

コミュニティ構造を考慮した貪欲なネットワーク生成手法

Greedy network model considering community structure

臼井 翔平*¹ 鳥海 不二夫*¹
Shohei Usui Fujio TORIUMI

*¹ 東京大学大学院工学系研究科
School of Engineering The University of Tokyo

The complex networks are used in various fields. It is difficult to prepare a large quantity of various networks. Making datasets is necessary. Our goal is producing arbitrary types of networks from the viewpoint of particular features. In this paper, We introduce our model Greedy Growth Model(GGM), and extend GGM focusing on community structures. We use new feature, which is the number of community. We experiment with networks produced by several network models. Producing network with the same number of community is succeeded, but it is not said that these are same community structure. It is necessary to consider the modularity.

1. はじめに

近年, 多くの現象がネットワークを介する事によって理解されるようになり, ネットワーク分析が多くの分野で注目されている. 例えば, 感染症は人の接触関係の上で拡散され, 交通渋滞は道路網の上で発生し, 電気は電力網の上を流れる. したがって, これらの現象を考える際, 多くの研究者はネットワークを扱う事になる.

しかし, これらの研究において, ネットワークの構造は軽視されている. 多くの場合, Erdős-Rényi モデル [1] によるランダムネットワークか, Watts-Strogatz モデル [2] によるスモールワールドネットワークか, Barabási and Albert モデル [3] によるスケールフリーネットワークが用いられる. これらのモデルは 10 年以上前に提案されたモデルであり, 現実のネットワークを説明するにはシンプルすぎる. 一方で, 現実のネットワークを大量に集める事は困難である. したがって, 様々な構造のネットワークが大量に含まれたデータセットが人工的に生成する必要がある. 現在提案されているネットワーク生成モデルから作られるネットワークは似通った特徴を持つため, ネットワークの多様性を確保することは困難である. 本研究の目的は, 任意の構造を持つネットワークを生成できるネットワーク生成モデルの提案である.

任意の構造を持つネットワーク生成手法として, 我々は Greedy Growth Model(GGM) を提案した. GGM は, 目標特徴を予め設定し, その特徴と同じ特徴を持つネットワークを生成するモデルである. 本研究では, GGM にコミュニティに関する特徴を導入し, コミュニティ構造を考慮した任意の特徴を持つネットワークを生成する.

2. 提案モデル

2.1 概要

本研究では, Greedy Growth Model(GGM) にコミュニティに関する特徴を導入し, コミュニティ構造を考慮した任意の特徴を持つネットワークを生成する. 本稿ではまず, 本研究で用いる特徴を説明し, その後 GGM の解説を行う.

2.2 ネットワーク構造の特徴

本研究では, ネットワークの構造特徴を以下に示す特徴によりネットワークを定量的に表現する.

- (1) 平均経路長 L [2]
- (2) クラスタリング係数 C [2]
- (3) 次数相関 r [4]
- (4) 次数分布 $p(k)$
- (5) コミュニティ数 s

2.2.1 次数分布 : $p(k)$

本研究では, 次数分布 $p(k)$ を β 分布 $q(k)$ によって与え, これを特徴とする.

$$q(k) = \frac{x^\alpha(1-x)^\beta}{B(\alpha, \beta)} \quad (1)$$

$$B(\alpha, \beta) = \int t^{\alpha-1}(1-t)^{\beta-1} \quad (2)$$

ここで, α 及び β は β 分布のパラメータである. また, x は $0 \leq x \leq 1$ であるため, x を以下の式で与える.

$$x = \frac{k}{k_{max}} \quad (3)$$

β 分布 $q(k)$ は確率密度関数であるため, 以下の処理を行う.

$$q'(k) = \frac{q(k)}{\sum_{k'} q(k')} \quad (4)$$

k_{max} は以下の式を満たす最大の値とする.

$$\sum kNq'(k) \leq 2M \quad (5)$$

なお, N はノード数であり, M はリンク数である. 本研究では, EM アルゴリズムを用いて, ネットワークの次数分布から β 分布を推定する.

β 分布は α 及び β を様々な値にする事で様々な形状の分布を実現する. この β 分布の形状に次数分布 $p(k)$ を近づける事で, 様々な次数分布 $p(k)$ を持つネットワークを生成する.

連絡先: 東京大学工学系研究科

〒 113-8654 東京都文京区本郷 7-3-1

E-mail: usui@crimon.q.t.u-tokyo.ac.jp

2.2.2 コミュニティ数 : s

コミュニティ数 s は、ネットワークをコミュニティ分割した時のコミュニティの数である。本研究では、コミュニティ分割を拡張 louvain 法 [5] によって行う。これにより、抽出されたコミュニティ数を特徴とする。

2.3 Greedy Growth Model

GGM では、ネットワーク成長モデルのプロセスを用いて、ノードの追加とリンクの追加を繰り返し行い、ネットワークを成長させていく。毎ステップのノードの追加、またはリンクの追加の際に複数のネットワーク候補を生成する。その候補の中から評価関数によって最良のものを選択し、これを次のステップのネットワークとして採用する。これを繰り返し行うことによって要求した特徴に近づきながらネットワークを成長させる。

2.3.1 ネットワーク成長モデル

一般的に、ネットワーク成長モデルでは、毎ステップ以下のいずれかの操作を行い、ネットワークを徐々に成長させていく。

- ノードの追加
- リンクの追加

本モデルではノードの追加を最初に行い、ノード数が必要数に達した時点で、リンクの追加を行う。リンク数が必要数まで達した時点で終了する。

2.3.2 候補ネットワークの生成

ノードの追加とリンクの追加のそれぞれにおいて、候補ネットワークを生成する手法について解説する。

ノードの追加

ノードの追加では、ネットワークに新規ノードを 1 つ追加する。さらに既存のノードの中から 1 つのノードを選択し、追加した新規ノードと選択した既存ノードの間にリンクを 1 つ生成する。この際既存のネットワークからランダムにノードを選択する。選択されたノードと新規ノードの間にリンクを生成したネットワークを候補ネットワークとする。この操作を m 回繰り返す事によって、 m 個の候補ネットワークを生成する。

リンクの追加

リンクの追加では既存のノードの中から 2 つのノードを選択し、その 2 つのノードの間にリンクを 1 つ生成する。まず、既存のノードから 1 つのノードをランダムに選択する。次に、2 つ目のノードはランダム選択か、一つ目のノードから 2 ステップ先のノードの中からランダムで選択する。2 つの選択方法は同じ確率で行われるようにする。この操作によって選択された 2 つのノード間にリンクを結ぶ事によって候補ネットワークを生成する。この操作を m 回繰り返す事によって、 m 個の候補ネットワークを生成する。

2.3.3 評価関数

生成した候補ネットワークの中から要求指標値に近いネットワークを選択するためには、ネットワーク間の指標値の差を定量的に評価する必要がある。そこでまず、使用する指標に関する距離を以下のようにそれぞれ定義する [6]。

$$D_L = \left(\frac{L_i - L_a}{\sigma_L}\right)^2, D_C = \left(\frac{C_i - C_a}{\sigma_C}\right)^2$$

$$D_r = \left(\frac{r_i - r_a}{\sigma_r}\right)^2, D_m = \left(\frac{m_i - m_a}{\sigma_m}\right)^2$$

また、生成中のネットワークの次数分布 $p(k)$ と β 分布 $p'(k)$ との距離をユークリッド距離によって定義する。

$$D_{p(k)} = \frac{\sum (p(k) - q'(k))^2}{\sigma_{p(k)}^2}$$

ただし、 $L_i, C_i, r_i, p(k)$ はそれぞれ生成中のネットワークの指標値であり、 $L_a, C_a, r_a, q'(k)$ はそれぞれ要求指標値である。各項の分母 $\sigma_L, \sigma_C, \sigma_r, \sigma_{p(k)}$ は各指標の標準偏差を表す。標準偏差は本章の検証に用いるデータセットから算出する。さらに、それぞれの指標の距離の和をネットワーク間相異度 D とする。

$$D = \sum_j D_j \quad (6)$$

D の値が小さい程、類似した特徴を持つネットワークであると言える。

3. 評価実験

3.1 概要

本節ではいくつかのデータセットを用いて、無向ネットワークに対する提案手法の有効性を検証する。既存手法で生成したネットワークから特徴ベクトルを抽出し、複数の特徴ベクトルを表現する事によって、提案手法が様々な特徴のネットワークを生成可能である事を示す。

3.2 既存モデルによる検証

ここでは、既存のネットワーク生成モデルによって生成されたネットワークに対して、評価実験を行い、提案手法の有効性を検証する。

3.2.1 データセット

本節では、以下のネットワーク生成モデルを利用し、ネットワークを生成する。

- BA モデル [3]
- CNN モデル [7, 8]
- Fitness モデル [9, 10]
- Goh のモデル [11]
- Holm-Kim モデル [12]

なお、ノード数は 1000、リンク数は 10000 とする。各モデルで 10 個ずつネットワークを生成し、そのネットワークを生成できるかどうかを各生成手法の提案手法で検証する。また、一つのネットワークにつき 3 回ずつ試行を行い、平均を算出する。これによって様々なネットワークを生成可能である事を示す。なお、ここでは $m = 30$ とする。

3.2.2 実験結果

生成したネットワークと対象ネットワークとのベクトル間距離を求め、図 1 に閾値 T よりも小さいネットワークの数を算出した。これより、Gho モデルで生成されたネットワークの再現については、0.2 より小さい距離で全てのネットワークを生成できている事がわかる。また、Fitness モデル、HolmKim モデル、BA モデルによって生成したネットワークは 0.4 以下の距離で全てのネットワークを生成できている。例として表 1 に、BA モデルで生成したネットワークと、それを目標として Greedy に生成したネットワークの特徴量を示す。この 2 つのネットワークはわずかに特徴の差はあるものの、大きくは離れていない事がわかる。図 2 に、2 つのネットワークの次数分布をそれぞれ示す。次数分布においても、わずかに特徴の差はあるものの、大きく離れていない事がわかる。次に、コミュニティについて検討する。2 つのネットワークのコミュニティ数を見ると共に 11 であり、コミュニティ数に差はない。それぞ

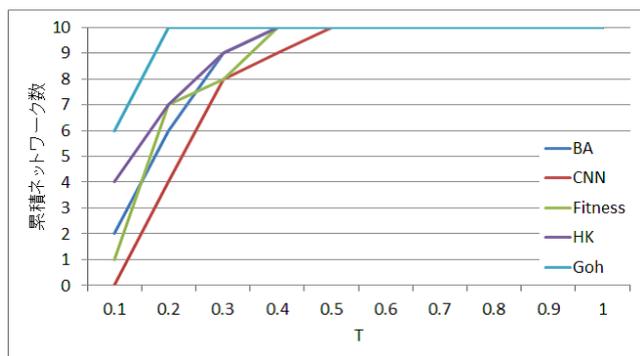


図 1: 閾値 T 以下で生成できたネットワークの累積数

表 1: BA モデルで生成したネットワークと提案モデルで生成したネットワークの特徴

	L	C	r	s	D
目標	2.66	0.118	-0.00922	11	-
生成	2.71	0.142	0.0502	11	0.126

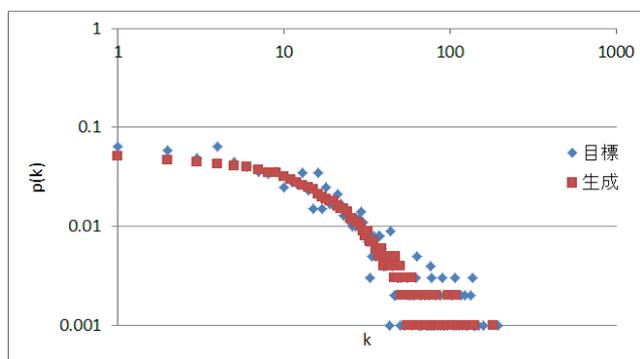
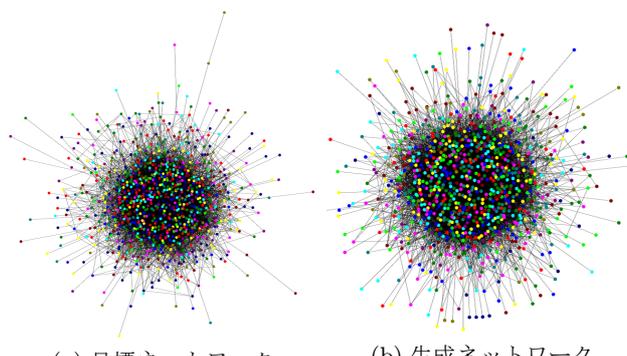


図 2: BA モデルで生成したネットワークと提案モデルで生成したネットワークの度数分布

れのモジュラリティ Q [13] を見てみると、目標ネットワークが $Q = 0.143$ で、生成ネットワークが $Q = 0.150$ であり、同程度にコミュニティ分割される事がわかる。したがって、2つのネットワークは同程度のコミュニティ構造と言える。最後に、**図 3** にそれぞれのネットワークを示す。それぞれ、コミュニティ毎に色分けされている。これらは視覚的に見ても近い構造をとっている事がわかる。しかし、一般に $0.3 \leq Q$ であれば、コミュニティに分割されているといわれるので、この2つのネットワークはコミュニティを持たないと言える。これより、コミュニティを持たないネットワークに関しては問題なく再現可能である事が証明された。

次にコミュニティを持つネットワークである CNN モデルに着目する。本研究で CNN モデルによって生成した 10 個のネットワークのモジュラリティの平均は $Q = 0.261$ であり、中には $0.3 \leq Q$ であるネットワークも存在する。CNN モデルのネットワークは 0.5 より小さい距離で全てのネットワークを生成できている事がわかる。 $0.3 \leq Q$ であるネットワークの内 1 つを例として示す。表 2 に CNN モデルで生成したネットワークと提案モデルで生成したネットワークの特徴を示す。目標ネットワークよりも生成ネットワークではクラスタ係数が小さい事がわかる。これは CNN モデルで生成したネットワークを目標として生成したネットワークの全てに共通する特徴であり、コ



(a) 目標ネットワーク (b) 生成ネットワーク

図 3: ネットワーク図

表 2: CNN モデルで生成したネットワークと提案モデルで生成したネットワークの特徴

	L	C	r	s	D	Q
目標	3.38	0.473	-0.00134	8	-	0.312
生成	3.44	0.336	0.0136	9	0.211	0.212

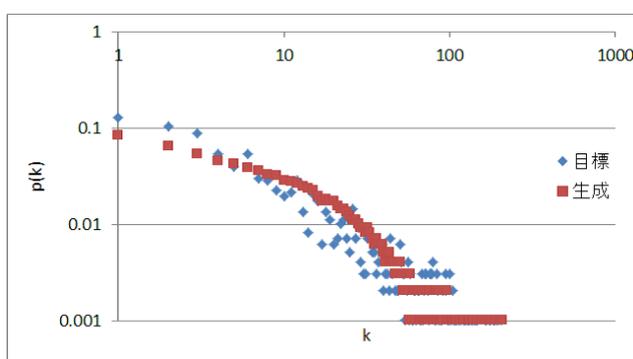


図 4: CNN モデルで生成したネットワークと提案モデルで生成したネットワークの度数分布

ミュニティ数を考慮する事でクラスタ係数を高くする事が困難になるという事を示唆している。これにより、精度が低いと考えられる。

また、**図 4** に、2つのネットワークの度数分布をそれぞれ示す。度数分布でも、わずかに特徴の差はあるものの、大きく離れていない事がわかる。次にコミュニティについて着目する。2つのネットワークのコミュニティ数には差はない。しかし、モジュラリティ Q を見てみると、値に大きな差が存在する。**図 6** にそれぞれのネットワークを示す。これらのネットワーク図を見ると、コミュニティ数は変わらないが、コミュニティの構造には差が存在する事が見て取れる。この差は、モジュラリティの差によって生じていると考えられる。本研究ではネットワークの生成にあたって、モジュラリティは考慮せず、コミュニティの個数だけを考慮しているため、このような差が生じたと考えられる。しかし、様々なコミュニティ構造を作るためには、モジュラリティも考慮したネットワーク生成が必要である事が示唆される。

4. まとめ

本研究では、任意の構造特徴を持つネットワーク生成手法に、コミュニティ構造に関する指標を導入し、精度よくネット

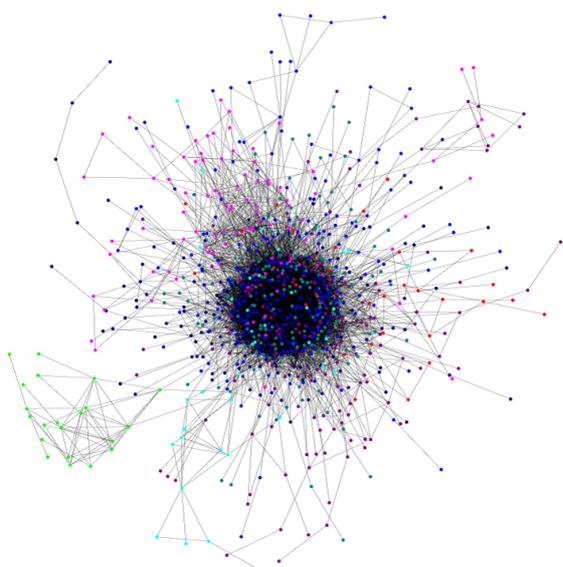


図 5: 目標ネットワーク

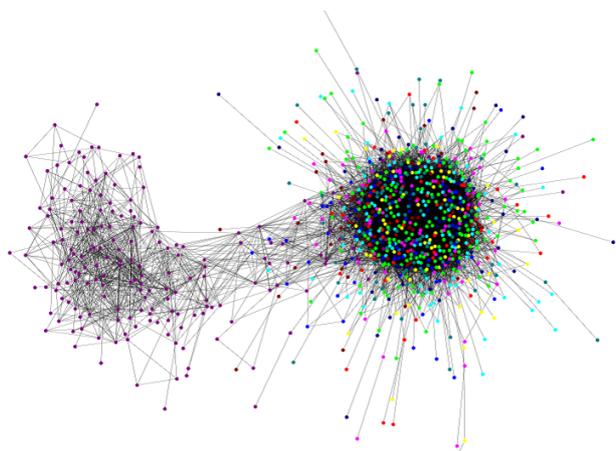


図 6: 生成ネットワーク

ワークを生成できるかを検証した。特徴としては、コミュニティの数をを用いた。検証の結果、コミュニティの数は近いネットワークを作成する事に成功した、しかし、モジュラリティの差によってコミュニティ構造が異なり、様々なコミュニティ構造を生成するためにはモジュラリティも考慮する事が必要である事が示唆された。

今後の課題としては、モジュラリティを考慮した上でモデルの拡張を行う事を挙げる。また、ノードの選択手法や候補生成数を変化させ、最適なものを分析する事を挙げる。

参考文献

- [1] B. Bollobas. *Random Graphs*. Cambridge University Press, 2001.
- [2] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, No. 393, pp. 440–442,

1998.

- [3] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, Vol. 286, No. 5439, pp. 509–512, 1999.
- [4] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, Vol. 67, No. 2, p. 026126, February 2003.
- [5] Hiroaki Shiokawa, Yasuhiro Fujiwara, and Makoto Onizuka. Fast algorithm for modularity-based graph clustering. In Marie desJardins and Michael L. Littman, editors, *AAAI*. AAAI Press, 2013.
- [6] 鳥海不二夫, 石田健, 石井健一郎. Snsにおけるネットワーク成長モデルの提案 (コミュニティ形成支援, 特集: インターネット技術とその応用論文). *電子情報通信学会論文誌. D, 情報・システム*, Vol. 93, No. 7, pp. 1135–1143, jul 2010.
- [7] Alexei Vazquez. Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations. *arXiv:cond-mat/0211528 v2*.
- [8] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, Vol. 88, pp. 128701+, Jun 2002.
- [9] G. Bianconi and A. L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, Vol. 54, pp. 436–442, 2001.
- [10] Ginestra Bianconi and Albert-László Barabási. Bose-einstein condensation in complex networks. *Phys. Rev. Lett.*, Vol. 86, pp. 5632–5635, Jun 2001.
- [11] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.*, Vol. 87, p. 278701, Dec 2001.
- [12] Petter Holme and Beom J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, Vol. 65, No. 2, p. 026107, 2002.
- [13] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review*, Vol. E 69, No. 026113, 2004.