

時系列グラフを活用する着目ノードに特化したリンク予測

山口 寛人[†] 前川 政司[†] 佐々木 勇和[†] 鬼塚 真[†]

[†] 大阪大学情報科学研究科 〒 565-0871 大阪府吹田市山田丘 1-5

E-mail: †{yamaguchi.hiroto,maekawa.seiji,sasaki,onizuka}@ist.osaka-u.ac.jp

あらまし グラフのリンク予測は、Web 広告の推薦や消費者の購買予測など幅広い分野で利用されている。中でも、SNS 上で特定の企業アカウント (ノード) に対して将来のリンクを予測し、フォロワーの拡大を目指すというユースケースが知られている。しかし、既存のリンク予測手法はグラフ中の全リンクを予測することを目的としている。そのため、グラフ中のリンクを均等に扱って学習しており、特定の企業アカウントのリンク予測のようなユースケースにおいては十分な精度が期待できない。したがって、本稿では特定のノードに特化したリンク予測手法を提案する。提案手法では、予測対象のノードの時系列的なグラフ構造の変化情報を利用するとともに、予測対象のノードに特化した学習を行うためのリンクのサンプリングとロス関数の設計を行うことで、特定の1ノードに対するリンク予測精度向上を図る。実験では、特定の1ノードに対するリンク予測精度の評価を行う。これにより、予測対象のノードに特化した学習を行うことで特定の1ノードに対するリンク予測精度が向上することを示した。

キーワード リンク予測, 時系列グラフ, グラフマイニング

1 はじめに

グラフはノードと呼ばれる頂点とリンクと呼ばれる辺によってノード間のつながりを表現するデータ構造であり、社会学や生物学、コンピュータ科学など幅広い領域で用いられている [1]。このグラフを用いて表されるデータの例として、ソーシャルネットワーク [1]、消費者の購買履歴 [2]、道路交通ネットワーク [3] などが挙げられる。このようなグラフからコミュニティを抽出するクラスタリングやノードを分類するクラシフィケーション、将来のリンクを予測するリンク予測などのグラフ解析は様々なアプリケーションに利用することが可能であり、幅広く研究が行われている [4-6]。中でも、リンク予測は SNS 上の広告の推薦 [7] や将来の消費者の購買予測 [2]、道路の劣化予測 [3] などに利用可能であり、広く研究が行われている。実際に SNS 上の企業アカウントのフォロワー拡大を目的として、ソーシャルネットワーク上の企業アカウントに対して将来のリンクを予測し、その企業アカウントをフォローする確率が高いユーザーに対して企業アカウントのフォローを推薦するといったような SNS ビジネスが知られている [8]。

このようなリンク予測には2種類の応用例が考えられる。一つ目は、道路の劣化予測のようにグラフ全体のリンク予測が必要なものである。この場合、任意の2ノード間のリンクの存在を予測するために、既知のグラフのリンクを均等に学習する必要がある。二つ目は、前述した SNS 上の企業アカウントのフォロー推薦のように特定の1ノードに対してリンクを予測するものである。この場合、特定のノードのリンクについて予測を行う必要があるが、その他のリンクについては予測する必要はない。そのため、既知のグラフのリンクを均等に学習する必要はなく、特定のノードのリンク予測に特化した学習が必要である。

既存手法では、グラフ全体のリンク予測を目的とした深層

学習モデルが数多く提案されている [5, 6, 9-13]。これらの手法はグラフ全体の構造情報や属性情報を均等に学習することでリンク予測を行う。これらの手法では、jaccard 係数 [14] や Adamic adar [15] など既知のグラフ構造からノード間の類似度を表す指標に基づいて予測を行うヒューリスティックな方法に比べて高精度なリンク予測が可能である。中でも、時系列グラフを活用したリンク予測手法 [11-13] ではグラフ構造の時系列変化の傾向も学習することが可能であり、静的なグラフのみからリンクを予測手法と比べても高精度なリンク予測が可能であることが知られている。一方で、既存手法を特定の1ノードのリンク予測に適用する場合には十分な予測精度を期待できない。既存手法はグラフ中の任意のノード間のリンクに対する予測精度を上昇させるために、グラフ中のリンクを均等に学習しているため、特定の1ノードのリンク予測に対しては十分な精度が期待できない。したがって、1ノードのリンク予測においては、特定の1ノードに特化したモデルの学習を行う必要がある。

本稿では、時系列グラフを活用する特定の1ノードに特化したリンク予測手法を提案する。以降では、この予測対象となる1ノードを着目ノードと呼ぶ。提案手法では、着目ノードのリンク予測に特化した学習を行うために、着目ノード近傍のグラフ構造を重点的に学習するようにリンクのサンプリングを行う。さらに、着目ノードからの距離が近いリンクほど着目ノードのグラフ構造に与える影響が大きくなると考えられるため、サンプリングしたリンクに対して着目ノードからの距離に応じて重み付け学習を行う。これにより、着目ノード近傍のグラフ構造をより重視したリンク予測モデルの学習を実現する。

実験では、提案手法の着目ノードのリンク予測精度を既存手法と比較する。これにより、着目ノード近傍のグラフ構造をより重視してリンク予測モデルを学習することで着目ノードのリンク予測精度が向上することを確認した。また、提案手法のサンプリングに用いるパラメータを変化させてリンク予測精度を

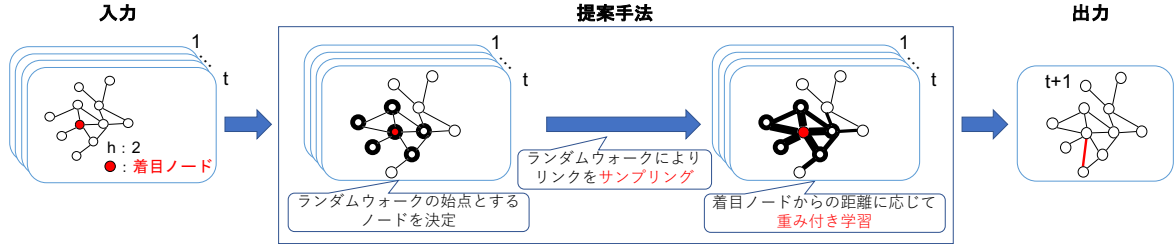


図 1: モデル概要図. 提案手法は, 入力として時刻 t までの時系列グラフと着目ノード, 学習の対象とするリンクのサンプリングに関するパラメータ h を入力し, 時刻 $t+1$ の着目ノードとその他の任意のノードとのリンクの存在を予測する. 提案手法は, ランダムウォークの始点とするノードを決定するステップとランダムウォークにより学習の対象とするリンクのサンプリングを行うステップ, サンプリングしたリンクに対し重み付き学習を行うステップの 3 つのステップからなる. 図は $h=1$ の場合の例を示している. この場合, 提案手法では, 最初に着目ノードから 1 ホップのノードをランダムウォークの始点として決定する. その後, ランダムウォークによりリンクをサンプリングすることで着目ノード近傍のグラフ構造を重点的にサンプリングする. 最後に, サンプリングしてきたリンクを着目ノードからの距離に応じて重み付けしたロス関数を用いてモデルを学習する.

測定することで, 着目ノードのリンク予測において有効なリンクのサンプリング方法を確認した.

本稿の構成は以下の通りである. 2 章で提案手法について詳細に述べた後, 3 章で実験結果を示し, 考察を行う. また, 4 章では既存のリンク予測に用いられる関連研究を紹介する. 最後に, 5 章で結論を述べる.

2 提案手法

時系列グラフを活用する着目ノードに特化したリンク予測手法を提案する. 提案手法の概要図を図 1 に示す. 提案手法は, 入力として時刻 t までの時系列グラフと着目ノード, 学習の対象とするリンクのサンプリングに関するパラメータ h を入力し, 時刻 $t+1$ の着目ノードとその他の任意のノードとのリンクの存在を予測する. 提案手法は, ランダムウォークの始点とするノードを決定するステップとランダムウォークにより学習の対象とするリンクのサンプリングを行うステップ, サンプリングしたリンクに対し重み付き学習を行うステップの 3 つのステップからなる.

2.1 問題定義

はじめに, 提案手法が扱う問題を定義する. 時系列グラフは $\mathbb{G} = \{G^1, \dots, G^T\}$ で表され, T はタイムステップ数とする. また, $G^t = (V, E^t)$ はタイムステップ t のグラフを表す. $V = \{v_1, \dots, v_n\}$ はノードセットを表し, 既存手法の問題定義と同様にノードセットは時刻変化しないものとする [12]. $E^t \subseteq V \times V$ はタイムステップ t のリンクセットを表し, それぞれのリンクは $e_{i,j}^t = (i, j) \in E^t$ と表す. 着目ノードは v_f と表す. 本稿では, タイムステップ $1 \sim t$ までの時系列グラフ $\mathbb{G} = \{G^1, \dots, G^t\}$ とリンクを予測する特定の 1 ノードを入力として, タイムステップ $t+1$ の着目ノードとその他の任意のノードとのリンクの存在を予測する.

2.2 提案手法のアプローチ

提案手法では, 着目ノードのリンク予測に特化した学習を行うために, 着目ノード近傍のグラフ構造を重点的に学習するよ

うなリンクのサンプリングとロス関数の設計を行う. リンクのサンプリングでは, 着目ノードから h ホップ先までのノードを始点とするランダムウォークで得られるリンクを学習の対象としてサンプリングを行う. これにより, 着目ノードから h ホップまでのグラフ構造をより重視したモデルの学習が可能となる. さらに, 着目ノードから h ホップまでのノードをランダムウォークの始点としてリンクをサンプリングすることで, 着目ノードから h ホップまでのグラフ構造に限定することなく, グラフ全体の構造についても考慮した学習が可能である. さらに, サンプリングしてきたリンクの中でも着目ノードからの距離が近いリンクほど着目ノードのグラフ構造に与える影響が大きくなると考えられるため, 提案手法ではサンプリングしてきたリンクを着目ノードからの距離に応じた重み付けをして学習を行うロス関数を設計する. 提案手法では, 着目ノードのリンク予測において適切な重み付けを検証するために以下の 3 つの重み付けした binary cross entropy loss を導入する.

$$L_d = \sum_{t=1}^T \sum_{v \in V_{sub}^t} \left(- \sum_{u \in \Gamma^t(v)} \frac{\log(\hat{e}_{v,u})}{\min(d_v, d_u) + 1} - w_n \cdot \sum_{u' \in P^t(v)} \frac{\log(1 - \hat{e}_{v,u'})}{\min(d_v, d_u) + 1} \right) \quad (1)$$

$$L_{root} = \sum_{t=1}^T \sum_{v \in V_{sub}^t} \left(- \sum_{u \in \Gamma^t(v)} \frac{\log(\hat{e}_{v,u})}{\sqrt{\min(d_v^t, d_u^t) + 1}} - w_n \cdot \sum_{u' \in P^t(v)} \frac{\log(1 - \hat{e}_{v,u'})}{\sqrt{\min(d_v^t, d_u^t) + 1}} \right) \quad (2)$$

$$L_{nor} = \sum_{t=1}^T \sum_{v \in V_{sub}^t} \left(- \sum_{u \in \Gamma^t(v)} \frac{\log(\hat{e}_{v,u})}{\frac{\min(d_v^t, d_u^t)}{\max(\min(d_a^t, d_b^t))} + 1} - w_n \cdot \sum_{u' \in P^t(v)} \frac{\log(1 - \hat{e}_{v,u'})}{\frac{\min(d_v^t, d_u^t)}{\max(\min(d_a^t, d_b^t))} + 1} \right) \quad (3)$$

ここで, V_{sub}^t は着目ノードから h ホップのサブグラフに含まれるノードセットを表し, $\Gamma^t(v)$ はタイムステップ t におけるノード v からのランダムウォークでノード v と共起するノ

ドセットを返す関数である。 $\hat{e}_{i,j}$ はノード i とノード j 間のリンクの存在確率の出力結果を表している。また、 $P^t(v)$ はタイムステップ t においてノード v とリンクを持っていないノードのネガティブサンプリング分布を表し、 w_n はネガティブサンプリングレートを表している。 d_i はノード i の着目ノードからの距離を表す。 L_d は単純に着目ノードからの距離をそのまま重みとして扱うロス関数である。 L_{root} は着目ノードから離れたリンクのロスを過剰に減少させないようにするために、距離による重みに対して平方根をとったロス関数である。 L_{nor} は、単純に着目ノードからの距離をそのまま重みとして扱う場合には、距離によって過剰に重み付けしてしまう可能性があるため、全体的な距離による重みを和らげるためにサンプリングしてきた各リンクの着目ノードからの距離を正規化したものを重みとするロス関数である。以上のアプローチにより、着目ノードからの距離がより近いグラフ構造を重視したリンク予測モデルの学習を行う。

2.3 提案手法アルゴリズム

提案手法では、既存の時系列グラフのリンク予測手法で用いられているモデルアーキテクチャを利用して、着目ノード近傍をより重点的に学習するためのリンクのサンプリングとロス関数を用いてモデルの学習を行う。具体的なリンクのサンプリングアルゴリズムをアルゴリズム 1 に示す。はじめに提案手法では、入力のパラメータ h に基づいて着目ノードから h ホップ先までのノードを探索し、リンクをサンプリングしてくるためのランダムウォークの始点とする。(1–5行目) その後、各始点からのランダムウォークで共起したノードとのリンクをポジティブリンクとしてサンプリングする。(8–17行目) また、各始点のランダムウォークで共起しなかったノードの中からランダムでサンプリングしてきたものをネガティブリンクとしてサンプリングを行う。(18–25行目) そして、これらのサンプリングされたリンクに対して、既存の時系列グラフのリンク予測手法で用いられているモデルアーキテクチャを利用し、2.2節で紹介した3つのロス関数のいずれかを用いて学習を行う。

3 実験

本稿では、以下の二つの実験を行う。はじめに、提案手法の着目ノードの予測精度と既存手法の予測精度を比較する。これにより、着目ノード近傍のグラフ構造をより重視してモデルを学習する提案手法の有効性を確認する。次に、提案手法のサンプリングに用いるパラメータを変化させてリンク予測精度を測定する。これにより、提案手法のパラメータが着目ノードのリンク予測精度に与える影響を確認する。

3.1 精度評価

3.1.1 データセット

本実験で用いるデータセットについての説明を以下に示す。
Enron¹：ヨーロッパの研究機関における電子メールの送受信

Algorithm 1 リンクのサンプリングアルゴリズム

Require: G^t, v_f, h

Ensure: Γ^t, P^t

```

# Search start nodes for random walks
1:  $V_{sub}^t = \{\}$ 
2: for  $v = 1$  to  $n$  do
3:   if  $d_v^t \leq h$  then
4:      $V_{sub}^t = V_{sub}^t \cup v$ 
5:   end if
6: end for
7:  $\Gamma^t = [ ] * |V_{sub}^t|$ 
   # Random walk sampling
8:  $walks = [ ] * |V_{sub}^t|$ 
9: for  $counter = 1$  to  $num\_walks$  do
10:   $random\_shuffle(V_{sub}^t)$ 
11:  for  $v$  in  $V_{sub}^t$  do
12:     $walks[v] = walks[v] \cup random\_walk(start\_node = v)$ 
13:  end for
14: end for
15: for  $v$  in  $V_{sub}^t$  do
16:   $\Gamma^t[v] = [(v, random\_sample(walks))$  for  $i = 1$  to  $sample\_size]$ 
17: end for
   # Negative sampling
18: for  $v$  in  $V_{sub}^t$  do
19:   while  $len(P^t[v]) < sample\_size$  do
20:      $u = random\_sample(range(n))$ 
21:     if  $u$  not in  $walks[v]$  then
22:        $P^t[v] = P^t[v] \cup u$ 
23:     end if
24:   end while
25: end for
26: return  $\Gamma^t, P^t$ 

```

履歴を表したグラフである。本実験では、最も送受信履歴の多かった年の半年間の履歴を1ヶ月ごとに区切り、6タイムステップの時系列グラフを作成した。

3.1.2 実験設定

本実験では、最新のタイムステップを除いたグラフを訓練データとし、最新のグラフの着目ノードに対するリンクを予測する。着目ノードは訓練データの中で最も合計のエッジ数が大きいものを着目ノードに設定する。また、静的グラフのリンク予測手法の精度評価では、データセットの情報量を保つために各タイムステップで出現するノードとエッジを一つのグラフに集約したグラフを用いて実験を行う。評価指標にはリンク予測で一般的に用いられている Area Under the ROC Curve (AUC) を用いる。また、提案手法は既存手法で最も高精度に時系列グラフのリンク予測を行うことができる DySAT のモデルアーキテクチャを利用する。

3.1.3 比較手法

我々は提案手法のリンク予測精度を以下の5手法と比較する。

- baseline：各手法の評価のための基準として設定した手

¹ : <http://networkrepository.com/ia-enron-email-dynamic.php>

データセット	baseline	Jaccard	node2vec	SEAL	DySAT	提案手法 L_d	提案手法 L_{root}	提案手法 L_{nor}
Enron	0.616	0.838	0.778	0.701	0.882	0.926	0.922	0.914

表 1: 着目ノードのリンク予測結果. 提案手法 L_d , 提案手法 L_{root} , 提案手法 L_{nor} はそれぞれ式 1, 式 2, 式 3 をを用いた場合の提案手法のことを示している. また, 提案手法のパラメータ h は $h = 2$ に設定した場合の結果を示す.

データセット名	ノード数	エッジ数	タイムステップ数
Enron	43661	16490~44440	6

表 2: データセット. ノード数は全タイムステップで累計のノード数である. エッジ数は全タイムステップでのエッジ数の値域を表している.

法である. 予測対象となるタイムステップのグラフ構造は直近のタイムステップのグラフ構造と類似している可能性が高いという観点から, 直近のタイムステップの着目ノードのリンクを予測結果とする手法である.

- jaccard 係数: 2 ノード間で共通する隣接ノードの割合が多いほどそのノード間でリンクを持ちやすいだろうという考えに基づき, 2 ノード間の共通隣接ノードの割合をリンクの存在確率のスコアとして扱うリンク予測手法である.

- node2vec: 深層学習を用いた静的グラフのグラフ埋め込み手法である. ランダムウォークによってグラフの隣接情報をサンプリングし, このランダムウォークを訓練データとしてノードの表現ベクトルを取得する.

- SEAL: 深層学習を用いた静的グラフのリンク予測手法である. 任意の 2 ノードから h ホップのサブグラフを取得することでグラフのサンプリングを行い, そのサブグラフから 2 ノード間にリンクが存在するかを予測するグラフニューラルネットワークモデルを学習する.

- DySAT: 深層学習を用いた時系列グラフのリンク予測手法である. 構造と時系列の二次元で自己注意ネットワークを構成することで, 構造的特徴と時系列変化の両方を考慮したリンク予測が可能である. この手法は, 既存の時系列グラフを活用したリンク予測モデルの中でも最も高精度なリンク予測が可能である.

3.1.4 実験結果

本実験の結果を表 1 に示す. 表 1 から, 3 つの提案手法全てが既存手法と比較して高精度な予測を達成していることが確認できる. このことから, 着目ノードのリンク予測において着目ノード近傍のグラフ構造を重視してモデルを学習する提案手法の有効性が確認できる. また, 3 つの提案手法の中ではロス関数 L_d を用いた手法が最も高精度な予測結果となっている. L_{root} や L_{nor} では着目ノードからの距離による重みに対して平方根をとる操作や正規化を行うことにより, 距離による重み付けを和らげる処理になっているため, 十分に着目ノード近傍のグラフ構造を重視して学習できておらず, L_d を用いた提案手法と比較して精度が低下しているものと考えられる. また, 時系列グラフのリンク予測手法である DySAT は他の静的グラフのリンク予測手法に比べて高精度な予測結果となっており, 時系列グラフを活用することの有効性が窺える.

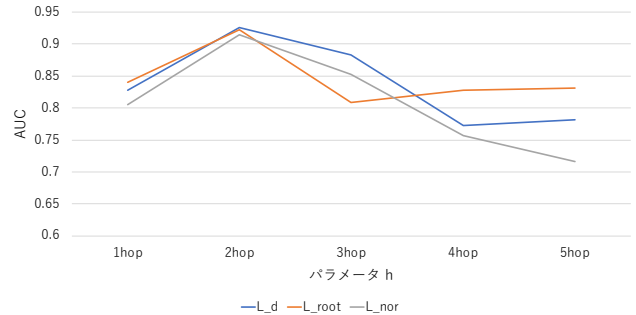


図 2: パラメータ h による着目ノードのリンク予測精度の変化. L_d , L_{root} , L_{nor} のロス関数を用いてそれぞれで $h = 1, 2, 3, 4, 5$ に設定した場合のリンク予測精度を図に示している.

3.2 パラメータセンシティブティ

提案手法では, ランダムウォークの始点とするノードとして着目ノードからの距離が h ホップ以内であるノードを選択する. 本節では, このパラメータ h を変更することにより提案手法の着目ノードのリンク予測精度に与える影響を確認する. L_d , L_{root} , L_{nor} のロス関数を用いてそれぞれで $h = 1, 2, 3, 4, 5$ に設定した場合のリンク予測精度を図 2 に示す. 図 2 からどのロス関数を用いた場合にも $h = 2$ とした場合に最も高い精度となっていることが確認できる. このことから, 着目ノードから 1 ホップのノードを始点とした場合には十分に近傍のグラフ構造をサンプリングすることができず, 精度が向上していないと考えられる. また, 着目ノードから 3 ホップ以上のノードをランダムウォークに始点とした場合には着目ノードから離れたリンクを多くサンプリングしてしまうため, 着目ノードの近傍のグラフ構造を十分に重視した学習を行うことができず精度が低下したと考えられる.

4 関連研究

リンク予測では既知または観測されたリンクに基づいて, 将来または未観測のリンクの存在を予測する. リンク予想に利用可能な手法は数多く提案されており, これらの手法は, (1) 静的なグラフのみを用いるリンク予測手法と (2) 時系列グラフを活用したリンク予測手法の二つのアプローチに分類することができる.

4.1 静的グラフのリンク予測

初期の静的グラフのリンク予測では common neighbor [16] や jaccard 係数 [14], Adamic adar [15], PageRank [17], SimRank [18] など, 既知のグラフ構造からノード間の類似度を示すような指標を計算し, それに基づいてリンクの存在を判断するヒューリスティックな手法が用いられていた. これらの指

Name	Formula	Order
common neighbors [16]	$ \Gamma(x) \cap \Gamma(y) $	first
Jaccard [14]	$\frac{ \Gamma(x) \cap \Gamma(y) }{ \Gamma(x) \cup \Gamma(y) }$	first
Adamic-Adar [15]	$\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log \Gamma(z) }$	second
PageRank [17]	$[\pi_x]_y + [\pi_y]_x$	high
SimRank [18]	$\gamma(x) \frac{\sum_{a \in \Gamma(x)} \sum_{b \in \Gamma(y)} \text{score}(a,b)}{ \Gamma(x) \cdot \Gamma(y) }$	high

表 3: ノード x, y 間の類似度を示す指標. $\Gamma(x)$ はノード x の近傍ノードセットを表し, $[\pi_x]_y$ はノード x からランダムウォークを行った際の y の定常分布確率を表す. SimRank のスコアは表中の式を再起的に計算することにより算出される. Order は各指標がノード x, y から何ホップ先までの情報を考慮しているかを表している. また, $\text{score}(a, b)$ はノード a, b 間の SimRank の値を表している. Order はそれぞれの指標が何ホップ先までの隣接情報を考慮しているかを示している.

標の詳細を表 3 に示す. common neighbor [16] や jaccard 係数 [14] は 2 ノード間の共通隣接ノードに着目した類似度指標である. 共通の隣接ノードが多いほどそのノード間でリンクを持ちやすいという考えに基づき, common neighbor では 2 ノード間の共通隣接ノードの数を, jaccard 係数では共通隣接ノードの割合をリンクの存在確率のスコアとして用いる手法である. また, Adamic-Adar [15] は上記の 2 つの共通隣接指標を改良した類似度指標である. 共通隣接ノードの中でもその他のノードとあまり隣接していないノードを重く評価するよう重み付けされた指標である. さらに, PageRank と SimRank はグラフ全体の構造を考慮した類似度指標である. PageRank [17] はグラフ中のノードをランダムに遷移するランダムウォークを行った場合に, ランダムウォークがあるノードに定常的に訪れる確率を示す指標である. ノード x からランダムウォークを行った場合に, あるノード y に訪れる確率が高いほどノード x と y は類似していることを示す. SimRank [18] は類似度の高いノードの隣接ノードは類似したノードであるという考えに基づいた類似度指標である. あるノード x, y の隣接ノード間の類似度の高いほどノード x, y の類似度も高くなるように類似度が再帰的に計算される.

しかし近年では, 深層学習を用いたグラフ埋め込み手法が数多く提案されており, これらの手法で得られたグラフ埋め込みがリンク予測に利用される [5, 9, 10]. node2vec [9] はランダムウォークによってグラフの隣接情報をサンプリングし, このランダムウォークを訓練データとしてノードの表現ベクトルを取得する. グラフ畳み込みネットワーク [5] は, 画像処理分野で広く用いられている畳み込みネットワークをグラフに拡張した手法である. 隣接ノードの属性を伝搬させることで構造的に類似するノードが類似する表現ベクトルを持つようにグラフ埋め込みを学習する. VGAE [10] はエンコーダにグラフ畳み込みネットワークを用いており, デコーダでは内積を用いて下のグラフを復元するような学習を行う. これらの手法はヒューリスティックなリンク予測手法に比べて高精度なリンク予測が可能である. これらの手法に加えて End-to-End でリンク予測を行う手法も提案されている [6, 19]. これらの手法では, 任意の 2 ノードから h ホップのサブグラフを抽出し, このサブグラフの

構造的特徴と属性的特徴からリンクの存在を二値分類問題として解く手法である. これらの手法は, グラフ埋め込み手法を用いたリンク予測と比較してもより高精度な予測が可能であることが知られている.

4.2 時系列グラフのリンク予測

グラフの時系列変化を学習することによる精度向上を目的として, 時系列グラフを活用したリンク予測手法の研究が進んでいる [11–13]. これらの手法では, 時系列グラフを活用することで時系列的なグラフ構造の変化を捉えた上でリンク予測を行うことが可能である. そのため, 静的グラフのみを用いたリンク予測手法と比較して高精度な予測を行うことが可能である. Dyn2AERNN [11] では, 時系列グラフの隣接情報を LSTM で学習し, 次の時刻のグラフを再現する学習を行う. これにより, 各ノードの時系列変化の傾向を捉えた上でリンクを予測している. また DySAT [12] では, 構造的な自己注意ネットワークと時間的な自己注意ネットワークを二次元的に組み合わせることにより, 各時刻の構造的特徴と時系列的な構造変化の特徴の両方を考慮してリンク予測を行っている. Evolve GCN [13] はグラフ畳み込みネットワークを時系列グラフへ拡張した手法である. この手法はグラフ埋め込み手法であり, このグラフ埋め込みを利用してリンク予測が可能である. しかし, これらの既存手法は全てグラフ中の全リンクを予測することを目的とした手法である. そのため, グラフ中のリンクを均等に学習しており, 着目ノードのリンク予測に適用する場合には十分に精度が期待できない. したがって, 着目ノードのリンク予測においては, 着目ノードのリンク予測に特化したモデルの学習を行う必要がある.

5 終わりに

本稿では, 着目ノードに特化したリンク予測手法を提案した. 提案手法では, 学習で利用できる情報量を増やすために, 時系列グラフを活用した. また, 着目ノードの構造変化が近傍のグラフ構造から受ける影響を適切に学習するために, 着目ノードのリンクとその近傍のリンクを距離に応じて重み付けしたロス関数を設計し, 既存のモデルアーキテクチャに適用した. これにより, 提案手法が着目ノードのリンク予測において最も高精度な予測が可能であることを確認した. 今後の課題としては, 異なる着目ノードを設定した場合のリンク予測精度の測定や複数のデータセットでの実験を行うことにより, 提案手法の汎用性を確かめる必要がある. また, リンク予測対象を着目ノードに限定することを活かして効率的に学習を行うことのできるモデルアーキテクチャの設計が挙げられる.

謝 辞

本研究は JSPS 科研費 JP20H00583 の助成を受けたものです.

文 献

- [1] Santo Fortunato. Community detection in graphs. *The Physics Reports*, Vol. 486, No. 3, pp. 75 – 174, 2010.
- [2] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *The TKDD*, Vol. 1, No. 1, p. 2-es, 2007.
- [3] Jia Li, Zhichao Han, Hong Cheng, Jiao Su, Pengyun Wang, Jianfeng Zhang, and Lujia Pan. Predicting path failure in time-evolving graphs. In *Proceedings of the SIGKDD*, p. 1279–1289, 2019.
- [4] Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, and Thomas A. J. Schweiger. Scan: A structural clustering algorithm for networks. In *Proceedings of the SIGKDD*, p. 824–833, 2007.
- [5] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the ICLR*, 2017.
- [6] Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. In *Proceedings of the NeurIPS*, pp. 5165–5175, 2018.
- [7] Xin Sun, Zenghui Song, Junyu Dong, Yongbo Yu, Claudia Plant, and Christian Böhm. Network structure and transfer behaviors embedding via deep prediction model. pp. 5041–5048, 2019.
- [8] Ashish Kumar, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan. From social to sale: The effects of firm-generated content in social media on customer behavior. *The Journal of Marketing*, Vol. 80, No. 1, pp. 7–25, 2016.
- [9] Aditya Grover and Jure Leskovec. Node2vec: Scalable feature learning for networks. In *Proceedings of the SIGKDD*, p. 855–864, 2016.
- [10] Thomas N. Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
- [11] Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *The Knowledge-Based Systems*, Vol. 187, p. 104816, 2020.
- [12] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the WSDM*, p. 519–527, 2020.
- [13] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B. Schardl, and Charles E. Leiserson. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI*, 2020.
- [14] Jaccard Paul. The distribution of the flora in the alpine zone. *The NEW PHYTOLOGIST*, Vol. 11, pp. 37–50, 1912.
- [15] Adamic Lada, A and Adar Eytan. Friends and neighbors on the web. *The Social Networks*, Vol. 25, No. 3, pp. 211 – 230, 2003.
- [16] M. E. J. Newman. Clustering and preferential attachment in growing networks. *The Physical Review E*, Vol. 64, No. 2, p. 025102, 2001.
- [17] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *The Computer Networks*, Vol. 30, No. 1, pp. 107–117, 1998.
- [18] Glen Jeh and Jennifer Widom. Simrank: A measure of structural-context similarity. In *Proceedings of the SIGKDD*, p. 538–543. Association for Computing Machinery, 2002.
- [19] Zheyi Liu, Darong Lai, Chuanyou Li, and Meng Wang. Feature fusion based subgraph classification for link prediction. In *Proceedings of the CIKM*, p. 985–994, 2020.