

情報伝播における潜在リンクの影響

Influence of Potential Edge on Information Diffusion

永田 勝也*¹
Katsuya NAGATA

塚本 鋭*²
Ei TSUKAMOTO

白山 晋*³
Susumu SHIRAYAMA

*¹ 東京大学工学部システム創成学科
Faculty of Engineering, the University of Tokyo

*² 東京大学大学院工学系研究科 *³ 東京大学人工物工学研究センター
School of Engineering, the University of Tokyo RACE, the University of Tokyo

It has become important to consider the effects of “word-of-mouth” or “trust-marketing” based on personal relationships in strategy of marketing. The word-of-mouth propagates along the network of the personal relationships, and often affects consumer purchasing behavior. It has been pointed out that characteristics of the propagation strongly depend on the network structure. In this paper, we study relationship between diffusion of word-of-mouth and network structure. Especially, we focus on the *potential edge* which a pair of nodes are not connected by an edge and have at least one common neighbor. Influence of potential edge on information diffusion is studied by numerical simulations using a simple information diffusion model on networks generated by several network models related to the potential edge.

1. はじめに

伝染病の伝播, 社会的合意形成といった, 現実社会の現象・事象の背後にはネットワークが存在し, その構造が影響を与えることが知られている. 噂の伝播やクチコミなどの情報伝播に関しても同様であり, ネットワーク構造の影響が考慮される[桑島 2005]. しかし, 既存研究では構造の分析が主であり, 情報伝播というネットワークの機能に対して構造が与える影響を調べたものは少ない. 実ネットワークを用いた情報伝播に関する実験ができたとしても結果の分析が困難なためである.

一方, 複雑ネットワークの分野では, 実ネットワークの性質を有する多くのネットワーク生成モデルが提案され, モデルによって形成されたネットワーク上で情報伝播のシミュレーションが行われている. そして, これにより, 分析が難しい実ネットワーク上の伝播現象に対して何かしらの知見を与えることができると考えられている. また, 効率的な情報伝播を実現するために, モデルでのネットワーク上で情報伝播の傾向を検討し, 実ネットワークに適用するという戦略も考えられる.

ネットワーク生成モデルとは, ネットワークをノードとリンク(エッジ)により表し, その生成規則を定めたもので, 様々なモデルが存在する[Boccaletti 2006]. 現実の, 特に社会ネットワークを再現するネットワークモデルの1つに, CNN (Connecting Nearest Neighbor)モデルがある[Vazquez 2003] [三井 2006]. このモデルの特徴は, 潜在リンク(Potential edge)という概念にある. 潜在リンクとは, ノード間でリンクが生じたときに, 一方のノードのもつ既にリンクされているノードとの間の関係性を示すものである(友人関係とすれば, 友人の友人との間の関係性である. 2ホップの関係性と呼ぶ場合もある).

この潜在リンクのリンク化という機構が, 社会ネットワークの構造の形成において重要な役割を果たしていることが示唆されている [Vazquez 2003] [三井 2006] [内田 2006]. しかし, その機構によって形成されるネットワークの構造が情報伝播のようなネ

ットワークの機能に与える影響についてはよくわかっていない.

そこで, 本稿では, この潜在リンクに注目し, 潜在リンクが情報伝播の効率性に与える影響を, 潜在リンクが生成機構の主要素であるネットワークモデルによって形成されたネットワークを用いて調べる. なお, 潜在リンクの直接的な影響を調べるものではないことを断っておく.

2. 提案手法

潜在リンクが生成機構の主要素であるいくつかのネットワークモデルによってネットワークを形成する. 数理モデルによってネットワーク上の情報伝播を模倣し, 数値シミュレーションによって情報伝播に対する潜在リンクの間接的な影響を調べる.

はじめに, 潜在リンクの存在する既存の3つのネットワークモデル(Vazquez のモデル[Vazquez 2003], 三井らのモデル[三井 2006], Yuta らのモデル[Yuta 2007])と, それらの組み合わせによる3つのモデルの計6つのモデルに対して情報伝播の影響を調べる. これらのモデルをネットワークモデルIで区別する.

次に, Vazquez モデルと三井らのモデルの混合モデルを考える. これをネットワークモデルIIとする.

2.1 ネットワークモデル I

Vazquez のモデル[Vazquez 2003], 三井らのモデル[三井 2006], Yuta らのモデル[Yuta 2007]を, それぞれ CNN モデル, M モデル, Y モデルと表記する.

(1) CNN モデル [Vazquez 2003]

以下の2つのプロセスを繰り返すことによってネットワークを生成する.

① 確率 p で新規ノードを生成し, 既存ノードのいずれかをランダムに選び, 新規ノードと結ぶ. 結ばれる既存ノードの既リンクの端点のノードと新規ノードの間を潜在リンクと呼ぶ.

② 確率 $(1-p)$ で①で生じた潜在リンクのいずれかをランダムに選びリンク化する.

(2) M モデル [三井 2006]

以下の2つのプロセスを繰り返すことによってネットワークを生成する.

① 確率 p で新規ノードを生成し, 既存ノードのいずれかをランダムに選び, 新規ノードと結ぶ.

連絡先: 白山 晋, 東京大学人工物工学研究センター, 千葉県柏市柏の葉 5-1-5, 04-7136-4263, 04-7136-4263, sirayama@race.u-tokyo.ac.jp

確率 $(1-p)$ で潜在リンクのいずれかをランダムに選びリンク化する。

CNN モデルとの違いは潜在リンクにある。本稿では、改めて、任意のノードと2ホップ先のノードとの関係を潜在リンクと定義する。

M モデルでの潜在リンクは、この意味でのものである。一方、CNN のモデルでは、新規ノード追加時に生じる新規ノードから2ホップ先のノードを潜在リンクとしている。本稿では、これを潜在リンク V(PE_V)とし、潜在リンクのリンク化の際に生じる潜在リンクを潜在リンク L(PE_L)として両者を区別する。CNN モデルとMモデルの違いを図1に示す。

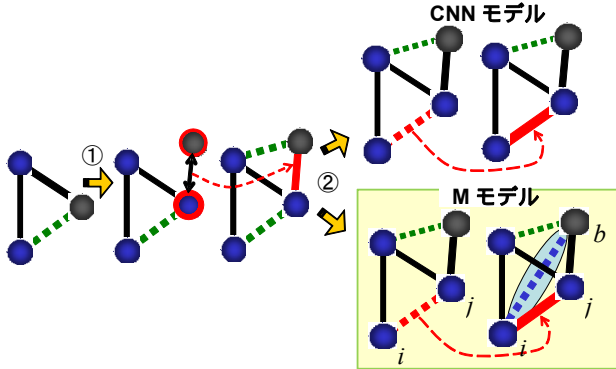


図1: モデル CNN とモデル M の生成メカニズムの違い

(3) Y モデル [Yuta2007]

以下の3つのプロセスを繰り返すことによってネットワークを生成する。

- ① 確率 p で新規ノードの生成 (潜在リンク V が生成)
 - ② 確率 u で潜在リンク V のリンク化
 - ③ 確率 $1-(p+u)$ で、既存の2ノードをランダムにリンク化
- ここで、③によって生じるリンクをショートカットと呼ぶことにする。また、ショートカットを SC と略記する。

(4) A モデル

以下の3つのプロセスを繰り返すことによってネットワークを生成する。M モデルに Y モデルのプロセス③を加えるものである。ただし、③において潜在リンクを生成しない。

- ① 確率 p で新規ノードの生成 (潜在リンク V が生成)
- ② 確率 u で潜在リンクのリンク化 (潜在リンク L が生成)
- ③ 確率 $1-(p+u)$ で、既存の2ノードをランダムにリンク化 (潜在リンクを生成しない)

(5) B モデル

以下の3つのプロセスを繰り返すことによってネットワークを生成する。プロセス③で生じる潜在リンクを潜在リンク S とする。 PE_S で示す。

- ① 確率 p で新規ノードの生成 (潜在リンク V が生成)
- ② 確率 u で潜在リンク V, L, S のリンク化 (潜在リンク L が生成)
- ③ 確率 $1-(p+u)$ で、既存の2ノードをランダムにリンク化 (潜在リンク S が生成)

(6) C モデル

以下の3つのプロセスを繰り返すことによってネットワークを生成する。

- ① 確率 p で新規ノードの生成 (潜在リンク V が生成)
- ② 確率 u で潜在リンク V, S のリンク化
- ③ 確率 $1-(p+u)$ で、既存の2ノードをランダムにリンク化 (潜在リンク S が生成)

以上の6つのモデルの違いは、

- (a) 新規ノード作成時に生じる潜在リンク (PE_V) と、潜在リンクの実リンク化時に生じる潜在リンク (PE_L) の区別の有無
- (b) ショートカット (SC) 作成の有無
- (c) ショートカット作成時の潜在リンク (PE_S) 生成の有無である。

これらのモデルの違いを表1に示す。

表1: モデル間の違い

	PE_V	PE_L	SC	PE_S
CNN	○	×	×	×
M	○	○	×	×
Y	○	×	○	×
A	○	○	○	×
B	○	○	○	○
C	○	×	○	○

2.2 ネットワークモデル II

ネットワークモデル I に属する6つのモデルの違いの中で、新規ノード作成時に生じる潜在リンク (PE_V) と、潜在リンクの実リンク化時に生じる潜在リンク (PE_L) の注目し、CNN モデルと M モデルの混合モデルによってネットワークを生成する。

以下の2つのプロセスを繰り返すことによってネットワークを生成する。

- ① 確率 p で新規ノードの生成 (PE_V が生成)
- ② 確率 $(1-p)$ で潜在リンク V, あるいは L のリンク化。
- ②' この際、確率 q で PE_V を、確率 $(1-q)$ で PE_L をリンク化 (いずれの場合も PE_L が生成)

このモデルでは、 $q = 1$ の場合、 PE_V のみがリンク化するため、CNN モデルと一致する。また、 $q = p$ のとき、M モデルに近づく。

2.3 情報伝播モデル

本稿では、ノードに $+1$ と -1 の2つの状態を与え、それが時間的に変化する下記の情報伝播モデルを扱う。

前節で述べたネットワーク生成モデルによって形成されたネットワークにおいて情報を伝播させることにより潜在リンクの影響を探る。

$\sigma_i(t)$ をノード i の時間 t での状態を示すものとし、次式で表されるルールによって状態を変化させる。

$$\sigma_i(t+1) = \begin{cases} \text{Sgn} \left\{ \sum_j^N a_{ij} \sigma_j(t) \right\} & : \sum_j^N a_{ij} \sigma_j(t) \neq 0 \\ \sigma_i(t) & : \sum_j^N a_{ij} \sigma_j(t) = 0 \end{cases} \quad (1)$$

ここで、 σ は $+1$, あるいは -1 の変数、 a_{ij} は隣接行列を表す。また、 N はノードの総数である。このモデルはイジングモデルの一種である。また、Uchida らの Glauber dynamics を扱ったモデル [Uchida 2007] と等価である。ただし、同期で扱う点が異なる。

状態 $+1$ のノードの割合を r とする。初期状態として、割合 r_0 になるように $+1$ をランダムに与え、式(1)に従って状態を更新する。情報伝播の効率率を、収束時間 (T_c) と収束時の $+1$ の割合 (r_∞) から算出する。

予備実験からは、ノードの割合が1つの値に収束する場合と、2つの値で振動する場合があることが分かっている。後者の場合は、2つの値の平均を r_∞ とする。

3. 実験と考察

3.1 ネットワークモデル I

はじめに、ネットワークモデル I について、潜在リンクが情報伝播に与える影響を調べる。

ノード数を $N = 10000$ とし、 p, u を与え、ネットワークを生成する。 r を 0.5 から 0.8 まで 0.1 刻みで与え、情報を伝播させる。一組のパラメータ (p, u, r) に対して、10 回の試行を行い、収束時間の平均 (T_c) と収束時の $+1$ の割合の平均 (r_∞) を求める。なお、ネットワークは試行毎に作成している。

図2に、 $p = 0.25, u = 0.65$ の場合における、CNN, M, Y, A, B の5つのモデルの次数分布を示す。 k は次数、 $p(k)$ は確率分布である。図からわかるように、これらの次数分布は、CNN, Y と M, A, B の2系列に分けられる。

CNN モデルと M モデルを抜き出したものを 図 3 に示す。SNS (Social Network Service) を例にすると、CNN モデルの次数分布は日本の mixi に近く[内田 2006][Yuta2007]、M モデルのものは、韓国の Cyworld に近い[Ahn2007]。ただし、ノード数に差があるため、低次数の領域が異なっている。

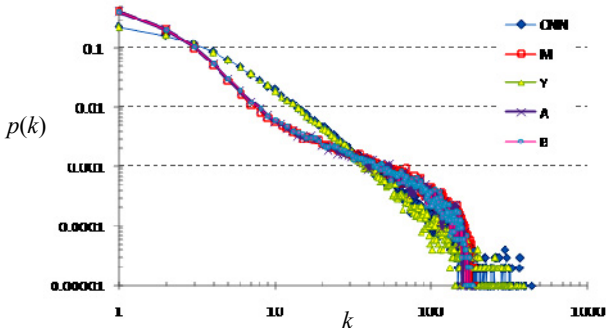


図2:次数分布 (CNN,M,Y,A,B)

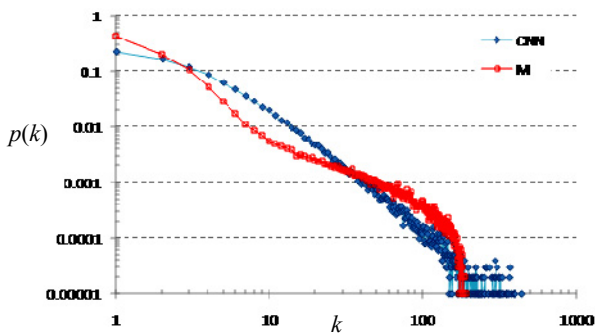


図3:次数分布 (CNN,M)

それぞれのネットワークに対して、2.3 節で述べた情報伝播モデルを用いて情報伝播のシミュレーションを行う。

図 4 に収束値を示す。この結果からは、CNN モデルと Y モデルから生成されたネットワークでの収束値が他の3つのモデルよりも大きいことがわかる。

次に、図 5 に各ネットワークでの収束時間を示す。図からは、M, A, B のモデルから作成されたネットワークでの収束時間が他の2つよりも短いことがわかる。また、CNN と Y, M と A, および B を比べると、ショートカットの存在が収束時間を短くしていることもわかる。

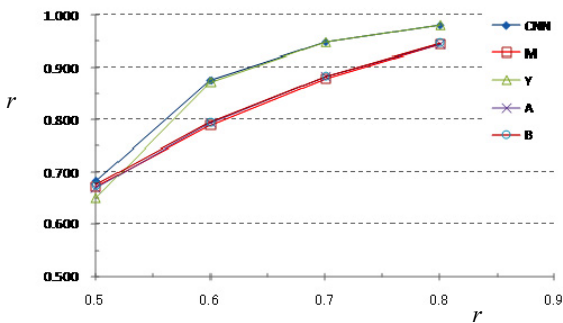


図4: 状態+1の初期割合と収束時の割合の関係

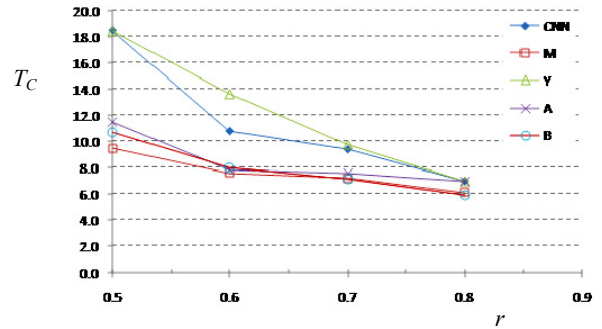


図5: 状態+1の初期割合と収束時間の関係

3.2 ネットワークモデル II

ネットワーク I の場合と同様に、ノード数を $N = 10000$ とし、 p, u を与え、ネットワークを生成する。 r を 0.6 から 0.8 まで 0.1 刻みで与え、情報を伝播させる。一組のパラメータ (p, q, r) に対して、50 回の試行を行い、収束時の +1 の割合の平均 (r_∞) と収束時間の平均 (T_c) を求める。なお、ネットワークは試行毎に作成している。以後、断りのない限り 50 回の試行を平均した結果を示す。

図 6 は、 $p = 0.25$ として、 q を変化させた場合の次数分布である ($q = 0.9, 0.7, 0.5, 0.3, 0.1, 0.01$ の場合を示す)。 q が小さくなるに従い、CNN モデルの次数分布から M モデルのものへと移行する。

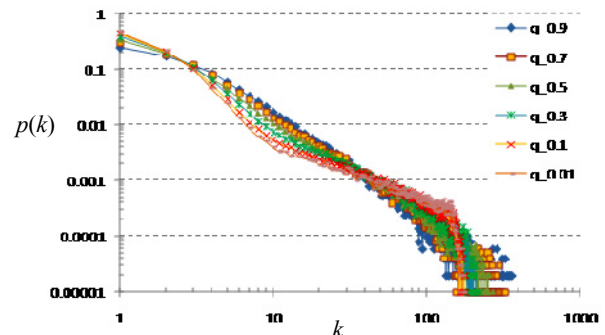


図6:混合モデルの次数分布

それぞれのネットワークに対して、2.3 節で述べた情報伝播モデルを用いて情報伝播のシミュレーションを行う。 q の値を 0.01 から 0.99 まで 0.01 刻みで変化させる。図 7 に初期の +1 の割合と収束状態の +1 の割合を示す ($q = 0.9, 0.7, 0.5, 0.3, 0.1, 0.01$ の場合)。また、初期の +1 の割合と、収束時間の関係を図 8 に示す。図 4 と図 7、図 5 と図 8 からは、混合モデルの、状態+1 の初期割合と収束時の割合、および収束時間が、CNN モデルと M モデルを連続的につないでいることがわかる。

CNN モデルの収束時の割合を $r_\infty^{\text{CNN}}(r)$ 、M モデルのものを $r_\infty^{\text{M}}(r)$ 、混合モデルのものを $r_\infty(r, q)$ とする。また、CNN モデルの収束時間を $T_c^{\text{CNN}}(r)$ 、M モデルのものを $T_c^{\text{M}}(r)$ 、混合モデルのものを $T_c(r, q)$ とする。数値実験の結果から、

$$r_\infty(r, q) \approx \frac{r_\infty^{\text{CNN}}(r) - r_\infty^{\text{M}}(r)}{1 - q_M} (q - q_M) + r_\infty^{\text{M}}(r) \quad (2)$$

$$T_c(r, q) \approx \frac{T_c^{\text{CNN}}(r) - T_c^{\text{M}}(r)}{1 - q_M} (q - q_M) + T_c^{\text{M}}(r) \quad (3)$$

となることがわかった。ここで、 q_M は、 $0 < q_M < 1$ で、 p に依存する。 $p = 0.25$ の場合は、約 0.2 であった。

一方、図 6 で示すように、混合モデルの次数分布は、CNN モデルと M モデルの次数分布を、パラメータ q によって線形結

合したものに近い。したがって、次数分布と情報伝播の特性である収束時の割合、および収束時間に強い相関が存在することがわかる。ただし、次数分布そのものと関係するかは現時点ではわかっていない。

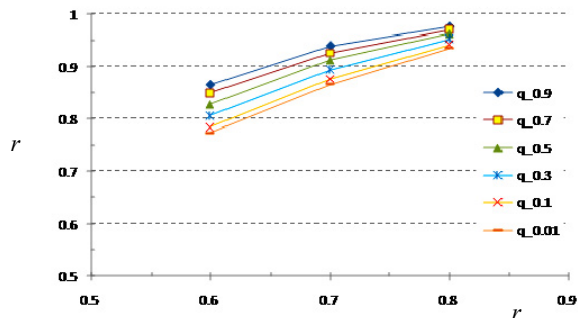


図 7: 状態+1の初期割合と収束時の割合の関係

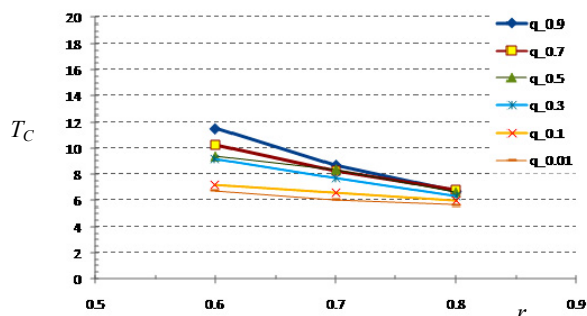


図 8: 状態+1の初期割合と収束時間の関係

次に、情報伝播の特性を調べた。本研究で採用した伝播モデルは単純なため多くのことはわからない。ただし、伝播モデルが単純なため、ネットワーク自体がもつ伝播特性が顕在化されうことを付記しておく。

ここでは、情報の拡がり方と伝わる速さから簡単な考察を行う。情報の拡がり方を示す指標 D を、単純に、状態+1の収束時の割合を初期割合で割ったものとする($D = r_c/r$)。伝わる速さを示す指標は、収束時間(T_c)を用いる。広く、早く伝われば情報伝播の効率が良いと考え、効率性を、 $E(\alpha, \beta) = D^\alpha / S^\beta$ とおく。いくつかの α と β の組について効率性を調べた。紙面の都合で結果を掲載できないので、講演時に示す。

3.3 まとめと考察

ネットワークIに対する結果において、収束値の観点では、潜在リンクのリンク化時に潜在リンクを作成しない方が、効率が良いことがわかる(図 4)。一方、収束時間の観点では、潜在リンクのリンク化時に潜在リンクを作成する方が、効率が良いことがわかる(図 5)。また、ショートカットの有無は収束時間を早める効果はあるが、今回のパラメータの範囲ではその影響は小さい。

ネットワークIIに対する結果において、収束値の観点では、新規ノード追加時にできた潜在リンク(PE_V)を優先的にリンク化させた方が効率がよいこと、その割合(q)を増加させると、ほぼ線形的に効率が良くなることがわかった。一方、収束時間は、潜在リンクのリンク化によってできる潜在リンク(PE_L)を優先的にリンク化させた方が短くなることがわかる。

情報伝播の効率に対して、 r への依存性が少ないものは、ネットワーク自体の伝播特性をより正確に表すものと考えられる。いくつかの α と β の組について調べた結果、 $\alpha = 3$, $\beta = 1$ のときに、 r への依存性が少なくなることがわかった。

ここで、現実世界における潜在リンクのリンク化の意味を考えてみる。人間関係としてみると、 PE_V のリンク化は、はじめに友人

となった人物の友人と、友人になることを表している。一方、 PE_L のリンク化は、初期には遠い関係だった人物とも友人になっていくことを表している。

ネットワークの成長速度は一定とする(ノードやリンクの増加速度は一定と仮定する)。潜在リンクのリンク化を介して成長するネットワークにおいて、時刻 t_1 と t_2 ($t_1 < t_2$) において情報を伝播させたとき、収束時間が長くなったとする。この場合、図 8 からは、 q の値が増加したものと考えられる。初期の友人の友人との交流が増えたといったことが推測される。収束値の増減についても同様である。所与の情報伝播の特性をもつようにネットワーク構造を変えることも考えられる。組織とすれば、新しく入ってきた人を、最初に友人になった人が友人に紹介する、などの最初の友人への依存度を大きくすれば、 q の値が大きいネットワークになる(q の値が大きいネットワークの情報伝播特性をもつ)。

4. おわりに

本研究で採用した情報伝播モデルは単純なため、一般的な意味では、情報伝播とネットワーク構造の関係性に関して詳細な知見を得ることは難しい。しかし、モデルが単純であることから、ネットワーク構造の違いによって生じる情報伝播の差が構造をより反映するものであると考えられる。

本稿で注目した次数分布はネットワーク成長の結果として現れ、構造上の特徴が顕在化されたものである。また、局所的な相互作用を反映するものである。ネットワークIIにおける情報伝播と次数分布の相関の強さは、副次的ではあるが、潜在リンクの影響を示唆していると考えている。

2ホップ内の伝播特性を変えることや、友人ができた時点のその友人の知人と、その他の友人の知人との伝播特性を変えるなどの方法によって、情報伝播に与える潜在リンクの直接的な影響について検討することが今後の課題である。

参考文献

- [桑島 2005] 桑島由美, 小林大祐: Web ネットワークにおけるクチコミ効果, 東京大学 COE ものづくり経営研究センター, MMRC-J-56, 2005.
- [Boccaletti2006] Boccaletti, S., Latora, V., Moreno, Y., Chavez M. and Hwang, D.-U.: Complex networks: Structure and dynamics, Physics Reports, 424, pp. 175 - 308, 2006.
- [Vazquez2003] V'azquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, Physical Review E 67, 056104, 2003.
- [三井 2006] 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—知能と複雑系, Vol.2006, No.2, pp.17-24, 2006.
- [内田 2006] 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.
- [Ahn2007] Ahn, Y.Y., Han, S., Kwak, H., Moon, S. and Jeong, H.: Analysis of topological characteristics of huge online social networking services, Proc. of the 16th international conference on World Wide Web, pp. 835-844, 2007.
- [Yuta2007] Yuta, K., Ono, N. and Fujiwara, Y.: A Gap in the Community-Size Distribution of a Large-Scale Social Networking Site, arXiv: physics/0701168v2, 2007.
- [Uchida2007] Uchida, M. and Shirayama, S.: Effect of Initial Conditions on Glauber Dynamics in Complex Networks, Physical Review E, 75, 046105, 2007.