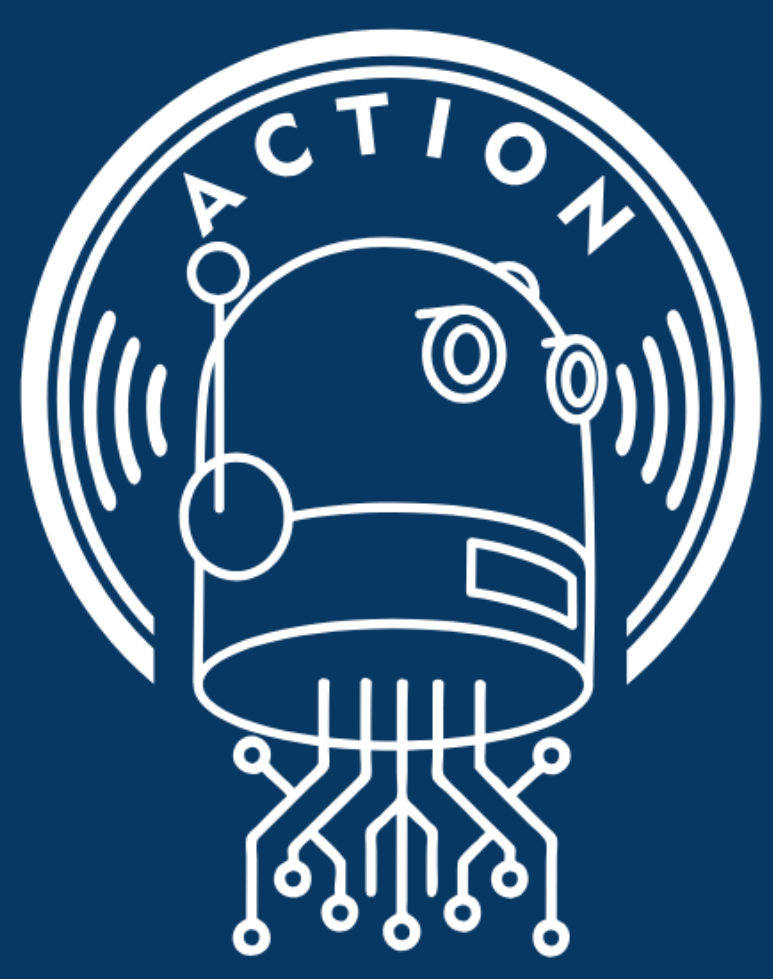




Bayesian Optimization-Based Explainer For Temporal Graph Neural Networks

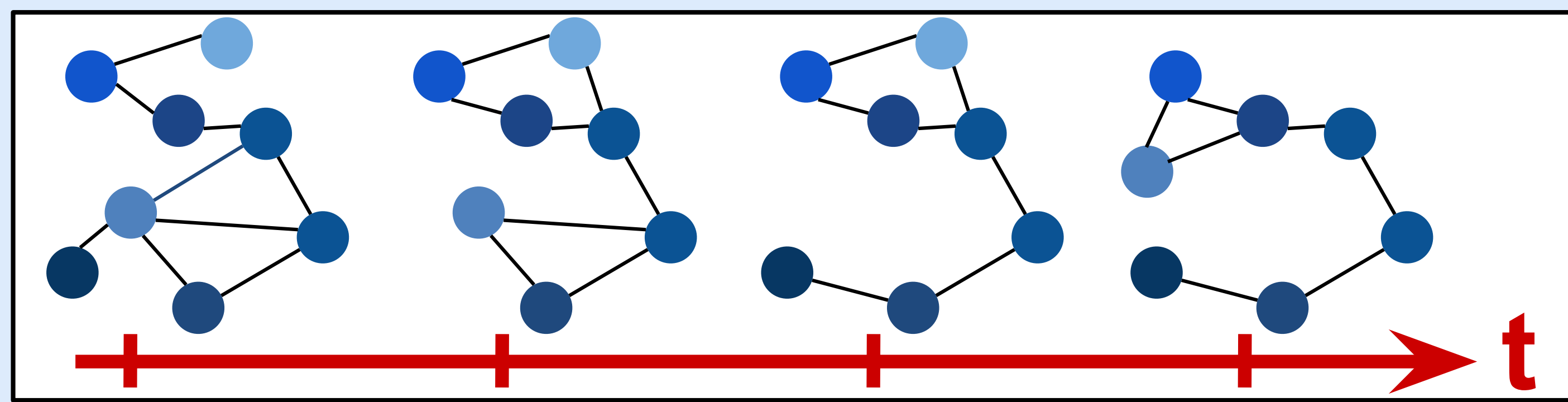
UC SANTA BARBARA

Chenrui Zhang¹, Varsha Narasiman¹, Ambuj K. Singh¹

1. Dynamo Lab, UC Santa Barbara

1 Background

- Temporal graphs change in structure or features over time.
- Temporal Graph Neural Networks (TGNNs) are designed to process dynamic data like temporal graphs.
- Applications of TGNNs: evolving user recommendations, traffic systems, and financial fraud detection.

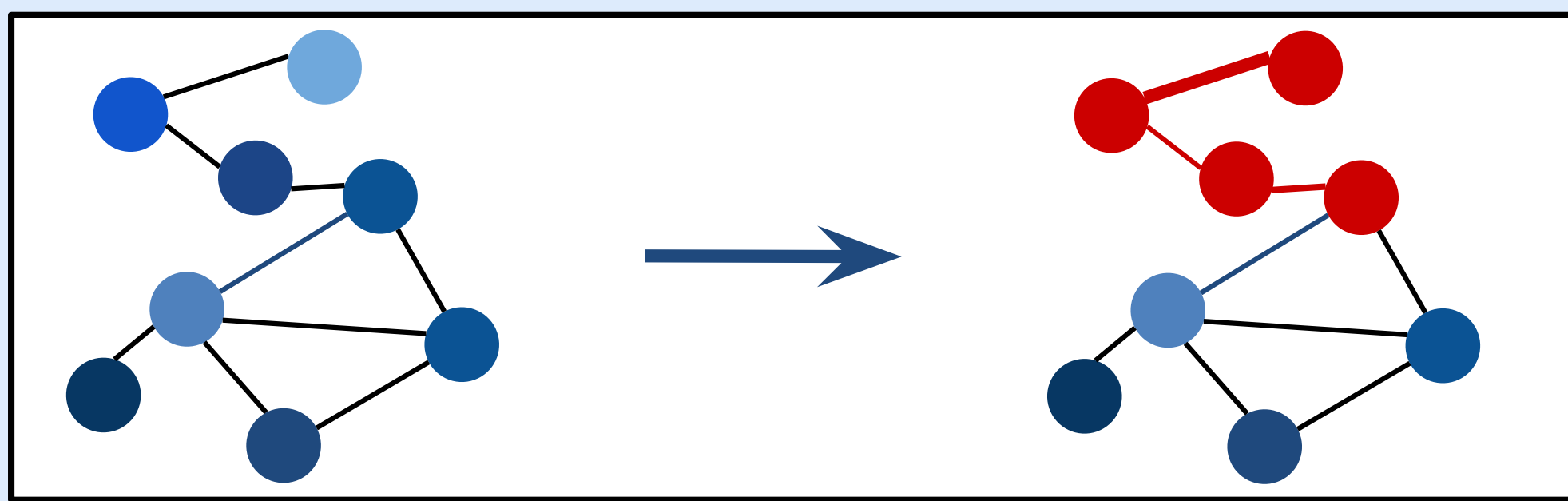


An example of a temporal graph

- Inner workings of Graph Neural Networks (GNNs), including TGNNs, remain obscure and not readable for humans.
- It is hard to trust the prediction of a TGNN unless it can identify the sequence of events that caused its decision.

2 Explainers

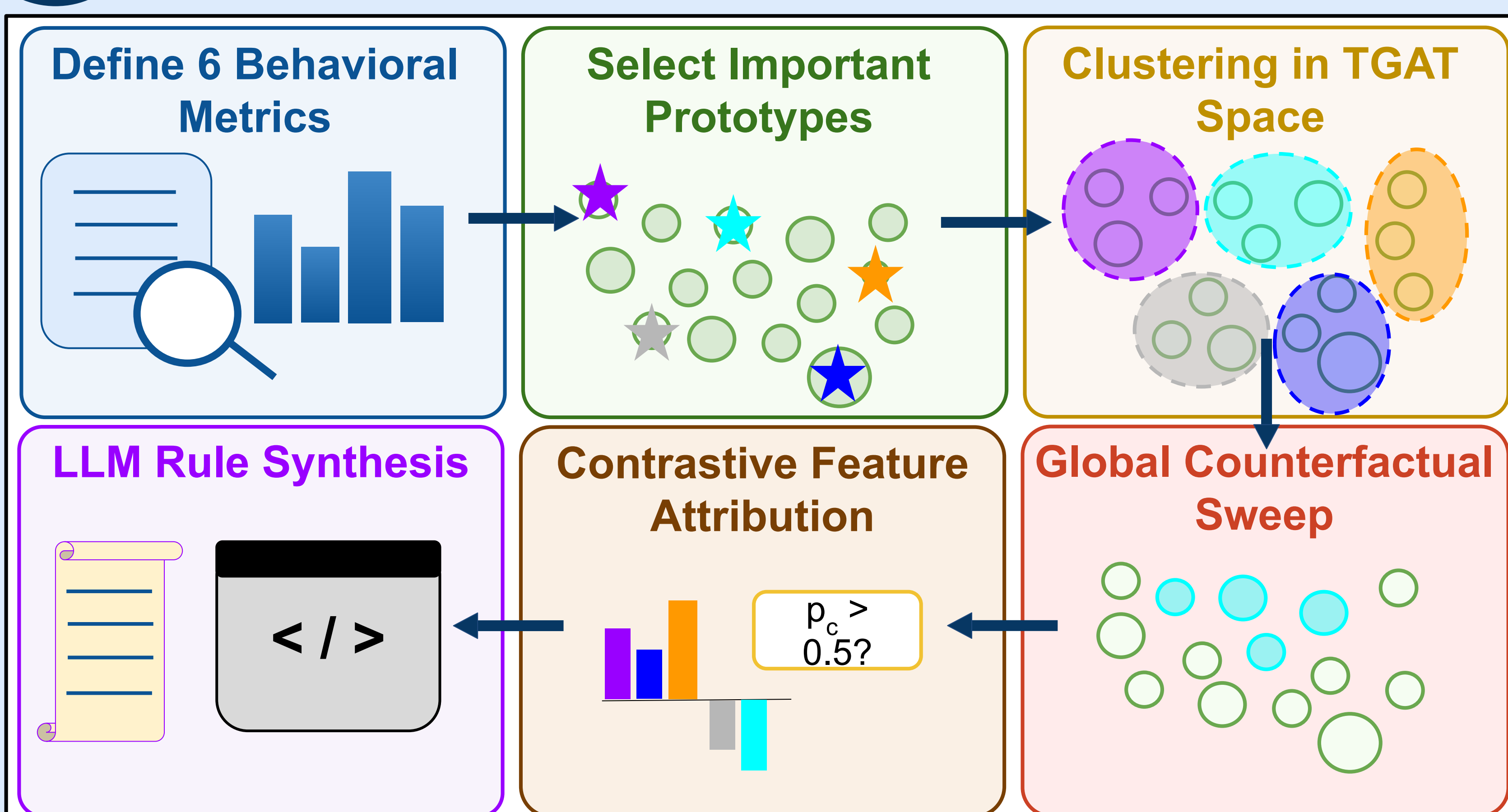
- GNN Explainers are graphs that identify the most key features the GNN used for prediction.



The highlighted part (in red) is the subgraph identified by the explainer.

- Factual explainers identify subgraphs that are sufficient to preserve the original prediction.
- Counterfactual explainers identify the minimal perturbation necessary to flip the original prediction.

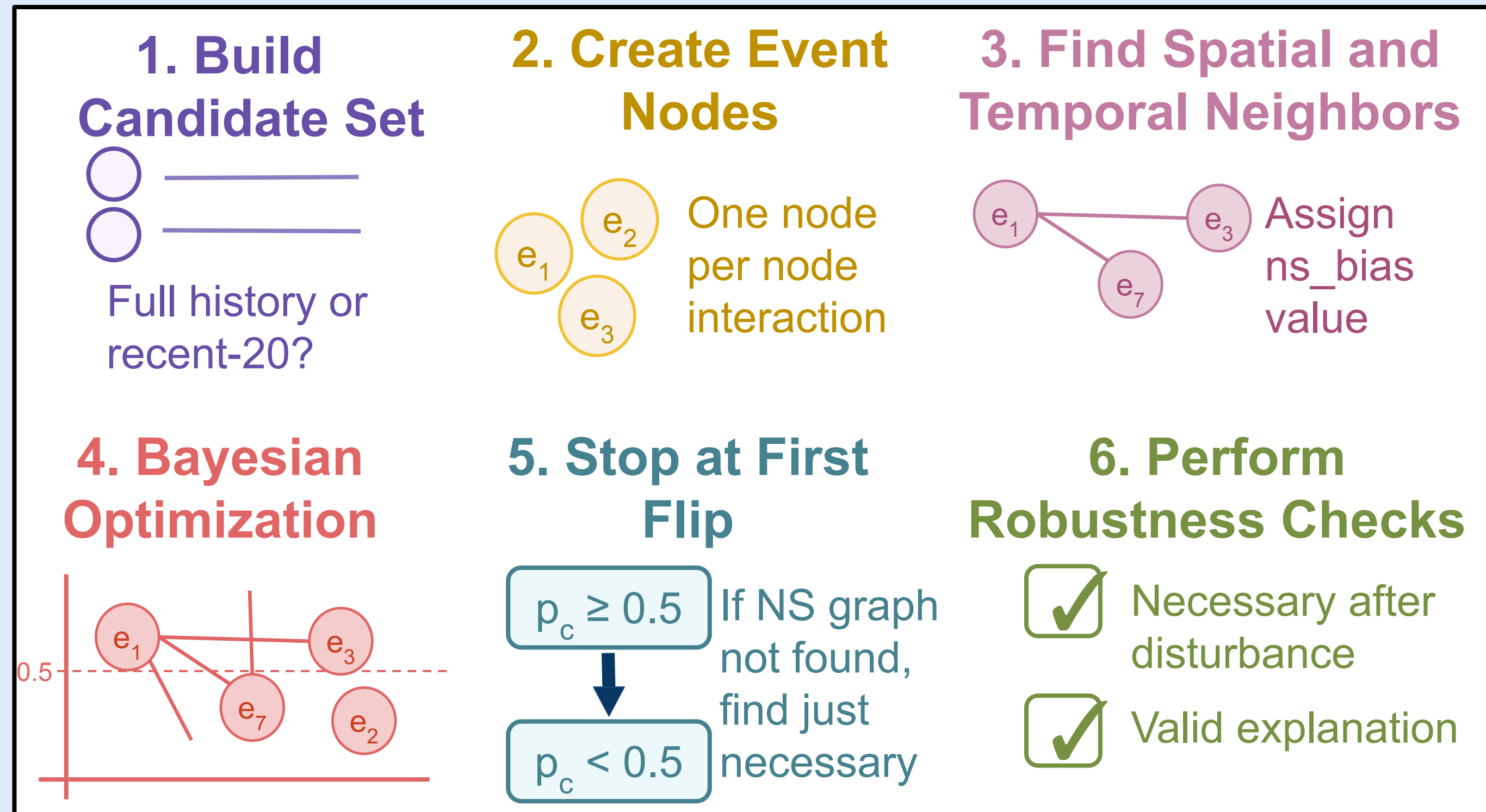
3 Existing GLoSS-T Pipeline



THE PROBLEM: How to produce global explanations for TGNNs that are semantically meaningful, sufficient, and necessary.

4 TNS-BO Method

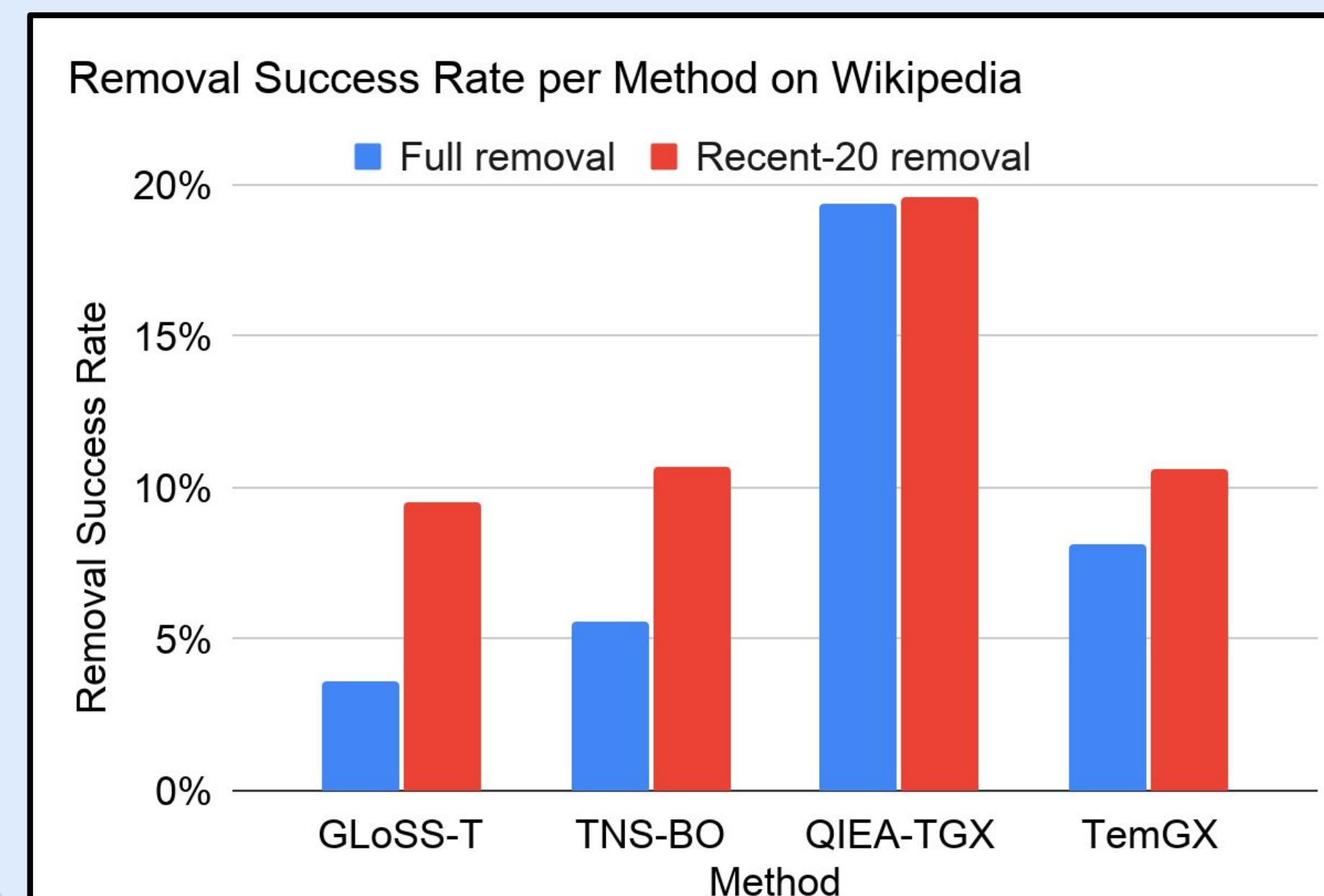
We propose the Temporal Necessity-Sufficient Bayesian Optimization (TNS-BO) method that identifies factual (sufficient) and counterfactual (necessary) subgraphs through a Bayesian Optimization search.



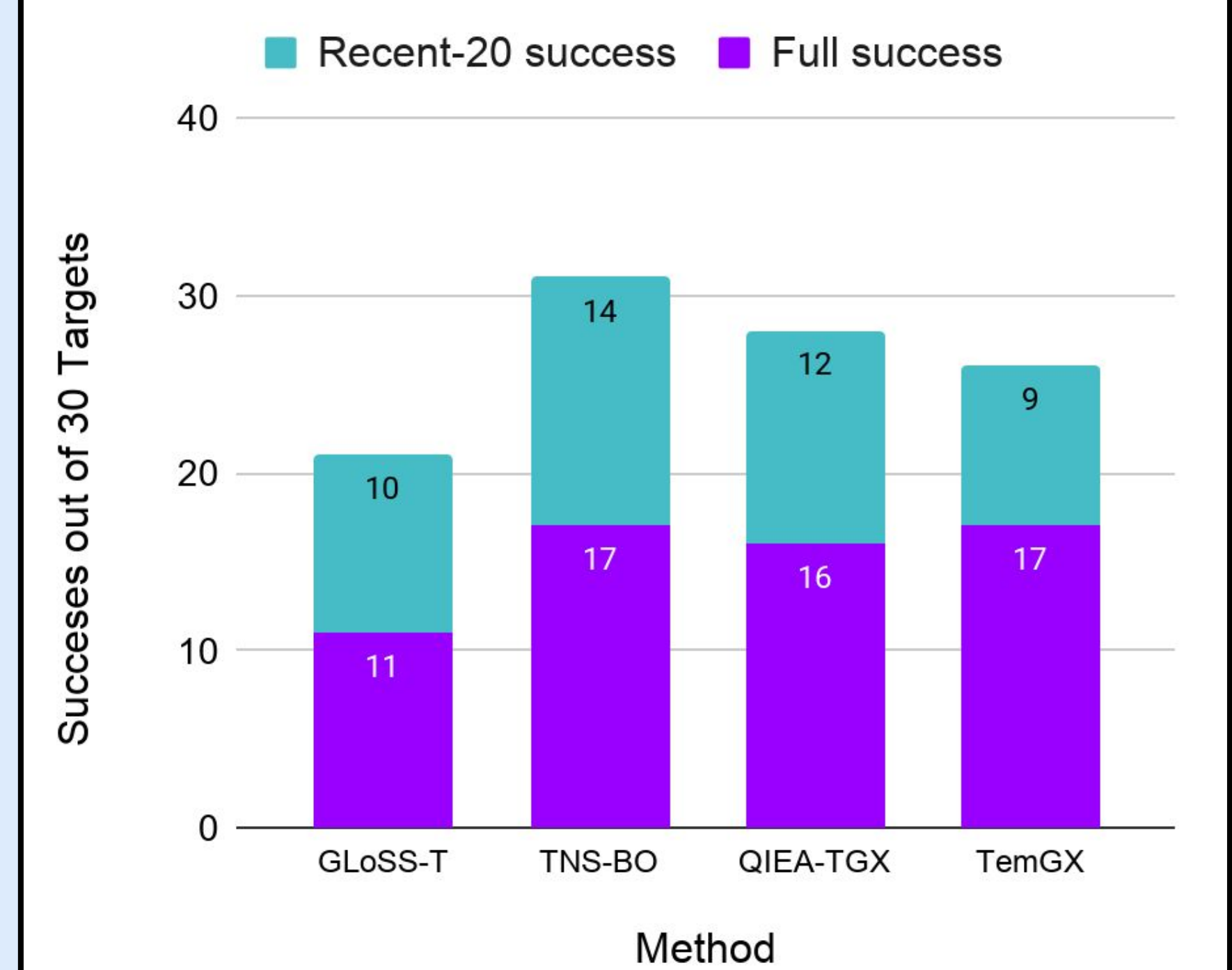
We introduce the NSBias technique, which weights the BO search to consider nodes with more spatial and temporal importance.

5 Results

This chart shows the pruning rate of the counterfactual graph when full and recent-20 histories are available.



Full success and Recent-20 success on Wikipedia



This chart shows number of targets, or nodes, for which a valid counterfactual was found.

6 Analysis

- TNS-BO with NSBias outperforms GLoSS-T [1], QIEA-TGX [2], and TemGX [3] on the Wikipedia dataset on target successes.
- Pruning the explanation remains a challenge, as QIEA-TGX outperforms TNS-BO.
- However, TNS-BO aims to find the necessary and sufficient graph, using the counterfactual graph only as a fallback, whereas other methods focus on finding the counterfactual graph.

7 Citations

- [1] Anonymous Author(s), "GLoSS: Natural Language Counterfactuals for GNNs via LLM Latent Steering," *Proceedings of the 40th Conference on Neural Information Processing Systems (NeurIPS)*, 2026, submitted.
- [2] Mitani, M. and Sasaki, Y., "Explaining Temporal Graph Neural Network via Quantum-Inspired Evolutionary Algorithm," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40, No. 29, pp. 24396–24404, 2026, doi:10.1609/aaai.v40i29.39621.
- [3] Lu, M., Che, H., Fan, Y., Liu, Q., Shao, F., Ge, T., Xiao, X., and Wu, Y., "Training-free Counterfactual Explanation for Temporal Graph Model Inference," *Proceedings of the Fourteenth International Conference on Learning Representations (ICLR)*, 2026. Available at: <https://openreview.net/forum?id=NqTYz3A8tQ>.