

ノード特性とコミュニティ情報を考慮した シナジーテックなネットワーク成長モデル

三好 栄次^{†1} 鈴木 育男^{†1}
山本 雅人^{†1} 古川 正志^{†1}

多くのネットワーク成長モデルの研究は、ノードの特性を考慮せず、トポロジ的構造のみに注目している。しかし実在する多くのネットワークは構成要素であるノードの特性にも影響を受ける。ノードの特性においても、ネットワークのコミュニティ構造形成に影響を与えていると考えられ、また、このコミュニティ構造はネットワークの成長に影響を与える。本研究では、既存のネットワーク成長モデルに対し、局所コミュニティ抽出方法を適用し、ネットワーク成長のルールにコミュニティ情報を付加した成長モデルの提案を行う。また、提案手法を適用したネットワークに対し、クラスタリング係数、パス長等の特徴量を観測し、ネットワークの成長過程について実際のネットワークデータと提案モデルのデータ比較を行う。

The synergistic network growth model in consideration of node characteristic and community information

EIJI MIYOSHI,^{†1} IKUO SUZUKI,^{†1}
MASAHITO YAMAMOTO^{†1}
and MASASHI FURUKAWA^{†1}

A lot of researches on network-growing models have not considered a characteristic of nodes. They have only paid attention to a topology structure. However, quite a few of existing networks are influenced by the characteristic of the nodes, which are the component of the network. It is thought that the characteristic of the nodes influences community structure appearance of the network and that the community structure influences a growing of network as well. In this study, existing networks are analyzed by the local community extraction. Then, a new network-growing model using the rule including the community information is proposed, while a new node is in turn added into the network and the network is growing. The growing process of the resultant network is examined by the conventional network feature value such as the average

shortest path, the clustering coefficient, etc. and is compared with ones of the real network.

1. 序 論

複雑ネットワークに関する研究の一分野として、ネットワークを構成している各要素間の結合構造に注目し、構成要素の局所的な結合方法を基にネットワークを生成することにより、ネットワークの構造の特徴を把握しようとする研究が数多くなされている。対象となるネットワークの要素によるネットワーク形成ルールの把握により、ネットワークを初期の状態から、成長させることが可能であり、さらにその先の状態の予測も可能となる。ネットワークの成長モデルの研究は、webのネットワーク形成の予測、人間同士の交友関係の予測など既存するネットワーク構造の今後の展開を予測する際に重要となる。これらのネットワークの成長モデルの研究には、社会ネットワークにおけるつながりの性質を捕らえようとするWSモデル¹⁾、ノードがネットワークへ連結する際、優先的選択を用いてスケールフリー性を表現したBarabási Albertモデル²⁾等が多くの研究で扱われている。また複雑ネットワークの構築、成長に関する研究³⁾⁴⁾⁵⁾も多くあり、実際に社会的ネットワークのデータを基に擬似的なネットワークを成長させ、解析する研究もある⁶⁾⁷⁾。ネットワーク成長モデルにおいて、ノード間でのリンクの張り方には、結合確率に次数を用いたり⁸⁾、隣接ノードとその近傍を利用している⁹⁾。また、コミュニティ構造を用いたノード間の相互作用を定義したモデルも研究されている¹⁰⁾。コミュニティ構造は、個人の趣味、嗜好等を基にネットワークを形成していると考えられる。実世界においては、ネットワークに存在するノードの部分集合がコミュニティ構造を形成し、ネットワークの成長過程では新規にネットワークに参入してきたノードが、コミュニティ情報を基にしてネットワークに新たにリンクを生成すると考えられる。しかし、これらのモデルでは、コミュニティの定義はランダムに結合したノード間と隣接しているノードが対象となっている。しかしながら、コミュニティを基にネットワークを構成するならば、ネットワークトポロジもコミュニティ構造を形成していると考えられる。コミュニティ構造の研究に関して、ネットワークのトポロジのみによってネットワークをコミュニティに分割する方法が多く提案されている¹¹⁾¹²⁾。例えば、リンクが密につな

^{†1} 北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

がったノードの集合コミュニティとして扱い、抽出する方法が提案されてきている。関連研究としては Girvan-Newman コミュニティ抽出¹³⁾がある。近年では大規模な実際のデータを対象とし分割を高速で行う Clauset Newman Moore アルゴリズム¹⁴⁾等の方法も研究されている。本研究では、既存のネットワーク成長モデルに対し、ネットワークトポロジを利用したコミュニティ抽出法を適用する。そして、ネットワークの成長のルールに、ネットワークトポロジから定義されるコミュニティの情報を付加した成長モデルの提案を行う。本論文の構成を以下に示す。第2章では比較対象となるネットワークについて説明し、第3章では関連研究を述べ、提案モデルについて説明する。第4章では数値計算実験について説明し、第5章では実験結果に対する考察を述べ、第6章で本研究をまとめる。

2. 対象とするデータ

本研究では、ソーシャルブックマーク「livedoor クリップ」^{*1}のデータを用いる。このサービスはネットワーク上に URL をタグをつけて保存しておくことができる。更に、各ユーザは他のユーザが保存した URL をチェックすることができる。ユーザはこの機能を用いて、趣味が似ているユーザや、有名なユーザが見ている URL を見たり、ブックマークが可能である。このブックマークに関するデータはネット上に公開されている。^{*2}

本研究では、このデータのうち以下の項目を用いてネットワークを定義する。

- ユーザ ID
- タグ
- タグ付けされた URL
- タグ付けしたユーザ

ユーザ間においてブックマークした URL が重複する数が多いほど趣味、嗜好が似ており、互いに影響を与えていると考えられる。そのため、今回ネットワークの定義はユーザが URL をブックマークする際、ユーザ間でブックマークした URL が一定数以上重複していればユーザ間にリンクを張ることにしている。各ユーザがタグをつけた URL に関して、ユーザ間でタグをつけた URL がどれだけ重複しているかの数を共参照数とする。

*1 <http://clip.livedoor.com/>

*2 <http://labs.edge.jp/datasets/>

3. 成長モデルの提案

本章では提案モデルの概要を説明する。提案モデルは既存モデルの拡張であるため、基となったモデルとコミュニティ構造の説明をした後、提案モデルの説明を行う。

3.1 CNN モデル

本研究に関連のある CNN モデル⁹⁾について述べる。CNN モデルとは、DEB モデル¹⁵⁾を拡張した成長モデルであり「友達の友達は友達」の関係に従ったネットワークモデルである。以下に CNN モデルのネットワーク成長過程を示す。以下において $V = \{v_i\}$ をノード集合とする

- (1) パラメータ $u (0 \leq u \leq 1)$ を設定する。
- (2) 各ステップにおいて、以下のアルゴリズムを実行する。
 - (a) 確率 $1 - \mu$ でネットワーク中に新しいノード v_j を V に追加し、すでにネットワーク中に存在するノード $v_i \in V$ をランダムにえらびノード v_i との間にリンクを張る。ノード v_j とノード v_i の近傍ノードの全てにポテンシャルリンクを設定する。
 - (b) 確率 μ でランダムにポテンシャルリンクを選び実際のリンクに変換する。ポテンシャルリンクとは、すでにネットワークに存在しているノード間で新しくリンクを張る場合、この2ノード間において共通のノードで張られる可能性があるリンクである。

CNN モデルは人間社会のネットワークにおける高密度なクラスタ構造を再現する仕組みとなっている。CNN モデルではランダムに選ばれたノードの近隣ノードとネットワークに新しく追加されたノードとの間のみポテンシャルリンクができる。これは紹介者の周りには共通の友人、コンテンツなどがある可能性が高いことを利用したもので、隣接ノードのリンク先とリンクを張る CNN モデルのようなノード間の相互作用は実際に多い。

3.2 局所的なコミュニティ

本研究で用いる、Clauset が提案した LocalModularity: R ¹⁶⁾を用いた局所コミュニティ抽出について説明する。局所的なコミュニティ解析では、あるノードを核として定めその核を中心にコミュニティを抽出し、ネットワーク全体の解析を必要としない。そのため計算量の面で大域的なコミュニティ分割より優れている。本研究ではコミュニティ構造を決定する R によってコミュニティ構造の決定を明確にする。以下に概要を示す。

グラフ全体を $G = (V, E)$ 、すでにコミュニティとなっているグラフの部分集合を $C \subseteq G$ 、

C に含まれず C に含まれるノードに対しリンクがある部分集合を $U \subseteq G$, C に含まれているノードにおいて U に含まれるノードとリンクがある部分集合を $B \subseteq C$ とする. B に関する隣接行列 $[B_{ij}]$ を以下のように定義する.

$$B_{ij} = \begin{cases} 1: \text{ノード } v_i \text{ と } v_j \text{ が繋がっており,} \\ \quad v_i \text{ か } v_j \text{ のどちらかは } B \text{ に属する} \\ 0: \text{それ以外} \end{cases} \quad (1)$$

B の隣接行列から以下の計算で LocalModularity である R を求める.

$$R = \frac{\sum_{ij} B_{ij} \delta(i, j)}{\sum_{ij} B_{ij}} = \frac{I}{T} \quad (2)$$

ここで $\delta(i, j)$ は, $v_i \in B$ かつ $v_j \in C$, または $v_i \in C$ かつ $v_j \in B$ のとき 1, それ以外の場合 0 となる. つまり, T は B の総リンク数, I は B のリンクにおいて U と繋がっていないリンクの数となる

Clauset による局所コミュニティ抽出¹⁶⁾ は, 1つの核となるノードから始め, LocalModularity を最大にするように C を構成する. アルゴリズムは以下ようになる.

- (1) $B = \emptyset, C = \emptyset$ とする.
- (2) 核となるノード v_0 を C に追加し B を v_0 とする. さらに v_0 の隣接ノード集合を U に追加する.
- (3) 以下の手順を C のノード数が決められた大きさになるまで繰り返す.
- (4) $v_j \in U$ に対して ΔR_j を計算する.
- (5) ΔR_j が最大となる v_j を C に追加し, v_j の隣接ノードのうち新規に現れたノードを U に追加する.
- (6) R を更新し, $R \leftarrow R + \Delta R_j$ にする.
- (7) v_j を B または C に追加し, v_j と隣接し, B に属するノードが C に追加されるか上記の条件により判定する.

ΔR_j は式 2 により導びいた式 3 で高速に計算できる.

$$\Delta R_j = \frac{x - Ry - z(1 - R)}{T - z + y} \quad (3)$$

ここで, x は v_j の B へのリンク数, y は v_j が C へ選ばれたとき T に加わる数 ($k_j - x$), z は v_j が C へ選ばれたとき T から減少する数である.

3.3 局所コミュニティを用いた生成アルゴリズム

前述したネットワーク全体の情報を用いたコミュニティ定義を適用したネットワーク生成モデルを用いると, ネットワークに新しく現れたノードはネットワークすべてのリンクの状態をすでに把握していることに等しい. そのためネットワーク未知の状態からの成長に適用はできない. 従って, 新しく現れたノードがネットワーク全体の情報を把握しないように考慮し, コミュニティ構造を抽出する必要がある. ここでは, Clauset の提案した LocalModularity を用いた局所コミュニティを CNN モデルに適用する. 提案するネットワークの生成方法は以下である.

- (1) 初期状態のネットワークをノード数を 1 とする
- (2) 以下の操作をノード数が一定数以上になるまで繰り返す.
 - (a) 確率 $1 - \mu$ でネットワーク中からランダムにノード v_i を選択し, 以下の手続きを実施する.
 - (i) ノード v_i を核とした局所コミュニティ(LC)を抽出する.
 - (ii) ネットワークにノード v_j を追加し, リンクを張る.
 - (iii) 抽出されたコミュニティに属するノードとノード v_j にポテンシャルリンクができる.
 - (b) 確率 μ でポテンシャルリンクを実際のリンクに変える.

提案方法の局所コミュニティ抽出は, コミュニティ抽出の出発点とするノードから遠く離れたノードの情報は必要なくなる. よって, ネットワーク全体のトポロジの影響を受けない. また, 同じノードで抽出されるコミュニティの範囲は, ネットワークが成長した場合でもノード付近のネットワークの構造が変わらない限り同じとなる.

この局所コミュニティの定義を用いて, 各ノード間において趣味や嗜好が近いノード同士が相互作用しながらネットワークを形成する過程を実現する.

4. 数値計算実験

4.1 実験目的

本章では, 提案モデルの妥当性を検証する. これは提案モデルと CNN モデルを livedoor クリップデータと比較して行う. それによって提案モデルが実データのネットワーク構造をどれくらい再現できるか検証する. 検証に用いる実データのネットワークの構成を複数用意することにより, 基となったモデルの CNN モデルとネットワーク構造がどの程度異なるかを確認する.

表 1 共参照ごとのノード数, リンク数

共参照数	ノード数	リンク数	クラスタリング係数	パス長
10	3351	203614	0.795	2.522
20	2133	90633	0.773	2.253
30	1747	54066	0.794	2.296
40	1480	37143	0.790	2.361
50	1272	28662	0.7966	2.459

表 2 終了時点のノード数, ポテンシャルリンクの顕在化確率 μ

共参照数	ノード数	μ
10	3351	0.984
20	2133	0.978
30	1747	0.961
40	1480	0.950
50	1272	0.946

4.2 実験手順

以下に示す実験を行う．livedoor クリップのデータの共参照数が 10,20,30,40,50 のネットワークを生成する．ネットワークを連結グラフの部分集合に分けたとき最大の次数が存在するネットワーク内で最大の連結成分であると考えられる．従って，それぞれの共参照ネットワークからネットワーク中において，最大の次数を持つノードが含まれている連結成分を取り出す．

各連結成分ごとの特徴量は表 1 となった．

表 1 のデータを基にポテンシャルリンクの顕在化確率および，終了条件を設定する．CNN モデル，提案手法を初期状態のノード数を 1 とし，各共参照ネットワークのノード数まで増えた時点で終了とする．各実験においてポテンシャルリンクの顕在化確率および，ネットワーク成長終了時点のノード数は表 2 となる．各実験におけるポテンシャルリンクの顕在化確率はネットワーク成長終了時に実ネットワークのリンクの密度に近くなるように設定した．これらのデータを基にネットワークを成長させる．

成長したネットワークから特徴量としてクラスタリング係数，平均パス長，ネットワーク直径を計算する．また，これらからネットワークのクラスタ性，スモールワールド性を確認する．提案モデルに関しては次数分布を計算し，その分布がべき乗側に従うかを観測し，スケールフリー性を確認する

4.3 実験結果

CNN モデル，提案モデルのパラメータ μ を変化させたネットワークを生成し，実ネット

表 3 各ネットワークデータの比較

	実データ	CNN	提案モデル		実データ	CNN	提案モデル
クラスタリング係数	0.795	0.490	0.277	クラスタリング係数	0.773	0.498	0.260
パス長	2.522	2.700	3.002	パス長	2.252	2.680	2.891
ネットワーク直径	7	12	9	ネットワーク直径	8	10	9

(a) 共参照数 10 顕在化確率 $\mu = 0.984$

(b) 共参照数 20 顕在化確率 $\mu = 0.978$

	実データ	CNN	提案モデル		実データ	CNN	提案モデル
クラスタリング係数	0.794	0.480	0.197	クラスタリング係数	0.791	0.474	0.185
パス長	2.296	3.003	2.984	パス長	2.361	3.126	3.059
ネットワーク直径	7	11	9	ネットワーク直径	7	11	9

(c) 共参照数 30 顕在化確率 $\mu = 0.961$

(d) 共参照数 40 顕在化確率 $\mu = 0.950$

	実データ	CNN	提案モデル
クラスタリング係数	0.498	0.482	0.181
パス長	2.459	3.076	3.080
ネットワーク直径	7	11	9

(e) 共参照数 50 顕在化確率 $\mu = 0.946$

ワークのデータとの比較を表 3 に示す．調べた特徴量はクラスタリング係数，平均最短パス長，ネットワーク直径および次数分布である．また実際のネットワークの成長過程において，ノード間のリンクの張られ方において何近傍先のノードにリンクが張られたかのデータを表 4 に示す．

5. 考察

表 4 の実データのリンクの結合から，ネットワークが成長していく際はノード間のパス長が近いノードにリンクが張られることが多いことがわかる．これは，ユーザが参照する URL は自分と隣接しているユーザの興味，嗜好の影響による相互作用を受けて決定されると考えられる．このユーザ間の相互作用をコミュニティ構造として利用するモデルを提案し数値計算を実施した．

この提案モデルは局所コミュニティによって顕在化確率が高くなると新規に追加された

表 4 リンクの結合先

近傍数	実データ	提案モデル	近傍数	実データ	提案モデル	近傍数	実データ	提案モデル
1	11297	3351	1	4003	2133	1	3005	1747
2	191714	55754	2	86555	25520	2	50965	9127
3	584	80628	3	101	38939	3	85	17300
4	15	56776	4	4	25223	4	10	13291
5	0	8027	5	0	3684	5	0	2670
6	0	405	6	0	205	6	0	227
7	0	12	7	0	7	7	0	10

(a) 共参照数 10

(b) 共参照数 20

(c) 共参照数 30

近傍数	実データ	提案モデル	近傍数	実データ	提案モデル
1	2277	1480	1	1774	1272
2	50965	6087	2	25537	4969
3	98	11621	3	72	9508
4	6	8279	4	5	6241
5	0	1679	5	0	1216
6	0	160	6	0	116
7	0	9	7	0	8

(d) 共参照数 40

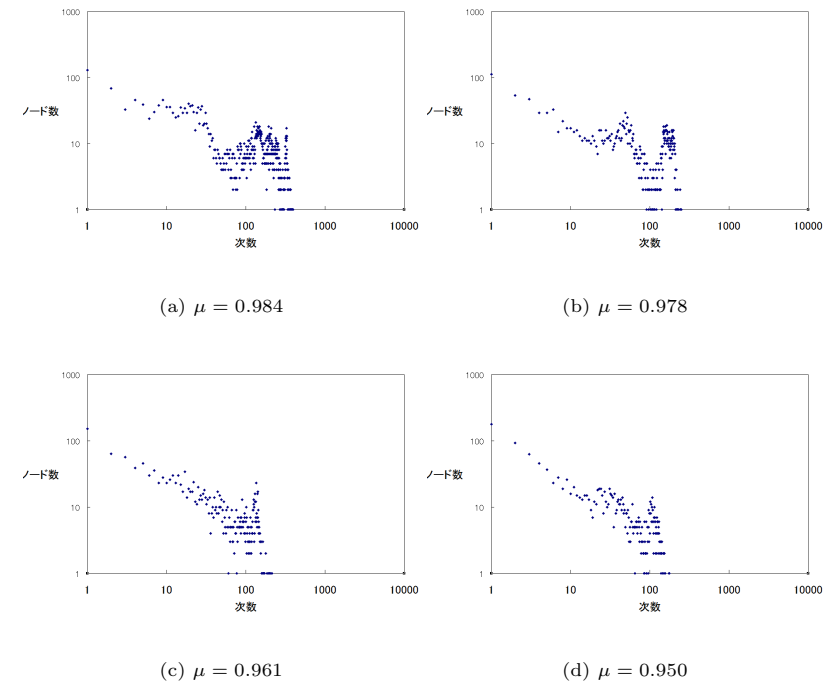
(e) 共参照数 50

ノードは多くのポテンシャルリンクを持つようになる。これは次数の大きなノードから局所コミュニティを抽出を開始した場合抽出される局所コミュニティのサイズは大きくなりやすく、次数の大きなノードは多くのノードとポテンシャルリンクを張ることになることが予測される。次数の小さなノードから始めた場合、*Localmodularity* によるコミュニティの定義によって、次数の大きなノードが局所コミュニティ内に多数存在することはあまりない。この局所コミュニティの抽出方法の特徴により、図 1 で示すように次数の中間値の次数のノードが増える結果となっている。

また、表 3 に示した平均パス長、ネットワーク直径が実データに近い数値を出していることから、提案モデルの局所コミュニティ定義によって決められるノードの間の距離は、実データにおけるユーザ間の距離と似ていると考えられる。

しかし、ポテンシャルリンクが実際のリンクに変わる際に選択されるリンクのパス長に関しては、実データではユーザ間のパス長が 2 のリンクが多いためほぼ CNN に近い動きを

図 1 次数分布

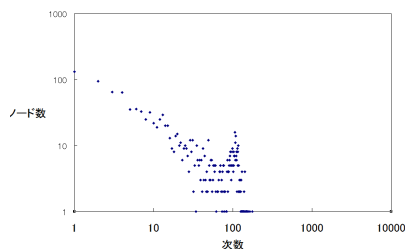


取っている。ネットワークを成長させる際提案モデルはポテンシャルリンクの選択は近傍の距離に関係なくランダムに選ばれているため、ユーザ間のパス長が 3 のリンクが多く選ばれている結果が多い。

6. 結 言

本研究では、ネットワークの生成モデルを拡張するために以下のことを行った。

- コミュニティ構造を考慮したネットワーク構築のモデル化を提案した
- 提案した生成モデルと基盤となった既存のネットワークの生成モデルとの比較を行い、構造の違いを考察した。
- モデルの比較に実際のネットワークデータを用いて、モデルの妥当性を検証した。



(e) $\mu = 0.946$

本研究では実ネットワークを模したネットワークの成長モデルではなく、既存の特徴に加え新たに「コミュニティ構造」を定義、既存モデルに導入した。このコミュニティ構造がユーザの趣味、嗜好および他のユーザからの影響を考慮した作用をネットワークの成長に与える。そのためコミュニティ構造を形成しながら成長するこのネットワークモデルは各ノードが他のノードの情報を利用し、自分の行動を決定するモデルである。

比較対象のネットワークのユーザの振る舞いに近いモデルを提案し、ネットワーク直径、パス長などの推移性は満たした。しかし、クラスタ性、結合におけるノード間のパス長によるリンク数の違いなどはまだ満たされていない。

しかし、本研究の提案モデルではポテンシャルリンクのユーザ間のパス長の長さ毎にパラメータを導入することも可能であり、それらのパラメータの追加および、パラメータの設定が今後の課題となる。

また、今回用いたネットワークの定義はユーザ間で重複している参照数が多いとそのユーザ同士の趣味、嗜好が近いという条件の下でネットワークを生成しているが、この条件を変えるとネットワークの構成が変わるためネットワークの条件の検討も今後の課題である。

参 考 文 献

- 1) Duncan.J.Watts: スモールワールド.
- 2) : Emergence of scaling in random networks, *ArXiv Condensed Matter e-prints* (1999).
- 3) Palla, G. and Lov' Multifractal network generator, *Proceedings of the National Academy of Sciences*, Vol.107, No.17, pp.7640–7645 (2010).

- 4) Chakrabarti, D. and Faloutsos, C.: Graph mining: Laws, generators, and algorithms, *ACM Comput. Surv.*, Vol.38, No.1, p.2 (2006).
- 5) 矢吹光佑, 矢吹太朗: 実ネットワークの大局的性質を再現する創発モデル, 情報処理学会研究報告 (2004).
- 6) Hofstad, R. V.D.: Random graphs and complex networks (2008).
- 7) : A survey of statistical network models, *ArXiv e-prints* (2009).
- 8) Bianconi, G. and Barabási, A.-L.: Bose-Einstein Condensation in Complex Networks, *Phys. Rev. Lett.*, Vol.86, No.24, pp.5632–5635 (2001).
- 9) Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Phys. Rev. E*, Vol.67, No.5, p.056104 (2003).
- 10) Mitsui, I., Uchida, M. and Shirayama, S.: An Evolving Network Model with a Community-Oriented Structure, *IPSI SIG Notes. ICS*, Vol.2006, No.2, pp.17–24 (2006-01-12).
- 11) Reichardt, J. and Bornholdt, S.: Statistical mechanics of community detection, *Phys. Rev. E*, Vol.74, No.1, p.016110 (2006).
- 12) Palla, G., Derenyi, I., Farkas, I. and Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society (2005).
- 13) Girvan, M. and Newman, M.E.: Community structure in social and biological networks., *Proc Natl Acad Sci U S A*, Vol.99, No.12, pp.7821–7826 (2002).
- 14) Clauset, A., Newman, M. E.J. and Moore, C.: Finding community structure in very large networks, *Phys. Rev. E*, Vol.70, No.6, p.066111 (2004).
- 15) Davidsen, J.: Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks, *Phys. Rev. Lett.*, Vol.88, No.12, p.128701 (2002).
- 16) Clauset, A.: Finding local community structure in networks, *Phys. Rev. E*, Vol.72, No.2, p.026132 (2005).