

Scale Free Network 上における伝搬速度限定モデルの 情報拡散シミュレーション

谷聖一 研究室 古池 琢也
Takuya Furuike

概要

ネットワーク上の情報伝播において、情報がネットワーク全体に行き渡るまでの時間は、情報を保持する頂点が隣接頂点の中から情報を伝える頂点を選択する方法に依存する。本演習では、情報を保持する頂点は 1 単位時間に 1 つの隣接頂点にのみ情報を伝えられる伝播速度限定モデルで、各頂点に与えられるネットワークに関する情報は隣接頂点の次数のみという条件のとき、どの頂点の選択方法を採用すれば、情報がネットワーク全体に行き渡るまでの平均時間が最小になるか、計算機実験を行った。また、情報を持つ頂点が情報を伝達する頂点を選択する方法として、一様分布に従って選択、次数の昇順に優先的選択、次数に降順に優先的選択という 3 つの方法で実験を行った。

1 はじめに

現実世界には、様々なネットワークが実在している。ネットワークとは、つながり全般を指し、私達の身の回りには、人間関係、交通網、道路網、インターネット網といった、多くのネットワークが関係している。それらのネットワークの繋がり方は複雑であり、かつ、ある一定の性質を見出すことができることから、「複雑ネットワーク」と呼ばれる。その性質とは、「スケールフリー性」、「スモールワールド性」、「クラスター性」である。複雑ネットワークの研究は、広範囲の分野において行われており、現実世界の様々な現象を説明する新たな枠組みとして注目を集めている ([1])。

ネットワークは、グラフを用いてモデル化できる。グラフとは、数学的に物事を研究するグラフ理論の分野の語法であり、ネットワークをグラフとして抽象化することで、ネットワークの研究が進められてきた。1998 年の Duncan J. Watts と S. H. Strogatz の複雑ネットワークの論文により、ネットワーク理論が急速に発展した ([1])。

このような解析対象の一つに、ネットワーク上の情報の伝播がある。近年、著しく私達の周りに定着した Twitter や Facebook などの SNS で、ニュースや噂が短時間で広範囲に伝達されることがあるが、これも情報伝播の 1 つである。

いかに効率良くネットワーク上に情報を広めるかということは重要な課題であり、様々な問題に応用することができる。病気の流行、コンピュータウイルス、噂の広がりなど多くの例が挙げられる。

ネットワーク上の情報伝播において、情報がネットワーク全体に行き渡るまでの時間は、情報を伝える頂点を選択する方法に依存する。また、多くのリンクを持つ頂点

は情報伝播において大きく影響する。その頂点に情報が伝われば、多くの頂点に情報拡散が見込まれるからだ。よって、優先的に多くのリンクを持つ頂点に向かって情報を伝播するほうが速く情報伝播を達成することができると考えられる。だが、隣接する頂点が多いため、隣接頂点から情報を受け取る機会も多いと考える。無駄な伝播は情報伝播において、効率が悪くなる原因となる可能性がある。ネットワークを介して、効率よく情報を広げることが、余計な伝播を減らすことが鍵であることがわかる。どのように情報を伝える頂点を選べば、効率よく情報を伝播できるかは、ネットワーク上に情報を伝播する様子を解明していく分野では重要である。

2012 年に、論文 ([2]) で、情報源が 1 単位時間に 1 つの頂点のみ情報を送ることができる場合、次数が少ない頂点を優先して情報を送る方法がネットワーク全体に無駄が少なく、効率よく情報を伝播できることを、ある仮定のもとに示した。その結果の妥当性を検証するため、([3], [4]) で、計算機実験が行われた。その結果は相反するものであった。

実験で使用したネットワークはどの程度スケールフリー性を有するのか、計算実験方法の妥当性、論文で使った仮定の妥当性を検討する必要がある。そこで、本演習では、実際に Scale Free Network を生成し、情報を伝播するシミュレーションを行うプログラムを作成し、伝播時間を計算する実験を行った。

2 グラフ理論の概念

ネットワークを取り扱うため、グラフ理論に関する準備を行う。グラフは、いくつかの点 (vertex) とそれらを結ぶ線分からなる図形である。線分は辺 (edge) と呼ばれ、どの 2 点を結んでいるかが重要である。従って、グラ

フ G は,

$$V = \{ v_1, v_2, \dots, v_N \}, E = \{ e_1, e_2, \dots, e_N \}$$

の点集合 V と辺集合 E からなる図形である. 1 本の辺は,

$$e_i = (v_a, v_b) \quad (a, b \text{ は } 1, 2, \dots, N \text{ のどれか})$$

と表される. なお, 点を頂点, またはノード, 辺を枝, またはリンクと呼ぶこともある. また, グラフ G の点 v に対して, v と接続する辺の個数を G における v の次数 (degree) という ([1],[5]).

図 1 のグラフは,

$$\begin{aligned} V &= \{ v_1, v_2, v_3, v_4, v_5 \} \\ E &= \{ (v_1, v_2), (v_1, v_6), (v_1, v_7), \\ &\quad (v_2, v_3), (v_2, v_4), (v_2, v_5), (v_6, v_8) \} \end{aligned}$$

のように表せる.

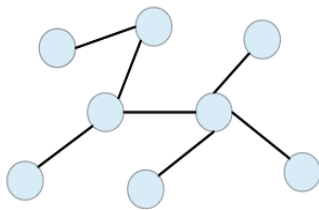


図 1: グラフ

3 Scale Free Network

複雑ネットワークの特徴の 1 つである「スケールフリー性」. 世の中の多くのネットワークの次数分布は, ある性質を有していることが確認されている. その性質は, ネットワーク上の次数分布がべき則であること. べき則とは,

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

を指す. (\propto は比例を表し, γ はべき指数と呼ぶ.)

直感的に説明すると, 一部の人には多くの友人がいる一方, 大部分の人は, 小数の友人しかいないという状況である. 地震の規模, 都市の大きさ, 生物の種類などの分布も, べき則となる.

つまり, Scale Free Network とは, 一部の頂点が大きな次数を有するが, 大部分の頂点は小さな次数を有するネットワークである ([1]).

Scale Free Network において, 次数が巨大な頂点のことをハブという. ハブはネットワークにおいて, 大きな影響を与えられられる.

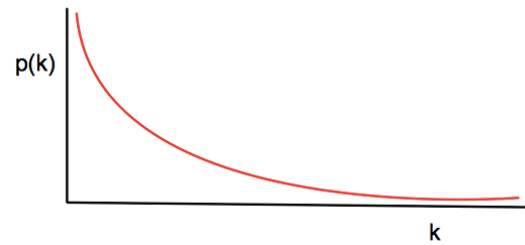


図 2: べき則

4 修正版 BA モデル

4.1 BA モデル

Scale Free Network を生成するモデルとして, BA モデルがある. BA モデルとは, 1999 年に, Barabási と Albert が提案したスケールフリー性を実現されるネットワークのモデルである. この BA モデルには, ネットワークの成長と優先的選択という 2 つの特徴がある.

- 成長...時間とともに頂点と枝が増えること.
- 優先的選択...新しく加わる頂点と結びつく既存の頂点を選ぶ際, その時点で次数の高い頂点と結びつきやすい.

この BA モデルを確率論の立場から解析はしようとすると, 条件を厳密に取り扱うのは一般に難しくなってしまう. そこで, 本演習では, BA モデルを簡略化し, 1 本ずつ辺を追加する修正版 BA モデルを用いて実験を行う.

4.2 修正版 BA モデル生成方法

1. 枝を保有しない既存のノードが 1 つ存在する.
2. 既存のグラフに対して, 頂点を 1 つ追加する.
3. 1 つの既存の頂点を選択し, 追加した頂点から枝をはる. 頂点が選択される確率は頂点の次数に比例.
4. 以後, 2 と 3 を繰り返す.

次数が高くなった頂点は, その後も新しい枝を獲得しやすくなり, ハブになりやすい. また, 次数獲得競争に一度破れると, たちまち新しい頂点を獲得する確率が低くなる.

5 伝播規則

5.1 用語

ソースノード ネットワーク上において伝播させる情報を持っている頂点のことを指す.

ターゲットノード ネットワーク上においてソースノードが次に伝播させる頂点として選択する頂点のこと。

5.2 初期状態

ネットワーク上の頂点はすべて情報を持っていない。また、最初の情報源をランダムで選ぶ。

5.3 伝播規則

ソースノードは、1 単位時間に隣接頂点の 1 つをターゲットノードとして選択し、情報を伝播する。一度情報を受け取った頂点はソースノードとなり、隣接頂点の中からターゲットノードを選択し、情報を発信し続ける。また、各ソースノードはターゲットノードを選択するとき、隣接頂点の度数だけがわかっていて、隣接頂点がすでに情報を受け取っているかの知識はもっていないとし、すでに情報を受け取った頂点にさえ情報を送る可能性がある。全ソースノードが、一斉にターゲットノードを選択し、情報を伝播する、これを 1 step とする。

5.4 ターゲットノードの選択方法

本研究では、ターゲットノードの選択方法として、3 つの方法を採用した。

- 一様分布に従って選択
- 頂点の度数に従って選択
- 頂点の度数の逆数に従って選択

隣接頂点の中から、頂点 a がターゲットノードとして選ばれる確率 q_a は、

$$q_a = \frac{\text{隣接頂点の重み}}{\text{ソースノードの隣接頂点の重みの合計}}$$

となる。

図 3 の頂点 A がソースノード、頂点 B, C, D がターゲットノード候補として考える。

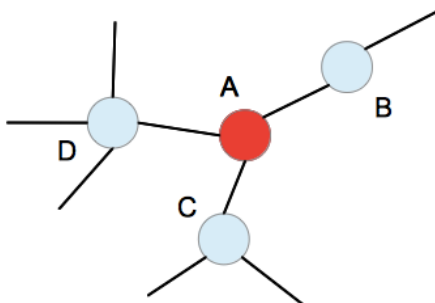


図 3: ネットワーク

一様分布に従って選択する方法は、ターゲットノードは均一な確率で選ばれる。重みは 1 とする。

図 3 において、各頂点が選ばれる確率は、

$$q_B = \frac{1}{3}, \quad q_C = \frac{1}{3}, \quad q_D = \frac{1}{3}$$

である。

頂点の度数に従って選択する方法は、次数が大きい頂点が選ばれる確率が高くなる。言い換えれば、次数が大きい頂点に優先的に情報を伝播する方法である。重みはソースノードの隣接頂点の度数の合計とする。

図 3 において、各頂点が選ばれる確率は、

$$q_B = \frac{2}{2+3+4} = \frac{2}{9}, \quad q_C = \frac{3}{9}, \quad q_D = \frac{4}{9}$$

である。

頂点の度数の逆数に従って選択する方法は、次数の小さい頂点が選ばれる確率が高くなる。次数が高いほどその逆数は小さくなるので、次数が小さい頂点に優先的に情報を伝播する方法である。重みはソースノードの隣接頂点の度数の逆数の合計とする。

図 3 において、各頂点が選ばれる確率は、

$$q_B = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}} = \frac{6}{13}, \quad q_C = \frac{4}{13}, \quad q_D = \frac{3}{13}$$

である。

6 実験概要

6.1 使用言語

シミュレーションを行うプログラムは、C++ を用いて作成した。

6.2 実験方法

修正版 BA モデルで、頂点数 1000 と 10000 のネットワークをそれぞれ 100 個用意した。このネットワークに対して、シミュレーションを行う。

情報を伝達する頂点をランダムに選択する際、rand 関数を使用する。異なる乱数生成の種を 50 個用意し、1 つのネットワークに対して、同じ乱数生成の種を用いて、一様分布に従って選択、頂点の度数に従って選択、頂点の度数の逆数に従って選択のそれぞれの伝播方法でシミュレーションを行った後、違う乱数生成の種を用いて、同様にシミュレーションを行う。つまり、100 個のネットワークに対して、それぞれ 50 回シミュレーションを行うことにした。

6.3 シミュレーションアルゴリズム

1. ネットワーク上からランダムにソースノードを 1 つ選択する.
2. 各ソースノードが情報を伝えるターゲットノードをそれぞれ 1 つ選択し, 情報を伝達する. 全ソースノードが同時に行う. この回数を 1 ステップとする.
3. ネットワーク上のすべての頂点がソースノードとなると, 終了.

7 実験結果

シミュレーションを行った結果を以下の図に表す.

図 4, 5, 6 は, 異なる 5 つのネットワークで同じ乱数の種を用いて行った伝播の様子である.

伝播の様子
頂点の次数の逆数に従って選択

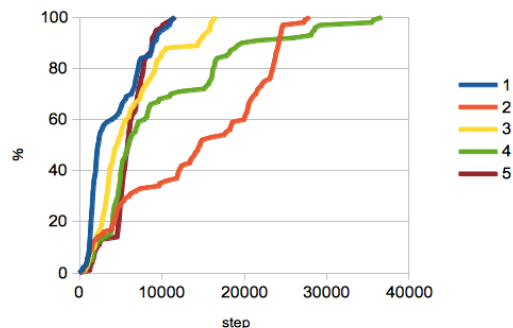


図 6 : 結果

3 つの伝播方法で行った結果を, 1 つの図にまとめると図 7 の様になる.

伝播の様子

一様分布に従って選択

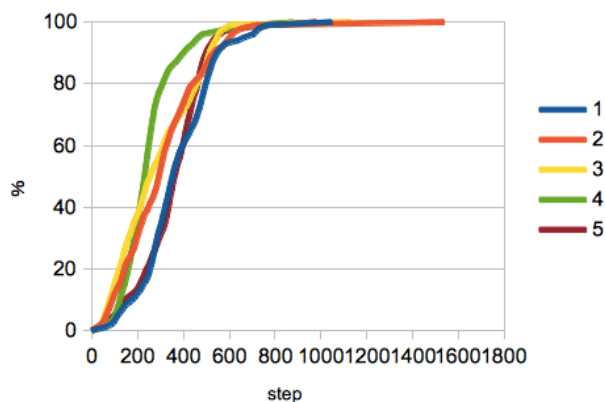


図 4 : 結果

伝播の様子

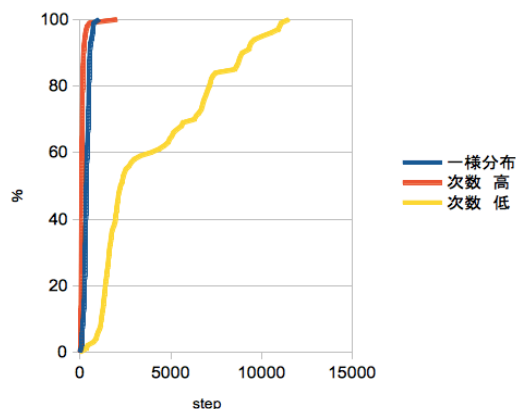


図 7 : 結果

図 8, 9, 10 は, 同じネットワークに対して, 5 つの乱数の種を用いて行った伝播の様子である.

伝播の様子

頂点の次数に従って選択

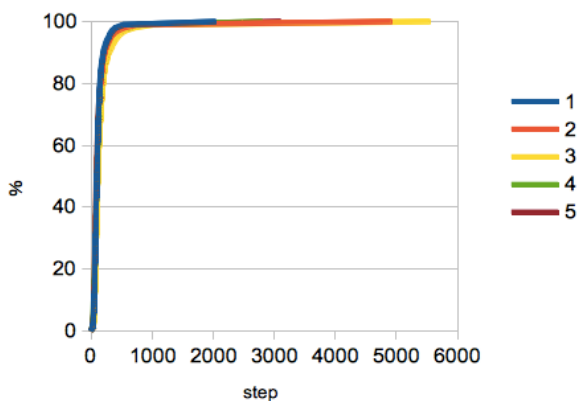


図 5 : 結果

伝播の様子

一様分布に従って選択

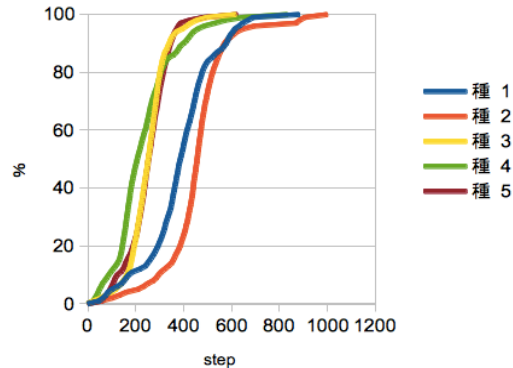


図 8 : 結果

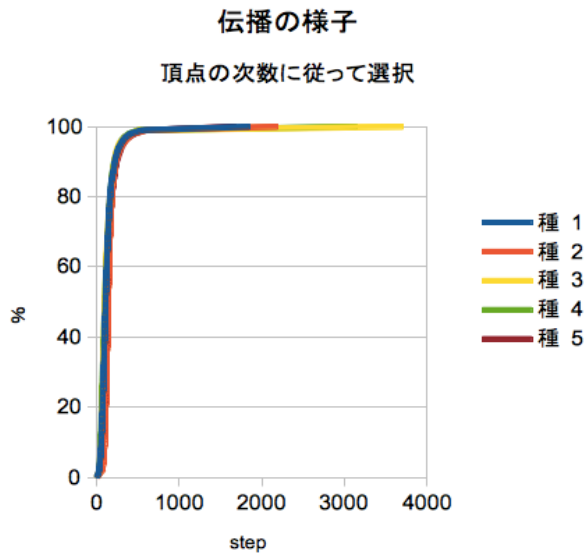


図 9：結果

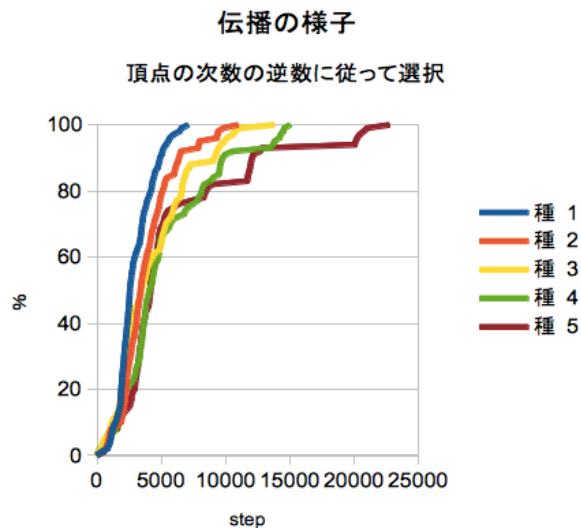


図 10：結果

図 4, 5, 6 により, 一様分布に従って選択する伝播方法が, 最も効率の良い伝播方法だという結果が得られた。頂点の次数に従って選択する伝播方法では, どのネットワークを見ても, 初めは効率良く情報が伝播されていたが, 全体に情報が伝わるまでには時間がかかることがわかる。頂点の次数の逆数に従って選択する方法は, どの伝播方法と比べても, 最も時間がかかり, 効率の悪い手法だった。

8 考察

今回のシミュレーションでは, 論文 ([2]) で主張されていた次数が少ない頂点を優先して情報を送る方法が, 最

も効率の良い伝播方法だという結果が得られなかった。

その原因を考えると, 今回のシミュレーションで用いたネットワークの頂点数が少ないということが考えられる。図 6 の結果を見ると, 突き刺さるように情報伝播が終わっていることがわかる。この結果は, より多くの頂点数のネットワークで情報伝播を行う際に, 違った結果が得られる可能性が考えられる。

また, 修正版 BA モデルでは, 論文の近似は成り立たない可能性がある。修正版 BA モデルは, 確かにスケールフリー性が実現されるネットワークのモデルだが, 生成されるネットワークは木構造になってしまう。現実世界のネットワークは木構造ではなく, クラスタが形成されることが多い。クラスタとは, 「自分の知人の知人が, 実は自分の直接の知人だった」という関係で, ネットワークに描くと三角形になることである。大抵の現実のネットワークにはクラスタが多くあり, 情報を伝播する経路も多いが, 今回のシミュレーションでは経路は限られてくる。確かに, 次数が少ない頂点を優先して情報を送る方法では, ネットワークの次数が少ない頂点に送りやすくなるが, その一方では, ハブに情報が伝わる確率が低いということである。ハブはネットワークにおいて大きく影響を与えるので, ハブに情報を伝わるまでのステップは, 余計な伝播であるといえる。

頂点を選択する際, 確率によって選択される。各選択方法の確率の検証では, 妥当な検証結果が得られていたが, 乱数については厳密に扱っていない部分がある。また, プログラム内の頂点を選択する処理を行う部分以外の実装が適切でなかった可能性も考えられる。

それらが適切だった場合, 論文の中で行われている近似は, 妥当でないということになる。どちらにしろ, 論文 ([2]) で行われた近似を解析する必要がある。

9 課題

頂点数を変更して, シミュレーションをすべきである。また, 本演習では修正版 BA モデルを採用したが, クラスタ性を考慮するなど, 現実世界に近いネットワークでシミュレーションを行う必要がある。さらに, 違うネットワークモデルでシミュレーションすることも大切である。例えば, Random Network と Small World Network がある。そして, 今回は伝播速度が限定されたモデルでシミュレーションを行ったが, 伝播規則を変更してシミュレーションを行うことを今後の目標としたい。

参考文献

- [1] 今野紀雄, 井手 勇介 複雑ネットワーク入門 (KS 理工学専門書) 2008/05/13.
- [2] Hiroshi Toyoizumi, Seiichi Tani, Naoto Miyoshi, Yoshio Okamoto,
Reverse preferential spread in complex networks.
- [3] Masahiro Obata 伝播速度限定モデルにおける Scale Free Network 上の情報拡散 Dynamics
- [4] Choy Yein Ann 伝播速度限定モデルにおける Scale Free ネットワーク上の情報拡散について
- [5] 加納幹雄 情報科学のためのグラフ理論 (朝倉書店)2001/02/20