This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3161201, IEEE Transactions on Control of Network Systems

Randomized Gossiping with Effective Resistance Weights: Performance Guarantees and Applications

Bugra Can, *Student Member, IEEE* , Saeed Soori, *Student Member, IEEE,* Necdet Serhat Aybat, *Member, IEEE,* Maryam Mehri Dehnavi, *Member, IEEE,* Mert Gürbüzbalaban, *Member, IEEE*

***Abstract*—The effective resistance between a pair of nodes in a weighted undirected graph is defined as the potential difference induced when a unit current is injected at one node and extracted from the other, treating edge weights as the conductance values of edges. The effective resistance is a key quantity of interest in many applications, e.g., solving linear systems, Markov Chains, and continuous-time averaging networks. We consider effective resistances (ER) in the context of designing randomized gossiping methods for the consensus problem, where the aim is to compute the average of node values in a distributed manner through iteratively computing weighted averages among randomly chosen neighbours. For barbell graphs, we prove that choosing wake-up and communication probabilities proportional to ER weights improves the averaging time corresponding to the traditional choice of uniform weights. For c-barbell graphs, we show that ER weights admit lower and upper bounds on the averaging time that improves upon the lower and upper bounds available for uniform weights. Furthermore, for graphs with a small diameter, we can show that ER weights can improve upon the existing bounds for Metropolis weights by a constant factor under some assumptions. We illustrate these results through numerical experiments where we showcase the efficiency of our approach on several graph topologies including barbell graphs, and small-world graphs. We also present an application of the ER gossiping to distributed optimization: we numerically verify that using ER gossiping within EXTRA and DPGA-W methods improves their practical performance in terms of communication efficiency.**

***Index Terms*—Distributed algorithms/control, networks of autonomous agents, optimization, randomized gossiping algorithms**

I. Introduction

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E}, w)$ be an undirected, weighted and connected graph defined by the set of nodes (agents) $\mathcal{N} = \{1, \ldots, n\}$, the set of edges $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, and the edge weights $w_{ij} > 0$ for $(i, j) \in \mathcal{E}$. Since $\mathcal{G}$ is undirected, we assume that both (i, j) and (j, i) refer to the same edge when it exists, and for all $(i, j) \in \mathcal{E}$, we set $w_{ji} = w_{ij}$. Identifying the weighted graph $\mathcal{G}$ as an electrical network in which each edge (i, j) corresponds to a branch of conductance w_{ij}, the effective resistance R_{ij} between a pair of nodes i and j is defined as the voltage potential difference induced between them when a unit current is injected at i and extracted at j. The effective resistance (ER), also known as the resistance distance, is a key quantity of interest to compute in many applications and algorithmic questions over graphs. It defines a metric on the graph providing bounds on its conductance [2], [3]. Furthermore, it is closely associated with the hitting and commute times of a random walk[1] on the graph $\mathcal{G}$ when the probability of a transition from i to $j \in \mathcal{N}_i$ is $w_{ij}/\sum_{j' \in \mathcal{N}_i} w_{ij'}$ where $\mathcal{N}_i \triangleq \{j \in \mathcal{N} : \ w_{ij} > 0\}$ denotes the set of neighboring nodes of $i \in \mathcal{N}$; therefore, it arises naturally for studying random walks over graphs and their mixing time properties [4], [5], [6], spectral approximation of graphs [7], continuous-time averaging networks including consensus problems in distributed optimization [4].

There exist *centralized* algorithms for computing or approximating effective resistances accurately which require global communication beyond local information exchange among the neighboring agents [8], [9], [10], [7], [11]. The references [8], [9], [10] develop key techniques for computing the effective resistances explicitly on specific network types. In particular, [9] addresses a class of graphs which are underlying networks of some symmetric association schemes whereas [8] considers two dimensional resistor networks. The reference [10] provides an algorithm for the calculation of the resistance between two arbitrary nodes in a distance-regular network and also provides analytical formulas. The works [7], [11] are based on computing or approximating the entries of the pseudoinverse $\mathcal{L}^\dagger$ of the graph Laplacian matrix $\mathcal{L}$, based on the identity:

$$R_{ij} = \mathcal{L}^\dagger_{ii} + \mathcal{L}^\dagger_{jj} - 2\mathcal{L}^\dagger_{ij}, \quad \forall \ (i, j) \in \mathcal{E}. \tag{1}$$

However, such centralized algorithms are *impractical* or *infeasible* for several key applications in multi-agent systems, e.g., randomized gossiping algorithms, for averaging the node values across the whole network, use only local communications between random neighbors (see [12], [13], [14]); this motivates the use of *distributed algorithms* for computing effective resistances which only rely on the information exchange among immediate neighbors. In these applications, communication among the agents is typically the bottleneck compared to the complexity of local computations of the agents; thus, it is crucial to develop distributed algorithms that are efficient in terms of the total number of communications required. To the best of the authors' knowledge, the first attempt for computing effective resistances in a decentralized way and also the first ER-based randomized gossiping algorithms appeared in [1]. The latter algorithms are asynchronous gossiping algorithms where each agents' wake-up and communication probabilities are chosen proportional to ER weights (see Section II for

B. Can and M. Gürbüzbalaban are with the Department of Management and Information Sciences, Rutgers Business School, Piscataway, NJ-08854, USA. S. Soori and M. M. Dehnavi are with the Department of Computer Sciences, University of Toronto, Toronto, Canada. N. S. Aybat is with the Industrial and Manufacturing Engineering Department Penn State University, University Park, PA 16802-4400, USA.

A short conference version of this work appeared in [1] at the IEEE GlobalSIP'17 conference.

[1]The hitting time is the expected number of steps of a random walk starting from i until it first visits j. The commute time C_{ij} is the expected number of steps required to go from i to j and from j to i back again.

details). Aybat and Gürbüzbalaban have shown in [1] that effective resistance (ER) weights can be computed at each agent locally with an efficient distributed algorithm, Distributed Randomized Kaczmarz (D-RK). Our paper is motivated by the numerical evidence presented in [1] that using ER weights has the potential to improve the performance of randomized gossiping algorithms on specific graphs. Since in [1] no rigorous performance guarantees for the use of ER weights were provided, here we focus on establishing the missing theoretical results that match the outstanding empirical behavior.

Contributions. First, in this paper, we provide theoretical guarantees on the ER-based randomized gossiping algorithms proposed in [1] for the consensus problem, where the objective is to compute the average of node values over a network in a decentralized manner [13]. A standard approach for solving the consensus problem is the *randomized uniform gossiping*, where each node keeps a local estimate of the average of node values and has the equal (uniform) probability of being activated to communicate with a randomly chosen neighbour to update its local estimate. However, this approach treats all the edges (equally) uniformly and can be slow in practice. To overcome this problem, in [1], ER-based randomized gossiping algorithms were proposed without any theoretical guarantees, in which the edges are being activated by non-uniform probabilities that are proportional to their effective resistances.

Our theoretical results presented in Section III (see Results 1, 2, and 3) explain the superior empirical behaviour of ER-based gossiping over the uniform gossiping observed in [1]. Briefly, we bound the time required to compute an inexact average using analysis based on conductance and spectral properties of the underlying weighted communication graph, and compare the bounds we obtained corresponding to the ER and uniform gossiping methods. We show that averaging time with ER weights is $\Theta(n)$ faster than that of uniform gossiping on a barbell graph where n is the number of agents. Furthermore, we also prove that for connected graphs with a small diameter, the averaging time with resistance weights can be faster than known performance bounds for the averaging time with gossiping based on Metropolis weights by a constant factor, see (Remark 12). Although our theoretical results are limited to these special graph types, we also provide numerical experiments on several other graph topologies which illustrate the performance improvements that can be obtained within ER-based gossiping. In our experiments, the effective resistances are first computed with the normalized D-RK algorithm of [1] and then used for ER-based gossiping.Our theoretical and numerical results show that ER weights are especially useful in the presence of "bottleneck edges" or clusters giving a graph cut leading to small graph conductance values. In the extended version [15], we have also numerically demonstrated that gossiping with ER weights performs better than uniform and Metropolis weights on a class of random graphs generated by *the stochastic block model* [16]. These graphs frequently arise in real-world applications, ranging from community detection to clustering, see e.g., [16] for a detailed discussion.

On a different note, Aybat and Gürbüzbalaban [1] introduced two methods to compute ER weights in a decentralized manner: D-RK and normalized D-RK –both converging linearly. In our experiments at Section V, we have adopted the normalized D-RK, upon proving that the convergence rate of normalized D-RK is better than D-RK (see the extended version of our paper in [15]); resolving a conjecture raised in [1].

Second, we consider the consensus optimization problem, where the agents connected on a network aim to collaboratively solve the optimization problem $\min_{x\in\mathbb{R}^p} f(x) \triangleq \sum_{i=1}^n f_i(x)$ where $f_i(x) : \mathbb{R}^p \to \mathbb{R}$ is a cost function only available to (node) agent i. This problem includes a number of key problems in supervised learning including distributed regression and logistic regression or more generally distributed empirical risk minimization problems [17], [18]. The consensus iterations are a building block of many existing state-of-the-art distributed consensus optimization algorithms such as the EXTRA and the distributed proximal gradient (DPGA-W) [19] algorithms for consensus optimization. We show through numerical experiments that our framework based on effective resistances can improve the performance of the EXTRA and DPGA-W algorithms for consensus optimization in terms of the total number of communications required. We believe our framework has far-reaching potential for improving the communication efficiency of many other distributed algorithms including distributed subgradient and ADMM methods, and this will be the subject of future work.

Related work. For consensus problems, there are some alternative methods to accelerate the commonly used consensus protocols. The approach in [20] is a synchronous algorithm combining Metropolis weights with a momentum averaging scheme. There are other approaches based on momentum averaging [21], [22], [23], min-sum splitting [24], and Chebyshev acceleration [25], [26], [27] to accelerate the convergence speed of the consensus methods. This paper is orthogonal to the momentum averaging-based approaches in the sense that it can be used in combination with the aforementioned momentum-based schemes, we refer the reader to the extended arXiv version of our paper [15] for details. There are also works that provide lower bounds on the distributed averaging time on a connected graph [13], [28], [29], [30]. In particular, it follows from these lower bounds that for the two-dimensional grid, even the best gossiping weights will not lead to an accelerated performance compared to baseline approaches. Indeed, for special graphs such as the two-dimensional grid, cycle graph or the line graph, ER weights will be similar to uniform weights due to the symmetries in the graph structure and consequently ER weights will not improve the performance compared to uniform weights. However, for graphs with asymmetries involving clusters or bottleneck edges along which the graph cut has low conductance, we expect ER weights lead to an improved performance, based on our numerical and theoretical results.

Outline. In Section II, we give a brief overview of randomized gossiping including uniform and ER-based gossiping methods. In Section III, we state our main contributions. In Section IV, we provide detailed arguments establishing the main results stated in Section III. In Section V, we provide numerical experiments illustrating that using ER weights can improve the performance of EXTRA and DPGA-

W algorithms for consensus optimization. In Section VI, we provide concluding remarks. Finally, we present some of the proofs and supporting results in Appendix A–B.

Notation. Let $|S|$ denote the cardinality of a set S, $\lfloor . \rfloor$ denote the floor function and $\mathbb{Z}_+$ be the set of nonnegative integers. We define $d_i \triangleq |\mathcal{N}_i|$ as the degree of $i \in \mathcal{N}$, and $m \triangleq |\mathcal{E}|$. Throughout the paper, $\mathcal{L} \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ denotes the weighted Laplacian of $\mathcal{G}$, i.e., $\mathcal{L}_{ii} = \sum_{j\in\mathcal{N}_i} w_{ij}$, $\mathcal{L}_{ij} = -w_{ij}$ if $j \in \mathcal{N}_i$, and equals to 0 otherwise. The diameter of a graph is $\mathcal{D} \triangleq \max_{i,j\in\mathcal{N}} d(i,j)$ where $d(i,j)$ is the shortest path on the graph between nodes i and j. The set $\mathbb{S}^n$ denotes the set of $n \times n$ real symmetric matrices. We use the notation $Z = [z_i]_{i=1}^n$ where z_i's are either the columns or rows of the matrix Z depending on the context. $\mathbf{1}$ is the column vector with all entries equal to 1, and $\mathbf{I}$ is the identity matrix. We let $||x||_p$ denote the L_p norm of a vector x for $p \geq 1$, and let $\|A\|_F$ denote the Frobenius norm of a matrix A. A square matrix A is *doubly stochastic* if all of its entries are non-negative and all its rows and columns sum up to 1. We say that a square matrix A is *weakly diagonally dominant* if it's diagonal entries A_{ii} satisfy the inequality $|A_{ii}| \geq \sum_{j\neq i} |A_{ij}|$ for every i. Let f and g be real-valued functions defined over positive integers. We say $f(n) = \mathcal{O}(g(n))$ if f is bounded above by g asymptotically, i.e., there exist constants $k_1 > 0$ and $n_0 \in \mathbb{Z}_+$ such that $f(n) \leq k_1 \cdot g(n)$ for all $n > n_0$. Similarly, we say $f(n) = \Omega(g(n))$ if there exist constants $k_2 > 0$ and $n_0 \in \mathbb{Z}_+$ such that $f(n) \geq k_2\, g(n)$ for every $n > n_0$; and we say $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = \mathcal{O}(g(n))$. Finally, $\log(x)$ denots the natural logarithm of x, and e_i is the i-th standard basis vector in $\mathbb{R}^n$ for $i = 1, 2, \ldots, n$.

II. Preliminaries

A. Randomized gossiping

Here we give an overview of randomized gossiping methods for the consensus problem. These methods can compute the average of node values over a network in an asynchronous and decentralized manner, for details see [13], [30].

Algorithm 1: Randomized Gossiping

```
1 Initialization: y^0 = [y_1^0, y_2^0, ..., y_n^0]^T ∈ R^n
2 for k ≥ 0 do
3     At time t_k, i ∈ N wakes up w.p. p_i = r_i / Σ_{j∈N} r_j
4     Picks j ∈ N_i randomly w.p. p_{j|i}
5     y_i^{k+1} ← (y_i^k + y_j^k)/2,   y_j^{k+1} ← (y_i^k + y_j^k)/2
```

Let $y^0 \in \mathbb{R}^n$ be a vector such that the i-th component y_i^0 represents the initial value at node $i \in \mathcal{N}$. The aim of the randomized gossiping algorithms is to have each node compute the average $\bar{y} \triangleq \sum_{i=1}^n y_i^0/n$ in a decentralized manner through an iterative procedure. At every iteration $k \in \mathbb{Z}_+$, each node $i \in \mathcal{N}$ possesses a local estimate y_i^k of the average to be computed and communicates with only randomly selected neighbors to update its estimate. The setup is that each node $i \in \mathcal{N}$ has an exponential clock ticking with rate $r_i > 0$ where the time between two ticks is exponentially distributed and independent of other nodes' clocks. A node wakes up when its clock ticks. Since all the clocks are independent, if a node wakes up at time $t_k \geq 0$, it is node i with probability (w.p.) $p_i \triangleq r_i / \sum_{j\in\mathcal{N}} r_j$. Given that the node i wakes up at time t_k, the conditional probability that it picks *one* of its neighbors $j \in \mathcal{N}_i$ to communicate with probability $p_{j|i} \in (0,1)$, where the probabilities $\{p_{j|i}\}_{j\in\mathcal{N}_i}$ are design parameters satisfying $\sum_{j\in\mathcal{N}_i} p_{j|i} = 1$. When either i wakes up and picks $j \in \mathcal{N}_i$ or vice versa, we say the edge (i,j) is activated. Once the edge (i,j) is activated, nodes i and j exchange their local variables y_i^k and y_j^k at time t_k and both compute the average $(y_i^k + y_j^k)/2$. This is illustrated in Algorithm 1 which admits an asynchronous implementation – see, e.g., [13].

Assuming *there are no self-loops for each $i \in \mathcal{N}$, let*

$$P_{ii} \triangleq 0;\ P_{ij} \triangleq p_i\ p_{j|i},\ \forall\ j \in \mathcal{N}_i;\ P_{ij} \triangleq 0,\ \forall j \in \mathcal{N} \setminus \mathcal{N}_i, \quad (2)$$

where P_{ij} is the (unconditional) probability that the edge (i,j) is activated by the node i. By definition, we have $\sum_{ij} P_{ij} \triangleq \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} P_{ij} = 1$. Let $\mathcal{A}(P)$ denote an asynchronous gossiping algorithm characterized by a probability matrix P as in (5) *for some set of probabilities $\{p_i\}_{i\in\mathcal{N}}$ and $\{p_{j|i}\}_{j\in\mathcal{N}_i}$ for $i \in \mathcal{N}$. The performance of $\mathcal{A}(P)$ is typically measured by the ε-averaging time, defined for any $\varepsilon > 0$ as*:

$$T_{ave}(\varepsilon, P) \triangleq \sup_{y^0\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \inf\left\{k : \mathbb{P}\left(\frac{\|y^k - \bar{y}\mathbf{1}\|}{\|y^0\|} \geq \varepsilon\right) \leq \varepsilon\right\}, \quad (3)$$

see, e.g., [13]. Suppose (i,j) is activated by node i, then we can write the update in Step 5 of the Algorithm 1 as

$$y^{k+1} = W^{(i,j)} y^k \quad \text{where} \quad W^{(i,j)} \triangleq I - \frac{(e_i - e_j)(e_i - e_j)^\top}{2}.$$

We also define

$$\overline{W}_P \triangleq \mathbb{E}_P[W^{(i,j)}] = \sum_{i,j\in\mathcal{N}} P_{ij} W^{(i,j)}, \quad (4)$$

which is the expected value of the random iteration matrix $W^{(i,j)}$ with respect to the distribution defined over $i \in \mathcal{N}$ and $j \in \mathcal{N}_i$. The following theorem from [13] shows that the second largest eigenvalue of $\overline{W}_P$ determines the ε-averaging time.

Theorem 1 ([13, Theorem 3])**.** *For a given $\mathcal{A}(P)$, the symmetric matrix $\overline{W}_P$ defined in* (4) *satisfies*

$$0.5\frac{\log(\varepsilon^{-1})}{\log([\lambda_{n-1}(\overline{W}_P)]^{-1})} \leq T_{ave}(\varepsilon, P) \leq 3\frac{\log(\varepsilon^{-1})}{\log([\lambda_{n-1}(\overline{W}_P)]^{-1})},$$

where $\lambda_{n-1}(\overline{W}_P)$ is the second largest eigenvalue of $\overline{W}_P$.

This result makes the connection between the convergence time of an asynchronous gossiping algorithm $\mathcal{A}(P)$ and the spectrum of the expected iteration matrix $\overline{W}_P$. It is therefore of interest to design P through carefully choosing the probabilities $\{p_i\}_{i\in\mathcal{N}}$ and $\{p_{j|i}\}_{j\in\mathcal{N}_i}$ for $i \in \mathcal{N}$ in order to get the best performance, i.e., the smallest ε-averaging time.

In this paper, we consider two different randomized gossiping algorithms: *uniform gossiping* and *ER gossiping* which differ in how the probabilities $\{p_i\}_{i\in\mathcal{N}}$ and $\{p_{j|i}\}_{j\in\mathcal{N}_i}$ for $i \in \mathcal{N}$ are selected. In particular, based on Theorem 1, we will study the second largest eigenvalue of the expected iteration matrix $\overline{W}_P$ corresponding to these two algorithms and compare their ε-averaging times.

B. Randomized uniform gossiping

In the *randomized uniform gossiping*, each node i wakes up with equal probability $p_i^u = \frac{1}{n}$, i.e., using uniform clock rates $r_i = r > 0$ for $i \in \mathcal{N}$. The superscript u stands for the uniform choice of clock rates. Then, node i picks the edge (i,j) with conditional probability $p_{j|i}^u = \frac{1}{d_i}$ for $j \in \mathcal{N}_i$; thus,

$$P_{ij}^u = p_i^u p_{j|i}^u = 1/(nd_i),$$

see, e.g., [31], [32]. One of the drawbacks of this approach is that it can be quite slow over graphs with a high *bottleneck ratio* [33] where, intuitively speaking, some "bottleneck edges" limit the spread of information over the underlying graph. A classical example of a graph with a high bottleneck ratio is the *barbell graph*. Barbell graphs are frequently studied within the consensus problem literature as they constitute a worst-case example in terms of both the mixing properties of random walks [5, Section 5] and the performance of distributed averaging algorithms (see, e.g., [4], [34]).

Barbell graphs consist of two complete subgraphs connected with an edge (see Figure 1). Let $K_{\tilde{n}}$ denote a complete graph with $\tilde{n}$ nodes, we will be denoting a barbell graph with $n = 2\tilde{n}$ nodes by $K_{\tilde{n}} - K_{\tilde{n}}$. Let (i^*, j^*) be the edge that connects the two complete subgraphs which we will be referring to as the *bottleneck edge*. This is the only edge that allows node values to be propagated between the two complete subgraphs; therefore, how frequently it is sampled is a key factor that determines the averaging time.

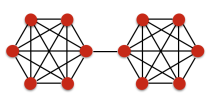

Fig. 1: Barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ with $n = 2\tilde{n} = 12$ nodes

The probability of sampling the *bottleneck edge* (i^*, j^*), with uniform weights can be computed explicitly:

$$P_{i^*j^*}^u = P_{j^*i^*}^u = \frac{1}{n}\frac{1}{d_{i^*}} = \frac{2}{n^2}. \tag{5}$$

This implies that it takes $\Theta(n^2)$ iterations in expectation to activate this edge, which is the underlying reason why the randomized uniform gossiping iterates converge slowly when n is large on the barbell graph. The effect of bottleneck edges on the performance of gossiping algorithms has been recently studied experimentally by Aybat and Gürbüzbalaban [1] on different topologies including the barbell and small-world graphs. The authors proposed *ER gossiping* where the edges are sampled with non-uniform probabilities proportional to effective resistances $\{R_{ij}\}_{(i,j)\in\mathcal{E}}$ and the numerical experiments in [1] showed that this can lead to significant performance improvement over graphs with bottleneck edges, such as barbell graphs. We next describe this method.

C. Effective-resistance (ER) gossiping

In the ER gossiping, each $i \in \mathcal{N}$ wakes up with probability $p_i^r = \frac{\sum_{j\in\mathcal{N}_i} R_{ij}}{2\sum_{(i,j)\in\mathcal{E}} R_{ij}}$, i.e., setting clock rate $r_i = \sum_{j\in\mathcal{N}_i} R_{ij}$ for $i \in \mathcal{N}$, and node i picks (i,j) with conditional probability $p_{j|i}^r = \frac{R_{ij}}{\sum_{j\in\mathcal{N}_i} R_{ij}}$ for all $j \in \mathcal{N}_i$; thus, ER gossiping corresponds to the unconditional probabilities

$$P_{ij}^r = p_i^r p_{j|i}^r = \frac{R_{ij}}{2\sum_{(i,j)\in\mathcal{E}} R_{ij}} = \frac{R_{ij}}{2(n-1)} = P_{ji}^r,$$

for all $(i,j) \in \mathcal{E}$ where the third equality follows from Foster's Theorem which says that $\sum_{(i,j)\in\mathcal{E}} R_{ij} = (n-1)$ – see, e.g., [35]. This choice of sampling probabilities can lead to bottleneck edges being more frequently sampled. We illustrate this fact on the barbell graph $(K_{\tilde{n}} - K_{\tilde{n}})$: Note that the unconditional probability of sampling the bottleneck edge (i^*, j^*) is given explicitly as

$$P_{i^*j^*}^r = P_{j^*i^*}^r = \frac{R_{i^*j^*}}{2(n-1)} = \frac{1}{2(n-1)}, \tag{6}$$

where $n = 2\tilde{n}$ and we used the fact that $R_{i^*j^*} = 1$ (see the proof of Lemma 17 for the derivation of (6)). Hence, comparing (5) and (6), we see that ER weights allow sampling of the bottleneck edge (i^*, j^*) more frequently, by a factor of $\Theta(n)$, than the uniform gossiping on $K_{\tilde{n}} - K_{\tilde{n}}$. Intuitively speaking, this is the reason why ER gossiping can be efficient on barbell graphs. Numerical experiments provided in [1] support this intuition where ER gossiping outperforms uniform gossiping over an unweighted barbell graph as well as small-world graphs, which are random graphs that arise frequently in real-world applications such as social networks.

Despite the empirical success of *ER gossiping* in practice, theoretical results supporting its practical performance have been lacking in the literature. The purpose of this paper is to provide rigorous convergence guarantees for ER gossiping algorithms on certain network topologies (see Section III for our main results' statements and Section IV for the proofs) and to present further numerical evidence that ER gossiping, beyond distributed averaging, can also improve the practical performance of distributed methods for consensus optimization (Section V). Indeed, in our analysis, we consider connected graphs characterized by their diameter $\mathcal{D} \in \mathbb{Z}_+$, barbell graphs and c-barbell graphs which are generalizations of barbell graphs. More specifically, a c-barbell graph $(K_{\tilde{n}}^c)$ for $c \geq 2$ is a path of c equal-sized complete graphs $(K_{\tilde{n}})$ [36]. In the special case, when $c = 2$, a c-barbell graph is equivalent to the barbell graph. We show that for these graphs, ER gossiping has provably better convergence properties than uniform gossiping in terms of ε-averaging times. Precise results will be stated in the next section.

III. Main Results

In this section, we state our main theoretical results: We provide performance bounds for the ER gossiping in terms of ε-averaging time $T_{ave}(\varepsilon, P^r)$. Our results highlight the performance improvements obtained with this approach.

Our first result concerns c-barbell graphs where we focus on the ε-averaging times of uniform and ER gossiping algorithms. To the best of our knowledge, for c-barbell graphs, an analytical formula for the second largest eigenvalue $\overline{W}_P$ is not analytically available; therefore, in our analysis we estimate this eigenvalue based on graph conductance techniques (see

Section IV-A for details) which leads to the following lower and upper bounds on the ε-averaging times.

Result 1. *Given $\varepsilon > 0$, and $\tilde{n}, c \in \mathbb{Z}_+$ such that $c \geq 2$, asynchronous randomized gossiping algorithms $\mathcal{A}(P^u)$ and $\mathcal{A}(P^r)$ on a c-barbell graph with $n = \tilde{n}c$ satisfy*

$$\Theta(c^2\tilde{n}^3 \log(1/\varepsilon)) \leq T_{ave}(\varepsilon, P^u) \leq \Theta(c^4\tilde{n}^6 \log(1/\varepsilon)), \quad (7)$$
$$\Theta(c^2\tilde{n}^2 \log(1/\varepsilon)) \leq T_{ave}(\varepsilon, P^r) \leq \Theta(c^4\tilde{n}^4 \log(1/\varepsilon)). \quad (8)$$

These bounds from Result 1 for the c-barbell graph show that, for any given precision $\varepsilon > 0$, using effective resistances one can improve upper and lower bounds on the averaging times by a factor of $\Theta(n)$ and $\Theta(n^2)$, respectively.

The next result shows that for the case of barbell graphs (when $c = 2$) the ER gossiping is in fact faster by a factor of $\Theta(n)$. The proof idea is based on computing the eigenvalues of $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ explicitly via exploiting symmetry group properties of barbell graphs and showing that the lower bounds in (7)–(8) are attained for $c = 2$.

Result 2. *Given $\varepsilon > 0$ and $n \in \mathbb{Z}_+$, let $n = 2\tilde{n}$. The ε-averaging times of asynchronous gossiping algorithms $\mathcal{A}(P^r)$ and $\mathcal{A}(P^u)$ on barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ satisfy the equality:*

$$T_{ave}(\varepsilon, P^r) = \Theta(1/n)\ T_{ave}(\varepsilon, P^u).$$

A natural question is whether it is possible to further improve the ER gossiping bounds for barbell graphs; however, in the next result, we show that this is not possible as long as the matrix P is symmetric –thus, ER gossiping is optimal. Finally, we also obtain ε-averaging bounds for a more general class of connected graphs depending on their diameters.

Result 3. *Given $\varepsilon > 0$ and $n \in \mathbb{Z}_+$, let $n = 2\tilde{n}$. Among all the gossiping algorithms $\mathcal{A}(P)$ with a symmetric P on the barbell graph, $K_{\tilde{n}} - K_{\tilde{n}}$, randomized ER gossiping leads to $T_{ave}(\varepsilon, P^r) = \Theta(n^2 \log(1/\varepsilon))$, which is optimal with respect to ε and n, and cannot be improved.*

In a more general setting, let $\mathcal{G}$ be a connected graph with diameter $\mathcal{D} \in \mathbb{Z}_+$. The ε-averaging time of $\mathcal{A}(P^r)$ satisfies

$$T_{ave}(\varepsilon, P^r) = \mathcal{O}(\mathcal{D}n^3) \log(\varepsilon^{-1}).$$

Remark 2. *The ε-averaging time of randomized gossiping with lazy Metropolis weights*[2] *on any graph is $\mathcal{O}(n^3 \log(1/\varepsilon))$; while, for the barbell graph, Metropolis weights perform similar to uniform weights; both require $\Theta(n^3 \log(1/\varepsilon))$ time which can be improved to $\Theta(n^2 \log(1/\varepsilon))$ by ER gossiping.*

Remark 3. *If the diameter $\mathcal{D} \leq 11$, our bounds for ER gossiping improve upon that of the randomized gossiping with lazy Metropolis weights by a (small) constant factor (see Remark 12). Note $\mathcal{D} = 3$ for barbell graphs and $\mathcal{D} \leq 11$ is also reasonable for mid-size small-world graphs which are random graphs that arise frequently in real-world applications [37]. For instance, Cont* et al. *[37] show that the diameter $\mathcal{D}$ of the randomized community-based small-world graphs admits $2\log(n)$ upper bound almost surely; hence, for these graphs $\mathcal{D} \leq 11$ almost surely for $n \leq 240$. Indeed, we empirically observe that randomly generated small-world graphs with parameters $n = \{5k : k = 1, \ldots, 5\}$ and $m = \lfloor 0.2(n^2 - n) \rfloor$ using the methodology described in the numerical experiments in Section V-A satisfy $\mathcal{D} \leq 5$ on average over 10^4 independent and identically distributed (i.i.d.) samples.*

[2]For lazy Metropolis weights see (15) and the paragraph after.

IV. PROOFS OF MAIN RESULTS

In order for both uniform and ER gossiping methods to have the same expected number of node wake-ups in a given time period, one should have $r_i = r = 2(n-1)/n$ for $i \in \mathcal{N}$ within the uniform gossiping model –recall that $r_i = \sum_{j\in\mathcal{N}_i} R_{ij}$ for $i \in \mathcal{N}$ for ER gossiping; hence, the rate of both Poisson processes will be the same, i.e., $\sum_{i\in\mathcal{N}} r_i = 2(n-1)$. We note that the number of clock ticks $k \in \mathbb{Z}_+$ can be converted to *absolute time* easily with standard arguments (simply dividing k by $\sum_{i\in\mathcal{N}} r_i$ to get the expected time of the k-th tick), e.g., see [13, Lemma 1]. This allows us to use the number of iterations (clock ticks) to compare asynchronous algorithms.

It can be easily verified that for a given $\mathcal{A}(P)$, the expected iteration matrix defined in (4) satisfies

$$\overline{W}_P = I - \tfrac{1}{2}D + \tfrac{1}{2}(P + P^\top), \quad (9)$$

where D is a diagonal matrix with i-th entry $D_i \triangleq \sum_{j\in\mathcal{N}_i}(P_{ij} + P_{ji})$. Note $W^{(i,j)}$ defined in Section II-A is a doubly stochastic, non-negative and weakly diagonally dominant matrix for all $i \in \mathcal{N}$ and $j \in \mathcal{N}_i$; therefore, $\overline{W}_P$, which is a convex combination of $W^{(i,j)}$ matrices, is also a doubly stochastic, non-negative and weakly diagonally dominant matrix. It follows then from the Gershgorin's Disc Theorem (see e.g. [38]) that all the eigenvalues of $\overline{W}_P$ are non-negative. Moreover, since $\overline{W}_P$ is a non-negative doubly stochastic matrix, its largest eigenvalue $\lambda_n(\overline{W}_P) = 1$. Plugging in P^u and P^r for P in this identity respectively leads immediately to the following result.

Lemma 4. *The matrices $\overline{W}_{P^r} = \mathbb{E}_{P^r}[W^{(i,j)}]$ and $\overline{W}_{P^u} = \mathbb{E}_{P^u}[W^{(i,j)}]$ satisfy the following identities:*

$$\overline{W}_{P^u} = I - \frac{1}{2}D^u + \frac{P^u + (P^u)^\top}{2}, \quad \overline{W}_{P^r} = I - \frac{1}{2}D^r + P^r,$$

where D^u and D^r are diagonal matrices satisfying $[D^u]_{ii} \triangleq \sum_{j\in\mathcal{N}_i}(P^u_{ij} + P^u_{ji})$, $[D^r]_{ii} = \frac{1}{(n-1)} R_i$ where $R_i \triangleq \sum_{j\in\mathcal{N}_i} R_{ij}$.

Recall the definition of $T_{ave}(\varepsilon, P)$ given in (3), i.e., ε-averaging time of an asynchronous gossiping algorithm $\mathcal{A}(P)$ characterized by a probability matrix P. According to Theorem 1, to compare uniform and ER gossiping methods introduced in Section II, it is sufficient to estimate the second largest eigenvalues of $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ and compare them. In the rest of this section, we discuss estimating the second largest eigenvalues of $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ based on the notions of graph conductance and hitting times when the eigenvalues are not readily available in closed form. We will also discuss some examples for which we can explicitly compute the eigenvalues.

It is worth emphasizing that since the matrices $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ are symmetric and doubly stochastic, they can both be viewed as the probability transition matrix of a *reversible Markov Chain* on the graph $\mathcal{G}$, both with a uniform stationary distribution. We saw that depending on the type of randomized

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3161201, IEEE Transactions on Control of Network Systems

gossiping, the sampling probabilities of the bottleneck edge can differ significantly –by a factor of $\Theta(n)$ on barbell graphs implied by (5) and (6). A similar effect can also be observed for the Markov chains defined by the transition probability matrices $\overline{W}_{P^u}$ and $\overline{W}_{P^r}$. In fact, by an explicit computation based on Lemma 4 (see Lemma 17 for details), we get

$$[\overline{W}_{P^u}]_{i^*j^*} = \frac{2}{n^2}, \quad [\overline{W}_{P^r}]_{i^*j^*} = \frac{1}{2(n-1)}.$$

That is, the probability of moving from one complete subgraph to the other is significantly larger (by a factor of $\Theta(n)$) for the Markov chain corresponding to $\overline{W}_{P^r}$ than that of the chain with $\overline{W}_{P^u}$. Intuitively speaking, this fact allows the ER-based chain to traverse between the complete subgraphs faster when n is large, leading to faster averaging over the nodes. This will be formalized and proven in the next subsection, where we study gossiping algorithms over *barbell* and *c-barbell graphs*.

A. Proof of Result 1 via conductance-based analysis

Probability transition matrices on graphs have been studied well; in particular, there are some combinatorial techniques to bound their eigenvalues based on *graph conductance* [5] as well as some algebraic techniques that allow one to compute all the eigenvalues explicitly exploiting symmetry groups of a graph [39] as we shall discuss in Section IV-B.

The notion of graph conductance is tied to a transition matrix W over a graph which corresponds to a reversible Markov chain admitting an arbitrary stationary distribution π. It can be viewed as a measure of how hard it is for the Markov chain to go from a subgraph to its complement in the worst case. The notion of graph conductance allows us to provide bounds on the mixing time of the corresponding Markov chain as we discuss below.

Definition 5 (Conductance). *Let W be the transition matrix of a reversible Markov chain[3] on the graph $\mathcal{G}$ with a stationary distribution $\pi = \{\pi_i\}_{i=1}^n$. The conductance Φ is defined as*

$$\Phi(W) \triangleq \min_{S\subset\mathcal{N}: S, S^c \neq \emptyset} \frac{\sum_{i\in S, j\in S^c} \pi_i W_{ij}}{\min\{\pi(S), \pi(S^c)\}} \tag{10}$$

where $\pi(S) \triangleq \sum_{i\in S} \pi_i$.

Given a transition matrix W, the relation between conductance $\Phi(W)$ and the second largest eigenvalue $\lambda_{n-1}(W)$ is well-known and given by the *Cheeger inequalities*:

$$1 - 2\Phi(W) \le \lambda_{n-1}(W) \le 1 - \Phi^2(W), \tag{11}$$

–see, e.g., [40, Proposition 6]. Therefore, larger conductance leads to faster averaging, i.e., shorter $T_{ave}(\varepsilon, P)$, in light of Theorem 1. In particular, we can get lower and upper bounds on the averaging time for both uniform and ER gossiping methods using the Cheeger's inequality. We study the performance bounds for these gossiping algorithms over *c-barbell graphs*; and our next result shows $\Theta(n)$ improvement on the conductance of effective resistance-based transition probabilities $\overline{W}_{P^r}$ compared to uniform probabilities $\overline{W}_{P^u}$ on a c-barbell graph with $n = c\tilde{n}$ nodes.

[3]That is $\pi_i W_{ij} = \pi_j W_{ji}$ for all $i, j \in \mathcal{N}$.

Proposition 6. *Given $\tilde{n}, c \in \mathbb{Z}_+$ such that $c \ge 2$, consider the two Markov chains on the c-barbell graph with $n = \tilde{n}c$ nodes defined by the transition matrices $\overline{W}_{P^u}$ and $\overline{W}_{P^r}$. Let $c_* = \left(\lfloor \frac{c}{2} \rfloor\right)^{-1}$. The conductance values are given by*

$$\Phi(\bar{W}_{P^u}) = \frac{c_*}{c\tilde{n}^3}, \quad \Phi(\bar{W}_{P^r}) = \frac{c_*}{2\tilde{n}(c\tilde{n}-1)}. \tag{12}$$

Remark 7. *Since a barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ is a special case of a c-barbell graph with $c = 2$ and $n = 2\tilde{n}$, Proposition 6 implies that $\Phi(\overline{W}_{P^u}) = \frac{4}{n^3}$ and $\Phi(\overline{W}_{P^r}) = \frac{1}{n(n-1)}$.*

Given the transition matrix W, by taking the logarithm of the Cheeger inequalities in (11), for $\Phi(W) \le 1/2$, we obtain

$$-\log(1-\Phi^2(W)) \le \log(\lambda_{n-1}^{-1}(W)) \le -\log(1-2\Phi(W)). \tag{13}$$

Then, choosing $W = \overline{W}_{P^u}$ and $W = \overline{W}_{P^r}$ above, applying Theorem 1 and Proposition 6 and noting $-\log(1-x) \approx x$ for x close to 0, leads to the lower and upper bounds on the averaging time of uniform and ER gossiping algorithms as shown in Result 1 of our main results section (Section III).

Although this analysis is also applicable to other graphs with low conductance, it does not typically lead to tight estimates, i.e., the lower and upper bounds do not match in terms of their dependency on n. In the next section, we show that for the case of barbell graphs, we get tight estimates on the averaging time by computing the eigenvalues of the averaging matrices $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ *explicitly*. More precisely, we will show in Proposition 9 that the lower bounds in (7)–(8) are tight for $c = 2$ in the sense that $T_{ave}(\varepsilon, P^u) = \Theta(n^3)$ and $T_{ave}(\varepsilon, P^r) = \Theta(n^2)$ and the effective resistance-based averaging is faster by a factor of $\Theta(n)$ which will imply Result 2.

B. Proof of Result 2 via spectral analysis

Eigenvalues of probability transition matrices defined on barbell graphs are studied in the literature. Consider the *edge-weighted* barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ with $n = 2\tilde{n}$ nodes, where $w = [w_{ij}]_{(i,j)\in\mathcal{E}}$ is the vector of edge weights that have positive entries. Suppose each node has a self-loop, e.g., see Fig. 2. Let (i^*, j^*) be the edge that connects the two complete subgraphs. The result [39, Prop. 5.1] gives an explicit formula for the eigenvalues of a probability transition matrix W with transition probabilities proportional to edge weights, i.e., $W_{ij} = w_{ij}/\sum_{j\in\mathcal{N}_i} w_{ij}$ where w_{ij} satisfy the following assumptions: $w_{i^*i^*} = w_{j^*j^*} = 0$, $w_{i^*j^*} = A$, $w_{i^*j} = w_{j^*i} = B$ for all $j \in \mathcal{N}_{i^*} \setminus \{j^*\}$ and $i \in \mathcal{N}_{j^*} \setminus \{i^*\}$, $w_{ij} = C$ for all (i,j) in each $K_{\tilde{n}}$ such that $i \ne j$ and $i, j \notin \{i^*, j^*\}$, and $w_{ii} = D$ for $i \in \mathcal{N} \setminus \{i^*, j^*\}$ for some $A, B, C, D > 0$. Note we cannot immediately use this result to compute the eigenvalues of the transition matrices $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ defined in Lemma 4. Mainly because all the diagonal entries of $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$ being strictly positive breaks the $w_{i*i*} = w_{j*j*} = 0$ assumption of [39, Prop. 5.1]. In Proposition 8, we adapt [39, Prop. 5.1] to our setting with some minor modifications to allow $w_{i^*i^*} = w_{j^*j^*} = G$ for any $G > 0$ so that it becomes applicable to $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$[4].

[4]We refer the reader to extended arXiv version of our paper [15] for further details of the proof.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3161201, IEEE Transactions on Control of Network Systems

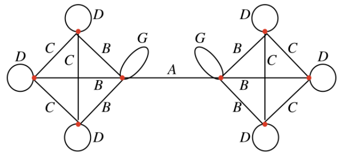

Fig. 2: An edge-weighted barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ with edge weights $A, B, C, D, G > 0$ for $\tilde{n} = 4$.

Proposition 8 (Generalization of Proposition 5.1 in [39])**.** *Consider the edge-weighted barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ with $n = 2\tilde{n}$ nodes. Let (i^*, j^*) be the edge that connects the two complete subgraphs. Assume that weights are of the form $w_{i^*i^*} = w_{j^*j^*} = G$, $w_{i^*j^*} = A$, $w_{i^*j} = w_{j^*i} = B$ for all $j \in \mathcal{N}_{i^*} \setminus \{j^*\}$ and $i \in \mathcal{N}_{j^*} \setminus \{i^*\}$, $w_{ij} = C$ for all (i,j) in each $K_{\tilde{n}}$ such that $i \neq j$ and $i, j \notin \{i^*, j^*\}$, and $w_{ii} = D$ for $i \in \mathcal{N} \setminus \{i^*, j^*\}$ for some $A, B, C, D, G > 0$. Consider the transition matrix W associated to this graph with entries $W_{ij} = w_{ij}/\sum_{j\in\mathcal{N}_i} w_{ij}$, then the eigenvalues of W are*

- $\lambda_a \triangleq 1$ *with multiplicity one,*
- $\lambda_b \triangleq -1 + \frac{A+G}{A+G+E} + \frac{F}{F+B}$ *with multiplicity one,*
- $\lambda_c \triangleq \frac{D-C}{B+F}$ *with multiplicity $n-4$,*
- $\lambda_\pm \triangleq \frac{1}{2}\Big(\frac{F}{B+F} + \frac{G-A}{A+E+G} \pm \sqrt{S}\Big)$,

where $E \triangleq (\tilde{n}-1)B$, $F \triangleq D + (\tilde{n}-2)C$ and $S \triangleq \big(\frac{F}{B+F} + \frac{G-A}{A+E+G}\big)^2 - \frac{4(FG-BE-AF)}{(B+F)(A+E+G)}$.

Based on this result, in Proposition 9, we characterize the second largest eigenvalue of the transition matrices $\overline{W}_{P^u}$ and $\overline{W}_{P^u}$ – the proof can be found in the Appendix.

Proposition 9. *Consider Markov chains on the barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ with transition matrices $\overline{W}_{P^r}$ and $\overline{W}_{P^u}$. The second largest eigenvalues of these matrices are given by*

$$\lambda_{n-1}(\overline{W}_{P^r}) = 1 - \Theta(\frac{1}{n^2}), \quad \lambda_{n-1}(\overline{W}_{P^u}) = 1 - \Theta(\frac{1}{n^3}).$$

Result 2 follows as a direct consequence of Proposition 9 and Theorem 1. Thus, we establish that that averaging time with resistance weights is $\Theta(n)$ faster on a barbell graph.

C. Proof of Result 3 via hitting and mixing times

Before giving a formal definition of the ε-mixing time, we introduce the total variation (TV) distance between two probability measures p and q defined on the set of nodes $\mathcal{N} = \{1, 2, \ldots, n\}$. TV distance between p and q is defined as $\|p-q\|_{TV} \triangleq \|p-q\|_1/2$. Given a Markov chain $\mathcal{M}$ with a probability transition matrix W and stationary distribution π, ε-mixing time is a measure of how many iterations are needed for the probability distribution of the chain to be ε-close to the stationary distribution in the TV distance. A related notion is the hitting time which is a measure of how fast the Markov chain travels between any two nodes.

Definition 10. *(Mixing time and hitting times) Given $\varepsilon > 0$ and a Markov chain with probability transition matrix W and stationary distribution π, the ε-mixing time is defined as*

$$T_{mix}(\varepsilon, W) \triangleq \inf_{k\geq 0}\Big\{ \sup_{p\geq 0: \|p\|_1=1} \|(W^k)^\top p - \pi\|_{TV} \leq \varepsilon \Big\},$$

and the hitting time $H_W(i \to j)$ is the expected number of steps until the Markov chain reaches j starting from i.

Mixing-times and averaging times are closely related. In fact, given probability transition matrix W, it is known that $T_{ave}(\varepsilon, W)$ and $T_{mix}(\varepsilon, \tilde{W})$ admit the same bounds up to $n \log n$ factors [13, Theorem 7] for $\tilde{W} = \frac{I+W}{2}$.[5] Hence, designing algorithms with a smaller mixing time, often leads to better algorithms for distributed averaging (see also [30]). It is also known that mixing time is closely related to hitting times [41, Theorem 1.1]. Next, we show the first part of Result 3, i.e., $T_{ave}(\varepsilon, P^r) = \Theta(n^2 \log(1/\varepsilon))$ is optimal among all $\mathcal{A}(P)$ with a symmetric P. Note P is symmetric implies that it is doubly stochastic. For large n and doubly stochastic P, by [13, Corollary 1], we have $T_{ave}(\varepsilon, P) = \Theta\Big(\frac{n\log(1/\varepsilon)}{1-\lambda_{n-1}(P)}\Big)$. On the other hand, Roch proved in [28, Section 3.3.1] that any symmetric doubly stochastic P matrix on the barbell graph with n nodes satisfies $\frac{1}{1-\lambda_{n-1}(P)} = \Omega(n)$. Inserting this estimate into the expression for the averaging time, we obtain $T_{ave}(\varepsilon, P) = \Omega\left(n^2 \log(1/\varepsilon)\right)$ for any $\mathcal{A}(P)$ with symmetric P on barbell graphs. We conclude that the averaging time of the ER-based gossiping on the barbell graph, which satisfies $T_{ave}(\varepsilon, P^r) = \Theta(n^2 \log(1/\varepsilon))$ by Proposition 9 and Theorem 1, is optimal with respect to its dependency to n and ε among all symmetric choices of the P matrix.

Next, given any connected graph $\mathcal{G}$, we obtain a bound on the second largest eigenvalue of the $\overline{W}_{P^r}$ and show that the averaging time with effective resistance weights $T_{ave}(\varepsilon, P^r) = \mathcal{O}\left(\mathcal{D}n^3 \log(1/\varepsilon)\right)$ where $\mathcal{D}$ is the diameter of the graph.

Theorem 11. *Let $\mathcal{G}$ be a graph with diameter $\mathcal{D}$. The second largest eigenvalue of $\overline{W}_{P^r}$ satisfies $\lambda_{n-1}(\overline{W}_{P^r}) \leq 1 - \frac{1}{6\mathcal{D}n^3}$.*

Proof: It follows from our discussion in Section IV that $\overline{W}_{P^r}$ is non-negative and doubly stochastic (see the paragraph before Lemma 4). Therefore, for analysis purposes, we can interpret $\overline{W}_{P^r}$ as the transition matrix of a Markov chain $\mathcal{M}$ whose stationary distribution π is the uniform distribution. Our analysis is based on relating the eigenvalues of $\overline{W}_{P^r}$ matrix to the hitting times of the Markov chain $\mathcal{M}$ where we follow the proof technique of [42, Lemma 2.1]. By Lemma 13 from the appendix, we get $H_{\overline{W}_{P^r}}(i \to j) \leq n\frac{2(n-1)}{R_{ij}}$ if $j \in \mathcal{N}_i$. For any graph, it is also known that[6] $\min_{i,j} R_{ij} \geq \frac{2}{n}$. Therefore, for any neighbors i and j, $H_{\overline{W}_{P^r}}(i \to j) \leq n^2(n-1)$. For any two vertices i and j not necessarily neighbors, $i \neq j$, let $v_0(= i), v_1, \ldots, v_\ell(= j)$ be the shortest path connecting i and j. Then, by the subadditivity property of hitting times, for any $i, j \in \mathcal{N}$, we obtain $H_{\overline{W}_{P^r}}(i \to j) \leq \ell n^2(n-1) \leq \mathcal{D}n^2(n-1)$. It follows from an analysis similar to [44] that

$$T_{mix}(\frac{1}{8}, \overline{W}_{P^r}) \leq 8 \max_{i,j\in\{1,\ldots,n\}} H_{\overline{W}_{P^r}}(i \to j) + 1 \leq 8\mathcal{D}n^3. \quad (14)$$

[5]Note [13, Theorem 7] uses absolute time whereas we used number of node wake-ups to define ϵ-averaging and ϵ-mixing times; therefore, we multiplied $\log(n)$ factor in [13, Theorem 7] by $\sum_{i\in\mathcal{N}} r_i = 2(n-1)$ to convert absolute times to number of node wake-ups.

[6]This follows directly from the Rayleigh's monotonicity rule [6] which says that if an edge is removed from a graph, effective resistance on any edge can only increase. Therefore, the complete graph provides a lower bound for R_{ij} where $R_{ij} = 2/n$ (see also [43]).

From [44, eqn. (12.12)], we also have

$$T_{mix}(\frac{1}{8}, \overline{W}_{P^r}) \geq \left(\frac{1}{1-\lambda_{n-1}(\overline{W}_{P^r})} - 1\right)\log(4).$$

Combining this with the estimate (14) implies directly $\lambda_{n-1}(\overline{W}_{P^r}) \leq 1 - 1/(6\mathcal{D}n^3)$, which proves the claim. ■

Metropolis vs ER gossiping: Given a connected $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, suppose there are no self-loops, i.e., $(i,i) \notin \mathcal{E}$ for $i \in \mathcal{N}$. Uniform weights $p^u_{j|i} = \frac{1}{d_i}$ can result in slow mixing on some graphs such as the barbell graph (see Proposition 9) or other graphs like lollipop graphs [5] which have both high degree and low degree nodes together. A popular alternative to uniform weights $\{p^u_{j|i}\}_{j\in\mathcal{N}_i}$ for $i \in \mathcal{N}$ is the Metropolis weights defined as

$$M_{ij} \triangleq \begin{cases} \frac{1}{\max(d_i,d_j)} & \text{if } (i,j) \in \mathcal{E}, \\ 1 - \sum_{j\in\mathcal{N}_i\setminus i} \frac{1}{\max(d_i,d_j)} & \text{if } i = j, \\ 0 & \text{else.} \end{cases} \tag{15}$$

Let M denote the matrix whose entries are the Metropolis weights M_{ij}. The weights determined by the matrix $\widetilde{M} \triangleq \frac{I+M}{2}$ are also popular in the distributed optimization practice [42] which is referred to as the lazy version of the Metropolis weights. The matrix $\widetilde{M}$ is symmetric and positive semi-definite, while the matrix M may have negative eigenvalues that can be close to -1 (therefore, using M can be problematic for the convergence of some distributed algorithms, see e.g. [45]). Combined with uniform wake-up of nodes, this leads to the following wake-up probabilities for the Metropolis weights based system: $P^{\widetilde{m}}_{ij} \triangleq \frac{1}{n}\widetilde{M}_{ij}$, and the associated matrix $\overline{W}_{P^{\widetilde{m}}} \triangleq \mathbb{E}_{P^{\widetilde{m}}}[W^{(i,j)}] = \sum_{ij} P^{\widetilde{m}}_{ij} W^{(i,j)}$. In particular, for any connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with n nodes, we have the following guarantees from [42, Lemma 2.1] on the lazy Metropolis weights:

$$\max_{i,j\in\mathcal{N}} H_{\widetilde{M}}(i \to j) \leq 12n^2, \lambda_{n-1}(\widetilde{M}) \leq 1 - \frac{1}{72n^2}. \tag{16}$$

By (9), we have also $\overline{W}_{P^{\widetilde{m}}} = (1-\frac{1}{n})I + \frac{1}{n}\widetilde{M}$. Therefore, from (16), we get the bound $\lambda_{n-1}(\overline{W}_{P^{\widetilde{m}}}) \leq 1 - \frac{1}{71n^3}$, for any connected graph $\mathcal{G}$. Therefore, we conclude from Theorem 1 that the ε-averaging time of Metropolis weights-based gossiping on any graph is $\mathcal{O}(n^3\log(1/\varepsilon))$ – again using the fact that $-\log(1-x) \approx x$ for x close to 0. That said, for barbell graphs, Metropolis weights perform similar to uniform weights; both require $\Theta(n^3\log(1/\varepsilon))$ time which is improved by the effective resistance-based weights to $\Theta(n^2\log(1/\varepsilon))$. This completes the proof of Result 3.

Remark 12. *Comparing the inequalities $\lambda_{n-1}(\overline{W}_{P^r}) \leq 1 - \frac{1}{6\mathcal{D}n^3}$ and $\lambda_{n-1}(\overline{W}_{P^{\widetilde{m}}}) \leq 1 - \frac{1}{71n^3}$, we see that for $\mathcal{D} \leq 11$, the upper bound on $\lambda_{n-1}(\overline{W}_{P^r})$ will be smaller than the upper bound for $\lambda_{n-1}(\overline{W}_{P^{\widetilde{m}}})$. Therefore, performance bounds obtained on the ε-averaging time through Theorem 1 for ER weights will be better than those of Metropolis weights by a (small) constant factor for $\mathcal{D} \leq 11$.*

V. Numerical Experiments

In this section, we demonstrate the benefits of using effective resistances for solving the consensus problem and also within DPGA-W [19] and EXTRA [45] algorithms for consensus optimization.

A. Consensus exploiting effective resistances

Gossiping algorithms have been studied extensively and there have been a number of approaches [30], [46], [47], [48], [49], [50], [51]. In light of Theorem 1, among all the algorithms $\mathcal{A}(P)$ with a symmetric P, the matrix P^{opt} that minimizes the second largest eigenvalue, i.e. $\lambda_{n-1}(\overline{W}_P)$, is the fastest. The gossiping algorithm $\mathcal{A}(P^{\text{opt}})$ with optimal choice of the probability matrix P^{opt} is called the Fastest Mixing Markov Chain (FMMC) in the literature [29]. In [13], Boyd et al. propose a distributed subgradient method to compute the matrix P^{opt}. This method requires a decaying step size and computation of the subgradient of the objective $\lambda_{n-1}(\overline{W}_P)$ with respect to the decision variable P at every iteration which itself requires solving a consensus problem at every iteration. This can be expensive in practice in terms of average number of communications required per node, and its convergence to P^{opt} can be slow with at most sublinear convergence rate [13]. In contrast, ER probabilities P^r are optimal for some graphs (such as the barbell graph, see Result 3) and can be computed efficiently with the normalized D-RK algorithm (see the arXiv version of our paper [15] for details) which admits linear convergence guarantees. Therefore, ER weights can serve as a computationally efficient alternative to optimal weights for consensus. For illustrating this point, we compare communication requirements per node for ER gossiping and FMMC on barbell and small-world graphs. This comparison consists of two stages: (i) pre-computation stage (where the probability matrices P^r and P^{opt} are computed up to a given tolerance) (ii) asynchronous consensus stage (where we run ER and FMMC with probability matrices P^r an P^{opt} obtained from the previous stage to solve a consensus problem).

First, we implement subgradient method with decaying step size $\alpha_k = R/k$ from [13] where R is tuned to the graph to achieve the best performance and stop the computation of matrix of FMMC at step k if the iterate P_k satisfies $\frac{||P_k - P^{opt}||_F}{||P^{opt}||_F} \leq \epsilon_1$ where ϵ_1 is the given precision level.[7] Similarly, we compute $\mathcal{L}^\dagger$ for ER and stop the normalized D-RK algorithm when the iterate X^k at step k satisfies $\frac{\|X^k - \mathcal{L}^\dagger\|_F}{\|\mathcal{L}^\dagger\|_F} \leq \epsilon_1$. Since the distributed subgradient method of [13] is based on synchronous computations, we also implemented the normalized D-RK algorithm with synchronous computations for fairness of comparison. We define the *communication* for a node as a contact with its neighbour either to compute an average of their state vectors or to update the matrix P_k at any iteration.

We compared both of the algorithms based on their communication performances on stage-i an stage-ii. In particular, we considered the number of communications required per node to obtain the matrix P_k for ER and FMMC at stage-i and at stage-ii, we generated 1000 instances of y^0_i to start consensus and compare the average number of

[7]The optimal probability matrix P^{opt} which serves as a baseline in the stopping criterion is estimated accurately by solving the semi-definite program (SDP) [13, eqn. (53)] directly using a centralized method, MOSEK, and computations required to solve this SDP is not counted as a part of the communication cost we report for FMMC in Tables I-II.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3161201, IEEE Transactions on Control of Network Systems

communications per node required to achieve y_i^k satisfying $\frac{||y^k-\bar{y}||}{||\bar{y}||} \leq \epsilon_2$ where ϵ_2 is the tolerance level. For the barbell graph, the initial state vector y_i^0 for consensus is sampled from the normal distribution $\mathbf{N}(500, 10)$ if $i \in \mathcal{N}_L$ and from $\mathbf{N}(-500, 10)$ if $i \in \mathcal{N}_R$ where tolerance levels are set to be $\epsilon_1 = \epsilon_2 = 0.01$. We also compare ER and FMMC on small-world graphs while the number of nodes n is varied with an edge density $\frac{2m}{n^2-n} \approx 0.4$ where m is the total number of edges. On small-world graphs we generated 1000 instances of y_i^0 drawn from $\mathbf{N}(0, 100)$ and stopped algorithms whenever tolerance levels $\epsilon_1 = \epsilon_2 = 0.05$ are obtained or the number of communications per node exceeded 10^6.

Graph	Method	Comm. per node (stage-i)	Comm. per node (stage-ii)
$K_5 - K_5$	ER	2.9 $\times 10^3$	81
	FMMC	1.28 $\times 10^5$	65
$K_{10} - K_{10}$	ER	8.4 $\times 10^4$	198
	FMMC	3.93 $\times 10^5$	130
$K_{20} - K_{20}$	ER	2.6 $\times 10^6$	433
	FMMC	6.4 $\times 10^6$	251
$K_{25} - K_{25}$	ER	7.9 $\times 10^6$	566
	FMMC	$> 10^7$	287

TABLE I: FMMC vs ER on the barbell graph.

Results for both of the graphs are reported in Tables I and II in which we compare the average communication per node in the pre-computation (stage-i) and in the consensus computation (stage-ii) where results are averaged over 1000 runs. On barbell graph, we observe that FMMC requires less communications at the second (consensus) stage as expected (as FMMC is based on the optimal matrix P^{opt}), but in terms of total communications (stage-i + stage-ii) ER outperforms FMMC. In the case of small-world graphs, computation of P^{opt} exceeded the maximum communication limit which caused FMMC to perform worse than ER in stage-ii (since the stage-i solution is not a precise approximation of P^{opt} anymore). We can say that ER performs better than FMMC in terms of total communications for both graph types.

Graph	Method	Comm. per node (stage-i)	Comm. per node (stage-ii)
$n = 5$	ER	6.4	41
	FMMC	41075.2	84
$n = 10$	ER	16.8	130
	FMMC	$> 10^6$	143
$n = 20$	ER	19.20	315
	FMMC	$> 10^6$	370
$n = 25$	ER	20.00	403
	FMMC	$> 10^6$	512

TABLE II: FMMC vs ER on the small-world graph

Fastest quantum gossiping (FQG), proposed by Jafarizadeh in [52], is an alternative key approach for choosing the optimal wake-up (p_i^f) and conditional communication probabilities ($p_{j|i}^f$) at each agent $i \in \mathcal{N}$ in a way to optimize the spectral gap of the expected iteration matrix. It requires solving an SDP similar to FMMC; however, it is a more general approach than FMMC as it can also support non-uniform wake-up probabilities. We compared our ER-based gossiping with FQG and Metropolis gossiping[8] on several graph topologies; due to the space limits, we provided the details in the extended arXiv version of our paper [15]. Our results suggest that the Metropolis weights require no pre-computation but they are the slowest in terms of the spectral gap. ER is faster than Metropolis but slower than FQG; but the advantage is that computing ER weights require less CPU time and also ER weights can be efficiently computed with a linearly convergent decentralized algorithm. Further experiments demonstrating the advantages of ER-based gossiping over gossiping with uniform and Metropolis weights can be found in the arXiv version of our paper [15].

B. Effective resistance-based DPGA-W and EXTRA

We implemented our ER-based communication framework into the state of the art distributed algorithms: DPGA-W [19] and EXTRA [45] to solve regularized logistic regression problems over a barbell graph $K_{\tilde{n}} - K_{\tilde{n}}$ with $n = 2\tilde{n}$ nodes: We minimize $\min_{x\in\mathbb{R}^p} \sum_{i=1}^n f_i(x)$ with

$$f_i(x) \triangleq \frac{1}{2n}\|x\|^2 + \frac{1}{N_s}\sum_{\ell=1}^{N_s} \log(1+\exp^{-b_{i\ell}a_{i\ell}^\top x}), \tag{17}$$

where N_s is the number of samples at each node, $\{(a_{i\ell}, b_{i\ell})\}_{\ell=1}^{N_s} \subset \mathbb{R}^p \times \{-1, 1\}$ for $i \in \mathcal{N}$ denote the set of feature vectors and corresponding labels. We let $p = 20$ and $N_s = 5$. For each $n \in \{20, 40\}$ and $\sigma \in \{1, 2\}$, we randomly generated 20 i.i.d. instances of the problem in (17) by sampling $a_{i\ell} \sim \mathbf{N}(\mathbf{1}, \sigma^2\mathbf{I})$ independently from the normal distribution and setting $b_{i\ell} = -1$ if $1/(1+e^{-a_{i\ell}^\top \mathbf{1}}) \leq 0.55$ and to $+1$ otherwise. Both algorithms are terminated after 10^4 iterations. For benchmark, we also solved each instance of (17) using MOSEK[9]. We initialized the iterates uniformly sampling each p components from the $[500, 510]$ interval for nodes in one $K_{\tilde{n}}$, and from $[-500, -490]$ for nodes in the the other $K_{\tilde{n}}$. The results for $n = 20$ and $n = 40$ are displayed in Fig. 3 and Fig. 4, respectively. We plotted relative suboptimality $\|\mathbf{x}^k - \mathbf{x}^*\| / \|\mathbf{x}^*\|$, function value sequence $\sum_{i\in\mathcal{N}} f_i(x_i^k)$ for the range $[0,\ 10^5]$, and consensus violation $\|\mathbf{x}^k - \bar{\mathbf{x}}^k\| / \sqrt{n}$, where k denotes the (synchronous) communication round counter – in each communication round neighboring nodes communicate among each other synchronously once – and $\mathbf{x}^k = [x_i^k]_{i\in\mathcal{N}}$ denotes the kth iterate; moreover, $\bar{\mathbf{x}}^k = \mathbf{1}\otimes\bar{x}^k$, $\bar{x}^k = \sum_{i\in\mathcal{N}} x_i^k/n$, $\mathbf{x}^* \triangleq \mathbf{1}\otimes x^*$ and x^* is the minimizer to (17), where $\otimes$ denotes the Kronecker product.

Both DPGA-W[10] and EXTRA uses a communication matrix W that encodes the network topology. DPGA-W uses node-specific step-sizes initialized at $\approx 1/L_i$ for $i \in \mathcal{N}$, where L_i denotes the Lipschitz constant of ∇f_i, we adopted the adaptive step-size strategy described in [19, Sec. III.D]; and for EXTRA, we choose the constant step-size, common for all nodes, as suggested in [45], i.e., we choose the step size as $2\lambda_{\min}(\tilde{W})/\max_{i\in\mathcal{N}} L_i$, where $\tilde{W} = (\mathbf{I}+W)/2$.

For both algorithms, we compared two choices of W: W^u based on uniform edge weights, and W^r based on effective

[8]In the Metropolis-based gossiping approach, each node i wakes up with uniform probabilities (i.e. $p_i^m = \frac{1}{n}$) and communicates with one of its neighbors $j \in \mathcal{N}_i \setminus \{i\}$ with probability $p_{j|i}^m = \frac{1}{\max\{d_i, d_j\}}$.

[9]http://docs.mosek.com/8.0/toolbox/index.html

[10]In DPGA-W stepsize parameter γ_i is set to $1/\|\omega_i\|$ for $i \in \mathcal{N}$ – see [19].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3161201, IEEE Transactions on Control of Network Systems

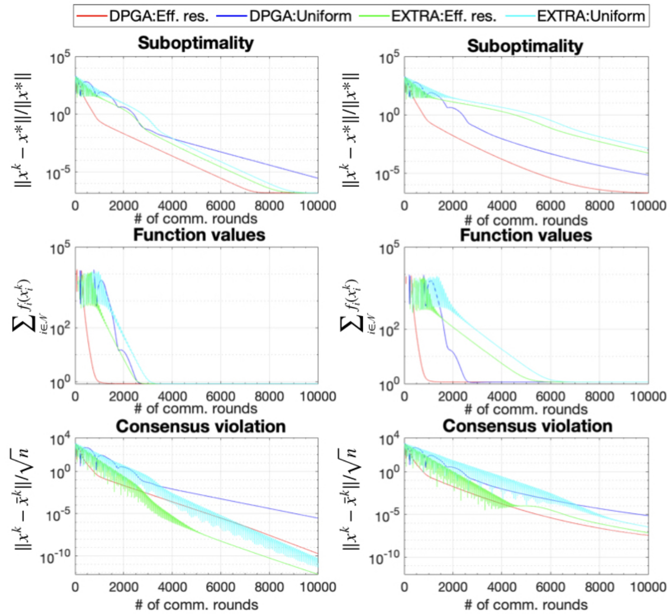

Fig. 3: The suboptimality, function value, and the difference from average versus the number of communication rounds, based on logistic regression using DPGA-W and EXTRA algorithms with resistance weights and uniform probability weights on barbell graph $K_{10} - K_{10}$. Data is sampled using $\sigma = 1$ (**Left**) and $\sigma = 2$ (**Right**).

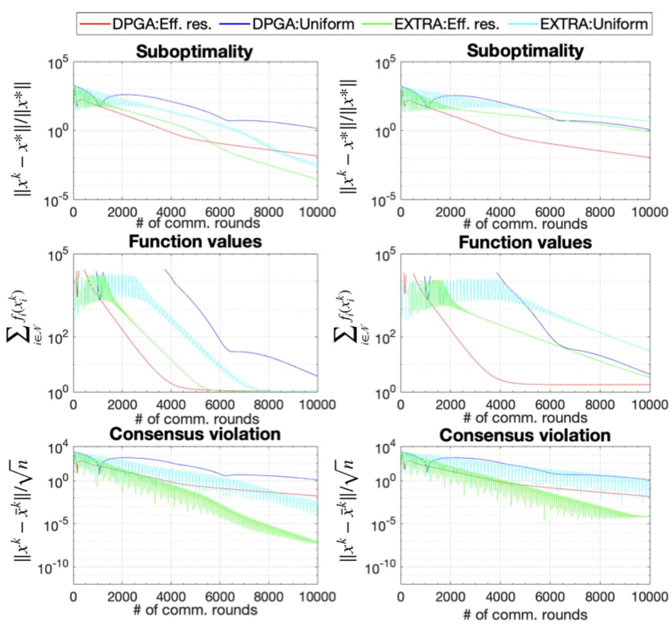

Fig. 4: The suboptimality, function value, and difference from average versus the average number of communication rounds, based on logistic regression using DPGA-W and EXTRA algorithms with resistance weights and uniform probability weights on $K_{20} - K_{20}$. Data is sampled using $\sigma = 1$ (**Left**) and $\sigma = 2$ (**Right**).

resistances. In DPGA-W, the graph Laplacian is adopted for uniform weights, i.e., $W^u = W^{u,\text{DPGA-W}} \triangleq \mathcal{L}$, while for the ER-based weights, we set $W^r = W^{r,\text{DPGA-W}}$ where $W_{ii}^{r,\text{DPGA-W}} \triangleq \sum_{j\in\mathcal{N}_i} R_{ij}$ for $i \in \mathcal{N}$ and $W_{ij}^{r,\text{DPGA-W}} = -R_{ij}$ for $(i,j) \in \mathcal{E}$ and 0 otherwise. For EXTRA, $W^{u,\text{EXTRA}} = \mathbf{I} - \mathcal{L}/\tau$ where $\tau = \lambda_{\max}(\mathcal{L})/2+\varepsilon$ where $\lambda_{\max}$ denotes the largest eigenvalue; on the other hand, $W^{r,\text{EXTRA}} = \mathbf{I} - W^{r,\text{DPGA-W}}/\tau$ where $\tau = \lambda_{\max}(W^{r,\text{DPGA-W}})/2 + \varepsilon$ for $\varepsilon = 0.01$.

Figures 3 and 4 illustrate the performance comparison of both DPGA-W and EXTRA algorithms with effective resistance and uniform weights in terms of suboptimality, convergence in function values and consensus violation for the barbell graph $K_{10} - K_{10}$ and $K_{20} - K_{20}$ respectively – the reported results are averages over the 20 problem instances. The subfigures on the left of Figures 3 and 4 are for noise level $\sigma = 1$ whereas those on the right are for $\sigma = 2$. In Figures 3 and 4, we observe that using ER weights improves upon the uniform weights for both EXTRA and DPGA-W methods consistently to solve the logistic regression problem in terms of suboptimality, function values and consensus violation significantly. We also observe that with noisier data, DPGA-W works typically faster than EXTRA in terms of function values and suboptimality. This is because when noise level σ gets larger, the local Lipschitz constant L_i of the nodes demonstrate higher variability, and DPGA-W adapts to this variability as it uses a step size that is different at each node in a way to adapt to L_i, whereas EXTRA uses a constant step size that is the same for all nodes. On the other hand, in terms of consensus violation, we see that EXTRA with ER weights typically outperforms DPGA-W with ER weights.

VI. Conclusions

We obtained a number of theoretical guarantees for gossiping with ER weights on c-barbell and barbell graphs, and also on arbitrary graphs with a small diameter. Our theoretical results are limited to these special graphs; however, we also showed that these methods are effective for solving the consensus problem in practice over barbell graphs and small-world graphs. We provided numerical experiments which demonstrate that using ER gossiping within EXTRA and DPGA-W methods improves their practical performance in terms of communication efficiency.

Appendix A
Proofs of Propositions 6 and 9

Proof of Proposition 6: The proof is based on finding the subset S of the vertex set of c-barbell graph that determines the conductance, i.e. that solves the minimization problem (10). First, for any given $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, the conductance of a subset $S \subset \mathcal{N}$ with respect to the probability transition matrix W is

$$\Phi_S(W) \triangleq \frac{1}{\pi(S)} \sum_{i\in S, j\in S^C} \pi_i W_{ij}. \tag{18}$$

Eq. (10) implies $\Phi(W) = \min_{S\subset\mathcal{N}:\pi(S)\in(0,1/2]} \Phi_S(W)$.[11] With slight abuse of notation, for a subgraph $\mathcal{H}_0$ with a vertex set S_0, we define $\Phi_{\mathcal{H}_0}(W) \triangleq \Phi_{S_0}(W)$. We say that a vertex set $S \subset \mathcal{N}$ on graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, w)$ is a *one-cut set* if its complement $\mathcal{N} \setminus S$ is a connected subgraph of $\mathcal{G}$. Similarly, we define *two-cut set* $S_2 \subset \mathcal{N}$ to be a set whose complement

[11] This follows after straightforward computations since the Markov chain with transition matrix W and stationary distribution π is reversible, i.e., $\pi(S)\Phi_S(W) = \pi(S^c)\Phi_{S^c}(W)$ for any S with $\pi(S) \in (0,1)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3161201, IEEE Transactions on Control of Network Systems

$\mathcal{N} \backslash S_2$ consists of two disjoint non-empty connected subgraphs $\mathcal{H}_1$ and $\mathcal{H}_2$ of $\mathcal{G}$. We define

$$\mathcal{G}_1 \triangleq \text{the left-most clique of the } c\text{-barbell graph.} \tag{19}$$

For $c_0 \in [2, c]$, we also define

$$\mathcal{G}_{c_0} \triangleq c_0\text{-barbell subgraph that includes the left-most } c_0 \text{ cliques of the } c\text{-barbell graph.} \tag{20}$$

Note that matrices W_{P^u} and W_{P^r} are symmetric and Markov chains with these transition matrices have the uniform distribution as a stationary distribution. Therefore, Lemmas 14 and 15 provided in Appendix B imply that a set S with minimal conductance should be a one-cut set and has to be given by the vertices of a subgraph $\mathcal{G}_{c_0}$ for some $c_0 \in [1, c]$ for both W_{P^u} and W_{P^r}. The conductance of one-cut subgraphs with respect to these transition matrices can be computed explicitly (see Proof of Lemma 15 for details):

$$\Phi_{\mathcal{G}_{c_0}}(\overline{W}_{P^u}) = \frac{1}{c_0}\frac{1}{c\tilde{n}^3}, \quad \Phi_{\mathcal{G}_{c_0}}(\overline{W}_{P^r}) = \frac{1}{2c_0\tilde{n}(c\tilde{n}-1)}. \tag{21}$$

Both of the expressions at (21) are minimized for the choice of $c_0 = \lfloor \frac{c}{2} \rfloor$. Therefore, the minimal conductance is attained for the subgraph $\mathcal{G}_{\lfloor \frac{c}{2} \rfloor}$. Plugging $c_0 = \lfloor \frac{c}{2} \rfloor$ into the expressions above yields the graph conductance values at (12). The bounds (7) and (8) follow from Theorem 1 and inequalities (13). ■

Proof of Proposition 9: It follows from Corollary 16 and Lemma 17 in Appendix B that the second largest eigenvalues of $\bar{W}_{P^u}$ and $\bar{W}_{P^r}$ are given by: $\lambda_{n-1}(\overline{W}_{P^u}) = 1 - \frac{8}{n^2(n-2)} + \Theta(\frac{1}{n^4})$ and $\lambda_{n-1}(\overline{W}_{P^r}) = 1 - \frac{1}{n(n-1)} - \Theta(\frac{1}{n^3})$. This implies directly $\lambda_{n-1}(\overline{W}_{P^r}) = 1 - \Theta(\frac{1}{n^2})$ and $\lambda_{n-1}(\overline{W}_{P^u}) = 1 - \Theta(\frac{1}{n^3})$, which completes the proof. ■

APPENDIX B
SUPPORTING RESULTS

Lemma 13. *[53, Eqn. (2.2)] Let W be the transition matrix of a Markov chain with stationary distribution π. Let j be a neighbor of i, i.e. $j \in \mathcal{N}_i$, then $H_W(i \to j) \leq (\pi_j W_{ji})^{-1}$.*

Lemma 14. *Consider a reversible Markov chain on a c-barbell graph with a uniform stationary distribution. Let $\mathcal{H}_0$ be a subgraph of $\mathcal{G}$ whose vertex set is a non-empty two-cut set $\mathcal{S}_0$ satisfying $|\mathcal{S}_0| \leq \frac{|\mathcal{N}|}{2}$. Then, there exists another subgraph $\tilde{\mathcal{H}}_0$ of $\mathcal{G}$ such that $\Phi_{\tilde{\mathcal{H}}_0}(W) < \Phi_{\mathcal{H}_0}(W)$.*

Proof: Let C_1 and C_2 be the vertex sets of two disjoint non-empty connected subgraphs within $\mathcal{N} \setminus \mathcal{S}_0$ satisfying $\mathcal{N} = C_1 \cup \mathcal{S}_0 \cup C_2$. Note $C_1 \cap C_2 = \emptyset$ implies either $|C_1 \cup \mathcal{S}_0| \leq \frac{|\mathcal{N}|}{2}$ or $|C_2| \leq \frac{|\mathcal{N}|}{2}$. Since the transition matrix W of a reversible Markov chain with a uniform stationary distribution is symmetric, the definition (18) implies $\Phi_{C_1 \cup \mathcal{S}_0}(W) = \Phi_{C_2}(W)$. Without loss of generality, choose $\tilde{\mathcal{H}}_0$ to be the subgraph with vertices $\tilde{\mathcal{S}}_0 = C_1 \cup \mathcal{S}_0$ s.t. $|C_1 \cup \mathcal{S}_0| \leq \frac{|\mathcal{N}|}{2}$ (otherwise, pick the subgraph with vertex set C_2 instead), then

$$\begin{aligned}\Phi_{\mathcal{H}_0}(W) &= \frac{1}{|\mathcal{S}_0|}\Big(\sum_{i \in \mathcal{S}_0, j \in C_1} W_{ij} + \sum_{i \in \mathcal{S}_0, j \in C_2} W_{ij}\Big) \\ &> \frac{1}{|\mathcal{S}_0|}\sum_{i \in \mathcal{S}_0, j \in C_2} W_{ij} > \frac{1}{|\tilde{\mathcal{S}}_0|}\sum_{i \in \tilde{\mathcal{S}}_0, j \in C_2} W_{ij} = \Phi_{\tilde{\mathcal{H}}_0}(W),\end{aligned}$$

which proves Lemma 14. ■

We will also need the following lemmas, whose proofs can be founded in the extended arXiv version [15] of our paper; the details of the proofs are omitted here due to space limitations.

Lemma 15. *Consider a Markov chain on a c-barbell graph with a probability transition matrix W. If $W = \overline{W}_{P^u}$ or $W = \overline{W}_{P^r}$, then for any subgraph $\mathcal{H}_0$ having a one-cut vertex set $\mathcal{S}_0$, there exists a subgraph $\mathcal{G}_{c_0}$ for some $c_0 \in [1, c]$ such that $\Phi_{\mathcal{G}_{c_0}}(W) \leq \Phi_{\mathcal{H}_0}(W)$ where $\mathcal{G}_{c_0}$ is defined by (19) and (20).*

Corollary 16. *Under the setting of Proposition 8, assume that the weight matrix w is normalized, i.e., $\sum_{j=1}^n w_{ij} = 1$ for all $i \in \mathcal{N}$. Then $W = w$ is doubly stochastic and the eigenvalues of W are $\lambda_a = 1$ with multiplicity one, $\lambda_b = -1 + (A+G) + F$ with multiplicity one, $\lambda_c = D - C$ with multiplicity $2\tilde{n} - 4$, and $\lambda_\pm = \frac{1}{2}\Big(F + G - A \pm \sqrt{S}\Big)$, where A, B, C, D, E, F, G and S are as in Proposition 8. Moreover, λ_+ satisfies*

$$\lambda_+ = \frac{1}{2}\Big(F + G - A + \sqrt{(F - G + A)^2 + 4BE}\Big), \tag{22}$$

and is the second largest eigenvalue, i.e. $\lambda_{n-1}(W) = \lambda_+$.

Lemma 17. *Consider the setting of Proposition 8:*

(i) If $W = \overline{W}_{P^u}$, then Proposition 8 applies with $A = A^u$, $B = B^u$, $C = C^u$, $D = D^u$ and $G = G^u$ where $A^u = \frac{2}{n^2}$, $B^u = \frac{n-1}{n^2(0.5n-1)}$, $C^u = \frac{2}{n(n-2)}$, $D^u = \frac{n^3-3n^2+2n+2}{n^2(n-2)}$, and $G^u = 1 - \frac{n+1}{n^2}$. The second largest eigenvalue of $\overline{W}_{P^u}$ is given by $\lambda_{n-1}(\bar{W}_{P^u}) = 1 - \frac{n^2+n-8}{2n^2(n-2)} + \frac{1}{8}\sqrt{S_n^u} = 1 - \frac{8}{n^2(n-2)} + \Theta(\frac{1}{n^4})$, for $S_n^u = \frac{4n^3+24n^2-156n+192}{(0.5n-1)^2 n^3}$.

(ii) If $W = \overline{W}_{P^r}$, then Proposition 8 applies with $A = A^r$, $B = B^r$, $C = C^r$, $D = D^r$ and $G = G^r$ where $A^r = \frac{1}{2(n-1)}$, $B^r = \frac{2}{n(n-1)}$, $C^r = \frac{2}{n(n-1)}$, $D^r = \frac{n^2-2n+2}{n(n-1)}$, and $G^r = 1 - \frac{1.5n-2}{n(n-1)}$. Moreover, the second largest eigenvalue of $\overline{W}_{P^u}$ is given by $\lambda_{n-1}(\bar{W}_{P^r}) = 1 - \frac{1}{(n-1)} + \frac{1}{2}\sqrt{S_n^r} = 1 - \frac{1}{n(n-1)} - \Theta(\frac{1}{n^3})$, for $S_n^r = \frac{4n-8}{n(n-1)^2}$.

REFERENCES

[1] N. S. Aybat and M. Gürbüzbalaban, "Decentralized computation of effective resistances and acceleration of consensus algorithms," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 538–542.

[2] D. J. Klein, "Resistance-distance sum rules," *Croatica chemica acta*, vol. 75, no. 2, pp. 633–649, 2002.

[3] D. J. Klein and M. Randić, "Resistance distance," *Journal of Mathematical Chemistry*, vol. 12, no. 1, pp. 81–95, 1993.

[4] A. Ghosh, S. Boyd, and A. Saberi, "Minimizing effective resistance of a graph," *SIAM review*, vol. 50, no. 1, pp. 37–66, 2008.

[5] D. Aldous and J. A. Fill, "Reversible Markov chains and random walks on graphs," 2014, unfinished monograph.

[6] P. G. Doyle and J. L. Snell, *Random walks and electric networks*. Mathematical Association of America, 1984.

[7] D. Spielman and N. Srivastava, "Graph sparsification by effective resistances," *SIAM J. on Computing*, vol. 40, no. 6, pp. 1913–1926, 2011.

[8] R. C. Mishra and H. Barman, "Effective resistances of two-dimensional resistor networks," *European Journal of Physics*, vol. 42, no. 1, p. 015205, Dec 2020.

[9] M. A. Jafarizadeh, R. Sufiani, and S. Jafarizadeh, "Calculating effective resistances on underlying networks of association schemes," *Journal of Mathematical Physics*, vol. 49, no. 7, p. 073303, Jul 2008.

[10] M. Jafarizadeh, R. Sufiani, and S. Jafarizadeh, "Calculating two-point resistances in distance-regular resistor networks," *Journal of Physics A*, vol. 40, pp. 4949–4972, 2006.

[11] R. B. Bapat, I. Gutmana, and W. Xiao, "A simple method for computing resistance distance," *Zeitschrift für Naturforschung A*, vol. 58, no. 9-10, pp. 494–498, 2003.

[12] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. on Auto. Control*, vol. 54, no. 1, p. 48, 2009.

[13] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Tran. on Networking (TON)*, vol. 14, no. SI, pp. 2508–2530, 2006.

[14] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[15] B. Can, S. Soori, N. S. Aybat, M. M. Dehnavi, and M. Gurbuzbalaban, "Randomized gossiping with effective resistance weights: Performance guarantees and applications," *arXiv preprint arXiv:1907.13110*, 2021. [Online]. Available: https://arxiv.org/pdf/1907.13110.pdf

[16] E. Abbe, "Community detection and stochastic block models: recent developments," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6446–6531, 2017.

[17] Y. Zhang and X. Lin, "DISCO: Distributed optimization for self-concordant empirical loss," in *International Conference on Machine Learning*, 2015, pp. 362–370.

[18] C.-P. Lee, C. H. Lim, and S. J. Wright, "A distributed quasi-Newton algorithm for empirical risk minimization with nonsmooth regularization," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1646–1655.

[19] N. S. Aybat, Z. Wang, T. Lin, and S. Ma, "Distributed linearized alternating direction method of multipliers for composite convex consensus optimization," *IEEE Tran. on Auto. Control*, vol. 63, no. 1, pp. 5–20, 2018.

[20] A. Olshevsky, "Linear time average consensus on fixed graphs and implications for decentralized optimization and multi-agent control," *arXiv preprint arXiv:1411.4186*, 2016.

[21] N. Loizou and P. Richtárik, "Accelerated gossip via stochastic heavy ball method," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2018, pp. 927–934.

[22] ——, "Revisiting randomized gossip algorithms: General framework, convergence rates and novel block and accelerated protocols," *arXiv preprint arXiv:1905.08645*, 2019.

[23] N. Loizou, M. Rabbat, and P. Richtárik, "Provably accelerated randomized gossip algorithms," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 7505–7509.

[24] P. Rebeschini and S. C. Tatikonda, "Accelerated consensus via min-sum splitting," in *Advances in Neural Information Processing Systems*, 2017, pp. 1374–1384.

[25] R. L. Cavalcante, A. Rogers, and N. Jennings, "Consensus acceleration in multiagent systems with the chebyshev semi-iterative method," 2011.

[26] J. H. Seidman, M. Fazlyab, G. J. Pappas, and V. M. Preciado, "A Chebyshev-accelerated primal-dual method for distributed optimization," in *IEEE Conf. on Decision and Control*. IEEE, 2018, pp. 1775–1781.

[27] J. Bu, M. Fazel, and M. Mesbahi, "Accelerated consensus with linear rate of convergence," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 4931–4936.

[28] S. Roch *et al.*, "Bounding fastest mixing," *Electronic Communications in Probability*, vol. 10, pp. 282–296, 2005.

[29] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM REVIEW*, vol. 46, pp. 667–689, 2003.

[30] D. Shah, "Gossip algorithms," *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1–125, 2009.

[31] Y. Afek, *Distributed Computing: 27th International Symposium, Jerusalem, Israel, Proceedings*. Springer, 2013, vol. 8205.

[32] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[33] F. Fagnani and P. Frasca, *Introduction to averaging dynamics over networks*. Springer, 2017, vol. 472.

[34] K. Jung, D. Shah, and J. Shin, "Distributed averaging via lifted Markov chains," *IEEE Trans. on Info. Theory*, vol. 56, no. 1, pp. 634–647, 2010.

[35] P. Tetali, "An extension of Foster's network theorem," *Combinatorics, Probability and Computing*, vol. 3, no. 3, pp. 421–427, 1994.

[36] K. Censor-Hillel and H. Shachnai, "Fast information spreading in graphs with large weak conductance," *SIAM Journal on Computing*, vol. 41, no. 6, pp. 1451–1465, 2012.

[37] R. Cont and E. Tanimura, "Small-world graphs: Characterization and alternative constructions," *Advances in Applied Probability*, vol. 40, no. 4, pp. 939–965, 2008.

[38] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2012, vol. 3.

[39] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, "Symmetry analysis of reversible Markov chains," *Internet Mathematics*, vol. 2, no. 1, pp. 31–71, 2005.

[40] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues of Markov chains," *The Annals of Applied Probability*, vol. 1, no. 1, pp. 36–61, 1991. [Online]. Available: http://www.jstor.org/stable/2959624

[41] Y. Peres and P. Sousi, "Mixing times are hitting times of large sets," *Journal of Theoretical Probability*, vol. 28, no. 2, pp. 488–519, 2015.

[42] A. Olshevsky, "Linear time average consensus on fixed graphs?" *IFAC-PapersOnLine*, vol. 48, no. 22, pp. 94–99, 2015.

[43] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, "The electrical resistance of a graph captures its commute and cover times," *Computational Complexity*, vol. 6, no. 4, pp. 312–340, 1996.

[44] D. A. Levin, Y. Peres, and E. L. Wilmer, "Markov chains and mixing times. with a chapter by James G. Propp and David B. Wilson," *American Mathematical Society, Providence, RI*, 2009.

[45] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

[46] Tsitsiklis and J. Nikolas, "Problems in decentralized decision making and computation," Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems, Tech. Rep., 1984.

[47] E. Zanaj, M. Baldi, and F. Chiaraluce, "Efficiency of the gossip algorithm for wireless sensor networks," in *15th Int. Conf. on Software, Telecommunications and Computer Networks*, Sep. 2007, pp. 1–5.

[48] A.-M. Kermarrec and M. van Steen, "Gossiping in distributed systems," *SIGOPS Oper. Syst. Rev.*, vol. 41, no. 5, pp. 2–7, Oct. 2007.

[49] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* IEEE, 2003, pp. 482–491.

[50] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, May 2008.

[51] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ser. MobiCom '99. New York, NY, USA: ACM, 1999, pp. 263–270.

[52] S. Jafarizadeh, "Gossip algorithm with nonuniform clock distribution: Optimization over classical and quantum networks," *Optimal Control Applications and Methods*, vol. 41, no. 2, pp. 616–639, 2020.

[53] S. Ikeda, I. Kubo, and M. Yamashita, "The hitting and cover times of random walks on finite graphs using local degree information," *Theoretical Computer Science*, vol. 410, no. 1, pp. 94–100, 2009.

Necdet Serhat Aybat is an Associate Professor at the Department of Industrial Engineering, Penn State University. He received the B.S. and M.S. degrees in industrial engineering from Bogazici University, Istanbul, Turkey, in 2003 and 2005, respectively, and his Ph.D. degree in operations research from Columbia University, New York, USA.

Bugra Can received the B.S. degrees in mathematics and economics from Koc University, Turkey in 2017. Currently, he is a Ph.D. student at the Department of Management Sciences and Information Systems (MSIS) of Rutgers University

Maryam Mehri Dehnavi is an Assistant Professor in the Computer Science department at the University of Toronto. She received her Ph.D. in Electrical and Computer Engineering from McGill University, Canada in 2013.

Mert Gürbüzbalaban is an Assistant Professor at Rutgers University working in optimization and network science. He received his B.Sc. degree in electrical engineering and mathematics from Bogazici University, Turkey in 2005, and Ph.D. in mathematics from New York University, USA in 2012.

Saeed Soori is a Ph.D. student at the University of Toronto in the Department of Computer Science. He received his B.S. degree in Electrical and Computer Engineering from University of Tehran in 2016.