

陽なコミュニティの情報伝播に与える影響

Influence of Explicit Community on Information Diffusion

永田勝也*1

Nagata Katsuya

白山晋*2

Shirayama Susumu

*1 東京大学大学院工学系研究科

School of Engineering, the University of Tokyo

*2 東京大学人工物工学研究センター

RACE, the University of Tokyo

In the field of complex network research, there are two types of communities, which play an important role in information diffusion. One is the community that is defined by the topological structure such as sparse and density of links. We call it the implicit community. The other is that which has concrete meanings or can be distinguished by a name like SIG (Special Interest Group). This is called by the explicit community. Although both communities affect directly the information diffusion on networks, few have been explored on the effect of the explicit community. In this paper, we analyze the influence of the explicit community on information diffusion. In addition, we show quantitative expression of the effect of the difference in which links and communities are generated.

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, その構造が現象に影響を与えることが知られている [1]. 噂の伝播やクチコミ (word-of-mouth) 等の情報伝播に関しても同様であり, ネットワーク構造が影響を与えている [8].

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.

現実のネットワークに近い, 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一つに, CNN (Connecting Nearest Neighbor) モデルがある [5, 6]. 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を示すものである. この潜在リンクの実リンク化という機構が, 社会ネットワーク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 [5, 6, 7]. 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に与える影響についてはよくわかっていない.

一方, 情報伝播において, コミュニティが影響を与えることが指摘されている. 複雑ネットワーク分析においても, ネットワーク構造の顕在化のための方法や方法論の中でコミュニティが扱われる [2]. しかしながら, 複雑ネットワーク分析で扱われるコミュニティの多くは, 「グループ内部のリンクが多く, グループ間のリンクが少ないノード集団」というような定義に見られるように, ネットワークに内在するクラスタ構造 (以降, 陰的なコミュニティと呼ぶ) のことを指すことが多い. 一方, SNS に代表される WWW 上に形成された社会的ネットワークにおける SIG (Special Interest Group) のような, 陽に形成されるコミュニティ (以後, 陽なコミュニティと呼ぶ) を扱っているわけではない.

一方, 陽なコミュニティは明示的に意味を与えられるものである. Newman ら [3] は, 映画と俳優, 取締役会と参加者といった対象に応用できる 2 部グラフに基づくネットワーク

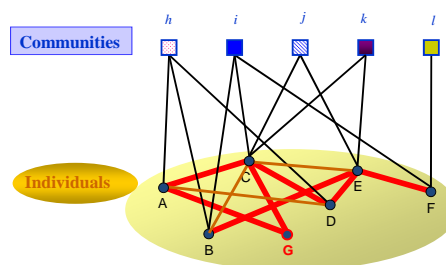


図 1: コミュニティを有するネットワーク

の形成モデルを提案している. また Noh ら [4] は, ノードと, ノードの所属関係を表すコミュニティの 2 部構造からなるネットワーク成長モデルを提案している. これらは, 陽なコミュニティを有するが, コミュニティとノードの関係は 2 部グラフ, さらにノードは必ずコミュニティと結びつけられる完全 2 部グラフの構造である. これらに対し三井らは, CNN モデルを拡張し, コミュニティを介しないリンクの形成やコミュニティに属さないノードを含めた, 陽なコミュニティを有するネットワーク成長モデルを提案している [6]. 図 1 はこのネットワークを模式的に示したものである. このようなコミュニティの構造とネットワーク上の情報伝播には何かしらの関係があると考えられるが, いずれの研究においても構造が情報伝播に与える影響は調べられていない.

以上をふまえて, 潜在リンクが生成機構の主要素であり, 陽なコミュニティを有するネットワークモデルによってネットワークを形成し, ネットワーク上での情報伝播に与えるコミュニティの影響を調べる.

2. ネットワーク生成モデル

本稿では, 陽なコミュニティを有するネットワークモデルとして, 三井らのモデルを基本とする [6]. 三井らのモデルでは, 2 種類の潜在リンクと 2 種類の潜在コミュニティリンクを有する. 潜在リンクの 1 つは, ネットワーク加入時に形成するもので, これを PE_V として表す. もう 1 つは, 潜在リンクのリンク化時に形成するもので, これを PE_L として表す. 潜在コミュニティリンクの 1 つは, ネットワーク加入時と新規コミュ

ティ形成時に形成するもので、 PCE_V として表す。もうひとつは、潜在リンクのリンク化時と、潜在コミュニティリンクのリンク化時に形成するもので、 PCE_L として表す。三井らのモデルでは、潜在リンクや潜在コミュニティリンクを、形成した過程によって区別することはできなかった。現実世界を考えると、形成過程に起因する区別はネットワーク形成に影響を与えると考えられる。

そこで本研究では、三井らの提案したネットワーク成長モデルを統合し、1つのネットワークモデルとして新たに提案する。以下に、本研究で提案する混合モデルの形成方法を示す。

2.1 混合モデル

u を新規ノードを追加する確率、 x を潜在リンクをリンク化する確率、 y を新規コミュニティを追加する確率とする。また、 q_1 を、潜在リンクをリンク化する際 PE_V を選択する確率とし、 q_2 を、潜在コミュニティリンクをリンク化する際 PCE_V を選択する確率とする。以下のいずれかを確率的に行うことを繰り返してネットワークを生成する。

- A) 確率 u で新規ノードを追加し、ランダムに選択した既存ノード i とリンクする。その際、新規ノードと選択ノード i の全ての隣接ノードとの辺を次のリンク候補 PE_V として保存する。また、選択ノード i の参加している全てのコミュニティと新ノードとの辺を、次のコミュニティリンク候補 PCE_V として保存する。
- B) 確率 x で、潜在リンクを1つリンク化する。その際、確率 q_1 で PE_V を選択し、確率 $1 - q_1$ で PE_L を選択してリンク化する。新しくリンクされたノード i, j について、ノード i とリンクされている全てのノードと、ノード j との辺を、次のリンク候補 PE_L とする。また、ノード j とリンクされている全てのノードとノード i との辺を、次のリンク候補 PE_L とする。ノード i とリンクされている全てのコミュニティと、ノード j との辺を、次のコミュニティリンク候補 PCE_L とする。また、ノード j とリンクされている全てのコミュニティと、ノード i との辺を、次のコミュニティリンク候補 PCE_L とする。
- C) 確率 y で、新規コミュニティを追加し、ランダムに選択した既存ノード i とリンクを張る。つまり、ノード i が新規コミュニティを作成したことになる。その際、新規コミュニティと、ノード i とリンクされている全てのノードとの辺を、次のコミュニティリンク候補 PCE_V として保存する。
- D) 確率 $1 - u - x - y$ で、潜在コミュニティリンクの1つをリンク化する。その際、確率 q_2 で PE_V を選択し、確率 $1 - q_2$ で潜在コミュニティリンク PE_L を選択してリンク化する。その際、該当コミュニティと、ノード i とリンクされている全てのノードとの辺を、次のリンク候補と PCE_L として保存する。

3. 情報伝播モデル

3.1 基本伝播モデル

情報伝播には、ある地点、ある人から移流拡散的に伝わっていくというモデルと、全体の合意形成のような、ある状態からある状態へと至る遷移を扱うモデルがある。

本研究では後者に注目し、ノードに+1,-1の2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。隣接行列を A_{ij} とし、 $\sigma_i(t)$ をノード i の時間 t での状態を示すものとする。

はじめに、局所的な場 $h_i(t)$ を次式で定める。

$$h_i(t) = \sum_j^N A_{ij} \sigma_j(t) \quad (1)$$

ここで、 N はノードの総数である。次に、ノード i の状態を式2で表されるルールによって変化させる。

$$\sigma_i(t+1) = \begin{cases} \text{sgn}\{h_i(t)\} & \text{if } h_i(t) \neq 0 \\ \sigma_i(t) & \text{if } h_i(t) = 0 \end{cases} \quad (2)$$

3.2 提案伝播モデル

ある状態からある状態への遷移には、直接リンクしたノードからの影響に加え、コミュニティを介した影響があると考えられる。本稿では、コミュニティの影響を考慮する伝播モデルを提案する。同一コミュニティに属するがリンクのないノード間の状態を局所的な場 $h_i(t)$ に反映させるモデルである。

はじめに、コミュニティ m に属するがリンクのないノードペア (i, j) の集合を IC_m として求める。このような集合全体を IC で表すと、

$$IC = \bigcup_{m \in C} IC_m \quad (3)$$

となる。ここで、 C はコミュニティ全体の集合、 m をその要素とする。また、 M をコミュニティの総数とする。2つのノードの間に共通のコミュニティがいくつかあった場合、1つのノードのペアとして扱うため、 IC から重複するペアノードを削除した集合を IC' とする。

次に、 IC' から、 $\alpha | IC'$ | 個の要素をランダムに抽出後、リンク化し、隣接行列 A_{ij} のネットワークに加える。

ここで、 $|IC'|$ は集合 IC' の要素数、 α は、 $0 \leq \alpha \leq 1$ のパラメータである。また、このようにして生じるネットワークの隣接行列を B_{ij} とする。

コミュニティが影響する伝播モデルは、局所的な場 $h_i(t)$ を

$$h_i(t) = \sum_j^N B_{ij} \sigma_j(t). \quad (4)$$

とし、ノード i の状態を式5で表されるルールによって変化させるものである。

$$\sigma_i(t+1) = \begin{cases} \text{sgn}\{h_i(t)\} & \text{if } h_i(t) \neq 0 \\ \sigma_i(t) & \text{if } h_i(t) = 0 \end{cases} \quad (5)$$

4. 実験

4.1 ネットワーク生成のパラメータ

ネットワーク生成のパラメータは、 u, x, y, q_1, q_2 の5つと、ノード数 N の6つである。これを、ネットワーク生成パラメータ空間 $G_c(N, u, x, y, q_1, q_2)$ とする。本研究において、 u, y を0.1とし、 x に関しては、三井らと同様に $x/u = 4$ に固定して実験を行う。また、ノード数 N は10000とする。潜在リンク及び潜在コミュニティリンクの形成過程の違いが情報伝播に与える影響を調べるため、 q_1, q_2 を0.1~1.0まで0.1刻みで与え、計100種類のネットワークを形成する。

4.2 情報伝播モデルのパラメータ

伝播モデルのパラメータは、コミュニティの影響度を示す α と、初期状態における+1の割合を表す r_0 である。本研究では、 α を0から1まで0.1刻みで変化させる。予備実験において $\alpha=0.1$ までの変化率が大きいことがわかっているため、 $\alpha=0.1$ までは0.01刻みで変化させる。 r_0 は0.6に固定する。

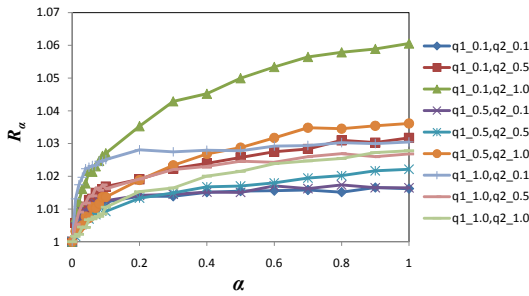


図 2: α 増加による収束値の変化

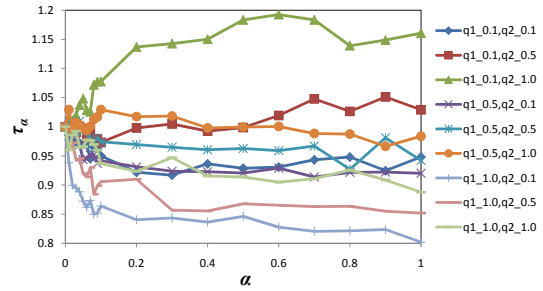


図 3: α 増加による収束値の変化

4.3 評価値

時間 t の $+1$ ($\sigma_i(t) = 1$) の割合を $r(t)$ とすると,

$$r(t) = \frac{1}{N} \sum_{i=1}^N \frac{(\sigma_i(t) + |\sigma_i(t)|)}{2} \quad (6)$$

となる。 r_0 は $r_0 = r(0)$ である。本研究においては情報伝播する際、十分時間が経過した際の $r(t)$ を $r(\cdot)$ として、これを収束値と呼び、 r で示す。収束値に至ったはじめの時刻を、収束時間と呼び、 T_c で示す。 r と T_c は、提案伝播モデルの場合、 q_1, q_2 と r_0, α に依存するため、 $r(q_1, q_2, r_0, \alpha)$ 、 $T_c(q_1, q_2, r_0, \alpha)$ と表記することにする。本研究では収束値と収束時間を観測する。1つのパラメータセットに対して200回の試行を行い、その平均をとる。

5. 実験結果

5.1 α 増加による影響

α の増加による、収束値、収束時間の変化を示す。図2に収束値増加率の結果の一部を示す。横軸は影響度 α 、縦軸は比率 $R_\alpha = r(q_1, q_2, 0.6, \alpha) / r(q_1, q_2, 0.6, 0)$ を示す。図3に収束時間増加率の結果の一部を示す。図の横軸は影響度 α 、縦軸は比率 $\tau_\alpha = T_c(q_1, q_2, 0.6, \alpha) / T_c(q_1, q_2, 0.6, 0)$ を示す。

図2から、いずれも α の増加にともなって収束値が増加していることがわかる。また、 α が0.1付近まではいずれも収束値の増加率が大きい、 α が0.1以上では、漸増を続けるものと、増加が抑制されているものが存在している。 q_2 が1の場合はいずれも、 α が0.1以上でも漸増し、 q_2 が0.1の場合は α が0.1以上で増加が抑制されている。

図3から、 α の増加にともなって、収束時間が減少するネットワークが多いことがわかる。一方で、収束時間が増加するネットワーク ($q_1 = 0.1, q_2 = 1.0$) が存在している。100種類のネットワーク全てについて調べたところ、 $q_1 < 0.3$ かつ $q_2 > 0.6$ の範囲のパラメータをもつネットワークにおいては、 α の増加にともなって、収束時間が明確に増加することがわかった。また、図3から、 $q_1 = 1.0, q_2 = 0.1$ のネットワークは、収束時間の減少率が大きいことがわかる。100種類のネットワーク全てについて調べたところ、 q_1 が大きく、 q_2 が小さいネットワークほど収束時間の減少率が大きいことがわかった。

5.2 q_1, q_2 の変化による影響

α を固定し、 q_1, q_2 を変化させた場合の収束値、収束時間の変化を示す。図4に $\alpha = 0.1$ の場合の収束値の結果を示す。奥行軸は q_1 、横軸は q_2 、縦軸は収束値を示す。図5に $\alpha = 0.1$ の場合の収束時間の結果を示す。奥行軸は q_1 、横軸は q_2 、縦軸は収束時間を示す。

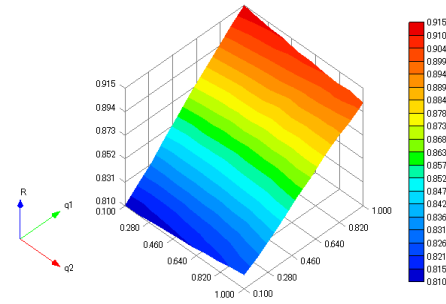


図 4: $\alpha = 0.1$ における収束値

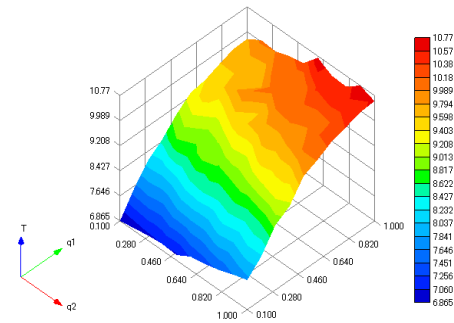


図 5: $\alpha = 0.1$ における収束時間

図4から、 q_1 が大きいほど収束値が増加する傾向があることがわかる。また、 q_1 が小さいときには、 q_2 が大きいネットワークほど収束値が増加している。一方で、 q_1 が大きいときには、 q_2 が小さいネットワークほど収束値が増加している。この傾向は、 α によらず表れた。また、 $r_0 = 0.6, \alpha = 0.1$ における収束値 $r(q_1, q_2, 0.6, 0.1)$ を q_1, q_2 の関数の近似式で表すと次式となる。

$$r(q_1, q_2, 0.6, 0.1) = 0.110q_1 + 0.015q_2 - 0.030q_1q_2 + 0.806 \quad (7)$$

図5から、 q_1, q_2 ともに大きいネットワークの収束時間が長く、 q_1, q_2 ともに小さいネットワークの収束時間が短いことがわかる。この傾向は、 α によらず表れた。また、 $r_0 = 0.6, \alpha = 0.1$ における収束時間 $T_c(q_1, q_2, 0.6, 0.1)$ を q_1, q_2 の関数の近似式で表すと次式となる。

$$T_c(q_1, q_2, 0.6, 0.1) = 7.34q_1 + 1.01q_2 - 3.82q_1^2 + 6.12 \quad (8)$$

5.3 統計量の仮定的変化

提案伝播モデルにおいては、仮想的に同じコミュニティ内のノード同士をリンクして情報伝播を行っている。そのため、

影響度 α の増加にともなって、クラスタ係数 $C(q_1, q_2, \alpha)$ 、平均経路長 $L(q_1, q_2, \alpha)$ は仮想的に変化する。 $\alpha = 0.1$ における仮想的な統計量はそれぞれ以下の多項式で近似することができた。

$$C(q_1, q_2, 0.1) = 0.329q_1 + 0.013q_2 - 0.073q_1^2 - 0.021q_1q_2 + 0.073q_1^2q_2 + 0.051 \quad (9)$$

$$L(q_1, q_2, 0.1) = -1.60q_1 + 0.042q_2 + 0.686q_1q_2 - 0.004 \quad (10)$$

6. 議論

パラメータ q_1 は、「潜在リンクのリンク時に、ネットワーク加入時に形成される潜在リンクが選択される確率」である。つまり q_1 が大きいネットワークは、「初期にリンクしたノードの周囲のノードと積極的にリンクを形成する」モデルから生じたものといえる。 q_1 が小さいネットワークは、「初期にリンクしたノードに固執せず、リンクをたどるようにして、徐々にリンクを形成する」モデルからのものである。

また、パラメータ q_2 は、「潜在コミュニティリンクのリンク時に、新規ノードもしくは新規コミュニティ形成時に形成される潜在コミュニティリンクが選択される確率」である。 q_2 が大きいネットワークにおいてコミュニティに加入するには、コミュニティを形成するノードと直接リンクしているか、ネットワーク加入時に、コミュニティに加入しているノードとリンクする必要がある。つまり、 q_2 が大きいネットワークは「コミュニティに加入しているノードは、そのコミュニティを形成したノードから近い距離にあるノードが多い」モデルから生じたものといえる。 q_2 が小さいネットワークは、「隣接ノードが加入しているコミュニティには、徐々に加入する」モデルからのものである。以上をふまえ、実験結果の考察をする。

5.1 節の、 α の増加による影響の結果から、陽なコミュニティ内での情報交換は、ネットワーク全体での収束値を増加させることがわかる。いずれのネットワークモデルにおいても、 $\alpha \leq 0.1$ での収束値の増加率が大きい原因は、伝播モデルにあると考えられる。本研究の伝播モデルでは、隣接ノードの2種の情報量が同じ場合には、次のステップにおいても同じ情報を保持する。コミュニティの影響が加わることによって、わずかも情報が多い方に遷移しやすくなったと考えられる。また、 q_2 が1の3つのネットワークでは、 α が大きいほど収束値が増加していた。このことから、 q_2 が大きい、つまりコミュニティを形成したノードから近い距離のノードがそのコミュニティに多く加入するモデルでは、コミュニティ内での情報交換が盛んになるほど収束値が増加する。一方で、 q_2 が0.1の3つのネットワークでは、 $0.1 \leq \alpha$ においては、収束値の増加が抑制されていた。隣接ノードが加入しているコミュニティに徐々に加入するモデルでは、一定以上コミュニティ内での情報交換しても、それ以上収束値はほとんど増加しない。

収束時間については、一部のネットワークを除き、コミュニティ内での情報交換はネットワーク全体の収束を早めた。一方、 $q_1 < 0.3$ かつ $q_2 > 0.6$ となるネットワークにおいては、コミュニティ内での情報交換によって収束時間は明確に増加した。このような q_1 が小さく、 q_2 が大きいネットワークは、コミュニティ形成に関しては、コミュニティ形成ノードから近い距離のノードがコミュニティに多く加入する一方、リンク形成に関しては、初期にリンクしたノードに固執せず、リンクをたどるようにして徐々にリンクを形成する。コミュニティ内での

情報交換によって、保持する情報が変化したノードの影響が、リンクをたどってネットワーク全体に行き渡るまでに時間がかかる構造になっていたと考えられる。

5.2 節の結果から、コミュニティ内での情報交換が一定の場合の、ネットワーク形成方法が与える影響について考察する。収束値の結果からは、 q_1 が大きいときには、 q_2 が小さいほど収束値が増加する。つまり、初期にリンクしたノードの周囲のノードと積極的にリンクを形成するネットワークにおいては、隣接ノードが加入しているコミュニティに徐々に加入すると、より多くのノードに情報を伝達することができる。一方で、 q_1 が小さいときには、 q_2 が大きいほど収束値が増加する。つまり、初期にリンクしたノードに固執せず徐々にリンクを形成するネットワークにおいては、コミュニティを形成したノードが、近い距離にあるノードをコミュニティに加入させるほど、より多くのノードに情報を伝達することができる。

収束時間の結果からは、 α が一定の場合、 q_1 によらず、 q_2 が小さいほど収束時間が短くなることがわかる。このことから、リンクの形成の仕方及び、コミュニティの影響度が一定の場合には、隣接ノードが加入しているコミュニティに徐々に加入するほど、ネットワーク全体の収束時間が減少するといえる。

また、5.2 節、5.3 節の結果から、 $r(q_1, q_2, 0.6, 0.1)$ 、 $T_c(q_1, q_2, 0.6, 0.1)$ 、 $C(q_1, q_2, 0.1)$ 、 $L(q_1, q_2, 0.1)$ はそれぞれ q_1, q_2 の関数として近似することができた。このことから、収束値や収束時間は、コミュニティの影響度を考慮した際の、仮想的な統計量であるクラスタ係数、平均経路長の関数として表せることが示唆される。

7. まとめ

本稿では、潜在リンクが生成機構の主要素であり、陽なコミュニティを有するネットワークモデルを提案し、ネットワーク上での情報伝播に与えるコミュニティの影響を調べた。その結果、情報伝播において、陽なコミュニティが与える影響と、リンク及びコミュニティ形成過程が与える影響を定性的に示した。それらの影響が何によってもたらされたかを明確に示していくことが課題である。

参考文献

- [1] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.U.: *Physics Reports*, Vol. 424, pp. 175–308 (2006)
- [2] Leicht, E.A., Newman, M.E.J.: *arXiv:physics*, No. 0709.4500v1 (2007)
- [3] Newman, M. E. J., Strogatz, S.H., and Watts, D.J.: *Physical Review E*, Vol. 64, No. 02618 (2001)
- [4] Noh, J. D., Jeong, H.-C., Ahn, Y.-Y. and Jeong, H.: *Physical Review E*, Vol. 71, No. 036131 (2005)
- [5] Vazquez, A.: *Physical Review E* 67, No. 056104 (2003)
- [6] 三井一平, 内田誠, 白山晋: 情報処理学会研究報告-知能と複雑系, Vol. 2006, No. 9, pp. 2840–2849 (2006)
- [7] 内田誠, 白山晋: 情報処理学会論文誌, Vol. 47, No. 9, pp. 2840–2849 (2006)
- [8] 林幸雄: 化学同人 (2007)