

2 点連結な直並列グラフ上の高速なランダムウォーク

穂坂 祐輔* 小野 廣隆† 定兼 邦彦† 山下 雅史†

*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研究院

1 はじめに

ランダムウォークとは、次に出現する位置が確率的に無作為に決定されるモデルであり、ブラウン運動や、証券株価の具体的モデル化に應用されている。グラフ上のランダムウォークとは、グラフ上の頂点にある粒子を遷移確率行列に従って隣接する頂点に遷移させていくモデルである。隣接頂点に等確率で遷移するランダムウォークを標準ランダムウォークという。グラフ $G = (V, E)$ 上のランダムウォークで、頂点 u から頂点 v まで粒子が移動するのに必要な遷移数の期待値を $H_G(u, v)$ とすると G の到達時間 H_G を $H_G = \max_{u, v \in V} H_G(u, v)$ と定義する。また頂点 u を出発した頂点が他の全ての頂点を訪れるまでに要する遷移数の期待値を、 $C_G(u)$ とすると、 G の全訪問時間を $C_G = \max_{u \in V} C_G(u)$ と定義する。ランダムウォークの高速化とは、到達時間、全訪問時間のより小さな遷移確率行列を与えることである。

標準ランダムウォークとは各頂点で粒子が隣接頂点に等確率で遷移するランダムウォークである。標準ランダムウォークの到達時間、全訪問時間については過去の研究により、様々なことが解明されている。まず Aleliusu[2] らが、頂点数 n 、辺数 m の一般のグラフにおいて全訪問時間の上限が $2m(n-1) = O(n^3)$ であることを示している。次に Aldous[1] が全訪問時間に関する同様の上限を示し、近年全訪問時間が最大で $4/27n^3$ であることが知られている。[3]

また [3] において、頂点数 n の木グラフ上の標準ランダムウォークが到達時間、全訪問時間が共に $O(n^2)$ であることが知られている。従って任意の連結グラフにおいて、その全域木上で標準ラ

ンダムウォークを行うような遷移確率行列を与えることで、到達時間、全訪問時間が共に $O(n^2)$ であるようなランダムウォークが設計でき、これは一般の連結なグラフにおける到達時間、全訪問時間の高速化の上限となっている。

著者は 2 点連結なグラフにおいて、全ての頂点がグラフ上のいずれかの閉路に含まれるという性質を利用することで、全訪問時間が一般の連結グラフの上限の $O(n^2)$ より低く出来る余地があると考えており、閉路を利用したランダムウォークの高速化手法を提案する。提案手法により、2 点連結な直並列グラフにおいては $O(n\sqrt{n} \log n)$ の全訪問時間を達成する遷移確率行列が存在することを発見した。

2 閉路を利用した高速化

2.1 ハミルトン閉路

ハミルトン閉路とはグラフの全ての頂点を一度ずつ通るような閉路を指す。ハミルトン閉路を持つグラフをハミルトングラフと呼ぶ。グラフ $G = (V, E)$ が以下のようなハミルトン閉路を持つとする。

$$v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n \rightarrow v_1 (v_i \in V : i = 1, 2, \dots, n) \tag{1}$$

このとき頂点 v_i から v_j への遷移確率 p_{ij} を以下のように決める。

$$p_{ij} = \begin{cases} 1(j = i + 1 \text{ または } i = n, j = 1) \\ 0(\text{その他}) \end{cases} \tag{2}$$

この遷移確率行列に従えば、グラフ G のどの頂点から出発してもハミルトン閉路を一周すればどの頂点にも到達できるので、到達時間及び全訪問時間は $n-1$ である。本発表ではハミルトングラフでのランダムウォークを基に、閉路を利用してランダムウォークを高速化することができることを示す。

2.2 閉路循環向き付け

2.2.1 定義

強連結グラフになるように無向グラフの辺に向きを付けることを閉路循環向き付けと定義する。また全ての無向グラフでこの向き付けが存在するとは限らない。

2.2.2 補題 2.1

任意の 2 点連結グラフで閉路循環向き付けが可能。

証明

任意の 2 点連結無向グラフ $G = (V, E)$ に対して、以下手順で各辺に向き付けをした有向グラフ G' を生成する。

手順 1: グラフ G から任意の閉路 $C(V_C, E_C)$ ($|V_C| = n_c$) を選び、閉路 C において、 $v_i \in V_C$: $i = 1, 2, \dots, n_c$ として、

$$v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_{n_c} \rightarrow v_1 \quad (3)$$

のように頂点が並んでいるとする。ここで閉路 C 上の頂点 v_i, v_{i+1} 間の辺 $e_{(i,i+1)}$ ($1 \leq i \leq n_c - 1$) および頂点 v_{n_c}, v_1 間の辺 $e_{(n_c,1)}$ に対して以下のように向きを付ける。

$$\begin{cases} e_{(i,i+1)} : v_i \rightarrow v_{i+1} \\ e_{(n_c,1)} : v_{n_c} \rightarrow v_1 \end{cases} \quad (4)$$

グラフ G の頂点が全て閉路 C に含まれていればここで終了する。

手順 2: グラフ G から両端のみが閉路 C に含まれる路 $P_0(V_{P_0}, E_{P_0})$ (頂点数 n_0) を選ぶ。両端の頂点を $v_1, v_{n_0} \in V_C$ とし、路上の頂点が

$$v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_{n_0} (v_i \in V_{P_0} : i = 1, 2, \dots, n_0) \quad (5)$$

のように並んでるとする。ここで路 P_0 上の辺 $e_{(i,i+1)}$ ($1 \leq i \leq n_0 - 1$) にたいして以下のように向き付けを行う。

$$e_{(i,i+1)} : v_i \rightarrow v_{i+1} \quad (6)$$

ここで G の部分グラフ $G' = (V', E')$ を以下のように決める。

$$V' = V_C \cup V_{P_0}, E' = E_C \cup E_{P_0} \quad (7)$$

手順 3: グラフ G から両端のみが G' に含まれている路 P_i を選び、手順 2 と同様にして辺の向きを決め、 $G' \cup P_i$ を新たに G' とする。 $V = V'$ となるまでこの操作を繰り返す。2 点連結グラフでは部分グラフ G' に路 P_i が接続しているとき、必ず路の両端が接続している。路の片方の端点しか接続していなければ、唯一の接続点を取り除けばグラフが非連結となり、2 点連結グラフではない。□

2.2.3 補題 2.2

閉路循環向き付けで生成された有向グラフは強連結である。

証明

帰納法により証明する。まず手順 1 で向き付けされた閉路 C が強連結であることは自明である。手順 2-3 でグラフ G' に路 P_i ($i \geq 0$) が付加された場合、強連結であると仮定し、新たに路 P_{i+1} を付加する場合でも強連結性が保たれることを示す。

路 P_{i+1} 上の頂点 $v_j, v_k \in V_{P_{i+1}}$ において、 $j < k$ であれば路 P_{i+1} 上を

$$v_j \rightarrow v_{j+1} \rightarrow \dots \rightarrow v_k \quad (8)$$

と遷移することで v_j から v_k は到達可能であり、特に v_1 からは路 P_{i+1} 上の全ての頂点に到達可

能であり, $v_{n_{i+1}}$ へは路 P_{i+1} 上の全ての頂点から到達可能である. G' は強連結なので G' 上の任意の頂点から $v_1 \in V' \cap P_{i+1}$ は到達可能であるので, 路 P_{i+1} 上の全ての頂点へも到達可能である. また G' の強連結性により, $v_{n_{i+1}} \in V' \cap P_{i+1}$ から G' 上の全ての頂点に到達可能なので, P_{i+1} 上の任意の頂点から G' 上の全ての頂点へ到達可能である. 以上より, 路 P_{i+1} 上の頂点 $v_j, v_k \in V_{P_i}$ において, $j > k$ の場合でも $v_j \rightarrow \dots \rightarrow v_{n_{i+1}} \rightarrow (G' \text{ 上の頂点}) \rightarrow v_1 \rightarrow \dots \rightarrow v_k$ と遷移することで v_j から v_k は到達可能である. 以上よりどの2頂点間もお互いに到達可能であり, P_{i+1} を加えた G' も強連結性が保たれている. □

2.3 閉路を利用したランダムウォーク

2点連結グラフ $G = (V, E)$ に閉路循環向き付けをして得られる有効強連結グラフ $G' = (V, E')$ に対し, 次のように頂点 $v_i \in V$ から $v_j \in V$ への遷移確率を決める.

$$p_{ij} = \begin{cases} \frac{1}{\text{outdeg}(v_i)} (e(i, j) \in E') \\ 0 (\text{その他}) \end{cases} \quad (9)$$

outdeg(v_i): 頂点 v_i の G' における出次数

この遷移確率行列に従うランダムウォークでは2点連結グラフを構成する路の数が定数であれば, $O(n)$ の全訪問時間を達成できる, しかしながら一般の2点連結グラフに対しては $O(n^2 \log n)$ の全訪問時間となる.

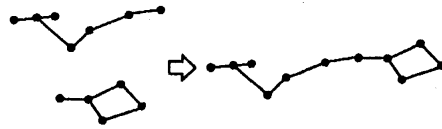
3 2点連結な直並列グラフ

3.1 直並列グラフ

直並列グラフとは, 直列操作と並列操作という2つの操作により再帰的に構成できるグラフである. 2つの端点に s, t の識別子が付けられたグラフ (two-terminal labeled graph TTLG) を (G, s, t) と表す. ここで $G = (V, E)$ は多重グラフである. また最小の TTLG は2つの頂点とそれを結ぶ1つの辺からなるグラフである.

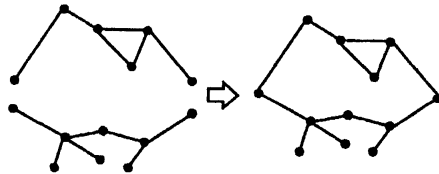
3.1.1 直列操作

2つの TTLG $((V_1, E_1), s_1, t_1), ((V_2, E_2), s_2, t_2)$ に対して $t_1 = s_2$ で直列操作を行うと, TTLG $((V_1 \cup V_2, E_1 \cup E_2), s_1, t_2)$ が得られる. ここで $V_1 \cap V_2 = \{t_1\}$ である.



3.1.2 並列操作

2つの TTLG $((V_1, E_1), s_1, t_1), ((V_2, E_2), s_2, t_2)$ に対して $s_1 = s_2, t_1 = t_2$ で並列操作を行うと, TTLG $((V_1 \cup V_2, E_1 \cup E_2), s_1, t_1)$ が得られる. ここで $V_1 \cap V_2 = \{s_1, t_1\}$ である.



3.2 直並列操作の拡張

直列操作, 並列操作を拡張した S^* 操作, P^* 操作を定義する. これらの操作により得られるグラフは直並列グラフである.

3.2.1 S^* 操作

グラフ $G = (V, E)$ に新しく頂点 w を追加し, 辺 $(u, v) \in E$ を取り除く. そして新たに辺 $(u, w), (w, v)$ を追加する.



3.2.2 P* 操作

グラフ $G = (V, E)$ 上の辺をなす 2 頂点 u, v に対し, 辺 (u, v) を 1 本追加する. ((u, v) は多重辺となる.)



3.3 2点連結グラフとの関連

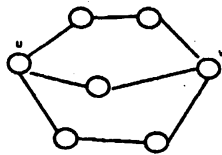
多重グラフが直並列グラフであることと, その多重グラフの 2 点連結成分が閉路に S^* 操作と P^* 操作を繰り返して得られることが等価であることが知られている. このことから多重グラフが二点連結な直並列グラフならば, その多重グラフが閉路に S^* 操作と P^* 操作を繰り返して得られることと等価である.

4 2点連結な直並列グラフ上のランダムウォークの高速化

直並列グラフのグラフ構造は再帰的に定義されているため, グラフを部分構造に分割し, 再帰的にグラフの遷移確率行列を決定していく.

4.1 2点連結な直並列グラフの基本構造

2点連結な直並列グラフの基本構造を 2 点 u, v とそれらをつなぐ $r (\geq 1)$ 本のパスからなる無向グラフと定義し, $B(u, v)$ と表す.



無向な直並列二点連結グラフ $G(V, E)$ が基本構造 $B(u, v)$ を含み, $B(u, v)$ が頂点 u, v のみで G の他の部分と接続しているとする.

G 上で以下を満たす遷移確率行列を考える.

- (1) 粒子は現在の頂点に滞在しない.

- (2) 粒子の遷移の方向はある閉路循環向き付けに従う.

- (3) 頂点 u から v への遷移確率に従ったパスが存在して, 全てが B に包含されている.

- (4) 頂点 v から u への遷移確率に従ったパスが存在して, 全てが B に包含されていない

- (5) 頂点 u では粒子が $\frac{1}{r}$ の確率で B に属する頂点に遷移する.

このとき $B(u, v)$ に対して, 以下のパラメータを定義する.

$c(B)$: 粒子が部分グラフ $B(u, v)$ の頂点を全て訪問するまでに, 頂点 u を通過する回数の期待値.

$t(B)$: 頂点 u にある粒子が部分グラフ $B(u, v)$ 上のパスを通り, 頂点 v に到達するまでの遷移数の期待値.

u, v 間に r 本のパスがあり, それぞれの路の長さの平均が l であるとき, 以下ようになる.

$$c(B) = rH_r$$

$$= r \sum_{i=1}^r \frac{1}{i}$$

$$\sim r \log r$$

$$t(B) = l$$

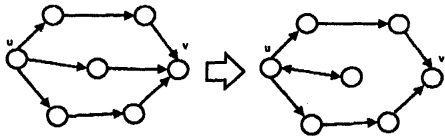
$B(u, v)$ に含まれる頂点数を n_B とする.

このとき $B(u, v)$ に含まれる $u-v$ パス $i (1 \leq i \leq r)$ で, その長さ l_i が閾値 $\sqrt{n_B}$ より短いものについて, 次のようにパス i 上の頂点での遷移確率を変更する.

パス i 上の頂点を u 側から $w_i^0, w_i^1, \dots, w_i^{l_i}$ とする. ただし $w_i^0 = u, w_i^{l_i} = v$ である.

- (1) $P(w_i^0, w_i^1) = \frac{1}{l_i}$
- (2) $P(w_i^j, w_i^{j+1}) = \frac{1}{2}, (1 \leq j \leq l_i - 2)$
- (3) $P(w_i^{l_i-1}, w_i^{l_i}) = 0$
- (4) $P(w_i^j, w_i^{j-1}) = \frac{1}{2}, (1 \leq j \leq l_i - 2)$
- (5) $P(w_i^{l_i-1}, w_i^{l_i-2}) = 1$

このような遷移確率の変更が行われたパスを艶と呼ぶ. また閾値以上の長さのパスを艶に変更する操作をパスの切断と呼ぶ.



各髭の長さは $\sqrt{n_B}$ 未満であり、各髭の端末点を訪れるまでに、粒子が頂点 u を訪れる回数の期待値は $O(\sqrt{n_B})$ である。また頂点 u 上の粒子が、髭以外のパスに侵入するまでに必要な遷移数の期待値は、全ての髭に含まれる頂点数が $O(n_B)$ なので $O(n_B)$ である。(補題 4.1)

一方、切断の行われぬパスの数は $\sqrt{n_B}$ 本以下であり、粒子がパス上の頂点を全訪問するまでにパス側に侵入する回数の期待値は $O(\sqrt{n_B} \log n_B)$ である。また粒子がパス側に侵入してから頂点 v に到達するまでの遷移数は $O(n_B)$ である。

以上より切断操作後のパラメータ $c(B), t(B)$ は以下ようになる。

$$c(B) = O(\sqrt{n_B} \log n_B)$$

$$t(B) = O(n_B)$$

4.1.1 補題 4.1

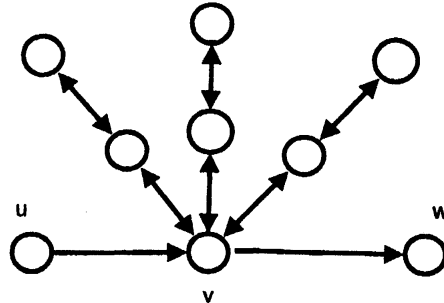
次のような無向グラフ $G = (V, E)$ があるとす

- (1) 頂点 u, v, w があり、辺 $(u, v), (v, w)$ が存在している。
- (2) 頂点 v に r 本の道グラフの端末点が接続している。
- (3) i 番目の道グラフを P_i とすると、辺の数が l_i で、頂点が $r_i^0, r_i^1, \dots, r_i^{l_i}$ の順に並んでおり、 $r_i^0 = v$ とする。

このとき次のような遷移確率を与える

- (1) $p(u, v) = 1, p(v, u) = 0$
- (2) $p(v, w) = \frac{1}{r+1}, p(w, v) = 0$
- (3) $p(v, r_i^1) = \frac{1}{r+1} \quad (1 \leq i \leq r)$
- (4) $p(r_i^j, r_i^{j+1}) = \frac{1}{2} \quad (1 \leq i \leq r, 1 \leq j \leq l_i - 1)$
- (5) $p(r_i^j + 1, r_i^j) = \frac{1}{2} \quad (0 \leq i \leq r, 1 \leq j \leq l_i - 1)$

$$l_i - 2) \quad (6) p(r_i^{l_i}, r_i^{l_i-1}) = 1 \quad (1 \leq i \leq r)$$



今、頂点 u にある粒子が、頂点 w に訪れる前に、頂点 $r_i^{l_i}$ を訪れるを K_i とすると、以下の式で表すことが出来る。

$$K_i = \frac{1}{r+1} \cdot \frac{1}{l_i} + A_i \cdot \frac{1}{r+1} \cdot \frac{1}{l_i} + A_i^2 \cdot \frac{1}{r+1} \cdot \frac{1}{l_i} + \dots$$

$$= \frac{1}{r+1} \cdot \frac{1}{l_i} \cdot \frac{1}{1 - A_i}$$

ここで A_i は粒子が i 番目以外の道グラフに侵入し、再度頂点 v を訪れる確率と、 i 番目の道グラフ進んで先端に到達せずに頂点 v に戻ってくる確率の和である。

$$A_i = \frac{r-1}{r+1} + \frac{1}{r+1} \cdot (1 - \frac{1}{l_i})$$

各道グラフの長さが 1 以上でなので、

$$\frac{r-1}{r+1} \leq A_i < \frac{r}{r+1}$$

よって K_i は、

$$\frac{1}{2l_i} \leq K_i < \frac{1}{l_i}$$

また頂点 u にある粒子が頂点 w に到達するまでの遷移数の期待値 D は、以下の式で計算できる。

$$D = \frac{1}{r+1} + \frac{1}{r+1} \cdot (D_1(l_1) + 1) + \dots + \frac{1}{r+1} \cdot (D_1(l_r) + 1)$$

$$= 1 + \frac{1}{r+1} \sum_{i=1}^r D_1(l_i)$$

ここで, $D_j(l_i)$ は頂点 r_i^j にある粒子が w に到達するまでの時間をとしており, $D_1(l_i)$ は以下の式で表せる.

$$D_1(l_i) = D + 2l_i + 1$$

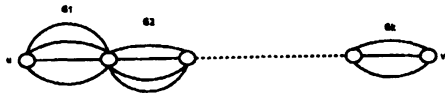
よって D の式から $D_1(l_i)$ を消去して,

$$D = 1 + \frac{r}{r+1}D + \frac{r}{r+1} + \frac{2}{r+1} \sum_{i=1}^r l_i$$

$$D = 1 + 2 \sum_{i=1}^r (l_i + 1)$$

4.2 部分構造の直列操作

2点連結な直並列グラフの k 個の部分構造 B_1, B_2, \dots, B_k を直列に接続して新しい部分構造 B を作ることを考える.



このとき各 $B_i (1 \leq i \leq k)$ のパラメータを $c(B_i), t(B_i)$ とし, それらを直列に連結してできる B のパラメータを $c(B), t(B)$ とする. $t(B)$ は以下の式で表すことができる.

$$\begin{aligned} t(B) &= \sum_{i=1}^k t(B_i) \\ &= O(n_B) \end{aligned}$$

また $c(B)$ については $\max_{1 \leq i \leq k} \{c(B_i)\}$ を達成する部分構造を B_j とすると, B_j が k 個直列に接続した部分構造を B^* とすると,

$$\begin{aligned} c(B) &\leq c(B^*) \\ &= O\left(\log \left(\frac{\sqrt{n_{B_j}}}{\sqrt{n_{B_j}-1}}\right) k \sqrt{n_{B_j}}\right) \\ &= O(\sqrt{n_{B_j}} \log k \sqrt{n_{B_j}}) \\ &= O(\sqrt{n_B} \log n_B) \end{aligned}$$

4.3 部分構造の並列操作

2点連結な直並列グラフの2つの部分構造 B_1, B_2 を並列に接続して新しい部分構造 B を作ることを考える. このとき B について以下のことが成り立つ.

$$(1) \text{ 髭の長さ: } \max\{\sqrt{n_{B_1}}, \sqrt{n_{B_2}}\} \leq \sqrt{n_B}$$

$$(2) \text{ 髭に含まれる頂点数: } n_B \text{ 個以下}$$

$$(3) \text{ パスの数: } \sqrt{n_{B_1}} + \sqrt{n_{B_2}} = O(\sqrt{n_B})$$

$$(4) \text{ パス長: } \max\{n_{B_1}, n_{B_2}\} \leq n_B$$

以上より B のパラメータ $c(B), t(B)$ は, 以下のようになる.

$$c(B) = O(\sqrt{n_B} \log n_B)$$

$$t(B) = O(n_B)$$

4.4 グラフ全体の全訪問時間

頂点 v から u へのパスを含む部分グラフ $G \setminus B$ も直並列操作で構築でき, パラメータ $c(G \setminus B), t(G \setminus B)$ が B と同様に求まり, 全訪問時間は $c(B) + c(G \setminus B)$ と $t(B) + t(G \setminus B)$ の積のオーダーで表せる.

したがって, 直並列操作で得られる頂点数 n の2点連結グラフでランダムウォークの全訪問時間を $O(n\sqrt{n} \log n)$ にする遷移確率行列が存在する.

5 まとめと今後の課題

本稿では2点連結な直並列グラフにおいて全訪問時間が $O(n\sqrt{n} \log n)$ を達成するような遷移確率行列が存在することを示した. しかしながら2点連結グラフ全般に対して, 全訪問時間が $O(n^2)$ より高速なランダムウォークを実現する遷移確率行列は分かってなく, 2点連結グラフ上のランダムウォークの高速化の実現, もしくは全訪問時間が $\Theta(n^2)$ であることの証明が今後の研究課題として挙げられる.

参考文献

- [1] D.J.Aldous: "On the time taken by random walks on finite groups to visit every state", *Z. Wahrsh. verw. Gebiet*, Vol. 62, pp. 361-393, 1983.
- [2] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász, and C. Rackoff: "Random walks, universal traversal sequences, and the complexity of maze problems", *Proc. 20th Ann. Symposium on Foundations of Computer Science*, pp. 218-223, 1979.
- [3] Satoshi Ikeda, Izumi Kubo, Norihiro Okumoto, and Masafumi Yamashita: "Impact of Local Topological Information on Random Walks on Finite Graphs", *Proceedings of Thirtieth International Colloquium on Automata, Languages and Programming*, pp. 1054-1067, 2003.