修士論文 要旨

ソフトモジュラリティ最大化問題の共定値問題への帰着

数理情報学専攻 48206234　松野　聖也
指導教員　武田　朗子　教授

1　はじめに

実ネットワークの大きな特徴の 1 つに密なコミュニティの存在があり、ネットワークの存在するありとあらゆる問題でコミュニティ検出は重要な問題となっている。モジュラリティはコミュニティ構造の質を測る関数として導入されて以降、コミュニティ検出の主要な方法として Newman, Girvan が導入したモジュラリティ最大化問題を解く手法が用いられている [1]。モジュラリティ最大化問題とは、重み付きグラフ $G=(V,E)$, 接続行列 $W=\{W_{ij}\}_{i,j}$, $w_i=\sum_j W_{ij}$,$w=\sum_i w_i$ と置いた時以下のような問題を指す。

$$\max_{p\in\mathbb{Z}^{n\times n}} \ \frac{1}{w}\sum_{i,j\in V}^{n}\sum_{k=1}^{n}\left(W_{ij}-\frac{w_i w_j}{w}\right)p_{ik}p_{jk}$$
$$\text{s.t.}\ \ \forall i\in V,\ \sum_{k=1}^{n}p_{ik}=1,\ p_{ik}\geq 0$$

Hollocou らはこのコミュニティへの各頂点の所属度合いを、従来の 0 か 1 かの 2 択だったモジュラリティ最大化問題に対し実数値をとることを許容したモジュラリティを定義し、ソフトモジュラリティと呼んだ [2]。その最大化問題は以下のように定義される。

$$\max_{p\in\mathbb{R}^{n\times n}} \ \frac{1}{w}\sum_{i,j\in V}^{n}\sum_{k=1}^{n}\left(W_{ij}-\frac{w_i w_j}{w}\right)p_{ik}p_{jk}$$
$$\text{s.t.}\ \ \forall i\in V,\ \sum_{k=1}^{n}p_{ik}=1,\ p_{ik}\geq 0$$

ソフトモジュラリティ最大化問題は従来のモジュラリティ (ハードモジュラリティと呼ばれる) 最大化問題と違い、個々のノードが複数の集団へ寄与することを許容する問題設定となっており、注目を集めている。今回はハードモジュラリティ最大化問題、ソフトモジュラリティ最大化問題の理論保証を目的として、再定式化を行った。

2　ハードモジュラリティ最大化問題の再定式化

クラスタリング問題では、クラスターの割り当てについて縦軸を頂点、横軸を該当クラスターに当てはめた行列に対応させて考える。ハードモジュラリティ最大化問題の場合は、X を用いて以下のように表現することができる。

$$\max_{Z_{\mathrm{mod}}}\ \mathrm{Tr}\langle B,\ Z_{\mathrm{mod}}\rangle,\ Z_{\mathrm{mod}}=XX^\top$$

Z_{mod} の満たす条件は非常に少なく、近似に用いることが難しくなっている。過去の研究ではモジュラリティ密度最大化問題 [3] を定義したり、標準化モジュラリティ最大化問題 [4] を定義したりすることで、k-means 最小化問題と同じ $X(X^\top X)^{-1}X^\top$ の形式に帰着させさせているケースもある。

今回は特徴が少なく、考えるのが難しいハードモジュラリティに使われる行列について以下の性質を示した。

定理 1. 行列 Z が X を割り当て行列として $Z=XX^\top$ と、ハードモジュラリティに用いられる行列の表記を用いて書き表せることは以下と同値である。

$$Z_{ii}=1\ \ \exists k\in \mathrm{Z}_+,\ \exists X=\{\boldsymbol{x}_1,\ldots \boldsymbol{x}_k\}\geq 0\in \mathrm{R}^{n\times k}$$
$$\text{s.t.}\ Z=XX^\top,\ \boldsymbol{x}_i^\top\boldsymbol{x}_j=0\ \forall i\neq j$$

この定理に関連して示唆できる内容として、$Z_{ii}=1$ という条件は実はそこまで本質的なものではないということである。実際、証明の際に重要だったのは、「合計が 1 かつ正の要素が 1 個しかない状況を作ること」であった。そこで、$Z_{ii}=1$ という条件を

$$X_{i1}+\cdots+X_{ik}=1\ \forall i\in[1,n]$$

としても同様の証明で等価性が示される。そしてさらに興味深いこととして、直交性を除いた条件

$$Z=XX^\top$$
$$X_{i1}+\cdots+X_{ik}=1\ \forall i\in[1,n]$$
$$X_{ij}\geq 0$$

については、ソフトモジュラリティ最大化問題を行列の内積として再定式化した時に課される条件を等価になることである。つまり、離散的な条件の本質に直交性があることが示唆できる。このような直交性の有無という形でソフトモジュラリティとハードモジュラリティの区別をした考え方は論文を見渡す限り無かった新しい考えとなる。

3 ソフトモジュラリティ最大化問題の再定式化

行列として表現されていた変数 p を、

$$\hat{\boldsymbol{p}} \in \mathbb{R}^{n^2},\ \hat{p}_{(j-1)n+i} = p_{ij} \quad i,j = [1,n]$$

と、ベクトル $\hat{\boldsymbol{p}}$ として表現し、行列 $\hat{W}$ を

$$\hat{W} = \begin{pmatrix} \bar{W} & & & & \\ & \bar{W} & & 0 & \\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & & \bar{W} \end{pmatrix} \tag{1}$$

$$\{\bar{W}\}_{ij} = W_{ij} - \frac{w_i w_j}{w}$$

とすれば、以下の 2 次計画問題に書き換えられる。

$$\begin{aligned} \max_{\hat{\boldsymbol{p}} \in \mathbb{R}^{n^2}} \quad & \hat{\boldsymbol{p}}^\top \hat{W} \hat{\boldsymbol{p}} \\ \text{s.t.} \quad & \boldsymbol{e}_i^\top \hat{\boldsymbol{p}} = 1 \quad \forall i = 1,\dots,n \\ & \hat{\boldsymbol{p}} \geq 0 \end{aligned} \tag{2}$$

この問題に [6] で用いられた再定式化を用いる。

$$\begin{aligned} \max_{X} \quad & \langle \hat{W}, X \rangle \\ \text{s.t.} \quad & \boldsymbol{e}_i^\top X \boldsymbol{e}_1 = 1 \quad \forall i = 1,\dots,n \\ & \boldsymbol{e}_i^\top X \boldsymbol{e}_i = 1 \quad \forall i = 1,\dots,n \\ & X \in C^*_{n^2} \end{aligned}$$

しかし、C^*_n は完全正値行列を指し、$\boldsymbol{e}_i$ は

$$(\boldsymbol{e}_i)_j = \begin{cases} 1 & j \equiv i \ (\mathrm{mod} n) \\ 0 & \text{o.w.} \end{cases}$$

と定義されるものとする。$(\boldsymbol{e_1}\boldsymbol{e_i}^\top + (\boldsymbol{e_1}\boldsymbol{e_i}^\top)^\top)/2 = E_{1i}$, $\boldsymbol{e_i}\boldsymbol{e_i}^\top = E_{ii}$ と表記すると、$\boldsymbol{e}_i^\top X \boldsymbol{e}_1 = 1 \Leftrightarrow \langle X, E_{1i} \rangle = 1$, $\boldsymbol{e}_i^\top X \boldsymbol{e}_i = 1 \Leftrightarrow \langle X, E_{ii} \rangle = 1$ となることからラグランジュ双対をとると、双対問題は

$$\begin{aligned} \min_{\lambda} \quad & \sum_{i=1}^{n} \lambda_{ii} + \sum_{j=2}^{n} \lambda_{1j} \\ \text{s.t.} \quad & -\hat{W} + \sum_{i} \lambda_{1i} E_{1i} + \sum_{i} \lambda_{ii} E_{ii} \in C_{n^2} \end{aligned} \tag{3}$$

と書くことができる。ここで、C_n は n 次元共正値行列の集合を指す。主計画問題も双対問題も狭義実行可能解は自明に持つため錐計画問題についての双対定理 [5] より、問題 (3) の最小値は問題 (2) の最大値に一致する。目的関数が $2n-1$ 次線形の共正値計画問題は、共正値計画問題は目的関数が 1 次 1 変数でさえ NP 困難となることが知られており、非常に複雑な問題設定であることが分かる。ここで、共正値性についてさらに分析していくと、共正値錐を緩和した、目的変数が 1 変数の問題へと再定式化が可能となる。

定理 2. ソフトモジュラリティ最大化問題

$$\begin{aligned} \max_{p \in \mathbb{R}^{n \times n}} \quad & \frac{1}{w} \sum_{i,j \in V} \sum_{k=1}^{n} \left(W_{ij} - \frac{w_i w_j}{w} \right) p_{ik} p_{jk} \\ \text{s.t.} \quad & \forall i \in V,\ \sum_{k=1}^{n} p_{ik} = 1,\ p_{ik} \geq 0 \end{aligned}$$

は、以下の問題を解くことと同値である。

$$\begin{aligned} \min \quad & \lambda \\ \text{s.t.} \quad & \min_{\boldsymbol{x} \in R} \quad -\sum_{i=1}^{n} \bar{\boldsymbol{x}}_i^\top \bar{W} \bar{\boldsymbol{x}}_i + \lambda y_1^2 \geq 0 \\ & R = \left\{ \boldsymbol{x} \in \mathbb{R}^{n^2}_{\geq 0} \quad \text{s.t.} \quad y_i = y_j \quad \forall i,j \in [1,n] \right\} \end{aligned}$$

ただし、$\bar{\boldsymbol{x}}_i = (x_{(i-1)n+1}, \dots, x_{in}), \boldsymbol{y}_i = \sum_{k=0}^{n-1} x_{kn+i}$ であり、$\hat{W}$ は (1) で定義されるものとする。

参考文献

[1] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113 (2004).

[2] A. Hollocou, T. Bonald, and M. Lelarge. Modularity-based sparse soft graph clustering. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, 323-332 (2019).

[3] Y. Izunaga, T. Matsui, and Y. Yamamoto. A doubly nonnegative relaxation for modularity density maximization. *Discrete Applied Mathematics*, 275, 69-78 (2020).

[4] M. Bolla. Penalized versions of the Newman-Girvan modularity and their relation to normalized cuts and k-means clustering. *Physical Review E*, 84(1), 016108 (2011).

[5] A. Bental, and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, Society for Industrial and Applied Mathematics (2001)

[6] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120, 479-495 (2009).