Approximating Generalized Network Design under (Dis)economies of Scale with Applications to Energy Efficiency*

YUVAL EMEK, Technion - Israel Institute of Technology, Israel
SHAY KUTTEN, Technion - Israel Institute of Technology, Israel
RON LAVI, Technion - Israel Institute of Technology, Israel
YANGGUANG SHI, Technion - Israel Institute of Technology, Israel

In a *generalized network design (GND)* problem, a set of *resources* are assigned (non-exclusively) to multiple *requests*. Each request contributes its weight to the resources it uses and the total *load* on a resource is then translated to the cost it incurs via a resource specific cost function. Motivated by *energy efficiency* applications, recently, there is a growing interest in GND using cost functions that exhibit *(dis)economies of scale ((D)oS)*, namely, cost functions that appear subadditive for small loads and superadditive for larger loads.

The current paper advances the existing literature on approximation algorithms for GND problems with (D)oS cost functions in various aspects: (1) while the existing results are restricted to *routing* requests in undirected graphs, identifying the resources with the graph's edges, the current paper presents a generic *approximation framework* that yields approximation results for a much wider family of requests (including various types of *Steiner tree* and *Steiner forest* requests) in both directed and undirected graphs, where the resources can be identified with either the edges or the vertices; (2) while the existing results assume that a request contributes the same weight to each resource it uses, our approximation framework allows for *unrelated* weights, thus providing the first non-trivial approximation for the problem of *scheduling unrelated parallel machines* with (D)oS cost functions; (3) while most of the existing approximation algorithms are based on convex programming, our approximation framework is fully *combinatorial* and runs in strongly polynomial time; (4) the family of (D)oS cost functions considered in the current paper is more general than the one considered in the existing literature, providing a more accurate abstraction for practical energy conservation scenarios; and (5) we obtain the first approximation ratio for GND with (D)oS cost functions that depends only on the parameters of the resources' technology and does not grow with the number of resources, the number of requests, or their weights. The design of our approximation framework relies heavily on Roughgarden's *smoothness* toolbox (JACM 2015), thus demonstrating the possible usefulness of this toolbox in the area of approximation algorithms.

CCS Concepts: • **Theory of computation → Approximation algorithms analysis**; **Mathematical optimization**; **Algorithm design techniques**;

*An extended abstract of this paper has appeared in the Proceedings of the 50th Annual ACM Symposium on the Theory of Computing (STOC 2018).

Authors' addresses: Yuval Emek, Technion - Israel Institute of Technology, Haifa, Israel, 3200003, yemek@technion.ac.il; Shay Kutten, Technion - Israel Institute of Technology, Haifa, Israel, 3200003, kutten@technion.ac.il; Ron Lavi, Technion - Israel Institute of Technology, Haifa, Israel, 3200003, ronlavi@technion.ac.il; Yangguang Shi, Technion - Israel Institute of Technology, Haifa, Israel, 3200003, shiyangguang@technion.ac.il.

XXXX-XXXX/2019/12-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Additional Key Words and Phrases: Approximation algorithms, generalized network design, (dis)economies of scale, energy consumption, real exponent polynomial cost functions, smoothness, best response dynamics

ACM Reference Format:
Yuval Emek, Shay Kutten, Ron Lavi, and Yangguang Shi. 2019. Approximating Generalized Network Design under (Dis)economies of Scale with Applications to Energy Efficiency. 1, 1 (December 2019), 34 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Generalized Network Design. An instance $\mathcal{I}$ of a *generalized network design (GND)* problem is defined over a finite set E of *resources* and N abstract *requests*. Each request $i \in [N]$ is served by choosing some *reply* $p_i \subseteq E$ from request i's *reply collection* $P_i \subseteq 2^E$. Serving request i with reply p_i contributes $w_i(e)$ units to the *load* l_e on resource e for each $e \in p_i$, where $w_i \in \mathbb{Z}^E_{\geq 1}$ is the *weight vector* associated with request i (specified in $\mathcal{I}$). We emphasize that our GND setting supports *unrelated* weights, that is, request i may contribute different weights to the load on different resources in p_i.

One should serve all the requests of the instance $\mathcal{I}$ with replies $p = \{p_i\}_{i\in[N]}$, satisfying $p_i \in P_i$ for every $i \in [N]$, under the objective of minimizing the *total cost* $C(p)$. This is defined as $C(p) = \sum_{e\in E} F_e(l_e)$, where $F_e : \mathbb{Z}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a resource *cost function* that maps the load $l_e = l^p_e = \sum_{i\in[N]:e\in p_i} w_i(e)$ induced by p on resource e to the cost incurred by that resource.

We restrict our attention to GND problems with *succinctly represented* requests, namely, requests whose reply collections P_i can be specified using poly($|E|$) bits. These requests are often defined by identifying the resource set E with the edge set of a (directed or undirected) graph $G = (V, E)$, giving rise to, e.g., the following request types.

- *Routing* requests. This type of requests is concerned with connecting a given source-target pair. Formally, in a directed or undirected graph, each routing request i is specified by a pair $(s_i, t_i) \in V \times V$ of *terminals*, and the reply collection P_i is defined to consist of all (s_i, t_i)-paths in G.
- *Multi-routing* requests. Formally, in a directed or undirected graph, each multi-routing request i is specified by a collection $D_i \subseteq V \times V$ of *terminal* pairs, and the reply collection P_i is defined to consist of all edge subsets $F \subseteq E$ such that the subgraph (V, F) admits an (s, t)-path for every $(s, t) \in D_i$ (useful for designing a multicast scheme).[1]
- *Set connectivity* (resp., *set strong connectivity*) requests. This type of requests is concerned with connecting a given set of terminals. Formally, in an undirected (resp., directed) graph, each set connectivity (resp,. set strong connectivity) request i is specified by a set $T_i \subseteq V$ of *terminals*, and the reply collection P_i is defined to consist of all edge subsets that induce on G a connected (resp., strongly connected) subgraph that spans T_i (useful for designing an *overlay network*).

Alternatively, one can identify the resource set E with the vertex set of a graph, obtaining the vertex variants of the aforementioned request types, or with any other combinatorial structure as long as it fits into the aforementioned setting.

(Dis)economies of Scale. The classic network design literature addresses scenarios where the higher the load on a resource is, the lower is the cost per unit load, thus making it advisable to share network resources among requests, commonly known as *buy-at-bulk* network design [4, 10, 18, 20]. More formally, the cost functions $F_e(\cdot)$ in buy-at-bulk network design are assumed to be *subadditive*,

[1] Notice that the multi-routing request given by D_i cannot be (trivially) reduced to $|D_i|$ (single-)routing requests since a reply F for the former contributes $w_i(e)$ units to the load on edge $e \in F$ "only once", even if this edge is used to connect multiple terminal pairs in D_i.

i.e., they exhibit *economies of scale*. Recently, there is a growing interest in investigating network design problems with *superadditive* cost functions (i.e., cost functions exhibiting *diseconomies of scale*) [5, 37] or even cost functions that may appear subadditive for small loads and superadditive for larger loads [5, 6, 9], referred to as cost functions exhibiting *(dis)economies of scale ((D)oS)* [6].

The (D)oS cost functions studied so far in the context of network design capture the *energy consumption* of network devices employing the popular *speed scaling* technique [3, 6, 7, 15, 22, 29, 38, 47] that allows the device to adapt its power level to its actual load. Given a global constant parameter $\alpha \in \mathbb{R}_{>1}$ (a.k.a. the *load exponent*), an energy consumption cost function for resource $e \in E$ is defined by setting

$$F_e(l_e) = \begin{cases} 0, & l_e = 0 \\ \sigma_e + \xi_e \cdot l_e^{\alpha}, & l_e > 0 \end{cases}, \tag{1}$$

where $\sigma_e \in \mathbb{R}_{\geq 0}$ (the *startup cost*) and $\xi_e \in \mathbb{R}_{>0}$ (the *speed scaling factor*) are parameters of e.

This paper improves the existing results on approximation algorithms for GND with energy consumption cost functions in various aspects (see Section 1.1). In fact, our results apply to a more general class of resource cost functions exhibiting (D)oS, referred to as *real exponent polynomial (REP)* cost functions. Given global constant parameters $q \in \mathbb{Z}_{\geq 1}$ and $\alpha_1, \ldots, \alpha_q \in \mathbb{R}_{>1}$, a REP cost function for resource $e \in E$ is defined by setting

$$F_e(l_e) = \begin{cases} 0, & l_e = 0 \\ \sigma_e + \sum_{j \in [q]} \xi_{e,j} \cdot l_e^{\alpha_j}, & l_e > 0 \end{cases}, \tag{2}$$

where $\sigma_e \in \mathbb{R}_{\geq 0}$ and $\xi_{e,1}, \ldots, \xi_{e,q} \in \mathbb{R}_{\geq 0}$ are parameters of e, constrained by requiring that $\xi_{e,j} > 0$ for at least one $j \in [q]$.[2]

On top of the theoretical interest in studying more general cost functions, there is also a practical motivation behind their investigation. While some of the theoretical literature on energy efficient network design considers the special case of (1) where $\sigma_e = 0$ (see Section 1.1), it has been claimed [6, 9] that the startup cost component is crucial for better capturing practical energy consumption structures. In fact, in realistic communication networks, even the energy consumption cost functions of (1) may not be general enough since a link often consists of several different devices (e.g., transmitter/receiver, amplifier, adapter), all of which are operating when the link is in use. As their energy consumption may vary in terms of the load exponents and speed scaling factors, the functions presented in (1) do not provide a suitable abstraction for the link's energy consumption and the more general REP cost functions (2) should be employed.

Approximation Framework. Our main contribution is a novel *approximation framework* for GND problems with REP resource cost functions. This framework yields an approximation algorithm when provided access to an appropriate oracle that we now turn to define. A *reply ϱ-oracle*, $\varrho \geq 1$, for a family $\mathcal{Q}$ of succinctly represented requests is an efficient procedure that gets as input a resource set E, the reply collection $R \subseteq 2^E$ (specified succinctly) of a request in $\mathcal{Q}$, and a function $\tau : E \to \mathbb{R}_{>0}$, referred to as a *toll function*, that maps every resource $e \in E$ to a positive real number $\tau(e)$. The output of the reply ϱ-oracle is some reply $r \in R$ that minimizes the total toll $\tau(r) = \sum_{e \in r} \tau(e)$ up to factor ϱ, i.e., it satisfies $\tau(r) \leq \varrho \cdot \tau(r')$ for every $r' \in R$. An *exact reply oracle* is a reply ϱ-oracle with $\varrho = 1$.

Notice that the optimization problem behind the reply oracle is *not* a GND problem: it deals with a *single* request (rather than multiple requests) and the role of the resource cost functions (combined with the weight vectors) is now taken by the (single) toll function. In particular, while

[2]The scenario where $\xi_{e,j} = 0$ for every $j \in [q]$ is beyond the scope of this paper and left open. See Section 5 for more details.

all the (specific) GND problems mentioned in this paper are intractable (to various extents of inapproximability [5, 14, 42]), the request classes corresponding to some of them admit exact reply oracles.

For example, routing requests (in directed and undirected graphs) admit an exact reply oracle implemented using, e.g., Dijkstra's shortest path algorithm [24, 25]. In contrast, set connectivity requests in undirected graphs, set strong connectivity requests in directed graphs, and multi-routing requests in undirected and directed graphs do not admit exact reply oracles unless P = NP as these would imply exact (efficient) algorithms for the *Steiner tree*, *strongly connected Steiner subgraph*, *Steiner forest*, and *directed Steiner forest* problems, respectively. However, employing known approximation algorithms for the latter (Steiner) problems, one concludes that: set connectivity requests in undirected graphs admit a reply ϱ-oracle for $\varrho \leq 1.39$ [16]; set strong connectivity requests in directed graphs admit a reply t^{ϵ}-oracle, where $t = |T|$ is the number of terminals [17]; multi-routing requests in undirected graphs admit a reply 2-oracle [2]; and multi-routing requests in directed graphs admit a reply $k^{1/2+\epsilon}$-oracle, where $k = |D|$ is the number of terminal pairs [19]. This means, in particular, that set connectivity replies and multi-routing replies in undirected graphs always admit a reply ϱ-oracle with a constant approximation ratio ϱ, whereas set strong connectivity replies and multi-routing replies in directed graphs admit such an oracle whenever $|T|$ and $|D|$ are fixed. The guarantees of our approximation framework are cast in the following theorem.

Theorem 1. *Consider some GND problem $\mathcal{P}$ with succinctly represented requests using REP resource cost functions as defined in (2). Suppose that the requests of $\mathcal{P}$ admit a reply ϱ-oracle $O_{\mathcal{P}}$. When provided with black-box access to $O_{\mathcal{P}}$, our approximation framework yields a randomized efficient approximation algorithm $\mathcal{A}_{\mathcal{P}}$ for $\mathcal{P}$ whose approximation ratio is*

$$O\left(\varrho^{\max_j \alpha_j} + \varrho \cdot \max_e \min_j \left(\frac{\sigma_e}{\xi_{e,j}}\right)^{1/\alpha_j}\right)$$

with high probability. Moreover, our approximation framework runs in strongly polynomial time, so if $O_{\mathcal{P}}$ is implemented to run in strongly polynomial time, then $\mathcal{A}_{\mathcal{P}}$ also runs in strongly polynomial time.

Notice that our approximation framework is fully combinatorial and does not rely on solving convex programs. We emphasize that when $\varrho = O(1)$, the approximation ratio promised in Theorem 1 becomes

$$O\left(1 + \max_e \min_j \left(\frac{\sigma_e}{\xi_{e,j}}\right)^{1/\alpha_j}\right)$$

which is free of any dependence on the number $|E|$ of resources, the number N of requests, and the weight vectors $\{w_i\}_{i\in[N]}$; rather, it depends only on the parameters $(\sigma_e, \xi_{e,j})$ of the network resources' technology (speed scaling in case $q = 1$). Notice that the hidden expressions in our O notations may depend on the parameters q and $\alpha_1, \ldots, \alpha_q$ assumed to be constants throughout this paper.

1.1 Comparison to Existing Results

GND with Routing Requests. The existing literature on (generalized) network design beyond subadditive resource cost functions [5, 6, 9, 37] focuses on routing requests, identifying the resources with the edges of a graph, and with the exception of [37], it is restricted to undirected graphs and *related* weights, i.e., $w_i(e) = w_i$ for every $e \in E$. In contrast, the current paper handles a wider class of request types over much more general combinatorial structures (including both directed and undirected graphs) and our approximation framework supports unrelated weights. Moreover,

Table 1. Comparison of the approximation algorithms for GND with routing requests (identifying the resources with the graph edges) under the energy consumption cost functions (1) and restrictions thereof.

paper	graphs	weights	algorithm	restrictions	approx. ratio
[5]	undir.	related	math. prog.	none	$O\left(1+\left(\max_e \frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}} \log^{\alpha-1} w_{\max}\right)$
					$O\left(N+\log^{\alpha-1} w_{\max}\right)$
[6]	undir.	related	math. prog.	$\frac{\sigma_e}{\xi_e}=\frac{\sigma_{e'}}{\xi_{e'}}$	$\mathrm{polylog}(N)\cdot O\left(\log^{\alpha-1} w_{\max}\right)$
[9]	undir.	$w_i=1$	combin.	$\xi_e=1$	$O\left(\log^{\alpha} N\right)$
[5]	undir.	related	math. prog.	$\sigma_e=0$	$O\left(\log^{\alpha-1} w_{\max}\right)$
[37]	un/dir.	unrel.	math. prog.	$\sigma_e=0$	$(1+\epsilon)\mathcal{B}_\alpha$
current	un/dir.	unrel.	combin.	none	$O\left(1+\left(\max_e \frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}}\right)$

the current paper addresses the general REP cost functions (2), whereas as stated beforehand, the existing literature addresses only the energy consumption cost functions (1) and special cases thereof (Table 1 summarizes the relevant approximation upper bounds).

Specifically, Makarychev and Sviridenko [37] consider purely superadditive cost functions by restricting (1) to $\sigma_e = 0$ for all $e \in E$, obtaining an approximation ratio of $(1+\epsilon)\mathcal{B}_\alpha$, where $\mathcal{B}_\alpha$ is the fractional Bell number with parameter α. This improves the prior $O\left(\log^{\alpha-1} w_{\max}\right)$ upper bound of Andrews et al. [5], where $w_{\max} = \max_{i\in[N]} w_i$. The case where the startup cost σ_e may be positive is addressed by Antoniadis et al. [9], obtaining an approximation ratio of $O\left(\log^{\alpha} N\right)$, but this result is limited to the uniform case where $w_i = 1$ for all $i \in [N]$.

As stated in [6, 9], for a more accurate abstraction of practical energy conservation scenarios, the cost function definition of (1) with positive startup costs and arbitrary (related) weights is unavoidable. In this setting, three different approximation ratios have been devised by Andrews et al.: $O\left(\left(1+\max_e \frac{\sigma_e}{\xi_e}\right)^{1/\alpha} \log^{\alpha-1} w_{\max}\right)$ and $O\left(N+\log^{\alpha-1} w_{\max}\right)$ in [5]; and $\mathrm{polylog}(N)\cdot\log^{\alpha-1} w_{\max}$ in [6].[3]

We emphasize that these three approximation ratios grow with the number N of traffic requests and/or the maximum weight $w_{\max}$, whereas the approximation ratio established in the current paper depends only on the parameters of the network resources' technology. Furthermore, the algorithms behind these approximation ratios are based on linear/convex programming and their (currently known) implementations do not run in strongly polynomial time (this is true also for the algorithm of [37]). In contrast, the approximation framework developed in the present paper is purely combinatorial with a strongly polynomial run-time.

Scheduling Unrelated Parallel Machines. While GND with routing requests and related weights is a classic problem by its own right, generalizing it to unrelated weights not also makes this abstraction suitable for a wider class of GND scenarios, but also captures the extensively studied problem of *scheduling unrelated parallel machines*. This problem can be represented as GND with routing requests over a graph consisting of two vertices and multiple parallel edges (referred to as

[3] Actually, in [6], the startup cost term in the cost function is somewhat restricted.

machines) between them. The earlier algorithmic treatment of this problem considers the objective of minimizing the ℓ_∞ norm (a.k.a. *makespan*) of the machines' load [34, 45].[4] Later on, the focus has shifted to minimizing the ℓ_p norm of the machines' load for $p \in (1, \infty)$ [12–14, 33, 37]. The state of the art approximation algorithm in this regard is the one developed by Kumar et al. [33] with a < 2 approximation ratio for all $p \in (1, \infty)$. Makarychev and Sviridenko [37] studied this problem for small values of p and designed a $\sqrt[p]{\mathcal{B}_p}$-approximation, improving upon the upper bound of [33] for the $p \in (1, 2]$ regime.

The ℓ_p norm optimization criterion corresponds to the energy consumption cost function (1) restricted to zero startup costs $\sigma_e = 0$ (energy efficiency is also the main motivation of [37]). In practice, however, machines' energy consumption typically incurs a positive startup cost [6, 9]. This motivated Khuller et al. [31, 35] to study a variant of unrelated parallel machine scheduling in which the (sub)set of activated machines should satisfy some budget constraint on the startup costs. To the best of our knowledge, the current paper presents the first non-trivial approximation algorithm for scheduling unrelated parallel machines that takes into account the (positive) machines' startup costs $\sigma_e > 0$ as part of the objective function.

1.2 Paper Organization.

The rest of the paper is organized as follows. Section 2 introduces the concepts and notations used in the design and analysis of the proposed approximation framework. Following that, a technical overview of the approximation framework's design and analysis is provided in Section 3. The actual approximation framework is presented in Section 4 and analyzed in Section 5– 8. Two variants of the proposed approximation framework, which are more feasible for a decentralized environment, are presented in Section 9. In Section 10, we establish additional bounds that demonstrate the tightness of certain components in the analysis. Finally, alternative approaches for designing GND approximation algorithms are discussed in Section 11. In particular, Section 11.2 discusses an alternative algorithm for the GND problem with routing requests using convex optimization and randomized rounding.

2 PRELIMINARIES

Throughout, we consider some GND problem $\mathcal{P}$ with succinctly represented requests using REP resource cost functions (2). Let

$$\mathcal{I} = \left\langle E, \{P_i, \{w_i(e)\}_{e \in E}\}_{i \in [N]}, \{\alpha_j\}_{j \in [q]}, \left\{\sigma_e, \{\xi_{e,j}\}_{j \in [q]}\right\}_{e \in E} \right\rangle$$

be some $\mathcal{P}$ instance. Let p^* be an optimal solution for $\mathcal{I}$ and $C^* = C(p^*)$ be its total cost.

GND Games and Cost Sharing Mechanisms. A key ingredient of the approximation framework designed in this paper is a *GND game* derived from instance $\mathcal{I}$. In this game, each request $i \in [N]$ is associated with a strategic *player* i that decides on the reply $p_i \in P_i$ serving the request. In the scope of this GND game, the reply $p_i \in P_i$ is referred to as the *strategy* of player i and the reply collection P_i is referred to as her *strategy space*. We let $P = P_1 \times \cdots \times P_N$ and refer to $p = (p_1, \ldots, p_N) \in P$ as the (players') *strategy profile*. Although the strategy profile p is a vector of replies, we may slightly abuse the notation and write $e \in p$ when we mean that $e \in \bigcup_{i \in [N]} p_i$.

The cost $F_e(l_e)$ of each resource $e \in E$ is divided among the players based on a *cost sharing mechanism (CSM)* $M = \left\{f_{i,e}(\cdot)\right\}_{i \in [N], e \in E}$, where $f_{i,e} : P \to \mathbb{R}_{\geq 0}$ is a *cost sharing function* that determines the *cost share* $f_{i,e}(p)$ incurred by player $i \in [N]$ for resource e under strategy profile

[4] This objective does not fit into the formulation of minimizing the sum of the resource cost functions considered in our paper.

p. Player i chooses her strategy p_i with the objective of minimizing her *individual cost* $C_i(p) = \sum_{e \in E} f_{i,e}(p)$, irrespective of the total cost $C(p) = \sum_{i \in [N]} C_i(p)$ (a.k.a. the *social cost*).

CSM $M = \left\{f_{i,e}(\cdot)\right\}_{i \in [N], e \in E}$ is said to be *budget-balanced* (cf. [21, 46]) if $\sum_{i \in [N]} f_{i,e}(p) = F_e(l_e^p)$ for every resource $e \in E$. It is said to be *separable* and *uniform* (cf. [21, 46]) if the cost share of player $i \in [N]$ in resource $e \in E$ satisfies (1) if $e \notin p_i$, then $f_{i,e}(p) = 0$; and (2) $f_{i,e}(p)$ is fully determined by $w_i(e)$ and by the multiset of weights of the (other) players using resource e. Notice that if M is separable and uniform, then $f_{i,e}(p)$ is independent of the identities and weights of the players using any resource $e' \neq e$. It may be convenient to write $f_{i,e}(S_e)$ instead of $f_{i,e}(p)$, where $S_e = \{j \in [N] \mid e \in p_j\}$, although, strictly speaking, $f_{i,e}(p)$ is also independent of the identities (rather than weights) of the players in $S_e - \{i\}$. Unless stated otherwise, all CSMs considered in this paper are budget-balanced and separable and uniform.

Best Response. Following the convention in the game theoretic literature, given some $i \in [N]$ and a strategy profile $p = (p_1, \ldots, p_N)$, let $p_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_N)$; likewise, let $P_{-i} = P_1 \times \cdots \times P_{i-1} \times P_{i+1} \times \cdots \times P_N$. Given some approximation parameter $\chi \geq 1$, strategy $p_i \in P_i$ is an *approximate best response (ABR)* of player i to $p_{-i} \in P_{-i}$ if $C_i(p_i, p_{-i}) \leq \chi \cdot C_i(p'_i, p_{-i})$ for every $p'_i \in P_i$. A *best response (BR)* is an ABR with approximation parameter $\chi = 1$.[5]

A *best response dynamic (BRD)* (resp., *approximate best response dynamic (ABRD)*) is an iterative procedure that given an initial strategy profile $p^0 \in P$, generates a sequence $p^1, p^2, \ldots$ of strategy profiles adhering to the rule that for every $t = 1, 2, \ldots$, there exists some $i \in [N]$ such that (1) $p^t_{-i} = p^{t-1}_{-i}$; and (2) p^t_i is a BR (resp., ABR) of player i to p^{t-1}_{-i}.

Strategy profile $p \in P$ is a *(pure) Nash equilibrium (NE)* of the GND game if p_i is a BR to p_{-i} for every $i \in [N]$. The *(pure) price of anarchy (PoA)* of the GND game is defined to be the ratio $C(p)/C^*$, where $p \in P$ is a NE strategy profile that maximizes the social cost $C(p)$.

Smoothness. The following definition of Roughgarden [43] plays a key role in our analysis: Given parameters $\lambda > 0$ and $0 < \mu < 1$, we say that the GND game is (λ, μ)-*smooth* if

$$\sum_i C_i(p'_i, p_{-i}) \;\leq\; \lambda C(p') + \mu C(p) \tag{3}$$

for any two strategy profiles $p, p' \in P$.[6] The game is said to be *smooth* if it is (λ, μ)-smooth for some $\lambda > 0$ and $0 < \mu < 1$.[7]

Potential Functions. Function $\Phi : P \to \mathbb{R}^+$ is said to be a *potential function* if for every $i \in [N]$ and for any two strategy profiles p and p' with $p_{-i} = p'_{-i}$, it holds that

$$\Phi(p') - \Phi(p) \;=\; C_i(p') - C_i(p)\,.$$

A game admitting a potential function is said to be a *potential game*. The potential function $\Phi(p)$ is said to be (A, B)-*bounded* for some parameters $A \geq 1$ and $B \geq 1$ if

$$\Phi(p)/A \;\leq\; C(p) \;\leq\; B \cdot \Phi(p)$$

for any strategy profile $p \in P$.

Additional Notation and Terminology. Throughout, we think of $\epsilon > 0$ as a sufficiently small (positive) constant and fix $\epsilon_1 = \frac{1+\epsilon}{1-\epsilon}$. A probabilistic event A is said to occur *with high probability (w.h.p.)* if $\mathbb{P}(A) \geq 1 - 1/(|E| + N)^b$, where b is an arbitrarily large constant.

[5] See [39] for a more detailed description of the BR notion.

[6] The original definition of Roughgarden [43], that applies for all cost minimization games, also requires that $C(p) = \sum_{i \in [N]} C_i(p)$, but this property is assumed to hold for all CSMs considered in the current paper, so we do not mention it explicitly.

[7] Examples of smooth games are given in [43, Section 2.3].

3 TECHNICAL OVERVIEW

The key concept in the design of our generic approximation framework is to *decouple* the combinatorial structure of the specific GND problem $\mathcal{P}$, captured by the request types (and encoded in the reply collections), from the (D)oS cost functions of the individual resources. Informally, the former is handled by the reply oracle $O_{\mathcal{P}}$ (specifically tailored for $\mathcal{P}$), whereas for the latter, we harness the power of Roughgarden's *smoothness* toolbox [43]. Since this toolbox was originally introduced in the context of game theory rather than algorithm design, we first transform the given $\mathcal{P}$ instance into a GND game by carefully choosing the CSM (more on that soon). The algorithm then progresses via a sequence of player individual improvements in the form of a BRD, where each BRD step is implemented by invoking $O_{\mathcal{P}}$ with a toll function constructed based on the current strategy profile $p \in P$, the choice of player $i \in [N]$, and her cost sharing functions $f_{i,e}(\cdot)$, $e \in E$ (Section 4).[8]

In order to establish the promised upper bound on the approximation ratio, we first analyze the smoothness parameters of the aforementioned GND game (Section 6) which allows us to bound its PoA, thus ensuring that the total cost $C(p)$ of any NE strategy profile $p \in P$ provides the desired approximation for the (global) optimum C^*. This part of the proof relies on introducing and analyzing a new class of *REP-expanded* CSMs (Section 6), interesting in its own right.

One may hope that a BRD of the GND game converges to a NE strategy profile $p \in P$, but unfortunately, the BRD need not necessarily converge, and even if it does converge, it need not necessarily be in polynomially many steps. Inspired by another component of the smoothness toolbox [43] (which is in turn inspired by [11]), we show (in Section 5) that if the game admits a bounded potential function, then after simulating the BRD for polynomially many steps, one necessarily encounters a strategy profile $p \in P$ that yields the promised approximation guarantee (although it is not necessarily a NE).

Does our GND game admit the desired bounded potential function? The answer to this question depends, once again, on the choice of a CSM. We therefore look for a CSM with three (possibly conflicting) considerations in mind: the game that it induces must admit a bounded potential function; it must be REP-expanded; and it must be efficiently computable. We prove that the *Shapley CSM* satisfies the first two conditions (Section 7 and 6, respectively) and although its exact computation is #P-hard, we manage to adapt the approximation scheme of [36], originally designed for superadditive cost functions, to accommodate the REP cost functions (2) with positive startup costs $\sigma_e > 0$ (Section 8). This presents another obstacle though since the original technique of [43] assumes (implicitly) that each step in the BRD is (as the definition implies) an exact BR. To overcome this obstacle, we show that an ABRD is still good enough for our needs (Section 5).

We believe that the construction described here demonstrates the usefulness of algorithmic game theory tools for algorithm design even for optimization problems that on the face of it, are not at all concerned with game theory. A similar concept is demonstrated by Cole et al. [23] who obtained an improved combinatorial algorithm for job scheduling on unrelated machines, with the objective of minimizing the weighted sum of completion times, based on the game theoretic tools developed in [11]. In comparison, we employ the smoothness toolbox [43] for the design and analysis of our approximation framework. It is the robustness of this toolbox that plays the key role in the generality of our framework that can be applied to a wide family of GND problems. This is in contrast to most of the existing approximation algorithms for such problems that rely on linear/convex programming and are therefore heavily tailored to one specific GND problem and much less generic.

[8] In this section (only), we assume for simplicity that $O_{\mathcal{P}}$ is an exact reply oracle.

4 ALGORITHM DESCRIPTION

Let $O_{\mathcal{P}}$ be a reply ϱ-oracle for the requests of the GND problem $\mathcal{P}$. Our goal is to design an approximation algorithm with black-box access to $O_{\mathcal{P}}$ as promised in Theorem 1. We shall refer to this approximation algorithm as `Alg-ABRD`.

Given an instance $\mathcal{I} = \left\langle E, \{P_i, \{w_i(e)\}_{e\in E}\}_{i\in[N]}, \left\{\alpha_j\right\}_{j\in[q]}, \left\{\sigma_e, \left\{\xi_{e,j}\right\}_{j\in[q]}\right\}_{e\in E}\right\rangle$ of $\mathcal{P}$, we first construct (conceptually) the GND game induced by $\mathcal{I}$ and a carefully chosen CSM $M = \left\{f_{i,e}(\cdot)\right\}_{i\in[N],e\in E}$. On top of the other properties of M that will be discussed in the next sections, we require that M is *poly-time ϵ-computable*, namely, that given $\mathcal{I}$, $p \in P$, and $i \in [N]$, it is possible to compute in time poly$(|E|, N)$ some *ϵ-cost shares* $\widetilde{f}_{i,e}(p)$, $e \in E$, that satisfy

$$(1-\epsilon)f_{i,e}(p) \leq \widetilde{f}_{i,e}(p) \leq (1+\epsilon)f_{i,e}(p)$$

w.h.p. Define the *ϵ-individual cost* $\widetilde{C}_i(p)$ to be the sum $\widetilde{C}_i(p) = \sum_{e\in E}\widetilde{f}_{i,e}(p)$, which means that

$$(1-\epsilon)C_i(p) \leq \widetilde{C}_i(p) \leq (1+\epsilon)C_i(p)$$

w.h.p. As we shall perform the computations of the ϵ-cost shares (and the ϵ-individual costs) poly$(|E|, N)$ times, all of them succeed w.h.p.; condition hereafter on this event.

To simplify the presentation, we assume that the values of the ϵ-cost shares $\widetilde{f}_{i,e}(p)$, $e \in E$, and the ϵ-individual costs $\widetilde{C}_i(p)$ have already been fixed before the algorithm's execution for all $i \in [N]$ and $p \in P$ in an (arbitrary) manner that satisfies the aforementioned ϵ-approximation inequalities; the algorithm then merely "exposes" some (poly$(|E|, N)$ many) of these values. The following lemma plays a key role in the design of `Alg-ABRD`.

Lemma 4.1. *If M is a poly-time ϵ-computable CSM, then there exists a randomized procedure that given $i \in [N]$ and $p_{-i} \in P_{-i}$, runs in time* poly$(|E|, N)$ *and computes a strategy $p_i \in P_i$ and the corresponding ϵ-individual cost $\widetilde{C}_i(p_i, p_{-i})$ such that $\widetilde{C}_i(p_i, p_{-i}) \leq \varrho \cdot \widetilde{C}_i(p'_i, p_{-i})$ for any $p'_i \in P_i$. This means in particular that p_i is an ABR of player i to p_{-i} with approximation parameter $\varrho\epsilon_1$.*[9]

Proof. Construct the toll function $\tau_{i,p_{-i}} : E \to \mathbb{R}_{\geq 0}$ by setting $\tau_{i,p_{-i}}(e)$ to be the ϵ-cost share $\widetilde{f}_{i,e}(S_e \cup \{i\})$, where $S_e = \{j \in [N] - \{i\} \mid e \in p_j\}$. This can be done in time poly$(|E|, N)$ since M is poly-time ϵ-computable. The assumption that M is separable and uniform guarantees that a reply $p_i \in P_i$ that minimizes the total toll $\sum_{e\in p_i}\tau_{i,p_{-i}}(e)$ up to factor ϱ satisfies $\widetilde{C}_i(p_i, p_{-i}) \leq \varrho \cdot \widetilde{C}_i(p'_i, p_{-i})$ for any $p'_i \in P_i$ and that the sum $\sum_{e\in p_i}\tau_{i,p_{-i}}(e)$ is the desired ϵ-individual cost. Such a reply p_i can be computed using the reply ϱ-oracle $O_{\mathcal{P}}$. □

Employing the procedure promised by Lemma 4.1, `Alg-ABRD` simulates an ABRD $p^0, p^1, \ldots$ of the GND game induced by $\mathcal{I}$ and M that includes at most T iterations for some $T =$ poly$(|E|, N)$ whose exact value will be determined later. This is done as follows (see also Pseudocode 1).

Set p^0 by taking p^0_i, $i \in [N]$, to be the strategy generated by $O_{\mathcal{P}}$ for the toll function τ^0_i defined by setting $\tau^0_i(e) = F_e(w_i(e))$. Assuming that p^{t-1}, $1 \leq t \leq T$, was already constructed, we construct p^t as follows. For $i \in [N]$, employ the procedure promised by Lemma 4.1 to compute an ABR p'_i of player i to p^{t-1}_{-i} and let $\delta^t_i = \widetilde{C}_i(p^{t-1}) - \epsilon_1 \cdot \widetilde{C}_i(p'_i, p^{t-1}_{-i})$. If $\delta^t_i \leq 0$ for all $i \in [N]$, then the ABRD stops, and we set $p^t = p^{t-1}$; in this case, we say that the ABRD *converges*. Otherwise, fix $\Delta^t = \sum_{i\in[N]}\delta^t_i$ and choose some player $i' \in [N]$ so that

$$\delta^t_{i'} > 0 \quad \text{and} \quad \delta^t_{i'} \geq \frac{1}{N}\Delta^t \tag{4}$$

[9] All subsequent occurrences of the term ABR (and ABRD) share the same approximation parameter $\varrho\epsilon_1$, hence we may refrain from mentioning this parameter explicitly.

ALGORITHM 1: Alg-ABRD

Input: A GND instance $\mathcal{I} = \left\langle E, \{P_i, \{w_i(e)\}_{e\in E}\}_{i\in[N]}, \left\{\alpha_j\right\}_{j\in[q]}, \left\{\sigma_e, \left\{\xi_{e,j}\right\}_{j\in[q]}\right\}_{e\in E}\right\rangle$.
Output: A profile $p \in P$ that is feasible for the given instance $\mathcal{I}$.
for $i \in [N]$ **do**
  set τ_i^0 to be the toll function defined by setting $\tau_i^0(e) = F_e(w_i(e))$ for every $e \in E$;
  set p_i^0 to be the output of oracle $\mathcal{O}_{\mathcal{P}}$ on E, P_i, and τ_i^0;
end
$t \leftarrow 0$;
while $t < T$ **do**
  $t \leftarrow t + 1$;
  for $i \in [N]$ **do**
    set p_i' to be an ABR of player i to p_{-i}^{t-1} with approximation parameter $\varrho\epsilon_1$;
    $\delta_i^t \leftarrow \widetilde{C}_i(p^{t-1}) - \epsilon_1 \cdot \widetilde{C}_i(p_i', p_{-i}^{t-1})$;
  end
  if $\delta_i^t \le 0$ *for all* $i \in [N]$ **then**
    $p^t \leftarrow p^{t-1}$;
    break;
  end
  $\Delta^t \leftarrow \sum_{i\in[N]} \delta_i^t$;
  pick some $j \in [N]$ such that $\delta_i^t > 0$ and $\delta_i^t \ge \Delta^t/N$;
  $p^t \leftarrow (p_j', p_{-j}^{t-1})$;
end
$t^* = \operatorname{argmin}_t C(p^t)$;
return p^{t^*};

to update her strategy, setting $p^t = (p_{i'}', p_{-i'}^{t-1})$ (the existence of such a player is guaranteed by the pigeonhole principle).

When the ABRD terminates (either because it has reached iteration $t = T$ or because it converged), Alg-ABRD chooses an iteration t^* such that the corresponding strategy profile p^{t^*} has a minimum total cost $C(p^{t^*})$ and outputs p^{t^*}. (Recall that in contrast to the player individual costs, the social cost can always be computed efficiently.)

5 ANALYZING ALG-ABRD

In this section, we begin our journey towards bounding the approximation ratio and run-time of Alg-ABRD as promised by Theorem 1. The analysis relies on a careful choice of the CSM $M = \left\{f_{i,e}(\cdot)\right\}_{i\in[N],e\in E}$. In particular, we are looking for a CSM whose induced GND game is smooth and admits a bounded potential function with the right choice of parameters.[10] The reason for that will be made clear in Theorem 5.3 whose proof relies on Lemma 5.1 and 5.2; the former provides an upper bound on the approximation ratio when the ABRD converges, whereas the latter is used to bound the number T of steps in the ABRD (and is the key to ensuring strongly polynomial run-time).

Lemma 5.1. *Suppose that the CSM M is chosen so that the induced GND game is (λ, μ)-smooth with $\mu < 1/(\varrho\epsilon_1^2)$. If the ABRD simulated in* Alg-ABRD *converges at step t for any $t \in [T]$, then the last*

[10] It is an open question whether there exists a CSM that induces a smooth game when $\xi_{e,j} = 0$ for every $j \in [q]$.

strategy profile p^t satisfies

$$C(p^t) \le \frac{\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu} \cdot C^* .$$

Proof. Recalling that we use p_i' to represent the ABR of player i to p^t, we observe that

$$\begin{aligned} C(p^t) &= \sum_i C_i(p^t) \\ &\le \frac{1}{1-\epsilon}\sum_i \widetilde{C}_i(p^t) \\ &\le \frac{1}{1-\epsilon}\epsilon_1 \sum_i \widetilde{C}_i(p_i', p_{-i}^t) \\ &\le \frac{1}{1-\epsilon}\epsilon_1 \cdot \varrho \sum_i \widetilde{C}_i(p_i^*, p_{-i}^t) \\ &\le \varrho\epsilon_1^2 \cdot \sum_i C_i(p_i^*, p_{-i}^t) \\ &\le \varrho\epsilon_1^2(\lambda \cdot C^* + \mu \cdot C(p^t)) , \end{aligned}$$

where the second and fifth transitions follow from the definition of ϵ-individual cost, the third transition holds since the algorithm converges at step t, the fourth transition holds following Lemma 4.1, and the sixth transition follows from the definition of (λ, μ)-smoothness. □

Lemma 5.2. *The initial strategy profile p^0 of* Alg-ABRD *satisfies $C(p^0) \le \varrho \cdot N^{\max_j \alpha_j} \cdot C^*$.*

Proof. The construction of p^0 guarantees that

$$\sum_{e\in p_i^0}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}(w_i(e))^{\alpha_j}\right] \le \varrho \cdot \sum_{e\in p_i^*}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}(w_i(e))^{\alpha_j}\right].$$

Therefore,

$$\begin{aligned} \sum_{i\in[N]}\sum_{e\in p_i^0}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}(w_i(e))^{\alpha_j}\right] &\le \varrho \cdot \sum_{i\in[N]}\sum_{e\in p_i^*}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}(w_i(e))^{\alpha_j}\right] \\ &\le \varrho \cdot \sum_{e\in p^*}\left[N\cdot\sigma_e + \sum_{j\in[q]}\xi_{e,j}\cdot\sum_{i:e\in p_i^*}(w_i(e))^{\alpha_j}\right] \\ &\le \varrho \cdot \sum_{e\in p^*}\left[N\cdot\sigma_e + \sum_{j\in[q]}\xi_{e,j}\cdot\Big(\sum_{i:e\in p_i^*}w_i(e)\Big)^{\alpha_j}\right] \\ &\le \varrho \cdot N\sum_{e\in p^*}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}\cdot\Big(\sum_{i:e\in p_i^*}w_i(e)\Big)^{\alpha_j}\right] \\ &= \varrho N \cdot C^* , \end{aligned}$$

where the third transition follows from the superadditivity and the last transition holds since $l_e^{p^*} > 0$ for every $e \in p^*$. Then,

$$\begin{aligned}
C(p^0) &= \sum_{e\in p^0}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}\Big(\sum_{i:e\in p_i^0} w_i(e)\Big)^{\alpha_j}\right] \\
&\le \sum_{e\in p^0}\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}\cdot N^{\alpha_j-1}\sum_{i:e\in p_i^0}(w_i(e))^{\alpha_j}\right] \\
&\le \sum_{e\in p^0}\left[\sum_{i:e\in p_i^0}\sigma_e + \sum_{j\in[q]}\xi_{e,j}\cdot N^{\alpha_j-1}\sum_{i:e\in p_i^0}(w_i(e))^{\alpha_j}\right] \\
&\le N^{\max_j \alpha_j-1}\sum_{e\in p^0}\sum_{i:e\in p_i^0}\left[\sigma_e + \sum_j \xi_{e,j}(w_i(e))^{\alpha_j}\right] \\
&= N^{\max_j \alpha_j-1}\sum_{i\in[N]}\sum_{e\in p_i^0}\left[\sigma_e + \sum_j \xi_{e,j}(w_i(e))^{\alpha_j}\right] \\
&\le \varrho N^{\max_j \alpha_j}\cdot C^*,
\end{aligned}$$

where the second transition holds because the convexity indicates that

$$\left(\frac{\sum_{i:e\in p_i^0} w_i(e)}{|S_e^{p^0}|}\right)^{\alpha_j} \le \frac{1}{|S_e^{p^0}|}\sum_{i:e\in p_i^0}(w_i(e))^{\alpha_j},$$

which means that

$$\Big(\sum_{i:e\in p_i^0} w_i(e)\Big)^{\alpha_j} \le |S_e^{p^0}|^{\alpha_j-1}\sum_{i:e\in p_i^0}(w_i(e))^{\alpha_j} \le N^{\alpha_j-1}\sum_{i:e\in p_i^0}(w_i(e))^{\alpha_j}.$$

The assertion follows. □

Theorem 5.3. *Suppose that the CSM M is chosen so that the induced GND game admits an (A,B)-bounded potential function Φ and is (λ,μ)-smooth with $\mu < 1/(\varrho\epsilon_1^2)$. Let $Q = \frac{2\epsilon_1 NA}{1-\varrho\epsilon_1^2\mu}$. If $T = \lceil Q\cdot\ln(ABN^{\max_j\alpha_j})\rceil$, then the output p^{t^*} of* `Alg-ABRD` *satisfies*

$$C(p^{t^*}) \le \frac{2\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu}\cdot C^*.$$

Proof. Lemma 5.1 ensures that the assertion holds if our ABRD converges at any step $t \le T$, so it is left to analyze the case where the ABRD does not converge. We say that profile p^t of the ABRD is *bad* if

$$C(p^t) > \frac{2\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu}\cdot C^*.$$

Claim 5.4. *For any $t < T$, if p^t is bad, then $\Phi(p^{t+1}) < (1-1/Q)\cdot\Phi(p^t)$.*

Proof. Fix

$$d^t = \frac{1}{1-\epsilon}\Big[\sum_{i\in[N]}\widetilde{C}_i(p^t) - \varrho\cdot\epsilon_1\sum_{i\in[N]}\widetilde{C}_i(p_i^*, p_{-i}^t)\Big]. \tag{5}$$

This means that

$$\begin{aligned} C(p^t) \;=\; \sum_{i\in[N]} C_i(p^t) \;&\le\; \frac{1}{1-\epsilon}\sum_{i\in[N]} \widetilde{C}_i(p^t) \\ &= \frac{\varrho\epsilon_1}{1-\epsilon}\sum_{i\in[N]} \widetilde{C}_i(p_i^*, p_{-i}^t) + d^t \\ &\le \varrho\epsilon_1^2 \sum_{i\in[N]} C_i(p_i^*, p_{-i}^t) + d^t \\ &\le \varrho\epsilon_1^2(\lambda\cdot C^* + \mu C(p^t)) + d^t\,. \end{aligned}$$

Therefore, $d^t \ge \left[1-\varrho\epsilon_1^2\mu\right] C(p^t) - \varrho\epsilon_1^2\lambda\cdot C^*$, hence, if p^t is bad, then d^t satisfies

$$d^t \;>\; \left[1-\varrho\epsilon_1^2\mu\right] C(p^t) - \frac{1-\varrho\epsilon_1^2\mu}{2} C(p^t) \;=\; \frac{1-\varrho\epsilon_1^2\mu}{2} C(p^t)\,. \tag{6}$$

Since the ABRD does not converge at step t, there exists a player i^t being selected to update its strategy. Recalling that the ABR of player i to p^t is denoted by p'_i, we observe that

$$\begin{aligned} \Phi(p^t) - \Phi(p^{t+1}) \;&= C_{i^t}(p^t) - C_{i^t}(p'_{i^t}, p^t_{-i^t}) \\ &\ge \frac{1}{1+\epsilon}\widetilde{C}_{i^t}(p^t) - \frac{1}{1-\epsilon}\widetilde{C}_{i^t}(p'_{i^t}, p^t_{-i^t}) \\ &= \frac{1}{1+\epsilon}\left[\widetilde{C}_{i^t}(p^t) - \epsilon_1\widetilde{C}_{i^t}(p'_{i^t}, p^t_{-i^t})\right] \\ &\ge \frac{1}{1+\epsilon}\cdot\frac{1}{N}\sum_{i\in[N]}\left[\widetilde{C}_i(p^t) - \epsilon_1\widetilde{C}_i(p'_i, p^t_{-j})\right] \\ &\ge \frac{1}{1+\epsilon}\cdot\frac{1}{N}\sum_{i\in[N]}\left[\widetilde{C}_i(p^t) - \varrho\cdot\epsilon_1\widetilde{C}_i(p_i^*, p^t_{-j})\right] \\ &= \frac{1-\epsilon}{1+\epsilon}\cdot\frac{d^t}{N} \\ &> \frac{1-\epsilon}{1+\epsilon}\cdot\frac{1}{2N}\left[1-\varrho\epsilon_1^2\mu\right] C(p^t) \\ &\ge \frac{1}{\epsilon_1}\cdot\frac{1}{2N}\left[1-\varrho\epsilon_1^2\mu\right]\frac{\Phi(p^t)}{A}\,, \end{aligned}$$

where the fourth transition follows from Eq. (4), the fifth transition holds since p'_{i^t} is the ABR promised by Lemma 4.1, which means that $\widetilde{C}_{i^t}(p'_{i^t}, p^t_{-i^t}) \le \varrho\cdot\widetilde{C}_{i^t}(p^*_{i^t}, p^t_{-i^t})$, the sixth and seventh transitions follow from Eq. (5) and Eq. (6), respectively, and the last transition holds because the potential function is assumed to be (A, B)-bounded. Therefore,

$$\Phi(p^{t+1}) \;<\; \Phi(p^t)\left(1 - \frac{1-\varrho\epsilon_1^2\mu}{2\epsilon_1 NA}\right) \;=\; (1-1/Q)\cdot\Phi(p^t)$$

as promised. ■ (Claim 5.4)

Since `Alg-ABRD` outputs the strategy profile with the minimum total cost among all the generated strategy profiles, this theorem holds if any of these strategy profiles is not bad.

Claim 5.5. *If all the $T+1$ strategy profiles in the ABRD are bad, then $C(p^T) < \varrho\cdot C^*$.*

Proof. Claim 5.4 implies that if all the $T+1$ profiles in the ABRD are bad, then

$$\Phi(p^T) < \left(1-\frac{1}{Q}\right)^T \Phi(p^0) = \left(1-\frac{1}{Q}\right)^{\left\lceil Q\cdot\ln\left(ABN^{\max_j \alpha_j}\right)\right\rceil} \Phi(p^0) \leq \frac{1}{ABN^{\max_j \alpha_j}}\Phi(p^0).$$

By the definition of the bounded potential function and by Lemma 5.2, we have

$$C(p^T) \leq B\cdot\Phi(p^T) < \frac{B}{ABN^{\max_j \alpha_j}}\Phi(p^0) \leq \frac{A}{AN^{\max_j \alpha_j}}C(p^0) \leq \frac{\varrho N^{\max_j \alpha_j}C^*}{N^{\max_j \alpha_j}} = \varrho\cdot C^*$$

which completes the proof. ■ (Claim 5.5)

Claim 5.6. $\varrho < \frac{2\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu}$.

Proof. For any cost minimization (λ,μ)-smooth game that has a (bounded) potential function, we have $\frac{\lambda}{1-\mu} \geq 1$. This is because the existence of a potential function implies the existence of a (pure) NE $p \in P$ with $C(p) \leq \frac{\lambda}{1-\mu}C^*$ [43]. Therefore, $\frac{2\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu} > \frac{2\varrho\lambda}{1-\mu} > \varrho$. ■ (Claim 5.6)

By combining Claims 5.5 and 5.6, we conclude that not all $T+1$ profiles are bad, thus completing the proof. □

Remark 5.7. *Roughgarden [43] proves that in the BRD of a (λ,μ)-smooth game, the number of strategy profiles whose cost is larger than $\frac{\lambda}{(1-\upsilon)(1-\mu)}\cdot C^*$ for some constant $\upsilon \in (0,1)$ is bounded by a polynomial. However, his proof depends on the exact values of the cost shares and exact best responses, both of which may be intractable in our GND setting.*

In the following sections, we search for a CSM whose induced GND game is (λ,μ)-smooth and admits an (A,B)-bounded potential function for parameters λ, μ, A, and B that when plugged into Theorem 5.3, yield the desired approximation ratio and run-time bounds.

6 SMOOTHNESS OF THE GND GAME

In this section, a rather wide class of CSMs, the *REP-expanded* CSMs, is presented and the smoothness parameters of the induced GND games are analyzed. This class is introduced because it includes every CSM that we investigate in the scope of this paper and provides a uniform way to study the smoothness of the GND games induced by these CSMs. The proof that an adequate potential function exists for (the GND game induced by) one of these CSMs is deferred to Section 7.

A CSM (for GND games) is said to be REP-expanded if the cost share $f_{i,e}(p)$ satisfies

$$f_{i,e}(p) \leq \sigma_e + \sum_{j\in[q]} \xi_{e,j}\cdot\left(z_{1,j}\left(l_e^p - w_i(e)\right)^{\alpha_j-1}\cdot w_i(e) + z_{2,j}\left(w_i(e)\right)^{\alpha_j}\right), \tag{7}$$

for any player $i \in [N]$, edge $e \in E$, and strategy profile $p \in P$, where $z_{1,j}$ and $z_{2,j}$ are non-negative constants that can only depend on α_j. For convenience, we also write Eq. (7) as

$$f_{i,e}(p) \leq \sigma_e + \sum_{j\in[q]} \xi_{e,j}\sum_{k=1}^{2} z_{k,j}\left(l_e^p - w_i(e)\right)^{x_{k,j}}\left(w_i(e)\right)^{y_{k,j}},$$

where $x_{1,j} = \alpha_{j-1}$, $y_{1,j} = 1$, $x_{2,j} = 0$, and $y_{2,j} = \alpha_j$. Note that the exponents $\{x_{1,j}, y_{2,j}\}_{j\in[q]}$ and the coefficients $\{z_{1,j}, z_{2,j}\}_{j\in[q]}$ are not necessarily integral.

Theorem 6.1. *Consider some REP-expanded CSM M. For any $\varrho \geq 1$, the GND game induced by M is guaranteed to be $\left(\gamma_\alpha + \lambda_\alpha\cdot\varrho^{\max_j \alpha_j-1}, 1/(2\varrho)\right)$-smooth, where $\gamma_\alpha = \max_{e\in E}\min_{j\in[q]}\left(\frac{1}{\alpha_j-1}\cdot\frac{\sigma_e}{\xi_{e,j}}\right)^{1/\alpha_j}$ and $\lambda_\alpha > 0$ is a positive constant that depends only on q and $\alpha_1,\ldots,\alpha_q$.*

Proof. Our goal in this proof is to show that (3) holds with $\lambda = \gamma_\alpha + \lambda_\alpha \cdot \varrho^{\max_j \alpha_j}$ and $\mu = 1/(2\varrho)$. We begin by observing that

$$\begin{aligned}
\sum_{i\in[N]} C_i(p'_i, p_{-i}) &= \sum_{i\in[N]} \sum_{e\in p'_i} f_{i,e}(p'_i, p_{-i}) \\
&\le \sum_{i\in[N]} \sum_{e\in p'_i} \left(\sigma_e + \sum_{j\in[q]} \xi_{e,j} \sum_{k\in\{1,2\}} z_{k,j} (l^p_e)^{x_{k,j}} (w_i(e))^{y_{k,j}} \right) \\
&\le \sum_{i\in[N]} \sum_{e\in p'_i} \sigma_e + \sum_{j\in[q],k\in\{1,2\}} z_{k,j} \sum_{e\in p'} \xi_{e,j} (l^p_e)^{x_{k,j}} \sum_{i:e\in p'_i} (w_i(e))^{y_{k,j}} \\
&\le \sum_{e\in p'} \sigma_e l^{p'}_e + \sum_{j\in[q],k\in\{1,2\}} z_{k,j} \sum_{e\in p'} \xi_{e,j} (l^p_e)^{x_{k,j}} (l^{p'}_e)^{y_{k,j}} ,
\end{aligned}$$

where the second transition follows by the definition of REP-expanded CSMs because when player i deviates to p'_i, the load on edge $e \in p'_i$ is at most $l^p_e + w_i(e)$ and the last transition holds because (1) $w_i(e) \ge 1$, hence $|\{i \in [N] : e \in p'_i\}| \le l^{p'}_e$ for any edge e; and (2) $(w_i(e))^{y_{k,j}}$ is a superadditive function of $w_i(e)$, hence $\sum_{i:e\in p'_i} (w_i(e))^{y_{k,j}} \le (\sum_{i:e\in p'_i} w_i(e))^{y_{k,j}} = (l^{p'}_e)^{y_k}$. The desired upper bound on $\sum_{e\in p'} \sigma_e l^{p'}_e + \sum_{j,k} z_{k,j} \sum_{e\in p'} \xi_{e,j} (l^p_e)^{x_{k,j}} (l^{p'}_e)^{y_{k,j}}$ is established in Claims 6.2 and 6.3.

Claim 6.2. $\sum_{e\in p'} \sigma_e l^{p'}_e \le \gamma_\alpha \cdot C(p')$

Proof. Define the function $g(x) = \frac{\sigma x}{\sigma + \xi x^\alpha}$ for arbitrary positive numbers $\sigma > 0$, $\xi > 0$ and $\alpha > 1$. Since its derivative is

$$g'(x) = \frac{\sigma}{(\sigma + \xi x^\alpha)^2} [\sigma - (\alpha - 1)\xi x^\alpha] ,$$

it attains its maximum for $x \ge 0$ at $x = \left(\frac{\sigma}{\xi(\alpha-1)}\right)^{1/\alpha}$. Therefore, for any $x \ge 0$, we have

$$\frac{\sigma \cdot x}{\sigma + \xi \cdot x^\alpha} = g(x) \le g\left((\sigma/\xi(\alpha-1))^{1/\alpha}\right) = \frac{\sigma \left(\frac{\sigma}{\xi(\alpha-1)}\right)^{1/\alpha}}{\sigma + \xi \frac{\sigma}{\xi(\alpha-1)}} = \left(\frac{1}{\alpha-1} \cdot \frac{\sigma}{\xi}\right)^{1/\alpha} \Big/ \left(1 + \frac{1}{\alpha-1}\right) .$$

Let $j^*_e \in \operatorname{argmin}_{j\in[q]} \left(\frac{1}{\alpha_j - 1} \cdot \frac{\sigma_e}{\xi_{e,j}}\right)^{1/\alpha_j}$. By the inequality above, we have

$$\sigma_e l^{p'}_e < \left(\frac{1}{\alpha_{j^*_e} - 1} \cdot \frac{\sigma_e}{\xi_{e,j^*_e}}\right)^{1/\alpha_{j^*_e}} \cdot \left[\sigma_e + \xi_{e,j^*_e} (l^{p'}_e)^{1/\alpha_{j^*_e}}\right] \le \left(\frac{1}{\alpha_{j^*_e} - 1} \cdot \frac{\sigma_e}{\xi_{e,j^*_e}}\right)^{1/\alpha_{j^*_e}} \cdot F_e(l^{p'}_e) .$$

The assertion follows since $\left(\frac{1}{\alpha_{j^*_e} - 1} \cdot \frac{\sigma_e}{\xi_{e,j^*_e}}\right)^{1/\alpha_{j^*_e}} \le \gamma_\alpha$ for every $e \in E$. ■ (Claim 6.2)

Fix $z_{\max} = \lceil \max_{j,k} z_{k,j} \rceil$ and let $\lambda_\alpha = (2 \cdot z_{\max})^{\max_j \alpha_j}$.

Claim 6.3. $\sum_{j=1}^{q} \sum_{k=1}^{2} z_{k,j} \sum_{e\in p'} \xi_{e,j} (l^p_e)^{x_{k,j}} (l^{p'}_e)^{y_{k,j}} \le \frac{C(p)}{2\varrho} + \lambda_\alpha \varrho^{\max_j \alpha_j - 1} C(p')$.

Proof. Let p and p' be any two profiles. First, consider the term $\sum_{j=1}^{q} \sum_{e\in p'} z_{1,j} \xi_{e,j} (l^p_e)^{\alpha_j - 1} \cdot l^{p'}_e$.

- If $l^p_e \le 2 z_{\max} \varrho \cdot l^{p'}_e$, then since $\alpha_j - 1$ for every j, we have

$$(l^p_e)^{\alpha_j - 1} \cdot l^{p'}_e \le (2 z_{\max} \varrho)^{\alpha_j - 1} \cdot (l^{p'}_e)^{\alpha_j} .$$

- If $l_e^p > 2z_{\max}\varrho \cdot l_e^{p'}$, then

$$(l_e^p)^{\alpha_j - 1} \cdot l_e^{p'} < \frac{1}{2z_{\max}\varrho}(l_e^p)^{\alpha_j}.$$

Therefore, we have

$$\begin{aligned}\sum_{j\in[q]}\sum_{e\in p'} z_{1,j}\xi_{e,j}(l_e^p)^{\alpha_j-1}\cdot l_e^p &= \sum_{j\in[q]}\sum_{e\in p'\cap p} z_{1,j}\cdot\xi_{e,j}(l_e^p)^{\alpha_j-1}\cdot l_e^p \\ &\le z_{\max}\sum_{j\in[q]}\sum_{e\in p\cap p'}\xi_{e,j}\left[(2z_{\max}\varrho)^{\alpha_j-1}\cdot(l_e^{p'})^{\alpha_j}+\frac{1}{2z_{\max}\varrho}(l_e^p)^{\alpha_j}\right] \\ &\le (2\varrho)^{\max_{j\in[q]}\alpha_j-1}(z_{\max})^{\max_{j\in[q]}\alpha_j}\vartheta(p')+\frac{1}{2\varrho}\vartheta(p),\end{aligned}$$

where $\vartheta(p)=\sum_{j\in[q]}\sum_{e\in p}\xi_{e,j}(l_e^p)^{\alpha_j}$, and $\vartheta(p')=\sum_{j\in[q]}\sum_{e\in p'}\xi_{e,j}(l_e^{p'})^{\alpha_j}$. The first transition above holds because for any $e\in p-p'$, the load $l_e^p=0$. Since $\sum_{j\in[q]}\sum_{e\in p'} z_{2,j}\xi_{e,j}(l_e^{p'})^{\alpha_j}\le z_{\max}\cdot\vartheta(p')$, we have

$$\begin{aligned}\sum_{j=1}^{q}\sum_{k=1}^{2} z_{k,j}\sum_{e\in p'}\xi_{e,j}(l_e^p)^{x_{k,j}}(l_e^{p'})^{y_{k,j}} &\le (2\varrho)^{\max_j\alpha_j-1}(z_{\max})^{\max_j\alpha_j}\vartheta(p')+\frac{1}{2\varrho}\vartheta(p)+z_{\max}\vartheta(p') \\ &\le \lambda_\alpha\cdot\varrho^{\max_j\alpha_j-1}\vartheta(p')+\frac{1}{2\varrho}\vartheta(p),\end{aligned}$$

where the second transition holds because $\varrho\ge 1$ and $\max_j\alpha_j>1$. Notice that

$$\vartheta(p)\le\sum_{e\in p}\left[\sigma_e+\sum_{j\in[q]}\xi_{e,j}\left(l_e^p\right)^{\alpha_j}\right]=C(p),\ \text{and}\ \vartheta(p')\le\sum_{e\in p'}\left[\sigma_e+\sum_{j\in[q]}\xi_{e,j}\left(l_e^{p'}\right)^{\alpha_j}\right]=C(p'),$$

which establishes the assertion. ■ (Claim 6.3)

Together, Claims 6.2 and 6.3 imply that

$$\sum_{i\in[N]} C_i(p_i',p_{-i}) \le \left(\gamma_\alpha+\lambda_\alpha\varrho^{\max_j\alpha_j-1}\right)C(p')+C(p)/(2\varrho),$$

so (3) indeed holds with $\lambda=\gamma_\alpha+\lambda_\alpha\varrho^{\max_j\alpha_j-1}$ and $\mu=1/(2\varrho)$. □

Since $1/(2\varrho)<1/(\varrho\epsilon_1^2)$ for sufficiently small $\epsilon>0$, it follows that we can employ Theorem 5.3 with the smoothness parameters $\lambda=\gamma_\alpha+\lambda_\alpha\cdot\varrho^{\max_j\alpha_j-1}$ and $\mu=1/(2\varrho)$ guaranteed by Theorem 6.1 to obtain the following corollary.

Corollary 6.4. *If M is an REP-expanded CSM, then the approximation ratio of* `Alg-ABRD` *is* $O\left(\varrho^{\max_j\alpha_j}+\varrho\cdot\max_e\min_j\left(\frac{\sigma_e}{\xi_{e,j}}\right)^{\frac{1}{\alpha_j}}\right)$.

We now turn to show that some natural and extensively studied CSMs are REP-expanded. Under the *proportional fair CSM* (see, e.g., [26, 32]), the cost share of player $i\in[N]$ in edge $e\in p_i$ is defined to be her share of the cost incurred by load l_e^p on edge e, proportional to her weight $w_i(e)$, namely $f_{i,e}(p)=\frac{w_i(e)}{l_e^p}F_e(l_e^p)$.

Lemma 6.5. *The proportional fair CSM is REP-expanded.*

Proof. Under the proportional fair CSM, the cost share of player i in edge e satisfies

$$\frac{w_i(e)}{l_e^p}\left[\sigma_e + \sum_{j=1}^{q} \xi_{e,j}(l_e^p)^{\alpha_j}\right] \leq \sigma_e + \sum_{j=1}^{q} \xi_{e,j} w_i(e) \cdot (l_e^p)^{\alpha_j - 1}$$
$$= \sigma_e + \sum_{j=1}^{q} \xi_{e,j} w_i(e) \cdot \left[(l_e^p - w_i(e)) + w_i(e)\right]^{\alpha_j - 1},$$

where the inequality holds because $l_e^p = \sum_{i':e\in p_{i'}} w_{i',e} \geq w_i(e)$. Consider the following two cases.

- $0 \leq l_e^p - w_i(e) \leq w_i(e)$: Since $\alpha_j - 1 > 0$, it follows that

$$\sigma_e + \sum_j \xi_{e,j} w_i(e) \cdot \left[(l_e^p - w_i(e)) + w_i(e)\right]^{\alpha_j - 1} \leq \sigma_e + \sum_j \xi_{e,j} w_i(e) \cdot (2 \cdot w_i(e))^{\alpha_j - 1}$$
$$= \sigma_e + \sum_j \xi_{e,j} 2^{\alpha_j - 1} (w_i(e))^{\alpha_j}.$$

- $l_e^p - w_i(e) > w_i(e)$: In this case,

$$\sigma_e + \sum_j \xi_{e,j} w_i(e) \cdot \left[(l_e^p - w_i(e)) + w_i(e)\right]^{\alpha_j - 1} < \sigma_e + \sum_j \xi_{e,j} 2^{\alpha_j - 1} w_i(e)(l_e^p - w_i(e))^{\alpha_j - 1}.$$

The assertion follows by taking $z_{1,j} = 2^{\alpha_j - 1}$, and $z_{2,j} = 2^{\alpha_j - 1}$ for every $j \in [q]$. □

Let $S_e = \{i \in [N] \mid e \in p_i\}$ and let π_e be a random permutation of S_e drawn from the uniform distribution. Under the *Shapley CSM* (see, e.g., [26, 32]), the cost share of player $i \in [N]$ in edge $e \in p_i$ is defined to be its expected marginal contribution if the players are added to e one-by-one in π_e order. More formally, taking $S_e^i(\pi_e) \subseteq S_e$ to denote the set of players that precede player i in π_e, the cost share of i in edge e under the Shapley CSM is defined to be $f_{i,e}(p) = \mathbb{E}\left[F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e} + w_i(e)\right) - F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e}\right)\right]$ if $e \in p_i$, otherwise $f_{i,e}(p) = 0$.

Before proving that the Shaply CSM is REP-expanded, we first define a function $h_e : 2^{[N]} \to \mathbb{R}_{\geq 0}$ by setting

$$h_e(X) = \sum_{j\in[q]} \xi_{e,j} \cdot \left(\sum_{i\in X} w_i(e)\right)^{\alpha_j}. \tag{8}$$

Lemma 6.6. *The function $h_e(X)$ is supermodular.*

Proof. By definition, $h_e(X)$ is supermodular if for any $X_1 \subset X_2 \subset X$ and $i' \in X \backslash X_2$,

$$\sum_{j=1}^{q} \xi_{e,j}\left[\Big(\sum_{i\in X_1\cup\{i'\}} w_i(e)\Big)^{\alpha_j} - \Big(\sum_{i\in X_1} w_i(e)\Big)^{\alpha_j}\right] \leq \sum_{j=1}^{q} \xi_{e,j}\left[\Big(\sum_{i\in X_2\cup\{i'\}} w_i(e)\Big)^{\alpha_j} - \Big(\sum_{i\in X_2} w_i(e)\Big)^{\alpha_j}\right].$$

Therefore, it suffices to prove that for any three non-negative numbers x_1, x_2 and y, $(x_1 + y)^{\alpha_j} - (x_1)^{\alpha_j} \leq (x_1 + x_2 + y)^{\alpha_j} - (x_1 + x_2)^{\alpha_j}$ holds for any $\alpha_j > 1$. Let $v(x_1, x_2, y) = (x_1 + x_2 + y)^{\alpha_j} - (x_1 + x_2)^{\alpha_j}$. Then we have

$$\frac{\partial v}{\partial x_2} = \alpha_j\left[(x_1 + x_2 + y)^{\alpha_j - 1} - (x_1 + x_2)^{\alpha_j - 1}\right].$$

Since $\alpha_j > 1$, $\frac{\partial v}{\partial x_2} > 0$ when $x_2 > 0$. Therefore, $v(x_1, x_2, y) \geq v(x_1, 0, y) = (x_1 + y)^{\alpha_j} - (x_1)^{\alpha_j}$. □

Lemma 6.7. *The Shapley CSM is REP-expanded.*

Proof. Under the Shapley CSM, the cost share of player $i \in [N]$ in edge $e \in p_i$ is

$$\begin{aligned} f_{i,e}(p) &\leq \mathbb{E}\left[\sigma_e + \sum_j \xi_{e,j}\left(\sum_{i' \in S_e^i(\pi_e)} w_{i',e} + w_i(e)\right)^{\alpha_j} - \sum_j \xi_{e,j}\left(\sum_{i' \in S_e^i(\pi_e)} w_{i',e}\right)^{\alpha_j}\right] \\ &\leq \sigma_e + \sum_j \xi_{e,j}\left(\sum_{i' \in (S_e - \{i\})} w_{i',e} + w_i(e)\right)^{\alpha_j} - \sum_j \xi_{e,j}\left(\sum_{i' \in (S_e - \{i\})} w_{i',e}\right)^{\alpha_j} \\ &= \sigma_e + \sum_j \xi_{e,j}(l_e^p)^{\alpha_j} - \sum_j \xi_{e,j}(l_e^p - w_i(e))^{\alpha_j}\,, \end{aligned}$$

where the first transition holds since $\sum_j \xi_{e,j} l^{\alpha_j} \leq F_e(l) \leq \sigma_e + \sum_j \xi_{e,j} l^{\alpha_j}$ for any $l \geq 0$ and the second transition follows from Lemma 6.6. Consider the following two cases.

- $l_e^p \leq 3 \cdot w_i(e)$: In this case,

$$\sigma_e + \sum_j \xi_{e,j}(l_e^p)^{\alpha_j} - \sum_j \xi_{e,j}(l_e^p - w_i(e))^{\alpha_j} \;\leq\; \sigma_e + \sum_j \xi_{e,j} \cdot 3^{\alpha_j}(w_i(e))^{\alpha_j}\,.$$

- $l_e^p > 3 \cdot w_i(e)$: By Newton's generalized binomial theorem, we have

$$\begin{aligned} &\sigma_e + \sum_j \xi_{e,j}(l_e^p)^{\alpha_j} - \sum_j \xi_{e,j}(l_e^p - w_i(e))^{\alpha_j} \\ &\leq \sigma_e + \sum_j \xi_{e,j} \sum_{k=0}^{\infty} \binom{\alpha_j}{k} (l_e^p - w_i(e))^{\alpha_j - k}(w_i(e))^k - \sum_j \xi_{e,j}(l_e^p - w_i(e))^{\alpha_j} \\ &= \sigma_e + \sum_j \xi_{e,j} \sum_{k=1}^{\infty} \binom{\alpha_j}{k} (l_e^p - w_i(e))^{\alpha_j - k}(w_i(e))^k \\ &< \sigma_e + \sum_j \xi_{e,j} \binom{\alpha_j}{\lfloor \frac{\alpha_j+1}{2} \rfloor} w_i(e)(l_e^p - w_i(e))^{\alpha_j - 1} \sum_{k=1}^{\infty} \left(\frac{w_i(e)}{l_e^p - w_i(e)}\right)^{k-1} \\ &< \sigma_e + \sum_j \xi_{e,j} \binom{\alpha_j}{\lfloor \frac{\alpha_j+1}{2} \rfloor} w_i(e)(l_e^p - w_i(e))^{\alpha_j - 1} \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} \\ &< \sigma_e + 2 \sum_j \xi_{e,j} \binom{\alpha_j}{\lfloor \frac{\alpha_j+1}{2} \rfloor} w_i(e)(l_e^p - w_i(e))^{\alpha_j - 1}\,, \end{aligned}$$

  where the third transition holds because for any $\alpha > 1$ and $k \in \mathbb{Z}^+$, the absolute value of $\binom{\alpha}{k}$ is at most $\binom{\alpha}{\lfloor \frac{\alpha+1}{2} \rfloor}$.

The assertion follows by taking $z_{1,j} = 3^{\alpha_j}$, and $z_{2,j} = 2\binom{\alpha_j}{\lfloor \frac{\alpha_j+1}{2} \rfloor}$ for every $j \in [q]$. □

7 THE POTENTIAL FUNCTION OF THE SHAPLEY COST SHARING MECHANISM

The next step is to prove that among the REP-expanded CSMs, there exists one that induces a GND game with an (A, B)-bounded potential function for sufficiently small $A, B \geq 1$. (Recall that by Theorem 5.3, this would provide an upper bound on the number of steps in the ABRD.) While we could not accomplish this task for the proportional fair CSM, the Shapley CSM turned out to be more successful.

It can be inferred from [44, Proposition 2.1] (see also [28, 32]) that the GND game induced by the Shapley CSM admits the potential function

$$\Phi(p) \;=\; \sum_{e\in p}\sum_{i\in S_e} f_{i,e}(S_e^i(\psi_e)\cup\{i\})\,, \tag{9}$$

where $S_e = \{j\in[N] \mid e\in p_j\}$, ψ_e is an arbitrary permutation of S_e, and $S_e^i(\psi_e)$ is the set of players that precede i in ψ_e. Note that in contrast to the random permutation π_e used in the definition of the Shapley CSM, the permutation ψ_e is an (arbitrary) deterministic permutation. The rest of this section is dedicated to proving that this potential function is $(\mathcal{H}_N, \lceil\max_j \alpha_j\rceil)$-bounded, where $\mathcal{H}_N$ is the N-th harmonic number. The following result is based on the function $h_e : 2^{[N]}\to\mathbb{R}_{\geq 0}$ defined in Eq. (8).

Lemma 7.1. *For any edge e and any permutation ψ_e, we have*

$$\sum_{i\in S_e} f_{i,e}(S_e^i(\psi_e)\cup\{i\}) \;=\; \sum_{k=1}^{|S_e|}\left(\frac{\sigma_e}{k} + \sum_{T\subseteq S_e,|T|=k}\frac{h_e(T)}{\binom{|S_e|}{k}\cdot k}\right).$$

Proof. By the definition of the Shapley CSM, the cost share of player i who uses edge e is

$$f_{i,e}(S_e) \;=\; \mathbb{E}\left[F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e} + w_i(e)\right) - F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e}\right)\right],$$

where π_e is a random permutation on S_e. For a fixed π_e and a fixed player i using edge e,

$$F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e} + w_i(e)\right) - F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e}\right) = \begin{cases}\sigma_e + h_e(\{i\}) - h_e(\emptyset)\,, & \text{if } S_e^i(\pi_e)=\emptyset\\ h_e(S_e^i(\pi_e)\cup\{i\}) - h_e(S_e^i(\pi_e))\,, & \text{otherwise}\end{cases}$$
$$= \mathbb{1}(S_e^i(\pi_e)=\emptyset)\sigma_e + h_e(S_e^i(\pi_e)\cup\{i\}) - h_e(S_e^i(\pi_e))\,,$$

where $\mathbb{1}(S_e^i(\pi_e)=\emptyset)$ denotes the indicator of the event $S_e^i(\pi_e)=\emptyset$. Since π_e is taken from the uniform distribution, it follows that $\mathbb{P}(S_e^i(\pi_e)=\emptyset) = \frac{1}{|S_e|}$ for any player i using edge e, thus

$$\mathbb{E}\left[F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e} + w_i(e)\right) - F_e\left(\sum_{i'\in S_e^i(\pi_e)} w_{i',e}\right)\right] \;=\; \frac{\sigma_e}{|S_e|} + \mathbb{E}\left[h_e(S_e^i(\pi_e)\cup\{i\}) - h_e(S_e^i(\pi_e))\right]. \tag{10}$$

Let $H_{i,e}(S_e) = \mathbb{E}[h_e(S_e^i(\pi_e)\cup\{i\}) - h_e(S_e^i(\pi_e))]$. Then, it can be inferred from Eq. (10) that for any player i using e,

$$f_{i,e}(S_e^i(\psi_e)\cup\{i\}) \;=\; \frac{\sigma_e}{|S_e^i(\psi_e)\cup\{i\}|} + H_{i,e}\left(S_e^i(\psi_e)\cup\{i\}\right).$$

Since $\sum_{i\in S_e}\frac{\sigma_e}{|S_e^i(\psi_e)\cup\{i\}|} = \sum_{k=1}^{|S_e|}\frac{\sigma_e}{k}$, it follows that

$$\sum_{i\in S_e} f_{i,e}(S_e^i(\psi_e)\cup\{i\}) \;=\; \sum_{k=1}^{|S_e|}\frac{\sigma_e}{k} + \sum_{i\in S_e} H_{i,e}\left(S_e^i(\psi_e)\cup\{i\}\right).$$

Notice that $H_{i,e}(S_e)$ can be viewed as the Shapley cost share of a player i who uses edge e in a network game where the cost of each edge e is $h_e(S_e)$. Since for every j, α_j is assumed to be larger

than 1, the ratio $\frac{h(S_e)}{l_e} = \frac{\sum_j \xi_{e,j}(l_e)^{\alpha_j}}{l_e}$ is non-decreasing with l_e. Kollias and Roughgarden [32, Proof of Proposition 3.2] prove that in such case, for any ψ_e,

$$\sum_{i \in S_e} H_{i,e}\left(S_e^i(\psi_e) \cup \{i\}\right) \;=\; \sum_{T \subseteq S_e, |T|=k} \frac{h_e(T)}{\binom{|S_e|}{k} \cdot k},$$

thus establishing the assertion. □

We are now ready to prove that the potential function $\Phi(p)$ of (9) is $(\lceil \max_j \alpha_j \rceil, \mathcal{H}_N)$-bounded.

Theorem 7.2. *The potential function $\Phi(p)$ of the GND game induced by the Shapley CSM satisfies $\frac{1}{\lceil \max_j \alpha_j \rceil} \cdot C(p) \le \Phi(p) \le \mathcal{H}_N \cdot C(p)$ for any strategy profile p.*

Proof. Let us first prove the lower bound on $\Phi(p)$. Since $e \in p$ implies that $|S_e| \ge 1$, we get

$$\Phi(p) \;=\; \sum_{e \in p} \sum_{k=1}^{|S_e|} \left(\frac{\sigma_e}{k} + \sum_{T \subseteq S_e, |T|=k} \frac{h_e(T)}{\binom{|S_e|}{k} \cdot k} \right) \ge \sum_{e \in p} \left(\sigma_e + \sum_{k=1}^{|S_e|} \sum_{T \subseteq S_e, |T|=k} \frac{h_e(T)}{\binom{|S_e|}{k} \cdot k} \right).$$

By the convexity of $\xi_{e,j} \cdot x^{\alpha_j}$, we conclude that

$$\Phi(p) \;\ge\; \sum_{e \in p} \left[\sigma_e + \sum_{j=1}^{q} \xi_{e,j} \sum_{k=1}^{|S_e|} \frac{1}{\binom{|S_e|}{k} \cdot k} \binom{|S_e|}{k} \left(\frac{\sum_{T \subseteq S_e, |T|=k} \sum_{i \in T} w_i(e)}{\binom{|S_e|}{k}} \right)^{\alpha_j} \right].$$

Since every player i is included in exactly $\binom{|S_e|-1}{k-1}$ subsets of S_e with k elements, it follows that

$$\begin{aligned}
\Phi(p) \;&\ge\; \sum_{e \in p} \left[\sigma_e + \sum_{j=1}^{q} \xi_{e,j} \sum_{k=1}^{|S_e|} \frac{1}{k} \left(\frac{\binom{|S_e|-1}{k-1} l_e^p}{\binom{|S_e|}{k}} \right)^{\alpha_j} \right] \\
&\ge\; \sum_{e \in p} \left[\sigma_e + \sum_{j=1}^{q} \xi_{e,j} \left(\frac{l_e^p}{|S_e|} \right)^{\alpha_j} \sum_{k=1}^{|S_e|} k^{\alpha_j - 1} \right] \\
&\ge\; \sum_{e \in p} \left[\sigma_e + \sum_{j=1}^{q} \xi_{e,j} \left(\frac{l_e^p}{|S_e|} \right)^{\alpha_j} \left(\frac{1}{|S_e|} \right)^{\lceil \alpha_j \rceil - \alpha_j} \sum_{k=1}^{|S_e|} k^{\lceil \alpha_j \rceil - 1} \right].
\end{aligned}$$

Then, we can derive from [8, Corollary 3.2] that

$$\begin{aligned}
\Phi(p) \;&\ge\; \sum_{e \in p} \left[\sigma_e + \sum_{j=1}^{q} \xi_{e,j} \left(\frac{l_e^p}{|S_e|} \right)^{\alpha_j} \left(\frac{1}{|S_e|} \right)^{\lceil \alpha_j \rceil - \alpha_j} \frac{|S_e|^{\lceil \alpha_j \rceil}}{\lceil \alpha_j \rceil} \right] \\
&\ge\; \frac{1}{\lceil \max_j \alpha_j \rceil} \sum_{e \in p} [\sigma_e + \sum_j \xi_{e,j} (l_e^p)^{\alpha_j}].
\end{aligned}$$

For the upper bound on $\Phi(p)$, by Lemma 7.1 and since $h_e(T)$ is a (set-wise) increasing function of T and $T \subseteq S_e$, we get

$$\Phi(p) \;=\; \sum_{e \in p} \sum_{k=1}^{|S_e|} \left(\frac{\sigma_e}{k} + \sum_{T \subseteq S_e, |T|=k} \frac{h_e(T)}{\binom{|S_e|}{k} \cdot k} \right) \le \sum_{e \in p} \left(\mathcal{H}_{|S_e|} \cdot \sigma_e + \sum_{k=1}^{|S_e|} \sum_{T \subseteq S_e, |T|=k} \frac{h_e(S_e)}{\binom{|S_e|}{k} \cdot k} \right).$$

As there are exactly $\binom{|S_e|}{k}$ subsets of S_e with k elements, it follows that

$$\begin{aligned}\Phi(p) &\leq \sum_{e\in p}\left(\mathcal{H}_{|S_e|}\cdot\sigma_e + \sum_{k=1}^{|S_e|}\frac{h_e(S_e)}{k}\right)\\ &= \sum_{e\in p}\left(\mathcal{H}_{|S_e|}\cdot\sigma_e + \mathcal{H}_{|S_e|}\cdot h_e(S_e)\right)\\ &= \sum_{e\in p}\mathcal{H}_{|S_e|}\cdot\left[\sigma_e + \sum_{j\in[q]}\xi_{e,j}\left(\sum_{i\in S_e} w_i(e)\right)^{\alpha_j}\right].\end{aligned}$$

Since $e\in p$ means that $l_e^p > 0$, we conclude that

$$\Phi(p) \leq \sum_{e\in p}\mathcal{H}_{|S_e|}F_e(l_e^p) \leq \mathcal{H}_N\cdot C(p)$$

which establishes the assertion. □

8 POLYNOMIAL-TIME ϵ-APPROXIMATION OF SHAPLEY COST SHARING

So far, we have proved that the Shapley CSM can satisfy the requirements on the smoothness and the potential function. It remains to show how to compute the ϵ-cost shares subject to the Shapley CSM in polynomial time for a sufficiently small $\epsilon > 0$. For the problem of computing the cost shares specified by the Shapley CSM, Liben-Nowell et al. [36] establish the following lemma.

Lemma 8.1 ([36]). *There exists an FPRAS (i.e., a randomized FPTAS), referred to as* `SV-Sample`*, for computing the ε-cost shares in any game subject to the Shapley CSM and* supermodular *monotone cost functions. In particular, given any $\varepsilon\in(0,1)$,* `SV-Sample` *generates an ε-cost share with probability at least $1-\frac{1}{2(TN|E|)^2}$ in $O\Big(\log(TN|E|)\cdot\Big[\frac{N^3}{\varepsilon^2}+\log(\log(TN|E|))\Big]\Big)$-time.*

Note that owing to the existence of the term σ_e, the cost function $F_e(l)$ is not supermodular. Now Lemma 6.6 comes to our help. Combining Lemma 8.1 and Lemma 6.6 with Eq. (10), we obtain an efficient procedure, named `Shapley-APX`, for computing the ϵ-cost share of any given player i on any resource e with respect to the Shapley CSM and the REP cost function.

More specifically, if $i\notin S_e$, then this procedure returns 0 as the cost share. Otherwise, `Shapley-APX` uses algorithm `SV-Sample` to obtain an ϵ-cost share $\theta_{i,e}$ for player i on resource e with respect to the Shapley CSM and the cost function $h_e(S_e)$. Finally, $\frac{\sigma_e}{|S_e|}+\theta_{i,e}$ is returned as the desired ϵ-cost share. By Eq. (10), Lemma 8.1 and Lemma 6.6, the following lemma trivially holds.

Lemma 8.2. *Procedure* `Shapley-APX` *computes an ϵ-cost share of a player i on resource e with probability at least $1-\frac{1}{2(TN|E|)^2}$ in $O\Big(\log(TN|E|)\cdot\Big[\frac{N^3}{\epsilon^2}+\log(\log(TN|E|))\Big]\Big)$-time.*

Theorem 8.3. *If* `Shapley-APX` *is employed to generate all the ϵ-cost shares used in* `Alg-ABRD`*, then w.h.p., every ϵ-cost share $\widetilde{f}_{i,e}(S_e)$ is an ϵ-approximation of the exact cost share $f_{i,e}(S_e)$.*

Proof. Recall that the ABRD contains at most T steps. In each step t, every player i needs to calculate $\widetilde{f}_{i,e}(S_e^t\cup\{i\})$ for every resource e to find her ABR. Therefore, procedure `Shapley-APX` is invoked at most $TN|E|$ times. The probability that this function generates a result that is not the ϵ-approximation of the exact cost share is at most $TN|E|\cdot\left[1-\left(1-\frac{1}{2(TN|E|)^2}\right)\right]=\frac{1}{2TN|E|}$. Therefore, this theorem follows. □

Using the facts that ϵ, λ_α (from Theorem 6.1), and $\varrho\epsilon_1^2/2$ are all constants, and $\mathcal{H}_N$ can be bounded by $O(\log N)$, the main result is established as summed up in the following theorem.

Theorem 8.4. *By plugging the Shapley CSM into* Alg-ABRD*, the total cost of the output profile is an* $O\left(\varrho \cdot \max_e \min_j \left(\frac{\sigma_e}{\xi_{e,j}}\right)^{\frac{1}{\alpha_j}} + \varrho^{\max_j \alpha_j}\right)$*-approximation of the optimal result with probability at least*

$$1 - O\left(\frac{1}{N^2|E|\log^2 N}\right).$$

The time complexity of the algorithm is

$$O\left(N^5 \log^2 N \cdot |E| \log^2(N|E|)\right).$$

9 IMPLEMENTATION IN A DECENTRALIZED ENVIRONMENT

The approximation algorithm Alg-ABRD that was developed up to now is centralized, and in particular two main aspects of the algorithm are incompatible with some common settings in game theory. The first aspect is that Alg-ABRD deterministically chooses a specific player for strategy update. Instead, if traffic requests were separate uncoordinated entities, it would make more sense that they decide to update their strategies in an *uncoordinated* way. The second aspect is that Alg-ABRD chooses the best profile it has seen during the ABRD. However, it is inappropriate in game theory to ask uncoordinated individual entities to "roll back" to a previous profile that might be more costly for some of them.

This section tackles these issues by providing two techniques for adapting algorithm Alg-ABRD to the game-theoretic settings. First, instead of choosing a specific player for updating the strategy, we now select the player uniformly at random. We believe that this better simulates the behaviors of uncoordinated players. Subsection 9.1 shows that this modification will still yield the same approximation ratio, with only a polynomial loss in the number of steps. Second, instead of choosing the best configuration in the sequence, subsection 9.2 analyzes the case where the last configuration is chosen. It is shown that the approximation ratio loses another $O(\log N)$ factor. Thus, while certainly inferior to the centralized algorithm, the game-theoretic version of Alg-ABRD still admits an approximately optimal outcome.

9.1 Randomized Selection and Decentralized Implementation

This subsection develops a random procedure, called *randomized player choosing* (RPC), for deciding the player to update her strategy in an uncoordinated way, using some techniques in the *leader election* protocol proposed in [1, 43].

Consider an arbitrary step $t \geq 1$ in the ABRD. We assume that all the players have the same view of p^{t-1}. Notice that this assumption trivially holds when $t = 1$, because p^0 is generated in a deterministic way, which means that every player i can easily simulate the computation for p^0_{-i} without any communication between players. Given the strategy profile p^{t-1}, each player i first finds her ABR p'_i to p^{t-1}_{-i} by the procedure promised in Lemma 4.1, and computes the value of δ^t_i (recall that $\delta^t_i = \widetilde{C}_i(p^{t-1}) - \epsilon_1 \cdot \widetilde{C}_i(p'_i, p^{t-1}_{-i})$).

Then, every player i generates an integer $Y^t_i \in [N]$ randomly and uniformly, and sends Y^t_i to all the other players. After receiving all the $N-1$ integers $\{Y^t_{i'}\}_{i' \in N \wedge i' \neq i}$, every player i calculates $Y^t = \left(\left(\sum_{i \in [N]} Y^t_i\right) \bmod N\right) + 1$, where mod refers to the modulo operator. It is easy to see that Y^t

follows the uniform distribution over $[N]$. If the player with index Y^t satisfies

$$\delta_i^t > 0, \tag{11}$$

then this player deviates to her ABR p'_{Y^t}, and send p'_{Y^t} to all the other players. Otherwise, player Y^t simply sends her current strategy $p_{Y^t}^{t-1}$ to the other players. After receiving the strategy from the chosen player Y^t, all the other players update p^{t-1} to p^t with the received strategy. In this way, it is guaranteed that in step $t+1$, all the players have the same view of p^t.

Lemma 9.1. *For any step $t \geq 1$, if the player selected for strategy update satisfies Eq. (11), we have $\Phi(p^{t-1}) - \Phi(p^t) > 0$.*

Proof. By the definition of the potential function,

$$\begin{aligned}
\Phi(p^{t-1}) - \Phi(p^t) &= C_j(p^{t-1}) - C_j(p_j^t, p_{-j}^{t-1}) \\
&\geq \frac{1}{1+\epsilon}\widetilde{C}_j(p^{t-1}) - \frac{1}{1-\epsilon}\widetilde{C}_j(p_j^t, p_{-j}^{t-1}) \\
&> \frac{1}{1+\epsilon}\epsilon_1\widetilde{C}_j(p_j^t, p_{-j}^{t-1}) - \frac{1}{1-\epsilon}\widetilde{C}_j(p_j^t, p_{-j}^{t-1}) \\
&= 0 .
\end{aligned}$$

The second transition holds by the definition of the ϵ-individual cost. The third one follows from Eq. (11). □

Let RPC-ABRD be a variation of Alg-ABRD that uses RPC to decide the player for updating the strategy, and runs in $T' = N \cdot T^2$ steps. Then we have the following result.

Theorem 9.2. *The output p^{t^*} of* RPC-ABRD *satisfies*

$$C(p^{t^*}) \leq \frac{2\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu} \cdot C^*$$

with probability at least $\frac{1}{2}$,

Proof. According to Lemma 5.1, we only need to consider the case where the ABRD does not converge. Partition the T' steps in ABRD into T *stages*, each of which contains $N \cdot T$ steps. We say a player i is *appropriate* for step t if δ_j^t satisfies Eq. (4). A step t is said to be appropriate if in this step an appropriate player is selected, and a stage is appropriate if it contains at least one appropriate step.

Claim 9.3. *With probability at least $\frac{1}{2}$, all the T stages are appropriate.*

Proof. If the ABRD does not converge, then the Pigeonhole Principle implies that there exists at least one appropriate player in each step. Procedure RPC ensures that each step is appropriate with probability at least $\frac{1}{N}$. Therefore, the probability that there exists no appropriate step in a stage should be at most

$$\left(1 - \frac{1}{N}\right)^{NT} \leq \left(\frac{1}{\exp(1)}\right)^T .$$

Hence, the probability that all the stages are appropriate is

$$\left(1 - \left(1 - \frac{1}{N}\right)^{NT}\right)^T \geq \left(1 - \left(\frac{1}{\exp(1)}\right)^T\right)^T > \left(1 - \left(\frac{1}{2}\right)^T\right)^T > \frac{1}{2} . \tag{12}$$

The last transition holds because $\left(\frac{1}{2}\right)^T + \left(\frac{1}{2}\right)^{1/T}$ decreases with T when $T > 1$. ■ (Claim 9.3)

Recall that a strategy profile p^t is said to be bad if $C(p^t) > \frac{2\varrho\epsilon_1^2\lambda}{1-\varrho\epsilon_1^2\mu} \cdot C^*$. Then we have the following proposition.

Claim 9.4. *For any appropriate stage $\Bbbk$, if all the profiles generated in this stage are bad, then $\Phi(p^{t^1_\Bbbk}) < \left(1 - \frac{1}{Q}\right)\Phi(p^{t^0_\Bbbk - 1})$, where $t^0_\Bbbk$ an $t^1_\Bbbk$ respectively represent the first and the last steps in stage $\Bbbk$.*

Proof. Since the player selected in each step does not update its strategy unless Eq. (11) is satisfied, Lemma 9.1 indicates that the potential function is non-increasing in all the steps. Let $t^*_\Bbbk$ be an arbitrary appropriate step in stage $\Bbbk$. Then we have

$$\Phi(p^{t^1_\Bbbk}) \le \Phi(p^{t^*_\Bbbk}) < \left(1 - \frac{1}{Q}\right)\Phi(p^{t^*_\Bbbk - 1}) \le \left(1 - \frac{1}{Q}\right)\Phi(p^{t^0_\Bbbk - 1}),$$

where the second transition follows from Claim 5.4. ■ (Claim 9.4)

Combining Claim 9.4 with the techniques in the proof of Theorem 5.3, it can be proved that if all the stages are appropriate, then at least one stage generates a profile that is not bad. □

Putting Theorem 9.2, Theorem 6.1, Theorem 7.2 and Theorem 8.3 together, we obtain the following result.

Corollary 9.5. RPC-ABRD *generates an* $O\left(\varrho^{\max_j \alpha_j} + \varrho \cdot \max_e \min_j \left(\frac{\sigma_e}{\xi_{e,j}}\right)^{1/\alpha_j}\right)$*-approximation solution with probability at least*

$$\frac{1}{2}\left(1 - O\left(\frac{1}{N^2|E|\log^2 N}\right)\right).$$

The time complexity of the revised algorithm is

$$O\left(N^7 \log^4 N \cdot |E| \log^2(N|E|)\right).$$

9.2 Output the Last Strategy Profile

Now let us study the approach of directly outputting the last strategy instead of the one with the minimum overall cost, p^{t^*}.

Theorem 9.6. *Suppose that the ABRD does not converge at any step t, then the cost corresponding to the last profile is at most $\lceil \max_j \alpha_j \rceil \mathcal{H}_N$ times larger than $C(p^{t^*})$, no matter whether* RPC *is employed or not.*

Proof. Let $t_{\max}$ be the last step of the ABRD. This theorem trivially holds when $t^* = t_{\max}$. Suppose that $t^* < t_{\max}$, then

$$C(p^{t_{\max}}) \le \lceil \max_j \alpha_j \rceil \Phi(p^{t_{\max}}) < \lceil \max_j \alpha_j \rceil \Phi(p^{t^*}) \le \lceil \max_j \alpha_j \rceil \mathcal{H}_N C(p^{t^*}).$$

The first transition and the last one hold since the potential function is $(\mathcal{H}_N, \lceil \alpha \rceil)$-bounded. The second transition follows from Eq. (4), Eq. (11) and Lemma 9.1, for the deterministic procedure of deciding the player for strategy update and RPC, respectively. □

Using the fact that $\mathcal{H}_N$ is bounded by $O(\log N)$, we get the following result.

Corollary 9.7. *Returning the strategy profile generated in the last step of the ABRD instead of p^{t^*} as the output will increase the approximation ratio by $O(\log N)$ times.*

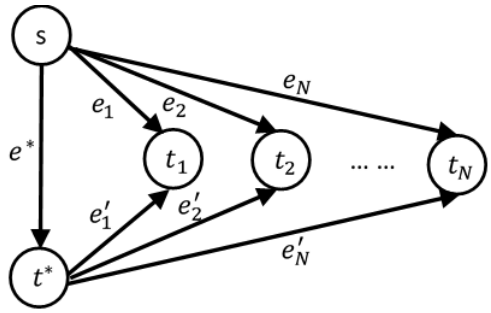

Fig. 1. A directed graph G_0 for proving the lower bound on the PoA.

10 POA OF THE GND GAME: UPPER BOUND AND LOWER BOUND

A byproduct obtained in this paper is a tight bound on the PoA of the GND games for a class of CSMs. In [43], it is proved that the PoA of a smooth cost minimization game is

$$\inf\left\{\frac{\lambda}{1-\mu}:\ \text{the game is } (\lambda,\mu)\text{-smooth}\right\}.$$

Such a PoA is said to be *robust* and can be extended to the mixed Nash equilibrium, correlated equilibrium, and coarse correlated equilibrium [43]. From Theorem 6.1, it can be inferred that:

Theorem 10.1. *For any REP-expanded CSM M, the induced GND game has a* robust *PoA of* $O\left(\max_{e\in E}\min_{j\in[q]}\left(\frac{\sigma_e}{\xi_{e,j}}\right)^{\frac{1}{\alpha_j}}\right)$.

Recall that our definition of CSMs requires that they are budget-balanced, namely, that

$$\sum_{i\in S_e} f_{i,e}(p) = F_e(l_e^p) \tag{13}$$

for any edge e. In the following, we prove based on this requirement that the upper bound in Theorem 10.1 is asymptotically tight. To that end, we restrict our attention to the GND problem with routing requests.

Definition 10.2 (GND problem with Routing Requests). In the GND problem with routing requests, the resources are represented by the set E of edges in a graph $G = (V, E)$, where V is the set of nodes. The feasible reply collection P_i of each request i is composed of the paths which connect the associated source-target node pair and contain no repeating edges.

Theorem 10.3. *For any (budget-balanced) CSM, there exists induced GND games with a PoA of* $\Omega\left(\max_e \min_j\left(\frac{\sigma_e}{\xi_{e,j}}\right)^{\frac{1}{\alpha_j}}\right)$.

Proof. Let $\mathcal{I}$ be an instance of a GND problem with routing requests defined on the directed graph $G_0 = (V_0, E_0)$ in Fig. 1. In particular, V_0 contains $N+2$ nodes $\{s, t^*, \{t_i\}_{i\in[N]}\}$. For each $i \in [N]$, there are a directed edge e_i from s to t_i, and a directed edge e'_i from t^* to t_i. For each request $i \in [N]$, the associated source-target pair is (s, t_i), and the weight $w_i(e) = 1$ for any $e \in E_0$. It is easy to see that the reply collection P_i of each request i contains two paths, $\{e^*, e'_i\}$ and $\{e_i\}$.

Define σ, ξ, and α to be three positive parameters such that $\left(\frac{\sigma}{\xi}\right)^{\frac{1}{\alpha}}$ is a large enough integer, and $\alpha > 1$. The number of requests N is assumed to be $\left(\frac{\sigma}{\xi}\right)^{\frac{1}{\alpha}}$. For the edges $e \in E_0$, the parameters in the associated REP cost functions are set as follows.

- $\sigma_{e^*} = \frac{N}{N+1} \cdot \sigma$, and $\xi_{e^*,1} = \frac{N}{N+1} \cdot \xi$.
- For every $i \in [N]$, $\sigma_{e_i} = \sigma$, and $\xi_{e_i,1} = \xi$.
- For every $i \in [N]$, $\sigma_{e'_i} = \frac{1}{N+1} \cdot \sigma$, and $\xi_{e'_i,1} = \frac{3}{N+1} \cdot \xi$.
- $\alpha_1 = \alpha$.
- For every $2 \le j \le q$, $1 < \alpha_j < \alpha_1$, $\xi_{e,j} < \frac{\xi_{e,1}}{qN^{\alpha_j}(N+1)}$.

With these settings, we have $\max_{e \in E_0} \min_{j \in [q]} \left(\frac{\sigma_e}{\xi_{e,j}}\right)^{\frac{1}{\alpha_j}} = \left(\frac{\sigma}{\xi}\right)^{\frac{1}{\alpha}}$.

Consider a GND game induced from $\mathcal{I}$ by an arbitrary budget-balanced CSM. Let p be a strategy profile where every player i chooses the path $p_i = \{e_i\}$. With any budget-balanced CSM, the cost of each player i must be

$$f_{i,e_i}(p) \;=\; F_{i,e_i}(1) \;=\; \sigma_{e_i} + \sum_{j=1}^{q} \xi_{e_i,j} \;<\; \sigma + \xi \cdot \frac{N+2}{N+1}\,.$$

The first transition follows from Eq. (13), since S_{e_i} only contains player i for every $i \in [N]$. The last transition holds since $N^{\alpha_j} > 1$ for every j. If any player i changes her strategy to $p'_i = \{e^*, e'_i\}$, with any budget-balanced CSM, her individual cost should be

$$\begin{aligned} f_{i,e^*}(p'_i, p_{-i}) + f_{i,e'_i}(p'_i, p_{-i}) &= F_{i,e^*}(1) + F_{i,e'_i}(1) \\ &> \sigma_{e^*} + \xi_{e^*,1} + \sigma_{e'_i} + \xi_{e'_i,1} \\ &= \frac{N}{N+1}\sigma + \frac{N}{N+1}\xi + \frac{1}{N+1}\sigma + \frac{3}{N+1}\xi \\ &= \sigma + \frac{N+3}{N+1}\xi\,, \end{aligned}$$

where the first equality still follows Eq. (13). Therefore, any player i cannot decrease her cost through a unilateral deviation. By definition, strategy profile p is a pure NE.

The total cost incurred by this NE is $N \cdot (\sigma_{e_i} + \sum_j \xi_{e_i,j}) > N(\sigma + \xi)$. In contrast, if every player chooses the path $\{e^*, e'_i\}$, the total cost should be

$$\begin{aligned} \sigma_{e^*} + \sum_j \xi_{e^*,j} \cdot N^{\alpha_j} + N \cdot (\sigma_{e'_i} + \sum_j \xi_{e'_i,j}) \;&<\; \frac{N}{N+1}c + \frac{N}{N+1}\xi \cdot N^{\alpha} \\ &\quad + N\left(\frac{1}{N+1}\sigma + \frac{3}{N+1}\xi\right) + \frac{2}{N+1}\xi \\ &= \frac{N}{N+1}\sigma + \frac{N}{N+1}\sigma + \frac{N}{N+1}\sigma + \frac{3N+2}{N+1}\xi \\ &< 3(\sigma + \xi)\,. \end{aligned}$$

Thus, the PoA is at least $\frac{N(\sigma+\xi)}{3(\sigma+\xi)} = \frac{N}{3}$. Since $N = (\frac{\sigma}{\xi})^{\frac{1}{\alpha}}$, this theorem follows. □

From Lemma 6.5, Lemma 6.7, and Theorem 10.1, it follows that for both the proportional fair CSM and the Shapley CSM, the induced GND games have a PoA of $O\left(\max_e \min_j (\sigma_e/\xi_{e,j})^{1/\alpha_j}\right)$. Then from Theorem 10.3, we know that these two natural CSMs are asymptotically optimal in the class of budget balanced CSMs, since they trivially follow Eq. (13).

11 ALTERNATIVE APPROACHES

Up to now, a set of game theoretic results, such as the smoothness parameters, have been established to investigate the performance of Alg-ABRD. Nevertheless, Alg-ABRD is not the only framework that can generate outputs with a desired approximation ratio. This section is dedicated to two alternative approaches for approximating the GND problem based on techniques developed in the existing literature in the context of routing requests.

11.1 Learning Based Algorithm for the GND problem with Routing Requests

In this part, we start to introduce a learning-based technique [43], which also utilizes the smoothness, and can guarantee a good approximation for the GND problem with routing requests when the optimal cost C^* of the input instance has a constant lower bound.

Definition 11.1 (Problem of Online Decision[30]). Consider an online problem where the input consists of a graph $G = (V, E)$ and a sequence of T' cost vectors $\left\{\tau^t = \{\tau^t(e)\}_{e\in E}\right\}_{t\in[T']}$, where $\tau^t(e) \in [0, 1]$. For each $t \in [T']$, this online problem requires a path r^t between a given source-target node pair without any knowledge of the cost vectors $\{\tau^t, \tau^{t+1}, \cdots, \tau^{T'}\}$. The objective is to minimize the REGRET, which is defined as

$$\texttt{REGRET} \;=\; \sum_{t=1}^{T'} \sum_{e\in r^t} \tau^t(e) - \min_{r': r' \text{ connects } \{s,t\}} \sum_{t=1}^{T'} \sum_{e\in p'} \tau^t(e)\,.$$

Lemma 11.2 (Follow the Perturbed Leader (FPL) [30]). *For the problem of online decision, there exists a randomized algorithm called FPL [30] that can compute every r^t in $O(|E| + |V|\log|V|)$-time such that the expectation of* REGRET *is no larger than $2|V|\sqrt{|E|T'}$.*

Using FPL as a subroutine, a learning based algorithm, referred to as Alg-L, is constructed as follows for the GND problem with routing requests. The first step is to transform the given problem instance $\mathcal{I}$ to a GND game by employing the proportional fair CSM, and divide every σ_e and every $\xi_{e,j}$ by a large enough number such that the cost share of any player on any edge is in the interval $[0, 1]$. Obviously, such a linear scaling on the cost functions $\{F_e\}$ does not influence the approximation ratio. Then, generate $T' = 4N^2|V|^2|E|$ strategy profiles $\{\bar{p}^t\}_{t\in[T']}$. For every t and every player i, the path $\bar{p}_i^t$ is obtained by running FPL with $\tau_i^{t'} = \left\{\{\tau_i^{t'}(e) = f_{i,e}(S_e^{t'}\cup\{i\})\}_{e\in E}\right\}_{t'\in[t-1]}$ as the input, where $f_{i,e}(\cdot)$ refers to the cost share determined by the proportional fair CSM. Note that with the proportional fair CSM, the *exact* cost share of each player on each edge can be obtained in constant time. Finally, choose one strategy profile t^* from $[T']$ randomly and uniformly, and output $\bar{p}^{t^*}$.

Lemma 11.3. *The algorithm* Alg-L *has a time complexity of $O(N^3|V|^2|E|^2\log|V|)$.*

Theorem 11.4. *Let C'^* be the optimal solution with respect to the linearly scaled cost functions. If the total cost of C'^* has a constant lower bound* LB*, then algorithm* Alg-L *guarantees an approximation ratio of $O\left(\max_{e\in E}\min_{j\in[q]}\left(\frac{\sigma_e}{\xi_{e,j}}\right)^{\frac{1}{\alpha_j}}\right)$ for the GND problem with routing requests.*

Proof. According to [41, Corollary 3.3], for a (λ, μ)-smooth game, by generating the strategy profiles $\{\bar{p}^t\}_{t\in[T']}$ through a randomized algorithm for the problem of online decision, the chosen strategy profile $\bar{p}^{t^*}$ guarantees that:

$$\mathbb{E}_{\bar{p}^{t^*}}[C'(\bar{p}^{t^*})] \;\le\; \frac{\lambda}{1-\mu}C'^* + \frac{1}{1-\mu}\frac{\sum_{i=1}^{N}\mathbb{E}_{\{\bar{p}_i^t\}_{t\in[T']}}[\texttt{REGRET}]}{T'}$$

where $C'(\bar{p}^{t^*})$ represents the total cost with respect to the scaled cost functions. By Theorem 6.1, the values of γ_α and λ_α are not influenced by the linear scaling on c_e and $\xi_{e,j}$. Thus,

$$\begin{aligned}\mathbb{E}_{\bar{p}^{t^*}}[C'(\bar{p}^{t^*})] &\leq 2(\gamma_\alpha + \lambda_\alpha)C'^* + \frac{2 \cdot \sum_{i=1}^{N} \mathbb{E}_{\{\bar{p}_i^t\}_{t\in[T']}}[\mathsf{REGRET}]}{T'} \\ &\leq 2(\gamma_\alpha + \lambda_\alpha)C'^* + \frac{N \cdot 4|V|\sqrt{|E|T'}}{T'} \\ &\leq 2\left(\gamma_\alpha + \lambda_\alpha + \frac{1}{\mathsf{LB}}\right)C'^*.\end{aligned}$$

The second transition above follows from Lemma 11.2. The third transition holds because it is assumed that $\mathtt{OPT}' \geq \mathtt{LB}$. Since LB is a constant, this theorem holds. □

Lemma 11.3 and Theorem 11.4 indicate that, Alg-L promises the same upper bound on the approximation ratio as Alg-ABRD for the special input instances with $C^{'*} \geq \mathtt{LB}$, and when the given graph has a small size while the number of requests is large, Alg-L has a better time complexity than Alg-ABRD. However, it remains unknown for us how to generalize Theorem 11.4 to the general case where there is no guarantee for the lower bound of the optimal solution. The critical issue here is that even before the linear scaling, the optimal result C^* can be arbitrarily small. This problem is left for future research.

11.2 Convex Programming and Rounding for the GND Problem with Routing Requests

The approach presented in this subsection was suggested to us by an anonymous reviewer for a special case of the GND problem. Specifically, like the approach presented in Section 11.1, this approach also addresses the GND problem with routing requests, but restricts the attention to the more specific case where the given graph $G = (V, E)$ is undirected and the weights of the requests are related. (Recall that related weights means that the weight of every request i satisfies $w_i(e) = w_i$ for every $e \in E$.) Furthermore, in this part, the cost function F_e of each edge e is assumed to be an energy consumption cost function Eq. (1), a specific (simpler) form of the REP cost functions Eq. (2) used in the other parts of this paper. We shall refer to this specific GND restriction as *energy efficient routing (EER)*.

Since this approach is based on convex programming and rounding, we shall refer to the resulting approximation algorithm as CPR. Throughout the following description of algorithm CPR, the notion *fractional solution* is often used. It refers to the solution obtained when the integral constraint is relaxed. Formally, for each request i in an EER problem, a fractional solution p specifies a finite set $p_i = \{p_{i,k}\}_{k\in[\mathcal{K}_i]}$ of paths $p_{i,k}$ connecting the source-target pair of request i, where each path $p_{i,k}$ is associated with a positive real number $y_{i,k} \in (0, 1]$ satisfying that $\sum_{k\in[\mathcal{K}]} y_{i,k} = 1$. For each $e \in E$, the load l_e^p incurred by a fractional solution p is defined to be $\sum_{i\in[N]} \sum_{k\in[\mathcal{K}_i]} \sum_{p_{i,k}:e\in p_{i,k}} w_i \cdot y_{i,k}$. The total cost of a fractional solution p is defined to be $\sum_{e\in E} F_e(l_e^p)$.

In algorithm CPR, every request $i \in [N]$ is first partitioned into a set $\mathcal{R}'_i$ of w_i *sub-requests*, where every sub-request i_j is associated with the same source-target node pair as the original request i, and has the same weight $w_{i_j} = 1$. Such a partition is feasible since the weights are assumed to be related. Let $\mathcal{R}' = \bigcup_i \mathcal{R}'_i$ be the set of all the sub-requests, $\mathcal{I}'$ be the instance obtained by replacing the set of requests in the given EER instance $\mathcal{I}$ with $\mathcal{R}'$, and $\widehat{\mathcal{I}}'$ be a variant instance which replaces

the energy consumption cost function F_e in $\mathcal{I}'$ with the following variant cost function

$$\widehat{F}_e(p') = \begin{cases} 0 & l_e^{p'} = 0 \\ \sigma_e + \xi_e \cdot \left((l_e^{p'})^\alpha + \sum_i (w_i)^{\alpha-1} \cdot l_e^{p'}(i) \right) & l_e^{p'} > 0 \end{cases}, \tag{14}$$

where p' is an arbitrary feasible path profile for $\mathcal{R}'$, $l_e^{p'}(i)$ is the load incurred by routing sub-requests in $\mathcal{R}'_i$ along e. Let $\widehat{p}^*$ be the *fractional* optimal solution of $\widehat{\mathcal{I}}'$. Then the following result can be proved in a similar way with [13].

Lemma 11.5. *$\widehat{C}(\widehat{p}^*) < 2 \cdot C^*$, where $\widehat{C}(\cdot)$ denotes the total cost with respect to Eq. (14).*

Proof. The path profile p^* also induces a feasible solution for $\widehat{\mathcal{I}}'$, which routes i_j along the path p_i^* for every $i \in [N]$. Let the total cost incurred by this feasible solution be $\widehat{C}(p^*)$, then,

$$\begin{aligned} \widehat{C}(\widehat{p}^*) \le \widehat{C}(p^*) &= \sum_{e \in p^*} \left[\sigma_e + \xi_e \cdot \left((l_e^{p^*})^\alpha + \sum_i (w_i)^{\alpha-1} \cdot l_e^{p^*}(i) \right) \right] \\ &= \sum_{e \in p^*} \left[\sigma_e + \xi_e \cdot \left((l_e^{p^*})^\alpha + \sum_{i: e \in p_i^*} (w_i)^\alpha \right) \right] \\ &\le \sum_{e \in p^*} \left[\sigma_e + \xi_e \cdot \left((l_e^{p^*})^\alpha + (l_e^{p^*})^\alpha \right) \right] \\ &< 2 \cdot \sum_{e \in p^*} \left(\sigma_e + \xi_e \cdot (l_e^{p^*})^\alpha \right) \\ &= 2 \cdot C^*. \end{aligned}$$

The second line holds since p^* is an integral solution of $\mathcal{I}$ [13]. The third line following the fact $\sum_{i: e \in p_i^*} w_i = l_e^{p^*}$ and the superadditivity of the power function. The fourth line holds since $\sigma_e > 0$. □

The next step of CPR is to utilize the convex programming based technique proposed in [5] to generate a solution for the instance $\widehat{\mathcal{I}}'$. In particular, it converts $\widehat{F}_e(p')$ to a convex cost function $\bar{F}_e(p') = W_e(p') + \xi_e \sum_i (w_i)^{\alpha-1} \cdot l_e^{p'}(i)$, where

$$W_e(p') = \begin{cases} \zeta_e \cdot (l_e^{p'}) & \text{if } l_e^{p'} \in \left[0, \max\left\{ 1, \left(\frac{\sigma_e}{(\alpha-1)\xi_e} \right)^{\frac{1}{\alpha}} \right\} \right] \\ \sigma_e + \xi_e \cdot (l_e^{p'})^\alpha & \text{if } l_e^{p'} > \max\left\{ 1, \left(\frac{\sigma_e}{(\alpha-1)\xi_e} \right)^{\frac{1}{\alpha}} \right\} \end{cases},$$

and

$$\zeta_e = \begin{cases} \sigma_e + \xi_e & \text{if } \left(\frac{\sigma_e}{(\alpha-1)\xi_e} \right)^{\frac{1}{\alpha}} < 1 \\ \alpha \xi_e \left(\frac{\sigma_e}{(\alpha-1)\xi_e} \right)^{1-\frac{1}{\alpha}} & \text{if } \left(\frac{\sigma_e}{(\alpha-1)\xi_e} \right)^{\frac{1}{\alpha}} \ge 1 \end{cases}.$$

By employing the convex cost function $\bar{F}_e(p')$ and relaxing the integral constraint, a solvable convex program is obtained. Following [5], CPR solves the convex program to obtain a fractional solution $\widehat{p}^\circ$ and then rounds it to an integral solution $\widehat{p}^\sharp$ through a random rounding procedure. In particular, for each sub-request i_j, a path $\widehat{p}_{i_j}$ is chosen from the set $\widehat{p}^\circ_{i_j} = \{\widehat{p}^\circ_{i_j,k}\}_{k \in [\mathcal{K}_{i_j}]}$ by taking the positive real numbers $\{y_{i_j,k}\}_{k \in \mathcal{K}_{i_j}}$ associated with $\{\widehat{p}^\circ_{i_j,k}\}_{k \in [\mathcal{K}_{i_j}]}$ as a probability distribution

[5]. This random rounding procedure guarantees that for every edge $e \in E$,

$$\mathbb{E}\left[\xi_e \sum_i (w_i)^{\alpha-1} \cdot l_e^{\widehat{p}^\sharp}(i)\right] = \xi_e \sum_i (w_i)^{\alpha-1} \cdot l_e^{\widehat{p}^\circ}(i) \text{ and } \mathbb{E}\left[W_e(\widehat{p}^\sharp)\right] \le O\left(\left(\frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}}\right) \cdot W_e(\widehat{p}^\circ). \quad (15)$$

Then we have

$$\sum_e \mathbb{E}\left[\bar{F}_e(\widehat{p}^\sharp)\right] \le O\left(\left(\max_e \frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}}\right) \cdot \sum_e \bar{F}_e(\widehat{p}^\circ). \quad (16)$$

Since $\widehat{p}^\circ$ is the optimal fractional solution of the convex program with respect to the cost function $\bar{F}_e(p')$, and $\bar{F}_e(p') \le \widehat{F}_e(p')$ for any edge e and any profile p' that is feasible for $\mathcal{R}'$ [5, Section IV.B]:

$$\sum_e \bar{F}_e(\widehat{p}^\circ) \le \sum_e \bar{F}_e(\widehat{p}^*) \le \sum_e \widehat{F}_e(\widehat{p}^*) = \widehat{C}(\widehat{p}^*) \quad (17)$$

Recall that $\widehat{C}(\widehat{p}^*)$ is the optimal fractional solution of $\widehat{\mathcal{I}}'$, Eq. (16) and Eq. (17) imply that:

Lemma 11.6. *The solution $\widehat{p}^\sharp$ is an $O\left(\left(\max_e \frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}}\right)$-approximation solution of the instance $\widehat{\mathcal{I}}'$.*

The last step of CPR is to convert $\widehat{p}^\sharp$ to an integral solution $p^\sharp$ that is feasible for the original instance $\mathcal{I}$, still by randomized rounding. In particular, to generate $p^\sharp$, each traffic request i in instance $\mathcal{I}$ should be routed along the path $\widehat{p}_{ij}^\sharp \in \widehat{p}^\sharp$ with probability $\frac{1}{w_i}$. Then we have

$$\begin{aligned}
\mathbb{E}\left[\sum_e F_e(p^\sharp)\right] &\le \mathbb{E}\left[\sum_{e\in\widehat{p}^\sharp} \left(\sigma_e + \xi_e (l_e^{p^\sharp})^\alpha\right)\right] \\
&= \mathbb{E}\left[\sum_{e\in\widehat{p}^\sharp} \sigma_e\right] + \mathbb{E}\left[\sum_{e\in\widehat{p}^\sharp} \xi_e (l_e^{p^\sharp})^\alpha\right] \\
&\le \mathbb{E}\left[\sum_{e\in\widehat{p}^\sharp} \sigma_e\right] + O(\alpha^\alpha)\mathbb{E}\left[\sum_{e\in\widehat{p}^\sharp} \xi_e (l_e^{\widehat{p}^\sharp})^\alpha + \sum_i (w_i)^{\alpha-1} l_e^{\widehat{p}^\sharp}(i)\right] \\
&\le O(\alpha^\alpha)\mathbb{E}\left[\sum_{e\in\widehat{p}^\sharp} \widehat{F}_e(\widehat{p}^\sharp)\right]. \quad (18)
\end{aligned}$$

The third transition above follows from [27, Section 5.2]. Combining Lemma 11.5, Lemma 11.6, and Eq. (18) gives the following result:

Theorem 11.7. *The algorithm* CPR *has an approximation ratio of $O\left(\left(\max_e \frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}}\right)$.*

Weight-scaling and Loss in Approximation Ratio. Algorithm CPR processes every sub-request independently, therefore its time complexity depends on the numeric value $\sum_i w_i$ which cannot be bounded by a polynomial of the instance size poly($|\mathcal{I}|$). A naive idea for overcoming this issue is to scale down and round the weights so that w_i is bounded by a polynomial of $|\mathcal{I}|$. Although this *weight-scaling* technique works well for some classic optimization problems such as the Knapsack problem, it may incur a significant loss in the approximation ratio for the EER problem.

Generally speaking, the weight-scaling technique can be described as a function WSF that maps each weight vector $\boldsymbol{w}$ to a weight vector $\boldsymbol{w'}$ of the same length. For any $N \in \mathbb{Z}_{\ge 1}$ and $\kappa > 1$, let

$\mathbf{1}_N$ and $\boldsymbol{\kappa}_N$ be two vectors of length N so that every element in $\mathbf{1}_N$ (resp. $\boldsymbol{\kappa}_N$) is 1 (resp. κ). A weight-scaling function is said to be κ-ambiguous if $\mathsf{WSF}(\mathbf{1}_N) = \mathsf{WSF}(\boldsymbol{\kappa}_N)$ for any $N \in \mathbb{Z}_{\geq 1}$.

For illustration, consider a weight-scaling function WSF which maps each given $\boldsymbol{w}$ to $\boldsymbol{w'}$ so that for each $i \in [|\boldsymbol{w}|]$, the element $\boldsymbol{w'}(i)$ is set to $\left\lceil \frac{\boldsymbol{w}(i) \cdot |\boldsymbol{w}|}{\varepsilon \cdot \max_i \boldsymbol{w}(i)} \right\rceil$, where $0 < \varepsilon < 1$ is a constant. A similar function is used in [40] to obtain an FPTAS algorithm for the Knapsack problem. It is easy to see that such a weight-scaling function is κ-ambiguous for any $\kappa > 1$.

To analyze the loss in approximation ratio that can be caused by a weight-scaling function, now construct an undirected graph $G_0 = (V_0, E_0)$ with respect to some positive number $\kappa > 1$. There are two nodes in G_0, u and v, which are connected by more than $\kappa^{\frac{\alpha}{2\alpha-1}}$ parallel edges. There exists a special edge e^* with cost parameters $\sigma_{e^*} = 1$ and $\xi_{e^*} = 1$, while for any other edge $e \in E - \{e^*\}$, $\sigma_e = \kappa^\alpha$ and $\xi_e = 1$. Following theorem shows that on this graph, a κ-ambiguous weight-scaling function leads to an approximation ratio that cannot be bounded by $O\left(\max_e \left(\frac{\sigma_e}{\xi_e}\right)^{\frac{1}{\alpha}}\right) = O(\kappa)$ when $\alpha > \frac{3+\sqrt{5}}{2}$.

Theorem 11.8. *By taking any κ-ambiguous weight-scaling function, the approximation ratio of any deterministic algorithm for the EER problem on the graph G_0 has a lower bound of $\Omega\left(\kappa^{\frac{\alpha\cdot(\alpha-1)}{2\alpha-1}}\right)$.*

Proof. Consider two input instances, $\mathcal{I}_1$ and $\mathcal{I}_2$, on the graph G_0. Each of these two instances contains $N = \kappa^{\frac{\alpha}{2\alpha-1}}$ requests with the source-target pair $\{u, v\}$. For every $i \in [N]$, the weights of player i in instance $\mathcal{I}_1$ and instance $\mathcal{I}_2$ are set to 1 and κ, respectively. Suppose that the vectors of weights in $\mathcal{I}_1$ and $\mathcal{I}_2$ are fed to a κ-ambiguous weight-scaling function to generate two new instances $\mathcal{I}_1^{\mathsf{MST}}$ and $\mathcal{I}_2^{\mathsf{MST}}$. By definition, no algorithm can distinguish $\mathcal{I}_1^{\mathsf{MST}}$ from $\mathcal{I}_2^{\mathsf{MST}}$. Hence, the following observation trivially holds.

Claim 11.9. *For any deterministic algorithm of the EER problem, the output generated for $\mathcal{I}_1^{\mathsf{MST}}$ is same as the one for $\mathcal{I}_2^{\mathsf{MST}}$.*

Let the output generated by a given deterministic algorithm of the EER problem for $\mathcal{I}_1^{\mathsf{MST}}$ be p. Denote the optimal solution of $\mathcal{I}_1$ (resp. $\mathcal{I}_2$) by $C^*(\mathcal{I}_1)$ (resp. $C^*(\mathcal{I}_2)$), and the total cost incurred by p for $\mathcal{I}_1$ (resp. $\mathcal{I}_2$) be $C(\mathcal{I}_1, p)$ (resp. $C(\mathcal{I}_2, p)$).

Claim 11.10. *The approximation ratio of the given deterministic algorithm is at least* $\max\left\{\frac{C(\mathcal{I}_1,p)}{C^*(\mathcal{I}_1)}, \frac{C(\mathcal{I}_2,p)}{C^*(\mathcal{I}_2)}\right\}$.

Claim 11.11. *For any profile p,* $\max\left\{\frac{C(\mathcal{I}_1,p)}{C^*(\mathcal{I}_1)}, \frac{C(\mathcal{I}_2,p)}{C^*(\mathcal{I}_2)}\right\} \geq \frac{1}{2}\kappa^{\frac{\alpha\cdot(\alpha-1)}{2\alpha-1}}$.

Proof. Suppose that x edges in $E - \{e^*\}$ are used by p. Then $C(\mathcal{I}_1, p) \geq x \cdot (\kappa^\alpha + 1)$. Since the cost of routing all requests through e^* in the instance $\mathcal{I}$ is $1 + \kappa^{\frac{\alpha^2}{2\alpha-1}}$,

$$\frac{C(\mathcal{I}_1, p)}{C^*(\mathcal{I}_1)} \geq \frac{x(\kappa^\alpha + 1)}{1 + \kappa^{\frac{\alpha^2}{2\alpha-1}}} \geq \frac{x}{2} \cdot \kappa^{\frac{\alpha(\alpha-1)}{2\alpha-1}} .$$

Noticing that the $x + 1$ edges used by p have the same value of the parameter ξ_e, we have:

$$C(\mathcal{I}_2, p) \geq (x+1)\left\lfloor \left(\frac{N}{x+1} \cdot \kappa\right)^\alpha \right\rfloor = \kappa^{\frac{\alpha(3\alpha-1)}{2\alpha-1}} \cdot (x+1)^{1-\alpha} .$$

By routing each request along a distinct edge, a solution with total cost $2\kappa^{\frac{2\alpha^2}{2\alpha-1}}$ can be obtained for the instance $\mathcal{I}_2$. Therefore,

$$\frac{C(\mathcal{I}_2, p)}{C^*(\mathcal{I}_2)} \geq \frac{(x+1)^{1-\alpha}}{2} \cdot \kappa^{\frac{\alpha(\alpha-1)}{2\alpha-1}} .$$

Note that $x \in \mathbb{Z}_{\geq 0}$. When $x \geq 1$, $\max\left\{\frac{C(\mathcal{I}_1,p)}{C^*(\mathcal{I}_1)}, \frac{C(\mathcal{I}_2,p)}{C^*(\mathcal{I}_2)}\right\} \geq \frac{C(\mathcal{I}_1,p)}{C^*(\mathcal{I}_1)} \geq \frac{1}{2}\kappa^{\frac{\alpha(\alpha-1)}{2\alpha-1}}$; while if $x = 0$, $\max\left\{\frac{C(\mathcal{I}_1,p)}{C^*(\mathcal{I}_1)}, \frac{C(\mathcal{I}_2,p)}{C^*(\mathcal{I}_2)}\right\} \geq \frac{C(\mathcal{I}_2,p)}{C^*(\mathcal{I}_2)} \geq \frac{1}{2}\kappa^{\frac{\alpha(\alpha-1)}{2\alpha-1}}$. ■ (Claim 11.11)

This proof is completed by combining the claims above. □

For any randomized algorithm for the EER instance, it should generate an output for the instance $\mathcal{I}_1$ with the same probability distribution over the path profiles as the one for the instance $\mathcal{I}_2$. Therefore, we have the following result.

Corollary 11.12. *With any κ-ambiguous weight-scaling function, the approximation ratio of any randomized algorithm for the EER problem on the graph G_0 has a lower bound of $\Omega\left(\kappa^{\frac{\alpha(\alpha-1)}{2\alpha-1}}\right)$.*

As mentioned earlier, the approximation ratio $\Omega\left(\kappa^{\frac{\alpha(\alpha-1)}{2\alpha-1}}\right)$ is worse than the approximation ratio promised by `Alg-ABRD` when $\alpha > \frac{3+\sqrt{5}}{2} \approx 2.618$. Furthermore, by Theorem 11.7, we cannot expect an approximation ratio better than the ratio promised by `Alg-ABRD` when $\alpha \leq 2.618$, either, if we combine `CPR` with a weight-scaling function.

ACKNOWLEDGMENTS

The work of Yuval Emek was supported in part by an Israeli Science Foundation grant number 1016/17. The work of Shay Kutten was supported in part by a grant from the ministry of science in the program that is joint with JSPS and in part by the BSF. The work of Ron Lavi was partially supported by the ISF-NSFC joint research program (grant No. 2560/17). The work of Yangguang Shi was partially supported at the Technion by a fellowship of the Israel Council for Higher Education.

The authors would like to thank Tim Roughgarden for very useful discussions and thank the referees for the helpful comments.

REFERENCES

[1] Ittai Abraham, Danny Dolev, and Joseph Y Halpern. 2013. Distributed protocols for leader election: A game-theoretic perspective. In *International Symposium on Distributed Computing*. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 61–75.

[2] Ajit Agrawal, Philip Klein, and R. Ravi. 1995. When Trees Collide: An Approximation Algorithm for the Generalized Steiner Problem on Networks. *SIAM J. Comput.* 24, 3 (1995), 440–456.

[3] Susanne Albers. 2010. Energy-efficient Algorithms. *Commun. ACM* 53, 5 (2010), 86–96.

[4] M. Andrews. 2004. Hardness of buy-at-bulk network design. In *45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, Rome, Italy, 115–124.

[5] Matthew Andrews, Antonio Fernández Anta, Lisa Zhang, and Wenbo Zhao. 2012. Routing for Power Minimization in the Speed Scaling Model. *IEEE/ACM Transactions on Networking* 20, 1 (2012), 285 –294. Preliminary version in INFOCOM'10.

[6] Matthew Andrews, Spyridon Antonakopoulos, and Lisa Zhang. 2016. Minimum-Cost Network Design with (Dis) economies of Scale. *SIAM J. Comput.* 45, 1 (2016), 49–66. Preliminary version in FOCS'10.

[7] Matthew Andrews, Antonio Fernández Anta, Lisa Zhang, and Wenbo Zhao. 2013. Routing and scheduling for energy and delay minimization in the power down model. *Networks* 61, 3 (2013), 226–237. Preliminary version in INFOCOM'10.

[8] Elliot Anshelevich, Anirban Dasgupta, Jon Kleinberg, Eva Tardos, Tom Wexler, and Tim Roughgarden. 2008. The price of stability for network design with fair cost allocation. *SIAM J. Comput.* 38, 4 (2008), 1602–1623. Preliminary version in FOCS'04.

[9] Antonios Antoniadis, Sungjin Im, Ravishankar Krishnaswamy, Benjamin Moseley, Viswanath Nagarajan, Kirk Pruhs, and Cliff Stein. 2014. Hallucination helps: energy efficient virtual circuit routing. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '14)*. Society for Industrial and Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1141–1153.

[10] Baruch Awerbuch and Yossi Azar. 1997. Buy-at-bulk network design. In *Proceedings 38th Annual Symposium on Foundations of Computer Science (FOCS'97)*. IEEE, Miami Beach, FL, USA, 542–547.

[11] Baruch Awerbuch, Yossi Azar, Amir Epstein, Vahab Seyed Mirrokni, and Alexander Skopalik. 2008. Fast Convergence to Nearly Optimal Solutions in Potential Games. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC '08)*. ACM, New York, NY, USA, 264–273.

[12] Baruch Awerbuch, Yossi Azar, Edward F. Grove, Ming-Yang Kao, P. Krishnan, and Jeffrey Scott Vitter. 1995. Load Balancing in the L_p Norm. In *36th Annual Symposium on Foundations of Computer Science FOCS*. IEEE, Milwaukee, WI, USA, 383–391.

[13] Yossi Azar and Amir Epstein. 2005. Convex Programming for Scheduling Unrelated Parallel Machines. In *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing STOC*. ACM, New York, NY, USA, 331–337.

[14] Yossi Azar, Leah Epstein, Yossi Richter, and Gerhard J Woeginger. 2004. All-norm approximation algorithms. *Journal of Algorithms* 52, 2 (2004), 120 – 133.

[15] Nikhil Bansal, Tracy Kimbrel, and Kirk Pruhs. 2007. Speed Scaling to Manage Energy and Temperature. *J. ACM* 54, 1 (2007), 3:1–3:39.

[16] Jarosław Byrka, Fabrizio Grandoni, Thomas Rothvoss, and Laura Sanità. 2013. Steiner tree approximation via iterative randomized rounding. *Journal of the ACM (JACM)* 60, 1 (2013), 6. Preliminary version in STOC'10.

[17] Moses Charikar, Chandra Chekuri, To-yat Cheung, Zuo Dai, Ashish Goel, Sudipto Guha, and Ming Li. 1998. Approximation Algorithms for Directed Steiner Problems. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 192–200.

[18] Moses Charikar and Adriana Karagiozova. 2005. On non-uniform multicommodity buy-at-bulk network design. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (STOC'05)*. ACM, New York, NY, USA, 176–182.

[19] Chandra Chekuri, Guy Even, Anupam Gupta, and Danny Segev. 2011. Set Connectivity Problems in Undirected Graphs and the Directed Steiner Network Problem. *ACM Trans. Algorithms* 7, 2 (2011), 18:1–18:17.

[20] Chandra Chekuri, Mohammad Taghi Hajiaghayi, Guy Kortsarz, and Mohammad R Salavatipour. 2010. Approximation algorithms for nonuniform buy-at-bulk network design. *SIAM J. Comput.* 39, 5 (2010), 1772–1798. Preliminary version in FOCS'06.

[21] Ho-Lin Chen, Tim Roughgarden, and Gregory Valiant. 2010. Designing network protocols for good equilibria. *SIAM J. Comput.* 39, 5 (2010), 1799–1832. Preliminary version in SODA'04.

[22] Ken Christensen, Pedro Reviriego, Bruce Nordman, Michael Bennett, Mehrgan Mostowfi, and Juan Antonio Maestro. 2010. IEEE 802.3 az: the road to energy efficient ethernet. *IEEE Communications Magazine* 48, 11 (2010), 50–56.

[23] Richard Cole, José R. Correa, Vasilis Gkatzelis, Vahab Mirrokni, and Neil Olver. 2011. Inner Product Spaces for MinSum Coordination Mechanisms. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing (STOC '11)*. ACM, New York, NY, USA, 539–548.

[24] E. W. Dijkstra. 1959. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* 1, 1 (1959), 269–271.

[25] Michael L. Fredman and Robert Endre Tarjan. 1987. Fibonacci Heaps and Their Uses in Improved Network Optimization Algorithms. *J. ACM* 34, 3 (1987), 596–615.

[26] Vasilis Gkatzelis, Konstantinos Kollias, and Tim Roughgarden. 2014. Optimal cost-sharing in weighted congestion games. In *International Conference on Web and Internet Economics*. Springer, Cham, 72–88.

[27] Anupam Gupta, Ravishankar Krishnaswamy, and Kirk Pruhs. 2011. Online Primal-Dual For Non-linear Optimization with Applications to Speed Scaling. *CoRR* abs/1109.5931 (2011), 16. arXiv:1109.5931 http://arxiv.org/abs/1109.5931

[28] Sergiu Hart and Andreu Mas-Colell. 1989. Potential, Value, and Consistency. *Econometrica: Journal of the Econometric Society* 57 (1989), 589–614.

[29] Sandy Irani and Kirk R. Pruhs. 2005. Algorithmic Problems in Power Management. *SIGACT News* 36, 2 (2005), 63–76.

[30] Adam Kalai and Santosh Vempala. 2005. Efficient algorithms for online decision problems. *J. Comput. System Sci.* 71, 3 (2005), 291–307.

[31] Samir Khuller, Jian Li, and Barna Saha. 2010. Energy Efficient Scheduling via Partial Shutdown. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms SODA*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1360–1372.

[32] Konstantinos Kollias and Tim Roughgarden. 2015. Restoring pure equilibria to weighted congestion games. *ACM Transactions on Economics and Computation* 3, 4 (2015), 21. Preliminary version in ICALP'11.

[33] V. S. Anil Kumar, Madhav V. Marathe, Srinivasan Parthasarathy, and Aravind Srinivasan. 2009. A Unified Approach to Scheduling on Unrelated Parallel Machines. *J. ACM* 56, 5, Article 28 (2009), 31 pages.

[34] J. K. Lenstra, D. B. Shmoys, and É. Tardos. 1990. Approximation Algorithms for Scheduling Unrelated Parallel Machines. *Math. Program.* 46, 3 (1990), 259–271.

[35] Jian Li and Samir Khuller. 2011. Generalized Machine Activation Problems. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms SODA*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 80–94.

[36] David Liben-Nowell, Alexa Sharp, Tom Wexler, and Kevin Woods. 2012. Computing shapley value in supermodular coalitional games. In *International Computing and Combinatorics Conference*. Springer Berlin Heidelberg, Berlin, Heidelberg, 568–579.

[37] Konstantin Makarychev and Maxim Sviridenko. 2014. Solving Optimization Problems with Diseconomies of Scale via Decoupling. In *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS '14)*. IEEE Computer Society, Washington, DC, USA, 571–580.

[38] Sergiu Nedevschi, Lucian Popa, Gianluca Iannaccone, Sylvia Ratnasamy, and David Wetherall. 2008. Reducing Network Energy Consumption via Sleeping and Rate-adaptation. In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation (NSDI'08)*. USENIX Association, Berkeley, CA, USA, 323–336.

[39] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. 2007. *Algorithmic game theory*. Vol. 1. Cambridge University Press Cambridge, New York, NY, USA.

[40] Debmalya Panigrahi. 2015. Design and Analysis of Algorithms. (Oct. 2015). https://www.cs.duke.edu/courses/fall15/cps232/scribe_notes/lec16.pdf

[41] Tim Roughgarden. 2013. CS364A: Algorithmic Game Theory Lecture# 17: No-Regret Dynamics. (2013).

[42] Tim Roughgarden. 2014. Barriers to Near-Optimal Equilibria. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS*. IEEE, Philadelphia, PA, USA, 71–80.

[43] Tim Roughgarden. 2015. Intrinsic robustness of the price of anarchy. *Journal of the ACM (JACM)* 62, 5 (2015), 32. Preliminary version in STOC'09.

[44] Tim Roughgarden and Okke Schrijvers. 2016. Network cost-sharing without anonymity. *ACM Transactions on Economics and Computation* 4, 2 (2016), 8. Preliminary version in SAGT'14.

[45] David B. Shmoys and Éva Tardos. 1993. An Approximation Algorithm for the Generalized Assignment Problem. *Math. Program.* 62, 3 (1993), 461–474.

[46] Philipp von Falkenhausen and Tobias Harks. 2013. Optimal cost sharing for resource selection games. *Mathematics of Operations Research* 38, 1 (2013), 184–208.

[47] F. Yao, A. Demers, and S. Shenker. 1995. A scheduling model for reduced CPU energy. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE, Milwaukee, WI, USA, 374–382.