Efficient Near-Optimal Testing of Community Changes in Balanced Stochastic Block Models
Aditya Gangrade\textsuperscript{a}, Praveen Venkatash\textsuperscript{b,*}, Bobak Nazer\textsuperscript{c}, Venkatesh Saligrama\textsuperscript{d}

\textsuperscript{a}Boston University
\textsuperscript{b}Carnegie Mellon University

\textsuperscript{a}under labeled 1 or -1, summarized as a partition \( x \in \{-1,1\} \)
• Sparse communities are balanced \( \sum_{v} x_{v} = 0 \)
• Edge within a community \( p_{v,w} = p_{v,v} \quad \text{and across communities} \quad p_{v,w} = p_{v,v} \)

\( \sum_{v} x_{v} = 0 \)

Community Goodness-of-Fit
• Community Goodness-of-Fit: We are given the

\[ \text{Partition } \pi \quad \text{and the } \quad \text{Graph from Partition } \pi \]

\[ \text{and would like to determine if } \pi = \pi' \text{ if the partitions differ in one or more positions.} \]
\[ \text{For which } \pi \text{ values can we find a mapping with vanishing minimax risk}\]

\[ \mathbb{D}(G, \pi) = \mathbb{E}(\mathcal{L}(x, \pi)) \]

Theorem
• Small changes, \( \pi \neq \pi' \) Unfavoring real possible \( (\pi' \subgraph \pi) \)
• Large changes, \( \pi \neq \pi' \) Favoring real possible \( (\pi \subgraph \pi') \)

Community Two-Sample Testing
• Community Two-Sample Testing: We are given the

\[ \text{Graph from Partition } \pi \quad \text{and the } \quad \text{Graph from Partition } \pi' \]

\[ \text{and would like to determine if } \pi = \pi' \text{ if the partitions differ in one or more positions.} \]
\[ \text{For which } \pi \text{ values can we find a mapping with vanishing minimax risk}\]

\[ \mathbb{D}(G, \pi) = \mathbb{E}(\mathcal{L}(x, \pi)) \]

Community Testing
Graph from Partition \( \pi \)

\[ H_{\pi} = 0 \]

Graph from Partition \( \pi' \)

\[ H_{\pi'} = 0 \]

Community Two-Sample Testing: Avoidability
• Avoidability: \( \pi \neq \pi' \)

\[ \mathbb{D}(G, \pi) = \mathbb{E}(\mathcal{L}(x, \pi)) \]

Community Two-Sample Testing: Converse
• Avoided-fit criteria applies by reminding \( \pi \) to be a testable \( \Delta_{2} \) for large changes

\[ \text{If } \text{not in list of balanced partitions } \]

\[ \text{Then } \text{is a non-trivial} \]

\[ \text{and across communities } \quad p_{v,w} = p_{v,v} \]

\[ \sum_{v} x_{v} = 0 \]

Community Goodness-of-Fit: Converse
• Avoided-fit of alternate \( q < v \leftarrow x \leftarrow \text{L}(x, \pi) \mathbb{P}(G|H_{\pi})(x) \)

\[ \text{Method: } \text{Pick a } q \text{ on the alternate. The Bayes risk } R_{q} \text{ of the problem with this prior is a lower bound on the minimax} \]

\[ \text{Risk for a community model converges prior on edge alternates} \]

\[ R_{2} \geq 1 - \mathbb{P}(G, \pi) + D_{2}\left(\pi, \pi' \right) \text{KL divergence} \]

\[ \text{Computing } R_{2} \text{ only needs to know } \mathbb{P}(G|H_{\pi})(x) \text{ which is discovered by the overlap of a } q \text{ and } \pi' \text{ as well as } \pi \]

\[ \text{Overlap is hypervolumetric distributed under our prior, which is substantially dominated by an appropriate Shannon distribution} \]

Community Goodness-of-Fit: Achievability
• Community Goodness-of-Fit: Achievability

\[ \text{Partition } \pi \quad \text{and the } \quad \text{Graph from Partition } \pi \]

\[ \text{and would like to determine if } \pi = \pi' \text{ if the partitions differ in one or more positions.} \]
\[ \text{For which } \pi \text{ values can we find a mapping with vanishing minimax risk}\]

\[ \mathbb{D}(G, \pi) = \mathbb{E}(\mathcal{L}(x, \pi)) \]

Community Recovery and Detection
• Let \( \mathbf{d} \) be the estimate given the graph \( G \)

\[ \text{Distortion with best testable } \mathbf{d} \quad \text{with } \mathbb{D}(G, \pi) = \mathbb{D}(G, \pi') \]

\[ \text{Exact Recovery: } \mathbf{d}(x, \pi) = 1 - x \]

\[ \text{Partial Recovery: } \mathbf{d}(x, \pi) = 1 - x \quad \text{for } x \in \{0,1\} \]

\[ \mathbb{P}(G|H_{\pi})(x) = \mathbb{P}(G|H_{\pi})(x') \]

\[ \text{Weak Recovery / Detection: } \mathbf{d}(x, \pi) = 1 - x \quad \text{for } x \leq \epsilon, \text{ and } G \text{ is detectable if only } \]

\[ \epsilon > \epsilon_{0} \]

Community Goodness-of-Fit: Partial Recovery
• Partial Recovery: \( \mathbf{d} \) graph from \( \pi' \) or \( \text{to } \mathbb{D}(G, \pi') \)

\[ \text{Graph from Partition } \pi' \]

\[ \mathbf{d} = 0 \quad \text{Signalling risk possible} \]

\[ \text{Graph from Partition } \pi \]

\[ \mathbf{d} > 0 \quad \text{Vanishing risk possible} \]

Community Two-Sample Testing: Converse
• Avoided-fit criteria applies by reminding \( \pi \) to be a testable \( \Delta_{2} \) for large changes

\[ \text{If } \text{not in list of balanced partitions } \]

\[ \text{Then } \text{is a non-trivial} \]

\[ \text{and across communities } \quad p_{v,w} = p_{v,v} \]

\[ \sum_{v} x_{v} = 0 \]

Community Goodness-of-Fit: Achievability
• Community Goodness-of-Fit: Achievability

\[ \text{Partition } \pi \quad \text{and the } \quad \text{Graph from Partition } \pi \]

\[ \text{and would like to determine if } \pi = \pi' \text{ if the partitions differ in one or more positions.} \]
\[ \text{For which } \pi \text{ values can we find a mapping with vanishing minimax risk}\]

\[ \mathbb{D}(G, \pi) = \mathbb{E}(\mathcal{L}(x, \pi)) \]

References