

コミュニティ情報を考慮した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モデルの研究が存在しているが、その多くはネットワークの構成要素であるノードの状態のみによりネットワークが成長する。しかし、実際のネットワーク成長はコミュニティによる影響も受ける。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、ネットワークの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生成によるコミュニティ構造の変化とともに、コミュニティ数の増減、クラスタリング係数などへの特徴量について時間経過による変化を観測し、ネットワークの成長過程について CNN モデルと提案モデルの比較を行う。

The network growth model that considered community information

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

There is a study of much network model, but, as for the most, a network grows up only by the state of the node that is the component of the network. However, the real network growth takes the influence by the community. In this study, We suggest the growth model that added community information to a rule of the growth of the network. In addition, with a change of the community structure by the network generation that applied suggestion technique. We observe a change by the progress about quantity of characteristic to numbers of increase and decrease of the number of the community, clustering origin in time, and we compare the suggestion model with CNN model about the growth process of the 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のルールを把握することでネットワークが初期の状態から、成長させることが可能であり、さらにその先の状態を予測ことも可能となる。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現在存在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。

また、形成されたネットワークにおいて、ネットワークの構造において重要な構成部分を抽出し、分析することにより、ネットワークの特徴を把握する研究もなされている。ネットワークの部分集合としてコミュニティを定義、抽出し、そこからネットワーク全体の特徴を考察する。こちらネットワークの構成を考察する上で重要な役割を果たす。

本研究では、既存のネットワーク成長モデルである CNN モデル¹⁾ に対し、コミュニティ抽出を適用し、ネットワークの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生成によるコミュニティ構造の変化とともに、コミュニティ数の増減、クラスタリング係数などへの特徴量について時間経過による変化を観測する。ネットワークの成長過程について既存手法と提案手法の比較を行う。

2. 関連研究

2.1 Connecting Nearest Neighbor モデル

CNN モデル¹⁾ は、DEB モデル²⁾ 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。このモデルは「スケールフリー性」、「クラスタ性」、「スモールワールド性」を満たすグラフを生成するアルゴリズムの一つである。

以下に CNN モデルの成長過程を示す。以下の説明において $V = v_i$, $E = e_{ij}$ をノード集合及びリンク集合とする

- (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。

^{†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

- (a) 確率 $1 - u$ でネットワーク中に新しいノード v_j を V 追加し,すでにネットワーク中に存在するノード $v_i, v_i \in V$ 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 の近傍ノード全てにポテンシャルリンクを設定する.
- (b) 確率 u 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友人からの紹介で張られる可能性があるリンクである.

しかし,このモデルは,コミュニティ構造を考慮しておらず,本研究の目的である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いう行動を行っていない.

2.2 Modularity

Aaron Clauset³⁾らが提案した方法であり,ネットワークの分割の良さを表す指標の Modularity である Q が高くなるような分割を目指すものである. Modularity はノード集合 V_i 内のノードの次数の総和とネットワークの全てのリンクとの割合を a_i ノード集合 V_i から V_m へのリンクの数とネットワーク中のリンクの総数 M の比を e_{lm} とすると

$$Q = \sum_{i \in 1 \dots I} Q_i = \sum_{i \in 1 \dots I} (e_{ii} - a_i^2) \quad (1)$$

となる.

2.3 Clauset Newman Moore アルゴリズム (CNM)

Clauset らのアルゴリズム³⁾は Modularity の変化量 ΔQ を用いることにより計算を高速化している.概要は以下のとおり

- (1) 初期状態は1ノードずつ別のコミュニティとしコミュニティごとに結合した場合の Q の差分

$$Q_{(i+j)} = Q_i + Q_j + 2e_{ij} - 2a_i a_j \quad (2)$$

- から $\Delta Q_{ij} = 2(e_{ij} - a_i a_j)$ を計算し最大のものを選ぶ.
- (2) ノード i, j が結合した場合ノード i, j と繋がっていたノード k に関して以下に従い ΔQ の更新をおこなう.

$$\Delta Q'_{jk} = \begin{cases} \Delta Q_{ik} + \Delta Q_{jk} : \text{if } k \text{ connected with } i, j \\ \Delta Q_{ik} - 2a_j a_k : \text{if } k \text{ connected with } i \\ \Delta Q_{jk} - 2a_i a_k : \text{if } k \text{ connected with } j \end{cases} \quad (3)$$

- (3) ΔQ がすべて負になるまで繰り返す

本研究では生成されたネットワークの評価において,ネットワーク中のリンクが高密度となっている部分集合を抽出する際 CNM 法を用いる. CNM 法によるコミュニティ抽出の評価をする際 Modularity を用いる.

2.4 局所的なコミュニティ抽出

本研究で用いる, Clauset が提案した LocalModularity⁴⁾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,核となるノード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 R によってコミュニティ構造を決定しているこの R によってコミュニティ構造の決定が明確となる.以下に概要を示す.

グラフ全体を $G = (N, E)$,すでにコミュニティとなっている部分を $C \subseteq G$, 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 $B \subseteq G$ とする.(図1参照) B に関する隣接行列を以下のように定義する.

$$B_{ij} = \begin{cases} 1 : \text{ノード } v_i, v_j \text{ が繋がっており,} \\ \quad v_i, v_j \text{ のどちらかは } B \text{ に属する} \\ 0 : \text{それ以外} \end{cases} \quad (4)$$

B の隣接行列から以下の計算で LocalModularity である R が求まる.

$$R = \frac{\sum_{ij} B_{ij} \delta(i, j)}{\sum_{ij} B_{ij}} = \frac{T}{I} \quad (5)$$

ここで $\delta(i, j)$ は, $v_i \in B$ かつ $v_j \in C$, または反対のとき 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⁴⁾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になる.

- (1) 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.
- (2) 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.
- (3) $v_j \in U$ に対して ΔR_j を計算する.

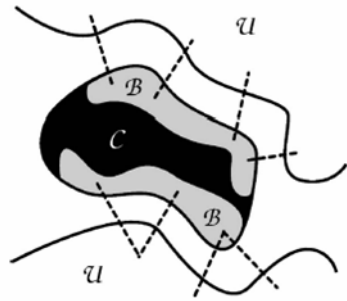


図1 局所コミュニティの概要⁴⁾

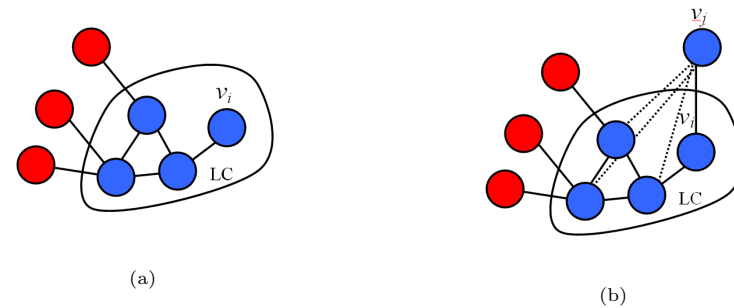


図2 提案手法の概要

(4) 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ものを U に追加する.

(5) R と B を更新する.

ΔR_j は式5により導かれる式を利用して式6で高速に計算できる.

$$\Delta R_j = \frac{x - Ry - z(1 - R)}{T - z + y} \quad (6)$$

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 から減る数である.

3. コミュニティ要素を加えたネットワークの生成

本研究では CNN モデルのネットワークの成長方法にコミュニティの概念を取り入れる. CNN モデルは基本的な3つの特徴をそそえているためモデルを改良しても, ある程度現実のネットワークの特徴に近いネットワークが生成できると考えられる. CNN モデルにおいて, ポテンシャルリンクが発生する範囲を隣接ノードからコミュニティに広げることにより同好の集団とは知り合いになりやすくなる条件を付加する.

Modularity をネットワーク生成に導入するには以下のような問題が生じる. GN 法,⁵⁾ NF 法⁶⁾ などネットワーク全体のコミュニティ分割を行う手法は, ネットワーク中に存在するノードのリンク状態すべてがどうなっているかを把握する必要がある. 例えば GN 法ならば, リンクを切断する際, リンクの媒介性を必要とする. これは各ノードの最短経路をすべて把握する必要がある. コミュニティを決定するまでにすべてのネットワーク状態を, 新し

く追加されたノードが把握している状態である.

この場合どのコミュニティにポテンシャルリンクが発生するかを決める前に, 新しく追加されたノードはネットワーク全体を把握している. そのためポテンシャルリンクの発生範囲はネットワークの部分集合であるコミュニティではなくネットワーク全体に発生する. また, 現実のネットワークを考えると, コミュニティは1ノードに対して2つ以上のコミュニティに属する場合が非常に多い. 例えばある人は, 大学に通学し, サークルに所属し, かつ4人家族である場合, 少なくとも人間関係のネットワークにおいては3つ以上のコミュニティに属している. しかし, ネットワーク全体のコミュニティ分割を用い, コミュニティを定義した場合には, 各ノードが属することができるコミュニティは必ず1つになってしまう.

3.1 局所コミュニティを用いた生成アルゴリズム

先ほども述べたが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把握していることに等しい. そのためネットワークの生成に適用はできない. 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 モデルに適用した. ネットワークの生成方法を以下のように提案する.

(1) 初期状態を決定する.

- (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 $1 - u$ でネットワーク中からランダムにノード v_i を選択する．
- (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．図 2.4(a)
 - (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(図 2.4(b) 点線部)ができる．
- (b) 確率 u 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．

今回の提案モデルでは，ネットワークの生成過程において

- 大きなコミュニティはより大きなコミュニティになり，小さなコミュニティは余り変化が無い状態
 - ポテンシャルリンクの発生は 2 ノード先のノード間同士だけではないためクラスタ性の犠牲
 - ネットワーク中のリンクの増加に伴う局所コミュニティの抽出範囲拡大
 - 局所コミュニティ抽出範囲拡大によるネットワーク直径，平均最短パスの減少
- などが考えられる．

4. 数値計算実験

4.1 ネットワークの生成

本研究の実験手順について説明する．CNN モデル，提案手法を初期状態のノード数を 1 とし，ノード数が 10000 になるまでネットワークを生成する．このとき提案手法において局所的コミュニティを抽出する範囲は最初に $\Delta R < 0$ となるまでとしている．それぞれの生成手法に対しノード数が 500 増えるごとに $Modularity : Q$ ，クラスタリング係数，パス長，直径，を計算．

4.2 実験結果

CNN, 提案手法による実験結果を示す．

図 3 は顕在化確率 μ を変化させたときの $Modularity$ の変化である．提案手法は CNN モデルと比べて $Modularity$ が高く CNM 法で分割されたコミュニティ内のノード同士の

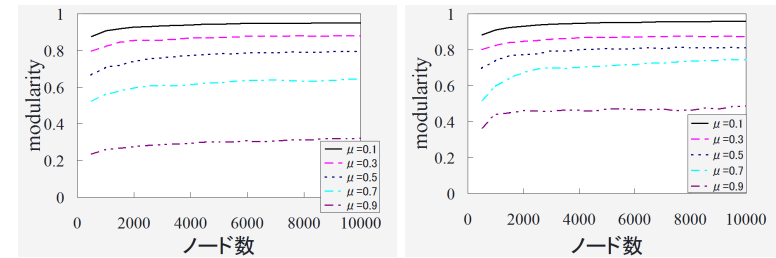


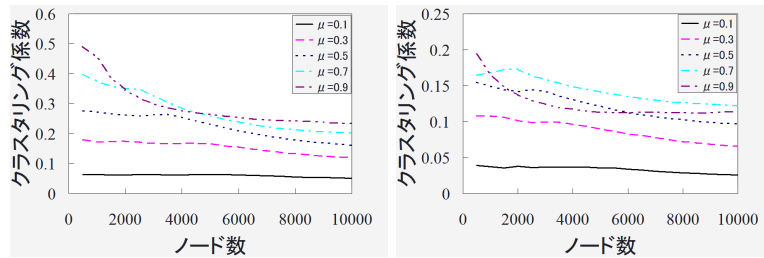
図 3 $Modularity$ の推移 (a):CNN モデル (b):提案手法

繋がりは強いものと考えられる．

またポテンシャルリンクの顕在化確率が増加した場合における $Modularity$ の減少の度合いは比較的低い．これは CNN モデルが 2 次近傍ノード同士にリンクを張るため，リンク数が増えるほど $Modularity$ で定義されるコミュニティ構造があいまいになってしまう．そのため $Modularity$ の値が低くなっていると考えられる．しかし，提案手法においてはポテンシャルリンクが顕在化するノードは，部分グラフの境界におけるコミュニティ内外へのリンク数の比で求められるコミュニティ定義によって決められる．したがってコミュニティ内リンク密度でコミュニティ分割が求められる CNM 法においては CNN モデルと比べ同じポテンシャルリンクの顕在化確率において $Modularity$ が高くなっている．

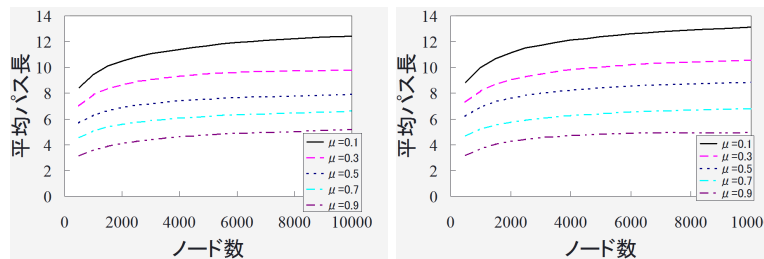
クラスタリング係数の推移は先に述べた $Modularity$ とは反対に提案手法では低い値となっている．これは先ほどの理由と同じで，CNN モデルが 2 次近傍のノードと，リンクを張っているためにクラスタリング係数が高くなっているのに対し，提案手法はポテンシャルリンクの存在範囲がコミュニティとなっており平均次数にもよるが 3 次近傍以上のノードとリンクが張られる可能性が高い．そのため $Modularity$ は高くなるがクラスタリング係数が低いといった現象が起こっている．

図 5 はネットワークの平均パス長の推移である．ネットワークが初期の段階，ポテンシャルリンクの顕在化確率が低い場合，CNN モデルにおいては，ハブとなるノードにリンクが集まりやすくなるため，提案手法と比べパス長が短い．しかし顕在化確率が高くなるにつれ，局所コミュニティの抽出範囲が増えていくため，CNN と比べて提案手法の平均パスは



(a) (b)

図 4 クラスタリング係数の推移 (a):CNN モデル (b):提案手法



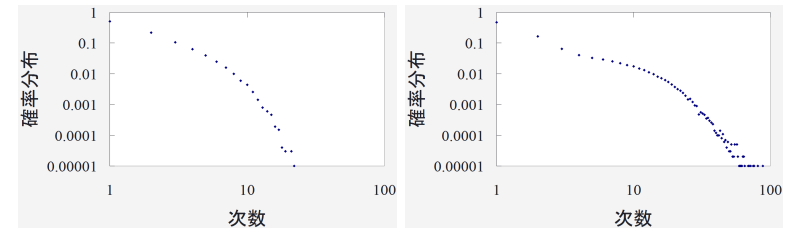
(a) (b)

図 5 平均パス長の推移 (a):CNN モデル (b):提案手法

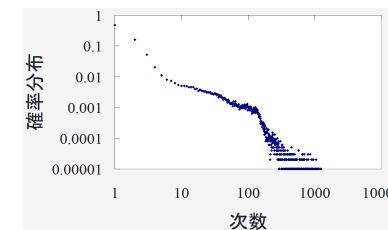
短くなる．これは提案手法においてより遠くのノードとリンクを張りやすくなるため平均パス長が短くなる．

図 6 は顕在化確率を変更した場合のネットワークの最終状態における次数の確率分布である．顕在化確率が低いと、分布に大きな違いは見られない．しかし、顕在化確率が高くなるにつれて提案手法の確率分布はべき分布にはならず、次数がある値を超えると存在する確率が急に下がる．

図 7 はポテンシャルリンクの顕在化確率が $\mu = 0.9$ において、ネットワークにノードが



(a) (b)



(c)

図 6 ポテンシャルリンクの顕在化確率が異なる状態での次数分布の比較 (a) $\mu = 0.1$ (b) $\mu = 0.5$ (c) $\mu = 0.9$

追加される際、追加されたノードとのポテンシャルリンクが張られる局所コミュニティのサイズの分布である．ポテンシャルリンクが発生するコミュニティのサイズの多くはノード数 1000 以下であり、その多くが 10 程度のかかなり小規模なコミュニティ構造となっている．

4.3 考 察

実験においてパラメータの変更および、手法との比較を行ったが、それぞれの比較において明確な良し悪しは無く．また、比較として用いる実際のネットワークにおいて、評価は異なる．

しかし、今回のコミュニティ構造を考慮したモデルにおいて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

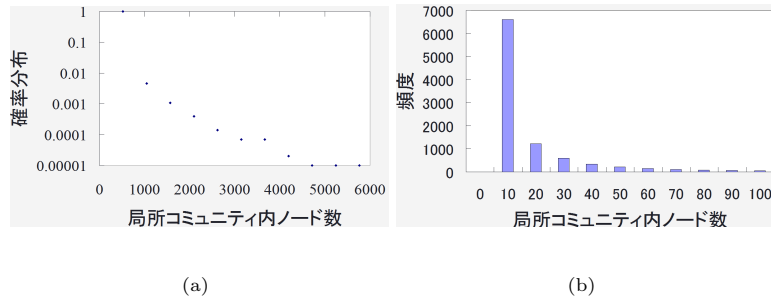


図7 ポテンシャルリンクの顕在化確率 $\mu = 0.9$ における局所コミュニティの抽出範囲

ンシャルリンクを張ることになる。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式によって、次数の大きなノードが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6で示すようにポテンシャルリンクの顕在化確率が高い場合において、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。

5. 結 言

本研究では、ネットワークの生成モデルを拡大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。
- コミュニティ構造をネットワークの生成に用いることによりネットワーク中のリンク密度における、パラメータの変化を解析した。

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が、現実のネットワークが持つ特徴を満たせていない部分が多く、課題としてパラメータ、成長ルールの追加およびコミュニティ構造の変更などが挙げられる。

参 考 文 献

- 1) Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Phys. Rev. E*, Vol.67, No.5, p.056104 (2003).
- 2) Davidsen, J., Ebel, H. and Bornholdt, S.: Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks, *Phys. Rev. Lett.*, Vol.88, No.12, p.128701 (2002).
- 3) 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).
- 4) Clauset, A.: Finding local community structure in networks, *Phys. Rev. E*, Vol.72, No.2, p.026132 (2005).
- 5) 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).
- 6) Newman, M. E.J.: Fast algorithm for detecting community structure in networks, *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Vol.69, No.6 (2004).