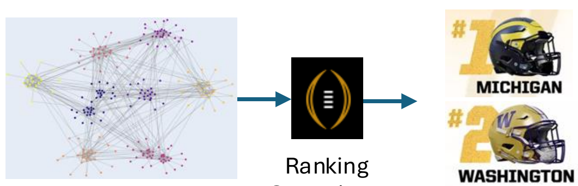
Explainable DiGCNs for Decomposition of Opaque Node Ranking Functions

Vishal Chandra
University of Michigan, MIT Lincoln Lab

Summary

The proposed method fits an interpretable set of topological components to explain black-box node ranking functions in digraphs.

Overview of Black-Box Ranking



College Football graph — Ranking Committee — Ranks

Citation graph — Semantic Scholar — Author Impact

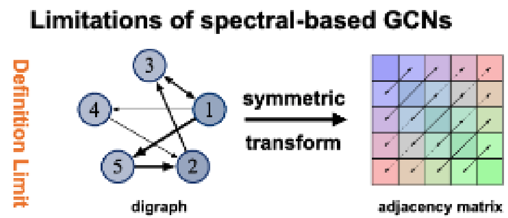
Two classes of opaque ranking. **Top**: no explicit formula exists, human intuition. **Bottom:** an explicit formula exists but is not public (was later found to be based on h-index).
More generally, there always exists a function

$$\mathcal{F} : G(N, E) \to \mathbb{R}^{|N|}$$

that we try to approximate interpretably

Model Selection

The function approximation task largely considers player-game networks, which are almost always digraphs.



Spectral methods assume adjacency symmetry or require complex edge representations.
DiGCN uses personalized PageRank to define digraph spectral convolution.

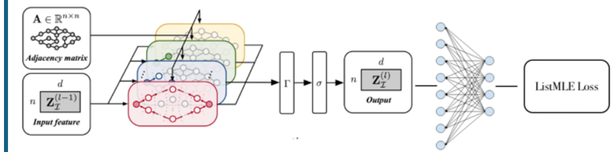
Learning to Rank

"Learning-to-Rank" (LTR) is a well-studied ML problem of fitting to rank-order statistics. That is, we do not have full information about the desired outputs, only their rank. This has been explored in graphs for ranking nodes (Ergashev et al.).

We employ ListMLE loss here to enable learning based only on black-box ranks.

See Xia et al. 2008 and Plackett-Luce model for details.

Explainable DiGCN for Node Ranking



Backbone: Standard DiGCN
Linear Layer: unbounded mapping from embedding to score space. Simpler activations may also miss inverse relationships between centrality and rank (ex. ranking the most isolated nodes)
Loss: ListMLE
Activation Mapping: Gradient-weighted mapping (similar to Grad-CAM) at each convolution scale to understand effects from local to global.

Data Generation & Brief Results

Borrowing from graph representation learning, we arbitrarily combined 3 standard centrality measures on the CORA dataset by SVD and rank the fused centrality measure.

Initial results show that larger receptive fields fit to global eigenvector centrality, while local fields fit well to betweenness and degree.