

クラスター性によるコミュニティ抽出と情報拡散速度

熱田 圭佑* 森口 一郎**

ネットワークのリンクが密な部分構造を抽出する手法として、クラスター性によるコミュニティ抽出法が提案されているが二重スターグラフ問題が未解決として残されている。本研究ではこの問題を、二重スターグラフの頂点ノードをコミュニティ抽出時に抽出しコミュニティ分割を行うことで解決した。さらに、コミュニティ構造が与える影響について情報拡散に着目し調べた。この結果、抽出されたコミュニティトポロジ情報を使いネットワーク全体に情報が拡散されるターン数を調べた結果、コミュニティサイズや所属コミュニティで拡散性が変化することと、ランダムな拡散よりコミュニティを使用した拡散手法が現実的に有効であることを明らかにした。

キーワード：コミュニティ抽出、二重スターグラフ、クラスター性、情報拡散速度、現実ネットワーク

Community Detection by Clustering Property and Diffusion Speed of Information

Keisuke ATSUTA* and Ichirou MORIGUCHI**

To detect partial structures with high density portion of links from arbitrary networks, Matsushima proposed a method using clustering property. However, the “overlapping star-graph” problem, has been remained unsolved in this technique. In this paper we resolved it by detecting overlapping star-graph nodes firstly before dividing networks to communities. Furthermore, we investigated the effect of community structures on information diffusion. By the topology information of the detected communities, the number of simulation turns that are sufficient for information to spread over the entire networks indicated that information diffusion speeds strongly depend on the size and the properties of these communities, and that the diffusion strategy by exploiting community structures is more effective and practically executable than the random diffusion strategy.

Keywords: community detection, overlapping star-graph, cluster property, information diffusion speed, real network

*東京情報大学 総合情報学部 情報システム学科 (2015年3月卒業)
2015年4月よりコムテック株式会社に所属
COMTEC Inc.

2015年6月23日受理

Tokyo University of Information Sciences, Faculty of Informatics, Department of Information Systems, Undergraduate Student
**東京情報大学 総合情報学部 情報システム学科
Tokyo University of Information Sciences, Faculty of Informatics, Department of Information Systems

1. はじめに

通信障害に強いネットワーク構築やマーケティングの向上など、ネットワークに関係する分野に役立てることを目的としたコミュニティ分析研究が近年活発である。コミュニティ分析には、ネットワーク全体から関係性が密な部分構造をコミュニティとして抽出する必要がある。このような研究の一例として、2004年に Newman と Girvan はリンクの割合を用いてネットワーク内部の密度が高い部分構造を抽出するコミュニティ分割法を提案している[1]。一方、松島は現実ネットワークの特徴の一つであるクラスタ性に着目し、ネットワーク全体のトポロジ情報を必要としない局所的コミュニティ抽出法を提案した[2][3]。しかし、重なりあう複数のコミュニティ構造を抽出できない二重スターグラフ問題が残っている。

本研究では、松島が提案した抽出法を改良し、問題として挙げられていた二重スターグラフ問題の解決を図った。そして、アドホックネットワークの構造モデルの1つである Random Geometric Network を使用し、二重スターグラフの分割とコミュニティ抽出の有効性を確かめた。

また、情報拡散についての研究の一つとして千石らはウィルスの拡散とネットワーク直径の関係性について調べた[4]。この研究では、ネットワークの直径からウィルスの拡散速度を求めているが、与えるノードによる拡散性の違いや拡散の特徴などは求められていない。

以上より本研究では、現実ネットワークの一つである Autonomous System ネットワーク[5]を用いてコミュニティがネットワーク上の情報拡散速度に与える影響について調べた。その結果、Cohen らが提案した知人手法[6]と同様の原理により、コミュニティに属しているノードをランダムに選択し隣接する同一コミュニティに属するノードに情報を与える手法が、最も効果的にネットワーク全体へ情報を拡散できることが判明した。

2. 使用したネットワークデータとネットワークモデル

本研究では、ネットワーク構造の違いによるコミュニティ構造の違いを比較するため、1種類の現実ネットワークデータと4種類の構造が異なるネットワークモデルを使用した(表1)。

まず、現実ネットワークデータとしてはインターネットの部分構造である Autonomous System (AS) ネットワークを使用した。この AS ネットワークは、CAIDA[7]から取得した2009年11月20日のデータであり、ノード数33326、平均リンク数2.25である。なお、本研究で使用するネットワークモデルはこの AS ネットワークのノード数と平均リンク数を基準に作成した。

シミュレーションネットワークモデルとしては、現実ネットワークの性質であるスモールワールド性、スケールフリー性、クラスタ性を持つモデルを3つ使用した。スモールワールド性は、任意に選ばれたノード同士がわずかなノードを介すだけで通信可能になる特徴であり、スケールフリー性とは、各ノードの持つリンク数を k とするとネットワーク全体のリンク数分布が $k^{-\beta}$ に比例するべき乗則に従う特徴である。また、クラスタ性は3つのノードが三角形のリンクを多く形成している特徴である。

まず、スモールワールド性を持つ Random Network (RN) モデル[5]、スモールワールド性とスケールフリー性を併せ持つ Barabási-Albert (BA) モデル[5]を使用し、クラスタ性の有無による抽出コミュニティの変化を AS ネットワークと比較した。この BA モデルは決められたリンク本数を持つ新規ノードを既存ノード群のリンク数を多く持つノードに優先的に接続することでネットワークを成長させるモデルである。そのため平均リンク数は整数とならず、AS ネットワークと同じ平均リンク数にならない。そこで、本研究では一定の確率 γ で、決められたリンク本数から更に1本のリンクを接続するように改良を加えた。今回は、初期

表1 現実ネットワーク (AS) と使用モデルの主な特徴

ネットワークデータ ネットワークモデル	スケールフリー性 (べき指数 β)	スモールワールド性	クラスター性 (C)
ASネットワーク	2.143	あり	0.224
BAモデル	2.831	あり	0.002
RNモデル	なし	あり	0.000
HKモデル	3.027	あり	0.224
	2.817		0.574
RGN	なし	なし	0.585

接続リンク数を1、 γ を0.25とすることでASネットワークと同じ平均リンク数を持つネットワークを作成した。

また、BAモデルにクラスター性を持つよう改良したHolme-Kim (HK) モデル[8]を使用し、クラスター性が抽出されるコミュニティに与える影響についてASネットワークとの比較を行った。

最後に、近年増加傾向である無線通信ネットワークがどのようなコミュニティ構造を持つのか比較するため、アドホックネットワークの構造モデルであるRandom Geometric Network (RGN) を使用した。RGNは無線マルチホップ通信に使用されるモデルであり、一定の電波到達範囲内のノードとリンクを形成するため、近くのノード同士が繋がりがやすくリンク関係が密になる。しかし、遠くのノードとは複数のノードを介さなければ通信ができないため、スモールワールド性がないネットワークである。このRGNは、直接リンクで繋がっていない任意のノード同士が途中ノードを中継することで通信可能となるジャイアントネットワークの生成に、平均リンク数が最低4.52以上必要である[9]。今回は、より現実ネットワークに近い通信可能なネットワークからコミュニティを抽出するために、ノード数をASネットワークに合わせた33,326とし、ジャイアントネットワーク生成に十分な平均リンク数10のRGNを使用した。さらに、RGNでは一定範囲内のノードとリンクを形成するため、複数のコミュニティ構造が現れやすいと考えられる。そこで、ノード

数1,000としたRGNを使用し、二重スターグラフの分割結果とコミュニティ抽出精度を改良前の手法と比較を行った。

3. クラスター性によるコミュニティ抽出

3.1 クラスター性

現実ネットワークは分岐だけからなるツリー構造であることは少なく、あるノードの近隣ノード同士がリンク関係にある三角形のサイクルを持つことが多い。例えば、ノードAとリンク関係にあるノードBが、同じようにノードAとリンク関係にあるノードCとリンク関係にあるとき、三角形ABCのリンク関係が成り立つ。このような三角形のリンク (クラスター) を多く含む構造をクラスター性が高いという。

このクラスター性を定量的に評価するため、クラスター係数を用いた。例えばノードAのクラスター係数は、ノードAの持つリンク本数から作成可能な三角形のリンクのうち、実際に持つ三角形のリンク数の割合で定義できる。また、ネットワーク全体のクラスター係数は、全ノードのクラスター係数を平均することで測定できる。よって、あるノード*i*のクラスター係数は、

$$C_i = \frac{\text{ノード}i\text{を頂点に持つ三角形の数}}{k_i(k_i - 1)/2}$$

で定義でき、ネットワーク全体のクラスター係数*C*は、

$$C = \frac{\sum_{i=1}^n C_i}{n}$$

で定義できる。ここで、 k_i はノード i の持つリンク数、 n はネットワークの全ノード数である。

現実ネットワークのクラスター係数は0.1~0.75程度であり[10]、上記式を用いて本研究で使用するネットワークモデルのクラスター係数を実際に計測したところ、ASネットワークは0.224、BAモデルRNモデルは共にほぼ0、RGNは0.585であった。またHKモデルは、ASネットワークとほぼ同じクラスター係数0.224のネットワークとクラスター係数を0.574に変化させたモデルを作成した。

3.2 松島の提案コミュニティ抽出手法

松島が提案した抽出手法は、ネットワーク全体のトポロジ情報を必要とせず、隣接するノードの構造情報だけでコミュニティ分割が可能な局所的抽出手法である。この手法はコミュニティ構造の特徴であるクラスター性を重視することにより、リンク関係が密な部分構造を抽出できる。例として図1のようにネットワーク内で三角形のリンク関係をもつノードを抽出し、コミュニティとして追加していくことで、コミュニティ構造を抽出することができる。こ

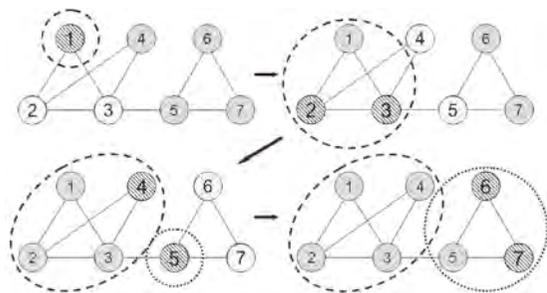


図1 松島のコミュニティ抽出手法流れ図

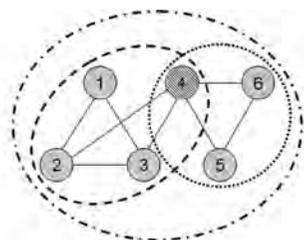


図2 二重スターグラフ問題

で図中の点線は、各ノードの所属するコミュニティを示しており、以後各コミュニティを $com1$ 、 $com2$ 、 $comj$...と表記する。

松島のコミュニティ抽出アルゴリズムでは、始めに、ネットワーク内からランダムに選ばれたノード①をコミュニティ ($com1$) に追加し、 $com1$ の頂点とする (図1)。次に、頂点に隣接するノード② (または③) が、 $com1$ に隣接するノード③ (または②) とリンク関係にあった場合②と③を $com1$ に追加する。その際、一定の閾値に達しないクラスター係数のノードはコミュニティに追加しないことでコミュニティを分割していく。そして頂点に隣接する全ノードのコミュニティ判定が完了後、頂点に隣接する②、③を $com1$ の新たな頂点とし、同様の抽出法を繰り返す。三角のリンク関係にないノードや、クラスター係数が閾値に達しないノードが隣接していた場合は、新規コミュニティ $comj$ の頂点とし抽出を繰り返す。全てのノードのコミュニティ判定終了で、コミュニティ分割完了とする。

しかし、上記松島の局所的抽出法には二つの問題点がある。一つ目は閾値の設定であり、閾値の決定にはネットワーク全体のクラスター係数が必要である。二つ目は、松島自身が挙げた二重スターグラフが発生する問題である[2]。二重スターグラフとは、図2のように、頂点ノード④が2つのコミュニティ ($com1$ 、 $com2$) の中心に存在し、どちらのコミュニティとも三角形のリンク関係を持つ状態である。この場合ノード④の属するコミュニティが抽出開始ノードによって変化し、複数のコミュニティを一つのコミュニティ ($com3$) として抽出することがある (図2)。このように抽出開始ノードによる違いで、コミュニティ構造が変化してしまう二重スターグラフ問題は、部分構造を抽出する手法の精度や構造解析に影響するため、解決策が必要である。

3.3 改良コミュニティ抽出手法

前述の二つの問題に対し、本研究では局所的

```

#コミュニティ抽出
community_detection()
start_node ← ランダムに選ばれたノード
comA[] ← start_node
WHILE comA[]に属するノードのコミュニティ抽出が完了
していない DO
lan ← 1
listA[] ← start_nodeの隣接ノードを格納
la ← listA[]のサイズ
WHILE lan ≠ la DO
listB[] ← listA[lan]の隣接ノードを格納
lb ← listB[]のサイズ
FOR i=1 to lb DO
flag ← 0
IF listA[] ⊇ listB[i] THEN
IF comA[] ⊄ listA[lan] AND comA[] ⊄ listB[i]
THEN
IF community_divide(listA[lan]) = 1 THEN
comA[] ← listA[lan]
ELSE
flag ← 1
ENDIF
IF community_divide(listB[i]) = 1 THEN
comA[] ← listB[i]
IF community_divide(listA[lan]) = 1 THEN
comA[] ← listA[lan]
flag ← 0
ENDIF
ELSE
comj[] ← listB[i]
ENDIF
IF flag = 1 THEN
comj[] ← listA[lan]
ENDIF
ENDIF
lan ← lan + 1
ENDWHILE
start_node ← comA[]
ENDWHILE

#コミュニティ分割
community_divide(num)
listC[] ← numの隣接ノードを格納
lc ← listC[]のサイズ
x ← 0
FOR s=1 to lc DO
IF comA[] ⊇ listC[s] THEN
x ← x + 1
ENDIF
ENDFOR
IF x < 2 THEN
RETURN with 0
ELSE
RETURN with 1
ENDIF
RETURN with 0

```

図3 改良コミュニティ抽出手法疑似コード

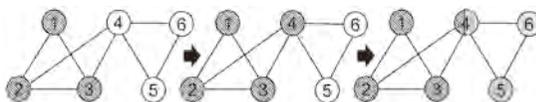


図4 改良手法のコミュニティ抽出流れ図

手法のみを用いて抽出を行うためにクラスター係数の閾値指定は行わないこととし、二重スターグラフ問題をコミュニティ抽出時に二重スターグラフの頂点を見つけ出し分割を行うことで解決した (図3) (図4)。

改良手法を用いて図4からコミュニティを抽出する場合、まずランダムに選ばれたノード (ノード①) を頂点とし、隣接するノードのうちクラスター関係にあるノードを同一コミュニティ ($comA$) に追加する。次に、ノード④を頂点とし $comA$ の抽出を行った際、隣接するノード (⑤、⑥) はクラスター関係にあるため $comA$ に追加する。この際、追加されたノードが属する $comA$ に対し、クラスター関係にあるノード数が閾値を越えているか判定する。もし、ノード (⑤、⑥) が $comA$ に属するノードに対しクラスター関係にあるノード数が閾値未満 (本研究ではノード数2とする) だった場合、コミュニティ判定が行われた際の頂点ノードを二重スターグラフノード (ノード④) と判断し、コミュニティ判定が行われたノード (⑤、⑥) と頂点ノード④を別コミュニティ ($comj$) に追加する。この改良により、二重スターグラフ頂点ノードを複数のコミュニティに属するノードとしコミュニティ分割を行い、二重スターグラフ問題を解決した。

また、ランダムに選択したノードを頂点としコミュニティ抽出 ($comA$) を行った際、クラスター関係のないノードが隣接関係にある場合は新規コミュニティの頂点ノードとし、現在抽出している $comA$ の抽出が全て完了したのち、新規コミュニティ ($comB$) の頂点ノードとし抽出を開始する。このように現在抽出しているコミュニティ頂点の隣接ノードが全て別コミュニティの頂点とならない限り、別コミュニティ

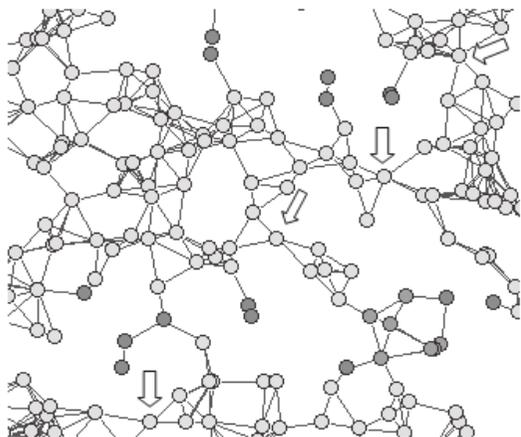


図5 二重スターグラフ解決前 RGNコミュニティ抽出

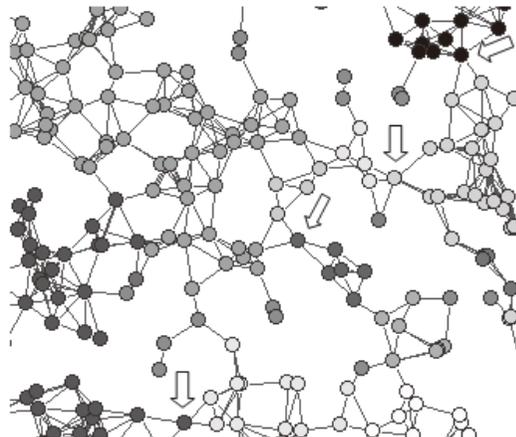


図6 二重スターグラフ解決後 RGNコミュニティ抽出

の抽出を行わないことで、二重スターグラフを含むコミュニティを抽出する際、開始頂点の違いによるコミュニティ構造が変化する問題も解決した。

3.4 コミュニティ抽出結果

本研究ではコミュニティ抽出によって抽出された各コミュニティを、コミュニティに属するノード数が m の場合はサイズ m のコミュニティと呼び、コミュニティに属さないノードを孤立ノードと呼ぶことにする。

3.4.1 RGNを用いた既存手法との比較

現在までRGNを用いたコミュニティ抽出研究は行われていなかったが、本研究では複数のコミュニティ構造や、二重スターグラフが存在することから手法の違いによるコミュニティ構造の比較のためにこのモデルを使用した。

まず、松島の抽出手法と改良後の抽出手法を比較するため、ノード数を1,000としたRGNを使用し二重スターグラフの分割とコミュニティ抽出の精度を調べた。ネットワークのノード数を1,000とすることでリンクを形成できる一定範囲内のノード数が減るため複数のコミュニティが発生し、複数のコミュニティにより二重スターグラフ構造が比較的多く発生すると考えられる。

2つの手法を使い抽出した図5、図6の結果

を比較すると、改良後の手法では二重スターグラフの頂点である矢印部でコミュニティの分割ができていたことが、所属するコミュニティの色で判断できる。よって、本研究の改良手法は、複数のコミュニティ構造が重なり二重スターグラフが存在するネットワークにも有効だと考えられる。

3.4.2 抽出されたコミュニティの詳細比較

続いて、抽出されたコミュニティ構造の比較も行った。改良後の手法では、コミュニティ数は61個であり、サイズ20以上のコミュニティが複数存在した。一方、松島の手法ではコミュニティを10個しか抽出できず、もっとも大きいコミュニティのサイズは600以上であった。

そこで、各手法により抽出されたコミュニティの平均クラスター係数とコミュニティの平均密度の比較を行った。 $comj$ の持つ平均クラスター係数は、

$$C_{comj} = \frac{\sum_{i \in comj} C_i}{comj \text{のサイズ}}$$

で定義できる。ここで $i \in comj$ はコミュニティ j に属するノードである。また、 $comj$ の密度 ρ_{comj} は、

$$\rho_{comj} = \frac{comj \text{に存在するリンク数}}{comj \text{のサイズ} (comj \text{のサイズ} - 1)}$$

で定義でき、全コミュニティの平均密度 P は、

$$P = \frac{\sum_{j=1}^N \rho_{conj}}{N}$$

で定義できる[3][11]。ここで N は全コミュニティ数である。

上記式を用いてRGNから抽出されたコミュニティのクラスター係数と密度を調べた結果、松嶋の手法ではクラスター係数が0.650、密度が0.602であった。一方改良手法ではクラスター係数が0.574、密度が0.697であった。二つを比較すると、改良後ではコミュニティのクラスター係数が低くなっているが、これは二重スターグラフを分割したため同一コミュニティに属するノード数が減り、二重スターグラフ頂点ノードが持つクラスター係数が平均に影響したと考えられる。逆に密度は改良後の方が高くなっており、二重スターグラフを分割することでネットワーク内からより密度の高いコミュニティ構造の抽出に成功しているといえる。さらに、複数の密度が高いコミュニティ構造の抽出は、RGNの持つ一定範囲内のノードとリンクを接続し成長する特徴とも一致する。これにより、本手法のコミュニティ抽出精度は高いと考えられる。

3.4.3 ASネットワーク

ASネットワークに対し改良手法によるコミュニティ抽出を行った結果、ネットワーク内からコミュニティを83個抽出できた。これらの抽出されたコミュニティにはサイズ11,111のコミュニティが1つ存在し、それ以外のほとんどのコミュニティはサイズ3であるため、ネットワークの約2/3 (22,010ノード)を孤立ノードが占めていることが判明した。これは、ASネットワークではネットワーク内で最大のサイズを持つコミュニティに属するノードにリンクが集中し、最大サイズのコミュニティに属さないノードはリンク本数が少なくなり、孤立ノードとなることを意味している。また、三角形リンクをコミュニティとして抽出する本手法と

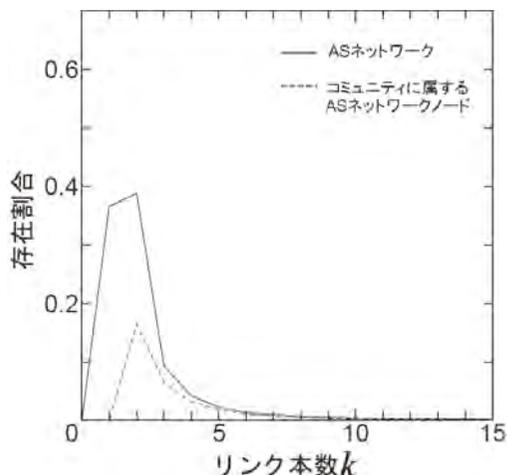


図7 コミュニティに属するノードのリンク数存在割合

ASネットワークのスケールフリー性により、コミュニティ内には、リンク数が非常に多いハブノードが複数存在すると考えられる。

さらに、孤立ノードのリンク数分布を調べたところ、ほぼ全てのリンク数が1か2であり、ネットワークに属するリンク数3以上のノードは、ほとんどが最大サイズのコミュニティに属していることがわかった(図7)。

3.4.4 BA、RN、HK、RGNモデル

BAモデルでは最大サイズ188の小さなコミュニティしか抽出できず、一方RNモデルではサイズ3のコミュニティが少数しか存在せず、ネットワークの99.8%以上が孤立ノードであった。これは、BAモデルRNモデルともにクラスター係数がほぼ0の値をとるため、リンクが密な部分構造がほとんどないことが原因と考えられる。しかし、ネットワーク内に存在するリンク関係が密な部分のコミュニティ抽出はできしており、クラスター性が0ではない限りコミュニティを効率的に抽出できることがわかった。

次に、ASネットワークに近いクラスター係数0.224を持つHKモデルから抽出したところ、コミュニティ数は2,994個であり、このうちサイズ10,978のコミュニティは1つのみで、それ以外のコミュニティはほとんどがサイズ3で

あった。コミュニティの最大サイズはASネットワークのもっともサイズが大きいコミュニティに近い結果であり、抽出されたサイズ3のコミュニティの数に比例し減少したがネットワークの約半数が孤立ノードであった。また、クラスター係数を0.574にしたHKモデルでは、抽出されたコミュニティ数が455となり、最大のコミュニティサイズが29,157と増え、孤立ノードが3,013と大きく減少した。

上記二つのHKモデルのコミュニティ数がASネットワークと異なるのは、各ノードの持つ最も少ないリンク本数である最小リンク数の影響が原因である。二つのネットワークを比較すると、ASネットワークの36.56%は最小リンク数が1であるのに対し、HKモデルの最小リンク数は2であった。クラスター性に着目した抽出法の場合、コミュニティの抽出には最低リンク本数が2本必要であり、最小リンク数2を持つHKモデルでは、小さなコミュニティ構造が多量に抽出されたと考えられる。

最後にRGNからコミュニティを抽出したところ、抽出されたコミュニティ数は29個で、最大のコミュニティサイズは33,096であり、ほぼ全てのノードが同一コミュニティに属していた。RGNではほぼ全てのノードが隣接するノードと三角リンクを密に形成している特徴と今回使用したモデルが平均リンク数5を持つためネットワークのクラスター係数と最大コミュニティサイズが大きくなったと考えられる。

これらの結果から、現実ネットワークの特徴であるスモールワールド性、スケールフリー性、クラスター性を持つネットワークでは、スケールフリー性を持つネットワークではリンク数が多いノードを中心としたコミュニティ構造がコミュニティの大部分を占めていることがわかった。また、クラスター性があるノードの大部分は最大サイズのコミュニティに属し、それ以外のノードの大部分は孤立ノードとなり、クラスター性が高くなるほどコミュニティの最大サイズが大きくなることが判明した。

4. 情報拡散速度

リンク関係が密なコミュニティ構造は情報拡散に大きく影響を与えていると考えられるため、本研究では、抽出されたコミュニティトポロジ情報を用いて情報拡散速度を調べた。また、現実ネットワークに対する効果的な情報拡散手法の比較を行い、コミュニティ構造が情報拡散に与える影響について調べた。

4.1 情報拡散速度シミュレーション方法

本研究では、情報付与ノードによる拡散速度の違いに着目し、効果的にネットワーク全体に情報を拡散できるノードを調べた。なお、シミュレーションにはSIモデルを使用し、情報を与えるノードによる情報拡散速度の比較を行った。SIモデルは生物間ウイルス感染解析に用いられるモデルの1つであり、ノードの状態をS (susceptible) : 健康状態、I (infected) : 感染状態の2通りで判断する[12]。感染方法は、Iのノードが隣接するSノードに感染活動を行い、感染確率 λ で健康ノードSはIに変化する。今回のシミュレーションでは、Iを情報所持状態、Sを情報未所持状態と想定し、情報伝播確率 λ を1.0とした。また、情報所持ノードが隣接ノードに対し情報拡散する一連の処理を1ターンとし、ネットワーク内の全ノードに情報が行き渡るまでのターン数を比較した。

情報を最初に付与するノードの選択法として、ランダム手法とCohenらが提案した知人手法[6]、コミュニティトポロジによる本研究の

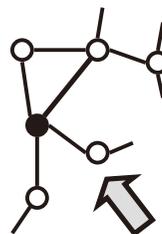


図8 知人手法の初期情報付与ノード選択

ノード (黒) : ランダム選択ノード
ノード (矢印) : 初期情報付与ノード

手法を用いた。このランダム手法は全ノードからランダムにノードを選択し情報を付与する手法であり、知人手法は図8のように、全ノードからランダムにノードを選択し（黒ノード）、隣接するノード群からランダムに選択したノード（矢印）に情報を付与する手法である。また本研究のコミュニティトポロジによる手法で使用するトポロジ情報として、コミュニティに属しているノード情報、コミュニティに属していないノード情報、二重スターグラフの頂点情報を使用し、これらのノードに情報を与え、ネットワーク全体に情報が拡散されるターン数を知人手法と比較した。

また、各手法をネットワーク全体の情報を必要とする大局的手法と全体トポロジを必要としない局所的手法に分類し、現実ネットワークであるASネットワークに対する有効な拡散法を調べた。さらに、初期情報付与ノードのコミュニティトポロジ情報の違いによる拡散ターン数の変化が、ネットワーク全体にどのように影響を与えているのか調べた。

4.2 コミュニティ構造による拡散速度の違い

ネットワークごとの拡散速度は、ネットワークモデルごとに5つの異なるネットワークデータを用意し、5回の計測をノード数(33,326)分行った。よって各ネットワークモデルの情報拡散速度試行回数は833,150回となっている。

始めに、スモールワールド性を持つネットワークモデルとASネットワークを用いてコミュニティ構造の違いによる情報拡散速度の違い

いを、ランダム手法を用いて全ノードに情報が拡散されるまでのターン数を平均することで求めた(表2)。

まず、もっとも拡散速度が遅かったのはRNモデルの11.910ターンであった。次にASネットワークは8.658ターン、BAは8.413であり、クラスター係数を変化させたHKモデル(C=0.224, 0.574)は8.952ターン、10.415ターンであった。この結果、BAとASネットワークではコミュニティサイズが全く異なるが拡散性はほとんど変化がないため、コミュニティサイズの違いは情報拡散に大きな影響を与えないと考えられる。しかし、クラスター係数を変化させコミュニティサイズとコミュニティ数が増減したHKモデルは拡散速度が変化している。ASネットワークとクラスター係数、コミュニティサイズが近似しているHKモデル(C=0.224)の拡散速度がASネットワークより遅いことからコミュニティ数が影響しているのではないかと考えられる。この結果から情報拡散にはクラスター性やコミュニティのサイズは大きく影響せずコミュニティ数が影響を与えていると考えられるが、今回使用した各モデルでは拡散速度に直接影響を与えている原因は求められなかった。

4.3 各初期情報付与手法の結果

次にASネットワークに対し複数の拡散手法を用いてコミュニティ構造の与える影響を調べ

表3 現実ネットワーク (AS) 上での各手法による情報拡散速度

初期情報付与選択手法	手法の分類	平均情報拡散ターン
ランダム手法	局所的手法	8.658
知人手法	局所的手法	7.902
コミュニティに属しているノードを選択	大局的手法	8.200
コミュニティに属していないノードを選択	大局的手法	8.846
二重スターグラフ頂点のノードを選択	大局的手法	7.800
コミュニティ知人手法	局所的手法	7.581

表2 現実ネットワーク (AS) と各モデルの情報拡散速度

ネットワーク	平均情報拡散ターン	最速情報拡散ターン	最遅情報拡散ターン
AS	8.658	7	12
BAモデル	8.413	6	10
RNモデル	11.910	10	15
HKモデル	8.952	7	11
(C=0.224, 0.574)	10.415	8	13

た(表3)。まず、各手法を比較する前に、ネットワーク内で最も効率的に情報拡散できるノードと、最も非効率的なノードの抽出を行った。その結果、ネットワーク内に情報を拡散するのに最も効率的なノードは7ターンでネットワーク全体に情報を拡散することができ、反対に非効率的なノードは12ターン必要であった。この効率的なノードは、ネットワーク内で0.0028%ほどしか存在しないため、ネットワークトポロジ情報がない状態で効率的に情報拡散できるノードを見つけることは難しいといえる。

4.3.1 局所的手法

Cohenらの知人手法は、情報を所持するノードに隣接するノードの情報のみを用いて情報拡散を行うため、局所的手法である。一方ランダム手法もトポロジを使用しないため局所的手法の一つである。

まずランダム手法を使用した場合、全体に情報拡散されるまで、平均8.658ターン必要であった。また、知人手法を使用した場合、平均7.902ターンで拡散し、約30%が7ターンでネットワーク全体に拡散した。一般に情報拡散ではリンク数が多いノードほど情報の拡散に優れているが、特にリンク数分布がべき乗則に従うネットワークでは、ランダムに選ばれたノードの隣接ノードはリンク数が非常に大きくなるのが原因である[6]。よって、リンク数が多いノードを選択する可能性が高い知人手法を用いた手法は有効な情報拡散手法の一つといえる。

4.3.2 大局的手法

各手法のうち、ネットワーク全体のトポロジ情報あるいは、一定数以上のノード情報を必要とする手法を大局的手法と呼ぶ。本研究のコミュニティトポロジ情報を用いた手法は、一度ネットワーク全体のコミュニティを抽出する必要があるため大局的手法である。

実際にコミュニティ抽出によるトポロジ情報をもとに情報を拡散したところ、コミュニティに属しているノードに情報を与えた場合は平均8.2ターン、コミュニティに属していないノード

は平均8.846ターン、二重スターグラフの頂点ノードは7.8ターンでネットワーク全体に情報が拡散した。二つの局所的手法と比較し、コミュニティに属しているノードは、ランダム手法より効果的にネットワーク全体へ情報拡散することができた。また、二重スターグラフの頂点ノードに情報を与える手法は知人手法よりも効果的であった。

各手法のネットワーク全体への情報の広がり方を比較したところ、所属コミュニティ内への情報拡散は非常に速いが、所属コミュニティ外に対する情報拡散はターン数がかかった。この結果、ASネットワークのようなクラスタ性のある現実ネットワークでは、コミュニティに属しているノードや、所属コミュニティと所属コミュニティ外の中心となる二重スターグラフの頂点ノードが、情報拡散に優れるノードだと判明した。しかし、二重スターグラフの頂点ノードの抽出やネットワーク全体のコミュニティ抽出には全ネットワークのトポロジ情報が必要であり、現実世界で実行可能な手法とはいえない。

4.3.3 提案情報拡散手法

本研究では、局所的手法の知人手法と局所的なコミュニティトポロジをあわせた改良手法を使い、現実的に有効な拡散手法提案した(コミュニティ知人手法)(図9、図10)。コミュニティ知人手法は、ランダムに選択されたノード(黒ノード)に隣接するノードのうち、同一のコミュニティに属しているノード(点線ノード)からノードをランダムに選択(矢印)し情報を与える方法である。また、隣接ノードに同一コミュニティが存在しない場合は知人手法と同様にランダムに隣接ノードを選択する。

シミュレーションの結果、情報拡散速度は平均7.581で、知人手法よりも効果的であり、またランダムに選ばれたノードのうち約50%が7ターンでネットワーク全体に情報を拡散させることができた。さらに、本研究で使用したコミュニティ抽出法は隣接するノードのリンク関

```

#コミュニティ知人手法
community()
start_node ← ランダムに選ばれたノード
comA[] ← start_node
lan ← 1
listA[] ← start_nodeの隣接ノードを格納
la ← listA[]のサイズ
WHILE lan ≠ la DO
  listB[] ← listA[lan]の隣接ノードを格納
  lb ← listB[]のサイズ
  FOR i=1 to lb DO
    IF listA[] ≧ listB[i] THEN
      IF comA[] ≠ listA[lan] AND comA[] ≠ listB[i]
THEN
        comA[] ← listA[lan]
        comA[] ← listB[i]
      ENDIF
    ENDIF
  ENDFOR
  lan ← lan + 1
ENDWHILE
l ← acquaintance_initiation(comAのサイズ)
IF l ≠ 0 THEN
  comA[l] ← 情報付与
ELSE
  listA[random(la)] ← 情報付与
ENDIF

#情報付与ノード選択
acquaintance_initiation(num)
IF num > 0 THEN
  RETUEN with random(num)
ELSE
  RETUEN with 0
ENDIF
RETUEN with 0

```

図9 コミュニティ知人手法疑似コード

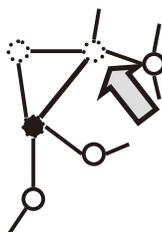


図10 コミュニティ知人手法の初期情報付与ノード選択

ノード (黒) : ランダム選択ノード
 ノード (点線) : コミュニティノード
 ノード (矢印) : 初期情報付与ノード

係のみを用いた手法であるため、コミュニティ知人手法は全ネットワークのトポロジ情報が必要としない局所的な手法であり、現実的に実現可能である有効な情報拡散手法といえる。

5. まとめ

本研究で挙げたクラスター性を用いた改良コミュニティ抽出手法は、クラスター性が0でないネットワークに対しコミュニティ抽出が可能である。本研究での抽出の結果、ASネットワークや構造が似ているネットワークは、ネットワーク内でもっとも大きいサイズのコミュニティに所属するノードと、コミュニティに所属しない孤立ノードで構成されていることがわかった。この改良手法はネットワークの一部情報のみを使用する局所的抽出手法であるため、現実的なコミュニティ抽出手法の一つだといえる。

ネットワーク全体に情報を拡散するシミュレーションでは、初期情報付与ノードの選択手法はCohenらの手法よりも、コミュニティトポロジ情報を用いた拡散手法のほうが有効であった。これは、所属コミュニティと所属コミュニティ外の中心となる二重スターグラフ頂点ノードが両方のコミュニティに接しているため、結果的にネットワーク全体へ効果的に情報を拡散できるからである。さらに、提案手法であるコミュニティ知人手法では、二重スターグラフの頂点ノードや多リンクノードに情報が付与されやすく、手法自体も局所的な手法であるため、ネットワーク全体のトポロジがわからなくとも情報拡散に適したノード抽出が可能である。

以上により、本研究で使用した改良コミュニティ抽出手法と、コミュニティ知人手法を使用することで、口コミを使ったマーケティングの有効な情報拡散や、緊急連絡のための情報拡散を効果的に行うことができると思われる。

今後の課題としては、

- 現実ネットワークに存在する片方向リンクや両方向リンクの方向性を考慮したコミュニ

ティを分割する手法の提案。

- 松嶋の提案コミュニティ抽出手法以外のコミュニティ抽出手法とのコミュニティ構造比較。
の二つが挙げられる。

【参考文献】

- [1] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks", *Phys. Rev.*, E69, 026113 (2004).
- [2] 松島祐：ネットワーク構造のクラスター性に着目した局所的コミュニティ抽出法に関する研究, 名古屋工業大学大学院修士論文 (2011).
- [3] 松島祐, 加藤竜丸, 武藤敦子, 犬塚信博：ネットワーク構造の推移性に着目した局所的クラスタリング手法の提案, 情報処理学会第74回全国大会 (2011).
- [4] 千石靖, 岡本栄司, 満保雅浩, 植松友彦：コンピュータウイルスの拡散と消滅の大局的振る舞いについて, *情報処理学会論文誌*, Vol. 37, No. 4, 579-587 (1996).
- [5] R. Pastor-Satorras and A. Vespignani, "Evolution and Structure of the Internet --A Statistical Physics Approach--", CAMBRIDGE UNIVERSITY PRESS (2004).
- [6] R. Cohen, S. Havlin and D. ben-Avraham, "Efficient Immunization Strategies for Computer Networks and Populations", *Phys. Rev. Lett.*, Vol. 91, No. 24, 247901 (2003).
- [7] CAIDA, Center for Applied Internet Data Analysis <http://www.caida.org/>, (2009. 11. 20).
- [8] 増田直紀, 今野紀雄：複雑ネットワーク--基礎から応用まで--, 近代科学社 (2010).
- [9] 遠野翔太, 森口一郎：アドホックネットワークでの浸透閾値を用いたウイルス蔓延抑制, *東京情報大学研究論集*, Vol. 17, No. 1, pp. 43-50 (2013).
- [10] R. Albert and AL. Barabási, "Statistical mechanics of complex networks", *Rev. Mod. Phys.*, Vol. 74, 47-97 (2002).
- [11] 齊藤正高：文系学生のためのネットワーク科学の初歩と Pajek の使用法 (1) <http://www.saitomas.net>
- [12] 本吉和馬, 森口一郎：局所ネットワーク構造を用いたウイルス蔓延抑制, *東京情報大学研究論集*, Vol. 15, No. 2, pp. 1-11 (2011).