STATISTICAL ANALYSIS OF GRAPH AND PREDICTION OF TRAINED ON STATIC TEMPORAL ATTENTION MODEL

Supervisor:
Harish Chandra Bhandari,
Lecturer,
Department of Mathematics
School of Science, Kathmandu University

Presented by:
Prastut Bhattarai (07)
Shaswat Bhushan Jangam (12)
Computational Mathematics (6th Semester),
Department of Mathematics
School of Science, Kathmandu University

BACKGROUND

- ➢ Initial work: Graph Neural Networks using PyTorch Geometric Temporal
  - A Graph Dataset was trained on Temporal Graph Attention Layers (Spatio-Temporal Graph Convolutional Networks)
- ➢ Challenge:
  - Hardware and time complexity issues with locally available CPU and GPU, resulting in the failure to train the model
- ➢ So,
  - The problem was approached using a pre-trained model
  - Then, the prediction results of the pre-trained model were analyzed

PROBLEM STATEMENT

Finding causes of low accuracy prediction of Graph Neural Network (GNN) model

- Mathematical Representation: $G(V, E, Ev, Xv(t))$
  - **V**: Set of nodes representing weather stations
  - **E**: Set of edges representing relationships between weather stations
  - **Xv(t)**: Set of node features representing the weather stations
  - **t**: Set of timestamps representing the temporal evolution
  - **Ev :** Set of edge features based on altitude of node
- Learned function f using GNN s.t $f\!:\!Xv(t1,t2,..,tn) \rightarrow R$
- Prediction of $Xv(t)$ i.e. future states of the graph ,using historical data and trained function f (parameters)
- Analyse the prediction of $Xv(t)$ with graph statistical parameters

OBJECTIVE

To analyze the prediction based on the statistical parameters of a graph, trained on a static temporal attention (STConv) model

Learn the impact of statistical parameters on model prediction.

Parameters to analyze:

1. Eccentricity
2. Betweenness Centrality
3. Closeness Centrality
4. Degree Centrality
5. Node Degree

Correlate these parameters with RMSE.

METHODOLOGY

Available Resources: ©Harish Chandra Bhandari

1. **Data Loading:**
   - Use of PyTorch (Static Graph Temporal Signal) data loader
   - Data details: ***x=[1,60,750,18], edge_index=[2,141004], edge_attr=[141004], y=[1,24,750,7]***
2. **Trained model (static temporal Convolutional):**
   - A trained model is available which predicts the future value for days in looking at past days historic data

METHODOLOGY CONTINUED…

Data Conversion:

- Convert PyTorch data loader to NetworkX graph G.
- Compute parameters:
  - Eccentricity
  - Closeness Centrality
  - Betweenness Centrality
  - Degree Centrality
  - Node Degrees

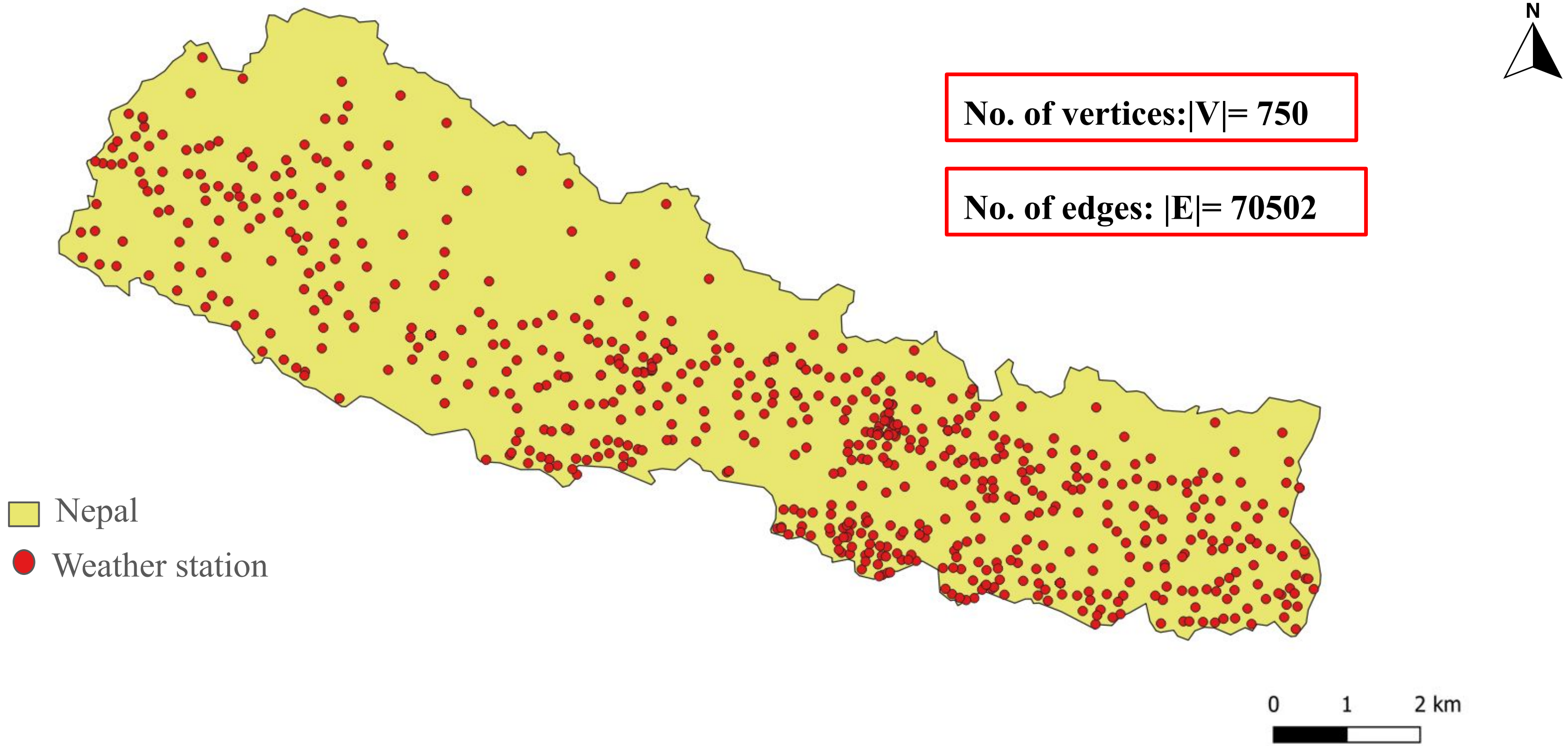
Data Analysis:

- Convert results to a pandas DataFrame for analysis.
- Plotted correlation using matplotlib.

STATISTICAL PARAMETERS

1. **Eccentricity:**
    a. Measures the distance to the farthest node.
    b. Nodes with higher eccentricity are further from center of network.
2. **Closeness Centrality:**
    a. Indicates how close a node is to all other nodes.
    b. Higher value indicates node that are on average closer to all other node in network.
3. **Betweenness Centrality:**
    a. Measures the number of times a node acts as a bridge.
    b. Higher value indicates node has more control over the flow of information in network.
4. **Degree Centrality:**
    a. Indicates the number of edges connected to a node.
5. **Node Degree:**
    a. The number of connections a node has.

Weather Stations Of Nepal Treated As Nodes



No. of vertices:|V|= 750

No. of edges: |E|= 70502

Figure 1: Weather stations of Nepal treated as nodes

ACHIEVEMENTS OVERVIEW

1. Extracted statistical parameter of a graph for each node

2. Analyzed the impact of statistical parameters on model predictions

3. Identified key parameters influencing prediction accuracy

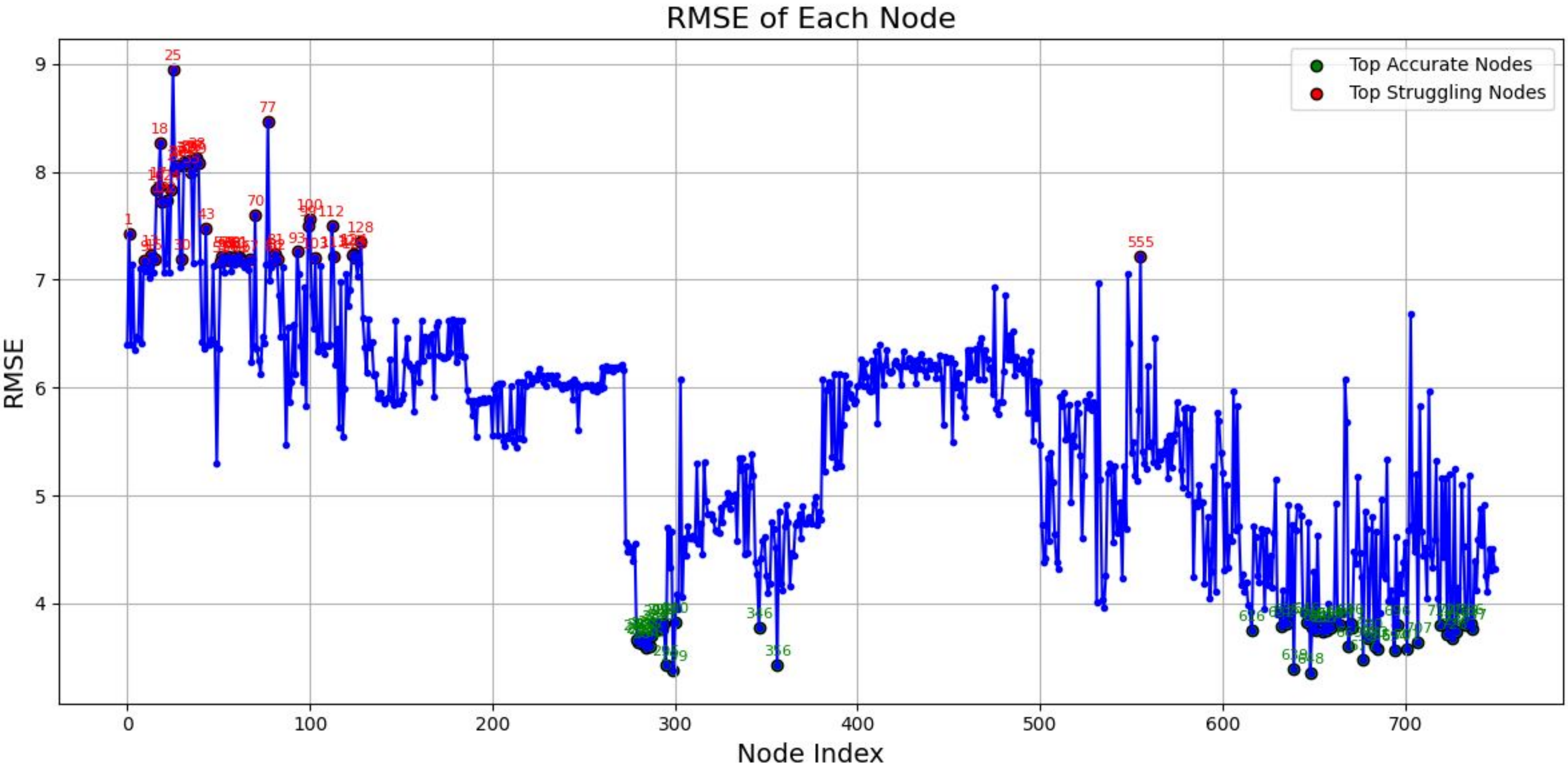


Figure 2: Root Mean Square Error (RMSE) of each node

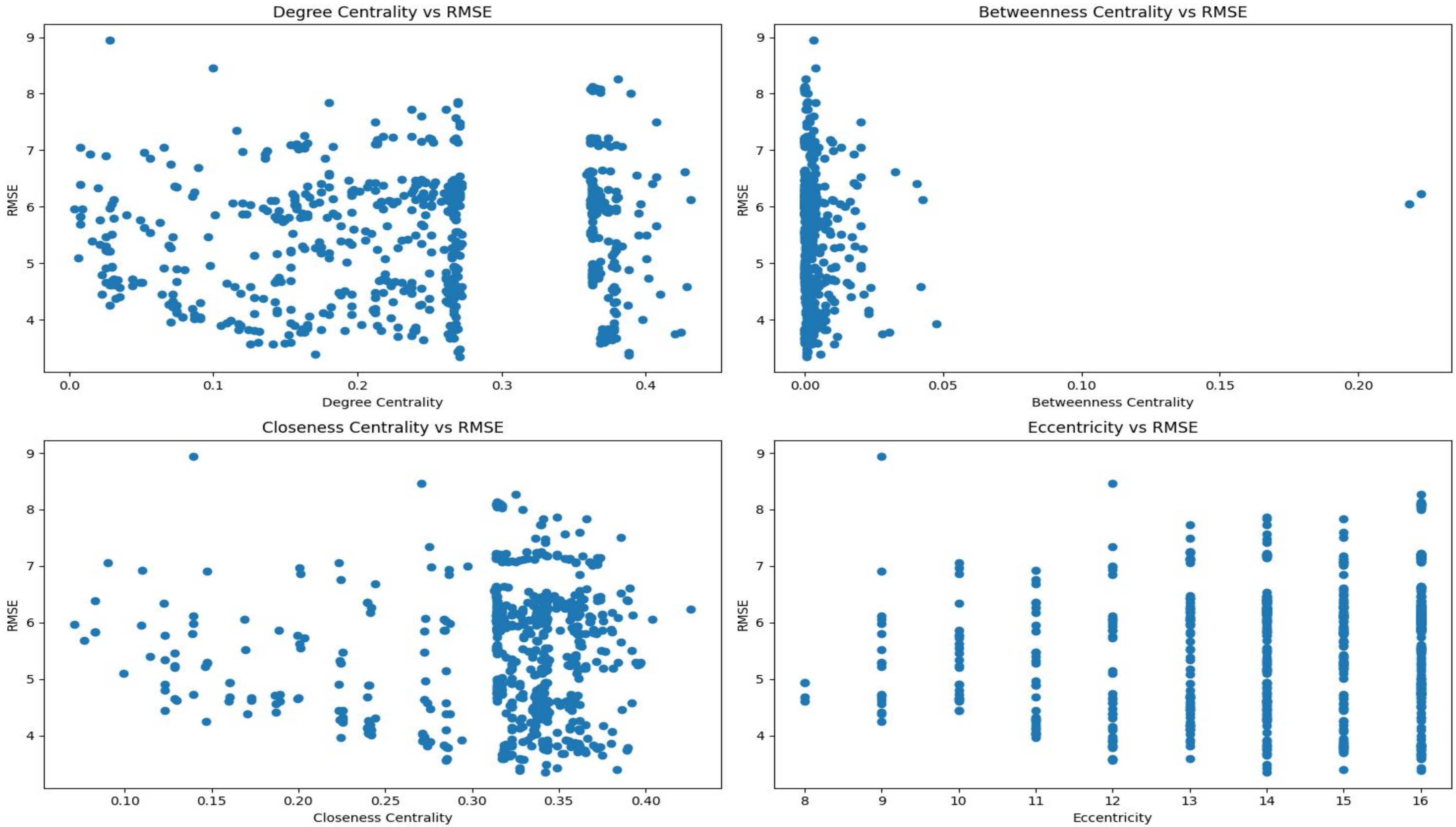


Figure 3: Correlation between Root Mean Square Error (RMSE) and statistical parameters

Top Accurate Node

Node	Eccentricity	Closeness_Centrality	Betweenness_Centrality	Degree_Centrality	Node_Degrees	altitude	RMSE
685	12	0.329665493	0.004658211	0.14953271	112	145	3.581660032
694	12	0.284899201	0.002425862	0.125500668	94	515	3.568217039
701	12	0.328364752	0.010542514	0.141522029	106	180	3.576862812
356	14	0.348858873	0.001229104	0.269692924	202	102	3.429826736
648	14	0.342009132	0.000688139	0.271028037	203	2353	3.353683472
677	14	0.34263495	0.000703272	0.271028037	203	1654	3.481489182
639	15	0.383512545	0.005780436	0.170894526	128	1435	3.396530151
284	16	0.317507418	0.0000359	0.368491322	276	1427	3.588338614
295	16	0.32736014	0.000831339	0.388518024	291	1736	3.425880194
299	16	0.32736014	0.000831339	0.388518024	291	1360	3.382719278

Table 1: Top accurate node based on Root Mean Square Error (RMSE) and its statistical parameters

Top Struggling Node

Node	Eccentricity	Closeness_Centrality	Betweenness_Centrality	Degree_Centrality	Node_Degrees	altitude	RMSE
25	9	0.139426657	0.003165253	0.028037383	21	140	8.947193146
77	12	0.270788142	0.003809603	0.100133511	75	1368	8.458855629
18	16	0.325086806	0.000316967	0.380507343	285	1519	8.26391983
38	16	0.314441646	0.000000452	0.363150868	272	71	8.129192352
37	16	0.314441646	0.000000452	0.363150868	272	1427	8.11057663
34	16	0.314441646	0.000000452	0.363150868	272	272	8.104405403
31	16	0.316300676	0.0000138	0.365821095	274	136	8.097679138
33	16	0.314441646	0.000000452	0.363150868	272	101	8.09544754
39	16	0.313782991	0.000000071	0.361815754	271	1427	8.081612587
32	16	0.317507418	0.0000359	0.368491322	276	109	8.079301834

Table 2: Top struggling node based on Root Mean Square Error (RMSE) and its statistical parameters

Top 50 Nodes With Lowest And Highest Root Mean Square Error (RMSE) Viewing Geographically

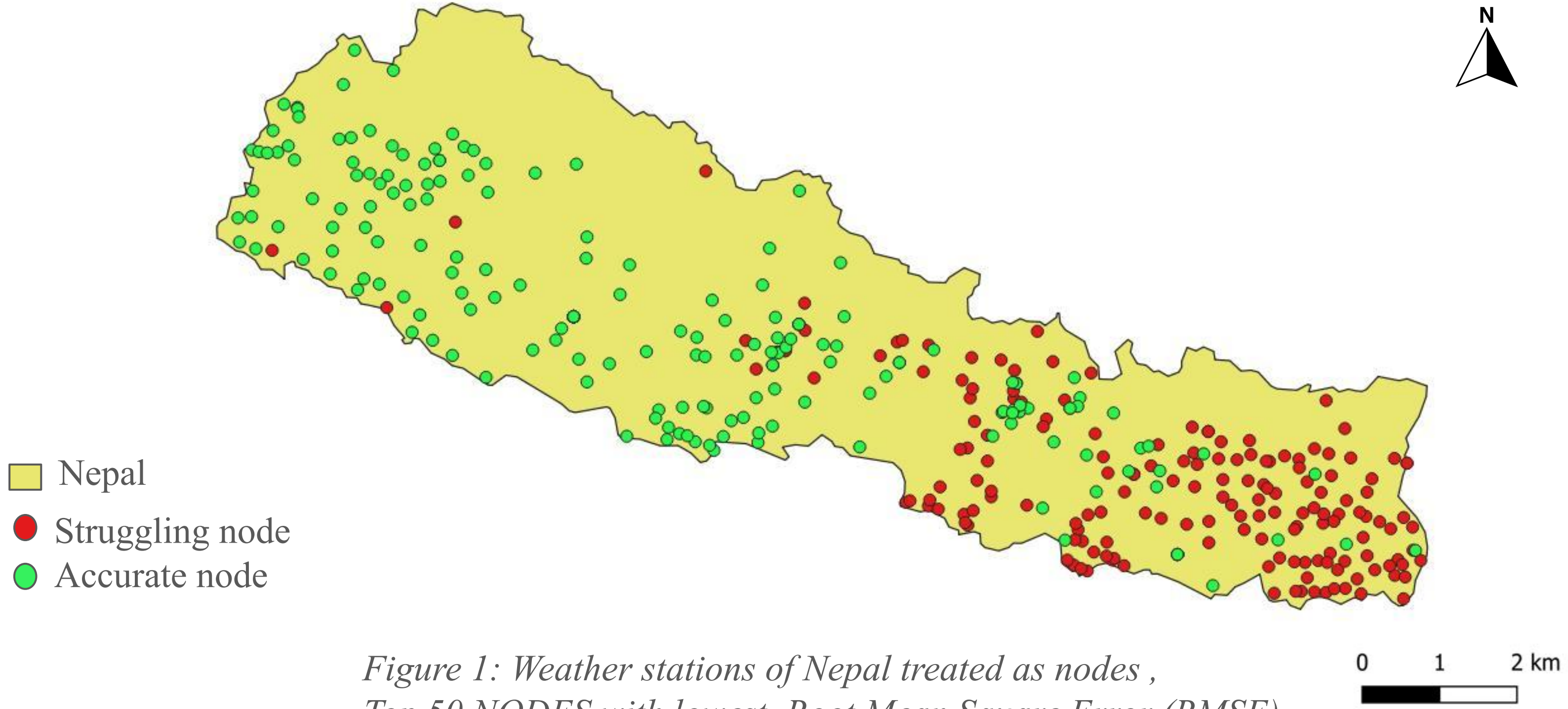


Figure 1: Weather stations of Nepal treated as nodes ,
Top 50 NODES with lowest Root Mean Square Error (RMSE)
Top 50 NODES with highest Root Mean Square Error (RMSE)

RESULT SUMMARY

1. Degree Centrality vs RMSE

- **Range of Degree Centrality**: Most nodes have a degree centrality between 0.0 and 0.4.
- **RMSE Distribution**: There is no clear trend that suggests a strong relationship between degree centrality and RMSE. Nodes with lower degree centrality (around 0.0 to 0.1) have RMSE values spread across the range, similar to those with higher degree centrality (around 0.3 to 0.4).

2.Betweenness Centrality vs RMSE

- **Range of Betweenness Centrality**: Most nodes have a very low betweenness centrality (close to 0.0), with a few outliers going up to 0.2.
- **RMSE Distribution**: Comparatively higher value of betweenness centrality for nodes with low RMSE.

3. Closeness Centrality vs RMSE

- **Range of Closeness Centrality**: Most nodes have closeness centrality between 0.1 and 0.4.
- **RMSE Distribution**: Similar to degree centrality, the RMSE values are spread across the range for different closeness centrality values. There is no clear pattern indicating a strong dependency of RMSE on closeness centrality.

4. Eccentricity vs RMSE

- **Range of Eccentricity**: Eccentricity values range from 8 to 16.
- **RMSE Distribution**: Nodes with lower eccentricity (8 to 11) tend to have more tightly clustered RMSE values, generally between 4.0 and 6.0. As eccentricity increases (12 to 16), the RMSE values become more spread out, ranging from 3.0 to 9.0. This suggests that nodes with higher eccentricity have more variable prediction accuracy, aligning with your initial observation.

CONCLUSION

Eccentricity: There is a noticeable pattern where nodes with higher eccentricity have a wider range of RMSE values, indicating more variability in prediction accuracy.

Degree, Betweenness, and Closeness Centrality: None of these centrality measures show a strong, clear relationship with RMSE. The RMSE values are spread across the range for different centrality values, suggesting that these measures are not primary determinants of prediction accuracy.

The observed results suggest that while some centrality measures and eccentricity can influence prediction accuracy, they do not provide a complete picture. The variability in RMSE values indicates that other factors, such as temporal dynamics, specific node attributes, and the nature of information flow within the network, play significant roles. Further investigation into these factors, possibly through analysing additional feature, could provide deeper insights into improving prediction accuracy.

REFERENCES

- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on DD-MMM-YYYY)
- >@inproceedings{rozembersky 2021 pytorch,author = {Benedek Rozemberczki and Paul Scherer and Yixuan He and George Panagopoulos and Alexander Riedel and Maria Astefanoaei and Oliver Kiss and Ferenc Beres and Guzman Lopez and Nicolas Collignon and Rik Sarkar},title = {{PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models}},year = {2021},book title={Proceedings of the 30th ACM International Conference on Information and Knowledge Management},pages = {4564–4573}
- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 922-929. https://doi.org/10.1609/aaai.v33i01.3301922