Clustering in Block Markov Chains

Jaron Sanders[1,2] Alexandre Proutière[1] Se Young Yun[3]

[1]KTH Royal Institute of Technology, Sweden

[2]Delft University of Technology, The Netherlands

[3]Korea Advanced Institute of Science and Technology, South Korea

Informs APS Conference 2019, Brisbane, Australia

Part I

Our idea and the motivation

Our idea: Can we do clustering in Markov Chains (MCs)?

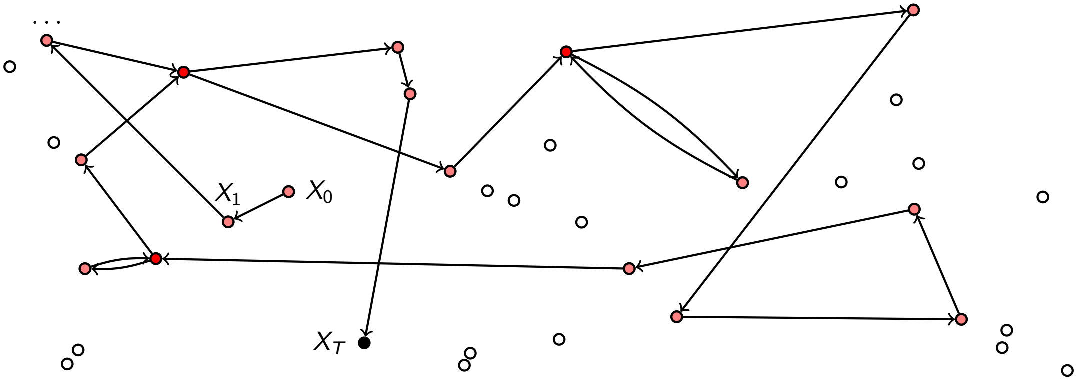


Figure: The goal of this paper is to infer the hidden cluster structure underlying a Markov chain $\{X_t\}_{t\geq 0}$, from one observation of a sample path $X_0, X_1, \ldots, X_T$ of length T.

The motivation

Clustering in MCs is motivated by *Reinforcement Learning (RL)* on large state spaces.

RL has recently received substantial attention due to its wide spectrum of applications (robotics, games, medicine, finance, etc), or more popularly said, *artificial intelligence*.

The motivation

Clustering in MCs is motivated by *Reinforcement Learning (RL)* on large state spaces.

RL has recently received substantial attention due to its wide spectrum of applications (robotics, games, medicine, finance, etc), or more popularly said, *artificial intelligence*.

In RL, the objective is to quickly identify an optimal control policy by *observing a trajectory of a Markov chain*.

Unfortunately, the time to learn the best policies using e.g. Q-learning *increases dramatically* with the number of states.

The motivation

Clustering in MCs is motivated by *Reinforcement Learning (RL)* on large state spaces.

RL has recently received substantial attention due to its wide spectrum of applications (robotics, games, medicine, finance, etc), or more popularly said, *artificial intelligence*.

In RL, the objective is to quickly identify an optimal control policy by *observing a trajectory of a Markov chain*.

Unfortunately, the time to learn the best policies using e.g. Q-learning *increases dramatically* with the number of states.

In practical problems however, different states may yield *similar* reward and exhibit *similar* transition probabilities. **In other words, states could maybe be clustered.**

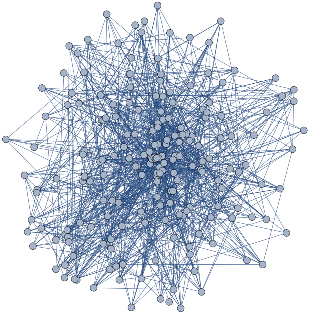
Part II

The literature and our model

Clustering in Stochastic Block Models (SBMs)

SBMs generate random graphs with groups of similar vertices.

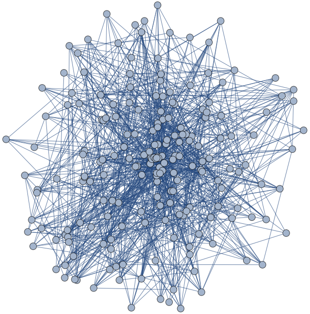
E.g. Suppose $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$. An edge is drawn between $x, y \in \mathcal{V}$ w.p. $p \in (0,1)$ if they belong to the same group, and w.p. $q \in (0,1), p \neq q$ otherwise.



Clustering in Stochastic Block Models (SBMs)

SBMs generate random graphs with groups of similar vertices.

E.g. Suppose $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$. An edge is drawn between $x, y \in \mathcal{V}$ w.p. $p \in (0,1)$ if they belong to the same group, and w.p. $q \in (0,1), p \neq q$ otherwise.



The goal is to infer the clusters from such an observed random graph.

Fundamental limits for clustering in SBMs in literature

Much literature exists on **when** and **how** we can cluster in SBMs.

[1] *"Community detection and SBMs: recent developments"*, Emmanuel Abbe, 2017 gives overview.

Fundamental limits for clustering in SBMs in literature

Much literature exists on **when** and **how** we can cluster in SBMs.

To start, many papers laid foundation for the discovery of the fundamental limits:[1]

Including: Holland, Laskey, Leinhardt 1983; Bui, Chaudhuri, Leighton, Sipser 1984; Boppana 1987; Dyer, Frieze 1989; Snijders, Nowicki 1997; Jerrum, Sorkin 1998; Condon, Karp 1999; Carson, Impagliazzo 2001; McSherry 2001; Bickel, Chen 2009; Rohe, Chatterjee, Yi 2011, and more.

[1] "*Community detection and SBMs: recent developments*", Emmanuel Abbe, 2017 gives overview.

Fundamental limits for clustering in SBMs in literature

Much literature exists on **when** and **how** we can cluster in SBMs.

To start, many papers laid foundation for the discovery of the fundamental limits:[1]

Including: Holland, Laskey, Leinhardt 1983; Bui, Chaudhuri, Leighton, Sipser 1984; Boppana 1987; Dyer, Frieze 1989; Snijders, Nowicki 1997; Jerrum, Sorkin 1998; Condon, Karp 1999; Carson, Impagliazzo 2001; McSherry 2001; Bickel, Chen 2009; Rohe, Chatterjee, Yi 2011, and more.

Theorem (Decelle, Krzakala, Moore, Zdeborova 2011; Massoulié 2014; Mossel, Neeman, Sly 2015)

If $p = a/n$, $q = b/n$, and $|\mathcal{V}_1| = |\mathcal{V}_2|$, then $a - b \geq \sqrt{2(a+b)}$ is a necessary and sufficient condition for the existence of algorithms that can detect the clusters.

Theorem (Abbe, Bandeira, Hall, 2014; Mossel, Neeman, Sly 2014)

If $p = a \ln n/n$, $q = b \ln n/n$, then $|\sqrt{a} - \sqrt{b}| > \sqrt{2}$ allows for exact recovery.

---

[1] *"Community detection and SBMs: recent developments"*, Emmanuel Abbe, 2017 gives overview.

Fundamental limits for clustering in SBMs in literature

Much literature exists on **when** and **how** we can cluster in SBMs.

To start, many papers laid foundation for the discovery of the fundamental limits:[1]

Including: Holland, Laskey, Leinhardt 1983; Bui, Chaudhuri, Leighton, Sipser 1984; Boppana 1987; Dyer, Frieze 1989; Snijders, Nowicki 1997; Jerrum, Sorkin 1998; Condon, Karp 1999; Carson, Impagliazzo 2001; McSherry 2001; Bickel, Chen 2009; Rohe, Chatterjee, Yi 2011, and more.

Theorem (Decelle, Krzakala, Moore, Zdeborova 2011; Massoulié 2014; Mossel, Neeman, Sly 2015)

If $p = a/n$, $q = b/n$, and $|\mathcal{V}_1| = |\mathcal{V}_2|$, then $a - b \geq \sqrt{2(a+b)}$ is a necessary and sufficient condition for the existence of algorithms that can detect the clusters.

Theorem (Abbe, Bandeira, Hall, 2014; Mossel, Neeman, Sly 2014)

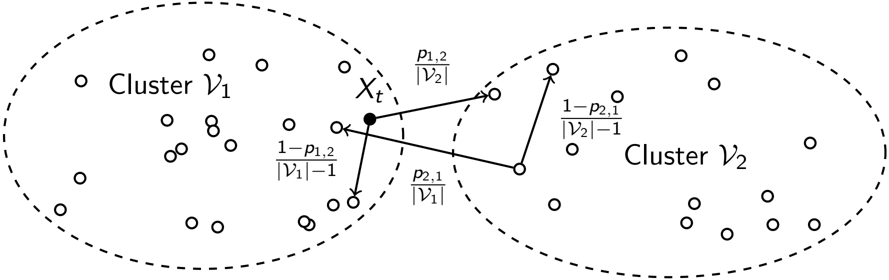
If $p = a \ln n/n$, $q = b \ln n/n$, then $|\sqrt{a} - \sqrt{b}| > \sqrt{2}$ allows for exact recovery.

In both cases, **efficient algorithms** were also developed that achieve the thresholds!

[1] *"Community detection and SBMs: recent developments"*, Emmanuel Abbe, 2017 gives overview.

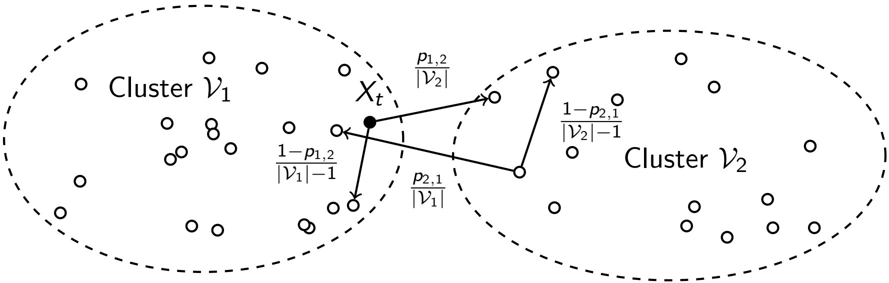
Clustering in Block Markov Chains (BMCs)

Our work also investigates **when** and **how** we can cluster, **but then in BMCs**!



Clustering in Block Markov Chains (BMCs)

Our work also investigates **when** and **how** we can cluster, **but then in BMCs**!



Let $\{X_t\}_{t\geq 0}$ be a BMC with parameters (n, α, p). Its transition matrix is given by

$$P_{x,y} \triangleq \frac{p_{\sigma(x),\sigma(y)}}{|\mathcal{V}_{\sigma(y)}| - \mathbb{1}[\sigma(x) = \sigma(y)]} \mathbb{1}[x \neq y] \quad \text{for all} \quad x, y \in \mathcal{V}.$$

Its equilibrium distribution will be denoted by Π_x for $x \in \mathcal{V}$.

Structure of the transition matrix

Here's an example transition matrix for $K = 3$ clusters:

$$
P = \left(\begin{array}{cc:ccc:ccccc}
0 & p_{1,1} & \frac{p_{1,2}}{3} & \frac{p_{1,2}}{3} & \frac{p_{1,2}}{3} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} \\
p_{1,1} & 0 & \frac{p_{1,2}}{3} & \frac{p_{1,2}}{3} & \frac{p_{1,2}}{3} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} & \frac{p_{1,3}}{5} \\
\hdashline
\frac{p_{2,1}}{2} & \frac{p_{2,1}}{2} & 0 & \frac{p_{2,2}}{2} & \frac{p_{2,2}}{2} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} \\
\frac{p_{2,1}}{2} & \frac{p_{2,1}}{2} & \frac{p_{2,2}}{2} & 0 & \frac{p_{2,2}}{2} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} \\
\frac{p_{2,1}}{2} & \frac{p_{2,1}}{2} & \frac{p_{2,2}}{2} & \frac{p_{2,2}}{2} & 0 & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} & \frac{p_{2,3}}{5} \\
\hdashline
\frac{p_{3,1}}{2} & \frac{p_{3,1}}{2} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & 0 & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} \\
\frac{p_{3,1}}{2} & \frac{p_{3,1}}{2} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,3}}{4} & 0 & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} \\
\frac{p_{3,1}}{2} & \frac{p_{3,1}}{2} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & 0 & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} \\
\frac{p_{3,1}}{2} & \frac{p_{3,1}}{2} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & 0 & \frac{p_{3,3}}{4} \\
\frac{p_{3,1}}{2} & \frac{p_{3,1}}{2} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,2}}{3} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & \frac{p_{3,3}}{4} & 0
\end{array}\right)
$$

Note the **block structure**, and that p must be a **stochastic matrix**.

Part III

Our main results

Main results

We obtain quantitative statements for

$$\mathcal{E} \triangleq \bigcup_{k=1}^{K} \hat{\mathcal{V}}_{\gamma^{\mathsf{opt}}(k)} \backslash \mathcal{V}_k \quad \text{where} \quad \gamma^{\mathsf{opt}} \in \arg \min_{\gamma \in \mathrm{Perm}(K)} \Big| \bigcup_{k=1}^{K} \hat{\mathcal{V}}_{\gamma(k)} \backslash \mathcal{V}_k \Big|.$$

Here, the sets $\hat{\mathcal{V}}_1, \ldots, \hat{\mathcal{V}}_K$ will always denote an approximate cluster assignment obtained from some clustering algorithm.

Main results

We obtain quantitative statements for

$$\mathcal{E} \triangleq \bigcup_{k=1}^{K} \hat{\mathcal{V}}_{\gamma^{\mathrm{opt}}(k)} \backslash \mathcal{V}_k \quad \text{where} \quad \gamma^{\mathrm{opt}} \in \arg \min_{\gamma \in \mathrm{Perm}(K)} \Big| \bigcup_{k=1}^{K} \hat{\mathcal{V}}_{\gamma(k)} \backslash \mathcal{V}_k \Big|.$$

Here, the sets $\hat{\mathcal{V}}_1, \ldots, \hat{\mathcal{V}}_K$ will always denote an approximate cluster assignment obtained from some clustering algorithm.

Remark

Throughout, we assume that K, α, p are fixed, and we study the asymptotic regime $n \to \infty$. Our clustering procedure will assume that K is known, and α, p unknown.

Information theoretical lower bound

Definition

For $\alpha \in \Delta^{K-1}$ and $p \in \mathbb{\Delta}^{(K-1)\times K}$, let

$$I(\alpha, p) \triangleq \min_{a \neq b} \left\{ \sum_{k=1}^{K} \frac{1}{\alpha_a} \left(\pi_a p_{a,k} \ln \frac{p_{a,k}}{p_{b,k}} + \pi_k p_{k,a} \ln \frac{p_{k,a} \alpha_b}{p_{k,b} \alpha_a} \right) + \left(\frac{\pi_b}{\alpha_b} - \frac{\pi_a}{\alpha_a} \right) \right\}.$$

Here π denotes the solution to $\pi^{\mathrm{T}} p = \pi^{\mathrm{T}}$.

Information theoretical lower bound

Definition

For $\alpha \in \Delta^{K-1}$ and $p \in \mathbb{\Delta}^{(K-1)\times K}$, let

$$I(\alpha, p) \triangleq \min_{a \neq b} \Big\{ \sum_{k=1}^{K} \frac{1}{\alpha_a} \Big(\pi_a p_{a,k} \ln \frac{p_{a,k}}{p_{b,k}} + \pi_k p_{k,a} \ln \frac{p_{k,a}\alpha_b}{p_{k,b}\alpha_a} \Big) + \Big(\frac{\pi_b}{\alpha_b} - \frac{\pi_a}{\alpha_a} \Big) \Big\}.$$

Here π denotes the solution to $\pi^{\mathrm{T}} p = \pi^{\mathrm{T}}$.

Theorem

Assume that $T = \omega(n)$. Then there exists a strictly positive and finite constant C independent of n such that: for any clustering algorithm

$$\mathbb{E}_P[|\mathcal{E}|] \geq C n \exp\Big(- I(\alpha, p) \frac{T}{n} (1 + o(1)) \Big).$$

Asymptotically accurate / exact detection

Conditions for asymptotically accurate detection
In view of our lower bound,

$$\mathbb{E}_P\Big[\frac{|\mathcal{E}|}{n}\Big] \geq C\exp\Big(-I(\alpha,p)\frac{T}{n}(1+o(1))\Big),$$

there may exist asymptotically *accurate* algorithms only if $I(\alpha,p) > 0$ and $T = \omega(n)$.

Asymptotically accurate / exact detection

Conditions for asymptotically accurate detection
In view of our lower bound,

$$\mathbb{E}_P\Big[\frac{|\mathcal{E}|}{n}\Big] \geq C \exp\Big(- I(\alpha, p)\frac{T}{n}(1 + o(1))\Big),$$

there may exist asymptotically *accurate* algorithms only if $I(\alpha, p) > 0$ and $T = \omega(n)$.

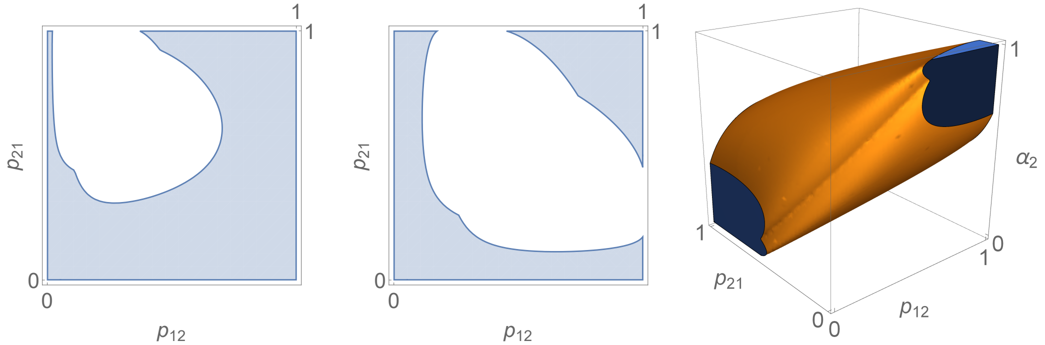
Conditions for asymptotically exact detection
Similarly,

$$\mathbb{E}_P[|\mathcal{E}|] \geq C \exp\Big(\ln n - I(\alpha, p)\frac{T}{n}(1 + o(1))\Big),$$

so necessary conditions for the existence of an asymptotically *exact* algorithm are $I(\alpha, p) > 0$ and $T - \frac{n \ln(n)}{I(\alpha,p)} = \omega(1)$. In particular, T must scale atleast as $n \ln n$.

Clustering in the critical regime

There is a **phase transition** in the *critical regime* $T = n \ln n$



$\alpha_2 = 1/4$ $\alpha_2 = 1/2$ Region where $I(\alpha, p) < 1$.

Figure: (left, middle) The parameters $(p_{1,2}, p_{2,1})$ in blue for which asymptotic exact recovery should be possible in the critical regime $T = n \ln n$ for $K = 2$ clusters. (right) The parameters $(\alpha_2, p_{1,2}, p_{2,1})$ for which asymptotic exact recovery is likely not possible, i.e., $I(\alpha, p) < 1$.

Procedure for cluster recovery

We have now established **necessary conditions** for asymptotically accurate and exact recovery, and identified **performance limits** satisfied by any clustering algorithm.

Procedure for cluster recovery

We have now established **necessary conditions** for asymptotically accurate and exact recovery, and identified **performance limits** satisfied by any clustering algorithm.

Next, we devised a clustering procedure that **reaches** these limits order-wise. Our procedure takes a sample path $X_0, X_1, \ldots, X_T$ as input and calculates

$$\hat{N}_{x,y} \triangleq \sum_{t=0}^{T-1} \mathbb{1}[X_t = x, X_{t+1} = y] \quad \text{for} \quad x, y \in \mathcal{V},$$

and then proceeds in two steps called:

- the *Spectral Clustering Algorithm (SCA)*, and
- the *Cluster Improvement Algorithm (CIA)*

Spectral Clustering Algorithm (SCA)

Input: n, K, and a trajectory $X_0, X_1, \dots, X_T$

Output: An approximate cluster assignment $\hat{\mathcal{V}}_1^{[0]}, \dots, \hat{\mathcal{V}}_K^{[0]}$, and matrix $\hat{N}$

begin

  for $x \leftarrow 1$ **to** n **do**

    for $y \leftarrow 1$ **to** n **do**

      $\hat{N}_{x,y} \leftarrow \sum_{t=0}^{T-1} \mathbb{1}[X_t = x, X_{t+1} = y]$;

    end

  end

  Calculate the trimmed matrices $\hat{N}_\Gamma$;

  Calculate the Singular Value Decomposition (SVD) $U \Sigma V^{\mathrm{T}}$ of $\hat{N}_\Gamma$;

  Order U, Σ, V s.t. the singular values $\sigma_1 \geq \sigma \geq \dots \geq \sigma_n \geq 0$ are in descending order;

  Construct the rank-K approximation $\hat{R} = \sum_{k=1}^{K} \sigma_k U_{\cdot,k} {V_{\cdot,k}}^{\mathrm{T}}$;

  Apply a K-means algorithm to $[\hat{R}, \hat{R}^\top]$ to determine $\hat{\mathcal{V}}_1^{[0]}, \dots, \hat{\mathcal{V}}_K^{[0]}$;

end

Algorithm 1: Pseudo-code for the Spectral Clustering Algorithm.

Performance of the SCA

Theorem

Assume that $T = \omega(n)$ *and* $I(\alpha, p) > 0$. *Then the proportion of misclassified states after the Spectral Clustering Algorithm satisfies:*

$$\frac{|\mathcal{E}|}{n} = O_{\mathbb{P}}\Big(\frac{n}{T} \ln \frac{T}{n}\Big) = o_{\mathbb{P}}(1).$$

Thus the SCA achieves asymptotically accurate detection whenever this is possible.

Question! But there's a huge problem. What does the SCA fail at?

Performance of the SCA

Theorem

Assume that $T = \omega(n)$ and $I(\alpha, p) > 0$. Then the proportion of misclassified states after the Spectral Clustering Algorithm satisfies:

$$\frac{|\mathcal{E}|}{n} = O_{\mathbb{P}}\Big(\frac{n}{T} \ln \frac{T}{n}\Big) = o_{\mathbb{P}}(1).$$

Thus the SCA achieves asymptotically accurate detection whenever this is possible.

Question! But there's a huge problem. What does the SCA fail at?

Answer. The bound fails to guarantee asymptotic exact recovery, even in the case $T = \omega(n \ln(n))$. We cannot guarantee that its recovery rate approaches Theorem 4's fundamental limit!

Cluster Improvement Algorithm (CIA)

Input: An approximate assignment $\hat{\mathcal{V}}_1^{[t]}, \ldots, \hat{\mathcal{V}}_K^{[t]}$, and matrix $\hat{N}$

Output: A revised assignment $\hat{\mathcal{V}}_1^{[t+1]}, \ldots, \hat{\mathcal{V}}_K^{[t+1]}$

begin

  $n \leftarrow \dim(\hat{N})$, $\mathcal{V} \leftarrow \{1, \ldots, n\}$, $T \leftarrow \sum_{x \in \mathcal{V}} \sum_{y \in \mathcal{V}} \hat{N}_{x,y}$;

  for $a \leftarrow 1$ **to** K **do**

    $\hat{\pi}_a \leftarrow \hat{N}_{\hat{\mathcal{V}}_a^{[t]}, \mathcal{V}} / T$, $\hat{\alpha}_a \leftarrow |\hat{\mathcal{V}}_a^{[t]}|/n$, $\hat{\mathcal{V}}_a^{[t+1]} \leftarrow \emptyset$;

    for $b \leftarrow 1$ **to** K **do**

      $\hat{p}_{a,b} \leftarrow \hat{N}_{\hat{\mathcal{V}}_a^{[t]}, \hat{\mathcal{V}}_b^{[t]}} / \hat{N}_{\hat{\mathcal{V}}_a^{[t]}, \mathcal{V}}$;

    end

  end

  for $x \leftarrow 1$ **to** n **do**

    $c_x^{\text{opt}} \leftarrow \arg\max_{c=1,\ldots,K} \left\{ \sum_{k=1}^{K} \left(\hat{N}_{x, \hat{\mathcal{V}}_k^{[t]}} \ln \hat{p}_{c,k} + \hat{N}_{\hat{\mathcal{V}}_k^{[t]}, x} \ln \frac{\hat{p}_{k,c}}{\hat{\alpha}_c} \right) - \frac{T}{n} \cdot \frac{\hat{\pi}_c}{\hat{\alpha}_c} \right\}$;

    $\hat{\mathcal{V}}_{c_x^{\text{opt}}}^{[t+1]} \leftarrow \hat{\mathcal{V}}_{c_x^{\text{opt}}}^{[t+1]} \cup \{x\}$;

  end

end

Algorithm 2: Pseudo-code for the Cluster Improvement Algorithm.

Performance of the CIA

Theorem

Assume that $T = \omega(n)$ *and* $I(\alpha, p) > 0$*. Then there exists a constant* $C > 0$ *such that for any* $t \geq 1$*, after* t *iterations of the Clustering Improvement Algorithm, initially applied to the output of the Spectral Clustering Algorithm, we have:*

$$\frac{|\mathcal{E}^{[t]}|}{n} = O_{\mathbb{P}}\Big(\mathrm{e}^{-t\left(\ln \frac{T}{n} - \ln \ln \frac{T}{n}\right)} + \mathrm{e}^{-C \frac{T}{n} I(\alpha, p)}\Big)$$

Performance of the CIA

Theorem

Assume that $T = \omega(n)$ and $I(\alpha, p) > 0$. Then there exists a constant $C > 0$ such that for any $t \geq 1$, after t iterations of the Clustering Improvement Algorithm, initially applied to the output of the Spectral Clustering Algorithm, we have:

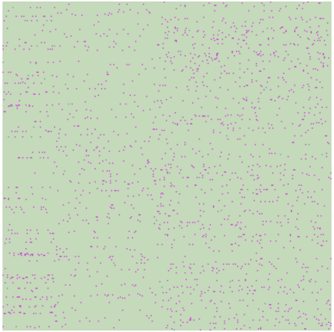
$$\frac{|\mathcal{E}^{[t]}|}{n} = O_{\mathbb{P}}\Big(\mathrm{e}^{-t\left(\ln \frac{T}{n} - \ln\ln \frac{T}{n}\right)} + \mathrm{e}^{-C\frac{T}{n}I(\alpha,p)} \Big)$$

Observe that for $t = \ln(n)$, the number of misclassified vertices after t applications of the CIA is at most of the order $n\mathrm{e}^{-C\frac{T}{n}I(\alpha,p)}$. Up to the constant C, this corresponds to Theorem 4's fundamental recovery rate limit.

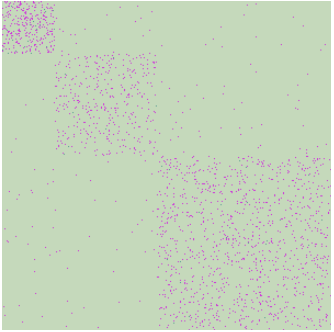
Plus, we have **asymptotically <u>exact</u> detection** when $T = \omega(n \ln n)$ and $I(\alpha, p) > 0$!

Let's start with an example – The observation and truth

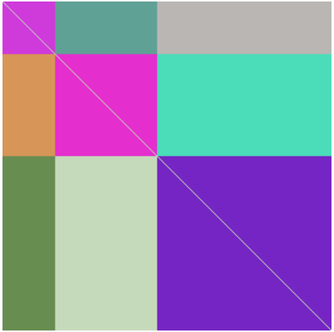
Consider $n = 300$ states grouped into three clusters of respective relative sizes $\alpha = (0.15, 0.35, 0.5)$. The transition rates between these clusters are defined by: $p = \big(0.9200, 0.0450, 0.0350;\ 0.0125, 0.8975, 0.0900;\ 0.0175, 0.0200, 0.9625\big)$.



(a) $\hat{N}$, unsorted

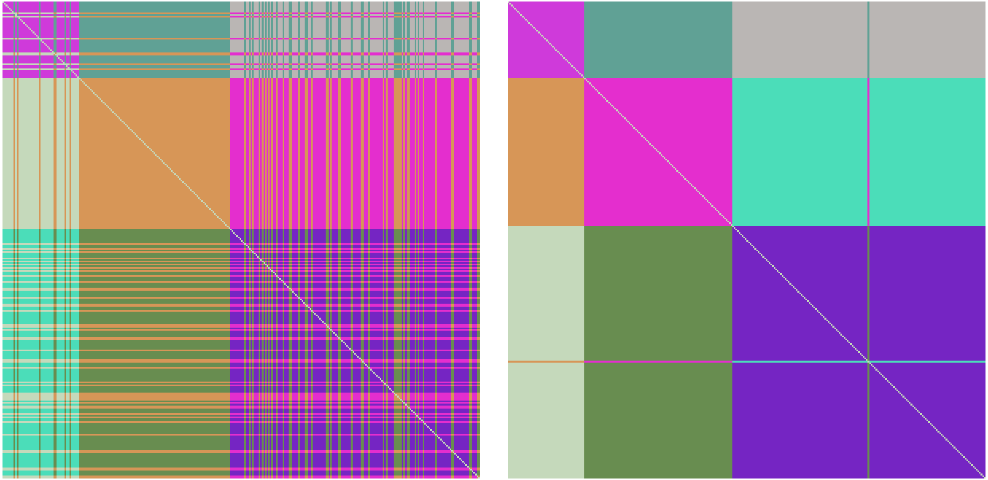


(b) $\hat{N}$, sorted



(c) P, sorted

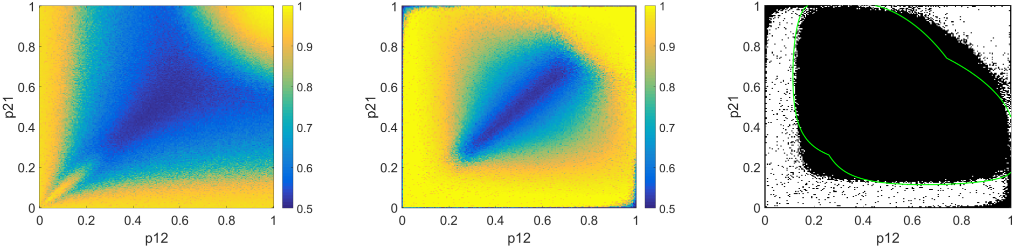
Let's start with an example – The procedure's 99.7% recovery



(a) Initial clustering. (b) Final clustering.

Our procedure in the critical regime

Consider $K = 2$, $\alpha_2 = \frac{1}{2}$, and $T = n \ln n$. Pascal Lagerweij (a MSc student) helped us numerically evaluate $\hat{\mathcal{F}}_1(\varepsilon) = \left\{ (p_{1,2}, p_{2,1}) \in (0,1)^2 \middle| \mathbb{E}_P \left[\frac{|\mathcal{E}^{[t]}|}{n} \right] \geq 1 - \varepsilon \right\}$.



After the SCA. After the CIA. $\hat{\mathcal{F}}_1(\varepsilon = 0.027)$

Figure: The average proportion of well-classified states for each rasterpoint $(p_{1,2}, p_{2,1}) \in (0,1)^2$, and numerical feasibility region of our clustering procedure (right), all in the critical regime $T = n \ln n$. The green line outlines the theoretical region $I(\alpha, p) \leq 1$ within which no algorithm exists able to asymptotically recover the clusters exactly.

Part IV

In conclusion

Let us summarize

Our paper "Clustering in Block Markov Chains":

- introduces Block Markov Chains (BMCs), a new interesting model;
- provides an information-theoretical lower bound for the detection error, tight conditions for asymptotically accurate detection and an almost tight condition for exact recovery;

Let us summarize

Our paper "Clustering in Block Markov Chains":

- introduces BMCs, a new interesting model;
- provides an information-theoretical lower bound for the detection error, tight conditions for asymptotically accurate detection and an almost tight condition for exact recovery;
- proposes an algorithm that almost reaches our information-theoretical lower bound;
- develops a new spectrum concentration bound for random matrices with *dependent* entries.

A preprint "Optimal Clustering Algorithms in Block Markov Chains" is available on **https://arxiv.org/abs/1712.09232**. This will soon be updated.