日本オペレーションズ・リサーチ学会
2022年 秋季研究発表会

モジュラリティ密度最大化に対する高速な列生成法の設計と計算困難性の解析

東京大学 *助田一晟 SUKEDA Issey
05000113 東京大学 宮内敦史 MIYAUCHI Atsushi
01308490 東京大学/理化学研究所 武田朗子 TAKEDA Akiko

1. はじめに

ネットワーク上のコミュニティ検出とは，ネットワークをまとまりらしい構造に分割する操作であり，実社会で様々な応用をもつ.

コミュニティ検出に対しては, 様々な最適化モデルが知られている．そこで用いられる評価関数の中で最もよく知られているものが Newman と Girvan [3] により提案されたモジュラリティであるが, その欠点の一部を克服する形で Li ら [2] によりモジュラリティ密度が提案された．ここで, $G = (V, E)$ を無向グラフとし，頂点部分集合 $S \subseteq V$ に対して，$E(S) = \{\{u, v\} \in E \mid u, v \in S\}$，$E(S, V \setminus S) = \{\{u, v\} \in E \mid u \in S, v \notin S\}$ を用意すると，モジュラリティ密度はグラフの頂点集合 V の分割 $\mathcal{C}$ に対して

$$D(\mathcal{C}) = \sum_{C \in \mathcal{C}} \frac{2|E(C)| - |E(C, V \setminus C)|}{|C|}$$

と定義される．モジュラリティ密度最大化問題に対しては, いくつかの厳密解法と数々の発見的解法が考案されてきた．一方で, その計算困難性はこれまで未解明であった.

本研究では, 密グラフ抽出の技法を取り入れた高速な列生成法を設計した．列生成法の設計では，補助問題と呼ばれる最適化問題に対する解法設計が鍵となるが，その目的関数の形から密グラフ抽出問題というネットワーク解析において重要な別の問題として捉えることが可能である．密グラフ抽出に対してよく知られた解法設計の戦略として貪欲ピーリングがある．そこで提案解法の設計では, 貪欲ピーリングを用いることで, 補助問題に対する高速な発見的解法を導入した．さらに，列生成法では補助問題を厳密に解く必要があるため，その計算コストを削減するための再定式化も行なった．現実のネットワークデータを用いた数値実験では, 提案解法がサイズの大きいグラフにおいて優位であることが示された.

また, 本研究では, 列生成法の設計に加え, 計算困難性の解析も行なった．具体的には, モジュラリティ密度最大化問題の変種が NP 困難であることを示した．また, 列生成法における補助問題が NP 困難であることも示した.

2. 一般的な列生成法のアルゴリズム

指数個の変数を持つ大規模線形計画問題（主問題）に厳密解を与えるアルゴリズムとして列生成法が知られている．列生成法では以下の手順に従って主問題の厳密解を出力する．まず, 主問題の双対問題（指数個の制約式を持つ）を用意する． 双対問題の制約式のうち一部のみを考慮した小さいサイズの問題 (RDP) を解き, 暫定的に得られた最適解 $\boldsymbol{\lambda}^*$ の元の双対問題における実行可能性を補助問題を解くことにより判定する．ここで補助問題とは, 目的関数を $\boldsymbol{\lambda}^*$ による双対問題の制約式の違反度合と設定した問題を指す．違反制約が発見された場合には, その制約式を (RDP) に追加し, 同様のステップを繰り返す．違反制約が 1 本も発見されなくなったら反復を終了し, $\boldsymbol{\lambda}^*$ を出力する．出力された解は元の双対問題の実行可能解となっているので, 最適性が保証される.

列生成法は反復計算回数が大きくなりやすく, また補助問題 1 回あたりの計算コストも小さくないことから, 収束が遅い側面がある．そこで, 補助問題を解く手前で発見的に違反制約を探索するステップを加えるのが一般的である.

3. 提案解法：高速な列生成法の設計

本研究ではモジュラリティ密度最大化問題を大規模整数計画問題として定式化した後, 連続緩和を施した問題を主問題と設定する.

そして, 以下の補助問題 (1) を密グラフ抽出問題として捉え直し, 貪欲ピーリングを適用する.

$$\begin{aligned} &\text{maximize} \quad \frac{2|E(S)| - |E(S, V \setminus S)|}{|S|} - \sum_{v \in S} \lambda_v^* \\ &\text{subject to} \quad S \in \mathcal{S}. \end{aligned} \tag{1}$$

貪欲ピーリングとは, 密グラフ抽出問題に対する解法設計の戦略であり, グラフの頂点集合 V から開始して順次最も貢献していない頂点を除いていく貪欲アルゴリズムである．貢献の大小を測る指標として,

表 1: 実験結果

ID	$(\lvert V\rvert, \lvert E\rvert)$	[1]	[4]	Ours
1	(24,38)	5.9±0.3	0.4	**0.3**
2	(34,78)	43.7±1.9	**0.7**	2.4
3	(62,159)	3261.4±316.5	**16.4**	42.8
4	(77,254)	13294.0±1991.1	**57.9**	119.7
5	(105,441)	> 86400	**365.4**	1886.3
6	(112,425)	> 86400	> 86400	**5681.1**
7	(115,613)	> 86400	70195.1	**5860.8**
8	(198,2742)	> 86400	> 86400	> 86400

補助問題の目的関数を $f_1(S)/|S| = \Sigma_{v\in S} f_2(v)/|S|$ と表した時の $f_1(S), f_2(v)$ を用いて, 以下の 2 種類

$$\text{寄与} : f_2(v)$$

$$\text{差分} : f_1(S) - f_1(S \setminus \{v\})$$

のそれぞれに重みづけを施したものを利用する.

さらに, 補助問題を厳密に解く際には, 目的関数の分母 $|S|$ を定数 k で固定した問題

$$\text{maximize} \quad \frac{2|E(S)| - |E(S, V \setminus S)| - |S| \sum_{v\in S} \lambda_v^*}{k}$$

を各 $k = 1, \ldots, |V|$ に対し解くよう改変を施す. これにより各反復につき $|V|$ 個の最適化問題を解く必要性が出てくるが, 非線形な整数計画問題を線形な整数計画問題に等価に改変してから解くので, 数理計画ソルバーでより高速に解くことが可能である. 加えて最適解が $|V|$ 個得られることとなり, 多様な違反制約を一度の反復計算の間に発見しやすくなっている.

4. 数値実験

紙面の都合上一部のみを報告する. インスタンスは 8 種類の実際のネットワークデータを利用した. 実験結果を表 1 に示す. 表中の数字は列生成アルゴリズムを用いて最適解を得るまでの計算時間（秒）を示している. ただし CGII+ILS [1] と Ours は主問題の最適解を, Sato and Izunaga [4] は緩和前のモジュラリティ密度最大化問題の最適解を出力することに留意されたい. また, CGII+ILS [1] の結果はランダムネスを含んでいるので, 5 回の試行の平均 ± 標準偏差を記載した. 提案解法は比較的サイズが大きいインスタンスに対し有効性が認められた. 特に先行研究において 24 時間の時間制限超過により最適解を求めることが出来ていなかったインスタンスに対し, 2 時間未満の計算時間で最適解を得た.

5. 計算困難性の解析

5.1. モジュラリティ密度最大化問題の困難性

モジュラリティ密度最大化問題の NP 困難性については, Li ら [2] をはじめ複数の文献で言及されてきたが, その根拠は与えられてこなかった. モジュラリティ密度最大化問題はその定式化から十分密度の高いグラフに対しては分割をしない解（頂点全体）を出力する. 本研究では, その自明な解を除外した問題を考え, 以下の定理を示した.

定理 1. 自明な解 ($|\mathcal{C}| = 1$) を除外したモジュラリティ密度最大化問題

$$\begin{aligned} &\text{maximize} && D(\mathcal{C}) \\ &\text{subject to} && |\mathcal{C}| \geq 2 \end{aligned}$$

は正則グラフ上に限っても NP 困難である.

NP 困難だと知られている 3-正則グラフ上の最大カット問題を帰着することで示した.

5.2. 補助問題の困難性

続いて, 列生成法で現れる補助問題の困難性を報告する. ここでは $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$ は問題の入力として与えられるものとする.

定理 2. 列生成法の補助問題

$$\begin{aligned} &\text{maximize} && \frac{2|E(S)| - |E(S, V \setminus S)|}{|S|} - \sum_{v\in S} \lambda_v \\ &\text{subject to} && S \subseteq V \end{aligned}$$

は $\lambda_v > 0\ (\forall v \in V)$, $(n-4)$-正則グラフ上に限っても NP 困難である.

NP 困難だと知られている $(n-4)$-正則グラフ上の最大クリーク問題を帰着することで示した.

参考文献

[1] R. de Santiago, L. C. Lamb, Exact computational solution of modularity density maximization by effective column generation, Computers & Operations Research 86 (2017) 18–29.

[2] Z. Li, S. Zhang, R.-S. Wang, X.-S. Zhang, L. Chen, Quantitative function for community detection, Physical Review E 77 (2008) 036109.

[3] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E 69 (2004) 026113.

[4] K. Sato, Y. Izunaga, An enhanced MILP-based branch-and-price approach to modularity density maximization on graphs, Computers & Operations Research 106 (2019) 236–245.