Markov Chains in the Slow Mixing Regime

Meysam Asadi, Kevin Oshiro, Ramezan Paravi, Narayana Santhanam, Changlong Wu
{masadi,kevinro,paravi,nsanthan,wuchangl}@hawaii.edu
Department of Electrical Engineering, University of Hawaiʻi Mānoa

Center for Science of Information
NSF Science and Technology Center

Objective

We observe a length-n sample generated by an unknown, stationary ergodic Markov process over a finite alphabet $\mathcal{A}$. Our goal is to provide sufficient conditions on length-n sample such that:

- Naive estimates of transition probabilities be accurate.
- Naive estimates of stationary probabilities be accurate.
- Provide deviation bounds which are *entirely data dependent*.

We also apply the estimation results for stationary probabilities to modify the Coupling From the Past algorithm for detecting communities in a graph.

Estimation Challenges

- The process could have long memory.
- The process could be slow mixing.

Natural Approach

- Memory is unknown *a-priori*.
- Approximate with a *coarser* Markov process:
  - ✓ Memory size is k_n for some known k_n.
  - ✓ Choose $k_n = \alpha_n \log n$ for some $\alpha_n = \mathcal{O}(1)$.
  - ✓ Leads to a consistent estimator as n grows.
- Call the coarser model the *Aggregated Model*.

Naive Estimates

Computation of naive estimators:

- Suppose sample is $Y_1^n = 1101010100$.
- Let $Y_{-\infty}^0 = \cdots 00$.
- Interested in aggregated parameters at depth 2.
- ✓ For instance, $\hat{P}(1|10) = \frac{3}{4}$ and $\hat{P}(0|10) = \frac{1}{4}$.

$$\cdots 00, 110\,1\,0\,1\,0\,1\,0\,0$$

- No reason such estimates make sense, since sample is *not* generated from aggregated model.

Dependencies Die Down

Considering our physical motivation, we assume

- Influence of prior symbols die down as we look further.
- Assume original process belongs to $\mathcal{M}_d$.
  - ✓ Does not imply memory is bounded.
  - ✓ No influence on mixing properties.

Good States

Combining universal compression results and the fact that dependencies die down:

- Identify a set $\tilde{G} \subseteq \mathcal{A}^{k_n}$ of *good* states that have occurred frequently enough in the sample.
- Any string $\mathbf{w} \in \tilde{G}$ is amenable to concentration results for conditional probabilities .

Stationary Probabilities

- Stationary probabilities are sensitive function of transition probabilities.

For deviation bounds, we consider the restriction of $\{Y_n\}_{n\geq 1}$ to $\tilde{G}$. Call it $\{Z_m\}_{m\geq 1}$.

- ✓ $\{Z_m\}_{m\geq 1}$ can be characterized using *stopping times* and by itself a Markov process.
- ✓ Let $\tilde{n}$ be the total count of good states in the sample. Define $V_m \triangleq \mathbb{E}[N_{\mathbf{w}}|Z_0, Z_1, \cdots Z_m]$.
- ✓ $\{V_m\}_{m=0}^{\tilde{n}}$ is a Doob Martingale.
- ✓ Bound Martingale differences by coupling argument.
- ✓ Using Azuma's Inequality for deriving concentration results.

Theorem

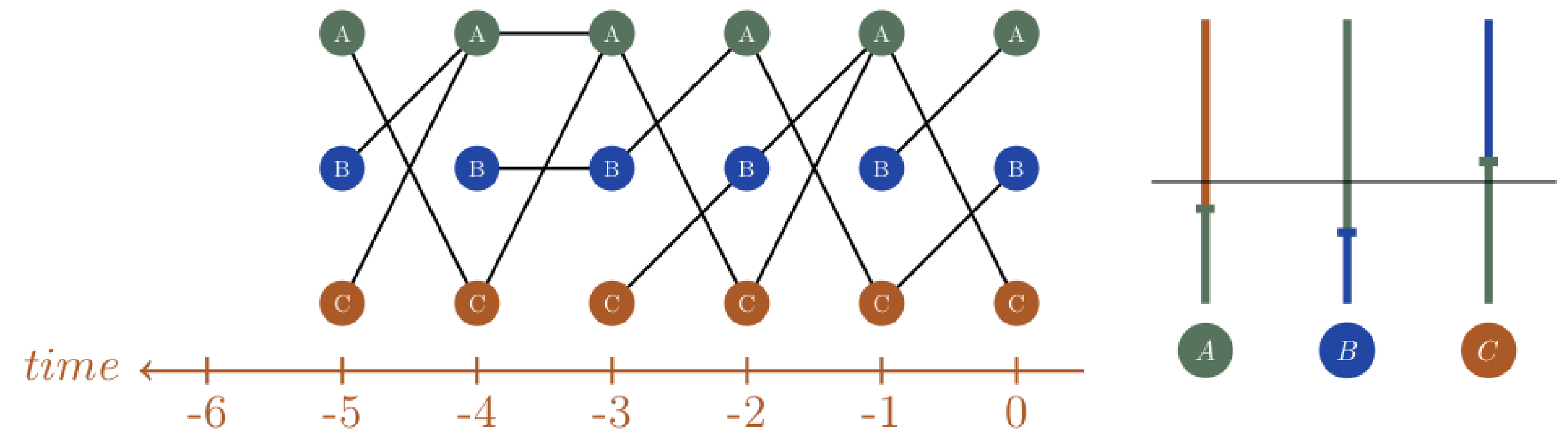
If $\{Z_m\}_{m\geq 1}$ is aperiodic, then for any $t > 0$, $Y_{-\infty}^0$ and $\mathbf{w} \in \tilde{G}$ we have

$$\mathcal{P}(|N_{\mathbf{w}} - \tilde{n}\frac{\mu(\mathbf{w})}{\mu(\tilde{G})}| \geq t|Y_{-\infty}^0) \leq 2\exp\left(-\frac{(t-\mathcal{B})^2}{2\tilde{n}\mathcal{B}^2}\right),$$

where $\mathcal{B}$ is entirely data dependent.

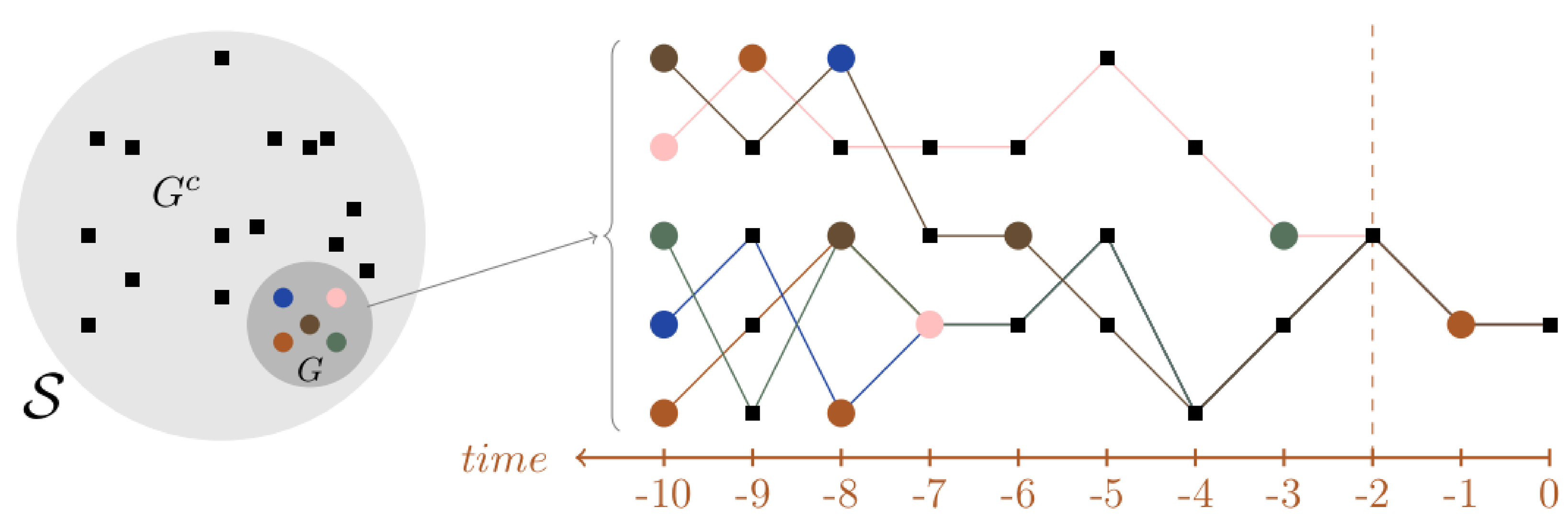
Coupling From The Past [Propp & Wilson, 1996]

- Run coupled Markov chains, one from each state $s \in \mathcal{S}$, and evolve the chains backwards in time.
- When all chains coalesce to a single state at time 0, that state is an exact sample from the stationary distribution.

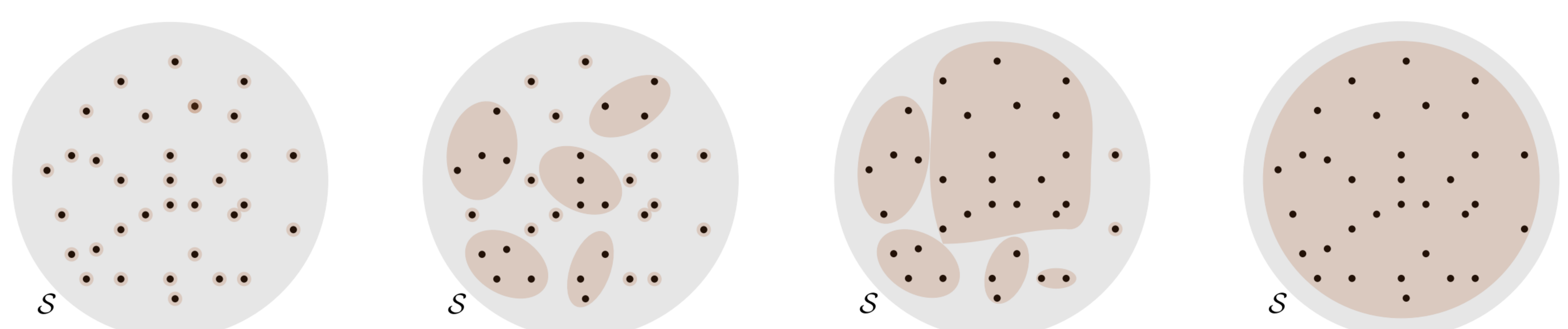


Community Detection

- Start $r = |\mathcal{S}|$ markov chains, one at each state $s \in \mathcal{S}$.
- Perform a random walk, simulating the chains backwards in time, using CFTP.
- Identify a set of critical times T, where chains have partially coalesced, each giving a clustering $\mathcal{C}$.
- Output the clustering $\mathcal{C}$ with the lowest cost $\mathcal{J}(\mathcal{C})$.



Partial Coalescence



Future Work

- The stationary probability results are sufficient to say that some estimates are approximately accurate with high confidence. A natural, but perhaps difficult, question is whether we can give necessary conditions on how the data must look for a given estimate to be accurate.
- The current algorithm performs well on small Stochastic Block and LFR models, but we would like to adapt it to work on larger LFR models.

*This work was supported by NSF grants CCF-1065632, CCF-1018984, EECS-1029081, CCF-1619452 and the Center for Science of Information. The authors also thank A. Kavcic and M. Mahoney for helpful discussions.