2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Adaptive sampling for online learning spectral properties of networks

Mohammed Abdullah, Yezekael Hayel, Alexandre Reiffers-Masson and Thierry Chonavel

Abstract— **Recently, the area of decision and control has been interested in studying the connectivity of large-scale networks. As networks under study are large, to have a complete knowledge of the network is impossible, whereas little but representative information is available with an efficient exploration scheme. Machine learning approaches were presented and used to tackle this difficulty to hold it up. In this regard, we present and prove the convergence of an efficient algorithm that converges to the Fielder vector when the topology is initially unknown and the only accessible information is gathered by a random walk process throughout the entire network. The Rayleigh quotient optimization problem and the notion of stochastic approximation are the foundations of our technique. We consider multiple sampling strategies that are categorized under random walks, as well as adapting another sampling approach that are considered random walk, the Gibbs sampling, and it showed better results. Finally, we demonstrate its performance on different network topologies.**

I. Introduction

Estimating spectrum properties of large scale graphs is a very hot topic in Computer and Network Science communities. In particular, the connectivity of large scale networks is essential for the study of their performance. Especially, in the context of overlay networks [PDL14] and ad-hoc wireless networks [SB14] such understanding is critical. Machine learning algorithms based on the Power iteration or the Rayleigh quotient techniques have been used to estimate graph spectral properties [UH12]. When the graph is unknown, these techniques have been considered recently in [ARM21] by coupling with a random walk exploration of the graph.

Due to the large size of the network, it is impossible to compute explicitly node metrics such as centrality measures. Having this difficulty in mind, in this work we propose to design an efficient effective online learning algorithm that explores a large graph and estimates at the same time a specific structural property of the underlying graph, the algebraic connectivity, also called Fiedler value [GY04]. In particular, we build a controlled random walk process on the network such that an estimator of the Fiedler vector, the eigenvector associated to the Fiedler value, is updated at each step of the stochastic process. This vector, as well as the eigenvalue associated with, has a relation with structural properties of the network. In particular, the Fiedler value is positive if and only if the graph is connected. Also, the Fiedler vector can be used to find a partition of the nodes [BMS14]. Consider a connected graph consisting of a union of categories. In some problems, only some categories are considered to estimate a particular metric. This can be done optimally using stratified sampling with full knowledge about the graph, and as the graph is explored via crawling, a method has been proposed in [KGBM11] to deal with such issues. The sampling is named Sampling Weighted Random Walk (S-WRW), as it was built by a weighted random walk that approximates the stratified sampling, starting from an ideal solution under WIS and taking into account graph exploration, then performing a simple weighted random walk and collect samples. This method was applied to the social graph of Facebook, studying a social phenomena related to college members. The authors have used S-WRW to sample 13-15 times more college members than a simple Random walk exploration. A normalized stochastic version of Oja's algorithm was presented in [BM12]. Given a connected graph, with Laplacian matrix L, a stochastic approximation scheme based on the Oja's algorithm is studied to estimate the first eigenvalues and the corresponding first eigenvectors. The convergence of this scheme has been demonstrated, and with some adaptation, it has been applied to study the Spectral Decomposition of a Markov Chain [BM12]. But this scheme and the previous ones assume a global knowledge of the entire network topology which is not the case in our approach.

mohammed.abdullah@telecom-sudparis.eu
yezekael.hayel@univ-avignon.fr
alexandre.reiffers-masson@imt-atlantique.fr
thierry.chonavel@imt-atlantique.fr

Contributions: An effective algorithm that converges to the Fielder vector is proposed in this paper, and its theoretical and analytical convergence is studied. This algorithm is related to stochastic approximation algorithms in the space of manifolds with Markovian noise, where we use the online learning technique on the Rayleigh quotient method and the proof of the convergence of our algorithm is done by the mean of the Poisson equation. Given that the information required to learn the fielder vector is obtained by a random walk. To hasten convergence, we also use a new sampling scheme driven by the spectral properties of the graph and employ other sampling distributions that all fall within the category of random walks using the spectral properties of the network. Section II introduces the model, the notations and the Rayleigh Quotient scheme which is the core of our algorithm. A stochastic approximation based algorithm is presented in section III and the main theorem that proves its convergence is given. Several sampling methods based on random walks are defined in section IV and several numerical illustrations about their performance in section V. Finally we conclude the paper in section VI.

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

II. Model

Let $G := (V, E)$ be an undirected unweighted graph, where $V := \{1, \dots, n\}$ is the set of nodes and E is the set of edges. The adjacency matrix of G is the matrix $\boldsymbol{A} := [a_{ij}]_{i,j=1:n}$, where the ij-entry of $\boldsymbol{A}$ is given by $a_{ij} := 1_{(i,j)\in E}$. The neighborhood of a node i is the set $\mathcal{N}(i) := \{j \in V \mid a_{ij} = 1\}$. We assume that for all node i, $a_{ii} = 0$. The degree of a node i is given by $deg(i) := \sum_{i'=1}^{n} a_{ii'}$ and the degree matrix is equal to $\boldsymbol{D} := [\delta_{ij} deg(i)]_{i,j=1:n}$ where $\delta_{..}$ is the Kronecker delta operator. We also define the Laplacian matrix as $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$. The spectrum of the Laplacian matrix $\boldsymbol{L}$ will be used to measure the connectivity of the graph. Let $\{\lambda_i\}_{i=1:n}$ represent the eigenvalues of $\boldsymbol{L}$, with $\lambda_1 < \lambda_2 < \lambda_3 \leq \ldots \leq \lambda_n$, and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ the corresponding unit norm eigenvectors. More precisely, we are interested in the estimation of the second smallest eigenvalue λ_2 of $\boldsymbol{L}$ denoted by λ^*. We assume that the graph is connected and therefore $\lambda_1 = 0$ and λ^* is strictly positive [VS10]. λ^* is called the algebraic connectivity (or Fiedler value) and we denote by v^* it's associated eigenvector. One well-known efficient method to compute this eigenvector is based on the Rayleigh Quotient function.

A. Raleigh Quotient

For any symmetric matrix $\boldsymbol{W}$, the Rayleigh quotient $r_{\boldsymbol{W}} : \mathbb{R}^n \to \mathbb{R}$ is the smooth function defined by:

$$r_{\boldsymbol{W}}(\boldsymbol{x}) \quad = \quad \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} \|\boldsymbol{x}\|^{-2}. \tag{1}$$

Let $U^{n-1} := \{\boldsymbol{x} \in \mathbb{R}^n, |\|\boldsymbol{x}\| = 1\}$ be the unit sphere in $\mathbb{R}^n$. Then by considering the restriction of the Rayleigh quotient on the unit sphere U^{n-1}, the eigenvalues of a real symmetric matrix are characterized using the Rayleigh Quotient [UH12] and in particular:

$$\lambda_{\min}(\boldsymbol{W}) = \min_{\boldsymbol{x} \in U^{n-1}} r_{\boldsymbol{W}}(\boldsymbol{x}).$$

Therefore, the smallest eigenvalue of $\boldsymbol{W}$ can be obtained by solving an optimization problem on the manifold U^{n-1}. One approach to solve this optimization problem is to apply the gradient flow in order to search for the dominant eigenvector of $\boldsymbol{W}$. This leads to the well known Rayleight Quotient Gradient Flow defined as the following ordinary differential equations:

$$\dot{\boldsymbol{x}} = \frac{h(x(t))}{||x(t)||^2},\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \in U^{n-1}, \tag{2}$$

with $h(\boldsymbol{x}) = -(\boldsymbol{W} - r_{\boldsymbol{W}}(\boldsymbol{x}) I_n)\boldsymbol{x}$. It can be proved that under specific conditions ($\lambda_{\min} < \lambda_2$, $\boldsymbol{W}$ is a symmetric matrix) that $\boldsymbol{x}$ converges to the eigenvector associated to the eigenvalue $\lambda_{\min}$ (see [UH12]).

B. Space reduction

Recall that $\lambda_1 = 0$ and $\boldsymbol{v}_1 = \frac{1}{\sqrt{n}} \mathbb{1}$, with $\mathbb{1} = [1, \dots, 1]^T$. Then, tracking $\boldsymbol{v}_2$ can be done equivalently by tracking the eigenvector associated with the largest eigenvalue of the matrix $\boldsymbol{I} - \frac{1}{n}\mathbb{1}\mathbb{1}^T - \epsilon \boldsymbol{L}$ for $\epsilon < \lambda_n^{-1}$. This method is called deflation [PDL14]. When λ_n is large this could lead to numerical issues. Then a reduction technique is proposed in [ARM21] where instead of looking at the second smallest eigenvalue of $\boldsymbol{L}$, the problem is to find the smallest eigenvalue of the matrix $\boldsymbol{S}$ such that

$$\boldsymbol{S} = \boldsymbol{Q}^T \boldsymbol{L} \boldsymbol{Q},$$

with $\boldsymbol{Q} = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_{n-1}] \in \mathbb{R}^{n \times n-1}$ and

$$\boldsymbol{q}_k = \frac{1}{\sqrt{k(k+1)}}[1, \dots, 1, -k, 0, \dots, 0]^T$$

(k entries are equal to $\sqrt{k(k+1)}^{-1}$). The eigenvalues of matrix $\boldsymbol{S}$ are $\lambda_2 < \lambda_3 \leq \ldots \leq \lambda_n$ and corresponding unit norm eigenvectors $(\boldsymbol{w}_2, , \dots, \boldsymbol{w}_n) := (\boldsymbol{Q}^T \boldsymbol{v}_2, \dots, \boldsymbol{Q}^T \boldsymbol{v}_n)$. Then the Rayleight quotient method is applied directly on matrix $\boldsymbol{S}$.

III. Stochastic approximation based algorithm

In this work, a discrete time random walk Y_k is used to explore the graph step-by-step. This is what we call Markovian sampling. At each time step k, the stochastic process Y is in vertex $i_k \in V$ and the neighboring nodes $\mathcal{N}(i_k)$ are observable. Then, after the k-th step of the random walk process, the σ-algebra $\mathcal{H}(k)$ is given by $\mathcal{H}(k) := \{i_1, \mathcal{N}(i_1), \dots, i_k, \mathcal{N}(i_k)\}$. At instants k, line i_k of $\boldsymbol{L}$ is observed and denoted by l_{i_k}. Then matrix $\boldsymbol{L}$ is replaced by the following instantaneous matrix $\boldsymbol{L}_k = (\boldsymbol{0} \,|\, \boldsymbol{l}_{i_k} \,|\, \boldsymbol{0})^T$, where only line i_k of $\boldsymbol{L}_k$ is nonzero. Then, the instantaneous version of matrix $\boldsymbol{S}$ is $\boldsymbol{S}_k = \boldsymbol{Q}^T \boldsymbol{L}_k \boldsymbol{Q}$. Note that the stochastic process Y_k is a Markov process and we denote by $q_{\boldsymbol{x}}(i_{k+1}, i_k) := P(Y_{k+1} = i_{k+1} | Y_k = i_k, \boldsymbol{x})$ its transition kernel. This transition kernel is state dependent, it means that the transition probability of the stochastic process Y can depend on current estimator vector $\boldsymbol{x}$. Under a specific exploration scheme u, note that the Markov process Y is assumed to be ergodic and the control is such that the Markov process is irreducible and aperiodic. Then there is a unique stationary distribution denoted by Γ_u.

Uniform Random walk, or any other types of exploration scheme u induced a bias in the discovery process. If the bias is known it can be corrected by an appropriate re-weighting of the measured vector using Hansen-Hurwitz estimator [KGBM11]. Since we have that $\boldsymbol{S} = \sum_{i=i}^{k} \boldsymbol{S}_i$, for any time step k, the Hansen-Hurwitz estimator will be for any exploration scheme u:

$$\boldsymbol{S}' = \frac{1}{k} \sum_{i=1}^{k} \frac{\boldsymbol{S}_i}{\Gamma_u(i)}$$

where $\Gamma_u(i)$ is the probability of visiting node i in the stationary distribution using exploration scheme u and this estimator is unbiased. Thus in order to correct the bias it is sufficient to divide the stochastic approximation scheme by the stationary distribution $\Gamma_u(i)$ when the exploration process is at node i.

The Rayleigh quotient method states that the minimum the Rayleigh quotient function corresponds to the smallest eigenvector of the matrix. Moreover, as the observation of the

matrix is noisy, a stochastic approximation (SA) algorithm can be used to update the estimator of the smallest eigenvector $\boldsymbol{v}_2$. Considering the retraction function $R_x(y) := \frac{x+y}{||x+y||}$, our SA scheme can be written as:

$$\boldsymbol{x}_{k+1} = R_{x_k}\left(\epsilon_k \frac{H(\boldsymbol{x}_k, Y_k)}{\Gamma_u(Y_k)}\right),$$

with $H(\boldsymbol{x}_k, Y_k) = -(\boldsymbol{S}_k - r_{\boldsymbol{S}_k}(\boldsymbol{x}_k)I_n)\boldsymbol{x}_k$, $\boldsymbol{x}_0$ is the initial point, $\boldsymbol{x}_k$ are the iterates and ϵ_k is the kth step-size. Note that our framework corresponds to a Stochastic Approximation scheme with Markovian noise which lives on the unit sphere manifold. Proofs of convergence under specific assumptions for the SA scheme with Markovian noise but not on a manifold and without Markovian noise on manifold have been done in recent works [CLP13] and [Sha21] respectively. But the proof of convergence of this SA scheme with Markovian noise and on a manifold has never been proved. That is the aim of the following theorem. First, the following conditions have to be satisfied:

1) **Step-size:**The step size $\epsilon_k = \frac{1}{k}$ verifies $\sum_{k=0}^{\infty} \epsilon_k = \infty$ and $\sum_{k=0}^{\infty} \epsilon_k^2 < \infty$ by using the p-series test it states that a series of the form $\sum(\frac{1}{n^p})$ converges if $p > 1$ and diverges if $p \leq 1$.
2) **Stationary Distribution of Markov chain:** For every $\boldsymbol{x} \in U^{n-1}$, we assume that the Markov chain generated by the transition matrix $\pi(\boldsymbol{x}) := [[\pi(i, j, \boldsymbol{x})]]_{i,j}$, where $\mathbb{P}(Y_{+1} = j \mid Y_k = i, \boldsymbol{x}) =: \pi(i, j, \boldsymbol{x})$ is an irreducible and aperiodic Markov chain. The unique stationary distribution to $\pi(\boldsymbol{x})$ is denoted by Γ_u where u describe the exploration scheme.
3) **Poisson equation:** We assume that for every $\boldsymbol{x}$, the solutions of the Poisson equation given by:
$$(1 - \pi_{i,\boldsymbol{x}})v_{\boldsymbol{x}}(\cdot) = H_1(\boldsymbol{x}, \cdot) - \textstyle\sum_j H_1(\boldsymbol{x}, j)\Gamma_u(i)$$
$$\text{where } \pi_{i,\boldsymbol{x}} v_{\boldsymbol{x}}(\cdot) = \textstyle\sum_j v_{\boldsymbol{x}}(j)\pi(i, j, \boldsymbol{x})$$
are such that $\sup \|v_{\boldsymbol{x}}(y)\| \leq C$ with C independent of y and $v.(y)$ is a smooth vector filed with U^{n-1} ($\|(v_{\boldsymbol{x}}(y)) - v_{\boldsymbol{x}'}(y)\| \leq L' d(\boldsymbol{x}, \boldsymbol{x}')$ for some $L' > 0$).

Theorem 1. *If the assumptions above mentioned are satisfied, then, for any exploration scheme u, the following stochastic approximation scheme:*

$$\boldsymbol{x}_{k+1} = R_{\boldsymbol{x}_k}(\epsilon_k H_1(\boldsymbol{x}_k, Y_k)), \tag{3}$$

with $H_1(\boldsymbol{x}_k, Y_k) = \frac{H(\boldsymbol{x}_k, Y_k)}{\Gamma_u(Y_k)}$ converges as $k \to +\infty$ to the eigenvector $\boldsymbol{w}_2$ associated to the smallest eigenvalue of $\boldsymbol{S}$.

Proof: The main result of the proof is to prove that the stochastic approximation scheme is behaving asymptotically as a reparametrized version of the Rayleigh quotient flow. Once we have been able to prove this fact, then the conclusion is directly from the fact that the Rayleigh quotient is converging to the Fielder vector, in our case.

Note that $\|\boldsymbol{x}_k\| < \infty$ for all every k because $\boldsymbol{x}_k \in U^{n-1}$. To simplify the proof, we will assume that $\boldsymbol{x}_{k+1} \in \tilde{M} \subset U^{n-1} \setminus \{(-1, 0, \cdots, 0)\}$, with $\tilde{M}$ being compact. If we want to relax this assumption, we should construct the same atlas as the one proposed in [Sha21]. We assume that the bijection $\Psi : \tilde{M} \to \mathbb{R}^{n-1}$ associated with $\tilde{M}$ is given by the stereographic projection $\psi(\boldsymbol{x}) = \frac{\boldsymbol{x}_{-1}}{1+\boldsymbol{x}_1}$, where $\boldsymbol{x}_{-1} := [x_2, \ldots, x_n]$. Note that $D\psi(x)$ is bounded for all $\boldsymbol{x} \in \tilde{M}$.

(Step 1) The first step of the proof is to rewrite our stochastic approximation using the Poisson equation: The Poisson equation ($H_1(\boldsymbol{x}_k, y_k) - h_1(\boldsymbol{x}_k) = v_{\boldsymbol{x}_k}(y_k) - \sum_{y'} v_{\boldsymbol{x}_k}(y')\pi_{\boldsymbol{x}_k}(y_k, y')$) implies that

$$\begin{aligned} H_1(\boldsymbol{x}_k, y_k) - h_1(\boldsymbol{x}_k) &= v_{\boldsymbol{x}_k}(y_k) - \pi_{\boldsymbol{x}_k} v_{\boldsymbol{x}_k}(y_k) \\ &= \underbrace{v_{\boldsymbol{x}_k}(y_{k+1}) - \pi_{\boldsymbol{x}_k} v_{\boldsymbol{x}_k}(y_k)}_{L^1_{k+1}} \\ &+ \underbrace{v_{\boldsymbol{x}_k}(y_k) - v_{\boldsymbol{x}_{k+1}}(y_{k+1})}_{L^2_{k+1}} \\ &+ \underbrace{v_{\boldsymbol{x}_{k+1}}(y_{k+1}) - v_{\boldsymbol{x}_k}(y_{k+1})}_{L^3_{k+1}}. \end{aligned}$$

Therefore, our stochastic approximation scheme can be rewritten as:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \mathcal{R}_{\boldsymbol{x}_k}[\epsilon_k H_1(\boldsymbol{x}_k, y_k)] \\ &= \mathcal{R}_{\boldsymbol{x}_k}[\epsilon_k h_1(\boldsymbol{x}_k) + \epsilon_k(H_1(\boldsymbol{x}_k, y_k) - h_1(\boldsymbol{x}_k, y_k))] \\ &= \mathcal{R}_{\boldsymbol{x}_k}[\epsilon_k h_1(\boldsymbol{x}_k) + \epsilon_k(L^1_{k+1} + L^2_{k+1} + L^3_{k+1})]. \end{aligned}$$

(Step 2) Local parametrization of the stochastic approximation and the O.D.E: In this second step, using the local parametrization $\Psi(\cdot)$, we study the error between the O.D.E and the stochastic approximation in $\mathbb{R}^{n-1}$ and not in $\tilde{M}$. We define $\hat{\boldsymbol{x}}_k = \psi(\boldsymbol{x}_k)$, $\hat{h}_1(\hat{\boldsymbol{x}}_k) = D\psi(\boldsymbol{x}_k)[h_1(\boldsymbol{x}_k)]$ and $\hat{L}^i_{k+1} = D\psi(\boldsymbol{x}_k)[L^i_{k+1}]$. Let us also define $t(k) := \sum_{k'=0}^{k-1} \epsilon_{k'}$ and $\overline{\boldsymbol{x}}(t)$ the interpolated version of $\hat{\boldsymbol{x}}_k$ (i.e. $\overline{\boldsymbol{x}}(t(k)) = \hat{\boldsymbol{x}}_k$ and $\overline{\boldsymbol{x}}$ is linear by part. We now consider the ODE on the manifold expressed using the local parametrization $\hat{\boldsymbol{x}}^{t_k}(t) = \psi(\boldsymbol{x}^{t_k}(t))$, where $\hat{\boldsymbol{x}}^{t_k}(t_k) = \hat{\boldsymbol{x}}_k$ and $\hat{\boldsymbol{x}}^{t_k}(t)$ is a solution of $\dot{\boldsymbol{x}} = h_1(\boldsymbol{x}(t))$. Note that we have $\dot{\hat{\boldsymbol{x}}} = \hat{h}_1(\hat{\boldsymbol{x}})$.

$$\hat{\boldsymbol{x}}_{k+1} = \psi(\mathcal{R}_{\boldsymbol{x}_k}[\epsilon_k(h_1(\boldsymbol{x}_k) + L^1_{k+1} + L^2_{k+1} + L^3_{k+1})])$$

By using the local rigidity property of a retraction, the fact that $\mathcal{R}_{\boldsymbol{x}}(0_{\boldsymbol{x}}) = \boldsymbol{x}$ and the linearity of $D\psi(\boldsymbol{x})$ we obtain, using a simple Taylor approximation we obtain that:

$$\begin{aligned} &\psi(\mathcal{R}_{\boldsymbol{x}_k}[\epsilon_k(h_1(\boldsymbol{x}_k) + L^1_{k+1} + L^2_{k+1} + L^3_{k+1})]) \\ &= \hat{\boldsymbol{x}}_k + \epsilon_k \hat{h}_1(\hat{\boldsymbol{x}}_k) + \epsilon_k(\hat{L}^1_{k+1} + \hat{L}^3_{k+1} + \hat{L}^3_{k+1}) + \mathcal{O}(\epsilon_k). \end{aligned}$$

We can now write $\hat{\boldsymbol{x}}_{k+m}$ as follows:

$$\hat{\boldsymbol{x}}_{k+m} = \hat{\boldsymbol{x}}_k + \sum_{u=k}^{m-1}(\epsilon_u \hat{h}_1(\hat{\boldsymbol{x}}_u)) + \delta_{k,u} + \mathcal{O}(\epsilon_k^2), \tag{4}$$

where:

$$\delta_{k,k+m} = \underbrace{\sum_{u=k}^{m-1} \epsilon_u \hat{L}^1_{u+1}}_{\delta^1_{k,k+m}} + \underbrace{\sum_{u=k}^{m-1} \hat{L}^2_{u+1}}_{\delta^2_{k,k+m}} + \underbrace{\sum_{u=k}^{m-1} \epsilon_u \hat{L}^3_{u+1}}_{\delta^3_{k,k+m}} \tag{5}$$

Note that the error term is easy to $\mathcal{O}(\epsilon_k^2)$ because converging to 0 when k is going to infinity. Let $t_{k+m} \in [t_k, t_k + T]$ for all $T > 0$. We are interested to study the behavior of $\sup_{t\in[s,s+T]} \|\overline{\boldsymbol{x}}(t) - \hat{\boldsymbol{x}}^{t_k}(t)\|$ for all T when s tends to infinity. We can study such term by simply restricting the proof to the $\{t_k\}$. The error due to the linear interpolation is easy to handle by following the same steps as in [Sha21]. For all k, we have by definition:

$$\begin{aligned}
&\|\overline{\boldsymbol{x}}(t) - \hat{\boldsymbol{x}}^{t_k}(t_{k+m})\| \leq \|\hat{\boldsymbol{x}}_k - \hat{\boldsymbol{x}}^{t_k}(t_k)\| \\
&+\|\delta^1_{k,k+m}\| + \|\delta^2_{k,k+m}\| + \|\delta^3_{k,k+m}\| \\
&+ \int_{t_k}^{t_{k+k}} \|\hat{h}_1(\hat{\boldsymbol{x}}^{t_k}(t)) - \hat{h}_1(\hat{\boldsymbol{x}}^{t_k}([t]))\|dt \\
&+ \sum_{u=k}^{m-1} \epsilon_u \|\hat{h}_1(\hat{\boldsymbol{x}}_u) - \hat{h}_1(\hat{\boldsymbol{x}}^{t_k}(t_u))\|
\end{aligned}$$

In the rest of the proof we will focus on studying the convergence of the stochastic term $\delta^i_{k,k+m}$ when k tends to infinity. The rest of the proof, including of the convergence of the interpolation term is standard $\int_{t_k}^{t_{k+k}} \|\hat{h}_1(\hat{\boldsymbol{x}}^{t_k}(t)) - \hat{h}_1(\hat{\boldsymbol{x}}^{t_k}([t]))\|dt$ and can be found in [Sha21].

(Step 3) Convergence of the stochastic terms:

$$\begin{aligned}
\delta^1_{k,k+m} &= \sum_{u=k}^{m-1} \epsilon_u D\psi_n(\boldsymbol{x}_u)[L^1_{u+1}] \\
&= \sum_{u=k}^{m-1} \epsilon_u D\psi_u(\boldsymbol{x}_u)[v_{\boldsymbol{x}_u}(y_{u+1}) \\
&\quad -\pi_{\boldsymbol{x}_u} v_{\boldsymbol{x}_u}(y_u)]
\end{aligned}$$

We can notice that $v_{\boldsymbol{x}_k}(y_{k+1}) - \pi_{\boldsymbol{x}_k} v_{\boldsymbol{x}_k}(y_k)$ is a zero mean martingale with respect to the filtration $\mathcal{F}_k$ and by the linearity of $D\psi(.)$ we get that $\hat{L}^1_{u+1}$ is martingale difference sequence. Also, we have $\|v_{\boldsymbol{x}_k}(y_{k+1}) - v_{\boldsymbol{x}_k}(y_k)\| \leq \|v_{\boldsymbol{x}_k}(y_{k+1})\| + \|v_{\boldsymbol{x}_k}(y_k)\|$ and $||v_{\boldsymbol{x}_k}(y_k)|| \leq C$. This implies that $\mathbb{E}[\|D\psi(\boldsymbol{x}_k)L^1_{k+1}\|^2|\mathcal{F}_k] \leq \|D\psi(\boldsymbol{x}_k)\|^2\mathbb{E}[\|L^1_{k+1}\|^2|\mathcal{F}_k] < \infty$. Hence, using the martingale convergence theorem and the fact that the step-size are square summable, we get that $\lim_{k\to\infty} \sup_m \delta^1_{k,k+m} = 0 \ \ a.s.$

One can write $D\psi(\boldsymbol{x}) =: A(\boldsymbol{x})$ as a matrix depending on $\boldsymbol{x}$. We get the following following bound on $\delta^2_{k,k+m}$:

$$\begin{aligned}
\|\delta^2_{k,k+m}\| &= \|\sum_{u=k}^{m-1} \epsilon_u A(\boldsymbol{x}_u)[L^2_{u+1}] + \mathcal{O}(\epsilon_{k+m-1})\| \\
&= \|\sum_{u=k}^{m-1} \epsilon_u A(\boldsymbol{x}_u)[v_{\boldsymbol{x}_u}(y_u) - (v_{\boldsymbol{x}_{u+1}}(y_{u+1}))\| \\
&\leq \epsilon_k \|A(\boldsymbol{x}_k)[v_{\boldsymbol{x}_k}(y_k)]\| \\
&\quad +\epsilon_{m+1}\|A(\boldsymbol{x}_{m-1})v_{\boldsymbol{x}_m}(y_m)\| \\
&\quad +\|\sum_{u=1}^{m-1}(\epsilon_u A(\boldsymbol{x}_u) - \epsilon_{u-1}A(\boldsymbol{x}_{u-1})[v_{\boldsymbol{x}_u}(y_u)]\|.
\end{aligned}$$

From the fact that $\|A(\boldsymbol{x})[v_{\boldsymbol{x}}(y)]\|$ is bounded we get that $\epsilon_k\|A(\boldsymbol{x}_k)[v_{\boldsymbol{x}_k}(y_k)]\| + \epsilon_{m+1}\|A(\boldsymbol{x}_{m-1})v_{\boldsymbol{x}_m}(y_m)\| \xrightarrow[k\to\infty]{} 0$.

Moreover, using the Taylor expansion and the fact that $\boldsymbol{x}_u = \boldsymbol{x}_{u-1} - \epsilon_{u-1}H_1(\boldsymbol{x}_{u-1}, y_{u-1})$ we get that: $A(\boldsymbol{x}_u) = A(\boldsymbol{x}_{u-1}) - \epsilon_{u-1}\nabla A(\boldsymbol{x}_{u-1})H_1(\boldsymbol{x}_{u-1}, y_{u-1}) + \mathcal{O}(\epsilon^2_{u-1})$

Let denote the last term $\delta^{2'}_{n,n+m}$, we obtain the following upper bound:

$$\begin{aligned}
\delta^{2'}{}_{k,k+m} &= \|\sum_{u=1}^{m-1}(\epsilon_u - \epsilon_{u-1}A(\boldsymbol{x}_u)[v_{\boldsymbol{x}_u}(y_u)] \\
&\quad -\epsilon_u\epsilon_{u-1}\nabla A(\boldsymbol{x}_{u-1})H_1(\boldsymbol{x}_{u-1}, y_{u-1})\| \\
&\leq \sum_{u=1}^{m-1}(\epsilon_u - \epsilon_{u-1})\|A(\boldsymbol{x}_u)[v_{\boldsymbol{x}_u}(y_u)]\| \\
&\quad +\epsilon_u\epsilon_{u-1}\|\nabla A(\boldsymbol{x}_{u-1})H_1(\boldsymbol{x}_{u-1}, y_{u-1})\| \\
&\leq \sum_{u=1}^{m-1}(\epsilon_u - \epsilon_{u-1})\|A(\boldsymbol{x}_u)[v_{\boldsymbol{x}_u}(y_u)]\| \\
&\quad +\epsilon^2_u\|\nabla A(\boldsymbol{x}_{u-1})H_1(\boldsymbol{x}_{u-1}, y_{u-1})\|
\end{aligned}$$

And all these terms we converge to zero as k approaches infinity. thus we get the following result:

$$\|\delta^2_{k,k+,m}\| \leq \sum_{u=k}^{m-1} \epsilon^2_u Cst \xrightarrow[k\to\infty]{} 0$$

Finally, by the fact that $v.(y)$ is L'-Lipschitz we get:

$$\begin{aligned}
\|\delta^3_{k,k+m}\| &= \|\sum_{u=k}^{m-1} \epsilon_u A(\boldsymbol{x}_u)[(v_{\boldsymbol{x}_{u+1}}(y_{u+1})) - v_{\boldsymbol{x}_u}(y_{u+1})]\| \\
&= \sum_{u=k}^{m-1} \epsilon_u \|A(\boldsymbol{x}_u)\|\|(v_{\boldsymbol{x}_{u+1}}(y_{u+1})) - v_{\boldsymbol{x}_u}(y_{u+1})\| \\
&\leq \sum_{u=k}^{m-1} L'\epsilon_u \|A(\boldsymbol{x}_u)\| d(\boldsymbol{x}_{u+1}, \boldsymbol{x}_u) \leq C\epsilon_u,
\end{aligned}$$

where the last inequality is coming the fact that H_1 is a smooth vector field. We get the following result:

$$\|\delta^3_{n,n+,m}\| \leq \sum_{u=n}^{m-1} \epsilon^2_u Cst \xrightarrow[n\to\infty]{} 0.$$

■

The previous theorem shows that any exploration scheme u can be used to estimate the smallest eigenvector $\boldsymbol{w}_2 := \boldsymbol{Q}^T\boldsymbol{v}_2$ of matrix $\boldsymbol{S}$. Then, the transformation $\boldsymbol{v}_2 = (Q^T)^{-1}\boldsymbol{w}_2$ is used to obtain the fielder vector of matrix $\boldsymbol{L}$. The next section illustrates the stochastic approximation scheme considering diverse sampling methods u.

IV. Sampling methods

Three different sampling methods are described. The first one is the only naive one as the next step of the random walk is totally random on the connected nodes and does not consider the exploration process so far. The two other random walk schemes are smarter as they have specific properties depending on the values of the indices of the Fielder vector explored so far that hold the properties of the graph.

A. Uncontrolled Markovian sampling

The first sampling studied in this paper is the uncontrolled Markovian sampling. In this context, we assume that the network is explored using a standard random walk. Let $\{Y_k\}$ be a finite state space Markov chain that captures which node is observed at time k. The transition matrix associated to this Markov chain is given by:

$$\mathbb{P}(Y_{k+1} = j \mid Y_k = i) := \begin{cases} \frac{1}{\mid N(i) \mid}, \text{ if } j \in N(i) \\ 0, \text{ otherwise.} \end{cases}$$

In this case, the stationary distribution is given by

$$\Gamma_u(i) = \frac{|N(i)|}{\sum_{j=1} |N(j)|}, \ \forall i.$$

This exploration scheme is interesting as such, even if it is not driven by the current estimate of the Fiedler vector. Indeed, random walks sampling schemes are a fundamental in networks applications and distributed systems, due their local and lightweight nature (see [SMP15], [DSNPT13]). Moreover random walk are naturally designed to handle dynamic networks. The uniform random walk may be stuck in one cluster and stays there without leaving for a large number of iterations and thus losing a lot of information. Given this and the bias generated by this sample, the uniform random walk is not an efficient sample distribution that can be used for all types of networks. This keeps us with introducing another Markov chain distributions.

B. Metropolis-Hastings sampling

The aim is to build a particular sampling such that each node is sampled the same number of times in average. Such a sample is well-known in Markov theory and can be achieved via the Metropolis-Hastings algorithm [Hit03]. With such exploration process, the stationary distribution should verify:

$$\Gamma_u(i) = \frac{1}{N}, \ \forall i.$$

The Metropolis-Hastings algorithm gives the following random walk transition probabilities:

$$\mathbb{P}(Y_{k+1} = j \mid Y_k = i) := \begin{cases} 0, \text{ if } j \notin N(i), \\ \frac{1}{|N(i)|} \min\left(\frac{|N(i)|}{|N(j)|}, 1\right), \text{ if } j \in N(i) \\ 1 - \sum_{j \neq i} \mathbb{P}(Y_{k+1} = j \mid Y_k = i), \text{ if } j = i \end{cases}$$

The main interesting property of this exploration scheme is that each node is visited in average the same number of times and thus there is a limited risk that the exploration process stays very large in a particular region of the network. However, this scheme tends to visit all nodes evenly, even those that are very isolated, and thus not very important for global connectivity.

C. Gibbs sampling

Inspired by [BMS14], we propose a random walk sampling algorithm driven by the current estimate of $\boldsymbol{x}_k$ of the Fiedler vector. To mimic the set-up proposed in [BMS14], we first define a function $\Psi : V \times \mathbb{R}^n \to \mathbb{R}$ which associates to every node i, and for a given $\boldsymbol{x}$ a value $\Psi(i, \boldsymbol{x})$ defined as follow:

$$\Psi(i, \boldsymbol{x}) := -\frac{1}{N(i)} \sum_{j \in N(i)} (x_i - x_j)^2. \tag{6}$$

If we assume that $\boldsymbol{x}$ is the true Fiedler vector, then the node with the lowest $\Psi(i, \boldsymbol{x})$ will be the one with, on average, the highest difference of its own value x_i, with respect to its neighbors. Of course other definitions of $\Psi(i, \boldsymbol{x})$ can be considered. The goal of the random walk designed in [BMS14] is to have a stationary distribution that is peaking to the argmin of $\Psi(\cdot, \boldsymbol{x})$. Coming back to our case, this would mean that we will be able to design a random walk that will focus on nodes with the lowest value of $\Psi(i, \boldsymbol{x})$. As a result of the fact that nodes that create links between two or more clusters will have the lowest value of the function ψ, we will enhance the likelihood of visiting nodes that connect the graph's clusters, increasing the likelihood of moving from one cluster to another and thus not being get stuck in one cluster for a long time. The online algorithm can visit all of them and quickly learn more about the characteristics of the graph. We assume that the transition probability of the Markov chain $\{Y_k\}$ is given by:

$$\mathbb{P}(Y_{k+1} = j \mid Y_k = i) := \begin{cases} \frac{1}{\mid N(i) \mid} e^{-[\Psi(j, \boldsymbol{x}_k) - \Psi(i, \boldsymbol{x}_k)]_+} \text{ if } j \in N(i) \\ 1 - \frac{1}{\mid N(i) \mid} \sum_j e^{-[\Psi(j, \boldsymbol{x}_k) - \Psi(i, \boldsymbol{x}_k)]_+} \text{ if } j = i \\ 0, \text{ otherwise,} \end{cases}$$

with $[x]_+ := \max(x, 0)$. This Markov chain has the unique stationary distribution:

$$\Gamma_u(i) = \frac{|N(i)|}{Z} e^{\sum_j (x_i - x_j)^2},$$

where $Z = \sum_{j=1}^{I} |N(j)| e^{\sum_j (x_i - x_j)^2}$ is a normalization factor. This exploration scheme is designed to explore the network based on the current estimation of the connectivity measure of each node. We next illustrate the performance of the different schemes on two types of network topology.

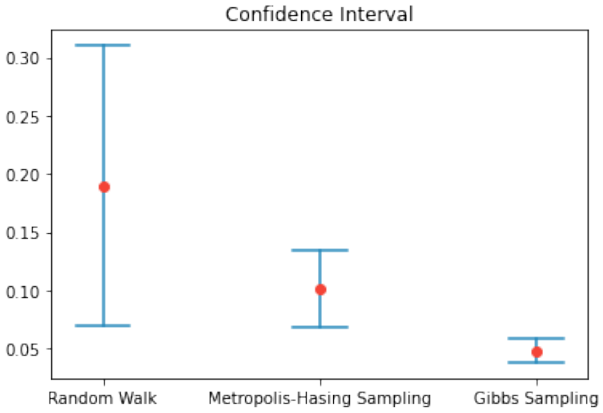
V. Numerical illustrations

We select two real-world networks and a random social one for our evaluation. The networks are Karate, Open-Street and a random one, having from 33 to 179 nodes and from 156 to 860 edges. The numerical experiments done here are based on using the stochastic approximation scheme with different Markovian sampling: Uniform Random walk, Metropolis-Hastings, and Gibbs samplings. The results are compared with the exact solution computed offline. The Laplacian matrix (only one row can be observed at each time slot $k \in \mathbb{N}$), the sampling distribution, the number of iterations, and the step-size $\epsilon_k = \frac{1}{k+1}$.

Network	Social	Geometric	karate
Number of nodes	50	200	33
Number of edges	600	860	156

A. Convergence of the scheme

In the first numerical experiment, we aim to study the convergence of the stochastic approximation scheme for all Markovian samples, as well as plotting the confidence intervals for these samples. To plot the confidence interval, 10 simulations were performed for each sampling distribution with 5×10^5 iterations. We restrict this study to the social network. In the following figure, the x-axis represents the sampling distribution used, and the y-axis represents the range of the error estimated by the stochastic scheme using the sampling distributions after 5×10^5.



B. Extension to other types of network

The second numerical study applies the same scheme to other networks and compares the performance there. Two important issues are considered in this context: the effect of using a different Markov chain distribution and its performance on different graphs. This keeps us with 9 simulations (3 networks and 3 sampling distributions). The error in each simulation is given after 10^5 and 10^6 iterations, for better comparison.

TABLE I

ERROR NORM CALCULATED AFTER 10^5 AND 10^6 ITERATIONS USING DIFFERENT SAMPLING TECHNIQUES ON DIFFERENT NETWORKS

Networks		Social	Geometric	karate
Random Walk	10^5 It	0.3	0.95	0.22
	10^6 It	0.15	0.9	0.1
Metropolis-Hastings	10^5 It	0.4	0.45	0.06
	10^6 It	0.1	0.25	0.02
Gibbs	10^5 It	0.16	0.35	0.035
	10^6 It	0.02	0.2	0.005

It is clear that the error varies when different sampling distributions are used, all of which are categorized as random walks. According to the table above, using the Gibbs distribution is the best among all, demonstrating its efficiency on various types of networks, including networks without clusters (Karate network), networks with 2-4 clusters (Social network), and networks with more than 5 clusters (Geometric network), helping thus to reduce the time of convergence.

VI. CONCLUSIONS AND PERSPECTIVES

In this paper, we have proposed an efficient exploration of random walks in order to estimate online the global connectivity of large-scale networks. Our main result is the proof of the convergence of our scheme using stochastic approximation techniques over manifolds. We have shown numerically that different sampling methods (uniform, Metropolis-Hastings, and Gibbs) lead to different performances in terms of convergence time. Then, a first natural extension of this work is to study in detail the convergence time of our exploration process depending on the scheme used and the type of network under study. Another interesting aspect is to consider a dynamic network in which nodes and/or links appear and disappear in a dynamic fashion. Therefore it could be very interesting to understand how our scheme can track a global connectivity measure like the algebraic connectivity for such a dynamic network. Finally, another metric like the Ricci curvature of the network can be estimated instead of the global connectivity.

REFERENCES

[ARM21] Y. Hayel A. Reiffers-Masson, T. Chonavel. Estimating fiedler value on large networks based on random walk observations. In *in proceedings of ICASSP*, 2021.

[BM12] V. Borkar and S.P. Meyn. Oja's algorithm for graph clustering, markov spectral decomposition, and risk sensitive control. *Automatica*, 48(10):2512–2519, 2012.

[BMS14] Vivek S Borkar, Rahul Makhijani, and Rajesh Sundaresan. Asynchronous gossip for averaging and spectral ranking. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):703–716, 2014.

[CLP13] Richard Combes, Jie Lu, and Alexandre Proutière. An introduction to stochastic approximation. *Literature*, pages 1–9, 2013.

[DSNPT13] Atish Das Sarma, Danupon Nanongkai, Gopal Pandurangan, and Prasad Tetali. Distributed random walks. *Journal of the ACM (JACM)*, 60(1):1–31, 2013.

[GY04] L. Gross and J. Yellen. *Handbook of Graph Theory*. 2004.

[Hit03] David B Hitchcock. A history of the metropolis–hastings algorithm. *The American Statistician*, 57(4):254–257, 2003.

[KGBM11] Maciej Kurant, Minas Gjoka, Carter T Butts, and Athina Markopoulou. Walking on a graph with a magnifying glass: stratified sampling via weighted random walks. pages 281–292, 2011.

[PDL14] S. Barbarossa P. Di Lorenzo. Distributed estimation and control of algebraic connectivity over random graphs. *IEEE Transactions on Signal Processing*, 62(21):5615–5628, 2014.

[SB14] P. Di Lorenzo S. Barbarossa, S. Sardellitti. Distributed detection and estimation in wireless sensor networks. *Academic Press Library in Signal Processing*, 2:329–408, 2014.

[Sha21] Suhail M Shah. Stochastic approximation on riemannian manifolds. *Applied Mathematics & Optimization*, 83(2):1123–1151, 2021.

[SMP15] Atish Das Sarma, Anisur Rahaman Molla, and Gopal Pandurangan. Distributed computation in dynamic networks via random walks. *Theoretical Computer Science*, 581:45–66, 2015.

[UH12] J. B Moore U. Helmke. Optimization and dynamical systems. 2012.

[VS10] Maarten Van Steen. Graph theory and complex networks. *An introduction*, 144, 2010.