MSc in Advanced Computer Science
FHS Computer Science; Mathematics and Computer Science;
Computer Science and Philosophy.

GRAPH REPRESENTATION LEARNING

Michaelmas Term 2023

Submission deadline 12 noon, Wednesday 3rd January 2024, via Inspera.

There is a total of 100 marks available for this paper, you should attempt all parts of the paper.

NB: You must not discuss this examination paper with anyone.

Graph Representation Learning

Please use the official LaTeX template to type your answers available in Moodle. Please respect the notation from Table 1 in your answers whenever applicable:

Table 1: Notation.

σ	An element-wise non-linearity.
t	Iteration, or layer t.
$d^{(t)}$	The dimension of a vector at iteration t.
d	The dimension of a vector, and an abbreviation for $d^{(0)}$.
$\mathbf{1}^d \in \mathbb{R}^d$	A d-dimensional vector of all 1's.
$\mathbb{I}^d \subseteq \mathbb{R}^d$	The set of d-dimensional one-hot vectors.
$\mathbb{B}$	Boolean domain $\{0, 1\}$.
$\boldsymbol{b}^{(t)} \in \mathbb{R}^d$	A bias vector.
$\boldsymbol{x}_u \in \mathbb{R}^d$	The feature of a node $u \in V$.
$\boldsymbol{h}_u^{(t)} \in \mathbb{R}^{d^{(t)}}$	The representation of a node $u \in V$ at layer t.
$\boldsymbol{z}_u = \boldsymbol{h}_u^{(T)} \in \mathbb{R}^{d^{(T)}}$	The final representation of a node $u \in V$ after T layers/iterations.
$\boldsymbol{W}_x^{(t)} \in \mathbb{R}^{d^{(t+1)} \times d^{(t)}}$	Learnable parameter matrix at layer t.
MLP	A multilayer perceptron with ReLU as nonlinearity.

Question 1

In this question, we consider simple, undirected graphs $G = (V_G, E_G, \boldsymbol{X}_G)$ with matching node features $\boldsymbol{X}_G \in \mathbb{R}^{|V_G| \times d}$ and write $\boldsymbol{X}_G = [\boldsymbol{x}_1^\top \dots \boldsymbol{x}_n^\top]$, where each vector $\boldsymbol{x}_i = \boldsymbol{h}_i^{(0)}$ represents the initial feature of a node $i \in V_G$. We denote by $\mathcal{N}(u)$ the set of neighbors of a node u in V_G.

Let $0 < t \leq T$ for some $T \in \mathbb{Z}^+$, and consider the model architecture $\mathcal{S}$, which is composed of T layers of the form:

$$\boldsymbol{h}_u^{(t)} = \mathsf{MLP}^{(t)}\left(\boldsymbol{W}_s^{(t)} \boldsymbol{h}_u^{(t-1)} + \boldsymbol{W}_n^{(t)} \sum_{v \in \mathcal{N}(u)} \boldsymbol{h}_v^{(t-1)} + \boldsymbol{b}^{(t)}\right)$$

Let $0 < t \leq T$ for some $T \in \mathbb{Z}^+$, and consider the model architecture $\mathcal{M}$, which is composed of T layers of the form:

$$\boldsymbol{h}_u^{(t)} = \mathsf{MLP}^{(t)}\left(\boldsymbol{W}_s^{(t)} \boldsymbol{h}_u^{(t-1)} + \boldsymbol{W}_n^{(t)} \sum_{v \in \mathcal{N}(u)} \frac{\boldsymbol{h}_v^{(t-1)}}{|\mathcal{N}(u)|} + \boldsymbol{b}^{(t)}\right)$$

After $T \in \mathbb{Z}^+$ layers, we obtain the final node embeddings $\boldsymbol{h}_u^{(T)} = \boldsymbol{z}_u$ for each node $u \in V_G$. For brevity, we will write $\tilde{\mathcal{X}}$ to refer to a model which is a result of setting the parameters in a given model architecture $\mathcal{X}$.

(a) Let G be a graph, where $\boldsymbol{x_u} = \boldsymbol{1}^d$ for all $u \in V_G$. Identify a function f which associates with every graph G a mapping $f(G) : V_G \to \mathbb{B}$, such that there is *no* parametrization of $\mathcal{S}$ satisfying $f(G)(u) = \tilde{\mathcal{S}}(G)(u)$ for all graphs G, and for all nodes $u \in V_G$. (5 marks)

(b) In this part, we are interested in graph-level functions and we assume that the pooling function is given by:

$$\boldsymbol{z}_G = \max_{u \in V_G} \boldsymbol{z}_u.$$

Let G be a bounded-degree graph, where $\boldsymbol{x}_u \in [0,1]^d$ for all $u \in V_G$, and let f be a function such that $f(G) \in \mathbb{B}$.

Prove or disprove: If there exists a parameterization of $\mathcal{S}$ satisfying $f(G) = \tilde{\mathcal{S}}(G)$ for all bounded-degree graphs G then there exists a parameterization of $\mathcal{M}$ satisfying $f(G) = \tilde{\mathcal{M}}(G)$ for all bounded-degree graphs G. (6 marks)

(c) Let G be a bounded-degree graph, where $\boldsymbol{x_u} \in \mathbb{I}^d$, for all $u \in V_G$.

Prove or disprove: For any single-layer parameterization of $\mathcal{M}$, there exists a single-layer parameterization of $\mathcal{S}$ such that the node representations computed by $\tilde{\mathcal{M}}$ and $\tilde{\mathcal{S}}$ are identical for all bounded-degree graphs G, and for all nodes $u \in V_G$. (9 marks)

Question 2

In this question, we consider simple, undirected graphs $G = (V_G, E_G, \boldsymbol{X}_G)$ where the nodes are divided into two disjoint sets: $V_G = R \cup P$, $R \cap P = \varnothing$. We consider matching node features $\boldsymbol{X}_G \in \mathbb{R}^{|V_G| \times d}$ and write $\boldsymbol{X}_G = [\boldsymbol{x}_1^\top \dots \boldsymbol{x}_n^\top]$ as before. We distinguish the neighbors of a node by their node types and write:

$$\mathcal{N}_R(u) = \{v \mid (u,v) \in E_G, \underset{v}{\cancel{u}} \in R\} \qquad \textit{and} \qquad \mathcal{N}_P(u) = \{v \mid (u,v) \in E_G \underset{v}{\cancel{,}} u \in P\}$$

We focus on model architecture $\mathcal{A}$ composed of T layers of the form:

$$\boldsymbol{h}_u^{(t)} = \mathsf{MLP}^{(t)}\Big(\boldsymbol{W}_s^{(t)}\boldsymbol{h}_u^{(t-1)} + \sum_{v \in \mathcal{N}_R(u)} \boldsymbol{W}_R^{(t)}\boldsymbol{h}_v^{(t-1)} + \sum_{v \in \mathcal{N}_P(u)} \boldsymbol{W}_P^{(t)}\boldsymbol{h}_v^{(t-1)} + \boldsymbol{b}^{(t)}\Big),$$

where $0 < t \leq T$, and we assume initial node features given as $\boldsymbol{x}_u \in \mathbb{I}^d$ for all $u \in V_G$.

(a) Introduce an iterative test κ as a variation of the 1-dimensional Weisfeiler Leman algorithm which aligns with $\mathcal{A}$. Given a graph $G = (V_G, E_G, c)$ with an initial coloring $c(G) : V_G \to \mathbb{I}^d$, this test should compute, for every node $u \in V_G$, and every iteration t, a code $\kappa^{(t)}(G)$ satisfying the properties stated in parts (b) and (c). (3 marks)

(b) Prove: For all graphs G with initial node features $\{\boldsymbol{x}_u = c(G)(u) | u \in V_G\}$, for all T layer models of $\mathcal{A}$, for all nodes $u, v \in V_G$, and for all $0 \leq t \leq T$:

$$\boldsymbol{h}_u^{(t)} \neq \boldsymbol{h}_v^{(t)} \Rightarrow \kappa^{(t)}(G)(u) \neq \kappa^{(t)}(G)(v).$$

(7 marks)

(c) Prove: For all graphs G with initial node features $\{\boldsymbol{x}_u = c(G)(u) | u \in V_G\}$, for all nodes $u, v \in V_G$, and for all choices of $T \in \mathbb{Z}^+$, there exists a model of $\mathcal{A}$ such that for all $0 \leq t \leq T$:

$$\kappa^{(t)}(G)(u) \neq \kappa^{(t)}(G)(v) \Leftrightarrow \boldsymbol{h}_u^{(t)} \neq \boldsymbol{h}_v^{(t)}.$$

(10 marks)

(d) Let G be a graph and let $\mathcal{Q}(u) = \mathcal{N}_R(u) \cup \mathcal{N}_P(u)$ for all $u \in V_G$. Consider the model architecture $\mathcal{B}$ which is composed of T layers of the form:

$$\boldsymbol{h}_u^{(t)} = \mathsf{MLP}^{(t)}\Big(\boldsymbol{W}_s^{(t)}\boldsymbol{h}_u^{(t-1)} + \sum_{v \in \mathcal{Q}(u)} \boldsymbol{W}_Q^{(t)}\boldsymbol{h}_v^{(t-1)} + \sum_{v \in V_G} \boldsymbol{W}_V^{(t)}\boldsymbol{h}_v^{(t-1)} + \boldsymbol{b}^{(t)}\Big),$$

where $0 < t \leq T$. Formally compare the expressive power of model architectures $\mathcal{A}$ and $\mathcal{B}$ in terms of their power to distinguish node representations in two graphs G_1 and G_2.
(10 marks)

Question 3

In this question, you are expected to investigate the properties of graph neural networks with a specific focus on their *expressive power*. Your task is to propose and conduct experiments based on certain desiderata of your choice pertaining to the dimension of expressiveness and report your findings. The study can focus on many different dimensions of your choice, including (but not limited to):

- Analyzing different choices of model components and their implications on the expressive power,
- Comparing two (or more) model architectures in the context of expressive power, and analyzing their strengths or weaknesses,

- Proposing an extension (or simplification) of an existing graph neural network architecture, and studying their expressive power,
- Analysing the expressive power of variations of graph neural networks on different types of input graphs (e.g., multi-relational, planar, directed graphs), or by putting restrictions on inputs (e.g., feature distributions).

The report should not exceed *three* pages using the margins and font size of the provided LaTeX template. The report should include a link to your anonymous code-base (i.e., a github repository). The study will be assessed based on the following criteria: (i) motivation, clarity, and presentation (15 marks), (ii) originality and novelty (12 marks), (iii) coherence and depth of the study (7 marks), (iv) relations to the concepts discussed throughout the course (6 marks),(v) scholarship, i.e., its framing in the existing literature (5 marks) (vi) the presence of a balanced critical self-evaluation (5 marks). (*Please ensure that the resources are used in a balanced and well-justified manner: The study is not evaluated according to the number of experiments, or datasets, but according to its merits in accordance to the outlined criteria.*)

(a) State a *small* research question related to expressive power of graph neural networks and explain your motivation for the proposed study. *(Example: How does model architecture A compare to model architecture B in terms of expressive power?)*

(b) Propose means to study the research question and give an outline of the overall approach, clearly identifying the goals of the chosen study. *(Example: An experimental setup which aims to test a hypothesis through, e.g., a synthetic dataset or using existing datasets, as well as a theoretical justification for the study.)*

(c) Clearly state your methodology, empirical setup, hyper-parameters, and the assumptions underpinning the study.

(d) Report your empirical/theoretical/conceptual findings, identifying whether or not the results support the initial hypothesis. You can use visuals, figures, and tables to present your findings in a structured manner. You are expected to relate and compare your results to the relevant literature and to the concepts from the course.

(e) Provide a detailed discussion relating the results to the original motivation, as well as a critical perspective regarding the study.

(f) The report should conclude with an outlook stating any additional studies that need to be conducted to reach more conclusive statements.

(50 marks)

LAST PAGE