

ランダム複体とパーシステントホモロジー

九州大学・マス・フォア・インダストリ研究所 白井朋之
Tomoyuki Shirai
Institute of Mathematics for Industry, Kyushu University

1 はじめに

Erdős-Rényi グラフやスケールフリーネットワークなどのランダムグラフはその重要性から理論・応用の両面で多くの研究がある。このことは二項関係を前提とするグラフ上のネットワーク研究の観点からは自然である。また多項関係を扱うものにハイパーグラフや単体複体があるが、ハイパーグラフは単体複体に比べればグラフの感覚に近いこともあり、ランダムハイパーグラフについては早い時期からランダムグラフとの類似の考察がなされている。一方、Erdős-Rényi グラフの連結性に関する閾値の問題が 1959 年に Erdős-Rényi [4] で既に議論されているのに対して、同様の問題がランダム複体に対して議論されて結果が得られたのは、講演者の知る限り 2006 年の Linial-Meshulam [14] が始めではないかと思う。

また、2000 年頃より位相的データ解析の立場からパーシステントホモロジーが研究され始めて、タンパク質の構造解析、画像解析、材料科学など種々の分野に応用されている。パーシステントホモロジーは通常ホモロジー論に時間軸を取り入れたものとも理解できて、ホモロジー類の生成消滅時刻の情報を考察することができる (cf. [3, 7, 10, 18]).

本稿では、ランダムな重みをもつ完全グラフ上の最小全域木に関する問題から始めて、全域木の高次元版である全域非輪体の定義、パーシステントホモロジーの生存時間とベッチ数の積・最小全域非輪体の関係、Kruskal-Katona の定理などについて述べた後、Erdős-Rényi ランダムグラフ過程の複体版 (の一つ) である Linial-Meshulam 複体過程のパーシステントホモロジー (または同値であるが最小全域非輪体の重み) について [8] で得られた結果について議論する。

1.1 最小全域木と Kruskal のアルゴリズム

まず、最小全域木について復習しておく。 $G = (V, E)$ を連結グラフ、 $w : E \rightarrow \mathbb{R}$ を辺の重みとして (G, w) を重みつきグラフという。最小全域木 (minimum spanning tree) とは重みつきグラフ (G, w) の全域木 T の中で重みの和 (以後、単に T の重みという)

$$\text{wt}(T) := \sum_{e \in T} w(e)$$

が最小となるものをいう。 $W(G) := \min_{T \in \mathcal{S}^{(1)}(G)} \text{wt}(T)$ は最小全域木の重みである。ただし、 $\mathcal{S}^{(1)}(G)$ は G の全域木の全体集合である。

RIMS 研究集会「デザイン、符号、グラフおよびその周辺」@京大 RIMS on July 8-10, 2015.
本研究は科研費挑戦的萌芽研究 No.26610025, 基盤研究 (B)No.26287019 の助成を受けたものです。
本稿は平岡裕章氏 (東北大 AIMR) との共同研究に基づく。

注意 1.1. Cayley の定理 (1859) より完全グラフ K_n 内の全域木の総数は $|S^{(1)}(K_n)| = n^{n-2}$ となることが知られている。

与えられた重みつき連結グラフ (G, w) に対して, 最小全域木を見つけるアルゴリズムはいくつか知られている。ここでは, 後との関連で Kruskal のアルゴリズム [13] について述べる。簡単のために w の値はすべて違うものとする。

- 辺の重みの小さいものから順に辺を並べて e_1, e_2, \dots, e_m ($m = |E|$) とおく。辺の部分集合列 $\{T_k\}_{k=0}^m$ を以下のルールで構成する。

$$T_0 = \emptyset,$$

$$T_k = \begin{cases} T_{k-1} \sqcup \{e_k\}, & T_{k-1} \sqcup \{e_k\} \text{ がサイクルをもたない,} \\ T_{k-1}, & T_{k-1} \sqcup \{e_k\} \text{ がサイクルをもつ.} \end{cases}$$

- 定義より $\{T_k\}_{k \geq 0}$ は単調非減少列で, ある $k_0 \leq m$ に対して $|T_{k_0}| = |V| - 1$ となりそれ以後は $T_k = T_{k_0}$ ($k \geq k_0$) となる。この T_{k_0} が最小全域木を与える。

1.2 Frieze の定理

$K_n = (V_n, E_n)$ を n 点上の完全グラフとする。 K_n の各辺 $e \in E_n$ に独立に確率変数 $w(e)$ を与える。 $\{w(e)\}_{e \in E_n}$ はすべて $[0, 1]$ 上の一様分布に従うとする。 $w : E_n \rightarrow [0, 1]$ を重みとみなすと, (K_n, w) は重みつき完全グラフとなる。一様分布が連続分布であることより, $\{w(e)\}_{e \in E_n}$ の値は確率 1 ですべて異なるので, 前節のアルゴリズムはそのまま適用できる。この (ランダム) 重みつき完全グラフに対する最小全域木の重みを $W_n := W(K_n)$ とおくと確率変数となる。Frieze は確率変数 W_n について以下のことを示した。

定理 1.2 (Frieze [6]). $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ は Riemann のゼータ関数とする。このとき,

$$\mathbb{E}[W_n] \rightarrow \zeta(3) = 1.20206 \dots$$

注意 1.3. 例えば, 完全グラフをある程度一般のグラフへ拡張, 一様分布の一般の連続分布への拡張, 漸近展開などこの定理は種々の方向に一般化されている (cf. [1]).

注意 1.4. Janson ([12]) は中心極限定理を示した。

$$\sqrt{n}(W_n - \zeta(3)) \rightarrow N(0, \sigma^2)$$

ただし, $\sigma^2 = 6\zeta(4) - 4\zeta(3) = 1.68571 \dots$

本稿の目標は Frieze の定理の高次元への拡張を議論することである。

1.3 Erdős-Rényi ランダムグラフ過程と Kruskal アルゴリズム

Frieze の定理の設定を見直してみよう. 完全グラフ $K_n = (V_n, E_n)$ にランダムな重み $\{w(e)\}_{e \in E_n}$ が与えられたときに, これに Kruskal のアルゴリズムを適用することを考える. ランダムな重みを辺の生成時間とみなして, 時刻 $t = w(e)$ において辺 e を加える操作を試みる. もしも辺 e を加えてサイクルができなければそれを辺部分集合としてつけくわえる. これを繰り返していくと最終的に最小全域木が得られるのであった. 一方, 時刻 $w(e)$ に従って辺を加えていく (開いていく) 操作は

$$E_n(t) = \{e \in E_n : w(e) \leq t\}, \quad G_n(t) = (V_n, E_n(t))$$

とあらわすことができる. $\mathbb{G} = \{G_n(t)\}_{0 \leq t \leq 1}$ はランダムグラフの増大列を与えるが, これが Erdős-Rényi ランダムグラフ過程とよばれる (右連続) 確率過程である. $G_n(t)$ は n 個の孤立点からなるグラフ $G_n(0)$ からスタートして, 最終的に完全グラフ $G_n(1) = K_n$ でおわる. 各時間 t における $G_n(t)$ は完全グラフの各辺を独立に確率 t で残し, 確率 $1-t$ で取り除いてできるいわゆる Erdős-Rényi グラフ $G(n, t)$ と一致する. 通常は辺の出現確率を p として Erdős-Rényi グラフは $G(n, p)$ とあらわされる. ここでは, 出現確率 p を時間 t とみなしていることに相当する.

ランダムな重みをもつ重みつき完全グラフ (K_n, w) に Kruskal のアルゴリズムを適用することは, Erdős-Rényi ランダムグラフ過程を走らせておいて, サイクルを形成する辺を無視して最終的に残る全域木を取り出すことに対応する. ランダムグラフ $G(n, p)$ の問題を考える際には, ランダムグラフ $G(n, p)$ をすべて集めた (カップリングした) 確率過程 $\{G(n, p)\}_{0 \leq p \leq 1}$ を考えるとよいことがあるが, 上で構成した $\mathbb{G} = \{G_n(t)\}_{0 \leq t \leq 1}$ は「増加確率過程」として自然にそのカップリングが実現されている.

1.4 パーシステントホモロジー

次節でパーシステントホモロジーの定義は述べるが, パーシステントホモロジーは時間軸の入ったホモロジー論と言える. 単体複体のホモロジー理論は一つの単体複体の位相的な性質を調べる理論であるが, パーシステントホモロジーは単体複体の増加列 (フィルトレーション) に対して位相的な性質の変化まで調べる理論として定式化されている. パーシステントホモロジーは大雑把には以下のような図式で得られ, フィルトレーションの位相的な情報がパーシステント図として視覚化される.

- input 1: 点配置データ
- ↓ via Čech or Rips-Vietoris 複体 etc.
- input 2: フィルトレーション = 単体複体の増加列
- ↓
- output: パーシステント図

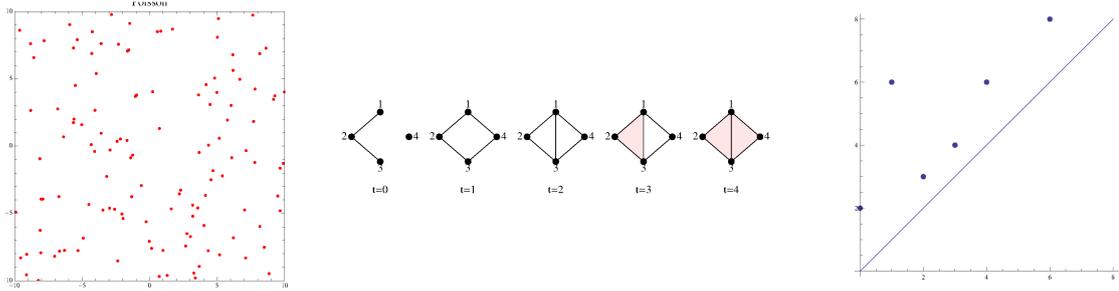


図 1: 点配置データ, フィルトレーション, パーシステント図.

input としては点配置データ (input1) としてもよいし, フィルトレーション (input2) から始めてもよい. input1 からスタートする場合には何かしらの手続き (例えば, Čech 複体や Rips-Vietris 複体をとる) でフィルトレーションを構成すればよい.

本稿ではフィルトレーションからスタートすることを考える. 特にランダムなフィルトレーションから始めると結果としてランダムなパーシステント図が得られる. これは点過程とみなすのが自然である ([8]). こうして得られる点過程の極限定理も興味深い問題であるが, 本稿ではパーシステント図の定める点過程のある汎関数 (生存時間の和) の極限定理を考える.

2 単体複体とパーシステントホモロジー

2.1 単体複体

V を有限集合とするとき, $\sigma \subset V$ が $|\sigma| = k+1$ のとき, σ は k -単体 (k -simplex) もしくは k -face という. このとき, $\dim(\sigma) = k$ とあらわす.

定義 2.1. X を V の非空部分集合の族とする. (V, X) が抽象的単体複体であるとは条件

(1) $\{v\} \subset X$ for all $v \in V$, (2) $\sigma \in X, \tau (\neq \emptyset) \subset \sigma \implies \tau \in X$

をみたすときをいう. 条件 (2) は「 X は非空の部分集合を取る操作に関して閉じていること」を意味する. $d = \max_{\sigma \in X} \dim(\sigma)$ のとき (V, X) を d 次元単体複体という.

以下, しばしば V を省略して X を単体複体という.

例 2.2. 2次元単体複体 $X = \{1, 2, 3, 4, 12, 13, 14, 23, 34, 123\}$ を考える.

X は facet $\{14, 34, 123\}$ (包含関係に関する極大な単体) で定まる. 実際, 定義 2.1(2) の条件をみたすように facet の非空部分集合をすべて取ればよい.

$$X = \underbrace{\{1, 2, 3, 4\}}_{X_0}, \underbrace{\{12, 13, 14, 23, 34\}}_{X_1}, \underbrace{\{123\}}_{X_2}$$

$X^{(1)}$: a graph

ここで, $X_k = \{\sigma \in X : \dim(\sigma) = k\}$ は X の k -face の全体とし, $X^{(k)} = \{\sigma \in X : \dim(\sigma) \leq k\}$ を X の k -スケルトンという.

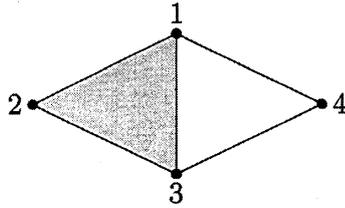


図 2: 2次元単体複体の例

注意 2.3. 1次元単体複体はグラフであるから, 1-スケルトンは単体複体のベースになっているグラフである.

定義 2.4. X が $V = [n] = \{1, 2, \dots, n\}$ 上の d 次元完全単体複体であるとは, 次元が d 以下のすべての単体が含まれているときをいい, $K_n^{(d)}$ とあらわす. 特に 1次元完全複体 $K_n^{(1)}$ は完全グラフ K_n に対応する.

$k = 1, 2, \dots, \dim X$ とする. $T \subset X_k$ に対して X の k 次元部分複体を

$$X_T := X^{(k-1)} \sqcup T$$

と定義する. ($T = \emptyset$ のときは $k-1$ 次元部分複体.)

2.2 全域木の概念の単体複体への拡張

グラフ $G = (V, E)$ における全域木は以下の性質をもつ辺部分集合 $T \subset E$ として特徴つけられる.

1. G の部分グラフ (V, T) は連結.
2. G の部分グラフ (V, T) はサイクルをもたない.
3. $|T| = |V| - 1$.

注意 2.5. (i) 条件 1 と 2 はそれぞれ 0 次と 1 次のホモロジーが自明であることを意味する.
(ii) 3 条件のうち 2 つが成立すれば残りも自動的に成立する.

単体複体における全域木の対応物は様々な定義が知られているので, 比較のためにいくつかを紹介する. いずれも上の全域木の特徴付けに近い形で定義されるが微妙に違いもある.

以下, $\tilde{H}_*(X)$ は \mathbb{Z} -係数の被約ホモロジー群として, $\tilde{\beta}_*(X) = \text{rank } \tilde{H}_*(X)$ とする. $\tilde{\beta}_*(X)$ は \mathbb{Q} -係数の被約ホモロジー群 $\tilde{H}_*(X, \mathbb{Q})$ の次元に等しい.

定義 2.6 ([2]). X を d 次元単体複体とし $(d-1)$ - \mathbb{Q} -acyclic であると仮定する. つまり, $\tilde{\beta}_i(X) = 0$ ($i = 0, 1, \dots, d-1$). X の k 次元部分複体 $T \subset X$ ($k = 1, \dots, d$) が以下の条件をみたすとき k 次元 simplicial spanning tree であるという.

1. $\mathcal{T}^{(k-1)} = X^{(k-1)}$.
2. $\tilde{H}_k(\mathcal{T}) = 0$.
3. $\tilde{\beta}_{k-1}(\mathcal{T}) = 0$.
4. $|\mathcal{T}_k| = |X_k| - \tilde{\beta}_k(X) + \tilde{\beta}_{k-1}(X)$.

考え方はほぼ上の 2.6 と同様であるが, k 次元の部分のみに着目して簡単に以下のような定義も考えられる. 以下の定義は G. Kalai [11] が完全単体複体について定義したものに相当する.

定義 2.7 ([8]). X は $\tilde{\beta}_{k-1}(X^{(k)}) = \tilde{\beta}_{k-2}(X^{(k-1)}) = 0$ をみたすとする. $T \subset X_k$ が以下の条件をみたすとき k 次元全域非輪体 (k -spanning acycle) であるという.

1. $\tilde{H}_k(X_T) = 0$.
2. $|\tilde{H}_{k-1}(X_T)| < \infty$.

注意 2.8. X が $\tilde{\beta}_{k-1}(X^{(k)}) = \tilde{\beta}_{k-2}(X^{(k-1)}) = 0$ をみたさないときは, 上の 2 条件をみたす T は存在しないこともある. 例えば, 球面の三角形分割からなる 2 次元単体複体においては, 2-単体をひとつだけ除いたものが 2 次元全域非輪体になるが, 種数が 1 以上の向きつけられた閉曲面の三角形分割には定義 2.7 の意味での 2 次元全域非輪体は存在しない.

上の注意に述べたような状況を回避するためには例えば以下のような定義もあり得る. 以下はマトロイドの枠組みで考えており, 階数最大の部分集合 (base) に対応するものである.

定義 2.9 ([16]). X を k -次元単体複体とする. $T \subset X_k$ が以下の条件をみたすとき k -base であるという.

1. $\tilde{H}_k(X_T) = 0$.
2. T は次の意味で極大. 任意の $\sigma \in X_k \setminus T$ に対して $\tilde{H}_k(X_{T \cup \{\sigma\}}) = 1$.

注意 2.10. 境界作用素 $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$ の行列表現のランク最大のベクトルの組に対応する k -単体の集合 T が上の定義の k -base に対応してその全体はマトロイドをなす.

以下では, 全域非輪体の定義は定義 2.7 に従う. 完全単体複体 $K_n^{(n-1)}$ は可縮なので $\tilde{\beta}_{k-1}(X^{(k)}) = \tilde{\beta}_{k-2}(X^{(k-1)}) = 0$ などの条件は自動的にみたしている.

上にも述べたように d 次元全域非輪体の種々の定義の元になる定義は完全単体複体 $K_n^{(n-1)}$ の場合に Kalai[11] によって与えられ, 注意 1.1 で述べた Cayley の定理を以下の形に拡張している.

定理 2.11 ([11]). $\mathcal{S}_n^{(d)}$ を n 点上の $(n-1)$ 次元完全単体複体 $X := K_n^{(n-1)}$ の d 次元全域非輪体の全体のなす集合とする. このとき

$$\sum_{T \in \mathcal{S}_n^{(d)}} |\tilde{H}_{d-1}(X_T)|^2 = n^{\binom{n-2}{d}}$$

が成立する.

注意 2.12. 定義 2.7 より全域非輪体 T に対して $|\tilde{H}_{d-1}(X_T)| < \infty$ であり, ねじれ群の位数をあらわす. 特に $d = 1$ の場合は任意の全域木 T に対して $|\tilde{H}_{d-1}(X_T)| = 1$ であるので左辺は $|S_n^{(1)}|$ をあらわし, 注意 1.1 で述べた Cayley の定理を意味する.

例 2.13. $d = 2$ のときを考える. $n = 4, 5$ のとき $\tilde{H}_1(T) = 0$ である. よって,

$$|S_4^{(2)}| = 4^{\binom{4-2}{2}} = 4, \quad |S_5^{(2)}| = 5^{\binom{5-2}{2}} = 125$$

$n = 6$ のときは 12 個の 2 次元全域非輪体に対して $\tilde{H}_1(T) \cong \mathbb{Z}_2$, つまり $|\tilde{H}_1(T)|^2 = 4$ となるので,

$$|S_6^{(2)}| = 46620 \neq 6^{\binom{6-2}{2}} = 6^6 = 46656$$

となる. 例えば, 下図で与えられる射影平面の三角形分割はそのような 2 次元全域非輪体の一つの例を与える.

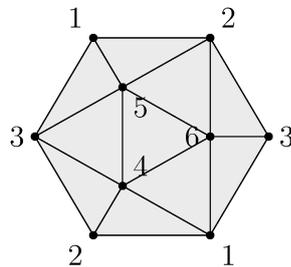


図 3: 射影平面の三角形分割

2.3 パーシステントホモロジー

単体複体 (V, X) の部分複体の増大列 $\mathbb{X} = (X(t))_{t \geq 0}$ もしくは $\mathbb{X} = (X(t))_{t=0,1,2,\dots}$ をフィルトレーションという. 前者の連続パラメータの場合は右連続性を仮定する. つまり, $s < t$ のとき $X(s) \subset X(t)$ で, $X(t+) = \bigcap_{t < u} X(u) = X(t)$ が成り立つとする. また, ある T が存在して $X(t) = X(T) (\forall t \geq T)$ を仮定する. そのような初めての時刻を飽和時刻という. t が連続パラメータの場合は連続時間, t が離散パラメータの場合は離散時間という.

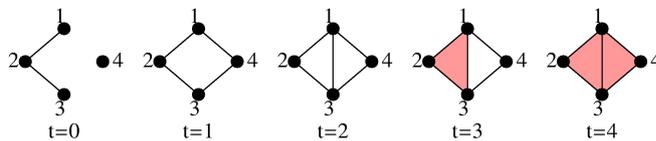


図 4: 離散時間フィルトレーションの例

定義 2.14 (生成時間 (birth time)). フィルトレーション $\mathbb{X} = (X(t))_{t \in \mathbb{T}}$ が与えられたとき,

$$t(\sigma) = t \quad \text{if} \quad \begin{cases} \sigma \in X(t) \setminus X(t-1) & (X(-1) = \emptyset), \quad \mathbb{T} = \{0, 1, 2, \dots\} \\ \sigma \in X(t+) \setminus X(t) & \mathbb{T} = [0, \infty) \end{cases}$$

を単体 σ の生成時間という.

単体複体 (V, X) の部分複体のフィルトレーション $\mathbb{X} = (X(t))_{t=0,1,2,\dots}$ が与えられたとする. ここでは, 簡単のために離散時間フィルトレーションを考えることにする. 通常ホモロジー群の定義と同様に, 鎖群, 境界準同型, パーシステントホモロジー群という順番で定義していく. パーシステントホモロジーに関する詳しい内容は [10](離散時間フィルトレーションの場合), 連続時間フィルトレーションの場合の定義などについては [8] を参照のこと.

定義 2.15 (k -鎖群 $C_k(\mathbb{X})$). \mathbb{F} は体とする. フィルトレーション \mathbb{X} に対する k -鎖群を

$$C_k(\mathbb{X}) := \bigoplus_{t=0}^{\infty} C_k(X(t)), \quad C_k(X(t)) := \left\{ \sum_{\sigma \in X(t)_k} a_{\sigma} \langle \sigma \rangle : a_{\sigma} \in \mathbb{F} \right\}$$

とする. ここで, $\langle \sigma \rangle$ は向き付けられた単体. \mathbb{F} -係数多項式環 $\mathbb{F}[x]$ の $C_k(\mathbb{X})$ への作用は

$$x \cdot (c_0, c_1, c_2, \dots) = (0, c_0, c_1, c_2, \dots)$$

を自然に線形に拡張して定義する. $x \cdot C_k(X(t)) \subset C_k(X(t+1))$ となるので, $C_k(\mathbb{X})$ は次数付き $\mathbb{F}[x]$ -加群となる. x の作用は時間 1 のシフトに対応する.

写像 $\iota_t : C_k(X(t)) \rightarrow C_k(\mathbb{X})$ を $\langle \sigma \rangle \mapsto (0, 0, \dots, 0, \overbrace{\langle \sigma \rangle}^t, 0, 0, \dots)$ と定義する. この写像を用いて

$$\Xi_k := \left\{ \langle \langle \sigma \rangle \rangle := \iota_{t(\sigma)} \langle \sigma \rangle, \sigma \in \mathbb{X}_k := \bigcup_{t \geq 0} (X(t))_k \right\}$$

と定義すると, Ξ_k は $C_k(\mathbb{X})$ の次数付き $\mathbb{F}[x]$ -加群としての基底を与える. 例えば, 図 4 の例では $C_1(\mathbb{X})$ の基底は

$$\begin{aligned} \langle \langle 12 \rangle \rangle &= (\langle 12 \rangle, 0, 0, 0, 0, \dots), & \langle \langle 13 \rangle \rangle &= (0, 0, \langle 13 \rangle, 0, 0, \dots), \\ \langle \langle 14 \rangle \rangle &= (0, \langle 14 \rangle, 0, 0, 0, \dots), & \langle \langle 23 \rangle \rangle &= (\langle 23 \rangle, 0, 0, 0, 0, \dots), \\ \langle \langle 34 \rangle \rangle &= (0, \langle 34 \rangle, 0, 0, 0, \dots) \end{aligned}$$

で与えられ, $\text{rank } C_1(\mathbb{X}) = 5$ である.

定義 2.16 (境界作用素). $\partial_k(x) : C_k(\mathbb{X}) \rightarrow C_{k-1}(\mathbb{X})$ は以下のように定義される.

$$\partial_k(x) \langle \langle \sigma \rangle \rangle := \sum_{j=0}^k (-1)^j x^{t(\sigma) - t(\sigma_j)} \langle \langle \sigma_j \rangle \rangle.$$

σ_j は $\sigma = \langle v_0 \dots v_k \rangle$ のとき $\sigma_j = \langle v_0 \dots \hat{v}_j \dots v_k \rangle$ をあらわす.

注意 2.17. (i) $\sigma_j \subset \sigma$ であるから定義内の x の指数 $t(\sigma) - t(\sigma_j)$ は常に非負である。
(ii) $\langle\langle \sigma \rangle\rangle$ を $\langle \sigma \rangle$ とみなして, 形式的に $x = 1$ とおけば通常ホモロジー論の境界作用素に一致する。

ホモロジー群の場合と同様に $\partial_{k-1}(x) \circ \partial_k(x) = 0$ がなりたつことが簡単にチェックできるので,

$$B_k(\mathbb{X}) := \text{Im } \partial_{k+1}(x) \subset Z_k(\mathbb{X}) := \ker \partial_k(x)$$

がわかる。

定義 2.18. k 次のパーシステントホモロジー群を $H_k(\mathbb{X}) := Z_k(\mathbb{X})/B_k(\mathbb{X})$ と定義する。

一般に有限生成次数付き $\mathbb{F}[x]$ 加群の構造定理は以下のように与えられる。

定理 2.19. \mathbb{F} を体とし, $M = \bigoplus_{i \geq 0} M_i$ は次数付き $\mathbb{F}[x]$ -加群とする。このとき,

$$M \cong \bigoplus_{i=1}^p ((x^{b_i})/(x^{d_i})) \oplus \bigoplus_{i=p+1}^{p+q} (x^{b_i}).$$

ただし, $b_i, d_i \in \mathbb{Z}_{\geq 0}$, $b_i < d_i$ である。

よって, 定理 2.19 をパーシステントホモロジー群 $H_k(\mathbb{X})$ に適用すれば,

$$H_k(\mathbb{X}) \cong \bigoplus_{i=1}^p ((x^{b_i})/(x^{d_i})) \oplus \bigoplus_{i=p+1}^{p+q} (x^{b_i})$$

を得る。 b_i, d_i はそれぞれ i 番目の k 次パーシステントホモロジー類の生成時間と消滅時間をあらわす。 $i = p+1, \dots, p+q$ の場合は時刻 b_i で生成したホモロジー類は消滅しないことを意味する。重要なことは

$$\underbrace{\mathbb{X}}_{\text{input}} \implies \underbrace{\bigcup_{i=1}^p \{(b_i, d_i)\} \cup \bigcup_{i=p+1}^{p+q} \{(b_i, \infty)\}}_{\text{output}} \quad (2.1)$$

という図式が得られることである。これを実現する計算ソフトウェアも多く開発されている。ソフトウェアについては例えば PHAT (Persistent Homology Algorithm Toolbox [19]) などを参照せよ。

(2.1) において, 各 (b_i, d_i) に対する $\ell_i := d_i - b_i$ は i 番目のホモロジー類の生存時間を意味する。 $p+1 \leq i \leq p+q$ の場合は $\ell_i = \infty$ と理解する。このとき, $L_k = \sum_{i=1}^p \ell_i$ を k 次パーシステントホモロジー類の生存時間の和という。この $\{(b_i, d_i)\}_{i=1}^p$ の情報をパーシステント図やバーコードなど視覚的にあらわすと便利である [7, 10]。

定理 2.20 ([8]). X は有限な単体複体とし, $d \in \{1, 2, \dots, \dim X\}$ とする。 $\mathbb{X} = \{X(t)\}_{t \in [0, \infty]}$ を X の連続フィルトレーションとし, L_k を k 次パーシステントホモロジーの生存時間の和とすると以下の等式が成り立つ。

$$L_{d-1} = \int_0^\infty \tilde{\beta}_{d-1}(t) dt. \quad (2.2)$$

ただし, $\tilde{\beta}_i(t)$ は複体 $X(t)$ の i 次元被約 Betti 数である.
 さらに, $\tilde{\beta}_{d-1}(X^{(d)}) = \tilde{\beta}_{d-2}(X^{(d-1)}) = 0$ をみたすとき

$$L_{d-1} = \min_{T \in \mathcal{S}^{(d)}} \text{wt}(T) - \max_{S \in \mathcal{S}^{(d-1)}} \text{wt}(X_{d-1} \setminus S). \tag{2.3}$$

ただし, $\mathcal{S}^{(k)}$ は k 次元全域非輪体のなす集合, $S \subset X^{(k)}$ に対して $\text{wt}(S) = \sum_{\sigma \in S} t(\sigma)$ をあらわす.

注意 2.21. (2.3) の右辺は $\min_{T \in \mathcal{S}^{(d)}} \text{wt}(T) + \min_{S \in \mathcal{S}^{(d-1)}} \text{wt}(S) - \text{wt}(X_{d-1})$ ともあらわされるので, 単体複体の $(d-1)$ 次元と d 次元の最小全域非輪体の重みに相当する量であることがわかる. 特に $d=1$ のときは, 最小全域木の重みに等しい. 次節の注意 2.23 も参照のこと.

2.4 d -Linial-Meshulam ランダム複体過程

$V = \{1, 2, \dots, n\}$ に対して, $\Delta_{n-1} := K_n^{(n-1)}$ を $(n-1)$ 次元完全単体複体とする (定義 2.4). つまり, V 自身を facet とする単体複体とする. $(\Delta_{n-1})_d$ は上の定義より V の d -単体の全体である. 各 $\sigma \in (\Delta_{n-1})_d$ に対して, 独立な一様確率変数 $t(\sigma) \in [0, 1]$ を与えて,

$$X^{(d)}(t) = K_n^{(d-1)} \sqcup \{\sigma \in (\Delta_{n-1})_d : t(\sigma) \leq t\}$$

とおくと, 各 d 毎に, $\mathbb{X}^{(d)} := \{X^{(d)}(t), 0 \leq t \leq 1\}$ は単調非減少な d 次元単体複体に値をとる確率過程となる. Linial-Meshulam ([14]) によって各 t 毎定義されるランダム複体 $X^{(d)}(t)$ は詳しく考察されている. ここでは $\mathbb{X}^{(d)}$ のことを d -Linial-Meshulam ランダム複体過程 (以後, $\text{LM}^{(d)}$ -process) とよぶ. $d=1$ の場合は, $X^{(1)}(0) = K_n^{(0)} = V$ で, $(\Delta_{n-1})_1$ の各元が辺に対応するので, $\mathbb{X}^{(1)}$ はいわゆる Erdős-Rényi ランダムグラフ過程である.



図 5: Erdős-Rényi graph process ($d=1$)

$d=2$ のときは, 完全グラフ $X^{(2)}(0) = K_n^{(1)}$ からスタートして, $\binom{n}{3}$ 個ある 2-face (三角形) をランダムに加えていき, $X^{(2)}(1) = K_n^{(2)}$ で終了する. 一般に, $\text{LM}^{(d)}$ -process は $X^{(d)}(0) = K_n^{(d-1)}$ からスタートして, $\binom{n}{d+1}$ 個ある d -face をランダムに加えていき, $X^{(d)}(1) = K_n^{(d)}$ で終了する.

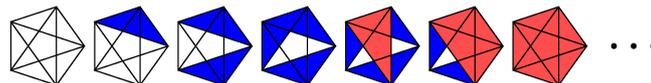


図 6: 2-Linial-Meshulam process

注意 2.22. 定義より $\text{LM}^{(d)}$ -process の時刻 t における $X^{(d)}(t)$ の k 次元 Betti 数は $\beta_k(t) \equiv 0$ ($k = 0, 1, \dots, d-2$) である. また, $(d-1)$ 次元 Betti 数 $\beta_{d-1}(t)$ は単調減少で,

$$\beta_{d-1}(0) = \binom{n-1}{d}, \quad \beta_{d-1}(1) = 0$$

である.

注意 2.23. $\text{LM}^{(d)}$ -process においては, 確率 1 で (2.1) において $q = 0$ となるので, $l_i, i = 1, 2, \dots, p$ はすべて有限 (定義より $d_i \leq 1$) である. よって, $L_k = \sum_{i=1}^p l_i \leq p$ である. 一方, $\dim(\sigma) \leq d-1$ ならば $t(\sigma) = 0$ なので, (2.3) において第 2 項目はすべて 0 となるので (2.3) は以下のようにシンプルになる.

$$L_{d-1} = \min_{T \in \mathcal{S}^{(d)}} \text{wt}(T). \quad (2.4)$$

右辺は $\text{LM}^{(d)}$ -process における最小全域非輪体 (minimum spanning acycle) の重みを与えていることになり, Frieze の考えている問題の自然な拡張を考えていることになる. つまり, 最小全域非輪体の問題とパーシステントホモロジーに関連があることが明らかになった.

3 Linial-Meshulam 過程のパーシステントホモロジーと最小全域非輪体

以下の定理 3.1 は 1 節で述べた Frieze の最小全域木の重みの漸近挙動の一般化に相当する.

定理 3.1. $d \in \mathbb{N}$ とする. L_{d-1} を n 点上の $\text{LM}^{(d)}$ -process \mathbb{X}^d の $(d-1)$ 次パーシステントホモロジーの生存時間の和とする. $n \rightarrow \infty$ のとき

$$\mathbb{E}[L_{d-1}] = \mathbb{E}[\min_{T \in \mathcal{S}^{(d)}} \text{wt}(T)] = O(n^{d-1}) \quad (3.1)$$

となる.

注意 3.2. 極限 $\lim_{n \rightarrow \infty} n^{-(d-1)} \mathbb{E}[L_{d-1}]$ については [8] で議論している.

$\mathbb{E}[L_{d-1}]$ の上からの評価には (2.2) を下からの評価には (2.3) を用いる.

(i) 下からの評価. 注意 2.23 で述べたように, $\text{LM}^{(d)}$ -process $\mathbb{X}^{(d)} = \{X^{(d)}(t)\}$ では任意の $(d-1)$ -face η に対して $t(\eta) = 0$ であるから, (2.4) にあるように T を $(d-1)$ 次元最小全域非輪体とすると $L_{d-1} = \text{wt}(T)$ となる. さらに $L_{d-1} = \text{wt}(T) \geq \sum_{i=1}^{|T|} u_i$ である. ここで, $0 \leq u_1 \leq u_2 \leq \dots \leq u_{\binom{n}{d}} \leq 1$ は $\{t(\sigma), \sigma \in (\Delta_{n-1})_d\}$ を小さい順に並べたものである. 順序統計量の平均の公式と $|T| = \binom{n-1}{d}$ であることより

$$\mathbb{E}[L_{d-1}] \geq \sum_{i=1}^{|T|} \mathbb{E}[u_i] = \sum_{i=1}^{|T|} \frac{i}{\binom{n}{d+1} + 1} = O(n^{d-1}).$$

(ii) 上からの評価. $\mathbb{E}[L_{d-1}]$ の上からの評価には (2.2) を用いる. そのためには $\tilde{\beta}_{d-1}(t)$ の評価が必要となる. 例えば, $\tilde{\beta}_{d-1}(t)$ は $(d-1)$ -単体の個数 $\binom{n}{d}$ で自明に上から評価される. これを用いると,

$$\mathbb{E}[L_{d-1}] = \int_0^1 \tilde{\beta}_{d-1}(t) dt \leq \int_0^1 \binom{n}{d} dt = O(n^d)$$

となるが, これでは大雑把すぎる. 次節ではもう少し構造に踏み込んだ Betti 数の評価に必要な Kruskal-Katona の定理について述べる.

3.1 Kruskal-Katona の定理

$[n] = \{1, 2, \dots, n\}$, $\mathcal{F} \subset 2^{[n]}$ とする. \mathcal{F} に対して

$$\partial\mathcal{F} := \{E \subset [n] : \exists F \in \mathcal{F} \text{ s.t. } E \subset F \text{ and } |F \setminus E| = 1\}$$

と定義する.

$[n]$ の k 点部分集合 $\binom{[n]}{k} = \{F \in 2^{[n]} : |F| = k\}$ とするとき, $\mathcal{F} \subset \binom{[n]}{k}$ かつ $|F| = m$ をみたく \mathcal{F} に対して $|\partial\mathcal{F}|$ の上界と下界について何が言えるか? 上界については, \mathcal{F} の元が互いに素ならば, 明らかに $|\partial\mathcal{F}| \leq km$ をみたく. 下界については Kruskal-Katona の定理が知られている. 定理の主張を述べるために自然数の k -カスケード表現について述べる.

定義 3.3 (k -カスケード表現). $m \in \mathbb{N}$ とする. $k \in \mathbb{N}$ に対してある $a_k > a_{k-1} > \dots > a_s \geq 1$ が存在して

$$m = \binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \dots + \binom{a_s}{s} \quad (3.2)$$

と一意的に表現される. これを自然数 m の k -カスケード表現という.

例 3.4. $k=3$ の場合, 例えば以下のようになる.

$$\begin{aligned} 3 &= \binom{3}{3} + \binom{2}{2} + \binom{1}{1}, & 4 &= \binom{4}{3}, & 5 &= \binom{4}{3} + \binom{2}{2}, \\ 6 &= \binom{4}{3} + \binom{2}{2} + \binom{1}{1}, & 7 &= \binom{4}{3} + \binom{3}{2}, & 8 &= \binom{4}{3} + \binom{3}{2} + \binom{1}{1}, \\ 9 &= \binom{4}{3} + \binom{3}{2} + \binom{2}{1}, & 10 &= \binom{5}{3}, \dots \end{aligned}$$

定理 3.5 (Kruskal-Katona). $\mathcal{F} \subset \binom{[n]}{k}$ かつ $|\mathcal{F}| = m$ とし, m の k -カスケード表現が (3.2) で与えられているとする. このとき, $|\partial\mathcal{F}|$ の下界は

$$|\partial\mathcal{F}| \geq \binom{a_k}{k-1} + \binom{a_{k-1}}{k-2} + \dots + \binom{a_s}{s-1} \quad (3.3)$$

で与えられる.

例 3.6. 例 3.4 の場合 ($k = 3$) について考えると, 定理 3.5 により

$$|\partial\mathcal{F}| \geq \begin{cases} 3, & m = 1 \\ 5, & m = 2 \\ 6, & m = 3, 4 \\ 8, & m = 5 \\ 9, & m = 6, 7 \\ 10, & m = 8, 9, 10, \dots \end{cases}$$

となる. 2 単体の集合 \mathcal{F} を facet とする単体複体 X を考えると, $|\partial\mathcal{F}|$ は X の 1-単体の個数に等しいので, 上の不等式は与えられた個数 m の 2-単体でなるべくタイトに単体複体をつくったときの辺の本数の下界を与える. 例えば, $m = 4$ のときは 2-単体を四面体の形にするとともに辺の本数を少なく 6 本にできる.

上の不等式は $|\partial\mathcal{F}|$ のタイトな評価を与えるが種々の評価のためには若干使いにくい. 以下は Lovász による簡易版である. $x \geq k$ に対して $\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}$ は k 次多項式と考える. $x \geq k$ で $\binom{x}{k}$ は単調増加なので, $\binom{x}{k} = m$ となる $x \geq k$ が一意的に定まる.

定理 3.7 (Kruskal-Katona の定理の Lovász 版). 上と同様の設定のもと以下の不等式が成り立つ. $\mathcal{F} \subset \binom{[n]}{k}$ かつ $|\mathcal{F}| = m = \binom{x}{k}$ ($x \geq k$) とする. このとき,

$$|\partial\mathcal{F}| \geq \binom{x}{k-1}$$

定理 3.5 と定理 3.7 の証明については例えば [5] を参照せよ.

Kruskal-Katona の定理とほぼ同様の思想で, 以下の定理が示されている.

定理 3.8 ([15]). Y を d 次元の単体複体とし, $|Y_d| = \binom{x}{d+1}$ ($x \geq d+1$) とする. このとき, 境界作用素 $\partial_d: C_d(Y) \rightarrow C_{d-1}(Y)$ に対して不等式

$$\text{rank } \partial_d \geq \binom{x-1}{d}$$

が成り立つ. 特に, Y が n 点上の単体複体の場合は

$$\text{rank } \partial_d \geq \frac{d+1}{n} |Y_d|$$

が成り立つ.

3.2 Betti 数の評価

以下では, 一般に Y を n 点上の d 次元の単体複体とし, $(d-1)$ 次完全単体複体 $K_n^{(d-1)}$ を含むとする.

Y に d -単体を加えたときに Y の $(d-1)$ -Betti 数は「変化しない」もしくは「1 減少する」かのいずれかである。これは境界作用素の言葉では以下のように言い換えられる。 $\partial_d: C_d(Y) \rightarrow C_{d-1}(Y)$ から $\partial_d: C_d(Y \cup \{\sigma\}) \rightarrow C_{d-1}(Y \cup \{\sigma\})$ としたときに rank が「変化しない」もしくは「1 増加する」かのいずれかである。このあたりの状況を調べるために以下の量を考える。

$$\mathcal{R}_d(Y) = \{\sigma \in Y_d : \tilde{\beta}_{d-1}(Y \cup \{\sigma\}) = \tilde{\beta}_{d-1}(Y) - 1\}.$$

このとき、定理 3.8 より以下の評価が得られる。

命題 3.9. $\tilde{\beta}_{d-1}(Y) \leq \frac{d+1}{n} |\mathcal{R}_d(Y)|.$

証明. [8] を参照のこと. □

(簡単のために) $\text{LM}^{(d)}$ -process $\mathbb{X}^{(d)} = \{X^{(d)}(t)\}$ を $\mathbb{Y} = \{Y_t\}$ と書くと、(2.2) と命題 3.9 より

$$\mathbb{E}[L_{d-1}] = \int_0^1 \mathbb{E}[\tilde{\beta}_{d-1}(Y_t)] dt \leq \frac{d+1}{n} \int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)| dt$$

となる。

$\text{LM}^{(d)}$ -process のカップリングの議論などにより以下の不等式を得る。

命題 3.10. $\text{LM}^{(d)}$ -process に対して

$$\int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)| dt \leq 8 \binom{n}{d}$$

証明. [8] を参照のこと. □

命題 3.9 と命題 3.10 をあわせれば定理 3.1 は得られる。

$$\mathbb{E}[L_{d-1}] = \int_0^1 \mathbb{E}[\tilde{\beta}_{d-1}(Y_t)] dt \leq \frac{d+1}{n} \int_0^1 \mathbb{E}|\mathcal{R}_d(Y_t)| dt \leq \frac{d+1}{n} \cdot 8 \binom{n}{d} = O(n^{d-1}).$$

3.3 最後に

本稿では、Linial-Meshulam 複体過程のパーシステント図の汎関数である生存時間の和について議論した。これは最初に述べた Frieze の最小全域木の重みに関する極限定理の高次元への一般化として自然なものである。このように一見無関係に見える最小全域非輪体の重みとパーシステントホモロジーの間の自然な関係は [8] で初めて指摘された。ここで述べたのは Linial-Meshulam 複体過程に関連した最小全域非輪体の重みに対する大数の法則であるが、他のランダム複体過程に対する同様の問題や注意 1.4 で述べた中心極限定理の一般化も重要な問題である。また、パーシステント図自体の点過程としての極限定理も興味深く今後の研究が待たれる。

参考文献

- [1] C. Cooper, A. Frieze, N. Ince, S. Janson and J. Spencer. On the length of a random minimum spanning tree. arXiv:1208.5170v2.
- [2] A. Duval, C. J. Klivans and J. L. Martin. Simplicial matrix-tree theorems. *Trans. Amer. Math. Soc.* **361** (2009), 6073–6114.
- [3] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological Persistence and Simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.
- [4] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen* **6** (1959), 290–297.
- [5] P. Frankl. A new short proof for the Kruskal-Katona theorem. *Discrete Math.* **48** (1984), 327–329.
- [6] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Math.* **10** (1985), 47–56.
- [7] R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.* **45** (2008), 61–75.
- [8] Y. Hiraoka and T. Shirai. Minimum spanning acycle and lifetime of persistent homology in the Linial-Meshulam process. <http://arxiv.org/abs/1503.05669>
- [9] Y. Hiraoka and T. Shirai. ランダム複体とパーシステントホモロジー (RIMS 講究録 1952).
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/1952.html>
- [10] 平岡裕章, タンパク質構造とトポロジー—パーシステントホモロジー群入門—. 共立出版, 2013年.
- [11] G. Kalai. Enumeration of Q -acyclic simplicial complexes. *Israel J. Math.* **45** (1983), 337–351.
- [12] S. Janson. The Minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Struct. Alg.* **7** (1995), 337–355.
- [13] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* **7** (1956), 48–50.
- [14] N. Linial and R. Meshulam. Homological connectivity of random 2-complexes. *Combinatorica* **26** (2006), 475–487.
- [15] N. Linial, I. Newman, Y. Peled and Y. Rabinovich. Extremal problems on shadows and hypercuts in simplicial complexes. arXiv:1408.0602v2.

- [16] R. Lyons. Random complexes and ℓ^2 -Betti numbers. *J. Topology Anal.* **1** (2009), 153–175.
- [17] R. Meshulam and N. Wallach. Homological connectivity of random k -dimensional complexes. *Random Struct. Alg.* **34** (2009), 408–417.
- [18] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.* **33** (2005), 249–274.
- [19] PHAT (Persistent Homology Algorithm Toolbox).
<https://code.google.com/p/phat/>