

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

谷 研究室 蔡 延安

Choy Yein Ann

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。情報を保持する頂点は単位時間に 1 つの隣接頂点にのみ情報を伝えられる伝搬速度限定モデルで、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である場合には、次数の逆数に比例した確率分布に従い情報を伝達する頂点を選択する方法が、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ということが、ある仮定の元で証明されている。本研究では、その仮定が妥当か、また、証明された結果が妥当かを検証するため、計算機実験を行った。

1 はじめに

1.1 背景

現実世界では、私たちの身の回りに様々なネットワークの知見が存在している。ネットワーク解析では、グラフ理論などを用いて、様々な研究が行われている。それらのネットワークは多様であり、巨大で複雑な構造していることが多く、複雑ネットワークと呼ばれる [1]。例えば、人間関係、航空網、商売のマーケティングなどある。複雑ネットワークでは、一定の共通する性質が見出すことができる。それらの性質は、「スモールワールド性」、「スケールフリー性」、「クラスター性」などである。ネットワーク上の情報拡散についても、様々な研究が行われている。どのようにすれば情報を効率よくネットワーク上に拡散できるのかは一つの重なる課題である。例えば、コンピュータウイルスの拡散、遺伝子の情報拡散、噂の広がりなどである。

ネットワークにおいて、大きな次数を持つ頂点は多く情報を伝播できると考えられる。一方、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く、これらの頂点は情報を受け取りやすいと考えられる。無駄な伝播は情報拡散において、効率が悪くなる原因となる可能性がある [2]。情報源が 1 単位時間に一つの頂点にのみ情報を流せる場合には、情報源が次数が少ない頂点を優先して情報を流す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と主張している。

1.2 目的

論文 [2] では、ある仮定に基づき、次数の逆数に比例した確率分布に従い情報を伝達する頂点を選択する方法が、

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ことが証明されている。本研究では、その仮定が妥当か、また、証明された結果が妥当かを検証するため、Scale Free ネットワークを生成することで計算機実験を行う。また、比較のため、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、次数の昇順に選択、次数に降順に選択という 3 つの方法に対しても計算機実験を行う。なお、計算機実験プログラムは C++ を用いて実装した。

1.3 構成

本稿は以下のようになっている。まず、第 2 節では、Scale Free ネットワークについて説明する。そして、第 3 節では、情報拡散の過程に用いる伝播速度限定モデルについて紹介する。第 4 節では、実験方法について述べる。第 5 節では、行ったシミュレーションの結果を示し、考査する。本研究のまとめは第 6 節に述べる。

2 Scale Free ネットワーク

Scale Free ネットワークとは、複雑ネットワークの中の一つネットワークである。実在する Scale Free ネットワークの例は、インターネット、映画俳優の共演ネットワークとタンパク質の反応ネットワークが挙げられる。Scale Free ネットワークの生成モデルはいくつかある。例えば、閾値モデルと BA モデルなどである。

2.1 グラフ理論の概念

ネットワークを考える上で、グラフ理論の知識が必要である。その必要な知識を簡単に説明する。ここでは、グラフ G を例にとって考える。グラフ G は頂点と辺からなる。鉄

道網をイメージした方が分かりやすいの可能性がある。

- 頂点：駅
- 辺：異なる駅を結ぶ路線
- 次数：頂点から出ている路線の個数

図 1 において赤い点は頂点を表し、赤い頂点同士を結ぶ直線は辺を表している。この場合は、頂点 A の次数が 3 になる。

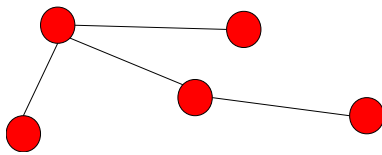


図 1: グラフ G

2.2 スケールフリー性

スケールフリー性とは、少数の頂点が多く of 辺を持ち、多数の頂点がわずかな辺を持つ性質である。例えば、鉄道網の関係から見ると、一部の駅は非常に多く路線を持っているが、多く駅を持つ路線は少ないである。また、次数分布はベキ分布に従っていることになる。

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

2.3 代表的な生成モデル — BA モデル

Scale Free ネットワークを生成する最も代表的なモデルは BA モデルである。BA モデルは、1999 年に Barabási と Albert によって提案され、2 人の頭文字をとって BA モデルと呼ばれる。BA モデルにおいては、以下の 2 つ重要な概念がある。

- 成長：頂点と辺がネットワークに次々と追加されていくことである。
- 優先的選択：大きい次数を持つ頂点の次数はさらに大きくなる傾向があることである。

ここでは、BA モデルで Scale Free ネットワークを生成するアルゴリズムを [1] に従って、説明する。

1. $n > 1$ 個の頂点を用いて完全グラフを作る。
2. 新しいノードを 1 個を追加する。そのノードから、既に存在している n 個のノードに対してリンクを張る。この時リンクが張られる確率は、それぞれのノードのその時点の次数 k に比例することになる。

すなわち、

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

の確率である。

3. ステップ 2 をノードの決められた数になるまで繰り返す。

3 情報拡散の過程 — 伝播速度限定モデル

情報拡散では、様々な分野で様々な研究を行っている。論文 [2] の提案した伝播速度限定モデルもその一つである。

3.1 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。

3.2 伝播規則

情報源が使える情報は以下の通りである：

1. 情報源は近傍頂点の次数のみ与えられる。
2. 情報源は以下を記憶することができる。
 - (a) 伝播させたい情報を送った頂点、
 - (b) 伝播させたい情報を送ってもらった頂点、

情報伝播規則について、情報源は、1 単位時間に、情報源が使える情報のみを用いて、隣接する頂点の 1 つに情報を伝える。また、伝播させたい情報をもった頂点は情報源になる。

情報源の伝播させたい情報を送られる近傍頂点、すなわち、ターゲット頂点 b を選ばれる確率は、

$$q_b = \frac{\text{ターゲット頂点 } b \text{ の重み}}{\text{情報源のターゲット頂点の重みの合計}} \quad (1)$$

である。式 (1) に重みの代わりより、Inverse Control と No Control の 2 つの比較方法があがられる。

1. No Control: 重みは 1 する。
2. Inverse Control: 重みはその頂点の次数の逆数とする。

No Control では、重みを 1 になると、ターゲット頂点は均一な確率で選ばれる。すなわち、ランダムで選ばれると考えられる。一方、Inverse Control は次数の逆数を使うので、次数が小さいターゲット頂点を選ばれる確率が高くなる。例えば、図 2 により、黒い頂点は情報源である。

ターゲット頂点は A,B,C であり,それぞれの次数は 4,3,2 とする. 頂点 C を選ばれる確率は $(1/2)/(1/4 + 1/3 + 1/2)=6/13$, 頂点 A,B の $3/13$ と $4/13$ より選ばれる確率が高くなる.

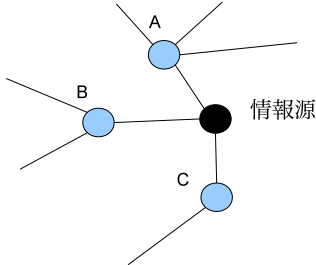


図 2: ネットワーク

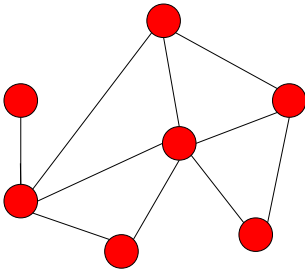
3.3 仮説について

論文 [2] における頂点を選ばれる確率の近似方法について述べる.

まず, n サイズのネットワークがある. n_k は次数 k の頂点の個数とする. 次数 k の分布は

$$P\{K = k\} = \frac{n_k}{n} \quad (2)$$

である.

図 3: n サイズのネットワーク

そのネットワークの頂点同士の隣接を切断する. 頂点同士は自分の次数の情報のみ持つことになる. すなわち, 隣接を切断した後の次数分布 $P_u\{K=k\}=P\{K=k\}$ となる.

また, 辺 e を選んだ場合の K_e は隣接先の頂点の次数とする. 次数 i の頂点から辺 e を選んだ場合, その隣接先の頂点の次数が k となる確率は以下の通りである.

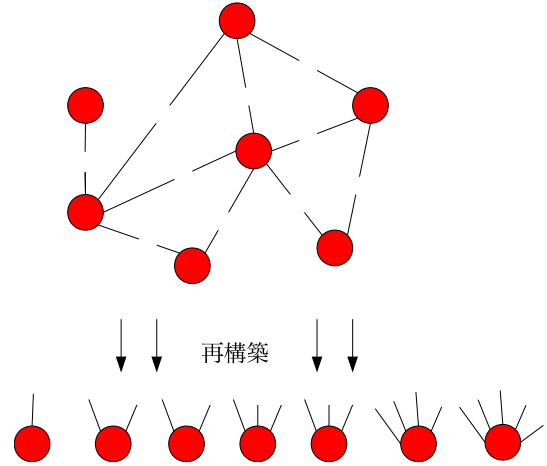
$$P_u\{K_e = k \mid K = i\} = \begin{cases} \frac{kn_k}{(\sum_{j=1}^n jn_j) - 1} & (k \neq i) \\ \frac{kn_k - 1}{(\sum_{j=1}^n jn_j) - 1} & (k = i) \end{cases} \quad (3)$$

簡単に言うと, ある頂点の次数が i の時, 隣接先の頂点の次数が k となる確率は式 (3) で与えられる. 一方, 全体か

ら見ると隣接先の頂点の次数が k の時, 選ばれる確率は式 (3) から,

$$\begin{aligned} P_u\{K_e = k\} &= \frac{n - n_k}{n} \frac{kn_k}{(\sum_{j=1}^n jn_j) - 1} + \frac{n_k}{n} \frac{kn_k - 1}{(\sum_{j=1}^n jn_j) - 1} \\ &= \frac{(k - 1/n)n_k}{nE[K] - 1} \approx \frac{kP\{K = k\}}{E[K]} \end{aligned} \quad (4)$$

計算できる.

図 4: n サイズのネットワークの切断と再構築

ここには, 2つ以上のネットワークを分裂, 異なる2つ頂点をが多重辺を持つ, 頂点が自己ループをなす問題点の可能性はあるが, ネットワークのサイズが大きければ大きいほど, そのような確率が小さくなることになるので, 無視できることを考えられる.

w_k は辺が次数 k と接続する重みとする. 情報が伝播する時, 各情報源がターゲット頂点を選ぶ確率は

$$\frac{w_k}{\text{次数 } k \text{ である頂点の全近傍頂点の重みの合計}} \quad (5)$$

である.

ターゲット頂点の次数が k を選ばれる確率は, (3) を用いて,

$$q_k = \frac{w_k P_u\{K_e = k\}}{\sum_j w_j P_u\{K_e = j\}} = \frac{kw_k n_k}{\sum_j jw_j n_j} \quad (6)$$

が計算できる.

No Control において,

$$q_k = P_u\{K = k\}$$

が計算できし, Inverse Control のにおいて,

$$q_k = P_u\{K_e = k\}$$

が計算できる.

4 実験方法

頂点数 100 万のグラフ G_i を BA モデルで 100 個を生成する. 各 $G_i (i \in \{1, 2, \dots, 100\})$ にたいして前章を述べた伝播規則を利用して, 次のようなシミュレーションを 100 回ずつ実施する.

- s_1 : No Control でターゲット頂点を選び
- s_2 : Inverse Control でターゲット頂点を選ぶ
- s_3 : 単にターゲット頂点の次数を昇順から選び
- s_4 : 単にターゲット頂点の次数を降順から選び

ただし, どのグラフ $G_i (i \in \{1, 2, \dots, 100\})$ のどのシミュレーション $s_j (j \in \{1, 2, 3, 4\})$ における, $k (k \in \{1, 2, \dots, 100\})$ のシミュレーションにおいて, 乱数の seed は k を用いる.

5 シミュレーションの結果

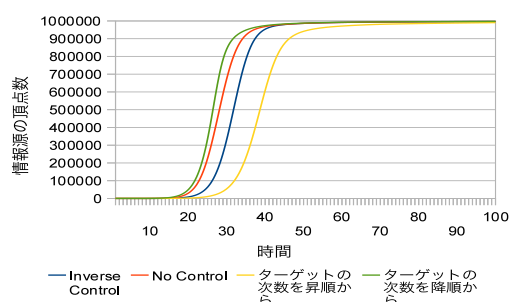


図 5: Scale Free ネットワーク上の情報拡散

参考文献

- [1] 今野紀雄, 井手勇介 著 「複雑ネットワーク入門」 (2008)
- [2] Hiroshi Toyoizumi, Seiichi Tani : Optimal Spread over Finite-Size Network. (2008)
<http://www.neeton.ac.uk/programmes/SCS/Toyoizumi,%20H.pfd>

図 2 により, No Control は [2] の理論とは違って, Inverse Control より情報を効率よく拡散する. また, 次数の昇順に選べば, 一番効率悪く拡散する. 一番効率いいのは単にターゲット頂点の次数の降順に選ぶ方法である.

約 15 単位時間まで情報拡散の効率はほぼ同様である. それ以後は境目が現れた. 次数の昇順に選ぶ方法以外の方法は, 約 40 単位時間にネットワーク全体の 95% に情報を拡散した. 15 から 40 単位時間の内に, 次数の降順に選ぶのは情報を効率的拡散になる. ネットワーク全体を約 80% に情報を拡散するとに各伝播方法の時間差が大きくなる.

6 まとめ

本研究では, Scale Free ネットワークに対して, 論文 [2] の伝播規則を用いて, Inverse Control, No Control, ターゲット頂点を昇順と降順から選ぶの 4 種類の選択方法で伝播シミュレーションを行った. 研究目的は情報拡散の効率を検証することである. 本研究の計算機実験では, Inverse Control は No Control より効率的であるという結果は得られなかった. 実装したプログラム中の確率を計算する部分に対する検証が終わってない. そのため, 評価はまだ定まっていない.

今後の課題として, まず確率の計算の検証する必要がある. また, 違うネットワーク上にシミュレーションすることをあげたい. 例えば, Random ネットワークと Small World ネットワークである. そして, 伝播規則の変更に興味深いであろう. 例えば, 情報源が記録することができないことと確率に次数が大きい近傍頂点を選ぶことである.