Forty-Fifth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 26-28, 2007

Probabilistic switching and convergence rate in consensus problems

Pedram Hovareshti, John S. Baras and Vijay Gupta

Abstract— **We consider the convergence problem of consensus seeking agents and the effect of probabilistic switching on the convergence rate. Using a probabilistic setting, we develop a framework to study the convergence rate for different classes of graph topologies, including Small World networks. We show that by making the effective diameter of the underlying graph small, probabilistic switching can be helpful in the design of fast consensus algorithms.**

I. Introduction

The study of distributed algorithms has been the subject of extensive research in recent years by the control systems community. There has been great interest in determining how agents with low capabilities can achieve global objectives using only local information. Consensus problems arise in many applications of collaborative control and their properties have been studied extensively in recent years (e.g. see [2], [20], [15], [11] and the references therein). Load balancing in multiple processes [7], averaging in sensor networks and Gossip algorithms [28], [27], [4], motion coordination problems [18], are some of the applications in which discrete consensus algorithm have arisen. Since the algorithm is decentralized, it is important to: (i) analyze the effect of graph topology on the convergence rate, (ii) design topologies with high convergence rate, and (iii) take into account the behavior of the algorithm in the presence of link/edge losses.

This paper is concerned with the first two problems mentioned, i.e. the analysis of the effect of graph topology on the convergence rate and the design of 'good' topologies. In the second section the related background which suggests the use of probabilistic switching is discussed. In the third section the probabilistic framework is provided. We use some existing results from the literature in conjunction with our results to build up our framework. We provide bounds for the convergence rate in the probabilistic setting. In the last section we examine our bounds for three classes of graphs. First we show that switching the neighbors in a ring structure results in a drastic improvement of the convergence rate. Then we study switching in Erdöš-Renyi Random graphs. We then return to the case of Small World graphs (which provided the inspiration for the present study) and analyze its behavior in the probabilistic framework.

Research supported by the National Aeronautics and Space Administration under award No NCC8235, and by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011.

The authors are with the Institute for Systems Research, University of Maryland College Park, MD 20742, USA.(hovaresp@umd.edu, baras@umd.edu, vijay@cds.caltech.edu)

II. Background and Motivation

Consider the discrete time consensus equation:

$$x(k+1) = F(k)x(k) \tag{1}$$

in which $F(k)$ is a nonnegative stochastic matrix. Assume that the algorithm is designed to work with a given graph topology G_1 and its corresponding matrix is F_1. The entry F_{ij} denotes the weight that node i applies to the values that it gets from node j, and its value may be time dependent. If there is no link between nodes i and j, $F_{ij} = 0$. Convergence of consensus algorithms have been extensively studied and different sufficient conditions have been proposed (see [11] and the references therein). The rate of convergence of the equation (1) is a function of the weights, as well as the graph topology. References [28], [23] have studied the problem of optimizing the rate of convergence by changing the weights when the topology is fixed.

In this paper we restrict our attention to two sets of weights which arise naturally in many applications and instead focus on changing the graph topology to get higher convergence rates. A natural set of weights arises from discretizing the continuous version of the consensus algorithm with time scale h. This results in a weight matrix F which has the value h as its ij entry, if there is a communication between the nodes i and j, and 0 otherwise. The resulting weight matrix is symmetric and doubly stochastic and can be represented as $F = I - hL(G)$, where $L(G)$ is the graph Laplacian matrix. Furthermore if we select $h < \frac{1}{2.d_{max}}$ then all the eigenvalues of the F matrix are non-negative [20]. This helps us in the analysis of Section 3. Another set of weights which we have considered in this paper comes from Vicsek's model for leaderless coordination, [15], [25]. Vicsek's model assigns weights to neighbors of a node in a way in which each node performs a local averaging in its neighborhood. This corresponds also to a natural random walk on a graph with self loops. For Vicsek's Model $F = (I+D)^{-1}(A+I)$, where A is the adjacency matrix of the graph G and D is the diagonal matrix with each node's degree on the corresponding diagonal.

In the case of fixed graph topology, the second largest eigenvalue modulus (SLEM) of the corresponding F matrix determines the convergence speed. This is because,

$$x(\infty) - x(t) = (F^{\infty} - F^{t})x(0) \tag{2}$$

If F is a primitive stochastic matrix, according to the Perron-Frobenius theorem [21], $\lambda_1 = 1$ is a simple eigenvalue with a right eigenvector $\mathbf{1}$ and a left eigenvector π such that $\mathbf{1}^T\pi = 1$, $F^{\infty} = \mathbf{1}\pi^T$ and if $\lambda_2, \lambda_3, ..., \lambda_r$ are the other eigenvalues of F, ordered in a way such that $\lambda_1 = 1 > |\lambda_2| \geq |\lambda_3| \geq ... \geq |\lambda_r|$. If m_2 is the algebraic multiplicity of λ_2, then

$$F^t = F^\infty + O(t^{m_2-1}|\lambda_2|^t) = \mathbf{1}\pi^T + O(t^{m_2-1}|\lambda_2|^t) \quad (3)$$

where $O(f(t))$ represents a function of t such that there exists $\alpha, \beta \in R$, with $0 < \alpha \leq \beta < \infty$, such that $\alpha f(t) \leq O(f(t)) \leq \beta f(t)$ for all t sufficiently large. This shows that the convergence of the consensus protocol is geometric, with relative speed equal to SLEM. We denote by $\mu = 1 - SLEM(G)$ the spectral gap of a graph, so graphs with higher spectral gaps converge more quickly. If the matrix F is symmetric, its SLEM can be written as its norm restricted to the subspace orthogonal to $\mathbf{1} = [111...1]^T$. However, the F matrices are not symmetric in general. In fact although the underlying graph structure is symmetrical, the weight that each node applies to another node is determined by its own degree. In general the SLEM of F matrices are not easily computable.

For the general case where topology changes are also included, Blondel *et al* [2] showed that the joint spectral radius of a set of matrices derived from F matrices determines the convergence speed. For Σ a finite set of $n \times n$ matrices, their joint spectral radius is defined as:

$$\rho = \limsup_{t\to\infty} \max_{A_1,\ldots,A_t \in \Sigma} ||A_t ... A_1||^{1/t} \quad (4)$$

Calculation of the joint spectral radius of a set of matrices is a mathematically hard problem and is not tractable for large sets of matrices. Using ergodic coefficients of blocks of matrices as in [13] can provide us with geometric rates. However, it is worthwhile to notice that graphs with well-connected nodes guarantee fast convergence. This is a direct result of the Cheeger inequality, which relates the spectral gap of an F matrix to the conductance of the corresponding graph [5]. Switching over such topologies will also result in good convergence speed; see our earlier work in [1].

Since agents usually have energy constraints, the number of agents with which they communicate is limited. Therefore an important design issue is to find topologies which induce certain performance, provided that the number of the links each agent can establish is less than an upper bound.

Watts and Strogatz [26] introduced and studied a simple tunable model that can explain behavior of many real world complex networks. Their "Small World" model takes a regular lattice and replaces the original edges by random ones with some probability $0 \leq \phi \leq 1$. It is conjectured that dynamical systems coupled in this way would display enhanced signal propagation and global coordination, compared to regular lattices of the same size. The intuition is that the short (i.e. direct) paths between distant parts of the network cause high speed spreading of information which may result in fast global coordination. Olfati-Saber [19] studied continuous time consensus protocols on small world networks and proposed some conjectures. We have used a variant of the Newman-Moore-Watts [17], the improved form of the $\phi-$model originally proposed by Watts and Strogatz [1]. The model starts with a ring of n nodes, each connected by undirected nodes to its nearest neighbors to a range k. Shortcut links are added -rather than rewired- between randomly selected pairs of nodes, with probability ϕ per link on the underlying lattice; thus there are typically $nk\phi$ shortcuts. In reference [1] we actually forced the number of shortcuts to be equal to $nk\phi$ (comparable to the Watts $\phi-$model) and used Vicsek's weights. We studied different choices of initial lattices. Figure 1 shows the effect of added shortcuts to a base ring $C(1000, 5)$.

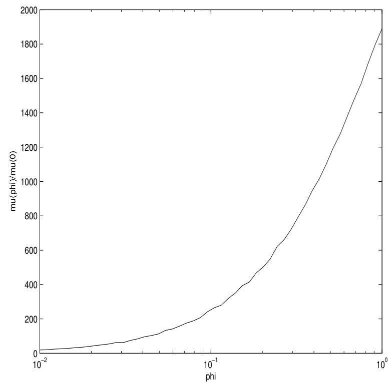

Fig. 1. Spectral gap gain for $(n, k) = (1000, 5)$

Our Simulation results show that adding a small number of links to a ring-structured graph should result in high convergence rate. However analytical verification of this result is difficult. Here we try to justify our result using a "mean field" approach and perturbation analysis. In the ϕ model of Small World graphs and its variants, a regular lattice is considered and m shortcuts are added randomly where m is equal to a proportion ϕ of the lattice's initial edges. In the present analysis, following [14], we reflect the effect of shortcuts by adding "small " nonzero positive numbers to the entries of F corresponding to non-adjacent nodes of the lattice; a method we also used in [1] This corresponds to using lots of shortcuts with negligible weights on them. Although by adding a uniform perturbation the topology of the graph is not respected, the analysis gives insight on random communication patterns for Small World networks. We state the result for the case where the base lattice is a ring but the result can be extended to $C(n, k)$ for other ks. We show the results for Vicsek's weights. It can be verified that the same results hold for weight matrices of the form $F = I - hL$.

We follow the perturbation approach to Small World networks proposed by Higham [14]. Consider the base lattice to have a ring topology on n nodes, $G(n, 2)$ and the corresponding F matrix F_0. This can be also viewed as a random walk with self loops. This is similar to a particular case of our base circulant matrix F_0. Therefore the base matrix is:

$$F_0 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & ... & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & ... & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & ... & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \frac{1}{3} & 0 & 0 & ... & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad (5)$$

We know that:

Corollary 2.1: $SLEM(F_0) = \frac{1}{3}[1 + 2cos(\frac{2\pi}{n})]$. Furthermore it has multiplicity at least 2.

Now we perturb the nonzero entries of the matrix F_0 by $\epsilon = \frac{K}{n^\alpha}$ for fixed $K > 0$ and $\alpha > 1$ in the limit $N \to \infty$, to get the perturbed matrix F_ϵ:

$$F_\epsilon = \begin{pmatrix} \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \epsilon & \epsilon & ... & \epsilon & \frac{1}{3} - \frac{(n-3)\epsilon}{3} \\ \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \epsilon & ... & \epsilon & \epsilon \\ \epsilon & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & ... & \epsilon & \epsilon \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \epsilon & \epsilon & \epsilon & ... & \frac{1}{3} - \frac{(n-3)\epsilon}{3} & \frac{1}{3} - \frac{(n-3)\epsilon}{3} \end{pmatrix}$$

We call the "shortcuts" created this way $\epsilon-$shortcuts. F_ϵ is also a circulant matrix. The representer —citepeba07 of this circulant is

$$p_a(z) = \frac{1}{3} - \frac{(n-3)\epsilon}{3} + (\frac{1}{3} - \frac{(n-3)\epsilon}{3})z + \quad (6)$$
$$\epsilon z^2 + \epsilon z^3 + ... + \epsilon z^{n-2} + (\frac{1}{3} - \frac{(n-3)\epsilon}{3})z^{n-1}$$

So, the eigenvalues of this matrix are $\lambda_i(\epsilon) = p_a(\omega^{i-1})$. For this matrix the largest eigenvalue is 1. Using a similar argument the SLEM can be calculated to be equal to:

$$\lambda_2(\epsilon) = (\frac{1}{3} - \frac{n}{3}\epsilon)(1 + 2cos\frac{2\pi}{n}) \quad (7)$$

Thus we can state the following proposition:

Proposition 2.1: Let $\epsilon = \frac{K}{n^\alpha}$, $\alpha \geq 1$.

- For $\alpha > 3$, the effect of ϵ-shortcuts on convergence rate is negligible. $\alpha = 3$ is the onset of the effectiveness of shortcuts.
- For $\alpha = 2$, the shortcuts are dominantly decreasing SLEM.
- For $\alpha = 1$, almost all of the nodes communicate effectively and thus the SLEM is very small.

Proof

For large n we can write:

$$\lambda_2(F_\epsilon) = \frac{1}{3} + \frac{2}{3}cos\frac{2\pi}{n} - \frac{n\epsilon}{3} - \frac{2n}{3}cos\frac{2\pi}{n} \Rightarrow$$

$$\lambda_2(F_\epsilon) = 1 - \frac{4\pi^2}{3n^2} + o(\frac{1}{n^4}) - n\epsilon + \frac{4\pi^2\epsilon}{3n} + o(\frac{1}{n^3}) \quad (8)$$

The first three terms are the contributions of the base lattice and the rest are the contributions of the perturbation. Comparing this to the SLEM of the base lattice

$$\lambda_2(F_0) = \frac{1}{3}(1 + 2cos\frac{2\pi}{n}) = 1 - \frac{4\pi^2}{3n^2} + o(\frac{1}{n^3}) \quad (9)$$

yields the following results.

For the base lattice, the spectral gap decreases as fast as n^2. If ϵ is $o(n^\alpha), \alpha > 3$, then terms coming from the lattice are dominant, and therefore the shortcuts does not affect the spectral gap. For $\alpha = 3$ the terms regarding the shortcuts will be of the same degree as the terms from the base and for k large enough, the SLEM starts decreasing from the corresponding lattice SLEM. For $\alpha = 2$ the terms regarding the shortcuts are dominant and the SLEM has considerably decreased compared to the base lattice. Only for the case of $\alpha = 1$ the spectral gap does not vanish as $n \to \infty$.

As observed above $\epsilon-$shortcuts are loosely analogous to the shortcuts in the ϕ-model. Since the Small World model is a probabilistic model, we anticipate that adding small weights is analogous to choosing graphs with low probability shortcuts. In the following section we will formalize this idea. To this end we need a framework for studying consensus problems with probabilistic switching.

A recent book by Durrett [9] addresses the mixing time of Markov Chains on small world graphs, which is closely related to the subject of this paper. A similar approach has also been used by the authors of [24].

III. Probabilistic framework

Consider n nodes and a finite set of n by n stochastic matrices $\mathbb{F} = \{F_1, ..., F_m\}$. At each time, graph G_i is selected with probability p_i. The choice of graph topology at each time is identically distributed and independent of the graph topologies selected at previous times. This framework includes a vast host of applications, including the case of i.i.d. link losses. Conditions for convergence of probabilistic consensus schemes and the rate of convergence have been studied in [24], [12], [4], [10]. Here we mention a result from [10]. Using results from [6], the authors of [10] have obtained the following, which constitutes a sufficient condition for high probability convergence of probabilistic consensus.

Let $Q(t) = F(t-1)...F(0)$.

Theorem 3.1: (Theorem 3.1 in [10]) The algorithm achieves probabilistic consensus if and only if for every two nodes i and j, $P(\mathbb{E}_{i,j}) = 1$, where $\mathbb{E}_{i,j} = \{\exists k, \exists t, Q_{ik}(t)Q_{jk}(t) > 0\}$.

Let $\overline{F} = E[F(t)]$ and $\overline{G}$ be the corresponding graph. Then:

Corollary 3.1: (Corollary 3.2 in [10]) Assume that for any node i, $F_{ii}(t) > 0$ almost surely. If $\overline{G}$ is strongly connected, then $F(t)$ achieves probabilistic consensus.

Using the fact that in our set of graphs there exists at least one connected graph and the communication is considered to be symmetric, we get the following simpler condition for convergence.

Proposition 3.1: The consensus algorithm with i.i.d link losses and symmetric communication ((i, j) is an edge iff (j,i) is an edge) converges with probability 1, if $\mathbb{F}$ is such that there is a positive probability that any two nodes communicate.

Proof: First note that since F_1 is a strongly connected graph, by definition $\mathbb{F}$ satisfies the condition that there is a positive probability that any two nodes communicate. The communication is symmetric and there exist a positive α such that for all k and i, $f_{ii}(k) \geq \alpha > 0$. Furthermore $f_{ij}(k) \in \{0\} \bigcup [\alpha, 1]$, and $\sum_{j=1}^{n} f_{ij}(k) = 1$. By Theorem 2 of [2], the algorithm converges if the graph $(N, \bigcup_{s\geq t} E(s))$ is strongly connected for all $t \geq 0$. But this is the case with probability 1. The reason is that for any two node i and j there is a nonzero probability that there exist a chain of edges that connects these two nodes at each time. If we denote this probability by q_c, and pick an arbitrary time t_0, then the probability that starting from t_0 the graph $(N, \bigcup_{s\geq t} E(s))$ is connected is given by: $lim_{k\to\infty} 1-(1-q_c)^k \to 1$. Since the link drops are i.i.d. the argument is valid starting at any t_0. Thus consensus will be reached with probability 1. ∎

To calculate the rate of convergence we consider the following Lyapunov function:

$$V(\mathbf{x}) = \frac{1}{n}[\sum_{i\neq j} E(||x_i - x_j||^2)] = E[\mathbf{x}^T(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n})\mathbf{x}] = E[\mathbf{x}^T \hat{L}\mathbf{x}]$$

in which $\hat{L} = (\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{n})$ is the Laplacian of the complete graph. Note that $\hat{L}^2 = \hat{L}$. If we consider $P(\mathbf{x}) = E[\hat{L}\mathbf{x}\mathbf{x}^T\hat{L}]$, then we will have $V(\mathbf{x}) = Tr(P(\mathbf{x}))$ and each entry of P converges to zero as consensus is reached. Let $G(x) = E[\mathbf{x}\mathbf{x}^T]$, then $P(\mathbf{x}) = \hat{L}G(\mathbf{x})\hat{L}$ and the rate of convergence of P to zero is equal to the rate of convergence of G to a constant matrix. We have:

$$G(k+1) = E[\mathbf{x}(k+1)\mathbf{x}(k+1)^T] = E[F(k)\mathbf{x}(k)\mathbf{x}(k)^T F(k)^T]$$

Therefore by vectorizing the matrix $G(k+1)$ we get,

$$Vec(G(k+1)) = E[F(k) \otimes F(k)]Vec(E[\mathbf{x}(k)\mathbf{x}(k)^T])$$

So, we get the iteration:

$$Vec(G(k+1)) = E[F(k) \otimes F(k)]Vec(G(k)) \qquad (10)$$

Therefore if in equation (10), $Vec(G(k+1))$ converges to a vector of equal entries with some rate, x will also reach consensus with the same rate. Equation (10) is a linear iteration. In the case of i.i.d. switchings it is time invariant, i.e. $A = E[F(k) \otimes F(k)]$ is a constant matrix. Since A is stochastic, the convergence of the equation (10) is governed by its second largest eigenvalue modulus (SLEM). [4] shows that in the case of symmetric weight matrices, the rate of convergence of equation (10) is equal to the SLEM of the matrix $E[F(k) \otimes F(k)]$.

In the following, we exploit certain graph topologies to determine this rate. We consider weight matrices of the form $F = I - hL$, with $h < \frac{1}{2d_{max}}$. These assumptions cause the F matrices to be symmetric with nonnegative eigenvalues. This simplifies the analysis. The following example illustrates the method.

Example 3.1: **Complete graph with probabilistic link losses**

First consider the case of two nodes with a link which drops with probability p. Therefore: $A = E[F \otimes F] = (1-p)(F_1 \otimes F_1) + p(I_2 \otimes I_2)$, where:

$$F_1 = \begin{pmatrix} 1-h & h \\ h & 1-h \end{pmatrix}$$

Let Γ_2 denote the two dimensional forward shift permutation matrix,

$$\Gamma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

.

Since A commutes with $\Gamma_2 \otimes I$, A is a block circulant matrix. Since A also commutes with $A \otimes \Gamma_2$, the blocks themselves are also circulant matrices. Therefore A is a block circulant matrix with circulant blocks and its eigenvalues are simply computable given the eigenvalues of each block [8]. It can be easily verified that the rate of convergence in this example is equal to $\lambda_2 = (1-p)(1-2h) + p$.

Now consider the case of a complete graph on n vertices, where the link losses occur with probability $1-p$. This means that at each time instance an Erdos-Renyi random graph $G(n, 1-p)$ is present. Therefore it is as if at each time we are selecting randomly among the set of $G(n, 1-p)$ random graphs. We are interested in calculating the second largest eigenvalue of

$$A = E[(I_n - hL) \otimes (I_n - hL)] = I_n - hI_n \otimes \bar{L} - h\bar{L} \otimes I_n + h^2 E[L \otimes L]$$

We show that

Proposition 3.2: The matrix $A = E[(I_n - hL) \otimes (I_n - hL)]$ can be transformed to a block-circulant matrix by a permutation.

The exact permutation matrix and the structure of the block-circulant matrix are given throughout the proof.

Proof: Let Γ denote the forward shift permutation matrix. The matrix A can be written as:

$$\begin{pmatrix} A_{11} & A_{12} & ... & A_{1n} \\ A_{21} & A_{22} & ... & A_{2n} \\ . & . & ... & . \\ . & . & ... & . \\ A_{n1} & A_{n2} & ... & A_{nn} \end{pmatrix}$$

Each block is a symmetric $n \times n$ matrix and $A_{ij} = A_{ji}$. The block A_{ij} contains contributions from the Laplacian of graphs that have (i,j) as a link. Furthermore, because of the symmetry, the set of all possible graphs can be decomposed into classes of cyclic isomorphic graphs, each of which contains circular shifts of a given graph, i.e. if the graph G is in isomorphic class i, then so are the graphs $\Gamma G \Gamma^*, \Gamma^2 G \Gamma^{2*}, ..., \Gamma^n G \Gamma^{n*}$. Therefore for all i and j we have the following result modulo n:

$$A_{ij} = \Gamma A_{i+1,j+1}\Gamma^* = \Gamma A_{i+1,j+1}\Gamma^{n-1}$$

Now take the block-diagonal matrix

$$J = \begin{pmatrix} I_n & 0 & ... & 0 \\ 0 & \Gamma_n & ... & 0 \\ . & . & ... & . \\ 0 & 0 & ... & \Gamma_n^{n-1} \end{pmatrix}$$

We will now show that the matrix $K = JAJ^*$ is block circulant. To see this, notice that

$$K = \begin{pmatrix} K_{11} & K_{12} & ... & K_{1n} \\ K_{21} & K_{22} & ... & K_{2n} \\ . & . & ... & . \\ . & . & ... & . \\ K_{n1} & K_{n2} & ... & K_{nn} \end{pmatrix}$$

where the ij^{th} block $K_{ij} = \Gamma^{i-1}A_{ij}\Gamma^{j-1*}$ modulo n. Therefore modulo n we have:

$$\begin{aligned} K_{i+1,j+1} &= \Gamma^i A_{i+1,j+1}\Gamma^{j*} \\ &= \Gamma^{i-1}\Gamma A_{i+1,j+1}\Gamma^*\Gamma^{j-1*} \\ &= \Gamma^{i-1}A_{ij}\Gamma^{j-1*} \\ &= K_{ij} \end{aligned}$$

Therefore the matrix K is block circulant. Therefore using block Fourier matrices, the matrix K can be block-diagonalized to

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & ... & 0 \\ 0 & \Lambda_2 & ... & 0 \\ . & . & ... & . \\ 0 & 0 & ... & \Lambda_n \end{pmatrix}$$

where Λ_k is unitarily similar to

$$K_{11} + \omega^k K_{12} + \omega_{2k} K_{13} + ... + \omega^{(n-1)k} K_{1n}$$

■

A fundamental observation is that in general calculating the Kronecker product is difficult and requires $n^2 \times n^2$ matrix operations for n agents, unless the dimensions of the F matrices are small and certain symmetry conditions are present. Even under symmetric conditions the operations are not trivial and require large matrix eigenvalue calculations. Therefore we are interested in finding lower and upper bounds on the rate of convergence. The following proposition provides such bounds:

Proposition 3.3: Let $A = E[(I - hL) \otimes (I - hL)]$, $\bar{L} = E[L]$ and λ_i denote the i^{th} largest eigenvalue of a matrix, i.e.

$$\lambda_1(\bar{L}) \geq \lambda_2(\bar{L}) \geq ... \geq \lambda_n(\bar{L})$$

then $1-h\lambda_{n-1}(\bar{L}) \leq \lambda_2(A) \leq 1-h\lambda_{n-1}(\bar{L})+h^2\lambda_n(E[L\otimes L])$.

The result given by proposition (3.3) indicates that for finding bounds on the convergence rate of probabilistic consensus algorithms on a set of matrices $\mathbb{F}$, we should

1) Find the exact value or bounds for $\lambda_{n-1}(E[L])$,
2) Find the exact value or bounds for $\lambda_1(E[L \otimes L])$

As the examples in the next section will show $\lambda_{n-1}(E[L])$ can be computed for many different classes of graphs. To find bounds on $E[L\otimes L]$, we use Jensen's inequality and the fact that all the Laplacian eigenvalues of a graph are twice the maximum degree of graph. Therefore we get:

$$\lambda_1(E[L\otimes L]) \leq E(\lambda_1([L\otimes L])) = E[(\lambda_1(L))^2] \leq 4E[d^2_{max}].$$

Proposition (3.3) can be proved using a result in Matrix perturbation theory. We quote the result as given in [22], Theorem 4.8 and Corollary 4.9 with some changes in notation:

Theorem 3.2: [22] Let A_0 be a Hermitian matrix with eigenvalues

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$$

and $A = A_0 + E$ be a Hermitian perturbation of A with eigenvalues

$$\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq ... \geq \tilde{\lambda}_n$$

Furthermore, let the eigenvalues of E be

$$\epsilon_1 \geq \epsilon_2 \geq ... \geq \epsilon_n$$

then for $i = 1, 2, ..., n$

$$\tilde{\lambda_i} \in [\lambda_i + \epsilon_n, \lambda_i + \epsilon_1] \tag{11}$$

Proof: [22], pp 200-201. ■

We can then proceed to the proof of Proposition (3.3)

Proof: (of Proposition (3.3) Let

$$A = E[(I-hL)\otimes(I-hL)] = I-hI\otimes\bar{L}-h\bar{L}\otimes I+h^2E[L\otimes L]$$

,

$$A_0 = I - hI \otimes \bar{L} - h\bar{L} \otimes I$$

and

$$E = h^2E[L \otimes L]$$

The Kronecker product of the Laplacian has the smallest eigenvalue equal to zero. $E[L\otimes L]$ preserves the property that its row sums all equal zero. Therefore, $\epsilon_n = 0$. Furthermore,

$$\lambda_i(A_0) = \lambda_i(I - hI \otimes \bar{L} - h\bar{L} \otimes I) = 1 - h\lambda_{n-i}(\bar{L})$$

Therefore

$$\lambda_2(A) \in [1 - h\lambda_{n-1}(\bar{L}), 1 - h\lambda_{n-1}(\bar{L}) + h^2\lambda_1(E[L \otimes L])]$$

■

IV. Improving the Convergence Rate by Probabilistic Switching

In this section we use the framework developed in the previous sections to compare the convergence rate in the presence of probabilistic switching with that of fixed topology for three different classes of matrices. The aim is to show that in cases where switching results in reducing the effective diameter of the graph, the convergence in the switching case is order of magnitude faster than the convergence rate of the fixed topology. In our last example, we will return to the case of Small World networks and will give an analytical description of the Small World effect.

Example 4.1: **Changing neighbors in a ring topology** We consider n agents placed on a ring with n empty slots. Being on a ring, constraints each node's neighbors to the agent to its left and the agent to its right. We consider a naive model of motion in which at each time each agent changes its position uniformly at random and ends up in a random slot. Each slot contains one node at each time. This is equivalent to the assumption that at each time agents randomly choose their neighbors bi-directionally. We compare the rate of convergence of this scheme with the case in which there is a fixed ring topology. In the case of the fixed ring topology G_0, it can be easily verified that the second largest eigenvalue modulus of the iteration matrix F_0 is equal to $\lambda_2 = 1-2h\cos(\frac{2\pi}{n})$ and therefore the spectral gap is $2h[1-\cos(\frac{2\pi}{n})] = 4h\sin^2(\frac{\pi}{n})$. For large n this quantity is approximately equal to $\frac{4\pi^2 h}{n^2}$ and varies as n^{-2}. To compute the $\bar{L}$ matrix, we use the fact that $\bar{L}_{ii}$ equals the expected number of agent i's neighbors, which is 2 in this case. $-\bar{L}_{ij}$ is equal to the probability that the agents i and j are neighbors. Therefore:

$$\bar{L} = \begin{pmatrix} 2 & -\frac{2}{n-1} & \dots & -\frac{2}{n-1} \\ -\frac{2}{n-1} & 2 & \dots & -\frac{2}{n-1} \\ . & . & \dots & . \\ -\frac{2}{n-1} & -\frac{2}{n-1} & \dots & 2 \end{pmatrix} \qquad (12)$$

It can be easily seen that $\lambda_{n-1}(\bar{L}) = 2 + \frac{2}{n-1}$. Therefore $\lambda_2(I - h\bar{L}) = 1 - 2h - \frac{2h}{n-1}$. Also notice that the maximum degree of each node is 2. Using Proposition 3.3 we get:

$$h(2+\frac{2}{n-1}) - 16h^2 \leq 1 - \lambda_2(A) \leq h(2+\frac{2}{n-1})$$

Taking a constant $h < \frac{1}{8}$, and denoting the spectral gap $1-\lambda_2$ by S.G., it can be seen that:

$$\frac{S.G.(fixed)}{S.G.(switching)} \leq \frac{4h\sin^2(\frac{\pi}{n})}{4h + \frac{2h}{n-1} - 16h^2}$$

For the limit of large n the numerator varies as n^{-2} whereas the denominator varies as n^{-1}, therefore the ratio approaches zero with increasing n. It can be seen that even the lower bound of the spectral gap of the switching case shows order of magnitude improvement compared to the spectral gap of the fixed topology.

Example 4.2: **Erdos-Renyi Random graphs** Consider the case in which there exists a link between any two nodes with probability $q \in (0,1]$. The existence of a link between any two nodes is therefore random and independent of the other connections. In this example we compare the convergence rates between two cases. In one case a fixed random graph is used for all the time instants. In the other case the random graph changes with time in a way that at each time the choice of the random graph at each time is independent of the choice of the random graph at other time instants. For the former case we use a high probability bound due to Fiedler, reported in [16], [12]. For the latter case we calculate the bounds given by Proposition 3.3. In [16], [12] the authors report a high probability result for the second smallest eigenvalue of the Laplacian of a random graph. For the limit of large n and for $\epsilon \in (0,2)$:

$$\lim_{n\to\infty} Pr\{qn - \sqrt{(2+\epsilon)q(1-q)n\log n} <$$
$$\lambda_{n-1}(L(G(n,q))) < qn - \sqrt{(2-\epsilon)q(1-q)n\log n}\} = 1.$$

Therefore if a fixed random topology is used at all times then with high probability:

$$1 - hqn + h\sqrt{(2-\epsilon)q(1-q)n\log n} < \lambda_2[I - hL_{fixed}] <$$
$$1 - hqn + h\sqrt{(2+\epsilon)q(1-q)n\log n}$$

On the other hand, if at different times we switch between different random graphs, then $\bar{L}_{11} = E[deg(1)]$. Consider $X_2, X_3, ..., X_{n-1}$ be independent Bernoulli random variables. X_i is the indicator of existence of a link between nodes 1 and i. Therefore $L_{ii} = L_{11} = q(n-1)$. Also, $\bar{L}_{ij} = E[-\mathbf{1}\{(i,j) \in E(G)\}] = -q$. Hence:

$$\bar{L} = q\begin{pmatrix} n-1 & -1 & \dots & -1 \\ -1 & n-1 & \dots & -1 \\ . & . & \dots & . \\ -1 & -1 & \dots & n-1 \end{pmatrix}$$

and $\lambda_{n-1}(\bar{L}) = qn$. Using Proposition 3.3, we get:

$$1 - hnq < \lambda_2(E[(I-hL)\otimes(I-hL)]) <$$
$$1 - hnq + h^2\lambda_{max}[E(L\otimes L)] < 1 - hnq + 4h^2 E[d_{max}^2]$$

An interesting case arises when $q = \Theta(\frac{\log n}{n})$. In this case, with high probability, the random graph is connected'[3]. For a given large n, taking $q = k\frac{\log n}{n}$ with $k \geq 2$,

$$S.G.(fixed) < hqn - h\sqrt{(2+\epsilon)q(1-q)n\log}$$
$$\sim (k - \sqrt{2k})h\log n$$

On the other hand for the switching case we have:

$$S.G.(Switching) > hnq = kh\log(n)$$

Both of the bounds are of the same order and therefore it can be seen in the regime where the random graph is connected, its diameter is small enough so that switching among graphs of the same family cannot help improving the convergence rate.

Example 4.3: **Small World graphs** We now return to our analysis of small world graphs. We consider a ring structure and assume that at each time each agent can establish shortcuts with small probability ϵ. By using the probabilistic framework of Proposition 3.3 we study the effect of the shortcut probability on the convergence rate. To this end we take $\epsilon \propto n^{-\alpha}$ for $\alpha = 1, 2, 3, ...$ and observe the variations of the convergence rate for different choices of α. Contrary to our "mean field" approach of previous sections, this approach has an actual physical meaning.

The iteration matrix of the fixed ring graph in this setting is:

$$I - hL_{fix} = \begin{pmatrix} 1-2h & h & 0 & \dots & h \\ h & 1-2h & h & \dots & 0 \\ . & . & . & . & . \\ h & 0 & \dots & h & 1-2h \end{pmatrix}$$

Its second largest eigenvalue modulus can be easily calculated to be $\lambda_2(F_0) = 1 - 2h + 2h\cos(\frac{2\pi}{n})$ and its spectral gap is equal to:

$$S.G.(fixed) = 2h(1 - \cos(\frac{2\pi}{n}))$$

In the limit of large n, the spectral gap varies as n^{-2}.

To calculate the expected Laplacian matrix, notice that for all i, j with $|i-j| = 1$ (i.e. the adjacent neighbors of the base ring), $\bar{L}_{ij} = -1$ and for all i, j with $i \neq j$ and $|i - j| \neq 1$ we have, $\bar{L}_{ij} = -\epsilon$. Furthermore we use the fact that $\bar{L}_{ii}$ is the expected number of the neighbors of node i to get:

$$\bar{L}_{ii} = 2 + (n - 3)\epsilon$$

It can be verified that

$$\lambda_{n-1}(\bar{L}) = 2 + (n - 3)\epsilon - 2(1 - \epsilon)\cos(\frac{2\pi}{n})$$

Using Proposition 3.3 we get the following bounds for the switching case:

$$\begin{aligned}&\lambda_2(E[(I - hL) \otimes (I - hL)]) \in \\ &[1 - h(2 + (n - 3)\epsilon - 2(1 - \epsilon)\cos(\frac{2\pi}{n})), \\ &1 - h(2 + (n - 3)\epsilon - 2(1 - \epsilon)\cos(\frac{2\pi}{n})) + 4h^2 E[d_{max}^2]]\end{aligned}$$

and therefore:

$$\begin{aligned}&S.G.(Switching) \in \\ &[h(2 + (n - 3)\epsilon - 2(1 - \epsilon)\cos(\frac{2\pi}{n})) - 4h^2 E[d_{max}^2], \\ &h(2 + (n - 3)\epsilon - 2(1 - \epsilon)\cos(\frac{2\pi}{n}))]\end{aligned}$$

For $\epsilon = n^{-\alpha}$ and $\alpha = 1, 2, 3, ...$ using a Chernoff bound, it can be easily shown that for the limit of large n almost surely:

$$d_{max} < \log n$$

Taking $h \propto n^{-1}$, we first compare the spectral gap of the fixed topology with the upper bound for the spectral gap of the switching case. In the fixed case for large n, the dominant term is the n^{-3} term. In the switching case, the lower bound for the limit of large n is approximately equal to: $h[n^{1-\alpha} - n^{-\alpha} + 2\pi^2 n^{-2} - 2n^{-\alpha-2}] = o(n^{-\alpha})$ For $n \geq 3$ the dominant term is also the n^{-3} term. So, at its best performance adding shortcuts with probability less than or equal n^{-3} cannot help increasing the spectral gap. The lower bound on the spectral gap of the switching case in the limit of large n is approximately equal to: $h[n^{1-\alpha} - n^{-\alpha} + 2\pi^2 n^{-2} - 2n^{-\alpha-2}] - \frac{(\log n)^2}{n^2}$. If we take $\alpha = 1$, then this bound is $o(n^{-1})$, which is order of magnitude better than the ring case. If we take $\alpha = 2$ then the lower bound is not tight enough to make any statement about the comparison of the two regimes. However, because of the result we got for the $\alpha > 3$ case, we can conclude that $\alpha = 2$ is the onset of Small World phenomena.

V. Conclusions and Future Work

In this paper we developed a probabilistic framework for studying the effect of probabilistic switching on the convergence rate of consensus algorithms. We found bounds for the convergence rate and using 3 different classes of graph topologies, showed that the convergence rate can be improved if the effective diameter of the graph is reduced via probabilistic switching. We addressed the convergence of consensus algorithms on Small World graphs by two methods and showed that the Small World phenomena happens provided that the shortcut probabilities or weights are larger than a threshold. Future work addresses finding tighter bounds on the convergence of probabilistic switching and studying the effect of probabilistic switching when switching happens as a result of the motion of the nodes rather than being a design parameter.

References

[1] J.S. Baras and P. Hovareshti. Effects of graph topology on performance of distributed algorithms for networked control and sensing. *Proceedings of the Workshop on Networked Distributed Systems for Intelligent Sensing and Control, Kalamata, Greece, (http://med.ee.nd.edu/)*, 2007.

[2] V. Blondel, J. Hendrickx, A.Olshevsky, and J. Tsitsiklis. Convergence in multiagent coordination, consensus and flocking. *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 2996–3000, 2005.

[3] Bella Bollobas. *Modern Graph Theory*. Springer Verlag, 1998.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2005.

[5] P. Bremaud. *Markov chains, Gibbs Fields, monte Carlo Simulations and Queues*. Springer Verlag, 2nd edition, 1999.

[6] R. Cogburn. On products of random stochastic matrices. *Contemporary mathematics*, 50:199–213, 1986.

[7] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *Journal on parallel and distributed computing*, (7):279–301, 1989.

[8] P. J. Davis. *Circulant Matrices*. Wiley, 1979.

[9] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2007.

[10] F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. *submitted*, 2007. available at http://calvino.polito.it/ fagnani/quantcontrol/Random

[11] L. Fang and P. Antsaklis. On communication requirements for multi-agents consensus seeking. *Proceedings of Workshop NESC05: University of Notre Dame, Lecture Notes in Control and Information Sciences (LNCIS), Springer*, 331:53–68, 2006.

[12] Y. Hanato and M. Mesbahi. Agreement over random networks. *Proceedings of IEEE Conference on Decision and Control*, 2004.

[13] D. J. Hartfiel. *Nonhomogeneous Matrix products*. World Scientific, 2002.

[14] D.J. Higham. A matrix perturbation view of the small world phenomenon. *Siam Journal on Matrix Analysis and Applications*, 25(2):429–444, 2003.

[15] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

[16] M. Juvan and B. Mohar. Laplace eigenvalues and bandwidth-type invariants of graphs. *Journal of Graph Theory*, 17(3):393 – 407, 1993.

[17] M.E.J. Newman, D.J. Watts, and S.H. Strogatz. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84:3201–3204, 2000.

[18] W. Ren, R. W. Beard, and T. W. McLain. Coordination variables and consensus building in multiple vehicle systems. *"Proceedings of the Block Island Workshop on Cooperative Control, Lecture Notes in Control and Information Sciences series,Springer-Verlag"*, 309:171–188, 2004.

[19] R. Olfati Saber. Ultrafast consensus in small-world networks. *Proceedings of the American Control Conference*, 4:2371–2378, 2005.

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

[21] E. Seneta. *Nonnegative Matrices and Markov Chains*. Springer, 2nd edition, 1981.

[22] G.W. Stewart and J. Sun. *Matrix permutation theory*. Academic press, 1990.

[23] J. Sun, S. Boyd, L. Xiao, , and P. Diaconis. The fastest mixing markov process on a graph and a connection to a maximum variance unfolding problem. *"SIAM Review"*, 48(4):681–699, 2006.

[24] A. Tahbaz-Salehi and A. Jadbabaie. Small world phenomenon, rapidly changing markov chains, and average consensus algorithm. *Proceedings of 46th IEEE Conference on Decision and Control , accepted*, 2007.

[25] T. Vicsek, A. Czirok, E. Ben Jakob, I. cohen, and O.Schochet. Novel type of phase transitions in a system of self-driven particles. *Phys. Rev. Lett.*, 75:1226–1299, 1995.

[26] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[27] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Proceedings of 42th IEEE Conference on Decision and Control*, pages 4997–5002, 2003.

[28] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. *Proceedings of International Conference on Information Processing in Sensor Networks*, pages 63–70, 2005.