

コミュニティ構造を有するネットワーク成長モデル

○三井一平 (東京大学) 内田誠 (東京大学) 白山晋 (東京大学)

多くのネットワークモデルが提唱されているが、ほとんどが人同士の繋がりのみをモデル化したものである。しかし、社会的ネットワークは、人同士の繋がりだけでなく、人と社会的コミュニティとの関係によっても規定される。例えば、SNSにおいては、コミュニティと呼ばれるグループが存在し、ネットワーク形成に重要な役割を果たしている。また、近年、いくつかのネットワークにおいて明らかにされつつある潜在的なコミュニティ構造にも影響しているものと考えられる。本研究では、SNSのコミュニティを陽なコミュニティ構造と考え、これを有するネットワークに対して、局所的・全体的な相互作用を考慮した、ネットワーク成長モデルを提案する。

An Evolving Network Model with a Community-Oriented Structure

* I. Mitsui (University of Tokyo), M. Uchida (University of Tokyo) and
S. Shirayama (RACE, University of Tokyo)

In the field of complex network research, many network models have been proposed, but most of them focused on modeling of link structure among each individual. On the contrary, social networks are composed not only of such individual links but also of social communities. For example, one of the main functions on SNS is the *Community*, which is like SIG (Special Interest Group) and also is explicit community. It is considered that the community in SNS will play an important role in formation of the network. Besides, this community may affect inherent community structure that has been found in some networks. In this paper, we focus on this community, and propose an evolving network model in an explicit community.

1. はじめに

ネットワークにおけるコミュニティの形成やその構造を調べるという研究は重要である。社会ネットワーク分析において盛んに行われている⁽¹⁾。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⁽²⁾⁽³⁾⁽⁴⁾の中でコミュニティが扱われることが多い。しかしながら、このコミュニティの多くは、ネットワークに内在するクラスター構造にもとづくものであり、SNSに代表される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を扱っているわけではない。さらに、そのコミュニティを含むネットワークの形成や成長、またネットワーク上の事象に与える影響を調べた研究はほとんどない⁽⁵⁾。

本稿では、VazquezのConnecting Nearest Neighborモデル⁽⁹⁾を拡張することにより、陽なコミュニティ構造を有するネットワークに対して、局所的・全体的な相互作用を考慮した、ネットワーク成長モデルを提案する。

2. SNSにおけるコミュニティ

近年、SNS(Social Networking Service)やブログな

どのウェブ上の新しいコミュニケーションサービスに関心が集まっている⁽⁶⁾。特に、ミクシィ(mixi)⁽⁷⁾に代表されるSNSの発展は目覚しく、近い将来において、情報サービス媒体の主要な構成要素となる可能性がある。SNSでは、コミュニケーションに対するいくつかの機能が提供され、利用者間のコミュニケーションの促進が図られる。

SNSにおける知人関係は、利用者が他の利用者に対して登録申請し、許可されることにより生まれる。利用者をノードとし、この2者間の相互の承認をリンクとみなせば、SNS内にネットワークが形成されることになる。

SNS内に各利用者のホームページがあり、その利用者の知人一覧が掲載されている。さらに、誰が自分のページにアクセスしてきたかの履歴を知ることができる。これらの機能はリンクの生成に寄与していると考えられる。一方、SIGと呼ばれる利用者の興味に応じた陽なコミュニティが存在する。コミュニティ毎にホームページがあり、BBSなどの機能があるほか、参加している利用者の一覧を見ることができる。利用者はコミュニティの作成や参加ができる。また、利用者のページでは参加コミュニティの一覧を見ることができる。Fig1に、コミュニティと利用者のリンク構造の模

式図を示す。

コミュニティのもつ、特定の話題に関心のある利用者を集約する機能が、情報伝播を考える上でも重要である。多くの利用者は各コミュニティを通じて情報収集および伝達を行っているものと考えられている。したがって、SNSにおいては、コミュニティ内での交流を通して、新たなリンクが生成されることや、リンクが生成されたことで知人の知人がコミュニティに参加する可能性がある。

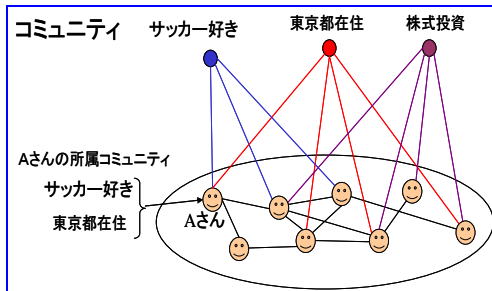


Fig.1: Schema of network structure

本研究では、SNSの運営企業から提供された実データを対象として、陽なコミュニティがネットワーク形成（成長）に果たす役割について分析する。なお、このSNSは、招待制であり、利用者プロフィールは社交系、ネットワークの活動形態はオンライン中心である。また、提供されたデータには一切の個人情報含まれていない。

3. 提案モデル

3.1 CNNモデル

現実世界に存在するネットワークの特徴を再現するネットワークのモデルが数多く提案されている⁽⁸⁾。内田らは、SNSに対する実データを分析し、いくつかのネットワークモデルから生成されるデータの統計量や、局所的な構造との比較を行った⁽¹⁰⁾。結果として、CNN (Connecting Nearest-Neighbor)モデル⁽⁹⁾と呼ばれるものが、実データの特徴を表していることを示している。そこで、本稿において、このCNNモデルを起点として、ネットワークモデルを考察する。

CNNモデルは、Vazquezによって提案された、「友達の友達は、友達になる」という考えにもとづくモデルである。このモデルは、

- 確率 p で新ノードを追加し、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となる潜在的リンクとして保存する。
- 確率 $1-p$ で、潜在的リンクのどれか1つを実際にリンクする。

のいずれかを確率的に繰り返しながら、ネットワ

ークを成長させるものである。

新しく追加されるノードは必ず既存ノードのどれか一つとリンクを形成する。ここで、ノードをSNSの視点で見ると、既存ノードに招待されているとの見方も可能である。本稿では、このCNNモデルを発展させる形で、ノードとコミュニティを生成する3つのモデルを提案する。

3.2 モデル1

モデル1では、1ノードのみの初期状態から、以下のいずれかを確率的に行うことを繰り返す。

- 確率 p で新ノードを追加し、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となる潜在的リンクとして保存する。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、次のリンク候補となる潜在的コミュニティリンクとして保存する。
- 確率 q で、潜在的リンクのどれか1つに実際にリンクする。
- 確率 r で、新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 が新コミュニティを作成したことになる。その際、新コミュニティと、ノード i とリンクされている全てのノードとの辺を、次のリンク候補となる潜在的コミュニティリンクとして保存する。
- 確率 $1-p-q-r$ で、潜在的コミュニティリンクのどれか1つに実際にリンクを張る。

3.3 モデル2

モデル1では、知人の知人と知り合う可能性及び知人の参加しているコミュニティに入る可能性は、新しくネットワークに加入するときのみに限定されていた。モデル2は、知人の知人と知り合う可能性は、新しく知人ができるたびに発生するという考えに基づいている。したがって、モデル1に、潜在リンク及び潜在コミュニティリンクはリンクが張られる度に生成される、というプロセスを加えている。

モデル2では、1ノードのみの初期状態から、以下のいずれかを確率的に行うことを繰り返す。

- のみを示す。A)、C)はモデル1と同じである。太字はモデル1との違いを示している。

B') 確率 q で、潜在的リンクのどれか1つに実際にリンクする。その際、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、次のリンク候補となる潜在的リンクとして、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、次のリンク候補となる潜在的リンクとして、それぞれ保存する。

また、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、次のリンク候補となる潜在的コミュニティリンクとして、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、次のリンク候補となる潜在的コミュニティリンクとして、それぞれ保存する。

D') 確率 $1-p-q-r$ で、潜在的コミュニティリンクのどれか1つに実際にリンクを張る。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、次のリンク候補となる潜在的コミュニティリンクとして保存する。

3.4 モデル 3

モデル 3 は、コミュニティ内でも知人ができるという考え方にもとづき、モデル 2 に、同じコミュニティ内のノード間で潜在リンクができるというプロセスを加えたものである。

モデル 3 は、1 ノードのみの初期状態から、以下のいずれかを確率的に行うことを繰り返す。

D)のみを示す。A), B), C)はモデル 2 と同じである。太字でモデル 2 との違いを示す。

D'') 確率 $1-p-q-r$ で、潜在的コミュニティリンクのどれか1つに実際にリンクを張る。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、次のリンク候補となる潜在的コミュニティリンクとして保存する。また、**新しくコミュニティに参加したノード i とそのコミュニティに参加している既存ノードとの辺を潜在的リンクとして保存する。**

4. 数値実験

4.1 SNS ネットワークの基本統計量

本研究では、実際の SNS のデータを用い、各モデルの数値実験結果の比較検証を行う。分析に用いた SNS ネットワークのデータは、利用者同士の知人関係を表わすリンクの情報からなる。また、コミュニティと利用者を紐づけるデータもある。なお、データからは個人やコミュニティを特定できる情報は一切消去されており、位相情報のみで

ある。

データの規模は、利用者数 363,819、リンク数(有向) 3,813,702、コミュニティ数 90,795、利用者 とコミュニティとのリンク数 5,817,255 である。利用者は平均して 16.18 のコミュニティに参加し、また 1 つのコミュニティには平均して 64.29 の利用者が所属している。また、標準偏差はそれぞれ 45.88、210.78 であり、平均に比べて大きい。

ネットワークは非連結であるが、全体の 99% 以上の利用者が最大連結成分に含まれるため、最大連結成分(利用者数 360,802、リンク数 3,809,228)のみを分析対象とした。

4.2 パラメータの設定

はじめに、比較検証する SNS と、平均次数、平均コミュニティ参加数、ノード数とコミュニティ数の比が等しくなるようなパラメータを見つける。SNS のデータは、平均次数が約 10、平均コミュニティ参加数が約 16、ノード数とコミュニティ数の比は、約 4:1 である。形成されるネットワークがそのような値に近づくようにパラメータを設定する。

例えば、各モデルで示した繰り返しを n 回行えば、確率的に np 個のノードと、 $n(p+q)$ 個のリンクが生じる。1 つのリンクは次数を 2 つ増やすことになるので、平均次数は、 $2n(p+q)/np$ である。これを 10 とすると、 $q/p = 4$ となる。

同様に、 nr 個のコミュニティができるので、ノード数とコミュニティ数の比は、 $p/r = 4$ となる。また、平均コミュニティ参加数は $n(1-p-q)/np$ 個となる。これを 16 とすると、 $p=1/21$ となる。

以上から、パラメータを、 $p=1/21$ 、 $q=4/21$ 、 $r=1/84$ と設定する。また、終了条件は、最大ノード数で与え、以下の数値実験では、18000 としている。これらの値は、全てのモデルで共通である。

4.3 数値実験の結果及び考察

4.3.1 基本統計量

各種統計量を Table 1 に示す。前述した確率パラメータにより、各種統計量が制御可能であることが確かめられた。

また、平均頂点間距離 L 及び、クラスタ率 C を Table 2 に示す。各モデルとも、小さい L および、比較的大きな C となっている。ただし、モデル 2 とモデル 3 の C に関しては、SNS と比べると、小さな値である。この点は、今後の検討課題である。

Table 1: Basic statistics for SNS and models

	利用者数(a)	コミュニティ数(b)	a/b
SNS	363819	90795	4.00
モデル 1	18000	4411	4.08
モデル 2	18000	4505	4.00
モデル 3	18000	4504	4.00

	平均次数	平均コミュニティ参加数
SNS	10.55	16.18
モデル 1	9.92	15.65
モデル 2	10.07	16.18
モデル 3	9.85	15.66

Table 2: Mean geodesic distance and clustering coefficient for SNS and models

	平均頂点間距離	クラスタ率
SNS	5.48	0.33
モデル 1	6.32	0.40
モデル 2	6.92	0.073
モデル 3	4.39	0.007

4.3.2 次数分布

次に、次数分布についての分析結果を示す。SNS及びモデル1から3の結果を、それぞれFigs.2-5に示す。

はじめに、SNSについて調べると、Fig.2に示すように区分的な冪乗分布になっていることがわかる。次数を k とし、次数 k の確率密度を $p(k)$ とすると、区分的に $p(k) \propto k^{-\gamma}$ が成り立っている。そこで、次数10以下を低次領域、次数50以上を高次領域とし、それ以外を中次の領域として、諸量を調べることにする。

SNSにおける冪指数 γ は、低次では0.97、中次では1.90、高次では2.63と、次数が高くなるほど冪指数が大きくなるという特徴がある。

一方、各モデルとも、区分的な冪状分布となっているが、冪指数はモデルによって大きく異なる(Figs.4-6)。まとめたものをTable 3に示す。

モデル1の結果は、内田らの報告⁽¹⁰⁾にあるようにSNSの次数分布をよく表している。ただし、このノード数では高次領域での構造が現れない。モデル3は、コミュニティに対するモデルを加えているが、モデル1と同様に、次数分布でみる限りは、SNSとモデル3のネットワーク構造には類似性があるものと考えられる。詳細に調べるために、SNSとモデル3の次数分布を比較する(Fig.6)。モデル3のノード数がSNSの20分の1のため、高次

領域の右側の挙動が異なるが、全体的によく一致していることがわかる。

Table 3: Exponent of the power law for degree distribution in low-, middle- and high-degree regions

	低次	中次	高次
SNS	0.97	1.90	2.63
モデル 1	1.03	2.04	—
モデル 2	2.11	1.07	—
モデル 3	1.10	1.81	3.76

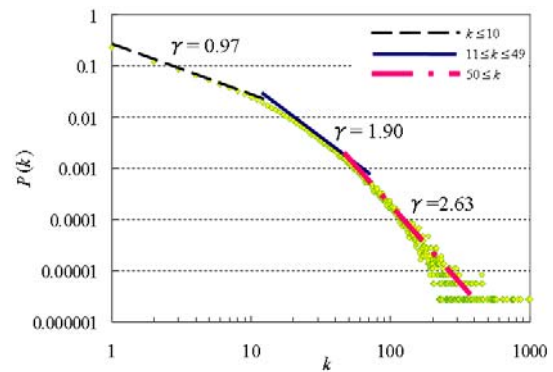


Fig.2: Probability distribution $P(k)$ of degree k for SNS

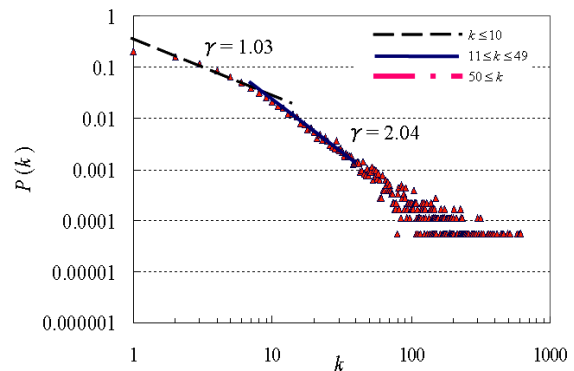


Fig.3: Probability distribution $P(k)$ of degree k for model 1

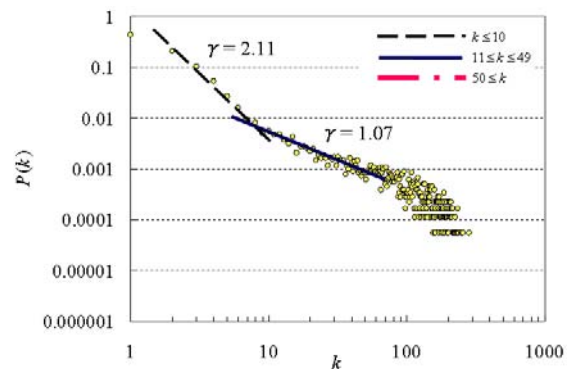


Fig.4: Probability distribution $P(k)$ of degree k for model 2

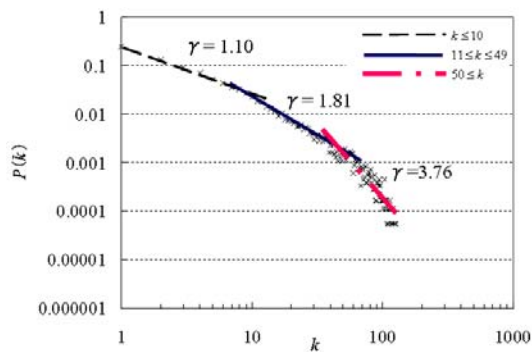


Fig.5: Probability distribution $P(k)$ of degree k for model3

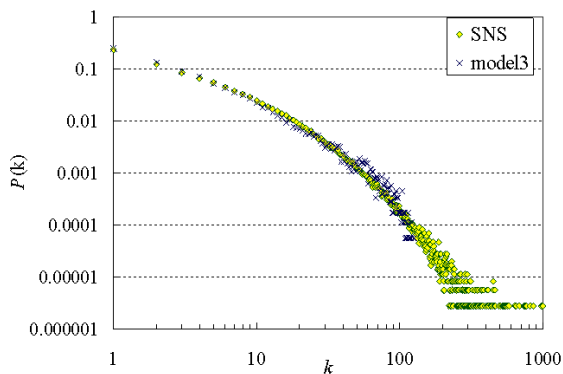


Fig.6: Comparison between SNS and model3

4.3.2 コミュニティ参加数

コミュニティ参加数（利用者が参加しているコミュニティの数）の確率分布についての分析結果を Figs.7-11 に示す.

Fig.7 に示すように、SNS では、次数分布と同様に区分型の冪乗分布になっていることがわかる。そこで、コミュニティ参加数 10 以下を低次領域、コミュニティ参加数 50 以上を高次領域とし、それ以外を中次の領域として、諸量を調べることにする。冪指数は、0.62（低次）、1.34（中次）、2.36（高次）と、コミュニティ参加数が多くなるほど冪指数も大きくなっていることがわかる。

一方、各モデルについては、低中高次領域において冪状分布となっているが、冪指数はモデルにより大きく異なる。各領域での冪指数を Table 4 にまとめる。

モデル 1 は、コミュニティ参加数 13 でピークが現れており、SNS の結果とは乖離しているといえる。モデル 2 は、コミュニティ参加数による違いは比較的小さく、冪指数は 1.10 となっている。

モデル 3 の冪指数は、0.91（低次）、1.27（中次）、

2.00（高次）となっており、SNS 同様、コミュニティ参加数が多くなるほど冪指数も大きくなっている。また、SNS とモデル 3 のコミュニティ参加数の確率分布を比較したものを Fig.11 に示す。類似した分布となっており、SNS とモデル 3 はコミュニティ参加数においても類似性が示唆される。

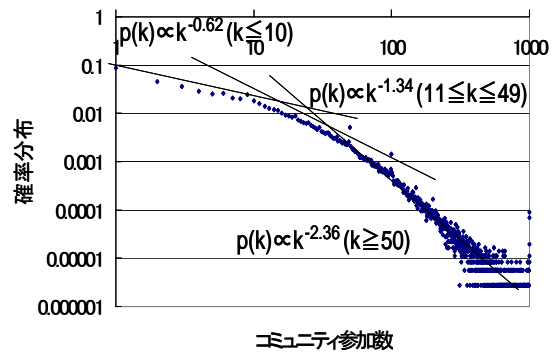


Fig.7: Probability distribution of number of community entry for SNS

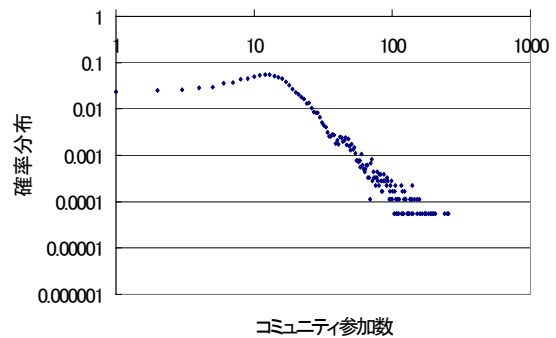


Fig.8: Probability distribution of number of community entry for model1

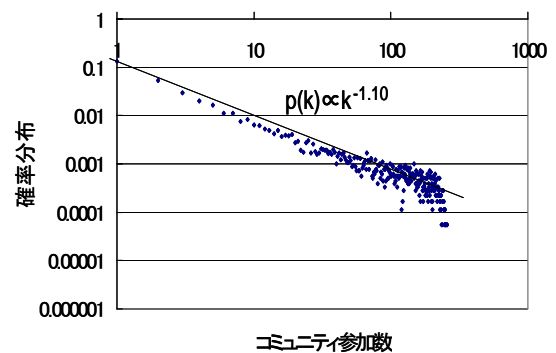


Fig.9: Probability distribution of number of community entry for model2

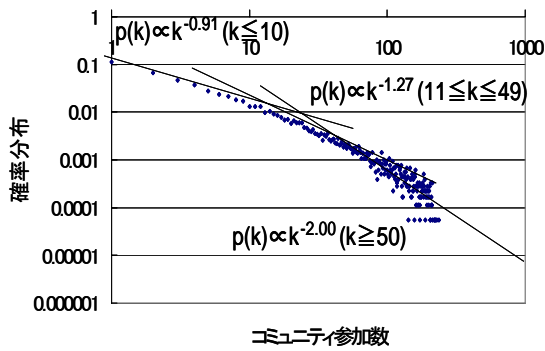


Fig.10: Probability distribution of number of community entry for model3.

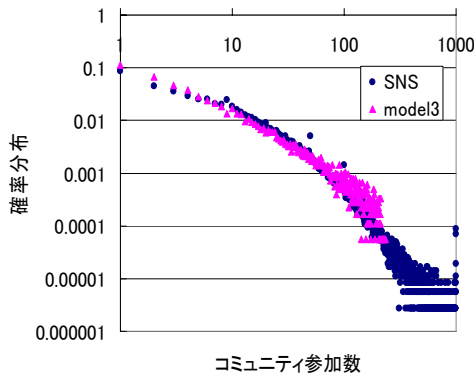


Fig.11: Comparison between SNS and model3

Table 4: Exponent of the power law for probability distribution of number of community entry in low-, middle- and high-degree regions

	低次	中次	高次
SNS	0.62	1.34	2.36
モデル 1	—	—	—
モデル 2	1.10	1.10	1.10
モデル 3	0.91	1.27	2.00

4.3.3 コミュニティ規模

コミュニティ規模（各コミュニティに所属している利用者の数）の確率分布を、Figs.12-16 に示す。また、同様に区分的な冪分布になっているものと考え、冪指数を Table 5 にまとめる。

SNS における、コミュニティ規模が大きくなるほど冪指数も大きくなるという性質を再現しているものは、モデル 3 のみである。そこで、SNS とモデル 3 のコミュニティ規模の確率分布を比較する (Fig.16)。低次領域には差異があるが、全体の

傾向は近く、SNS とモデル 3 はコミュニティ参加数においても類似性があると考えられる。

Table 5: Exponent of the power law for probability distribution of community size in low-, middle- and high-degree regions

	低次	中次	高次
SNS	—	1.27	1.41
モデル 1	1.57	1.57	2.36
モデル 2	—	0.76	—
モデル 3	0.80	0.80	1.23

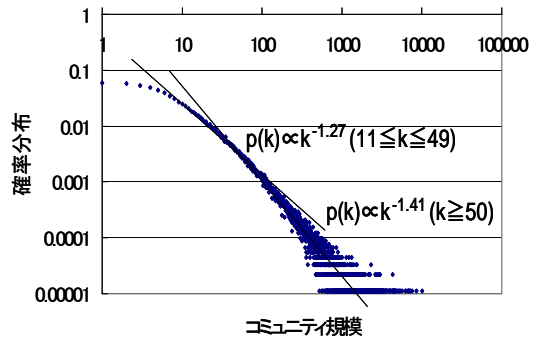


Fig.12: Probability distribution of community size for SNS

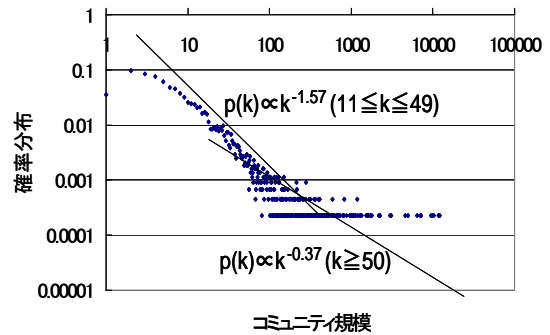


Fig.13: Probability distribution of community size for model1.

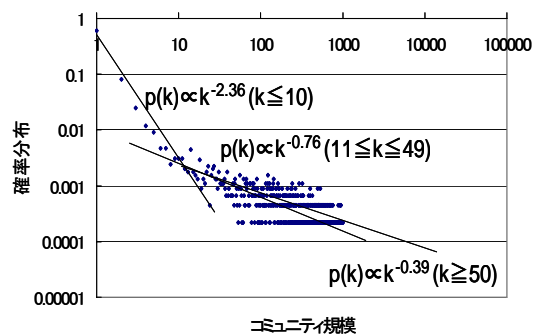


Fig.14: Probability distribution of community size for model2.

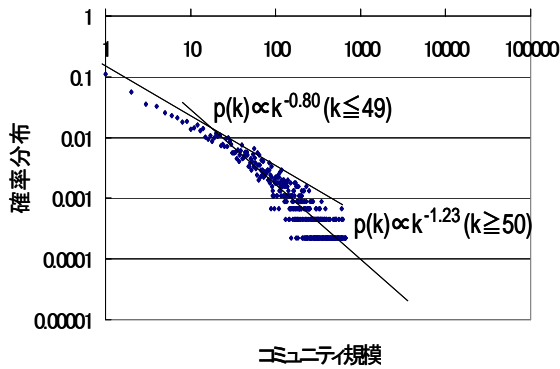


Fig.15: Probability distribution of community size of model3

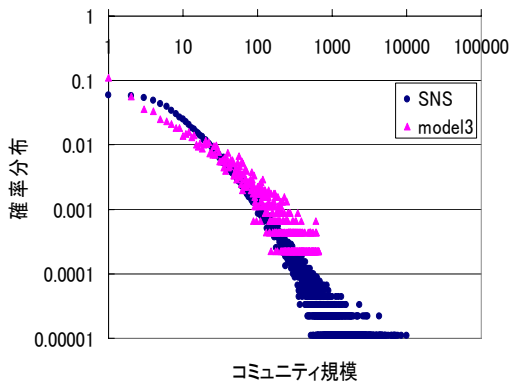


Fig.16: Comparison between SNS and model3

4.3.4 離心性 (farness)

最後に、SNS、及びモデル 1 から 3 に対し、近接中心性の逆数を横軸とし、出現頻度を縦軸としたグラフを Figs.17,18 に示す。近接中心性 C_c は、 n 個のノードからなる連結グラフにおいて、ある頂点 v_i に対して、以下で定義される指標である。

$$C_c(v_i) = \frac{n-1}{\sum_{j \neq i} d_G(v_i, v_j)}$$

ここで、 $d_G(v_i, v_j)$ は 2 頂点 v_i と v_j 間の最短のグラフ距離を示す。本稿では、 C_c の逆数を、farness (正確には normalized farness) と呼ぶことにする。

SNS においては、farness が 4 から 8 に集中することが特徴的である (Fig.17)。モデル 1, 2 については、4 から 12 に広く分布している。また、振動パターンとなることが特徴である (Fig.18)。

一方、モデル 3 の farness の分布は 3 から 6 に集中しており、そのパターンも SNS の性質に近いも

のとなっている。このことから、モデル 3 は SNS に近いネットワーク構造をもっていると考えられる。

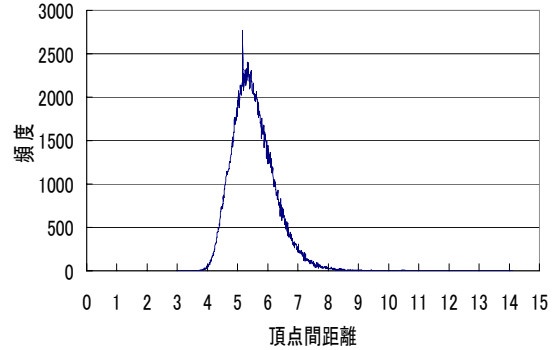


Fig.17: Histogram of the normalized farness for SNS

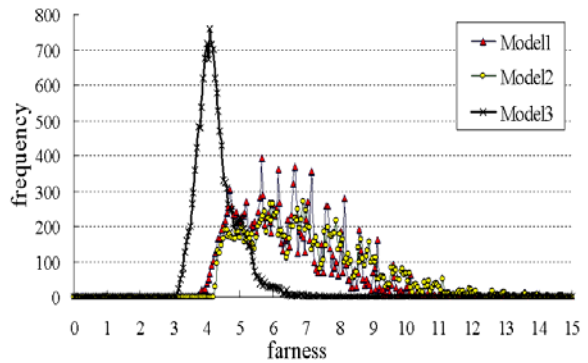


Fig.18: Histogram of the normalized farness for model 1, model 2 and model 3

5. 結論

SNS の実データを用いて、コミュニティの規模及び利用者のコミュニティ参加数に対して調べた結果、次数分布と同様な低次、中次、高次領域に対する区分型の冪乗の分布が現れることがわかった。

この構造を再現するような、陽なコミュニティ構造を有するネットワークに対するネットワーク成長モデルを考察した。

CNN モデルを拡張することにより、3つのモデルを作り、モデルを用いた数値実験により、平均次数、平均コミュニティ参加数、ノード数とコミュニティ数の比等の統計量が再現できることを確認した。さらに、モデル 3 が SNS という実ネットワーク構造の特徴と高い類似性があることを示し

た。モデル3は、モデル2に「コミュニティ内で知人になる」というプロセスを加えただけであることから、対象とした SNS 内においてこのプロセスが働いて、ネットワークの形成にコミュニティが寄与している可能性が示唆された。

6. 今後の展望

モデル3においては SNS ネットワークの特徴を多く再現できたものの、クラスタ率は大きく乖離している。今回再現できた特長に加え、クラスタ率も高く維持できるようなモデルを考案していきたい。

また、本研究で分析した、陽なコミュニティは、いくつかのネットワークにおいて明らかにされつつある潜在的なコミュニティ構造にも影響しているものと考えられる。今後、陽なコミュニティと潜在的なコミュニティの関係性について分析していきたい。

参考文献

- (1) 金光 淳：社会・ネットワーク分析の基礎，勁草書房，2003
- (2) M. E. J. Newman M.E.J.: Detecting community structure in networks. Eur. Phys. J. B 38, 2004, pp.321-330
- (3) Clauset, A., Newman, M.E.J., and Moore, C.: Finding community structure in very large networks, Physical Review, E **70**, 066111, 2004
- (4) Li, C. and Maini, P. K.: An evolving network model with community structure,
- (5) Dong, J., Noh, J.D., Jeong, H-C., Ahn, Y-Y. and Jeong, W.: Growing network model for community with group structure, Physical Review E **71**, 036131, 2005
- (6) 藤巻潤一：ネットコミュニティビジネスの動向，大和総研・経営情報サーチ 2005/冬，Dec. 2004, pp.29-43
<http://www.dir.co.jp/research/report/hitech/041201hitech.pdf>
- (7) mixi: <http://www.emercury.co.jp>
- (8) 増田直紀，今野紀雄：複雑ネットワークの科学，産業図書，2005
- (9) Vazquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, Physical Review E **67**, 056104, 2003
- (10) 内田 誠，白山 晋：SNS のネットワーク構造と情報伝播現象の分析，第 1 回横幹連合コンファレンス講演論文集，CD-ROM 論文集，Nov. 2005