

BA モデルで生成されたネットワークと同じ次数列を持つ ネットワークの Uncorrelated 性について

谷聖一研究室 香取 秀柄・逸見 優聡

Shuheï Katori, Masatoshi Henmi

概要

有限ネットワーク上において、次数が大きな頂点ほど大きな影響力を有すると思われる。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せ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拓する方法が最小となることが証明されている。この定理を適用するにはネットワークが Uncorrelated 性を持つ必要があるが、その判定法は、報告者の知る限りでは確立されていない。本報告では、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拓する方法が効率の良いことと Uncorrelated 性を持つことの関連性を検討するため、BA モデルおよび修正 BA モデルで生成したネットワークと同じ次数列を持つネットワークを生成し、それらに対して情報伝搬実験を行った。

1 はじめに

1.1 背景

ネットワークとは「つながり」全般を指す。グラフ理論では、ネットワークは頂点と枝（辺）で表す。我々の身の回りに存在する多くの関係はネットワークとして捉えることができ、鉄道網や噂話はその例である。駅や人は頂点、駅間を結ぶ路線や人同士の関係は枝とみなせる。

現実世界に存在するネットワークは非常に大規模で複雑な振る舞いを生じる特徴があり、それらに共通して見られる普遍的な性質をもつことから「複雑ネットワーク」として研究されるようになった。ネットワークの問題はグラフとして抽象化することで、数学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])。

1.2 目的

緊急連絡はより早く全体に周知させたいことや、ウィルス感染はより流行を遅らせたいなど、情報の伝搬においてその速度制御が必要になることがある。このように、ネットワーク上に情報をいかに効率よく伝搬できるかという問題は重要な課題であり、様々なことに応用することができる。情報伝搬の速度は情報を伝搬する相手を選択するモデルに依存する。

論文 ([2]) 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拓する方法が最小となる。

そこで本演習に取り組む前に、BA モデルで生成したネットワーク上で成り立つかを検証するため、BA モデルでネットワークを生成し情報伝搬実験を行った。すると、枝数の小さな一部のネットワークにおいて「次数の逆数に比例した確率分布に従い乱拓する方法 (以下 IC とする)」の伝搬速度が速くないという結果を得た。対象のネットワークが Uncorrelated 性を備えているかによって異なる結果となった可能性があると考えられるため、特徴量を比較し Uncorrelated 性と関係する特徴量を探ることにした。

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、グラフ理論の概念について述べる。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、生成モデルについて述べる。5 節では、伝搬速度限定モデルについて述べる。6 節では、Uncorrelated ネットワークについて述べる。7 節では、実験方法について述べる。8 節では、実験結果について述べる。9 節では、今後の課題について述べる。

2 グラフ理論の概念

ここではネットワークを取り扱うため必要となるグラフ理論の概念に関して説明を行う。グラフ G は集合 V と集合 $E \subseteq V \times V$ からなり、

$$G = (V, E)$$

と定義される。 V の要素を点、頂点、ノードなどと呼び、 E の要素を辺、枝、リンクなどと呼ぶ。以下に友人関係を例にして紹介する。

- 頂点：人
- 辺：人と人の間の交友関係
- 次数：人のもつ交友関係の数

図 1 において水色の点 (頂点) は人を表し、人間を結ぶ直線 (辺) は交友関係を示している。この場合頂点 v_4 の次数は 6 となる。

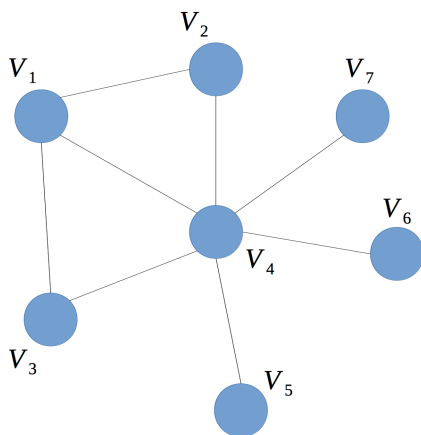


図 1: 友人関係

図 1 のグラフは、

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

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

のように表せる。

3 スケールフリー・ネットワーク

次数分布がべき則であるネットワークをスケールフリー・ネットワークという。

次数分布がべき則であるとは

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

を満たしていることである。(γ はべき指数と呼ぶ。)

つまりスケールフリー・ネットワーク上では、一部の頂点のみ次数が大きく、その他の多数の頂点は次数が小さい。そしてこの膨大な頂点と接続している頂点のことをハブという。世の中の多くのネットワークはこの性質を有していることが確認されている。またネットワークにまつわる驚きの多くが、この性質から出てくる ([1], [3])。

4 生成モデル

スケールフリー・ネットワークを生成するモデルは幾つか提唱されており、その 1 つとして BA モデルがあげられる。BA モデルとは 1999 年に Barasi と Albert によって提唱されたネットワーク生成モデルで、2 人の頭文字をとって BA モデルと呼ばれる ([1])。そして木構造のネットワークを生成するために BA モデルに修正を加えたモデルを修正 BA モデルという。本節では BA モデルおよび修正 BA モデルについて述べる。

4.1 BA モデルのアルゴリズム

BA モデルでネットワークを生成するアルゴリズムは以下のとおりである ([4])。

1. $m > 1$ 個の頂点からなる完全グラフを置く。
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 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 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)

3. 指定の頂点数になるまで Step2 を繰り返す。

BA モデルの特徴は既に多くの枝を持つ頂点がその後も新たな枝を獲得しやすく、少ない枝を持つ頂点はその後も新たな枝を獲得しにくいことである ([1])。

4.2 修正 BA モデルのアルゴリズム

修正 BA モデルでネットワークを生成するアルゴリズムは以下のとおりである ([5])。

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、新しい枝が張られる確率は、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.

修正 BA モデルも BA モデルと同様に、既に多くの枝を持つ頂点がある後も新たな枝を獲得しやすく、少ない枝を持つ頂点がある後も新たな枝を獲得しにくい特徴を持つ。そして生成されるネットワークは木構造となる。

5 伝搬速度限定モデル

情報伝搬を行うにあたって、情報を保持し他の頂点に情報を発信する頂点をソースノード、情報を送信しようと選択された頂点をターゲットノードと定義する。

- ソースノードは単位時間 (1 ステップ) に 1 つの隣接頂点をターゲットノードとして選択し情報伝搬が可能
- 情報を受けとった頂点は新たなソースノードとなる
- 各頂点は隣接頂点の次数を知識として有する

本演習では、更に「初期状態でのソースノードは 1 つ」、「各頂点が過去にどの頂点へ情報を送ったか、どの頂点から情報を受け取ったかの記憶を持たない」という条件を加えたモデルを使用した。

5.1 ターゲットノード選択確率の定義

ターゲットノードを選択する際の確率として次の式が定義されている。

$$q(b; a) = \frac{w_b}{\sum_{i=1}^{k_a} w_{a_i}} \quad (1)$$

w_i は頂点 i の重み、 k_i は頂点 i の次数、 a_i は頂点 a の隣接頂点の i 番目である。式 (1) はソースノードが a のとき隣接頂点 b がターゲットノードとして選択される確率である。

5.2 ターゲットノード選択方法

次数が低い頂点を優先するのが最適であると証明されているが、それは以下の 3 通りの選択方法を比較したものである。そして戦略によって式 (1) の w_i の値が変化する。ソースノードの隣接頂点の中から

1. 一様分布に従い乱拓 ($w_i = 1$)
2. 次数の大きな頂点を優先的に乱拓 (w_i : 頂点 i の次数)
3. 次数の小さな頂点を優先的に乱拓 (w_i : 頂点 i の次数の逆数)

6 Uncorrelated ネットワーク

ここでは、Uncorrelated ネットワークについて説明する。Uncorrelated ネットワークとはネットワークにおける様々な研究でよく用いられる性質の 1 つであるが論文や文献によって定義が異なる。そもそも「correlate」とは「互いに関係を持つ」という意味で「uncorrelate」とはその反意語なので「無相関の、お互いに関係をもたない」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」ネットワークということの意味する。

伝搬速度限定モデルにおいて、ソースノード a によって隣接頂点 b がターゲットノードとして選択される確率は式 (1) で求まることを 3 節で紹介した。論文 ([2]) では、与えられたネットワーク N が Uncorrelated ネットワークであると仮定し、式 (1) を N の次数分布を保持したままランダムに枝をつなげ直して得られるネットワークでの期待値

$$q_b = \frac{k_b w_b}{\sum_{j=1}^n k_j w_j} \quad (2)$$

で近似している。式 (2) は全頂点の中から頂点 b がターゲットノードとして選択される確率となっており、ソースノード a の選択に依らない。

7 実験方法

1. BA モデルで初期完全グラフの頂点数を 2 個, 3 個, 5 個, 10 個, 15 個として全頂点数 10000 個のネットワークを各 20 個生成する。また修正 BA モデルで全頂点数 10000 個のネットワークを 20 個生成する。組み換え元ネットワークは合計で 120 個となる。
2. 生成された各ネットワークに元にして、次数分布を保持したまま枝の組み替えを行い 100 個ずつネットワークを生成する。組み換え後のネットワークは合計で 1 万 2 千個となる。
3. 各ネットワークに対して次の特徴量を求める。
 - 半径
 - 直径
 - 平均距離
 - 次数相関
4. 各ネットワークに対して情報伝搬実験を行い IC 伝搬効率を測る。

なおネットワークの生成、ネットワークの組み替え、特徴量の調査のプログラムには C++ 言語を使用し、グラフ出力には gnuplot を使用した。

7.1 ネットワーク組み替え

一様分布に従って枝を繋ぎかえる方法では「連結・自己ループを含まない・多重辺を含まない」という条件を満たすネットワークが生成される割合が小さ過ぎる。実際に 20 万回程度の繋ぎ換えを試したが、上記の条件を満たすネットワークは全く生成されなかった。この方法で 1 万個のネットワークを生成するのは困難である。

他に、同じ次数分布のネットワークを全列挙し、それらの内から乱拓する方法が考えられる。木構造のネットワークに関しては、指定した次数列を持つ順序なし木を一つ当たり $O(n^3)$ で列挙する方法が提案されているが、この方法で頂点数 10000 のネットワークを全列挙するためには膨大な計算時間を要するので困難である ([6])。また、指定された次数列をもつ順序なしグラフを一つ当たり $O(m)$ で列挙する方法が提案されている ([7])。 m はグラフの枝数である。この方法ではネットワーク一つ当たり要する計算時間は大きくないが、組み替えて得られるネットワーク総数が大きいため全列挙するには膨大な計算時間が必要となる。よってこの方法も困難である。

そこで、偏りのある確率分布に従うが「連結・自己ループを含まない・多重辺を含まない」という条件を満たすネットワークを生成できる、独自に考案した次の方法によって枝の組み替えを実現した。なお、これらの方法によるネットワーク生成がどのような確率分布に従うかの確認はできていない。

まず、木構造のネットワークの組み替え方法について説明する。

木構造のネットワークの組み替え

1. ネットワークの次数 1 の頂点を除く全頂点の頂点番号を配列に格納。
2. 格納した配列の要素をランダムにシャッフル。
3. 配列の先頭から順に新しいネットワークに頂点を加えていく。その際、枝を繋ぐ先を宙に浮いた枝の端点の中からランダムに 1 つ選ぶ。残りの枝が接続する相手頂点は未定とする。
4. 新しいネットワークの宙に浮いた枝の端点に次数 1 の頂点をつなげる。

この方法では多重辺と自己ループを含まない木構造の

ネットワークが生成される。木構造であるため、多重辺と自己ループを含まなければ連結なネットワークである。よって、この方法で生成されるネットワークは「連結・自己ループを含まない・多重辺を含まない」という条件を満たす。

次に、木構造でないネットワークの組み替え方法について説明する。

木構造でないネットワークの組み替え

1. ネットワークの各頂点の頂点番号をそれぞれの次数の数だけ配列に格納。
2. 格納した配列の要素をランダムにシャッフル。
3. シャッフルした配列の隣り合う番号の頂点同士を枝で繋ぐ。ただし、それらの頂点を枝で繋ぐことで自己ループまたは多重辺が生じる場合、それらの番号を新たな配列に格納。
4. 配列が空になるまで Step2, 3 を繰り返す。100 回繰り返しても配列が空にならない場合、その組み換えを失敗とする。
5. Step4 で配列が空になったら、枝組み替え後のネットワークが連結となっているか確認。ネットワークが連結ならその組み替えを成功とし、非連結なら失敗とする。

この方法には失敗する 2 つの場合がある。Step2 で格納されている配列要素をどのように並べ替えても自己ループまたは多重辺が生じる場合と、枝組み替え後のネットワークが非連結である場合である。

まず、前者について説明する。初め、シャッフル回数の上限を設定せずに以上の組み替え実験を行った。すると Step2 で格納されている配列要素をどのように並べ替えても自己ループまたは多重辺が生じる場合、以降でシャッフルを繰り返しても条件を満たすネットワークは生成されないため、無限ループとなりプログラムが終了しなかった。この無限ループを防ぐには、現時点で自己ループと多重辺が生じない配列要素の並べ方が存在するか確かめ、存在しなければそれを終了条件とすればよい。しかし、これには配列要素に対して順列を全列挙する必要がある、配列のサイズが大きいときに計算量が膨大となるため困難である。従って、他の方法で無限ループを防ぐことを考えた。初めに行ったシャッフル回数の上限なしの組み替え実験より、シャッフル回数が 30 回を超える場合のほぼ全てにおいて、シャッフル回数が 1000 回を超えても配列が空にならないということがわかった。

そこで、十分なシャッフル回数として 100 回を上限に設定し、それ以上の組み替えを要する場合はプログラムを終了することにした。

次に、後者について説明する。ネットワークが連結であるかの確認は、ある頂点から他のすべての頂点への経路が存在するかを確認することによって行っている。なお、BA モデルで生成したネットワークが非連結となることはほとんどない ([5])。

7.2 特徴量の計測

ネットワークの半径、直径、平均距離、次数相関について述べる。 $G = (V, E)$ において、頂点 v から各頂点への距離の最大値を G の頂点 $v \in V$ の eccentricity という。この eccentricity の最大値を G の直径、最小値を G の半径という。また eccentricity が最小である頂点を中心という。

平均距離は任意の 2 頂点間の距離の平均である。

次数相関は隣接する 2 点の次数が似る度合いを測る概念である。次数 k の頂点の隣接頂点の平均次数は

$$k_n(k) \equiv \frac{1}{(k_i = k \text{ である頂点の数})} \sum_{\substack{i=1; \\ k_i=k}}^N \frac{1}{k_i} \sum_{\substack{j=1; \\ (v_i, v_j) \in E}}^N k_j$$

で求められる。ここで、 k_i は頂点 v_i の次数、 $\sum_{i=1; k_i=k}^N$ は次数が k である頂点について和をとる、 $\sum_{j=1; (v_i, v_j) \in E}^N$ は v_i と隣接する v_j について和をとる、という意味である。 $k_n(k)$ が k とともに増える傾向があれば正の次数相関があり、減る傾向があれば負の次数相関があるという。

7.3 IC 伝搬効率

先に紹介した 3 つのターゲット選択方法で 100 回の情報伝搬実験を行い、全頂点が情報源となるまでに要したステップ数を比較する。IC の総ステップ数が最も小さかった割合を百分率で表し IC 伝搬効率とする。

8 実験結果

ここでは実験結果の一部を紹介する。なお、本演習ではネットワーク特徴量として半径、直径、平均距離、次数相関を求めたが、同じ次数分布のネットワークの半径と直径にほとんど違いがなかったため、本報告では平均距離と次数相関についてのみ紹介する。また、本演習で生成したデータやグラフは <http://www.tani.cs.chs.nihon-u.ac.jp/g-2015/n2/> で公開している。

8.1 次数相関

図 2～7 は横軸に次数、縦軸に隣接頂点の平均次数をとり次数相関を示している。図 2～6 は BA モデルで生

成したネットワークの次数相関、図 7 は修正 BA モデルで生成したネットワークの次数相関である。

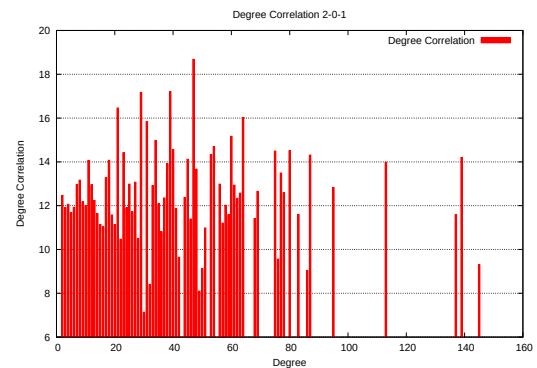


図 2:追加する枝 2 本

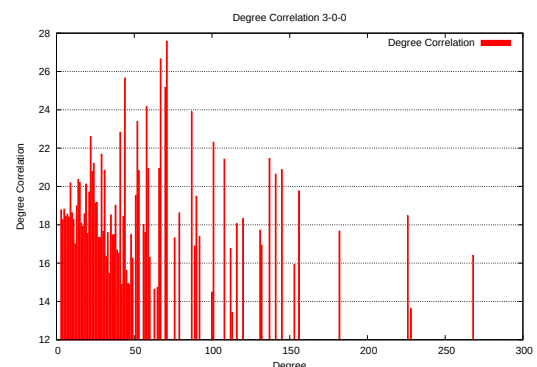


図 3:追加する枝 3 本

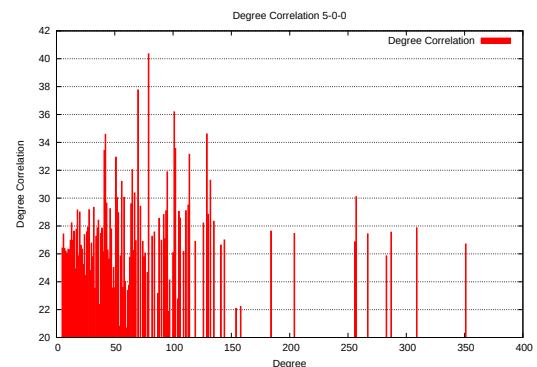


図 4:追加する枝 5 本

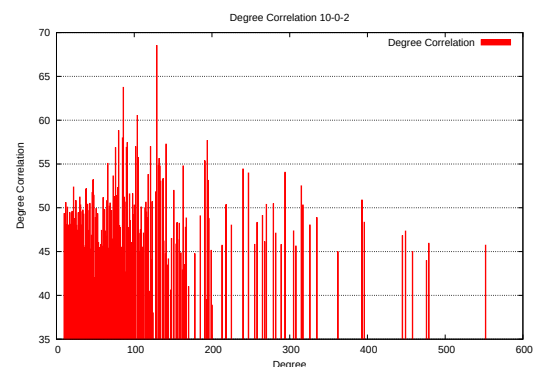


図 5:追加する枝 10 本

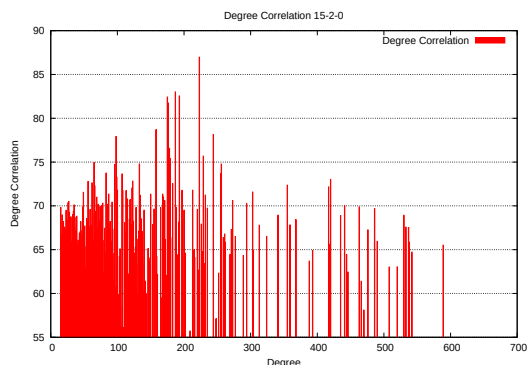


図 6:追加する枝 15 本

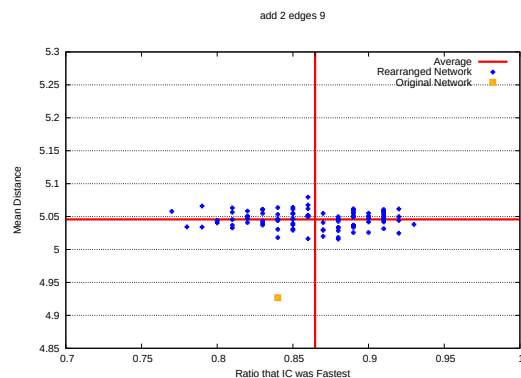


図 8:追加する枝 2 本

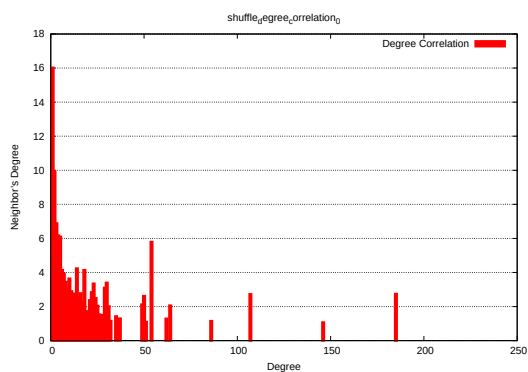


図 7:修正 BA

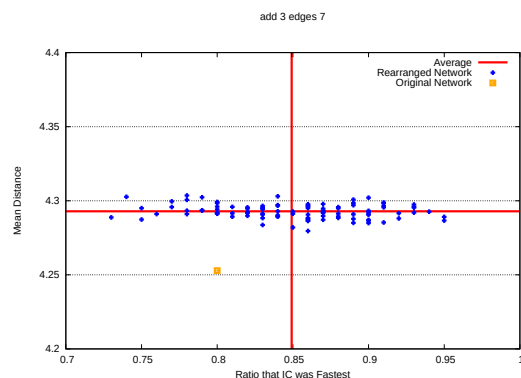


図 9:追加する枝 3 本

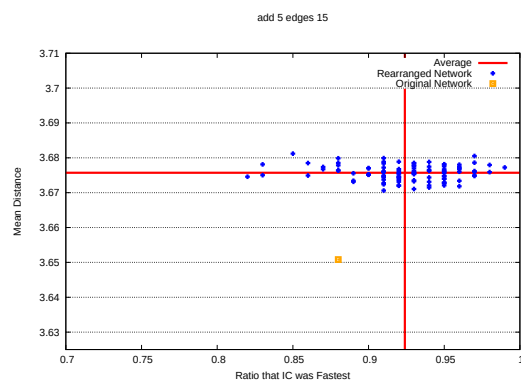


図 10:追加する枝 5 本

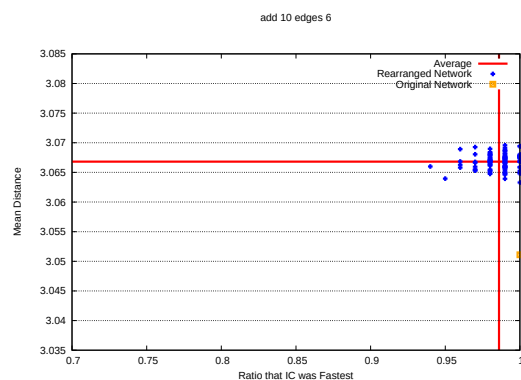


図 11:追加する枝 10 本

BA モデルで生成したネットワークの次数相関は、次数の小さな頂点の隣に次数の小さな頂点がいやすく、ハブの隣にハブがいやすい。修正 BA モデルで生成したネットワークの次数相関は、次数の小さな頂点の隣にハブがいやすく、ハブの隣に次数の小さな頂点がいやすい。

8.2 IC 伝搬効率と平均距離の関係

修正 BA で生成したネットワークでは IC 伝搬効率が常に 0% という結果を得たため、IC 伝搬効率は平均距離に依らないことがわかった。そのため、修正 BA で生成したネットワークについては省略する。

図 8～12 は横軸に IC 伝搬効率、縦軸に平均距離をとり、点はあるネットワークを示している。オレンジ色の点が組み替え元ネットワークで、青色の点が組み替え後ネットワークである。そして赤色の線は同じ次数分布のネットワーク 100 個における、IC 伝搬効率の平均と平均距離の平均を示している。

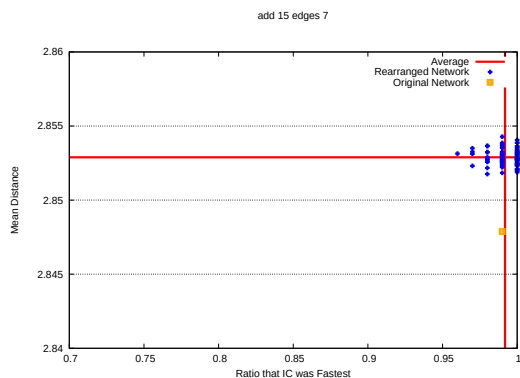


図 12:追加する枝 15 本

これから以下のことがわかる.

- 平均距離は大きくても 5.3 以下.
- IC 伝搬効率は低くても 0.7 以上.
- 同じ次数分布のネットワークの平均距離は高々0.1 程度.
- 組み替え後ネットワークの平均距離は組み替え元ネットワークの平均距離より大きくなる傾向.

そして, 上で得た結果を次数列ごとの平均値についてまとめたものが図 13~17 である. 赤色の点はある次数列を示しており, 上図の赤線の交点と等しい.

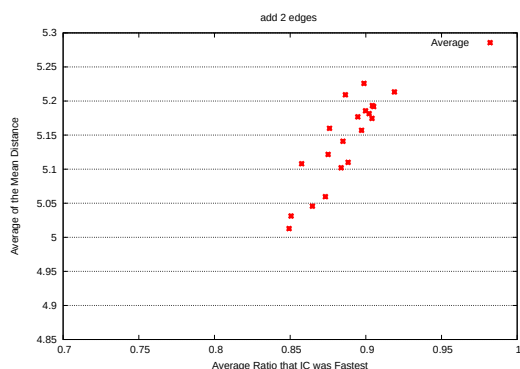


図 13:追加する枝 2 本

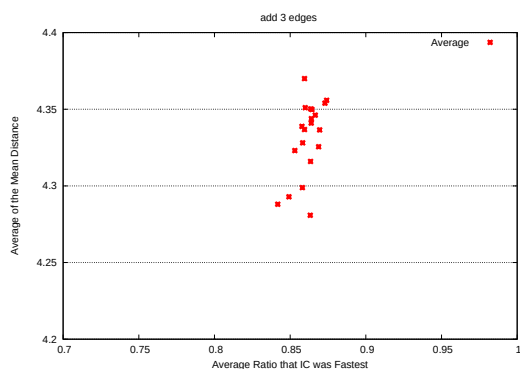


図 14:追加する枝 3 本

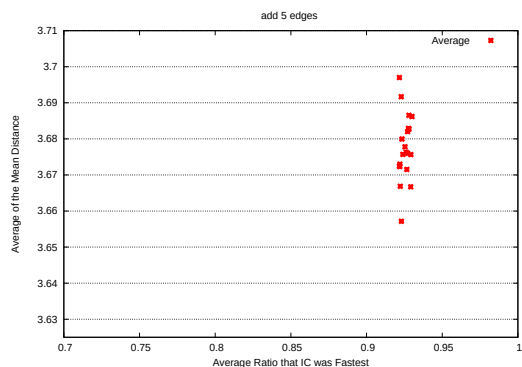


図 15:追加する枝 5 本

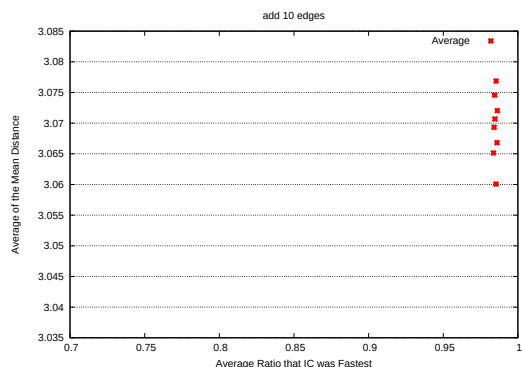


図 16:追加する枝 10 本

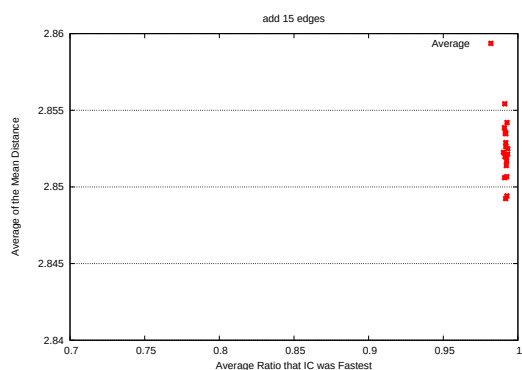


図 17:追加する枝 15 本

これから以下のことがわかる.

- 枝数が多いほど IC 伝搬効率が低い傾向.
- 追加する枝 2 本で生成したネットワークでは平均距離が大きいと IC 伝搬効率が低い傾向.
- 枝数が多いと IC 伝搬効率は平均距離に依らない.

8.3 考察

ハブは隣接頂点の 1 つからターゲットとして選択される確率が低くても, 多数の隣接頂点が情報源ならば少なくともその内の 1 つからターゲットとして選択される確率は低くならない. そのため, IC による伝搬が有効と成り得ると考えられる. しかし, 木構造のネットワークで

初期情報源を 1 つとして情報伝搬実験を行う場合、複数の情報源が情報を持たない同一頂点をターゲットとして選択することはない。従って、修正 BA で生成したネットワークでは IC 伝搬効率が低いのだと考えられる。

BA モデルで生成したネットワークではハブ同士が隣接しやすい。組み替え元ネットワークで隣接していたハブ同士が、組み替え後ネットワークで次数の小さい頂点を介して繋がる場合、最短経路がそれらのハブを経由する 2 頂点間の距離は大きくなる。そのような場合が稀に起こるため、組み替え後ネットワークの平均距離が組み替え元ネットワークの平均距離よりわずかに大きくなる可能性がある。

枝数が多いほど IC 伝搬効率が低い傾向が見られた。

そして、追加する枝 2 本で生成したネットワークでは平均距離が大きいと IC 伝搬効率が低い傾向が見られた。これらを合わせると、枝数が多いほど IC 伝搬効率が低く、同じ枝数のネットワークに関しては平均距離が小さいほど IC 伝搬効率が低いということが考えられる。

9 今後の課題

本演習では BA モデルによってネットワーク生成を行ったが、追加する枝数を 1 大きくすると総枝数がほぼ総頂点数分増えることになるため、追加する枝数の違いによる変化に飛躍があった。従って、追加する枝数を指定した上限以下のランダムな整数とする BA モデルにより生成し、枝数と IC 伝搬効率の関係をより詳細に調べ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] Hiroshi Toyozumi, Seiichi Tani, Naoto Miyoshi, Yoshio Okamoto, Reversr preferential spread in complex networks.
- [3] Amaral, L.A.N. et al, "Classes of small-world networks", Proceedings of the National Academy of Sciences of the United States of America 97, No. 21, pp. 11149-11152 (2000).
- [4] Barabási, A.-L., and Albert, R., "Emergence of scaling in random networks", Science 286, pp509-512 (1999).
- [5] 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).
- [6] 石川雅信 中野真一 指定した次数列を持つ順序なし木の高速列挙 (2014) 一般社団法人電子情報通信学会 電子情報通信学会論文誌 D Vol.J96-d No.11 pp2710-2715.
- [7] 佐藤圭介, 中野真一 指定された次数列をもつグラフの列挙 (2008) 一般社団法人電子情報通信学会 電子情報通信学会論文誌 A Vol.J91-d No.7 pp716-725.