グラフ分割の検出限界

川本達郎 〈産業技術総合研究所人工知能研究センター　kawamoto.tatsuro@aist.go.jp〉

グラフ分割，と言ってもあまり馴染みがないかもしれない．グラフ分割は，基本的には計算機科学・統計科学の対象であり，自然科学とは毛色が異なる面も存在するが，統計力学的なアプローチでの研究が近年も活発に進められている．

グラフ分割は，頂点と枝で構成されるグラフ（ネットワーク）から，いくつかの部分グラフに分割し，マクロなグループ構造を抽出する問題である．グラフ分割には，純粋に最適化問題としての定式化と，推論問題としての定式化の2種類が存在する．純粋な最適化問題とは，例えばグラフ上のコスト（もしくはエネルギー）最小化問題で，与えられた制約のもと，通信コストや消費電力等のコストが最小になるようにグループ分けする問題である．推論問題としてのグラフ分割とは，例えば人間関係・出版物の引用関係・遺伝子間関係・画像や地図上の要素間関係等のデータから，類似した特徴を持つグループを抽出するという問題である．

前者はコストの定義が明確で，とにかく最もコストが小さくなる解を見つけることが至上命題である．一方後者は，何をコストや制約とすべきかは明確ではなく，そもそも絶対的な正解というものが存在しない．自らコスト関数をデザインして評価する必要があり，そのためにはグラフデータがどのように生成されたものなのかという統計性も考慮する必要がある．前者に比べて後者はやたらと曖昧な問題であるが，そうは言っても社会データや画像データから特徴抽出をしたいというニーズは確固として存在する．

さらにグラフ分割は，通常，計算困難な問題になっており，最適解を求めることは技術的に困難である．そのため，使用しているアルゴリズムによってどの程度最適化がちゃんとできているのかの評価も難しい．

問題設定の曖昧性と計算困難性という，推論問題としてのグラフ分割が抱える二重の困難は，混沌とした研究の流れを生んだ．自然科学の問題と異なり，自らデザインしたコスト関数が実験結果を説明するという要請もないため，2000年代にはコミュニティ検出というテーマで膨大な数の手法提案論文が出版された．

しかし，手法だけ提案し，評価はいい加減にやっておけば良いというのでは科学として成り立たない．得られた分割結果はどの程度“正しい”のだろうか．このような推論問題に対しての真面目な取り扱いは，統計的有意性をちゃんと評価するということである．

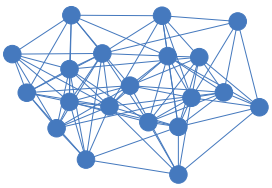
正解としてのグループ構造を持つ人工的なグラフデータを，ランダムグラフモデルによって生成してみたとき，特定のアルゴリズムが出力する分割結果がどれだけ正解を当てられるかを考えよう．正解としてのグループ構造を一様な構造に近づけていくと次第にアルゴリズムの正解率が下がっていき，あるところで，完全ランダムに分割した場合と同程度の正解率しか得られなくなってしまう．すなわち，統計的に有意な解が得られなくなる．この限界を（アルゴリズミックな）検出限界と呼び，グラフサイズが無限大の極限で，相転移として捉えることができる．

興味深いことに，グラフが十分にスパースなときは，どれだけはっきりとしたグループ構造のグラフを生成したとしても検出限界を超えてしまうことがある．この場合，アルゴリズムは完全にその機能を失ってしまうため，「どうせ正解がないのだから」という言い訳が通用しなくなり，出力結果は“正しい”とは言えなくなる．このように，理論的な後ろ盾（やそれがないこと）を増やしていくことで，より精密な議論が可能になる．

用語解説

グラフ分割：
グラフを，指定された数の部分集合（部分グラフ）に分割（分類）すること．

グラフ（ネットワーク）：
離散的なオブジェクト（人や物・文章・事象など）の（離散的な）関係性を，頂点集合とその間を結ぶ枝で表現したもの．



コミュニティ検出：
グラフ分割が与えられた数にグラフを分割するのに対し，分割数の推定までを含めた問題をコミュニティ検出と呼ぶ．また，密につながった頂点集合に分割することも定義に含まれることが多い．

統計的有意性：
得られた結果が，ランダムなイベントの結果生じた偶然ではないと判断できる場合，「統計的に有意である」と言う．

1. はじめに

グラフ分割とは，「頂点集合とそれらを結ぶ枝の集合で与えられたグラフ（ネットワーク）を，マクロなグループ構造が抽出されるように，いくつかの部分グラフに分割する問題」である．ある種の最小カット問題という，純粋な最適化問題としての定義もあるが，ここではむしろ上記のような比較的広い意味での推論問題として扱う．また，グラフ分割と似たような用語として，グラフクラスタリング，コミュニティ検出，node classification など，微妙に想定している問題の範囲や設定が異なるものが多数存在する．ここでは，グラフと，それをいくつに分けるかという分割数が入力として与えられたもとで，マクロなグループ構造として密につながった頂点集合を抽出する問題を考える．

図1を見ていただきたい．このグラフにはグループ構造があると言えるだろうか．見た目では左右に頂点の“塊”があり，それぞれの“塊”の内部では密に枝がつながっているように見える．しかしそれは，グループ構造に見えるように単に頂点の配置を工夫して描画しただけかもしれない．何をもってこのグラフにグループ構造があるかないかを判断したらよいのだろうか．また，アルゴリズムによって抽出したグループ構造が“正しい”かどうかを，どう評価したら良いのだろうか．

グラフ分割は，教師なし学習と呼ばれる問題の一種であり，そもそも明示的な正解というものが存在しない．「どうせ正解がないのだから“正しい”かどうかという議論は存在しない」と思われるかもしれない．しかし，もし図1で，左右の頂点の“塊”の間には数本しか（もしくはまったく）枝が張られていなかった場合，「このグラフにはグループ構造があるかどうかは判定できない」と言うのにも無理があるのではないだろうか．教師なし学習には絶対的な正解がないというのは確かだが，まさにそういうものを評価するために統計科学というものが存在する．ここでは，1つのグラフだけに注目して評価するのではなく，背後にあるグラフのアンサンブルを考えたもとで，実現しているグラフや推論結果が統計的に有意かどうかを判定しよう，という立場をとる．

グラフ分割についての統計的有意性の研究は，昔から統計学や理論計算機科学の分野で脈々と進められてきたが，統計力学的な研究も活発に進められている．本稿ではその一部として，最尤法に基づく簡単なアルゴリズムの統計的有意性を，スピングラスの枠組みで評価できることを紹介する．まず，想定するグラフアンサンブルをランダムグラフモデルを用いて導入し（2節），そのモデルのもとでの最尤法とその貪欲アルゴリズムを紹介する（3節）．4節では，ランダムグラフモデルのパラメーター空間の中で，貪欲アルゴリズムの出力が完全にランダムになってしまう限界（検出限界）を導出する方法の概略を述べる．この検出限界を超えた領域にあるグラフアンサンブルにおいては，貪欲アルゴリズムによって統計的に有意なグループは抽出できない，もしくは原理的に存在しない，ということが分かる．これは，統計的仮説検定のような，いわゆるp値を出すスタイルの評価ではないが，グラフ分割について，より精密な議論を可能にしてくれる．

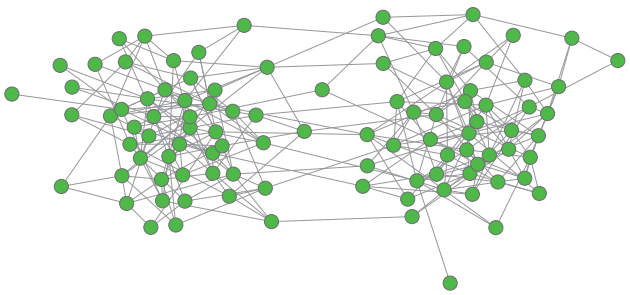


図1　Stochastic block model から生成されたグラフインスタンス．

2. ランダムグラフモデル

最も基本的なグラフ $G=(V,E)$ は，頂点集合 $V(|V|=N)$ と枝の集合 $E(|E|=M)$ の組で与えられ，隣接行列 A という $N\times N$ 行列で表現される．隣接行列の各要素は頂点対に対応しており，頂点 i と j が枝で繋がっていれば $A_{ij}=1$，繋がっていなければ $A_{ij}=0$ という成分を持つ．枝に向きがなければ隣接行列は対称行列になっている．隣接行列の対角成分は自己ループを示すが，ここでは自己ループはないと仮定する．頂点 i に付いている枝の数 c_i を次数と呼ぶ（$c_i=\sum_{j=1}^{N}A_{ij}$）．

隣接行列がランダム行列の場合，対応するアンサンブルは，ランダムグラフモデルとなる．例えば，対称行列の制約のもと行列要素が確率 q で1，確率 $1-q$ で0を持つランダム行列に対応するものは，Erdős–Rényi ランダムグラフモデルと呼ばれている．各頂点の次数 c_i の期待値が定数に抑えられるようなスパースグラフの場合には，結合確率を $q=O(1/N)$ とスケーリングする．モデルから実際に生成されるグラフを（グラフ）インスタンスと呼ぶ．

Erdős–Rényi モデルは，どの頂点にも個性がない一様モデルだが，グラフ分割の性能を議論する場合には，グループ構造を持つランダムグラフモデルを考えたい．そのような正準モデルとして，stochastic block model（SBM）がよく扱われる．図1は，実はこのモデルから生成されたインスタンスである．このモデルでは，各頂点にはグループラベルが与えられ，同じグループに属する頂点は，統計的に等価な結合パターンを持つ．本稿では簡単のため，グループ数は2の場合のみを考える．頂点 i が所属するグループは，イジング変数を用いて $s_i\in\{-1,+1\}$ と表現し，全頂点のグループ割り当てを $|s\rangle$ と表す．ここでは $|s\rangle$ を，埋め込まれた正解と呼ぶ．最も素朴な SBM では，同じグループの頂点対は確率 ρ_{in}，異なるグループに属する頂点対は確率 $\rho_{\mathrm{out}}(<\rho_{\mathrm{in}})$ で枝が張られるようにし，同じグループの頂点同士がより密に繋がったグラフが典型的に生成されるようにする（$\rho_{\mathrm{in}}=\rho_{\mathrm{out}}$ とすることで，SBM は Erdős–Rényi モ

デルに帰着する）．このモデルが生成するグラフの隣接行列Aの確率分布$P(A)$は，

$$P(A|\rho_{\rm in},\rho_{\rm out},|s\rangle)=\prod_{i<j(s_i=s_j)}\rho_{\rm in}^{A_{ij}}(1-\rho_{\rm in})^{1-A_{ij}} \times\prod_{i<j(s_i\neq s_j)}\rho_{\rm out}^{A_{ij}}(1-\rho_{\rm out})^{1-A_{ij}} \quad (1)$$

と書ける．本稿ではスパースグラフを扱うため，$\rho_{\rm in}$や$\rho_{\rm out}$は$O(1/N)$である場合を考える．

この素朴なモデルは，尤もらしいランダムグラフモデルに見えるかもしれない．しかし，このモデルは，各頂点の次数のアンサンブル平均がどれも等しく，また頂点数Nが大きい極限では次数分布がポアソン分布に従うことに注意されたい．一方，現実のグラフデータでポアソン分布のような指数減衰型の次数分布を持っているものは稀であることが知られている．従って，式(1)のままでは，あまり現実的なモデルとは言えない．

この問題を克服するため，次数補正（degree-corrected）SBM[1]というものが考案された．詳細は割愛するが，次数補正SBMから生成されるグラフインスタンスのアンサンブル平均をとると，各頂点の次数が指定した値になっているように拘束条件が付けられている．次数補正SBMは，この次数制約のもとで，所望のグループ構造を典型的に実現させるランダムグラフモデルとなっている．このモデルは，上記の素朴なSBMを特殊な場合として含んでいる．

3. グラフ分割の方法

グラフ分割の問題は，入力として与えられた隣接行列Aに対し，頂点集合の尤もらしいグループラベル$|\hat{s}\rangle$を見つける問題である．これを実現する方法の一つとして，前節で述べたSBMを仮定した最尤推定（尤度関数最大化）が考えられる．式(1)を尤度関数として見たとき，グループ構造を表すパラメーターとして$\rho_{\rm in}$と$\rho_{\rm out}$，そして推定すべきグループラベルとしてのイジングベクトル$|\hat{s}\rangle$を引数に持つ関数となっている．この最尤推定は，離散変数を含む非凸最適化問題となっているため，容易に解くことはできないことに注意されたい．

ここで，思い切って$\rho_{\rm in}$と$\rho_{\rm out}$を決め打ちした値に固定すると，SBMの最尤法は，ありとあらゆるイジング変数の組み合わせの中から尤度関数を最大化するだけの問題となる．実はその尤度関数は，モジュラリティ関数と呼ばれる，実データ解析で非常によく使われてきた評価関数と等価になっている．[2] モジュラリティ関数にもいくつか種類があるが，通常使われる関数は，前節の次数補正SBMの尤度関数になっている．[*1]

少し脱線するが，モジュラリティ関数最大化問題を考える際は，同時にグループの数も推定する場合が多い．[*2] しかし，このグループ数推定方法は，ベイズ推論のそれとは機構が異なる．これは，モジュラリティ関数最大化が本質的には最尤法であることからもお分かりいただけるだろう．考慮しているモデル空間が厳しく制限されているために，グループ数推定が実現できているのだが，ここでは抽出したいグループ数は入力として与えられているとするため，これ以上は立ち入らない．

グラフ分割問題を解くアルゴリズムとして，近年ではグループラベルを隠れ変数として扱うベイズ推論の方法が特に活発に研究されている．しかし，実データ解析においては，モジュラリティ関数最大化による方法の人気は根強く，特に貪欲アルゴリズムによる最大化法は非常によく用いられてきた．貪欲アルゴリズムとは，（2分割の場合）イジング変数を1つもしくは少数選び，モジュラリティ関数の値が上がるのであれば変数をフリップする（状態が更新されなくなるまで繰り返す）という，局所更新アルゴリズムの総称である．これは単純で分かりやすく，高速なアルゴリズムである．特に有名なものとして，Louvain法[3]がある．

ここで冒頭の問いに戻る．貪欲アルゴリズムは，一体どれくらい“正しく”機能するのだろうか．より高級な手法と比較して，実質的に性能として劣るのだろうか．

4. 貪欲アルゴリズムの検出限界

モジュラリティ関数をエネルギー関数と見なすと，現在の問題は，イジング変数の状態空間での（上に凸な）エネルギーランドスケープのなかで，最大エネルギー状態を探索するということになる．図2は，2^{N-1}個の状態の一部をcurvilinear component analysisという多様体学習の方法を用いて2次元座標上にマップし，エネルギーランドスケープを可視化したものである．もしグラフがはっきりとしたグループ構造を持っていれば，図2の上図のように，単峰に近いランドスケープになっていると想像される．一方，一様に近いグラフの場合は，図2の下図のように，非常に多峰的になっていると思われる．エネルギーランドスケープの多峰性が非常に強いとき，貪欲アルゴリズムでは最大エネルギー状態を見つけるのが現実的に不可能となり，アルゴリズムとして機能しなくなるはずである．

この振る舞いをより正確に表現するために，準安定状態というものを考える．準安定状態は，「貪欲アルゴリズムのどのような1ステップ更新によっても，エネルギーを上げられないイジング状態」と定義する．ある状態が準安定状態かどうかは，使用するアルゴリズムの詳細に依存するが，ここでは最も簡単な，頂点のイジング変数をランダムに1つずつ更新していくアルゴリズムを考える．非常に多

[*1] 歴史的には，モジュラリティ関数はグループ構造の強さを示す指標として導入され，実データと帰無モデル（通常Chung–Luモデルが採用される）との隣接行列の差を測ったものとして定義された．

[*2] モジュラリティ関数最大化は，コミュニティ検出の方法としてよく知られている．コミュニティ検出という用語は，グループ数の推定を含めた推論問題を指す．また，通常は密につながった頂点集合を抽出することが想定されている．

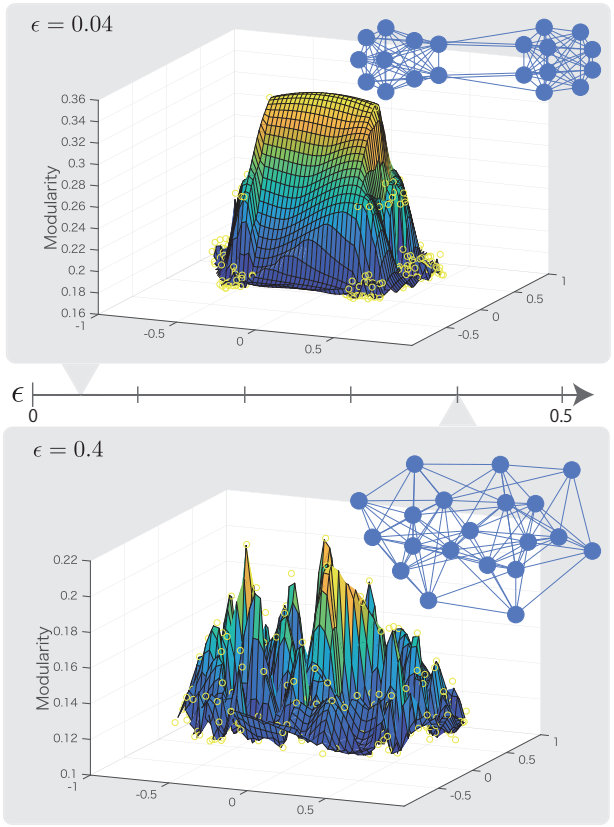


図2　モジュラリティ関数のエネルギーランドスケープの例．グラフインスタンスは，素朴なSBMから生成されている．埋め込まれた正解として，10個の頂点をグループ1に，10個の頂点をグループ2に割り当て，$\epsilon \equiv \rho_{\mathrm{out}}/\rho_{\mathrm{in}}$が小さい場合と大きい場合に，平均次数がおよそ9になるように生成したインスタンスが，上下の図の右上に載せたグラフである．これらのグラフに対するモジュラリティ関数の値を，イジング変数の状態空間の一部について3次元プロットした．実装には，文献4の著者らによって公開されているコードを使用した．

峰的になっている状態は，「準安定状態の個数が（頂点数Nについて指数関数的に）膨大にある」と表現することができる．[*3]

モジュラリティ関数最大化問題は，SBMを仮定した最尤法だったので，以下では，素朴なSBMから生成したグラフを貪欲アルゴリズムで解いたときに，どれくらい準安定状態が発生するかの評価を簡単に紹介する．[5] ここでは，SBMのパラメーター空間で，準安定状態が指数的にたくさん存在する領域を検出不可能領域（スピングラス相）と呼び，その境界を検出限界（detectability limit）と呼ぶ．

4.1　Tanaka–Edwardsの方法

ランダム行列で与えられるエネルギー関数の準安定状態を数える方法の一つとして，Tanaka–Edwardsの方法[6]がある．この方法をSBMでのモジュラリティ関数最大化問題に用いる．

グラフを2分割する場合のモジュラリティ関数は，以下のように表現される．

$$Q(|\hat{s}\rangle) = \langle \hat{s}|B|\hat{s}\rangle = \mathrm{const.} + \sum_{i,j(i\neq j)} \hat{s}_i B_{ij} \hat{s}_j . \tag{2}$$

ここで，Bは次のように定義される$N\times N$行列である．

$$B \equiv A - \frac{1}{2M}|c\rangle\langle c|. \tag{3}$$

$|c\rangle$は，各頂点の次数を並べた次数ベクトルである．一つの頂点iのグループラベルを更新したときの$Q(|\hat{s}\rangle)$の変化は，

$$(-\hat{s}_i)\sum_{j(\neq i)} B_{ij}\hat{s}_j - \hat{s}_i \sum_{j(\neq i)} B_{ij}\hat{s}_j = -2\hat{s}_i \sum_{j(\neq i)} B_{ij}\hat{s}_j . \tag{4}$$

となる．すなわち，更新によってエネルギーが上がらないという条件は，

$$\lambda_i = s_i \sum_{j(\neq i)} B_{ij} s_j \tag{5}$$

を満たす非負のλ_iが存在すると表現することができる．準安定状態は，この条件がすべての頂点について成立している場合ということになる．従って，準安定状態の数$\mathcal{N}_m$は，以下の積分で表現できる．

$$\mathcal{N}_m = \sum_{\{\hat{s}_i\}} \prod_{i=1}^{N} \int_0^\infty \mathrm{d}\lambda_i \delta\left(\lambda_i - \hat{s}_i \sum_{j(\neq i)} B_{ij}\hat{s}_j\right), \tag{6}$$

ここで，$\delta(\cdot)$はDiracのデルタ関数である．

式(6)の$\mathcal{N}_m$は，一つのグラフインスタンスについての準安定状態の個数だが，我々はこれをSBMで生成されるグラフ（式(1)に従うランダム隣接行列）についてアンサンブル平均（ここでは$[\cdots]_A$と表現する）をとった量$[\mathcal{N}_m]_A$を評価したい．具体的には，

$$[\mathcal{N}_m]_A \sim \mathrm{e}^{Nf} \tag{7}$$

という漸近形を仮定したもとで，fが正の値をとる領域では検出不可能領域になっていると評価する．

計算の詳細は割愛し，ポイントを述べるに留めよう．元々のTanaka–Edwardsの論文ではGaussianランダム行列を考えていたため，式(6)のような積分が比較的簡単に実行できる．一方今回の問題では，隣接行列Aとして2値のスパースランダム行列を考えているため，そのまま計算すると，積分が発散してしまう．しかしこの発散は本質的なものではなく，式(5)として与えられた，本質的には離散の条件を，式(6)ではデルタ関数という特異性の強い関数で表現してしまったことに由来する．逆に言えば，発散をもたらしているデルタ関数の高周波成分は無視してしまっても問題ないはずなので，そのような近似のもと上記の積分を実行する．

計算すると，SBMのパラメーター空間で，図3のような相図を描くことができる．cとϵ（$\equiv\rho_{\mathrm{out}}/\rho_{\mathrm{in}}$）は，それぞれ平均次数とグループ構造の強さの逆数を示すパラメーターである．図中の濃淡は，数値実験によって得られた，埋め込まれた正解と比較した正解率（$\mathrm{overlap}\equiv 1/2 + \langle s|\hat{s}\rangle/2N$）を表している．ここでは，$s_i = 1$である頂点の数と$s_i = -1$である頂点の数が等しい場合を考えているため，$|\hat{s}\rangle$が完全にランダムな場合，正解率は1/2となる．黄色の実線が，

[*3] もちろん図2自体は，どのようにスピン変数を2次元座標に落とし込むかの詳細に依存するため，図2の凸凹が正確に準安定状態かどうかを示しているわけではない．

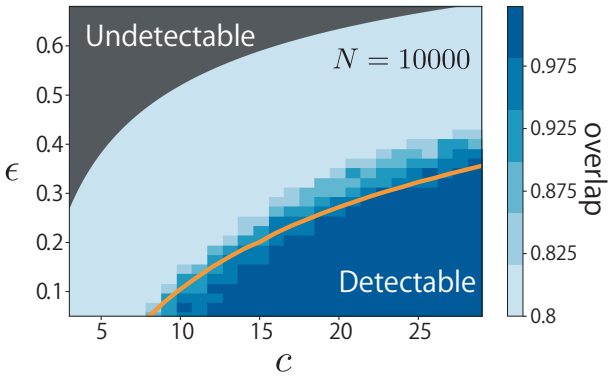


図3　SBMのパラメーター空間での検出限界の相図．色の濃淡は数値実験によって得られた正解率（overlap）を表す．正解率が0.8を下回るものは，0.8の場合と同じ色を割り当てている．

式(7)において$f=0$の条件から得られる相境界である．この境界より左上の領域が，貪欲アルゴリズムの検出不可能領域である．[*4] 境界より右下は検出可能領域であり，多峰性が強くないため，正解率が1/2を上回ると評価する．左上の灰色領域は，情報理論的に検出不可能な領域で，（貪欲アルゴリズムに限らず）どんなアルゴリズムを用いても正解率が1/2となる．この境界は，本節の内容とは独立に導出されるもので，厳密な証明が存在する．[7)]

4.2　貪欲アルゴリズムが正当化される条件

図3を見ると，平均次数cが7を下回ったところでは，どんなにはっきりとしたグループ構造（$\epsilon\approx 0$）のグラフであっても検出不可能領域になっていることが分かる．[*5] 検出可能領域は，考えているグループの頂点サイズの相対的なバランスにもよる．ただ，同じくらいのグループサイズの構造を抽出しようと思ったとき，最尤法として仮定しているモデルから実際にデータが生成されている場合でさえ，必ずアルゴリズムはどこかの準安定状態に典型的にトラップされてしまうという結果は，貪欲アルゴリズムを平均次数cが7より小さい実データについて（統計推論の意味で）使うことには，後ろ盾がないことを示唆している．

5.　おわりに

本稿では，推論問題としてのグラフ分割について，アルゴリズムの検出限界という視点から，どのように統計的有意性を評価できるかの一部を紹介した．SBM上の検出限界の議論は，他のどんなアルゴリズムにも存在するが，どのように性能が失われるかの機構は，アルゴリズムによって異なるため，その理論解析には異なる技術を必要とする．また，ここでは詳しく触れなかったが，アルゴリズミックな限界の他に，SBM自体の情報理論的な検出限界（図3の左上の灰色領域）や，SBMのラベルをすべて正しく当てられるかを評価する理論限界（strong recovery）なども研究されている．[7)] ただし，グラフ分割の理論的研究は，実学としてのグラフ分割とは，大きな乖離があることも注意しておきたい．

アルゴリズムの開発や，理論的な性能評価の進展も重要だが，そもそもグラフ分割について精密な議論ができるようなデータ収集（実験）も重要である．例えば，「空手クラブネットワーク」[8)]という，ベンチマークとしてよく使われる有名なデータセットがある．これは，ある空手クラブの関係者という小さな社会の中で，インタビューによってすべての関係者の人間関係をグラフ表現したものである．別の例として，Twitter上でのフォロー関係を表すグラフデータが挙げられる．ユーザーアカウントを頂点，フォロー関係を（有向）枝として表現したもので，これも人間関係を表していると考えられる．どちらもアルゴリズムを走らせれば何らかの分割結果は得られるだろう．しかし，前者はグラフデータの対象や生成過程がよく制御・把握されている一方，後者においては，制御したり正確に把握することが困難である．SBMのようなシンプルなモデルの仮定のもと統計的評価を行うことは，空手クラブネットワークのような“実験室環境”のデータの場合には正当化されるだろうが，明らかに多様な生成過程があるデータに対して同様の評価を行ったとしても，その結果がどの程度有用かは疑問である．

理論解析と照らし合わせられるレベルの“実験室環境”のデータ収集法と，理論評価が可能な範囲の両者を近づけていくことが，より精密で価値のあるデータ科学を実現するためには必要だと，筆者は考えている．

[*4] 図3を見ると，黄色の実線を超えても正解率が0.8程度の大きな値を取っているようにも見えるが，これは臨界線付近では正解率の揺らぎが非常に大きくなっていることによるものである．図3は，100インスタンスでの正解率の平均値を濃淡で描画したが，その揺らぎかたを見てみると，ここでの計算による評価は妥当な結果となっていることが確認できる．詳しくは文献5のSupplemental Materialを参照されたい．

[*5] 精密には，$c\approx 6.6$，$\epsilon\approx 0$のとき，式(7)において$f=0$となる．

参考文献

1）B. Karrer and M. E. J. Newman, Phys. Rev. E **83**, 016107 (2011).
2）M. E. J. Newman, Phys. Rev. E **94**, 052315 (2016).
3）V. D. Blondel et al., J. Stat. Mech. **2008**, P10008 (2008).
4）B. H. Good, Y.-A. de Montjoye, and A. Clauset, Phys. Rev. E **81**, 046106 (2010).
5）T. Kawamoto and Y. Kabashima, Phys. Rev. E **99**, 010301(R) (2019).
6）F. Tanaka and S. F. Edwards, J. Phys. F **10**, 2769 (1980).
7）E. Abbe, J. Mach. Learn. Res., **18** (177), 1 (2018).
8）W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

（2020年6月1日原稿受付）

Detectability Limit of Graph Partitioning

Tatsuro Kawamoto

abstract:　Graph partitioning as an inference problem has been an important topic in multiple fields of science. In this article, we derive a performance limit called the algorithmic detectability limit on graph partitioning using a technique developed in statistical physics. This limit is a phase transition point beyond which an algorithm completely loses the ability to identify the group structure that is assumed in a random graph model.