



Title	コミュニティの対応関係を考慮した二部モジュラリティによるコミュニティ分割
Author(s)	原田, 恵雨; 鈴木, 育男; 山本, 雅人 他
Citation	日本ソフトウェア科学会誌「コンピュータソフトウェア」, 28(1), 127-134
Issue Date	2011-01-25
Doc URL	https://hdl.handle.net/2115/50647
Rights	ここに掲載した著作物の利用に関する注意 : 本著作物の著作権は日本ソフトウェア科学会に帰属します。本著作物は著作権者である日本ソフトウェア科学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」に従うをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Japan Society for Software Science and Technology (JSSST). This material is published on this web site with the agreement of the JSSST. Please be complied with Copyright Law of Japan if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof.
Type	journal article
File Information	CF28_127.pdf



コミュニティの対応関係を考慮した 二部モジュラリティによるコミュニティ分割

原田 恵雨 鈴木 育男 山本 雅人 古川 正志

本論文では、二部グラフのコミュニティ対応関係を考慮したコミュニティ分割を定量的に評価する指標である二部モジュラリティを提案し、すべてのコミュニティ分割のパターンに対する二部モジュラリティの分布を調査することにより、その妥当性を示す。さらに、提案した二部モジュラリティを用いたコミュニティ分割手法を提案し、人工的に生成した二部グラフに対し適用することで、その有効性を示す。

In this paper, we propose a new bipartite modularity which is a measure to evaluate community structure considering correspondence of the community relation quantitatively. The measure has validated by the value distribution of all patterns of community structure on a simple bipartite graph. Furthermore, we proposed the method to extract community structure using proposed measure. We show the availability of the method by applying it to some artificial bipartite graph models.

1 はじめに

現実世界のネットワークには2つの部集合で構成される二部グラフで表現できる関係が数多く存在する。例えば、人間と好きな食べ物の関係、文書と単語の共起関係等である。一般に、二部グラフの解析をする際は簡便的に一部グラフに変換することがある。この変換は一方の部集合に属するノード間のリンクを他方へのリンク関係に基づき生成するものである。二部グラフのコミュニティ分割を行う際も、一部グラフに変換した後に既存のコミュニティ分割手法を適用する方法がとられてきた。しかし、一部グラフに変換することによって二部グラフ特有の情報が失われることが指摘されている[6]。その後、二部グラフを一部グラフに変換することなしに、コミュニティ分割の良

さを評価する指標や、それを用いたコミュニティ分割手法が提案されてきた[1][5][11]。しかし、Barber[1]は、両方の部集合にまたがるコミュニティ分割を評価しているため、一般に性質の異なるノード同士が同じコミュニティに属することになる点が問題である。また、Guimeràら[5]は一方の部集合のコミュニティ分割のみを評価しているため、結果的に一部グラフに変換した後に既存指標で評価することと同じとなるため、二部グラフの特徴を用いているとはいえない。村田ら[11]は他方の部集合のコミュニティ分割も考慮した評価指標を提案しているが、各コミュニティが1つのコミュニティとの対応関係しか考慮していない。

本論文では、二部グラフのコミュニティ対応関係を考慮したコミュニティ分割を定量的に評価する指標である二部モジュラリティを提案し、すべてのコミュニティ分割のパターンに対する二部モジュラリティの分布を調査することにより、その妥当性を示す。さらに、提案した二部モジュラリティを用いたコミュニティ分割手法を提案し、人工的に生成した二部グラフに対し適用することで、その有効性を示す。

また、鈴木ら[9]は二部モジュラリティについてほぼ同じ指標を同時期に提案していたことを挙げておく。

Community Extraction by Bipartite Modularity
Considering Community Relations.

Keiu Harada, Ikuo Suzuki, Masahito Yamamoto,
Masashi Furukawa, 北海道大学大学院情報科学研究科,
Graduate School of Information Science and Tech-
nology, Hokkaido University.

コンピュータソフトウェア, Vol.28, No.1 (2011), pp.127–134.
[研究論文] 2009年12月26日受付.

2 関連研究

2.1 一部グラフにおけるコミュニティ

ネットワーク G はノード数 n のノード集合 V と、リンク数 m のリンク集合 E からなり、 $G = (V, E)$ で表す。またノード $v \in V$ の次数を k_v で表す。隣接行列 A の要素 $A_{i,j}$ は以下の式で表される。

$$A_{i,j} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{それ以外} \end{cases} \quad (1)$$

コミュニティ数 N のコミュニティ集合を $C = \{c_1, c_2, \dots, c_N\}$ とし、ノード v が属するコミュニティ番号を $c(v)$ とするとき、コミュニティ c_i は $c_i = \{v | (\delta(c(v), i) = 1)\}$ と表される。ここで δ はクロネッカのデルタである。ただし、すべてのノードは必ず一つのコミュニティのみに属するものとする。

コミュニティ c_i と c_j の間のリンク密度を e_{ij} とし、以下の式で表す。

$$e_{ij} = \frac{1}{2m} \sum_{v \in c_i} \sum_{w \in c_j} A_{v,w} \quad (2)$$

また、コミュニティ c_i の次数密度 a_i をネットワーク内の全ノードの次数の総和に対する c_i に属するノードの次数の総和の割合とし、以下の式で表す。

$$a_i = \frac{1}{2m} \sum_{v \in c_i} k_v \quad (3)$$

ネットワーク G と G のコミュニティ集合 C が与えられたとき、Newman らはモジュラリティ Q を以下の式で定義した。

$$Q = \sum_{i=1}^N e_{ii} - a_i^2 \quad (4)$$

これは、コミュニティ内リンク密度とコミュニティに属するノードの次数に応じたリンク密度の期待値との差を評価している。Newman らはこのモジュラリティ Q が極大となるコミュニティ集合を得る (コミュニティ分割) 手法を提案している [8]。

2.2 二部グラフにおけるコミュニティ

二部グラフは $G = (V_A, V_B, E)$ と表し、二つの部集合 V_A, V_B とそれらの間のリンク集合から構成される。ただし、各部集合内ではリンクをもたず、

$n = |V_A| + |V_B|$, $m = |E|$ とする。

二部グラフにおけるコミュニティ分割の良さを評価する指標や、それを用いたコミュニティ分割手法はこれまで様々なものが提案されている。

Barber [1] は、Newman のモジュラリティを二部グラフに拡張し、二部モジュラリティ Q^B を以下のように定義した。

$$Q^B = \frac{1}{2m} \sum_{i=1}^n \sum_{j=1}^n (A_{ij} - P_{ij}) \delta(c(v_i), c(v_j)) \quad (5)$$

ここで、 P_{ij} はランダム二部グラフにおけるノード v_i, v_j にリンクが存在する確率であり、以下の式で表される。

$$P_{ij} = \begin{cases} \frac{k_i k_j}{m} & v_i, v_j \text{ が異なる部集合に属する場合} \\ 0 & \text{それ以外} \end{cases} \quad (6)$$

Barber は Q^B を最大化するコミュニティ分割を固有ベクトルを用いた手法によって求めている。

しかし、一般に異なる部集合に属するノードは性質も異なることが多く、それらのノードが同じコミュニティに属するコミュニティ分割は適切とはいえない。

Guimerà ら [5] は同一部集合でコミュニティ集合を定義し、ノード間の共起度 (リンクしているノードの重複度) に基づき二部モジュラリティを定義している。しかし、他方の部集合のコミュニティ集合に全く依存しない問題がある。

村田 [11] はこの問題を解決するため、異なる部集合のコミュニティ間に対応関係をもたせた二部モジュラリティ Q^M を以下のように定義した。

$$Q^M = \sum_i (e_{ij} - a_i a_j), \quad j = \underset{k}{\operatorname{argmax}} (e_{ik}) \quad (7)$$

ここで、 e_{ij}, a_i, a_j は式 (2), (3) で用いているものと同じである。 Q^M はあるコミュニティを評価する際に最も対応するコミュニティとのリンク密度のみ評価の対象としている。つまり、各コミュニティが1つのコミュニティとの対応関係しか考慮していないことになる。しかし、複数のコミュニティとの対応関係を考慮した方がより適切であると考えられる。

3 二部グラフにおけるコミュニティ

3.1 二部モジュラリティの提案

本論文では、各部集合内でコミュニティ集合を定義し、かつ、異なる部集合のすべてのコミュニティとの対応関係を考慮した二部モジュラリティ Q^H を提案する。

Q^H は Newman のモジュラリティと類似した性質を持たせるため、また、複数のコミュニティ対応関係を考慮するため、以下のような性質を持つことが必要条件となる。

1. 一方の部集合の全ノードが 1 つのコミュニティに属する場合、0 となる。
2. 全コミュニティが 1 つのノードだけを含む場合、0 以上の非常に小さい値となる。ただし、この場合が最大の Q^H となることもある。
3. コミュニティ内ノードの共起度が高いほど高い値を示す。
4. 全てのコミュニティペアの対応関係を計算に含める。

各部集合のコミュニティ集合を C_A, C_B とするとき、二部グラフのコミュニティ集合を $C = C_A \cup C_B$ で表す。このとき Q^H を以下のように定義する。

$$Q^H = \sum_{i=1}^{|C_A|} \sum_{j=1}^{|C_B|} \alpha_{ij} (e_{ij} - a_i a_j) \quad (8)$$

従来のモジュラリティではコミュニティ内、もしくは最も対応するコミュニティ間のリンク密度のみを評価していたが、全てのコミュニティ間のリンク密度を評価することによって複数の対応関係を持つコミュニティ集合を評価している。これによって、条件 4 が満たされる。式中の $(e_{ij} - a_i a_j)$ の部分は、モジュラリティと同じく、コミュニティ間のリンク密度とリンク密度の期待値との差を計算している。これが対応関係を持つコミュニティ間では正の値になり、持たないコミュニティでは負の値になる。全ての $(e_{ij} - a_i a_j)$ を単に総和してしまうと、対応関係を持たないコミュニティ間も総和に含まれるため、総和の値は、どのようなコミュニティ集合であっても 0 となり、定数になってしまう。そこで、対応関係の度合いに応じて重み付

けすることでこれを解決する。コミュニティ c_i, c_j 間対応度 α_{ij} を、 c_i, c_j 間リンク密度に対するコミュニティ c_i, c_j それぞれの次数密度の比の平均値として、以下のように定義する。

$$\alpha_{ij} = \frac{1}{2} \left(\frac{e_{ij}}{a_i} + \frac{e_{ij}}{a_j} \right) \quad (9)$$

α_{ij} を $(e_{ij} - a_i a_j)$ に乗算することにより、対応度の大きいコミュニティ間ほど Q^H に貢献することになる。一方、リンクが存在しないコミュニティ間は Q^H に全く貢献しない。

3.2 二部モジュラリティ Q^H の特徴解析

与えられた二部グラフに対してどのようなコミュニティ分割が最も大きな Q^H をもつのかを調査するため、すべてのコミュニティ分割のパターンについて Q^H の値の分布を調べる。その上で、大きな Q^H をもつコミュニティ分割がどのような性質をもつのかを考察する。ここでは簡単のため $|V_A| = |V_B| = 6$ の二部グラフを扱う。 n 個のノードを直和分割するコミュニティ分割のパターン数は Bell Number [2] として知られている。今回の二部グラフではコミュニティ分割のパターン数は各部集合に 203 通りであるため、両方の部集合では $203^2 = 41209$ 通りとなる。

図 1 に対象にした二部グラフを示す。(a) は上の部集合の左右 2 個ずつと下の部集合の中央 2 個のノード間にリンクが密に存在し、次数分布に偏りがある。また、(b) はすべてのノードの次数が同じで構造的にノード間の相違がない二部グラフとなっている。

二部グラフ (a) のコミュニティ分割パターンのすべてについて、コミュニティ数に対する Q^H の分布を図 2 に示す。コミュニティ数が 2, 3 の場合、一方の部集合はノードが必ずすべて 1 つのコミュニティに属していることになるため、 Q^H は 0 となる。また、二部グラフ (a) の最も Q^H が高いコミュニティ分割を図 3(a) に示すが、コミュニティ間のリンクが密となっているため、最も妥当なコミュニティ分割であるといえる。一方、 Q^H が 2 番目に高いコミュニティ分割も、上の部集合の中央のコミュニティが分離されたものであるため、これも妥当なコミュニティ分割の 1 つといえる。また、各ノードがそれぞれコミュニティ

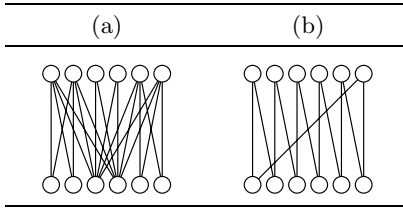
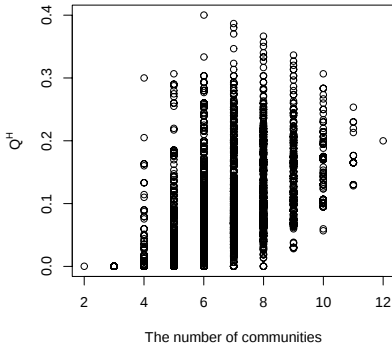
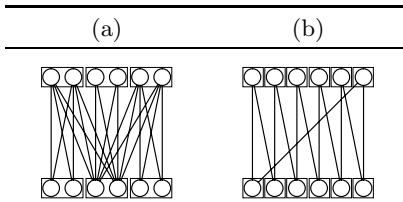
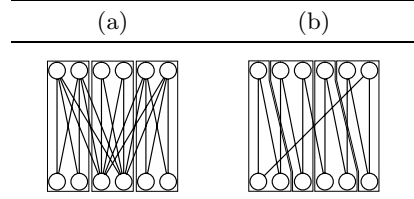
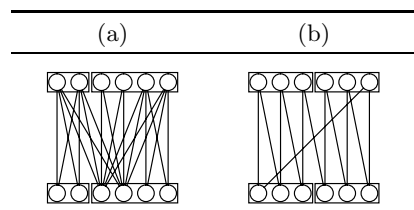


図 1 対象二部グラフ

図 2 二部グラフ (a) のコミュニティ分割パターンのコミュニティ数に対する Q^H 分布図 3 最大 Q^H を持つコミュニティ集合

となるコミュニティ分割の Q^H は最大値と比べて低くなっている．二部グラフ (b) の最も Q^H が高いコミュニティ分割は各ノードがそれぞれコミュニティとなる場合であった (図 3(b))．これは，すべてのノードに構造的な相違がないため，複数のノードが 1 つのコミュニティとなるとリンク密度が均一でなくなり最も Q^H が高いコミュニティ分割とはならないためである．

対象二部グラフにおいて，他の指標 Q^B, Q^M が最も大きくなるコミュニティ分割を図 4, 5 にそれぞれ示す． Q^B は，二部グラフ (a) では Q^H と同様に，それぞれが 2 ノードの 6 つのコミュニティ集合となる場合 Q^B が最大となった．二部グラフ (b) では部集合をまたいでそれぞれが 3 ノードの 4 つのコミュニ

図 4 最大 Q^B を持つコミュニティ集合図 5 最大 Q^M を持つコミュニティ集合

ティ集合となる場合に Q^B が最大となった．ただし， Q^B はそもそもコミュニティ内ノードに部集合の区別をしていないため，必然的にコミュニティ対応関係は一對一となり，複数のコミュニティ対応関係は考慮されていない．一方， Q^M が最大となるのは，二部グラフ (a) では，各部集合が 2 ノードと 4 ノードの 2 つのコミュニティ集合となるときであり，二部グラフ (b) では，各部集合が 3 ノードずつの 2 つのコミュニティとなるときであった． Q^M では，最も大きな対応関係を持つコミュニティのみを考慮しているため，他を無視してでも 1 つのコミュニティ間リンク密度を高めている．このため， Q^M も複数のコミュニティ対応関係を考慮していないとえる．それに対して， Q^H は先に挙げた必要条件を満たし，複数のコミュニティ対応関係を考慮している．これより，提案した二部モジュラリティ Q^H は妥当な指標であることが示された．

4 コミュニティ分割手法

本節では，二部モジュラリティ Q^H が大きな値となるコミュニティ分割を求めるための手法を提案する．しかし，前述の実験で示したように，コミュニティ分割のパターン数は膨大な数があるため，ここではコミュニティ分割を Newman らが開発した CNM 法 [3] を二部モジュラリティ Q^H を扱えるように拡張する．したがって，必ずしも Q^H の最大値をとるコミュニ

ティ分割が得られるわけではない．

4.1 アルゴリズム

提案するコミュニティ分割手法は，各ノードが 1 つのコミュニティである状態から始め，コミュニティ c_i, c_j を結合したときの Q^H の増分 ΔQ_{ij}^H が最大となる c_i, c_j を結合していく手法である．アルゴリズムの途中で ΔQ_{ij}^H が負となり， Q^H が下がることもあるため， Q^H が最大値をとるコミュニティ分割を保存しておく必要がある．

アルゴリズムは以下のとおりである．

1. 各ノードが 1 つのコミュニティであるコミュニティ分割を初期解とし， Q^H を求める．
2. 各部集合内のすべてのコミュニティ間でコミュニティ c_i, c_j を結合したときの Q^H の増分 ΔQ_{ij}^H を計算する．
3. 最大の ΔQ_{ij}^H であるコミュニティ c_i, c_j を結合し， $Q^H \leftarrow Q^H + \Delta Q_{ij}^H$ とする．
4. 1 つのコミュニティに各部集合内のすべてのノードが属するまで手順 2～3 を繰り返す．
5. ここまでの過程で， Q^H が最大であったコミュニティ分割を解とする．

4.2 ΔQ^H の計算

c_i, c_k が結合すると変化する Q^H の増加分である ΔQ_{ik}^H は，以下のように異なる部集合のコミュニティに関係する $\Delta Q_{ik}^{H(1)}$ と，同じ部集合のコミュニティに関係する $\Delta Q_{ik}^{H(2)}$ との和で表される．ただし， C_{opp} は結合したコミュニティとは異なる部集合のコミュニティ集合である．

$$\Delta Q_{ik}^H = \Delta Q_{ik}^{H(1)} + \Delta Q_{ik}^{H(2)} \quad (10)$$

$\Delta Q_{ik}^{H(1)}$ と $\Delta Q_{ik}^{H(2)}$ はそれぞれ以下のように計算される．

$$\Delta Q_{ik}^{H(1)} = \sum_{j=1}^{|C_{opp}|} \frac{e_{ij}(e_{kj} - a_k a_j) + e_{kj}(e_{ij} - a_i a_j)}{2a_j} \quad (11)$$

$$\Delta Q_{ik}^{H(2)} = \frac{-1}{2a_i a_k (a_i + a_k)} \sum_{j=1}^{|C_{opp}|} (e_{ij} a_k - e_{kj} a_i)^2 \quad (12)$$

4.3 ΔQ^H 計算の高速化

基本的にコミュニティが結合されると各コミュニティ間の ΔQ^H を計算しなおす必要がある．しかし，結合したコミュニティに属するノードとの距離が遠いコミュニティ間の ΔQ^H はコミュニティの結合前と比べて変化しないため，再度計算する必要はない．また，計算する必要があるコミュニティ間の ΔQ^H も以下の工夫をすることで高速化できる．

コミュニティ c_l, c_m が結合し， c'_l になったとき，異なる部集合に属する c_i, c_k 間の ΔQ_{ik}^H は式 (11), (12) より，以下のように計算され，更新される．

$$\begin{aligned} \Delta Q_{ik}^H &\leftarrow \Delta Q_{ik}^H \\ &\quad - \frac{(e_{im} a_l - e_{il} a_m)(e_{km} a_l - e_{kl} a_m)}{a_l a_m (a_l + a_m)} \\ &\quad - \frac{(e_{kl} a_i - e_{il} a_k)(e_{km} a_i - e_{im} a_k)}{a_i a_k (a_i + a_k)} \end{aligned} \quad (13)$$

この式はコミュニティ c_l, c_m に属するノードからリンクしているノードが属するコミュニティ間のみ適用する．ただし， c_i, c_k が結合して初めて計算される場合は，式 (10) を用いて ΔQ_{ik}^H を計算する．それ以外のコミュニティ間 ΔQ_{ik}^H は変化しないため計算する必要がない． c'_l と同じ部集合に属するコミュニティ間の ΔQ^H も計算する必要があるが，式の簡単化は困難であるため，式 (10) を用いて再計算する．ただし， c'_l に属するノードから 2 近傍のノードが属するコミュニティ c_x との ΔQ_{lx}^H のみ計算すればよい．

5 コミュニティ分割手法の有効性評価

提案するコミュニティ分割手法の有効性を示すため，本論文ではコミュニティ構造が明らかな二部グラフに対して提案手法を適用し，既存手法との比較を行う．本実験での比較手法は提案手法の基となった CNM 法とする．ただし，CNM 法を二部グラフに直接適用することはできないので，与えられた二部グラフを重み付き一部グラフに変換して，CNM 法を適用する．

5.1 対象とする二部グラフ

一般にコミュニティ集合が明らかになっているケースは稀である．しかし，提案手法の有効性を評価する

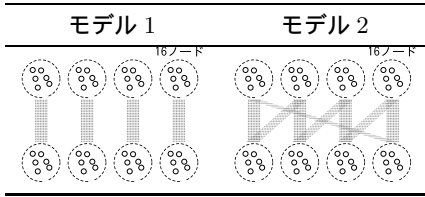


図6 実験で用いるコミュニティ対応関係

ために、コミュニティ構造が明らかな二部グラフを人工的に生成することとした。以下にその生成手順を示す。

1. 各部集合につきコミュニティ数4のコミュニティ集合を用意する。
2. 各コミュニティには16個のノードを割り当てる。
3. 各コミュニティ間に対応関係を定める。対応関係を図6に示す。
4. 各コミュニティに属するすべてのノードについて以下を4回行う。
 - (a) 確率 p_{in} で対応するコミュニティに、確率 $1 - p_{in}$ で対応しないコミュニティに属するノードをランダムに選びリンクを張る。

このモデルは、 $p_{in} = 1$ の場合、コミュニティ対応関係のあるノード間のみリンクが張られるため、コミュニティ構造が明らかになる。 p_{in} が小さくなると徐々にコミュニティ構造が曖昧になる。モデル1の場合は $p_{in} = 0.25$ のとき、モデル2の場合は $p_{in} = 0.5$ のとき、対応関係によらずランダムにリンクが張られることになる。

5.2 評価指標

Newman[7] や Danon[4] が用いている指標として、Fraction of Correctly Identified (FCI) がある。これは正しいコミュニティ分割 $R = \{r_1, r_2, \dots, r_{|R|}\}$ に対する得られたコミュニティ分割 $O = \{o_1, o_2, \dots, o_{|O|}\}$ の精度を測る評価指標であり、 $N_{ij} = |r_i \cap o_j|$ とするとき、以下のように計算する。

$$FCI = \frac{\sum_k |R_k| N_{kl}}{\sum_i |R_i| \sum_j |O_j| N_{ij}}, \quad l = \underset{m}{\operatorname{argmax}}(N_{km}) \quad (14)$$

ただし、 $l = \underset{m}{\operatorname{argmax}}(N_{km})$ に対応する l が複数存在する場合は、 l を未定義とし、 $N_{kl} = 0$ とする。FCI

が大きいほど、 O は正しいコミュニティ集合に近いといえる。すなわち、そのようなコミュニティ集合 O を抽出した手法は高精度であることを意味する。FCI に併せてコミュニティ数による評価も行う。

5.3 実験手順

モデル1, モデル2それぞれについて p_{in} を0から1まで0.1ずつ変化させて各値につき100個ずつの二部グラフを生成する。提案手法とCNM法を適用し、得られたコミュニティ分割を上で述べた評価指標の平均値によって評価する。

5.4 結果

図7左にモデル1に対するFCIの結果を示す。最もコミュニティ構造が明らかである $p_{in} = 1$ の場合、どちらの手法もFCIはほぼ1に近づき、 p_{in} が徐々に小さくなり二部グラフがランダムグラフに近づくにつれて低い評価となった。すべての p_{in} について僅かに提案手法の方が高い評価を得ていた。t検定により、平均値の差の検定を行った。帰無仮説を「2つの手法間に差はない」としたときに、それ以下ならば帰無仮説が棄却される値である P 値が $p_{in} \geq 0.5$ において0.01未満となり、コミュニティ構造が明らかな場合には、二手法間には有意水準1%で十分に有意な差があるといえた(図8左)。コミュニティ数の結果(図7右)では、CNM法がすべての p_{in} について正しいコミュニティ数8であるのに対して、提案手法では二部グラフがランダムグラフになるほど大きくなることが分かる。 p_{in} が小さい値では、ランダムな二部グラフに近い状態になっており、コミュニティの対応関係も明確ではなくノード間の共起関係もほとんどない。この状態ではリンク密度に偏りがなく、コミュニティ構造が曖昧であるにもかかわらず、CNM法はコミュニティ数の少ない分割を求めている。それに対し、提案手法は二部グラフがランダムグラフになるにつれてコミュニティ数が増えている。したがって、提案手法の方がCNM法と比較して有効な分割を求めているといえる。

図9左にモデル2に対するFCIの結果を示す。モデル2は $p_{in} = 1$ の状態でもモデル1と比べてコ

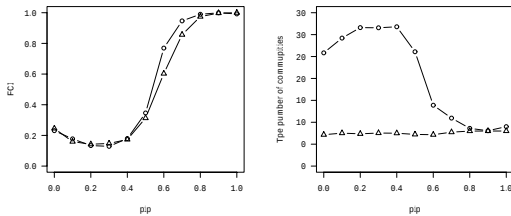


図 7 モデル 1 に対する FCI(左), コミュニティ数(右).
(○ は提案手法, △ は一部グラフ変換+CNM 法)

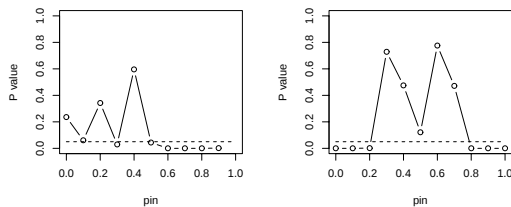


図 8 モデル 1, モデル 2 に対する FCI の t 検定 (P 値)

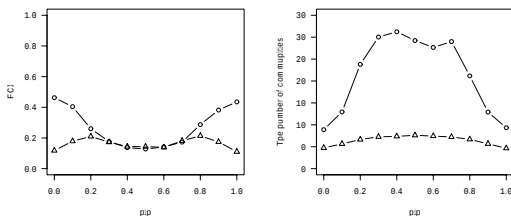


図 9 モデル 2 に対する FCI(左), コミュニティ数(右).
(○ は提案手法, △ は一部グラフ変換+CNM 法)

コミュニティ対応関係が複雑になっており, コミュニティ分割がより困難となる. そのため, CNM 法はかなり低い評価となっているが, 提案手法ではそれと比べて高い評価となっている. t 検定では, P 値が $p_{in} \leq 0.2, p_{in} \geq 0.8$ において 0.01 未満となった. したがって, モデル 2 でもまた, コミュニティ構造が明らかな場合に, 二手法間に有意水準 1% で有意な差があるといえた (図 8 右). また, コミュニティ数の結果 (図 9 右) では, CNM 法がすべての p_{in} について正しいコミュニティ数 8 の半分となるが, 提案手法では $p_{in} = 1$ については正しいコミュニティ数が得られ, $p_{in} = 0.5$ に近づくにつれて, モデル 1 と同様にコミュニティ数が増える. したがって, モデル 2 でも, 提案手法の方が CNM 法と比較して有効な分割

を求めているといえる.

以上の結果から, 提案手法の有効性が示された.

また, 提案手法が, 与えられた二部グラフに対する Q^H の最大値に対して, どれくらいの値の Q^H を求めることができるかを検証する実験を行った. 各部集合のノード数が 6 である二部グラフをリンク数が 10 本から 26 本まで 2 本間隔でそれぞれ 100 個 (計 900 個) 生成し, その内 42.7% で, 最適解を求めることができた. また, 最適解に対して, 平均 0.996 倍の Q^H となるコミュニティ分割を求めることができた.

5.5 計算量の評価

提案手法の大規模データへの適用可能性を評価するために, ノード数, リンク数の異なる二部グラフに対して適用したときの計算時間を計測した. また, 高速化の有効性を検証するために, 高速化を施していない提案手法も併せて計測を行った. 計測に用いた二部グラフは, CNN モデル[10]を二部グラフ用に拡張したものである. 生成アルゴリズムを以下に示す.

1. ノード数 1, リンク数 0 の二部グラフ G を用意する.
2. 終了条件を満たす場合, アルゴリズムを終了する.
3. 確率 $1/2$ で手順 7 へ進む.
4. 二部グラフにノード v_{new} を追加する. 追加される部集合は, それぞれ $1/2$ の確率でランダムに選択する.
5. 追加された部集合とは異なる部集合の中から, ランダムに 1 つノード v_{sel} を選び, v_{new} とリンクを生成する.
6. v_{new} から, v_{sel} の二次近傍ノード集合へポテンシャルリンクを生成し, 2 へ戻る.
7. G 上に存在するポテンシャルリンクの中から, ランダムに 1 つ選んでリンクとし, 2 へ戻る.

このモデルでは, ノード数の増加に対するリンク数の増加が一定であり, 現実の二部グラフと近い特徴をもつ.

結果を図 10 にノード数と計算時間の両対数グラフとして示す. グラフが直線となり, その傾きから, 計算時間がノード数の約 1.5 乗に比例することがわか

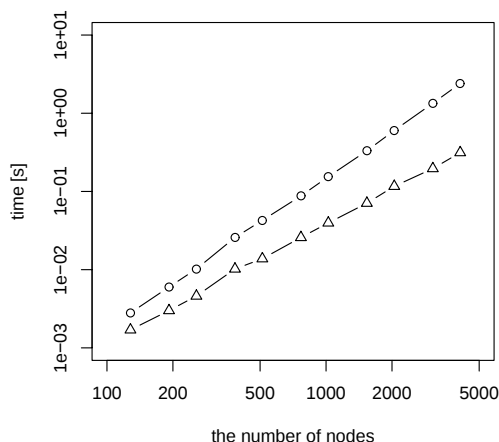


図 10 ノード数の変化に応じた計算時間の変化
(△:提案手法, ○:高速化を施していない提案
手法)(計算機スペック: Intel Xeon @ 2.27G)

る。したがって、2000 万ノード程度の二部グラフでは、約 3 時間で計算が終了すると予想され、大規模データに対する有効性があると考えられる。しかし、高速化を施さない場合の計算時間はノード数の約 2 乗に比例し、2000 万ノード程度の二部グラフでは、約 700 日の計算時間が必要であると予想され、大規模データに対する有効性はないと考えられる。これより、高速化による有効性が示された。

6 おわりに

コミュニティ対応関係を考慮したコミュニティ分割の良さを評価する二部モジュラリティ Q^H を提案した。そして、2 種類の二部グラフに対して、すべてのコミュニティ分割のパターンに対する Q^H を求め、その特徴を調べることにより、 Q^H は二部グラフのコミュニティ分割を評価する妥当な指標であることを示した。

また、一部グラフのコミュニティ分割において有効である CNM 法を基に、 Q^H が極大となるコミュニティ分割を得る手法も提案した。コミュニティ構造の明確さを制御するパラメータを導入し、そのパラメータを変化させることにより様々な二部グラフを多数

生成し、既存手法と比較して提案手法の有効性を示した。

今後の課題として、 Q^H を用いたコミュニティ分割手法を現実の二部グラフへ適用することが挙げられる。

参考文献

- [1] Barber, M. J.: Modularity and community detection in bipartite networks, *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Vol. 76, No. 6(2007), 066102.
- [2] Bell, E. T.: Exponential Numbers, *The American Mathematical Monthly*, Vol. 41, No. 7(1934), pp. 411–419.
- [3] Clauset, A., Newman, M. E. J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol. 70, No. 6(2004), 66111.
- [4] Danon, L., Díaz-Guilera, A., Duch, J. and Arenas, A.: Comparing community structure identification, *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 9(2005), P90080.
- [5] Guimerà, R., Sales-Pardo, M., and Amaral, L. A. N.: Module identification in bipartite and directed networks, *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Vol. 76, No. 3(2007), 036102.
- [6] Latapy, M., Magnien, C. and Vecchio, N. D.: Basic notions for the analysis of large two-mode networks, *Social Networks*, Vol. 30, No. 1(2008), pp. 31–48.
- [7] Newman, M. E. J.: Fast algorithm for detecting community structure in networks, *Phys. Rev. E*, Vol. 69(2004), 066133.
- [8] Newman, M. E. J. and Girvan, M.: Finding and evaluating community structure in networks, *Phys. Rev. E*, Vol. 69(2004), 026113.
- [9] Suzuki, K. and Wakita, K.: Extracting Multifacet Community Structure from Bipartite Networks, *Computational Science and Engineering, IEEE International Conference on*, Vol. 4(2009), pp. 312–319.
- [10] Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Phys. Rev. E*, Vol. 67, No. 5(2003), 056104.
- [11] 村田剛志: 2 部ネットワークにおけるモジュラリティ, *JWEIN08*, 2008, pp. 30–35.