

# 重要ノード特定のためのドメイン変化に対して頑健なグラフニューラルネットワーク

Robust Graph Neural Networks for Identifying Important Nodes Across Different Domains



津川 翔 (Sho Tsugawa, Ph. D.)  
筑波大学 システム情報系 准教授 博士(工学)  
(Associate Professor, University of Tsukuba, Institute of Systems and Information Engineering)  
電子情報通信学会 計算社会科学会 IEEE 他  
受賞：電気通信普及財団 テレコム学際研究賞奨励賞 (2024 年)、電子情報通信学会コミュニケーションクオリティ研究賞 (2023 年) Best Tutorial Paper Award of IEICE Communications Society (2020 年) 他

研究専門分野：ネットワーク科学 計算社会科学

グラフニューラルネットワーク (GNN) は、グラフ構造を有するデータに対する深層学習のモデルとして様々な応用が期待され、活発に研究されている[1]。例えば、交通トラフィックの予測や、化合物の機能推定、商品推薦など様々な応用において GNN の有効性が報告されている。

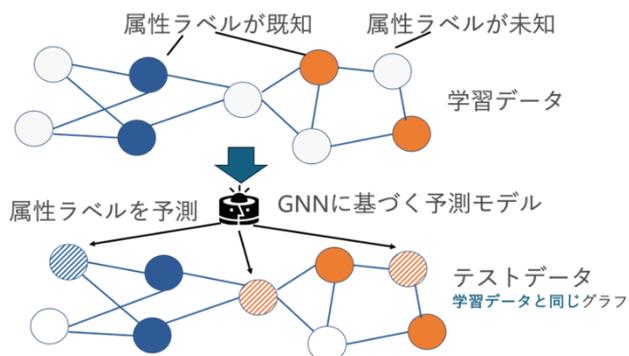


図1 GNNによる典型的なノードラベル推定問題

## あらまし

グラフニューラルネットワーク (GNN) は、グラフ構造を有するデータに対する深層学習の手法として様々な応用が期待されている。基本的な GNN を用いる場合、あるグラフで学習したモデルを、別のグラフに転用することは一般には難しい。本研究は、あるグラフで学習した GNN モデルを異なるドメインの別のグラフの予測に用いるための技術を開発することを目指す。特にグラフにおける重要ノードの特定タスクを題材として、GNN モデルを異なるグラフに適用するための方法を検討する。ソーシャルメディアにおけるインフルエンサーが既知であるデータセットを用いてインフルエンサー予測モデルを学習する。学習したモデルをインフルエンサーが未知である別のソーシャルメディアのデータセットに適用し、インフルエンサー予測の精度を評価する。実験の結果、Unsupervised Domain Adaptive Graph Convolutional Networks (UDA-GCN) と呼ばれるモデルを用いることで、あるグラフで学習したモデルを別のソーシャルグラフ上のインフルエンサー特定に転用可能であることを示す。

## 1. はじめに

GNN の典型的な利用例であるノード属性の推定タスクでは、与えられたグラフの構造とグラフ中の一部のノードに付与された既知のラベル情報から、ノード属性が未知のノードの属性ラベルを予測する(図 1)。例えば、ソーシャルメディアユーザーへの広告配信の効果を高めるために、ユーザーの興味のあるジャンルを予測するタスクを考える。GNN を用いてこのタスクを解く場合、ソーシャルメディアユーザーをノード、ユーザー間のフォロー関係をリンクとして表現したソーシャルグラフが与えられる。さらに、一部のユーザーについては、どのジャンルに興味を持つかを表現した正解のラベル情報が与えられる。これらの情報を教師データとして用いて、GNN は属性ラベルが未知であるユーザーの興味ジャンルを表すラベルを予測するモデルを学習する。

基本的な GNN は学習時に用いるグラフと、予測時のグラフは同一であることを前提とするため、あるグラフで学習した GNN モデルを、別のグラフに適用することは一般には難しい。例えば、本プロジェクトの基礎検討において、学習時に用いるグラフにランダムなリンクの追加や削除などのノイズを加えただけでも、GNN

# 重要ノード特定のためのドメイン変化に対して頑健なグラフニューラルネットワーク

Robust Graph Neural Networks for Identifying Important Nodes Across Different Domains

の予測精度が低下することが示されている[2]。そのため、学習時とは全く構造が異なるグラフにGNNを適用することは一般には難しい。一方、あるグラフでは豊富な教師ラベルのデータが得られるが、別のグラフでは教師ラベルのデータが得られないようなケースも考えられる。そこで、GNNを学習時と異なるグラフの予測に用いるための手法を明らかにすることは重要な研究課題の1つとなっている。

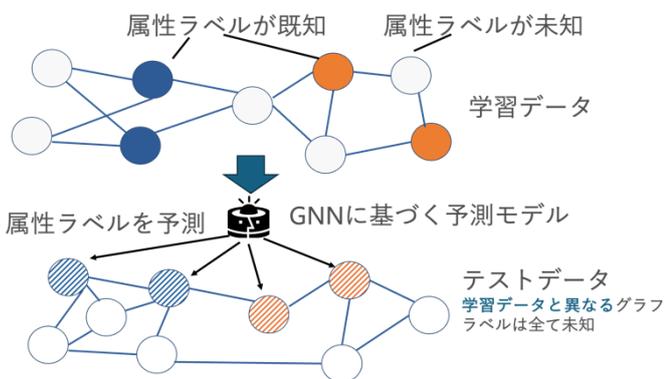


図2 学習元と適用先のグラフが異なる場合のGNNによるノードラベル推定問題

本研究は、あるグラフで学習したGNNモデルを異なるドメインの別のグラフの予測に用いるための技術を開発することを目指し、特にグラフにおける重要ノードの特定タスクを題材として、GNNモデルを異なるグラフに適用するための方法を検討する(図2)。ソーシャルメディアにおけるユーザー間の関係を表現したソーシャルグラフから、インフルエンサーと呼ばれる影響力の強いユーザーを特定する問題を考える。複数のソーシャルグラフのデータセットを用いて、あるデータセットで学習したモデルを別のデータセットの予測に用いる実験を行う。まず、予測対象のグラフと同一のグラフで学習したモデルと、予測対象のグラフとは異なるグラフで学習したモデルの予測精度を比較する。これによって、異なるグラフで学習したモデルを用いてインフルエンサー特定の問題を解くことがどの程度難しい問題であるかを明らかにする。さらに、異なるグラフ間でのドメイン適応\*1が可能なGNNである

Unsupervised Domain Adaptive Graph Convolutional Networks (UDA-GCN) [3]を用いて、あるグラフで学習したモデルで他のグラフのインフルエンサーを高精度に特定することを目指す。

## 2. 研究の方法

本研究では、インフルエンサー特定問題を、ソースドメイングラフが与えられたときに、ターゲットドメイングラフの各ユーザーのラベルを予測する二値分類問題として定式化する。この問題を解くために、ソースドメイングラフの構造とノードがインフルエンサーであるかどうかを表すラベル情報が学習データとして利用可能であることを仮定する。学習データを用いて、ターゲットドメイングラフの各ユーザーがインフルエンサーであるかどうかを予測する分類器を構築する。

本研究では文献[4]で用いた影響力指標に基づき、各ノードがインフルエンサーであるかどうかを表すラベルを定義する。具体的には、ソーシャルグラフに含まれる全ノードの内、影響力指標の高い上位ノードをインフルエンサー、残りを非インフルエンサーとする。本稿で用いる影響力指標は、直接的な影響力および間接的な影響力である(図3)。ユーザの直接的な影響力は、そのユーザの投稿を閲覧したユーザーの総数である。間接的な影響力は、そのユーザの投稿を閲覧したユーザのうち、そのユーザを直接フォローしているユーザを除いたものの数である。本稿では、直接的影響力および間接的影響力が上位1%のユーザーをインフルエンサーとした場合の結果を示す。

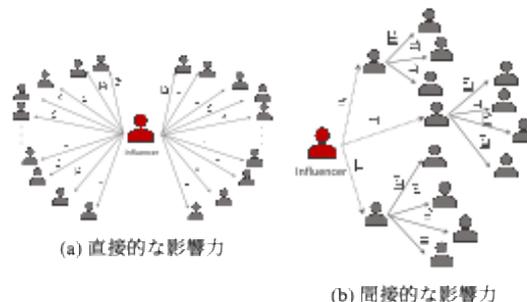


図3 影響力指標のイメージ

# 重要ノード特定のためのドメイン変化に対して頑健なグラフニューラルネットワーク

Robust Graph Neural Networks for Identifying Important Nodes Across Different Domains

本研究では、Twitter (現在の X) ユーザー間のソーシャルグラフとそのツイートおよびリツイートを含む 3 種類のデータセットを用いる。データセットの概要は表 1 の通りである。Ordinary データセットは、2018 年 11 月 19 日から 2018 年 11 月 25 日に投稿された英語ツイートからランダムに抽出されたツイートとそれらのリツイートのデータセットである。Nepal データセットは、2015 年のネパール地震に関するツイートおよびリツイートのデータセットである。Higgs データセットは、2012 年の Higgs 粒子発見に関連するツイートとリツイートのデータセットである。いずれのデータセットにも、ツイートあるいはリツイートに参加したユーザー間のフォロー関係を表現したソーシャルグラフの情報が含まれている。

これらの 3 つそれぞれのデータセットでインフルエンサー特定モデルを学習し、学習元とは異なる適用先のデータセットにモデルを適用してその予測精度を評価する。なお、比較のため、学習元と同じデータセットを適用先とした場合の評価も行う。

予測モデルの構築には、ドメイン適応可能な GNN である UDA-GCN と様々な機械学習に広く用いられる LightGBM[5]を用いる。UDA-GCN、LightGBM のどちらのモデルにおいても、各ノードの特徴量として、代表的な中心性指標\*2 である入次数中心性、出次数中心性、固有ベクトル中心性、媒介中心性、近接中心性、PageRank、k-core 指標を用いた。各データセットでモデルを学習する際、70%のユーザーを学習データとして、残りの 30%のユーザーをバリデーションデータとして用いた。なお学習データにおけるインフルエンサーと非インフルエンサーの比率が不均衡であるため、LightGBM モデルにおいては、インフルエンサーと非インフルエンサーの割合が均等になるようにアンダーサンプリングして学習した。UDA-GCN モデルにおいては、インフルエンサークラスに対する損失関数の重みを大きくして学習した。

評価指標には、分類モデルの評価に広く用いられる F 値を用いる。F 値は適合率と再現率の調和平均として定義される。ここで適合率は、インフルエンサーであると予測したユーザーのうち、実際にインフルエン

サーであったユーザーの割合、再現率は、実際にインフルエンサーであるユーザーのうち、モデルがインフルエンサーであると予測したユーザーの割合である。

## 3. 実験結果

構築したモデルの予測精度を表す F 値を図 4 および図 5 に示す[6]。図の凡例における括弧内の O、N、H はそれぞれ、Ordinary、Nepal、Higgs データセットでモデルを学習したことを表す。O、N、H の記号の付かない凡例は、学習元のデータと予測対象が同一である場合の結果を示している。なお、学習元と予測対象が同一の場合、学習データとテストデータの比率は 7:3 とした。

これらの結果より、適用先とは異なるグラフで学習したモデルを適用先の予測に用いる場合、学習元と適用先が同一である場合よりも予測精度が低くなることわかる。特に LightGBM で構築した予測モデルを異なるグラフに適用する場合、予測精度が著しく低くなることわかる。このことは、異なるグラフへインフルエンサー特定モデルを転用することの困難さを示している。

一方、UDA-GCN と LightGBM を比較すると、学習元と適用先が異なる場合においては、UDA-GCN の方が予測精度が高い傾向にあることがわかる。特に Higgs データセットにおける直接的な影響力に基づくインフルエンサーを特定する場合や、Nepal データセットにおける間接的な影響力に基づくインフルエンサーを特定する場合においては、UDA-GCN の予測精度は、学習元と適用先が同一である場合と遜色ないことがわかる。この結果から、UDA-GCN は、異なるグラフで学習したモデルを別のグラフに適用する際の有望な方法であることが示唆される。ただし、予測対象が Ordinary データセットの場合には、異なるグラフで学習した UDA-GCN モデルの予測精度は、同一のグラフで学習した場合と比べて低い値であり、改善の余地も示唆される。

# 重要ノード特定のためのドメイン変化に対して頑健なグラフニューラルネットワーク

Robust Graph Neural Networks for Identifying Important Nodes Across Different Domains

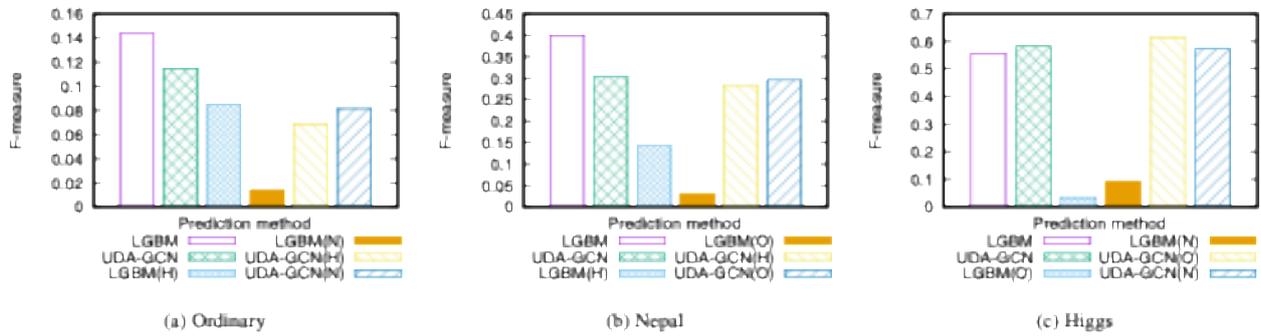


図 4 直接的な影響力に基づくインフルエンサーの予測精度 [6]

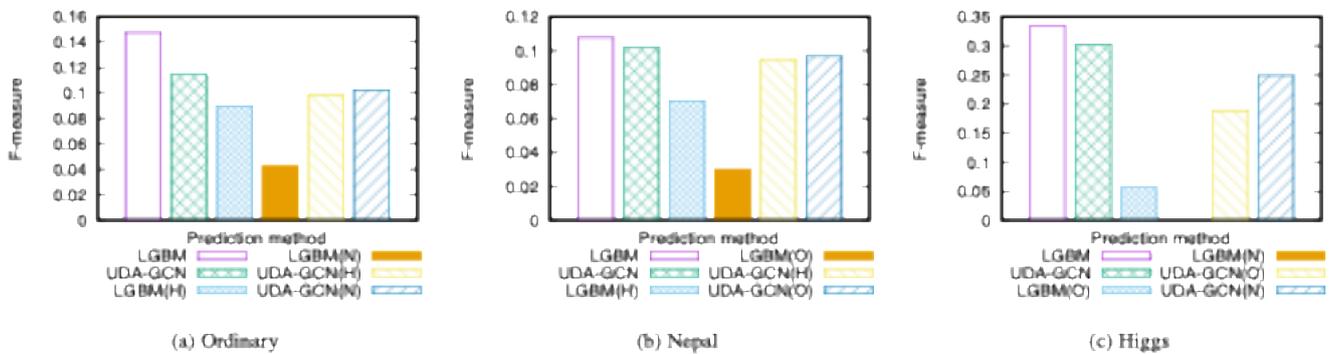


図 5：間接的な影響力に基づくインフルエンサーの予測精度 [6]

表 1 データセットの概要

	Ordinary	Nepal	Higgs
ユーザー数	111,000	270,032	456,626
総リツイート数	558,168	464,466	328,132
平均フォロワー数	28.6	68.5	32.5

## 4. 将来展望

本研究の将来展望として主に、インフルエンサー特定タスク以外での UDA-GCN の有効性の評価と、GNN モデルのドメイン適用可能性の理論的・実験的評価を考えている。本稿で紹介したインフルエンサー特定は、あくまで GNN の 1 つの利用例である。他のタスクにおける UDA-GCN の有効性を評価することや、本研究の知

見を出発点としてモデル自体の改良を行うことが重要な課題であると考えている。また、本研究を通じて、学習元と適用先のデータの組合せによって、ある程度の予測精度が達成できる場合とできない場合があることがわかっている。どのような要因がグラフをまたがるモデルの転用可能性に影響を与えるのかを明らかにすることで、グラフの変化に頑健なモデルの開発に貢献したいと考えている。

## おわりに

本研究は、あるグラフで学習した GNN モデルを学習元と異なる他のグラフに転用する手法を確立することを目指している。この目的を達成するため、特にソーシャルグラフにおけるインフルエンサーを特定するタスクに焦点を当てて、ドメイン適応可能な GNN である UDA-GCN の有効性を評価した。あるソーシャルグラフ

# 重要ノード特定のためのドメイン変化に対して頑健なグラフニューラルネットワーク

Robust Graph Neural Networks for Identifying Important Nodes Across Different Domains

において学習した UDA-GCN モデルを、適用先のグラフに用いた場合と、適用先のグラフで学習した予測モデルをそのまま用いた場合で予測精度を比較した。その結果、UDA-GCN は、適用先の学習データを用いないにも関わらず、適用先の学習データを用いて構築したモデルと比較可能な精度を達成できることを示した。このことは、予測対象のグラフにおけるラベル付きデータが得られない場合には、他のグラフで学習した UDA-GCN モデルを用いることが有望であることを示唆している。ただし、学習元と適用先のデータセットの組合せによっては、UDA-GCN の予測精度にも改善の余地があることも示されている。そのため、今後は、UDA-GCN モデルがドメイン適応可能である条件を理論的・実験的に明らかにし、グラフの変化に対してより頑健な GNN モデルの開発を進める計画である。

## 用語解説

- \*1 **ドメイン適応**・・・機械学習において、学習データと、機械学習モデルの適用先のデータの性質にずれがある場合にはモデルの予測精度が低下してしまう。学習元と適用先のデータのずれを補正して、予測モデルを構築する技術がドメイン適応と呼ばれる。
- \*2 **中心性指標**・・・グラフにおけるノードの重要度を定量化するための指標。

## 参考文献

[1] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)

[2] S. Ando, and S. Tsugawa, "Effects of random errors on graph convolutional networks," in Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS 2022), pp.3285–3294, Jan. 2022

[3] Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. "Unsupervised

domain adaptive graph convolutional networks." in Proceedings of The Web Conference 2020 (WWW 2020), pp. 1457–1467, 2020.

[4] 橋本 祥, 津川 翔, 塚本圭一郎, 猪狩慎太郎, "グラフニューラルネットワークを用いたエゴネットワークからのインフルエンサー推定に関する検討," 電子情報通信学会技術研究報告 (CQ2021-91), pp.82–87, Jan. 2022.

[5] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tiejian Liu. "LightGBM: A highly efficient gradient boosting decision tree." Advances in Neural Information Processing Systems, Vol. 30, , 2017.

[6] 田原 幸太, 津川 翔, "異なるドメインに適用可能なインフルエンサー特定手法に関する一検討," 電子情報通信学会技術研究報告 (CQ2023-12), pp.24–29, Jul. 2023.

## 関連文献

[7] 田原 幸太, 津川 翔, "異なるドメインへ適用可能なインフルエンサー特定手法のフォロワーサイズ別のランキング性能の評価," 電子情報通信学会技術研究報告 (CQ2023-76), pp. 32-37, Mar. 2024.

[8] 田原 幸太, 津川 翔, "ソーシャルネットワークにおける情報拡散者特定手法の異なるドメインへの汎化性能の評価," 電子情報通信学会ソサイエティ大会講演論文集 (B-11-18), Sep. 2023.

この研究は、令和2年度SCAT研究助成の対象として採用され、令和3~5年度に実施されたものです。