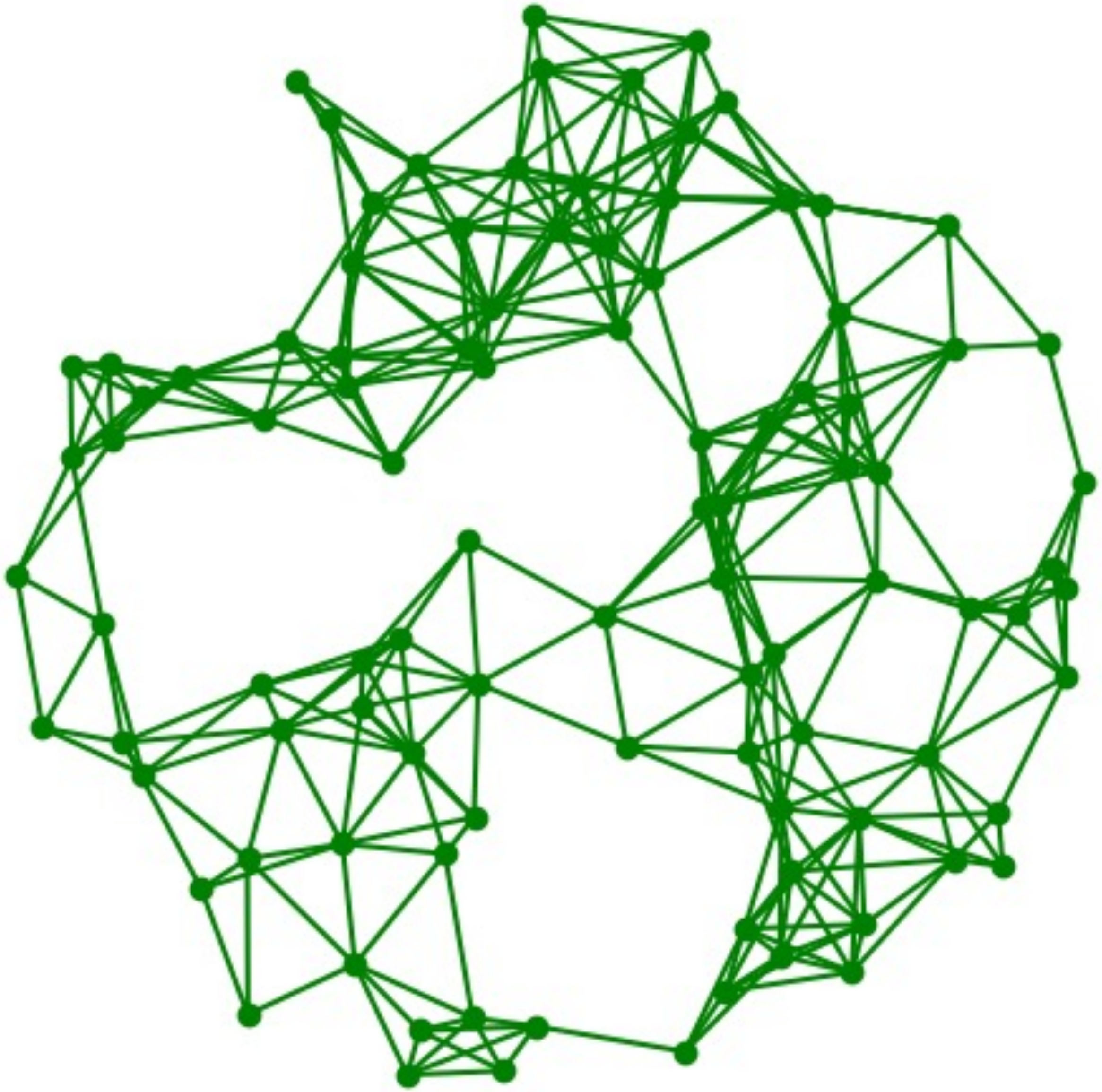
Dynamics and inference for voter model processes

Milan Vojnovic
Department of Statistics

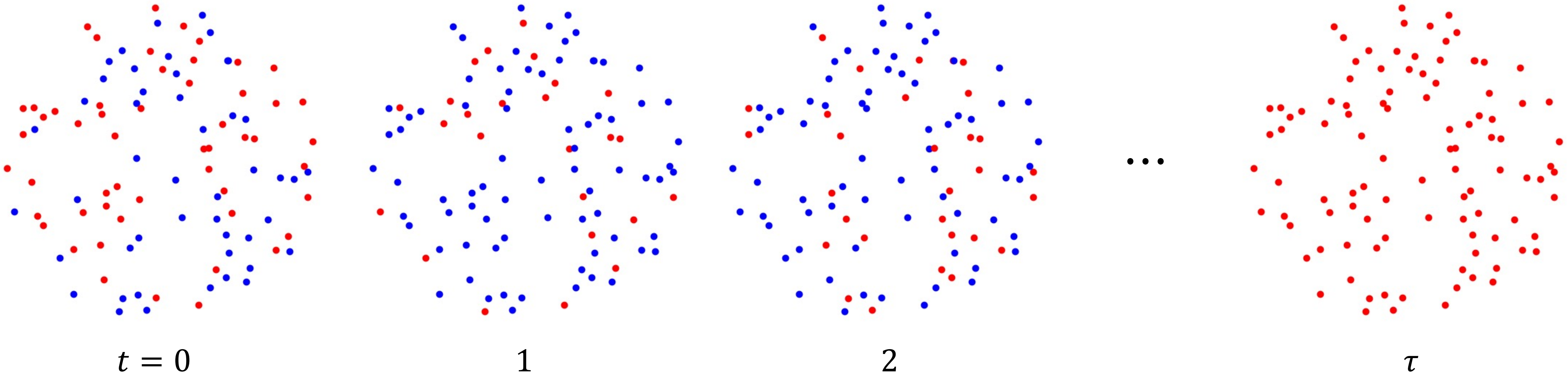
Joint work with Kaifang Zhou

LSE Department of Statistics Research Showcase, June 5-6, 2023

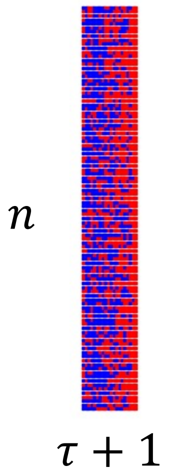


Nodes engage in pairwise interactions and upon interaction update their states
Interactions restricted by a communication graph

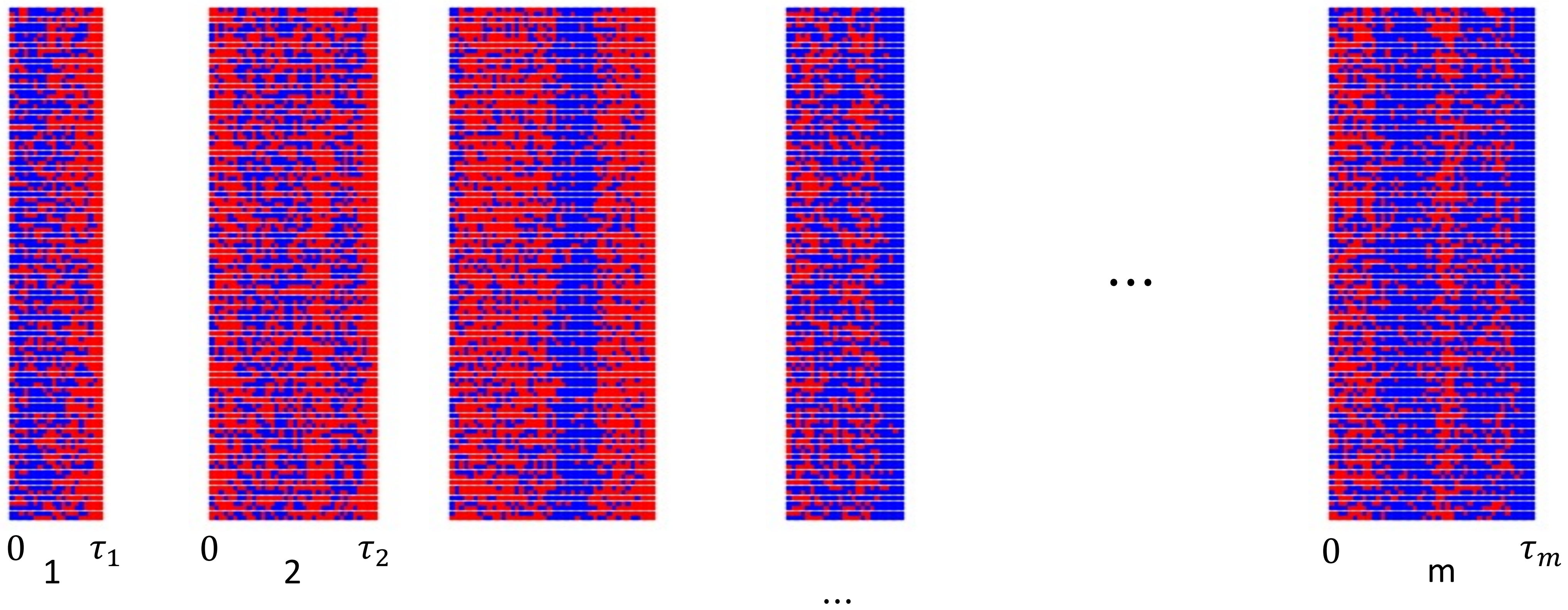
Observations: node states



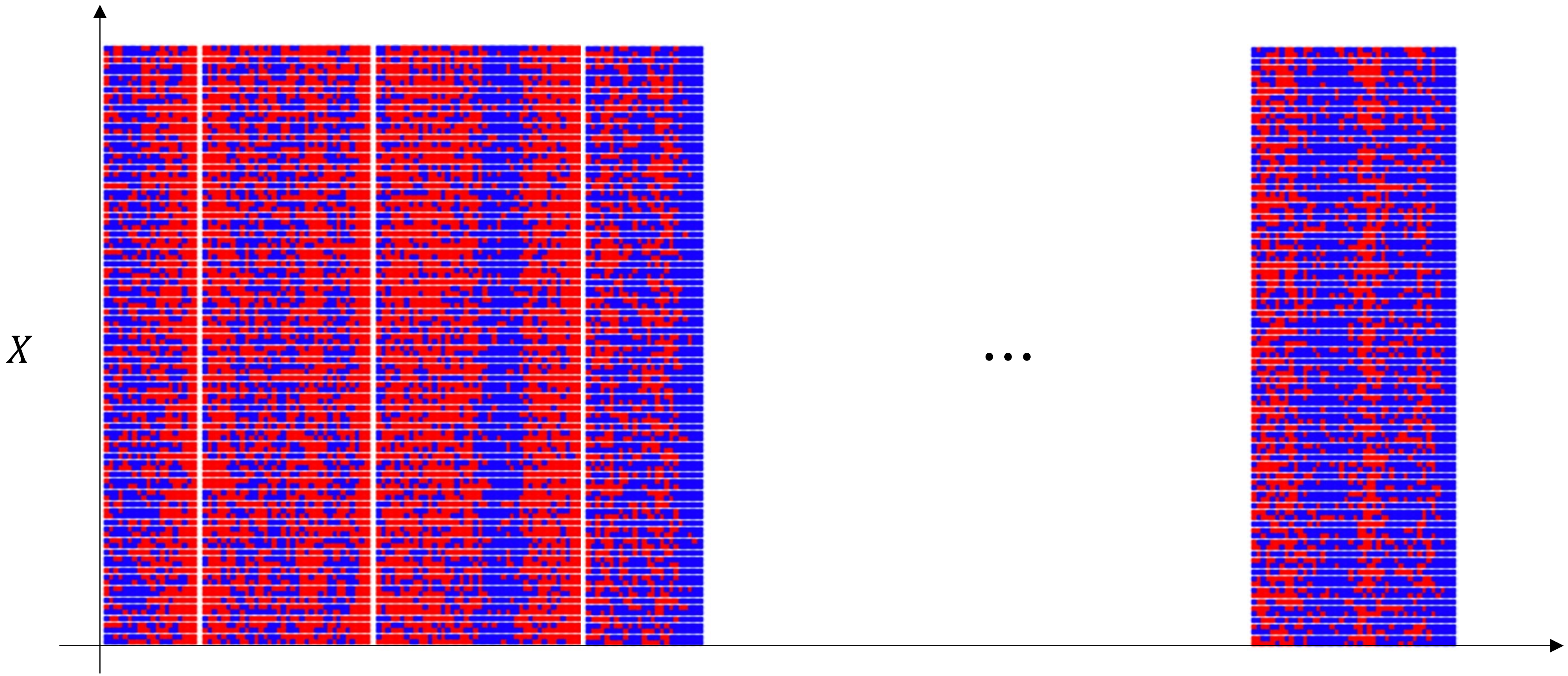
Observations:



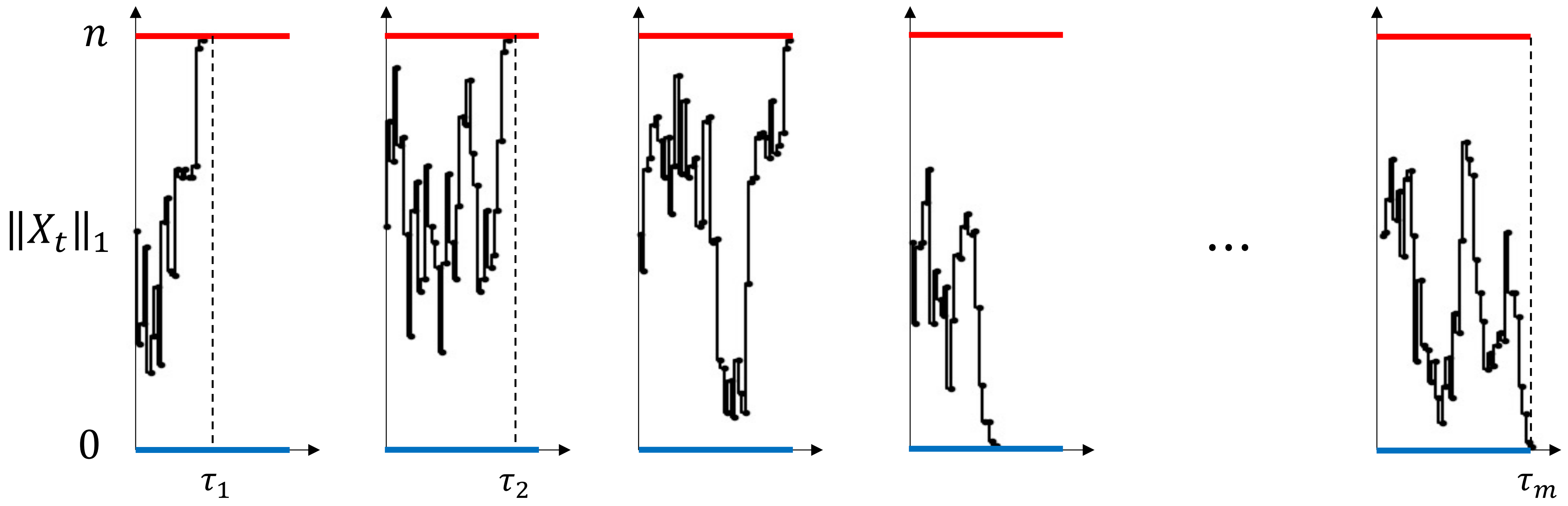
Observations: node states (cont’d)



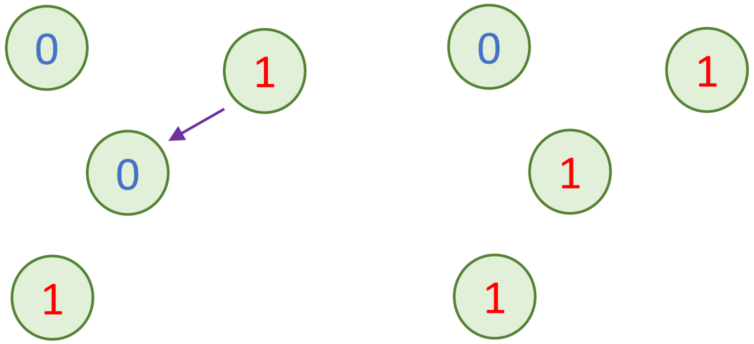
Stitching absorbing process realizations



Consensus times



Voter model



- Each node state of value 0 or 1
- At an interaction time a node adopts the state of randomly sampled neighbor

- Classical model: Holley and Liggett (1975), Liggett (1985), …

- Studied under different assumptions about interaction time instances

- Our focus:
    - Discrete-time model, in each time step every node updates its state
    - Random neighbor selection with probabilities A where the u-th row is the sampling probability distribution of node u

Dynamics and inference

- Much work has been devoted to studying dynamics of voter model processes
  - Hitting probabilities of absorption states (consensus), assuming convergence is to a consensus state
  - Hitting time (consensus time)

- Much less is known about parameter estimation (node sampling probabilities) from observed data

Voter model process

- Initial state $X_0 \sim \mu$, and

$$X_{t+1,u} \mid X_t \sim \mathrm{Ber}(a_u^\top X_t) \quad \text{for} \quad t = 0,1,\ldots, \quad u \in \{1,\ldots,n\}$$

where $A = (a_1, \ldots, a_n)^\top$ is the model parameter

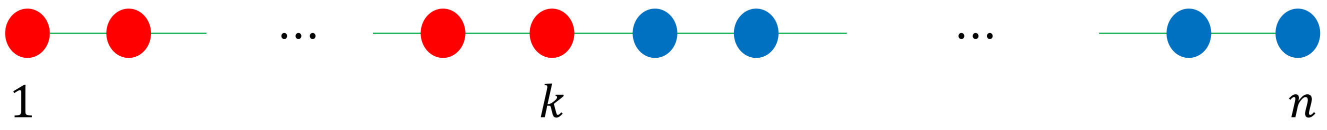
- Or, equivalently,

$$X_{t+1} = Z_{t+1} X_t \text{ for } t = 0,1,\ldots,$$

where $Z_1, Z_2, \ldots$ are i.i.d. random stochastic matrices in $\{0,1\}^{n\times n}$, $\mathbf{E}[Z_1] = A$

Parameter estimation is "hard"

- Path example
    - At each time step a random node initiates interaction
    - Communication graph is a path
    - Initial state: k nodes on one end of path in state 1, other nodes in state 0



- An interaction is *informative* only if initiated by a node with disagreeing neighbors

- Expected number of informative interactions = $k\left(\log\left(\frac{n}{k}\right)+\Theta(1)\right)$

- Number of unknown parameters: $\Theta(n)$

Challenges

- Number of observations is a priori random for any fixed number m of voter model process realizations

- Existing work focused on inference for stationary stochastic processes for a fixed number of observation points

- Some related work
  - High-dimensional generalized linear autoregressive models: Hall et al (2019)
  - Sparse multivariate Bernoulli processes in high dimensions: Pandit et al (2019)
  - Network vector autoregression: Zhu et al (2017)
  - Inferring graphs from cascades: Pouget-Abadie and Horel (2015)

Limit to a consensus state

- **Thm** [Hassin and Peleg, 2001] For any A corresponding to adjacency of a nonbipartite graph, for any initial state x,

$$\lim_{t\to\infty} \mathbf{P}[X_t = \mathbf{1}] = 1 - \lim_{t\to\infty} \mathbf{P}[X_t = \mathbf{0}] = \pi^\top x$$

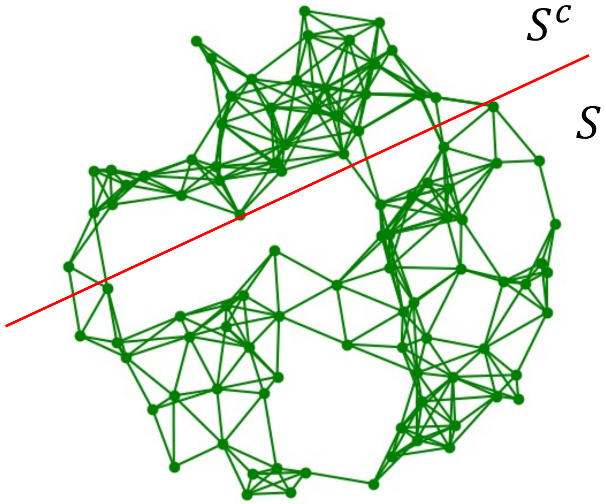
  where π is the stationary distribution for A, i.e. $\pi^\top = \pi^\top A$

- Consensus states $C = \{\mathbf{0}, \mathbf{1}\}$

Consensus time

- Hassin and Peleg (2001): $\mathbf{E}[\tau] = O(m(G)\log(n))$ where $m(G)$ is the worst-case expected meeting time for two random walks on G
- Berenbrink et al (2016): $\mathbf{E}[\tau] = O\left(\frac{1}{\Phi(G)}\frac{d(V)}{d_{\min}}\right)$ for lazy random walk
- Kanade et al (2019): $m(G) = O\left(\frac{1}{\Phi(G)} n d_{\max} \log(d_{\max})\right)$ for lazy random walks
- Lazy random walk: with probability ½ moves to a randomly chosen neighbor and otherwise remains at the current node

Graph conductance



- Graph conductance of graph $G = (V, E)$,

$$\Phi(G) = \min_{S \subset V: 0 < |S| < n} \frac{|E(S, S^c)|}{\min\{d(S), d(S^c)\}}$$

where $E(S, S^c)$ is the set of edges with vertices in S and S^c, $d(S)$ is the sum of degrees of nodes in S

- Cheeger's inequality: $\frac{\lambda_2}{2} \leq \Phi(G) \leq \sqrt{2\lambda_2}$ where λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix

$$L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

Consensus time (cont'd)

- Cooper and Rivera (2016):

$$\mathbf{E}[\tau] \le \frac{64}{\Psi_A}$$

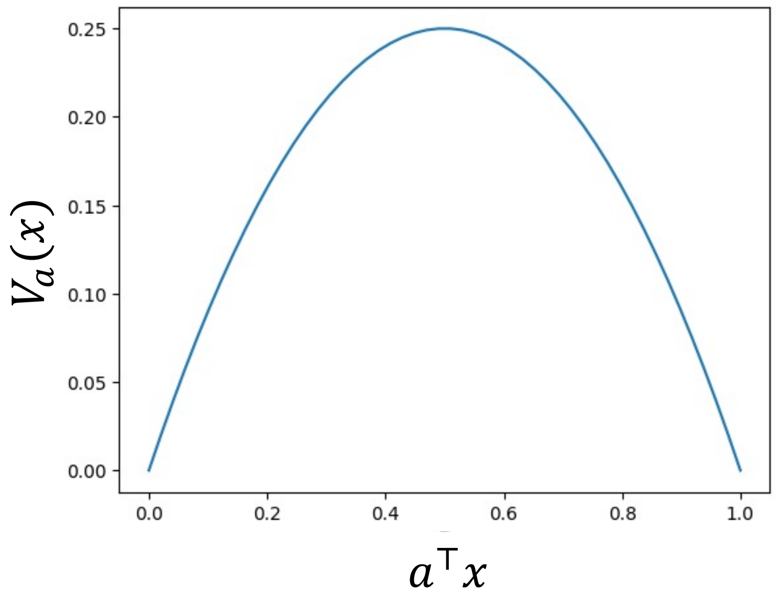
where

$$\Psi_A = \pi^* \widetilde{\Psi}_A$$

and

$$\widetilde{\Psi}_A = \min_{x \in \{0,1\}^n \setminus C} \frac{E\left[\left|\sum_{u=1}^{n} \pi_u (x_u - \sum_{v=1}^{n} Z_{u,v} x_v)\right|\right]}{\min\{\pi^\top x, 1 - \pi^\top x\}}$$

Expected consensus time bound



- **Thm** For every initial state $x \in \{0,1\}^n$,

$$\mathbf{E}_x^0[\tau] \leq \frac{1}{\Phi_A} \log\left(\frac{1}{2\pi^*}\right)$$

where

$$\Phi_A = \min\left\{\frac{\sum_{u=1}^n \pi_u^2 V_{a_u}(x)}{\pi^\top x(1 - \pi^\top x)} : x \in \{0,1\}^n, x \notin C\right\}$$

and $\pi^* = \min\{\pi_u : u = 1, \dots, n\}$

Comments on Φ_A

- Fact: $0 < \Phi_A \leq 1$

- For A according to graph G, i.e. $a_{u,v} = 1/d_u$ for $(u, v) \in E$

$$\Phi_A = \min_{S \subset V : 0 < |S| < n} \frac{|E_2(S, S^c)|}{d(S)d(S^c)}$$

where $E_2(S, S^c)$ is the set of paths of length equal to two edges, connecting a vertex in S and a vertex in S^c

Examples

- Complete graph K_n:

$$\Phi_A = \frac{n-2}{(n-1)^2} = \frac{1}{n}(1 + o(1))$$

- Cycle C_n:

$$\Phi_A = 4\frac{1}{n^2}(1 + o(1))$$

Relations between $\Phi_A, \Phi(G), \Psi_A$

- Assume $a_{u,v} = \frac{1}{2} 1_{\{u=v\}} + \frac{1}{2} \frac{1}{d_u} 1_{\{(u,v)\in E\}}$ (lazy random walk)

- Then,

$$\frac{1}{\Phi_A} \leq 2 \frac{d(V)}{d_{\min}} \frac{1}{\Phi(G)} \quad \text{and} \quad \frac{1}{\Phi(G)} \leq \frac{1}{\tilde{\Psi}_A}$$

- Hence,

$$\frac{1}{\Phi_A} \leq 2 \frac{1}{\Psi_A}$$

Exponential moment bound

- **Thm** For any $x \in \{0,1\}^n$ such that $x \notin C$ and any $\theta \in \mathbf{R}$ such that $(1 - \Phi_A)e^{\theta} \leq 1$, we have

$$\mathbf{E}_x^0\left[e^{\theta\tau}\right] \leq \frac{V_\pi(x)}{\min_{z \in \{0,1\}^n \setminus C} V_\pi(z)}$$

Proof: Follows from a general result for Markov chain hitting times.

Let $\tau_S = \min\{t > 0: X_t \in S\}$

Assume that $V: \mathcal{X} \to [1, \infty)$ is a measurable function that satisfies, for some set C and $\lambda < 1$, $\mathbf{E}[\,V(X_1) \mid X_0 = x\,] \leq \lambda V(x)$ for all $x \notin C$. Then, $\mathbf{E}_x[\lambda^{-\tau_C}] \leq V(x)$.

A probability bound

- **Thm** Let $\tau_1, \dots, \tau_m$ be consensus times of m independent realizations of voter model processes with parameter A with independent initial states according to arbitrary distributions. Then, for any $a \geq 0$,

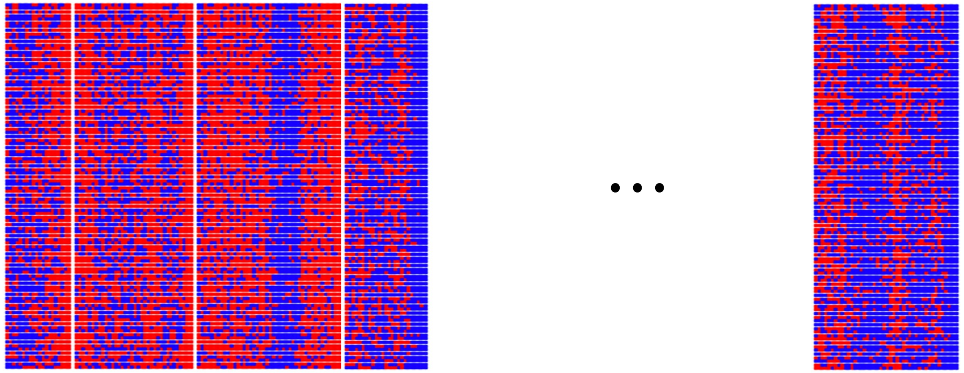
$$\mathbf{P}^0\left[\sum_{i=1}^{m} \tau_i \geq ma\right] \leq \left(\frac{\mathbf{E}^0[V_\pi(X_0)]}{\min_{z\in\{0,1\}^n\setminus C} V_\pi(z)}(1-\Phi_A)^a\right)^m$$

It follows that for any $\delta \in (0,1]$, with probability at least $1-\delta$,

$$\sum_{i=1}^{m} \tau_i \leq \frac{1}{\Phi_A}\left(m \log\left(\frac{1}{2\pi^*}\right) + \log\left(\frac{1}{\delta}\right)\right)$$

Parameter estimation

- Data: $X = \left(X_0^{(1)}, \dots, X_{\tau_1}^{(1)}, \dots, X_0^{(m)}, \dots, X_{\tau_m}^{(m)}\right)^{\top}$



- We consider maximum likelihood estimation:

$$\hat{A} \in \arg\min_{A \in \Theta} \{\mathcal{L}(A; X)\}$$

where $\mathcal{L}(A; X) = \underbrace{-\ell(A; X)}_{\text{negative log-likelihood function}} + \underbrace{\lambda_m \|A\|_{1,1}}_{\text{regularization}}$

$\|B\|_{p,q} = \left(\|b_1\|_q^p, \dots, \|b_n\|_q^p\right)^{1/p}$

Parameter estimation bound

- **Thm** Consider the voter model process with parameter A^* with support size s and $a^*_{u,v} \geq \alpha$ whenever $a^*_{u,v} > 0$ for some $\alpha > 0$. Assume that $\hat{A}$ is a minimizer of $\mathcal{L}(A; X)$ with the regularization parameter

$$\lambda_m = 2\sqrt{2}\frac{c_{n,\pi^*}}{\alpha\sqrt{\Phi_{A^*}}}\sqrt{m}$$

and m is sufficiently large (precise condition omitted). Then, for some constant $c > 0$, with probability at least $1 - 5/n$,

$$\|\hat{A} - A^*\|_F^2 \leq c\frac{s c_{n,\pi^*}^2}{\alpha^2(\Phi_{A^*}\mathbf{E}^0[\tau])^2\lambda_{\min}(\mathbf{E}[X_0X_0^\top])^2}\Phi_{A^*}\frac{1}{m}$$

where

$$c_{n,\pi^*}^2 = \left(\log\left(\frac{1}{2\pi^*}\right) + \log(2n^3)\right)\log(4n^3)$$

Proof sketch

- Proof is based on the framework of M-estimators with decomposable regularizers (Negahban et al 2012, Wainwright 2019)

- **Thm** Assume that loss function $\mathcal{L}(A; X)$ has the regularization parameter such that

  (C1) $\lambda_m \geq 2\|\nabla \ell(A^*)\|_\infty$

  and

  (C2) for some $S \subseteq V^2$, $-\ell(A; X)$ satisfies the restricted strong convexity (RSC) condition relative to A^* and S with curvature $\kappa > 0$ and tolerance γ^2

  Then,

$$\|\hat{A} - A^*\|_F^2 \leq 9|S| \left(\frac{\lambda_m}{\kappa}\right)^2 + \left(2\gamma^2 \frac{1}{m} + 4\|A^*_{S^c}\|_{1,1}\right) \frac{\lambda_m}{\kappa}$$

RSC condition

- A loss function $\mathcal{L}$ is said to satisfy the RSC relative to A^* and S with curvature $\kappa > 0$ and tolerance γ^2 if

$$\mathcal{E}(\Delta) \geq \kappa \|\Delta\|_F^2 - \gamma^2 \text{ for all } \Delta \in \mathcal{C}(S; A^*)$$

where

$$\mathcal{E}(\Delta) = \mathcal{L}(A^* + \Delta) - \mathcal{L}(A^*) - \nabla\mathcal{L}(A^*)^\top vec(\Delta) \qquad \text{(first-order Taylor error)}$$

and

$$\mathcal{C}(S; A^*) = \left\{\Delta \colon \|\Delta_{S^c}\|_{1,1} \leq 3\,\|\Delta_S\|_{1,1} + 4\left\|A^*_{S^c}\right\|_{1,1}\right\}$$

Condition (C1)

- **Lem** For any $\delta \in (0,1]$ and any $m \geq 1$ independent realizations of the voter model process with parameter A^* and initial value distribution μ, with probability at least $1 - \delta$,

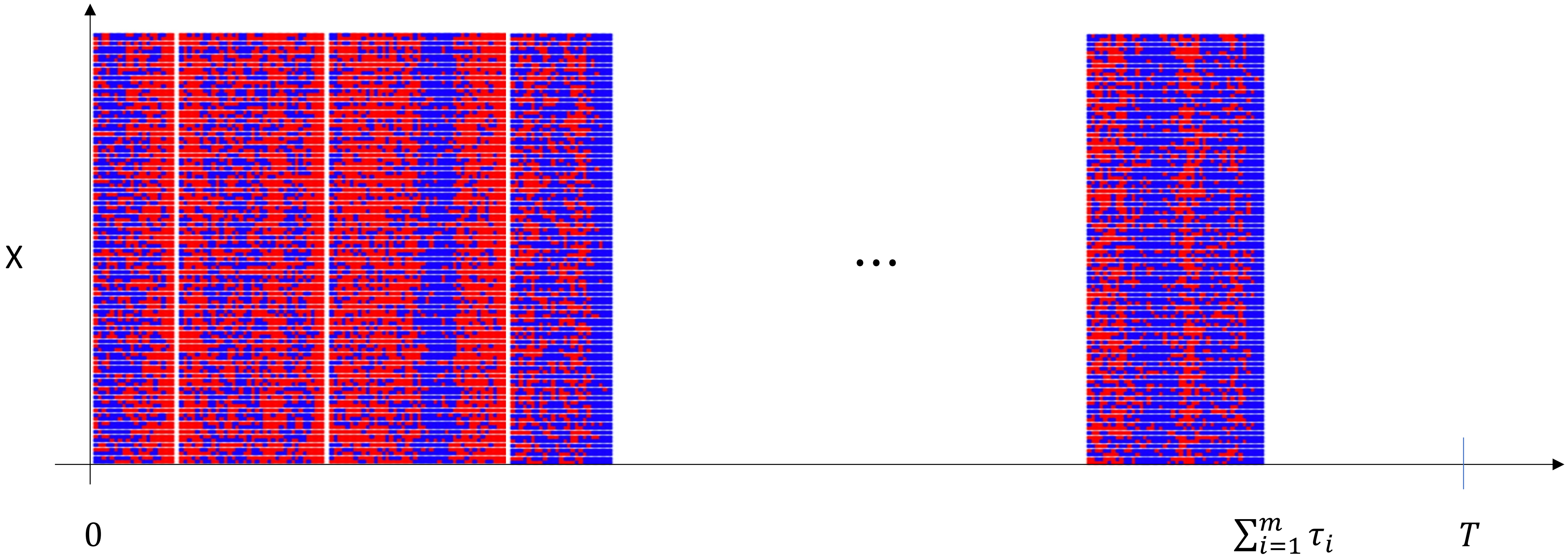
$$\|\nabla \ell(A^*)\|_\infty \leq \sqrt{2}\frac{1}{\alpha}\frac{1}{\sqrt{\Phi_{A^*}}}\sqrt{m}c_{n,\delta,\pi^*}(m)$$

where

$$c_{n,\delta,\pi^*}(m)^2 = \left(\log\left(\frac{1}{2\pi^*}\right) + \frac{1}{m}\log\left(\frac{2n^2}{\delta}\right)\right)\log\left(\frac{4n^2}{\delta}\right)$$

Proof: Using a truncation argument, consensus time probability tail bound, and Azuma-Hoeffding's inequality for bounded-difference martingale sequences

Truncation argument in a picture



$$\mathbf{P}\left[E_{\sum_{i=1}^{m} \tau_i}\right] \leq \mathbf{P}[E_T] + \mathbf{P}\left[\sum_{i=1}^{m} \tau_i \geq T\right]$$

Some event

Condition (C2)

- Show

$$\mathcal{E}(\Delta) \geq h(\Delta; X) \coloneqq \sum_{i=1}^{m} \sum_{t=0}^{\tau_i - 1} \sum_{u=1}^{n} \left(\Delta_u^\top X_t^{(i)} \right)^2$$

- Then show **(C2')**: $h(\Delta; X)$ satisfies the RSC condition with high probability

Condition (C2')

- **Step 1:** $\mathbf{E}^0[h(\Delta; X)] \geq \kappa_1 \|\Delta\|_F^2$ for all Δ where

$$\kappa_1 \leq m\mathbf{E}^0[\tau]\lambda_{\min}(\mathbf{E}[X_0 X_0^\top])$$

- **Step 2:** For any $\delta \in (0,1/2]$, any S such that $|S| \leq s$ and any $\Delta \in \mathcal{C}(S; A^*)$, $h(\Delta; X) \geq \frac{\kappa_1}{2}\|\Delta\|_F^2$ with probability at least $1-\delta$ provided that

$$m \geq \frac{s^2}{\Phi_{A^*}} \frac{1}{\mathbf{E}^0[\tau]^2 \lambda_{\min}(\mathbf{E}[X_0 X_0^\top])^2} c_{\delta,\pi^*}(m)$$

where

$$c_{\delta,\pi^*}(m) = 8\left(\log\left(\frac{1}{2\pi^*}\right) + \frac{1}{m}\log\left(\frac{2}{\delta}\right)\right)\log\left(\frac{2}{\delta}\right)$$

Condition (C2') cont'd

- **Step 3:** Show that

$$\mathbf{P}\left[h(\Delta; X) \geq \kappa' \|\Delta\|_F^2 - \gamma'^2 \text{ for all } \Delta \in \mathcal{C}(S; A^*)\right] \geq 1 - \frac{4}{n}$$

To show this, we apply some set covering arguments and combine with the bound in Step 2

Conclusion

- Shown a new bound on consensus time, in expectation and probability
- Shown new parameter estimation bounds for absorbing voter model processes, obtained by leveraging the consensus time bounds