深度圖神經網路過度擠壓及過度平滑之減緩方法研究

侯冠宇[1]，劉立頌[2]，王亮潛[3]

國立中正大學電機工程系

[1] E-mail: willy20140201@alum.ccu.edu.tw

[2] E-mail: aliu@ee.ccu.edu.tw

[3] E-mail: g12415170@alum.ccu.edu.tw

摘要

在圖神經網路的應用中，加深網路層數可以讓模型學習到更複雜的特徵，並得到更大範圍的結構訊息。然而加深圖神經網路會面臨過度擠壓和過度平滑的問題。過去研究專注於減緩過度擠壓或是過度平滑問題上，而本研究以組合這些方法來提供同時解決這兩個問題的方案。文中探討不同資料集特色並檢驗不同組合方法的效能，所使用的方法有Virtual edge、JKNet、SDRF、PairNorm、GroupNorm、GCNII 等方法。結果顯示，Virtual Edge+GCNII 在同質性高的資料集中表現最佳，而 JKNet+GCNII 則在同質性低的資料集中效果最佳。此外，本研究通過量化節點表示的平均距離進行定量分析，進一步證實了方法的有效性。

關鍵字：圖神經網路、過度擠壓、過度平滑

Abstract

In the application of graph neural networks (GNNs), increasing the number of layers allows the model to learn more complex features and capture broader structural information. However, deepening GNNs introduces challenges such as over-squashing and over-smoothing. Previous studies have focused on mitigating either over-squashing or over-smoothing, while this study combines these methods to provide a solution that addresses both problems simultaneously. The paper explores the characteristics of different datasets and evaluates the performance of various combined methods, including Virtual Edge, JKNet, SDRF, PairNorm, GroupNorm, and GCNII. The results show that the Virtual Edge + GCNII method performs best in high-homogeneity datasets, while JKNet + GCNII is most effective in low-homogeneity datasets. Additionally, this study also conducted a quantitative analysis by measuring the average distance of node representations, further confirming the effectiveness of the proposed methods.

Keywords：GNN, over-squashing, over-smoothing

1. 緒論

在現實世界中也有很多資料屬於非序列資料，例如社交網路、分子結構、交通網路等，這些資料通常使用由節點和邊組成的圖表示，若使用傳統的深度學習方法容易失去重要的拓樸訊息。而圖神經網路 (Graph Neural Network, GNN) 是專門處理圖資料的神經網路架構，其相比於過去的神經網路架構可以對節點的特徵和節點之間的關係有更好的處理[1]。然而相較一般神經網路，GNN 深度通常相對較淺，只有少數幾層。這與 GNN 所使用的訊息傳遞方式有關，與全連接網路和卷積神經網路不同，GNN 使用訊息傳遞神經網路 (Message Passing Neural Network) [2]作為訊息傳遞的框架，如 GCN [3]。在實際應用中可以觀察到，GNN 的層數越深，模型雖然能夠學習到更複雜的特徵，但可能反而導致性能下降。這是由於訊息傳遞神經網路傳播方式造成的兩個現象：過度擠壓 (over-squashing) [4]和過度平滑 (over-smoothing) [5]。

過度擠壓是指圖中節點經過遠距離傳播後造成訊息被壓縮導致失真，而過度平滑指的是隨著神經網路層數增加，所有節點特徵表示會收斂到相近的數值，造成分類困難。這兩個原因導致深度 GNN 性能的下降。本研究提出組合方法同時改善深度 GNN 過度擠壓及過度平滑之問題，能夠有效學習圖上節點的遠程依賴關係，並防止節點表示的過度擠壓和過度平滑。

本論文分為五個章節，第二章首先介紹 GNN 的傳播方式並過度擠壓及過度平滑問題；第三章探討九種不同組合方法之效能，而第四章提出對不同資料集有效方法的驗證；最後在第五章提出總結。

2. 文獻探討

2.1. 過度擠壓

過度擠壓可以解釋為來自指數成長的感受野 (receptive field) 的資訊被壓縮為固定長度的節點向量。因此，模型難以獲得來自遠端節點的訊息，並且僅從訓練資料中學習短程訊息。不同於卷積神經網路，GNN 的感受野是動態的，隨著網路層數的增加而指數增長。這是因為 GNN 的每一層都會聚合來自其相鄰節點的資訊，而這些相鄰節點又可以聚合來自自身相鄰的資訊。然而，過大的感受野會導致大量節點的資訊被壓縮到固定大小的向量中，從而引發過度擠壓的問題。尤其是來自遠距離節點的資訊，由於經過多層訊息傳遞神經網路的多次聚合，被擠壓得尤為嚴重，導致 GNN 無法有效傳遞遠距離節點的訊息。

Gilmer 等人提出了一種名為 Virtual Edge [2] 的方法。為了捕捉遠距離依賴關係，Virtual Edge 方法在輸入圖中增加一個虛擬節點，並使其與圖中所有節點相連。它們能夠在每次訊息傳遞過程中收集和分發全局訊息，從而優化訊息的傳遞，特別是在跨越長距離的依賴關係方面。

Topping 等人提出一種開創性的圖重新佈線方法，稱為 Stochastic Discrete Ricci Flow (SDRF) [6]，旨在透過基於曲率的圖重新佈線策略來解決 GNN 中過度擠壓的問題。他們認為負曲率的邊是過度擠壓發生的原因。透過增加額外的邊，SDRF 減輕了過度擠壓的問題，並改善圖上的訊息流動。

Xu 等人使用跳躍連接的方式來緩解過度擠壓，提出 JKNet (Jumping Knowledge Network) [7]。JKNet 透過跳躍連接在聚合資訊時會跳過一些相鄰節點。因此，在聚合步驟中被壓縮的資訊較少。

比較這三種方法，在圖重新佈線的方法中，相較於 SDRF，Virtual Edge 方法在設計和實施上相對簡單，而 SDRF 能在保留圖結構的情況下能夠有效地緩解過度擠壓問題。JKNet 透過跳躍連接，有效地保留了不同層級的訊息，使得節點表示在深度網路中不會被壓縮，從而減緩過度擠壓問題。同時 JKNet 的結構易於實施，能夠在保持計算效率的同時，提升模型性能。

2.2. 過度平滑

GNN 中的過度平滑問題指的是隨著網路深度的增加，節點表示變得越來越相似，導致模型在下游任務的表達能力下降的現象。造成過度平滑的原因是由於訊息傳遞網路的架構，每個節點相互連接，隨著網路的深度增加，每個節點會收到越來越多其他節點的資訊，各個節點的感受野可能會包含來自相同的節點資訊，並且隨著網路深度加深，感受野中資訊相似的程度越高，造成節點表示變得難以區分，進而導致模型分辨能力下降。

過去減緩過度平滑問題的方法可以大致分為歸一化 (normalization) [8]及改變傳播機制。歸一化方法在一般的情境指的是調整資料的尺度，使其壓縮在相似的範圍內，從而提高數據的一致性和穩定性。在 GNN 中，歸一化的目的是為了維持每一層節點表示的距離，避免節點表示隨著網路層數加深而收斂成相似的數值，從而提高性能的穩定性。這類方法包括 Pair Normalization (PairNorm) [8] 及 Group Normalization (GroupNorm) [9] 等方法。PairNorm 通過保持節點表示的距離恆定，來防止特徵過於相似，從而保持節點表示的多樣性。GroupNorm 則將節點分成多的小組，對每個小組內的節點特徵獨立進行歸一化，減少節點特徵在傳播過程中的平滑效應。

改變傳播機制的方法重點在於通過改變 GNN 內部訊息傳遞和更新機制來減緩過度平滑問題，從而提高模型的性能和穩定性。GCNII (Graph Convolutional Network via Initial residual and Identity mapping) [10]通過在每一層加入初始特徵和恆等映射來保持初始特徵在深層傳播過程中的重要性，從而減少過度平滑的影響。

PairNorm 和 GroupNorm 的優點在於它們可以應用在任何 GNN 架構中，插入在 GNN 層中間，作為一個歸一化層使用，這使得這兩種方法在處理過度平滑問題時具有極大的靈活性和適應性，能夠在不同 GNN 架構中發揮作用。GCNII 使用簡單的架構改進來減緩過度平滑，避免了額外的學習參數。因此在本文選擇 PairNorm、GroupNorm 和 GCNII

進行後續的研究，這些方法能夠更直接地針對過度平滑問題進行處理，並且具有更廣泛的適用性。

3. 研究方法

本節將介紹本論文的研究架構，以圖 1 所示，其他內容會在各章節介紹。

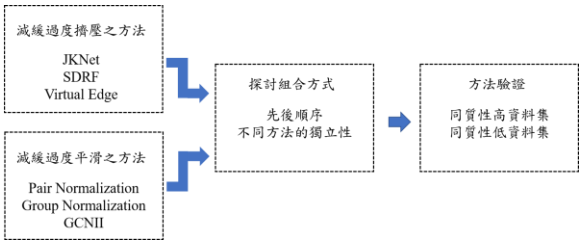

圖 1. 研究架構圖

3.1. 減緩過度擠壓之方法

3.1.1. Virtual Edge

Virtual Edge 方法 [2] 使用一個連接到所有節點的虛擬節點，使節點之間的訊息傳遞距離縮短為兩次跳躍 (2 hops)。訊息首先從源節點傳遞到虛擬節點，然後再從虛擬節點傳遞到目標節點。添加虛擬節點的目的是為了減少訊息在傳遞過程中因聚合相鄰節點而產生的壓縮。在 GNN 中，訊息會因為訊息傳遞神經網路的機制，在每次傳遞時被壓縮，這可能導致訊息丟失，虛擬邊使訊息能夠跳過部分聚合步驟，從而減少訊息壓縮的次數。

3.1.2. JKNet

Xu 等人對於通用的 GNN 聚合方法提出疑問，質疑一個固定的訊息傳遞結構是否能為所有節點和任務提供最佳的表示[7]。深度 GNN 可能導致節點表示過於平滑化，同時造成過度擠壓，而淺層 GNN 可能導致不穩定或資訊聚合不足。為此，作者提出跳躍知識網路對 GNN 提出了兩個架構改進，即跳躍連接和可選擇的聚合機制。

跳躍知識網路的運作機制與常見的訊息傳遞網路一樣，每一層都透過聚合前一層的鄰域來增加節點的影響距離，每一層隱藏層的節點表示會透過跳躍連接傳遞至最後一層，跳躍知識網路的最後一層將會聚合來自中間每一層隱藏層的節點表示，以此來獲得不同層，即不同距離的節點資訊。同時跳躍連接避免了遠距離節點經過多層訊息傳遞網路，也就是多次鄰域聚合，所造成的資訊擠壓。

3.1.3. SDRF

Topping 等人以曲率的角度理解 GNN 中普遍存在的過度擠壓問題[6]。通過對圖中節點曲率進行深入分析，提出 SDRF 作為一種圖重新佈線方法，其目的是為了通過重新分配圖中資訊流動的路徑來緩解過度擠壓問題。為了評估圖上的曲率，作者提出平衡福爾曼曲率 (Balanced Forman Curvature) 的概念。平衡福爾曼曲率可以有效地度量圖中節點的曲率，從而對圖中的曲率提供更深入的理解，其公式如下式 (1) 所示。

$$\mathrm{Ric}(i,j) = \frac{2}{d_i} + \frac{2}{d_j} - 2 + 2\frac{|\#\triangle(i,j)|}{\max\{d_i,d_j\}} + \frac{|\#\triangle(i,j)|}{\min\{d_i,d_j\}} + \frac{(\gamma_{\max})^{-1}}{\max\{d_i,d_j\}}\left(|\#^i_{\square}| + |\#^j_{\square}|\right) \tag{1}$$

其中 i,j 表示圖中節點，d 表示節點的度數 (degree)，$\#\triangle(i,j)$ 指的是基於邊 i~j 形成三角形的節點集合，$\#^i_{\square}(i,j)$ 指的是透過邊 i~j 形成四角形的節點集合，且四角形內部不存在對角線，而 $\gamma_{\max}(i,j)$ 指的是基於邊 i~j 形成四邊形的最大數量。

透過平衡福爾曼曲率可以評估圖中各處的節點，文獻[6]通過實驗分析，證明了負曲率是導致過度擠壓的主要原因。而作者提出的 SDRF 方法旨在透過福爾曼曲率評估圖結構的曲率，在曲率最小的邊周圍增加額外的輔助邊，以此來增加該負曲率邊周圍的曲率，達成減緩過度擠壓的效果。

3.2. 減緩過度平滑之方法

3.2.1. PairNorm

PairNorm 方法 [9]之關鍵概念是透過保持總節點對 (node pair) 之間的距離恆定，從而減緩過度平滑造成節點表示差異越來越小的問題。作者首先定義 TPSD (Total Pairwise Squared Distance)來度量節點表示之間的距離，如下式 (2) 所示。

$$TPSD(X) = \Sigma_{(i,j)\in e}\|x_i - x_j\|_2^2 + \Sigma_{(i,j)\notin e}\|x_i - x_j\|_2^2 \tag{2}$$

其中 i, j 表示圖中節點，e 表示邊，x_i 表示節點 i 的節點表示。TPSD 前項表示相連節點對之間的總距離，後項表示不相連節點對之間的總距離。PairNorm 通過保持 GNN 中每一層的 TPSD 值恆定，縮小相連點之間的距離，同時將不相連節點之間的距離擴大，從而實現更好的分類效果。

3.2.2. GroupNorm

Zhou 等人推廣在卷積神經網路中使用的歸一化方法，提出適用於 GNN 中的 GroupNorm 層[9]。其與 PairNorm 一樣，為適用於大部分 GNN 的歸一化方法，用來減緩隨著 GNN 層數增加，節點表示變得難以區分的過度平滑問題。作者表示，並非所有資料集中同類別的節點都相連在一起，若按照 PairNorm 將相連結點拉近，不相連結點推遠的分群方式可能導致不同類別的節點被分為同一類別，進而導致次佳的分類結果。GroupNorm 的做法是先將節點表示相近的節點分類為同一群組，期望解決 PairNorm 可能將不同類別分類為同一群組的問題，接著在不同組內獨立做歸一化，分離不同組之間的節點分布，並減緩過度平滑的問題。

3.2.3. GCNII

GCNII 架構 [10]以兩種方式擴展了原始的 GCN [3]架構。首先，它加入初始殘差連接，將輸入的特徵資訊傳遞至每個隱藏層，這確保了最後的節點表示至少包含來自輸入層的部分訊息，即使堆疊了許多層，輸入層的訊息也不會完全丟失。其次引入了隱藏層權重的恆等映射，將每層的權重矩陣加入一個單位矩陣，其可以確保一個深層的 GCNII 模型至少可以達到與其淺層版本相同的性能[5]。原始 GCN 的節點表示如式 (3) 所示:

$$H^{(l+1)} = \sigma(\tilde{P}H^{(l)}W^{(l)}) \quad (3)$$

其中$\tilde{P}$為圖卷積矩陣，而 GCNII 則是在前項 $\tilde{P}H^{(l)}$ 加入輸入特徵$H^{(0)}$，並在後項權重矩陣$W^{(\ell)}$中加入單位矩陣I_n。

3.3. 減緩過度擠壓及過度平滑之組合方法

雖然過度擠壓及過度平滑同時存在於深度 GNN，但過去的研究大多集中於單獨解決這些問題。本研究認為在解決深度 GNN 中的過度擠壓及過度平滑問題時，僅使用單一方法難以有效應對這兩個問題。因此本研究認為通過組合不同的方法來同時解決這兩個問題，以提高深度 GNN 的效能是可行的方案。在探索不同組合方法中，須注意不同組合的先後順序，在緩解過度擠壓及過度平滑的方法中，有些方法必須先使用，有些方法後使用，有些則是兩個方法同時並行使用。

本研究選擇三種減緩過度擠壓及三種減緩過度平滑問題的方法，共可以組合成 9 種不同的組合方法。在減緩過度擠壓的方法中，Virtual Edge 及 SDRF 皆是透過調整輸入的圖資料本身來達到減緩過度擠壓的效果，因此在資料輸入模型之前完成。這兩個方法在組合對抗過度平滑的方法時需要先使用，然後再使用 PairNorm、GroupNorm 及 GCNII 來對抗過度平滑。

JKNet 的作法是將每一層的 GNN 透過跳躍連接將每層 GNN 的輸出連接至最後一層，再透過所選擇的聚合函數聚合。因此這種方法是調整 GNN 層與層之間的連接方式，需要與主要的 GNN 架構，如 GCN 並行使用。而 PairNorm 及 GroupNorm 屬於 GNN 使用的歸一化方法，對每一層的節點表示進行歸一化，用來對抗深度 GNN 過度平滑的問題，其應用於每一層 GNN 之後。

GCNII 的做法是在 GCN 中加入初始殘差及恆等映射，這種方法是調整 GNN 內部的訊息保留和特徵融合機制來提升模型效能。由於這種調整是在網路內部進行，使 GCNII 具有較強的獨立性，可以方便地與其他方法結合使用。

3.4. 同質性探討

在本文中，將探討不同組合方法在不同性質資料集上的表現。根據本研究的實驗，圖神經網路在不同性質資料集上會面臨不同的挑戰，因此適合的組合方法也會有所不同。我們將同質性(Homophily)的度量定義為整個圖中每個節點的相鄰節點中與其標籤相同的節點比例的平均值[11]。

4. 實驗與驗證

本研究將使用同質性高的 CiteSeer 資料集和同質性低的 Cornell 資料集來驗證這些推薦方法。

4.1. 資料集介紹

CiteSeer 資料集包含了 3312 篇科學論文，這些論文被分類為 6 個類別包含代理人、人工智慧、資料庫、資訊檢索、機器學習、人機互動。資料集中的節點代表論文，邊代表引用關係。每個節點有 3703 個特徵，每個特徵是詞向量，用來描述論文的內容。Cornell 資料集屬於 WebKB 的子資料集。WebKB 資料集是卡內基美隆大學從各大學電腦科

學系蒐集的網頁資料集。其中節點代表網頁，邊是它們之間的超連結。節點共分為五個類別，包含學生、職員、教員、課程、專案。每個節點有 1703 種特徵，用來描述網頁的內容。

4.2. 實驗設計

本小節將以 GCN 模型作為基礎架構測試 9 種不同組合方法在同質性高及同質性低的資料集上的表現，以推薦在不同情況下適用的組合方法。

4.2.1. 實驗一 CiteSeer 資料集方法驗證

圖 2 為 9 種不同組合方法在同質性高的 CiteSeer 資料集中的表現。可以發現 Virtual Edge+GCNII 和 SDRF+GCNII 兩種組合方法的表現相近，均顯示出較高的模型效能。然而，SDRF 的方法牽涉到基於曲率的圖重新佈線，這一過程的時間複雜度高達 $O(N^3)$，其中 N 是節點的數量。這是因為 SDRF 需要計算所有節點對的曲率，並進行多次迭代以優化圖結構。相對地，Virtual Edge 的方法僅需要將一個虛擬節點加入現有圖中，並將其連接到所有其他節點，其時間複雜度為 O(N)。在考慮到計算效率和實際應用的可行性，本研究選擇將 Virtual Edge+GCNII 作為在同質性高的資料集下的推薦組合方法。

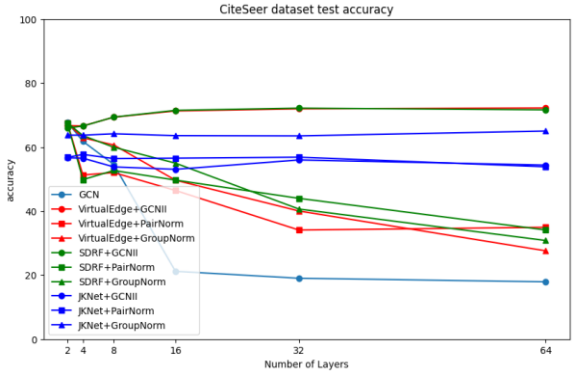

圖 2. 九種組合方法在 CiteSeer 資料集的準確率

4.2.2. 實驗二 Cornell 資料集方法驗證

圖 3 為 9 種不同組合方法與比較基準模型 GCN 在同質性低的 Cornell 資料集上的表現。在此資料集中，9 種不同組合方法的準確率相較同質性高的 CiteSeer 資料集來的相近。在模型深度較深時，即 32 到 64 層，大部分的組合方法準確率位在 20% 至 45%之間。唯獨 JKNet+GCNII 的準確率最高，能將準確率維持 50%以上，推測 JKNet 的跳躍連接特性能有效避免資訊過度擠壓，而 GCNII 的初始殘差連接與恆等映射技術則進一步增強了模型對深層訊息處理的能力，使得此組合不僅能夠應對過度擠壓，也優化了訊息的長距離傳播，從而在學習任務中達到更高的準確率。綜合評估，JKNet+GCNII 在所有層數上的表現最穩定，並且本研究的目標著重在 GNN 層數加深時的表現，因此本研究以 JKNet+GCNII 作為在同質性低的資料集上的最佳組合方法。

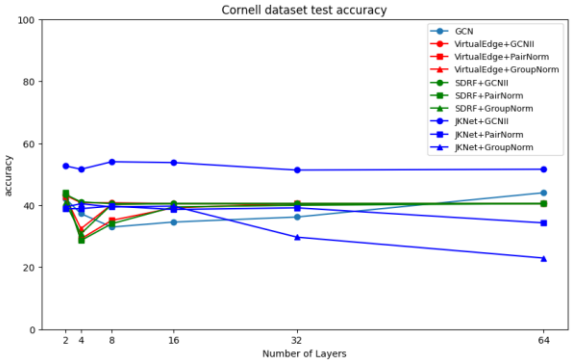

圖 3. 九種組合方法在 Corenell 資料集的準確率

4.2.3. 實驗三 Virtual Edge+GCNII 方法

本研究將使用餘弦相似度來檢驗 Virtual Edge+GCNII 組合方法在不同層數下節點表示的平均距離 (Mean Average Distance, MAD)。MAD 的數值越小，代表整體 GNN 節點表示之間的距離較短，節點表示過於相似，造成過度平滑的發生；反之數值越大則意味著節點表示之間的距離較遠，模型分辨的能力較高。這個實驗設計旨在檢驗該方法對抗過度平滑的能力。圖 4 中 Virtual Edge+GCNII 組合方法能有效對抗這種過度平滑的趨勢。該方法不僅保持了節點之間的一定距離，使節點特徵在多層結

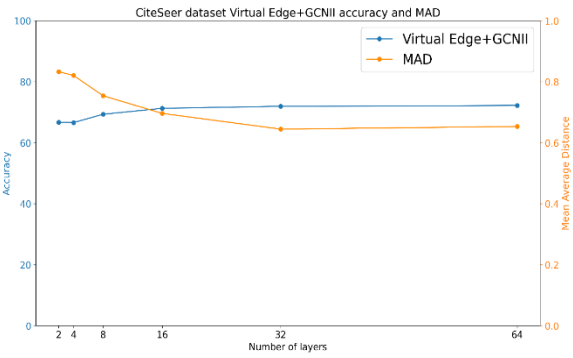

圖 4. Virtual Edge+GCNII 在 CiteSeer 資料集的準確率與 MAD 比較圖

構中保持了一定的差異性，而且隨著網路層深度的增加，模型的準確率漸漸的提升，從 66.69%提高到 72.20%。這表明 Virtual Edge+GCNII 不僅能夠減緩過度平滑，且能夠利用網路深度的優勢。

4.2.4. 實驗四 JKNet+GCNII 方法

本研究以餘弦相似度來檢驗 JKNet+GCNII 組合方法在不同層數下節點表示的平均距離，這個實驗設計旨在檢驗該方法對抗過度平滑的能力。本實驗所使用的準確率數據與實驗二是在相同的 Cornell 資料集上運行的相同數據組，以確保比較的一致性和有效性。觀察圖 5，JKNet+GCNII 組合方法 MAD 下降僅從 2 層的 0.7309 下降至 0.6135，由此可見此組合方法能夠有效維持節點表示的多樣性，從而對抗過度平滑問題。並且透過 JKNet 獲取不同層節點表示的能力，可以提高模型對遠距離節點資訊的學習能力，在任何層數保持準確率的穩定。

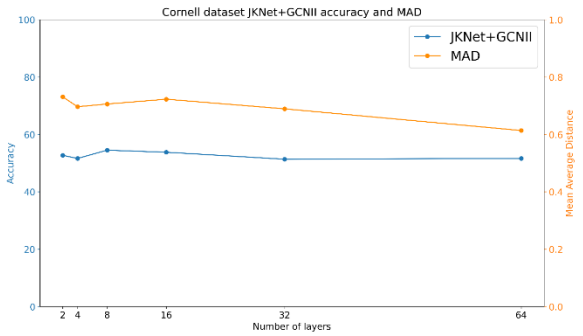

圖 5. JKNet+GCNII 在 Cornell 資料集的準確率與 MAD 比較圖

5. 結論與未來展望

本研究致力於探討 GNN 中的兩個主要問題：過度擠壓及過度平滑。這些問題在深度 GNN 中尤為突出，會導致訊息傳遞的瓶頸和節點表示的同質化，從而影響模型表現。為了解決這些問題，本研究提出兩種推薦組合方法：Virtual Edge+GCNII 和 JKNet+GCNII，並通過實驗驗證了這兩個方法在不同性質資料集上的有效性。實驗結果顯示，Virtual Edge+GCNII 和 JKNet+GCNII 均能有效維持節點表示的多樣性，並提升模型準確率，在 64 層時準確率分別提升 54.31%及 7.57%，且兩者在不同層數皆能表現出良好的穩定性。透過推薦的組合方法，成功減緩過度擠壓及過度平滑，加深了 GNN。

6. 參考文獻

[1] F. Scarselli et al., “The Graph Neural Network Model,” IEEE Trans. Neural Netw., vol. 20, no. 1, pp. 61–80, Jan. 2009

[2] J. Gilmer et al., “Neural message passing for quantum chemistry,” in *ICML*, PMLR, 2017, pp. 1263–1272.

[3] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv:1609.02907*, 2016.

[4] A. Derrow-Pinion et al., “Eta prediction with graph neural networks in google maps,” in *Proceedings of the 30th ACM International CIKM*, 2021, pp. 3767–3776.

[5] T. K. Rusch et al., “A survey on oversmoothing in graph neural networks,” *arXiv:2303.10993*, 2023.

[6] J. Topping et al., “Understanding over-squashing and bottlenecks on graphs via curvature,” *arXiv*:2111.14522, 2021.

[7] K. Xu et al., “Representation learning on graphs with jumping knowledge networks,” in *ICML*, PMLR, 2018, pp. 5453–5462.

[8] L. Zhao and L. Akoglu, “Pairnorm: Tackling oversmoothing in GNNs,” *arXiv:1909.12223*.

[9] K. Zhou et al., “Towards deeper graph neural networks with differentiable group normalization,” *NeurIPS* , vol. 33, pp. 4917–4928, 2020.

[10] M. Chen et al., “Simple and deep graph convolutional networks,” in *ICML*, PMLR, 2020, pp. 1725–1735.

[11] H. Pei et al., “Geom-gcn: Geometric graph convolutional networks,” arXiv preprint arXiv:2002.05287, 2020.

7. 致謝

感謝國科會在研究經費上的補助，計畫編號：112-2221-E-194-014-MY3