

# リンク予測における時間減衰の効果について

## On the effect of time decay function on link prediction

志村海<sup>1\*</sup> 白川真一<sup>2</sup> 大原剛三<sup>2</sup>  
Kai Shimura<sup>1</sup> Shinichi Shirakawa<sup>2</sup> Kouzou Ohara<sup>2</sup>

<sup>1</sup> 青山学院大学大学院理工学研究科

<sup>1</sup> Graduate School of Science and Engineering, Aoyama Gakuin University

<sup>2</sup> 青山学院大学理工学部

<sup>2</sup> College of Science and Engineering, Aoyama Gakuin University

**Abstract:** In this paper, we address the link prediction problem in complex network analysis, where the aim is predicting links most likely to be added to a given network in the future. Especially, we focus on the approach based on structural properties of a network such as the number of neighboring nodes, and take into account time decay factors for existing links in computing link prediction measurements. Indeed, two kinds of time decay functions are considered in this paper: exponential decay function and power-law decay function. We experimentally investigate the effects of these time decay functions on link prediction and show they can improve the predictive accuracy if a training period of sufficient length is given.

## 1 はじめに

近年、注目を集めている研究分野の1つに複雑ネットワークがある。その研究対象は、実世界における多くの実体間の関係を表現する、巨大、かつ複雑なネットワークであり、それらに共通する諸性質を明らかにすることが主要な目的となっている。たとえば、人間関係、インターネット、食物網、神経ネットワークなど実世界における様々な関係性が複雑ネットワークとして表現され得る。これら個々のネットワークの性質は、従来からそれぞれの応用分野で研究されてきたが、1990年代後半におけるスモールワールドモデル [1] やスケールフリーネットワークモデル [2] の提案を機に、それらの複雑ネットワークに共通する性質への関心が高まり、その研究はこの十数年の間に急速な発展を遂げてきた。

そのような複雑ネットワークに関する研究の1つにリンク予測がある [3, 4]。これは、現時点のネットワークに対して、将来、新たに発生するリンクを予測することを目的としたものである。たとえば、研究者をノード、論文の共著関係をリンクとした共著者ネットワークを対象とした場合、将来のリンクを予測することは、将来の論文の共著関係を予測すること、すなわち新たな共同研究者を推定することに相当する。また、SNS

(Social Networking Service) における友達推薦も、同じ問題と捉えることができる。一方、ネットワークに新たなノードが加わった際に、どのノード間にリンクが生じやすいかを予測することもリンク予測の対象となるが、本稿では、ネットワーク中のノードが増減しない場合のリンク予測問題について議論する。

個々のノードペア間のリンクの有無を予測する方法に関しては、ノード情報に基づくアプローチ、構造情報に基づくアプローチ、およびそれらを組み合わせたアプローチなどがあるが [4]、本稿ではその中でも特に構造情報に基づくアプローチに着目する。ここでいう構造情報とは、ノード周辺に存在するリンク構造のことを指し、そこから導かれる指標の値に基づいて2つのノード間にリンクが存在しそうかどうかを判定する。たとえば、2つのノードが多くの隣接ノードを共有している場合、そうでないノード間よりも、それらのノード間の方がリンクが生じやすいと考えることができる。そのようなリンク予測指標としては、共通隣接ノード指標 [5]、Jaccard 係数 [6]、Adamic/Adar [7] などが利用され、利用する指標値の降順にノードのペアを並べ、ある閾値以上となるノード間にリンクがあると予測する [3]。

一方、社会ネットワークなどを想定した場合、予測の基準となる時点で存在しているリンクをすべて同等に扱って上記のような指標を計算することは必ずしも妥当ではない。たとえば、共著者関係であれば、研究対象の変化や所属する組織の変化などの要因により、共同

\*連絡先：青山学院大学大学院理工学研究科  
〒252-5258 神奈川県相模原市中央区淵野辺 5-10-1  
E-mail: c5614131@aoyama.jp

研究者は変化し得るため、古い共著関係を表すリンクは直近の共著関係を表すリンクよりその重要性は低いものとしてもよいと考えられる。そのような考えの下、本研究ではリンク予測指標の計算時に、既存のリンクに対する時間減衰重みを考慮することにより、その予測精度がどのように変化するかを調べた。時間減衰重みとして利用可能な関数としては幾つか考えられるが、ここではそれらのうち指数関数的減衰とべき則減衰という代表的な2つについて考える。また、実際の共著者ネットワークを用い、それらの時間減衰重みがリンク予測の精度に与える影響を実験的に検証する。

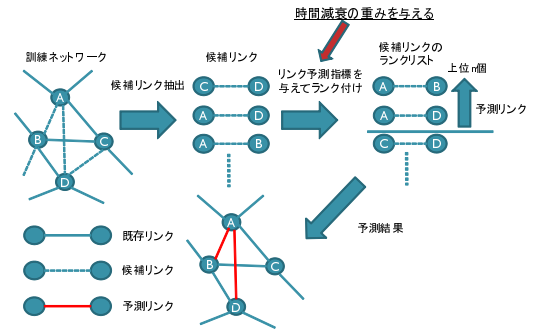


図 1: リンク予測法の概要

## 2 構造情報に基づくリンク予測

### 2.1 リンク予測法の概要

本研究では、文献 [3] と同じ方法でリンクを予測した場合における時間減衰重みの効果を検証する。そのリンク予測手順の概要を図 1 に示す。まず、予測の基準となる時点より以前を訓練期間、そより後をテスト期間とし、訓練期間中に生じたリンクからなるネットワークを訓練ネットワーク、さらにテスト期間中に生じたリンクも加えたネットワークをテストネットワークとする。そして、訓練ネットワークにおいてリンクの存在しないノードペア（以下、候補ペアと呼ぶ）に対してリンク予測指標値を計算し、その値の上位  $n$  個の候補ペア間に、テスト期間中にリンクが生じると予測する。ここで、 $n$  はテスト期間中に実際に生じたリンク数である。従って、予測結果の  $n$  本のリンクのうち、実際にテストネットワークにおいて存在したリンクの割合を予測精度として評価指標に用いる。ただし、文献 [3] では、すべてのノードを対象にするのではなく、訓練ネットワーク、テストネットワークいずれにおいても次数が 3 以上のノード（以下、コアノードと呼ぶ）のみを対象としており、本稿でもそれに従うものとする。この結果、訓練ネットワークには存在しないが、テストネットワークには存在するようなノードは考慮されないことに注意されたい。

### 2.2 予測指標

リンク予測指標としては、文献 [3] で用いられたもののうち、共通隣接ノード指標 [5]、Jaccard 係数 [6]、Adamic/Adar [7]、Preferential attachment [2]、および Katz $\beta$  [8] を用いる。以下に、2つのノード  $x, y$  に対するそれぞれの定義を示す。

#### 共通隣接ノード指標

$$COM(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

多くの隣接ノードを共有するノードペアの間には、そうでないノードペアよりも新たなリンクが発生しやすいという考えの下、ノード  $x$  と  $y$  の共通隣接ノード数をその値とする指標である。ここで、 $\Gamma(x)$  はノード  $x$  の隣接ノード集合を表す。

#### Jaccard 係数

$$JAC(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (2)$$

基本的な考え方は共通隣接ノード指標と同様であるが、その値を2つのノードの隣接ノードの総数（重複を除く）で割ることで得られる、共通隣接ノードの割合をその値とする指標である。

#### Adamic/Adar

$$ADA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log|\Gamma(z)|} \quad (3)$$

ノード  $x, y$  の共通隣接ノードが他の隣接ノードを多く持たないほど、両者の類似性/共通性が高く、リンクが生じやすいという考えに基づいた指標であり、ノード  $x, y$  の共通隣接ノード  $z$  の隣接ノード数  $|\Gamma(z)|$  が少ないほど大きな値を取る。

#### Preferential attachment

$$PRE(x, y) = |\Gamma(x)| \cdot |\Gamma(y)| \quad (4)$$

スケールフリーネットワークの生成モデルに基づき、隣接ノードの多い2つのノード間には新しいリンクが生成されやすいという考えの下、2つのノード  $x, y$  の隣接ノード数の積をその値とする指標である。

#### Katz $\beta$

$$KAT(x, y) = \sum_{\ell=1}^{\infty} \beta^{\ell} \cdot |\text{paths}_{xy}^{<\ell>}| \quad (5)$$

2つのノード  $x$  と  $y$  を結ぶパスが多ければ多いほど、両者の共通性が高いという考えの下、2ノード  $x$ ,  $y$  間の長さ  $\ell$  のパス数の加重和をその値とする指標である。ここで、 $paths_{xy}^{<\ell>}$  は  $x$ ,  $y$  間の長さ  $\ell$  のパス集合を表し、 $\beta$  は  $[0,1)$  の減衰係数であり、パスが長くなるほど重みが小さくなるように定義されている。ここで、 $I$  を単位行列、 $M$  をネットワークの隣接行列としたとき、ノード  $v_i$ ,  $v_j$  間の指標値は  $(I - \beta M)^{-1} - I$  により得られる行列の  $(i, j)$  成分として求めることができる。なお、文献 [3] では、 $x$ ,  $y$  間に多重リンクがある場合に、それらを重複して数えるものと、重複カウントしないものの2種類を用いている。本稿では、前者を  $KAT_W$ 、後者を  $KAT_{UW}$  と表記する。

次節では、以上のリンク予測指標に対して、既存リンクに対する時間減衰重みを考慮した定義を与える。

### 3 時間減衰を考慮したリンク予測

#### 3.1 時間減衰重み

本研究では、リンクに対する代表的な時間減衰重みとして、指数関数減衰とべき則減衰の2つを考える。以下では、対象リンクの生成時刻を  $t$ 、リンク予測時の時刻を  $t'$  とし、その間の時間差を  $\Delta t = t' - t$  とする。このとき、本稿では次式で定義される  $w_e(\Delta t)$  を指数関数減衰重みとして利用する。

$$w_e(\Delta t) = e^{-\lambda(\Delta t)} \quad (6)$$

ここで、 $\lambda$  は  $\lambda > 0$  となるパラメータである。

一方、べき則減衰重みは、同様に  $\Delta t$  と  $\lambda$  を用いて以下のように定義される  $w_p$  を用いる。

$$w_p(\Delta t) = (\Delta t)^{-\lambda} \quad (7)$$

実際には、どのような時間減衰が適切かは対象とするネットワークの性質に依存すると考えられるため、本稿では、それぞれの重みを個別に適用し、そのリンク予測結果を実験的に比較する。

#### 3.2 リンク予測指標への時間減衰重みの導入

上記の時間減衰重みによりその値が重みづけられるのはネットワーク中のリンクである。このことから、実際には隣接行列の要素を時間減衰重みにより重みづけることになる。そのため、時間減衰重みを考慮した隣接行列を最初に定義する。以下では、 $w(\Delta t)$  を指数関数減衰重み  $w_e(\Delta t)$ 、べき則減衰重み  $w_p(\Delta t)$  のい

れかを表すものとする。まず、簡単のために多重リンクが存在しない場合を考える。この場合、時間減衰重み  $w(\Delta t)$  で重みづけられた隣接行列  $M^w$  の  $(i, j)$  成分  $m_{ij}^w$  は元の隣接行列  $M$  の  $(i, j)$  成分  $m_{ij}$  を用いて以下のように定義できる。

$$m_{ij}^w = m_{ij} \times w(\Delta t_{ij}) \quad (8)$$

ここで、 $\Delta t_{ij}$  は、ノード  $v_i$ ,  $v_j$  間のリンク  $e_{ij}$  の生成時刻を  $t_{ij}$ 、現在の時刻を  $t'$  としたとき、 $\Delta t_{ij} = t' - t_{ij}$  と定義される。

次に、 $KAT_W$  の計算に必要となる多重リンクがある場合を考える。ノード  $v_i$  と  $v_j$  間に存在するリンクの本数を  $n_{ij}$  としたとき、元の隣接行列  $M$  の  $(i, j)$  成分が  $m_{ij} = n_{ij}$  と与えられるものとする。このとき、ノード  $v_i$  と  $v_j$  間の  $n_{ij}$  本のリンクのうち  $k$  番目のリンクの生成時刻を  $t_{ij}^k$  とすると、時間減衰重みで重みづけられた隣接行列  $M^{wm}$  の  $(i, j)$  成分  $m_{ij}^{wm}$  は  $n_{ij}$  本のリンクの重みの総和として以下のように定義する。

$$m_{ij}^{wm} = \sum_{k=1}^{n_{ij}} w(\Delta t_{ij}^k) \quad (9)$$

ここで、 $\Delta t_{ij}^k = t' - t_{ij}^k$  である。

以下では、これらを用いて時間減衰重みを考慮した各予測指標の定義について述べる。

#### 時間減衰重みつき共通隣接ノード指標

$$COM^w(x, y) = \frac{1}{2} \sum_{z \in \Gamma(x, y)} (m_{xz}^w + m_{yz}^w) \quad (10)$$

ここで、 $\Gamma(x, y) = \Gamma(x) \cap \Gamma(y)$  である。時間減衰重みを考慮しない場合の共通隣接ノード指標は、ノード  $x$ ,  $y$  の共通隣接ノード数であった。これは、ノード  $x$  と  $y$  間の長さ2のパスの総数に相当する。したがって、ここでは  $x$  と  $y$  の共通隣接ノード  $z$  に対して、 $x$  と  $z$  間のパスに対する重み、および  $y$  と  $z$  間のパスに対する重みの合計に  $1/2$  を乗じた値、すなわちそれらの平均をすべての共通隣接ノードに対して合計した値を指標値とする。

#### 時間減衰重みつき Jaccard 係数

$$\begin{aligned} JAC^w(x, y) \\ = \frac{COM^w(x, y)}{COM^w(x, y) + \sum_{z \in \Gamma(x, y)} \max(m_{xz}^w, m_{yz}^w)} \end{aligned} \quad (11)$$

ここで、 $\bar{\Gamma}(x, y) = \Gamma(x) \cup \Gamma(y) - \Gamma(x) \cap \Gamma(y)$  である。時間減衰重みを考慮しない Jaccard 係数は、ノード  $x$  と  $y$  の共通隣接ノード数を両者の隣接ノードの重複を含まない総数  $|\Gamma(x) \cup \Gamma(y)|$

で割った値であった。この  $|\Gamma(x) \cup \Gamma(y)|$  の値は、ノード  $x$  と  $y$  それぞれを端点とする長さ 1 のパスからもう一方の端点が重複するものを除いたものの数に相当する。このことから、 $COM^w(x, y)$  にノード  $x$  もしくは  $y$  から非共通隣接ノード  $z \in \bar{\Gamma}(x, y)$  へのパスに対する時間減衰重みの総和を加えたもので  $COM^w(x, y)$  を割った値を、時間減衰重み付きの Jaccard 係数とする。上記における  $\max(m_{xz}^w, m_{yz}^w)$  では、 $z \in \bar{\Gamma}(x, y)$  に対して、 $m_{xz}^w$  と  $m_{yz}^w$  のいずれかは 0 となることに注意されたい。

#### 時間減衰重みつき Adamic/Adar

$$ADA^w(x, y) = \sum_{z \in \Gamma(x, y)} \frac{1}{\log \sum_i m_{zi}^w} \quad (12)$$

時間減衰重みを考慮しない Adamic/Adar におけるノード  $x$  と  $y$  の共通隣接ノード  $z$  の隣接ノード数  $|\Gamma(z)|$  は、ノード  $z$  を端点とする長さ 1 のパスの総数に相当することから、各パスに対する時間減衰重みの総和  $\sum_i m_{zi}^w$  をその代わりに用いた値を時間減衰重みつき Adamic/Adar とする。

#### 時間減衰重みつき Preferential Attachment

$$PRE^w = \sum_{z \in \Gamma(x)} m_{xz}^w \cdot \sum_{z \in \Gamma(y)} m_{yz}^w \quad (13)$$

これまでの議論と同様に、ノード  $x$  もしくは  $y$  に対する隣接ノード数は、各ノードとその隣接ノードを端点とする長さ 1 のパスの総数に相当することから、各パスに対する時間減衰重みの総和をその代わりに用いた上記の式の値を時間減衰重みつき Preferential Attachment の値とする。

#### 時間減衰重みつき Katz $\beta$

$$KAT^w(x, y) = \sum_{\ell=1}^{\infty} \beta^\ell \cdot |paths_{xy}^{w<\ell>}| \quad (14)$$

ここで、 $paths_{xy}^{w<\ell>}$  は、2 ノード  $x, y$  間の時間減衰重みで重みづけられた長さ  $\ell$  のパス集合とする。本研究では、実際には、 $(I - \beta M^w)^{-1} - I$  により求められる行列の値を多重リンクを考慮しない  $KAT_{UW}$  に対して時間減衰重みを考慮した値として  $KAT_{UW}^w$  と表し、 $(I - \beta M^{wm})^{-1} - I$  により求められる行列の値を多重リンクを考慮した  $KAT_W$  に対して時間減衰重みを考慮した値として  $KAT_W^w$  と表す。

以上、いずれの場合においても、時間減衰重みがすべて 1 の場合、すなわち  $\lambda = 0$  の場合、時間減衰重み付きの指標値は、時間減衰重みを考慮しない場合の指標値と一致する。

## 4 評価実験

### 4.1 実験設定

本実験では、文献 [3] で用いられた、スタンフォード大学の Web サイト<sup>1</sup> で公開されている学術論文データセットである HEP-TH から構築した共著者ネットワークを用いた。HEP-TH は、論文投稿サイト arXiv<sup>2</sup> の HEP-TH (高エネルギー物理学分野) に 1993 年から 2003 年の間に投稿された 25,998 本の論文の情報をまとめたものであり、論文タイトル、著者名、キーワード、アブストラクトなどの情報が投稿月ごとにまとめられている。実際には、同 Web サイトでは、このデータセットから生成された共著者ネットワークの構造データが公開されているが、リンクに出版年の情報が付加されておらず、時間減衰重みを考慮する本研究ではそのまま利用できないため、元データから論文の出版年もその論文の共著者間のリンクに属性とした付与したネットワークデータを改めて生成した。そのため、名寄せ処理などの違いから、文献 [3] で用いられたネットワークとは構造が異なることに注意されたい。

リンク予測手順は、2 節で述べた通り文献 [3] に従い、訓練期間とテスト期間は表 1 に示す 2 種類とした。表 1 における実験 1 の設定は、文献 [3] で用いられた期間と同じであり、精度の比較を目的としている。これに対して実験 2 の設定では、訓練期間を倍にすることで、時間減衰重みの効果がリンク予測精度にどのように影響するのかを調べる。なお、本実験では、年を時間の単位とし、予測指標としては、3 節で述べた  $COM^w$ 、 $JAC^w$ 、 $ADA^w$ 、 $PRE^w$ 、 $KAT_{UW}^w$ 、 $KAT_W^w$  を用いた。Katz $\beta$  のパラメータ  $\beta$  の値は、文献 [3] と同様に 0.05、0.005、0.0005 とし、時間減衰重みのパラメータ  $\lambda$  は、 $[0.0, 0.9]$  の範囲で 0.1 刻みで変化させた。

評価指標は、2 節で述べたように訓練ネットワークに存在しないリンクの中でテストネットワークに存在すると予測したリンクのうち、実際にテストネットワークに存在するものの割合とする。形式的には、訓練ネットワーク、テストネットワークいずれにおいても次数が 3 以上であるコアノードの集合を  $Core$  とし、訓練ネットワークを頂点集合  $V$ 、リンク集合  $E_{old}$  を用いて  $G_{old} = (V, E_{old})$ 、テストネットワークを同様に  $G_{test} = (V', E_{test})$  と表すと、 $Core \subseteq V$  かつ  $Core \subseteq V'$  であり、予測指標を計算すべき候補リンク (ノードペア) の集合は  $(V \times V - E_{old}) \cap (Core \times Core)$  となる。一方、正解となる候補リンクの集合は、 $E_{new}^* = E_{new} \cap (Core \times Core)$  となる。ただし、 $E_{new} = (V \times V - E_{old}) \cap E_{test}$  である。実際には、 $n = |E_{new}^*|$  としたとき、上記の候補リンク集合に含まれる候補リンクのうち、その予測

<sup>1</sup><http://snap.stanford.edu/data/>

<sup>2</sup><http://arxiv.org/>

表 1: 実験で用いたデータの組み合わせ

	訓練データ			テストデータ		
	年度	著者	共著関係	年度	著者	共著関係
実験 1	1994-1996	3,825	6,011	1997-1999	4,050	7,130
実験 2	1993-1999	6,205	12,859	2000-2002	4,685	8,809

表 2: 実験 1 (訓練期間 [1994, 1996], テスト期間 [1997, 1999]) における各リンク予測指標による予測精度 (ランダム予測の精度 0.00321 を 1 とした場合の比率)

	$COM^w$	$JAC^w$	$ADA^w$	$PRE^w$	$KAT_W^w$			$KAT_{UW}^w$		
					$\beta=0.05$	$\beta=0.005$	$\beta=0.0005$	$\beta=0.05$	$\beta=0.005$	$\beta=0.0005$
$\lambda=0.0$ (時間減衰なし)	25.22	25.88	25.71	5.67	15.67	<b>27.33</b>	27.00	25.33	25.00	25.00
指数関数減衰										
$\lambda=0.1$	25.26	<b>26.33</b>	25.33	6.00	19.00	23.67	24.33	25.33	25.67	26.00
$\lambda=0.2$	25.26	<b>27.00</b>	26.00	6.33	21.00	25.00	25.00	25.67	26.33	26.33
$\lambda=0.3$	25.26	26.67	<b>27.83</b>	6.33	21.67	24.33	24.33	25.33	26.00	26.00
$\lambda=0.4$	25.32	27.00	<b>27.66</b>	6.33	22.67	24.67	25.00	25.00	26.33	26.33
$\lambda=0.5$	25.70	<b>27.33</b>	<b>27.33</b>	6.00	23.33	23.67	23.67	27.00	26.00	26.00
$\lambda=0.6$	24.88	26.67	<b>27.30</b>	6.00	24.00	24.33	24.33	25.00	26.67	26.33
$\lambda=0.7$	25.11	<b>26.33</b>	<b>26.33</b>	6.00	24.33	25.33	25.67	23.00	24.00	23.67
$\lambda=0.8$	25.12	<b>25.67</b>	<b>25.67</b>	6.00	24.33	25.33	25.33	23.67	23.33	23.33
$\lambda=0.9$	25.19	<b>26.00</b>	23.00	5.67	25.00	25.00	25.00	23.00	23.00	23.00
べき則時間減衰										
$\lambda=0.1$	25.26	<b>26.33</b>	25.67	5.67	16.00	24.00	24.00	24.67	25.67	25.67
$\lambda=0.2$	25.26	<b>26.33</b>	25.33	6.00	16.67	23.66	24.00	24.33	25.67	26.00
$\lambda=0.3$	25.26	<b>27.00</b>	25.66	6.33	18.66	24.66	24.66	24.66	26.33	26.00
$\lambda=0.4$	25.26	<b>27.00</b>	25.64	6.33	18.33	24.33	25.00	25.33	26.33	26.33
$\lambda=0.5$	25.26	<b>27.33</b>	25.66	6.33	19.00	23.67	23.67	25.67	26.33	26.33
$\lambda=0.6$	25.26	<b>27.00</b>	26.00	6.00	20.00	23.67	23.67	26.00	26.33	26.33
$\lambda=0.7$	25.32	27.22	<b>27.33</b>	6.00	21.00	23.33	24.00	25.00	25.67	25.67
$\lambda=0.8$	25.70	<b>27.33</b>	<b>27.33</b>	6.00	21.00	23.33	23.67	24.00	26.00	26.00
$\lambda=0.9$	24.82	27.00	<b>27.33</b>	6.00	21.67	24.00	24.33	23.67	25.00	26.33

指標値が上位  $n$  個までのリンク集合  $R$  を予測結果とすることから, その予測精度は  $|R \cap E_{new}^*|/n$  で与えられる.

## 4.2 実験結果と考察

まず, 実験 1 の結果を表 2 にまとめる. 表 2 では, 文献 [3] と同様に, 各リンク予測指標に対する結果の値として, 候補リンクの中からランダムにリンクを予測した場合の予測精度 0.00321 を 1 としたときの予測精度の比率を示している. ネットワーク構造が若干異なるものの, 時間減衰重みを考慮しない場合 ( $\lambda = 0.0$ ) の予測精度は文献 [3] と同程度となっており, ランダム予測より高い精度を達成できていることがわかる. 一方, 時間減衰重みを適用した場合,  $KAT_W^w$  の  $\beta = 0.005$ ,  $\beta = 0.0005$  以外は, いずれも時間減衰重みを加味しない場合と比べて予測精度が向上していることがわか

る. ただし, その向上率はそれほど高くなく, また, パラメータ  $\lambda$  の値によって変動していることがわかる. 全体的には, 指数関数減衰重みと比較してべき則減衰重みのほうが同程度の改善率を達成するのにより大きな  $\lambda$  を設定する必要があるといえる. これは, べき則減衰が指数減衰よりも緩やかに減衰し, かついずれの減衰重みにおいても  $\lambda$  の値が小さいほど減衰効果が小さいために, べき則減衰重みでは指数減衰重みと同程度の効果を得るためには, より大きな  $\lambda$  の値が必要であったためと考えられる. 一方,  $KAT_W$  に関しては, 時間減衰重みを加味すると予測精度が下がることが確認された. これは,  $KAT_W$  では比較的高く評価される短いパスにも時間減衰重みが適用され, その指標値における価値が過剰に減少するためであると考えられる.  $KAT_W$  では多重リンクを考慮しているため, その影響が多重リンクを考えない  $KAT_{UW}$  より大きくなったものと考えられる.

表 3: 実験 2 (訓練期間 [1993, 1999], テスト期間 [2000, 2002]) における各リンク予測指標による予測精度 (ランダム予測の精度 0.00008 を 1 とした場合の比率)

$\lambda=0.0$ (時間減衰なし)	$COM^w$	$JAC^w$	$ADA^w$	$PRE^w$	$KAT_W^w$			$KAT_{UW}^w$		
	60.64	<b>69.81</b>	67.67	9.35	$\beta=0.05$ 6.47	$\beta=0.005$ 46.07	$\beta=0.0005$ 47.51	$\beta=0.05$ 35.27	$\beta=0.005$ 56.87	$\beta=0.0005$ 56.87
指数関数減衰										
$\lambda=0.1$	68.39	71.99	67.82	10.07	30.23	64.07	66.95	52.55	<b>73.43</b>	<b>73.43</b>
$\lambda=0.2$	71.37	77.03	63.35	9.35	23.75	64.07	64.07	68.39	<b>77.75</b>	77.03
$\lambda=0.3$	76.75	<b>79.19</b>	56.51	10.79	48.95	68.39	68.39	70.55	77.03	76.31
$\lambda=0.4$	70.76	<b>72.72</b>	54.00	11.52	53.28	69.12	69.12	70.56	72.00	72.00
$\lambda=0.5$	77.66	<b>79.19</b>	55.43	10.07	60.47	68.39	69.11	68.39	71.99	71.99
$\lambda=0.6$	<b>77.76</b>	<b>77.76</b>	51.12	10.08	61.92	66.96	66.96	72.72	72.00	72.00
$\lambda=0.7$	76.65	<b>77.76</b>	44.64	10.80	63.36	69.12	69.12	67.68	69.84	69.84
$\lambda=0.8$	75.60	<b>77.76</b>	43.92	11.52	62.64	68.40	68.40	66.24	69.12	69.84
$\lambda=0.9$	75.99	<b>77.03</b>	54.71	12.23	63.35	68.39	69.11	70.56	68.39	69.11
べき則減衰										
$\lambda=0.1$	68.15	<b>70.55</b>	69.11	10.07	25.19	61.19	61.91	42.47	66.95	67.67
$\lambda=0.2$	68.15	<b>73.43</b>	67.67	10.79	20.15	63.35	62.63	51.11	66.95	69.83
$\lambda=0.3$	68.39	<b>75.59</b>	66.95	10.07	37.43	64.07	64.79	52.55	73.43	41.39
$\lambda=0.4$	67.03	74.88	66.24	10.01	13.68	62.20	62.64	60.48	74.16	<b>75.60</b>
$\lambda=0.5$	71.24	<b>77.03</b>	64.79	9.35	23.03	62.63	62.63	64.79	75.59	75.59
$\lambda=0.6$	73.20	<b>78.48</b>	61.92	10.08	41.04	66.96	66.24	64.80	75.60	77.76
$\lambda=0.7$	72.50	<b>77.04</b>	59.76	10.08	8.64	65.52	66.24	63.36	71.28	72.00
$\lambda=0.8$	73.04	<b>80.64</b>	58.32	11.52	19.44	67.68	68.40	62.64	71.28	70.56
$\lambda=0.9$	75.59	<b>79.19</b>	58.31	10.07	26.63	68.39	69.11	62.63	72.71	72.71

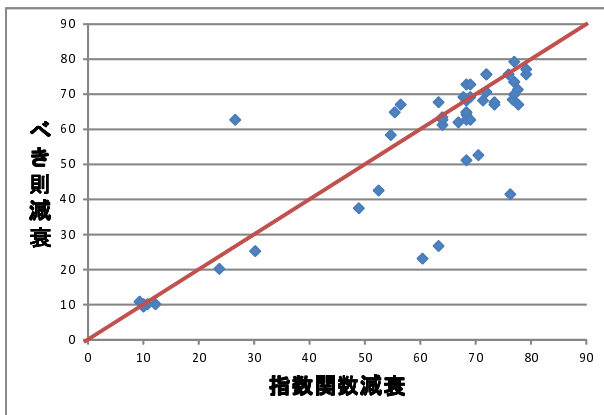


図 2: 実験 2 における散布図による正解率の比較

次に、実験 2 の結果を表 2 と同様に表 3 にまとめる。本実験では、ランダム予測の予測精度は 0.00008 であった。実験 1 の結果と比べると、多くの指標において、時間減衰重みを加味した場合の予測精度が大幅に向上していることがわかる。これは、訓練期間が実験 1 と比べて長くなり、リンク予測において評価を下げるべき古い共著関係の価値が適切に減衰されたためと考えられる。また、実験 1 の結果と同様に、指数関数減衰重み

のほうがべき則減衰重みと比べてより小さな  $\lambda$  の値で大きな改善率の向上を達成していることがわかる。ここで、指数減衰重みとべき則減衰重みのどちらがリンク予測精度を高くする傾向にあるのかを定量的に調べるために、同じ  $\lambda$  に対する結果 (ランダム予測に対する予測精度の比率) を、指数関数減衰重みの結果の値を  $x$  軸、べき則減衰重みの結果の値を  $y$  軸に取ってプロットした散布図を図 2 に示す。対角線より下の指数関数減衰重みの予測精度が高い領域には、対角線より上のべき則減衰重みの予測精度が高い領域と比較すると、プロットされている結果が多いことから、今回のデータセットに関しては、べき則減衰よりも指数関数減衰の方がリンクの価値の時間減衰をより適切に表していると言える。

## 5 結論

本研究では、複雑ネットワークにおけるリンク予測問題に対して、既存のリンクに時間減衰重みを導入し、そのリンク予測における影響を調べた。具体的には、従来研究で用いられたネットワークの構造情報に基づくリンク予測指標に対し、指数関数減衰重みとべき則減衰重みを導入し、共著者ネットワークを対象に予測精度の変化を実験的に検証した。その結果、訓練期間をある程度長く取ることで、時間減衰効果が明確に現れ

るようなり，時間減衰重みを加味しないよりも加味した方が，多くの予測指標においてその予測精度が大きく改善され得ることが明らかとなった．また，指数関数減衰重みとべき則減衰重みの比較という観点からは，いずれも予測精度の改善に寄与するものの，年を時間の単位とした共著者ネットワークに関しては，指数関数減衰重みのほうが小さいパラメータ値で高い改善率を達成できることがわかった．

今後の課題としては，より多様なネットワークデータを用いたさらなる検証実験が挙げられる．本研究では，一種類の共著者ネットワークしか用いなかったが，今後は，他の共著者ネットワーク，ならびに他の社会ネットワーク，他のドメインにおける複雑ネットワークなどを対象に，時間減衰重みがリンク予測に与える影響をより詳細に検証する必要がある．また，今回は予測対象を訓練ネットワークに含まれるノードペアに限定したが，実際のネットワークでは新たなノードが生じることもあるため，今後はそのようなノードの変化も考慮した場合のリンク予測に関しても，時間減衰の効果を検証する必要がある．

## 参考文献

- [1] D. J. Watts and S. H. Strogatz: “Collective dynamics of ‘small-world’ networks,” *Nature*, Vol.393, pp.440–442(1998).
- [2] A.-L. Barabasi and R. Albert: “Emergence of Scaling in Random Networks,” *Science*, Vol.286, pp.509–512(1999).
- [3] D. L. Nowell, J. Kleinberg: “The link prediction problem for social networks,” In *proc. of CIKM2003*, pp.556–559(2003).
- [4] 鹿島久嗣: “ネットワーク構造予測”，*人工知能学会誌*, Vol.22, No.3, pp.334–351(2007)
- [5] M. Newman: “Clustering and preferential attachment in growing networks,” *Physical Review E*, Vol.64, No.2, 025102(2001)
- [6] C. Manning and H. Schuetze: “Foundations of statistical natural language processing,” MIT Press (1999).
- [7] L. Adamic and E. Adar: “Friends and neighbors on the web,” *Social Networks*, Vol.25, No.3, pp.211–230(2003).
- [8] L. Katz: “A new status index derived from sociometric analysis,” *Psychometrika*, Vol.18, No.1, pp.39–43 (1953).