Geometric deep learning and node classification: An application of Graph Convolutional Networks to citation networks

Yifan Qian[1], Paul Expert[2], Pietro Panzarasa[1], and Mauricio Barahona[2]

[1]School of Business and Management, Queen Mary University of London, London, UK
[2]Department of Mathematics, Imperial College London, London, UK

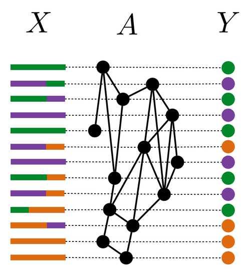
1. Introduction

1.1 Geometric deep learning with graphs

- The adoption of deep learning in the field of network science has been lagging behind until very recently.
- Geometric deep learning (GDL) [1] refers to a fairly broad set of emerging techniques attempting to generalize deep neural models to graphs.
- Recently, the method of Graph Convolutional Networks (GCNs) [2], which uses additional information from available graphs, has been shown to perform particularly well in classification tasks.

1.2 Ingredients of GCNs

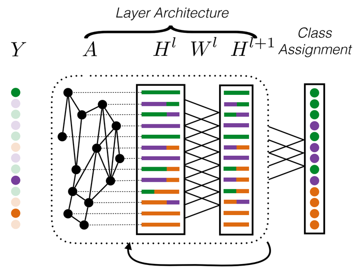
GCNs classification is a transductive semi-supervised machine learning method that relies on three main ingredients:



- an $N \times D$ dimensional feature matrix $\mathbf{X}$, where each row characterizes an object with a (high) dimensional vector;
- an $N \times N$ graph $\mathbf{A}$, representing the interactions between the N objects to be classified; and
- an $N \times K$ 'ground truth' assignment matrix $\mathbf{Y}$, which is used in the training stage and for the *a posteriori* evaluation of the classification performance.

1.3 Layer-wise propagation rule in GCNs

$$\boxed{H^{l+1} = \sigma\left(\hat{A} H^l W^l\right)}$$



- $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$.
- $\tilde{A} = A + I_N$ and $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$.
- W^l is a layer-specific trainable weight matrix.
- H^l is the matrix of activations in the l^{th} layer where $\boxed{H^0 = X}$.
- $\sigma(\cdot)$ is an activation function.

2. Motivation

Can additional information from the graph always be beneficial to the performance of GCNs?

We consider three limiting cases of GCNs:

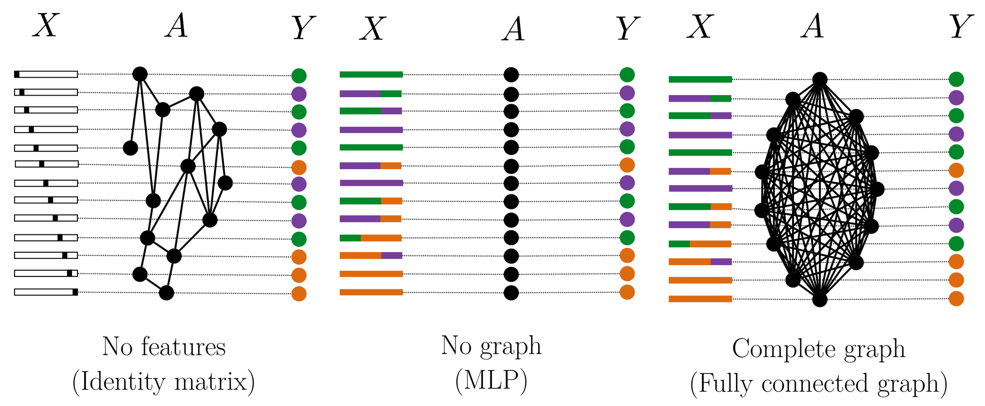
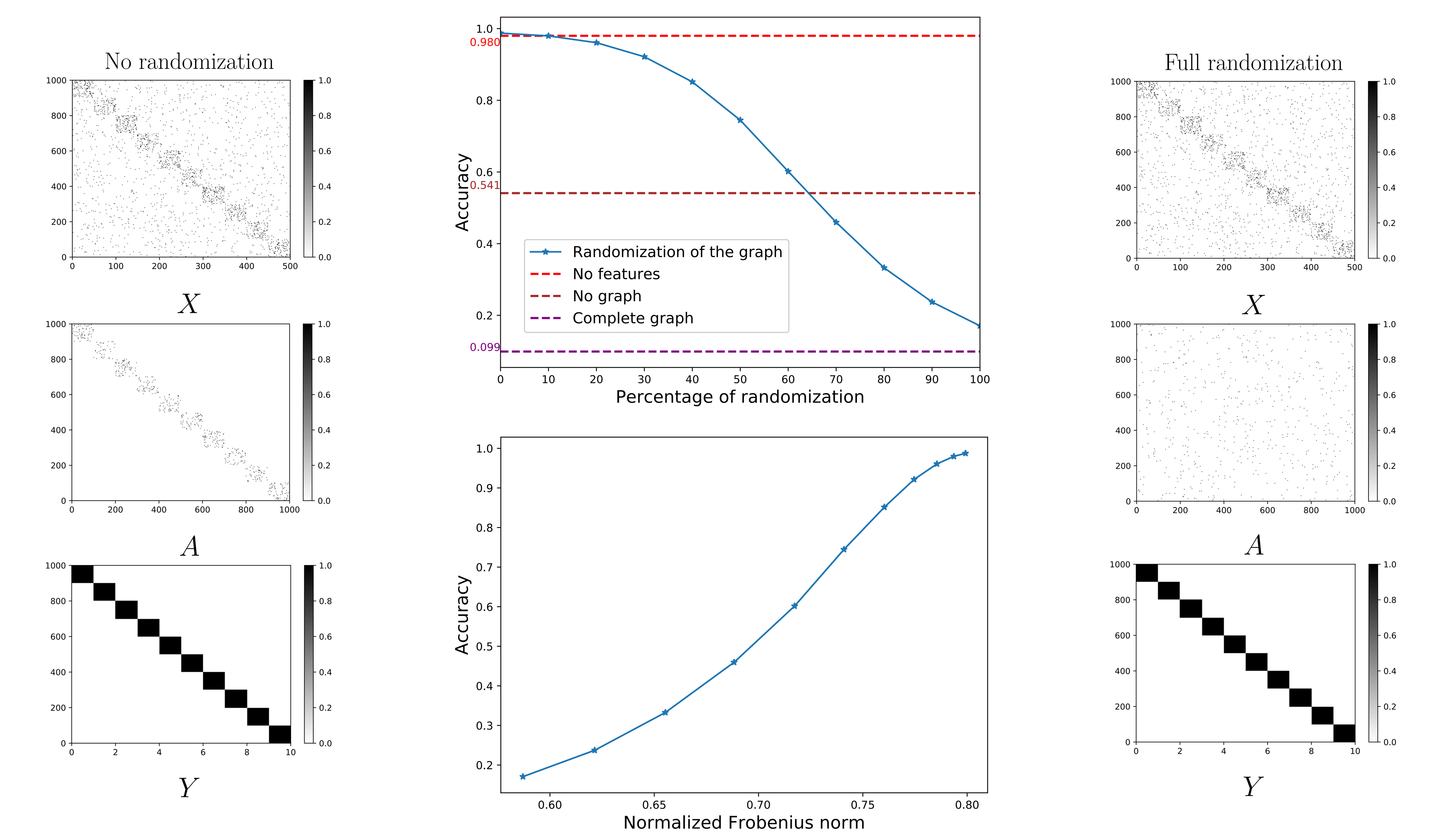


Table 1: Limiting cases of GCNs in CORA and Wikipedia

Data set	Nodes	Edges	Classes	Features	Cases	Accuracy
CORA	2,485	5,069	7	1,433	**GCNs**	**0.810 ± 0.007**
					No features (i.e., identity matrix)	0.631 ± 0.003
					No graph (i.e., MLP)	0.543 ± 0.001
					Complete graph (i.e., fully connected graph)	0.154 ± 0.008
Wikipedia	20,525	215,771	12	100	GCNs	0.358 ± 0.001
					No features (i.e., identity matrix)	0.213 ± 0.007
					No graph (i.e., MLP)	**0.442 ± 0.008**
					Complete graph (i.e., fully connected graph)	O.O.M.

Information from the graph can potentially increase the performance of GCNs (e.g., CORA), but this is not always the case (e.g., Wikipedia)!

Graph randomization in the toy model



3. Hypothesis and goal

- **Hypothesis**:
  A certain degree of alignment among X, A and Y is needed to obtain good performance of GCNs, and any degradation in the information content leads to worsened performance.
- **Goal**:
  Linking the classification performance of GCNs with the alignment of features, the graph, and ground truth.

4. Randomization: Testing the hypothesis

- Randomizing the graph (by rewiring edges while keeping the degree distribution unchanged).
- Randomizing the features (by swapping the feature vectors at random).

Data set:

- A toy model (a synthetic stochastic block model graph).

Table 2: Statistics of the toy model

Data set	Nodes	Edges	Classes	Features
Toy model	1,000	2,568	10	500

5. Quantifying the alignment

Proposing a synthetic measure of spectral alignment based on principal angles [3] among subspaces spanned by the features X, the Laplacian of the graph A and the ground truth Y.

- Alignment matrix S:

$$S = \begin{bmatrix} \cos(\theta_{X\_X}) & \cos(\theta_{X\_A}) & \cos(\theta_{X\_Y}) \\ \cos(\theta_{A\_X}) & \cos(\theta_{A\_A}) & \cos(\theta_{A\_Y}) \\ \cos(\theta_{Y\_X}) & \cos(\theta_{Y\_A}) & \cos(\theta_{Y\_Y}) \end{bmatrix}$$

- Frobenius norm $\|S\|$:

$$\|S\| = \sqrt{\sum_{i=1}^{3} \sum_{j=1}^{3} |S_{ij}|^2}$$

- Normalized Frobenius norm $\|S_n\|$:

$$\boxed{\|S_n\| = \frac{\|S\| - \|S\|_{min}}{\|S\|_{max} - \|S\|_{min}}}$$

where $0 \leq \|S_n\| \leq 1$. The larger $\|S_n\|$, the better the alignment.

Constructing subspaces for X, A and Y:

- PCA for features:
  $X \longrightarrow \mathcal{F} \in R^{N \times k_X}$
- Eigendecomposition for the Laplacian of the graph:
  $A \longrightarrow \mathcal{L} \longrightarrow \mathcal{U} \in R^{N \times k_A}$
- PCA for ground truth:
  $Y \longrightarrow \mathcal{C} \in R^{N \times k_Y}$

6. Two subsets of Wikipedia

Table 3: Statistics of Wikipedia data sets

Data set	Nodes	Edges	Classes	Features	Modularity
Wikipedia	20,525	215,771	12	100	2.98
Wikipedia1	2,414	8,285	5	100	3.95
Wikipedia2	16,216	164,784	5	100	2.97

Wikipedia1 = [Health, Mathematics, Nature, Sports, Technology]
Wikipedia2 = [Culture, Geography, History, Society, People]

Table 4: Summary of results in Wikipedia, Wikipedia1 and Wikipedia2

Data set	Normalized Frobenius norm	Cases	Accuracy
Wikipedia	**0.063**	GCNs	0.358 ± 0.001
		No features (i.e., identity matrix)	0.214 ± 0.007
		No graph (i.e., MLP)	**0.442 ± 0.008**
		Complete graph (i.e., fully connected graph)	O.O.M.
Wikipedia1	**0.444**	**GCNs**	**0.860 ± 0.004**
		No features (i.e., identity matrix)	0.840 ± 0.004
		No graph (i.e., MLP)	0.773 ± 0.008
		Complete graph (i.e., fully connected graph)	0.172 ± 0.142
Wikipedia2	**0.086**	GCNs	0.539 ± 0.001
		No features (i.e., identity matrix)	0.395 ± 0.003
		No graph (i.e., MLP)	**0.592 ± 0.005**
		Complete graph (i.e., fully connected graph)	O.O.M.

Normalized Frobenius norm corresponds to (i) $k_X = k_Y = 12$, and $k_A = 512$ for Wikipedia, and (ii) $k_X = k_Y = 5$, and $k_A = 512$ for Wikipedia1 and Wikipedia2.

7. Conclusion

- We have confirmed that a certain degree of alignment of the features, the graph, and the ground truth is needed to obtain good performance of GCNs, and any degradation in the information content leads to worsened performance.
- Our findings establish a direct geometric relationship between the performance of the GCNs classification and the spectral alignment of the features, the graph, and the ground truth.
- This allows us to deepen our understanding of the synergy between graphs and feature vectors in machine learning.

References

[1] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst.
Geometric deep learning: going beyond Euclidean data.
IEEE Signal Processing Magazine, 34(4):18–42, 2017.

[2] Thomas N. Kipf and Max Welling.
Semi-supervised classification with graph convolutional networks.
In *International Conference on Learning Representations (ICLR)*, 2017.

[3] Gene H Golub and Charles F Van Loan.
Matrix Computations, volume 3.
JHU Press, 2012.

Contact information

Scan me

Yifan Qian (PhD student)

- Research interests: Computational social science, Complex networks, Machine learning
- Email: y.qian@qmul.ac.uk
- Twitter: @qian_yifan
- Website: https://haczqyf.github.io/