

スケール・フリー・ネットワーク上での情報伝搬の効率

松久保 潤[†] 林 幸雄[†]

[†]北陸先端科学技術大学院大学 知識科学研究科

1. はじめに

俳優間の共演関係, 論文の引用関係, 線虫の神経回路網, 電力網, 知人関係, ルータ, 電子メールの送受信, 及び Web ページ間のリンク関係などの現実に存在する多くのネットワークは, スケール・フリーと呼ばれる構造的特徴をもっている. すなわちスケール・フリー・ネットワークでは, 辺数(次数)に対する頂点数の分布がべき乗則に従う. この特徴は正則ネットワークやランダム・ネットワーク[1]と異なる.

一方, これらのネットワークはランダムな攻撃への耐故障性は高いが, Hub(高い次数の頂点)への攻撃に弱いことが近年明らかにされた[2], [3]. また, 情報伝搬の観点からも, Hub はスケール・フリー・ネットワーク上で, 多くの頂点に情報を中継する役割をもっているため, Hub が機能なくなるとネットワークが細分され, ネットワーク自体が機能しなくなる.

そのため, 上述のようなネットワーク上で効率的に Hub を発見することが重要な課題となる. 本報告では, スケール・フリー・ネットワーク上での情報伝搬のシミュレーションを行い, Hub を比較的早く発見方法について議論する.

2. 情報伝搬シミュレーション

まず, スケール・フリー・ネットワークの生成法を述べる. べき乗則に従うネットワークを生成するモデルに必要な性質は,

- 頂点と辺を追加しながらネットワークが成長すること
- 辺を多く持つ頂点ほど辺が繋がりがやすいことの 2 つである. Kumar によって提案された(,)モデル[4]は, 出次数と入次数のそれぞれに任意のべき係数をもつ有向グラフを確率的に生成できる. Web に関する実測値[4]に従って, 出次数, 及び入次数のべき係数をそれぞれ $\alpha_{in} = -2.1$, $\alpha_{out} = -2.4$ とした. また, 平均の辺数を 8 本として, 10000, 50000, 100000 の頂点数のネットワークを 5 個ずつ生成した. それぞれのネットワーク上で, ランダムな 10 個の開始点から幅優先探索によるクローラをシミュレートし, 訪問した頂点に対して, Hub の累積獲得数, 新たに得られた頂点, 及び

重複した頂点の累積数の平均値をそれぞれ求めた.

図 1 は, 出次数の降順にネットワークの総頂点数の 0.5% を Hub として, 訪問した頂点数に対する Hub の累積獲得数を示す. 横軸はネットワークの総頂点数に対する訪問した頂点数の割合(%), 縦軸は Hub の総数に対する Hub の累積獲得数の割合(%)である. 図 1 において, 訪問した頂点数が 20% となる近傍で, Hub の累積獲得数が 80% を超えている. また, Hub は伝搬を開始する頂点に関係なく, 発見されやすい傾向をもつことが分かった.

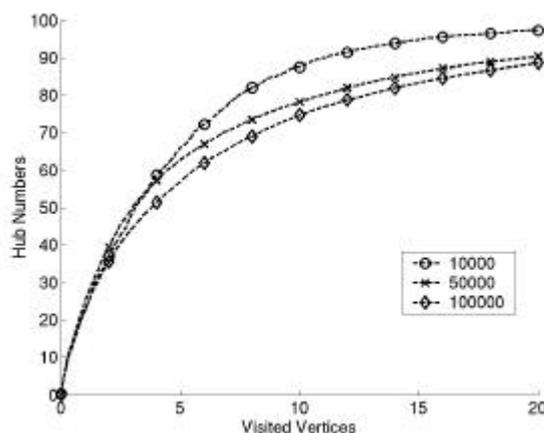


Fig. 1: Hub の累積獲得数の割合

図 2 は, 訪問した頂点数に対する新たに得られた頂点数の累積数である. 横軸, 及び縦軸はそれぞれ, ネットワークの総頂点数に対する訪問した頂点数の割合(%), および新たに得られた頂点の累積獲得数(%)である. 図 2 では, ネットワークの総頂点数に関係なく, 訪問した頂点数が 20% となったとき, 新たに得られた頂点の累積数は約 80% となっている. これ以降, 新たに得られた頂点の累積数は緩やかに増加している. このことから, Hub の発見後, 残りの少数の頂点を発見するために, 高いコストが掛かることが分かる.

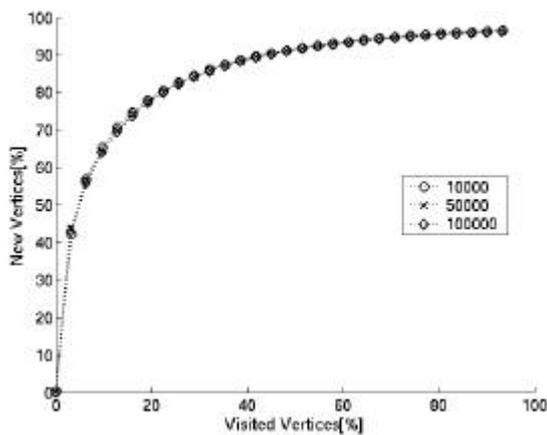


Fig.2訪問した頂点数に対する新たに得られた頂点の累積数

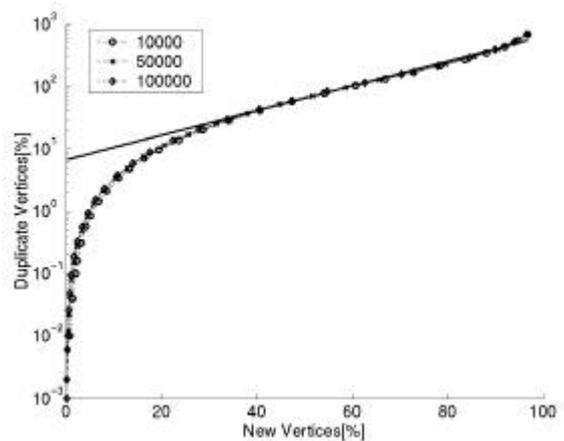


Fig.4新たに得られた頂点の累積数に対する重複した頂点の累積数 (縦軸のみ対数プロット)

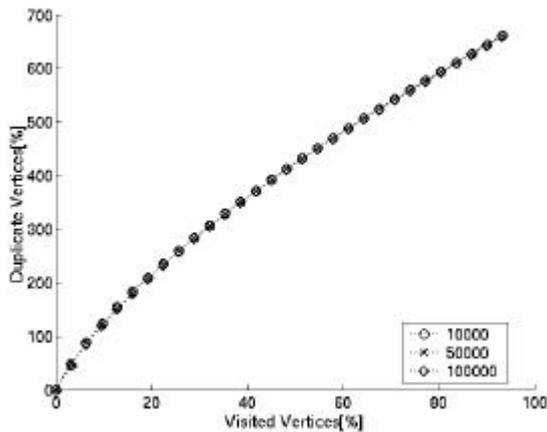


Fig.3訪問した頂点数に対する重複した頂点の累積数

図3は、訪問した頂点数に対する重複した頂点数の累積数を示す。横軸、及び縦軸はそれぞれ、ネットワークの総頂点数に対する訪問した頂点数の割合(%), 及び重複した頂点の累積獲得数(%)である。図3から、スケール・フリー・ネットワーク上の多くの頂点は、複数の頂点から接続されているため、重複して発見される可能性が高いことが分かる。逆に、このことはネットワークの接続性が頂点のランダムな故障に対して強いことを示している。すなわち、任意の頂点間に多くの経路が存在し得る。これは正則ネットワークと異なる性質である。

図4は、新たな頂点数に対する重複した頂点数の累積数を示している。図4は伝搬が進むとその累積数が指数関数的に増加することを示している。また、訪問した頂点数がネットワーク全体の頂点数の3%となった

とき、新たに得られた頂点数と重複した頂点数の累積数が等しくなった。このとき、それぞれの累積数はネットワークの総頂点数の約40%であった。

上述の性質は、探索を開始する頂点、ネットワークの総頂点数に依存しなかった。

3. おわりに

本報告では、人工的に生成したスケール・フリー・ネットワーク上での幅優先探索によるクローラを考え、情報伝搬の効率を調べた。シミュレーションの結果、スケール・フリー・ネットワーク上では、伝搬を開始する頂点に関係なく、多くの頂点にリンクをもつ Hub を比較的早く発見できることが分かった。この性質はスケール・フリー・ネットワークの構造に強く依存するものと考えられる。

参考文献

- [1] Watts, D. J. and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature 393, pp.440-42, 1998.
- [2] Kim B.J., "Path finding strategies in scale-free networks", Phys. Rev. E, Vol.65, 027103, 2001.
- [3] 辻野,「生命のごとく成長するウェブの向かう先は」, <http://www5.ocn.ne.jp/~report/news/networktopology.htm>.
- [4] Kumar R, et al, "Extracting large-scale knowledge base from the web," IEEE Int. Conf. on VLDB, Edinburgh, Scotland, 1999.