局所密分解の一般化

松下 祐介

東京大学理学部情報科学科

yskm24t@is.s.u-tokyo.ac.jp

2018 年 6 月 30 日

1　概要

まず、Tatti & Gionis (2015) の研究について紹介しよう。Tatti & Gionis (2015) では、無向グラフの密度関数を (辺数)/(頂点数) として、無向グラフの点誘導部分グラフで**局所密** (locally dense) という性質を満たすものを列挙する**局所密分解** (locally-dense decomposition) が考案され、厳密アルゴリズムと線形時間近似アルゴリズムが与えられている。局所密集合の族は包含関係について全順序部分集合をなし、**コア分解** (core decomposition) と類似の特長をもつが、内側の局所密集合ほど密度が濃いという性質が必ず成立する（Tatti & Gionis (2015) の図 1 からわかるように、コア分解は必ずしもこの性質を満たさない）という点でコア分解より性質が良い。また多くの場合、局所密分解ではコア分解に比べてかなり細かく分かれた階層が得られる（Tatti & Gionis (2015) の表 3）。

さて、ここではさらに一般化して、頂点集合の冪集合の代わりに任意の有限束上で考えることにして、分子の「**質量関数**」を辺数の代わりに任意の単調優モ

ジュラ関数、分母の「**サイズ関数**」を頂点数の代わりに任意の単調劣モジュラ関数（ただし、いずれの関数も最小値は 0 であり、質量関数は最大元でのみ最大値を取り、サイズ関数は最小元でのみ最小値を取る）としても諸性質が成立することを示す。また、独自に興味深い性質の発見、証明の改善、アルゴリズムの改善などをおこなった。

そして、定義域が有限集合の冪集合である場合について一般に局所密分解のための多項式時間アルゴリズムが得られることについて触れ、さらに辺の密度に限らず多重集合上の任意点数の**パターン**の密度にも局所密分解の枠組みが使えることを示した。辺の密度については Tatti & Gionis (2015) での議論を詳しく説明し直した。

実装としては、辺の密度に関する局所密分解を書いた。最大フロー最小カットには Dinic のアルゴリズムを用いた。

2 基本概念

2 元以上持つ（最小元と最大元が異なる）有限束 $\mathcal{L}$ を自由に選ぶ。最小元を $\bot$、最大元を $\top$ と書くことにする。束の演算は $\sqsubseteq, \sqsubset, \sqcap, \sqcup$ で表すことにする。この有限束の元は**コンテナ**と呼ぶことにしよう。**質量関数** $m\colon \mathcal{L} \to \mathbb{R}_{\geq 0}$ と**サイズ関数** $s\colon \mathcal{L} \to \mathbb{R}_{\geq 0}$ で、以下を満たすものを任意に取って考えることにする。[*1]

- m, s は単調、すなわち $X \sqsubseteq Y$ なら $m(X) \leq m(Y)$ かつ $s(X) \leq s(Y)$。
- m は**優モジュラ** (supermodular)、すなわち $m(X) + m(Y) \leq m(X \sqcap Y) + m(X \sqcup Y)$。
- s は**劣モジュラ** (submodular)、すなわち $s(X) + s(Y) \geq s(X \sqcap Y) + s(X \sqcup Y)$。

[*1] 質量関数、サイズ関数という名称は自分で考えた。物理における密度のアナロジーである。

- $m(\bot) = s(\bot) = 0$。[*2]
- $X \sqsubset \top$ ならば $m(X) < m(\top)$ であり、$Y \sqsupset \bot$ ならば $s(Y) > s(\bot) = 0$ である。[*3]

特に、Tatti & Gionis (2015) では有限無向グラフ $G = (V, E)$ を考え、$\mathcal{L}$ を V の冪集合の束として $m(X) := |E(X)|,\ s(X) := |X|$ と定義した場合が論じられている。

定義 1. $s(A) > 0$ であるとき、A の**密度**を $d(A) := m(A)/s(A)$ として定義する。

$A \sqsubseteq B$ であるとき、B の A に対する**外質量**、**外サイズ**をそれぞれ $m_A(B) := m(B) - m(A),\ s_A(B) := s(B) - s(A)$ として定義し、B の A に対する**外密度** (outer density) をそれらの比 $d_A(B) := m_A(B)/s_A(B)$ として定義する。特に、$d_\bot(B) = d(B)$ が成立することに注意。

ただし、密度および外密度の計算において、(正の数)$/0 = \infty,\ 0/0 = 0$ とする。$\mathbb{R} \cup (\infty)$ は上に有界な全順序集合をなすと考える（$\infty < \infty$ は成立しないとする）。

次の補題は以下の議論で多用される。[*4]

[*2] これは外密度 $d_\bot(X)$ が密度 $d(X)$ と一致することを要請するための制約にすぎない。

[*3] この条件が仮に満たされていなくても、m は優モジュラであるので、$m(X)$ を最大化する X 全体は $\mathcal{L}$ の部分束をなし、特にそのような X の中で束で最小のもの $\top_*$ が取れ、同様に $s(Y)$ を最小化する（すなわち $s(Y) = 0$ となる）Y の中で束で最大のもの $\bot^*$ が取れる。

$\mathcal{L}$ が有限集合 V の冪集合 2^V である場合を考えると、質量の優モジュラ性より $x \notin \top_*$ であることと x が質量の増加に全く寄与しない（すなわち、いかなる集合 $X \in 2^V$ についても $m(X \cup \{x\}) = m(X)$ である）ことは同値であるし、サイズの劣モジュラ性より $x \in \bot^*$ であることと x がサイズの増加に全く寄与しないことも同値である。

ゆえに、$\bot^* \sqsubseteq X \sqsubseteq \top_*$ の範囲で考えるようにすることは特に実用上の問題にならないと思う。

[*4] Tatti & Gionis (2015) ではこれら 2 つの補題ははっきり述べられていない。ここでの補題 2 は、Tatti & Gionis (2015) の命題 3 の証明の中で暗に使われている。ここでの補題 3 は、Tatti & Gionis (2015) の補題 6 と補題 7 で部分的に述べられている。

補題 2. $d_{A\sqcap B}(B) \geq d_A(A \sqcup B)$ が成立する。

証明. m の優モジュラ性、s の劣モジュラ性より明らか。 □

補題 3. $A \sqsubset B \sqsubset C$ であり $(m_A(B), s_A(B)), (m_A(C), s_A(C)), (m_B(C), s_B(C)) \neq (0, 0)$ であるとき、以下は同値。(i) $d_A(B) \gtreqless d_B(C)$。(ii) $d_A(B) \gtreqless d_A(C)$。(iii) $d_A(C) \gtreqless d_B(C)$。ただし、複号は同順。

証明. 分数の中間値の性質より明らかである。 □

3 局所密分解

まず、いくつか道具を導入する。

定義 4. パラメータ $\alpha \in \mathbb{R}_{\geq 0}$ を取る関数 $\delta_\alpha(X) := m(X) - \alpha s(X)$ を考えよう。$\delta_\alpha(X)$ は優モジュラであるから、$\delta_\alpha(X)$ を最大化する X 全体は $\mathcal{L}$ の部分束をなす。そのような X のうち束で最小のものを $\Delta_*(\alpha)$、束で最大のものを $\Delta^*(\alpha)$ と書くことにする。

Tatti & Gionis (2015) では $\Delta^*(\alpha)$ の側のみを（F という表記で）導入しているが、以下では $\Delta_*(\alpha)$ の側を主に扱う。

$\Delta_*(\alpha)$, $\Delta^*(\alpha)$ はサイズに関して α 倍のペナルティーを与えた上で質量をどれだけ大きくできるかということだから、実用的な観点からも非常に自然な問いに答えるものだということがわかるだろう。

以下の補題が重要である。[*5]

[*5] Tatti & Gionis (2015) ではこの補題は命題 11 の証明中で使われているが、補題として明示的に設けることで、局所密の定義がよりわかりやすくなるし、命題 14 の証明にも使えるようになった。

補題 5. 任意の $\alpha \in \mathbb{R}_{\geq 0}$ を取る。$X \sqsubset \Delta_*(\alpha) \sqsubseteq Y$ であるならば $d_X(\Delta_*(\alpha)) > \alpha \geq d_{\Delta_*(\alpha)}(Y)$ である。

証明. $\delta_\alpha(\Delta_*(\alpha)) > \delta_\alpha(X)$, $\delta_\alpha(Y) \leq \delta_\alpha(\Delta_*(\alpha))$ より $m_X(\Delta_*(\alpha)) - \alpha s_X(\Delta_*(\alpha)) > 0$, $m_{\Delta_*(\alpha)}(Y) - \alpha s_{\Delta_*(\alpha)}(Y) \leq 0$ であるから、$d_X(\Delta_*(\alpha)) > \alpha \geq d_{\Delta_*(\alpha)}(Y)$ が得られる。[*6] □

この補題より $\Delta_*(\alpha)$ が満たす性質として、局所密という概念が考えられる。

定義 6. コンテナ A が**局所密** (locally dense) であるというのを、$X \sqsubset A \sqsubseteq Y$ であるとき $d_X(A) > d_A(Y)$ が成立することとして定義する。

以下の綺麗な性質が得られる。

命題 7. A, B が局所密であるならば、$A \sqsubseteq B$ と $A \sqsupseteq B$ のいずれかが成立する。[*7]

証明. 対称性より $d_{A \sqcap B}(A) \leq d_{A \sqcap B}(B)$ としてよい。補題 2 より $d_{A \sqcap B}(B) \leq d_A(A \sqcup B)$ が成立するから、$d_{A \sqcap B}(A) \leq d_A(A \sqcup B)$ が得られる。A は局所密であるから、$A \sqcap B = A$ すなわち $A \sqsubseteq B$ でなくてはならない。 □

定義 8. 先ほどの命題を踏まえて、局所密なコンテナ全体を $D_0 \sqsubset D_1 \sqsubset \ldots \sqsubset D_K$ というように番号付けることにする。このように局所密なコンテナの族を得ることを**局所密分解** (locally-dense decomposition) と呼ぶことにする。

$\bot$ は $X \sqsubset \bot$ なる X が取れないから局所密となる。ゆえに $D_0 = \bot$ である。

$d_X(\top) > 0 = d_\top(\top)$ が成立するから、$\top$ は局所密となる。ゆえに $D_K = \top$ である。

[*6] ここで $\alpha \geq d_{\Delta_*(\alpha)}(Y)$ を成立させるために、$0/0 = 0$ と定義したのである。

[*7] これは Tatti & Gionis (2015) で命題 3 として述べられている内容である。

補足 9. 局所密の定義より $\infty > d_{D_0}(D_1) > d_{D_1}(D_2) > \cdots > d_{D_{K-1}}(D_K) > 0$ が成立するということに注意。この性質を拡張して、$d_{D_{-1}}(D_0) := \infty$, $d_{D_K}(D_{K+1}) := 0$ と定義することにしよう（これ以外の形で D_{-1}, D_{K+1} は登場しない）。

補足 10. 局所密の定義より $1 \leq i \leq K-1$ について $d(D_i) = d_{\perp}(D_i) > d_{D_i}(D_{i+1})$ であるから、補題 3 より $d(D_1) > d(D_2) > \cdots > d(D_K)$ が成立することがわかる。これは、コア分解では見られない重要な性質である。

補題 5 より任意の $\alpha \in \mathbb{R}_{\geq 0}$ について $\Delta_*(\alpha)$ は局所密であったが、実は任意の局所密なコンテナはこの形で書ける。

命題 11. 任意の $i = 0, \ldots, K$ について、$d_{D_{i-1}}(D_i) > \alpha \geq d_{D_i}(D_{i+1})$ であれば $D_i = \Delta_*(\alpha)$ が成立する。[*8]

証明. $i+1 \leq j \leq K$ に対し、補足 9 より $\alpha \geq d_{D_{j-1}}(D_j)$ であるから、$m_{D_{j-1}}(D_j) - \alpha s_{D_{j-1}}(D_j) \leq 0$ が成立するので、$\delta_\alpha(D_{j-1}) \geq \delta_\alpha(D_j)$ が得られる。$1 \leq j \leq i$ に対し、補足 9 より $d_{D_{j-1}}(D_j) > \alpha$ であるから、同様に $\delta_\alpha(D_{j-1}) < \delta_\alpha(D_j)$ が得られる。ゆえに $\cdots < \delta_\alpha(D_{i-1}) < \delta_\alpha(D_i) \geq \delta_\alpha(D_{i+1}) \geq \ldots$ であるから、$\Delta_*(\alpha) = D_i$ が得られる。 □

補足 12. 以上より、局所密であることと $\Delta_*(\alpha)$ の値域に属することは同値であるとわかった。

また、局所密分解 $D_0 \sqsubset D_1 \sqsubset D_2 \sqsubset \ldots \sqsubset D_{K-1} \sqsubset D_K$ が得られれば $d_{D_0}(D_1)$, $d_{D_1}(D_2), \ldots, d_{D_{K-1}}(D_K)$ もわかるから、$\Delta_*(\alpha)$ の関数形が得られたことになる。

ここで、以下の概念を命題 14 のために補助的に導入する。

---

*8 これは Tatti & Gionis (2015) の命題 9 の主張を Δ^* の代わりに Δ_* に合うように微修正したものである。

定義 13. コンテナ A は任意の $X \sqsubset A$ について $(m_X(A), s_X(A)) \neq (0,0)$ を満たしているとき、**凝縮**しているといわれる。

特に、局所密なコンテナは凝縮している。

なお、m, s のいずれかが狭義単調である場合はすべてのコンテナが凝縮していることになる。

さて、命題 11 を使うと、隣り合う番号の局所密集合について興味深い性質が得られる。

命題 14. $i = 0, \ldots, K-1$ に対し、D_{i+1} は $d_{D_i}(Y)$ を最大化する凝縮した $Y \sqsupseteq D_i$ のうちで束で最大のものである。[*9]

証明. $\alpha = d_{D_i}(D_{i+1})$ とすると、命題 11 より $D_i = \Delta_*(\alpha)$ であり、さらに補題 5 より任意の $Y \sqsupseteq \Delta_*(\alpha) = D_i$ について $d_{D_i}(D_{i+1}) = \alpha \geq d_{\Delta_*(\alpha)}(Y) = d_{D_i}(Y)$ であるから、D_{i+1} は $d_{D_i}(Y)$ を最大化する。

よって、$d_{D_i}(Y)$ を最大化する凝縮した C が任意に与えられたときに $C \sqsubseteq D_{i+1}$ となることを示せばよい。

矛盾を導くために $C \not\sqsubseteq D_{i+1}$ と仮定する。$C \sqcap D_{i+1} \sqsubset C$ となるから、C の凝縮性より $(m_{C \sqcap D_{i+1}}(C), s_{C \sqcap D_{i+1}}(C)) \neq (0,0)$ (*) が得られる。$d_{D_i}(C \sqcap D_{i+1}) \leq d_{D_i}(C)$ であるから、補題 3 と (*) より $d_{C \sqcap D_{i+1}}(C) \geq d_{D_i}(C)$ が得られる。ゆえに、補題 2 を使うと $d_{D_{i+1}}(C \sqcup D_{i+1}) \geq d_{C \sqcap D_{i+1}}(C) \geq d_{D_i}(C) = d_{D_i}(D_{i+1})$ が得られるが、これは D_{i+1} の局所密性に反する。 □

系 15. D_1 は、密度最大の凝縮しているコンテナのうち、束で最大のものである。

[*9] D_{i+1} は $d_{D_i}(Y)$ を最大化するという主張は Tatti & Gionis (2015) で命題 5 として紹介されており、これを最大化する凝縮したコンテナのうちで束で最大となるという性質は自分で気づいた。Tatti & Gionis (2015) の証明はかなり複雑なものになっているが、ここでは Δ_* を使うことで簡潔に示している。

同様に、以下も成立する。

命題 16. $i = 1, \ldots, K$ に対し、D_{i-1} は $d_X(D_i)$ を最小化する $X \sqsubset D_i$ のうちで束で最小のものである。

証明. 命題 11 より、任意の十分小さい $\epsilon > 0$ に対して $\alpha = d_{D_{i-1}}(D_i) - \epsilon$ としたときに $D_i = \Delta_*(\alpha)$ であり、さらに補題 5 より任意の $X \sqsubset D_i = \Delta_*(\alpha)$ について $d_{D_{i-1}}(D_i) - \epsilon = \alpha < d_X(\Delta_*(\alpha)) = d_X(D_i)$ であるから、$\epsilon \to 0$ を考えると $d_{D_{i-1}}(D_i) \leq d_X(D_i)$ が得られる。ゆえに、D_{i+1} は $d_X(D_i)$ を最小化する。

よって、$d_{D_i}(Y)$ を最小化する C が任意に与えられたときに $D_{i-1} \sqsubseteq C$ となることを示せばよい。

そうでないと仮定すると、$D_{i-1} \sqcap C \sqsubset D_{i-1}$ であるから D_{i-1} の局所密性より $d_{D_{i-1} \sqcap C}(D_{i-1}) > d_{D_{i-1}}(D_i) = d_C(D_i)$ である。補題 2 より $d_C(D_{i-1} \sqcup C) \geq d_{D_{i-1} \sqcap C}(D_{i-1})$ が得られるから $d_C(D_{i-1} \sqcup C) > d_C(D_i)$ であるので、補題 3 より $d_{D_{i-1} \sqcup C}(D_i) > d_C(D_i)$ となり C が $d_{D_i}(Y)$ を最小化するという仮定に反する。 □

また、次の性質はアルゴリズムの設計上重要である。

命題 17. $0 \leq a < b \leq K$ が与えられているとき、$\alpha := d_{D_a}(D_b)$, $D_c := \Delta_*(\alpha)$ とすると、$a + 1 < b$ ならば $a < c < b$ が、$a + 1 = b$ ならば $c = a$ が成立する。[*10]

証明. 命題 11 を用いる。$a + 1 < b$ の場合は補足 9 と補題 3 より $d_{D_a}(D_{a+1}) > d_{D_a}(D_b) > d_{D_{b-1}}(D_b)$ が成立するから、$a < c < b$ が満たされる。$a + 1 = b$ なら $d_{D_a}(D_{a+1}) = d_{D_a}(D_b)$ より $c = a$ が得られる。 □

[*10] Tatti & Gionis (2015) では $\alpha := d_{D_a}(D_b) + |V|^{-2}$ として Δ^* を使う方針を取っているが、こうすると流量を有理数として正確に計算する場合には分母が大きくなりすぎるので、$\alpha := d_{D_a}(D_b)$ とするのが良いと思う。

さて、これで厳密アルゴリズムが得られるわけであるが、一旦少し脱線した話題について述べておこうと思う。

3.1 $\Delta^*(\alpha)$ を使う場合

$\Delta_*(\alpha)$ の代わりに $\Delta^*(\alpha)$ を使う場合について考えよう。両者はかなり似た性質を示すことがわかり、特に m, s のいずれかが狭義単調である場合は有限個の α を除いて $\Delta_*(\alpha) = \Delta^*(\alpha)$ となる。

$\Delta_*(\alpha)$ の場合と同様にして得られる証明は適宜省いた。

密度や外密度の定義において、0/0 を 0 でなくて ∞ と定義したものを、$d(Y)$ や $d_X(Y)$ の代わりに $\hat{d}(Y)$ や $\hat{d}_X(Y)$ と書くことにする。

こうすると、後は同様に議論することができる。

なお、局所密の代わりに**飽和局所密**と呼ぶことにした。のちほど凝縮に対応するものとして飽和という概念を導入するので、この名前を付けた。元の局所密はここでは**凝縮局所密**と呼ぶことにしよう。

補題 18. 任意の $\alpha \in \mathbb{R}_{\geq 0}$ を取る。$X \sqsubseteq \Delta^*(\alpha) \sqsubset Y$ であるならば $\hat{d}_X(\Delta^*(\alpha)) \geq \alpha > \hat{d}_{\Delta^*(\alpha)}(Y)$ である。

定義 19. コンテナ A が**飽和局所密**であるというのを、$X \sqsubseteq A \sqsubset Y$ であるとき $\hat{d}_X(A) > \hat{d}_A(Y)$ が成立することとして定義する。

命題 20. A, B が飽和局所密であるならば、$A \sqsubseteq B$ と $A \sqsupseteq B$ のいずれかが成立する。

定義 21. 先ほどの命題を踏まえて、飽和局所密なコンテナ全体を $\hat{D}_0 \sqsubset \hat{D}_1 \sqsubset \ldots \sqsubset \hat{D}_{\hat{K}}$ というように番号付けることにする。このように飽和局所密なコンテナの族を得ることを**飽和局所密分解**と呼ぶことにする。

$\hat{D}_0 = \bot$ と $\hat{D}_{\hat{K}} = \top$ が成立する。

補足 22. 飽和局所密の定義より $\infty > \hat{d}_{\hat{D}_0}(\hat{D}_1) > \hat{d}_{\hat{D}_1}(\hat{D}_2) > \cdots > \hat{d}_{\hat{D}_{\hat{K}-1}}(\hat{D}_{\hat{K}}) > 0$ が成立するということに注意。この性質を拡張して、$\hat{d}_{\hat{D}_{-1}}(\hat{D}_0) := \infty$, $\hat{d}_{\hat{D}_{\hat{K}}}(\hat{D}_{\hat{K}+1}) := 0$ と定義することにしよう（これ以外の形で $\hat{D}_{-1}$, $\hat{D}_{\hat{K}+1}$ は登場しない）。

補足 23. $\hat{d}(\hat{D}_1) > \hat{d}(\hat{D}_2) > \cdots > \hat{d}(\hat{D}_{\hat{K}})$ が成立する。

命題 24. 任意の $i = 0, \ldots, \hat{K}$ について、$\hat{d}_{\hat{D}_{i-1}}(\hat{D}_i) \geq \alpha > \hat{d}_{\hat{D}_i}(\hat{D}_{i+1})$ であれば $\hat{D}_i = \Delta^*(\alpha)$ が成立する。なお、$\Delta^*(0) = \hat{D}_{\hat{K}} = \top$ である。

補足 25. 以上より、飽和局所密であることと $\Delta^*(\alpha)$ の値域に属することは同値であるとわかった。

また、飽和局所密分解 $\hat{D}_0 \sqsubset \hat{D}_1 \sqsubset \hat{D}_2 \sqsubset \ldots \sqsubset \hat{D}_{\hat{K}-1} \sqsubset \hat{D}_{\hat{K}}$ が得られれば $\hat{d}_{\hat{D}_0}(\hat{D}_1), \hat{d}_{\hat{D}_1}(\hat{D}_2), \ldots, \hat{d}_{\hat{D}_{\hat{K}-1}}(\hat{D}_{\hat{K}})$ もわかるから、$\Delta^*(\alpha)$ の関数形が得られたことになる。

実はここから、凝縮局所密分解と飽和局所密分解が綺麗に対応することがわかる。

補題 26. 異なる $\alpha_1, \alpha_2 \in \mathbb{R}_{\geq 0}$ について $\Delta_*(\alpha_1) = \Delta_*(\alpha_2)$ かつ $\Delta^*(\alpha_1) = \Delta^*(\alpha_2)$ が成立するならば、前者を D、後者を $\hat{D}$ として $m_D(\hat{D}) = s_D(\hat{D}) = 0$ である。

証明. $\Delta_*(\alpha), \Delta^*(\alpha)$ の定義より、$\delta_\alpha(D) = \delta_\alpha(\hat{D})$ すなわち $m(D) - \alpha s(D) = m(\hat{D}) - \alpha s(\hat{D})$ が $\alpha = \alpha_1, \alpha_2$ で成立する。ゆえに、これは α についての恒等式である。つまり、$m(D) = m(\hat{D})$ かつ $s(D) = s(\hat{D})$ が成立する。 □

系 27. $\hat{K} = K$ であり、$i = 0, \ldots, K$ について $d_{D_i}(D_{i+1}) = \hat{d}_{\hat{D}_i}(\hat{D}_{i+1})$ および

$m_{D_i}(\hat{D}_i) = s_{D_i}(\hat{D}_i) = 0$ が成立する。

特に、m, s のいずれかが狭義単調であるとき、$i = 0, \ldots, K$ について $D_i = \hat{D}_i$ が成立し、それゆえ α が $d_{D_0}(D_1), \ldots, d_{D_{K-1}}(D_K)$ 以外のとき $\Delta_*(\alpha) = \Delta^*(\alpha)$ となる。この状況においては、凝縮局所密であることと飽和局所密であることは同値である。

さて、再び $\Delta_*(\alpha)$ と同様の性質を見ていこう。先ほどの凝縮に対応するものを飽和と呼ぶことにした。

定義 28. コンテナ A は任意の $Y \sqsupset A$ について $(m_A(Y), s_A(Y)) \neq (0, 0)$ を満たしているとき、**飽和**しているといわれる。

特に、飽和局所密なコンテナは飽和している。

なお、m, s のいずれかが狭義単調である場合はすべてのコンテナが飽和していることになる。

命題 29. $i = 1, \ldots, K$ に対し、$\hat{D}_{i-1}$ は $\hat{d}_X(\hat{D}_i)$ を最小化する飽和した $X \sqsubset \hat{D}_i$ のうちで束で最小のものである。

命題 30. $i = 0, \ldots, K-1$ に対し、$\hat{D}_{i+1}$ は $\hat{d}_{\hat{D}_i}(Y)$ を最大化する $Y \sqsupseteq \hat{D}_i$ のうちで束で最大のものである。

系 31. $\hat{D}_1$ は、密度最大のコンテナのうち、束で最大のものである。

命題 32. $0 \leq a < b \leq K$ が与えられているとき、$\alpha := d_{\hat{D}_a}(\hat{D}_b)$, $\hat{D}_c := \Delta^*(\alpha)$ とすると、$a + 1 < b$ ならば $a < c < b$ が、$a + 1 = b$ ならば $c = b$ が成立する。

3.2 質量とサイズに狭義単調凸・凹関数を被せた場合

局所密分解のために導入した制約を満たす質量関数とサイズ関数 m, s が与えられているとき、任意の狭義単調凸関数 $f : \mathbb{R} \to \mathbb{R}$ と狭義単調凹関数 $g : \mathbb{R} \to \mathbb{R}$（ただし $f(0) = g(0) = 0$ とする）について $m'(X) := f(m(X)),\ s'(X) := g(s(X))$ としたときに $m',\ s'$ も質量関数、サイズ関数として制約を満たしている。

m, s について $d_X(Y),\ \delta_\alpha(X),\ \Delta_*(X),\ D_i,\ K$ といった記法を導入するが、m', s' については $d'_X(Y),\ \delta'_\alpha(X),\ \Delta'_*(\alpha),\ D'_i,\ K'$ と書くことにしよう。

実は以下の性質が成り立つ。

命題 33. $\{D'_0, D'_1, \ldots, D'_{K'}\} \subseteq \{D_0, D_1, \ldots, D_K\}$。

証明. 任意の $\alpha \in \mathbb{R}_{\geq 0}$ を取り、$D' := \Delta'_*(\alpha)$ としよう。これが m, s について局所密であることを示せばよい。関数 $f(x)$ の $x = m(D')$ での接線（の一つ）を $ax + b$ としよう。また、関数 $g(x)$ の $x = s(D')$ での接線（の一つ）を $cx + d$ としよう。このとき、$a > 0$ に注意すると $\delta'_\alpha(X) = f(m(X)) - \alpha g(s(X)) \geq a \cdot m(X) + b - c\alpha \cdot s(X) - d\alpha = a\delta_{\frac{c}{a}\alpha}(X) + b - d\alpha$ であり、$X = D'$ のとき等号が成立するから、$D' = \Delta'_*(\alpha)$ を踏まえると D' は $\delta_{\frac{c}{a}\alpha}(X)$ を最大化する X の中で束で最小である、すなわち $D' = \Delta_*(\frac{c}{a}\alpha)$ である。ゆえに、D' は m, s について局所密である。 □

ゆえに、$D_0, \ldots, D_K$ が求まっている状況下で $D'_0, \ldots, D'_{K'}$ を求めるには、命題 16 に基づくと、以下の要領で求めればよい。

- $\top$ は m', s' についても局所密であるから、結果のリストに $\top$ を加え、$D' := \top$ とする。
- 以下を $D' = \bot$ となるまで繰り返す。

– $D_i \sqsubset D'$ なる D_i で、$d'_{D_i}(D')$ が一番小さくなるもののうち、束で最小になるものを新たな D' とし、これを結果のリストの先頭に加える。

- 最終的に得られるリストは、$D'_0, \ldots, D'_{K'}$ のリストと一致する。

4　局所密分解の厳密アルゴリズム

Tatti & Gionis (2015) で述べられているとおり、命題 17 を使うと、局所密分解について以下のアルゴリズムが考えられる。

ただし、計算量の改善のため $\Delta_*(\alpha, A, B) := \mathrm{argmax}_{A \sqsubseteq X \sqsubseteq B}\, \delta_\alpha(X)$（この argmax は同点のものについて束で最小のものを取る）とした。ここを $\Delta_*(\alpha)$ としても以下のアルゴリズムで得られる X は変化しない。

アルゴリズム 1. ExactLD および ExactLDBody

保証: $\bot, \top$ を除くすべての局所密なコンテナを出力する。

function ExactLD()

  ExactLDBody($\bot, \top$);

end function

制約: $A \sqsubset B$ であり、A, B は局所密である。

保証: $A \sqsubset X \sqsubset B$ なるすべての局所密な X を束で小さい順に一度ずつ出力し、停止する。

function ExactLDBody(A, B)

  $\alpha \leftarrow d_A(B)$;

  $C \leftarrow \Delta_*(\alpha, A, B)$;

  if $A \sqsubset C$ **then**

    ExactLDBody(A, C);

    output C;

    ExactLDBody(C, B);

  end if

end function

さて、問題は、$\Delta_*(\alpha, A, B)$ を効率良く計算できるかどうかである。なお、$A \sqsubset B$, $\alpha > 0$, $\Delta_*(\alpha, A, B) = \Delta_*(\alpha)$ の場合のみ考えればよい。また、計算量解析において、ExactLDBody の呼び出しの最大深さを L としよう。

まず、定義域が有限集合の冪集合である場合の一般論について述べ、次に具体例（辺の密度、パターンの密度）について説明する。

4.1 定義域が有限集合の冪集合である場合の一般論

$\Delta_*(\alpha, A, B) = \operatorname{argmax}_{A \sqsubseteq X \sqsubseteq B} \delta_\alpha(X) = \operatorname{argmin}_{A \sqsubseteq X \sqsubseteq B}(-\delta_\alpha(X))$ であるから、これは（$A \sqsubseteq X \sqsubseteq B$ の部分束上の）劣モジュラ関数の最小化問題に帰着される。

劣モジュラ関数の始域が有限集合 U の冪集合のなす束である場合、劣モジュラ関数の計算量と $|U|$ についての多項式時間で劣モジュラ関数の最小化問題が解けることが知られている（Iwata et al. (2001)）。ゆえに、$\mathcal{L}$ が U の冪集合として考えられる場合、ExactLDBody の呼び出し回数は $O(|U|)$ であるから、ExactLD が全体で $\delta_\alpha(X)$ の計算時間と $|U|$ についての多項式時間で動作することが保証される。

4.2 辺の密度

これは Tatti & Gionis (2015) で述べられている内容と概ね同じである。ただし、計算量解析だけはより詳しくおこなっている。

頂点に正の費用、辺に正の重みがついている有限無向グラフ $G = (V, E)$ を考える。ただし、次数が 0 の頂点は存在しないとする。頂点の費用関数を c、辺の重み関数を w とする。頂点の集合 $W \subseteq V$、辺の集合 $F \subseteq E$ に対しても費用、重みの総和として $c(W)$, $w(F)$ を使うことにする。$E(X, Y) := \{ (u, v) \in E \mid u \in X, v \in Y \}$, $E(X) := E(X, X)$, $\deg_w(v; X) := w(\{ (u, v) \in E \mid u \in X \})$ と定義する。

さて、$\mathcal{L}$ は V の冪集合として、$m(X) := w(E(X))$, $s(X) := c(X)$ としよう。これは局所密分解ができる状況である。

このとき、$X \subseteq B$ のもとで $\delta_\alpha(X)$ の最大化は $2w(E(B)) - 2\delta_\alpha(X) = 2w(E(B)) - 2w(E(X)) + 2\alpha c(X) = w(E(X, B - X)) + 2\alpha c(X) + \sum_{v \in B - X} \deg_w(v; B)$ の最小化

と同値であるから、以下のようにして最小カット問題に帰着できる。[*11]

アルゴリズム **2.** 辺の密度に対する $\Delta_*(\alpha, A, B)$（改善前）

制約: $A \subset B \subseteq V,\ \alpha > 0,\ \Delta_*(\alpha, A, B) = \Delta_*(\alpha)$。

function $\Delta_*(\alpha, A, B)$

辺に非負の容量がついた有向グラフ $G' = (\{s, t\} + B,\ E')$ で、辺集合と各辺の容量が

- $E(B)$ の要素に対応する辺（容量は w を引き継ぐ）。
- s と各 $v \in B$ の間の容量 $\deg_w(v; B)$ の辺。
- t と各 $v \in B$ の間の容量 $2\alpha c(v)$ の辺。

により与えられるものを構成する。

G' において（s を始点、t を終点とする）最小カット問題を解き、最小カットを与える（s 側の）点集合で包含関係で最大のもの $C \subseteq B$ を出力する（そのためには最大フローを構成して、残余グラフで s から辿り着ける頂点全体の集合を C とすればよい）。

end function

最小カットを与える集合として A を含むものしか考えないから、計算量を削減するために以下のような工夫ができる。[*12] 以下、$B - A$ を B_A と略記することにする。

---

*11 この帰着は Goldberg (1984) によるものである。

*12 この改善法は Tatti & Gionis (2015) の手法である。

アルゴリズム 3. 辺の密度に対する $\Delta_*(\alpha, A, B)$（改善後）

制約: $A \subset B \subseteq V,\ \alpha > 0,\ \Delta_*(\alpha, A, B) = \Delta_*(\alpha)$。

function $\Delta_*(\alpha, A, B)$

辺に非負の容量がついた有向グラフ $G' = (\{s, t\} + B_A,\ E')$ で、辺集合と各辺の容量が

- $E(B_A)$ の要素に対応する辺（容量は w を引き継ぐ）。
- s と各 $v \in B_A$ の間の容量 $\deg_w(v; A) + \deg_w(v; B)$ の辺。
- t と各 $v \in B_A$ の間の容量 $2\alpha c(v)$ の辺。

により与えられるものを構成する。

G' において最小カット問題を解き、最小カットを与える（s 側の）点集合で包含関係で最小のものを $C' \subseteq B_A$ として、$A + C'$ を出力する。

end function

G' は頂点数 $O(|B_A|)$、辺数 $O(|E(B_A)| + |B_A|)$ である。一般にグラフの最大フロー最小カットは Orlin (2013) より $O((頂点数)(辺数))$ で解けることが知られているから、ここでの最大フロー最小カットは $O\big(|B_A|(|E(B_A)| + |B_A|)\big)$ で動く。

全体で効率的に計算するには、ExactLDBody における $d_A(B)$ の計算と、$\Delta_*(\alpha, A, B)$ における $\deg_w(v; A) + \deg_w(v; B)$ の計算を工夫することが重要である。具体的に言えば、常に $\deg_w(v; A)$ と $\deg_w(v; B)$ を持ち回り、A が C に、あるいは B が C に変わったときに差分を上手く計算するようにする。そうすると、ExactLD 全体では $O\big(L|V|(|E| + |V|)\big)$ で動くことが保証される。[*13] 現実のグラフでは最大フロー最小カットは理論保証から期待される速度よりもずっと速く動作することが多い。こうした考察から、大規模なグラフに対しても通常

[*13] Tatti & Gionis (2015) では L というパラメータを導入しておらず、K もオーダー表記には用いていないので、$O\big(|V|^2(|E| + |V|)\big)$ で押さえられるという控えめな解析となっている。

は現実的な速さで動作すると考えられ、実際そのような実験結果が得られている。詳しくは Tatti & Gionis (2015) の表 1、あるいはここでの実験結果の表 1 を参照してほしい。

4.3 パターンの密度

辺の密度の場合を一般化したものである。

点の有限多重集合 V を取る。各点 $x \in V$ に対し、x の V での重複度を n として、$i = 1, \ldots, n$ について x を i 個得たときの正の費用 $c(x^i)$ が与えられている。$c(x^i)$ は i について単調であるとする。これをもとに、一般に点の多重集合 $W \subseteq V$ に対して各種類の要素の費用の総和として $c(W)$ を考える。また、**パターン**の有限集合 Π があり、各パターン π には任意サイズの点の多重集合 $P(\pi) \subseteq V$ と正の重み $w(\pi)$ が紐付けられている。ただし、$V = \bigcup_{\pi \in \Pi} P(\pi)$ が成立するとする。$X \subseteq V$ に対して $\Pi(X) := \{\, \pi \in \Pi \mid P(\pi) \subseteq X \,\}$ と書くことにしよう。パターンの集合 $\Xi \subseteq \Pi$ に対しても重みの総和として $w(\Xi)$ を使うことにする。

ここで、$\mathcal{L}$ は V の冪集合として、$X \subseteq V$ に対し $m(X) := w(\Pi(X)),\ s(X) := c(X)$ と定義したときを考えよう。これは局所密分解ができる状況である。

以下の方法で多重集合 X を同じサイズの集合 $X^\dagger$ に変換することにしよう：$x \in X$ の多重度が n であるとき、$X^\dagger$ には $x^1, \ldots, x^n$ を入れる。

$X \subseteq B$ のもとで、$\delta_\alpha(X)$ の最大化は $m(B) - \delta_\alpha(X) = m_X(B) + \alpha c(X)$ の最小化と同値であるので、以下のようにして最小カットに帰着できる。[*14]

[*14] この帰着は Tsourakakis (2015) のアルゴリズム 6（k-クリークの密度についての手法）を参考にして考案したが、s から各 π への辺を設けるところが異なる。また、多重集合についての扱いを独自に加えた。

1 点あるいは 2 点のパターンについてはパターンに対応する頂点を設けない手法がある（辺の密度が 2 点からなるパターンの密度であることに注意）。ただ、ここでそれを明示的に書くと記述が複雑になるので、省略した。

アルゴリズム 4. パターンの密度に対する $\Delta_*(\alpha, A, B)$（改善前）

制約: $A \subset B \subseteq V,\ \alpha > 0,\ \Delta_*(\alpha, A, B) = \Delta_*(\alpha)$。

function $\Delta_*(\alpha, A, B)$

辺に正の容量がついた有向グラフ $G = (\{s, t\} + B^\dagger + \Pi(B), E)$ で、辺集合と各辺の容量が

- 各 $\pi \in \Pi(B)$ に対して、π から $(P(\pi))^\dagger$ の各点への容量無限大の辺と、s から π への容量 $w(\pi)$ の辺。
- 各 $x^i, x^{i+1} \in B^\dagger$ について、x^{i+1} から x^i への容量無限大の辺。
- t から各 $x^i \in B^\dagger$ への容量 $\alpha c(x^i)$ の辺。

により与えられるものを構成する。

G において最小カット問題を解き、最小カットを与える点集合のうち包含関係で最小のものを C' として、$C' \cap B$ を出力する。

end function

最小カットにおいては各 π について $p \in P(\pi)$ がみな s 側にあることと π が s 側にあることが同値である、ということが重要である。

最小カットを与える集合として A を含むものしか考えないから、計算量を削減するために、辺の密度の場合と同様、以下のような工夫ができる。*15

ただし、$B - A$ を B_A と略記し、$\Pi_X(Y) := \Pi(Y) - \Pi(X)$ とした。

*15 先程の Tatti & Gionis (2015) の改善法を参考にして自分で考えた。

アルゴリズム 5. パターンの密度に対する $\Delta_*(\alpha, A, B)$（改善後）

制約: $A \subset B \subseteq V,\ \alpha > 0,\ \Delta_*(\alpha, A, B) = \Delta_*(\alpha)$。

function $\Delta_*(\alpha, A, B)$

辺に正の容量がついた有向グラフ $G = (\{s, t\} + (B_A)^\dagger + \Pi_A(B),\ E)$ で、辺集合と各辺の容量が

- 各 $\pi \in \Pi_A(B)$ に対して、π から $(P(\pi) \setminus A)^\dagger$ の各点への容量無限大の辺と、s から π への容量 $w(\pi)$ の辺。
- 各 $x^i, x^{i+1} \in (B_A)^\dagger$ について、x^{i+1} から x^i への容量無限大の辺。
- t から各 $x^i \in (B_A)^\dagger$ への容量 $\alpha c(x)$ の辺。

により与えられるものを構成する。

G において最小カット問題を解き、最小カットを与える点集合のうち包含関係で最小のものを C' として、$A + (C' \cap B_A)$ を出力する。

end function

$\Pi_X^\dagger(Y) := \{\, (\pi, p) \mid \pi \in \Pi_X(Y),\ p \in (P(\pi))^\dagger \,\}$ とすると、ここでの G は頂点数 $O(|\Pi_A(B)| + |B_A|)$、辺数 $O(|\Pi_A^\dagger(B)| + |B_A|)$ のグラフであるから、最小カットは $O\big(|\Pi_A(B)| + (|B_A|)(|\Pi_A^\dagger(B)| + |B_A|)\big)$ で動くことがわかる。

効率的に計算するには、ExactLDBody における $d_A(B)$ の計算を工夫すれば良い。そうすると、ExactLD 全体では $O\big(L(|\Pi| + |V|)(|\Pi^\dagger| + |V|)\big)$ で動くことが保証される（$\Pi^\dagger := \Pi_\varnothing^\dagger(V)$ とした）。

5 実験

無向グラフの辺の密度について局所密分解を行うアルゴリズムを C++ で書いた。様々なグラフについて L の大きさを確かめることが主な目的である。最大フロー最小カット部分は Dinic のアルゴリズムで書いているが、それなりに速

く動作する。ソースコードは src というフォルダに置いた。

Stanford Large Network Dataset Collection から取ったグラフのデータについて東京大学理学部情報科学科の CSC クラスタ（CPU は Intel® Xeon® CPU E5-2687W v3 3.10GHz である）において実験をおこなった。出力は out というフォルダに置いた。データ本体はファイルサイズが大きすぎるので置いていない。

実験の結果は以下の通り。規模の大きいグラフに対しても現実的な時間で動いていることがわかる。また、L はかなり小さいということがわかる。

表 1: 実験結果

グラフ名	$\|V\|$	$\|E\|$	K	L	実行時間
ca-GrQc	5,247	28,980	108	10	0.149 s
com-amazon	334,863	925,872	1,256	14	44.715 s
com-youtube	1,134,890	2,987,624	918	14	50.535 s
as-skitter	1,696,415	11,095,298	3,501	16	362.862 s
com-lj	3,997,962	34,681,189	8,122	17	3170.55 s

参考文献

Goldberg, A. V. (1984), ‘Finding a maximum density subgraph’, *University of California at Berkeley Technical Report* .

Iwata, S., Fleischer, L. & Fujishige, S. (2001), ‘A combinatorial strongly polynomial algorithm for minimizing submodular functions’, *Journal of the ACM* .

Orlin, J. B. (2013), ‘Max flows in $O(nm)$ time, or better’, *STOC ’13 Proceedings of* *the 45th Annual ACM Symposium on Theory of Computing* .

Tatti, N. & Gionis, A. (2015), 'Density-friendly graph decomposition', *WWW '15 Proceedings of the 24th International Conference on World Wide Web* .

Tsourakakis, C. E. (2015), 'The k-clique densest subgraph problem', *WWW '15 Proceedings of the 24th International Conference on World Wide Web* .