**Optimal Signaling of Content Accuracy: Engagement vs. Misinformation**

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Operations Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Candogan, Ozan; The University of Chicago, Booth School of Business Drakopoulos, Kimon; University of Southern California Marshall School of Business, Data Sciences and Operations</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Games/group decisions, Networks/graphs, Applications &lt; Networks/graphs</td>
</tr>
</tbody>
</table>

**Abstract:**

This paper studies information design in social networks. We consider a setting, where agents’ actions exhibit positive local network externalities. There is uncertainty about the underlying state of the world, which impacts agents’ payoffs. The platform can choose a signaling mechanism that sends informative signals to agents upon realization of this uncertainty, thereby influencing their actions. We investigate how the platform should design its signaling mechanism to achieve a desired outcome. Although this abstract setting has many applications, we discuss our results in the context of a specific one: misinformation in social networks. Agents in a social network engage with content that is possibly erroneous. Their payoff is based on the direct satisfaction from engaging, the disutility from engaging with erroneous content, and the positive externality that they derive from engaging with the same content as their peers in the underlying social network. The platform can commit to a signaling mechanism that sends agents informative signals based on the realization of the error of the content, and influence agents’ engagement decisions.

We show that the optimal (in terms of engagement/misinformation) signaling mechanisms have a simple threshold structure: the platform recommends that agents “Engage” with the content if its error is below a threshold and recommends “Do not engage” otherwise. For the mechanism that maximizes engagement, these thresholds depend on agents’ network positions, which we capture through a novel centrality measure that we introduce. Surprisingly, in the case where the platform seeks only to minimize misinformation (regardless of the induced engagement), public signal mechanisms with identical thresholds across agents are optimal. This is in contrast to the engagement maximization setting, where when agents are heterogeneous in terms of their network positions, public signal mechanisms induce substantially lower engagement than the optimal mechanisms. We also study the frontier of the engagement/misinformation levels that can be achieved via different mechanisms and characterize when public signal mechanisms can achieve optimal tradeoffs. Finally, we
| supplement our theoretical findings with numerical simulations on a Facebook subgraph. |
Optimal Signaling of Content Accuracy: Engagement vs. Misinformation*

Ozan Candogan
University of Chicago, Booth School of Business. E-mail: ozan.candogan@chicagobooth.edu

Kimon Drakopoulos
Marshall School of Business, Data Sciences and Operations. E-mail: drakopou@marshall.usc.edu

This paper studies information design in social networks. We consider a setting, where agents’ actions exhibit positive local network externalities. There is uncertainty about the underlying state of the world, which impacts agents’ payoffs. The platform can choose a signaling mechanism that sends informative signals to agents upon realization of this uncertainty, thereby influencing their actions. We investigate how the platform should design its signaling mechanism to achieve a desired outcome. Although this abstract setting has many applications, we discuss our results in the context of a specific one: misinformation in social networks. Agents in a social network engage with content that is possibly erroneous. Their payoff is based on the direct satisfaction from engaging, the disutility from engaging with erroneous content, and the positive externality that they derive from engaging with the same content as their peers in the underlying social network. The platform can commit to a signaling mechanism that sends agents informative signals based on the realization of the error of the content, and influence agents’ engagement decisions.

We show that the optimal (in terms of engagement/misinformation) signaling mechanisms have a simple threshold structure: the platform recommends that agents “Engage” with the content if its error is below a threshold and recommends “Do not engage” otherwise. For the mechanism that maximizes engagement, these thresholds depend on agents’ network positions, which we capture through a novel centrality measure that we introduce. Surprisingly, in the case where the platform seeks only to minimize misinformation (regardless of the induced engagement), public signal mechanisms with identical thresholds across agents are optimal. This is in contrast to the engagement maximization setting, where when agents are heterogeneous in terms of their network positions, public signal mechanisms induce substantially lower engagement than the optimal mechanisms. We also study the frontier of the engagement/misinformation levels that can be achieved via different mechanisms and characterize when public signal mechanisms can achieve optimal tradeoffs. Finally, we supplement our theoretical findings with numerical simulations on a Facebook subgraph.

Key words: Social Networks, Information Design, Misinformation, Fake News, Online Platforms

* We would like to thank Kostas Bimpikis, John Birge, Rene Caldentey, Odilon Camara, Matthew Jackson, Asuman Ozdaglar, Yiagios Papanastasiou, Raman Randhawa, Daniela Saban, Nicolas Stier-Moses for their constructive feedback. We also are grateful to the seminar participants at Stanford University, Facebook, and Informs Annual Meeting 2017 for their comments.
1. Introduction

In social networks, individuals’ actions influence the actions of their peers. The prevalence of online social networks provides us with a detailed understanding of the interactions among individuals, and hence the aforementioned influence structure. Motivated by the availability of social network information, recent literature has focused on understanding how an agent’s position in the underlying social network shapes her actions (see, e.g., Ballester et al. (2006)). Intuitively, if network positions of agents impact their decisions, a platform can use the knowledge of the underlying social network to induce a desired behavior. Leveraging this intuition, a growing literature in Operations has shed light on how a seller can target agents according to their network position with appropriate discounts/markups (e.g., Candogan et al. (2012)), with advertising (e.g., Bimpikis et al. (2016)), or with seeding strategies (e.g., Kempe et al. (2003); Zhou and Chen (2016)) to influence agents’ purchase decisions and improve the seller’s profits.

In this paper, we focus on a different lever that a platform can use to induce desired behavior in social networks: information. In particular, we focus on a setting where agents’ actions exhibit **positive local network externalities**: agents derive higher utility from taking certain actions, if their peers in the underlying social network also take the same actions. There is an unknown state of the world that impacts agents’ payoffs. A priori both the agents and the platform are uninformed about the true state. The platform can commit to a **signaling mechanism** that upon realization of the uncertainty sends signals to agents about the underlying state. Appropriately chosen signaling mechanisms can alter agents’ expected payoffs and impact the outcome of their interaction. This paper asks the following fundamental question: How should the platform choose its signaling mechanism and design information so as to induce a desired outcome?

The abstract setting described above has many applications, ranging from product recommendations to signaling of content accuracy in online social networks. For instance, consider an online social network promoting products that exhibit positive network externalities (e.g., online games) to its users. The quality of each product is ex ante unknown, and the platform can choose a mechanism that upon realization of the quality sends informative signals to agents about the product. By doing so, the platform can influence agents’ purchase decisions, and in turn, e.g., increase the total sales, the average quality of the product sold, or a weighted combination of the two – depending on which objective is of higher importance.

Another application of our framework is to the problem of misinformation in online social networks. These networks have an immense reach and significant impact on individuals’ beliefs and opinions: 62% of American adults are informed about the news through online social media (Gotfrieand Shearer (2016)). However, this has led to the intensification of a troubling phenomenon:
the spread of false information. This phenomenon takes many forms, ranging from deliberate disinformation campaigns to inadvertent mistakes in reporting and associated accumulation of misinformation. To deal with this phenomenon, social networking platforms have explored different interventions, ranging from disrupting the financial incentives of spammers to labeling or censoring articles that have been identified as containing false information [Mosseri 2016].

In this work, we consider a setting where content (articles, videos, etc.) with ex ante unknown accuracy becomes available on a platform. The platform can choose a signaling mechanism that sends informative signals to agents about the accuracy of the content once it is realized[1]. An example of such an intervention is labeling some content as containing misinformation[2]. Given these signals, agents decide on whether to engage with (i.e., read, comment on, discuss) the content by taking into account the following factors: the satisfaction from engaging, the disutility from engaging with erroneous content, and the network externalities. The last refers to the additional utility that an agent derives from engaging with the same content as her peers in the underlying social network.

The engagement decisions depend on the received signals and, therefore, the platform can design the signaling mechanism to influence the agents and consequently achieve desired performance metrics. We focus on two metrics for analyzing the equilibrium outcome for a given signaling mechanism: (i) the expected error of the content that agents engage with and (ii) the expected number of agents who engage, which quantify (and are henceforth referred to as) misinformation and engagement, respectively. We refer to the tuple of misinformation and engagement associated with a mechanism as the mechanism’s performance level. We study optimal mechanisms that maximize engagement or minimize misinformation, as well as the frontier of achievable performance levels.

We would like to emphasize that mathematically, this setting is equivalent to the one where the platform signals the quality of a product to agents so as to influence their purchase decisions. In the remainder of the paper, we focus on the misinformation application, and discuss our results in this context.

1.1. Main Results

In general, the signaling mechanisms of the platform are mappings from the set of possible error levels (associated with the content) to a message (or signal) in a message space. The space of possible

[1] The information on the accuracy of the content may be available through third-party fact-checking services or various algorithmic tools.

[2] Incidentally, flagging some content as disputed is one of the main interventions that online social networks employed to address hoaxes and fake news in the wake of the 2016 U.S. elections; see [Mosseri 2016].
mechanisms is large, as it encompasses all possible message spaces and all possible such mappings. Despite this generality, our first contribution establishes that optimal mechanisms (that maximize engagement, minimize misinformation, or more generally optimize a weighted combination of engagement and misinformation) exhibit natural properties and have a very simple structure (see Section 3). In particular, optimal mechanisms are straightforward. That is, their message spaces consist of possible actions of agents and, effectively, given the error, the platform recommends an action to each agent: “Engage” or “Do not engage.” The optimal mechanisms are straightforward in the sense that given a recommendation, agents always find it optimal to follow that recommendation. Moreover, optimal mechanisms are threshold mechanisms: for each agent, the platform commits to a threshold for the error. If the error realization is below the threshold, the platform recommends that the agent engage, and otherwise it recommends that she not engage.

These two properties allow for tractable design and analysis of the optimal mechanisms, as they reduce the platform’s decision problem to choosing thresholds for each agent. If an agent’s threshold is high, she engages with more content but is also exposed to more misinformation in expectation. A key property that is particularly noteworthy is that in general, in an optimal mechanism, agents may have different thresholds and hence receive different signals. In other words, signals are private and depend on agents’ positions in the network. We refer to such straightforward threshold mechanisms as private signal mechanisms.

As our second contribution, we introduce two classes of benchmark mechanisms that have a simpler structure than private signal mechanisms, and we characterize the associated performance levels. The first benchmark that we analyze is public signal mechanisms where the platform commits to the same threshold for all agents. Consequently, these mechanisms always send the same signal to all agents: if the error is below the public threshold, the platform recommends that the agents “Engage,” and otherwise the recommendation is “Do not engage.” The second benchmark is a special case of public signal mechanisms, which we refer to as no-intervention mechanisms. In these mechanisms, the platform’s signals are not informative (e.g., signals that always recommend that all agents engage). We show that the performance levels associated with public signal mechanisms (and hence no-intervention mechanisms) critically depend on the structure of the underlying network and admit a geometric characterization in terms of the $k$-cores of the underlying network.

Our third contribution is a characterization of mechanisms that maximize engagement and minimize misinformation (Section 4). As discussed above, optimal mechanisms always take the form of a private signal mechanism. However, it is not clear a priori how optimal thresholds relate to agents’ network positions and if/when simpler benchmark mechanisms suffice to achieve (approximate) optimality. We shed light on this question and establish the surprising fact that while engagement-maximizing mechanisms assign higher thresholds (and hence higher error levels in expectation) to
more central agents, the optimal misinformation-minimizing mechanisms is a public signal mechanism that assigns the same threshold to all agents regardless of their positions in the network. Interestingly, our results indicate that no intervention leads to both higher levels of misinformation and lower levels of engagement, and therefore that appropriate signaling mechanisms could allow the platform to improve its operations on both fronts.

Our last contribution is to show that the structural properties of the aforementioned optimal mechanisms extend to other mechanisms that capture the best possible tradeoff between engagement and misinformation, i.e., that achieve performance on the Engagement/Misinformation frontier of performance levels (Section 5). In particular, we show that for small engagement and misinformation levels, public signal mechanisms achieve performance on the frontier while for higher engagement and misinformation levels private signal mechanisms are necessary for optimality.

1.2. Related Literature

Our work touches upon three different strands of literature, as we explain next.

**Bayesian Persuasion and Information Design:** Seminal papers [Brocas and Carrillo (2007), Rayo and Segal (2010), and Kamenica and Gentzkow (2011)], study the problem of a “sender” trying to persuade a “receiver” to take an action that would not be incentive compatible for the latter to follow if she were perfectly informed. Ex ante the “sender” and “receiver” are both uninformed. The “sender” commits to an informative signal about the state of the world. The “receiver” observes a realization of this signal and takes an action. These papers consider situations where there is a single “sender” and a single “receiver” as well as multiple “senders” and a single “receiver”, as examined by Gentzkow and Kamenica (2017), and provide conditions under which the “sender” can benefit from choosing appropriate signaling mechanisms. Furthermore, Alonso and Camara (2016) consider the case of one “sender” and many “receivers” in a voting scenario. In our work, we build on this framework and consider one “sender” (platform) and many “receivers” (users) who are connected through social ties, as represented by a network. To the best of our knowledge, our paper is the first to study information design for a network of agents. Furthermore, while straightforward mechanisms turn out to be optimal in other settings as well (e.g., Kamenica and Gentzkow (2011)), we provide an even stronger refinement of optimal mechanisms in our context: that of a straightforward threshold mechanism.

There is an emerging and growing literature on the operations of online platforms and information design. This literature has so far focused on optimal design of contests (e.g., Bimpikis et al. (2015)), design of crowdfunding campaigns (e.g., Alaei et al. (2016)), and inspection and information disclosure in online platforms (e.g., Papanastasiou et al. (2017)), among others. Our paper
adds to this literature by studying how an online social networking platform can design signals provided to agents so as to achieve various tradeoffs between engagement and misinformation.

Finally, the value of releasing information to agents has been explored in various contexts within the field of operations management: Guo and Zipkin (2007), Jouini et al. (2011), Allon et al. (2011), and Lingenbrink and Iyer (2017) explore the different effects of information revelation in the context of queuing, while Bergemann et al. (2015) study mechanisms for revealing information to a firm facing pricing decisions. Zhou and Chen (2016) explore how a product of unknown quality should be seeded in a social network so as to increase the influence of initial adopters on the rest, and in turn to increase the total consumption of the product. This setting can be interpreted as revealing the full information about a product to a subset of agents, in the hope of influencing the decisions of the rest.

Misinformation: Questions related to misinformation have been explored in several fields. Acemoglu et al. (2010) have studied how the presence of agents who influence others but are not influenced by them impacts information aggregation and the spread of misinformation in societies. The problem of controlling contagious misinformation has been studied by Budak et al. (2011) and Nguyen et al. (2012). Methods for identifying erroneous content have been proposed by Kumar et al. (2016) and Tambuscio et al. (2015). The impact of misinformation on political outcomes has been studied by Allcott and Gentzkow (2017). Papanastasiou (2017) considers a dynamical model of news consumption and studies the operations of a platform that tries to detect fake news while observing the propagation of each post on the platform. In contrast to the prior work, our work studies the control of misinformation after it has been detected. Specifically, we assume that the platform chooses a signaling mechanism that upon realization of errors sends signals to agents to persuade them to engage with it or not (depending on the performance metric).

Network games with externalities: More broadly, in our model the interaction between agents is characterized by positive externalities; i.e., agents derive higher utility from engaging with content if their friends also engage with it. Network games with positive network externalities have been studied in the context of consumption models by Katz and Shapiro (1985); Ballester et al. (2006); Candogan et al. (2012), and coordination in global games by Morris and Shin (2002); Angeletos et al. (2006a,b). Our paper builds on such a model but adds a second layer of information design on top of it. The role of this layer is twofold: coordination around a desired outcome and persuasion of agents about the quality of each post.

2. Model

We consider a set of $n$ agents $V = \{1, \ldots, n\}$ interacting through an online social network, which is represented by a graph $G = (V, E)$. We assume that the underlying graph is undirected and
unweighted. The edges of $G$ represent “friendship” relationships in the social network. We use $N_i$ to denote the set of friends of node $i$, i.e., $N_i = \{ j \in V : (i,j) \in E \}$. We denote the degree of node $i \in V$ by $d_i = |N_i|$ and the minimum and maximum degrees in the network by $d_{\text{min}} = \min_{i \in V} d_i$ and $d_{\text{max}} = \max_{i \in V} d_i$, respectively. We denote by $G$ the adjacency matrix of $G$. The $(i,j)$th entry of this matrix, denoted by $g_{ij}$, belongs to $\{0,1\}$ and encodes whether agents $i$ and $j$ are connected in the underlying network.

In this paper, we study agents’ engagement decisions with some content (e.g., a news article, opinion piece, blog post, etc.) that is available on the social network (or platform, as we refer to it in the rest of the paper). The content on the platform is possibly erroneous. We denote the associated error (or inaccuracy) by $\epsilon$. The error $\epsilon$ is uniformly distributed in $[0,\alpha]$. In the special case where $\alpha = 1$, $\epsilon$ can be interpreted as the probability that the content is fake.

Each agent $i$ on the platform chooses a binary action $x_i \in \{0,1\}$: engage ($x_i = 1$) or not engage ($x_i = 0$) with the content. Motivated by the social aspects of engaging with content in online social networks, we focus on a model where the engagement decision of an agent impacts the payoffs of her friends who also engage with the content. Specifically, if the error of the content is $\epsilon$, the agent’s realized payoff is given by

$$\tilde{U}_i(x_i,x_{-i}) = 1(x_i = 1) \cdot \left( v - b\epsilon + \sum_j g_{ij} x_j \right). \quad (1)$$

Thus, the agent $i$ has zero payoff if she decides not to engage with the content. Otherwise, her payoff consists of three terms: (i) a constant payoff term $v$ corresponding to the satisfaction that she (unilaterally) derives from engaging with the content, (ii) a negative term that captures the disutility that the agent incurs due to engaging with erroneous content, and (iii) a network externality term that captures the additional utility that she derives from engaging with content with which her friends also engage. We assume that agents do not a priori observe the error level $\epsilon$ of the available content or the actions of the other agents; hence, they cannot directly maximize their payoffs in (1), and instead, they choose actions so as to maximize their expected payoffs.

Note that in this paper, we do not explicitly model heterogeneity across agents in terms of biases, or taste for certain types of content (e.g., some agents inherently being more favorably disposed towards content promoting right/left-leaning policies). Such heterogeneity among agents could in principle affect their engagement behavior and hence could also be taken into account. However, in order to single out the impact of the network structure on agents’ engagement behavior, we omit other sources of heterogeneity from our model and defer them to future extensions of this work.

The platform can use technology, fact-checking, and other resources to accurately calculate the error $\epsilon$ of the content. Using the knowledge of the error, the platform can send a signal $s_i$ to
each agent $i \in V$ in order to influence agents’ engagement decisions. In general, the signal that the platform sends can be random (even conditioned on the error of the content). We denote the underlying probability space by $(\Omega, \Sigma, \mathbb{P})$, where $\Omega$ denotes the set of possible outcomes, $\Sigma$ denotes the $\sigma$-algebra, and $\mathbb{P}$ denotes the probability measure. Using this notation, the signaling mechanism of the platform can formally be defined as follows.

**Definition 1.** A **signaling mechanism** is a collection of joint probability density functions $F(\cdot; \epsilon)$ of the random variables $(s_1, \ldots, s_n)$ for $\epsilon \in [0, \alpha]$. We denote the range of possible signals for agent $i$ by $S_i$, i.e., $s_i : \Omega \rightarrow S_i$.\footnote{If $\{S_i\}$ are discrete, $F(\cdot; \epsilon)$ denotes a probability mass function. We also allow for signal spaces that have continuous components as well as atoms.}

In other words, for each agent $i \in V$ and error level $\epsilon \in [0, \alpha]$, the signaling mechanism specifies a probability distribution from which the signal $s_i$ of each agent $i \in V$ is drawn. We use the shorthand notation $F$ to denote the signaling mechanism of the platform and focus on mechanisms for which $s_i^{-1}(z) \in \Sigma$ for all $z \in S_i$ and $i \in V$. We denote the set of all such ($\Sigma$-measurable) mechanisms by $\mathcal{F}$.\footnote{In what follows we will study expected number of agents who engage, and the induced misinformation for a given mechanism. For these quantities to be well-defined mild technical assumptions such as measurability are needed. Such technical assumptions are stated and discussed in Appendix A.1.} Note that the signals that agents receive may be correlated with each other as well as the error. However, an agent does not observe the signals received by the remaining agents.

Given a signaling mechanism $F \in \mathcal{F}$, each agent $i$ chooses a strategy $\mu_i : S_i \rightarrow \{0, 1\}$. We use $\mu = (\mu_1, \ldots, \mu_n)$ to denote the strategy profile (i.e., the collection of strategies of all agents). That is, conditional on the received signal, each agent on the platform chooses a binary action $x_i = \mu_i(s_i)$. More concretely, suppose agent $i$ receives signal $s_i = s$; then, conditional on her signal, her expected payoff is given as follows:

$$U_i(\mu_i, \mu_{-i}, s) := 1(\mu_i(s) = 1) \cdot \mathbb{E} \left[ v - b \epsilon + \sum_j g_{ij} \mu_j(s_j) | s_i = s \right]. \quad (2)$$

Here, the expectation is taken over the error $\epsilon$ as well as the signals $s_{-i}$ received by agents other than $i$. Note that payoff realization of $i$ depends on the signal $s_j$ of another agent $j \neq i$, as $s_j$ impacts the action of agent $j$, i.e., $x_j = \mu_j(s_j)$.

Under the platform’s signaling mechanism $F$, agents face an incomplete information game, and the solution concept that we use to analyze this game is that of a Bayesian Nash Equilibrium.\footnote{More generally, one could consider randomized strategies where $\mu_i : S_i \rightarrow \Delta(\{0, 1\})$, where $\Delta(\cdot)$ denotes probability distributions over its argument. For ease of exposition, we restrict attention to pure strategies.}
DEFINITION 2. Given a signaling mechanism $F$, a strategy profile $\mu$ is a Bayesian Nash Equilibrium if, for all agents $i \in V$, we almost surely have:

$$
\mu_i(s_i) \in \arg\max_{x \in \{0,1\}} \mathbb{1}(x = 1) \cdot \mathbb{E} \left[ v - b \epsilon + \sum_j g_{ij}(s_j) | s_i \right],
$$

We denote by $Q(F)$ the set of Bayesian Nash equilibria of the game under signaling mechanism $F$.

When designing the signaling mechanism $F$, the platform essentially determines the amount of information that each agent receives regarding the underlying error $\epsilon$. For example, the platform could reveal all the available information by choosing $S_i = [0, \alpha]$ and setting $s_i = \epsilon$. Depending on the performance metric that the platform seeks to optimize, different signaling mechanisms could be optimal. In this work, we focus on two important performance metrics that interact with each other in nontrivial ways: engagement and misinformation. For a particular mechanism $F \in \mathcal{F}$ and an induced equilibrium $Q \in Q(F)$, the engagement (denoted by $E(Q)$) captures the expected number of agents who engage with the content on the platform, i.e.,

$$
E(Q) \triangleq \mathbb{E} \left[ \sum_{i \in V} x_i \right].
$$

In contrast, misinformation (denoted by $M(Q)$) is related to the quality of the content on the platform, and is given by the expected total error of the content that agents engage with,

$$
M(Q) \triangleq \mathbb{E} \left[ \sum_{i \in V} \epsilon x_i \right],
$$

where, in both cases, the expectation is taken with respect to the error $\epsilon$ and the signals $(s_1, \ldots, s_n)$. We refer to $P(Q) \triangleq (E(Q), M(Q))$ as the performance associated with equilibrium $Q$. Similarly, we denote the performance levels achievable by a mechanism $F$ as follows:

$$
\mathcal{P}(F) \triangleq \{P(Q) | Q \in Q(F)\} = \{(E(Q), M(Q)) | Q \in Q(F)\}.
$$

6 Alternatively Bayesian Nash Equilibrium could be defined by requiring that $\mu_i \in \arg\max_{\mu'_i} \mathbb{E}[U_i(\mu'_i, \mu_{-i}, s)]$ for all $i$, where the expectation is taken over all signals and error realizations, and $\mu'_i : S_i \rightarrow \{0,1\}$. It can be shown that when the signal space is finite (and all signals in $S_i$ are realized with nonzero probability), this implies that (3) surely holds. When the signal space is infinite, it suffices to impose this condition almost surely, see, e.g., [Mas-Colell et al. (1995)]. That is, $\mathbb{P}(\text{3} \text{ holds}) = 1$.

7 It is also worth noting that in the case of infinite signal spaces, the conditional expectation is still well defined in our context since from the definition of signaling mechanisms, it follows that signals and error admit a joint distribution.

8 Note that it is also necessary to take the expectation over equilibrium probabilities of choosing different actions if randomized strategies are allowed.
Observe that there is a clear tradeoff between these two metrics. In the extreme case where none of the agents engages with the content, misinformation is minimal, but engagement is zero. At the other extreme, where all agents always engage and the engagement is maximized, misinformation is also maximized. In order to study this tradeoff and design optimal signaling mechanisms, we focus on the natural cases where a platform would like to minimize misinformation and/or maximize engagement.

Specifically, we consider the problem of the platform that chooses its signaling mechanism to maximize a weighted combination of these two metrics:

\[
\max_{F \in F} W(F),
\]

where

\[
W(F) = \max_{Q \in \mathcal{Q}(F)} \gamma_e E(Q) - \gamma_m M(Q),
\]

and \(\gamma_e, \gamma_m \geq 0\) are fixed exogenous constants.

We use the Sender-Preferred Subgame Perfect Equilibrium to analyze the interaction between the platform and the agents, as defined below.

**Definition 3.** A **Sender-Preferred Subgame Perfect Equilibrium** consists of a signaling mechanism \(F\), a strategy profile \(\mu\), and beliefs \(\nu\) denoting agents’ posterior beliefs about \(\epsilon\) after respectively observing signals \(s\), such that

(i) Beliefs are derived using the Bayes rule, whenever possible,

(ii) Strategies \(\mu\) are optimal given the beliefs \(\nu\), \(i \in V\), i.e., \(3\) almost surely holds,\(^9\)

(iii) Mechanism \(F\) is optimal, i.e., \(F \in \arg \max_{F' \in F} W(F')\),

(iv) \(\mu \in \arg \max_{Q \in \mathcal{Q}(F)} \gamma_e E(Q) - \gamma_m M(Q)\).

The notion of Sender-Preferred Equilibrium, is commonly used in the Bayesian Persuasion literature (see Kamenica and Gentzkow (2011)), and essentially assumes that out of all the equilibria that may potentially arise given a signaling mechanism, the sender (the platform in our case) will choose the most desirable equilibrium with respect to her own utility (cf. condition (iv) in the definition above).

Note that in the Bayesian Persuasion literature in general, it is assumed that the sender (platform) designs and commits to a signaling mechanism prior to realization of the underlying uncertainty (i.e., error in our setting). Subsequently, based on this realization of the uncertainty, the

---

\(^9\)Kamenica and Gentzkow (2011) mainly focus on settings where the message space has finite cardinality and require the optimality of \(\mu(z)\) for all signals \(z \in S_i\). Since we allow for more general message spaces, we impose this condition almost surely.
sender sends signals to the receiver(s) (or agents in our setting) consistently with this mechanism. Given these signals, agents update their posterior belief about the underlying uncertainty and choose their actions. The platform’s problem is to choose the signaling mechanism so as to ensure a desired outcome as a result of agents’ actions.

Remark: As discussed earlier, the model of this section has applications other than the engagement-misinformation problem presented here. For instance, consider a setting where agents decide whether to consume products that exhibit local network externalities (similar to Candogan et al. (2012)). Suppose that the qualities of products are ex ante unknown, and the platform commits to a signaling mechanism that sends informative signals to agents once quality is realized. Choosing this mechanism appropriately, the platform may want to maximize total sales, or the average quality of the product sold (or any affine combination of the two). Denoting by $\epsilon, x, E(Q), M(Q)$ respectively the quality of the product, the consumption decision of agent $i$, the total sales induced at equilibrium $Q$, and the negative of the average quality of the product sold at $Q$, the problem of the platform can be precisely modeled as in this section.

3. Signaling Mechanisms

In this section, we first establish that optimal signaling mechanisms always admit a simple threshold structure and that agents always have the incentive to follow the signals provided by the optimal mechanism. Then, for comparison purposes, we focus on simple but natural threshold mechanisms where all agents receive the same signal and we discuss the structure of the equilibria that arise in this case.

3.1. Optimal Signaling Mechanisms

In general, the space of possible signaling mechanisms is (uncountably) infinite, and hence, the problem of the platform given in (4) is potentially very challenging. Despite this fully general structure, through a sequence of lemmata, we establish that it suffices to focus our search for optimal signaling mechanisms on a subclass that we refer to as the straightforward threshold mechanisms.

We start by formally defining straightforward mechanisms.

**Definition 4.** Consider a mechanism $F \in \mathcal{F}$ such that the signal space is $S_i = \{0, 1\}$ for all $i \in V$. We say that an associated equilibrium $Q \in \mathcal{Q}(F)$ is a **straightforward equilibrium** if the signal that agent $i$ receives ($s_i$) is equal to her equilibrium action, i.e., $\mu_i(s_i) = s_i$ for all $s_i \in S_i$, $i \in V$. A signaling mechanism $F$ with signal space $S_i = \{0, 1\}$ is referred to as a **straightforward mechanism** if it admits a straightforward equilibrium, i.e., if some $Q \in \mathcal{Q}(F)$ is straightforward.
Our first lemma establishes that in (1), without loss of optimality, we can restrict our attention to the class of straightforward mechanisms. Note that the result is actually stronger, as it establishes that the performance of any mechanism (as opposed to only the optimal ones in (1)) can be matched using a straightforward mechanism.

**Lemma 1.** Consider a signaling mechanism that achieves engagement equal to $L$ and total misinformation equal to $M$. Then, there exists a straightforward mechanism that achieves total engagement equal to $L$ and total misinformation equal to $M$ at its straightforward equilibrium.

We establish this result by showing that for any given mechanism, the platform can group all of the signals that lead to engagement by each agent $i$ and just relabel them as a recommendation to “Engage” ($s_i = 1$). Similarly, the platform can group all of the signals that lead to non-engagement and relabel them as the recommendation “Do not engage” ($s_i = 0$). The two mechanisms are effectively the same, and they lead to the same outcome for the same error realization. Moreover, the latter one is straightforward, implying that a straightforward mechanism achieves the same performance as the former.

Next, we provide a further refinement of the set of optimal mechanisms by establishing that it suffices to restrict attention to straightforward mechanisms that admit a threshold structure.

**Definition 5.** A signaling mechanism $F$ is a **threshold mechanism** with thresholds $\{\sigma_i\}_{i \in V}$ if we have $\sigma_i \in [0, \alpha]$, $S_i = \{0, 1\}$ for all $i \in V$, and

$$s_i(\omega) = 1 \iff \epsilon(\omega) \leq \sigma \quad \text{for all } \omega \in \Omega \text{ and } i \in V.$$ 

In words, a threshold mechanism is a signaling mechanism where agent $i$ receives the signal $s_i = 1$ if and only if the error $\epsilon$ is less than or equal to an agent-specific threshold $\sigma_i$. Note that this can simply be interpreted as a recommendation mechanism for the platform: if the error is less than $\sigma_i$, the platform recommends engagement, and otherwise it recommends agent $i$ not to engage.

Intuitively, a straightforward mechanism that is not a threshold mechanism can be improved by recommending “Engage” for error realizations smaller than a threshold and “Do not engage” for larger error realizations, while choosing the threshold so that the (total) probability of recommending each action remains unchanged. This mechanism is still straightforward, since the expected error is smaller than that of the initial straightforward mechanism under an “Engage” recommendation and larger under a “Do not engage” recommendation. Consequently, under the new mechanism, the total engagement remains the same, and the induced misinformation is smaller (since agents engage with content for lower error realizations in expectation). Our next result formalizes this construction and establishes that an optimal signaling mechanism can be obtained by restricting attention to straightforward threshold mechanisms.
Theorem 1. Given any signaling mechanism that achieves engagement equal to $L$ and total misinformation equal to $M$, there exists a straightforward threshold mechanism that achieves total engagement equal to $L$ and total misinformation weakly lower than $M$ at its straightforward equilibrium.

Consequently, if an optimal solution to (4) exists, then there exists a straightforward threshold mechanism whose straightforward equilibrium solves (4).

We conclude this section by deriving a set of conditions for thresholds that ensure straightforwardness of threshold mechanisms. In particular, we consider a threshold mechanism with thresholds $\{\sigma_i\}_{i \in V}$ and suppose that all agents $j \neq i$ follow the signals of the platform, i.e., $x_j = s_j = 1(\epsilon \leq \sigma_j)$. We explore what conditions on $\sigma_i$ guarantee that agent $i$ also finds it optimal to follow the signal of the platform.

Assume that $\sigma_i > 0$ and consider the recommendation to “Engage” ($s_i = 1$). The expected utility of agent $i$ from engaging consists of three terms: the satisfaction term $v$; the error disutility term, which is equal to $-bE[\epsilon | s_i = 1] = -bE[\epsilon | \epsilon \leq \sigma_i] = -b\frac{\sigma_i}{2}$; and the externality term, which is equal to

$$\sum_{j \in V} g_{ij} P(x_j = 1 | s_i = 1) = \sum_{j \in V} g_{ij} P(s_j = 1 | s_i = 1) = \sum_{j \in V} g_{ij} \frac{\min\{\sigma_i, \sigma_j\}}{\sigma_i},$$

where we make use of the fact that agents $j \neq i$ follow the signal of the platform. Therefore, it is incentive compatible for agent $i$ to follow the “Engage” recommendation when $v - \frac{\sigma_i^2}{2} + \sum_{j \in V} g_{ij} \frac{\min\{\sigma_i, \sigma_j\}}{\sigma_i} \geq 0$, or equivalently:

$$\sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}. \quad (5)$$

If $\sigma_i = 0$, then almost surely agent $i$ receives the signal $s_i = 0$. Hence, for equilibrium characterization, her action under the “Engage” recommendation ($s_i = 1$) is irrelevant. In addition, for this case, (5) trivially holds.

Similarly, assume that $\sigma_i < \alpha$, and consider the recommendation “Do not engage” ($s_i = 0$). Once again, agent $i$’s utility from engaging consists of three terms: the satisfaction term $v$; the penalty term, which is now equal to

$$-bE[\epsilon | s_i = 0] = -bE[\epsilon | \epsilon > \sigma_i] = -b\frac{\alpha + \sigma_i}{2};$$
and the externality term, which is equal to
\[
\sum_{j \in V} g_{ij} \mathbb{P}(x_j = 1 \mid s_i = 0) = \sum_{j \in V} g_{ij} \mathbb{P}(s_j = 1 \mid s_i = 0) = \sum_{j \in V} g_{ij} \frac{\max\{\sigma_i, \sigma_j\} - \sigma_i}{\alpha - \sigma_i}.
\]

Therefore, it is incentive compatible for agent \(i\) to follow the “Do not engage” recommendation when
\[
\alpha^2 - \sigma_i^2 \geq \frac{2}{b} v(\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b} d_i \sigma_i. \quad (6)
\]

If, on the other hand, \(\sigma_i = \alpha\), then agent \(i\) surely receives the signal \(s_i = 1\). Hence, in this case, the agent’s action under the engage recommendation is not relevant for equilibrium characterization. Moreover, in this case (6) trivially holds.

Assume that a threshold mechanism where thresholds satisfy (5) and (6) is given. The above discussion establishes that this mechanism is straightforward. Conversely, if these inequalities are violated, then at least one agent finds it optimal to not always follow the recommendation. Thus, we have proved the following characterization of threshold mechanisms that are straightforward:

**Proposition 1.** Let \(F\) be a threshold mechanism with thresholds \(\{\sigma_i\}_{i \in V}\). \(F\) is straightforward if and only if (5) and (6) hold for all \(i\).

In summary, using the results of this section, the set of optimal mechanisms has been simplified significantly: there exists an optimal mechanism that is a straightforward threshold mechanism. Moreover, using Proposition 1 the problem can be further simplified to searching over thresholds that satisfy simple inequalities. We should note here that for the optimal mechanism, the thresholds \(\sigma_i\) may be different for each agent, which may lead to difficulties in terms of implementation. With this observation in mind and for comparison purposes, in the subsequent sections, we introduce and analyze two classes of threshold mechanisms: (i) the public signal mechanisms, where the threshold \(\sigma_i\) is the same across all agents, and hence, the recommendation to engage is the same for all agents, and (ii) the no-intervention mechanisms where the signal from the platform bears no information on the true error of the content.

**Remark 1:** For simplicity of exposition, we restricted attention to pure equilibria, i.e., \(\mu_i : S_i \rightarrow \{0, 1\}\). An inspection of the proofs of this section reveals that this assumption is not used in our derivations and hence is not needed for the results of this section.

**Remark 2:** In this paper, we focus on a natural objective for the platform, that is to maximize a weighted sum of engagement and misinformation. Alternatively, one can consider the case of a platform that maximizes social welfare (sum of agents’ utilities). The characterization of optimal mechanisms given this section is still valid in this case.
Remark 3: Conceivably, the platform has the option to censor erroneous content (content with large $\epsilon$). In this case, as opposed to sending a “Do not engage” signal to agents, the platform simply eliminates any opportunity to engage. Mathematically, this is equivalent to removing the constraint (10) from our optimization formulations.

3.2. The No-Intervention and Public Signal Mechanisms

In this section, we discuss public signal mechanisms, where all agents have the same thresholds, i.e., $\sigma_i = \sigma$ for all $i \in V$. Note that in these mechanisms, all agents receive the same signal $s_i = s \in \{0, 1\}$ for all $i \in V$ and $\epsilon \in [0, \alpha]$. Thus, effectively, there is a public signal that coordinates the actions of all agents.

We emphasize that if $\sigma = \alpha$ (similarly $\sigma = 0$), then platform’s signal is uninformative. That is, the platform either recommends always engaging or (almost surely) never does so. In these settings, it can be seen that the equilibrium outcome will (almost surely) be identical to the one in the absence of a platform sending signals. Thus, we refer to the associated equilibria as the no-intervention equilibria. In subsequent sections, we analyze this special case of public signal mechanisms as a benchmark to understand how the intervention of the platform alters the engagement outcome.

In the remainder of this section, we first establish that agents’ equilibrium strategies under public signal mechanisms can be obtained through the equilibria of auxiliary supermodular games. Then, we leverage the connection to supermodular games to characterize the equilibrium performance levels for these mechanisms.

Realized payoffs of agents, which are mappings from actions $\{x_i\}_i$ to reals, were introduced in (1). When studying public signal mechanisms (with $\sigma \in (0, \alpha)$), it is convenient to introduce the following functions, which capture the expected payoffs from different actions conditional on the public signal of the platform:

$$u_i(x_i, x_{-i}, s) \triangleq \begin{cases} 
1(x_i = 1) \cdot \mathbb{E} \left[ v - b \sigma + \sum_j g_{ij} x_j | \epsilon \leq \sigma \right] & \text{for } s = 1, \\
1(x_i = 1) \cdot \mathbb{E} \left[ v - b \sigma + \sum_j g_{ij} x_j | \epsilon > \sigma \right] & \text{for } s = 0.
\end{cases}$$

(7)

Note that these functions are still mappings from actions of agents to reals, and are closely related to realized payoffs. In particular, we have the following: $u_i(x_i, x_{-i}, s) = \mathbb{E}[\hat{U}_i(x_i, x_{-i}) | \epsilon \leq \sigma]$ if $s = 1$, and $u_i(x_i, x_{-i}, s) = \mathbb{E}[\hat{U}_i(x_i, x_{-i}) | \epsilon > \sigma]$ if $s = 0$. Moreover, expressing actions in terms of strategies, i.e., $x_i = \mu_i(s)$, we recover the payoff functions in (2).

\[^{10}\text{Here, we restrict attention to } \sigma \in (0, \alpha) \text{ to avoid conditioning on measure zero events. If } \sigma \in \{0, \alpha\}, \text{ the signals are uninformative as the public signal mechanism either almost surely recommends engagement or no engagement. In such cases, for our analysis of performance mechanism induced under public signal mechanisms, it suffices to restrict attention to signals that have a non-zero probability of being sent by the platform.}\]
Observe that in public signal mechanisms, either all agents receive the “Engage” signal or no agent receives the “Engage” signal. Thus, depending on the signal, two perfect information games with different payoffs are induced. In particular, let $G_0$ and $G_1$ be normal-form games with a set of players $V$ and strategy sets $\{0, 1\}$ for each $i \in V$. Assume that the payoffs of agents in $G_0$ are given by $\{u_i(x_i, x_{-i}, 0)\}$ and that in $G_1$, they are given by $\{u_i(x_i, x_{-i}, 1)\}$. In our setting, if the public signal is $s = 1$, then agents play $G_1$, and otherwise they play $G_0$.

We next establish that $G_0$ and $G_1$ are supermodular games. We start with a definition of supermodular games, adapted to our setting.

**Definition 6.** Let $H = (N, \{\bar{S}_i\}, \{\bar{u}_i\})$ be a game with set of players $N$, compact set of strategies $\bar{S}_i \subseteq \mathbb{R}$ and continuous payoff functions $\bar{u}_i : x_i \in N \bar{S}_i \rightarrow \mathbb{R}$ for each $i \in N$. We say that $H$ is a supermodular game if $\bar{u}_i(x_i, x_{-i})$ has increasing differences in $(x_i, x_{-i})$ for $x_i \leq x'_i$ and $x_{-i} \leq x'_{-i}$ (where the last inequality is entrywise).

**Lemma 2.** $G_0$ and $G_1$ are supermodular games.

The supermodularity of the two games allows us to obtain an insightful geometric characterization of the performance levels achievable by public signal (and therefore no-intervention) mechanisms. In particular, our next result provides a connection between the aforementioned performance levels and the primitives of our model, namely, the agents’ utility parameters and the underlying network. Interestingly, the geometry of the achievable performance levels depends on the $k$-core of the graph, i.e., the maximal induced subgraph in which every node has a degree of at least $k$.

**Theorem 2.** Let $F$ be a public signal mechanism with threshold $\sigma$. Let $k_0$, $k_1$ respectively be the smallest integers satisfying $k_0 \geq b\frac{\alpha + \sigma}{2} - v$ and $k_1 \geq b\frac{\sigma}{2} - v$. Denote by $A_0$ and $A_1$ respectively the nodes of the $k_0$-core and the $k_1$-core of the underlying network. Let $\text{conv}(\cdot)$ denote the convex hull operator, and denote by $C(F) \overset{\Delta}{=} \text{conv}(P(F))$ the convex hull of performance levels achieved by $F$.

The set $C(F) \supseteq P(F)$ admits the following characterization:

(i) If $v \leq \frac{b\sigma}{2}$, then $C(F)$ is given by

$$C(F) = \text{conv}\left(\left\{0, 0\right\} \cup \left\{(A_1|\frac{\sigma}{\alpha}, A_1|\frac{\sigma^2}{2\alpha}), (A_0|\frac{\alpha - \sigma}{\alpha}, A_0|\frac{\alpha^2 - \sigma^2}{2\alpha}), (A_0|\frac{\alpha - \sigma}{\alpha} + A_1|\frac{\sigma}{\alpha}, A_1|\frac{\sigma^2}{2\alpha} + A_0|\frac{\alpha^2 - \sigma^2}{2\alpha})\right\}\right).$$

Moreover, if $0 < \sigma < \alpha$ and $A_0, A_1 \neq \emptyset$, then $C(F)$ is a subset of $\mathbb{R}^2$ with nontrivial area.

---

Note that by construction, the strategy sets of agents are sublattices of $\mathbb{R}$, and payoff functions are continuous. Both of these assumptions can be relaxed, e.g., it suffices to consider more general sublattices (of $\mathbb{R}^{m_i}$ for $m_i \geq 1$), and semicontinuity of payoffs in an agent’s own actions (together with continuity in other agents’ actions) suffice.
(ii) If \( \frac{b\sigma}{2} < v \leq \frac{b(\sigma + \alpha)}{2} \), then \( \mathcal{C}(F) \) is a line segment and is given by
\[
\mathcal{C}(F) = \text{conv}\left( \left\{ \left( |V| \frac{\sigma}{\alpha}, |V| \frac{\sigma^2}{2\alpha} \right), \left( |V| \frac{\sigma}{\alpha} + |A_0| \frac{\alpha - \sigma}{\alpha}, |V| \frac{\sigma^2}{2\alpha} + |A_0| \frac{\alpha^2 - \sigma^2}{2\alpha} \right) \right\} \right).
\]

(iii) If \( \frac{b(\sigma + \alpha)}{2} < v \), then \( \mathcal{C}(F) \) is a singleton given by \( \{(|V|, |V| \frac{\sigma}{2})\}\).

Note that Theorem 2 implies that in general, the equilibria induced by public signal mechanisms may or may not be unique. For example, if \( v > \alpha b \), part (iii) implies that there is a unique equilibrium where all agents engage, regardless of \( \sigma \). On the other hand, if say \( \alpha b/2 < v \leq \alpha b \), then we may have multiple equilibria. Their performance depends on the particular choice of \( \sigma \) and is given by part (ii) or (iii) of Theorem 2.

Finally, the reader can verify that depending on the value of \( \sigma \), cases (ii) and (iii) can both apply for particular values of the primitives \( \alpha, b, v \). The same is true for cases (i) and (ii). On the other hand, cases (i) and (iii) never co-exist for any values of the primitives. Therefore, in words, the set of achievable performance levels can consist of:

(i) \textit{diamonds} and \textit{lines}: cases (i) and (ii), respectively,

(ii) \textit{lines} and \textit{points}: cases (ii) and (iii), respectively.

We conclude this section by providing an example that illustrates the different cases described above.

![Figure 1](image-url)  
A network with 12 nodes.

Example 1. Consider the network \( (\mathcal{G}) \) given in Figure 1. This network consists of 12 nodes: half of the nodes are organized in a complete graph (hereafter subgraph \( \mathcal{G}_L \)), the other half constitute a cycle (hereafter \( \mathcal{G}_R \)), and two (dashed) edges connect the two halves. Denote the set of nodes of \( \mathcal{G}_L \) and \( \mathcal{G}_R \) respectively by \( V_L \) and \( V_R \).
As discussed earlier, given a network $\mathcal{G}$, there do not exist parameters $\alpha, b, v$ such that the cases (i) and (iii) in Theorem 2 jointly emerge. Thus, we analyze the network in Figure 1 for two sets of parameters, where respectively cases (i), (ii) and (ii), (iii) jointly emerge.

In Figure 2(a), we focus on parameters $\alpha = 4, b = 2, v = 2$. Theorem 2 suggests that for these parameters, when $\sigma \geq 2$, equilibrium characterization is given as in case (i), whereas when $\sigma < 2$, case (ii) provides a characterization (see Appendix A.2 for detailed derivations). Note that when $\sigma = 4$, the equilibrium characterization is given as in case (i), yet due to the choice of problem parameters, the region of performance levels collapses to a line segment.

Figure 2(b), on the other hand, considers $v = 6$, while keeping $\alpha, b$ the same. In this case, for $\sigma \geq 2$ and $\sigma < 2$, respectively Theorem 2(ii) and (iii) provide the equilibrium characterization. In the latter case, the set of equilibrium performance levels becomes a singleton, depicted by ‘x’ in the figure. Interestingly, even in case (ii) (e.g., for $\sigma = 2$ or $\sigma = 4$), the equilibrium performance levels may collapse to a singleton. Otherwise, they correspond to the line segments in Figure 2(b).

In this section, we derived the main properties of optimal signaling mechanisms and their equilibria, and we obtained a precise characterization of the performances that are achievable by public signal mechanisms. In the remaining sections, we use these findings to design optimal signaling mechanisms under different assumptions on the primitives, and we compare their engagement/misinformation to that of the public signal mechanisms.
4. Engagement Maximization vs. Misinformation Minimization

In this section, we consider two extreme cases for the platform’s optimization problem (4): maximizing engagement (i.e., $\gamma_m = 0, \gamma_c = 1$) and minimizing misinformation (i.e., $\gamma_c = 0, \gamma_m = 1$). We show that the optimal intervention mechanisms in these two cases are fundamentally different. Section 4.1 focuses on the engagement maximization problem and establishes that the platform benefits from choosing different thresholds for different agents. In addition, in Section 4.1, we propose a novel centrality measure that captures how thresholds should be set for each agent according to her position in the underlying network. In stark contrast, Section 4.2 focuses on the misinformation minimization problem and establishes that public signal mechanisms are optimal and that choosing different thresholds is not beneficial for the platform.

4.1. Engagement Maximization

The problem of a platform that wishes to maximize engagement, irrespective of the resulting misinformation, is given as follows:

$$\max_{F \in \mathcal{F}, Q \in \mathcal{Q}(F)} E(Q). \quad (8)$$

We first characterize the optimal mechanisms solving this problem, and then we compare them to the maximum engagement that can be achieved by public signal mechanisms.

4.1.1. Optimal Mechanism

Theorem 1 implies that whenever a solution to (4) exists, there exists a straightforward threshold mechanism whose straightforward equilibrium solves this problem. Since (8) is a special case of (4), it follows that we can find its optimal solution by solving the following problem:

$$\max_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \sigma_i$$

s.t. $\sigma_i^2 \leq \frac{2}{b} v\sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}$

$\alpha^2 - \sigma_i^2 \geq \frac{2}{b} v(\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b} d_i \sigma_i$

$$0 \leq \sigma_i \leq \alpha. \quad (10)$$

Here, (11) guarantees that $\{\sigma_i\}_{i \in V}$ correspond to valid thresholds. As established in Proposition 1, the constraints (9) and (10) ensure that these thresholds induce a straightforward mechanism. The engagement at the straightforward equilibrium $Q$ of a straightforward threshold mechanism is equal to

$$E(Q) = E \left[ \sum_{i \in V} x_i \right] = E \left[ \sum_{i \in V} s_i \right] = \frac{1}{\alpha} \sum_{i \in V} \sigma_i.$$
Hence, the optimal thresholds obtained from (OPT-E) yield the straightforward threshold mechanism that maximizes engagement. As discussed earlier, this mechanism also solves (8). For the remainder of this section, we will refer to the optimization problem above as the *engagement maximization problem*. We point out that this is a convex optimization problem and, as such, can be solved efficiently.

The solution of the engagement maximization problem is governed by the following fundamental tradeoff: on the one hand, the platform wants to increase the thresholds (and hence the induced engagement), while on the other hand, the agents prefer smaller thresholds that ensure a lower disutility term when the recommendation is “Engage”. Clearly, when $b$ is small, the agents do not incur significant disutility from engaging with erroneous content, and the second effect disappears. In this case, to maximize engagement, the platform sets the thresholds to the highest possible levels.

**Lemma 3.** If $b \leq \frac{2}{a}(v + d_{\text{min}})$, then setting $\sigma_i = \alpha$ induces an optimal mechanism, and the corresponding engagement is equal to $n$.

Lemma 3 establishes that whenever agents do not incur too much disutility from engaging with erroneous content (i.e., $b$ is small), the problem of maximizing engagement is trivial: the platform always recommends engagement. In other words, the corresponding optimal mechanism is a no-intervention mechanism.

In contrast, when this disutility parameter is nontrivial (i.e., $b$ is large), in the absence of any intervention by the platform, agents may cease to engage with the content, which in turn reduces the overall engagement. As we establish in our next lemma, in such cases, it is optimal for the platform to intervene by setting (some) thresholds to levels strictly below $\alpha$.

**Lemma 4.** Assume that $b > \frac{2}{a}(v + d_{\text{min}})$. Let $\{\sigma_i\}_{i \in V}$ be an optimal solution of (OPT-E).

(i) There exists $i \in V$ such that $\sigma_i < \alpha$.

(ii) The engagement of the corresponding straightforward equilibrium $Q$ is strictly smaller than $n$, i.e., $E(Q) < n$.

(iii) Constraint (11) is not relevant in the sense that in the engagement maximization problem, the set of optimal solutions remains the same if this constraint is relaxed.

(iv) The optimal solution of (OPT-E) is unique.

**Fixed-point characterization.** In the remainder of this subsection, we provide insights into the structure of the optimal signaling mechanism and the resulting engagement. To this end, we focus on a special regime where a stronger version of the assumption of Lemma 4 holds. In particular, we assume the following:

\[ b > \frac{2}{\alpha}(v + d_{\text{max}}). \]
This strengthening of the assumption of Lemma 4 is interesting, as it allows us to focus on cases where agents incur significant disutility from misinformation and, hence, intervention by the platform is most critical. In particular, as we establish next, in this regime, $\sigma_i < \alpha$ for all $i \in V$, and consequently, for all agents, the platform recommends no engagement with the most erroneous content.

**Lemma 5.** Assume that $b > \frac{2}{\alpha}(v + d_{\text{max}})$. Let $\{\sigma_i\}_{i \in V}$ be the optimal solution to the engagement maximization problem. Then,

(i) (10) is not binding for any $i \in V$,

(ii) $0 < \sigma_i < \alpha$ for all $i \in V$,

(iii) $\sigma_i^2 = \frac{2}{b} \nu \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}$.

From a prescriptive point of view, Lemma 5 has a clear interpretation: when agents incur large disutility from engaging with erroneous content, an engagement-maximizing platform finds it optimal to recommend no engagement to all agents for large error realizations (part (ii)). In this regime, if the platform were to always recommend engagement to some agents, then, realizing that the expected disutility from engagement is high, these agents would not engage with the content at all. Observe that part (iii) of the lemma implies that the thresholds $\sigma_i$ are set to the largest possible values that make agents indifferent between engaging and not engaging when the recommendation is “Engage”. That is, the no-engagement recommendation is made as infrequently as possible, without violating agents’ incentives to follow the “Engage” recommendation. Note that since $\sigma_i$ are set as high as possible, conditional on receiving a “Do not engage” recommendation, agents can infer that the content is highly erroneous. Thus, it should come as no surprise that agents find it strictly optimal to follow the “Do not engage” recommendation, as can be seen from the first part of the lemma.

From a technical point of view, the most important consequence of Lemma 5 is that it transforms the optimization problem into a fixed-point equation:

$$\sigma_i^2 = \frac{2}{b} \nu \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}. \quad (12)$$

We proceed by showing that (12) has a unique nonzero solution, which also is the optimal solution of (OPT-E). Moreover, while in general, fixed-point equations need not admit tractable solutions, we establish that this is not the case for (12). In particular, we provide an efficient algorithm for the solution of this fixed-point equation.

**Proposition 2.** Suppose $b > \frac{2}{\alpha}(v + d_{\text{max}})$. 
(i) \( \{12\} \) has a unique solution where \( \sigma_i \in (0, \alpha] \) for all \( i \in V \), which is the optimal solution of \( \text{OPT-E} \).

(ii) Algorithm 1 terminates with this solution of \( \{12\} \).

(iii) If the algorithm sets \( \sigma_i \) before \( \sigma_j \), then \( \sigma_i \leq \sigma_j \).

Algorithm 1

A recursive algorithm for computing the fixed point of \( \{12\} \).

S1. Initialize with \( V = \{1, 2, \ldots, n\} \) and \( a_i = \frac{v}{b} \) for all \( i \in V \).

S2. For each \( i \in V \), obtain unique solution \( y_i^* \geq 0 \) to the equation

\[
y_i = \frac{v}{b} + \sqrt{a_i + \frac{2}{b} y_i \sum_{j \in V} g_{ij}}.
\]

(13)

Fix any \( j \in \arg\min_{i \in V} y_i^* \).

S3. Set \( \sigma_j = y_j^* \), \( V := V \setminus \{j\} \).

If \( V = \emptyset \), go to S4. Otherwise, for \( i \in V \), set \( a_i := a_i + \frac{2}{b} g_{ij} \sigma_j \). Go to S1.

S4. Return \( \sigma \).

In addition to providing a simple algorithm for solving \( \{12\} \) and hence \( \text{OPT-E} \), Proposition 2 also characterizes the agents whose thresholds are larger/smaller than others'. However, this characterization is implicit, as it is in terms of the order that the algorithm determines the thresholds. We next leverage the fixed-point characterization of optimal thresholds to explicitly relate agents' thresholds to their network position.

**Optimal thresholds and the network structure.** Our first result that relates optimal thresholds to agents’ network position focuses on degrees and provides bounds on thresholds in terms of degrees.

**Lemma 6.** Let \( \{\sigma_i\}_{i \in V} \) be the solution to \( \{12\} \) with \( \sigma_i > 0 \) for all \( i \in V \). Then,

(i) for \( i \in \arg\min_{j \in N_i} \sigma_j \), we have

\[
\sigma_i = \frac{2}{b} (v + d_i),
\]

(ii) \( \sigma_i \geq \frac{2}{b} (v + d_{\min}) \),

(iii) \( \sigma_i \leq \frac{2}{b} (v + d_i) \).

For agents who have the smallest thresholds, the fixed-point equation \( \{12\} \) simplifies to a linear equation, allowing for the closed-form characterization in this lemma. For the remaining agents, \( \{12\} \) can be used to obtain an explicit bound on the thresholds. Both types of characterizations provided in Lemma 6 feature agents’ degrees, suggesting that network structure can significantly impact the optimal thresholds, especially in settings where agents have vastly different degrees.
Intuitively, agents who have the smallest degrees benefit the least from network externalities. As such, they have the least incentive to engage with the content. Hence, unless their thresholds and, consequently, the expected error from engaging with the content (conditional on an “Engage” recommendation) are small, they find it optimal not to engage. This observation suggests that the platform may find it optimal to set the thresholds of these agents lower than the rest. Using the characterization in Lemma 6, our next corollary establishes that this indeed is the case.

**Corollary 1.** Let $F$ be the optimal threshold mechanism, with thresholds $\{\sigma_i\}_{i \in V}$. Suppose that $d_k \leq d_i$ for all $i \in V$. Then, $\sigma_k \leq \sigma_i$ for all $i \in V$.

In order to further explore the impact of the network structure on the optimal thresholds, we next focus on a particular class of networks where all agents have the same degree $k$, also known as $k$-regular graphs. As the next lemma suggests, for such networks, sending private signals to agents does not benefit the platform for maximizing engagement, and public signal mechanisms are optimal.

**Lemma 7.** If $G$ is $k$-regular and $2(v+k)/b < \alpha$, then optimal thresholds are given by $\sigma_i = 2(v+k)/b$ for all $i \in V$.

Intuitively, in $k$-regular graphs, nodes are “symmetric” in terms of their network position. Thus, for such graphs, choosing thresholds differently for each node should not help the platform increase engagement. Our lemma formalizes this intuition.

On the other hand, when the nodes are “asymmetric” in terms of their network position, some agents may be more “central” than the rest, and it may be optimal for the platform to choose different thresholds for different agents. Our discussion so far may suggest that the degree of each node determines the corresponding optimal threshold. However, by solving (12) for simple graphs it can be readily seen that degrees alone are not sufficient to characterize optimal thresholds or their rankings (e.g., nodes with smaller degrees may have larger or smaller thresholds). We illustrate this with an example in Appendix B.2. To shed light on the relationship between optimal thresholds and network positions, we next introduce a well-studied network centrality measure and we argue that the fixed points of equation (12) can inherently be interpreted as a (nonlinear) analogue of this centrality measure.

**Definition 7.** For a network with adjacency matrix $G$ and scalar $\gamma$, the **Bonacich centrality** vector of parameter $\gamma$ is given by

$$K(G,\gamma) = (I - \gamma G)^{-1}1$$

provided that $(I - \gamma G)^{-1}$ is well defined and nonnegative.
In the remainder of this section, whenever we write \( K(G, \gamma) \), we will assume that \( (I - \gamma G)^{-1} \) is well defined and nonnegative.

Fix \( \gamma \), and let \( \kappa = K(G, \gamma) \) denote the Bonacich centrality vector associated with this parameter and adjacency matrix \( G \). Observe that the definition of Bonacich centrality implies that \( (I - \gamma G)\kappa = \mathbf{1} \), or equivalently, for each agent \( i \), we have:

\[
\kappa_i = 1 + \gamma \sum_{j \in V} g_{ij} \kappa_j. \tag{14}
\]

Thus, Bonacich centrality of agents can equivalently be viewed as a solution of the above fixed equation. On the other hand, restricting attention to positive solutions (as in Proposition 2), (12) can equivalently be expressed as follows:

\[
\sigma_i = \frac{2\nu}{b} + 2 \frac{b}{\nu} \sum_{j \in V} g_{ij} \min \left\{ 1, \frac{\sigma_j}{\sigma_i} \right\}. \tag{15}
\]

Defining \( \sigma^e_i \triangleq \sigma_i \frac{b}{2\nu} \) for all \( i \in V \), we obtain:

\[
\sigma^e_i = 1 + \frac{1}{\nu} \sum_{j \in V} g_{ij} \min \left\{ 1, \frac{\sigma^e_j}{\sigma^e_i} \right\}. \tag{16}
\]

Thus, we conclude that \( \{\sigma^e_i\}_{i \in V} \) solve a fixed-point equation similar to (14). The only difference is that the right-hand side of (16) is nonlinear, unlike for (14). In particular, in (16), the “contribution” of \( j \) to its neighbor \( i \)'s \( \sigma^e_i \) is capped at one and scaled down by \( \sigma^e_i \). Despite these differences and motivated by the resemblance between (14) and (16), we define the unique solution of (16) where \( \sigma^e_i > 0 \) for all \( i \in V \) as engagement centrality.  

It turns out that engagement centrality can be formally related to other centrality measures. In particular, exploiting the resemblance between (14) and (16), our next result shows that engagement centralities of agents can be upper/lower bounded by their Bonacich centralities (with different parameters).

**Proposition 3.** Assume that \( b > \frac{2}{\alpha}(v + d_{\text{max}}) \), and let \( \sigma^e = \{\sigma^e_i\}_{i \in V} \) denote the engagement centrality vector. Then, provided that \( (I - \frac{2}{\beta b} G)^{-1} \) is well-defined and nonnegative, we have

\[
K \left( G, \frac{2}{\beta b} \right) \geq \sigma^e \geq K \left( G, \frac{2}{\alpha b} \right),
\]

where \( \beta = 2(v + d_{\text{min}})/b \).

\[\text{12} \] Here, uniqueness follows since (12) has a unique fixed point with positive \( \{\sigma_i\} \) and \( \sigma^e_i = \sigma_i \frac{b}{2\nu} \).
Summarizing, in this section, we have established that a platform that seeks to maximize engagement should employ different thresholds for different agents and hence it should possibly rely on private signals. Moreover, the thresholds of agents who are more central (according to engagement centrality) are higher. A natural question at this point is whether private signaling mechanisms that treat agents differently lead to substantially greater engagement than mechanisms that rely on public signals. We study this question in the next two subsections.

4.1.2. Public Signal and No-Intervention Mechanisms  We state our result by using the notations \( A_0(\sigma) \) and \( A_1(\sigma) \) to denote the sets \( A_0 \) and \( A_1 \) introduced in Theorem 2 for a given value of \( \sigma \), respectively. Similarly, we use \( k_0(\sigma) \) and \( k_1(\sigma) \) to denote the scalars \( k_0 \) and \( k_1 \) introduced in this theorem for the same \( \sigma \). Observe that these definitions imply that \( A_0(\sigma) \subset A_0(\sigma') \) and \( A_1(\sigma) \subset A_1(\sigma') \) when \( \sigma' \leq \sigma \). In addition, by construction, we have \( A_0(\sigma) \subset A_1(\sigma) \subset V \). As discussed in Theorem 2 for a given public signal mechanism with threshold \( \sigma \), there may be multiple equilibria with different performance levels. Throughout this subsection, among all equilibria induced by \( \sigma \), we focus on the one that has the maximum engagement.

Proposition 4. The maximum engagement level under public signal mechanisms is given by:

- \(|V|\), if \(2(v + d_{\text{min}})/\alpha \geq b\). This can be achieved by setting \( \sigma = 0 \) or \( \sigma = \alpha \).
- \(\max\{\max_{0 \leq \sigma \leq 2v/b} f_L(\sigma), \max_{2v/b \leq \sigma \leq \alpha} f_R(\sigma)\}\), if \(2(v + d_{\text{min}})/\alpha < b\). Here,

\[
  f_L(\sigma) := |V|\frac{\sigma}{\alpha} + |A_0(\sigma)|\frac{\alpha - \sigma}{\alpha} \quad \text{and} \quad f_R(\sigma) := |A_0(\sigma)|\frac{\alpha - \sigma}{\alpha} + |A_1(\sigma)|\frac{\sigma}{\alpha},
\]

respectively correspond to the maximum engagement achievable for a fixed \( \sigma \leq 2v/b \) and \( \sigma \geq 2v/b \).

An implication of this result is that to maximize engagement, the platform may find it optimal not to reveal any information (by setting \( \sigma = 0 \) or \( \sigma = \alpha \)). Thus, no-intervention mechanisms may be optimal for some range of parameters (when \(2(v + d_{\text{min}})/\alpha \geq b\)), consistent with Lemma 3. For the remaining range of parameters, the platform finds it optimal to choose the common threshold \( \sigma \) in a way that depends on the network structure. If the network features densely connected components, then \( A_0(\sigma) \) and \( A_1(\sigma) \) have large cardinality for large \( \sigma \). In this case, the platform finds it optimal to set \( \sigma > 2v/b \), and otherwise it chooses \( \sigma \leq 2v/b \). Intuitively, when the network is densely connected, the network externalities are large, and hence, even when the expected error is large, agents find it optimal to engage with content. Thus, the platform can set \( \sigma \) to be large and, despite high expected error, achieve high engagement.

A main disadvantage (other than performance) of using public signal mechanisms is lack of robustness. Note that while \(|A_0(\sigma)|\) and \(|A_1(\sigma)|\) decrease monotonically in the common threshold \( \sigma \), this is not necessarily the case for the corresponding engagement (i.e., functions \( f_L(\cdot) \) and \( f_R(\cdot) \))
are non-monotone). Moreover, slight perturbation of the optimal $\sigma$ can lead to significant drops in engagement. We illustrate this point in our next example.

**Example 2.** Consider the network in Example 1. Assume the parameters are as given in Figure 2, i.e., $\alpha = 4, b = 2, v = 2$. The maximum engagement achieved for different values of the public signal threshold $\sigma$ is as given in Figure 3(a). In this example, it can be seen that the engagement is non-monotone in $\sigma$. In addition, the maximal engagement is achieved when $\sigma \in \{0, \alpha\}$. Thus, no-intervention mechanism maximizes engagement, and under this mechanism, all agents engage with the content. This is a byproduct of the fact that disutility parameter $b$ is small. Next, consider the same example after setting $b = 4$. The corresponding engagement levels are now given in Figure 3(b). Observe that now no-intervention mechanisms do not maximize engagement. Moreover, we still observe non-monotonicity in engagement as a function of $\sigma$. Perhaps more interestingly, in this example, the optimal $\sigma$ is equal to 2. Yet, a slight increase in the optimal threshold reduces engagement roughly by a factor of two, indicating an important drawback of public signal mechanisms: a lack of robustness.

To understand why engagement is non-monotone in $\sigma$ for public signal mechanisms, one needs to consider the two-fold role of a signal in this setting: the signal is an informational device (since it gives information on the error of content) and a coordination device (since it is the same across agents). Although the impact of the first role on engagement is “continuous” in $\sigma$, the coordination aspect (as captured by $A_0(\sigma)$ and $A_1(\sigma)$) is discontinuous, a finding that is consistent with similar discontinuous outcomes observed in the coordination games literature (see Angeletos et al. (2006a))
and references therein). Thus, while the former role leads to continuously increasing engagement in $\sigma$, the latter leads to discontinuous drops, and their aggregate effect is non-monotone.

In Appendix D we formalize our observations from the previous example and provide a more refined characterization of the public signal mechanisms that maximize engagement (among all public signal mechanisms). Using this characterization, we also provide an efficient algorithm for obtaining engagement-maximizing public signal mechanisms.

4.1.3. Performance comparison Our findings in Sections 4.1.1 and 4.1.2 naturally bring up the question of comparing the engagement achieved by the optimal signaling mechanism, the optimal public signal mechanism, and the no-intervention mechanisms.

In this section, we first focus on $k$-regular graphs. As discussed earlier (and established in Lemma 7), for these graphs, public signal mechanisms are optimal among all mechanisms. However, it is not clear if, for these graphs, not intervening at all already guarantees high engagement. Our first corollary argues that for $k$-regular graphs, the benefit of intervention can range from being extremely significant to being non-existent.

**Corollary 2.** Let $E_O$, $E_P$, and $E_N$ respectively denote the maximum engagement that can be obtained by the optimal (private) signal mechanism, a public signal mechanism, and a no-intervention mechanism. Moreover, assume that the underlying network is $k$-regular. Then

(i) if $b \leq 2(v + k)/\alpha$, then $E_N = E_P = E_O = |V|$,

(ii) if $b > 2(v + k)/\alpha$, then

$$E_N = 0 < E_P = E_O = \frac{2(v + k)}{\alpha b}.$$ 

This corollary implies that no intervention may lead to a drastic reduction in the amount of engagement on the platform depending on the connectivity ($k$) of the underlying network and the parameters of agents’ payoffs. For example, when agents are not as well connected (small $k$) and $b > 2(v + k)/\alpha$, in the absence of intervention, the platform observes no engagement, whereas with an appropriately designed public signal mechanism, an amount of engagement proportional to the size of the network can be obtained.

In Corollary 2 since we focus on $k$-regular graphs, there is no gap between public and private signaling mechanisms in terms of engagement. We next analyze more general graphs and explore whether these mechanisms induce very different engagement levels.

**Corollary 3.** Let $E_O$ and $E_P$ respectively denote the maximum engagement that can be obtained by the optimal (private) signaling mechanism and a public signal mechanism. Consider a graph that consists of two disjoint graphs with $n/2$ nodes each: a ring graph where each node is connected to its
immediate neighbor and a ring graph where each node is connected to its $2m$ immediate neighbors, where $m > 1$. Moreover, assume that $b > 2(v + 2m)/\alpha$. Then,

$$E_O = 2n \frac{v + m + 1}{\alpha b}, \quad \text{while} \quad E_P = n \cdot \max \left\{ \frac{2(v + 2)}{\alpha b}, \frac{v + 2m}{\alpha b} \right\}.$$  

This corollary readily implies that when agents in the network are heterogeneous in terms of their network positions, the engagement achieved under the best public signal mechanism can be substantially lower than that under the optimal private signaling mechanism. To see this, note that in the corollary by choosing $m = 2$ and $v$ arbitrarily small, we immediately obtain examples where $E_O/E_P \approx 3/2$, i.e., the optimal private signaling mechanism leads to 50% more engagement.

In conclusion, we have established that an engagement-maximizing platform greatly benefits from agent-specific signaling of content accuracy, which exposes more central agents to content with higher expected error. In the next section, we switch our focus to misinformation minimization and obtain a strikingly different result.

4.2. Misinformation Minimization

Theorem 1 implies that it suffices to restrict attention to straightforward threshold mechanism to minimize misinformation. Thus, following a similar approach to the one that we took for engagement maximization, we restrict attention to such mechanisms. Once again, using Proposition 1, it follows that the thresholds of the straightforward threshold mechanism that minimizes misinformation can be obtained by solving the following problem:

$$\min_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \frac{\sigma_i^2}{2}$$

s.t. $\sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}$

$\alpha^2 - \sigma_i^2 \geq \frac{2}{b} v (\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b} d_i \sigma_i$

$0 \leq \sigma_i \leq \alpha,$

where the objective captures the expected misinformation of the straightforward equilibrium $Q$ of the underlying threshold mechanism, since for such a mechanism we have:

$$M(Q) = \mathbb{E} \left[ \sum_{i \in V} \epsilon x_i \right] = \mathbb{E} \left[ \sum_{i \in V} \epsilon s_i \right] = \mathbb{E} \left[ \sum_{i \in V} \epsilon 1(\epsilon \leq \sigma_i) \right] = \frac{1}{\alpha} \sum_{i \in V} \frac{\sigma_i^2}{2}.$$
Surprisingly, the main result of this section, which is stated next, implies that at the optimal solution of this problem, all agents have the same threshold, i.e., misinformation is minimized via public signal mechanisms.

**Theorem 3.** At the optimal solution of \( (\text{OPT-M}) \), for all \( i \in V \), we have

\[
\sigma_i = \begin{cases} 
0 & \text{if } \alpha \geq \frac{2v}{b}, \\
\min\{\alpha, \frac{2v}{b} - \alpha\} & \text{otherwise.}
\end{cases}
\]

Theorem 3 implies that when agents’ disutility from engaging with inaccurate content (parameter \( b \)) is large, the platform can minimize misinformation by not recommending any content for engagement (\( \sigma_i = 0 \)). Thus, in this case no intervention is optimal. This is because, an agent’s expected utility from engaging with inaccurate content is significant, and hence, in the absence of any intervention, agents find it optimal not to engage with the content, which in turn leads to minimal misinformation.

On the other hand, if the disutility \( b \) from engaging with inaccurate content is not excessively large (so that \( \alpha < \frac{2v}{b} < 2\alpha \)), then the platform finds it optimal to intervene by setting \( \sigma_i = \frac{2v}{b} - \alpha \). Employing a no-intervention mechanism in this case would lead to nonnegative expected payoff from always engaging with the content, thereby leading to significant levels of (engagement and) misinformation. However, with proper signaling, the platform can help agents identify cases where the error of the content is substantial and avoid engagement with highly inaccurate content. Thus, intervention helps for moderate levels of disutility parameter \( b \) by allowing agents to coordinate better on accurate content.

In the extreme case where \( b \) is very small (so that \( 2v/b \geq 2\alpha \)), the platform sets \( \sigma_i = \alpha \) and hence again uses a no-intervention mechanism. In this regime, agents’ direct utility from engaging with the content outweighs the disutility that they incur when the content is inaccurate. Thus, regardless of the platform’s intervention, the agents find it optimal to always engage. In this case, any mechanism (including the no-intervention mechanism) is optimal.

Theorem 3, in addition to prescribing the optimal intervention mechanism, provides two important and counterintuitive insights: (i) for the problem of minimizing misinformation, the platform should always rely on identical recommendations for all agents, and (ii) the minimum level of misinformation for the platform does not depend on the underlying network structure. Both findings are striking since intuitively, one might expect private signal mechanisms to perform strictly better when agents are heterogeneous in terms of their network position (as in the engagement maximization case). Similarly, it might be expected that denser networks suffer more from high...
mispertual since, due to network externalities, agents have a higher incentive to engage with content even when the error is not small.

The optimality of public signal mechanisms can be attributed to the fact that they enable agents to coordinate their actions in a way that makes network effects irrelevant. More precisely, to minimize misinformation, the platform needs to ensure that agents do not engage with content that has high error. In public signal mechanisms (with sufficiently small $\sigma$), when the platform recommends that agents not engage, all agents can coordinate on this action and cease to engage. In this case, an agent cannot unilaterally deviate and improve her payoff since the other agents do not engage and network externalities do not play a role. Thus, thanks to this coordination effect, public signal mechanisms suffice to guarantee that no agent engages for large error realizations. Choosing $\sigma$ appropriately, this is sufficient to minimize the overall misinformation.

We next compare (optimal) public signal mechanisms with no-intervention mechanisms and quantify the improvement due to intervention.

**Corollary 4.** If $\frac{2\alpha}{v} \leq \alpha$ or $\frac{2\alpha}{v} \geq 2\alpha$, then no-intervention mechanisms are optimal. Otherwise, let $M_P$, $M_N$ respectively denote the misinformation associated with the optimal public signal mechanism and no-intervention mechanism respectively. Then,

$$M_N = \left| V \right| \frac{\alpha}{2}, \text{ and } M_P = \left| V \right| \frac{\left( \frac{2\alpha}{v} - \alpha \right)^2}{2\alpha}.$$

This corollary implies that for moderate levels of disutility parameter $b$ (so that $\alpha < 2v/b < 2\alpha$), no-intervention mechanisms may lead to substantially more misinformation than public signal mechanisms. In particular, in this regime, we have

$$\frac{M_N}{M_P} = \left( \frac{\alpha}{2v/b - \alpha} \right)^2.$$

Thus, for $2v/b \approx \alpha$, the misinformation ratio becomes unbounded, clearly highlighting the substantial effect of intervention on misinformation levels.

### 4.3. Numerical Studies

In this section, we numerically evaluate the performance (in terms of maximum engagement and minimum misinformation) of the different signaling mechanisms (optimal/public signal/no-intervention) presented in this section. As an additional benchmark, we consider a mechanism where the platform reveals the true error $\epsilon$ of the content publicly to all agents (for any realization of $\epsilon$). Characterization of equilibria for this mechanism is very similar to that of public signal mechanisms and is outlined in Appendix E.
For this purpose, we focus on an induced subnetwork of the Facebook network, provided by Leskovec and Krevl (2014). The subnetwork consists of 4,039 nodes and 88,234 edges. The average degree in the subnetwork is 43.7, and degrees vary between 1 and 1,045. Note that nodes in the subnetwork may have connections in the full network that do not belong to the subnetwork. Hence, the degree distribution in the subnetwork is biased towards lower values. We plot the degree distribution of this network in Figure 4 and provide a visualization of the network in Figure 6. As is obvious from these figures, the Facebook subnetwork is far from a regular graph, and nodes are substantially differentiated based on their position in the network.

4.3.1. Numerical Studies: Engagement Maximization. In Figure 5(a) we compare the engagement obtained for the Facebook subnetwork under different mechanisms. In particular, we focus on the optimal mechanism of Section 4.1.1, the public signal mechanism with the highest level of engagement, the no-intervention mechanism, and finally the true error revelation mechanism (of Appendix E). We plot the engagement obtained under these mechanisms for different levels of the misinformation disutility parameter $b$. For our simulations, we normalize $\alpha$ to 1 and set $v = 20$.

Our results indicate that for low levels of $b$, under all of these mechanisms, all agents always engage. In particular, for the optimal/public signal/no-intervention mechanisms this seems to be the case for $b \leq 2(v + d_{\text{min}})/\alpha = 42$, consistent with our Lemma 3. On the other hand, for larger levels of $b$, the engagement that can be achieved under these mechanisms significantly differs. Figure 5(a) suggests that under the no-intervention mechanism, the engagement that the platform can achieve quickly decreases as $b$ increases. The decrease in engagement is more gradual for public signal, true error revelation mechanisms, and the optimal mechanism. For large values of $b$, the
public signal and true error revelation mechanisms behave similarly, and the gap between these mechanisms and the optimal mechanism turns out to be significant (e.g., for the rightmost data point, the ratio of engagement for the optimal and public signal mechanisms is 2.6).

(a) Maximum engagement under different mechanisms.  
(b) Misinformation under different mechanisms.

Figure 5  Numerical comparison of the performance of different mechanisms on the Facebook subgraph.

Under the optimal mechanism, relative to the threshold of the public signal mechanism that achieves the greatest engagement, some agents will have lower thresholds, and others will have higher thresholds. To illustrate this, we focus on disutility parameter $b = 200$. In Figure 6, we provide a visualization of the network obtained using Gephi, with agents whose thresholds correspond to the highest %10 and lowest %10 respectively in red and blue (while shading the remaining agents). Consistent with the results of Section 4.1.1, the agents who appear to be more tightly connected and central in the network have higher thresholds.

4.3.2. Numerical Studies: Minimizing Misinformation  We next focus on minimizing misinformation for the Facebook subnetwork. Assume once again that $\alpha = 1, v = 20$. We vary $b$ and plot how the misinformation levels under the optimal mechanism, true error revelation mechanism, and the no-intervention mechanism change as a function of $b$. As established in Theorem 3, public signal mechanisms are optimal in terms of minimizing engagement; hence, in this section, we do not separately discuss them.

In agreement with our theoretical characterizations of Theorem 3 and Corollary 4, Figure 5(b) suggests that when an agent’s disutility from engaging with inaccurate content is moderate, the platform can significantly reduce misinformation with appropriate signaling. However, for large or small levels of disutility, the platform does not need to intervene. Finally, we observe that mechanisms revealing true error are significantly suboptimal relative to public signal mechanisms.

13 See Bastian et al. (2009) for details on using Gephi for network visualization.
for the purpose of misinformation minimization. This is the case, whenever the disutility parameter $b$ is non trivial.

5. Engagement/Misinformation Frontier

There is an important tradeoff between (lowering) misinformation and (increasing) engagement: achieving high engagement requires agents to engage more intensively with content, including highly erroneous content, which in turn induces high misinformation. The results in Sections 4.1 and 4.2 characterize the platform’s optimal signaling mechanisms in cases where the platform wants to purely maximize engagement (ignoring any implications for misinformation) or minimize misinformation (ignoring any implications for engagement). These mechanisms turn out to be qualitatively very different. In this section, our objective is to understand how the platform can obtain an optimal tradeoff between misinformation and engagement. To this end, we focus on the following optimization problem:

$$M^*(E) \triangleq \min_{F \in \mathcal{F}, Q \in \mathcal{Q}(F)} M(Q)$$

subject to $E(Q) \geq E$.

Observe that (17) provides mechanisms that induce a minimum level of misinformation while guaranteeing a certain level of engagement. We refer to $E \in [0, \alpha]$ as the engagement requirement.

Consider the curve

$$\{(E, M^*(E)) \mid n \geq E \geq 0\}.$$
It can be readily seen that given a point on this curve, there is no mechanism in $F$ that achieves lower misinformation without lowering engagement (or higher engagement without increasing misinformation). Thus, these points achieve the best possible tradeoffs between engagement and misinformation. We refer to this curve as the Engagement/Misinformation frontier (E/M frontier).

We start our discussion by first analyzing the optimal mechanisms that solve (17) and characterizing the E/M frontier. Then, we restrict our attention to the public signal and the no-intervention mechanisms and compare their performance levels to those of the E/M frontier.

5.1. Optimal Mechanism

Once again, to obtain optimal mechanisms solving (17), we restrict attention to straightforward threshold mechanisms and focus on the following problem:

\[
M^*(E) = \min_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \sigma_i^2
\]

s.t. $\sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}$

\[\alpha^2 - \sigma_i^2 \geq \frac{2}{b} v (\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b} d_i \sigma_i\] (OPT-F)

\[\sum_{i \in V} \sigma_i \geq \alpha E\]

\[0 \leq \sigma_i \leq \alpha.\]

The constraints other than the third one, as before, guarantee that the optimal solution yields thresholds of a straightforward threshold mechanism (see Proposition 1). The third constraint is the engagement requirement constraint from (17). For the remainder of this section, we denote the solution to (OPT-F) for a given $E \geq 0$ by $\sigma_i^*(E)$.

If we remove the two families of incentive compatibility constraints (i.e., the first two constraints), the resulting problem is fairly easy to solve: we set all the thresholds equal to $\alpha E/n$ (whenever $E \leq n$). This provides a lower bound to $M^*(E)$, as the next lemma formalizes.

**Lemma 8.** $M^*(E) \geq \frac{\alpha}{2n} E^2$.

We proceed by providing a particular parameter regime where the optimal signaling mechanism can achieve the lower bound of Lemma 8. To obtain this result, we show that for this regime, incentive compatibility constraints in (OPT-F) are irrelevant and can be relaxed.

14 Strictly speaking, this problem is not of the form (4), so it does not directly follow from Theorem 1 that it suffices to restrict attention to straightforward threshold mechanisms. However, the theorem implies that given a signaling mechanism, a straightforward threshold mechanism with weakly lower misinformation and the same amount of engagement can always be obtained. Thus, it still suffices to search for optimal mechanisms by restricting attention to this class of mechanisms.
Lemma 9. If \( E \in \left[ \max \left\{ 0, \frac{n}{\alpha} \left( \frac{2v}{b} - \alpha \right) \right\}, \min \left\{ n, \frac{2n}{\alpha} \frac{d_{\text{min}} + v}{b} \right\} \right] \), then
\[
M^*(E) = \frac{\alpha}{2n} E^2, \quad \text{and} \quad \sigma_i^*(E) = \frac{\alpha E}{n} \quad \text{for all } i \in V.
\]

Interestingly, this lemma implies that when the engagement requirement belongs to the given interval, there is a corresponding public signal mechanism with appropriately chosen thresholds that solves (OPT-F). Hence, the lemma identifies a regime where the E/M frontier is achieved via public signal mechanisms. This is due to the fact that in this regime, the platform wants to ensure a moderate engagement level while minimizing misinformation. Consistent with the results of Section 4.2, public signal mechanisms facilitate coordination among agents on the no-engagement action for high levels of error and achieve low misinformation and moderate engagement.

In general, the minimum misinformation that can be achieved on the platform (i.e., the optimal objective of (OPT-M)) is strictly greater than zero. This can be seen from Theorem 3 by noting that the misinformation-minimizing thresholds can be strictly positive and hence lead to nontrivial engagement and misinformation. Thus, as the engagement requirement in (OPT-F) decreases, eventually, the corresponding constraint becomes irrelevant, and the platform’s problem becomes equivalent to (OPT-M). In this case, the platform finds it optimal to implement misinformation-minimizing mechanisms. Based on the results of Section 4.2, public signal mechanisms are optimal for this purpose, and hence they belong to the E/M frontier. Our next lemma characterizes this regime.

Lemma 10. Suppose \( E \in [0, n] \) is such that \( E \leq \max \left\{ 0, \frac{n}{\alpha} \left( \frac{2v}{b} - \alpha \right) \right\} \).

- If \( \frac{2v}{b} \leq \alpha \), then \( M^*(E) = \sigma_i^*(E) = 0 \) for all \( i \in V \).
- Otherwise,
  \[
  M^*(E) = \frac{n}{2\alpha} \left( \min \left\{ \alpha, \frac{2v}{b} - \alpha \right\} \right)^2, \quad \text{and} \quad \sigma_i^*(E) = \min \left\{ \alpha, \frac{2v}{b} - \alpha \right\} \quad \text{for all } i \in V.
  \]

Lemmata 9 and 10 jointly characterize the E/M frontier for \( E \in [0, \min \left\{ n, \frac{2n}{\alpha} \frac{d_{\text{min}} + v}{b} \right\}] \). Note that if \( b \leq 2(d_{\text{min}} + v)/\alpha \), then the aforementioned interval contains \( [0, n] \), and hence, the lemmata above fully characterize the E/M frontier. In other words, this implies that when the disutility parameter \( b \) is small, public signal mechanisms are sufficient to obtain all performance levels on the E/M frontier and to achieve optimal tradeoffs between engagement and misinformation.

We next focus on complementary settings where \( b > 2(d_{\text{min}} + v)/\alpha \) and engagement levels higher than \( 2n(d_{\text{min}} + v)/(\alpha b) \) could be feasible. For \( E > 2n(d_{\text{min}} + v)/(\alpha b) \), it is not possible to solve (OPT-F) in closed form and obtain the E/M frontier explicitly. Still, in the remainder of this section, we explore the structure of the E/M frontier in this regime and provide a characterization.
Lemma 11. Suppose $b > 2(d_{\min} + v)/\alpha$ and the network is not regular. Let $m$ denote the number of minimum-degree nodes, i.e., $m = |\{i \mid d_i = d_{\min}\}|$. Then, there exists $\delta > 0$ such that, for all $E \in \left(\frac{2n d_{\min} + v}{b}, \frac{2n d_{\min} + v}{\alpha b} + \delta\right)$, we have

$$M^*(E) = \frac{2m}{\alpha} \left(\frac{v + d_{\min}}{b}\right)^2 + \frac{1}{2\alpha(n-m)} \left(\alpha E - m\frac{2(v + d_{\min})}{b}\right)^2.$$

Lemma 11 suggests that when $E$ is slightly above the level where sending a public signal is optimal, the E/M frontier depends on the degree distribution of the underlying network. In particular, the shape of the E/M frontier is still quadratic in $E$, but the coefficient of the quadratic term is inversely proportional to the number of non-minimum-degree agents $(n - m)$. In other words, when there are fewer high-degree agents on the platform, increasing engagement becomes more costly in terms of misinformation. This is because agents with small degrees do not benefit as much from network externalities, and hence, the platform cannot increase their engagement significantly. Thus, when the engagement requirement is high, to satisfy this requirement, the platform finds it optimal to recommend engagement to high-degree nodes for highly erroneous content (by setting their thresholds higher than those of the minimum-degree nodes – and hence employing private signal mechanism). When there are fewer higher-degree agents, this effect becomes more pronounced.

We conclude this section by providing another result that relates the E/M frontier to the underlying network structure. Clearly, the engagement cannot be increased indefinitely. Specifically, there exists some $\bar{E}$ for which there does not exist a mechanism that supports engagement higher than $\bar{E}$, i.e.,

$$\bar{E} = \sup\{E : \text{OPT-F is feasible}\}.$$

This quantity corresponds to the maximum engagement that can be achieved by a mechanism in $F$, and can be readily characterized by our results in Section 4.1.1. Our next corollary provides the point on the E/M frontier that corresponds to $\bar{E}$, and follows immediately from Propositions 2 and 3.

Corollary 5. Assume that $b > \frac{2}{\alpha}(v + d_{\max})$. Then,

(i) $\bar{E} = \sum_{i \in V} \sigma_i / \alpha$, where $\{\sigma_i\}_{i \in V}$ is the unique solution to (12).

(ii) $\frac{4\alpha^2}{\beta} \sum_i K_i^2(G, \frac{2}{\beta}) \geq M^*(\bar{E}) \geq \frac{4\alpha^2}{h \beta} \sum_i K_i^2(G, \frac{2}{\alpha b})$, where $\beta = 2(v + d_{\min})/b$, and $K_i(G, \gamma)$ denotes the $i$th entry of the Bonacich centrality vector with parameter $\gamma$.

This result implies that the misinformation of the “last” point on the E/M frontier depends on the sum of squares of the Bonacich centralities of the users, which in turn highlights the impact of the network structure on the misinformation-engagement tradeoff.
5.2. Public Signal and No-Intervention Mechanisms

In general, the performance level achieved by a public signal mechanism (or a no-intervention mechanism) will not be on the E/M frontier. Lemmata 9 and 10 identify some exceptions to this. The main result of this section, presented below, complements these lemmata by focusing on the regime where the disutility parameter is large (and the intervention by the platform is necessary), and establishing that outside the regime identified in the lemmata, the public signal mechanisms are not on the E/M frontier.

Proposition 5. Assume that $b > \frac{2}{\alpha}(v + d_{\text{max}})$. Then, a public signal mechanism can achieve a performance level that is on the E/M frontier, if and only if $E \leq \frac{a}{n} \min\{\frac{2}{b}(v + d_{\text{min}}), \alpha\}$.

Thus, we conclude that in general the public mechanisms do not achieve the optimal tradeoff between engagement and misinformation. Due to lack of a closed-form solution to (OPT-F) for $E \geq \frac{a}{n} \min\{\frac{2}{b}(v + d_{\text{min}}), \alpha\}$, we continue our exploration of the E/M frontier of public signal mechanisms through numerical studies.

5.3. Numerical Studies: E/M Frontier

We conclude this section by illustrating the E/M frontier for the Facebook subnetwork considered in Section 4.3. We assume that $\alpha = 1, v = 20$, and $b \in \{30, 100\}$. Recall that if $b < v/\alpha = 20$, then Figure 5(b) implies that under mechanisms that minimize misinformation, all agents engage with the content. Since this is true for mechanisms that minimize misinformation, it follows that under any mechanism, all agents always engage. Thus, in this case, for all engagements such that $E \leq n$, the corresponding misinformation at the E/M frontier is given by $M^*(E) = 4039/2$ (using the fact that there are 4039 agents).

When $b \in \{30, 100\}$, we have $b > v/\alpha$, and the E/M frontier has a more interesting structure. For $b = 30$ and $b = 100$, the E/M frontier is respectively illustrated in Figures 7 and 8. In these figures, in addition to the E/M frontier, we also plot the convex hull of the performance levels that can be obtained using public signal mechanisms and the no-intervention mechanism.

For $b = 30$, we see that the no-intervention mechanism can implement only a single performance level. This point corresponds to an outcome where all agents always engage and is on the E/M frontier. The public signal mechanisms, on the other hand, can achieve any point on the E/M frontier. For $b = 100$, we see that the results qualitatively change. In this case, the convex hull of the performance levels of no-intervention mechanisms constitute a line segment that has one end point at the origin. This point corresponds to an outcome where no agent engages, and it is on the E/M frontier. Moreover, for $b = 100$, the public signal mechanisms can achieve the optimal tradeoff between engagement and misinformation for low engagement requirement levels. However,
these mechanisms fail to increase engagement as much as the mechanisms that rely on agent-specific thresholds do. Hence, they cannot cover the end of the E/M frontier associated with high engagement requirements.

6. Conclusions

In this paper, we studied the problem of a platform that is informed about the accuracy of the available content and wishes to minimize the misinformation on the platform while maximizing engagement. Agents on the platform are uninformed about the accuracy of the content, and they decide whether to engage with it based on their satisfaction from engaging, the potential disutility from engaging with erroneous content, and the externality obtained from engaging with the same content as their friends. However, the platform can send signals to agents about the accuracy of the content to influence their behavior.

We establish that the platform’s optimal signaling mechanisms are straightforward threshold mechanisms, i.e., it is sufficient for the platform to choose thresholds for agents and recommend engagement if and only if the error associated with the content is lower than this threshold. The mechanisms that maximize engagement rely on agent-specific thresholds and possibly send different engagement recommendations to agents depending on their network position (as captured through a novel centrality measure). On the other hand, if the platform wants to minimize misinformation regardless of the impact on engagement, public signal mechanisms (which have identical thresholds for all agents) are optimal, and the optimal level of misinformation is independent of the network
structure. We also characterize how the platform can ensure an optimal tradeoff between engagement and misinformation, by studying the E/M frontier of performance levels, and identifying regimes where public signal mechanisms achieve optimal tradeoffs.

To the best of our knowledge, this paper is the first to study information design for a social network. Our model captures the main tradeoffs faced by a platform that can send signals to agents regarding the quality of the available content and opens up a number of interesting future directions. Among them, we mention the following:

**Time component:** Our paper focuses on a static model where the platform commits to a signaling mechanism before observing the accuracy of the content and the users simultaneously decide whether to engage. A reasonable alternative is one where agents sequentially make decisions, i.e., some agents act first and influence their friends, their friends influence their own friends, and so on. We believe that understanding how a platform can achieve the desired tradeoff between engagement and misinformation in such a dynamic environment, possibly by dynamically adjusting the signals that it sends, is a challenging but exciting future direction.

**Error uncertainty:** In this paper, we have assumed that the platform is completely informed about the error of the content. This assumption is an approximation to reality, where a platform potentially has a more accurate estimate of the error than the agents, and our results could be extended to such settings. A closely related but different direction pertains to learning. In particular, suppose that the platform does not precisely know the error and that agents also have noisy signals about the accuracy of the content. If agents sequentially make engagement decisions, the platform can learn from their revealed engagement decisions whether the content is of high quality or not. Furthermore, the platform can use the learned information to dynamically adjust its mechanism. This setting involves key ideas from learning, networks, and information design, which we believe constitutes a fruitful research direction.

**Idiosyncratic preferences:** In this work, we have assumed that all agents receive the same satisfaction \( v \) from engaging with the content. On the other hand, the satisfaction from engaging with particular content may be different across agents and can be captured by a more general satisfaction term \( v_i \) for each agent \( i \) that depends on other (a priori known or unknown) characteristics of the content, such as bias, ideological positioning, etc. We decided not to add this ingredient to our analysis, as we primarily are interested in understanding the role of the network structure in the optimal mechanisms and induced performances. That being said, given the polarization that is observed in online social media (where users often engage with content that is close to their own “type”), an extension of the framework that we developed in this paper to settings with heterogeneous content preferences could potentially lead to important practical insights.
Applications of information design in social and economic networks: Information is an important lever many platforms can play with to induce desired outcome in social networks. The engagement-misinformation problem studied in this paper exemplifies the significant improvement the platforms can achieve in their operations by appropriately using this lever. We find it an exciting future avenue of research to explore other applications of information design in social and economic networks. We anticipate that there will be wide variety of domains where these tools can be employed, ranging from social networks to transportation networks.

References


Appendix

A. Proofs for Section 3

In this section, we first briefly discuss some technical assumptions we impose on signaling mechanisms. Then, we provide proofs of claims stated in Section 3.

A.1. Technical Assumptions on Signaling Mechanisms

Recall that the underlying probability space is given by \((\Omega, \Sigma, P)\). We assume that \(\Omega = \Omega_e \times \Omega_r\), where \(\Omega_e = [0, \alpha]\) captures the possible realizations of the error and \(\Omega_r\) denotes the possible outcomes associated with the randomization device that the platform uses to generate the (randomized) signals. Furthermore, we assume that \(\Sigma\) is the smallest sigma algebra containing \(\Sigma_e \times \Sigma_r\), where \(\Sigma_e\) is the Lebesgue \(\sigma\)-algebra associated with \(\Omega_e = [0, \alpha]\) and \(\Sigma_r\) is the \(\sigma\)-algebra associated with \(\Omega_r\). Similarly, the probability measure \(P\) is assumed to be a product measure between \(U[0, \alpha]\), and \(P_r\) – the probability measure corresponding to the randomization device. Note that using this product measure assumption together with Fubini’s theorem, for any \(\tilde{S} \subset S_i\), we obtain:

\[
\mathbb{P}(s_i \in \tilde{S}) = \int \mathbb{1}(s_i(\omega) \in \tilde{S})d\mathbb{P} = \int_0^\alpha \frac{1}{\alpha} \int \mathbb{1}(s_i(z, \omega_r) \in \tilde{S})d\mathbb{P}_r dz
\]

where \(s_{-i}\) denotes the signals of all agents but \(i\), and similarly \(S_{-i} = \times_{k \neq i} S_k\). Thus, the product measure assumption allows for expressing the probability of observing a set of signals in terms of the underlying mechanism in a straightforward way.

A.2. Proofs

Proof of Lemma 1. Let \(F \in \mathcal{F}\) denote a given mechanism, and assume that equilibrium \(Q \in Q(F)\) achieves engagement \(L\) and misinformation \(M\). Denote by \(S_i^k\) the set of signals that at equilibrium \(Q\) induce the action \(x_i = k\) for \(k \in \{0, 1\}\), i.e., \(S_i^k = \{s \in S_i : \mu_i(s) = k\}\). Consider the mechanism \(F^* \in \mathcal{F}\), for which the signals \(s_i^*: \Omega \to S_i\) are such that

\[
s_i^*(\omega) = 1, \quad \text{for } \omega \text{ such that } s_i(\omega) \in S_i^1 \text{ and }
\]

\[
s_i^*(\omega) = 0, \quad \text{for } \omega \text{ such that } s_i(\omega) \in S_i^0.
\]

In other words, all signals that induce action 1 (action 0) in mechanism \(F\) are collected to a single recommendation \(s_i^* = 1 (s_i^* = 0)\) in \(F^*\).

We next prove that the constructed mechanism \(F^*\) is straightforward, i.e., it is an equilibrium for each agent \(i\) to use the strategy \(\mu_i^*(s_i^*) = s_i^*\). Assume first that \(\mathbb{P}(s_i \in S_i^k) > 0\) and hence \(\mathbb{P}(s_i^* = 1) = 1 - \mathbb{P}(s_i \in S_i^0) > 0\) and \(\mathbb{P}(s_i^* = 0) = 1 - \mathbb{P}(s_i \in S_i^1) > 0\). Thus, \(\mu_i^*(s_i^*) = 1\) for all \(s_i^* \in S_i^1\) and \(\mu_i^*(s_i^*) = 0\) for all \(s_i^* \in S_i^0\).
these inequalities immediately imply that

\[ 0 \leq \mathbb{E} \left[ v - b \epsilon + \sum_{j} g_{ij} \mu_{j}(s_{j}) \mid s_{i} \in S_{i}\right] = \mathbb{E} \left[ v - b \epsilon + \sum_{j} g_{ij} \mu_{j}^{*}(s_{j}^{*}) \mid s_{i} \in S_{i}\right]. \quad (18) \]

and similarly

\[ 0 \geq \mathbb{E} \left[ v - b \epsilon + \sum_{j} g_{ij} \mu_{j}(s_{j}) \mid s_{i} \in S_{i}\right] = \mathbb{E} \left[ v - b \epsilon + \sum_{j} g_{ij} \mu_{j}^{*}(s_{j}^{*}) \mid s_{i} \in S_{i}\right]. \quad (19) \]

Noting that \( \{\omega | s_{i}(\omega) \in S_{i}\} = \{\omega | s_{i}^{*}(\omega) = 1\} \) and \( \{\omega | s_{i}(\omega) \in S_{i}\} = \{\omega | s_{i}^{*}(\omega) = 0\} \), by construction these inequalities immediately imply that

\[ \mathbb{E} \left[ v - b \epsilon + \sum_{j} g_{ij} \mu_{j}^{*}(s_{j}^{*}) \mid s_{i}^{*} = 0\right] \leq 0 \leq \mathbb{E} \left[ v - b \epsilon + \sum_{j} g_{ij} \mu_{j}^{*}(s_{j}^{*}) \mid s_{i}^{*} = 1\right]. \quad (20) \]

Thus, it follows that in \( F^{*} \), if agent \( i \) receives a signal to engage (not engage), then her pay-off is maximized by engaging (not engaging) with the content. Hence, \( \{\mu_{i}^{*}\} \) is a straightforward equilibrium and \( F^{*} \) is a straightforward mechanism.

If \( \mathbb{P}(s_{i} \in S_{i}^{k}) = 0 \) for some \( k, i \), then, under \( F^{*} \), the corresponding agents almost surely receive either signal 0 or 1 and, to characterize the equilibrium, only one of the inequalities in (20) suffices. In addition, only one of (18) and (19) hold for equilibrium \( Q \) and establishes the optimality of following the aforementioned signal. Thus, straightforwardness of \( F^{*} \) readily follows.

Finally, by construction, the engagement and the misinformation achieved by the straightforward equilibrium of the new mechanism \( F^{*} \) are the same as in the equilibrium \( Q \) of \( F \) since in both mechanisms, all agents engage for precisely the same \( \omega \in \Omega \). Hence, the claim follows. \( \square \)

**Proof of Theorem** The second part of the theorem follows immediately from the first part, i.e., if a straightforward threshold mechanism yields same level of engagement as a given mechanism and lower misinformation, then whenever a solution to (4) exists, there exists a solution that is a straightforward threshold mechanism. Thus, we focus on establishing the first part.

Given an arbitrary mechanism \( F_{0} \), by Lemma the it follows that there exists a straightforward mechanism \( F \) with signals \( \{s_{i}\} \) whose straightforward equilibrium achieves the same level of en-
engagement and misinformation. Note that $F$ need not be a threshold mechanism itself. For each $i \in V$, we construct the following thresholds $\sigma_i$:

$$\sigma_i = \alpha \mathbb{P}(s_i = 1).$$

We denote by $\hat{F}$ the corresponding threshold mechanism, with signals $\{\hat{s}_i = 1(\epsilon \leq \sigma_i)\}$. We claim that $\hat{F}$ is a straightforward threshold mechanism whose straightforward equilibrium achieves the same level of engagement as $F$ (and hence $F_0$), but yields weakly lower misinformation. In what follows, we first assume that $1 > \mathbb{P}(s_i = 1) > 0$ for all $i$. We then treat the case where $\mathbb{P}(s_i = 1) \in \{0, 1\}$ separately.

Recall that signaling mechanism $F$ is a collection of probability distributions $F(\cdot; z)$ of random variables $s = (s_1, \ldots, s_n)$ for $z \in [0, \alpha]$. Moreover, $F$ under consideration is $\Sigma$-measurable, i.e., $s_i^{-1}(y_i) \in \Sigma$ for any $y_i \in S_i = \{0, 1\}$. For the remainder of this proof, for any $y_i \in S_i$, we define:

$$F_i(y_i; z) := \sum_{y_{-i} \in S_{-i}} F(y_i, y_{-i}; z),$$

where $S_{-i} = \times_{k \neq i} S_k$ denotes the range of $s_{-i}$. Note that $F_i(y_i; z)$ captures the probability that agent $i$ receives signal $y_i \in S_i$ when the error level is $\epsilon = z$.

We start by showing that straightforwardness of $F$ implies that $\hat{F}$ is also straightforward. To this end, we first focus on the case when the platform sends a signal $\hat{s}_i = 1$ to agent $i$, and we establish that when agents $j \neq i$ follow strategies $\hat{\mu}_j(\hat{s}_j) = \hat{s}_j$, the agent $i$ finds it optimal to follow the platform’s signal and engage. Then, we show that when the platform sends a signal $\hat{s}_i = 0$, the agent finds it optimal not to engage. Jointly, these claims imply that the strategy profile $\hat{\mu}_j(\hat{s}_j) = \hat{s}_j$ for all $j \in V$ is a straightforward equilibrium.

Assume that $j \neq i$ follow strategies $\hat{\mu}_j(\hat{s}_j) = \hat{s}_j$, and consider the case where agent $i$ receives signal $\hat{s}_i = 1$. We claim that for agent $i$, the expected value of $\epsilon$ conditioned on $\hat{s}_i = 1$ from the mechanism $\hat{F}$ is weakly smaller than that from the mechanism $F$ under the same recommendation (i.e., conditioned on $s_i = 1$). To establish this result, we make use of two auxiliary lemmata, which are stated next:

**Lemma 12.** Consider a function $g : [0, \alpha] \to [0, 1]$, and let us write

$$A = \int_0^\alpha g(x) dx.$$

Then,

$$\int_0^A x dx \leq \int_0^\alpha x g(x) dx \leq \int_\alpha^{A} x dx. \quad (22)$$
Proof: Both inequalities follow by observing that the middle integral in (22) is mini-
mized/maximized when $g(\cdot)$ is a threshold-type function (i.e., equal to zero below/above a threshold and equal to one otherwise). □

Lemma 13. The functions $F_i(\cdot;z)$ satisfy the following:

\[
\int_0^\alpha F_i(1;z)dz = \sigma_i, \tag{23}
\]
\[
\int_0^\alpha F_i(0;z)dz = \alpha - \sigma_i, \tag{24}
\]
\[
\int_0^\alpha zF_i(1;z)dz = \sigma_i \mathbb{E}[\epsilon | s_i = 1], \tag{25}
\]
\[
\int_0^\alpha zF_i(0;z)dz = (\alpha - \sigma_i) \mathbb{E}[\epsilon | s_i = 0]. \tag{26}
\]

Proof: Recall that $(\Omega, \Sigma, \mathbb{P})$ is such that $\Omega = \Omega_e \times \Omega_r$, $\Sigma$ is the smallest sigma algebra containing $\Sigma_e \times \Sigma_r$, and $\mathbb{P}$ is a product measure between $U[0, \alpha]$, and $\mathbb{P}_r$. Denote each $\omega \in \Omega$ more explicitly by the tuple $\omega = (z, \omega_r)$ where $z \in \Omega_e$, $\omega_r \in \Omega_r$, and observe that $\epsilon(\omega) = \epsilon(z, \omega_r) = z$. Using Fubini’s theorem for any $A \in \Sigma_e$ we have

\[
\mathbb{P}(s_i = 1, \epsilon \in A) = \int 1(s_i(\omega) = 1, \epsilon(\omega) \in A) d\mathbb{P} = \int_\alpha 1 \alpha \int 1(s_i(z, \omega_r) = 1, z \in A) d\mathbb{P}_r dz
\]

\[
= \int_A \frac{1}{\alpha} F_i(1;z)dz. \tag{27}
\]

Rearranging terms and using this identity for $A = [0, \alpha]$ yields

\[
\int_0^\alpha F_i(1;z)dz = \alpha \mathbb{P}(s_i = 1) = \sigma_i. \tag{28}
\]

Repeating the same argument, we also obtain

\[
\mathbb{P}(s_i = 0, \epsilon \in A) = \int_A \frac{1}{\alpha} F_i(0;z)dz \text{ and } \tag{29}
\]
\[
\int_0^\alpha F_i(0;z)dz = \alpha \mathbb{P}(s_i = 0) = \alpha - \sigma_i. \tag{30}
\]

From (28) and (30) equalities in (23) and (24) follow.

Next observe that

\[
\mathbb{E}[\epsilon | s_i = 1] = \frac{\mathbb{E}[\epsilon 1(s_i = 1)]}{\mathbb{P}(s_i = 1)} = \frac{\alpha \mathbb{E}[\epsilon 1(s_i = 1)]}{\sigma_i}. \tag{31}
\]
On the other hand, once again using Fubini’s theorem, we obtain
\[
\mathbb{E}[\epsilon | s_i = 1] = \int \epsilon(\omega) \mathbf{1}(s_i(\omega) = 1) d\mathbb{P} = \int \frac{1}{\alpha} \int \mathbf{1}(s_i(z, \omega_r) = 1) d\mathbb{P}, dz = \int \frac{1}{\alpha} z F_i(1; z) dz.
\]
Together with (31), this implies (25).
Finally, following the same approach, we have
\[
\mathbb{E}[\epsilon | s_i = 0] = \frac{\alpha \mathbb{E}[\epsilon | s_i = 0]}{\mathbb{P}(s_i = 0)} = \frac{\alpha \mathbb{E}[\epsilon | s_i = 0]}{\alpha - \sigma_i},
\]
and also
\[
\mathbb{E}[\epsilon | s_i = 0] = \int \epsilon(\omega) \mathbf{1}(s_i(\omega) = 0) d\mathbb{P} = \int \frac{1}{\alpha} \int z \mathbf{1}(s_i(z, \omega_r) = 0) d\mathbb{P}, dz = \int \frac{1}{\alpha} z F_i(0; z) dz.
\]
Hence, using the last equality together with (32), we also have (26), and the claim follows.

Using these auxiliary lemmata, we are ready to characterize the outcome when the platform sends signal \(s_i = 1\) under mechanism \(\hat{F}\). Observe that we have
\[
\mathbb{E}[\epsilon | \hat{s}_i = 1] = \frac{\sigma_i}{2} = \frac{\int_0^{\sigma_i} z dz}{\sigma_i} \leq \frac{\int_0^{\sigma_i} z F_i(1; z) dz}{\sigma_i} = \frac{\sigma_i \mathbb{E}[s_i = 1]}{\sigma_i} = \mathbb{E}[\epsilon | s_i = 1],
\]
Here, the inequality follows from Lemma 12 and the fact that \(\int_0^{\sigma_i} F_i(1; z) dz = \sigma_i\), as established in Lemma 13. The third equality also follows from Lemma 13.

In addition, for each pair \(i, j \in V\), using (23), we write
\[
\mathbb{P}(s_j = 1 | s_i = 1) = \frac{\mathbb{P}(s_j = 1, s_i = 1)}{\mathbb{P}(s_i = 1)} = \frac{\min\{\mathbb{P}(s_i = 1), \mathbb{P}(s_j = 1)\}}{\sigma_i} \leq \frac{\min\{\int_0^{\sigma_i} F_i(1; z) dz, \int_0^{\sigma_i} F_j(1; z) dz\}}{\sigma_i} = \frac{\min\{\sigma_i, \sigma_j\}}{\sigma_i} = \mathbb{P}(s_j = 1 | \hat{s}_i = 1),
\]
where the inequality follows from the fact that \(\mathbb{P}(s_j = 1, s_i = 1) \leq \mathbb{P}(s_j = 1)\) and \(\mathbb{P}(s_j = 1, s_i = 1) \leq \mathbb{P}(s_i = 1)\), for all \(i, j \in V\).

It follows from (33) and (34) that when the platform sends a signal to engage, the payoff from engaging is higher under \(\hat{F}\) than it is under \(F\). This is the case because under \(\hat{F}\), both the disutility due to error is lower and the network externality due to having connections who engage with the same content is higher.

Next, consider the case where \(\hat{s}_i = 0\), i.e., the platform sends a “Do not engage” signal to agent \(i\) in mechanism \(\hat{F}\). We can express the associated expected error as follows:
\[
\mathbb{E}[\epsilon | \hat{s}_i = 0] = \frac{\alpha + \sigma_i}{2} = \frac{\int_0^{\alpha} \sigma d\sigma}{\alpha - \sigma_i} \geq \frac{\int_0^{\alpha} z F_i(0; z) dz}{\alpha - \sigma_i} = \mathbb{E}[\sigma | s_i = 0].
\]
Here, the inequality follows from Lemma 12 and the fact that \( \int_0^\alpha F_i(1; z)dz = \alpha - \sigma_i \), as established in Lemma 13. The last equality is also a direct consequence of Lemma 13.

Similarly, for each pair \( i, j \in V \), we obtain

\[
\mathbb{P}(s_j = 1 | s_i = 0) = \frac{\mathbb{P}(s_j = 1, s_i = 0)}{\mathbb{P}(s_i = 0)} = \frac{\mathbb{P}(s_j = 1) - \mathbb{P}(s_j = 1, s_i = 1)}{\alpha - \sigma_i} \geq \frac{\sigma_j - \min\{\mathbb{P}(s_j = 1), \mathbb{P}(s_i = 1)\}}{\alpha - \sigma_i} = \sigma_j - \min\{\sigma_i, \sigma_j\} = \frac{\max\{\sigma_i, \sigma_j\} - \sigma_i}{\alpha - \sigma_i} = \mathbb{P}(\hat{s}_j = 1 | \hat{s}_i = 0).
\]

(36)

Here, the inequality follows again from the fact that \( \mathbb{P}(s_j = 1, s_i = 1) \leq \mathbb{P}(s_j = 1) \) and \( \mathbb{P}(s_j = 1, s_i = 1) \leq \mathbb{P}(s_i = 1) \), for all \( i, j \in V \). The fourth equality employs Lemma 13.

It follows from (35) and (36) that when the platform sends a signal to not to engage, the payoff from engaging is lower under \( \hat{F} \) than it is under \( F \). This is the case, because under \( \hat{F} \) both the disutility due to error is larger and the network externality due to having connections who engage with the same content is smaller.

Since under \( \hat{F} \) the payoff due to engagement under the engage (do not engage) signal is larger (lower) than \( F \), and since \( F \) is straightforward, it follows that \( \hat{F} \) is also straightforward. We conclude the proof by establishing that the engagement achieved at a straightforward equilibrium \( \hat{Q} \) of \( \hat{F} \) is the same as that achieved at a straightforward equilibrium \( Q \) of \( F \). However, the misinformation associated with \( \hat{Q} \) is weakly lower than that associated with \( Q \).

To see this, first note that the total engagement level for the original mechanism satisfies

\[
E(Q) = \sum_{i \in V} \mathbb{P}(s_i = 1) = \sum_{i \in V} \frac{\sigma_i}{\alpha} = \sum_{i \in V} \mathbb{P}(\hat{s}_i = 1) = E(\hat{Q}),
\]

where the first and last equalities follow from the definition of engagement and straightforwardness of equilibria \( Q \) and \( \hat{Q} \), the second equality follows from (21) and the third equality follows from the construction of the thresholds.

Second, observe that

\[
M(Q) = \sum_{i \in V} \mathbb{E}[\epsilon | s_i = 1] = \sum_{i \in V} \int_0^\alpha \frac{1}{\sigma_i} F_i(1; z)dz \geq \sum_{i \in V} \frac{1}{\sigma_i} \int_0^{\sigma_i} zdz = \sum_{i \in V} \mathbb{E}[\epsilon | \hat{s}_i = 1] = M(\hat{Q}),
\]

where the first and last equalities follow from the definition of \( M(\cdot) \) and the fact that \( Q \) and \( \hat{Q} \) are straightforward equilibria, the second equality follows from Lemma 13, the third equality follows from the definition of threshold mechanisms, and the inequality follows from Lemma 12
Thus, we conclude that the straightforward equilibrium of the mechanism \( \hat{F} \) yields the same engagement and weakly lower misinformation. Hence, the claim follows.

Finally, observe that if \( P(s_i = 1) \in \{0, 1\} \) for some \( i \), then \( \sigma_i \in \{0, \alpha\} \). Hence, almost surely both \( F \) and \( \hat{F} \) send the same signals to such agents. Since \( F \) has a straightforward equilibrium where these agents find it optimal to follow the signal they receive, in \( \hat{F} \) aforementioned agents find it optimal to follow their signals as well. Thus, applying the argument above to the remaining agents, it follows that \( \hat{F} \) is a straightforward mechanism. In addition, as before, the engagement associated with these agents is the same in both mechanisms and the misinformation is weakly lower. Hence, the claim follows as before.

\( \square \)

**Proof of Lemma 2.** Observe that in both games, the strategies of agents belong to \( \{0, 1\} \), and hence, the strategy sets are compact. Since the domains of the payoff functions are discrete, continuity follows trivially. Moreover, the payoffs exhibit increasing differences, i.e.,

\[
u_i(1, x_{-i}, s) - u_i(0, x_{-i}, s) \leq u_i(1, x_{-i}', s) - u_i(0, x_{-i}', s),
\]

where \( x_{-i}' \geq x_{-i} \) and \( s \in \{0, 1\} \). The fact that the inequality holds can trivially be seen by noting that when \( x_{-i}' \geq x_{-i} \), the network externality term is weakly larger. Thus, it follows that for both \( s = 0 \) and \( s = 1 \), the induced game is supermodular.

\( \square \)

**Proof of Theorem 2.** To prove the result, we first characterize the equilibrium strategies under mechanism \( F \) in terms of the pure strategy equilibria of the associated supermodular games. Then we use this characterization to study the performance levels of \( F \) at different equilibria and their convex hull.

If the platform chooses a threshold \( \sigma \), when \( \epsilon \leq \sigma \), agents play \( G_1 \), and otherwise they play \( G_0 \). Let \( X_0 \) denote the set of pure equilibria of \( G_0 \) and \( X_1 \) denote the set of pure equilibria of \( G_1 \). With some abuse of notation we denote any equilibrium \( X \) of \( X_i \ (i \in \{0, 1\}) \) with a vector in \( \mathbb{R}^n \), where the \( j \)th entry of the vector, \( X_j \), denotes the equilibrium action of agent \( j \). By Lemma 2, it follows that both of these sets admit a largest and a smallest equilibrium (with respect to the usual order in \( \mathbb{R}^n \)). We denote the largest/smallest equilibria in \( X_i \) respectively by \( X_i^\ast \) and \( X_i^\dagger \) for \( i \in \{0, 1\} \). Using the vector representation of equilibria, for any \( X \in X_i \), we have \( X_i^\dagger \leq X \leq X_i^\ast \), where the inequalities are entrywise. We start by expressing any equilibrium \( Q \in Q(F) \) in terms of \( X^0 \in X_0 \) and \( X^1 \in X_1 \).
Lemma 14. Suppose that the platform uses a public signaling mechanism $F$ with threshold $\sigma \in (0, \alpha)$. Denote the equilibrium strategy profile for any $Q \in Q(F)$ by $\mu^Q := \{\mu_i^Q\}$, and note that agents’ equilibrium actions are given by $Y_0 = \mu^Q(0)$ and $Y_1 = \mu^Q(1)$ for signals $s = 0$ and $s = 1$, respectively. Let $\mathcal{Y} = \{(Y_0, Y_1) | Y_0 = \mu^Q(0), Y_1 = \mu^Q(1), Q \in Q(F)\}$. Then, we have:

$$\mathcal{Y} = \mathcal{X} \triangleq \{(X^0, X^1) | X^0 \in \mathcal{X}_0, X^1 \in \mathcal{X}_1\}.$$

Proof: Consider any $Q \in Q(F)$. We claim that $\mu^Q(0)$ is an equilibrium in $\mathcal{G}_0$. Suppose that this is not the case and that agent $i$ can unilaterally deviate to $x_i'$ and strictly improve her payoff in $\mathcal{G}_0$, i.e., $u_i(x_i', \mu_{-i}^Q(0), 0) > u_i(\mu_i^Q(0), \mu_{-i}^Q(0), 0)$. Let $\mu_i'$ denote a strategy of agent $i$ in the original game, such that $\mu_i'(0) = x_i'$ and $\mu_i'(1) = \mu_i^Q(1)$. We have

$$U_i(\mu_i', \mu_{-i}^Q, 0) = u_i(x_i', \mu_{-i}^Q(0), 0) > u_i(\mu_i^Q(0), \mu_{-i}^Q(0), 0) = U_i(\mu_i^Q, \mu_{-i}^Q, 0),$$

and $U_i(\mu_i', \mu_{-i}^Q, 1) = U_i(\mu_i^Q, \mu_{-i}^Q, 1)$. Thus, we obtain a contradiction to the fact that $Q$ is an equilibrium under mechanism $F$. Following the same approach, it also follows that $\mu^Q(1)$ is an equilibrium in $\mathcal{G}_1$. Hence, we conclude that $\mathcal{Y} \subset \mathcal{X}$.

Conversely, consider any $X^0 \in \mathcal{X}_0, X^1 \in \mathcal{X}_1$. Let $\mu$ be the strategy profile that satisfies $\mu(0) = X^0$, and $\mu(1) = X^1$. Suppose that this is not an equilibrium of the original game under mechanism $F$ and that agent $i$ strictly improves her payoff by deviating to strategy $\mu_i'$. Then, agent $i$ strictly improves her expected payoff under signal $s = 0$ or $s = 1$ by using $\mu_i'$, i.e., $U_i(\mu_i', \mu_{-i}, s) > U_i(\mu_i, \mu_{-i}, s)$. Consider the case $s = 0$; then deviating from $X^0$, using $x_i' = \mu_i'(s)$ strictly improves the payoff of agent $i$ in $\mathcal{G}_0$, contradicting the assumption that $X^0$ is an equilibrium. The case of $s = 1$ similarly contradicts the assumption that $X^1$ is an equilibrium. Thus, it follows that $\mathcal{Y} \subset \mathcal{X}$. Since we also have $\mathcal{Y} \subset \mathcal{X}$, the claim follows. \qed

Having characterized the structure of the set of equilibria that are induced by public signaling mechanisms, we now proceed to analyze the structure of the performance levels that are achievable by these mechanisms. Consider a signaling mechanism $F$ and an equilibrium $Q \in Q(F)$. Recall that the performance of $Q$ is given by $P(Q) = (E(Q), M(Q))$, and performance levels achievable at an equilibrium of $F$ are denoted by $\mathcal{P}(F) = \{P(Q) | Q \in Q(F)\}$.

Suppose that $\sigma \in (0, \alpha)$. As established in Lemma 14, the equilibrium actions of agents in a public signal mechanism $F$ are given as product sets of equilibria of corresponding supermodular games $\mathcal{G}_0$ and $\mathcal{G}_1$. We next establish that $\mathcal{P}(F)$ is contained in the convex hull of four equilibria associated with these supermodular games. Before proving our result, we define these equilibria:
Lemma 15. Suppose that \( \sigma \in (0, \alpha) \), then

\[
\mathcal{P}(F) \subset \mathcal{C}(F) = \mathcal{C}'(F) \triangleq \text{conv}\{ P(Q^{(B,B)}), P(Q^{(B,W)}), P(Q^{(W,B)}), P(Q^{(W,W)}) \}.
\]

Proof: To prove the claim, it suffices to establish that \( \mathcal{C}(F) = \mathcal{C}'(F) \). From the definition of \( \mathcal{C}(F) \), it can be seen that this is equivalent to having \( P(Q) \in \mathcal{C}'(F) \) for all \( Q \in \mathcal{Q}(F) \).

Consider \( Q \in \mathcal{Q}(F) \), and using Lemma 14, let \( (X^0, X^1) \in \mathcal{X}_0 \times \mathcal{X}_1 \) denote the corresponding actions of agents under signals \( s = 0 \) and \( s = 1 \), respectively. Using these sets, engagement and misinformation associated with \( Q \) can be explicitly given as follows:

\[
E(Q) = p_1|X^1| + p_0|X^0|
\]

\[
M(Q) = p_1|X^1|\frac{\sigma}{2} + p_0|X^0|\frac{\alpha + \sigma}{2},
\]

where \( p_1 \triangleq \mathbb{P}(s = 1) \), \( p_0 \triangleq 1 - p_1 \), and with abuse of notation, we denote by \( |X^k| \) the number of nonzero entries of \( X^k \in \{0,1\}^n \) for \( k \in \{0,1\} \).

Suppose that \( P(Q) \) does not belong to \( \mathcal{C}'(F) \). Then, by the separating hyperplane theorem, there exist scalars \( \gamma_e, \gamma_m \in \mathbb{R} \) such that

\[
\gamma_e E(Q) - \gamma_m M(Q) > \gamma_e E' - \gamma_m M'
\]

for any \( (E', M') \in \mathcal{C}'(F) \). Observe that

\[
\gamma_e E(Q) - \gamma_m M(Q) = |X^1|\left(\gamma_e p_1 - \gamma_m p_1 \frac{\sigma}{2}\right) + |X^0|\left(\gamma_e p_0 - \gamma_m p_0 \frac{\alpha + \sigma}{2}\right).
\]
Let $Y^1 = X^1$ if $(\gamma_e p_1 - \gamma_m p_1 \frac{\sigma}{2}) \geq 0$, and $Y^1 = X^1$ otherwise. Similarly, let $Y^0 = X^0$ if $(\gamma_e p_0 - \gamma_m p_0 \frac{\alpha + \sigma}{2}) \geq 0$, and $Y^0 = X^0$. Observe that by Lemma 14 we have $X^1 \leq Y^1 \leq X^1$ and, similarly, $X^0 \leq Y^0 \leq X^0$. Thus, it follows that

$$\gamma_e E(Q) - \gamma_m M(Q) \leq |Y^1|((\gamma_e p_1 - \gamma_m p_1 \frac{\sigma}{2}) + |Y^0|((\gamma_e p_0 - \gamma_m p_0 \frac{\alpha + \sigma}{2})).$$

On the other hand, by construction, $(Y^0, Y^1)$ is the equilibrium actions for one of the equilibria $Q^{B,B}$, $Q^{B,W}$, $Q^{W,B}$, $Q^{W,W}$. Thus, we obtain a contradiction to (37), and the claim follows. □

Lemma 15 characterizes equilibrium performance levels for $\sigma \in (0, \alpha)$. Note that if $\sigma \in \{0, \alpha\}$ agents almost surely receive the signal $s = 0$, or $s = 1$. Suppose $\sigma = 0$, and agents almost surely receive signal $s = 0$. In this case, equilibrium performance levels coincide with the equilibria of $\mathcal{G}_0$. Thus, it can be seen that $Q^{W,W}$, $Q^{B,W}$, $Q^{B,B}$, $Q^{W,W}$ remain to be equilibria for public signal mechanisms with $\sigma = 0$. Moreover, in this case we have $P(Q^{W,W}) = P(Q^{W,B})$, and $P(Q^{B,W}) = P(Q^{B,B})$. Finally, at all pure equilibria, a subset of agents always engage and another subset never engages. Hence, equilibrium performance levels belong to the convex hull of the best/worst equilibria of $\mathcal{G}_0$. These observations imply that equilibrium performance levels once again belong to $C'(F)$, i.e., $C(F) = C'(F)$. Recalling that when $\sigma = \alpha$ agents almost surely receive signal $s = 1$, and following a similar argument, it follows that in this case as well $C(F) = C'(F)$.

Thus, for any $\sigma \in [0, \alpha]$, equilibrium performance levels for mechanism $F$, denoted by $P(F)$, belong to the convex hull of performance levels associated with equilibria: $Q^{B,B}$, $Q^{B,W}$, $Q^{W,B}$, $Q^{W,W}$. To finish the proof of Theorem 2 we explicitly characterize the performances associated with $Q^{W,W}$, $Q^{W,B}$, $Q^{B,W}$, $Q^{B,B}$.

(i) First, we focus on the case where $\nu \leq b\sigma/2$. Consider $\mathcal{G}_0$, and assume that no neighbors of some agent $i$ engage with the content. Then, agent $i$’s payoff from engaging is given by $\nu - b^2z/2 < \nu - b\sigma/2 \leq 0$, and she finds it optimal not to engage. Since $i$ is arbitrary, having no agents engage is an equilibrium. Given that the actions of each agent belong to $\{0, 1\}$, we conclude that the worst equilibrium $X^0$ of $\mathcal{G}_0$ is such that no agents engage.

Similarly, consider $\mathcal{G}_1$, and assume that no neighbors of agent $i$ engage with the content. Then, agent $i$’s payoff from engaging is given by $\nu - b\sigma/2 \leq 0$. In this case as well, the agent finds it optimal not to engage. As before, having no agents engage is an equilibrium. This constitutes the worst equilibrium $X^1$ of $\mathcal{G}_1$.

Having characterized the worst equilibria of $\mathcal{G}_0$ and $\mathcal{G}_1$, we next focus on the best equilibria. First consider $\mathcal{G}_1$. We claim that it is an equilibrium to have all agents in $A_1$ engage and agents in $A_1^c$ not engage. To see this, note that for the described strategy profile for $i \in A_1$, the payoff from engaging is given by $\nu - b\sigma/2 + \sum_{j \in A_1} g_{ij} \geq \nu - b\sigma/2 + k_1 \geq 0$, where the first equality follows from the definition
of $A_1$, and the last one follows from the definition of $k_1$. Thus, agent $i$ finds it optimal to engage. Conversely, for $i \notin A_1$ the payoff from sharing is given by $v - b_2^s + \sum_{j \in A_1} g_{ij} \leq v - b_2^s + (k_1 - 1) < 0$, where the first inequality follows since $i \notin A_1$ implies that $i$ has $k_1 - 1$ or fewer neighbors in $A_1$. The second inequality follows from the definition of $k_1$. Thus, we conclude that agent $i \notin A_1$ finds it optimal not to engage. Hence, it indeed is an equilibrium to have only the agents in $A_1$ engage.

We next argue that this is the best equilibrium of $G_1$. However, suppose that this is not true, in which case there is an equilibrium where agents in $A \supset A_1$ engage. Let $j \in A \setminus A_1$ be an agent for whom $\sum_{k \in A} g_{jk}$ is minimized among all agents in $A \setminus A_1$. Observe that $\sum_{k \in A} g_{jk} \leq k_1 - 1$, as otherwise we obtain a contradiction to the fact that $A_1$ is the maximal induced subgraph where agents have at least $k_1$ neighbors. Observe that the payoff of agent $j$ is given by $v - b_2^s + \sum_{k \in A} g_{jk} \leq v - b_2^s + (k_1 - 1) < 0$. Thus, agent $j$ is better off not engaging and hence, we obtain a contradiction. We conclude that at the best equilibrium of $G_1$, only agents in $A_1$ engage.

The best equilibrium of $G_0$ admits an identical characterization after replacing $-b_2^s$ with $-b_2^{\sigma + \alpha}$ in the above argument (and using $A_0$ in place of $A_1$), accounting for the fact that in $G_0$, the misinformation penalty term in the payoff is $-b_2^{\sigma + \alpha}$, as opposed to $-b_2^s$. We conclude that the best equilibrium of $G_0$ is such that only agents in $A_0$ engage.

Given best/worst equilibria of $G_0$ and $G_1$, we next characterize the engagement and misinformation associated with $Q^{(B,B)}, Q^{(B,W)}, Q^{(W,B)}, Q^{(W,W)}$. Observe that the platform sends the signal $s = 1$ with probability $\sigma/\alpha$ and the signal $s = 0$ with probability $(\alpha - \sigma)/\alpha$. Using this, together with the characterization of best/worst equilibria given above, we obtain the following:

- $Q^{(W,W)}$: Associated equilibrium actions are $X_0 = 0$ and $X_1 = 0$ for $s = 0$ and $s = 1$, respectively. Thus, we have $E(Q^{(W,W)}) = 0$ and $M(Q^{(W,W)}) = 0$.

- $Q^{(W,B)}$: Associated equilibrium actions are $X_0 = 0$ and $X_1 = e_{A_1}$ for $s = 0$ and $s = 1$, where $e_{A_1}$ is a vector that has one for entries $i \in A_1$ and zero otherwise. Thus, we have $E(Q^{(W,B)}) = \frac{\alpha}{\alpha} |A_1|$ and $M(Q^{(W,B)}) = \frac{\alpha^2}{2\alpha} |A_1|$. $Q^{(B,W)}$: Associated equilibrium actions are $X_0 = e_{A_0}$ and $X_1 = 0$ for $s = 0$ and $s = 1$, where $e_{A_0}$ has one for entries $i \in A_0$ and zero otherwise. Thus, we have $E(Q^{(B,W)}) = \frac{\alpha - \sigma}{\alpha} |A_0|$ and $M(Q^{(B,W)}) = \frac{\alpha^2 - \sigma^2}{2\alpha} |A_0|$.

- $Q^{(B,B)}$: Associated equilibrium actions are $X_0 = e_{A_0}$ and $X_1 = e_{A_1}$. Thus, we have $E(Q^{(B,B)}) = \frac{\sigma}{|A_1|} + \frac{\alpha - \sigma}{\alpha} |A_0|$ and $M(Q^{(B,B)}) = \frac{\alpha^2}{2\alpha} |A_1| + \frac{\alpha^2 - \sigma^2}{2\alpha} |A_0|$.

Observe that when $0 < \sigma < \alpha$, these points are not collinear provided that $A_0, A_1 \neq \emptyset$. This can be seen by noting that $\frac{M(Q^{(B,B)})}{E(Q^{(B,B)})} = \frac{\sigma}{2}$, whereas $\frac{M(Q^{(W,W)})}{E(Q^{(W,W)})} = \frac{\alpha + \sigma}{2}$. Thus, the region covered by this equilibria has nontrivial area.

(ii) Next, we consider the case with $\frac{\alpha \sigma}{2} < v \leq \frac{\beta (\alpha + \sigma)}{2}$. Note that $v > \frac{\beta a}{2}$ immediately implies that in $G_1$, agents always have nonnegative payoff from engaging, even in the absence of network externalities. Thus, $X^1 = X^1 = e$, where $e$ denotes the vector of ones.
On the other hand, in $G_0$, the payoffs from engaging are given by $v - b \frac{x + \alpha}{2} + \sum_j g_{ij} x_j$. Thus, in the absence of network effects, we have $v - b \frac{x + \alpha}{2} \leq 0$. Hence, in the worst equilibrium, nobody engages and we obtain $X_0 = 0$. Also, following the same argument as before, in the best equilibrium, only agents in $A_0$ engage, i.e., $X_0 = e_{A_0}$. Summarizing, when $\frac{b\alpha}{2} < v \leq \frac{b(\alpha + \alpha)}{2}$ we have the following:

- $Q^{(W,W)} = Q^{(W,B)}$: Associated equilibrium actions are $X^0 = 0$ and $X^1 = e$ respectively for $s = 0$ and $s = 1$. Thus, we have $E(Q^{(W,W)}) = E(Q^{(W,B)}) = |V|^\frac{\alpha}{2}$ and $M(Q^{(W,W)}) = M(Q^{(W,B)}) = |V|^\frac{\sigma^2}{2\alpha}$.

- $Q^{(B,W)} = Q^{(B,B)}$: Associated equilibrium actions are $X^0 = e_{A_0}$ and $X^1 = e$ respectively for $s = 0$ and $s = 1$. Thus, we have $E(Q^{(B,W)}) = E(Q^{(B,B)}) = |V|^\frac{\alpha}{2} + |A_0|\frac{\alpha - \alpha}{\alpha}$ and $M(Q^{(B,W)}) = M(Q^{(B,B)}) = |V|^\frac{\sigma^2}{2\alpha} + |A_0|\frac{\alpha - \alpha}{\alpha}$. □

(iii) Suppose next that $v > \frac{b(\alpha + \alpha)}{2}$. As in case (ii), since $v > \frac{b\alpha}{2}$, in $G_1$, agents always have incentive to engage and hence we have $X^0 = X^1 = e$. Consider $G_0$. Since $v > \frac{b(\alpha + \alpha)}{2}$, similar to $G_1$, all agents always have incentive to engage even in the absence of any neighbors who engage. Thus, we also have $X^0 = X^0 = e$. Hence, in this case, we have $Q = Q^{(W,W)} = Q^{(W,B)} = Q^{(B,W)} = Q^{(B,B)}$ and $E(Q) = |V|$, $M(Q) = |V|^\frac{\sigma^2}{\alpha} + |V|^\frac{\alpha - \alpha}{\alpha} = |V|^\frac{\sigma^2}{2\alpha}$. □

Details of Example 1. When $\sigma = 3$, the characterization in (i) applies. We have $k_0 = 5$, $k_1 = 1$, and thus, $A_0 = V_L$, whereas $A_1 = V_L \cup V_R$. As given in Theorem 2 in this case, the convex hull of equilibrium performance levels ($C(F)$) has nontrivial area. This corresponds to the shaded region in the figure. On the other hand, for $\sigma = 4$, while the characterization is still given by Theorem 2(i), we have $k_0 = 6, k_1 = 2$. Thus, $A_0 = \emptyset$. Hence, $C(F)$ collapses to the line segment in the figure associated with $\sigma = 4$. Similarly, for $\sigma = 1$, we have another line segment in Figure 2(a). This one is characterized by Theorem 2(ii). □

B. Proofs and Examples of Section 4

B.1. Proofs

Proof of Lemma 3. Suppose $\sigma_i = \alpha$ for all $i \in V$. Observe that for this choice of $\{\sigma_i\}$, (10) and (11) trivially hold, whereas (9) holds since the lemma assumes $b \leq \frac{2}{\alpha} \left(v + d_{\min}\right)$. Thus, the constructed solution is feasible in the engagement maximization problem. The corresponding engagement is $\frac{1}{\alpha} \sum_{i \in V} \sigma_i = n$. The optimality of the constructed mechanism follows since $n$ is also the largest achievable engagement. □

Proof of Lemma 4. (i) It is straightforward to verify that if $\sigma_i = \alpha$ for all $i \in V$, constraint (9) is violated under the assumptions of the lemma for any minimum degree node.

(ii) Follows immediately from (i).
(iii) Assume that $(\hat{\sigma}_i)_{i \in V}$ is an optimal solution of the relaxed problem (where (11) is not imposed). To prove the claim, it suffices to establish that this solution always satisfies (11). For the purposes of contradiction, assume that this is not the case. Then, either there exists some $i \in V$ such that $\hat{\sigma}_i > \alpha$ or there exists some $i^* \in V$ such that $\hat{\sigma}_{i^*} < 0$.

First suppose that there exists some $i \in V$ for which $\hat{\sigma}_i > \alpha$. Then, from constraint (10), we obtain

$$\alpha^2 - \hat{\sigma}_i^2 \geq \frac{2}{b} v(\alpha - \hat{\sigma}_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\hat{\sigma}_i, \hat{\sigma}_j\} - \frac{2}{b} d_i \hat{\sigma}_i \geq \frac{2}{b} v(\alpha - \hat{\sigma}_i),$$

where the last inequality follows from $\max\{\hat{\sigma}_i, \hat{\sigma}_j\} \geq \hat{\sigma}_i$. Therefore, since $\hat{\sigma}_i > \alpha$, we get

$$\hat{\sigma}_i \leq \frac{2v}{b} - \alpha \leq 0,$$

where the last inequality follows from the assumptions of the lemma. Thus, it follows that $\hat{\sigma}_i \leq \alpha$ for all $i \in V$.

Next, assume that there exists $i^* \in V$ for which $\hat{\sigma}_{i^*} < 0$. Then, using the relation $\min\{\hat{\sigma}_{i^*}, \hat{\sigma}_j\} \leq \hat{\sigma}_{i^*}$ in (9) and canceling out common terms, we obtain $\hat{\sigma}_{i^*} \geq \frac{2v}{b} + \frac{2}{b} \sum_{j \in V} g_{i^*j} \geq \frac{2v}{b} + \frac{2}{b} d_{i^*} \geq 0$. Hence, we reach a contradiction, and it follows that $\hat{\sigma}_i \geq 0$ for all $i \in V$.

We conclude that even after (11) is relaxed, the optimal solution satisfies this constraint. Hence, the set of optimal solutions of the engagement maximization problem and the following problem (obtained after relaxing (11)) coincide:

$$\max_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \sigma_i,$$

$$s.t. \quad f_i(\sigma) \leq 0,$$

$$g_i(\sigma) \leq 0,$$

where

$$f_i(\sigma) = \sigma_i^2 - \frac{2}{b} v \sigma_i - \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\},$$

$$g_i(\sigma) = -\alpha^2 + \sigma_i^2 + \frac{2}{b} v(\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b} d_i \sigma_i.$$

(iv) We first establish that (R) has a unique solution. For the purposes of contradiction, assume that there are two solutions, $\sigma$ and $\hat{\sigma}$ such that $\sigma_i \neq \hat{\sigma}_i$ for some $i \in V$. Note that since (R) is a
concave maximization problem, its optimal set is convex. Therefore, the point

\[ \sigma^m = \frac{\sigma + \hat{\sigma}}{2} \]

is also optimal. Note that the functions \( f_i(\sigma) \) and \( g_i(\sigma) \) can both be expressed as the sum of a strictly convex term \( (\sigma_i^2) \) and convex functions of \( \{\sigma_j\} \). Therefore, using the feasibility conditions

\[ f_i(\sigma) \leq 0, \quad f_i(\hat{\sigma}) \leq 0, \quad g_i(\sigma) \leq 0, \quad g_i(\hat{\sigma}) \leq 0, \]

we obtain

\[ f_i(\sigma^m) < 0, \quad g_i(\sigma^m) < 0. \]

Hence, \( \sigma^m \) is an optimal solution for which the constraints are not binding. Therefore, for small enough \( \delta \), \( (\sigma^m + \delta)_{i \in V} \) is also a feasible solution for which the objective is strictly larger than that of \( \sigma^m \), contradicting the optimality of \( \sigma^m \). Hence, it follows that \( (R) \) has a unique solution. Together with (iii), this implies that the engagement maximization problem also has a unique solution. □

**Proof of Lemma 5.** (i) We use the identity

\[ \max\{\sigma_i, \sigma_j\} = \sigma_i + \sigma_j - \min\{\sigma_i, \sigma_j\} \]

to rewrite (10) as

\[ \sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\} + \alpha^2 - \frac{2}{b} v \alpha - \frac{2}{b} \sum_{j \in V} g_{ij} \sigma_j. \]

Note that \( \sigma_j \leq \alpha \), and therefore,

\[ \alpha^2 - \frac{2}{b} v \alpha - \frac{2}{b} \sum_{j \in V} g_{ij} \sigma_j \geq \alpha \left[ \alpha - \frac{2}{b} v - \frac{2}{b} \sum_{j \in V} g_{ij} \right] \geq \alpha \left[ \alpha - \frac{2}{b} v - \frac{2}{b} d_{\text{max}} \right] > 0, \]

where the last inequality follows from the assumptions of the lemma. Therefore, by (9)

\[ \sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\} < \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\} + \alpha^2 - \frac{2}{b} v \alpha - \frac{2}{b} \sum_{j \in V} g_{ij} \sigma_j, \]

and hence, constraint (10) is not binding.
(ii) Assume that the optimal thresholds \{\sigma_i\} are such that \sigma_i = \alpha for some \(i \in V\). Then, (9) implies that
\[
\alpha^2 \leq \frac{2}{b}v\alpha + \frac{2}{b}\sum_{j \in V} g_{ij}\alpha \leq \frac{2}{b}v\alpha + \frac{2}{b}d_{\max}\alpha,
\]
which contradicts with the assumption of the lemma. Hence, \(\sigma_j < \alpha\) for all \(j\).

Assume next that \(\sigma_{i^*} = 0\) for some \(i^* \in V\). We claim that for small enough \(\delta > 0\), the solution \(\tilde{\sigma}_i = \sigma_i + 1(i = i^*)\delta\) is feasible and yields strictly larger engagement, contradicting the optimality of \{\sigma_i\}.

From part (i) the constraints (10) are not binding, and therefore, we do not need to check them for feasibility for small enough \(\delta\). Constraints (11) are trivially satisfied by the feasibility of \{\sigma_i\} and the fact that \(\sigma_{i^*} = 0\). Furthermore, the constraints (9) are satisfied for all \(i \in V \setminus \{\{i^*\} \cup N_{i^*}\}\). This is because for such \(i\) and any \(k \in \{i\} \cup N_i\), we have \(\tilde{\sigma}_k = \sigma_k\). Hence, the constraints (9) corresponding to such \(i\) remain intact after the update of thresholds to \{\tilde{\sigma}_j\}. For all \(i \in N_{i^*}\), the right hand side of the constraint (9) (weakly) increases, and therefore, the constraint is satisfied. We conclude by verifying that (9) holds for \(i^*\). Indeed, for small enough \(\delta\), we have
\[
\delta^2 \leq \frac{2}{b}v\delta \leq \frac{2}{b}v\delta + \frac{2}{b}\sum_{j \in V} g_{ij}\min\{\delta, \tilde{\sigma}_j\},
\]
and hence, the constraint continues to hold under \{\tilde{\sigma}_j\}. Thus, we conclude that the constructed solution is feasible. It can be readily seen that it leads to a strictly higher objective than \{\sigma_j\}, contradicting its optimality. Hence, we have \(\sigma_j > 0\) for all \(j \in V\).

(iii) From parts (i) and (ii), we conclude that only (9) can be binding at the optimal solution. Assume that for some agent \(i \in V\), this constraint is not binding, i.e.,
\[
\sigma_i^2 < \frac{2}{b}v\sigma_i + \frac{2}{b}\sum_{j \in V} g_{ij}\min\{\sigma_i, \sigma_j\}.
\]

Then, \(\sigma_i\) can be increased by some \(\delta > 0\) without violating any constraint in the family (9), as the relevant constraint is not binding for \(i\), and for \(i \neq j\), increasing \(\sigma_i\) relaxes the aforementioned constraint. Since the remaining constraints are also not binding at the optimal solution, it follows that by increasing \(\sigma_i\), a new feasible solution with higher engagement is obtained. This contradicts optimality of \{\sigma_i\}; hence, we conclude that \(\sigma_i^2 = \frac{2}{b}v\sigma_i + \frac{2}{b}\sum_{j \in V} g_{ij}\min\{\sigma_i, \sigma_j\}\) for all \(i \in V\). □

**Proof of Proposition 2** (i) Suppose (12) has a solution where \(\sigma_j > 0\) for all \(j \in V\), and let \(\sigma_i \leq \sigma_j\) for all \(i\). Since \(\min\{\sigma_i, \sigma_j\} = \sigma_i\) for all \(j\) and \(\sigma_i > 0\), (12) implies that \(\sigma_i = 2(v + d_{\min})/b\). Hence, to
study solutions of (12) where $\sigma_j \in (0, \alpha]$ for all $j \in V$, it suffices to restrict attention to

$$X = \left[\frac{2(v + d_{\min})}{b}, \alpha\right]^n.$$  

We next show that (12) indeed has a unique solution in $X$, which in turn is its unique solution in $(0, \alpha]^n$. Using simple algebra and restricting attention to positive solutions, we rewrite (12) as follows:

$$\sigma_i = \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}}. \quad (38)$$

Therefore, all positive solutions of (12) are fixed points of the operator, $T(\sigma) = (T_i(\sigma))_{i \in V}$, where

$$T_i(\sigma) = \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}}. \quad (39)$$

We next establish that for $\sigma \in X$, we have $T(\sigma) \in X$. Note that this is equivalent to establishing $T_i(\sigma) \leq \alpha$ and $T_i(\sigma) \geq \frac{2(v + d_{\min})}{b}$ for any $\sigma \in X$ and $i \in V$.

We first prove that $T_i(\sigma) \leq \alpha$ for any $\sigma \in X$ and $i \in V$. To that end, first observe that

$$T_i(\sigma) \leq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j \in V} g_{ij} \alpha} \leq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \alpha d_{\max}}.$$  

Thus, to show that $T_i(\sigma) \leq \alpha$, it suffices to establish that $\frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \alpha d_{\max}} \leq \alpha$. Since $\alpha \geq v/b$, the former inequality is true, if and only if $\sqrt{\frac{v^2}{b^2} + \frac{2}{b} \alpha d_{\max}} \leq \alpha - \frac{v}{b}$, which after taking the square of both sides is equivalent to the following inequality:

$$\frac{v^2}{b^2} + \frac{2}{b} \alpha d_{\max} \leq \alpha^2 + \frac{v^2}{b^2} - 2 \alpha \frac{v}{b}.$$  

Cancelling out common terms and rearranging, this is equivalent to

$$\frac{2}{b} \alpha (d_{\max} + v) \leq \alpha^2.$$  

Note that by the assumptions of the proposition, this inequality trivially holds, and hence, we conclude that $T_i(\sigma) \leq \alpha$ for any $\sigma \in X$ and $i \in V$.  


Next, we prove that \( T_i(\sigma) \geq \frac{2(v+d_{\min})}{b} \) for any \( \sigma \in X \) and \( i \in V \). Observe that

\[
T_i(\sigma) \geq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j \in V} g_{ij} \left( \frac{2(v+d_{\min})}{b} \right) \geq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j \in V} g_{ij} \left( \frac{2(v+d_{\min})}{b} \right) \geq \frac{2}{b} \left( \frac{v}{b} + \frac{d_{\min}}{b} \right) = \frac{2(v+d_{\min})}{b} .}
\]

(40)

Hence, we obtain \( T_i(\sigma) \geq \frac{2(v+d_{\min})}{b} \), as claimed. Thus, after restricting the domain of \( T(\cdot) \) to \( X \), we see that its image is also contained in \( X \), i.e., \( T : X \to X \).

We use \( S \) to denote the set of fixed points of this operator:

\[
S = \{ \sigma \in X : T(\sigma) = \sigma \}.
\]

Let \( \geq \) denote the natural partial order in \( X \), i.e., for \( \hat{\sigma}, \sigma' \in X \), we say \( \hat{\sigma} \geq \sigma' \) if \( \hat{\sigma}_i \geq \sigma'_i \) for all \( i \in V \). Moreover, we denote by \( \hat{\sigma} > \sigma' \) the relations \( \hat{\sigma} \geq \sigma' \) and \( \hat{\sigma} \neq \sigma' \).

Since \( X \) is a compact subset of \( \mathbb{R}^n \), it follows that \((X, \geq)\) constitutes a complete lattice. It follows from (39) that when \( \hat{\sigma} \geq \sigma' \), we have \( T(\hat{\sigma}) \geq T(\sigma') \), i.e., \( T(\cdot) \) is an increasing operator on \((X, \geq)\). Using Tarski’s fixed point theorem (see, e.g., Topkis (2011)), we conclude that \( S \) is a (nonempty) complete lattice. We next establish that \( S \) is a singleton.

For the purposes of contradiction, assume that \( \hat{\sigma}, \tilde{\sigma} \in S \), where \( \hat{\sigma} \) is the smallest point of the complete lattice \( S \), and \( \hat{\sigma} > \tilde{\sigma} \). Without loss of generality, we index nodes in increasing order of \( \hat{\sigma}_i \), i.e., \( \hat{\sigma}_1 \leq \hat{\sigma}_2 \leq \ldots \leq \hat{\sigma}_n \), and prove by induction that \( \hat{\sigma}_i = \tilde{\sigma}_i \) for \( i = 1, \ldots, n \). Let us start with \( i = 1 \) and observe that

\[
\left( \hat{\sigma}_1 - \frac{v}{b} \right)^2 - \left( \tilde{\sigma}_1 - \frac{v}{b} \right)^2 = \left( T_1(\hat{\sigma}) - \frac{v}{b} \right)^2 - \left( T_1(\tilde{\sigma}) - \frac{v}{b} \right)^2 = \frac{2}{b} \sum_{j \in V} g_{ij} \left( \min\{ \hat{\sigma}_1, \tilde{\sigma}_j \} - \min\{ \hat{\sigma}_1, \tilde{\sigma}_j \} \right) \leq \frac{2}{b} \sum_{j \in V} g_{ij} (\tilde{\sigma}_1 - \hat{\sigma}_1),
\]

(41)

where the inequality follows from the fact that \( \hat{\sigma}_1 \geq \min\{ \hat{\sigma}_1, \tilde{\sigma}_j \} \) and \( \hat{\sigma}_1 = \min\{ \hat{\sigma}_1, \tilde{\sigma}_j \} \). Therefore,

\[
(\hat{\sigma}_1 - \tilde{\sigma}_1) \left( \hat{\sigma}_1 + \tilde{\sigma}_1 - \frac{2v}{b} \right) \leq \frac{2}{b} \sum_{j \in V} g_{ij} (\tilde{\sigma}_1 - \hat{\sigma}_1).
\]

If \( \tilde{\sigma}_1 \neq \hat{\sigma}_1 \), it has to be the case that \( \hat{\sigma}_1 > \tilde{\sigma}_1 \) (since, by construction, \( \hat{\sigma} \) is larger), and therefore,

\[
\hat{\sigma}_1 + \tilde{\sigma}_1 - \frac{2v}{b} \leq \frac{2d_1}{b}.
\]
Observe that \( \hat{\sigma}_1 \) is the smallest element of \( \hat{\sigma} \), and therefore, by (12), we obtain
\[
\hat{\sigma}_1^2 = \frac{2}{b} v \hat{\sigma}_1 + \frac{2}{b} \sum_{j \in V} g_{ij} \hat{\sigma}_1.
\]

Hence, \( \hat{\sigma}_1 = 2(v + d_1)/b \), and consequently, \( \hat{\sigma}_1 \leq 0 \), which is a contradiction since \( \hat{\sigma} \in S \subseteq X \). Let us proceed with the induction by assuming that for some \( k > 1 \) it is the case that \( \hat{\sigma}_i = \hat{\sigma}_i \) for all \( i \leq k - 1 \). Using \( T_k(\hat{\sigma}) = \hat{\sigma}_k \), we have
\[
\left( \frac{\hat{\sigma}_k - v}{b} \right)^2 = \frac{v^2}{b^2} + \frac{2}{b} \sum_{j < k} g_{kj} \min\{\hat{\sigma}_k, \hat{\sigma}_j\} + \frac{2}{b} \sum_{j \geq k} g_{kj} \min\{\hat{\sigma}_k, \hat{\sigma}_j\}
\leq \frac{v^2}{b^2} + \frac{2}{b} \sum_{j < k} g_{kj} \hat{\sigma}_j + \frac{2}{b} \sum_{j \geq k} g_{kj} \min\{\hat{\sigma}_k, \hat{\sigma}_j\}.
\]
Using \( T_k(\hat{\sigma}) = \hat{\sigma}_k \) we also obtain
\[
\left( \frac{\hat{\sigma}_k - v}{b} \right)^2 = \frac{v^2}{b^2} + \frac{2}{b} \sum_{j < k} g_{kj} \min\{\hat{\sigma}_k, \hat{\sigma}_j\} + \frac{2}{b} \sum_{j \geq k} g_{kj} \min\{\hat{\sigma}_k, \hat{\sigma}_j\}
= \frac{v^2}{b^2} + \frac{2}{b} \sum_{j < k} g_{kj} \hat{\sigma}_j + \frac{2}{b} \sum_{j \geq k} g_{kj} \hat{\sigma}_k,
\]
where the last equality uses the fact that \( \hat{\sigma}_1 \leq \hat{\sigma}_2 \leq \cdots \leq \hat{\sigma}_n \).

Using these observations, we get
\[
(\hat{\sigma}_k - \hat{\sigma}_k) \left( \hat{\sigma}_k + \hat{\sigma}_k - \frac{2v}{b} \right) = \left( \frac{\hat{\sigma}_k - v}{b} \right)^2 - \left( \frac{\hat{\sigma}_k - v}{b} \right)^2
\leq \frac{2}{b} \sum_{j \geq k} g_{kj} \min\{\hat{\sigma}_k, \hat{\sigma}_j\} - \frac{2}{b} \sum_{j \geq k} g_{kj} \hat{\sigma}_k
\leq \frac{2}{b} \sum_{j \geq k} g_{kj} \hat{\sigma}_k - \frac{2}{b} \sum_{j \geq k} g_{kj} \hat{\sigma}_k
= (\hat{\sigma}_k - \hat{\sigma}_k) \frac{2}{b} \sum_{j \geq k} g_{kj},
\]
Here, the first inequality follows from (42) and (43), whereas the second inequality uses the fact that \( \min\{\hat{\sigma}_k, \hat{\sigma}_j\} \leq \hat{\sigma}_k \). Assume that \( \hat{\sigma}_k \neq \hat{\sigma}_k \). Then, since \( \hat{\sigma} > \hat{\sigma} \) it has to be the case that \( \hat{\sigma}_k > \hat{\sigma}_k \).
This in turn implies that the above inequality can be written as
\[
\hat{\sigma}_k + \hat{\sigma}_k - \frac{2v}{b} - \frac{2}{b} \hat{d}_k \leq 0,
\]
where \( \hat{d}_k = \sum_{j \geq k} g_{kj} \).
We claim that
\[ \hat{\sigma}_k \geq \frac{v}{b} + \frac{2}{b}d_k. \]

To see this, note that by equation (43), we have
\[ (\hat{\sigma}_k - \frac{v}{b})^2 = \frac{v^2}{b^2} + \frac{2}{b} \sum_{j<k} g_{kj} \hat{\sigma}_j + \frac{2}{b}d_k \hat{\sigma}_k, \]
which can be equivalently written as
\[ f(\hat{\sigma}_k) = \frac{2}{b} \sum_{j<k} g_{kj} \hat{\sigma}_j, \]  
where
\[ f(x) = (x - \frac{v}{b})^2 - \frac{v^2}{b^2} - \frac{2}{b}d_k x. \]

It is straightforward to verify that \( f(x) \) has a root at 0 and a root at \( \frac{v}{b} + \frac{2}{b}d_k \). By convexity, in between these points, \( f(x) \leq 0 \). Therefore, the unique positive solution to (46) is larger than \( \frac{v}{b} + \frac{2}{b}d_k \), i.e., \( \hat{\sigma}_k \geq \frac{v}{b} + \frac{2}{b}d_k \). Thus, (45) implies that \( \hat{\sigma}_k \leq 0 \). Hence, we obtain a contradiction to the fact that \( \hat{\sigma} \in S \), and it follows that \( \hat{\sigma}_k = \hat{\sigma}_k \).

By induction, we conclude that \( \sigma = \hat{\sigma} \), and hence, \( S \) is a singleton. This in turn implies that (12) has a single solution where \( \sigma_k \in (0, \alpha] \) for all \( k \). The optimality of this solution follows directly from Lemma 5.

(ii) Observe that (13) always admits a unique solution. This is because its left hand side is linear, and right hand side is strictly concave in \( y_i \), and for \( y_i = 0 \) the left hand side is strictly smaller. Moreover, at every visit to S3, the algorithm decreases the cardinality of set \( V \) by one. Thus, it follows that the algorithm terminates in finite time. Without loss of generality, assume that the algorithm determines \( \sigma = \{\sigma_i\}_{i \in V} \) in the following order: \( \sigma_1, \ldots, \sigma_n \).

We first establish that \( \sigma \in X \). Assume not. Then, there are two cases possible: \( \sigma_i > \alpha \) or \( \sigma_i < \frac{2(v + d_{\min})}{b} \) for some \( i \). Denote the smallest \( i \) that satisfies these inequalities with \( i_1 \) and \( i_2 \), respectively. It follows from the algorithm that
\[ \sigma_{i_1} = \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j<i_1} g_{ij} \sigma_j + \frac{2}{b} \sigma_{i_1} \sum_{j \geq i_1} g_{ij}} \leq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sigma_{i_1} \sum_{j \in V} g_{ij}} \leq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sigma_{i_1} d_{\max}}, \]
where we use the fact that when the algorithm sets \( i_1 \), \( a_{i_1} = \frac{v^2}{b^2} + \frac{2}{b} \sum_{j<i_1} g_{ij} \sigma_j \) and \( \sigma_j \leq \alpha < \sigma_{i_1} \) for \( j < i_1 \) by definition of \( i_1 \). On the other hand, rearranging terms, the previous inequality yields
\[ \sigma_{i_1}^2 + \frac{v^2}{b^2} - 2 \sigma_{i_1} \frac{v}{b} \leq \frac{v^2}{b^2} + \frac{2}{b} \sigma_{i_1} d_{\max}. \]
After canceling out the common terms and rearranging, this can be equivalently written as follows:

$$\sigma_{i_1}(\sigma_{i_1} - 2\frac{v}{b}) \leq \frac{2}{b} \sigma_{i_1} d_{\max}.$$ 

This in turn implies $\sigma_{i_1} \leq \frac{2v}{b} + \frac{2}{b} d_{\max} < \alpha$, where the last inequality follows by the assumptions of the proposition. Hence, we reach a contradiction to $\sigma_{i_1} > \alpha$ and conclude $\sigma_{i_1} \leq \alpha$ for all $i$.

Next, we show that $\sigma_{i} \geq \frac{2(v + d_{\min})}{b}$ for all $i \in V$. Observe that we have

$$\sigma_{i_2} = \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j < i_2} g_{ij} \sigma_j + \frac{2}{b} \sigma_{i_2} \sum_{j \geq i_2} g_{ij} \geq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sigma_{i_2} \sum_{j \in V} g_{ij} \geq \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sigma_{i_2} d_{\min}},$$

where we use the fact that when the algorithm sets $i_2$, $a_{i_2} = \frac{v}{b} + \frac{2}{b} \sum_{j < i_2} g_{ij} \sigma_j$ and $\sigma_j \geq \frac{2v + d_{\min}}{b} > \sigma_{i_2}$ for $j < i_2$ by definition of $i_2$. Rearranging terms, the previous inequality yields

$$\sigma_{i_2}^2 + \frac{v^2}{b^2} - 2\frac{v}{b} \frac{v}{b} \geq \frac{v^2}{b^2} + \frac{2}{b} \sigma_{i_2} d_{\min}.$$ 

After canceling out the common terms and rearranging, we obtain

$$\sigma_{i_2} - 2\frac{v}{b} > \frac{2}{b} d_{\min}.$$ 

Hence, we obtain a contradiction to $\sigma_{i_2} < \frac{2}{b} (v + d_{\min})$. Thus, it follows that $\sigma_{i} \geq \frac{2}{b} (v + d_{\min})$ for all $i$. Hence, $\sigma \in X$, as claimed.

We conclude the proof by showing that the constructed $\sigma \in X$ corresponds to the unique solution of (12) where $\sigma_i > 0$ for all $i$. Let $a_{i_k}^t$ denote the value of $a_i$ in the $k$th visit of the algorithm to Step S2 (or prior to the updates in Step S3). Consider the $\sigma$ constructed by the algorithm; we argue that it is a fixed point of $T(\cdot)$, i.e.,

$$\sigma_i = T_i(\sigma) = \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j} g_{ij} \min\{\sigma_i, \sigma_j\}.$$ 

Observe that this condition can be written as follows:

$$\sigma_i = \frac{v}{b} + \sqrt{\frac{v^2}{b^2} + \frac{2}{b} \sum_{j < i} g_{ij} \sigma_j + \frac{2}{b} \sum_{j \geq i} g_{ij} \sigma_i}.$$ 

Consider the $i$th visit of the algorithm to step S3 (where $\sigma_i$ is determined). It can be seen that before the update in Step S3, $a_i^t = \frac{v^2}{b^2} + \frac{2}{b} \sum_{k < i} g_{ik} \sigma_k$. Using this observation (47), can be restated.
\[ \sigma_i = \frac{v}{b} + \sqrt{\frac{a_i^i + 2 \sum_{j \geq i} g_{ij} \sigma_i}{b}}. \]

On the other hand, this condition is trivially satisfied since by construction, \( y_i^* \) solves this equation in Step S2 (in the \( i \)th visit of the algorithm to this step) and \( \sigma_i = y_i^* \). Thus, we conclude that \( T_i(\sigma) = \sigma_i \). Since \( i \) is arbitrary, we conclude that the \( \sigma \) obtained by the algorithm is a fixed point of \( T(\cdot) \). Moreover, as established before, \( \sigma \in X \). Hence, using part (i) we conclude that the algorithm terminates with the unique fixed point of \( T(\cdot) \) in \( X \), which also corresponds to the unique solution of (12) where \( \sigma_i > 0 \) for all \( i \in V \).

(iii) Without loss of generality, again assume that the algorithm determines \( \sigma \) in the following order: \( \sigma_1, \ldots, \sigma_n \). Let \( a_i^k \) and \( a_{i+1}^k \) respectively denote the value of \( a_i \) and \( a_{i+1} \) in \( k \)th visit of the algorithm to Step S2 (or prior to their update in Step S3). Similarly, let \( V^k \) denote the set \( V \) in \( k \)th visit of the algorithm to Step S2.

Observe that by definition, we have

\[ \sigma_i = \frac{v}{b} + \sqrt{\frac{a_i^i + 2 \sum_{j \in V_i} g_{ij} \sigma_i}{b}}. \quad (48) \]

Moreover, using the fact \( a_{i+1}^{i+1} = a_{i+1}^i + \frac{2}{b} \sigma_i g_{i+1,i} \), we obtain

\[ \sigma_{i+1} = \frac{v}{b} + \sqrt{\frac{a_{i+1}^{i+1} + \sum_{j \in V^{i+1}} g_{ij} \sigma_{i+1}}{b}} = \frac{v}{b} + \sqrt{\frac{a_{i+1}^i + \frac{2}{b} \sigma_i + \sum_{j \in V^{i+1}} g_{ij} \sigma_{i+1}}{b}}. \quad (49) \]

Suppose for contradiction that \( \sigma_{i+1} < \sigma_i \). Consider the equation

\[ \frac{v}{b} + \sqrt{\frac{a_{i+1}^i + \frac{2}{b} \sigma_i + \sum_{j \in V^{i}} g_{ij} \sigma_{i+1}}{b}} = y \]

Let \( y^* > 0 \) denote the unique solution of this equation. Observe that \( y^* \geq \sigma_i \), as otherwise, in the \( i \)th visit of the algorithm to Step S2, we get \( y^* = y_{i+1}^* < y_i^* \) and obtain a contradiction to the fact that at step \( i \), we have \( i \in \arg \min_k \in V \), \( y_k^* \). Since we also assumed \( \sigma_{i+1} < \sigma_i \), we have the following relation:

\[ y^* \geq \sigma_i > \sigma_{i+1}. \quad (50) \]

Define the following functions with domains \( \mathbb{R}_+ \):

\[ f(x) := \frac{v}{b} + \sqrt{\frac{a_{i+1}^i + \frac{2}{b} x + \sum_{j \in V^{i}} g_{i+1,j} - x}{b}}. \]
$$g(x) := \frac{v}{b} + \sqrt{\frac{a_{i+1}}{b} \sigma_i + \frac{2}{b} \sum_{j \in V_{i+1}} g_{i+1,j} - x.}$$

Observe that $f(0) > 0$, $f(y^*) = 0$, and $f(x) = 0$ has a unique solution since it is strictly concave and $f(0) > 0$. Using these observations, together with the fact that $y^* \geq \sigma_i$, also implies that $f(y^*) \leq f(\sigma_i)$. Similarly, observe that by (49), we have $g(\sigma_{i+1}) = 0$. Moreover, $g(0) > 0$ and similar reasoning as before yields $g(x) = 0$ has a unique solution. Based on these observations, together with $\sigma_i > \sigma_{i+1}$, we conclude that $0 = g(\sigma_{i+1}) > g(\sigma_i)$. Finally, observe that by construction, $f(\sigma_i) = g(\sigma_i)$.

Collecting all these observations, we conclude that $0 = f(y^*) \leq f(\sigma_i) = g(\sigma_i) < g(\sigma_{i+1}) = 0$.

Thus, we obtain a contradiction and $\sigma_i \leq \sigma_{i+1}$. The claim follows since $i$ is arbitrary.

**Proof of Lemma 6**

(i) Consider $i$ such that $\sigma_i \leq \sigma_j$ for all $j \in N_i$. Writing (12) for such an $i$ (and recalling that $\sigma = \{\sigma_j\}_{j \in V}$ is a solution to (12)), we obtain:

$$\sigma_i^2 = \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \sigma_i.$$ 

Since $\sigma_i > 0$, this implies

$$\sigma_i = \frac{2}{b} (v + d_i),$$

as claimed.

(ii) Follows directly from (i).

(iii) By equation (12), we have

$$\sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} d_i \sigma_i,$$

where the inequality follows after observing $\min \{\sigma_i, \sigma_j\} \leq \sigma_i$ for all $i, j \in V$. Since $\sigma_j > 0$ for all $j$, the claim follows from this inequality.

**Proof of Corollary 1**

Observe that by Lemma 6(iii), $\sigma_k \leq \frac{2}{b} (v + d_k) = \frac{2}{b} (v + d_{\min})$. On the other hand, by Lemma 6(ii), we also have $\sigma_k \geq \frac{2}{b} (v + d_{\min})$. These inequalities imply that $\sigma_k = \frac{2}{b} (v + d_{\min})$. On the other hand, Lemma 6(ii) implies that $\sigma_i \geq \frac{2}{b} (v + d_{\min}) = \sigma_k$ for all $i \in V$, and the claim follows.
Proof of Lemma 7. It can readily checked that $\sigma_i = 2(v + k)/b$ for all $i \in V$ is a solution to (12). Proposition 2 implies that the aforementioned $\{\sigma_i\}$ constitute thresholds of an engagement maximizing mechanism. Hence, the claim follows. □

Proof of Proposition 3. Recall that $\sigma_i^* = \sigma_i \frac{b}{2v}$ for all $i \in V$, where $\{\sigma_i\}_{i \in V}$ is the solution of (12) where each $\sigma_i$ is positive. Equivalently, by Proposition 2 these thresholds constitute an optimal solution of the engagement maximization problem. Before showing the connection between the engagement centrality and Bonacich centrality, we present an algebraic lemma that will be of use.

**Lemma 16.** Let $\{\sigma_i\}_{i \in V}$ denote the optimal solution of (OPT-E). We have $\frac{\sigma_i \sigma_j}{\alpha} \geq \min\{\sigma_i, \sigma_j\} \geq \frac{\sigma_i \sigma_j}{\alpha}$, where $\beta = 2(v + d_{\min})/b$.

**Proof:** Without loss of generality, assume that $\sigma_i \leq \sigma_j$. Then, $\min\{\sigma_i, \sigma_j\} = \sigma_i$. Note that $\sigma_j$ is trivially no larger than $\alpha$, i.e., $\sigma_j \leq \alpha$. Therefore, 

$$\frac{\sigma_i \sigma_j}{\alpha} \leq \sigma_i = \min\{\sigma_i, \sigma_j\}.$$ 

Similarly, by Lemma 6, $\sigma_j \geq \beta$, from which we obtain 

$$\frac{\sigma_i \sigma_j}{\beta} \leq \sigma_i = \min\{\sigma_i, \sigma_j\}.$$ 

Together with the previous inequality this establishes the claim. □

From Equation (12), for any $i \in V$, we have 

$$\sigma_i^2 = \frac{2}{b}v\sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\}.$$ 

By Lemma 16 we obtain 

$$\frac{2}{b}v\sigma_i + \frac{2}{b} \sum_{j} g_{ij} \frac{\sigma_i \sigma_j}{\beta} \geq \sigma_i^2 \geq \frac{2}{b}v\sigma_i + \frac{2}{b} \sum_{j} g_{ij} \frac{\sigma_i \sigma_j}{\alpha}.$$ 

Since by Lemma 6 we have $\sigma_j \geq \beta > 0$ for all $j$, the last inequality can be equivalently written as follows: 

$$\frac{2}{b}v + \frac{2}{b} \sum_{j} g_{ij} \frac{\sigma_j}{\beta} \geq \sigma_i \geq \frac{2}{b}v + \frac{2}{b} \sum_{j} g_{ij} \frac{\sigma_j}{\alpha}.$$ 

Denoting by $\sigma$ the vector of $\{\sigma_i\}$, in vector form, this inequality be expressed as 

$$\frac{2}{b}v\mathbf{1} + \frac{2}{b\beta} G\sigma \geq \sigma \geq \frac{2}{b}v\mathbf{1} + \frac{2}{b\alpha} G\sigma,$$  (51)
Proof of Corollary 2. □
The claim follows.

The maximum engagement is given by 

\[
\max f(1) = \frac{2v}{b} \left( I - \frac{2}{\beta b} G \right)^{-1}
\]

The claim follows after observing that the quantities on the left and right correspond to Bonacich centrality vectors with different parameters. □

Proof of Proposition 4. Observe that if \(2(v + d_{\min})/\alpha \geq b\), Lemma 3 implies that engagement can be maximized using a public signal mechanism with \(\sigma = \alpha\). Observe that this mechanism sends the same signal, regardless of the realization of \(\epsilon\), and is uninformative. Hence, agents’ strategies remain optimal, if the platform were to use instead \(\sigma = 0\) – another mechanism with uninformative signals. Thus, the first part of the claim follows.

Suppose next that \(2(v + d_{\min})/\alpha < b\). Consider maximal equilibria in cases (i) and (ii) of Theorem 2. For \(\alpha \geq \sigma \geq 2v/b\), case (i) of the theorem applies, and it follows that for any given \(\sigma\) satisfying these inequalities, the maximum corresponding engagement is given by \(f_R(\sigma)\). Since 

\[
v/b - \alpha < 0,
\]

Theorem 2 implies that for \(\sigma\) satisfying \(0 \leq \sigma < 2v/b\), the maximum corresponding engagement is given by \(f_L(\sigma)\). Note that when \(\sigma = 2v/b\), \(A_1(\sigma) = V\), and hence, \(f_L(\sigma) = f_R(\sigma)\). Therefore, we conclude that the maximum engagement level for \(\sigma\) satisfying \(0 \leq \sigma \leq 2v/b\) (where both inequalities are weak), can also be given by \(f_L(\sigma)\). Combining these observations, it follows that for \(2v/b \leq \alpha\) the maximum engagement is given by 

\[
\max \left\{ \max_{0 \leq \sigma \leq 2v/b} f_L(\sigma), \max_{2v/b \leq \sigma \leq \alpha} f_R(\sigma) \right\}.
\]

Also note that \(|A_0(\cdot)|\) and \(|A_1(\cdot)|\) and hence \(f_L(\cdot)\) and \(f_R(\cdot)\) are left-continuous by definition. Moreover, \(f_L(\cdot)\) and \(f_R(\cdot)\) are increasing in the intervals where they are continuous (i.e., where \(|A_0(\cdot)|\) and \(|A_1(\cdot)|\) are constant). Thus, the maximum 

\[
\max \left\{ \max_{0 \leq \sigma \leq 2v/b} f_L(\sigma), \max_{2v/b \leq \sigma \leq \alpha} f_R(\sigma) \right\}
\]

is well-defined, and the claim follows. □

Proof of Corollary 2. (i) Follows immediately from Lemma 3.

(ii) In this case, \(k < k_0(0) = \left[ \frac{\alpha}{2} - v \right]\). Hence, \(A_0(0) = \emptyset\), and by Proposition 4, we conclude that under the no intervention mechanism (\(\sigma = 0\)), we obtain an engagement of 

\[f_L(0) = 0.\]

Observe that Lemma 7 implies that setting \(\sigma_i = (v + k)/b\) for all \(i \in V\) maximizes engagement under the assumption of part (ii) of the corollary. Hence, public signal mechanisms are optimal,
and the maximum engagement achieved for the public signal and the optimal signaling mechanisms is given as $2\frac{n}{\alpha} \frac{v+k}{b}$.

**Proof of Corollary 3.** Observe that for the optimal mechanism, the problem decouples over the two disjoint components of the graph. By construction, one component is 2-regular, whereas the other is $2m$-regular. Lemma 7 suggests that the optimal threshold is such that for any $i$ in the first component, we have $\sigma_i = 2(v+2)/b$, and for any $i$ in the second component, we have $\sigma_i = 2(v+2m)/b$. Thus, corresponding overall engagement is given by

$$E_\sigma = \frac{n}{2\alpha} \frac{2(v+2)}{b} + \frac{n}{2\alpha} \frac{2(v+2m)}{b} = 2\frac{n}{\alpha b} (v + m + 1).$$

To characterize the engagement under the optimal public signal mechanism, we next introduce additional notation: Let $V_1$ denote the set of nodes that belong in the ring graph where each node is connected to its immediate neighbor, and $V_2$ denote the set of nodes that belong in the ring graph, where each node is connected to its $2m$ immediate neighbors.

By the assumptions of the corollary, for all $\sigma \geq 0$, we have

$$b \frac{\sigma + \alpha}{2} - v > 2m.$$

Hence, $k_0(\sigma) > 2m$ and $A_0(\sigma) = \emptyset$ for all $\sigma \geq 0$. On the other hand, we have $k_1(\sigma) \leq 2$ and $A_1(\sigma) = V$ for $\sigma \leq 2(v+2)/b$, $2 < k_1(\sigma) \leq 2m$ and $A_1(\sigma) = V_2$ for $2(v+2)/b < \sigma \leq 2(v+2m)/b$, and finally $k_1(\sigma) > 2m$ and $A_1(\sigma) = \emptyset$ for $\sigma > 2(v+2m)/b$. Based on these observations, together with Proposition 4, we characterize engagement levels for different values of $\sigma$, as follows:

(i) if $\sigma \leq 2(v+2)/b$, then $A_0(\sigma) = \emptyset$ and $A_1(\sigma) = V$, and hence, $|A_0(\sigma)| = 0$, while $|A_1(\sigma)| = n$. In this regime, we have $f_L(\sigma) = f_R(\sigma) = \frac{\alpha \sigma}{2\alpha}$, which gives the corresponding engagement level.

(ii) if $2(v+2)/b < \sigma \leq 2(v+2m)/b$, then $A_0(\sigma) = \emptyset$ and $A_1(\sigma) = V_2$, and hence, $|A_0(\sigma)| = 0$ and $|A_1(\sigma)| = n/2$. Since $\sigma > 2v/b$, the corresponding engagement level is given by $f_R(\sigma) = \frac{\alpha \sigma}{2\alpha}$.

(iii) if $2(v+2m)/b < \sigma$, then $A_1(\sigma) = \emptyset$, and therefore, $|A_0(\sigma)| = 0$ and $|A_1(\sigma)| = 0$. Since $\sigma > 2v/b$, the corresponding engagement level is given by $f_R(\sigma) = 0$.

The largest engagement in case (i) is achieved for $\sigma = 2(v+2)/b$ and is given by $\frac{2n(v+2)}{\alpha b}$, and the largest engagement in case (ii) is achieved for $\sigma = 2(v+2m)/b$ and is given by $\frac{n(v+2m)}{\alpha b}$. In case (iii), the engagement is always equal to zero. This observation readily implies that

$$E_P = n \cdot \max \left\{ \frac{2(v+2)}{\alpha b}, \frac{v+2m}{\alpha b} \right\},$$
and the claim follows. □

**Proof of Theorem 3.** First note that if $\alpha \geq \frac{2}{b}v$, then setting $\sigma_i = 0$ for all $i$ is feasible in (OPT-M). This leads to an objective value of zero, which, after observing the nonnegativity of the objective, implies that the constructed solution is optimal. Thus, we conduct our analysis restricting attention to $\alpha < \frac{2}{b}v$. We claim that in this regime, the optimal solution is given by $\sigma_i = \min\{\alpha, \frac{2}{b}v - \alpha\}$ for all $i$. We prove the claim by first relaxing (OPT-M) and showing that the aforementioned solution is optimal for this relaxation. Then, we argue that this solution is also feasible, and hence optimal, in (OPT-M).

We start by relaxing the first constraint of (OPT-M) and considering the following problem:

$$
\begin{align*}
\min_{\sigma_1, \ldots, \sigma_n} & \quad \frac{1}{\alpha} \sum_{i \in V} \sigma_i^2 \\
\text{s.t.} & \quad \alpha^2 - \sigma_i^2 \geq \frac{2}{b}v(\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b}d_i \sigma_i \\
& \quad 0 \leq \sigma_i \leq \alpha.
\end{align*}
$$

(M2)

Observe that

$$
\frac{2}{b}v(\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max\{\sigma_i, \sigma_j\} - \frac{2}{b}d_i \sigma_i \geq \frac{2}{b}v(\alpha - \sigma_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \sigma_i - \frac{2}{b}d_i \sigma_i = \frac{2}{b}v(\alpha - \sigma_i).
$$

We use this observation to further relax the first constraint of (M2) by replacing the right-hand side of the constraint with $\frac{2}{b}v(\alpha - \sigma_i)$:

$$
\begin{align*}
\min_{\sigma_1, \ldots, \sigma_n} & \quad \frac{1}{\alpha} \sum_{i \in V} \sigma_i^2 \\
\text{s.t.} & \quad \alpha^2 - \sigma_i^2 \geq \frac{2}{b}v(\alpha - \sigma_i) \\
& \quad 0 \leq \sigma_i \leq \alpha.
\end{align*}
$$

(M3)

Note that this problem decouples over $i$. Thus, solving the following problem for each $i$ yields an optimal solution to (M3):

$$
\begin{align*}
\min_{\sigma_i} & \quad \frac{1}{\alpha} \sigma_i^2 \\
\text{s.t.} & \quad \alpha^2 - \sigma_i^2 \geq \frac{2}{b}v(\alpha - \sigma_i) \\
& \quad 0 \leq \sigma_i \leq \alpha.
\end{align*}
$$

(M4)
Suppose that the optimal solution is such that $\sigma^*_i < \alpha$. Then, feasibility implies that $\alpha^2 - (\sigma^*_i)^2 \geq \frac{2}{b} v (\alpha - \sigma^*_i)$, or equivalently

$$\alpha + \sigma^*_i \geq \frac{2}{b} v.$$

This implies that $\sigma^*_i \geq \frac{2}{b} v - \alpha$. Since the objective is strictly increasing in $\sigma_i$ (for $\sigma_i \geq 0$) and $\frac{2}{b} v - \alpha > 0$, we conclude that if the optimal solution is not $\sigma^*_i = \alpha$, then it is given by $\sigma^*_i = \frac{2}{b} v - \alpha$. Moreover, since the objective is decreasing in $\sigma_i$, we conclude that if $\frac{2}{b} v - \alpha \leq \alpha$, then setting $\sigma^*_i = \frac{2}{b} v - \alpha$ yields a feasible and optimal solution. Otherwise, $\sigma^*_i = \alpha$ yields an optimal solution.

More compactly, we express the optimal solution to (M4) and hence (M3) as

$$\sigma^*_i = \min \{ \alpha, \frac{2}{b} v - \alpha \}.$$  \hspace{1cm} (52)

Since (M3) is obtained by relaxing (OPT-M), if (52) is feasible in this problem, then it is optimal.

Observe that under the solution (52), the last constraint of (OPT-M) holds by construction. In addition, we have $\alpha^2 - (\sigma^*_i)^2 \geq \frac{2}{b} v (\alpha - \sigma^*_i)$. This inequality trivially follows if $\sigma^*_i = \alpha$. Otherwise, observe that $\sigma^*_i = \frac{2}{b} v - \alpha$, and hence,

$$\alpha^2 - (\sigma^*_i)^2 = (\alpha - \sigma^*_i)(\alpha + \sigma^*_i) = \frac{2}{b} v (\alpha - \sigma^*_i).$$

On the other hand, we also have

$$\frac{2}{b} v (\alpha - \sigma^*_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \max \{ \sigma^*_i, \sigma^*_j \} - \frac{2}{b} d_i \sigma^*_i = \frac{2}{b} v (\alpha - \sigma^*_i) + \frac{2}{b} \sum_{j \in V} g_{ij} \sigma^*_i - \frac{2}{b} d_i \sigma^*_i = \frac{2}{b} v (\alpha - \sigma^*_i).$$

Using this observation together with the previous inequality, implies that the second constraint of (OPT-M) trivially holds under the constructed solution. Finally, observe that

$$\frac{2}{b} v \sigma^*_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min \{ \sigma^*_i, \sigma^*_j \} \geq \frac{2}{b} v \sigma^*_i > \alpha \sigma^*_i \geq (\sigma^*_i)^2.$$

Thus, the first constraint of (OPT-M) also trivially holds for the solution (52). Since this solution is optimal for a relaxation of (OPT-M) and also satisfies all the constraints of (OPT-M), we conclude that it is optimal in this problem.

Summarizing, if $\alpha \geq \frac{2}{b} v$, then the optimal solution to (OPT-M) is given by $\sigma_i = 0$ for all $i$. Otherwise, it is given by the solution in (52). Hence, the claim follows. \(\square\)

**Proof of Corollary 4** The fact that no intervention mechanisms are optimal when $\frac{2}{b} v \leq \alpha$ and $\frac{2}{b} v \geq 2\alpha$ readily follows from Theorem 3.
When $\alpha < \frac{2v}{b} < 2\alpha$, the misinformation under the no intervention mechanism is given by Theorem 2(iii) and equals $|V|^{\frac{1}{2}}$, as claimed. By Theorem 3, the optimal public signal mechanism features a threshold of $\sigma = \frac{2v}{b} - \alpha$. The corresponding minimal misinformation is given by $|V|^{\frac{1}{2}} = |V|^{\frac{(2\alpha^2 - \alpha^2)}{2\alpha}}$. 

\[\Box\]

B.2. Example: Optimal thresholds vs. Network Structure

In this section, we provide an example that illustrates that in the engagement maximization problem nodes that have larger degree may have smaller or larger thresholds than the rest.

Example 3. Assume that $G$ is a star graph with $k$ leaves. Assume further that $\frac{2}{b}(v + k) < \alpha$. The leaves have the minimum degree, and therefore, by Corollary 1, the optimal thresholds corresponding to the leaves are the smallest. By part (i) of Lemma 6, it follows that for a leaf node $l$ the threshold is given by $\sigma_l = \frac{2}{b}(v + 1)$. For the center, the unique positive solution of (12) yields

\[\sigma_c = \frac{v}{b} + \frac{\sqrt{4k(v + 1) + v^2}}{b}.\]

Thus, in this example agents with larger degree have larger thresholds.

Next, consider a graph that consists of two different connected components: a star with $k$ leaves and a $(k - 1)$-regular graph. Here, each node of the $(k - 1)$-regular graph has a smaller degree than the center of the star. It can be readily checked that (OPT-E) decouples over disconnected components of the graph. Lemma 7 implies that the optimal threshold for each node of the $(k - 1)$-regular graph is equal to

\[\sigma_r = \frac{2}{b}(v + k - 1).\]

Thus, when $k$ is large enough, $\sigma_r > \sigma_c$ even though the degrees of the nodes in the regular graph are smaller than the degree of the center of the star.

C. Proofs of Section 5

Proof of Lemma 8. Observe that if $E > n$, then (OPT-F) is infeasible; hence $M^*(E) = \infty$, and the claim trivially holds. Assume that $E \leq n$.

Consider the following relaxation of (OPT-F) obtained by omitting the first two constraints:

\[
\min_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \frac{\sigma_i^2}{2}
\]

s.t. $\sum_{i \in V} \sigma_i \geq \alpha E$

\[0 \leq \sigma_i \leq \alpha \quad \text{for all } i \in V.\]
Since this is an optimization problem with a continuous objective over a compact domain, it admits an optimal solution. Observe that this problem is symmetric in \( \{\sigma_i\} \), i.e., if \( \{\sigma_1, \ldots, \sigma_i, \ldots, \sigma_j, \ldots, \sigma_n\} \) is an optimal solution, so is \( \{\sigma_1, \ldots, \sigma_j, \ldots, \sigma_i, \ldots, \sigma_n\} \). Furthermore, by convexity of the problem (and hence the set of optimal solutions), it admits an optimal solution such that \( \sigma_i = \sigma \) for all \( i \). Setting \( \sigma_i = \sigma \) for all \( i \) in (53), we obtain the following problem:

\[
\min_{\sigma} \quad \frac{n}{2\alpha} \sigma^2 \\
\text{s.t.} \quad \sigma \geq \frac{\alpha E}{n}, \\
0 \leq \sigma \leq \alpha.
\]

Observe that since the objective is increasing in \( \sigma \), it follows that the optimal solution to this problem is \( \sigma = \frac{\alpha E}{n} \leq \alpha \). Thus, we conclude that (53) admits a solution \( \sigma = \sigma_i = \frac{\alpha E}{n} \) for all \( i \). The corresponding objective value is \( \frac{n}{2\alpha} \sigma^2 = \frac{\alpha E^2}{2n} \). Since (53) is obtained by relaxing (OPT-F), it follows that \( M^*(E) \geq \frac{\alpha E^2}{2n} \), as claimed.

\[\square\]

**Proof of Lemma 9**. Fix some \( E \) that satisfies the constraints of the lemma. Set \( \sigma_i = \sigma = \alpha E/n \) for all \( i \in V \). We claim that this is a feasible solution in (OPT-F). To see this, note that when \( \sigma_i = \sigma \) for all \( i \in V \), the first constraint of this problem reduces to

\[\sigma^2 \leq \frac{2}{b} \nu \sigma + d_i \sigma \]

for any \( i \in V \). On the other hand, by construction, \( 0 \leq \sigma = \alpha E/n \leq \frac{2}{b} (d_{\text{min}} + \nu) \), and hence, this inequality trivially holds. Similarly, when \( \sigma_i = \sigma \) for all \( i \in V \), the second constraint of (OPT-F) reduces to

\[\alpha^2 - \sigma^2 \geq \frac{2}{b} \nu (\alpha - \sigma).\]  

(54)

On the other hand, the constraints on \( E \) imply that \( \sigma = \alpha E/n \leq \alpha \) and \( \sigma \geq \frac{2}{b} \alpha - \alpha \), and hence (54) holds. The third constraint readily follows from the construction. Finally, the last constraint follows since, as discussed before, \( \sigma \leq \alpha \), and due to the fact that \( E \geq 0 \), we have \( \sigma \geq 0 \).

Observe that the constructed solution leads to a total misinformation of \( M^*(E) = \frac{1}{2\alpha} \sum_{i \in V} \sigma_i^2 = \alpha E^2/2n \). Therefore, by Lemma 8, it is optimal, and \( \sigma_i^*(E) = \sigma = \alpha E/n \) for all \( i \in V \).

\[\square\]

**Proof of Lemma 10**. If \( 2\nu/b \leq \alpha \), then \( E = 0 \). Setting \( \sigma_i^*(E) = 0 \) for all \( i \in V \), we obtain a feasible solution of (OPT-F). Since the objective is lower bounded by zero, it follows that this is an optimal solution with \( M^*(E) = 0 \), and the claim trivially follows.
Suppose instead $2v/b > \alpha$. Let

$$E = \frac{n}{\alpha} \min \left\{ \alpha, \frac{2v}{b} - \alpha \right\},$$

and observe that for all $E \leq E_{\text{min}}$,

$$M^*(E) \leq M^*(E),$$

since the optimization problem $\text{(OPT-F)}$ corresponding to engagement requirement $E$ is a relaxed version of that corresponding to $E_{\text{min}}$. Observe that by setting $\sigma_i = \frac{\alpha E}{n} = \min \left\{ \alpha, \frac{2v}{b} - \alpha \right\}$ for all $i \in V$, we obtain a feasible solution of an instance of $\text{(OPT-F)}$ with $E \leq E_{\text{min}}$. The fact that the constructed solution satisfies the first three constraints of $\text{(OPT-F)}$ can be readily checked by substituting $\sigma_i = \min \left\{ \alpha, \frac{2v}{b} - \alpha \right\}$ in these constraints. The last constraint holds since by construction $\sigma_i \leq \alpha$, and the assumptions of the lemma guarantee that $2v/b - \alpha \geq 0$, which in turn ensures $\sigma_i \geq 0$ for all $i \in V$.

Since the optimal solution of $\text{(OPT-F)}$ induces a weakly lower objective value than the constructed solution, for $E \leq E_{\text{min}}$, we have

$$M^*(E) \leq \frac{1}{2\alpha} \sum_{i \in V} \sigma_i^2 = \frac{\alpha}{2n} E^2.$$  \hspace{1cm} (56)

On the other hand, by Theorem 3, the mechanism with $\sigma_i = \alpha E/n$ for all $i$ minimizes misinformation (even in the absence of engagement constraints), and hence, for any $E \leq E_{\text{min}}$, we have

$$M^*(E) \geq \frac{\alpha}{2n} E^2.$$  \hspace{1cm} (56)

Combining this inequality with (56), we obtain $M^*(E) = M^*(E) = \frac{\alpha}{2n} E^2$. Moreover, since $\sigma_i = \alpha E/n$ is feasible in $\text{(OPT-F)}$ for $E \leq E_{\text{min}}$ and achieves this objective value, we conclude that $\sigma_i^*(E) = \sigma_i = \alpha E/n$ for $E \leq E_{\text{min}}$. \hspace{1cm} \square

**Proof of Lemma 11** Let us denote by $K$ the set of minimum degree nodes

$$K = \{ i \in V : d_i = d_{\text{min}} \}.$$  

Observe that since the network is not regular, we have $K \neq V$ and $m < n$.

For notational convenience, we write

$$E = \frac{2n(d_{\text{min}} + v)}{b\alpha}.$$
Note that for any feasible solution of the program \((\text{OPT-F})\) and any \(i \in K\), we have
\[\sigma_i^2 \leq \frac{2}{b}v\sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij} \min\{\sigma_i, \sigma_j\} \leq \frac{2}{b}v\sigma_i + \frac{2}{b} \sum_{j \in V} g_{ij}\sigma_i,\]
which implies
\[\sigma_i \leq \frac{2}{b}v + \frac{2}{b}d_{\min}.\]

Therefore, the following program
\[
\min_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \frac{\sigma_i^2}{2}
\text{ s.t. } \sigma_i \leq \frac{2}{b}v + \frac{2}{b}d_{\min}, \text{ for all } i \in K
\]
\[
(57)
\]
is a relaxation of \((\text{OPT-F})\). Since the objective is increasing in \(\{\sigma_i\}\), it can be seen that at the optimal solution of \((57)\), the engagement requirement will be binding. Hence, explicitly imposing this constraint, we obtain the following equivalent program:
\[
\min_{\sigma_1, \ldots, \sigma_n} \frac{1}{\alpha} \sum_{i \in V} \frac{\sigma_i^2}{2}
\text{ s.t. } \sigma_i \leq \frac{2}{b}v + \frac{2}{b}d_{\min}, \text{ for all } i \in K,
\sum_{i \in V} \sigma_i = \alpha E,
\]
\[
(58)
\]
We next obtain an optimal solution to the above program and establish that it is feasible in \((\text{OPT-F})\). Note that since the above program is a relaxation of \((\text{OPT-F})\), this readily implies the optimality of the constructed solution in the latter program.

Note that \((58)\) is symmetric among \(\{\sigma_i\}_{i \in K}\) as well as \(\{\sigma_i\}_{i \in V \setminus K}\). Since the problem is convex, we conclude that there exists an optimal solution where \(\sigma_i = \sigma_L\) for all \(i \in K\) and \(\sigma_i = \sigma_H\) for all \(i \in V \setminus K\). Feasibility implies that for these solutions we have \(m\sigma_L + (n-m)\sigma_H = \alpha E\). Using these observations, \((58)\) can be written as the following single parameter optimization problem:
\[
\min_{\sigma_L \in [0, 2(v+d_{\min})/b]} \frac{1}{2\alpha} \left( m\sigma_L^2 + \frac{(\alpha E - m\sigma_L)^2}{n-m} \right).
\]
The optimal solution of the above for a given \(E\) is equal to
\[
\sigma_L(E) := \min\left\{ \frac{2(v + d_{\min})}{b}, \frac{\alpha E}{n} \right\},
\]
where the second term is obtained from the first order conditions of the objective. For $E \geq E$, 
\[
\frac{2(v + d_{\text{min}})}{b} = \frac{E}{n} \leq \frac{E}{n},
\]
and therefore
\[
\sigma_L(E) = \frac{2(v + d_{\text{min}})}{b} \quad \text{and} \quad \sigma_H(E) = \frac{1}{n-m}\left(\alpha E - m\frac{2(v + d_{\text{min}})}{b}\right).
\]
Furthermore, we note that by (59), $\sigma_L(E) \leq \alpha E/n$, which yields
\[
\sigma_L(E) \leq \sigma_H(E),
\]
and
\[
\sigma_H(E) = \sigma_L(E).
\]

We conclude the proof by showing that $\sigma_i = \sigma_L(E)$, $i \in K$ and $\sigma_i = \sigma_H(E)$, $i \in V \setminus K$ is actually feasible in (OPT-F), when $E \in [E, E+\delta)$ for $\delta$ small enough. We start by checking the first family of constraints. By (60), $\sigma_i \leq \sigma_j$ for all $i \in K$, $j \in V$, and therefore, for all $i \in K$, the first constraint can be written as
\[
\sigma_L(E)^2 \leq \frac{2}{b} v\sigma_L(E) + \frac{2}{b} \sum_{j \in V} g_{ij}\sigma_L(E) = \frac{2(v + d_{\text{min}})}{b}\sigma_L(E),
\]
which trivially holds, by the definition of $\sigma_L(E)$. For $i \in V \setminus K$, note that when $E = E$, by (61), we get
\[
\sigma_H(E)^2 = \frac{2}{b} v\sigma_H(E) + \frac{2}{b} d_{\text{min}}\sigma_H(E) < \frac{2}{b} v\sigma_H(E) + \frac{2}{b} \sum_{j \in V} g_{ij}\sigma_H(E),
\]
and therefore, by continuity, the constraint continues to hold for $E \in [E, E+\delta)$ and $\delta$ small enough. Observe that for all $i \in V$ and $E = E$, by using (61), the second constraint of (OPT-F) reduces to
\[
\alpha^2 - \sigma_L(E)^2 > \frac{2}{b} v(\alpha - \sigma_L(E)),
\]
where the inequality is strict since the lemma assumes $\alpha > 2(d_{\text{min}} + v)/b$. Therefore, by continuity, the constraint continues to hold for $E \in [E, E+\delta)$ and $\delta$ small enough. The engagement requirement constraint (i.e., the third constraint of (OPT-F)) follows from the definition of $\sigma_L(E)$ and $\sigma_H(E)$. Finally, the last constraint follows from the assumption of the lemma and the observation that
\[
\sigma_H(E) = \sigma_L(E) < \alpha.
which, by continuity of \( \sigma_L(E), \sigma_H(E) \), implies that the constraint continues to hold for \( E \in [E, E + \delta) \) and \( \delta \) small enough.

Thus, we conclude that the constructed solution is optimal in [OPT-F] or \( E \in [E, E + \delta) \) and \( \delta \) small enough. The objective associated with this solution can be given by

\[
M^*(E) = \frac{m}{2\alpha} \left( \frac{\sigma_L(E)^2}{\alpha} + \frac{(n - m)\sigma_H(E)^2}{\alpha} \right) + \frac{m}{2}(n - m) \left( \frac{1}{\alpha} \right) \sum_{i \in E} \left( \frac{2(v + d_{\min})}{b} \right)^2 + \left( \frac{1}{n - m} \right) \left( \frac{2(v + d_{\min})}{b} \right)^2.
\]

\[\Box\]

**Proof of Proposition 5** In order to prove Proposition 6, we will make use of the following auxiliary results.

**Lemma 17.** (i) Suppose \( \bar{\sigma}_i > 0 \), and \( \bar{\sigma}_i^2 = \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_j g_{ij} \min \{ \bar{\sigma}_i, \bar{\sigma}_j \} \). Then for \( \sigma_i \in [0, \bar{\sigma}_i] \), we have \( \sigma_i^2 \leq \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_j g_{ij} \min \{ \sigma_i, \sigma_j \} \).

(ii) Suppose \( \bar{\sigma}_i < \alpha \), and \( \alpha^2 - \sigma_i^2 = \frac{2}{b} v (\alpha - \sigma_i) + \frac{2}{b} \sum_j g_{ij} \max \{ \sigma_i, \sigma_j \} - \frac{2}{b} d_i \sigma_i \). Then, for \( \sigma_i \in [\bar{\sigma}_i, \alpha] \), we have \( \alpha^2 - \sigma_i^2 \geq \frac{2}{b} v (\alpha - \sigma_i) + \frac{2}{b} \sum_j g_{ij} \max \{ \sigma_i, \sigma_j \} - \frac{2}{b} d_i \sigma_i \).

**Proof:** (i) Let \( f_L(\sigma_i) = \sigma_i^2 \), and \( f_R(\sigma_i) = \frac{2}{b} v \sigma_i + \frac{2}{b} \sum_j g_{ij} \min \{ \sigma_i, \sigma_j \} \). Observe that \( f_L(0) = f_R(0) \) and \( f_L(\sigma_i) = f_R(\sigma_i) \). Since \( f_L(\cdot) \) is strictly convex, and \( f_R(\cdot) \) is a concave function, this immediately implies that \( f_L(\sigma_i) \leq f_R(\sigma_i) \) for \( \sigma_i \in [0, \bar{\sigma}_i] \), and the claim follows.

(ii) Let \( f_L(\sigma_i) = \alpha^2 - \sigma_i^2 \), and \( f_R(\sigma_i) = \frac{2}{b} v (\alpha - \sigma_i) + \frac{2}{b} \sum_j g_{ij} \max \{ \sigma_i, \sigma_j \} - \frac{2}{b} d_i \sigma_i \). Observe that \( f_L(\alpha) = f_R(\alpha) \) and \( f_L(\sigma_i) = f_R(\sigma_i) \). Since \( f_L(\cdot) \) is strictly concave, and \( f_R(\cdot) \) is a convex function, this immediately implies that \( f_L(\sigma_i) \geq f_R(\sigma_i) \) for \( \sigma_i \in [\bar{\sigma}_i, \alpha] \), and the claim follows.

We next characterize how given a feasible solution of [OPT-M], another feasible solution can be obtained by perturbing the original solution. Since the constraints of [OPT-F] are a superset of those of [OPT-M], our result also allows us to reason about the feasible solutions of [OPT-F].

**Lemma 18.** Suppose \( \{ \sigma_i \} \) is feasible in [OPT-M]. Let \( \bar{\mathcal{S}} \) = \( \{ i \mid \sigma_i \geq \sigma_j \ \forall j \in V \} \), and \( \mathcal{S} = \{ i \mid \sigma_i \leq \sigma_j \ \forall j \in V \} \). Assume that \( \bar{\mathcal{S}} \neq \mathcal{S} \).

(i) Let \( \bar{\sigma}_i \) be such that \( \bar{\sigma}_i = \sigma_i - \delta \) for \( i \in \bar{\mathcal{S}} \) and \( \sigma_i = \bar{\sigma}_i \) for \( i \notin \mathcal{S} \), for an arbitrarily small \( \delta > 0 \). Then \( \{ \bar{\sigma}_i \} \) is also feasible in [OPT-M].

(ii) \( \{ \bar{\sigma}_i \} \) such that \( \bar{\sigma}_i = \sigma_i + \delta \) for \( i \in \mathcal{S} \) and \( \sigma_i = \bar{\sigma}_i \) for \( i \notin \bar{\mathcal{S}} \), for an arbitrarily small \( \delta > 0 \). If \( \sigma_j = 0 \) for \( j \in \mathcal{S} \), then \( \{ \bar{\sigma}_i \} \) is also feasible in [OPT-M].
Proof: First note that updating the $\{\sigma_i\}$ as in (i) or (ii) does not change the ranking of agents in terms of their $\{\sigma_i\}$, i.e., if $\sigma_i \geq \sigma_j$ then $\hat{\sigma}_i \geq \hat{\sigma}_j$. Consequently the agents who achieve minima/maxima in the first two constraints of $\text{(OPT-M)}$ do not change. Since $\mathcal{S} \neq \mathcal{S}$, we have that $\sigma_i > 0$ for $i \in \mathcal{S}$ and $\sigma_j < \alpha$ for $i \in \mathcal{S}$3. Thus the updates still guarantee that $\hat{\sigma}_i \in [0, \alpha]$ for arbitrarily small $\delta$. To prove the claim we only focus on first and second family of constraints of $\text{(OPT-M)}$. Note that for each agent $i \in V$ we have one constraint in each family, where the left hand side is only a function of $\sigma_i$, and the right hand side involves either $\min \{\sigma_i, \sigma_j\}$ or $\max \{\sigma_i, \sigma_j\}$ terms. We refer to these respectively as the first/second constraint for agent $i$.

(i) First consider $j \notin \mathcal{S}$. Observe that updating $\{\sigma_i\}$ as stated relaxes the second constraint of $\text{(OPT-M)}$ for agent $j \notin \mathcal{S}$. On the other hand for such $j$, we have $\sigma_j < \max \{\sigma_i\}$. Thus, the first constraint is not impacted for $j \notin \mathcal{S}$. Hence, the update does not violate first/second constraint for $j \notin \mathcal{S}$.

Consider $j \in \mathcal{S}$. Observe that if the first and second constraint were not binding for agent $j$ under $\{\sigma_i\}$, then the update does not violate feasibility. Suppose the first constraint was binding, i.e., $\sigma_j = \hat{\sigma}_j$ for $j \notin \mathcal{S}$. On the other hand for such $j$, we have $\sigma_j < \max \{\sigma_i\}$. Thus, the first constraint is satisfied for any $\hat{\sigma}_j \in [0, \sigma_j]$, and in particular it is satisfied for $\hat{\sigma}_j = \sigma_j$. Note that this result continues to hold if $\sigma_j$ is updated for all agents in $\mathcal{S}$ jointly, since $\sigma_{j_1} = \sigma_{j_2}$ and $\sigma_j = \sigma_j$ for $j_1, j_2 \notin \mathcal{S}$.

Similarly, suppose that the second constraint was binding, i.e.,

$$\alpha^2 - \sigma_j^2 = 2v_b (\alpha - \sigma_j) + \frac{2}{b} \sum_{k \in V} g_{jk} \max \{\sigma_j, \sigma_k\} - \frac{2}{b} d_j \sigma_j = \frac{2v_b}{b} (\alpha - \sigma_j). \quad (62)$$

Observe that for $i \notin \mathcal{S}$, we have

$$\alpha^2 - \sigma_i^2 \geq 2v_b (\alpha - \sigma_i) + \frac{2}{b} \sum_{k \in V} g_{ik} \max \{\sigma_i, \sigma_k\} - \frac{2}{b} d_i \sigma_i \geq \frac{2v_b}{b} (\alpha - \sigma_i). \quad (63)$$

Let $h(\sigma) = \alpha^2 - \sigma^2 - \frac{2v}{b} (\alpha - \sigma)$. Observe that $h(\cdot)$ is strictly concave. We have from (62) and (63) that $h(\sigma_j) = 0$, and $h(\sigma_i) \geq 0$. By concavity, we have for any $\sigma \in [\sigma_i, \sigma_j]$ that $h(\sigma) \geq \min \{h(\sigma_j), h(\sigma_i)\} \geq 0$. Thus, it follows that the second constraint is satisfied for any $\hat{\sigma}_j \in [\sigma_i, \sigma_j]$, and in particular it is satisfied for $\hat{\sigma}_j = \sigma_j$. Thus, we conclude that the constructed solution is feasible in $\text{(OPT-M)}$.

(ii) First consider $j \notin \mathcal{S}$. Observe that updating $\{\sigma_i\}$ as stated relaxes the first constraint in $\text{(OPT-M)}$ for $j \notin \mathcal{S}$. On the other hand for such $j$, we have $\sigma_j < \min \{\sigma_i\}$. Thus, the second constraint for $j \notin \mathcal{S}$ is not impacted. Hence, the update does not impact the first/second constraint for $j \notin \mathcal{S}$.

Consider $j \in \mathcal{S}$. Observe that if the first/second constraint was not binding for agent $j$ under $\{\sigma_i\}$, then the update does not violate feasibility. Suppose the second constraint was binding, i.e.,
\( \alpha^2 - \sigma_j^2 = \frac{2v}{b} (\alpha - \sigma_j) + \frac{2}{b} \sum_{k \in V} g_{jk} \max \{ \sigma_j, \sigma_k \} - \frac{2}{b} d_j \sigma_j. \) By Lemma 17, it follows that the second constraint is satisfied for any \( \tilde{\sigma}_j \in [\sigma_j, \alpha] \) when the threshold of only agent \( j \) is updated, and in particular it is satisfied for \( \tilde{\sigma}_j = \sigma_j \). As before, this result continues to hold if \( \sigma_j \) is jointly updated for all \( j \in S \). Note that since \( \sigma_j = 0 \), the minimum constraint was binding, i.e.,

\[
\sigma_j^2 = \frac{2v}{b} \sigma_j + \frac{2}{b} \sum_{k \in V} g_{jk} \min \{ \sigma_j, \sigma_k \} = 0. \tag{64}
\]

On the other hand, as a result of the aforementioned update the left hand side becomes \( \delta^2 < \frac{2v}{b} \delta \). Thus, it follows that the minimum constraint is satisfied for \( \tilde{\sigma}_j \). Hence, we conclude that the constructed solution is feasible in \((\text{OPT-M})\).

Lemma 19. Suppose \( E > 0 \), and let \( \{ \sigma_i \} \) be an optimal solution to \((\text{OPT-F})\). Then \( \sigma_i > 0 \) for all \( i \in V \).

Proof: Let \( S = \{ j \mid \sigma_j = 0 \} \). Since \( E > 0 \), by feasibility in \((\text{OPT-F})\), it follows that \( V \setminus S \neq \emptyset \).

Consider another threshold mechanism, where \( \tilde{\sigma}_i = \sigma_i + \delta \) for \( i \in S \) and arbitrarily small \( \delta > 0 \). Assume that \( \tilde{\sigma}_j = \sigma_j \) for \( j \notin S \). Note that the constraints of \((\text{OPT-F})\) consist of those of \((\text{OPT-M})\) and an additional constraint \( \sum_i \sigma_i \geq E \alpha \). The latter constraint is trivially satisfied by \( \{ \tilde{\sigma}_i \} \) since \( \tilde{\sigma}_i \geq \sigma_i \), whereas Lemma 18 implies that the rest of the constraints also hold under \( \{ \tilde{\sigma}_i \} \). Thus, we conclude that \( \{ \tilde{\sigma}_i \} \) is feasible in \((\text{OPT-F})\).

Let \( \tilde{S} = \{ i \mid \tilde{\sigma}_i \geq \tilde{\sigma}_k, \forall k \in V \} \), and observe that since \( V \setminus S \neq \emptyset \), we have \( \tilde{S} \cap S = \emptyset \), and \( \tilde{S} \neq \emptyset \). We next construct a third threshold mechanism with thresholds \( \bar{\sigma}_i = \tilde{\sigma}_i \) for \( i \notin \tilde{S} \), and \( \bar{\sigma}_j = \bar{\sigma}_j - \bar{\delta} \) for \( j \in \tilde{S} \), where \( \bar{\delta}|\tilde{S}| = \delta|S| \). Lemma 18 implies that (for \( \bar{\delta}, \delta \) small enough) these thresholds also satisfy the constraints of \((\text{OPT-F})\) other than \( \sum_i \sigma_i \geq E \alpha \). On the other hand, by construction, we have \( \sum_i \bar{\sigma}_i = \sum_i \sigma_i \geq E \alpha \). Thus, we conclude that \( \{ \bar{\sigma}_i \} \) is feasible in \((\text{OPT-F})\).

Observe that

\[
\sum_i \sigma_i^2 - \sum_i \bar{\sigma}_i^2 = \sum_{i \in S} (\sigma_i^2 - \bar{\sigma}_i^2) + \sum_{i \notin S} (\sigma_i^2 - \bar{\sigma}_i^2) = \sum_{i \in S} (\sigma_i^2 - (\sigma_i + \delta)^2) + \sum_{i \notin S} (\sigma_i^2 - (\sigma_i - \delta)^2) = |S| \delta^2 - 2\delta \sum_{i \in S} \sigma_i - |\tilde{S}| \delta^2 + 2\delta \sum_{i \notin S} \sigma_i. \tag{65}
\]

Observe that \( \sigma_i = \sigma \) for all \( i \in S \) and similarly \( \sigma_i = \bar{\sigma} \) for all \( i \notin \tilde{S} \), for some \( \bar{\sigma} > \sigma \). Thus, \( (65) \) implies that

\[
\sum_i \sigma_i^2 - \sum_i \bar{\sigma}_i^2 = |S| \delta^2 - |\tilde{S}| \delta^2 + 2(|S| \delta \bar{\sigma} - |S| \delta \sigma) > 0, \tag{66}
\]

for arbitrarily small \( \delta, \bar{\delta} \). Here, the last inequality follows by noting that

\[
|\tilde{S}| \delta \bar{\sigma} - |S| \delta \sigma = |\tilde{S}| \delta (\bar{\sigma} - \sigma) + \sigma (|\tilde{S}| \delta - |S| \delta) = |\tilde{S}| \delta (\bar{\sigma} - \sigma) > 0.
\]
Feasibility of $\{\bar{\sigma}_i\}$ in (OPT-F) together with (66) implies that $\{\sigma_i\}$ is not optimal in (OPT-F).

Thus, we obtain a contradiction. Hence $S = \emptyset$, and the claim follows. □

**Lemma 20.** Let $F$ be a public signal mechanism with public signal $s$, and threshold $\sigma < \alpha$. Denote by $Q \in Q(F)$ the corresponding equilibrium that solves (17) for a given $E \geq 0$. Then, the set of agents for which $x_i = 1$ when $s = 0$ and $x_i = 0$ when $s = 1$ is empty.

**Proof:** Let $\{\mu_i\}_{i \in V}$ denote agents’ strategies at equilibrium $Q$, and define

$$S := \{i \in V | \mu_i(0) = 1 \text{ and } \mu_i(1) = 0\}. \quad (67)$$

Consider a new mechanism $\hat{F}$ with signals $\{\hat{s}_i\}_{i \in V}$ such that for all $\omega \in \Omega$ and $i \in V$, we have:

$$\hat{s}_i(\omega) = x \iff \mu_i(s(\omega)) = x,$$

where $x \in \{0, 1\}$. Note that mechanism $\hat{F}$ is a straightforward mechanism that implements the equilibrium $Q$ as argued in the proof of Lemma 1.

Note that, by the definition (67) of $S$, for all $i \in S$,

$$\hat{s}_i = 1 \iff \epsilon(\omega) > \sigma.$$

We now construct a third mechanism $\tilde{F}$, for which

$$\tilde{s}_i(\omega) = \hat{s}_i(\omega)$$

for all $\omega \in \Omega$ and $i \in V \setminus S$, and

$$\tilde{s}_i = 1 \iff \epsilon(\omega) \leq \alpha - \sigma,$$

for all $\omega \in \Omega$ and $i \in S$. We claim that $\tilde{F}$ has a straightforward equilibrium $\tilde{Q}$. Indeed, since under the new mechanism engagement signals are associated with lower error levels, we have

$$\mathbb{E}[\epsilon | \tilde{s}_i = 1] \leq \mathbb{E}[\epsilon | \hat{s}_i = 1],$$

$$\mathbb{E}[\epsilon | \tilde{s}_i = 0] \geq \mathbb{E}[\epsilon | \hat{s}_i = 0],$$

for all agents $i \in V$. In addition, since agents in $S$ now engage when $\epsilon(\omega) \leq \alpha - \sigma$

$$\sum_{j \in V} g_{ij} \mathbb{E}[\tilde{s}_j | \tilde{s}_i = 1] \geq \sum_{j \in V} g_{ij} \mathbb{E}[\hat{s}_j | \hat{s}_i = 1],$$
while
\[
\sum_{j \in V} g_{ij} \mathbb{E}[\tilde{s}_j | \tilde{s}_i = 0] \leq \sum_{j \in V} g_{ij} \mathbb{E}[\tilde{s}_j | \tilde{s}_i = 0].
\]

Therefore, since it is incentive compatible for all agents in \( V \) to follow the recommendation of \( \hat{F} \), it is also incentive compatible to follow the recommendation of \( \bar{F} \). Hence, \( \hat{Q} \) is a straightforward equilibrium. By construction
\[
E(\hat{Q}) = E(Q),
\]
but
\[
M(Q) - M(\hat{Q}) = \sum_{i \in S} \frac{1}{\alpha} \left( \int_{\sigma}^{\alpha} zdz - \frac{(\alpha - \sigma)^2}{2} \right) = \sum_{i \in S} \frac{1}{\alpha}(\alpha - \sigma)\sigma > 0.
\]

On the other hand,
\[
M(Q) \leq M(\hat{Q}),
\]
since the performance level of \( Q \) is on the E/M frontier, and therefore, \( S = \emptyset \). \( \square \)

We are now ready to proceed to the proof of Proposition 5. One direction follows readily from Lemmata 9 and 10. We now prove that if \( E > \bar{E} \), where
\[
\bar{E} = \frac{n}{\alpha} \min \left\{ \frac{2}{b}(v + d_{\min}), \alpha \right\}
\]
then public signals cannot induce performance that is on the E/M frontier. By the assumption of the proposition, \( \frac{2}{b}(v + d_{\min}) < \alpha \). Thus, we focus on the case where
\[
\bar{E} = \frac{2n}{b\alpha}(v + d_{\min}) < n.
\]

Consider some \( E' > \bar{E} \). Assume, for the purposes of contradiction that there exists some public signal mechanism \( \bar{F} \) with threshold \( \bar{\sigma} \) that admits an equilibrium \( Q \) on the E/M Frontier. If \( \bar{\sigma} < \alpha \), by Lemma 20, the set of agents for which \( x_i = 1 \) when \( s = 0 \) and \( x_i = 0 \) when \( s = 1 \) is empty. Therefore, there are three sets of agents:
\[
S_1 = \{ i : \mu_i(0) = 1, \mu_i(1) = 1 \},
S_2 = \{ i : \mu_i(0) = 0, \mu_i(1) = 1 \},
S_3 = \{ i : \mu_i(0) = 0, \mu_i(1) = 1 \}.
\]

Note that if \( \bar{\sigma} = \alpha \) agents trivially belong in one of the aforementioned sets.

Following the steps of Theorem 1 the threshold mechanism \( \bar{F} \) with thresholds \( \bar{\sigma}_i = \alpha \) for \( i \in S_1 \), \( \bar{\sigma}_i = 0 \) for \( i \in S_2 \) and \( \bar{\sigma}_i = \bar{\sigma} \) for all \( i \in S_3 \) is a straightforward threshold mechanism whose
straightforward equilibrium achieves the same performance level. Since \( Q \) is on the E/M Frontier, \( \{\tilde{\sigma}_i\} \) solves (OPT-F). By Lemma 19, \( S_2 = \emptyset \). By the incentive compatibility constraint (5), for all agents \( i \in S_1 \) we get
\[
\alpha^2 \leq \frac{2}{b} v \alpha + \frac{2}{b} d_i \alpha \leq \frac{2}{b} v \alpha + \frac{2}{b} d_{\max} \alpha,
\]
which contradicts with the assumption of the lemma, and hence \( S_1 = \emptyset \). Therefore, \( S_3 = V \). Observe that by feasibility, we have \( \sum_i \tilde{\sigma}_i = n\tilde{\sigma} \geq \alpha E' \). On the other hand, from the incentive compatibility constraint (5) and the fact that \( \tilde{\sigma}_i = \sigma \) for all \( i \in S_3 = V \), we get
\[
\tilde{\sigma}^2 \leq \frac{2}{b} \tilde{\sigma} v + \frac{2}{b} d_i \tilde{\sigma},
\]
for all \( i \in V \), and hence
\[
\tilde{\sigma}^2 \leq \frac{2}{b} \tilde{\sigma} v + \frac{2}{b} d_{\min} \tilde{\sigma}.
\]
Rearranging terms, this yields \( \tilde{\sigma} \leq \frac{2}{b}(v + d_{\min}) = \alpha \frac{E'}{n} < \alpha \frac{E'}{n} \). Thus, we obtain a contradiction to \( n\tilde{\sigma} \geq \alpha E' \), and the claim follows. \( \square \)

D. Engagement Maximization for Public Signal Mechanisms

In this section we provide an algorithm for finding the engagement-maximizing public signal mechanism. Our algorithm relies on the following proposition.

Proposition 6. \( (i) \) Let \( B_0 = \{\sigma | (b\frac{z+\alpha}{2} - v) \in \{0, 1, \ldots, n\}\} \), \( B_1 = \{\sigma | (b\frac{z}{2} - v) \in \{0, 1, \ldots, n\}\} \), and \( B = B_1 \cup B_2 \). Consider \( \sigma_1, \sigma_2 \in B \) such that \( (\sigma_1, \sigma_2) \cap B = \emptyset \). The maximum achievable engagement is an increasing function of \( \sigma \) in \( (\sigma_1, \sigma_2) \).

\( (ii) \) If engagement is discontinuous at some \( \sigma \in B \), then engagement achievable at \( \sigma + \delta \) is strictly lower than that at \( \sigma \) for any \( \delta > 0 \).

\( (iii) \) Moreover, maximum engagement (among all public signal mechanisms) is achieved by some \( \sigma \in B \).

Proof: Consider the characterization of public signal mechanisms in Theorem 2. Observe that \( B_0 \) contains the \( \sigma \) where \( k_0 \) changes, i.e., \( k_0 \) associated with \( \sigma + \delta \) is different than the one associated with \( \sigma \). Similarly, \( B_1 \) contains the \( \sigma \) where \( k_1 \) changes. Thus, it follows that \( |A_0(\sigma)| \) and \( |A_1(\sigma)| \) are constant for \( \sigma \in (\sigma_1, \sigma_2) \), where \( \sigma_1, \sigma_2 \) are as defined in the statement of the proposition. Thus, Theorem 2 implies that the maximum engagement that corresponds to a given \( \sigma \) changes continuously in \( (\sigma_1, \sigma_2) \). Moreover, since \( |V| \geq |A_1(\sigma)| \geq |A_0(\sigma)| \), it follows that the corresponding engagement is weakly increasing in \( \sigma \in (\sigma_1, \sigma_2) \). Hence, we conclude that the optimal engagement in this interval is achieved for \( \sigma = \sigma_2 \). Since this is true for any such interval, we conclude that optimal engagement is obtained for \( \sigma \in B \).
Suppose that maximum engagement changes discontinuously at \( \sigma \in B \). Then, it must be the case that \( |A_i(\sigma)| \neq |A_i(\sigma + \delta)| \) for arbitrarily small \( \delta > 0 \) for \( i = 0 \) or \( i = 1 \). However, since \( A_i(\sigma) \subseteq A_i(\sigma') \) for \( \sigma' \leq \sigma \), we conclude that \( |A_i(\sigma)| > |A_i(\sigma + \delta)| \). Hence, if engagement changes discontinuously at \( \sigma \), then it has to be the case that engagement achievable at \( \sigma + \delta \) is strictly lower than that at \( \sigma \), as claimed. \( \square \)

This finding leads to two important conclusions. First, if the platform resorts to using a public signal mechanism, the corresponding threshold must be designed with extreme care, as the performance is not robust to perturbations (as formalized in part (i)). Second, in general, non-monotonicity suggests that it may be difficult to find the optimal \( \sigma \). However, Proposition 6 suggests that when searching over optimal \( \sigma \), it suffices to restrict attention to finitely many public thresholds that belong to \( B \). Since \( |B| \leq 2n + 2 \), this immediately implies a simple algorithm (Algorithm 2) for characterizing the thresholds that maximize engagement among all public signal mechanisms.

**Algorithm 2** An algorithm for computing public signal threshold \( \sigma \) that maximizes engagement.

**S1.** If \( 2v/b > \alpha \), set \( \sigma = 0 \), and go to Step S3. Otherwise, compute \( B_1, B_2 \), and set \( B = B_1 \cup B_2 \).

**S2.** Compute \( \sigma_1 = \arg \max_{\sigma \in B|\sigma < 2v/b} f_L(\sigma) \) and \( \sigma_2 = \arg \max_{\sigma \in B|\sigma \geq 2v/b} f_R(\sigma) \), and denote the corresponding optimal values respectively by \( f^*_L \) and \( f^*_R \). Set \( \sigma = \sigma_1 \) if \( f^*_L > f^*_R \), and set \( \sigma = \sigma_2 \) otherwise.

**S3.** Return \( \sigma \).

### E. Revealing Error Publicly

In this section we consider an alternative benchmark where the platform reveals the true error \( \epsilon \) of the content publicly to all the agents (for any realization of \( \epsilon \)). We refer to such mechanisms as *true error revelation* mechanisms, and explore the expected engagement and misinformation under these mechanisms.

Observe that under true error revelation the agents play a game where their payoffs are as given in (1). It can be readily checked that for any given \( \epsilon \) the induced game is a supermodular game. Hence, our approach in Section 3.2 can be used to characterize the equilibria.

In particular, for any given \( \epsilon \), the induced game has best and worst equilibria, which we respectively denote by \( X^- \) and \( X^+ \). These equilibria can be characterized following the approach in the proof of Theorem 2. In this theorem, for a given threshold \( \sigma \), when agents receive the public signal \( s_i = 1 \), they play a supermodular game where the expected error is given by \( \sigma/2 \), and the error disutility is given by \( -b\sigma/2 \) (similarly when \( s_i = 0 \), they play a supermodular game where the error disutility is \( -b(\sigma + \alpha)/2 \). The corresponding best and worst equilibria are characterized in terms of the \( k \)-cores of the underlying network. When \( \epsilon \) is publicly revealed, as opposed to focusing on
the expected error $\sigma/2$, agents base their decision on the error $\epsilon$ of the content. Thus, repeating the steps in the proof of Theorem 2 (and replacing $\sigma/2$ with $\epsilon$ in the payoffs), the following result can be established:

**Lemma 21.** Let $k_\epsilon$ be the smallest integer satisfying $k_\epsilon \geq b\epsilon - v$, and $A_{k_\epsilon}$ denote the $k_\epsilon$-core of the underlying network.

- If $v \leq b\epsilon$, in $X^\epsilon$ agents in $A_{k_\epsilon}$ engage and in $X^\epsilon$ no agent engages.
- If $v > b\epsilon$ then both in $X^\epsilon$ and $X^\epsilon$ all agents engage.

Observe that this result allows for characterizing the maximum engagement/minimum misinformation that can be obtained when the platform uses true error revelation mechanisms. In particular, to characterize maximum expected engagement (minimum expected misinformation), it suffices to focus on the best (worst) equilibrium $X^\epsilon (X^\epsilon)$ for every $\epsilon$ and take an expectation over $\epsilon$.

Let $\epsilon_k := (k - 1 + v)/b$ for $k \geq 1$. Observe that by the lemma given above for $\epsilon \in (\epsilon_k, \epsilon_{k+1}]$ at the best equilibrium agents in the $k$-core of the underlying graph engage. Thus, using the convention $\epsilon_0 = 0$ (and $A_0 = V$), the expected maximum engagement under true error revelation mechanism can be given as follows:

$$\sum_{k=0}^{\infty} \frac{\epsilon_{k+1} - \epsilon_k}{\alpha} |A_k|,$$

and the corresponding expected misinformation is given by:

$$\sum_{k=0}^{\infty} \frac{\epsilon_{k+1}^2 - \epsilon_k^2}{2\alpha} |A_k|.$$

Similarly, expected minimum engagement under this mechanism is $|V|\epsilon_1^2/\alpha$, and the corresponding minimum expected misinformation is given by $|V|\epsilon_1^4/2\alpha$. Note that in these expressions, the terms $\{\epsilon_k\}$ are functions of the disutility parameter $b$, as can be seen from the definition of these terms.

Thus, we conclude that if the platform were to reveal the true error, the maximum expected engagement achievable under this mechanism can be expressed in terms of the $k$-cores of the underlying network and $\{\epsilon_k\}$. Similarly, the minimum misinformation can be characterized in terms of the number of nodes in the underlying network and $\epsilon_1$. For the numerical examples of Section 4.3, following this observation we first find the $k$-cores of the network and then use them to compute the induced maximum expected engagement and minimum expected misinformation, as outlined above.