Deterministic Gossiping

Gossip algorithms can provide information exchange and computation for autonomous vehicles in a group, where each vehicle must make estimates and decisions, while ensuring consensus at the group level.

By Ji Liu, Student Member IEEE, Shaoshuai Mou, Student Member IEEE, A. Stephen Morse, Life Fellow IEEE, Brian D. O. Anderson, Life Fellow IEEE, and Changbin Yu, Senior Member IEEE

ABSTRACT | For the purposes of this paper, “gossiping” is a distributed process whose purpose is to enable the members of a group of autonomous agents to asymptotically determine, in a decentralized manner, the average of the initial values of their scalar gossip variables. This paper discusses several different deterministic protocols for gossiping which avoid deadlocks and achieve consensus under different assumptions. First considered is \( T \)-periodic gossiping which is a gossiping protocol which stipulates that each agent must gossip with the same neighbor exactly once every \( T \) time units. Among the results discussed is the fact that if the underlying graph characterizing neighbor relations is a tree, convergence is exponential at a worst case rate which is the same for all possible \( T \)-periodic gossip sequences associated with the graph. Many gossiping protocols are request based which means simply that a gossip between two agents will occur whenever one of the two agents accepts a request to gossip placed by the other. Three deterministic request-based protocols are discussed. Each is guaranteed to not deadlock and to always generate sequences of gossip vectors which converge exponentially fast. It is shown that worst case convergence rates can be characterized in terms of the second largest singular values of suitably defined doubly stochastic matrices.

KEYWORDS | Consensus; distributed averaging; nonhomogeneous Markov chains; stochastic matrices

I. INTRODUCTION

There has been considerable interest recently in developing algorithms for distributing information among the members of a group of sensors or mobile autonomous agents via local interactions. Notable among these are those algorithms intended to cause such a group to reach a consensus in a distributed manner [1]–[7]. One particular type of consensus process which has received much attention lately is called distributed averaging [8]. In its simplest form, distributed averaging deals with a network of \( n > 1 \) agents and the constraint that each agent is able to communicate only with certain other agents called agent \( i \)’s neighbors. Neighbor relations are described by a simple, connected graph \( N \) in which vertices correspond to agents and edges indicate neighbor relations. Thus, the neighbors of an agent \( i \) have the same labels as the vertices in \( N \) which are adjacent to vertex \( i \). Initially, each agent \( i \) has or acquires a real number \( y_i \) which might be a measured temperature or something similar. The distributed averaging problem is to devise a protocol which will enable each agent to compute the average \( y_{avg} = (1/n) \sum_{i=1}^{n} y_i \) using only information acquired from its neighbors. There are many variants of this problem. For example, instead of real numbers, the \( y_i \) may be integer-valued [9]. Another variant assumes that the edges of \( N \) change over time [10]. This paper considers the case when the \( y_i \) are real numbers and \( N \) does not depend on time.

As noted in [8], the distributed averaging problem can be solved, in principle, by “flooding”; that is, by propagating across the network over time the values of all of the \( y_i \). Armed with knowledge of all of these values, each agent
is thus able to compute $y_{avg}$. A more sophisticated approach to the problem is for each agent to use a linear iterative update rule of the general form

$$x_i(t + 1) = w_{ij}x_i(t) + \sum_{j \in \mathcal{N}_i} w_{ij}x_j(t), \quad x_i(1) = y_i$$

where $t$ is a discrete time index, $x_i(t)$ is agent $i$'s current estimate of $y_{avg}$, the $w_{ij}$ are real-valued weights, and $\mathcal{N}_i$ is the set of labels of the neighbors of agent $i$. In [8], several methods are proposed for choosing the $w_{ij}$. One particular choice, which defines what has come to be known as the Metropolis algorithm, requires only local information to define the $w_{ij}$. Algorithms of this type, which require each agent to communicate with all of its neighbors on each iteration, are sometimes called broadcast algorithms.

An alternative approach to distributed averaging, which typically does not involve broadcasting, exploits a form of “gossiping” [11] specifically tailored to the distributed averaging problem. The idea of gossiping is very simple. A pair of neighbors with labels $i$ and $j$ are said to gossip at time $t$ if both $x_i(t + 1)$ and $x_j(t + 1)$ are set equal to the average of $x_i(t)$ and $x_j(t)$. Each agent is allowed to gossip with at most one neighbor at one time. Under appropriate assumptions, algorithms which possess this simple property can be shown to solve the distributed averaging problem. Gossiping algorithms do not necessarily involve broadcasting and thus have the potential to require less transmissions per iteration than broadcast algorithms. Of course, one would not expect gossip algorithms to converge as fast as broadcast algorithms.

Gossiping might find application in many contexts. For example, suppose that a spatially distributed network of temperature sensors has been deployed in such a way so that each sensor can communicate with nearby sensors (i.e., neighbors). Suppose that at some specific time $t_0$ all sensors take readings and use these readings as initial estimates of the average temperature across the network. At subsequent clock times, each sensor then passes its current estimate to one of its neighbors who in turn uses this estimate to update its own estimate with the goal of ultimately arriving at the average value of the temperature across the network. Gossiping is a process for recursively carrying out these computations.

Implementation of any gossiping protocol necessarily involves some degree of centralization. In particular, for the aforementioned sensor network averaging task to make sense each sensor must be instructed by a centralized manager to take a temperature reading at the same time $t = t_0$ as the rest. In some cases, it may be useful to take centralization further. For example, to facilitate gossiping it may be helpful in some instances to centrally define a network-wide sensor ordering by assigning offline to each sensor a unique priority number with the understanding that each sensor can make use of the priority numbers of its neighbors to carry out its role in the gossiping process [12]. Another idea requiring some degree of centralization is to assign offline to each agent, a specific sequence of times at which the agent may gossip (Section IV-A). One might carry centralization one step further by specifying offline for each agent which neighbor the agent is to gossip with at each clock time (Section III). This of course could be done only in a network whose population of sensors does not change over time in a predictable manner. For networks whose members change with time, one might consider other ideas and assumptions. For example, for the temperature sensing network, one might try implementing a gossiping protocol which assumes that at each clock time each sensor can acquire the current temperature estimates of all of its neighbors [13]. A refinement of this protocol which requires each agent to acquire the current temperature of only a subset of its neighbors at each clock time is discussed in Section IV-A. In the end, which assumptions make sense depends on the specific application.

Although gossiping is a form of consensus, it differs from consensus in several important ways. First, the goal of consensus is to agree on the value of some quantity whereas the goal of gossiping is to compute the average of the initial values of the $x_i$, henceforth called gossip variables. Second, unlike consensus processes, gossiping processes are invariably designed so that the sum total of all gossip variables remains constant from clock time to clock time. This has a simple but important consequence: If the sum total of all gossip variables remains constant and a consensus is reached in that all gossip variables converge to the same value, then this value must be the average of the initial values of all gossip variables in the network. In gossiping, the sum total of all gossip variables is kept constant by requiring agents to always gossip in pairs using averaging. Although this simple idea keeps constant the sum of gossip variables across the network, the idea comes with a price in that a deadlock may well occur unless specific precautions are built into the protocol to preclude this.

The specific sequence of gossip which occurs during a given gossiping process might be determined either probabilistically [11], [14] or deterministically [12], [15], depending on the problem of interest. Deterministic gossiping protocols are intended to guarantee that under all conditions, a consensus will be achieved asymptotically whereas probabilistic protocols aim at achieving consensus asymptotically with probability one. Both approaches have merit. The probabilistic approach is typically somewhat easier both in terms of algorithm development and convergence analysis. On the other hand, the deterministic approach forces one to consider worst case scenarios and has the potential of yielding algorithms which may outperform those obtained using the probabilistic approach. This paper takes the deterministic approach.

Of particular interest is the rate at which a sequence of agent gossip variables converge to a common value. The convergence rate question for more general deterministic
consensus problems has been studied in [16] and [17]. In [11] and [14], the convergence rate question is addressed for gossiping algorithms in which the sequence of gossip pairs under consideration is determined probabilistically. A modified gossiping algorithm intended to speed up convergence is proposed in [18] without proof of correctness, but with convincing experimental results. The algorithm has recently been analyzed in [19]. Recent results concerning convergence rates appear in [15], [20], and [21] for periodic gossiping and in [22]–[24] for deterministic periodic gossiping. This paper presents a more comprehensive treatment of the work in [15] and [24].

A typical gossiping process can usually be modeled as a discrete time linear system of the form $x(t + 1) = M(t)x(t)$, $t = 1, 2, \ldots$, where $x$ is a vector of agent gossip variables $x_i$ and each value of $M(t)$ is a specially structured doubly stochastic matrix (Section II). Roughly speaking, a finite sequence of gossip pairs is “complete” if the corresponding set of edges in $\mathbb{N}$ forms a connected spanning subgraph. A complete gossip sequence is minimally complete if there is no other complete gossip sequence of shorter length (Section II-B). An infinite sequence of gossiping is “repetitively complete” with period $T$ if each successive subsequence of gossip of length $T$ in the sequence is complete. The gossip variable sequences associated with repetitively complete gossip sequences converge exponentially fast (Section II-C). Repetitively complete gossip sequences which are also periodic with period $T$ are treated in Section III. The worst case convergence rate of any such sequence is determined by $T$ and by the second largest eigenvalue (in magnitude) of the stochastic matrix which the gossipes define over a period. In the case when $\mathbb{N}$ is a tree and the sequence of gossipes over a period is minimally complete, the value of this eigenvalue does not depend on the order in which gossipes over a period take place. A proof of this surprising result is given in [25].

Most gossiping protocols are “request-based.” By request-based gossiping is meant a gossiping process in which a gossip occurs between two agents whenever one of the two accepts a request to gossip placed by the other (Section IV-A). An agent’s “event times” are the times at which the agent makes requests to gossip. In Section IV-A, a request-based protocol is given which generates repetitively complete (and thus exponentially convergent) gossip sequences under the assumption that the event times of each agent are different than the event times of all of its neighbors. A more refined repetitively complete gossiping protocol not requiring this assumption is discussed in Section IV-A. The protocol is inspired by ideas put forth in [13].

It is shown in Section IV-C that the worst case convergence rate of a repetitively complete gossip sequence with period $T$ can be characterized in terms of a suitably defined seminorm of the stochastic matrix $S$ determined by the subsequence of gossipes occurring over a given period. A specific goal of this paper is to find a seminorm with respect to which $S$ is a contraction. The role played by seminorms in characterizing convergence rate is explained in Section IV-C. Three different types of seminorms are considered in Section IV-C. Each is compared to the well-known coefficient of ergodicity which plays a central role in the study of convergence rates for nonhomogeneous Markov chains [26]. Somewhat surprisingly, it turns out that a particular Euclidean seminorm on $\mathbb{R}^{n \times n}$ has the required property—namely that in this seminorm, the stochastic matrix $S$ determined by any complete gossip sequence is a contraction (Section IV-D). This particular seminorm turns out to be the second largest singular value of $S$.
A. Doubly Stochastic Matrices

Each single-gossip matrix is a nonnegative matrix whose row sums and column sums all equal one. Matrices with these two properties are called doubly stochastic. Note that the type of doubly stochastic matrices which characterize single gossips (i.e., single-gossip matrices) has two additional properties—it is symmetric and its diagonal entries are all positive. The same is true for the type of doubly stochastic matrices which characterize multigossips. Doubly stochastic matrices are special types of “stochastic matrices” where by a stochastic matrix is meant a nonnegative $n \times n$ matrix whose row sums all equal one. It is easy to see that a nonnegative matrix $S$ is stochastic if and only if $S1 = 1$ where $1 \in \mathbb{R}^n$ is a column vector whose entries are all ones. Similarly, a nonnegative matrix $S$ is doubly stochastic if and only if $S1 = 1$ and $S1 = 1$. Using these characterizations it is easy to prove that the class of stochastic matrices in $\mathbb{R}^{n \times n}$ is compact and closed under multiplication as is the class of doubly stochastic matrices in $\mathbb{R}^{n \times n}$. It is also true that the class of nonnegative matrices in $\mathbb{R}^{n \times n}$ with positive diagonal entries is closed under multiplication.

Mathematically, reaching a consensus by means of an infinite sequence of gossips or multigossips modeled by a corresponding infinite sequence of gossip matrices $M(1), M(2), M(3), \ldots$ means that the sequence of matrix products $M(1), M(2)M(1), M(3)M(2)M(1), \ldots$ converges to a matrix of the form $1c$. It turns out that if convergence occurs, the limit matrix $1c$ is also a doubly stochastic matrix; this means that $c = (1/n)1^T$ and consequently that all $n$ gossip variables will have converged to the average of their initial values. This particular fact further distinguishes a gossiping process from a more general consensus process, since in a consensus process the value to which all consensus variables typically converge is not necessarily the average of their initial values.

B. Gossiping Sequences

By a gossiping sequence is meant a sequence of individual gossips corresponding to some or all of the edges in a given neighbor graph $N$. Corresponding to any given sequence of gossips $(i_1, j_1), (i_2, j_2), \ldots$ is a sequence of single-gossip matrices $P_{i_1j_1}, P_{i_2j_2}, \ldots$ whose product $\cdots P_{i_tj_t}P_{i_{t-1}j_{t-1}} \cdots$ defines the mapping which assigns to any given initial vector of gossip variables, the vector of gossip variables which results from the gossips in the sequence. We call any such matrix product a gossip matrix. It is thus clear that a given neighbor graph has associated with it a family of gossip matrices whose members are all products of all combinations of single-gossip matrices of all lengths. These are the gossip matrices determined by $N$. Conversely, any given sequence of individual gossips (or corresponding product of single-gossip matrices) induces a spanning subgraph of $N$ whose edges correspond to the gossips in the sequence. We say that a gossip sequence or corresponding gossip matrix is complete if the graph the gossips in the sequence induce is a connected spanning subgraph within $N$. A gossip sequence and corresponding gossip matrix is minimally complete, if it is complete and if there is no other complete gossip sequence of shorter length. It is easy to see that a nonredundant gossip sequence is minimally complete if and only if the subgraph of $N$ that it induces is a minimal spanning tree in $N$. In the special but important case when $N$ is itself a tree $T$, more can be said. In this case, a minimally complete gossip sequence is one in which, for each edge in $T$, there is exactly one corresponding individual gossip in the sequence.

The preceding ideas extend in a natural way to sequences of multigossips. Corresponding to any given sequence of multigossips $\gamma_1, \gamma_2, \ldots$ is a sequence of gossip matrices $Q_1, Q_2, \ldots$ where $Q_t$ is the product of the single-gossip matrices of the individual gossips in the $t$th multigossip in the sequence. The product $\cdots Q_2Q_1$ thus defines the mapping which assigns to any given initial vector of gossip variables the vector of gossip variables which results from the multigossips in the sequence. Clearly any such matrix product is also a product of single-gossip matrices and thus is a bona fide gossip matrix.

Extending the concept of nonredundancy, we say that a multigossip sequence is nonredundant if no individual gossip occurs in more than one multigossip in the sequence. Nonredundant multigossip sequences are clearly of finite length in that the length of each is no larger than the number of edges of $N$. The graph induced by a multigossip sequence $\Sigma$, written $N_{\Sigma}$, is the spanning subgraph of $N$ whose edges correspond to all of the gossips in all of the multigossips in the sequence. A multigossip sequence

\[ A \text{ gossip sequence is nonredundant if each gossip in the sequence occurs at most once.} \]
is complete if the graph $N\Sigma$ which it induces is a connected spanning subgraph of $N$. $\Sigma$ is minimally complete if it is complete and if the sum total of all the single gossips in all the multigossips in the sequence is no larger than the sum total of all the single gossips in all the multigossips in any other complete multigossip sequence. It is clear that if $\Sigma$ is minimally complete, then the subgraph $N\Sigma$ it induces must be a minimal spanning tree. On the other hand, if $\Sigma$ is nonredundant and $N\Sigma$ is a minimal spanning tree of $N$, then $\Sigma$ must be minimally complete.

**C. Convergence**

Roughly speaking, if over a period $T$ a complete multigossip sequence has occurred, then each agent in the group will have been “in touch” with each other agent at least indirectly. It is not surprising then that complete multigossip subsequences over successive periods in an infinitely long sequence should be sufficient for all gossip variables in a gossiping process to converge to a common value. Prompted by this, let us call an infinite sequence of multigossips repetitively complete with period $T$ if each successive subsequence of multigossips of length $T$ in the sequence is complete. The following theorem implies that repetitive complete multigossip sequences converge exponentially fast.

**Theorem 1:** Let $M(1), M(2), M(3), \ldots$ denote the gossiping matrices corresponding to an infinite sequence of multigossips which is repetitively complete with period $T$. Suppse that the vector of gossip variables $x(t)$ evolves according to $x(t + 1) = M(t)x(t), t \geq 1$. There exists a real nonnegative number $\lambda < 1$ such that for each initial value of $x(1)$, all $n$ gossip variables converge to the average value

$$\frac{1}{n} \sum_{i=1}^{n} x_i(1)$$

as fast as $X$ converges to zero.

There are several different ways to prove this theorem; see, for example, [3], [6], [7], and [23]. It turns out that this theorem is a simple consequence of more technical results which are of interest in their own right and which will be stated and proved in Section IV.

The theorem also applies to more general gossiping algorithms in which simple averaging between agent pairs is replaced with averaging based on convex combinations. All that is required is that the averaging rule determines doubly stochastic matrices $M(\cdot)$ which take values in a compact set.

**D. Tree Graphs**

In graph theory, tree graphs (i.e., graphs without cycles) often lead to significant simplifications. This is also the case with gossiping. Let us note that a tree graph has the property that removal of any one edge $(i, j)$ results in a disconnected graph. This means that if $N$ is a tree, a necessary condition for a finite sequence of gossips to be complete is that over the period during which the gossips occur, each agent must gossip with each of its neighbors at least once. It is clear that the converse is also true; i.e., if each agent gossips with each of its neighbors at least once during a given period, then the sequence of gossips which took place over that period must be complete. It is easy to see that in the case when $N$ is a tree, a gossip sequence is minimally complete if and only if a gossip between each agent and each of its neighbors occurs exactly once in the sequence. Equivalently a gossip matrix $G$ for a graph $N$, which is a tree, is minimally complete if and only if $G$ is a product of all of the single-gossip matrices associated with $N$.

**III. PERIODIC GOSSIPING**

About the easiest way to guarantee a repetitively complete multigossip sequence is to use a protocol which generates an infinite multigossip sequence which on the one hand is “periodic” and on the other is complete on each successive period. Prompted by this, let us agree to call an infinite sequence of multigossips periodic with period $T$ if each multigossip in the sequence occurs once every $T$ time units; such a sequence is periodically complete if each subsequence consisting of $T$ consecutive multigossips is complete. It is clear that any periodically complete multigossip sequence is repetitively complete. The converse of course is not true.

**A. Convergence Rate**

Corresponding to any $T$-periodic sequence of multigossips is an infinite sequence of gossip matrices; such a matrix sequence is periodic with period $T$ in that each matrix within the sequence repeats itself every $T$ time units. Suppose that $M(1), M(2), \ldots$ is such a $T$-periodic sequence. If $x(t + 1) = M(t)x(t), t \geq 1$, it is clear that $x((i + 1)T + 1) = Nx((iT + 1), \ i \geq 0$, where $N = M(T)M(T - 1) \ldots M(1)$. Thus, $x((iT + 1) = N^i x(1), \ i \geq 0$, which means that both the convergence and convergence rate of the gossip variables in a periodic multigossip sequence are completely determined by properties of $N$. Note that $N$ is a doubly stochastic matrix because each of the matrices in the product defining it is doubly stochastic, and because the class of $n \times n$ doubly stochastic matrices is closed under multiplication. Now because $N$ is stochastic, it has an eigenvalue at 1 and its spectral radius is 1 [27]. We are interested in the case when $\lim_{i \rightarrow \infty} N^i = (1/n) 11^T$ which is clearly just when all eigenvalues other than the one eigenvalue with

---

2If this were not so then there would have to be at least two distinct paths between $i$ and $j$ which contradicts the requirement that a tree be acyclic.
value 1 have magnitudes strictly less than 1. This is precisely the property of a complete gossip matrix.

Theorem 2: A gossip matrix is complete if and only if the magnitudes of all of its eigenvalues, with the exception of a single eigenvalue of value 1, are strictly less than 1.

A proof of this theorem will be given in Section IV-D.

For any doubly stochastic matrix $S$, let $\rho(S)$ denote the magnitude of the second largest eigenvalue (in magnitude) of $S$. It is clear that the rate at which $x$ converges to $1_{\text{avg}}$ is $\rho^{1/T}(N)$. In the case when $N$ is a tree, $\rho(N)$ turns out to be the same for all minimally complete gossip matrices determined by $N$. This somewhat surprising fact is a direct consequence of the following theorem which is the main result of [25].

Theorem 3: Let $\mathcal{E} = e_1, e_2, \ldots, e_k$ be the sequence of edges of $N$ labeling a nonredundant, complete gossip sequence. Let $N_{\mathcal{E}}$ denote the spanning subgraph of $N$ whose edges are the edges in $\mathcal{E}$. Let $G(\mathcal{E})$ be the group (under composition) consisting of the identity map on $\{1, 2, \ldots, k\}$ together with all maps $\pi : \{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\}$ generated by all permutations which satisfy one of the following conditions:
1) $\pi$ is a cyclic permutation of $\{1, 2, \ldots, k\}$;
2) $\pi$ is a permutation of $\{1, 2, \ldots, k\}$ which for some $i < k$, interchanges $i$ and $i + 1$ provided that in $N_{\mathcal{E}}$, either
   a) $e_i$ and $e_{i+1}$ are not incident on the same vertex or
   b) $e_i$ and $e_{i+1}$ are incident on the same vertex but neither edge is contained in any cycle of $N_{\mathcal{E}}$.

For each $\pi \in G(\mathcal{E})$, let $G_\pi$ denote the gossip matrix induced by the edge sequence $e_{\pi(1)}, e_{\pi(2)}, \ldots, e_{\pi(k)}$. Then, the characteristic polynomial of $G_\pi$ is the same for all $\pi \in G(\mathcal{E})$.

A proof of this theorem can be found in [25]. See also [21] for an alternative proof.

Note that if $N_{\mathcal{E}}$ is a tree, then condition 2 in Theorem 3 implies that for any two successive integers $i$ and $i + 1$ in $\{1, 2, \ldots, k\}$, there is a permutation in $G(\mathcal{E})$ which interchanges $i$ and $i + 1$. A simple induction thus proves that if $N_{\mathcal{E}}$ is a tree, $G(\mathcal{E})$ is the set of all permutations on $\{1, 2, \ldots, k\}$. Therefore, in this case, the characteristic polynomial of $G_\pi$ is the same for all permutations of $\{1, 2, \ldots, k\}$.

Suppose that $N$ is a tree. Then, because of completeness, $N_{\mathcal{E}} = N$, and therefore, $G(\mathcal{E})$ is the set of all permutations on $\{1, 2, \ldots, k\}$. Thus, if $N$ is a tree, Theorem 3 implies that $\rho(N)$ is the same for all minimal complete gossip matrices determined by $N$. This conclusion is not implied by Theorem 3 if $N$ is not a tree.

As stated, the theorem is only for sequences of single gossips. However, the same theorem also applies, with virtually the same proof, to sequences of multigossips. This is because for purposes of analysis, any multigossip can be viewed as a sequence of noninteracting single gossips arranged in any order.

B. Multigossiping
It is clear from the preceding that the rate at which the gossip variables of a periodically complete gossiping sequence converge depends not only on $\rho(N)$ but also on $T$. For example, suppose that $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \ldots, \gamma_T, \gamma_1, \gamma_2, \ldots$ is an infinite periodically complete gossip sequence with period $T$. Suppose in addition that $\gamma_1, \gamma_2, \gamma_3$ are noninteracting gossips. Then, these three gossips might be executed simultaneously as a multigossip $\{\gamma_1, \gamma_2, \gamma_3\}$, rather than sequentially, at the beginning of each period without in any way affecting the complete gossip matrix $N$ corresponding to the original subsequence $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \ldots, \gamma_T$. In other words, rather than executing the $T$-periodic sequence $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \ldots, \gamma_T, \gamma_1, \gamma_2, \ldots$, the group could execute the periodic sequence $\{\gamma_1, \gamma_2, \gamma_3\}, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8, \gamma_9, \ldots$ without changing the value of $\rho(N)$. The key point here is that this sequence has period $T - 2$ rather than $T$. Thus, by using multigossiping, the worst case convergencerate for this gossiping process would be reduced from $\rho^{1/T}(N)$ to $\rho^{1/(T-2)}(N)$. It is obvious that, in general, to get faster convergence, one would want to implement multigossiping sequences using the smallest number of distinct multigossips possible. For the case when $N$ is a tree and the original subsequence $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \ldots, \gamma_T$ is minimally complete, we know that the order of the gossips in the sequence can be changed without changing $\rho(N)$. In this case, the minimal number of multigossips needed to implement the original sequence would be the same as the minimal number of colors needed to color the edges of $N$ subject to the constraint that no two edges incident on any vertex have the same color, for edges of the same color would then correspond to those gossips which could be implemented together as a single multigossip. Edge coloring is a basic problem in graph theory [28]. The minimal number of colors required to color a graph subject to this constraint is called the chromatic index. Vizing’s theorem states that the chromatic index of a neighbor graph $N$ is either $d$ or $d + 1$ where $d$ is the maximum vertex degree of $N$ [29]. Moreover, if $N$ is a tree, the chromatic index is $d$ because of König’s theorem [28]. In other words, if $N$ is a tree with maximum vertex degree $d$, it is possible to construct a periodic sequence of multigossips with period $T = d$ which converges as fast as the sequence $\rho^{1/d}(N)$, $\rho^{2/d}(N), \rho^{3/d}(N), \ldots$ converges to zero where $N$ is any minimally complete gossip matrix for the graph.

IV. REQUEST-BASED GOSSIPING
Request-based gossiping is a gossiping process in which a gossip occurs between two agents whenever one of the two accepts a request to gossip placed by the other. The aim of this section is to discuss this process.

In a request-based gossiping process, a given agent $i$ may gossip with one of its neighbor’s at time $t$ only if it is either an “event time” of agent $i$ or an “event time” of its
neighbor which has made a request to gossip with agent $i$. By an event time of agent $i$ is meant a time at which agent $i$ may place a request to gossip with one of its neighbors. By an event time interval of agent $i$ is meant the interval of time between two successive event times of agent $i$. For obvious reasons, we assume that the lengths of agent $i$’s event time intervals are all bounded above by a finite positive number $T_i$. We write $T_i$ for the set of event times of agent $i$ and $T$ for the union of the event time sequences of all $n$ agents.

Conflicts leading to deadlocks can arise if an agent who has placed a request to gossip, at the same time receives a request to gossip from another agent. It is challenging to devise rules which resolve such conflicts while at the same time ensuring exponential convergence of the gossiping process. One way to avoid such conflicts is to assign event times offline so that no agent can receive a request to gossip at any of its own event times. There are several ways to do this which will be discussed below.

From time to time, agent $i$ may have more than one neighbor to which it might be able to make a request to gossip with. Also from time to time, agent $i$ may receive more than one request to gossip. While in such situations decisions about who to place requests with or whose request to accept can be randomized, in this paper, we will examine only completely deterministic strategies. To do this we will assume that each agent has ordered its neighbors in $N_i$ according to some priorities so when a choice occurs between neighbors, agent $i$ will always choose the one with highest priority.

Consider first the situation when the event times of each agent and each agent’s neighbor priorities are chosen offline and are fixed throughout the gossiping process. Assume that the event times are chosen so that no agent can receive a request to gossip at any of its own event times. Our aim is to show that this arrangement can be problematic. The following protocol illustrates this.

Protocol I: At each event time $t \in T$ the following rules apply for each $i \in \{1, 2, \ldots, n\}$.
1) If $t \in T_i$, agent $i$ places a request to gossip with that neighbor whose priority is the highest.
2) If $t \not\in T_i$, agent $i$ does not place a request to gossip.
3) Each agent $i$ receiving one or more requests to gossip must gossip with that requesting neighbor whose priority is the highest.
4) If $t \not\in T_i$ and agent $i$ does not receive a request to gossip, it does not gossip.

The following example shows that this simple strategy will not necessarily lead to a consensus. Suppose that $N$ is a path graph with edges $(a, b), (b, c), (c, d)$. Assume that agents $a$ and $b$ have distinct event times and that agents $a$ and $c$ have the same event times as do agents $b$ and $d$; note that this guarantees that no agent can receive a request to gossip at any of its own event times. To avoid ambiguities in decision making, suppose that agent $b$ assigns a higher priority to $a$ than to $c$ and agent $c$ assigns a higher priority to $d$ than to $b$. Let $t$ be an event time of agents $a$ and $c$. Then, at this time $a$ places a request to gossip with $b$ and $c$ places a request to gossip with $d$. Since $b$ and $d$ receive no other requests, gossips take place between $a$ and $b$ and between $c$ and $d$. Alternatively, if $t$ is an event time of agents $b$ and $d$, then at this time, $b$ places a request to gossip with $a$ and $d$ places a request to gossip with $c$. Since $a$ and $c$ receive no other requests, gossips again take place between $a$ and $b$ and between $c$ and $d$. Thus, under no conditions is there ever a gossip between $b$ and $c$, so the gossiping process will never reach a consensus. The reader may wish to verify that simply changing the priorities will not rectify this situation: For any choice of priorities, there will always be at least one gossip needed to reach a consensus, which will not take place.

The preceding example illustrates that fixed priorities can present problems. In what follows we take an alternative approach.

In the light of Theorem 1 it is of interest to consider gossiping protocols which generate repetitively complete gossip sequences. Towards this end, let us agree to say that an agent $i$ has completed a round of gossiping after it has gossiped with each neighbor in $N_i$, at least once. Thus, a finite gossiping sequence for the entire group which has occurred over an interval of length $T$ will be complete if over the same period each agent in the group completes a round. In fact in the case when $N$ is a tree, the only way such a sequence could be complete is if over the same period each agent in the group completes a round.

For the protocols which follow it will be necessary for each agent $i$ to keep track of where it is in a particular round. To do this, agent $i$ makes use of a recursively updated neighbor queue $q_i(t)$ where $q_i(\cdot)$ is a function from $T$ to the set of all possible lists of the $m_i$ labels in $N_i$, the neighbor set of agent $i$. Roughly speaking, $q_i(t)$ is a list of the labels of the neighbors of agent $i$ at time $t$ which defines the queue of neighbors at time $t$ which are in line to gossip with agent $i$. The updating of $q_i(t)$ is straightforward: If neighbor $j$ gossips with agent $i$ at time $t$, the updated queue $q_i(t + 1)$ is obtained by moving agent $j$’s label from its current position in $q_i(t)$, to the end of the queue. If on the other hand, agent $i$ does not gossip at time $t$, $q_i(t + 1) = q_i(t)$.

A. Protocols

As noted earlier, it is helpful to have event time assignments which guarantee that no agent can receive a request to gossip at any of its own event times. One easy way to accomplish this is to use event time assignments which satisfy the following assumption.

Distinct Event Times Assumption: For each distinct pair of integers $i$ and $j$ in $\{1, 2, \ldots, n\}$, $T_i$ and $T_j$ are disjoint sets.

Note that the assumption implies that at any fixed event time in $T$, only one agent can receive a request to
gossip, and moreover, that this agent will receive exactly one such request. A simple protocol which ensures exponential convergence under this assumption is as follows.

**Protocol II:** Suppose that the distinct event times assumption holds. At each event time \( t \in T \), the following rules apply for each \( i \in \{1, 2, \ldots, n\} \).

1) If \( t \in T_i \), agent \( i \) places a request to gossip with that neighbor whose label is in the front of the queue \( q_i(t) \).
2) If \( t \notin T_i \), agent \( i \) does not place a request to gossip.
3) Each agent \( i \) receiving a request to gossip must gossip with the neighbor placing the request.
4) If \( t \notin T_i \) and agent \( i \) does not receive a request to gossip, it does not gossip.

The behavior of Protocol II can be easily explained as follows. For \( i \in \{1, 2, \ldots, n\} \), let \( d_i \) be the number of neighbors of agent \( i \), or equivalently, the degree of vertex \( i \) in \( N \). It is clear that with the distinct event times protocol, each agent \( i \) will complete a round in a time interval containing no more than \( d_i \) event time intervals of agent \( i \). The length of the time interval large enough to contain \( d_i \) successive event time intervals of agent \( i \) for all \( i \in \{1, 2, \ldots, n\} \) is the maximum of the times \( T_i d_i \), \( i \in \{1, 2, \ldots, n\} \). Thus, the sequence of gossips which occur on an interval of this length must necessarily be complete. We have proved the following.

**Proposition 1:** Suppose that the distinct event times assumption holds and that all agents in the group adhere to Protocol II. Then, the infinite sequence of gossips generated will be repetitively complete with period

\[
T = \max_i(d_i T_i).
\]

One shortcoming of Protocol II is that it does not allow for multigossiping. Another is that the distinct event times assumption on which the protocol depends is somewhat stringent. It is possible to relax this assumption and still devise a protocol which ensures exponential convergence. The relaxed assumption is as follows.

**Distinct Neighbor Event Times Assumption:** For each \( i \in \{1, 2, \ldots, n\} \) and each \( j \in N_i \), \( T_i \) and \( T_j \) are disjoint sets.

Thus, this assumption holds, the event times of each agent are distinct from the event times of all of its neighbors. The assignment of event times which satisfy this assumption is mathematically identical to the classic “vertex coloring problem” from graph theory [28]. Note that the distinct neighbor event times assumption stipulates that no two adjacent vertices on the neighbor graph \( N \) can have the same event times. The rule defining vertex coloring of \( N \) stipulates that no two adjacent vertices can have the same color. The least number of different colors required to vertex color \( N \) is called the chromatic number of \( N \) [28]. Brooks’ theorem states that this number is bounded above by the maximum degree of \( N \), except for complete graphs and for graphs with cycles of odd length in which cases the bound is one plus the maximum degree of \( N \) [30]. Thus, in all cases the largest number of distinct event time sequences which would need to be assigned to \( N \) to satisfy the distinct neighbor event times assumption is no greater than one plus the maximum vertex degree of \( N \).

Under the distinct neighbor event times assumption, it is possible to ensure exponential convergence with the following protocol which is a refinement of Protocol II.

**Protocol III:** Suppose that the distinct neighbor event times assumption holds. At each event time \( t \in T \) the following rules apply for each \( i \in \{1, 2, \ldots, n\} \).

1) If \( t \in T_i \), agent \( i \) places a request to gossip with that neighbor whose label is at the front of the queue \( q_i(t) \).
2) If \( t \notin T_i \), agent \( i \) does not place a request to gossip.
3) Each agent \( i \) receiving one or more requests to gossip must gossip with that requesting neighbor whose label is closest to the front of the queue \( q_i(t) \).
4) If \( t \notin T_i \) and agent \( i \) does not receive a request to gossip, it does not gossip.

Just as with the distinct event times protocol, it is possible to derive a worst case bound on the time it takes for all agents to complete a round of gossiping. As a first step towards this end, fix \( i \) and suppose that at some given event time \( t_0 \in T_i \), \( j \) is the leading label in the queue \( q_i(t_0) \). According to the preceding protocol, from this event time forward agent \( i \) must repeatedly place requests with agent \( j \) to gossip at successive event times in \( T_i \), until gossiping between the two takes place. Meanwhile, at these same event times, neighbor \( j \) will be receiving requests to gossip from neighbor \( i \) and possibly some other neighbors. In the worst case, when label \( i \) is at the end of \( q_i(t_0) \) at time \( t_0 \), it will take at most \( d_i \) event time intervals of agent \( i \) for label \( i \) to advance to the front of agent \( j \)’s queue. This means that agents \( i \) and \( j \) are guaranteed to gossip at least once within any time interval containing no more than \( d_i \) event time intervals of agent \( i \). If agent \( i \) has only one neighbor, then the round is complete in at most \( T_i d_i \) time units. On the other hand, if agent \( i \) has more than one neighbor, agent \( i \) then begins to place requests to gossip with the agent whose label \( k \) was second from the front in the queue \( q_i(t_0) \). But at this time label \( i \) might be, in the worst case, at the end of the queue for agent \( k \). By the same reasoning as before, it will take at worst an additional \( d_k \) successive event time intervals of agent \( i \) for agents \( i \) and \( k \) to gossip. In other words, agent \( i \) is guaranteed to have gossiped at least once with both agent \( j \) and agent \( k \) within any time interval containing no more than
\(d_i + d_k\) event time intervals of agent \(i\). By repeating this argument for all labels in the queue \(q_i(t_q)\), one reaches the conclusion that agent \(i\) is guaranteed to complete a round of gossiping with all of its neighbors in any time interval containing \(\sum_{j \in N_i} d_j\) event time intervals of agent \(i\). An upper bound on the length of any such interval is thus \(\sum_{j \in N_i} d_j\).

From the preceding it is clear that within any interval of time containing either \(\sum_{j \in N_i} d_j\) event times of agent \(i\) or \(\sum_{j \in N_k} d_j\) event time intervals of agent \(k\), neighbors \(i\) and \(k\) will gossip at least once; thus neighbors \(i\) and \(k\) will gossip at least once in any time interval of length \(\min\{T_i \sum_{j \in N_i} d_j, T_k \sum_{j \in N_k} d_j\}\). The maximum of this amount of time over all agent pairs is thus an upper bound on the amount of time it takes for all neighbor pairs to gossip at least once. But completeness of a gossip sequence is assured if the sequence contains a gossip for each possible neighbor pair. We have therefore proved the following proposition.

**Proposition 2:** Let \(E\) denote the set of all edges \((i, j)\) in \(\mathbb{N}\). Suppose that the distinct neighbor event times assumption holds and that all agents in the group adhere to Protocol III. Then, the infinite sequence of gossiping generated will be repetitively complete with period

\[
T = \max_{(i, j) \in E} \min \left\{ T_i \sum_{j \in N_i} d_j, T_k \sum_{j \in N_k} d_j \right\}.
\]

A disadvantage of Protocol III is that it requires the distinct neighbor event times assumption. This assumption can only be satisfied by offline assignment of event times for each agent, and in some applications such an offline assignment may be undesirable. In a recent doctoral thesis [13], a clever gossiping protocol is proposed which does not require the distinct neighbor event times assumption. The protocol avoids deadlocks and achieves consensus exponentially fast. A disadvantage of this protocol is that it requires each agent to obtain the values of all of its neighbors’ gossip variables at each clock time. Thus, if communication cost is an important issue, this protocol may not be satisfactory even though only local information is required. By exploiting one of the key ideas in [13] together with the notion of an agent’s neighbor queue \(q_i(t)\) defined earlier, it is possible to obtain a gossiping protocol which also avoids deadlocks and achieves consensus exponentially fast but without requiring each agent to obtain the value of more than one of its neighbors’ gossip variables at each clock time. Our aim below is to outline this protocol.

**Protocol IV:** In the following, agent \(i\)'s preferred neighbor at time \(t\) is that agent whose label \(i^*(t)\) is in the front of the queue \(q_i(t)\). Between clock times \(t\) and \(t + 1\) each agent \(i\) performs the steps enumerated below in the order indicated. Although the agents’ actions need not be precisely synchronized, it is understood that for each \(k \in \{1, 2, 3\}\) all agents complete step \(k\) before any embark on step \(k + 1\).

1) First Transmission: Agent \(i\) places a request to gossip with its preferred neighbor by sending both its label \(i\) and its gossip value \(x_i(t)\) to agent \(i^*(t)\). At the same time agent \(i\) receives requests to gossip (i.e., the labels and corresponding gossip values) from all of those neighbors which have agent \(i\) as their current preferred neighbor. Let \(R_i(t)\) denote the set of labels of these requesting neighbors.

2) Second Transmission: Agent \(i\) sends its current gossip value \(x_i(t)\) to those neighbors which have agent \(i\) as their current preferred neighbor, namely the neighbors of agent \(i\) with labels in \(R_i(t)\).

3) Acceptances:

   a) If agent \(i\) has not placed a request to gossip but has received at least one request to gossip, then agent \(i\) sends an acceptance to that particular requesting neighbor whose label is closest to the front of the queue \(q_i(t)\).

   b) If agent \(i\) has either placed a request to gossip or has not received any request to gossip, then agent \(i\) does not send out an acceptance.

4) Gossip variable and queue updates:

   a) If agent \(i\) either sends an acceptance to or receives an acceptance from neighbor \(j\), then agent \(i\) gossips with neighbor \(j\) by setting

   \[
x_i(t + 1) = \frac{x_i(t) + x_j(t)}{2}.
\]

Agent \(i\) updates its queue by moving \(j\), and any labels \(k \in \{i^*(t)\} \cup R_i(t)\) for which \(x_k(t) = x_i(t)\) from their current positions in \(q_i(t)\) to the end of the queue while maintaining their relative order.

b) If agent \(i\) has not sent out an acceptance or received one, then agent \(i\) does not update the value of \(x_i(t)\). In addition, \(q_i(t)\) is updated by moving any labels \(k \in \{i^*(t)\} \cup R_i(t)\) for which \(x_k(t) = x_i(t)\) from their current positions in \(q_i(t)\) to the end of the queue while maintaining their relative order.

It is possible to show that every gossip sequence generated by this protocol is repetitively complete with period no greater than the number of edges of \(\mathbb{N}\) [31]. It follows from Theorem 1 that any sequence of gossip vectors generated by this protocol is exponentially convergent.
1) Convergence Rate: Recall that a gossiping sequence is repetitively complete with period \( T \) if each successive subsequence of gossips of length \( T \) in the sequence is complete; and if, in addition, each gossip in the sequence repeats once every \( T \) time units, the sequence is periodic with period \( T \). As was noted in Section III-A, for a repetitively complete sequence of gossiping matrices \( M(1), M(2), \ldots \) which is \( T \)-periodic, the convergence rate of the product \( M(T)M(T-1) \cdots M(1) \) as \( t \to \infty \) is determined by \( T \) and by the eigenvalue of \( N = M(T)M(T-1) \cdots M(1) \) which is second largest in magnitude. For gossiping sequences which are repetitively complete but not periodic this is no longer true. Such sequences are closely related to what are called “nonhomogeneous Markov chains” for which there is a substantial literature [26]. Notwithstanding this, the following question remains. What determines the convergence rate of a repetitively complete gossip sequence which is not necessarily periodic? This is the question which will be considered next. We will tackle the question in two steps. First, in Section IV-B, we will discuss certain relevant basic properties of stochastic matrices. Then, in Section IV-C, we will study several types of “seminorms” appropriate to the analysis of nonhomogeneous Markov chains. Finally, in Section IV-D, we will show that a certain seminorm provides exactly what is needed to characterize the convergence rate of a repetitively complete gossip sequence.

B. Stochastic Matrices

Since gossip matrices are stochastic matrices, a natural starting point for the study of convergence rates of gossiping sequences is a review of some of the basic concepts associated with stochastic matrices. We begin with the idea of a graph of a stochastic matrix.

1) Graph of a Stochastic Matrix: Many properties of a stochastic matrix can be usefully described in terms of an associated directed graph determined by the matrix. The graph of nonnegative matrix \( M \in \mathbb{R}^{n \times n} \), written \( \gamma(M) \), is a directed graph on \( n \) vertices with an arc from vertex \( i \) to vertex \( j \) just in case \( m_{ij} \neq 0 \); if \( (i,j) \) is such an arc, we say that \( i \) is a neighbor of \( j \) and that \( j \) is an observer of \( i \). Thus, \( \gamma(M) \) is that directed graph whose adjacency matrix is the transpose of the matrix obtained by replacing all nonzero entries in \( M \) with ones.

2) Connectivity: There are various notions of connectivity which are useful in the study of the convergence of products of stochastic matrices. Perhaps the most familiar of these is the idea of “strong connectivity.” A directed graph is strongly connected if there is a directed path between each pair of distinct vertices. A directed graph is weakly connected if there is an undirected path between each pair of distinct vertices. There are other notions of connectivity which are also useful in this context. To define several of them, let us agree to call a vertex \( i \) of a directed graph \( G \), a root of \( G \) if for each other vertex \( j \) of \( G \), there is a directed path from \( i \) to \( j \). Thus, \( i \) is a root of \( G \), if it is the root of a directed spanning tree of \( G \). We will say that \( G \) is rooted at \( i \) if \( i \) is in fact a root. Thus, \( G \) is rooted at \( i \) just in case each other vertex of \( G \) is reachable from vertex \( i \) along a directed path within the graph. \( G \) is strongly rooted at \( i \) if each other vertex of \( G \) is reachable from vertex \( i \) along a directed path of length 1. Thus, \( G \) is strongly rooted at \( i \) if \( i \) is a neighbor of every other vertex in the graph. By a rooted graph \( G \) is meant a directed graph which possesses at least one root. A strongly rooted graph is a graph which has at least one vertex at which it is strongly rooted. Note that a nonnegative matrix \( M \in \mathbb{R}^{n \times n} \) has a strongly rooted graph if and only if it has a positive column. Note that every strongly connected graph is rooted and every rooted graph is weakly connected. The converse statements are false. In particular there are weakly connected graphs which are not rooted and rooted graphs which are not strongly connected.

3) Composition: Since we will be interested in products of stochastic matrices, we will be interested in graphs of such products and how they are related to the graphs of the matrices comprising the products. For this we need the idea of “composition” of graphs. Let \( G_p \) and \( G_q \) be two directed graphs with vertex set \( V \). By the composition of \( G_p \) with \( G_q \), written \( G_q \circ G_p \), is meant the directed graph with vertex set \( V \) and arc set defined in such a way so that \((i,j)\) is an arc of the composition just in case there is a vertex \( k \) such that \((i,k)\) is an arc of \( G_p \) and \((k,j)\) is an arc of \( G_q \). Thus, \((i,j)\) is an arc in \( G_q \circ G_p \) if and only if \( i \) has an observer in \( G_p \) which is also a neighbor of \( j \) in \( G_q \). Note that composition is an associative binary operation; because of this, the definition extends unambiguously to any finite sequence of directed graphs \( G_1, G_2, \ldots, G_k \) with the same vertex set.

Composition and matrix multiplication are closely related. In particular, the graph of the product of two nonnegative matrices \( M_1, M_2 \in \mathbb{R}^{n \times n} \) is equal to the composition of the graphs of the two matrices comprising the product. In other words, \( \gamma(M_2M_1) = \gamma(M_2) \circ \gamma(M_1) \).

If we focus exclusively on graphs with self-arcs at all vertices, more can be said. In this case, the definition of composition implies that the arcs of both \( G_p \) and \( G_q \) are arcs of \( G_q \circ G_p \); the converse is false. The definition of composition also implies that if \( G_p \) has a directed path from \( i \) to \( k \) and \( G_q \) has a directed path from \( k \) to \( j \), then \( G_q \circ G_p \) has a directed path from \( i \) to \( j \). These implications are consequences of the requirement that the vertices of the graphs in question all have self-arcs. It is worth emphasizing that the union of the arc sets of a sequence of graphs \( G_1, G_2, \ldots, G_k \) with self-arcs must be contained in the arc set of their composition. However, the converse is not true in general and it is for this reason that composition rather than union proves to be the more useful concept for our purposes.
4) **Convergability:** It is of obvious interest to have a clear understanding of what kinds of stochastic matrices within an infinite product guarantee that the infinite product converges. There are many ways to address this issue and many existing results. Here we focus on just one issue.

Let $S$ denote the set of all stochastic matrices in $\mathbb{R}^{n\times n}$ with positive diagonal entries. Call a compact subset $M \subset S$ **convergable** if for each infinite sequence of matrices $M_1, M_2, M_3, \ldots$ from $M$, the sequence of products $M_1 M_2 M_3 \ldots$ converges exponentially fast to a matrix of the form $1c$. Convergability can be characterized as follows.

**Theorem 4:** Let $R$ denote the set of all matrices in $S$ with rooted graphs. Then, a compact subset $M \subset S$ is convergable if and only if $M \subset R$.

The theorem implies that $R$ is the largest subset of $n \times n$ stochastic matrices with positive diagonal entries whose compact subsets are all convergable. $R$ itself is not convergable because it is not closed and thus not compact.

**Proof of Theorem 4:** The fact that any compact subset of $R$ is convergable is more or less well known from the work reported in [32]; the statement also follows from Proposition 11 of [33]. To prove the converse, suppose that $M \subset S$ is convergable. Then, by continuity, every sufficiently long product of matrices from $M$ must be a matrix with a positive column. Therefore, the graph of every sufficiently long product of matrices from $M$ must be strongly rooted. It follows from Proposition 5 of [33] that $M$ must be a subset of $R$.

Although doubly stochastic matrices are stochastic, convergability for classes of doubly stochastic matrices has a different characterization than it does for classes of stochastic matrices. Let $D$ denote the set of all doubly stochastic matrices in $S$. In the following, we will prove Theorem 5.

**Theorem 5:** Let $W$ denote the set of all matrices in $D$ with weakly connected graphs. Then, a compact subset $M \subset D$ is convergable if and only if $M \subset W$.

The theorem implies that $W$ is the largest subset of $n \times n$ doubly stochastic matrices with positive diagonal entries whose compact subsets are all convergable. Like $R$, $W$ is not convergable because it is not compact.

An interesting set of stochastic matrices in $S$ whose compact subsets are known to be convergable is the set of all “scrambling matrices.” A matrix $S \subset S$ is **scrambling** if for each distinct pair of integers $i$ and $j$, there is a column $k$ of $S$ for which $s_{ik}$ and $s_{jk}$ are both nonzero [26]. In graph theoretic terms $S$ is a scrambling matrix just in case its graph is “neighbor shared” where by neighbor shared we mean that each distinct pair of vertices in the graph share a common neighbor [33]. Convergability of compact subsets of scrambling matrices is tied up with the concept of the coefficient of ergodicity [26] which for a given stochastic matrix $S \in S$ is defined by

$$\tau(S) = \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} |s_{ik} - s_{jk}|.$$  

(1)

It is known that $0 \leq \tau(S) \leq 1$ for all $S \in S$ and that

$$\tau(S) < 1$$

(2)

if and only if $S$ is a scrambling matrix. It is also known that

$$\tau(S_2 S_1) \leq \tau(S_2) \tau(S_1), \quad S_1, S_2 \in S.$$  

(3)

It can be shown that (2) and (3) are sufficient conditions to ensure that any compact subset of scrambling matrices is convergable. But $\tau(\cdot)$ has another role. It provides a worst case convergence rate for any infinite product of scrambling matrices from a given compact set $C \subset S$. In particular, it can be easily shown that as $i \to \infty$, any product $S_i S_{i-1} \cdots S_2 S_1$ of scrambling matrices $S_i \in C$ converges to a matrix of the form $1c$ as fast as $\lambda$ converges to zero where

$$\lambda = \max_{S \in C} \tau(S).$$

This preceding discussion suggests the following question. Can analogs of the coefficient of ergodicity satisfying formulas like (2) and (3) be found for the set of stochastic matrices with rooted graphs or perhaps for the set of doubly stochastic matrices with weakly connected graphs? In the following, we will provide a partial answer to this question for the case of stochastic matrices and a complete answer for the case of doubly stochastic matrices. Our approach will be to appeal to certain types of seminorms of stochastic matrices.

**C. Seminorms**

Let $\| \cdot \|_p$ be the induced $p$-norm on $\mathbb{R}^{m \times n}$. We will be interested in $p = 1, 2, \infty$. Note that for a nonnegative matrix $A$

$$\|A\|_1 = \max \text{ column sum } A$$

$$\|A\|_2 = \sqrt{\mu(A^T A)}$$

$$\|A\|_\infty = \max \text{ row sum } A$$

where $\mu(A^T A)$ is the largest eigenvalue of $A^T A$; that is, the square of the largest singular value of $A$. For $M \in \mathbb{R}^{m \times n}$
define
\[
|M|_p = \min_{c \in \mathbb{R}^{|X|}} \|M - 1c\|_p.
\]

As defined, \(| \cdot |_p\) is nonnegative and \(|M|_p \leq \|M\|_p\); clearly \(|\mu M|_p = |\mu| |M|_p\) for all real numbers \(\mu\) so \(| \cdot |_p\) is “positively homogeneous” [27]. Let \(M_1\) and \(M_2\) be matrices in \(\mathbb{R}^{m \times n}\) and let \(c_0, c_1,\) and \(c_2\) denote values of \(c\) which minimize \(\|M_1 + M_2 - 1c\|_p\), \(\|M_1 - 1c\|_p\), and \(\|M_2 - 1c\|_p\), respectively. Note that
\[
|M_1 + M_2|_p = \|M_1 + M_2 - 1c_0\|_p
\leq \|M_1 + M_2 - 1(c_1 + c_2)\|_p
\leq \|M_1 - 1c_1\|_p + \|M_2 - 1c_2\|_p
= |M_1|_p + |M_2|_p.
\]

Thus, the triangle inequality holds. These properties mean that \(| \cdot |_p\) is a seminorm. \(| \cdot |_p\) behaves much like a norm. For example, if \(N\) is a submatrix of \(M\), then \(\|N\|_p \leq |M|_p\). However, \(| \cdot |_p\) is not a norm because \(|M|_p = 0\) does not imply \(M = 0\); rather it implies that \(M = 1c\) for some row vector \(c\) which minimizes \(\|M - 1c\|_p\). For our purposes, \(| \cdot |_p\) has a particularly important property.

Lemma 1: Suppose \(\mathcal{M}\) is a subset of \(\mathbb{R}^{n \times n}\) such that \(M1 = 1\) for all \(M \in \mathcal{M}\). Then
\[
|M_2M_1|_p \leq |M_2|_p|M_1|_p.
\]

We say that \(| \cdot |_p\) is submultiplicative on \(\mathcal{M}\).

Proof of Lemma 1: Let \(c_0, c_1,\) and \(c_2\) denote values of \(c\) which minimize \(\|M_2M_1 - 1c\|_p\), \(\|M_1 - 1c\|_p\), and \(\|M_2 - 1c\|_p\), respectively. Then
\[
|M_2M_1|_p = \|M_2M_1 - 1c_0\|_p
\leq \|M_2M_1 - 1(c_2M_1 + c_1 - c_21c_1)\|_p
= \|M_2M_1 - 1c_2M_1 - M_21c_1 + 1c_21c_1\|_p
= \|(M_2 - 1c_2)(M_1 - 1c_1)\|_p
\leq \|(M_2 - 1c_2)\|_p\|(M_1 - 1c_1)\|_p
= |M_2|_p|M_1|_p.
\]

Thus, (4) is true.

We say that \(M \in \mathbb{R}^{n \times n}\) is semicontractive in the \(p\)-norm if \(|M|_p < 1\). In view of Lemma 1, the product of semicontractive matrices in \(\mathcal{M}\) is thus semicontractive. The importance of these ideas lies in the following fact.

Proposition 3: Suppose \(\mathcal{M}\) is a subset of \(\mathbb{R}^{n \times n}\) such that \(M1 = 1\) for all \(M \in \mathcal{M}\). Let \(p\) be fixed and let \(\mathcal{M}\) be a compact set of semicontractive matrices in \(\mathcal{M}\). Let
\[
\lambda = \sup_{M \in \mathcal{M}} |M|_p.
\]

Then, for each infinite sequence of matrices \(M_i \in \mathcal{M}\), \(i \in \{1, 2, \ldots\}\), the matrix product
\[
M_1M_2 \cdots M_i
\]
converges as \(i \to \infty\) as fast as \(\lambda^i\) converges to zero, to a rank one matrix of the form \(1c\).

Proof of Proposition 3: Clearly \(|M|_p \leq \lambda\), \(M \in \mathcal{M}\). Moreover, \(\lambda < 1\) because each \(M \in \mathcal{M}\) is semicontractive and because \(\mathcal{M}\) is compact. Write \(M_i = 1c_i + T_i\), \(i \geq 1\), where \(c_i\) is a value of \(c\) which minimizes \(\|M_i - 1c\|_p\). For \(i \geq 1\) set \(X_i = M_iM_{i-1} \cdots M_1\) and \(Y_i = T_iT_{i-1} \cdots T_1\). Clearly \(|M_i|_p = \|T_i\|_p\), \(i \geq 1\), so
\[
\|T_i\|_p \leq \lambda^i, \quad i \geq 1.
\]

A simple computation yields
\[
X_k = Y_k + \sum_{i=1}^{k} 1c_iY_{i-1}, \quad k \geq 1
\]
where \(Y_0 = I\). Note also that because of (5), \(Y_k\) tends to zero as \(k \to \infty\). We claim that the sequence \(\sum_{i=1}^{k} 1c_iY_{i-1}, k \geq 1\), has a limit. To prove that this is so it is enough to show that \(\sum_{j=1}^{k} 1c_iY_{i-1}, k \geq 1\), is a Cauchy sequence. Towards this end observe that
\[
\sum_{i=1}^{j} 1c_iY_{i-1} - \sum_{i=1}^{j} 1c_iY_{i-1} = \sum_{i=k+1}^{j+k} 1c_iY_{i-1}, \quad j \geq 1.
\]

Moreover
\[
\|1c_iY_{i-1}\|_p \leq d\lambda^{-1}, \quad i \geq 1
\]
where $d = \sup_{i \geq 1} \|1_{i}c\|_p$. Therefore, for $j \geq 1$

$$\left\| \sum_{i=k+1}^{j+k} 1_{c}Y_{i-1} \right\|_p \leq d \sum_{i=k+1}^{j+k} \lambda^{i-1} = d \lambda^{k} \sum_{i=1}^{j} \lambda^{i-1}$$

$$\leq d \lambda^{k} \sum_{i=1}^{\infty} \lambda^{i-1} = d \frac{\lambda^{k}}{1 - \lambda}.$$ 

Therefore

$$\left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - \sum_{i=1}^{k} 1_{c}Y_{i-1} \right\|_p \leq d \frac{\lambda^{k}}{1 - \lambda}, \quad j, k \geq 1 \quad (7)$$

which shows that $\sum_{i=1}^{k} 1_{c}Y_{i-1}$, $k \geq 1$, is a Cauchy sequence. Therefore, the sequence $\sum_{i=1}^{k} 1_{c}Y_{i-1}$, $k \geq 1$, has a limit which we denote by $1_{c}$.

Note next that for $j, k \geq 1$

$$\left\| \sum_{i=1}^{k} 1_{c}Y_{i-1} - 1_{c} \right\|_p = \left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - 1_{c} \right\|_p - \left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - \sum_{i=1}^{k} 1_{c}Y_{i-1} \right\|_p \quad (8)$$

so

$$\left\| \sum_{i=1}^{k} 1_{c}Y_{i-1} - 1_{c} \right\|_p \leq \left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - 1_{c} \right\|_p - \left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - \sum_{i=1}^{k} 1_{c}Y_{i-1} \right\|_p.$$ 

In view of (7)

$$\left\| \sum_{i=1}^{k} 1_{c}Y_{i-1} - 1_{c} \right\|_p \leq \left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - 1_{c} \right\|_p + d \frac{\lambda^{k}}{1 - \lambda}.$$ 

But $\left\| \sum_{i=1}^{j+k} 1_{c}Y_{i-1} - 1_{c} \right\|_p$ tends to zero as $j \to \infty$ so

$$\left\| \sum_{i=1}^{k} 1_{c}Y_{i-1} - 1_{c} \right\|_p \leq d \frac{\lambda^{k}}{1 - \lambda}, \quad k \geq 1. \quad (8)$$

To proceed observe that (6) implies that

$$\|X_{k} - 1_{c}\|_p \leq \|Y_k\|_p + \left\| \sum_{i=1}^{k} 1_{c}Y_{i-1} - 1_{c} \right\|_p.$$ 

From this, (5), and (8), it follows that

$$\|X_{k} - 1_{c}\|_p \leq \lambda^{k} \left( 1 + \frac{d}{1 - \lambda} \right), \quad k \geq 1.$$ 

This completes the proof. 

1) Case $p = 1$: We now consider in more detail the case when $p = 1$. For this case, it is possible to derive an explicit formula for the seminorm $|M|_1$ of a nonnegative matrix $M \in \mathbb{R}^{n \times n}$.

**Proposition 4:** Let $q$ be the unique integer quotient of $n$ divided by 2. Let $M \in \mathbb{R}^{n \times n}$ be a nonnegative matrix. Then

$$|M|_1 = \max_{i \in \mathcal{L}} \left\{ \sum_{i \in \mathcal{L}_i} m_{ij} - \sum_{i \in \mathcal{S}_i} m_{ij} \right\}$$

where $\mathcal{L}_i$ and $\mathcal{S}_i$ are, respectively, the row indices of the $q$ largest and $q$ smallest entries in the $i$th column of $M$.

This result is a direct consequence of the following lemma and the definition of $| \cdot |_1$.

**Lemma 2:** Let $q$ denote the unique integer quotient of $n$ divided by 2. Let $y$ be a nonnegative $n$-vector and write $\mathcal{L}$ and $\mathcal{S}$ for the row indices of the $q$ largest and $q$ smallest entries in $y$, respectively. Then

$$|y|_1 = \sum_{i \in \mathcal{L}} y_i - \sum_{i \in \mathcal{S}} y_i$$

where $y_i$ is the $i$th entry in $y$.

**Proof of Lemma 2:** Let $a$ denote the $n$-vector whose entries $a_1, a_2, \ldots, a_n$ are the entries of $y$ relabeled so that $a_1 \leq a_2 \leq \cdots \leq a_n$. Then

$$\sum_{i \in \mathcal{L}} y_i = \sum_{i > (q+r)} a_i \quad \text{and} \quad \sum_{i \in \mathcal{S}} y_i = \sum_{i \leq q} a_i$$

where $r$ is the unique integer remainder of $n$ divided by 2.
Moreover
\[ \|y - \mathbf{1}x\|_1 = \|a - \mathbf{1}x\|_1, \quad x \in \mathbb{R}. \]

Therefore, to prove the lemma, it is enough to show that
\[ |a|_1 = \min_{x \in \mathbb{R}} \|a - \mathbf{1}x\|_1 = \sum_{i>(q+r)} a_i - \sum_{i \leq q} a_i. \tag{9} \]

Suppose \( n \) is even in which case \( n = 2q \) and \( r = 0 \). If \( q = 1 \), then \( \min_{x \in \mathbb{R}} (|a_1 - x| + |a_2 - x|) = a_2 - a_1 \) in which case (9) holds. Suppose \( q > 1 \). For fixed \( k \in \{1, 2, \ldots, q - 1\} \) and any value of \( x \) located in the interval \([a_k, a_{k+1}]\), it must be true that \( |a_i - x| = x - a_i \) for \( i \leq k \) and \( |a_i - x| = a_i - x \) for \( i \geq k + 1 \). Since \( k < q \), the number of values of \( i \) such that \( i \geq k + 1 \) is greater than the number of values of \( i \) such that \( i \leq k \). Thus, for \( x \in [a_k, a_{k+1}] \), the sum \( \sum_{j=1}^{n} |a_i - x| \) is a linear polynomial in \( x \) and the coefficient of \( x \) is negative. Since this is true for all \( k \in \{1, 2, \ldots, q - 1\} \), the sum \( \sum_{j=1}^{n} |a_i - x| \) is a decreasing function of \( x \) on the union of the intervals \([a_k, a_{k+1}]\), \( k = 1, 2, \ldots, q - 1 \); i.e., on \([a_k, a_{k+1}]\). By similar reasoning \( \sum_{j=1}^{n} |a_i - x| \) is an increasing function of \( x \) on the interval \([a_k, a_{k+1}]\). Meanwhile, for values of \( x \in [a_k, a_{k+1}] \), clearly \( |a_i - x| = x - a_i \) and \( |a_{q+1} - x| = a_{q+1} - x \), so the sum \( \sum_{j=1}^{n} |a_i - x| \) is a constant. But \( \sum_{j=1}^{n} |a_i - x| \) is a continuous function of \( x \). Therefore, \( \sum_{j=1}^{n} |a_i - x| \) is nonincreasing for \( x \leq a_k \) and nondecreasing for \( x \geq a_{q+1} \). Therefore, a value of \( x \) which minimizes \( \sum_{j=1}^{n} |a_i - x| \) is \( x = a_k \). Equation (9) follows at once.

Now suppose \( n \) is odd in which case \( r = 1 \). For fixed \( k \in \{1, 2, \ldots, q\} \) and any value of \( x \) located in the interval \([a_k, a_{k+1}]\), it must be true that \( |a_i - x| = x - a_i \) for \( i \leq k \) and \( |a_i - x| = a_i - x \) for \( i \geq k + 1 \). Since \( k < q \), the number of values of \( i \) such that \( i \geq k + 1 \) is greater than the number of values of \( i \) such that \( i \leq k \). Thus, for \( x \in [a_k, a_{k+1}] \), the sum \( \sum_{j=1}^{n} |a_i - x| \) is a linear polynomial in \( x \) and the coefficient of \( x \) is negative. Since this is true for all \( k \in \{1, 2, \ldots, q\} \), the sum \( \sum_{j=1}^{n} |a_i - x| \) is an increasing function of \( x \) on the union of the intervals \([a_k, a_{k+1}]\), \( k = 1, 2, \ldots, q \); i.e., on \([a_k, a_{k+1}]\). By similar reasoning \( \sum_{j=1}^{n} |a_i - x| \) is an increasing function of \( x \) on the interval \([a_k, a_{k+1}]\). Therefore, the unique value of \( x \) which minimizes \( \sum_{j=1}^{n} |a_i - x| \) is \( x = a_{q+1} \). Equation (9) follows at once. \( \blacksquare \)

Consider now the case when \( M \) is a doubly stochastic matrix \( S \). Then, the column sums of \( S \) are all equal to \( 1 \). This implies that \( |S|_1 \leq 1 \) because \( |S|_1 \leq \|S\|_1 = 1 \). The column sums all equaling one also imply that
\[ \sum_{i \in C_j} s_{ij} + rm_j + \sum_{i \in S_j} s_{ij} = 1, \quad j \in \{1, 2, \ldots, n\} \]

where \( m_i \) is the median\(^3\) of the \( n \) entries in the \( j \)th column of \( S \). Therefore
\[ |S|_1 = \max_{j \in \{1, 2, \ldots, n\}} \left\{ 2 \sum_{i \in C_j} s_{ij} + rm_j - 1 \right\}. \]

This means that \( S \) is semicontractive in the one-norm just in case
\[ \sum_{i \in C_j} s_{ij} + \frac{r}{2} m_j < 1, \quad j \in \{1, 2, \ldots, n\}. \]

We are led to the following result.

**Theorem 6:** Let \( q \) be the unique integer quotient of \( n \) divided by 2. Let \( S \in \mathbb{R}^{n \times n} \) be a doubly stochastic matrix. Then, \( |S| \leq 1 \). Moreover, \( S \) is a semicontraction in the one-norm if and only if the number of nonzero entries in each column of \( S \) exceeds \( q \).

Note that the doubly stochastic matrix
\[
\begin{bmatrix}
0.5 & 0.125 & 0.125 & 0.125 & 0.125 \\
0.5 & 0.125 & 0.125 & 0.125 & 0.125 \\
0 & 0.25 & 0.125 & 0.125 & 0.125 \\
0 & 0.25 & 0.125 & 0.125 & 0.125 \\
0 & 0.25 & 0.125 & 0.125 & 0.125 \\
\end{bmatrix}
\]
has a weakly connected graph but is not a semicontraction for \( p = 1 \). Thus, this particular seminorm is not as useful as we would like for gossiping problems.

It is possible to compare this seminorm with the coefficient of ergodicity. Observe that while the preceding matrix is not a semicontraction it is a scrambling matrix. Thus, for this example, \( \tau(S) = |S|_1 = 1 \). On the other hand, there are also doubly stochastic matrices which are semicontractions but which are not scrambling matrices. An example of this is the matrix
\[
\begin{bmatrix}
0.5 & 0 & 0 & 0 & 0.5 & 0 \\
0 & 0.5 & 0 & 0 & 0 & 0.5 \\
0.125 & 0.125 & 0.25 & 0.25 & 0.125 & 0.125 \\
0.125 & 0.125 & 0.25 & 0.25 & 0.125 & 0.125 \\
0.125 & 0.125 & 0.25 & 0.25 & 0.125 & 0.125 \\
0.125 & 0.125 & 0.25 & 0.25 & 0.125 & 0.125 \\
\end{bmatrix}
\]

Thus, for this example, \( |S|_1 < \tau(S) = 1 \), which means that there are situations when it may be more advantageous to

\(^3\)The median of a finite set of real numbers is the “middle value” of the set. More precisely, suppose that \( A \) is a set of \( n \) real numbers which are \( a_1 \leq a_2 \leq \cdots \leq a_n \) and let \( q \) and \( r \) be, respectively, the unique integer quotient and remainder of \( n \) divided by 2. If \( n \) is odd, the median of \( A \) is \( a_{q+1} \).

If \( n \) is even, the median of \( A \) is defined to be the average of \( a_q \) and \( a_{q+1} \).
use the seminorm $| \cdot |_1$ to compute convergence rates than to appeal to the coefficient of ergodicity.

2) Case $p = \infty$: Note that in this case $|S|_\infty \leq 1$ for any stochastic matrix because $|S|_\infty \leq \|S\|_\infty = 1$. Although not at all obvious, it turns out that $|S|_\infty$ equals the well-known coefficient of ergodicity discussed earlier and defined by (1). This is an immediate consequence of Proposition 5 which is stated below. Unfortunately, the last example in the preceding section shows that there are doubly stochastic matrices with weakly connected graphs which are not scrambling matrices. Thus, this particular seminorm is also not useful for our purposes.

**Proposition 5:** Let $A \in \mathbb{R}^{n \times n}$ be a nonnegative matrix. Then

$$|A|_\infty = \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} |a_{ik} - a_{jk}|.$$ 

The proof of Proposition 5 depends on the following lemma.

**Lemma 3:** Suppose $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$ is a set of $m > 1$ row vectors in $\mathbb{R}^{1 \times n}$. Let $d(x, y)$ denote the metric

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|, \quad x, y \in \mathbb{R}^{1 \times n}$$

where $x_i$ and $y_i$ are the $i$th entries of $x$ and $y$, respectively. Then

$$\min_{c \in \mathbb{R}^{1 \times n}} \max_i d(c, a_i) = \frac{1}{2} \max_{j,k} d(a_j, a_k).$$

**Proof of Lemma 3:** For any $j$ and $k$, and any row vector $c \in \mathbb{R}^{1 \times n}$, $d(a_j, c) + d(c, a_k) \geq d(a_j, a_k)$ because of the triangle inequality. Clearly $\max_i d(c, a_i) \geq (1/2)(d(a_j, c) + d(c, a_k))$, so

$$\max_i d(c, a_i) \geq \frac{1}{2} d(a_j, a_k)$$

for all $j, k$. In particular, $\max_i d(c, a_i) \geq (1/2) \max_{j,k} d(a_j, a_k)$. Since this is true for all $c$

$$\min_{c} \max_i d(c, a_i) \geq \frac{1}{2} \max_{j,k} d(a_j, a_k).$$

To complete the proof, it must be shown that

$$\min_{c} \max_i d(c, a_i) \leq \frac{1}{2} \max_{j,k} d(a_j, a_k).$$

For this it is enough to show that there exists a closed ball $\mathcal{B}$ in $\mathbb{R}^{1 \times n}$ with radius $r = (1/2) \max_{j,k} d(a_j, a_k)$ such that $\mathcal{B} \supset \mathcal{A}$. Towards this end let $\mathcal{B}$ denote the ball

$$\mathcal{B} = \left\{ x : x \in \mathbb{R}^{1 \times n}, \sum_{i=1}^{n} |x_i| \leq r \right\}.$$ 

The boundary of $\mathcal{B}$, $\partial \mathcal{B}$, consists of $2^{n-1}$ pairs of parallel $(n-1)$-hyperplanes. Consider one of these pairs, $\mathcal{H}_1 : x_1 + x_2 + \cdots + x_n = r$ and $\mathcal{H}_2 : x_1 + x_2 + \cdots + x_n = -r$. For any $y \in \mathcal{H}_1$ and $z \in \mathcal{H}_2$

$$\sum_{i=1}^{n} |y_i - z_i| \geq \sum_{i=1}^{n} (y_i - z_i) = 2r$$

which shows that the distance of the two hyperplanes equals $2r$. Thus, there must exist a real number $s$ such that every $a_i \in \mathcal{A}$ lies between or on the two parallel hyperplanes

$$x_1 + x_2 + \cdots + x_n = s + r$$

$$x_1 + x_2 + \cdots + x_n = s - r.$$ 

Using the same arguments, the other $2^{n-1} - 1$ pairs of parallel hyperplanes also can be shown to exist. Therefore, $\mathcal{A}$ can be contained in a closed set bounded by the $2^{n-1}$ pairs of parallel $(n-1)$-hyperplanes which is a closed ball with radius $(1/2) \max_{j,k} d(a_j, a_k)$ in $\mathbb{R}^{1 \times n}$. 

3) Case $p = 2$: For the case when $p = 2$ it is also possible to derive an explicit formula for the seminorm $|M|_2$ of a nonnegative matrix $M \in \mathbb{R}^{n \times n}$. Towards this end note that for any $x \in \mathbb{R}^n$, the function

$$g(x, c) = x' (M - 1c)' (M - 1c) x$$

$$= x' M x - 2x' M' 1 c x + n(c x)^2$$

We are indebted to Chun-Yi Sun (Department of Mathematics, Yale University) for pointing this out to us.
attains its minimum with respect to \( c \) at

\[
c = -\frac{1}{n} M.
\]

This implies that

\[
|M|_2 = \|PM\|_2 = \sqrt{\mu(M'P'PM)}
\]

where \( P = I - (1/n)11' \) and, for any symmetric matrix \( T \), \( \mu(T) \) is the largest eigenvalue of \( T \). We are led to the following result.

**Proposition 6:** Let \( M \in \mathbb{R}^{n \times n} \) be a nonnegative matrix. Then, \( |M|_2 \) is the largest singular value of the matrix \( PM \) where \( P \) is the orthogonal projection on the orthogonal complement of the span of \( 1 \).

Now suppose that \( M \) is a doubly stochastic matrix \( S \). Then, \( SS' \) is also doubly stochastic and \( 1'S = 1' \). The latter and Proposition 6 imply that

\[
|S|_2 = \sqrt{\mu\left( SS' - \frac{1}{n} 11' \right)}.
\]

More can be said.

**Lemma 4:** If \( S \) is doubly stochastic, then \( \mu( SS' - (1/n)11') \) is the second largest eigenvalue of \( SS' \).  

**Proof of Lemma 4:** Since \( SS' \) is symmetric it has orthogonal eigenvectors one of which is \( 1 \). Let \( 1, x_2, \ldots, x_n \) be such a set of eigenvectors with eigenvalues \( 1, \lambda_2, \ldots, \lambda_n \). Then, \( SS1 = 1 \) and \( SSx_i = \lambda_i x_i, i \in \{ 2, 3, \ldots, n \} \). Clearly \( (SS' - (1/n)11')1 = 0 \) and \( (SS' - (1/n)11')x_i = \lambda_i x_i, i \in \{ 2, 3, \ldots, n \} \). Since 1 is the largest eigenvalue of \( SS' \) it must therefore be true that the second largest eigenvalue \( SS' \) is the largest eigenvalue of \( SS' - (1/n)11' \).

We summarize as follows.

**Theorem 7:** For \( p = 2 \), the seminorm of a doubly stochastic matrix \( S \) is the second largest singular value of \( S \).

There is another way to think about what this theorem implies. Promoted by the work in [8] and [23], suppose one wants to measure in the sense of a 2-norm \( \| \cdot \| \), how much closer an \( n \)-vector \( x \) gets to the average vector \( z = (1/n)11'x \) when it is multiplied by a doubly stochastic matrix \( S \). In other words how does the norm \( \|Sx - z\| \) compare with \( \|x - z\| \)? To address this question, note first that \( x - z \in \mathcal{O} \) where \( \mathcal{O} \) is the orthogonal complement of the span of \( 1 \). Note next that

\[
\|Sx - z\|^2 = \|S(x - z)\|^2 \leq \sup_{y \in \mathcal{O}} y'S'y/\|y\|^2.
\]

But \( \sup_{y \in \mathcal{O}} y'S'y/\|y\|^2 \) is the second largest eigenvalue of \( SS' \) which in turn is the square of the second largest singular value of \( S \). In other words, \( \|Sx - z\| \leq |S|_2\|x - z\| \). Thus, \( Sx \) is always as close to the average vector \( z \) as \( x \) is and is even closer if \( |S|_2 \) is a contraction.

In the light of Theorem 7, we are now in a position to characterize in graph-theoretic terms those doubly stochastic matrices with positive diagonal entries which are semicontractions for \( p = 2 \).

**Theorem 8:** Let \( S \) be a doubly stochastic matrix with positive diagonal entries. Then, \( |S|_2 \leq 1 \). Moreover, \( S \) is a semicontraction in the 2-norm if and only if the graph of \( S \) is weakly connected.

To prove this theorem we need several concepts and results. Let \( G \) denote a directed graph and write \( G' \) for that graph which results when the arcs in \( G \) are reversed; i.e., the dual graph. Call a graph symmetric if it is equal to its dual. Note that in the case of a symmetric graph, the three properties of being rooted, strongly connected, and weakly connected are equivalent. Note also that if \( G \) is the graph of a nonnegative matrix \( M \) with positive diagonal entries, then \( G' \) is the graph of \( M' \) and \( G' \circ G \) is the graph of \( M'M \).

**Lemma 5:** A directed graph \( G \) with self-arcs at all vertices is weakly connected if and only if \( G' \circ G \) is strongly connected.

**Proof of Lemma 5:** Since \( G \) has self-arcs at all vertices it so does \( G' \). This implies that the arc set of \( G' \circ G \) contains the arc sets of \( G \) and \( G' \). Thus, for any undirected path in \( G \), if vertices \( i \) and \( j \) have a corresponding directed path in \( G' \circ G \) between the same two vertices. Thus, if \( G \) is weakly connected, \( G' \circ G \) must be strongly connected.

Now suppose that \((i, j)\) is an arc in \( G' \circ G \). Then, because of the definition of composition, there must be a vertex \( k \) such that \((i, k) \) is an arc in \( G \) and \((k, j) \) is an arc in \( G' \). This implies that \((i, k) \) and \((j, k) \) are arcs in \( G \). Thus, \( G \) has an undirected path from \( i \) to \( j \). Now suppose that \((i, v_1), (v_1, v_2), \ldots, (v_r, j) \) is a directed path in \( G' \circ G \) between \( i \) and \( j \). Between each pair of successive vertices along this path there must therefore be an undirected path in \( G \). Thus, there must be an undirected path in \( G \) between \( i \) and \( j \). It follows that if \( G' \circ G \) is strongly connected, then \( G \) is weakly connected.
Lemma 6: Let $T$ be a stochastic matrix with positive diagonal entries. If $T$ has a strongly connected graph, then the magnitude of its second largest eigenvalue is less than 1. If, on the other hand, the magnitude of the second largest eigenvalue of $T$ is less than 1, then the graph of $T$ is weakly connected.

Proof of Lemma 6: Suppose that the graph of $T$ is strongly connected. Then, via Theorem 6.2.24 of [27], $T$ is irreducible. Thus, there is an integer $k$ such that $(I + T)^k > 0$. Since $T$ has positive diagonal entries, this implies that $T^k > 0$. Therefore, $T$ is primitive [27]. Thus, by the Perron–Frobenius theorem [26], $T$ has only one eigenvalue of maximum modulus. Since the spectral radius of $T$ is 1 and 1 is an eigenvalue, the magnitude of the second largest eigenvalue of $T$ must be less than 1.

To prove the converse, suppose that $T$ is a stochastic matrix whose second largest eigenvalue in magnitude is less than 1. Then

$$\lim_{i \to \infty} T^i = 1c$$

for some row vector $c$. Suppose that the graph of $T$ is not weakly connected. Therefore, if $q$ denotes the number of weakly connected components of the graph, then $q > 1$. This implies that $T = PDP$ for some permutation matrix $P$ and block diagonal matrix $D$ with $q$ blocks. Since $D = PT P'$, $D$ is also stochastic. Thus, each of its $q$ diagonal blocks is stochastic. Since $T^i$ converges to $1c$, $D^i$ must converge to a matrix of the form $1c$. But this is clearly impossible because $1c$ cannot have $q > 1$ diagonal blocks.

Proof of Theorem 8: Let $S$ be a doubly stochastic matrix with positive diagonal entries. Then, $S$ is irreducible. Thus, there is an integer $k$ such that $(I + S)^k > 0$. Since $S$ has positive diagonal entries, this implies that $S^k > 0$. Therefore, $S$ is primitive [27]. Thus, by the Perron–Frobenius theorem [26], $S$ has only one eigenvalue of maximum modulus. Since the spectral radius of $T$ is 1 and 1 is an eigenvalue, the magnitude of the second largest eigenvalue of $S$ must be less than 1.

In particular, $S$ is a doubly stochastic matrix with a weakly connected graph but it is not a scrambling matrix.

D. Contraction Coefficient

By the contraction coefficient of a gossip matrix $M$ is meant the seminorm $|M|_2$. The main result we want to prove is as follows.

Theorem 9: A gossip matrix $M$ is complete if and only if its contraction coefficient is less than one.

Before turning to a proof of this theorem, let us consider its consequences. As in the hypothesis of Theorem 1, let $M(1), M(2), M(3), \ldots$ denote the gossiping matrices corresponding to an infinite sequence of single gossips which is repetitively complete with period $T$. Our aim is to explain how fast the matrix product $M(t)M(t - 1) \cdots M(1)$ converges to $(1/n)11^T$, and in so doing provide a proof of Theorem 1. Towards this end note first that each $M(t) \in \mathcal{P}$, the set of all $n \times n$ single-gossip matrices. Let $C$ denote the set of all complete gossip matrices which are products of exactly $T$ single-gossip matrices from $\mathcal{P}$. Note that $C$ is a compact set because $\mathcal{P}$ is. For each $S \in C$, let $|S|_2$ denote the contraction coefficient of $S$. Then, in view of Theorem 9, $|S|_2 < 1, S \in C$. Therefore, the nonnegative number

$$\mu = \max_C |S|_2$$

is less than one. Next observe that since $M(1), M(2), M(3), \ldots$ corresponds to a repetitively complete gossip sequence, each matrix $N_i = M(i)M((i - 1)T + 1), i \geq 1$, must be in $C$; thus, $|N_i|_2 \leq \mu, i \geq 1$. It follows from Proposition 3 that as $i \to \infty$, the matrix product $N_iN_{i-1} \cdots N_1$ converges to $(1/n)11^T$ as fast as $\mu^i$ converges to zero. Thus, if we define $\lambda = \mu^{1/T}$, then as $i \to \infty$, the
matrix product $M(t)M(t-1)\cdots M(1)$ also converges to $(1/n)11'$, in this case as fast as $\lambda'$ converges to zero. The final assertion of Theorem 1 follows at once.

We now turn to the proof of Theorem 9. For this we need several preliminary results.

**Lemma 7:** Let $G$ and $H$ be directed graphs on the same $n$ vertices. Suppose that both graphs have self-arcs at all vertices. If there is an undirected path from $i$ to $j$ in $H \circ G$, then there is an undirected path from $i$ to $j$ in the union of $H$ and $G$.

**Proof of Lemma 7:** First suppose that there is an undirected path of length one between vertices $i$ and $j$ in $H \circ G$. Then, either $(i, j)$ or $(j, i)$ must be an arc in $H \circ G$. Without loss of generality suppose that $(i, j)$ is an arc in $H \circ G$. Then, because of the definition of composition, there must be a vertex $k$ such that $(i, k)$ is an arc in $G$ and $(k, j)$ is an arc in $H$. This implies that $(i, k)$ and $(k, j)$ are arcs in $H \cup G$. Thus, $H \cup G$ has a directed path from $i$ to $j$. Therefore, $H \cup G$ has an undirected path from $i$ to $j$. Now suppose that $(i, v_1), (v_1, v_2), \ldots, (v_g, j)$ is an undirected path in $H \circ G$ between $i$ and $j$. Between each pair of successive vertices along this path there must therefore be an undirected arc in $H \circ G$. Therefore, between each pair of successive vertices along this path there must be an undirected arc in $H \cup G$. Thus, there must be an undirected path in $H \cup G$ between $i$ and $j$.

It is obvious that the preceding lemma extends from two graphs to a finite set of directed graphs. In the following, we appeal to this extension without special mention. The next result characterizes completeness of a gossip matrix in terms of a property of its graph.

**Lemma 8:** A gossip matrix is complete if and only if its graph is weakly connected.

**Proof of Lemma 8:** Since $M$ is a gossip matrix, there exist single-gossip matrices $P_1, P_2, \ldots, P_m$ such that $M = P_mP_{m-1}\cdots P_1$. Write $(j, k)$ for the edge in $N$ associated with $P_i$ and let $G$ denote the spanning subgraph of $N$ whose edges are $(j_1, k_1), (j_2, k_2), \ldots, (j_m, k_m)$. Since each $P_i$ is a single-gossip matrix, its graph $\gamma(P_i)$ must contain arcs from $j_i$ to $k_i$ and from $k_i$ to $j_i$. This means that the union of the $\gamma(P_i)$, written $U$, will be weakly connected if and only if $G$ is a connected graph. Therefore, weak connectivity of $U$ is equivalent to $M$ being complete.

In general, the composition of two directed graphs contains the union of the two graphs whenever the two graphs in question have self-arcs at all vertices. This means that $U$ must be a subgraph of $\gamma(M)$ because each vertex of each $\gamma(P_i)$ has a self-arc. But it has just been shown that completeness of $M$ implies weak connectivity of $U$. It follows that completeness of $M$ must also imply weak connectivity of $\gamma(M)$.

Suppose next that $\gamma(M)$ is weakly connected. Then, $U$ must be weakly connected because of Lemma 7. Therefore, $M$ must be complete.

**Proof of Theorem 9:** Since $M$ is a product of matrices with positive diagonal entries, $M$ has positive diagonal entries. By Lemma 8, $M$ is complete if and only if $\gamma(M)$ is weakly connected. By Theorem 8, any doubly stochastic matrix with positive diagonal entries has a contraction coefficient less than one just in case the graph of the matrix is weakly connected. Therefore, $M$ is complete if and only if it has a contraction coefficient less than one.

As it stands the proof of Theorem 1 rests on the assumption that $\gamma(M)$ has a weakly connected graph. Nedić et al. [23] derives a result very similar to Theorem 1 under what at first glance appears to be a more restrictive assumption, namely that $\gamma(M)$ has a strongly connected graph. But because doubly stochastic matrices constitute a special class of stochastic matrices, the assumption of strong connectivity turns out to be not restrictive at all. Here is why.

**Lemma 9:** The graph $G$ of a doubly stochastic matrix $D$ is strongly connected if and only if it is weakly connected.\(^5\)

The proof of Lemma 9 which follows is based on ideas from [26] and [34]. Let $G$ be a directed graph with vertex set $\{1, 2, \ldots, n\}$. Call a vertex $j$ reachable from $i$ if either $j = i$ or if there is a directed path from $i$ to $j$. Call a vertex $i$ essential if $i$ is reachable from all vertices which are reachable from $i$.

**Lemma 10:** Every directed graph has at least one essential vertex.

**Proof of Lemma 10:** Suppose that $G$ has $n > 0$ vertices. If $G$ has an isolated vertex $i$, then $i$ is essential. Consider next the case when $G$ has no isolated vertices in which case $n > 1$. Suppose that $G$ has no essential vertices. Then, for each vertex $i$ there must be a vertex $j$ which is reachable from $i$ but from which $i$ is not reachable. It follows that it is possible to construct a sequence of vertices $i_1, i_2, \ldots, i_m$ of any length $m > 1$ such that $i_{j+1}$ is reachable from $i_j$, $j \in \{1, 2, \ldots, m-1\}$ but not conversely. But for $m > n$ at least one vertex must appear in the list twice, say in positions $j$ and $k > j$. This implies that vertex $j + 1$ is reachable from $j$ and conversely which is a contradiction. Therefore, $G$ must have an essential vertex.

To proceed, let us say that vertices $i$ and $j$ are mutually reachable if each is reachable from the other. Mutual reachability is clearly an equivalence relation on $V$ which partitions $V$ into the disjoint union of a finite number of equivalence classes. Note that if $i$ is an essential vertex of $G$, then every vertex in the equivalence class of $i$ is also essential. Thus, every directed graph possesses at least one mutually reachable equivalence class whose members are all essential.

\(^5\)It is clear that strong connectivity of $G$ implies weak connectivity of $G$. The converse was conjectured by John Tsitsiklis in a private communication.
Proof of Lemma 9: Strong connectivity clearly implies weak connectivity. We prove the converse. Suppose \( G \) is weakly connected. In view of the proceeding, \( G \) has at least one mutually reachable equivalence class \( \mathcal{E} \) whose members are all essential. If \( \mathcal{E} = \mathcal{V} \), then \( G \) is obviously strongly connected. Thus, to prove the lemma, it is enough to show that \( \mathcal{E} = \mathcal{V} \). Suppose the contrary, namely that \( \mathcal{E} = \{ \pi_1, \pi_2, \ldots, \pi_m \} \) is a strictly proper subset of \( \mathcal{V} \). Let \( \pi \) be any permutation map for which \( \pi(i) = j, j \in \{1, 2, \ldots, m\} \), and let \( P \) be the corresponding permutation matrix. Then, clearly
\[
P'^D = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}
\]
and \( P'^D \) is doubly stochastic. Since \( P'^D \) is doubly stochastic, the column sums of \( A \) must all equal one as must the row sums of the submatrix \( [A \ B] \). But the transformation \( D \rightarrow P'^D \) corresponds to a relabeling of the vertices of \( G \), so the graph of \( P'^D \) must also be weakly connected. This means that \( B \) cannot be the zero matrix. Therefore, the sum of the row sums of \( A \) must be less than \( m \). But this contradicts the fact that the sum of the column sums of \( A \) equals \( m \). Therefore, \( \mathcal{E} = \mathcal{V} \).

We are now in a position to prove Theorem 2.

Proof of Theorem 2: Suppose that \( M \) is a complete gossip matrix. In view of Lemmas 8 and 9, \( \gamma(M) \) is a strongly connected graph. Thus, by Lemma 6, the magnitude of the second largest eigenvalue of \( M \) is less than 1.

To prove the converse, now suppose that \( M \) is a gossip matrix whose second largest eigenvalue in magnitude is less than one. By Lemma 6 the graph of \( M \) is therefore weakly connected. Thus, by Lemma 8, \( M \) is complete.

V. CONCLUDING REMARKS

Let \( N \) be a given neighbor graph. Recall that a complete gossip sequence is minimal if there is no shorter sequence of gossips which is complete. It is easy to see that a complete gossip sequence will be minimal if and only if the gossip graph it induces is a minimal spanning tree in \( N \). For a given neighbor graph there can be many minimal spanning trees and consequently many minimally complete gossip sequences. Moreover, there can be differing second largest singular values and different second largest eigenvalues (in magnitude) for the different doubly stochastic matrices associated with different complete minimal sequences. A useful exercise then would be to determine those complete minimal sequences whose associated singular values or second largest eigenvalues (in magnitude) are as small as possible.

One of the problems with the idea of gossiping, which apparently is not widely appreciated, is that it is difficult to devise provably correct gossiping protocols which are guaranteed to avoid deadlocks without making restrictive assumptions. The research in this paper and in [12] and [13] contributes to our understanding of this issue and how to deal with it.

REFERENCES

C. Yu, B. D. O. Anderson, S. Mou, J. Liu, A. S. Morse, B. D. O. Anderson, and
Science and Engineering.

National Academy of Engineering and the Connecticut Academy of
Lecturer of the IEEE Control Systems Society and a member of the
from the IEEE Control Systems Society in 1993 and 2005, the American
systems, and, most recently, coordination and control of large grouping
control, urban transportation, vision-based control, hybrid and nonlinear
network synthesis, optimal control, multivariable control, adaptive
is currently the Dudley Professor of Engineering. His main interest is in
Yale University, New Haven, CT, faculty, where he
Research Center, Cambridge, MA, he joined the
Theory and Application (OCTA), NASA Electronics
engineering.

M.S. degree from the University of Arizona,
B.S. degree from Cornell University, Ithaca, NY, the
University, New Haven, CT, in 2007, where he is currently
working towards the Ph.D. degree in electrical engineering.
His research interests are in the areas of
multiagent systems and cooperative control.

Ji Liu (Student Member) was born in Shanghai,
China, in 1983. He received the B.S. degree in
information engineering from Shanghai Jiao Tong
University, Shanghai, China, in 2006 and the M.S.
degree in electrical engineering from Yale Univer-
sity, New Haven, CT, in 2007, where he is currently
working towards the Ph.D. degree in electrical engineering.
His research interests are in the areas of
multiagent systems and cooperative control.

Shaoshuai Mou (Student Member) was born in
Yantai, Shandong Province, China, in 1984. He
received the B.Eng. and M.Eng. degrees in aero-
space engineering from Harbin Institute of Tech-
nology, Harbin, China, in 2006 and 2008,
respectively, and the M.S. degree in electrical
engineering from Yale University, New Haven, CT,
in 2010, where he is currently working towards the
Ph.D. degree in electrical engineering under the
supervision of Prof. Morse.
He worked as a Research Associate in the University of Hong Kong,
Hong Kong, in 2008. His major research interests include distributed
averaging in sensor networks and formation control of autonomous agents.

A. Stephen Morse (Life Fellow, IEEE) received the
B.S. degree from Cornell University, Ithaca, NY, the
M.S. degree from the University of Arizona,
Tucson, and the Ph.D. degree from Purdue Univer-
sity, West Lafayette, IN in 1967, all in electrical
engineering.
Following three years at the Office of Control
Theory and Application (OCTA), NASA Electronics
Research Center, Cambridge, MA, he joined the
Yale University, New Haven, CT, faculty, where he is
currently the Dudley Professor of Engineering. His main interest is in
network synthesis, optimal control, multivariable control, adaptive
control, urban transportation, vision-based control, hybrid and nonlinear
systems, and, most recently, coordination and control of large grouping
of mobile autonomous agents.
Dr. Morse received the George S. Axelby Outstanding Paper Award from
the IEEE Control Systems Society in 1993 and 2005, the American
Automatic Control Council’s Best Paper Award (twice), and the IEEE
Technical Field Award for Control Systems in 1999. He is a Distinguished
Lecturer of the IEEE Control Systems Society and a member of the
National Academy of Engineering and the Connecticut Academy of
Science and Engineering.

Brian D. O. Anderson (Life Fellow, IEEE) was born
in Sydney, Australia. He was educated in mathe-
matics and electrical engineering at Sydney
University, Sydney, N.S.W., Australia, and received the Ph.D. degree in electrical engineering from
Stanford University, Stanford, CA, in 1966.
He is a Distinguished Professor at the Australian National
University, Canberra, A.C.T., Australia, and
Distinguished Researcher at the National ICT
Australia (NICTA). His current research interests
are in distributed control, sensor networks, and econometric modeling.
Dr. Anderson is a Fellow of the Australian Academy of Science, the
Australian Academy of Technological Sciences and Engineering, the
Royal Society, and a foreign associate of the National Academy of Engineering. His awards include the IEEE Control Systems Award of 1997,
the 2001 IEEE James H. Mulligan, Jr. Education Medal, and the Bode Prize
of the IEEE Control System Society in 1992, as well as several IEEE best
top paper prizes.

Changbin (Brad) Yu (Senior Member, IEEE) received the B.Eng. degree with first class honors
in computer engineering from Nanyang Techno-
lological University, Singapore, in 2004 and the
Ph.D. degree in information engineering from the
Australian National University, Canberra, A.C.T.,
Australia, in 2008.
Since then he has been on an academic staff at
the Australian National University and Adjunct
Researcher at the National ICT Australia (NICTA).
His current research interests include control of autonomous formations,
multiagent systems, mobile sensor networks, human-robot interaction
and graph theory.
Dr. Yu had won a competitive Australian Postdoctoral Fellowship (APD) in 2007 and a prestigious Queen Elizabeth II Fellowship (QEI) in
2010. He was also a recipient of Australian Government Edeavour Asia
Award (2005). Inaugural Westpac Australian Chinese Students New
Age Award (2006), Chinese Government Outstanding Overseas Stud-
dents Award (2006), Asian Journal of Control Best Paper Award
(2006-2009), etc.

ABOUT THE AUTHORS

REFERENCE