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
Author(s)	平山，亮
Citation	
Issue Date	2006-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/2037
Rights	
Description	Supervisor:上原　隆平，情報科学研究科，修士

修 士 論 文

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

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

平山 亮

2006 年 9 月

修 士 論 文

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

指導教官　上原 隆平 助教授

審査委員主査　上原 隆平 助教授
審査委員　金子 峰雄 教授
審査委員　平石 邦彦 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

410204 平山 亮

提出年月: 2006 年 8 月

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スケールフリーグラフが注目されている．また, 局所情報を用いたランダムウォークは, 既存のシンプルランダムウォーク 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, スケールフリーグラフ上で, 局所情報を用いたランダムウォークの最適なパラメータを実験的に求め, 既存のシンプルランダムウォークに対する優位性を示した．

目 次

第 1 章　はじめに **1**

第 2 章　グラフモデル **3**

2.1 グラフ 3

2.1.1 無向グラフ 3

2.1.2 有向グラフ 3

2.1.3 疎なグラフ 3

2.1.4 密なグラフ 4

2.1.5 次数 4

2.1.6 セルフループ 4

2.1.7 多重辺 5

2.1.8 ウォーク 5

2.1.9 トレイル 5

2.1.10 パス 5

2.1.11 到達可能 5

2.1.12 連結 6

2.1.13 強連結 6

2.1.14 弱連結 6

2.1.15 距離 6

2.1.16 直径 6

2.1.17 半径 6

2.2 ランダムグラフ 6

2.2.1 $\mathcal{G}(n, M)$ モデル 7

2.2.2 $\mathcal{G}(n, p)$ モデル 7

2.3 スケールフリーグラフ 7

2.3.1 スケールフリーネットワーク 7

2.3.2 Barabási-Albert の Preferential attachment モデル 8

2.3.3 有向スケールフリーグラフ 8

第 3 章　ランダムウォークモデル **10**

3.1 マルコフチェーン 10

3.1.1 マルコフチェーン 10

3.1.2 既約性 . . . 10
3.1.3 非周期性 . . . 10
3.1.4 定常分布 . . . 10
3.2 ランダムウォーク . . . 11
3.2.1 シンプルランダムウォーク . . . 11
3.2.2 シンプルランダムウォークの定常分布 . . . 12
3.2.3 ヒッティングタイム . . . 12
3.2.4 カバータイム . . . 12
3.2.5 Cooper-Frieze による理論的解析 . . . 13
3.3 局所情報を用いたランダムウォーク . . . 13
3.3.1 池田らのランダムウォーク . . . 13
3.3.2 ポテンシャル関数の改良 . . . 14
3.3.3 有向スケールフリーグラフ上のランダムウォーク . . . 15
3.4 その他のランダムウォーク . . . 15
3.4.1 ページランク . . . 15

第 4 章 計算機実験 **16**
4.1 実装 . . . 16
4.2 無向スケールフリーグラフ上における実験結果 . . . 16
4.2.1 次数べき乗型を用いた場合 . . . 17
4.2.2 次数の対数べき乗型を用いた場合 . . . 25
4.3 有向スケールフリーグラフ上における実験結果 . . . 35
4.3.1 逆辺をたどることを許した場合 . . . 35

第 5 章 おわりに **46**

謝辞 **47**

参考文献 **49**

第1章　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ランダムウォークと呼ぶ. ランダムウォークは, それ自身が, 興味深い研究の対象であるとともに, 乱択アルゴリズムの設計と解析など, 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 [13].

近年, 池田らによって, 新しいランダムウォークのモデル [15][16] が提案された. 通常のランダムウォークでは, 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, 従来のシンプルランダムウォーク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 $\beta = 0.5$ のとき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, スケールフリーグラフ [3] が注目を集めている. これは従来の Erdős-Rényi[5] による一様な構造を持つランダムグラフモデルとは違い, 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. スケールフリーグラフでは 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 スケールフリーグラフ モデルがすでに知られている.

本研究では, こうした スケールフリーグラフの非一様性を積極的に利用することで, 池田らの新しいランダムウォークをより効率化することを目標とした.

スケールフリーグラフモデルにはいくつかの異なるモデルが知られているが, 最も広く受け入れられている preferential attachment グラフモデル [3][7] 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, 池田らのランダムウォークモデルにおけるポテンシャル関数のパラメータ β, スケールフリーグラフにおけるパラメータ m, 及びグラフのサイズ n との関係を実験的に評価する.

次に, ポテンシャル関数の改良を試みる. スケールフリーグラフでは, 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, 池田らのランダムウォーク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う.

Preferential attachment グラフモデルでは, 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 グラフモデルの一般化である Bollobás らのモデル [6] が知られている. しかしこうした power law に従った ランダムな有向グラフでは, 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 ランダムウォークモデルとして, “back button” を持ったランダムウォークモデル [11] 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[1] 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 を満たす有向なスケールフリーグラフでは「有向辺を逆に辿ることができるランダムウォーク」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, 池田らのランダムウォークモデルにおけるポテンシャル関数に, 逆方向に対して確率 $\gamma(0 < \gamma \leq 1)$ で辿ることを許したランダムウォーク提案し, 効率の良いパラメータを実験的に確かめる.

本論文において，以下各章は次のように構成される. 第 2 章 (p.3) では, グラフモデルについて説明する. 特に, スケールフリーグラフについては, 有向グラフを含むいくつかのモデルについて説明する. 第 3 章 (p.10) では, ランダムウォークモデルについて説明する. 局所情報を用いたランダムウォークモデルについては, 池田らのランダムウォークモデルについて説明するとともに, それを改良したランダムウォークモデルについても提案する. 第 4 章 (p.16) では, 各モデルにおける計算機実験の結果を示し, 考察を行う. 第 5 章 (p.46) では, まとめと今後の課題について述べる.

第2章　グラフモデル

2.1　グラフ

ここでは, グラフの基本的な定義, 性質について説明する [4][12][18].

2.1.1　無向グラフ

無向グラフ $G(V,E)$ は, 頂点集合 V と辺集合 E からなる. 辺 $e \in E$ は, 頂点 $v_i, v_j \in V$ の非順序対で, (v_i, v_j) として表される. 頂点 v_i と v_j の間に辺があるとき, v_i と v_j は隣接していると言い, v_i との間に辺を持つ全ての頂点を v_i の隣接点と言う.

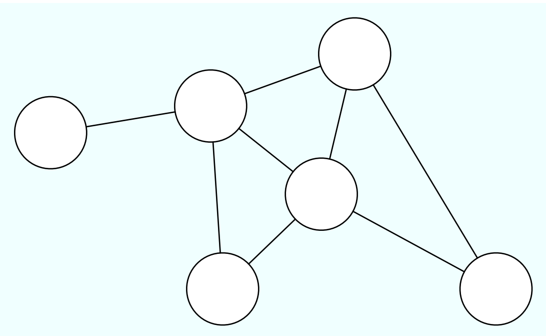

図 2.1: 無向グラフの例

2.1.2　有向グラフ

有向グラフ $G(V,E)$ は, 頂点集合 V と辺集合 E からなる. 辺 $e \in E$ は, 頂点 $v_i, v_j \in V$ の順序対で, (v_i, v_j) として表される. 頂点 v_i から v_j への辺があるとき, v_j は v_i に隣接していると言い, v_i からの辺を持つ全ての頂点を v_i の隣接点と言う.

2.1.3　疎なグラフ

グラフ G において, 辺の数 m が頂点の数 n にくらべて小さいとき, (通常は $m = O(n)$ のとき,) G は疎なグラフであると言う.

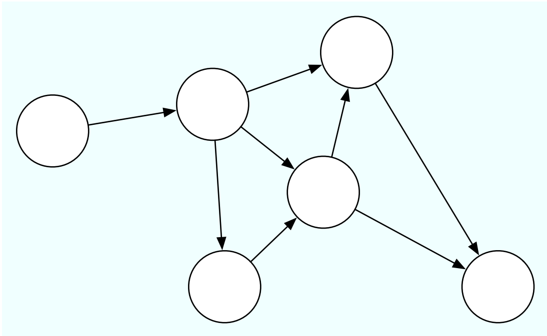

図 2.2: 有向グラフの例

2.1.4 密なグラフ

グラフ G において, 辺の数 m が頂点の数 n にくらべて大きいとき, (通常は $m = \Omega(n^2)$ のとき,) G は密なグラフであると言う.

2.1.5 次数

無向グラフにおいて, 頂点 v と接続する辺の数を頂点 v の次数と呼び, $deg(v)$ と表す.

有向グラフにおいて, 頂点 v に入ってくる辺の数を頂点 v の入次数, 頂点 v から出て行く辺の数を頂点 v の出次数と呼び, それぞれ $in\_deg(v)$, $out\_deg(v)$ と表す. また, $deg(v)$ は, 頂点 v と接続する辺の数を頂点 v の次数を表し, $deg(v) = in\_deg(v) + out\_deg(v)$ となる.

2.1.6 セルフループ

同じ頂点 v を結ぶ辺 (v, v) をセルフループと呼ぶ.

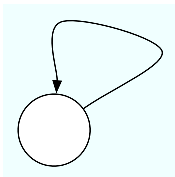

図 2.3: セルフループの例

2.1.7　多重辺

頂点 u と頂点 v とを接続する複数の辺が存在するとき, それらを多重辺と呼び, それらの辺は平行であると言う.

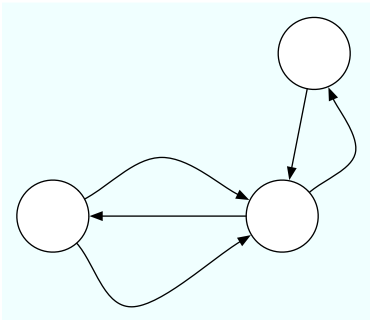

図 2.4: 多重辺の例

2.1.8　ウォーク

ウォークは, 頂点から始まり頂点で終わる, 頂点と辺の連 $v_0, e_1, v_1, \cdots, v_{n-1}, e_n, v_n$ で, 各々の辺は, その両側の頂点を結ぶ辺である.

2.1.9　トレイル

トレイルは, 全ての辺が別々であるウォークである.

2.1.10　パス

パスは, 全ての頂点が別々であるウォークである.

2.1.11　到達可能

頂点 v_i から頂点 v_j へのパスがあるとき, v_j は v_i から到達可能であると言う.

2.1.12　連結

無向グラフ G において, どの頂点もそれ以外の各頂点から到達可能であるとき G は連結であると言い, そのようなグラフを連結グラフと呼ぶ.

有向グラフ G においては, G に対応する無向グラフが連結ならば, G は連結であると言う.

2.1.13　強連結

有向グラフにおいて, 頂点 v_i と頂点 v_j との間に, v_i から v_j への辺と v_j から v_i への辺がともに存在するとき, v_i と v_j は強連結であると言う.

2.1.14　弱連結

無向グラフとみなしたときに v_i と v_j との間に辺が存在し, v_i と v_j が強連結でないとき, v_i と v_j は弱連結であると言う.

2.1.15　距離

グラフ G において, 2 頂点 v_i, v_j の最短パスを $v_i v_j$ の距離と呼び, $d_G(v_i, v_j)$ と表す. ただし, そのようなパスが存在しないとき, $d_G(v_i, v_j) = \infty$ とする.

2.1.16　直径

グラフ G における, 任意の 2 頂点の距離の最大値を G の直径と呼び, $\mathrm{diam}(G)$ で表す.

2.1.17　半径

グラフ G において, そこから他の頂点までの距離の最大値が最も小さくなる頂点を, G の中心と呼ぶ. そのときの距離の最大値を G の半径と呼び, $\mathrm{rad}(G)$ で表す.

2.2　ランダムグラフ

ランダムグラフの理論 [4][5][12] は, Erdős と Rényi によってその基礎が築かれた. Erdős は, 頂点集合 $V = \{1, 2, \ldots, n\}$ の n 頂点のグラフ全体の集合を確率空間 $\mathcal{G}^n$ として定義し, グラフ理論の極値問題に取り組む上で, 確率論的手法が様々な場合において有効であり, 実際にグラフを構築することなく, 欲しい条件を満たすグラフの存在性を示した. この確

率的手法は, グラフ理論のみならず, 応用範囲の広い証明技法として, 離散数字の他の分野においても広く用いられるようになった.

ここでは, 最もよく知られている $\mathcal{G}(n,M)$ モデルと $\mathcal{G}(n,p)$ モデルについて説明する.

2.2.1　$\mathcal{G}(n,M)$ モデル

$\mathcal{G}(n,M)$ の全てのグラフは, n 個の頂点と M 本の辺を持つ. 完全グラフ K_n は, $N=\binom{n}{2}$ 本の辺を持つので $0 \le M \le N$ で, $\mathcal{G}(n,M)$ は $\binom{N}{M}$ 個の要素からなり, 全ての要素は等しい確率 $\binom{N}{M}^{-1}$ で生成される. $G_M = G_{n,M}$ を $\mathcal{G}(n,M)$ 空間から生成されるランダムグラフ, H をある固定された M 本の辺を持つ n 頂点のグラフであるとすると

$$\mathbf{Pr}(G_M = H) = \binom{N}{M}^{-1}. \tag{2.1}$$

2.2.2　$\mathcal{G}(n,p)$ モデル

$\mathcal{G}(n,p)$ の全てのグラフは, n 個の頂点を持ち, 各々の頂点間には, 確率 p $(0<p<1)$ で辺が存在する. 完全グラフ K_n の部分グラフの数は, 2^N 個であるが, $\mathcal{G}(n,p)$ は, その 2^N 個の部分グラフを全て要素に持つ. $G = G_{n,p}$ を $\mathcal{G}(n,p)$ 空間から生成されるランダムグラフ, H をある固定された m 本の辺を持つ n 頂点のグラフであるとすると

$$\mathbf{Pr}(G_p = H) = p^m(1-p)^{N-m}. \tag{2.2}$$

$\mathcal{G}(n,\frac{1}{2})$ では, $\mathcal{G}^n$ 空間における任意の 2 つのグラフが生成される確率は等しい. また, $M \sim pN$ のとき, $\mathcal{G}(n,M)$ モデルと $\mathcal{G}(n,p)$ モデルは非常に似た振る舞いをする.

$\mathcal{G}(n,p)$ モデルの次数分布 $P(k)$ は, 平均値 $\bar{k} = p(n-1)$ を持つ二項分布

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{2.3}$$

となる. さらに, $n \to \infty$ のとき, ポアソン分布

$$P(k) \sim e^{-\bar{k}} \frac{\bar{k}}{k!} \tag{2.4}$$

になる.

2.3　スケールフリーグラフ

2.3.1　スケールフリーネットワーク

World Wide Web を始めとする多くの実在のネットワークは, その頂点の次数分布 $P(k)$ が, ある定数 γ による, べき法則 (power law) 分布

$$P(k) \sim k^{-\gamma} \tag{2.5}$$

に従うという共通の性質を持つ．次数 k はいくらでも大きなサイズをとりうることから，この性質を "スケールフリー" と呼ぶ．実際に，WWW，インターネット，神経細胞，社会ネットワークなどのネットワークは，スケールフリーネットワークであることが報告されている．

従来の Erdős-Rényi のランダムグラフモデルではこれらのスケールフリーネットワークを再現することが難しかった．近年，このスケールフリーを実現するランダムグラフモデルが注目されるようになり，いくつかの異なるスケールフリーグラフモデル [3][17] が提案されている．

2.3.2 Barabási-Albert の Preferential attachment モデル

本論文では，いくつかの異なるスケールフリーグラフモデルの中で，最も広く受け入れられている，Barabási と Albert による preferential attachment モデル [3] を採用する．

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．

Preferential attachment モデルは，Bollobás らによって次のように一般化されている [7]．

- 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．

$$\mathbf{Pr}(u = v_i) = \begin{cases} \frac{d_{t-1}(v_i)}{2t-1} & \text{if } v_i \neq v_t, \\ \frac{1}{2t-1} & \text{if } v_i = v_t. \end{cases} \tag{2.6}$$

  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，$t, t-1, t-2, \cdots, t-m+1$ 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

$$P(k) = \frac{2m^3}{k^3} \tag{2.7}$$

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．

2.3.3 有向スケールフリーグラフ

本研究では，Bollobás らよって提案された次の有向スケールフリーグラフモデル [6] を採用する．このグラフモデルの基本的なアイデアは Barabási らの preferential attachment モデルと同じであるが，Web グラフなどが持つ，入次数と出次数で異なるべき乗の指数をとるという性質を満たした有向スケールフリーグラフを生成することができる．

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 つ辺を張る. w は (2.8) 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. w, v はそれぞれ (2.9), (2.8) 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 (2.9) の確率分布に従ってランダムに選択される.

$$\mathbf{Pr}(w = v_i) = \frac{d_{in}(v_i) + \delta_{in}}{t + \delta_{in} n(t)} \tag{2.8}$$

$$\mathbf{Pr}(v = v_i) = \frac{d_{out}(v_i) + \delta_{out}}{t + \delta_{out} n(t)} \tag{2.9}$$

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, α, β, γ, δ_{in}, δ_{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て, Broder らによると $\gamma_{IN} = 2.1$,$\gamma_{OUT} = 2.7$ である [8]. 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.

第3章　ランダムウォークモデル

3.1　マルコフチェーン

3.1.1　マルコフチェーン

ランダムウォークについて考えるとき, しばしば, マルコフチェーンという概念が使われる [14][18]. マルコフチェーン $\mathcal{M}$ は, 遷移確率行列 $\boldsymbol{P}$ で有限状態集合 S を表す散時間確率過程である. マルコフチェーンは, 1 つの状態から, 離散時刻 $t = 1, 2, \ldots$ で状態遷移していく. すなわち, 全ての $i, j \in S$ に対して, $0 \leq p_{ij} \leq 1$ かつ $\sum_j p_{ij=1}$ である. また, マルコフチェーンの未来の振る舞いは現在の状態にだけ依存し, どのようにして現在の状態になったかには依存しない (*memoryless property*) ので, 遷移確率 p_{ij} は, 現在の状態 i だけに依存することになる.

3.1.2　既約性

有限状態集合 $S = \{s_1, s_2, \ldots, s_k\}$, 状態遷移確率行列 $\boldsymbol{P}$ のマルコフチェーン $\mathcal{M}$ について, 全ての $s_i, s_j \in S$ に関して s_i から s_j への遷移, s_j から s_i への遷移が共に可能ならば, マルコフチェーン $\mathcal{M}$ は既約であると言う. マルコフチェーン $\mathcal{M}$ が既約であるならば, 任意の $s_i, s_j \in S$ に関して, $[\boldsymbol{P}^n]_{i,j} > 0$ となる n を見つけることができる.

3.1.3　非周期性

状態 i の周期 $d(i)$ は, 状態が s_i からスタートして, s_i に戻ってくるまでの数の最大公約数として定義される. $d(i) > 1$ のとき, 状態 s_i は周期的であり, $d(i) = 1$ のとき, 状態 i は非周期である. マルコフチェーン $\mathcal{M}$ の全ての $s_i \in S$ に関して, 状態 s_i が非周期であるとき, マルコフチェーン $\mathcal{M}$ は非周期的であると言う.

3.1.4　定常分布

時刻 t におけるマルコフチェーン $\mathcal{M}_G$ の状態確率分布 $q^{(t)} = (q_1^{(t)}, q_2^{(t)}, \ldots, q_k^{(t)})$ は, 列ベクトルで, その i 成分 $q_i^{(t)}$ は時刻 t で状態が i である確率を表す. 時刻 $t+1$ における確率

分布 $\boldsymbol{q}^{t+1}$ は, 遷移確率行列 $\boldsymbol{P}$ と時刻 t における確率分布 $\boldsymbol{q}^t$ によって次のように表される.

$$\boldsymbol{q}^{t+1} = \boldsymbol{q}^t \boldsymbol{P} \tag{3.1}$$

遷移確率行列 $\boldsymbol{P}$ のマルコフチェーンの定常分布は,

$$\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{P} \tag{3.2}$$

を満たす確率分布 $\boldsymbol{\pi}$ である. 時刻 t における定常分布は, 時刻 $t+1$ においても変化しないので, この定常分布は, マルコフチェーンの定常状態を表している.

マルコフチェーン $\mathcal{M}$ が既約, 非周期的であるとき, マルコフチェーン $\mathcal{M}$ は, ただ 1 つの定常分布 $\boldsymbol{\pi}$ を持つ.

3.2 ランダムウォーク

3.2.1 シンプルランダムウォーク

ランダムウォークは, 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.

シンプルランダムウォークムは, 最も標準的なランダムウォークで, 単にランダムウォークと呼ばれることもある. 連結グラフ $G(V, E)$ 上のシンプルランダムウォークは以下のような, 連続する離散ステップである.

- 頂点 $v_0 \in V$ をスタートする.
- 頂点 v_0 から, v_0 の隣接点の中から同じ確率で選択された頂点 $v_1 \in N(v_0)$ へ移動する.
- 頂点 v_1 から, v_1 の隣接点の中から同じ確率で選択された頂点 $v_2 \in N(v_1)$ へ移動する. 以下同様に繰り返す.

各々のステップにおいてどの頂点が選択されるかは, その前にどの頂点が選択されているか, ということとは完全に独立である.

無向連結グラフ $G = (V, E)$ 上のシンプルランダムウォークをマルコフチェーン $\mathcal{M}_G$ でモデル化すると, 状態集合 S は頂点集合 V, 遷移確率行列 $\boldsymbol{P}$ は $P = (p_{uv})_{u,v \in V} \in [0, 1]^{V \times V}$ となり, 頂点 u から頂点 v への遷移確率 p_{uv} は

$$p_{uv} = \begin{cases} \frac{1}{deg(u)} & \text{if } v \in N(u) \ , \\ 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

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, $deg(u) = |N(u)|$ である.

3.2.2 シンプルランダムウォークの定常分布

連続グラフで, 無向グラフかつ 2 部グラフでないグラフ $G(V, E)$ 上のシンプルランダムウォークの定常分布について考える [14][18].

補題 3.1. グラフ G 上のシンプルランダムウォークにおいて, 全ての $v \in V$ における定常分布 π_v は,

$$\pi_v = \frac{deg(v)}{2m} \tag{3.4}$$

である. ただし, $m = |E|$.

証明. グラフ G は, 連結グラフなので既約であり, 無向グラフありかつ 2 部グラフでないので非周期的であるので, マルコフチェーン $\mathcal{M}_G$ は, ただ 1 つの定常分 $\boldsymbol{\pi}$ を持つ. したがって, $\pi_v = \frac{deg(v)}{2m}$ であれば, $\boldsymbol{\pi P} = \boldsymbol{\pi}$ が成り立つことを示せばよい. $\tilde{\boldsymbol{\pi}} = \boldsymbol{\pi P}$ とすると,

$$\begin{aligned}
\tilde{\pi}_v &= \sum_{u \in N(v)} \pi_u p_{uv} \\
&= \sum_{u \in N(v)} \frac{deg(u)}{2m} \times \frac{1}{deg(u)} \\
&= \sum_{u \in N(v)} \frac{1}{2m} \\
&= \frac{deg(v)}{2m}.
\end{aligned}$$

□

3.2.3 ヒッティングタイム

ヒッティングタイム (hitting time) h_{uv} は, 連結グラフ上のランダムウォークで頂点 u をスタートとして, 最初に頂点 v に訪れるまでに必要なステップ数の期待値である. ここで, 全ての $v \in V$ に対して h_{vv} は,

$$h_{vv} = \frac{1}{\pi_v} = \frac{2m}{deg(v)} \tag{3.5}$$

である.

3.2.4 カバータイム

頂点 u におけるカバータイム (cover time) $\mathcal{C}_u$ は, 連結グラフ G 上のランダムウォークで, u をスタートとして, 全ての頂点を訪れるまでに必要なスッテプ数の期待である. (最大) カバータイム $\mathcal{C}_G$ は,

$$\mathcal{C}_G = \max_{u \in V} \mathcal{C}_u \tag{3.6}$$

である．また, 平均カバータイム $\bar{\mathcal{C}}_G$ は,

$$\bar{\mathcal{C}}_G = \sum_{u \in V} \frac{\mathcal{C}_u}{n} \tag{3.7}$$

である．本研究の計算機実験においては, この平均カバータイムをランダムウォークにおける指標として用いることにする．

シンプルランダムウォーク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, $n/2$ 頂点のクリークと残りの頂点のパスからなる lollipop グラフ L_n 上で $\Omega(n^3)$ で抑えられるので, シンプルランダムウォークのカバータイムは $\Theta(n^3)$ である [15].

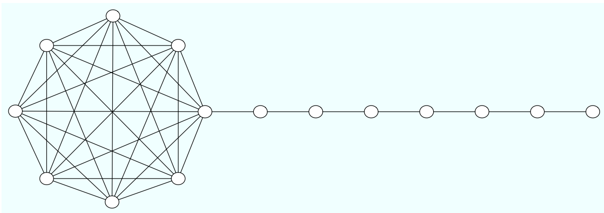

図 3.1: L_{15}

3.2.5 Cooper-Frieze による理論的解析

Cooper と Frieze は理論的解析により, n 頂点, m 本づつ辺が付け加えられる preferenial attachment グラフ $G_m(n)$ におけるシンプルランダムウォークで, カバータイム $\mathcal{C}_{G_m(n)}$ が, 高い確率で,

$$\mathcal{C}_{G_m(n)} \sim \frac{2m}{m-1} n \log n = O(n \log n) \tag{3.8}$$

となることを示している [9][10].

3.3 局所情報を用いたランダムウォーク

3.3.1 池田らのランダムウォーク

従来のシンプルランダムウォークは遷移確率がそれぞれの頂点 u における次数 $deg(u)$ のみに依存しているが, 池田らはさらなる局所情報を利用することによって, ランダムウォークのヒッティングタイムやカバータイムを改善できるのではないかと考え, 隣接点の次数

情報 $deg(v \in N(u))$ を利用した, 新しい局所情報を用いたランダムウォーク [16] を提案した. このモデルにおける遷移行列 $P^{(\beta)} = (p_{uv}^{(\beta)})_{u,v \in V}$ は

$$p_{uv}^{(\beta)} = \frac{\exp[-\beta U(u,v)]}{\sum_{\omega \in N(u)} \exp[-\beta U(u,\omega)]} \text{ for } v \in N(u) \tag{3.9}$$

によって定義される. この遷移確率は $\beta = 0$ のとき, 従来のシンプルランダムウォーク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. また, この分布関数は, 統計力学においてギブス分布として知られている.

池田らは, 最初にポテンシャル関数 $U(u,v)$ を

$$U(u,v) = \log(\max\{deg(u), deg(v)\}) \tag{3.10}$$

と定義したが, 次のポテンシャル関数 $U(u,v)$ の方がより (定数倍程度ではあるが) より良い結果を与えることを示している.

$$U(u,v) = U(v) = \log(deg(v)) \tag{3.11}$$

式 (3.9) に式 (3.11) を代入すると,

$$p_{uv}^{(\beta)} = \frac{deg(v)^{-\beta}}{\sum_{\omega \in N(u)} deg(\omega)^{-\beta}} \text{ for } v \in N(u) \tag{3.12}$$

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.

一般のグラフに対しては $\beta = 1/2$ のときカバータイム $\mathcal{C}^{1/2}$ が

$$\mathcal{C}^{1/2} = O(n^2 \log n) \tag{3.13}$$

であることを示している [15][16].

3.3.2 ポテンシャル関数の改良

本研究では, 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のさらなる効率化を目指して, ポテンシャル関数の改善を試みた. スケールフリーグラフでは, 頂点の次数分布はベキ則にしたがう. そのため, 次数分布は指数関数的に極端に偏っている. この事実から, 次数分布を線形に補正することでより良い振る舞いをするランダムウォークのポテンシャル関数を設計できると予想した. そこで, ポテンシャル関数 $U(u,v)$ を

$$U(u,v) = U(v) = \log\log(deg(v)) \tag{3.14}$$

と定義した. このとき,

$$p_{uv}^{(\beta)} = \frac{(\log deg(v))^{-\beta}}{\sum_{\omega \in N(u)} (\log deg(\omega))^{-\beta}} \text{ for } v \in N(u) \tag{3.15}$$

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.

3.3.3 有向スケールフリーグラフ上のランダムウォーク

本論文では池田らのランダムウォークモデルで, 逆方向に対して確率 $\gamma(0 < \gamma \leq 1)$ で辿ることを許したランダムウォークを提案する.

ここでは, 有向グラフ上の次数を (3.16) で定義し, 遷移確率は, 順辺に対しては (3.17), 逆辺に対しては (3.18) とした.

$$deg(v) = in\_deg(v) + \gamma \cdot out\_deg(v) \tag{3.16}$$

$$p_{OUTuv}^{(\beta,\gamma)} = \frac{deg(v)^{-\beta}}{\sum_{\omega \in OUT(u)} deg(\omega)^{-\beta} + \sum_{\omega \in IN(u)} \gamma \cdot deg(\omega)^{-\beta}} \tag{3.17}$$

$$p_{INuv}^{(\beta,\gamma)} = \frac{\gamma \cdot deg(v)^{-\beta}}{\sum_{\omega \in OUT(u)} deg(\omega)^{-\beta} + \sum_{\omega \in IN(u)} \gamma \cdot deg(\omega)^{-\beta}} \tag{3.18}$$

3.4 その他のランダムウォーク

3.4.1 ページランク

最も人気の高い検索エンジンの 1 つである google では, ページランクと呼ばれる Web ページの価値を評価するアルゴリズムにおいて, ランダムウォーク (マルコフチェーン) の理論を利用している.

u をある Web ページすると, ページ u の (シンプルな) ページランク R_u は次のように定義される.

$$R_u = c \sum_{v \in IN(u)} \frac{R(v)}{out\_deg(v)} \tag{3.19}$$

このページランクは理論的には, シンプルランダムウォークの定常分布と一致することが示されている [19].

第4章　計算機実験

4.1　実装

実装には, C++言語を使用した. C++クラスライブラリーである boost Libralies[2] を用いて, 乱数発生アルゴリズムには mt19937 を使用した. グラフライブラリーとしては, boost graph Libralies を利用した. 計算機は, 主に情報科学センターの SGI Altix3700 (Intel® Itanium® 2 プロセッサ ×128) を使用した.

4.2　無向スケールフリーグラフ上における実験結果

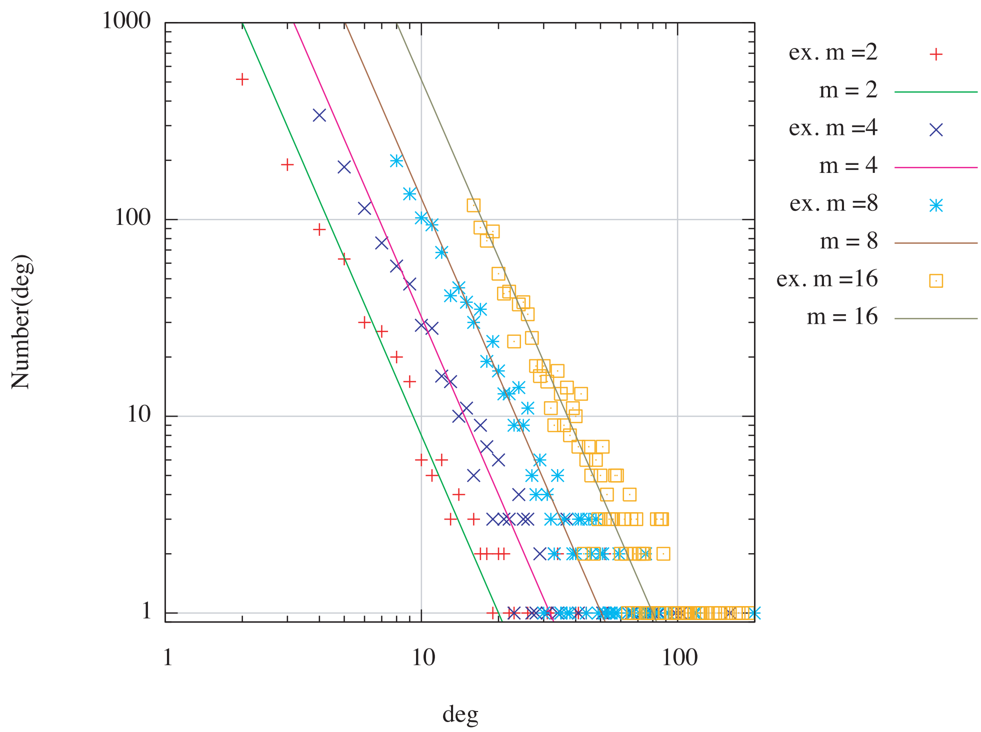

図 4.1: 無向スケールフリーグラフの次数分布 ($n = 1000$)

図 4.1 は, preferential attachment グラフモデルの次数分布をパラメータ m ごとに理論値及び実験値 (ex.) を表したものである.

4.2.1 次数べき乗型を用いた場合

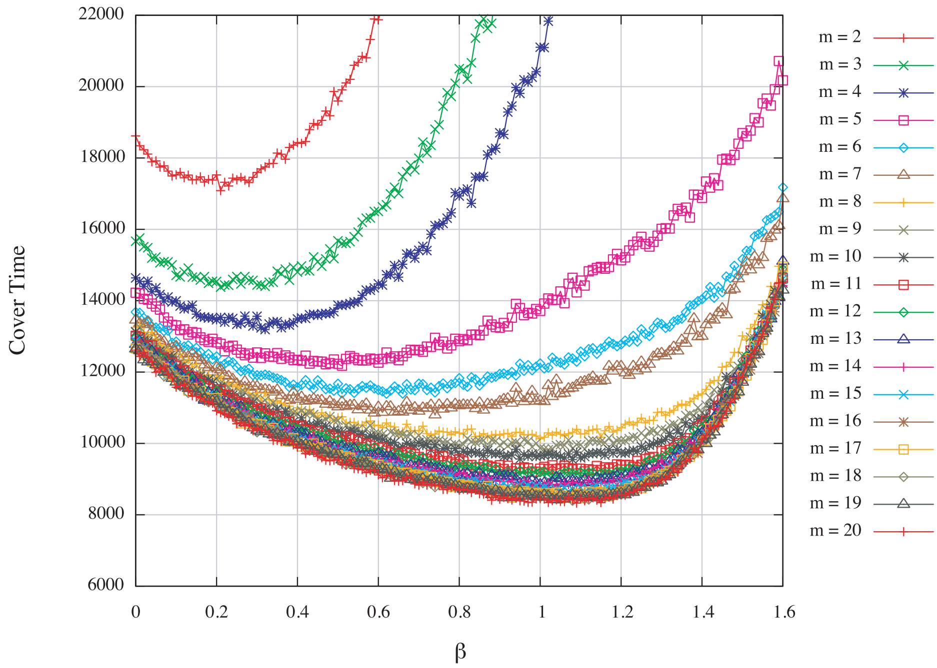

図 4.2: 次数べき乗型 β vs. $\bar{C}$ $(n = 1000)$

図 4.2,4.3 は, $n = 1000$ のときの各 m の局所情報を用いたランダムウォークのパラメータ β に対するカバータイムである. m によって, カバータイムの最低値を与える β が異ることがわかる. $\beta = 0$ のときの従来のシンプルランダムウォークよりも, 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.

図 4.4 は, 図 4.2 の最小値を与える β を最小自乗法により求め, 頂点数 n, グラフパラメータ 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.

図 4.7 は, $n = 1000$ のときの各 m に対する最適な β を表したものである. $m = 2$ から $m = 10$ にかけて, 最適な β の値は, 線形に単調増加していき, $m = 10$ 以降では, 最適な β の値は 1 程度にに収束していることがわかる. 図 4.5, 4.6 は, その最適な β に対するカバータイムをグラフパラメータ m の関数として表示したものである.

図 4.8 は, $n = 1000$ のときに, シンプルランダムウォークに対する局所情報を用いたランダムウォークの改善率を, m と β の関数として表したものである. また, 図 4.9 は, 各 m のときのシンプルランダムウォークに対する局所情報を用いたランダムウォークの改善率を表している. 2 つの図から, $m = 10$ にかけては, m が大きくなるに従って改善の割合は良くなり, $m = 10$ 以降では, 最適な β の値が収束すると同時に, 改善率も 0.35 程度に収束していくことがわかる.

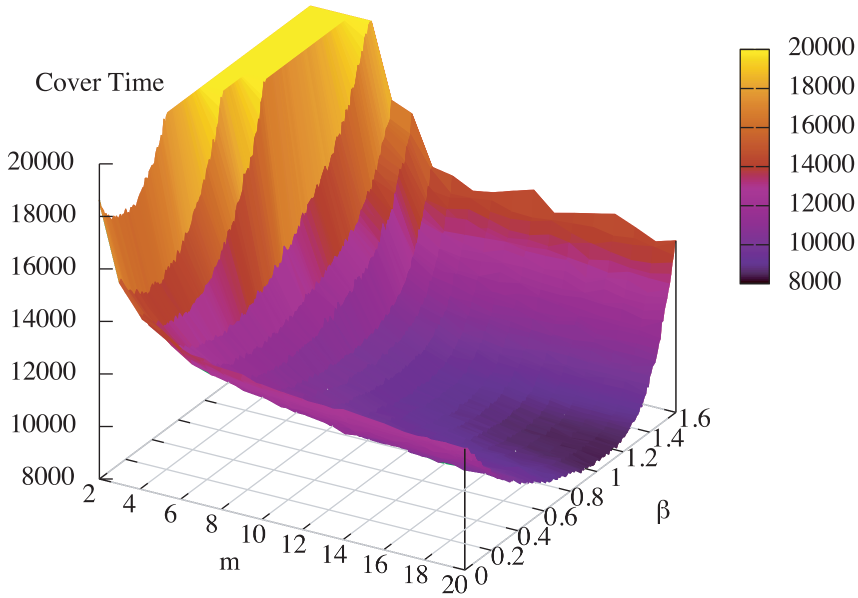

図 4.3: 次数べき乗型 m vs. β vs. $\bar{C}$ $(n = 1000)$

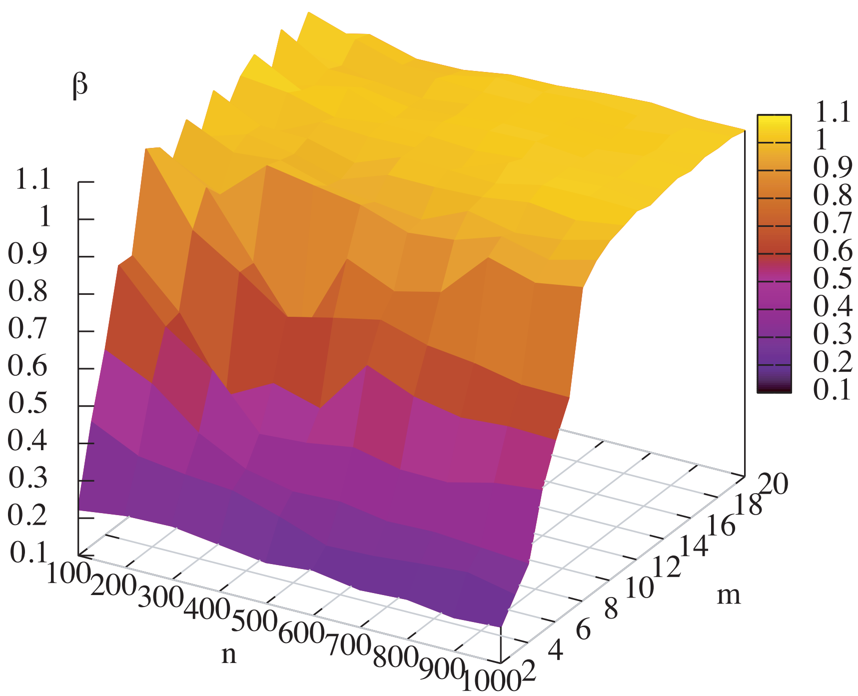

図 4.4: 次数べき乗型 n vs. m vs. 最適な β

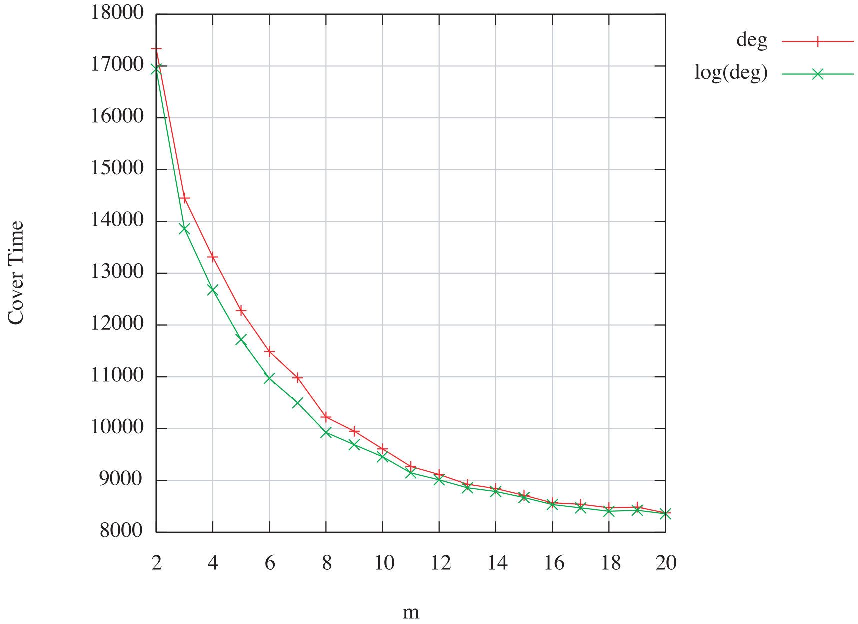

図 4.5: 次数べき乗型と次数の対数べき乗型 m vs. 最適な β の $\bar{\mathcal{C}}$ $(n = 1000)$

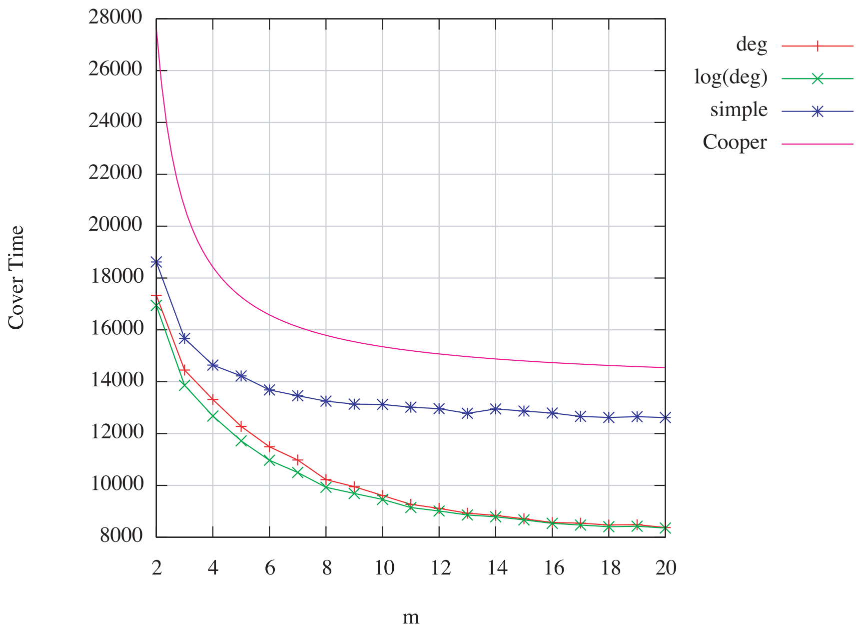

図 4.6: 次数べき乗型, 次数の対数べき乗型, シンプルランダムウォーク $(\beta = 0)$ 及び Cooper et al. のシンプルランダムウォークに対する理論的上界 m vs. 最適な β の $\bar{\mathcal{C}}$ $(n = 1000)$

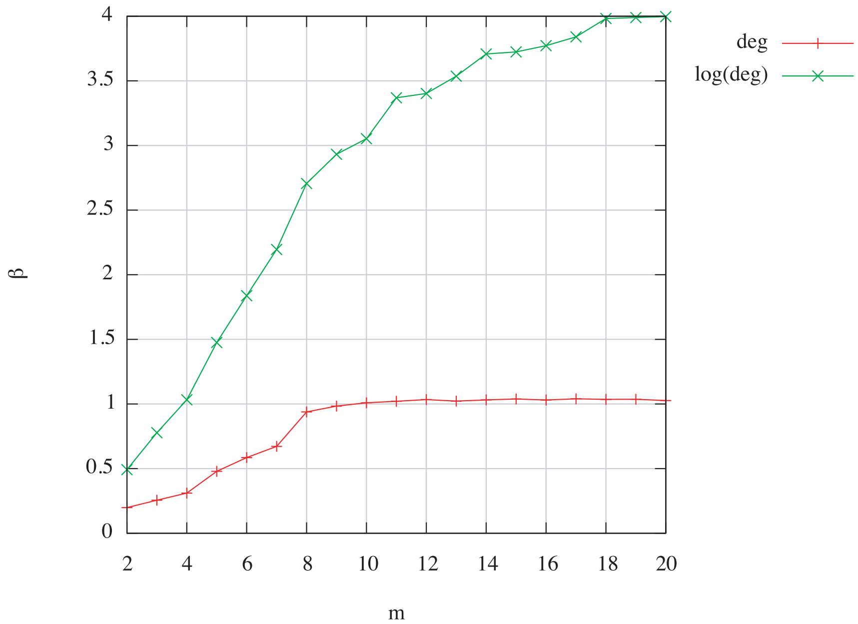

図 4.7: 次数べき乗型と次数の対数べき乗型 m vs. 最適な β $(n = 1000)$

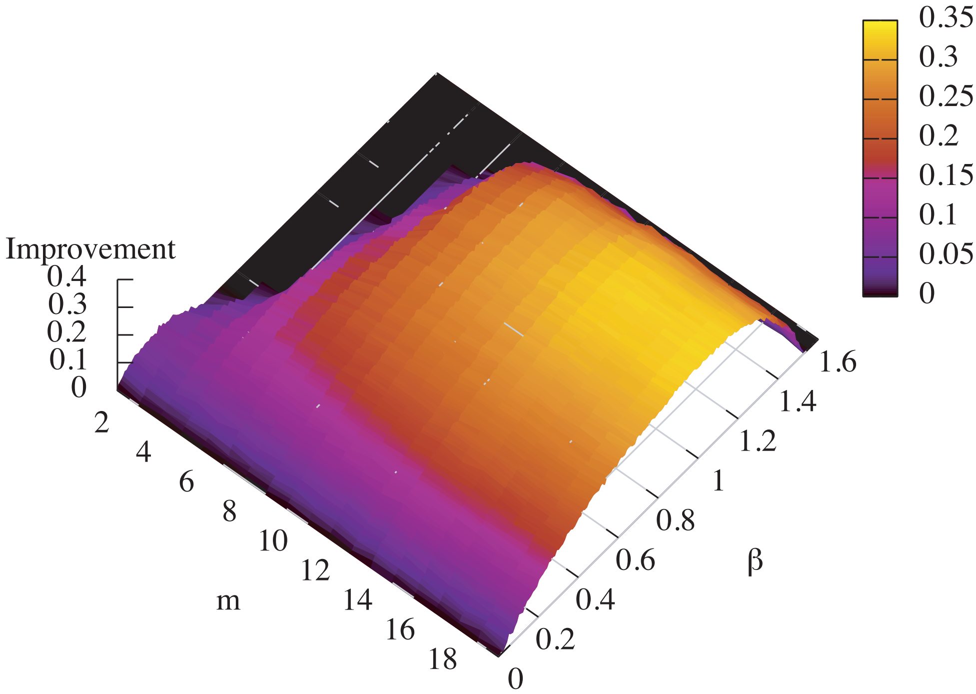

図 4.8: 次数べき乗型 m vs. β vs. シンプルランダムウォーク $(\beta = 0)$ に対する改善率 $(n = 1000)$

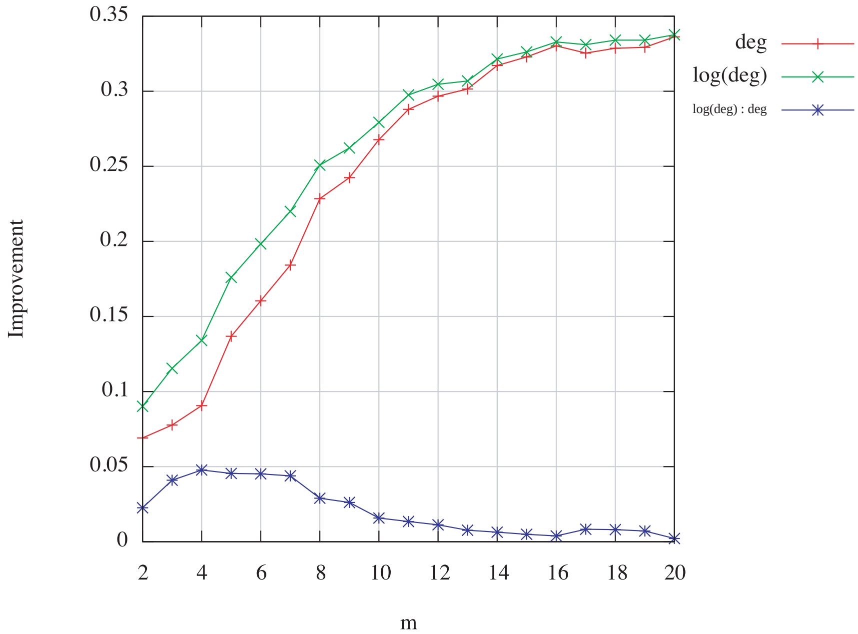

図 4.9: 次数べき乗型と次数の対数べき乗型 m vs. シンプルランダムウォーク ($\beta = 0$) に対する改善率 ($n = 1000$)

表 4.1: 次数べき乗型 各 m に対する最適な β

m	**最適なβ**
2	0.1983
4	0.3103
8	0.9388
20	1.026

図 4.10 は, 頂点数 $n = 1000$, グラフパラメータ $m = 2$ のときの, 各次数の頂点に対する訪問数の平均を表したものである. 表 4.1 より, $m = 2$ のときの最適な β は, $\beta = 0.1983$ である. これは, 最小自乗法より求められた値であるので, その値に近い $\beta = 0.20$ のときと, それ以外のいくつかの β の値で比較している. 最適値に近い $\beta = 0.20$ では, $\beta = 0$ のシンプルランダムウォークに対して, ほぼ全ての次数での訪問数が下回っていることが確認できる. 特に, 次数の高い頂点での訪問数が, 目立って改善されている. $\beta = 1.00, 1.60$ のときは, 均一化していてより良い結果を与えるように思われるが, 小さな次数での訪問数が, シンプルランダムウォークに比べて多くなってしまっていることが, 全体としてカバータイムを悪化させている. 図 4.1 からもわかるように, 次数が小さくなるにしたがって頂点数は指数関数的に増加するので, 小さな次数への訪問数の増加は, カバータイムに対して大きな影響を与える.

図 4.11 は, スケールフリーグラフのパラメータ $m = 4$ のときの, 各次数の頂点に対する訪問数の平均を表したものである. 表 4.1 より, $m = 4$ のときの最適な β は, $\beta = 0.3103$ であり, その値に近い $\beta = 0.31$ のときと, それ以外のいくつかの β の値で比較している. $m = 2$ ときと同じく, 最適値に近い $\beta = 0.31$ では, $\beta = 0$ のシンプルランダムウォークに対して, ほぼ全ての次数での訪問数が下回り, $\beta = 1.00, 1.60$ のときは, 小さな次数での訪問数が, シンプルランダムウォークに比べて増加している.

図 4.12,4.13 は, $m = 8, 20$ のときの, 各次数の頂点に対する訪問数の平均を表したものである. 表 4.1 より, $m = 8$ のときの β の最適値は $\beta = 0.9388$, $m = 20$ のときの最適値は $\beta = 1.026$ である. それぞれの最適値に近い $\beta = 0.94, 1.03$ のときの訪問数は共に, 小さな次数で, わずかにシンプルランダムウォーク ($\beta = 0$) の訪問数を越えているが, 全体的には均一な分布で, シンプルランダムウォークの訪問数を下回っている.

$m = 2, 4$ の小さな m のときに比べて, $m = 8, 20$ の大きな m のときは, カバータイム及びその改善率が良い結果を与えている. 小さな m のとき, スケールフリーグラフは疎なグラフとなり, 直径が大きいためにカバータイムは悪化する. また, このとき, 次数の高い頂点はハブとなっていて, 次数の低い頂点の多くは, そのハブのどれかと接しているため, 全ての頂点を訪れるために, 次数の高い頂点への訪問数を大きく下げることは難しい. そのため, 最適な β に対しても図 4.10, 4.11 のような分布になる.

大きな m のとき, スケールフリーグラフは密なグラフで直径は比較的小さく, カバータイムは小さな m と比べて改善する. 同時に, 小さな次数の頂点は, いくつかのハブと接していたり, 小さい次数同士が接している確率がより高くなるので, 必ずしも高い次数の頂点を多く訪問しなくても, グラフ全体をカバーできる. その結果, 図 4.12, 4.13 のように, 最適な β のとき, シンプルランダムウォークの偏った分布から, 訪問数を均一化することが可能であると考えることができる.

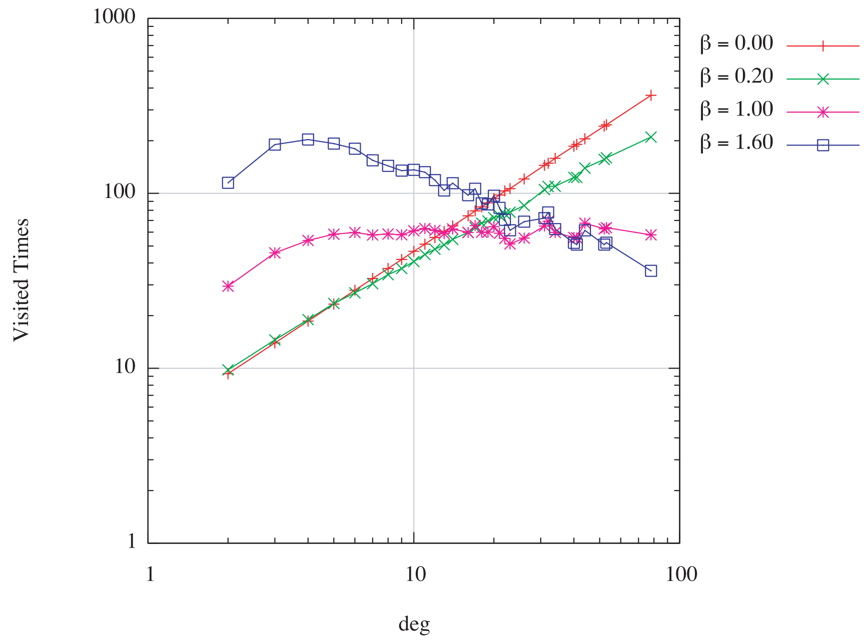

図 4.10: 次数べき乗型 各次数の頂点に対する訪問数 ($n = 1000,\ m = 2$)

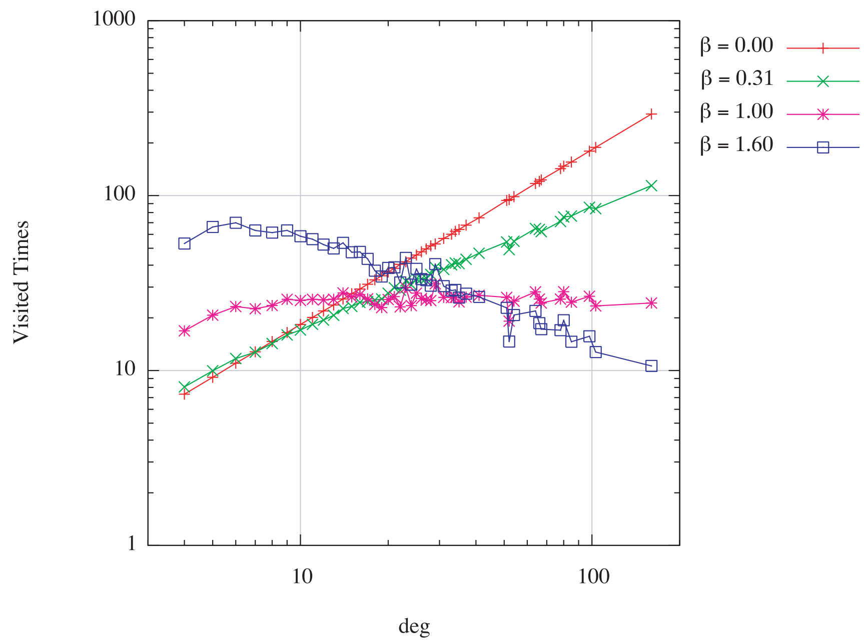

図 4.11: 次数べき乗型 各次数の頂点に対する訪問数 ($n = 1000,\ m = 4$)

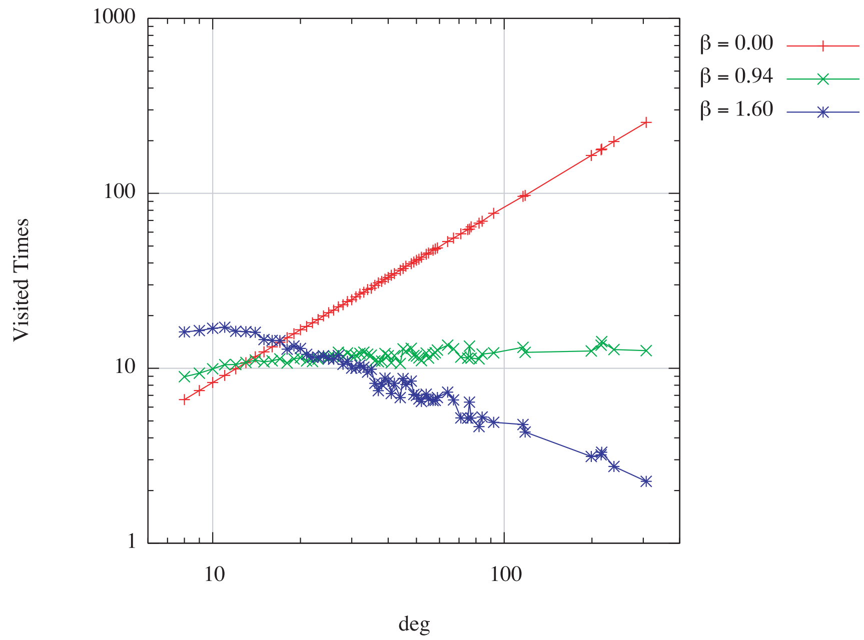

図 4.12: 次数べき乗型 各次数の頂点に対する訪問数 ($n = 1000$, $m = 8$)

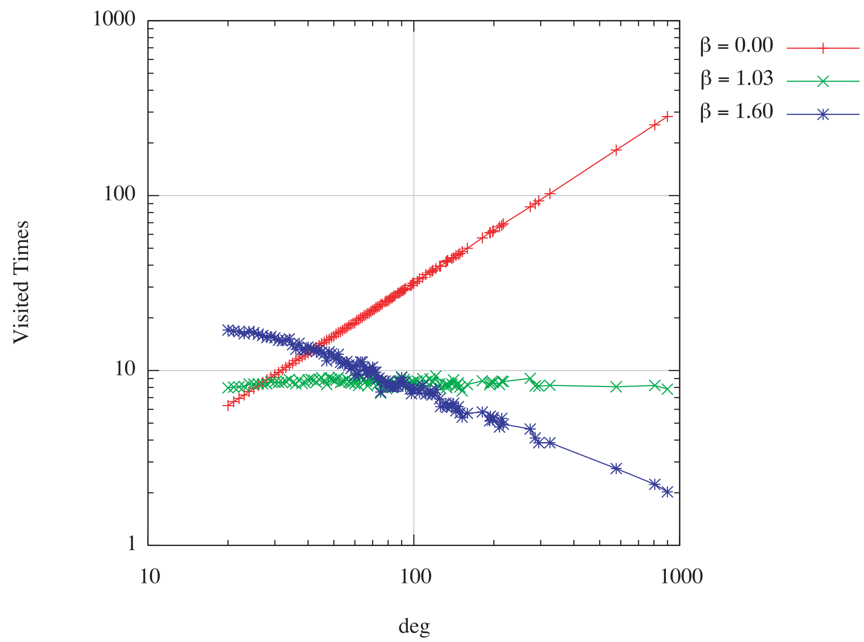

図 4.13: 次数べき乗型 各次数の頂点に対する訪問数 ($n = 1000$, $m = 20$)

4.2.2 次数の対数べき乗型を用いた場合

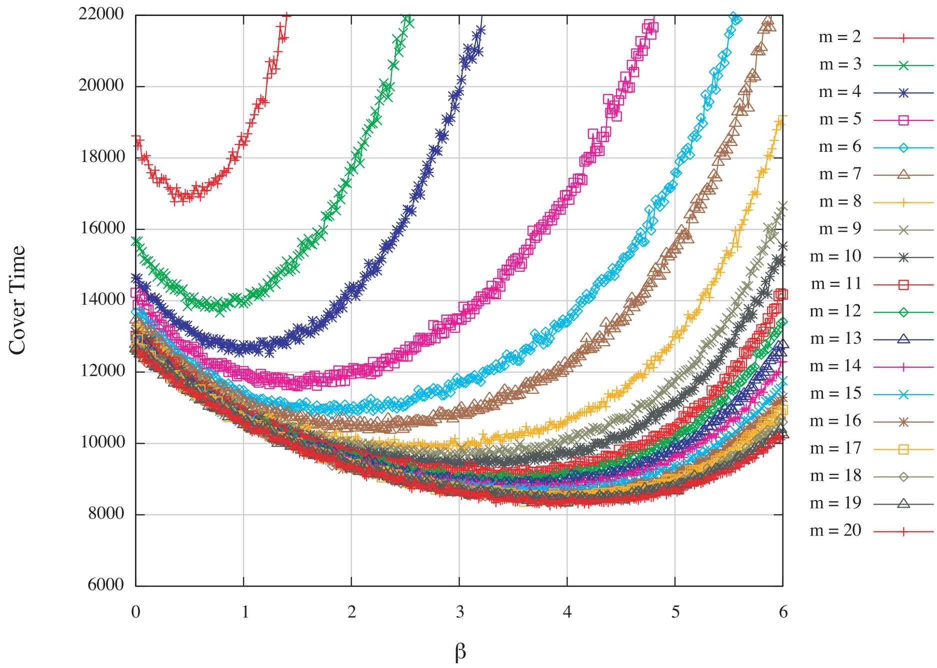

図 4.14: 次数の対数べき乗型 β vs.$\bar{\mathcal{C}}$ $(n = 1000)$

図 4.14,4.15 は, 次数の対数べき乗型の局所情報を用いたランダムウォークにおける, $n = 1000$, 各 m の β に対するカバータイムを表したものである. 図 4.2,4.3 と比較して, 最適な β の値が大きくなっていることがわかる. これは, 次数のべき乗が, 次数の対数のべき乗に置き換わることによって, 重みが弱められているので, β を大きくすることでキャンセルしていると考えられる.

図 4.16 は, 図 4.4 と同じく, 最小値を与える β を最小自乗法により求め, 頂点数 n, グラフパラメータ m の関数として表示したものである. 図 4.7, 4.16 より, 次数の対数べき乗型の最適な β は, 次数べき乗型に比べて大きな値をとる. また, m が増加していくときの, β が収束する速さも, 次数べき乗型に比べて遅く, $m = 20$ のときでも, 4 程度である.

図 4.17, 4.18, 4.19, 4.20 は, それぞれ $m = 2, 4, 8, 20$, $n = 1000$ のとき, 次数べき乗型と次数の対数べき乗型の β に対するカバータイムを比較したものである.

図 4.21 は, $n = 1000$ のときに, シンプルランダムウォークに対する局所情報を用いたランダムウォークの改善率を, m と β の関数として表したものである. 図 4.9, 4.21 から, $m = 2$ から $m = 8$ にかけて, 次数べき乗型に対する改善の割合は良くなっていることがわかる. 最も改善の効果が大きい, $m = 4$ から $m = 7$ では 0.05 程度の改善率を示している. m が増加するにしたがって, 次数べき乗型に対する改善率は減少し, 次数べき乗型と次数の対数べき乗型のカバータイムは, ほぼ一致するものになる.

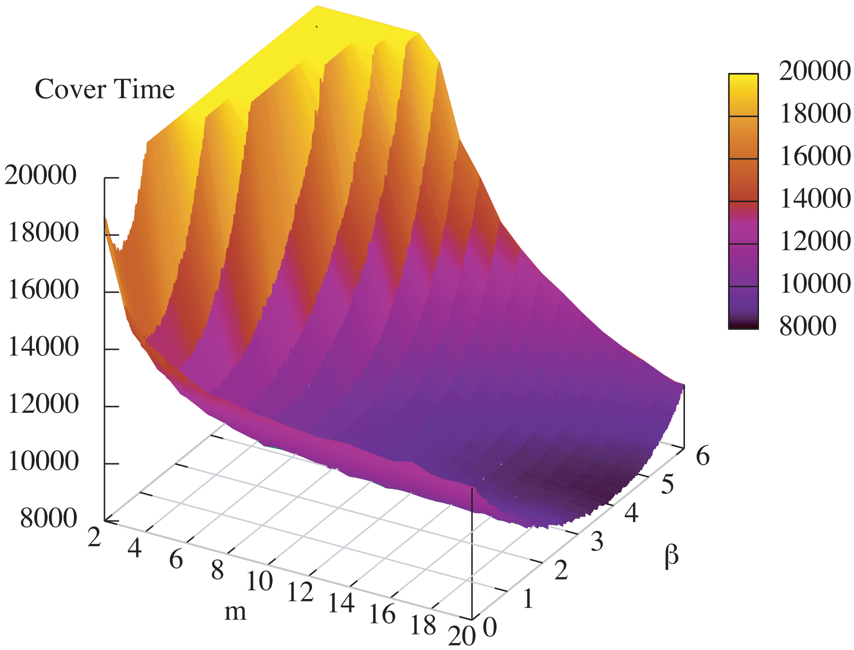

図 4.15: 次数の対数べき乗型 m vs. 最適な β vs. $\bar{C}$ $(n = 1000)$

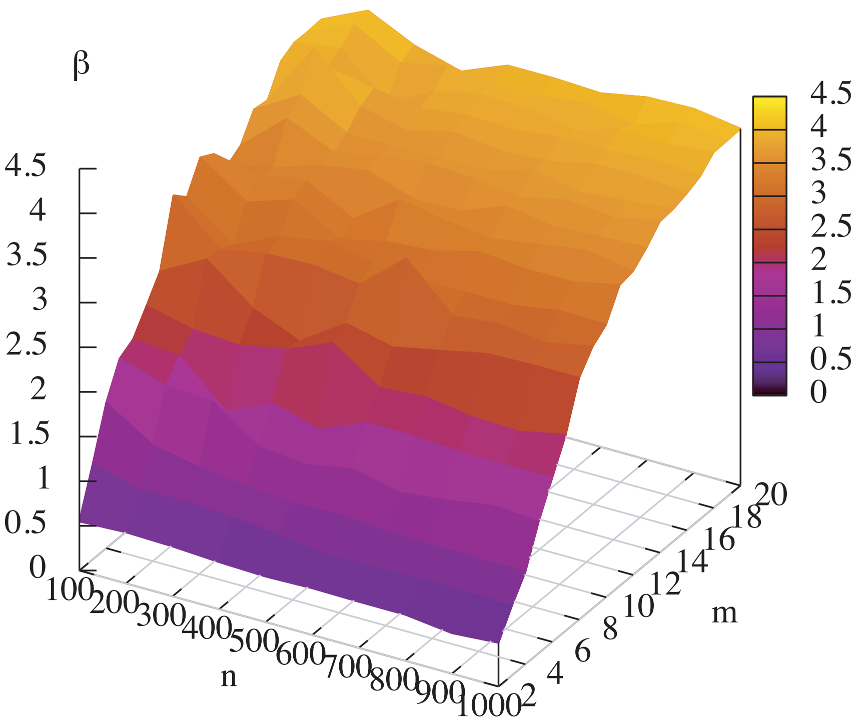

図 4.16: 次数の対数べき乗型 n vs. m vs. 最適な β

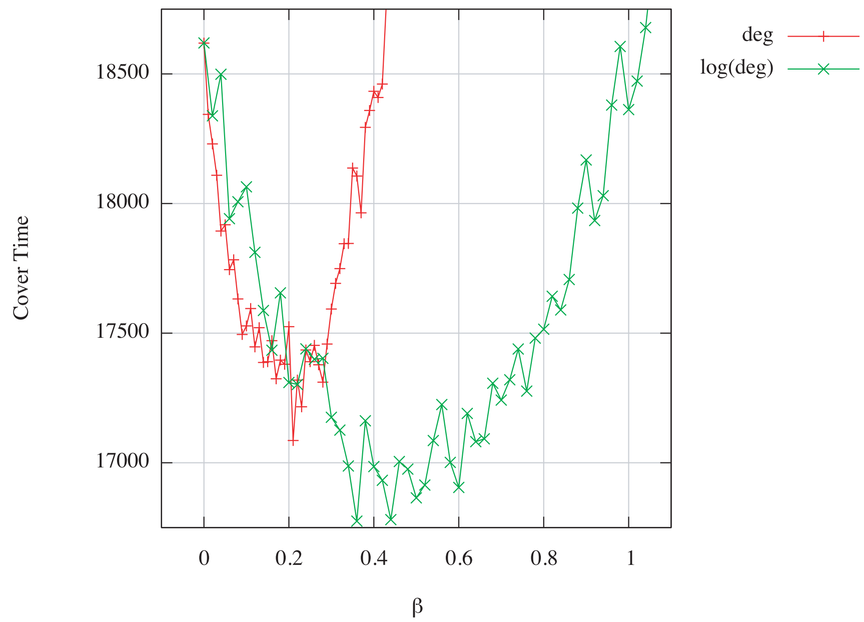

図 4.17: 次数べき乗型と次数の対数べき乗型 $(n = 1000, m = 2)$

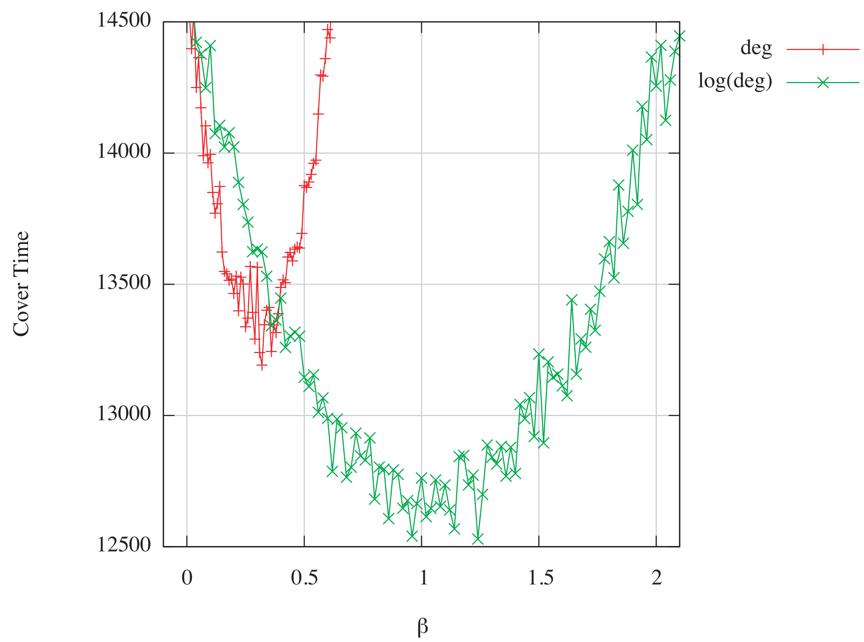

図 4.18: 次数べき乗型と次数の対数べき乗型 $(n = 1000, m = 4)$

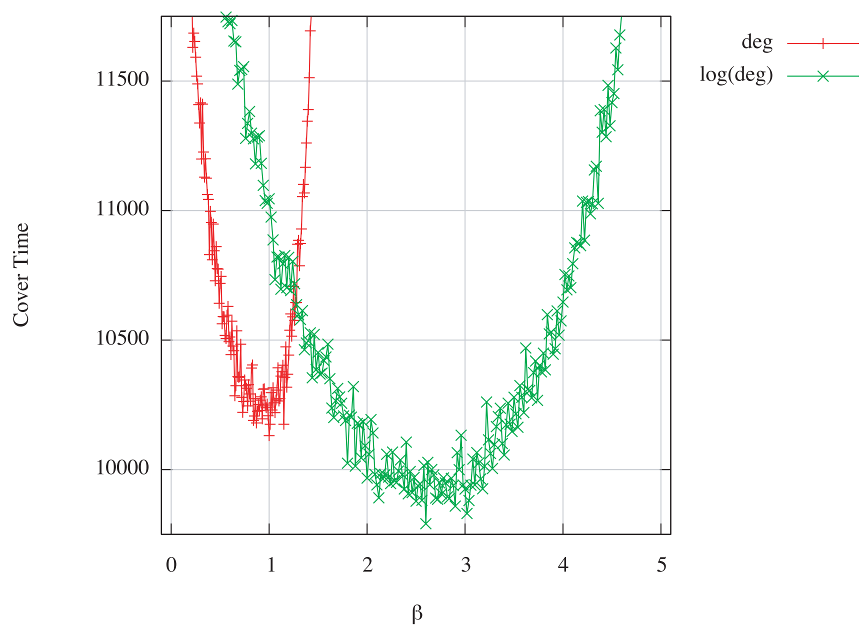

図 4.19: 次数べき乗型と次数の対数べき乗型 $(n = 1000, m = 8)$

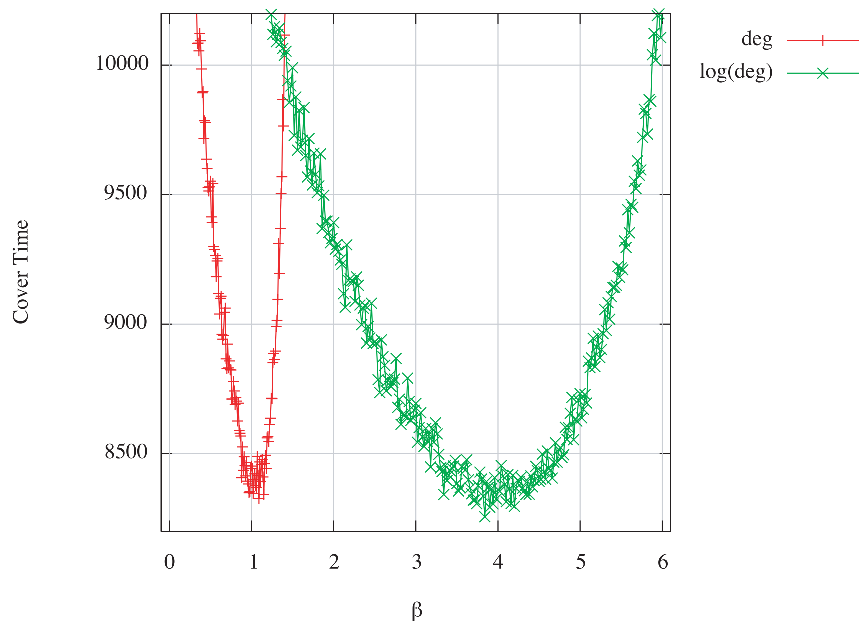

図 4.20: 次数べき乗型と次数の対数べき乗型 $(n = 1000, m = 20)$

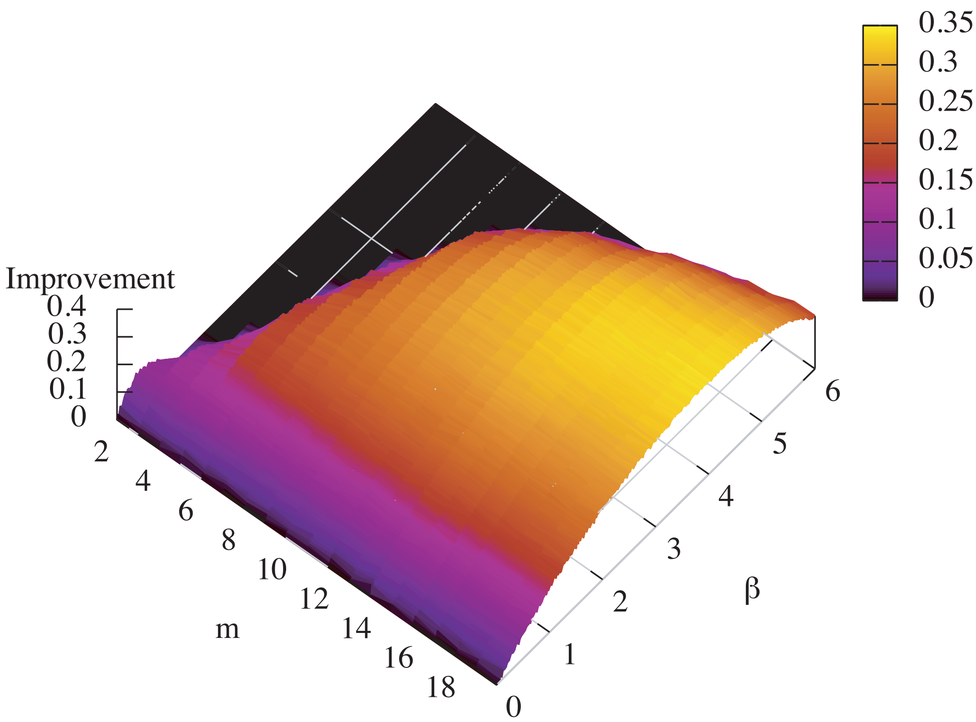

図 4.21: 次数の対数べき乗型 m vs. β vs. シンプルランダムウォーク ($\beta = 0$) に対する改善率 ($n = 1000$)

表 4.2: 次数の対数型の各 m に対する最適な β

m	最適なβ
2	0.4918
4	1.032
8	2.706
20	3.997

図 4.22 は, 頂点数 $n = 1000$, グラフパラメータ $m = 2$ のときの, 各次数の頂点に対する訪問数の平均を表したものである. 表 4.2 より, $m = 2$ のときの最適な β は, $\beta = 0.4913$ である. これは, 最小自乗法より求められた値であるので, その値に近い $\beta = 0.50$ のときと, それ以外のいくつかの β の値で比較している. また, 図 4.23 は, このときの, 次数べき乗型と次数の対数べき乗型の訪問数を比較したものである. 最適値に近い $\beta = 0.50$ では, 次数べき乗型に対して, 特に小さな次数を除くほぼ全ての次数での訪問数がわずかであるが下回っていることが確認できる.

図 4.24 は, スケールフリーグラフのパラメータ $m = 4$ のときの, 各次数の頂点に対する訪問数の平均を表したものである. 表 4.2 より, $m = 4$ のときの最適な β は, $\beta = 1.032$ であり, その値に近い $\beta = 1.04$ のときと, それ以外のいくつかの β の値で比較している. また, 図 4.25 は, このときの, 次数べき乗型と次数の対数べき乗型の訪問数を比較したものである. $m = 2$ ときと同じく, 最適値に近い $\beta = 1.04$ では, $\beta = 0$ のシンプルランダムウォークに対して, 特に小さな次数を除くほぼ全ての次数での訪問数が下回り, $m = 2$ のときよりもさらに改善の割合は増加している.

図 4.26,4.28 は, $m = 8, 20$ のときの, 各次数の頂点に対する訪問数の平均を表したものである. 表 4.2 より, $m = 8$ のときの β の最適値は $\beta = 2.706$, $m = 20$ のときの最適値は $\beta = 3.997$ である. それぞれの最適値に近い $\beta = 2.70, 4.00$ のときの訪問数は共に, 小さな次数で, わずかにシンプルランダムウォーク ($\beta = 0$) の訪問数を越えているが, 全体的には均一な分布で, シンプルランダムウォークの訪問数を下回っている. また, 図 4.27,4.29 は, このときの, 次数べき乗型と次数の対数べき乗型の訪問数を比較したものである. $m = 8, 20$ のとき, 高い次数で次数の対数べき乗型の訪問数が, 次数べき乗型より大きくなっており, 低い次数では, わずかに次数べき乗型の訪問数が上回っている. 図 4.9 から, $m = 2$ のときと, $m = 8$ のときの改善率は同程度であることがわかるが, $m = 8$ のときの訪問数の分布図 4.27 は, $m = 2$ のときの訪問数の分布図 4.23 とは大きく異なり, $m = 20$ のときの訪問数の分布 4.29 に近い分布となっていることがわかる. 図 4.5, 4.9 から, m の値が大きくなっていくにしたがい, 次数べき乗型と次数の対数べき乗型のカバータイムは, 同じ値に近づいていくことがわかるが, 訪問数の分布から, そのランダムウォークの振る舞いは, 異なっていると考えることができる.

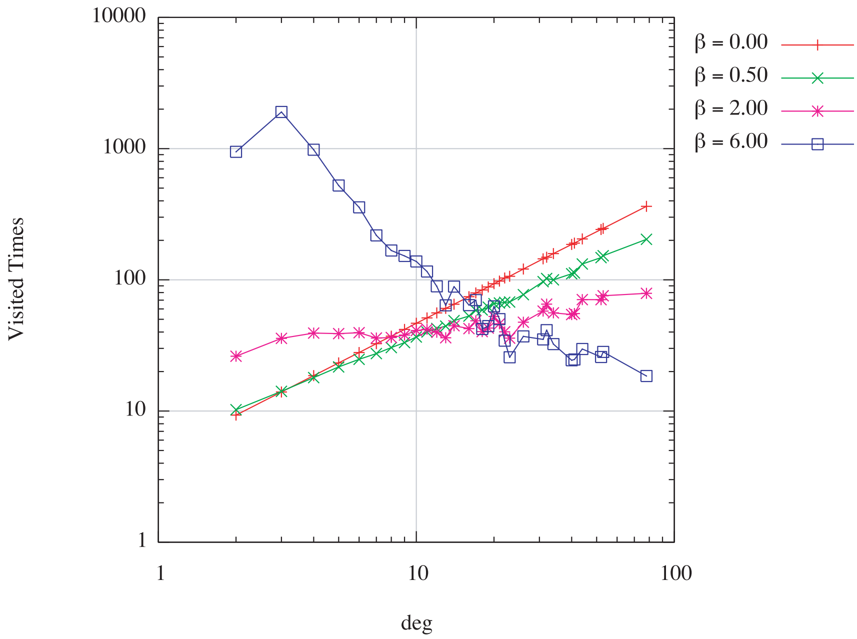

図 4.22: 次数の対数べき乗型 各次数の頂点に対する訪問数 ($n = 1000$, $m = 2$)

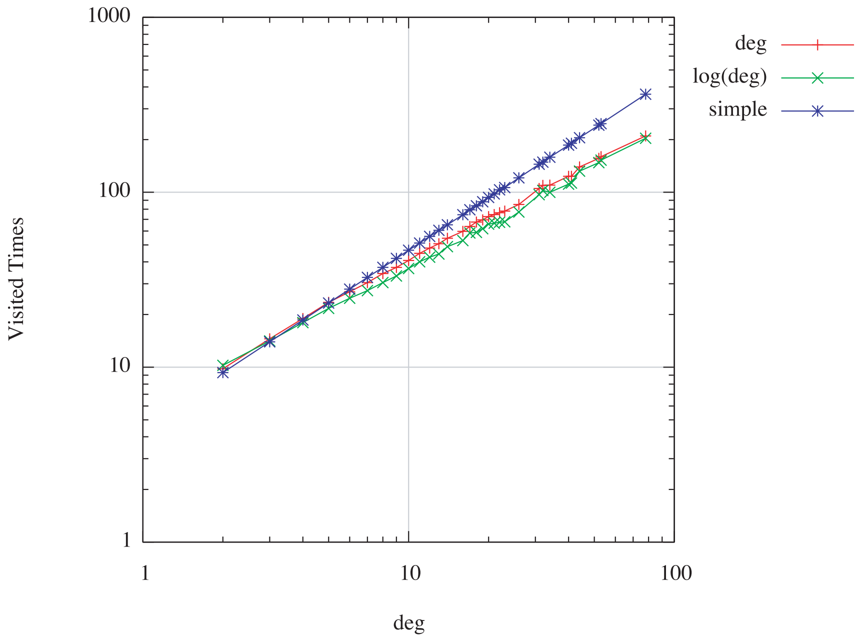

図 4.23: 次数べき乗型と次数の対数べき乗型における最適な β での各次数の頂点に対する訪問数の比較 ($n = 1000$, $m = 2$)

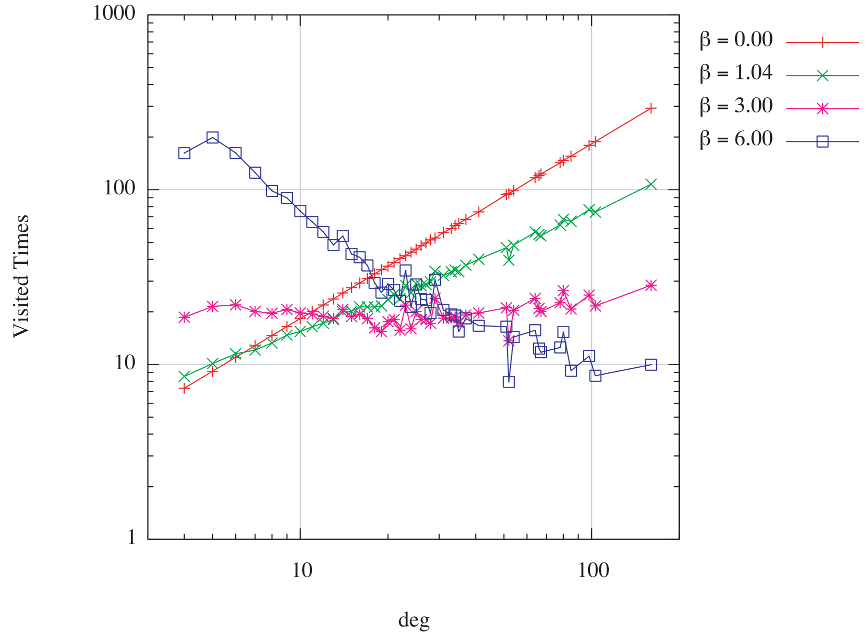

図 4.24: 次数の対数べき乗型 各次数の頂点に対する訪問数 ($n = 1000$, $m = 4$)

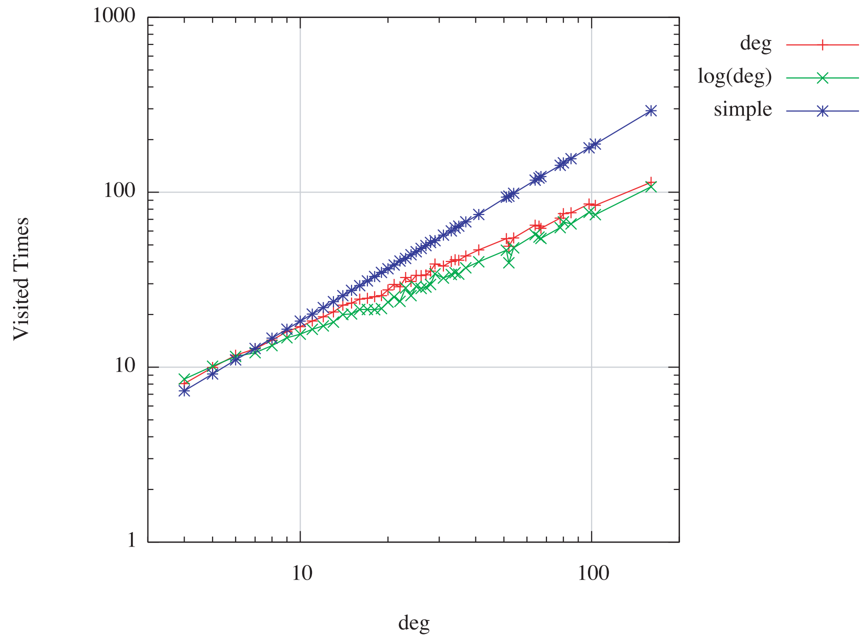

図 4.25: 次数べき乗型と次数の対数べき乗型における最適な β での各次数の頂点に対する訪問数の比較 ($n = 1000$, $m = 4$)

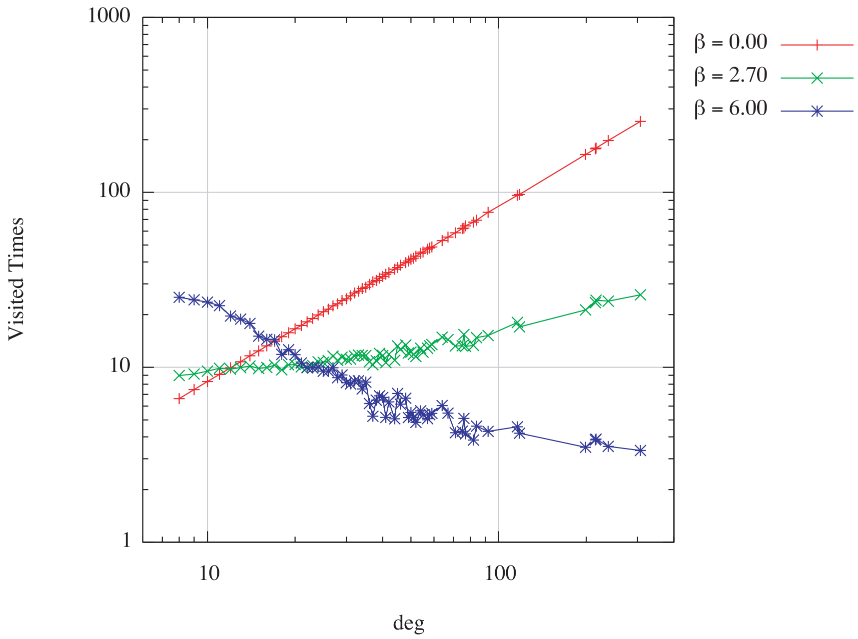

図 4.26: 次数の対数べき乗型 各次数の頂点に対する訪問数 ($n = 1000$, $m = 8$)

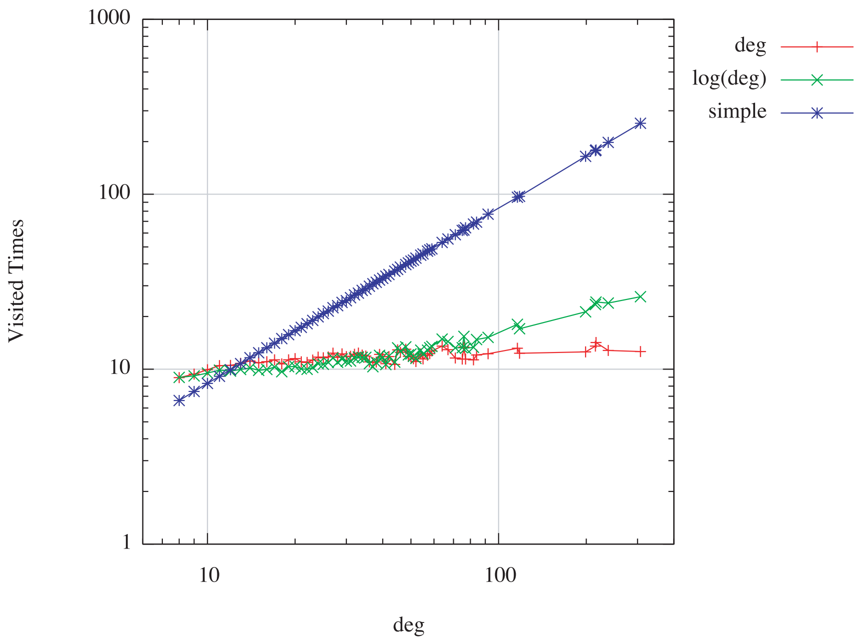

図 4.27: 次数べき乗型と次数の対数べき乗型における最適な β での各次数の頂点に対する訪問数の比較 ($n = 1000$, $m = 8$)

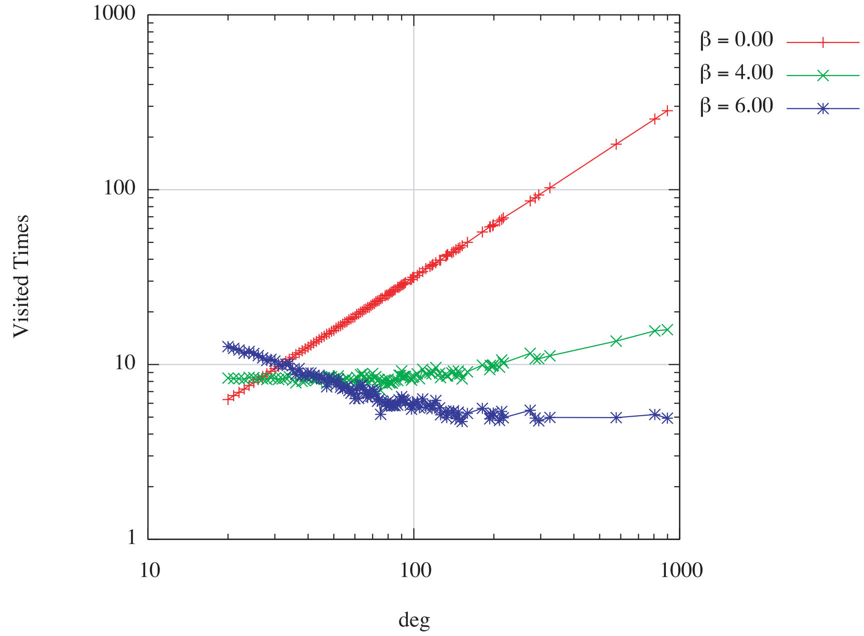

図 4.28: 次数の対数べき乗型 各次数の頂点に対する訪問数 ($n = 1000$, $m = 20$)

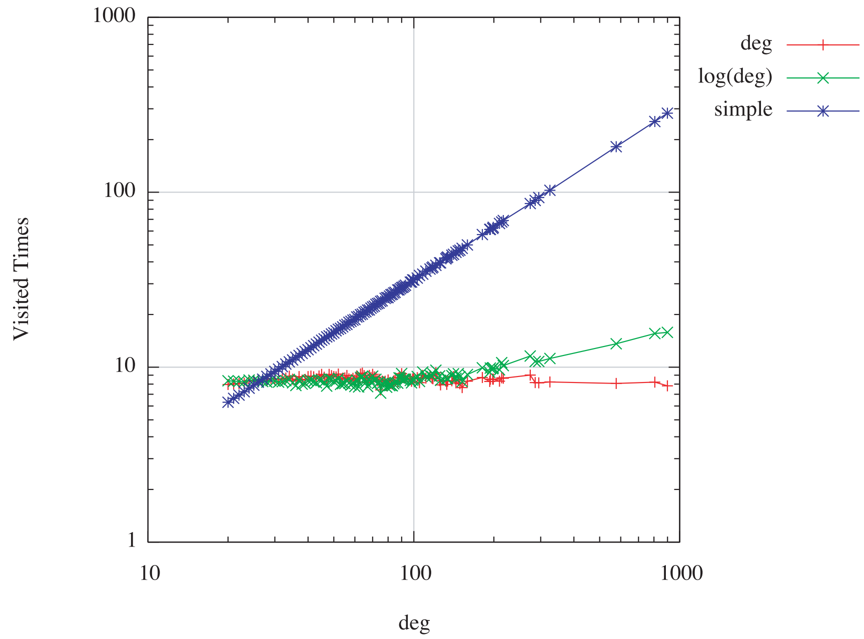

図 4.29: 次数べき乗型と次数の対数べき乗型における最適な β での各次数の頂点に対する訪問数の比較 ($n = 1000$, $m = 20$)

4.3　有向スケールフリーグラフ上における実験結果

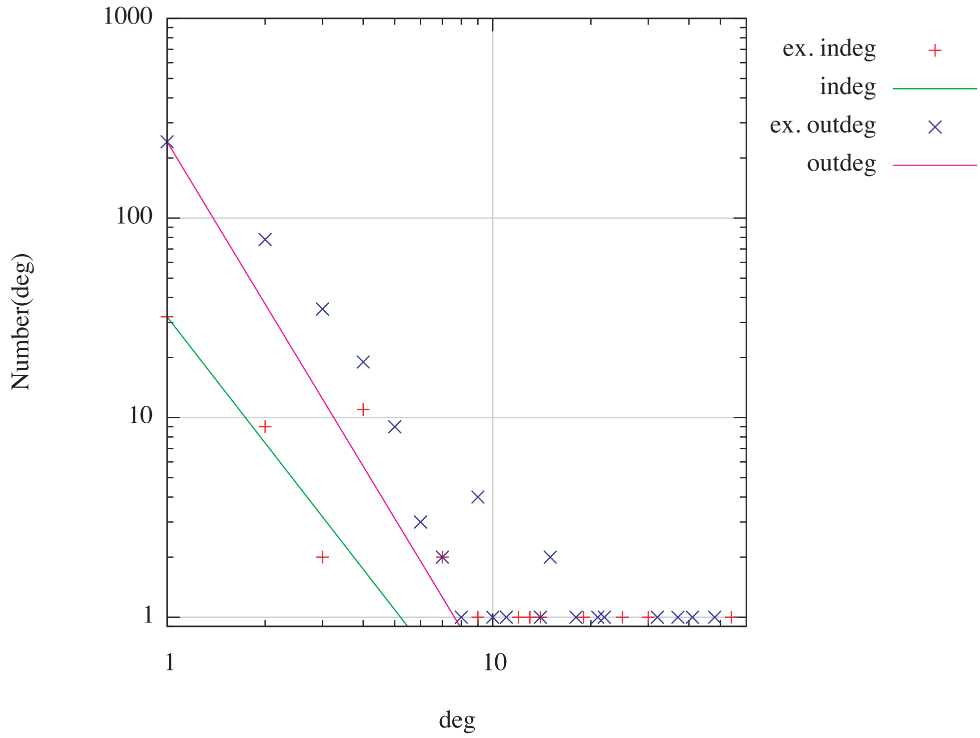

図 4.30: 有効スケールフリーグラフの次数分布 ($t = 1000$,$n \sim 410$, $\alpha = 0.41$, $\beta = 0.59$, $\gamma = 0$, $\delta_{in} = 0.24$, $\delta_{out} = 0$)

図 4.30 は, $t = 1000, \alpha = 0.41, \beta = 0.59, \gamma = 0, \delta_{in} = 0.24, \delta_{out} = 0$ のときの Bollobás らの有向スケールフリーグラフ [6] の入次数分布, 出次数分布の理論値及び実験値 (ex.) を表したものである.

4.3.1　逆辺をたどることを許した場合

Bollobás らの有向 スケールフリーグラフにおいて, 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 $t = 1000, \alpha = 0.41, \beta = 0.59, \gamma = 0, \delta_{in} = 0.24, \delta_{out} = 0$ 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.

このとき, $t = 1000$ で生成されるグラフの頂点数 n は, $n \sim 410$ と小さいことに注意する必要がある. $t = 1000$ 程度としたのは, $\gamma = 0.01$ のときにおいてでさえ, カバータイムが最大で 1000^2 程度となるからである.

図 4.31 は, $t = 1000$ のときの各 γ の逆辺をたどることを許した局所情報を用いたランダムウォークのパラメータ β に対するカバータイムである. 特に, $\gamma = 0.01$ のときの局所情

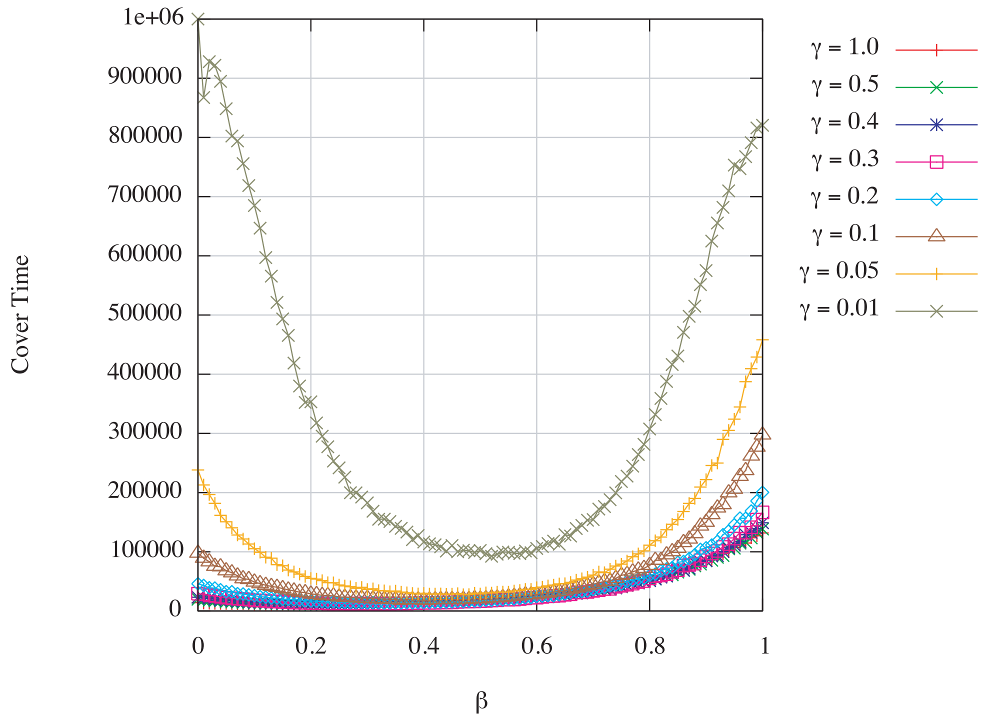

図 4.31: 逆辺をたどることを許した型 β vs. $\bar{C}$ ($t = 1000$,$n \sim 410$, $\alpha = 0.41$, $\beta = 0.59$, $\gamma = 0$, $\delta_{in} = 0.24$, $\delta_{out} = 0$)

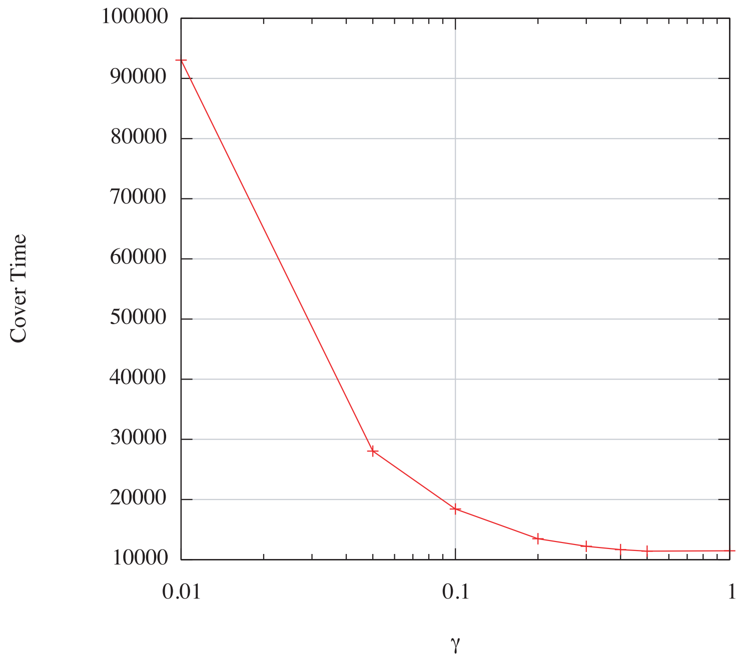

図 4.32: 逆辺をたどることを許した型 γ vs. $\bar{C}$ ($t = 1000$,$n \sim 410$)

報を用いないランダムウォーク ($\beta = 0$) のカバータイムが 1000^2 程度と悪化していることがわかる．図 4.32 は, その最適な β に対するカバータイムをグラフパラメータ γ の関数として表示したものである．

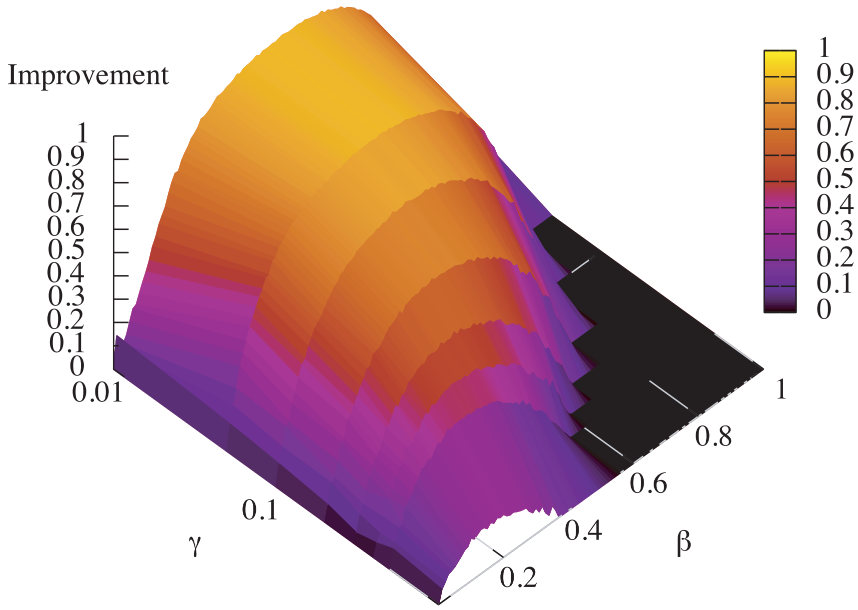

図 4.33: 逆辺をたどることを許した型 γ vs. β vs. $\beta = 0$ に対する改善率 ($t = 1000$,$n \sim 410$, $\alpha = 0.41$, $\beta = 0.59$, $\gamma = 0$, $\delta_{in} = 0.24$, $\delta_{out} = 0$)

図 4.33 は, $t = 1000$ のときに, $\beta = 0$ のときのカバータイムに対するの改善率を, γ と β の関数として表したものである．また, 図 4.34 は, 各 γ のときの, 最適な β のときの改善率を表している．図 4.31,4.32,4.33,4.34 より, γ の値が小さくなるにしたがって, $\beta = 0$ のときのカバータイムは, 指数関数的に悪化していくが, 最適な β のカバータイムはある程度低い値に抑えられており, その結果, 改善率が向上していると考えられる．

図 4.35 は, $t = 1000$ のときの各 γ に対する最適な β を表したものである．最適な β の値は, 指数関数的に増加していることがわかる．

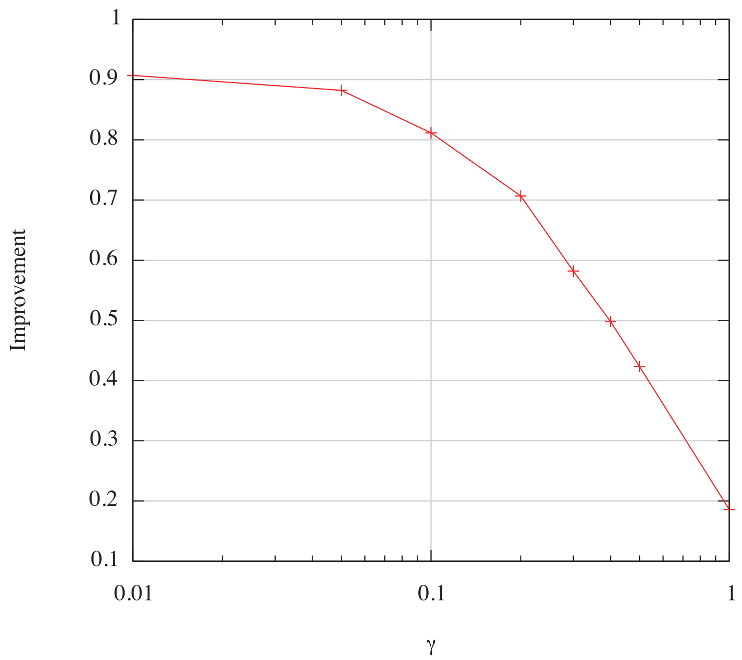

図 4.34: 逆辺をたどることを許した型 γ vs. $\beta = 0$ に対する改善率 ($t = 1000$,$n \sim 410$)

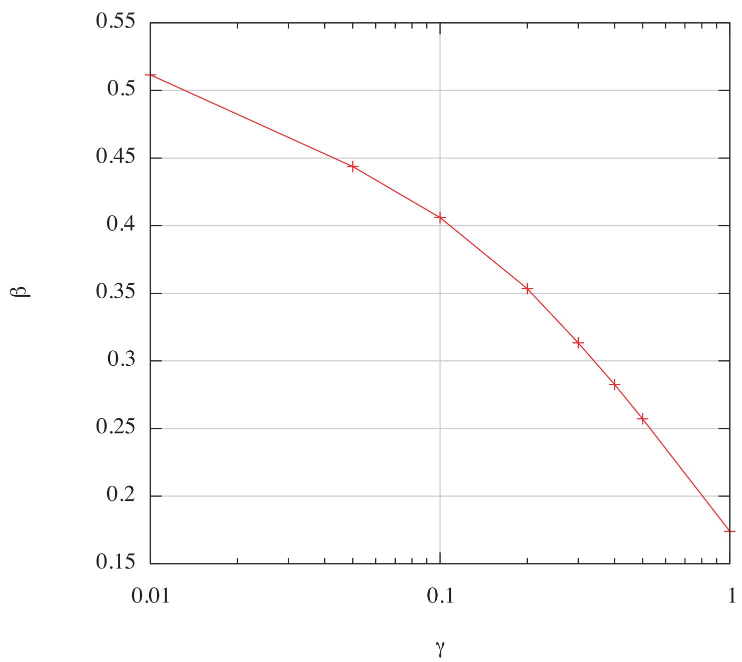

図 4.35: 逆辺をたどることを許した型 γ vs. 最適な β ($t = 1000$,$n \sim 410$)

表 4.3: 逆辺をたどることを許した型 各 γ に対する最適な β

γ	**最適なβ**
0.01	0.5116
0.1	0.4060
0.5	0.2571
1.0	0.1740

図 4.36,4.37,4.38 は, 頂点数 $t = 1000$, $\gamma = 0.01$ のときの, 各次数, 入次数, 出次数の頂点に対する訪問数の平均を表したものである. 表 4.3 より, $\gamma = 0.01$ のときの最適な β は, $\beta = 0.5116$ である. その値に近い $\beta = 0.51$ のときと, それ以外のいくつかの β の値で比較している. 全ての次数, 入次数, 出次数での訪問数が $\beta = 0$ のときを大きく下回っていることが確認できる.

図 4.39,4.40,4.41 は, 頂点数 $t = 1000$, $\gamma = 0.10$ のときの, 各次数, 入次数, 出次数の頂点に対する訪問数の平均を表したものである. 表 4.3 より, $\gamma = 0.10$ のときの最適な β は, $\beta = 0.4060$ である. その値に近い $\beta = 0.41$ のときと, それ以外のいくつかの β の値で比較している. $\gamma = 0.01$ のときと同じく, 全ての次数, 入次数, 出次数での訪問数が $\beta = 0$ のときを大きく下回っていることが確認できる.

図 4.42,4.43,4.44 は, $\gamma = 0.50$ のときの, 各次数, 入次数, 出次数の頂点に対する訪問数の平均を表したものである. 表 4.3 より, $\gamma = 0.50$ のときの最適な β は, $\beta = 0.2571$ である. その値に近い $\beta = 0.26$ のときと, それ以外のいくつかの β の値で比較している. 全ての次数, 入次数, 出次数での訪問数が $\beta = 0$ のときを下回っているが, 特に, 大きな次数, 入次数, 出次数での訪問数が $\beta = 0$ のときを大きく下回っていることがわかる.

図 4.45,4.46,4.47 は, $\gamma = 1.00$ のときの, 各次数, 入次数, 出次数の頂点に対する訪問数の平均を表したものである. 表 4.3 より, $\gamma = 0.50$ のときの最適な β は, $\beta = 0.1740$ である. その値に近い $\beta = 0.17$ のときと, それ以外のいくつかの β の値で比較している. 小さい次数, 入次数, 出次数での訪問数は $\beta = 0$ のときと同程度であるが, 特に大きな次数, 入次数, 出次数での訪問数が $\beta = 0$ のときを下回っていることがわかる.

有効スケールフリーグラフ上で, 逆辺をたどることを許した型では, γ が減少するにしたがって, カバータイムが悪化し, 改善率は向上した. これは, m が増加するにしたがって, カバータイム, 改善率共に向上した, 無向スケールフリーグラフ上の場合と逆であることに注意したい.

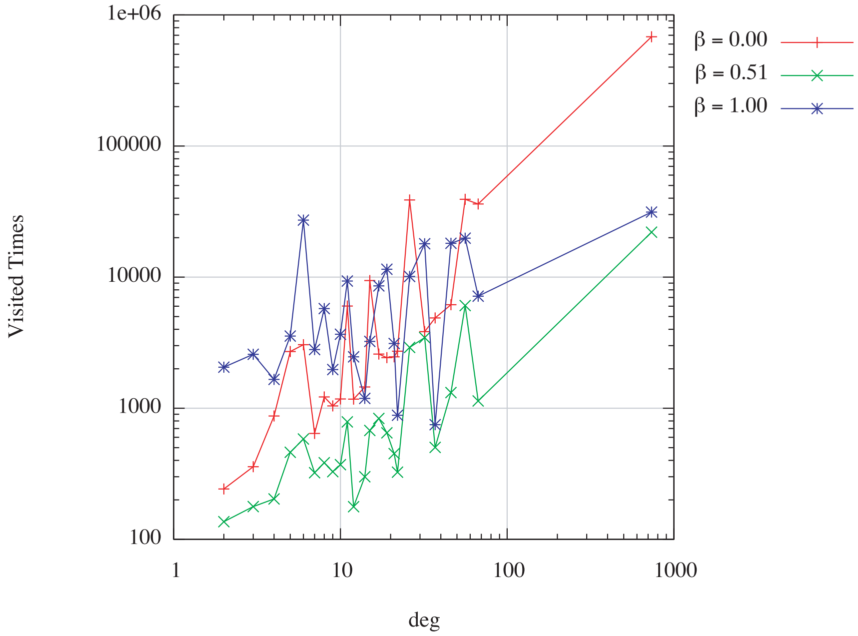

図 4.36: 逆辺をたどることを許した型 各次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.01$)

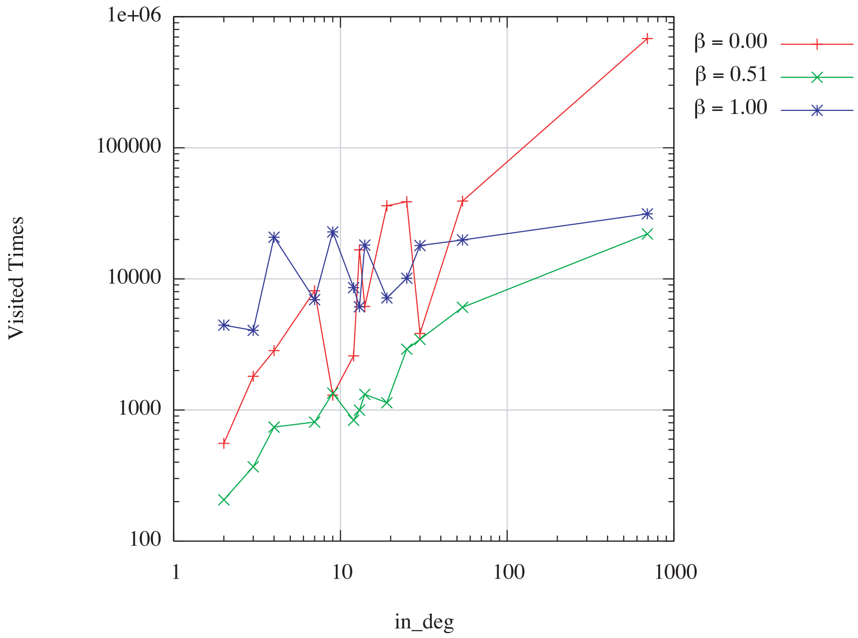

図 4.37: 逆辺をたどることを許した型 各入次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.01$)

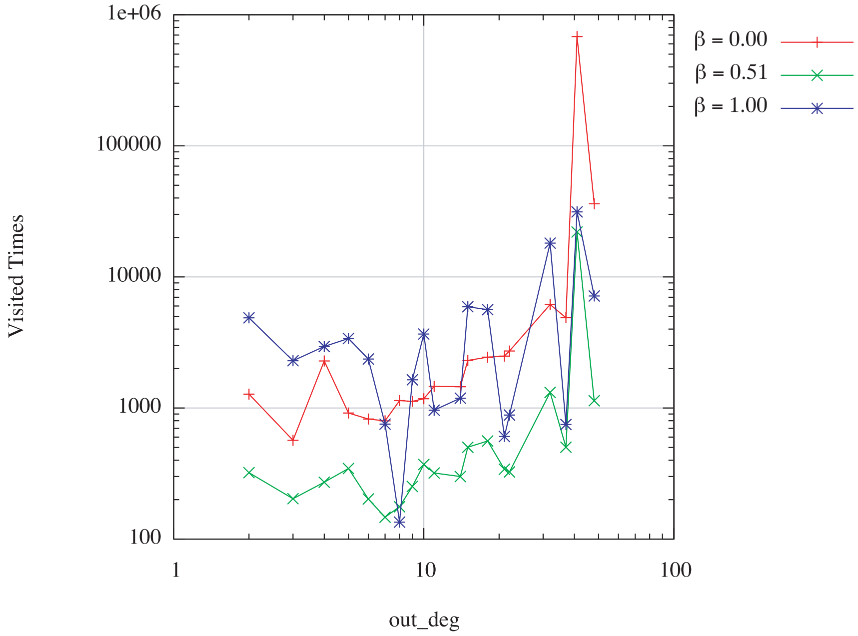

図 4.38: 逆辺をたどることを許した型 各出次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.01$)

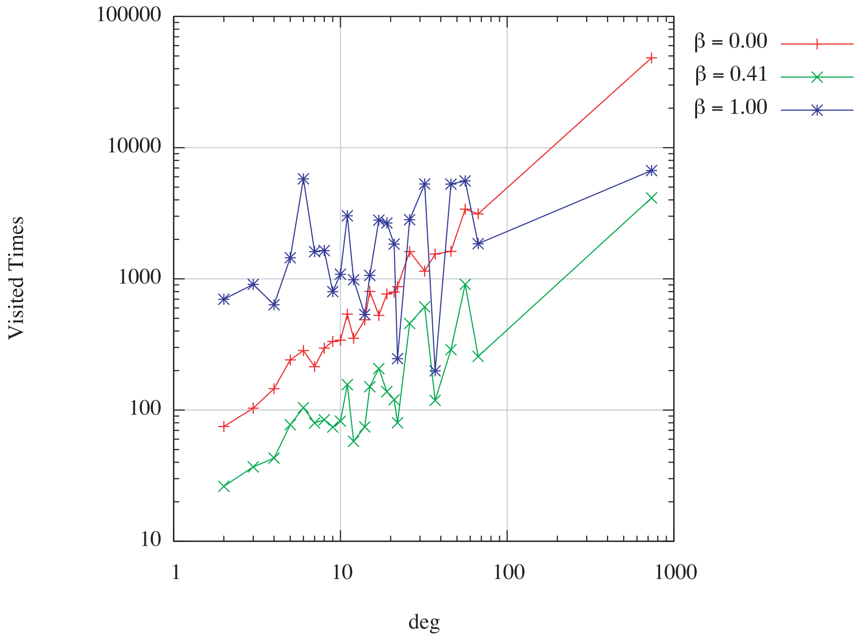

図 4.39: 逆辺をたどることを許した型 各次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.10$)

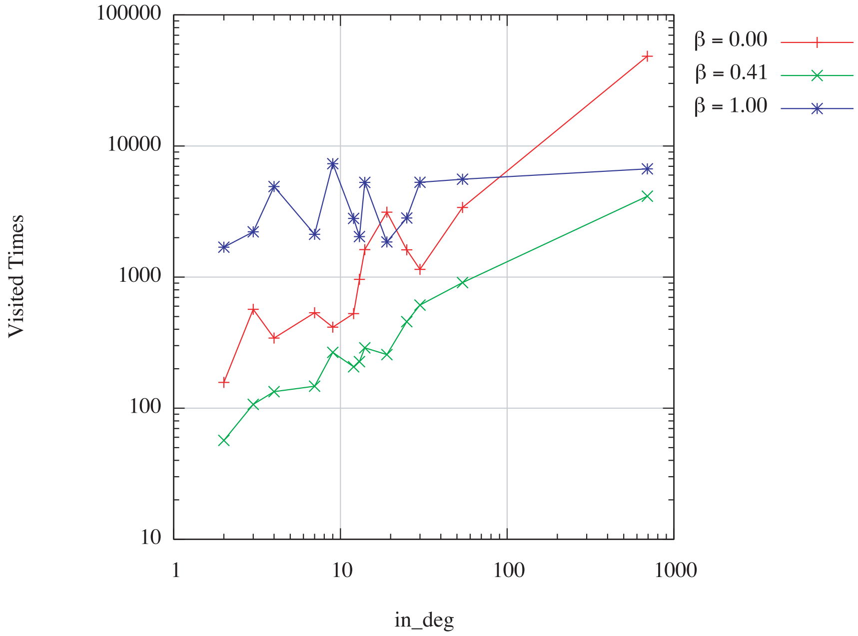

図 4.40: 逆辺をたどることを許した型 各入次数の頂点に対する訪問数 $(t = 1000, \gamma = 0.10)$

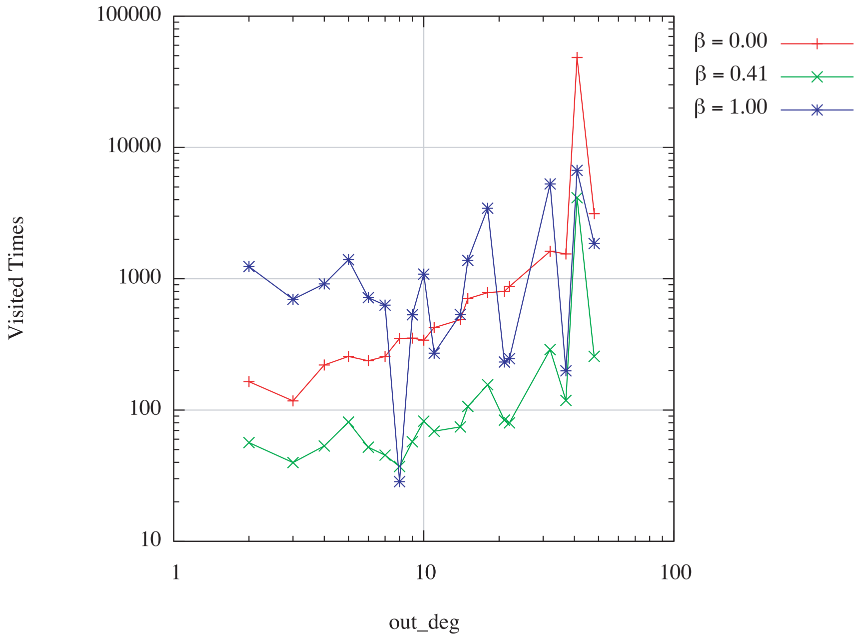

図 4.41: 逆辺をたどることを許した型 各出次数の頂点に対する訪問数 $(t = 1000, \gamma = 0.10)$

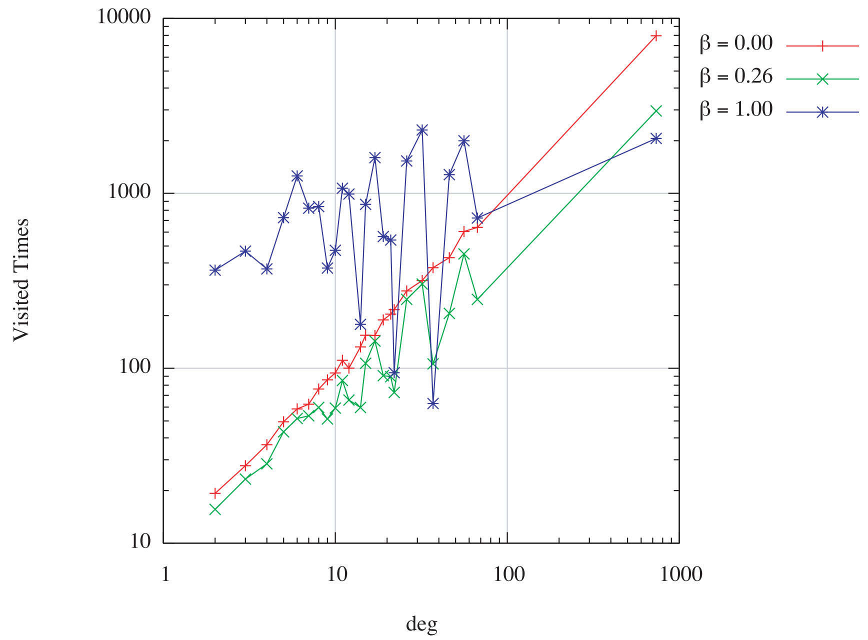

図 4.42: 逆辺をたどることを許した型 各次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.50$)

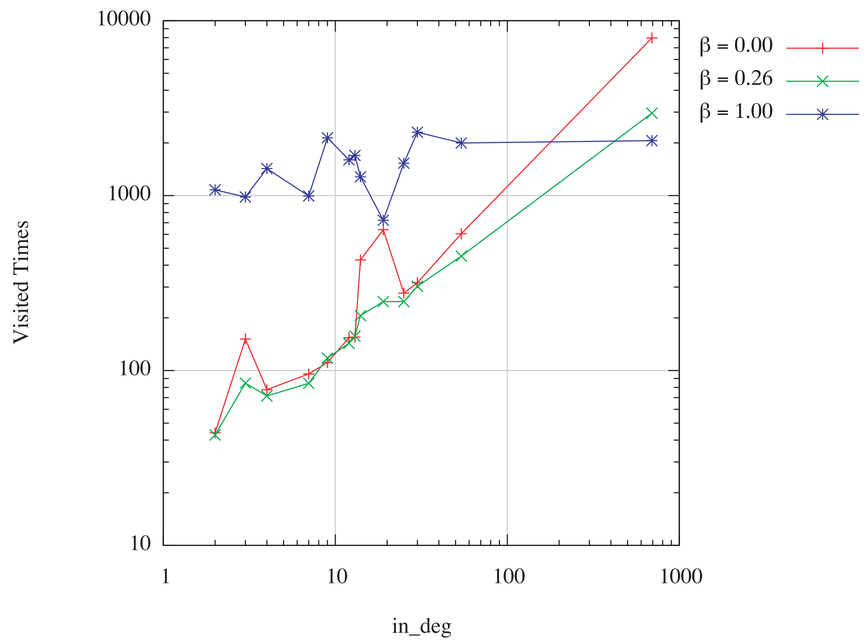

図 4.43: 逆辺をたどることを許した型 各入次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.50$)

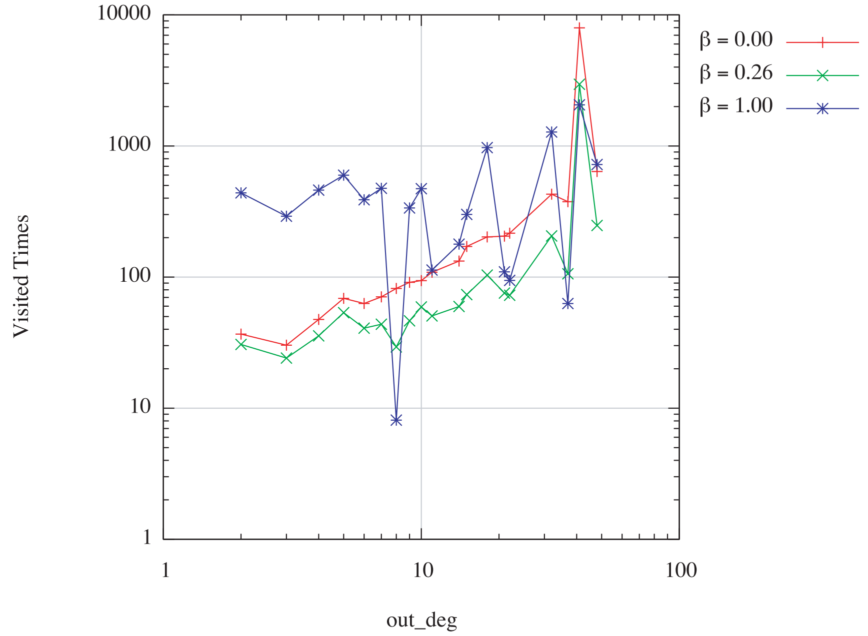

図 4.44: 逆辺をたどることを許した型 各出次数の頂点に対する訪問数 ($t = 1000, \gamma = 0.50$)

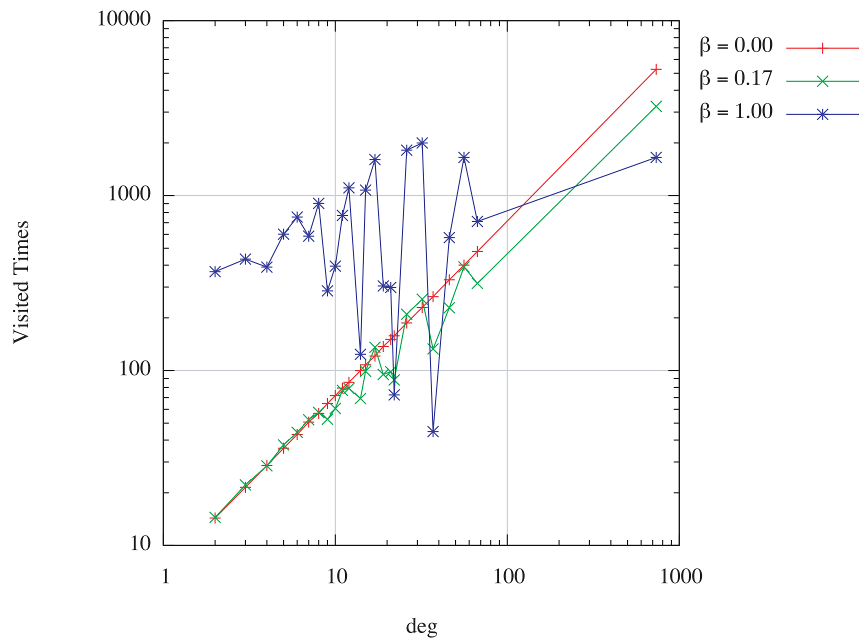

図 4.45: 逆辺をたどることを許した型 各次数の頂点に対する訪問数 ($t = 1000, \gamma = 1.00$)

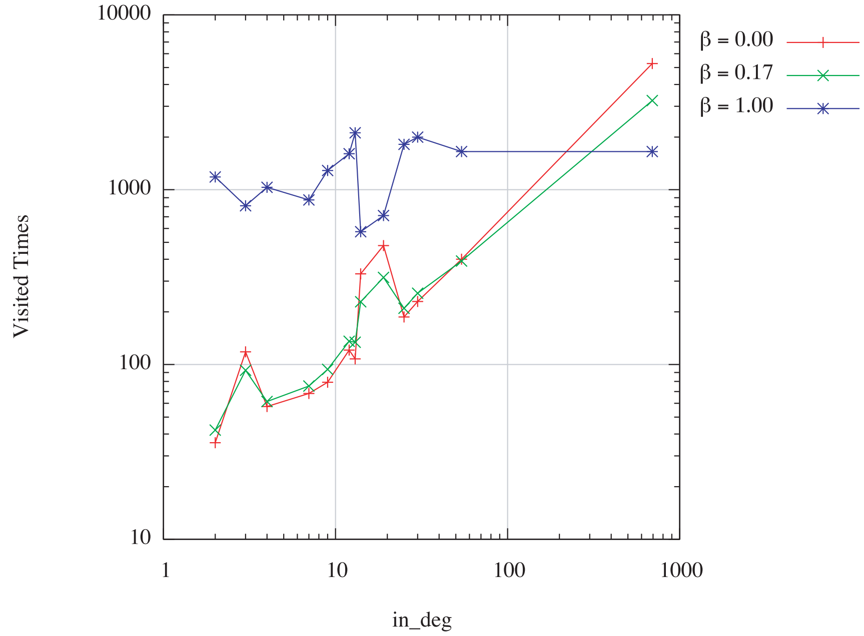

図 4.46: 逆辺をたどることを許した型 各入次数の頂点に対する訪問数 ($t = 1000, \gamma = 1.00$)

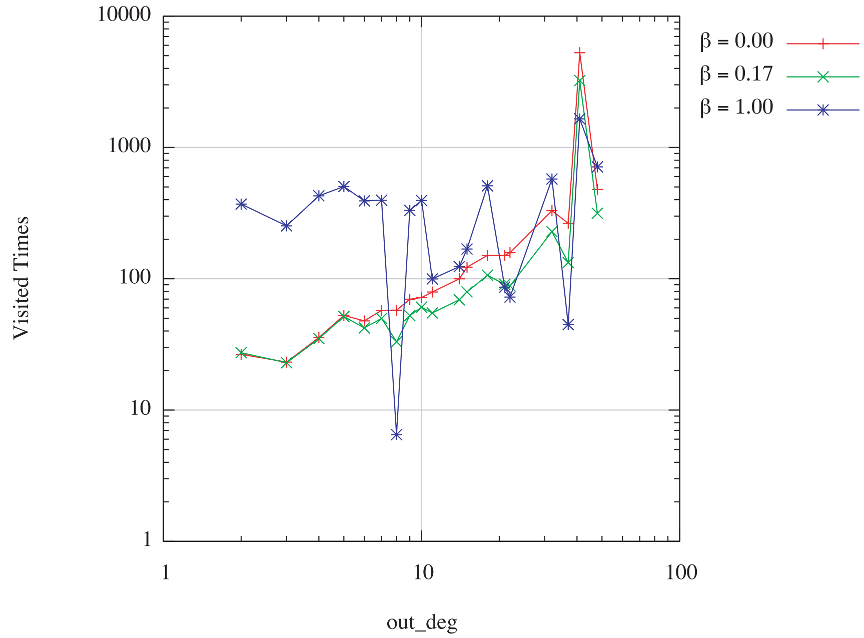

図 4.47: 逆辺をたどることを許した型 各出次数の頂点に対する訪問数 ($t = 1000, \gamma = 1.00$)

第5章　おわりに

本論文では, スケールフリーグラフ上での効率の良い ランダムウォークを実験的に示した.

スケールフリーグラフ上の従来のシンプルランダムウォークに関して, 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 [9][10]. 彼らの手法を拡張して, スケールフリーグラフ上の池田らの ランダムウォーク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，上原隆平助教授には, 常に熱心にご指導いただき, 深く感謝いたします. 浅野哲夫教授, 元木光雄助手, 寺本幸生氏をはじめとする情報基礎学講座の学生の皆様, 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．最後に, 研究生活を支援してくれた家族に感謝の意を記したいと思います.

参考文献

[1] [online]Available from: `http://www.google.com/`.

[2] [online]Available from: `http://www.boost.org/`.

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[4] B. Bollobás. *Modern Graph Theory*. Number 184 in Graduate Texts in Mathematics. Springer, New York NY, 1998.

[5] B. Bollobás. *Random Graphs*. Cambridge studies in advanced mathematics, 2001.

[6] B. Bollobás. Directed scale-free graphs. In *ACM Symposium on Discrete Algorithms (SODA)*, pages 132–139, 2003.

[7] 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.

[8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. In *Proceedings of the 9th International World Wide Web Conference*, pages 247–256, 2000. Available from: `http://www.almaden.ibm.com/cs/k53/www9.final/`.

[9] C. Cooper and A. Frieze. The cover time of the preferential attachment graph, manuscript, 2005.

[10] 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.

[11] 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.

[12] A. Gibbons. *Algorithmic Graph Theory*. Cambridge University Press, 1985.

[13] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. OXFORD UNIVERSITY PRESS, third edition, 2001.

[14] O. Haggstrom. Finite markov chains and algorithmic applications. In *London Mathematical Society*. 2002.

[15] S. Ikeda, I. Kubo, N. Okumoto, and M. 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.

[16] 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.

[17] N. Masuda, H. Miwa, and N. Konno. Geographical threshold graphs with small-world and scale-free properties. *PHYSICAL REVIEW E*, 71:036108, 2005.

[18] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, Cambridge, UK, 1995.

[19] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Computer Science Department, Stanford University, 1998. Available from: `http://dbpubs.stanford.edu/pub/1999-66`.