

Accelerated Gossip in Networks of Given Dimension Using Jacobi Polynomial Iterations*

Raphaël Berthier[†], Francis Bach[†], and Pierre Gaillard[†]

Abstract. Consider a network of agents connected by communication links, where each agent holds a real value. The gossip problem consists in estimating the average of the values diffused in the network in a distributed manner. We develop a method for solving the gossip problem that depends only on the spectral dimension of the network, that is, in the communication network set-up, the dimension of the space in which the agents live. This contrasts with previous work that required the spectral gap of the network as a parameter, or suffered from slow mixing. Our method shows an important improvement over existing algorithms in the nonasymptotic regime, i.e., when the values are far from being fully mixed in the network. Our approach stems from a polynomial-based point of view on gossip algorithms, as well as an approximation of the spectral measure of the graphs with a Jacobi measure. We show the power of the approach with simulations on various graphs, and with performance guarantees on graphs of known spectral dimension, such as grids and random percolation bonds. An extension of this work to distributed Laplacian solvers is discussed. As a side result, we also use the polynomial-based point of view to show the convergence of the message passing algorithm for gossip of Moallemi and Van Roy on regular graphs. The explicit computation of the rate of the convergence shows that message passing has a slow rate of convergence on graphs with small spectral gap.

Key words. gossip, consensus, averaging, multiagent, distributed, polynomial iterations, Jacobi orthogonal polynomials, Laplacian solvers

AMS subject classifications. 90C35, 68M14, 68W15

DOI. 10.1137/19M1244822

1. Introduction. The averaging problem, or gossip problem, is a fundamental primitive of distributed algorithms. Given a network composed of agents and undirected communication links between them, we assign to each agent v a real value ξ_v , called an observation. The goal is to design an iterative communication procedure allowing each agent to know the average of the initial observations in the network as quickly as possible.

The landmark paper [6] suggests the natural following protocol to solve the averaging problem: at each iteration, each agent replaces his current observation by some average of the observations of its neighbors in the network. We will refer to this method in the following by the term *simple gossip*. More precisely, we are given a weight matrix $W = (W_{v,w})_{v,w \in V}$, called the gossip matrix, indexed by the vertices $v \in V$ of the network graph, satisfying the

*Received by the editors February 13, 2019; accepted for publication (in revised form) October 28, 2019; published electronically February 6, 2020.

<https://doi.org/10.1137/19M1244822>

Funding: This work was supported by the European Research Council (grant SEQUOIA 724063) and by the DGA.

[†]Inria and École Normale Supérieure, PSL Research University, 75006 Paris, France (raphael.berthier@inria.fr, francis.bach@inria.fr, pierre.gaillard@inria.fr).

property that $W_{v,w}$ is nonzero only if $v \sim w$, that is v and w are connected in the graph. Then the simple gossip iteration writes

$$(1.1) \quad x_v^0 = \xi_v, \quad x_v^{t+1} = \sum_{w:w \sim v} W_{v,w} x_w^t, \quad t \geq 0.$$

The paper [6] proves the linear convergence of the observations to their average.

However, the rate of the linear convergence was shown to worsen significantly in many networks of interest as the size of the network increases. More precisely, define the diameter D of the network as the largest number of communication links needed to connect any two agents. While, obviously, D steps of averaging are needed for any gossip method to spread information in the network, the simple gossip method may require up to $\Theta(D^2)$ communication steps to estimate the average, as for instance on the line graph, the two-dimensional grid, or the random geometric graph (see [8] or [6, section IV.A]). To reach the $O(D)$ bound, a diverse set of ideas were proposed, including second-order recursions [7, 25], message passing algorithms [18], lifted Markov chain techniques [28], methods using Chebychev polynomial iterations [2, 27], inspiration arising from advection-diffusion processes [26], and geographic gossip [10], a method that uses the knowledge of the location of the agents on a field. To the best of our knowledge, all of these accelerated methods assume that the agents hold additional information about the network graph, such as its spectral gap. For instance, second-order methods typically take the form (see [7])

$$(1.2) \quad \begin{aligned} x_v^0 &= \xi_v, & x_v^1 &= \sum_{w:w \sim v} W_{v,w} x_w^0, \\ x_v^{t+1} &= \omega \sum_{w:w \sim v} W_{v,w} x_w^t + (1 - \omega) x_v^{t-1}, & t &\geq 1, \end{aligned}$$

where ω is some simple function of the spectral gap γ , that is the distance between the largest and the second largest eigenvalues of W . This iteration obtains optimal asymptotic convergence on many graphs, with a relaxation time of the linear convergence on the order of $1/\sqrt{\gamma}$ as $\gamma \rightarrow 0$.

In this paper, we develop a gossip method based not on the spectral gap γ but on the density of eigenvalues of W near the upper edge of the spectrum. Looking at the upper part of the spectrum at a broader scale allows us to improve the local averaging of the gossip algorithm in the regime $t < 1/\sqrt{\gamma}$. This improvement is worthwhile as the spectral gap γ can get arbitrarily small in large graphs. For instance, in the case of the line graph or the two-dimensional grid, the relaxation time $1/\sqrt{\gamma}$ is of the order of the diameter D of the graph. Thus the regime $t < 1/\sqrt{\gamma}$ can be relevant for applications.

Remarkably, the spectral density of W near the upper edge can be described by a very natural parameter: the spectral dimension d . The network is of spectral dimension d if the number of eigenvalues of W in $[1 - \Lambda, 1]$ is of the order of $\Lambda^{d/2}$ for small Λ ($\gamma \ll \Lambda \ll 1$); see section 5.3 for rigorous definitions. We will see with examples that this definition coincides with our intuition of the dimension of the graph, which is the dimension of the manifold on which the agents live. For instance, the grid with nodes \mathbb{Z}^d where the nodes at distance 1 are connected is a graph of dimension d . Thus the parameter d is much easier to know than the spectral gap γ .

In real-world situations, the practitioner reasonably knows if the network on which she implements the gossip method is of finite dimension, and if so, she also knows the dimension d . In this paper, we argue that she should run a second-order iteration with time-dependent weights

$$(1.3) \quad \begin{aligned} x_v^0 &= \xi_v, & x_v^1 &= a_0 \sum_{w:w \sim v} W_{v,w} x_w^0 + b_0 x_v^0, \\ x_v^{t+1} &= a_t \sum_{w:w \sim v} W_{v,w} x_w^t + b_t x_v^t - c_t x_v^{t-1}, & t &\geq 1, \end{aligned}$$

where the recurrence weights a_t, b_t, c_t are given by the formulas

$$(1.4) \quad \begin{aligned} a_0 &= \frac{d+4}{2(2+d)}, & b_0 &= \frac{d}{2(2+d)}, \\ a_t &= \frac{(2t+d/2+1)(2t+d/2+2)}{2(t+1+d/2)^2}, & b_t &= \frac{d^2(2t+d/2+1)}{8(t+1+d/2)^2(2t+d/2)}, \\ c_t &= \frac{t^2(2t+d/2+2)}{(t+1+d/2)^2(2t+d/2)}, & t &\geq 1. \end{aligned}$$

The motivation for these choices of weights a_t, b_t, c_t is not obvious at first sight. It follows from a *polynomial-based* point of view on gossip algorithms: it consists in seeing the iterations (1.1), (1.2), and (1.3) as sequences P_0, P_1, P_2, \dots of polynomials in the gossip matrix W . The correspondence is given by the relation $x^t = P_t(W)\xi$ where $x^t = (x_v^t)_{v \in V}$ and $\xi = (\xi_v)_{v \in V}$. This approach is inspired by similar work done in the resolution of linear systems [12] and on the load balancing problem [9]. The choice of an iteration is reframed as the choice of a sequence of polynomials, and the performance of the resulting gossip method depends on the spectrum of W . As the dimension of the graph gives the rate of decrease of the spectral density near the edge of the spectrum, it suggests the sequence of polynomials one should take: we choose a parametrized sequence of polynomials called *Jacobi polynomials* that is well known in the literature on orthogonal polynomials (see Definition SM2.2 of the Jacobi polynomials). This actually leads to the iteration (1.3), which we call the *Jacobi polynomial iteration*.

The Jacobi polynomial iteration (1.3) improves the convergence of the gossip method in the transitive phase $t < 1/\sqrt{\gamma}$ but loses the optimal rate of convergence of second-order gossip because it does not use the spectral gap γ . We argue that in most applications of gossip methods, the asymptotic rate of convergence is not of practical importance, especially if the transient time is long. However, we also build a gossip iteration that uses both parameters d and γ and achieves both the efficiency in the nontransitive regime and the fast rate of convergence.

This resolution of the gossip problem with inner-product free polynomial-based iterations is new, and could lead to other interesting algorithms on other types of graphs. Here, the phrase “inner-product free” comes from the literature on polynomial-based iterations for linear systems [12] and refers to the fact that recurrence coefficients a_t, b_t, c_t are computed without using the gossip matrix W (but parametrized using the knowledge of d). Indeed, as the

knowledge of the gossip matrix W is distributed across the graph, it would be a challenging distributed problem to compute the recurrence coefficients if they depended on W .

Although our work is inspired by iterative methods for linear systems, the Jacobi iteration that we developed for gossip can be transposed into a new idea in this literature which can be useful for the distributed resolution of Laplacian systems over multiagent networks.

Outline of the paper. Section 2 sets some notation used in the remainder of the paper. In section 3, we give simulations in different types of networks of dimensions 2 and 3. We show that the recursion (1.3) brings important benefits over existing methods in the nonasymptotic regime, i.e., when the observations are far from being fully mixed in the graph.

In sections 4–5, we develop the derivation of the Jacobi polynomial iteration. Section 4 describes an optimal way to design polynomial-based gossip algorithms, following the lines of [12, 9], and discusses its feasibility. Section 5 uses the notion of spectral dimension of a graph to inspire the practical Jacobi polynomial iteration (1.3).

In section 6, we present the adaptation of the Jacobi polynomial iteration to the case where the spectral gap γ of W is given to improve the asymptotic rate of convergence.

In section 7, we describe the parallel between gossip methods and iterative methods for linear systems and discuss the contributions that our work can bring to the distributed resolution of Laplacian systems over networks.

Code. The code that generated the simulations is available online [4].

2. Problem setting. A network of agents is modeled by an undirected finite graph $G = (V, E)$, where V is the set of vertices of the graph, or agents, and E the set of edges, or communication links. We assume each agent v holds a real value ξ_v . Our goal is to design an iterative algorithm that quickly gives each agent the average $\bar{\xi} = (1/n) \sum_{v \in V} \xi_v$, where $n = |V|$ is the number of agents. A fundamental operation to estimate the average $\bar{\xi}$ consists in averaging the observations of neighbors in the network. We formalize this notion using a gossip matrix.

Definition 2.1. A gossip matrix $W = (W_{v,w})_{v,w \in V}$ on the graph G is a matrix with entries indexed by the vertices of the graph satisfying the following properties:

- W is nonnegative: for all $v, w \in V$, $W_{v,w} \geq 0$.
- W is supported by the graph G : for all distinct vertices v, w such that $W_{v,w} > 0$, $\{v, w\}$ must be an edge of G .
- W is normalized: for all $v \in V$, $\sum_{w \in V} W_{v,w} = 1$.
- W is symmetric: for all $v, w \in V$, $W_{v,w} = W_{w,v}$.

If W is a gossip matrix and $x = (x_v)_{v \in V}$ is a set of values stored by the agents v , the product Wx is interpreted as the computation by each agent v of a weighted average of the values x_w of its neighbors w in the graph (and of its own value x_v). This average is computed simultaneously for all agents v ; indeed, in this paper we deal only with *synchronous* gossip. Note that we do not need the symmetry assumption on W to interpret W as an averaging operation. This assumption is usual in gossip frameworks as it allows one to use the spectral theory for W , on which our analysis relies heavily. It appears, for instance, in the works [6, 7, 25].

In a d -regular graph G ($\forall v, \deg v = d$), a typical gossip matrix is $W = A(G)/d =$

$(\mathbf{1}_{\{\{v,w\} \in E\}}/d)_{v,w \in V}$, where $A(G)$ is the adjacency matrix of the graph. More generally, if the graph has all vertices of degree bounded by some quantity d_{\max} , then a natural gossip matrix is

$$(2.1) \quad W = I + \frac{1}{d_{\max}}(A - D),$$

where D is the degree matrix, which is the diagonal matrix such that $D_{v,v} = \deg v$.

Note that any gossip matrix W is an operator on \mathbb{R}^V bounded by 1. Indeed, if $x \in \mathbb{R}^V$, by Jensen's inequality,

$$\|Wx\|_2^2 = \sum_{v \in V} x_v^2 = \sum_{v \in V} \left(\sum_{w \in V} W_{v,w} x_w \right)^2 \leq \sum_{v \in V} \sum_{w \in V} W_{v,w} x_w^2 = \sum_{w \in V} x_w^2 = \|x\|_2^2.$$

Definition 2.2 (spectral gap). Denote by $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -1$ the real eigenvalues of the symmetric matrix W . As W is normalized, $W\mathbf{1} = \mathbf{1}$; we can take $\lambda_1 = 1$, which corresponds to the eigenvector $\mathbf{1} = (1, \dots, 1)$. We define

1. the spectral gap $\gamma = 1 - \lambda_2$ as the distance between the two largest eigenvalues of W ;
2. the absolute spectral gap $\tilde{\gamma} = \min(1 - \lambda_2, \lambda_n + 1)$ as the difference between the moduli of the two largest eigenvalues of W in magnitude.

We now discuss different iterations for the gossip problem.

Simple gossip. Simple gossip is a natural algorithm solving the gossip problem that consists in repeatedly averaging values in the graph [6]. More precisely, we choose a gossip matrix W on the graph G , initialize $x^0 = \xi = (\xi_v)_{v \in V}$, and, at each communication round t , compute

$$(2.2) \quad x^{t+1} = Wx^t.$$

Note that the latter equation is simply a compact rewriting of (1.1). We can rewrite this iteration as $x^t = W^t \xi$. Note that in this last equation, we used the notation $.^t$ to denote both the index of x and the power of the square matrix W . We will frequently make use of the indexation $.^t$ when vectors indexed by the vertices (or the edges) also depend on time.

We describe the speed of convergence of this method using ideas from [6].

Proposition 2.3. Let ξ be an arbitrary family of initial observations and x^t the iterates of simple gossip defined by (2.2). Denote by $\tilde{\gamma}$ the absolute spectral gap of W . Then

$$\limsup_{t \rightarrow \infty} \|x^t - \bar{\xi} \mathbf{1}\|_2^{1/t} \leq 1 - \tilde{\gamma}.$$

Moreover, the upper bound is reached if and only if there exists an eigenvector u of W , corresponding to an eigenvalue of magnitude $1 - \tilde{\gamma}$, such that $\langle \xi, u \rangle \neq 0$.

Proof. Let $u^1 = \mathbf{1}/\sqrt{n}, u^2, \dots, u^n$ be the eigenvectors of W associated to the eigenvalues

$\lambda_1, \dots, \lambda_n$, normalized such that $\|u^i\|_2 = 1$. Then

$$\begin{aligned} \|x^t - \bar{\xi}\mathbf{1}\|_2^2 &= \|W^t \xi - \bar{\xi}\mathbf{1}\|_2^2 = \left\| \sum_{i=1}^n \lambda_i^t \langle \xi, u^i \rangle u^i - \langle \xi, u^1 \rangle u^1 \right\|_2^2 = \left\| \sum_{i=2}^n \lambda_i^t \langle \xi, u^i \rangle u^i \right\|_2^2 \\ &\leq \sum_{i=2}^n \|\lambda_i^t \langle \xi, u^i \rangle u^i\|_2^2 = \sum_{i=2}^n \lambda_i^{2t} \langle \xi, u^i \rangle^2 \leq (1 - \tilde{\gamma})^{2t} \sum_{i=2}^n \langle \xi, u^i \rangle^2. \end{aligned}$$

Shift-register gossip. Several acceleration schemes of gossip [7, 25] store some past iterates to compute higher-order recursions (that thus depend on powers of W). For instance, the shift-register iteration of [7] is of the form

$$(2.3) \quad x^0 = \xi, \quad x^1 = W\xi, \quad x^{t+1} = \omega Wx^t + (1 - \omega)x^{t-1},$$

where ω is a parameter that needs to be tuned.

Proposition 2.4 (from [16, Theorem 2]). *Let ξ be an arbitrary family of initial observations and x^t the iterates of shift-register gossip defined in (2.3) with parameter*

$$\omega = 2 \frac{1 - \sqrt{\tilde{\gamma}(1 - \tilde{\gamma}/4)}}{(1 - \tilde{\gamma}/2)^2},$$

where $\tilde{\gamma}$ is the absolute spectral gap of the gossip matrix W . Then

$$\limsup_{t \rightarrow \infty} \|x^t - \bar{\xi}\mathbf{1}\|_2^{1/t} \leq 1 - 2 \frac{\sqrt{\tilde{\gamma}(1 - \tilde{\gamma}/4)} - \tilde{\gamma}/2}{1 - \tilde{\gamma}}.$$

Moreover, the upper bound is reached if and only if there exists an eigenvector u of W , corresponding to an eigenvalue of magnitude $1 - \tilde{\gamma}$, such that $\langle \xi, u \rangle \neq 0$.

The important consequence of this result is that the rate of convergence of the shift-register method behaves like $1 - 2\sqrt{\tilde{\gamma}} + o(\sqrt{\tilde{\gamma}})$ as $\tilde{\gamma} \rightarrow 0$. This differs from simple gossip where the rate of convergence behaves like $1 - \tilde{\gamma}$. This means that in graphs with a small spectral gap, shift-register enjoys a much better rate of convergence than simple gossip: this is why we say that shift-register enjoys an *accelerated* rate of convergence as opposed to simple gossip which has a *diffusive* or *unaccelerated* rate. This effect on the asymptotic rate of convergence can be seen in Figures 2 and 3.

Polynomial gossip. More abstractly, we define a polynomial gossip method as any method combining the past iterates of the simple gossip method:

$$(2.4) \quad x^t = P_t(W)\xi,$$

where P_t is a polynomial of degree smaller than or equal to t satisfying $P_t(1) = 1$. The constraint $P_t(1) = 1$ ensures that $x^t = \bar{\xi}\mathbf{1}$ if all initial observations are the same, i.e., $\xi = \bar{\xi}\mathbf{1}$. The constraint $\deg P_t \leq t$ ensures that the iterate x^t can be computed in t time steps. Simple gossip corresponds to the particular case of the polynomial $P_t(\lambda) = \lambda^t$. Shift-register gossip is a polynomial gossip method whose corresponding polynomials can be expressed using the

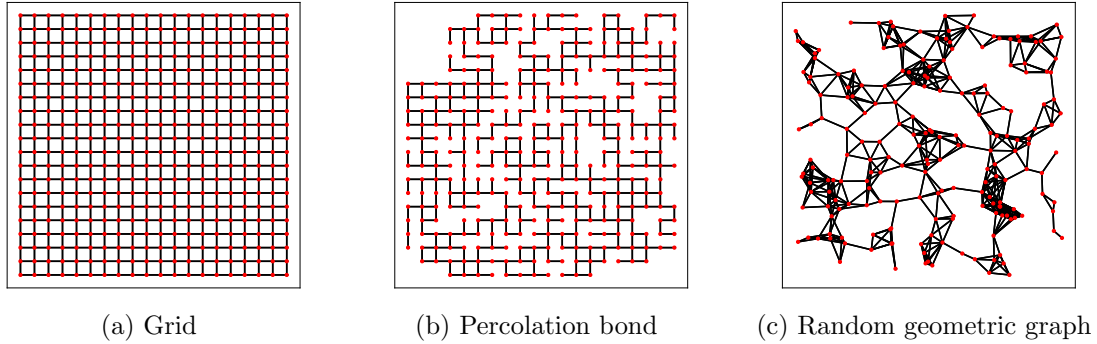


Figure 1. The three types of 2D graphs considered in simulations.

Chebyshev polynomials (see Proposition SM8.4). The method (1.3) will be derived as the polynomial iteration corresponding to some Jacobi polynomials.

In this paper, we design polynomial gossip methods whose polynomials P_t , $t \geq 0$, satisfy a second-order recursion. This key property ensures that the resulting iterates $x^t = P_t(W)\xi$ can be computed recursively.

3. Simulations: Comparison of simple gossip, shift-register gossip, and the Jacobi polynomial iteration. In this section, we run our methods on grids, percolation bonds, and random geometric graphs; the latter is a widely used model for real-world networks [23, section 1.1]. In each case, we consider both the two-dimensional (2D) structure and its three-dimensional (3D) counterpart. We refer the reader to Figure 1 for visualizations of the 2D structures and to section SM1 for details about the parameters used.

We compare our Jacobi polynomial iteration (1.3) with the simple gossip method (1.1) and the shift-register algorithm (1.2). We found experimentally that the behavior of the shift-register algorithm was typical of methods based on the spectral gap such as the splitting algorithm of [25] and the Chebyshev polynomial acceleration scheme [2, 27]; to avoid redundancy we do not present the similar behavior of these methods. We also compare with local averaging, which is given by the formula

$$x_v^t = \frac{1}{|B_t(v)|} \sum_{w \in B_t(v)} \xi_w,$$

where $|B_t(v)|$ denotes the ball in G , centered in v , of radius t , for the shortest path distance. Note that local averaging does not correspond in general to any computationally cheap iteration, as opposed to the algorithms we present here. Thus it should not be considered as a gossip method, but rather as a lower bound on the performance achievable by any gossip method. (This is made fully rigorous in the statistical gossip framework of section SM7.)

In our simulations, we change the graph G that we run our algorithms on, but we always sample $\xi_v \sim_{i.i.d.} \mathcal{N}(0, 1)$, $v \in V$, and measure the performance of gossip methods through the quantity $\|x^t - \bar{\xi}\mathbf{1}\|_2 / \sqrt{n}$. Thus the performance of the algorithms is random because the initial values ξ_v are random, and also because percolation bonds and random geometric graphs are

random. The results we present here are averaged over 10 realizations of the graph and the initial values, which is sufficient to give stable results.

Tuning. The optimal tuning of the shift-register gossip method as a function of the spectral gap was determined in [16, Theorem 2]; it is given by the formula (2.4); this is the tuning that we use in our simulations. The Jacobi polynomial iteration is tuned by choosing $d = 2$ in the 2D grid, 2D percolation bonds, and 2D random geometric graphs, and $d = 3$ for their 3D analogues.

Interpretation of the results. The results of the simulations are exposed in Figure 2. The qualitative picture remains the same across different graphs. Simple gossip performs better than shift-register gossip in a first phase, but in a large t asymptotic, simple gossip converges slowly whereas shift-register gossip converges quickly. The Jacobi polynomial iteration enjoys the quick diffusion of simple gossip in the first phase and reaches the full mixing before shift-register gossip. As a consequence, the Jacobi polynomial iteration gets considerably closer to the local averaging optimal bound, especially in very regular structures like grids.

These results should be mitigated with the large t asymptotic: in Figure 3, we show the comparison of gossip methods on a longer time scale, in linear and log-scale y -axis. We only present the results on the 2D grid as they are typical of the behavior on other structures. We observe that shift-register gossip enjoys a much better asymptotic rate of convergence than simple gossip and the Jacobi polynomial iteration.

Methods that use the spectral gap are designed to achieve the best possible asymptotic (see [7], [25]); thus the above observation is not surprising. These methods, however, fail in the nonasymptotic regime, where they are outperformed by the Jacobi polynomial iteration and simple gossip. We believe that in applications where a high precision on the average is not needed, the Jacobi polynomial iteration brings important improvements over existing methods, let alone the fact that it is considerably easier to tune. However, in section 6, we present a Jacobi polynomial iteration that uses the spectral gap of the gossip matrix to obtain the accelerated convergence rate.

4. Design of best polynomial gossip iterations. We now turn to the design of efficient polynomial iterations of the form $x^t = P_t(W)\xi$. An important result of this section is that the best iterates of this form can be computed in an online fashion as they result from a second-order recurrence relation.

The approach presented in this section is similar to [9, section 3.3], although therein it is applied to the slightly different problem of load balancing. We repeat here the derivations as we take a slightly different approach: here we derive the best polynomial P_t with fixed W and ξ , while in [9] the matrix W is fixed, but a polynomial P_t efficient uniformly over ξ is sought. We then discuss why the resulting recursion may be impractical. The next section introduces some approximation of the impractical scheme that leads to the practical iteration (1.3).

Our measure of performance of a polynomial gossip iteration is the sum of squared errors over the agents of the network:

$$\mathcal{E}(P_t) = \sum_{v \in V} (x_v^t - \bar{\xi})^2 = \|x^t - \bar{\xi}\mathbf{1}\|_2^2 = \|P_t(W)\xi - \bar{\xi}\mathbf{1}\|_2^2.$$

Denote by $\lambda_1, \lambda_2, \dots, \lambda_n$ the real eigenvalues of the symmetric matrix W and by u^1, u^2, \dots, u^n

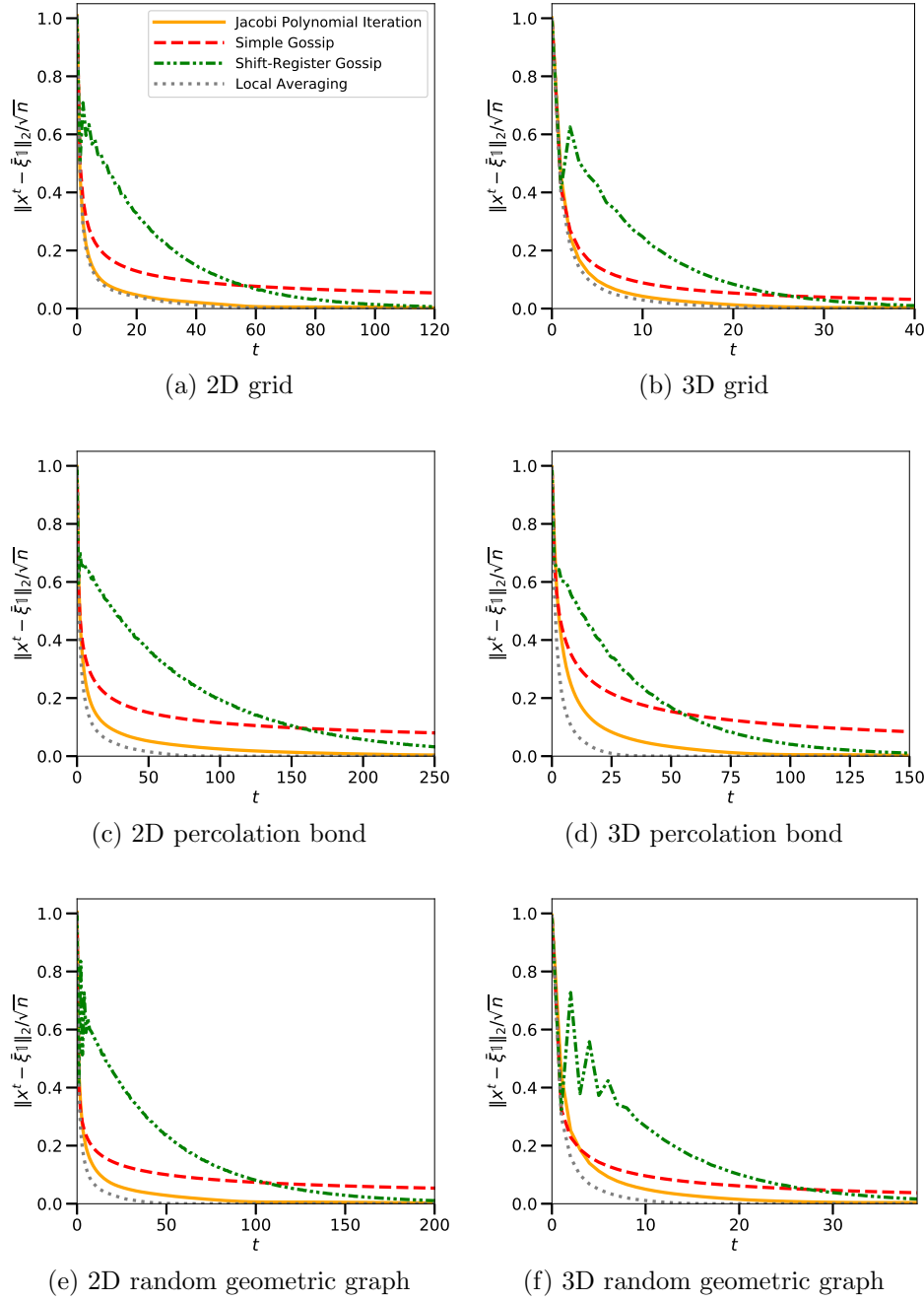


Figure 2. Performance of different gossip algorithms running on graphs with an underlying low-dimensional geometry, as measured by $\|x^t - \xi \mathbf{1}\|_2 / \sqrt{n}$.

the associated eigenvectors, normalized such that $\|u^i\|_2 = 1$. The diagonalization of W gives

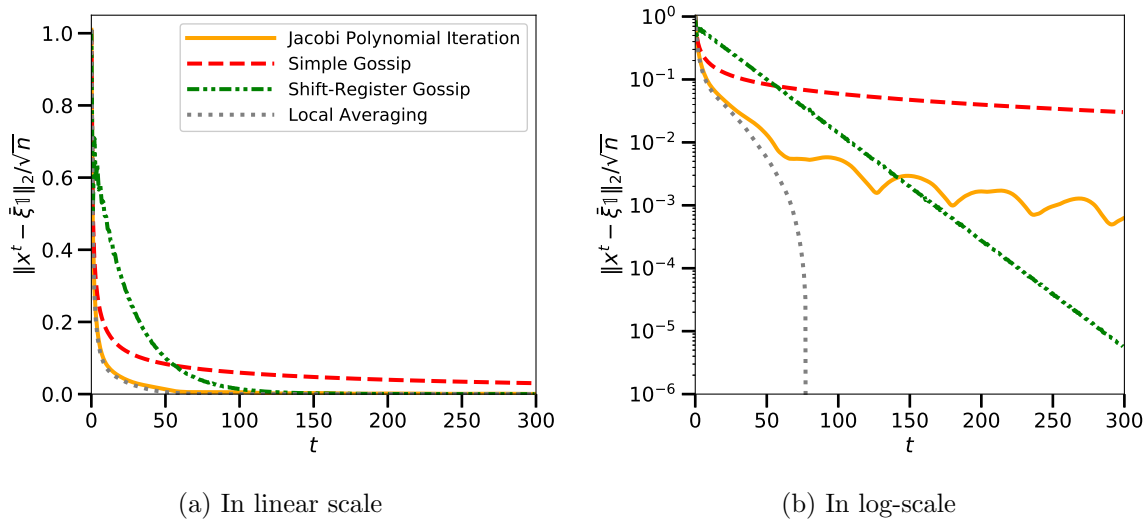


Figure 3. Performance of different gossip algorithms running on the 2D grid.

the new expression of the error

$$(4.1) \quad \mathcal{E}(P_t) = \sum_{i=2}^n \langle \xi, u^i \rangle^2 P_t(\lambda_i)^2 = \int_{-1}^1 P_t(\lambda)^2 d\sigma(\lambda), \quad d\sigma(\lambda) = \sum_{i=2}^n \langle \xi, u^i \rangle^2 \delta_{\lambda_i},$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical scalar product on \mathbb{R}^n and δ_λ is the Dirac mass at λ .

The polynomial π_t minimizing the error $\mathcal{E}(P_t)$ must be chosen as

$$(4.2) \quad \pi_t \in \underset{P(1)=1, \deg P \leq t}{\operatorname{argmin}} \int_{-1}^1 P(\lambda)^2 d\sigma(\lambda).$$

We now show that the sequence of best polynomials $\pi_0, \pi_1, \pi_2, \dots$ can be computed as the result of a second-order recursion, which leads to a second-order gossip method, whose coefficients depend on σ . As noted in [7], having iterates x^t that satisfy a low-order recurrence relation is valuable as it ensures that they can be computed online with limited memory cost. In order to prove this property for our iterates, we use that these polynomials are orthogonal with respect to (w.r.t.) some measure τ .

Definition 4.1 (orthogonal polynomials w.r.t. τ). Let τ be a measure on \mathbb{R} with finite moments. Endow the set of polynomials $\mathbb{R}[X]$ with the scalar product

$$\langle P, Q \rangle_\tau = \int_{\mathbb{R}} P(\lambda) Q(\lambda) d\tau(\lambda).$$

Denote by $T \in \mathbb{N} \cup \{\infty\}$ the cardinal of the support of τ . Then there exists a family $\pi_0, \pi_1, \dots, \pi_{T-1}$ of polynomials, such that for all $t < T$, $\pi_0, \pi_1, \dots, \pi_t$ form an orthogonal basis of $(\mathbb{R}_t[X], \langle \cdot, \cdot \rangle_\tau)$, where $\mathbb{R}_t[X]$ denotes the set of polynomials of degree smaller than or equal to t . In other words, for all $s, t < T$,

$$\deg \pi_t = t, \quad \langle \pi_s, \pi_t \rangle_\tau = 0 \quad \text{if } s \neq t.$$

$\pi_0, \pi_1, \dots, \pi_{T-1}$ is called a sequence of orthogonal polynomials w.r.t. τ . Moreover, the family of orthogonal polynomials $\pi_0, \pi_1, \dots, \pi_{T-1}$ is unique up to a rescaling of each of the polynomials.

An extensive reference on orthogonal polynomials is the book [29]. An introduction from the point of view of applied mathematics can be found in [13]. In section SM2, we recall the results from the theory of orthogonal polynomials that we use in this paper. The next proposition states that the optimal polynomials sought in (4.2) are orthogonal polynomials.

Proposition 4.2. *Let σ be some finite measure on $[-1, 1]$, and let $T \in \mathbb{N} \cup \{\infty\}$ be the cardinal of $\text{Supp } \sigma - \{1\}$. For $0 \leq t \leq T - 1$, the minimizer π_t of*

$$\min_{P(1)=1, \deg P \leq t} \int_{-1}^1 P(\lambda)^2 d\sigma(\lambda)$$

is unique. Moreover, π_0, \dots, π_{T-1} is the unique sequence of orthogonal polynomials w.r.t. $d\tau(\lambda) = (1 - \lambda)d\sigma(\lambda)$ normalized such that $\pi_t(1) = 1$.

This result is well known and usually stated without proof [21, sections 3, 4.1], [22, section 2]; we give the short proof in section SM4. In the following, the phrase “the orthogonal polynomials w.r.t. τ ” will refer to the unique family of orthogonal polynomials w.r.t. τ and normalized such that $\pi_t(1) = 1$.

Remark 4.3. When T is finite and $t \geq T$, finding a minimizer of $\int_{-1}^1 P(\lambda)^2 d\sigma(\lambda)$ over the set of polynomials such that $P(1) = 1$, $\deg P \leq t$ is trivial. Indeed, one can consider the polynomial

$$\pi_T(\lambda) = \frac{\prod_{\lambda' \in \text{Supp } \sigma - \{1\}} (\lambda - \lambda')}{\prod_{\lambda' \in \text{Supp } \sigma - \{1\}} (1 - \lambda')}$$

which is of degree T , satisfies $\pi_T(1) = 1$, and $\int_{-1}^1 \pi_T(\lambda)^2 d\sigma(\lambda) = \sigma(\{1\})$. This is the best value that a polynomial P of any degree, such that $P(1) = 1$, can get.

A fundamental result on orthogonal polynomials states that they follow a second-order recursion.

Proposition 4.4 (three-term recurrence relation, from [29, Theorem 3.2.1]). *Let π_0, \dots, π_{T-1} be a sequence of orthogonal polynomials w.r.t. some measure τ . There exist three sequences of coefficients $(a_t)_{1 \leq t \leq T-2}$, $(b_t)_{1 \leq t \leq T-2}$, and $(c_t)_{1 \leq t \leq T-2}$ such that, for $1 \leq t \leq T - 2$,*

$$\pi_{t+1}(\lambda) = (a_t \lambda + b_t) \pi_t(\lambda) - c_t \pi_{t-1}(\lambda).$$

The classical proof of this proposition is given in section SM2.1. Taking σ to be the spectral measure of (4.1) in Proposition 4.2, we get that the best polynomial gossip algorithm is a second-order method whose coefficients are determined by the graph G , the gossip matrix W , and the vertex v . Indeed, as π_0, \dots, π_{T-1} is a family of orthogonal polynomials, there exist coefficients a_t, b_t, c_t such that

$$\pi_{t+1}(\lambda) = (a_t \lambda + b_t) \pi_t(\lambda) - c_t \pi_{t-1}(\lambda),$$

and thus

$$\pi_{t+1}(W) = a_t W \pi_t(W) + b_t \pi_t(W) - c_t \pi_{t-1}(W).$$

Decomposing $\pi_1(\lambda) = a_0 \lambda + b_0$ and applying the previous relation in ξ gives the second-order recursion for the best polynomial estimators $x^t = \pi_t(W)\xi$:

$$(4.3) \quad x^0 = \xi, \quad x^1 = a_0 W \xi + b_0 \xi, \quad x^{t+1} = a_t W x^t + b_t x^t - c_t x^{t-1}.$$

Note that the dependence of the gossip method on the graph G , the gossip matrix W , and the vertex v is entirely hidden in the coefficients a_t, b_t, c_t . Thus the choice of the coefficients is central. In [9], it is argued that the coefficients can be computed in a “preprocessing step.” Indeed, the coefficients can be computed in a centralized or decentralized manner, at the cost of extra communication steps. The gossip method that consists in computing the optimal coefficients a_t, b_t, c_t and running (4.3) will be referred to as *parameter-free polynomial iteration*, as it does not require any tuning of parameters, by analogy with the terminology used in polynomial methods for the resolution of linear systems (see [12, section 6]). It corresponds to the optimal polynomial iteration. For a detailed exposition on the parameter-free polynomial iteration and a discussion of its practicability, see section SM5.

However, in dynamic networks that are constantly changing, it is not a valid option to keep repeating the preprocessing step to update the coefficients a_t, b_t, c_t . Our approach consists in observing that there are sequences of coefficients like (1.4) that, despite not being optimal, work reasonably well on a large set of graphs. This implies that even if the details of the graph are not known to the algorithmic designer, she can make a choice of coefficients that have a fair performance.

More formally, we approximate the true spectral measure σ of the graph with a simpler measure $\tilde{\sigma}$, whose associated polynomials have known recursion coefficients a_t, b_t, c_t . We show that in some cases, substituting the orthogonal polynomials w.r.t. σ with the ones orthogonal to $\tilde{\sigma}$ does not worsen the efficiency of the gossip method much. In the next sections, we argue for two choices of the approximating measure $\tilde{\sigma}$. The first uses only the spectral dimension d of the network and gives the Jacobi polynomial iteration (1.3). The second uses both the spectral dimension d and the spectral gap γ of W and gives the Jacobi polynomial iteration with spectral gap.

Figure 4 reproduces Figure 3 and adds the performance of the parameter-free polynomial iteration and the Jacobi polynomial iteration with spectral gap. It shows that in linear scale, the performance of the parameter-free polynomial iteration is indistinguishable from the performance of the Jacobi polynomial iterations with or without spectral gap, which are obtained through approximations of the spectral measure σ . However, the figure in log-scale shows that the asymptotic convergence of the methods depends on the coarseness of the approximation. The relevance of this asymptotic convergence to the practice depends on the application.

Remark 4.5. The shift-register iteration $x^t = P_t(W)\xi$ defined in (2.3) can be seen as a best polynomial gossip iteration with some approximating measure. Indeed, the polynomials P_t , $t \geq 0$, are the orthogonal polynomials w.r.t. some measure whose support is strictly included in $[-1, 1]$ (see Proposition SM8.5).

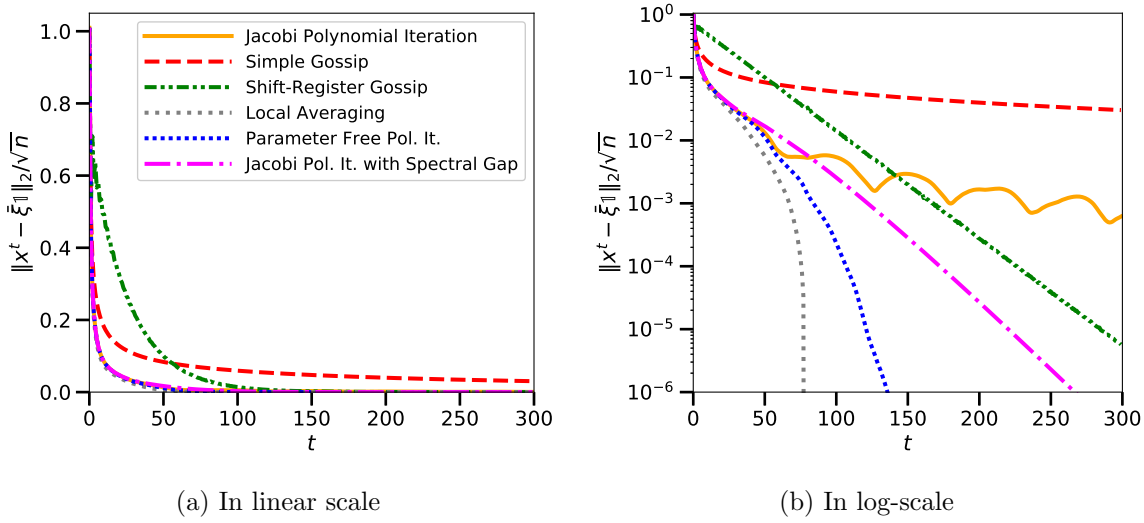


Figure 4. Performance of different gossip algorithms running on the 2D grid.

5. Design of polynomial gossip algorithms for graphs of given spectral dimension.

5.1. The dimension d and the rate of decrease of the spectral measure near 1. We now assume that we are given a graph G on which we would like to run the optimal polynomial gossip algorithm (4.3). However, we do not know the spectral measure σ or the coefficients a_t, b_t, c_t . In this section, we give a heuristic motivating an approximation $\tilde{\sigma}$ of the spectral measure σ using only the dimension d of the graph. The heuristic is supported by the simulations of section 3 and some rigorous theoretical support in section SM7.

Our approximation is given by the following nonrigorous intuition:

$$(5.1) \quad \text{the graph } G \text{ is of dimension } d \Leftrightarrow \sigma([1 - \Lambda, 1]) \approx C\Lambda^{d/2} \quad \text{as } \Lambda \ll 1,$$

for some constant C . Of course, we have neither defined the dimension of a graph nor given a rigorous signification of the symbols \approx and \ll . We come back to these questions in section 5.3, but for now we assume that the reader has an intuitive understanding of these notions and finish drawing the heuristic picture.

Intuition (5.1) describes the repartition of the mass of σ near 1. This mass near 1 challenges the design of polynomial methods as the gossip polynomials P are constrained to satisfy $P(1) = 1$ while minimizing $\int P^2 d\sigma$. Moreover, eigenvalues of a graph close to 1 are known to describe the large-scale structure of the graph and thus must be central in the design of gossip methods. The traditional design of gossip algorithms considered the spectral gap γ between 1 and the second largest eigenvalue, a quantity that typically gets very small in large graphs. Intuition (5.1) also describes the behavior of the spectrum near 1, but on a larger scale than the spectral gap. It describes how the set of the largest eigenvalues is distributed around 1.

5.2. The Jacobi iteration for graphs of given dimension. When a spectral measure satisfies the edge estimate (5.1), we approximate it with a measure satisfying the same estimate,

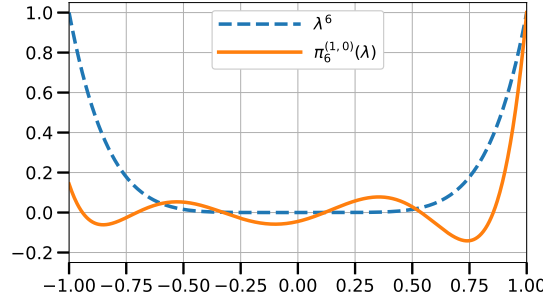


Figure 5. Comparison of the Jacobi polynomial $\pi_6^{(1,0)}(\lambda)$ with the polynomial of simple gossip λ^6 .

namely

$$d\tilde{\sigma}(\lambda) = (1 - \lambda)^{d/2-1} \mathbf{1}_{\{\lambda \in (-1,1)\}} d\lambda.$$

Note that we do not elaborate on the normalization of the approximate measure $d\tilde{\sigma}$ as it is only used to define an orthogonality relation between polynomials, in which the normalization does not matter. The orthogonal polynomials w.r.t. the modified spectral measure $(1 - \lambda)d\tilde{\sigma}(\lambda) = (1 - \lambda)^{d/2} \mathbf{1}_{\{\lambda \in (-1,1)\}} d\lambda$ and their recursion coefficients are known as they correspond to the well-studied Jacobi polynomials [29, Chapter IV]:

$$(5.2) \quad \begin{aligned} a_0^{(d)} &= \frac{d+4}{2(2+d)}, & b_0^{(d)} &= \frac{d}{2(2+d)}, \\ a_t^{(d)} &= \frac{(2t+d/2+1)(2t+d/2+2)}{2(t+1+d/2)^2}, & b_t^{(d)} &= \frac{d^2(2t+d/2+1)}{8(t+1+d/2)^2(2t+d/2)}, \\ c_t^{(d)} &= \frac{t^2(2t+d/2+2)}{(t+1+d/2)^2(2t+d/2)}. \end{aligned}$$

These coefficients are derived in section SM6.2. This approximation of the spectral measure gives the practical recursion

$$(5.3) \quad x^0 = \xi, \quad x^1 = a_0^{(d)} W \xi + b_0^{(d)} \xi, \quad x^{t+1} = a_t^{(d)} W x^t + b_t^{(d)} x^t - c_t^{(d)} x^{t-1},$$

which only depends on d . It is just a rewriting of the Jacobi polynomial iteration (1.3) given in the introduction of this paper. The Jacobi polynomial $\pi_t^{(d/2,0)}(\lambda)$ such that $x^t = \pi_t^{(d/2,0)}(W)\xi$ is plotted in Figure 5 with $d = 2$ and $t = 6$, along with the polynomial λ^6 associated with simple gossip. The Jacobi polynomial is smaller in magnitude near 1.

5.3. Spectral dimension of a graph. In this section, we discuss the meaning of intuition (5.1). There are several definitions of the dimension of a graph.

When referring to the dimension of a graph, many authors actually refer to some quantity d that has been used in the construction of the graph. An example is the d -dimension grid $\{1, \dots, n\}^d$. Another example consists in removing edges in \mathbb{Z}^d with probability $1 - p$, independently of one another. The resulting graph G is called a percolation bond [14]. It is natural to consider that this graph is of dimension d . A more complicated example is the

random geometric graph: choose $d \geq 1$, sample n points uniformly in the d -dimensional cube $[0, 1]^d$, and connect with an edge all pairs of points closer than some chosen distance $r > 0$. It is natural to say that this random geometric graph is d -dimensional as it is the dimension of the surface it is built on.

Mathematicians have developed more intrinsic definitions of the dimension of a graph [11]; here we use the notion of *spectral dimension*. This definition is of interest only for infinite graphs $G = (V, E)$. Here, we consider only locally finite graphs, meaning that each node has only a finite number of neighbors. As with Definition 2.1, one can define a gossip matrix W with entries indexed by $V \times V$. If G is infinite, W is a doubly infinite array, but with only a finite number of nonzero elements in each line and column as the graph is locally finite.

The spectral dimension of a graph G is defined using a random walk on the graph—typically the simple random walk on G —but here we consider the lazy random walk with transition matrix $\tilde{W} = (I + W)/2$. (We take the *lazy* random walk to avoid periodicity issues.)

Definition 5.1 (spectral dimension). Denote by p_t the probability that the lazy random walk, when started from v , returns at v at time t . The spectral dimension of the graph, if it exists and is finite, is the limit

$$d_s = d_s(G, W, v) = -2 \lim_{t \rightarrow \infty} \frac{\ln p_t}{\ln t}.$$

If the graph is connected and W is the transition matrix of the simple random walk, this definition does not depend on the choice of the vertex v . Motivations for this definition are as follows.

Proposition 5.2. The spectral dimension of (\mathbb{Z}^d, W) with $W = A(\mathbb{Z}^d)/d$ is d .

Proof. The return probability p_t of the lazy random walk on \mathbb{Z}^d is equivalent to $C/t^{d/2}$ for some constant C . It is, for instance, a consequence of the local central limit theorem for random walks on \mathbb{Z}^d [15, Theorem 2.1.1]. Thus the spectral dimension of \mathbb{Z}^d is d . ■

The spectral dimension of a graph is related to the decay of the spectrum of W near 1.

Definition 5.3 (spectral measure of a possibly infinite graph). Let G be a graph and W its gossip matrix. Fix $v \in V$. As W is an auto-adjoint operator, bounded by 1, acting on $\ell^2(V)$, there exists a unique positive measure $\sigma = \sigma(G, W, v)$ on $[-1, 1]$, called the spectral measure, such that for all polynomial P ,

$$\langle e_v, P(W)e_v \rangle_{\ell^2(V)} = \int_{-1}^1 P(\lambda) d\sigma(\lambda).$$

For a deeper presentation of spectral graph theory, see [19] and references therein. Note that when the graph G is finite, it is easy to check that the spectral measure is the discrete measure $\sigma(G, W, v) = \sum_{i=1}^n (u_v^i)^2 \delta_{\lambda_i}$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of W and u^1, \dots, u^n are the associated normalized eigenvectors. However, when the graph G is infinite, the spectrum may exhibit a continuous part w.r.t. the Lebesgue measure.

Proposition 5.4 (the spectral dimension is the spectral decay). Let G be a graph, W a gossip matrix on G , and v a vertex. We denote by $d_s = d_s(G, W, v)$ the spectral dimension and by

$\sigma = \sigma(G, W, v)$ the spectral measure. Then the limit $\lim_{\Lambda \rightarrow 0} \ln \sigma([1 - \Lambda, 1]) / \ln \Lambda$ exists and is finite if and only if d_s exists and is finite. In that case,

$$\lim_{\Lambda \rightarrow 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} = \frac{d_s}{2}.$$

This proposition gives a rigorous equivalent to intuition (5.1). It uses the spectral dimension of the graph, which is an intrinsic property of the graph and turns out to coincide with our intuition of the dimension of a graph in examples of interest. Note that in section 4, the spectral measure σ is defined as $d\sigma(\lambda) = \sum \langle \xi, u^i \rangle^2 \delta_{\lambda_i}$, whereas in this section it is defined for finite graphs as $d\sigma(\lambda) = \sum (u_v^i)^2 \delta_{\lambda_i}$. Roughly speaking, intuition (5.1) is valid for the former if ξ projects evenly on all eigenvectors u^i . This is the case if ξ has random i.i.d. components, for instance; this is used in section SM7.

Proof of Proposition 5.4. We first assume that $l = \lim_{\Lambda \rightarrow 0} \ln \sigma([1 - \Lambda, 1]) / \ln \Lambda$ exists and is finite. We show that d_s exists and that $l = d_s/2$. To this end, we define

$$\underline{d}_s = -2 \limsup_{t \rightarrow \infty} \frac{\ln p_t}{\ln t}, \quad \bar{d}_s = -2 \liminf_{t \rightarrow \infty} \frac{\ln p_t}{\ln t},$$

where p_t is defined as in Definition 5.1. Note that

$$(5.4) \quad p_t = \left\langle e_v, \left(\frac{I + W}{2} \right)^t e_v \right\rangle \stackrel{(\text{Definition 5.3})}{=} \int \left(\frac{1 + \lambda}{2} \right)^t d\sigma(\lambda).$$

Proof that $\bar{d}_s/2 \leq l$. Consider $l_+ > l$. Then there exists constants $c_1, c_2 > 0$ such that for all $\Lambda \in [0, 2]$,

$$\sigma([1 - \Lambda, 1]) \geq c_1 \Lambda^{l_+} = c_2 \sigma^{(l_+ - 1, 0)}([1 - \Lambda, 1]),$$

where $\sigma^{(l_+ - 1, 0)}(d\lambda) = (1 - \lambda)^{l_+ - 1} d\lambda$. Then

$$\begin{aligned} p_t &\stackrel{(5.4)}{=} \int_{[-1, 1]} \left(\frac{1 + \lambda}{2} \right)^t d\sigma(\lambda) \stackrel{(\text{Lemma SM3.1})}{\geq} c_2 \int_{-1}^1 \left(\frac{1 + \lambda}{2} \right)^{t-1} (1 - \lambda)^{l_+ - 1} d\lambda \\ &\stackrel{(u = (1 + \lambda)/2)}{=} c_3 \int_0^1 u^t (1 - u)^{l_+ - 1} du = c_3 B(t + 1, l_+) \stackrel{(\text{SM3.2})}{\underset{t \rightarrow \infty}{\sim}} \frac{c_4}{t^{l_+}} \end{aligned}$$

for some constant $c_3, c_4 > 0$. Thus $\liminf_{t \rightarrow \infty} \frac{\ln p_t}{\ln t} \geq -l_+$, i.e., $\bar{d}_s/2 \leq l_+$. This being true for all $l_+ > l$, this proves $\bar{d}_s/2 \leq l$.

Proof that $\underline{d}_s/2 \geq l$. Consider $l_- < l$. Then there exist constants C_1, C_2 such that for all $\Lambda \in [0, 2]$,

$$\sigma([1 - \Lambda, 1]) \leq C_1 \Lambda^{l_-} = C_2 \sigma^{(l_- - 1, 0)}([1 - \Lambda, 1]),$$

where $\sigma^{(l_- - 1, 0)}(d\lambda) = (1 - \lambda)^{l_- - 1} d\lambda$. Then

$$\begin{aligned} p_t &\stackrel{(5.4)}{=} \int_{[-1, 1]} \left(\frac{1 + \lambda}{2} \right)^t d\sigma(\lambda) \stackrel{(\text{Lemma SM3.1})}{\leq} C_2 \int_{-1}^1 \left(\frac{1 + \lambda}{2} \right)^t (1 - \lambda)^{l_- - 1} d\lambda \\ &\stackrel{(u = (1 + \lambda)/2)}{=} C_3 \int_0^1 u^t (1 - u)^{l_- - 1} du = C_3 B(t + 1, l_-) \stackrel{(\text{SM3.2})}{\underset{t \rightarrow \infty}{\sim}} \frac{C_4}{t^{l_-}} \end{aligned}$$

for some constants C_3, C_4 . Thus $\limsup_{t \rightarrow \infty} \frac{\ln p_t}{\ln t} \leq -l_-$, which implies $\underline{d}_s/2 \geq l_-$. This being true for all $l_- < l$, this proves $\underline{d}_s/2 \geq l$.

Finally, we have proven $l \leq \underline{d}_s/2 \leq \bar{d}_s/2 \leq l$. Thus the limit $d_s = -2 \lim_{t \rightarrow \infty} \ln p_t / \ln t$ exists and is equal to $2l$.

Conversely, we assume now that d_s exists and is finite. We show that $l = \lim_{\Lambda \rightarrow 0} \ln \sigma([1 - \Lambda, 1]) / \ln \Lambda$ exists and that $l = d_s/2$. To this end, we define

$$\underline{l} = \liminf_{\Lambda \rightarrow 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda}, \quad \bar{l} = \limsup_{\Lambda \rightarrow 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda}.$$

Proof that $\underline{l} \geq d_s/2$. For any $t \in \mathbb{N}$, we have

$$\mathbf{1}_{\{\lambda \geq 1 - \Lambda\}} \leq \left(1 - \frac{\Lambda}{2}\right)^{-t} \left(\frac{1 + \lambda}{2}\right)^t;$$

thus, by integrating against $d\sigma(\lambda)$,

$$\begin{aligned} \sigma([1 - \Lambda, 1]) &\leq \left(1 - \frac{\Lambda}{2}\right)^{-t} \int \left(\frac{1 + \lambda}{2}\right)^t \sigma(d\lambda), \\ \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} &\geq \frac{\ln \int \left(\frac{1 + \lambda}{2}\right)^t \sigma(d\lambda)}{\ln t} \frac{\ln t}{\ln \Lambda} - \frac{t \ln \left(1 - \frac{\Lambda}{2}\right)}{\ln \Lambda}. \end{aligned}$$

We choose $t(\Lambda) = \lfloor \Lambda^{-1} \rfloor$. Then we get

$$\underline{l} = \liminf_{\Lambda \rightarrow 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} \geq -\frac{d_s}{2}(-1) - 0 = \frac{d_s}{2}.$$

Proof that $\bar{l} \leq d_s/2$. For any $t \in \mathbb{N}$, we have $((1 + \lambda)/2)^t - (1 - \Lambda/2)^t \leq \mathbf{1}_{\{\lambda \geq 1 - \Lambda\}}$; thus, by integrating against $d\sigma(\lambda)$,

$$\int \left(\frac{1 + \lambda}{2}\right)^t d\sigma(\lambda) - \left(1 - \frac{\Lambda}{2}\right)^t \leq \sigma([1 - \Lambda, 1]).$$

Let $d > d_s$. There exists a constant $c > 0$ such that $\int ((1 + \lambda)/2)^t d\sigma(\lambda) \geq c/t^{d/2}$. Then

$$\ln \left(\frac{c}{t^{d/2}} - \left(1 - \frac{\Lambda}{2}\right)^t \right) \leq \ln \sigma([1 - \Lambda, 1]).$$

Let $\alpha > 1$. We choose $t(\Lambda) = \lceil \Lambda^{-\alpha} \rceil$. Then

$$\left(1 - \frac{\Lambda}{2}\right)^{t(\Lambda)} = \exp \left(t(\Lambda) \ln \left(1 - \frac{\Lambda}{2}\right) \right) \leq \exp \left(-\frac{t(\Lambda)\Lambda}{2} \right) \leq \exp \left(-\frac{1}{2} \Lambda^{1-\alpha} \right)$$

decreases superpolynomially fast as $\Lambda \rightarrow 0$. Since $ct(\Lambda)^{-d/2} \underset{\Lambda \rightarrow 0}{\sim} c\Lambda^{\alpha d/2}$, it yields

$$\bar{l} = \limsup_{\Lambda \rightarrow 0} \frac{\ln \sigma([1 - \Lambda, 1])}{\ln \Lambda} \leq \frac{\alpha d}{2}.$$

As this is true for all $\alpha > 1, d > d_s$, we have $\bar{l} \leq d_s/2$.

Finally, we have proven that $d_s/2 \leq \underline{l} \leq \bar{l} \leq d_s/2$. Then the limit $l = \lim_{\Lambda \rightarrow 0} \ln \sigma([1 - \Lambda, 1]) / \ln \Lambda$ exists and $l = d_s/2$. ■

Proposition 5.4 allows us to prove the following generalization of Proposition 5.2.

Proposition 5.5 (spectral dimension of the supercritical percolation cluster). *Let G_0 be a supercritical percolation bond in \mathbb{Z}^d with edge probability $p \in (p_c, 1]$; i.e., a.s., there is an infinite connected component G in G_0 . Endow G with the gossip matrix $W = I + (A - D)/(2d)$, where A and D are, respectively, the adjacency and the degree matrices of G . Fix $v \in \mathbb{Z}^d$. Then a.s. on the event $\{v \in G\}$, $d_s(G, W, v) = d$.*

This proposition suggests that the spectral dimension is unrelated to the small-scale structure of the graph.

Proof of Proposition 5.5. The return probabilities of the random walk on the supercritical percolation cluster have rather been studied in continuous time. The continuous-time random walk is defined as follows: the random walk at w waits at an exponential time of parameter 1 before picking a site w' out of the $2d$ neighboring sites uniformly randomly. If there is an edge in the percolation configuration between w and w' , the random walk jumps to w' ; otherwise it stays in w and starts again. Denote by X_t the continuous-time random walk and by \mathbb{P}_w the probability w.r.t. this random walk when it is started from some vertex w .

Lemma 5.6. *There exist two constants $c = c(d, p)$, $C = c(d, p) > 0$ such that, a.s. on the set $\{v \in G\}$, there exists a random time t_0 such that for $t \geq t_0$,*

$$\frac{c}{t^{d/2}} \leq \mathbb{P}_v(X_t = v) \leq \frac{C}{t^{d/2}}.$$

Proof. The upper bound is proved in [17, Theorem 1.2]. As noted in [5, Lemma 5.1], the lower bound can be proved using a central limit theorem on X_t ; we repeat the argument here as our random walk differs slightly from theirs. As X_t is reversible w.r.t. the uniform measure on G ,

$$\mathbb{P}_v(X_{2t} = v) = \sum_{w \in G} \mathbb{P}_v(X_t = w) \mathbb{P}_w(X_t = v) = \sum_{w \in G} \mathbb{P}_v(X_t = w)^2.$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned} \mathbb{P}_v(\|X_t - v\|_2 \leq \sqrt{t})^2 &= \left(\sum_{x \in G} \mathbf{1}_{\{\|x-v\|_2 \leq \sqrt{t}\}} \mathbb{P}_v(X_t = x) \right)^2 \\ &\leq \left| \{x \in G : \|x - v\|_2 \leq \sqrt{t}\} \right| \left(\sum_{w \in G} \mathbb{P}_v(X_t = w)^2 \right) \\ &\leq C_1 t^{d/2} \mathbb{P}_v(X_{2t} = v) \end{aligned}$$

for some constant C_1 . Now, using [1, Theorem 1.1(a)], there exists a deterministic variance σ^2 such that the law of $(X_t - v)/\sqrt{t}$ converges a.s. on the event $\{v \in G\}$ to a centered Gaussian with variance σ^2 . Thus there exist a deterministic constant $c_1 > 0$ and a random time t_1 such that for $t \geq t_1$, $\mathbb{P}_t(\|X_t - v\|_2 \leq \sqrt{t})^2 \geq c_1$. This finishes the proof of the lower bound. ■

We now finish the proof of the proposition using Lemma 5.6. If μ^t denotes the law of X_t , we have $\frac{d}{dt} \mathbb{E}[\mu^t] = (W - I)\mu^t$. This yields $\mu^t = e^{t(W-I)}\mu^0$, which implies

$$\mathbb{P}_v(X_t = v) = \langle \delta_v, \mu^t \rangle = \langle \delta_v, e^{t(W-I)} \delta_v \rangle \stackrel{\text{(Definition 5.3)}}{=} \int e^{t(\lambda-1)} d\sigma(\lambda).$$

As a consequence, Lemma 5.6 translates into bounds on the Laplace transform of σ : a.s. on $\{v \in G\}$, for t large enough,

$$\frac{c}{t^{d/2}} \leq \int e^{t(\lambda-1)} d\sigma(\lambda) \leq \frac{C}{t^{d/2}}.$$

Some bounds on the spectral density of σ near 1 easily follow (see [20, Lemma 4.5]): there exist constants $c', C' > 0$ such that a.s. on $\{v \in G\}$, for Λ small enough,

$$c'\Lambda^{d/2} \leq \sigma([1-\Lambda, 1]) \leq C'\Lambda^{d/2}.$$

The proof is finished using Proposition 5.4. ■

In section SM7, we prove some performance guarantees of the Jacobi polynomial iteration (1.3) under the assumption that the graph has spectral dimension d . As a corollary, we get performance results on two types of infinite graphs: the d -dimensional grid \mathbb{Z}^d and supercritical percolation bonds in dimension d . This supports that the iteration (1.3) is robust to local perturbations of a graph.

6. The Jacobi polynomial iteration with spectral gap. In this section, we adapt the Jacobi polynomial iteration to the case where the spectral gap γ of the gossip matrix W is given. This allows us to obtain accelerated asymptotic rates of convergence, which compete with the state-of-the-art accelerated algorithms for gossip.

We assume that we are given the spectral dimension d of the graph, which determines the density of eigenvalues near 1, and the spectral gap $\gamma = 1 - \lambda_2(W)$, the distance between the largest and the second largest eigenvalues. Given these parameters, we can approximate the spectral measure of W with

$$d\tilde{\sigma}(\lambda) = ((1-\gamma) - \lambda)^{d/2-1} \mathbf{1}_{\{\lambda \in (-1, 1-\gamma)\}} d\lambda.$$

Following the recommendation of Proposition 4.2, this means that we should consider the polynomial iteration associated with the orthogonal polynomials w.r.t. $(1-\lambda)d\tilde{\sigma}(\lambda) = (1-\lambda)((1-\gamma) - \lambda)^{d/2-1} \mathbf{1}_{\{\lambda \in (-1, 1-\gamma)\}} d\lambda$. We do not know how to compute the recurrence formula for this measure; thus we used the orthogonal polynomials w.r.t. $((1-\gamma) - \lambda)d\tilde{\sigma}(\lambda) = ((1-\gamma) - \lambda)^{d/2} \mathbf{1}_{\{\lambda \in (-1, 1-\gamma)\}} d\lambda$, which is a rescaled version of a Jacobi measure. The corresponding polynomial method is called the *Jacobi polynomial iteration with spectral gap*.

A recursive formula for orthogonal polynomials w.r.t. $((1-\gamma) - \lambda)d\tilde{\sigma}(\lambda)$ is derived in section SM6.3. Taking $\alpha = d/2$ and $\beta = 0$ in equations (SM6.3), and using the coefficients

$a_t^{(d)}, b_t^{(d)}, c_t^{(d)}$ defined in (5.2), we get the recursion

$$\begin{aligned}
 x^t &= \frac{y^t}{\delta_t}, \\
 y^0 &= \xi, \quad \delta_0 = 1, \\
 y^1 &= a_0^{(d,\gamma)} W \xi + b_0^{(d,\gamma)} \xi, \quad \delta_1 = a_0^{(d,\gamma)} + b_0^{(d,\gamma)}, \\
 (6.1) \quad y^{t+1} &= a_t^{(d,\gamma)} W y^t + b_t^{(d,\gamma)} y^t - c_t^{(d,\gamma)} y^{t-1}, \quad t \geq 1, \\
 \delta_{t+1} &= \left(a_t^{(d,\gamma)} + b_t^{(d,\gamma)} \right) \delta_t - c_t^{(d,\gamma)} \delta_{t-1}, \quad t \geq 1, \\
 a_t^{(d,\gamma)} &= a_t^{(d)} \left(1 - \frac{\gamma}{2} \right)^{-1}, \quad b_t^{(d,\gamma)} = b_t^{(d)} + \frac{\gamma}{2} \left(1 - \frac{\gamma}{2} \right)^{-1} a_t^{(d)}, \quad t \geq 0, \\
 c_t^{(d,\gamma)} &= c_t^{(d)}, \quad t \geq 1.
 \end{aligned}$$

Theorem 6.1 (asymptotic rate of convergence). *Let $\gamma > 0$ be a lower bound on the spectral gap of the gossip matrix W , and let d be any positive real. Let $\xi = (\xi_v^t)_{v \in V}$ be any family of initial observations and $x^t = (x_v^t)_{v \in V}$ be the sequence of iterates generated by the Jacobi polynomial iteration with spectral gap (6.1). Then*

$$\limsup_{t \rightarrow \infty} \|x^t - \bar{\xi} \mathbf{1}\|_2^{1/t} \leq \frac{1 - \gamma/2}{(1 + \sqrt{\gamma/2})^2}.$$

This shows that the Jacobi polynomial iteration with spectral gap enjoys linear convergence. The asymptotic rate of convergence is equivalent to $1 - \sqrt{2\gamma}$ as $\gamma \rightarrow 0$. This justifies that we obtain an accelerated asymptotic rate of convergence that compares with the state-of-the-art accelerated gossip methods (see Figure 4).

Proof of Theorem 6.1. In this section, we use the notation of section SM6.3. As $x^t = \pi_t^{(d/2,0,\gamma)}(W)\xi$, we have

$$(6.2) \quad \|x^t - \bar{\xi} \mathbf{1}\|_2^2 = \sum_{i=2}^n \langle \xi, u^i \rangle^2 \pi_t^{(d/2,0,\gamma)}(\lambda_i)^2 \leq \|\xi - \bar{\xi} \mathbf{1}\|_2^2 \left(\sup_{\lambda \in [-1, 1-\gamma]} |\pi_t^{(d/2,0,\gamma)}(\lambda)| \right)^2,$$

where $\lambda_2, \dots, \lambda_n$ are the eigenvalues of W different from 1 that lie in $[-1, 1-\gamma]$ by definition of γ , and u^2, \dots, u^n are the corresponding normalized eigenvectors.

$$\begin{aligned}
 \sup_{\lambda \in [-1, 1-\gamma]} |\pi_t^{(d/2,0,\gamma)}(\lambda)| &\leq \frac{1}{|P_t^{(d/2,0,\gamma)}(1)|} \sup_{\lambda \in [-1, 1-\gamma]} |P_t^{(d/2,0,\gamma)}(\lambda)| \\
 &= \frac{1}{|\pi_t^{(d/2,0)}(\varphi_\gamma^{-1}(1))|} \sup_{\lambda \in \varphi_\gamma^{-1}([-1, 1-\gamma])} |\pi_t^{(d/2,0)}(\lambda)| \\
 &= \frac{1}{\left| \pi_t^{(d/2,0)} \left(\frac{1+\gamma/2}{1-\gamma/2} \right) \right|} \sup_{\lambda \in [-1, 1]} |\pi_t^{(d/2,0)}(\lambda)| \\
 (6.3) \quad &= \frac{1}{\left| P_t^{(d/2,0)} \left(\frac{1+\gamma/2}{1-\gamma/2} \right) \right|} \sup_{\lambda \in [-1, 1]} |P_t^{(d/2,0)}(\lambda)|,
 \end{aligned}$$

where $P_t^{(\alpha,\beta)}$ is the Jacobi polynomial; see section SM6.2. By Proposition SM2.6,

$$(6.4) \quad \sup_{\lambda \in [-1,1]} |P_t^{(d/2,0)}(\lambda)| = \binom{t+d/2}{t} \underset{t \rightarrow \infty}{\sim} t^{d/2},$$

and by Proposition SM2.9 applied in $x = \frac{1+\gamma/2}{1-\gamma/2}$, there exists $c > 0$ such that

$$(6.5) \quad P_t^{(d/2,0)}\left(\frac{1+\gamma/2}{1-\gamma/2}\right) \underset{t \rightarrow \infty}{\sim} ct^{-1/2} \left(\frac{(1+\sqrt{\gamma/2})^2}{1-\gamma/2}\right)^t.$$

Combining (6.3), (6.4), and (6.5), we get that there exists a constant C such that

$$\sup_{\lambda \in [-1,1-\gamma]} |\pi_t^{(d/2,0,\gamma)}(\lambda)| \leq Ct^{(d+1)/2} \left(\frac{1-\gamma/2}{(1+\sqrt{\gamma/2})^2}\right)^t,$$

and we conclude using (6.2). ■

Note that the asymptotic rate of convergence does not depend on d . However, the choice of d may have an important effect during the nonasymptotic phase $t < 1/\sqrt{\gamma}$. In this phase, the spectral gap γ can be neglected in the approximation of the spectral measure, and it is important that the densities of eigenvalues of σ and $\tilde{\sigma}$ match near the upper edge of the spectrum. This is why one should choose d as the spectral dimension of the graph.

In sections 5 and 6, we have used the polynomial point of view to build gossip algorithms suited to our priors on the graph structure (spectral dimension and spectral gap). In the supplementary material (section SM9), we reverse-engineer the message passing gossip iteration of [18] through the polynomial point of view. We show that this algorithm can be interpreted as an inner-product free polynomial iteration corresponding to a tree prior. This point of view allows us to derive convergence rates of the message passing gossip on regular graphs. This suggests that the polynomial point of view can be used more generally to analyze existing gossip algorithms.

7. The parallel between the gossip methods and distributed Laplacian solvers. There is a natural parallel between gossip methods and iterative methods that solve linear systems. Loosely speaking, simple gossip corresponds to gradient descent on the quadratic minimization problem associated to the linear system, shift-register gossip to Polyak's heavy-ball method, and the parameter-free polynomial iteration to the conjugate gradient algorithm (see [12] or [24] for references on these subjects). In this parallel, the fact that we can reach perfect gossip in n steps (see Remark 4.3) translates into the finite convergence of the conjugate gradient algorithm in a number of iterations equal to the dimension of the ambient space. In the distributed resolution of linear systems, the problem that the recursion coefficients a_t, b_t, c_t cannot be computed in a centralized manner has also appeared, and it has motivated the development of inner-product free iterations.

The Jacobi polynomial iterations presented above were motivated by the facts that (a) the parameter-free polynomial iteration is not feasible in the distributed setting of gossip, and (b) the gossip matrix W exhibits a structure due to the low-dimensional manifold on which the agents live. Interestingly, the literature on multiagent systems deals with some minimization

problems with the same properties. Examples are given by the estimation of quantities on graphs from relative measurements, in which the agents $v \in V$ try to estimate some quantity $x_v, v \in V$, defined over the graph, from noisy relative measurements over the edges of the graph:

$$\xi_{v,w} = x_v - x_w + \eta_{v,w}, \quad \{v, w\} \in E.$$

This problem has applications in network localization, where the x_v are the positions of the agents and the $\xi_{v,w}$ come from measurements of the distances and directions between the neighbors. It also has similar applications in time synchronization of clocks over networks, where x_v is the offset of the clock of node v , and to motion consensus, where x_v is the speed of agent v . For an introduction to estimation on graphs from relative measurements and its applications, see [3] and references therein. Note that the quantities x_v can only be determined up to a global constant from the measurements; either we seek the true solution up to a constant only, or we assume that some agents know their true value.

A natural approach to solving the problem is to determine estimates y_v of x_v that minimize

$$\frac{1}{2} \sum_{v,w} W_{v,w} (\xi_{v,w} - (y_v - y_w))^2,$$

where $W_{v,w}$ are some weights on the edges of the graph. Indeed, this corresponds to finding the maximum likelihood estimator if the noise $\eta_{v,w}$ is i.i.d. Gaussian and $W_{v,w}$ is the inverse variance of $\eta_{v,w}$. The above minimization problem is a quadratic problem whose covariance matrix is the Laplacian $I - W$. It can be solved using gradient descent or spectral gap based accelerations like the heavy-ball method. However, the conjugate gradient algorithm cannot be applied here as it involves centralized computations. The Jacobi polynomial iterations developed in this paper can be adapted to this situation to develop accelerations exploiting the structure of the Laplacian $I - W$. Experimenting with how this performs in real-world situations is left for future work.

8. Conclusion. Gossip methods based on the spectral gap were designed to improve the slow convergence rate of simple gossip. However, these methods are paradoxically bad at averaging locally in the intermediate regime before consensus is reached. In this paper, we propose another acceleration of simple gossip based on (i) the polynomial-based point of view, which designs iterations that are efficient at all times, and (ii) the Jacobi approximation, which uses prior information on the spectral dimension of the graph, a more natural property than the spectral gap.

It would be interesting for future work to better understand the Jacobi polynomial iteration in the asynchronous setting, i.e., when a randomized gossip matrix is used, as this setting is closer to practical cases.

In general, this paper advocates for the use of the polynomial point of view to design a gossip algorithm, as it allows us to use different types of prior information about the graph (spectral gap, spectral dimension, tree-like structure, etc.) and gives tools to prove the convergence of the designed algorithms.

Acknowledgment. The authors thank an anonymous referee for improving the clarity of this work on the questions of robustness and practicability of this work.

REFERENCES

- [1] S. ANDRES, M. T. BARLOW, J.-D. DEUSCHEL, AND B. M. HAMBLY, *Invariance principle for the random conductance model*, Probab. Theory Related Fields, 156 (2013), pp. 535–580.
- [2] M. ARIOLI AND J. SCOTT, *Chebyshev acceleration of iterative refinement*, Numer. Algorithms, 66 (2014), pp. 591–608.
- [3] P. BAROOAH AND J. P. HESPAHNA, *Estimation from relative measurements: Electrical analogy and large graphs*, IEEE Trans. Signal Process., 56 (2008), pp. 2181–2193.
- [4] R. BERTHIER, *Code for the Paper: Accelerated Gossip in Network of Given Dimension*, <https://github.com/raphael-berthier/jacobi-polynomial-iterations>, 2019.
- [5] M. BISKUP, *Recent progress on the random conductance model*, Probab. Surv., 8 (2011), pp. 294–373.
- [6] S. BOYD, A. GHOSH, B. PRABHAKAR, AND D. SHAH, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.
- [7] M. CAO, D. A. SPIELMAN, AND E. M. YEH, *Accelerated gossip algorithms for distributed computation*, in Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computation, 2006, pp. 952–959.
- [8] P. DIACONIS AND L. SALOFF-COSTE, *Moderate growth and random walk on finite groups*, Geom. Funct. Anal., 4 (1994), pp. 1–36.
- [9] R. DIEKMANN, A. FROMMER, AND B. MONIEN, *Efficient schemes for nearest neighbor load balancing*, Parallel Comput., 25 (1999), pp. 789–812.
- [10] A. D. DIMAKIS, A. D. SARWATE, AND M. J. WAINWRIGHT, *Geographic gossip: Efficient averaging for sensor networks*, IEEE Trans. Signal Process., 56 (2008), pp. 1205–1216.
- [11] B. DURHUUS, *Hausdorff and spectral dimension of infinite random graphs*, Acta Phys. Polon., 40 (2009), pp. 3509–3532.
- [12] B. FISCHER, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, John Wiley & Sons, Chichester, UK, B. G. Teubner, Stuttgart, Germany, 1996.
- [13] W. GAUTSCHI, *Orthogonal Polynomials: Computation and Approximation*, Oxford University Press, New York, 2004.
- [14] G. GRIMMETT, *What is percolation?*, in Percolation, Springer, Berlin, 1999, pp. 1–31.
- [15] G. F. LAWLER AND V. LIMIC, *Random Walk: A Modern Introduction*, Cambridge Stud. Adv. Math. 123, Cambridge University Press, Cambridge, UK, 2010.
- [16] J. LIU, B. D. O. ANDERSON, M. CAO, AND A. S. MORSE, *Analysis of accelerated gossip algorithms*, Automatica J. IFAC, 49 (2013), pp. 873–883.
- [17] P. MATHIEU AND E. REMY, *Isoperimetry and heat kernel decay on percolation clusters*, Ann. Probab., 32 (2004), pp. 100–128.
- [18] C. C. MOALLEMI AND B. VAN ROY, *Consensus propagation*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2005, pp. 899–906.
- [19] B. MOHAR AND W. WOESS, *A survey on spectra of infinite graphs*, Bull. London Math. Soc., 21 (1989), pp. 209–234.
- [20] P. MÜLLER AND P. STOLLMANN, *Spectral asymptotics of the Laplacian on supercritical bond-percolation graphs*, J. Funct. Anal., 252 (2007), pp. 233–246.
- [21] P. NEVAI, *Géza Freud, orthogonal polynomials and Christoffel functions. A case study*, J. Approx. Theory, 48 (1986), pp. 3–167.
- [22] P. G. NEVAI, *Orthogonal Polynomials*, AMS, Providence, RI, 1979.
- [23] M. PENROSE, *Random Geometric Graphs*, Oxford Stud. Probab. 5, Oxford University Press, Oxford, UK, 2003.
- [24] B. T. POLYAK, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, New York, 1987.
- [25] P. REBESCHINI AND S. TATIKONDA, *Accelerated consensus via min-sum splitting*, in Proceedings of the 31st Conference on Advances in Neural Information Processing Systems, Long Beach, CA, 2017, pp. 1374–1384.
- [26] S. SARDELLITTI, M. GIONA, AND S. BARBAROSSA, *Fast distributed average consensus algorithms based on advection-diffusion processes*, IEEE Trans. Signal Process., 58 (2010), pp. 826–842.

- [27] K. SCAMAN, F. BACH, S. BUBECK, Y. T. LEE, AND L. MASSOULIÉ, *Optimal algorithms for smooth and strongly convex distributed optimization in networks*, in Proceedings of the 34th International Conference on Machine Learning, Vol. 70, Sydney, Australia, 2017, pp. 3027–3036.
- [28] D. SHAH, *Gossip algorithms*, Found. Trends® Networking, 3 (2009), pp. 1–125.
- [29] G. SZEGÖ, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ. 23, AMS, Providence, RI, 1939.