}^{\mathsf{IR}}$.
- $\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}})\right] \leq (\mathsf{OPT}^{\mathsf{IR}})^2 + \mathsf{OPT}^{\mathsf{IR}}$.
- $\mathbb{E}\left[\mathsf{Cut}(\widetilde{\mathsf{DS}}, V \setminus \widetilde{\mathsf{DS}})\right] \geq n - 2\mathsf{OPT}^{\mathsf{IR}}$.
- $\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}})\right] \leq (1+\gamma)\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]$.

The first three bounds follow from Lemma 5.2. The last follows from the assumption on $\mathcal{D}$.

Conditions for persuasiveness. By Definition 3.1 and Lemma 3.2, in order for the signaling scheme above to be persuasive, we need to satisfy:

- For $\theta \in \{\alpha, 1-\epsilon\}$,
$$(1-\theta)\cdot \mathsf{Num}_\theta = \mathsf{Contrib}_\theta \tag{14}$$
- For $\theta = 0$, Equation (14) holds with "$=$" replaced with "$\leq$".

In the rest of the proof, we carefully pick the parameters of the signaling scheme to satisfy the conditions above, while ensuring that the resulting cost satisfies the desired upper bound.

Pick α to handle $\theta = \alpha$. We first examine Equation (14) when $\theta = \alpha$. Since we send α to $V \setminus \widetilde{\mathsf{DS}}$ with probability $\frac{q}{1+p+q}$, the left-hand side of Equation (14) is given by

$$(1-\alpha)\cdot\frac{q}{1+p+q}\,\mathbb{E}\left[|V\setminus\widetilde{\mathsf{DS}}|\right] = (1-\alpha)\cdot\frac{q}{1+p+q}(n-\mathsf{OPT}^{\mathsf{IR}}).$$

The right-hand side, on the other hand, is given by

$$\frac{q}{1+p+q}\cdot\alpha\cdot\mathbb{E}\left[\mathsf{Induced}(V\setminus\widetilde{\mathsf{DS}}) - (n - |\widetilde{\mathsf{DS}}|)\right] = \frac{q}{1+p+q}\cdot\alpha\cdot\left[\mathbb{E}\left[\mathsf{Induced}(V\setminus\widetilde{\mathsf{DS}})\right] - (n-\mathsf{OPT}^{\mathsf{IR}})\right].$$

Therefore, for $\theta = \alpha$, Equation (14) reduces to

$$(1-\alpha)(n-\mathsf{OPT}^{\mathsf{IR}}) = \alpha\cdot\left[\mathbb{E}\left[\mathsf{Induced}(V\setminus\mathsf{DS})\right] - (n-\mathsf{OPT}^{\mathsf{IR}})\right],$$

which holds if we pick $\alpha = \frac{n-\mathsf{OPT}^{\mathsf{IR}}}{\mathbb{E}\left[\mathsf{Induced}(V\setminus\widetilde{\mathsf{DS}})\right]}$. Note that this choice of α is valid, since we have

$$\mathbb{E}\left[\mathsf{Induced}(V\setminus\widetilde{\mathsf{DS}})\right] \geq \mathbb{E}\left[|V\setminus\widetilde{\mathsf{DS}}|\right] = n - \mathsf{OPT}^{\mathsf{IR}} \geq 0,$$

which implies $\alpha \in [0,1]$.

Pick ϵ to handle $\theta = 1-\epsilon$. When $\theta = 1-\epsilon$, the left-hand side of Equation (14) is

$$\epsilon\cdot\left[\frac{1}{1+p+q}\cdot\mathbb{E}\left[\left|\widetilde{\mathsf{DS}}\right|\right] + \frac{p}{1+p+q}\cdot\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right] = \frac{1}{1+p+q}\cdot\epsilon\cdot\left(\mathsf{OPT}^{\mathsf{IR}} + p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right).$$

The right-hand side is equal to

$$(1-\epsilon)\cdot\left[\frac{1}{1+p+q}\cdot\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}}) - |\widetilde{\mathsf{DS}}|\right] + \frac{p}{1+p+q}\cdot\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}}) - |\widetilde{\mathsf{IS}}|\right]\right].$$

Then, Equation (14) at $\theta = 1 - \epsilon$ is equivalent to

$$\epsilon \cdot \left(\mathsf{OPT}^{\mathsf{IR}} + p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right) = (1-\epsilon) \cdot \left[\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}}) - |\widetilde{\mathsf{DS}}|\right] + p \cdot \mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}}) - |\widetilde{\mathsf{IS}}|\right]\right],$$

which is satisfied if we pick

$$\epsilon = \frac{\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}})\right] - \mathsf{OPT}^{\mathsf{IR}} + p \cdot \mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}}) - |\widetilde{\mathsf{IS}}|\right]}{\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}})\right] + p \cdot \mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}})\right]} \in [0,1].$$

Pick q to handle $\theta = 0$ case. Finally, at $\theta = 0$, the left-hand side of Equation (14) is equal to

$$\begin{aligned}
&\frac{1}{1+p+q} \cdot \left[\mathbb{E}\left[|V \setminus \widetilde{\mathsf{DS}}|\right] + p \cdot \mathbb{E}\left[|V \setminus \widetilde{\mathsf{IS}}|\right] + q \cdot \mathbb{E}\left[|\widetilde{\mathsf{DS}}|\right]\right] \\
= &\frac{1}{1+p+q} \cdot \left[(n - \mathsf{OPT}^{\mathsf{IR}}) + p\left(n - \mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right) + q\mathsf{OPT}^{\mathsf{IR}}\right].
\end{aligned}$$

The right-hand side is given by

$$\frac{1}{1+p+q} \cdot \left[(1-\epsilon) \cdot \mathbb{E}\left[\mathsf{Cut}(\widetilde{\mathsf{DS}}, V \setminus \widetilde{\mathsf{DS}})\right] + p \cdot (1-\epsilon) \cdot \mathbb{E}\left[\mathsf{Cut}(\widetilde{\mathsf{IS}}, V \setminus \widetilde{\mathsf{IS}})\right] + q \cdot \alpha \cdot \mathbb{E}\left[\mathsf{Cut}(\widetilde{\mathsf{DS}}, V \setminus \widetilde{\mathsf{DS}})\right]\right].$$

Recall that we have $\mathbb{E}\left[\mathsf{Cut}(\widetilde{\mathsf{DS}}, V \setminus \widetilde{\mathsf{DS}})\right] \geq n - 2\mathsf{OPT}^{\mathsf{IR}}$. We also trivially relax the $\mathbb{E}\left[\mathsf{Cut}(\widetilde{\mathsf{IS}}, V \setminus \widetilde{\mathsf{IS}})\right]$ term to 0. Recall that at $\theta = 0$, it suffices for the left-hand side of Equation (14) to be upper bounded by the right-hand side. This can be guaranteed if we have

$$(n - \mathsf{OPT}^{\mathsf{IR}}) + p\left(n - \mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right) + q\mathsf{OPT}^{\mathsf{IR}} \leq (1-\epsilon) \cdot (n - 2\mathsf{OPT}^{\mathsf{IR}}) + q \cdot \alpha \cdot (n - 2\mathsf{OPT}^{\mathsf{IR}}),$$

which is equivalent to

$$q \cdot [\alpha(n - 2\mathsf{OPT}^{\mathsf{IR}}) - \mathsf{OPT}^{\mathsf{IR}}] \geq (n - \mathsf{OPT}^{\mathsf{IR}}) + p\left(n - \mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right) - (1-\epsilon) \cdot (n - 2\mathsf{OPT}^{\mathsf{IR}}).$$

We claim that it is without loss of generality to assume that $\alpha(n - 2\mathsf{OPT}^{\mathsf{IR}}) - \mathsf{OPT}^{\mathsf{IR}} \geq \alpha n/2$; otherwise, as we show at the end of the proof, a cost of $O(\mathsf{OPT}^{\mathsf{IR}})$ can be trivially achieved. Under this additional assumption, it is, in turn, sufficient to satisfy

$$\frac{1}{2} q\alpha n \geq \epsilon n + (1 - 2\epsilon)\mathsf{OPT}^{\mathsf{IR}} + p\left(n - \mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right).$$

Therefore, we can always pick q to satisfy the condition at $\theta = 0$, such that $q\alpha n$ is at most $O(\mathsf{OPT}^{\mathsf{IR}} + (\epsilon + p)n)$.

Upper bound the cost by optimizing p. The cost of our signaling scheme is clearly

$$\frac{1}{1+p+q} \cdot \left[(1-\epsilon)\,\mathbb{E}\left[|\widetilde{\mathsf{DS}}|\right] + p \cdot (1-\epsilon)\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right] + q \cdot \alpha\,\mathbb{E}\left[|V \setminus \widetilde{\mathsf{DS}}|\right]\right] \leq \mathsf{OPT}^{\mathsf{IR}} + p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right] + q\alpha n.$$

By our choice of q, the $q\alpha n$ term is at most $O(\mathsf{OPT}^{\mathsf{IR}} + (\epsilon + p)n)$, so the cost is also upper bounded by $O(\mathsf{OPT}^{\mathsf{IR}} + (\epsilon + p)n)$.

Now, recall our choice of

$$\epsilon = \frac{\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}})\right] - \mathsf{OPT}^{\mathsf{IR}} + p \cdot \mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}}) - |\widetilde{\mathsf{IS}}|\right]}{\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}})\right] + p \cdot \mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}})\right]}.$$

Also recall that $\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{DS}})\right] \leq (\mathsf{OPT}^{\mathsf{IR}})^2 + \mathsf{OPT}^{\mathsf{IR}}$ and $\mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}})\right] \leq \mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right] \cdot (1+\gamma)$. Note that the denominator above is trivially lower bounded by $p \cdot \mathbb{E}\left[\mathsf{Induced}(\widetilde{\mathsf{IS}})\right] \geq p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]$. This gives

$$\epsilon \leq \frac{(\mathsf{OPT}^{\mathsf{IR}})^2 + p \cdot \left(\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right] \cdot (1+\gamma) - \mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]\right)}{p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]} = \gamma + \frac{(\mathsf{OPT}^{\mathsf{IR}})^2}{p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]}.$$

It follows that our cost is at most

$$O(\mathsf{OPT}^{\mathsf{IR}} + (\epsilon + p)n) \preceq \mathsf{OPT}^{\mathsf{IR}} + \gamma n + pn + \frac{n(\mathsf{OPT}^{\mathsf{IR}})^2}{p\,\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]} \preceq \mathsf{OPT}^{\mathsf{IR}} + \gamma n + \frac{n \cdot \mathsf{OPT}^{\mathsf{IR}}}{\sqrt{\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]}},$$

where the second step holds if we set $p = \frac{\mathsf{OPT}^{\mathsf{IR}}}{\sqrt{\mathbb{E}\left[|\widetilde{\mathsf{IS}}|\right]}}$.

When $\alpha(n - 2\mathsf{OPT}^{\mathsf{IR}}) - \mathsf{OPT}^{\mathsf{IR}} \geq \alpha n/2$ does not hold. Finally, we show that if the assumption $\alpha(n - 2\mathsf{OPT}^{\mathsf{IR}}) - \mathsf{OPT}^{\mathsf{IR}} \geq \alpha n/2$ is violated, we can achieve an $O(\mathsf{OPT}^{\mathsf{IR}})$ cost easily.

Note that we must have $\mathsf{OPT}^{\mathsf{IR}} > \alpha n/6$ in this case; otherwise, we would have

$$\alpha(n - 2\mathsf{OPT}^{\mathsf{IR}}) - \mathsf{OPT}^{\mathsf{IR}} \geq \alpha[n - 2 \cdot (\alpha n/6)] - \alpha n/6 \geq \alpha n - \alpha n/3 - \alpha n/6 = \alpha n/2,$$

a contradiction. The second step above holds since $\alpha \in [0, 1]$. Also, we may assume that $\mathsf{OPT}^{\mathsf{IR}} \leq n/2$; otherwise, any persuasive signaling scheme would give a cost of at most $n = O(\mathsf{OPT}^{\mathsf{IR}})$.

Now, recall that $\alpha = \frac{n - \mathsf{OPT}^{\mathsf{IR}}}{\mathbb{E}\left[\mathsf{Induced}(V \setminus \widetilde{\mathsf{DS}})\right]}$. Then, $\mathsf{OPT}^{\mathsf{IR}} > \alpha n/6$ and $\mathsf{OPT}^{\mathsf{IR}} \leq n/2$ together imply

$$\mathsf{OPT}^{\mathsf{IR}} > \frac{n(n - \mathsf{OPT}^{\mathsf{IR}})}{6 \cdot \mathbb{E}\left[\mathsf{Induced}(V \setminus \widetilde{\mathsf{DS}})\right]} \geq \frac{n^2}{12 \cdot \mathbb{E}\left[\mathsf{Induced}(V \setminus \widetilde{\mathsf{DS}})\right]} \geq \frac{n^2}{12(n + 2m)},$$

where m is the total edge weight in the graph. Then, by Corollary 3.5, we can achieve a cost of

$$\frac{n^2}{n + 2m} \leq 12\mathsf{OPT}^{\mathsf{IR}} = O(\mathsf{OPT}^{\mathsf{IR}}).$$

□

F.2 Proof of Theorem 7.2

Proof of Theorem 7.2. Let m be the total edge weight in the graph, and $\gamma > 0$ be a parameter to be chosen later. Define $\mathcal{D} \in \Delta(2^V)$ as the distribution of the random set obtained from including each $v \in V$ independently with probability $r := \min\left\{\frac{\gamma n}{2m}, 1\right\}$. Clearly, we have $\mathbb{E}_{S \sim \mathcal{D}}[|S|] = rn$ and

$$\mathop{\mathbb{E}}_{S \sim \mathcal{D}}[\mathsf{Induced}(S)] = rn + 2r^2 m = \left(1 + \frac{2rm}{n}\right) \cdot \mathop{\mathbb{E}}_{S \sim \mathcal{D}}[|S|] \leq (1 + \gamma) \mathop{\mathbb{E}}_{S \sim \mathcal{D}}[|S|].$$

By Lemma 7.1, there is a persuasive signaling scheme with a cost of, up to a constant factor, at most

$$\mathsf{OPT}^{\mathsf{IR}} + \gamma n + \frac{n \cdot \mathsf{OPT}^{\mathsf{IR}}}{\sqrt{\min\left\{\frac{\gamma n^2}{2m}, n\right\}}} \preceq \mathsf{OPT}^{\mathsf{IR}} + \gamma n + \mathsf{OPT}^{\mathsf{IR}} \cdot \max\left\{\sqrt{m/\gamma}, \sqrt{n}\right\}.$$

Under this choice of $\gamma = (\mathsf{OPT}^{\mathsf{IR}})^{2/3} m^{1/3} n^{-2/3}$, that above cost reduces to

$$(\mathsf{OPT}^{\mathsf{IR}})^{2/3} m^{1/3} n^{1/3} + \mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n}.$$

In addition, recall from Corollary 3.5 that we can easily achieve a cost of $O(n^2/m)$. Therefore, we can always achieve the minimum between these two and obtain an upper bound of

$$\begin{aligned}\min\left\{(\mathsf{OPT}^{\mathsf{IR}})^{2/3} m^{1/3} n^{1/3} + \mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n}, \frac{n^2}{m}\right\} &\leq \min\left\{(\mathsf{OPT}^{\mathsf{IR}})^{2/3} m^{1/3} n^{1/3}, \frac{n^2}{m}\right\} + \mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n} \\ &\preceq n^{3/4} \cdot (\mathsf{OPT}^{\mathsf{IR}})^{1/2} + \mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n}.\end{aligned}$$

Note that the second term above dominates the first only if $\mathsf{OPT}^{\mathsf{IR}} = \Omega(\sqrt{n})$, at which point both terms are greater than the trivial cost of n. Therefore, we can further simplify the bound to $O\left(n^{3/4} \cdot (\mathsf{OPT}^{\mathsf{IR}})^{1/2}\right)$. □

F.3 Proof of Theorem 7.3

Before proving the theorem, we first present a lemma that lower bounds the size of the maximum independent set in a graph.

Lemma F.1. *A graph with n vertices and m edges contains an independent set of size $\Omega\left(\min\{n^2/m, n\}\right)$.*

Proof. Consider the greedy algorithm that keeps adding the vertex with the smallest degree, and then removing the vertex along with its neighbors. As long as the number of remaining vertices, n', is at least $n/2$, the smallest degree is at most $\frac{2m}{n'} \leq 4m/n$, so each iteration removes at most $4m/n + 1$ vertices. The maximal independent set obtained from this greedy approach has size at least $\frac{n/2}{4m/n+1} = \Omega(\min\{n^2/m, n\})$. □

Now we are ready to prove Theorem 7.3.

Proof of Theorem 7.3. Let m denote the total edge weight of the graph. Since the edge weights are lower bounded by δ, there are at most m/δ edges in the graph. By Lemma F.1, the graph contains an independent set IS of size $\Omega(\min\{n^2/(m/\delta), n\}) = \Omega(\min\{\delta n^2/m, n\})$.

Let $\mathcal{D} \in \Delta(2^V)$ be the degenerate distribution at IS. We have $\mathbb{E}_{S\sim\mathcal{D}}\left[\mathsf{Induced}(S)\right] = \mathbb{E}_{S\sim\mathcal{D}}\left[|S|\right] = |\mathsf{IS}|$. Then, applying Lemma 7.1 with $\gamma = 0$ gives a persuasive signaling scheme with a cost of

$$O\left(\mathsf{OPT}^{\mathsf{IR}} + \frac{n \cdot \mathsf{OPT}^{\mathsf{IR}}}{\sqrt{\min\{\delta n^2/m, n\}}}\right) = O\left(\mathsf{OPT}^{\mathsf{IR}} \cdot \left(\sqrt{n} + \sqrt{m/\delta}\right)\right).$$

Again, we use the fact that we can achieve a cost of $O(n^2/m)$ (Corollary 3.5). This give an upper bound of

$$O\left(\min\left\{\mathsf{OPT}^{\mathsf{IR}} \cdot \left(\sqrt{n} + \sqrt{m/\delta}\right), \frac{n^2}{m}\right\}\right) = O\left(\left(n \cdot \mathsf{OPT}^{\mathsf{IR}}\right)^{2/3} \delta^{-1/3} + \mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n}\right).$$

Note that the $\mathsf{OPT}^{\mathsf{IR}} \cdot \sqrt{n}$ term dominates the first term only if $\mathsf{OPT}^{\mathsf{IR}} = \Omega(\sqrt{n})$, at which point the first term is already $\Omega(n)$. Therefore, the upper bound can be simplified to $O\left((n \cdot \mathsf{OPT}^{\mathsf{IR}})^{2/3}\delta^{-1/3}\right)$. □

F.4 Proof of Lemma 7.5

Proof of Lemma 7.5. For the sake of contradiction, assume $\|W\boldsymbol{\theta}\|_1 < n + \frac{\|\boldsymbol{\theta}\| - \mathsf{OPT}}{\mathsf{OPT}}$. Therefore, it must be the case that for all $i \in [n]$,

$$1 \le (W\boldsymbol{\theta})_i < 1 + \frac{\|\boldsymbol{\theta}\|_1 - \mathsf{OPT}}{\mathsf{OPT}} = \frac{\|\boldsymbol{\theta}\|_1}{\mathsf{OPT}}.$$

Recall from the proof of Lemma 5.7 that OPT is characterized by the linear program in eq. (primal LP), and its dual is given by eq. (dual LP). We construct dual variables $\boldsymbol{\phi}^\star$ as follows, where $R_{\max}$ is the largest value among all coordinates of $W\boldsymbol{\theta}$:

$$\boldsymbol{\phi}^\star \triangleq \frac{1}{R_{\max}}\boldsymbol{\theta}, \qquad R_{\max} \triangleq \max_{i\in[n]}(W\boldsymbol{\theta})_i < \frac{\|\boldsymbol{\theta}\|_1}{\mathsf{OPT}}.$$

Note that $\boldsymbol{\phi}^\star$ is a feasible solution of eq. (dual LP) because

$$W\boldsymbol{\phi}^\star = \frac{1}{R_{\max}}(W\boldsymbol{\theta}) \le \mathbf{1}.$$

Connecting to the optimal objective value $\mathsf{OPT} = \min_{\boldsymbol{\theta}} \mathbf{1}^\top\boldsymbol{\theta}$ of eq. (primal LP), we have

$$\begin{aligned}
\mathsf{OPT} &= \min_{\boldsymbol{\theta}} \mathbf{1}^\top\boldsymbol{\theta} \ge \max_{\boldsymbol{\phi}} \mathbf{1}^\top\boldsymbol{\phi} \ge \mathbf{1}^\top\boldsymbol{\phi}^\star && \text{(weak duality)}\\
&= \frac{1}{R_{\max}}\left(\mathbf{1}^\top\boldsymbol{\theta}\right) = \frac{1}{R_{\max}} \cdot \|\boldsymbol{\theta}\|_1 && \text{(definition of } \boldsymbol{\phi}^\star)\\
&> \mathsf{OPT}, && \left(R_{\max} < \tfrac{\|\boldsymbol{\theta}\|_1}{\mathsf{OPT}}\right)
\end{aligned}$$

a contradiction! Therefore, we must have $\|W\boldsymbol{\theta}\|_1 \ge n + \frac{\|\boldsymbol{\theta}\|_1 - \mathsf{OPT}}{\mathsf{OPT}}$. □

F.5 Proof of Theorem 7.4

Proof of Theorem 7.4. Let $\boldsymbol{\theta}^\star \in [0,1]^V$ be an optimal IR solution with cost $\mathsf{OPT}^{\mathsf{IR}}$, $\boldsymbol{\theta} \in [0,1]^V$ be the optimal stable solution with cost $\mathsf{OPT}^{\mathsf{stable}}$, and $\alpha \in [0,1]$ be a signal value to be defined later. By Lemma 5.7, the fact that $\|\boldsymbol{\theta}\|_1 = \mathsf{OPT}^{\mathsf{stable}} > \mathsf{OPT}^{\mathsf{IR}} \ge \mathsf{OPT}$ implies that $\boldsymbol{\theta}$ must be wasteful.

Let $\mathcal{D}_A \in \Delta(\{0,\alpha\}^V)$ be the distribution defined by first sampling a random set $S \subseteq V$ by including each vertex v independently with probability $\theta_v^\star$, and then labeling each $v \in V$ with $s_v = \alpha \cdot \mathbb{1}[v \in S]$. Let $\mathcal{D}_{\boldsymbol{\theta}^\star} \in \Delta(2^V)$ denote the distribution of such a random set S. In addition, we define $\mathcal{D}_B \in \Delta([0,1]^V)$ as the degenerate distribution supported on $\{\boldsymbol{\theta}\}$. Let

$$\mathcal{S} = \{\theta_v : v \in V\} \cup \{\alpha\}$$

be the signal space including all distinct values in $\boldsymbol{\theta}$ and α. Note that we must have $0 \in \mathcal{S}$, because a stable solution can only be wasteful when there exists v such that $\theta_v = 0$; otherwise, the stability condition implies $W\boldsymbol{\theta} = \mathbf{1}$, i.e., $\boldsymbol{\theta}$ is not wasteful. Thus, we have $|\mathcal{S}| \le n+1$.

We define a parametrized family of distributions $\mathcal{D}_\epsilon \in \Delta(\mathcal{S}^V)$ for $\epsilon \in [0,1]$ as

$$\forall s \in \mathcal{S}^V, \qquad \mathcal{D}_\epsilon(s) = \epsilon \cdot \mathcal{D}_A(s) + (1-\epsilon) \cdot \mathcal{D}_B(s).$$

Recall from Lemma 3.2 that the signaling scheme $\mathcal{D}_\epsilon$ is persuasive if for all $\theta \in \mathcal{S}$, we have

$$\Delta_\theta = \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v = \theta\right]\left(\sum_{v'\in N(v)} W_{v,v'}s_{v'} - (1-\theta)\right)\right] \geq 0, \tag{15}$$

and the above is tight for all $\theta \in \mathcal{S}\setminus\{0\}$. We now analyze Δ_θ for each $\theta \in \mathcal{S}$.

- **Case 1.** $\theta \in \mathcal{S}\setminus\{0,\alpha\}$. Since $\boldsymbol{\theta}$ is stable, for any v such that $\theta = \theta_v > 0$, the feasibility condition must be tight, i.e.,

  $$u_v(\boldsymbol{\theta}) = \theta + \sum_{v'\in N(v)} W_{v,v'}\theta_{v'} = 1.$$

  This, together with the fact that $\mathcal{D}_B$ is a degenerate distribution on $\boldsymbol{\theta}$, implies

  $$\Delta_\theta^{\mathcal{D}_B} = \sum_{v\in V}\mathbb{1}\left[\theta_v = \theta\right]\left(\sum_{v'\in N(v)} W_{v,v'}\theta_{v'} - (1-\theta)\right) = 0.$$

  Therefore, we have

  $$\Delta_\theta^{\mathcal{D}_\epsilon} = (1-\epsilon)\Delta_\theta^{\mathcal{D}_B} = 0.$$

- **Case 2.** $\theta = \alpha$. If $\alpha = \theta_v$ for some v, the analysis in the previous case gives us $\Delta_\alpha^{\mathcal{D}_B} = 0$. On the other hand,

  $$\begin{aligned}\Delta_\alpha^{\mathcal{D}_A} &= \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_A}\left[\sum_{v\in V}\mathbb{1}\left[s_v = \alpha\right]\left(\sum_{v'\in N(v)} W_{v,v'}s_{v'} - (1-\alpha)\right)\right] \\ &= \mathop{\mathbb{E}}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[\sum_{v\in S}\left(\alpha\sum_{v'\in S} W_{v,v'} - 1\right)\right] && (s_v = \alpha\cdot\mathbb{1}\left[v\in S\right]) \\ &= \alpha\cdot\mathop{\mathbb{E}}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[\mathsf{Induced}(S)\right] - \mathsf{OPT}^{\mathsf{IR}}. && (\mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[|S|\right] = \mathsf{OPT}^{\mathsf{IR}})\end{aligned}$$

  Therefore, to guarantee $\Delta_\alpha^{\mathcal{D}_A} = 0$, we set

  $$\alpha = \frac{\mathsf{OPT}^{\mathsf{IR}}}{\mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[\mathsf{Induced}(S)\right]}. \tag{16}$$

  Note that $\alpha \leq 1$ because $\mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[\mathsf{Induced}(S)\right] \geq \mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[|S|\right] = \mathsf{OPT}^{\mathsf{IR}}$. In addition, we have

  $$\Delta_\alpha^{\mathcal{D}_\epsilon} = \epsilon\Delta_\alpha^{\mathcal{D}_A} + (1-\epsilon)\Delta_\alpha^{\mathcal{D}_B} = 0.$$

- **Case 3.** $\theta = 0$. We first analyze the slack in $\mathcal{D}_A$.

  $$\begin{aligned}\Delta_0^{\mathcal{D}_A} &= \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{D}_A}\left[\sum_{v\in V}\mathbb{1}\left[s_v = 0\right]\left(\sum_{v'\in N(v)} W_{v,v'}s_{v'} - 1\right)\right] \\ &= \mathop{\mathbb{E}}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}\left[\sum_{v\in V\setminus S}\left(\alpha\sum_{v'\in S} W_{v,v'} - 1\right)\right] && (s_v = \alpha\cdot\mathbb{1}\left[v\in S\right])\end{aligned}$$

$$
\begin{aligned}
&= \alpha \cdot \mathop{\mathbb{E}}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}[\mathsf{Cut}(S, V\setminus S)] - (n - \mathsf{OPT}^{\mathsf{IR}}) && (\mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}[|V\setminus S|] = n - \mathsf{OPT}^{\mathsf{IR}}) \\
&= \mathsf{OPT}^{\mathsf{IR}} \cdot \frac{\mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}[\mathsf{Cut}(S, V\setminus S)]}{\mathbb{E}_{S\sim\mathcal{D}_{\boldsymbol{\theta}^\star}}[\mathsf{Induced}(S)]} - (n - \mathsf{OPT}^{\mathsf{IR}}) && \text{(choice of } \alpha \text{ in eq. (16))} \\
&\geq \mathsf{OPT}^{\mathsf{IR}} \cdot \frac{n - 2\mathsf{OPT}^{\mathsf{IR}}}{(\mathsf{OPT}^{\mathsf{IR}})^2 + \mathsf{OPT}^{\mathsf{IR}}} - (n - \mathsf{OPT}^{\mathsf{IR}}) && \text{(last two properties in Lemma 5.2)} \\
&= -\frac{\mathsf{OPT}^{\mathsf{IR}}(n - \mathsf{OPT}^{\mathsf{IR}} + 1)}{\mathsf{OPT}^{\mathsf{IR}} + 1}.
\end{aligned}
$$

For $\Delta_0^{\mathcal{D}_B}$, since $(W\boldsymbol{\theta})_v - 1 = 0$ for all coordinates where $\theta_v > 0$, we have

$$
\begin{aligned}
\Delta_0^{\mathcal{D}_B} &= \sum_{v\in V} \mathbb{1}[\theta_v = 0]\left(\theta_v + \sum_{v'\in N(v)} W_{v,v'}\theta_{v'} - 1\right) \\
&= \sum_{v\in V} \mathbb{1}[\theta_v = 0]\left((W\boldsymbol{\theta})_v - 1\right) \\
&= \sum_{v\in V} \mathbb{1}[\theta_v = 0]\left((W\boldsymbol{\theta})_v - 1\right) + \sum_{v\in V} \mathbb{1}[\theta_v > 0]\left((W\boldsymbol{\theta})_v - 1\right) \quad (\theta_v > 0 \implies (W\boldsymbol{\theta})_v = 1) \\
&= \|W\boldsymbol{\theta}\|_1 - n \\
&\geq \frac{\|\boldsymbol{\theta}\|_1 - \mathsf{OPT}}{\mathsf{OPT}} = \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{\mathsf{OPT}},
\end{aligned}
$$

where the inequality follows from Lemma 7.5. Therefore, for the mixed distribution $\mathcal{D}_\epsilon$, we have

$$
\begin{aligned}
\Delta_0^{\mathcal{D}_\epsilon} &= \epsilon\Delta_0^{\mathcal{D}_A} + (1-\epsilon)\Delta_0^{\mathcal{D}_B} \\
&\geq -\epsilon \cdot \frac{\mathsf{OPT}^{\mathsf{IR}}(n - \mathsf{OPT}^{\mathsf{IR}} + 1)}{\mathsf{OPT}^{\mathsf{IR}} + 1} + (1-\epsilon)\frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{\mathsf{OPT}}.
\end{aligned}
$$

Therefore, for the signaling scheme $\mathcal{D}_\epsilon$ to be persuasive, it suffices to choose $\epsilon \in [0,1]$ that such that

$$
\begin{aligned}
&-\epsilon \cdot \frac{\mathsf{OPT}^{\mathsf{IR}}(n - \mathsf{OPT}^{\mathsf{IR}} + 1)}{\mathsf{OPT}^{\mathsf{IR}} + 1} + (1-\epsilon) \cdot \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{\mathsf{OPT}} \geq 0 \\
\iff \epsilon \leq\ & \frac{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}}{\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT} + \frac{\mathsf{OPT}\cdot\mathsf{OPT}^{\mathsf{IR}}(n - \mathsf{OPT}^{\mathsf{IR}}+1)}{\mathsf{OPT}^{\mathsf{IR}}+1}} = \frac{\mathsf{PoS} - 1}{\mathsf{PoS} - 1 + \frac{\mathsf{OPT}^{\mathsf{IR}}(n - \mathsf{OPT}^{\mathsf{IR}}+1)}{\mathsf{OPT}^{\mathsf{IR}}+1}},
\end{aligned}
$$

which justifies the choice of ϵ in Equation (10). The cost of this signaling scheme is

$$
\mathsf{Cost}(\mathcal{D}_\epsilon) = \epsilon\alpha \mathop{\mathbb{E}}_{S\sim\mathcal{D}_A}[|S|] + (1-\epsilon)\mathsf{OPT}^{\mathsf{stable}} \leq \mathsf{OPT}^{\mathsf{stable}} - \epsilon(\mathsf{OPT}^{\mathsf{stable}} - \mathsf{OPT}^{\mathsf{IR}}),
$$

where we have used $\alpha \leq 1$ and $\mathbb{E}_{S\sim\mathcal{D}_A}[|S|] = \mathsf{OPT}^{\mathsf{IR}}$.

□

G Omitted Proofs from Section 8

G.1 Proof of Theorem 8.1

Proof of Theorem 8.1. We prove the theorem using the example from Figure 4: Set $n = 3k + 2$ for some integer k. Let $C_1, C_2, \ldots, C_k$ be k cliques, each of size 3. Let v_1, v_2 be two additional vertices.

Each v_i is connected to each other and all the vertices in $C_{1:k}$. All edges have weight 1/2. The optimal solution is 2, achieved when both v_1 and v_2 play 1 while the other vertices play 0.

Fix any binary signaling scheme $\mathcal{D} \in \Delta(2^{[n]})$. Consider the distribution $\mathcal{D}'$ over $\vec{x} = (x_3, x_2, x_1, x_0, y) \in \{0, 1, \ldots, k\}^4 \times \{0, 1, 2\}$ induced by $\mathcal{D}$ through the following procedure:

- Sample $S \sim \mathcal{D}$.
- For $i \in \{0, 1, 2, 3\}$, let x_i be the number of cliques in $C_1, \ldots, C_k$ that contain exactly i vertices in S.
- Let $y = |S \cap \{v_1, v_2\}|$.

Note that $\mathcal{D}'$ is a sufficient statistics of $\mathcal{D}$ in terms of the persuasiveness of the induced binary signaling scheme: When $\mathcal{D}$ is the degenerate distribution at $S \subseteq [n]$, the resulting $\vec{x} = (x_3, x_2, x_1, x_0, y)$ satisfies

$$\begin{cases} |S| = 3x_3 + 2x_2 + x_1 + y \\ \mathsf{Cut}(S, V \setminus S) = 3x_3 + 3x_2 + 2x_1 - \frac{y}{2}(3x_3 + x_2 - x_1 - 3x_0) + \frac{\mathbb{1}[y=1]}{2} \\ \mathsf{Induced}(S) = |S| + 3x_3 + x_2 + y(3x_3 + 2x_2 + x_1) + \mathbb{1}[y = 2] \end{cases} \tag{17}$$

By Lemma 3.4, we have

$$\frac{\mathbb{E}_{\mathcal{D}}[\mathsf{Cut}(S, V \setminus S)]}{n - \mathbb{E}_{\mathcal{D}}[|S|]} \geq \frac{\mathbb{E}_{\mathcal{D}}[\mathsf{Induced}(S)]}{\mathbb{E}_{\mathcal{D}}[|S|]} = \frac{1}{\alpha},$$

where α is the value of the non-zero signal.

Substituting Equation (17) into the above condition, subtracting 1 from both sides, and using the fact that $n = 3(x_0 + x_1 + x_2 + x_3) + 2$, we obtain

$$\begin{aligned} &\frac{\mathbb{E}_{\mathcal{D}'}\left[3x_3 + 2x_2 + \frac{x_1 y}{2} + y - 3x_0(1 - \frac{y}{2}) - \frac{y}{2}(3x_3 + x_2) - \left(2 - \frac{\mathbb{1}[y=1]}{2}\right)\right]}{n - \mathbb{E}_{\mathcal{D}}[|S|]} \\ &\quad \geq \frac{\mathbb{E}_{\mathcal{D}'}[3x_3 + x_2 + y(3x_3 + 2x_2 + x_1) + \mathbb{1}[y = 2]]}{\mathbb{E}_{\mathcal{D}}[|S|]}. \end{aligned} \tag{18}$$

We compare the numerators of both sides of Equation (18) when (x_3, x_2, x_1, x_0, y) is deterministic. When $y \in \{0, 2\}$, we have $2 \cdot \mathbb{1}[y = 2] \geq y$, and it follows that

$$2 \cdot [3x_3 + x_2 + y(3x_3 + 2x_2 + x_1) + \mathbb{1}[y = 2]] \geq 3x_3 + 2x_2 + \frac{x_1 y}{2} + y,$$

which is, in turn, lower bounded by the numerator on the left-hand side of Equation (18). When $y = 1$ and $x_1 + x_2 + x_3 \geq 1$, we have

$$\begin{aligned} 2 \cdot [3x_3 + x_2 + y(3x_3 + 2x_2 + x_1) + \mathbb{1}[y = 2]] &= 12x_3 + 6x_2 + 2x_1 \\ &= (11x_3 + 5x_2) + x_1 + (x_1 + x_2 + x_3) \\ &\geq 3x_3 + 2x_2 + \frac{x_1 y}{2} + y, \end{aligned}$$

which is again lower bounded by the numerator on the left-hand side of Equation (18). Finally, in the remaining case that $x_1 = x_2 = x_3 = 0$, $x_0 = k$, and $y = 1$, the numerator on the right-hand side of Equation (18) is 0, while the numerator on the left is given by

$$y - 3x_0(1 - y/2) - (2 - \mathbb{1}[y = 1]/2) = -\frac{3}{2}k - \frac{1}{2} < 0.$$

Let LHS and RHS denote the left- and right-hand sides of Equation (18), respectively. Then, the above case analysis shows that

$$2 \cdot \frac{\mathbb{E}_{\mathcal{D}}\left[|S|\right]}{n - \mathbb{E}_{\mathcal{D}}\left[|S|\right]} \cdot \text{RHS} \geq \text{LHS} \geq \text{RHS},$$

which implies $\mathbb{E}_{\mathcal{D}}\left[|S|\right] \geq n/3 = \Omega(n)$.

Recall that by Lemma 3.4, the cost of the binary scheme is given by $\alpha \cdot \mathbb{E}_{\mathcal{D}}\left[|S|\right]$. Therefore, it is left to show $\alpha = \Omega(1)$. To prove $\alpha = \Omega(1)$, it suffices to upper bound $\frac{\mathbb{E}_{\mathcal{D}}[\mathsf{Induced}(S)]}{\mathbb{E}_{\mathcal{D}}[|S|]}$ by a constant. We have

$$\begin{aligned}\frac{\mathbb{E}_{\mathcal{D}}\left[\mathsf{Induced}(S)\right]}{\mathbb{E}_{\mathcal{D}}\left[|S|\right]} - 1 &= \frac{\mathbb{E}_{\mathcal{D}'}\left[3x_3 + x_2 + y(3x_3 + 2x_2 + x_1) + \mathbb{1}\left[y = 2\right]\right]}{\mathbb{E}_{\mathcal{D}}\left[|S|\right]} \\ &\leq \frac{\mathbb{E}_{\mathcal{D}'}\left[9x_3 + 5x_2 + 2x_1 + \mathbb{1}\left[y = 2\right]\right]}{\mathbb{E}_{\mathcal{D}'}\left[3x_3 + 2x_2 + x_1 + y\right]} \leq 3.\end{aligned}$$

Therefore, $\alpha = \frac{\mathbb{E}_{\mathcal{D}}[|S|]}{\mathbb{E}_{\mathcal{D}}[\mathsf{Induced}(S)]} \geq \frac{1}{4} = \Omega(1)$. This lower bounds the cost of the binary scheme by $\alpha \cdot \mathbb{E}_{\mathcal{D}}\left[|S|\right] = \Omega(n)$. □

G.2 Projection to a low-dimensional space

Consider an identity-independent persuasive signaling scheme specified by $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ with signal space $\mathcal{S}$. The scheme naturally induces a distribution $\mathcal{D}'$ over $(x, y, \alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathcal{S}^{k+2}$ in the following way:

- First, we sample a set of signals from $\mathcal{D}_\varphi$.
- Let x and y be the signals sent to the two center vertices.
- Choose one of the k^2 cliques uniformly at random, and set $\alpha_1, \ldots, \alpha_k$ to the signals sent to the k vertices in that clique.

Define random variable $\mathsf{sum} := \sum_{i=1}^{k} \alpha_i$. It is easy to verify that the cost of the signaling scheme is given by

$$\mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[x + y + k^2 \cdot \mathsf{sum}\right],$$

so our goal is to prove that $\mathbb{E}_{\mathcal{D}'}\left[\mathsf{sum}\right] = \Omega(1)$.

Recall from Definition 3.1 that Δ_θ is the following quantity:

$$\Delta_\theta = \mathsf{Contrib}_\theta - (1 - \theta) \cdot \mathsf{Num}_\theta,$$

which is the total amount of slack at signal value θ. Lemma 3.2 states that a valid scheme must satisfy $\Delta_0 \geq 0$ and $\Delta_\theta = 0$ for every $\theta > 0$.

Our first step is to re-write each Δ_θ as an expectation over the distribution of $(x, y, \alpha_1, \ldots, \alpha_k)$. This reduces the dimensionality of the signaling scheme from $n = k^3 + 2$ to $k + 2$, and slightly simplifies the notations in the remainder of the proof.

Lemma G.1 (informal)**.** *The slacks $(\Delta_\theta)_{\theta\in\mathcal{S}}$ can be equivalently defined as the expected outcome of the following procedure:*

- *Draw $(x, y, \alpha_1, \ldots, \alpha_k) \sim \mathcal{D}'$.*

- $\Delta_x \leftarrow \Delta_x + \frac{y+k^2\mathsf{sum}}{2} - (1-x)$.
- $\Delta_y \leftarrow \Delta_y + \frac{x+k^2\mathsf{sum}}{2} - (1-y)$.
- *For every* $i \in [k]$, $\Delta_{\alpha_i} \leftarrow \Delta_{\alpha_i} + k^2 \cdot \left[\frac{\mathsf{sum}+x+y+\alpha_i}{2} - 1\right]$.

Informally, the k^2 factors in the above account for the fact that every clique is only sampled with probability $1/k^2$.

Below is a more formal statement of the lemma above.

Lemma G.2 (Formal version of Lemma G.1)**.** *Let $\mathcal{D}_\varphi \in \Delta(\mathcal{S}^V)$ be the distribution that specifies a signaling scheme for the graph. Formally, the scheme draws*

$$s = (s_{c,1}, s_{c,2}, s_{1,1}, \ldots, s_{1,k}, s_{2,1}, \ldots, s_{2,k}, \ldots, s_{k^2,1}, \ldots, s_{k^2,k}) \sim \mathcal{D}_\varphi,$$

labels the two centers with $s_{c,1}$ and $s_{c,2}$, and labels the j-th vertex in the i-th clique with $s_{i,j}$.

Let $\mathcal{D}' \in \Delta(\mathcal{S}^{k+2})$ be the projection of $\mathcal{D}_\varphi$, i.e., $\mathcal{D}'$ is the distribution of

$$(s_{c,1}, s_{c,2}, s_{i,1}, s_{i,2}, \ldots, s_{i,k})$$

when $s \sim \mathcal{D}_\varphi$ and i is drawn uniformly at random from $[k^2]$.

Then, for any $\theta \in \mathcal{S}$, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{s\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v = \theta\right] \cdot \frac{1}{2} \sum_{u\in N(v)} s_u\right] - (1-\theta) \cdot \mathop{\mathbb{E}}_{s\sim\mathcal{D}_\varphi}\left[\sum_{v\in V} \mathbb{1}\left[s_v = \theta\right]\right] \\
= &\mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[\mathbb{1}\left[x = \theta\right] \cdot \left(\frac{y + k^2\mathsf{sum}}{2} - (1-x)\right)\right] + \mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[\mathbb{1}\left[y = \theta\right] \cdot \left(\frac{x + k^2\mathsf{sum}}{2} - (1-y)\right)\right] \\
&+ \sum_{i=1}^{k} \mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[\mathbb{1}\left[\alpha_i = \theta\right] \cdot k^2 \cdot \left(\frac{\mathsf{sum} + x + y + \alpha_i}{2} - 1\right)\right].
\end{aligned}$$

Proof. It suffices to prove the identity for deterministic values of $s_{c,1}$, $s_{c,2}$, and $(s_{i,j})_{(i,j)\in[k^2]\times[k]}$, i.e., when $\mathcal{D}_\varphi$ is degenerate. The general case then follows from taking an expectation. To this end, we show that each $s_{c,1}$, $s_{c,2}$, and $s_{i,j}$ contributes the same amount to both sides of the equation. For simplicity, we will assume that the k^3+2 entries of s are distinct; the general case follows from the same argument.

Contribution from $s_{c,1}$ and $s_{c,2}$. When $\theta = s_{c,1}$, the left-hand side is given by

$$\frac{1}{2} \sum_{u\in N((c,1))} s_u - (1 - s_{c,1}) = \frac{s_{c,2}}{2} + \frac{1}{2} \sum_{i=1}^{k^2} \sum_{j=1}^{k} s_{i,j} - (1 - s_{c,1}).$$

The right-hand side is

$$\begin{aligned}
\mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[\frac{y + k^2\mathsf{sum}}{2} - (1-x)\right] &= \frac{s_{c,2}}{2} - (1 - s_{c,1}) + \frac{k^2}{2} \mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[\mathsf{sum}\right] \\
&= \frac{s_{c,2}}{2} - (1 - s_{c,1}) + \frac{k^2}{2} \cdot \frac{1}{k^2} \sum_{i=1}^{k^2} \sum_{j=1}^{k} s_{i,j},
\end{aligned}$$

which is equal to the left-hand side.

By symmetry, the contributions from $s_{c,2}$ to both sides are also equal.

Contribution from $s_{i,j}$. When $\theta = s_{i,j}$, the left-hand side reduces to

$$\frac{1}{2}\sum_{u\in N((i,j))} s_u - (1-s_{i,j}) = \frac{s_{c,1}+s_{c,2}}{2} + \sum_{j'\in[k]\setminus\{j\}} \frac{s_{i,j'}}{2} - 1 + s_{i,j} = \frac{s_{c,1}+s_{c,2}+s_{i,j}}{2} + \sum_{j'\in[k]} \frac{s_{i,j'}}{2} - 1,$$

while the right-hand side is given by

$$\mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}\left[\mathbb{1}\left[\alpha_j = s_{i,j}\right]\cdot k^2 \cdot \left(\frac{\mathsf{sum}+x+y+\alpha_j}{2} - 1\right)\right].$$

Note that $\alpha_j = s_{i,j}$ holds only when $\alpha_1,\ldots,\alpha_k$ are equal to $s_{i,1},\ldots,s_{i,k}$, which happens with probability $1/k^2$ by definition of $\mathcal{D}'$. Thus, the expression above can be simplified to

$$\frac{1}{k^2}\cdot k^2 \cdot \left(\frac{\sum_{j'=1}^{k} s_{i,j'} + s_{c,1}+s_{c,2}+s_{i,j}}{2} - 1\right) = \frac{s_{c,1}+s_{c,2}+s_{i,j}}{2} + \sum_{j'\in[k]} \frac{s_{i,j'}}{2} - 1,$$

which is exactly the left-hand side. This completes the proof. □

G.3 Choice of the test function

Recall that our goal is to show that any distribution $\mathcal{D}'$ over $(x, y, \alpha_1, \ldots, \alpha_k)$ induced by a valid signaling scheme $\mathcal{D}_\varphi$ must satisfy $\mathbb{E}_{\mathcal{D}'}[\mathsf{sum}] = \Omega(1)$. In order for $\mathcal{D}'$ to be induced by a valid scheme, it must satisfy $\Delta_0 \geq 0$ and $\Delta_\theta = 0$ for every $\theta \in \mathcal{S}\setminus\{0\}$, where $(\Delta_\theta)_{\theta\in\mathcal{S}}$ are obtained from $\mathcal{D}'$ via Lemma G.1. While the number of constraints might be large, we will carefully take only a few linear combinations of them, such that they are sufficient for lower bounding $\mathbb{E}_{\mathcal{D}'}[\mathsf{sum}]$.

In particular, we note that for every function $f : \mathcal{S} \to \mathbb{R}$ with $f(0) \geq 0$, we must have

$$\sum_{\theta\in\mathcal{S}} f(\theta)\cdot\Delta_\theta \geq 0.$$

The hope is that we can choose a few simple functions f such that, after plugging x, y, and α_i into the inequality above, we obtain a good lower bound on $\mathbb{E}_{\mathcal{D}'}[\mathsf{sum}]$.

The constant function. We start with the most simple choice of $f(\theta) \equiv 1$. By Lemma G.1, the contribution of the combination $(x, y, \alpha_1, \ldots, \alpha_k)$ to $\sum_{\theta\in\mathcal{S}} f(\theta)\cdot\Delta_\theta = \sum_{\theta\in\mathcal{S}}\Delta_\theta$ is given by:

$$\begin{aligned}
&\left[\frac{y+k^2\mathsf{sum}}{2} - (1-x)\right] + \left[\frac{x+k^2\mathsf{sum}}{2} - (1-y)\right] + k^2\cdot\sum_{i=1}^{k}\left[\frac{\mathsf{sum}+x+y+\alpha_i}{2} - 1\right] \\
&= \frac{3}{2}(x+y) + k^2\mathsf{sum} - 2 + k^3\cdot\left[\frac{\mathsf{sum}+x+y}{2} - 1\right] + \frac{k^2}{2}\mathsf{sum} \\
&= \left(\frac{1}{2}k^3 + \frac{3}{2}k^2\right)\mathsf{sum} + \left(\frac{1}{2}k^3 + \frac{3}{2}\right)(x+y) - (k^3+2).
\end{aligned}$$

By linearity, the condition $\sum_{\theta\in\mathcal{S}}\Delta_\theta \geq 0$ can be written as

$$\left(\frac{1}{2}k^3 + \frac{3}{2}k^2\right)\mathop{\mathbb{E}}_{\mathcal{D}'}[\mathsf{sum}] + \left(\frac{1}{2}k^3 + \frac{3}{2}\right)\mathop{\mathbb{E}}_{\mathcal{D}'}[x+y] \geq (k^3+2).$$

Note that this condition alone is not enough for lower bounding $\mathbb{E}_{\mathcal{D}'}[\mathsf{sum}]$, since the above can be easily satisfied by setting $\mathsf{sum} \equiv 0$ and $x, y \equiv 1$.

Getting rid of signal 1. Later, we will consider another choice of $f(\theta) = \frac{\theta}{1-\theta}$, which is ill-defined at $\theta = 1$. To avoid this issue, we now take a detour and show that sending signal 1 is essentially useless, so it is without loss of generality that $1 \notin \mathcal{S}$.

Note that whenever signal 1 is sent to a vertex, all the neighbouring vertices must receive signal 0. This is because, according to Lemma G.1, whenever Δ_1 is changed, the change is always non-negative. Therefore, to ensure $\Delta_1 = 0$, there cannot be any non-zero contribution to vertices that receive signal 1. In other words, all neighbours of such vertices must receive signal 0.

Therefore, the support of $\mathcal{D}'$ must be contained in $(\mathcal{S} \setminus \{1\})^{k+2} \cup \{e_1, e_2, \ldots, e_{k+2}\}$, where e_i is the vector with 1 at the i-th coordinate and zeros elsewhere. Using Lemma G.1, we can verify that the contribution of each e_i to Δ_0 is non-positive, while Δ_θ is unaffected for $\theta > 0$. In particular, when either $x = 1$ or $y = 1$, Δ_0 increases by

$$\left(\frac{1}{2} - 1\right) + k \cdot k^2 \cdot \left(\frac{1}{2} - 1\right) = -\frac{k^3 + 1}{2} < 0.$$

When one of the α_i's is equal to 1, Δ_0 gets increased by

$$2 \cdot \left(\frac{k^2}{2} - 1\right) + (k-1) \cdot k^2 \cdot \left(\frac{1}{2} - 1\right) = -\frac{1}{2}k^3 + \frac{3}{2}k^2 - 2 \leq 0.$$

Therefore, we may let $\mathcal{D}''$ be the restriction of $\mathcal{D}'$ to $(\mathcal{S} \setminus \{1\})^{k+2}$. As argued above, $\mathcal{D}''$ is still valid in the sense that $(\Delta_\theta)_{\theta \in \mathcal{S} \setminus \{1\}}$ induced by $\mathcal{D}''$ (according to Lemma G.1) still satisfies $\Delta_0 \geq 0$ and $\Delta_\theta = 0$ for all $\theta > 0$.

We still need to show that the restriction to $(\mathcal{S} \setminus \{1\})^{k+2}$ does not significantly increase the cost, i.e., $\mathbb{E}_{\mathcal{D}''}[\mathsf{sum}]$ should be $O(1) \cdot \mathbb{E}_{\mathcal{D}'}[\mathsf{sum}]$. To this end, it suffices to prove that $\mathcal{D}'((\mathcal{S} \setminus \{1\})^{k+2})$ is lower bounded by $\Omega(1)$, so that the normalization does not blow up the cost. We start by claiming that

$$\mathcal{D}'(\{e_3, e_4, \ldots, e_{k+2}\}) = \Pr_{(x,y,\alpha) \sim \mathcal{D}'}[\alpha_1 = 1 \vee \alpha_2 = 1 \vee \cdots \vee \alpha_k = 1] \leq \frac{1}{10}.$$

This holds because, otherwise, we have $\mathbb{E}_{(x,y,\alpha) \sim \mathcal{D}'}[\mathsf{sum}] \geq \frac{1}{10} \cdot 1 = \Omega(1)$, and we are done. Furthermore, we claim that

$$\mathcal{D}'(\{e_1, e_2\}) = \Pr_{(x,y,\alpha) \sim \mathcal{D}'}[x = 1 \vee y = 1] \leq \frac{1}{10}.$$

Earlier, we showed that when either $x = 1$ or $y = 1$, $\sum_{\theta \in \mathcal{S}} \Delta_\theta$ is decreased by $\frac{k^3+1}{2}$. Furthermore, in general, each fixed value of $(x, y, \alpha_1, \ldots, \alpha_k)$ increases $\sum_{\theta \in \mathcal{S}} \Delta_\theta$ by

$$\left(\frac{1}{2}k^3 + \frac{3}{2}k^2\right)\mathsf{sum} + \left(\frac{1}{2}k^3 + \frac{3}{2}\right)(x + y) - (k^3 + 2) \leq \left(\frac{1}{2}k^3 + \frac{3}{2}k^2\right)\mathsf{sum} + 1.$$

This gives

$$\begin{aligned}
0 \leq \sum_{\theta \in \mathcal{S}} \Delta_\theta \leq & -\Pr_{(x,y,\alpha) \sim \mathcal{D}'}[x = 1 \vee y = 1] \cdot \frac{k^3 + 1}{2} \\
& + \mathop{\mathbb{E}}_{(x,y,\alpha) \sim \mathcal{D}'}\left[\mathbb{1}[x \neq 1 \wedge y \neq 1] \cdot \left[\left(\frac{1}{2}k^3 + \frac{3}{2}k^2\right)\mathsf{sum} + 1\right]\right] \\
\leq & -\Pr_{(x,y,\alpha) \sim \mathcal{D}'}[x = 1 \vee y = 1] \cdot \frac{k^3 + 1}{2} + \left(\frac{1}{2}k^3 + \frac{3}{2}k^2\right) \mathop{\mathbb{E}}_{(x,y,\alpha) \sim \mathcal{D}'}[\mathsf{sum}] + 1,
\end{aligned}$$

which, together with $\Pr[x = 1 \vee y = 1] \geq 1/10$, would imply $\mathbb{E}[\mathsf{sum}] = \Omega(1)$. Therefore, without loss of generality, we may assume that $\mathcal{D}'((\mathcal{S} \setminus \{1\})^{k+2}) \geq 1 - \frac{1}{10} - \frac{1}{10} = \frac{4}{5}$. Therefore,

$$\mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}''}[\mathsf{sum}] \leq \frac{5}{4} \mathop{\mathbb{E}}_{(x,y,\alpha)\sim\mathcal{D}'}[\mathsf{sum}].$$

In particular, proving $\mathbb{E}_{\mathcal{D}''}[\mathsf{sum}] = \Omega(1)$ would imply $\mathbb{E}_{\mathcal{D}'}[\mathsf{sum}] = \Omega(1)$.

Another choice of f. Now, we assume that $1 \notin \mathcal{S}$, and examine $\sum_{\theta\in\mathcal{S}} f(\theta) \cdot \Delta_\theta$ when $f(\theta) = \frac{\theta}{1-\theta}$. Again, Lemma G.1 implies that the contribution of a fixed set of values $(x, y, \alpha_1, \ldots, \alpha_k)$ to $\sum_{\theta\in\mathcal{S}} \frac{\theta}{1-\theta} \cdot \Delta_\theta$ is given by

$$\frac{x}{1-x} \cdot \left[\frac{y + k^2\mathsf{sum}}{2} - (1-x)\right] + \frac{y}{1-y} \cdot \left[\frac{x + k^2\mathsf{sum}}{2} - (1-y)\right] + \sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_i} \cdot k^2 \left[\frac{\mathsf{sum} + x + y + \alpha_i}{2} - 1\right].$$

The sum of the first two terms above can be simplified into

$$\frac{xy}{2} \cdot \left(\frac{1}{1-x} + \frac{1}{1-y}\right) + \frac{k^2\mathsf{sum}}{2} \cdot \left(\frac{x}{1-x} + \frac{y}{1-y}\right) - (x+y),$$

while the last summation can be re-written as

$$k^2 \sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_i} \cdot \left[\frac{\mathsf{sum} + x + y - 1}{2} - \frac{1-\alpha_i}{2}\right] = \frac{k^2}{2}(\mathsf{sum} + x + y - 1) \sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_i} - \frac{k^2}{2}\mathsf{sum}.$$

G.4 Verify the dual feasibility

We will show that, for any finite signal space $\mathcal{S} \subset [0, 1)$ and every fixed choice of $(x, y, \alpha_1, \ldots, \alpha_k) \in \mathcal{S}^{k+2}$, it holds that

$$\mathsf{sum} + \beta_2 \cdot \sum_{\theta\in\mathcal{S}} \frac{\theta}{1-\theta} \cdot \Delta_\theta \geq \Omega(1) + \beta_1 \cdot \sum_{\theta\in\mathcal{S}} \Delta_\theta, \tag{19}$$

where $\beta_2 = \frac{1}{4k}$, $\beta_1 = \frac{1}{2k^2}$, and the $\Omega(1)$ notation hides a positive universal constant that does not depend on $(x, y, \alpha_1, \ldots, \alpha_k)$, as long as k is sufficiently large.

Assuming that eq. (19) holds, taking an expectation and rearranging shows that

$$\mathop{\mathbb{E}}_{\mathcal{D}'}[\mathsf{sum}] \geq \Omega(1) + \beta_1 \cdot \mathbb{E}\left[\sum_{\theta\in\mathcal{S}} \Delta_\theta\right] - \beta_2 \cdot \mathbb{E}\left[\sum_{\theta\in\mathcal{S}} \frac{\theta}{1-\theta} \cdot \Delta_\theta\right] \geq \Omega(1),$$

which implies the $\Omega(k^2) = \Omega(n^{2/3})$ lower bound on the cost and proves Theorem 8.2.

Now we plug $\beta_1 = \frac{1}{4k}$, $\beta_2 = \frac{1}{2k^2}$, as well as the expressions for $\sum_{\theta\in\mathcal{S}} \Delta_\theta$ and $\sum_{\theta\in\mathcal{S}} \frac{\theta}{1-\theta} \cdot \Delta_\theta$ into eq. (19). It is sufficient to prove that

$$\mathsf{sum} + \frac{xy}{8k}\left(\frac{1}{1-x} + \frac{1}{1-y}\right) + \frac{k}{8}\mathsf{sum} \cdot \left(\frac{x}{1-x} + \frac{y}{1-y}\right) - \frac{x+y}{4k} + \frac{k(\mathsf{sum} + x + y - 1)}{8} \sum_{i=1}^{k} \frac{\alpha_i}{1-\alpha_i} - \frac{k}{8}\mathsf{sum}$$
$$\geq \Omega(1) + \left(\frac{k}{4} + \frac{3}{4}\right)\mathsf{sum} + \left(\frac{k}{4} + \frac{3}{4k^2}\right)(x+y) - \left(\frac{k}{2} + \frac{1}{k^2}\right).$$

Since $x+y \le 2 = O(1)$, every additive term of form $(x+y)/\mathrm{poly}(k)$ or $1/\mathrm{poly}(k)$ can be absorbed into the $\Omega(1)$ gap (for all sufficiently large k). Thus, it suffices to show

$$\begin{aligned}&\frac{k}{2} + \frac{xy}{8k}\left(\frac{1}{1-x} + \frac{1}{1-y}\right) + \frac{k}{8}\mathsf{sum}\cdot\left(\frac{x}{1-x} + \frac{y}{1-y}\right) + \frac{k(\mathsf{sum}+x+y-1)}{8}\sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_i} \\ &\ge \Omega(1) + \left(\frac{3}{8}k - \frac{1}{4}\right)\mathsf{sum} + \frac{k}{4}(x+y).\end{aligned} \tag{20}$$

We will consider the following three different cases:

- **Case 1:** $x+y<1$ and $\mathsf{sum}+x+y-1 \ge 0$. Since $x, y, \alpha_1, \ldots, \alpha_k \in [0,1)$, we have

$$\frac{x}{1-x} + \frac{y}{1-y} \ge x+y \quad \text{and} \quad \sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_i} \ge \sum_{i=1}^{k}\alpha_i = \mathsf{sum}.$$

  Furthermore, we relax the $\frac{xy}{8k}\left(\frac{1}{1-x} + \frac{1}{1-y}\right)$ term in eq. (20) to 0. Then, it suffices to prove the following:

$$\frac{k}{2} + \frac{k}{8}\mathsf{sum}\cdot(x+y) + \frac{k(\mathsf{sum}+x+y-1)}{8}\cdot\mathsf{sum} \ge \Omega(1) + \left(\frac{3}{8}k - \frac{1}{4}\right)\mathsf{sum} + \frac{k}{4}(x+y). \tag{21}$$

  For fixed sum, both sides of eq. (21) are affine in $x+y$, so it suffices to verify it at $x+y=1$ and $x+y = \max\{1-\mathsf{sum}, 0\}$. At $x+y=1$, eq. (21) gets reduced to

$$\frac{k}{8}\mathsf{sum}^2 - \left(\frac{k}{4} - \frac{1}{4}\right)\mathsf{sum} + \frac{k}{4} \ge \Omega(1),$$

  which is true since the left-hand side is equal to $\frac{k}{8}(\mathsf{sum}-1)^2 + \frac{\mathsf{sum}}{4} + \frac{k}{8} \ge \frac{k}{8} \ge \frac{1}{8}$.

  At $x+y = \max\{1-\mathsf{sum}, 0\}$, if $\mathsf{sum} \le 1$ (so that $x+y = 1-\mathsf{sum}$), eq. (21) is equivalent to

$$\frac{\mathsf{sum}}{4} + \frac{k}{4}\left(1 - \frac{\mathsf{sum}^2}{2}\right) \ge \Omega(1),$$

  which is true since $\mathsf{sum} \le 1$ implies $1 - \frac{\mathsf{sum}^2}{2} \ge 1/2$. In the other case that $\mathsf{sum} > 1$, we have $x+y=0$, and this reduces eq. (21) to

$$\frac{k}{2} + \frac{k}{8}\mathsf{sum}(\mathsf{sum}-1) - \left(\frac{3}{8}k - \frac{1}{4}\right)\mathsf{sum} \ge \Omega(1),$$

  which always holds since the left-hand side is equal to $\frac{k}{8}(\mathsf{sum}-2)^2 + \frac{\mathsf{sum}}{4} \ge \frac{\mathsf{sum}}{4} > \frac{1}{4}$.

- **Case 2:** $x+y<1$ and $\mathsf{sum}+x+y-1<0$. Again, we may apply the relaxation

$$\frac{xy}{8k}\left(\frac{1}{1-x} + \frac{1}{1-y}\right) + \frac{k}{8}\mathsf{sum}\cdot\left(\frac{x}{1-x} + \frac{y}{1-y}\right) \ge \frac{k}{8}\mathsf{sum}\cdot(x+y).$$

  However, since the factor $\mathsf{sum}+x+y-1$ is now negative, the last term on the left-hand side of eq. (20) (namely, $\frac{k(\mathsf{sum}+x+y-1)}{8}\sum_{i=1}^{k}\frac{\alpha_i}{1-\alpha_i}$) is minimized when α is a permutation of $(\mathsf{sum}, 0, 0, \ldots, 0)$. Thus, we need to prove the following inequality:

$$\frac{k}{2} + \frac{k}{8}\mathsf{sum}(x+y) + \frac{k(\mathsf{sum}+x+y-1)}{8}\cdot\frac{\mathsf{sum}}{1-\mathsf{sum}} \ge \Omega(1) + \left(\frac{3}{8}k - \frac{1}{4}\right)\mathsf{sum} + \frac{k}{4}(x+y). \tag{22}$$

Again, it suffices to verify the above at $x+y=0$ and $x+y=1-\mathsf{sum}$, respectively. At $x+y=0$, eq. (22) reduces to

$$\frac{k}{2}\cdot(1-\mathsf{sum})+\frac{1}{4}\mathsf{sum}\geq\Omega(1),$$

which is true since $\frac{k}{2}\cdot(1-\mathsf{sum})+\frac{1}{4}\mathsf{sum}\geq\frac{1}{4}\cdot(1-\mathsf{sum})+\frac{1}{4}\mathsf{sum}=\frac{1}{4}$.

At $x+y=1-\mathsf{sum}$, we get exactly the same inequality as the "$x+y=\max\{1-\mathsf{sum},0\}$ and $\mathsf{sum}\leq 1$" part of Case 1, which has already been verified.

- **Case 3:** $x+y\geq 1$ (and thus, $\mathsf{sum}+x+y-1\geq 0$). In this case, we claim that the minimum of both $\frac{xy}{8k}\left(\frac{1}{1-x}+\frac{1}{1-y}\right)$ and $\frac{x}{1-x}+\frac{y}{1-y}$ are achieved at $x=y\geq 1/2$. Write $x=\mu+\delta$ and $y=\mu-\delta$ for $\mu=\frac{x+y}{2}\in[1/2,1)$. We have

$$xy\cdot\left(\frac{1}{1-x}+\frac{1}{1-y}\right)=(\mu^2-\delta^2)\cdot\frac{2-(\mu+\delta)-(\mu-\delta)}{(1-\mu-\delta)(1-\mu+\delta)}=(2-2\mu)\cdot\left[1+\frac{\mu^2-(1-\mu)^2}{(1-\mu)^2-\delta^2}\right],$$

which is minimized at $\delta=0$, since $\mu\in[1/2,1)$ guarantees that $2-2\mu>0$ and $\mu^2-(1-\mu)^2\geq 0$.

Similarly, $\frac{x}{1-x}+\frac{y}{1-y}$ can be written as

$$\frac{1}{1-x}+\frac{1}{1-y}-2=\frac{2-2\mu}{(1-\mu)^2-\delta^2}-2,$$

which is also minimized at $\delta=0$.

Thus, it suffices to prove eq. (20) for the $x=y$ case, i.e., for all $x\in[1/2,1)$,

$$\frac{k}{2}+\frac{1}{4k}\cdot\frac{x^2}{1-x}+\frac{k}{4}\mathsf{sum}\cdot\frac{x}{1-x}+\frac{k(\mathsf{sum}+2x-1)}{8}\cdot\mathsf{sum}\geq\Omega(1)+\left(\frac{3}{8}k-\frac{1}{4}\right)\mathsf{sum}+\frac{k}{2}x.$$

Consider the function

$$\begin{aligned}g(x,s)&\coloneqq\frac{k}{2}+\frac{1}{4k}\cdot\frac{x^2}{1-x}+\frac{k}{4}s\cdot\frac{x}{1-x}+\frac{k(s+2x-1)}{8}\cdot s-\left(\frac{3}{8}k-\frac{1}{4}\right)s-\frac{k}{2}x\\&=\frac{k}{8}s^2-\left(\frac{1}{2}k-\frac{1}{4}-\frac{k}{4}x-\frac{k}{4}\cdot\frac{x}{1-x}\right)s+\frac{k}{2}(1-x)+\frac{1}{4k}\cdot\frac{x^2}{1-x}.\end{aligned}$$

We need to prove that $g(x,s)\geq\Omega(1)$ holds for all $x\in[1/2,1)$ and $s\geq 0$. For any fixed x, $g(x,s)$ is quadratic in s, and we have

$$\inf_{s\geq 0}g(x,s)=\begin{cases}g(x,0), & s^*(x)<0,\\ g(x,s^*(x)), & s^*(x)\geq 0,\end{cases}$$

where

$$s^*(x)\coloneqq\frac{1}{k/4}\left(\frac{1}{2}k-\frac{1}{4}-\frac{k}{4}x-\frac{k}{4}\cdot\frac{x}{1-x}\right)=2-x-\frac{x}{1-x}-\frac{1}{k}.$$

The case that $s^*(x)<0$ is easy, since for any $x\geq 1/2$,

$$g(x,0)=\frac{k}{2}(1-x)+\frac{1}{4k}\cdot\frac{x^2}{1-x}\geq 2\sqrt{\frac{k}{2}(1-x)\cdot\frac{1}{4k}\cdot\frac{x^2}{1-x}}=\frac{2x}{\sqrt{8}}\geq\frac{1}{2\sqrt{2}}=\Omega(1).$$

To handle the $s^*(x) \geq 0$ case, we note that

$$\inf_{s \geq 0} g(x, s) = g(x, s^*(x)) = g(x, 0) - \frac{k}{8}[s^*(x)]^2,$$

since the minimum of the quadratic function $f(x) = ax^2 - bx + c$ where $a > 0$ is given by $c - \frac{b^2}{4a} = f(0) - a(x^*)^2$, achieved at $x^* = \frac{b}{2a}$. We will prove that whenever $s^*(x) \geq 0$,

$$g(x, 0) \geq \frac{k}{4}[s^*(x)]^2. \tag{23}$$

This then implies

$$\inf_{s \geq 0} g(x, s) = g(x, 0) - \frac{k}{8}[s^*(x)]^2 \geq g(x, 0) - g(x, 0)/2 = \Omega(1),$$

where the last step follows from our previous argument for $g(x, 0) = \Omega(1)$.

Since $2 - 1/k - x - x/(1 - x) = s^*(x) \geq 0$, we have $x + x/(1 - x) \leq 2$, which further implies $x \leq 2 - \sqrt{2} \leq 3/5$. Then, $g(x, 0)$ can be lower bounded as follows:

$$g(x, 0) = \frac{k}{2}(1 - x) + \frac{1}{4k} \cdot \frac{x^2}{1 - x} \geq \frac{k}{2}(1 - x) \geq \frac{k}{5}.$$

Furthermore, $x \in [1/2, 1)$ implies that

$$s^*(x) = 2 - \frac{1}{k} - x - \frac{x}{1 - x} \leq 2 - 0 - \frac{1}{2} - 1 = \frac{1}{2}.$$

Therefore, eq. (23) follows from

$$g(x, 0) \geq \frac{k}{5} > \frac{k}{4} \cdot (1/2)^2 \geq \frac{k}{4}[s^*(x)]^2.$$

This finishes the proof for Case 3.