Performance metrics in the average consensus problem: a tutorial

E. Lovisari, S. Zampieri [1]

Abstract

The average consensus algorithm is a distributed procedure which allows a network of agents to agree on the average of a set of initial values. The computation occurs through local exchange of information only, namely the information exchange takes place only between agents which are neighbors with respect to a graph representing the system communication architecture. Several performance metrics have been proposed for the evaluation of this algorithm. Particularly interesting and challenging is to relate them to the communication topology. Different performance metrics may yield different answers in comparing alternative communication topologies. In this paper, we present a few performance metrics and we show how these metrics are related to the communication topology. In particular, when available, we present bounds which permit to relate performance and topology for general graphs, for graphs with symmetries, called d-dimensional tori, and for geometric graphs.

Keywords: linear consensus, graph theory, sensor networks, Cayley graphs, geometric graphs, distributed estimation

1 Introduction

Distributed algorithms constitute a growing field of research within many scientific communities. The possibility of deploying a large networks of small, simple and cheap but intelligent units forces researchers to enlarge their point of view from centralized, fast and fully designable procedures, to leaderless algorithms in which information can flow only according to a constrained communication architecture. Interesting examples come from coordinated control of mobile vehicles [1, 2, 3, 4, 5, 6, 7]. In this scenario some agents are required to move in a possibly unknown environment maintaining a formation suitable to their current objective. The absence of a leader can be a big advantage in this application. Indeed, if the environment is dangerous, we want to avoid that the failure of the leader yields the disruption of the entire formation, neither we want use a time–consuming leader–election procedure. Instead, it is often preferable that all the agents are on the same hierarchical level. Moreover, in other cases, a centralized control is undesirable when a large environment or a slow communication cause delays which could destabilize the system. Finally, if the number of agents is large, the computational effort required to design a centralized control could be unbearable. These requirements motivate the interest on distributed algorithms. The task to be accomplished, whatever it is, must be distributed over the network, and each agent contributes to build up the solution. Clearly,

[1]E. Lovisari and S. Zampieri are with the Department of Information Engineering, University of Padova, [lovisari, zampi]@dei.unipd.it

The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n°257462 HYCON2 Network of excellence and n°223866 FeedNetBack.

only in particular cases it is possible, via a distributed algorithm, to obtain the same level of performance as in the centralized case.

Similar features characterize several other applications, such as distributed estimation [8, 9, 10, 11, 12], load balancing [13, 14], sensor calibration for sensor networks [15, 16, 17, 18, 19], distributed optimization [20, 21, 22, 23, 24, 25, 26], distributed demodulation [27, 28]. In general, in all these problems a set of agents is given and these agents can communicate with each other. The information exchange is modeled via a communication graph. Each node of the graph represents an agent, an edge represents the possibility for an agent to receive information from another one. In certain cases, the communication graph is allowed to be time-variant. This is used to model asynchronous working conditions or scenarios in which two agents are not always able to communicate, because of possible packet losses. The richness of such scenarios is clear, since the enormous number of states that characterize such systems opens new frontiers in research in terms of complexity of algorithms, of tasks to be accomplished, and so on. Vice versa, the same complexity yields an extreme difficulty for the designers, since many problems are known to be NP-hard when the communication constraints are imposed.

This difficulty forced researchers to spend much effort on the design of simple algorithms for simple tasks. One of these algorithms is the average consensus algorithm. In general, an average consensus algorithm is a distributed strategy in which each agent has a real number. The goal for the agents is to compute the average of those numbers. One way to accomplish this task is to label each agent, gather the values via multi-hop communication in the network, and combine them in the same way by each node. However, this implies that each agent has to memorize the values of all the other nodes, which can be hard if the network is large. The average consensus algorithm is an iterative procedure, much simpler and easier to be implemented, based on a linear local iterative data fusion. On the one hand, this method is in general less efficient than the methods based on multi-hop communications, but on the other hand, it is more robust to both communication drops and node failures.

1.1 The average consensus algorithm

The average consensus algorithm consists in the following discrete time linear algorithm. Assume that each node u has a real number y_u and a state $x_u(t)$ which is initialized to y_u, namely $x_u(0) = y_u$. Assume that u updates $x_u(t)$ according to the iteration

$$x_u(t+1) = p_{uu}(t)x_u(t) + \sum_{v \in \mathcal{N}_u} p_{uv}(t)x_v(t)$$

where $\mathcal{N}_u := \{v \in V \setminus \{u\} \ : \ (v,u) \in \mathcal{E}\}$ is the set of neighbors of the agent u. We impose that $p_{uv}(t) \geq 0$ and that $p_{uu}(t) + \sum_{v \in \mathcal{N}_u} p_{uv}(t) = 1$ or, in words, that $x_u(t+1)$ is a convex combinations of all the states available to the agent u. If we stack all the states $x_u(t)$'s into a vector $\boldsymbol{x}(t) \in \mathbb{R}^N$ and we introduce the matrix $P(t) \in \mathbb{R}^{N \times N}$ having entries equal to $p_{uv}(t)$ in position u, v, we can rewrite the previous iteration in the following compact form

$$\boldsymbol{x}(t+1) = P(t)\boldsymbol{x}(t). \tag{1}$$

Notice that $P(t)$ is stochastic, namely it has non-negative entries and $P(t)\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes a column vector with all entries equal to 1.

In this paper we will limit to time-invariant consensus algorithms, namely to the case in which the iteration is the following

$$\boldsymbol{x}(t+1) = P\boldsymbol{x}(t) \tag{2}$$

where P is a stochastic matrix consistent with the graph $\mathcal{G}$. If we assume that all the diagonal entries of P are strictly positive and if we assume that the graph associated with P is strongly connected (these notions will be introduced below), then P is primitive and thus, by Frobenius-Perron theorem [29], all the eigenvalues of P lie inside the open unit disk, except one which is equal to 1. Moreover the eigenvalue 1 has algebraic multiplicity 1. Let $\boldsymbol{\pi} \in \mathbb{R}^N$ be a left eigenvector of P associated with the eigenvalue 1. Since it can be proved that all the entries of $\boldsymbol{\pi}$ are positive, it can be assumed with no loss of generality that $\boldsymbol{\pi}^T\mathbf{1} = 1$. Under the above assumptions we can argue that

$$P^t \overset{t\to\infty}{\longrightarrow} \mathbf{1}\boldsymbol{\pi}^T$$

This implies that

$$x_u(t) \overset{t\to\infty}{\longrightarrow} \alpha, \forall\, u = 1, \ldots, N$$

with $\alpha = \sum_{u=1}^N \pi_u x_u(0) = \sum_{u=1}^N \pi_u y_u$. We see in this way that the agents reach consensus on the value of their states $x_u(t)$, since they all converge to the same number α, called the consensus value. Notice that, if P is doubly stochastic, namely if both P and P^T are stochastic, then $\boldsymbol{\pi} = \frac{1}{N}\mathbf{1}$ and so the consensus value α is equal to the average of the initial states. In this case we have the average consensus.

While the pioneering work [30] studied the consensus algorithm for applications to distributed estimation, much work has been done in the last decade motivated by the formation control problem [31, 4]. A large number of papers have been proposed for studying the convergence of this algorithm. In particular, it has been highlighted its relations with Markov chains via the stochastic matrix P, and proposed control-oriented criteria (e.g., Nyquist criterion) in order to ensure its convergence [3]. Many papers included in the model more realistic scenarios, such as packet-drop communication and delays [32, 33, 34, 35, 36, 37] and some others have been devoted to the study of randomized consensus algorithms, namely consensus algorithms in which $P(t)$ is a matrix valued stochastic process [38, 39, 40, 41, 42].

As the classical theory suggests (see Section 4), the typical trajectory which the states draw while approaching the consensus value is exponential in time. Many papers have been devoted to the study of the exponential rate of convergence, both for structured graphs [43, 44, 45, 46, 47, 48, 49, 50], and in terms of optimization problems [51, 52, 53, 54, 55, 56]. Since consensus is often considered an algorithm which has a direct application to many sensor network problems, much effort has been spent trying to understand how consensus behaves under some typical constraints in communications, such as quantization of information and noisy channels [57, 58, 59, 60, 61, 62, 36, 37]. To conclude, an increasing interest is devoted to the problem of designing an effective distributed Kalman-type estimation algorithm. The consensus averaging procedure is applied in many cases either as a tool in a distributed version of the Kalman filter

or as a tool to build a state observer which is then required to satisfy a minimum variance criterion [8, 9, 10, 11, 63, 64, 8].

One relevant issue concerning consensus algorithms is related to the choice of a performance metric which allows the comparison of several possible choices of the matrix P and of the associated graph $\mathcal{G}_P$. Several different performance metrics have been proposed in the literature. One important aspect of the performance evaluation is how the communication topology, namely the graph, influences performance. More precisely, taken two different graph topologies $\mathcal{G}_1$ and $\mathcal{G}_2$, one can wonder if it may happen that $\mathcal{G}_1$ is better than $\mathcal{G}_2$ if compared through a certain performance metric, but the reverse occurs when the comparison is done through another performance metric. In other words the question is whether all the relevant performance metrics are coherent in comparing different topologies or not. Recent papers gave some partial answers to this question, and the answer is that topology does play a fundamental role, which should be an important guidance in the design process.

1.2 Paper contents

The aim of this paper is to give a tutorial on this subject focusing on the following three performance metrics:

- The exponential rate of convergence of the consensus algorithm to its asymptotic value.
- A time-dependent H_2 type performance index arising when using the consensus algorithm for distributed estimation.
- An H_2 type performance index arising when evaluating the effect of the additive noise to the asymptotic value of the consensus value.

The goal is to evaluate how these indices are influenced by the network topology. In wireless sensor networks, which is the main application area of consensus algorithms, the network topology is strongly related to the sensors positions. For this reason we restrict our attention to the class of network topologies described by geometric graphs. A geometric graph can be seen as a perturbation of a regular grid. More precisely in a geometric graph the nodes of the graph are deployed in an area and the connections between nodes have to satisfy some constraints which, roughly speaking, prevent nodes that are too far apart to be connected. This graph model, proposed in [65], can be seen as the deterministic counterpart of the random geometric graph [66].

Specifically, we propose bounds able to suggest how the performance indices scale with the number of nodes, when the network topology is described either by a general geometric graph, or by a geometric graph with symmetries, called a d-dimensional torus, or sometimes Cayley graphs. For tori, the results are easier to be obtained and, in some cases, only results of those graphs are in fact available. We preferred to give the proof only for the bounds on the rate of convergence performance index. This proof, which is complicated, is largely inspired by the proof of the analogous result given in [40] for the random geometric graph model. We preferred to present again the proof here, both because in this paper we are considering a slightly different model of geometric graph, and because in some points we found the original proof in [40] quite difficult to follow.

2 Some preliminaries on graphs and matrices

A graph is a quadruple $\mathcal{G} = (V, \mathcal{E}, s, t)$ where V is called the set of nodes, $\mathcal{E}$ is called the set of edges, and s and t are two functions $s : \mathcal{E} \to V$ and $f : \mathcal{E} \to V$. If $s(e) = u$ we say that the edge e starts in u, or that u is the tail of e. If $t(e) = v$ we say that the edge e ends in v, or that v is the head of e. In this paper we consider only graphs in which there do not exist different edges having the same tail and head. Hence, if an edge e is such that $s(e) = u$ and $f(e) = v$, then we can simply denote it as $e = (u, v)$ and write $\mathcal{G} = (V, \mathcal{E})$, where $\mathcal{E}$ is a subset of $V \times V$. We say that a graph is undirected if $(u, v) \in \mathcal{E} \iff (v, u) \in \mathcal{E}$. If this does not occur, then the graph is called directed, or a digraph. A path in the graph $\mathcal{G}$ from a node u to a node v is a set of edges $(u_i, v_i) \in \mathcal{E}$, $i = 1, \dots, l$, such that $u_1 = u$, $v_l = v$ and $v_i = u_{i+1}$ for all $i = 1, \dots, l-1$. The integer l is called the length of the path. The graphical distance between u and v, denoted with the symbol $d_{\mathcal{G}}(u, v)$ is given by the minimum length of the paths from u to v. A graph is said to be strongly connected if there exists a path from u to v for any pair $(u, v) \in V^2$.

Given a node $u \in V$, we denote by $\mathcal{N}_u^{in} = \{v : e = (v, u), \exists\, e \in \mathcal{E}\}$ the in-neighbor set of u, namely the set of nodes such that there exists an edge starting in such nodes and ending in u. Analogously, we denote by $\mathcal{N}_u^{out} = \{v : e = (u, v), \exists\, e \in \mathcal{E}\}$ the out-neighbor set of u, namely the set of node such that there exists an edge ending in such nodes and starting from u. Clearly, for an undirected graph the two notions coincide, so we will generically talk about the neighbor set $\mathcal{N}_u$ of u.

Given a graph $\mathcal{G} = (V, \mathcal{E})$, we will always denote by $N := |V|$ and $M := |\mathcal{E}|$ respectively the number of nodes and edges[2]. We define the incidence matrix $A \in \{0, \pm 1\}^{M \times N}$ as follows

$$A_{eu} = \begin{cases} -1 & \text{if } u = t(e) \\ 1 & \text{if } u = s(e) \\ 0 & \text{otherwise} \end{cases},$$

so the e-th row, related to directed edge e, has a -1 in correspondence with the ending node, and a 1 in correspondence with the starting node. By construction it is clear that $A\mathbf{1} = 1$. By Theorem 8.3.1 in [67] we know that the left kernel of A has dimension equal to $N - c$ where c is the number of connected components of the graph. We will always refer to connected graphs, thus the kernel of A has dimension 1 and it is generated by $\mathbf{1}$.

Another matrix related to a graph is the adjacency matrix $F \in \{0, 1\}^{N \times N}$, which is defined as

$$F_{uv} = \begin{cases} 1 & (u, v) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}.$$

For undirected graphs, the adjacency matrix is symmetric.

Given a matrix $P \in \mathbb{R}^{N \times N}$ we define the graph associated with P, denoted by the symbol $\mathcal{G}_P$, as the graph with vertex set $V = \{1, 2, \dots, N\}$ and edge set $\mathcal{E}_P := \{(u, v) \in V \times V \ : \ P_{u,v} \neq 0\}$. Moreover we say that P is consistent with a graph $\mathcal{G}$ if $\mathcal{G}_P$ is a subgraph of $\mathcal{G}$, namely they have the same vertex set and the edge set of $\mathcal{G}_P$ is a subset of the edge set of $\mathcal{G}$

[2]The symbol $|X|$ denotes the cardinality of the set X.

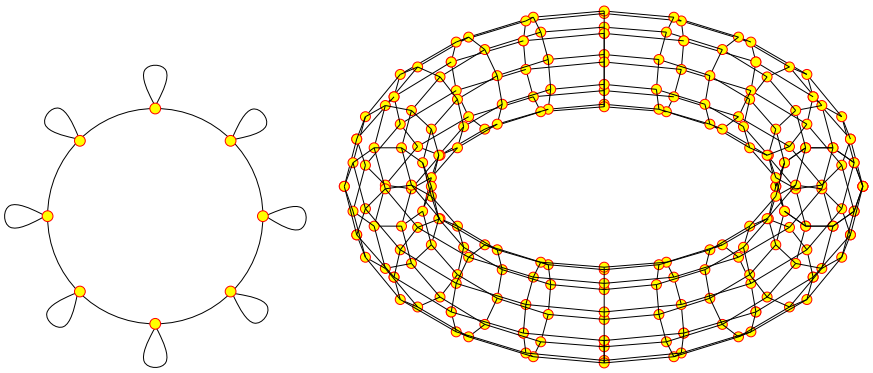

Figure 1: Two examples of d-dimensional tori.

2.1 Cayley matrices and d-dimensional tori

Cayley matrices and tori are highly structured matrices and graphs which present a number of symmetries, and that are defined through groups.

Definition 2.1. *Let G be an finite Abelian group of order $N = |G|$. A matrix $P \in \mathbb{R}^{G\times G}$ is said to be a Cayley matrix over the group G if*

$$P_{i,\,j} = P_{i+h,\,j+h},\ \forall\, i,\ j,\ h \in G. \tag{3}$$

It is easy to see [68] that there exists a function $\boldsymbol{g} : G \to \mathbb{R}$, called generator, such that $P_{ij} = \boldsymbol{g}(i-j)$. Note that $\boldsymbol{g}$ can be read from any row of P. A graph $\mathcal{G}$ is a d-dimensional torus if its adjacency matrix is a Cayley matrix. With this definition it is obvious that the graph associated with a Cayley matrix is automatically a torus, or more precisely a d_g-regular lattice on a d-dimensional torus, where d_g is the number of neighbors of each node. A d-dimensional torus is completely determined by giving the group G and a set $S \subseteq G$. Indeed, the set of edges $\mathcal{E}$ of a torus is such that $(i,\,j) \in \mathcal{E}$ if and only if $j - i \in S$.

In Fig. 1 two tori are presented. On the left, $G = \mathbb{Z}_8$ and $S = \{\pm 1,\, 0\}$ generate the circle with $N = 8$ nodes, in which each agent communicates with the node on the left and on the right. On the right, $G = \mathbb{Z}_{20} \times \mathbb{Z}_{10}$ and $S = \{(-1,0),\,(1,0),\,(0,1),\,(0,-1)\}$ generates the torus with $N_1 = 20$ circles of $N_2 = 10$ nodes each, where each agent communicates with the nodes on the left, on the right, above and below.

Notice that any finite Abelian group G is isomorphic to the group $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$, for some $n_1,\,\ldots,\,n_d \in \mathbb{N}$. In order to simplify the notation, in this paper we will restrict to d-dimensional tori with respect to groups of the type $\mathbb{Z}_n^d$. Moreover we will always assume that there is a positive constant δ (small enough compared with n) such that[3] $(i,\,j) \in \mathcal{E}$ only if $||j - i|| \leq \delta$. This constraint describes the assumption that a node can not communicate with nodes that are too "far" away from it. Another parameter will play an important role in the sequel. If we assume that the d-dimensional torus is connected, its connectivity

[3]Here we are assuming that the entries of i, j in $\mathbb{Z}_n$ are represented by the integers $-n/2+1,\ldots,-1,0,1,\ldots,n/2$ in case n is even or by the integers $-(n-1)/2,\ldots,-1,0,1,\ldots,(n-1)/2$ in case n is odd

implies that there exists a path from the node $\mathbf{0} = (0, \ldots, 0) \in \mathbb{Z}_n^d$ to the node $\boldsymbol{e}_i \in \mathbb{Z}_n^d$, which is the vector with all entries equal to zero except the entry in position i which is equal to 1. Let l_i be the graphical distance between $\mathbf{0}$ and $\boldsymbol{e}_i$ and let $l := \max_i\{l_i\}$. In many practical cases we have that $l = 1$. This parameter will play an important role when we will need to bound the graphical distance between nodes in tori.

2.2 Geometric graphs

Roughly speaking a geometric graph is a perturbation of a regular grid in d dimension, for instance by removing or adding nodes or/and edges. There exist several different mathematical models for networks deployed in a real environment. Some define certain geometric parameters which characterize the graph [65]. Others give either probabilistic rules to deploy agents in an area [66, 69] or rules to choose the edges connecting nodes [70, 71, 72]. All these models fall in the general class of the so called ad-hoc networks, or proximity–induced graphs. Here we have chosen the one proposed in [65, 73] because it is, in our opinion, simpler, rather general and, compared to the definition of random geometric graph model (see [66]), it requires no probabilistic rule.

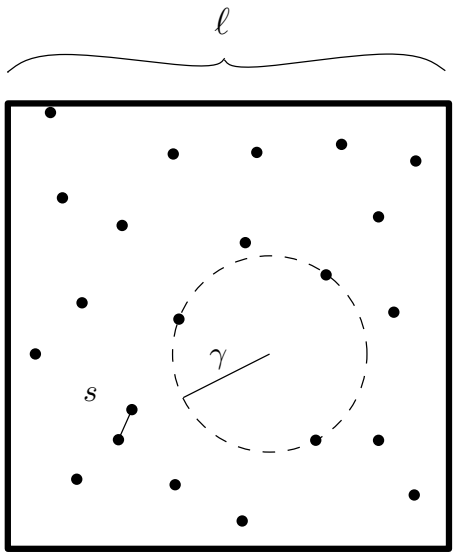

Figure 2: An example of geometric graph in 2 dimensions.

Consider an hypercube $Q \subset \mathbb{R}^d$ with edge of length ℓ, namely $Q = [0, \ell]^d \subseteq \mathbb{R}^d$. Let $\mathcal{G} = (V, \mathcal{E})$ be a undirected graph such that $V \subset Q$ and $|V| = N$. For such graphs embedded in $\mathbb{R}^d$, we denote by $d_E(u, v)$ the Euclidean distance among any couple of nodes $(u, v) \in V \times V$, namely if u and v are considered to denote also the points in $\mathbb{R}^d$ we have

$$d_E(u, v) = \sqrt{\sum_{k=1}^{d} (u_k - v_k)^2}. \tag{4}$$

Following [73, 74, 75], the following parameters can be defined:

- the minimum Euclidean distance between any two nodes

$$s = \min_{u,\, v \in V,\, u \neq v} \{d_{\mathrm{E}}(u,\, v)\}\,; \tag{5}$$

- the maximum Euclidean distance between any two connected nodes

$$r = \max_{(u,\, v) \in \mathcal{E}} \{d_{\mathrm{E}}(u,\, v)\}\,; \tag{6}$$

- the radius γ of the largest ball centered in Q containing no nodes of the graph

$$\gamma = \sup \{r |\, B(x,\, r) \cap V = \emptyset,\, \exists\, x \in Q\}\,; \tag{7}$$

- the minimum ratio between the Euclidean distance of two nodes and their graphical distance

$$\rho = \min \left\{ \frac{d_{\mathrm{E}}(u,\, v)}{d_{\mathcal{G}}(u,\, v)} |\, (u,\, v) \in V \times V \right\}. \tag{8}$$

Notice that, if $\mathcal{G}$ is connected, then ρ is well defined and positive[4]. Such a graph is called a geometric graph with parameters $(N, d, \ell, s, r, \gamma, \rho)$. An example of geometric graph is pictured in Fig. 2.

2.3 De Bruijn's graphs

De Bruijn's graphs constitute a very particular class of graphs. Nevertheless we introduce them here because they are fast mixing graphs in which, if the weights are chosen in a clever manner, it is possible to reach consensus in a minimum finite number of steps. Related to this nice behavior, de Bruijn's graphs have been proposed as the optimal architecture for multi-processor networks due to their fast information spreading properties, inspiring scientific papers [76] as well as patents [77]. As a matter of fact, instead, in ad-hoc networks the De Bruijn is hardly implemented due to the long-range communication it requires. We choose it as an example to illustrate an extreme behavior of the performance costs under analysis. The paper [44] uses de Bruijn's graphs to solve linear consensus and gives detailed and general results on these graphs and provides several useful properties and characterizations. We will restrict here to a particular case. Let k and n be two positive integers, and consider the graph $\mathcal{G}$ whose adjacency matrix is the following

$$F = \mathbf{1} \otimes I \otimes \mathbf{1}^T$$

where the column vector $\mathbf{1}$ is k dimensional, the identity I is k^{n-1} dimensional and the symbol $\otimes$ denotes Kronecker product. The graph $\mathcal{G}$ is called de Bruijn's graph [78], and it is displayed in Figure 3 in the case $k = 2$, $n = 3$.

[4]When $\mathcal{G}$ is not connected, then we assume by definition that $\rho = 0$.

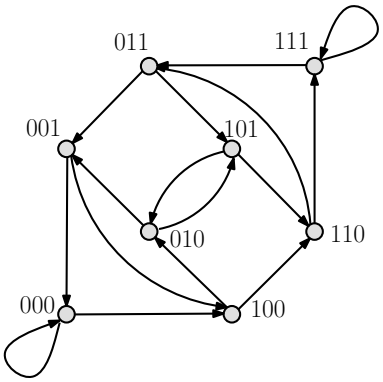

Figure 3: A de Bruijn graph with $N = 8$ nodes.

3 Electrical networks

It has been firstly stated in [65] the remarkable and deep analogy between reversible Markov chains and electrical networks. We focus our attention on Markov chains with symmetric P, which are a particular case of reversible chains.

Define a resistive electrical network as a pair $(\mathcal{G}, C)$, or equivalently $(\mathcal{G}, R)$, where

- $\mathcal{G}$ is an undirected graph (without self-loops), with N vertices, and M edges;
- C and R are two functions $C : \mathcal{E} \to [0, +\infty)$ and $R : \mathcal{E} \to [0, +\infty)$. They both associate with each edge of the graph a strictly positive number, called respectively the conductance and the resistance of the edge, one the inverse of the other.

Actually, we can assume that any non-existing edge has zero conductance or infinite resistance.

Recall that in an undirected graph to any connected pair of nodes corresponds two directed edges, one for each direction. For this reason we start from a map $C : \mathcal{E} \to [0, +\infty)$ such that $C((u, v)) = C((v, u))$. We define the matrix $\mathcal{C} \in \mathbb{R}^{M \times M}$ as a diagonal matrix with diagonal entries equal to the conductances on the edges, namely $\mathcal{C}_{ee} = C(e)$ for all $e \in \mathcal{E}$. We denote the Laplacian of the network by $L_C := A^T \mathcal{C} A$. Recalling that $A\mathbf{1} = 0$, we immediately have $L_C \mathbf{1} = 0$. Using the notion of incidence matrix, it is immediate to obtain

$$[L_C]_{uv} = [A^T \mathcal{C} A]_{uv} = \begin{cases} 2c_u & \text{if } u = v \\ -2C(e) & \text{if } (u, v) = e \in \mathcal{E} \\ 0 & \text{if } (u, v) \notin \mathcal{E} \end{cases}, \tag{9}$$

where $c_u := \sum_{e|u=t(e)} C(e)$ is the sum of all the conductances of the edges ending in u.

We define the effective resistance between two nodes u and v in the electrical network $(\mathcal{G}, C)$ as the quantity

$$\mathcal{R}_{uv}(\mathcal{G}, C) = \frac{v_u - v_v}{I}, \tag{10}$$

where v_u and v_v are the potentials at the nodes u and v when we inject a current of value I into u and we extract the same from v.

We are now going to obtain the effective resistance between two nodes in terms of a matrix depending on L_C. This procedure is quite well known, see for example [79]. Consider any vector $\boldsymbol{i} \in \mathbb{R}^N$ such that $\mathbf{1}^T\boldsymbol{i} = 0$, and assume to inject (or extract if negative) the current i_u into the node u of the network, for any node. Since $\mathbf{1}^T\boldsymbol{i} = 0$, the total current injected into and extracted from the network is zero and this is a necessary condition for the problem to be well-posed. The current $\boldsymbol{i}$ injected into the network induces the potentials $\boldsymbol{v} \in \mathbb{R}^N$ at the nodes and the currents $\boldsymbol{j} \in \mathbb{R}^M$ on the edges. Potentials and currents must satisfy both Kirchhoff's current law and Ohm's law, which can be represented by the following linear system

$$\begin{cases} A^T\boldsymbol{j} = \boldsymbol{i} \\ \mathcal{C}A\boldsymbol{v} = \boldsymbol{j} \end{cases} \tag{11}$$

which in turn implies $L_C\boldsymbol{v} = A^T\mathcal{C}A\boldsymbol{v} = \boldsymbol{i}$. We solve the electrical equations of the network if we are able to find $\boldsymbol{v}$ and $\boldsymbol{j}$ from $\boldsymbol{i}$. It is clear that $\boldsymbol{v}$ is not uniquely determined by (11), since, if $\boldsymbol{v}$ is a solution, then also $\boldsymbol{v} + \alpha\mathbf{1}$ is a solution. We have to impose a constraint on $\boldsymbol{v}$ in order to ensure the solution to be unique. We will assume in this paper that $\boldsymbol{v}$ has zero mean, namely we impose $\boldsymbol{v}^T\mathbf{1} = 0$. In this way we can rewrite (11) and the constraint on $\boldsymbol{v}$ in the following matrix form

$$\begin{bmatrix} L_C & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{i} \\ 0 \end{bmatrix}. \tag{12}$$

In order to find $\boldsymbol{v}$, we introduce the so called Green matrix of L_C, or of the electrical network, which is the unique matrix X_C such that

$$\begin{cases} L_C X_C = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \\ \mathbf{1}^T X_C = 0 \end{cases}. \tag{13}$$

We can write X_C in closed formula as follows

$$X_C = (L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T)^{-1} - \frac{1}{N}\mathbf{1}\mathbf{1}^T. \tag{14}$$

This formula holds since $L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T$ is positive definite and thus invertible. Indeed, we have that

$$\boldsymbol{x}^T(L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T)\boldsymbol{x} = \boldsymbol{x}^T A^T\mathcal{C}A\boldsymbol{x}^T + \frac{1}{N}(\mathbf{1}^T\boldsymbol{x})^2,$$

where the right hand is the sum of two nonnegative terms. Since $\mathcal{C} > 0$, in order the sum to be zero, the vector $\boldsymbol{x}$ must satisfy

$$\begin{cases} A\boldsymbol{x} = 0 \\ \mathbf{1}^T\boldsymbol{x} = 0 \end{cases}.$$

Since the first constraint implies $\boldsymbol{x} = \alpha\mathbf{1}$ (recall that, since the graph is connected, $\ker A = \operatorname{span}\{\mathbf{1}\}$), the second implies $\alpha = 0$, yielding in this way that $L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T$ is positive definite.

The following lemma rewrites this expression in terms of power series.

Lemma 3.1. *Let $\alpha \geq \frac{1}{2}\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the symmetric positive semi-definite matrix L_C. Then*

$$X_C = \frac{1}{\alpha}\sum_{t\geq 0}\left((I - \frac{1}{\alpha}L_C)^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right)$$

Proof. For any $\alpha \in \mathbb{R}\setminus\{0\}$, we can rewrite (14) as

$$X_C = \frac{1}{\alpha}\left[(\frac{1}{\alpha}L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T)^{-1} - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right].$$

The condition $\alpha \geq \frac{1}{2}\lambda_{\max}$ ensures that all the eigenvalues of $I - \frac{1}{\alpha}L_C - \frac{1}{N}\mathbf{1}\mathbf{1}^T$ lie in the interval $(-1, 1)$, so we can write in power series

$$(\frac{1}{\alpha}L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T)^{-1} = \sum_{t\geq 0}\left(I - \frac{1}{\alpha}L_C - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right)^t.$$

The thesis now follows from the fact that $L_C\mathbf{1} = 0$. □

By construction, the Green matrix can be directly used to obtain $\boldsymbol{v}$ in (12), yielding

$$\boldsymbol{v} = X_C\boldsymbol{i}. \tag{15}$$

Assume now that $\boldsymbol{i} = \boldsymbol{e}_u - \boldsymbol{e}_v$. This corresponds to inject 1 Ampere into the node u and extract 1 Ampere from the node v. Since $v_u - v_v = \boldsymbol{v}^T\boldsymbol{i}$, by definition (10) we can conclude that

$$\mathcal{R}_{uv}(\mathcal{G}, C) = \boldsymbol{v}^T\boldsymbol{i} = (\boldsymbol{e}_u - \boldsymbol{e}_v)^T X_C(\boldsymbol{e}_u - \boldsymbol{e}_v).$$

Consider now a symmetric, stochastic, primitive matrix $P \in \mathbb{R}^N$. We build an electrical network taking N nodes and setting $C((u, v)) = P_{uv}$. If $P_{uv} = 0$, we assume that there is not an edge between u and v. Notice that this definition implies that there are possibly nonzero conductances in the self loops.

From this definition of the electrical network, it is easy to obtain the following relation between P and the Laplacian L_C of the network

$$P = I - \frac{1}{2}L_C.$$

This relation implies that the largest eigenvalue of L_C satisfies $\lambda_{\max} \leq 4$. Thus we can use Lemma 3.1 with $\alpha = 2$, obtaining

$$X_C = \frac{1}{2}\sum_{t\geq 0}\left((I - \frac{1}{2}L_C)^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right),$$

and, in terms of P,

$$X_C = \frac{1}{2}\sum_{t\geq 0}\left(P^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right).$$

We can thus compute the effective resistance between two nodes u and v in the network build from P as follows

$$\mathcal{R}_{uv}(\mathcal{G},\, P) = \frac{1}{2}(\boldsymbol{e}_u - \boldsymbol{e}_v)^T X (\boldsymbol{e}_u - \boldsymbol{e}_v),$$

where we define

$$X := \sum_{t\geq 0} \left(P^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T \right),$$

and where we use the notation $\mathcal{R}_{uv}(\mathcal{G},\, P)$ in order to underline the role of P.

This relation, which is of interest by itself, will be used in order to analyze the H_2 performance index (Section 7).

4 The rate of convergence

The most classical performance index for the evaluation of the convergence of an iterative algorithm is the speed of convergence of the algorithm output towards its asymptotic value. It is well known from classical system theory that the rate of convergence of the state $\boldsymbol{x}(t)$ to its asymptotic value is exponential. The exponential rate of convergence is then defined as

$$R := \lim_{t\to\infty} (||\boldsymbol{x}(t) - \boldsymbol{x}(\infty)||)^{1/t}$$

where $||\cdot||$ denotes the 2-norm of a vector. Assume we are given a primitive stochastic matrix P. By standard linear algebra it can be seen that $R = \rho(P)$ where $\rho(P)$ is the essential spectral radius of P

$$\rho(P) := \max\{|\lambda| :\, \lambda \in \Lambda(P) \setminus \{1\}\}, \tag{16}$$

where $\Lambda(P)$ is the set of all the eigenvalues of P. The relation between the essential spectral radius of P and the topology of the graph $\mathcal{G}_P$ associated with P is a problem which has been widely studied both in the Markov chains community and in the community studying the theory of graphs. In Markov chain theory $\rho(P)$ is related to the so called mixing time of the Markov chain having P as the transition matrix [80]. Spectral graph theory instead studies the geometric properties of graphs using the so-called Laplacian matrix L of the graph. This matrix is related to P via $P = I - L$, so that $\Lambda(P) = 1 - \Lambda(L)$ and the spectral properties of the two matrices essentially coincide. Once we sort in increasing order the eigenvalues of L

$$0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_{N-1}$$

the value λ_2 is known as the Fiedler eigenvalue [81], or algebraic connectivity of L. Its characteristics are closely related to those of $\rho(P)$. An extensive treatment of these and many other graph–theoretic topics can be found in the book [82].

Example: de Bruijn's graphs

Consider the de Bruijn's graph with $N = k^n$ we defined previously, and assume that each node uniformly weights all its neighbors, namely the consensus matrix is

$$P = \frac{1}{k}\mathbf{1} \otimes I \otimes \mathbf{1}^T$$

where the column vector $\mathbf{1}$ is k dimensional and the identity I is k^{n-1} dimensional. Notice that P defined in this way is not symmetric but it is doubly stochastic and that the number of neighbors of each agent is exactly k. A simple computation [44] shows that

$$P^h = \frac{1}{N}\mathbf{1}\mathbf{1}^T, \forall\, h \geq k$$

where column vector $\mathbf{1}$ is here N dimensional. This means that with this matrix the state converges to consensus in at most $n = \log_k N$ steps. It is possible to show that there exists no $k^n \times k^n$ stochastic matrix for which we have faster convergence. Notice finally that P has the minimum possible essential spectral radius since $\rho(P) = 0$.

4.1 Bounds on the convergence rate for general graphs

One of the major issues in research on consensus algorithms and Markov chains is to understand how to bound the essential spectral radius of P in terms of geometric parameters of the network.

In the sequel we will briefly recall some classical results (see [83, 84, 85, 86, 87]). For simplicity, as already said, we will restrict our attention to symmetric stochastic matrices. If P is symmetric, then its eigenvalues are real, and in the sequel we will assume that they are ordered in such a way that $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1}$. Notice that in this case we have that

$$\rho(P) = \max\{\lambda_1, -\lambda_{N-1}\}$$

Therefore the second largest eigenvalue λ_1 will play an important role in determining bounds on $\rho(P)$. Indeed, if we find bounds on λ_1 and on $-\lambda_{N-1}$ we can get an upper bound on $\rho(P)$ too. Applying Gershgorin circle theorem we can argue that

$$-\lambda_{N-1} \leq 1 - 2\max_i\{P_{ii}\}.$$

Therefore, the difficulty in finding bounds on $\rho(P)$ essentially stands in finding bounds on λ_1. This eigenvalue will be denoted with the symbol $\lambda_1(P)$.

The well-known Rayleigh-Ritz theorem (see [88]) proves to be a helpful tool in bounding $\lambda_1(P)$. Rayleigh-Ritz theorem, in our case of P symmetric, coincides with the following variational characterization of

$$1 - \lambda_1(P) = \min\left\{\frac{\boldsymbol{x}^T(I-P)\boldsymbol{x}}{\boldsymbol{x}^T\Omega\boldsymbol{x}}, \boldsymbol{x} \neq \alpha\mathbf{1}\right\}, \tag{17}$$

where $\Omega = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T$ is the projector over span $\{\mathbf{1}\}^\perp$.

This characterization is the basis of several results relating geometric parameters of the graph associated with the stochastic matrix P to its second largest eigenvalue. We will briefly review two among the most important ones, namely the Poincarè and the Cheeger inequalities. For the proof of both of them, we refer to [83].

4.1.1 Poicarè inequality

Let P be a symmetric stochastic matrix and let $\mathcal{G}_P = (V, \mathcal{E})$ be the undirected graph associated with P. For any couple $(u, v) \in V \times V$, $u \neq v$, let γ_{uv} be a

path from u to v. Namely, γ_{uv} is a set of edges $\gamma_{uv} = \{e_0, \ldots, e_l\}$ such that $e_0 = (u, u_1)$, $e_i = (u_i, u_{i+1})$ $\forall\, i = 1, \ldots, l-1$ and $e_l = (u_l, v)$. We will assume that in a path a vertex can be touched many times, while an edge may appear at most once. We define the following weighted length of the path γ_{uv}

$$|\gamma_{uv}|_P = \sum_{e \in \gamma_{uv}} P(e)^{-1}$$

where e are the edges forming γ_{uv} and $P(e) = p_{z,w}$ if $e = (z, w)$.

Let Γ be a collection of such paths, one for each pair (u, v). We associate to Γ the following quantity

$$\kappa = \kappa(\Gamma) = \max_{e \in \mathcal{E}} \{ \sum_{\gamma_{uv} \ni e} |\gamma_{uv}|_P \} \tag{18}$$

namely κ is the maximum, as e varies over $\mathcal{E}$, of the sum of $|\gamma_{uv}|_P$ for all the paths γ_{uv} in which e appears as an edge.

This value has an immediate, intuitive, interpretation. We take, for any path, the measure of the resistance to the flow of information through that path. Then, we maximize it over the edges of the graph, obtaining thus a measure of the bottleneck in the network. This bottleneck influences the rate of convergence to the asymptotic distribution of the states as stated in the following theorem.

Theorem 4.1. *The second largest eigenvalue of P satisfies*

$$\lambda_1(P) \leq 1 - \frac{N}{\kappa}, \tag{19}$$

with κ defined in (18).

This inequality is fundamentally an edge-perspective bound. It links geometric properties of paths along the network with the rate of convergence. Intuitively, less information can flow along the paths considered, the slowest is the convergence.

4.1.2 Cheeger inequality

In this section the interest is switched from paths to "surfaces", giving the definition of Cheeger ratio, as well as the relation between such quantity and the second largest eigenvalue. Unfortunately, even if the computation of the bound is someway simpler, this approach has been proved to offer less effective results over large families of graphs if compared with the Poincarè inequality [86].

Let P be a symmetric stochastic matrix and let $\mathcal{G}_P = (V, \mathcal{E})$ be the undirected graph associated with P. Take a proper subset $S \subseteq V$ of the nodes. It is rather intuitive that the flow of information from the set S to its complement $S^C = V \setminus S$ is linked to the transition probability from S to S^C. We can thus consider the conditional expectation of crossing the boundary of S given that we started from S, and minimize it over any possible set S. We obtain in this way the so called Cheeger ratio

$$h(P) = \min_{S : |S| \leq \frac{N}{2}} \left\{ \frac{P(S \times S^C)}{|S|} \right\}, \tag{20}$$

where $S \subseteq V$ and $P(S \times S^C) = \sum_{(x,\, y) \in S \times S^C} P_{xy}$.

This quantity can be used in order to derive both an upper and a lower bound on the second largest eigenvalue, as stated in the following result.

Theorem 4.2. *The second largest eigenvalue of P satisfies the following inequalities*

$$1 - 2h(P) \leq \lambda_1(P) \leq 1 - h(P)^2, \tag{21}$$

with $h(P)$ defined in (20).

4.2 Bounds on the convergence rate for Cayley matrices and d-dimensional tori

In this section we will present the results about the rate of convergence for the class of Cayley matrices and d-dimensional tori. Details and proofs can be found for example in [43, 85, 40]. Consider the class of Cayley matrices and d-dimensional tori with respect to the group $\mathbb{Z}_n^d$ with given d and n. Notice that $N = n^d$. Assume that the graphs belonging to this class have the self loops and that they are connected. Recall from section Sect. 2.1 the definition of l and δ.

The following theorem provides an upper bound on $\lambda_1(P)$. We don't give the proof of this result because it follows the same lines of the proof of Theorem 5.2 which treats a more general case.

Theorem 4.3. *Let P be a symmetric stochastic Cayley matrix with respect to the group $\mathbb{Z}_n^d$ whose associate graph $\mathcal{G}_P$ is in the previous class of d-dimensional tori characterized by the parameter h as described above. Assume that all the nonzero entries of P lie in an interval $[p_{\min}, p_{\max}]$. Then*

$$\lambda_1(P) \leq 1 - C\frac{1}{n^2}$$

where C is a strictly positive constant depending on d, l and $p_{\min}$.

The following theorem instead provides a lower bound on $\lambda_1(P)$. We give the proof of this result because this theorem will be instrumental in the proof of the more general result given in Theorem 5.1.

Theorem 4.4. *Let P be a symmetric stochastic Cayley matrix with respect to the group $\mathbb{Z}_n^d$ whose associate graph $\mathcal{G}_P$ is in the previous class of d-dimensional tori characterized by the parameter δ as described above. Then*

$$1 - C\frac{1}{n^2} \leq \rho(P)$$

where C is a strictly positive constant depending on d and δ.

Proof. First observe that, since $\rho(P) = \max\{\lambda_1, -\lambda_{N-1}\}$, it is enough to prove that

$$\lambda_1 \geq 1 - C'\frac{1}{n^2}$$

From the properties of the Cayley matrices [89] we can argue that the eigenvalues of P are given, for any $h = (h_1, \ldots, h_d) \in \mathbb{Z}_n^d$ by the formula

$$\lambda_h = \sum_{k \in \mathbb{Z}_n^d} P_{k,0} \cos\left(\frac{2\pi}{n} k^T h\right)$$

Take $h = e_i$, where e_i is the canonical vector defined above (see Sect. 2.1). Then, using the fact that $\cos x \geq 1 - x^2/2$ and the definition of the generator $\boldsymbol{g}$ of the matrix, we have

$$\lambda_{e_i} = \sum_{k \in \mathbb{Z}_n^d} \boldsymbol{g}(k) \cos\left(\frac{2\pi}{n} k_i\right) \geq \sum_{k \in \mathbb{Z}_n^d} \boldsymbol{g}(k) \left(1 - \frac{2\pi^2}{n^2} k_i^2\right) = 1 - \left(\sum_{k \in \mathbb{Z}_n^d} \boldsymbol{g}(k) k_i^2\right) \frac{2\pi^2}{n^2}$$

This in turn implies the thesis as follows

$$\begin{aligned}
\rho(P) &\geq 1 - \min_{i=1,\ldots,d} \left(\sum_{k \in \mathbb{Z}_n^d} \boldsymbol{g}(k) k_i^2\right) \frac{2\pi^2}{n^2} \\
&\geq 1 - \frac{1}{d} \sum_{i=1}^{d} \left(\sum_{k \in \mathbb{Z}_n^d} \boldsymbol{g}(k) k_i^2\right) \frac{2\pi^2}{n^2} \\
&= 1 - \frac{1}{d} \left(\sum_{k \in \mathbb{Z}_n^d} \boldsymbol{g}(k) ||k||^2\right) \frac{2\pi^2}{n^2} \\
&\geq 1 - \frac{2\pi^2 \delta^2}{d} \frac{1}{n^2}
\end{aligned}$$

□

It is possible to obtain a similar result if we consider instead of Cayley matrices, more general matrices which are consistent with d-dimensional tori. We start from the following theorem providing an upper bound on $\lambda_1(P)$. We don't give the proof of this result because it follows the same lines of the proof of Theorem 5.2 which treats a more general case.

Theorem 4.5. *Let P be a symmetric stochastic matrix whose associate graph $\mathcal{G}_P$ is in the previous class of d-dimensional tori characterized by the parameter h as described above. Assume that all the nonzero entries of P lie in an interval $[p_{\min}, p_{\max}]$. Then*

$$\lambda_1(P) \leq 1 - C \frac{1}{n^2}$$

where C is a strictly positive constant depending on d, l and $p_{\min}$.

We give finally a lower bound on $\rho(P)$. For this result we consider the case in which the stochastic matrices are not restricted, as in the rest of the paper, to be symmetric, but we give the result for the more general class of reversible matrices. This because this theorem in this form will be a step in the proof of the more general result given in Theorem 5.1.

A stochastic matrix P is called reversible [83] with respect to the diagonal matrix Π if

$$P^T \Pi = \Pi P$$

This means that P is a self adjoint operator with respect to the inner product $\langle x, y \rangle := x^T \Pi y$. Notice that a stochastic matrix is symmetric if and only if it is reversible with respect to the identity matrix. Notice moreover that the eigenvalues of a reversible stochastic matrix P are real, and that if $\boldsymbol{\pi}^T = \mathbf{1}^T \Pi$

and $\boldsymbol{\pi}^T\mathbf{1} = 1$, then $\boldsymbol{\pi}^T$ is the normalized left eigenvalue of P associated to 1, coherently with the notations adopted so far. We will assume that the eigenvalues of P are ordered in such a way that $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1}$.

Theorem 4.6. *Let P be a stochastic matrix which is consistent with a graph $\mathcal{G}$ which belongs to the previous class of d-dimensional tori characterized by the parameter δ as described above. Assume moreover that P is reversible with respect to a diagonal matrix $\Pi = \mathrm{diag}\,(\pi_1, \ldots, \pi_N)$. Then*

$$1 - C\frac{1}{n^2} \leq \rho(P)$$

where C is a strictly positive constant depending on d, δ and $\pi_{\min}/\pi_{\max}$, where $\pi_{\min} := \min_i\{\pi_i\}$ and $\pi_{\max} := \max_i\{\pi_i\}$.

Proof. First observe that, since $\rho(P) = \max\{\lambda_1, -\lambda_{N-1}\}$, it is enough to prove that

$$\lambda_1(P) \geq 1 - C'\frac{1}{n^2}$$

Observe moreover that (17) can be adapted to reversible stochastic matrices as follows

$$\mu_1 = \min\left\{\frac{\boldsymbol{x}^T\Pi L\boldsymbol{x}}{\boldsymbol{x}^T(\Pi - \boldsymbol{\pi}\boldsymbol{\pi}^T)\boldsymbol{x}},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}$$

where $L := I - P$ and μ_1 is its second largest eigenvalue. Consider now for any $h \in G := \mathbb{Z}_n^d$ the operator σ_h over the matrices in $\mathbb{R}^{G\times G}$ defined by letting $\sigma_h(P)_{i,j} := P_{i+h,j+h}$. Notice that

$$\bar{P} := \sum_{h\in G}\sigma_h(\Pi P)$$

is still compatible with the graph $\mathcal{G}$ and moreover it is a symmetric Cayley matrix with respect to the group G. Let $1 = \bar{\lambda}_0 > \bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_{N-1}$ be the eigenvalues of $\bar{P}$. Let moreover $\bar{L} := I - \bar{P}$ and $\bar{\mu}_1$ be its second largest eigenvalue. Then we have that

$$\begin{aligned}
\bar{\mu}_1 &= \min\left\{\frac{\boldsymbol{x}^T\bar{L}\boldsymbol{x}}{\boldsymbol{x}^T\Omega\boldsymbol{x}},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}\\
&= \min\left\{\frac{1}{\boldsymbol{x}^T\Omega\boldsymbol{x}}\boldsymbol{x}^T\left(\sum_{h\in G}\sigma_h(\Pi L)\right)\boldsymbol{x},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}\\
&\geq \sum_{h\in G}\min\left\{\frac{1}{\boldsymbol{x}^T\Omega\boldsymbol{x}}\boldsymbol{x}^T\sigma_h(\Pi L)\boldsymbol{x},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}\\
&= \sum_{h\in G}\min\left\{\frac{\boldsymbol{x}^T\Pi L\boldsymbol{x}}{\boldsymbol{x}^T\Omega\boldsymbol{x}},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}\\
&= N\min\left\{\frac{\boldsymbol{x}^T\Pi L\boldsymbol{x}}{\boldsymbol{x}^T\Omega\boldsymbol{x}},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}\\
&\geq N^2\pi_{\min}^2\min\left\{\frac{\boldsymbol{x}^T\Pi L\boldsymbol{x}}{\boldsymbol{x}^T(\Pi - \boldsymbol{\pi}\boldsymbol{\pi}^T)\boldsymbol{x}},\ \boldsymbol{x} \neq \alpha\mathbf{1}\right\}\\
&\geq \left(\frac{\pi_{\min}}{\pi_{\max}}\right)^2\mu_1
\end{aligned}$$

The last inequalities are motivated by the fact that

$$\begin{aligned}\boldsymbol{x}^T(\Pi - \boldsymbol{\pi}\boldsymbol{\pi}^T)\boldsymbol{x} &= \frac{1}{2}\sum_{uv}(x_u - x_v)^2\pi_u\pi_v \geq \pi_{\min}^2\frac{1}{2}\sum_{uv}(x_u - x_v)^2 \\ &= N\pi_{\min}^2\boldsymbol{x}^T\Omega\boldsymbol{x}\end{aligned}$$

and that $\pi_{\max} \geq \frac{1}{N}$.

Considering that, by Theorem 4.4 we have that $\bar{\mu} \leq C/n^2$ for some constant C, we get that the previous inequality implies that

$$\lambda_1 = 1 - \mu_1 \geq 1 - \left(\frac{\pi_{\max}}{\pi_{\min}}\right)^2\bar{\mu}_1 \geq 1 - \left(\frac{\pi_{\max}}{\pi_{\min}}\right)^2\frac{C}{n^2}$$

□

5 Bounds on the convergence rate for geometric graphs

In this section we will analyze the rate of convergence, namely the second largest eigenvalue in absolute value, of a generic symmetric stochastic matrix whose associated graph is a geometric graph. Our aim is to obtain a lower and an upper bound for such a quantity, and the tools used will be a Poincarè inequality-type for the upper bound, and the state–aggregation approach for Markov Chains for the lower bound. The procedure is similar to the one proposed in [40], where the authors study the random geometric graphs [66] in dimension d and show that the rate of convergence in such graphs is with high probability the same as the rate of convergence of a d-dimensional grid. Here we obtain a similar result for our deterministic model of geometric graphs. We start from the lower bound given in the following theorem whose proof is given in the appendix.

Theorem 5.1. *Let P be a symmetric stochastic matrix whose associate graph $\mathcal{G}_P$ is a geometric graph with parameters $(N, d, \ell, s, r, \gamma, \rho)$, where we assume that $\ell \geq 4\gamma$. Then*

$$1 - \frac{C}{N^{2/d}} \leq \rho(P) \tag{22}$$

where C is a strictly positive constant depending on the parameters d, s, r, γ and ρ but not on ℓ, N.

We will give now an upper bound on the second largest eigenvalue of a symmetric stochastic P whose associated graph is a geometric graph and whose entries lie in an interval $[p_{\min}, p_{\max}]$. The proof of this theorem, given in the appendix, makes use of the Poincarè inequality given in Eq. 19.

Theorem 5.2. *Let P be a symmetric stochastic matrix whose associate graph $\mathcal{G}_P$ is a geometric graph with parameters $(N, d, \ell, s, r, \gamma, \rho)$, where we assume that $\ell \geq 4\gamma$. We assume moreover that all the nonzero entries of P belongs to $[p_{\min}, p_{\max}]$. Then*

$$\lambda_1(P) \leq 1 - \frac{C}{N^{2/d}} \tag{23}$$

where C is a strictly positive constant depending on the parameters $p_{\min}$, d, s, r, γ and ρ but not on ℓ, N.

5.1 Applications

As a first example of application of the previous theorem we can consider the regular grid of $N = n^2$ nodes on the plane. In this case the nodes are deployed in a square of edge length equal to $\ell = n - 1$ and have coordinates i, j with $i, j \in \{0, \ldots, n-1\}$. In this scenario, it is clear that the distance among any pair of nodes is 1 and so we have that $s = 1$. Moreover the communication range is $r = 1$ and the disks which do not contain any node have radius which is at most $\gamma = \sqrt{2}/2$. Moreover, given a pair of nodes, it can be seen that the minimum of the ratio between the nodes Euclidean distance and graphical distance is $\rho = \sqrt{2}/2$. The regular grid it thus a geometric graph with these parameters, and so, if P is a symmetric stochastic matrix having the regular grid as its associated graph, then we can apply Theorem 5.2 which yields the well known result

$$\lambda_1 \leq 1 - C_r \frac{1}{N},$$

where C_r is a strictly positive constant depending only on the minimum value of the entries of P associated with the edges of the grid.

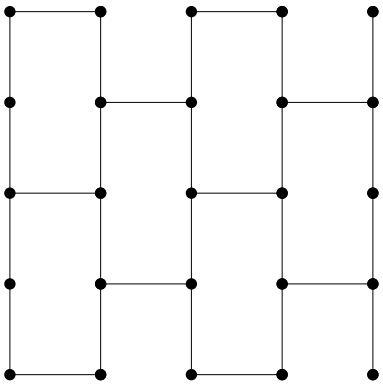

Figure 4: Perturbation of the regular grid.

Consider now a perturbation of such a grid. We take the same set of nodes, in the same positions, and we modify the communication topology as illustrated in Fig. 4. It is clear that in this case r, s and γ remain the same as in the previous example. Only the parameter ρ changes. It can be seen that ρ is determined by any two nodes at Euclidean distance 1 which are not neighbours in the perturbed grid. The application of Theorem 5.2 yields

$$\rho(P) \leq 1 - C_p \frac{1}{N}.$$

where C_p is a strictly positive constant depending only on the minimum value of the entries of P associated with the edges of the grid. Notice that in both the grid and in the perturbed grid the number of node is $N = n^2$. On the other hand in the regular grid we have $2n(n-1)$ edges, while in the perturbed grid we have $\frac{3}{2}n(n-1)$ edges. As shown in the proof of Theorem 5.2, the constants C_r and C_p are proportional to ρ and so $C_p = \frac{1}{\sqrt{2}} C_r$. This means that, in this

case, even dropping one fourth of the edges yields a rate of convergence with the same behavior in N and a different constant.

6 A time-dependent H_2 performance index

The application of the consensus algorithm to static estimation highlights a new performance index which is of H_2 type and will be called time dependent in order to distinguish it from another H_2 type performance index which will be introduced in the next section. More details on the time-dependent H_2 performance index can be found in [90].

We assume that N sensors are deployed in an environment and that they are all able to measure a certain value θ, which is corrupted by noise, namely each sensor i obtains

$$y_i = \theta + n_i,\ i = 1,\ \dots,\ N.$$

We assume that n_i are independent random variables with zero mean and variance σ^2. In order to obtain an estimate of θ, we run a consensus setting $x_i(0) = y_i$ in order to obtain average of y_i's. However such an average can be obtained only asymptotically, while at tim t each sensor knows $x_i(t)$ which can be considered its of θ at time t. A natural performance index in this context is the time t estimation error.

While for the rate of convergence, the increase of the number of agents in general yields a performance degradation due to the decrease of the graph connectivity, intuitively we expect that this is not true for the estimation error. Indeed in this case a larger number of sensors should cause, on the one hand, a more difficult communication, but, on the other, a better estimate due to the presence of more measurements. In this section we try to correctly highlight this trade-off.

In order to do this, we take as measure of performance the variance of the difference between the estimate $\boldsymbol{x}(t)$ and the true value θ, normalized over the number of agents, namely we want to compute

$$J_{H_2}(P,\, t) = \frac{1}{N}\mathbb{E}\left[\boldsymbol{e}(t)^T\boldsymbol{e}(t)\right] \tag{24}$$

where $\boldsymbol{e}(t) = \boldsymbol{x}(t) - \theta\mathbf{1}$. Observe that, since P in the consensus algorithm is symmetric and stochastic, then $\mathbb{E}[\boldsymbol{e}(t)] = \theta$, namely the estimator is unbiased.

It turns out that

$$J_{H_2}(P,\, t) = \frac{\sigma^2}{N}\operatorname{Tr} P^{2t} = \frac{\sigma^2}{N}\sum_{\lambda\in\Lambda(P)}\lambda^{2t}.$$

Example: de Bruijn's graphs

In the case of de Bruijn's graphs with $N = n^k$, we have

$$J_{H_2}(P,\, t) = \begin{cases} \frac{1}{n^t}, & 0 \le t < k \\ \frac{1}{N}, & t \ge k \end{cases} \tag{25}$$

so the finite time convergence of the algorithm to the steady state can be seen also in such performance index. Notice that $J_{H_2}(P,\, t) \ge \frac{1}{N}$ for any P and any

t, so that de Bruijn's graphs are optimal for this index as they are concerning the rate of convergence.

Notice that, differently from the rate of convergence, this performance index depends on all the eigenvalues of P. No bounds for this performance index are known for general matrices P. Interesting bounds instead have been found in the special case in which P is a Cayley matrix.

6.1 Bounds on the time-dependent H_2 performance index for Cayley matrices

Before proposing the general result, we prefer to present a simple example. Consider the following consensus matrix

$$P = \begin{bmatrix} 1/3 & 1/3 & 0 & \cdots & \cdots & 1/3 \\ 1/3 & 1/3 & 1/3 & \cdots & \cdots & 0 \\ 0 & 1/3 & 1/3 & 1/3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1/3 & 1/3 & 1/3 \\ 1/3 & \cdots & \cdots & 0 & 1/3 & 1/3 \end{bmatrix} \tag{26}$$

whose associate graph is shown in Fig. 1. The essential spectral radius of P is $\rho(P) = 1 - C/N^2$. This shows that, as N grows, the convergence of the algorithm tends to be very slow. Nonetheless we expect that, in case of distributed estimation, the presence of more sensors should instead improve performance. Figure 5 depicts $J_{H_2}(P, t)$ as a function of t, for various values of N.

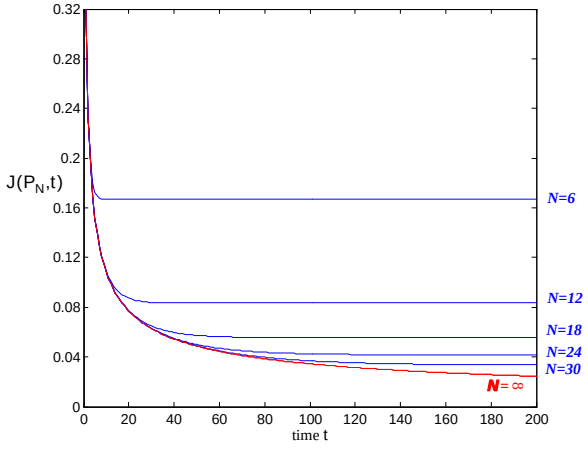

Figure 5: Time–dependent H_2 performance index for a Cayley matrix.

For any fixed N, we have evolutions which exponentially converge (with rate $\sim 1-C/N^2$) to the constant value $1/N$. The different curves become lower as N grows, and their envelope, which corresponds to the limit for $N \to \infty$, converges to zero for $t \to \infty$.

The general bounds we will give in the theorem below show that indeed the asymptotic behavior of $J_{H_2}(P,\, t)$ in this example is given by $\max\left\{\frac{1}{N}, \frac{1}{\sqrt{t}}\right\}$. In particular, for $N = \infty$, we have that $J_{H_2}(P,\, t)$ converges to zero as $1/\sqrt{t}$. This result shows that increasing N does not have the disadvantages predicted by observing that the essential spectral radius of P tends to 1 as N tends to infinity. Nevertheless, a further look at Figure 5, gives a caveat against the choice of too large values of N. Indeed, when the number of iterations t is finite, there is a bound on the number of nodes being truly useful, after which there is no improvement in adding new nodes. This is intuitive, as at time t there is no way for a node to use information coming from other agents further than t steps apart.

We give now the general result which has been proposed in [90].

Consider the class of Cayley matrices with respect to the group $\mathbb{Z}_n^d$ with given d and n. Assume that the associated d-dimensional tori have the self loops and that they are connected. Recall from section Sect. 2.1 the definition of l and δ.

Theorem 6.1. *Let P be a symmetric stochastic matrix whose associate graph $\mathcal{G}_P$ is in the previous class of d-dimensional tori characterized by the parameters l and δ as described above. Assume that all the nonzero entries of P lie in an interval $[p_{\min},\, p_{\max}]$. Then*

$$C' \max\left\{\frac{1}{N},\, \frac{1}{t^{d/2}}\right\} \leq J_{H_2}(P,\, t) \leq C'' \max\left\{\frac{1}{N},\, \frac{1}{t^{d/2}}\right\}, \tag{27}$$

where C', C'' are a strictly positive constant depending on d, l, δ and $p_{\min}$.

7 H_2 performance index

The performance index we are going to present in this section arises in the analysis of the consensus algorithm corrupted with additive noise [91, 92, 57]. Assume that the consensus iteration is corrupted by additive noise as follows

$$\boldsymbol{x}(t+1) = P\boldsymbol{x}(t) + \boldsymbol{w}(t),$$

where $\boldsymbol{w}(t)$ is an i.i.d. process with zero mean and variance $\mathbb{E}[\boldsymbol{w}(s)\boldsymbol{w}(t)^T] = R\delta(s-t)$, where $\delta(t)$ is the Kroenecker delta function and R is a positive semidefinite matrix. For sake of simplicity, assume $R = I$. Assume moreover $\boldsymbol{x}(0)$ and $\boldsymbol{w}(s)$ are uncorrelated for all $s \geq 0$. Define the dispersion of $\boldsymbol{x}(t)$ around its center of mass as

$$\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \boldsymbol{x}_A(t) := \boldsymbol{x}(t) - \frac{1}{N}\mathbf{1}\mathbf{1}^T\boldsymbol{x}(t) = \Omega\boldsymbol{x}(t)$$

where $\Omega = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T$. The vector $\tilde{\boldsymbol{x}}(t)$ describes the distance from consensus. In this case a reasonable performance metrics is given by

$$J_{H_2}(P) = \frac{1}{N}\limsup_{t\to\infty} \mathbb{E}[||\tilde{\boldsymbol{x}}(t)||^2],$$

Observe that

$$\tilde{\boldsymbol{x}}(t) = P^t\tilde{\boldsymbol{x}}(0) + \sum_{i=0}^{t-1} P^{t-1-i}\Omega\boldsymbol{w}(i)$$

and thus, if $\boldsymbol{x}(0)$ is zero mean, $\mathbb{E}[\tilde{\boldsymbol{x}}(t)] = 0, \forall\, t \geq 0$. Denote by $\Sigma(t) = \mathbb{E}[\tilde{\boldsymbol{x}}(t)\tilde{\boldsymbol{x}}(t)^T]$. Notice that $J_{H_2}(P) = \operatorname{Tr}\Sigma(\infty)$. The matrix $\Sigma(t)$ can be computed as follows

$$\Sigma(t) = P^t\Sigma(0)(P^T)^t + \sum_{i=0}^{t-1} P^i\Omega\Omega^T(P^T)^i.$$

Since $\Omega\Omega^T = \Omega$ and since $P^t\tilde{\boldsymbol{x}}(0)$ converges to zero, we can argue that

$$\Sigma(\infty) = \sum_{t\geq 0} P^t\Omega(P^T)^t.$$

and hence

$$J_{H_2}(P) = \frac{1}{N}\operatorname{Tr}\sum_{t\geq 0} P^t\Omega(P^T)^t. \tag{28}$$

Since P is symmetric, the previous performance index can be rewritten as

$$J_{H_2}(P) = \frac{1}{N}\sum_{\lambda\in\Lambda(P),\,\lambda\neq 1}\frac{1}{1-\lambda^2}. \tag{29}$$

This index describes a different performance feature of the standard consensus algorithm. In classic control a way to describe the transient performance of a control strategy is through the position of the dominant eigenvalues. This is analogous of taking the essential spectral radius of P as a performance index for the consensus algorithm. In control however there are other indices used for the transient evaluation, such as, for instance, the L_2 norm of the tracking error trajectory. We can do something similar for the consensus algorithm, namely we can take the usual consensus iteration (1), and take the index

$$\frac{1}{N}\sum_{t\geq 0}\mathbb{E}||\boldsymbol{x}(t) - \boldsymbol{x}(\infty)||_2^2 = \boldsymbol{x}(0)^T\left(\frac{1}{N}\sum_{t\geq 0}P^t\Omega(P^T)^t\right)\boldsymbol{x}(0),$$

If we assume that the initial state $\boldsymbol{x}(0)$ is a random vector with $\mathbb{E}[\boldsymbol{x}(0)\boldsymbol{x}(0)^T] = I$, then the previous index is equal to $J_{H_2}(P)$.

7.1 Bounds on the H_2 performance index for Cayley matrices

Concerning d-dimensional tori, from [92] we know the following interesting result. Consider the class of Cayley matrices with respect to the group $\mathbb{Z}_n^d$ with given d and n. Assume that the associated d-dimensional tori have the self loops and that they are connected. Recall from section Sect. 2.1 the definition of l and δ, and define

$$f(N,\, d) := \begin{cases} N, & d = 1 \\ \log N & d = 2 \\ 1, & d \geq 3 \end{cases}.$$

Theorem 7.1. *Let P be a symmetric stochastic matrix whose associate graph $\mathcal{G}_P$ is in the previous class of d-dimensional tori characterized by the parameters l and δ as described above. Assume that all the nonzero entries of P lie in an interval $[p_{\min}, p_{\max}]$. Then*

$$C' f(N, d) \leq J_{H_2}(P) \leq C'' f(N, d) \tag{30}$$

where C', C'' is a constant depending on d and l, on δ and on $p_{\min}$.

7.2 Electric analogy and bounds on the H_2 performance index for geometric graphs

Recall from Section 3 that there is a relation between a symmetric stochastic matrix and a suitable electrical network. In particular, by exploiting the relation between the effective resistance and the Green matrix of the electric network, we obtain the following theorem. For the details and the proof of the theorem see [93]. A similar result, applied to a different problem, can be found also in [94].

Theorem 7.2. *Given a stochastic, symmetric, primitive matrix P, then*

$$J_{H_2}(P) = \overline{\mathcal{R}}(\mathcal{G}_{P^2}, P^2) := \frac{1}{N^2} \sum_{u \neq v} \mathcal{R}_{uv}(\mathcal{G}_{P^2}, P^2). \tag{31}$$

Namely, $J_{H_2}(P)$ is the average of the effective resistances in a network build from P^2.

The result of this theorem can be significantly simplified using the following argument. Assume that all the nonzero entries of P lie in an interval $[p_{\min}, p_{\max}]$ and that the maximum degree of the agents in the network is d_g. By exploiting the properties of the effective resistance (see [65, 73]) it is possible to bound from above and from below the average effective resistance $\overline{\mathcal{R}}(\mathcal{G}_{P^2}, P^2)$ in the following way

$$c_1 \overline{\mathcal{R}}(\mathcal{G}_P) \leq \overline{\mathcal{R}}(\mathcal{G}_{P^2}, P^2) \leq c_2 \overline{\mathcal{R}}(\mathcal{G}_P),$$

whence the performance index is bounded by

$$c_1 \overline{\mathcal{R}}(\mathcal{G}_P) \leq J_{H_2}(P) \leq c_2 \overline{\mathcal{R}}(\mathcal{G}_P). \tag{32}$$

In these inequalities c_1 and c_2 denote two strictly positive constants depending only on $p_{\min}$, $p_{\max}$ and d_g, and $\overline{\mathcal{R}}(\mathcal{G}_P)$ denotes the average effective resistance in an electrical network with graph $\mathcal{G}_P$ and having all resistances set to 1 Ohm. This result is particulary interesting because we can focus our attention to the graph $\mathcal{G}_P$, rather than on the particular matrix P. Indeed, we can achieve in this way the following result, in which again

$$f(N, d) = \begin{cases} N, & d = 1 \\ \log N & d = 2 \\ 1, & d \geq 3 \end{cases}.$$

Theorem 7.3. *Let P be a symmetric stochastic matrix whose associate graph $\mathcal{G}_P$ is a geometric graph with parameters $(N, d, \ell, s, r, \gamma, \rho)$. We assume moreover that all the nonzero entries of P belongs to $[p_{\min}, p_{\max}]$. Then*

$$C' f(N,\, d) \leq J_{H_2}(P) \leq C'' f(N,\, d) \tag{33}$$

where C' and C'' are two constants depending on d, on the parameters s, r, γ, ρ, d_g, and on $p_{\min}$, and C'' depending on $p_{\max}$ too.

8 Conclusions

This paper is devoted to gather many results developed by the scientific community in the last years concerning the evaluation of the performances of consensus algorithms. It was shown that many performance indices can be defined in relation to the specific applications in which the consensus algorithm is employed. Moreover, it is investigated the influence of the communication topology on these performance indices. Geometric graph topologies are considered both in the general case and in the more structured case in which the graphs satisfy some symmetries, namely for d-dimensional tori. While for some performance indices, there is a quite complete understanding, for others the research is still at its first steps and many important questions remain still open, especially in understanding the influence of more general classes of graph topologies.

A Proofs of Theorem 5.1

The proof of this theorem is very difficult and to present it we need to introduce some notation and some lemmas. Given a connected graph $\mathcal{G}$ let

$$\rho(\mathcal{G}) := \min\{\rho(P) :\ P \text{ primitive symmetric stochastic matrix consistent with } \mathcal{G}\}$$

namely $\rho(\mathcal{G})$ denotes the minimum $\rho(P)$ when P varies in the set of symmetric, stochastic, primitive matrices consistent with $\mathcal{G}$. Notice that the matrix attaining such minimum might be not unique. It is clear that lower bound in the thesis of the theorem is proved if we prove that

$$\rho(\mathcal{G}) \geq 1 - \frac{C}{N^{2/d}}. \tag{34}$$

Now we give a lemma which will be useful in the proof of the theorem.

Lemma A.1. *Let $P, M \in \mathbb{R}^{N\times N}$ be symmetric stochastic matrices and assume that P is primitive. Then*

$$\rho(MPM) \leq \rho(P).$$

Proof. Notice first that

$$\rho(MPM) = \max\left\{\frac{|y^T MPMy|}{y^T y},\, y \perp \mathbf{1},\, y \neq 0\right\}$$

$$\rho(P) = \max\left\{\frac{|x^T Px|}{x^T x},\, x \perp \mathbf{1},\, x \neq 0\right\}$$

To prove that $\rho(MPM) \leq \rho(P)$ it is sufficient to prove that for any $y \neq 0$ such that $\mathbf{1}^T y = 0$, there exists $x \neq 0$ such that $\mathbf{1}^T x = 0$ such that

$$\frac{|y^T MPMy|}{y^T y} \leq \frac{|x^T Px|}{x^T x}.$$

In finding x we distinguish two cases:

1. In case $My \neq 0$, we let $x := My$. Notice that in this case we have that $x \neq 0$ and that $\mathbf{1}^T x = \mathbf{1}^T My = \mathbf{1}^T y = 0$. Notice finally that, since all the eigenvalues of M are in $[-1, 1]$, then $y^T MMy \leq y^T y$ and so

$$\frac{|y^T MPMy|}{y^T y} \leq \frac{|y^T MPMy|}{y^T M^2 y} = \frac{|x^T Px|}{x^T x}.$$

2. In case $My = 0$, we let x to be any nonzero vector such that $\mathbf{1}^T x = 0$.

□

Proof of Theorem 5.1

Let $h := \lfloor \ell/2\gamma \rfloor$ and tessellate the hypercube Q into $H := h^d$ identical hypercubes with edge of length $\bar{\ell} := \ell/h$. Notice that, since $\ell/h \geq 2\gamma$ then each of such hypercubes contains at least one node of $\mathcal{G}$. This implies that $h^d \leq N$. Notice moreover that, from the assumption that $\ell \geq 4\gamma$, we can argue that

$$\bar{\ell} = \frac{\ell}{\lfloor \ell/2\gamma \rfloor} \leq \frac{\ell}{\ell/2\gamma - 1} \leq 4\gamma$$

We can define now a graph $\mathcal{G}_{\mathcal{L}} = (\mathcal{Q}, \mathcal{L})$ having all hypercubes as nodes and having as edges all the pairs (q', q'') of hypercubes such that, either $q' = q''$ (self-loops), or there exists $(u, v) \in \mathcal{E}$ with $u \in q'$ and $v \in q''$. Notice that $\mathcal{G}_{\mathcal{L}}$ is a subgraph of a d-dimensional torus over the group $\mathbb{Z}_n^d$. Notice moreover that if two nodes x, $y \in \mathbb{Z}_n^d$ are connected, then there exists $(u, v) \in \mathcal{E}$ with $u \in q_x$ and $v \in q_y$, with q_x and q_y the hypercubes correspondent to x and y respectively. Identify now $x \in \mathbb{Z}_n^d$ with the center of its hypercube q_x, and the same for y. We can argue that

$$||x - y||\bar{\ell} = d_E(x, y) \leq d_E(x, u) + d_E(u, v) + d_E(v, y) \leq r + \bar{\ell}\sqrt{d}$$

From this we can argue that $x, y \in \mathbb{Z}_n^d$ are connected only if $||x - y|| \leq \delta$ where $\delta = r/2\gamma + \sqrt{d}$.

Build now a new graph, $\tilde{\mathcal{G}} = (V, \tilde{\mathcal{E}})$, which has the same set of nodes as $\mathcal{G}$, and has an edge connecting any couple of nodes u, v if and only if one of the following two conditions hold:

1. u, v lie in the same hypercube,
2. $u \in q'$ and $v \in q''$ and $(q', q'') \in \mathcal{L}$.

Notice that, by construction, we have that $\mathcal{E} \subseteq \tilde{\mathcal{E}}$, namely $\mathcal{G}$ is embedded in $\tilde{\mathcal{G}}$ so that

$$\rho(\tilde{\mathcal{G}}) \leq \rho(\mathcal{G})$$

since minimization in $\tilde{\mathcal{G}}$ is subjected to a smaller set of constraints. Our aim is now to lower bound $\rho(\tilde{\mathcal{G}})$, since this will yield in turn a lower bound for $\rho(\mathcal{G})$ as well.

Let $n_i \geq 1$, $i = 1, 2, \ldots, H$, be the number of nodes inside the i-th hypercube so that $\sum_{i=1}^{H} n_i = N$. Assume that the nodes of $\tilde{\mathcal{G}}$ are ordered in such a way that the nodes $n_{i-1}+1, n_{i-1}+2, \ldots, n_i$ end in the i-th hypercube. Introduce the following matrices, $D := \mathrm{diag}\,(n_1, \ldots, n_H)$ and

$$S := \begin{bmatrix} \mathbf{1}_{n_1} & 0 & \ldots & 0 & 0 \\ 0 & \mathbf{1}_{n_2} & \ldots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \ldots & 0 & \mathbf{1}_{n_H} \end{bmatrix} \in \{0,\, 1\}^{N \times H}$$

Notice that $M := SD^{-1}S^T$ is stochastic. Assume now that $P^{\tilde{\mathcal{G}}}$ be a symmetric, stochastic, primitive matrix consistent with $\tilde{\mathcal{G}}$ such that $\rho(P^{\tilde{\mathcal{G}}}) = \rho(\tilde{\mathcal{G}})$. By the definition of M it is easy to see that

$$MP^{\tilde{\mathcal{G}}}M = \begin{bmatrix} p_{11}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^T & p_{12}\mathbf{1}_{n_1}\mathbf{1}_{n_2}^T & \ldots & p_{1H}\mathbf{1}_{n_1}\mathbf{1}_{n_H}^T \\ p_{12}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^T & p_{22}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T & \ldots & p_{2H}\mathbf{1}_{n_2}\mathbf{1}_{n_H}^T \\ \vdots & & & \vdots \\ p_{1H}\mathbf{1}_{n_H}\mathbf{1}_{n_1}^T & p_{2H}\mathbf{1}_{n_H}\mathbf{1}_{n_2}^T & \ldots & p_{HH}\mathbf{1}_{n_H}\mathbf{1}_{n_H}^T \end{bmatrix}.$$

where p_{ij} are the elements of the matrix $\tilde{P} = D^{-1}S^T P^{\tilde{\mathcal{G}}} S D^{-1}$. By definition of $\tilde{\mathcal{G}}$ we have that $MP^{\tilde{\mathcal{G}}}M$ is compatible with $\tilde{\mathcal{G}}$ and so $\rho(MP^{\tilde{\mathcal{G}}}M) \geq \rho(\tilde{\mathcal{G}}) = \rho(P^{\tilde{\mathcal{G}}})$. On the other hand, by lemma A.1 we have that $\rho(MP^{\tilde{\mathcal{G}}}M) \leq \rho(P^{\tilde{\mathcal{G}}})$, proving in this way that $\rho(MP^{\tilde{\mathcal{G}}}M) = \rho(\tilde{\mathcal{G}})$.

Let $\bar{P} := D^{-1}S^T P^{\tilde{\mathcal{G}}} S = \tilde{P} D$. This is a stochastic matrix. It can be shown that

$$\Lambda(MP^{\tilde{\mathcal{G}}}M) \setminus \{0\} \subseteq \Lambda(\bar{P}) \subseteq \Lambda(MP^{\tilde{\mathcal{G}}}M) \tag{35}$$

To prove (35) first observe that, if $\mu \in \Lambda(\bar{P})$ then there exists a nonzero vector v such that $\bar{P}v = \mu v$. If we let $\tilde{v} := Sv \neq 0$, then

$$MP^{\tilde{\mathcal{G}}}M\tilde{v} = SD^{-1}S^T P^{\tilde{\mathcal{G}}} S D^{-1} S^T S v = S\bar{P}v = S\mu v = \mu\tilde{v},$$

and so $\mu \in \Lambda(MP^{\tilde{\mathcal{G}}}M)$. Assume conversely that $\mu \in \Lambda(MP^{\tilde{\mathcal{G}}}M) \setminus \{0\}$. Then $MP^{\tilde{\mathcal{G}}}M\tilde{v} = \mu\tilde{v}$ for some nonzero vector $\tilde{v}$. Let $v := S^T\tilde{v}$. Observe that v is nonzero because otherwise we would have that $\mu = 0$. Observe finally that

$$\mu v = \mu S^T\tilde{v} = S^T MP^{\tilde{\mathcal{G}}}M\tilde{v} = S^T S D^{-1} S^T P^{\tilde{\mathcal{G}}} S D^{-1} S^T \tilde{v} = \bar{P}v$$

This implies that $\mu \in \Lambda(\bar{P})$.

Notice that (35) implies that $\rho(\bar{P}) \leq \rho(\tilde{\mathcal{G}}) \leq \rho(\mathcal{G}) \leq \rho(P)$. It remains to find a lower bound of $\rho(\bar{P})$. Notice first that, if we define $\Pi := N^{-1}D$, then we have that

$$\bar{P}^T\Pi = \Pi\bar{P}$$

This means that $\bar{P}$ is a reversible stochastic matrix with respect to D. Notice moreover that $\bar{P}$ is compatible with the graph $\mathcal{G}_{\mathcal{L}}$ which is a subgraph of a d-dimensional torus with respect to the group $\mathbb{Z}_h^d$ in which two nodes $x, y \in \mathbb{Z}_n^d$

are connected only if $||x - y|| \leq \delta$, where $\delta = r/2\gamma + \sqrt{d}$. Therefore we can apply Theorem 4.6 to obtain

$$\rho(\bar{P}) \geq 1 - \left(\frac{\pi_{\max}}{\pi_{\min}}\right)^2 \frac{C}{h^2}.$$

where C depends on d, γ and r.

Recall finally that, since each hypercube contains at least a node, then $h^d \geq N$. Notice moreover that $\pi_{\max}/\pi_{\min} \leq \max_i\{n_i\}$. Observe that the volume of the d-dimensional sphere of radius $s/2$ is $A_d \left(\frac{s}{2}\right)^d$, where $A_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$. Then, by definition of s, we have that

$$n_i A_d \left(\frac{s}{2}\right)^d \leq \bar{\ell}^d$$

and so

$$n_i \leq \frac{1}{A_d}\left(\frac{2\bar{\ell}}{s}\right)^d \leq \frac{1}{A_d}\left(\frac{8\gamma}{s}\right)^d$$

where we used the fact that $\bar{\ell} \leq 4\gamma$. We can conclude that

$$\rho(\mathcal{G}) = \rho(\tilde{\mathcal{G}}) = \rho(\bar{P}) \geq 1 - \frac{1}{A_d}\left(\frac{8\gamma}{s}\right)^d \frac{C}{N^{2/d}}.$$

B Proofs of Theorem 5.2

For simplicity we will develop the proof only for $d = 2$, showing at the end how it is possible to generalize the result to any dimension.

For any pair of nodes x, y we build a path γ_{xy} connecting x and y as follows (see Fig. 6). We first tessellate the square Q as we did in the previous proof, namely by letting $h := \lfloor \ell/2\gamma \rfloor$ and by tessellating Q into h^2 identical squares with edge of length $\bar{\ell} := \ell/h$. Notice that, since $\ell/h \geq 2\gamma$ then each of such squares contains at least one node of $\mathcal{G}_P$. As in the previous proof, we have that $h^2 \leq N$ and that $\bar{\ell} \leq 4\gamma$.

Assume that, for any square q, we choose one of its nodes, denoted by u_q, as its representative node in such a square. Moreover, for any node v let u_v be the representative of the square x belongs to.

Take the two nodes x, y we want to connect. Assume that x is below with respect to y. Link, via the shortest path, x to u_x, the representative of the square q_x. Then, start proceeding upward, linking via the shortest path the representatives of each of the neighbor squares till we reach a square which is at the same height of the square q_y containing y. Then, proceed left or right with in a similar way. At a certain point, we will reach the representative u_y of q_y, and then we will link u_y with y via the shortest path.

Let Γ be the collection of these paths, one for each pair x, y. We want to give an estimate of $\kappa(\Gamma)$, where $\kappa(\Gamma)$ is defined (18). By definition of $|\gamma_{xy}|_P$, given (7), we have that

$$|\gamma_{xy}|_P = \sum_{e \in \gamma_{xy}} \frac{1}{P_e} \leq \frac{1}{p_{\min}} |\gamma_{xy}|,$$

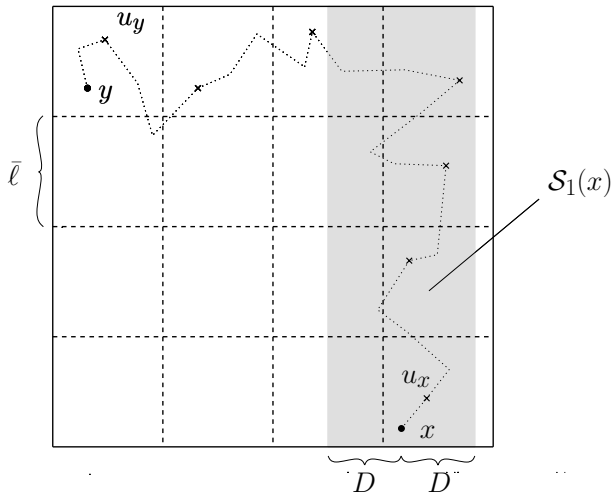

Figure 6: Building γ_{xy}. For each square, we choose a representative, which is marked by a cross. We proceed from representative to representative upwards and then leftwards (or rightwards). We can exit from the squares, but not from stripes of width $2D$.

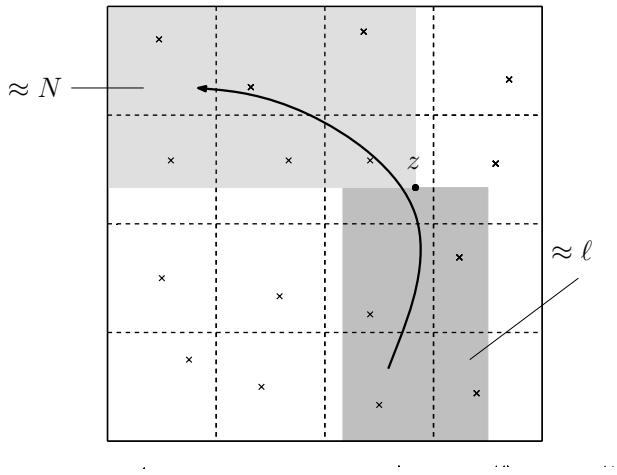

Figure 7: Maximum usage of a node. Given z, we can use it if we start from a node in the dark grey below it, which is a subset of $\mathcal{S}_1(z)$. The number of nodes in is region is proportional to $\ell \approx N^{1/d}$. Once we used z, we can possibly reach any node in the soft grey region. The number of nodes in this region is, in the worst case, proportional to the number of nodes in the network, N.

where $|\gamma_{xy}|$ is the length of the path γ_{xy}, namely the cardinality of $\{e \in \mathcal{E} : e \in \gamma_{xy}\}$. We want to upper bound $|\gamma_{xy}|$. We will start by bounding the graphical distance between two representatives u and v of neighbor squares. Observe that $d_E(u,v) \leq \sqrt{5}\bar{\ell} \leq 4\sqrt{5}\gamma$, and so

$$d_G(u,v) \leq \frac{d_E(u,v)}{\rho} \leq 4\sqrt{5}\frac{\gamma}{\rho}$$

Through similar arguments it can be proved that

$$d_G(x,u_x) \leq \frac{d_E(x,u_x)}{\rho} \leq 4\sqrt{2}\frac{\gamma}{\rho}, \qquad d_G(y,u_y) \leq \frac{d_E(y,u_y)}{\rho} \leq 4\sqrt{2}\frac{\gamma}{\rho}$$

Now, observe that a path in Γ touches at most $2h$ squares, so including the starting and the ending points, we have that

$$|\gamma_{xy}| \leq 8\sqrt{5}\frac{\gamma}{\rho}h + 8\sqrt{2}\frac{\gamma}{\rho} \leq 8(\sqrt{5}+\sqrt{2})\frac{\gamma}{\rho}h \tag{36}$$

Now, given an edge e, we have that

$$\sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_P \leq 8(\sqrt{5}+\sqrt{2})\frac{\gamma}{p_{\min}\rho}h|\{\gamma_{xy}|\gamma_{xy} \ni e\}|. \tag{37}$$

Hence, it remains to bound the maximum number of paths which cross a certain edge. To do this we will bound the number of paths which cross a certain node (see Fig. 7). Assume that a node z belongs to γ_{xy}. We want to understand where z has to be with respect to x, y. Let

$$D := 4\sqrt{5}\frac{r\gamma}{\rho} + 4\gamma \geq \sqrt{5}\frac{r\bar{\ell}}{\rho} + \bar{\ell}$$

which is a constant depending only on γ, ρ and r. We show now that

$$z \in \mathcal{S}_1(x) \cup \mathcal{S}_2(y) \tag{38}$$

where for any point $w \in Q$ we define

$$\mathcal{S}_1(w) := \{w' \in Q \ : \ |w_1 - w'_1| \leq D\}, \qquad \mathcal{S}_2(w) := \{w' \in Q \ : \ |w_2 - w'_2| \leq D\}$$

and where w_1 and w_2 denote the two coordinates of the point $w \in Q$. Namely, $S_1(w)$ is a "vertical stripe" centered in w and of width $2D$, $S_2(w)$ is the horizontal analogous.

We distinguish various cases:

1. If z belongs to the shortest path between x and the representative u_x of the square x belongs to, then since the path from x to u_x is the shortest possible, $d_G(x,z) \leq d_G(x,u_x) \leq \sqrt{2}\bar{\ell}/\rho$. This implies

$$d_E(x,z) \leq r d_G(x,z) \leq \sqrt{2}\frac{r\bar{\ell}}{\rho} \leq D$$

and hence $|x_1 - z_1| \leq d_E(x,z) \leq D$.

2. In an analogous way it can be shown that, in case z belongs to the shortest path between y and u_y, then $|y_2 - z_2| \leq D$.

3. If z belongs to the shortest path between u and v, which are representatives of neighbor squares in the vertical portion of the path from x, then $|x_1 - u_1| \leq \bar{\ell}$ and moreover, since $d_G(u,z) \leq d_G(u,v) \leq \sqrt{5}\bar{\ell}/\rho$, we can argue that
$$d_E(u,z) \leq r d_G(u,z) \leq \sqrt{5}\frac{r\bar{\ell}}{\rho}$$
and so $|x_1 - z_1| \leq |x_1 - u_1| + |u_1 - z_1| \leq \bar{\ell} + d_E(u,z) \leq D$.

4. In an analogous way it can be shown that, in case z belongs to the shortest path between u and v, which are representatives of neighbor squares in the horizontal portion of the path to y, we have that $|y_2 - z_2| \leq D$.

Now we are able to bound the number of paths γ_{xy} which cross z. Indeed, if γ_{xy} crosses z then (38) holds and so we can distinguish to cases:

(i) Assume that $z \in \mathcal{S}_1(x)$. This holds if and only if $x \in \mathcal{S}_1(z)$. Then the number of γ_{xy} such that $x \in \mathcal{S}_1(z)$ is upper bounded by
$$\frac{8}{\pi}\frac{D\ell}{s^2}N.$$
In fact, such value is less than or equal to the number of nodes in the stripe $\mathcal{S}_1(z)$ multiplied by the total number of nodes N. The first quantity can be upper bounded as before by the area of the stripe, $2D\ell$, over the area of a sphere of radius $\frac{s}{2}$.

(ii) Assume in this case that $z \notin \mathcal{S}_1(x)$. Then the number of γ_{xy} such that $x \notin \mathcal{S}_1(z)$ can be bounded as follows. Indeed observe from (38) that, if $x \notin \mathcal{S}_1(z)$, then $y \in \mathcal{S}_2(z)$, and so the number of γ_{xy} such that $x \notin \mathcal{S}_1(z)$ is less than or equal to the number of nodes in $\mathcal{S}_2(z)$ multiplied by the total number of nodes N. This value is upper bounded by the same number above.

Putting together the two cases, the number of paths γ_{xy} which cross z is upper bounded by
$$\frac{8}{\pi}\frac{D\ell}{s^2}N$$
Finally, by considering this last bound, with the bounds (37) and (36) we obtain that
$$\kappa(\Gamma) \leq 8(\sqrt{5}+\sqrt{2})\frac{\gamma}{p_{\min}\rho}h\frac{8}{\pi}\frac{D\ell}{s^2}N \leq \frac{64(\sqrt{5}+\sqrt{2})\gamma D}{\pi p_{\min}\rho s^2}4\gamma N^2.$$
where we used the fact that $\ell = \bar{\ell}h \leq 4\gamma h$ and the fact that $h^2 \leq N$.

To conclude, by exploiting the Poicaré inequality, we have
$$\lambda_1 \leq 1 - \frac{N}{\kappa(\Gamma)} \leq 1 - C\frac{1}{N},$$
where C is a constant depending on the geometric parameters s, γ, r and ρ and on $p_{\min}$.

In the general case, the whole reasoning still holds. What differs is the value of many numerical constants, since for example the Euclidean distance among the representatives of two neighbor hypercubes is in general $d_E(u, v) \leq \sqrt{3+d}$. The most important difference lies however in the fact that now $h^d \leq N$, whence $\ell \leq 4\gamma N^{1/d}$. This yields

$$\kappa(\Gamma) \leq C' h \ell N \leq C N^{1+2/d},$$

and thus Poincaré inequality implies

$$\lambda_1 \leq 1 - C\frac{N}{N^{1+2/d}} = 1 - C\frac{1}{N^{2/d}}.$$

References

[1] J. Cortes, S. Martinez, and F. Bullo. Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions. *Automatic Control, IEEE Transactions on*, 51(8):1289–1298, 2006.

[2] F. Bullo, J. Cortes, and S. Martinez. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009.

[3] J.A. Fax and R.M. Murray. Information flow and cooperative control of vehicle formations. *Automatic Control, IEEE Transactions on*, 49(9):1465 – 1476, sept. 2004.

[4] A. Jadbabaie, Jie Lin, and A.S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *Automatic Control, IEEE Transactions on*, 48(6):998 – 1001, june 2003.

[5] Wei Ren, R.W. Beard, and E. M. Atkins. Information conesnsus in multivehicle cooperative control. *Control System Magazine, IEEE*, 27(2):71–82, may 2005.

[6] H. G. Tanner, A. Jadbabaie, and G.J. Pappas. Flocking in fixed and switching networks. *Automatic Control, IEEE Transaction on*, 52(5):863–868, may 2005.

[7] M. Franceschelli, M. Egerstedt, A. Giua, and C. Mahulea. Constrained invariant motions for networked multi-agent systems. In *Proc of. American Control Conference 2009. ACC '09*, pages 5749–5754, june 2009.

[8] R. Olfati-Saber. Distributed kalman filter with embedded consensus filters. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*, pages 8179 – 8184, 12-15 2005.

[9] R. Olfati-Saber. Distributed kalman filtering for sensor networks. In *Proc. of 46th Conference on Decision and control. CDC'07*, dec. 2007.

[10] A. Speranzon, C. Fischione, and K. H. Johansson. Distributed and collaborative estimation over wireless sensor networks. In *Proc. of 45th Conference on Decision and control. CDC'06*, pages 1025–1030, dec. 2006.

[11] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proc. of 4th International Symposium on Information Processing in Sensor Networks. IPSN '05*, pages 63–70, april 2005.

[12] D. P. Spanos, R. Olfati-Saber, and R. M. Murray. Approximate distributed kalman filtering in sensor networks with quantifiable performance. In *Proc. of 4th International Symposium on Information Processing in Sensor Networks. IPSN '05*, pages 133–139, april 2005.

[13] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *J. Parallel Distrib. Comput.*, 7(2):279–301, 1989.

[14] J. E. Boillat. Load balancing and poisson equation in a graph. *Concurrency: Practice and Experience*, 7(4):289–313, may 1990.

[15] Saurabh Ganeriwal, Ram Kumar, and Mani B. Srivastava. Timing-sync protocol for sensor networks. In *Proc. of the 1st international conference on Embedded networked sensor systems (SenSys '03)*, pages 138–149, 2003.

[16] S. Bolognani, S. Del Favero, and L. Schenato. Distributed sensor calibration and least-square parameter identification in wsns using consensus algorithms. In *Proc. of 46th Annual Allerton Conference on Communication, Control and Computing.*, pages 1191–1198, sept. 2009.

[17] L. Schenato and G. Gamba. A distributed consensus protocol for clock synchronization in wireless sensor network. In *Proc. of 47th Conference on Decision and Control. CDC'07*, pages 2289–2294, dec. 2007.

[18] G. Xiong and S. Kishore. Discrete-time second-order distributed consensus time syhchronization algorithm for wireless sensor networks. *J. Wirel. Commun. Netw. EURASIP*, 1(1):1–12, jan. 2009.

[19] A. Fagiolini, S. Martini, and A. Bicchi. Set-valued consensus for distributed clock synchronization. In *Proc. of IEEE Conference on Automation Science and Engineering. CASE 2009*, pages 116–121, aug. 2009.

[20] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained consensus and optimization in multi-agent networks. *Automatic Control, IEEE Transactions on*, 55(4):922 –938, april 2010.

[21] A. Nedic and A. Ozdaglar. Distibuted subgradient methods for multiagent optimization. *Automatic Control, IEEE Transactions on*, 54(1):48–61, jan. 2009.

[22] Z. Minghui and S. Martinez. On distributed optimization under inequality and equality constraints via penalty primal-dual methods. In *Prof of. American Control Conference. ACC'10*, pages 2434–2439, july 2010.

[23] B. Johansson, M. Rabi, and M. Johansson. A simple peer-to-peer algorithm for distribured optimization in sensor networks. In *Prof of. 46th IEEE Conference on Decision and Control. CDC'07*, pages 4705–4710, dec. 2007.

[24] S. Sundhar Ram, A. Nedic, and V. V. Veeravalli. Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, 20(2):691–717, 2009.

[25] B. Johansson, C. M. Carretti, and M. Johansson. On distributed optimization using peer-to-peer communications in wireless sensor networks. In *Proc. of 5th Annual IEEE Communication Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks. SECON '08*, pages 497–505, june 2008.

[26] B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, 20(3):1157–1170, 2009.

[27] Hao Zhu, A. Cano, and G.B. Giannakis. Distributed demodulation using consensus averaging in wireless sensor networks. In *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pages 1170 –1174, 26-29 2008.

[28] Hao Zhu, A. Cano, and G.B. Giannakis. Distributed consensus-based demodulation: algorithms and error analysis. *IEEE Transactions on Wireless Communications*, 9(6):2044–2054, june 2010.

[29] F.R. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, 1959.

[30] J. Tsitsiklis. *Problems in decentralized decision making and computation*. PhD thesis, Department of EECs, MIT, 1984.

[31] R. Olfati-Saber and R.M. Murray. Consensus protocols for networks of dynamic agents. In *American Control Conference, 2003. Proceedings of the 2003*, volume 2, pages 951 – 956, june 2003.

[32] R. Olfati-Saber, J.A. Fax, and R.M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *Automatic Control, IEEE Transactions on*, 49(9):1520 – 1533, sept. 2004.

[33] R. Olfati-Saber, J.A. Fax, and R.M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215 – 233, jan. 2007.

[34] F. Fagnani and S. Zampieri. Average consensus with packet drop communication. *Siam Journal on Control and Optimization*, 48:102–133, 2009.

[35] P.-A. Bliman and G. Ferrari-Trecate. Average consensus problems in networks of agentws with delayed communications. In *Proc. of 44th IEEE Conference on Decision and Control and European Control Conference. CDC-ECC '05*, pages 7066–7071, dec. 2005.

[36] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor netowks with imperfect communcation: Link failures and channel noise. *IEEE Transactions on Signal Processing*, 57(1):355–369, jan. 2009.

[37] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks: Quantized data and random link failures. *IEEE Transactions on Signal Processing*, 58(3):1383–1400, march 2010.

[38] Wei Ren and R.W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *Automatic Control, IEEE Transactions on*, 50(5):655 – 661, may 2005.

[39] S. Boyd, A. Ghogh, B. Prabhakar, and D. Shah. Analysis and optimization of randomized gossip algorithms. In *Proc. of 43th IEEE Conference on Decision and Control. CDC'04*, volume 5, pages 5310–5315, dec. 2004.

[40] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *Information Theory, IEEE Transactions on*, 52(6):2508 – 2530, june 2006.

[41] S. Patterson, B. Bamieh, and A. El Abbadi. Distributed average consensus with stochastic communication failures. In *Proc. of 46th IEEE Conference on Decision and Control. CDC'07*, pages 4215–4220, dec. 2007.

[42] Y. Hatano and M. Mesbahi. Agreement over random networks. *IEEE Transactions on Automatic Control*, 50(11):1867–1872, nov. 2005.

[43] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Automatica*, 44:671–684, 2008.

[44] J.C. Delvenne, R. Carli, and S. Zampieri. Optimal strategies in the average consensus problem. *Systems & Control Letters*, 58:759–765, 2009.

[45] A. Olshevsky and J. N. Tsitsiklis. Convergence rates in distributed consensus and averaging. In *Proc. of 45th IEEE Conference on Decision and Control. CDC'06*, pages 3387–3392, dec. 2006.

[46] V. D. Blondel, J.M. Hedrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *Proc. of 44th IEEE Conference on Decision and Control and European Control Conference. CDC-ECC '05*, pages 2996–3000, dec. 2005.

[47] R. Olfati-Saber. Algebraic connectivity ratio of ramanujan graphs. In *Proc. of American Control Conference. ACC'07*, pages 4619–4624, july 2007.

[48] A. Tahbaz-Saleh and A. Jadbabaie. Small world phenomenon, rapidly mixing markov chains, and averaging consensus algorithms. In *Proc. of 46th IEEE Conference on Decision and Control. CDC'07*, pages 276–281, dec. 2007.

[49] R. Olfati-Saber. Ultrafast consensus in small-world networks. In *Proc. of American Control Conference. ACC'05*, pages 2371–2378, june 2005.

[50] S. Boyd, P. Diaconis, P. Parrillo, and L. Xaio. Fastest mixing markov chain on graphs with symmetries. *SIAM Journal on Optimization*, 20(2):792–819, 2009.

[51] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 52, 2004.

[52] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing markov chain on a graph. *Siam Review*, 2004.

[53] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, 2007.

[54] B. Johansson, A. Speranzon, M. Johansson, and K. Henrik. On decentralized negotiation of optimal consensus. *Automatica*, 44:1175–1179, april 2008.

[55] M. M. Zavlanos, D. E. Koditschek, and G. J. Pappas. A distributed dynamical scheme for fastest mixing markov chains. In *Proc. of American Control Conference. ACC'09*, pages 1436–1441, june 2009.

[56] K. Jung, D. Shah, and J. Shin. Distributed averaging via lifted markov chains. *IEEE Transactions on Information Theory*, 56(1):634–647, jan. 2010.

[57] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri. Average consensus on networks with quantized communication. *International Journal of Robust and Nonlinear Control*, 19:1787 – 1816, 2008.

[58] R. Rajagopal and M. Wainwright. Network-based consensus averaging with general noisy channels. *CoRR, abs/0805.0438*, 2008. on-line available.

[59] A. Kashyap, T. Basar, and R. Srikant. Quantized consensus. *Automatica*, 43:192–1203, 2007.

[60] J. Lavaei and R.M. Murray. On quantized consensus by means of gossip algorithm - part i: Convergence proof. In *American Control Conference, 2009. ACC '09.*, pages 394 –401, 10-12 2009.

[61] J. Lavaei and R.M. Murray. On quantized consensus by means of gossip algorithm - part ii: Convergence time. In *American Control Conference, 2009. ACC '09.*, pages 2958 –2965, 10-12 2009.

[62] M. Huang and J. H. Manton. Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. *SIAM Journal on Control and Optimization*, 48(1):134–161, 2009.

[63] P. Alriksson and A. Rantzer. Distributed kalman filtering using weighted averaging. In *Prooceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, 2006.

[64] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed kalman filtering based on consensus strategies. *Selected Areas in Communications, IEEE Journal on*, 26(4):622 –633, may 2008.

[65] P.G. Doyle and J.L. Snell. *Random Walks and Electric Networks*. Carus Monographs. Mathematical Association of America, 1984.

[66] P. Gupta and P.R. Kumar. The capacity of wireless networks. *Information Theory, IEEE Transactions on*, 46(2):388 –404, mar. 2000.

[67] C. Godsil and G. Royle. *Algebraic Graph Theory*. Graduate Texts in Mathematics. Springer, 2001.

[68] R. Carli. *Topics on the average consensus problem*. PhD thesis, PhD school in Information Engineering – University of Padua, 2008.

[69] M. Franceschetti and R. Meester. *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge Series in Statistical and Probabilistic Mathematics, 2007.

[70] J. Cortes, S. Martinez, and F. Bullo. Spatially-distributed coverage optimization and control with limited-range interactions. In *ESAIM. Control, Optimisation & Calculus of Variations*, pages 691–719, 2004.

[71] J.W. Jaromczyk and G.T. Toussaint. Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, 80(9):1502 –1517, sep 1992.

[72] Godfried T. Toussaint. Some unsolved problems on proximity graphs, 1991.

[73] P. Barooah and J. Hespanha. Estimation from relative measurements: Error bounds from electrical analogy. *ICISIP*, 2005.

[74] P. Barooah and J. Hespanha. Estimation from relative measurements: Electrical analogy and large graphs. *IEEE Transactions on Signal Processing*, 56(6):2181–2193, june 2008.

[75] P. Barooah and J. Hespanha. Error scaling laws for linear optimal estimtation from relative measurements. *IEEE Transactions on Information Theory*, 55(12):5661–5673, dec. 2009.

[76] M.R. Samatham and D.K. Pradhan. The de bruijn multiprocessor network: A versatile parallel processing and sorting network for vlsi. *IEEE Transactions on Computers*, 38(4):567–581, april 1984.

[77] M.R. Samatham. Augmented multiprocessor networks, July 1992. US Patent Number 5134690.

[78] N.G. de Bruijn. A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 46, 1946.

[79] F.Y. Wu. Theory of resistor networks: the two-point resistance. *J. Phys. A: Math. Gen.*, 2004.

[80] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

[81] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[82] F.R.K. Chung. *Spectral Graph Theory*, volume 92 of *Regional conference series in mathematics*. American Mathematical Society, 1997.

[83] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[84] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *The Annals of Probability*, 21(4):2131–2156, 1993.

[85] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible markov chains. *The Annals of Probability*, 3(3):696–730, 1993.

[86] J. Fulman and E.L. Wilmer. Comparing eigenvalue bounds for markov chains: When does poincare beat cheeger? *The Annals of Applied Probability*, 9(1):1–13, 1999.

[87] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.

[88] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[89] E. Behrends. *Introduction to Markov Chain (with special enphasys on rapid mixing)*. Vieweg Verlag, 1999.

[90] F. Garin and S. Zampieri. Performance of consensus algorithms in large-scale distributed estimation. In *European Control Conference*, 2009.

[91] B. Bamieh, M. Jovanovic, P. Mitra, and S. Patterson. Coherence in large-scale networks: Dimension dependent limitations of local feedback. *Submitted to Transactions on Automatic Control*, 2009.

[92] R. Carli, F. Garin, and S. Zampieri. Quadratic indices for the analysis of consensus algorithms. In *Proceedings of Information Theory and Applications Workshop*, pages 96–104, August 2009.

[93] E. Lovisari, F. Garin, and S. Zampieri. A resistance-based approach to consensus algorithm performance analysis. *Proc. of 49th IEEE Conference on Decision and Control. CDC'10*, pages 5714 –5719, dec. 2010.

[94] A. Ghosh, S. Boyd, and A. Saberi. Minimizing effective resistance of a graph. *SIAM review*, 50(1):37–66, 2008.