定常ネットワークにおけるベキ分布

和田　信也
（株）ソニー・コンピュータエンタテインメント
東京都港区南青山 2 - 6 - 21
E-mail: swada@rd.scei.sony.co.jp

ノードの成長性とエッジの選択的接続からノードの持つエッジ数の分布がベキ分布になることはよく知られている。[1, 2, 3] これに対して、我々はノード及びそのノードが持つエッジを取り除き、それらを再び選択的接続によってネットワークに接続することで、エッジ数の時間平均が一定になるようなネットワークモデルを考えた。この総エッジ数の時間平均が一定のモデルで、ベキ分布が出現する可能性について考察した。さらに、我々のモデルを計算クラスタを構成する計算ノードの管理方法に当てはめられないか考察した。

Power-law distribution of connectivities from non-evolutional networks

SHINYA WADA
Sony Computer Entertainment Inc.
2 - 6 - 21, Minamiaoyama Minato-ku, Tokyo
E-mail: swada@rd.scei.sony.co.jp

In the previous work [1, 2, 3] , Barabasi et. al. have shown that following two key features are essential for the property of power-law distributions of connectivities. First, networks continuously grown, and second, new vertices connect preferentially to highly connected vertices. We introduce a new network model which the time average of the number of edges does not change. To keep the number of edges constant, nodes and their edges are subtracted for some time interval from the network. We show that these dis-connection and prefential re-connection feature make our model possible to have a power-low distribution of the edge number with cut-off. Finally, we consider this model to apply the computer cluster model.

1　はじめに

Barabasi らは World Wide Web のハイパーリンクの構造を調べ、興味深い結果を得た。彼らは、各ノードの持つリンク数の分布を詳細に調べ、非常に僅かな数の多数のエッジを持つノードから少数のエッジを持つ多くのノードまで、エッジ数の分布が非常に広い範囲にわたっていることを発見した。その分布 $P(k)$ は、

$$P(k) \sim k^{-\gamma} \tag{1}$$

(ただし、k はノードの持つエッジ数、γ は定数) と表されるベキ分布になる。さらに彼らは、このような分布を持つネットワーク構成するモデルを考えた。そのモデルを使い、ノードが新しく追加されるような成長するネットワークで、新しく付け加わったノードはより多くのエッジを持つノード対して接続する確率が高いという２つの条件があればネットワークにベキ分布が生じることを生じることを示した。そして、彼らはこのネットワークを scale-free ネットワークと名づけた。

我々は、Barabasi らのネットワークモデルを変形してノードは接続されるだけでなく、ある一定の確率で取り除かれ、そのノードと接続されていた全てのエッジも取り除かれるモデルを考えた。その際に、ネットワークを構成する総エッジ数の時間平均が一定であり、総エッジ数に上限があるような準定常なモデルでも、Barabasi らと同様の選択的接続を導入することでネットワークにベキ分布が出現することを示した。このようなモデルを考えた理由は、ノードをクラスタ、エッジをクラスタを構成する計算ノードに読み替えると、総計算ノード数が一定の下で、計算ノード数がベキ分布するクラスタを構成できるのではないかと考えたからである。

2　ネットワークモデルと計算機クラスタ

Barabasi らは、ネットワークとはノードと、ノードを接続するエッジから構成されるとした。そして、彼らの scale-free ネットワークのモデルではエッジを持たない m_0 個のノードからスタートし、時間ステップごとにノードが付け加わっていき、加わったノードは、既にネットワークに存在するノードのうちの m

個のノードとエッジにより接続される。(成長性) 新しく加わったノードがエッジによって接続するノードを選ぶ基準として、ノードの持つエッジの数に比例する接続確率を導入した。(選択的接続) ノード i と接続する確率は、このノードが持つリンクの数を k_i として、

$$\Pi_i = \Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{2}$$

とした。このとき、ネットワークには式 (1) のようなベキ分布が生じることが示されており、ネットワークの成長性とエッジの選択的接続がベキ分布を生じさせる要因であるとされている。

我々は、このBarabasiらのモデルを利用して、インターネットに点在する計算機を共有して大きな計算力として利用するGrid環境において、プロセスを効率よく処理する方法についての考察を試みた。従来の分散処理では、LANで接続されたほぼ等しい性能の計算機に対して、プロセスを実行するクライアントから一方的に計算を行うサーバー群へのプロセス割り当て要求のみが発生していた。この場合は、ＣＰＵ資源やメモリ資源などの計算資源の管理は比較的容易で、さまざまな大きさのプロセスやサブプロセスへの資源の割り当ては比較的効率よく行いやすいと考えられる。しかしながら、インターネット上に点在するさまざまな計算機でGridを構成する場合には、Grid内でインターネットを介して多数の計算機によるプロセス割り当てが常に行われる。このようなオープンなネットワーク上では、全ての計算資源の把握は困難である。従って、各計算ノードが自律的に集合し、様々なサイズのプロセスに対応できるように計算資源がクラスタ化し計算の終了したノードは速やかに別のプロセスの実行のために割り当てられなければならない。このような、Grid上でプロセスを効率よく処理するための方法は、様々考えられているが [5]、我々は計算ノードが簡単なルールに基づいて互いに接続し、クラスタを構成し計算プロセスの終了後さらに別のクラスタを構成するというモデルを考察したいと考えた。

そこで、まずBarabasiらのモデルのノード、エッジと計算ノード及びクラスタの対応を考えた。図1は、Barabasiらのネットワークモデルを上段に我々のクラスタモデルの下段に、左から順にタイムステップの小さいものから並べたものである。Barabasiらのモデルのノードは大きな円で、エッジはノードを結ぶラインで表されている。それに対して、クラスタモデルでは、大きな円は、計算クラスタ、小さな円が計算ノードを表している。(下段のクラスタモデルでは、図中には計算ノードを結ぶ点線があるが、これはBarabasiらのモデルとの類似性を分かりやすくするためのもので実際には存在せず、小円で表される計算ノードのみが存在する) つまり、Barabasiらのモデルで言う、ノードはクラスタモデルでは、計算クラスタと解釈し、各ノードから出るエッジはノード間を結ぶリンクではなく、計算クラスタに含まれる

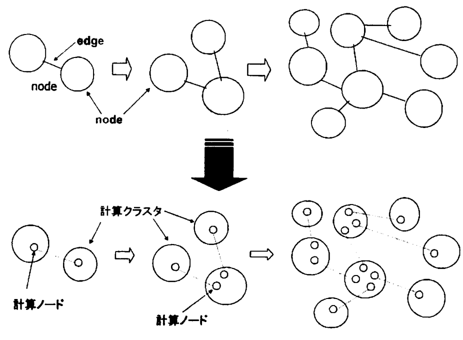

図 1: 計算機クラスタモデル

計算ノードと解釈するのである。このように対応付けると、ノードのもつエッジ数は、計算クラスタの持つ計算ノードの数に完全に一致し、Barabasiらのモデルをそのまま使用する事ができる。これにより、成長性は新しいクラスタの生成 (新しい計算ノードが付け加わる)、選択的接続はどのクラスタに新たに計算ノードが付け加わるかの確率がクラスタのサイズに依存すると解釈することができる。

このような対応を行ったときに、問題となるのは計算機クラスタの場合には、計算ノードの数は有限で一般的には、計算機クラスタに参加する計算機数の最大数はほぼ一定である。これは、ネットワークモデルではエッジ数は一定であることに相当する。元のスケールフリーネットワークをつくるモデルでは成長性により、総エッジ数も増加するのみであったが、総エッジ数を一定に保ちながら、成長性を考える必要がある。そこで、我々のモデルで計算ノード数を一定に保ったまま、成長性と選択的接続の性質を持たせるために、クラスタ化されたノードは計算が終了するといったん全て分離され、計算機クラスタから離脱し、その後再び計算クラスタを形成することを考えた。これをBarabasiらのモデルの立場で言うと、時間ステップとともにあるノードが取り除かれ、そのノードの持つエッジが一斉に外れる。同時に、時間ステップごとにノードが付け加えられ、エッジを接続するのだが、接続されるエッジの数は取り除かれたエッジの数を超えないようになっており、取り除かれるエッジ数と付け加えられるエッジ数の十分に長い時間での平均は等しいようにするのである。

次章で、この変形したBarabasiらのモデルについて詳細に検討する。

3　エッジ数一定のモデル

エッジ数が一定のモデルで、エッジ数の分布をBarabasiらの平均場近似モデル [1] に従って解析的に検討してみる。初期状態として、ノード数 m_0、各ノードが k_0 本のエッジを等しく持っているとする。

すると、全エッジ数は、$m_0 k_0$ 本であり、これを最大エッジ数とする。この初期状態からスタートし、ある確率で、ノードが取り除かれてそのノードの持つエッジも取り除かれる。そして、一定時間ごとに新たに m 本のエッジを持つノードが付け加えられ、m 本のエッジは選択的接続、式 (2) に従ってノード i に再接続される。十分長い時間後、k 本のエッジを持つノード数の分布 $P(k)$ は一定になり、取り除かれるノードの持つエッジ数に対して新たに付け加えられるエッジ数の増加頻度が高い場合は、エッジ数が増加していくが、総エッジ数はある最大のエッジ数を超えないものとする。(あるタイムステップでネットワークに参加する総エッジ数が最大エッジ数に達している場合には、新たにノードは付け加わらないとする。) このとき、時刻 τ_i で付け加えられたノード i を考える。このノードが次に取り除かれるまでの時間間隔での、エッジ数の変化について考える。このノードの持つエッジ数 k_i とすると Barabasi らのモデルと同様に、

$$\frac{\partial k_i}{\partial \tau} = m\Pi(k_i) \tag{3}$$

と書ける。もし、取り除かれるノードがあり、それがノード i と接続されている場合は k_i はその分だけ減少するはずであるが、今は短い間隔で k_i の時間依存性を考えているのでこの減少分はノード数が十分多い場合は、一時的なもので時間依存性は無視できると考えられる。また、クラスタのモデルの場合には、この取り除かれるノードと接続しているエッジは取り除かないというモデルも考えられ、その場合にも、式 (3) は、同様に成り立つ。時刻 τ でのエッジ数のトータルは、時刻 τ_i でのエッジ総数を N_i として、

$$\sum k \;=\; N_i + 2m\tau - \text{(dis-connected edge)} \tag{4}$$

つまり、総エッジ数は $m_0 k_0$ で一定であるが、ある局所的な時間では、ネットワークに参加しているエッジ数は $2m\tau$ の時間依存性を持って増加しているようにみえる。つまり、ネットワークの成長性と同様の現象が見られる局所時間について考えるのである。この局所時間の間では、エッジ数の一定性が破れて、

$$\sum k = N + 2m\tau \tag{5}$$

となっていると考えられる。ただし、N_i + (dis-connected edge) を時間依存性のない定数として扱い N と表した。従って、式 (2) は、

$$\frac{\partial k_i}{\partial \tau} = \frac{m k_i}{N + 2m\tau} \tag{6}$$

となり、$k_i(t_i) = m$ なので、

$$k_i(\tau) = m\left(\frac{N + 2m\tau}{N + 2m\tau_i}\right) \tag{7}$$

が得られる。時刻、τ_i での全ノード数を M_i とすると、Barabasi らと同様の議論により、

$$P(k) = \frac{N_i m}{k^3(M_i + \tau)} + \frac{2m^2\tau}{k^3(M_i + \tau)} \tag{8}$$

となり、τ の大きいところでは、Barabasi らの結果と同様に、

$$P(k) \sim \frac{1}{k^3} \tag{9}$$

と振舞う。

上記のようなシナリオが正しいためには、ノードが取り除かれる頻度を T_d、取り除かれるノードの持つエッジ数を N_d とし、ノードが付け加わる頻度を T_a とすると、

$$T_d \times N_d < T_a \times m \tag{10}$$

が成り立たないと、エッジの数が減っていき定常的にならない。

この解析から分かるように、十分長い時間後の分布は初期状態のエッジ数の分布に依存しないことが分かる。また、局所的な時間で選択性、と成長性が得られるようにエッジを取り除くことが重要であることが分かる。

次章では、いくつかの条件の下で数値シミュレーションを行い、上の結果を確かめる。

4 数値計算

我々は、上記の解析を確認するために計算機シミュレーションを行った。モデルとして、最大エッジ数 10^4 で、毎タイムステップごとにノードが一つづつ増えていき、(ただし、ネットワークにある総エッジ数が既に最大エッジ数の時には、そのタイムステップではノードは追加されない。) $m = 1$ であるモデルを考えた。また、ノードが取り除かれる確率は、2 タイムステップに 1 回の割合で、全てのノードの中からランダムに選んだ一つのノードを取り除くこととする。また、初期状態では、円状に並んだノードが互いに隣のノードと結びついている状態とする。つまり、初期ノード数は 5000、各ノードは全て 2 つのエッジを所有し総エッジ数 = 最大エッジ数 = 10^4 の状態である。図 2 は、ネットワークに参加するノード数の時間変化である。タイムステップが小さい時には、取り除かれるほとんどのノードは、エッジを 2 つ持っているのに対して、接続されるノードはエッジを一つしか持たないので、エッジは余っていることになり、ノード数は増加していく。しかし、十分長い時間ステップ後には、エッジ数の分布が変化したことにより、ノード数がほぼ一定のネットワークになっていることが分かり、局所的に成長性が実現していると考えられる。また、図 2 の中の図は、タイムステップ 20000 から 20200 の 200 ステップを拡大したものであるが、短い時間間隔では、ノード数がダイナミックに増減していることが分かる。図 3 は、総エッジ数の時間変化である。総エッジ数も最大エッジ数である 10000 を最大にして、長い時間平均ではほぼ一定であるが、局所的には大きく上下していることがわかり、式 (4) の時間依存性が局所的に見られる。最後に図 4 は、時間ステップ 10^5 回 (三

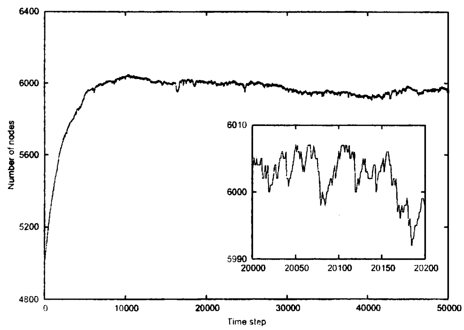

図 2: 総ノード数の時間変化

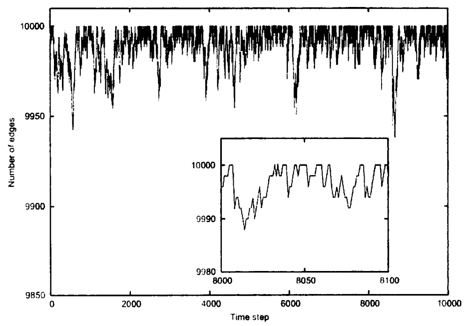

図 3: 総エッジ数の時間変化

角のプロット）及び 10^6 回（丸印のプロット）後の異なるエッジ数を持つノードの分布を両対数グラフでプロットしたものである。点線は、k^{-3} のラインであるが、十分長い時間後にはほぼベキ分布となり、式 (9) によく一致していることがわかる。

5 まとめ

ノード、エッジからなるネットワークモデルで局所的な時間間隔でノードの分離及び接続が行われノード間を結ぶエッジが選択的接続をする場合には、ノード数やエッジ数がほぼ一定なネットワークでも十分長い時間の後にベキ分布が出現することがわかった。このことから、単純な分離・再接続のルールにより様々なノード数分布をもつクラスタをつくることができる可能性があることがわかる。今回は、全ての場合について検討することができなかったが、様々な初期条件、分離・接続確率などのパラメータの調整により、投入されるプロセス群に最適化できる分布を持つクラスタが作成できるならば、Grid を構成する各ノードをより高い効率で使用することが可能であるかもしれない。しかしながら、これを実際の Grid コンピューティングに応用するためには、実際にこのクラスタでどのようなプロセス群が効率よく処理でき、計算機資源の利用効率を高められるかを

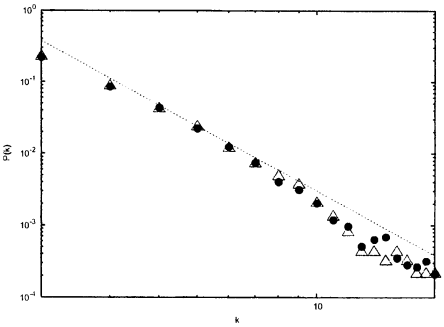

図 4: 異なるエッジ数を持つノードの分布

具体的に検討する必要がある。

参考文献

[1] Albert-László Barabási, Réka Albert and Hawoong Jeong, "Mean-field theory for scale-free random networks", Physica A 272 (1999) 173-187

[2] Albert-László Barabási, Réka Albert and Hawoong Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web", Physica A 281 (2000) 69-77

[3] Réka Albert, Albert-László Barabási, "Statistical Mechanics of Complex Networks", Reviews of Modern Physics 74 (2002) 47

[4] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks", Proc. National Academy of Sciences 97 (2000) 11149-11152

[5] 山本　寛, 川原　憲治, 滝根　哲哉, 尾家 祐二" グリッドコンピューティングにおける分割プロセス割り当て方法の検討とその基礎性能解析", 電子情報通信学会、信学技法 IN2002-8(2002-05) 43-48