

ネットワーク上のランダム・ウォークの到達時間について

理化学研究所 脳科学総合研究センター 増田 直紀¹

1 はじめに

現実世界のネットワークは、正方格子、ケーリー・ツリー、ランダム・グラフなどの簡単なグラフで表せないような、複雑かつ秩序ある構造を持つ [2]。このようなネットワーク上のランダム・ウォークは、例えばネットワーク上のサーチという応用と関連している [4]。サーチとは、ネットワーク上を歩いて目的地に到達しようとするものであり、WWWの検索エンジンのイメージがあてはまる。

本発表では、次数（各頂点が持つ枝の数）の分布がランダム・ウォークの振舞いに与える影響について紹介する [3]。ネットワークのモデルとしては、ゴルトン・ワトソン木（一般化ランダム・グラフ）を用いる。このグラフの特徴は、無限に広がり、ループがないことである。ループがないことによって、ランダム・ウォークに限らずネットワーク上で起こる諸事情の解析は非常に簡単になる。ただし、実際のネットワークにはループが多くある（クラスター性）ので、その点では現実性を損なう。

2 モデルと解析結果

図1に模式的に示すように、ゴルトン・ワトソン木では頂点次数 k が頂点ごとに独立にある確率密度 p_k に従って分布している。 p_k はネットワークの次数分布である。ただ、枝をたどった先には次数の大きい頂点（ハブ）がついていやすい事実を取り入れる場合は、 p_k を $kp_k / \sum_{k'} k' p_{k'}$ でおきかえる必要がある [2, 第8章]。ここでは、この修正を考慮しない。

グラフ上のある1点を原点 O と定義して、 O を起点とする離散時間ランダム・ウォークを考える。グラフにループがないので、ウォーカーが O に初めて帰還するのは、必ず偶数時間後である。 q_n を、時刻 $2n$ でウォーカーが初めて O に帰る確率とする。 $\{q_n\}$ の生成関数を $Q(z)$ とする。すなわち、

$$Q(z) \equiv \sum_{n=0}^{\infty} q_n z^n, \quad (1)$$

¹masuda@brain.riken.jp : 本発表は横浜国立大学・今野紀雄氏との共同研究に基づく

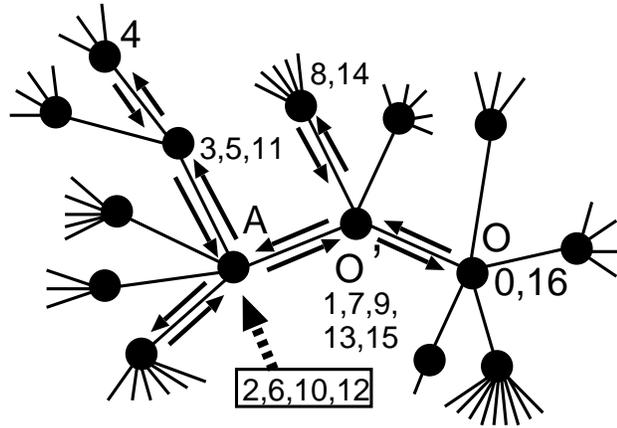


図 1: ゴルトン・ワトソン木上のランダム・ウォーク

ただし、 $q_0 = 0$ に注意する。

ランダム・ウォーカーが O から出発して O に初めて帰還するまでの時間について、近似公式を導きたい。そのためには、再帰的な関係に注目する。図 1 で考えよう。 O からスタートして、時刻 $2n$ で初めて O に戻るとする。図 1 では $n = 8$ である。時刻 1 でウォーカーが O に隣接する頂点 O' に動くとする。 O' の次数は $\{p_k\}$ に従って分布している。時刻 $2n (= 16)$ で O に戻ってくるには、時刻 $2n - 1 (= 15)$ で O' にいないとしない。そうなってれば、最後のステップで確率 $1/k$ で O' から O へ戻れる。時刻 1 と時刻 $n - 1$ の間 ($2n - 2$ ステップだけある) は、ウォーカーは O には訪れずに歩き続ける。 O を訪れないということは、 O' を根とする (O と反対側の) 部分木の中を歩いていなければならない。 O' から一步動くときに、 O に戻らずにこのような部分木に入る確率は $(k - 1)/k$ である。この場合、ウォーカーはいずれ O' に戻ってきて、1 回の「部分木ウォーク」が終わる。時刻が $2n - 1$ より小さかったら、再び「部分木ウォーク」を始める。

O' を起点とする「部分木ウォーク」の回数を a とする。図 1 の場合は、 $a = 4$ である。各「部分木ウォーク」の長さはやはり偶数なので、 $2n_i$ ($1 \leq i \leq a$) とする (図 1 では $n_1 = 3$, $n_2 = 1$, $n_3 = 2$, $n_4 = 1$)。 $2n_i$ の和は $2n - 2 (= 14)$ となる。

ここで、再帰性を利用して、各部分木ウォークの長さ (O' から出発して O' に戻る) の分布は、元のランダム・ウォークが O から出発して O に初めて戻るまでの時間の分布に等しいと考える。実際には、 a 個の部分木は独立に選ばれないが (O' の同じ近傍が選ばれた場合)、これを独立と考えることが近似である。実際には、ネットワークの平均次数が小さすぎない限り、近似誤差は小さい。

今まで述べたことを定式化すると、以下のようになる。

$$\begin{aligned}
q_n &= \sum_{k=1}^{\infty} p_k \sum_{a=0}^{n-1} \sum_{\sum_{i=1}^a n_i = n-1, n_i \geq 0, 1 \leq i \leq a} \frac{k-1}{k} q_{n_1} \frac{k-1}{k} q_{n_2} \cdots \frac{k-1}{k} q_{n_a} \frac{1}{k} \\
&= \sum_{k=1}^{\infty} \frac{p_k}{k} \sum_{a=0}^{\infty} \left(\frac{k-1}{k} \right)^a \sum_{\sum_{i=1}^a n_i = n-1, n_i \geq 0, 1 \leq i \leq a} \prod_{a'=1}^a q_{n_{a'}}, \quad (2)
\end{aligned}$$

式 (2) より

$$Q(z) = z \sum_{k=1}^{\infty} \frac{p_k}{k - (k-1)Q(z)} \quad (3)$$

とわかる。 $Q(1) = 1$ は式 (3) を満たすが、ゴルトン・ワトソン木上のランダム・ウォークが過渡的 $Q(1) < 1$ であるという事実に反するので除く。

$Q(1) < 1$ となる解を探すために式 (3) の右辺を展開すると

$$Q(z) = \frac{zM[Q(z)]}{Q(z)} \quad (4)$$

を得る。ただし、

$$M(z) \equiv \sum_{n=1}^{\infty} m_n z^n, \quad m_n \equiv \sum_{k=1}^{\infty} \frac{(k-1)^{n-1}}{k^n} p_k \quad (5)$$

はネットワーク構造にのみ依存する量である。ラグランジュの反転公式を用いると

$$Q(z) = \sum_{n=1}^{\infty} \frac{z^n}{n!} \left\{ \frac{d^{n-1}}{du^{n-1}} \left[\left(\frac{M(u)}{u} \right)^n \right] \right\}_{u=0}, \quad (6)$$

$$M(u)^n = \sum_{n'=n}^{\infty} u^{n'} \sum_{\lambda \vdash n', \sum_{l=1}^{\infty} i_{\lambda}(l) = n} \frac{n!}{\prod_{l=1}^{\infty} i_{\lambda}(l)!} \prod_{l=1}^{\infty} m_l^{i_{\lambda}(l)}, \quad (7)$$

となる。 $\sum_{\lambda \vdash n'}$ は n' の自然数への分割の種類に関する和である。一般論として、分割 λ は、 $\lambda = (1^{i_{\lambda}(1)} 2^{i_{\lambda}(2)} \dots)$ と表される。1 が λ に $i_{\lambda}(1)$ 個含まれ、2 が $i_{\lambda}(2)$ 個含まれ... という意味である。分割の定義より $\{i_{\lambda}(1), i_{\lambda}(2), \dots\}$ ($\lambda \vdash n'$) は $\sum_{l=1}^{\infty} l i_{\lambda}(l) = \sum_{l=1}^{n'} l i_{\lambda}(l) = n'$ を満たす。式 (7) によると、分割の中でも n パーツあるものだけについての和がとられる。式 (7) を用いて式 (6) を評価すると以下のように q_n の生成関数、ひいては q_n が得られる。

$$Q(z) = \sum_{n=1}^{\infty} z^n \sum_{\lambda \vdash 2n-1, \sum_{l=1}^{\infty} i_{\lambda}(l) = n} c_{\lambda}^{(n)} \prod_{l=1}^{\infty} m_l^{i_{\lambda}(l)}, \quad c_{\lambda}^{(n)} = \frac{(n-1)!}{\prod_{l=1}^{\infty} i_{\lambda}(l)!}. \quad (8)$$

3 例

(A) ケーリー・ツリーは、各頂点の次数が d であるゴルトン・ワトソン木である。 $p_k = \delta_{k,d}$ を式 (3) に代入すると、 $Q(z) = 1/(d - (d-1)Q(z))$ となる。 $Q(1) = 1$ でない解は、 $Q(z) = (d - \sqrt{d^2 - 4(d-1)z})/2(d-1)$. カタラン数 $D_n \equiv {}_{2n}C_n/(n+1)$ の生成関数 $S(z) = (1 - \sqrt{1-4z})/2z$ との類似を利用すると $q_n = (d-1)^{n-1}D_{n-1}/d^{2n-1}$ とわかる。

(B) ランダム・グラフでは、次数分布は $p_k = \lambda^k e^{-\lambda}/k!$ のポアソン分布に従う。この場合は数値計算をできる。ケーリー・ツリーの結果から類推してわかるのだが、ランダム・グラフでは q_n は n について指数的に減る。 n が増えると q_n が急速に小さくなるので、 q_n を直接の数値計算で求めることは現実上不可能である。本研究のヤング図形 (整数の分割) の表式を使うと、大きな n までの q_n が求まる。近似誤差は概して小さい。

(C) バラバシとアルバートが提案した B A モデル [2, 第 6 章] では $p_k = \mathcal{N}k^{-3}$ ($k \geq m$)、 $p_k = 0$ ($k < m$) である (\mathcal{N} は正規化定数)。B A モデル上のランダム・ウォークが出発点に初めて戻ってくるまでの時間はやはり指数的に減衰することが、数値計算の結果として言える。なお、減衰の仕方は同程度の平均次数を持つケーリー・ツリーよりもランダム・グラフの方が遅く、B A モデルではさらに遅い。同じ平均次数の場合、次数分布の非一様性が大きいほど原点に帰ってきやすいことが示唆される。なお、パーコレーション、コンタクト・プロセス等の伝搬過程一般について、次数分布の散らばりが大きいほど伝播が起こりやすいことが知られている [2, 第 8 章]。このことと、次数分布が散らばるほどウォーカーが再帰しやすいことは、双対性で結びついている [1, 第 6 章]。

参考文献

- [1] R. B. Schinazi, *Classical and Spatial Stochastic Processes* (Birkhäuser, Boston, 1999).
- [2] 増田直紀, 今野紀雄. 複雑ネットワークの科学, 産業図書 (2005).
- [3] N. Masuda, N. Konno. Return times of random walk on generalized random graphs. *Physical Review E* **69**, 066113 (2004).
- [4] L. A. Adamic, R. M. Lukose, A. R. Puniyani, B. A. Huberman. Search in power-law networks. *Physical Review E* **64**, 041635 (2001).