Tracking a Markov-Modulated Stationary Degree Distribution of a Dynamic Random Graph

Maziyar Hamdi, Vikram Krishnamurthy, *Fellow, IEEE*, and George Yin, *Fellow, IEEE*

***Abstract*— This paper considers a Markov-modulated duplication-deletion random graph where at each time instant, one node can either join or leave the network; the probabilities of joining or leaving evolve according to the realization of a finite state Markov chain. Two results are presented. First, motivated by social network applications, the asymptotic behavior of the degree distribution is analyzed. Second, a stochastic approximation algorithm is presented to track empirical degree distribution as it evolves over time. The tracking performance of the algorithm is analyzed in terms of mean square error and a functional central limit theorem is presented for the asymptotic tracking error. Also, a Hilbert-space-valued stochastic approximation algorithm that tracks a Markov-modulated probability mass function with support on the set of nonnegative integers is analyzed.**

***Index Terms*— Adaptive algorithms, complex networks, degree distribution, Markov-modulated random graphs, power law, social networks, stochastic approximation algorithms.**

I. Introduction

STOCHASTIC approximation algorithms have several applications in diverse areas such as target tracking, change detection, communication systems, and economics [1]–[7]. The ubiquitous use of stochastic approximation algorithms is mainly due to their ability to track a time-varying unknown parameter of a system; this is called "tracking capability", see [5]. In this paper, motivated by social network applications, we consider a class of stochastic approximation algorithms to track a time-varying probability mass function that evolves according to a finite-state Markov chain whose transition matrix is close to identity. In the context of social network analysis, the time-varying probability mass function which we aim to track is the expected degree distribution of a dynamic random graph.

Dynamic random graphs have been widely used to model social networks, biological networks [8] and Internet graphs [9]. Such dynamic models can be viewed as a sequence of graphs where the random graph at each time may depend on all the earlier graphs (snapshots of the evolving graph at earlier times) [9]. Motivated by analyzing social networks, we introduce *Markov-modulated duplication-deletion random graphs*[1] where at each time instant, nodes can either be added to or eliminated from the graph with probabilities that change according to a finite-state Markov chain. Such graphs mimic social networks where the interactions between nodes evolve over time according to a Markov process that undergoes infrequent jumps. An example of such a social network is the friendship network among residents of a city, where the dynamics of the network change in the event of a large festival. A class of stochastic approximation algorithms are used to track the expected degree distribution of such Markov-modulated dynamic graphs.

A. Why Analyze the Degree Distribution?

The degree of a node in a network (also known as the connectivity) is the number of connections the node has in that network. The most important measure that characterizes the structure of a network (specially when the size of the network is large and the connections—adjacency matrix of the underlying graph—are not given) is the *degree distribution* of the network. The degree distribution can further be used to investigate the diffusion of information or disease through social networks [10]–[12]. The existence of a "giant component"[2] in complex networks can be studied using the degree distribution. The size and existence of a giant component has important implications in social networks in terms of modeling information propagation and spread of human disease [13]–[15].

B. Main Results and Paper Organization

Sec.II describes the construction of Markov-modulated duplication-deletion random graphs. Sec.III provides an asymptotic degree distribution analysis for the non-Markov modulated case of two different scenarios: (i) fixed size duplication-deletion random graph, and (ii) infinite

Manuscript received February 27, 2013; revised June 6, 2014; accepted July 12, 2014. Date of publication August 7, 2014; date of current version September 11, 2014. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and in part by the Social Sciences and Humanities Research Council of Canada. G. Yin was supported by the Air Force Office of Scientific Research, Arlington, VA, USA. This paper was presented at the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing.

M. Hamdi and V. Krishnamurthy are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: maziyarh@ece.ubc.ca; vikramk@ece.ubc.ca).

G. Yin is with the Department of Mathematics, Wayne State University, Detroit, MI 48202 USA (e-mail: gyin@math.wayne.edu).

Communicated by G. Moustakides, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2346183

[1]The duplication-deletion procedure for Markov-modulated random graphs is described in Sec.II.

[2]A giant component is a connected component of size $O(\eta)$, where η is the number of vertices in the graph. If the average degree of a random graph is strictly greater than one, then there exists a unique giant component with probability one [9], and the size of this component can be computed from the expected degree distribution.

duplication-deletion random graph. Theorem 3.1 in Sec.III-A asserts that the expected degree distribution of the fixed size Markov-modulated random graph at each time can be computed in terms of the expected degree distribution of the graph at the previous time and the dynamics of the graph via recursive equation (8). Sec.III-B extends the results of Sec.III-A to infinite random graphs. Theorem 3.2 parameterizes the degree distribution of such a graph by the *power law exponent* which depends on the dynamics of the graph.

Sec.IV considers the problem of adaptively estimating the degree distribution of a fixed size Markov-modulated duplication-deletion random graph given observations of the degree distribution. A stochastic approximation algorithm is presented for tracking the degree distribution as it evolves over time. In particular, Sec.IV presents three results regarding the tracking performance of the stochastic approximation algorithm:

- *Mean square error analysis:* Theorem 4.1 analyzes the asymptotic mean square error between the expected degree distribution and the estimate obtained via the stochastic approximation algorithm. Deriving this result uses error bounds on two-time scale Markov chains and perturbed Lyapunov function methods.
- *Weak convergence analysis:* Theorem 4.2 shows that the asymptotic behavior of the stochastic approximation algorithm converges weakly to the solution of a switched Markovian ordinary differential equation.
- *Functional central limit theorem for scaled tracking error:* Finally, Theorem 4.3 investigates the asymptotic behavior of the scaled tracking error. Similar to [16], it is shown that the interpolated scaled tracking error converges weakly to the solution of a switching diffusion process.

Sec.V extends the results of Sec.IV to infinite (denumerable) duplication-deletion random graphs where the number of nodes in the graph (and so the support of degree distribution) is no longer fixed and increases over time. A Hilbert-space-valued stochastic approximation algorithm is proposed to track the degree distribution of the infinite graph with support on the set of non-negative integers. To study the tracking performance of such a Hilbert-space-valued stochastic approximation algorithm, limit system characterization and asymptotic analysis of scaled tracking error are provided. Numerical examples are presented in Sec.VI.

C. Related Works

For a comprehensive development of stochastic approximation algorithms, see [1], [5], [17]. Tracking capability of regime switching stochastic approximation algorithms is further investigated in [18]. For the applications of stochastic approximation algorithms in interactive sensing and decision making see [19], [20]. Also, see [17], [21]–[23] for the performance analysis of adaptive algorithms in discrete-time. A detailed exposition of random graphs is provided in [24]. The dynamics of random graphs are investigated in the mathematics literature, for example, see [9], [25], [26] and the reference therein. In [27], a duplication model is proposed where at each time step a new node joins the network. However, the probabilities of joining in this model do not evolve over time. In [9], it is shown that the degree distribution of such networks satisfies a power law.

II. Markov-Modulated Dynamic Random Graph of Duplication-Deletion Type

This section outlines the construction of Markov-modulated dynamic random graphs of duplication-deletion type. Let $n = 0, 1, 2, \ldots$ denote discrete time. Denote by θ_n a discrete-time Markov chain with state space

$$\mathcal{M} = \{1, 2, \ldots, m\}, \tag{1}$$

and initial probability distribution π_0.

Assumption 2.1: The Markov chain θ_n evolves according to the transition matrix

$$A^\rho = I + \rho Q. \tag{2}$$

Here, I is an $m \times m$ identity matrix, ρ is a small positive real number, and $Q = [q_{ij}]$ is an irreducible[3] generator of a continues-time Markov chain satisfying

$$q_{ij} > 0 \ \text{ for } \ i \neq j, \ \text{ and } \ Q\mathbf{1} = \mathbf{0}, \tag{3}$$

where $\mathbf{1}$ and $\mathbf{0}$ represent column vectors of ones and zeros, respectively. The transition probability matrix A^ρ is therefore close to identity matrix. Here and henceforth, we refer to such a Markov chain θ_n as a "slow" Markov chain. The initial distribution π_0 is assumed independent of ρ.

A Markov-modulated duplication-deletion random graph is parameterized by the 7-tuple $(m, A^\rho, \pi_0, r, p, q, \mathcal{G}_0)$. Here, p and q are m-dimensional vectors with elements $p(i)$ and $q(i) \in [0, 1]$, $i = 1, \ldots, m$, where $p(i)$ denotes the connection probability, and $q(i)$ denotes the deletion probability. Also, $r \in [0, 1]$ denotes the probability of the duplication step, and $\mathcal{G}_0$ denotes the initial graph at time 0. In general, $\mathcal{G}_0$ can be any finite simple connected graph. For simplicity, assume that $\mathcal{G}_0$ is a simple connected graph with size η_0. The duplication-deletion random graph is constructed via the duplication-deletion Procedure 1.[4]

The Markov-modulated random graph generated by the duplication-deletion Procedure 1 mimics social networks where the interactions between nodes evolve over time due to the underlying dynamics (state of nature) such as seasonal variations (e.g., the high school friendship social network evolving over time with different winter/summer dynamics). In such cases, the connection/deletion probabilities p, q evolve with time. Procedure 1 models these time variations as a finite state Markov chain θ_n with transition matrix A^ρ.

[3]The irreducibility assumption implies that there exists a unique stationary distribution $\pi \in \mathbb{R}^{m\times 1}$ for this Markov chain such that $\pi' = \pi' A^\rho$.

[4]In Procedure 1, Step 1 is executed with probability r. Then, regardless of execution of Step 1, Step 2 is implemented. For convenience in the analysis, assume that a node generated in the duplication step cannot be eliminated in the deletion step immediately after its generation. Also, nodes whose degrees change in the edge-deletion part of Step 2, remain unchanged in the duplication part of Step 2 at that time instant. Finally, to prevent formation of isolated nodes, assume that the neighbor of a node with degree one cannot be eliminated in the deletion step. Note also that the duplication step in Step 2 ensures that the graph size does not decrease.

Procedure 1 Markov-Modulated Graph Parameterized by $(m, A^\rho, \pi_0, r, p, q, \mathcal{G}_0)$

At time n, given the graph $\mathcal{G}_n$ and Markov chain state θ_n, simulate the following events:

Step 1: Duplication step: With probability r implement the following steps:

- Choose node u from graph $\mathcal{G}_n$ randomly with uniform distribution.
- *Vertex-duplication*: Generate a new node v.
- *Edge-duplication*:
  - Connect node u to node v. (A new edge between u and v is added to the graph.)
  - Connect each neighbor of node u with probability $p(\theta_n)$ to node v. These connection events are statistically independent.

Step 2: Deletion Step: With probability $q(\theta_n)$ implement the following steps:

- *Edge-deletion*: Choose node w randomly from $\mathcal{G}_n$ with uniform distribution. Delete node w along with the connected edges in graph $\mathcal{G}_n$.
- *Duplication Step*: Choose a node from graph x from $\mathcal{G}_n$ randomly and implement *Vertex-duplication* and *Edge-duplication* processes as described in Step 1.

Step 3: Denote the resulting graph by $\mathcal{G}_{n+1}$. Generate θ_{n+1} (Markov chain) using transition matrix A^ρ. Set $n \to n+1$ and go to Step 1.

III. Asymptotic Degree Distribution Analysis for Non-Markov Modulated Case

This section presents degree distribution analysis for duplication-deletion random graphs generated according Procedure 1 for the non-Markov modulated case, i.e., $m = 1$. The stationary degree distribution obtained in Sec.III-A below will be used in the Markov modulated case. The results in this section constitute a minor extension of [9] to the duplication-deletion random graphs.

Notation: At each time n, let η_n denote the number of nodes of graph $\mathcal{G}_n$. Also, let F_n be a η_n dimensional vector such that its i-th element, F_n^i, denotes the number of vertices of graph $\mathcal{G}_n$ with degree i. Clearly $F_n'\mathbf{1} = \eta_n$ where $\mathbf{1}$ denotes the vector of ones. Here, $'$ is used to denote the transpose of a vector or matrix. Define the "empirical vertex degree distribution" as

$$G_n = (G_n^i, i = 1, 2, \ldots), \text{ where } G_n^i = \frac{F_n^i}{\eta_n}. \tag{4}$$

Note that G_n can be viewed as a probability mass function since all of its elements are non-negative and $G_n'\mathbf{1} = 1$.

A. Fixed Size Random Graph

This subsection analyzes the evolution of the expected degree distribution for a fixed size duplication-deletion random graph generated according to Procedure 1 with $r = 0$, $m = 1$. (Recall r denotes the probability of Step 1 in Procedure 1.) Therefore, the number of vertices in the graph remains fixed, i.e., $\eta_n = \eta_0$ for $n = 0, 1, 2, \ldots$. Theorem 3.1 below gives a recursion for the expected degree distribution of the fixed size Markov-modulated duplication-deletion random graph.

Theorem 3.1: Consider the fixed size duplication-deletion random graph generated according to Procedure 1, where $r = 0$, $m = 1$. Let $\overline{G}_n$ denote the expected degree distribution of nodes at time n. Then, $\overline{G}_n$ satisfies the recursion

$$\overline{G}_{n+1} = (I + \frac{1}{\eta_0}L')\overline{G}_n, \tag{5}$$

where L is a generator matrix[5] with elements (for $1 \le i, j \le \eta_0$):

$$l_{ji} = \begin{cases} 0, & j < i-1, \\ qp^{i-1} + q\big(1 + p(i-1)\big), & j = i-1, \\ iqp^{i-1}(1-p) - q\big(i + 2 + pi\big), & j = i, \\ q\binom{i+1}{i-1}p^{i-1}(1-p)^2 + q(i+1), & j = i+1, \\ q\binom{j}{i-1}p^{i-1}(1-p)^{j-i+1}, & j > i+1. \end{cases} \tag{6}$$

Proof: The proof is presented in Appendix A. ■

Theorem 3.1 shows that evolution of the expected degree distribution in a fixed size Markov-modulated duplication-deletion random graph satisfies (5). One can rewrite (5) as

$$\overline{G}_{n+1} = B_{\eta_0}'\overline{G}_n, \text{ where } B_{\eta_0} = I + \frac{1}{\eta_0}L. \tag{7}$$

Since L is a generator matrix, B_{η_0} can be considered as the transition matrix of a slow Markov chain. It is also straightforward to show that B_{η_0} is irreducible and aperiodic.[6] Hence, there exists a unique stationary distribution $\overline{G} = (\overline{G}^i, i = 1, 2, \ldots)$ such that

$$\overline{G} = B_{\eta_0}'\overline{G}. \tag{8}$$

The stationary distribution $\overline{G}$ is the stationary expected degree distribution of a fixed size duplication-deletion random graph generated according to Procedure 1 where $r = 0$.

B. Power Law Exponent for Infinite Duplication-Deletion Random Graph

The degree distribution analysis provided in the previous subsection was for a fixed size random graph generated according to the duplication-deletion Procedure 1 with $r = 0$. This section extends this analysis to infinite duplication-deletion random graphs (obtained by choosing $r = 1$). Assume that $\mathcal{G}_0$ is an empty set. Since $r = 1$, at time n, the graph $\mathcal{G}_n$ has n nodes. By employing the same approach as in the proof of Theorem 3.1, it will be shown that the infinite duplication-deletion random graph without Markovian dynamics satisfies a power law. An expression is further derived for the power law exponent.

Definition 3.1 (Power Law Distribution): The degree distribution $G = (G^i, i = 1, 2, \ldots)$, of a graph $\mathcal{G}$ has a power

[5] That is, each row adds to zero and each non-diagonal element of L is positive.

[6] It is straightforward to show that all elements of $(B_{\eta_0})^{\eta_0}$ are strictly greater than zero. Therefore, B_{η_0} is irreducible and aperiodic.

law distribution[7] if there exists an integer i^* such that for all $i \geq i^*$,

$$\log G^i = \alpha - \beta \log i$$

where α is a constant[8] and $\beta > 1$. Parameter β is called the *power law exponent*.

The power law is satisfied in many networks such as WWW-graphs, peer-to-peer networks, phone call graphs, co-authorship graph and various massive online social networks (e.g. Yahoo, MSN, Facebook) [29]–[33]. The following theorem states that the graph generated according to Procedure 1 with $r = 1$ and $m = 1$ satisfies a power law.

Theorem 3.2: Consider the infinite random graph with Markovian dynamics $\mathcal{G}_n$ obtained by Procedure 1 with 7-tuple $(1, 1, 1, 1, p, q, \mathcal{G}_0)$ with the expected degree distribution $\overline{G}_n$. Then, if $\log p + p < \frac{q}{1+q} < p$, the expected degree of nodes in $\mathcal{G}_n$ has a power law distribution with exponent $\beta > 1$. The power law exponent is computed from

$$(1 + q)(p^{\beta-1} + p\beta - p) = 1 + \beta q. \tag{9}$$

Here, p and q are the probabilities defined in duplication and deletion steps, respectively.

Proof: The proof is similar to that of Theorem 3.1 with some modifications, see [34]. Here, we only present an outline of the proof which is comprised of two steps: (i) finding the power law exponent, and (ii) showing that the degree distribution converges to a power law with the computed exponent as $n \to \infty$. To find the power law exponent, we derive a recursive equation for the number of nodes with degree $i + 1$ at time $n + 1$, denoted by F_{n+1}^{i+1}, in terms of the degrees of nodes in graph $\mathcal{G}_n$. Then, rearranging this recursive equation yields an equation for the power law exponent. To prove that the degree distribution satisfies a power law, we show that $\lim_{n\to\infty} \sum_{k=1}^{i} \mathbf{E}\{F_n^k\} = \sum_{k=1}^{i} ck^{-\beta}$, where $\beta > 1$ is the power law exponent computed in the first step and F_n^k is the k-th element of F_n. ■

Theorem 3.2 asserts that the infinite duplication-deletion random graph without Markovian dynamics generated by Procedure 1 satisfies a power law and provides an expression for the power law exponent. The significance of this theorem is that it ensures, with use of one single parameter (the power law exponent), we can describe the degree distribution of graphs with relatively large number of nodes. The above result slightly extends [8], [27], where only a duplication model was considered. Theorem 3.2 allows us to explore characteristics (such as searchability, diffusion, and existence/size of the giant component) of large networks which can be modeled with the infinite duplication-deletion random graphs.

[7] There is a difference between "power law" and "power law distribution". Power law is a functional relationship between two parameters where one parameter is proportional to the power of another, i.e., $x \propto y^{-\beta}$, where β can be any real number. In comparison, the exponent of a power law distribution is strictly greater that one [28]. Otherwise, the probability distribution does not add up to one.

[8] The normalization constant α is computed from $\alpha = -\log[\zeta(\beta, i^*)]$, where $\zeta(\beta, i^*) = \sum_{k=i^*}^{\infty} k^{-\beta}$ denotes the incomplete Riemann ζ-function.

IV. Estimating (Tracking) the Degree Distribution of the Fixed Size Markov-Modulated Duplication-Deletion Random Graph

In Sec.II, an expression was given for the unique stationary degree distribution $\overline{G}$ for the non-Markov modulated case, see (8). In this section, we consider fixed size Markov modulated duplication deletion random graphs. Consider Procedure 1 and assume that there are m possible stationary degree distributions, namely $\mathbf{G} = \{\overline{G}(1), \overline{G}(2), \ldots, \overline{G}(m)\}$ corresponding to the m states of a Markov chain. Here each $\overline{G}(i)$ is computed using (8) where the corresponding parameters $p(i), q(i)$ are used. At each time n, a stationary distribution $\overline{G}(\theta_n) \in \mathbf{G}$ is chosen where θ_n evolves according to an m-state Markov chain as described in Sec.II. We assume that the stationary degree distribution of the graph is sampled by a network administrator. How can the network administrator track the expected degree distribution of the fixed size Markov-modulated duplication deletion random graph without knowing the dynamics of the graph? The motivation for tracking the stationary degree distribution stems from social networks where the dynamics of the degree distribution evolve on a faster time scale than the Markov chain θ_n. Therefore, it suffices to track $\overline{G}(\theta_n)$ given observations.

At each time n, the network administrator samples a node from the graph based on degree distribution $\overline{G}(\theta_n)$ and records its degree $y_n(\theta_n)$. Let $Y_n(\theta_n) = \mathbf{e}_{y_n(\theta_n)}$ denote the observation vector where $\mathbf{e}_i \in \mathbb{R}^{\eta_0 \times 1}$ is the i-th standard unit vector. Such a sampling procedure can be time correlated. Therefore, we allow $Y_n(\theta_n)$ to be a mixing process with thee following assumption:

Assumption 4.1: For each $\theta \in \mathcal{M}$, the sequence $\{y_n(\theta)\}$ is stationary ϕ-mixing with sufficiently fast mixing rate such that the sequence $\{y_n(1), \ldots, y_n(m)\}$ is independent of $\{\theta_n\}$ and that for each $\theta \in \mathcal{M}$, $\{Y_n(\theta)\}$ is a stationary ϕ-mixing sequence with mixing rate ψ_n satisfying $\sum_{j=0}^{\infty} \psi_j^{1/2} < \infty$.

Remark 4.1: Because $\{y_n(\theta)\}$ is a stationary ϕ-mixing sequence for each $\theta \in \mathcal{M}$, $\{Y_n(\theta)\}$ is a bounded sequence of ϕ-mixing process for each $\theta \in \mathcal{M}$ [35, p. 82] (see also [36, p.170]). The stationarity implies that

$$\mathbf{E}Y_n(\theta) = \mathbf{E}Y_1(\theta) = \sum_{i=0}^{\infty} \mathbf{e}_i P(y_1(\theta) = i) = \sum_{i=0}^{\infty} \overline{G}^i(\theta)\mathbf{e}_i = \overline{G}(\theta) \tag{10}$$

The mixing rate given requires that for any positive integers i and j,

$$\begin{aligned} &\|\mathbf{E}_k I_{\{y_n(\theta)=i\}} - \overline{G}^i(\theta)\| \leq \psi_{n-k} \quad \text{for} \quad n \geq k, \\ &\|\mathbf{E}[I_{\{y_l(\theta)=j\}} - \overline{G}^j(\theta)][I_{\{y_n=i\}} - \overline{G}^i(\theta)]\| \leq \psi_{n-l}^{1/2}\psi_{l-k}^{1/2} \\ &\qquad\qquad\qquad \text{for any} \quad k < l < n, \end{aligned} \tag{11}$$

where $\mathbf{E}_k$ denotes the conditional expectation on the past data up to time k (i.e., condition on the σ-algebra $\mathcal{F}_k$ generated by $\{Y_j(\theta) : j \leq k\}$) and $I\{\cdot\}$ denotes the indicator function. Here, $\|\cdot\|$ is used to denote the Euclidean norm.

The analysis in this paper can be generalized to include certain non-stationary cases for the observation process $\{y_n(\theta)\}$.

For example, for each $\theta \in \mathcal{M}$, suppose $\{\zeta_n(\theta)\}$ is an ergodic finite state Markov chain.[9] Let $y_n(\theta) = \widetilde{f}(\zeta_n(\theta))$. The n-step transition probability matrix of the Markov chain converges to a matrix (with identical rows consisting of its stationary distribution) with exponential rate. Then it can be verified similar to [36, p. 178] that $y_n(\theta)$ is mixing. Although (10) does not hold, the analysis using mixing inequalities can still be obtained.

Given the observation sequence $Y_n(\theta_n),\ n = 0, 1, 2, \ldots$, the aim is to adaptively estimate $\overline{G}(\theta_n)$ via the following stochastic approximation algorithm with (small positive) constant step-size ε:

$$\widehat{G}_{n+1} = \widehat{G}_n + \varepsilon \left(Y_n(\theta_n) - \widehat{G}_n\right), \quad \widehat{G}_0 = \mathbf{e}_1 \tag{12}$$

To summarize, the evolution of the slow Markov chain θ_n and stochastic approximation algorithm (12) form a two-time-scale Markovian system as follows when $\rho, \varepsilon = o(\frac{1}{\eta_0})$

$$\begin{cases} \text{True system: } \overline{G}(\theta_n) \in \{\overline{G}(1), \ldots, \overline{G}(m)\}, \\ \qquad\qquad \text{where } \theta_n \text{ evolves according to } A^\rho = I + \rho Q, \\ \text{Algorithm: } \widehat{G}_{n+1} = \widehat{G}_n + \varepsilon \left(Y_n(\theta_n) - \widehat{G}_n\right), \\ \qquad\qquad Y_n(\theta_n) = \mathbf{e}_{y_n(\theta_n)}, \text{ where } y_n(\theta_n) \sim \overline{G}(\theta_n). \end{cases} \tag{13}$$

Note that the stochastic approximation algorithm (12) does not assume any knowledge of the Markov-modulated dynamics of the graph. The Markov chain assumption for the random graph dynamics is only used in our convergence and tracking analysis. By means of the stochastic approximation (12), the network administrator can track the stationary expected degree distribution $\overline{G}(\theta_n)$.

A. Tracking Error of the Stochastic Approximation Algorithm

The goal here is to analyze how well algorithm (12) tracks the empirical degree distribution of the fixed size Markov-modulated duplication-deletion graph. Define the tracking error as $\widetilde{G}_n = \widehat{G}_n - \overline{G}(\theta_n)$. Theorem 4.1 below shows that the difference between the sample path and the stationary degree distribution is small—implying that the stochastic approximation algorithm can successfully track the Markov-modulated node distribution given the noisy measurements. We again emphasize that no knowledge of the Markov chain parameters are required in the algorithm. It also finds the order of this difference in terms of ε and ρ.

Theorem 4.1: Consider the random graph $(m, A^\rho, \pi_0, p, q, r, \mathcal{G}_0)$. Suppose $\rho^2 \ll \varepsilon$ and Assumptions 2.1 and 4.1 hold.[10] Then, for sufficiently large n, the tracking error of the stochastic approximation algorithm (12) is

$$\mathbf{E}\|\widetilde{G}_n\|^2 = O\left(\varepsilon + \rho + \frac{\rho^2}{\varepsilon}\right). \tag{14}$$

Proof: The proof uses the perturbed Lyapunov function method and is provided in Appendix B. ■

[9]Respondent driven sampling (RDS) was introduced in [37] as an approach for sampling from hidden populations in social networks. RDS has been selected by the U.S. Centers for Disease Control and Prevention as part of the HIV behavioral surveillance system. RDS can be viewed as a form of Markov Chain Monte Carlo sampling [38].

[10]In this paper, we assume that $\rho = O(\varepsilon)$. Therefore, $\rho^2 \ll \varepsilon$.

Remark 4.2: Most existing literature analyzes stochastic approximation algorithms for tracking a parameter that evolves according to a "slowly time-varying" sample path of a continuous-valued process so that the parameter changes by small amounts over small intervals of time. When the rate of change of the underlying parameter is slower than the adaptation rate of the stochastic approximation algorithm (e.g., a slow random walk), the mean square tracking error can be analyzed as in [1], [17], [21]–[23], and [39]. In comparison, our analysis covers the case where the underlying parameter evolves with discrete jumps that can be arbitrarily large in magnitude on short intervals of time. Also, the jumps occur on the same time scale as the speed of adaptation of the stochastic approximation algorithm. We explicitly consider this Markovian time-varying parameter in our mean square error and weak convergence analysis.

As a corollary of Theorem 4.1, we obtain the following mean square error convergence result.

Corollary 4.1: Under the conditions of Theorem 4.1, if $\rho = O(\varepsilon)$,

$$\lim_{n\to\infty} \mathbf{E}\|\widetilde{G}_n\|^2 = O(\varepsilon).$$

Therefore,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \mathbf{E}\|\widetilde{G}_n\|^2 = 0.$$

■

B. Limit System of Regime-Switching Ordinary Differential Equations

The following theorem asserts that the sequence of estimates generated by the stochastic approximation algorithm (12) follows the dynamics of a Markov-modulated ordinary differential equation (ODE).

Before proceeding with the main theorem below, let us recall a definition.

Definition 4.1 (Weak Convergence): Let Z_k and Z be $\mathbb{R}^r$-valued random vectors. We say Z_k *converges weakly* to Z ($Z_k \Rightarrow Z$) if for any bounded and continuous function $f(\cdot)$, $Ef(Z_k) \to Ef(Z)$ as $k \to \infty$.

Weak convergence is a generalization of convergence in distribution to a function space.[11]

Theorem 4.2: Consider the Markov-modulated random graph generated according to Procedure 1, and the sequence of estimates $\{\widehat{G}_n\}$, generated by the stochastic approximation algorithm (12). Suppose Assumptions 2.1 and 4.1 hold and $\rho = O(\varepsilon)$. Define the continuous-time interpolated process

$$\widehat{G}^\varepsilon(t) = \widehat{G}_n,\ \theta^\varepsilon(t) = \theta_n \quad \text{for } t \in [n\varepsilon, (n+1)\varepsilon). \tag{15}$$

Then, as $\varepsilon \to 0$, $(\widehat{G}^\varepsilon(\cdot), \theta^\varepsilon(\cdot))$ converges weakly to $(\widehat{G}(\cdot), \theta(\cdot))$, where $\theta(\cdot)$ is a continuous-time Markov chain with generator Q, $\widehat{G}(\cdot)$ satisfies the Markov-modulated ODE

$$\frac{d\widehat{G}(t)}{dt} = -\widehat{G}(t) + \overline{G}(\theta(t)), \quad \widehat{G}(0) = \widehat{G}_0 \tag{16}$$

and $\overline{G}(\theta) \in \mathbf{G}$. ■

[11]We refer the interested reader to [1, Ch. 7] for further details on weak convergence and related matters.

The above theorem asserts that the limit system associated with the stochastic approximation algorithm (12) is a Markovian switched ODE (16). As mentioned in Sec.I, this is unusual since typically in averaging of stochastic approximation algorithms, convergence occurs to a deterministic ODE. The intuition behind this is that the Markov chain evolves on the same time-scale as the stochastic approximation algorithm. If the Markov chain evolved on a faster time-scale, then the limiting dynamics would be a deterministic ODE weighed by the stationary distribution of the Markov chain. If the Markov chain evolved slower than the dynamics of the stochastic approximation algorithm, then the asymptotic behavior would also be a deterministic ODE with the Markov chain being a constant.

C. Scaled Tracking Error

Next, we study the behavior of the scaled tracking error between the estimates generated by the stochastic approximation algorithm (12) and the expected degree distribution. The following theorem states that the tracking error should also satisfy a switching diffusion equation and provides a functional central limit theorem for this scaled tracking error. Let $\nu_k = \frac{\widehat{G}_k - \overline{G}(\theta_k)}{\sqrt{\varepsilon}}$ denote the scaled tracking error.

Theorem 4.3: Suppose Assumptions 2.1 and 4.1 hold. Define $\nu^\varepsilon(t) = \nu_k$ for $t \in [k\varepsilon, (k+1)\varepsilon)$. Then, $(\nu^\varepsilon(\cdot), \theta^\varepsilon(\cdot))$ converges weakly to $(\nu(\cdot), \theta(\cdot))$ such that $\nu(\cdot)$ is the solution of the following Markovian switched diffusion process

$$\nu(t) = -\int_0^t \nu(s)ds + \int_0^t \Sigma^{\frac{1}{2}}(\theta(\tau))d\omega(\tau). \tag{17}$$

Here, $\omega(\cdot)$ is an $\mathbb{R}^{\eta_0}$-dimensional standard Brownian motion. The covariance matrix $\Sigma(\theta)$ in (17) can be explicitly computed as

$$\Sigma(\theta) = Z(\theta)' D(\theta) + D(\theta) Z(\theta) - D(\theta) - \overline{G}(\theta)\overline{G}'(\theta). \tag{18}$$

Here, $D(\theta) = \text{diag}(\overline{G}(\theta))$ and $Z(\theta) = \big(I - B_{\eta_0}(\theta) + \mathbf{1}\overline{G}'(\theta)\big)^{-1}$, where $\overline{G}(\theta) \in \mathbf{G}$. For each $\theta \in \mathcal{M}$, $B_{\eta_0}(\theta)$ is computed using (7) where the corresponding parameters $p(i), q(i)$ are used. ■

For general switching processes, we refer to [40]. In fact, more complex continuous-state dependent switching rather than Markovian switching are considered there. Equation (18) reveals that the covariance matrix of the tracking error depends on $B_{\eta_0}(\theta)$ and $\overline{G}(\theta)$ and, consequently, on the parameters p and q of the random graph. Recall from Sec.II that $B_{\eta_0}(\theta)$ is the transition matrix of the Markov chain which models the evolution of the expected degree distribution in duplication-deletion random graphs and can be computed from Theorem 3.1.

V. Estimating the Degree Distribution of Infinite Duplication-Deletion Random Graphs

This section has two results: First, the results of Sec. IV are extended to infinite random graphs without Markovian dynamics generated according to Procedure 1. Second, we show how this analysis can be extended to Markov-modulated probability mass functions with denumerable support. The analysis is non-standard, since it is formulated on a Hilbert space.

A. Infinite Random Graphs Without Markovian Dynamics

Consider the infinite duplication-deletion random graph without Markovian dynamics generated according to Procedure 1 with 7-tuple $(1, 1, 1, 1, p, q, \mathcal{G}_0)$. In this section, let G_n represent the degree distribution of the infinite graph with support on the set of non-negative integers; its elements are denoted by $G_n^i, i = 0, 1, 2, \ldots$. Recall from Sec.III-B that, the size of such a graph increments at each time by one and thus the size of the graph at time n is equal to n; that is $\eta_n = n$. Therefore, the maximum degree of the graph at time n cannot exceed $n-1$ and $G_n^j = 0$ for $j \geq n$. Similar to the proof of Theorem 3.1, the following theorem asserts that the expected degree distribution of the infinite duplication-deletion random graph satisfies a recursive equation.

Theorem 5.1: Consider the infinite duplication-deletion random graph without Markovian dynamics generated according to Procedure 1 with 7-tuple $(1, 1, 1, 1, p, q, \mathcal{G}_0)$, Let $\overline{G}_n = \mathbf{E}\{G_n\}$ denote the expected degree distribution of nodes with support on the set of non-negative integers. Then, $\overline{G}_n$ satisfies the following recursion

$$\overline{G}_{n+1} = \overline{G}_n + \frac{1}{n} L^{(n)\prime}\overline{G}_n, \tag{19}$$

where $L^{(n)}$ is a generator matrix of infinite size with elements:

$$l_{ji}^{(n)} = \begin{cases} (1+q)\left(p^{i-1} + 1 + p(i-1)\right), \\ \qquad\qquad j = i-1, 1 \leq i, j \leq n \\ (1+q)\left(ip^{i-1}(1-p) + 1 + pi\right) - q(i+1), \\ \qquad\qquad j = i, 1 \leq i, j \leq n \\ (1+q)\binom{i+1}{i-1} p^{i-1}(1-p)^2 + q(i+1), \\ \qquad\qquad j = i+1, 1 \leq i, j \leq n \\ (1+q)\binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}, \\ \qquad\qquad j > i+1, 1 \leq i, j \leq n \\ 0, \qquad\qquad \text{otherwise} \end{cases} \tag{20}$$

Proof: The proof is similar to the proof of Theorem 3.1 and is omitted due to the lack of space. ■

Remark 5.1: Theorem 3.2 in Sec.III-B asserts that the expected degree distribution converges to a power law probability distribution $\overline{G}$ with exponent $\beta > 1$, if $\log p + p < \frac{q}{1+q} < p$; that is $\lim_{n\to\infty} \overline{G}_n^i = \frac{i^{-\beta}}{\zeta(\beta)}$. We assume that the dynamics of the degree distribution evolve on a faster time scale than the stochastic approximation algorithm. Therefore, it suffices to track the stationary degree distribution $\overline{G}$ given observations. At each time n, the network administrator samples from the graph and records the degree of a randomly chosen vertex of the graph which is denoted by y_n. Let $Y_n = \mathbf{e}_{y_n}$ denote the observation vector. Here, $\mathbf{e}_i$ is the i-th standard unit vector with support on the set of non-negative integers (i.e., $\mathbf{e}_i = (0, \ldots, 1, \ldots)' \in \mathbb{R}^\infty$). The following stochastic approximation algorithm is used to estimate the expected degree distribution of the graph from such samples.

$$\widehat{G}_{n+1} = \widehat{G}_n + \varepsilon \left(Y_n - \widehat{G}_n\right). \tag{21}$$

Here, $\varepsilon > 0$ denote a small positive step size and $\widehat{G}_0 = \mathbf{e}_1$. Therefore, (21) is a Hilbert-space-valued stochastic approximation algorithms. By means of the stochastic approximation (21), the network administrator can track the expected degree distribution of the infinite graph whose size increases over time.

Define

$$\widehat{G}^\varepsilon(t) = \widehat{G}_n \quad \text{for} \quad t \in [n\varepsilon, n\varepsilon + \varepsilon).$$

Then $\widehat{G}^\varepsilon(\cdot) \in D([0,\infty) : \ell_2)$ the space of functions defined on $[0,\infty)$ taking values in $\ell_2 = \{z \in \mathbb{R}^\infty : \sum_{i=0}^{\infty} \|z_i\|^2 < \infty\}$ such that the functions are right continuous and have left limits endowed with the Skorohod topology. Here, we obtain a weak convergence result of the interpolated sequence of iterates. Theorem 5.2 below asserts that the mean square tracking error is bounded and shows that the sequnce of estimates obtained by (21) converge to the solution of an ODE. Before proceeding to the main theorem, we shall use the following conditions.

Theorem 5.2: Suppose Assumption 4.1 holds with the modification that $m = 1$, i.e., there is no Markovian dynamics. Define $\widetilde{G}_n = \overline{G} - \widehat{G}_n$ Then, $\lim_{n\to\infty} \mathbf{E}\|\widetilde{G}_n\|^2 = O(\varepsilon)$. Also, $\widehat{G}^\varepsilon(\cdot)$ is tight in $D([0,\infty) : \ell_2)$. Any convergent subsequence has a limit $\widehat{G}(\cdot)$ that is the solution of the differential equation

$$\frac{d\widehat{G}(t)}{dt} = \overline{G} - \widehat{G}(t), \;\; \widehat{G}(0) = \mathbf{e}_1. \tag{22}$$

Proof: The proof is presented in Appendix E. The proof of the theorem is divided into several steps, which uses techniques in stochastic approximation [1] but with the modification that ℓ_2 is a Hilbert space (see [41], [42]). ∎

The above result concerns $n \to \infty$ and $\varepsilon \to 0$, εn remains to be bounded. We next obtain a result with $\varepsilon \to 0$, $n \to \infty$, $\varepsilon n \to \infty$.

Corollary 5.1: Consider $\widehat{G}^\varepsilon(\cdot + t_\varepsilon)$, where $t_\varepsilon \to \infty$ as $\varepsilon \to 0$. Under the condition of Theorem 5.2, $\widehat{G}^\varepsilon(\cdot + t_\varepsilon) \to \overline{G}$ in probability as $\varepsilon \to 0$.

Proof: Note that $\{\widehat{G}_k\}$ is tight. Define $\widehat{G}^{\varepsilon,\text{large}}(\cdot) = \widehat{G}^\varepsilon(\cdot + t_\varepsilon)$. Using the same approach, we can show that $\{\widehat{G}^{\varepsilon,\text{large}}(\cdot)\}$ is tight. We extract a weakly convergent subsequence of $(\widehat{G}^{\varepsilon,\text{large}}(\cdot), \widehat{G}^{\varepsilon,\text{large}}(\cdot - T))$ with limit denoted by $(\widehat{G}(\cdot), \widehat{G}_T(\cdot))$. We note that $\widehat{G}(0) = \widehat{G}_T(T)$ and that $\widehat{G}_T(0)$ belongs to a set that is bounded in probability. Writing it in variational form, we obtain

$$\begin{aligned}\widehat{G}_T(T) &= e^{-T}\widehat{G}_T(0) + \int_0^T e^{-(T-t)}\overline{G}dt \\ &= e^{-T}\widehat{G}_T(0) + \overline{G} - \int_T^\infty e^{-t}dt \\ &\to \overline{G} \;\; \text{as} \;\; T \to \infty.\end{aligned}$$

The desired result then follows. ∎

To study the rate of variation of estimation error, we define the sequence of scaled estimation error $\nu_n = (\widehat{G}_n - \overline{G})/\sqrt{\varepsilon}$. Theorem 5.3 asserts that the scaled estimation error satisfy a differential equation and provides a weak convergence results for it.

Theorem 5.3: Suppose assumptions of Theorem 5.2 hold. Then, for sufficiently small ε there is an N_ε such that $\mathbf{E}\{\langle \nu_n, \nu_n\rangle\} = O(1)$ for all $n > N_\varepsilon$. Define the sequence of continuous-time interpolation of estimation error as

$$\nu^\varepsilon(t) = \nu_n \text{ for } t \in [(n - N_\varepsilon)\varepsilon, (n - N_\varepsilon + 1)\varepsilon).$$

Under the assumptions of Theorem 5.2, $\{\nu^\varepsilon(\cdot)\}$ is tight in $D([0,\infty); \ell_2)$. Moreover, suppose that $\nu^\varepsilon(0)$ converges weakly to $\nu(0)$, $\nu^\varepsilon(\cdot)$ converges weakly to $\nu(\cdot)$ such that $\nu(\cdot)$ is the solution of the following stochastic differential equation

$$d\nu(t) = -\nu(t)dt + dW(t). \tag{23}$$

Here, $W(t) = \sum_{i=0}^{\infty} W_i(t)\mathbf{e}_i$ and the covariance operator is given by

$$\begin{aligned}\mathbf{E}\langle W(t), \upsilon\rangle\langle W(t), z\rangle &= t\langle z, \Gamma\upsilon\rangle \\ &= t\sum_{i=0}^{\infty} \sigma_i^2 \langle \mathbf{e}_i, \upsilon\rangle\langle \mathbf{e}_i, z\rangle \;\; \text{for} \;\; \upsilon, z \in \ell_2,\end{aligned}$$

where $W_i(\cdot)$ is a real-valued Wiener process with covariance $t\sigma_i^2$ and

$$\sigma_i^2 = \mathbf{E}[\langle Y_0 - \overline{G}, \mathbf{e}_i\rangle]^2 + 2\sum_{j=1}^{\infty} \mathbf{E}\langle Y_0 - \overline{G}, \mathbf{e}_i\rangle\langle Y_j - \overline{G}, \mathbf{e}_i\rangle.$$

Proof: The proof is presented in Appendix F. ∎

Note that the covariance σ_i^2 depends on the stationary expected degree distribution $\overline{G}$ and thus is a function of the power law exponent β.

B. Markov-Modulated Probability Mass Functions With Denumerable Support

Here, we extend the above results to the problem of tracking a time-varying probability mass function with infinite support. The aim is to track a probability mass function with support on the set of non-negative integers that evolves according to a slow Markov chain θ_n with m states and initial probability distribution π_0. The state space $\mathcal{M}$, and the transition probability matrix A^ρ of the underlying Markov chain θ_n are defined in (1) and (2), respectively. For each $\theta \in \mathcal{M}$, let

$$\overline{G}(\theta) = [\overline{G}^1(\theta), \overline{G}^2(\theta), \ldots]', \tag{24}$$

be a probability mass function with support on the set of non-negative integers such that $\sum_{i=1}^{\infty} \overline{G}^i(\theta) = 1$ and $\overline{G}^i(\theta) \propto i^{-\beta_\theta}$, where $\beta_\theta > 1$. When the underlying Markov chain θ_n jumps from one state to another within $\mathcal{M}$, $\overline{G}(\theta_n)$ switches accordingly.

At each time n, we sample $y_n(\theta_n)$ from PMF $\overline{G}(\theta_n)$; that is $y_n(\theta_n) \sim \overline{G}(\theta_n)$. Let $Y_n(\theta_n) = \mathbf{e}_{y_n(\theta_n)}$ denote the observation vector. To estimate $\overline{G}(\theta_n)$, the following constant step size stochastic approximation algorithm is deployed

$$\widehat{G}_{n+1} = \widehat{G}_n + \varepsilon(Y_n(\theta_n) - \widehat{G}_n). \tag{25}$$

Here $\varepsilon > 0$ denotes a small positive step size and $\widehat{G}_0 = \mathbf{e}_1$. We further assume that the Markov chain is slowly changing in that the rate of changes is an order slower than that of adaptation (25); that is $\rho = \varepsilon^2$.

To analyze the asymptotic properties of the stochastic approximation algorithm, we define the sequence of continuous-time interpolation $\widehat{G}^\varepsilon(t) = \widehat{G}_n$ for $t \in [n\varepsilon, n\varepsilon + \varepsilon)$. Similar to what have been obtained thus far for the non-Markovian case, with the details omitted, we obtain the following weak convergence results. Theorem 5.4 states that the sequence of estimates obtained via Hilbert-space-valued stochastic approximation algorithm (25) converges weakly to the solution of an ODE which depends on the initial distribution of the underlying Markov chain.

Theorem 5.4: Suppose Assumptions 2.1 and 4.1 hold. Then $\widehat{G}^\varepsilon(\cdot)$ is tight in $D([0,\infty) : \ell_2)$. Any convergent subsequence has a limit $\widehat{G}(\cdot)$ that is the solution of the differential equation

$$\frac{d\widehat{G}(t)}{dt} = \sum_{\theta=1}^{m} \overline{G}(\theta) p_\theta - \widehat{G}(t), \ \widehat{G}(0) = \mathbf{e}_1, \tag{26}$$

where

$$\overline{G}(\theta) = \sum_{i=0}^{\infty} \overline{G}^i(\theta)\mathbf{e}_i, \ \text{and} \ \ (p_\theta : \theta \le m) = \pi_0$$

is the initial probability distribution of Markov chain.

Proof: The proof is presented in Appendix G. ∎

Furthermore, we can obtain the following corollary. The proof is similar to that of Corollary 5.1 and thus omitted.

Corollary 5.2: Consider $\widehat{G}^\varepsilon(\cdot + t_\varepsilon)$, where $t_\varepsilon \to \infty$ as $\varepsilon \to 0$. Under the condition of Theorem 5.4, $\widehat{G}^\varepsilon(\cdot + t_\varepsilon) \to \underline{G} = \sum_{\theta=1}^m p_\theta \overline{G}(\theta)$ in probability as $\varepsilon \to 0$.

Redefine $\nu_n = (\widehat{G}_n - \underline{G})/\sqrt{\varepsilon}$. It can be shown that there exists N_ε such that the sequence $\{\nu_n : n \ge N_\varepsilon\}$ is tight. Next, redefine $\nu^\varepsilon(t) = \nu_n$ for $t \in [\varepsilon(n - N_\varepsilon), \varepsilon(n - N - \varepsilon) + \varepsilon)$. With a little more effort, we can also obtain the associated rates of convergence result, which is stated in the next theorem.

Theorem 5.5: Suppose Assumptions 2.1 and 4.1 hold. Then, $\{\nu^\varepsilon(\cdot)\}$ is tight in $D([0,\infty); \ell_2)$. Moreover, suppose that $\nu^\varepsilon(0)$ converges weakly to $\nu(0)$, then $\nu^\varepsilon(\cdot)$ converges weakly to $\nu(\cdot)$ such that $\nu(\cdot)$ is the solution of the following stochastic differential equation (SDE)

$$du(t) = -\nu(t)dt + \sum_{\theta=1}^{m} p_\theta dW(\theta, t), \tag{27}$$

where for each $\theta \in \mathcal{M}$, $W(\theta, \cdot)$ is a Wiener process as given in Theorem 5.3.

Proof: The proof is similar to the proof of Theorem 5.3 with modifications similar to those of the proof of Theorem 5.4. ∎

VI. Numerical Examples

In this section, numerical examples are given to illustrate the results from Sec.II, and Sec.IV.

The main conclusions are:

(i) The infinite duplication-deletion random graph without Markovian dynamics generated by the duplication-deletion Procedure 1 satisfies a power law as stated in Theorem 3.2; see Example 1.

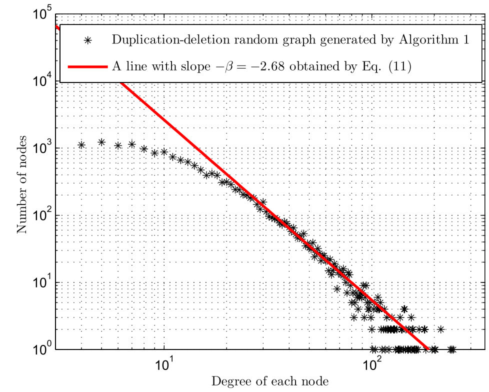

Fig. 1. The degree distribution of the duplication-deletion random graph satisfies a power law. The parameters are specified in Example 1 of Sec.VI.

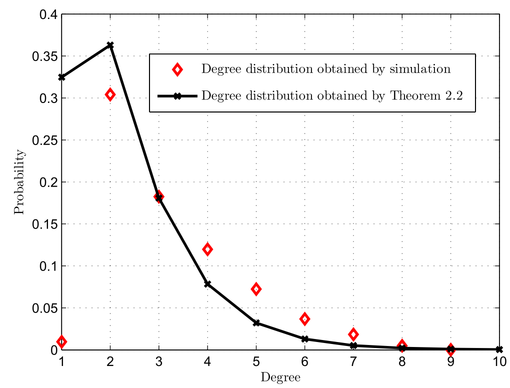

Fig. 2. The degree distribution of the fixed size duplication-deletion random graph. The parameters are specified in Example 2 of Sec.VI.

(ii) The degree distribution of the fixed size duplication-deletion random graph generated by the duplication-deletion Procedure 1 can be computed from Theorem 3.1. When η_0 (the size of the random graph) is sufficiently large, numerical results show that the degree distribution satisfies a power law as well; see Example 2.

(iii) The estimates obtained by stochastic approximation algorithm (12) follow the expected probability distribution precisely without information about the Markovian dynamics; see Example 3.

Example 1: Consider an infinite duplication-deletion random graph without Markovian dynamics (so $m = 1$) generated by Procedure 1 with $p = 0.5$ and $q = 0.1$. Theorem 3.2 implies that the degree sequence of the resulting graph satisfies a power law with exponent computed using (9). Fig.1 displays the un-normalized degree distribution on a logarithmic scale. The linearity in Fig.1 (excluding the nodes with very small degree), implies that the resulting graph from duplication-deletion process satisfies a power law. As can be seen in Fig.1, the power law is a better approximation for the middle points compared to both ends.

Example 2: Consider the fixed size duplication-deletion random graph generated by Procedure 1 with $r = 0$, $\eta_0 = 10$, $p = 0.4$, and $q = 0.1$. We consider $m = 1$ (no Markovian dynamics) to illustrate Theorem 3.1. Fig. 2 depicts the

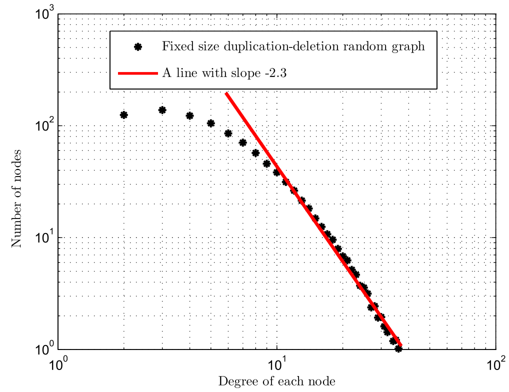

Fig. 3. Degree distribution of the fixed size duplication-deletion random graph satisfies a power law when η_0 is sufficiently large. The parameters are specified in Example 2 of Sec.VI.

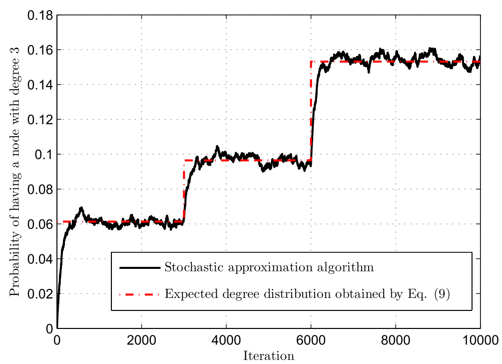

Fig. 4. The estimates obtained by SA algorithm (12) follows the expected PMF precisely with no knowledge of the Markovian dynamics. The parameters are specified in Example 3.

normalized degree distribution of the fixed size duplication-deletion random graph obtained by Theorem 3.1. As can be seen in Fig. 2, the computed degree distribution is close to that obtained by simulation. The numerical results show that the degree distribution of the fixed size random graph also satisfies a power law for some values of p when the size of random graph is sufficiently large. Fig. 3 shows the number of nodes with specific degree for the fixed size random graph obtained by Procedure 1 with $r = 0$, $\eta_0 = 1000$, $p = 0.4$, and $q = 0.1$ on a logarithmic scale for both horizontal and vertical axes.

Example 3: Consider the fixed size Markov-modulated duplication-deletion random graph generated by Procedure 1 with $r = 0$ and $\eta_0 = 500$. Assume that the underlying Markov chain has three states, $m = 3$. We choose the following values for probabilities of connection and deletion: state (1): $p = q = 0.05$, state (2): $p = 0.2$ and $q = 0.1$, and state (3): $p = 0.4$, $q = 0.15$. The sample path of the Markov chain jumps at times $n = 3000$ from state (1) to state (2) and $n = 6000$ from state (2) to state (3). As the state of the Markov chain evolves, the expected degree distribution, $\overline{G}(\theta)$, obtained by (8) evolves over time. The corresponding values for the expected degree distribution for nodes of degree $i = 3$ are displayed in Fig. 4 using a dotted line. The estimated probability mass function, $\widehat{G}_n$, obtained by the stochastic approximation algorithm (12) is plotted in Fig. 4 using a solid line. The figure shows that the estimates using by the stochastic approximation algorithm (12) follow the expected degree distribution (8) satisfactorily even though the algorithm has no information about the underlying Markovian dynamics.

VII. Conclusion

Markov-modulated duplication-deletion random graphs are analyzed in terms of their degree distribution. When the size of graph is fixed ($r = 0$) and ρ is small, the expected degree distribution of the Markov-modulated duplication-deletion random graph can be computed from (5) for each state of the underlying Markov chain. This result allows us to express the structure of network (degree distribution) in terms of the dynamics of the model. We also showed that, the infinite duplication-deletion random graph without Markovian dynamics generated according to Procedure 1 ($r = 1, m = 1$) satisfies a power law with component computed from (9). The importance of this result is that a single parameter (power law component) characterizes the structure of a possibly very large dynamic network.

Also a stochastic approximation algorithm was presented to adaptively estimate the degree distribution of random graphs. The stochastic approximation algorithm (12) does not assume knowledge of the Markov-modulated dynamics of the graph. Theorem 4.1 showed that the tracking error of the stochastic approximation algorithm is small and is in order of $O(\varepsilon)$. As a result of this bound, we showed that the scaled tracking error weakly converges to a diffusion process. Motivated by the analysis of social networks, we presented a Hilbert-space-valued stochastic approximation algorithm to estimate the expected degree distribution of the infinite duplication-deletion random graph without Markovian dynamics. The asymptotic behaviour of such an algorithm is analyzed in terms of the power law degree distribution. Finally, we extended the analysis to a Hilbert-space-valued stochastic approximation algorithm that aims to track a Markov-modulated probability mass function with denumerable support. Using weak convergence methods, it was shown that the estimates obtained via such an algorithm converge weakly to the solution of an ordinary differential equation. It was also shown that the interpolated sequence of scaled tracking error converges weakly to the solution of a stochastic differential equation.

Appendix

A. Proof of Theorem 3.1

The proof is based on the proof of [9, Lemma 4.1, Ch. 4, p. 79]. To compute the expected degree distribution of the Markov-modulated random graph, we find a relation between the number of nodes with specific degree at time n and the degree distribution of the graph at time $n-1$. Recall that the i-th element of F_n, F_n^i, denotes the number of vertices with degree i at time k. Given the resulting graph at time n, the aim is to find the expected number of nodes with degree $i+1$ at time $n+1$. The following events can occur that result in a node with degree $i+1$ at time $n+1$:

- *Degree of a node with degree i increments by one in the duplication step (Step 1 of the duplication-deletion*

Procedure 1) and remains unchanged in the deletion step (Step 2):

- A node with degree i is chosen at the duplication step as a parent node and remains unchanged in the deletion step. The probability of occurrence of such an event is
$$r\left(1-\frac{q(i+1)+q(1+pi)-q(1+pi)(i+1)/\eta_n}{\eta_n}\right)\frac{F_n^i}{\eta_n};$$
the probability of choosing a node with degree i is $\frac{F_n^i}{\eta_n}$ and the probability of the event that this node remains unchanged in the deletion step is[12]
$$1-\frac{q(i+1)+q(1+pi)-q(1+pi)(i+1)/\eta_n}{\eta_n}.$$
- One neighbor of a node with degree i is selected as a parent node; the parent node connects to its neighbors (including the node with degree i) with probability p in the edge-duplication part of Step 1. The probability of such an event is
$$r\frac{F_n^i pi}{\eta_n}\left(1-\frac{q(i+2)+q(1+p(i+1))-q(1+p(i+1))(i+2)/\eta_n}{\eta_n}\right).$$
Note that the node whose degree is incremented by one in this event should remain unaffected in Step 2; the probability of being unchanged in Step 2 for such a node is
$$1-\frac{q(i+2)+q(1+p(i+1))-q(1+p(i+1))(i+2)/\eta_n}{\eta_n}.$$

- *A node with degree $i+1$ remains unchanged in both Step 1 and Step 2 of Procedure 1:*
  - Using the same argument as above, the probability of such an event is
$$F_n^{i+1}\left(1-q\frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{\eta_n}}{\eta_n}\right)\times\left(1-r\frac{p(i+1)+1}{\eta_n}\right).$$
- *A new node with degree $i+1$ is generated in Step 1:*
  - The degree of the most recently generated node (in the vertex- duplication part of Step 1) increments to $i+1$; the new node connects to "i" neighbors of the parent node and remains unchanged in Step 2. The probability of this scenario is
$$r\left(1-q\frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{\eta_n}}{\eta_n}\right)\times\sum_{j\geq i}\frac{F_n^j}{\eta_n}\binom{j}{i}p^i(1-p)^{j-i}.$$
- *Degree of a node with degree $i+2$ decrements by one in Step 2:*
  - A node with degree $i+2$ remains unchanged in the duplication step and one of its neighbors is eliminated in the deletion step. The probability of this event is
$$q\left(\frac{i+2}{\eta_n}\right)\left(1-\frac{p(i+2)+1}{\eta_n}\right).$$
- *A node with degree $i+1$ is generated in Step 2:*
  - The degree of the node generated in the vertex-duplication part of duplication step within Step 2 increments to $i+1$. The probability of this event is
$$q\sum_{j\geq i}\frac{1}{\eta_n}F_n^j\binom{j}{i}p^i(1-p)^{j-i}.$$
- *Degree of a node with degree i increments by one in Step 2:*
  - A node with degree i remains unchanged in Step 1 and its degree increments by one in the duplication part of Step 2. The corresponding probability is
$$\frac{q(1+pi)}{\eta_n}\left(1-\frac{1+pi}{\eta_n}\right).$$

Let Ω denote the set of all arbitrary graphs and $\mathcal{F}_n$ denote the sigma algebra generated by graphs $\mathcal{G}_\tau$, $\tau\leq n$. Considering the above events that result in a node with degree $i+1$ at time $n+1$, the following recurrence formula can be derived for the conditional expectation of F_{n+1}^{i+1}:

$$\begin{aligned}\mathbf{E}\{F_{n+1}^{i+1}|\mathcal{F}_n\}
&=\left(1-q\frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{\eta_n}}{\eta_n}\right)\\
&\times\left(1-r\frac{p(i+1)+1}{\eta_n}\right)F_n^{i+1}\\
&+r\left(1-\frac{q(i+1)+q(1+pi)-q(1+pi)(i+1)/\eta_n}{\eta_n}\right)\\
&\times\left(\frac{1+pi}{\eta_n}\right)F_n^i+r\left(1-q\frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{\eta_n}}{\eta_n}\right)\\
&\times\sum_{j\geq i}\frac{F_n^j}{\eta_n}\binom{j}{i}p^i(1-p)^{j-i}\\
&+q\sum_{j\geq i}\frac{F_n^j}{\eta_n}\binom{j}{i}p^i(1-p)^{j-i}+q\left(\frac{i+2}{\eta_n}\right)\end{aligned}$$

[12]The deletion step (Step 2 of Procedure 1) comprises an edge-deletion step and a duplication step. The probability that the degree of node with degree i changes in the edge-deletion step is $\frac{q(i+1)}{\eta_n}$; either this node or one of its neighbors should be selected in the edge-deletion step. Also given that the degree of this node dose not change in the edge-deletion step, if either this node or one of its neighbor is selected in the duplication step (within Step 2) then the degree of this node increments by one with probability $\frac{1+pi}{\eta_n}$. Therefore, the probability that the degree of a node of degree i remains unchanged in Step 2 is
$$1-\frac{q(i+1)+q(1+pi)-q(1+pi)(i+1)/\eta_n}{\eta_n}.$$
Note that for simplicity in our analysis, it is assumed that the nodes whose degrees change in the edge-deletion part of Step 2, remain unchanged in the duplication part of Step 2 at that time instant. Also, the new node, which is generated in the vertex-duplication step of Step 1, remains unchanged in Step 2.

$$\times\left(1-\frac{p(i+2)+1}{\eta_n}\right)F_n^{i+2}+\frac{q(1+pi)}{\eta_n}\left(1-\frac{1+pi}{\eta_n}\right)F_n^i. \tag{28}$$

Let $\overline{F}_n^i=\mathbf{E}\{F_n\}$. By taking expectation of both sides of (28) with respect to trivial sigma algebra $\{\Omega,\emptyset\}$, the smoothing property of conditional expectations yields:

$$\begin{aligned}
&\overline{F}_{n+1}^{i+1}\\
&=\left(1-q\frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{\eta_n}}{\eta_n}\right)\\
&\times\left(1-r\frac{p(i+1)+1}{\eta_n}\right)\overline{F}_n^{i+1}\\
&+r\left(1-\frac{q(i+1)+q(1+pi)-\frac{q(1+pi)(i+1)}{\eta_n}}{\eta_n}\right)\\
&\times\left(\frac{1+pi}{\eta_n}\right)\overline{F}_n^i+r\left(1-q\frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{\eta_n}}{\eta_n}\right)\\
&\times\sum_{j\ge i}\frac{\overline{F}_n^j}{\eta_n}\binom{j}{i}p^i(1-p)^{j-i}+q\sum_{j\ge i}\frac{1}{\eta_n}\overline{F}_n^j\binom{j}{i}p^i(1-p)^{j-i}\\
&+q\left(\frac{i+2}{\eta_n}\right)\left(1-\frac{p(i+2)+1}{\eta_n}\right)\overline{F}_n^{i+2}\\
&+\frac{q(1+pi)}{\eta_n}\left(1-\frac{1+pi}{\eta_n}\right)\overline{F}_n^i.
\end{aligned}\tag{29}$$

Assuming that size of the graph is sufficiently large, each term like $\frac{\overline{F}_n^i}{\eta_n^2}$ can be neglected. Eq. (29) can be written as

$$\begin{aligned}
\overline{F}_{n+1}^{i+1}&=\left(1-\frac{q(i+2)+(r+q)(p(i+1)+1)}{\eta_n}\right)\overline{F}_n^{i+1}\\
&+\left(\frac{(1+pi)(r+q)}{\eta_n}\right)\overline{F}_n^i+q\left(\frac{i+2}{\eta_n}\right)\overline{F}_n^{i+2}\\
&+q\sum_{j\ge i}\frac{1}{\eta_n}\overline{F}_n^j\binom{j}{i}p^i(1-p)^{j-i}.
\end{aligned}\tag{30}$$

Using (29), we can write the following recursion for the $(i+1)$-th element of $\overline{G}_{n+1}$:

$$\begin{aligned}
\overline{G}_{n+1}^{i+1}&=\left(\frac{\eta_n-\big(q(i+2)+(r+q)(p(i+1)+1)\big)}{\eta_{n+1}}\right)\overline{G}_n^{i+1}\\
&+\left(\frac{(1+pi)(r+q)}{\eta_{n+1}}\right)\overline{G}_n^i+q\left(\frac{i+2}{\eta_{n+1}}\right)\overline{G}_n^{i+2}\\
&+q\sum_{j\ge i}\frac{1}{\eta_{n+1}}\overline{G}_n^j\binom{j}{i}p^i(1-p)^{j-i}.
\end{aligned}\tag{31}$$

Since the probability of duplication step $r=0$, the number of vertices does not increase. Thus, $\eta_n=\eta_0$ and (31) can be written as

$$\begin{aligned}
\overline{G}_{n+1}^{i+1}&=\left(1-\frac{1}{\eta_0}\big(q(i+2)+q(p(i+1)+1)\big)\right)\overline{G}_n^{i+1}\\
&+\frac{1}{\eta_0}\Big((1+pi)q\overline{G}_n^i+\frac{1}{\eta_0}q(i+2)\overline{G}_n^{i+2}\Big)\\
&+\frac{1}{\eta_0}q\sum_{j\ge i}\overline{G}_n^j\binom{j}{i}p^i(1-p)^{j-i}.
\end{aligned}\tag{32}$$

It is clear in (32) that the vector $\overline{G}_{n+1}$ depends on elements of $\overline{G}_n$. Using matrix notation, (32) can be expressed as

$$\overline{G}_{n+1}=\left(I+\frac{1}{\eta_0}L'\right)\overline{G}_n \tag{33}$$

where L is defined as (6).

To prove that L is a generator, we need to show that $l_{ii}<0$ and $\sum_{i=1}^{\eta_0}l_{ki}=0$. Accordingly,

$$\begin{aligned}
\sum_{i=1}^{\eta_0}l_{ki}&=-(q(k+1)+q(1+pk))+(1+pk)q\\
&\quad+qk+q\sum_{k\le i-1}\binom{k}{i-1}p^{i-1}(1-p)^{k-i+1}\\
&=-q+q\sum_{k\le i-1}\binom{k}{i-1}p^{i-1}(1-p)^{k-i+1}.
\end{aligned}\tag{34}$$

Let $m=i-1$. Then, (34) can be rewritten as

$$\begin{aligned}
\sum_{i=1}^{\eta_0}l_{ik}&=-q+q\sum_{m=0}^{k}\binom{k}{m}p^m(1-p)^{k-m}\\
&=-q+q(1-p)^k\sum_{m=0}^{k}\binom{k}{m}\left(\frac{p}{1-p}\right)^m.
\end{aligned}\tag{35}$$

Knowing that $\sum_{m=0}^{k}\binom{k}{m}a^m=(1+a)^k$, (35) can be written as

$$\sum_{i=1}^{\eta_0}l_{ik}=-q+q(1-p)^k\left(\frac{1}{1-p}\right)^k=0. \tag{36}$$

Also, it can be shown that $l_{ii}<0$. Since $p^{i-1}\le 1$, $p^{i-1}<1+\frac{2}{i}+p+p^i$. Consequently, $iqp^{i-1}(1-p)-q(i+2+ip)<0$. Therefore, $l_{ii}<0$ and the desired result follows.

B. Proof of Theorem 4.1

Define the Lyapunov function $V(x)=(x'x)/2$ for $x\in\mathbb{R}^{N_0}$. Use $\mathbf{E}_n$ to denote the conditional expectation with respect to the σ-algebra $\mathcal{H}_n$ generated by $\{Y_j(\theta_j),\theta_j,\quad j\le n\}$. Then,

$$\begin{aligned}
&\mathbf{E}_n\{V(\widetilde{G}_{n+1})-V(\widetilde{G}_n)\}\\
&=\mathbf{E}_n\Big\{\widetilde{G}_n'[-\varepsilon\widetilde{G}_n+\varepsilon\left(Y_n(\theta_n)-\overline{G}(\theta_n)\right)+\overline{G}(\theta_n)-\overline{G}(\theta_{n+1})]\Big\}\\
&\quad+\mathbf{E}_n\Big\{\|-\varepsilon\widetilde{G}_n+\varepsilon\left(Y_n(\theta_n)-\overline{G}(\theta_n)\right)+\overline{G}(\theta_n)-\overline{G}(\theta_{n+1})\|^2\Big\}
\end{aligned}\tag{37}$$

where $Y_n(\theta_n)$ and $\overline{G}(\theta_n)$ are vectors in $\mathbb{R}^{N_0}$ with elements $Y_n^i(\theta_n)$ and $\overline{G}(\theta_n)^i$, $1\le i\le N_0$, respectively. Due to the Markovian assumption and the structure of the transition matrix of θ_n, defined in (2),

$$\begin{aligned}\mathbf{E}_n\{\overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\} &= \mathbf{E}\{\overline{G}(\theta_n) - \overline{G}(\theta_{n+1})|\theta_n\} \\ &= \sum_{i=1}^{m} \mathbf{E}\{\overline{G}(i) - \overline{G}(\theta_{n+1})|\theta_n = i\} I\{\theta_n = i\} \\ &= \sum_{i=1}^{m} \left[\overline{G}(i) - \sum_{j=1}^{m} \overline{G}(j) A_{ij}^{\rho}\right] I_{\{\theta_n = i\}} \\ &= -\rho \sum_{i=1}^{m}\sum_{j=1}^{m} \overline{G}(j) q_{ij} I\{\theta_n = i\} = O(\rho) \end{aligned} \quad (38)$$

where $I\{\cdot\}$ denotes the indicator function. Similarly, it is easily seen that

$$\mathbf{E}_n\{\|\overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\|^2\} = O(\rho). \quad (39)$$

Using K to denote a generic positive value (with the notation $KK = K$ and $K + K = K$), a familiar inequality $ab \le \frac{a^2+b^2}{2}$ yields

$$O(\varepsilon\rho) = O(\varepsilon^2 + \rho^2). \quad (40)$$

Moreover, we have $\|\widetilde{G}_n\| = \|\widetilde{G}_n\| \cdot 1 \le (\|\widetilde{G}_n\|^2 + 1)/2$. Thus,

$$O(\rho)\|\widetilde{G}_n\| \le O(\rho)\left(V(\widetilde{G}_n) + 1\right). \quad (41)$$

Then, detailed estimates lead to

$$\begin{aligned}\mathbf{E}_n &\left\{\left\| -\varepsilon\widetilde{G}_n + \varepsilon\left(Y_n(\theta_n) - \overline{G}(\theta_n)\right) + \overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\right\|^2\right\} \\ &\le K\mathbf{E}_n\Big\{\varepsilon^2\|\widetilde{G}_n\|^2 + \varepsilon^2\|(Y_n(\theta_n) - \overline{G}(\theta_n)\|^2 \\ &+ \varepsilon^2\left\|\widetilde{G}_n'\left(Y_n(\theta_n) - \overline{G}(\theta_{n+1})\right)\right\| \\ &+ \varepsilon\|\widetilde{G}_n'\left(\overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\right)\| \\ &+ \varepsilon\|\left(Y_n(\theta_n) - \overline{G}(\theta_n)\right)'\left(\overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\right)\|\Big\} \\ &+ \mathbf{E}_n\{\|\overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\|\}^2. \end{aligned} \quad (42)$$

It follows that

$$\begin{aligned}\mathbf{E}_n&\left\{\left\| -\varepsilon\widetilde{G}_n + \varepsilon\left(Y_n(\theta_n) - \overline{G}(\theta_n)\right) + \overline{G}(\theta_n) - \overline{G}(\theta_{n+1})\right\|^2\right\} \\ &= O(\varepsilon^2 + \rho^2)(V(\widetilde{G}_n) + 1). \end{aligned} \quad (43)$$

Furthermore,

$$\begin{aligned}&\{V(\widetilde{G}_{n+1}) - V(\widetilde{G}_n)\} \\ &= -2\varepsilon V(\widetilde{G}_n) + \varepsilon\mathbf{E}_n\{\widetilde{G}_n'[Y_n(\theta_n) - \overline{G}(\theta_n)]\} \\ &\quad + \mathbf{E}_n\{\widetilde{G}_n'[\overline{G}(\theta_{n+1}) - \overline{G}(\theta_n)]\} + O(\varepsilon^2 + \rho^2)(V(\widetilde{G}_n) + 1). \end{aligned} \quad (44)$$

To obtain the desired bound, define V_1^{ρ} and V_2^{ρ} as follows:

$$\begin{aligned}V_1^{\rho}(\widetilde{G}, n) &= \varepsilon\textstyle\sum_{j=n}^{\infty} \widetilde{G}'\mathbf{E}_n\{Y_j(\theta_j) - \overline{G}(\theta_j)\}, \\ V_2^{\rho}(\widetilde{G}, n) &= \textstyle\sum_{j=n}^{\infty} \widetilde{G}'\mathbf{E}_n\{\overline{G}(\theta_j) - \overline{G}(\theta_{j+1})\}. \end{aligned} \quad (45)$$

It can be shown that

$$\begin{aligned}|V_1^{\rho}(\widetilde{G}, n)| &= O(\varepsilon)(V(\widetilde{G}) + 1), \\ |V_2^{\rho}(\widetilde{G}, n)| &= O(\rho)(V(\widetilde{G}) + 1). \end{aligned} \quad (46)$$

Define $W(\widetilde{G}, n)$ as

$$W(\widetilde{G}, n) = V(\widetilde{G}) + V_1^{\rho}(\widetilde{G}, n) + V_2^{\rho}(\widetilde{G}, n). \quad (47)$$

This leads to

$$\begin{aligned}\mathbf{E}_n\{W(\widetilde{G}_{n+1}, n+1) - W(\widetilde{G}_n, n)\} &= \mathbf{E}_n\{V_1^{\rho}(\widetilde{G}_{n+1}, n+1) - V_1^{\rho}(\widetilde{G}_n, n)\} \\ &\quad + \mathbf{E}_n\{V(\widetilde{G}_{n+1}) - V(\widetilde{G}_n)\} \\ &\quad + \mathbf{E}_n\{V_2^{\rho}(\widetilde{G}_{n+1}, n+1) - V_2^{\rho}(\widetilde{G}_n, n)\}. \end{aligned} \quad (48)$$

Moreover,

$$\begin{aligned}\mathbf{E}_n\{W(\widetilde{G}_{n+1}, n+1) - W(\widetilde{G}_n, n)\} &= -2\varepsilon V(\widetilde{G}_n) \\ &\quad + O(\varepsilon^2 + \rho^2)(V(\widetilde{G}_n) + 1). \end{aligned} \quad (49)$$

Equation (49) can be rewritten as

$$\begin{aligned}&\mathbf{E}_n\{W(\widetilde{G}_{n+1}, n+1) - W(\widetilde{G}_n, n)\} \\ &\le -2\varepsilon W(\widetilde{G}_n, n) + O(\varepsilon^2 + \rho^2)(W(\widetilde{G}_n, n) + 1). \end{aligned} \quad (50)$$

If ε and ρ are chosen small enough, then there exists a small λ such that $-2\varepsilon + O(\rho^2) + O(\varepsilon^2) \le -\lambda\varepsilon$. Therefore, (50) can be rearranged as

$$\begin{aligned}\mathbf{E}_n\{W(\widetilde{G}_{n+1}, n+1)\} &\le (1 - \lambda\varepsilon)W(\widetilde{G}_n, n) \\ &\quad + O(\varepsilon^2 + \rho^2). \end{aligned} \quad (51)$$

Taking expectation of both sides yields

$$\begin{aligned}\mathbf{E}\{W(\widetilde{G}_{n+1}, n+1)\} &\le (1 - \lambda\varepsilon)\mathbf{E}\{W(\widetilde{G}_n, n)\} \\ &\quad + O(\varepsilon^2 + \rho^2). \end{aligned} \quad (52)$$

Iterating on (52) then results

$$\begin{aligned}\mathbf{E}\{W(\widetilde{G}_{n+1}, n+1)\} &\le (1 - \lambda\varepsilon)^{n-N_\rho}\mathbf{E}\{W(\widetilde{G}_{N_\rho}, N_\rho)\} \\ &\quad + \sum_{j=N_\rho}^{n} O(\varepsilon^2 + \rho^2)(1 - \lambda\varepsilon)^{j-N_\rho}. \end{aligned} \quad (53)$$

As a result,

$$\begin{aligned}\mathbf{E}\{W(\widetilde{G}_{n+1}, n+1)\} &\le (1 - \lambda\varepsilon)^{n-N_\rho}\mathbf{E}\{W(\widetilde{G}_{N_\rho}, N_\rho)\} \\ &\quad + O\left(\varepsilon + \rho^2/\varepsilon\right). \end{aligned} \quad (54)$$

If n is large enough, one can approximate $(1 - \lambda\varepsilon)^{n-N_\rho} = O(\varepsilon)$. Therefore,

$$\mathbf{E}\{W(\widetilde{G}_{n+1}, n+1)\} \le O\left(\varepsilon + \frac{\rho^2}{\varepsilon}\right) \quad (55)$$

Finally, using (46) and replacing $W(\widetilde{G}_{n+1}, n+1)$ with $V(\widetilde{G}_{n+1})$, we obtain

$$\mathbf{E}\{V(\widetilde{G}_{n+1})\} \le O\left(\rho + \varepsilon + \frac{\rho^2}{\varepsilon}\right). \quad (56)$$

C. Sketch of the Proof of Theorem 4.2

Since the proof is similar to [18, Th. 4.5], we only indicate the main steps in what follows and omit most of the verbatim details.

Step 1: First, we show that the two component process $(\widehat{G}^\varepsilon(\cdot), \theta^\varepsilon(\cdot))$ is tight in $D([0,T] : \mathbb{R}^{\eta_0} \times m)$. Using techniques similar to [43, Th. 4.3], it can be shown that $\theta^\varepsilon(\cdot)$ converges weakly to a continuous-time Markov chain generated by Q. Thus, we mainly need to consider $\widehat{G}^\varepsilon(\cdot)$. We show that

$$\lim_{\Delta\to 0} \limsup_{\varepsilon\to 0} \mathbf{E}\left[\sup_{0\le s\le \Delta} \mathbf{E}_t^\varepsilon \left\|\widehat{G}^\varepsilon(t+s) - \widehat{G}^\varepsilon(t)\right\|^2\right] = 0 \quad (57)$$

where $\mathbf{E}_t^\varepsilon$ denotes the conditioning on the past information up to t. Then, the tightness follows from the criterion [35, p. 47].

Step 2: Since $(\widehat{G}^\varepsilon(\cdot), \theta^\varepsilon(\cdot))$ is tight, we can extract weakly convergent subsequence according to the Prohorov theorem; see [1]. To figure out the limit, we show that $(\widehat{G}^\varepsilon(\cdot), \theta^\varepsilon(\cdot))$ is a solution of the martingale problem with operator L_0. For each $i \in \mathcal{M}$ and continuously differential function with compact support $f(\cdot, i)$, the operator is given by

$$\begin{aligned} L_0 f(\widehat{G}, i) &= \nabla f'(\widehat{G}, i)[-\widehat{G} + \overline{G}(i)] \\ &\quad + \sum_{j\in\mathcal{M}} q_{ij} f(\widehat{G}, j), \ \ i \in \mathcal{M}. \end{aligned} \quad (58)$$

We can further demonstrate the martingale problem with operator L_0 has a unique solution (in the sense of in distribution). Thus, the desired convergence property follows.

D. Sketch of the Proof of Theorem 4.3

The proof comprises of four steps as described below:

Step 1: First, note

$$\begin{aligned} \nu_{n+1} &= \nu_n - \varepsilon\nu_n + \sqrt{\varepsilon}(y_{n+1} - \mathbf{E}\overline{G}(\theta_n)) \\ &\quad + \frac{\mathbf{E}[\overline{G}(\theta_n) - \overline{G}(\theta_{n+1}]}{\sqrt{\varepsilon}}. \end{aligned} \quad (59)$$

The approach is similar to that of [18, Th. 5.6]. Therefore, we will be brief.

Step 2: Define an operator

$$\begin{aligned} \mathcal{L} f(\nu, i) &= -\nabla f'(\nu, i)\nu + \frac{1}{2}\text{tr}[\nabla^2 f(\nu, i)\Sigma(i)] \\ &\quad + \sum_{j\in\mathcal{M}} q_{ij} f(\nu, j), \ \ i \in \mathcal{M}, \end{aligned} \quad (60)$$

for function $f(\cdot, i)$ with compact support that has continuous partial derivatives with respect to ν up to the second order. It can be shown that the associated martingale problem has a unique solution (in the sense of in distribution).

Step 3: It is natural now to work with a truncated process. For a fixed, but otherwise arbitrary $r_1 > 0$, define a truncation function

$$q^{r_1}(x) = \begin{cases} 1, & \text{if } x \in S^{r_1}, \\ 0, & \text{if } x \in \mathbb{R}^{\eta_0} - S^{r_1}, \end{cases}$$

where $S^{r_1} = \{x \in \mathbb{R}^{\eta_0} : \|x\| \le r_1\}$. Then, we obtain the truncated iterates

$$\begin{aligned} \nu_{n+1}^{r_1} &= \nu_n^{r_1} - \varepsilon\nu_n^{r_1} q^{r_1}(\nu_n^{r_1}) + \sqrt{\varepsilon}(y_{n+1} - \mathbf{E}\overline{G}(\theta_n)) \\ &\quad + \frac{\mathbf{E}[\overline{G}(\theta_n) - \overline{G}(\theta_{n+1}]}{\sqrt{\varepsilon}} q^{r_1}(\nu_n^{r_1}). \end{aligned} \quad (61)$$

Define $\nu^{\varepsilon,r_1}(t) = \nu_n^{r_1}$ for $t \in [\varepsilon n, \varepsilon n + \varepsilon)$. Then, $\nu^{\varepsilon,r_1}(\cdot)$ is an r-truncation of $\nu^\varepsilon(\cdot)$; see [1, p. 284] for a definition. We then show the truncated process $(\nu^{\varepsilon,r_1}(\cdot), \theta^\varepsilon(\cdot))$ is tight. Moreover, by Prohorov's theorem, we can extract a convergent subsequence with limit $(\nu^{r_1}(\cdot), \theta(\cdot))$ such that the limit $(\nu^{r_1}(\cdot), \theta(\cdot))$ is the solution of the martingale problem with operator $\mathcal{L}^{r_1}$ defined by

$$\begin{aligned} \mathcal{L}^{r_1} f^{r_1}(\nu, i) &= -\nabla\prime f^{r_1}(\nu, i)\nu + \frac{1}{2}\text{tr}[\nabla^2 f^{r_1}(\nu, i)\Sigma(i)] \\ &\quad + \sum_{j\in\mathcal{M}} q_{ij} f^{r_1}(\nu, j) \end{aligned} \quad (62)$$

for $i \in \mathcal{M}$, where $f^{r_1}(\nu, i) = f(\nu, i) q^{r_1}(\nu)$.

Step 4: Letting $r_1 \to \infty$, we show that the un-truncated process also converges and the limit, denoted by $(\nu(\cdot), \theta(\cdot))$, is precisely the martingale problem with operator $\mathcal{L}$ defined in (62). The limit covariance can further be evaluated as in [18, Lemma 5.2].

E. Proof of Theorem 5.2

The proof of the theorem is divided into several steps and uses techniques in stochastic approximation [1] but with the modification that ℓ_2 is a Hilbert space (see [41], [42]). Whenever possible, we only indicate the main idea and refer to the literature of stochastic approximation.

Step 0: Note that (22) has a unique solution for each initial condition since it is linear in $\widehat{G}(\cdot)$.

Step 1: Preliminary estimates. From (21), we obtain that for $0 < \varepsilon < 1$, the elements of $\widehat{G}_n$ are non-negative and add up to one. Thus, $\widehat{G}_n$ is bounded.

In addition, define $V(\widehat{G}) = \frac{1}{2}\langle \widehat{G} - \overline{G}, \widehat{G} - \overline{G}\rangle$, which can be thought of as a Lyapunov function. Then using perturbed Lyapunov function argument [1], it can be shown

$$\mathbf{E}V(\widehat{G}_n) = O(\varepsilon). \quad (63)$$

Step 2: Tightness of $\{\widehat{G}^\varepsilon(\cdot)\}$. Henceforth, we often use t/ε and $(t+s)/\varepsilon$ to denote $\lfloor t/\varepsilon \rfloor$ and $\lfloor (t+s)/\varepsilon \rfloor$, the integer parts of t/ε and $(t+s)/\varepsilon$, respectively. By using the boundedness of $\{\widehat{G}_n\}$ established in the first step together with the Hölder inequality, we have for each $0 < T < \infty$, any $t \ge 0$, any $0 < \delta$, any $0 < s \le \delta$, and $\varepsilon > 0$,

$$\begin{aligned} \mathbf{E}_t^\varepsilon \|\widehat{G}^\varepsilon(t+s) - \widehat{G}^\varepsilon(t)\|^2 &\le \mathbf{E}_t^\varepsilon \left| \varepsilon \sum_{j=t/\varepsilon}^{(t+s)/\varepsilon - 1} [Y_j - \widehat{G}_j] \right|^2 \\ &\le K\varepsilon\left(\frac{t+s}{\varepsilon} - \frac{t}{\varepsilon}\right) \le Ks, \end{aligned}$$

where $K > 0$ is independent of ε and $\mathbf{E}_t^\varepsilon$ denotes the conditional expectation with respect to $\mathcal{F}_t^\varepsilon$. Thus

$$\begin{aligned} \mathbf{E}_t^\varepsilon \|\widehat{G}^\varepsilon(t+s) - \widehat{G}^\varepsilon(t)\|^2 &\le Ks, \\ \lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathbf{E}[\sup_{0<s\le\delta} \mathbf{E}_t^\varepsilon \|\widehat{G}^\varepsilon(t+s) - \widehat{G}^\varepsilon(t)\|^2] &= 0. \end{aligned} \quad (64)$$

The tightness criterion (see [35, Th. 3, p. 47] with $\mathbb{R}^r$ replaced by ℓ_2; see also [41]) enables us to conclude that $\{\widehat{G}^\varepsilon(\cdot)\}$ is tight in $D([0,\infty) : \ell_2)$.

Step 3: Characterization of the limit process. Since $\{\widehat{G}^\varepsilon(\cdot)\}$ is tight, by Prohorov's theorem, we can extract a convergent subsequence. Select such a sequence and still denote it by $\widehat{G}^\varepsilon(\cdot)$ with limit denoted by $\widehat{G}(\cdot)$. By using the Skorohod representation, with a slight abuse of notation, we may assume that $\widehat{G}^\varepsilon(\cdot)$ converges to $\widehat{G}(\cdot)$ w.p.1 and the convergence is uniform on any bounded time interval. We shall show that $\widehat{G}(\cdot)$ is a solution of the martingale problem with operator

$$\mathcal{L}f(\widehat{G}) = \langle \nabla f(\widehat{G}), [\overline{G} - \widehat{G}]\rangle$$

for any $f(\cdot) \in C_0^1(\ell_2 : \mathbb{R})$ (collection of real-valued C^1 functions defined on ℓ_2 with compact support). We need to show that

$$f(\widehat{G}(t)) - f(\widehat{G}(0)) - \int_0^t \mathcal{L}f(\widehat{G}(\tau))d\tau \quad \text{is a marginale.}$$

To prove the martingale property, we pick out any bounded and continuous function $h(\cdot)$ defined on ℓ_2, any $T < \infty$, any $0 < t, s \le T$, any positive integer κ, and $t_{l_1} \le t$ for any $l \le \kappa$. To derive the desired property, we need only show that

$$\begin{aligned}&\mathbf{E}h(\widehat{G}(t_{l_1}) : l_1 \le \kappa)\\ &\times\left(f(\widehat{G}(t+s)) - f(\widehat{G}(t)) - \int_t^{t+s}\mathcal{L}f(\widehat{G}(\tau))d\tau\right) = 0. \quad (65)\end{aligned}$$

To prove (65), we work with the process indexed by ε. First, by the weak convergence and the Skorohod representation,

$$\begin{aligned}&\lim_{\varepsilon\to 0}\mathbf{E}h(\widehat{G}^\varepsilon(t_{l_1}) : l_1 \le \kappa)[f(\widehat{G}^\varepsilon(t+s)) - f(\widehat{G}^\varepsilon(t))]\\ &\quad = \mathbf{E}h(\widehat{G}(t_{l_1}) : l_1 \le \kappa)[f(\widehat{G}(t+s)) - f(\widehat{G}(t))]. \quad (66)\end{aligned}$$

Choose a sequence of integers $\{m_\varepsilon\}$ such that $m_\varepsilon \to \infty$ as $\varepsilon \to 0$ but $\Delta_\varepsilon = \varepsilon m_\varepsilon \to 0$. Next, we note

$$\begin{aligned}&f(\widehat{G}^\varepsilon(t+s)) - f(\widehat{G}^\varepsilon(t))\\ &= f(\widehat{G}_{(t+s)/\varepsilon}) - f(\widehat{G}_{t/\varepsilon})\\ &= \sum_{lm_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1}[f(\widehat{G}_{lm_\varepsilon + m_\varepsilon}) - f(\widehat{G}_{lm_\varepsilon})]\\ &= \varepsilon\sum_{lm_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1}\Big\langle \nabla f(\widehat{G}_{lm_\varepsilon}), \sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1}[Y_j - \widehat{G}_j]\Big\rangle + o(1)\\ &= \sum_{lm_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1}\Delta_\varepsilon\Big\langle \nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon}\sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1}[Y_j - \widehat{G}_j]\Big\rangle + o(1),\end{aligned}$$

where $o(1) \to 0$ in probability as $\varepsilon \to 0$. The stationarity and the mixing condition imply that

$$\begin{aligned}\frac{1}{m_\varepsilon}\sum_{j=lm_\varepsilon}^{lm_\varepsilon+m_\varepsilon-1}\mathbf{E}_{lm_\varepsilon}Y_j \to \mathbf{E}Y_0 &= \sum_{i=0}^{\infty}\mathbf{e}_i P(y_0 = i)\\ &= \overline{G} \quad \text{(in probability).}\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathbf{E}h(\widehat{G}^\varepsilon(t_{l_1}) : l_1 \le \kappa)\Bigg[\sum_{lm_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1}\Delta_\varepsilon\Big\langle\nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon}\sum_{j=lm_\varepsilon}^{lm_\varepsilon+m_\varepsilon-1}Y_j\Big\rangle\Bigg]\\ &= \mathbf{E}h(\widehat{G}^\varepsilon(t_{l_1}) : l_1 \le \kappa)\\ &\quad\times\Bigg[\sum_{lm_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1}\Delta_\varepsilon\Big\langle\nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon}\sum_{j=lm_\varepsilon}^{lm_\varepsilon+m_\varepsilon-1}\mathbf{E}_{lm_\varepsilon}Y_j\Big\rangle\Bigg]\\ &\to \mathbf{E}h(\widehat{G}(t_{l_1}) : l_1 \le \kappa)\\ &\quad\times\left(\int_t^{t+s}\langle\nabla f(\widehat{G}(\tau)), \overline{G}\rangle d\tau\right) \quad \text{as} \quad \varepsilon \to 0. \quad (67)\end{aligned}$$

Likewise,

$$\begin{aligned}&\mathbf{E}h(\widehat{G}^\varepsilon(t_{l_1}) : l_1 \le \kappa)\\ &\quad\times\Bigg[-\sum_{lm_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1}\Delta_\varepsilon\Big\langle\nabla f(\widehat{G}_{lm_\varepsilon}), \frac{\sum_{j=lm_\varepsilon}^{lm_\varepsilon+m_\varepsilon-1}\widehat{G}_j}{m_\varepsilon}\Big\rangle\Bigg]\\ &\to \mathbf{E}h(\widehat{G}(t_{l_1}) : l_1 \le \kappa)\Bigg[-\int_t^{t+s}\langle\nabla f(\widehat{G}(\tau)), \widehat{G}(\tau)\rangle d\tau\Bigg]. \quad (68)\end{aligned}$$

Combing (66)–(68), (65) follows. ∎

F. Proof of Theorem 5.3

In the proof of Theorem 5.3, we use several lemmas and propositions described below. From (21),

$$\nu_{n+1} = \nu_n - \varepsilon\nu_n + \sqrt{\varepsilon}(Y_n - \overline{G}). \quad (69)$$

Lemma 7.1: Under assumption Theorem 5.2, for sufficiently small ε, there is an N_ε such that $\mathbf{E}V(\nu_n) = O(1)$ for all $n \ge N_\varepsilon$.

Proof: The proof uses a perturbed Lyapunov function argument. ∎

To proceed, recall the definition of covariance operator and Wiener process [42], [44] on ℓ_2. A covariance Γ of an ℓ_2-valued random variable y is an operator from ℓ_2 to ℓ_2 defined by $\Gamma v = EY\langle v, y\rangle$ for any $v \in \ell_2$. A process $W(\cdot)$ is a zero mean (stationary increment) ℓ_2-valued Wiener process if there are mutually independent real-valued, zero mean, Wiener processes $\{W_i(\cdot)\}$ with covariances $t\rho_i$ satisfying $\sum_{i=0}^{\infty}\rho_i < \infty$ and there is an orthonormal sequence $\{\beta_i\}$ with $\beta_i \in \ell_2$ such that $W(t) = \sum_{i=0}^{\infty}W_i(t)\beta_i$. For $v, z \in \ell_2$, the covariance operator of $W(t)$ is defined by

$$\mathbf{E}\langle W(t), v\rangle\langle W(t), z\rangle = t\langle z, \Gamma v\rangle = t\sum_{i=0}^{\infty}\rho_i\langle\beta_i, v\rangle\langle\beta_i, z\rangle. \quad (70)$$

Lemma 7.2: Assume the conditions of Theorem 5.2. For any natural number $i \in \mathbb{N}$, define

$$W_i^\varepsilon(t) = \sqrt{\varepsilon}\sum_{j=0}^{t/\varepsilon-1}\langle Y_j - \overline{G}, \mathbf{e}_i\rangle.$$

Then $W_i^\varepsilon(\cdot)$ converges weakly to a real-valued Wiener process $W_i(\cdot)$ with covariance $t\sigma_i^2$, where

$$\sigma_i^2 = \mathbf{E}[\langle Y_0 - \overline{G}, \mathbf{e}_i\rangle]^2 + 2\sum_{j=1}^{\infty}\mathbf{E}\langle Y_0 - \overline{G}, \mathbf{e}_i\rangle\langle Y_j - \overline{G}, \mathbf{e}_i\rangle. \quad (71)$$

Proof: Note that with the use of inner product in ℓ_2, $\{\langle Y_n - \overline{G}, \mathbf{e}_i\rangle\}$ is a real-valued mixing sequence with mean 0. The desired convergence follows from the functional invariance principle for mixing process; see [1, Ch. 7] (see also [36], [41]). ■

Lemma 7.3: Under the conditions of Lemma 7.2, for $i \neq l$, $EW_i^\varepsilon(t)W_l^\varepsilon(t) = 0$. As a result, the limit Wiener processes $W_i(\cdot)$ and $W_l(\cdot)$ are independent.

Proof: It is straightforward that

$$\begin{aligned}\mathbf{E}W_i^\varepsilon(t)W_l^\varepsilon(t) &= \varepsilon\mathbf{E}\sum_{k=0}^{t/\varepsilon-1}\sum_{j=0}^{t/\varepsilon-1}\langle Y_j - \overline{G}, \mathbf{e}_i\rangle\langle Y_k - \overline{G}, \mathbf{e}_l\rangle\\ &= \varepsilon\mathbf{E}\sum_{k=0}^{t/\varepsilon-1}\sum_{j=0}^{t/\varepsilon-1}\langle Y_j - \overline{G}, \mathbf{e}_i\mathbf{e}_l'(Y_k - \overline{G})\rangle\\ &= 0 \quad \text{since} \quad \mathbf{e}_i\mathbf{e}_l' = 0 \in \mathbb{R}^{\infty\times\infty}.\end{aligned}$$

Since $\mathbf{E}W_i^\varepsilon(t) = 0$, we conclude that $\Sigma(W_i^\varepsilon(t), W_l^\varepsilon(t)) = 0$. Consequently, $\Sigma(W_i(t), W_l(t)) = 0$, and as a result $W_i(t)$ and $W_l(t)$ are independent Wiener processes. ■

Proposition 7.4: Under the conditions of Lemma 7.2, define

$$W^\varepsilon(t) = \sqrt{\varepsilon}\sum_{j=0}^{t/\varepsilon-1}[Y_j - \overline{G}]. \tag{72}$$

Then $W^\varepsilon(\cdot)$ converges weakly to $W(\cdot)$ such that

$$W(t) = \sum_{i=0}^{\infty} W_i(t)\mathbf{e}_i, \tag{73}$$

and the covariance operator is given by

$$\begin{aligned}\mathbf{E}\langle W(t), v\rangle\langle W(t), z\rangle &= t\langle z, \Gamma v\rangle\\ &= t\sum_{i=0}^{\infty}\sigma_i^2\langle \mathbf{e}_i, v\rangle\langle \mathbf{e}_i, z\rangle \quad \text{for} \quad v, z \in \ell_2,\end{aligned} \tag{74}$$

where σ_i^2 is defined in (71).

Proof: In view of the definition of (72), for any $\delta > 0$, $t > 0$, $0 < s \le \delta$, with $\mathbf{E}_t^\varepsilon$ denotes the conditional expectation with respect to $\mathcal{F}_t^\varepsilon$, using the mixing properties, we can show that

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\Big[\sup_{0\le\delta\le s}\mathbf{E}_t^\varepsilon\langle W^\varepsilon(t+s) - W^\varepsilon(t), W^\varepsilon(t+s) - W^\varepsilon(t)\rangle\Big] = 0.$$

Thus $W^\varepsilon(\cdot)$ is tight in $D([0,\infty);\ell_2)$. We can extract any weakly convergent subsequence and denote the limit by $W(\cdot)$. We next characterize its limit.

Again, using (72)

$$W^\varepsilon(t) = \sum_{i=0}^{\infty} W_i^\varepsilon(t)\mathbf{e}_i = \sqrt{\varepsilon}\sum_{i=0}^{\infty}\sum_{j=0}^{t/\varepsilon-1}\langle Y_j - \overline{G}, \mathbf{e}_i\rangle\mathbf{e}_i.$$

Therefore, for each $l \in \mathbb{N}$,

$$\mathbf{E}[\langle W^\varepsilon(t), \mathbf{e}_l\rangle]^2 = \mathbf{E}\|W_l^\varepsilon(t)\|^2 = t\sigma_l^2.$$

By virtue of exponential decay property of $\overline{G}^i \propto i^{-\beta}$, $\sum_{l=0}^{\infty}\sigma_l^2 < \infty$. By Lemma 7.2, $W_i^\varepsilon(\cdot)$ converges weakly to $W_i(\cdot)$. By virtue of Lemma 7.3, $W_i(\cdot)$ are independent Wiener processes. In view of the definition of Wiener process on ℓ_2, we conclude that $W^\varepsilon(\cdot)$ converges weakly to $W(\cdot)$ such that (73) holds. In addition, the structure of the covariance operator (74) is obtained. ■

We proceed to obtain the desired weak convergence of $\nu^\varepsilon(\cdot)$. Since the stochastic differential equation (23) is linear, there is a unique solution for each initial condition. The rest of the proof is similar to the finite dimensional counter part with necessary modifications similar to that of the proof of Theorem 5.2.

G. Proof of Theorem 5.4

Before proceeding to the main proof, we first state a preliminary result. The proofs of the assertions below can be found in [43, Ths. 3.6 and 4.3] and are thus omitted.

Lemma 7.5: Under Assumption 2.1, the following claims hold:

(a) Denote $p_n^\rho = [P(\theta_n^\rho = 1), \ldots, P(\theta_n^\rho = m)]$ and the n-step transition probability by $(A^\rho)^n$ with A^ρ given in (2) with $\rho = \varepsilon^2$. Then

$$\begin{aligned}p_n^\rho &= p(\rho n) + O(\rho + \rho^{-k_0 t/\rho}),\\ (A^\rho)^{n-n_0} &= \Xi(\rho n, \rho n_0) + O(\rho + e^{-k_0(t-t_0)/\rho}),\end{aligned} \tag{75}$$

where $p(t) \in \mathbb{R}^{1\times m}$ and $\Xi(t, t_0) \in \mathbb{R}^{m\times m}$ are the continuous-time probability vector and transition matrix satisfying

$$\begin{aligned}\frac{dp(t)}{dt} &= p(t)Q, \quad p(0) = p_0,\\ \frac{d\Xi(t,t_0)}{dt} &= \Xi(t,t_0)Q, \quad \Xi(t_0,t_0) = I,\end{aligned} \tag{76}$$

with $t_0 = \rho n_0$ and $t = \rho n$.

(b) $\theta^\rho(\cdot)$ converges weakly to $\theta(\cdot)$, a continuous-time Markov chain generated by Q.

To analyze the algorithm, the techniques developed in the proof of Theorem 5.2 are used along with the ideas and methods developed in [45]. The developments are similar in the approach and the results, but are more complex due to the added switching process. For example, with modifications, Step 1 in the proof of Theorem 5.2 can still be carried out. Also Step 2 can be proved. So the sequence $\{\widehat{G}^\varepsilon(\cdot)\}$ is tight.

To characterize the limit, we still use martingale averaging techniques. We shall only highlight the main difference here. In carrying out the analysis similar to that of Step 3 in the proof of Theorem 5.2, we will encounter the following term

$$\begin{aligned}&\mathbf{E}h(\widehat{G}^\varepsilon(t_{l_1}) : l_1 \le \kappa)\\ &\quad\times\Big[\sum_{lm_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1}\Delta_\varepsilon\langle\nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon}\sum_{j=lm_\varepsilon}^{lm_\varepsilon+m_\varepsilon-1}Y_j(\theta_j)\rangle\Big]\\ &= \mathbf{E}h(\widehat{G}^\varepsilon(t_{l_1}) : l_1 \le \kappa)\\ &\quad\times\Big[\sum_{lm_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1}\Delta_\varepsilon\langle\nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon}\sum_{j=lm_\varepsilon}^{lm_\varepsilon+m_\varepsilon-1}\mathbf{E}_{lm_\varepsilon}Y_j(\theta_j)\rangle\Big]\end{aligned}$$

$$= \mathbf{E}h(\widehat{G}^{\varepsilon}(t_{l_1}) : l_1 \leq \kappa) \times \Bigg[\sum_{lm_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} \Delta_\varepsilon \Big\langle \nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon} \sum_{\theta=1}^{m} \sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1} \mathbf{E}_{lm_\varepsilon} Y_j(\theta) I_{\{\theta_j = \theta\}} \Big\rangle \Bigg]. \quad (77)$$

Since $Y_j(\theta)$ and θ_j are independent, we have

$$\frac{1}{m_\varepsilon} \sum_{\theta=1}^{m} \sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1} \mathbf{E}_{lm_\varepsilon} Y_j(\theta) I_{\{\theta_j = \theta\}} \quad (78)$$

$$= \frac{1}{m_\varepsilon} \sum_{\imath_0=1}^{m} \sum_{\theta=1}^{m} \sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1} \mathbf{E}_{lm_\varepsilon} Y_j(\theta) P(\theta_j = \theta | \theta_{lm_\varepsilon} = \imath_0) I_{\{\theta_{lm_\varepsilon} = \imath_0\}}. \quad (79)$$

For each $\theta \in \mathcal{M}$, the averaging of $Y_j(\theta)$ can be carried out as in Case 1. We concentrate on the term involving Markov chain. By virtue of Lemma 7.5, noting $\rho = \varepsilon^2$ and using (75), we have

$$[A^\rho]^{j - lm_\varepsilon} = \Xi(\varepsilon^2 j, \varepsilon^2 lm_\varepsilon) + O(\varepsilon^2 + e^{-k_0(\varepsilon^2 j - \varepsilon^2 lm_\varepsilon)/\varepsilon^2}).$$

Because we are working with (25) and the stepsize is ε. In the interval $[l\Delta_\varepsilon, l\Delta_\varepsilon + \Delta_\varepsilon)$ with $\Delta_\varepsilon = \varepsilon m_\varepsilon$, it is readily seen that $\Xi(\varepsilon^2 j, \varepsilon^2 lm_\varepsilon) \to \Xi(0,0) = I$ as $\varepsilon \to 0$. As a result,

$$P(\theta_j = \theta | \theta_{lm_\varepsilon} = \imath_0) + o_\varepsilon(1) = \delta_{\imath_0, \theta} + o_\varepsilon(1) = \begin{cases} 1, & \text{if } \imath_0 = \theta \\ 0, & \text{otherwise} \end{cases} + o_\varepsilon(1),$$

where $o(1) \to 0$ as $o_\varepsilon(1) \to 0$ as $\varepsilon \to 0$. Putting the above estimates in (78), we obtain the limit in probability of

$$\frac{1}{m_\varepsilon} \sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1} \mathbf{E}_{lm_\varepsilon} Y_j(\theta) P(\theta_j = \theta | \theta_{lm_\varepsilon} = \imath_0) I_{\{\theta_{lm_\varepsilon} = \imath_0\}}$$

is the same as that of

$$\sum_{i=1}^{\infty} \mathbf{e}_i \theta_i(\theta) \delta_{\imath_0 \theta} I_{\{\theta^{\varepsilon^2}(\varepsilon^2 lm_\varepsilon) = \theta_0\}}.$$

This further leads to that as $\varepsilon \to 0$,

$$\begin{aligned} &\mathbf{E}h(\widehat{G}^{\varepsilon}(t_{l_1}) : l_1 \leq \kappa) \\ &\quad \times \Bigg[\sum_{lm_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} \Delta_\varepsilon \Big\langle \nabla f(\widehat{G}_{lm_\varepsilon}), \frac{1}{m_\varepsilon} \sum_{j=lm_\varepsilon}^{lm_\varepsilon + m_\varepsilon - 1} Y_j(\theta_j) \Big\rangle \Bigg] \\ &\to \mathbf{E}h(\widehat{G}(t_{l_1}) : l_1 \leq \kappa) \\ &\quad \times \Bigg[\int_t^{t+s} \big\langle \nabla f(\widehat{G}(\tau)), \theta_i(\theta) P(\theta(0) = \theta) \big\rangle d\tau \Bigg] \\ &= \mathbf{E}h(\widehat{G}(t_{l_1}) : l_1 \leq \kappa) \Bigg[\int_t^{t+s} \big\langle \nabla f(\widehat{G}(\tau)), \theta_i(\theta) \mathbf{e}_i p_\theta \big\rangle d\tau \Bigg]. \end{aligned}$$

■

REFERENCES

[1] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications* (Stochastic Modeling and Applied Probability), vol. 37, 2nd ed. New York, NY, USA: Springer-Verlag, 2003.

[2] X. G. Doukopoulos and G. V. Moustakides, "Blind adaptive channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1716–1725, Jul. 2006.

[3] X. G. Doukopoulos and G. V. Moustakides, "Fast and stable subspace tracking," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1452–1465, Apr. 2008.

[4] X. Xie and R. J. Evans, "Multiple target tracking and multiple frequency line tracking using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 39, no. 12, pp. 2659–2676, Dec. 1991.

[5] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York, NY, USA: Springer-Verlag, 2012.

[6] G. G. Yin and V. Krishnamurthy, "LMS algorithms for tracking slow Markov chains with applications to hidden Markov estimation and adaptive multiuser detection," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2475–2490, Jul. 2005.

[7] V. Krishnamurthy and T. Ryden, "Consistent estimation of linear and non-linear autoregressive models with Markov regime," *J. Time Ser. Anal.*, vol. 19, no. 3, pp. 291–307, May 1998.

[8] F. Chung, L. Lu, T. G. Dewey, and D. J. Galas, "Duplication models for biological networks," *J. Comput. Biol.*, vol. 10, no. 5, pp. 677–687, Oct. 2003.

[9] F. Chung and L. Lu, *Complex Graphs and Networks*. Arlington, VA, USA: NSF, 2006.

[10] D. López-Pintado, "Diffusion in complex social networks," *Games Econ. Behavior*, vol. 62, no. 2, pp. 573–590, Mar. 2008.

[11] F. Vega-Redendo, *Complex Social Networks* (Economic Society Monographs). Cambridge, U.K.: Cambridge Univ. Press, 2007.

[12] M. O. Jackson, *Social and Economic Networks*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[13] S. Eubank *et al.*, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, May 2004.

[14] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, p. 208701, Oct. 2002.

[15] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. S1, pp. 2566–2572, 2002.

[16] V. Krishnamurthy, K. Topely, and G. Yin, "Consensus formation in a two-time-scale Markovian system," *Multiscale Model. Simul.*, vol. 7, no. 4, pp. 1898–1927, 2009.

[17] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance* (Information and System Sciences). Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[18] G. Yin, V. Krishnamurthy, and C. Ion, "Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization," *SIAM J. Optim.*, vol. 14, no. 4, pp. 1187–1215, 2004.

[19] V. Krishnamurthy, O. N. Gharehshiran, and M. Hamdi, "Interactive sensing and decision making in social networks," *Found. Trends Signal Process.*, vol. 7, nos. 1–2, pp. 1–196, 2014.

[20] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin, "Distributed tracking of correlated equilibria in regime switching noncooperative games," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2435–2450, Oct. 2013.

[21] G. V. Moustakides, "Exponential convergence of products of random matrices: Application to adaptive algorithms," *Int. J. Adapt. Control Signal Process.*, vol. 12, no. 7, pp. 579–597, 1998.

[22] L. Guo, L. Ljung, and G.-J. Wang, "Necessary and sufficient conditions for stability of LMS," *IEEE Trans. Autom. Control*, vol. 42, no. 6, pp. 761–770, Jul. 1997.

[23] J. Reason and W. Ren, "Estimating the optimal adaptation gain for the LMS algorithm," in *Proc. 32nd IEEE Conf. Decision Control*, Dec. 1993, pp. 1587–1588.

[24] R. Durrett, *Random Graph Dynamics* (Statistical and Probabilistic Mathematics). Cambridge, U.K.: Cambridge Univ. Press, 2007.

[25] P. Erdös and A. Rényi, "On the evolution of random graphs," *Publication Math. Inst. Hungarian Acad. Sci.*, vol. 5, pp. 17–61, 1960.

[26] E. Lieberman, C. Hauert, and M. A. Nowak, "Evolutionary dynamics on graphs," *Nature*, vol. 433, no. 7023, pp. 312–316, Jan. 2005.

[27] R. Pastor-Satorras, E. Smith, and R. V. Solé, “Evolving protein interaction networks through gene duplication,” *J. Theoretical Biol.*, vol. 222, no. 2, pp. 199–210, May 2003.

[28] M. E. J. Newman, “Power laws, Pareto distributions and Zipf’s law,” *Contemporary Phys.*, vol. 46, no. 5, pp. 323–351, 2005.

[29] A.-L. Barabási and A. Reka, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[30] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, “The degree sequence of a scale-free random graph process,” *Random Struct. Algorithms*, vol. 18, no. 3, pp. 279–290, 2001.

[31] C. Cooper and A. Frieze, “A general model of web graphs,” *Random Struct. Algorithms*, vol. 22, no. 3, pp. 311–335, 2003.

[32] J. Shrager, T. Hogg, and B. A. Huberman, “A graph-dynamic model of the power law of practice and the problem-solving fan-effect,” *Science*, vol. 242, no. 4877, pp. 414–416, 1988.

[33] S. H. Strogatz, “Exploring complex networks,” *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.

[34] M. Hamdi, V. Krishnamurthy, and G. Yin. (Feb. 2013). “Tracking the empirical distribution of a Markov-modulated duplication-deletion random graph.” [Online]. Available: http://arxiv.org/abs/1303.0050

[35] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, With Applications to Stochastic Systems Theory*. Cambridge, MA, USA: MIT Press, 1984.

[36] P. Billingsley, *Convergence of Probability Measures*, vol. 493. New York, NY, USA: Wiley, 2009.

[37] D. D. Heckathorn, “Respondent-driven sampling: A new approach to the study of hidden populations,” *Soc. Problems*, vol. 44, no. 2, pp. 174–199, May 1997.

[38] S. Goel and M. J. Salganik, “Respondent-driven sampling as Markov chain Monte Carlo,” *Statist. Med.*, vol. 28, no. 17, pp. 2209–2229, Jul. 2009.

[39] A. Benveniste, M. Métivier, and P. P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York, NY, USA: Springer-Verlag, 1990.

[40] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications* (Stochastic Modeling and Applied Probability), vol. 63. New York, NY, USA: Springer-Verlag, 2010.

[41] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, vol. 282. New York, NY, USA: Wiley, 2009.

[42] H. J. Kushner and A. Shwartz, “Stochastic approximation in Hilbert space: Identification and optimization of linear continuous parameter systems,” *SIAM J. Control Optim.*, vol. 23, no. 5, pp. 774–793, 1985.

[43] G. Yin and Q. Zhang, *Discrete-Time Markov Chains: Two-Time-Scale Methods and Applications* (Applications of Mathematics), vol. 55. New York, NY, USA: Srpinger-Verlag, 2005.

[44] R. F. Curtain and A. J. Pritchard, *Infinite Dimensional Linear Systems Theory* (Lecture Notes in Control and Information Sciences), vol. 8. New York, NY, USA: Springer-Verlag, 1978.

[45] G. Yin, Y. Sun, and L. Y. Wang, “Asymptotic properties of consensus-type algorithms for networked systems with regime-switching topologies,” *Automatica*, vol. 47, no. 7, pp. 1366–1378, 2011.

Maziyar Hamdi was born in 1986. He received his B.Sc. from Sharif University of Technology, Tehran, Iran in 2007 and M.Sc. from the University of British Columbia, Vancouver, Canada in 2010, both in Electrical and Computer Engineering. He is currently working toward his Ph.D. degree at Statistical Signal Processing Lab., University of British Columbia. His current research interests include stochastic optimization, social learning, statistical signal processing on random graphs, and interactive sensing and decision making in social networks.

Vikram Krishnamurthy (S’90–M’91–SM’99–F’05) was born in 1966. He received his Ph.D. from the Australian National University, Canberra, in 1992. He is currently a professor and holds the Canada Research Chair at the Department of Electrical and Computer Engineering, University of British Columbia, Canada. His current research interests include social networks, computational game theory, stochastic optimization and scheduling. In 2009 and 2010 he served as Distinguished lecturer for the IEEE signal processing society. He has served as Editor-in-Chief of IEEE JOURNAL SELECTED TOPICS IN SIGNAL PROCESSING. In 2013 he was awarded an honorary doctorate from KTH (Royal Institute of Technology), Sweden. He is coauthor of the monograph *Interactive Sensing and Decision Making in Social Networks* published in *Foundations and Trends in Signal Processing* in 2014.

George Yin (F’02) received the Ph.D. in Applied Mathematics from Brown University and joined Wayne State University in 1987. His research interests include stochastic systems and applications. He was Co-Chair of SIAM Conference on Control and Its Application 2011, Co-organizer of 2006 IMA PI conference and 2005 IMA Workshop on Wireless Communications, and Co-Chair of 1996 AMS-SIAM Summer Seminar in Applied Mathematics and 2003 AMS-IMS-SIAM Summer Research Conference. He has been serving as Chair, Vice Chair, and Program Director of SIAM Activity Group on Control and Systems Theory in the past five years. In addition to chairing a number of award committees, he is an associate editor of *SIAM Journal on Control and Optimization*, was an associate editor of *Automatica* and IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and is on the editorial board of a number of other journals. He was president of Wayne State University’s Academy of scholars.