

Almost sure exponential convergence to consensus of random gossip algorithms

Giorgio Picci^{1,*} and Thomas J. Taylor²

¹*Department of Information Engineering, University of Padova, Italy*

²*Department of Mathematics and Statistics, Arizona State University, Tempe, AZ, USA*

SUMMARY

In this paper, we provide a proof of almost sure exponential convergence to consensus for a general class of ergodic edge selection processes. The proof is based on the multiplicative ergodic theorem of Oseledec and also applies to continuous time gossip algorithms. An example of exponential convergence in a non ergodic case is also discussed. Copyright © 2012 John Wiley & Sons, Ltd.

Received 27 August 2011; Revised 6 February 2012; Accepted 23 April 2012

KEY WORDS: consensus; multiplicative ergodic theorem

1. DEDICATION

Chris Byrnes brought both of us to Arizona State University in the mid-1980s. Our collaboration started there in 1986 and has continued (although in a somewhat scattered way) until recently. We owe him a great deal, both scientifically and from a human point of view.

We dedicate this paper to his memory.

2. CONSENSUS FOR RANDOM GOSSIP ALGORITHMS

Consider a finite set of nodes representing say wireless sensors or distributed computing units, exchanging information for the purpose of forming a common estimate of some physical variable x . At time t , node i has a local estimate of the variable, denoted by $x_i(t)$ and as communication with a neighboring node j takes place, the two nodes i and j exchange information on their estimates $x_i(t)$, $x_j(t)$ and update them in some rational way so as to achieve consensus, namely at the successive time instant $x_i(t)$, $x_j(t)$ are both processed in such a way that the new estimates of the two nodes will be closer and eventually led to coincide (at least) asymptotically as $t \rightarrow \infty$.

We shall assume that only neighboring nodes can communicate. Stated more precisely, let $G := (V, E)$ be the graph with vertices $V = \{v_i; i = 1, 2, \dots, d\}$ representing the nodes, and let these nodes be connected by n edges in a set $E := \{e_k = v_{i_k} v_{j_k}\}$ where the pairs (i_k, j_k) range in some subset of neighboring pairs (i, j) ; we shall assume that, in each (discrete) time interval, communication can occur only between the two nodes belonging to the same edge. It is implicit in the preceding description that the communication relation between two nodes is mutual (i.e., reflexive), so that our graph will be undirected.

*Correspondence to: Giorgio Picci, Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy.

†E-mail: picci@dei.unipd.it

The communicating edges are chosen according to some stationary random mechanism. We shall visualize this by introducing a stationary *edge process* $\mathbf{e} := \{\mathbf{e}(t); t \in \mathbb{Z}\}$ taking values in E . The simplest communication scheme occurs when, in each time slot, there is just one (randomly chosen) couple of neighboring nodes exchanging information. This is what is normally referred to as a *random gossip algorithm*, see for example [1–3]. The case when many edges can simultaneously communicate is also of interest in some applications and will be discussed shortly in Section 8.

The main questions about gossip algorithms are: proving global convergence to consensus; that is, $\lim_{t \rightarrow \infty} x_i(t) - x_j(t) = 0$ for all i, j s, and estimating the rate of convergence. It is known that, in some cases, the rate may be very slow and make the scheme of little practical value.

There is a vast and rapidly growing literature on this subject at the boundary of control, communication and computer sciences, and we shall refer the reader to [4] for a very readable historical survey and a description of many possible application areas. Recent papers discussing convergence are [5–7] and [8, 9].

Main results of the paper

As far as we have been able to see, many results in the literature seem to deal with convergence in L^2 ; that is, convergence of the first two moments. Although this implies convergence in probability, in general it does not guarantee almost sure (a.s.) convergence; that is, that (almost) all sample paths of the x process governed by a random gossip algorithm will converge to consensus. In this sense, a.s. convergence is a ‘practical’ issue. Also, the edge process is normally taken to be a i.i.d. process or a Markov chain, which does not seem to have a cogent physical justification. Here, based on the results reported in a preliminary conference version of this paper [10], we instead provide a proof of a.s. convergence to consensus for an extremely general class of edge selection processes. Our main convergence result is based on Oseledec multiplicative ergodic theorem [11, 12] exploiting the doubly stochastic character of the updating algorithm. The result may seem to belong to the asymptotic analysis of Markov chains in a random environment, as initiated by Wolfowitz, Cogburn *et al.*, see [13–16] but in fact turns out to have little to do with this area because, on the one hand, the random Markov matrices analyzed here are rather special, namely doubly stochastic and irreducible, so that the quest for an invariant measure becomes a trivial matter. On the other hand, the rate of convergence results that we obtain under an extremely general stochastic evolution assumption, seem to be out of reach by the general Markov chain techniques exposed in these papers.

One more novelty in this paper is the treatment of continuous time gossip algorithms. This will be dealt with in Section 7.

3. MODELING OF GOSSIP ALGORITHMS

Our notion of gossip algorithms is that while two nodes v_i and v_j are in communication, they exchange information and refine their estimates. The mutual information allows them to maneuver their coordinates toward a consensus, and each successive contact enables a further exponential refinement of this consensus.

We shall model this adjustment in discrete time by a simple symmetric linear relation

$$\begin{aligned} x_i(t+1) &= x_i(t) + p(x_j(t) - x_i(t)) \\ x_j(t+1) &= x_j(t) + p(x_i(t) - x_j(t)) \end{aligned} \quad (1)$$

where p is some positive gain parameter modeling the speed of adjustment. For stability, we need to impose that $|1 - 2p| \leq 1$ and hence $0 \leq p \leq 1$. On the whole coordinate vector $x(t) \in \mathbb{R}^d$, this dynamics corresponds to $x(t+1) = A(e)x(t)$, the $d \times d$ real matrix $A(e)$ depending on the selected edge e at that particular time instant; specifically, when $e = v_i v_j$:

$$A(e) = I_d - p(1_{v_i} - 1_{v_j})(1_{v_i} - 1_{v_j})' \quad (2)$$

where 1_{v_i} denotes the vector that takes the value 1 in the i^{th} entry and zero otherwise. Note that each matrix $A(e)$ is a symmetric doubly stochastic matrix. If $p \neq 1/2$, the value $1 - 2p$ is a simple

eigenvalue, which is associated to the eigenvector $(1_{v_i} - 1_{v_j})$, and the codimension one subspace $(1_{v_i} - 1_{v_j})^\perp$ is the eigenspace of the eigenvalue 1. Boyd *et al.* [2] use the value $p = \frac{1}{2}$ that makes $A(e)$ into a rank one projection. In this case system, (1) is of ‘deadbeat’ type and two communicating nodes reach consensus in one step.

In general, our gossip algorithm generates stochastic trajectories in $\mathbb{R}^d$ according to the random dynamical system $x(t+1) = A(\mathbf{e}(t))x(t)$ or,

$$x(t) = \prod_{s=0}^{t-1} A(\mathbf{e}(s))x_0, \quad \mathbf{e}(s) \in E \quad (3)$$

where the ordering of matrices in the product is from largest time index on the left to smallest index on the right.

4. PARAContracting MATRICES

In the following, the matrix norm will be taken to be the spectral norm (i.e., the largest singular value).

Definition 4.1

An $n \times n$ matrix A is called *paracontracting* if $\|A\| \leq 1$ and

$$0 \neq x \in [\text{Ker}(I - A)]^\perp \Rightarrow \|Ax\| < \|x\| \quad (4)$$

where for $x \in \mathbb{R}^n$, $\|x\|$ denotes Euclidean norm.

The following facts are taken from Nelson and Neumann’s paper [17]

Proposition 4.1

If $A \in \mathbb{R}^{n \times n}$ is paracontracting, then

$$[\text{Ker}(I - A)]^\perp = \text{Im}(I - A). \quad (5)$$

If A and B are paracontracting, so is AB and

$$\text{Ker}(I - AB) = \text{Ker}(I - A) \cap \text{Ker}(I - B). \quad (6)$$

Moreover,

$$\text{Ker}(I - A^*A) = \text{Ker}(I - A). \quad (7)$$

Note that each $A(e)$ is paracontracting as it trivially satisfies $A(e)x \neq x \implies \|A(e)x\| < \|x\|$. It follows from the proposition earlier that any finite product $\prod_i A(e_i)$ and hence the right-ordered product of random matrices

$$B = B(\mathbf{e}) := \prod_{s=1}^N A(\mathbf{e}(s))$$

is also paracontracting. In addition, the subspace $\{x : x = Bx\} = \text{Ker}(I - B)$ is equal to $\bigcap_s \text{Ker}(I - A(\mathbf{e}(s)))$.

For a paracontraction B , the orthogonal complement of the null space $\text{Ker}(I - B)^\perp$ is an invariant subspace of B . The largest singular value of B on this invariant subspace is the norm $\|B|_{\text{Ker}(I - B)^\perp}\|$. This number, which we will denote $\gamma(B)$, is called the *contraction factor* of a paracontraction.

Let L_V denote the vertex space of G (i.e., the vector space of functions $f : V \rightarrow \mathbb{R}$, isomorphic to $\mathbb{R}^d$) and similarly, L_E the edge space of G (isomorphic to $\mathbb{R}^n$); let $\mathbf{1} \in L_V$ denote the constant function and $\mathbf{1}^\perp$ denote the subspace of L_V perpendicular to $\mathbf{1}$ under the dot product (i.e., the functions with zero average). Let $\|\cdot\|$ denote the Euclidean norm on L_V . Remember that a Markov matrix M has rows, which sum to 1; that is, satisfies $M\mathbf{1} = \mathbf{1}$, hence, has eigenvalue 1 associated with the eigenvector $\mathbf{1}$.

The following lemma is trivial, but we will prove it anyway.

Lemma 4.1

Let $G = (V, E)$ be a graph. The subspace $\text{span}\{1_{v_i} - 1_{v_j} : (v_i v_j) \in E\} = \mathbf{1}^\perp$ iff G is connected.

Proof

Consider the ‘difference’ operator $D : L_V \rightarrow L_E$ defined by $Df(e) = f(v_i) - f(v_j)$ if $e = (v_i v_j) \in E$ (choose an arbitrary assignment of sense for each edge) and let $Df := \text{vec}\{Df(e) \mid e \in E\}$. The rows of the matrix representation of the map $f \rightarrow Df$ are precisely the vectors $(1_{v_i} - 1_{v_j})'$ with $e = (v_i v_j) \in E$. The statement of the lemma amounts to the statement that the rowspan of this matrix is $\mathbf{1}^\perp$, which is equivalent to the statement that $\ker(D) = \text{span}\{\mathbf{1}\}$. Thus, $\ker(D)$ is the set of functions in L_V for which $f(v_i) - f(v_j) = 0$ whenever $(v_i v_j) \in E$. But this set of functions is constant on each connected component of G , hence, is made of constant functions exactly when G is connected. $\square$

Let e range on any connected subgraph $G' = (V, E')$ with $E' \subseteq E$. The product B is also Markov, hence, has the eigenvector $\mathbf{1}$ with eigenvalue 1. Each $\text{Ker}(I - A(e)) = (1_{v_i} - 1_{v_j})^\perp$ when $e = (v_i v_j)$, so $\text{Ker}(I - B) = \text{span}\{\mathbf{1}\}$ (is one dimensional) as soon as $\mathbf{1}^\perp = \text{span}\{1_{v_i} - 1_{v_j} : (v_i v_j) \in E'\}$. But by Lemma 4.1, this is the case iff $G' = (V, E')$ is connected. Hence, we have $\|Bx\| < \|x\|$ for all $x \in \mathbf{1}^\perp$. Because $\text{span}\{\mathbf{1}\}$ and $\mathbf{1}^\perp$ are (orthogonal) complementary invariant subspaces for any finite product of the form $B(E') := \prod_{e_i \in E'} A(e_i)$, we can write the last inequality as $\|B(E', \pi)\|_{\mathbf{1}^\perp} < 1$, for which we shall normally use the shorthand $\|B(E', \pi)\|_{\mathbf{1}^\perp} < 1$. In conclusion, we have the following.

Corollary 4.1

Let $G' = (V, E')$ with $E' \subseteq E$ be any connected subgraph of G . Let $\{e_i : 1 \leq i \leq n'\}$ be an ordering of E' , and let π denote a permutation of $\{1, 2, \dots, n'\}$. Let $B(E', \pi) = \prod_{i=1}^{n'} A(e_{\pi i})$, where the product is ordered from right to left. Then $\|B(E', \pi)\|_{\mathbf{1}^\perp} < 1$.

5. CONVERGENCE OF THE GOSSIP ALGORITHM

This section will be the basis of our discrete time results.

Let $\Omega = E^\mathbb{N}$ be the space of all semi-infinite sequences taking values in E , and let $\sigma : \Omega \rightarrow \Omega$ denote the shift map: $\sigma(e_0, e_1, e_2, \dots, e_n, \dots) = (e_1, e_2, \dots, e_n, \dots)$. Let $ev_k : \Omega \rightarrow E$ denote the evaluation on the k^{th} term. Let μ denote an ergodic shift invariant probability measure on Ω , so that the edge process $\mathbf{e}(k) : \omega \rightarrow ev_k(\omega)$ is ergodic. Special cases, which may be helpful to visualize are, $\mu =$ product measure so that $\mathbf{e}(k)$ is iid, or a Markov measure defined on cylinder sets by $\mu([e_0, e_1, \dots, e_n]) = \pi_{e_0} P_{e_0 e_1} P_{e_1 e_2} \dots P_{e_{n-1} e_n}$ for a transition matrix P with state space E and stationary distribution π . However, what we shall do works for general ergodic processes.

In both the continuous and discrete time versions of this problem, a controllability-like result is required for convergence. Lemma 5.1 next is such a controllability condition. It establishes a condition under which the (stationary) inhomogeneous Markov chain with time T transition matrix $A(\mathbf{e}_T) \dots A(\mathbf{e}_2)A(\mathbf{e}_1)$ has positive probability to reach any element of the vertex set from any vertex.

Lemma 5.1

Suppose that $G = (V, E)$ is connected and let $\mathbf{e}(t) : \Omega \rightarrow E$ be a stationary ergodic process for which the probability $p_e := \mu\{\mathbf{e}(0) = e\}$ is strictly positive on $E' \subset E$ where $G' = (V, E')$ is a connected subgraph of G ; that is, $\forall e \in E', p_e > 0$. Then, there is a deterministic time T such that a sample trajectory of the process $\mathbf{e}$ restricted to $[0, T]$, visits all edges in E' with positive probability and

$$\mu\{\omega \in \Omega : \|A(\mathbf{e}(T)) \dots A(\mathbf{e}(1))\|_{\mathbf{1}^\perp} < 1\} > 0.$$

Proof

By Corollary 4.1, $\|A(e_T) \cdots A(e_1)\|_{1^\perp} < 1$ iff for $E' = \bigcup_{i=1}^T \{e_i\} \subseteq E$, the subgraph $G' = (V, E')$ is connected and this in turn happens if and only if E' contains a spanning tree. Hence, we just need to show that for T large enough, there is positive probability that a sample trajectory contains a spanning tree. Intuitively, if this was not so, the graph could not be connected. A formal proof of this fact is as follows.

Let $f(T) = \mu\left(\left\{\omega : \bigcup_{i=1}^T \{\mathbf{e}(i)\} \text{ contains a spanning tree}\right\}\right)$. Note that $f(T)$ is monotonic increasing, because the constraint ' $\bigcup_{i=1}^T \{\mathbf{e}(i)\}$ contains a spanning tree' is progressively weaker as T increases. Clearly $f(T) \geq 0$ for all T . Either $f(T) = 0$ for all T , or there is a T for which $f(T)$ is positive.

By assumption $p_e = \mu(\mathbf{e}(1) = e)$; is positive for all $e \in E'$. Let $g(T) = \mu\left(\left\{\omega : \bigcup_{i=1}^T \{\mathbf{e}(i)\} = E'\right\}\right)$. Clearly, $0 \leq g(T) \leq f(T)$. We will prove that $f(T) > 0$ for some finite deterministic T , by proving the same for $g(T)$.

Consider the indicator function $\delta(e, \cdot) : E \rightarrow \{1, 0\}$, defined by $\delta(e, x) = 1$ if $x = e$ and zero otherwise. The time average of the stochastic process $\delta(e, \mathbf{e}(t))$ satisfies Birkhoff's ergodic theorem (see e.g. [18]), which amounts to

$$\mu\left(\left\{\omega : \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) = p_e\right\}\right) = 1.$$

From this follows $\mu\left(\left\{\omega : \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right) = 1$, which in turn implies $\mu\left(\left\{\omega : \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right) = 1$, where the latter limit is taken in the extended sense that the sum increases to a positive number or positive infinity. From this, it follows

$$\begin{aligned} 1 &= \mu\left(\bigcap_{e \in E'} \left\{\omega : \lim_{T \rightarrow \infty} \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right) \\ &= \mu\left(\bigcap_{e \in E'} \bigcup_{T > 0} \left\{\omega : \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right) \\ &= \mu\left(\bigcup_{T > 0} \bigcap_{e \in E'} \left\{\omega : \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right), \end{aligned}$$

where the third equality follows from distribution of finite intersection over infinitary unions, and the second equality from the fact that the mapping $T \mapsto \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i))$ is monotonic. Now, the family of sets $T \mapsto \bigcap_{e \in E'} \left\{\omega : \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}$ is monotonic increasing, so by sigma additivity

$$1 = \lim_{T \rightarrow \infty} \mu\left(\bigcap_{e \in E'} \left\{\omega : \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right),$$

although $\mu\left(\bigcap_{e \in E'} \left\{\omega : \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right) = 0$ for T small enough. Thus, there is a finite T for which $\mu\left(\bigcap_{e \in E'} \left\{\omega : \sum_{i=0}^{T-1} \delta(e, \mathbf{e}(i)) > 0\right\}\right) > 0$. But the latter probability is exactly $\mu\left(\left\{\omega : \bigcup_{i=1}^T \{\mathbf{e}_i\} = E'\right\}\right)$.

Hence, we have shown that for T large enough, there is a positive probability that a sample trajectory will visit all edges in E' and will therefore contain a spanning tree. $\square$

Remark 5.1

Tahbaz-Salehi and Jadbabaie [5] offer an alternative controllability condition, that the directed graph of the Markov matrix $\mathbb{E} A(\mathbf{e})$ should contain a directed spanning tree. This condition is equivalent to the controllability condition of Lemma 5.1 (which is a version Lemma 4.1 of [10]), subject to our narrower class of models more specifically aligned to modeling a gossip process: the digraph of the expectation of our Markov matrix $\mathbb{E} A(\mathbf{e})$ contains a spanning tree iff p_e is strictly positive for all e in the edge set of a strongly connected graph. Indeed, by the Birkhoff theorem,

$$\mathbb{E} A(\mathbf{e}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n A(\mathbf{e}(k))$$

a.s., which is equal to $\sum_{e \in E} A(e) p_e$ by stationarity. In particular, the (i, j) -entry of $\mathbb{E} A(\mathbf{e})$ is positive iff $p_e > 0$ for the edge $e = v_i v_j$, so that the matrix $\mathbb{E} A(\mathbf{e})$ is irreducible iff p_e is positive on the edge set of a connected graph, that is, iff p_e is positive on a spanning tree.

It is a fact that a power of an ergodic transformation need not be ergodic. The following elementary lemma describes what happens when this occurs.

Lemma 5.2

Suppose that σ is an ergodic μ -preserving transformation, and that σ^T is not ergodic. Then, the sigma algebra $\mathcal{I}$ of σ^T invariant sets is generated mod null sets by a partition $\Omega = \bigsqcup_{i=0}^{K-1} \Omega_i$ such that $(\sigma^T)^{-1} \Omega_i = \Omega_i$, $\mu(\Omega_i) = \frac{1}{K}$ for $0 \leq i \leq K-1$, K divides T and $\sigma^{-1} \Omega_i = \Omega_{i+1} \bmod K$. In addition, σ^T restricted to each Ω_i is ergodic with respect to the conditional measure $\mu(\cdot | \Omega_i)$.

Proof

Because σ^T is not ergodic, it has invariant sets of positive measure less than one. Let K be the smallest positive power of σ , which has the same invariant sets. Suppose that the measurable set $U \subset \Omega$ is one such set: $(\sigma^K)^{-1} U = U$ and $0 < \mu(U) < 1$. Then, because σ^i and σ^K commute, each set $(\sigma^i)^{-1} U$ is a σ^K -invariant set, as are the result of any set operation performed on these sets. In particular, $\bigcup_{i=0}^{K-1} \sigma^{-i} U$ is a σ -invariant set of positive measure, and hence

$$1 = \mu \left(\bigcup_{i=0}^{K-1} \sigma^{-i} U \right) \leq \sum_{i=0}^{K-1} \mu(\sigma^{-i} U) = K \mu(U),$$

so that $1 > \mu(U) \geq \frac{1}{K}$. If for some i , the σ^K -invariant sets $U \cap \sigma^i U, U/\sigma^i U$ have positive measure, they must satisfy the same inequality. From the preceding formula, we can obtain $K \mu(U) \geq 1 \geq K \mu(U) - \sum_{i < j} \mu(\sigma^i U \cap \sigma^j U)$. As U cannot be a σ -invariant set, if $\mu(U) > \frac{1}{K}$, it follows that there is a i with $1 \leq i \leq K-1$ such that $0 < \mu(U \cap \sigma^i U), \mu(U/\sigma^i U) < \mu(U)$. Thus, $\mu(U \cap \sigma^i U) \leq \mu(U) - \frac{1}{K}$, so that $U \cap \sigma^i U$ is also a σ^K -invariant set of measure at least $\frac{1}{K}$ less than that of U . Clearly, this process can be iterated a finite number of times to obtain an σ^K -invariant set having the requisite properties. Because $\sigma \Omega_i = \Omega_{i+1} \bmod K$, the requirement $\sigma^T \Omega_i = \Omega_i$ implies $K|T$. Because each Ω_i has no σ^T -invariant subsets of positive measure strictly smaller than $\frac{1}{K}$, $\sigma^T|_{\Omega_i}$ is ergodic. $\square$

Now, consider the function $C : \Omega \times \mathbf{Z} \rightarrow \mathbb{R}^{n \times n}$ defined as

$$C(\omega, t) := \prod_{i=0}^{t-1} A(\mathbf{e}_i(\omega)) = \prod_{i=0}^{t-1} A(e v_0 \sigma^i \omega)$$

which, by stationarity of $\mathbf{e}$, obeys the composition rule $C(\omega, t+s) = C(\sigma^t \omega, s) C(\omega, t)$ with $C(\omega, 0) = I$. Such a function is called a *matrix cocycle*.

Theorem 5.1 (Oseledec's multiplicative ergodic theorem [11, 12, 19])

Let μ be a shift invariant probability measure on Ω and suppose that the shift map $\sigma : \Omega \rightarrow \Omega$ is ergodic and that $\log^+ \|C(\omega, t)\|$ is in L^1 . Then, the limit

$$\Lambda = \lim_{t \rightarrow \infty} (C(\omega, t)'C(\omega, t))^{\frac{1}{2t}} \quad (8)$$

exists with probability one, is symmetric and nonnegative definite, and is μ a.s. independent of ω . Let $\lambda_1 < \lambda_2 < \dots < \lambda_k$ for $k \leq d$ be the distinct eigenvalues of Λ , let U_i denote the eigenspace of λ_i , and let $V_i = \bigoplus_{j=1}^i U_j$. Then, for $u \in V_i - V_{i-1}$,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \|C(\omega, t)u\| = \log(\lambda_i). \quad (9)$$

The numbers λ_i are called the *Lyapunov exponents* of C . Their role is explained by the corollary next.

Corollary 5.1

Let $p \neq \frac{1}{2}$. For $u \in V_i - V_{i-1}$, and for every $\epsilon > 0$ there is a random constant K_ϵ such that

$$\|C(\omega, t)u\| < K_\epsilon(\lambda_i + \epsilon)^t \|u\|. \quad (10)$$

Proof

Note that (9) is equivalent to

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{\|C(\omega, t)u\|}{\|u\|} = \log(\lambda_i),$$

and hence also to

$$\lim_{t \rightarrow \infty} \left| \frac{1}{t} \log \frac{\|C(\omega, t)u\|}{\|u\|} - \log(\lambda_i) \right| = 0$$

which is equivalent to

$$\lim_{t \rightarrow \infty} \frac{1}{t} \left| \log \frac{\|C(\omega, t)u\|}{\|u\|\lambda_i^t} \right| = 0.$$

Now, this means that for every $\delta > 0$, there exists an almost sure finite random $N_\delta > 0$ such that for all $t \geq N_\delta$,

$$\frac{1}{t} \left| \log \frac{\|C(\omega, t)u\|}{\|u\|\lambda_i^t} \right| < \delta,$$

hence, for all $t \geq N_\delta$,

$$\left| \log \frac{\|C(\omega, t)u\|}{\|u\|\lambda_i^t} \right| < \delta t.$$

Let

$$k_\delta := \max_{0 \leq t \leq N_\delta} \left| \log \frac{\|C(\omega, t)u\|}{\|u\|\lambda_i^t} \right|,$$

which is a.s. finite by our assumption that $p \neq \frac{1}{2}$. It follows that for all $t \geq 0$ (it only makes sense to consider nonnegative t) that

$$\left| \log \frac{\|C(\omega, t)u\|}{\|u\|\lambda_i^t} \right| \leq k_\delta + \delta t,$$

from which it follows that for all t

$$\frac{1}{K_\delta} \lambda_i^t e^{-\delta t} \|u\| \leq \|C(\omega, t)u\| \leq K_\delta \lambda_i^t e^{\delta t} \|u\|$$

where $K_\delta := e^{k_\delta}$. Now, define $\epsilon = \lambda_i e^\delta - \lambda_i$, and note that $\lambda_i e^\delta - \lambda_i > \lambda_i - \lambda_i e^{-\delta}$, so that we obtain

$$\frac{1}{K_\delta} (\lambda_i - \epsilon)^t \|u\| < \|C(\omega, t)u\| < K_\delta (\lambda_i + \epsilon)^t \|u\|.$$

Because we can solve for δ as a function of ϵ , we can write K_ϵ instead of K_δ . $\square$

Thus, the Lyapunov exponents of this matrix cocycle control the exponential rate of convergence (or non-convergence) to consensus. We can say several things about these Lyapunov exponents just on the basis of the linear geometry of this problem. This is based on the fact that the matrices $A(e)$ are doubly stochastic as are any matrix products of them, $C(\omega, t)$ in particular. It follows that the subspace of constant functions on V , denoted $\mathbb{R}\mathbf{1}$, as well as the mean zero functions in $\mathbf{1}^\perp$, are invariant under the action of this cocycle *and* of its transpose. Thus, these subspaces are also invariant under the limiting matrix Λ of the Oseledec's theorem, and there is a Lyapunov exponent associated with the subspace $\mathbb{R}\mathbf{1}$ which, it is not difficult to see, is one.

There are also $n-1$ Lyapunov exponents associated with the subspace $\mathbf{1}^\perp$, so characterizing these is now of interest.

For $x \in \mathbb{R}^n$, we shall use the symbol $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$. The main convergence result follows.

Theorem 5.2

Let $G = (V, E)$ be a connected graph and let $\mathbf{e}$ be an ergodic stochastic process taking values in E . Suppose that $p_e > 0$ for all $e \in E'$, where $G' = (V, E')$ is a connected subgraph of G . Let $p \neq \frac{1}{2}$ and assume that the gossip algorithm (1) is initialized at $x(0) = x_0$. Then, there is a (deterministic) constant $|\lambda| < 1$ and a (random) constant K_λ such that

$$\|x(t) - \bar{x}_0 \mathbf{1}\| < K_\lambda \lambda^t \|x_0 - \bar{x}_0 \mathbf{1}\| \quad (11)$$

μ -almost surely.

Proof

The fact that $C(\omega, t)$ is a Markov matrix for each t implies that $C(\omega, t)\mathbf{1} = \mathbf{1}$ and $\|C(\omega, t)\| = 1$, so the largest Lyapunov exponent is equal to 1. Let $(x_0 - \alpha \mathbf{1}) + \alpha \mathbf{1}$ be the orthogonal decomposition of x_0 with respect to the orthogonal direct sum of invariant subspaces $\mathbb{R}^n = \mathbf{1}^\perp \oplus \mathbb{R}\mathbf{1}$. Then, $\alpha = \bar{x}_0$. Moreover, $x(t) = C(\omega, t)x_0$ implies that

$$x(t) - \bar{x}_0 \mathbf{1} = C(\omega, t)|_{\mathbf{1}^\perp} (x_0 - \bar{x}_0 \mathbf{1})$$

hence, by the multiplicative ergodic theorem, there exists a λ such that $\|x(t) - \bar{x}_0 \mathbf{1}\| < K_\lambda \lambda^t \|x_0 - \bar{x}_0 \mathbf{1}\|$. The second largest Lyapunov exponent λ_2 is the greatest lower bound of the set of λ for which this assertion is true. Thus, the rest of the proof amounts to showing that the logarithm of the second largest Lyapunov exponent is less than zero.

Because the edge process $\mathbf{e}(t)$ is ergodic and $p_e > 0$ for all edges in E' , it follows from Lemma 5.1 that there is a deterministic time T , and a positive probability that $\|C(\omega, T)\|_{\mathbf{1}^\perp} < 1$. Let k be the largest multiple of T , which is smaller than t and let $\Delta t := t - kT$. Then, for times $t \geq T$,

$$\begin{aligned} \|C(\omega, t)\|_{\mathbf{1}^\perp} &\leq \|C(\sigma^{kT} \omega, \Delta t)\|_{\mathbf{1}^\perp} \prod_{j=0}^{k-1} \|C(\sigma^{jT} \omega, T)\|_{\mathbf{1}^\perp} \\ &\leq \prod_{j=0}^{k-1} \|C(\sigma^{jT} \omega, T)\|_{\mathbf{1}^\perp}. \end{aligned}$$

Let us consider first the case when σ^T is ergodic. Then, if $\log \|C(\omega, T)\|_{1^\perp} \in L^1$, the Birkhoff theorem implies that a.e.

$$\sum_{i=0}^{k-1} \log \|C(\sigma^{iT}\omega, T)\|_{1^\perp} = k \left(\int \log \|C(\omega, T)\|_{1^\perp} d\mu(\omega) + o(k) \right)$$

and $\lim_{t \rightarrow \infty} \sum_{i=0}^{k-1} \log \|C(\sigma^{iT}\omega, T)\|_{1^\perp} = -\infty$ otherwise. Because $\log \|C(\omega, T)\|_{1^\perp} \leq 0$ always, and $\log \|C(\omega, T)\|_{1^\perp} < 0$ with positive probability, the result then follows from (9), which implies $\mathbb{E} \log \|C(\omega, t)\|_{1^\perp} = \log \lambda_2$.

Suppose, however, that σ^T is not ergodic. Because σ is ergodic, it follows from Lemma 5.2 that there is a finite partition of $\Omega = \bigsqcup_i \Omega_i$ by disjoint σ^T -invariant sets $\{\Omega_i\}_{i=0}^{K-1}$ such that K divides T , $\sigma \Omega_i = \Omega_{i+1} \bmod K$ and $\mu(\Omega_i) = \frac{1}{K}$ for $i = 0, \dots, K-1$. In addition, $(\sigma^T, \mu(\cdot|_{\Omega_i}))$ is ergodic, and $\hat{C}(\omega, \tau) := C(\omega, \tau T)$ is a cocycle for $\sigma^T|_{\Omega_i}$. Moreover, there is an index, say j , such that $\int_{\Omega_j} \log \|C(\omega, T)\|_{1^\perp} d\mu(\omega|_{\Omega_j}) < 0$, and repeating the earlier reasoning, we have that on Ω_j the largest Lyapunov exponent λ of $\hat{C}$ is negative. But, for each nonnegative $i < K$, there is a time $s(i) < T$ such that $\sigma^{s(i)}\Omega_i = \Omega_j$. Thus, for $\omega \in \Omega_i$,

$$\begin{aligned} C(\omega, t) &= C \left(\sigma^{\lfloor \frac{t-s(i)}{T} \rfloor T + s(i)} \omega, t - s(i) - \left\lfloor \frac{t-s(i)}{T} \right\rfloor T \right) \\ &\quad \times \hat{C} \left(\sigma^{s(i)} \omega, \left\lfloor \frac{t-s(i)}{T} \right\rfloor T \right) C(\omega, s(i)), \end{aligned}$$

so that $\|C(\omega, t)\|_{1^\perp} \leq \left\| \hat{C} \left(\sigma^{s(i)}, \left\lfloor \frac{t-s(i)}{T} \right\rfloor T \right) \right\|_{1^\perp} \leq K \lambda^{t/T}$ where K is a constant. $\square$

6. EXCHANGEABLE EDGE PROCESSES

Ergodicity is not necessary for convergence. There is a class of non-ergodic edge processes for which convergence can be shown quite directly. These are the *exchangeable processes*. The edge process $\{\mathbf{e}(t)\}$ is called *exchangeable* if the equality in distribution

$$\{\mathbf{e}(1), \mathbf{e}(2), \dots, \mathbf{e}(n), \dots\} \stackrel{d}{=} \{\mathbf{e}(\pi(1)), \mathbf{e}(\pi(2)), \dots, \mathbf{e}(\pi(n)), \dots\}$$

holds for each finite permutation π of $\{1, 2, \dots, n, \dots\}$, that is, each permutation for which $\#\{t : \pi(t) \neq t\} < \infty$. By this property, any finite sequence of edges of the graph has equal probability of being selected. For connected graphs, which possess a certain degree of symmetry, this is a simple and quite natural random communication mechanism.

It is well known [20, 21] that any exchangeable process is a mixture of i.i.d. processes in the sense that the random variables $\{\mathbf{e}(1), \mathbf{e}(2), \dots, \mathbf{e}(n), \dots\}$ are conditionally independent given a certain σ -algebra $\mathcal{F}$ and

$$P(\mathbf{e}(i) \in A_i, 1 \leq i \leq n | \mathcal{F}) = \prod_i P(\mathbf{e}(i) \in A_i | \mathcal{F}) \quad (12)$$

where $P(\mathbf{e}(i) \in A | \mathcal{F}) := \mu(A | \mathcal{F})$, called the *directed conditional measure* of the process, does not depend on i . It is known that the smallest σ -algebra for which (12) holds is the tail σ -algebra $\mathcal{T}$ of the process. When $\mathcal{T}$ is trivial, then $\mathbf{e}$ is just an i.i.d. process. Conditions under which $\mathcal{T}$ is finite have been investigated by Finesso and Prosdociami [22]. In this case, the probability space admits a finite partition $\{\Omega_i : i = 1, \dots, N\}$ with induced probabilities $\mu_i := \mu(\cdot | \Omega_i)$, which are ergodic, each being the distribution of an i.i.d. process. In general, however, i ranges on an uncountable set.

Hence, for exchangeable edge processes, a similar asymptotic result to Theorem 5.2 holds and $x(t) \rightarrow \bar{x}_0 \mathbf{1}$ exponentially fast although in this case, the Lyapunov exponent λ depends on the particular set of the partition the process is started from.

7. CONTINUOUS TIME RESULTS

Our continuous-time model is very similar; we suppose that evolution toward consensus is effectuated by the differential equation $\dot{x}(t) = A(\mathbf{e}_t)x(t)$, where $A(e) = -c(1_{v_i} - 1_{v_j})(1_{v_i} - 1_{v_j})'$ for $e = v_i v_j$, c is a parameter that determines the rate of convergence to consensus of the coordinates x_i and x_j , $\mathbf{e}$ denotes a non-colliding edge set and $A(\mathbf{e}) = \sum_{e \in \mathbf{e}} A(e)$. Each matrix $A(\mathbf{e})$ is symmetric, nonpositive on the diagonal and nonnegative off, and has rows that sum to zero. The value $-2c$ is a simple eigenvalue associated to the eigenvector $(1_{v_i} - 1_{v_j})$, and the codimension one subspace $(1_{v_i} - 1_{v_j})^\perp$ is the eigenspace of the eigenvalue 0, that is, is the null space of $A(e)$.

Let $\omega = (\mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_k, \dots)$ denote the sequence of edge sets representing communication between nodes, and let $\mathbb{L} = (l_0, l_1, \dots, l_k, \dots)$ be the sequence of lengths of time intervals these edge sets were occupied, and let $t_k = \sum_{j=0}^k l_j$ denote the instant at which the continuous time process $\mathbf{e}_t$ changes value from $\mathbf{e}_k$ to $\mathbf{e}_{k+1}$. In the discrete time case, all of the l_k can be taken to be 1. In the continuous time case, they may take real values between 0 and a random positive 'roof' function $\tau : \Omega \rightarrow \mathbb{R}_+$, which we assume has finite expectation with respect to μ . When $t_n \leq t < t_{n+1}$, the solution of the differential equation is of the form

$$x_t = \exp((t - t_n)A(\mathbf{e}_{n+1})) \prod_{i=0}^n \exp(l_i A(\mathbf{e}_i)) x_0. \quad (13)$$

On the set $\Omega \times \mathbb{R}$, we may consider the quotient $\hat{\Omega}$ by the equivalence relation $(\omega, t) \equiv (\sigma\omega, t - \tau(\omega))$; we may also view the quotient space as the set $\{(\omega, t) : 0 \leq t \leq \tau(\omega)\}$, in which each point $(\omega, \tau(\omega))$ is glued to the point $(\sigma\omega, 0)$. The natural action of $\mathbb{R}$ on $\Omega \times \mathbb{R}$ passes to a flow $\sigma^t : \hat{\Omega} \rightarrow \hat{\Omega}$, and the measure

$$\mu_\tau = \frac{\mu \times dt}{\int \tau(\omega) d\mu(\omega)}$$

is invariant and ergodic under σ^t , iff μ is ergodic under σ . The idea here is that there is a continuous time stochastic process on edge sets, which resides in the edge set $\mathbf{e} = ev_0(\omega)$ for a length of time $\tau(\omega)$. From time zero up to time t , the edges $\mathbf{e}_0 = ev_0(\omega)$, $\mathbf{e}_1 = ev_1(\omega)$, $\dots$, $\mathbf{e}_k = ev_k(\omega)$ are occupied for times $l_0 = t_0 = \tau(\omega)$, $l_1 = t_1 - t_0 = \tau(\sigma\omega)$, $\dots$, $t - \sum_{i=0}^{k-1} \tau(\sigma^i \omega)$ if t satisfies $\sum_{i=0}^{k-1} \tau(\sigma^i \omega) < t \leq \sum_{i=0}^k \tau(\sigma^i \omega)$. It is well known that if the distribution of $\tau(\omega)$ is exponential with respect to the conditional distribution $\mu(\cdot | e)$, and if μ is a Markov measure, then the edge process is a continuous time Markov chain.

Lemma 7.1

Let $G' = (V, E')$ and π be as discussed in Lemma 4.1, and let $\mathbf{t} = (l_1, \dots, l_{m'})$ be strictly positive. Let $B(E', \pi, \mathbf{t}) = \prod_{i=1}^{m'} \exp(t_i A(e_{\pi i}))$, where the product is ordered from right to left. Then, $\|B(E', \pi, \mathbf{t})\|_{1^\perp} < 1$.

Proof

The proof is exactly as in the proof of Lemma 4.1. □

A continuous time matrix cocycle on $\hat{\Omega}$ is a function $C : \hat{\Omega} \times \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$ such that $C(\omega, 0) = I$ and $C(\omega, t + s) = C(\sigma^t \omega, s) C(\omega, t)$.

In the continuous time case, we have

$$C(\omega, t) = \exp \left(\left(t - \sum_{i=0}^{n(\omega)} \tau(\sigma^i \omega) \right) A(e_{v_0} \sigma^{n(\omega)+1} \omega) \right) \prod_{i=0}^{n(\omega)} \exp(\tau(\sigma^i \omega) A(e_{v_0} \sigma^i \omega)),$$

where $n(\omega)$ is the largest integer n such that $t - \sum_{i=0}^n \tau(\sigma^i \omega) > 0$. The continuous time analog of Theorem 5.2 is as follows.

Theorem 7.1

Let $G = (V, E)$ be a connected graph and let $\mathbf{e}$ be an ergodic continuous time stochastic process taking values on E . Suppose that the support of the membership function defined by $\mathbf{e}$ is all of E . Let the gossip algorithm (13) be initialized at $x(0) = x_0$. Then, there is a (deterministic) constant $\lambda < 0$ and a (random) constant K such that $\|x(t) - \bar{x}_0 \mathbf{1}\| < K e^{\lambda t} \|x_0 - \bar{x}_0 \mathbf{1}\|$ μ -almost surely.

8. MULTIPLE GOSSIP AND PACKET DROPS

One may envisage situations where m ($1 < m \leq n$) edges can exchange information in the same time slot, see for example [23], although the coordinate pairs x_i, x_j still evolve according to the algorithm (1). If there is no packet drop when a node receives more than one call, this setting is equivalent to simple gossiping on a m -fold Cartesian product graph $G^m := (V^m, E^m)$ where V^m is the m -fold Cartesian product of the set of vertices of the graph G identified by a fixed arbitrary ordering and E^m is the set of admissible m -tuples of edges, $\mathbf{e} := \{e_1, e_2, \dots, e_m \mid e_i \in E\}$. The A matrix governing the m simultaneous adjustments occurring in one time slot will be a product of the type

$$A(\mathbf{e}) = \prod_{k=1}^m A(e_k) \quad (14)$$

corresponding to a m -tuple of edges $\mathbf{e} := \{e_1, e_2, \dots, e_m\}$ being selected sequentially in that order by an m -dimensional stationary edge process $\mathbf{e}^m$ whose i -th component is governed by the m -th power σ^m of the stationary shift σ of $\mathbf{e}(t)$. Theorem 5.2 still applies. Because the product graph has higher degree, it may be expected that the rate of convergence would be higher than for simple gossiping. This point, however, has not been investigated.

When packet collision occurs, it is necessary to introduce some consistency constraints in the model to account for possible conflicts. For example, communication calls from nodes, which are already busy (or to busy nodes), should result in packet drops and no communication taking place. In other words, it should not be allowed that in the same time slot, two selected edges belong to the same node. Formally, the occurrence of edges $v_i v_j, v_k v_h$ for which either $i = k$ or $j = h$ should have probability zero or, whenever this happens, the updating step of the x_i 's should be canceled. One should assume that all adjustment algorithms (1) active in the same time slot, involve distinct pairs of nodes (and hence distinct edges as well).

Define a subset of edges $E' \subseteq E$ to be *non-colliding* iff no two edges $e, e' \in E'$ are incident on a common vertex. Let E_* denote the collection of all non-colliding subsets of edges $E_* = \{E' \subseteq E \text{ is non-colliding}\}$. Hence, the A matrix governing the m simultaneous adjustments occurring in one time slot will be a product of the type

$$A(\mathbf{e}) = \prod_{k=1}^m A(e_k) \quad (15)$$

corresponding to a m -tuple of non-colliding edges $\mathbf{e} := \{e_1, e_2, \dots, e_m\} \in E_*$. All $A(e_k)$ matrices occurring in the product (15) have the 2×2 non-identity block in disjoint pairs of locations and hence *commute*. Consequently, it is immaterial how the simultaneous communicating links are ordered in (15). In this multiple gossip algorithm, each matrix $A(\mathbf{e})$ is still symmetric and doubly stochastic. The value $1 - 2p$ is an eigenvalue of multiplicity m (both algebraic and geometric), which is associated to the m orthogonal eigenvectors $\{(1_{v_i} - 1_{v_j}) ; v_i v_j \in \mathbf{e}\}$, and the $d - m$ -dimensional subspace $\{(1_{v_i} - 1_{v_j}) ; v_i v_j \in \mathbf{e}\}^\perp$ is the eigenspace of the eigenvalue 1. $A(\mathbf{e})$ has only two distinct singular values, the largest (of multiplicity $d - m$) equal to one and the second (of multiplicity m) equal to $[1 - 4p(1 - p)]^{1/2}$. The latter is, in fact, the second largest singular value of any

of the m factors $A(e_k)$. Hence, as long as $\mathbf{e}$ is not the empty set of edges, $A(\mathbf{e})$ and its elementary factors have the same contraction factor

$$\gamma(A(\mathbf{e})) = \gamma(A(e_k)) = [1 - 4p(1 - p)]^{1/2}.$$

In these models, the edge process $\mathbf{e}$ selects a fixed deterministic number m of admissible edges at each time step. They could be rendered more realistic by allowing m to be random. For instance, a number of vertexes could be chosen at each time step according to some distribution, and each chosen vertex initiates a call to a neighboring vertex by choosing one of its edges.

In this setting, we may interpret $A(\mathbf{e})$ as a conditional quantity given that the occurring number of links is m . This approach leads to a more complicated analysis and we shall not pursue this generalization here. Naturally, m can be interpreted just as the average number of active links per time step.

9. CONCLUSIONS

In this paper, we have provided a proof of almost sure convergence to consensus for an extremely general class of edge selection processes. The proof also applies to continuous time gossip algorithms. Computational schemes for estimating the Lyapunov exponents (and hence the rate of convergence) is an issue, which we plan to address in the near future.

REFERENCES

1. Jadbabaie A, Lin J, Morse AS. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control* 2003; **48**:988–1001.
2. Boyd S, Ghosh A, Prabhakar B, Shah D. Randomized gossip algorithms. *IEEE Transactions on Information Theory* 2006; **52**:2508–2530.
3. Tahbaz-Salehi A, Jadbabaie A. On consensus over random networks. *4th Annual Allerton Conference on Communication, Control and Computing*, Allerton House, University of Illinois, Sept. 2006; 1315–1321.
4. Olfati-Saber R, Fax J. A, Murray R. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE* 2007; **95**:215–233.
5. Tahbaz-Salehi A, Jadbabaie A. Consensus over ergodic stationary graph processes. *IEEE Transactions on Automatic Control* Jan. 2010; **55**(1):225–230.
6. Lorenz J, Lorenz DA. On conditions for convergence to consensus. *IEEE Transactions on Automatic Control* 2010; **55**(7):1651–1656.
7. Fagnani F, Zampieri S. Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas of Communications* May 2008; **26**(4):634–649.
8. Kempe D, Kleinberg J, Demers A. Spatial gossip and resource allocation protocols. *Journal of the ACM* 2004; **51**:943–967.
9. Olfati-Saber R, Murray R. Consensus problems in networks of agents with switching topology and time delays. *IEEE Transactions on Automatic Control* 2004; **49**:1520–1533.
10. Picci G, Taylor T. Almost sure convergence of random gossip algorithms, Proc. of the 46th IEEE Decision and Control (CDC) Conference, New Orleans Lu, Dec 2007; 282–287.
11. Oseledec VI. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trudy Moskovskogo Matematicheskogo Obshchestva* 1968; **19**:179–210.
12. Ruelle D. Ergodic theory on differentiable dynamical systems. *IHES Publicationes Mathematicae* 1979; **50**: 275–320.
13. Wolfowitz J. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society* 1963; **15**:733–736.
14. Cogburn R. The ergodic theory of Markov chains in random environment. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 1984; **66**:66–128.
15. Cogburn R. On direct convergence and periodicity for transition probabilities of Markov chains in random environments. *Annals of Probability* 1990; **18**:642–654.
16. Orey S. Markov chains with stochastically stationary transition probabilities. *Annals of Probability* 1991; **19**:907–928.
17. Nelson S, Neumann M. Generalizations of the projection method with applications to SOR theory for Hermitian positive semidefinite linear systems. *Numerische Mathematik* 1987; **51**:123–141.
18. Walters P. *An Introduction to Ergodic Theory*. Springer Graduate Texts in Math: New York, 1982.
19. Ludwig A. *Random Dynamical Systems*. Springer-Verlag: New York, 1998.

20. de Finetti B. La Prevision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 1937; **7**:1–68.
21. Aldous D. Exchangeability and related topics École d'Été de Probabilités de Saint-Flour XIII - 1983. *Springer Lecture Notes in Mathematics* 1985; **1117**:1–198.
22. Finesso L, Prosdocimi C. Stochastic realization of binary exchangeable processes. *Proc. 19th Int. Symposium on the Mathematical Theory of Networks and Systems (MTNS 2010)*, Budapest, Hungary, 2010; 1859–1864.
23. Sanghavi S, Hajek B, Massoulié L. Gossiping with multiple messages. *IEEE Transactions on Information Theory* 2007; **53**:4640–4654.