

SNS におけるネットワーク成長のモデル化

Proposal of network growth model on Social network services

鳥海 不二夫*¹ 石田 健*¹ 石井 健一郎*¹

Fujio Toriumi

Ken Ishida

Kenichiro Ishii

*¹名古屋大学 大学院情報科学研究科

Nagoya University

As a part of the steady growth of new network communication tools, the expansion of Social Network Services is becoming a phenomenon impacting societies all over the world. Along with the increased interests in SNSs, their network structures are receiving a lot of attention. In this paper, we proposed a growth model of SNS networks. We performed numerical simulation to verify the validity of the proposed model. We compared networks generated by the proposed model with data from two actual SNSs. We also generated networks by the proposed model, and compared it with actual data sets. Results of the simulations showed that the proposed model is able to represent the network structures of the actual SNSs.

1. はじめに

近年、SNS 上でユーザが形成するネットワークへの関心が高まっており、SNS ネットワークを表現する様々なモデルが提案されている [9, 10]。しかし、これまでに SNS 固有の特徴に着目したモデルは提案されていない。そこで本研究では、SNS の特徴を表現可能な SNS ネットワークの成長モデルを提案する。複数 SNS の実データを用いた検証シミュレーションにより、提案モデルによって従来までのモデルよりも精度良く SNS を表現可能であることを確認する。

2. モデル化対象 SNS

本研究では、以下の 2 つの SNS を対象としてモデル化を行った。

- 西千葉地域 SNS 「あみっぴい*¹」
- 盛岡地域 SNS 「モリオネット*²」

これらの SNS に関するデータは、それぞれ NPO 法人 TRY-WARP および盛岡市情報企画室から提供を受けている。なお、本データを扱うにあたり個人情報の扱いには特に注意し、全てのデータから個人情報を除去したものが提供されている。

本研究では、提供を受けたデータのうち、以下の項目を用いる。

1. ユーザ (ユーザ ID, 参加日時)
2. ユーザ間リンク (リンク ID, リンク元ユーザ ID, リンク先ユーザ ID, 構築日時)

これらのデータを用いることで、SNS のネットワークがどのように成長したかについて時系列を追って分析することが可能である。

3. SNS 成長モデルの提案

3.1 概要

本章では、SNS の成長を表現するモデルを提案する。提案モデルでは、SNS のネットワーク構造に着目し、実際の SNS に類似したネットワークの生成を目的とする。提案モデルは、既存のネットワークモデルを拡張したモデルである。そのため、まずはいくつかの既存のネットワークモデルについて概要を述べた後、提案モデルの詳細について述べる。次節以降では、以下の 3 つのネットワークモデルについて解説する。

- Fitness モデル [3]
- CNN モデル [8]
- TF モデル [4]

3.2 Fitness モデル

Fitness モデルは、ネットワークに新たなノードが追加される際に、特定のノードが優先的にリンクを結びモデルである [2, 3]。ノード同士がリンクを形成する確率 Π_i は各ノードの「次数 k_i 」と「適応度 (fitness) η_i 」に比例し、

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \quad (1)$$

となる。ここで、次数 k_i はノードのリンク数を表し、適応度 η_i は「他のノードと次数を競う能力」を表す実数値である。このように適応度に重みをつけて確率的に接続ノードを決定することを、優先的選択と呼ぶ。

SNS においては、「SNS に適応したユーザが積極的に友人を招待する」ことに対応する。

3.3 Connecting nearest neighbor (CNN) モデル

CNN モデルは、「友達の友達は、友達になる」という考えに基づいたモデルである [8]。このモデルでは、「新規ノードの追加」、「新規リンクの作成」の 2 つの操作を確率的に繰り返すことによってネットワークを構築する。各操作が行われる確率は、パラメータ p によって制御し、その値は、生成するネットワークのノード数、リンク数に応じて決定される。

SNS においては、「友人の友人と友人になる」ことに対応する。

連絡先: 鳥海不二夫, 名古屋大学 大学院情報科学研究科, 愛知県名古屋市千種区不老町, 052-789-3812, tori@is.nagoya-u.ac.jp

*¹ <http://amippy.jp>

*² <https://sns.city.morioka.lg.jp>

3.4 Triad formation(TF) モデル

TF モデルは、ネットワーク形成に「トライアドフォーメーション」と呼ばれる操作を取り入れたモデルである [4]。

このモデルでは、新規ノードがネットワークに接続される際に、被選択ノードの隣接ノードからランダムに選んだ m_t 個のノードと、新規ノードとの間にリンクを作成する。これにより、新規ノード、被選択ノード、被選択ノードの隣接ノードの 3 ノード間にトライアド (三角形) が形成される。

SNS においては、「招待者の友人とすぐに友人関係を結ぶ」ことに対応する。

3.5 提案モデル

3.5.1 概要

本研究では、Fitness モデル, CNN モデル, TF モデルの 3 つのモデルの考えを組み合わせた上で独自の拡張を行った PFT モデルを提案する。

3.5.2 モデルパラメータ

本研究で提案する PFT モデルは、「新規ノードの追加」「新規リンクの作成」の 2 つの操作を確率的に繰り返すことによりネットワークを構築する。確率 p で新規ノードの追加、確率 $(1-p)$ で新規リンクの作成を行う。パラメータ p の値は、生成するネットワークのノード数、リンク数に応じて一意に決定される。また、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、以下の 2 つのパラメータを持つ。

- 優先的近傍接続確率 p_{pc} ($0 \leq p_{pc} \leq 1$)
- トライアドフォーメーション確率 p_{tf} ($0 \leq p_{tf} \leq 1$)

PFT モデルは、これらの 2 つのパラメータを変化させることにより、各 SNS 固有の特徴を表現可能である。

3.5.3 モデル詳細

図 1 に、PFT モデルによるネットワーク生成手続きのフローチャートを示す。ここでは、モデルの各手続きをフローチャートと対応させて解説する。以下の手続き中の (a), (b.1) 等の表記は、フローチャート内の表記と対応している。

- (a) 新規ノードの追加 (確率 p)
 - (a.1) 接続先ノードの選択
新規ノードとリンクで繋がれるノードを、ノードのリンク数と適応度に比例した優先的選択によって決定する。
 - (a.2) 潜在的リンクの作成
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「潜在的リンク」を作成する。
 - (a.3) トライアドフォーメーション
確率 p_{tf} でトライアドフォーメーションを行う。すなわち、被選択ノードの隣接ノードの中からランダムに選んだノードと、新規ノードとの間にリンクを作成する。
- (b) 新規リンクの作成 (確率 $(1-p)$)
 - (b.1) 優先的選択を行う場合 (確率 p_{pc})
優先的選択によってノードを選択し、そのノードが持つ潜在的リンクの中から 1 つをランダムに選択する。
 - (b.2) 優先的選択を行わない場合 (確率 $1-p_{pc}$)
ネットワーク内の「潜在的リンク」からランダムに 1 つ選択する。
 - (b.3) 選択された潜在リンクを実リンク化する

3.5.4 ネットワーク構築手順

PFT モデルを用いて実際にネットワークを生成する手順を以下にまとめる。

1. 生成したいネットワークのノード数 N , リンク数 M を決定する。
2. 優先的近傍接続確率 p_{pc} を決定する。 ($0 \leq p_{pc} \leq 1$)
3. トライアドフォーメーション確率 p_{tf} を決定する。 ($0 \leq p_{tf} \leq 1$)
4. 各パラメータを適用して、3.5.3 の操作を繰り返す。
5. ノード数が N になった時点で終了する。

4. 妥当性検証シミュレーション

4.1 シミュレーション目的

本章では、PFT モデルを用いたシミュレーションを行い、モデルの妥当性を検証する。PFT モデルによって生成したネットワークと SNS の実データを比較することにより、PFT モデルが SNS のネットワーク構造を再現できることを確認する。なお、モデルによって再現する対象 SNS として複数の実データを用いることにより、PFT モデルが多様な SNS を再現可能か否かを確認する。

4.2 シミュレーション手法

本シミュレーションでは、PFT モデルが SNS のネットワーク構造を再現可能であるか否かを、モデルのパラメータ p_{tf} , p_{pc} を変化させたシミュレーションにより確認する。パラメータを離散的に変化させ、取りうるすべての組み合わせについてシミュレーションを行う。なお、他モデルとの性能比較のため、過去の研究により SNS のネットワーク生成に適していると言われる「CNNR モデル [9]」との比較も行う。

4.3 評価指標

モデルによって生成したネットワークと、実データのネットワークが類似しているか否かを定量的に判断するために、ネットワーク間の距離を示す指標を用いる。

モデルを用いて生成したネットワーク G_i と、実データのネットワーク G_R の間の距離 D をクラスタリング係数 C [6], 平均経路長 L [6], 同類選択性 r [5, 7], ベキ指数 γ [1] を用いて以下の式で定義する。

$$D(G_i, G_R) = \frac{|L_i - L_R|}{\sigma_L} + \frac{|C_i - C_R|}{\sigma_C} + \frac{|r_i - r_R|}{\sigma_r} + \frac{|\gamma_i - \gamma_R|}{\sigma_\gamma} \quad (2)$$

ここで、 L_i, C_i, r_i, γ_i はモデルによって生成したネットワークの評価値であり、 L_R, C_R, r_R, γ_R は実データの評価値を表す。また、各項の分母 $\sigma_L, \sigma_C, \sigma_r, \sigma_\gamma$ 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、距離計算の式に用いる 4 つの指標のスケールが異なるため、各指標を正規化するための係数である。

$D(G_i, G_R)$ が十分に小さければ、 G_i, G_R は類似したネットワーク構造をもつと考えることができる。

4.4 シミュレーション条件

シミュレーション条件を表 1 に示す。まず、ノード数 N , リンク数 M は再現する SNS に合わせて設定する。各ノードの適応度 η は、0.0~1.0 の範囲の一様分布で設定する。また、トライアドフォーメーション確率 p_{tf} と優先的近傍接続確率 p_{pc} は 0.0~1.0 の範囲において 0.1 間隔で変化させる。各パラメータで 10 回シミュレーションを行い、その平均を結果とする。

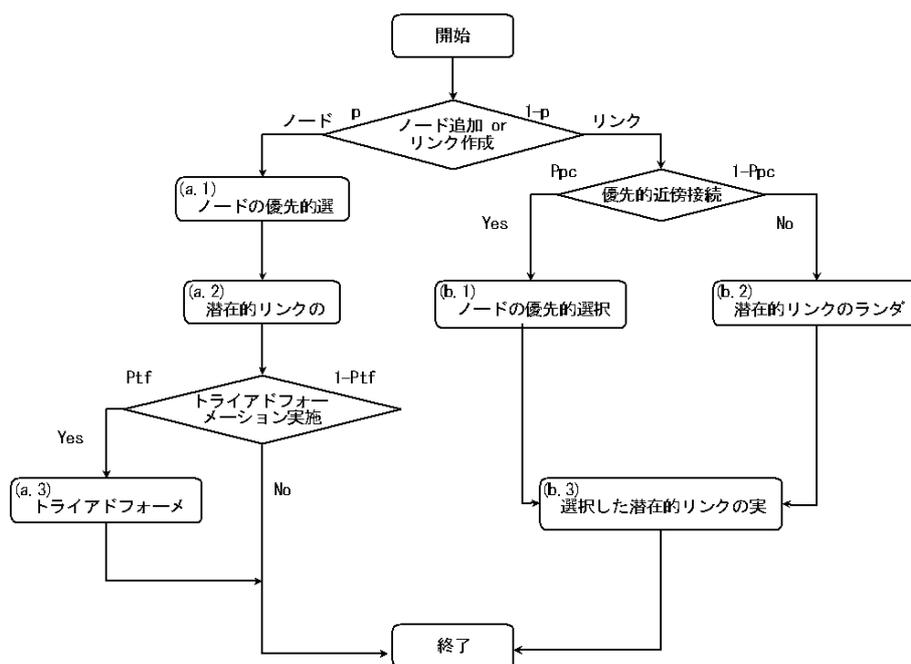


図 1: 提案モデルフローチャート

表 1: シミュレーション条件

ノード数 N , リンク数 M	実データと同様
適応度 η の分布	一様分布 ($0.0 \leq \eta \leq 1.0$)
トライアドフォーメーション確率 p_{tf}	$p_{tf} = \{0.0, 0.1, \dots, 1.0\}$
優先的近傍接続確率 p_{pc}	$p_{pc} = \{0.0, 0.1, \dots, 1.0\}$
パラメータの刻み	0.1
試行回数	10 回

4.5 ネットワーク再現シミュレーション結果

4.5.1 あみっぴいのネットワーク

提案モデルのパラメータ p_{tf}, p_{pc} を変化させて、ネットワークを生成した。シミュレーションの結果、実データとの距離 D が小さいネットワークを生成した上位 3 パラメータを表 2 に示す。

表 2: 上位 3 パラメータ (あみっぴい)

	p_{tf}	p_{pc}	距離 D
1	0.0	0.9	3.802
2	0.3	1.0	3.965
3	0.1	1.0	4.085

表 2 に示したパラメータによって生成されたネットワークの各種ネットワーク指標を表 3 に示す。ここで、PFT(1)~PFT(3) は、それぞれ表 2 における上位 1~3 位のパラメータによって生成されたネットワークである。

表 3 より、PFT モデルは実データに近い平均経路長、クラスタリング係数を再現できている。同類選択性については、わずかに実データよりも大きい値を示しているが、負の同類選択性を再現出来ている。一方で、CNNR モデルは負の同類選択性を表現できていない。これらより、PFT モデルによってあみっぴいの実データに近いネットワーク構造を再現可能であるといえる。

4.5.2 モリオネットのネットワーク

提案モデルのパラメータ p_{tf}, p_{pc} を変化させて、ネットワークを生成した。シミュレーションの結果、実データとの距離 D が小さいネットワークを生成した上位 3 パラメータを表 4 に示す。

表 4: 上位 3 パラメータ (モリオネット)

	p_{tf}	p_{pc}	距離 D
1	0.0	0.7	2.973
2	0.0	0.5	3.221
3	0.0	0.6	3.240

表 4 に示した 3 パラメータによって生成されたネットワークの各種ネットワーク指標を表 5 に示す。表 5 の PFT(1)~PFT(3) は、それぞれ表 4 における上位 1~3 位のパラメータによって生成されたネットワークである。なお、実データのネットワーク指標と、CNNR モデルによって生成したネットワークの指標もあわせて示す。

表 4 より、PFT モデルでは、負の同類選択性、高いクラスタリング係数、短い平均経路長を再現できており、どの値も実データに近い値を示していることがわかる。一方、CNNR モデルでは平均経路長の値が実際の値より大きく、負の同類選択性も再現出来ていない。以上の結果より、PFT モデルによって CNNR モデルよりもモリオネットの実データに近いネットワーク構造を再現可能であるといえる。

表 3: あみっぴいの実データとモデルによって生成されたネットワークの比較

	あみっぴい	CNNR	PFT(1)	PFT(2)	PFT(3)
平均経路長 L	3.10	5.28	3.09	3.06	3.09
クラスタリング係数 C	0.443	0.420	0.402	0.477	0.430
同類選択性 r	-0.255	0.234	-0.199	-0.196	-0.177
ベキ指数 γ	1.169	1.211	1.261	1.262	1.267
距離 D	-	25.73	3.801	3.965	4.085

表 5: モリオネットの実データとモデルによって生成されたネットワーク

	モリオネット	CNNR	PFT(1)	PFT(2)	PFT(3)
平均経路長 L	2.81	3.70	2.72	2.82	2.68
クラスタリング係数 C	0.415	0.446	0.463	0.436	0.470
同類選択性 r	-0.237	0.043	-0.212	-0.156	-0.241
ベキ指数 γ	1.023	-1.134	1.086	1.109	1.086
距離 D	-	14.13	2.973	3.221	3.240

以上のシミュレーションにより、提案モデルによって実 SNS のネットワーク構造を従来のモデルよりも高い精度で再現可能であることがわかった。

5. 結論

5.1 まとめ

本研究では、SNS 上でユーザが形成するネットワークの成長モデルを提案した。提案モデルを用いたシミュレーションにより、モデルの妥当性検証を行った。検証の結果、提案モデルによって SNS のネットワーク構造を表現可能であることを確認した。提案モデルを用いることで、SNS がこれまで成長してきた過程や、SNS の今後を予測することが可能となると期待される。

5.2 今後の課題

今後の課題としては、詳細な時系列データの分析を行うことが挙げられる。また、ネットワーク構造だけでなく、日記やコメントなどを通じて行われるユーザ間コミュニケーションに着目していくことも今後の課題である。本研究では SNS 上の友人ネットワークのみに着目したが、SNS 内でユーザが交換するコメントネットワークを分析することも考えられる。分析結果に基づいて、どのような運営が SNS の活性化に寄与するのかを明らかにし、SNS 運営者に対して適切な SNS 運営手法の提案を行うことも重要な課題である。

本研究ではシミュレーションの対象として地域 SNS を扱ったが、提案モデルはパラメータ設定の変更によって他の SNS への適用も可能であると考えられる。今後、他の SNS を対象としたシミュレーションを行い、モデルの妥当性及び一般性を検証する事により、さらに正確な分析が可能になると期待される。

謝辞

本研究を行う上で、あみっぴいのデータを提供していただいた NPO 法人 TRYWARP、NPO 法人 TRYWARP 代表理事 虎岩雅明氏、および、モリオネットのデータを提供いただいた、盛岡市情報企画室および岩手大学の五味壮平准教授、吉田等明准教授に感謝いたします。

参考文献

- [1] A.L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, Vol. 286, No. 5439, p. 509, 1999.
- [2] G. Bianconi and A.L. Barabási. Bose-Einstein Condensation in Complex Networks. *Physical Review Letters*, Vol. 86, No. 24, pp. 5632–5635, 2001.
- [3] G. Bianconi and A.L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, Vol. 54, No. 4, pp. 436–442, 2001.
- [4] P. Holme and B.J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, Vol. 65, No. 2, p. 26107, 2002.
- [5] MEJ Newman. Mixing patterns in networks. *Physical Review E*, Vol. 67, No. 2, p. 26126, 2003.
- [6] MEJ Newman. The Structure and Function of Complex Networks. *SIAM Review*, Vol. 45, No. 2, pp. 167–256, 2004.
- [7] MEJ Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, Vol. 68, No. 3, p. 36122, 2003.
- [8] A. Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, Vol. 67, No. 5, p. 56104, 2003.
- [9] K. Yuta, N. Ono, and Y. Fujiwara. A Gap in the Community-Size Distribution of a Large-Scale Social Networking Site. *Arxiv preprint physics/0701168*, 2007.
- [10] 三井一平, 内田誠, 白山晋. コミュニティ構造を有するネットワーク成長モデル. 情報処理学会研究報告, 2006-ICS-142(3), 2006.