

グラフ上の線形 Cover Time ランダムウォーク実現の必要条件

野中 良哲 (Yoshiaki Nonaka) *

小野 廣隆 (Hiroataka Ono) †

定兼 邦彦 (Kunihiko Sadakane) †

山下 雅史 (Masafumi Yamashita) †

* 九州大学大学院システム情報科学府

† 九州大学大学院システム情報科学府

Graduate School of Information Science and Electrical Engineering Kyushu University

1 はじめに

1.1 定義

グラフ $G = (V, E)$ 上のランダムウォークとは、以下のようなプロセスである。

1. ある頂点 $v_0 \in V$ から出発する。
2. v_0 の隣接点 $v_1 \in N(v_0)$ をランダムに選択し、 v_1 に移動する。これを 1 ステップとする。
3. v_1 に対しても 2 と同様に隣接する頂点に移動し、以下同様に繰り返す。

通常単にランダムウォークといえば、隣接する頂点に等確率で遷移するという標準ランダムウォーク (定義 1.1) を指す。

定義 1.1

$$p_{uv} = \begin{cases} \frac{1}{\deg(u)} & (u \in N(v)) \\ 0 & (\text{otherwise}) \end{cases} \quad u, v \in V$$

定義 1.2 (Hitting Time) グラフ $G = (V, E)$ におけるランダムウォークで、頂点 $u \in V$ から出発した粒子が遷移確率行列 P に従ってランダムに移動し、頂点 $v \in V$ に初めて到達するまでに要する遷移数の期待値を $H_G^P(u, v)$ とする。このとき G の Hitting Time (H_G^P) を以下のように定義する。

$$H_G^P = \max_{u, v \in V} H_G^P(u, v)$$

定義 1.3 (Cover Time) グラフ $G = (V, E)$ におけるランダムウォークで、頂点 $u \in V$ から出発した粒子遷移確率行列 P に従ってランダムに移動し、全ての頂点を訪問するまでに要する遷移数の期待値を $C_G^P(u)$ とする。このとき G の Cover Time (C_G^P) を以下のように定義する。

$$C_G^P = \max_{u \in V} C_G^P(u).$$

Hitting Time, Cover Time はランダムウォークの速さの指標であり、これらの値が小さい程高速なランダムウォークであるといえる。

池田ら [4] は、任意の遷移確率に対する頂点数 n の道グラフにおける Hitting Time, Cover Time の下界について以下のように示している。

定理 1.4 $P_n = (V, E)$ を頂点数 n の道グラフとし、 $v_1, v_2, \dots, v_n \in V$ (v_1, v_n は端点) とする。このとき、任意の $\mu \in M^+(\Omega)$ に対し以下の式が成り立つ。

$$\frac{1}{2}(n^2 - 1) \leq C_{P_n}^P \leq 2H_{P_n}^P$$

1.2 隣接頂点情報を扱うランダムウォーク

1.2.1 定義

任意の n 個の頂点をもつ有限無向グラフ $G = (V, E)$ において、遷移確率行列 $P^{(\beta)} = (p_{uv}^{(\beta)})_{u, v \in V}$ を以下のように定義する。

定義 1.5

$$p_{uv}^{\beta} = \frac{\exp[-\beta U(u, v)]}{\sum_{w \in N(u)} \exp[-\beta U(u, w)]}$$

$$u \in V, v \in N(u)$$

ここで β は実数とする。また、 $U(u, w)$ は以下のように定義される potential function である。

$$U(u, v) = U(v) = \log \deg(v)$$

$$v \in N(u), u \in V$$

池田ら [4] は、このランダムウォークにおいて任意のグラフに対して Hitting Time が $O(n^2)$ 、Cover Time が $O(n^2 \log n)$ であることを示している。既存の研究により、グラフの隣接情報を用いることでランダムウォークが高速化されることが分かった。ここで、より離れた頂点の持つ情報を利用すればさらにランダムウォークを高速化できると考えられる。実際、ハミルトングラフにおいては、 $n-1$ ステップで全ての頂点を訪問する遷移確率を定義可能である。一方でパスグラフのように、Cover Time に $\Omega(n^2)$ を要するグラフも存在する。

2 以前の結果

本研究の目的は、グラフ全体の構造を利用した高速化を考えることで、近傍情報を用いた高速化の限界を明らかにすることにある。特に、頂点数 n に対して線形な Cover Time を実現可能なグラフの属するクラスに関心がある。

発表者 [7] は、以前 Tree 上のランダムウォークにおける Cover Time の下界を示した、ハミルトニアン成分分解を提案し、線形 Cover Time を持つための十分条件を示した。

2.1 Tree における Cover Time

定理 2.1 Tree 上でのランダムウォークに対する Cover Time は $\Omega(n \log n)$

この定理により、Tree はいかなる遷移確率を用いても線形な Cover Time を持ち得ないことがあきらかになった。

2.2 ハミルトニアン成分分解

補題 2.2 グラフ $G = (V, E)$ に対し、ある整数 $k > 0$ および以下のような V の分割 V_1, V_2, \dots, V_k が存在する。

- $|V_i| = 1$, または
- $|V_i| \geq 3$ かつ V_i の頂点を全て通る閉路が存在する。

ただし、 $1 \leq i \leq k$ 。

この補題を用いて、グラフに対するハミルトニアン成分分解を定義する。

定義 2.3 グラフ $G = (V, E)$ が与えられたとき、 V に対して補題 2.2 の条件を満たすような分割 V_1, V_2, \dots, V_k を与える。このとき、 V_i のなす閉路を \mathcal{H}_i とし、これをハミルトニアン成分と定義する。閉路にはある向きを定める。両端の頂点がそれぞれ異なるハミルトニアン成分に属する枝から、ハミルトニアン成分が木構造をなすように、ハミルトニアン成分同士を接続する枝の集合 $E_t \subset E$ を定義する。

グラフ G に対するハミルトニアン成分分解は、分割 V_1, V_2, \dots, V_k と E_t により定義される。

グラフに対してハミルトニアン成分分解を適用し、その上でのランダムウォークを考える。ハミルトニアン成分分解のサイズと Cover Time に関して、以下の定理、系が成り立つ。

定理 2.4 頂点数 n のグラフがサイズ k のハミルトニアン成分分解を持つとき、Cover Time が $O(nk)$ となるランダムウォークを定義できる。

系 1 頂点数 n のグラフが定数サイズのハミルトニアン成分分解を持つとき Cover Time が $O(n)$ となるランダムウォークを定義できる。

3 二点連結成分分解と線形 Cover Time の必要条件

系 1 はグラフが線形 Cover Time を持つための十分条件を与えているが, 必要十分条件との間には大きなギャップがある. 実際, 定数サイズのハミルトニアン成分分解を持たないが, Cover Time が $O(n)$ となるような遷移確率を定義可能なグラフが存在することが分かっている.

3.1 二点連結成分分解

補題 3.1 グラフ $G = (V, E)$ に対し, ある整数 $k > 0$ および以下のような V の分割 B_1, B_2, \dots, B_k が存在する.

- $|B_i| = 1$, または
- $|B_i| \geq 3$ かつ 任意の $x, y \in B_i$ の間に, B_i に含まれる頂点のみからなる 2 つ以上の点素なパスが存在する,

ただし, $1 \leq i \leq k$.

この補題を用いて 2 点連結成分分解を以下のよう
に定義する.

定義 3.2 グラフ $G = (V, E)$ が与えられたとき, V に対して補題 3.1 の条件を満たすような分割 B_1, B_2, \dots, B_k を与える. このとき, 各 B_i を 2 点連結成分と定義する. ここで, $B_i, B_j (i \neq j)$ について, 以下を満たすならば, B_i, B_j を 1 つの 2 点連結成分とする.

$$\forall x \in B_i, y \in B_j \exists \omega_1, \omega_2 \in \mathcal{P}(x, y) (\omega_1 \cap \omega_2 = \emptyset)$$

ただし, $\mathcal{P}(x, y)$ は頂点 x から頂点 y へのパスの集合である. この操作を合成できる 2 点連結成分がなくなるまで行う.

これは先に提案したハミルトニアン成分分解を拡張したものといえる. 図 1 は簡単な 2 点連結成分分解の例である. このように, グラフに対して 2 点連結成分分解を与えると, 2 点連結成分は Tree 構造をなす.

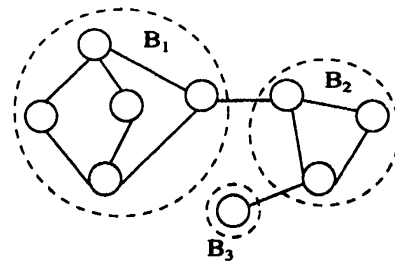


図 1: 2 点連結成分分解の例

3.2 線形 Cover Time の必要条件

グラフ G がサイズ k の 2 点連結成分分解を持ち, 最大の 2 点連結成分が α 個の頂点からなっているとす
る. また, 全ての 2 点連結成分をそれぞれ 1 点に縮約することで得られる Tree を T_G とし, その直径を W とする. このとき以下の定理が成り立つ.

定理 3.3 グラフが線形 Cover Time のランダムウォークをもつための必要条件は,

$$\begin{aligned} \alpha &= \Omega(\log n) \\ k &= O\left(\frac{n}{\log n}\right) \\ W &= O(\sqrt{n}) \end{aligned}$$

であることである.

証明 後に示す補題 3.4 を繰り返し適用すると, $C_G \geq C_{T_G}$ が得られる. $k \geq n/\alpha$ なので,

$$\begin{aligned} C_G &\geq C_{T_G} = \Omega\left(\frac{n}{\alpha} \log \frac{n}{\alpha}\right) \\ &= \Omega\left(\frac{n}{\alpha} (\log n - \log \alpha)\right) \\ &= \Omega\left(\frac{n}{\alpha} \log n\right) \end{aligned}$$

より, $C_G = O(n)$ となるためには, $\alpha = \Omega(\log n)$ である必要がある. よって, $k = O\left(\frac{n}{\log n}\right)$. また, 直径 W の Tree には長さ W のパスが存在する. 長さ W のパスグラフの Cover Time が $\Omega(W^2)$ であることから, $W = O(\sqrt{n})$ が得られる.

補題 3.4 任意のグラフ $G = (V, E)$ に対し, ある遷移確率行列 P を与えたときの Cover Time を C_G^P とする. このとき G 上のある 2 頂点 $u_0, v_0 \in V$ を 1 つの頂点 w に縮約したグラフ $G' = (V', E')$ に対して, $C_G^P \geq C_{G'}^{P'}$ となるような遷移確率行列 P' が存在する.

証明は付録参照.

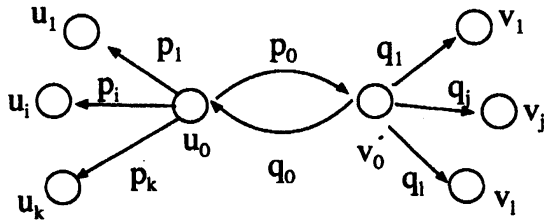


図 2: 縮約前

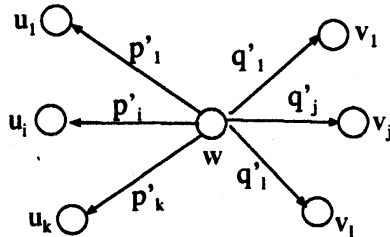


図 3: 縮約後

4 おわりに

本発表では, 以前提案したハミルトニアン成分分解の拡張として 2 点連結成分分解を定義し, それによってグラフが線形な Cover Time を持つための必要条件を示した. しかし, 以前求めた十分条件との間には未だ大きなギャップが存在する. 特に 2 点連結グラフ自体の Cover Time について詳細な解析を行う必要がある.

参考文献

- [1] D.J.Aldous, "On the token by random walks on finite groups to visit every state", Z.Wahrsch. verw. Gebiete 62 361-1983.
- [2] R.Aleliunas, R.M Karp, R.J. Lipton, L.Lovaasz, and C.Rackoff, "Random walks, universal traversal sequences, and the complexity of maze problems", Proc. 20th IEEE Ann.Symposium on Foundations of Computer Science, 218-223, 1979.
- [3] A.Z.Broder, and Karlin, "Bounds on covering thmes", In 29th IEEE Annual Symposium on Foundations of Computer science, 479-487 1988.
- [4] Satoshi Ikeda, Izumi Kubo, Norihiro Okumoto, Masafumi Yamashita "Impact of Local Topological Information on Random Walks on Finite Graphs", Proceedings of Thirtieth International Colloquium on Automata, Languages and Programming, 1054-1067, 2003.
- [5] Jeff D. Kahn, Nathan Linial, Noam Nisan, Michael E.Saks "On the Cover Time of Random Walks on Graphs", Journal of Theoretical Probability, Vol.2, No. 1:121-128, 1989.
- [6] P.Matthews, "Covering Problems for Markov Chain", The Annals of Probability Vol.16, No3, 1215-1228, 1988.
- [7] 野中 良哲, 小野 廣隆, 定兼 邦彦, 山下 雅史, "ランダムウォーク高速化とその限界", 2007 年度夏の LA シンポジウム [14].
- [8] Norihiro Okumoto "A study on Random Walks of Tokens on Graphs", Dissertaion for the degree of Master of Engineering Graduate School of Engineering Hiroshima University, 1996

付録. 補題 3.4 証明

証明 2つの頂点 u_0, v_0 は隣接していると考え、その隣接点集合を $N(u_0) = \{v_0, u_1, u_2, \dots, u_k\}, N(v_0) = \{u_0, v_1, v_2, \dots, v_l\}$ とする。 u_0, v_0 から各隣接点への遷移確率を図 2 のように、それぞれ以下で表す。

$$\begin{aligned} p(u_0, v_0) &= p_0, & p(u_0, u_i) &= p_i (i = 1, 2, \dots, k) \\ p(v_0, u_0) &= q_0, & p(v_0, v_j) &= q_j (j = 1, 2, \dots, l) \end{aligned}$$

この時、 u_0 からの u_1 への Hitting Time $H(u_0, u_1)$ を考える。ただし、 $K = \{1, 2, \dots, k\}, L = \{1, 2, \dots, l\}$ 。頂点 u_i から出発し、 $\{u_1, u_0, v_0\}$ のいずれかの頂点へ最初に訪れる確率をそれぞれ、 r_{i1}, r_{iu}, r_{iv} とする。また、このときに要した遷移数の期待値をそれぞれ h_{i1}, h_{iu}, h_{iv} とする。同じく、頂点 v_j から出発し、 $\{u_1, u_0, v_0\}$ のいずれかの頂点へ最初に訪れる確率をそれぞれ、 s_{j1}, s_{ju}, s_{jv} とする。また、このときに要した遷移数の期待値をそれぞれ k_{j1}, k_{ju}, k_{jv} とする。

このとき、 $H(u_0, v_1)$ は以下の式で表される。

$$H(u_0, u_1) = \frac{p_1 + p_0 + \frac{q_0}{1 - \sum_{j \in L} q_j s_{jv}} + \sum_{i \in K \setminus \{1\}} p_i (1 + \hat{h}_i) + \frac{1}{1 - \sum_{j \in L} q_j s_{jv}} \sum_{j \in L} q_j (1 + \hat{k}_j)}{p_1 + \sum_{i \in K \setminus \{1\}} p_i r_{i1} + \frac{(p_0 + \sum_{i \in K \setminus \{1\}} p_i r_{iv}) \sum_{j \in L} q_j s_{j1}}{1 - \sum_{j \in L} q_j s_{jv}}} \quad (1)$$

同様に

$$H(v_0, v_1) = \frac{q_1 + q_0 + \frac{p_0}{1 - \sum_{i \in K} p_i r_{iu}} + \sum_{j \in L \setminus \{1\}} q_j (1 + \hat{k}_j) + \frac{1}{1 - \sum_{i \in K} p_i r_{iu}} \sum_{i \in K} p_i (1 + \hat{h}_i)}{q_1 + \sum_{j \in L \setminus \{1\}} q_j s_{j1} + \frac{(q_0 + \sum_{j \in L \setminus \{1\}} q_j s_{ju}) \sum_{i \in K} p_i r_{i1}}{1 - \sum_{i \in K} p_i r_{iu}}} \quad (2)$$

ただし、 $\hat{h}_i = r_{i1} h_{i1} + r_{iu} h_{iu} + r_{iv} h_{iv}$, $\hat{k}_j = s_{j1} k_{j1} + s_{ju} k_{ju} + s_{jv} k_{jv}$ である。次に、縮約後の w から u_1 への Hitting Time $H'(w, u_1)$ を考える。縮約に関係無い部分の遷移確率は変更しないとすると、 $H'(w, u_1)$ は以下の式で表される。

$$H'(w, u_1) = \frac{p'_1 + \sum_{i \in K \setminus \{1\}} p'_i \{1 + \hat{h}_i\} + \sum_{j \in L} q'_j \{1 + \hat{k}_j\}}{p'_1 + \sum_{i \in K \setminus \{1\}} p'_i r_{i1} + \sum_{j \in L} q'_j s_{j1}} \quad (3)$$

同様に

$$H'(w, v_1) = \frac{q'_1 + \sum_{j \in L \setminus \{1\}} q'_j \{1 + \hat{k}_j\} + \sum_{i \in K} p'_i \{1 + \hat{h}_i\}}{q'_1 + \sum_{j \in L \setminus \{1\}} q'_j s_{j1} + \sum_{i \in K} p'_i r_{i1}} \quad (4)$$

ここで縮約後の遷移確率を $p'_i = \alpha p_i, q'_j = \beta q_j$ と定義する。ただし、 $(\alpha, \beta > 0, \beta = \varepsilon \alpha)$ である。すると、式

(3), 式(4) は次のように表される.

$$H'(w, u_1) = \frac{p_1 + \sum_{i \in K \setminus \{1\}} p_i \{1 + \hat{h}_i\} + \varepsilon \sum_{j \in L} q_j \{1 + \hat{k}_j\}}{p_1 + \sum_{i \in K \setminus \{1\}} p_i r_{i1} + \varepsilon \sum_{j \in L} q_j s_{j1}} \quad (5)$$

$$H'(w, v_1) = \frac{q_1 + \sum_{j \in L \setminus \{1\}} q_j \{1 + \hat{k}_j\} + \frac{1}{\varepsilon} \sum_{i \in K} p_i \{1 + \hat{h}_i\}}{q_1 + \sum_{j \in L \setminus \{1\}} q_j s_{j1} + \frac{1}{\varepsilon} \sum_{i \in K} p_i r_{i1}} \quad (6)$$

$H(u_0, u_1) \geq H'(w, u_1)$ かつ $H(v_0, v_1) \geq H'(w, v_1)$ を満たす ε が存在することを示せばよい. 式(1) および, 式(5) より, ε が以下の区間にあるとき, $H(u_0, u_1) \geq H(w, u_1)$ を満たす.

$$\frac{p_0 + \sum_{i \in K \setminus \{1\}} p_i r_{i0}}{1 - \sum_{j \in L} q_j s_{j0}} \leq \varepsilon \leq \frac{1}{1 - \sum_{j \in L} q_j s_{jv}} \quad (7)$$

また, 式(2) および, 式(6) より, ε が以下の区間にあるとき, $H(v_0, v_1) \geq H(w, v_1)$ を満たす.

$$1 - \sum_{i \in K} p_i r_{iu} \leq \varepsilon \leq \frac{1 - \sum_{i \in K} p_i r_{iu}}{q_0 + \sum_{j \in L \setminus \{1\}} q_j s_{ju}} \quad (8)$$

ここで,

$$\frac{1}{1 - \sum_{j \in L} q_j s_{jv}} \geq 1 \geq 1 - \sum_{i \in K} p_i r_{iu}$$

また

$$\begin{aligned} \frac{1 - \sum_{i \in K} p_i r_{iu}}{q_0 + \sum_{j \in L \setminus \{1\}} q_j s_{ju}} &\geq \frac{1 - \sum_{i \in K} p_i (1 - r_{i1} - r_{iv})}{q_0 + \sum_{j \in L} q_j (1 - s_{j1} - s_{jv})} \\ &\geq \frac{p_0 + \sum_{i \in K \setminus \{1\}} p_i r_{i0}}{1 - \sum_{j \in L} q_j s_{jv}} \end{aligned}$$

これは, 式(7) が表す区間と, 式(8) が表す区間に重なりがあることを示している. すなわち条件を満たすような ε が必ず存在する. このような ε が決まると, $\alpha(p_0 - 1) + \beta(q_0 - 1) = 1$ から縮約後の遷移確率を定義できる. p_0, q_0 の一方もしくは両方が 0 の場合も同様に Cover Time を小さくするような遷移確率を定義できることが示される. u_1, v_1 として適切な頂点を選ぶと, 他の隣接点への Hitting Time も小さくすることが示される.