

BA モデルで生成されたネットワークにおける 次数と情報伝搬速度の関係について

谷聖一 研究室 枝國 雄太

Yuta Edakuni

概要

有限ネットワーク上の情報伝搬において、ネットワーク全体に情報が行き渡るまでの時間は、情報を保持する頂点が隣接する頂点の中から情報の伝搬先を選択する方法に依存する。単位時間に 1 つの頂点にのみ情報を伝搬できる伝搬速度限定モデルにおいて、Uncorrelated 性を持つネットワーク上では、次数の逆数に比例する確率分布に従い伝搬先の頂点を選択する方法が、ネットワーク全体に情報が行き渡るまでの平均時間が最小になると証明されている。過去の演習では、計算機実験を行い、修正 BA モデルを用いて生成したネットワークではこの結果が成り立たないことが報告されている。その原因として、修正 BA モデルで作成したネットワークが Uncorrelated 性を持たないことがあるのではないかと推測される。本演習では、BA モデルを用いてネットワークを生成する際、追加する頂点の次数を変化させることで情報の伝搬速度がどのように変化するかに着目し伝搬実験を行った。また情報を保持する頂点が隣接点との情報のやりとりの記憶をするかしないかという点にも着目し伝搬実験を行った。

1 はじめに

1.1 背景

私達の身の回りには多種多様なネットワークが存在している。例をあげると WWW、人間関係、交通網、論文の共著関係、食物連鎖の関係など、現実世界の関係を抽象化するとネットワークとして表せることが多い。現実世界に存在するネットワークは非常に膨大で複雑であるが、その中にある一定の法則をもつことから「複雑ネットワーク」と名付けられる。ネットワークの問題は頂点と辺を割り振ることでグラフ理論としてモデル化し、数理的に解析することが可能である。現実世界に多く存在し、数理的に解析が可能ということで、ネットワークの研究は幅広く行われてきた ([1])。

グラフ理論は 1736 年にオイラーという学者がケーニヒスベルクという町にある 7 つの橋をすべて 1 度だけ通り元の場所に戻ってこれるかという、一筆書き問題を考察した際に橋の位置関係をグラフに置き換えて考えたことが起源とされている ([2])。そこから研究の歩みは遅かったが 1998 年の Duncan J. Watts と S. H. Strogatz の研究 ([3]) により「スモールワールドモデル」という数学モデルを考案され、その性質が情報学以外にも社会学、経済学、生物学などの幅広い分野においてのネットワークにも共通してあてはまることから「複雑ネットワーク」として新たな飛躍を遂げた。

1.2 目的

論文 [4] で次が示された。

定理 1. 記憶無し伝搬速度限定モデルにおいて、ネットワークが Uncorrelated 性を持つなら、ネットワーク全体に情報が伝搬されるまでの平均時間は、情報を伝搬させる隣接頂点を選択する際に次数の逆数に比例した確率分布に従い乱択する方法が最小となる。

そこで 2011 年度の谷研究室卒業演習 [5] では、この結果の直用範囲を検証するために修正 BA モデルでスケールフリーネットワークを生成し計算機実験を行った。しかし、修正 BA モデルで生成したネットワークでは、「次数の逆数に比例した確率分布に従い情報伝搬する頂点を乱択する方法」の伝搬速度は速くなかった。その原因として修正 BA モデルで生成されたネットワークのスケールフリー性が低い可能性が考えられる。そこで 2012 年度の谷研究室卒業演習 [6] では、修正 BA モデルで生成したネットワークがどの程度のスケールフリー性を持つのか検証を行った、その結果、十分なスケールフリー性を持つという結果が得られた。2013 年度の谷研究室卒業演習 [7] では、BA モデル、修正 BA モデルで生成されたネットワークが定理 1 の仮定である Uncorrelated 性を十分に備えているかどうかを検証を試みたが、Uncorrelated 性を持つかどうか明確な結論が得られなかった。

修正 BA モデルで生成したネットワークで、「次数の逆数に比例した確率分布に従い情報伝搬する頂点を乱択する方法」の伝搬速度が速くなかった原因は、ネットワー

ク内の枝数が少ないことにより生成したネットワークの Uncorrelated 性が低いことが原因の可能性がある。実際に枝数の多いネットワークを作成し Uncorrelated 性を検証すること望ましいが、現状では Uncorrelated 性を明確に検証する方法が確立されていない。そこで本演習では BA モデルを生成する際、追加する頂点の次数を変化させると情報伝搬速度にどのような変化が起こるか検証する情報伝搬実験を行った。

1.3 構成

本稿は以下のような構成になっている。2 節では、グラフ理論の準備を行う。3 節では、スケールフリーネットワークについて述べる。4 節では、Uncorrelated 性について述べる。5 節では、伝搬速度限定モデルについて述べる。6 節では、実験方法について述べる。7 節では、実験結果について述べる。8 節では、今後の課題について述べる。

2 グラフ理論の用語

ネットワークについて論議したり、解析したりする場合、グラフ理論の知識が必要になってくる。ここでは必要な知識について簡単に説明する。グラフは、いくつかの点 (vertex) とそれらを結ぶ線分 (edge) からなる図形である。グラフ G は頂点集合 V と辺集合 E からなり、

$$G = (V, E)$$

で定義される。 V の要素を頂点、点、接点などと呼び、 E の要素を辺、枝、リンクと呼ぶこともある、2つの点が枝で直接結ばれている状態を2頂点は隣接しているという ([1])。

現実世界の関係を抽象化するとグラフで表せることを以下に WWW を例として紹介する。

- 頂点：ページ
- 辺：リンク
- 次数：ページのリンク数

図 1 において赤色の点 (頂点) はページを表し、ページとページを結ぶ直線 (辺) はリンクを示している。

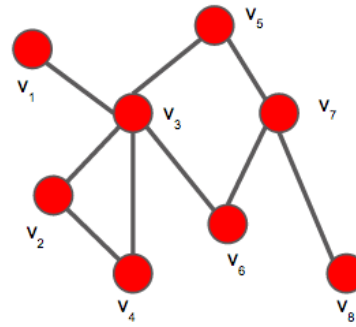


図 1: グラフ G

図 1 のグラフは

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$$

$$E = \{(v_1, v_3), (v_2, v_3), (v_2, v_4), (v_3, v_4), (v_3, v_5), (v_3, v_6), (v_5, v_7), (v_6, v_7), (v_7, v_8)\}$$

のように表せる。

3 スケールフリーネットワーク

3.1 スケールフリー性

現実世界の多くのネットワークはスケールフリー性を持ったスケールフリーネットワークである。スケールフリー性とは複雑ネットワークにおける 1 つの特徴である。その特徴は次数分布がべき則に従っているということである。次数分布がべき則に従っているというのは

$$P(k) \propto k^{-\gamma}, (\gamma > 0)$$

を満たしていることである ([1])。 (γ はべき指数と呼ぶ) つまり少数の頂点が枝を多く持っており (多数の頂点と隣接している)、多数の頂点がわずかな枝しか持たない (少数の頂点としか隣接していない) ということである。スケールフリーネットワークにおいて枝を多く持っている頂点のことをハブと呼ぶ。厳密に何本以上枝を持っていればハブであるという一般的な定義はない。

3.2 BA モデル

スケールフリーネットワークを生成するモデルはいくつか提唱されている。その中でも代表的なものの 1 つとして BA モデルが挙げられる。BA モデルは 1999 年に Barabasi と Albert によって提唱された ([8]) ネットワークのモデルであり、2 人の頭文字をとって BA モデルと呼ばれる。BA モデルには、「成長」と「優先的選択」という 2 つの特徴がある。「成長」というのは時間が経過するにつれて、頂点が次々とネットワークに加わっていき

ネットワークが大きくなっていくことである。「優先的選択」というのは新しく加わった頂点はその時点で次数が高い頂点と結ばれるということである。したがって一度次数が高くなった頂点は続いて枝を獲得しやすくなり、逆に一度次数獲得競争に敗れるとその後は新しい枝を獲得して他の頂点を抜いてハブになることは困難である。

3.3 BA モデルのアルゴリズム

ここでは BA モデルのアルゴリズムについて説明する。BA モデルの作り方にはいくつかあるが、本演習で用いたネットワーク作成アルゴリズムを説明する。

1. $m > 1$ 個の頂点からなる完全グラフを置く。(初期段階で m を定めると以後固定)
2. 新しい頂点を 1 個を追加し、既に存在している頂点のうちの m 個の頂点に対して、枝を張る。この時、新しい枝が張られる確率は、各頂点のその時点での次数 k_i に比例する。すなわち、

$$\frac{k_i}{\sum_{j=1}^n k_j}, (1 \leq i \leq n)$$

の確率である。(n は頂点数, k_i : 頂点 i の次数)

3. 指定の頂点数になるまで Step2 を繰り返す。

新しく追加される点の次数は m で固定なのでネットワーク全体の最低次数は m となる。

3.4 修正 BA モデル

修正 BA モデルとは生成されたネットワークの特徴量がより解析しやすいように BA モデルに修正を加えたモデルである。

3.5 修正 BA モデルのアルゴリズム

1. 枝を保有しない既存の頂点を 1 個置く。
2. 新しい頂点を 1 個追加し、既に存在している頂点のうちの 1 つに枝を張る。この時、新しい枝が貼られる確率は、各頂点のその時点での次数に比例する。
3. 指定の頂点数になるまで Step2 を繰り返す。

4 Uncorrelated 性

ここでは定理 1 で仮定している Uncorrelated 性について説明する。

Uncorrelated 性とはネットワークにおける様々な研究で用いられる性質の 1 つであるが、論文や文献によって定義が異なる。「Uncorrelated」とは「無相関の、無関係の」という意味である。そこから Uncorrelated ネットワークとは「局所性がない、どこも同じような構造をしている」ということを意味する。Uncorrelated 性を持つネットワークは、統計的、平均的に解析できることが多い。

5 伝搬速度限定モデル

5.1 ソースノードとターゲットノードの定義

ソースノードとターゲットノードの定義を以下のように行う。

- ソースノード：情報を持っている頂点のことである。ソースノードは隣接頂点にのみ情報を伝搬することができる。
- ターゲットノード：隣接するソースノードから情報の伝搬先として選択された頂点のことである。

ソースノードのことを情報源ともいう。

5.2 伝搬速度限定モデルの定義

伝搬速度限定モデルの定義は以下である。

- ソースノードは単位時間に 1 つの隣接点をターゲットノードとし情報伝搬が可能
- 情報を受けとった頂点は新たなソースノードとなる
- ソースノードは隣接点の次数のみわかる (既に情報を持っているかどうかは不明)

本演習で行った伝搬実験ではこの定義に「初期状態でのソースノードは 1 つ」という条件を加えたモデルを使用した

5.3 ターゲットの選択

本演習で用いたターゲットの選択方法は以下の 3 つである。

- 一様分布に従い乱択
- 次数が大きいものを優先的に乱択
- 次数が小さいものを優先的に乱択

ソースノード S がノード T をターゲットとして選択する確率は以下である。

$$q(S; T) = \frac{T \text{ の重み}}{S \text{ の隣接点の重みの総和}}$$

この式の重みはターゲットの選択方法によって次のように変化する。

一様分布に従い乱択した場合、重みは 1 になる

次数が大きいものを優先的に乱択した場合、重みは次数となる

次数が小さいものを優先的に乱択した場合、重みは次数の逆数となる

6 実験方法

6.1 記憶の有無

論文 [4] では「情報源は伝搬元と伝搬先の記憶を保持しない」という仮定であった。しかし演習報告 [5] では「情報源は伝搬元と伝搬先の記憶を保持する」という仮定で情報伝搬実験を行っていた。本演習では記憶を「保持する」「保持しない」の違いにも着目して、記憶有り無し両方の情報伝搬実験を行った。

6.2 情報伝搬シミュレーション

本演習では次の情報伝搬シミュレーションを行った

入力：ネットワーク N, 記憶の有無

処理：記憶の有無に関して指定された方法の伝搬速度限定モデルで、初期情報源をランダムに選び、すべてのノードが情報源となるまで情報伝搬をシミュレーションし情報源数を記録

6.3 実験方法

本実験で用いるネットワークは頂点数 10000 の BA モデル、修正 BA モデルを用いて作成した。BA モデルを生成する際、追加する枝の数は 2 本, 3 本, 5 本, 10 本, 15 本, 20 本, 30 本で固定して作成した。この方法で各モデル 100 個ずつネットワーク生成し、記憶有り無しそれぞれで伝搬速度限定モデルで情報伝搬シミュレーションし、情報源数の時間変化を記録した。本実験でのターゲット選択方法は一様分布に従い乱択、次数の小さいものを優先的に乱択、次数の大きいものを優先的に乱択で行った。

ネットワーク生成と情報伝播実験を行うプログラムは C++ で作成し、伝搬結果のグラフ出力には gnuplot を使用した。また実験プログラムの管理には bash スクリプトを用いた。

7 実験結果

ここでは実験結果の一部を紹介する。

7.1 修正 BA モデルの結果

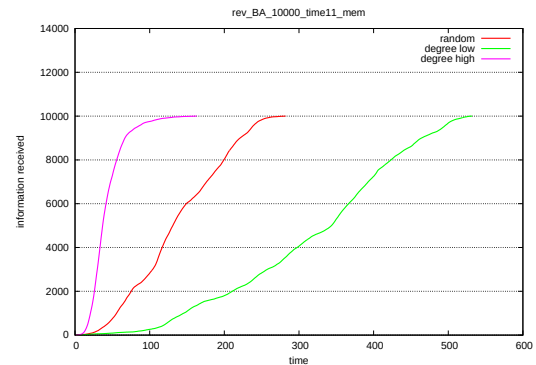


図 2:記憶有り

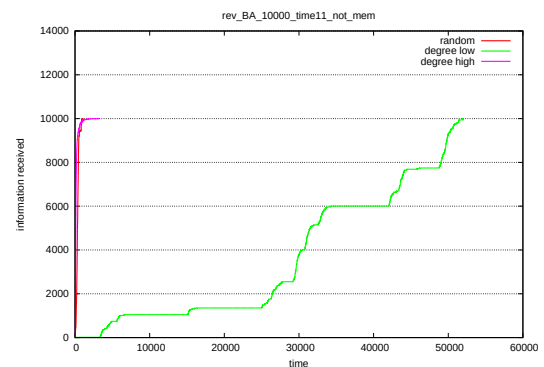


図 3:記憶無し

修正版 BA モデルでは記憶有り無し両方とも次数が小さいものを優先的にした乱択が伝搬速度が最も遅くなっている。

7.2 BA モデルの結果

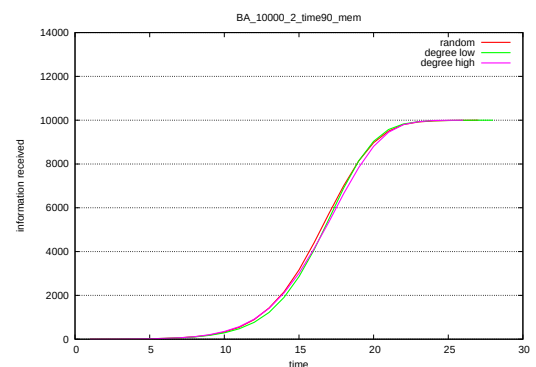


図 4:追加する枝 2 本 記憶有り

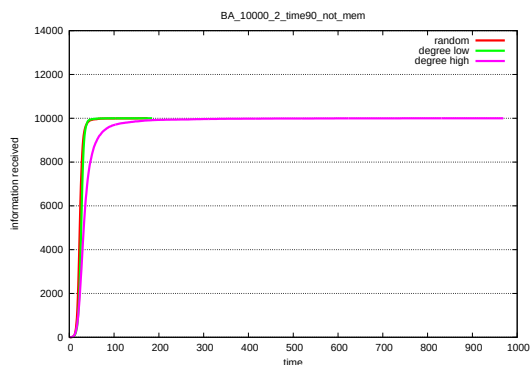


図 5:追加する枝 2 本 記憶無し

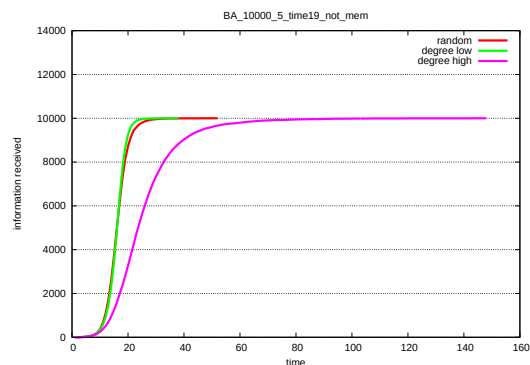


図 9:追加する枝 5 本 記憶無し

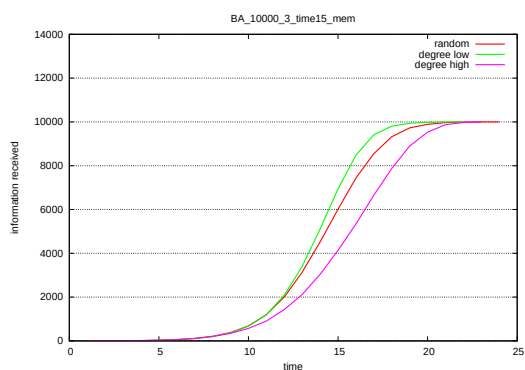


図 6:追加する枝 3 本 記憶有り

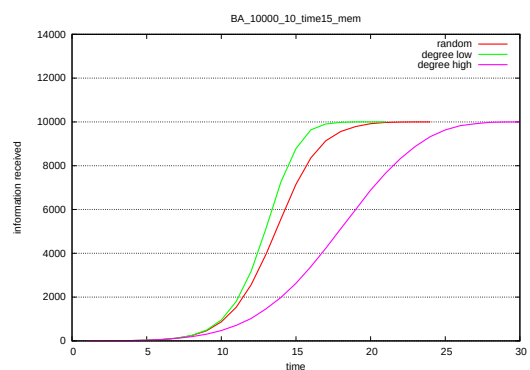


図 10:追加する枝 10 本 記憶有り

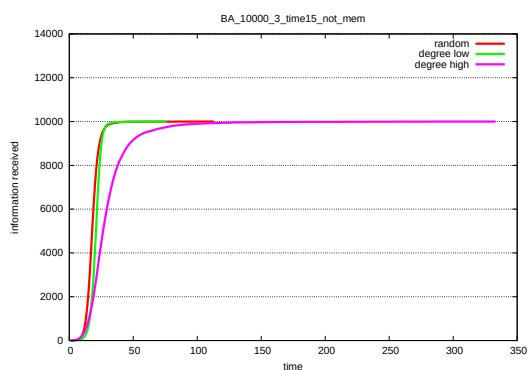


図 7:追加する枝 3 本 記憶無し

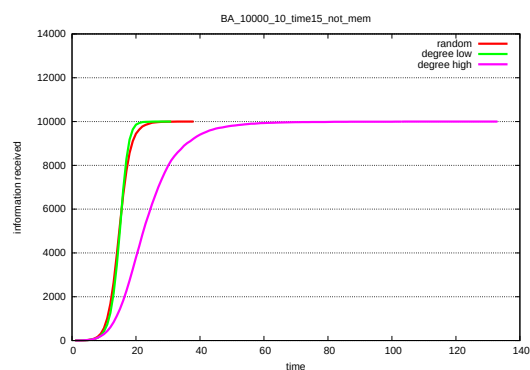


図 11:追加する枝 10 本 記憶無し

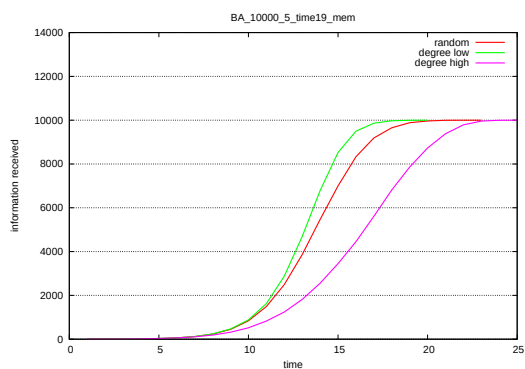


図 8:追加する枝 5 本 記憶有り

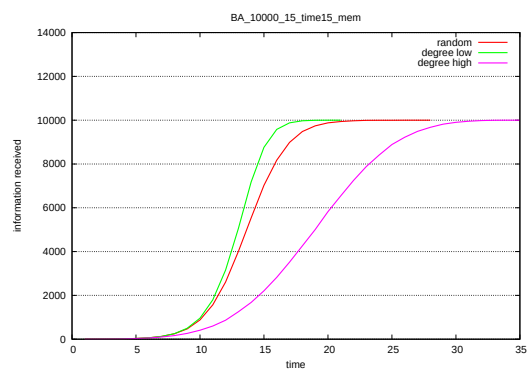


図 12:追加する枝 15 本 記憶有り

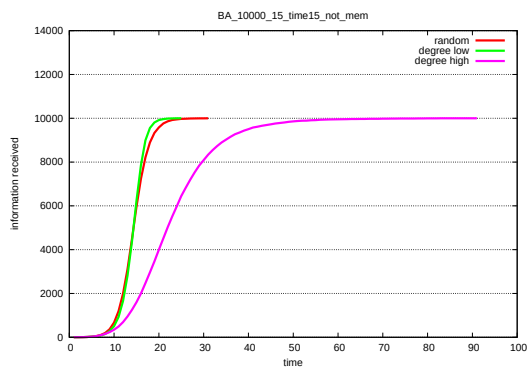


図 13:追加する枝 15 本 記憶無し

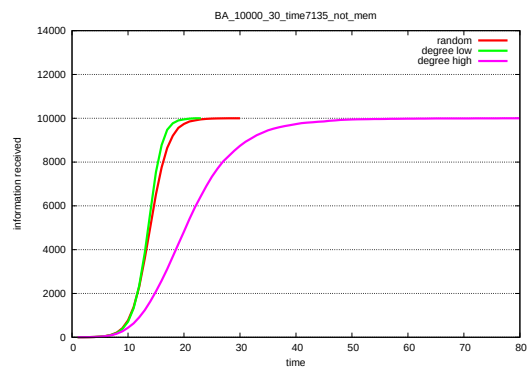


図 17:追加する枝 30 本 記憶無し

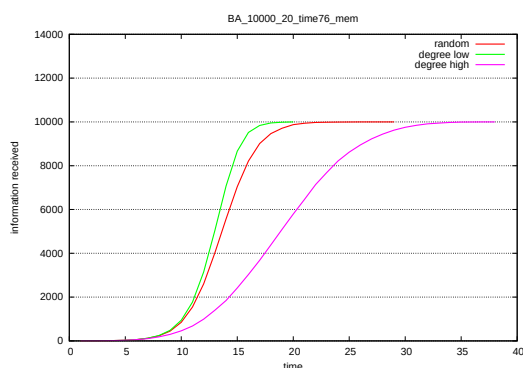


図 14:追加する枝 20 本 記憶有り

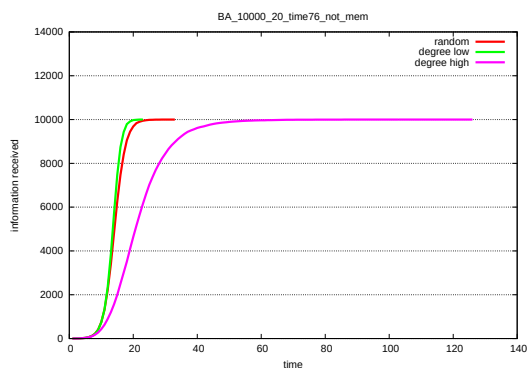


図 15:追加する枝 20 本 記憶無し

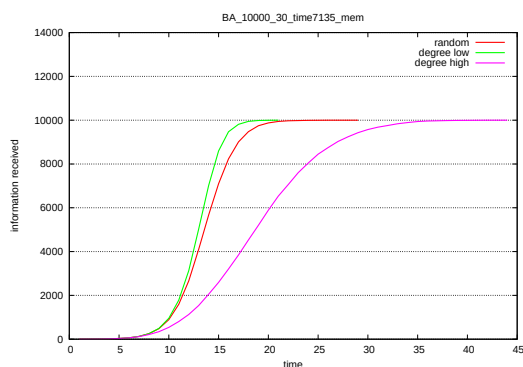


図 16:追加する枝 30 本 記憶有り

記憶有り無しにかかわらず枝数が増えると次数が小さい頂点を優先した乱択と他の乱択との伝搬速度差が大きくなっていることがわかる。また枝数が増えると「次数の小さいものを優先した乱択」の伝搬速度が速くなるという結果になっている。

7.3 枝数による伝搬結果の変移

表 1 は本実験で次数の小さい頂点を優先した乱択が最も速く伝搬できた回数を追加する枝数ごとにまとめた表である。

表 1: 成立したものをまとめた表

追加する枝数 m	記憶有り	記憶無し
m=1	0	0
m=2	6798	8385
m=3	9790	8474
m=5	9923	9282
m=10	9998	9928
m=15	9999	9962
m=20	10000	9977
m=30	9999	9983

表 1 を見てわかるように記憶有り無し両方、枝数が多いネットワークほど次数の小さいものを優先とした乱択が最も速く伝搬できた回数が多くなっている。

図 18 は枝数の変化による平均情報伝搬速度の変化を表したグラフである。

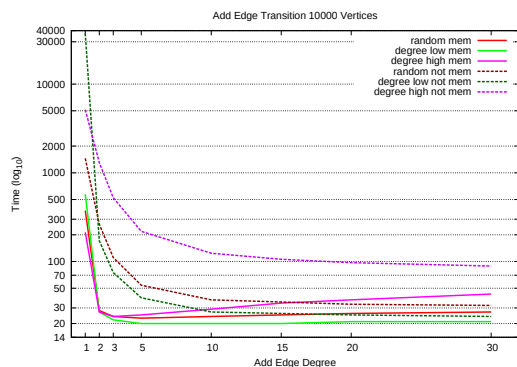


図 18: 枝数による平均情報伝搬速度の変化

図 18 を見ると初めは次数が小さいものを優先とした乱択が最も平均伝搬時間が長いですが、枝数が増えると最終的には記憶有り無し両方次数が小さいものを優先とした乱択が最も平均伝搬時間が短くなっている。

7.4 考察

今回の実験で修正 BA モデルで作成したネットワークはどのネットワークも次数の小さな頂点を優先的にする乱択が平均情報伝搬速度が最も遅いという結果が出た。これは初期情報源の数が 1 つであったことが原因の可能性がある。伝搬速度限定モデルにおいて次数が小さい頂点を優先した乱択で伝搬を行う場合、1 つの情報源がハブを伝搬先として選ぶ確率が低くなるが、ハブの隣接点のほとんどが情報源になっている場合、1 つの情報源から伝搬先として選ばれる確率が低くともどれか 1 つから伝搬先として選ばればよいのでハブに情報が伝搬される確率は低くなる。修正 BA モデルで生成されるネットワークは木構造であるので任意の 2 頂点間の経路が一意に定まる。したがって初期情報源が 1 つの場合複

数の頂点から同時に伝搬先として選択されることがないので、ハブに情報が伝わる確率が極端に低くなるのではないかと考えられる。また枝数の少ないネットワークはハブにより連結構造を保っていることも可能性の 1 つとしてあげられる。したがってハブに情報が伝わらないとネットワーク全体に情報が伝わらないのではないかと考えられる。

また今回の実験で枝数が増えると次数が小さい頂点を優先した乱択が最も速く伝搬できるネットワークが増えるという結果が出た。これは枝数が増え経路が増えることによりハブを経由せずとも大域に情報が伝搬されやすくなり、ハブの隣接点に情報が伝わりやすくなるからだと考えられる。

考察をまとめると伝搬速度限定モデルで次数が小さい頂点を優先した乱択で伝搬を行う場合、ハブを経由せずとも情報が大域に伝わるネットワークが最も速く伝搬できるのではないかと考えられる。これは Uncorelated 性に関係する可能性がある。

8 今後の課題

今回、枝数の多いネットワークは次数が小さい頂点を優先して情報伝搬する方法が平均伝搬時間が最も短くなるという結果が得られた。しかし生成したネットワークの Uncorelated 性を検証はしなかったので、生成したネットワークの Uncorelated 性を検証し、枝数が多いネットワークだと、次数が小さい頂点を優先して情報伝搬する方法が平均伝搬時間が最も短くなる理由が Uncorelated 性にあるかどうかを検証することが今後の課題である。

参考文献

- [1] 今野紀雄, 増田直紀 複雑ネットワーク 基礎から応用まで (近代科学社) 2010/4/30.
- [2] Guido Caldarelli, Michele Catanzaro 著, 高口太郎 訳, 増田直樹 監修 ネットワーク科学 つながりが解き明かす 世界のかたち (丸善出版) 2014/4/25
- [3] Duncan J. Watts, Steven H. Strogatz (1998) Collective dynamics of 'small-world' networks. Nature 393, 440-442.
- [4] Hiroshi Toyozumi, Seiichi Tani, Naoto Miyoshi, Yoshio Okamoto, Reversr preferential spread in complex networks.
- [5] 小幡 正博 作, 伝播速度限定モデルにおける Scale Free Network 上の情報拡散 Dynamics (http://www.tani.cs.chs.nihon-u.ac.jp/g-2012/kujira1220/obata_resume.pdf) 2015/2/12/20:53 UTC
- [6] 田中勇歩 作, 修正版 BA モデルで生成したネットワークのスケールフリー性判定計算実験 (http://www.tani.cs.chs.nihon-u.ac.jp/g-2013/ufo/resume/tanaka_resume.pdf) 2015/2/12/20:53 UTC

- [7] 阿部光太郎 大谷舞 作, BA モデルおよび修正 BA モデルで生成したネットワークの Uncorrelated 性について (http://www.tani.cs.chs.nihon-u.ac.jp/g-2014/nw/network_resume.pdf) 2015/2/12/20:53 UTC
- [8] Albert-Lszl Barabasi and Rka Albert (1999) Emergence of Scaling in Random Networks. Science 286, 509.