# Adaptive Universal Generalized PageRank Graph Neural Network

Eli Chien[1] Jianhao Peng[1] Pan Li[2] Olgica Milenkovic[1]

ECE, University of Illinois Urbana-Champaign, USA[1]
CS, Purdue University, USA[2]

March 21, 2021

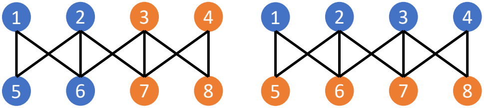
Our code can be found at :https://github.com/jianhao2016/GPRGNN

# Fundamental weaknesses of existing GNNs

- Most of the existing GNN architectures have two fundamental weaknesses, as they repetitively stacking GNN layers and only use final layers as output.
  - Universality issue restricts their learning ability on general graph-structured data.
  - Over-smoothing problem prevents them from being "deep models".

# Universality issue

- Most GNNs only work on homophilic (associative) graphs.
- Homophily principle: nodes from the same class tend to form edges.
- Counterexamples (heterophily/weak homophily):
  - Dating graphs: preferential connection with people of the opposite sex.
  - Protein structure: different classes of amino acids are more likely to connect.
- Design GNN that works in both cases: Universality.




(a) Homophily (b) Heterophily

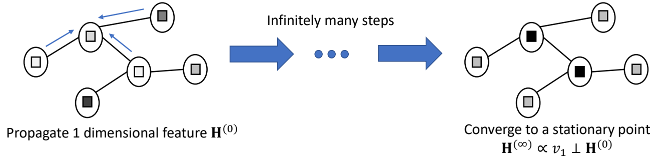
# Over-smoothing problem:

- Practical GNNs are mostly shallow (2-4 layers): better empirical performance.
- If one drop all non-linearity and stack infinitely many layers in GCN:

$$H^{(\infty)}_{\text{GCN}} = \tilde{A}^{\infty}_{\text{sym}} XW = v_1 v_1^T XW,$$

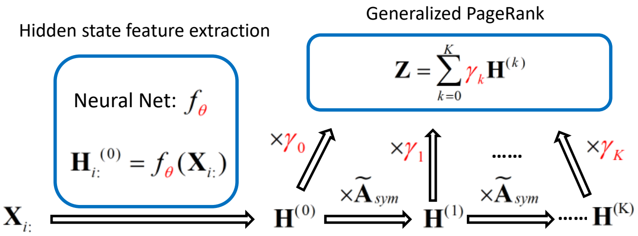
where $v_1$ is the first eigenvector of $\tilde{A}_{\text{sym}}$.

- Each row of $H^{(\infty)}_{\text{GCN}}$ becomes identical (up to a scalar independent of $X$).
- $\Rightarrow$ Lose feature information.



# Our solution: Generalized PageRank GNN (GPR-GNN)

- GPR-GNN further prevents the over-fitting problem and has interpretability on its learned weights.
- Theory of GPR: we analyzed the statistical properties of GPR in (Li et al. 2019[1]).
- Special cases of GPR: Personalized PageRank (PPR), Heat-kernel PageRank (HPR) ... etc.




[1] Pan Li*, Eli Chien*, and Olgica Milenkovic. Optimizing generalized pagerank methods for seed-expansion community detection. In NeurIPS, 2019.

# Why GPR-GNN is universal?

- Equivalence of GPR and polynomial graph filtering $\left(\sum_{k=0}^{K} \gamma_k \tilde{A}_{\text{sym}}^k\right)$
- Learning optimal GPR weights $\Leftrightarrow$ optimal polynomial graph filter coefficients.
- Polynomial filter can approximate arbitrary filter (not restricting to high-pass!).
- Our theorem shows that
  - If all GPR weights are positive: Low-pass filter
  - $\gamma_k = (-\alpha)^k, \alpha \in (0, 1)$: High-pass filter
  - PPR(APPNP/SGC), and HPR are low-pass filter.

# Why GPR-GNN resolves over-smoothing?

Key idea: If the $k^{\text{th}}$ step $H^{(k)}$ is not helpful (increase training loss)$\Rightarrow$ the magnitude of $\gamma_k$ should decay (done by the optimizer).

### Theorem (Informal)

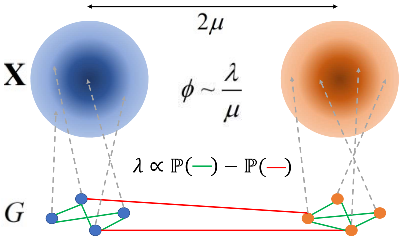
*Assume the graph $G$ is connected and the training set contains nodes from each of the classes. Also assume that $k'$ is large enough so that the over-smoothing effect occurs for $\mathrm{H}^{(k)}, \forall k \geq k'$ which dominate the contribution to the final output $\mathrm{Z}$. Then, the gradients of $\gamma_k$ and $\gamma_k$ are identical in sign for all $k \geq k'$.*

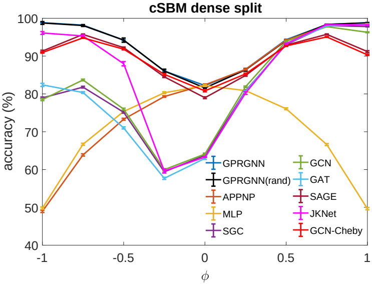
# Experiments: Synthetic data

Test GNNs on graphs with arbitrary levels of homophily and heterophily: contextual Stochastic Block Model (cSBM).

Node features X: Gaussian vectors; Graph $G$: from SBM. $\phi \in [-1, 1]$

- $|\phi| \to 1$: Graph topology gets more informative and node features become less informative.
- Sign of $\phi$: $+$ for homophily and $-$ for heterophily.



# More experiments

- Real-world datasets: in general GPR-GNN has the best performance (especially in heterophily data sets).
- Interpretability: Learned GPR weights match homophily/heterophily of graph data sets.
- Escaping over-smoothing: Even start with over-smoothing scenario, GPR-GNN can escape from it and learn meaningful results.
- These can be found in our poster and paper.

Thanks for your attention!

Feel free to ask questions at our poster!

Visit our poster at Poster Session 9

May 5, 2021, 5 p.m.-7 p.m. (PDT)