

共同出願特許ネットワークにおける 高クラスタリング係数再現モデルの検証

井 上 寛 康

Investigation of a Model Generating Clustering Coefficient as High as that of Joint Patent Application Networks

INOUE Hiroyasu

目 次

1. はじめに
2. 社会科学と経済物理学の違いによる研究の位置づけ
3. 共同出願特許ネットワーク
4. ネットワークの分析
5. 従来の再現モデルにおける問題点の検証
6. 再現モデル改善点の検証
7. まとめ

Abstract

This paper investigates a model that generates networks with clustering coefficient as high as those in networks based on joint patent applications. A model has been proposed that generates a network structure that is similar to those of joint patent applications. However, the model is not sufficient to achieve a high clustering coefficient. The clustering coefficient is an index of network structures that shows the existence of groups that are densely linked. Our discoveries are summarized here as follows. (1) To improve the existing model, replacements of links are effective based on the exponential probability in degrees. (2) In the improvement, the degree distribution of the generated network has a different figure from the original network. In the above investigation, an original joint application network was necessary, and it was created from about five million Japanese patents issued between 1993 and 2002.

キーワード：特許，ネットワーク，クラスタリング係数，共同出願

Keywords：Patent, Network, Clustering Coefficient, Joint Application

1. はじめに

イノベーションが集中的に起きる地域は産業クラスターと呼ばれ、経済成長に重要であるとして政策面からもサポートされている。産業クラスターを構成するのは、道路を初めとする社会インフラ、組織¹⁾そしてそれら組織間のつながり（ネットワーク²⁾）とされている。この組織間のネットワークには、取引、株所有、役員派遣などのさまざまなネットワークがある [1] [2] が、本研究では、協調的にイノベーションを生み出す上で重要な組織間の共同研究開発に焦点をあてる。

組織間の共同研究開発から構成されるネットワークがどのような構造を持ち、どのように成長するのは、地域経済およびそれをコントロールする政策にとって極めて重要である。このネットワークを理解するにあたっては、共同研究開発の長期間、大規模なデータが必要である。そこで本研究では、共同出願特許のデータを用いる。

共同出願特許は組織間の共同研究開発の結果の一部にすぎないが、本研究の主眼は、各組織間の厳密な共同研究開発活動について調べるのではなく、それが構成するネットワークの構造の把握にある。その共同研究開発のネットワークの構造は、共同出願特許ネットワークの構造から十分推測できると考える。

多くの先行研究では組織間ネットワークの構造がスモールワールド [3] やスケールフリー [4] であるとよいという議論がなされているが、ネットワークの構造が個別組織やそれらの集団にどのような影響を与えるのかについては、いまだ定かとはいえない。そのため本研究では、ネットワークの構造がいかに組織の活動に影響するのか、という議論は行わない。

本研究の目的は、ネットワークの構造の分析とそれを再現する物理学的なモデルの検証にある。本研究ではこれまでに、共同出願特許の分析 [5] を行い、これを再現するモデルの提案 [6] を行ってきた。本論文で特に焦点をあてるのは、ネットワークにおける集積具合を示す指標、クラスタリング係数である。この指標は、ネットワークにおいて密にリンクを保持するノードの集団を表す。密なリンクは社会ネットワークにおけるインタラクションの濃さと単純に理解でき、ネットワーク構造の理解に極めて重要な観点といえる。従来のモデルでは、高いクラスタリング係数を再現することができなかったため、本研究では、その改善を行ってきた [7]。本論文では、その改善が未だ不足であることを指摘し、

¹⁾ 本原稿で組織とは、法人格を持つものとする。

²⁾ ここでネットワークとは、結節点（ノード）とそれらの間のつながり（リンク）から構成されるものとする。

今後新たなモデルを提案するにあたって、どのような改善を行うべきかについて述べる。

2. 社会科学と経済物理学の違いによる研究の位置づけ

本論文の一般的な読者を想定すると、社会科学ではなく経済物理学 [8] の研究のフレームワークに従っている本論文の内容は、理解が困難である可能性がある。そこで本章をもうけて、本論文がどのような問題に、どのような形でアプローチしようとしているのかについて述べ、1章を補足する。

一般的な社会科学の研究のフレームワークにおいては、社会におけるあるモデルを想定し、そのモデルを裏付けるデータの分析結果をもって研究の帰結とする。一方で、本論文を含む一連の研究、すなわち、経済物理学の研究手法はこの点で決定的に異なる。それは、これまでの社会科学に不足していたモデルの数理的あるいは数値解析的検証というプロセスを追加するものである。このことにより、あるモデルが本当にその社会の現象を表出しているのかどうか検証可能となる。このような検証を可能とするには、統計的に十分な数のデータとそれを処理する計算機の発達が必要であるが、近年において、大規模データの整備は進み、計算機の発達も顕著である。

本論文の内容は上記のような研究上の背景の中で、以下のような問題を扱う。すなわち、企業の間で特許を共同出願する際、そのネットワークはどのように構築されていくのか、である。そして、これがわかることで、イノベーションの発生過程の理解を進めることができる。また本論文ではモデルを検証する中で特にクラスティング係数の再現性を改善することに主眼がある。

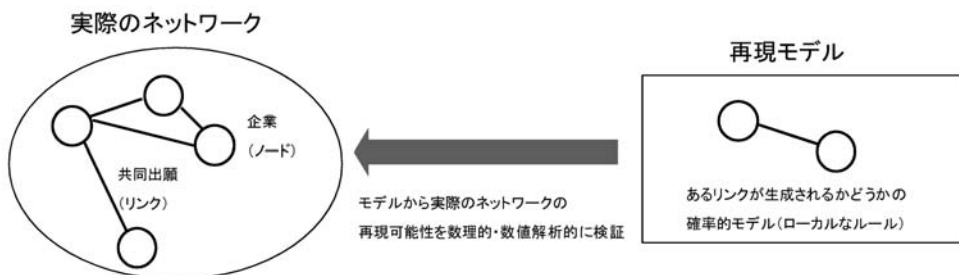


図1 実際のネットワークと再現モデルの関係：データを分析するだけでなく、実際にモデルが現実を再現可能であるか検証する点で従来の社会科学とは異なる。

3. 共同出願特許ネットワーク

3. 1 特許データ

本研究では特許を元にネットワークを構築し、その分析とモデルの検証を行う。このネットワーク構築の元となるデータは、日本の公開特許公報および特許公報において1993年1月から2002年12月の10年間に記載された4,998,464件の特許である。

特許には、出願人や発明者の氏名・名称や住所、特許の内容となる請求項など、いくつかの項目がある。それらのうち、出願人の名称や住所については、変更の度に過去の特許をさかのぼって更新する必要がある。このように変更に従えば、別の出願人と判断されるため、ある出願人に関する特許を正確に数えることができない。

本研究では上記のような変更に従っていない特許データではなく、変更の追従をしたデータベース（TamadaDatabase [9]）を利用する。後述の3.2節のネットワークの作成の方法に現れる特許上の項目は、出願人の名称（正確には‘氏名又は名称’）と住所（正確には‘住所又は居所’）、および発明者の住所（同上）である

3. 2 共同出願特許ネットワークの作成方法

本論文で分析を行う共同出願特許ネットワークの作成方法について説明する。ノードは組織の拠点ごとに作る。その識別子は組織名と住所の組み合わせである。リンクは上記のノードの間で共同出願が1つでもあるとき生成される。リンクの重複は無視される¹⁾。

ほとんどの特許において、出願人は組織であり、出願人名は組織名を含んでいる。そしてその住所は本社を指している。実際の発明が行われた拠点を求めるためには、発明者の住所を用いる必要がある。

上記に基づいた具体的なネットワーク作成アルゴリズムを以下に述べる。

ネットワーク作成アルゴリズム

すべての特許について以下を実行する。

1. ノードを作成する。

¹⁾ 共同出願1つでリンクとするかはアドホックな面がある。複数の共同出願によってリンクとする場合などは今後の課題である。

表1 ネットワーク基礎データ

拠点による分離	あり	なし
ノード数	54,197	24,767
リンク数	154,205	105,088

1.1 出願人名にフィルタ²⁾を用いて組織名を得る。出願人名が個人³⁾の場合はその出願人を無視する。出願人名すべてが個人のとき、何もせず次の特許に移る。組織名と対応する住所の組をノードの識別子（複数組ありうる）とし、これらをノードの候補にする。

1.2 発明者の住所を用いて組織の他の住所（拠点）を得る。住所に1.1で得た組織名のいずれかを含んでいる場合は、該当する組織名と町丁目までに縮めた住所をノードの識別子とし、それをノードの候補に追加する。このような処理が必要なのは、特許中に発明者がどの組織に属するか陽には示されていないためである。このとき重複は無視する。

1.3 特に1.1で作成された各々のノードの候補は次の条件を満たすときに削除される。

- すべての発明者は住所に組織名を持っている。
- いずれの発明者もそのノードの候補の住所を持っていない。

これはつまり、すべての発明者が組織の本社としての拠点ではなく、支社などの拠点にすることが判明しているときには、本社としての拠点はこの発明に関係していないとしてノードの候補からはずすことを意味する。

2. リンクを作成する。

1により作成されたノードの候補を新たなノードとし、それらの間に完全グラフの形で

²⁾ 以下の法人格を表す文字列を除く。（これらは特許に出現したものであり、法人格を表す文字列のすべてではない。）それらの文字列は株式会社、有限会社、特定非営利活動法人、財団法人、学校法人、独立行政法人、国立大学法人、医療法人、社団法人、医療法人社団である。これらの文字列がなければ法人格ではないとする。

³⁾ 上述のフィルタにより法人格でないとされたもの。

リンクを張る。ただし、この特許以前に生成されたノードと識別子が一致する場合は、ノードを新たに生成せず、その既存のノードを用いる。

4. ネットワークの分析

作成された共同出願特許ネットワークを分析する。図1は、共同出願特許ネットワークのノード数とリンク数の基礎データを示している。本論文では、出願人を拠点の数で分離したが、分離しなかった場合についても示してある。分離によってノード数は2.2倍、リンク数は1.5倍になっている。分離しなかった場合でも、ネットワーク構造に差異がないこともありうるが、本論文の分析では、発明が実際に起きた場所に関心があるため、拠点の分離により発明の起きた場所に近づけている。

図2はネットワークに関する各分布について示したものである。まず、(a)は次数分布を示している。次数とはノードに接続されているリンクの数である。ある次数を持つノードの分布をプロットしたのがこの次数分布である。この図において、横軸は次数、縦軸は順位であり、両対数である。

順位は累積確率と同様の意味合いがあるため、順位について説明する前に、先に累積確率について説明する。累積確率分布とは、 k 以上の次数をもつノードが出現する確率である。多くの文献では、累積確率分布ではなく確率密度関数 ($p(k)$) を用いて次数分布を議論しているが、確率密度関数と累積確率分布の関係は、次数の連続性を仮定すると

$$P(k) = \int_k^{\infty} dk p(k)$$

である。したがって、もし $p(k)$ がべき分布、すなわち

$$p(k) \propto k^{-\gamma}$$

ならば、

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

である。

次に順位と累積確率の関係について述べる。同じ値のデータに同じ順位を与えるが、そのときの順位は一般的に (1つ前の順位 + 1) である。しかしここでは (1つ前の順位 + その値のデータの数) とする。この定義ならば各順位を全体のデータの数で割ったときに、累積確率と一致する。本論文では、累積確率で表現できるところは、すべて順位で統一し

である。この理由は、（本論文なら距離などの）連続的な値に対して確率を求めるときに、（例えば10km 以上20km 未満などの）そのデータの区間を設定する必要がある。この区間を任意に変えることで、べき分布を初めとする関係性が存在するかのように設定できてしまう。本論文はこのような不明瞭さを避け、より厳密な累積確率が議論できるように順位で統一している。

図2 の (a) では1993年から2002年までの累積データでネットワークを作った場合の次数分布を示している。この図からいずれのプロットも累積確率が直線でフィットできることを示している。すなわちべき分布である。よってスケールフリーネットワークの1つであるということが出来る。最小自乗法による直線の傾きは -1.3 である。この値は近似であり、より正確な値を求めることもできる [10]。しかし、本論文ではこの値で十分議論可能である。この傾き ($-\gamma + 1 = -1.3$) は、ほとんどのスケールフリーネットワークにおいて $-2 < (-\gamma + 1) < -1$ である [4] ので、このネットワークが典型的な傾きのスケールフリーネットワークであるとわかる。

また図2 の (b) は、1平方 km あたりのノード数を密度とし、その分布をプロットしている。横軸は密度、縦軸は順位であり、両対数である。このプロットは直線に近い形であるため、べき分布と思われる。

次に、図2の (c) は、リンクの距離の分布をプロットしている。リンクの距離とはリンクの張られている2つのノードの間の、地球を楕円体とみなしたときの球面上の距離である。住所から緯度・経度を得ることで、この距離が求められる。横軸はリンクの距離、縦軸は順位であり、横軸を対数とする片対数である。ただし1 km 以下は省略してある。直線に近い形であることから、リンクの距離の対数に対して累積確率が比例していると推測される。すなわち、リンクの距離の確率分布は距離に反比例すると推測される。

共同研究開発などを通じた科学的知見の交流には、地理的近接性が重要であるという研究が、Jaffe などを初めとしてこれまで多く行われてきた [11] が、個別の共同研究開発の距離は確認されていなかった。一方で、本論文では図2の (c) のように、はっきりと各共同研究開発でのリンクの距離がプロットされ、距離に対する減衰性がわかったことは意義がある。

クラスティング係数 [3] とは、あるノードにつながったノードの集合において、それら集合の間に実現できるリンクのうち、どれだけが実現されているかの割合である。すなわち、次数 k のある頂点 i のクラスティング係数 C_i は、

$$C(i) = \frac{e_i}{kC_2} \quad (1)$$

である。ただし、 e_i はノード i につながっているノードの集合の間に存在するリンク数であり、 C は組み合わせの略記である。また、グラフに含まれるノード全体の C_i の平均を C とする。すなわち、

$$C = \frac{1}{n} \sum_{i=1}^n C_i$$

である。

図2の(d)は、次数が同じノードについてクラスタリング係数を平均し、それをプロットしたものである。この図において、縦軸は平均クラスタリング係数、横軸は次数である。クラスタリング係数の性質上、次数が大きくなるほどその値は低下するのが一般的である。また、全ノードの平均クラスタリング係数は、次数0, 1のノードを除いて求めた場合、0.53である。これらのノードを含んだ場合は、0.11である。この数値は自然界や人の社会に存在する他のネットワークと比べて非常に高く、この共同出願特許ネットワークがかなり特異な構造を持つことがわかる。

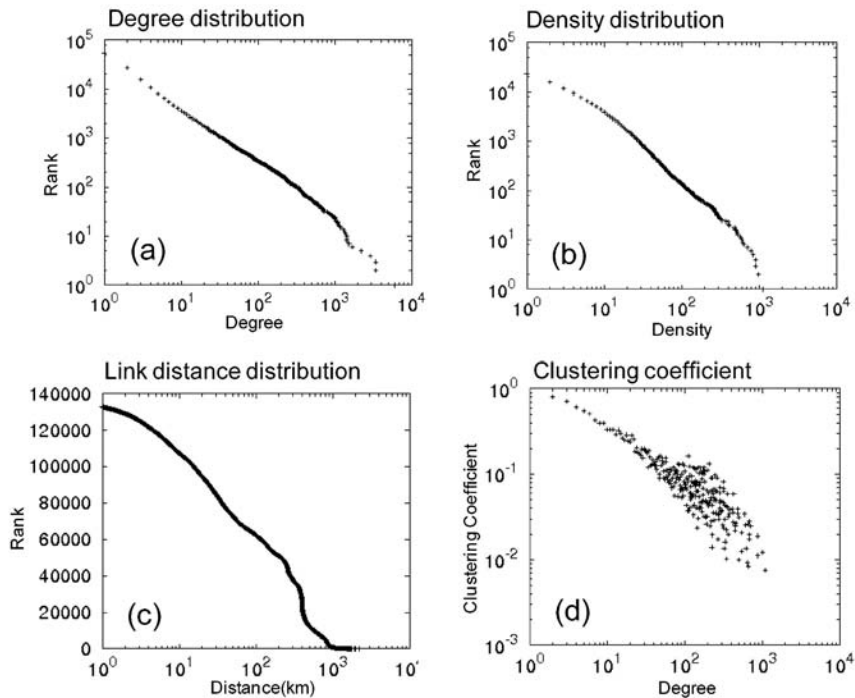


図2 各分布のデータ：(a) 次数分布。(b) ノードの密度。(c) リンクの距離の分布。(d) 次数あたりの平均クラスタリング係数。

この高いクラスタリング係数の再現モデルはこれまでに存在していなかったが、そのようなモデルを本論文で検証する。

5. 従来の再現モデルにおける問題点の検証

優先的選択型成長モデル [12] は、スケールフリーネットワーク、すなわち次数のべき分布を再現するための方法として広く知られている。初期設定を含む細かな部分は割愛するが、そのモデルでは、新しいノードが加えられ、新たなリンクを張るとき、そのリンクがノード i と張られる確率は

$$\Pi(k_i) = k_i / \sum_i k_i \quad (2)$$

となる。ただし、 k は次数である。このとき次数分布は

$$p(k) \propto k^{-3}$$

となる。式 (2) のより一般的な形として

$$\Pi(k_i) \propto k_i^\alpha$$

があり、べき分布の指数を変化させることができる。しかしながら、このモデルにおいては距離が考慮されておらず、距離の影響について検討するためには、モデルの中に距離を含める必要があった。

本研究ではこれまでに優先的選択型成長モデルを、距離を組み入れるために最小限に変更したモデル [13, 14, 15] を用いて、共同出願特許ネットワークの再現を検証してきた [6]。その検証においては、実際のノードの地理的な分布を用いて、それらノードに対してリンクを形成し、次数分布とリンクの距離の分布が再現されるかを検討した。そのモデルにおいては、次数分布とリンクの距離の分布について再現性を確認した。しかしながら、各ノードのクラスタリング係数については優先的選択型成長モデルと同様にほとんど0であることがわかり、共同出願特許ネットワークを十分に再現するに至っていなかった。

優先的選択型成長モデルを改善し、高いクラスタリング係数を実現できるようにしたモデル [16] が存在する。このモデルを参考に、上述したこれまでのモデルに対して、さらに改善したモデルを本研究で提案してきた [7]。そのモデルは以下になる。

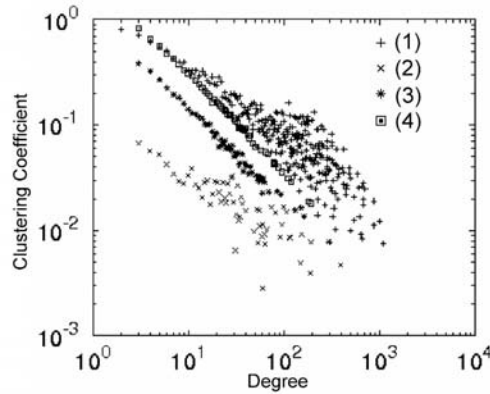


図3 実際のネットワークと再現モデルの次数あたり平均クラスタリング係数：(1) 実際のネットワーク。(2) $pTF = 0$ 。(3) $pTF = 0.5$ 。(4) $pTF = 1.0$ 。

ネットワーク再現モデル

1. m_0 個のノードからなる完全グラフから始める。
2. ノードを新たに1つ増やす。4章で分析した共同出願特許ネットワークにおいて、実在するノードの住所を、新たなノードの住所としてランダムに与える。
3. 新たに増やしたノードから1本のリンクを生成する。このとき重複を許さない。新たなノードを i とし、リンクが張られる先のノード j は次の確率で選ばれる。

$$\Pi(k_j, d_{ij}) \propto k_j^a / d_{ij}^\sigma$$

ただし、 k_j は j の次数、 d_{ij} はノード i と j の間の距離である。また a と σ は定数である。

4. 新たに増やしたノードから $(m - 1)$ 本のリンクを順に生成する。各リンクにつき、 $(1 - pTF)$ の確率で3. と同様にリンクを生成する。それ以外 (pTF の確率) のときは、直前にリンクを張ったノードとつながったノードをランダムに選んで、ノード i からリンクを生成する。該当するノードが1つもないときは3. と同様のプロセスを行う。
5. 必要なノード数に達していなければ2. へ戻る。

このモデルでは pTF を変化させ、クラスタリング係数の再現性との関係を調べた。図3は数値解析的に求めたクラスタリング係数のデータである。横軸は次数であり、縦軸はその次数に対応するノードのクラスタリング係数の平均である。そして、(1) 実際のネッ

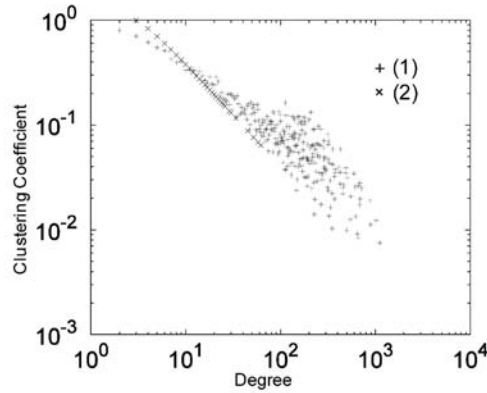


図4 実際のネットワークとノード非活性モデルのクラスタリング係数：(1) 実際のネットワーク。(2) ノード非活性モデル

トワーク，(2) $pTF = 0$ ，(3) $pTF = 0.5$ ，(4) $pTF = 1.0$ がプロットされている。なお， $m_0 = 3$ ， $m = 3$ ， $a = 1$ ， $\sigma = 1$ ，最終のノード数は3,000個である。図からわかるように， pTF が大きいほど，全体のクラスタリング係数も大きい。また，3つの候補の中では $pTF = 1.0$ が実際のネットワークに近い。

これ以前のモデルは， $pTF = 0$ のときに対応する。前述したように，次数とリンクの距離の分布（図2の（a）および（c））をよく再現できることは確認したが，全体の平均クラスタリング係数，および次数あたりの平均クラスタリング係数は非常に小さく，全く再現できていなかった。

再現モデルでは次数2以下のノードは存在しないため，次数3以上のノード全体での平均クラスタリング係数を求めると，上述の（1）から（4）に対して，0.53, 0.05, 0.30, 0.62である。これからもわかるとおり $pTF=1.0$ がもっとも実際のネットワークに近い。ただし，平均クラスタリング係数はノード数に対してべき乗で減少することがわかっており，実際のネットワークがノード数54,197， $pTF=1.0$ のネットワークがノード数3,000であることから考えると，相당한開きがあることがわかる。

なお，次数分布およびリンクの距離の分布（図2の（a）と（c））における再現性が，このモデルにおいて変わっていないことが重要であるが，全く変化はなかった。優先的選択型成長モデルに対してクラスタリング係数を調整できるよう拡張したモデルにおいて，次数分布は変化しないことが証明されている [16] が，距離を含んだ優先的選択型成長モデルに対して，クラスタリング係数を調整できるよう拡張したこのモデルにおいて，次数とリンクの距離の分布は変化しないということが，数値解析上は確認できた，ということ

になる。

平均クラスタリング係数が実際のネットワークに及ばないだけでなく、次数あたりの平均クラスタリング係数の分布も十分に再現できていない。実際のネットワークにおいて顕著であるのは、大きな次数のノードにおいても、高いクラスタリング係数を持つことである。次数あたり平均クラスタリング係数におけるモデル間の違いを検証するため、最小二乗法によって求められたべき指数の大きさについて議論する。実際のネットワークのべき指数は -0.67 であり、上述の $pTF = 1.0$ では -0.93 である。図3からもわかるが、大きな次数のノードになるほど、 $pTF = 1.0$ のモデルは低いクラスタリング係数となっている。式(1)において、クラスタリング係数は次数の負の2乗に比例するが、実際のネットワークのべき指数は -0.67 であり、 -2 よりもかなり大きい。

本研究がこれまでに提案してきた再現モデルは以上であるが、スケールフリーネットワークを高いクラスタリング係数で実現する他のモデルが存在する。それはノードの非活性化を用いたモデル [17] であり、次のようなアルゴリズムで作成する。

ノード非活性モデル

1. m 個の活性ノードからなる完全グラフを作成する。
2. 新しいノードを1つ追加する。
3. 新しいノードからすべての活性ノードへリンクを張る。
4. 以下の確率で既存活性ノードを非活性化する。

$$P(k_j) = \frac{\gamma - 1}{a + k_j} \quad (3)$$

ただし、 k_j はノード j の次数、 $\gamma - 1 = (\sum_{l \in A} \frac{1}{a + k_l})^{-1}$ 、 A はすべてのノードの集合、 a は定数である。

5. 新しいノードを活性ノードとする。
6. 必要なノード数に達していなければ2.へ戻る。

このモデルにおいて、初期ノード m を3、非活性化の定数 a を3に設定し、本論文の他の検証と同様に、3000ノード、9000リンクを作成した場合の結果を図4に示す。図の意味は図3と同じである。べき指数は、非活性化モデルの場合もやはり -0.93 と実際のネットワークに比べるとかなり小さい。これは a をいかに変えても同様の傾向である。

実際のネットワークに見られるような、高い次数での高いクラスタリング係数の実現が可能な既存のモデルは、上記からまだ存在しないといえる。

6. 再現モデル改善点の検証

5章での、従来の再現モデルの問題点を受け、本論文ではどのようにモデルを改善すればよいのかを検証する。具体的には、従来モデルによって作成されたネットワークにおけるリンクを、クラスタリング係数が上がるように付け替えたとき、実際のネットワークとどれくらい乖離するのかを、数値解析的に検証する。クラスタリング係数を上げるには、リンク数を増やす方法も考えられるが、ノード数とリンク数の比は実際のネットワークと同じである必要があるため、リンクを付け替える方法をとる。

本論文では、既存のネットワークに対して以下のようなリンク付け替えアルゴリズムを適用する。

1. ネットワーク中のノードからランダムに1つ選び、ノード i とする。
2. ノード i につながれたノードのうち、次の割合 p のノードの集合 B を選ぶ、

$$p = b k^\delta \quad (4)$$

ただし、選ばれるノードは次数が2以上であるとする。次数が2以上のノードが足りないときは、不足分を無視する。

3. ノードの集合 B からランダムにノードを1つ選びそれをノード j とする。
4. ノード j をノードの集合 B から取り除く。
5. ノード j とつながっているが、ノード j とつながっていないノードをランダムに選び、ノード l とする。該当するノードがなければ7.へ飛ぶ。
6. ノード j とノード l にリンクを張る。
7. ノードの集合 B が空でなければ3.へ飛ぶ。
8. すべてのノードがノード i として選ばれるまで、1.に戻る。

上記のアルゴリズムにより、クラスタリング係数が上がるように、リンクを付け替えることができる。各ノード i について付け替える割合 p が式 (4) のようになるのは、平均クラスタリング係数と次数あたり平均クラスタリングの傾きを調整することと関係がある。すなわち、 b は平均クラスタリング係数を、 δ は次数あたり平均クラスタリング係数の傾きをそれぞれ調整する。 b が平均クラスタリング係数に対応する理由は明らかである。一方で、 δ が次数あたり平均クラスタリング係数に対応する理由については説明が必要である。5章での議論では次数あたり平均クラスタリング係数の分布は、

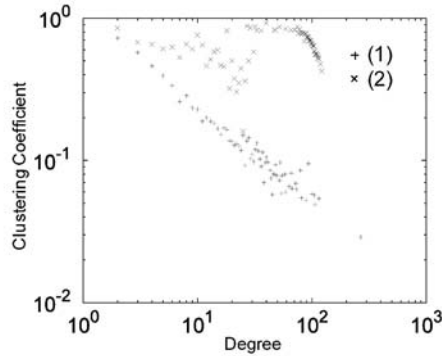


図5 リンク付け替え後の次数あたり平均クラスタリング係数：(1) $\beta = 1$ かつ $\xi = 0$,
(2) $\beta = 2.5$ かつ $\xi = 0.75$

$$\overline{p_{C_k}} \propto k^\xi \quad (5)$$

のような形で表現できると推測した。ただし、次数を k 、次数あたり平均クラスタリング係数の確率を $\overline{p_{C_k}}$ とする。従来のモデルによるネットワークのべき指数を ξ_1 、実際のネットワークのべき指数を ξ_2 とすると、その差は $k^{\xi_1 - \xi_2}$ に比例する。すなわち、次数に対するべきによって次数あたり平均クラスタリング係数を調整しなければならないことから、式 (4) が適当であると推測した。

元となるネットワークとして、5章におけるクラスタリング係数を向上する前のモデル（すなわち $p_{TF} = 0$ のモデル）を用いて、上記のアルゴリズムを適用した結果を示したのが、図5である。図5においては、 $\beta = 1$ かつ $\xi = 0$ 、および $\beta = 2.5$ かつ $\xi = 0.75$ の場合の次数あたり平均クラスタリング係数がプロットされている。図5を見ると、 $\beta = 1$ かつ $\xi = 0$ は実際のネットワークとは全く一致していない。やはり $\overline{p_{C_k}}$ を定数とするのはよくないことがわかる。一方で、 $\beta = 2.5$ かつ $\xi = 0.75$ はべき分布のようにみえる。実際のネットワークの次数あたり平均クラスタリング係数の分布におけるべき指数は -0.67 、修正前の元のネットワークは -0.93 であるが、このネットワークのべき指数は -0.64 である。このことから、このネットワークが修正前の元のネットワークにおいて、次数あたり平均クラスタリング係数において実際のネットワークに近付いていることがわかる。

この修正されたネットワークが、実際のネットワークの他の指標、すなわち、次数分布およびリンクの距離の分布において差があるかどうか、今後のモデル検討において重要となる。

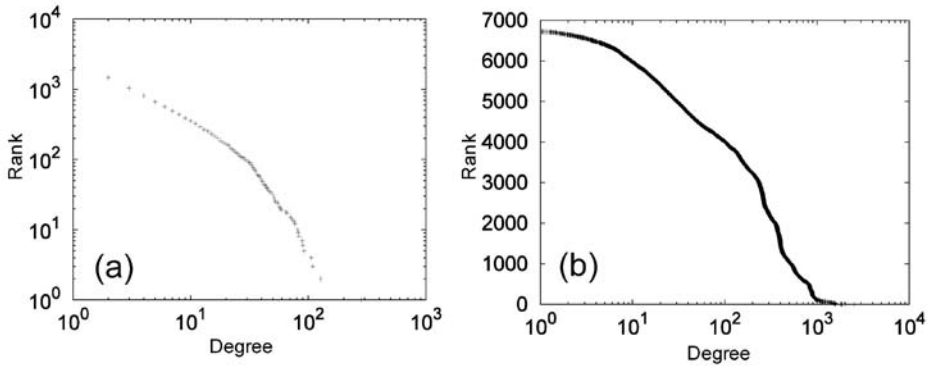


図6 $\beta = 2.5$ かつ $\xi = 0.75$ によって修正されたネットワークの (a) 次数分布, (b) リンクの距離の分布。

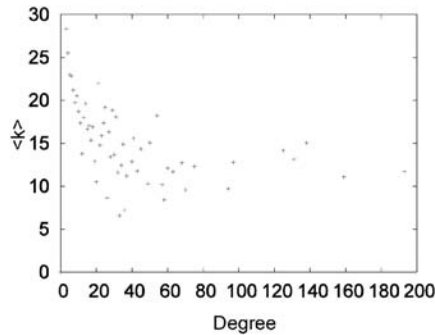


図7 修正前のネットワークの次数相関

図6は、左が次数分布、右がリンクの距離の分布である。これらの図の意味は図2と同じである。図2と比較するとリンクの距離の分布は同様に反比例しているが、次数分布においてはべき分布が破れているように見える。特に次数の 10^1 から 10^2 付近で分布が増加していると思われる。この原因は、ネットワークの修正方法と次数相関から説明できる。

次数相関とは、ノードにつながれたノードの次数を平均し、その平均値を次数ごとに平均したものである。図7はその次数相関を示している。縦軸が $\langle k \rangle$ で次数相関の値であり、横軸は次数である。この図からわかるように、次数20より小さいあたりの低い次数のノードは、そのノードの次数よりも大きな次数のノードとつながっている傾向にあり、それ以上の次数になるとほぼ次数10で一定である。

リンクの付け替えにおいては、1本のリンクを付け替えるときに次数の下がるノードと上がるノードが存在する。各リンクが付け替えにおいて選ばれる確率は一樣であることが

ら、あるノードの次数が下がる確率はそのノードの次数に比例する。このことは、次数分布において全体の傾きが小さくなることに対応する。一方で、式(4)からわかるように、あるノードの次数が上がる確率は、小さい次数のノードにつながっているノードほど大きい。これは逆に次数が小さいノードは次数が上がらないことを意味する。以上から、次数分布全体の傾きは小さくなり、次数の小さいところでの確率分布が小さくなることから、図6のような形になると考えられる。

以上の議論から、今後新しい再現モデルを提案する際に、本章で行ったネットワークの修正を取り込むならば、高いクラスタリング係数が再現できる一方で、次数分布において違いが生じることがわかる。したがって、今後再現モデルを改善するにあたっては、この次数分布の変化に対策する必要があるということがわかった。

7. まとめ

本論文では、共同出願特許ネットワークにおける高クラスタリング係数を再現できるようなモデルの実現に向け、従来研究の不足点を指摘し、改善の方向性を指摘した。

検証の結果、本研究で提案してきた従来のモデルでは、ノード全体の平均クラスタリング係数だけでなく、特に次数に対するクラスタリング係数の分布の傾きにおいて再現できていないことがわかった。そこで、どのようにモデルを改善すれば高いクラスタリング係数を実現できるかを探るため、次数に対するべき乗の確率でクラスタリング係数が上がるようにリンクを付け替えたところ、実際のネットワークに近付くことができることを発見した。一方で、単純な反映では次数分布における不一致を生じるため、これを回避するような改善が今後必要であるということがわかった。

共同出願特許ネットワークを再現できるモデルの存在を示したことは、価値があると考えられる反面、企業や大学、公的研究機関の共同研究開発は、技術分野の一致を初めとする複雑な条件の上で成り立っていると考えられる。したがって、本論文で議論したモデルで扱われる説明変数と、共同研究開発に対する社会科学的な分析で必要とされる説明変数とが一致するのかどうかについては、今後議論されねばならない。

そのときに注意しなければならないのは、局所的に共同研究開発の成立の因果関係を探ったとしても、その因果関係を元にして共同研究開発ネットワークの全体の構造が再現できないのであれば、それは誤ったモデルにすぎないということである。したがって、全体のネットワークの構造を把握し、それを再現できるモデルを模索することと、局所的なインタラクションを検証してリンクの生成過程を把握することは同時に進めるべき研究の

対象である。本研究が指向しているように、数理的解析が困難な社会シミュレーションではなく、実際のデータに基づく物理学的なモデルによってミクロとマクロをつなぐような研究を行うことは、今後の社会科学にとって大きな意味を持つと考える。

謝辞

本研究を進めるにあたり、国際電気基礎技術研究所の相馬亘氏、関西学院大学大学院の玉田俊平太氏に議論いただいた。ここに謝意を記す。本研究は科研費（20730268）の助成を受けたものである。

参考文献

- [1] W. Souma. 経済における複雑ネットワークー日本の経済ネットワークは特殊か？ー. 人工知能学会, 20 (3) : 289-295, 2005.
- [2] 相馬 亘. 株所有ネットワークのシンプルなモデル. 情報処理学会論文誌, 47 (3) : 850-856, 2006.
- [3] D.J. Watts and S.H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440-442, 1998.
- [4] A.L. Barabási and Z.N. Oltvai. Network biology: Understanding the cell’s functional organization. *Nature Reviews Genetics*, 5:101-113, 2004.
- [5] H. Inoue, W. Souma, and S. Tamada. Spatial characteristics of joint application networks in japanese patents. *Physica A*, 383:152-157, 2007.
- [6] 井上寛康, 相馬 亘, and 玉田俊平太. 共同研究開発の過程理解に向けた特許のネットワーク分析とモデル構築. 情報処理学会論文誌, 49 (4), 2008. to be published.
- [7] 井上寛康. イノベーションネットワークと地理モデル. 青山秀明・相馬亘・藤原義久 共編著, ネットワーク科学への招待ー世界の“つながり”を知る科学と思考, pages 128-134, 2008.
- [8] 青山秀明, 家富 洋, 池田裕一, 相馬 亘, and 藤原義久. 経済物理学. 共立出版, 2008.
- [9] S. Tamada, Y. Naitou, F. Kodama, K. Gemba, and J. Suzuki. Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, 68 (2) : 289-302, 2006.
- [10] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *arXiv:0706.1062*, 2007.

- [11] R. Ponds, F. van Oort, and K. Frenken. The geographical and institutional proximity of scientific collaboration networks. *Regional Science*, 86 (3) : 423-444, 2007.
- [12] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509-512, 1999.
- [13] S. Yook, H. Jeong, and A.L. Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences*, 99 (21) : 13382-13386, 2002.
- [14] S.S. Manna and P. Sen. Modulated scale-free network in the euclidean space. *Physical Review E*, 66, 066114, 2002.
- [15] S.S. Manna and A. Kabakcioglu. Scale-free network on euclidean space optimized by rewiring of links. *Journal of Physics A: Mathematical and General*, 36 (19) : L279-L285, 2003.
- [16] P. Holme and B.J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65, 026107, 2002.
- [17] K. Klemm and V.M. Eguiluz. Highly clustered scale-free networks. *Physical Review E*, 65, 036123, 2002.