Performance metrics in the consensus problem: a Survey ⋆

Lovisari E. * **Zampieri S.** **

* *Dipartimento di Ingegneria dell'Informazione, Università di Padova, Italy, (e-mail: lovisari@dei.unipd.it)*
** *Dipartimento di Ingegneria dell'Informazione, Università di Padova, Italy, (e-mail: zampi@dei.unipd.it)*

Abstract
Linear consensus is a distributed algorithm which ensures that, under certain assumptions, a set of agents reach asymptotically the same opinion over a certain variable. Moreover this occurs with only a local exchange of information, namely the information exchange takes place only between agents which are neighbors in 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 performance to the communication topology. Different performance metrics may yield to different answers in comparing alternative communication topologies. In this paper, we present a number of possible performance metrics. Moreover, we show how these metrics are related to the communication topology. In particular, when available, we present bounds by which it is possible to relate performance and topology for general graphs, for graphs with symmetries, called Cayley graphs, and for geometric graphs.

1. INTRODUCTION

Distributed algorithms are a growing field of research in many scientific communities. The availability of an enormous number of small, simple and cheap agents forces researchers to enlarge their point of view from centralized, fast and fully designable procedures, to algorithms which should ensure a certain behavior without any leader, and in which information can flow only according to a constrained communication architecture. Interesting examples come from coordinated control Cortes et al. (2006). In this scenario some agents are required to move in a possibly unknown environment maintaining a formation suitable for their current objective. The absence of a leader can be a big advantage in this application. Indeed, if the environment is dangerous, we don't want the failure of the leader to yield the disruption of the entire formation, neither we want a possible, slow, procedure of election of a new leader. Instead, we want all the agents to be on (almost) the same hierarchical level. Moreover, no centralized control is desired, for many reasons: 1) if the environment is large and communication is slow, delays could well destabilize the system and 2) if the number of agents is large, the computational effort required to design the control could take too much time. This couple of requirements gives the name to 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, 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 different applications, for example distributed estimation Olfati-Saber (2005), load balancing Cybenko (1989), sensor calibration for sensor networks Ganeriwal et al. (2003), distributed optimization Nedic et al. (2010), distributed demodulation Zhu et al. (2008).

In general, in all these problems it is given a set of agents which can communicate each other. The information exchange is modeled via a communication graph $\mathcal{G} = (V, \mathcal{E})$, where $V = \{1, 2, \dots, N\}$ and $\mathcal{E} \subseteq V \times V$. Each node of the graph represents an agent, an edge (i, j) belonging to $\mathcal{E}$ represents the possibility for the agent j to receive information from agent i. In certain cases, the communication graph is allowed to be time-variant. This is used in order to model asynchronous working conditions or scenarios in which two agents are not always able to communicate, due to possible packet loss. 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 simple algorithms for simple tasks. One of the most studied of these algorithms is the consensus algorithm. In general, a consensus algorithm is a distributed strategy in which each agent is initialized with a certain real number. The

⋆ The research leading to these results has received funding from the European Community's Seventh Framework Programme under agreement. FP7-ICT-223866-FeedNetBack.

goal for the agents is to reach, possibly in finite time, a consensus value based on a suitable fusion of the initial values that the agents possess. This fusion is obtained by an information exchange. One way to accomplish this task is to label each agent, gather the values via multi-hop in the network, and combine them in the same way by each node. However, this implies that each agent needs to memorize the value of any other node, which can be hard if the network is large.

A different strategy is to use a discrete time linear algorithm consistent with the communication constraints. Assume that each node u possesses a state $x_u(t)$ which is initialized to a real number y_u, namely $x_u(0) = y_u$, and 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 (not containing u itself). We impose that $p_{uu}(t) \geq 0$, $p_{uv}(t) \geq 0$ 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 in 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 and zero elsewhere, 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 consistent with the graph $\mathcal{G}$, namely that $\mathcal{G}_{P(t)}$ is a subgraph of $\mathcal{G}$, where we define the graph $\mathcal{G}_{P(t)} = (V, \mathcal{E}_{P(t)})$, by letting $\mathcal{E}_{P(t)} := \{(u,v) \in V \times V \ : \ P(t)_{u,v} \neq 0\}$. The graph $\mathcal{G}_{P(t)}$ is called the graph associated with $P(t)$. Moreover $P(t)$ is row–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 row–stochastic matrix consistent with a graph $\mathcal{G}$. By Frobenius-Perron theorem Gantmacher (1959), if we choose P to be aperiodic and irreducible [1], then all the eigenvalues of P are strictly inside the unitary disk, except just one which is equal to 1. Moreover the eigenvalue 1 has algebraic multiplicity 1. Let $\boldsymbol{\pi}^T$ 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 $\alpha = \sum_{u=1}^N \boldsymbol{\pi}_u \boldsymbol{x}_u(0) = \sum_{u=1}^N \boldsymbol{\pi}_u y_u$. Notice that, if P is doubly stochastic, namely both P and P^T are row–stochastic, then $\boldsymbol{\pi} = \frac{1}{N}\mathbf{1}$ and so in this case the consensus point α is equal to the average of the initial states.

[1] P is aperiodic if the greatest common divisor of the lengths of all cycles in its associated graph $\mathcal{G}_P$ is one. The presence of a self-loop implies aperiodicity. P is irreducible if $\mathcal{G}_P$ is strongly connected, namely, for all $u, v \in V$, there exists a path connecting u to v.

One of the pioneering works in distributed algorithms for estimation is Tsitsiklis (1984), and much work has been done along the entire following decade. At the beginning of the last decade, motivated by a problem of formation control constrained by local information only, a number of papers proposed linear consensus as an effective algorithm to solve the problem Olfati-Saber and Murray (2003); Jadbabaie et al. (2003), showing convergence of the algorithm, relation with Markov chains via the row–stochastic matrix P, and proposing control-oriented criteria (e.g., Nyquist criterion) in order to ensure stability Fax and Murray (2004).

Many papers included in the model more realistic scenarios, such as packet-drop communication and delays Olfati-Saber et al. (2004, 2007); Fagnani and Zampieri (2009). A number of papers was also devoted to the study of randomized consensus algorithms Ren and Beard (2005); Boyd et al. (2006).

As the classical theory suggests (see Section 4), the typical trajectory which the states draw while approaching consensus value is exponential in the time. Many papers have been devoted to study the exponential rate of convergence, both for structured graphs Carli et al. (2008b); Delvenne et al. (2009), and in terms of optimization problems Xiao and Boyd (2004); Boyd et al. (2004).

Since moreover 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 Frasca et al. (2008); Rajagopal and Wainwright (2008); Kashyap et al. (2007); Lavaei and Murray (2009a,b).

To conclude, an increasing interest is devoted to the problem of designing an effective distributed Kalman-type estimation algorithm. A possible, simple way, has been found by an application of the consensus averaging procedure Alriksson and Rantzer (2006); Carli et al. (2008a); Olfati-Saber (2005).

One relevant issue concerning consensus algorithms is related to the choice a performance metrics which allows the comparison among 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 graphs 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.

The aim of this paper is to give a survey on this subject focusing on four performance metrics: the rate of convergence to asymptotic value, L^2-norm of the error

between the current state and its final value, the H^2-norm of the discrete time system which models the consensus algorithm and a cost related to an application of consensus to distributed Kalman filtering. Each of these metrics will be presented as the solution of a specific problem, and each will be computed in relation to some important families of graphs, namely general graphs, graphs with symmetries (called Cayley graphs) and geometric graphs.

2. SOME PRELIMINARIES ON GRAPH THEORY

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 does not exist different edges having the same tail and head. Hence, if an edge E is such that $s(e) = u$ and $f(e) = u$, then we can simply write $e = (u, v)$ 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 $(u, v) \in \mathcal{E} \iff (v, u) \in \mathcal{E}$. A graph is directed, or it is a digraph, if it is not directed.

Given a node $u \in V$, we denote by $\mathcal{N}_u^{in} = \{v : e = (v, u), \exists e \in \mathcal{E}\} = s(f^{-1}(u))$ [2] 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 $\mathcal{N}_u^{out} = \{v : e = (u, v), \exists e \in \mathcal{E}\} = f(s^{-1}(u))$ 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 $\mathcal{G} = (V, \mathcal{E})$, let $N := |V|$ and $M := |\mathcal{E}|$. 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. 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.

We will be interested in weighted graphs, namely couples $(\mathcal{G}, w)$ where $\mathcal{G}$ is a graph and w is a function $w : \mathcal{E} \to \mathbb{R}$ which associates to each edge a value. For example, if $\mathcal{G}$ is the graph underlying a Markov chain, the function $w(e) = w(u, v)$ represents the transition probability from u to v. For the consensus algorithm the value $w(u, v)$ represents the weight the node u gives to the information it receives from node v, so that with the above notation we have $w(u, v) = p_{uv}$. Finally, if we consider an electrical network, the value $w(u, v)$ can be interpreted as the resistance, or the conductance, of the edge (u, v).

[2] Here $f^{-1}(u)$ is the preimage of u.

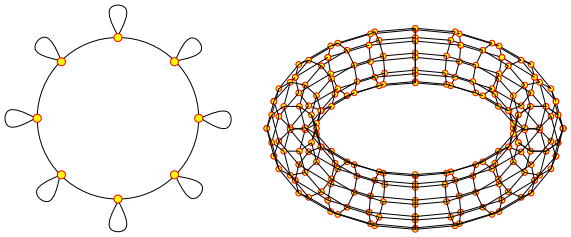

Figure 1. Two examples of Cayley graphs.

2.1 Cayley graphs

Cayley matrices and Cayley graphs 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 Abelian finite 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 Carli (2008) that there exists a function $\boldsymbol{g} : G \to \mathbb{R}$ 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 Cayley graph 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 Cayley graph. A Cayley graph is completely determined by giving the group G and a set $S \subseteq G$. Indeed, the set of edges $\mathcal{E}$ of a Cayley graph is such that $(i, j) \in \mathcal{E}$ if and only if $j - i \in S$.

In Fig. 1 two Cayley graphs are presented. On the left, $G = \mathbb{Z}_7$ 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 Abelian group G is isomorphic to the group $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$. In order to simplify the notation, in this paper we will restrict to Cayley graphs with respect to groups of the type $\mathbb{Z}_n^d$. More precisely we will consider families of Cayley graphs with respect to that group in which n varies while the dimension d is fixed and a positive δ is fixed (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 each node can not communicate with nodes that are too "far" from it.

2.2 Geometric graphs

Roughly speaking a geometric graphs is a perturbation of a regular grid in d dimension, for instance by moving the nodes and removing or adding some edges. There exists several different mathematical definitions of geometric graphs Doyle and Snell (1984); Gupta and Kumar (2000); Franceschetti and Meester (2007). Here we have chosen the model proposed in Doyle and Snell (1984); Barooah and Hespanha (2005) because it is, in our opinion, simpler and rather general.

[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

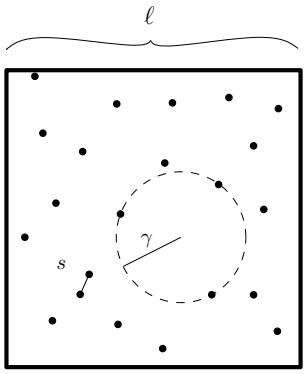

Figure 2. An example of geometric graph in 2 dimensions.

Consider an hypercube $Q \subset \mathbb{R}^d$ with edge length ℓ, namely $Q = [0, \ell]^d \subseteq \mathbb{R}^d$. Let $\mathcal{G} = (V, \mathcal{E})$ be a connected undirected graph such that $V \subset Q$ and $|V| = N$.

Following Doyle and Snell (1984); Barooah and Hespanha (2005), the following parameters can be defined:

- the minimum Euclidean distance [4] between any two nodes
$$s = \inf_{u,\, v \in V,\, u \neq v} \{d_{\mathrm{E}}(u,\, v)\}\,; \tag{4}$$
- the maximum Euclidean distance between any two connected nodes
$$r = \sup_{(u,\, v) \in \mathcal{E}} \{d_{\mathrm{E}}(u,\, v)\}\,; \tag{5}$$
- the radius γ of the largest ball centered in Q not containing any node of the graph
$$\gamma = \max\{r |\, B(x,\, r) \cap V = \emptyset,\, \forall\, x \in Q\}\,; \tag{6}$$
- the minimum ratio between the Euclidean distance of two nodes and their graphical distance
$$\rho = \min\left\{ \frac{d_{\mathrm{E}}(u,\, v)}{d_{\mathcal{G}}(u,\, v)} |\, (u,\, v) \in V \times V \right\}. \tag{7}$$

Such a graph is called a geometric graph with parameters $(Q,\, s,\, r,\, \gamma,\, \rho,\, N)$. The problems we will consider will involve class of geometric graphs with increasing number N of nodes and with increasing length ℓ of the hypercube edge, but with bounded parameters s, r, γ and ρ.

2.3 De Bruijn's graphs

De Bruijn's graphs constitutes a very particular class of graphs. Nevertheless we introduce them here because they are fast mixing digraphs in which, if the weights are chosen in a clever manner, it is possible to reach consensus in a minimum finite number of steps. The paper Delvenne et al. (2009) 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 whose adjacency matrix is the following

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

where the column vector $\mathbf{1}$ is k dimensional and the identity I is k^{n-1} dimensional. Its associated graph, $\mathcal{G}_F$,

[4] Given a generic graph $\mathcal{G} = (G,\, \mathcal{E})$ and two nodes $u,\, v \in G$ deployed in $\mathbb{R}^d$, we will denote by $d_{\mathrm{E}}(u,\, v)$ the Euclidean distance between u and v (in $\mathbb{R}^d$), and with $d_{\mathcal{G}}(u,\, v)$ their graphical distance (in $\mathcal{G}$), namely the length of the shortest path between them.

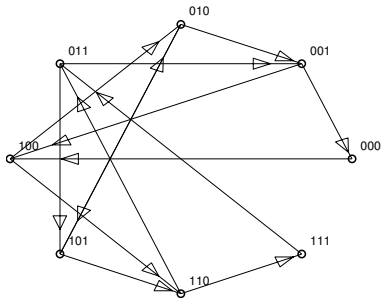

Figure 3. A de Bruijn graph with $N = 8$ nodes.

is called de Bruijn's graph de Bruijn (1946), and it is displayed in Figure 3 in the case $k = 2$, $n = 3$.

3. ELECTRICAL NETWORKS

It has been firstly stated in Doyle and Snell (1984) the remarkable and deep analogy between reversible Markov chains and electrical networks. Given a Markov chain in the form of its transition matrix P and its left invariant measure $\boldsymbol{\pi}^T$ (namely, the vector such that $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ and $\boldsymbol{\pi}^T \mathbf{1} = 1$), we say that it is reversible if it holds

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

where Π is the diagonal matrix with diagonal entries equal to the entries of $\boldsymbol{\pi}$. We focus our attention on Markov chains with symmetric P, which are clearly 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 a 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 $L_C := A^T \mathcal{C} A$. Notice that $L_C \mathbf{1} = 0$. Using the notion of incidence matrix, it is immediate to obtain that

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

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

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

$$\mathcal{R}_{uv}(\mathcal{G},\, C) = \frac{v_u - v_v}{I}, \tag{8}$$

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 Wu (2004). 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. Since potentials and currents mush satisfy both Kirchhoff's current law and Ohm's law, it can be obtained the following system

$$\begin{cases} A^T\boldsymbol{j} = \boldsymbol{i} \\ \mathcal{C}A\boldsymbol{v} = \boldsymbol{j} \end{cases} \tag{9}$$

which 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 (9), 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 (9) and the constraint on $\boldsymbol{v}$ in the following matrix form

$$\begin{bmatrix} L_C & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{i} \\ 0 \end{bmatrix}. \tag{10}$$

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

$$\begin{cases} L_C X_C = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \\ X_C\mathbf{1} = 0 \end{cases}. \tag{11}$$

We can write X_C in closed formula as follows

$$X_C = (L_C + \frac{1}{N}\mathbf{1}\mathbf{1}^T)^{-1} - \frac{1}{N}\mathbf{1}\mathbf{1}^T. \tag{12}$$

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 (12) 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 (10), yielding

$$\boldsymbol{v} = X_C\boldsymbol{i}. \tag{13}$$

Assume now that $\boldsymbol{i} = e_u - e_v$. This corresponds to inject 1 Ampere into the node v and extract 1 Ampere from the node u. Since $v_u - v_v = \boldsymbol{v}^T\boldsymbol{i}$, by definition (8) we can conclude that

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

Consider now a symmetric, row–stochastic, irreducible and aperiodic 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. However, it can be seen that they to not contribute to L_C.

From this definition of the electrical network, it is easy to obtain the following relation between P and 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 $\alpha = 2$, thus 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,

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

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

$$\mathcal{R}_{uv}(\mathcal{G}, P) = \frac{1}{2}(e_u - e_v)^T X(e_u - e_v),$$

where we define

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

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 two performance metrics, namely the rate of convergence and the H_2 cost (resp. Section 4 and Section 6).

4. THE RATE OF CONVERGENCE

The most classical performance index evaluating 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 an aperiodic and irreducible 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

$$\rho(P) = \max\{|\lambda| : \lambda \in \Lambda(P) \setminus \{1\}\}, \tag{14}$$

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 spectral graph community. In Markov chain theory $\rho(P)$ is related to the so called mixing time of the Markov chain having P as the transition matrix Levin et al. (2008). Spectral graph theory instead studies the geometric properties of weighted graphs using the so-called Laplacian L matrix 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. An extensive treatment of these topics can be found in Chung (1997).

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 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 k steps. It is possible to show moreover that there does not exists $n^k \times n^k$ 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 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 Diaconis and Stroock (1991), Diaconis and Saloff-Coste (1993) Fulman and Wilmer (1999) and Jerrum and Sinclair (1989)). For simplicity we will restrict our attention to symmetric stochastic matrices. If P is symmetric, then its eigenvalues are real. In the sequel we will assume that the eigenvalues of P are ordered in such a way that $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1}$. Notice that in this case we have that

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

If we find upper bounds on λ_1 and on $-\lambda_{N-1}$ we can get an upper on $\rho(P)$. Applying Gershgorin circle theorem we can argue that

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

Therefore, finding an upper bound on $\rho(P)$ reduces to finding an upper bound on λ_1, the second largest eigenvalue of P. To this aim the well-known Rayleigh-Ritz theorem (see Horn and Johnson (1990)) proves to be a helpful tool. Rayleigh-Ritz theorem in our case coincides with the following variational characterization of λ_1

$$\lambda_1 = \max\left\{\frac{\boldsymbol{x}^T P \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}, \boldsymbol{x} \neq 0 \text{ and } \boldsymbol{x} \perp \mathbf{1}\right\}. \tag{15}$$

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 Diaconis and Stroock (1991).

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 $(x, y) \in V^2$, $x \neq y$, let γ_{xy} be a path from x to y. Namely, γ_{xy} is a set of edges $\gamma_{xy} = \{e_0, \ldots, e_l\}$ such that $e_0 = (x, u_1)$, $e_i = (u_i, u_{i+1})$ $\forall\, i = 1, \ldots, l-1$ and $e_l = (u_l, y)$. 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_{xy}

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

where e are the edges forming γ_{xy} and $P(e) = P_{u,v}$ if $e = (u, v)$.

Let Γ be a collection of such paths, one for each couple (x, y). We associate to Γ the following quantity

$$\kappa = \kappa(\Gamma) = \max_{e \in \mathcal{E}}\{\sum_{\gamma_{xy} \ni e} |\gamma_{xy}|_P\} \tag{16}$$

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

This value has an immediate, intuitive, interpretation. We take, for any edge, the measure of the flow of information through that edge. Then, we minimize it over the edges, 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

$$\lambda_1 \leq 1 - \frac{N}{\kappa}, \tag{17}$$

with κ defined in (16).

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

In the following 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 Fulman and Wilmer (1999).

Cheeger inequality 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 in information from the set S to its complement $S^C = V \backslash S$ is linked to the probability transition 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

$$h(P) = \min_{S:|S|\leq \frac{N}{2}} \left\{ \frac{P(S \times S^C)}{|S|} \right\}, \tag{18}$$

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

$$1 - 2h(P) \leq \lambda_1 \leq 1 - h(P)^2, \tag{19}$$

with $h(P)$ defined in (18).

An electrical based inequality The electrical analogy proposed in Section 3 offers a bound on the second largest eigenvalue in terms of electrical quantities.

In fact, from Rayleigh-Ritz theorem we have that

$$\lambda_1 = \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{\boldsymbol{i}^T P \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\}.$$

Observe that the vector $\boldsymbol{i}$ which maximizes the quantity $\boldsymbol{i}^T P \boldsymbol{i}$ can be chosen as the normalized eigenvector relative to λ_1. Rayleigh-Ritz theorem can be applied to the matrix P^t too, and we obtain

$$\lambda_1^t = \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{\boldsymbol{i}^T P^t \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\}.$$

Since the eigenvector of P^t relative to the eigenvalue λ_1^t is the same for any power of P, the maximum is reached exactly for the same $\boldsymbol{i} \in \mathbb{R}^N$. So we can write

$$\begin{aligned}
\frac{1}{1-\lambda_1} &= \sum_{t\geq 0} \lambda_1^t = \sum_{t\geq 0} \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{\boldsymbol{i}^T P^t \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\} \\
&\overset{(a)}{=} \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{\sum_{t\geq 0} \boldsymbol{i}^T P^t \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\} \\
&\overset{(b)}{=} \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{\boldsymbol{i}^T (\sum_{t\geq 0} P^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T)\boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\} \\
&= \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{\boldsymbol{i}^T X \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\} \\
&= \max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{2\boldsymbol{i}^T X_C \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\},
\end{aligned}$$

where (a) holds because the maximum is attained at the same vector $\boldsymbol{i}$ and (b) holds because $\boldsymbol{i}^T \mathbf{1} = 0$, and where we recall that

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

Since we know that $\boldsymbol{v} = X_C \boldsymbol{i}$ is the potential at the nodes obtained by injecting i_u Ampere into the node u, $\forall\, u \in V$, and that the total power dissipated in the electrical network is $P = \boldsymbol{v}^T \boldsymbol{i}$, we end up with

$$\lambda_1 = 1 - \frac{1}{2P_{max}}$$

where P_{max} is the maximum dissipated power in the network by injecting a vector $\boldsymbol{i}$ of currents such that $||\boldsymbol{i}|| = 1$.

We can moreover give a bound using the maximum effective resistance $\mathcal{R}_{max} := \max\{\mathcal{R}_{u,v} : u, v \in V\}$. Notice in fact that

$$\begin{aligned}
\max_{\boldsymbol{i}:||\boldsymbol{i}||=1} \{2\boldsymbol{i}^T X_C \boldsymbol{i} : \boldsymbol{i} \perp \mathbf{1}\} &\geq \max_{(u,\, v)} \{(e_u - e_v)^T X_C (e_u - e_v)^T\} \\
&= \mathcal{R}_{max},
\end{aligned}$$

since the maximization on the right is made over $\boldsymbol{i} = \frac{1}{\sqrt{2}}(e_u - e_v)$, so in a smaller set. So we can argue that

$$\lambda_1 \geq 1 - \frac{1}{\mathcal{R}_{max}}.$$

However, this estimate proves to be not very effective for our purposes, as we will see in the following section.

4.2 Bounds on the convergence rate for Cayley graphs

In this section we will present the results about the rate of convergence for the class of graphs known as Cayley graphs. Details and proofs can be found for example in Carli et al. (2008b); Diaconis and Saloff-Coste (1993). Consider a class of Cayley graphs with respect to the group $\mathbb{Z}_n^d$ where we assume the dimension d fixed and n variable. Let δ be a positive constant such that $(i,\, j)$ is an edge of the graph only if $||j - i|| \leq \delta$. Assume that in the Cayley graphs there are the self loops and assume moreover that they are strongly connected.

Theorem 4.3. Let P be a stochastic Cayley matrix with respect to the group $\mathbb{Z}_n^d$ whose associate graph is in the previous class of Cayley graphs. Assume that all the nonzero entries of P lie in an interval $[p_{\min},\, p_{\max}]$. Then

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

where the two strictly positive constant C' and C depend only on d, δ and on $p_{\min}$ and $p_{\max}$, but not on n.

We can see here that the result obtained in the previous section using the maximum effective resistance does not give a good estimate of λ_1. Consider in fact a Cayley matrix with respect to the group $\mathbb{Z}_N$. It is rather intuitive that the maximum effective resistance coincides with the effective resistance between the two nodes that are at the maximum graphical distance, and that this grows in N as $\mathcal{R}_{max} = kN$, where k is a constant. The bound using the effective resistance is hence

$$\lambda_1 \geq 1 - \frac{1/k}{N}$$

while the correct order is given by Theorem 4.3, and is

$$\lambda_1 \geq 1 - \frac{C}{N^2}.$$

4.3 Bounds on the convergence rate for geometric graphs

In this section we analyze the essential spectral radius of stochastic matrices consistent with a geometric graph

characterized by the parameters $(Q,\, s,\, r,\, \gamma,\, \rho,\, N)$. The analysis is similar to that proposed in Boyd et al. (2006). In this paper, the authors consider the well known random geometric graph in dimension d, which is a probabilistic model for geometric graphs Gupta and Kumar (2000). The authors show that the essential spectral radius of stochastic matrices consistent with such graphs is, with high probability, the same as the essential spectral radius of stochastic matrices consistent with d-dimensional regular grids. Notice that the regular grids behave with respect to the essential spectral radius similarly to Cayley graphs. Here we show that a similar result holds for the class of geometric graphs proposed here.

A lower bound The following theorem offers a lower bound for the optimal stochastic matrix consistent with a geometric graph $\mathcal{G}$, namely the matrix having the lowest essential spectral radius. Since this bound holds for the optimal stochastic matrix, it will hold for any stochastic matrix consistent with $\mathcal{G}$.

Theorem 4.4. Consider a geometric graph $\mathcal{G}$ characterized by the parameters $(Q,\, s,\, r,\, \gamma,\, \rho,\, N)$. Then, if $P_{\mathcal{G}}$ is the row–stochastic, irreducible and aperiodic matrix consistent with $\mathcal{G}$ which minimizes the essential spectral radius, and we denote by $\rho(P^{\mathcal{G}})$ its value, it holds

$$\rho(P^{\mathcal{G}}) \geq 1 - \frac{C}{N^{2/d}} \tag{20}$$

where C is a constant only dependent on the parameters $(s,\, r,\, \gamma,\, \rho)$, and on the nonzero entries of $P^{\mathcal{G}}$.

An upper bound The following theorem offers an upper bound for the essential spectral radius of a stochastic matrix consistent with a geometric graph. We assume that all the nonzero entries of the stochastic matrix P lie in an interval $[p_{\min},\, p_{\max}]$.

Theorem 4.5. Consider a geometric graph $\mathcal{G}$ characterized by the parameters $(Q,\, s,\, r,\, \gamma,\, \rho,\, N)$. Then, if P is a row–stochastic, irreducible and aperiodic matrix consistent with $\mathcal{G}$, it holds

$$\rho(P) \leq 1 - \frac{C'}{N^{2/d}} \tag{21}$$

where C' is a constant only dependent on the parameters $(s,\, r,\, \gamma,\, \rho)$, and on $p_{\min}$.

5. TIME-DEPENDENT H_2 COST

To introduce the time-dependent H_2 cost we will refer to a possible practical application of consensus, related to static estimation. More details can be found in Garin and Zampieri (2009). 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 gets y_i which is

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

We assume that $n_i \sim \mathcal{N}(0,\, \sigma^2)$. In order to improve the value of the sensor estimates, we run a consensus setting $x_i(0) = y_i$ in order to average away the noise. Each sensor knows $x_i(t)$ which can be considered an estimate that i has of the parameter θ. We want to analyze the evolution of the estimation error as a function of the time t, which can be considered as a rough evaluation of the computational and communication resources employed by the algorithm. While for the rate of convergence, the increase of the number of agents always yields a performance degradation, this is not automatically true for the performance index considered in this section. 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. 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 state $\boldsymbol{x}(t)$ and the true value θ, normalized over the number of agents, namely we want to compute

$$J_{H_2}(P,\, t) = \frac{1}{N}\mathbb{E}\left[\boldsymbol{e}(t)^T\boldsymbol{e}(t)\right] \tag{22}$$

where $\boldsymbol{e}(t) = \boldsymbol{x}(t) - \theta\mathbf{1}$. Observe that this is a good performance metrics because $\mathbb{E}[e(t)] = \theta$, so the estimator is unbiased.

It turns out that

$$J_{H_2}(P,\, t) = \frac{1}{N}\,\mathrm{Tr}\,(P^T)^t P^t = \frac{1}{N}\sum_{\lambda\in\Lambda(P)} |\lambda|^{2t}$$

where the second equality holds if P is normal (e.g., if it is symmetric).

No bounds for this cost are known for P consistent with a generic graph. It is nonetheless known, in some special cases, how to use the structure of the graph in order to obtain bounds on such cost.

Example: de Bruijn's graphs

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

$$J_{H_2}(P,\, t) = \begin{cases} \dfrac{1}{n^t}, & 0 \leq t < k \\ \dfrac{1}{N}, & t \geq k \end{cases} \tag{23}$$

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

5.1 Bounds on the H_2 time-dependent cost for Cayley graphs

We start from a simple example. Consider the following consensus matrix

$$P = \begin{bmatrix} 1/3 & 1/3 & 0 & \cdots & \cdots & 1/3 \\ 1/3 & 1/3 & 1/3 & \cdots & \cdots & 0 \\ 0 & 1/3 & 1/3 & 1/3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1/3 & 1/3 & 1/3 \\ 1/3 & \cdots & \cdots & 0 & 1/3 & 1/3 \end{bmatrix} \tag{24}$$

whose associate graph is shown in Fig. 1. The essential spectral radius of P is $1 - C/N^2$. 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 4 depicts $J_{H_2}(P,\, t)$ as a function of t, for various values of N.

For any fixed N, we have evolutions which exponentially converge (with rate $\sim (1 - C/N^2)$) to a constant value

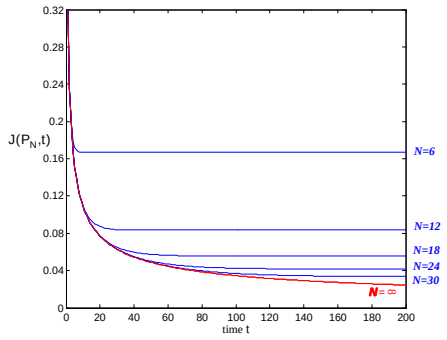

Figure 4. H_2 time–dependent cost, an example for a Cayley graph.

$1/N$. 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 bounds stated below for general Cayley matrices 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, $\lim_{N\to\infty} J_{H_2}(P, t)$ 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 4, 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 Garin and Zampieri (2009). Consider the group $G = \mathbb{Z}_n^d$. Consider a class of Cayley graphs with respect to the group $\mathbb{Z}_n^d$ where we assume the dimension d fixed and n variable. Let δ be a positive constant such that (i, j) is an edge of the graph only if $||j - i|| \leq \delta$ Assume that in the Cayley graphs there are the self loops and assume moreover that they are strongly connected.

Theorem 5.1. Let P be a stochastic Cayley matrix with respect to the group $\mathbb{Z}_n^d$ whose associate graph is in the previous class of Cayley graphs. Then

$$C \max\left\{\frac{1}{N}, \frac{1}{t^{d/2}}\right\} \leq J_{H_2}(P, t) \leq C' \max\left\{\frac{1}{N}, \frac{1}{t^{d/2}}\right\}, \tag{25}$$

where two constants, C and C' depend on d, δ and on the minimum and the maximum values of the nonzero entries of P, but not on n and t.

6. H_2 COST

The performance cost we are going to present and analyze in this section arises in a consensus problem with additive noise Bamieh et al. (2009); Carli et al. (2009); Frasca et al. (2008). Consider a consensus iteration corrupted by additive noise

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

in which $\boldsymbol{w}(t)$ is an i.i.d. process with zero mean and variance $\mathbb{E}[\boldsymbol{w}(s)\boldsymbol{w}(t)^T] = R\delta_{st}$, where δ_{st} is the Kroenecker delta function. For sake of simplicity, assume $R = I$. Assume moreover $\boldsymbol{x}(0)$ and $w(s)$ are uncorrelated for all $s \geq 0$. Define the error from the actual average 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 $\tilde{\boldsymbol{x}}(t)$ is the dispersion of $\boldsymbol{x}(t)$ around its center of mass and so it describes the distance from consensus. 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 $\Sigma(t) = \mathbb{E}[\tilde{\boldsymbol{x}}(t)\tilde{\boldsymbol{x}}(t)^T]$ the variance of the error. 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

$$J_{H_2}(P) = \frac{1}{N}\operatorname{Tr}\sum_{t\geq 0} P^t\Omega(P^T)^t. \tag{26}$$

If P is symmetric, the previous cost can be rewritten as

$$J_{H_2}(P) = \frac{1}{N}\sum_{\lambda\in\Lambda(P),\,\lambda\neq 1} \frac{1}{1-|\lambda|^2}. \tag{27}$$

This cost can be used to describe a different feature of the consensus algorithm. In classic control a way to describe the transient performance of a control system is through the position of the dominant eigenvalue. This is analogous to taking the essential spectral radius of P as a performance index of the consensus algorithm. In control however there are other choices of indices for evaluating the transient of a control system, such as, for instance, the L_2 norm of the error. If we do the same for the consensus algorithm, namely we take the usual consensus iteration (1), we assume that the initial state $\boldsymbol{x}(0)$ is a random variable with $\mathbb{E}[\boldsymbol{x}(0)\boldsymbol{x}(0)^T] = I$ and we take the index

$$\frac{1}{N}\sum_{t\geq 0}\mathbb{E}[\boldsymbol{x}(t) - \boldsymbol{x}(\infty)]_2^2 = \frac{1}{N}\sum_{t\geq 0}\mathbb{E}[\tilde{\boldsymbol{x}}(t)]_2^2,$$

It can be shown that this cost coincides with

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

which, in case P is symmetric, is the same as (26).

6.1 Bounds on the H_2 cost for Cayley graphs

Concerning Cayley graphs, from Carli et al. (2009) we know the following interesting result. Consider the group

$G = \mathbb{Z}_n^d$. Consider a class of Cayley graphs with respect to the group $\mathbb{Z}_n^d$ where we assume the dimension d fixed and n variable. Let δ be a positive constant such that (i, j) is an edge of the graph only if $||j - i|| \leq \delta$ Assume that in the Cayley graphs there are the self loops and assume moreover that they are strongly connected.

Theorem 6.1. Let P be a stochastic Cayley matrix with respect to the group $\mathbb{Z}_n^d$ whose associate graph is in the previous class of Cayley graphs. Then

$$C_d f(N, d) \leq J_{H_2}(P) \leq C'_d f(N, d) \tag{28}$$

where

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

where the two strictly positive constants, C_d and C'_d depend on d, S and on the minimum and the maximum values of the nonzero entries of P, but not on n.

6.2 Electric analogy and bounds on the H_2 cost 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 Lovisari et al. (2010). A similar result, applied to a different problem, can be found also in Ghosh et al. (2008).

Theorem 6.2. Given a stochastic, symmetric, aperiodic and irreducible matrix P, the associated LQ cost defined in (26) is equal to

$$J_{H_2}(P) = \overline{\mathcal{R}}(\mathcal{G}_{P^2}, P^2) := \frac{1}{N^2} \sum_{u \neq v} \mathcal{R}_{uv}(\mathcal{G}_{P^2}, P^2). \tag{29}$$

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 indeed 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 Doyle and Snell (1984); Barooah and Hespanha (2005)) 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 cost is bounded by

$$c_1 \overline{\mathcal{R}}(\mathcal{G}_P) \leq J_{H_2}(P) \leq c_2 \overline{\mathcal{R}}(\mathcal{G}_P). \tag{30}$$

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 whose graph is that associated with 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.

Theorem 6.3. Assume P to be a stochastic, symmetric, aperiodic and irreducible matrix, associated with a geometric graph $\mathcal{G}_P = (V, \mathcal{E})$ characterized by the parameters $(Q, s, r, \gamma, \rho, N)$. Assume moreover that all the nonzero entries of P lie in an interval $[p_{\min}, p_{\max}]$. Then

$$C_1 f(N, d) \leq J_{H_2}(P) \leq C_2 f(N, d) \tag{31}$$

where

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

and where the two strictly positive constants C_1 and C_2 depend only on Qs, r, γ, ρ, $p_{\min}$ and $p_{\max}$.

The proof of the theorem is based on the following reasoning. Consider a geometric graph which satisfies the assumptions in Theorem 6.3. Then there exist two regular Cayley-type graphs, $\mathcal{L}_1$ and $\mathcal{L}_2$ (intuitively, a rougher and a finer version of $\mathcal{G}_P$), such that

$$k_1 + q_1 \overline{\mathcal{R}}(\mathcal{L}_1) \leq J_{H_2}(P) \leq k_2 + q_2 \overline{\mathcal{R}}(\mathcal{L}_2), \tag{32}$$

where q_1, q_2, k_1 and k_2 depend on s, r, γ, ρ, and on $p_{\min}$ and $p_{\max}$ only. Theorem 6.3 is thus derived from Theorem 6.2 and Theorem 6.1 by estimating $\overline{\mathcal{R}}(\mathcal{L}_1)$ and $\overline{\mathcal{R}}(\mathcal{L}_2)$ as the cost of Cayley graphs.

7. KALMAN FILTER

In this section we present again an estimation algorithm based on consensus, but, differently from what we did in the previous sections, we switch from the static to the dynamic case. This application of the consensus algorithm can be found in Carli et al. (2008a). In this paper, the authors consider a set of N agents measuring a random walk corrupted by noise

$$\begin{cases} x(t+1) = x(t) + w(t) \\ y_i(t) = x(t) + n_i(t) \end{cases} \tag{33}$$

where $w(t)$ and $n(t)$ are zero mean i.i.d. gaussian processes with $\mathbb{E}[w(t)^2] = q$ and $\mathbb{E}[n_i(t)^2] = r$. We assume moreover that $x(0)$ is a zero mean gaussian random variable with $\mathbb{E}[x(0)^2] = \sigma^2$ and that $\mathbb{E}[w(t)x(0)] = \mathbb{E}[n_i(t)x(0)] = 0$ for all $t \geq 0$ and for all $i = 1, \ldots, N$.

We assume each agent i has in its memory an estimate $\hat{x}_i(t)$ of $x(t)$ and that the agents can communicate only through an assigned communication graph $\mathcal{G} = (V, \mathcal{E})$. In order to update its estimate $\hat{x}_i(t)$ using measurement $y_i(t)$ and merging information from the other nodes, three steps are proposed:

- Prediction step: each agent propagates its state one step ahead, obtaining $\hat{x}_i^P(t) = \hat{x}_i(t)$;
- Estimation step: each agent updates its estimate using its measurement by computing a convex combination of $\hat{x}_i^P(t)$ and $y_i(t)$
$$\hat{x}_i^E(t) = (1 - \ell)\hat{x}_i^P(t) + \ell y_i(t)$$
- Communication step: in the centralized algorithm, it can be shown that the optimal way to combine all the estimates is to average over them. In the proposed algorithm, the agents combine the estimates performing a number of a consensus iterations
$$\hat{\boldsymbol{x}}(t+1) = P^m \hat{\boldsymbol{x}}^E(t)$$
where we stacked all the estimates $\hat{x}_i^E(t)$ and $\hat{x}_i(t+1)$ in the vectors $\hat{\boldsymbol{x}}^E(t)$ and $\hat{\boldsymbol{x}}(t+1)$. The value m is the number of times the agents are allowed to exchange information between two measurements.

Notice that for this model the prediction step is irrelevant. The estimation is thus described by the following iteration

$$\hat{\boldsymbol{x}}(t+1) = (1-\ell)P^m\hat{\boldsymbol{x}}(t) + \ell P^m\hat{\boldsymbol{y}}(t). \qquad (34)$$

At first glance, one could take as performance cost just the second eigenvalue of P and assume m is large enough. In this way, during the time interval between two measurements, the estimates are able to converge to the centralized estimate, which is the optimum.

However, a more natural performance index can be considered. Let $\tilde{\boldsymbol{x}}(t) := \hat{\boldsymbol{x}}(t) - x(t)\mathbf{1}$ be the error between the estimates and the value to be estimated. Let $\Sigma(t) := \mathbb{E}[\tilde{\boldsymbol{x}}(t)\tilde{\boldsymbol{x}}(t)^T]$ be the error covariance matrix. Then we define

$$J_K(P,m) := \limsup_{t\to\infty} \operatorname{Tr}\Sigma(t).$$

We try now to evaluate $J_K(P,m)$. Assume for sake of simplicity $\mathbb{E}[x(0)] = 0$ and $\mathbb{E}[x(0)^2] = 1$. Then is easy to see that $\mathbb{E}[\tilde{\boldsymbol{x}}(t)] = 0$ and that $\Sigma(t)$ evolves according to the following equation

$$\Sigma(t) = (1-\ell)^{2t}\mathbf{1}\mathbf{1}^T + q\sum_{s=0}^{t-1}(1-\ell)^{2s}\mathbf{1}\mathbf{1}^T + r\ell^2\sum_{s=0}^{t}(1-\ell)^{2s}P^{(s+1)m}(P^T)^{(s+1)m}.$$

Therefore we have that

$$J_K(P,m) = \frac{q}{1-(1-\ell)^2} + \frac{r\ell^2}{N}\sum_{s\geq 0}(1-\ell)^{2s}\operatorname{Tr}P^{(s+1)m}(P^T)^{(s+1)m}$$

and so, using the fact that P is symmetric, we can argue that

$$J_K(P,m) = \frac{q}{1-(1-\ell)^2} + \frac{r\ell^2}{N}\sum_{\lambda\in\Lambda(P)}\frac{|\lambda|^{2m}}{1-(1-\ell)^2|\lambda|^{2m}}.$$

Observe now that, similarly to what happen to the H_2 costs, $J_K(P,m)$ depends on the entire spectrum, not only on the second largest eigenvalue. Observe moreover that the parameter m in the index $J_K(P,m)$ plays a similar role played by the parameter t in the cost $J_{H_2}(P,t)$ presented in Section 5. In fact, by the trivial fact that $0 \leq |\lambda|^{2m} \leq 1$, $\forall\,\lambda\in\Lambda(P)$, we can easily obtain

$$J_K(P,m) \geq \frac{q}{1-(1-\ell)^2} + \frac{r\ell^2}{N}\sum_{\lambda\in\Lambda(P)}|\lambda|^{2m}$$

$$J_K(P,m) \leq \frac{q}{1-(1-\ell)^2} + \frac{r\ell^2}{N(1-(1-\ell)^2)}\sum_{\lambda\in\Lambda(P)}|\lambda|^{2m}$$

and so we can conclude that there exists three constants depending only on the variances of process and measurement noises, and on the gain ℓ, such that

$$k_1 + q_1 J_{H_2}(P,\,m) \leq J_K(P,m) \leq k_1 + q_2 J_{H_2}(P,\,m).$$

For this reason we can use the bounds we know for $J_{H_2}(P,t)$ for also analyzing the cost $J_K(P,m)$.

REFERENCES

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

Bamieh, B., Jovanovic, M., Mitra, P., and Patterson, S. (2009). Coherence in large-scale networks: Dimension dependent limitations of local feedback. *Submitted*.

Barooah, P. and Hespanha, J. (2005). Estimation from relative measurements: Error bounds from electrical analogy. *ICISIP*.

Boyd, S., Diaconis, P., and Xiao, L. (2004). Fastest mixing markov chain on a graph. *iam Review*.

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

Carli, R. (2008). *Topics on the average consensus problem*. Ph.D. thesis, PhD school in Information Engineering – University of Padua.

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

Carli, R., Fagnani, F., Speranzon, A., and Zampieri, S. (2008b). Communication constraints in the average consensus problem. *Automatica*, 44, 671–684.

Carli, R., Garin, F., and Zampieri, S. (2009). Quadratic indices for the analysis of consensus algorithms. In *Proceedings of Information Theory and Applications Workshop*, 96–104. [10.1109/ITA.2009.5044929] [http://dx.medra.org/10.1109/ITA.2009.5044929].

Chung, F. (1997). *Spectral Graph Theory*, volume 92 of *Regional conference series in mathematics*. American Mathematical Society.

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

Cybenko, G. (1989). Dynamic load balancing for distributed memory multiprocessors. *J. Parallel Distrib. Comput.*, 7(2), 279–301. doi: http://dx.doi.org/10.1016/0743-7315(89)90021-X.

de Bruijn, N. (1946). A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 46.

Delvenne, J., Carli, R., and Zampieri, S. (2009). Optimal strategies in the average consensus problem. *Systems & Control Letters*, 58, 759–765.

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

Diaconis, P. and Saloff-Coste, L. (1993). Comparison techniques for random walk on finite groups. *The Annals of Probability*, 21(4), 2131–2156.

Doyle, P. and Snell, J. (1984). *Random Walks and Electric Networks*. Carus Monographs. Mathematical Association of America.

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

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

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

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

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

Ganeriwal, S., Kumar, R., and Srivastava, M.B. (2003). Timing-sync protocol for sensor networks. In *Proc. of the 1st international conference on Embedded networked sensor systems (SenSys '03)*, 138–149. Los Angeles, CA, USA. doi:http://doi.acm.org/10.1145/958491.958508.

Gantmacher, F. (1959). *The Theory of Matrices*. Chelsea Publishing Company.

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

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

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

Horn, R. and Johnson, C. (1990). *Matrix Analysis*. Cambridge University Press.

Jadbabaie, A., Lin, J., and Morse, A. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *Automatic Control, IEEE Transactions on*, 48(6), 998 – 1001.

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

Kashyap, A., Basar, T., and Srikant, R. (2007). Quantized consensus. *Automatica*, 43, 192–1203.

Lavaei, J. and Murray, R. (2009a). On quantized consensus by means of gossip algorithm - part i: Convergence proof. In *American Control Conference, 2009. ACC '09.*, 394 –401.

Lavaei, J. and Murray, R. (2009b). On quantized consensus by means of gossip algorithm - part ii: Convergence time. In *American Control Conference, 2009. ACC '09.*, 2958 –2965.

Levin, D., Peres, Y., and Wilmer, E. (2008). *Markov Chains and Mixing Times*. American Mathematical Society.

Lovisari, E., Garin, F., and Zampieri, S. (2010). A resistance-based approach to consensus algorithm performance analysis. Submitted to CDC10.

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

Olfati-Saber, R. (2005). Distributed Kalman filter with embedded consensus filters. In *Proceedings of 44th IEEE Conf. on Decision and Control and 2005 European Control Conference*, 8179–8184.

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

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

Olfati-Saber, R. and Murray, R. (2003). Consensus protocols for networks of dynamic agents. In *American Control Conference, 2003. Proceedings of the 2003*, volume 2, 951 – 956.

Rajagopal, R. and Wainwright, M. (2008). Network-based consensus averaging with general noisy channels. *CoRR, abs/0805.0438*. On-line available.

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

Tsitsiklis, J. (1984). *Problems in decentralized decision making and computation*. Ph.D. thesis, Department of EECs, MIT.

Wu, F. (2004). Theory of resistor networks: the two-point resistance. *J. Phys. A: Math. Gen.*

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

Zhu, H., Cano, A., and Giannakis, G. (2008). Distributed demodulation using consensus averaging in wireless sensor networks. In *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, 1170 –1174.