Small World Phenomenon, Rapidly Mixing Markov Chains, and Average Consensus Algorithms

Alireza Tahbaz-Salehi and Ali Jadbabaie

***Abstract*— In this paper, we demonstrate the relationship between the diameter of a graph and the mixing time of a symmetric Markov chain defined on it. We use this relationship to show that graphs with the small world property have dramatically small mixing times. Based on this result, we conclude that addition of independent random edges with arbitrarily small probabilities to a cycle significantly increases the convergence speed of average consensus algorithms, meaning that small world networks reach consensus orders of magnitude faster than a cycle. Furthermore, this dramatic increase happens for any positive probability of random edges. The same argument is used to draw a similar conclusion for the case of addition of a random matching to the cycle.**

I. Introduction

In their famous 1998 paper, Watts and Strogatz [1] introduced a simple model which captured a very important phenomenon observed in many real-world networks: the fact that any pair of vertices can be connected with a short path over the network. This property, often known as the *small world phenomenon*, is a feature of many large-scale social, biological and technological networks such as co-authorship network of scientists [2], power networks in the western United States [1], the structure of the World Wide Web [3], the network of actors who played in the same movies [4], [5] and neural networks [6]. The reader can refer to [7] for a survey on complex networks with the small world property.

Watts and Strogatz's paper resulted in a huge amount of interest among researchers. In order to capture different aspects of complex networks various models were proposed and studied in detail. For example, a simple model closely related to Watts-Strogatz model was proposed by Newman *et al.* [8] followed by its generalization by Kleinberg [9], [10]. Different properties of these models have been the subject of research over the past few years as well. Properties such as degree distribution, average path length, diameter, clustering effects and search for short paths are among properties investigated.

In the control community, these models became subjects of interest following a paper by Olfati-Saber [11], in which the author conjectures (based on some numerical simulations) that a continuous-time consensus algorithm over the small world model of Watts and Strogatz exhibits faster convergence (on average). He also conjectures that a phase transition takes place in the expected speed of convergence when the edge-rewiring probability is increased. Using similar simulations, the authors of [12] make the same conclusions for discrete-time consensus algorithms.

In this paper we show that the speed of convergence of a consensus algorithm (either discrete or continuous-time) over a cycle graph is directly affected by its diameter, and as a result, small world models have considerable faster convergence as predicted by the simulations. In order to show this, we use the model proposed by Newman *et al.* [8] which is closely related to Watts-Strogatz model. Our results are presented in the context of mixing rate of Markov chains rather than convergence of consensus algorithms. The obvious connection between the two guarantees that all the conclusions drawn for the long-run behavior of Markov chains hold for the convergence of average consensus algorithms as well.

On one hand, we show that the fastest mixing Markov chain possible on a cycle with n vertices mixes in order of $\Omega(n^2)$ time steps. On the other hand, by using simple arguments from theory of random graphs, we demonstrate that addition of an Erdős-Rényi random graph $G(n, \epsilon/n)$ to the cycle decreases the mixing time of a specific symmetric Markov chain on that graph to $\mathcal{O}(n \log^3 n)$, a considerably smaller value. In other words, for large enough n, the addition of random edges with probability ϵ/n results in "ultrafast" average consensus algorithms, no matter how small ϵ is. This also implies that if n is large enough, this dramatic change in mixing time occurs even when the randomly added edges have very small probabilities. Moreover, using similar arguments and a result by Bollabás and Chung [13], we show that Markov chains on a cycle with a random *matching* mix in time at most $\mathcal{O}(n \log^2 n)$ asymptotically almost surely.

Over the course of writing this paper, we realized that one can achieve much tighter bounds on the mixing rate of Markov chains on Newman's model, using Cheeger's inequality instead of inequalities based on the diameter. This was shown recently by Durrett in [14]. In this paper we present his results and use them to improve our bounds.

The paper is organized as follows: Some background on Markov chains and their mixing properties is presented in section II. Different graph models with the small world property are described in section III. Section IV contains our main results on the mixing time of Markov chains on Newman's graph and its comparison with the fastest mixing Markov chain on the cycle. The mixing rates of Markov chains on cycles with random matchings are investigated in section V, followed by application of Durrett's results to

This research is supported in parts by the following grants: ARO/MURI W911NF-05-1-0381, ONR/YIP N00014-04-1-0467 and NSF-ECS-0347285.

Alireza Tahbaz-Salehi and Ali Jadbabaie are with GRASP Laboratory and Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6228, USA. {atahbaz,jadbabai}@seas.upenn.edu

our problem in section VI. Finally, section VII contains our conclusions.

II. Markov Chains on Undirected Graphs

Consider a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, 2, \cdots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ such that $(i, j) \in \mathcal{E}$ whenever $(j, i) \in \mathcal{E}$. Assume that $(i, i) \in \mathcal{E}$ for all $i \in \mathcal{V}$, i.e. all vertices have self-loops. The unweighted adjacency and degree matrices of the graph are denoted by A and D respectively. To any edge (i, j) of the graph, we assign a nonnegative weight w_{ij}. These weights are assigned such that $w_{ij} = w_{ji}$ and $\sum_{j \in \mathcal{N}_i} w_{ij} = 1$, where $\mathcal{N}_i$ is the set of neighbors of vertex i. As a result, the weight matrix $W = [w_{ij}]$ satisfies

$$W \geq 0 \quad , \quad W\mathbf{1} = \mathbf{1} \quad , \quad W = W^T$$

where $\geq$ means element-wise nonnegativity and $\mathbf{1}$ is the vector of all ones.

One can define a discrete-time Markov chain on the vertices of $\mathcal{G}$ as follows. The state of the chain at time $k \in \{0, 1, 2, \cdots\}$, denoted by $X(k)$, is a vertex in $\mathcal{V}$ and the weight associated to each edge in the graph is the probability with which X makes a transition between two adjacent vertices. In other words, the transition from state i to state j happens with probability w_{ij}, the weight of edge (i, j). If $\pi(k)$ with elements defined as $\pi_i(k) = \mathbb{P}(X(k) = i)$ is the probability distribution of the state at time k, the state distribution satisfies the recursion $\pi(k+1)^T = \pi(k)^T W$. Since W is symmetric and $W\mathbf{1} = \mathbf{1}$, we conclude that the uniform distribution $(1/n)\mathbf{1}$ is a stationary distribution for the Markov chain. If the chain is irreducible and aperiodic (which is the case here because of self-loops and connectivity of the graph), then for all initial distributions, $\pi(k)$ converges to the unique stationary distribution $(1/n)\mathbf{1}$ as k increases.

The main object of study in this paper is speed of convergence of $\pi(k)$ to the uniform distribution, which depends on the eigenstracture of the probability transition (weight) matrix W. Since W is a symmetric and *primitive* stochastic matrix, all its eigenvalues are real and by Perron-Frobenius theorem [15] they satisfy

$$1 = \lambda_1(W) > \lambda_2(W) \geq \cdots \geq \lambda_n(W) > -1$$

It is well-known that the rate of convergence to the stationary distribution is governed by the second largest eigenvalue modulus of W defined as $\mu(W) = \max_{i=2,\cdots,n}\{|\lambda_i(W)|\} = \max\{\lambda_2(W), |\lambda_n(W)|\}$. To make this statement more precise, let i be the initial state and define the *total variation distance* between the distribution at time k and the stationary distribution as

$$\Delta_i(k) = \frac{1}{2}\sum_{j \in \mathcal{V}} |W_{ij}^k - \frac{1}{n}|.$$

The rate of convergence to the stationary distribution is measured using the following quantity known as the *mixing time*[1].

$$T_{mix} = \max_i \min\{k : \Delta_i(k') < \frac{1}{e} \text{ for all } k' \geq k\}$$

The following theorem indicates the relationship between the mixing time of a Markov chain and the second largest eigenvalue modulus of its probability transition matrix [16].

Theorem 1: The mixing time of a Markov chain with symmetric transition probability matrix W and second largest eigenvalue modulus μ satisfies

$$\frac{\mu}{2(1-\mu)}(1 - \ln 2) \leq T_{mix} \leq \frac{1 + \log n}{1 - \mu}.$$

Therefore, the speed of convergence of the Markov chain to its stationary distribution is determined by the value of $1-\mu$ known as the *spectral gap*; the larger the spectral gap, the faster the convergence.

In a close relationship with symmetric Markov chains on a graph lies the discrete-time version of the distributed *average consensus algorithm*, which guarantees the asymptotic achievement of a consensus over a network of agents. In the setup of consensus algorithms, a state $x_i(0)$ is assigned to vertex i at time $k = 0$ and the following distributed update is performed:

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

where $W = [w_{ij}]$ is the matrix of edge weights of the graph. By using an argument similar to the Markov chain case, one can show that over a connected graph, (1) converges to $\frac{1}{n}\mathbf{1}^T x(0)$ and the speed of convergence is determined by the spectral gap of the weight matrix [17], [18]. More precisely, the mixing time of a Markov chain with transition probability matrix W quantifies the convergence time of a discrete-time consensus algorithm with weight matrix W to the average. Thus, in this paper we only state our results in terms of the terminology of Markov chain theory.

III. Small World Phenomenon

Small world phenomenon is a feature of many complex networks where arbitrary vertices are connected by short paths. In other words, graphs with the small world property have a relatively small diameter compared to their number of vertices[2]. This phenomenon is observed in many social, biological and physical complex networks including power networks in the western United States, the structure of the World Wide Web, a co-authorship network of scientists, and the network of actors who played in the same movies.

There are various graph models which are known to have this property. In [19] the authors show that the diameter of Erdős-Rényi random graph $G(n, p)$ (a graph on n vertices and independent edges with probability p) scales as $\log n$ if $np = c > 1$, where c is a constant not depending on

[1]Note that in this definition the choice of $1/e$ is arbitrary and any other small number can be used to define the mixing time.

[2]Some authors define the small world phenomenon based on the behavior of the average path length between any two vertices of the graph instead of the diameter [1], [8], [11]. In this paper, unless otherwise noted, we use the diameter to define this phenomenon.

n.[3] This means that for a large enough random graph, its diameter is considerably smaller than the number of vertices. On the other hand, Erdős-Rényi random graphs do not exhibit another characteristic of complex networks, known as the *clustering* effect, which requires that two neighbor-shared vertices be neighbors with high probability.

In 1998 Watts and Strogatz proposed a simple model which serves as a tradeoff between low dimensional lattices and random graphs and displays both the small-diameter and clustering properties [1]. The Watts-Strogatz model consists of an initial cycle of n nodes, denoted by $\mathcal{C}_{n,k}$, in which each node is connected to its nearest neighbors up to $k \ll n$ hops away. Then every edge is rewired with probability p by randomly changing one of its end-points. In fact, one can assume that by varying p from 0 to 1, the resulting graph interpolates between a $(2k+1)$-regular cycle and a random graph. In their ground breaking paper, Watts and Strogatz [1] numerically show that for small values of p (for example $0.001 < p < 0.01$) the average path length of the graph drops significantly, while the local neighborhoods remain almost unchanged and as a result, the graph shows a high level of clustering. This model has been the subject of a recent study in [11] and [20], where based on numerical simulations it is conjectured that the consensus algorithm (or equivalently the Markov chain) defined on the Watts-Strogatz model has a faster expected convergence rate, compared to the consensus problem defined on the $(2k+1)$-regular cycle.

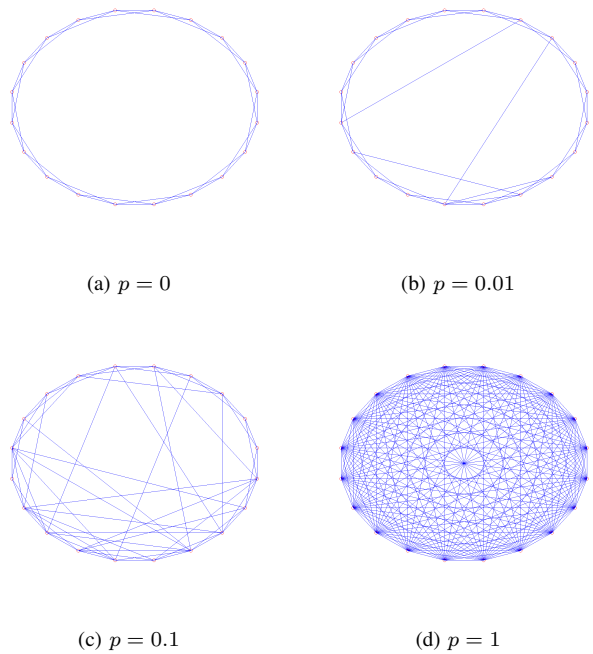
A closely related model was introduced in [8] by Newman *et al.* which consists of addition of a random graph to the $(2k+1)$-regular cycle. In their model, instead of rewiring edges, independent random edges are added to the graph with probability p. In other words, one can consider Newman's model as the union of cycle $\mathcal{C}_{n,k}$ with an Erdős-Rényi random graph $G(n,p)$. As p varies from 0 to 1, the resulting graph interpolates between a $(2k+1)$-regular cycle and the complete graph. Fig. 1 depicts graphs obtained from Newman's model for different values of p. Because of the way it is constructed, this model has properties similar to the Watts-Strogatz model for small values of p, including the small world property and high clustering. In fact, by approximating the cycle with a continuum of nodes, the authors in [8] show that Newman's model exhibits a small average path length after adding very few random edges (i.e. for small values of p).

One can achieve the same conclusion by using the following powerful theorem of Flaxman and Frieze, without approximating $\mathcal{C}_{n,k}$ with a continuum of vertices:

Theorem 2: Let ϵ be any positive constant. For any connected graph $\mathcal{G}$ with n vertices and bounded maximum degree, let $\tilde{\mathcal{G}} = \mathcal{G} \cup G(n, \epsilon/n)$. Then the diameter of $\tilde{\mathcal{G}}$ is at most $c_\epsilon \log n$ with probability converging to one as n increases, where c_ϵ is a constant that depends on ϵ but not on n.

Proof: The proof can be found in [21]. ■

[3]For a disconnected graph, the diameter is defined to be the largest diameter of all connected components of the graph. In this paper, all the graphs that we deal with are connected.



(a) $p = 0$ (b) $p = 0.01$

(c) $p = 0.1$ (d) $p = 1$

Fig. 1. The small world model of Newman *et al.* for different edge addition probabilities with parameters $n = 20$ and $k = 2$.

The above theorem indicates that due to connectivity of the underlying cycle $\mathcal{C}_{n,k}$ in Newman's model, adding a random $G(n, \epsilon/n)$ results in a graph with diameter scaling at most as fast as $\mathcal{O}(\log n)$ asymptotically almost surely, a significantly smaller value compared to the diameter of the cycle which is $\lfloor n/k \rfloor$.

The Newman's model has the advantage of being much easier to analyze compared to the rewiring model of Watts and Strogatz. Therefore, we use it as our model of a small world network. Moreover, from now on, we will use the logarithmic scaling of the diameter as the definition of the small world phenomenon. More formally,

Definition 1: The graph $\mathcal{G}$ has the small world property if $D(\mathcal{G}) = \mathcal{O}(\log n)$, where $D(\mathcal{G})$ is the diameter of $\mathcal{G}$ and n is the size of its vertex set.

At the end of this section we would like to emphasize that models that exhibit small world properties are not limited to the ones above. For example, another widely used model, was first introduced by Kleinberg [9]. This model, which exhibits both the small world and clustering properties, has a two dimensional grid as its non-random underlying structure to which random edges are added with a non-uniform probability distribution [9], [10].For a complete review of small world models the reader can refer to [7].

IV. Rapidly Mixing Markov Chains on Small World Graph

In this section we present our main results, where we show that a random walk on Newman's small world graph $\mathcal{C}_{n,1} \cup G(n,p)$ has orders of magnitude faster mixing rate than the

fastest mixing Markov chain on the regular cycle $\mathcal{C}_{n,1}$.[4] Our results are based on theorem 2 and a well-known lower bound on the diameter of a graph in terms of its algebraic connectivity or second smallest Laplacian eigenvalue. The results of this section can be used to explain some of the observations made in [11] and [20]. In order to achieve this, first we need to address the problem of finding the fastest mixing Markov chain (FMMC) on a graph. For a given graph $\mathcal{G}$, FMMC is the problem of finding the weight matrix W that minimizes the mixing time of the Markov chain defined on $\mathcal{G}$ with transition probability matrix W (or equivalently, maximizes the spectral gap of W). This was first formulated as the following optimization problem by Boyd *et al.* in [22]:

$$\begin{array}{ll} \text{Maximize} & 1-\mu(W) \\ \text{Subject to} & W \geq 0, \quad W\mathbf{1}=\mathbf{1}, \quad W=W^T \\ & W_{ij}=0, \quad (i,j)\notin\mathcal{E} \end{array} \tag{2}$$

In [22], the authors show that (2) is a convex optimization problem and can be solved using standard numerical semi-definite programming techniques. In the particular case of a cycle, because of symmetry of $\mathcal{C}_n$, we can solve FMMC analytically. The following lemma provides the asymptotic behavior of the optimal spectral gap for the FMMC problem on the cycle $\mathcal{C}_n$.

Lemma 1: The spectral gap of the fastest mixing Markov chain on the n-vertex cycle $\mathcal{C}_n$ satisfies $1-\mu^*=\Theta(1/n^2)$.

Proof: Because of the structure of $\mathcal{C}_n$, for all $1\leq i\leq n$ the i-th row of the weight matrix W has only three non-zero elements, corresponding to vertex i and its two neighbors. Due to symmetry, the optimal weight assignments must satisfy

$$\begin{aligned} w_{i,i+1} &= x \qquad \forall i \\ w_{ii} &= 1-2x \qquad \forall i \end{aligned}$$

where the indices are modulo n. Hence, the FMMC problem (2) can be reformulated as

$$\begin{array}{ll} \text{Maximize} & 1-\mu(W) \\ \text{Subject to} & W=(1-3x)I+xA \\ & 0\leq x\leq \frac{1}{2} \end{array}$$

which is a one-dimensional convex optimization problem. With some algebra one can easily show that for even n, the second largest eigenvalue modulus as a function of x is given by

$$\mu(x)=\max\left\{1-4x\sin^2\left(\frac{\pi}{n}\right), 4x-1\right\}$$

and therefore, the optimal spectral gap is equal to

$$1-\mu^*=\frac{2\sin^2(\pi/n)}{1+\sin^2(\pi/n)}=\Theta(1/n^2)$$

finishing the proof. The case of odd n is similar. ■

Having solved the FMMC problem for $\mathcal{C}_n$, our next step is to find an upper bound on the mixing time of the Markov chains on Newman's small world graph. The next two lemmas are required in order to compute such a bound. The first one provides a relationship between the second smallest Laplacian eigenvalue of a graph, known as the *algebraic connectivity*, and its diameter, while the second lemma states that the maximum degree of a sparse Erdős-Rényi random graph is at most $\log n$ with probability tending to 1 as n grows.

Lemma 2: For a graph $\mathcal{G}$ with n vertices, its second smallest Laplacian eigenvalue, $\lambda_{n-1}(L)$, imposes a lower bound on the diameter of the graph

$$D(\mathcal{G})\geq\frac{4}{n\lambda_{n-1}(L)}. \tag{3}$$

Proof: See [23]. ■

Lemma 3: In the Erdős-Rényi random graph $G(n,\epsilon/n)$, where $\epsilon>0$ is a constant not depending on n, the maximum degree almost surely satisfies $d_{\max}\leq\log n$ as $n\to\infty$.

Proof: For a given vertex i, define the random variable $Y_j=\mathbb{I}_{\{(i,j)\in\mathcal{E}(\mathcal{G})\}}$, where $\mathbb{I}_B$ is the indicator of event B. This implies that the degree of vertex i, which is equal to $d_i=Y_1+\cdots+Y_n$, has a binomial distribution. Using a Chernoff bound, we have

$$\begin{aligned} \mathbb{P}(d_i>\log n) &= \mathbb{E}\mathbb{I}_{\{d_i>\log n\}} \\ &\leq e^{-2\log n}\mathbb{E}\left(e^{2d_i}\right) \\ &= n^{-2}\left(\mathbb{E}e^{2X_1}\right)^n \\ &\approx \exp(\epsilon(e^2-1))n^{-2}. \end{aligned}$$

As a result,

$$\mathbb{P}(d_{\max}>\log n)\leq n\mathbb{P}(d_i>\log n)=\delta/n$$

for some $\delta>0$. The right hand side approaches zero as n approaches infinity and the desired result follows. ■

We now present our main result.

Theorem 3: Let $\epsilon>0$ be a constant not depending on n. Denote the mixing time of the Markov chain defined on $\tilde{\mathcal{G}}=\mathcal{C}_n\cup G(n,\epsilon/n)$ with transition probability matrix

$$W=I+\frac{1}{2\log n}(A-D) \tag{4}$$

by $T_{mix}(\tilde{\mathcal{G}})$ and the mixing time of the fastest mixing Markov chain on $\mathcal{C}_n$ with $T^*_{mix}(\mathcal{C}_n)$. Then,

$$\begin{aligned} T_{mix}(\tilde{\mathcal{G}}) &= \mathcal{O}(n\log^3 n) \\ T^*_{mix}(\mathcal{C}_n) &= \Omega(n^2). \end{aligned}$$

Proof: The second equality can be proved using lemma 1 and the lower bound on the mixing time in theorem 1. To prove the first equality, note that the eigenvalues of W are related to the eigenvalues of the Laplacian matrix $L=D-A$:

$$\lambda_k(W)=1-\frac{1}{2\log n}\lambda_{n-k+1}(L).$$

Therefore, based on lemma 3 and the fact that all Laplacian eigenvalues are less than or equal to twice the maximum degree of the graph [24], one can conclude that all the eigenvalues of W are nonnegative asymptotically almost surely and as a result $\mu(W)=\lambda_2(W)$. Hence, (3) implies

$$1-\mu(W)=\frac{1}{2\log n}\lambda_{n-1}(L)\geq\frac{2}{(n\log n)D(\tilde{\mathcal{G}})} \quad \text{a.s.}$$

[4] In the rest of the paper we only address the case of $k=1$ and for simplicity we denote $\mathcal{C}_{n,1}$ by $\mathcal{C}_n$. Similar results hold for cycles with $k\ll n$.

for large enough n. At this point, theorem 2 guarantees the small world property of Newman's graph, i.e. $D(\tilde{\mathcal{G}}) = \mathcal{O}(\log n)$. Therefore, the upper bound in theorem 1 implies that the Markov chain on $\tilde{\mathcal{G}}$ with transition probability matrix W mixes in $\mathcal{O}(n \log^3 n)$. ■

Theorem 3 states that the mixing time of the Markov chain with transition probability matrix as in (4) on Newman's graph $\tilde{\mathcal{G}}$ is orders of magnitude smaller than the mixing time of the fastest mixing Markov chain on the cycle, no matter how small ϵ is. Equivalently, in order to achieve an "ultrafast" convergence to a consensus on a cycle of nodes, one can augment the graph with random edges with arbitrarily small probabilities. When n is large, the convergence rate achieved by this is arbitrarily better than using the best possible weight matrix W on $\mathcal{C}_n$.

A. Continuous-Time Consensus Algorithms

The above arguments can be extended to the continuous-time variant of the consensus algorithms, or a continuous-time, finite Markov chain. It is well-known that for a connected graph with Laplacian matrix $L = D - A$, the linear dynamical system

$$\dot{x}(t) = -Lx(t) \tag{5}$$

converges and the limit is equal to $\frac{1}{n}\mathbf{1}^T x(0)$, i.e. each entry of the state vector converges to the average of the initial condition [25]. The speed of convergence of (5) to the consensus is dominated by $\lambda_{n-1}(L)$, the algebraic connectivity of the graph [26]:

$$\sup_{x(0)} \|x(t) - \frac{1}{n}\mathbf{1}\|_{tv} \leq \frac{1}{2}\sqrt{n}e^{-t\lambda_{n-1}}.$$

In other words, the larger the algebraic connectivity, the faster the convergence. Therefore, (3) can be used to lower bound the second smallest Laplacian eigenvalue of Newman's random graph. More precisely, it implies that the ratio of algebraic connectivity of $\tilde{\mathcal{G}}$ and $\mathcal{C}_n$ satisfies

$$\frac{\lambda_{n-1}(L(\tilde{\mathcal{G}}))}{\lambda_{n-1}(L(\mathcal{C}_n))} \geq \frac{(n \log n)^{-1}}{c_\epsilon \sin^2(\pi/n)} = \Omega\left(\frac{n}{\log n}\right)$$

with probability tending to 1 as $n \to \infty$. As a result, adding random edges with probability ϵ/n increases the speed of convergence to consensus significantly. This provides an explanation for the increase observed for the expected algebraic connectivity ratio in [11]'s simulations. Also note that this happens regardless of the value of ϵ for large enough n.

V. Addition of a Random Matching

An alternative way to generate a small world graph on the cycle is the addition of a random *matching* to it. A matching on $\mathcal{C}_n$ is defined as a partition of the vertex set $\mathcal{V} = \{1, 2, \ldots, n\}$ into disjoint pairs (plus a singleton if n is odd). More precisely, a matching is the addition of $\lfloor n/2 \rfloor$ edges to the cycle such that the degree of each vertex is at most 3. The set of all such graphs is our probability space and we assume all the graphs in the probability space are equiprobable.

An interesting result due to Bollabás and Chung [13] states that the diameter of a cycle plus a random matching, with probability tending to 1 satisfies

$$\log_2 n - c \leq D(\mathcal{G}) \leq \log_2 n + \log_2 \log n + c$$

where c is a small constant (at most 10). Therefore, similar to section IV, by applying (3) one can lower bound the algebraic connectivity of such a graph. As a result, theorem 1 implies that the symmetric Markov chain with probability transition matrix $W = I + 1/6(A - D)$ on an n-cycle with a random matching mixes in time at most $\mathcal{O}(n \log^2 n)$, a significantly smaller time comparing to mixing time of the fastest mixing Markov chain on $\mathcal{C}_n$. Note that since in this case the maximum degree is 3, all the eigenvalues of W are non-negative and therefore there is no need to scale the transition probabilities with $\log n$.

This can be generalized to the cycles $\mathcal{C}_{n,k}$ for $k = \mathcal{O}(\log n)$ in a straightforward manner: start with $\mathcal{C}_{n,k}$, remove all the edges except the ones connecting a vertex to its two nearest neighbors (i.e. transforming it to $\mathcal{C}_n$) and substitute the $2(k-1)$ removed edges with one single random edge. As a result, the convergence speed of average consensus algorithms (either discrete or continuous-time) increases dramatically, although we have significantly reduced the total number of edges of the graph.

VI. Bounds Based on Cheeger's Inequality

By using an eigenvalue inequality tighter than (3), one can improve the results of sections IV and V and provide better estimates of the mixing rates on small world graphs. This was recently done by Durrett in [14]. By using Cheeger's inequality, the author provides tight bounds on the spectral gap of the probability transition matrix of a "lazy" natural random walk[5] on Newman's graph. These results can be used to bound the spectral gap of symmetric Markov chains as well.

Given a Markov chain with stationary distribution π define its *conductance* as

$$h = \min_{\substack{S \subset \mathcal{V} \\ \pi_S \leq 1/2}} \frac{Q(S, S^c)}{\pi_S},$$

where $Q(A, B) = \sum_{i \in A, j \in B} \pi_i W_{ij}$ and $\pi_A = \sum_{i \in A} \pi_i$ for $A, B \subseteq \mathcal{V}$. The quotient in the above definition is merely the conditional probability that the chain in equilibrium escapes from the subset S of the state space in one step, given that it is initially in S. Thus, h measures the ability of the chain to escape from any small region of the state space, and hence make a rapid progress to the equilibrium [16]. In the case that the transition probability matrix W is symmetric and all the edge weights are equal to w (except possibly for the self-loops), the expression for the conductance can be simplified

$$h = w \min_{\substack{S \subset \mathcal{V} \\ |S| \leq n/2}} \frac{e(S, S^c)}{|S|},$$

[5] A lazy natural random walk is a Markov chain which stays in its current state with probability $1/2$ and picks a neighbor uniformly at random with probability $1/2$.

where $e(S, S^c)$ is the number of edges connecting S to S^c and $|S|$ is number of vertices in S.

The following lemma, known as Cheeger's inequality, bounds the spectral gap in terms of conductance [16] and is the key to computing the mixing time.

Lemma 4 (Cheeger's inequality): The second largest eigenvalue of a symmetric and primitive stochastic matrix W satisfies

$$\frac{h^2}{2} \leq 1 - \lambda_2(W) \leq 2h, \tag{6}$$

where h is the conductance of a Markov chain with transition probability matrix W.

Note that when all the eigenvalues of W are non-negative, one can replace λ_2 in (6) with μ, the second largest eigenvalue modulus. Clearly, this holds if we pick W as in (4).

It is shown in [14] that for all $S \subset \mathcal{V}$ the inequality

$$\frac{e(S, S^c)}{vol(S)} \geq c/\log n,$$

holds for Newman's small world graph $\tilde{\mathcal{G}} = \mathcal{C}_n \cup G(n, \epsilon/n)$ with probability tending to 1 as $n \to \infty$. We can now use this result to prove the next theorem.

Theorem 4: The Markov chain on $\tilde{\mathcal{G}}$ with transition probability matrix as in (4) mixes in time at most $\mathcal{O}(\log^5 n)$.

Proof: Since the minimum degree of the graph is at least 2, we have $vol(S) \geq 2|S|$. Thus, $e(S, S^c) \geq 2c|S|/\log n$ holds for any $S \subset \mathcal{V}$ if n is chosen large enough, which implies that $h \geq c/\log^2 n$ asymptotically almost surely. Combining this with theorem 1 and (6) finishes the proof. ■

A similar method can be used to determine tighter bounds on mixing time of a symmetric lazy random walk on a cycle with a random matching. In fact, in [14], Durrett shows that the conductance of such a Markov chain is bounded away from zero, and therefore by lemma 4, it mixes in time at most $\mathcal{O}(\log n)$ almost surely.

VII. Conclusions

Using the lower bound on algebraic connectivity of a graph in terms of its diameter, we demonstrated that a symmetric Markov chain on Newman's small world graph mixes in time at most $\mathcal{O}(n \log^3 n)$ with probability tending to 1, a significantly smaller value in comparison to the fastest mixing symmetric Markov chain on the regular cycle which mixes in time $\Omega(n^2)$. We also demonstrated that mixing time of a symmetric Markov chain on a cycle with a random matching is $\mathcal{O}(n \log^2 n)$. From these we showed that the convergence speed of average consensus algorithms (both discrete and continuous-time) can be increased dramatically by adding random edges or a random matching to the cycle. Finally, using a recent result by Durrett, we showed that with a bit of more work one can achieve even better bounds using conductance and Cheeger's inequality. Specifically, it is shown that the mixing times are reduced to $\mathcal{O}(\log^5 n)$ for Newman's model and $\mathcal{O}(\log n)$ for cycles with a random matching.

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[2] M. E. J. Newman, "Scientific collaboration networks: I. Network construction and fundamental results," *Physical Review E*, vol. 64, no. 1, p. 016131, 2001.

[3] A. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A: Statistical Mechanics and its Applications*, vol. 281, pp. 69–77, 2000.

[4] A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small world networks," *Proceedings of National Academy of Science*, vol. 97, pp. 11 149–11 152, 2000.

[5] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, pp. 268–276, 2001.

[6] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, "The structure of the nervous system of the nematode Caenorhabditis elegans," *Philosophical Transactions of the Royal Society of London*, vol. 314, no. 1165, pp. 1–340, 1986.

[7] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, p. 167, 2003.

[8] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-field solution of the small-world network model," *Physical Review Letters*, vol. 84, no. 14, pp. 3201–3204, Apr 2000.

[9] J. Kleinberg, "The small world phenomenon: An algorithmic approach," in *Proceedings of 32nd annual ACM Symposium on Theory of Computing*. New York, NY: ACM Press, 2000, pp. 163–170.

[10] ——, "Small world phenomena and dynamics of information," in *Advances in Neural Information Processing Systems (NIPS) 14*, 2001.

[11] R. Olfati-Saber, "Ultrafast consensus in small world networks," in *Proceedings of the 2005 American Control Conference*, june 2005, pp. 2371–2378.

[12] P. Hovareshti and J. S. Baras, "Consensus problems on small world graphs: A structural study," Institute for Systems Research, College Park, MD, Tech. Rep., 2006.

[13] B. Bollabás and F. R. K. Chung, "The diameter of a cycle plus a random matching," *SIAM Journal of Discrete Mathematics*, vol. 1, no. 3, pp. 328–333, 1988.

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

[15] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic Press, 1979.

[16] A. J. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatorics, Probability and Computing*, vol. 1, pp. 351–370, 1992.

[17] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, Dec. 2005, pp. 2996–3000.

[18] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, 2005.

[19] F. Chung and L. Lu, "The average distance in a random graph with given expected degrees," *Proceedings of the National Academy of Sciences*, vol. 99, no. 25, pp. 15 879–15 882, 2002.

[20] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, 2007.

[21] A. D. Flaxman and A. M. Frieze, "The diameter of randomly perturbed digraphs and some applications," *Random Structures & Algorithms*, vol. 30, no. 4, pp. 484–504, 2007.

[22] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Review, problems and techniques section*, vol. 46, no. 4, pp. 667–689, Dec 2004.

[23] B. Mohar, "Eigenvalues, diameter and mean distance in graphs," *Graph and Combinatorics*, vol. 7, no. 1, pp. 53–64, 1991.

[24] ——, "The Laplacian spectrum of graphs," *Graph Theory, Combinatorics, and Applications*, vol. 2, pp. 871–898, 1991.

[25] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[26] 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*, vol. 48, no. 4, pp. 681–699, 2006.