Conformal Edge-Weight Prediction in Latent Space

Akash Choudhuri[1], Yongjian Zhong[1], Mehrdad Moharrami[1], Christine Klymko[2], Mark Heimann[2], Jayaraman J Thiagarajan[3], Bijaya Adhikari[1]

[1]University of Iowa, [2]Lawrence Livermore National Laboratory, [3]Apple Inc.

Lawrence Livermore National Laboratory

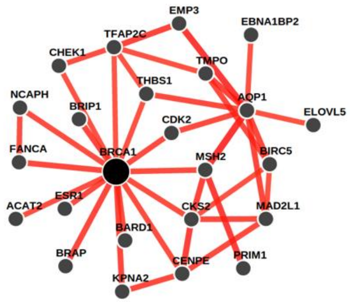
The University of Iowa

Overview

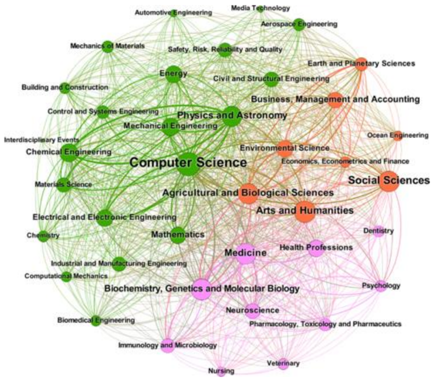
- Motivation
- Problem Formulation
- Our Approach
- Experiments
- Conclusion

Graphs are Everywhere

Transportation Networks

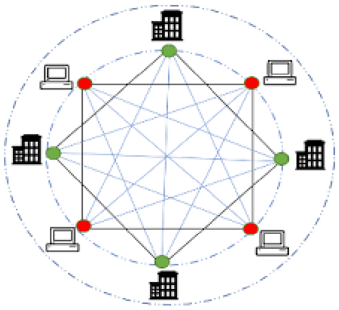


Biological Networks

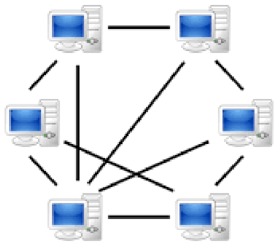


Collaboration Networks

Social Networks



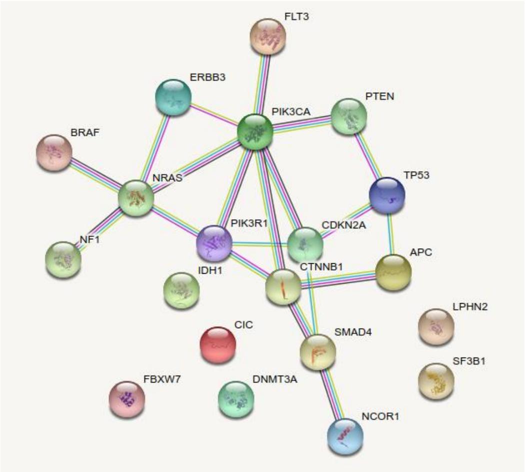
Blockchain Networks



P2P Networks

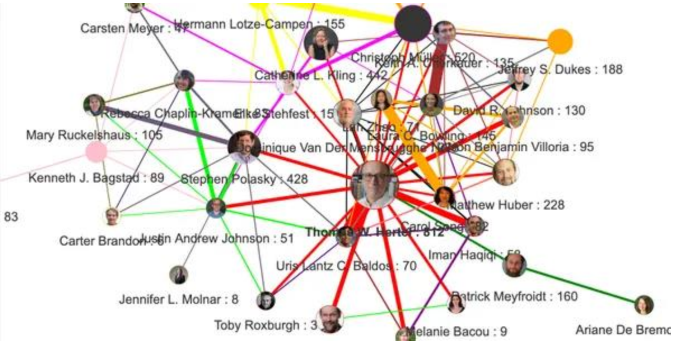
Predictive Tasks on Graphs

- Biological Networks:
  - Node Level: What is the function of a given protein?
  - Edge Level: What is the strength of interaction between a pair of proteins?



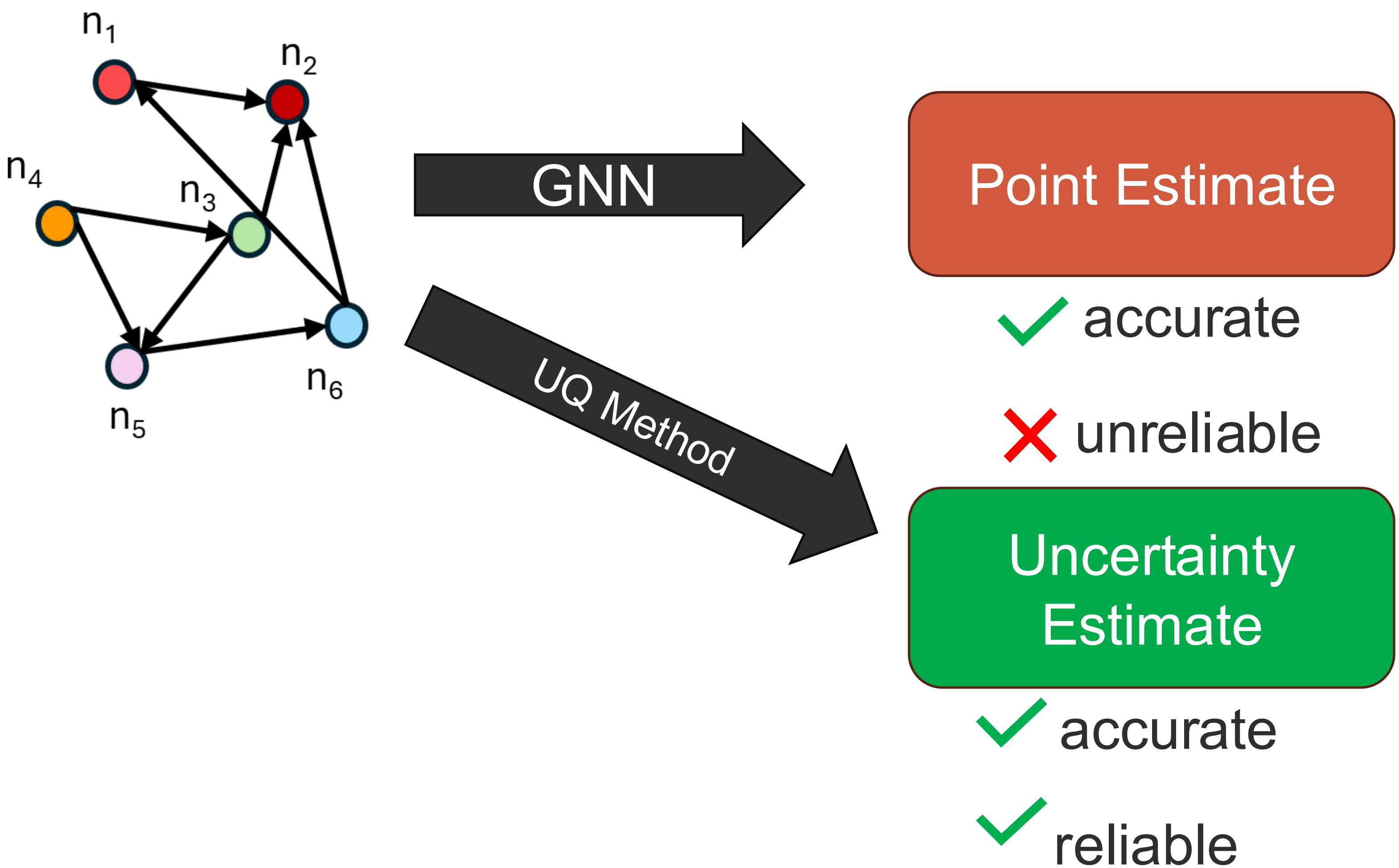
https://string-db.org/cgi/help?sessionId=b3v9M4cmNN9X

- Collaboration Networks:
  - Node Level: What is the subfield of work of a researcher?
  - Edge Level: How many publications are co-authored by 2 researchers?



https://www.sciencecollaborations.net/

Need for Uncertainty Quantification

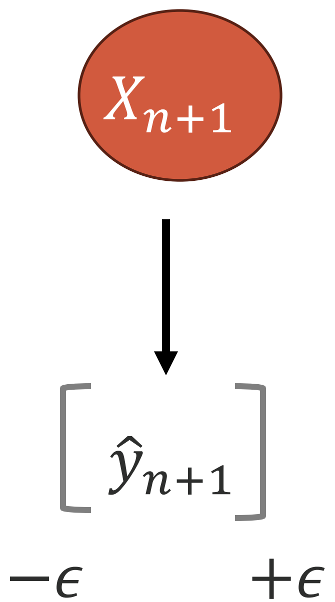


The University of Iowa

Standard Approach for Uncertainty Quantification

Test Sample		Model Point Prediction		Prediction Band
X_{n+1}	$\longrightarrow$	$\hat{y}_{n+1}$	$\longrightarrow$	$C(X_{n+1})$

Node Level



- No statistical guarantees
- Many assumptions
- Prediction bands may not typically cover the true label (a.k.a. miscoverage)

Conformal Inference

- Assuming the test point X_{n+1} is **exchangeable** with the other points, the coverage guarantee with the true label y_{n+1} holds:

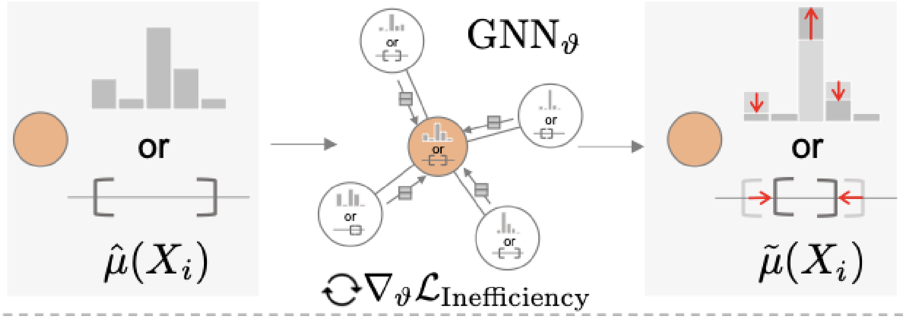
$$\mathbb{P}(y_{n+1} \in C_{1-\alpha}(X_{n+1})) \geq 1 - \alpha$$

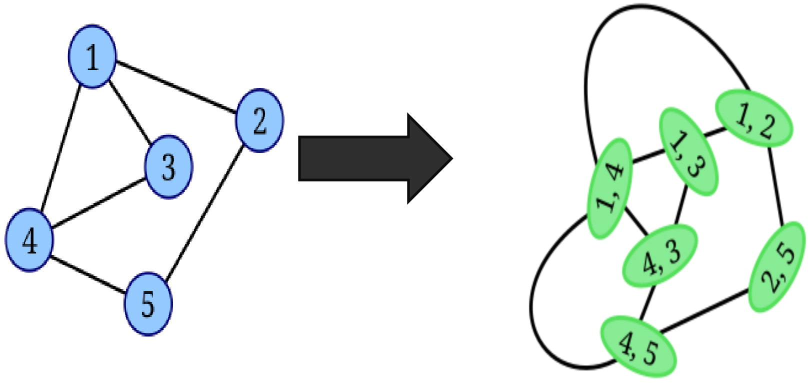
- Provides a statistical guarantee of coverage for the predicted bands

- This guarantee is:
  - Distribution free
  - Model agonistic

Need for Conformal Edge Weight Inference

- Prior works primarily on nodes
- **[Huang et al., 2023]:**
  - Exchangeable and differentiable topology-aware loss on the calibration **node data**
  - Minimize the size of the prediction set while maintaining coverage
- Conformal line-graph inference:
  - Will not scale to larger graphs
  - Ambiguity about edge features



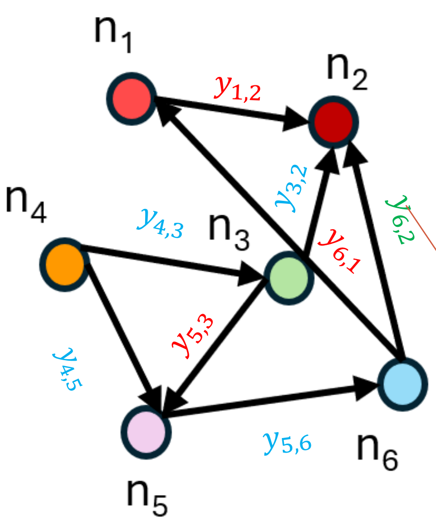


Overview

- Motivation
- Problem Formulation
- Our Approach
- Experiments
- Conclusion

Problem Formulation

$G = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{Y})$



Given:

Training Data: $D_{\text{train}} = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{Y}_{\text{train}})$

Calibration Data: $D_{\text{cal}} = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{Y}_{\text{cal}})$

Estimate:

Testing Data: $D_{\text{test}} = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{Y}_{\text{test}})$

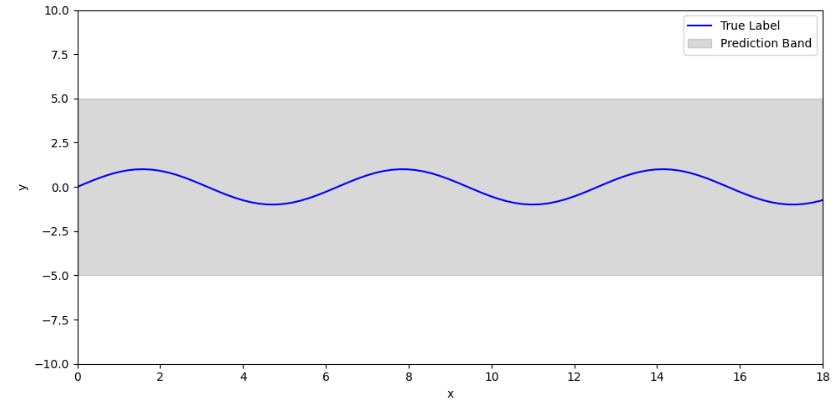
Prediction set

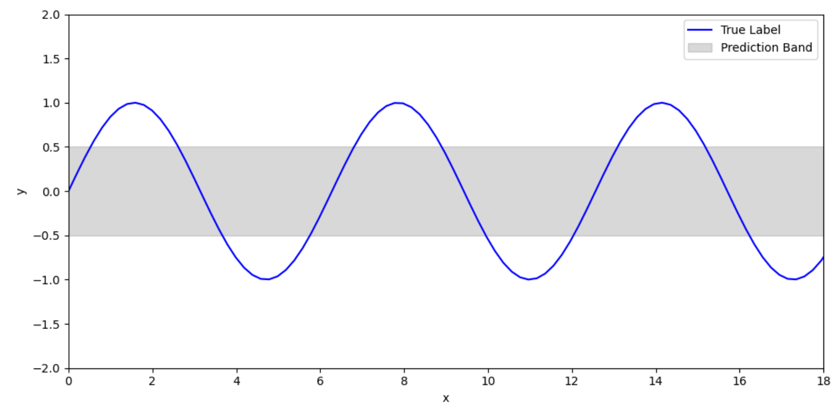
$$\left[\ \hat{y}_{6,2}\ \right]$$

$-\epsilon$ $+\epsilon$

Challenges

- Trivially predict infinitely large bands
  - Smallest possible band
- Trivially predict very small bands
  - Given a significance level (α), true labels of test point present $(1-\alpha)$% of time





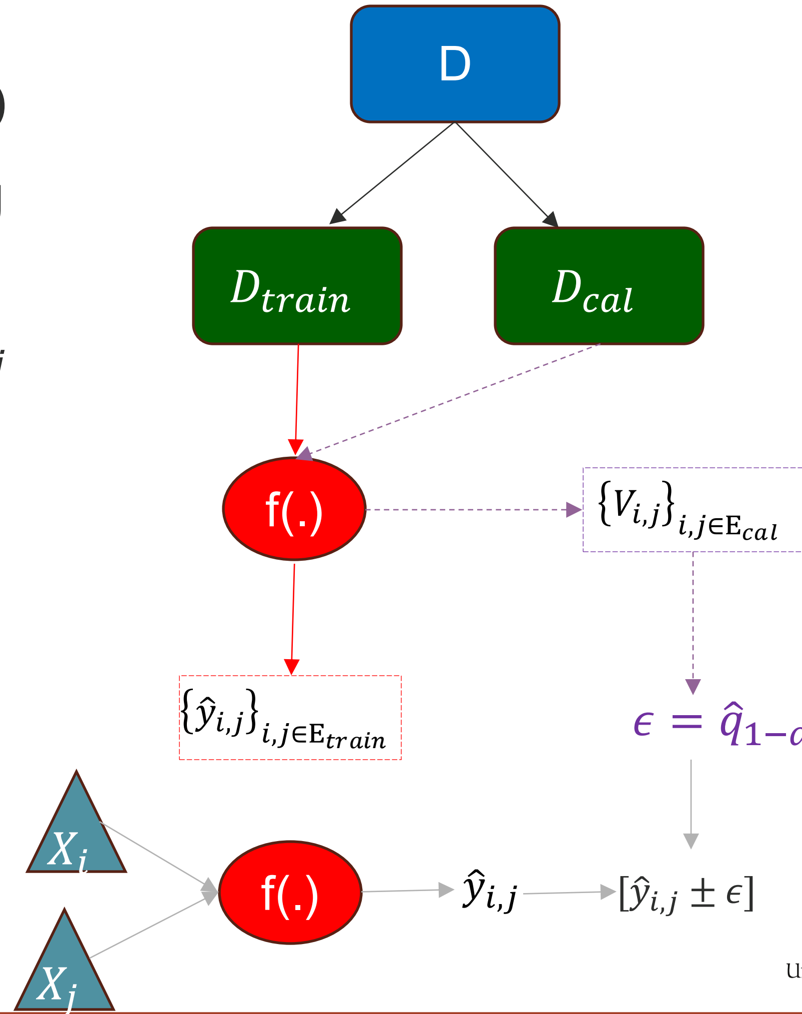
The University of Iowa

Overview

- Motivation
- Problem Formulation
- Our Approach
- Experiments
- Conclusion

Overview of Conformal Inference (Vanilla CP)

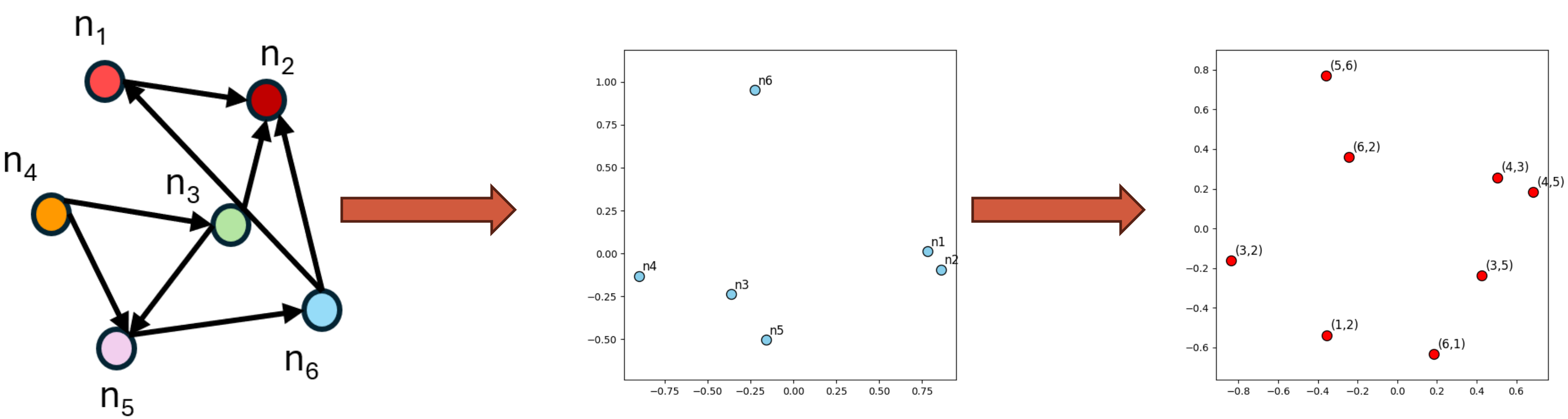
- Split the data D into training (D_{train}) and calibration fold (D_{cal})
- Train a base ML model $f(.)$ using (D_{train})
- For each edge (i,j) with weight $y_{i,j}$ in E_{cal}:
  - Compute the Non-Conformity (NC) Score V$(X_i, X_j, y_{i,j})$ by some norm-based score (like MSE)
- Calculate the $\hat{q}_{1-\alpha}$ of the NC score distribution
- For a test edge (i,j) with node features (X_i, X_j), return prediction band $\hat{C}(\hat{y}_{i,j}) \leq \hat{q}_{1-\alpha}$



The University of Iowa

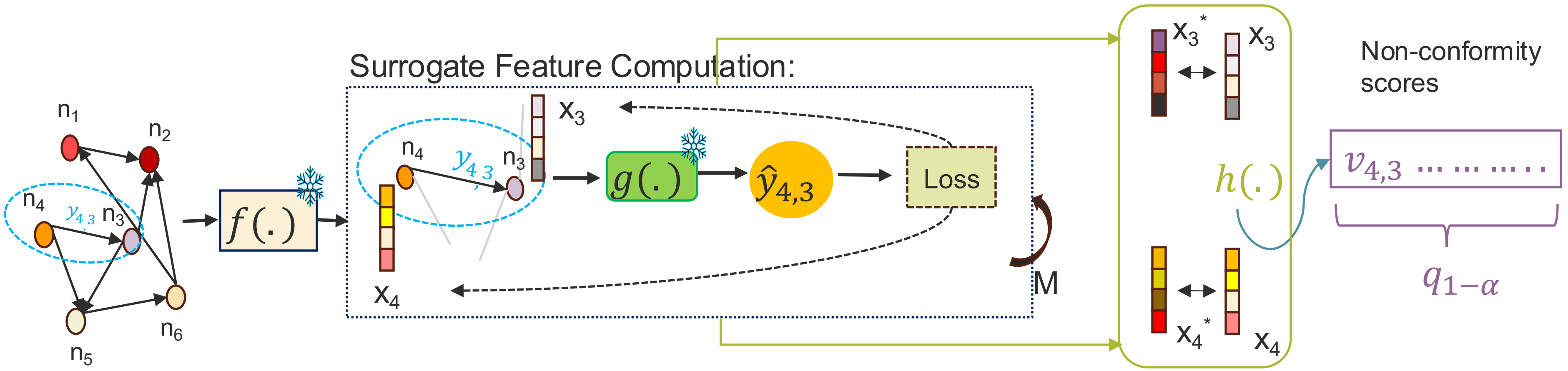
Our Idea- Edge-CP

- Instead of computing the NC score on the edge space, compute the score on the node latent space
- Find a way to translate the node NC score to the edge space
- We will need a proxy of true node label on the node feature space (surrogate node embedding)



Our Approach

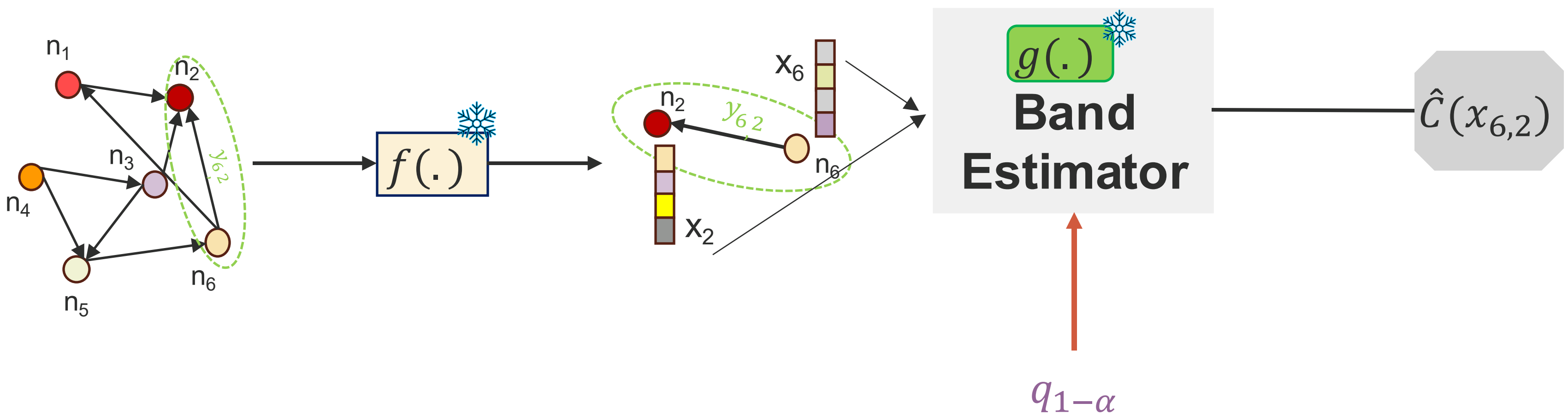
- Obtain surrogate node embeddings
- Compute the binary operator function h(.)



The University of Iowa

Prediction Set for Test Edge

- Define a Band Estimator to translate the node feature NC score to edge space
- We use Neural Network Robustness Certification Methods like:
  - IBP[1]
  - CROWN[2]



[1] Gowal, Sven, et al. "On the effectiveness of interval bound propagation for training verifiably robust models." arXiv preprint arXiv:1810.12715 (2018).
[2] Zhang, Huan, et al. "Efficient neural network robustness certification with general activation functions." *Advances in neural information processing systems* 31 (2018).

Effectiveness of Edge-CP

Lemma: If the node latent feature space satisfies the properties:

- The loss of information for Edge-CP in node space is bounded by the information obtained in the edge space
- The Band Estimation expands the differences between individual length and their quantiles
- Given a calibration set, the quantile of the band length is stable in both feature space and output space

Then Edge-CP provably outperforms Vanilla CP with shorter predictive band length

The University of Iowa

Overview

- Motivation
- Problem Formulation
- Our Approach
- Experiments
- Conclusion

The University of Iowa

Datasets

- **Human Protein-Protein Physical Interaction Network (HS-PI):** Biological protein-protein interaction dataset
- **Astrophysics Collaboration Network (astro-ph):** Collaboration network of scientists on the astrophysics archive during 1995-1999
- **Condensed-Matter Physics Collaboration Network (cond-mat):** Collaboration network of scientists on the condensed matter archive during 1995-1999
- **Benzodiazepine Receptor (BZR) Network (BZR-MD):** Interactions between 405 ligands of the Benzodiazepine Receptor (BZR)

Name	|V|	|E|	Label Range
HS-PI	17,849	633,460	(1.77-4.90)
cond-mat	16,264	47,594	(0.05-22.33)
astro-ph	16,046	121,251	(0.01-16.50)
BZR-MD	6,520	137,734	(1.14-16.64)

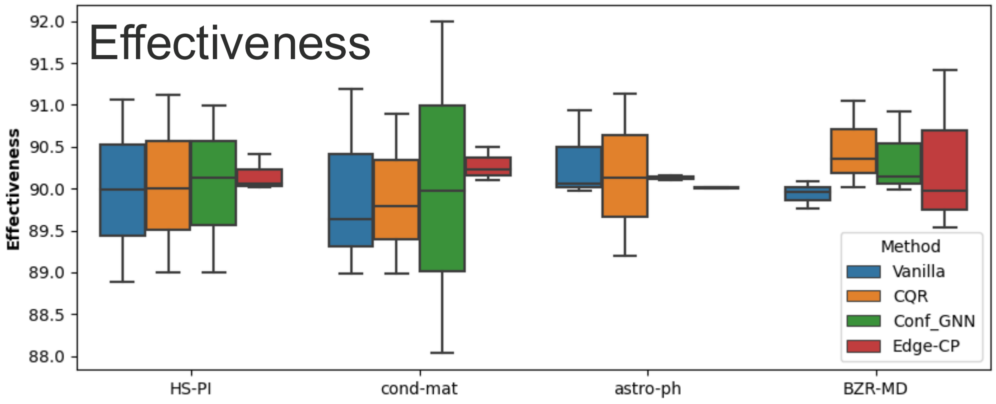
The University of Iowa

Performance Metrics

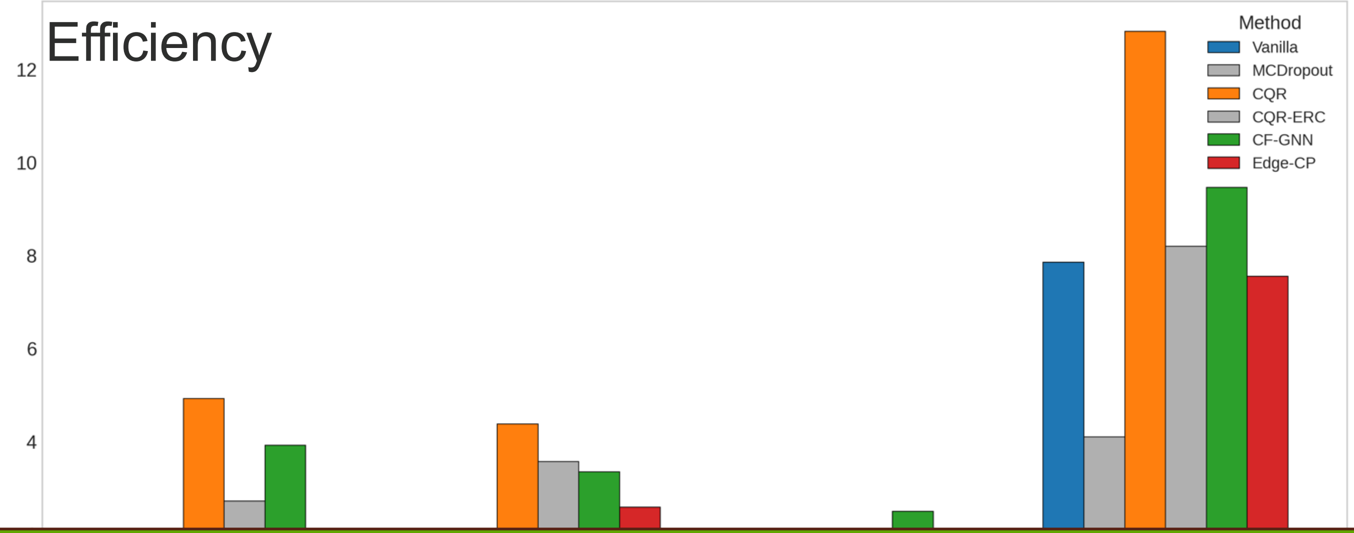
- **Effectiveness:** Empirical probability that a test point falls into the predicted confidence band. **We expect that the prediction sets will cover (1−α)% of the true test levels for a given significance level α**
- **Efficiency:** Average length of the prediction sets. **Lower efficiency is better**
- **Size Stratified Coverage (SSC):** Proportion of prediction sets containing the true label, stratified by prediction set size, to assess calibration across different confidence levels. **We want it to be close to (1−α)%**

The University of Iowa

Results



Close to 90

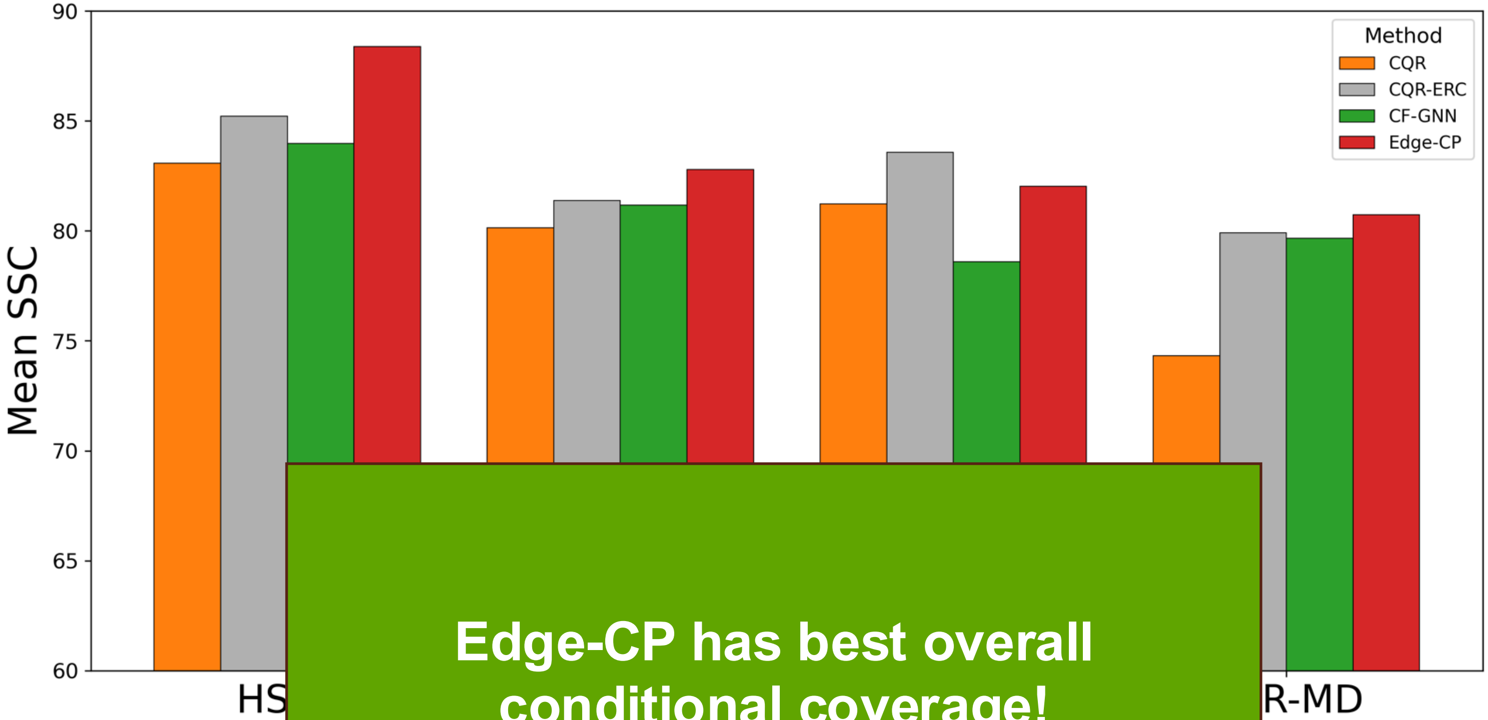


Lower is better

Edge-CP consistently reaches required effectiveness while having the best efficiency among all methods!

The University of Iowa

Results- Conditional Coverage



Higher is better

Edge-CP has best overall conditional coverage!

Overview

- Motivation
- Problem Formulation
- Our Approach
- Experiments
- Conclusion

Conclusion

- Edge-CP extends conformal inference in GNNs to weighted edge prediction tasks
- Edge-CP leverages the latent node embeddings to construct a NC score
- Edge-CP outperforms all baselines
- Future Directions:
  - Explore the validity of this method in higher-order interaction structures like hypergraphs
  - Extend the coverage guarantee to locally valid coverage

Thank You

Akash Choudhuri

Yongjian Zhong

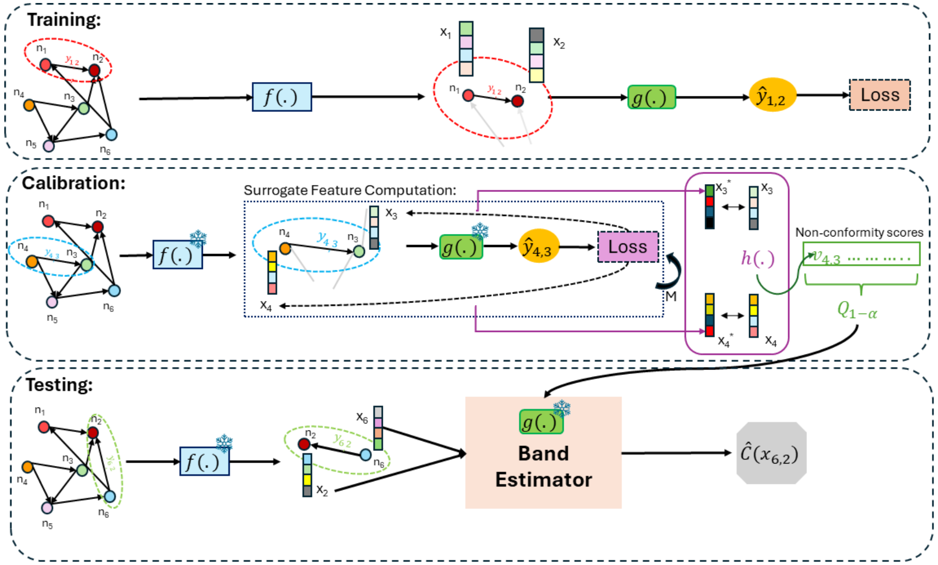
Mehrdad Moharrami

Christine Klymko

Mark Heimann

Jayaraman J Thiagarajan

Bijaya Adhikari



Code: https://github.com/Soothysay/Edge-CP.git

More details and results are in the paper!

Contact:
akash-choudhuri@uiowa.edu