Spectral norm bounds for block Markov chain random matrices

Albert Senen–Cerda, Jaron Sanders

Department of Mathematics and Computer Science

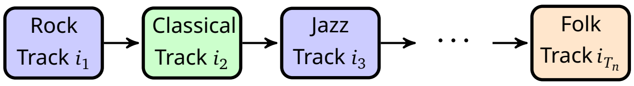
arXiv:2111.06201 [math.PR]

Abstract: This paper quantifies the asymptotic order of the largest singular value of a centered random matrix built from the path of a Block Markov Chain (BMC) [5]. Given a path $X_0, X_1, \dots, X_{T_n}$ started from equilibrium, we construct a random matrix $\hat{N}$ that records the number of transitions between each pair of states. We prove that if $\omega(n) = T_n = o(n^2)$, then $\|\hat{N} - \mathbb{E}[\hat{N}]\| = \Omega_{\mathbb{P}}(\sqrt{T_n/n})$. We also prove that if $T_n = \Omega(n \ln n)$, then $\|\hat{N} - \mathbb{E}[\hat{N}]\| = O_{\mathbb{P}}(\sqrt{T_n/n})$ as $n \to \infty$; and if $T_n = \omega(n)$, a sparser regime, then $\|\hat{N}_\Gamma - \mathbb{E}[\hat{N}]\| = O_{\mathbb{P}}(\sqrt{T_n/n})$. Here, $\hat{N}_\Gamma$ is a regularization that zeroes out entries corresponding to jumps to and from most-often visited states. Together this establishes that the order is $\Theta_{\mathbb{P}}(\sqrt{T_n/n})$ for BMCs.

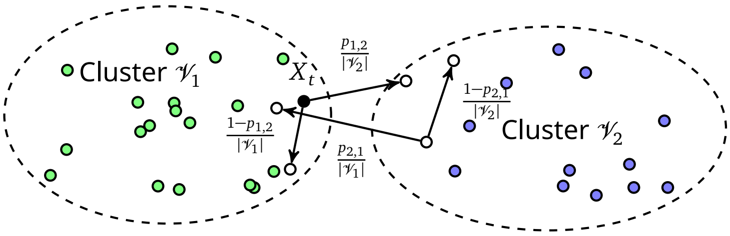
Motivation

The goal in typical community detection problems is to infer which state belongs to which cluster from the edges of a graph where a hidden community structure exist. However, sometimes observations of the edges are not independent as the next example shows.

Example 1: Among n music tracks we choose a genre at random and then a track in that genre. The choice usually depends on the previous tracks we have listened to so far.



Block Markov Chains (BMCs)[5] allow for *time dependency* in its observations. An example of a BMC with two clusters:



Example 2: For a transition from cluster $i \in \{1,2\}$, cluster 1 or 2 is chosen first with probability p_{i1} and $1-p_{i1}$ respectively. Next state from the cluster is chosen uniformly at random. This process generates a trajectory $(X_t)_{t=1}^{T_n}$.

The frequency matrix of a BMC

Suppose we have n states. $\hat{N}$ records the transitions between each pair of states in a trajectory of length T_n, where

$$\hat{N}_{xy} = \sum_{t=0}^{T_n-1} \mathbb{1}[X_t = x, X_{t+1} = y] \quad \text{for} \quad x, y \in [n]. \qquad (1)$$

We study $\|\hat{N} - \mathbb{E}[\hat{N}]\|$ as $n \to \infty$ because:

1. Bounds on this object provide performance guarantees for spectral clustering algorithms [3, 4].

2. Weakly dependent entries of $\hat{N}$ prevent from directly using typical bounding methods making the problem interesting.

3. Sparsity of $\hat{N}$ is determined by the length of the trajectory T_n. Different regimes are expected, as is with Erdös–Rényi random graphs (ERRGs) [1, 2].

Spectral Norm of $\hat{N} - \mathbb{E}(\hat{N})$

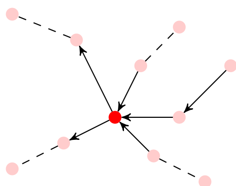
Our first result is a lower bound:

Proposition: *If $\omega(n) = T_n = o(n^2)$, then there exist constants $\mathfrak{b}, \mathfrak{c}_{\mathfrak{b}} > 0$ independent of n and an integer $n_0 \in \mathbb{N}_+$ such that for all $n \geq n_0$,*

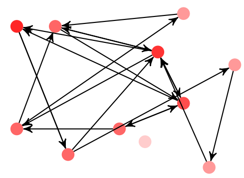
$$\mathbb{P}(\|\hat{N} - \mathbb{E}[\hat{N}]\| > \mathfrak{b}\sqrt{T_n/n}) \geq 1 - \exp(-\mathfrak{c}_{\mathfrak{b}}(T_n\ n)). \qquad (2)$$

Our second result is an order-wise matching upper bound to $\|\hat{N} - \mathbb{E}[\hat{N}]\|$. However, we have to regularize $\hat{N}$ depending on the regimes:

Sparse — $T_n = o(n \ln n)$



Dense — $T_n = \Omega(n \ln n)$



In the sparse regime, we set to zero states that are most visited (see reddest state above) yielding a trimmed matrix $\hat{N}_\Gamma$.

Theorem: *The following holds:*
(a) If $T_n = \Omega(n \ln n)$, then $\|\hat{N} - \mathbb{E}[\hat{N}]\| = O_{\mathbb{P}}(\sqrt{T_n/n})$.
(b) If $T_n = \omega(n)$ and Γ^c is a set of size $\lfloor n e^{-T_n/n} \rfloor$ containing the states with highest number of visits, i.e., with the property that $\min_{y \in \Gamma^c} \hat{N}_{[n],y} \geq \max_{y \in \Gamma} \hat{N}_{[n],y}$, then

$$\|\hat{N}_\Gamma - \mathbb{E}[\hat{N}]\| = O_{\mathbb{P}}(\sqrt{T_n/n}). \qquad (3)$$

Proof sketch

We use an ϵ-net argument and separate contributions to the norm in two terms. We then leverage concentration inequalities for Markov chains and exploit the low rank structure of BMCs to bound the terms. This proof method draws inspiration from an analogous result in sparse ERRG [2].

Conclusion

In BMCs, $\|\hat{N} - \mathbb{E}[\hat{N}]\| = \Theta_{\mathbb{P}}(\sqrt{T_n/n})$, a first step towards understanding the limiting distribution of the spectrum.

References

[1] P. Erdős and A. Rényi. On random graphs. i. *Publ. Math. Debrecen*, 6:290–297, 1959.
[2] U. Feige and E. Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005.
[3] R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE transactions on information theory*, 56(6):2980–2998, 2010.
[4] J. Lei, A. Rinaldo, et al. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015.
[5] J. Sanders, A. Proutière, S.-Y. Yun, et al. Clustering in block markov chains. *Annals of Statistics*, 48(6):3488–3512, 2020.