

Title	スケールフリーグラフ上における局所情報を用いたランダムウォークについて
Author(s)	平山, 亮; 上原, 隆平
Citation	情報処理学会研究報告 : アルゴリズム研究会, 2006-AL(107-6): 31-37
Issue Date	2006-07
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/3273
Rights	<p>社団法人 情報処理学会, 平山亮 / 上原隆平, 情報処理学会研究報告 : アルゴリズム研究会, 2006(71), 2006, 31-37. ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.</p>
Description	

スケールフリーグラフ上における 局所情報を用いたランダムウォークについて

平山 亮[†], 上原 隆平[‡]
北陸先端科学技術大学院大学 情報科学研究科

概要

近年, WWW やインターネットをモデル化できるものとして scale free graph が注目されている。また, 局所情報を用いた random walk は, 既存の random walk よりもネットワークを効率よくカバーすることが知られている。本論文では, scale free graph 上で, 局所情報を用いた random walk の最適なパラメータを実験的に求め, 既存の random walk に対する優位性を示した。

キーワード: スケールフリーグラフ, ランダムウォーク。

Random Walks by Local Topological Information on Scale-free Graphs

Ryo HIRAYAMA[†] and Ryuhei UEHARA[‡]
School of Information Science, JAIST

Abstract

Recently, scale free graph model which reproduces WWW graph and internet graph etc. has attracted a lot of attention. Random walk using local topological information is known as a model which covers vertices of network efficiently. In this paper, we study the optimal parameter of a random walk using local topological information on a scale free graph and show that the random walks have advantage against standard random walks.

Keywords: Scale free graph, random walk.

1 はじめに

与えられたグラフ構造の上をランダムに移動するプロセスを random walk と呼ぶ。Random walk は Markov chain など広範な応用を持つ要素技術として, 幅広く研究されている。

近年, 池田らによって, 新しい random walk のモデル [8][9] が提案された。通常の random walk では, それぞれの隣接点に同じ確率で遷移するのに対し, 新しいモデルでは, 隣接点の近傍情報を利用する。比較的単純な拡張であるにも関わらず, 従来の Markov chain よりもずっと効率良くネットワーク全体をカバーすることが知られている。より正確には, 彼らの新しいランダムウォークモデルでは, 近傍の情報を利用したポテンシャル関数によって決まる遷移確率 $p^{(\beta)}$ にしたがって, 次に移動する先を決める。池田らの結果は一般のグラフにおける最悪の場合を考えるため, グラフ構

[†]hira-r@jaist.ac.jp

[‡]uehara@jaist.ac.jp

造には特別な仮定は置いていない。 β は隣接点を評価する割合を設定する定数であり、一般のグラフの場合は $\beta = 0.5$ の時に良い性能が得られることが知られている。

一方、WWWやインターネットをモデル化できるものとして、scale free graphが注目を集めている。これは従来のErdős-Rényiによる一様な構造を持つRandom Graphモデル[2]とは違い、非一様な構造を持っており、さまざまな現実の社会ネットワークをモデル化していると考えられている。scale free graphではpower lawと呼ばれる法則が成立しており、このpower lawを実現できるいくつかのscale free graphモデルがすでに知られている。

本研究では、こうしたscale free graphの非一様性を積極的に利用することで、池田らの新しいrandom walkをより効率化することを目標とした。

Scale free graphモデルにはいくつかの異なるモデルが知られている。まず始めに、最も広く受け入れられているpreferential attachment graphモデルを採用する。このモデルでは、頂点の次数を制御するための整数パラメータ m を調節することで、疎なグラフから密なグラフまで生成することができる。また生成されるグラフは無向グラフとなる。

まず始めに、池田らのrandom walkモデルにおけるポテンシャル関数のパラメータ β と、scale free graphにおけるパラメータ m とグラフのサイズとの関係を実験的に評価する。実験の結果、 β の最適値は、グラフのサイズにはほとんど依存せず、 m の関数とみなすことができることがわかった。本論文では具体的な m に対する β の最適値と、その場合の改善の割合を実験的に示した。

Scale free graphでは、頂点の次数の分布はpower lawに従う。そのため、次数の分布は指数関数的に極端に偏っている。この事実から、次数の分布を線形に補正することでより良い振舞をするポテンシャル関数が設計できる、と予想される。この予想から、池田らのrandom walkにおけるポテンシャル関数を次数の関数ではなく次数の対数の関数として実験を行なった。その結果、小さな m に対しては通常の次数のポテンシャル関数よりもさらに効率が良くなることを示すことができた。

Preferential attachment graphモデルでは、生成されるグラフは無向グラフとなる。これは、例えばWebグラフなどのネットワークのモデルとしては不適切である。power lawに従った有向グラフを生成するモデルとしてpreferential attachment graphモデルの一般化であるBollobásらのモデルが知られている。しかしこうしたpower lawに従ったrandomな有向グラフでは、強連結性が一般には保証できない。むしろ入次数を持たない多数の頂点を含んでしまうので、到達性すら確保できない。これに対するrandom walkモデルとして、“Back Button”を持ったRandom walkモデル[7]がFaginらによって提案されている。また、代表的なサーチエンジンGoogleでは、Webページ P から「そのページ P にリンクを張っているページ」を逆にたどることができる。以上のことから、power lawを見たす有向なscale free graphでは「有向辺を逆に辿ることができるrandom walk」を考えることが応用上重要である、と考えられる。

本論文ではBollobásらのpower lawに従う有向グラフモデル上で、池田らのrandom walkモデルにおけるポテンシャル関数に、有向辺を順方向に辿る確率と逆方向に辿る確率とを取り入れたrandom walkを提案し、効率の良いパラメータを実験的に確かめた。

2 Random Walk とは

2.1 既存の Random Walk

random walkは、与えられたグラフ構造の上で確率的に選択された頂点に移動していくプロセスである。従来の標準的なrandom walkでは、次に移動する頂点は隣接点中から同じ確率で選択

される。したがって、グラフ $G = (V, E)$ の遷移行列 $P = (p_{uv})_{u,v \in V} \in [0, 1]^{V \times V}$ は

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

によって定義される。ただし $N(u)$ は頂点 u の隣接頂点集合で、 $\text{deg}(u) = |N(u)|$ である。

カバータイムとは、連結グラフ上の random walk で全ての頂点を訪れるまでに必要なステップ数の期待値である。標準的な random walk におけるカバータイムの上界は、 $n = |V|$, $m = |E|$ を用いて $2m(n-1) = O(n^3)$ で表される。また、下界は lollipop グラフ L_n 上で $\Omega(n^3)$ で与えられるので、従来の random walk の一般のグラフにおけるカバータイムは $\Theta(n^3)$ である。

2.2 池田らの Random Walk

従来の random walk は遷移確率がそれぞれの頂点 u における次数 $\text{deg}(u)$ のみに依存しているのに対して、池田らによる新しい random walk [9] は、隣接点の次数情報 $\text{deg}(v \in N(u))$ を利用する。このモデルにおける遷移行列 $P^{(\beta)} = (p_{uv}^{(\beta)})_{u,v \in V}$ は

$$p_{uv}^{(\beta)} = \frac{\exp[-\beta U(u, v)]}{\sum_{\omega \in N(u)} \exp[-\beta U(u, \omega)]} \text{ for } v \in N(u) \quad (2)$$

によって定義される。この遷移確率は $\beta = 0$ のとき、従来の標準的な random walk の遷移確率と一致し、このモデルが従来のモデルの自然な拡張であることがわかる。

ここで、 $U(u, v)$ はポテンシャル関数である。池田らは、

$$U(u, v) = U(v) = \log(\text{deg}(v)) \quad (3)$$

と定義し、一般のグラフに対しては $\beta = 1/2$ のときカバータイムが $O(n^2 \log n)$ であることを示している。

3 Scale free graph とは

3.1 Scale free graph のもつ性質

近年、WWW、引用文献、神経細胞など現実社会に存在するネットワークを再現できるとして、scale-free power law に従う次数分布を持つグラフモデルが注目されている。Erdős と Rényi による従来からの標準的なランダムグラフモデルである $G(n, p)$ [2] は、辺の出現確率 p 、"固定された" 頂点数 n を持ち、その次数分布は、平均値 $(n-1)p$ を持つポアソン分布または正規分布に従う。従来のランダムグラフモデルと上記のネットワークの間にはギャップがあった。上記のネットワークは、"富む者はますます富む" という言葉が知られているように、次数分布は power law 分布に従うとされ、次数 d の確率分布 $P(d)$ は

$$P(d) \propto d^{-\gamma} \quad (4)$$

となる。ここで、 γ はネットワークの大きさ（頂点数）に依存しない定数で例えば web ページでは $\gamma = 2.1 \pm 0.1$ である。次数分布がこのような scale-free power law に従うグラフモデルを一般に scale free graph モデルと呼ぶ。

3.2 Preferential attachment graph とは

scale free graph モデルにはいくつかの異なるモデル [1][10] が存在しているがその中でも Barabási と Albert による preferential attachment graph モデル [1] は最も広く受けいられているモデルである。このグラフモデルは、グラフに新しい頂点を加えていき、新しい頂点は次数の大きい頂点との間に優先的に辺を張るということを繰り返しグラフを生成していく。

本研究では、Bollobás らによって一般化された次のモデルを採用する [4]。

- 各時刻 t で、頂点 v_t を加え、 v_t とある頂点 u の間に 1 つの辺を張る。 u は次の確率分布に従ってランダムに選択される。

$$\Pr(u = v_i) = \begin{cases} \frac{d_{t-1}(v_i)}{2t-1} & \text{if } v_i \neq v_t, \\ \frac{1}{2t-1} & \text{if } v_i = v_t. \end{cases} \quad (5)$$

ここで、 $d_{t-1}(v)$ は時刻 $t-1$ の終わりでの頂点 v の次数を表す。

- m は任意の定数で、時刻 $t \equiv 0 \pmod{m}$ のとき、 $t, t-1, t-2, \dots, t-m+1$ で加えられた m 頂点を 1 つの頂点にまとめる。

このグラフの次数分布は

$$P(d) = \frac{2m^3}{d^3} \quad (6)$$

となることが示されている。また、このモデルは非常に簡潔なモデルではあるが、セルフループや多重辺が存在する可能性があることには注意しなければならない。

3.3 有向 scale free graph モデル

本研究では、Bollobás らによって提案された次の有向 scale free graph モデル [3] を採用する。各時刻 t で、次のルールにしたがって、グラフは成長していく。

- 確率 α で、頂点 v を加え、 v からある頂点 w へ 1 つの辺を張る。 w は (7) の確率分布に従ってランダムに選択される。
- 確率 β で、ある頂点 v からある頂点 w へ 1 つの辺を張る。 v, w はそれぞれ (8), (7) の確率分布に従ってランダムに選択される。
- 確率 γ で、頂点 w を加え、ある頂点 v から w へ 1 つの辺を張る。 v は (8) の確率分布に従ってランダムに選択される。

$$\Pr(w = v_i) = \frac{d_{in}(v_i) + \delta_{in}}{t + \delta_{in}n(t)} \quad (7)$$

$$\Pr(v = v_i) = \frac{d_{out}(v_i) + \delta_{out}}{t + \delta_{out}n(t)} \quad (8)$$

ここで、 $d_{out}(v)$ は頂点 v の出次数、 $d_{in}(v)$ は頂点 v の入次数、 $n(t)$ は時刻 t での頂点数を表し、 $\alpha, \beta, \gamma, \delta_{in}, \delta_{out}$ は非負の実数で、 $\alpha + \beta + \gamma = 1$ である。

Web グラフは入次数と出次数で異なるべき乗の指数をとることが測定されているが、それを再現する各パラメータの値は、 $\alpha = 0.41, \beta = 0.49, \gamma = 0.1, \delta_{in} = 0, \delta_{out} = 0$ または $\alpha = 0.41, \beta = 0.59, \gamma = 0, \delta_{in} = 0.24, \delta_{out} = 0$ である。

4 実験結果

4.1 次数ベースのポテンシャル関数を用いた場合

式(2)に式(3)を代入すると,

$$p_{uv}^{(\beta)} = \frac{\deg(v)^{-\beta}}{\sum_{\omega \in N(u)} \deg(\omega)^{-\beta}} \text{ for } v \in N(u) \quad (9)$$

となり, 遷移確率は各頂点の次数のべき乗型分布関数に従うことがわかる.

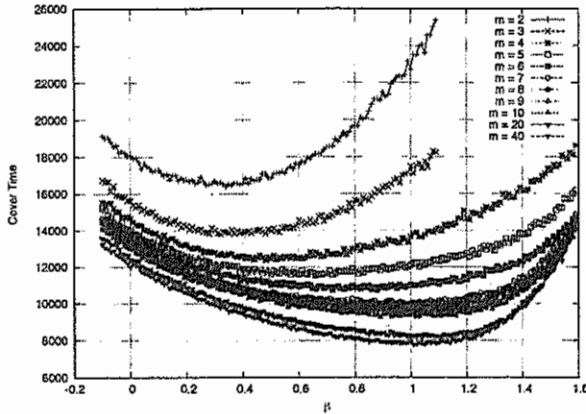


図 1: $n = 1000$

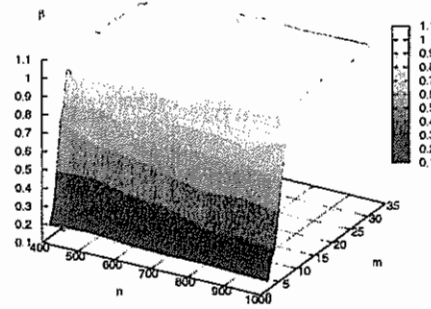


図 2: n vs m vs β

図1は $n = 1000$ の時, いくつかの m に対する結果である. $\beta = 0$ のときの従来の random walk よりも, 最適な β のときはカバータイムが改善されていることがわかる. m によって, カバータイムの最低値を与える β が異なり, m が大きくなるに従って改善の割合は良くなっているのがわかる.

図2は, 図1の最小値を与える β を最小自乗法により求め, n, m の関数として表示したものである. その最適値は, グラフのサイズ n にはほとんど依存せず, m の関数とみなすことができることがわかった.

4.2 次数の対数ベースのポテンシャル関数を用いた場合

次にポテンシャル関数の改善を試みた.

$$U(u, v) = U(v) = \log \log(\deg(v)) \quad (10)$$

と定義すると,

$$p_{uv}^{(\beta)} = \frac{(\log \deg(v))^{-\beta}}{\sum_{\omega \in N(u)} (\log \deg(\omega))^{-\beta}} \text{ for } v \in N(u) \quad (11)$$

となり, 遷移確率は各頂点の次数の対数のべき乗型分布関数に従うことがわかる.

図3は $n = 800, m = 2$ のときの, 次数のべき乗型, 次数の対数べき乗型の比較である. その結果, 元の random walk よりも効率が良くなっていることがわかる.

図4は $n = 800$ のときの, 各 m における次数のべき乗型, 次数の対数べき乗型の比較である. m が比較的小さな値のとき, このポテンシャル関数の改善は効果があることがわかる.

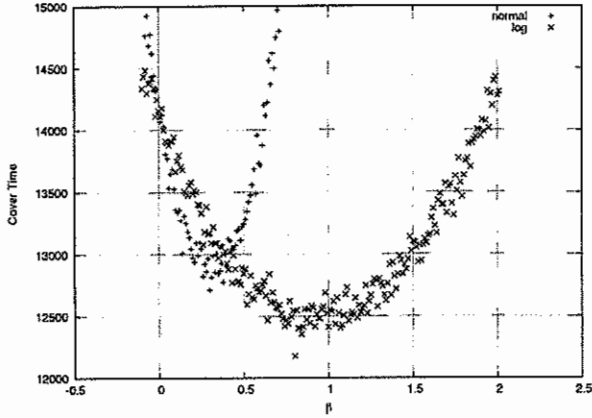


図 3: $n = 800, m = 2$

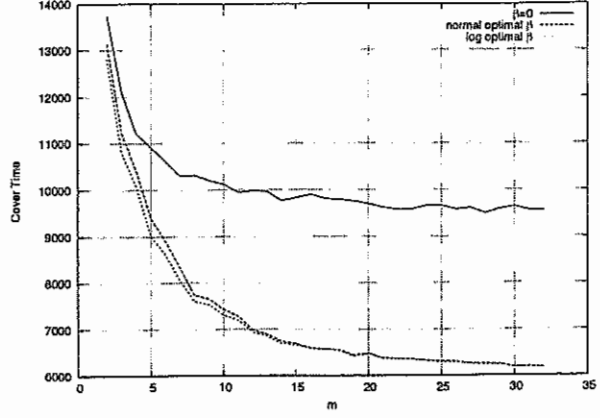


図 4: $n = 800$

4.3 有向 scale free graph モデル上の逆辺をたどることを許した Random Walk

本論文では池田らの random walk モデルで、逆方向に対して確率 $\gamma (0 < \gamma \leq 1)$ で辿ることを許した random walk を提案する。

ここでは、有向グラフ上の次数を (12) で定義し、遷移確率は、順辺に対しては (13)、逆辺に対しては (14) とした。

$$\text{deg}(v) = \text{indegree}(v) + \gamma \cdot \text{outdegree}(v) \quad (12)$$

$$p_{OUTuv}^{(\beta, \gamma)} = \frac{\text{deg}(v)^{-\beta}}{\sum_{\omega \in OUT(u)} \text{deg}(\omega)^{-\beta} + \sum_{\omega \in IN(u)} \text{deg}(\omega)^{-\beta}} \quad (13)$$

$$p_{INuv}^{(\beta, \gamma)} = \frac{\gamma \cdot \text{deg}(v)^{-\beta}}{\sum_{\omega \in OUT(u)} \text{deg}(\omega)^{-\beta} + \sum_{\omega \in IN(u)} \gamma \cdot \text{deg}(\omega)^{-\beta}} \quad (14)$$

Bollobás らの有向 scale free graph[3] において入次数, 出次数に関して Web グラフを再現するとされたパラメータ $\alpha = 0.41, \beta = 0.59, \gamma = 0, \delta_{in} = 0.24, \delta_{out} = 0$ に対して、各 γ における最適な β の値を求め、隣接点の次数情報を使わない random walk ($\beta = 0$) の結果と比較した。

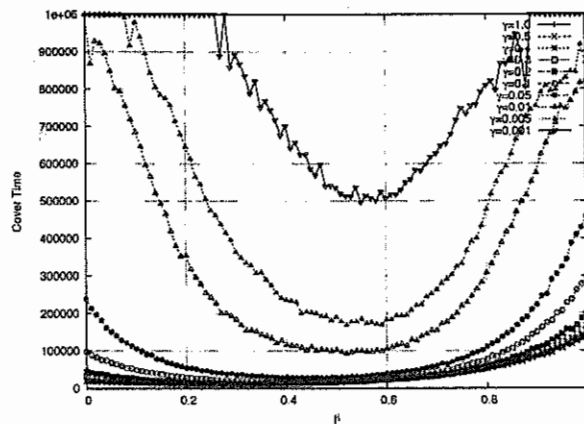


図 5: $t = 1000, \alpha = 0.41, \beta = 0.59, \gamma = 0, \delta_{in} = 0.24, \delta_{out} = 0$

図5は $t = 1000, \alpha = 0.41, \beta = 0.59, \gamma = 0, \delta_{in} = 0.24, \delta_{out} = 0$ のときの、各 γ における結果である。

γ の値が0に近づくとしたがって、局所情報を用いない random walk ($\beta = 0$) と比べて最適な β の cover time の改善の幅が大きくなることがわかる。一方、 γ の値が1 (無向グラフ) に近づくとしたがって、局所情報を用いない random walk と最適な β の cover time の差が相対的に小さくなることがわかる。

5 おわりに

本論文では、scale free graph 上での効率の良い random walk を実験的に示した。scale free graph 上の通常の random walk に関して、Cooper, Frieze はカバータイムなどに対する理論的な解析を行なって、正確な上界や下界を得ている [5][6]。彼らの手法を拡張して、scale free graph 上の池田らの random walk の理論的な解析を行うことが今後の課題である。

謝辞

本研究の遂行にあたり、九州大学大学院システム情報科学研究院山下雅史教授と、山下研究室の定兼邦彦助教授、小野廣隆助手、来見田裕一氏には、数多くの有益な助言やご支援をいただき、深く感謝します。また本研究は文部科学省科学研究費補助金 (基盤研究 (C)18244120) の支援を受けて行なった。

参考文献

- [1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [2] B. Bollobás. *Random Graphs*. Cambridge studies in advanced mathematics, 2001.
- [3] B. Bollobás. Directed scale-free graphs. In *ACM Symposium on Discrete Algorithms (SODA)*, pages 132–139, 2003.
- [4] B. Bollobás, O. Rordan, J. Spencer, and G. Tusnady. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18:279–290, 2001.
- [5] C. Cooper and A. Frieze. The cover time of the preferential attachment graph, manuscript, 2005.
- [6] C. Cooper and A. Frieze. The Cover Time of Two Classes of Random Graphs. In *ACM Symposium on Discrete Algorithms (SODA)*, pages 961–970, 2005.
- [7] R. Fagin, A. R. Karlin, J. Kleinberg, P. Raghavan, S. Rajagopalan, R. Rubinfeld, M. Sudan, and A. Tomkins. Random walks with “back buttons”. In *ACM Symposium on Theory of Computing (STOC)*, pages 484–493, 2000.
- [8] Ikeda, Kubo, Okumoto, and Yamashita. Impact of local topological information on random walks on finite graphs. In *Annual International Colloquium on Automata, Languages and Programming (ICALP)*, pages 1054–1067, 2003.
- [9] S. Ikeda, I. Kubo, and M. Yamashita. Reducing the hitting and the cover times of random walks on finite graphs by local topological information. In *Proceedings of the International Conference on VLSI (VLSI)*, pages 203–207, 2003.
- [10] N. Masuda, H. Miwa, and N. Konno. Geographical threshold graphs with small-world and scale-free properties. *PHYSICAL REVIEW E*, 71:036108, 2005.