

## 成長型ネットワーク向け安定コミュニティ分割手法の提案

大和田 純 吉井 伸一郎 古川 正志  
北海道大学 北海道大学 北海道大学

本研究では、成長する複雑ネットワークの一例として Web 上のブログのネットワークを扱い、効率的な情報抽出のためのコミュニティ分割手法の比較検討を行う。まず、既存のコミュニティ分割手法を動的に変化するネットワークに対して適用し、各分割手法を分割コミュニティのモジュール度 (Modularity) と安定度 (Stability) の観点から評価する。様々なネットワークトポロジと、それに対する各コミュニティ分割手法の振る舞いを観察し、得られた知見をもとにブログネットワークに適したコミュニティ分割手法を考察する。

## Stable Dividing Method for Network Community on Evolving Networks

Jun Ohwada Shinichiro Yoshii Masashi Furukawa  
Hokkaido University Hokkaido University Hokkaido University

Recently, many studies have been made on complex networks and finding their community by means of dividing them based on their network topology. However these studies usually are interested in static networks rather than evolving one. This paper describes how efficient the existing methods are for dynamically changing networks. Then, we propose an evaluation score of divided communities modularity and stability, and apply for various kinds of complex network topology as well as blog network.

### 1 はじめに

複雑なシステムをネットワーク的視点から構成要素とその要素間の関係で捉え、要素の中身を解析することなくトポロジのみによってネットワークをコミュニティに分割する手法が数多く提案されている [1, 2, 3]。一方で、ここ数年の Web の急速な成長に伴い、Web 上の膨大な情報の中からユーザにとって有益なものを取り出す方法が求められている。しかし、Web は絶えず成長を続ける動的なネットワークであり、従来の静的なネットワークを対象とした分割手法では、ネットワークの変化に応じた効率的な情報抽出は難しい。また、Web のネットワークは、古典的なランダムグラフや正則グラフとは大きく異なる複雑な構造を持つことが知られているが、既存の手法は様々なネットワークトポロジに対する有効性は考慮されていない。

そこで本研究では、既存のコミュニティ分割手法の成長する複雑ネットワークに対する有効性を議論するために、分割コミュニティのモジュール度 (Modularity) と安定度 (Stability) で評価する手法を提案する。特にネットワークのトポロジと分割手法の組み合わせによって決まる評価値に着目し、種々の分割手法がどういったトポロジの成長するネットワークに対して有効かを明らかにする。

コミュニティ分割の対象とする成長型ネットワークとして、BA モデル [4] ネットワークや CNN モデル [5] ネットワークといった理論モデルに加え、Web 上の実際のデータをもとに構築したブログのネットワークを扱った。理論モデルでの実験結果と比較した実データでの実験結果についても考察し、より実践的なコミュニティ分割手法を検討する。

## 2 背景

研究の背景について述べる。

### 2.1 コミュニティ

「コミュニティ」という言葉には、元来より「地域共同体」という物理的に近いものの集合体を指す意味がある。また、ことさら Web においては、特定のトピックを取り上げてコミュニケーションを図るために自発的に集まったユーザ団体の意味で用いられることが多く、物理的な近さよりも、興味や趣味の近さで集まる集合体としての意味合いが強い。

一方、ネットワーク中のコミュニティとは密なリンクで繋がるノードの集合である。言い換えれば、強い関係性を持ったノードの集まりである。ネットワークに含まれる各ノードは絶対的な位置を持たず、他のノードとの関係がリンクによって相対的な位置として表現されている。

コミュニティ分割とは、その名の通り、リンクの情報をういてネットワーク中のノードをいくつかのサブネットワークに分けることを指す。ここで、分割されたサブネットワークを分割コミュニティと呼ぶ。ネットワーク研究の分野では尤もらしい分割コミュニティの分け方を見付ける数々の手法が提案されている [1, 2, 3]。

### 2.2 ブログネットワーク

ブログ (Weblog, Blog) は従来の Web サイトと異なっていくつかの特徴を持つ。主な特徴を以下に挙げる。

1. エントリ (entry) と呼ばれる記事が時系列に管理されている。通常は新しいエントリほどページのトップなど目に付きやすい位置に表示される
2. エントリごとに個別の不変な URL (Permalink) が与えられる
3. コメント、トラックバックといったコミュニケーションのための機能を持つ

ブログは、ツールの日本語化とともに 2002 年頃から国内でも急速に普及した。ブロガー (Blogger) の層は幅広く、その内容も多岐に渡る。総務省の発表によると、2006 年 3 月末現在の日本でのブロガー数は 868 万人に達するといわれており、今後もさらに増えていくと予想される [6]。特徴 1 のためにブログには新しい話題が取り上げられていることが多く、ユーザによる頻繁な訪問とコミュニケーションが起こる。そしてブログを広く普及させた特徴として特筆すべきは特徴 2、特徴 3 に挙げ

た項目である。ブロガーはひとつのエントリについてひとつの話題を書くことができ、エントリにはコメントとトラックバックという 2 つのコミュニケーション機能が備わっている。このためブログの閲覧者は、自分の興味を持ったエントリに対して感想を述べたり、さらにはそのエントリを自身のブログで取り上げて話題を発展させたりと、双方向のコミュニケーションを築くことができる。

ブログ同士がエントリ間のハイパーリンクやトラックバックで繋がると、ブログのネットワークが形成される。ブログをノードとし、ハイパーリンクやトラックバックをリンクとしてネットワークを構成すると、ブログネットワークは時間経過によって成長するネットワークである。

2 つのブログがハイパーリンクやトラックバックで繋がるとき、それら 2 つのブログ間には共通の興味、関心、あるいは話題などがあると考えられる。このことから、ブログネットワークはブロガーたちの興味関心ネットワークと言い換えることができる。ブロガーの興味や関心が似ていれば似ているほど、2 つのブログはネットワーク上で近い位置に配置される。密なリンク構造を持ったブログネットワーク中のコミュニティは、共通の興味を持ったブログの集合であると言える。ブログネットワークから意味のあるコミュニティを見付けられれば、特定の話題に関する密度の高い情報源として利用することができるだろう。

## 3 コミュニティ分割手法

本研究で扱う数種類のコミュニティ分割手法を説明する。

コミュニティ分割手法は、扱うパラメータによって大きく 2 種類に分けられる。ひとつは、対象となるネットワークの大局的パラメータを用いて分割を行うもの、もうひとつは、ネットワークの局所的パラメータを用いて分割を行うものである。

例えば Newman らは、Freeman らによって提唱されたノードのビトウィーンネス (Betweenness) [7] をリンクに拡張したリンク・ビトウィーンネス (Link Betweenness) を基に分割を進める手法を提案した [1]。あるリンクについてのリンク・ビトウィーンネスは、ネットワーク全体の構造から算出されるので、Newman らの手法は大局的パラメータを用いた手法である。別の例で、Radicchi らは、ノードのクラスタリング係数 (Clustering

Coefficient) をリンクに拡張したリンクのクラスタリング係数を基に分割を進める手法を提案した [2]。あるリンクについてのクラスタリング係数は、そのリンクと接続しているノード、及び近傍のリンクの情報のみから決まるので、Radicchi らの手法は局所的パラメータを用いた手法である。両者の特徴を比較してみると、まず前者は、ネットワークにわずかでも変化が起きた場合、すべてのパラメータを再計算する必要があるため、その度にコミュニティ分割を行うと計算量が膨大になる。対して後者は、変化が起きた箇所の近傍についてのみ計算を行えばよいので、過去の情報を再利用してコミュニティ分割を実行できる。

また、分割のアプローチによって手法を分類すると、トップダウンのアプローチとボトムアップのアプローチに分けることができる。

前述の Newman らの手法と Radicchi らの手法は、どちらもひとつのネットワークを出発点とし、徐々にネットワークを非連結にしてサブネットワークを抽出する。このようなアプローチをトップダウンのアプローチとする。これとは対称的に、Clauset らが提案した手法 [3] では、最初にそれぞれのノードをひとつのコミュニティとみなし、コミュニティ内でのリンクが密になるようにコミュニティ同士を結合させていく。Clauset らの手法が取るようなアプローチをボトムアップのアプローチとする。

本研究では、先に示したように異なる特徴を持つ分割手法として、Newman らの手法、Radicchi らの手法、Clauset らの手法、Radicchi らの手法に一部変更を加えた比較手法の 4 種類を成長する複雑ネットワークに適用し、比較・検討の対象とする。

### 3.1 比較手法

Radicchi らの提案手法はネットワークの局所的パラメータを利用するため高速であり、以下に示す 2 つの特徴を持つ。

1. ネットワーク中に次数 1 のノードがあると、多くの場合で次数 1 のノードを 1 つのコミュニティとして抽出する
2. 定量化したコミュニティの定義が厳しすぎるため、手順の早い段階で分割がストップする

特に複雑ネットワークにおいてはこれらの特徴は顕著に結果として表れる。図 1 はノード数 150 のネットワークを Radicchi らの手法を用いて分割した結果である。

同じ色が与えられたノードは同一のコミュニティに属しており、この例では 2 つの分割コミュニティが抽出された。さらなる分割を行うと分割コミュニティがコミュニティ定義式を満たさなくなってしまうので、図に示す時点で分割がストップしている。もっと極端な例では、分割がまったく行われずに、ネットワーク全体をひとつの分割コミュニティとして結果を返すこともある。このような結果では、他の手法との分割コミュニティの比較は難しい。

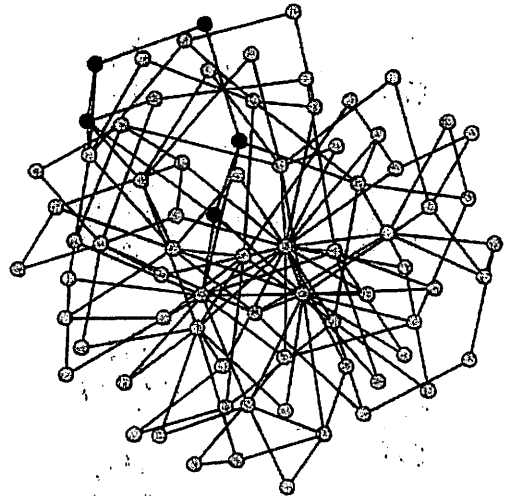


図 1 手順の早い段階で分割がストップした例

そこで、Radicchi らの手法に変更を加え、特徴比較のための新手法を提案し、以降の図表中ではこれを「Radicchi+」と表記する。具体的な変更は下記の 2 つである。

- 変更 1 次数 1 のノードに接続されたリンクのクラスタリング係数を 1 より大きい値とする
- 変更 2 分割コミュニティの評価に定量的な定義式を用いず、モジュール度が最大となるところまで分割を進める

最後に、検討の対象とするコミュニティ分割手法の特徴を表 1 にまとめる。

## 4 コミュニティの安定度評価手法の提案

ネットワークが成長し、新たなノードやリンクが追加されれば、それに伴ってコミュニティ分割の結果が変わるのは明らかである。変わらずに形を保ち続ける分割コミュニティもあれば、成長前には別々のコミュニティに

表1 各コミュニティ分割手法の特徴

分割手法	用いるパラメータ	アプローチ
Newman	大局的	トップダウン
Clusset	大局的	ボトムアップ
Radicchi	局所的	トップダウン
Radicchi+	局所的	トップダウン

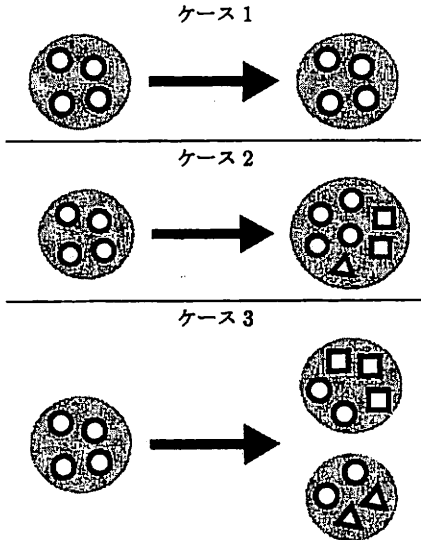


図2 あるコミュニティを主体として見た変化のパターン

分けられていたノード間にリンクが発生し、成長後に同一のコミュニティに分類される例もある。逆に、同一のコミュニティに属していた2ノードが、ネットワーク構造の変化によって成長後には別々のコミュニティに分離されることもある。分割結果として得られるコミュニティがどういった状態になるときにそれを安定とするかを、本研究では各分割コミュニティから見た変化のパターンを3つのケースに分け、それらの評価を考える。

- ケース1 ネットワークの成長後、自身と等しいノード集合から成る分割コミュニティが抽出された場合
- ケース2 ネットワークの成長後、自身を部分集合とするような分割コミュニティが抽出された場合
- ケース3 ネットワークの成長後、自身の構成要素であるノードが、2つ以上の分割コミュニティに分かれた場合

それぞれのケースを図2を用いて説明する。図中の丸、四角、三角はネットワーク中のノード、影のついている部分がひとつのコミュニティを表す。矢印の左側が主体となる分割コミュニティの成長前の形であり、矢印の右側が主体となるコミュニティに属するノードが成長後にどのようなコミュニティに属しているかを表す。ケース1では、主体となる分割コミュニティの形がそのまま維持されているので、これを「安定である」と評価する。続いてケース2では、主体となる分割コミュニティはより大きな分割コミュニティの一部となっており、捉え方によっては分割コミュニティが形を失っているとも解釈できる。本研究では、主体となる分割コミュニティ内のノード同士の関係性は維持されているものとし、このケースも「安定である」と評価する。最後にケース3を取り扱い、このときは主体となる分割コミュニティに含まれるノード間の関係が失われてしまっているので「安定ではない」と評価する。

以上を踏まえ、コミュニティ分割手法によって抽出される分割コミュニティ安定度を以下のように定量化する。

$$S_t = \frac{1}{m+n} \sum_{i=1}^m \max_{1 \leq j \leq n} f(C_t^i, C_{t+1}^j) + \frac{1}{m+n} \sum_{j=1}^n \max_{1 \leq i \leq m} f(C_{t+1}^j, C_t^i)$$

ここで、

$$f(A, B) = \begin{cases} 1 & : \text{if } A \subseteq B \\ -1 & : \text{otherwise} \end{cases}$$

とする。式中の  $m$  はステップ  $t$  において抽出された分割コミュニティ数、 $n$  はステップ  $t+1$  において抽出された分割コミュニティ数、 $C_t^i$  はステップ  $t$  における  $i$  番目の分割コミュニティ、 $C_{t+1}^j$  はステップ  $t+1$  における  $j$  番目の分割コミュニティをそれぞれ表す。ステップ  $t$  において抽出されるコミュニティと、ネットワーク成長後のステップ  $t+1$  で抽出されるコミュニティを比較して安定度を計測する。

安定度の計算には、ステップ  $t$  において抽出される分割コミュニティを主体としてステップ  $t+1$  における分割コミュニティと比べるばかりではなく、ステップ  $t+1$  の分割コミュニティを主体としてステップ  $t$  での分割コミュニティと比較し、その場合の評価を計算に加えるようにした。ネットワークの成長の前後で、1つのコミュニティが2つのコミュニティに分裂することと、2つのコミュニティが1つのコミュニティに統合することは、

表 2 実験時に用いたパラメータ

パラメータ	値
BA モデルの $m_0$	2
BA モデルの $m$	2
CNN モデルの $u$	0.5
ランダム選択成長モデルの $m_0$	2
ランダム選択成長モデルの $m$	2

表 3 各ネットワークの性質

ネットワークの種類	SW 性	SF 性
BA モデル	無し	有り
CNN モデル	有り	有り
ランダム選択成長モデル	無し	無し

コミュニティの変化という点では本質的に同じである。従って、変化の方向は安定を考える上で考慮すべきではない。今回の実験ではネットワークはノードとリンクの追加のみで成長させたが、ノードやリンクの除去による変化を考えたときも、ここで提案した式を適用できる。

## 5 実験と結果

本研究で行った実験について説明する。

### 5.1 実験設定と手順

まず、実験の設定について説明する。本研究では、成長するネットワークの理論モデルとして BA モデル、CNN モデル [5]、BA モデルにおける優先的選択 (Preferential Attachment) をランダム選択に置き換えたランダム選択成長モデルを採用した。

実験時に設定した各モデルのパラメータを表 2 に示す。これらのパラメータは、比較を容易にするためにすべてのモデルの成長過程でノード数とリンク数の比がほぼ等しくなるように設定した。加えて、表 2 に示したパラメータで各モデルを十分なステップ成長させた状態でのネットワークのスモールワールド (SW) 性、スケールフリー (SF) 性の有無を表 3 に示す。なお、今回の実験では最終成長ステップ  $T$  は 500 とした。

続いて、実験の手順について説明する。手順は次に示す通りである。

1.  $t = 0$  の初期状態となるネットワークを生成する
2. 各ネットワークに対して 4 つの手法でコミュニティ

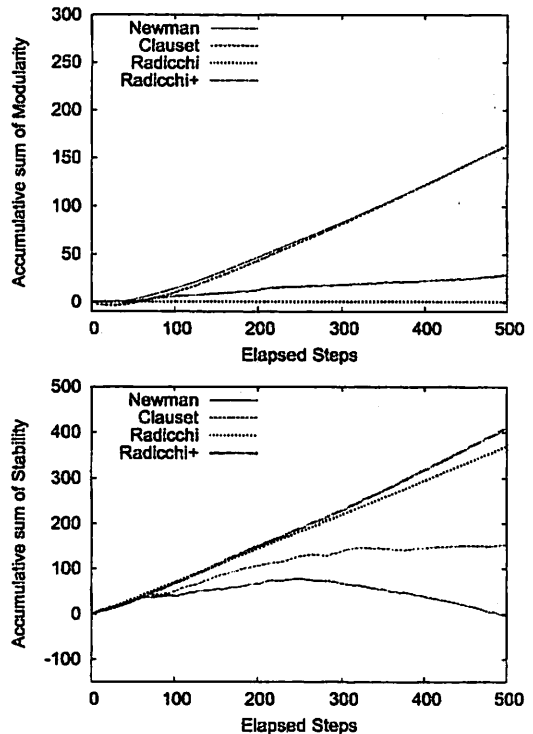


図 3 BA モデルでの結果

分割を行い、コミュニティ分割結果から手法ごとにモジュール度  $Q_t$  と安定度  $S_t$  を計算し、累計を求める

3. ネットワークを 1 ステップ成長させ、 $t \leftarrow t + 1$  とする
4.  $t$  が  $T$  ならば終了する。そうでなければ手順 2 へ戻る

以上の設定と手順で実験を行った。5.2 節、5.3 節にて結果を示し、考察を加える。

### 5.2 理論モデルでの実験結果

理論モデルから生成したネットワークに対する実験結果を図 3~図 5 に示す。

BA モデルネットワーク (図 3) では、Newman らの手法、及び Clauset らの手法で高いモジュール度と低い安定度を示した。それとは逆に、Radicchi らの手法と Radicchi+ の手法では低いモジュール度と高い安定度を示した。モジュール度と安定度の双方が高い理想的な手法は、今回の実験では発見されなかった。ネットワー

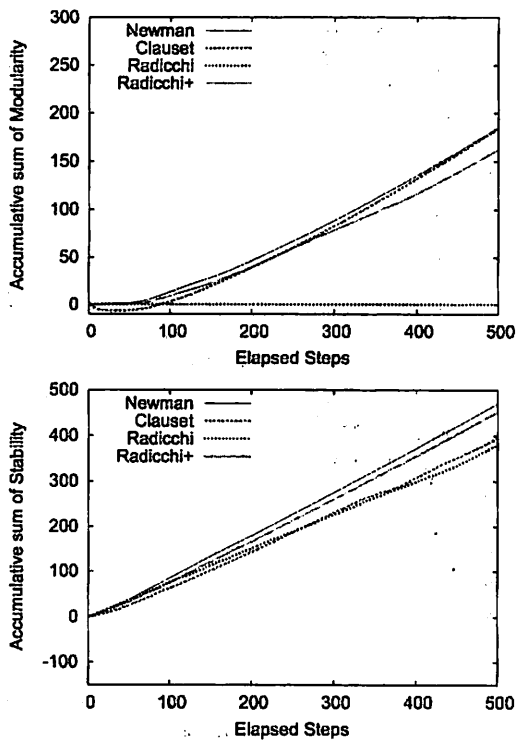


図4 CNNモデルでの結果

クの大局的なパラメータを基に各ステップでモジュール度を最大化するために細かく分割を行ってしまうと、成長の前で得られる分割コミュニティは形が維持されず、安定度は下がってしまう。一方、局所的なパラメータを用いて各ステップで大まかな分割を行うと、安定度は高くなるがモジュール度は低くなる。BAモデルネットワークのようなスモールワールド性を持たないネットワークには潜在的に密なコミュニティ構造はないので、このような結果になると考えられる。

2つ目に、CNNモデルネットワーク(図4)では、Newmanらの手法、Clauasetらの手法、Radicchi+の手法でモジュール度・安定度、共に高い値を示した。Radicchiらの手法は安定度こそ他の3手法と並ぶほどに高くなるが、モジュール度はほぼ0のまま増加しない。

また、BAモデルネットワークでの結果と比較して、モジュール度と安定度の双方で高い値を示す分割手法の存在が特徴的である。スモールワールド性を持つCNNモデルネットワークは、密なリンク構造を有しているため、多くの分割手法で安定した分割コミュニティが現れ

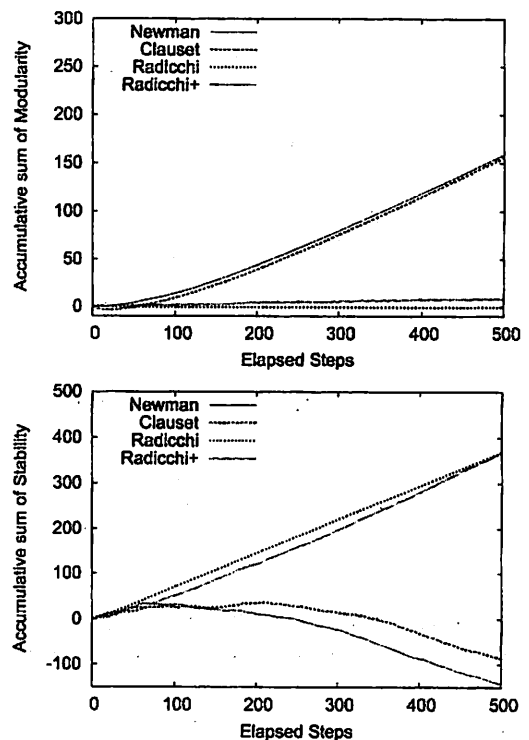


図5 ランダム選択成長モデルでの結果

やすい。

3つ目に、ランダム選択成長モデルネットワーク(図5)では、Newmanらの手法とClauasetらの手法で高いモジュール度と低い安定度、Radicchiらの手法とRadicchi+の手法で低いモジュール度と高い安定度となった。この結果はBAモデルネットワークでの結果と一致しており、密なリンク構造を持たないネットワークでは、分割コミュニティのモジュール度と安定度はトレードオフの関係にある。また、BAモデルネットワークとランダム選択成長モデルネットワークの違いとしてスケールフリー性の有無が挙げられるが(図3参照)、今回の実験の結果に相違点は見られなかった。

### 5.3 ブログネットワークでの実験結果

理論モデルにおける実験結果を踏まえ、ブログネットワークでの実験結果を考察する。結果を確認する前に、まずは今回用いたブログネットワークの特性を明らかにする。表4は、500ステップ経過時の各成長モデルネットワークとブログネットワークの平均最短パス長( $L$ )と平均クラスタリング係数( $C$ )を比較のために並べたも

表 4 500 ステップ経過時の各モデルの  $L$  と  $C$

ネットワーク	$L$	$C$
BA モデル	3.6619	0.0509
CNN モデル	4.9573	0.2981
ランダム選択成長モデル	4.1466	0.0162
プログネットワーク	3.3105	0.1250

のである。この表から、プログネットワークがスモールワールド性を有していると確認できる。

ここで、結果を図 6 に示す。プログネットワークでは、CNN モデルネットワークと同様の結果となった。モジュール度は Newman らの手法、Clauset らの手法、Radicchi+ の手法で高い値となり、安定度は全ての手法で高い値を示した。プログネットワークと CNN モデルネットワークに共通の性質はスモールワールド性を持つことであり、2つのタイプのネットワークには安定なコミュニティ構造があると伺える。対象とするネットワークとは無関係に同様の結果を示す Radicchi らの手法を例外とすれば、プログネットワークでは分割手法の特徴に依らず、安定度したもってもらしいコミュニティ分割を実現できる。

実験結果から得られた以上の考察をまとめると、以下のようになる。

- 大局的なパラメータを用いたコミュニティ分割手法は、モジュール度の高い分割を行う
- 局所的なパラメータを用いたコミュニティ分割手法は、安定度の高い分割を行う
- スモールワールド性を持たないネットワークにおいては、モジュール度と安定度はトレードオフの関係にある
- スモールワールド性を持つネットワークにおいては、コミュニティ分割手法の特徴に依らず、モジュール度と安定度の双方が高い分割が可能である
- ネットワークの持つスケールフリー性が分割コミュニティに与える影響は、今回の実験では読み取れなかった
- 同じく高いモジュール度と安定度を得られるコミュニティ分割手法であれば、計算時間を短縮できる局所的なパラメータを用いた手法が望ましい

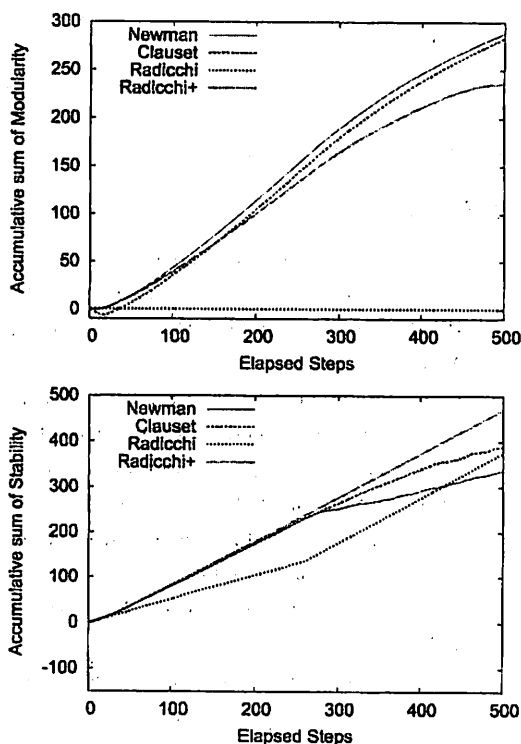


図 6 プログネットワークでの結果

## 6 おわりに

本研究では、様々なトポロジの成長する複雑ネットワークを対象とし、コミュニティ分割手法の比較検討を行った。その際、結果として得られる分割コミュニティから分割手法の安定の度合いを計る指標として安定度を導入し、モジュール度と組み合わせる評価を行った。

数種類のコミュニティ分割手法を特徴ごとに分け、複雑なネットワークのトポロジに対して各分割手法がどのように振る舞うかを観察した。プログネットワークはスモールワールド性を持ち、密なリンク構造を有しているので、分割手法に依存せず安定なコミュニティを抽出することができる。よって、大規模なプログネットワークからの情報抽出を考えるのであれば、計算の実行時間を考慮したコミュニティ分割手法を採用するべきである。局所的なパラメータを用いる分割手法であれば、ネットワークの変化に対して生じる再計算のコストが小さいので、十分に高速なアルゴリズムを適用すれば逐次的なコミュニティ分割も可能であろう。

今後は、ネットワークの自己相似性と分割コミュニティのサイズ分布についても調査を行い、より多角的にコミュニティ分割手法の特徴を調べるためにさらなる評価指標を導入する。本研究では扱わなかった分割手法も評価対象として追加していきたい。

## 参考文献

- [1] M.Girvan and M.E.J.Newman. "*Community structure in social and biological networks*". PNAS, June 11, 2002, vol.99, no.12, 7821-7826, 2002.
- [2] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. "*Defining and identifying communities in networks*". PNAS March 2, 2004 vol.101 no.9, 2004.
- [3] Aaron Clauset, M.E.J.Newman, and Christopher Moore. "*Finding community structure in very large networks*". Physical Review E, 70, 066111, 2004.
- [4] A.-L. Barabási, R.Albert. "*Emergence of scaling in random networks*". Science, 286, 509-512, 1999.
- [5] Alexei Vázquez. "*Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations*". Physical Review E 67, 056104, 2003.
- [6] 総務省. "ブログ及びSNSの登録者数(平成18年3月末現在)". [http://www.soumu.go.jp/s-news/2006/060413\\_2.html](http://www.soumu.go.jp/s-news/2006/060413_2.html), 2006.
- [7] L.C.Freeman. "*Centrality in social networks*". Conceptual clarification, Social Networks, Vol.1, pp.215-239 (1979).