# Optimal Clustering for Block Markov Chains

Jaron Sanders[†,*] Alexandre Proutière[†]

[*]Dept. of Quantum and Computer Engineering, TU Delft, [†]Dept. of Automatic Control, KTH

*This research was performed at the **KTH Royal Institute of Technology**, Stockholm, Sweden. We now continue our collaboration, working from TU Delft and KTH, respectively.*

## Relevancy and Future Research

The ability to accurately discover all hidden relations between items that share similarities is of paramount importance to a wide range of disciplines. Clustering algorithms are employed throughout social sciences, biology, computer science, economics, and physics. The reason these techniques are prevalent is that once clusters have been identified, any subsequent analysis or optimization procedure benefits from a powerful reduction in dimensionality.

Our paper is a world's first on extending the classical Stochastic Block Model (SBM) type of results, which are valid for random graphs, into the domain of biased random walks. It opens up many new exciting research directions:

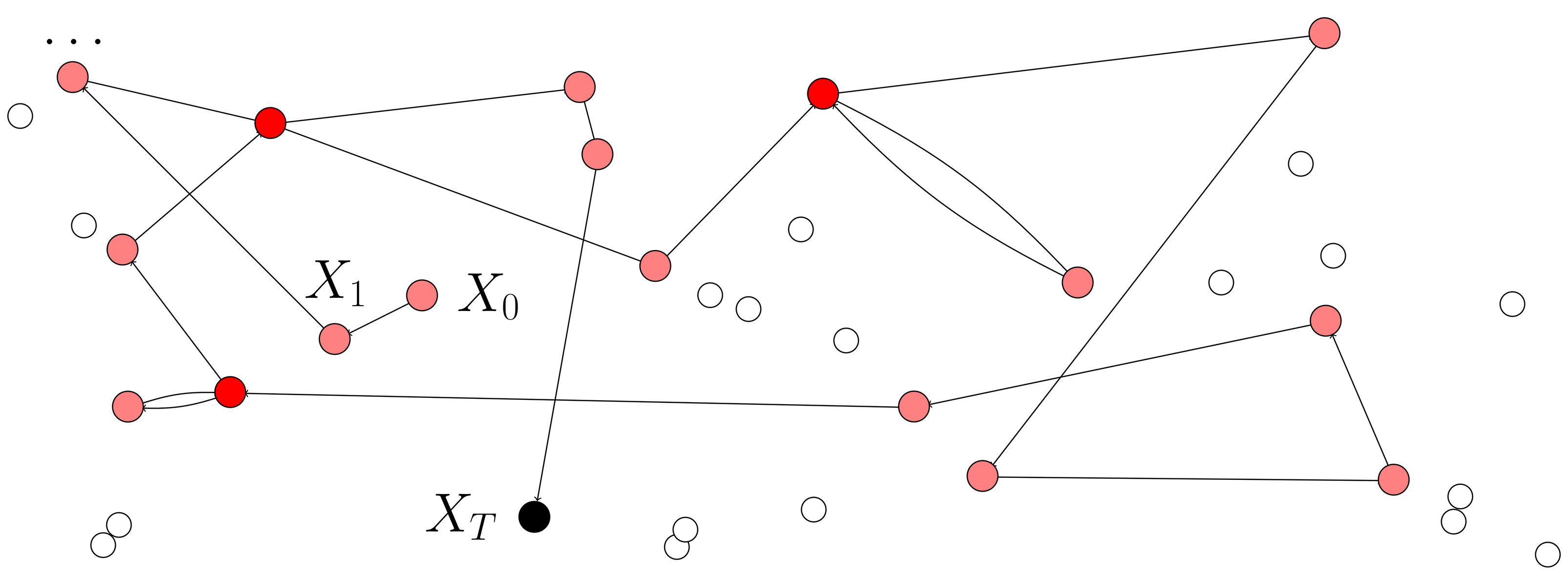
- We hope to extend the techniques developed here for an uncontrolled Block Markov Chain (BMC) to the more general case of controlled Markov chains, and devise reinforcement learning algorithms that will efficiently exploit an underlying cluster structure.
- We require further investigating into how eigenvalues of *biased* random matrices concentrate around its spectrum. The reason is that BMCs introduce mathematical challenges due to the time-dependencies of samples. We aim to extend fundamental insights like Wigner's Semicircle Law.
- Finite BMCs live on countable state spaces. We now have the ambitious goal to generalize our clustering algorithm to continuous state spaces.

## Summary

Our paper considers cluster detection in BMCs. These Markov chains are characterized by a block structure in their transition matrix. More precisely, the $n$ possible states are divided into a finite number of $K$ groups or clusters, such that states in the same cluster exhibit the same transition rates to other states. One observes a trajectory of the Markov chain, and the objective is to recover, from this observation only, the initially unknown clusters. In this paper we devise a clustering procedure that **accurately** ($\approx 100\%$ recovery), **efficiently** (= from the shortest paths), and **provably** (so mathematically guaranteed!) detects the clusters. We first derive a fundamental information-theoretical lower bound on the detection error rate satisfied under any clustering algorithm. This bound identifies the parameters of the BMC, and trajectory lengths $T$, for which it is possible to accurately detect the clusters. We next develop two clustering algorithms that can together accurately recover the cluster structure from the shortest possible trajectories, whenever the parameters allow detection. These algorithms thus reach the fundamental detectability limit, and are optimal in that sense.

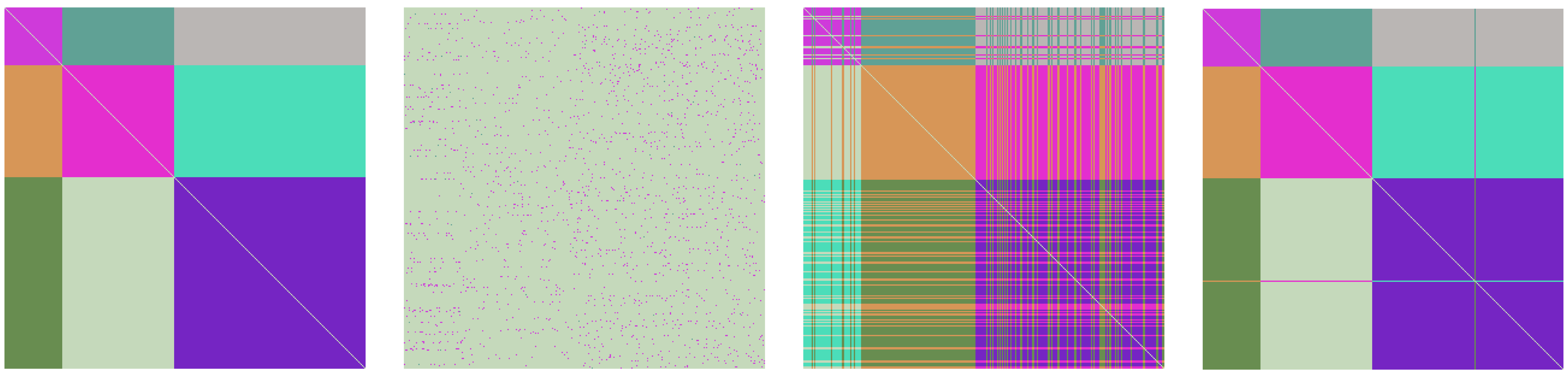
## The Algorithms

Our clustering procedure consists of two steps. First, we cluster the majority of states roughly correctly through a rank-$K$ approximation of a random matrix corresponding to the empirical transition rates between any pair of states, and a subsequent application of a $K$-means algorithm. Next, we exploit the just-learned rough structure *and* the sample path to move each individual state into the cluster the state most likely belongs to through a recursive, local maximization of a log-likelihood ratio.




*Our goal is to infer the hidden cluster structure underlying a Markov chain from one observation of a sample path $X_0, X_1, \dots, X_T$.*

## A Demonstration



*(a) Unknown kernel $P$.* *(b) Noisy observation.* *(c) Initial clustering.* *(d) Final clustering.*

In these simulations to your left, we generated a path of length $T = n^{1.025} \ln n \approx 1973$ for $n = 300$ and $K = 3$. Once our algorithm finished, $99.7\%$ of all states were accurately clustered.

## The Fundamental Detectability Limit

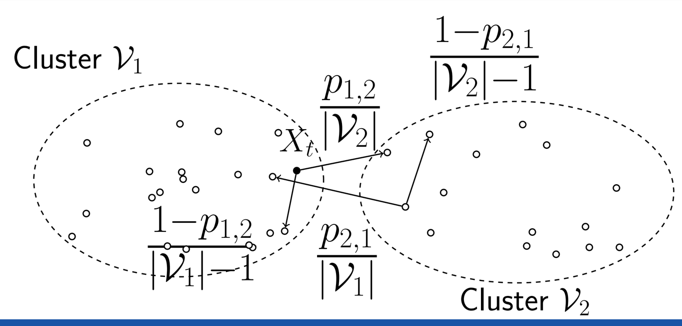
Let $\pi$ solve $\pi^{\mathrm{T}} p = \pi^{\mathrm{T}}$, and for any stochastic vector $\alpha$ and matrix $p$, set $I(\alpha, p) \triangleq \min_{a \neq b} \Big\{ \sum_{k=1}^{K} \frac{1}{\alpha_a} \Big( \pi_a p_{a,k} \ln \frac{p_{a,k}}{p_{b,k}} + \pi_k p_{k,a} \ln \frac{p_{k,a} \alpha_b}{p_{k,b} \alpha_a} \Big) + \Big( \frac{\pi_b}{\alpha_b} - \frac{\pi_a}{\alpha_a} \Big) \Big\}$.

**"Lower bound":** If $I(\alpha, p) > 0$, then $\exists$ strictly positive, finite constants $C, J(\alpha, p)$ independent of $n$ such that $\mathbb{E}_P \Big[ \frac{|\mathcal{E}|}{n} \Big] \geq C \exp \Big( -J(\alpha, p) \frac{T}{n} + o\Big(\frac{T}{n}\Big) \Big)$ for any clustering algorithm.

**"Upper bound":** If $I(\alpha, p) > 0$, if $\exists_{0 < \eta \neq 1} : \max_{a,b,c=1,\dots,K} \{ p_{b,a}/p_{c,a}, p_{a,b}/p_{a,c} \} \leq \eta$, $\|\hat{N} - N\| = O_{\mathbb{P}}(f(n,T))$ for some $f(n,T) = o(T/n)$, and if $\|\hat{P} - P\| = O_{\mathbb{P}}(g(n,T))$ for some $g(n,T) = o(1)$, then $\exists$ a clustering algorithm that misclassifies $|\mathcal{E}| = o_{\mathbb{P}}(1)$ states.

As a consequence, a necessary condition for the existence of an asymptotically exact clustering algorithm, i.e., such that $\mathbb{E}_P[|\mathcal{E}|] = o(1)$, is $T = \omega(n \ln n)$.

TU Delft Delft University of Technology




*We invite all people who are interested in doing research with us to **contact us**. If you e.g. are a MSc student looking for a PhD position or a PhD candidate looking for a postdoc position, or wish to collaborate and jointly apply for funding, tell us!*