

# 異種ネットワーク間リンクの ペアワイズ類似度に基づくリンク予測法

Link prediction over heterogeneous networks using link pairwise similarity

長谷川 聡\*<sup>1</sup>      佐久間 淳\*<sup>2\*3</sup>  
Satoshi Hasegawa      Jun Sakuma

\*<sup>1</sup>筑波大学情報学群情報科学類      \*<sup>2</sup>筑波大学システム情報系  
College of Information Science, University of Tsukuba.      Faculty of Engineering, Information and Systems

\*<sup>3</sup>科学技術振興機構さきがけ  
Japan Science and Technology Agency

There is a link as a link between nodes belonging to the different network other than a link between nodes in the network, in the network data structure. Such as a link, there is the ties of between accounts belonging to different social network. Link prediction of a link as between networks is important task. The link prediction method of the link between the networks is proposed to use pair-wise similarity. Existing pair-wise similarity has been pointed out scalability and prediction accuracy. In this paper, we present a new pair-wise similarity using the link between the networks, scalable and good prediction accuracy.

## 1. はじめに

ネットワーク構造を持ったデータには、個々のネットワーク内のノード同士の関係だけでなく、ネットワークをまたいだノード同士の関係が存在する。異なるソーシャルネットワークのアカウント同士の結びつきなどがそれにあたる。これは実世界のある人間が作成した異なるソーシャルネットワークのアカウント同士が、その人間を通して結びつく関係である。このような関係を用いると、異なるソーシャルネットワークを介した友達推薦などが可能となる。他にも、人物ネットワークと商品のネットワークは、ある人物が商品を購入することにより結びつきが発生する。友達の購入した商品を推薦したりなど、様々な推薦を行えることから、異なるネットワークのノード同士の関係の予測は、実世界において重要な課題であると考えられる。本稿では、このようなネットワークをまたいだノード間のエッジの有無の予測を取り扱う。

ノード間のエッジの有無の予測を行う手法の1つとして、ノードペア同士の類似度(ペアワイズ類似度)を用いた学習法が提案されている [1][2]。ペアワイズ類似度を用いた学習法として、ペアワイズカーネル法や、リンク伝播法が提案されている。共通して適切なペアワイズ類似度を定義することによりネットワークデータに限らず、2項関係を持つデータに適用することができる。その反面、ノードペア同士の類似度を定義するため、ノード同士の類似度と比べ、記憶容量が必要となり、スケーラビリティの課題が生じる。

本稿では、ネットワークをまたいだノード間のエッジの有無の予測をペアワイズ類似度を用いて取り扱う。ペアワイズ類似度を求める方法は、これまでに2手法が知られている。1つはすべてのノードペアに対し類似度を定義し、もう1つは限られたノードペアに対し類似度を定義する方法である。前者は密な類似度であることから、空間計算量が大きく、後者は、含まれる情報量が少なく疎な類似度であり、予測精度の面で不安が残る。この点から、記憶容量についてスケーラブルかつ予測精度が良い類似度を作ることが課題となる。そこで本稿では、既存

のペアワイズ類似度では用いられていない重要な情報であるネットワーク間エッジを新たに考慮し、ペアワイズ類似度の課題を克服する、スケーラブルかつ予測精度が良いペアワイズ類似度を提案する。

## 2. ネットワーク間リンク予測

ネットワークをまたいだノード間のエッジの有無を、ネットワーク間リンク予測と呼ぶことにする。まず一般的なネットワーク間リンク予測問題を定義し、その後応用事例として、脱匿名化問題について示す。

### 2.1 問題定義

ネットワーク間リンク予測問題を定式化する。2ネットワーク  $G_1 = (V_1, E_1)$ ,  $G_2 = (V_2, E_2)$  を考える。ネットワーク  $G_1$  において、ノードを  $v_1^{(i)} \in V_1$ ,  $v_1^{(i)}$  と  $v_1^{(j)}$  の間のエッジを  $e_{12}^{(ij)}$  と表記する(ネットワーク  $G_2$  についても同様に表記する)。  $G_1$ ,  $G_2$  のノードペア  $v_1^{(i)}, v_2^{(j)}$  間のエッジを  $e_{12}^{(ij)} \in E_{12}$  とする ( $E_{12} \subseteq V_1 \times V_2$  である)。さらに、ノード集合  $V_1, V_2$  の大きさはそれぞれ  $|V_1| = N$ ,  $|V_2| = M$  とする。ネットワーク間のノードペアの関係、すなわちノード集合  $V_1$  とノード集合  $V_2$  の関係を  $N \times M$  の行列  $\mathbf{F}$  で表す。行列  $\mathbf{F}$  の要素を以下で与える。

$$\mathbf{F}_{ij} \equiv \begin{cases} 1 & e_{12}^{(ij)} \in E_{12} \\ 0 & e_{12}^{(ij)} \notin E_{12} \\ 0.5 & e_{12}^{(ij)} \text{の存在が未知} \end{cases}$$

また既知の  $G_1, G_2$  間のエッジの集合を  $\mathcal{O}_{12} = \{e_{12}^{(ij)} | \mathbf{F}_{ij} \neq 0.5\}$  とする。ネットワーク間リンク予測問題とは、既知であるそれぞれのネットワークのエッジ集合  $E_1, E_2$  およびネットワーク間のエッジ集合  $\mathcal{O}_{12}$  を用い、存在が未知なネットワーク間のエッジ  $e_{12}^{(ij)} \notin \mathcal{O}_{12}$  を予測する問題である。

### 2.2 脱匿名化

脱匿名化問題は、近年急成長を遂げたソーシャルネットワークサービス(以下 SNS)におけるプライバシーの問題である。SNS は大きく2種類に分類することができる。1つは実

連絡先: 長谷川 聡, 筑波大学情報学群情報科学類,  
hasegawa(at)mdl.cs.tsukuba.ac.jp

名登録アカウントの多いもの、もう1つは匿名登録アカウントの多いものである。前者としては facebook、後者としては twitter などが有名である。インターネット利用者の多くがこれらの SNS を利用しており、実名アカウントと匿名のアカウント、両方のアカウントを持つユーザも多い。この SNS において、利用者のプライバシー情報の漏洩が問題となりつつある。

利用者のプライバシーは基本的には自身が設定するプライバシーポリシーで制御可能であるが、SNS が表示する他人とのリンク構造や発言などの解析により、本来非開示のユーザの個人情報が漏えいする可能性がある。複数 SNS 間の脱匿名 (De-anonymization) とは、異なる SNS 上の匿名のアカウントと、実名のアカウントを実世界の人物を介して結びつけることを示す [3]。

匿名アカウントの SNS を  $G_1 = (V_1, E_1)$ 、実名アカウントの SNS を  $G_2 = (V_2, E_2)$  とする。  $V_1, V_2$  は各 SNS のアカウントの集合を表し、  $E_1, E_2$  は各 SNS でのアカウントの関係 (例えば友達関係) を表すものとする。  $G_1$  のアカウント  $v_1^{(i)} \in V_1$ 、  $G_2$  のアカウント  $v_2^{(j)} \in V_2$  が同一人物のアカウントの場合、  $v_1^{(i)}$  と  $v_2^{(j)}$  が結びつくとし、エッジ  $e_{12}^{(ij)}$  が存在すると定義する。脱匿名化問題は、未だエッジの存在が不明な異なる SNS のアカウント同士のペアの間のエッジを予測する問題と捉えることができ、ネットワーク間リンク予測の枠組みであるといえる。

### 3. edge conditional ペアワイズ類似度の提案

ペアワイズ類似度を用いたネットワーク間リンク予測において、ペアワイズ類似度の設計は予測精度に大きく影響する。本稿では、ネットワーク間に張られるエッジを考慮したペアワイズ類似度を提案する。またそれらが既存のペアワイズ類似度とどのように異なるかを比較する。

#### 3.1 既存のペアワイズ類似度

既存のペアワイズ類似度として、クロネッカー積類似度 [1]、クロネッカー和類似度 [2] が提案されている。ネットワーク  $G_1$  のノード集合  $V_1$  のノード同士の類似度を類似度行列  $\mathbf{W}_1$ 、ネットワーク  $G_2$  のノード集合  $V_2$  のノード同士の類似度を類似度行列  $\mathbf{W}_2$  と表し、ペアワイズ類似度の定式化を行う。

##### 3.1.1 クロネッカー積類似度

クロネッカー積類似度 [1] は、任意の2つのノードペアに対し、それぞれの対応する要素がすべて類似している場合、ノードペア同士が類似していると定義する。

$$\mathbf{W}_{i,j,l,m} \equiv \mathbf{W}_{1i,l} \mathbf{W}_{2j,m}$$

ここで、  $\mathbf{W}_{1i,l}$  は、類似度行列  $\mathbf{W}_1$  の  $v_1^{(i)}$ 、  $v_1^{(l)}$  の要素を表す ( $\mathbf{W}_{2j,m}$  も同様)。ノードペア  $v_1^{(i)}$ 、  $v_1^{(l)}$  が似ておりかつノードペア  $v_2^{(j)}$ 、  $v_2^{(m)}$  が似ているとき、ノードペア  $v_1^{(i)}$ 、  $v_2^{(j)}$  と  $v_1^{(l)}$ 、  $v_2^{(m)}$  が似ているとするものである。これは以下のクロネッカー積で書き表される。

$$\mathbf{W} \equiv \mathbf{W}_2 \otimes \mathbf{W}_1 \quad (1)$$

##### 3.1.2 クロネッカー和類似度

クロネッカー積は、すべてのノードペアの組み合わせに対し類似度を定義することから密な類似度行列を生成し、多くの記憶容量を必要とする。そこで Kashima ら [2] は、2つのノードペアの片方が同一で、もう片方が類似している場合、ノードペ

ア同士が類似していることを想定した類似度を提案した。

$$\mathbf{W}_{i,j,l,m} = \mathbf{W}_{1i,l} \delta(j=m) + \delta(i=l) \mathbf{W}_{2j,m}$$

ここで  $\delta$  は  $\delta$  内の式が成立していれば1、していなければ0を返す関数である。片方のノードが同一で  $v_2^{(j)}$ 、  $v_2^{(m)}$  が似ているとそのノードペア同士は似ているとする。これは以下のように2つの類似度行列のクロネッカー和で表される。

$$\begin{aligned} \mathbf{W} &\equiv \mathbf{W}_2 \oplus \mathbf{W}_1 \\ &= \mathbf{W}_2 \otimes \mathbf{I}_1 + \mathbf{I}_2 \otimes \mathbf{W}_1 \end{aligned} \quad (2)$$

ここで、  $\mathbf{I}_1$ 、  $\mathbf{I}_2$  はそれぞれ、  $N \times N$ 、  $M \times M$  の単位行列である。

#### 3.2 提案ペアワイズ類似度

クロネッカー積は、すべてのノードペア同士に対し類似度を定義するため、密なペアワイズ類似度行列を与え、計算負荷が高い。クロネッカー和類似度は、ペアワイズ類似度行列は疎であるため記憶容量の消費は少ないが、多くのノードペア同士は類似度が与えられず、予測にとって情報が不足するケースがある。またどちらのペアワイズ類似度も、重要な情報であるネットワーク間のエッジを考慮していない。

そこで、重要な要素である既知のネットワーク間のエッジを用いることにより、与えられた情報がある程度保持しつつ、スパース性を持つ新たなペアワイズ類似度、edge conditional ペアワイズ類似度を提案する。

この edge conditional ペアワイズ類似度は、クロネッカー和類似度に加え、ネットワーク間のエッジがある場合、それぞれのネットワークのペア同士が類似しているなら、ノードペア同士が似ていると定義する。

$$\begin{aligned} \mathbf{W}_{i,j,l,m} &\equiv \mathbf{W}_{1i,l} \delta(j=m) + \delta(i=l) \mathbf{W}_{2j,m} \\ &+ \mathbf{W}_{1i,l} \tau(v_1^{(i)}, v_2^{(j)}) + \mathbf{W}_{2j,m} \tau(v_1^{(i)}, v_2^{(j)}) \\ &+ \tau(v_1^{(l)}, v_2^{(m)}) \mathbf{W}_{2j,m} + \tau(v_1^{(l)}, v_2^{(m)}) \mathbf{W}_{1i,l} \end{aligned}$$

ここで  $\tau(v_1^{(i)}, v_2^{(j)})$  は、  $v_1^{(i)}$  と  $v_2^{(j)}$  間にエッジが存在する場合1を、存在しない場合0を返す関数である。既に  $v_1^{(i)}$ 、  $v_2^{(j)}$  にエッジがあるとき、  $v_1^{(i)}$ 、  $v_1^{(l)}$  が似ているもしくは、  $v_2^{(j)}$ 、  $v_2^{(m)}$  が似ていれば、ノードペアが似ているとするものである。

これは先ほど述べたとおり、クロネッカー和に、既に結びついているエッジも考慮したものとなっている。これを行列の形で表すと、

$$\begin{aligned} \mathbf{W} &\equiv \mathbf{W}_2 \otimes \mathbf{I}_1 + \mathbf{I}_2 \otimes \mathbf{W}_1 \\ &+ \frac{1}{|\mathbf{E}_{12}|} \sum_{e_{12}^{(ij)} \in \mathbf{E}_{12}} \{(\mathbf{W}_2 * \mathbf{T}_2^{(j)}) \otimes \mathbf{T}_1^{(i)} \\ &+ \mathbf{T}_2^{(j)} \otimes (\mathbf{T}_1^{(i)} * \mathbf{W}_1)\} \end{aligned} \quad (3)$$

となる。ただし、  $\mathbf{T}_1^{(i)}$ 、  $\mathbf{T}_2^{(j)}$  はそれぞれ  $N \times N$ 、  $M \times M$  の正行列であり、以下を満たす行列の要素を持つものである。

$$\begin{aligned} \mathbf{T}_1^{(i)}_{l,m} &\equiv \begin{cases} \frac{1}{\sqrt{|\mathbf{E}_{12}|}} & l=m \text{ を除く, } l=i \text{ か } m=i \text{ の場合} \\ 0 & \text{otherwise} \end{cases} \\ \mathbf{T}_2^{(j)}_{l,m} &\equiv \begin{cases} \frac{1}{\sqrt{|\mathbf{E}_{12}|}} & l=m \text{ を除く, } l=j \text{ か } m=j \text{ の場合} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

### 3.3 各ペアワイズ類似度の空間計算量について

クロネッカー積類似度は、すべてのノードペア同士に対し類似度を与えるため、空間計算量が  $O(N^2M^2)$  必要である。それに対しクロネッカー和類似度は、片方のノードペアが同一の場合のときのみ類似度を定義するため、空間計算量が  $O(MN^2 + NM^2)$  である。提案した edge conditional ペアワイズ類似度は、既知のネットワーク間エッジの数を  $L$  とすると空間計算量は  $O(NM^2 + MN^2 + LNM)$  となる。また  $L \ll N, M$  において、空間計算量は  $O(NM^2 + MN^2)$  となり、クロネッカー和類似度と同等になる。  $L$  の数に依存してスパース性が変化することから、空間計算量を考慮して類似度を用いることができる利点がある。

## 4. 提案ペアワイズ類似度のリンク伝播法への適用

ペアワイズ類似度を用いた学習法の1つとして、半教師付き学習法のリンク伝播法 [2] がある。「2組のノードペアが互いに似ているならば、それらノードペアのエッジの存在の有無も似ている」という仮説を正則化項としてモデル化したもので、リンク伝播法の目的関数は、

$$J(\hat{\mathbf{F}}) = \frac{\mu}{2} \text{vec}(\hat{\mathbf{F}})^T \mathbf{L} \text{vec}(\hat{\mathbf{F}}) + \frac{1}{2} \|\text{vec}(\hat{\mathbf{F}} * \mathcal{G}) - \text{vec}(\mathbf{F})\|_2^2$$

となる。 $\hat{\mathbf{F}}_{ij}$  はエッジの存在の有無の確からしさを表す。 $\mathbf{L}$  はグラフラプラシアン  $\mathbf{L} \equiv \mathbf{D} - \mathbf{W}$  である ( $\mathbf{D}$  は対角行列で、対角要素が  $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$  を満たす要素を持つ行列)。また  $\text{vec}$  は行列の列ベクトルを積み上げて構成するベクトルを表す。 $*$  は hadamard 積 (行列の要素ごとの掛け算) を表す。 $\mathcal{G}$  は、要素  $\mathcal{G}_{ij}$  が、 $e_{12}^{(ij)} \in \mathcal{O}_{12}$  なら 1, それ以外なら 0 の、 $N \times M$  の行列である。この目的関数を  $\text{vec}(\mathbf{F})$  で微分し最小化を行い、極値を求めると

$$(\mu \mathbf{L} + \text{diag}(\text{vec}(\mathcal{G}))) \text{vec}(\hat{\mathbf{F}}) = \text{vec}(\mathbf{F}) \quad (4)$$

となり、この式を解くことにより解が求まる。

$\mathbf{L} \text{vec}(\hat{\mathbf{F}})$  は巨大な行列とベクトルとの積になるため、計算コストがかかる。そこで edge conditional ペアワイズ類似度などのクロネッカー積演算を含むペアワイズ類似度を、リンク伝播法へ適用する際、vec トリックを用いることにより計算コストを削減する。vec トリック [4] とは、クロネッカー積とベクトルとの積を、単なる行列の積に置き換えるものである。これにより大きな記憶容量を必要とするクロネッカー積を含む計算を、小さな行列の積に置き換える。edge conditional ペアワイズ類似度の  $\mathbf{L} \text{vec}(\hat{\mathbf{F}})$  は以下のように表される。

$$\begin{aligned} \mathcal{L}^{\text{PROP}}(\hat{\mathbf{F}}) &\equiv \mathbf{L} \text{vec}(\hat{\mathbf{F}}) \\ &= \left( \text{vec}(\mathbf{I}_1 \hat{\mathbf{F}} \mathbf{D}_2) + \text{vec}(\mathbf{D}_1 \hat{\mathbf{F}} \mathbf{I}_2) \right) \\ &+ \frac{1}{|\mathbf{E}_{12}|} \sum_{e_{12}^{(ij)} \in \mathbf{E}_{12}} \{ \text{vec}(\mathbf{T}_1^{(i)} \hat{\mathbf{F}} (\mathbf{D}_2 * \mathbf{T}_2^{(j)})) \} \\ &+ \text{vec}((\mathbf{D}_1 * \mathbf{T}_1^{(i)}) \hat{\mathbf{F}} \mathbf{T}_2^{(j)}) \\ &- \left( \text{vec}(\mathbf{I}_1 \hat{\mathbf{F}} \mathbf{W}_2) + \text{vec}(\mathbf{W}_1 \hat{\mathbf{F}} \mathbf{I}_2) \right) \\ &+ \frac{1}{|\mathbf{E}_{12}|} \sum_{e_{12}^{(ij)} \in \mathbf{E}_{12}} \{ \text{vec}(\mathbf{T}_1^{(i)} \hat{\mathbf{F}} (\mathbf{W}_2 * \mathbf{T}_2^{(j)})) \} \\ &+ \text{vec}((\mathbf{W}_1 * \mathbf{T}_1^{(i)}) \hat{\mathbf{F}} \mathbf{T}_2^{(j)}) \end{aligned}$$

edge conditional ペアワイズ類似度をリンク伝播法に適用した予測アルゴリズムを Algorithm 1 に示す。

**Algorithm 1** edge conditional ペアワイズ類似度を用いたリンク伝播法アルゴリズム

**Input:**  $\mathbf{F}, \mathcal{G}, \mathbf{W}_1, \mathbf{W}_2, \sigma, e$

**Output:**  $\hat{\mathbf{F}}$

```

1:  $\mathbf{F}(0) = \mathbf{F}$ ,  $\text{vec}(\mathbf{R}(0)) = -\mu \mathcal{L}^{\text{PROP}}(\mathbf{F}(0))$ ,  $\mathbf{P}(0) = \mathbf{R}(0)$ 
2: for  $t = 0, 1, 2, \dots$  do
3:    $\mathbf{Q}(t) = \mu \mathcal{L}^{\text{PROPOSAL}}(\mathbf{P}(t)) + \mathcal{G} * \mathbf{P}(t)$ 
4:    $\alpha(t) = \frac{\text{vec}(\mathbf{R}(t)) \cdot \text{vec}(\mathbf{P}(t))}{\text{vec}(\mathbf{P}(t)) \cdot \text{vec}(\mathbf{Q}(t))}$ 
5:    $\mathbf{F}(t+1) = \mathbf{F}(t) + \alpha(t) \mathbf{P}(t)$ 
6:    $\mathbf{R}(t+1) = \mathbf{R}(t) - \alpha(t) \mathbf{Q}(t)$ 
7:    $\beta(t) = \frac{\|\mathbf{R}(t+1)\|_2^2}{\|\mathbf{R}(t)\|_2^2}$ 
8:   if  $\frac{\|\mathbf{R}(t+1)\|_2}{\|\mathbf{R}(0)\|_2} < e$  then
9:     return  $\mathbf{F}(t+1)$ 
10:  end if
11:   $\mathbf{P}(t+1) = \mathbf{R}(t+1) + \beta(t) \mathbf{P}(t)$ 
12: end for
    
```

## 5. 実験

既存のペアワイズ類似度と提案ペアワイズ類似度の予測精度を比較するため、ネットワーク間リンク予測の実験を行う。既存のペアワイズ類似度のリンク伝播法のアルゴリズムは [2] を用い、提案ペアワイズ類似度は、Algorithm 1 を用いた。実験データは、人工的に作成したデータおよび、facebook と twitter から収集したデータを用いる。前者は、単一ネットワークデータを 2 ネットワークに分け、ネットワーク間のエッジを付与するものである。ネットワークデータとして、ロビラ・イビルヒリ大学内の学生の e-mail のやり取りを表したノード数 1,133, エッジ数 10,903 を持つネットワークデータを用いた (以下、URV データセットと呼ぶ) [5]。後者は、クローリングにより収集したデータである。facebook がノード数 802, エッジ数 10466, twitter がノード数 3132, エッジ数 322394 の規模のデータであり、そのうち 82 ノードが facebook と twitter の結びつきを持つ。このデータを fbtw データセットと呼ぶ。

類似度行列  $\mathbf{W}_1, \mathbf{W}_2$  は、それぞれのネットワークの拡散カーネル [6] を用いる。また各実験共に 25 回平均の Top- $k$  Precision 及び それにもとづく AUC を算出し、予測精度として用いる。Top- $k$  Precision は、予測値を降順に並べた際、テストデータが上位  $k$  番目までに含まれている割合を表す。

### 人工的なデータを用いた実験

URV データセットをノードの重なり割合 0.5, エッジの重なり割合 0.5 として 2 つのネットワークに分割したデータに対し、予測精度の評価実験を行う。データは、以下の手順に従って作成した。

1.  $G = (V, E)$  を用意する。
2. ノード集合  $V$  からそれぞれ部分集合  $V_A, V_B, V_C$  をランダムに作成する。ここで  $V_A, V_B, V_C$  はそれぞれサイズが  $0.25|V|, 0.5|V|, 0.25|V|$  となるよう、 $V$  からランダム非復元抽出する。
3.  $V_A, V_B, V_C$  から集合  $V_1 = V_A \cup V_B, V_2 = V_B \cup V_C$  を得る。これらを 2 ネットワークのノードとする。

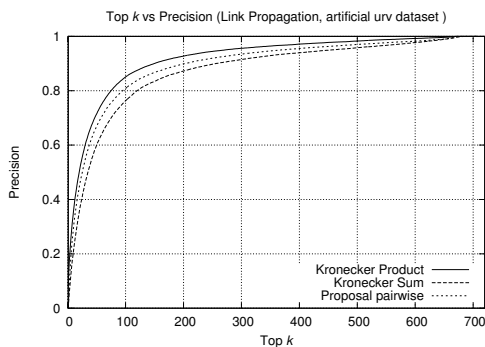


図 1: 人工 URV データセットに対する, リンク伝播法の予測精度. Precision が 1 になるまで Top  $k$  を変化させた結果.

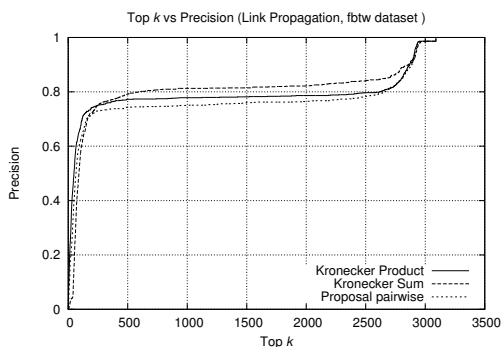


図 2: fbtw データセットに対する, リンク伝播法の予測精度. Precision が 1 になるまで Top  $k$  を変化させた結果.

4. グラフ  $G$  のノード集合  $V_1$  による誘導部分グラフ  $G_1 = (V_1, E_1)$ , を得る ( $G_2$  も同様).
5. グラフ  $G_1, G_2$  のノード集合  $V_B$  による誘導部分グラフのエッジ集合  $E_1 \cap E_2$  から,  $0.5|E_1 \cap E_2|$  本のエッジをランダム非復元抽出し削除する.
6. 作成したグラフ  $G_1, G_2$  のノードのうち, 孤立ノードを削除する.

こうして作成したデータを人工 URV データセットと呼ぶことにする. 人工 URV データセットのノード集合  $V_B \subseteq V_1, V_B \subseteq V_2$  同士にエッジを張り, そのうち 50 本のエッジを訓練データ, 残りをテストデータとし予測精度を測る実験を行った. 結果を図 1 及び表 1 に示す.

クロネッカー積類似度, 提案ペアワイズ類似度, クロネッカー和類似度の順に, AUC の値が高いことがわかる. クロネッカー和類似度より精度が高い理由として, ネットワーク間エッジの周辺の類似度も定義したことにより, 予測に優れた情報が増えたためと考えられる.

**fbtw データセットを用いた実験**

fbtw データセットを用いて実験を行う. 今回の実験では訓練データ数を 40 とし, 既存のペアワイズ類似度および提案ペアワイズ類似度を用いたリンク伝播法の予測精度を求めた. 予測精度の結果を図 2 及び表 1 に示す.

AUC は, クロネッカー和類似度, クロネッカー積, 提案ペアワイズ類似度の順に高いことがわかる. しかし,  $k$  がお

表 1: 各ペアワイズ類似度の空間計算量及び, 人工 URV 及び fbtw データセットの AUC

pairwise similarity	space complexity	AUC	
		URV	fbtw
kronecker prod	$O(N^2M^2)$	0.91853	0.78644
kronecker sum	$O(N^2M + M^2N)$	0.87294	0.80481
proposal	$O(N^2M + M^2N + LNM)$	0.89626	0.76128

よそ 200 より小さい場合は, クロネッカー積類似度, 提案ペアワイズ類似度の予測精度が高く,  $k$  がおよそ 250 以上では, クロネッカー和類似度が良い結果となり,  $k$  が小さい場合においてはクロネッカー積や提案ペアワイズ類似度が有利と言える. 提案ペアワイズ類似度が  $k$  がおよそ 200 以上の際に, 予測精度の向上が伸び悩んでいる. この結果については現在検討中である.

**6. まとめ**

本稿では, ネットワーク間エッジを用いたペアワイズ類似度を提案した. 人工的に作ったデータに対し, このペアワイズ類似度は, スパース性を持ちつつ予測精度も比較的高いということがわかった. 今後の課題として, fbtw データセットでの結果の検討, 3 ネットワーク以上のネットワーク間リンクの予測へのペアワイズ類似度の適用, 予測精度の向上に有効なペアワイズ類似度の設計などがある.

**謝辞**

本研究は JST さきがけ「知の創成と情報社会」の助成を受けたものである.

**参考文献**

- [1] A. Ben-Hur and W.S. Noble. Kernel methods for predicting protein-protein interactions. *Bioinformatics*, Vol. 21, pp. i38-i46, 2005.
- [2] H. Kashima, T. Kato, Y. Yamanishi, M. Sugiyama, and K. Tsuda. Link propagation: A fast semi-supervised learning algorithm for link prediction. In *Proceedings of the SIAM International Conference on Data Mining*, pp. 1099-1110, 2009.
- [3] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. *IEEE Symposium on Security and Privacy*, Vol. 0, pp. 173-187, 2009.
- [4] S.V.N. Vishwanathan, Karsten M. Borgwardt, and Nicol N. Schraudolph. Fast computation of graph kernels. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pp. 1449-1456. MIT Press, Cambridge, MA, 2007.
- [5] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, Vol. 68, No. 6, p. 065103, 2003.
- [6] R.I. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *in Proceedings of the Nineteenth International Conference on Machine Learning (ICML)*, pp. 315-322, 2002.