

ネットワーク上の陽なコミュニティとクラスター構造の関係性について

○三井一平 (東京大学) 内田誠 (東京大学) 白山晋 (東京大学)

近年、ネットワークモデルが盛んに提唱されているが、その多くは人同士の繋がりのみをモデル化している。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趣味や属性のグループが存在し、SNS 内のネットワーク形成に重要な役割を果たしていると考えられる。一方、様々なネットワークに対して、潜在的にクラスター構造が存在することが指摘され、その構造を顕在化するための方法や方法論が提案されている。本研究では、陽なコミュニティを有する SNS のネットワークにおける、コミュニティとクラスター構造の関係性を明らかにする。

A Study of Relationship between Explicit and Inherent Community Structure on Networks

* 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 reveal the relationship between the community and inherent community structure.

1. はじめに

近年、ネットワークモデルが盛んに提唱されているが、その多くは人同士の繋がりのみをモデル化している。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 (Social Networking Service) においては、コミュニティと呼ばれる趣味や属性のグループが存在し、SNS内のネットワーク形成に重要な役割を果たしていると考えられる。しかしながら、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が、そのコミュニティを含むネットワークの形成や成長、さらには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⁾。一方、様々なネットワークに対して、潜在的にクラスター構造が存在することが指摘され、その構造を顕在化するための方法や方法論⁽²⁾⁽³⁾⁽⁴⁾が提案されている。

著者らは、陽なコミュニティ構造を有するSNSのネットワークにおける、コミュニティと、Newmanら⁽³⁾によって提案されたクラスタリング手法によって顕在化するクラスター構造の関係を明らかにするために、コミュニティを有するネット

ワーク成長モデルの提案を行った⁽¹¹⁾。具体的には、VazquezのConnecting Nearest Neighborモデル⁽¹⁾を拡張したモデルを比較しながら、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3つ提案した。

本稿では、陽なコミュニティ構造と、ネットワークに内在するクラスター構造の関係を調べることで、モデルの正当性を検証し、モデルを発展させるための指針について検討する。本稿では、ネットワークに内在するクラスター構造のことを、湯田らにしたがって、GNS (ギルバン・ニューマン・サブストラクチャー) と呼ぶことにする。

2. 陽なコミュニティ

ミクシイ(mixi)⁽⁷⁾に代表されるSNSの会員数の増加は顕著であり、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指摘されている⁽⁶⁾。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。その一つがコミュニティというものである。SNSにおけるコミュニティとは、利用者の興味に応じて、SNS上で利用者自身が形成する集まりのことである。コミ

コミュニティを形成するサービスは、多くのSNSのシステムに実装されている。例えばmixiでは、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、参加しているユーザーの一覧を見ることができる。ユーザーはコミュニティの作成や、コミュニティへの参加ができるようになっていく。

一般にSNSは活動形態によって、オンライン中心のもの、オフラインでのイベント中心のものに分類される。藤巻によれば、オンラインでの関係性は、社会的特性よりも共通の関心に基づいていとされ⁽⁶⁾、個人の関係性を考える上で、コミュニティの考察が必要になる。また、オフラインのイベント中心のSNSでも、コミュニティに対するイベントのように、活動という観点ではコミュニティが中心になる場合も多いとされる。

このように、SNSにおけるコミュニティが有する、特定の話題に関心のある利用者を集約する機能が、利用者の交流を考える上で重要になる。多くの利用者が各コミュニティを通じて情報収集および伝達を行い、新たな知人関係(リンク)が形成されることや、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ということが考えられる。本研究では、このような、明示的に示すことのできるコミュニティを、*陽なコミュニティ*と呼ぶことにする。

3. コミュニティ構造を有するネットワーク成長モデル

本稿では、先述した陽なコミュニティとネットワークに内在するクラスター構造の間に関係性があるものと仮定し、コミュニティがネットワークに内在するクラスター構造に与える影響について考察する。本章では、文献(11)で提案したモデルについてまとめている。

著者らは、SNSに対する実データを分析し、CNN(Connecting Nearest-Neighbor)モデル⁽¹¹⁾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た^{(9),(12)}。既報においては、このCNNモデルを起点として、ネットワークモデルを考察した。

CNNモデルは、Vazquezによって提案されたもので、「友達の友達は、友達になる」という考えの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。これを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。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つのモデルを提案した⁽¹¹⁾。提案したモデルの概要を以下に示す。

モデル1は、CNNモデルと同様の原理でコミュニティも形成していくモデルである。

モデル2は、「新たにリンクができたときは、必ず潜在リンクができる」というプロセスをモデル1に加えたモデルである。

モデル3は、「同じコミュニティ内の人同士は友達になりやすい」という効果をモデル2に加えたモデルである。

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すリンクの情報からなる。また、コミュニティと利用者をつ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 3,813,702(有向)、コミュニティ数 90,795、利用者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 数値実験の結果及び考察

4.2.1 基本統計量

SNS,各モデル及びBAモデルの各種統計量を、Table1, 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。モデル3とBAモデルの C 及び L は近い値となっていることも特徴といえる。

Table 1: Basic statistics for SNS and models

	利用者数(a)	コミュニティ数(b)	a/b
SNS	360802	90795	4.00
モデル1	18000	4492	4.01
モデル2	18000	4378	4.11
モデル3	18000	4616	3.90
BA	18000		

Table 1 (cont.)

	平均次数	平均コミュニティ参加数
SNS	10.55	16.18
モデル1	10.10	16.06
モデル2	9.98	15.79
モデル3	10.00	15.99
BA	8.00	

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

	平均頂点間距離 L	クラスタ率 C
SNS	5.48	0.33
モデル1	6.30	0.40
モデル2	7.04	0.077
モデル3	4.37	0.007
BA	4.07	0.004

4.2.2 次数分布

著者らのこれまでの研究⁽¹¹⁾における分析結果の一例として次数分布とコミュニティ参加数の分析結果(次項)を示す。まず、次数分布についての分析結果の代表例を示す(Figs.1,2)。

はじめに、SNSについて調べると、Fig.1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(Table 3)。

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

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(Table 3)。

モデル1の結果は、内田らの報告⁽⁹⁾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.3)。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。

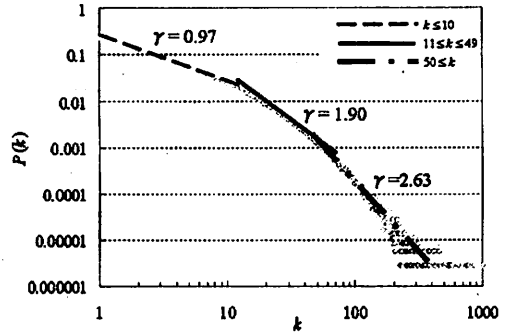


Fig.1: Probability distribution $p(k)$ of degree k for SNS

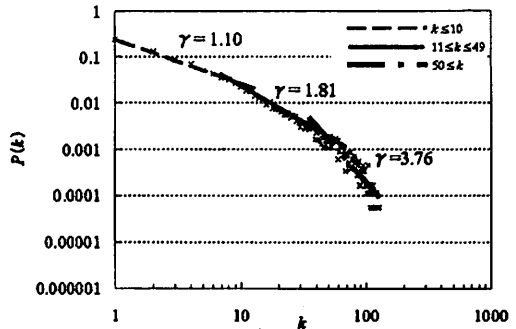


Fig.2: Probability distribution $p(k)$ of degree k for model3

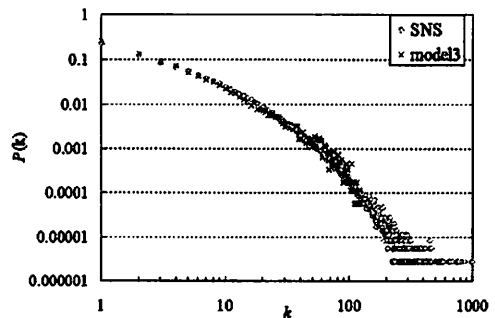


Fig.3: Comparison between SNS and model3

4.2.3 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)の確率分布の代表例を Figs.4-5

に示す。Fig.4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、賭量を調べることにする。

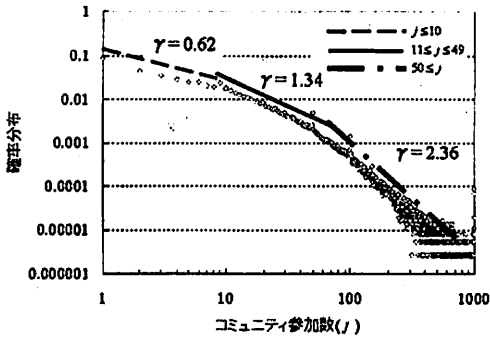


Fig.4: Probability distribution of number of community entry for SNS

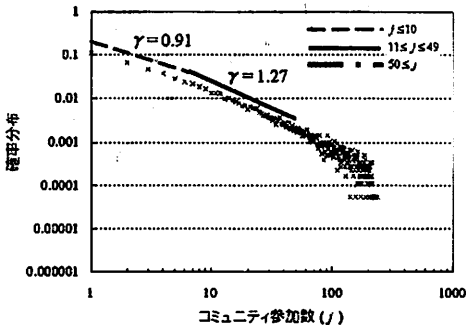


Fig.5: Probability distribution of number of community entry for model3

SNSの場合、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。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。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(Fig.5)。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.6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。

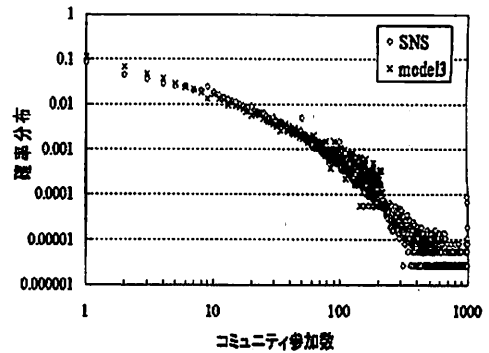


Fig.6: Comparison between SNS and model3

4.2.4 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.11,12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.7)。

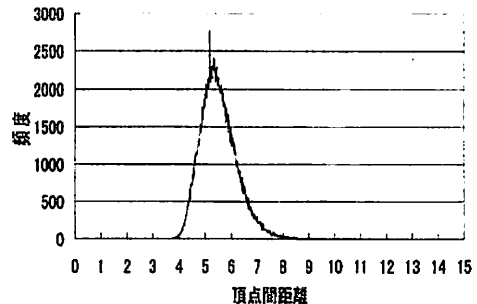


Fig.7: Histogram of the normalized farness for SNS

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的である (Fig.8)。

一方、モデル3のfarnessの分布は、BAモデルとmodel1,2の中間的なものとなっている。

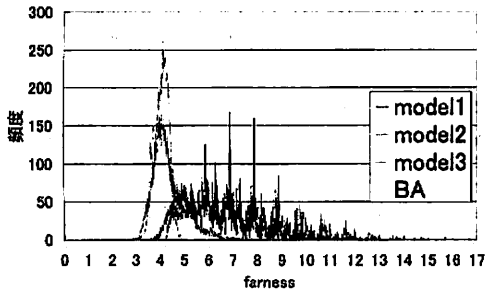


Fig.8: Histogram of the normalized farness for model 1, model 2, model 3, and BA model

4.2.5 ギルバン・ニューマン・サブストラクチャー (GNS)

次に, Newman のクラスタリング手法⁽³⁾によって, ユーザーネットワークの GNS を求める. SNS 及び各モデル, BA モデルの結果をそれぞれ Table4, Fig.9~Fig.13 に示す.

Table 4: Number and mean size of GNS

	GNS の数	平均サイズ
SNS	3930	91.8
model1	391	46.0
model2	1185	15.1
model3	81	222
BA	21	857

SNS について, 3,932 個のクラスターに分割できた. 最も大きなクラスターは, 88,615 人を含み, 全体の 24.4% を占めている. 規模が二番目, 三番目のクラスターは, それぞれ全体の 17.7%, 17.2% を占めており, 3 つの大きなクラスターと, 多くの小さなクラスターが存在することが特徴である. また, Fig.13 に示す GNS の規模とランクとの関係は, 湯田らの結果⁽¹⁰⁾と一致しており, 3 つの領域に分かれ, 中層構造のない独特な分布となっている.

モデル 1, 2 は, 冪乗分布となっている. モデル 3 は, SNS 同様, 中層構造のない複数の領域に分かれた分布となっている. 一方, モデル 3 と BA モデルを比較すると, GNS の数は同じオーダーとなっており, また, グラフについても類似性が見受けられ, モデル 3 と BA モデルの GNS の構造が近いことが示唆される.

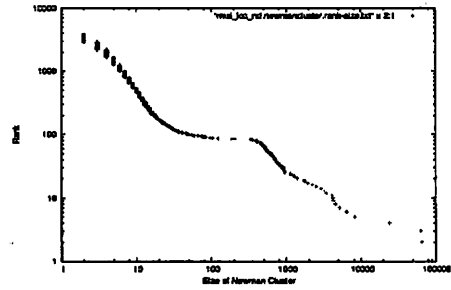


Fig.9: Rank-size plot of GNS on SNS

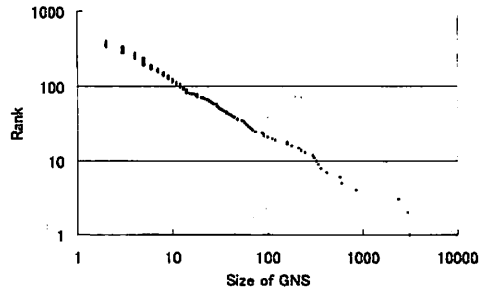


Fig.10: Rank-size plot of GNS on model1

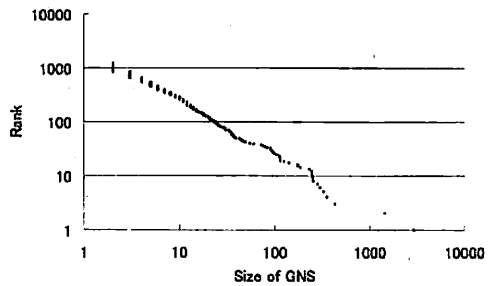


Fig.11: Rank-size plot of GNS on model2

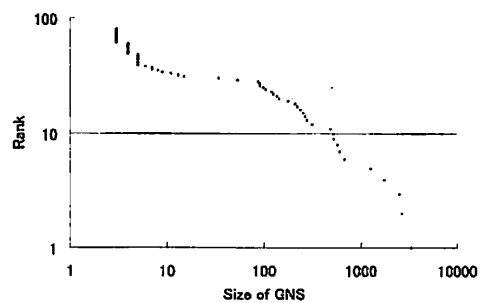


Fig.12: Rank-size plot of GNS on model3

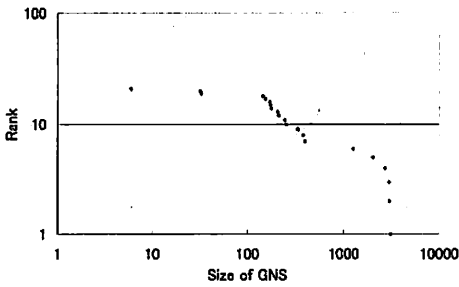


Fig.13 :Rank-size plot of GNS on BA

続いて、各コミュニティが、いくつの GNS で構成されているかについて、SNS 及び各モデルに対し、Fig.14, Fig.15, Table5 に示す。

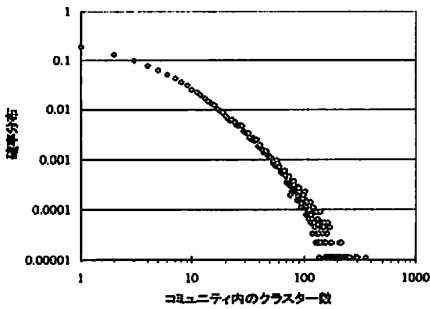


Fig.14: Probability distribution of the number of clusters obtained by GNS within each community

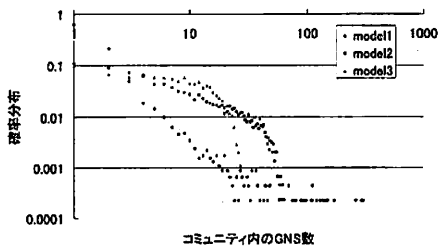


Fig.15: Probability distribution of the number of GNS obtained by models within each community

	単一の GNS	単一の GNS (複数人限定)	4つ以下の GNS
SNS	18.7%	13.5%	49.6%
model1	63.8%	61.8%	92.2%
model2	11.0%	1.39%	33.7%
model3	10.8%	1.03%	28.5%

SNS において、単一の GNS で構成されるコミュニティは 16,905 個あり、全体の 18.7%を占める。複数人で構成されるコミュニティに限定すると、13.5%が単一の GNS から成り立っている。また、約半数のコミュニティは 4つ以下の GNS で構成されている。GNS の総数と比較すると、コミュニティの多くは非常に少ない数の GNS によって構成されていることから、コミュニティと GNS に相関関係があるといえる。

続いて、各モデルについて分析する。単一の GNS で構成されるコミュニティは、各モデルそれぞれ、全体の 63.8%, 11.0%, 10.8%を占める。複数人で構成されるコミュニティに限定すると、それぞれ、61.8%, 1.39%, 1.03%が単一の GNS から成り立っている。また、4つ以下の GNS で構成されるコミュニティは、それぞれ、92.2%, 33.7%, 28.5%となっている。以上から、モデル1においては、コミュニティと GNS に深い関連性があるといえる一方、モデル2とモデル3については、コミュニティと GNS の関連性は低いことがわかる。

コミュニティの規模(参加者の数)と、そのコミュニティにおける最大の GNS が占める割合を比較する。Fig.16 に SNS の結果を、Fig.17 に各モデルの結果を示す。

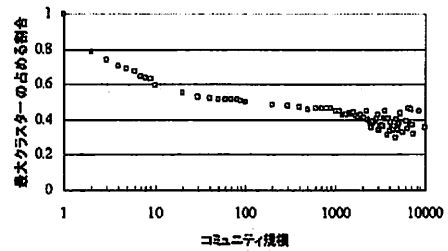


Fig.16: Relationship between size of community on SNS and ratio of the maximum GNS size on the size of each community

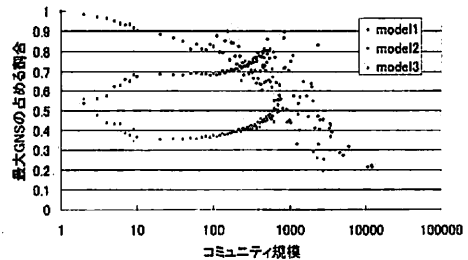


Fig.17: Relationship between size of community on models and ratio of the maximum GNS size on the size of each community

はじめに、SNS の結果について分析する。SNS における最大の GNS のユーザーが均等にコミュニティに加入した場合、各コミュニティにおいて占める割合は 41% となることが判っている。Fig.5 の、コミュニティ規模が比較的小さい領域において、最大 GNS の割合は 4 割より高くなっている。このことから、比較的小規模のコミュニティに参加しているユーザー同士は、同一の GNS に属している傾向があることがわかる。

次に、各モデルについて分析する。各モデルにおける最大の GNS のユーザーが均等にコミュニティに加入した場合、各コミュニティにおいて占める割合は、37.8%, 71.3%, 39.8% となる。モデル 1 については、同様な傾向が見られ、SNS と近い性質を持つと推察される。一方、モデル 2, 3 については、均等にコミュニティに加入した場合の割合に近いことから、コミュニティと GNS の関係性は弱いと考えられ、SNS と異なる性質をもっているといえる。

GNS は、その定義から、親しい友人同士などが互いに密に接続し、そうではないユーザーとはあまり接続しないことで形成されると考えられる。そのため、参加者の多くが同一の GNS に属するコミュニティは、友人同士で構成されたコミュニティと推測できる。もしくは、あるコミュニティに共に参加しているユーザー同士に、コミュニティでの交流により、互いに密な接続関係が生じた可能性も考えられる。SNS の成長モデルを考える上では、この点を考慮することが重要になるだろう。

5. 結論

SNS の実データを用いて、陽なコミュニティを有する SNS のネットワークにおける、コミュニティと GNS の関係性について分析した。その結果、コミュニティと GNS には相関関係があることが示された。そして、SNS におけるユーザーネットワークの形成に、コミュニティが影響を与えている可能性が示唆された。

提案したモデルにおいて、コミュニティと GNS についても分析を行い、SNS の結果と比較した。モデル 3 と SNS は、Fig.1~Fig.6 で示した量に関しては高い類似性があり、SNS の性質を良く再現しているといえる。一方、クラスター率や、コミュニティと GNS の分析において、両者は異なる性質をもつことが示され、モデル 3 は SNS を十分には再現していないことがわかった。以上から、コミュニティを有するネットワークモデルの検証には、コミュニティについての分析も行う必要があると

いえる。

一方、モデル 3 と BA モデルは、クラスター率や離心性、GNS の分析において、類似した性質をもつことがわかった。モデル 3 は、モデル 2 に「コミュニティ内で知り合う効果」を加えたモデルであり、BA モデルは、ネットワーク全体において優先的選択を行うモデルである。従って、「コミュニティ内で知り合う効果」が「ネットワーク全体において優先的選択を行うこと」に相当していると考えられる。

6. 今後の展望

「全体的に優先的選択を行うこと」は、その性質上、クラスター率 C を下げると考えられる。

よって、モデル 3 における「コミュニティ内で知り合う効果」がクラスター率 C を下げているといえる。以上から、SNS とモデル 3 の結果が合わない一因は、「コミュニティ内で知り合う効果」が強すぎることにありと考える。そこで、相対的に弱くすること等により、クラスター率 C を上げることが可能となり、SNS に類似した性質をより多く有するモデルを作ることができると考える。今後、モデル 3 を改良することにより、コミュニティと GNS の関係性も再現するようなネットワークモデルを提案していきたい。

参考文献

- (1) Vazquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Physical Review E* 67, 056104, 2003
- (2) 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/041201hit ech.pdf>
- (7) mixi: <http://www.emercury.co.jp>

- (8) Newman, M.E.J: The structure and function of complex networks, SIAM Review 45, 2003, pp.167-256
- (9) 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
- (10) 湯田 聡夫, 小野直亮, 藤原義久: ソーシャル・ネットワーキング・サービスにおける人的ネットワークの構造, 情報処理学会論文誌, 47 卷 3 号, 2006, pp. 865-874
- (11) 三井一平, 内田 誠, 白山 晋: コミュニティ構造を有するネットワーク成長モデル, 情報処理学会研究報告「知能と複雑系」No.2006-ICS-142-4, January 2006, pp.17-24
- (12) 内田 誠, 白山 晋: SNS のネットワーク構造の分析とモデル推定, 情報処理学会論文誌, 0747 卷 9 号, 2006