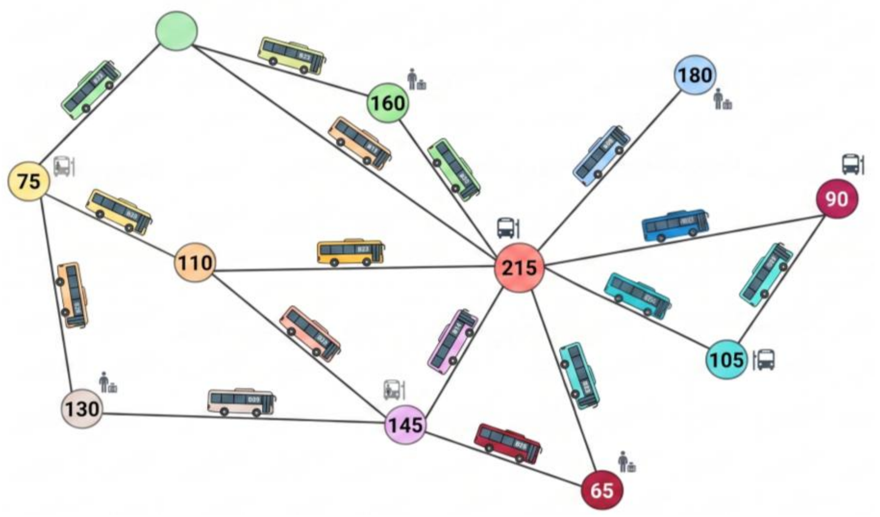
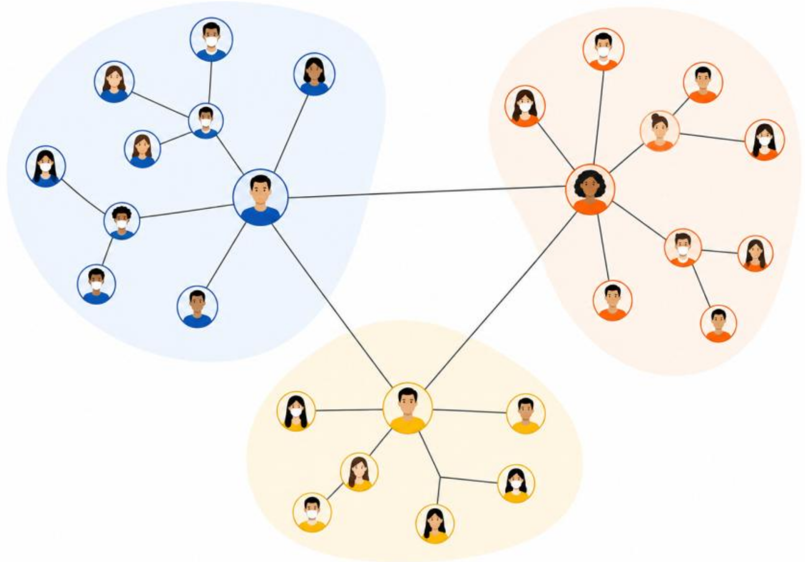
Conformal Inference for time series over graphs

Sonakshi Dua, Gonzalo Mateos, Sundeep Prabhakar Chepuri
Indian Institute of Science
University of Rochester

ICASSP2026 Barcelona

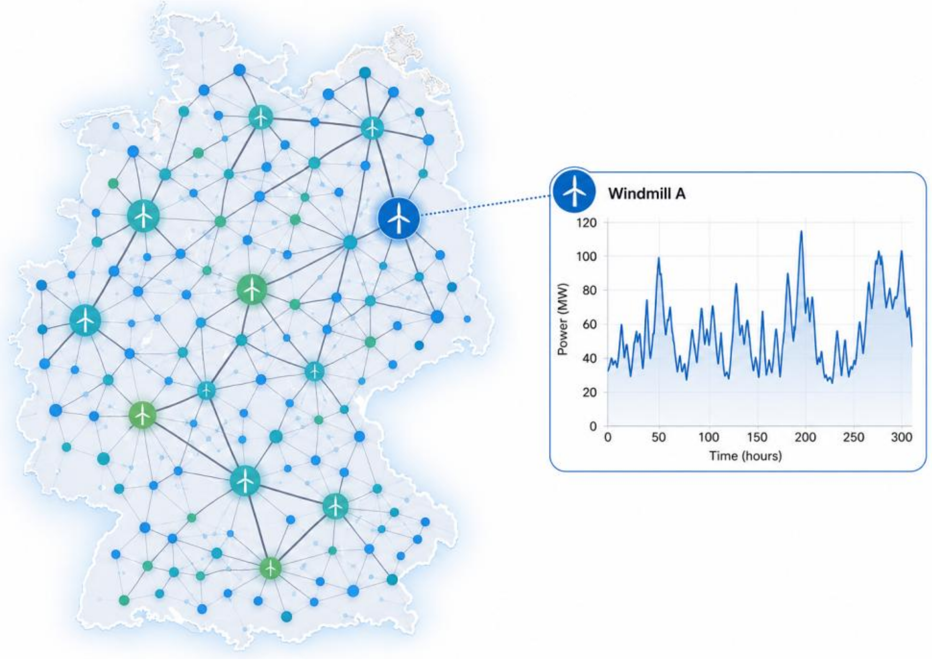


Passenger footfall at different bus stations



Evolving epidemic network

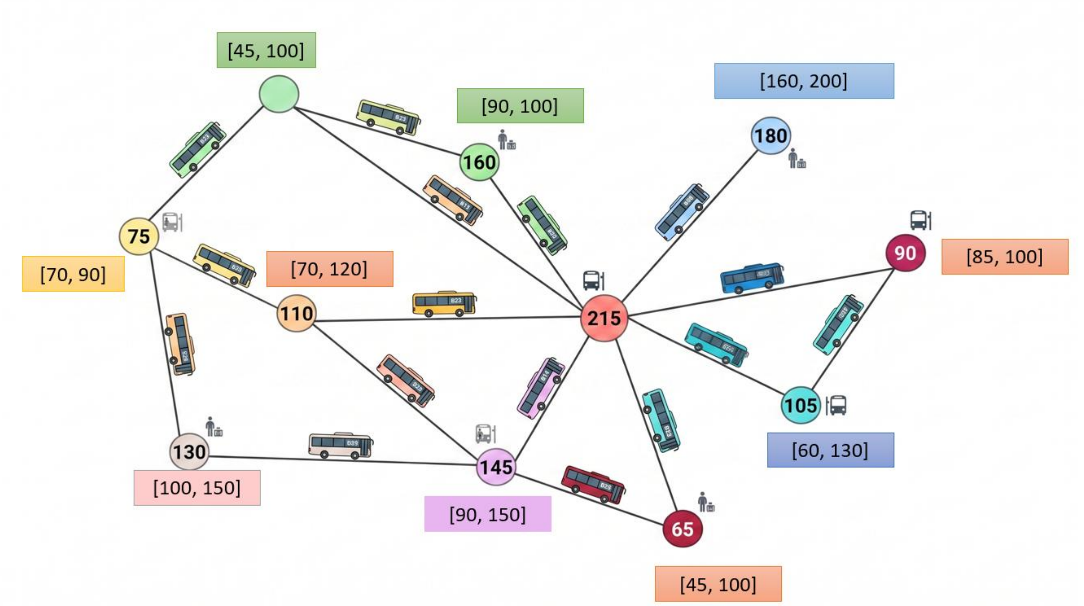
Temperature over different cities



Hourly energy output of windmills

Going beyond point predictions

Prediction set examples on Imagenet *[1]*



Prediction intervals for regression

Distribution-free wrapper around **pre-trained** neural net with **finite-sample coverage guarantee:**

$$P(y \in C(x) \geq 1 - \alpha)$$

Minimal assumption of exchangeability between calibration and test data required

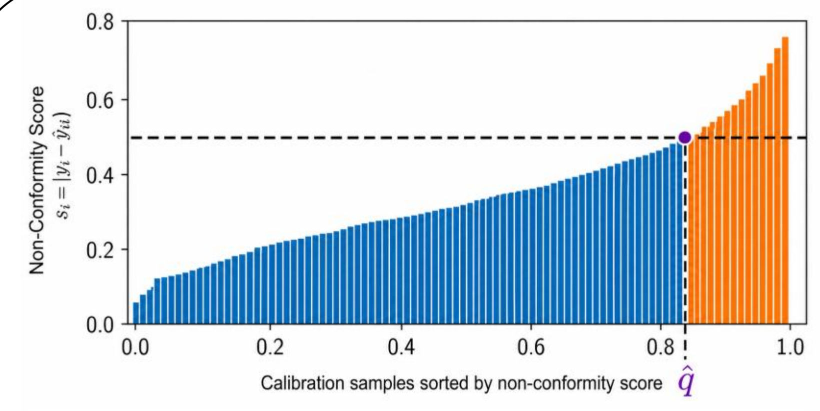
[1] Angelopoulos, A. N. & Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends in Machine Learning, 16(4), 494–591.

Conformal Prediction

Exchangeability : Random variables $W_1, \ldots, W_n$ for $n \geq 1$ are said to be exchangeable if

$$(W_1, \ldots, W_n) \stackrel{d}{=} (W_{\pi(1)}, \ldots, W_{\pi(n)}),$$

for any permutation π.



$\hat{q} = \text{Quantile}_{(1-\alpha)(1+\frac{1}{n})}\left(\{s(x_i)\}_{i=1}^{n}\right)$

Non-conformity scores

$s(x_i) = |\, y - f(x_i) \,|$

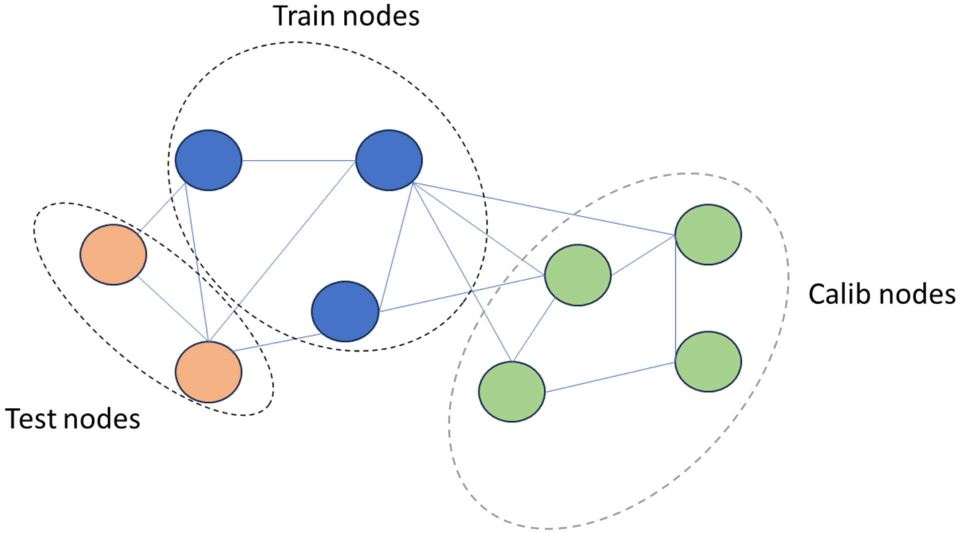
Calibration dataset (held-out data)

x_{test} → $f(.)$ Pre-trained model → $\hat{y}$ → + → $[\hat{y} - \hat{q}, \hat{y} + \hat{q}]$

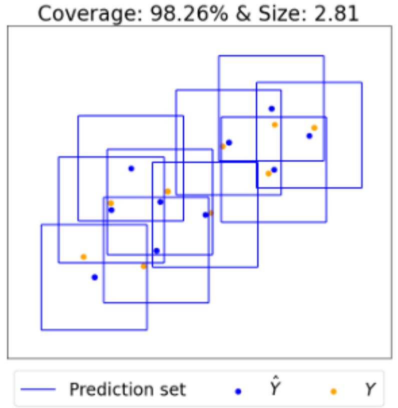
$\hat{q}$

State of the Art

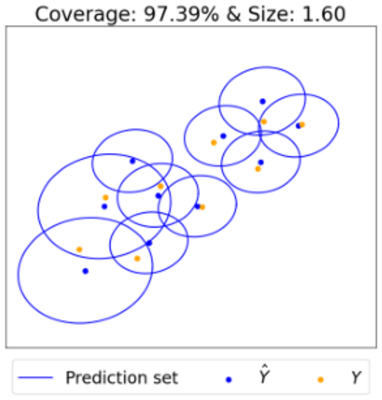
- CP over static graphs : DAPS *[2]*, CF-GNN *[3]*



- Multidimensional time series: Multidim SPCI *[4]*



Coordinate wise set



Ellipsoidal set [4]

[2] H. Zargarbashi, S., Antonelli, S. & Bojchevski, A.. (2023). Conformal Prediction Sets for Graph Neural Networks. Proceedings of the 40th International Conference on Machine Learning.
[3] Huang, K., Jin, Y., Candes, E., & Leskovec, J. (2023). Uncertainty quantification over graph with conformalized graph neural networks. *Advances in Neural Information Processing Systems*, *36*, 26699-26721.
[4] Xu, C., Jiang, H., & Xie, Y. (2024). Conformal prediction for multi-dimensional time series by ellipsoidal sets. *arXiv preprint arXiv:2403.03850*.

Problem Statement

Graph G is static and at each time step t,
graph data : $G_t = (x_t, G)$ with signal $x_t \in R^{N \times K}$ and target $y_t \in R^N$.

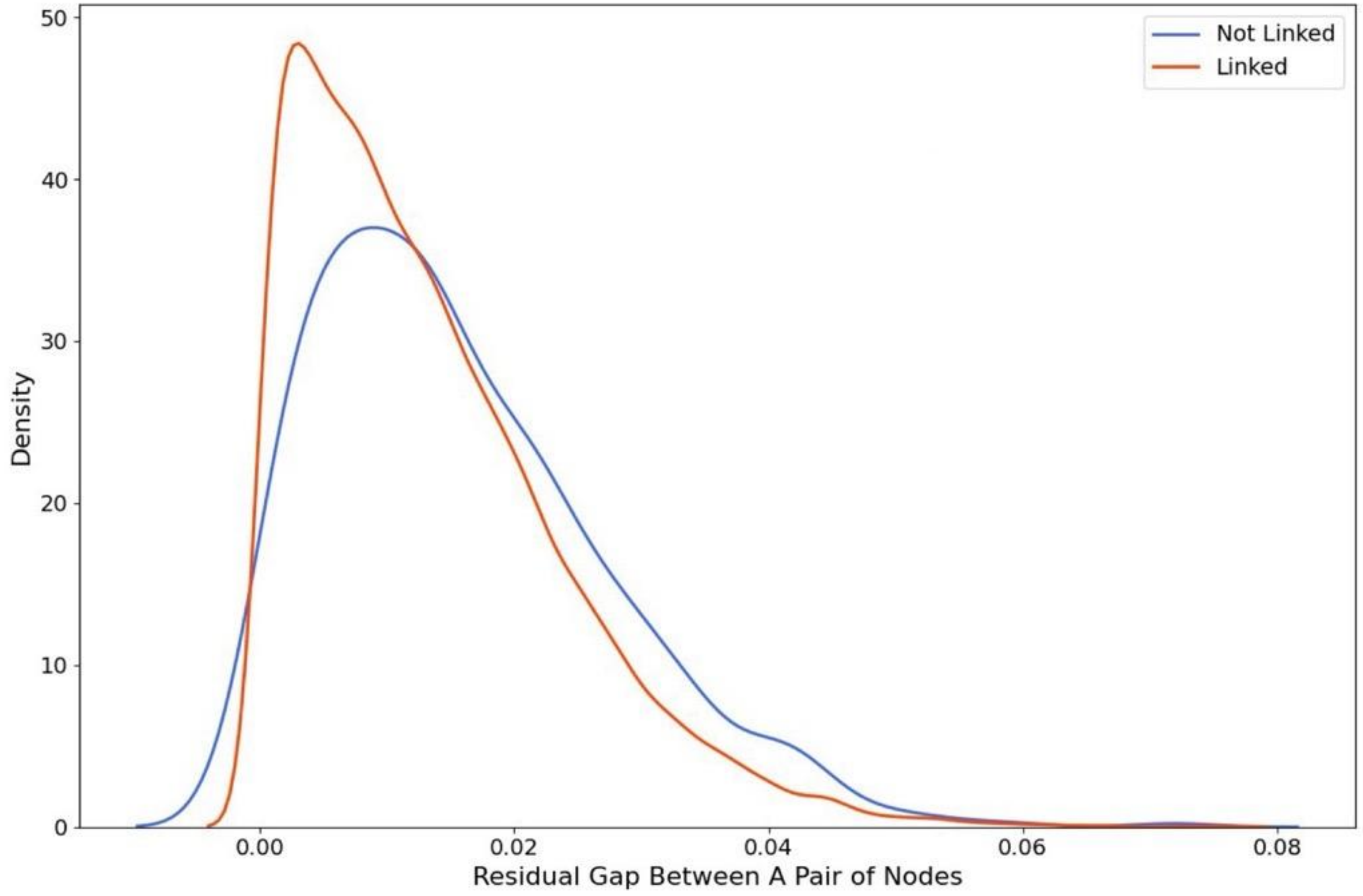
AIM : Estimate a meaningful (in terms of efficiency) prediction set $C_t(G_t)$ such that

$$\boldsymbol{P\left(y_t \in C(G_t)\right) \geq 1 - \alpha}, \forall \text{ t}$$

PROPSED SOLUTION:

- Taking motivation from DAPS, refine the scores by using a graph filter.
- Utilize the refined scores to obtain a scalar non-conformity score and build ellipsoidal sets as proposed by *[4]*

Graph aware non-conformity scores



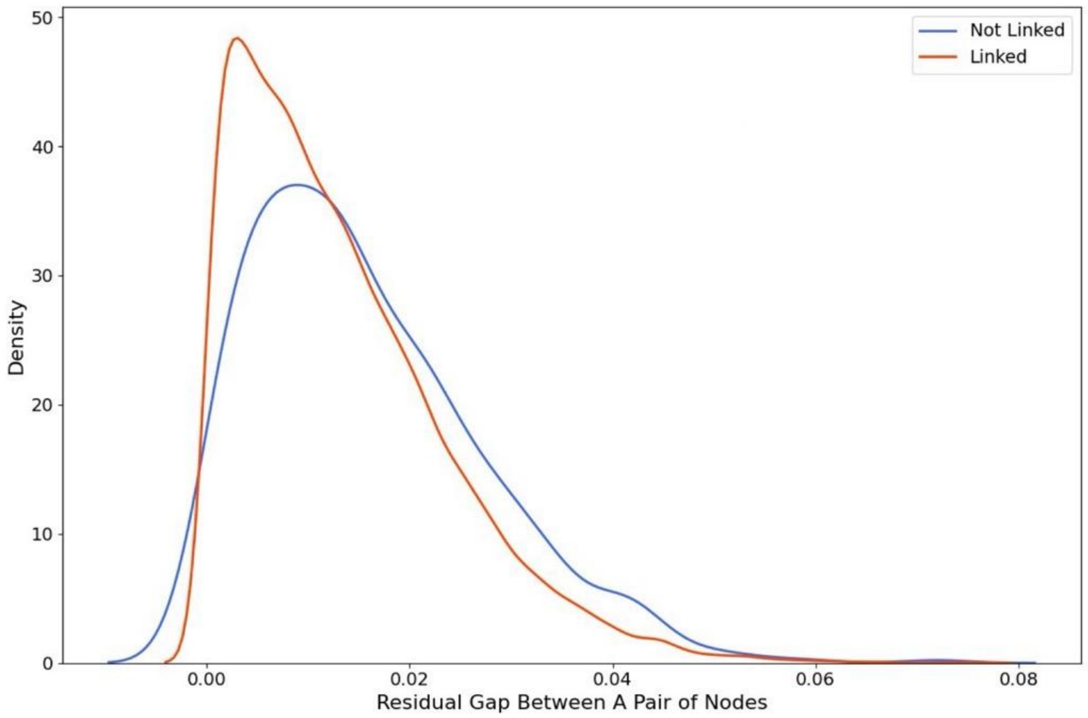
Residuals are smooth over the graph!

Graph aware non-conformity scores

$$\begin{bmatrix} \varepsilon_{t,1} \\ . \\ . \\ . \\ \varepsilon_{t,n} \end{bmatrix} \quad \boldsymbol{\varepsilon_t} = \boldsymbol{y}_t - f(\boldsymbol{G}_t)$$

$$\boldsymbol{e_t}(\boldsymbol{y}_t) = \boldsymbol{H}\boldsymbol{\varepsilon}_t = [(1-\tau)\boldsymbol{I} + \tau(\boldsymbol{D}^{-1}\boldsymbol{A})]\boldsymbol{\varepsilon}_t$$

hyper-parameter



Graph aware non-conformity scores

$$\boldsymbol{e}_t(\boldsymbol{y}_t) = \boldsymbol{H}\boldsymbol{\varepsilon}_t = [(1-\tau)\boldsymbol{I} + \tau(\boldsymbol{D}^{-1}\boldsymbol{A})]\boldsymbol{\varepsilon}_t$$

$$\begin{bmatrix} \varepsilon_{t,1} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{t,n} \end{bmatrix} \rightarrow \begin{bmatrix} \mathrm{e}_{t,1} \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{e}_{t,n} \end{bmatrix} \rightarrow s_t(\boldsymbol{y}_t)$$

Squared Mahalanobis distance as [4]

$$s_t(\boldsymbol{y}_t) = (\boldsymbol{e}_t - \bar{\boldsymbol{e}})^{\boldsymbol{T}} \boldsymbol{\Sigma}_{\boldsymbol{G}}^{-1}(\boldsymbol{e}_t - \bar{\boldsymbol{e}})$$

$$\Sigma_G = E[(e_t - \bar{e})(e_t - \bar{e})^T]$$

Ellipsoidal Prediction Sets

$$C_t(\boldsymbol{G}_t) = \{\boldsymbol{y}_t | s_t(y_t) \leq Q_t(1-\alpha)\}$$
$$= f(\boldsymbol{G}_t) + B\left(\sqrt{Q_t(1-\alpha)}, \bar{\boldsymbol{e}}, \boldsymbol{\Sigma}_{\boldsymbol{G}}\right)$$

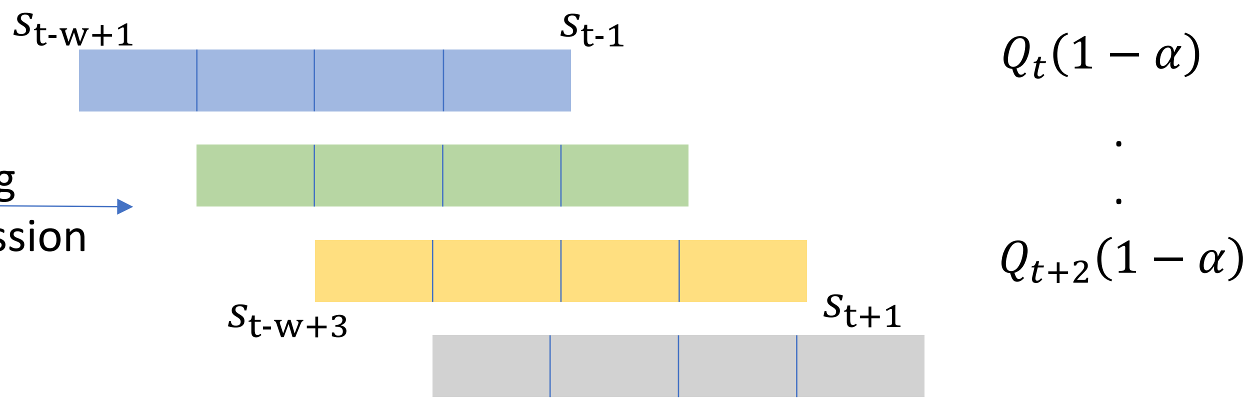
$$B(r, c, \Sigma) = x: (x-c)^T \Sigma^{-1} (x-c) \leq r$$

[4] Xu, C., Jiang, H., & Xie, Y. (2024). Conformal prediction for multi-dimensional time series by ellipsoidal sets. *arXiv preprint arXiv:2403.03850*.

Ellipsoidal Prediction Sets

$$C_t(\boldsymbol{G}_t) = \{\boldsymbol{y}_t | s_t(y_t) \leq Q_t(1-\alpha)\}$$
$$= f(\boldsymbol{G}_t) + B\left(\sqrt{Q_t(1-\alpha)}, \bar{\boldsymbol{e}}, \boldsymbol{\Sigma}_{\boldsymbol{G}}\right)$$

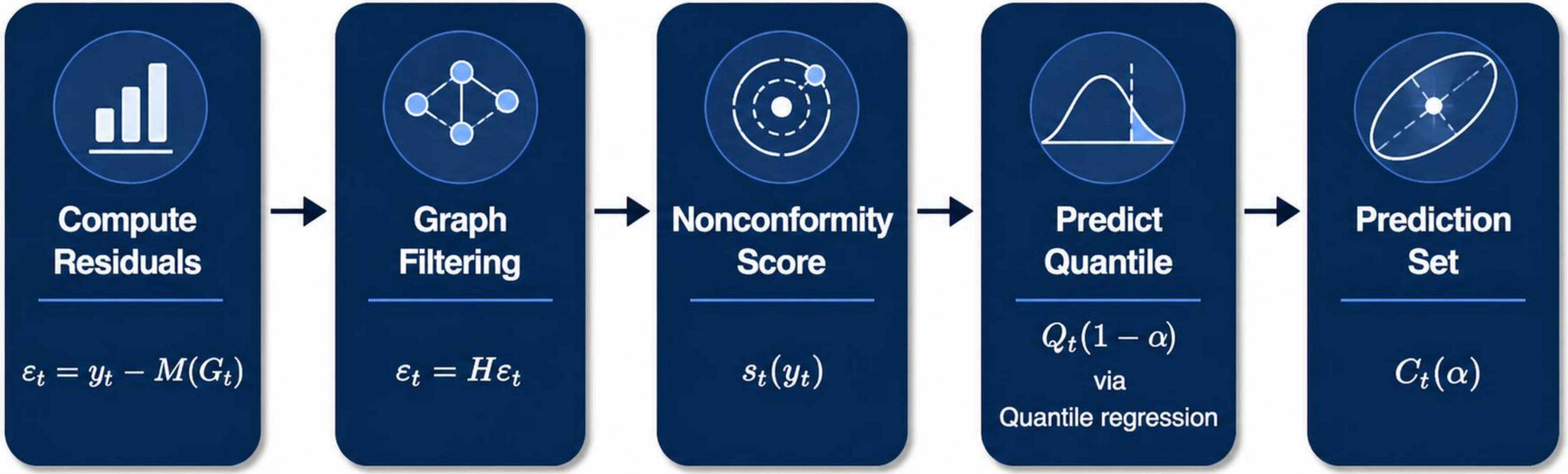
estimated using Quantile regression



$$B(r, c, \Sigma) = x: (x-c)^T \Sigma^{-1} (x-c) \leq r$$

[4] Xu, C., Jiang, H., & Xie, Y. (2024). Conformal prediction for multi-dimensional time series by ellipsoidal sets. *arXiv preprint arXiv:2403.03850*.

Pipeline



Compute Residuals
$\varepsilon_t = y_t - M(G_t)$
Graph Filtering
$\varepsilon_t = H\varepsilon_t$
Nonconformity Score
$s_t(y_t)$
Predict Quantile
$Q_t(1-\alpha)$
via
Quantile regression
Prediction Set
$C_t(\alpha)$

Theoretical Results

Ellipsoidal Volume Shrinkage

If $\frac{Q_t(1-\alpha)}{Q_t'(1-\alpha)} \approx 1$, then $Vol(B_G) \leq e^{-\eta\tau} Vol(B')$, for some η $>$ 0, where τ is the graph filter coefficient

Proof Sketch: $$log\left[\frac{Vol(B_G)}{Vol(B')}\right] = log\ det(\boldsymbol{H}) + N/2\left(log\left(\frac{Q_t(1-\alpha)}{Q_t'(1-\alpha)}\right)\right)$$

log det(**H**) = log det $((1 - \tau)I + \tau D^{-1}A)$

$= \sum_{i=1}^{N} \log(1 - \tau(1 - \lambda_i))$

$\leq -\tau \sum_{i=1}^{N} (1 - \lambda_i)$

log det(**H**) $\leq$ -τη or det(**H**) $\leq$ $\boldsymbol{e^{-\eta\tau}}$

Theoretical Results

Coverage Guarantee

Assume that the true covariance matrix $\sum_G$, is known and positive definite with minimum eigenvalue at least $\lambda > 0$, the filtered residuals are i.i.d over time and CDF of the true non-conformity score is Lipschitz, then we have

$$|P(y_{t+1} \in C_{T+1}(\alpha)|G_{T+1}) - (1-\alpha)| \leq 12\sqrt{\frac{log(16T)}{T}} + L\left(\frac{\delta_T}{\sqrt{\lambda}} + \delta_T\right)$$

where, T is the training data size, δ_T is bound on residual error, and L depends on Lipschitz constant.

as established in Corollary 4.14 of *[4]*

[4] Xu, C., Jiang, H., & Xie, Y. (2024). Conformal prediction for multi-dimensional time series by ellipsoidal sets. *arXiv preprint arXiv:2403.03850*.

Experimental Results

$\alpha = 0.1$

Method	Metric	Wiki Maths	MontevideoBus	Chickenpox Hungary
Graph-agnostic	Coverage	0.903 ± 0.006	0.91 ± 0.0122	0.89 ± 0.0167
	Volume	$5.19 \times 10^3 \pm 2010.298$	$3.09 \times 10^3 \pm 247.325$	$2.74 \times 10^2 \pm 14.842$
Graph-aware (proposed)	Coverage	0.897 ± 0.010	0.912 ± 0.008	0.89 ± 0.018
	Volume	$\mathbf{1.46 \times 10^3 \pm 135.769}$	$\mathbf{1.56 \times 10^3 \pm 880.327}$	$\mathbf{1.25 \times 10^2 \pm 70.851}$

$\alpha = 0.05$

Method	Metric	Wiki Maths	MontevideoBus	Chickenpox Hungary
Graph-agnostic	Coverage	0.954 ± 0.005	0.948 ± 0.004	0.916 ± 0.011
	Volume	$8.51 \times 10^3 \pm 1458.303$	$1.406 \times 10^4 \pm 205.985$	$1.6 \times 10^2 \pm 14.153$
Graph-aware (proposed)	Coverage	0.952 ± 0.004	0.952 ± 0.008	0.924 ± 0.005
	Volume	$\mathbf{2.04 \times 10^3 \pm 238.579}$	$\mathbf{2.7 \times 10^3 \pm 112.472}$	$\mathbf{1.29 \times 10^2 \pm 20.144}$

Experimental Results

Wiki Maths Dataset	$r = 1$		$r = 5$		$r = 10$	
Method	Coverage	Volume	Coverage	Volume	Coverage	Volume
Graph-agnostic	0.903 ± 0.006	$5.19 \times 10^3 \pm 2010.29$	0.889 ± 0.012	$1.20 \times 10^4 \pm 754.91$	0.875 ± 0.008	$9.40 \times 10^3 \pm 5218.74$
Graph-aware (proposed)	0.897 ± 0.010	$\mathbf{1.46 \times 10^3 \pm 135.77}$	0.885 ± 0.0129	$\mathbf{2.40 \times 10^3 \pm 694.09}$	0.867 ± 0.002	$\mathbf{3.50 \times 10^3 \pm 595.01}$

Results over one-step prediction and multi-step prediction (r being the number of steps)

Experimental Results

Wiki Maths Dataset	$w = 10$		$w = 50$		$w = 100$	
Method	Coverage	Volume	Coverage	Volume	Coverage	Volume
Graph-agnostic	$0.903 \pm .006$	$5.19 \times 10^3 \pm 2010.298$	0.892 ± 0.005	$4.80 \times 10^3 \pm 1135.54$	0.885 ± 0.014	$2.85 \times 10^3 \pm 1260.33$
Graph-aware (proposed)	0.897 ± 0.01	$\mathbf{1.46 \times 10^3 \pm 135.769}$	0.896 ± 0.009	$\mathbf{1.27 \times 10^3 \pm 182.67}$	0.886 ± 0.002	$\mathbf{1.22 \times 10^3 \pm 345.89}$

Results over different window lengths to predict the quantiles

Conclusions

- Applied conformal prediction to graph time series
- Refined nonconformity scores using graph topology
- Derived theoretical guarantees on volume reduction
- Achieved smaller prediction sets while maintaining coverage
- Future work: recursive threshold updates, dynamic graphs, heterophilic settings

Thank you!

We thank SPS for the travel grant!