

コミュニティ構造に基づくノード表現を用いた ネットワーク分断法

吉田 哲也^{1,a)} 山田 佑¹

概要: 本稿では効果的なネットワーク分断を実現するためにネットワークのコミュニティ構造に着目し、コミュニティ間を媒介する境界ノードを同定して除去するネットワーク分断法を提案する。従来のコミュニティ構造に基づく手法では教師情報に対応するコミュニティラベルが必要となるという課題がある。本稿ではネットワークのノードに着目し、コミュニティ構造に基づくノード表現を構築することにより、コミュニティラベルを用いずにコミュニティ間を媒介する境界ノードの同定を実現する。提案法を人工ネットワークおよび実世界のネットワークに適用して評価し、他手法との比較を通じてその有効性を示す。

キーワード: ネットワーク分断, コミュニティ構造, ノード表現, モジュラリティ

Network Immunization via Community Structure based Node Representation

TETSUYA YOSHIDA^{1,a)} YUU YAMADA¹

Abstract: We propose an approach for immunization of networks via modularity based node representation. Since epidemics (e.g. virus) are propagated among groups of nodes (communities) in a network, network immunization has often been conducted by removing nodes with large score (e.g., centrality) so that the major part of the network can be protected from the contamination. Since communities are often interwoven through intermediating nodes, we propose to identify such nodes based on the community structure of a network and remove them for immunization. By regarding the community structure in terms of nodes, we construct a vector representation of each node based on a quality measure of communities for node partitioning. Preliminary experiments are conducted over synthetic and real-world networks, and compared with other centrality based immunization strategies.

Keywords: modularity, node profile, eigenmap

1. はじめに

現在、様々な種類のネットワークが情報交換の基盤として活用されている。コンピュータネットワークや人間同士の交友関係ネットワーク、近年ではソーシャルネットワークサービス上での人間関係なども広く活用されている。ネットワークにおける接続関係を通じて遠隔地にある

情報などの活用が可能である反面、インフルエンザ等の感染症やコンピュータウイルス、悪質なデマなどの蔓延にもつながる恐れがある。

人間をノード、接触関係をリンクとするネットワークにおけるウイルスを通じた感染症の蔓延防止を考えると、ウイルスに感染した人にワクチンを接種してネットワークから除外することが望ましい。しかし、ネットワークのサイズに比べて使用可能なワクチンの数が少ないことが多い。このため、ワクチンを接種する対象(人)を効率的に選択し、できるだけネットワークの大部分が感染しないようにネットワークを分断して守ることが重要となる。

¹ 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan
^{a)} yoshida@meme.hokudai.ac.jp

本稿では効果的なネットワーク分断を実現するためにネットワークのコミュニティ構造に着目し、コミュニティ間を媒介する境界ノードを同定して除去するネットワーク分断法を提案する。ネットワークのリンクに着目する従来の手法と異なり [6], [9], 本稿ではネットワークのノードに着目してコミュニティ構造をとらえる。このためにノード分割に基づくコミュニティ発見における評価指標として広く用いられるモジュラリティ [5] に基づいてノード表現 (ノードベクトル) を構築し、構築したノードベクトルの分布に基づくノードスコアを提案する。スコアの高いノードはコミュニティ間を媒介する境界ノードに対応し、このノードをネットワークから除外してコミュニティ間での伝播が少なくなるようにネットワークを分断する。

提案法を人工ネットワークおよび実世界のネットワークに適用し、ネットワークの中心性に基づく他手法との比較を通じてその有効性を確認した。特に、提案法では教師情報に対応するコミュニティラベルを要求することなくコミュニティ構造を活用したネットワーク分断が可能となる。

2 節で関連研究を紹介し、3 節で提案法の詳細を説明する。4 節で他手法との評価実験を報告し、提案法の有効性を議論する。5 節でまとめと今後の展望を述べる。

2. ネットワーク分断法

2.1 準備

本稿では、行列は太字の大文字、ベクトルは太字のイタリック小文字で表記し、 A_{ij} で行列 A の第 ij 要素を表す。tr は行列のトレースを表し、 A の転置を A^T で表す。要素が全て 1 である n 次元ベクトルを $\mathbf{1}_n$ と表記する。また、ベクトル a から生成される対角行列を $diag(a)$ と表記する*1。

ネットワーク G のノード数を n 、リンク数を m とする。ネットワーク分析では無向で自己ループのない単純グラフを扱うことが多いため [7]、本稿でもこのクラスのネットワークを扱う。 G の接続関係を表現する隣接行列を $A \in \{0, 1\}^{n \times n}$ と表記し、ノード i, j がリンクで接続する場合に $A_{ij} = 1$ であり、接続しない場合は $A_{ij} = 0$ である。 G の隣接行列 A に対して $k = A\mathbf{1}_n$ を次数ベクトルと呼び、要素 k_i は i 番目のノードの次数を表す。

2.2 ノード除去によるネットワーク分断

ウイルスなどによる感染症はネットワークのノード間での接触を通じて蔓延することが多いため、感染したノードとの接触を許すとネットワーク全体にウイルスが蔓延する恐れがある。ネットワークのノードを極力感染から保護するためには感染したノードをネットワークから除外し、最大連結成分 (LCC) などのネットワークの主要部分に対する感染防止を行うことが望ましい (図 1 参照)。

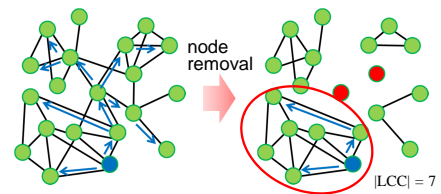


図 1 ノード除去を通じたネットワーク分断

Fig. 1 Network immunization

ネットワーク分断では感染などの伝播に主要な役割を果たすノードを同定して除去することが多い。そのようなノードは何らかの意味でネットワークにおける中心的なノードであると仮定して、従来のアプローチでは中心性の大きなノードをネットワークから逐次的に除去することが多い。

2.3 ノード中心性

これまでに様々なネットワークの中心性が提案されてきた [3], [7]。ネットワーク G の中心性とは G の構造 (隣接関係) から計算されるものであり、値が大きいほど G において中心的な役割を果たすと解釈される。

多くのリンクを持つノードはネットワークのハブに対応すると考えられるため、中心性の代表例としてノードの次数 (リンクの数) に基づく次数中心性がある。他方、情報はノード間でのリンクの連なりであるパスに沿ってネットワーク中で伝播するため、ノード間の最短パスの数に基づく概念として媒介中心性がある。

Google のページランクと同様、ネットワークの隣接行列の固有ベクトルに基づく中心性として固有ベクトル中心性がある。この中心性では、次数中心性のように直接接続するノードのみではなく、ネットワーク上で離れたノードからの影響も考慮した値が計算される。

また、固有ベクトル中心性の摂動法による近似計算法 [8] や、外部から与えられたコミュニティラベルをもとにコミュニティ間の接続関係を表現する隣接行列に対して近似計算法 [8] を適用する手法も提案されている [4]。しかし、正しいコミュニティラベルを求めること自体が困難な問題として知られている [2]。

3. コミュニティ構造に基づくノードスコア

3.1 ノードベクトル

ノード分割に基づくコミュニティ発見ではモジュラリティと呼ばれる指標が標準的に用いられてきた [5]。モジュラリティの高い分割ほどコミュニティ構造を反映した良い分割と考えられるが、モジュラリティの最大化は以下の行列の最大固有値に対応する固有ベクトルの計算に帰着できることが示されている [6]。

$$B = A - P \quad (1)$$

*1 $diag(a)$ の第 i 対角要素は a の第 i 要素である。

ここで $P_{ij} = k_i k_j / 2m$ であり (k は次数ベクトル), 行列 B はモジュラリティ行列と呼ばれる.

式 (1) の行列 B のスペクトル分解を考え, 正の降順固有値に対応する q 本の固有ベクトルを用いた下記の近似分解を考える.

$$B \simeq \Lambda U U^T \quad (2)$$

ここで $U = [u_1, u_2, \dots, u_q]$ であり, 固有ベクトル u_j を固有値の降順に並べたものである. また, Λ は対応する固有値を並べた対角行列である. 式 (2) に基づき, ノードのデータ表現として下記の行列が提案された [6].

$$X = U \Lambda^{1/2} \quad (3)$$

行列 X の各行がネットワーク中の第 i ノードのデータ表現 (ベクトル表現) に対応する. 以下では, ネットワーク中の第 i ノードを $x_i \in \mathbb{R}^q$ (式 (3) の第 i 行に対応) と表現し, この表現をノードベクトルと呼ぶ.

上記のベクトル表現に基づき, コミュニティ中心性と呼ばれる以下の指標が提案された [6] *2.

$$cc(x_i) = x_i^T x_i \quad (4)$$

コミュニティの中心に対応するノードほど式 (4) のコミュニティ中心性は大きな値となる.

3.2 ノードベクトルの分布

3.1 節のノードベクトルには以下の性質が成り立つ.

命題 1. ノードベクトルに対し以下の性質が成り立つ.

$$\sum_{i=1} x_i = \mathbf{0}_q \quad (5)$$

証明は省略する. 命題 1 より, ノードベクトルの中心は \mathbb{R}^q の原点に一致することになる. この性質に基づき, ノードベクトルに対して以下の性質が成り立つと仮定する.

a) ノードベクトルはコミュニティの中心付近に指向性を持って分布する

b) コミュニティ間繋ぐ役割を果たすノードが存在する

a) における指向性とは, 同じコミュニティに属しているノードは q 次元空間において類似した方向を向くことを意味する. b) は, 複数のコミュニティにまたがるノード (境界ノードと呼ぶ) はコミュニティの密集している方向の間を向くことを意味する.

上記の検証のために, 図 2 に示すような 5 つのノードを介してコミュニティが接続する人工ネットワークを生成し, ノードベクトルの性質を調べた. 生成したネットワーク中のノードを大きくコミュニティノードと境界ノードに分け, さらにコミュニティ内のノードを中心, 中間, 周辺に分類した.

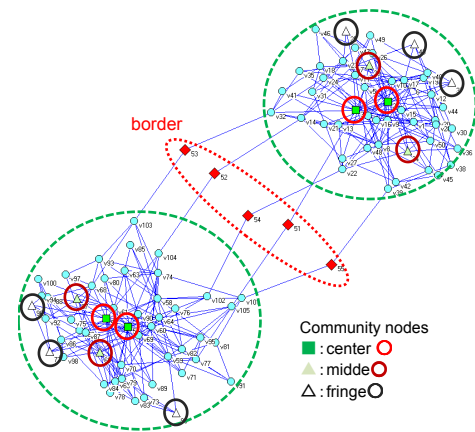


図 2 人工ネットワークにおけるノードの種類
Fig. 2 Node types in a synthetic network.

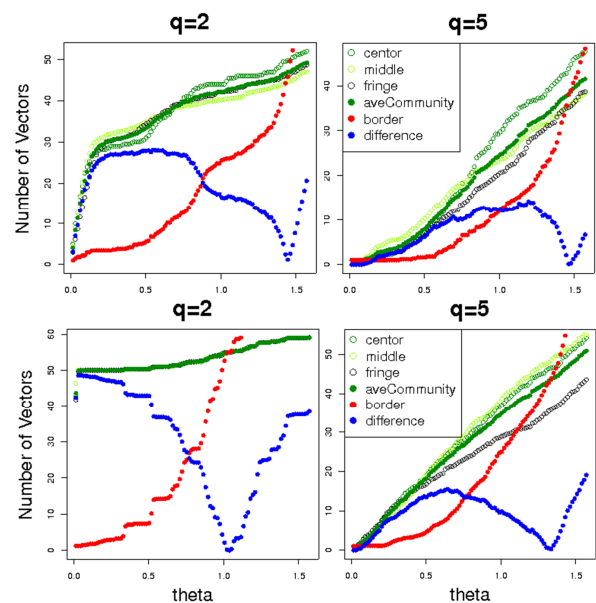


図 3 ノードベクトルの分布 (上段: $c=2$, 下段: $c=3$)

Fig. 3 Distribution of node vectors (upper: $c=2$, lower: $c=3$)

各ノードベクトルに対し, 角度の観点から近傍にあるノードベクトルの分布に対する予備実験を行った. 実験ではコミュニティ数 c は 2, 3 とし, 固有ベクトル数 q を 2 から 10 まで変化させて近傍のベクトル数の分布を調べた.

結果を図 3 に示す. 図で, 横軸は近傍ベクトルまでの角度 (単位はラジアン), 縦軸は角度 θ までの範囲にある近傍ベクトル数である. 凡例で緑は中心ノード, 黄緑は中間ノード, 黒は周辺ノード, 赤は境界ノードである. 濃い緑はコミュニティノードの平均を表し, 青はコミュニティノード (濃い緑) と境界ノード (赤) との差分を表す.

図 3 の結果より, コミュニティノード (緑, 黄緑, 黒線) は同様な傾向を示し, 角度の変化に対して急激に近傍ベクトル数が増えており, 上記 a) の傾向が見られることがわかる. 他方, 境界ノードでは θ が小さい場合, 近傍ベクトル数はコミュニティノードに比べて少なくなっている. 特に, 差分を表す青線はコミュニティノードと境界ノードで

*2 文献 [6] では, x_i の二乗和 ($x_i^T x_i$) と L_2 ノルム ($\|x_i\|$) のどちらもコミュニティ中心性と呼ばれる.

近傍のノードベクトルの分布が大きく異なることを示し、上記 b) の傾向が見られると言える。

3.3 ベクトル逆密度

3.2 節で述べた近傍ベクトル数の分布に基づき、ノードベクトル相互の角度に基づくノードスコアを提案する。コミュニティ間を媒介する役目を果たす境界ノードは近傍ベクトル数が少ないという性質を活用するためにノード対 (i, j) に対するノードベクトルの角度 θ_{ij} を計算し、頂角 θ_{ij} の直角三角形を回転して得られる単位円錐（頂角 θ_{ij} の対辺が円錐の底面に対応）の内部を通るベクトル数を計測する。上記のアイデアを以下のように定式化する。

$$\mathbf{D} = \text{diag}(\|\mathbf{x}_1\|, \dots, \|\mathbf{x}_n\|) \quad (6)$$

$$\mathbf{X}_1 = \mathbf{D}^{-1} \mathbf{X} \quad (7)$$

$$\Theta = \cos^{-1}(\mathbf{X}_1 \mathbf{X}_1^T) \quad (8)$$

$$f(\Theta_{ij}, \theta) = \begin{cases} 1 & (\Theta_{ij} < \theta) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

$$ivd(\mathbf{x}_i) = \frac{1}{\sum_j^n f(\Theta_{ij}, \theta)} \quad (10)$$

ここで式 (6) の \mathbf{D} はベクトル $(\|\mathbf{x}_1\|, \dots, \|\mathbf{x}_n\|)^T$ から生成される対角行列であり、 θ は閾値である。式 (10) の $ivd(\cdot)$ の値がノードスコアに対応する。

式 (3) の行列 \mathbf{X} の各行を \mathbf{D}^{-1} でスケールリングすることにより、式 (7) の行列 \mathbf{X}_1 の各行は L_2 ノルムで 1 に正規化されることになる。このため、正規化した \mathbf{X}_1 を用いて式 (8) のようにノードベクトル間の角度を計算できる。式 (9) の関数 f は指示関数であり、閾値 θ 未満の角度 θ_{ij} の個数を集計する。最後に、3.2 節で述べたように境界ノードでは角度の観点から近傍ノードベクトル数が比較的小ないため、角度 θ の単位円錐の内部を通るベクトル数の逆数をとることでスコアを計算する。式 (10) の $ivd(\cdot)$ は単位円錐の底面あたりのベクトル数の逆数に対応するため、このスコアをベクトル逆密度 (IVD) と呼ぶ。

3.4 コミュニティ中心性を反映したベクトル逆密度

一般に情報伝播を仲介するハブに相当するノードの除去はネットワーク分断に効果的と考えられる。たとえば、2.3 節で述べた次数中心性や固有ベクトル中心性ではハブに対するスコアが大きくなり、従来からネットワーク分断に活用されてきた。しかし、式 (3) で構築するノードベクトルは 3.2 節で述べたようにコミュニティ中心付近に分布するためハブに対応するノードの近傍ベクトル数が大きくなり、3.3 節の IVD ではそのノードスコアが小さくなってしまいう課題がある。

この理由として、式 (3) で構築するノードベクトルの方向性は IVD で活用されているが、その大きさは活用されていないことが挙げられる。ノードベクトルのノルム（ある

表 1 人工ネットワーク

Table 1 Synthetic networks

データ名	ノード数	エッジ数 (平均)
CL_2.5.1	105	708
CL_2	100	355.7
CL_3	150	548.1

表 2 実ネットワーク

Table 2 Real-world networks

データ名	ノード数	エッジ数
karate	34	78
dolphins	62	159
polbooks	105	441

いは 2 乗ノルム) はコミュニティにおける中心性を表すと考えられ、式 (4) のコミュニティ中心性として定義されていた [6]。このコミュニティ中心性を IVD にも反映することにより、以下のノードスコアを定義する。

$$ccivd(\mathbf{x}_i) = cc(\mathbf{x}_i) \times ivd(\mathbf{x}_i) \quad (11)$$

以下では、式 (11) のスコアをコミュニティ中心性を反映したベクトル逆密度 (CCIVD) と呼ぶ。

4. 評価実験

4.1 実験設定

4.1.1 対象データ

提案法を人工ネットワーク (表 1) と実ネットワーク (表 2) に適用して評価した。人工ネットワークはコミュニティ構造を人為的に埋め込んだネットワークであり、まずコミュニティ構造に対応する部分ネットワーク Barabási-Albert (BA) モデル [1] で生成した^{*3}。生成した部分ネットワークから、3.2 節のように 2 つのコミュニティを 5 つのノードを介して接続したネットワーク (CL_2.5.1) と、コミュニティ間をリンクで直接接続したネットワーク (CL_2, CL_3) を使用した^{*4}。人工ネットワークはランダムに生成されるため、表 1) に示す各ネットワークを 10 個生成し、その平均の結果を報告する。また、表 2 のネットワークは GML (graph markup language) で表現され公開されている。

4.1.2 評価尺度

感染症の蔓延防止を考えた場合、分断後のネットワークの最大連結成分 (LCC) の大きさが小さいほど良いネットワーク分断法と見なされる [4]。ノード数 n のネットワーク G に対し、スコア (あるいは中心性) の高いノードをネットワークから逐次的に除去する際、ネットワークに残るノード数の割合 p に対する LCC の大きさの割合 S を評価した。これらは以下で定義される。

$$p = \frac{\# \text{remaining nodes}}{n}, \quad S = \frac{|LCC|}{n} \quad (12)$$

^{*3} BA モデルでの初期次数は 4 とした。

^{*4} CL_2, CL_3 のコミュニティ数はそれぞれ 2, 3 である。

ここで $|LCC|$ は LCC でのノード数を表す。図 1 に示すように、 S が小さいほどネットワーク全体への蔓延を防止できるため良い分断法とみなされる。

4.1.3 比較手法

他手法との比較として、2.3 節で述べた中心性に基づくネットワーク分断法を評価した。各手法では中心性の高いノードを逐次的に選択してネットワークから除去した。

D: 次数中心性

B: 媒介中心性

EVC: 固有ベクトル中心性

CC: コミュニティ中心性

ModC: コミュニティラベルを活用する手法 [4]

提案法におけるパラメータ (q, θ) は予備実験に基づき設定した。

4.1.4 ネットワーク分断手順

ネットワークから逐次的にノードを除去する際、スコア (中心性) の計算には下記の 2 つの戦略を用いた。

単計算: 初期ネットワークに対して一度だけ評価

再計算: ノード除去ごとに、その時点のネットワークに対して評価

再計算は計算時間はかかるが最新のネットワークの状況を反映したスコアを活用できるという利点がある。ModC はスコアの再計算に基づく手法であるため単計算は評価しなかった。なお、ModC で用いるコミュニティラベルは fastgeedy 法 [5] で与えた。

4.2 人工ネットワークに対する結果

人工ネットワークに対する結果 (10 回平均) を図 4 に示す (上段: 単計算, 下段: 再計算)。図では横軸が未除去ノードの割合 p , 縦軸が LCC の割合 S である。灰色 (x) は D, 黒 (+) は B, 黄緑 () は EVC, 青 () は CC, 水色 () は ModC に対応する。提案するノードスコアは (緑は IVD, 赤は CCIVD) で示す。各手法に “R” を付けたものは再計算に対応する。

図 4 に示すように、単計算、再計算ともに提案する IVD は他手法と遜色ない性能を示した。特に、CL_2.5.1 では IVD (緑) で LCC の割合 S の急激な低下が見られた。他方、コミュニティ中心性を活用したにもかかわらず CCIVD は IVD に及ばなかった。また、外部から与えられたコミュニティラベルに基づいてコミュニティ構造を活用する ModC と比較すると、ノードがあまり除去されていない場合 ($p > 0.85$) では提案法は及ばなかったが、 p が小さい場合 (多くのノードが除去された後) には上回った。

4.3 実ネットワークに対する結果

表 2 の実ネットワークに対する結果を図 5 に示す (上段: 単計算, 下段: 再計算)。単計算に対しては、提案する IVD, CCIVD とともに B とほぼ同等な性能を示した。再計

算に対しては、CCIVD は B とほぼ同等な結果となり、コミュニティ中心性を活用する CCIVD は IVD を上回った。これは、コミュニティ中心性の大きなハブノードの除去は実ネットワークでは効果的であるためと考えられる。人工ネットワークの場合と同様に、ModC はネットワークが非連結な部分ネットワークに分断後はあまり LCC の割合 S 低下しなかったが、提案法では分断後も低下し続けた。

5. おわりに

本稿では効果的なネットワーク分断を実現するためにネットワークのコミュニティ構造に着目し、コミュニティ間を媒介する境界ノードを同定して除去するネットワーク分断法を提案した。ネットワークのノードに着目してコミュニティ構造をとらえ、ノード分割に基づくコミュニティ発見における評価指標に基づいてノード表現 (ノードベクトル) を構築し、構築したノードベクトルの分布に基づくノードスコアを提案した。

提案法を人工ネットワークおよび実世界のネットワークに適用し、ネットワークの中心性に基づく他手法との比較を通じてその有効性を確認した。しかし、提案法におけるパラメータ (q, θ) の決定規範に対しては課題が残る。今後はネットワークの構造に基づくパラメータの決定規範を検討するとともに、大規模な実ネットワークへの適用と評価を行う予定である。

謝辞 本研究の一部は文部科学省科研費 (No.24300049)、カシオ科学振興財団、豊田理化学研究所の補助による。

参考文献

- [1] Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol. 286, pp. 509–512 (1999).
- [2] Brandes, U., Delling, D., Gaertler, M., Görke, R., Hofer, M., Nikoloski, Z. and Wagner, D.: On Modularity Clustering, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 20, No. 2, pp. 172–188 (2008).
- [3] Freeman, L. C.: Centrality in Social Networks Conceptual Clarification, *Social Networks*, pp. 215–239 (1979).
- [4] Masuda, N.: Immunization of networks with community structure, *New Journal of Physics*, Vol. 11, p. 123018 (2011).
- [5] Clauset, A., Newman, M. E. J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol. 70, No. 6, p. 066111 (2004).
- [6] Newman, M.: Finding community structure using the eigenvectors of matrices, *Physical Review E*, Vol. 76, No. 3, p. 036104 (2006).
- [7] Newman, M.: *Networks: An Introduction*, Oxford University Press (2010).
- [8] Restrepo, J. G., Ott, E. and Hunt, B. R.: Characterizing the Dynamical Importance of Network Nodes and Links, *Physical Review Letters*, Vol. 97, p. 094102 (2006).
- [9] 吉田哲也: ネットワークのノード情報を考慮した正則化モジュラリティ固有空間法, 情報処理学会論文誌: 数理モデル化と応用, Vol. 5, No. 1, pp. 66–72 (2012).

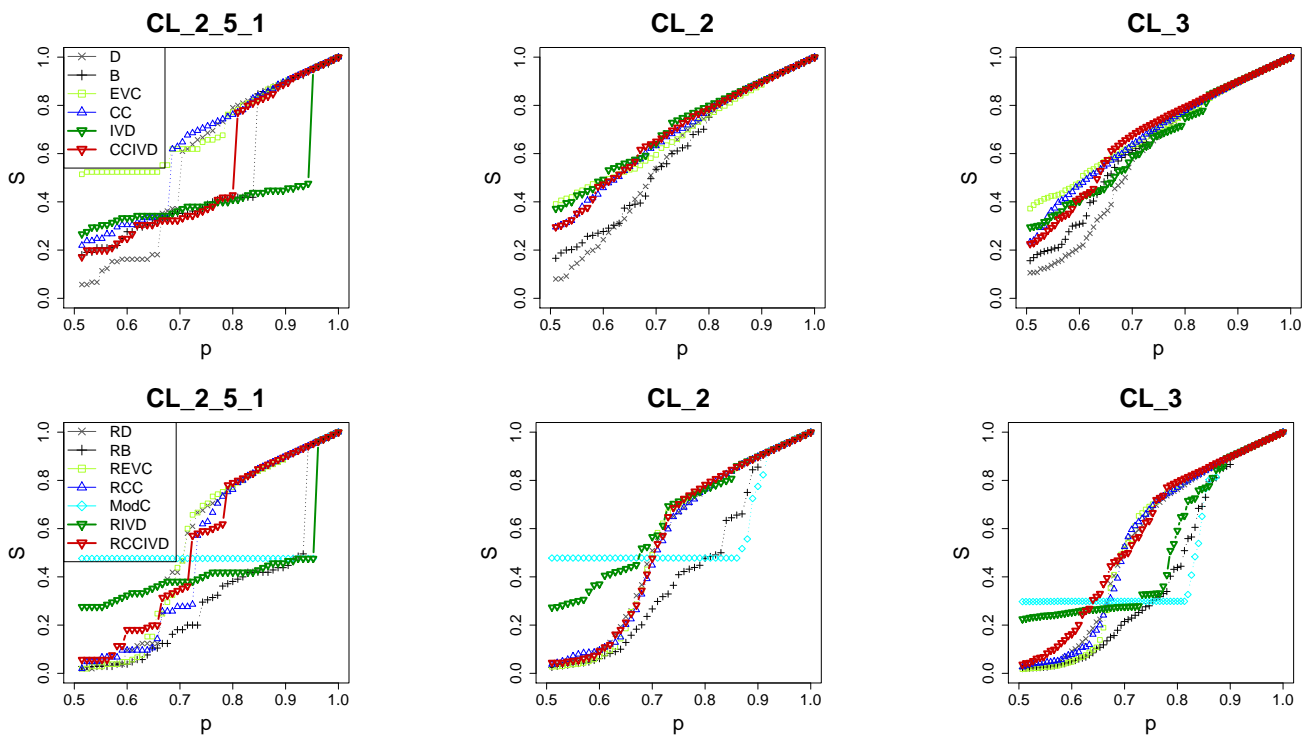


図 4 人工ネットワークに対する結果 (上段: 単計算, 下段: 再計算)
 Fig. 4 Results of synthetic networks (upper: single, lower: recal)

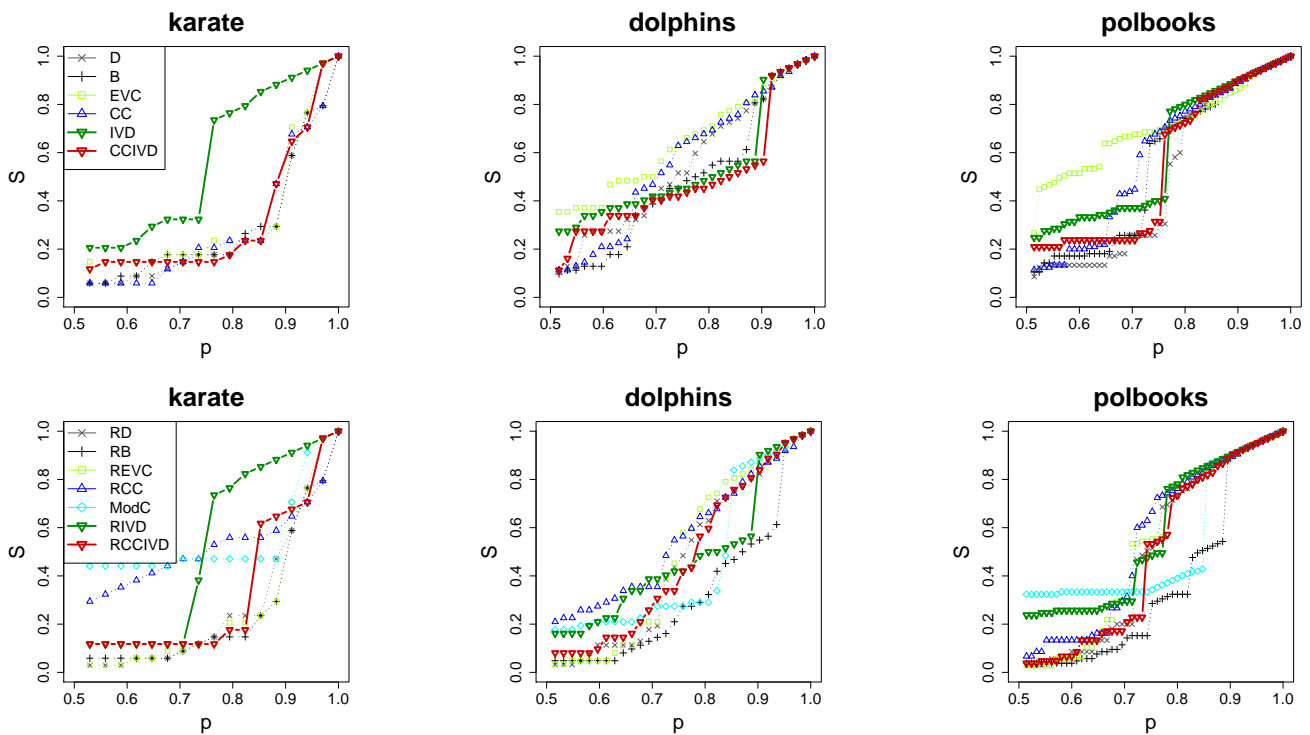


図 5 実ネットワークに対する結果 (上段: 単計算, 下段: 再計算)
 Fig. 5 Results of real-world networks (upper: single, lower: recal)