

複数のコミュニティ構造を考慮した SNS の成長モデル

A Network Growth Model of Social Network Services with Multiple Communities

西郷泰広¹ 篠沢佳久¹

Yasuhiro Saigo¹, Yoshihisa Shinozawa¹

¹慶應義塾大学 理工学部 管理工学科

¹Faculty of Science and Technology, Keio University

概要: 本研究においては、SNS の成長過程において、現実の人間関係に近いコミュニティとネットワーク上での人間関係という形成過程の異なるコミュニティ構造を導入した SNS の成長モデルを提案する。そして実存の SNS と提案モデルによって生成したネットワークとの定量的分析を行なった結果、実存の SNS に近いネットワークを生成することができたことを示す。

Abstract: In this paper, we propose a network growth model of SNS. The communities play an important role in the formation of the networks. We adopt a network growth model with two type of the community structure. We generate the networks by our model and compare them with actual networks. The results of our experiments show that the networks generated by our model are similar to the actual networks.

1. はじめに

近年、コンピュータネットワークの発展によって、SNS (Social Networking Service) や Twitter など、新しいネットワークコミュニケーションツールが注目されている。本研究においては、SNS による友人ネットワークの成長過程を明らかにし、そのモデル化を考案することを目的とする。特にコミュニティ構造を考慮した上で、複数個のネットワーク成長モデルを組み合わせたモデルを提案する。

提案する成長モデルにおいては、SNS の成長過程を考慮して、ネットワークの成長を二つの段階に分けて考える。まず、第一段階においては、CNNR (Connecting Nearest Neighbors with Random linkage) モデルに管理者優先接続を加えたモデルによってネットワークを形成する。第二段階においては、ノード追加、リンク追加およびコミュニティ追加を確率的に行なうことによってネットワークを形成する。

特にコミュニティの追加においては、現実の人間関係に近いコミュニティとネットワーク上での人間関係という形成過程の異なるコミュニティ構造を導入することを提案する。

そして実存の SNS 「モリオネット」および「あみっぴい」を対象として、ネットワークを生成し、ネットワーク指標による定量的分析を行なった結果、提案モデルにおいては、実存の SNS に近いネットワークを生成することができることを示す。

2. 関連研究

SNS に関する研究はこれまで数多く行われてきた。現実社会のネットワークの性質を明らかにする分析手法の一つに複雑ネットワーク分析[1]が挙げられる。この複雑ネットワーク分析の手法とネットワーク上での情報の伝播という機能に着目し、SNS のネットワーク構造モデルを推定した研究[2]や、既存のネットワーク成長モデルを複数個組み合わせたモデルを用いてネットワークを形成し、どの組み合わせのネットワーク成長モデルが実存の SNS の友人ネットワークを最も精度良く再現したかを分析した研究[3]などがある。

このように SNS の成長過程を明らかにしようとする研究については、数多くの研究が存在するが、本研究においては、特に二つの既存研究[4][5]を基礎としている。まず既存研究[4]においては、ソーシャルネットワークは、人同士の繋がりだけでなく、人と社会的コミュニティとの関係によっても規定されるという前提のもと、コミュニティ構造を有するネットワークに対して、局所的、全体的な相互作用を考慮したネットワーク成長モデルを提案している。また既存研究[5]においては、3 種類のネットワークモデルを組み合わせ、SNS 固有の特徴が表現可能な SNS の成長モデルを提案している。

そこで本研究においては、これら二つの既存研究[4][5]の長所を生かし、複数のコミュニティ構造を考

慮した上で、ネットワークの成長を二段階に分け、複数個のネットワークモデルを組み合わせることによって、実存の SNS と性質上類似したネットワークを形成することを試みる。

3. ネットワークの成長モデル

3.1 提案モデルの概要

SNS の友人ネットワークを形成する過程においては、「新規ユーザーの追加」＝「ノードの追加」、「友人関係を構築」＝「リンクの作成」という二つの操作を繰り返していくことになる。この二つの操作の手順によって、さまざまな形態のネットワークを形成することが可能となる。

本研究においては、より SNS 固有の特徴を表現したネットワークを生成するため、既存研究[5][6]を基礎として、以下に示す複数個のネットワーク成長モデルを利用する。

- CNNR モデル
- 管理者優先接続モデル
- Fitness モデル
- TF (トライアドフォーメーション) モデル

さらには本研究においては、ネットワークの成長過程において、「コミュニティの追加」という操作も考慮する。既存の SNS のコミュニティについては、知人関係によるコミュニティと同じ趣味や思想を持つ人同士という二種類のコミュニティが存在する。知人関係によるコミュニティに属するユーザーについては、「同じようなリンクを張るノード」としてとらえることができる。一方で、同じ趣味や思想を持つ人同士のコミュニティに属するユーザーについては、趣味などが同じ人同士は、知人関係によってリンク接続される必要がないため、「異なったリンクを張るノード」としてとらえることができる。また、一人のユーザーが複数個のコミュニティに属することも多々ある。

そこで本研究においては、この二種類のコミュニティを導入することを試みる。知人関係によるコミュニティについては、「同じようなリンクを張るノード」という性質を持っている。そのため、類似したリンクを張っているノードを一つのコミュニティとしてとらえる。これについては、ノード間の接続関係を表現した隣接行列を対象として、k-means 法によるクラスタリングを行ない、その結果を利用する。

一方で、同じ趣味や思想の持つ人同士のコミュニティについては、「異なったリンクを張るノード」という性質を持つため、ランダムにノードを選択しコ

ミュニティを作成する。そして一つのノードが複数個のコミュニティに属することも可能とする。図 1 に提案モデルの概要を示す。

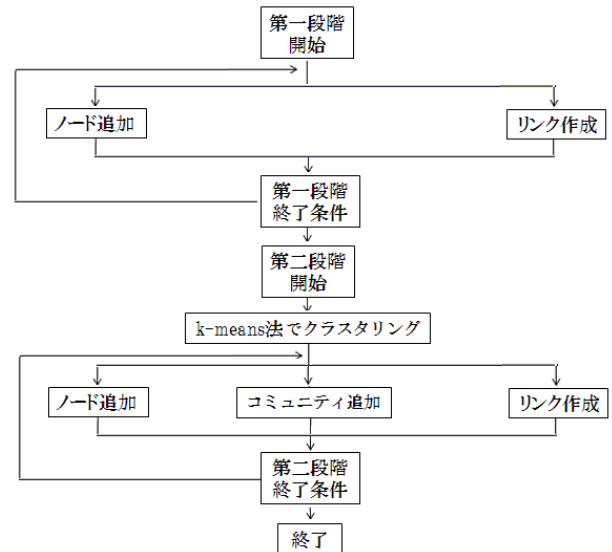


図 1 提案モデルの概要

図 1 に示すように提案モデルは二段階から構成される。第一段階は、コミュニティがまだ存在しないネットワークの初期状況を形成することを目的としている。第一段階においては、CNNR モデル[7]に管理者優先接続を加えたモデルを使ってネットワークの初期段階を形成する。管理者優先接続[6]とは、管理者とユーザーが優先されて接続されるモデルのことで、ネットワークの初期段階においては、新規ユーザーを増やすために管理者がユーザーを積極的に SNS に招待する仕組みを表したモデルである。管理者優先接続を導入することにより、管理者に招待されたユーザーは登録と同時に管理者と接続されるため、管理者にリンクが集中しやすくなる。このように管理者優先接続を CNNR モデルに導入し、ネットワークを形成することで、ネットワークの初期段階を形成するのに適したモデルではないかと考えた。

次に第二段階は、ネットワークが比較的大規模化し、複数個のコミュニティが存在する状況を形成することを目的としている。第二段階においては、「ノードの追加」「リンクの作成」に加えて、「コミュニティの追加」の操作を繰り返すことによって、ネットワークを形成する。

コミュニティについては、前述したように友人関係によるコミュニティと同じ趣味を持つ人同士のコミュニティという性質の異なるコミュニティを追加していくことで、より現実の SNS と類似したネットワークが形成されると考えた。

また「ノードの追加」「リンクの作成」についても、

第一段階とは異なり, CNNR モデルだけではなく, Fitness モデル, TF モデルといった複数のモデルを組み合わせることにより, ネットワークを形成する.

3.2 提案モデルの第一段階

提案モデルの第一段階におけるネットワーク成長の流れを図2に示す. 第一段階においては(1)新規ノードの追加, (2)新規リンクの追加をノード数が T 個となるまで繰り返す. 第一段階における各操作の詳細を以下に示す.

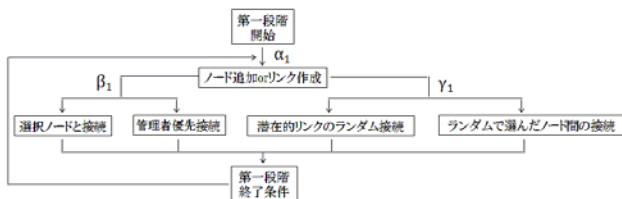


図2 第一段階におけるネットワーク成長の流れ

(1) 新規ノードの追加 (確率 α_1)

確率 α_1 で以下の新規ノードの追加を行なう. これは次の2つの操作 (1-1) (1-2) を, さらにそれぞれの確率にしたがって行なう.

(1-1) 選択ノード接続 (確率 β_1)

確率 β_1 でネットワークに新規ノード i を追加し, ネットワーク内からランダムに選択したノード j ($j \neq i$) と, 新規ノード i との間にリンクを作成する. さらに, ノード j のすべての隣接ノードと, 新規ノード i の間に「潜在的リンク」を作成する.

(1-2) 管理者優先接続 (確率 $1 - \beta_1$)

確率 $1 - \beta_1$ でネットワークに新規ノード i を追加し, 新規ノード i と管理者との間にリンクを作成する. さらに, 管理者のすべての隣接ノードと, 新規ノード i の間に「潜在的リンク」を作成する.

(2) 新規リンクの作成 (確率 $1 - \alpha_1$)

確率 $1 - \alpha_1$ で以下の新規リンクの作成を行なう. これは次の2つの操作 (2-1) (2-2) を, さらにそれぞれの確率にしたがって行なう.

(2-1) 潜在的リンクの実リンク化 (確率 γ_1)

確率 γ_1 で「潜在的リンク」の一つを選択し, 実リンク化する.

(2-2) ランダム接続 (確率 $1 - \gamma_1$)

確率 $1 - \gamma_1$ でランダムに選んだノード間にリンクを張る.

3.3 提案モデルの第二段階

ネットワーク成長の第二段階においては, まず k -means 法を用いて, 「友人関係のコミュニティ」を抽出し, 新規コミュニティとして追加する. そして (1) 新規ノードの追加, (2) 新規コミュニティの追加, (3) リンク作成を確率的に行ない, ノード数が S 個となるまで, ネットワークを成長させていく.

(1) 新規ノードの追加 (確率 α_2)

確率 α_2 で新規ノードの追加を行なう. これはさらに以下の操作 (1-1) (1-2) (1-3) (1-4) をそれぞれの確率にしたがって行なう. 新規ノードの追加の流れを図3に示す.

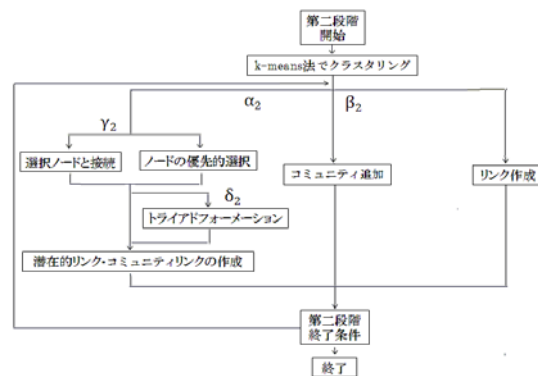


図3 第二段階における新規ノードの追加

(1-1) 選択ノード接続 (確率 γ_2)

確率 γ_2 でネットワークに新規ノード i を追加し, ネットワーク内からランダムに選択したノード j ($j \neq i$) と, 新規ノード i との間にリンクを作成する. 次に確率 δ_2 で操作 (1-3) を, 確率 $1 - \delta_2$ で操作 (1-4) を行なう.

(1-2) ノードの優先接続 (確率 $1 - \gamma_2$)

確率 $1 - \gamma_2$ でネットワークに新規ノード i を追加し, 新規ノード i と次数の高いノード j ($j \neq i$) を優先的に接続する (Fitness モデル). 次に確率 δ_2 で操作 (1-3) を, 確率 $1 - \delta_2$ で操作 (1-4) を行なう.

(1-3) トライアドフォーメーション

確率 δ_2 で新規ノード i と接続したノード j の隣接ノード k を接続し, トライアドフォーメーション[5]を実施する. 次に操作 (1-4) を行なう.

(1-4) 潜在的リンクおよび潜在的コミュニティリンク作成

操作 (1-1) で新規ノード i と接続したノード j の隣

接ノードと潜在的リンクを作成する。また新規ノード i と接続したノードが参加しているすべてのコミュニティとの間に潜在的コミュニティリンクを作成する。

(2) 新規コミュニティの追加 (確率 β_2)

確率 β_2 で新規コミュニティの追加を行なう。これはさらに以下の操作 (2-1) (2-2) (2-3) をそれぞれの確率にしたがって行なう。新規コミュニティの追加の流れを図 4 に示す。

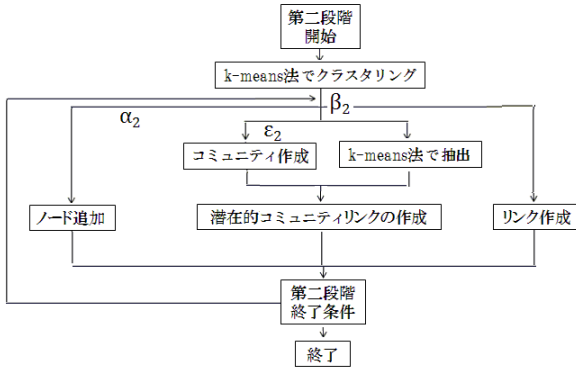


図 4 第二段階における新規コミュニティの追加

(2-1) ユーザーによるコミュニティ追加 (確率 ϵ_2)
確率 ϵ_2 でランダムに選択したユーザーが新コミュニティを追加し、 m 人のユーザーと新コミュニティとの間に実コミュニティリンクを作成する。次に操作 (2-3) を行なう。

(2-2) k-means 法によるコミュニティ抽出 (確率 $1 - \epsilon_2$)
確率 $1 - \epsilon_2$ で k-means 法を用いてコミュニティ抽出を行ない、新たに k 個の新コミュニティを追加する。操作 (2-3) を行なう。

(2-3) 新コミュニティに参加した各ユーザーのすべての隣接ノードと新コミュニティとの間に潜在的コミュニティリンクを作成する。

(3) リンク作成 (確率 $1 - \alpha_2 - \beta_2$)

確率 $1 - \alpha_2 - \beta_2$ で以下の新規リンクの作成を行なう。これはさらに次の操作 (3-1) (3-2) (3-3) をそれぞれの確率にしたがって行なう。新規リンクの追加の流れを図 5 に示す。

(3-1) 潜在的リンクの実リンク化 (確率 ζ_2)
確率 ζ_2 で「潜在的リンク」の一つを選択し、実リンク化する。ただし、選択する潜在的リンクは確率 θ_2 でランダムに選択され、確率 $1 - \theta_2$ で優先的選択に

より選択される。

(3-2) 潜在的コミュニティリンクの実リンク化 (確率 η_2)

確率 η_2 で「潜在的コミュニティリンク」の一つを選択し、実リンク化する。ただし、選択する潜在的コミュニティリンクは確率 ι_2 でランダムに選択され、確率 $1 - \iota_2$ で優先的選択により選択される。また、接続したコミュニティと接続したノードとの間に潜在的コミュニティリンクを作成する。

(3-3) ランダム接続 (確率 $1 - \zeta_2 - \eta_2$)

確率 $1 - \zeta_2 - \eta_2$ でランダムに選んだノード間にリンクを張る。

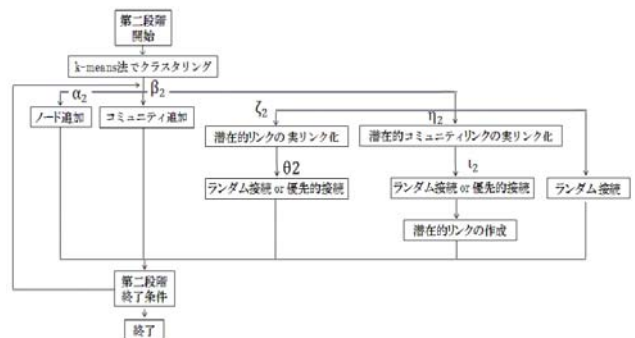


図 5 第二段階における新規リンクの追加

以上の第一段階、第二段階の流れに従って、既存の SNS と性質上類似したネットワークを形成することを試みる。

4. 評価実験

4.1 評価実験の概要

提案モデルを用いたシミュレーションを行ない、提案モデルの妥当性を検証する。提案モデルを用いて作成したネットワークと実在する SNS のネットワーク指標データを比較することにより、提案モデルが SNS のネットワーク構造を再現できることを検証する。なお、再現対象の SNS としては、性質の異なったものを複数用意し、提案モデルによって、多様な SNS の再現が可能かどうかを検討する。本研究では、先行研究[3][5]をもとに、以下の 2 つの SNS を対象として分析を行なった。

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

対象とした SNS のネットワーク指標を表 1 に示す。

表1 対象とした SNS のネットワーク指標

	モリオネット	あみっぴい
ノード数	661	2609
リンク数	4617	11352
コミュニティ数	50	133
平均経路長	2.81	3.10
クラスタリング係数	0.415	0.443
べき指数	1.023	1.169
同類選択性	-0.237	-0.255

ここでモリオネットとあみっぴいのコミュニティ数に関しては、それぞれの2011年1月時点でのノード数とコミュニティ数の比率から算出した。

4.2 パラメータの設定方法

提案モデルにおいて SNS を生成する際、設定すべきパラメータは全部で以下の13個である。

- 第一段階
 - (a) ノード追加を選択する確率 (α_1)
 - (b) 選択ノード接続を選択する確率 (β_1)
 - (c) 潜在的リンクを実リンク化する確率 (γ_1)
- 第二段階
 - (d) ノード追加を選択する確率 (α_2)
 - (e) コミュニティ追加を選択する確率 (β_2)
 - (f) 選択ノード接続を選択する確率 (γ_2)
 - (g) トライアドフォーメーションの実施確率 (δ_2)
 - (h) ユーザーによるコミュニティの追加確率 (ϵ_2)
 - (i) 潜在的リンクを実リンク化する確率 (ζ_2)
 - (j) 潜在的コミュニティリンクを実リンク化する確率 (η_2)
 - (k) 潜在的リンクのランダム選択接続の確率 (θ_2)
 - (l) 潜在的コミュニティリンクのランダム選択接続を選択する確率 (ι_2)
 - (m) 第一段階と第二段階の区切り (T)

このうち、(c) 確率 γ_1 に関しては、先行研究[3]において、「ランダム接続が採用されているモデルにおけるランダム接続確率は0.04に固定した」という記述があることから、本研究においても第一段階のランダム接続確率は0.04にするため、確率 γ_1 は0.96に固定した。同様に第二段階のランダム接続確率を0.04になるように $\zeta_2 + \eta_2 = 0.96$ として固定した。

他のパラメータの設定方法については、まず初めに理論上決定できるパラメータについては優先的に決定する。モリオネットとあみっぴいのノード数とリンク数の関係からパラメータをいくつか決定する。

はじめにノード数とリンク数の関係を選んだのは、他のネットワーク指標が各パラメータの値によって変化するものであるのに対し、リンク数はいくつかのパラメータの値によって、その理想値が決定するからである。ここで決められるパラメータとは確率 $\alpha_1, \alpha_2, \beta_2, \delta_2, \zeta_2, T$ の6個である。

この6個のパラメータを一定の刻み幅で変化させ、リンク数の理想値を算出し、モリオネットとあみっぴいの実際のリンク数と比較し、パラメータ候補となるものを決定した。

そしてそれ以外のパラメータについては、パラメータを一定の刻み幅で変化させ、ネットワーク生成のシミュレーションを行ない、シミュレーションによって算出したネットワーク指標と実際の SNS のネットワーク指標を比較することでパラメータを決定する。対象とした SNS は主にモリオネットを用いた。

4.3 提案モデルの評価

以上のように各パラメータを設定し、提案モデルによってネットワークを生成した。そして生成したネットワークにおけるネットワーク指標と対象の SNS におけるネットワーク指標と比較することで、提案モデルの有効性を評価する。

- モリオネットのネットワーク指標
提案モデルによって、生成されたモリオネットのネットワーク指標を表2に示す。

表2 モリオネットと生成されたネットワークの比較

	モリオネット	提案モデル
リンク数	4617	4295
平均経路長	2.81	2.80
クラスタリング係数	0.415	0.451
べき指数	1.023	1.067
同類選択性	-0.237	-0.034

表2には実存の SNS のネットワーク指標(平均経路長, クラスタリング係数, べき係数, 同類選択性)と提案モデルによって生成されたネットワークの指標を示す。表2より、提案モデルでは、平均経路長, べき指数だけでなく、リンク数においても実際の指標に近い値を示している。一方でクラスタリング係数については、若干大きくなってしまった。また同類選択性についても、負の値は再現できているが、実際値より若干ずれた値となってしまった。しかしながら、提案モデルによって、モリオネットについては、実データに近いネットワーク構造を再現可能

であると言える。

● あみっぴいのネットワーク指標

あみっぴいの場合、モリオネットを形成する際に求めたパラメータをそのまま適用し、ネットワークを生成した。提案モデルによって生成されたネットワーク指標を表3に示す。

表3 あみっぴいと生成されたネットワークの比較

	あみっぴい	提案モデル
リンク数	11352	19914
平均経路長	3.10	3.211
クラスタリング係数	0.443	0.399
べき指数	1.169	1.327
同類選択性	-0.255	-0.008

表3より、提案モデルでは、平均経路長やべき指数およびリンク数が実データよりも大きい値になってしまっているが、かなり近い値を示している。クラスタリング係数については、若干小さくなってしまった。同類選択性については、モリオネットの結果と同様に、負の値は再現できているが、実際値より若干ずれた値となってしまった。

総合的には、モリオネットの方が再現できているようにとらえることができるが、これは、モリオネットの場合においては、最も良い結果の出たパラメータを採用しているからである。

本研究の提案モデルでは、対象とする SNS が変わってもパラメータの値を変化させていない。第一段階、第二段階における他パラメータも考慮することによって、あみっぴいの場合も、実存する SNS と類似したネットワークを生成できるものと期待できる。

5. まとめ

本研究においては、SNS による友人ネットワークの成長過程を明らかにし、そのモデル化を考案することを目的とした。

考案した成長モデルにおいては、実存の SNS に成長過程を考慮して、ネットワークの成長を二つの段階に分けて考えた。まず、第一段階においては、CNNR モデルに管理者優先接続を加えたモデルによって、ネットワークを生成する。第二段階においては、2 種類のコミュニティを導入した成長モデルを考案した。そして実存の SNS 「モリオネット」および「あみっぴい」を対象として、提案モデルの各パラメータを設定し、ネットワークを生成した。

ネットワーク指標による定量的分析の結果、あみ

っぴいに関しては、平均経路長やべき指数およびリンク数の値が若干大きくなってしまったが、実存の指標に近いネットワークを生成することができた。また、モリオネットに関しては、ネットワーク指標は実ネットワークとほぼ同じ値となり、提案モデルの有効性を示すことができた。

今後は、対象 SNS ごとに適切なパラメータの値を求める手法を模索し、より実存の SNS に近いネットワークを生成できるモデルの構築を試みていく予定である。

謝辞

研究開始時におきまして、研究に協力していただきました本郷寛君に深く感謝の意を表します。

参考文献

- [1] 安田雪：実践ネットワーク分析—関係を解く理論と技法，新曜社（2001）
- [2] 内田誠，白山晋：SNS のネットワーク構造の分析とモデル推定，情報処理学会論文誌，Vol.47，No.9，pp.2840-2849（2006）
- [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モデルの提案，電子情報通信学会論文誌，Vol.J93-D，No.7，pp.1135-1143（2010）
- [4] 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，pp.17-24（2006）
- [5] 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，The 23rd Annual Conference of the Japanese Society for Artificial Intelligence（2009）
- [6] 神谷達幸，鳥海不二夫，石井健一郎：パラメータ推定を利用した SNS 成長モデルの提案，合同エージェンツワークショップ&シンポジウム 2010 論文集（2010）
- [7] K. Yuta, N. Ono, and Y. Fujiwara. : A Gap in the Community-Size Distribution of a Large-Scale Social Networking Site, Arxiv preprint physics（2007）
- [8] 鳥海不二夫，石田健，石井健一郎：地域 SNS のネットワーク構造分析，電子情報通信学会技術研究報告，AI2008-09，pp.33-38（2008）
- [9] 松尾豊，安田雪：SNS における関係形成原理-mixi のデータ分析，人工知能学会論文誌，Vol.22，No.5，pp.531-541（2007）
- [10] 本郷寛：コミュニティ構造を考慮した SNS のネットワーク成長モデル，慶應義塾大学理工学部卒業論文（2011）