Refereed Conference paper

ネットワークの局所構造における成長パターンの発見とその応用

甲谷 優[-1] 西田京介[†1] 藤村 考[†1]

ネットワークの局所構造における成長パターンに着目したネットワークを分析する手法を提案する．本手法によりこれまでの分析では理解することが困難であったネットワークの生成過程に関する性質が明らかになり，さらにどのノード間に将来エッジが発生するかを予測することができる．エッジの発生の予測は「リンク予測」問題と呼ばれ，近年のソーシャルネットワーキングサービス等で友人推薦のコア技術となっており注目されている．本稿ではソーシャルメディアを用いて実験を行い，提案法により明らかになるソーシャルネットワークの性質と，リンク予測アルゴリズムとしての性能について報告する．

Discovering Evolutionary Patterns of Substructures in Complex Networks and its Applications

Yutaka Kabutoya,[†1] Kyosuke Nishida [†1]
and Ko Fujimura[†1]

We propose an entirely new approach to understanding complex networks; called dynamic network motifs (DNMs), it is significantly different from existing network analysis approaches in terms of focusing on the evolution of substructures in a graph. The concept has the potential to uncover previously unknown properties of graphs and allows us to infer whether new edges among nodes the present in a snapshot of the graph are likely to occur. This inference, called "link prediction", is significant because it yields various useful applications, such as interaction recommenders for social networking services. Experiments on snapshots of social networks based on real log data sets of social media show the following two significant results: First, DNM analysis has the potentials described above. Next, a new link predictor based on DNMs outperforms existing link predictors in terms of prediction accuracy.

1. はじめに

様々な分野でネットワーク分析に関する研究は数多くなされてきたが，近年は計算機の処理能力の向上から，部分ネットワーク (ネットワークのノードの部分集合と，それらの間のエッジからなるネットワーク) に関するマイニングが可能となった[10),13)]．

部分ネットワーク分析の 1 つに，ネットワークモチーフ分析と呼ばれる手法が存在する．ネットワークモチーフとは，ネットワーク中に特徴的に現れる部分ネットワーク構造と定義されており，ネットワ―ク理解のための新たな分析手法として注目を集めている．Milo ら[7),8)] はネットワークが時間とともに成長してもそのネットワークモチーフは変わらないことを発見し，ネットワークモチーフはネットワークの成長に関する制約を表すと言及している．しかしながらネットワークモチーフはある時点での静的なネットワークのスナップショットに対する分析であり，ネットワークの時間にともなう成長には着目できていない．

一方で，近年は部分ネットワークの成長パターンに関する研究も始まっている．たとえば Inokuchi ら[4)] は効率のよい部分ネットワークの成長パターンの発見に関して研究を行っているが，未だに部分ネットワークの成長パターンに関する統計的分析に関する研究はなされていない．そのため，どの部分ネットワークの成長パターンが重要なのかについて議論することは困難である．

我々は文献[14)] にて部分ネットワーク中のエッジの発生，再発生に関する統計的分析手法を提案した．本稿では，前述の近年の研究の流れを踏まえエッジの再発生は省き成長 (新しいエッジの発生) に関するもののみに着目し，ネットワークの成長過程における統計的に重要な部分ネットワークの成長パターン (**動的ネットワークモチーフ**と定義する) の発見手法を提案する．3 ノードの部分ネットワークの場合，エッジの発生と再発生のパターンは 54 通りであるが，そのうち本稿で着目する成長パターンは 24 通りである．

また我々[14)] は動的ネットワークモチーフのリンク予測への応用も提案している．リンク予測とはある時点でのネットワークのスナップショットが与えられた際に，そのネットワーク中のどの 2 ノード間に将来エッジが発生するかを推定する問題である．リンク予測技術は SNS における友人推薦機能[⋆1]をはじめとしてさまざまなアプリケーションに利用されてい

†1 日本電信電話株式会社 NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation

⋆1 たとえば，mixi (`http://mixi.jp`) が友人推薦機能を提供している

る．従来のリンク予測アルゴリズムの多くはネットワークの大域的特性に基づいている[*1]が，我々の手法はその局所構造に基づいている．本稿では文献[14] 記載のヒューリスティックな手法とは異なり，局所構造の変化パターンのうち成長に関するもののみを素性として 2 ノード間のエッジの発生を確率モデル化する手法を提案する．理論的な基礎を与えることで，より高精度なリンク予測が可能となると考えたためである．

本稿では様々なソーシャルメディアからネットワークを抽出し，動的ネットワークモチーフ分析と従来のネットワークモチーフ分析の比較を行う．また，リンク予測実験を行い，提案法が従来法よりも高精度にリンク予測できることを示す．

2. 従来法：ネットワークモチーフ

提案手法の説明に入る前に，その基礎であるネットワークモチーフ分析について言及する．今，あるネットワークのスナップショットを $\boldsymbol{G} = \langle \boldsymbol{V}, \boldsymbol{E} \rangle$ とする．ここでネットワーク $\boldsymbol{G}$ は有向であってもなくてもかまわない．

まず，サイズ n パターンを，ノード数が n で，連結であり，互いに同型でない全ネットワークと定義する．このとき，有向なネットワークの場合サイズ 3 パターンは 13，サイズ 4 パターンは 199 得られることが Milo ら[3] によって発見されている．

ここで与えられたネットワーク $\boldsymbol{G}$ においてあるパターンの出現頻度が等価なランダムネットワーク $\{\hat{\boldsymbol{G}}_k\}_{k=1}^{K}$ と比較して統計的に有意に高い場合，そのパターンをネットワークモチーフと定義する．ここで $\boldsymbol{G}$ と等価なランダムネットワークは，各ノードの接続するエッジ数，双方向エッジ数が同じになるように発生させる．

Milo ら[7] の提案している各パターンの統計量は以下のように算出することができる．まず，$\boldsymbol{G}$ 中のパターン i の出現頻度を m_i とする．また，ランダムネットワーク $\hat{\boldsymbol{G}}_k$ 中のパターン i の出現頻度を $\hat{m}_{ik}$ とする．このとき $\boldsymbol{G}$ におけるパターン i の統計量は以下に示すように Z スコアで与えられる．

$$z_i = \frac{m_i - \mu_i}{\sigma_i}, \tag{1}$$

ただし $\mu_i = \frac{1}{K}\sum_{k=1}^{K} \hat{m}_{ik}$, $\sigma_i^2 = \frac{1}{K}\sum_{k=1}^{K}(\hat{m}_{ik} - \mu_i)^2$.

このときベクトル $\boldsymbol{z} = \{z_i\}_{i=1}^{I}$ の長さを 1 とするように正規化することで，ネットワークモチーフプロファイルを得る．

[*1] たとえばクラスタ性に基づいた手法[1),9)]，スケールフリー性に基づいた手法[2)] が存在する

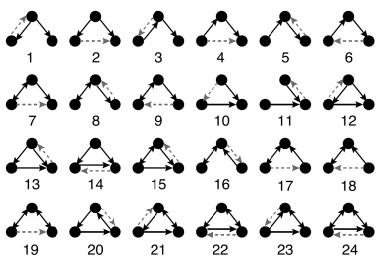

図 1 全 24 通りのサイズ 3 成長パターン
Fig. 1 All 24 size-three evolutionary patterns.

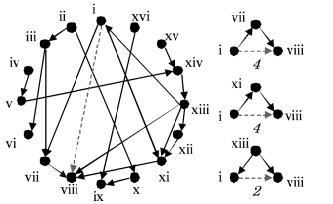

図 2 ノードペア ⟨i, viii⟩ は ⟨i, vii, viii⟩ と ⟨i, xi, viii⟩, ⟨i, xiii, viii⟩ の 3 つの成長に対応する
Fig. 2 Pair ⟨i, viii⟩ yields three triad-evolution, ⟨i, vii, viii⟩, ⟨i, xi, viii⟩, and ⟨i, xiii, viii⟩.

3. 提案法：動的ネットワークモチーフ

3.1 動的ネットワークモチーフ分析

本研究はネットワークの成長パターンに着目した手法であり，ネットワークが動的であれば期間 t から $t+1$ 内にノード $\boldsymbol{V}$ 間に発生したエッジが対象となる．本稿では静的なネットワークにも対応するために，それぞれのエッジ $e \in \boldsymbol{E}$ がネットワーク $\boldsymbol{G}_{\setminus e} = \langle \boldsymbol{V}, \boldsymbol{E} - \{e\} \rangle$ に対する新たなエッジの発生とみなす．

まずは n ノードの連結な部分ネットワーク中における新たな 1 本のエッジの発生をサイズ n 成長パターンと定義する．既存のエッジを実線で，新たなエッジの発生を点線で表わすと，有向ネットワークにおける全 24 通りのサイズ 3 成長パターンは図 1 のようになる．今扱う成長パターンが J 通りであるとする．

一般に，1 本のエッジの発生は複数の部分ネットワークの成長とみなすことができる．例を図 2 に示す．ここでエッジの発生 $\langle u, v \rangle$ に対応する成長パターン j の数を $\langle u, v \rangle$ の j に関する *EF* (*Evolution Frequency*) と定義し，x_{uvj} で表わす．図 2 の場合，$x_{\text{iviii}4} = 2$ となる．サイズ n 成長パターンに着目した場合，x_{uvj} は u，v と，$\boldsymbol{V}$ 中の u，v 以外の $(n-2)$ ノードの全組合せに対して，それらからなる部分ネットワークを検証することにより得ることができる．このとき $\boldsymbol{G}$ 内の成長パターン j の頻度は以下のようにして得られる．

$$n_j = \sum_{\langle u,v \rangle \in \boldsymbol{E}} x_{uvj}. \tag{2}$$

ここで与えられたネットワーク $\boldsymbol{G}$ において等価なランダムネットワーク $\{\hat{\boldsymbol{G}}_k\}_{k=1}^{K}$ と比

較してある成長パターンの頻度が統計的に有意に高い場合，その成長パターンを動的ネットワークモチーフと定義する．等価なランダムネットワークはネットワークモチーフ分析と同様の手法で発生させる．$\hat{\boldsymbol{G}}_k$ における成長パターン j の頻度を $\hat{n}_{jk}$ とするとき，$\boldsymbol{G}$ における成長パターン j の統計的な重みは以下に示すように Z スコアで与えられる．

$$\zeta_j = \frac{n_j - \nu_j}{\tau_j}, \tag{3}$$

ただし $\nu_j = \frac{1}{K}\sum_{k=1}^{K}\hat{n}_{jk}$，$\tau_j^2 = \frac{1}{K}\sum_{k=1}^{K}(\hat{n}_{jk} - \nu_j)^2$．

ベクトル $\boldsymbol{\zeta} = \{\zeta_j\}_{j=1}^{J}$ の長さを 1 とするように正規化することで，動的ネットワークモチーフプロファイルを得る．

3.2 成長パターンに基づくリンク予測

現在のネットワークが $\boldsymbol{G} = \langle \boldsymbol{V}, \boldsymbol{E} \rangle$ であるとき，将来のネットワークの $\boldsymbol{V}$ に関する部分ネットワークを $\boldsymbol{G}' = \langle \boldsymbol{V}, \boldsymbol{E}' \rangle$ とする．このとき $\boldsymbol{V}$ の 2 ノードの全順列から $\boldsymbol{E}' - \boldsymbol{E}$ を予測する問題がリンク予測である．

ここでエッジの発生は対応する EF 値に依存すると仮定する．

$$P(y=1|u,v) \propto P(y=1|\boldsymbol{x}) = P(y=1|x_1,\ldots,x_J), \tag{4}$$

ただし $\langle u,v \rangle \in \boldsymbol{E}'$ の場合 $y=1$，そうでない場合 $y=0$．ここで確率 $P(y=1|\boldsymbol{x})$ は以下のように推定することができる．

$$P(y=1|\boldsymbol{x}) = \frac{\left|\{\langle u,v\rangle \in \boldsymbol{E} | x_{uv1} = x_1 \wedge \ldots \wedge x_{uvJ} = x_J\}\right| + \delta}{\left|\{\langle u,v\rangle \in \boldsymbol{V}\times\boldsymbol{V} | x_{uv1} = x_1 \wedge \ldots \wedge x_{uvJ} = x_J\}\right| + 2\delta}, \tag{5}$$

ただし δ はスムージングパラメータである．しかしこのようにモデル化すると，EF 値の最大値が W である場合パラメータ数が $O(W^J)$ に達し，頑健なパラメータ推定が困難であるという問題点がある．

そこでまずはベイズの定理から u から v にエッジが発生する確率を，以下に示す尤度で近似する．

$$P(y=1|\boldsymbol{x}) = \frac{P(y=1)P(\boldsymbol{x}|y=1)}{\sum_{y'=0}^{1}P(y')P(\boldsymbol{x}|y')} \propto \frac{P(\boldsymbol{x}|y=1)}{P(\boldsymbol{x}|y=0)}, \tag{6}$$

このとき以下に示す 3 つの方法でパラメータ数を減らし，頑健な推定を行う．

(1) 独立性の仮定．
エッジが発生したという条件のもと成長パターン j の EF 値と j' の EF 値が独立であると仮定すると，下式

$$P(\boldsymbol{x}|y=1) = P(x_1,\ldots,x_J|y=1) = \prod_{j=1}^{J}P(x_j|y=1), \tag{7}$$

のように分解でき，パラメータ数を $O(JW)$ まで減らすことができる．エッジが発生したという条件に対する成長パターン j の EF 値の事後確率は以下のように推定できる．

$$P(x_j|y=1) = \frac{\left|\{\langle u,v\rangle \in \boldsymbol{E} | x_{uvj} = x_j\}\right| + \delta}{\left|\boldsymbol{E}\right| + \delta W}. \tag{8}$$

$P(\boldsymbol{x}|y=0)$ も同様に頑健に推定することが可能である．

(2) ベルヌーイ試行の仮定．
ベルヌーイ試行の仮定は文書分類においてもよく用いられている[3]．各ノードペアは各成長パターンが出現するか否かのベルヌーイ試行であると仮定すると，

$$P(\boldsymbol{x}|y=1) = \prod_{j=1}^{J}P(b_j=1|y=1)^{b_j}(1-P(b_j=1|y=1))^{1-b_j}, \tag{9}$$

ここで b_j は $x_j > 0$ の場合 1，それ以外の場合は 0 をとる．このようにベルヌーイ試行を仮定する場合，各成長パターン j の EF 値は考慮せず出現したか否かだけを考慮するため，パラメータ数は $O(J)$ まで減らすことができる．エッジの発生に対する成長パターン j の出現の事後確率は以下のように推定できる．

$$P(b_j=1|y=1) = \frac{\left|\{\langle u,v\rangle \in \boldsymbol{E} | x_{uvj} > 0\}\right| + \delta}{\left|\boldsymbol{E}\right| + 2\delta}. \tag{10}$$

(3) 多項分布の仮定．
多項分布の仮定もベルヌーイ試行と同様文書分類においてよく用いられている[11]．各ノードペアは $\sum_j x_j$ 回の独立した成長の結果だと仮定する．各成長はその成長パターンの出現確率 $P(b_j=1|y)$ で生じるものとしている．このとき

$$P(\boldsymbol{x}|y=1) \propto \prod_{j=1}^{J}\frac{P(b_j=1|y=1)^{x_j}}{x_j!}. \tag{11}$$

この場合のパラメータ数も $O(J)$ で，頑健な推定が可能である．ここでエッジの発生

表 1　ソーシャルネットワークの基本的な統計量
Table 1　General statistics of social networks created from datasets.

	Nodes	Edges	Density	Mutual edges	Clustering coefficient
Blog	3,079	46,695	0.507%	17,469	30.38%
Flickr	3,721	175,402	1.331%	35,938	30.50%
Email	5,122	21,346	0.086%	654	4.511%

に対する成長パターン j の出現の事後確率は以下のように推定する．

$$P(b_j = 1|y = 1) = \frac{\sum_{\langle u,v\rangle \in \boldsymbol{E}} x_{uvj} + \delta}{\sum_{j'=1}^{J} \sum_{\langle u,v\rangle \in \boldsymbol{E}} x_{uvj'} + \delta J}. \quad (12)$$

4. 実　　験

4.1 データセット

提案法を評価するため，以下に示す 3 種類のソーシャルメディアから得たネットワークデータを対象として実験を行った．ただし，いずれのデータからも接続しているエッジ数が 3 未満のノードは省いた．

(1) *Blog.* 本データセットは 2008 年 9 月 21 日から 2009 年 5 月 1 日の期間クロールした日本のブログ記事に基づいており，各ノードはブログサイト，各エッジはトラックバックを表す．

(2) *Flickr.* 本データセットは Flikcr API[*1] を用いてクロールした Flickr のメタデータのデータベースであり，各ノードはユーザで，各エッジはコメントのやりとりを表す．

(3) *Email.* 本データセットは Enron 社のメールログであり，研究目的でのみ利用が許可されている[*2]．各ノードはユーザで，各エッジはメールのやりとりを表す．

表 1 に各ソーシャルネットワークの統計量を示す．Flickr が他のメディアと比較して密なネットワークで，また Email が比較的疎なネットワークであることがわかる．

4.2 ネットワークモチーフ分析と動的ネットワークモチーフ分析

今回の実験ではサイズ 3 の部分ネットワークに注目する．

まず，図 3 は 3 つのソーシャルネットワークのネットワークモチーフプロファイルである．Clique (図 3 中のトライアッド 13) に着目すると，Blog，Flickr，Email いずれにおいても 3 人のユーザがお互いにコミュニケーションを取り合っていることが多い．また two mutual dyads (図 3 中のトライアッド 8) に着目すると，Email は他の 2 つのメディアに較べて 2 人と相互にインタラクションしているユーザは存在するがその 2 人に直接インタラクションがないことが多い．これについては，例えば社内のメールでは上司と部下は連絡を取り合うが部下の間では連絡を取り合わないというような直観的な説明が可能である．

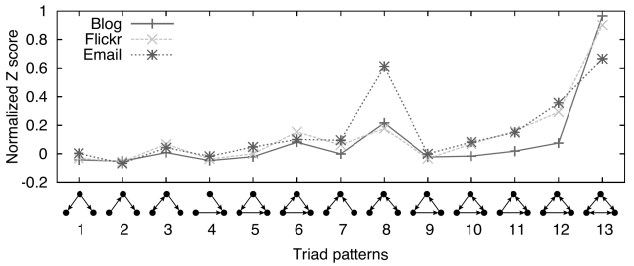

図 3　ソーシャルネットワークのネットワークモチーフプロファイル
Fig. 3　Network motif profiles for social networks.

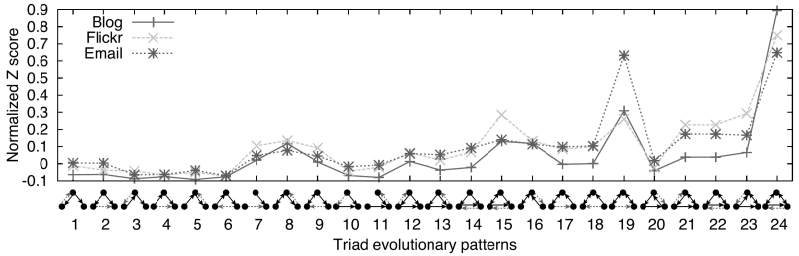

図 4　ソーシャルネットワークの動的ネットワークモチーフプロファイル
Fig. 4　Dynamic network motif profiles for social networks.

次に，図 4 は 3 つのソーシャルネットワークの動的ネットワークモチーフプロファイルである．今回は静的なネットワークに対して行った分析であったため，得られた動的ネットワークモチーフプロファイルがネットワークモチーフプロファイルとほとんど変わらないと

*1 http://www.flickr.com/services/api
*2 http://www.cs.cmu.edu/~enron

いう結果になった．したがって，時刻 t までのネットワーク $\boldsymbol{G}$ に対する時刻 t から $t+1$ までのエッジの発生を対象とした，動的なネットワークに対する分析をする必要がある．

4.3 動的ネットワークモチーフに基づくリンク予測

4.3.1 評 価 方 法

動的ネットワークモチーフのリンク予測精度を評価するために，10 分割交差検定を行った．すなわち，データセット中のエッジを 10 分割し，9 つをトレーニングセット $\boldsymbol{E}$ とし，残りの 1 つをテストセット $\boldsymbol{E}' - \boldsymbol{E}$ とした．

評価手順としては各リンク予測手法にて $\boldsymbol{V} \times \boldsymbol{V} - \boldsymbol{E}$ 中の各ノードペア $\langle u, v \rangle$ に対してエッジが発生する確率 $P(y = 1|u, v)$ を算出し，上位 $\left|\boldsymbol{E}' - \boldsymbol{E}\right|$ 件のノードペア集合 $\hat{\boldsymbol{\Upsilon}}$ を得る．

評価尺度は Liben-Nowell ら[6] と同様で，以下のように各リンク予測器の出力中の正答率をまったくランダムなリンク予測の正答率の理論値と比較することで得られる．

$$\text{Performance} = \frac{\hat{\boldsymbol{\Upsilon}} \cap (\boldsymbol{E}' - \boldsymbol{E})}{\hat{\boldsymbol{\Upsilon}}} \Big/ \frac{\boldsymbol{V} \times \boldsymbol{V} \cap (\boldsymbol{E}' - \boldsymbol{E})}{\boldsymbol{V} \times \boldsymbol{V}}. \tag{13}$$

4.3.2 比 較 手 法

以下に示す 7 つの手法を比較し，提案手法の有効性を示す．

(1) *Our method.* 提案手法．

$$P(y = 1|u, v\rangle = P(y = 1|\boldsymbol{x}_{uv}), \tag{14}$$

右辺は式 (6) によって計算する．今回の実験では扱う部分ネットワークのノード数は 3 とし，スムージングパラメータは $\delta = 1$ とした．

(2) *Common neighbors.* 共通隣接ノード数．NewMan ら[9] は 2 ノードの近接性をそれらに共通して隣接するノード数で与えた．

$$P(y = 1|u, v\rangle \propto \left|\Gamma_u \cap \boldsymbol{\Gamma}_v\right|, \tag{15}$$

ただし $\boldsymbol{\Gamma}_u = \{w \in \boldsymbol{V} | \langle u, w\rangle \in \boldsymbol{E}\} \cup \{w \in \boldsymbol{V} | \langle w, u\rangle \in \boldsymbol{E}\}$，すなわち u に隣接するノード集合を意味する．

(3) *Jaccard.* Liben-Nowell ら[6] は 2 ノード間の近接性を共通隣接ノードの数ではなくジャカード係数により与えることで，次数 (隣接するノード数) が小さいノードであってもリンク予測が可能な手法を提案した．

$$P(y = 1|u, v) \propto \frac{\left|\boldsymbol{\Gamma}_u \cap \boldsymbol{\Gamma}_v\right|}{\left|\boldsymbol{\Gamma}_u \cup \boldsymbol{\Gamma}_v\right|}. \tag{16}$$

(4) *Adamic/Adar.* 次数 (隣接するノード数) が大きいノードはいずれのノードとも隣接している可能性が高い．そこで，Adamic と Adar[1] は各共通隣接ノードにその次数にペナルティを与えることでより高精度なリンク予測器を実現した．

$$P(y = 1|u, v) \propto \sum_{w \in \boldsymbol{\Gamma}_u \cap \boldsymbol{\Gamma}_v} \frac{1}{\log\left|\boldsymbol{\Gamma}_w\right|}. \tag{17}$$

(5) *Preferential attachment.* 優先的選択指標．Barabasi ら[2] と Newman ら[9] によって提案された手法で，ネットワークのスケールフリー性から導かれる「次数が多いノードほど隣接しやすい」という仮定に基づいている．

$$P(y = 1|u, v) \propto \left|\boldsymbol{\Gamma}_u\right| \cdot \left|\boldsymbol{\Gamma}_v\right|. \tag{18}$$

(6) *Katz.* 2 ノード間にパスが存在すればするほど，それらは近接していると考えられる．また，それらのパス長が短ければ短いほど近接している．そこで Katz[5] は 2 ノードの近接性を，それらの間のパスについて長さで指数的に減衰する重みの総和で与えた．

$$P(y = 1|u, v) \propto \sum_{l=1}^{\infty} \beta^l \left|\boldsymbol{P}_{uv}^{\langle l\rangle}\right|, \tag{19}$$

ただし $\boldsymbol{P}_{uv}^{\langle l\rangle}$ は u から v への長さ l のパスの集合を意味する．重み β は $\{0.05, 0.005, 0.0005\}$ の 3 つの候補から，トレーニングセットを使ってさらに交差検定を行い，リンク予測の performance を最大にするものを選択した．

(7) *Random walk with restart.* Tong ら[12] が採用している手法で，PageRank の概念に基づいた手法である．u から出発して，$1 - \alpha$ の確率で今いるノードに隣接するいずれかのノードに移り α の確率で u に戻るというランダムウォークを繰り返す．このときの v の滞在確率を $P(y = 1|u, v)$ とみなす．α は $\{0.01, 0.05, 0.15, 0.30, 0.50\}$ の 5 つの候補から，トレーニングセットを使ってさらに交差検定を行い，リンク予測の performance を最大にするものを選んだ．

4.3.3 結　　果

表 2 に比較手法のリンク予測の性能を示す．太字は各ネットワークに対するもっとも高いリンク予測精度を表している．Blog と Email においては，提案法は他の比較手法よりも統計

表 2　ソーシャルネットワークに対する比較手法の性能
Table 2　Performance of compared methods for social networks.

Predictor		Blog	Flickr	Email
Probability that a random prediction is correct		0.052%	0.136%	0.009%
Our method	conditional independence	792.2	74.2	391.5
	multi-variate Bernoulli	**816.2**	74.8	**409.6**
	multinomial, EF attributes	635.4	83.9	101.3
Common neighbors		279.6	85.2	234.4
Jaccard's coefficient		69.0	77.8	8.0
Adamic/Adar		289.5	**85.5**	213.9
Preferential attachment		61.5	76.1	72.9
Katz	$\beta = 0.05$	156.4	51.1	148.7
Random walk with restart	$\alpha = 0.01$	76.0	7.5	318.8

的に有意に高精度であった (符号検定，$p < .01$)．一方で Flickr においては Adamic/Adar が最も高精度であったものの，提案法の精度も比較的高いという結果になった．このように，動的ネットワークモチーフに基づくリンク予測は有効であることがわかった．

また提案法に着目すると，ベルヌーイ試行の仮定に基づく手法はどのネットワークに対しても有効だが，多項分布の仮定に基づく手法は Email のような疎なネットワークでは有効ではないが Blog，Flickr の密なネットワークで有効であることがわかった．

5. まとめと今後の課題

動的ネットワークモチーフというネットワークの局所構造の特徴的な成長に着目し，複雑ネットワークを分析する手法を提案した．ソーシャルネットワークに提案法を適用した結果，提案法の結果はネットワークモチーフ分析の結果とあまり変わらなかった．これは，実験の設定が静的なネットワークに対する分析になっていたことが原因であった．

また，動的ネットワークモチーフのリンク予測への応用をベイズの定理により実現した．本稿ではソーシャルネットワークを用いた実験を通じて，提案法が従来法よりも高精度にリンク予測できることを示した．

本稿では静的なネットワークに対してのみ分析を行ったが，時間とともに成長するような動的なソーシャルネットワークに適用し分析を行う必要がある．具体的には，SNS におけるユーザ間のつながりが時間とともにどう成長していくのか提案法により分析を行うことを検討している．また，今回は 3 ノードに関する成長パターンに着目して分析を行ったが，4 ノード以上の部分ネットワークに関する分析も行っていく予定である．

参 考 文 献

1) Adamic, L. and Adar, E.: Friends and neighbors on the web, *Social Networks*, Vol.25, No.3, pp.211–230 (2003).
2) Barabási, A., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T.: Evolution of the social network of scientific collaborations, *Physica A: Statistical Mechanics and its Applications*, Vol.311, No.3–4, pp.590–614 (2002).
3) Domingos, P. and Pazzani, M.: On the optimality of the simple Bayesian classifier under zero-one class, *Machine Learning*, Vol.29, No.2, pp.103–130 (1997).
4) Inokuchi, A. and Washio, T.: A fast method to mine frequent subsequences from graph sequence data, *Proceedings of the 8th IEEE International Conference on Data Mining*, pp.303–312 (2008).
5) Katz, L.: A new status index derived from sociometric analysis, *Psychometrika*, Vol.18, No.1, pp.39–43 (1953).
6) Liben-Nowell, D. and Kleinberg, J.: The link-prediction problem for social networks, *Journal-American Society for Information Science and Technology*, Vol.58, No.7, pp.1019–1031 (2007).
7) Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. and Alon, U.: Superfamilies of evolved and designed networks, *Science*, Vol.303, No.5663, pp.1538–1542 (2004).
8) Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network motifs: simple building blocks of complex networks, *Science*, Vol.298, No.5594, pp.824–827 (2002).
9) Newman, M.: Clustering and preferential attatchment in growing networks, *Physical Review E*, Vol.64, No.2, p.025102 (2001).
10) Nijssen, S. and Kok, J.: A quickstart in frequent structure mining can make a difference, *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.647–652 (2004).
11) Schneider, K.: On word frequency information and negative evidence in Naive Bayes text classification, *Proceedings of the 4th International Conference on Advances in Natural Language Processsing*, Berlin, Heidelberg, pp.474–485 (2004).
12) Tong, H., Faloutsos, C. and Pan, J.: Random walk with restart: fast solutions and applications, *Knowledge and Information Systems*, Vol.14, No.3, pp.327–346 (2008).
13) Yan, X. and Han, J.: gSpan: Graph-based substructure pattern mining, *Proceedings of the 2002 IEEE International Conference on Data Mining*, pp.721–724 (2002).
14) 甲谷　優，川島晴美，藤村　考：QA コミュニティの成長パターンに基づく回答者への質問推薦，日本データベース学会論文誌，Vol.8, No.1, pp.89–94 (2009).