Incentive Design on Networks*

Arup Daripa
Birkbeck University of London
a.daripa@bbk.ac.uk

Sandeep Kapur
Birkbeck University of London
s.kapur@bbk.ac.uk

Marco Pelliccia
Heriot-Watt University
m.pelliccia@hw.ac.uk

December 2022

Abstract

We consider a productive network formed by agents with heterogeneous privately-known productivity types. Under convex cost of link formation, bilateral links create endogenous externalities. We calculate the "rearrangement cost" imposed on the entire network by each subset of other agents an agent can link to, and the consequent social opportunity cost of such subsets. We construct a mechanism to implement the socially optimal network, which specifies transfers based on the calculated opportunity cost, and a allocation to each agent consisting of network links. Once agents form allocated links, network connections might reveal further information to an agent on types of other agents. As a robustness check we allow pairwise adjustments under full information at the stage after forming the allocated network. We show that the mechanism is strict ex post incentive compatible, implements the optimal network as a pairwise-stable network and does not have a budget deficit.

JEL CLASSIFICATION: D85

KEYWORDS: Convex cost of link formation, Rearrangement cost, Ex post network implementation

*We thank two anonymous referees and the associate editor for comments on a previous version.

1 Introduction

Productive activity is often the outcome of collaboration. Much of scientific research involves interaction among individuals or research units. Similarly, joint ventures are a common form of corporate organization for the production of goods and services. In this paper we study the problem of incentive design on productive networks and construct a mechanism to implement the socially optimal network.

We consider a network formed by a finite population of agents who differ in their innate productivity, which we call their "type." Each bilateral link generates output that is strictly increasing in the types of the two agents, and each agent receives an equal share of that output. Links are costly to form. Crucially, we assume that for any agent, the total cost of forming connections is increasing and convex in the number of links formed, that is, the marginal cost of link formation is increasing in the number of links.

Convexity of cost implies that any network formed on the basis of private incentives is typically not socially optimal. There may be direct inefficiencies in bilateral link formation between agents who differ in the number of existing links, implying a divergence between private and social optima.[1] A second reason for departure from social optimality in the presence of convex costs is more subtle, as it relates to the impact of a set of links on the entire network. With convex costs, when agent i connects to a set of other agents, the marginal cost of all agents in the set are driven up. These agents must then use this higher marginal cost to assess the desirability of every one of their other existing connections (simply because they can save this marginal cost by dropping any of those connections). Therefore an agent j forming a link with i might drop her link with k, who might, in turn, now find it optimal to form a link with ℓ, and so on. It follows that the impact of the formation of a link ripples across the entire network, generating both negative and positive effects across other agents. Agents i and j do not take into account such externalities when deciding to form a link, implying that privately incentive compatible networks may fail to be socially optimal.

These inefficiencies create a scope for incentive design over networks. The first task is to quantify the complex structure of externalities that can arise from links and calculate

[1] Suppose total output of a link between i and j is 1, shared equally. The marginal cost of i (with no existing links) is 0.1, and that for j is 0.6 (with some existing links). The link might be part of the socially optimal network (output 1 exceeds combined marginal costs of 0.7), but j would reject any such link.

the opportunity cost of links. Note that with convex cost, we cannot consider the impact of single links in isolation. Instead we must evaluate the externality generated by the entire set of connections an agent makes. A contribution of our paper is to elucidate a method to calculate the "rearrangement cost" imposed on the entire network by any set of connections of an agent, which precisely captures the externality. Adding direct costs to this, we get the opportunity cost of each such set.

We specify a mechanism that uses these calculations to specify transfers and network allocations to implement the socially optimal network as an equilibrium network, and show that the total transfers are non-positive, so that the "budget" (of zero) is not exceeded. A novel aspect of a mechanism that specifies a network allocation is that agents can adjust network links even after receiving an allocation. We clarify these notions below.

An agent i can potentially connect to any subset of other agents, which we call a "connector set" of i. In a direct mechanism, agents report privately-known types. Based on the reports, the designer calculates the socially optimal network, allocates the associated connector sets to each individual, and announces transfers for all possible connector sets. The transfers reflect the opportunity cost of such sets as calculated from reports. The agents then form the allocated links. However, they might adjust these links if new information arises *once a network is formed*. It seems reasonable to allow for the possibility that the agents learn the types of their partners, so that if an allocated link proves unprofitable given the true type of the relevant partner, it would be severed. Network connections may also facilitate learning types of agents beyond immediate partners. If the network outcome is subject to deviations from subsequent information discovery that arise *because of the presence of a network*, the mechanism cannot be considered robust. To address the problem, we specify an adjustment process under full information *after the formation of an allocated network*, adopting pairwise-stability (Jackson and Wolinsky, 1996) under full information (PWS-FI) as the network equilibrium notion.

We use the notion of strict ex post network implementation, which requires the following two conditions to hold for implementing any network G. First, the mechanism is strict ex post incentive compatible (Bergemann and Morris, 2005, 2009): for any agent, truthful reporting is a strict best response to similar reporting by others. Second, given truthful reporting by all agents, the network allocated is G and it is PWS-FI.

We show that the second condition holds when G is socially optimal. To show the first, we need to consider the network equilibrium that arises under any report profile. Following a report profile, transfers and allocation are specified. If the allocated network is PWS-FI, it is a network equilibrium. We show this is the case under truthful reporting. However, under misreports, the allocated network could have scope for changes. We specify a pairwise adjustment process starting from the allocated network and define the network equilibrium, which exists under every report profile (and consequent transfers/allocation). Given network equilibria that arise following reports, we then show that at the reporting stage, the mechanism is ex post incentive compatible.

Our work makes three contributions. We consider the question of mechanism design on a network, with consequent non-standard issues of network readjustments after the allocation is specified. Second, we consider convex costs which generate endogenous externalities. We calculate the rearrangement cost and the resulting opportunity cost imposed on the network by each connector set of an agent. Using these calculations, we specify a mechanism that internalizes the externality and the total transfers do not exceed the budget (of zero). Finally, strict ex post implementation implies a detail-free result. However, while this relaxes the distribution-reliance of standard mechanism design, we do follow the standard practice of focusing on the truthtelling equilibrium, so our implementation is partial.

1.1 Related literature

Our setting involves preference interdependence arising through the fact that output results from network links, therefore the payoff of any agent depends on types of linked partners. In this setting, we use the notion of ex post incentive compatibility (Bergemann and Morris, 2005) that achieves partial robust implementation. A detail-free full implementation approach is considered in Bergemann and Morris (2009), who show that this can be achieved if preference interdependence is not too strong. A more recent paper by Ollár and Penta (2017) shows an important route out of such restrictions while achieving full implementation. They consider strategic externalities, which are potentially manipulable. Strong strategic externalities give rise to multiplicity of equilibria. However, introducing relatively mild moment-based conditions on beliefs (thus departing a little from robustness of ex post implementation) Ollár and Penta (2017) show that the designer is able to weaken strategic externalities, ensuring uniqueness. Whether a related approach might be useful for full implementation in network design under convex costs is an interesting question that we hope to address in future.

We consider incentives of agents who form a network, which naturally relates to the literature on networks. However, we are not aware of any direct antecedents of a mechanism design approach on networks under convex costs.

A large literature discusses the impact of exogenous externalities in network settings.[2] In the context of strategic network formation models, externalities are usually defined as direct costs/benefits experienced by some agents as result of the formation/severance of links by other pairs. In our setting externality is endogenous, arising through convex cost.

Jackson and Wolinsky (1996) study a *co-author* model. This is a network formation model with negative externality as players' return from being connected to a specific partner is negatively proportional to their own degrees – the formation of a new link negatively impacts the payoff of any existing partner. As individual agents do not internalize this cost, private and social optimality differ. In view of such problems, Bloch and Jackson (2007) study the role of transfers between agents in the formation of networks in a complete information setting. They highlight how, in the presence of negative externalities, it is possible to induce the formation of efficient equilibrium networks by allowing transfers that are conditioned on the final network structure. In the setting of these papers, externalities of a link are exogenous and either a cost or a benefit on the payoff of agents not directly involved in the link. We consider a mechanism design exercise in an incomplete information scenario (with possible full-information adjustments after an allocation is specified). Further, externalities of any set of links are indirect impacts arising through optimal readjustment of links by agents not directly involved, and the impact could be positive for some agents and negative for others at the same time.

The next section sets out our model. Section 3 calculates the rearrangement cost and opportunity cost of connector sets. Section 4 characterizes the socially optimal network. Section 5 presents the mechanism and describes the network equilibrium that arises after the specification of allocation and transfers. Section 6 defines the notion of network implementation and presents the implementation result. Section 7 concludes.

[2] Jackson (2010) provides a succinct overview.

2 The model

There is a finite set of agents $N = \{1, \dots, n\}$. Agents privately observe their own type θ_i, drawn independently from $\Theta_i \equiv [\underline{\theta}, \overline{\theta}] \subseteq \mathbb{R}^+$. We use standard notation $\theta_{-i} \in \Theta_{-i} = \times_{j \neq i} \Theta_j$ and $\theta \in \Theta = \times_{i \in N} \Theta_i$ for profiles. As an agent is associated with own type, we use i and θ_i interchangeably.

Agents can form productive bilateral links with others. A network $G(N, L)$ comprises a set of nodes N and a set of links L. Links are undirected, so that for any pair $i, j \in N$ we have $ij \in L$ if and only if $ji \in L$. Agents cannot form links with themselves, so $ii \notin L$ for any $i \in N$. Let Γ be the set of all possible undirected networks of size $|N|$.[3] Starting with any network G, let $G - ij$ denote the network where a link between i and j is removed, and similarly, let $G + ij$ denote a network where this link is added.

Let $\Psi(N \backslash \{i\})$ be the power set of $N \backslash \{i\}$. This is the set of all subsets of agents that i could potentially connect to. We refer to its typical element ψ_i as a **connector set** of agent i.

Note that a connector set, which does not include i, is distinct from a **neighborhood** of i, which does. Agent i's neighborhood $\{i, \psi_i\}$ in G is the set $\{\psi_i \cup \{i\} : ij \in L \text{ for all } j \in \psi_i\}$. Clearly, one and only one connector set is associated with each neighborhood. The **degree** of agent i in network G is $d_i(G) = |\psi_i|$, the cardinality of the connector set of i in G.

A link between types θ_i and θ_j produces output $g(\theta_i, \theta_j) \in \mathbb{R}^+$ for each agent.[4] Here $g(\cdot, \cdot)$ is positive, strictly increasing, and symmetric in both arguments. Let $Y(\{i, \psi_i\}|\theta) = \sum_{j \in \psi_i} 2g(\theta_i, \theta_j)$ be the output generated by the neighborhood $\{i, \psi_i\}$ under type profile θ and let $Y(G|\theta) = \sum_{i \in N} Y(\{i, \psi_i\}|\theta)/2$ be the total output generated by a network G.

Forming links is costly. Let $C(k)$ denote the total cost for an agent forming $k \in 1, \dots, n-1$ links and assume $C(0) = 0$. Then $\mathrm{MC}_k \equiv C(k) - C(k-1)$ indicates the marginal cost for the k-th link. We assume increasing marginal cost, or $\mathrm{MC}_k \geqslant \mathrm{MC}_{k-1}$ for any k, with strict inequality for some values of k.

The **net social value** of network G is

$$V(G|\theta) = Y(G|\theta) - \sum_{i \in N} C(d_i(G)). \tag{2.1}$$

For a given profile θ, a **socially optimal network** $G^*(\theta)$ is the network architecture that

[3] Since each agent is a node in a network, we use the terms "agent" and "node" interchangeably.
[4] An asymmetric split would not change the results qualitatively.

maximizes the net social value,

$$G^*(\theta) = \arg\max_{G \in \Gamma} V(G|\theta). \tag{2.2}$$

Note that given a finite set of agents, any network is of finite size. Therefore $V(G)$ is well-defined and finite for any $G \in \Gamma$. Further, the set Γ is itself a finite set. Therefore a socially optimal network $G^*(\theta)$ exists for any $\theta \in \Theta$.

Agent i's utility in G given profile θ is $v_i(G|\theta) = Y(\{i, \psi_i\}|\theta)/2 - C(d_i(G)) + t_i$ where the first two terms define i's *valuation* while the third term is any *transfer* paid to agent i. The tuple $\langle N, (\Theta_i, v_i)_{i \in N} \rangle$ is common knowledge among agents.

2.1 Network stability without intervention

To see the impact of endogenous externalities arising through increasing marginal costs, we need a notion of network stability without intervention, then compare privately stable and socially optimal networks.

For this purpose we use the notion of pairwise stability (PWS) defined by Jackson and Wolinsky (1996). However, given that agents form links with others, and there is incomplete information, we need to specify the information environment in which PWS holds. Since we want a *benchmark* under private incentives, we assume that agents learn all types. This serves as a benchmark by removing any inefficiency due to information problems (e.g., two types who could add a mutually beneficial link but do not do so because of lack of information). Any inefficiency that remains is therefore due to the endogenous externalities through convex cost, which is the problem we address later.

We therefore adopt *PWS under full information* as the non-intervention benchmark of network stability. This is defined below.

Definition 1. *A network G is pairwise stable under full information (PWS-FI) if the following conditions hold*

(i) *If types θ_i and θ_j are linked, $v_i(G|\theta) \geqslant v_i(G - ij|\theta)$ and $v_j(G|\theta) \geqslant v_j(G - ij|\theta)$, and*

(ii) *If types θ_i and θ_j are not linked, if $v_i(G + ij|\theta) \geqslant v_i(G|\theta)$, then $v_j(G + ij|\theta) < v_j(G|\theta)$.*

In words, a network structure is PWS-FI if, having learnt the types of all other agents, no agent would gain from severing an existing link and no pair of agents would find it profitable to form an additional link.

In the presence of convex cost of link formation, the set of pairwise stable networks do not generally coincide with the set of socially optimal networks. We present two examples below.

2.2 Example 1

Suppose $g(\theta_i, \theta_j) = \sqrt{\theta_i + \theta_j}$. The social output from the link ij is $2\sqrt{\theta_i + \theta_j}$. Four agents draw types $\theta_1 = 4$, $\theta_2 = 3$, $\theta_3 = 2$, and $\theta_4 = 1$. The marginal cost of connections is $\mathrm{MC}_1 = 1$ and $\mathrm{MC}_2 = 6$. With these parameter values every agent derives positive net benefit from forming a link, but the high marginal cost of the second link is prohibitively large for all agents.

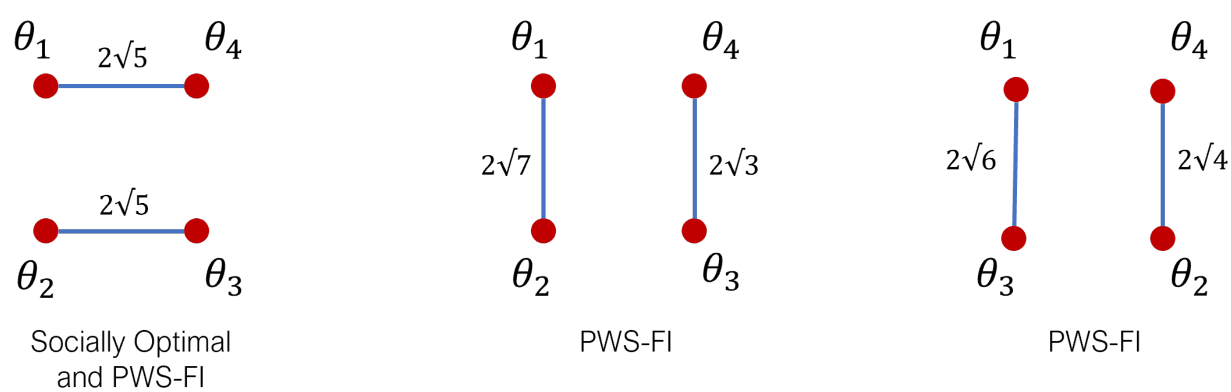

Figure 1: All three networks are PWS-FI but only the first from the left is socially optimal. The labels of the links indicate the total output generated by each link.

Without intervention, there are three PWS-FI configurations, each involving a pairs of links (see Figure 1). But, only one of them is socially optimal: the costs are the same for all three, but it is easy to check that aggregate output $2\sqrt{5} + 2\sqrt{5}$ for the first network from the left exceeds that for the other two.

2.3 Example 2

Suppose $g(\theta_i, \theta_j) = \theta_i \theta_j$. Four agents draw types $\theta_1 = 4$, $\theta_2 = 3$, $\theta_3 = 2.5$, $\theta_4 = 2$. The marginal cost of links are $\mathrm{MC}_1 = 2$, $\mathrm{MC}_2 = 7$, and $\mathrm{MC}_3 = 15$.

The socially optimal network is not PWS-FI (see Figure 2). Links between θ_1 and θ_2 and between θ_1 and θ_3 are socially and privately desirable. Beyond that, it is socially beneficial for θ_2 to connect to θ_4 instead of θ_3, as the net social value of the former, $2g(\theta_2, \theta_4) - MC_2 - MC_1 = 12 - 7 - 2 = 3$, exceeds the net social value of the latter, $2g(\theta_2, \theta_3) - 2MC_2 = 15 - 14 = 1$. The figure on the left shows the socially optimal network. However, from θ_2's individual perspective, a link with θ_3 has payoff $g(\theta_2, \theta_3) - MC_2 = 7.5 - 7 > 0$ while that with θ_4 has payoff $g(\theta_2, \theta_4) - MC_2 = 6 - 7 < 0$. Therefore starting from the socially optimal network, agent 2 would sever the link with agent 4, implying the socially optimal network is not PWS-FI. The figure on the right shows the PWS-FI network.

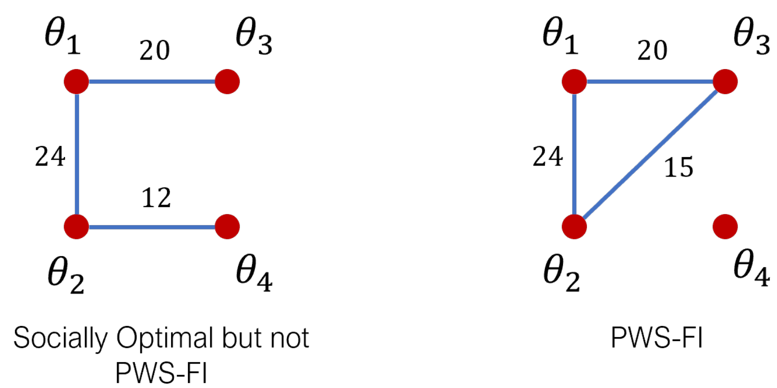

Figure 2: The network on the left is socially optimal but not PWS-FI: the link (θ_2, θ_4) is not profitable for type θ_2. On the other hand, link (θ_2, θ_3) on the right is individually profitable for both types involved but not socially optimal.

3 Rearrangement cost and opportunity cost of links

When would a particular connector set ψ_i of i be part of a socially optimal network? The social benefit of adding agent i connected to ψ_i is clear: it is simply the added output $Y(\{i, \psi_i\}|\theta)$. Calculating the opportunity cost of ψ_i is more complicated. This is what we do in this section.

A part of opportunity costs of ψ_i are direct link-formation costs incurred by i as well as by each agent in ψ_i. The more interesting component is the costs imposed on others when i connects to ψ_i through link changes across the network. We call this the *rearrangement cost* of ψ_i.[5]

To calculate the rearrangement cost, we proceed as follows. Exclude i from the set of agents. Suppose agents other than i are optimally arranged. Now ask, if we connect i with ψ_i, how much disruption would it cause for others? To answer this, still consider a network of agents other than i, but with the additional constraint that agents in ψ_i have an extra "shadow" link so that their marginal costs are raised by one extra degree. Derive the optimal network with the constraint in place. The difference in the value of the two optimal networks – unconstrained optimum of agents in N_{-i} versus their constrained optimum – gives us the rearrangement cost of ψ_i. Adding direct costs, we get the opportunity cost. Note that embedded in these calculations for i is that others are always arranged optimally. This is why when we compare the benefit minus opportunity cost of different connector sets, the one with the largest difference is the one that should be part of the socially optimal network, which answers the question posed at the top of the section.

3.1 The rearrangement cost of a connector set

Consider the problem of constructing any network starting from an empty one, in which all nodes are disconnected. Let d_i^0 denote the degree of an agent i before any link is formed. As we start from an empty network, $d_i^0 = 0$ for all $i \in N$. As $1, 2, \ldots, k$ links are formed for i, the link formation cost of i then rises as $C(1) = MC_1$, $C(2) = MC_1 + MC_2$, $\ldots$, $C(k) = MC_1 + \ldots + MC_k$.

Now consider an artificial case where we repeat the same process but where for a subset of agents, we have a virtual starting degree $d_i^0 = 1$: that is, these agents start with one external "shadow" link: links for which output contribution is ignored in the computation of the total output generated by the network. This implies that when considering links for any such agent, the cost of forming $1, 2, \ldots, k$ links are MC_2, $MC_2 + MC_3$, $\ldots$, $MC_2 +$

[5] Note that if marginal costs are constant, rearrangement cost associated with any set of links is zero. Indeed, in this case we can simply consider bilateral links: whether a link between two agents should optimally form does not depend on the presence of other links.

$\ldots + MC_{k+1}$. Note that for agent i with degree $d_i(G)$, we can write

$$MC_2 + \ldots + MC_{d_i(G)+1} = C(d_i(G) + 1) - C(1).$$

Let H denote the set of agents for whom $d_i^0 = 1$, while $d_i^0 = 0$ for agents in $N \backslash H$. Let $(G|H)^*$ denote the *constrained* socially optimal network. This is given by

$$(G|H)^* = \arg\max_{G \in \Gamma} \left(Y(G) - \sum_{i \in H} \Big(C(d_i(G) + 1) - C(1) \Big) - \sum_{i \in N \backslash H} C(d_i(G)) \right).$$

When determining the constrained optimal network $(G|H)^*$, costs are artificially raised for nodes in H. Crucially, note that once we have derived $(G|H)^*$, its net social value is determined with costs considered in the usual way. Thus $V((G|H)^*)$ is given by equation (2.1), with G replaced by $(G|H)^*$.

We now define formally the network rearrangement cost of a connector set ψ_i of agent i. Remove i from N and calculate the constrained socially optimal network denoted by $(G_{-i}|\psi_i)^*$. This is the optimal network in which the connection costs of agents of the connector set ψ_i are artificially burdened by an additional shadow connection. Clearly, the value of this network is weakly lower than that of the unconstrained optimal network without i, denoted by G^*_{-i}.[6] The difference between the two net values captures the loss in value from the constraint. This loss is precisely due to the difference in network connections across the network with or without the extra links for the set ψ_i, capturing the rearrangement cost of ψ_i.

Definition 2 (Rearrangement cost). *The rearrangement cost generated by the connector set ψ_i of agent i, denoted by $RC_i(\psi_i)$, is given by*

$$RC_i(\psi_i) \equiv V(G^*_{-i}) - V((G_{-i}|\psi_i)^*).$$

The rearrangement costs evaluates, for a network that excludes agent i, the reduction in the net value of the optimal network if a set ψ_i of agents are constrained to own a shadow link each.

[6]Alternatively, using the same notation as for constrained networks, the unconstrained socially optimal network without i is $G^*_{-i} = (G_{-i}|\varnothing)^*$, that is the constraint set is the empty set.

Note that two different agents (say i or j) connecting to the same subset of other agents (that is, the same connector set) can generate different rearrangement costs as the precise link changes depend on which other agents are in the network ($N \backslash \{i\}$ and $N \backslash \{j\}$ are different sets). Therefore the rearrangement cost depends on the identity of the removed agent, accounting for the subscript i in RC_i. However, the rearrangement cost $\mathrm{RC}_i(\psi_i)$ does not depend on the type θ_i of agent i. This feature is crucial to our mechanism, as it ensures that the social opportunity cost of any ψ_i does not depend on the type reported by i.

3.2 The social opportunity cost of a connector set

The opportunity cost sums rearrangement cost and direct costs to form new links.

Let G_{ψ_i} denote the network $(G_{-i}|\psi_i)^* \cup \{i, \psi_i\}$. This is the constrained optimal network with the shadow links replaced by actual links with i. The degree of any agent $j \in \psi_i$ in this network is $d_j(G_{\psi_i}) = d_j((G_{-i}|\psi_i)^*) + 1$.

Definition 3 (**Social opportunity cost**). *The social opportunity cost of the connector set ψ_i, denoted by $OC_i^S(\psi_i)$, is the sum of direct and indirect costs, given by*

$$OC_i^S(\psi_i) = C(|\psi_i|) + \sum_{j \in \psi_i} MC_{d_j(G_{\psi_i})} + RC_i(\psi_i). \tag{3.1}$$

The first two terms are the direct private costs of forming connections between i and agents ψ_i. Note that for i, the cost of connecting to the connector set ψ_i is $C(|\psi_i|)$. For $j \in \psi_i$, the relevant cost of connecting to i is the largest marginal cost $MC_{d_j(G_{\psi_i})}$. This is the amount j would save if the ij link were dropped. The third term is the rearrangement cost of the connector set ψ_i as derived above.

Below we calculate the rearrangement cost and opportunity cost for the examples introduced earlier.

3.2.1 Example 1 (continued)

Consider the connector sets for agent 1 in Example 1 (see Section 2.2). The socially optimal network requires links between the pairs: $\{(\theta_1, \theta_4), (\theta_2, \theta_3)\}$. Since the assumed structure of marginal costs prohibits any agent having more than one link in any optimal network, we can restrict attention to singleton connector sets, $\{\{\theta_2\}, \{\theta_3\}, \{\theta_4\}\}$. We also ignore the empty connector set, $\varnothing$, for which rearrangement cost is obviously zero.

Let us calculate $RC_1(\{\theta_2\})$, the rearrangement cost for $\psi_1 = \{\theta_2\}$.

• Remove θ_1 from the population. Calculate the resulting socially optimal network G^*_{-1}. This has θ_2 and θ_3 connected and leaves θ_4 disconnected (network on the left in Figure 3). The net social value of this network is $2\sqrt{\theta_2 + \theta_3} - 2MC_1 = 2\sqrt{5} - 2$.

• Now calculate the socially optimal network $(G^*_{-1}|\{\theta_2\})$. Remove θ_1 as before, but now constrain θ_2 to own a shadow link. With the increased marginal cost of a second link for θ_2, a link between θ_2 and θ_3 is now sub-optimal, and instead optimally replaced by a link between θ_3 and θ_4 (see the network on the right in Figure 3). This network has net social value $2\sqrt{\theta_3 + \theta_4} - 2MC_1 = 2\sqrt{3} - 2$.

• The rearrangement cost is $RC_1(\{\theta_2\}) = 2(\sqrt{5} - \sqrt{3})$.

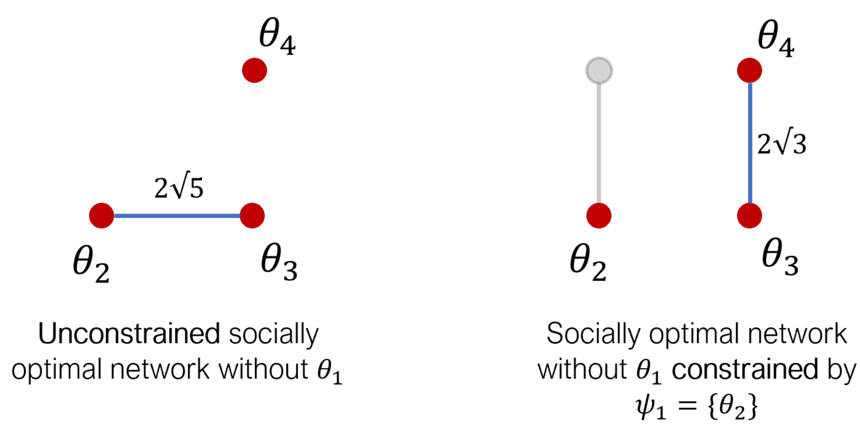

Figure 3: With θ_1 removed, it is optimal for θ_2 to connect with θ_3 (left). If θ_2 is constrained to form a shadow link (in gray), its connection with θ_3 becomes sub-optimal (dashed link), and the constrained optimal network would connect θ_3 and θ_4 instead. The rearrangement cost measures the loss in value due to this change in structure.

The opportunity cost of $\psi_1 = \{\theta_2\}$ is then simply the sum of the direct cost of linking θ_1 to θ_2 and the rearrangement cost calculated above:

$$OC^S_1(\{\theta_2\}) = 2MC_1 + RC_1(\{\theta_2\}) = 2 + 2(\sqrt{5} - \sqrt{3}).$$

Similar calculation shows that $OC^S_1(\{\theta_3\}) = 2 + 2(\sqrt{5} - \sqrt{4})$. The rearrangement cost of

$\psi_1 = \{\theta_4\}$ is zero, as θ_4 is disconnected in the optimal network without 1, so the value of the network is unaffected by the imposition of a shadow link on θ_4. The opportunity cost for this connector set then simply coincides with the direct costs of link formation: we have $OC_1^S(\{\theta_4\}) = 2$.

3.2.2 Example 2 (continued)

In Example 2, the socially optimal network has links between the pairs $\{(\theta_1, \theta_2), (\theta_1, \theta_3), (\theta_2, \theta_4)\}$. We focus here on the connector sets of agent 1 that have at most two elements and ignore the empty connector set.

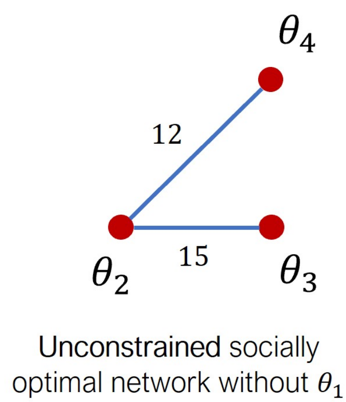

Figure 4

Consider the connector set $\psi_1 = \{\theta_2\}$. In the absence of agent 1, the unconstrained optimal network has θ_2 connected to θ_3 and to θ_4, with net social value $2(\theta_2\theta_3 + \theta_2\theta_4) - 2MC_1 - MC_1 - MC_2 = 14$ (see Figure 4). When θ_2 is constrained to own one shadow link, the link between θ_2 and θ_4 is optimally replaced by a link between θ_3 and θ_4, reducing net social output to $2(\theta_2\theta_3 + \theta_3\theta_4) - 2MC_1 - MC_1 - MC_2 = 12$ (see diagram on the left in Figure 5). The rearrangement cost then is $RC_1(\{\theta_2\}) = 14 - 12 = 2$ and the associated opportunity cost is

$$OC_1^S(\{\theta_2\}) = C(1) + MC_2 + RC_1(\{\theta_2\}) = 2 + 7 + 2 = 11.$$

Consider, next, the costs associated with 1 connecting to connector set $\psi_1 = \{\theta_2, \theta_3\}$. If both θ_2 and θ_3 were constrained to own one shadow link each, the resulting (constrained) optimal network would have only one connection, the link between θ_2 and θ_4, and a net

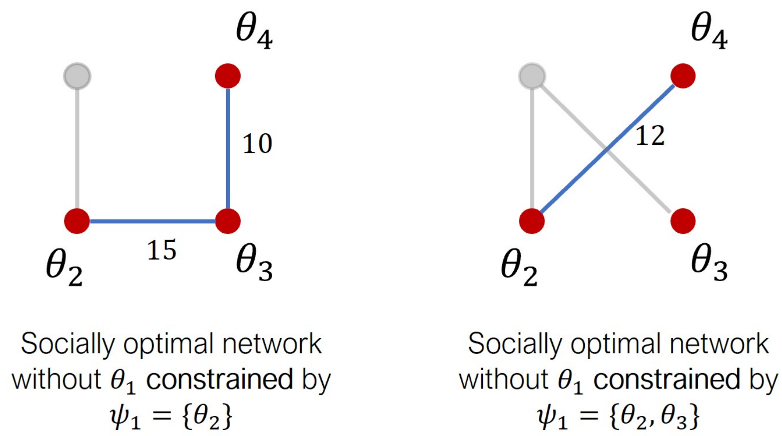

Figure 5

social value $2\theta_2\theta_4 - 2MC_1 = 12 - 4 = 8$ (see diagram on the right in Figure 5). Once again, comparing this to the net output of the unconstrained network without θ_1, we get rearrangement cost $RC_1(\{\theta_2, \theta_3\}) = 14 - 8 = 6$.

Note that in the network on the right in Figure 5, once we activate the shadow links by connecting θ_1 to $\{\theta_2, \theta_3\}$, both θ_1 and θ_2 have two links each, and θ_3 has one, hence the relevant direct costs are $C(2)$ for θ_1, MC_2 for θ_2 and MC_1 for θ_3. Therefore the opportunity cost is

$$OC_1^S(\{\theta_2, \theta_3\}) = C(2) + MC_2 + MC_1 + RC_1(\{\theta_2\}) = 9 + 7 + 2 + 6 = 24.$$

4 The socially optimal network

We have identified the opportunity cost of any connector set of an agent. Using this, we can now re-define the socially optimal network in terms of optimal connector sets.

The result below follows in a straightforward manner from our construction of opportunity cost, which takes into account the impact of a connector set on the entire network. The net social value of a connector set of agent i is the extra output minus its opportunity cost, as shown in the last section. Therefore, in the socially optimal network, each i is linked to the connector set that maximizes the net social value across all connector sets of i. The result is crucial in that it allows us to restate the socially optimal network simply by comparing the net social values of connector sets of each agent. We then use this formulation to construct the mechanism for implementing the socially optimal network.

Lemma 1. *For any $\theta \in \Theta$, suppose agent $i \in N$ is connected to the connector set ψ_i^* in network $G^*(\theta)$. $G^*(\theta)$ is socially optimal if and only if, for all $i \in N$*

$$Y(\{i, \psi_i^*\}) - OC_i^S(\psi_i^*) \geqslant Y(\{i, \psi_i\}) - OC_i^S(\psi_i) \tag{4.1}$$

for all $\psi_i \in \Psi(N \backslash \{i\})$.

Proof: Suppose G^* is socially optimal but for some $i \in N$ there is $\psi_i' \in \Psi(N \backslash \{i\})$ such that $Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*) < Y(\{i, \psi_i'\}) - \mathrm{OC}_i^S(\psi_i')$. Since $\mathrm{OC}_i^S(\psi_i')$ captures the full social cost of ψ_i' by construction, the right hand side is the social value of ψ_i'. Since the social value of ψ_i' is higher, a change in the neighborhood of i to $\{i, \psi_i'\}$ would improve net social value. But then G^* is not socially optimal, which gives us a contradiction.

Next, suppose condition (4.1) holds for all $i \in N$ and all $\psi_i \in \Psi(N \backslash \{i\})$. Suppose that G^* is not socially optimal. Then there is some optimal network G' with i connected to $\psi_i' \neq \psi_i^*$ for some i. If for all such i we have $Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*) = Y(\{i, \psi_i'\}) - \mathrm{OC}_i^S(\psi_i')$, that implies if G' is optimal, then G^* must also be optimal, a contradiction. If, on the other hand, for any i with $\psi_i' \neq \psi_i^*$, we have $Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*) > Y(\{i, \psi_i'\}) - \mathrm{OC}_i^S(\psi_i')$, then clearly we can improve social value by changing the connector set of i from ψ_i' to ψ_i^*, implying that G' is not socially optimal, a contradiction. □

Note that if for one agent the net social value is negative for all nonempty connector sets, the optimal connector set would be the empty set, guaranteeing zero net social value.

4.1 Example 2 (continued)

The example demonstrates the construction of a socially optimal network using net social values of connector sets.

The net social value of the neighborhood formed by θ_1 connecting to $\psi_1 = \{\theta_2, \theta_3\}$ is $2(\theta_1\theta_2 + \theta_1\theta_3) - \mathrm{OC}_1^S(\{\theta_2, \theta_3\})$. The output is 44 and the opportunity cost is 24 (as calculated in Section 3.2.2), so the net value is 20. Using similar calculations, Table 1 below reports the net social value of all neighborhoods formed by connecting each agent to all possible connector sets for them, excluding any connector set involving more than two

Net Social Value \ ψ_i	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_4\}$	$\{\theta_1,\theta_2\}$	$\{\theta_1,\theta_3\}$	$\{\theta_1,\theta_4\}$	$\{\theta_2,\theta_3\}$	$\{\theta_2,\theta_4\}$	$\{\theta_3,\theta_4\}$	$\{\varnothing\}$
$Y(\{1,\psi_1\}) - OC_1^S(\psi_1)$	-	13	11	9	-	-	-	**20**	19	15	0
$Y(\{2,\psi_2\}) - OC_2^S(\psi_2)$	9	-	6	3	-	10	**11**	-	-	4	0
$Y(\{3,\psi_3\}) - OC_3^S(\psi_3)$	**7**	6	-	1	5	-	5	-	2	-	0
$Y(\{4,\psi_4\}) - OC_4^S(\psi_4)$	1	**2**	0	-	-4	-3	-	0	-	-	0

Table 1: This shows the net social value of neighborhoods for Example 2. Highlighted cells indicate the connector sets with highest net social value for each type. The corresponding neighborhoods form the socially optimal network.

players.[7] Using the maximum in each row (highlighted cells), we can construct the socially optimal network (as in Figure 2).

4.2 Properties of the socially optimal network

Let us first comment on the existence and uniqueness of a socially optimal network for any given profile of types. As noted in Section 2, the socially optimal network exists. Further, uniqueness holds under weak additional restrictions on g as shown in the result below.[8] The proof is relegated to the appendix.

Proposition 1. *Suppose for any $\theta \in \Theta$, the cross-partial derivative of g is non-zero. Then the socially optimal network is generically unique.*

In what follows, we assume that the cross-partial derivative of g is non-zero for all $\theta \in \Theta$.

The next condition is the same as the responsive social choice function definition in Bergemann and Morris (2009).

Definition 4. ***(Responsive Socially Optimal Network)*** *The socially optimal network is responsive if for all $i \in N$, and all $\theta_i' \neq \theta_i$, there exists θ_{-i} such that*

$$\psi_i^*(\theta_i, \theta_{-i}) \neq \psi_i^*(\theta_i', \theta_{-i}).$$

[7] Recall that high marginal cost for a third connection rules out private or social optimality of three links for any agent.

[8] As stated earlier, g is assumed to be positive, strictly increasing and symmetric. The additional restriction essentially rules out linear functions such as $g(\theta_i, \theta_j) = \theta_i + \theta_j$. It is easy to see that in this case, different configurations could result in the same net value.

Responsiveness requires that a change in the report of i changes the socially optimal network allocation for i for some reports of other agents. This would, for example, fail if the marginal costs are increasing but so low that the complete network is optimal for all types of any agent in some sub-interval of the set of types. We assume that the socially optimal network is responsive.

5 The mechanism

We now specify a direct mechanism that asks agents to report their types and specifies an allocation that consists of a connector set for each agent and also specifies transfers for any possible connector set.

As noted in the introduction, once an allocated network is formed, new information can arise and allow for deviation opportunities. To make our implementation result robust to such information discoveries *after formation of the allocated network*, we define an adjustment process under full information in Section 5.3, and define the concept of network equilibrium.

The mechanism imposes two weak restrictions on network adjustments. First, the fact that agents can adjust network connections allows some scope for misreporting and then correcting the consequence later. To prevent these trivial deviations, the transfer design adds (arbitrarily small) fines in the transfers for any final choice of connector set that deviates from the allocation.

The second restriction has to do with the adjustment process itself and is clarified in section 5.3 after we specify the allocation and transfers.

5.1 The sequence of moves

Notation The mechanism is denoted by $\mathcal{M}$. We use variables with a "hat" to denote reports and report-based calculations. The sequence of moves is as follows.

1. The agents are asked to report types. Let $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)$ denote the vector of reports.

2. Given the report vector, the designer announces a connector set allocation for each announced type and transfers for all possible connector sets associated with each agent.

3. Agents form allocated connections. After allocated connections are formed, adjustments are allowed take place. We specify the adjustment process and the network equilibrium notion.

4. Given the neighborhoods formed by each agent in the network equilibrium, they receive transfers.

5.2 Connector set allocation and transfers

Given the report vector $\widehat{\theta}$, the designer calculates

- the opportunity cost $\widehat{\mathrm{OC}}_i^S(\psi_i)$ for each $i \in N$, and each $\psi_i \in \Psi(N \backslash \{i\})$ and
- the output each ψ_i would generate when connected to i, given by $\widehat{Y}(\{i, \psi_i\}) = \sum_{j \in \psi_i} g(\widehat{\theta}_i, \widehat{\theta}_j)$.

Allocation of i The allocation of i is a connector set $\widehat{\psi}_i^*$ such that

$$\widehat{Y}(\{i, \psi_i^*\}) - \widehat{\mathrm{OC}}_i^S(\psi_i^*) \geqslant \widehat{Y}(\{i, \psi_i\}) - \widehat{\mathrm{OC}}_i^S(\psi_i)$$

for any $\psi_i \in \Psi(N \backslash \{i\})$. From Lemma 1, this is the socially optimal connector set of i as calculated from the report vector $\widehat{\theta}$.

Transfer to i The designer announces a function $T_i : \Theta \times \Psi(N \backslash \{i\}) \mapsto \Re$ for each agent $i \in N$. That is, for any reported type profile, and for any connector set ψ_i of i, the function specifies a real number.

The transfer to i for any connector set ψ_i is given by

$$T_i(\psi_i | \widehat{\theta}) - \mathbb{1}(\widehat{\psi}_i^*)\varepsilon$$

where $T_i(\psi_i | \widehat{\theta})$ is specified below and $\varepsilon > 0$ is a (possibly arbitrarily small) penalty for deviations from the optimal connector set allocation: if $\psi_i = \widehat{\psi}_i^*$ then the indicator variable $\mathbb{1}(\widehat{\psi}_i^*) = 0$, and $\mathbb{1}(\widehat{\psi}_i^*) = 1$ otherwise.

Given report $\widehat{\theta}$, and given transfers, the payoff of agent i connecting to the connector set

ψ_i is given by[9]

$$v_i(\psi_i|\theta,\widehat{\theta}) = \frac{Y(\{i,\psi_i\})}{2} - C(|\psi_i|) + T_i(\psi_i|\widehat{\theta}) - \mathbb{1}(\widehat{\psi}_i^*)\varepsilon. \tag{5.1}$$

Note that the payoff of i from connecting to any connector set ψ_i involves the true output $Y(\{i,\psi_i\})$ of the neighborhood $\{i,\psi_i\}$. The report vector affects only the transfers and allocation specified by the mechanism.

Specifying the function T_i For any connector set ψ_i, the function T_i has two components. The first component compensates i for the direct cost of connections in the set; the second component is a "tax": the agent is asked to pay an amount that corresponds to the social opportunity cost (as calculated from reports) of the connector set. T_i is then given by

$$T_i(\psi_i|\widehat{\theta}) = C(|\psi_i|) - \frac{\widehat{\mathrm{OC}}_i^S(\psi_i)}{2}.$$

Note that the opportunity cost of the connector set ψ_i is calculated from reports of agents other than i, and the report of i plays no role in the calculation.

Finally, given report vector $\widehat{\theta}$, the transfer-augmented payoff of agent i under mechanism $\mathcal{M}$ from linking to the connector set ψ_i is given by equation (5.1), which can be rewritten as

$$v_i(\psi_i|\theta,\widehat{\theta}) = \frac{1}{2}\Big(Y(\{i,\psi_i\}) - \widehat{\mathrm{OC}}_i^S(\psi_i)\Big) - \mathbb{1}(\widehat{\psi}_i^*)\varepsilon. \tag{5.2}$$

5.3 Network adjustments and network equilibrium

Finally we specify the network adjustment process starting from the allocated network and the network equilibrium concept. As noted in the introduction, it seems plausible that once the allocated network is formed, agents learn the types of their partners. Network relations may also facilitate learning about types of agents beyond partners. Such new information *after the allocated network is formed* might lead to link readjustments. To ensure robustness to any such information discovery, we specify an adjustment process under full information at the stage after the formation of the allocated network.

[9] The payoff depends on the neighborhood $\{i,\psi_i\}$. Since this is clear from the context, we economize by abusing notation slightly and put ψ_i as the argument.

The mechanism imposes two weak restrictions on network adjustments. First, as noted in Section 5.2 above, the transfer design adds (arbitrarily small) fines for any final choice of connector set that deviates from the allocation.

Second, to characterize fully the network equilibrium after the specification of allocation and transfers for each possible report profile, we need an additional restriction. After misreports, the allocation might not be pairwise stable under full information, and agents would adjust links in a pairwise manner (sever unprofitable links, join mutually beneficial links). This process might lead to a cycle which does not converge to a pairwise stable network. To be able to define the network equilibrium under all possible reports, we impose the restriction that K rounds of adjustments are allowed, where K is a finite positive integer (that can be arbitrarily large).[10]

Note that at the stage of network adjustments, the mechanism payoff from any connector set for any agent is fixed since transfers for each connector set and allocations are already specified. In other words, any change in ψ_i for i changes the payoff according to a fixed menu, and has no implication for agents whose connector sets are not changing in the process. Therefore there are no endogenous externalities at this stage.

5.3.1 Network adjustment process

The following definition of *a pairwise improving path* is adapted from Jackson and Watts (2001). Starting from any network, a link that leads to gains by both agents (at least one agent gains strictly) is added, and any agent unilaterally severs a link if this strictly improves own payoff.

Let ψ_i^k denote the connector set of agent i in network G^k.

Definition 5. *A pairwise improving path from a network G to network G' is a finite sequence of networks $G^0, \ldots, G^T$ where $G^0 = G$ and $G^T = G'$ such that for any $k \in \{0, \ldots, T-1\}$ either*

i. $G^{k+1} = G^k - ij$ *for some* ij *such that* $v_i(\psi_i^k \backslash \{j\} | \theta, \widehat{\theta}) > v_i(\psi_i^k | \theta, \widehat{\theta})$, *or*

ii. $G^{k+1} = G^k + ij$ *for some* ij *such that* $v_i(\psi_i^k \cup \{j\} | \theta, \widehat{\theta}) > v_i(\psi_i^k | \theta, \widehat{\theta})$ *and* $v_j(\psi_j^k \cup \{i\} | \theta, \widehat{\theta}) \geqslant v_j(\psi_j^k | \theta, \widehat{\theta})$.

[10] This restriction is not required under an alternative stronger condition on link adjustment. We show this in the appendix.

The next result follows immediately.

Lemma 2. *A network is PWS-FI if and only if no pairwise improving path can emanate from it.*

Proof. Suppose no pairwise improving path can start from a network G. This implies that for any pair $i, j \in N$ such that $ij \in G$, the network $G - ij$ does not satisfies condition (i) in Definition 5 above and for any pair $i, j \in N$ such that $ij \notin G$, the network $G + ij$ does not satisfies condition (ii) above. Since this is true for any pair i, j, it also coincides with the Definition 1 of PWS-FI, thus G must be PWS-FI. Suppose now that G was PWS-FI. Then by definition no pair $i, j \in N$ such that $ij \in G$ can gain from severing their connection and no pair $i, j \in N$ such that $ij \notin G$ can both gain from connecting (with strict gain for at least one of the agents). This means that there is no link ij such that $G + ij$ or $G - ij$ can satisfy, respectively, conditions (i) or (ii) of Definition 5. Thus, no pairwise improving path can emanate from a PSW-FI network. □

5.3.2 Network equilibrium

As noted the start of section 5.3, we impose the following restriction. The mechanism allows K rounds of pairwise adjustments where K is any finite (possibly arbitrarily large) number.

We now define the network equilibrium given any profile of types θ and after any profile of reports $\widehat{\theta}$. Let G^0 denote the allocated network formed by the connector set allocations ψ_i^0 for $i \in \{1, \ldots, N\}$ in period 0 (allocation stage). Let G^k denote the network where $G^0, \ldots, G^k$ is a pairwise improving path.

Definition 6. ***(Network Equilibrium)*** *Under the mechanism $\mathcal{M}$ and given any $K > 0$, a network G is an equilibrium network under any $(\theta, \widehat{\theta})$ if either*

- $G = G^k$ *for* $0 \leqslant k < K$ *and* G *is PWS-FI, or*
- $G = G^K$.

Remark 1. ***(Network restrictions under the mechanism)*** *As the above specifications make clear, two relatively weak restrictions are imposed by the mechanism on network adjustments starting from the allocated network: a fine (can be arbitrarily small) for deviating from the allocation, which prevents trivial deviations (say report high type then sever links) and a finite number* *of periods (can be arbitrarily large) allowed for adjustment in a pairwise fashion, which allows us to characterize the network equilibrium following any vector of reports (and consequent transfers/allocation) by ruling out cycles.*

Remark 2. ***(Alternative adjustment process)*** *Suppose the designer is able to impose a stronger restriction on the adjustment process: starting from the allocation, any link severance must also be mutually agreed upon to be allowed by the designer. In this case, a network equilibrium exists under any report profile without requiring the K-rounds restriction. We show this in Appendix A.2.*

From the definition of network equilibrium we know that under any vector of types θ and reports $\widehat{\theta}$, the equilibrium network exists and is given by G^k for some $k \in \{0, 1, \ldots, K\}$. This proves the following result.

Lemma 3. *Let $\Gamma_S(\theta, \widehat{\theta}|\mathcal{M})$ be the set of equilibrium networks that can arise under the mechanism $\mathcal{M}$ given type profile θ and report profile $\widehat{\theta}$. This set is non-empty for any θ and $\widehat{\theta}$.*

6 Ex post network implementation

6.1 Definitions

Next, we need a notion of equilibrium at the reporting stage. Note the difference between this and the network equilibrium notion. After reporting, agents form the allocated network, and having formed a network, they can uncover information and act on it. So any network adjustments are under full information. However, at the reporting stage agents only know own type so any *reporting stage equilibrium* is under incomplete information.

6.1.1 Ex post incentive compatibility at the reporting stage

As clarified in the previous section, after any profile of reports and formation of consequent allocated network, agents arrive at some equilibrium network. Given such network equilibria after any profile of reports, we consider incentives at the reporting stage.

Following the definition of Bergemann and Morris (2005), we say that $\mathcal{M}$ is (strict) ex post incentive compatible if for any θ, and given any choice of equilibrium network after each report profile (and consequent transfers/allocation), truthful reporting is (strictly) optimal if others report truthfully.

Let $\Gamma_S(\theta, \widehat{\theta}_i, \widehat{\theta}_{-i}|\mathcal{M})$ be the set of equilibrium networks that can arise under the mechanism $\mathcal{M}$ given type profile θ and report profile $\widehat{\theta}$. From Lemma 3, the set is non-empty.

Let G be the network formed by the connector set allocations under mechanism $\mathcal{M}$. We say that **G is allocated under** $\mathcal{M}$**.**

Recall that the payoff of agent i from the connector set ψ_i, given θ and the report vector $\widehat{\theta}$, under mechanism $\mathcal{M}$ is $v_i(\psi_i|\theta, \widehat{\theta})$ given by equation (5.2). In what follows, it helps to also explicitly refer to the payoff arising under a specific network that forms after any profile of types and reports, so we include the relevant network as an argument in the payoff function.

Definition 7. ***(Strict ex post incentive compatibility)*** *Suppose $\widehat{\theta}_i \neq \theta_i$. Consider $G \in \Gamma_S(\theta, \theta|\mathcal{M})$ with associated connector set ψ_i for i and $\widehat{G} \in \Gamma_S(\theta, \widehat{\theta}_i, \theta_{-i}|\mathcal{M})$ with associated connector set $\widehat{\psi}_i$ for i. $\mathcal{M}$ is strict ex post incentive compatible if for **any** choice of $(G, \widehat{G})$ the following holds for all $i \in N$ and all $\theta \in \Theta$*

$$v_i(\psi_i|\theta, \theta, G) > v_i(\widehat{\psi}_i|\theta, (\widehat{\theta}_i, \theta_{-i}), \widehat{G})$$

for any $\widehat{\theta}_i \neq \theta_i$.

6.1.2 Strict ex post network implementation

Definition 8. ***(Strict Ex Post Network Implementation)*** *A network G is strict ex post implemented under mechanism $\mathcal{M}$ if the following conditions hold:*

- *$\mathcal{M}$ is strict ex post incentive compatible.*
- *Given truthful reporting, G is allocated under $\mathcal{M}$, and G is PWS-FI.*

6.2 Result on Implementation

Proposition 2. *For any $\theta \in \Theta$, the socially optimal network $G^*(\theta)$ is strict ex post implemented under $\mathcal{M}$.*

Proof: The proof proceeds through the following lemmas.

Lemma 4. *For any $\theta \in \Theta$, given truthful reporting, the allocated network is $G^*(\theta)$ and it is PWS-FI.*

Proof: Given the report vector, the mechanism specifies that the designer calculates the socially optimal neighborhood for i. Therefore, under truthful reporting $G^*(\theta)$ is allocated. Let us now show that $G^*(\theta)$ is PWS-FI.

From condition (4.1), we know that for any agent i, the connector set ψ_i^* under the socially optimal network satisfies

$$Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*) > Y(\{i, \psi_i\}) - \mathrm{OC}_i^S(\psi_i)$$

for any $\psi_i \in \Psi(N \backslash \{i\})$ such that $\psi_i \neq \psi_i^*$. The strict inequality follows from the assumption that g has non-zero cross partials and Proposition 1.

Since all agents report truthfully, the calculated opportunity cost of any connector set is the same as the true opportunity cost, and $G^*(\theta)$ is allocated. From equation (5.2), the payoff of any agent i by setting $\psi_i = \psi_i^*$ is

$$\frac{1}{2}(Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*)).$$

Since $\psi_i = \psi_i^*$, and the allocation is also ψ_i^*, $\mathbb{1}(\psi_i^*) = 0$ implying that the last term in equation (5.2) is zero. Therefore individual and social payoffs coincide, and given truthful reporting, ψ_i^* maximizes the payoff of i. Since the socially optimal network is unique, the maximization is strict.

It follows that given truthful reporting, i has no incentive to deviate from the connector set allocation by forming a neighborhood other than $\{i, \psi_i^*\}$. In particular this implies that there is no incentive to deviate by either forming a link with an agent not in the allocation or by severing a link with an agent in the allocation. Therefore the allocated network $G^*(\theta)$ is PWS-FI for any $\theta \in \Theta$. $\square$

The next result shows that the payoff of any agent i from any *given* connector set is unaffected by the report of i.

Lemma 5. *Consider any report vector $\widehat{\theta}$. For any $i \in N$, given any connector set $\psi_i \in \Psi(N \backslash \{i\})$, the payoff $v_i(\psi_i|\theta, \widehat{\theta})$ of i from connecting to ψ_i does not depend on $\widehat{\theta}_i$.*

Proof: From equation (5.2),

$$v_i(\psi_i|\theta, \widehat{\theta}) = \frac{1}{2}(Y(\{i, \psi_i\}) - \widehat{\mathrm{OC}}_i^S(\psi_i)) - \mathbb{1}(\widehat{\psi}_i^*)\varepsilon.$$

Note that $Y(\{i, \psi_i\})/2 = \sum_{j \in \psi_i} g(\theta_i, \theta_j)$. In other words, the output depends on true types of the agent and partners, irrespective of what any agent may report. Next, consider the tax term, $\widehat{\mathrm{OC}}_i^S(\psi_i)$. The rearrangement cost of ψ_i is calculated excluding i, and does not depend on either the report or the true type of i. The opportunity cost simply adds link costs to rearrangement cost, therefore the same applies to the opportunity cost. Finally, for any given ψ_i, $\mathbb{1}(\widehat{\psi}_i^*)$ is either 1 or 0, and this does not depend on the report of i. □

The next result established that $\mathcal{M}$ is strict ex post incentive compatible.

Lemma 6. *Under the specified mechanism $\mathcal{M}$, for any type profile θ, if other agents report truthfully, truthful reporting is the unique best response of an agent.*

Proof: From Lemma 4, under report profile (θ_i, θ_{-i}) the connector set allocation for i is the true socially optimal ψ_i^*. In this case the payoff of i given by $\frac{1}{2}(Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*))$. Further, since G^* is the unique socially optimal network,

$$Y(\{i, \psi_i^*\}) - \mathrm{OC}_i^S(\psi_i^*) > Y(\{i, \psi_i\}) - \mathrm{OC}_i^S(\psi_i) \tag{6.1}$$

for any $\psi_i \in \Psi(N \backslash \{i\})$ such that $\psi_i \neq \psi_i^*$.

Now suppose agents other than i report θ_{-i}, but i reports $\widehat{\theta}_i \neq \theta_i$. Let $\widehat{\psi}_i^*$ be the connector set allocation under report profile $(\widehat{\theta}_i, \theta_{-i})$. From the assumption that the socially optimal network is responsive (section 4.2), $\widehat{\psi}_i^*$ can be different from ψ_i^* allocated under report profile (θ_i, θ_{-i}).

From Lemma 5 above, we know that the report of i does not affect the payoff of i from any given ψ_i. Since opportunity cost of any ψ_i depends on the reports of agents other than i, and since other agents report truthfully, we have $\widehat{\mathrm{OC}}_i^S(\psi_i) = \mathrm{OC}_i^S(\psi_i)$. Therefore, from any report profile $(\widehat{\theta}_i, \theta_{-i})$, given allocation $\widehat{\psi}_i^*$, the payoff of i after any adjustments is bounded above by

$$\max_{\psi_i \in \Psi(N \setminus \{i\})} \frac{1}{2}\Big(Y(\{i, \psi_i\}) - \mathrm{OC}_i^S(\psi_i)\Big) - \mathbb{1}(\widehat{\psi}_i^*)\varepsilon$$

If this is maximized at $\psi_i = \widehat{\psi}_i^*$, and the neighborhood $\{i, \widehat{\psi}_i^*\}$ is part of a Network Equilibrium (Definition 6), $\mathbb{1}(\widehat{\psi}_i^*) = 0$, implying the last term is zero. But from equation (6.1), the payoff is strictly lower than that from ψ_i^*. If this is maximized after adjustments at a $\psi_i' \neq \widehat{\psi}_i^*$, and $\{i, \psi_i'\}$ is part of a Network Equilibrium, $\mathbb{1}(\widehat{\psi}_i^*) = 1$. Therefore the upper bound of payoff i after adjustments is lower than the payoff from ψ_i^* by at least ε.

Therefore, when others report truthfully, if agent i misreports and receives an allocation $\widehat{\psi}_i^* \neq \psi_i^*$, this leads to a strictly lower payoff for i compared to truthtelling. Since $\widehat{\psi}_i^*$ can be different from ψ_i^* as noted above, misreporting is strictly suboptimal for i if others report truthfully. □

Proof of Proposition 2 (completed): Lemma 4 and Lemma 6 together show that under the mechanism, G^* satisfies the conditions for strict ex post implementation (Definition 8). This completes the proof. □

6.3 Budget balance

We show that the transfers required for strict ex post implementation of the socially optimal network generate a budget surplus. The budget deficit for the socially optimal connector set ψ_i^* is given by

$$D(\psi_i^*) = C(|\psi_i^*|) - \frac{1}{2}\mathrm{OC}_i^S(\psi_i^*). \tag{6.2}$$

Note that since the socially optimal network is being strict ex post implemented, $G^*(\theta)$ is allocated and also PWS-FI. This implies that any fines are out-of-equilibrium and therefore do not apply to budget calculations.

Proposition 3. *Strict ex post implementation of the socially optimal network generates a budget surplus:*

$$\sum_{i \in N} D(\psi_i^*) \leqslant 0$$

where strict inequality holds if marginal costs of link formation are increasing, and equality holds if marginal costs are constant.

Proof: Recall, from section 3.2, that $G_{\psi_i} \equiv (G_{-i}|\psi_i)^* \cup \{i, \psi_i\}$. With $\psi_i = \psi_i^*$, G_{ψ_i} is simply the optimal network G^*. The degree of i is then $d_i(G^*)$, given by $|\psi_i^*|$. Since throughout this proof we only refer to the socially optimal network, we simply refer to the degree of an agent i as d_i.

From equation (3.1),

$$\mathrm{OC}_i^S(\psi_i^*) = C(d_i) + \sum_{j \in \psi_i^*} MC_{d_j} + \mathrm{RC}_i(\psi_i^*).$$

It follows that

$$\sum_{i \in N} D(\psi_i^*) = \frac{1}{2} \sum_{i \in N} \left(C(d_i) - \sum_{j \in \psi_i^*} MC_{d_j} - \mathrm{RC}_i(\psi_i^*) \right).$$

Since $\mathrm{RC}_i(\psi_i^*) \geqslant 0$, we have

$$-\sum_{i \in N} \mathrm{RC}_i(\psi_i^*). \leqslant 0$$

Note that if marginal costs are constant, rearrangement costs are zero since forming a link does not have any impact on the costs of forming other links.

Next, consider the term $\sum_{i \in N} \left(C(d_i) - \sum_{j \in \psi_i^*} MC_{d_j} \right)$. Note that MC_{d_j} appears once for each i such that $j \in \psi_i^*$. Therefore, in the overall sum, MC_{d_j} appears d_j times. This implies

$$\sum_{i \in N} \left(\sum_{j \in \psi_i^*} MC_{d_j} \right) = \sum_{j \in N} d_j MC_{d_j} = \sum_{i \in N} d_i MC_{d_i}.$$

Let $AC_i \equiv C(d_i)/d_i$. Then

$$\sum_{i \in N} \left(C(d_i) - \sum_{j \in \psi_i^*} MC_{d_j} \right) = \sum_{i \in N} \left(d_i AC_i - d_i MC_{d_i} \right) = \sum_{i \in N} d_i \left(AC_i - MC_{d_i} \right) \leqslant 0.$$

where the last inequality follows from the fact that AC_i is the average link cost for i and MC_{d_i} is the highest marginal link cost incurred by i. Given that marginal costs of link formation are increasing, $AC_i - MC_{d_i} \leqslant 0$, with strict inequality if costs are not constant. This completes the proof. □

6.4 Constant marginal cost of link formation

As noted at the outset, a constant marginal cost is a simple special case of increasing marginal costs. Recall from equation (3.1) that the opportunity cost of i connecting to a connector set ψ_i is given by

$$\mathrm{OC}_i^S(\psi_i) = C(|\psi_i|) + \sum_{j \in \psi_i} MC_{d_j(G_{\psi_i})} + \mathrm{RC}_i(\psi_i).$$

Suppose that marginal cost of link formation is constant, given by $c > 0$. In this case adding a link does not change the marginal cost calculation for any other link, that is, no externalities arise from the formation of a link. This implies there is no rearrangement cost (the third term above is zero). Further, each of the first two terms simplify to $c|\psi_i|$, implying that $\mathrm{OC}_i^S(\psi_i) = 2c|\psi_i|$. Since the opportunity cost of a connector set scales up by its cardinality, we can simply consider the incentive to form individual links, where each link has the opportunity cost $2c$. In this case whether a link should be part of an optimal network is independent of other links.

The fact that under a constant marginal cost the units of analysis can be simplified from the elements of the set of connector sets to individual links allows further characterization of the socially optimal network. While we omit the formal proof, it is easy to see that if a link between θ_i and θ_j is part of a socially optimal network, then so is a link between θ_i and any other type higher than θ_j. This implies that any non-empty socially optimal network is a connected nested split graph where the neighborhoods of lower types are nested in that of higher types.

7 Conclusion

We implement a social choice function that takes the form of a productive network formed by agents with privately known productivity types and convex cost of link formation. The latter gives rise to endogenous externalities across the entire network, and the impact could be negative on some agents and positive for yet others. Such externalities imply that networks stable under private incentives may not maximize social value, creating scope for a mechanism to implement the socially optimal network. The first question towards design of incentives is to calculate the externality or "rearrangement cost." With convex cost, the relevant unit for such calculation is any connector set of an agent, which is a subset of other agents. The rearrangement cost of a connector set ψ_i of i then shows how the value of optimal arrangement of agents other than i changes when i is connected to ψ_i. Adding direct link costs to the rearrangement cost, we obtain the social opportunity cost for each connector set of an agent. Crucially, this allows deriving optimal network connections from net values of neighborhoods, where each neighborhood is an agent connected to a particular connector set. This formulation proves critical in setting up incentives.

The mechanism uses the opportunity cost calculated from reports of agents to construct transfers, and gives a network allocation to each agent that is simply the connector set of each agent associated with the calculated optimal network.

In the context of network implementation, there is an additional problem that does not arise in standard mechanisms. Once the allocated network is formed, it is reasonable to assume that agents learn the types of their partners. This might lead them to sever links allocated based on reports, but subsequently turn out to be suboptimal under further information generated in the network. The network might enable agents to also learn types of others beyond their partners, giving rise to further opportunities for deviation from the allocated network. To ensure that our implementation is robust to such network-generated information, we consider the equilibrium notion of pairwise stability under full information (PWS-FI), and require that any implemented network arising through the allocation or from further pairwise adjustments satisfies PWS-FI.

Given PWS-FI network outcome after specification of allocation and transfers following any reports, we then show that at the reporting stage, the mechanism is strict ex post incentive compatible: for any profile of types, truthful reporting is a strict mutual best response. Accordingly we define the notion of strict ex post implementation of a network as ex post incentive compatibility at the reporting stage and formation of PWS-FI network once the transfers and allocations are specified given reports. Ex post implementation is similar to Bergemann and Morris (2005, 2009) and implements in a detail-free manner. However, our implementation through a direct mechanism is partial rather than full implementation (Bergemann and Morris, 2009, Ollár and Penta, 2017). Whether reasonable network restrictions allow ex post full implementation of the socially optimal network in convex cost environments is a question we hope to address in future research.

A Appendix

A.1 Proof of Proposition 1

We prove by induction. Note that $n \geqslant 3$. First, consider $n = 3$. Without loss of generality rename the agents so that the types are $\theta_1 > \theta_2 > \theta_3$.

The socially optimal network can have one, two or three links. If there is only one link, it must be between θ_1 and θ_2, since any other combination would have the same total cost but lower output. The two-link optimal network must have the links between θ_1, θ_2 and θ_1, θ_3 (any other combination produces a lower output without changing total cost). The only three-link network is the complete network.

Suppose the optimal network is not unique. Then at least any two structures of the above must have the same net social value. But the total value of the optimal one-link structure does not depend on θ_3 while it does for the other two optimal structures, so the optimal one-link structure can generate same total value as any of the other optimal structures only with probability 0 (i.e., for a specific value of θ_3).

So suppose the total value generated by the optimal two-link and three-link networks is the same. Then it must be that the link between θ_2, θ_3 generates 0 net social value, $2g(\theta_2, \theta_3) - 2MC_2 = 0$ (no rearrangement cost here as we are just adding links at each stage). But this is non-generic. Therefore with $n = 3$ we have a generically unique social optimal network.

Now suppose we have a generically unique socially optimal network for $n = k$. Consider $n = k + 1$. Let i denote the $k + 1$-th agent being added.

Start from the optimal network under k players. Connect i to some neighborhood $\{i, \psi_i\}$. Note that $\sum_{j \in \psi_i} g(\theta_i, \theta_j)$ is the marginal gross social value added by the presence of agent i linked to ψ_i.

Now suppose there are two socially optimal networks with associated neighborhoods $\{i, \psi_i\}$ and $\{i, \widetilde{\psi}_i\}$ for all i. Note that for the two optimal networks to be different, there must be some agents for which these neighborhoods are not identical. So, without loss of generality, suppose $\psi_i \neq \widetilde{\psi}_i$, that is, there is at least one agent s such that $s \in \psi_i$ but $s \notin \widetilde{\psi}_i$.

Since we have a unique socially optimal network for $n = k$, but two socially optimal networks for $n = k+1$, it must be that the marginal net social value added by the inclusion of the $k+1$-th agent is same across the neighborhoods,

$$\sum_{j \in \psi_i} g(\theta_i, \theta_j) - OC_i^S(\psi_i) = \sum_{j \in \widetilde{\psi}_i} g(\theta_i, \theta_j) - OC_i^S(\widetilde{\psi}_i)$$

which implies

$$\sum_{j \in \psi_i} g(\theta_i, \theta_j) - \sum_{j \in \widetilde{\psi}_i} g(\theta_i, \theta_j) = OC_i^S(\psi_i) - OC_i^S(\widetilde{\psi}_i) \tag{A.1}$$

The right hand side is independent of θ_i. Therefore, for the above to hold generically, we need the left hand side to be invariant with respect to θ_i, that is

$$\sum_{j \in \psi_i} \frac{\partial g(\theta_i, \theta_j)}{\partial \theta_i} = \sum_{j \in \widetilde{\psi}_i} \frac{\partial g(\theta_i, \theta_j)}{\partial \theta_i}$$

As noted above, since $\psi_i \neq \widetilde{\psi}_i$, there is $s \in \psi_i$ such that $s \notin \widetilde{\psi}_i$. Since the cross partial of g is not zero, any change in θ_s would therefore change the left hand side, breaking the above equality. In other words, the above equality is non-generic, implying that condition A.1 does not hold generically.

Therefore if generic uniqueness of optimal network is true for $n = k$ then it is true for $n = k+1$. Since this is true for $n = 3$, it is true for all $n > 3$ as well. □

A.2 Alternative adjustment process and network equilibrium

In Remark 2 in Section 5.3, we noted that an alternative adjustment process can be specified which places a greater constraint on link severance but does not require a limit on adjustment rounds. We specify the details below for this case.

Starting from the allocated network, consider an alternative adjustment process. Both formation and severance now requires mutual consent. Thus, implicit in the process is an enforcement of allocated links by the designer that is stronger than the other case analyzed in the paper.

We define a strong PWS-FI as follows.

Definition 9. *A network G is Strong Pairwise Stable under Full Information (Strong PWS-FI) if the following conditions hold*

(i) *If types θ_i and θ_j are linked, $v_i(\psi_i|\theta,\widehat{\theta}) > v_i(\psi_i \backslash \{j\}|\theta,\widehat{\theta})$ and $v_j(\psi_j|\theta,\widehat{\theta}) \geqslant v_j(\psi_j \backslash \{i\}|\theta,\widehat{\theta})$, and*

(ii) *If types θ_i and θ_j are not linked, then $v_i(\psi_i \backslash \{j\}|\theta,\widehat{\theta}) > v_i(\psi_i|\theta,\widehat{\theta})$ and $v_j(\psi_j \backslash \{i\}|\theta,\widehat{\theta}) \geqslant v_j(\psi_j|\theta,\widehat{\theta})$.*

A strong pairwise improving path is defined as follows.

Definition 10. *A strong pairwise improving path from a network G to network G' is a finite sequence of networks $G^1,\ldots,G^T$ where $G^1 = G$ and $G^T = G'$ such that for any $k \in \{1,\ldots,T-1\}$ either*

i. *$G^{k+1} = G^k - ij$ for some ij such that $v_i(\psi_i^k \backslash \{j\}|\theta,\widehat{\theta}) > v_i(\psi_i^k|\theta,\widehat{\theta})$ and $v_j(\psi_j \backslash \{i\}|\theta,\widehat{\theta}) \geqslant v_j(\psi_j|\theta,\widehat{\theta})$, or*

ii. *$G^{k+1} = G^k + ij$ for some ij such that $v_i(\psi_i^k \cup \{j\}|\theta,\widehat{\theta}) > v_i(\psi_i^k|\theta,\widehat{\theta})$ and $v_j(\psi_j^k \cup \{i\}|\theta,\widehat{\theta}) \geqslant v_j(\psi_j^k|\theta,\widehat{\theta})$.*

The next result follows. Essentially, the stronger adjustment process precludes cycles, so that starting from any allocated network not already Strong PWS-FI, the adjustment process converges to a Strong PWS-FI in a finite number of periods.

Proposition 4. *Under mechanism $\mathcal{M}$, given any profile of types and reports $(\theta,\widehat{\theta})$, a Strong PWS-FI exists.*

Proof. Consider a network G allocated by $\mathcal{M}$ under $(\theta,\widehat{\theta})$. If there are no strong pairwise improving paths emanating from G, G must be Strong PWS-FI. Suppose there exists at least one improving path from G. Let us show that we reach a Strong PWS-FI after $M \geq 1$ iterations of strong pairwise improving adjustments, where M is a finite integer. The proof is immediate if a cycle does not arise. To see a cycle is not possible, suppose on the contrary that we have an improving cycle. Let $G \equiv G^0$. It must be that there is a sequence of networks such that G^{k-1}, G^k, G^0 is a strong pairwise improving path for $k \geqslant 1$. Since a strong pairwise improvement is a Pareto improvement for the agents involved, with strict improvement for at least one agent, and does not affect payoffs of any other agents, we have $V(G^0) < \ldots < V(G^k) < V(G^0)$ which is a contradiction. This completes the proof. $\square$

Definition 11. ***(Network Equilibrium)*** *Under the mechanism $\mathcal{M}$ a network G is an equilibrium under any $(\theta, \widehat{\theta})$ if G is Strong PWS-FI.*

References

BERGEMANN, D. AND S. MORRIS (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813. 2, 3, 22, 29

——— (2009): "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76, 1175–1204. 2, 3, 16, 29

BLOCH, F. AND M. O. JACKSON (2007): "The formation of networks with transfers among players," *Journal of Economic Theory*, 133, 83–110. 4

JACKSON, M. O. (2010): *Social and economic networks*, Princeton university press. 4

JACKSON, M. O. AND A. WATTS (2001): "The Existence of Pairwise Stable Networks," *Seoul Journal of Economics*, 14, 299–322. 20

JACKSON, M. O. AND A. WOLINSKY (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74. 2, 4, 6

OLLÁR, M. AND A. PENTA (2017): "Full implementation and belief restrictions," *American Economic Review*, 107, 2243–77. 3, 29