

# Singular Value Approximation and Sparsifying Random Walks on Directed Graphs

AmirMahdi Ahmadinejad <i>Amazon</i> Seattle, USA ahmadi@alumni.stanford.edu	John Peebles <i>Apple</i> Cupertino, USA peebles@apple.com	Edward Pyne <i>MIT</i> Cambridge, USA epyne@mit.edu	Aaron Sidford <i>Stanford University</i> Palo Alto, USA sidford@stanford.edu	Salil Vadhan <i>Harvard University</i> Cambridge, USA salil_vadhan@harvard.edu
--	---	--	---	---

**Abstract**—In this paper, we introduce a new, spectral notion of approximation between directed graphs, which we call *singular value (SV) approximation*. SV-approximation is stronger than previous notions of spectral approximation considered in the literature, including spectral approximation of Laplacians for undirected graphs [ST04], standard approximation for directed graphs [CKP<sup>+</sup>17], and unit-circle (UC) approximation for directed graphs [AKM<sup>+</sup>20]. Further, SV approximation enjoys several useful properties not possessed by previous notions of approximation, e.g., it is preserved under products of random-walk matrices and bounded matrices.

We provide a nearly linear-time algorithm for SV-sparsifying (and hence UC-sparsifying) Eulerian directed graphs, as well as  $\ell$ -step random walks on such graphs, for any  $\ell \leq \text{poly}(n)$ . Combined with the Eulerian scaling algorithms of [CKK<sup>+</sup>18], given an arbitrary (not necessarily Eulerian) directed graph and a set  $S$  of vertices, we can approximate the stationary probability mass of the  $(S, S^c)$  cut in an  $\ell$ -step random walk to within a multiplicative error of  $1/\text{polylog}(n)$  and an additive error of  $1/\text{poly}(n)$  in nearly linear time. As a starting point for these results, we provide a simple black-box reduction from SV-sparsifying Eulerian directed graphs to SV-sparsifying undirected graphs; such a directed-to-undirected reduction was not known for previous notions of spectral approximation.

*Index Terms*—graph algorithms

## I. INTRODUCTION

Random walks on graphs play a central role in theoretical computer science. In algorithm design, they have found a wide range of applications including, maximum flow [CKM<sup>+</sup>11], [LRS13], [KLOS14], [vdBGJ<sup>+</sup>22], [vdBLL<sup>+</sup>21], sampling random spanning trees [KM09], [MST15], and clustering and partitioning [AM85], [KVV04], [ACL06], [OSV12]. Correspondingly, new algorithmic results on efficiently accessing properties of random walks have the potential for broad implications. In particular, in complexity theory, such algorithms have attracted attention as a promising approach to derandomizing space-bounded computation [SZ99], [Rei08], [RTV06], [AKM<sup>+</sup>20].

In this paper we consider the well-studied problem of *estimating the  $\ell$ -step random walk on a directed graph*. Given

A.A. the work was started prior to joining Amazon and does not relate to Amazon. E.P. is supported by an Akamai Presidential Fellowship. A.S. is supported by a Microsoft Research Faculty Fellowship, NSF CAREER Award CCF-1844855, NSF Grant CCF-1955039, a PayPal research award, and a Sloan Research Fellowship. S.V. is supported by a Simons Investigator Award.

a strongly connected, weighted, directed graph  $G = (V, E, w)$ , its associated random walk matrix  $\mathbf{W} \in \mathbb{R}^{V \times V}$ , and an integer  $\ell > 0$ , we seek to approximate key properties of the  $\ell$ -step random walk,  $\mathbf{W}^\ell$ , more efficiently than we could computing  $\mathbf{W}^\ell$  explicitly. For example, we may wish to estimate individual entries of  $\mathbf{W}^\ell$ , the conductance or cut probabilities of subsets of vertices, or (expected) hitting times between pairs of vertices.

In recent years, *graph sparsification* has emerged as a powerful approach for efficiently solving such problems. When the graph is undirected, we look for *spectral sparsifiers* of the Laplacian  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , where  $\mathbf{D}$  is the diagonal matrix of degrees and  $\mathbf{A}$  is the adjacency matrix. It is known that for all  $\epsilon \in (0, 1)$ , that there exist  $\epsilon$ -spectral sparsifiers with sparsity  $\tilde{O}(|V|\epsilon^{-2})$ ; that is, a Laplacian matrix  $\tilde{\mathbf{L}}$  with  $\tilde{O}(|V|\epsilon^{-2})$  non-zero entries such that

$$(1 - \epsilon)x^\top \mathbf{L}x \leq x^\top \tilde{\mathbf{L}}x \leq (1 + \epsilon)x^\top \mathbf{L}x \text{ for all } x \in \mathbb{R}^V. \quad (1)$$

Spectral sparsifiers can be computed in nearly linear time [ST04], [SS08], [BSS12], [PS14]. Normalizing such a sparsifier  $\tilde{\mathbf{L}}$  by  $\mathbf{D}^{-1/2}$  on both sides, we obtain a spectral approximation of the *normalized Laplacian*  $\mathbf{D}^{-1/2}\tilde{\mathbf{L}}\mathbf{D}^{-1/2}$ , which directly gives information about random walks because it is equivalent (up to a change of basis) to the *random-walk Laplacian*,  $\mathbf{L}\mathbf{D}^{-1} = \mathbf{I} - \mathbf{W}$ . Indeed, from any spectral sparsifier  $\tilde{\mathbf{L}}$  satisfying Equation (1), we can approximate any desired cut  $(S, S^c)$  in the original graph in nearly linear time by evaluating  $x^\top \tilde{\mathbf{L}}x$  for  $x$  equal to the indicator vector of  $S$ . Furthermore, there are nearly linear-time algorithms for computing sparse  $\epsilon$ -spectral sparsifiers corresponding to the  $\ell$ -step random walk, i.e., sparsifiers of the weighted graph whose random-walk Laplacian is  $\mathbf{I} - \mathbf{W}^\ell$ , for any polynomial length  $\ell$  [CCL<sup>+</sup>15], [MRSV21].

Obtaining analogous results for sparsifying  $\mathbf{I} - \mathbf{W}^\ell$  for directed graphs has been more challenging. For a directed graph, we consider the *directed Laplacian* [CKP<sup>+</sup>16]  $\mathbf{L} = \mathbf{D}_{\text{out}} - \mathbf{A}^\top$  where  $\mathbf{D}_{\text{out}}$  is the associated diagonal matrix of out-degrees and  $\mathbf{A}^\top$  is the transpose of the associated weighted adjacency matrix,  $\mathbf{A}$ . In comparison to their symmetric counterparts for undirected graphs, nearly linear-time sparsification algorithms (which approximate more than the associated undirected graph) were developed more recently [CKP<sup>+</sup>17], [CGP<sup>+</sup>18] and have yet to be extended to handle long random walks.

Here we describe challenges in sparsifying  $\mathbf{I} - \mathbf{W}^\ell$  for directed graphs.

*a) Unknown Stationary Distribution.:* While the kernel of an undirected Laplacian matrix is the all ones vector, computing the kernel of a directed Laplacian matrix  $\mathbf{L}$  corresponds to computing the stationary distribution  $\pi$  of the random walk on the directed graph ( $\mathbf{L}\mathbf{D}_{\text{out}}^{-1}\pi = 0$ ). Without explicitly knowing the kernel, it is not known how to efficiently perform any kind of useful sparsification or approximately solve linear systems in  $\mathbf{L}$ . This difficulty was overcome in [CKP<sup>+</sup>16], [AJSS19] which provide reductions from solving general directed Laplacian systems to the case where the graph is *Eulerian*, meaning that every vertex has the same in-degree as out-degree. In Eulerian graphs, the stationary distribution is simply proportional to the vertex degrees and the all ones vector is both the left and right kernels of the associated directed Laplacian.

*b) Defining Approximation.:* Undirected Laplacians  $\mathbf{L}$  are symmetric and positive semidefinite (PSD), i.e.,  $x^\top \mathbf{L}x \geq 0$  for all  $x$ . This leads to the natural Spielman–Teng [ST04] definition of multiplicative approximation given in Equation (1). That is, we say that  $\tilde{\mathbf{L}}$  is an  $\varepsilon$ -approximation of  $\mathbf{L}$  if  $(1 - \varepsilon)\mathbf{L} \preceq \tilde{\mathbf{L}} \preceq (1 + \varepsilon)\mathbf{L}$ , where  $\preceq$  is the Löwner order on PSD matrices. However, even though Laplacians of directed graphs are potentially asymmetric, the quadratic form  $x^\top \mathbf{L}x$  depends only on a symmetrization of the Laplacian ( $x^\top \mathbf{L}x = x^\top ((\mathbf{L} + \mathbf{L}^\top)/2)x$ ). Consequently, the quadratic form discards key information about the associated directed graph (e.g. the quadratic form of a directed cycle and an undirected cycle are the same). Thus, defining approximation for directed graphs (even Eulerian ones) is more challenging than for undirected graphs and a more complex notion of approximation was introduced in [CKP<sup>+</sup>16]. This additional complexity requires designing new sparsification algorithms that take into account the directedness of the graph.

*c) Preservation under Powering.:* Even for undirected graphs, the standard definition of spectral approximation in Equation (1) is not preserved under powering. That is,  $\mathbf{I} - \tilde{\mathbf{W}} \approx \mathbf{I} - \mathbf{W}$  does not imply that  $\mathbf{I} - \tilde{\mathbf{W}}^2 \approx \mathbf{I} - \mathbf{W}^2$ . Indeed, in graphs that are bipartite (and connected),  $\mathbf{I} - \mathbf{W}^2$  has a two-dimensional kernel, corresponding to the  $\pm 1$  eigenvalues of  $\mathbf{W}$ , whereas  $\mathbf{I} - \mathbf{W}$  has only a one-dimensional kernel. Standard spectral approximation requires perfect preservation of the kernel of  $\mathbf{I} - \mathbf{W}$ , but not of  $\mathbf{I} - \mathbf{W}^2$ . Graphs that are nearly bipartite (i.e., where  $\mathbf{W}$  has an eigenvalue near  $-1$ ) can also experience a large loss in quality of approximation when squaring.

Cheng et al. [CCL<sup>+</sup>15] addressed this issue by (implicitly) strengthening spectral approximation to require that  $\mathbf{I} + \tilde{\mathbf{W}} \approx \mathbf{I} + \mathbf{W}$  in addition to  $\mathbf{I} - \tilde{\mathbf{W}} \approx \mathbf{I} - \mathbf{W}$ . This notion of approximation enabled algorithms for sparsifying  $\mathbf{I} - \mathbf{W}^\ell$  for undirected graphs in randomized near-linear time [CCL<sup>+</sup>15], [MRSV21] and deterministic logspace [MRSV21], [DMVZ20]. For directed graphs, the problem comes not just from bipartiteness, but general periodic structures (e.g. a directed cycle), which give  $\mathbf{W}$  complex eigenvalues on or

near the unit circle. This led Ahmadi et al. [AKM<sup>+</sup>20] to propose the notion of *unit-circle (UC) approximation*, which amounts to requiring that  $\mathbf{I} - z\tilde{\mathbf{W}}$  approximate  $\mathbf{I} - z\mathbf{W}$  for all complex numbers  $z$  of magnitude 1, with respect to the standard notion of approximation for directed graphs proposed in [CKP<sup>+</sup>16]. UC approximation has the property that it is preserved under taking arbitrary powers, with no loss in the approximation error. As such, sparsification techniques for UC and stronger notions must exactly preserve periodicity.

*d) Preservation of Periodic Structures.:* Sparsifying directed graphs under UC approximation is more challenging due to the need to preserve periodic structures in the graph, which can be easily lost or introduced by common sparsification techniques such as random sampling [SS08] or patching to fix degrees. Thus in [AKM<sup>+</sup>20], it was only shown how to partially sparsify the *square* of a graph; that is obtain a graph with random-walk matrix  $\tilde{\mathbf{W}}^2$  such that  $\mathbf{I} - \tilde{\mathbf{W}}^2$  UC-approximates  $\mathbf{I} - \mathbf{W}^2$ , but has fewer edges than the true square  $\mathbf{W}^2$ . Still, the number of edges in  $\tilde{\mathbf{W}}^2$  is larger than in  $\mathbf{W}$  by at least a constant factor, so if we iterate to obtain a sparse approximation of  $\mathbf{W}^\ell$ , the number of edges will grow by a factor of  $c^{\log \ell} = \text{poly}(\ell)$  and our approximations will quickly become dense. This was affordable in the deterministic logspace algorithms of [AKM<sup>+</sup>20], but is not in our setting of nearly linear time.

*Our Work.:* In this paper we provide several tools for overcoming these challenges, advancing both algorithmic and structural tools regarding graphs sparsification. First, we introduce a new notion of directed graph approximation called *singular value (SV) approximation*. We then show that this notion of approximation strictly strengthens unit-circle approximation and show that it has a number of desirable properties, such as preservation under not only powers but arbitrary products of random-walk matrices, and implying approximation of stationary probabilities of all cuts. Then we provide an efficient near linear-time randomized algorithm for computing nearly linear-sized SV-sparsifiers for arbitrary Eulerian directed graphs; this implies the first proof that nearly linear-sized UC-sparsifiers exist for Eulerian directed graphs. As a starting point for this result, we provide a simple reduction from SV-sparsifying Eulerian directed graphs to SV-sparsifying *undirected graphs*; no such reduction was known for the previous, weaker forms of spectral approximation of directed graphs, and shows that SV approximation is a significant strengthening even for undirected graphs.

Combined with the Eulerian scaling algorithms of [CKK<sup>+</sup>18], we obtain an algorithm for approximating the stationary probabilities of cuts (as well as “uncuts”) in random walks on arbitrary directed graphs, which we define as follows:

**Definition 1.1** (Cut values). For a strongly connected, weighted digraph  $G$  on  $n$  vertices, let  $\mu$  be the unique stationary distribution of the random walk on  $G$ , and let  $\mu_{\text{edge}}$  be the stationary distribution on edges  $(i, j)$  of  $G$  (i.e., pick  $i$  according to  $\mu$  and  $j$  by following an edge from  $i$  proportional to its weight). For subsets  $S$  and  $T$  of vertices, define:

- $\text{Cut}_G(S, T) = \Pr_{(i,j) \sim \mu_{\text{edge}}} [i \in S, j \in T]$ ,
- $\text{Cut}_G(S) = \text{Cut}_G(S, S^c) = (\text{Cut}_G(S, S^c) + \text{Cut}_G(S^c, S))/2$ , and
- $\text{Uncut}_G(S) = (\text{Cut}_G(S, S) + \text{Cut}_G(S^c, S^c))/2$ .

If  $G$  has random-walk matrix  $\mathbf{W}$ , we may write  $\text{Cut}_{\mathbf{W}}$  and  $\text{Uncut}_{\mathbf{W}}$  instead of  $\text{Cut}_G$  and  $\text{Uncut}_G$ .

**Definition I.2** (Powering). For a weighted digraph  $G$  with adjacency matrix  $\mathbf{A}$ , out-degree matrix  $\mathbf{D}_{\text{out}}$ , and random-walk matrix  $\mathbf{W} = \mathbf{A}\mathbf{D}_{\text{out}}^{-1}$ , we write  $G^\ell$  for the weighted digraph with adjacency matrix  $(\mathbf{A}\mathbf{D}_{\text{out}}^{-1})^\ell \cdot \mathbf{D}_{\text{out}}$  (and thus out-degree matrix  $\mathbf{D}_{\text{out}}$  and random-walk matrix  $\mathbf{W}^\ell$ ).

With these definitions, the main application of our SV sparsification results is the following:

**Theorem I.3** (informal). *There is a randomized algorithm that, given a strongly connected  $n$ -node  $m$ -edge directed graph  $G$  with integer edge weights in  $[1, U]$ , a walk length  $\ell$ , an error parameter  $\varepsilon > 0$ , and lower bound  $s$  on the minimum stationary probability of the random walk on  $G$ , runs in time  $O((m + n\varepsilon^{-2}) \cdot \text{poly}(\log(U\ell/s)))$  and outputs an  $O(n\varepsilon^{-2} \cdot \text{poly}(\log(U\ell/s)))$ -edge graph  $H$  such that for every two sets  $S, T$  of vertices, we have:*

$$\begin{aligned} & |\text{Cut}_H(S, T) - \text{Cut}_{G^\ell}(S, T)| \\ & \leq \frac{\varepsilon}{2} \cdot \sqrt{\min\{\text{Cut}_{G^\ell}(S), \text{Uncut}_{G^\ell}(S)\}} \\ & \quad \cdot \sqrt{\min\{\text{Cut}_{G^\ell}(T), \text{Uncut}_{G^\ell}(T)\}}. \end{aligned}$$

In particular:

$$(1 - \varepsilon) \cdot \text{Cut}_{G^\ell}(S) \leq \text{Cut}_H(S) \leq (1 + \varepsilon) \cdot \text{Cut}_{G^\ell}(S),$$

and

$$(1 - \varepsilon) \cdot \text{Uncut}_{G^\ell}(S) \leq \text{Uncut}_H(S) \leq (1 + \varepsilon) \cdot \text{Uncut}_{G^\ell}(S).$$

Note that when  $U, \ell \leq \text{poly}(n)$ ,  $s \geq 1/\text{poly}(n)$ , and  $\varepsilon \geq 1/\text{poly}(\log n)$ , our algorithm runs in time  $\tilde{O}(m)$ . For comparison, note that, given a set  $S$ , we can estimate the cut value for  $S$  using random walks in time roughly  $\tilde{O}(\ell/(\varepsilon^2 \text{Cut}_G(S)))$ , which is slower when  $\ell/\text{Cut}_G(S)$  is  $m^{1+\Omega(1)}$ . (Note that  $\text{Cut}_G(S)$  can be as small as  $1/m$ .)

It is also worth comparing to the following approaches that yield *high-precision* estimates (i.e. replacing multiplicative error  $\varepsilon$  with polynomially small additive error):

- Use matrix powering via repeated squaring to compute  $\mathbf{W}^\ell$ . This takes time  $n^\omega \cdot \log(\ell)$ , where  $\omega$  is the matrix multiplication exponent. This is slower than our algorithm assuming  $\omega > 2$  or  $m \leq n^{2-\Omega(1)}$ .
- Use the algorithm of [CKK<sup>+</sup>18] to obtain a high-precision estimate of the stationary distribution  $\mu$  of  $G$  in time  $\tilde{O}(m)$ , and then use repeated matrix-vector multiplication to compute  $\mathbf{W}^\ell \mu$ . This takes time  $\tilde{O}(m\ell)$ , so is slower than our algorithm except when  $\ell = \text{polylog}(n)$ .
- Use the algorithm of Ahmadi et al. [AKM<sup>+</sup>20]. This also gives high-precision estimates of  $\mathbf{W}^\ell$ , and does so in nearly logarithmic space, but the running time is superpolynomial. A running time of  $\Omega(m \cdot \ell)$  seems

inherent in the approach as it works by reducing to solving a directed Laplacian system of size  $m \cdot \ell$ .

It remains an interesting open problem to estimate any desired entry of  $\mathbf{W}^\ell$  to high precision in nearly linear time.

*e) Other Work on SV Approximation.*: The definition of SV approximation and some of our results on it (obtained in collaboration between Jack Murtagh and the authors) were first presented in the first author’s PhD thesis [Ahm20] in August 2020. Independently, Kelley [Kel21] used a variant of SV approximation to present an alternative proof of a result of [HPV21], who used unit-circle approximation to prove that the Impagliazzo–Nisan–Wigderson pseudorandom generator [INW94] fools permutation branching programs. Golowich and Vadhan [GV22] used SV approximation and some of our results (presented in the aforementioned thesis) to prove new pseudorandomness properties of expander walks against permutation branching programs. Most recently, Chen, Lyu, Tal, and Wu [CLTW22] have used a form of SV approximation to present alternative proofs of the results of [AKM<sup>+</sup>20], [PV21].

#### A. Singular-Value Approximation

In this paper, we present a stronger and more robust notion for addressing the challenge of defining approximation between directed graphs. Specifically, we introduce a novel definition of approximation for asymmetric matrices, which we call *singular-value approximation* (or *SV approximation*).

For simplicity in the rest of this introduction, we focus on the case of regular directed graphs, i.e. directed graphs where for some value  $d \geq 0$ , every vertex has in-degree  $d$  and out-degree  $d$ . (In the case of digraphs with non-negative edge weights, we obtain the in- and out-degrees by summing the incoming or out-going edge weights at each vertex.) However, all of our results generalize to Eulerian digraphs and some generalize to wider classes of complex matrices.

To introduce SV approximation, let  $\mathbf{A}$  be the adjacency matrix of a  $d$ -regular digraph, i.e.,  $\mathbf{A}$  is a non-negative real matrix where every row and column sum equals  $d$ . Then the (in- and out-) degree matrix is simply  $d\mathbf{I}$ . Dividing by  $d$ , it is equivalent to study approximation of the *random-walk matrix*  $\mathbf{W} = \mathbf{A}/d$ , which is doubly stochastic, and has degree matrix  $\mathbf{I}$ .

**Definition I.4** (SV approximation for doubly stochastic matrices). For doubly stochastic matrices  $\mathbf{W}, \tilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$  we say that  $\tilde{\mathbf{W}}$  is an  $\varepsilon$ -singular-value (SV) approximation of  $\mathbf{W}$ , written  $\tilde{\mathbf{W}} \stackrel{sv}{\approx}_\varepsilon \mathbf{W}$ , if for all “test vectors”  $x, y \in \mathbb{R}^n$ , we have

$$\begin{aligned} & \left| x^\top (\tilde{\mathbf{W}} - \mathbf{W}) y \right| \\ & \leq \frac{\varepsilon}{4} \cdot [x^\top (\mathbf{I} - \mathbf{W}\mathbf{W}^\top) x + y^\top (\mathbf{I} - \mathbf{W}^\top \mathbf{W}) y]. \end{aligned} \quad (2)$$

This formulation of SV-approximation is one of several equivalent formulations we provide in the full paper. We can equivalently define SV approximation between doubly stochastic matrices by requiring Equation (3) to hold for all

complex test vectors  $x, y \in \mathbb{C}^n$ . SV-approximation can also be defined equivalently by replacing condition (3) with

$$\begin{aligned} & \left| x^\top (\widetilde{\mathbf{W}} - \mathbf{W}) y \right| & (4) \\ & \leq \frac{\varepsilon}{4} \cdot \sqrt{[x^\top (\mathbf{I} - \mathbf{W}\mathbf{W}^\top) x] \cdot [y^\top (\mathbf{I} - \mathbf{W}^\top \mathbf{W}) y]}. & (5) \end{aligned}$$

These two formulations, (3) and (5), differ only in using the geometric mean or the arithmetic mean of the terms involving  $x$  and  $y$  on the right-hand side. The formulation in terms of the geometric mean implies the one in terms of the arithmetic mean (since the geometric mean is no larger than the arithmetic mean); the converse follows by optimizing over scalar multiples of  $x$  and  $y$  (as was done in e.g. [CKP<sup>+</sup>17], [AKM<sup>+</sup>20]). Both formulations can be rewritten more simply by noting that  $x^\top (\mathbf{I} - \mathbf{W}\mathbf{W}^\top) x = \|x\|^2 - \|x^\top \mathbf{W}\|^2$  and  $y^\top (\mathbf{I} - \mathbf{W}^\top \mathbf{W}) y = \|y\|^2 - \|\mathbf{W}y\|^2$ , but the description in terms of quadratic forms will be more convenient for comparison with previous notions of approximation. In the full paper, we provide more general definitions of SV approximation which also apply to unnormalized directed Laplacians and even to complex matrices.

We prove that SV approximation is strictly stronger than previous notions of spectral approximation considered in the literature, even for undirected graphs, and enjoys several useful properties not possessed by the previous notions. Most notably, there is a simple black-box reduction from SV-sparsifying Eulerian directed graphs to SV-sparsifying undirected graphs; no such reduction is known for prior notions of asymmetric spectral approximation.

Furthermore, we give efficient algorithms for working with SV approximation. These include nearly linear-time algorithms for SV-sparsifying undirected and hence also Eulerian directed graphs (Theorem I.11), as well as random-walk polynomials of directed graphs (Theorem I.11). We also show that a simple repeated-squaring and sparsification algorithm for solving Laplacian systems also works for Eulerian digraphs whose random-walk matrix is normal (i.e., unitarily diagonalizable), if we use SV-sparsification at each step (Theorem I.12). Prior Laplacian solvers for Eulerian graphs are more complex. We elaborate on these results in the next several subsections.

### B. Comparison to Previous Notions of Approximation

Let us compare Theorem I.4 to previous definitions of approximation.

a) *Undirected spectral approximation.*: Let's start with the undirected case, where  $\mathbf{W} = \mathbf{W}^\top$ . In this case, it can be shown that we can without loss of generality restrict Theorem I.4 to  $x = y$ , obtaining the following:  $\widetilde{\mathbf{W}} \stackrel{sv}{\approx}_\varepsilon \mathbf{W}$  requires that for all  $x \in \mathbb{R}^n$ ,

$$\left| x^\top (\widetilde{\mathbf{W}} - \mathbf{W}) x \right| \leq \frac{\varepsilon}{2} \cdot [x^\top (\mathbf{I} - \mathbf{W}^2) x]. \quad (6)$$

In contrast, the standard definition of spectral approximation (introduced by Spielman and Teng [ST04]), which we denote

by  $\widetilde{\mathbf{W}} \approx_\varepsilon \mathbf{W}$ , is equivalent to requiring that for all  $x \in \mathbb{R}^n$ , we have

$$\left| x^\top (\widetilde{\mathbf{W}} - \mathbf{W}) x \right| \leq \varepsilon \cdot [x^\top (\mathbf{I} - \mathbf{W}) x]. \quad (7)$$

To compare inequalities (6) and (7), we write  $x = \sum_i c_i v_i$ , where  $v_1, \dots, v_n$  is an orthonormal eigenbasis for  $\mathbf{W}$  with associated eigenvalues  $\lambda_1, \dots, \lambda_n$ . Since  $\mathbf{W}$  is stochastic,  $|\lambda_i| \leq 1$  for all  $i \in [n]$ , the right-hand side of SV inequality (6) becomes

$$\frac{\varepsilon}{2} \cdot \sum_{i \in [n]} c_i^2 \cdot (1 - \lambda_i^2),$$

whereas the right-hand side of ST inequality (7) becomes

$$\varepsilon \cdot \sum_{i \in [n]} c_i^2 \cdot (1 - \lambda_i).$$

Since each  $|\lambda_i| \leq 1$ , the fact that SV approximation implies ST approximation then follows from

$$(1 - \lambda_i^2) = (1 - \lambda_i)(1 + \lambda_i) \leq 2(1 - \lambda_i).$$

However, we also see that inequality (6) can be much stronger than inequality (7) when  $\mathbf{W}$  has eigenvalues  $\lambda_i$  close to -1 (e.g. in a bipartite graph with poor expansion) because then  $1 - \lambda_i^2$  is close to 0, but  $1 - \lambda_i$  is bigger than 2. More generally, inequality (6) requires that  $\widetilde{\mathbf{W}}$  approximates  $\mathbf{W}$  very well on every test vector  $x$  that is concentrated on the eigenvectors whose eigenvalues have *magnitude* close to 1, whereas inequality (7) only requires close approximation on the (signed) eigenvalues that are close to 1.

We remark that another way of ensuring  $\widetilde{\mathbf{W}}$  preserves unit singular values is to replace  $\mathbf{W}^2$  in the SV inequality (6) with the matrix  $|\mathbf{W}|$  where we replace all eigenvalues of  $\mathbf{W}$  with their absolute value rather than their square,<sup>1</sup> so that we have:

$$x^\top (\mathbf{I} - |\mathbf{W}|) x = \sum_{i \in [n]} c_i^2 \cdot (1 - |\lambda_i|).$$

Using  $|\mathbf{W}|$  instead of  $\mathbf{W}^2$  results in an equivalent definition up to a factor of 2 in  $\varepsilon$ , and  $\mathbf{W}^2$  turns out to be convenient to work with.<sup>2</sup> This viewpoint also explains why we stop at  $\mathbf{W}^2$  in the definition and don't explicitly use higher powers; it is simply a convenient proxy for  $|\mathbf{W}|$ , which captures all powers. Indeed, for all  $k \in \mathbb{N}$

$$\mathbf{I} - \mathbf{W}^{2k} \preceq 2 \cdot (\mathbf{I} - |\mathbf{W}|) \preceq 2 \cdot (\mathbf{I} - \mathbf{W}^k).$$

b) *Directed spectral approximation.*: Turning to previous notions of spectral approximation for directed graphs, *standard approximation* [CKP<sup>+</sup>17] generalizes the definition of Spielman and Teng [ST04] by saying  $\widetilde{\mathbf{W}} \approx_\varepsilon \mathbf{W}$  if for all  $x, y \in \mathbb{R}^n$

$$\left| x^\top (\widetilde{\mathbf{W}} - \mathbf{W}) y \right| \leq \frac{\varepsilon}{2} \cdot [x^\top (\mathbf{I} - \mathbf{W}) x + y^\top (\mathbf{I} - \mathbf{W}) y]. \quad (8)$$

<sup>1</sup>Another way of describing  $|\mathbf{W}|$  is as the psd square root of the psd matrix  $\mathbf{W}^2$ .

<sup>2</sup> $|\mathbf{W}| = (\mathbf{W}^2)^{1/2}$ , i.e.,  $|\mathbf{W}|$  is the PSD square root of  $\mathbf{W}^2$ . In Definition I.4, we could similarly replace  $\mathbf{W}\mathbf{W}^\top$  and  $\mathbf{W}^\top \mathbf{W}$  with their PSD square roots and obtain a definition that is equivalent up to a factor of 2 in  $\varepsilon$ .

Equivalently, we can require that for all  $x, y \in \mathbb{R}^n$ ,

$$\left| x^\top (\widetilde{\mathbf{W}} - \mathbf{W}) y \right| \leq \varepsilon \cdot \sqrt{[x^\top (\mathbf{I} - \mathbf{W}) x] \cdot [y^\top (\mathbf{I} - \mathbf{W}) y]}. \quad (9)$$

It can be shown that for undirected graphs, standard approximation is equivalent to the condition of Equation 7, so we will also refer to it as standard approximation.

The use of different left and right test vectors  $x$  and  $y$  on the left-hand side is crucial for capturing the asymmetric information in  $\widetilde{\mathbf{W}}$  and  $\mathbf{W}$ . As before, if  $x$  or  $y$  is concentrated on eigenvectors of  $\mathbf{W}$  whose eigenvalues are close to 1, then the right-hand side of ST inequality (9) is close to 0 and  $\widetilde{\mathbf{W}}$  must approximate  $\mathbf{W}$  very well. However, like the standard undirected ST inequality (7), not much is required on eigenvalues near -1. Moreover, asymmetric matrices can have eigenvalues that are not real and are equal to or close to complex numbers of magnitude 1. For example, the eigenvalues of a directed  $n$ -cycle are the complex  $n$ 'th roots of unity.

To address this issue, *unit-circle (UC) approximation* [AKM<sup>+</sup>20], written  $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\varepsilon \mathbf{W}$ , requires that for all complex test vectors  $x, y \in \mathbb{C}^n$ , we have

$$\left| x^* (\widetilde{\mathbf{W}} - \mathbf{W}) y \right| \leq \frac{\varepsilon}{2} \cdot [ \|x\|^2 + \|y\|^2 - |x^* \mathbf{W} x + y^* \mathbf{W} y| ]. \quad (10)$$

That is, we take the complex magnitude of the terms involving  $\mathbf{W}$  on the right-hand side of ST inequality (8). That way, if  $x$  and  $y$  are concentrated on eigenvectors of  $\mathbf{W}$  that have eigenvalue near some complex number  $\mu$  of magnitude 1, we require that  $\widetilde{\mathbf{W}}$  approximates  $\mathbf{W}$  very well. For example, consider the case where  $\mathbf{W}$  is normal, i.e., has an orthonormal basis of complex eigenvectors  $v_1, \dots, v_n$  and with corresponding complex eigenvalues  $\lambda_1, \dots, \lambda_n$ . Then if we write  $x = \sum_i c_i v_i$  and  $y = \sum_i d_i v_i$ , the right-hand side of UC inequality (10) becomes:

$$\frac{\varepsilon}{2} \cdot \sum_{i \in [n]} (|c_i|^2 + |d_i|^2) - \left| \sum_{i \in [n]} (|c_i|^2 + |d_i|^2) \cdot \lambda_i \right|. \quad (11)$$

If  $x$  and  $y$  are concentrated on eigenvalues  $\lambda_i \approx \mu$  where  $|\mu| = 1$ , then this expression will be close to 0. Unit-circle approximation has valuable properties not enjoyed by standard approximation, in particular being preserved under powering: If  $\widetilde{\mathbf{W}}$  is an  $\varepsilon$ -UC approximation of  $\mathbf{W}$ , then for every positive integer  $k$ ,  $\widetilde{\mathbf{W}}^k$  is an  $O(\varepsilon)$ -UC approximation of  $\mathbf{W}^k$ ; note that the quality of approximation does not degrade with  $k$ . This property was crucial for the results of [AKM<sup>+</sup>20].

However, UC approximation has two limitations compared to SV approximation. First, UC expression (11) is only small if  $x$  and/or  $y$  is concentrated on eigenvalues that are all close to the *same* point  $\mu$  on the complex unit circle. Even in the undirected case, if  $x$  and  $y$  are mixtures of eigenvectors with eigenvalue close to 1 and eigenvalue close to -1, then there will be cancellations in the second term of UC expression (11) and the result will not be small. Second, some properties of asymmetric matrices are more directly captured by *singular*

*values* than eigenvalues, since singular values treat the domain and codomain as distinct. For example, the second-largest singular value of  $\mathbf{W}$  equals 1 if and only if there is a probability distribution  $\pi$  on vertices that does not mix at all in one step (i.e.,  $\|\mathbf{W}\pi - u\| = \|\pi - u\|$ , where  $u = \mathbf{1}/n$  and  $\pi \neq u$ ), but the latter can hold even when all nontrivial eigenvalues have magnitude strictly smaller than 1.

To see how SV approximation addresses these limitations, let  $\sigma_1, \dots, \sigma_n \geq 0$  be the singular values of  $\mathbf{W}$ , let  $u_1, u_2, \dots, u_n \in \mathbb{C}^n$  the corresponding left-singular vectors of  $\mathbf{W}$ , and let  $v_1, \dots, v_n \in \mathbb{C}^n$  the corresponding right-singular vectors. If we write  $x = \sum_i c_i u_i$  and  $y = \sum_i d_i v_i$ , then the right-hand side of SV inequality (3) becomes:

$$\frac{\varepsilon}{4} \cdot \left[ \sum_{i \in [n]} (|c_i|^2 + |d_i|^2) \cdot (1 - \sigma_i^2) \right]. \quad (12)$$

Consequently, SV-approximation requires high-quality approximation if  $x$  is concentrated on left-singular vectors of singular value close to 1 and/or  $y$  is concentrated on right-singular vectors of singular value close to 1. (For the ‘‘or’’ interpretation, use the formulation of SV approximation in terms of inequality (5).) To compare with UC expression (11), let us consider what happens with a *normal* matrix, where  $u_i = v_i$  and  $\sigma_i = |\lambda_i|$ . In this case, SV expression (12) amounts to bringing the absolute value of UC expression (11) inside the summation (and squaring, which only makes a factor of 2 difference), to avoid cancellations between eigenvalues of different phases.

Furthermore, for non-normal matrices, SV approximation retains the asymmetry of  $\mathbf{W}$  even on the right-hand side, by always using  $x$  on the left of  $\mathbf{W}$  (thus relating to its decomposition into left singular vectors) and  $y$  on the right of  $\mathbf{W}$  (thus relating to its decomposition into right singular vectors). Indeed, this feature allows us to even extend the definition of SV approximation to non-square matrices.

Following the above intuitions, we prove that SV approximation is indeed strictly stronger than the previous notions of approximation, even for undirected graphs:

**Theorem 1.5.** *For all doubly stochastic matrices  $\mathbf{W}$  and  $\widetilde{\mathbf{W}}$ , if  $\widetilde{\mathbf{W}} \overset{svn}{\approx}_\varepsilon \mathbf{W}$ , then  $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\varepsilon \mathbf{W}$  (and hence  $\widetilde{\mathbf{W}} \approx_\varepsilon \mathbf{W}$ ). On the other hand, for every  $n \in \mathbb{N}$  there exist random walk matrices  $\widetilde{\mathbf{W}}, \mathbf{W}$  for  $n$ -node undirected graphs such that  $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_{O(1/\sqrt{n})} \mathbf{W}$ , but it is not the case that  $\widetilde{\mathbf{W}} \overset{svn}{\approx}_{.3} \mathbf{W}$ .*

Since UC approximation implies standard approximation, we likewise separate SV from standard approximation. Finally, we note that our separation implies that several useful properties enjoyed by SV approximation, such as preservation under products, are not satisfied by UC approximation.

### C. Properties of SV Approximation

SV approximation enjoys a number of novel properties not known to be possessed by previous notions of spectral approximation. Most striking is the fact that directed approximation

reduces to undirected approximation. To formulate this, we define the symmetric lift of a matrix:

**Definition I.6.** Given  $\mathbf{W} \in \mathbb{C}^{m \times n}$ , let the *symmetric lift* of  $\mathbf{W}$  be defined as

$$\text{slift}(\mathbf{W}) \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0}^{n \times n} & \mathbf{W}^* \\ \mathbf{W} & \mathbf{0}^{m \times m} \end{bmatrix}.$$

Graph theoretically, the symmetric lift of  $\mathbf{W}$  is the following standard operation: Given our directed graph  $G$  on  $n$  vertices with random-walk matrix  $\mathbf{W}$ , we lift  $G$  to an undirected bipartite graph  $H$  with  $n$  vertices on each side, where we connect left-vertex  $i$  to right-vertex  $j$  if there is a directed edge from  $i$  to  $j$  in  $G$ . Then  $\text{slift}(\mathbf{W})$  is the random-walk matrix of  $H$ .

**Theorem I.7.** Let  $\mathbf{W}$  and  $\widetilde{\mathbf{W}}$  be doubly stochastic matrices. Then  $\mathbf{W} \stackrel{sv}{\approx}_\varepsilon \widetilde{\mathbf{W}}$  if and only if  $\text{slift}(\widetilde{\mathbf{W}}) \stackrel{sv}{\approx}_\varepsilon \text{slift}(\mathbf{W})$ .

Thus, for the first time (as far as we know), sparsification of directed graphs reduces directly to sparsification of undirected graphs. It would be very interesting to obtain a similar reduction for other algorithmic problems in spectral graph theory, such as solving Laplacian systems.

Another novel property of SV approximation is that it is preserved under products:

**Theorem I.8.** Let  $\mathbf{W}_1, \dots, \mathbf{W}_k$  and  $\widetilde{\mathbf{W}}_1, \dots, \widetilde{\mathbf{W}}_k$  be doubly stochastic matrices such that  $\widetilde{\mathbf{W}}_i \stackrel{sv}{\approx}_\varepsilon \mathbf{W}_i$  for each  $i \in [k]$ . Then  $\widetilde{\mathbf{W}}_1 \widetilde{\mathbf{W}}_2 \cdots \widetilde{\mathbf{W}}_k \stackrel{sv}{\approx}_{\varepsilon + O(\varepsilon^2)} \mathbf{W}_1 \mathbf{W}_2 \cdots \mathbf{W}_k$ .

Notably the approximation error does not grow with the number  $k$  of matrices being multiplied. This property does not hold for UC approximation, only the weaker property of preservation under powering, i.e.,  $\mathbf{W}_1 = \mathbf{W}_2 = \cdots = \mathbf{W}_k$  and  $\widetilde{\mathbf{W}}_1 = \widetilde{\mathbf{W}}_2 = \cdots = \widetilde{\mathbf{W}}_k$ .

In addition, SV approximation is preserved under multiplication on the left and right by arbitrary matrices of bounded spectral norm. Indeed, it can be seen as the ‘‘closure’’ of standard approximation under this operation (up to a factor of 2).

**Theorem I.9.** The following hold for all doubly stochastic matrices  $\mathbf{W}$  and  $\widetilde{\mathbf{W}}$ :

- 1) If  $\widetilde{\mathbf{W}} \stackrel{sv}{\approx}_\varepsilon \mathbf{W}$  then for all complex matrices  $\mathbf{U}$  and  $\mathbf{V}$  of spectral norm at most 1, we have  $\mathbf{U}\widetilde{\mathbf{W}}\mathbf{V} \stackrel{sv}{\approx}_\varepsilon \mathbf{U}\mathbf{W}\mathbf{V}$ , and hence  $\mathbf{U}\widetilde{\mathbf{W}}\mathbf{V} \approx_\varepsilon \mathbf{U}\mathbf{W}\mathbf{V}$ .
- 2) If for all complex matrices  $\mathbf{U}$  and  $\mathbf{V}$  of spectral norm at most 1, we have  $\mathbf{U}\widetilde{\mathbf{W}}\mathbf{V} \approx_\varepsilon \mathbf{U}\mathbf{W}\mathbf{V}$  then  $\widetilde{\mathbf{W}} \stackrel{sv}{\approx}_{2\varepsilon} \mathbf{W}$ .

Since  $\mathbf{U}\mathbf{W}\mathbf{V}$  and  $\mathbf{U}\widetilde{\mathbf{W}}\mathbf{V}$  need not be doubly stochastic matrices, Theorem I.9 uses the generalization of SV approximation to more general matrices, which can be found in the full version of the paper.

Recall that standard spectral sparsifiers [ST04] are also *cut sparsifiers* [BK00]. That is, if  $\widetilde{G}$  is an  $\varepsilon$ -approximation of  $G$ , then for every set  $S$  of vertices, the weight of the cut  $S$  in  $\widetilde{G}$  is within a  $(1 \pm \varepsilon)$  factor of the weight of  $S$  in  $G$ . Indeed, if

we take the test vector  $x$  to be the characteristic vector of the set  $S$  in inequality (7), we obtain

$$|\text{Cut}_{\widetilde{G}}(S) - \text{Cut}_G(S)| \leq \varepsilon \cdot \text{Cut}_G(S), \quad (13)$$

where  $\text{Cut}(\cdot)$  is as in Theorem I.1.

Similarly, we can obtain a combinatorial consequence of SV approximation, by taking  $x$  to be a characteristic vector of a set  $S$  of vertices and taking  $y$  to be a characteristic vector of a set  $T$  of vertices. This yields:

**Proposition I.10.** Let  $\widetilde{\mathbf{W}}$  and  $\mathbf{W}$  be doubly stochastic  $n \times n$  matrices and suppose that  $\widetilde{\mathbf{W}} \stackrel{sv}{\approx}_\varepsilon \mathbf{W}$ . Then for every two subsets  $S, T \subseteq [n]$ , we have

$$\begin{aligned} & |\text{Cut}_{\widetilde{\mathbf{W}}}(S, T) - \text{Cut}_{\mathbf{W}}(S, T)| \\ & \leq \frac{\varepsilon}{2} \cdot \sqrt{\text{Cut}_{\mathbf{W}\mathbf{W}^\top}(S) \cdot \text{Cut}_{\mathbf{W}^\top\mathbf{W}}(T)}. \end{aligned}$$

Note that  $\mathbf{W}^\top\mathbf{W}$  (resp.,  $\mathbf{W}\mathbf{W}^\top$ ) is the transition matrix for the *forward-backward walk* (resp. backward-forward walk), namely where we take one step using a forward edge of the graph followed by one step using a backward edge.

Let us interpret Proposition I.10. First, consider the case that  $\mathbf{W} = \mathbf{J}$ , the matrix with every entry equal to  $1/n$  (the random-walk matrix for the complete graph with self-loops). Then the distribution  $\mu_{\text{edge}}$  on pairs  $(i, j)$  in the definition of  $\text{Cut}_{\mathbf{W}}$  (Theorem I.1) has  $i$  and  $j$  as uniform and independent vertices, and the same is true for  $\text{Cut}_{\mathbf{W}\mathbf{W}^\top}$  and  $\text{Cut}_{\mathbf{W}^\top\mathbf{W}}$ . Thus, Proposition I.10 says:

$$\begin{aligned} & |\text{Cut}_{\widetilde{\mathbf{W}}}(S, T) - \mu(S) \cdot \mu(T)| \\ & \leq \frac{\varepsilon}{2} \cdot \sqrt{\mu(S) \cdot (1 - \mu(S)) \cdot \mu(T) \cdot (1 - \mu(T))}, \end{aligned}$$

where  $\mu(S) = |S|/n$  and  $\mu(T) = |T|/n$  are the stationary probabilities of  $S$  and  $T$ , respectively. This amounts to a restatement of the Expander Mixing Lemma (cf., [Vad12, Lemma 4.15]); indeed  $\widetilde{\mathbf{W}} \stackrel{sv}{\approx}_\varepsilon \mathbf{J}$  if and only if  $\widetilde{\mathbf{W}}$  is a spectral expander with all nontrivial singular values at most  $\varepsilon/2$ .

Next, let’s consider the case that  $T = S^c$ . Since  $\text{Cut}_{\mathbf{W}\mathbf{W}^\top}(S) = \text{Cut}_{\mathbf{W}\mathbf{W}^\top}(S^c)$ , SV approximation implies that:

$$|\text{Cut}_{\widetilde{\mathbf{W}}}(S) - \text{Cut}_{\mathbf{W}}(S)| \leq \frac{\varepsilon}{2} \cdot \sqrt{\text{Cut}_{\mathbf{W}\mathbf{W}^\top}(S) \cdot \text{Cut}_{\mathbf{W}^\top\mathbf{W}}(S)} \quad (14)$$

We claim that (14) is stronger than the standard notion of a cut approximator (13). Indeed, it can be shown that

$$\text{Cut}_{\mathbf{W}\mathbf{W}^\top}(S) \leq 2 \cdot \text{Cut}_{\mathbf{W}}(S),$$

and similarly for  $\text{Cut}_{\mathbf{W}^\top\mathbf{W}}(T)$ . The reason is that if a backward-forward walk crosses between  $S$  and  $S^c$ , then it must cross between  $S$  and  $S^c$  in either the first step or in the second step. Similar reasoning shows that

$$\text{Cut}_{\mathbf{W}\mathbf{W}^\top}(S) \leq 2 \cdot \text{Uncut}_{\mathbf{W}}(S),$$

and similarly for  $\text{Cut}_{\mathbf{W}^\top\mathbf{W}}(T)$ . Thus SV approximation also implies:

$$|\text{Uncut}_{\widetilde{G}}(S) - \text{Uncut}_G(S)| \leq \varepsilon \cdot \text{Uncut}_G(S), \quad (15)$$

Thus, we conclude that an SV-approximator not only approximates every cut to within a small additive error that is scaled by the weight of the cut edges (as in (13)), but also scaled by the weight of the *uncut* edges.

#### D. Algorithmic Results

Even though SV approximation is stronger than previously considered notions of spectral approximation, we show that it still admits sparsification:

**Theorem I.11.** *There is a randomized nearly-linear time algorithm that given a regular directed graph  $G$  with  $n$  vertices and  $m$  edges, integer edge weights in  $[0, U]$ , and random-walk matrix  $\mathbf{W}$ , and  $\varepsilon > 0$ , whp outputs a weighted graph  $\tilde{G}$  with at most  $O(n\varepsilon^{-2} \cdot \text{poly}(\log(nU)))$  edges such that its random-walk matrix  $\tilde{\mathbf{W}}$  satisfies  $\tilde{\mathbf{W}} \stackrel{sv}{\approx}_\varepsilon \mathbf{W}$ .*

A more general theorem that also applies to Eulerian digraphs is stated in the full version of the paper. Prior to this work, it was open whether or not even UC-sparsifiers with  $O(n \cdot \text{poly}(\log n, 1/\varepsilon))$  edges existed for all unweighted regular digraphs. Instead, it was only known how to UC-sparsify powers of a random walk matrix in such a way that the number of edges increases by at most a polylogarithmic factor compared to the original graph (rather than decrease the number of edges) [AKM<sup>+</sup>20].

By Theorem I.7, it suffices to prove Theorem I.11 for undirected bipartite graphs. We obtain the latter via an undirected sparsification algorithms based on expander partitioning [ST04]. It remains an open question whether algorithms based on edge sampling can yield SV approximation or unit circle approximation, even in undirected graphs. The standard approach to spectral sparsification of undirected graphs via sampling, namely keeping each edge independently with probability proportional to its effective resistance [SS08], does not work for SV or UC approximation. For example, this method does not exactly preserve degrees, which we show is necessary for SV sparsification.<sup>3</sup>

However, we remark that the work of Chu, Gao, Peng, Sawlani, and Wang [CGP<sup>+</sup>18] does yield something closer to sparsification via *degree preserving* sampling for standard approximation [CKP<sup>+</sup>17] but not unit circle approximation. They show that if one has a directed graph and decomposes it into short “cycles” without regard for the direction of the edges on the cycle, then one can sparsify by randomly eliminating either the clockwise or counterclockwise edges on each such cycle. We build on their procedure and use it to obtain SV sparsification (and hence, unit circle) by showing that this technique obtains SV approximation, even if the cycles are not short, as long as (a) all the cycles are within expanding subgraphs, and (b) the cycles alternate between forward and backward edges. (Note that such alternating cycles in a di-

rected graph correspond to ordinary cycles in the undirected lift given by Theorem I.7.)

Given Theorem I.11, we obtain our algorithm for longer walks (Theorem I.3) as follows:

- 1) First, we show that we can SV-sparsify the *squares* of random-walk matrices of Eulerian digraphs; we follow the approach of [CKK<sup>+</sup>18] by locally sparsifying the bipartite complete graphs that form around each vertex when squaring, and then applying Theorem I.11 to globally sparsify further. We likewise show the “de-randomized square” approach used in [RV05], [PS14], [MRSV21], [AKM<sup>+</sup>20] gives a square sparsifier.
- 2) Then we SV-sparsify arbitrary powers of 2 by repeatedly squaring and sparsifying, using the fact that SV approximation is preserved under powering. During this process, we need to ensure that the ratio between the largest and smallest edge weights remains bounded. We do this by restricting to graphs that have second-largest singular value bounded away from 1 by  $1/\text{poly}(nU\ell)$ , which allows us to discard edge weights that get too small and make small patches to preserve degrees. We can achieve this assumption on the second-largest singular value by adding a small amount of laziness to our initial graph.
- 3) Then to sparsify arbitrary powers  $\mathbf{W}^\ell$ , we can multiply sparsifiers for the powers of 2 appearing in the binary representation of  $\ell$ . For example, to get a sparsifier for  $\mathbf{W}^7$ , we multiply sparsifiers for  $\mathbf{W}^4$ ,  $\mathbf{W}^2$ , and  $\mathbf{W}^1$ , sparsifying and eliminating small edge weights again in each product. The use of SV approximation plays an important role in the analysis of this algorithm, because it has the property that the product of the approximations of the powers still approximates the product of the true powers (Theorem I.8).
- 4) Given Theorem I.11, we obtain Theorem I.3 for general directed graphs by using [CKK<sup>+</sup>18] to compute a high-precision estimate of the stationary distribution, which allows us to construct an Eulerian graph whose random-walk matrix closely approximates that of the original graph. SV-sparsifying the  $\ell$ 'th power of the Eulerian graph gives us a graph all of whose Cut and Uncut values approximate the  $\ell$ 'th power of our input graph. The use of [CKK<sup>+</sup>18] to estimate the stationary distribution and the introduction of laziness to  $\mathbf{W}$  both incur a small additive error  $\delta$ , but we can absorb that into  $\varepsilon$  by setting  $\delta = 1/\text{poly}(nU/s)$  and observing that  $\text{Cut}_{G^\ell}(S)$  and  $\text{Uncut}_{G^\ell}(S)$  are at least  $1/\text{poly}(nU/s)$  (if nonzero).

Our final contribution concerns algorithms for solving directed Laplacian systems. The recursive identities used for solving undirected Laplacian systems, while behaving nicely with respect to PSD approximation, do not behave as nicely with respect to the previous approximation definitions for directed graphs. This led to different, more sophisticated recursions with a more involved analysis of the error [CKP<sup>+</sup>17], [CKK<sup>+</sup>18], [AKM<sup>+</sup>20], [KMG22]. We make progress towards simplifying the recursion and analysis of solving di-

<sup>3</sup>Unlike standard spectral approximation, degrees cannot be fixed just by adding self loops; indeed, self-loops ruin bipartiteness and periodicity, which are properties that UC and SV approximation retain (as they are captured by eigenvalues like -1 or other roots of unity).

rected Laplacian linear systems in the following way. We show that a simpler recursion, a variant of the one used by Peng and Spielman [PS14] (Equation (17) below), and a simpler analysis suffice if the directed Laplacian is normal (i.e., unitarily diagonalizable) and we perform all sparsification with respect to SV approximation. Note that this result is the only result in our paper that relies on a normality assumption; the aforementioned sparsification results hold for all Eulerian directed graphs.

**Theorem I.12.** *For a doubly stochastic normal matrix  $\mathbf{W} \in \mathbb{R}^{n \times n}$  with  $\|\mathbf{W}\| \leq 1$ , let  $\mathbf{W} = \mathbf{W}_0, \dots, \mathbf{W}_{k-1}$  be a sequence of matrices such that for  $\varepsilon \leq 1/4k$  we have*

$$\mathbf{W}_i \stackrel{sv}{\approx}_\varepsilon \mathbf{W}_{i-1}^2 \quad \forall 0 < i < k, \quad (16)$$

and

$$\mathbf{P}_i = \frac{1}{2} [\mathbf{I} + (\mathbf{I} + \mathbf{W}_i)\mathbf{P}_{i+1}(\mathbf{I} + \mathbf{W}_i)] \quad \forall 0 \leq i < k \quad (17)$$

defining the Peng-Spielman squaring recursion.

Then, for a matrix  $\mathbf{P}_k$ , such that  $\left\| (\mathbf{I} - \mathbf{W}^{2^k})^{\frac{1}{2}} [\mathbf{P}_k - (\mathbf{I} - \mathbf{W}^{2^k})^+] (\mathbf{I} - \mathbf{W}^{2^k})^{\frac{1}{2}} \right\| \leq O(k\varepsilon)$ , we have

$$\begin{aligned} & \|\mathbf{P}_0(\mathbf{I} - \mathbf{W}) - \mathbf{I}\|_{\mathbf{B}} \\ &= \left\| (\mathbf{I} - \mathbf{W})^{\frac{1}{2}} [\mathbf{P}_0 - (\mathbf{I} - \mathbf{W})^+] (\mathbf{I} - \mathbf{W})^{\frac{1}{2}} \right\| \leq O(k^2\varepsilon) \end{aligned}$$

where  $\mathbf{B} = ((\mathbf{I} - \mathbf{W})^{1/2})^*(\mathbf{I} - \mathbf{W})^{1/2}$ .

Theorem I.12 says that that we can compute a good preconditioner  $\mathbf{P}_0$  for the Laplacian  $\mathbf{I} - \mathbf{W}$  by repeatedly computing SV-approximate squares (eq. 16) and use the simple recurrence (eq. 17, starting with a preconditioner  $\mathbf{P}_k$  for a sufficiently large power of  $\mathbf{W}$ ). Generally,  $\mathbf{P}_k$  is easy to obtain for a large enough  $k = O(\log n)$  since  $\mathbf{W}^{2^k}$  is well-approximated by a complete graph (assuming the original graph is connected and aperiodic).

### E. Open Problems

One open problem is to determine whether or not it is possible to obtain linear-sized sparsifiers. Recall that undirected graphs have sparsifiers with respect to standard spectral approximation that have only  $O(n/\varepsilon^2)$  nonzero edge weights [BSS12]. If this result could be extended to obtain linear-sized SV-sparsifiers of undirected graphs, we would also have linear-sized sparsifiers for directed graphs by Theorem I.7, which would be a new result even for standard approximation [CKP<sup>+</sup>17].

### REFERENCES

- [ACL06] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local graph partitioning using pagerank vectors. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, 21-24 October 2006, Berkeley, California, USA, Proceedings, pages 475–486. IEEE Computer Society, 2006.
- [Ahm20] AmirMahdi Ahmadinejad. *Computing Stationary Distributions: Perron Vectors, Random Walks, and Ride-Sharing Competition*. Stanford University, 2020.
- [AJSS19] AmirMahdi Ahmadinejad, Arun Jambulapati, Amin Saberi, and Aaron Sidford. Perron-frobenius theory in nearly linear time: Positive eigenvectors, m-matrices, graph kernels, and other applications. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 1387–1404. SIAM, 2019.
- [AKM<sup>+</sup>20] AmirMahdi Ahmadinejad, Jonathan A. Kelner, Jack Murtagh, John Peebles, Aaron Sidford, and Salil P. Vadhan. High-precision estimation of random walks in small space. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1295–1306. IEEE, 2020.
- [AM85] Noga Alon and V. D. Milman.  $\lambda_1$ , isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.
- [BK00] András A. Benczúr and David R. Karger. Augmenting undirected edge connectivity in  $\delta(n^2)$  time. *J. Algorithms*, 37(1):2–36, 2000.
- [BSS12] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.
- [CCL<sup>+</sup>15] Dehua Cheng, Yu Cheng, Yan Liu, Richard Peng, and Shang-Hua Teng. Spectral sparsification of random-walk matrix polynomials. *CoRR*, abs/1502.03496, 2015.
- [CGP<sup>+</sup>18] Timothy Chu, Yu Gao, Richard Peng, Sushant Sachdeva, Saurabh Sawlani, and Junxing Wang. Graph sparsification, spectral sketches, and faster resistance computation, via short cycle decompositions. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 361–372. IEEE Computer Society, 2018.
- [CKK<sup>+</sup>18] Michael B. Cohen, Jonathan A. Kelner, Rasmus Kyng, John Peebles, Richard Peng, Anup B. Rao, and Aaron Sidford. Solving directed laplacian systems in nearly-linear time through sparse LU factorizations. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 898–909. IEEE Computer Society, 2018.
- [CKM<sup>+</sup>11] Paul F. Christiano, Jonathan A. Kelner, Aleksander Madry, Daniel A. Spielman, and Shang-Hua Teng. Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 273–282. ACM, 2011.
- [CKP<sup>+</sup>16] Michael B. Cohen, Jonathan A. Kelner, John Peebles, Richard Peng, Aaron Sidford, and Adrian Vladu. Faster algorithms for computing the stationary distribution, simulating random walks, and more. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 583–592. IEEE Computer Society, 2016.
- [CKP<sup>+</sup>17] Michael B. Cohen, Jonathan A. Kelner, John Peebles, Richard Peng, Anup B. Rao, Aaron Sidford, and Adrian Vladu. Almost-linear-time algorithms for markov chains and new spectral primitives for directed graphs. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 410–419. ACM, 2017.
- [CLTW22] Lijie Chen, Xin Lyu, Avishay Tal, and Hongxun Wu. Personal communication, November 2022.
- [DMVZ20] Dean Doron, Jack Murtagh, Salil P. Vadhan, and David Zuckerman. Spectral sparsification via bounded-independence sampling. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPICs*, pages 39:1–39:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.
- [GV22] Louis Golowich and Salil Vadhan. Pseudorandomness of expander random walks for symmetric functions and permutation branching programs. In Anindya De, editor, *Proceedings of the*



- 37th Computational Complexity Conference (CCC '22), volume 234 of *LIPICs*, pages 27:1–27:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. Full version posted as ECCC TR22-024.
- [HPV21] William M. Hoza, Edward Pyne, and Salil P. Vadhan. Pseudorandom generators for unbounded-width permutation branching programs. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCs 2021, January 6-8, 2021, Virtual Conference*, volume 185 of *LIPICs*, pages 7:1–7:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.
- [INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In Frank Thomson Leighton and Michael T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 356–364. ACM, 1994.
- [Kel21] Zander Kelley. An alternative proof that inw fools permutation programs. Unpublished manuscript, February 2021.
- [KLOS14] Jonathan A. Kelner, Yin Tat Lee, Lorenzo Orecchia, and Aaron Sidford. An almost-linear-time algorithm for approximate max flow in undirected graphs, and its multicommodity generalizations. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 217–226. SIAM, 2014.
- [KM09] Jonathan A. Kelner and Aleksander Madry. Faster generation of random spanning trees. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 13–21. IEEE Computer Society, 2009.
- [KMG22] Rasmus Kyng, Simon Meierhans, and Maximilian Probst Gutenberg. Derandomizing directed random walks in almost-linear time. *CoRR*, abs/2208.10959, 2022.
- [KVV04] Ravi Kannan, Santosh S. Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, 2004.
- [LRS13] Yin Tat Lee, Satish Rao, and Nikhil Srivastava. A new approach to computing maximum flows using electrical flows. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 755–764. ACM, 2013.
- [MRSV21] Jack Murtagh, Omer Reingold, Aaron Sidford, and Salil P. Vadhan. Deterministic approximation of random walks in small space. *Theory Comput.*, 17:1–35, 2021.
- [MST15] Aleksander Madry, Damian Straszak, and Jakub Tarnawski. Fast generation of random spanning trees and the effective resistance metric. In Piotr Indyk, editor, *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 2019–2036. SIAM, 2015.
- [OSV12] Lorenzo Orecchia, Sushant Sachdeva, and Nisheeth K. Vishnoi. Approximating the exponential, the lanczos method and an  $\tilde{O}(m)$ -time spectral algorithm for balanced separator. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1141–1160. ACM, 2012.
- [PS14] Richard Peng and Daniel A. Spielman. An efficient parallel solver for SDD linear systems. *STOC*, 2014.
- [PV21] Edward Pyne and Salil Vadhan. Pseudodistributions That Beat All Pseudorandom Generators (Extended Abstract). In *Proceedings of the 36th Annual Computational Complexity Conference (CCC)*, pages 33:1–33:15, 2021.
- [Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, 2008.
- [RTV06] Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Pseudorandom walks on regular digraphs and the RL vs. L problem. In Jon M. Kleinberg, editor, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 457–466. ACM, 2006.
- [RV05] Eyal Rozenman and Salil P. Vadhan. Derandomized squaring of graphs. In Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th International Workshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005, Proceedings*, volume 3624 of *Lecture Notes in Computer Science*, pages 436–447. Springer, 2005.
- [SS08] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 563–568. ACM, 2008.
- [ST04] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In László Babai, editor, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 81–90. ACM, 2004.
- [SZ99] Michael E. Saks and Shiyu Zhou.  $\text{BP}_{\text{Hspace}(s)} \text{subsetq dspace}(s^{3/2})$ . *J. Comput. Syst. Sci.*, 58(2):376–403, 1999.
- [Vad12] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012.
- [vdBGJ<sup>+</sup>22] Jan van den Brand, Yu Gao, Arun Jambulapati, Yin Tat Lee, Yang P. Liu, Richard Peng, and Aaron Sidford. Faster maxflow via improved dynamic spectral vertex sparsifiers. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 543–556. ACM, 2022.
- [vdBLL<sup>+</sup>21] Jan van den Brand, Yin Tat Lee, Yang P. Liu, Thatchaphol Saranurak, Aaron Sidford, Zhao Song, and Di Wang. Minimum cost flows, mdps, and l1-regression in nearly linear time for dense instances. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 859–869. ACM, 2021.