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]

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:

![Diagram of two clusters with transitions between tracks](image)

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

$$\hat{N}_{xy} = \sum_{t=0}^{T_n-1} 1[x_t = x, x_{t+1} = y] \text{ for } x, y \in [n].$$

We study $\|\hat{N} - E(\hat{N})\|$ 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 Erdős–Rényi random graphs (ERRGs) [1, 2].

Spectral Norm of $\hat{N} - E(\hat{N})$
Our first result is a lower bound:

**Proposition:** If $\omega(n) = T_n = o(n^2)$, then there exist constants $b, \epsilon_0 > 0$ independent of $n$ and an integer $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$P(\|\hat{N} - E(\hat{N})\| > b\sqrt{T_n/n}) \geq 1 - \exp(-\epsilon_0(T_n/n)).$$

Our second result is an order-wise matching upper bound to $\|\hat{N} - E(\hat{N})\|$. 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 $\hat{N}_T$.

**Theorem:** The following holds:

(a) If $T_n = \Omega(n \ln n)$, then $\|\hat{N} - E(\hat{N})\| = O_\beta(\sqrt{T_n/n})$.

(b) If $T_n = o(n)$ and $\Gamma$ 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 $\min_{y \in \Gamma} \hat{N}_{[n],y} \geq \max_{y \notin \Gamma} \hat{N}_{[n],y}$, then

$$\|\hat{N}_T - E(\hat{N})\| = O_\beta(\sqrt{T_n/n}).$$

**Proof sketch**
We use an $\epsilon$-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, $\|\hat{N} - E(\hat{N})\| = O_\beta(\sqrt{T_n/n})$, 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. L. Fienberg and E. Ofek. Spectral techniques applied to sparse random graphs. Random Structures & Algorithms, 19(2):251–275, 2001.
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(2):1255–1277, 2015.
5. J. Sanders, A. Proufèdre, S.-Y. Yun, et al. Clustering in block markov chains. Annals of Statistics, 48(6):3488–3512, 2020.