A Network Approach to Public Goods

Matthew Elliott† Benjamin Golub‡
(Job Market Paper)

November 15, 2012

Abstract

We study settings where each agent can exert costly effort that creates nonrival, heterogeneous benefits for some of the others. For example, municipalities can forgo consumption to reduce pollution. How do asymmetries in the environment affect the prospects for efficient cooperation? We approach this question by analyzing a network that describes the marginal benefits agents can confer on one another. The first set of results explains how the largest eigenvalue of this network measures the marginal gains available from cooperating; as an application, we describe the players whose participation is essential to achieving any Pareto improvement on an inefficient status quo. Next, we examine mechanisms all of whose equilibria are Pareto efficient and individually rational; an outcome is called robust if it is an equilibrium outcome in every such mechanism. Robust outcomes exist and correspond to the Lindahl public goods solutions. The main result is a characterization of effort levels at these outcomes in terms of players’ centralities in the benefits network. It entails that an outcome is robust if and only if agents contribute in proportion to how much they value the efforts of those who help them.

Keywords: collective action, externalities, public goods, network centrality, implementation theory, Lindahl equilibrium

---

*Some results previously circulated in a working paper titled “A Network Centrality Approach to Coalitional Stability,” which, unlike this paper, studied equilibria of repeated interaction resistant to coalitional deviations. The constant guidance of Nageeb Ali, Abhijit Banerjee, Matthew O. Jackson, Andy Skrzypacz, and Bob Wilson has been essential. We are indebted to Arun Chandrasekhar, Gabriel Carroll, Sylvain Chassang, Jessica Hui, Anil Jain, Juuso Toikka, and Ariel Zucker for detailed readings and extensive comments. For their helpful suggestions we thank Kyle Bagwell, Federico Echenique, Glenn Ellison, Alex Frankel, Andrea Galeotti, Hari Govindan, Sanjeev Goyal, Scott Kominers, David Kreps, Jacob Leshno, Jon Levin, Mihai Manea, Vikram Manjunath, Stephen Morris, Roger Myerson, Muriel Niederle, Michael Ostrovsky, Michael Powell, Phil Reny, John Roberts, Larry Samuelson, Ilya Segal, Joe Shapiro, Moshe Tennenholtz, Leeat Yariv, Muhamet Yildiz, Xiao Yu Wang, Glen Weyl, and Alex Wolitzky.

†Division of the Humanities and Social Sciences, Caltech. Email: melliott@hss.caltech.edu, web: http://www.hss.caltech.edu/~melliott/.

‡Fellowships in Economics, History, and Politics, Harvard; and J-PAL, MIT. Email: bgolub@mit.edu, web: http://www.mit.edu/~bgolub/
1 Introduction

Consider a situation in which agents can provide public goods to one another – each can do something at a private cost that creates nonrival benefits for others. For example, municipalities can sacrifice some consumption to pollute less. A municipality cannot be excluded from enjoying the benefits of others’ sacrifices, but it may not value everyone’s contributions equally: the river pollution of an upstream neighbor matters more than that of a downstream neighbor. How do such asymmetries in the environment affect the potential gains from cooperation? Which efficient solutions are particularly robust in the presence of strategic behavior? And how do the asymmetries relate to inequality of contributions at those outcomes? This paper argues that insight into these questions can be gained by studying simple statistics of a network that describes the benefits agents can confer on one another.

An example helps to fix ideas. Suppose that there are three towns: X, Y and Z, located as shown in Figure 1. In the course of production, each town generates both air and water pollution. The prevailing winds are from west to east, and so X’s air pollution affects Y and, to a lesser degree, Z. Town Y’s air pollution affects only Z, and no one is affected by Z’s air pollution. A river flows westward, so Z’s water pollution affects X but not Y, which is located away from the river. Each town can reduce its pollution by forgoing some of its own production; this entails private costs, but creates both private benefits and external ones, whose incidence is heterogeneous across beneficiaries. Define $a_i \geq 0$ to be the amount of pollution reduction that town $i$ undertakes relative to some status quo, with a higher action corresponding to less pollution. Each agent, $i$, has a payoff depending on all the actions: $u_i(a_X, a_Y, a_Z)$.

In such a world, the equilibrium outcome will typically not be Pareto efficient in the absence of a way to internalize externalities. Suppose that initially the towns are in an inefficient equilibrium and come together seeking to achieve a Pareto improvement through conditional cooperation. That is, each is willing to commit to an institution whereby it reduces pollution and others reciprocate.\footnote{We focus on favor-trading through the provision of public goods and abstract from side transfers of private goods. This keeps the modeling clean and is relevant for many practical negotiations.}

First, there is a question of how much there is to be gained from such an institution, and how the answer can be estimated using data on the marginal costs and benefits.
of effort at the status quo, which may be the only data available. To address this question, we propose and justify a new measure of the marginal return to concerted investment in public goods. The measure is the largest eigenvalue of the benefits network, in which a link from Y to X measures the marginal value of Y’s effort to X. Our results on this have two important implications. First, cycles in the benefits network (e.g., X can help Y, who can help Z, who can help X) are critical to achieving Pareto improvements on the status quo; the network’s largest eigenvalue quantifies the returns available from such cycles. Second, there is a simple matrix algorithm for finding the players who are essential to a negotiation – in the sense that without their participation, there can be no Pareto improvement on the status quo. They are the ones whose removal causes a sufficiently large disruption of cycles in the benefits network, as measured by the decrease in its largest eigenvalue.

The benefits network is also crucial to our study of robust, efficient outcomes. The formal approach we take to defining robustness is based on implementation theory. Consider a hypothetical designer of the institution that the towns will use to address their collective action problem. At the design stage, she does not know the preferences that the players will have. Nevertheless, she wants to leave them with a reliable mechanism for determining levels of contribution – one such that, under whatever preferences they might have, all Nash equilibria in the mechanism will be Pareto efficient and unanimously preferred to the status quo. An outcome is called robust if it is an equilibrium outcome of every reliable mechanism. Such outcomes always exist in our setting and turn out to be precisely the Lindahl solutions – defined by an analogue of Walrasian equilibrium with individualized pricing of public goods.

Our main result characterizes these robust outcomes in terms of the benefits network. Essentially, an outcome is a Lindahl solution if and only if each agent’s effort level $a_i$ satisfies the equation

$$a_i = \sum_{j \neq i} B_{ij} a_j,$$

where $B_{ij}$ measures the marginal value to $i$ of $j$’s effort at the outcome $a$. Thus, players contribute in proportion to how much they value the efforts of those who help them. There are two substantive consequences. First, it is the benefits an agent receives (rather than those he confers) that determine his level of effort at a robust outcome. Second, by repeatedly plugging (1) into itself, we find that the players contributing the most are those who are most “central” in the benefits network, in the sense that they receive strong direct and indirect benefit flows from others. We flesh out these implications below.

---

2 This designer could represent the towns themselves at an ex ante stage where they make the rules for their future interactions.

3 These may be viewed as predictions that can be made assuming only that the institution chosen is reliable. See Section 4, as well as the discussion of related literature below, for details. For the formal result, a continuity requirement is also imposed on the reliable mechanisms.

4 Imagine that each agent must pay a personalized tax on each unit provided of public goods that he values; the tax is specific both to the good and to the agent. He also receives the taxes others pay for the public good he supplies. A Lindahl outcome is a profile of public goods provision that is optimal for each agent at the given tax schedule. A formal definition appears in Section 4.2.

5 The rest of the introduction offers a more detailed discussion, and Section 5 states formal results.
Figure 2: Marginal benefits for a given action profile \( \mathbf{a} \) can be represented as a network (a). Town \( i \) can help town \( j \) by mitigating pollution when air or water carries pollution from \( i \) to \( j \) (b).

The remainder of the introduction describes the key concepts and results in more detail, and also discusses connections to prior work.

**The Benefits Network** Let us return to the example of the towns. Assume for this discussion that the net marginal cost to an agent of increasing his own action, holding all else fixed, is unity.\(^6\) At action profile \( \mathbf{a} \), define \( \mathbf{B}(\mathbf{a}) \), the benefits matrix, as follows. Let \( B_{ij}(\mathbf{a}) \) denote the marginal benefit that \( i \) receives when \( j \) (different from \( i \)) increases his action, and let \( B_{ii}(\mathbf{a}) = 0 \) for all \( i \). This matrix can also be represented as a directed, weighted graph, where the towns are the nodes and the links correspond to strictly positive entries of the matrix. For instance, the link directed to \( Y \) from \( X \) has weight \( B_{YX}(\mathbf{a}) \), which measures the marginal benefit that accrues to \( Y \) when \( X \) increases its action at a profile \( \mathbf{a} \). The network corresponding to our example is depicted in Figure 2a, only the directed links having nonzero weights are shown. There is a link from town \( i \) to town \( j \) if pollutants flow in that direction (see Figure 2b).

**Pareto Efficient Outcomes and the Magnitude of Inefficiency** We begin with basic results on diagnosing and measuring inefficiency; these results do not depend on a selection of a particular Pareto improvement on the status quo, and do not involve Lindahl outcomes.

Proposition 1 in Section 3.1 shows that an interior action profile \( \mathbf{a} \) is Pareto efficient if and only if 1 is a largest eigenvalue of \( \mathbf{B}(\mathbf{a}) \). The reason for this is as follows. The matrix \( \mathbf{B}(\mathbf{a}) \) is a linear system describing how investments translate into returns at the margin. Consider a particular sequence of investments: in Figure 2a, \( Y \) can increase his action slightly and provide a benefit to \( Z \). Then \( Z \), in turn, can “pass forward” some of the resulting increase in his utility, investing effort to

---

\(^6\) The important assumption is that each agent has (more than) exhausted the net benefits available by unilaterally increasing his own action. That restricts the analysis to the interesting regime where there is a collective action problem. As for the choice of 1 as the marginal cost, it just streamlines some definitions (see Sections 2 and 3 for the general case).
help X. Finally, X can similarly create benefits for Y, completing a cycle. If they can all receive back more than they invest in such an adjustment, then the starting point is not Pareto efficient. It is precisely in such cases that the linear system \( B(a) \) is “expansive”: there is scope for everyone to get out more than they put in. And expansive systems are characterized by having a largest eigenvalue exceeding 1. If the largest eigenvalue of \( B(a) \) is less than 1, then everyone can be made better off by reducing investment. As a result, the Pareto-efficient outcomes have a benefits matrix with a largest eigenvalue exactly equal to 1. Section 3.1 derives this result. Section 3.4.2 formalizes the idea that cycles in the benefits network are crucial to Pareto improvements.

The main substantive implication of this result is that it allows us to analyze, in terms of the benefits network, which agents are essential to successful negotiations and to give a simple matrix criterion for finding them (Sections 3.4.4 and 3.4.5). These crucial agents are ones whose removal breaks important cycles in the benefits network, dropping the spectral radius of \( B(a) \) from above 1 to below 1. Section 8.1 applies these results in analyzing the temptation to free-ride.

Beyond characterizing Pareto efficiency, we would like to quantify the magnitude of inefficiency at the margin – defined loosely as the return on the best marginal investment in public goods. If agents’ preferences are denominated in a transferable numeraire (say, dollars), then it is easy to quantify this: just find the small change in contribution levels that maximizes the ratio \( \rho \) of marginal benefits to marginal costs in dollars. According to Coasian logic, all players agree that such an adjustment is the best one. There is another way of looking at this measure of inefficiency: suppose there are frictions in the economy, so that to sacrifice $1 for the sake of pollution reduction, a town must also waste an additional $t (e.g., due to polluting industries’ lobbying). That makes the real cost per unit of effective contribution equal to \( \tau = 1 + t \). Then the return on investment \( \rho \) in the frictionless economy is equal to the friction coefficient \( \tau \) that would have to prevail in the economy with frictions in order to make the status quo Pareto efficient.

If agents’ preferences are not quasilinear in any numeraire, utility is no longer transferable and the first approach to quantifying inefficiency no longer makes sense. We show, however, that the second way – the one based on a hypothetical friction – is portable easily to a world without such a numeraire (or without the possibility of transfers at all). Indeed, Section 3.2 derives that the measure of inefficiency defined via \( \tau \) is equal to the largest eigenvalue of the benefits matrix. Therefore, that invariant is useful not only for diagnosing inefficiency but also for rigorously quantifying its magnitude. Section 3.4.6 discusses implications when exact values of marginal costs and benefits are not known, but bounds are available.

**Characterizing the Lindahl Outcomes through Network Centrality** Theorem 1 in Section 4 formally establishes an equivalence between robust outcomes as defined in our discussion above and the Lindahl public goods solution. The main result of the paper characterizes these robust outcomes in terms of eigenvector centrality in the benefits network. We now outline this result.
What is Eigenvector Centrality? In general, suppose we are given a network, represented by a matrix $M$, where the nonnegative number $M_{ij}$ captures the strength of the link $ij$ (and is zero if that link is absent); this may differ from the strength of the link $ji$. A centrality measure is an assignment of a number to each node measuring how important that node is based on the connections in the network.\(^7\)

A widely studied measure is one that takes as a starting point the notion that an important node is one that is connected to other important nodes. A (right-hand) eigenvector centrality of the network $M$ is simply a nonnegative, nonzero vector $e$ whose entries sum to 1, with the property that for each $i$,

$$e_i = \gamma \sum_j M_{ij} e_j$$

for some proportionality constant $\gamma > 0$. If a vector $e$ satisfying these conditions exists and is uniquely determined, then we say $e_i$ is the eigenvector centrality of node $i$ according to matrix $M$. The equation says that the centrality of agent $i$ is proportional to a weighted sum of the centralities of others, where $e_j$ is weighted by $M_{ij}$ – the strength of the connection $ij$. By repeatedly plugging the equation into itself, it can be seen that the most central agents are the ones with strong direct and indirect connections to others in the network $M$ (see Section 6.4 for a precise statement of this).

Perhaps the most famous application of this idea is the PageRank measure introduced as a part of Google’s early algorithms to rank search results (Brin and Page, 1998). In that context, the nodes are web pages and $M_{ij}$ is the number of hyperlinks that go to page $i$ from page $j$, normalized by the total number of hyperlinks on page $j$. Then equation (2) says that a page is important if it is linked to by other important pages. Other applications include identifying those sectors in the macroeconomy that contribute the most to aggregate volatility via a network of intersectoral linkages (Acemoglu et al., 2012); the measurement of intellectual influence (Palacios-Huerta and Volij, 2004); and many others. We discuss the prior economic applications most closely related to our work – by Ballester, Calvó-Amengol, and Zenou (2006) and by Bramoullé, Kranton, and d’Amours (2011) – in the section on related literature below.

Centrality measures are intuitive and have useful mathematical properties, so it worth knowing when they can be useful in economic analysis. In all prior work we are aware of, conditions such as (2) are obtained in economic models by positing a setting with particular parametric structure (linear-quadratic, Cobb-Douglas, etc.) on preferences or production. Our contribution is a nonparametric connection between network centrality and an economic solution concept.

The Characterization Result. In our main result, we give a full characterization of the Lindahl outcomes in terms of eigenvector centrality.

As before, we assume for this introductory discussion that the marginal cost of

\(^7\)There are very many ways to define such a measure, depending on the interpretation of the network and the notion of importance that is relevant in a given context. A survey of many measures with references is provided by Jackson (2008, Section 2.2.4).
increasing one’s action, holding all else equal, is 1 for every agent. An eigenvector centrality action profile is defined by the property that

\[ a_i = \sum_j B_{ij}(a)a_j, \] (3)

for every i: player i’s action is the weighted sum of actions of other players j, where the weight on \( a_j \) captures how much i cares about the effort of j at the margin. This condition states that the action profile itself is a right-hand eigenvector of the marginal benefits matrix with eigenvalue \( \gamma \) equal to 1 – and this is intimately related to the efficiency characterization discussed above.

The main result of the paper, Theorem 2 in Section 5.2, asserts that an interior action profile is a Lindahl outcome if and only if it is an eigenvector centrality action profile. Intuitively, this means that the Lindahl outcomes are characterized by the condition that the agents contributing a lot to the public goods are the ones who are benefiting a lot, at the margin, from the efforts of others who are contributing. The conceptual import of this result is that it provides a general connection between efficient, robust outcomes of public goods problems and the theory of network centrality.

**Applications and Consequences.** The first application of the characterization is solving for the Lindahl outcome in an environment with global public goods, where each agent’s benefits depend only on the sum of others’ actions (Section 5.3). Here we find a very simple characterization of Lindahl outcomes: an agent contributes a fraction of the overall public good quantity equal to his marginal rate of substitution between the public good and private consumption.

Because the condition (3) involves \( B(a) \), which varies as a changes, we also build intuition by considering (in Section 6) special cases where the fixed point defined by (3) can be expressed explicitly in terms of an exogenous network that does not depend on a. Substantively, the theme of these results is that it is the intensity of benefits that an agent receives, and not of those he generates, that determine his effort level. However, it is not just i’s own links in the benefits network that matter: chains of benefits of the form “k helps j who helps i” are crucial to determining i’s centrality in the benefits network, and thus i’s effort level at Lindahl outcomes.

In studying these special cases, we also provide price-theoretic foundations for three of the most important centrality measures that have been studied in the networks literature.

**Related Work** This paper is at the intersection of the theory of public goods provision and the theory of economic networks. A recent literature has studied Nash equilibria of one-shot games in networks when best responses are linear in others’ actions and has related those equilibria to eigenvalue and centrality conditions. Key papers include Ballester, Calvó-Amengol, and Zenou (2006) on skill investment with externalities, and Bramoullé, Kranton, and d’Amours (2011) on local public goods;

---

8See Section 5 for the general case.

9In terms equation (2), this definition of an eigenvector centrality action profile requires \( \gamma = 1 \).
they offer more comprehensive surveys of the literature.\textsuperscript{10} A fundamental contrast is that our work focuses on institutions that implement Pareto efficient public goods outcomes, rather than on Nash behavior in the basic game without additional structure. From a technical perspective, these prior papers can be seen as addressing the question: “In which economic settings are eigenvalue and network centrality methods useful?” We provide a new answer that is complementary to the prior work. Since the Lindahl solution is formulated in terms of prices, the main theorem of the present paper can be seen as making the first general link between network centrality and price theory (as well as implementation theory). A final contrast is that our main characterization result does not impose functional form assumptions on preferences.

There is a large literature in implementation theory on game forms that yield Walrasian and Lindahl outcomes. Samuelson (1954) doubted that the price-theoretic public goods solutions first contemplated by Wicksell (1896) and Lindahl (1919) could be implemented despite agents’ incentives to manipulate a social institution. This inspired the early literature on implementation theory and mechanism design, with Groves and Ledyard (1977) showing that a game could be designed yielding Pareto efficient public goods provision in every Nash equilibrium. Key papers in the subsequent literature on implementing Walrasian and Lindahl outcomes include Hurwicz (1979a), Hurwicz (1979b), and Hurwicz, Maskin, and Postlewaite (1995). A broad survey is presented by Jackson (2001). The paper most important to us in this line of work is Hurwicz (1979a) who first suggested strategic foundations for Lindahl outcomes based on more primitive criteria – Pareto efficiency and individual rationality. We rely on the insights of this paper to establish what makes Lindahl outcomes special in our world.\textsuperscript{11}

Complementary foundations for Lindahl outcomes based on equilibria of bargaining games are offered by Dávila, Eeckhout, and Martinelli (2009) and Penta (2011); these works are part of a broader literature on Walrasian bargaining – see Yildiz (2003) and Dávila and Eeckhout (2008).

**The Plan of the Paper** Section 2 lays out the basic assumptions and notation. Section 3 collects all results on efficiency and on measuring inefficiency. Section 4 formally presents the implementation theory framework; defines the Lindahl solution; and states the theorem on its unique robustness. Section 5 presents the main theorem, characterizing the Lindahl outcomes as the eigenvector centrality action profiles. Section 6 considers special cases of the characterization to show how network structure affects robust outcomes; this also yields new economic foundations for three important network centrality measures. Section 7 ties up the theory; its main purpose is to explain the proof of the characterization result. Finally, Section 8 discusses limitations and extensions.

\textsuperscript{10}Most recently, Allouch (2012) has studied a network version of the setting introduced in the seminal paper of Bergstrom, Blume, and Varian (1986) on the voluntary (static Nash) private provision of public goods. Generalizing results of Bramoullé, Kranton, and d’Amours (2011), he derives comparative statics of public goods provision using network centrality tools.

\textsuperscript{11}For technical reasons we must provide our own proofs; e.g., Hurwicz (1979a) assumes positive endowments of all private goods, an assumption that does not have a reasonable analogue in our setting.
2 Framework

2.1 The Environment

Each member $i$ of a set $N = \{1, 2, \ldots, n\}$ of players (or agents) simultaneously chooses an effort level, or action$^{12}$, $a_i \in \mathbb{R}_+$. Taking a higher action should be interpreted as doing more of something that helps the other agents – for instance, mitigating pollution. Each player has a utility function $u_i : \mathbb{R}_+^n \rightarrow \mathbb{R}$; player $i$’s payoff when the action profile $a$ is played is $u_i(a)$.

2.2 Main Assumptions

Unless otherwise noted, the following assumptions are maintained throughout the paper.

Each $u_i : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is concave and continuously differentiable. The main substantive assumptions about the payoffs are:

Assumption 1 (Costly Actions). Each player finds it costly to invest effort, holding others’ actions fixed: $\frac{\partial u_i}{\partial a_i}(a) < 0$ for any $a \in \mathbb{R}_+^n$ and $i \in N$.

Assumption 2 (Positive Externalities). Increasing any player’s action level weakly benefits all other players: $\frac{\partial u_i}{\partial a_j}(a) \geq 0$ for any $a \in \mathbb{R}_+^n$ whenever $j \neq i$.

In brief, these conditions make the game a generalized prisoners’ dilemma. The unique Nash equilibrium of a game in which players choose their actions $a_i$ entails that everyone contributes nothing – $a_i = 0$ for each $i$ – even though other outcomes may Pareto dominate this outcome.

Two additional technical assumptions are useful:

Assumption 3 (Connectedness of Benefits). If $a \in \mathbb{R}_+^n$ and $M$ is a nonempty proper subset of $N$, then there exist $i \in M$ and $j \notin M$ (which may depend on $a$) such that $\frac{\partial u_i}{\partial a_j}(a) > 0$.

This posits that it is not possible to find an outcome and partition society into two nonempty groups such that, at that outcome, one group does not care about the effort of the other at the margin.

Finally, we assume that the set of points where everybody wants to scale up actions is bounded. To state this, we introduce a few definitions. Under a preference profile $u$, action profile $a' \in \mathbb{R}_+^n$ Pareto dominates another profile $a \in \mathbb{R}_+^n$ if $u_i(a') \geq u_i(a)$ for all $i \in N$, and the inequality is strict for some $i$. We say $a'$ strictly Pareto dominates $a$ if $u_i(a') > u_i(a)$ for all $i \in N$. Finally, $a$ is Pareto efficient if no other action profile Pareto dominates it.

---

$^{12}$We use $\mathbb{R}_+$ (respectively, $\mathbb{R}_{++}$) to denote to the set of nonnegative (respectively, positive) real numbers. We write $\mathbb{R}_+^n$ (respectively, $\mathbb{R}_{++}^n$) for the set of vectors $v$ with $n$ entries such that each entry $v_i$ is in $\mathbb{R}_+$ (respectively, $\mathbb{R}_{++}$). When we write an inequality between vectors, e.g. $v > w$, that means the inequality holds coordinate by coordinate, i.e. $v_i > w_i$ for each $i \in N$. 
Assumption 4 (Bounded Improvements). The set
\[
\{ \mathbf{a} \in \mathbb{R}_+^n : \text{there is an } s > 1 \text{ so that } s \mathbf{a} \text{ strictly Pareto dominates } \mathbf{a} \}
\]
is bounded.\(^{13}\)

2.3 Key Notions

The Jacobian We write \( \mathbf{u} = (u_1, u_2, \ldots, u_n) \) for a profile of utility functions. To keep track of the marginal costs and benefits of actions, we define the Jacobian, \( J(\mathbf{a}; \mathbf{u}) \), to be the \( n \)-by-\( n \) matrix whose \((i,j)\) entry is
\[
J_{ij}(\mathbf{a}; \mathbf{u}) = \frac{\partial u_i}{\partial a_j}(\mathbf{a}).
\]

The Benefits Matrix A close relative of this matrix turns out to be useful. The benefits matrix \( B(\mathbf{a}; \mathbf{u}) \) is defined as follows:
\[
B_{ij}(\mathbf{a}; \mathbf{u}) = \begin{cases} 
J_{ij}(\mathbf{a}; \mathbf{u}) - J_{jj}(\mathbf{a}; \mathbf{u}) & \text{if } i \neq j \\
0 & \text{otherwise}
\end{cases}
\]

The quantity \( B_{ij}(\mathbf{a}; \mathbf{u}) \) captures the marginal benefit an increase in \( j \)'s action provides to \( i \) per marginal unit of \( j \)'s cost.\(^{14}\) We call it the benefits matrix because it measures how much pleasure \( j \) can confer on \( i \) by enduring one unit of pain. Suppose \( \mathbf{u} \) satisfies the assumptions of Section 2.2. Since \( J_{jj}(\mathbf{a}; \mathbf{u}) < 0 \) by Assumption 1, this matrix is well-defined. By Assumption 2, it is entrywise nonnegative. Assumption 3 is equivalent to the statement that this matrix is irreducible\(^{15}\) at every \( \mathbf{a} \).

In discussing both the Jacobian and the benefits matrix, when there is no ambiguity about what \( \mathbf{u} \) is, we suppress it.

The Spectral Radius (Largest Eigenvalue) For any nonnegative matrix \( \mathbf{M} \), we define \( r(\mathbf{M}) \) as the magnitude of a largest eigenvalue of \( \mathbf{M} \), also called the spectral radius. By the Perron-Frobenius Theorem (see Section 2.4 below for a formal statement), any such matrix has a real eigenvalue equal to the maximum of all eigenvalues' magnitudes. Thus, we may equivalently think of \( r(\mathbf{M}) \) as the largest eigenvalue of \( \mathbf{M} \) on the real line.

This quantity can be thought of as a single measure of the “size” of the matrix. When applied to the matrix \( \mathbf{B} \), it will quantify, in a sense we will make precise, the marginal gains that can be generated by cooperating.

\(^{13}\)This condition is substantially weaker than assuming that the set of Pareto efficient outcomes is bounded, which fails to be true in many reasonable environments.

\(^{14}\)One might be concerned that the definition of this matrix is affected by changes such as multiplying one player’s utility by a scalar. Reassuringly, this type of change does not change any results. On this point, see Remark 1 in Section 3.1 and Remark 7 in Section 5.2.

\(^{15}\)A matrix \( \mathbf{M} \) is irreducible if it is not possible to find a nonempty subset \( S \) of indices so that \( M_{ij} = 0 \) for every \( i \in S \) and \( j \notin S \).
2.4 The Perron-Frobenius Theorem

The key mathematical tool we use is the Perron-Frobenius Theorem, a fundamental result about nonnegative matrices which will be applied to analyze the benefits matrix $B(a)$. The facts stated in this section are necessary only for proofs, and skipping this section (or any of the proofs in the paper) will not interrupt the flow. Nevertheless, since this powerful theorem underlies so much of the analysis, we state it here for easy reference.

Theorem (Perron-Frobenius\textsuperscript{16}). Let $M$ be an irreducible square matrix with no negative entries and spectral radius $r(M)$. Then:

(i) The real number $r(M)$ is an eigenvalue of $M$.

(ii) There is a vector $p$ (called a Perron vector) with only positive entries such that $Mp = r(M)p$.

(iii) If $v$ is a nonzero vector with nonnegative entries such that $Mv = qv$ for some $q \in \mathbb{R}$, then $v$ is a positive scalar multiple of $p$, and $q = r(M)$.

Statement (i) says that there is always a real, positive eigenvalue of the matrix $M$ that is at least as large as any other eigenvalue. Statement (ii) says that this eigenvalue, $r(M)$, is associated with a strictly positive eigenvector, the Perron vector. Statement (iii) says that any nonnegative eigenvector of the matrix $M$ is a scalar multiple of the Perron vector, and is associated with the special eigenvalue $r(M)$.

Note that because a matrix has exactly the same eigenvalues as its transpose, all the same statements are true, with the same eigenvalue $r(M) = r(M^\top)$, when we replace $M$ by its transpose $M^\top$. This observation yields a left-hand Perron eigenvector of $M$, i.e. a row vector $w$ such that $wM = r(M)w$.

3 Efficiency and the Spectral Radius

The thesis of this paper is that we can gain insight about the economics of the public goods problem by constructing, for any outcome $a$ under consideration, a network in which the agents are nodes and the weighted links among them measure the benefits available by increasing actions. The matrix corresponding to this network is $B(a)$, defined in Section 2.3 above. (Throughout this section, we fix a profile $u$ of utility functions and suppress it in the notation for the Jacobian and benefits matrix.)

This section offers support for our thesis by showing that an important statistic of this network – the size of the largest eigenvalue – can be used to diagnose whether an outcome is Pareto efficient (Section 3.1), and if not, to measure the magnitude of the inefficiency. This measurement can be done in two ways: via a hypothetical friction (Section 3.2) or by considering egalitarian marginal improvements (Section 3.3). After presenting these general results, we discuss interpretations (especially in terms of cycles) and applications in Section 3.4.

\textsuperscript{16}Meyer (2000, Section 8.3) has a comprehensive exposition of this theorem, its proof, and related results. Whether the Perron-Frobenius Theorem is all these statements or just (i) is a question on which conventions vary, but we will use the name as an umbrella term for all these facts; which one is relevant will be clear from context.
3.1 A Characterization of Pareto Efficiency

3.1.1 The Result

**Proposition 1.** Under the assumptions in Section 2.2, an interior action profile \( a \in \mathbb{R}^n_+ \) is Pareto efficient if and only if the spectral radius of \( B(a) \) is 1.

**Proof of Proposition 1:** For any nonzero \( \theta \in \mathbb{R}^n_+ \), define \( P(\theta) \), the Pareto problem with weights \( \theta \) as:

\[
\text{maximize } \sum_{i \in N} \theta_i u_i(a) \text{ subject to } a \in \mathbb{R}^n_+.
\]

It is well-known that concavity of all the \( u_i \) guarantees that the set of Pareto efficient points coincides with the set of solutions to this problem as \( \theta \) ranges over all the nonnegative vectors not equal to \( 0 \).

Suppose that an interior action vector \( a^* \) is Pareto efficient, and therefore solves \( P(\theta) \) for a vector \( \theta \). Since \( a^* \) is interior and the \( u_i \) are concave, that solution satisfies the system of first-order conditions \( \theta J(a^*) = 0 \). Using the definition of \( B \) (and writing \( I \) for the \( n \)-by-\( n \) identity matrix) this is equivalent to each of the following two statements:

\[
\theta (B(a^*) - I) = 0
\]

\[
\theta B(a^*) = \theta.
\]

Thus \( B(a^*) \) has an eigenvalue of 1 (with corresponding left-hand eigenvector \( \theta \)). Since \( B(a^*) \) is irreducible by Assumption 3 and has only nonnegative entries, the Perron-Frobenius Theorem applies to this matrix. That theorem (part (iii) of the statement in Section 2.4) says that the only eigenvalue of \( B(a^*) \) that is associated with a nonnegative eigenvector (in this case, the relevant eigenvector is \( \theta \)) is the largest real eigenvalue. Thus, 1 must be a largest eigenvalue of \( B(a^*) \).

Conversely, if \( B(a^*) \) has a largest eigenvalue of 1, then the Perron-Frobenius Theorem guarantees the existence of a nonnegative left-hand eigenvector \( \theta \) such that \( \theta B(a^*) = \theta \), so the first-order conditions of the Pareto problem are satisfied for weights \( \theta \) by the calculation above. By the assumption of concave utilities, it follows that \( a^* \) solves the Pareto problem for weights \( \theta \) (i.e., the first-order conditions are sufficient for optimality). \( \square \)

An intuition for the result based on passing benefits around a cycle and thinking of \( B(a) \) an expansive linear system was presented in the introduction.

**Remark 1.** The condition that the largest eigenvalue of \( B(a) \) is 1 is independent of how different players’ cardinal utilities are measured – as, of course, it must be, since Pareto efficiency is an ordinal notion. To see how the benefits matrix changes under reparameterizations of cardinal utility, suppose we define, for each \( i \in N \), new utility functions \( \hat{u}_i(a) = f_i(u_i(a)) \) for some smooth, strictly increasing functions \( f_i \). If we let \( \hat{B} \) be the benefits matrix obtained from these new utility functions, then 1 is a largest eigenvalue of \( \hat{B}(a) \) if and only if it is a largest eigenvalue of \( B(a) \), though
the corresponding eigenvectors may differ. Indeed, the move from \( u_i \) to \( \hat{u}_i \) induces a similarity transformation of \( B(a) \), so that \( \hat{B}(a) = F(u(a))B(a)F(u(a))^{-1} \), where \( F(u) \) is a diagonal matrix with \( F_{ii}(u_i) = f'_i(u_i) \). Therefore, \( \hat{B}(a) \) has the same set of eigenvalues as \( B(a) \).

From the perspective of maximizing a social planner’s welfare function (as in the proof), the invariance of the efficiency characterization can be seen as follows: to recover the same maximizer under utilities \( \hat{u} \), we simply need to adjust the Pareto weights \( \theta \).\(^{17}\)

Proposition 1 focuses on interior outcomes. One particular boundary outcome – the profile \( a = 0 \) in which nobody contributes – is particularly important. We can characterize when it is Pareto efficient without assuming irreducibility of the benefits matrix. A proof appears in Section A.2.

**Proposition 2.** Under Assumptions 1 and 2, the outcome \( 0 \) is Pareto efficient if and only if \( r(B(0)) \leq 1 \).

### 3.1.2 An Interpretation of the First-Order Conditions

It is worth remarking on an interpretation of the eigenvector way of looking at the Pareto efficiency condition. It says that if the social planner is using a weight vector \( \theta \), then for each \( i \),

\[
\theta_i = \sum_{j \neq i} \theta_j B_{ji}(a).
\]

Thus, at a Pareto efficient outcome, an agent gets a Pareto weight in proportion to how much his effort helps others; in proportion to how important those others are; and in proportion to how many others he helps. This makes sense: the social planner would not make an important player work unless it yielded something valuable (evaluated according to the planner’s weights).

In network language, the Pareto weights \( \theta \) are proportional to the eigenvector centrality of the network \( B(a)^T \); the transpose is used here because we defined eigenvector centrality as a right-hand eigenvector of a matrix, while \( \theta \) is a left-hand eigenvector of \( B(a) \).

### 3.2 Quantifying Inefficiency: The Frictions Approach

The size of the largest eigenvalue of the benefits matrix does not merely diagnose inefficiency, but also measures its magnitude. In particular, we will show that to deter a social planner from wanting to change actions at a profile \( a \), effort must be taxed at a rate equal to the largest eigenvalue of \( B(a) \) minus one.

To formalize this, this subsection focuses on a setting where costs and benefits are separable. Suppose that for each \( i \),

\[
u_i(a) = v_i(a_{-i}) - k(a_i),
\]

\(^{17}\)In particular, set the new vector of Pareto weights to \( \hat{\theta} = \theta F(u(a))^{-1} \).
where $a_{-i}$ is the vector of actions of players excluding $i$. Suppose also that the resulting utility function satisfies the assumptions Section 2.2. Here $v$ measures the value agent $i$ gets from others’ contributions, and $k$ is his own cost of contributing. Construct a related economy in which frictions (e.g., political economy costs) dissipate some of the effort agents contribute to the public goods. More precisely, we posit that agent $i$, in order to confer benefits equivalent to contributing $a_i$ in the original game, must multiply his costs by $\tau$, for some $\tau \geq 0$. The utility functions under this assumption are given, for each $i$, by

$$u^{(\tau)}_i(a) = v_i(a_{-i}) - \tau k(a_i).$$

The quantity $t = \tau - 1$ may be thought of as a tax rate on effort. Then we have the following corollary of Proposition 1.

**Corollary 1.** Fix a positive real number $\tau$. The interior action profile $a$ is a Pareto efficient outcome under $u^{(\tau)}$ if and only if $\tau$ is a largest eigenvalue of $B(a; u)$.

**Proof** This follows from Proposition 1 by observing that, for all $a \in \mathbb{R}^n_+$, it holds that $B(a; u^{(\tau)}) = \tau^{-1}B(a; u)$. \\Thus, to rationalize an outcome $a$ as being Pareto efficient, a tax rate of $t = r(B(a)) - 1$ on effort is required; effort must be taxed at a rate equal to the largest eigenvalue of $B(a)$ minus one. Correspondingly, if $r(B(a)) < 1$, then effort must be subsidized at the margin to render the outcome $a$ efficient. This is a sense in which the size of the largest eigenvalue of $B(a)$ is a quantitative measure of inefficiency.

**Remark 2.** The spectral radius is a continuous function of the matrix entries (see, e.g., Wilkinson, 1965). It is reassuring that our measure of inefficiency varies continuously in the data of the problem.

### 3.3 Quantifying Inefficiency: The Value of Egalitarian Improvements

There is another way to think of the spectral radius of $B(a)$ as measuring inefficiency. In brief, if we consider a small improvement on the status quo which equalizes everyone’s benefit-to-cost ratio, then that ratio will be equal to the spectral radius of $B(a)$.

Let $\Delta_n$ denote the simplex in $\mathbb{R}^n_+$ defined by $\Delta_n = \{a \in \mathbb{R}^n_+: \sum_i a_i = 1\}$.

**Definition 1.** The **bang for the buck** vector $b(a, d)$ at an action profile $a$ along a direction $d \in \Delta_n$ defined by

$$b_i(a, d) = \frac{\sum_{j: j \neq i} J_{ij}(a) d_j - J_{ii}(a) d_i}{J_{ii}(a) d_i}.$$ 

\[18\] Formally, since $\tau$ can be between 0 and 1, the setup also allows for effort being subsidized by an outside benefactor; in that case, $t < 0$. 

13
The thought experiment behind this definition is as follows: suppose that we start at point \( a \) and consider increasing each agent \( i \)'s effort by \( \epsilon d_i \), for some small \( \epsilon > 0 \). Then \( b_i(a, d) \) measures how much agent \( i \) receives in benefits (from others' increases in effort) per unit of cost incurred in increasing his own action\(^{19} \). In other words, it is equal to the ratio \( \frac{i's \ marginal \ benefit}{i's \ marginal \ cost} \) evaluated at \( a \) with respect to this hypothetical deviation from \( a \).

We say a direction \( d \in \Delta_n \) is egalitarian if every entry of \( b_i(a, d) \) is the same. That is, each agent’s sacrifice (relative to free-riding on this marginal improvement by leaving his own effort unchanged) converts into benefits at the same rate. This can also be viewed as a Rawlsian direction: it maximizes the minimum return any agent obtains on investment per unit of cost.\(^{20} \) If agents negotiate over small improvements and fairness or equality concerns are important, egalitarian directions may be a relevant benchmark because agents intrinsically care about them. However, and perhaps surprisingly, egalitarian directions also arise naturally in the context of solutions to public goods problems with more fundamental strategic foundations (see Remark 6 in Section 5.2).

**Proposition 3.** At any \( a \), there is a unique egalitarian direction \( d^{eg}(a) \). Every entry of \( b(a, d^{eg}(a)) \) is equal to the spectral radius of \( G(a) \).

When the spectral radius of \( G(a) \) exceeds 1, then by increasing actions in the direction given by \( d^{eg}(a) \), an egalitarian Pareto improvement can be achieved, and the benefit-to-cost ratio of that change is equal to \( r(G(a)) \) for every agent. When the spectral radius of \( G(a) \) is less than 1, then by decreasing actions in the direction \( -d^{eg}(a) \), each agent foregoes \( r(G(a)) \) units of benefit per unit of cost saved, and that yields is a Pareto improvement as well.

### 3.4 Interpretations and Applications

While the spectral radius of the benefits matrix has the economic interpretations laid out in the previous subsections, it is useful to give some intuition as to what this quantity corresponds to in the structure of a network. One may also wonder how it may be estimated based on incomplete data. This section addresses these issues.

Section 3.4.1 gives a very simple example in which the benefits matrix is a cycle and the spectral radius measures the magnitude of the benefits that are generated along that cycle. Section 3.4.2 lays out the general facts about how cycles in the benefits network relate to its spectral radius, fleshing out the claims in the introduction about the importance of cycles to enabling Pareto improvements.

Section 3.4.3 shows a monotonicity result: when links in the benefits network get weaker or disappear, its spectral radius decreases. Sections 3.4.4 and 3.4.5 explore an

\(^{19}\)Formally, \( i \)'s cost is defined as the difference in \( i \)'s payoff between (i) leaving his own action unchanged at \( a_i \), (ii) increasing his own action by \( \epsilon d_i \) (with \( a_j \) unchanged for every \( j \neq i \)). Similarly, \( i \)'s benefit is defined as the difference in his payoff between (iii) all other \( j \) increasing actions by \( \epsilon d_j \) while \( i \) holds his own action fixed at \( a_i \) and (iv) nobody increasing actions.

\(^{20}\)For a proof, see Section A.1
important special case, discussing the decrease in a group’s potential for cooperation when some of its members are unable to contribute. These sections focus particularly on the question of which agents are essential, in the sense that their absence eliminates the potential for any Pareto improvements on the status quo. Finally, Section 3.4.6 applies the monotonicity result to bounding the gains from cooperation under imperfect measurement of marginal costs and benefits.

3.4.1 A Simple Example

Suppose $N = \{1, 2, 3\}$ and

$$B(0) = \begin{bmatrix} 0 & 0 & 7 \\ 5 & 0 & 0 \\ 0 & 6 & 0 \end{bmatrix}.$$ 

This network is depicted in Figure 3. It can be computed that $r(B(0)) = (5 \cdot 6 \cdot 7)^{1/3} \approx 5.94$. Thus, there is substantial potential for cooperation in this network. The spectral radius of $B(0)$ is the geometric mean of weights in the benefits network along the cycle – the product of these weights to the power $1/\ell$, where $\ell$ is the cycle’s length. We state without proof that this is always the case when there is just one cycle in the network. Facts 1 below is a general bound that builds on this observation.

![Figure 3: The network of the example. An arrow from from $i$ to $j$ indicates that $B_{ji} > 0$ (i.e. benefits flow from $i$ to $j$) and the link is labeled by the weight $B_{ji}$.](image)

3.4.2 Interpreting Largest Eigenvalues: Cycles of Mutual Benefit

As noted in the introduction, if we can find a cycle of players such that each can help the next, then that creates scope for cooperation. The larger the benefits that can be passed along a cycle, the greater that scope. This section formalizes the idea that the spectral radius of a matrix can be interpreted generally as measuring the intensity of such cycles.

**Definitions** A (directed) cycle of length $\ell$ in the matrix $M$ is a sequence

$$(c(1), c(2), \ldots, c(\ell))$$
of elements of $N$ (player indices), where $c(1) = c(\ell)$ – i.e., the cycle starts and ends at the same node such that $M_{c(t)c(t+1)} > 0$ for each $t \in \{1,\ldots,\ell - 1\}$.\footnote{Nodes other than the first and last may also be repeated in this sequence. The convention on length – that the cycle $i,j,i$ is said to be of length 3 even though it involves only two “steps” – is slightly awkward but useful for consistency with the discussion in Section 6.3.} Let $\mathcal{C}(\ell; \mathbf{M})$ be the set of all cycles of length $\ell$ in matrix $\mathbf{M}$. For any nonnegative matrix $\mathbf{M}$, define the value of a cycle $c \in \mathcal{C}(\ell; \mathbf{M})$ as

$$v(c; \mathbf{M}) = \prod_{t=1}^{\ell-1} M_{c(t)c(t+1)}.$$  \hfill (4)

This is the number obtained by taking the product of all weights along the cycle $c$ in the weighted directed graph defined by $\mathbf{M}$. Define

$$V(\ell; \mathbf{M}) = \sum_{c \in \mathcal{C}(\ell; \mathbf{M})} v(c; \mathbf{M}).$$

This is the sum of the values of all cycles of length $\ell$ in matrix $\mathbf{M}$.

Having made these definitions, we can show how the spectral radius of the benefits matrix is related to the total value of such cycles.

**A Note on the Directions of Links and Cycles** Before this, it is worth remarking on a potentially confusing wrinkle of the notation. Since the thing that is naturally thought of as traveling among players is the positive externalities they confer on each other, we think of the strength of the link directed from $i$ to $j$ as $B_{ji} = \partial u_j / \partial a_i$ – i.e., how much $i$ can help $j$ (recall Figure 2 or Figure 3, for example). However, when a matrix $\mathbf{B}$ is represented as a weighted directed graph, the typical convention is to label an arrow from $i$ to $j$ by $B_{ij}$. These two conventions conflict. This problem goes away if we instead focus on $\mathbf{B}^T$, the transpose of $\mathbf{B}$, when we draw the benefits network or discuss cycles in it. For example, if we consider the cycle $c = (1,2,3)$, then $v(c; \mathbf{B}^T) = B_{21}B_{32}B_{13}$ multiplies three factors: how much 1 helps 2, how much 2 helps 3, and how much 3 helps 1.

**The Results** First, any cycle provides a simple lower bound on the spectral radius of a matrix. A proof of this fact is in Section A.8.

**FACT 1.** For any nonnegative matrix $\mathbf{M}$ and any cycle $c$ of length $\ell$ in $\mathbf{M}$,

$$r(\mathbf{M}) \geq v(c; \mathbf{M})^{1/\ell}.$$  

Moreover, the total value of long cycles provides an asymptotically exact estimate to the spectral radius in the following sense:

**FACT 2.** For any nonnegative matrix $\mathbf{M}$,

$$r(\mathbf{M}) = \limsup_{\ell \to \infty} V(\ell; \mathbf{M})^{1/\ell}.$$
Proof of Fact 2: Note that $V(\ell; M) = \text{trace}(M^\ell)$. With this replacement, the fact is standard – see, e.g. Milnor (2001).

From these formulas we can conclude that $V(\ell; B(a)^T)$, the total magnitude of benefit cycles\(^{22}\) at action profile $a$, grows roughly as $e^{\ell \log r(B(a))}$. Thus, the logarithm of the spectral radius of the benefits matrix can be thought of as the growth rate of the cycle values as we increase the cycles’ length.

Intuitively, benefit networks with an imbalanced structure, in which it is rare for the beneficiaries of one agent’s effort to be able to directly or indirectly “give back”, will have a lower spectral radius and, in the senses we have articulated earlier in this section, less scope for cooperation. Section 3.4.5 gives a concrete example where this insight is key to determining which agents are essential.

3.4.3 Comparing Marginal Returns

We recall a standard fact from linear algebra (see, e.g., Debreu and Hernstein 1953, Theorem I\(^{*}\)).

**FACT 3.** If $B$ and $\hat{B}$ are two nonnegative matrices such that $B \geq \hat{B}$, then $r(B) \geq r(\hat{B})$.

A key economic consequence of this is a version of diminishing marginal returns: the spectral radius of the benefits matrix decreases as actions increase.

**PROPOSITION 4.** If $a' \geq a$, then $r(B(a')) \leq r(B(a))$.

**Proof of Proposition 4:** By the concavity of the $u_i$, the inequality $B(a') \leq B(a)$ holds (where the inequality between matrices means entrywise inequality).\(^{23}\) Then the fact above implies the result.

To formalize the interpretation of this as diminishing marginal returns, recall that according to Section 3.3, the spectral radius measures the returns to egalitarian investments at the margin.

3.4.4 How Essential is a Player?

The efficiency results provide a simple way of quantifying how essential any given player is to the negotiations. Suppose for a moment that a given player exogenously may or may not be able to participate in an institution to negotiate an outcome that Pareto dominates the status quo. If he is not able, then his action is set to the

\(^{22}\)Here we use that the spectral radius does not change if we take the transpose. Of course, it is perfectly legitimate to think of directed cycles in $B$, without transposing; this way of looking at things simply leads to a different interpretation. In that case, the cycle $(1, 2, 3)$ is a cycle of dependence – 1 depends on the help of 2, who depends on the help of 3.

\(^{23}\)This is established as follows. Fix any $i$. Concavity of $u_i$ implies that the Hessian of $u_i$ in $a$ is negative semidefinite. Therefore, each diagonal entry of this Hessian is nonpositive, and the principal minor of the Hessian corresponding to leaving only two indices $i$ and $j$ is nonpositive. Then it follows that $-J_{ii}(a)$ is increasing in $a_i$ and $J_{ij}(a)$ is decreasing in $a_j$. But $i$ and $j$ were arbitrary, and we can get from $a$ to $a'$ by increasing actions one by one. That implies the needed inequality.
status quo level of $a_i = 0$. How much does such an exclusion hurt the prospects for cooperation by the rest of the society?

Without player $i$, the benefits matrix at the status quo of 0 is equal to the original $B(0)$ without row and column $i$; equivalently, each entry in that row and column may be set to 0. Call a matrix constructed that way $B[−i](0)$. By Proposition 3, the spectral radius of $B[−i](0)$ is smaller than that of $B(0)$. The most dramatic case is one in which the spectral radius of $B(0)$ exceeds 1, but the spectral radius of $B[−i](0)$ is less than 1. Then by Proposition 3 on egalitarian improvements, a Pareto improvement on 0 exists when $i$ is present. But by Proposition 2, none exists when 0 is absent.

This argument shows that player $i$’s participation is essential to achieving any Pareto improvement on the status quo precisely when his removal changes the spectral radius of the benefits matrix at the status quo from being greater than 1 to being less than 1.

More generally, the difference $r(B(0)) − r(B[−i](0))$ measures how much the potential for cooperation at the margin is damaged when player $i$ does not participate. The strategic implications of this are discussed in Section 8.1.

### 3.4.5 A More Elaborate Example: Who is Essential?

We now build on the example of Section 3.4.1 to illustrate what it means for a player to be essential. Suppose $N = \{1, 2, 3, 4\}$ and

$$B(0) = \begin{bmatrix} 0 & 0 & 7 & 0.5 \\ 5 & 0 & 6 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0 \end{bmatrix}. \quad (5)$$

See Figure 4 for a graphical depiction of this benefits matrix (panel b) and a comparison with the example of Section 3.4.1.

![Figure 4](image_url)

(a) The example of Section 3.4.1. (b) The benefits network of equation (5).

Figure 4: The benefit flows in (b) differ from those of (a) in two ways. First, the arrow from 2 to 3 has been flipped, destroying the directed cycle $1 \rightarrow 2 \rightarrow 3$. Second, a new player (#4) has been added, with bilateral links of weight 0.5 in the benefits matrix to all other players.
The import of the example is that player #4, even though he confers and receives the smallest benefits, is the only essential player. Without him, there are no cycles at all and the spectral radius of the corresponding benefits matrix $B^{[-4]}(0)$ is 0 (by Fact 2). On the other hand, when he is present but any one other player $i \neq 4$ is absent, then there is always a cycle whose edges multiply to more than 1, and the spectral radius of $B^{[-i]}(0)$ exceeds 1 (by Fact 1). Thus, the participation of a seemingly “small” player in negotiations can make an essential difference to the ability to improve on the status quo when that player completes cycles in the benefits network.

3.4.6 Bounding the Marginal Gains from Cooperation

In practical situations, the entries of the matrix $B(a)$ may be known imperfectly, and some of them may not be known at all. Can an analyst give useful advice about how inefficient an outcome is in this circumstance?

We now explain the sense in which such advice under imperfect information is possible. If we can find an estimator $\hat{B}(a)$ such that $\hat{B}_{ij}(a) \leq B_{ij}(a)$, then $r(\hat{B}(a))$ provides a lower bound on $r(B(a))$ by the monotonicity of the spectral radius (recall Section 3.4.3 above). In particular, if we have good theoretical reasons to believe that externalities are nonnegative for all pairs of players $i$ and $j$ but have no way to estimate some of the marginal benefits, then we can simply set $\hat{B}_{ij} = 0$ for those pairs.

By the results of Sections 3.2 and 3.3, a lower bound on $r(B(a))$ has practical meaning; if we have such an estimate $\hat{\tau}$, then we know that a tax on contributions of at least $\hat{\tau} - 1$ is required to rationalize that outcome as Pareto efficient. We know that there is an egalitarian vector that yields a bang for the buck of at least $\hat{\tau}$ at the current action profile. Thus, even partial knowledge of the network of benefit flows can be very useful. However, the results in Section 3.4.2 do tell us that for these estimates to yield a conclusion that the status quo is inefficient, there must be some cycles in the estimated benefits matrix.

Since in actual estimation, $\hat{B}(a)$ is a random variable, there is a question of formulating a probabilistic version of the monotonicity result. This is a direction for future work.

4 Lindahl Outcomes: Foundations

This section briefly reviews the framework and notation of implementation theory (Section 4.1) and formally defines Lindahl outcomes (Section 4.2). The purpose of the section is to state Theorem 1 in Section 4.3, which characterizes the Lindahl outcomes as the robustly implementable ones, in the sense discussed in the introduction.

4.1 The Implementation Theory Framework

We consider a designer of an institution who can leave future participants with a game to play. She does not know what preferences they will have. She believes they will play an equilibrium, but does not know which equilibrium. She would like to
design the game so that all its equilibria have some desirable properties. What can she do? Which outcomes can she achieve?

Let $\mathcal{U}_A$ be the set of all functions $u : \mathbb{R}^n_+ \to \mathbb{R}$. We denote by $\succeq_u$ and $\succ_u$ the weak and strict preference orderings, respectively, induced by $u \in \mathcal{U}_A$. The domain of possible preference profiles$^{24}$ is a set $\mathcal{U} \subseteq \mathcal{U}_A^n$.

A game form is a tuple $H = (\Sigma_1, \ldots, \Sigma_n, g)$ where:

- $\Sigma_i$ is a set of strategies that agent $i$ can play; we write $\Sigma = \prod_{i \in N} \Sigma_i$;
- $g : \Sigma \to \mathbb{R}^n_+$ is the outcome function that maps strategy profiles to action profiles.

**Definition 2.** In a game form $H = (\Sigma_1, \ldots, \Sigma_n, g)$, a strategy profile $\sigma \in \Sigma$ is a Nash equilibrium for preference profile $u \in \mathcal{U}$ if for any $\tilde{\sigma}_i \in \Sigma_i$, it holds that $g(\sigma) \succeq_u g(\tilde{\sigma}_i, \sigma_{-i})$. We define $\Sigma^*(H, u)$ to be the set of all such $\sigma$.

A social choice correspondence $F : \mathcal{U} \Rightarrow \mathbb{R}^n_+$ maps each preference profile to a nonempty set of outcomes. Any game form for which equilibrium existence is guaranteed$^{25}$ naturally induces a social choice correspondence: its Nash equilibrium outcome correspondence $F_H(u) = g(\Sigma^*(H, u))$. The set $F_H(u)$ describes all the outcomes the participants with preferences $u$ can end up with if they are left with a game form $H$ and they play some Nash equilibrium. We say that $F_H$ is the social choice correspondence that the game form $H$ implements$^{26}$. A social choice correspondence is said to be implementable if there is some game form $H$ that implements it.

There are two basic normative criteria we impose on such correspondences.

**Definition 3.** A social choice correspondence $F$ is Pareto efficient if, for any $u \in \mathcal{U}$ and $a \in F(u)$, the profile $a$ is Pareto efficient under $u$.

**Definition 4.** A social choice correspondence $F$ is individually rational if, for any $u \in \mathcal{U}$ and $a \in F(u)$, it holds that $g(a) \succeq_u 0$.

An individually rational social choice correspondence is one that leaves every player no worse off than the status quo. For more on how to interpret this condition, see Section 8.1.

There is also a technical condition – upper hemicontinuity.

**Definition 5.** A social choice correspondence $F$ is upper hemicontinuous if, for any sequence $(u^{(k)})$ of preference profiles converging compactly$^{27}$ to $u$, and any sequence of outcomes $(a^{(k)})$ such that $a^{(k)} \in F(u^{(k)})$ for every $k$ and $a^{(k)} \to a$, we have $a \in F(u)$.

This condition has some normative appeal in that a social choice correspondence not satisfying upper hemicontinuity is sensitive to arbitrarily small changes in preferences that may be difficult for the agents themselves to detect.

$^{24}$The standard approach (e.g. Maskin, 1999) is to work with preference relations. We use sets of utility functions to avoid carrying around two parallel notations.

$^{25}$Otherwise, we can still talk about the correspondence, but it will not be a social choice correspondence, which is required to be nonempty-valued.

$^{26}$To be more precise, this is full Nash implementation. Since we only consider this kind of implementation, we drop the adjectives.

$^{27}$That is, converging uniformly on every compact set.
4.2 Definition of Lindahl Outcomes

Informally, a Lindahl outcome is an analogue, in a public goods setting, of a Walrasian equilibrium allocation. Instead of standard prices for private goods, there are personalized taxes and subsidies: each player pays a tax for every public good he enjoys (in proportion to how much of that public good is produced), and receives a personalized subsidy (financed by others’ taxes) per unit of effort he invests in the public good he provides. More formally:

**Definition 6.** An action profile \( a^* \) is a *Lindahl outcome* for a preference profile \( u \) if there is an \( n \)-by-\( n \) matrix \( P \) so that the following conditions hold for every \( i \):

(i) The inequality

\[
\sum_{j:j\neq i} P_{ij}a_j \leq a_i \sum_{j:j\neq i} P_{ji}
\]

is satisfied when \( a = a^* \);

(ii) for any \( a \) such that \( BB_i(P) \) is satisfied, we have \( a^* \succeq_{u_i} a \).

The *Lindahl correspondence* \( L : U \rightrightarrows \mathbb{R}^n_+ \) is defined by

\[
L(u) = \{ a \in \mathbb{R}^n_+ : a \text{ is a Lindahl outcome for } u \}.
\]

In terms of interpretation, \( P_{ij} \) is the price \( i \) pays \( j \) for the effort of \( j \), and \( \sum_{j:j\neq i} P_{ji} \) is the “wage” or subsidy player \( i \) receives per unit of effort. Thus, \( BB_i(P) \) is simply a budget balance condition: expenditures are no greater than income.

The definition says that an action profile is a Lindahl outcome if there are some prices so that the amounts of externalities produced and consumed are optimal for each player, subject to budget balance at those prices. For any \( u \) satisfying the assumptions of Section 2.2, the set \( L(u) \) is nonempty – see Section 7.1.

This definition, though simple, does not relate in an obvious way to realistic markets; Samuelson (1954) elaborates on this. It is the implementation-theoretic rationale that we are about to formally state that renders the Lindahl outcomes worthy of study.

4.3 The Hurwicz Rationale for Lindahl Outcomes

Fix \( U \). Let \( F \) be the set of implementable social choice correspondences \( F : U \rightrightarrows \mathbb{R}^n_+ \) that are Pareto efficient, individually rational, and upper hemicontinuous. For any \( u \in U \), define the set of outcomes in *every* such correspondence:

\[
R(u) = \bigcap_{F \in \mathcal{N}} F(u).
\]

This defines a correspondence \( R : U \rightrightarrows \mathbb{R}^n_+ \). We call this the *robustly attainable* correspondence.

For any correspondence \( F : U \rightarrow \mathbb{R}^n_+ \), we define the interior part of \( F \) by

\[
F^\text{int}(u) = F(u) \cap \mathbb{R}^n_{++}.
\]
We now state the characterization of the robustly attainable correspondence.

**Theorem 1.** Suppose $\mathcal{U}$ is the set of all preference profiles satisfying the assumptions of Section 2.2, and the number of players $n$ is at least 3. Then the interior part of the robustly attainable correspondence is equal to that of the Lindahl correspondence: $R^\text{int} = L^\text{int}$.

As we note in Section 7.1 below, $L^\text{int}$ is nonempty. A proof of Theorem 1 can be found in Section 7.2. As we discuss formally in Section A.3, it is not necessary to assume that $\mathcal{U}$, the set of all preference profiles the designer considers possible, is equal to the set of all preference profiles satisfying the assumptions of Section 2.2. It suffices, for example, to assume that $\mathcal{U}$ contains a neighborhood of every linear utility function. Actually, $\mathcal{U}$ may be even smaller than this.

The above theorem is an analogue of Theorem 3 of Hurwicz (1979a). Because the environment studied in that paper (with assumptions such as nonzero endowments of all private goods) is not readily adapted to ours, we prove the result separately, using Hurwicz’s insights combined with Maskin’s Theorem.

### 5 The Main Result: A Network Centrality Characterization of the Lindahl Outcomes

Having explained what makes Lindahl outcomes special from an implementation perspective, we present our main results on their characterization. Section 5.1 states two equivalent definitions of eigenvector centrality action profiles. Section 5.2 is devoted to Theorem 2, which asserts, in essence, that the Lindahl outcomes are the same as the eigenvector centrality action profiles. Section 5.3 gives a simple first application: characterizing the outcome when there is a global public good whose production is additive and symmetric in the players’ effort levels. Section 5.4 derives a general inequality on contribution levels at Lindahl outcomes: agents who enjoy all public goods more also provide more themselves.

#### 5.1 Eigenvector Centrality Action Profiles

**5.1.1 Definition**

Fix a preference profile $\mathbf{u}$ throughout this subsection and the next. The main theorem characterizes Lindahl outcomes as ones satisfying a network centrality condition. In the introduction, we discussed this condition in the special case where $\frac{\partial u_i}{\partial a_i}(\mathbf{a}) = -1$ for each $\mathbf{a} \in \mathbb{R}_+^n$ and $i \in N$. To state the condition more generally, we first define the *contribution* of player $i$ at an action profile $\mathbf{a}$ as

$$c_i(\mathbf{a}) = -a_i \cdot \frac{\partial u_i}{\partial a_i}(\mathbf{a}).$$

One interpretation of this is as follows. Suppose $i$ subcontracted taking the action $a_i$ to an identical copy of herself. If $i$ paid her twin a wage in utiles equal to the twin’s marginal cost, then $i$’s total wage bill for this contract would be $c_i(\mathbf{a})$. Note that by
concavity of the $u_i$, the function $c_i$ is increasing in $a_i$ (holding the other components of $a$ fixed).

We can now define eigenvector centrality action profiles.

**Definition 7.** An action profile $a \in \mathbb{R}_+^n$ is an eigenvector centrality action profile if $a \neq 0$ and $B(a)c(a) = c(a)$.

The name comes from the fact that $c(a)$ is, according to the condition, a right-hand eigenvector of $B(a)$ with eigenvalue 1. In other words, for each $i \in N$,

$$
c_i = \sum_j B_{ij}c_j. \tag{7}
$$

Equation (7) asserts that each player’s contribution is a weighted sum of the other players’ contributions, where the weight on $c_j$ is proportional to the marginal benefits that $j$ provides to $i$.

In the case where $\frac{\partial u_i}{\partial a_i}(a) = -1$ for each $a \in \mathbb{R}_+^n$ and $i \in N$, the contribution is simply equal to the action, so $c(a) = a$. In that case, $a$ is an eigenvector centrality action profile if, for each $i \in N$,

$$
a_i = \sum_{j \neq i} B_{ij}a_j.
$$

**5.1.2 An Alternative Definition: Scaling-Indifference**

An eigenvector centrality action profile is equivalently characterized by a convenient property called scaling-indifference, which we define and then explain.

**Definition 8.** An action profile $a \in \mathbb{R}_+^n$ satisfies scaling-indifference (or is scaling-indifferent) if $a \neq 0$ and $J(a)a = 0$.

The vector $J(a)v$ gives the derivatives of utilities in $\epsilon$ when actions are changed from $a$ to $a + \epsilon v$ for some vector $v \in \mathbb{R}^n$. That is, to a first-order approximation, $u(a + \epsilon v) \approx u(a) + \epsilon J(a)v$. Suppose now that actions are scaled by $1 + \epsilon$, for some small real number $\epsilon$; this corresponds to setting $v = a$. If $J(a)a = 0$, then all players are indifferent, at the margin, to this small proportional perturbation in everyone’s actions.

By Lemma 2, proved in Section A.4 (a very simple calculation), the vector $a$ is an eigenvector centrality action profile if and only if it satisfies scaling-indifference.

**5.2 The Characterization**

We now present our main theorem, characterizing Lindahl outcomes.

**Theorem 2.** Under the assumptions in Section 2.2, the following are equivalent for a nonzero $a \in \mathbb{R}_+^n$:

(i) $a$ is a Lindahl outcome;

(ii) $J(a)a = 0$ – i.e., $a$ is scaling-indifferent;
(iii) \( B(a)c(a) = c(a) \), i.e., \( a \) is an eigenvector centrality action profile.

The implications (ii) \( \Rightarrow \) (iii) \( \Rightarrow \) (i) hold for any nonzero \( a \in \mathbb{R}_n^+ \).

The key reasons behind this characterization are discussed in Section 7.3. The formal proof appears in Section A.5.

**Remark 3.** The issue of when the profile \( 0 \) is a Lindahl outcome is treated in Proposition 7 in Section A.6; the most important fact is that \( 0 \) is a Lindahl outcome if and only if it is a Pareto efficient outcome.

**Remark 4.** The fact that the implication (i) \( \Rightarrow \) (ii) is proved for interior \( a \) is not particularly restrictive in the following sense. If \( a \) is a Lindahl outcome, then it continues to be one when the agents taking action 0 are removed from the economy. Then the characterization above applies, with all matrices restricted to the remaining agents.

The main theorem implies some simple consequences, tying up loose ends from the earlier exposition.

**Remark 5.** By Theorem 2, at any interior Lindahl outcome \( a \), the matrix \( B(a) \) has a nonnegative right eigenvector \( c(a) \), and therefore, by the Perron-Frobenius Theorem (part (iii)), a spectral radius of 1. Proposition 1 therefore implies the Pareto efficiency of \( a \). Of course, the standard proof of the First Welfare Theorem also goes through: see, e.g., Foley (1970).

**Remark 6.** The scaling-indifference condition is equivalent to the assertion that \( d = a/\sum_i a_i \) is an egalitarian direction at \( a \) with corresponding bang for the buck \( b(a, d) = 1 \) (recall Section 3.3). Therefore, we can alternatively characterize Lindahl outcomes as Pareto efficient ones under which moving in the direction given by \( a \) itself is egalitarian.

**Remark 7.** Note that the condition \( J(a)a = 0 \) does not change if we form new utility functions according to the transformation discussed in Remark 1 in Section 3.1. The Jacobian of the “new” utilities \( \tilde{u} \) is, by the chain rule, simply \( F(u(a))J(a) \), where \( F(u) \) is a diagonal matrix with \( F_{ii}(u_i) = f_i'(u_i) \). Since \( F(u(a)) \) is nonsingular for the transformation being considered, it follows that \( J(a)a = 0 \) if and only if \( F(u(a))J(a)a = 0 \).

**Remark 8.** The condition \( J(a)a = 0 \) is a system of \( n \) equations in \( n \) unknowns (the coordinates of \( a \)). By a standard Sard’s Theorem argument, this entails that for generic utility functions satisfying our assumptions, the set of solutions will be of dimension zero in \( \mathbb{R}_n^+ \). Therefore, the set of Lindahl outcomes is typically small, as is usually the case with sets of market equilibria.

**5.3 Example: A Global Public Good**

The purpose of the general theory is, of course, to study environments with arbitrary heterogeneity in marginal costs and benefits across pairs. But to get a feeling for the
mechanics of the characterization, it is helpful to consider a simple special case in which the public good is global.

Suppose each player’s preferences are separable in costs and benefits, and the benefits depend only on the sum of all effort levels:

\[ u_i(\mathbf{a}) = v_i \left( \sum_{j \in N} a_j \right) - k_i(a_i), \]

where each \( v_i \) is continuously differentiable, increasing, and concave, while \( k_i \) is continuously differentiable, increasing, and convex. To ensure that Assumption 1 holds, we assume \( v'(0) - k'(0) < 0 \). We can think of \( v_i \) as being the value of air quality and \( k_i \) as being the costs of foregone consumption.

To characterize the interior Lindahl outcomes, we use Theorem 2(ii). First, observe that:

\[ J_{ij}(\mathbf{a}) = \begin{cases} 
    v'_i \left( \sum_{h \in N} a_h \right) & \text{if } i \neq j \\
    -k'_i(a_i) & \text{if } i = j.
\end{cases} \]

Then the condition \( \mathbf{J}(\mathbf{a}) \mathbf{a} = 0 \) can be rewritten:

\[ \frac{a_i}{\sum_{j \in N} a_j} = \frac{v'_i \left( \sum_{j \in N} a_j \right)}{k'_i(a_i)} \]  

(8)

Thus, each agent contributes a share\(^{28}\) of the total public good provision equal to his marginal rate of substitution between clean air and consumption.

**5.4 A Simple Inequality**

The formula of Theorem 2(iii) equates an agent’s contribution at a Lindahl outcome to quantities involving the marginal value of incoming benefits. This can be used to derive a general result on relative contributions: if, at every outcome, agent \( i \) receives marginal benefits at least as great as agent \( j \) from the effort of any other \( k \) (i.e. if \( B_{ik}(\mathbf{a}) \geq B_{jk}(\mathbf{a}) \) for all \( \mathbf{a} \in \mathbb{R}^n_+ \) and all \( k \)) then \( i \) always makes a weakly greater contribution than \( j \) at any Lindahl outcome \( \mathbf{a}^* \) (i.e., \( c_i(\mathbf{a}^*) \geq c_j(\mathbf{a}^*) \)). In the case where \( \frac{\partial u_i}{\partial a_i}(\mathbf{a}) = -1 \) for each \( \mathbf{a} \in \mathbb{R}^n_+ \) and \( i \in N \), the contribution is simply equal to the action, so agents who value all public goods more have higher action levels. This generalizes (8) above.

**6 Explicit Formulas for Lindahl Outcomes**

The general characterization of Lindahl outcomes in terms of network centrality makes no parametric assumptions on preferences. The cost of this generality is that the characterization is implicit: an eigenvector centrality action profile \( \mathbf{a} \) is defined with respect to a matrix, \( \mathbf{B}(\mathbf{a}) \), that depends on \( \mathbf{a} \) itself. To build intuition, it is useful to

\(^{28}\) The right-hand side is less than 1 because our assumptions entail that \( v'(0) - k'(0) < 0 \) and that \( v - k \) is a concave function.
study cases in which the Lindahl outcomes can be characterized explicitly in terms of exogenous parameters.

To this end, the present section examines a parametric family of preferences in which explicit formulas for the Lindahl outcomes are available. We study three different regimes within this parametric family, and in each of them there is a connection between Lindahl outcomes and an important centrality measure that has been studied independently in the networks literature. In the first regime, actions are equal to degree centralities. In the second, actions are equal to Bonacich centralities. And in the third regime, actions approximate eigenvector centralities. In each case, the centralities are defined relative to an exogenous network, and all the results are special cases of Theorem 2.

The propositions of this section may also be viewed as microfoundations for the network centrality measures in terms of price equilibria, because the Lindahl outcomes are defined in terms of prices (recall Definition 6 in Section 4.2). Each result described below says that for particular preferences, the “market” levels of public good provision defined by Lindahl are equal to centralities according to a particular measure. Of course, in our framework the Lindahl outcomes are motivated primarily by their implementation-theoretic foundations, but the connection to prices is also worth noting.

Substantively, the import of this section is that in each case, what matters for the level of i’s public good provision at a Lindahl outcome is the extent to which i is the recipient of benefits. In the simplest case – Section 6.2 – i’s effort level is simply the sum of coefficients that determine how much each of his neighbors can help him. In Sections 6.3 and 6.4, what matters is not only links carrying benefits from other agents to i, but chains or walks of favor-giving links, involving multiple agents, that terminate at i.

6.1 A Parametric Family of Preferences and an Explicit Formula for Lindahl Outcomes

Let G and H be nonnegative matrices with zeros on the diagonal and r(G) < 1. For each i ∈ N let

\[ u_i(a) = -a_i + \sum_j [G_{ij}a_j + H_{ij}\log a_j]. \]  

Then c(a) = a and we compute that the benefits matrix B(a) is given by:

\[ B_{ij}(a) = \begin{cases} G_{ij} + \frac{H_{ij}}{a_j} & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases} \]  

---

29 We will introduce each measure without assuming background in network topics.
30 These should be viewed as functions \( u_i : \mathbb{R}_+^n \to \mathbb{R} \cup \{-\infty\} \), with \( 0 \cdot \log 0 \) understood as 0. In other words, preferences should be completed by continuity to the extended range. No result in the paper is affected by this slight departure from the framework of Section 2.
Defining $h_i = \sum_j H_{ij}$, the condition for $a$ to be an eigenvector centrality action profile – i.e., $c(a) = B(a)c(a)$ – boils down to

$$ a = h + Ga $$

or

$$ a = (I - G)^{-1}h. \tag{10} $$

The vector $a$ is well-defined and nonnegative\(^{31}\) by the assumption that $r(G) < 1$. This formula will yield all of the results in this section.

### 6.2 Degree Centrality

The most basic measure of connectedness is the total weight of links that a node has.

**Definition 9.** For a nonnegative matrix $M$ the degree centrality in $M$ is defined by the following equation for each $i$:

$$ \delta_i(M) = \sum_j M_{ij} $$

Suppose $G$ is the zero matrix and $H = M$ for a matrix $M$ with zeros on the diagonal. Then it follows immediately from (10) that, under the preferences of (9) with these parameter values, we have

$$ a = \delta_i(M). $$

In other words, when costs are linear in one’s own action and benefits are logarithmic in others’ actions, then an agent $i$’s contribution is determined by how much he benefits from everyone else’s effort at the margin: the sum of coefficients $H_{ij}$ as $j$ ranges across the other agents. The agents who are particularly dependent on the rest are the ones who are contributing the most. Note that how much others depend on $i$ does not figure at all in this formula. If $i$’s help suddenly becomes much more valuable to others, then his action does not change, though his share of the overall contribution goes down (because others’ action levels go up).

Thus, even the simplest case already yields some stark comparative statics in levels of provision.

### 6.3 Bonacich Centrality

The second centrality measure we consider is due to Bonacich (1987).\(^{32}\)

**Definition 10.** For a nonnegative matrix $M$ and a constant $\alpha < 1/\rho(M)$, the Bonacich centrality of parameter $\alpha$ in $M$ is defined by

$$ \beta(M, \alpha) = [I - \alpha M]^{-1} 1, $$

\(^{31}\)See Ballester, Calvó-Armengol, and Zenou (2006, Section 3).

\(^{32}\)An important antecedent was discussed by Katz (1953).
where \(1\) is the column vector of ones.

Dropping the arguments, the defining equation says that for every \(i\), we have:

\[
\beta_i = 1 + \alpha \sum_j M_{ij} \beta_j.
\]

Thus, every node gets a baseline level of centrality (one unit) and then additional centrality in proportion to the centrality of those it is linked to.

Now, returning to the preferences defined by equation (9), let \(G = \alpha M\) for any \(M\) with zeros on the diagonal and \(\alpha < 1/r(M)\). If we fix any \(H\) such that each row sums to 1, then (10) implies

\[
a = \beta(M, \alpha).
\]

**An Interpretation via Walks**

To explain the meaning of this result, we make some definitions reminiscent of Section 3.4.2, which discussed cycles.

A walk of length \(\ell\) in the matrix \(M\) is a sequence \((w(1), w(2), \ldots, w(\ell))\) of elements of \(N\) (player indices) such that \(M_{w(t)w(t+1)} > 0\) for each \(t \in \{1, 2, \ldots, \ell - 1\}\).\(^{33}\) Let \(W_i(\ell; M)\) be the set of all walks of length \(\ell\) in \(M\) such that \(w(1) = i\). For a matrix \(M\), define the value of a walk \(w\) of length \(\ell\) exactly as in equation (4) of Section 3.4.2. It is the number obtained by taking the product of all weights along the walk \(w\):

\[
v(w; M) = \prod_{t=1}^{\ell-1} M_{w(t)w(t+1)}.
\]

Define

\[
V_i(\ell; M) = \sum_{w \in W_i(\ell; M)} v(w; M).
\]

This is the sum of the values of all walks of length \(\ell\) starting at \(i\). Then we have:

**Fact 4.** \(\beta_i(M, \alpha) = 1 + \sum_{\ell=1}^{\infty} \alpha^\ell V_i(\ell; M).\)

Fact 4 is established, e.g., in Ballester, Calvó-Armengol, and Zenou (2006, Section 3). Thus, the Bonacich centrality is equal to one plus a weighted sum of values of all walks, with longer walks downweighted exponentially.

Walks in \(G\) emanating from \(i\) correspond to chains of dependence: \(i\) depends on the help of \(j\) who depends on the help of \(k\), etc. The weight of such a walk is the product along the chain of the strengths of the dependence relationships. According to equation (11) and Fact 4, a player at a Lindahl outcome in this environment contributes in proportion to the total strength of such dependence chains that he initiates. Another way to say it (recalling the discussion about the directions of arrows in Section 3.4.2) is that \(i\) contributes in proportion to the total strength of chains of benefit flow terminating at \(i\). In contrast to the case of degree centrality treated in the previous section, it is not only \(i\)'s immediate neighborhood that matters, but also how much \(i\)'s neighbors benefit from their neighbors, etc.

\(^{33}\)As with cycles, defined in Section 3.4.2, nodes can be repeated in this sequence. Note also that a cycles of length \(\ell\) is a walk of length \(\ell\).
6.4 Eigenvector Centrality

Eigenvector centrality was defined in the introduction, and is a key notion throughout the paper. Indeed, one may ask why this section is necessary – after all, doesn’t Theorem 2 establish a general connection between eigenvector centrality and Lindahl outcomes?

The answer is that Theorem 2 characterizes \( \mathbf{a} \) through an \textit{endogenous} eigenvector centrality condition – a condition that depends on \( \mathbf{B}(\mathbf{a}) \). In this section, we study the special case in which action levels approximate eigenvector centralities defined according to an \textit{exogenous} network.

Fix any nonnegative, irreducible matrix \( \mathbf{M} \). By the Perron-Frobenius Theorem, \( \mathbf{M} \) has a unique eigenvector centrality vector \( \mathbf{e} \), with entries summing to 1, so that \( \mathbf{e} = \gamma \cdot \mathbf{M} \mathbf{e} \), where \( \gamma = 1/r(\mathbf{M}) \).

Take any matrix \( \mathbf{H} \) such that each row sums to 1. For any \( z < 1 \), let \( \mathbf{G} = \frac{z}{r(\mathbf{M})} \mathbf{M} \).

By the result just established in Section 6.3,

\[
\mathbf{a} = \beta \left( \mathbf{M}, \frac{z}{r(\mathbf{M})} \right).
\]

The key fact that allows us to connect outcomes to \textit{eigenvector} centrality as opposed to Bonacich centrality is this: as \( z \uparrow 1 \), it holds that \( a_i/a_j \to e_i/e_j \) for every \( i, j \). That is, each player’s share of the total of all actions converges to his eigenvector centrality according to \( \mathbf{M} \). For a proof, see Theorem 3 of Golub and Lever (2010).

Because eigenvector centrality arises as a limit of Bonacich centrality, the walks interpretation above is relevant: indeed, the eigenvector centrality of \( i \) can be seen as the fraction of the very long chains of benefits that terminate at \( i \) (as opposed to some other player).

7 Tying up the Theory

This section focuses on the proofs of the two most important results – Theorem 1 and Theorem 2 – as well as on the existence of Lindahl outcomes.

First, Section 7.1 states the existence of outcomes satisfying the three equivalent conditions of Theorem 2, and in particular the existence of Lindahl outcomes. Section 7.2 uses Theorem 2 to prove Theorem 1 (the theorem on what makes Lindahl outcomes special from the implementation perspective). Section 7.3 sketches the proof of our main result, Theorem 2; the full proof appears in the appendix.

7.1 Existence of Lindahl Outcomes

For completeness, we state a result on the existence of Lindahl outcomes in our setting.\textsuperscript{34}

\textsuperscript{34}See Foley (1970) for a general existence proof. Note, however, that adapting Foley’s environment to ours – in which, for example, there are no \textit{ex ante} upper bounds on public good provision – is not entirely trivial.
**Proposition 5.** If \( u \) satisfies the assumptions of Section 2.2, then \( L(u) \) is nonempty, i.e., there is a Lindahl outcome for this preference profile. Moreover, either \( 0 \in L(u) \) or there is a Lindahl outcome in which all actions are strictly positive.

The proof appears in Section A.7, and works by directly showing the existence of scaling-indifferent outcomes; existence of Lindahl outcomes then follows by Theorem 2. The existence of scaling-indifferent outcomes may be of some independent interest, and is established without reference to prices, since the condition \( J(a)a = 0 \) does not involve prices.

### 7.2 Proof of Theorem 1 on the Hurwicz Rationale for Lindahl Outcomes

In this section, we give a proof of Theorem 1 in Section 4.3, which asserts that the only robustly attainable outcomes are the Lindahl outcomes.

To do this, we briefly recall Maskin’s Theorem. Assuming that the number of agents \( n \) is at least 3 and that a social choice correspondence \( F \) satisfies no veto power\(^{35} \) (a condition that is vacuously satisfied in our setting), then \( F \) is implementable if and only if it satisfies Maskin monotonicity.

**Definition 11.** A social choice correspondence \( F : U \Rightarrow \mathbb{R}^n_+ \) satisfies Maskin monotonicity if, for all \( a^* \in \mathbb{R}^n_+ \) and all \( \hat{u}, u \in U \), if \( a^* \in F(\hat{u}) \) and

\[
\forall i \in N, \forall a \in \mathbb{R}^n_+, \quad a^* \succeq_{\hat{u}} a \Rightarrow a^* \succeq_{u} a
\]

then \( a^* \in F(u) \).

**Proof of Theorem 1:** First we show that\(^{37} \) \( R \subseteq L \) (from which it follows immediately that \( R^{\text{int}} \subseteq L^{\text{int}} \)). By definition of \( R \), it suffices to show that \( L \subseteq N \); i.e. that \( L \) is an individually rational, Pareto efficient, and upper hemicontinuous social choice correspondence. First, a social choice correspondence must be nonempty-valued; Proposition 5 in Section 7.1 guarantees that \( L \) complies. By Assumption 3, the no veto power condition is vacuous in our setting. It is verified immediately from Definition 6 that \( L \) satisfies Maskin monotonicity. Thus, \( L \) is implementable by Maskin’s Theorem. Also, \( L \) is individually rational since, by definition of a Lindahl outcome, each agent prefers a Lindahl outcome to \( 0 \), which is always feasible. By the standard proof of the First Welfare Theorem, \( L \) is Pareto efficient (see, e.g., Foley, 1970). Similarly, the standard argument for the upper hemicontinuity of equilibria in preferences transfers to our setting.

Now assume \( F \) is implementable, Pareto efficient, individually rational, and upper hemicontinuous. Fix \( u \in U \) and an interior \( a^* \in L(u) \). We will show \( a^* \in F(u) \).

\(^{35}\)A social choice correspondence \( F : U \Rightarrow \mathbb{R}^n_+ \) satisfies no veto power if, for every \( u \), whenever there is an \( a \) such that \( a \succeq_{u} a' \) for all \( i \neq i' \) and all \( a' \in \mathbb{R}^n_+ \), then \( a \in F(u) \).

\(^{36}\)In words, if an alternative \( a \) was selected by \( F \) under \( u \) and then we change those preferences so that \( a \), in each agent’s preference, defeats all the same alternatives as under \( u \) and perhaps some others, then \( a \) is still selected under \( \hat{u} \).

\(^{37}\)For two correspondences \( F, F^\dagger : U \rightarrow \mathbb{R}^n_+ \), we write \( F \subseteq F^\dagger \) if for every \( u \in U \), it holds that \( F(u) \subseteq F^\dagger(u) \). In this case, we say that \( F \) is a sub-correspondence of \( F^\dagger \).
Define
\[ \hat{u}(a) = J(a^*; u)a. \]

Lemma 1 in Section A.3 states that since \( F \) is individually rational, Pareto efficient, and upper hemicontinuous, it follows that \( a^* \in F(\hat{u}). \)\(^{38}\) Note that for all \( a \in R^n_+ \), we have
\[ \hat{u}(a^*) - \hat{u}(a) = J(a^*; u)(a^* - a) \leq u(a^*) - u(a) \]
by concavity of \( u \), so (12) holds. Since \( F \) is implementable, it satisfies Maskin monotonicity, so we conclude that \( a^* \in F(u) \).

### 7.3 Centrality Characterization of Lindahl Outcomes: Proof Outline

It will be convenient to use in this section an equivalent definition of Lindahl outcomes:

**Definition 12.** An action profile \( a^* \) is a *Lindahl outcome* for a preference profile \( u \) if there exists an \( n \)-by-\( n \) matrix \( P \) with each column summing to zero so that the following conditions hold for every \( i \):

(i) The inequality
\[ \sum_{j \in N} P_{ij} a_j \leq 0 \quad (\hat{BB}_i(P)) \]
is satisfied when \( a = a^* \);

(ii) for any \( a \) such that \( \hat{BB}_i(P) \) is satisfied, we have \( a^* \succeq_u a \).

Given a Lindahl outcome defined as in Definition 6, set \( P_{ii} = -\sum_{j:j \neq i} P_{ij} \) to find prices satisfying the new definition.\(^{39}\) Conversely, the prices of Definition 12 work in Definition 6 without modification, since the original definition does not involve the diagonal terms of \( P \) at all.

The key to Theorem 2 is that the system of equations \( J(a^*)a^* = 0 \) contains within it, secretly, all the conditions defining a Lindahl outcome. It allows us to extract Pareto weights that support the outcome \( a^* \) as efficient, and using those Pareto weights and the Jacobian, we can construct prices that support \( a^* \) as a Lindahl outcome.

Now in more detail: suppose \( J(a^*)a^* = 0 \). First, because this equation implies\(^{40}\) \( B(a^*)c(a^*) = c(a^*) \), the benefits matrix \( B(a^*) \) has a spectral radius of 1 and thus, by Proposition 1, the profile \( a^* \) is a Pareto efficient outcome. So there are Pareto weights \( \theta \) such that \( a^* \) solves \( P(\theta) \). That implies that the first-order conditions of

---

\(^{38}\)The proof of that lemma constructs a sequence of preference profiles \((\hat{u}^{(k)})\) converging to \( \hat{u} \) such that individual rationality and Pareto efficiency alone force the set \( F(\hat{u}^{(k)}) \) to converge to \( a^* \). Then by upper hemicontinuity of \( F \), it follows that \( F(\hat{u}) \) contains \( a^* \).

\(^{39}\)In essence, \(-P_{ii}\) is the wage of agent \( i \), equal to the sum of taxes paid for his effort.

\(^{40}\)The equivalence of (ii) and (iii) in Theorem 2 comes about via a simple manipulation (see Lemma 2 in Section A.4).
the Pareto problem \( \mathcal{P}(\theta) \) defined in the proof of Proposition 1 are satisfied: for each \( j \), we have

\[
\sum_{i \in N} \theta_j J_{ij}(a^*) = 0.
\]

If we set \( P_{ij} = \theta_j J_{ij}(a^*) \), then that implies \( \sum_i P_{ij} = 0 \) for each \( j \). Thus, the price matrix is a legitimate candidate for Definition 12. Moreover, the budget balance condition \( \hat{B}_i(P) \) is satisfied at these prices, since

\[
\sum_{j \in N} P_{ij} a_j = \theta_i \sum_{j \in N} J_{ij}(a^*) a_j^* = 0.
\]

The first equality is by construction of the prices, and the second is by the assumption that \( J(a^*) a^* = 0 \). Thus, each agent is exhausting his budget. It only remains to see that each agent is optimizing at prices \( P \). The essential reason for this is that price ratios are equal to marginal rates of substitution by construction:

\[
\frac{P_{ij}}{P_{ik}} = \frac{\theta_i J_{ij}(a^*)}{\theta_i J_{ik}(a^*)} = \frac{J_{ij}(a^*)}{J_{ik}(a^*)}.
\]

The choice of prices \( P_{ij} = \theta_j J_{ij}(a^*) \) is intuitive: \( P_{ij} \) is proportional to both how much \( i \) benefits from \( j \)'s help at the margin, as well as to how much the social planner values \( j \)'s welfare. The more the social planner cares about \( j \), the more expensive it is to buy his effort.

8 Concluding Discussion

8.1 Free-Riding

Throughout, we have studied the implementation problem under the requirements of Pareto efficiency and individual rationality. The latter is defined in terms of leaving everyone no worse off than at the status quo. In many mechanism design and implementation problems, the individual rationality requirement is motivated by the notion that participation is voluntary. For example, if a player may end up worse off after an auction than before he entered, then he may not wish to participate, paying nothing and having no chance of winning the item. The value of this outcome can be normalized to 0 because it is independent of what happens in the mechanism without the agent.

In our context, the notion of opting out is more complicated. Since the goods in question are public goods, a player who “opts out” of a mechanism to provide them often cannot be excluded from enjoying the benefits of an arrangement reached in his absence – for example, one that leads to cleaner air or water. Thus, a player may wish to “opt out” in hopes of enjoying public goods provided by others without investing any effort himself.

In the analysis so far, we have implicitly assumed that players can commit to participate as part of their general power to commit to the rules of an institution –
the function $g$ in the language of Section 4.1. To explore the extent to which this is a binding assumption, we enrich the environment, giving each player an inalienable decision over whether to participate – one that he cannot make commitments about or delegate.\footnote{See Kartik and Tercieux (2012) for a discussion of inalienability in the context of implementation with evidence.} If an agent chooses to participate, he can commit to the rules of the subsequent game, as in Section 4.1.

Formally, we assume that the designer of Section 4.1 announces a family of game forms, $(H_P)_{P \subseteq N}$, one for every possible set $P$ of participants. Players first simultaneously and unilaterally choose whether to participate, and the set $P$ of those who chose to participate play the game $H_P$, just as in Section 4.1.

The idea of this section is that we can provide a sufficient condition under which no player wishes to use his power to opt out. Recall Section 3.4.4, in which we showed that player $i$'s participation is essential to achieving any Pareto improvement on the status quo precisely when his removal changes the spectral radius of the benefits matrix at the status quo from being greater than 1 to being less than 1. If each player has this property, then his decision not to participate leaves the remaining players with no Pareto improvement on the status quo. Thus, assuming each $H_P$ always implements Pareto efficient outcomes that are at least as good as 0 for each of the remaining participants, the only outcome a player can hope for after opting out is 0.

In short, we have argued that if every player is essential, then each player expects that opting out will lead to the outcome 0. Therefore, individual rationality as we defined it in our main analysis is equivalent to voluntary participation.

There is a simple matrix characterization of what it means for every player to be essential: the spectral radius of $B(0)$ exceeds 1, and the spectral radius of every submatrix obtained by removing one row and one column is less than 1.

This argument yields only a crude sufficient condition and suggests avenues for future work, of which we mention one. What if opting out decisions are made in equilibrium? An interesting framework is one in which they are sequential and players behave farsightedly, anticipating the effects of their opting out on others’ incentives to participate in providing the public goods. How can we compute, given any $B(0)$, the largest stable coalition not at risk of being unraveled by rational sequences of decisions to opt out? (See Genicot and Ray [2003] for a related analysis in a risk-sharing game.)

### 8.2 Lindahl Outcomes and Coalitional Deviations: The Core

Aumann (1959) suggested a way to define the core in games with externalities. He considered an outcome to be outside the core if a coalition could, by choosing a different action profile, improve the welfare of every one of its members no matter what the others did. Recall that in our setting, externalities across players are always nonnegative by Assumption 2, so in this case the uniformly worst punishment by the outsiders is to play 0. The outcomes that do not admit such a deviation are the core outcomes.

**Definition 13.** An action profile $a$ is in the $\beta$-core under a preference profile $u$ if
there is no nonempty coalition \( M \subseteq N \) and no other action profile \( a' \) of this game so that:

(i) \( a'_i = 0 \) for all \( i \notin M \);

(ii) for each \( i \in M \) we have \( a' \succ_i a \).

Under nonnegative externalities, the Lindahl outcomes are in the \( \beta \)-core (Shapley and Shubik, 1969; Foley, 1970).

Leaving no coalition worse off than they could be without outsiders’ help may be a desirable normative property. It is also important, intuitively speaking, for making sure that coalitions do not have incentives to secede. However, developing a full non-cooperative theory of play under the possibility of coalitional deviations is a very challenging and important direction for future work. See Bernheim, Peleg, and Whinston (1987) and Horniaček (1996) for some partial progress.

8.3 The Informational Assumptions of the Implementation Theory Framework

In the setup of Section 4.1, we follow the standard approach of complete information implementation in focusing on the Nash equilibria of the game form \( H \) given agents’ true preferences. This implicitly assumes that agents know enough about each other’s strategies to play a Nash equilibrium of the game\(^{42} \). On the other hand, the institutions in which they interact – the rules of the game – are long-run, and are designed to be robust to many realizations of preferences.

We do not view the assumption that agents play Nash equilibria of the complete information game as very restrictive or inappropriate when the negotiators represent municipalities or countries.\(^{43} \) In those cases, decisions about strategy are made by institutions (e.g. governments) whose agents communicate internally, and consequently the intentions of each government are difficult to conceal from counterparties. Relatedly, the parties often base their valuations on a more or less common pool of information. For instance, when the public good is environmental, then estimates of the economic sacrifices needed for abatement and the harms from pollution – which jointly determine the payoffs – are made based on information and models regarding the environment and the economy that are essentially public\(^{44} \).

Nevertheless, a setting with some incomplete information is certainly more realistic than the stark complete information benchmark, and further study of it is a direction for future work.\(^ {45} \)

\(^{42} \) Viewed as elements of the strategy spaces \( \Sigma_i \), not as maps from privately known types to such strategies.

\(^{43} \) Roberts (2008) briefly critiques the informational assumptions of complete information Nash implementation when the agents in question are many citizens in a public finance setting.

\(^{44} \) We thank Joe Shapiro for suggesting this point. Of course, the information and the models may themselves involve a great deal of uncertainty. This alone need not create a problem for complete information implementation because it is uncertainty common to the agents, and therefore does not create informational asymmetries – as long as the agents know the preferences rationalizing counterparties’ decisions. Implicitly, our analysis is carried out conditioning on a realization of this uncertainty.

\(^{45} \) See Tian (1996) for a study of implementing a rational expectations Lindahl equilibrium and
References


Jackson (2001) for a survey of incomplete-information implementation theory.


A Omitted Proofs and Additional Results

A.1 Proof of the Bang for the Buck Proposition

This section proves Proposition 3 and also gives a formal statement and proof of the Rawlsian interpretation of an egalitarian direction: that it maximizes the minimum return on investment.

**Proof of Proposition 3 in Section 3.3:** Fix $a$ and denote by $r$ the spectral radius of $B(a)$. Since $B(a)$ is nonnegative and irreducible, the Perron-Frobenius Theorem guarantees that it has a right-hand eigenvector $v$ such that

$$B(a)v = rv. \quad (13)$$

Define $d_{eg}(a) = -v_i / J_{ii}(a)$ for each $i$. Then algebraic manipulation shows (13) is equivalent to $b(a, d_{eg}(a)) = r1$, where 1 is the column vector of all ones. This shows there is an egalitarian improvement that generates a bang-for-the-buck of $r$ (the spectral radius of $B(a)$) for everyone.

Now suppose $d \in \Delta_n$ is any egalitarian direction, i.e. for some $b$ we have

$$b(a, d) = b1. \quad (14)$$

Define $w$ by $w_i = -J_{ii}(a)d_i$ for each $i$, and note that then (14) implies

$$B(a)w = bw. \quad (15)$$

By the Perron-Frobenius Theorem (statement (iii)), the only real number $b$ and vector $w \in \Delta_n$ satisfying (15) are $b = r$ and $w = v$, which implies $d = d_{eg}(a)$. Proposition 6. The egalitarian direction $d_{eg}(a)$ maximizes $\min_i b_i(a, d)$ among all $d \in \Delta_n$.

**Proof of Proposition 6:** Fix $a$ and, for any $d$, define $w(d)$ by $w_i(d) = -J_{ii}(a)d_i$. Then $b_i(a, d) = [B(a)w(d)]_i / w_i(d_i)$. Now we claim that the $r(B(a))$ is the maximum of

$$\min_i \frac{[B(a)v]_i}{v_i}$$

as $v$ ranges over nonzero, nonnegative vectors. Indeed, this is precisely what the Collatz-Wielandt formula (Meyer, 2000, equation 8.3.3) says. Thus, the minimum bang for the buck is never greater than $r(B(a))$. We saw in the proof of Proposition 3 above that the bang for the buck in the egalitarian direction $d_{eg}(a)$ is equal to $r(B(a))$ for everyone. This completes the proof.

A.2 The Zero Action Profile as an Efficient Outcome

**Proof of Proposition 2 in Section 3.1:** First we prove the result in the case when $B(0)$ is irreducible.

If $r(B(0)) > 1$, then Proposition 3 yields an egalitarian direction at 0 with bang for the buck exceeding 1; this is a Pareto improvement at 0.
If 0 is not Pareto efficient, there is an \( a' \in \mathbb{R}_+ \) such that \( u_i(a') \geq u_i(0) \) for each \( i \), with strict inequality for some \( i \). Using Assumption 3, namely the irreducibility of \( B(a') \), as well as the continuity of the \( u_i \), we can find an \( a'' \) with all positive entries so that \( u_i(a'') > u_i(0) \) for all \( i \). Let \( v \) denote the derivative of \( u(\zeta a'') \) in \( \zeta \) evaluated at \( \zeta = 0 \). This derivative is strictly positive in every entry, since (by convexity of the \( u_i \)) the entry \( v_i \) must exceed \( [u_i(a'') - u_i(0)]/a'' \). By the chain rule, \( v = J(0)a'' \). From the fact that \( v \) is positive, it is immediate to deduce that there is a positive vector \( w \) so that \( B(0)w > w \). And from this it follows by the Collatz-Wielandt formula (Meyer, 2000, equation 8.3.3) that the spectral radius of \( B(0) \) exceeds 1.

Now assume \( B(0) \) is reducible.

If \( r(B(0)) > 1 \), then the same is true when \( B(0) \) is replaced by one of its irreducible blocks, and in that case a Pareto improvement on 0 (involving only the agents in the irreducible block taking positive effort) is found as above.

Now suppose 0 is not Pareto efficient. There is an \( a' \in \mathbb{R}_+ \) such that \( u_i(a') \geq u_i(0) \) for each \( i \), with strict inequality for some \( i \). Let \( P = \{ i : a'_i > 0 \} \) be the agents taking positive actions at \( a' \). And let \( \hat{B}(0) \) be obtained by restricting \( B(0) \) to \( P \) (i.e. by throwing away rows and columns not corresponding to indices in \( P \)). For each \( i \in P \), there is a \( j \in P \) such that \( \hat{B}_{ij}(0) > 0 \); otherwise, then \( i \) would be worse off than at 0. Therefore, each \( i \in P \) is on a cycle in \( \hat{B}(0) \). And it follows that for each \( i \in P \) there is a set \( P_i \subseteq P \) such that \( \hat{B}(0) \) is irreducible when restricted to \( P_i \). Next, applying the argument of footnote 46 to each such \( P_i \) separately, we can find \( a'' \) such that \( u_i(a'') > u_i(0) \) for each \( i \in P \). From this point we can argue as above to conclude that \( r(\hat{B}(0)) > 1 \). Since \( \hat{B}(0) \) is a submatrix of \( B(0) \), Fact 3 on the monotonicity of the spectral norm implies that \( r(B(0)) > 1 \).

**A.3 Details Relating to Theorem 1**

As noted in Section 4.3, the Hurwicz rationale for the Lindahl outcomes is more general than is stated there. In this section, we formalize this. Let \( \mathcal{A} \) be the set of preference profiles \( u \) satisfying the assumptions of Section 2.2.

Fix any probability measure \( \mu \) on \( \mathbb{R}_+^n \) that is absolutely continuous with respect to Lebesgue measure – in particular, one assigning positive measure to every open set. We define a metric \( d \) on \( \mathcal{A} \) by \( d(u, \hat{u}) = \int_{\mathbb{R}_+^n} |u - \hat{u}| d\mu \). That metric induces a topology on the function space \( \mathcal{A} \).

**Definition 14.** A set of preferences \( \mathcal{U} \subseteq \mathcal{A} \) is called rich if, for every \( u \in \mathcal{U} \) and

---

46 Suppose otherwise and let \( a'' \) be chosen so that \( u(a'') - u(0) \geq 0 \) (note this is possible, since \( a'' = a' \) satisfies this inequality) and so that the number of 0 entries in \( u(a'') - u(0) \) is as small as possible. Let \( S \) be the set of \( i \) for which \( u_i(a'') - u_i(0) > 0 \). Then by irreducibility of benefits, we can find \( j \in S \) and \( k \notin S \) such that \( J_{kj}(0) > 0 \). Define \( a'''_j = a''_j + \epsilon \) and \( a'''_i = a''_i \) for all \( i \neq j \). If \( \epsilon > 0 \) is chosen small enough, then by continuity of the \( u_i \) we have \( u_i(a''') - u_i(0) > 0 \) for all \( i \in S \), but also \( u_k(a''') - u_k(0) > 0 \), contradicting the choice of \( a''' \).

47 Recall the definition in Section 3.4.2.

48 The Collatz-Wielandt formula does not assume irreducibility.
there is a preference profile \( \hat{\mathbf{u}} \in \mathcal{U} \) defined by

\[
\hat{\mathbf{u}}(\mathbf{a}) = J(\mathbf{a}^*; \mathbf{u})\mathbf{a}
\]  

(16)

and a neighborhood of \( \hat{\mathbf{u}} \) relative to \( \mathcal{A} \) is contained in \( \mathcal{U} \).

Richness of \( \mathcal{U} \) requires that for every preference profile \( \mathbf{u} \in \mathcal{U} \) and every \( \mathbf{a}^* \in \mathbb{R}^n_+ \), there are preferences in \( \mathcal{U} \) that are linear over outcomes and have the same *marginal* tradeoffs as \( \mathbf{u} \) does at \( \mathbf{a}^* \), as well as a neighborhood of these preferences. To take a simple example, \( \mathcal{A} \) itself is rich.

**Theorem 3.** Suppose \( \mathcal{U} \) is rich and the number of players, \( n \), is at least 3. Then the interior part of the robustly attainable correspondence is equal to that of the Lindahl correspondence: \( R^{\text{int}} = L^{\text{int}} \).

The proof is exactly as in Section 7.2. The only thing that remains to do is to establish the lemma used in that proof under the hypothesis that \( \mathcal{U} \) is rich.

**Lemma 1.** Fix \( \mathbf{u} \) satisfying the assumptions of Section 2.2 and an interior \( \mathbf{a}^* \in L(\mathbf{u}) \). Define \( \hat{\mathbf{u}} \) as in (16). Suppose \( F : \mathcal{U} \Rightarrow \mathbb{R}^n_+ \) is a Pareto efficient, individually rational, and upper hemicontinuous social choice correspondence. If \( \mathcal{U} \) is rich, then \( \mathbf{a}^* \in F(\hat{\mathbf{u}}) \).

**Proof of Lemma 1:** For \( \gamma > 0 \), and \( i \in N \), define \( \hat{u}_i^{[\gamma]} : \mathbb{R}^n_+ \rightarrow \mathbb{R} \) by

\[
\hat{u}_i^{[\gamma]}(\mathbf{a}) = J_{ii}(\mathbf{a}^*; \mathbf{u}) \left( \frac{\gamma + a_i}{a_i^*} \right)^{1+\gamma} + \sum_{j \neq i} J_{ij}(\mathbf{a}^*; \mathbf{u})a_j.
\]

This is just an adjustment obtained from \( \hat{\mathbf{u}} = \hat{\mathbf{u}}^{[0]} \) by building some convexity into the costs. Note that for all \( \gamma \) close enough to 0, the profile \( \hat{\mathbf{u}} \) is in \( \mathcal{U} \) by the richness assumption.

Choose \( \mathbf{a}^{(k)} \in F(\hat{\mathbf{u}}^{[1/k]}) \); this is legitimate since \( F \) is a social choice correspondence, and hence nonempty-valued. We will show that by the properties of \( F \), a subsequence of the sequence \( (\mathbf{a}^{(k)})_k \) converges to \( \mathbf{a}^* \). Then by upper hemicontinuity of \( F \), it will follow that \( \mathbf{a}^* \in F(\hat{\mathbf{u}}^{[0]}) \), as desired. The trickiest part of the argument is showing that the \( \mathbf{a}^{(k)} \) lie in some compact set, so we can extract a convergent subsequence; it will then be fairly easy to show that the limit point of this sequence is \( \mathbf{a}^* \).

Let \( \text{IR}^{[\gamma]} \) be the set of individually rational points under \( \hat{\mathbf{u}}^{[\gamma]} \) and let \( \text{PE}^{[\gamma]} \) be the set of Pareto efficient points under \( \hat{\mathbf{u}}^{[\gamma]} \). Let \( a_{\text{max}}^* = \max_i a_i^* \) and define the box \( K = [0, 2a_{\text{max}}^*]^n \).

**Claim 1.** For all \( k \), the point \( \mathbf{a}^{(k)} \) is either in \( K \) or on the ray

\[
Z = \{ \mathbf{a} \in \mathbb{R}^n_+ : J(\mathbf{a}^*; \mathbf{u})\mathbf{a} = 0 \}.
\]

---

49Of course, it suffices that there be some \( \hat{\mathbf{u}} \in \mathcal{U} \) that generates the same preferences for each agent.
To show the claim, we first establish that

$$\text{IR}^0 = Z.$$  

The proof is as follows. First note that $$\hat{u}^0(a) = J(a^*; u)a.$$ There cannot be an a such that $$J(a^*; u)a$$ is nonnegative in all entries and positive in some entries.\(^{50}\) Thus, if $$J(a^*; u)a$$ is nonzero, it must have some negative entries; i.e. $$\hat{u}_i^0(a) < 0$$ for some $$i$$, and then $$a \notin \text{IR}^0,$$ contradicting the fact that $$F$$ is individually rational.

Note that for $$a$$ outside the box $$K,$$ we have for small enough $$\gamma$$

$$\hat{u}^\gamma(a) \leq \hat{u}^0(a).$$

From this and the fact that $$\hat{u}^\gamma(0) = 0$$ for all $$\gamma,$$ we have the relation $$\text{IR}^\gamma \cap K^c \subseteq \text{IR}^0 \cap K^c.$$ The claim follows.

It is easily checked\(^{51}\) that if $$a \in Z$$ and $$a > a^*,$$ then for $$\gamma > 0$$ we have $$r(B(a; \hat{u}^\gamma)) < r(B(a^*; u)) = 1,$$ where the latter equality holds by Remark 5. Therefore, by Proposition 1, no point on the ray $$Z$$ outside $$K$$ is Pareto efficient. This combined with Claim 1 shows that $$\text{IR}^\gamma \cap \text{PE}^\gamma \subseteq K,$$ and therefore (since $$F$$ is Pareto efficient and individually rational) it follows that $$a^{(k)} \in K$$ for all $$k.$$

As a result we can find a sequence $$(j(k))_k$$ such that the sequence $$(a^{(j(k))})_k$$ converges to some $$\bar{a}.$$ Define $$\bar{a}^{(k)} = a^{(j(k))}$$ and set $$\hat{u}^{(k)} = \hat{u}^{[1/j(k)]}.$$ Note that the $$\hat{u}^{(k)}$$ converge uniformly to $$\hat{u}^0$$ on $$K$$ and, indeed, on any compact set (thus, they converge compactly to $$\hat{u}^0$$). By upper hemicontinuity of $$F,$$ it follows that $$\bar{a} \in F(\hat{u}^0).$$ It remains only to show that $$\bar{a} = a^*,$$ which we now do.

If $$\bar{a} \notin Z,$$ then it is easy to see that for large enough $$k,$$ we would have $$\hat{u}^{(k)}(\bar{a}^{(k)}) < 0$$ for some $$i.$$ This would contradict the hypothesis that $$F$$ is individually rational. Thus, $$\bar{a}^{(k)} \to \zeta a^*$$ for some $$\zeta \geq 0.$$ If $$\zeta = 0$$ then inspection of the definition of $$\hat{u}^\gamma$$ shows that eventually $$\bar{a}^{(k)}$$ is not Pareto efficient for preferences $$u^{(k)}.$$ But that contradicts the Pareto efficiency of $$F.$$ So assume $$\zeta > 0.$$ In that case we would have:

$$J_{ij}(\bar{a}^{(k)}; u^{(k)}) \to \begin{cases} \zeta J_{ij}(a^*; u) & \text{if } j = i \\ J_{ij}(a^*; u) & \text{otherwise.} \end{cases}$$

Thus,

$$B(\bar{a}^{(k)}; u^{(k)}) \to \zeta B(a^*; u).$$

Recall from Remark 5 that $$B(a^*; u) = 1.$$ Since the spectral radius is linear in scaling the matrix and continuous (recall Remark 2), it follows that

$$r(B(\bar{a}^{(k)}; u^{(k)})) \to \zeta,$$

\(^{50}\)Otherwise, $$a^*$$ would not have been Pareto efficient under $$u;$$ moving in the direction $$a$$ would have yielded a Pareto improvement. But $$a^*$$ is Pareto efficient – see Remark 5.

\(^{51}\)We do a very similar calculation below.
By the Pareto efficiency of $F$, we know that $r(B(\bar{a}^{(k)}; u^{(k)})) = 1$ whenever $\bar{a}^{(k)}$ is interior\textsuperscript{52}, which holds for all large enough $k$ since $\zeta \neq 0$. Thus $\zeta = 1$. It follows that $\bar{a} = a^*$ and the proof is complete.

### A.4 Some Relationships Between the Jacobian and the Benefits Matrix

**Lemma 2.** Under the assumption in Section 2.2, for any $a \in \mathbb{R}_+^n$:

1. $\theta B(a) = \theta$ is equivalent to $\theta J(a) = 0$;
2. $B(a)c(a) = c(a)$ is equivalent to $J(a)a = 0$;
3. $B(a)c(a) = rc(a)$ for $r > 1$ implies $J(a)a > 0$; and
4. $B(a)c(a) = rc(a)$ for $r < 1$ implies $J(a)a < 0$.

**Proof of Lemma 2:** Let $D(a)$ be a diagonal matrix with $D_{ii}(a) = -J_{ii}(a)$. Recall that the diagonal entries are strictly positive by Assumption 1, so this matrix is invertible. Note that $B = JD^{-1} + I$. Multiplying on the left by $\theta$ gives (i).

Dropping the argument $a$ on all the matrices, for all $r > 0$:

\[
Bc = rc \\
\Leftrightarrow (B - I)c = (r - 1)c \\
\Leftrightarrow JD^{-1}Da = (r - 1)c \\
\Leftrightarrow Ja = (r - 1)c.
\]

Taking $r = 1$ shows (ii); taking $r > 1$ shows (iii); and taking $r < 1$ shows (iv).

### A.5 Proof of the Centrality Characterization of Lindahl Outcomes

**Proof of Theorem 2 in Section 5.2:** First, we show (i) implies (ii). Suppose $a^* \in \mathbb{R}_+^n$ is a Lindahl outcome. Let $P$ be the matrix of prices in the modified definition of a Lindahl outcome (Definition 12). Consider the following program for each of $i \in M$, denoted by $\Pi_i(P)$:

\[
\max u_i(a) \text{ subject to } a \in \mathbb{R}_+^n \text{ and } \widehat{BB}_i(P).
\]

By definition of a Lindahl outcome, $a^*$ solves $\Pi_i(P)$. By Assumption 3, which says there is always some $a_j$ whose increase makes $i$ better off, the budget balance constraint $\widehat{BB}_i(P)$ is satisfied with equality, so that $Pa^* = 0$. Because $a^*$ is interior, the gradient of the maximand $u_i$ must be orthogonal to the constraint set given by $\widehat{BB}_i(P)$. In other words, row $i$ of $J(a^*)$ is parallel to row $i$ of $P$. These facts together imply $J(a^*)a^* = 0$.

\textsuperscript{52}Recall Remark 5 again.
The equivalence of (ii) and (iii) is a simple computation and holds for any \( a \in \mathbb{R}_+^n \) – see Lemma 2 in Section A.4.

Finally, we claim that (iii) implies (i). Suppose \( B(a^*)c(a^*) = c(a^*) \). We will show \( a^* \) is a Lindahl outcome.\(^{53}\) The main challenge is to construct the matrix \( P \) in Definition 12. In the proof of the implication (i) \( \Rightarrow \) (ii), we saw that the Jacobian and \( P \) at a Lindahl outcome are intimately related – row \( i \) of \( P \) is some scalar \( \theta_i \) times row \( i \) of the Jacobian \( J(a) \). Inspired by this, we guess prices \( P_{ij} = \theta_i J_{ij}(a^*) \) for some scalars \( \theta \) and show that these prices work. The right scalars turn out to be given by the weights \( \theta \) such that \( a^* \) solves the Pareto problem \( P(\theta) \).

More formally: since \( a^* \) is nonzero and Assumption 1 ensures that \( J_{ii}(a^*) < 0 \) for each \( i \), we know that \( c(a^*) \) is a right eigenvector of \( B(a^*) \) with eigenvalue 1. By the Perron-Frobenius Theorem, there is then a row vector \( \theta \in \mathbb{R}_+^n \), i.e. one with only positive entries, so that

\[
\theta B(a^*) = \theta.
\]

From Lemma 2 in Section A.4 it follows that\(^{54}\) \( \theta J(a^*) = 0 \). Define the matrix \( P \) by \( P_{ij} = \theta_i J_{ij}(a^*) \) and note that for all \( j \in N \) we have

\[
\sum_{i \in N} P_{ij} = \sum_{i \in N} \theta_i J_{ij}(a^*) = [\theta J(a^*)]_j = 0,
\]

where \([\theta J(a^*)]_j\) refers to entry \( j \) of the vector \( \theta J(a^*) \).

Again by Lemma 2, the condition \( B(a^*)c(a^*) = c(a^*) \) implies \( J(a^*)a^* = 0 \) and each row of \( P \) is just a scaling of the corresponding row of \( J(a^*) \). We therefore have:

\[
Pa^* = 0.
\]

We claim that, for each \( i \), the vector \( a^* \) solves \( \Pi_i(P) \). This is because the gradient of \( u_i \) at \( a^* \), which is row \( i \) of \( J(a^*) \), is normal to the constraint set by construction of \( P \), and by (18) above, \( a^* \) satisfies the constraint \( BB_i(P) \). The claim then follows by the concavity of \( u_i \). \( \square \)

### A.6 The Zero Action Profile as a Lindahl Outcome

**Proposition 7.** The following are equivalent:

(i) \( r(B(0)) \leq 1 \);

(ii) \( 0 \) is a Pareto efficient action profile;

(iii) \( 0 \) is a Lindahl outcome.

**Proof of Proposition 7:** Proposition 2 establishes the equivalence between (i) and (ii).

\(^{53}\)If any entry of \( a^* \) is nonzero, then every entry is, because \( B(a) \) is irreducible by Assumption 3. So the implication (iii) \( \Rightarrow \) (i) requires only that \( a^* \) be a nonzero element of \( \mathbb{R}_+^n \).

\(^{54}\)To see that these are Pareto weights, recall the proof of Proposition 1.
(ii) $\Rightarrow$ (iii): The construction of prices is exactly analogous to the proof of Theorem 2 in Section A.5; the only difference is that rather than the Pareto weights, we use Pareto weights adjusted by the Lagrange multipliers on the binding constraints $a_i \geq 0$.

(iii) $\Rightarrow$ (ii): The standard proof of the First Welfare Theorem goes through without modification; see, e.g. Foley (1970).

A.7 Existence of a Lindahl Outcome

**Proof of Proposition 5 in Section 7.1:** We will use the Kakutani Fixed Point Theorem to find a Lindahl outcome. Define $Y = \{ a \in \mathbb{R}_+^n : \min_i [J(a) a_i] > 0 \}$, the set of action profiles $a$ at which everyone has positive gains from scaling. It is easy to check that $Y$ is convex.\(^{55}\) Also, $Y$ is bounded by Assumption 4. Thus, $\overline{Y}$, the closure of $Y$, is compact.

Define the correspondence $F : \overline{Y} \to \overline{Y}$ by

$$F(a) = \{ z a \in \overline{Y} : z \geq 0 \text{ and } \min_i [J(z a) a_i] \leq 0 \}. $$

Finally, define the correspondence $G : \overline{Y} \to \overline{Y}$ by

$$G(a) = F(d^{eg}(a)).$$

The function $d^{eg}$ is continuous (Wilkinson, 1965, pp. 66–67), and $F$ is clearly upper hemicontinuous, so it follows that $G$ is upper hemicontinuous. Finally, from the definitions of $Y$ and $F$ it follows that $F$ is nonempty-valued.\(^{56}\) Since $\overline{Y}$ is a compact and convex set, the Kakutani fixed point theorem implies that there is an $a \in \overline{Y}$ such that $a \in F(d^{eg}(a))$. Writing $\hat{a} = d^{eg}(a)$, this means that there is some $z \geq 0$ such that $\min_i [J(z \hat{a}) \hat{a}_i] \leq 0$. Let $a^* = z \hat{a}$. We will argue that $a^*$ is a Lindahl outcome.

Suppose $a^* \neq 0$. Then by continuity of the function $z \mapsto J(z \hat{a}) \hat{a}$, there is some $i$ for which we have $[J(a^*) \hat{a}]_i = 0$ – i.e., some player’s marginal benefit to scaling is equal to marginal cost. Since $\hat{a}$ is egalitarian and $\min_i [J(a^*) \hat{a}_i] = 0$, the same must be true for all players, and therefore $J(a^*) \hat{a} = 0$. Since $\hat{a}$ and $a^*$ are parallel, we deduce $J(a^*) a^* = 0$ and Theorem 2 shows that $a^*$ is a Lindahl outcome. Assumption 3 – irreducibility – implies that $a^* \in \mathbb{R}_+^n$.

If $a^* = 0$, then the bang for the buck $b(0, \hat{a})$ of starting at 0 and moving in the egalitarian direction $\hat{a}$ is no greater than 1 – otherwise, $F(d^{eg}(a))$ would not contain

\(^{55}\)Given $a, a' \in Y$ and $\lambda \in [0, 1]$, consider any $a'' = \lambda a + (1 - \lambda) a'$. Note that for all $i \in N$ and $\epsilon \geq -1$

$$u_i((1 + \epsilon)a'') \geq \lambda u_i((1 + \epsilon)a) + (1 - \lambda) u_i((1 + \epsilon)a')$$

by concavity of the $u_i$. Differentiating in $\epsilon$ at $\epsilon = 0$ yields the result.

\(^{56}\)Toward a contradiction, take $a$ such that $F(a)$ is empty. Let $z$ be the maximum $z$ such that $z a \in \overline{Y}$. Since $J(z a) a > 0$ it follows that for all $i$,\(^{57}\)

$$\frac{du_i((1 + \epsilon)a)}{d\epsilon} \bigg|_{\epsilon = 0} > 0,$$

from which it follows that $(z + \delta) a \in Y$ for small enough $\delta$. This contradicts the choice of $z$ (recalling the definition of $Y$).
\(a^* = 0\). By Proposition 3, it follows that \(r(B(0)) \leq 1\). Then \(0\) is a Lindahl outcome by Proposition 7.

### A.8 Proofs of Minor Facts

**Proof of Fact 1 in Section 3.4.2:** Consider the submatrix \(L\) obtained by leaving only indices in the cycle \(c\). Note \(L^\ell\) has diagonal entries, and hence row sums, bounded below by \(v(c;M)\). Then by Meyer (2000, p. 497) it follows that \(\rho(L^\ell) \geq v(c;M)\), and by Fact 3 in Section 3.4.3 that \(\rho(M^\ell) \geq v(c;M)\). The desired result then follows because \(\rho(M^\ell) = \rho(M)^\ell\) for any nonnegative matrix. (By Perron-Frobenius we know that the largest eigenvalue is always real and positive; moreover, \(\lambda\) is an eigenvalue of \(M\) if and only if \(\lambda^\ell\) is an eigenvalue of \(M^\ell\).)