Geometric deep learning with graphs: Linking the classification performance of Graph Convolutional Networks with the alignment of graphs and features

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

[1]School of Business and Management,
Queen Mary University of London, London, UK

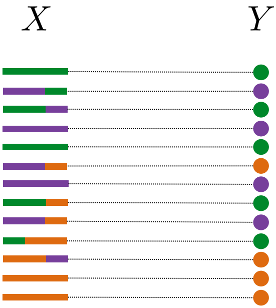
[2]Department of Mathematics,
Imperial College London, London, UK

Machine Learning in Network Science
NetSci18 Satellite, Paris

12th June 2018

Background

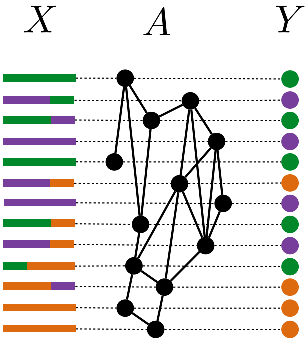
Supervised classification



Ingredients of Multilayer perceptron (MLP)

- a feature matrix $\mathbf{X} \in R^{N \times C}$.
- a ground truth assignment matrix $\mathbf{Y} \in R^{N \times F}$.

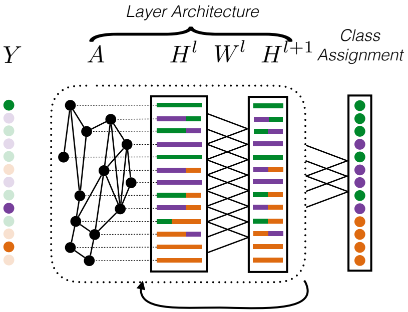
Semi-supervised node classification



Ingredients of Graph Convolutional Network (GCN) (Kipf and Welling, 2017)

- a feature matrix $\mathbf{X} \in R^{N \times C}$.
- an adjacency matrix $\mathbf{A} \in R^{N \times N}$.
- a ground truth assignment matrix $\mathbf{Y} \in R^{N \times F}$.

Layer-wise propagation rule in GCN



Architecture of GCN

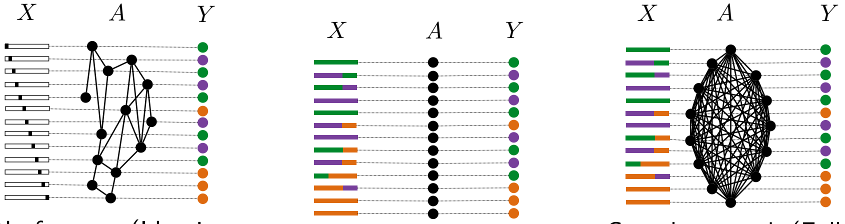
$$\boxed{H^{l+1} = \sigma\left(\hat{A}H^l W^l\right)}$$

- $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$.
- $\tilde{A} = A + I_N$ 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.

Motivation

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

We consider three limiting cases of GCN:



No features (Identity matrix) | No graph (MLP) | Complete graph (Fully connected graph)

Motivation (continued)

Data set	Nodes	Edges	Classes	Features	Cases	Accuracy
CORA	2,485	5,069	7	1,433	**GCN**	**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

Data set	Nodes	Edges	Classes	Features	Cases	Accuracy
Wikipedia	20,525	215,771	12	100	GCN	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 GCN (e.g., CORA), but this is not always the case (e.g., Wikipedia)!

Hypothesis and goal

- **Hypothesis**:

  A certain degree of alignment among X, A and Y is needed to obtain good performance of GCN, and any degradation in the information content leads to worsened performance.

- **Goal**:

  Linking the classification performance of GCN with the alignment of features, the graph, and ground truth.

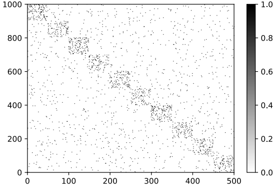
Randomization: Testing the hypothesis

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

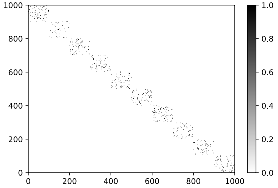
Data set:

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

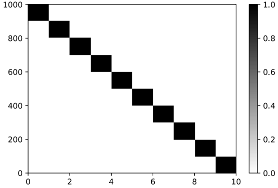
Data set	Nodes	Edges	Classes	Features
Toy model	1,000	2,568	10	500



X



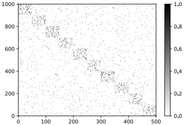
A



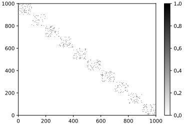
Y

Graph randomization

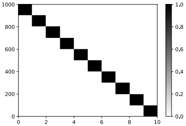
No randomization



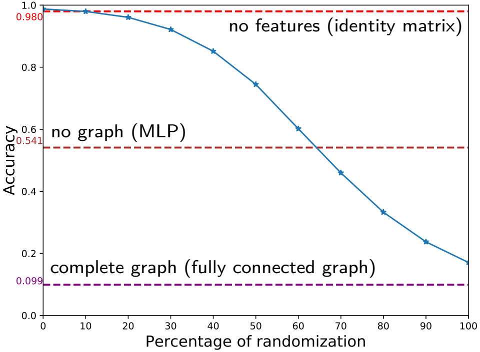
X



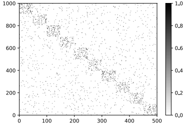
A



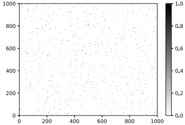
Y



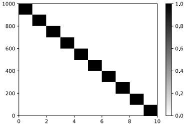
Full randomization



X



A



Y

Quantifying the alignment among X, A and Y

Proposing a synthetic measure of spectral alignment based on principal angles (Golub and Van Loan, 2012) among subspaces spanned by the features, the Laplacian of the graph and the ground truth.

Alignment matrix S:

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

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.

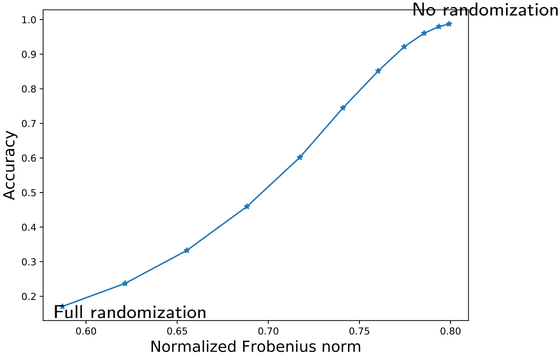
Constructing subspaces

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:
  $Y \longrightarrow \mathcal{C} \in R^{N \times k_Y}$

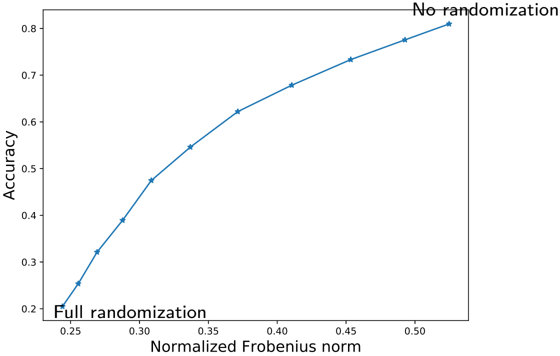
Frobenius norm vs GCN performance: Toy model

Toy model:
$k_X = 10$
$k_A = 512$
$k_Y = 10$



Frobenius norm vs GCN performance: CORA

CORA:
$k_X = 7$
$k_A = 512$
$k_Y = 7$



Two subsets of Wikipedia

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]

Two subsets of Wikipedia (Continued)

Data set	Normalized Frobenius norm	Cases	Accuracy
Wikipedia	0.063	GCN	0.358 ± 0.001
		No features (i.e., identity matrix)	0.021 ± 0.007
		No graph (i.e., MLP)	$\mathbf{0.044 \pm 0.008}$
		Complete graph (i.e., fully connected graph)	O.O.M.
Wikipedia1	0.444	**GCN**	$\mathbf{0.860 \pm 0.004}$
		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	GCN	0.539 ± 0.001
		No features (i.e., identity matrix)	0.395 ± 0.003
		No graph (i.e., MLP)	$\mathbf{0.592 \pm 0.005}$
		Complete graph (i.e., fully connected graph)	O.O.M.

Summary

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

Golub, G. H. and Van Loan, C. F. (2012). *Matrix Computations*, volume 3. JHU Press.

Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.