



Title	コミュニティ構造を用いたネットワーク成長モデル
Author(s)	三好, 栄次; 鈴木, 育男; 山本, 雅人 他
Citation	情報処理北海道シンポジウム講演論文集, 2009, 37-40
Issue Date	2009-10-03
Doc URL	https://hdl.handle.net/2115/51058
Rights	ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。
Type	journal article
File Information	Hokkaidosympo20093740.pdf



コミュニティ構造を用いたネットワーク成長モデル

三好栄次 鈴木育男 山本雅人 古川正志

(北海道大学大学院情報科学研究科)[†]

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。その多くは、ネットワーク中の次数分布やクラスタリング係数などの特徴量によりネットワークを分類することに主眼が置かれている。

それ以外にも複雑ネットワークの研究分野では、構成要素の局所的なルールに基づいたネットワークの成長モデルの研究も盛んにおこなわれている。この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といった現存するネットワークの将来予測をするうえで重要な研究である。これまでの、ネットワークの成長モデルでは、成長段階でネットワーク全体を把握するような情報（ネットワーク中の全ノードの次数など）を付加することによる、追加ノードのリンク先を決定する手法が多い。

しかし、現実のネットワークにおいては、その規模も大きく追加される瞬間のネットワーク全体の情報を把握できることは不可能に近い。この意味で、これまでの成長モデルは、現実のネットワークの成長を模倣したモデルとは言い難い。

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。

- (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。

実際の世界において、ネットワークに存在するノードがコミュニティ構造を形成する場合は多い。また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、このモデルでは、コミュニティ構造を考慮した成長モデルではない。

2.2 ネットワークの特徴量

本研究で用いる、Clausetが提案したLocalModularity: $R[3]$ を用いた局所コミュニティ抽出について説明する。

本研究においてコミュニティとは、密なリンクで繋がるノード集合であり、関係性の強いノード集合によって構成される。

ネットワークの構造においてこのコミュニティを抽出し、分析することにより、ネットワークの特徴を把握する。このコミュニティ構造を用いることにより、情報推薦などのサービスに応用できる。

局所的なコミュニティ抽出では、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 に含まれていないノード集合において、 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。 B に関する隣接行列を以下のように定義する。

miyoshi@complex.eng.hokudai.ac.jp

[†] 札幌市北区北14条西9丁目北海道大学大学院情報科学研究科

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

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

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

ここで $\delta(i, j)$ は, $v_i \in B$ かつ $v_j \in C$, または $v_j \in B$ かつ $v_i \in C$ 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 [3] 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のようになる.

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.
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 2 により導かれる式を利用して式 3 で高速に計算できる.

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

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 から減る数である.

本研究においては, コミュニティを意図的に決めるのではなく, 数式に従いネットワークトポロジから決めることでコミュニティの設定の簡略にした.

他にも GN 法, [4] NF 法 [5] などネットワーク全体のコミュニティ分割を行う手法があるが, これらの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ため, 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ことに等しいためネットワークの成長モデルには適用できない.

3 提案モデル

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 つの特徴をそろ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.

ここでは, Clauset の提案した LocalModularity を用いた局所コミュニティ抽出を 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.
 - 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 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.

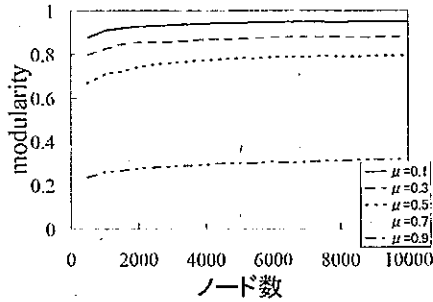
4 数値計算実験

4.1 ネットワークの生成

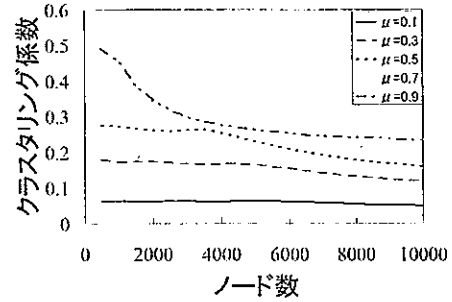
本研究の実験手順について説明する. CNN モデル, 提案手法を初期状態のノード数を 1 とし, ノード数が 10000 になるまでネットワークを生成する. このとき提案手法において局所的コミュニティを抽出する範囲は最初に $\Delta R < 0$ となるまでとしている. それぞれの生成手法に対しノード数が 500 増えるごとに Modularity: Q , クラスタリング係数, パス長, 直径, を計算.

4.2 実験結果

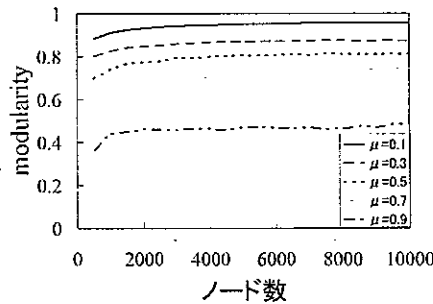
CNN, 提案手法による実験結果を示す.



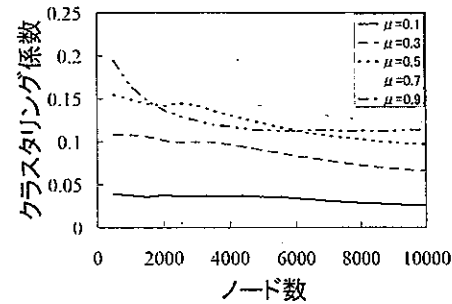
(a) CNN モデル



(a) CNN モデル



(b) 提案手法



(b) 提案手法

Fig. 1 Modularity の推移

Fig. 2 クラスタリング係数の推移

図1は顕在化確率 μ を変化させたときのModularity[6]の変化である。Modularityとはネットワークの分割の良さを表す指標の1つである提案手法は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の減少の割合は比較的低い。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。

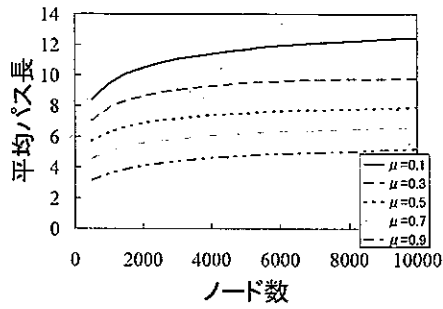
図2はクラスタリング係数の推移である。クラスタリング係数の推移は提案手法では低い値となっている。これは、CNNモデルが2次近傍のノードと、リンクを張っているためにクラスタリング係数が高くなっているのに対し、提案手法はポテンシャルリンクの存在範囲がコミュニティとなっており平均次数にもよるが3次近傍以上のノードとリンクが張られる可能性が高い。そのためクラ

スタリング係数が低くなる現象が起こっている。

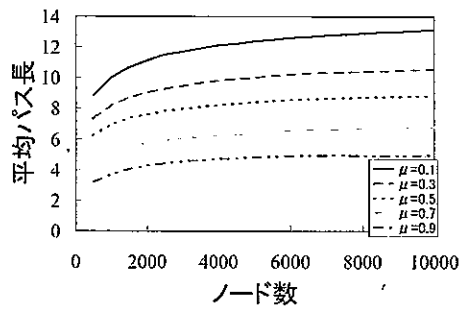
図3はネットワークの平均パス長の推移である。ネットワークが初期の段階、ポテンシャルリンクの顕在化確率が低い場合、CNNモデルにおいては、ハブとなるノードにリンクが集まりやすくなるため、提案手法と比べパス長が短い。しかし顕在化確率が高くなるにつれ、抽出される局所コミュニティのサイズが増えていくため、CNNと比べて提案手法の平均パスは短くなる。これは提案手法においてより遠くのノードとリンクを張りやすくなるため平均パス長が短くなる。図4は顕在化確率を変更した場合のネットワークの最終状態における次数の確率分布である。顕在化確率が低いと、分布に大きな違いは見られない。しかし、顕在化確率が高くなるにつれて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。

4.3 考察

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。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



(a) CNN モデル



(b) 提案手法

Fig. 3 平均パス長の推移

がコミュニティ内に多数存在することはほとんどない。

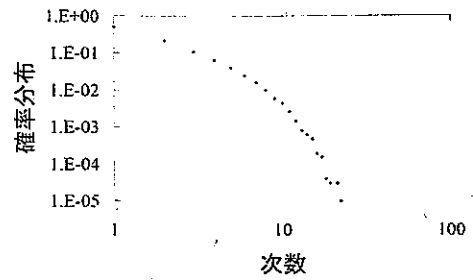
この局所コミュニティの抽出方法の特徴により、図 4 で示すようにポテンシャルリンクの顕在化確率が高い場合において、次数の中間値の次数のノードが増える結果となっている。

5 結言

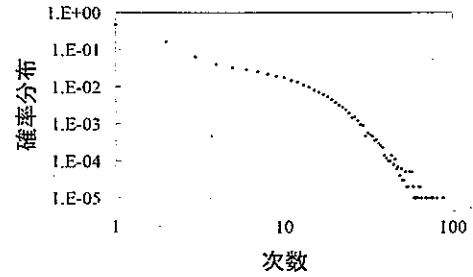
本研究では、ネットワークの生成モデルに関して、以下のことを行った。

- 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。

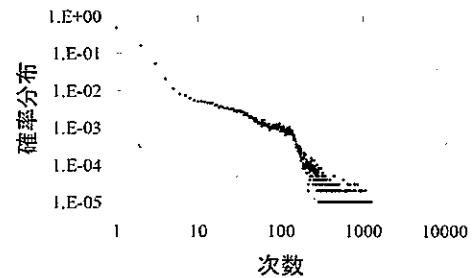
本研究では実ネットワークを模したネットワークの成長モデルではなく、既存の特徴に加え新たに「コミュニティ構造」を定義、既存モデルに導入したが、現実のネットワークが持つ特徴を満たせていない部分が多く、課題としてパラメータ、成長ルールの追加およびコミュニティ構造の変更などが挙げられる。



(a) $\mu = 0.1$



(b) $\mu = 0.5$



(c) $\mu = 0.9$

Fig. 4 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

参考文献

- [1] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, May 2003.
- [2] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks, Mar 2002.
- [3] Aaron Clauset. Finding local community structure in networks, Aug 2005.
- [4] M. Girvan and M. E. Newman. Community structure in social and biological networks., June 2002.
- [5] M. E. J. Newman. Fast algorithm for detecting community structure in networks, 2004.
- [6] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks, Dec 2004.