

# Identifying the Best Agent in a Network

Leonie Baumann<sup>‡</sup>

5th December 2017

*Abstract* This paper develops a mechanism for a principal to allocate a prize to the most valued agent when agents have a knowledge network. The principal does not know any agent's value but any two linked agents know each other's values. Agents compete for the prize and send costless private messages about their own value and the values of others they know to the principal. Agents can lie only to a certain extent and only lie if it increases their chances of winning the prize. A mechanism that determines each agent's chances of winning for any possible message profile is proposed. We show that with this mechanism, there exists an equilibrium such that the most valued agent wins with certainty if every agent has at least one link; if the network is a star or complete, then the most valued agent wins with certainty in every equilibrium.

*Keywords* network, knowledge network, information network, mechanism design, principal-agents problem, allocation problem

*JEL Classification Codes* C72, D82, D83

---

\*Cambridge INET-Institute, Faculty of Economics, University of Cambridge, United Kingdom. Email: lb714@cam.ac.uk.

<sup>†</sup>I am especially grateful to Anke Gerber and Ariel Rubinstein for comments and discussion. I also thank Matt Elliott, Sanjeev Goyal, Claudia Herresthal, Benny Moldovanu, Mikhail Safronov, participants in the Networks Reading Group at the Cambridge INET-Institute, in the 3rd Annual Conference on Network Science and Economics, the CTN Annual Workshop 2017, the ESEM 2017, the Vfs Annual Conference 2017, and seminar participants at the University of Málaga. While working on this paper, I received financial support from the University of Hamburg and the Cambridge INET-Institute.

# 1 Introduction

This paper considers a mechanism design problem in a network setting: A principal has to assign a prize to one agent from a set of heterogeneously valued agents. She does not know the exact value of any agent and she wishes to assign the prize to the most valuable. Every agent would like to get the prize and knows her own value for the principal. So far, this is a standard allocation problem in the mechanism design literature which has many real-world interpretations: A government wants to privatize a public service and has to select one firm out of many applying firms to provide the service from then on; an employer has to select one applicant out of all applicants for a position; a committee has to assign the prize to one individual among all nominees. Agents might have knowledge about the values of other applicants: two firms operating in the same field might have a good estimate about each others' costs; two people who have worked on a joint project might know each others' abilities. Such an information structure among agents can be interpreted as a *knowledge network* in which links between agents signify knowledge. The general concern of this paper is to propose a mechanism for the principal to identify the highest value agent with certainty in such a setting.

In the model introduced in this paper, agents exactly know their own value and the values of agents to whom they are linked in the knowledge network. We restrict our attention to knowledge networks in which every agent is linked to at least one other agent. The principal does not know any exact value, but the distribution of values and the network are common knowledge. Every agent sends a private costless message about her information to the principal. An agent's message contains a statement about her own value (*application*) and statements about the value of each of the agents she is linked to (*references*). Agents can lie about a true value, but only to a commonly known certain extent. This accounts for the intuition that statements must be to some extent credible and verifiable. For example, a scholar with a PhD in Economics cannot pass off as a PhD in Biology as her degree certificate specifies Economics as a field. However, it is possible to

lie about her specialization within the field of Economics. Likewise, a junior applicant cannot pass off as a senior, but it is possible to lie about certain abilities of the junior.

Before agents send their messages, the principal publicly announces a mechanism according to which she will select an agent for the prize given a profile of messages. This mechanism defines a probability of being selected for every agent for every message profile. Knowing the mechanism, agents engage in a static Bayesian game. They simultaneously choose and send their messages. Each agent selects her message such that it maximizes her expected probability of being selected given her own type and all other agents' strategies. We introduce a *truth-telling assumption*: If the true message maximizes an agent's expected probability of being selected, then she strictly prefers to tell the truth over lying. The solution concept applied to this static game among agents is Bayesian Nash equilibrium. Finally, given the profile of messages sent by the agents, the principal selects an agent according to the announced mechanism.

The paper proposes a mechanism which guarantees partial implementation for any knowledge network in which every agent is linked to at least one other agent. This means if every agent is linked to at least one other agent, then the proposed mechanism induces a static Bayesian game for which there exists an equilibrium strategy profile such that the principal selects the best agent with probability 1. Moreover, the mechanism achieves full implementation for knowledge networks which are complete or a star. This means if the knowledge network is complete or a star, then every equilibrium strategy profile is such that the principal selects the best agent with probability 1. However, full implementation is not achieved for every knowledge network; the circle network of four agents is an example for this.

The paper contributes to the literature on mechanism design for allocation, implementation and persuasion problems with one principal, multiple agents and (partially) verifiable private information. Most closely related papers in this field are Lipman and Seppi (1995), Glazer and Rubinstein (2001), Ben-Porath and Lipman (2012), and Kartik and Tercieux (2012), all of which assume that agents are symmetrically or fully informed about the

state of the world, and Deneckere and Severinov (2008), in which agents can prove something about their own type, but not about other agents' types. We augment this literature by accounting for a knowledge network among agents. This allows us to capture symmetric as well as diverse asymmetric information situations among agents. Moreover, it enables agents to not only prove something about their own type but also about other agents' types. In our setup, partial proofs about other agents' types are necessary to always identify the best agent.

The literature which combines mechanism design and network theory is scarce. Renou and Tomala (2012) study the implementation of social choice functions for different communication networks between agents and a mechanism designer; an agent's ex-ante information is independent of the network and her message space is determined by the designer. Dziubiński et al. (2016) analyze optimal network protection against an attacker when the network is unknown to the defender and imperfectly to the nodes in the network. The value of protecting a certain node depends on this node's network position. In our paper, the communication system between agents and the designer is fixed: every agent can only communicate to the principal. Moreover, an agent's value for the principal is independent of her network position. In our setup the network is common knowledge and determines the information structure among agents and their message space.

Note that we do not study strategic communication and information transmission between agents, but between agents who have a knowledge network and an uninformed principal who wants to extract information from the network. The literature who investigates strategic communication and information transmission among agents in communication networks includes Hagenbach and Koessler (2010), Garcia (2012), Galeotti et al. (2013), Acemoglu et al. (2014), Bloch et al. (2014), Patty and Penn (2014), Calvó-Armengol et al. (2015), Currarini and Feri (2015), Förster (2015), and Wolitzky (2015).

In section 2 and 3, we introduce the model and discuss the partial verifiability of information. In section 4, we construct the mechanism and provide examples to highlight the role of references and of the truth-telling assumption. In section 5, we first propose a strategy profile of agents, second, we

show that given the strategy profile the constructed mechanism selects the best agent with certainty, and third, we prove that the proposed strategy profile is a Bayesian Nash equilibrium of the static game among agents. In section 6, it is shown that if the knowledge network is complete or a star, then every Bayesian Nash equilibrium of the game induced by the mechanism is such that the best agent is selected with certainty. However, we also provide an example of a knowledge network for which there exists an equilibrium such that the best agent is not selected with certainty.

## 2 The Model

A principal has to assign a prize to one agent out of a set of agents  $N = \{1, \dots, n\}$  where  $n \geq 3$ . The prize is valuable to every agent and each agent is differently suited to receive the prize. The principal knows which properties the ideal candidate should have and agent  $i$ 's value for the principal is measured by her absolute distance,  $d_i$ , to the ideal. Thus an agent matching the ideal would have  $d_i = 0$ . Regarding the distribution of distances across agents, we assume that each agent's distance is an independent random draw from a distribution with strictly positive density  $f$  over the unit interval  $[0, 1]$ .<sup>1</sup> The profile of distances will be denoted by  $d = (d_1, \dots, d_n)$ . The lower  $d_i$ , the better agent  $i$  fits the principal's purpose. We restrict our attention to cases in which no two agents have the same distance. This means  $d_i \neq d_j$  for  $i \neq j$ .<sup>2</sup> The principal does not know the distance of any agent but only the distribution of distances. Every agent exactly knows her own distance.

Agents are organized in a network characterized by the set of links  $L$  which exist among them. A link between agent  $i$  and agent  $j$  is denoted by

---

<sup>1</sup>The assumption that distances have an upper bound is without loss of generality. The proposed mechanism could be easily adjusted to the case without an upper bound and all our results would still hold.

<sup>2</sup>Concentrating on cases in which no two agents have the same distance is for analytical convenience only. With this we exclude special cases which would complicate the analysis significantly, and would not add further insights. Although we do not show it in this paper, we strongly believe that our results continue to hold if the distances of any two agents can be the same.

$ij$ . We say that two agents  $i$  and  $j$  are linked if link  $ij \in L$ . The network is undirected, thus if  $ij \in L$ , then  $ji \in L$ . An agent  $j$  to whom agent  $i$  is linked is called a *neighbor* of agent  $i$ . The set of all neighbors of agent  $i$  is  $N_i := \{j \mid ij \in L\}$ . Links signify familiarity: Besides knowing her own exact distance, an agent also knows the exact distances of her neighbors. Regarding the distances of agents who are not her neighbors, agent  $i$  only knows the distribution of distances across these agents. We assume in the following that every agent has at least one neighbor. This is an important assumption as for identifying the best agent for all  $d$  it will be necessary that every agent is known to at least one other agent. Note that this assumption does not require the network to be connected. The network is common knowledge. This means everybody knows about which other agents  $j$  an agent  $i$  is informed.

For given  $d$  and  $L$ , agents do not only differ in their own distances but also in who their neighbors are and which distances their neighbors have. These aspects together determine an agent's type. Consequently, an agent  $i$ 's type  $t_i$  is defined as the tuple of her own distance and all her neighbors' distances,  $t_i := (d_i, (d_j)_{j \in N_i})$ , and the type profile as  $t := (t_1, \dots, t_n)$ .

After distances and thus types have realized, every agent sends a costless, private message about her information to the principal. The message from agent  $i$  is denoted by  $m_i$  and the profile of messages from all agents by  $m := (m_1, \dots, m_n)$ . The profile of messages from all agents except  $i$  will be denoted by  $m_{-i}$ . Agent  $i$ 's message contains statements about her own distance and about the distance of each of her neighbors. The statement about her own distance is  $m_{ii}$  and is called *application*; the statement about the distance of her neighbor  $j \in N_i$  is  $m_{ij}$  and is called agent  $i$ 's *reference* about  $j$ . Thus,  $m_i = (m_{ii}, (m_{ij})_{j \in N_i})$ . Agents can lie about their own distance and about a neighbor's distance but only to a certain extent: It is common knowledge that each statement  $m_{ik}$  must be an element of  $[\max\{0, d_k - b\}, \min\{d_k + b, 1\}]$  where  $b \in (0, \frac{1}{2})$  is the maximum possible lie about the distance of agent  $k$ . Hence, an agent's message space depends on her type.

The principal chooses a function  $\pi$  which specifies a probability  $\pi_i$  of being selected for every agent  $i$  for every possible  $m$ . More specifically, she

chooses  $\pi(m) = (\pi_1(m), \dots, \pi_n(m))$  such that  $\pi_i(m) \in [0, 1]$  for every  $i$  and  $\sum_{i \in N} \pi_i(m) = 1$  for every possible  $m$ . The fact that selection probabilities sum up to 1 given  $m$  accounts for the assumption that the principal has to assign the prize and cannot destroy it.<sup>3</sup> The principal publicly announces  $\pi$  and commits to it before agents send their messages. The function  $\pi$  is the mechanism which the principal designs and commits to and according to which the principal selects an agent when she receives message profile  $m$ .

Knowing  $\pi$ , agents engage in a static Bayesian game. The solution concept applied will be Bayesian Nash equilibrium. Each agent chooses a message strategy  $\hat{m}_i$  such that her message  $\hat{m}_i(t_i)$  given any of her possible types  $t_i$  maximizes her expected probability of being selected,  $\pi_i^e$ , given every other agent  $j$ 's message strategy,  $\hat{m}_j$ . The strategy profile will be denoted by  $\hat{m}$  and the profile of every agent's strategy except agent  $i$ 's by  $\hat{m}_{-i}$ . We introduce an assumption on an agent's preference for truth-telling. If the truthful message maximizes  $\pi_i^e$  for agent  $i$  of type  $t_i$ , then she tells the truth given  $t_i$ , this means  $\hat{m}_{ik}(t_i) = d_k$  for all  $k$ . On the other hand, we will say that an agent with type  $t_i$  is lying if  $\hat{m}_{ik}(t_i) \neq d_k$  for some  $k$ . Thus given her type an agent's first priority is to maximize her chances of being selected, and, second, she cares about truth-telling. After agents have chosen their strategies, distances and hence types realize. An agent of type  $t_i$  sends message  $m_i = \hat{m}_i(t_i)$  and the message profile the principal receives is  $m = \hat{m}(t)$ .

The model summarizes as follows. First, the principal chooses, announces and commits to  $\pi$ . Second, agents engage in the static Bayesian game. They simultaneously choose their message strategies, then their types are realized and the corresponding messages are privately send to the principal. Third, having received  $m$  the principal selects an agent according to  $\pi(m)$ .

The goal of this paper is to construct  $\pi$  such that the principal selects the best agent with probability 1. For any distance profile  $d$ , the *best* agent is agent  $i$  with  $d_i = \min d$ . The best agent will be called the *global minimum*.

In section 3, we briefly explain the partial verifiability of information in our model.

---

<sup>3</sup>This assumption implies that this is not a separable environment as defined by Jackson et al. (1994).

### 3 Partial Verifiability of Information

Before receiving the message profile  $m$ , the principal only knows the distribution of distances among agents. The message profile reveals partial information about the true distances of agents. The principal can infer from any message  $m_{ik}$  that  $d_k \in [\max\{0, m_{ik} - b\}, \min\{m_{ik} + b, 1\}]$  because she is aware of the maximum possible lie  $b$ . This means that if  $d_i < d_j$  for agent  $i$  and agent  $j$  and  $ij \in L$ , then  $i$  proves that she is better than  $j$  by sending  $m_{ij} - m_{ii} > 2b$ : The principal knows that  $d_i \leq m_{ii} + b$  and  $m_{ij} - b \leq d_j$ . If  $m_{ij} - m_{ii} > 2b$  which means that  $m_{ij} - b > m_{ii} + b$ , then the principal also knows that  $d_i \leq m_{ii} + b < m_{ij} - b \leq d_j$ . Thus, in this case, the principal is sure that  $i$  is better than  $j$ . For certain configurations of  $d_i$  and  $d_j$ ,  $i$  can always prove that she is better. For example, if  $b < d_i < d_j < 1 - b$ ,  $i$  can always send  $m_{ii}$  and  $m_{ij}$  such that  $m_{ij} - m_{ii} > 2b$ .

Agent  $i$  does not prove that she is better than agent  $j$ , if  $m_{ij} - m_{ii} < 2b$ . In this case, the principal cannot be sure that  $i$  is better than  $j$  because  $d_j < d_i$  with  $d_i, d_j \in [\max\{0, m_{ij} - b\}, \min\{m_{ii} + b, 1\}]$  is possible. For certain configurations of  $d_i$  and  $d_j$ ,  $i$  can never prove that she is better. For example, if  $d_i < d_j < b$ ,  $i$  cannot send  $m_{ii}$  and  $m_{ij}$  such that  $m_{ij} - m_{ii} > 2b$ . She creates the largest difference between her application and her reference about  $j$  by choosing  $m_{ii} = 0$  and  $m_{ij} = d_j + b$  for which  $m_{ij} - m_{ii} < 2b$ .

Observe that the partial verifiability of information due to the limits to lying implies that an agent  $i$  with type  $t_i$  can never imitate all other possible types  $t_j$ . Thus the revelation principle does not apply in this model.

In section 4, we first define the mechanism  $\pi$ . Second, we give examples to provide intuition for the mechanism and to highlight the role of references and of the truth-telling assumption.

### 4 The Mechanism

The following mechanism  $\pi$  defines the outcome of the static Bayesian game among agents for any message profile  $m$  with  $\pi_i(m)$  being the probability for agent  $i$  to be selected given  $m$ . We show in section 5 that  $\pi$  as defined

below induces a game for which there exists a Bayesian Nash equilibrium with strategy profile  $\hat{m}$  such that the principal selects the global minimum with probability 1 for any  $t$ . In section 6, we prove that every equilibrium of the game induced by  $\pi$  is such that the principal selects the global minimum with probability 1 for any  $t$ , if the knowledge network is complete or a star.

Definition 1 introduces  $\pi$ . The cardinality of a set  $x$  will be denoted by  $|x|$ .

**Definition 1.** Define  $\pi(m)$  for any message profile  $m$  as follows.

When the principal receives message profile  $m$ , she first identifies all agents who send the best application. Let these agents be the set  $B_1(m)$ . Formally,  $i \in B_1(m)$  if and only if  $m_{ii} = \min_{j \in N} m_{jj}$ .

Second, the principal selects a set  $B_2(m)$  from  $B_1(m)$ .

If the best application is greater than zero, then all agents in  $B_1(m)$  are in  $B_2(m)$ . Formally, if  $\min_{j \in N} m_{jj} > 0$ , then  $B_2(m) = B_1(m)$ .

If the best application is equal to zero, then the principal turns to the references about the agents in  $B_1(m)$  to construct  $B_2(m)$ . For each  $i \in B_1(m)$ , the principal identifies the worst reference about  $i$ :  $\bar{r}_i = \max_{j \in N_i} m_{ji}$ . Then, the principal determines who receives the least bad worst reference (the min-max reference) among agents in  $B_1(m)$ . This means she identifies every agent  $i$  for whom  $\bar{r}_i = \min_{k \in B_1(m)} \bar{r}_k$ . These agents then form the set  $B_2(m)$ . Formally, if  $\min_{j \in N} m_{jj} = 0$ , then  $i \in B_2(m)$  if and only if  $i \in B_1(m)$  and  $\bar{r}_i = \min_{k \in B_1(m)} \bar{r}_k$ .

Third, the principal selects a set  $B_3(m)$  from  $B_2(m)$ . An agent  $i \in B_2(m)$  is in  $B_3(m)$  if and only if at least one of the following two conditions is satisfied:

1) Agent  $i$ 's application conflicts with a reference about her or her reference about a neighbor conflicts with this neighbor's application, this means  $m_{ii} \neq m_{ji}$  or  $m_{ij} \neq m_{jj}$  for some  $j \in N_i$ .

2) Agent  $i$ 's message proves that she is better than each of her neighbors, this means  $m_{ij} - m_{ii} > 2b$  for all  $j \in N_i$ .

Having constructed  $B_1(m)$ ,  $B_2(m)$  and  $B_3(m)$ , the principal determines  $\pi_i(m)$  for each  $i \in N$ .

If  $B_3(m)$  is not empty, then the principal selects every  $i \in B_3(m)$  with  $\pi_i(m) = \frac{1}{|B_3(m)|}$ .

If  $B_3(m)$  is empty and  $B_2(m)$  is not a singleton, then the principal selects every  $i \in B_2(m)$  with  $\pi_i(m) = \frac{1}{|B_2(m)|}$ .

If  $B_3(m)$  is empty and  $B_2(m) = \{i\}$ , then the principal selects  $i$  with  $\pi_i(m) = 0$ . If there exists  $j \notin N_i$ , the principal selects every  $j \notin N_i$  with  $\pi_j(m) = \frac{1}{|N| - |N_i| - 1}$ . If every  $j$  is in  $N_i$  and there exists  $j$  with  $m_{ij} - m_{ii} > 2b$ , then the principal selects every  $j$  with this property with the same  $\pi_j(m) = p$  such that  $p|J| = 1$  where  $J = \{j \in N_i | m_{ij} - m_{ii} > 2b\}$ . If every  $j$  is in  $N_i$  and there does not exist  $j \in N_i$  with  $m_{ij} - m_{ii} > 2b$ , then the principal selects every  $j \in N_i$  with  $\pi_j(m) = \frac{1}{|N| - 1}$ .

Observe that  $B_2(m)$  is not empty for any  $m$ . There is always at least one agent who sends the best application among all agents such that  $B_1(m)$  is not empty. Moreover, either  $\min_i m_{ii} = 0$  or  $\min_i m_{ii} > 0$  and at least one agent  $i \in B_1(m)$  obtains  $\bar{r}_i = \min_{k \in B_1(m)} \bar{r}_k$ .

Next, we provide some intuition for the mechanism and highlight the importance of the truth-telling assumption. Assume the principal commits to  $\pi$ . Then, for any  $m$ , the principal first selects all agents with the best application for  $B_1(m)$ ; for  $B_2(m)$  the principal selects all agents from  $B_1(m)$ , if the best application is larger than zero; if the best application is zero, then the principal compares the references about agents in  $B_1(m)$  and selects those with the min-max reference for  $B_2(m)$ ; for  $B_3(m)$  the principal selects all agents from  $B_2(m)$  who prove that they are better than all their neighbors, or who conflict with some neighbor. For every equilibrium which we establish in sections 5 and 6, it will always be the case that the global minimum is in  $B_3(m)$ . The different cases in which  $B_3(m) = \emptyset$  serve as a punishment for agents who would deviate from equilibrium.

For the three examples below, assume  $n = 3$  and the line network  $L = \{12, 23\}$  as depicted in figure 1. In this network, 1 knows  $d_1$  and  $d_2$ ; 2 knows

$d_1$ ,  $d_2$  and  $d_3$ ; 3 knows  $d_2$  and  $d_3$ .

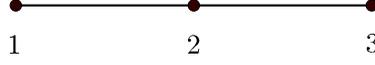


Figure 1: Line network with agents 1, 2 and 3.

**Example 1** Consider  $b = 0.2$  and the following realization of distances  $d = (0.5, 0.6, 0.4)$  as depicted in figure 2 below.

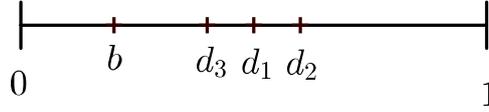


Figure 2:  $b = 0.2$  and  $d = (0.5, 0.6, 0.4)$  in the unit interval.

In this example, 3 is the global minimum. 1 and 3 both expect with positive probability to be the global minimum, and 2 knows that she is not the global minimum. Consider agents send the following message profile:

$$\begin{aligned} m_{11} &= d_1 - b = 0.3, \quad m_{12} = d_2 + b = 0.8; \\ m_{22} &= d_2 = 0.6, \quad m_{21} = d_1 = 0.5, \quad m_{23} = d_3 = 0.4; \\ m_{33} &= d_3 - b = 0.2, \quad m_{32} = d_2 + b = 0.8. \end{aligned}$$

1 and 3 exaggerate positively about themselves and negatively about 2, and 2 says the truth. The principal executes  $\pi$  such that  $B_3(m) = \{3\}$  and  $\pi_3(m) = 1$ . No agent has an incentive to deviate given the other agents' strategies: 1 and 3 maximize their expected probability of being selected and truth-telling would be strictly worse for each of them in expectation. 2 knows that she has no chance of being selected and tells the truth. We will be precise about the corresponding equilibrium strategy profile and provide a proper proof in the next section. For now, this example should only serve to illustrate certain properties of the model and the mechanism. Observe that given  $m$  it is possible to differentiate between agents on the basis of their applications only. Applications are sufficient to identify the best agent

in this case and  $B_1(m) = B_2(m)$  is justified.

**Example 2** Consider  $b = 0.2$  and  $d = (0.15, 0.3, 0.05)$  as in figure 3 below.

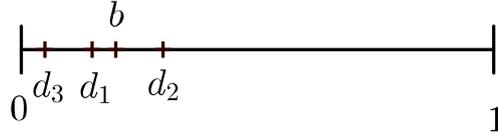


Figure 3:  $b = 0.2$  and  $d = (0.15, 0.3, 0.05)$  in the unit interval.

Let agents send the following message profile:

$$\begin{aligned} m_{11} &= 0, m_{12} = 0.5; \\ m_{22} &= 0.3, m_{21} = 0.15, m_{23} = 0.05; \\ m_{33} &= 0, m_{32} = 0.5. \end{aligned}$$

Again, 1 and 3 exaggerate positively about themselves and negatively about 2, and 2 says the truth. Executing  $\pi$  results into  $B_3(m) = \{3\}$  and  $\pi_3(m) = 1$ . No agent has an incentive to deviate, for the same reasons as above. Observe that given  $m$  it is not possible to differentiate between 1 and 3 on the basis of their applications only. Therefore the principal consults the references of 2 about 1 and 3. As 2 reports truthfully, the principal can now perfectly differentiate between 1 and 3. Thus,  $B_1(m)$  still includes all agents with the best application which is zero in this case, but  $B_2(m)$  only includes those who also get the least bad worst reference, in this case 3.

Note that 2 knows that she is selected with probability zero for any of her messages, given the strategies of 1 and 3, because she cannot send an application of zero. Thus if we did not introduce the truth-telling assumption, 2 would be indifferent between all of her possible messages. The truth-telling assumption lets 2 strictly prefer the truth over any other message. The principal can rely on the informativeness of 2's references because 2 is not lying. If 2 was indifferent between all her possible messages, there would exist other equilibrium message profiles for which 2 would be lying. Then,

the principal could not use 2's references to differentiate between 1 and 3 and she could not identify the best agent with certainty. Thus, if we dropped the truth-telling assumption, we would sacrifice the full implementation achieved in certain knowledge networks and could only sustain partial implementation.

The truth-telling assumption necessitates the introduction of  $B_3(m)$ . To see why, consider the next example.

**Example 3** Consider  $b = 0.2$  and  $d = (0.5, 0.3, 0.4)$  as depicted in figure 4.

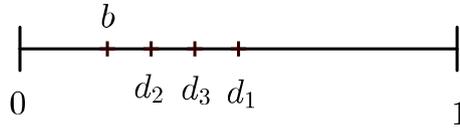


Figure 4:  $b = 0.2$  and  $d = (0.5, 0.3, 0.4)$  in the unit interval.

Thus 2 knows that she is the global minimum, and 1 and 3 know that they are not the global minimum. Let agents send the following message profile:

$$\begin{aligned} m_{11} &= 0.5, m_{12} = 0.3; \\ m_{22} &= 0.1, m_{21} = 0.7, m_{23} = 0.6; \\ m_{33} &= 0.4, m_{32} = 0.3. \end{aligned}$$

2 exaggerates positively about herself and negatively about 1 and 3, and 1 and 3 say the truth. Executing  $\pi$  results into  $B_3(m) = \{2\}$  and  $\pi_2(m) = 1$ . No agent has an incentive to deviate, by the same arguments as above. To understand the role of  $B_3(m)$ , assume for a moment that  $B_3(m)$  is not part of the mechanism, meaning that the principal ends the procedure after forming  $B_2(m)$  and selects agents in  $B_2(m)$  with equal probability. Then, for an agent to be selected, the principal only requires a best application, and, if this is zero, a min-max reference. In that case, given the above messages of 1 and 3, 2 is still selected with probability 1 if she deviates to the truth. Thus, given the truth-telling assumption, 2 has a strict incentive to deviate to the truth.

If, however, all three agents tell the truth, 3 has an incentive to deviate and to send  $m'_{33} = 0.2$ .

The purpose of  $B_3(m)$  is to force the global minimum to prove that she is better than all of her neighbors or to conflict with some neighbor. We eliminate her incentive to deviate to the truth, to support certain equilibria in which she does not say the truth. Given the truth-telling assumption which is necessary for full implementation with our mechanism,  $B_3(m)$  is needed to guarantee equilibrium existence.

In section 5, we propose a strategy profile of the game induced by  $\pi$  for  $n$  agents and for any network architecture. Subsequently, we show that the proposed strategy profile is a Bayesian Nash equilibrium of the game. The message profiles in the above examples correspond to this equilibrium strategy profile for  $n = 3$  in the line network given each respective  $d$ .

## 5 Partial Implementation

In this section, we first define a strategy profile  $\hat{m}$  for the static game following the announcement of  $\pi$ . Second, we show that if agents use  $\hat{m}$  and the principal executes  $\pi$ , then the principal selects the best agent with certainty for all  $t$ . Third, we show that  $\hat{m}$  is an equilibrium given  $\pi$ . Remember that by assumption, the following analysis and results only apply to knowledge networks in which every agent has at least one neighbor.

For the definition of  $\hat{m}$ , we introduce three type categories according to which we can classify all possible types.

An agent  $i$  who is better than all of her neighbors and who is not linked to all other agents will be called a *local minimum with partial information*. Thus, these are all agents with a type  $t_i$  such that  $|N_i| < n - 1$  and  $d_i < d_j$  for all  $j \in N_i$ . Such agents have *partial information* because they do not know every other agent. Agents 1 and 3 are both a local minimum with partial information in examples 1 and 2.

An agent  $i$  who is better than all of her neighbors and who is linked to all other agents will be called a *local minimum with full information*. Thus,

these are all agents with a type  $t_i$  such that  $|N_i| = n - 1$  and  $d_i < d_j$  for all  $j \in N_i$ . Such agents have *full information* because they know the distance of every other agent. Agent 2 is a local minimum with full information in example 3.

Note that the global minimum is always a local minimum. A local minimum with partial information might be the global minimum, and a local minimum with full information is always the global minimum.

The third type category includes every agent who has at least one neighbor who is better than her. Thus, these are all agents with a type  $t_i$  such that  $d_j < d_i$  for some  $j \in N_i$ . An agent in this category will be called *non-minimal*. Agent 2 is non-minimal in examples 1 and 2, and agents 1 and 3 are non-minimal in example 3.

In the following, an agent's message will only depend on her type category. Roughly speaking, we condense infinitely many types to just three types. At this point, we redefine an agent's type  $t_i$  as

$$t_i \in \{\text{local minimum with part. inf.}, \text{local minimum with full inf.}, \text{non-minimal}\}.$$

Intuitively, the strategy profile  $\hat{m}$  introduced below in definition 2 is such that every local minimum either proves that she is better than each of her neighbors or at least lies to the full extent about herself and her neighbors. Every non-minimal agent says the truth.

**Definition 2.** *The strategy profile  $\hat{m}$  with  $\hat{m}_i$  for every agent  $i$  is such that*

- *an agent who is non-minimal says the truth.*

*Formally, for  $t_i = \text{non-minimal}$*

$$\hat{m}_{ii}(t_i) = d_i \text{ and } \hat{m}_{ij}(t_i) = d_j \text{ for all } j \in N_i.$$

- *an agent  $i$  who is a local minimum with partial information sends the best possible application. Regarding her neighbors, she sends the worst possible references if some neighbor's distance is relatively close to her's, and she sends truthful references if all neighbors are sufficiently worse than her.*

Formally, for  $t_i = \text{local minimum}$  with partial information

$$\hat{m}_{ii}(t_i) = \max \{0, d_i - b\},$$

$\hat{m}_{ij}(t_i) = \min \{d_j + b, 1\}$  for all  $j \in N_i$ , if  $d_j - d_i \leq 2b$  for some  $j \in N_i$ ,  
and  $\hat{m}_{ij}(t_i) = d_j$  for all  $j \in N_i$ , if  $d_j - d_i > 2b$  for all  $j \in N_i$ .

- an agent who is a local minimum with full information sends the best possible application and the worst possible references if some neighbor's distance is sufficiently close to her's. If all her neighbors are sufficiently worse than her, she says the truth about herself and her neighbors.

Formally, for  $t_i = \text{local minimum}$  with full information

$\hat{m}_{ii}(t_i) = \max \{0, d_i - b\}$  and  $\hat{m}_{ij}(t_i) = \min \{d_j + b, 1\}$  for all  $j \in N_i$ ,  
if  $d_j - d_i \leq 2b$  for some  $j \in N_i$ ,

and  $\hat{m}_{ii}(t_i) = d_i$  and  $\hat{m}_{ij}(t_i) = d_j$  for all  $j \in N_i$ , if  $d_j - d_i > 2b$  for all  $j \in N_i$ .

**Proposition 5.1.** *If agents use strategy profile  $\hat{m}$  and the principal selects an agent according to  $\pi$ , then the principal selects the global minimum with probability 1 for every  $t$ .*

*Proof.* Let agent  $g$  be the global minimum. Remember first that  $g$  is a local minimum. Given  $\hat{m}$ , every local minimum sends her best possible application unless she has full information and all other agents' distances are far in which case she says the truth. This implies that if  $m_{gg} > 0$ , then  $m_{gg}$  is the unique best application and  $B_1(m) = B_2(m) = \{g\}$ . If  $m_{gg} = 0$ , then  $g \in B_1(m)$  and  $g$  uniquely receives the min-max reference among all agents in  $B_1(m)$  because the neighbors of every local minimum are non-minimal and send truthful references about all agents in  $B_1(m)$ . Thus  $B_2(m) = \{g\}$  as well. It is left to show that  $g$  is also in  $B_3(m)$ . The global minimum lies about her neighbors if some neighbor's distance is close. Then  $g$  conflicts with her neighbors as her neighbors are non-minimal and send truthful applications. If all neighbors are far, then the global minimum says the truth about her neighbors and proves to be better than all of her neighbors. Hence,  $g$  is also in  $B_3(m)$  and  $\pi_g(m) = 1$  for all  $t$ .  $\square$

Next it is shown that the strategy profile  $\hat{m}$  is a Bayesian Nash equilibrium of the static game following the announcement of  $\pi$ . First, we define the equilibrium conditions.

**Definition 3.** *The strategy profile  $\hat{m}$  is an equilibrium if, given  $\hat{m}_{-i}$ , for every agent  $i$  and for all  $t_i$*

1.  $\hat{m}_i(t_i)$  maximizes  $\pi_i^e$ , and
2.  $\hat{m}_i(t_i)$  is the truth if the truth maximizes  $\pi_i^e$ .

If  $\hat{m}_i$  is such that both of the above conditions are satisfied for all  $t_i$  for every agent  $i$ , then no agent  $i$  with preferences as outlined in section 2 has an incentive to deviate from  $\hat{m}_i$  given all other agents' strategies, and  $\hat{m}$  is a Bayesian Nash equilibrium strategy profile.

**Proposition 5.2.** *The strategy profile  $\hat{m}$  is an equilibrium strategy profile of the static Bayesian game induced by  $\pi$ .*

*Proof.* We show in turn that, given  $\hat{m}_{-i}$ ,  $\hat{m}_i(t_i) = m_i$  maximizes  $\pi_i^e$  and is the truth if the truth maximizes  $\pi_i^e$  for  $t_i = \text{local minimum with full information}$ ,  $t_i = \text{local minimum with partial information}$ , and  $t_i = \text{non-minimal}$ .

Consider first that agent  $i$  is a local minimum with full information. Then agent  $i$  knows that she is the global minimum and that  $\pi_i(m) = 1$ . Trivially, there does not exist another message  $m'_i$  which increases  $\pi_i^e$ . If  $m_i$  is not the truth, then the distance of some neighbor  $j \in N_i$  is close to  $d_i$ . By unilaterally deviating to the truth  $m'_i$ , agent  $i$  would not conflict with any of her neighbors anymore because her neighbors are non-minimal and say the truth. At the same time, agent  $i$  would not prove to be better than all of her neighbors because  $d_j - d_i \leq 2b$  for some  $j \in N_i$  and hence  $i \notin B_3(m'_i, m_{-i})$  but  $B_2(m'_i, m_{-i}) = \{i\}$ . Then agent  $i$  would be selected with probability 0 and deviating to the truth would strictly decrease  $\pi_i^e$ .

Consider second that agent  $i$  is a local minimum with partial information. Then agent  $i$  expects with positive probability to be the global minimum and

that  $\pi_i(m) = 1$ . If she is the global minimum,  $m_i$  clearly maximizes  $\pi_i$ . If she is not the global minimum, then  $\pi_i(m) = 0$  and another local minimum agent  $g$  who is not her neighbor is the global minimum. For any other  $m'_i \neq m_i$ , first,  $g$  still sends the overall best application because  $m_{ii}$  is already  $i$ 's best application, second,  $g$  still receives the min-max reference among agents in  $B_1(m'_i, m_{-i})$  because  $i$  can neither change the references about herself nor about  $g$  or any other agent in  $B_1(m)$  who are all local minima, and third,  $g$  still conflicts with every neighbor  $j \in N_g$  for whom  $d_j - d_g \leq 2b$ . Hence for any  $m'_i \neq m_i$ ,  $B_3(m'_i, m_{-i}) = \{g\}$  and  $m_i$  also maximizes  $\pi_i$ , if agent  $i$  is not the global minimum. Thus  $m_i$  maximizes  $\pi_i^e$  for agent  $i$ .

Next, observe that  $\pi_i^e$  strictly decreases if  $m_i$  is not the truth and agent  $i$  deviates to the truth  $m'_i$ . Agent  $i$  expects with strictly positive probability that she is the global minimum and that there exists another local minimum  $j$  with  $d_j \in (d_i, d_i + b)$  who sends  $m_{jj} = \max\{0, d_j - b\}$ . Suppose this is indeed the case. If  $d_i > 0$  and  $i$  deviates to  $m'_{ii} = d_i$ , then her application is worse than  $j$ 's and  $B_3(m'_i, m_{-i}) \neq \emptyset$  but  $i \notin B_3(m'_i, m_{-i})$  such that  $\pi_i(m'_i, m_{-i}) = 0$ . If  $d_i = 0$ , agent  $i$  only does not say the truth in case she has a neighbor  $k$  with  $d_k - d_i \leq 2b$ . By deviating to the truth, agent  $i$  still sends the best application and receives the min-max reference but neither conflicts with a neighbor nor proves better than all her neighbors. Thus  $B_3(m'_i, m_{-i}) = \emptyset$  and  $B_2(m'_i, m_{-i}) = \{i\}$  such that  $\pi_i(m'_i, m_{-i}) = 0$ .

Consider third that agent  $i$  is non-minimal. Agent  $i$  knows that she is not the global minimum  $g$  and that  $\pi_i(m) = 0$ . We show that there does not exist a message  $m'_i$  such that  $\pi_i^e(m'_i, m_{-i}) > 0$ .

First, observe that there does not exist  $m'_i$  such that  $i \notin B_2(m'_i, m_{-i})$  and hence  $i$  never gets selected with  $\pi_i(m'_i, m_{-i}) > 0$  as a member of  $B_2(m'_i, m_{-i})$  or  $B_3(m'_i, m_{-i})$  for any  $m'_i$ . If  $d_i > b$ , then agent  $i$ 's best feasible application is strictly worse than the best application of the global minimum and hence  $i \notin B_1(m'_i, m_{-i})$  for any  $m'_i$ . If  $d_i \leq b$ , then agent  $i$  can choose  $m'_{ii} = 0$  and, consequently, will be in  $B_1(m'_i, m_{-i})$ . However,  $i \notin B_2(m'_i, m_{-i})$  for any  $m'_i$ : If  $g$  is a neighbor of  $i$ , then the maximum reference about  $i$  is  $\bar{r}_i = d_i + b$  and  $i$  can increase the maximum reference about  $g$  to at most  $m'_{ig} = d_g + b$  which

is still strictly less than  $\bar{r}_i$ . If  $g$  is not a neighbor of  $i$ , then  $i$  cannot increase the maximum reference about  $g$  and  $\bar{r}_g = d_g < \bar{r}_i \in \{d_i, d_i + b\}$  for any  $m'_i$ .

Second, observe that there does not exist  $m'_i$  such that  $i$  gets selected with  $\pi_i(m'_i, m_{-i}) > 0$  if  $B_3(m'_i, m_{-i}) = \emptyset$  and  $B_2(m'_i, m_{-i}) = \{j\}$  where  $j \neq i$ . For agent  $j$  to be in  $B_2(m'_i, m_{-i})$  and not in  $B_3(m'_i, m_{-i})$ ,  $j$  must send the best application, must receive the min-max reference among all agents in  $B_1(m'_i, m_{-i})$  in case the best application is zero, and must not conflict with any neighbor where  $m_{jk} - m_{jj} \leq 2b$  for some  $k \in N_j$ , given  $(m'_i, m_{-i})$ . The only candidate agents  $j$  to be in  $B_2(m'_i, m_{-i})$  are those who did already sent  $m_{jj} = \min_{k \in N} m_{kk}$  before any deviation of agent  $i$ . These are the global minimum  $g$  and other local minima  $l$  with  $d_l \leq b$ . Before any deviation of  $i$ ,  $g$  and every  $l$  conflict with each of their neighbors  $k \in N_j$  for whom  $m_{jk} - m_{jj} \leq 2b$  for  $j = l, g$ . In order for  $j \in B_2(m'_i, m_{-i})$  and  $j \notin B_3(m'_i, m_{-i})$ ,  $i$  must deviate to  $m'_i$  such that  $j$  does not conflict any more with any of her neighbors  $k \in N_j$  and  $m_{jk} - m_{jj} \leq 2b$  for some  $k$ . For such  $m'_i$  to exist,  $i$  must be a neighbor of  $j$  and  $m_{ji} - m_{jj} \leq 2b$ . Assume such  $m'_i$  exists. If there exists some agent  $k \neq i$  who is not a neighbor of  $j$ , then still  $\pi_i(m'_i, m_{-i}) = 0$ . If every agent  $k \neq i$  is a neighbor of  $j$  and there is some  $k$  for whom  $m_{jk} - m_{jj} > 2b$ , then still  $\pi_i(m'_i, m_{-i}) = 0$ . If every agent  $k \neq i$  is a neighbor of  $j$  and  $m_{jk} - m_{jj} \leq 2b$  for all  $k$ , then such  $m'_i$  cannot exist: By assumption,  $|N_j| \geq 2$  because  $|N| \geq 3$ . So even if agent  $i$  chooses a message such that she does not conflict with agent  $j$ , there is at least one other agent  $k$  with whom agent  $j$  is conflicting and  $j$  cannot be in  $B_2(m'_i, m_{-i})$  without being in  $B_3(m'_i, m_{-i})$ .

Thus,  $\pi_i^e(m'_i, m_{-i}) = 0$  for every  $m'_i \neq m_i$  and the true message  $m_i$  maximizes  $\pi_i^e$ .

Then  $\hat{m}$  is an equilibrium. □

## 6 Full Implementation

In this section, we first show that for a complete network in which every agent has full information or for a star network in which one agent has full

information every equilibrium of the Bayesian game induced by  $\pi$  is such that the principal selects the global minimum with probability 1. Second, we show that the circle network with four agents is an example for which the principal does not select the global minimum with probability 1 in every equilibrium given  $\pi$ .

The network is complete if  $ij \in L$  with  $i \neq j$  for all  $i \in N$  and all  $j \in N$ . The network is a star if there exists one agent  $i \in N$  such that  $ij \in L$  for all  $j \in N$  and  $jk \notin L$  for all  $k \neq i$ . We will call the agent who is linked to every other agent in the star network the center  $c$ . The network is a circle with four agents if it takes the following form:

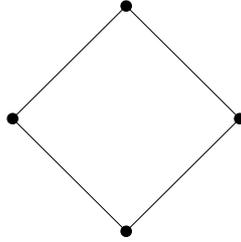


Figure 5: Circle network with four agents.

**Proposition 6.1.** *If the network is complete, then every equilibrium of the Bayesian game induced by  $\pi$  is such that the principal selects the global minimum with probability 1.*

*Proof.* Let the network be complete such that every agent has full information. Thus the global minimum agent  $g$  knows  $t$ ,  $m$ , and  $\pi(m)$ . Suppose to the contrary that the principal has committed to  $\pi$  and that  $\hat{m}$  is an equilibrium strategy profile such that  $\hat{m}(t) = m$  and  $\pi_g(m) < 1$  for  $t$ .

If  $d_g > b$  and  $g$  deviates to  $m'_g$  with  $m'_{gg} = d_g - b$  and  $m'_{gj} \neq m_{jj}$  for some  $j \neq g$ , then consequently  $g$  uniquely sends the best application and conflicts with some neighbor. Hence,  $B_3(m'_g, m_{-g}) = \{g\}$  and  $\pi_g(m'_g, m_{-g}) = 1$ .

If  $d_g \leq b$  and  $g$  deviates to  $m'_g$  with  $m'_{gg} = 0$  and  $m'_{gj} > d_g + b$  such that  $m'_{gj} \neq m_{jj}$  for all  $j \neq g$ , then consequently  $g$  sends the best application and uniquely receives the min-max reference because  $g$ 's reference about every agent  $j \neq g$  is strictly greater than any reference about  $g$ . Moreover,  $g$

conflicts with her neighbors. Hence,  $B_3(m'_g, m_{-g}) = \{g\}$  and  $\pi_g(m'_g, m_{-g}) = 1$ .

Then, however,  $\hat{m}$  is not an equilibrium strategy profile because  $g$  can deviate to some  $m'_g$  which strictly increases  $\pi_g^e$ .  $\square$

**Proposition 6.2.** *If the network is a star, then every equilibrium of the Bayesian game induced by  $\pi$  is such that the principal selects the global minimum with probability 1.*

*Proof.* Let the network be a star such that the center agent has full information and every other agent only knows the center agent. First, we show by contradiction that there does not exist an equilibrium strategy profile  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) = \emptyset$ . Second, we show by contradiction that every equilibrium strategy profile  $\hat{m}$  is such that  $\hat{m}(t) = m$  and  $B_3(m) = \{g\}$  for all  $t$ . Given these two results, we can then conclude that every equilibrium strategy profile  $\hat{m}$  is such that  $\hat{m}(t) = m$  and  $\pi_g(m) = 1$  for all  $t$  which proves proposition 6.2.

First suppose that  $\hat{m}$  is an equilibrium strategy profile such that  $\hat{m}(t) = m$ ,  $c \in B_2(m)$  and  $B_3(m) = \emptyset$  for  $t$  where  $c$  is the center of the star. From  $c \in B_2(m)$  and  $B_3(m) = \emptyset$  follows that  $\pi_c(m) < 1$ . Since  $c$  has full information, she knows  $t$ ,  $m$  and  $\pi(m)$ .

If all  $j \neq c$  are also in  $B_2(m)$ , then  $\pi_c(m) = \frac{1}{n} \leq \frac{1}{3}$ . In this case  $c$  can deviate to  $m'_c$  with  $m'_{cc} = m_{cc}$ ,  $m'_{ck} \neq m_{kk}$  for exactly one  $k \neq c$  and  $m'_{cj} = m_{jj}$  for all  $j \neq k, c$ . After the deviation,  $B_3(m'_c, m_{-c}) = \{c, k\}$  and  $\pi_c(m'_c, m_{-c}) = \frac{1}{2} > \pi_c(m)$  if  $m_{cc} = m_{kk} > 0$  because  $c$  and  $k$  are now the only agents sending the best application and conflicting with a neighbor, or  $B_3(m'_c, m_{-c}) = \{c\}$  and  $\pi_c(m'_c, m_{-c}) = 1$  if  $m_{cc} = m_{kk} = 0$  because  $c$  is now the only agent sending the best application, receiving the min-max reference and conflicting with a neighbor.

If not all  $j \neq c$  are in  $B_2(m)$ , then  $c$  can deviate to  $m'_c$  with  $m'_{cc} = m_{cc}$ ,  $m'_{cj} \neq m_{jj}$  for  $j \notin B_2(m)$  and  $m'_{cj} = m_{jj}$  for all  $j \in B_2(m)$ . After the deviation,  $B_3(m'_c, m_{-c}) = \{c\}$  and  $\pi_c(m'_c, m_{-c}) = 1$ .

Then, however,  $\hat{m}$  cannot be an equilibrium strategy profile because if

$\hat{m}(t) = m$ ,  $c \in B_2(m)$  and  $B_3(m) = \emptyset$  for  $t$ , then  $c$  can deviate to some  $m'_c$  which strictly increases  $\pi_c^e$ . Hence, there does not exist an equilibrium strategy profile  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$ ,  $c \in B_2(m)$  and  $B_3(m) = \emptyset$ .

Second suppose that  $\hat{m}$  is an equilibrium strategy profile such that  $\hat{m}(t) = m$ ,  $c \notin B_2(m)$  and  $B_3(m) = \emptyset$  for  $t$ . Then there exists  $j \neq c$  such that  $j \in B_2(m)$  and  $\pi_j(m) \leq \frac{1}{2}$  because  $B_2(m)$  is never empty. As  $c$  has full information,  $c$  knows that  $\pi_c(m) = 0$  and must say the truth because  $\hat{m}$  is an equilibrium strategy profile. This implies that  $j$  also says the truth because she would otherwise conflict with  $c$  and the outcome would not be  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ . Thus,  $j$ 's message is such that  $m_{jj} = d_j$ ,  $m_{jc} = d_c$  and  $m_{jc} - m_{jj} \leq 2b$  and  $j$  expects with positive probability that  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ .

Assume first that  $d_j > 0$ . If indeed  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ , then  $j$  can deviate to  $m'_j$  with  $m'_{jj} < m_{jj}$  such that  $m'_{jj} \neq \hat{m}_{cj}(t_c)$  for all  $t_c$  and  $m'_{jc} = m_{jc}$ . The deviation results into  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$ . If it is not the case that  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ , then  $j$  does not do worse by deviating to such  $m'_j$ . If  $j \in B_3(m)$ , then  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$ . If  $B_2(m) = \{k\}$  where  $k \neq j, c$  and  $B_3(m) = \emptyset$ , then either  $B_2(m'_j, m_{-j}) = B_2(m) = \{k\}$  and  $\pi_j(m'_j, m_{-j}) = \pi_j(m)$  or  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$ . In all other remaining sub-cases which could occur in equilibrium  $\pi_j(m) = 0$  and  $j$  always weakly improves by deviating to  $m'_j$ . Remember, also for the following paragraphs, we showed earlier that  $B_2(m) = \{c\}$  and  $B_3(m) = \emptyset$  which could imply  $\pi_j(m) > 0$  does not occur in equilibrium.

Assume second that  $d_j = 0$  and  $d_c < 1$ . If indeed  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ , then  $j$  can deviate to  $m'_j$  with  $m'_{jj} = m_{jj}$  and  $m'_{jc} > m_{jc}$  such that  $m'_{jc} \neq \hat{m}_{cc}(t_c)$  for all  $t_c$ . The deviation results into  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$ . If it is not the case that  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ , then  $j$  does not do worse by deviating to such  $m'_j$ . If  $j \in B_3(m)$ , then  $j \in B_3(m'_j, m_{-j}) \subseteq B_3(m)$  and  $\pi_j(m'_j, m_{-j}) \geq \pi_j(m)$ . If  $B_2(m) = \{k\}$  where  $k \neq j, c$  and  $B_3(m) = \emptyset$ , then  $B_2(m'_j, m_{-j}) = B_2(m)$  and  $\pi_j(m'_j, m_{-j}) = \pi_j(m)$ . For all other remaining sub-cases which could occur in equilibrium

$\pi_j(m) = 0$  and thus  $j$  always weakly improves by deviating to  $m'_j$ .

Observe that if  $d_j = 0$  and  $d_c = 1$ , then  $m_{jj} - m_{jc} = 1 > 2b$  and  $j$  does not expect with positive probability that  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ .

Then, however,  $\hat{m}$  is not an equilibrium strategy profile because if  $j$  expects with positive probability that  $j \in B_2(m)$  and  $B_3(m) = \emptyset$ , then there always exists a deviation for  $j$  which strictly increases  $\pi_j^e$ . Hence, there does not exist an equilibrium strategy profile  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $c \notin B_2(m)$  and  $B_3(m) = \emptyset$ .

Thus, there is no equilibrium strategy profile  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) = \emptyset$ . This completes the first part of the proof.

This first result implies that for every equilibrium strategy profile  $\hat{m}$  and every  $t$  with  $\hat{m}(t) = m$  it is true that  $\pi_i(m) > 0$  if and only if  $i \in B_3(m)$  for all  $i \in N$ .

Next, we show by contradiction that any equilibrium strategy profile  $\hat{m}$  is such that  $\hat{m}(t) = m$  and  $B_3(m) = \{g\}$  for all  $t$  where  $g$  is the global minimum.

Suppose first that  $\hat{m}$  is an equilibrium strategy profile and that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{c\}$  for  $t$  where  $c$  is the global minimum. From  $B_3(m) \neq \{c\}$  follows that  $\pi_c(m) < 1$ . If  $d_c > b$ , then  $c$  can deviate to  $m'_c$  with  $m'_{cc} = d_c - b$  and  $m'_{cj} \neq m_{jj}$  for all  $j$  such that  $B_3(m'_c, m_{-c}) = \{c\}$  and  $\pi_c(m'_c, m_{-c}) = 1$ . If  $d_c \leq b$ , then  $c$  can deviate to  $m'_c$  with  $m'_{cc} = 0$  and  $m'_{cj} > d_c + b$  such that  $m'_{cj} \neq m_{jj}$  for all  $j$  and thus  $B_3(m'_c, m_{-c}) = \{c\}$  and  $\pi_c(m'_c, m_{-c}) = 1$ . Then, however,  $\hat{m}$  is not an equilibrium. Hence, there does not exist an equilibrium strategy profile  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{c\}$  where  $c$  is the global minimum.

Suppose second that  $\hat{m}$  is an equilibrium strategy profile and that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{j\}$  for  $t$  where  $j \neq c$  is the global minimum. From  $B_3(m) \neq \{j\}$  follows that  $\pi_j(m) < 1$ . Given  $B_3(m) \neq \{j\}$  where  $j \neq c$  is the global minimum,  $j$  is a local minimum with partial information and expects with positive probability that  $B_3(m) \neq \{j\}$  and that she is the global minimum.

Assume first that  $d_j > b$ . If indeed  $B_3(m) \neq \{j\}$  where  $j \neq c$  is the global

minimum, then  $j$  can deviate to  $m'_j$  with  $m'_{jj} = d_j - b$  and  $m'_{jc} \neq \hat{m}_{cc}(t_c)$  for all  $t_c$  such that  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$ . If it is not the case that  $B_3(m) \neq \{j\}$  where  $j \neq c$  is the global minimum, then  $j$  does never do worse by deviating to  $m'_j$ . If  $j \in B_3(m)$ , then  $j$  still sends the best application which is greater than zero after her deviation to  $m'_j$ . Thus,  $j \in B_3(m'_j, m_{-j}) \subseteq B_3(m)$  and  $\pi_j(m'_j, m_{-j}) \geq \pi_j(m)$ . If  $j \notin B_3(m)$ , then  $\pi_j(m) = 0$ . In this case deviating to  $m'_j$  always weakly increases  $\pi_j$ .

Assume second that  $d_j \leq b < d_c$  and define  $j$ 's deviation as  $m'_j$  with  $m'_{jj} = 0$  and  $m'_{jc} > d_j + b$  such that  $m'_{jc} \neq \hat{m}_{cc}(t_c)$  for all  $t_c$ .

If  $B_3(m) \neq \{j\}$  and  $c \in B_3(m)$  where  $j \neq c$  is the global minimum, then  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$  because  $j$  uniquely sends the best application after deviating.

If  $B_3(m) \neq \{j\}$  and  $c \notin B_3(m)$  where  $j \neq c$  is the global minimum, then  $\pi_c(m) = 0$ . This implies that  $m_c$  must be the true message since  $\hat{m}$  is an equilibrium strategy profile and  $c$  has full information. If  $m_c$  is the truth and  $j$  deviates to  $m'_j$ , then  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$  because  $j$  as the global minimum receives the min-max reference from  $c$ .

If it is not the case that  $B_3(m) \neq \{j\}$  where  $j \neq c$  is the global minimum, then  $j$  does never do worse by deviating to  $m'_j$ . If  $j \in B_3(m)$ , then  $j \in B_3(m'_j, m_{-j}) \subseteq B_3(m)$  and  $\pi_j(m'_j, m_{-j}) \geq \pi_j(m)$ . If  $j \notin B_3(m)$ , then  $\pi_j(m) = 0$  and  $j$  always weakly improves by deviating to  $m'_j$ .

Up to here, we have shown that there does not exist an equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{j\}$  where  $j \neq c$  is the global minimum and  $d_j > b$  or  $d_j \leq b < d_c$ . The proof strategy was to define a profitable deviation for agent  $j$ , if  $d_j > b$  or  $d_j \leq b < d_c$  and she expected with positive probability to be the global minimum and that  $B_3(m) \neq \{j\}$ .

We use a different proof strategy to show that there does not exist an equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{j\}$  where  $j \neq c$  is the global minimum and  $d_j < d_c \leq b$ . First, we show that every equilibrium  $\hat{m}$  and  $t$  is such that  $\hat{m}(t) = m$  where an agent  $j \neq c$  with  $d_j < d_c \leq b$  sends  $m_{jj} = 0$  and  $m_{jc} > d_j + b$ . From this, we derive that every equilibrium  $\hat{m}$  and  $t$  is such that  $\hat{m}(t) = m$  where  $c$  says the truth if there exists an agent  $j$

with  $d_j < d_c \leq b$ . If an agent  $j$  with  $d_j < d_c \leq b$  exists, then such an agent  $j$  is the global minimum. From these three observations, we can conclude that there does not exist an equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{j\}$  where  $j \neq c$  with  $d_j < d_c \leq b$  is the global minimum.

Suppose first that  $\hat{m}$  is an equilibrium and that  $\hat{m}(t) = m$  for  $t$  where agent  $j$  with  $d_j < d_c \leq b$  sends  $m_{jj} > 0$ . Given  $d_j < d_c \leq b$ , agent  $j$  expects with strictly positive probability that  $d_j < d_c \leq b < d_k$  for all  $k \neq c, j$ . If indeed  $d_j < d_c \leq b < d_k$  for all  $k \neq c, j$ , then  $c$  sends  $m_c$  such that  $B_3(m) = \{c\}$  in order to maximize  $\pi_c^e$  given  $\hat{m}_{-c}$ . Then, however,  $j$  can deviate to  $m'_j$  with  $m'_{jj} = 0$  and  $m'_{jc} > d_j + b$  with  $m'_{jc} \neq \hat{m}_{cc}(t_c)$  for all  $t_c$  such that  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1 > \pi_j(m)$ . If it is not the case that  $d_j < d_c \leq b < d_k$  for all  $k \neq j$  and if  $j \in B_3(m)$  which is necessary for  $\pi_j(m) > 0$ , then deviating to  $m'_j$  leads to  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j(m'_j, m_{-j}) = 1$  as well. Thus, deviating to  $m'_j$  strictly increases  $\pi_j^e$  such that  $\hat{m}$  is not an equilibrium. Hence, there is no equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  where agent  $j$  with  $d_j < d_c \leq b$  sends  $m_{jj} > 0$ .

Suppose second that  $\hat{m}$  is an equilibrium and that  $\hat{m}(t) = m$  for  $t$  such that agent  $j$  with  $d_j < d_c \leq b$  sends  $m_{jj} = 0$  and  $m_{jc} < d_j + b$ . In case that  $d_j < d_c \leq b < d_k$  for all  $k \neq c, j$ , every agent  $k$  says the truth. If some  $k$  did not say the truth, she would have to expect with strictly positive probability to be  $B_3(m)$  and that  $\pi_k(m) > 0$ . Suppose  $k \in B_3(m)$  and that  $\pi_k(m) > 0$  for  $k$  with  $d_c \leq b < d_k$ . Then  $c$  would always deviate to a message  $m'_c$  such that  $B_3(m'_c, m_{-c}) = \{c\}$ . Thus, there is no equilibrium such that  $k$  with  $d_k > b \geq d_c$  is in  $B_3(m)$  and hence agent  $k$  expects with probability 0 to be in  $B_3(m)$  and says the truth. Then, given  $m_{-c}$ ,  $c$  sends  $m_{cc} = 0$  and some  $m_{cj} > m_{jc}$  such that  $B_3(m) = \{c\}$  to maximize  $\pi_c^e$ . However,  $j$  can deviate to  $m'_j$  with  $m'_{jj} = 0$  and  $m'_{jc} > d_j + b$  with  $m'_{jc} \neq \hat{m}_{cc}(t_c)$  for all  $t_c$  such that  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j$  strictly increases. If it is not the case that  $d_j < d_c \leq b < d_k$  for all  $k \neq c, j$  and if  $j \in B_3(m)$ , then deviating to  $m'_j$  implies  $j \in B_3(m'_j, m_{-j}) \subseteq B_3(m)$  and weakly increases  $\pi_j$  as well. Thus, deviating to  $m'_j$  strictly increases  $\pi_j^e$  such that  $\hat{m}$  is not an equilibrium. Hence, there is no equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  where agent  $j$  with  $d_j < d_c \leq b$  sends  $m_{jj} = 0$  and  $m_{jc} < d_j + b$ .

Suppose third that  $\hat{m}$  is an equilibrium and that  $\hat{m}(t) = m$  for  $t$  such that agent  $j$  with  $d_j < d_c \leq b$  sends  $m_{jj} = 0$  and  $m_{jc} = d_j + b$ . In case  $d_j < d_c \leq b < d_k$  for all  $k \neq c, j$ , every  $k$  says the truth and, given  $m_{-c}$ ,  $c$  sends  $m_{cc} = 0$  and  $m_{cj} = m_{jc}$  such that  $B_3(m) = \{j, c\}$  to maximize  $\pi_c^e$ . Then, however,  $j$  can deviate to  $m'_j$  with  $m'_{jj} = 0$  and  $m'_{jc} > d_j + b$  with  $m'_{jc} \neq \hat{m}_{cc}(t_c)$  for all  $t_c$  such that  $B_3(m'_j, m_{-j}) = \{j\}$  and  $\pi_j$  strictly increases. If it is not the case that  $d_j < d_c \leq b < d_k$  for all  $k \neq j$  and if  $j \in B_3(m)$ , then deviating to  $m'_j$  implies  $j \in B_3(m'_j, m_{-j}) \subseteq B_3(m)$  and weakly improves  $j$  as well. Thus, deviating to  $m'_j$  strictly increases  $\pi_j^e$  such that  $\hat{m}$  is not an equilibrium. Hence, there is no equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  where agent  $j$  with  $d_j < d_c \leq b$  sends  $m_{jj} = 0$  and  $m_{jc} = d_j + b$ .

Thus, every equilibrium  $\hat{m}$  and  $t$  is such that  $\hat{m}(t) = m$  where an agent  $j \neq c$  with  $d_j < d_c \leq b$  sends  $m_{jj} = 0$  and  $m_{jc} > d_j + b$ . This implies that every equilibrium  $\hat{m}$  and  $t$  is such that  $\hat{m}(t) = m$  and  $\pi_c(m) = 0$  if there exists an agent  $j$  with  $d_j < d_c \leq b$  because the best application is zero and the worst reference about  $c$  is worse than any feasible reference about  $j$ . As  $c$  has full information and knows that  $\pi_c(m) = 0$ , she says the truth. Thus, every equilibrium  $\hat{m}$  and  $t$  is such that  $\hat{m}(t) = m$  where  $c$  says the truth, if there exists an agent  $j$  with  $d_j < d_c \leq b$ .

If there exists an agent  $j$  with  $d_j < d_c \leq b$ , then such an agent  $j$  is the global minimum  $g$ . Thus, for every equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and an agent  $j$  with  $d_j < d_c \leq b$  exists, the global minimum sends  $m_{gg} = 0$  and receives the min-max reference since  $c$  says the truth. Moreover, the global minimum conflicts with  $c$ . Thus,  $B_3(m) = \{g\}$ .

Hence, there does not exist an equilibrium  $\hat{m}$  and  $t$  such that  $\hat{m}(t) = m$  and  $B_3(m) \neq \{j\}$  where  $j \neq c$  with  $d_j < d_c \leq b$  is the global minimum.

Thus, if the knowledge network is a star, then every equilibrium  $\hat{m}$  and  $t$  is such that  $\hat{m}(t) = m$ ,  $B_3(m) = \{g\}$  and  $\pi_g(m) = 1$  where  $g$  is the global minimum.  $\square$

The mechanism  $\pi$  does not guarantee full implementation for every knowledge network. The following example identifies a network and an equilibrium

strategy profile  $\hat{m}'$  such that the principal does not select the global minimum with probability 1.

**Example 4** Let  $n = 4$  and let the network be a circle as shown in figure 6.

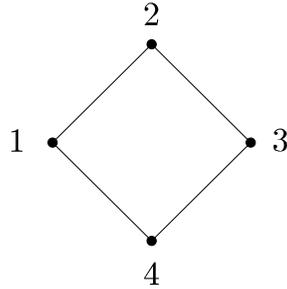


Figure 6: Circle network with  $n = 4$ .

Let  $\hat{m}'$  be such that

- a local minimum sends her best possible application. Regarding her neighbors, she sends the worst possible references if some neighbor's distance is relatively close to her's, and she sends truthful references if all her neighbors are sufficiently worse than her.

Formally, if  $d_i < d_j$  for all  $j \in N_i$ , then

$$\hat{m}'_{ii}(t_i) = \max \{0, d_i - b\},$$

$$\hat{m}'_{ij}(t_i) = \min \{d_j + b, 1\} \text{ for all } j \in N_i, \text{ if } d_j - d_i \leq 2b \text{ for some } j \in N_i,$$

$$\text{and } \hat{m}'_{ij}(t_i) = d_j \text{ for all } j \in N_i, \text{ if } d_j - d_i > 2b \text{ for all } j \in N_i.$$

- a non-minimal agent  $i$  who has  $d_i > b$  or a neighbor  $j$  with  $d_j = 0$  says the truth.

Formally, if  $d_i > d_j$  for some  $j \in N_i$  and if  $d_i > b$  or  $d_j = 0$  for some  $j \in N_i$ , then

$$\hat{m}'_{ii}(t_i) = d_i \text{ and } \hat{m}'_{ij}(t_i) = d_j \text{ for all } j \in N_i.$$

- a non-minimal agent  $i$  who has  $d_i \leq b$  and no neighbor  $j$  with  $d_j = 0$  sends her best possible application, a reference equal to  $b$  about every

neighbor  $j$  with  $d_j \leq b$  and a truthful reference about a neighbor with  $d_j > b$ .

Formally, if  $d_i > d_j$  for some  $j \in N_i$  and if  $d_i \leq b$  and  $d_j = 0$  for no  $j \in N_i$ , then

$$\hat{m}'_{ii}(t_i) = 0,$$

$$\hat{m}'_{ij}(t_i) = b \text{ for all } j \in N_i \text{ with } d_j \leq b,$$

$$\text{and } \hat{m}'_{ij}(t_i) = d_j \text{ for all } j \in N_i \text{ with } d_j > b.$$

**Lemma 6.3.** *If the network is a circle with  $n = 4$ , then strategy profile  $\hat{m}'$  is an equilibrium of the static Bayesian game induced by  $\pi$ .*

*Proof.* Let the network be a circle with  $n = 4$ . We show in turn that given  $\hat{m}'_{-i}$ ,  $\hat{m}'_i(t_i) = m_i$  maximizes  $\pi_i^e$  and is the truth if the truth maximizes  $\pi_i^e$  for  $t_i$  such that agent  $i$  is

1. a local minimum.
2. non-minimal and has  $d_i > b$  or a neighbor  $j \in N_i$  with  $d_j = 0$ .
3. non-minimal, has  $d_i \leq b$  and no neighbor  $j \in N_i$  with  $d_j = 0$ .

First, let  $t$  be such that agent  $i$  is a local minimum and  $\hat{m}'(t) = m$ . Suppose agent  $i$  is the global minimum  $g$ .

If  $j \notin N_i$  has  $d_j > b$ , then  $B_3(m) = \{i\}$  because  $m_{ii} = \min_{k \in N} m_{kk}$ ,  $\bar{r}_i = \min_{k \in N} \bar{r}_k$  and  $i$  conflicts with neighbor  $j \in N_i$  if  $m_{ij} - m_{ii} \leq 2b$ . This implies  $\pi_i(m) = 1$ .

If  $j \notin N_i$  has  $d_j \leq b$ , then  $i$  and  $j \notin N_i$  both send the best application 0. No  $j \in N_i$  is in  $B_2(m)$  because  $i$  sends reference  $m_{ij} = \min\{d_j + b, 1\}$  about each  $j \in N_i$  if one of them is sufficiently close to  $i$ . If both neighbors of  $i$  send truthful references, then  $B_3(m) = \{i\}$  and hence  $\pi_i(m) = 1$ . If a neighbor  $j \in N_i$  sends references equal to  $b$ , then  $B_3(m) = \{i, j\}$  for  $j \notin N_i$  and  $\pi_i(m) = \frac{1}{2}$ . In the latter case, there does not exist  $m'_i \neq m_i$  such that  $j \notin B_3(m'_i, m_{-i})$  for  $j \notin N_i$  and thus such that  $\pi_i(m'_i, m_{-i}) > \pi_i(m)$ .

Suppose  $i$  is not  $g$ . Then  $j \notin N_i$  is  $g$  because the neighbors of  $i$  are non-minimal.

If  $d_i > b$ , or if  $d_i \leq b$  and both neighbors of  $i$  send truthful references, then  $B_3(m) = \{g\}$  and  $\pi_i(m) = 0$  and there does not exist  $m'_i \neq m_i$  such that  $B_3(m'_i, m_{-i}) \neq \{g\}$ . If  $d_i \leq b$  and a neighbor  $j \in N_i$  sends references equal to  $b$ , then  $B_3(m) = \{g, i\}$  and  $\pi_i(m) = \frac{1}{2}$  and there does not exist  $m'_i \neq m_i$  such that  $g \notin B_3(m'_i, m_{-i})$ . Hence, in each case, there exists no  $m'_i \neq m_i$  such that  $\pi_i(m'_i, m_{-i}) > \pi_i(m)$ .

Thus,  $\hat{m}'_i(t_i) = m_i$  maximizes  $\pi_i^e$  given  $\hat{m}'_{-i}$ .

Next, we show that if  $\hat{m}'_i(t_i) = m_i$  is not the truth, then deviating to the truth strictly decreases  $\pi_i^e$  given  $\hat{m}'_{-i}$ . A local minimum  $i$  with  $d_i > 0$  assigns a strictly positive probability to herself being  $g$  and  $j \notin N_i$  being a local minimum with  $d_j \in (d_i, d_i + b)$ . If  $i$  is  $g$  and  $j \notin N_i$  is a local minimum with  $d_j \in (d_i, d_i + b)$ , then  $i \in B_3(m)$  and  $\pi_i(m) > 0$ . In this case, a deviation by  $i$  to the truth results into  $i \notin B_1(m'_i, m_{-i})$  and  $B_3(m'_i, m_{-i}) = \{j\}$  with  $j \notin N_i$  and thus  $\pi_i(m'_i, m_{-i}) = 0$ . A local minimum  $i$  with  $d_i = 0$  only does not say the truth if  $d_j - d_i \leq 2b$  for some  $j \in N_i$ . Given  $m$ ,  $B_3(m) = \{i\}$  and  $\pi_i(m) = 1$ . If  $i$  deviates to the truth, then  $B_2(m'_i, m_{-i}) = \{i\}$  and  $B_3(m'_i, m_{-i}) = \emptyset$  and thus  $\pi_i(m'_i, m_{-i}) = 0$ .

Thus, if  $\hat{m}'_i(t_i) = m_i$  is not the truth, then deviating to the truth strictly decreases  $\pi_i^e$  given  $\hat{m}'_{-i}$ .

Second, let  $t$  be such that agent  $i$  is non-minimal and has  $d_i > b$  or a neighbor  $j \in N_i$  with  $d_j = 0$  and  $\hat{m}'(t) = m$ . Given  $m$ ,  $i \notin B_3(m) \neq \emptyset$  and thus  $\pi_i(m) = 0$ .

If  $d_i > b$ , then there is no  $m'_i \neq m_i$  such that  $i \in B_1(m'_i, m_{-i})$  because  $m_{gg} < d_i - b$ . Moreover, there does not exist  $m'_i \neq m_i$  such that  $B_2(m'_i, m_{-i}) = \{j\}$  with  $j \notin N_i$  and  $B_3(m'_i, m_{-i}) = \emptyset$ : If  $j$  is  $g$  for  $j \notin N_i$ , then  $B_3(m'_i, m_{-i}) = \{g\}$  for all  $m'_i \neq m_i$ . If  $j$  is not  $g$  for  $j \notin N_i$ , then  $j \notin N_i$  is never in  $B_2(m'_i, m_{-i})$  for any  $m'_i \neq m_i$  because  $g$  sends  $m_{gg} = \max\{0, d_g - b\}$  and  $m_{gj} = \min\{d_j + b, 1\}$  for  $j \in N_g$  if  $d_j - d_g \leq 2b$ .

If  $i$  has a neighbor  $j$  with  $d_j = 0$ , then this neighbor is  $g$ . There does not exist  $m'_i \neq m_i$  such that  $i \in B_2(m'_i, m_{-i})$  because  $m_{gg} = 0$  and  $m_{gi} = \min\{d_i + b, 1\}$  if  $d_i \leq 2b$ . Moreover,  $j \notin N_i$  is never in  $B_2(m'_i, m_{-i})$  for any  $m'_i \neq m_i$  because  $g$  sends  $m_{gg} = 0$  and  $m_{gj} = \min\{d_j + b, 1\}$  for  $j \in N_g$  if

$d_j - d_g \leq 2b$ .

Thus, in both cases there does not exist  $m'_i \neq m_i$  such that  $\pi_i(m'_i, m_{-i}) > 0$  and  $\hat{m}'_i(t_i) = m_i$  which is the truth maximizes  $\pi_i^e$  given  $\hat{m}'_{-i}$ .

Third, let  $t$  be such that agent  $i$  is non-minimal, has  $d_i \leq b$  and no neighbor  $j$  with  $d_j = 0$  and  $\hat{m}'(t) = m$ .

If  $g \notin N_i$  and  $d_g > 0$ , then at least one neighbor  $j$  of  $i$  and  $g$  sends references equal to  $b$  because  $d_j < d_i$  such that  $B_3(m) = \{g, i\}$  and  $\pi_i(m) = \frac{1}{2}$ . For any  $m'_i \neq m_i$ ,  $g \in B_3(m'_i, m_{-i})$  and thus no  $m'_i \neq m_i$  increases  $\pi_i$ . If  $g \notin N_i$  and  $d_g = 0$ , then  $B_3(m) = \{g\}$  and  $\pi_i(m) = 0$ . For any  $m'_i \neq m_i$ ,  $B_3(m'_i, m_{-i}) = \{g\}$  because  $m_{gg} = 0$  and the neighbors of  $i$  and  $g$  say the truth. If  $g \in N_i$  and  $d_g > 0$ , then  $i \notin B_3(m) \neq \emptyset$  and  $\pi_i(m) = 0$ . There does not exist  $m'_i \neq m_i$  such that  $i \in B_2(m'_i, m_{-i})$  because  $m_{gg} = 0$  and  $m_{gi} = d_i + b$ . Moreover, there is no  $m'_i \neq m_i$  such that  $j \in B_2(m'_i, m_{-i})$  for  $j \notin N_i$  because  $m_{gg} = 0$  and  $m_{gj} > d_g + b$ . Thus, there exists no  $m'_i \neq m_i$  such that  $\pi_i(m'_i, m_{-i}) > \pi_i(m)$  and  $\hat{m}'_i(t_i) = m_i$  which is not the truth maximizes  $\pi_i^e$  given  $\hat{m}'_{-i}$ .

Next, we show that if agent  $i$  deviates to the truth, then  $\pi_i^e$  strictly decreases given  $\hat{m}'_{-i}$ . Agent  $i$  assigns strictly positive probability to  $g \notin N_i$  and  $d_g > 0$ . If  $g \notin N_i$  and  $d_g > 0$ , then  $B_3(m) = \{g, i\}$  and  $\pi_i(m) = \frac{1}{2}$ . Deviating to the true message results into  $B_3(m'_i, m_{-i}) = \{g\}$  and  $\pi_i(m'_i, m_{-i}) = 0$  because  $m'_{ii} > m_{gg}$ . Thus, deviating to the truth strictly decreases  $\pi_i^e$ .  $\square$

Suppose the type realization in the circle network with  $n = 4$  is such that all  $i \in N$  have  $d_i \leq b$  and  $d_g > 0$  and that the principal has committed to  $\pi$  and agents use  $\hat{m}'$ . Then  $B_3(m) = \{g, j\}$  where  $j \notin N_g$  and  $\pi_g(m) = \frac{1}{2}$ . This case occurs with a strictly positive probability. Thus,  $\hat{m}'$  is an equilibrium strategy profile for the circle network with  $n = 4$  for which the principal does not select the global minimum with probability 1 given  $\pi$ .

We summarize the results from example 4 in the following proposition.

**Proposition 6.4.** *Let  $n = 4$  and the network be a circle. Then  $\hat{m}'$  is an equilibrium such that there exists  $t$  for which the principal does not select the*

*global minimum with probability 1 given  $\pi$ .*

Proposition 6.4 raises the question for which networks the mechanism  $\pi$  achieves full implementation. This is not answered in this paper. Our conjecture is that  $\pi$  guarantees full implementation for every network in which at least one agent has full information. This means that there exists one agent who is linked to every other agent.

## 7 Conclusion

This paper introduces a model in which a principal has to assign a prize to one agent out of a set of heterogeneously valued agents. The principal is uninformed in the sense that she only knows the distribution of values across agents. Every agent would like to get the prize and she exactly knows her own value for the principal. Moreover, an agent might have knowledge about other competing agents. The distribution of information across agents is described by their knowledge network. In addition to her own value an agent exactly knows the values of other agents she is linked to in the network. A link signifies familiarity between two agents, e.g. two researcher who are coauthors have knowledge about each other's abilities.

We propose a mechanism for the principal to identify the best agent in such an environment, when 1) agents send private costless messages to the principal containing statements about their own value (application) and about values of other agents they know (references), 2) agents can lie about true values only to a commonly known extent, and 3) prefer to tell the truth if they cannot increase their expected probability of being selected by lying. The mechanism specifies a probability of being selected for every agent for any possible message profile. Intuitively, the proposed mechanism encourages agents who have a positive expectation of being selected and who cannot prove better than all their neighbors to exaggerate positively about themselves. Truthful references are necessary for the principal to distinguish between different agents who are all close to the ideal. We show that if every agent has at least one neighbor, this mechanism induces a static Bayesian

game among agents for which there exists an equilibrium such that the principal selects the best agent with certainty. In this equilibrium, agents who have a positive expectation of being the best prove that they are better than all their neighbors or lie positively about themselves and negatively about their neighbors. Agents who know that they are not the best say the truth. Moreover, we show that if the knowledge network is complete or a star, the mechanism guarantees that every equilibrium is such that the best agent is identified with certainty. However, we also provide an example of a knowledge network for which the mechanism does not achieve full implementation. A characterization of all knowledge networks for which full implementation is attained and an analysis of which properties determine whether full implementation on a given network is possible or not would be interesting to pursue in future work.

The fact that in our model references of competitors are valuable to distinguish between agents close to the ideal suggests that also in real world allocation problems it might be helpful to request statements from applicants about their competitors to identify the best applicant. Of course, our model speaks to some real world allocation problems more than to others. For example, modeling an agent's value for the principal as her distance to the ideal is suitable when an agent's value is measured on a continuous space and the ideal point is known. In the real world this is, for example, the case when an individual's abilities and qualifications are summarized in a score. Such scores are sometimes used by committees to evaluate and rank candidates for scholarships or other rewards. We would expect limits to manipulating the "true" score, which are reflected in the  $b$  of our model. Usually points attributed to an individual must be well reasoned, but of course it is possible to either emphasize the positive or negative aspects and by that distort an individual's true value.

Our assumption that  $b$  is common knowledge and a candidate perfectly knows her own value and the ones of her "neighbors" does not account for existing uncertainties and partial knowledge in the real world. Such uncertainty about an agent's own value and about her neighbors' values is an important extension of the model to investigate. Moreover, the knowledge

network probably is common knowledge in only few real world situations. This leads us to another interesting alteration of the model: to allow the existence of a link to be private knowledge and agents to be silent about neighbors. Agents might want to conceal that they have knowledge about another neighbor or might not want to make a statement about some neighbor if it improves their own expected probability of being selected. It is unclear if with such extensions we can still find a mechanism to identify the best agent with certainty.

## References

- Acemoglu, D., Bimpikis, K. and Ozdaglar, A. (2014), ‘Dynamics of information exchange in endogenous social networks’, *Theoretical Economics* **9**(1), 41–97.
- Ben-Porath, E. and Lipman, B. L. (2012), ‘Implementation with partial provability’, *Journal of Economic Theory* **147**(5), 1689–1724.
- Bloch, F., Demange, G. and Kranton, R. (2014), ‘Rumors and social networks’, *PSE Working Papers n2014-15, halshs-00966234* .
- Calvó-Armengol, A., Martí, J. and Prat, A. (2015), ‘Communication and influence’, *Theoretical Economics* **10**(2), 649–690.
- Currarini, S. and Feri, F. (2015), ‘Information sharing networks in linear quadratic games’, *International Journal of Game Theory* **44**(3), 701–732.
- Deneckere, R. and Severinov, S. (2008), ‘Mechanism design with partial state verifiability’, *Games and Economic Behavior* **64**(2), 487–513.
- Dziubiński, M., Sankowski, P. and Zhang, Q. (2016), Network elicitation in adversarial environment, in ‘International Conference on Decision and Game Theory for Security’, Springer, pp. 397–414.
- Förster, M. (2015), ‘Strategic communication under persuasion bias in social networks’, *Available at SSRN 2610768* .  
**URL:** <https://ssrn.com/abstract=2610768>
- Galeotti, A., Ghiglino, C. and Squintani, F. (2013), ‘Strategic information transmission networks’, *Journal of Economic Theory* **148**(5), 1751–1769.
- Garcia, D. (2012), ‘Communication and information acquisition in networks’, *MPRA Paper No. 55481* .  
**URL:** <http://mpra.ub.uni-muenchen.de/55481/>
- Glazer, J. and Rubinstein, A. (2001), ‘Debates and decisions: On a rationale of argumentation rules’, *Games and Economic Behavior* **36**(2), 158–173.

- Hagenbach, J. and Koessler, F. (2010), ‘Strategic communication networks’, *The Review of Economic Studies* **77**(3), 1072–1099.
- Jackson, M. O., Palfrey, T. R. and Srivastava, S. (1994), ‘Undominated Nash implementation in bounded mechanisms’, *Games and Economic Behavior* **6**(3), 474–501.
- Kartik, N. and Tercieux, O. (2012), ‘Implementation with evidence’, *Theoretical Economics* **7**(2), 323–355.
- Lipman, B. L. and Seppi, D. J. (1995), ‘Robust inference in communication games with partial provability’, *Journal of Economic Theory* **66**(2), 370–405.
- Patty, J. W. and Penn, E. M. (2014), ‘Sequential decision making and information aggregation in small networks’, *Political Science Research and Methods* **2**(02), 243–271.
- Renou, L. and Tomala, T. (2012), ‘Mechanism design and communication networks’, *Theoretical Economics* **7**(3), 489–533.
- Wolitzky, A. (2015), ‘Communication with tokens in repeated games on networks’, *Theoretical Economics* **10**(1), 67–101.