構造特徴とグラフ畳み込みを用いたネットワークの半教師あり学習

Semi-supervised learning on network using structure features and graph convolution

立花誠人 [1*]　村田剛志 [1]
Makoto Tachibana[1] Tsuyoshi Murata[1]

[1] 東京工業大学　情報理工学院　情報工学系
[1] Department of Computer Science, School of Computing, Tokyo Institute of Technology

Abstract: Since several types of data can be represented as graphs, there has been a demand for generalizing neural network models for graph data. Graph convolution is a recent scalable method for performing deep feature learning on attributed graphs by aggregating local node information over multiple layers. Such layers only consider attribute information of node neighbors in the forward model and do not incorporate knowledge of global network structure in the learning task. In this paper, we present a scalable semi-supervised learning method for graph-structured data which considers not only neighbors information, but also the global network structure. In our method, we add a term preserving the network structural features such as centrality to the objective function of Graph Convolutional Network and train for both node classification and network structure preservation simultaneously. Experimental results showed that our method outperforms state-of-the-art baselines for the node classification tasks in the sparse label regime.

1　はじめに

近年，SNS(ソーシャルネットワーキングサービス) の普及や，生物学における生体ネットワークの重要性の認識の高まりとともに，グラフ構造を持ったデータの解析が注目を集めている．グラフ構造とは,「ノード」とノード間を結ぶ「エッジ」から構成されるデータ構造である．グラフ構造データ解析の基本的なタスクとしては，ノードのクラスタリング，ラベル推定，ランキング，コミュニティ抽出，リンク予測などが挙げられる．このようなグラフ構造を持つデータの解析は簡単ではない．その理由は，機械学習手法の多くは，固定長の特徴ベクトルで表現されるデータを対象としているためである．近年，グラフ構造からノードの分散表現を獲得する手法である，ネットワークエンベディング [1] の研究が盛んに行われている．これらの手法で計算された分散表現を機械学習モデルの入力として用いることで，既存の手法よりもラベル推定や分類タスクにおいて高い精度を示すことがわかっている．また近年では，グラフ構造に対して畳み込み演算を定義することにより，グラフ構造データをニューラルネットワークの入力として直接用いる手法が研究されている．グラフ畳み込みネットワーク [2] は，グラフを入力として，1 層が 1 ホップ隣接したノードの情報の畳み込みを行うニューラルネットワークであり，半教師あり分類問題において，エンベディングベースの手法よりも高い精度を示している．この手法は，エンドツーエンド学習により柔軟な特徴抽出を可能にしているが，ノードの近傍情報しか考慮していないという問題がある．

そこで本研究では，従来のグラフ畳み込みネットワークの学習に加えて，グラフの構造特徴を用いる手法を提案する．ここで構造特徴とは，グラフ構造からノードごとに導出される指標のことであり，中心性や PageRank などがあげられる．グラフ畳み込みによるノードのラベル推定タスクの学習に加え，構造特徴を保持する学習を同時に行うことで，近傍情報とグラフ全体の構造の両方を考慮した学習が可能になる．論文の引用ネットワーク，ナレッジグラフなど，複数のデータセットを用いて実験を行い，半教師あり分類問題において，提案手法は既存手法と比べて高い精度を得ることを示した．

*連絡先：東京工業大学 情報理工学院 情報工学系
東京都目黒区大岡山 2 丁目 12-1 W8-59
E-mail: tachibana@net.c.titech.ac.jp

2 関連研究

2.1 ネットワークの半教師あり分類問題

本研究が対象とする，ネットワークの半教師あり分類問題について説明する. 半教師あり分類問題は，グラフ，ノードの特徴量，少数のノードのラベルが与えられ，残りのノードのラベルを推定することが目的である．グラフを $G = (V, E)$ とし，$G = (V, E)$ の隣接行列を $A \in \mathcal{R}^{N \times N}$，ただし $N = |V|$ とし，ノード集合を $V = \{v_1, ..., v_l, v_{l+1}, ..., v_n\}$ とする．ここで，l はラベルありデータ数であり，$x_1, ..., x_l$，$x_{l+1}, ..., x_n$ はそれぞれラベルありデータ，ラベルなしデータに対応するノードである．また，ラベルありノード $x_i (1 \leq i \leq l)$ に対する教師ラベルを $y_i \in L$ とする．ネットワークの半教師あり学習分類問題は，以下のように定義される．

- 入力：
  隣接行列：$A \in R^{N \times N}$,
  全ノードの特徴ベクトル：$X = \{x_1, ..., x_n\}$,
  l 個の教師ラベル： $y_1, ..., y_l$
- タスク：
  ラベルなしノード $v_{l+1}, ..., v_n$のラベル推定

ネットワークの半教師あり分類問題が，通常の教師あり学習と異なる点は，ラベル付けされたノードとラベル付けされていないノードのどちらの情報も，学習の際に用いることができる点である．そのため，グラフ構造やノードの特徴量をいかに利用するのかが重要となる．

2.2 ネットワークエンベディング

ネットワークエンベディングとは，ネットワーク構造からノードの低次元のベクトル表現を獲得する手法である．2014 年に DeepWalk[3] が提案されて以来，数多くの研究が行われている．エンベディングで得られた分散表現を機械学習モデルの入力とすることで，既存の手法よりもラベル推定や分類タスクの精度を大きく向上させることに成功している．DeepWalk は，ランダムウォークによって得られたノードの系列に対し，skip-gram モデルを用いることで分散表現を得る手法である．その他にも，LINE[4] や node2vec[5] など，skip-gram をベースにした数多くの派生研究が行われている．

エンベディングを用いて分類問題などのタスクを解く場合，分散表現を生成した後，機械学習モデルの学習を行うという二つのプロセスを行う必要がある．そのため，エンベディングによる手法では，特定のタスクに特化した学習を行うことは難しい．

2.3 グラフ畳み込み

近年，畳み込みニューラルネットワーク (CNN) が画像処理分野で大きな成功を収めている．このモデルの特徴である畳み込み層では，フィルタを移動させながら畳み込み演算を行うことで，移動普遍性や位置普遍性などの重要な特性を抽出することができる．この畳込み演算は，画像や音声と言った規則性のあるデータ構造を対象としている．一方，グラフはノードごとに接続関係が不定形であるため，通常の畳み込み演算をそのまま適用することができない．そこで近年，グラフ構造に対して畳み込みを定義する方法が研究されている．グラフに畳み込みを定義する手法は，spectral covolution と spatial convolution の 2 つに分類される．

2.3.1 spectral convolution

グラフ信号処理の考え方を用いて，グラフ畳み込みを定義する方法が spectral convolution である．[6] では，グラフ信号としてグラフ上のデータを扱い，グラフフーリエ変換を行うことで畳み込み演算を定義している．ここでグラフ上の信号とは、グラフの各ノードに対する d 次元ベクトルの割り当てのことを指す．グラフ信号 X と，パラメータ θ を用いたフィルター $g_\theta = diag(\theta)$ による畳み込み演算は以下のように定義される．

$$g_\theta * X = U g_\theta U^T X \tag{1}$$

ここで U は正規化グラフラプラシアン
$\Delta = I_n - D^{-1/2} A D^{-1/2}$ の固有値ベクトル行列である．上記の式は計算コストが大きいため，[7] ではチェビシェフ近似を用いることで計算量を削減している．

2.3.2 spatial convolution

上記のグラフ信号処理ベースの畳み込み演算は数学的に厳密な手法である一方，ループや多重エッジをもたない、重み付き無向グラフしか扱えないといった問題や，計算コストが大きいという問題がある．そこで，接続関係を用いて直接的に畳み込みを定義する方法が提案されている [8]．近傍ノードの信号を集約して足し合わせた後、得られた信号の情報を用いて自身の信号を更新するという操作により，畳み込み演算を定義している．ノード i における畳込み演算は以下の式で表される．

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i^k} \frac{1}{c_{i,j}} W_1^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right).$$

ここで $c_{i,j}$ は正規化のための定数，$\mathcal{N}_i^k$ は x_i から k 以下の距離でつながっている近傍ノードの集合を表す．

2.3.3 グラフ畳み込みネットワーク

グラフ畳み込みネットワーク (GCN)[2] は，spatial convolution を用いて，半教師あり分類問題を行うモデルである．1 層が 1 ホップ先のノードの情報を畳み込むという制約を与えることによって，計算量を抑えながらも高い精度を示している．グラフ畳み込みネットワークの第 i 層における出力は以下のように表される．

$$H^{(i)} = \sigma(D^{-1/2}\tilde{A}D^{-1/2}H^{(i-1)}W^{(i)}) \qquad (2)$$

ここで，$\tilde{A} = A + I_n$，$I_n \in \mathcal{R}^{n\times n}$ は単位行列，$\tilde{D}_{i,j} = \sum \tilde{A}_{i,j}$ である．ネットワークの入力 $Z^{(0)}$ は特徴ベクトル X，$W^{(i)}$ はネットワーク中の i 層における学習重み行列であり，$\sigma(\cdot)$ は ReLU 関数やシグモイド関数などの活性化関数を表すものとする．[2] では畳み込み層が 2 層のネットワークが用いられている．

2.4 グラフ構造特徴量の利用

古くから，複雑ネットワーク分野において，グラフの構造の特徴を測るためのさまざまな指標が提案されている．グラフ全体の性質を測るものとしては，コミュニティ分割の質を測る指標であるモジュラリティや，グラフ内の三角形の多寡を測る指標であるクラスタ係数などがある．個々のノードごとの重要度を測る指標としては，任意のノードペア間を媒介する度合いを示す媒介中心性，他のノードとの距離の近さを示す近接中心性，Web ページのランキング手法である PageRank などがある．これらの構造特徴を，グラフデータ解析のタスクに活用した研究が数多く存在する．

[9] では，グラフ畳み込みネットワークの学習において，モジュラリティを保持するように学習を行う手法を提案している．以下の目的関数を最適化することで，従来のグラフ畳み込みネットワークよりも半教師ありノード分類タスクの精度を向上させている．

$$L = L_0 - \lambda tr(H^T BH) * (\frac{1}{2e}) \qquad (3)$$

ここで，L_0 はラベル付けされたデータについての教師あり損失関数，H はコミュニティへの割り当て行列，$B_{ij} = A_{ij} - k_i k_j$，$k_i$ はノード i の次数，e は全エッジ数，λ は重み付けパラメータを表す．この手法は，学習時に毎回モジュラリティを計算する必要があり，大規模なグラフを扱う場合，計算コストが非常に大きくなってしまうという問題がある．

そこで本研究では，モジュラリティなどのグラフごとに算出される特徴ではなく，個々のノードごとに計算可能な中心性などの指標に着目する．そのようなグラフの構造特徴を保持する学習を同時に行うことで、精度の向上が期待できる。

3 提案手法

3.1 モデル概要

本研究では，グラフ畳み込みネットワーク [2] をベースとして，グラフの構造特徴を利用した学習を行うモデルを提案する．提案手法の構成を図 1 に示す．グラフ畳み込みネットワークの目的関数に，構造特徴を保持する項を追加することで，教師あり学習であるラベル予測に加え，教師なし学習である構造特徴の保持を同時に行うモデルとなっている．ラベル予測の学習には，通常のグラフ畳み込みネットワークを用いる．構造特徴保持のタスクは，グラフ畳み込みネットワークの中間出力から分岐して実行される．教師あり学習と教師なし学習，それぞれの損失関数が得られた後，それらをトレードオフパラメータ α によって重み付けして足し合わせたものを，最終的な目的関数として用いる．

$$L = (1-\alpha)L_{supervised} + \alpha L_{unsupervised} \qquad (4)$$

ここで，$L_{supervised}$ は教師あり学習の損失関数，$L_{unsupervised}$ は教師なし学習の損失関数を示す．学習時には，二つの目的関数それぞれについて勾配を計算することで最適化を行う．このように，2 つの学習を同時に最適化するニューラルネットワークモデルは，[9][10] でも用いられている．以下で，2 つの学習がそれぞれどのように実行されるかについて詳しく説明する．

3.2 ラベル予測学習

ラベル予測学習のためには，[2] で用いられているものと同様のグラフ畳み込みネットワークを用いる．隣接行列 A と特徴行列 X を入力とし，順伝搬ニューラルネットワークによって演算が行われる．ネットワークの i 層における出力は以下のように表される．

$$Z^{(i)} = \sigma(D^{-1/2}\tilde{A}D^{-1/2}Z^{(i-1)}W^{(i)}) \qquad (5)$$

ここで，$\tilde{A} = A + I_n$，$I_n \in R^{n\times n}$ は単位行列，$\tilde{D_{i,j}} = \sum \tilde{A_{i,j}}$ である．ネットワークの入力 $Z^{(0)}$ は特徴ベクトル X，$W^{(i)}$ はネットワーク中の i 層における学習重み行列であり，$\sigma(.)$ は ReLU 関数やシグモイド関数などの活性化関数を表すものとする．畳み込み層が 2 層のネットワークを用いて，活性化関数として ReLu 関数とシグモイド関数を用いた場合，出力は以下の式で表される．

$$Z = f(X,A) = \mathrm{softmax}(\hat{A}\mathrm{ReLU}(\hat{A}XW^{(0)})W^{(1)}) \qquad (6)$$

$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ である．ここで，$W^{(0)} \in \mathcal{R}^{C\times H}$ は入力層から隠れ層への重み行列であり，$W^{(1)} \in \mathcal{R}^{H\times F}$ は，隠れ層から出力層の重み行列である．ここで，C,

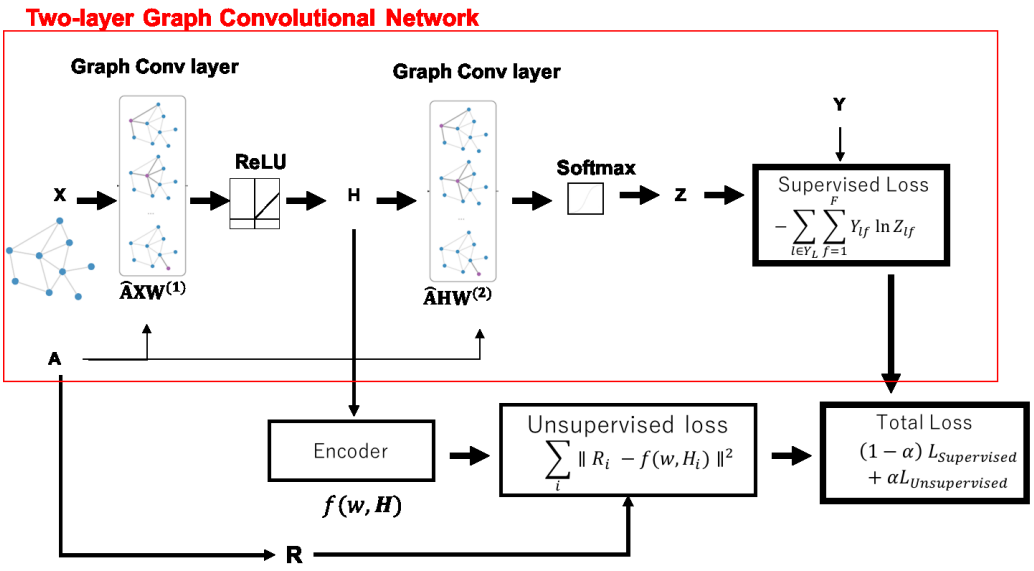

図 1: 提案手法の構成 (2 層グラフ畳み込みネットワークを用いたモデル)

H，F はそれぞれ入力層，中間層，出力層のサイズを表している．損失関数は，各ラベル付ノードのクロスエントロピーを用いて計算される．教師あり損失関数は，以下のようになる．

$$L_{supervised} = -\sum_{l\in Y_l}\sum_{f=1}^{F} Y_{lf} ln Z_{lf} \tag{7}$$

ここで Y_l はラベル付されたノードのセットを表す．

3.3 構造特徴保持学習

構造特徴保持学習の詳細について説明する．

3.3.1 構造特徴行列の計算

まずはじめに，入力された隣接行列 A から各ノードごとの構造特徴を計算する．構造特徴としては，近接中心性，媒介中心性，固有ベクトル中心性，次数中心性，PageRank など，ノードごとに算出可能な指標を用いる．それぞれの特徴量について計算し，それら全てを結合したものを構造特徴行列 $R \in R^{N\times S}$ とする．ここで N はノード数，S は用いる構造特徴の数を表す．この構造特徴行列の計算は，一度行ったら再計算をする必要がないため，モデル全体の計算量に対する影響は少ない．

3.3.2 特徴ベクトルの計算

グラフ畳み込みネットワークの中間出力 H から各ノードの特徴ベクトル $\hat{R}$ を計算する層を導入する．図 1 のように 2 層のグラフ畳み込みネットワークを用いる場合は，1 層目のグラフ畳み込み層の出力 H から，特徴ベクトルを計算することとなる．各ノードについて重みパラメータ ω を用いて特徴ベクトル $\hat{R}$ を以下のように計算する．

$$\hat{R}_i = f(w, H_i) \tag{8}$$

出力された特徴ベクトルが，グラフ構造から計算された構造特徴と等しくなるように学習を行うことで，構造特徴の保持を行う．グラフから計算した構造特徴と出力された特徴ベクトルの 2 乗誤差を用いて損失関数の計算を行う．

$$L_{unsupervised} = \sum_{i=1}^{N} ||R_i - f(w, H_i)||^2 \tag{9}$$

4 評価実験

4.1 モデル構成とパラメータ設定

提案手法の有効性を確認するため，ネットワークの半教師あり分類問題のタスクにおいて実験を行った．モデルは 2 層のグラフ畳み込みネットワークを使用し，隠れ層の活性化関数としては ReLU 関数，出力層の活性化関数としてソフトマックス関数を用いた．また，ドロップアウトと重み減衰を用いて正則化を行っている．重み行列は，Grolot の一様分布を用いて初期化し，最適化のための勾配降下法には Adam[11] を用いた．トレードオフパラメータ α は 0.5 とし，ラベル推測学習と構造特徴保持学習の損失を等しい割合で足し合わせたものを，最終的な目的関数として用いる．全ての実験は

Python3.6 で動作しており，ニューラルネットワークモデルの作成と最適化には Tensorflow を用いた．ネットワーク構造特徴の計算には，NetworkX を用いた．

本実験では，平等に比較を行うため，ハイパーパラメータの設定は [2] における実験で用いられているパラメータ設定と同じ条件で実験を行った．最大エポック数は 200，学習率は 0.01，ドロップアウト率は 0.5 とした．また，バリデーションセットの損失関数が 10 エポック以上にわたり減少しない場合は学習を打ち止めするようにした．隠れ層のサイズは 32 とし，正規化のために L2 正規化を用いた．また，構造特徴として，次数中心性，媒介中心性，近接中心性，固有ベクトル中心性の 4 つを連結したものを利用した．

4.2 データセット

データセットとして，論文の引用関係ネットワークの Cora，Citeseer，Pubmed，およびナレッジグラフの NELL を用いた．論文の引用関係ネットワークにおいては，ノードは論文，エッジは論文の引用関係，ラベルはその論文が属する分野を表す．また，論文の概要を Bag of Words を用いて抽出したものを特徴ベクトルとして用いている．学習用データとは別に，1000 個のラベル付データをテストセットとして用いて評価を行った．

4.3 比較手法

比較手法として，グラフ上でラベルを伝播させることでラベル予測を行う手法である LP[12]，ランダムウォークを用いたエンベディング手法である DeepWalk[3]，サンプリングを用いることで，ラベル情報とグラフ構造を同時に用いてエンベディングを行う手法である Planetoid[13]，本研究のベースとなっている，2 層グラフ畳み込みネットワーク (GCN)[2] を用いた．なお，GCN 以外の手法に関しては実験は行わず，[2] 内で示されている結果を引用した．

表 1: データセット

	ノード	エッジ	クラス	特徴量
Cora	2,708	5,429	7	1,433
Citeseer	3,327	4,732	6	3,703
Pubmed	19,717	44,338	3	500
NELL	9,891	13,142	210	5,414

5 実験結果

5.1 分類精度

学習時に用いる教師ラベルの数を，クラス数に応じて変えながら，分類精度の評価を行った．学習時に用いる 1 クラスごとの教師ラベルの数を変化させた場合の，分類精度を評価したものをそれぞれ示す．表 2-4 は，論文ネットワークデータにおける結果を，表 5 はナレッジグラフにおける結果を示している．GCN 以外の比較手法については，論文から引用した数値を用いたため，空欄となっている箇所がある．全てのデータセットにおいて，提案手法が最も高い精度を示している．とくに，教師ラベルの数が少ない場合において，提案手法は既存手法よりも大きく精度が向上している．これは，教師ラベルが少ない場合には，構造特徴を用いることで，学習時の情報の少なさを補っているためだと考えられる．

表 2: Cora データセットにおける予測精度

教師ラベル数	21	35	70	95	140
LP	–	–	–	–	68.0
DeepWalk	–	–	–	–	67.2
Planetoid	–	–	–	–	75.1
GCN	59.8	68.7	75.3	78.0	79.4
Proposed	**62.0**	**70.1**	**76.2**	**78.4**	**79.8**

表 3: Citeseer データセットにおける予測精度

教師ラベル数	18	30	60	90	120
LP	–	–	–	–	45.3
DeepWalk	–	–	–	–	43.2
Planetoid	–	–	–	–	64.7
GCN	38.3	46.1	58.6	66.3	68.8
Proposed	**42.3**	**49.7**	**63.9**	**67.6**	**69.1**

表 4: Pubmed データセットにおける予測精度

教師ラベル数	9	15	30	45	60
LP	–	–	–	–	63.0
DeepWalk	–	–	–	–	65.3
Planetoid	–	–	–	–	77.2
GCN	62.6	69.1	73.9	76.1	78.5
Proposed	**64.2**	**71.3**	**74.8**	**76.4**	**78.7**

表 5: NELL データセットにおける予測精度

教師ラベル数	105	210	630	1,050	2,100
GCN	28.4	53.0	60.4	64.6	68.0
Proposed	**34.0**	**57.6**	**63.5**	**66.9**	**69.6**

5.2 トレードオフパラメータの影響

提案手法では，トレードオフパラメータによって教師なし損失と教師あり損失の重み付けを行っている．このパラメータによって結果に大きな影響が出る可能性が考えられるため，α を変化させた場合の実験を行った．データセットとしては Cora と Citeseer の 2 種類を

用い，教師ラベルの数はクラスごとに 10 個とした．その他のパラメータは前節と同様の条件とし，パラメータ α は 0 から 1 まで 0.1 刻みで変化させたときの分類精度を図 2 に示す．横軸は α の値，縦軸は分類精度を表している．図からみてとれるように，α の値が極端に大きいか小さい場合でなければ，分類精度に大きな影響を与えないことがわかった．α が極端の値をとる場合，α=0 に近い場合はほとんどラベル情報のみを用いて学習し，α=1 に近い場合は，ほとんど構造特徴情報のみを用いて学習しているため，精度が落ちていると推測できる．一方パラメータを変えてもほとんど結果に影響を与えていない理由としては，学習を進めるにつれて，それぞれのタスクの出力が自然とスケーリングされるためだと考えられる．

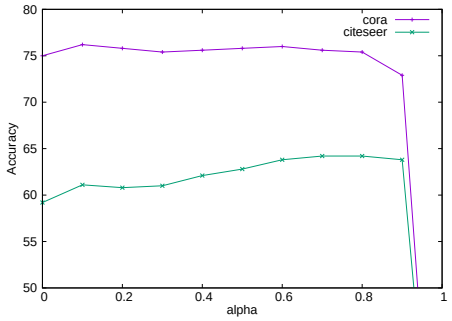

図 2: パラメータ α を変動させたときの分類精度

6 おわりに

本研究では，従来のグラフ畳み込みネットワークの学習に加えて，グラフの構造特徴を考慮する手法を提案した．ノードのラベル予測と構造特徴の保持を同時に学習することで，ノードの近傍情報とグラフ全体の構造特徴の両方を考慮した学習が可能となった．実験により，ネットワークの半教師あり分類問題のタスクにおいて既存手法との比較を行った．その結果，とくに教師ラベルが少ない場合において，従来手法を大きく上回る精度が出ることを確認した．今後の課題として，畳み込み層の数や用いる構造特徴を変化させることで，どのような影響が出るのかを調査する必要がある．また，今回は半教師あり分類問題を対象としたが，リンク予測やグラフ分類などの他のタスクへの適用を検討する予定である．

参考文献

[1] Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems, Vol. 151, pp.78-94, 2018.

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

[3] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’14, pp. 701-710. ACM, 2014.

[4] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In Proceedings of the 24th International Conference on World Wide Web, pp.1067-1077, 2015.

[5] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pp.855-864. ACM, 2016.

[6] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine, Vol. 30, pp. 83-98, 2013.

[7] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In Advances in Neural Information Processing Systems, pp. 3844-3852, 2016.

[8] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In International Conference on Machine Learning, pp.1263-1272, 2017.

[9] Tsuyoshi Murata, Naveed Afzal, Modularity Optimization as a Training Criterion for Graph Neural Networks. Proceedings of the 9th International Conference on Complex Networks In: Cornelius S., Coronges K., Goncalves B., Sinatra R., Vespignani A. (eds) Complex Networks IX. CompleNet 2018. Springer Proceedings in Complexity. pp.123-135, Springer, Cham, 2018.

[10] Jason Weston, Frederic Ratle, Hossein Mobahi, and Ronan Collobert. Deep learning via semi- supervised embedding. In Neural Networks: Tricks of the Trade, pp.639-655. Springer, 2012.

[11] D Kinga and J Ba Adam. A method for stochastic optimization. In International Conference on Learning Representations, 2015.

[12] Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning, pp.912-919, 2003.

[13] Zhilin Yang, William W Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. In Proceedings of the 33rd International Conference on International Conference on Machine Learning Volume 48, pp.40-48, 2016.