

SNSの特徴を考慮に入れたネットワークモデルの評価 Evaluation of the Network Model Considering Feature of SNS

工藤 匡浩[†] 松澤 智史[‡] 武田 正之[‡]

Tadahiro Kudo Tomofumi Matsuzawa Masayuki Takeda

1. はじめに

近年、コンピュータ技術の発展によって膨大なデータを容易に扱えるようになり、現実のネットワークの数理解析が活発に行われている。経済の物流システムや伝染病の伝播、生態系など現実社会の現象の背後にはネットワークが関わっており、ネットワークの構造が与える影響について多くの研究がされている。

複雑ネットワークの分野では、現実のネットワークの性質を持つネットワークモデルが提案され、ネットワークモデルを用いることにより、分析が困難である現実のネットワークの現象に対して新しい発見を得ることが期待されている。

また、SNS (Social Networking Service) やブログが Web 上のコミュニケーションサービスとして注目を集め、SNS を対象にしたネットワーク分析が行われ、SNS のユーザが形成するネットワークを表す様々なネットワークモデルが提案されている。SNS のネットワーク生成原理の理解は限定的ではあるが現実の人間関係を表現する可能性がある。

現在 SNS のネットワークを表現するモデルはいくつか提案されているが、従来の多くの SNS は「紹介制」と呼ばれる特徴を持つ。しかしながら、近年「登録制」と呼ばれる形態の SNS が増加し、従来のネットワークモデルでは、この登録制 SNS のネットワークを表現するには不十分である可能性がある。そこで本論文では、既存のネットワークモデルに対して実在する紹介制、登録制の SNS のネットワークを比較し、新たに登録制のネットワークモデルを表現する新しいモデルを提案し、評価する。

2. SNS とは

SNS とは、社会的ネットワーク (友人関係など) を Web 上で表現するサービスであり、ユーザは SNS を利用してお互いに交流を行う。代表的な SNS は、世界で最もユーザ数が多い MySpace や Facebook、日本においては、mixi や GREE などがある。

日本では、2004 年に mixi、GREE がオープンし、特に最もユーザ数が多い mixi は、2004 年 5 月に 1 万人、2005 年 8 月に 100 万人、2009 年 3 月には 1630 万人を超えるスピードで広まっている。

また、ウェブトラフィックを計測している Alexa Internet によると、My Space や Facebook は世界で人気のあるサイトとして共に世界 20 位以内にランキングされている*。

[†] 東京理科大学大学院 理工学研究科 情報科学専攻

Dept. of Information Sciences, Graduate School of Science and Technology Tokyo University of Science

[‡] 東京理科大学 理工学部 情報科学科

Dept. of Information Sciences, Tokyo University of Science

2.1 SNS の特徴

SNS では友人登録機能があり、友人関係を結ぶ際に互いの承認を必要とする場合と、承認を必要とせず一方的に結ぶ場合がある。友人登録機能によりユーザ同士の間には友人関係のリンクが形成される。本論文では、友人関係を重みの無い無向グラフとして取り扱う。

また、SNS には既に参加しているユーザから招待されない限り SNS に参加できない紹介制の SNS、招待を受ける必要がなく SNS に登録さえすれば参加できる登録制の SNS がある。図 1 のように友人関係をリンクとして表すと、紹介制の SNS では新しく SNS に参加するユーザは招待したユーザとリンクを結んだ状態で参加する。一方、登録制 SNS では誰ともリンクを結んでいない状態で SNS に参加する。

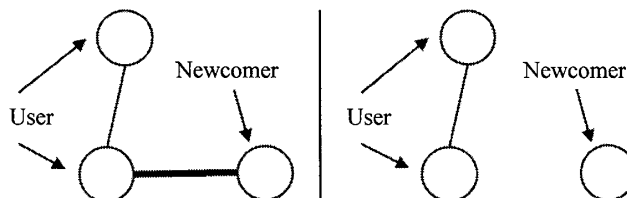


図 1: 紹介制 SNS (左) と登録制 SNS (右) の例

3. 関連研究

3.1 CNN モデル

Vazquez が提案した CNN (Connecting Nearest Neighbor) モデルは、成長するネットワークモデルの一部である[1]。CNN モデルは、新しい友人は今の友達の友達である可能性が高いという傾向をモデル化したものであり以下の手順で作成する。

- 確率 $1-p$ で新しいノードをネットワークに追加する。この際、既存のノードをランダムに 1 つ選びリンクで結び、潜在リンクを作成する。
- 確率 p で 1 つの潜在リンクをランダムに選び実リンク化する。

潜在リンク (Potential Edge) とは、新規ノードを追加する際に選ばれた既存のノードと既につながっているノードと新規ノードをつなぐリンクである。図 2 においてノード D を新規ノードとしノード A を選ばれた既存ノードとすると、ノード D はノード A とリンクを結ぶ。このとき、ノード A と既につながっているノード B、ノード C との

* 2010 年 4 月現在

間に結ぶリンクを潜在リンクと呼ぶ。

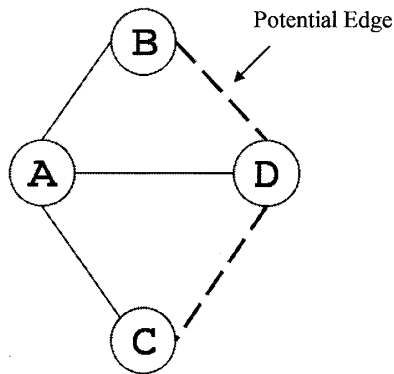


図2: 潜在リンク

3.2 CNNR モデル

Vazquez は CNN モデルを提案した際に、社会ネットワークには潜在リンクを顕在化する要素の他に、ランダムな結合があることを指摘していたがモデルには組み込んでいなかった。そこで湯田らは、CNN モデルにランダムな結合をとり入れた CNNR (CNN model with Random linkage) モデルを提案した[2]。CNNR は mixi のネットワーク成長を表したネットワークモデルであり、以下の手順で作成される。

- 確率 $1-p$ で新しいノードをネットワークに追加する。この際、既存のノードをランダムに 1 つ選びリンクで結び、潜在リンクを作成する。
- 確率 p で以下のうちの 1 つの操作を行う。
 - 確率 $1-q$ で 1 つの潜在リンクをランダムに選び実リンク化する。
 - 確率 q でランダムに選んだ 2 つのノードをリンクで結ぶ。

確率 q はランダムな結合を行う確率であり $q=0$ のとき CNN モデルとなる。

3.3 PFT モデル

PFT モデルは、鳥海らが SNS を表現するネットワークモデルとして提案したモデルである[3]。PFT モデルは以下の 3 つのネットワークモデルを組み合わせで作成する。

- 適応度モデル
- CNN モデル
- TF (トライアドフォーメーション) モデル

PFT モデルでは、適応度モデルを「SNS に適応したユーザが積極的に友人を招待する」要素、CNN モデルを「友人の友人と友人になる」要素、TF モデルを「招待者の友人とすぐに友人関係を結ぶ」要素としている。

PFT モデルはモデルをコントロールするパラメータとして優先的近傍接続確率 P_{pc} とトライアドフォーメーション確率 P_{tr} の 2 つのパラメータを持ち以下の手順で作成する。

- 確率 p で新しいノードをネットワークに追加する。新規ノードとつながれるノードはノードの次数と適応度に比例した優先的选择によって選択される。
- 被選択ノードと隣接するすべてのノードと新規ノードとの間に潜在リンクを作成する。
- 確率 P_{tr} でトライアドフォーメーションを行う。
- 確率 $1-p$ でリンクの作成を行う。
 - 確率 P_{pc} で優先的結合によりノードを選択し、そのノードが持つ潜在リンクの中から 1 つランダムに選び実リンク化する。
 - 確率 $1-P_{pc}$ でネットワーク内の潜在リンクからランダムに選択し実リンク化する。

それぞれのノード v_i が独立に確率密度 $P(\eta_i)$ に従って選ばれた適応度 η_i を持っているとするとき、ネットワーク中のノード数を N 、ノード v_i の次数を k_i として、優先的選択によりノードが選ばれる確率 $\Pi(v_i)$ は

$$\Pi(v_i) = \frac{\eta_i k_i}{\sum_{j=1}^N \eta_j k_j}$$

として表される。

トライアドフォーメーションとは新しく追加されたノードがネットワーク中に既に存在するノードを選びリンクを結ぶ際に、被選択ノード v_i の隣接ノードをランダムに m_i ($\leq k_i$) 個選びリンクを結ぶ操作である。

3.4 既存ネットワークモデルの問題点

これらの研究では SNS の特徴である「紹介制」や「登録制」の性質は考慮されておらず比較された SNS はすべて「紹介制」のものであった。また、登録制の SNS ではユーザが SNS に参加した時点では誰とも友人関係が結ばれていないため、ユーザ数がリンク数より多くなる可能性が有るが、既存のネットワークモデルでは SNS に新規参加するノードは既存のノードとリンクを結ぶため、登録制の SNS を表すことはできないと考えられる。

4. 提案手法

そこで、ノードの数がリンクの数より多くなるような場合を考慮に入れた、PFT モデルを拡張したモデル、CNN モデルを拡張したモデル、の 2 つを考える。

4.1 拡張 PFT モデル

PFT モデルを拡張したモデルを拡張 PFT モデルと呼び、以下の手順でネットワークを作成する。

- 確率 p で新しいノードをネットワークに追加する
 - 確率 P_{link} で優先的選択により既存ノードを選び新規ノードとリンクを結ぶ。また、被選択ノードと隣接するすべてのノードと新規ノードとの間に潜在リンクを作成する。
 - 確率 $1-P_{link}$ でどのノードともリンクを結ばない。
- 確率 $1-p$ でリンクの作成を行う。

- 確率 P_{pc} で優先的結合によりノードを選択し、そのノードが持つ潜在リンクの中から1つランダムに選び実リンク化する。
- 確率 $1-P_{pc}$ でネットワーク内の潜在リンクからランダムに選択し実リンク化する。

次数0のノードは優先的選択確率が0となってしまうため、次数0のノードは優先的選択確率では次数1として計算をする。 $P_{link}=1$ のとき、 $P_{tr}=0$ の PFT モデルとなる。

また、鳥海らの実験結果[3]から、トライアドフォーメーションを行う確率が0に近いとき、実 SNS に一番近い生成モデルとなったので、拡張 PFT モデルではトライアドフォーメーションを行わないものとする。

4.2 拡張 CNN モデル

既存のどのネットワークモデルに対しても次数0のノードをネットワーク中に増やすことのみで登録制 SNS をより表すことができるかを調べるために CNN モデルにも拡張を行う。CNN モデルを拡張したモデルを拡張 CNN モデルと呼び、以下の手順でネットワークを作成する。

- 確率 $1-p$ で新しいノードをネットワークに追加する。
- 確率 P_{link} で既存のノードをランダムに1つ選びリンクで結び、潜在リンクを作成する。
- 確率 $1-P_{link}$ でどのノードともリンクを結ばない。
- 確率 p で1つの潜在リンクをランダムに選び実リンク化する。

5. 実験

5.1 実験目的と方法

本論文では、SNS にはどのような生成原理が存在するか、また、既存のネットワークモデルがどの程度「紹介制」や「登録制」などの SNS の特徴を表現できるかを実 SNS と比較することによって確認する。実 SNS は mixi(紹介制)、So-net SNS 中の1つの SNS*(紹介制)、Any**(登録制)を対象とした。以後、So-net SNS 中の1つの SNS を So-net SNS と呼ぶ。

mixi の場合は、SNS に既に参加しているユーザから招待を受けて参加した新規ユーザは、招待者と友人関係を結ばれた状態で参加する。So-net SNS は管理人のみが招待する権利があり、新規ユーザは管理人と友人関係が最初に結ばれている。一方、「登録制」である Any では SNS に参加した時点では誰も友人関係が結ばれていない。また、友人関係を結ぶ際は3つの SNS はすべてユーザ同士の承認が必要となる。

それぞれの SNS においてユーザの SNS 上の友人関係のネットワークを取得する。今回、取得したネットワークは、mixi ではネットワーク全体の約 20%分であるノード数 3701697 個、リンク数 10843278 本、So-net SNS ではネットワーク全体のノード数 5187 個、リンク数 15555 本、Any ではネットワーク全体のノード数 170350 個、リンク

数 101922 本取得した。

5.2 比較するネットワークモデル

今回、比較するネットワークモデルは CNN モデル、CNNR モデル、PFT モデルを用いた。Any では拡張 PFT モデル、拡張 CNN の比較も行う。また、CNNR モデルのリンクを作成時にランダムな結合をする確率 q 、PFT モデルの優先的結合確率 P_{pc} 、トライアドフォーメーションを行う確率 P_{tr} は $[0.0, 1.0]$ の範囲を 0.1 間隔で変化させる。PFT モデルの適応度 η の分布は $0.0 \leq \eta \leq 1.0$ の一様分布とする。また CNN モデル、CNNR モデル、PFT モデルのリンクの追加を行う確率 p は、実 SNS のノード数 N とリンク数 M とすると $p = N/(N+M)$ として値を設定した[2]。So-net SNS では $p=0.25$ 、Any では $p=0.63$ 、mixi では $p=0.25$ となる。拡張 CNN モデル、拡張 PFT モデルのパラメータ P_{link} は、(Any の次数0のノード数) / (Any の全ノード数) から $P_{link}=0.73$ とした。

5.3 類似判定に用いる指標

モデルによって生成したネットワークと、実 SNS が類似しているか否かを判断するために分散を考慮したユークリッド距離 D を用いる。

分散を考慮したユークリッド距離 D は、モデルを用いて生成したネットワーク、実データのネットワークの間の距離 D を平均パス長 $L(=x_1)$ 、クラスタリング係数 $C(=x_2)$ 、次数相関 $r(=x_3)$ 、ベキ指数 $\gamma(=x_4)$ を用いて以下の式で定義される。

$$D = \sqrt{\sum_i \frac{(x_i - y_i)^2}{\sigma_i^2}}$$

x_i は実 SNS の各指標の値であり、 y_i はモデルによって生成したネットワークの各指標の値である。

また、 σ_i はモデルによって生成したネットワーク群におけるそれぞれの指標の標準偏差を表す。これは、4つの指標のスケールが異なるため、正規化を行うための係数である。

CNN モデルと拡張 CNN モデル、PFT モデルと拡張 PFT モデルの違いは「新規ノードを追加するとき既存ノードとリンクを結ぶか」、「どのノードともリンクを結ばずに追加するか」の分岐が存在するという点である。ここから、登録制 SNS ではノード数とリンク数の比率を近づけることにより距離 D が小さくなるかを調べることができる。また、拡張 PFT モデルは拡張 CNN モデルにノードとリンクを追加する際に優先的選択の要素を追加したものである。そのため、ノードやリンクの優先的選択が登録制 SNS において距離 D を小さくする要素であるかを調べることができる。

5.4 実験結果

CNNR モデルは q のパラメータを変化させ一番距離が小さくなった生成モデル、PFT モデルは P_{tr} 、 P_{pc} を変化させたとき一番距離が小さくなった生成モデルを示す。

ここで、 r を次数相関、 D を距離とする。

*ソネットエンタテインメント株式会社が提供する SNS 無料作成サービスにより作成された SNS である。

(2010年6月3日サービス終了予定)

**株式会社 Any が運営する SNS である。

<http://pc.myany.jp/>

5.4.1 紹介制 SNS

表 1 に So-net SNS の次数相関の値と距離を示し、表 2 に mixi の次数相関の値と距離を示す。

表 1: So-net SNS の次数相関の値と距離

ネットワークモデル	r	D
So-net SNS	-0.20	-
PFT モデル($P_{if}=0, P_{pc}=70$)	-0.15	6.77
CNN モデル	0.11	25.3
CNNR モデル($q=0$)	0.18	30.9

表 1 より PFT モデルが一番距離が小さくなっていることが分る。

表 2: mixi の次数相関の値と距離

ネットワークモデル	r	D
mixi	-0.10	-
PFT モデル($P_{if}=0, P_{pc}=60$)	-0.13	31.0
CNNR モデル($q=0$)	0.20	44.0
CNN モデル	0.11	47.1

表 2 より PFT モデルが一番距離が小さくなっていることが分る。また CNNR モデルにおいて $q=0$ から CNNR モデルは CNN モデルと同じである。

5.4.2 登録制 SNS

表 3 に Any の次数相関の値と距離を示す。

表 3: Any の次数相関の値と距離

ネットワークモデル	r	D
Any	-0.08	-
拡張 PFT モデル($P_{if}=0, P_{pc}=30$)	-0.02	5.32
PFT モデル($P_{if}=0, P_{pc}=30$)	-0.05	12.0
CNNR モデル($q=100$)	0.29	19.8
CNN モデル	0.28	26.7
拡張 CNN モデル	0.99	60.7

表 3 より拡張 PFT モデルが一番距離が小さくなっていることが分るが、拡張 CNN モデルが一番距離が大きくなっている。

6. 考察

6.1 既存ネットワークモデルの評価

3つの SNS ではユーザがもつ友人関係の数には負の相関がみられたが、負の次数相関を再現できたのは PFT モデルモデルだけであった。生成した PFT モデルは CNN モデルや CNNR モデルなどの他の生成モデルよりも実 SNS との距離が小さい値となっていることから「SNS に適応したユーザが積極的に SNS に人を招待する」という要素が SNS に存在している可能性を示唆している。実験の結果、紹

介制 SNS では PFT モデルが一番実 SNS に近いネットワークモデルだといえるが、トライアドフォーメーションを実施する確率が 0 のときの生成モデルが、今回調査したものの SNS に対しても近いネットワークモデルとなっているので「SNS に参加する際に招待者の友達とすぐに友人関係を結ぶ」という要素が SNS の特徴として必要か今後検討する必要がある。

6.2 提案手法の評価

登録制 SNS において、拡張 PFT モデルでは負の次数相関を表現することができ、距離 D も一番小さい値になっていることから拡張 PFT モデルは既存のネットワークモデルよりも登録制 SNS をよく表しているといえる。拡張 CNN モデル、拡張 PFT モデル以外のどのネットワークモデルでも、SNS に新規参加するノードは既存のノードとリンクを結ぶため、登録制の SNS を表すことはできず、この SNS の特徴を表すネットワークモデルは存在しない。しかし、ノード数がリンク数より多いことを表現できる拡張 CNN モデルは CNN モデルよりも距離 D が大きくなっている(表 3)。このことから、登録制の SNS を表す際、次数 0 のノードをネットワーク中に加えノード数とリンク数の比率を実 SNS に近づける操作が、ネットワークの距離 D を常に小さくするものとはいえないことが分かった。拡張 CNN モデルでは CNN モデルより距離 D が大きくなり、拡張 PFT モデルでは PFT モデルより距離 D が小さくなったことから、登録制の SNS には優先的選択の要素が関わっていることが考えられる。

7. まとめ

本論文では紹介制と登録制の SNS を 3 つ用いて、現在提案されているネットワークモデルの比較を行った。結果、PFT モデルが一番 SNS を表現できているモデルといえた。

しかし、どのネットワークモデルでも、SNS に新規参加するノードは既存のノードとリンクを結ぶため、登録制の SNS を表すことはできず、SNS の特徴を表すネットワークモデルは存在しなかった。そこで、PFT モデルに拡張を加えたモデルを提案し、この提案手法で生成したネットワークが他のネットワークモデルよりも実際の登録制 SNS のネットワークに近いことを示した。

参考文献

- [1] A. Vazquez, "Growing network with local rules: preferential attachment, clustering hierarchy and degree correlations", *Physical Review E*, Vol.67, 056104 (2003).
- [2] K. Yuta, N. Ono, Y. Fujiwara, "A Gap in the Community-Size Distribution of a Large-Scale Social Networking Site", *arXiv: physics/0701168v2* (2007).
- [3] 鳥海 不二夫, 石田 健, 石井 健一郎, "SNS におけるネットワーク成長のモデル化", 第 23 回人工知能学会全国大会, 2E1-OS5-2(2009).