

論文 / 著書情報
Article / Book Information

題目(和文)	複雑ネットワークの成長モデルの構築とその応用
Title(English)	
著者(和文)	三浦航
Author(English)	Wataru Miura
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第9419号, 授与年月日:2014年3月26日, 学位の種別:課程博士, 審査員:高安 美佐子,小野 功,樺島 祥介,出口 弘,寺野 隆雄
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第9419号, Conferred date:2014/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	要約
Type(English)	Outline

平成 25 年度知能システム科学専攻博士論文概要

複雑ネットワークの成長モデルの構築とその応用

三浦 航

提出年月日 2014 年 3 月 18 日

目次

第 1 章	序論	5
1.1	複雑ネットワーク研究の背景	5
1.2	優先的成長モデルの数理的背景	5
1.3	凝集現象の数理的背景	9
1.4	目的と構成	11
第 2 章	ノードの凝集を考慮した一般的成長モデルの構築	13
2.1	ノードの凝集を考慮した一般的成長モデルの構築	13
2.2	既存の成長モデルとの対応関係	17
2.3	線形な優先的成長モデルの統計的性質	17
第 3 章	複雑ネットワークの成長モデルの応用	39
3.1	企業間取引ネットワークの既存研究	39
3.2	解析に用いるデータの説明	41
3.3	企業間取引ネットワークの解析	43
3.4	企業間取引ネットワークへの成長モデルの応用	48
3.5	企業間取引ネットワークと成長モデルの改良	50
第 4 章	結論	63
4.1	本論文のまとめ	63
4.2	今後の展望	64
	参考文献	67

第 1 章

序論

この章では複雑ネットワークの研究背景や、本研究に関連する研究の背景および、本研究の目的と論文の構成について述べる。

1.1 複雑ネットワーク研究の背景

多くの要素から構成される系全体の振る舞いは、要素単独の振る舞いを考慮するだけでは理解できない場合が多い。近年、要素間の複雑な相互作用を記述し、系全体の振る舞いや現象を理解しようとする試みとして、複雑ネットワークの分野 [1-6] が注目を集めている。こうした要素 (ノード) と、要素間の相互作用 (リンク) からシステムが構成され、現象を複雑ネットワークとして捉えることができるものとしては WWW [7]、インターネット [8]、代謝物質の反応 [9]、タンパク質の反応 [10]、性的接触 [11]、論文の共著者 [12]、航空網 [13]、食物網 [14]、携帯電話の通話 [15] など、自然界から社会まで多岐にわたる。こうして要素間の関係性に注目した、複雑ネットワークを代表する性質の一つとして、1つのノードが持つリンク数 (次数) の分布がべき分布に従うという、スケールフリー性が知られている。

Barabási と Albert はこのスケールフリー性を生み出す原因が、ノードが新たにリンクを獲得する確率は次数に比例するという、優先的接続性にあると仮定し BA モデル [16, 17] を考案した。BA モデルはネットワークにノードを追加し続ける成長モデルである。しかしノードの素過程として、生成だけではなくノードの消滅やノード同士の結合など、様々な振る舞いがネットワーク上で実際には観測される。

BA モデル以降、多くのスケールフリー性を持つ複雑ネットワーク生成モデル [17-27] が考案された。次節でこれらのモデルについて、簡単に紹介を行う。

1.2 優先的成長モデルの数理的背景

BA モデルは、ネットワークの成長と優先的接続という 2つの効果によって、ネットワークのスケールフリー性を生み出すモデルである。BA モデル [16] のアルゴリズムは次

のようなものである。

BA モデルのアルゴリズム

1. 初期状態として, m_0 個のノードを用意する. ノード間のリンクの状態は任意であるが, 各ノードは次数 1 以上 ($k_i \geq 1$) を持つ.
2. 各ステップにおいて, リンクを m 本持つノードを 1 つネットワークに追加する. リンクの接続先は, ノード i へ接続する確率を $\Pi(k_i) = k_i / \sum_j k_j$ とし, m 個のノードを確率的に決定する.
3. ノード数が目的のサイズに達するまで, ステップ (2) を繰り返す.

この簡単なモデルにより, 次数分布のベキ指数が 3 のスケールフリーネットワークが生成される. 上記の BA モデルの特徴としては, まずネットワークにノードを追加し続ける, 成長モデルであるという点である. 次に, 新たなリンクの接続先をノード i の次数 k_i に比例した確率

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (1.1)$$

で決める, 優先的接続という概念を用いた点である.

BA モデルはシンプルでありながら, ベキ指数が 3 のスケールフリーネットワークしか生成できないなど, 現実のネットワークからかけ離れた点も多かったため, それ以降多くの複雑ネットワーク生成モデルが考案された. それらのモデルは, ネットワークが成長するモデル [17–22, 25–27] と, 成長しないモデル [23, 24] に大別することができる. 成長するモデルとはノード数が変化し (ほとんどの場合単調増加である), 次数分布が定常状態に達するモデルである. 成長しないモデルとは, ノード数が初期状態から変化せず, リンク数とその状態だけが変化していくモデルである. 任意の次数分布に従うネットワークを生成するコンフィグモデル [23] や, あらかじめノードに割り当てた重みに従って, ノード間にリンクを張る Goh モデル [24] などが, 成長しないモデルの中に含まれる. 次項では複雑ネットワークの成長モデルについて述べていく.

1.2.1 複雑ネットワークの成長モデル

最初の成長モデルである BA モデルは, 次数分布のベキ指数が 3 のネットワークしか生成できないモデルであったため, 以降考案されたモデルはそのほとんどがベキ指数を変化させることができる. もちろんこれらのモデルはそれぞれの現象に着目し, 現象に即して考案されたわけであるが, ここではモデルの数理面のみに焦点を当てる. 次数分布のベキ指数をコントロールするために, (i) 各事象の生起確率でコントロールするモデルと, (ii) 優先的接続の確率を工夫することでコントロールするモデルが存在する.

各事象の生起確率が変化したモデル

(i) のタイプに属するのは, Albert と Barabási ら自身による拡張 BA モデル [17] である. 彼らは次のように BA モデルを拡張した.

拡張 BA モデルのアルゴリズム

1. 初期状態として、 m_0 個のノードを用意する。ノード間のリンクの状態は任意である。
2. 各ステップにおいて、次のどれかの操作を行う。
 - (a) 確率 p で m 本 ($m \leq m_0$) の新しいリンクを追加する。リンクの片方の接続先のノードはランダムに選び、もう片方の接続先はノード i へ接続する確率を $\Pi(k_i) = (k_i + 1) / \sum_j (k_j + 1)$ として、優先的接続を行う。
 - (b) 確率 q で m 本のリンクを付け替える。ランダムにノード i とリンク l_{ij} を選択し、そのリンクを取り除く。ノード j を優先的接続の確率 $\Pi(k_j)$ を用いて決め、リンク $l_{ij'}$ として付け替える。
 - (c) 確率 $1 - p - q$ で新しいノードを追加し、 m 本のリンクの接続先を優先的接続の確率 $\Pi(k_i)$ を用いて決める。
3. 目的のサイズに達するまでステップ (2) を繰り返す。

この拡張 BA モデルでは、既存の BA モデルで考慮されているノードの生成以外にも、リンクの生成、リンクの付け替えという事象を考慮し、これらのバランスで次数分布のベキ指数を変化させることができるようになっている。この拡張 BA モデルでは、次数分布のベキ指数は $3 - 2q + (1 - p - q)/m$ (≥ 2) で決まる。

引用ネットワークのモデル (GNR モデル) [22] では、ネットワーク内に新規に生成したノードが確率 $1 - r$ で、ランダムに接続先のノードを選択しリンクを張る。しかし確率 r で、ランダムに選択したノードではなく、その隣接ノードを選ぶ。これは引用しようとした論文の、参考文献の論文を引用することに対応している。あるノードが次数 k となる確率分布を $P(k)$ とする (これを次数分布と呼ぶ) とき、その隣接ノードの次数分布は $kP(k)/\langle k \rangle$ で表される。これにより次数 k に比例した事実上の優先的接続を行い、次数分布のベキ指数を $1 + 1/r$ と変化させている。

以上のモデルはネットワークにノードを追加し続ける成長モデルである。しかし現実のネットワークでは、引用ネットワーク [28, 29] における論文数のようにノードが増加するばかりではなく、WWW [16, 18] でウェブページが削除されるように、ノードが減少することも考えられる。そこで Moore らはノードの生成に加え、ノードが消滅するモデル [27] を考案した。

Moore モデルのアルゴリズム

1. 初期状態として、 m_0 個のノードを用意する。
2. 各ステップで、ネットワークに 1 つの新規ノードを追加し、 m 本のリンクの接続先を優先的接続 $\Pi(k_i) = k_i / \sum_j k_j$ によって選択する。
3. ノードの追加と同時に、確率 r ($0 \leq r \leq 1$) でノードを取り除く。取り除く際には、そのノードが持つリンクも全て取り除く。

Moore らはこのモデルのシミュレーションを行なっただけでなく、マスター方程式を立

て次数分布の理論解を求めた。

ある時刻 t においてノード数を N 、次数 k となるノードの確率分布を $P(k, t)$ とおくと、次数 k のノード数は $NP(k, t)$ である。1ステップで次数 k のノード数の変化について考慮すると、

$$(N+1-r)P(k, t+1) - NP(k, t) = m \left(\frac{k-1}{N\langle k \rangle} NP(k-1, t) - \frac{k}{N\langle k \rangle} NP(k, t) \right) + \delta_{k,m} \quad (1.2)$$

$$+ r(k+1)P(k+1, t) - rkP(k, t) - rP(k, t)$$

となる。ここで、 $\sum_j k_j = N\langle k \rangle$ としている。1ステップでノード数 N は、確率 $1-r$ で1つ増加する。右辺一行目の初項は、新規ノードからリンクを接続され $k-1 \rightarrow k$ となるノード数を表し、第二項は $k \rightarrow k+1$ となるノード数、最後の項は次数 m の新規ノードの増加を表す。右辺二行目の初項は、隣接ノードが消滅し $k+1 \rightarrow k$ となるノード数、第二項は $k \rightarrow k-1$ となるノード数、最後の項は確率 r で選択され除去されるノードをそれぞれ表す。式 (1.2) において、 $r=0$ の場合は BA モデルに一致する。また平均次数 $\langle k \rangle$ は時刻 t で、リンク数の2倍をノード数で割ることにより、

$$\langle k \rangle = 2 \frac{(m-r\langle k \rangle)t}{(1-r)t} \quad (1.3)$$

となり、

$$\langle k \rangle = \frac{2m}{1+r} \quad (1.4)$$

と求まる。

定常状態においては、 $P(k, t+1) = P(k, t) = P(k)$ である。母関数変換を用いて式 (1.2) のマスター方程式を解くと $k \rightarrow \infty$ で、

$$P(k) \sim \begin{cases} k^{-(3-r)/(1-r)} & (r < 1) \\ k^{-3/4} e^{-2\sqrt{k}} & (r = 1) \end{cases} \quad (1.5)$$

となる。ノードの生成が消滅よりも起こる場合 ($r < 1$) に次数分布はベキ分布に従い、そのベキ指数は $(3-r)(1-r) (\geq 3)$ となる。すなわち当然ではあるが、ノードの消滅が起こらず生成のみが起こる場合 ($r = 0$) は、BA モデルと同じくベキ指数が3に従う次数分布のネットワークが生成される。しかし、 $r = 1$ でノードの生成と消滅が等しく起こる場合には、ベキ指数は発散し次数分布は拡張指数分布に従う。

Moore らのモデルでは、次数分布のベキ指数は3以上の値となる。これはノードの消滅時に、そのノードが持つリンクが全て消滅し、リンクのもう一方の接続先のハブのリンクが取り除かれやすいからである。したがってノードの消滅を考慮すると、ハブと大多数のノードの次数の差は広がりやすく、次数分布のベキ指数は大きな値となる。しかし現実のネットワークでは、次数分布のベキ指数が3未満のネットワークも多数存在する [7-9, 11, 13, 14].

優先的接続の確率が変化したモデル

優先的接続の確率を変えることにより次数分布のベキ指数を変化させるモデルとしては、Dorogovtsev, Mendes, Samukhin による DMS モデル [20] が知られている。彼らは BA モデルの優先的接続の確率を一般化し、 $\Pi(k_i) = (k_i + A) / \sum_j (k_j + A)$ とした。ここで A は各ノードの次数に依存しない魅力 (attractiveness) を表し、次数の小さいノードでも新たなリンクをある程度獲得できるようにしている。このモデルにおいて新規ノードをリンク数 m で生成する場合、ネットワークの次数分布のベキ指数は $3 + A/m$ で表される。 $-m \leq A < \infty$ で調整することで、ベキ指数が 2 以上のネットワークを生成することができる。

Krapivsky らは、優先的接続の確率を $\Pi(k_i) \sim k^\lambda$ とした、非線形な優先的接続を行う成長モデルを考案した [19]。当時、優先的接続というのは現実のネットワークから確認されたわけではなく、成長モデル上の概念に過ぎなかった。それ以降、優先的接続性が確認された現実のネットワークでも、優先的接続の確率の指数 $\lambda = 1$ とは限らない。後の章でも述べるが、現実のネットワークから推定される優先的接続の確率の指数は、論文の共著関係や俳優の共演関係のネットワークで 0.8 [30]、ニューヨークの服飾業の取引関係ネットワークで 1.2 [31] と、 $\lambda = 1$ ではないネットワークもあり様々である。Krapivsky らはモデルのマスター方程式を解くことで、次数分布の振る舞いを理論的に解析し、次数分布が $\lambda < 1$ で拡張指数分布、 $k \rightarrow \infty$ で漸近的に $\lambda = 1$ となる場合に指数が 2 以上のベキ分布、 $\lambda > 1$ でほぼ全てのリンクを独占するハブとなるノードが現れることを示した。彼らはこれをゾル-ゲル転移になぞらえて、ゲルノードと呼んでいる。

その他の優先的接続の確率が変化したモデルとしては、適応度という概念を導入した適応度モデル [21] がある。WWW のウェブページや論文の被引用回数など、後から登場し次数の小さなノードでも、短時間で多くのリンクを獲得することがある。適応度モデルではこうした特徴を適応度 η_i という量で表し、新たにリンクを獲得する優先的接続の確率を、 $\Pi(k_i) = \eta_i k_i / \sum_j \eta_j k_j$ とおく。すなわち新たにリンクを獲得する確率は、ノードの次数と適応度の積に比例する。適応度モデルでは、次数分布は各ノードに与えられる適応度 η_i の分布に依存し、必ずしもスケールフリーネットワークが生成されるとは限らない。

1.3 凝集現象の数理的背景

本節ではモデルの構築に必要な凝集現象の数理と、その研究背景について述べる。ベキ分布が観測される現象や、その背景に潜むメカニズムと数理構造について述べる。

1.3.1 粒子の注入・凝集現象

ベキ分布が観測される現象は多数知られているが、そのうちの一つに凝集現象 [32–37] がある。ミルクをかき混ぜるとやがてクリームになるように、粒子同士が凝集してサイズ

の大きなクラスターになる凝集現象の研究は、天体物理学 [38] や雲の物理 [39], 高分子重合 [40] など様々な分野で行われてきた。凝集現象ではクラスターの形状のフラクタル性などとともに、クラスターのサイズ分布について議論される。ここで言うサイズとは、クラスターの質量のことである。特にサイズ i とサイズ j のクラスター同士が凝集する確率 $K(i, j)$ と、サイズ分布の関係性について議論され、コロイドやエアロゾルの粒子が凝集して作るクラスターのサイズ分布はベキ分布に従うことが多い。

スモルコフスキー方程式

スモルコフスキー方程式 [41] は非可逆な凝集現象を表すのに使われる方程式である。もちろんクラスターの分裂という逆反応が起こることもあるが、ここで議論するのはこのような逆反応が無視できるタイムスケールにおいてである。サイズ k ($k = 1, 2, \dots$) の粒子の濃度を $c_k(t)$ とするとき、最も基本的な場合のスモルコフスキー方程式は

$$\dot{c}_k(t) = \frac{1}{2} \sum_{i+j=k} K(i, j)c_i(t)c_j(t) - c_k(t) \sum_{j=1}^{\infty} K(k, j)c_j(t) \quad (1.6)$$

と記述される。 $K(i, j)$ はサイズ i のクラスターと、サイズ j のクラスターが凝集する確率を表す。ここで右辺の初項はサイズが k になるクラスター同士の凝集を表し、第二項はサイズが k のクラスターとの凝集を表す。

この方程式は空間座標が含まれておらず、平均場近似の式であり、空間的に一様な系に対して適用される。空間的な阻害要素が存在する場合には通常、凝集確率 $K(i, j)$ にこうした要素が考慮されて組み込まれる [42]。系への注入や散逸、沈殿の有無などで、方程式 (1.6) はまた変化し、注入項や散逸項などを伴う [43–45]。

既存の凝集現象でよく考慮される凝集確率 $K(i, j)$ は、次のような性質を満たす [46, 47]。

$$K(ai, aj) = a^\lambda K(i, j) = a^\lambda K(j, i), \quad (1.7)$$

$$K(i, j) \simeq i^\mu j^\nu + i^\nu j^\mu \sim i^\mu j^\nu \quad (i \ll j, \lambda = \mu + \nu). \quad (1.8)$$

最初の性質はサイズ i, j に関して可換、かつ同次であるということである。次の性質は凝集確率 $K(i, j)$ が、サイズ i の μ 乗とサイズ j の ν 乗の積に比例するというものであり、van Dongen ら [46] は μ の値によって $\mu > 0$ でクラス I, $\mu = 0$ でクラス II, $\mu < 0$ でクラス III と凝集現象の分類を行なった。凝集過程はクラスター同士が反応するまでの時間が支配的 (反応律速凝集) か、クラスターが拡散するまでの時間が支配的 (拡散律速凝集) かで、その性質が異なる。クラス I では大きなサイズの粒子同士の凝集が支配的であり、高分子重合などに代表される。反応律速凝集がこのクラスに含まれる。クラス III では小さなサイズの粒子と大きなサイズの粒子の凝集が支配的であり、エアロゾルなどの拡散律速凝集がこのクラスに含まれる。クラス II では、凝集する粒子のサイズは無相関である。

空間構造を考慮した注入・凝集モデル

スモルコフスキー方程式は空間座標を含まない方程式である。そこで高安ら [48] は空間座標を導入した注入・凝集のモデルを導入した。このモデルは

$$c_i(t+1) = \sum_j w_{ij}(t)c_j(t) + I(i,t) \quad (1.9)$$

で表され、時刻 t におけるサイト i の粒子の質量 $c_i(t)$ について考慮する。式 (1.9) の w_{ij} は、サイト j から i へと遷移する粒子の割合を表し、 $I(i,t)$ はサイト i への注入を表す。数理モデルのシミュレーションや理論解析などから、系への小さな粒子の注入とこれらの粒子の凝集のバランスが釣り合うことで定常状態に達し、ベキ分布に従うサイズ分布が生成される [48–52] ことがわかってきた。

1.3.2 指数分布と指数関数

凝集現象の他にベキ分布が観測される数理構造として、指数分布と指数関数の組み合わせが挙げられる。ある変数 x と y が、

$$P(\geq x) = e^{-\alpha x} \quad (1.10)$$

$$y = y_0 e^{\beta x} \quad (1.11)$$

という関係にあるとする。すなわち変数 x は指数分布に従い、変数 y は x に対して指数関数的に増加する。このとき 2 番目の式より $x = \frac{1}{\beta} \log\left(\frac{y}{y_0}\right)$ は単調増加関数であるので、1 番目の式にこれを代入すると、

$$P(\geq y) = e^{-\frac{\alpha}{\beta} \log\left(\frac{y}{y_0}\right)} = \left(\frac{y}{y_0}\right)^{-\frac{\alpha}{\beta}} \quad (1.12)$$

と変数 y はベキ分布に従う。つまり、指数分布に従う変数を変数変換することにより、ベキ分布に従う変数を作り出すことができる。

1.4 目的と構成

BA モデルやその他のネットワーク生成モデルは、そのほとんどがネットワークにノードを追加し、ネットワークのノード数が単調増加するモデルである。またこれらのモデルでは、ノードの平均的な次数が時刻 t に対してベキ関数的に増加する [16, 17, 20, 21] ことが知られており、ノードの次数が指数関数的に成長するような場合を再現することはできない。Moore のモデルでは、BA モデルのノードの生成に加えノードの消滅を考慮しており、ネットワークのノード数が単調増加するわけではない点で、BA モデルよりも応用範囲が広い。しかしノードの生成、消滅だけでは次数分布のベキ指数が 3 以上のネットワークしか生成されず、しかもベキ指数が 3 のときは Moore モデルは BA モデルに一致する。

また、これらのモデルでは次数に線形比例する優先的接続を行うが、現実のネットワークでは優先的接続の指数が1であるとは限らない。

本論文ではこのような背景を踏まえ、ノード数が変化し次数が指数関数的に成長する、複雑ネットワークの成長モデルを構築する。成長モデル構築の際には、ノードの凝集をモデルに取り入れ、非線形な優先的接続を行うことで、複雑ネットワークの新しい一般的成長モデルを提案する。構築した成長モデルのモンテカルロシミュレーションを行い、パラメータを変化させることでモデルの振る舞いを調べる。のみならず1ステップでのノード数の遷移確率を考え、マスター方程式を立てて解くことにより、モデルの振る舞いを定量的に調べる。モデルのマスター方程式は、ノードの凝集について着目すると、凝集現象を記述するスモルコフスキー方程式とみなすことができる。

また構築した複雑ネットワークの成長モデルを、現実のネットワークに応用する。モデルを応用する対象として、日本企業約100万社から成る取引ネットワークを構築し、このネットワークの統計的性質をデータ解析より得る。提案した成長モデルが、現実のネットワークの統計的性質を再現できるか確認し、モデルの妥当性を検証する。本研究で構築したモデルは物理的にミニマルなモデルであり、当然のことながらそのまま企業という現実のネットワークに適用するには不十分である。現実の企業間取引ネットワークにはノードの生成や消滅に留まらず、取引先の変更や分社化など様々な要素が混在し、そうした要素もモデルを現実のネットワークに応用するには取り込んでいかなければならない。企業間取引ネットワークでは直接の取引相手だけではなく、取引相手がさらにどこと取引を行っているかという、取引相手の取引相手も重要である。これはネットワークで言うと、2リンクの関係に当たる。そこで企業間取引ネットワークの2リンクの構造を解析し、観測された性質を再現するために、構築したモデルの改良を行う。

本論文の構成は次のようになる。まず2章で、ノードの凝集を考慮し、非線形な優先的接続を用いた、複雑ネットワークの一般的成長モデルの構築を行う。2章では、このモデルの線形な優先的接続を行う場合について、モデルのマスター方程式を導出し、その解を求めシミュレーションを行うことで、モデルの統計的性質について詳しく調べる。3章では、成長モデルを現実のネットワークに応用することを試みる。応用例として、企業間取引ネットワークに着目する。用いる企業データと企業間取引ネットワークについて説明し、その統計的性質を調べる。その後一般的成長モデルが、あるパラメータ下で観測された企業間取引ネットワークの統計的性質を、再現できることを確認する。またモデルの再現度を高めるために、企業間取引ネットワークの2リンク構造について解析を行い、2リンク構造を再現するようなモデルの改良を行う。4章では本論文のまとめと今後の課題について述べる。

第 2 章

ノードの凝集を考慮した一般的成長モデルの構築

本章では、ノードの凝集効果を考慮し、非線形な優先的接続を用いた、複雑ネットワークの一般的な成長モデルの構築を行う。構築した成長モデルのシミュレーションを行い、パラメータを変化させてモデルの性質について調べる。また成長モデルの 1 ステップでの次数の変化を考慮し、ノード数についてのマスター方程式を立てて解くことで、モデルの性質を定量的に調べる。

2.1 ノードの凝集を考慮した一般的成長モデルの構築

本節では複雑ネットワークの成長モデルの構築を行う [53–56]。既存の BA モデルで考慮されているノードの追加、Moore モデルで考慮されているノードの消滅に加えて、本研究ではノードの凝集を考慮して成長モデルを構築する。既存の成長モデルでは、ノードがリンクを得る際に 1 本ずつしか得ることができない。ある程度の数のリンクを一度にまとめて得ることができるようにすることで、ノードの指数関数的な成長が可能になると考えられる。また、優先的接続を次数に線形比例した確率で行うのではなく、非線形性を取り入れることで、既存のモデルよりも様々な現象が観測されるようにする。

いま成長モデルを構成する素過程としてノードの消滅、追加、凝集について考慮する。この概念図を図 2.1 に示す。成長モデルのアルゴリズムを次のように導入し、モデルを構築する。

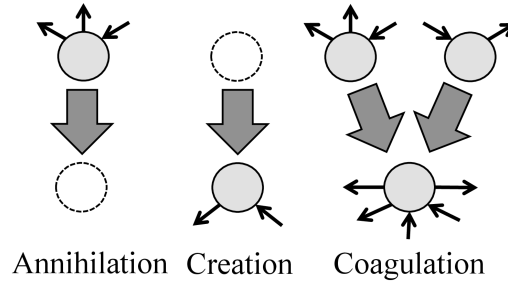


図 2.1 成長モデルを構成するノードの素過程. BA モデルで考慮されるノードの生成 (Creation), Moore モデルで考慮されるノードの消滅 (Annihilation) に加え, 本研究では新たにノードの凝集 (Coagulation) について考慮する.

提案する成長モデルのアルゴリズム

1. 初期状態として, N_0 個のノードを用意する. このとき, ノード間のリンクの状態は任意である.
2. 各ステップにおいて, 次のどれかの操作を確率的に行う.
 - (a) 確率 a でノードの消滅を行う. ランダムに選択したノードを 1 つ取り除き, ノードに付随するリンクも全て取り除く.
 - (b) 確率 b でノードの追加を行う. 入出次数が $1(k_{in} = k_{out} = 1)$ の新規ノードを 1 つネットワークに追加し, 新規ノードから出るリンクの接続先 (つまり既存ノードは入るリンクを受け取る) を確率 $\Pi(k_{i,in})$ で, 新規ノードに入るリンクの接続先 (既存ノードは出るリンクを受け取る) を確率 $\Pi(k_{i,out})$ で優先的に選択する.
 - (c) 確率 c でノードの凝集を行う. ランダムに選択したノードを 1 つ取り除き, 付随するリンクは全て確率 $\Pi(k_{i,in}, k_{i,out})$ で優先的に選択したノードに付け替える. リンクの付け替え先にリンクが既に存在し重複する場合には, 2 本をまとめて 1 本のリンクとする.
3. ネットワークにおけるノードの平均次数 $\langle k \rangle$ が定常状態に達するまで, 手順 2 を繰り返す.

ここで, $\Pi(k_{i,in}), \Pi(k_{i,out}), \Pi(k_{i,in}, k_{i,out})$ は次の式によってそれぞれ定義される, 優先的接続の確率である.

$$\Pi_b(k_{i,in}) = \frac{k_{i,in}^{\lambda_b} + A}{\sum_j (k_{j,in}^{\lambda_b} + A)}, \quad (2.1)$$

$$\Pi_b(k_{i,out}) = \frac{k_{i,out}^{\lambda_b} + A}{\sum_j (k_{j,out}^{\lambda_b} + A)}, \quad (2.2)$$

$$\Pi_c(k_{i,in}, k_{i,out}) = \frac{k_{i,in}^{\lambda_c} + k_{i,out}^{\lambda_c} + A}{\sum_j (k_{j,in}^{\lambda_c} + k_{j,out}^{\lambda_c} + A)}. \quad (2.3)$$

現実のネットワークでは, 常に次数に線形比例した優先的接続の確率を持つ (すなわち

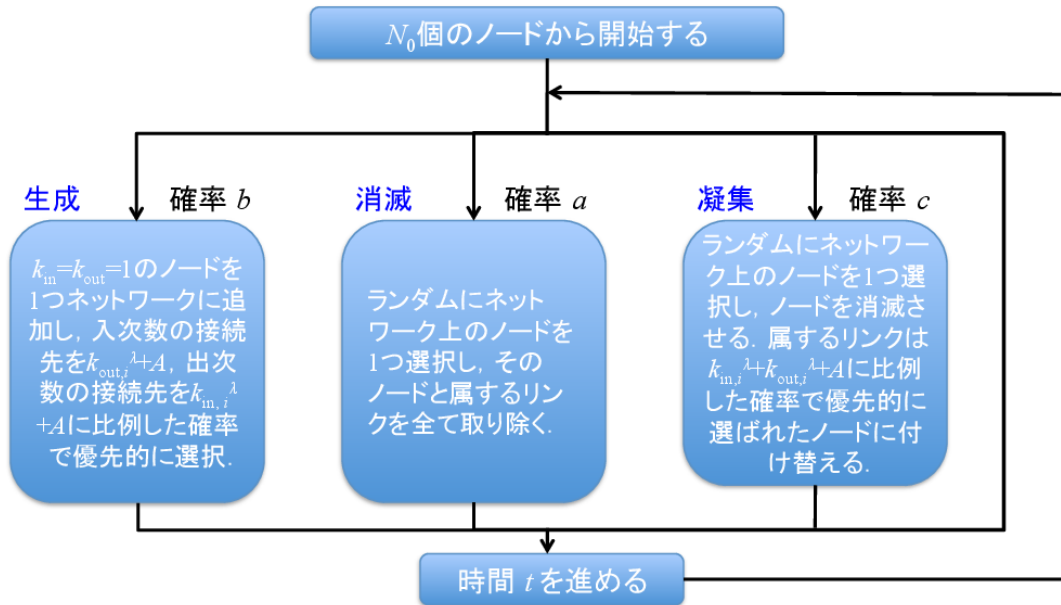


図 2.2 成長モデルのフローチャート。

$\lambda_b = \lambda_c = 1$) とは限らないため、上式では非線形な優先的接続の確率を採用している。またノードの生成時の優先的接続の確率の指数を λ_b , 凝集時の優先的接続の確率の指数を λ_c とし、ノードの生成時と凝集時で異なる指数を持つ。優先的接続の確率を $+A$ した値に比例させているのは、ノードの消滅時にリンクを除去することにより、次数が0のノードも存在しうるからである。こうした次数0のノードでも、次数に依存しない値 A に比例した確率で、リンクを得ることができるようにしている。

以上のアルゴリズムをフローチャートとして図 2.2 に示す。この成長モデルにおいて、1ステップでノードの消滅、生成、凝集のいずれかは行われることから $a+b+c=1$ であり、消滅時と凝集時にノード数は1つ減少し、生成時にノード数は1つ増加する。したがって $b = a+c = 0.5$ の時に、ネットワークの平均ノード数が一定となる準定常状態が実現される。それ以外の状態では $b > 0.5$ でノード数は増加し続け、 $b < 0.5$ でノード数は減少し続ける。

次に、この一般的成長モデルのマスター方程式を導出する。入次数についてマスター方程式を立てるが、いまモデルの入出リンクに対して対称性があるため、出次数についても全く同じ方程式となる。時刻 t で入次数が k となる確率を $P(k, t)$ とおき、時刻 t から $t+1$ に変化する際の、入次数 k のノード数 $NP(k, t)$ の変化について考える。モデルのマスター

方程式は次のようになる.

$$\begin{aligned}
(N + 2b - 1)P(k, t + 1) - NP(k, t) = & a \left[\frac{k+1}{N} NP(k+1, t) - \left(\frac{k}{N} + \frac{1}{N} \right) NP(k, t) \right] \\
& + b \left[\frac{(k-1)^{l_b} + A}{\sum_j (k_j^{l_b} + A)} NP(k-1, t) - \frac{k^{l_b} + A}{\sum_j (k_j^{l_b} + A)} NP(k, t) + \delta_{k,1} \right] \\
& + c \left[\sum_{k'=0}^k P(k-k', t) \frac{2k'^{l_c} + A}{\sum_j (2k_j^{l_c} + A)} NP(k', t) - \sum_{k'=0}^{\infty} P(k, t) \frac{2k'^{l_c} + A}{\sum_j (2k_j^{l_c} + A)} NP(k', t) \right. \\
& \left. - \sum_{k'=0}^{\infty} P(k', t) \frac{2k^{l_c} + A}{\sum_j (2k_j^{l_c} + A)} NP(k, t) \right]. \quad (2.4)
\end{aligned}$$

ここで左辺は、入次数 k となるノード数の変化を表す. 確率 b のノードの生成時にノード数は 1 増え, それ以外でノード数は 1 減る. 時刻 t においてノード数が N とすると, 1 ステップでノード数は $b - (1 - b) = 2b - 1$ 増えるので, 時刻 $t+1$ ではノード数が $N + 2b - 1$ となる. 右辺はそれらの変化が, ノードの消滅 (第一項), 生成 (第二項), 凝集 (第三項) のどの事象によってもたらされるか, 次数の遷移確率を表している. それぞれの項について説明していく.

ノードの消滅の際には, 確率 $1/N$ で次数 k のノードが倒産 ($k \rightarrow 0$), 確率 k/N で隣接ノードが倒産し次数が減る ($k \rightarrow k-1$), 確率 $(k+1)/N$ で隣接ノードが倒産し次数が減る ($k+1 \rightarrow k$) という 3 つの場合が考えられる. ここでカッコ内はその際の次数の変化を表す. ノードの生成の際には, 優先的接続により確率 $[(k-1)^{l_b} + A] / \sum_j (k_j^{l_b} + A)$ で新規ノードからリンクを受け取る ($k-1 \rightarrow k$), 同じく優先的接続により確率 $(k^{l_b} + A) / \sum_j (k_j^{l_b} + A)$ で新規ノードからリンクを受け取る ($k \rightarrow k+1$), 新規ノードにより次数 1 のノードが生成されるという, こちらも 3 つの場合が考えられる. ノードの凝集の際には, 次数 k のノードが消滅して次数 k' のノードと凝集する, 次数 k' のノードが消滅して次数 k のノードと凝集する, 次数 $k-k'$ のノードと次数 k' のノードが凝集するという, 3 つの場合が考えられる. 凝集の際には総次数 (入次数と出次数の和) に比例した優先的接続を行うが, これを入次数の 2 倍に比例するものと仮定する. するとそれぞれの場合が起こる確率は, $P(k, t) \frac{2k'^{l_c} + A}{\sum_j (2k_j^{l_c} + A)}, P(k', t) \frac{2k^{l_c} + A}{\sum_j (2k_j^{l_c} + A)}, P(k-k', t) \frac{2k'^{l_c} + A}{\sum_j (2k_j^{l_c} + A)}$ となる.

ノードの凝集でリンクを張り替える際に, リンクが重複する場合は 2 本のリンクをまとめて 1 本とするが, ここではリンクをまとめる確率を無視してマスター方程式を立てる. 本来, 和を取る際の $k'=0$ や $k'=k$ では, 次数が変化しないので式の中に含まれないが, ここでは計算の為に含めている. 結局, 項同士で相殺しているので, 結果には影響ない. ここで畳み込みの項がサイズが k になる粒子同士の凝集を表し, その他の和を取っている項はサイズが k から外れる粒子の凝集を表す. したがって方程式 (2.4) は, 粒子の凝集現象を記述するのに使われるスモルコフスキー方程式と同等であり, スモルコフスキー方程式にさらに系への注入, 消滅 (ノードの生成, 消滅) を考慮した方程式となっている. 以上が, 構築した一般的な成長モデルのマスター方程式である.

2.2 既存の成長モデルとの対応関係

前節で構築したモデルのパラメータを整理しておく。まずノード数の増減を決め、ネットワークが成長するか衰退するかを決める、ノードの生成確率 b が存在する。次にノードの凝集と消滅の割合を決める、ノードの凝集確率 c が存在する。生成確率 c と凝集確率 b を決めることで、消滅確率 a は $1 - b - c$ で一意に決まる。また優先的接続の線形、非線形性を決める指数 λ_b, λ_c が存在する。 $\lambda_b, \lambda_c = 0$ では接続先が次数に比例せずランダムになり、 λ_b, λ_c が大きな値を取ることで、次数の大きなノードほど新たにリンクを受け取る確率が高くなる、非線形性の強い優先的接続になる。優先的接続に関しては、次数に依存しない値である、魅力度 A というパラメータも存在する。そして、ネットワークのサイズに関わる、初期ノード数 N_0 が存在する。

構築したモデルは、様々な場合を含む一般的な成長モデルであり、パラメータが特定の値を取ることで、既存の成長モデルに一致する。すなわち $b = \lambda_b = \lambda_c = 1, a = c = A = 0$ で BA モデルに一致し、 $\lambda_b = \lambda_c = 1, c = A = 0$ で Moore モデルに一致する。さらに $\lambda_b = \lambda_c = 0, a = c = 0$ ではランダムグラフに一致する。Moore モデルではノードの生成に対する消滅の割合が r として表されるので、本研究のモデルでは a/b が Moore モデルの r に相当する。

本章ではまず $\lambda_b = \lambda_c = 1$ とした、線形な優先的接続を行う成長モデルの場合について、BA モデル、Moore モデルとの対応関係や、パラメータ N_0, c, b の影響について調べる。

2.3 線形な優先的成長モデルの統計的性質

本節では一般的な成長モデルのシミュレーションを、優先的接続の指数 $\lambda_b, \lambda_c = 1$ として固定し、それ以外の様々なパラメータの値を変化させて行う。まず成長モデルが特定のパラメータ下で、既存の BA モデルと Moore モデルに一致することを確認する。次に、平均ノード数が一定の場合の時間発展について調べ、その後にパラメータ依存性について確認する。成長モデルにおいて、入次数と出次数に対する操作は同じなので、入次数と出次数は同じ挙動を示す。したがってどちらを観測しても問題なく、以降は主に入次数について観測していく。

2.3.1 BA モデルに一致する場合 ($b = 1, c = 0, A = 0$)

まず $b = 1, c = 0, A = 0$ とした、ノードの生成のみが起こる場合について考え、本研究の成長モデルが BA モデルに一致することを確認する。このときモデルのマスター方程式は、式 (2.4) より次のようになる。

$$(N + 1)P(k, t + 1) - NP(k, t) = \frac{k - 1}{\sum_j k_j} NP(k - 1, t) - \frac{k}{\sum_j k_j} NP(k, t) + \delta_{k,1}. \quad (2.5)$$

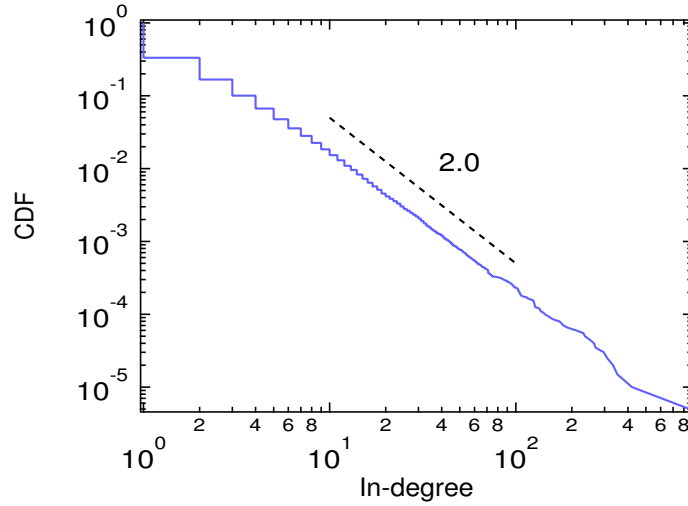


図 2.3 モデルが BA モデルに一致する場合の累積入次数分布.

成長モデルでは，新規ノードの入出次数を $1(k_{\text{in}} = k_{\text{out}} = 1)$ としているので， $\sum_j k_j = 2N$ となる．したがって定常状態においては，

$$P(k) \simeq \frac{k-1}{k+2} P(k-1) \quad (2.6)$$

$$\sim k^{-3}$$

となり，指数 3 のべき分布に従う．

この結果を確認するために，初期ノード数 $N_0 = 5$ ，シミュレーション時間 $T = 2 \times 10^5$ として，モデルのシミュレーションを行う．このとき毎ステップで，ネットワークに新たにノードが追加されるので，シミュレーション終了時のノード数 N は $N = 2.0 \times 10^5$ となっている．図 2.3 にシミュレーション結果の累積次数分布を示す．BA モデルでは指数 3 の次数分布を持つので，累積では指数 2 となり図 2.3 のフィッティングのようになる．前述のパラメータの成長モデルで，BA モデルの結果を再現出来ていることがわかる．

2.3.2 Moore モデルに一致する場合 ($c = 0, A = 0$)

次に $c = 0, A = 0$ として，ノードの生成，消滅が起こる場合について考え，本研究の成長モデルが Moore モデルに一致することを確認する．このときモデルのマスター方程式は，式 (2.4) より次のようになる．

$$(N + 2b - 1)P(k, t + 1) - NP(k, t) = (1 - b) \left[\frac{k+1}{N} NP(k+1, t) - \left(\frac{k}{N} + \frac{1}{N} \right) NP(k, t) \right] \quad (2.7)$$

$$+ b \left[\frac{k-1}{\sum_j k_j} NP(k-1, t) - \frac{k}{\sum_j k_j} NP(k, t) + \delta_{k,1} \right].$$

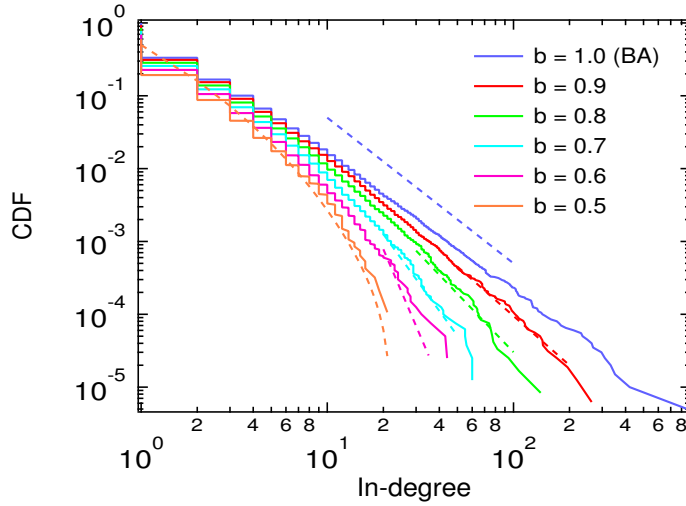


図 2.4 モデルが Moore モデルに一致する場合の入次数分布. フィッティングは式 (2.8) による.

式 (1.2) との右辺の比較により, $(1-b)/b$ が Moore モデルの r に対応していることがわかるので, このときの解は

$$P(k) \sim \begin{cases} k^{-(4-\frac{1}{b})/(2-\frac{1}{b})} & (b > 0.5) \\ k^{-3/4} e^{-2\sqrt{k}} & (b = 0.5) \end{cases} \quad (2.8)$$

となる. $b = 1$ のときは, 前項の BA モデルに一致する.

次に成長モデルのシミュレーションを行い, この結果を確認する. シミュレーション時間 $T = 2 \times 10^5$ としてモデルのシミュレーションを行うが, ノードの生成確率 b (消滅確率 a) がパラメータごとに異なるので, シミュレーション終了時のノード数 N はそれぞれのパラメータで異なる. ノード数 N は平均で,

$$N = N_0 + (2b - 1)T \quad (2.9)$$

に従う. 初期ノード数 N_0 は, ノードの生成確率 $b = 1, 0.9, 0.8$ で $N_0 = 5$, $b = 0.7, 0.6$ で $N_0 = 10$, $b = 0.5$ で $N_0 = 10^4$ である. $b = 0.5$ ではノードの生成と消滅が釣り合うので, この場合の初期ノード数だけは多くしなければならない. 生成確率 $b > 0.5$ では, ノードの生成回数が消滅回数を上回るので, ノード数は増加する.

シミュレーションによる次数分布を図 2.4 に示す. $b = 1$ では BA モデルに一致するので, 前項と同じ次数分布を示す. ノードの消滅が生成に対して増加するとともに, 隣接ノードが除去されることでハブのリンクも除去されるので, ハブの次数は減少する. すなわちべき指数は増加していくことがわかる. また式 (2.8) によるフィッティングが, シミュレーション結果と一致している. $c = 0, A = 0$ というパラメータでの成長モデルで, Moore モデルの結果を再現出来ていることがわかる.

2.3.3 ノードの消滅と生成が起こる場合 ($b = 0.5, c = 0, A = 1$)

本項では、ノードの消滅と生成が起こる場合 ($b = 0.5, c = 0, A = 1$) について考慮する。これは Moore モデルにおける $b = 0.5$ の場合とほぼ同じであるが、成長モデルにおいて凝集を考慮する場合の、前段階として重要である。

このパラメータでの成長モデルのマスター方程式は、式 (2.4) より次のようになる。

$$2N(P(k, t+1) - P(k, t)) = \frac{k+1}{N}NP(k+1, t) - \left(\frac{k}{N} + \frac{1}{N}\right)NP(k, t) + \frac{k}{\sum_j(k_j+1)}NP(k-1, t) - \frac{k+1}{\sum_j(k_j+1)}NP(k, t) + \delta_{k,1}. \quad (2.10)$$

式 (2.10) に、 $k_j = \langle k \rangle = 1$ を代入する^{*1}。定常状態において、左辺の値は 0 となる。以上より式 (2.10) を整理すると、

$$kP(k-1) - 3(k+1)P(k) + 2(k+1)P(k+1) + 2\delta_{k,1} = 0. \quad (2.11)$$

これを母関数変換 ($\phi(z) = \sum_k P(k)z^k$) を行うことにより解く。式 (2.11) の両辺に、 z^k をかけて k について和を取ると、

$$\left(z^2 \frac{d\phi}{dz} + z\phi\right) - 3\left(z \frac{d\phi}{dz} + \phi\right) + 2\frac{d\phi}{dz} + 2z = 0, \quad (2.12)$$

$$(1-z)(2-z)\frac{d\phi}{dz} - (3-z)\phi + 2z = 0 \quad (2.13)$$

となる。両辺に $(1-z)/(2-z)^2$ をかけてこれを解くと、

$$\phi(z) = \frac{2(2-z)}{(1-z)^2} \left[z - 3 + \frac{2}{2-z} + 3 \log(2-z) \right] \quad (2.14)$$

と $\phi(z)$ を求めることができる。あとはこれをテイラー展開すれば良く、 $k \geq 1$ で

$$P(k) = 6 \left[(k+2) \sum_{m=k+1}^{\infty} \frac{2^{-m}}{m} - 2^{-k} \right] \sim \frac{2^{-k}}{k+1} \quad (2.15)$$

となる。

$b = 0.5, c = 0, A = 1$ としてモデルのシミュレーションを行なった結果と、式 (2.15) の解を、図 2.5 に片対数プロットで示す。シミュレーションでは次数分布は指数関数的に減衰するが、理論解でもそうなっていることがわかる。

^{*1} これについては「平均次数の時間発展」という項で説明する。

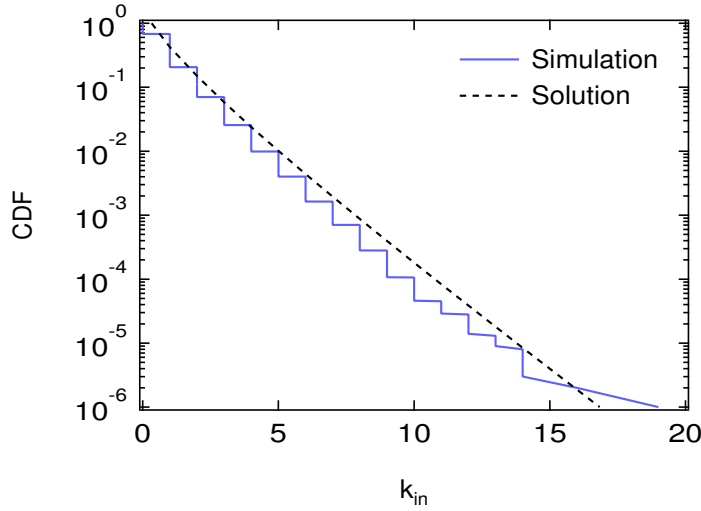


図 2.5 モデルのシミュレーション結果 (青) とマスター方程式の理論解 (黒).

2.3.4 ノードの凝集と生成が起こる場合 ($b = c = 0.5, A = 1$)

本項では、ノードの消滅と生成が起こる場合 ($b = c = 0.5, A = 1$) について考慮する。このパラメータでの成長モデルのマスター方程式は、式 (2.4) より次のようになる。

$$\begin{aligned}
 2N(P(k, t+1) - P(k, t)) &= \frac{k}{\sum_j (k_j + 1)} NP(k-1, t) - \frac{k+1}{\sum_j (k_j + 1)} NP(k, t) + \delta_{k,1} \\
 &\quad - \sum_{k'=0}^{\infty} P(k, t) \frac{2k'+1}{\sum_j (2k_j + 1)} NP(k', t) - \sum_{k'=0}^{\infty} P(k', t) \frac{2k+1}{\sum_j (2k_j + 1)} NP(k, t) \\
 &\quad + \sum_{k'=0}^k P(k-k', t) \frac{2k'+1}{\sum_j (2k_j + 1)} NP(k', t).
 \end{aligned} \tag{2.16}$$

$k_j = \langle k \rangle$ とし、 $k \gg 1$ として定常状態について考えると、

$$\frac{kP(k-1) - (k+1)P(k)}{\langle k \rangle + 1} + \frac{2}{2\langle k \rangle + 1} \left(\sum_{k'=0}^k k' P(k') P(k-k') - \sum_{k'=0}^{\infty} (k+k') P(k') P(k) \right) = 0 \tag{2.17}$$

となる。スモルコフスキー方程式で理論解を得ることができるのは、凝集確率が $1, k_1 k_2, k_1 + k_2$ に線形比例する時のみなので、凝集項に関して次のように式の対称化

を行う.

$$\begin{aligned}
& 2 \left(\sum_{k'=0}^k k' P(k') P(k-k') - \sum_{k'=0}^{\infty} (k+k') P(k') P(k) \right) \\
&= \sum_{k_1+k_2=k} k_1 P(k_1) P(k_2) + \sum_{k_1+k_2=k} k_2 P(k_1) P(k_2) - \sum_{k_1=0}^{\infty} (k+k_1) P(k_1) P(k) - \sum_{k_2=0}^{\infty} (k+k_2) P(k_2) P(k) \\
&= \sum_{k_1+k_2=k} (k_1+k_2) P(k_1) P(k_2) - \sum_{k_1=0}^{\infty} (k+k_1) P(k_1) P(k) - \sum_{k_2=0}^{\infty} (k+k_2) P(k_2) P(k)
\end{aligned} \tag{2.18}$$

さらに次数 k が $k \gg 1$ で連続であるとみなすことにより, 式 (2.17) は

$$\begin{aligned}
& \frac{kP(k-1) - (k+1)P(k)}{\langle k \rangle + 1} + \int_0^{\infty} dk_1 dk_2 [K(k_1, k_2) P(k_1) P(k_2) \delta(k - k_1 - k_2) \\
& \quad - K(k_1, k) P(k_1) P(k) \delta(k_2 - k - k_1) - K(k, k_2) P(k) P(k_2) \delta(k_1 - k_2 - k)] = 0
\end{aligned} \tag{2.19}$$

と表せる. ただし,

$$K(k_1, k_2) = \frac{k_1 + k_2}{2\langle k \rangle + 1} \tag{2.20}$$

であり, 凝集確率を両クラスターの和に比例させている.

この式に $P(k) = Ck^{-x}$ を代入し, 積分項の2番目と3番目の項に対して, それぞれ次の Zharov 変換 [57] と呼ばれる変数変換を行う.

$$(k_1, k_2) \rightarrow \left(\frac{kk'_1}{k'_2}, \frac{k^2}{k'_2} \right), \tag{2.21}$$

$$(k_1, k_2) \rightarrow \left(\frac{k^2}{k'_1}, \frac{kk'_2}{k'_1} \right). \tag{2.22}$$

すると2番目の積分項について式 (2.21) より,

$$\begin{aligned}
& \int_0^{\infty} dk_1 dk_2 K(k_1, k) P(k) P(k_1) \delta(k_2 - k - k_1) \\
&= C^2 \int_0^{\infty} dk_1 dk_2 K(k_1, k_2) \left(\frac{k}{k_2} \right)^{3-2x} (k_1 k_2)^{-x} \delta(k - k_1 - k_2).
\end{aligned} \tag{2.23}$$

3番目の積分項について式 (2.22) より,

$$\begin{aligned}
& \int_0^{\infty} dk_1 dk_2 K(k, k_2) P(k) P(k_2) \delta(k_1 - k_2 - k) \\
&= C^2 \int_0^{\infty} dk_1 dk_2 K(k_1, k_2) \left(\frac{k}{k_1} \right)^{3-2x} (k_1 k_2)^{-x} \delta(k - k_1 - k_2).
\end{aligned} \tag{2.24}$$

またノードの生成を表す項については,

$$kP(k-1) - (k+1)P(k) = C(x-1)k^{-x} \tag{2.25}$$

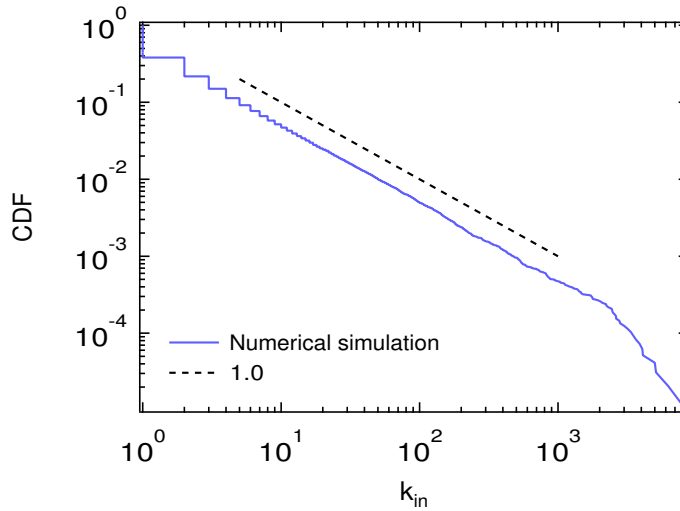


図 2.6 モデルのシミュレーション結果 (青) とマスター方程式の理論解 (黒).

となる。したがって式 (2.19) は、

$$\frac{C(x-1)k^{-x}}{\langle k \rangle + 1} + C^2 k^{3-2x} \int_0^\infty dk_1 dk_2 K(k_1, k_2) (k_1 k_2)^{-x} (k^{2x-3} - k_1^{2x-3} - k_2^{2x-3}) \delta(k - k_1 - k_2) = 0 \quad (2.26)$$

と計算できる。ここで式 (2.26) の積分は $x = 2$ のときに値が 0 になる。それと同時に k のべき指数は、ノードの生成の項が 2、凝集の項が 1 なので、凝集項が支配的であると言えることができる。

以上から $P(k) \sim k^{-2}$ であることがわかった。 $b = c = 0.5$ としてモデルのシミュレーションを行なった結果と、解として得られた指数 2 のべき分布の累積分布関数 (累積では指数が 1) を、図 2.6 に両対数プロットで示す。累積次数分布は理論解と同じく、傾き 1 のべき分布に従うことがわかる。マスター方程式を解く際に、リンクをまとめる確率を考慮しておらず、平均次数 $\langle k \rangle$ も与えていないが、次数分布のべき指数はそれらの要素に依存せず決まる。

2.3.5 平均ノード数が一定の場合 ($b = 0.5, A = 1$)

前項まで、ノードの消滅と生成が起こる場合 ($b = 0.5, c = 0, A = 1$) と、ノードの凝集と生成が起こる場合 ($b = c = 0.5, A = 1$) について、理論的にマスター方程式を解き、シミュレーション結果を示した。これらはモデルの両極端な場合であり、ノードの消滅と凝集の割合が変化することで、モデルは複雑な振る舞いを見せる。そこで本項では平均ノード数が一定の場合 $b = 0.5$ について考え、凝集確率 c を $[0, 0.5]$ で変化させることで、成長モデルの振る舞いを調べる。

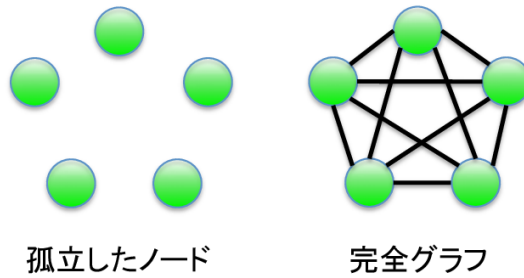


図 2.7 ノードの初期状態.

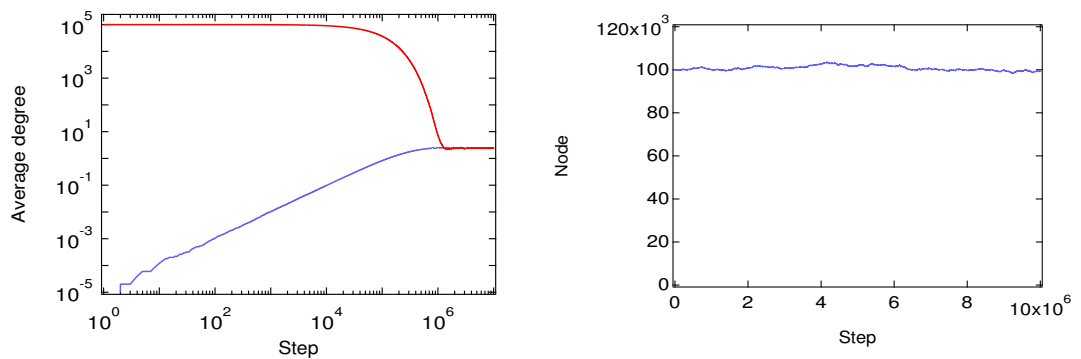


図 2.8 シミュレーション結果の時系列. (左) 平均入次数の時系列. 孤立したノード (青) と完全グラフ (赤) からシミュレーションを開始した結果を示す. (右) ノード数の時系列. ここでは孤立したノード (青) の結果を示しているが, 乱数が同じならば完全グラフでも同一の結果を示す.

ノードの消滅, 生成, 凝集が起こる場合 ($b = 0.5, c = 0.3, A = 1$)

まず, ノードの凝集を考慮した線形な優先的成長モデルの振る舞いを一例として知るために, ノードの消滅確率 $a = 0.2$, 生成確率 $b = 0.5$, 凝集確率 $c = 0.3$, 初期ノード数 $N_0 = 10^5$, シミュレーション時間 $T = 10^7$ として, 成長モデルのシミュレーションを行なった結果を示す. いま, ノードの初期状態としては図 2.7 の孤立したノードの状態と, 完全グラフを用意した.

まず平均入次数の時系列とノード数の時系列を, 図 2.8 に示す. 平均次数の時系列は孤立したノードでは 0 から増加して, 完全グラフでは初期ノード数が 10^5 なので 10^5 から減少して, 同一値に収束する. 収束値の大きさに関しては, 本項で後ほど議論する. ノード数 N の時系列は, いまノード数が増加・減少する確率がどちらも 0.5 なので単純ランダムウォークに従う. しかし, 初期ノード数 N_0 がシミュレーション時間 T に対して $N_0 \gg \sqrt{T}$ なので, N はシミュレーションを通じてほぼ一定となる. ノード数の時系列はシミュレーション時に乱数を固定すれば, リンクの初期状態には依存しない.

図 2.9 はネットワークのノードの年齢分布である. ノードの年齢 τ_i は, ノードがネット

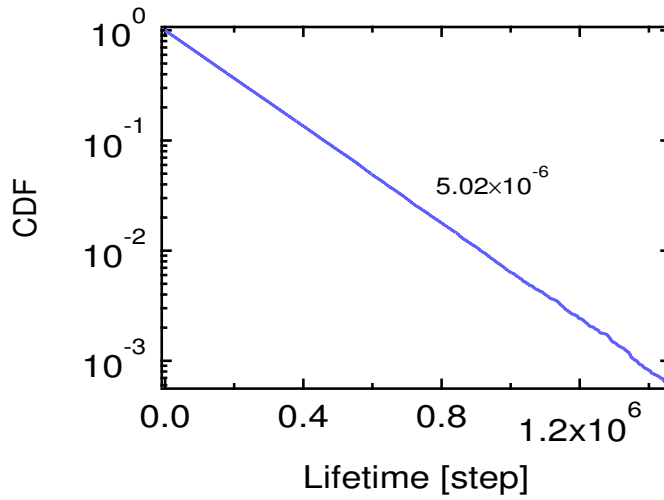
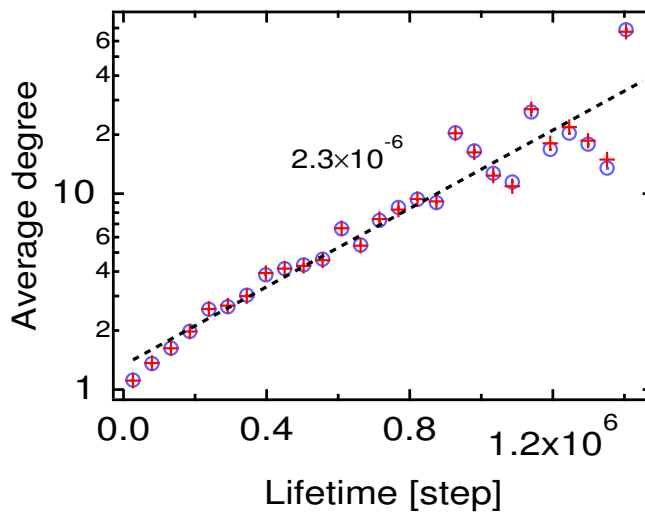


図 2.9 ノードの累積年齢分布.

図 2.10 ノードの年齢と平均入次数 (青), 平均出次数 (赤) の関係. 点線のフィッティングは指数 2.3×10^{-6} の指数関数を示す.

ワークに追加された時刻を t_i とすると,

$$\tau_i = T - t_i \quad (2.27)$$

で定義される. ノード数の時系列同様, ノードの年齢分布はリンクの初期状態に依存しない. Moore モデルや本研究の成長モデルでは, ノードの消滅を次数に関係なくランダムに行なっているので, 年齢分布は指数分布に従う. この指数分布の指数についても本項で, 後ほど定量的に求める.

次にノードの年齢と次数の関係を図 2.10 に片対数プロットで示す. これは年齢が同じノードに対して平均次数を計算したものであり, 平均次数は年齢に対して指数関数的に増加する. また入次数と出次数の成長率について, ほとんどその差異は見られない. 指数関

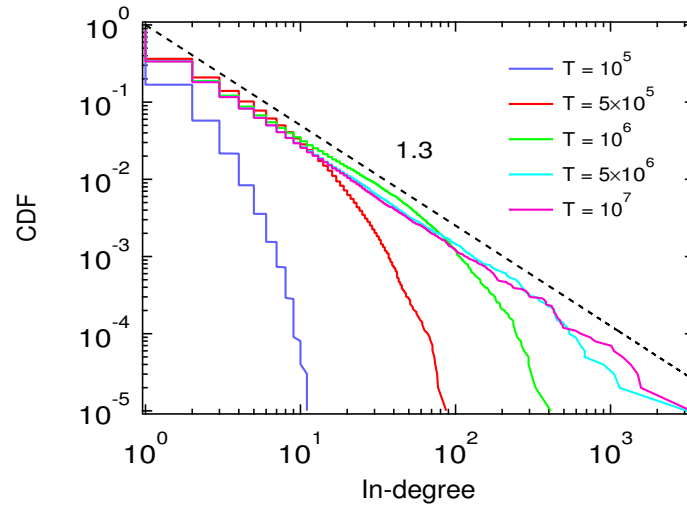


図 2.11 孤立ノードからシミュレーションを開始した入次数分布. それぞれ $t = 10^5$ (青), $t = 5 \times 10^5$ (赤), $t = 10^6$ (緑), $t = 5 \times 10^6$ (水色), $t = 10^7$ (紫) での累積入次数分布を示す. 点線は累積で指数が 1.3 のベキ分布.

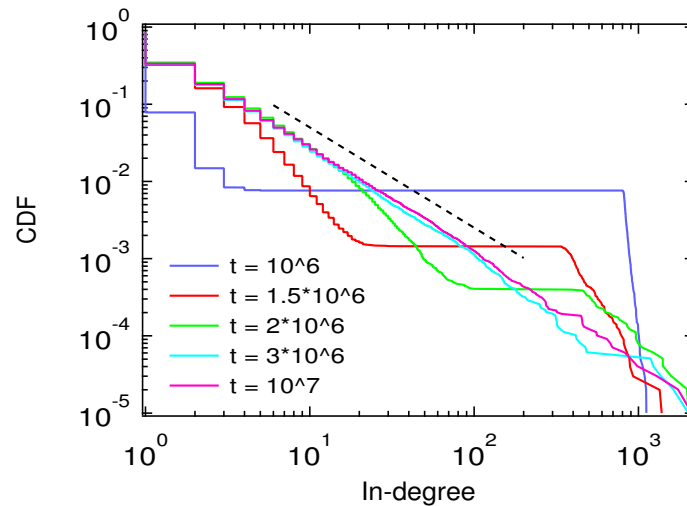


図 2.12 完全グラフからシミュレーションを開始した入次数分布. それぞれ $t = 10^6$ (青), $t = 1.5 \times 10^6$ (赤), $t = 2 \times 10^6$ (緑), $t = 3 \times 10^6$ (水色), $t = 10^7$ (紫) での累積入次数分布を示す. 点線は累積で指数が 1.3 のベキ分布.

数でプロットをフィッティングすると指数が 2.32×10^{-6} となるので, 10,000 ステップ後に次数はおおよそ 1.02 倍になる.

最後に孤立したノードからシミュレーションを開始した入次数分布と, 完全グラフからシミュレーションを開始した入次数分布を, それぞれ図 2.11 と図 2.12 に示す. どちらも累積で指数 1.3 のベキ分布に収束するが, 定常状態に至るまでの過程が異なる. 孤立したノードからシミュレーションを開始すると, ベキ分布のテイルが伸びていき, ハブが成長することでスケールフリーネットワークが達成される. 一方, 完全グラフからシミュレ

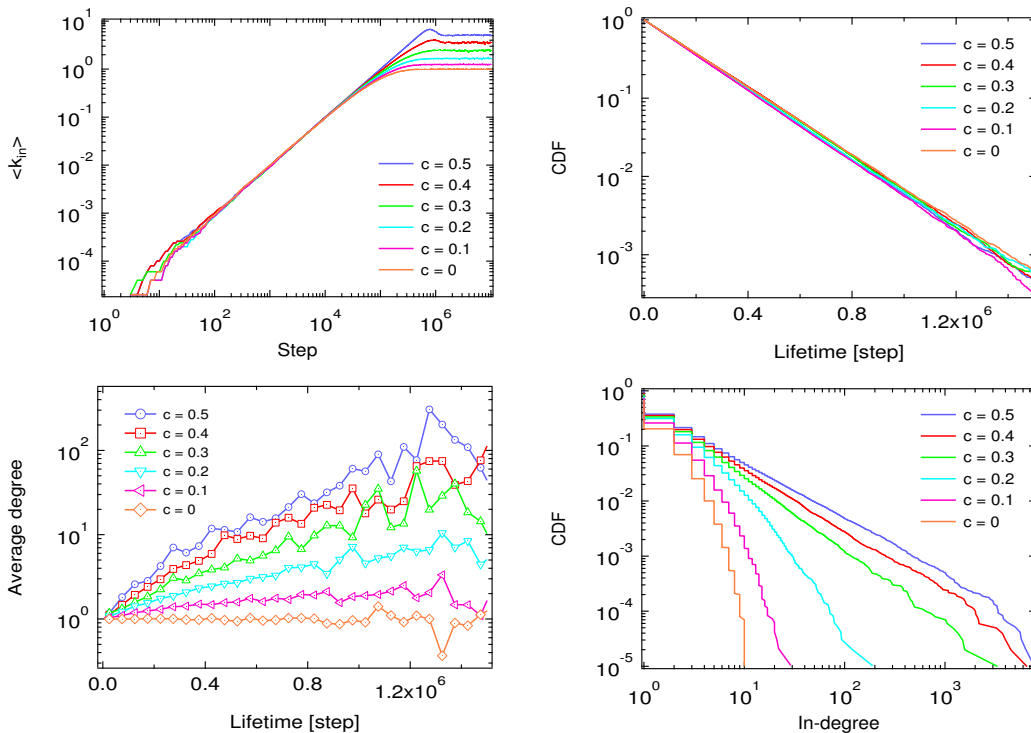


図 2.13 凝集確率 c を変化させた場合のシミュレーション結果. (左上) 平均入次数の時系列. (右上) ノードの年齢分布. (左下) ノードの年齢と平均入次数の関係. (右下) ノードの累積入次数分布. $c = 0.5$ (青), $c = 0.4$ (赤), $c = 0.3$ (緑), $c = 0.2$ (水色), $c = 0.1$ (紫), $c = 0$ (朱色) の場合について示し, 全て $N_0 = 10^5, T = 10^7$ である.

シミュレーションを開始した場合, 最初は全てのノードが入次数 $N_0 - 1$ を持つハブである. そこからノードの消滅によりハブが取り除かれ, 個数が減少する. 同時に隣接ノードがリンクを失うことで, ハブの次数は減少していく. またノードの生成により, ネットワークに次数の小さなノードが追加され, それらの新規ノードが成長していく. すなわち, 次数分布は次数の小さな領域と, 大きな領域の両側からべき分布に漸近していく. 以上がパラメータ $b = 0.5, c = 0.3, A = 1$ の場合のシミュレーション結果である.

凝集確率 c に対するパラメータ依存性

次にノードの凝集確率 c を変化させてシミュレーションを行う. ノードの生成確率を $b = 0.5$ で固定しているので, 凝集確率 $c = 0.5 - a$ を変化させるということは, ノードの消滅と凝集の割合を変化させるということである.

凝集確率を $c = 0, 0.1, 0.2, 0.3, 0.4, 0.5$ (このときそれぞれ $a = 0.5, 0.4, 0.3, 0.2, 0.1, 0$) と変化させたときのシミュレーション結果を図 2.13 に示す. まず平均次数について, ノードの消滅と凝集の割合が異なると収束値が異なる. ノードの凝集時には, リンクの付け替えでリンクが重複しない限り, リンク数は変化しない. 凝集の割合が増加するとノードの消滅の割合が減少するので, リンクの除去数も減少し平均次数は増加する.

年齢分布については, ノードの凝集と消滅時にノードが 1 つ減少するので, 両者の割合

が変化してもノード数は変化しない。したがって年齢分布もまた、消滅と凝集の割合によらず同じ分布を示す。年齢と次数の関係は、凝集確率が高いほどリンクをまとめて得る確率が高く、次数の成長が速い。次数の成長を指数関数的に増加させるために、成長モデルに凝集効果を取り入れたので、これはほぼ自明である。一方、凝集確率 $c = 0$ の場合（このとき消滅確率 $a = 0.5$ 、生成確率 $b = 0.5$ ）は、次数は年齢に対して成長せず一定のままである。新規ノードがネットワークに生成され、リンクが接続されることによる次数の増加と、隣接ノードが消滅しリンクが1本消えることによる次数の減少が、ちょうど釣り合っている。

最後に次数分布を図 2.13 の右下に示す。凝集確率が 0.5 から減少していくことで、次数分布はベキ分布から指数分布へと変化していく。またノードの凝集でリンクを付け替える際に、リンクが重複した場合には1本にまとめているため、凝集確率が大きい場合にはリンクをまとめる回数も多くなる。したがって次数分布にはカットオフが生まれ、次数が一定以上の大きさのノードは生成しにくいという効果を生み出している。

図 2.13 の右下の図から、次数分布がベキ分布に従う場合に、ベキ指数は凝集確率 c に依存していることがわかる。ある物理量がベキ分布に従うとき、ベキ指数は最尤法や最小二乗法を用いて推定されるが、ベキ指数はベキ分布に従う領域の下限値をいくつに設定するか大きく依存する。真の値よりも低くベキ分布の下限値を見積もると、ベキ指数は実際の値よりも小さくなり、高く見積もるとベキ分布に従うデータ数が少なくなり、指数推定の精度が低くなる。ベキ指数を推定する場合には、ベキ指数とベキ分布の下限値を同時に推定しなければならない。このベキ指数とベキ分布の下限値を同時に推定する方法はいくつか知られている [58–61] が、ここでは Clementi ら [58] の方法を用いてベキ指数の推定を行う。この方法はブートストラップ法を用いて最尤法でベキ指数を推定し、ベキ指数の平均二乗誤差が最も小さくなる値をベキ分布の下限値として採用する方法である。

いま凝集確率 c を $[0, 0.5]$ で変化させてシミュレーションを行い、得られた入次数分布に対してブートストラップの回数を 10,000 としてベキ分布の下限値を求め、その際のベキ指数 γ を推定する。この結果を図 2.14 に示す。 c が 0 に近い領域ではそもそも次数分布は指数分布に従うので、この方法で得られたベキ指数は参考にならない。しかし $c > 0.2$ ではベキ指数は 2 から減少し、 $c \rightarrow 0.5$ で $\gamma \rightarrow 1$ に収束していくことがわかる。ここでは累積分布の指数を推定しているので、実際の密度関数 $P(k)$ のベキ指数はこのとき、3 から 2 へと変化する。

平均次数の時間発展

平均入次数 $\langle k_{in} \rangle$ の時間発展について、1ステップでのリンク数の増減を考慮することにより求める。ネットワークのリンク数は、ノード数 N と平均入次数 $\langle k_{in} \rangle$ を用いて、 $N\langle k_{in} \rangle$ と表すことができる。ノードの消滅時（確率 a ）にリンクは $\langle k_{in} \rangle + \langle k_{out} \rangle = 2\langle k_{in} \rangle$ 本だけ消滅する。またノードの生成時（確率 b ）に $k_{in} = k_{out} = 1$ のノードをネットワークに追加するので、リンクは 2 本生成される。ノードの凝集時（確率 c ）には $\langle k_{in} \rangle + \langle k_{out} \rangle = 2\langle k_{in} \rangle$ 本

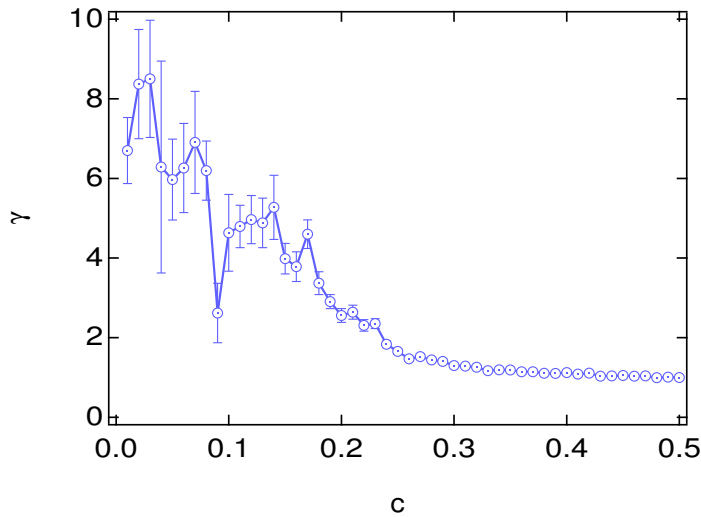


図 2.14 凝集確率 c と入次数分布のべき指数 γ の関係. エラーバーはべき指数 γ の標準偏差を表す.

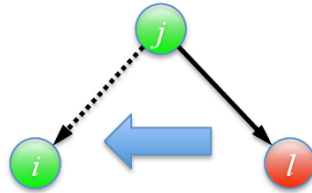


図 2.15 リンクが重複する場合の概念図. いまノードの凝集時に, 最初にランダムに選択したノードを l , 優先的接続で選択したノードを i とする. この際にリンク l_j をリンク l_i に付け替えるが, 既にリンク l_j が存在し重複する場合にはまとめて 1 本とする. リンクをまとめる確率 $p_n(k)$ とはすなわち, 入次数 k のノード i と出次数 k_j のノード j 間にリンクが存在する確率である.

のリンクを優先的接続により確率 $\frac{k_{i,\text{in}}+k_{i,\text{out}}+1}{\sum_j(k_{j,\text{in}}+k_{j,\text{out}}+1)} \simeq \frac{2k+1}{\sum_j(2j+1)}$ で入次数 k のノードに付け替えるが, このうち重複してリンクをまとめる確率を $p_n(k)$ とする. 以上より, 1 ステップでのリンク数の増減は次のようになる.

$$N \frac{d\langle k_{\text{in}} \rangle}{dt} = -2a\langle k_{\text{in}} \rangle + 2b - 2c\langle k_{\text{in}} \rangle \sum_k p_n(k) \frac{2k+1}{\sum_j(2j+1)} NP(k). \quad (2.28)$$

ここで $NP(k)$ は次数 k となるノードの数である.

リンクを付け替える際に重複したリンクをまとめる確率 $p_n(k)$ とは, 出次数 k_j と入次数 k のノード間に, リンクが存在する確率と解釈することができる. この考え方を図 2.15 に示す. リンクをまとめる確率 $p_n(k)$ は, リンク数が $N\langle k_{\text{in}} \rangle$ なので,

$$p_n(k) = \frac{kk_j}{N\langle k_{\text{in}} \rangle} \quad (2.29)$$

と考えられる. 式 (2.29) を確かめるために, $c = 0.3, N_0 = 10^5, T = 10^7$ としてモデルのシミュレーションを行い, ネットワークを生成する. 生成したネットワークから, 出次数 k_j

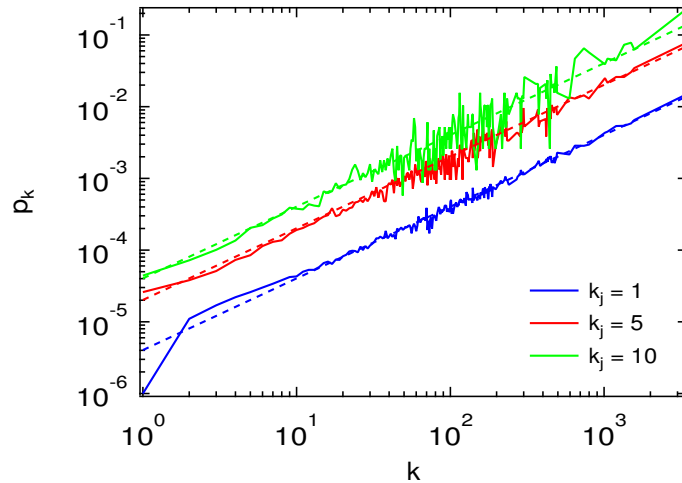


図 2.16 出次数 k_j と入次数 k のノード間にリンクが存在する確率 $p_n(k)$. $k_j = 1$ (青), $k_j = 5$ (赤), $k_j = 10$ (緑) の場合について示す. 点線のフィッティングは式 (2.29) による.

と入次数 k のノード間にリンクが張られている確率 $p_n(k)$ を測定し, この結果を図 2.16 に示す. 実線がシミュレーション結果, 点線が式 (2.29) によるフィッティングである. 隣接ノードの出次数 k_j が比較的小さな値ならば, 式 (2.29) による近似は有効である. しかし例えば $k_j = N - 1$ でノード j が全てのノードと接続しているような場合, $p_n(k)$ は k の値によらず 1 となるはずである. このように k_j が大きな値の場合は, 式 (2.29) は $p_n(k) > 1$ となってしまう, 近似が成り立たない.

$k_j = \langle k_{in} \rangle$ とし, 式 (2.29) を微分方程式 (2.28) に代入し, $b = 1/2, a = 1/2 - c$ として整理すると,

$$N \frac{d\langle k_{in} \rangle}{dt} = -(1 - 2c)\langle k_{in} \rangle + 1 - 2c\langle k_{in} \rangle \frac{2\langle k_{in}^2 \rangle + \langle k_{in} \rangle}{(2\langle k_{in} \rangle + 1)N} \quad (2.30)$$

となる. これは $\langle k_{in}^2 \rangle$ について不明なため, 解くことができない.

いまノードの凝集時にリンクをまとめる確率 $p_n(k)$ を無視し, $t = 0$ で $\langle k_{in} \rangle = 0$ (孤立したノードの状態) として方程式 (2.30) を解くと,

$$\langle k_{in} \rangle(t) = \frac{1}{1 - 2c} \left[1 - \exp\left(-\frac{1 - 2c}{N}t\right) \right] \quad (2.31)$$

となり, $1/(1 - 2c)$ という値に収束することがわかる. すなわち定常状態においてネットワークの平均次数は凝集確率 c にのみ依存し, ノード数 N には依存しない. ノード数が大きいと, 平均次数が収束するまでの時間が長くなる.

平均入次数 $\langle k_{in} \rangle$ の時系列を図 2.17 に, 凝集パラメータ c を変化させた時の平均入次数 $\langle k_{in} \rangle$ と $1/(1 - 2c)$ を図 2.18 に示す. 図 2.17 では, 実線でシミュレーション結果を, 点線で式 (2.31) による解を $c = 0.4$ まで表示している. また図 2.18 では, シミュレーションで得られた平均入次数 $\langle k_{in} \rangle$ を青で, 理論値 $1/(1 - 2c)$ を黒の点線で示している. これらの図より $c > 0.3$ の領域では, ノードの凝集時にリンクをまとめる確率 $p_n(k)$ が無視できな

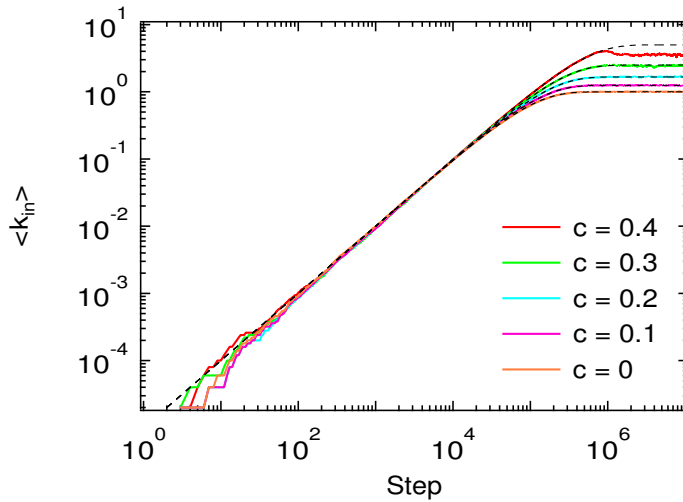


図 2.17 平均入次数 $\langle k_{in} \rangle$ の時系列. $c = 0.4$ (赤), $c = 0.3$ (緑), $c = 0.2$ (水色), $c = 0.1$ (紫), $c = 0$ (朱色) の場合について示す. 図 2.13(左上) とほぼ同じ図だが, 式 (2.31) でフィッティング (点線) している.

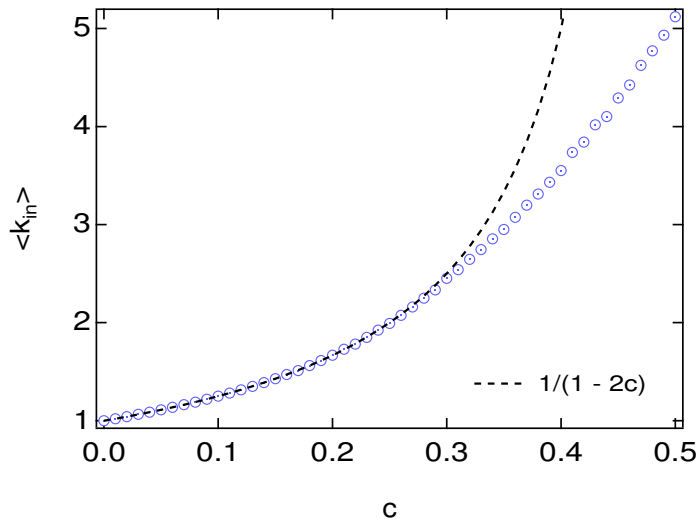


図 2.18 凝集確率 c と平均入次数 $\langle k_{in} \rangle$ (青), 理論値 $1/(1-2c)$ (黒) の関係.

くなり, $\langle k_{in} \rangle$ は理論値よりも小さな値を取ることがわかる.

ノードの年齢分布

ノードの年齢 (ネットワークに追加されてからの経過時間) を τ とおく. いまノードは, 消滅 (確率 a) か凝集 (確率 c) 時に 1 つランダムに選ばれて消滅し, それ以外の場合には τ が 1 増加する. したがって年齢の確率分布 $P(\tau)$ は,

$$P(\tau) = \frac{a+c}{N} \left(1 - \frac{a+c}{N}\right)^{\tau-1} \quad (2.32)$$

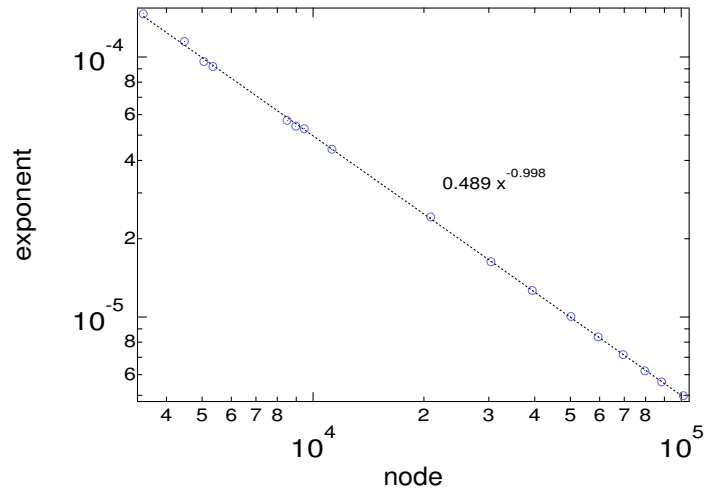


図 2.19 ノード数 N と指数分布でフィッティングした年齢分布の指数の関係.

に従う. ここで $a+c=1-b$ でもある. 上式よりノード数が十分に大きい時, ノードの年齢分布は指数 $(a+c)/N$ の指数分布となる.

$$P(\tau) \sim e^{-\frac{a+c}{N}\tau}. \quad (2.33)$$

例えばノードの生成確率 $b=0.5$, ノード数 $N=10^5$ ならば, 年齢分布の指数は 5×10^{-6} となり, 図 2.9 の指数と一致する. 誤差は $t=T$ でのノード数 N が, 初期ノード数 N_0 とは異なることによるものである.

平均ノード数が一定の場合 $b=a+c=0.5$ となるので, 年齢分布の指数はノード数の逆数に比例する. 図 2.19 はノード数を変えてシミュレーションを行い, 年齢分布を指数分布でフィッティングした時の指数を, ノード数に対してプロットしたものである. いま指数は $0.5/N$ におおよそ従い, 式 (2.33) の結果は正しい.

マスター方程式の数値計算解

今までと同様に, 式 (2.4) よりマスター方程式は,

$$\begin{aligned} P(k, t+1) - P(k, t) = & a \left[\frac{k+1}{N} P(k+1, t) - \left(\frac{k}{N} + \frac{1}{N} \right) P(k, t) \right] \\ & + b \left[\frac{(k-1)+1}{\sum_j (k_j+1)} P(k-1, t) - \frac{k+1}{\sum_j (k_j+1)} P(k, t) + \frac{\delta_{k,1}}{N} \right] \\ & + c \left[\sum_{k'=0}^k P(k-k', t) \frac{2k'+1}{\sum_j (2k_j+1)} P(k', t) - \sum_{k'=0}^{\infty} P(k, t) \frac{2k'+1}{\sum_j (2k_j+1)} P(k', t) \right. \\ & \left. - \sum_{k'=0}^{\infty} P(k', t) \frac{2k+1}{\sum_j (2k_j+1)} P(k, t) \right]. \end{aligned} \quad (2.34)$$

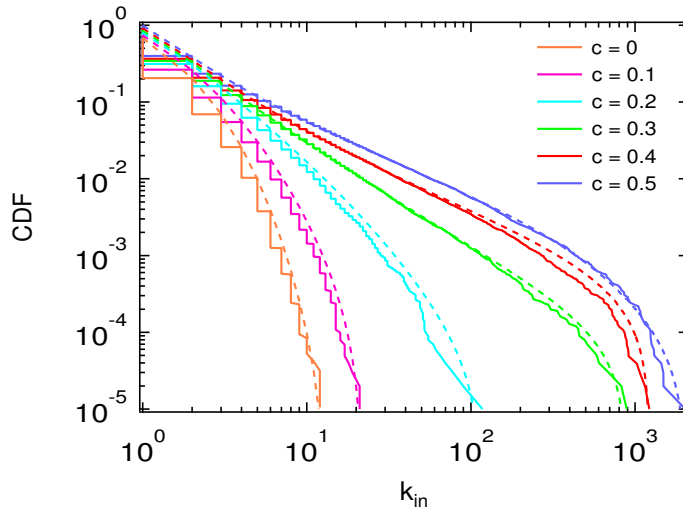


図 2.20 モデルのシミュレーション結果 (実線) とマスター方程式の数値計算解 (点線) による累積入次数分布. シミュレーションは $N_0 = 10^4, T = 10^6$ として行なった 10 回のアンサンブル平均であり, $c = 0.5$ (青), $c = 0.4$ (赤), $c = 0.3$ (緑), $c = 0.2$ (水色), $c = 0.1$ (紫), $c = 0$ (朱色) である.

既存研究でも粒子の注入, 凝集, 消滅を同時に考慮したスモルコフスキー方程式は理論的に解かれていない. そこで式 (2.34) を数値的に解くことを考える.

このマスター方程式に $b = 1/2, a = 1/2 - c, k_j = \langle k \rangle$ を代入して整理する. 定常状態において,

$$(1 - 2c)(k + 1)(P(k + 1) - P(k)) + \frac{kP(k - 1) - (k + 1)P(k)}{\langle k \rangle + 1} + \delta_{k,1} + \frac{2c}{2\langle k \rangle + 1} \left[\sum_{k'=0}^k (2k' + 1)P(k - k')P(k') - 2(k + \langle k \rangle + 1)P(k) \right] = 0 \quad (2.35)$$

ここで平均次数 $\langle k \rangle$ は式 (2.31) から求まるが, この式はノードの凝集時にリンクをまとめる確率 p_n を全く考慮していないので, c の値が大きいと適用することができない. いま $\langle k \rangle$ と最大次数 k_{\max} を, モデルのシミュレーション結果から得る. 境界条件として $P(k_{\max} + 1) = 0$ とし, シミュレーションから得た $\langle k \rangle$ を用いて, 方程式 (2.35) を $\{P(0), P(1), \dots, P(k_{\max})\}$ について連立させて解く.

初期状態を $P(0) = 1$ (孤立したノードの状態) とし, 平均次数 $\langle k \rangle$ を初期ノード数 $N_0 = 10^4$, シミュレーション時間 $T = 10^6$ として求めた時の数値計算解を, 10 回のアンサンブル平均として図 2.20 に示す. これはやはり前項のときと同様, マスター方程式におけるノードの凝集で, リンクをまとめる確率を全く考慮していない. しかし $P(k_{\max} + 1) = 0$ という境界条件を与えることで, マスター方程式の数値計算解は実際のシミュレーション結果を再現できることがわかる.

結局のところ, ノードの凝集時にリンクをまとめるという行為は, 次数分布のベキ指数には影響しない. ただ重複したリンクをまとめることで, ある一定以上の次数のノードが

生まれにくくなり、次数分布にはカットオフが生じる。本項ではこのリンクをまとめることによるカットオフの効果を、最大次数 k_{\max} を用い $P(k_{\max} + 1) = 0$ という境界条件として、別の形で取り入れている。

2.3.6 その他のパラメータ依存性

本項では、パラメータを様々な値に設定してシミュレーションを行い、モデルの特性について調べる。まず生成するネットワークの平均ノード数を一定 $b = 0.5$ とした上で、初期ノード数 N_0 依存性について調べる。次に、ノード数が増加しネットワークが成長する場合 ($b > 0.5$) と、ノード数が減少しネットワークが衰退する場合 ($b < 0.5$) について調べる。

初期ノード数 N_0 に対するパラメータ依存性

成長モデルで生成するネットワークのサイズと、そのネットワークの性質の関係を知りたい場合、ノード数 N を変化させてネットワークの性質を観測するのが妥当である。しかし、ノード数 N は単純ランダムウォークに従うため、コントロールするのが難しい。そこで初期ノード数 N_0 を変化させてシミュレーションを行う。その他のパラメータは $b = 0.5, c = 0.3$ である。全てのシミュレーションにおいて $T = 10^6$ としているため、ノード数の揺らぎとしては $N \approx N_0 \pm 10^3$ 程度である。

初期ノード数を $N_0 = 3 \times 10^3, 5 \times 10^3, 10^4, 5 \times 10^4$ と変化させたときのシミュレーション結果を図 2.21 に示す。平均次数の時系列は、シミュレーションの開始時点では乱数の影響を受けやすい。しかしノード数に依存せず同じ値に収束し、異なるのは収束するまでのステップ数(時間)である。初期ノード数が多くなるほど、シミュレーションが収束するまでの時間がかかる。

年齢分布と、年齢と次数の関係は、初期ノード数 N_0 によって大きく異なる。ネットワークのノード数が増加すると、あるノードがランダムに選択され消滅する確率 $1/N$ は低くなるが、逆にリンクを新たに得る確率も低くなる。そのため年齢分布の指数はノード数に対して減少し、年齢に対する次数の増加も遅くなる。図 2.21 の左下の図で、 $N = 3000$ や $N = 5000$ で年齢の長いノードは数が少ないため、年齢と次数の関係のプロットはその右端の領域で揺らぎが大きい。

最後に次数分布を図 2.21 の右下に示す。次数分布は全て同じ指数 1.3 のべき分布に従い、サイズが増加するとべき分布のテイルが伸びる。実はモデルにおいてノード数 N を変化させても、次数分布 $P(k)$ に影響はなく、時間 t が異なるだけである。式 (2.34) の簡

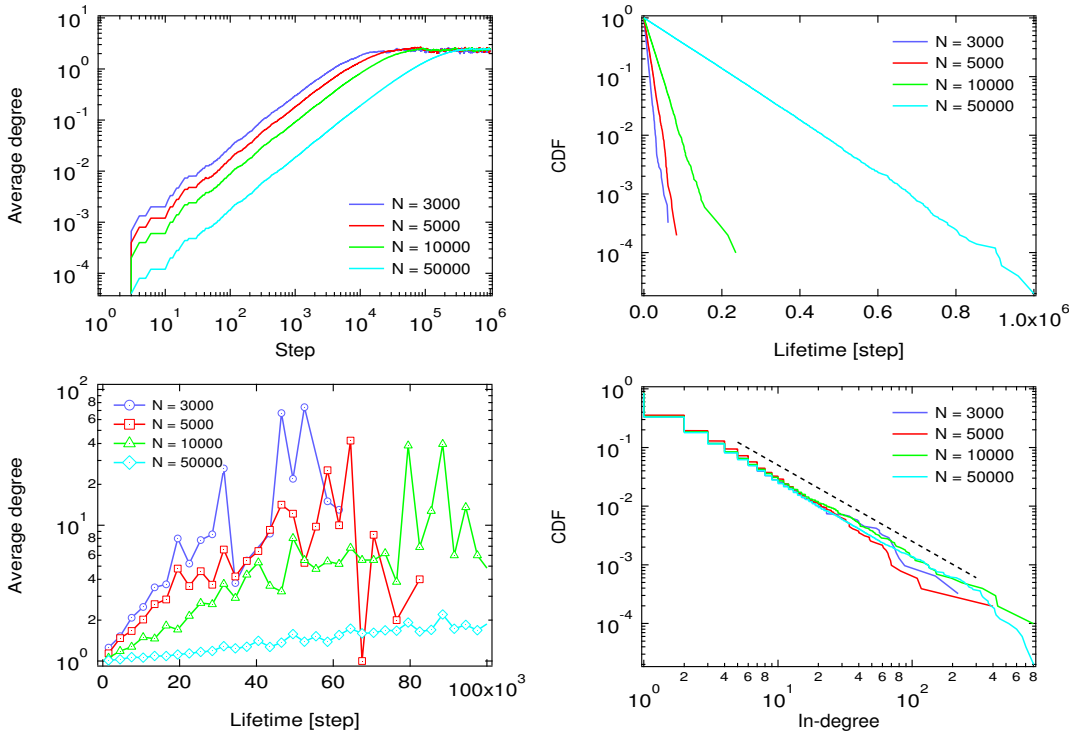


図 2.21 初期ノード数 N_0 を変化させた場合のシミュレーション結果. (左上) 平均入次数の時系列. (右上) ノードの年齢分布. (左下) ノードの年齢と平均入次数の関係. (右下) ノードの累積入次数分布. フィッティングは累積で指数 1.3 のべき分布を表す. $N_0 = 3 \times 10^3$ (青), $N_0 = 5 \times 10^3$ (赤), $N_0 = 10^4$ (緑), $N_0 = 5 \times 10^4$ (水色) の場合について示し, 全て $c = 0.3, T = 10^6$ である.

単な変形により,

$$\begin{aligned}
 N\dot{P}(k) = & a(k+1)(P(k+1) - P(k)) + b \left[\frac{k}{\langle k \rangle + 1} P(k-1) - \frac{k+1}{\langle k \rangle + 1} P(k) + \delta_{k,1} \right] \\
 & + c \left[\sum_{k'=0}^k P(k-k') \frac{2k'+1}{2\langle k \rangle + 1} P(k') - \sum_{k'=0}^{\infty} P(k) \frac{2k'+1}{2\langle k \rangle + 1} P(k') \right. \\
 & \left. - \sum_{k'=0}^{\infty} P(k') \frac{2k+1}{2\langle k \rangle + 1} P(k) \right]. \quad (2.36)
 \end{aligned}$$

となり, システムサイズに関する変数であるノード数 N は左辺にしか現れない. そこで $t \rightarrow Nt$ という変数変換を行うことにより, マスター方程式から N を消去することができる. すなわちノード数 N はシミュレーション時間にしか影響せず, 次数分布には影響を与えない. ノード数 N は式 (2.9) によって決まるため, 初期ノード数 N_0 , ノードの生成確率 b を変化させることによってノード数 N は変化するが, それらは次数分布には影響しない. そのため, 初期ノード数 N_0 , ノードの生成確率 b の値を変化させてシミュレーションを行っても, 同じような結果が得られる.

初期ノード数 N_0 やノード数 N は, どこまで小さな値でシミュレーションをできるだろうか. さらに $N_0 = 500, 600, 700, 800, 900, 1000$ の場合について $T = 10^6$ のシミュレー

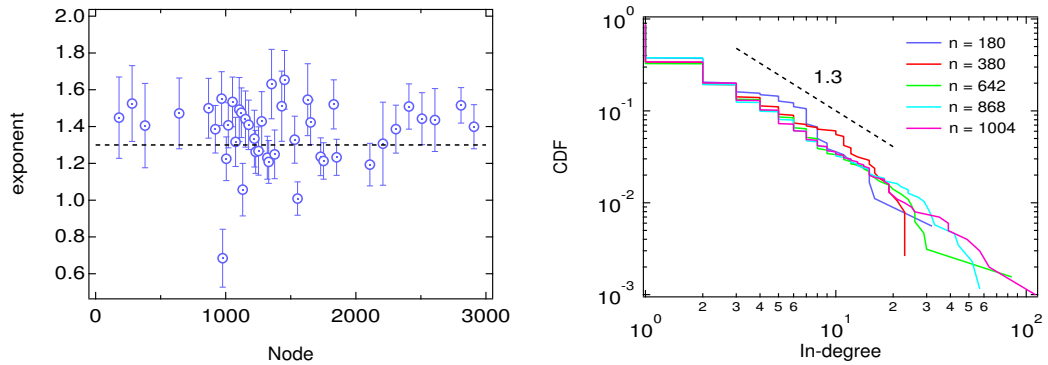


図 2.22 ノード数 N が小さい場合のシミュレーション結果. (左) ノード数 N と度数分布のベキ指数 (累積) の関係. 値はおよそ 1.3 を取るはずであるが, その付近に分布していることがわかる. (右) 各ノード数における度数分布.

シミュレーションを行い, シミュレーション終了後のノード数 N を記録する. また前述の Clementi ら [58] の方法を用いて, 生成したネットワーク度数分布のベキ指数を測定する. このノード数に対するベキ指数の値を, 図 2.22 に示す. $b = 0.5, c = 0.3$ のとき, 度数分布が累積でおよそ 1.3 のベキ分布に従うことを既に確認した. 図 2.22 の左を見ると, ノード数がいくつであっても, およそ 1.3 の付近でベキ指数が分布していることがわかる. また右図より, そのとき度数分布が実際にノード数によらず, ベキ分布に従うことがわかる. ノード数が 100 程度のネットワークであっても, その度数分布はベキ分布に従う.

生成確率 b を変化させた場合

次に, ノードの生成確率 b を変化させてシミュレーションを行う. 初期ノード数 $N_0 = 10^5$, シミュレーション時間 $T = 2 \times 10^6$ とし, この間のノード数の増減が $N - N_0 = -5 \times 10^4, -2 \times 10^4, 0, 2 \times 10^4, 5 \times 10^4$ となるように設定してシミュレーションをおこなう. 式 (2.9) より, それぞれ $b = 0.4875, 0.495, 0.5, 0.505, 0.5125$ となる. また消滅確率 a と凝集確率 c は, ベキ指数が累積で 1.3 のスケールフリーネットワークを生成するように, $a : c = 2 : 3$ として決める.

まずネットワークのノード数と, 平均入次数の時系列を図 2.23 に示す. ノード数の時系列から, シミュレーション時間 T の間に, $N - N_0$ が設定した通りに $\pm 2 \times 10^4, \pm 5 \times 10^4$ だけ変化していることがわかる. 一方, 平均入次数の時系列はノードの生成確率を変化させてもほぼ変化しない.

次に年齢分布と年齢と入次数の関係を図 2.24 に示す. 年齢分布では $b > 0.5$ で成長するネットワークだと, $b = 0.5$ の平均ノード数一定のネットワークよりも, ノードの年齢が長くなる. これはノード数が増加し, ノードの消滅確率が低下していることに由来する. 逆に $b < 0.5$ で衰退するネットワークだとノードの年齢は短くなる. 一方, 年齢と入次数の関係では $b > 0.5$ だと新たにリンクを得る確率が低くなるので, 入次数の成長は遅くなる. $b < 0.5$ だとノード数は減少し続けるので, 新たにリンクを得る確率は高くなり, 入次数の

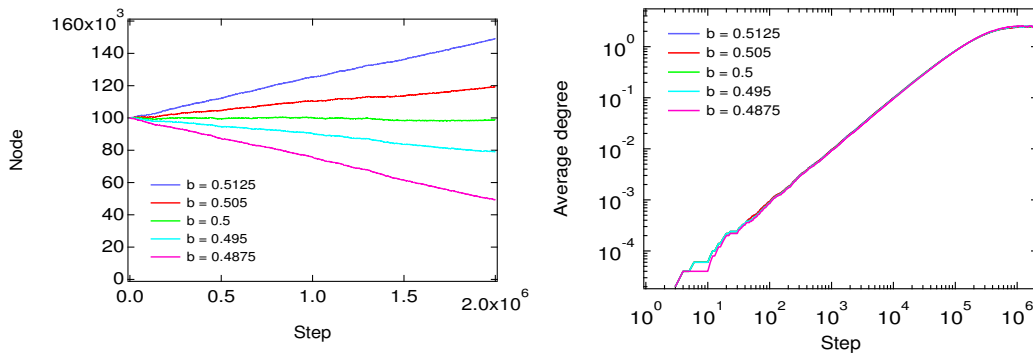


図 2.23 生成確率 b を変化した場合の時間発展. (左) ノード数と (右) 平均入次数の時系列を $b = 0.5125$ (青), $b = 0.505$ (赤), $b = 0.5$ (緑), $b = 0.495$ (水色), $b = 0.4875$ (紫) の場合についてそれぞれ示す.

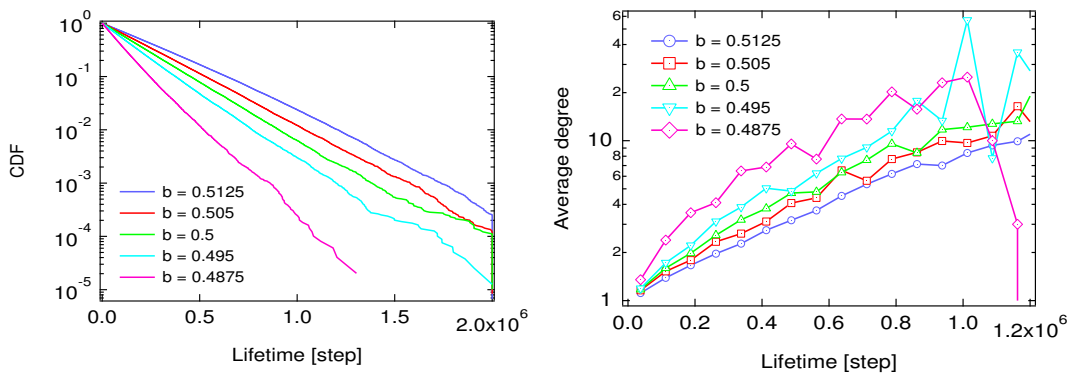


図 2.24 生成確率 b を変化した場合の (左) 累積年齢分布と (右) 年齢と平均入次数の関係. $b = 0.5125$ (青), $b = 0.505$ (赤), $b = 0.5$ (緑), $b = 0.495$ (水色), $b = 0.4875$ (紫) の場合についてそれぞれ示す.

成長が速くなる.

最後に、入次数の累積分布を図 2.25 に示す. 次数分布はどのネットワークでもノードの生成確率に依存せず、累積で指数 1.3 のべき分布に従う. 生成確率 b を変化させることで、ノード数が増加、減少し続ける. この場合のシミュレーション結果は、初期ノード数 N_0 を変化した場合のシミュレーション結果とほぼ同じである.

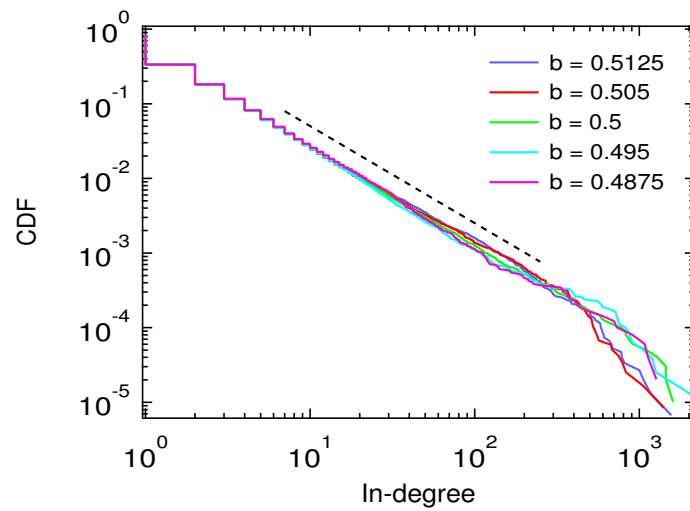


図 2.25 生成確率 b を変化させた場合の入次数の累積分布. $b = 0.5125$ (青), $b = 0.505$ (赤), $b = 0.5$ (緑), $b = 0.495$ (水色), $b = 0.4875$ (紫) の場合についてそれぞれ示す. フィッティングは累積で指数 1.3 のべき分布を表す.

第3章

複雑ネットワークの成長モデルの 応用

この章では、本研究で構築した複雑ネットワークの成長モデルを、現実のネットワークに応用することを考える。企業間取引ネットワークをその一例として挙げ、2章で構築したモデルのシミュレーション結果を、企業間取引ネットワークの統計的性質と比較する。

本章ではまず企業間取引ネットワークの既存研究について簡単に紹介する。企業間取引ネットワークは日本企業約100万社の仕入・販売先データを用いて構築し、構築したネットワークの性質について解析を行い、得られた結果について示す。また現実の企業間取引ネットワークにおいて優先的接続の確率の指数の測定を行う。

2章で構築したモデルを企業間取引ネットワークに応用するには、不十分な面が多い。隣接ノードにあたる取引先企業だけではなく、2リンク先の企業との関係性についても解析を行う。企業間取引ネットワークの2リンク構造の性質を成長モデルで再現することを試み、モデルの再現度を高める。

3.1 企業間取引ネットワークの既存研究

企業間取引ネットワークではノードが企業、リンクが企業間の取引によるお金の流れを表し、とりわけ東京商工リサーチ社調査の企業データを用いて、これまでも様々な研究がなされてきた [62–72]。企業間取引ネットワークを研究することにより、複雑ネットワークの構造や統計的性質だけではなく、マクロな経済現象をミクロなリンク構造から理解できる。Matteo ら [62] は、労働生産性と資本生産性の分布がイタリアとフランスでベキ分布に従っていることを観測した。彼らはこの性質を再現するために、企業間取引ネットワーク上でイノベーションが伝搬するという輸送モデルを導入し、ネットワークがスケールフリーであるとき、生産性の分布がベキ分布に従うことを発見した。実際に企業間取引ネットワークはスケールフリーネットワークであることが高安ら [64] によって確認されている。齋藤ら [63] は企業の売上と入次数が線形比例しないことを発見した。飯野ら [68] は企業間取引ネットワークのコミュニティ分析を行い、貪欲法などを用いてモ

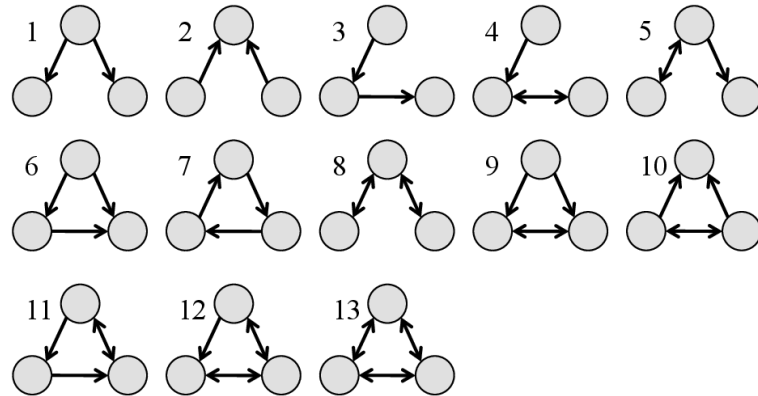


図 3.1 モチーフ解析で考慮される 3 ノードから構成される部分グラフ.

ジュラリティの最適化を行なった。抽出されたコミュニティを分子動力学の方法を用いて表示することにより、業種や地域などで同じコミュニティがクラスターとして近くに配置されやすいことを確認している。藤原と青山 [69] は連鎖倒産の影響を考慮し、その連鎖倒産するクラスターのサイズを、企業間取引ネットワーク上で連鎖倒産のシミュレーションを行い、また理論から示した。渡邊ら [71] は、企業同士の取引によるお金のやり取りを考慮し、企業間取引ネットワーク上でのお金の輸送により企業の売上を再現することを試みた。田村ら [70] によってその後、企業間の取引金額は両社の売上の積に非線形比例することがわかり、この複雑ネットワーク上の輸送モデルは拡張されようとしている。

3.1.1 モチーフ解析

大西ら [67] は企業間取引ネットワークのミクロな構造を調べるために、ネットワークのモチーフ解析 [73, 74] を行なった。モチーフ解析とは、無向ネットワークにおけるクラスター係数を有向ネットワークで一般化したものであり、ネットワークにどのような三角形のリンク構造が出現しやすいかを測定するものである。ここで考慮する三角形のリンク構造とは図 3.1 に示すようなものである。重要なモチーフを検出するために、各モチーフの出現回数はランダム化したネットワーク [75] と比較される。ランダム化したネットワークとは、ネットワークの次数分布を保存しながらリンクをランダムに抽出し、 $A \rightarrow B, C \rightarrow D$ という 2 本のリンクがある時に、 $A \rightarrow D, C \rightarrow B$ というリンクの付け替えを繰り返したものである。

各モチーフの統計的有意性を評価するために、ランダム化したネットワークを多数用意し、次の Z スコアを計算する。

$$Z_i = \frac{N_i - \langle N_i^{\text{rand}} \rangle}{\sqrt{\langle (N_i^{\text{rand}} - \langle N_i^{\text{rand}} \rangle)^2 \rangle}}. \quad (3.1)$$

ここで N_i はモチーフ i の出現回数、 N_i^{rand} は次数分布を保存してランダム化したネットワークでのモチーフ i の出現回数である。分母はランダム化したネットワークにおける

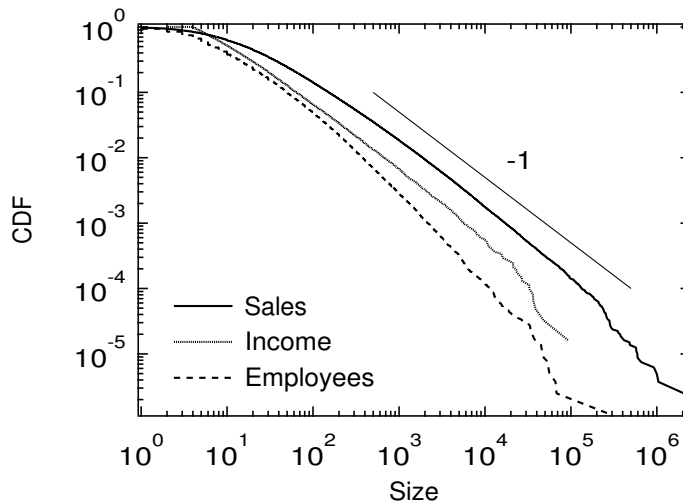


図 3.2 企業の売上 (Sales), 所得 (Income), 従業員数 (Employees) の累積分布. 売上と所得の単位は百万円.

モチーフ i の出現回数の標準偏差であり, Z スコアは観測対象のネットワークでモチーフ i が有意に多く出現するかどうかを評価する値である.

前述の大西らは, 図 3.1 の 1-3 の V 字型のモチーフと 13 のクリーク (完全グラフ) が, 企業間取引ネットワークに有意に多く出現することを発見した. 一方 6 や 7 のループ構造は観測されず, お金が 3 社で循環するループ構造よりも, V 字型やクリーク構造が企業間取引ネットワークでは構築されやすいと結論づけている.

以上のようにモチーフ解析を含め, 企業間取引ネットワークについてそのネットワークの構造や, ネットワーク上で起こる現象について様々な議論がなされているが, ネットワークのスケールフリー性を説明するモデルや理論は存在しない. そこで 2 章で構築した成長モデルを企業間取引ネットワークに応用し, スケールフリー性などの性質を説明することをこの章では目的とする.

3.2 解析に用いるデータの説明

本研究で用いるデータは, 経済産業研究所によって提供された, 東京商工リサーチ社の 2000 年から 2005 年の調査による企業データである. このデータは企業の売上や利益, 資本金, 従業員数, 工場数, 事業所数, 業種, 取り扱い品目など約 100 項目が記載され, 日本企業のほぼ全ての主要企業を含む. しかしデータに外国企業や金融機関, 政府機関などは含まれていない. このデータにおいて企業の規模を示す, 売上や所得, 従業員数などはベキ分布に従うことが知られており [64, 66], これを図 3.2 に示す. 企業の売上は累積で指数が 1 の, 従業員数は累積で指数が 1.3 のベキ分布に従い, 米国や英国で観測されている結果 [76, 77] と一致する.

本研究で用いたのは, 財務情報を示すデータではなく, 日本企業 961,318 社の取引関係

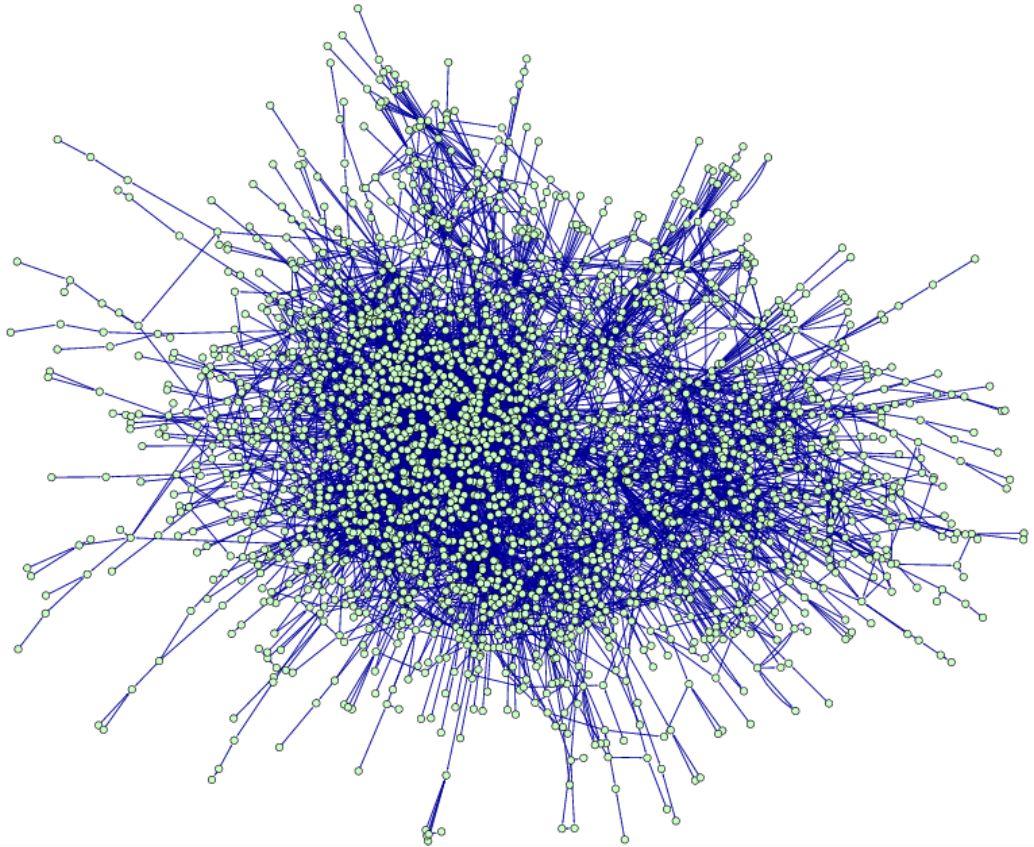


図 3.3 佐賀県内の取引から構成される企業間取引ネットワーク。

データ(仕入・販売)である。これは調査対象の企業に、主な仕入先と販売先を 24 社まで挙げてもらったものであり、各企業がどの企業からお金をもらい、どの企業にお金を払う取引をしているかということがわかる。東京商工リサーチ社の取引関係データには、各企業が仕入先と販売先として挙げられる企業数に 24 社という上限がある。しかし、自社が販売先として挙げなくても、取引先の企業が仕入先として挙げることはあり、逆もまた同じである。したがって、仕入と販売関係データの 2 つを組み合わせることで 24 社という上限を撤廃することができ、日本企業のお金の流れを見た取引関係ネットワークが構築可能となる。このネットワークでは企業をノードとし、ある企業から別の企業にお金の流れる取引をリンクとする。取引関係データに取引金額は記載されていないので、リンクに重みは存在しない。この企業間取引ネットワークはノード数 $N = 961,318$ 、リンク数 $L = 3,783,711$ の大規模な有向ネットワークである。

この企業間取引ネットワークの部分グラフとして、佐賀県内の企業間取引から構成されるネットワークを図 3.3 に示す。簡単のために、リンクの方向はここでは示していない。この佐賀県企業間取引ネットワークはノード数(企業数)2,760、リンク数(取引数)5,183 であり、これだけでも十分に複雑なリンク構造を持つ。

日本全体の企業間取引ネットワークの特徴を抽出するために、各企業の販売先の企業数(入次数)と、仕入先の企業数(出次数)について着目する。販売先との取引は、自分の

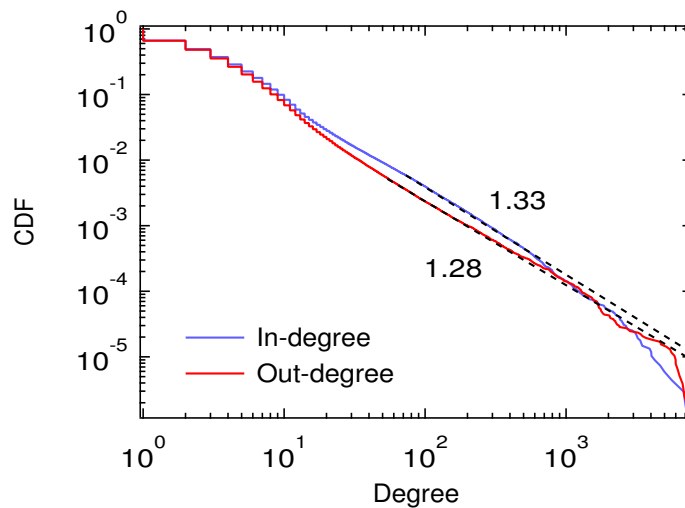


図 3.4 企業間取引ネットワークの累積度数分布。入次数 (青) と出次数 (赤) について示す。ベキ指数は入次数について 1.33, 出次数について 1.28 という結果になる。

ノードにお金を流すリンクを表し、仕入先との取引は、自分のノードからお金が流れるリンクを表す。企業間取引ネットワークにおける入次数と出次数の分布は、次の図 3.4 のようになる。入出次数の平均は 3.9, 二乗平均は入次数について 8.2×10^2 , 出次数について 9.9×10^2 となり、平均と二乗平均で大きく異なる値を持つ。累積度数分布はベキ分布に従い、企業間取引ネットワークは既存研究 [64] で知られている通り、スケールフリーネットワークであるということがわかる。累積分布でベキ指数が 1.3 なので、確率密度関数ではベキ指数が 2.3 ということになる。次節以降ではこの企業間取引ネットワークの特性について調べていく。

3.3 企業間取引ネットワークの解析

この節では前節で構築した企業間取引ネットワークの詳細なデータ解析を行い、ネットワークの統計的性質を得る。企業の年齢分布、年齢と次数の関係を観測し、優先的接続の確率の指数を測定する。

3.3.1 企業の年齢分布

本研究ではノードの消滅の効果について考慮するために、まず企業の年齢分布の解析を行なった。ネットワーク内に登場する各企業は、創業年の項目を持つ。データが 2005 年のデータであるので、 $2005 - (\text{創業年})$ を企業の年齢 τ として計算した。年齢 τ の累積分布 $P_{>}(\tau)$ を、図 3.5 に片対数プロットで示す。創業年のデータは明治元年 (1868 年) からしか存在しないので、データの年齢分布には 137 年でカットオフが生じる。40 年から 120 年の領域はよく指数分布に従い、この領域で指数 τ は 5.32×10^{-2} , すなわち時定数は 18.8 年となる。

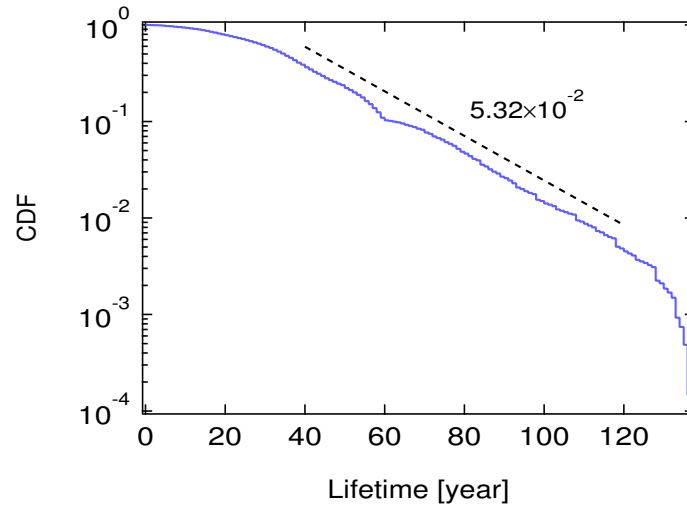


図 3.5 企業の累積年齢分布. フィッティングは傾き 5.32×10^{-2} の指数分布を表す.

信頼性工学によれば, 寿命の分布は指数分布以外にも, ガンマ分布やワイブル分布が代表的なものとして知られている. 特にワイブル分布はパラメータによって故障率を変化させることができ, 初期故障期 ($m < 1$), 偶発故障期 ($m = 1$), 摩耗故障期 ($m > 1$) の全てを再現することができる分布である. そこで, 年齢分布をワイブル分布でもフィッティングすることを試みる.

ワイブル分布では年齢が τ より大きくなる確率 (信頼度) が,

$$P_{>}(\tau) = \exp\left(-\frac{\tau^m}{\tau_0}\right) \quad (3.2)$$

で表され, これを変形すると

$$\ln \ln \frac{1}{P_{>}(\tau)} = m \ln \tau - \ln \tau_0 \quad (3.3)$$

となるので, この関係を利用して最小二乗法により m, τ_0 を求める. 結果 $m = 1.88, \tau_0 = 1.14 \times 10^3$ という値が得られ, いま $m > 1$ なので特性寿命として $\eta = \tau_0^{1/m} = 42.1$ 年となる. 年齢分布 $P(\tau)$ をワイブル分布でフィッティングした結果を図 3.6 に示す. 青のプロットが実データ, 黒の点線がフィッティングしたワイブル分布である. ワイブル分布では, あまり分布の裾の方で合わないということがわかる.

3.3.2 企業の年齢と次数の関係

BA モデルなど既存の成長モデルでは, ノードの年齢に対して次数はべき関数的に成長する. 企業ネットワークにおいて, 企業の取引数はどれぐらいのスピードで成長する (もしくは成長しない) のであろうか. 本項では, 企業の年齢と次数の関係について求めた. 各年齢に区切って, 平均次数を求めた関係を図 3.7 に示す. 年齢が 20 年以上の領域では,

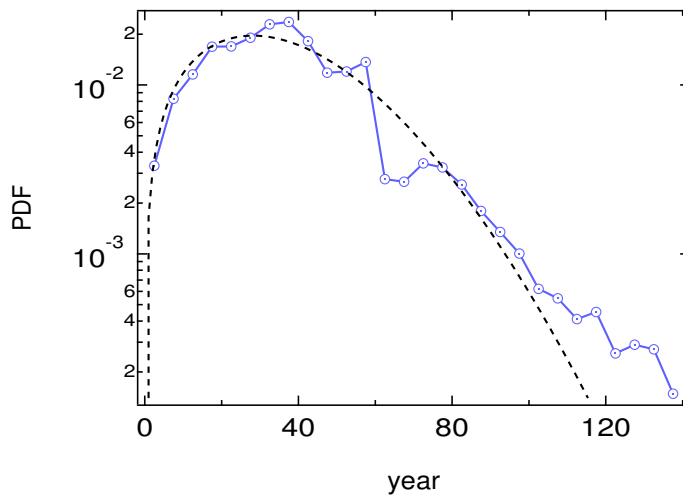


図 3.6 企業の年齢分布 (青) とワイブル分布 (黒).

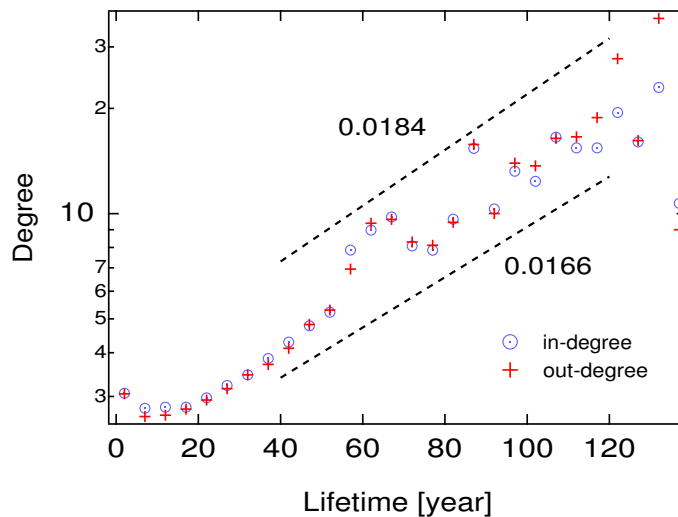


図 3.7 企業の年齢と入次数 (青), 出次数 (赤) の関係. 指数は入次数について 1.66×10^{-2} , 出次数について 1.84×10^{-2} となる.

次数は年齢に対して指数関数的に増加し、その指数は入次数について 1.66×10^{-2} 、出次数について 1.84×10^{-2} である。すなわち入出次数ともに、1年で1.02倍の成長を見せる。

あくまで各年齢での平均次数を観測しているだけなので、揺らぎも大きい。年齢ごとの次数もまたべき分布に従い、それぞれの平均次数で規格化することで図3.8のようになる。ここでは20年ごとの入出次数の分布を表示しているが、年齢が経過するに連れてべき指数が小さくなっていくことがわかる。したがって年齢とともに平均次数も増加するわけだが、指数関数的に増加するというのは自明ではない。

図3.7の0-20年では、年齢とともに平均次数が増加ではなく減少している。そこで年齢が0-20年の企業について、もう少し詳しく調べる。4年ごとに区切った0-20年の入次

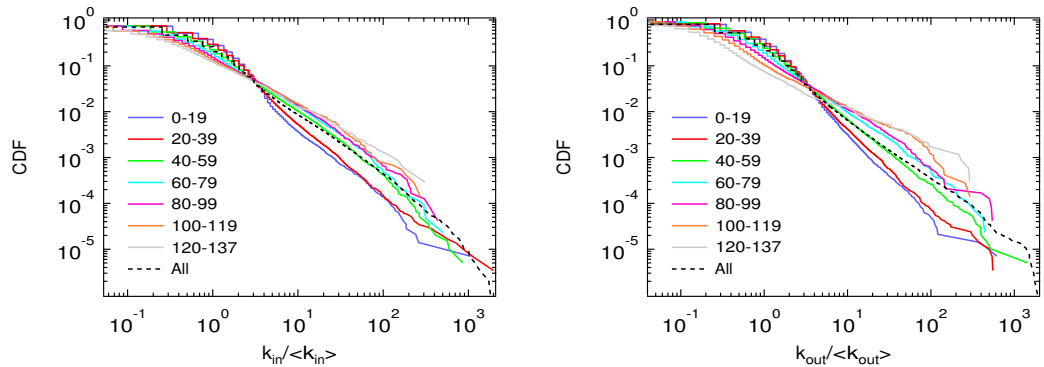


図 3.8 20 年ごとに区切り、各年齢の平均次数で規格化した年齢別累積次数分布。(左) 入次数, (右) 出次数を表し, 黒点線は全体の次数分布を表す。

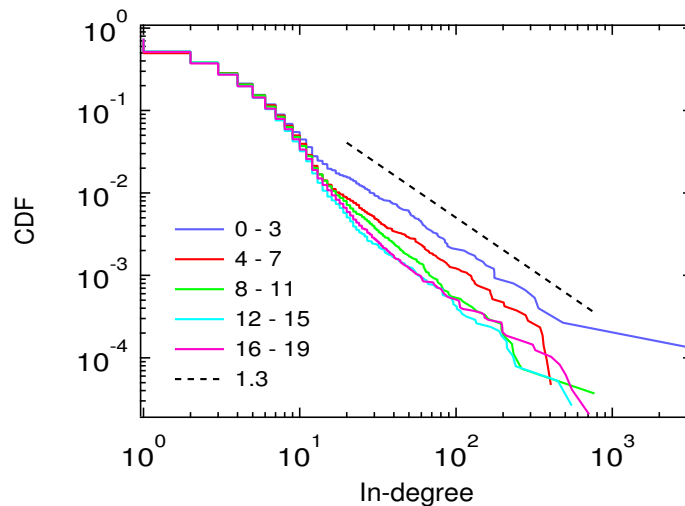


図 3.9 4 年ごとの区切りで観測した年齢別入次数分布。フィッティングは累積で指数 1.3 のべき分布。

数分布を図 3.9 に示す。すると、0-3 年の入次数分布のべき指数が一番値が小さく、そこからべき指数が増加していき指数分布に漸近するという現象が観測される。ここから、なぜ新規企業の次数分布がべき分布に従うのか、新規企業で次数の大きな企業はどのような企業なのか、という疑問が生まれる。

図 3.9 のテイルに位置するような、新規企業で次数の大きな企業はそのほとんどが営業譲渡、分社化、組織再編、経営統合、合併などによる企業である。すなわち創業年が新しいだけで、新規に市場に参入したような企業はほとんど存在しない。いま、2004 年以降に創業した企業は 1478 社存在し、このうち入次数が 10 以上なのは 88 社である。各企業の創業理由を調べ、先程述べたような理由で創業した企業を取り除くと、真の新規企業は 88 社中 24 社しか存在しない。この 2004 年以降に創業した企業の次数分布を図 3.10 に示す。合併や経営統合、名称変更などをした企業を除去する前の次数分布はべき分布に従う。一方、こうした既存の企業を除去した後の次数分布は指数分布に従い、その平均次数



図 3.10 新規参入企業の累積次数分布。(左) 企業の除去前(青), 除去後(赤)の累積次数分布の両対数プロット。(右) 除去後の累積次数分布(青)と傾き $1/2.17$ の指数分布(黒)の片対数プロット。

は 2.17 である。

東京商工リサーチ社調査のデータには、創業の経緯などに相当する項目は記載されていない。そのため年齢が小さな企業でも、企業間取引ネットワークに元々存在していて創業年が最近の企業なのか、それとも新たにネットワークに加わった企業なのかを区別するのは難しい。しかし個別企業に対して創業の経緯を調べ、次数分布を観測すると図 3.10 のように指数分布に従うことから、企業の次数分布は年齢とともに指数分布からベキ分布に変化し、ベキ指数もまた減少しテイルが広がっていきと考えられる。図 3.7 に示す年齢と平均次数の関係も、0-20 年では年齢に対して平均次数は減少するが、こうしたことを考慮すると全領域で平均次数は指数関数的に単調増加すると推測される。

3.3.3 優先的接続の確率の推定

本項では、企業間取引ネットワークの優先的接続の確率を計算し、その指数を推定する。優先的接続の確率は、次のように計算される。

優先的接続の確率の計算法

1. ネットワークのノードを、ネットワークに新しく追加されたノードと、それ以外のノードから構成されるネットワークの 2 つに分ける。
2. 新しいノードがネットワークに追加される時に、ネットワークのどのノードとリンクを張るか、接続先のノードの次数を測定する。
3. これを全ての新規ノードに対して行い、接続先の次数の確率分布 $Q(k)$ を測定する。
4. $Q(k)$ を次数 k のノード数 $N(k)$ で割り、優先的接続の確率 $\Pi(k_i)$ を得る。

以上である。 $Q(k)$ の確率分布は、次数 k のノード集合に接続する確率である。次数が 1 のノードが大きければ $Q(1)$ の値は大きくなるし、逆にハブはリンクを集めやすいがハブとなるノード数は少ないので、 $Q(k)$ の値自体は大きくなならない。1 つのノードが新たにリン

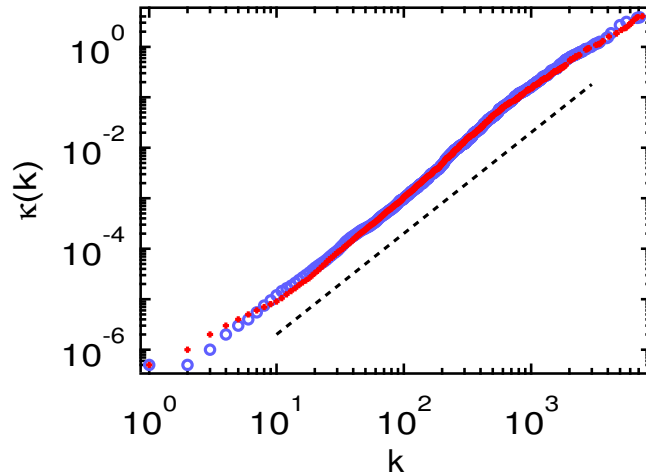


図 3.11 優先的接続の累積確率 $\kappa(k)$. 2004 年以降に創業した企業を新規参入企業とし、優先的接続の確率を計算した。入次数 (青) と出次数 (赤) について示し、フィッティングは指数 2 のべき乗を表す。

クを得る確率を計算するには、次数 k のノード数 $N(k)$ で割るといふ手順 4 が必要となる。

優先的接続の確率を求める手順としてはこれで十分であるが、次数の大きなノードは数が少ないため、実際にこの確率を計算すると値の大きな領域でプロットが揺らぐ。そこでここでは揺らぎを減らすために、Jeong ら [30] が導入した次の累積関数を計算する。

$$\kappa(k) = \int_0^k \Pi(k) dk \sim k^{\lambda+1} \quad (3.4)$$

つまり $\Pi(k_i)$ が k^{λ} に比例すると考え、その積分した値の指数 (真の値 +1) を見積もることで、優先的接続の確率の指数を推定する。

いま企業間取引ネットワークのデータは 2005 年のものなので、2004 年以降に創業した企業を新しい企業とし、ネットワークをそれ以前のネットワークと新しい企業の集合にわけ、そうして上式の $\kappa(k)$ を計算した図を、図 3.11 として示す。優先的接続の累積確率 $\kappa(k)$ は次数に比例する。指数は入次数に対して $\lambda_{in} = 1.09, \lambda_{out} = 1.05$ となり、ほぼ次数に線形比例した優先的接続が行われているということがわかる。

3.4 企業間取引ネットワークへの成長モデルの応用

本節では企業間取引ネットワークと 2 章で構築した成長モデルの比較を行う。本研究の成長モデルではノードの消滅、生成、凝集について考慮している。これらは企業間取引ネットワークにおける企業の倒産、新規参入、合併・統合といった事象に対応している。企業間取引ネットワークでは年齢分布は指数分布に従うので、単純に考えると企業は年齢に依存せず倒産しているとすることができる。成長モデルでもノードの年齢に依存せず、ノードの消滅時にノードをランダムに選択し除去している。またノードの生成・凝集時に

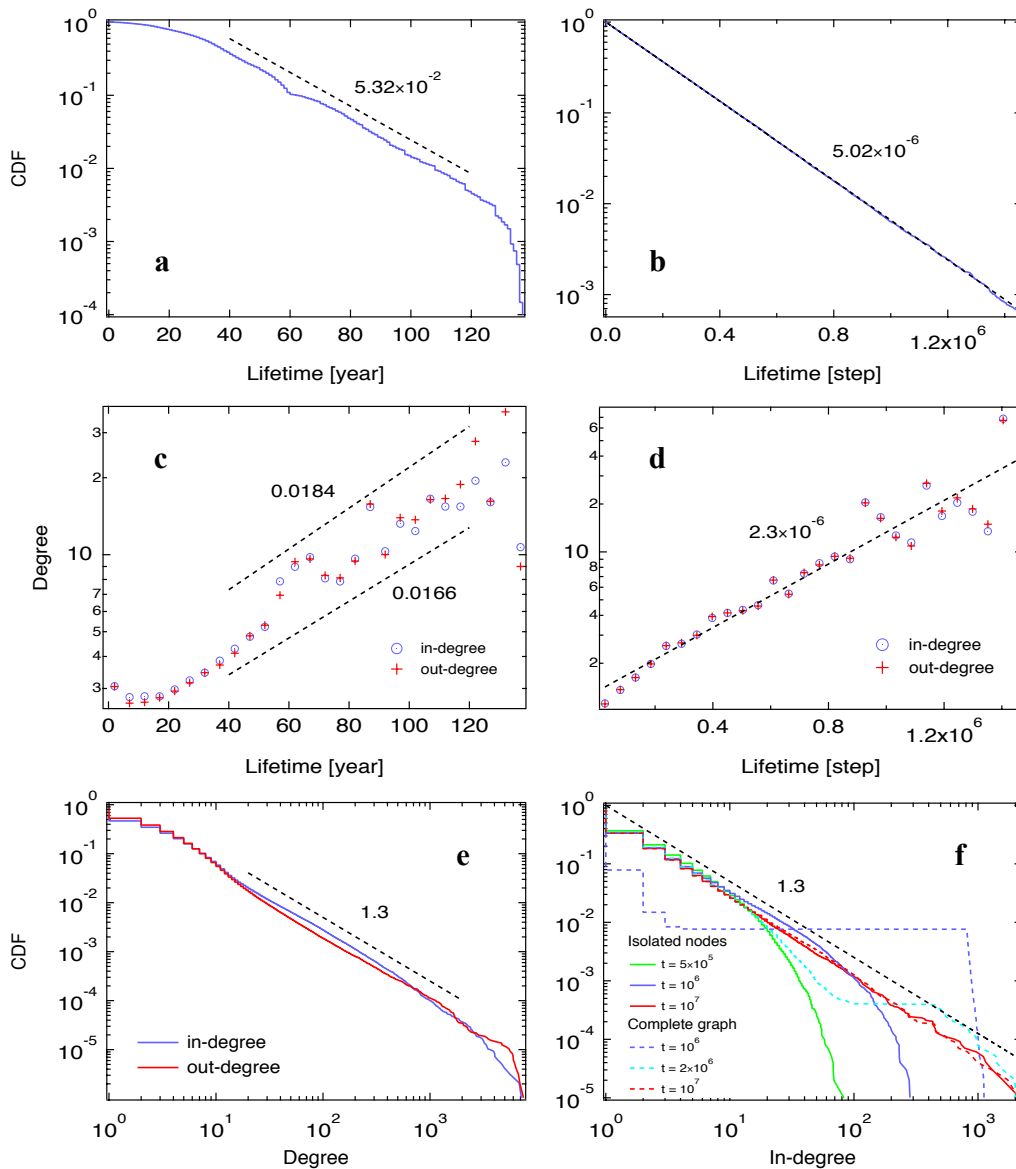


図 3.12 企業間取引ネットワークと成長モデルの統計的性質. (a,b) 企業間取引ネットワーク (a) と成長モデル (b) の年齢分布. (c,d) 企業間取引ネットワーク (c) と成長モデル (d) の年齢と次数の関係. 各年齢に区切って平均次数を計算している. (e,f) 企業間取引ネットワーク (e) と成長モデル (f) の累積次数分布. 成長モデルの方では孤立したノードと完全グラフの初期状態から開始した結果を表示している.

優先的接続で選択したノードにリンクを張るというのは、企業間取引ネットワークで観測された優先的接続とも矛盾がない。つまり構築した成長モデルは、企業間取引ネットワークの統計的性質を考慮した素過程を含む、最も単純化したモデルとなっている。

そこで凝集確率 $b = 0.5, c = 0.3, \lambda_b = \lambda_c = 1, A = 1$ として成長モデルのシミュレーションを行い、企業間取引ネットワークに成長モデルを応用した結果を、図 3.12 として示す。年齢分布はどちらも指数分布に従う。企業間取引ネットワークと成長モデルのシミュレ-

ション結果で年齢分布が一致すると考えると、両者の分布の指数の比較よりモデルにおける 10,600 ステップが1年に対応している。また年齢に対して次数はどちらも指数関数的に増加する。最後に成長モデルにおいて次数分布は、凝集：倒産の割合が3:2の時に指数が累積で1.3のベキ分布となり、現実のネットワークと一致するベキ指数を持つ。このように構築した成長モデルは、いま企業間取引ネットワークの性質をおおよそ再現することができる。しかしこの成長モデルは企業の分社化やその他の現実の企業活動を考慮しているわけではなく、ミニマルでマクロな性質を再現するモデルであるため、企業間取引ネットワークを再現できない部分も多い。その点について次節以降で説明する。

3.5 企業間取引ネットワークと成長モデルの改良

本節では、成長モデルを企業間取引ネットワークに応用する上で足りない点として、ネットワークの2リンク構造に着目する。企業間取引ネットワークでは、自社の直接の取引先だけでなく、取引先の取引先も重要である。そうした点を鑑み、2リンク構造の統計的性質として HITS アルゴリズムに着目し、この性質を再現できるように成長モデルを改良する。

3.5.1 HITS アルゴリズム

複雑ネットワークのリンク構造から重要なノードを見つけたいことがある。重要なノードを定量的に評価する方法として、中心性という指標が知られており、次数中心性や媒介中心性など様々な指標 [78–80] が提案されている。ここでは有向ネットワークで重要なノードを抽出する HITS アルゴリズムに注目する。HITS (Hypertext Induced Topic Search) アルゴリズムは、Kleinberg [80] によって考案され、Google などで用いられている PageRank [79] と同様、WWW のリンク構造のみから重要なページを抽出する検索アルゴリズムである。HITS アルゴリズムは、ウェブページから検索クエリに関連するページを抽出する段階と、抽出したページ間のハイパーリンクの構造から有用なページを見つけ出す段階の2段階にわけられる。本研究で用いるのは後者のみであり、ネットワークのリンク構造から重要なノードを見つけ出すアルゴリズムを、本論文では HITS と呼ぶ。

Kleinberg はウェブページから検索クエリに関連するページを抽出した際に、その中で有用なページの集合は入次数が大きいだけでなく、同じようなページの集合からリンクを受け取っていることを発見した。多くの有用なページにリンクを張っているページ自体には求める情報は存在しないが、どのページを参照すれば良いかを示すという点で、こうしたリンク集のページも重要である。そこで彼は有用なページの集合をオーソリティ、こうしたオーソリティのページにリンクを多数張るページの集合をハブとして、同時に見つけ出すアルゴリズムを考案した。HITS は再帰的なアルゴリズムであり、ネットワークのリンク構造において、良いハブから沢山リンクが張られているページが良いオーソリティ、良いオーソリティに沢山リンクを張っているページが良いハブとする。

具体的なアルゴリズムについて説明する。有向ネットワークでノード i からノード j へのリンクが存在する場合に $A_{ij} = 1$ とし、それ以外では値を 0 とする隣接行列 A_{ij} を考える。また、 i 番目のノードのオーソリティ度を a_i 、ハブ度を h_i とする。

HITS アルゴリズム

1. 初期状態として、全てのノードのオーソリティ度、ハブ度を 1 とする ($a_i = h_i = 1$).
2. 全てのノードに対してオーソリティ度 a_i を計算する.

$$a_i = \sum_{j=1}^N A_{ji} h_j \quad (3.5)$$

3. 計算されたオーソリティ度を用いて、次にハブ度 h_i を計算する.

$$h_i = \sum_{j=1}^N A_{ij} a_j \quad (3.6)$$

4. 値の規格化を $\sum_{i=1}^N a_i^2 = \sum_{i=1}^N h_i^2 = 1$ で行う.
5. 2. から 4. を値が収束するまで繰り返す.

上記のアルゴリズムにより、オーソリティ度 a_i が大きなノードが良いオーソリティ、ハブ度 h_i が大きいノードが良いハブとして抽出される。

式 (3.5) と式 (3.6) からオーソリティ度、ハブ度は次のように別々の式で表すことができる。

$$\mathbf{a} = \mathbf{A}^T \mathbf{A} \mathbf{a}, \quad (3.7)$$

$$\mathbf{h} = \mathbf{A} \mathbf{A}^T \mathbf{h}. \quad (3.8)$$

この式において、 $\mathbf{A}^T \mathbf{A}$ と $\mathbf{A} \mathbf{A}^T$ はそれぞれ共引用行列と書誌結合行列 [4] として知られている。共引用行列でノード i, j 間のリンク l_{ij} は、ノード i と j へのリンクを張っている共通のノード数であり、これは論文 i, j を同時に引用している論文数に対応している。一方、書誌結合行列でノード i, j 間のリンク l_{ij} は、ノード i と j から共通してリンクを張るノード数であり、これは論文 i, j に同時に引用されている論文数に対応している。両行列はどちらもリンク l_{ij} が 0 から $N-2$ までの値を取りうる (対角成分は無視する) 対称行列であり、リンクに重みがある無向ネットワークを表す。オーソリティ度とハブ度は、共引用行列と書誌結合行列の第一固有ベクトルとして得られる。

実用上はほとんど問題ないことであるが、ネットワークの構造によってはオーソリティ度とハブ度が初期状態に依存し、解が一意に求まらないという問題が Farahat ら [81] によって指摘されている。彼らはこの問題の解決法として、

$$e^A - E \simeq A + \frac{A^2}{2} \quad (3.9)$$

に対して HITS アルゴリズムを適用することを提案しているが、大規模ネットワークの場合これを計算することは不可能に近い。大西ら [66] はこの問題に対し、ネットワークの

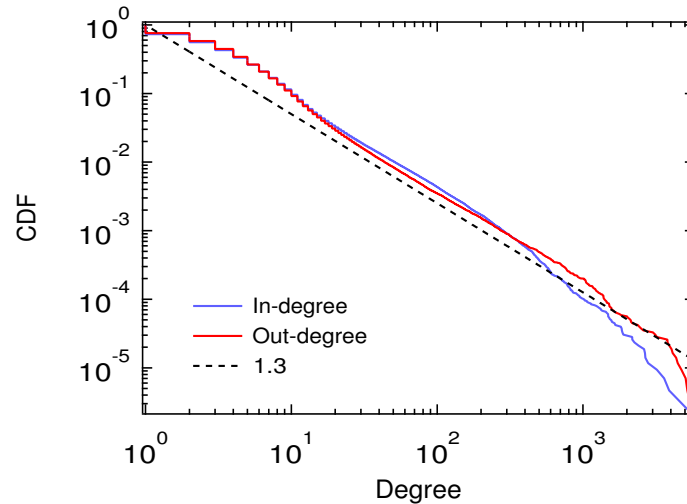


図 3.13 最大強連結成分の次数分布. 入次数 (青), 出次数 (赤) について示す. またフィッティングは累積で指数 1.3 のべき分布を示す.

隣接行列を $(1 - \epsilon)A + \epsilon/N$ と PageRank の計算におけるランダムジャンプの項を導入し、ネットワークを弱く連結させることでこの問題に対処した. 本研究では複雑ネットワークの最大強連結成分を抽出し、これに対して HITS アルゴリズムを適用することで、オーソリティ度とハブ度を求める.

強連結成分 [82] とは、ネットワークのリンクを辿って、お互いに到達可能なノードの集合である. つまり任意の 2 ノード間に、両方向に到達可能なリンクが存在している. この強連結成分の最大のクラスターが最大強連結成分であり、この部分ネットワークに対してオーソリティ度、ハブ度は一意に求まる.

3.5.2 HITS アルゴリズムの計算結果

いま企業間取引ネットワークの最大強連結成分は、ノード数 426,602, リンク数 2,583,620 であり、この次数分布を図 3.13 に示す. 最大強連結成分の次数分布もまた累積で指数 1.3 のべき分布に従い、元のネットワークとほぼ変わっていないということがわかる. この最大強連結成分に対して HITS アルゴリズムを適用する.

HITS アルゴリズムにより求めたオーソリティ度、ハブ度の累積分布を図 3.14 に示す. 次数分布が入次数と出次数で同じべき分布に従うのに対し、オーソリティ度、ハブ度の分布は全く異なる分布となる. オーソリティ度の分布は指数分布的な減衰を見せるのに対し、ハブ度の分布はテイルが広い分布となる. 参考までに PageRank の分布も図 3.15 に示す. これは全体のネットワークに対して、ランダムジャンプの確率を 0.15 として計算したものである. PageRank の分布はオーソリティ・ハブ度の分布と異なり、およそ次数と同じ値のべき指数を持つ、綺麗なべき分布に従うことがわかる.

また全体のネットワークに対して、ランダムジャンプの項を導入し HITS アルゴリズ

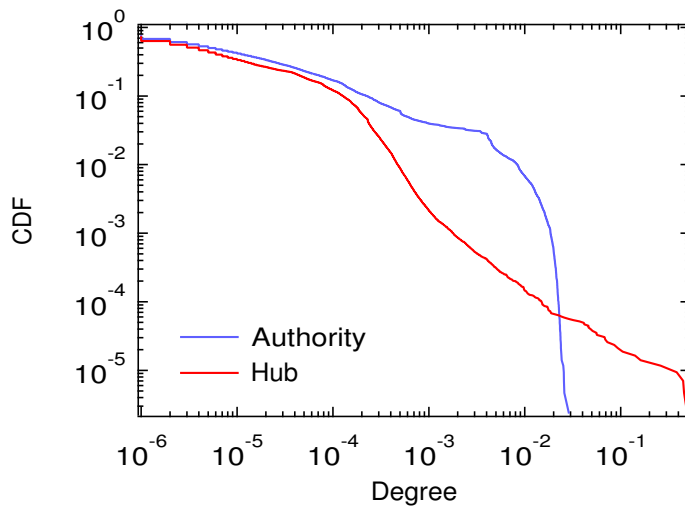


図 3.14 オーソリティ度 (青), ハブ度 (赤) の累積分布.

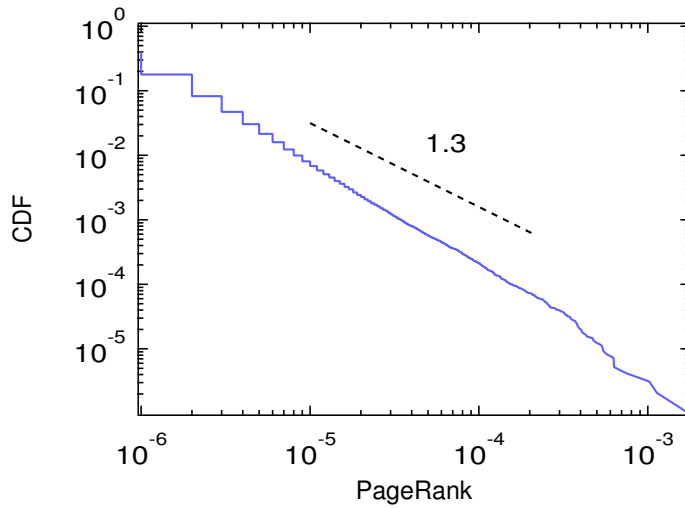


図 3.15 PageRank の累積分布. フィッティングは累積で指数 1.3 のべき分布を表す.

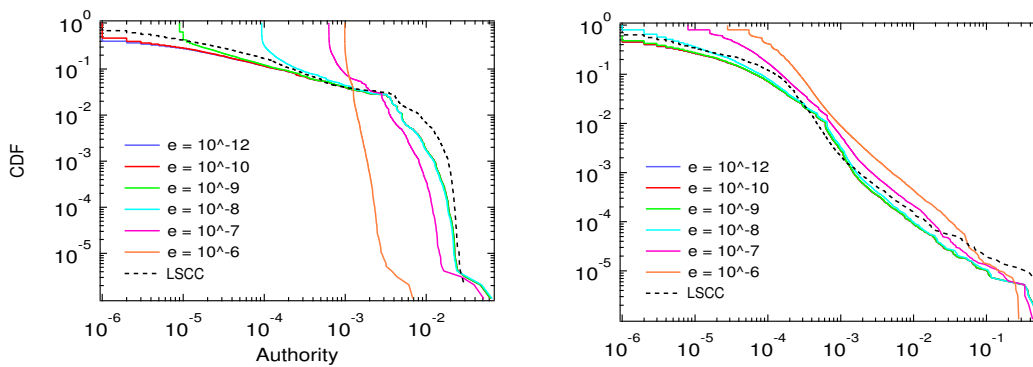


図 3.16 ランダムジャンプの項を導入して計算した HITS の結果. (左) オーソリティ度の累積分布, (右) ハブ度の累積分布. 最大強連結成分での結果を黒点線で示す.

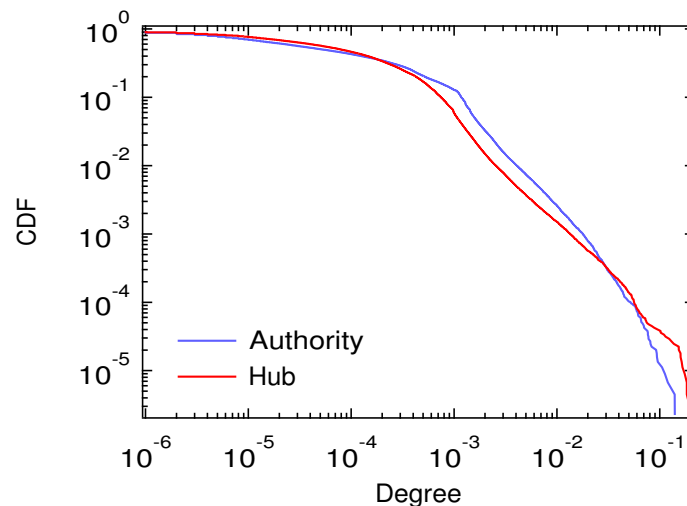


図 3.17 ランダム化した企業間取引ネットワークでのオーソリティ度 (青), ハブ度 (赤) の累積分布.

ムを計算した結果を, 図 3.16 に示す. これは隣接行列 A に対して $A' = (1 - e)A + e/N$ とし, A' という新たな隣接行列に対して HITS を計算したものである. 図 3.16 では $10^{-12} \leq e \leq 10^{-6}$ としている. 全ノード数がおよそ 10^6 なので, $e = 10^{-6}$ では全ノードから計 1 本のリンクを相互に張っている計算になる. $e \leq 10^{-9}$ ではほとんど分布は変化しない. この時, 最大強連結成分 (LSCC) の結果ともほぼ同じ分布となり, オーソリティ度とハブ度の分布は計算の手法にはよらないことがわかる.

次に企業間取引ネットワークのリンクをランダム化して HITS アルゴリズムの計算をおこなう. ランダム化の方法としては, ネットワークのリンクを 2 本ランダムに抽出し, $A \rightarrow B, C \rightarrow D$ というリンクを $A \rightarrow D, C \rightarrow B$ と付け替えることをリンク数 $\times 100$ 回繰り返した. この時ノードの次数は保存され, 全く変化しない. ランダム化したネットワークの最大強連結成分に対して, HITS アルゴリズムを計算した結果を図 3.17 に示す. この場合図 3.14 とは異なり, オーソリティ度とハブ度の分布はほぼ同一の分布を示す. リンクのつなぎ替えを行なったネットワークではオーソリティ度とハブ度の分布の非対称性が見られないということは, この非対称性は企業間取引ネットワークに特有の性質であるということができる.

表 3.1 と表 3.2 に, それぞれオーソリティ度, ハブ度の高い 10 企業を, 所属する業種名で載せる. オーソリティ度の高い企業には様々な業種の企業が含まれている. この中には総合商社 (各種商品卸売業) のように入次数が多く, かつオーソリティ度の高い企業もあれば, 建設下請け企業のように入次数は少ないがハブ度の高い企業と取引をおこない, オーソリティ度の高い企業も存在する. 一方, ハブ度の高い企業は全て総合工事業に属する企業である. いまハブ度を平方和が 1 となるように規格化しているので, 表 3.2 の 10 企業だけでハブ度の大部分を占有していることがわかる.

表 3.1 オーソリティ度の高いトップ 10 企業 (入次数, 出次数, オーソリティ度, ハブ度)

順位	オーソリティ度の高い企業
1	物品賃貸業 (312, 186, 2.89×10^{-2} , 1.91×10^{-3})
2	設備工事業 (144, 1390, 2.59×10^{-2} , 1.51×10^{-2})
3	各種商品卸売業 (5336, 4044, 2.56×10^{-2} , 1.86×10^{-2})
4	金属製品製造業 (1524, 515, 2.55×10^{-2} , 4.51×10^{-3})
5	建築材料, 鉱物・金属材料等卸売業 (152, 139, 2.52×10^{-2} , 3.05×10^{-3})
6	その他の事業サービス業 (11, 3, 2.42×10^{-2} , 1.15×10^{-4})
7	廃棄物処理業 (20, 11, 2.41×10^{-2} , 3.0×10^{-5})
8	職別工事業 (13, 8, 2.40×10^{-2} , 2.74×10^{-4})
9	廃棄物処理業 (16, 8, 2.38×10^{-2} , 1.59×10^{-4})
10	職別工事業 (24, 36, 2.38×10^{-2} , 6.92×10^{-4})

表 3.2 ハブ度の高いトップ 10 企業 (入次数, 出次数, オーソリティ度, ハブ度)

順位	ハブ度の高い企業
1	総合工事業 (172, 5553, 6.03×10^{-3} , 0.494)
2	総合工事業 (163, 5191, 2.88×10^{-3} , 0.458)
3	総合工事業 (165, 5056, 5.84×10^{-3} , 0.449)
4	総合工事業 (138, 4347, 5.67×10^{-3} , 0.383)
5	総合工事業 (121, 3129, 3.12×10^{-3} , 0.260)
6	総合工事業 (51, 2135, 1.12×10^{-3} , 0.162)
7	総合工事業 (64, 2085, 1.52×10^{-3} , 0.141)
8	総合工事業 (67, 2030, 4.23×10^{-3} , 0.105)
9	総合工事業 (44, 1425, 8.03×10^{-4} , 9.34×10^{-2})
10	総合工事業 (63, 1408, 1.35×10^{-3} , 8.56×10^{-2})

3.5.3 企業間取引ネットワークの 2 リンク構造

オーソリティのノードはハブのノードを介して相互作用し、ハブのノードはオーソリティのノードを介して相互作用する。この様子を図 3.18 に示す。図 3.18 で点線で表されている相互作用の強さは、2 ノードを仲介するノード数に比例し、相互作用が強くなればなるほどノード同士のオーソリティ度、ハブ度は大きくなる。この関係は式 (3.7) と式 (3.8) の、共引用行列 C と書誌結合行列 B で表され、それぞれ

$$C_{ij} = \sum_{k=1}^N a_{ki}a_{kj}, \quad (3.10)$$

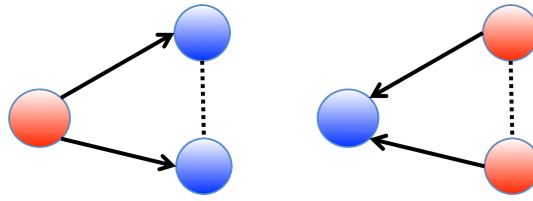


図 3.18 オーソリティ同士 (青いノード)、ハブ同士 (赤いノード) の相互作用の概念図

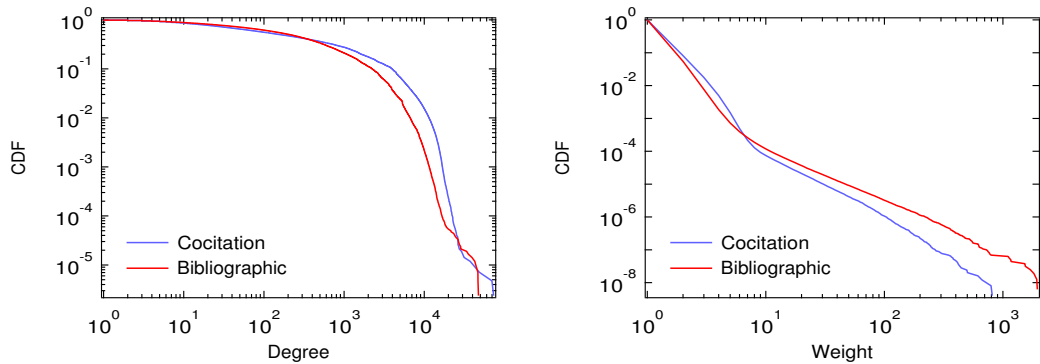


図 3.19 2リンク構造ネットワークの (左) 累積度数分布と (右) リンクの重みの累積分布。オーソリティネットワーク (青), ハブネットワーク (赤) について示す。

$$B_{ij} = \sum_{k=1}^N a_{ik} a_{jk} \quad (3.11)$$

となる。したがってオーソリティ度、ハブ度の構造を探るためには、共引用行列、書誌結合行列を隣接行列とする2リンク構造ネットワークの解析を行えば良い。いま、共引用行列を隣接行列として構成されるネットワークをオーソリティネットワーク、書誌結合行列を隣接行列として構成されるネットワークをハブネットワークと呼ぶ。このネットワークはリンクの重みが0から $N-2$ の無向ネットワークである。ただし、対角成分は0として自己ループは存在しないものとする。2リンク構造のネットワークを解析することは、モチーフ解析の項で紹介したV字型構造の詳細を調べることも関連している。

企業間取引ネットワークの最大強連結成分に対して、共引用行列と書誌結合行列を計算した。元々のネットワークはノード数に対してリンク数が少なく疎行列であるが、2リンク構造のオーソリティネットワークとハブネットワークはリンク数がそれぞれ248,741,138本と159,414,273本となり、元の隣接行列よりも密な行列となる。図3.19に度数分布と各リンクの重みの分布を示す。度数分布は 10^3 以下程度まではほぼ同じ分布であるが、値の大きな領域ではオーソリティネットワークの方がハブネットワークよりも度数が大きい。これは両ネットワークのリンク数からも理解できる結果となる。一方、各リンクの重みの分布を見ると、ハブネットワークの方がべき指数の値が小さな分布となる。すなわちオーソリティネットワークとハブネットワークを比較すると、ハブネットワークの方はオーソリティネットワークよりリンク数や度数が小さいが、各リンクの重みは大きな値を取る。

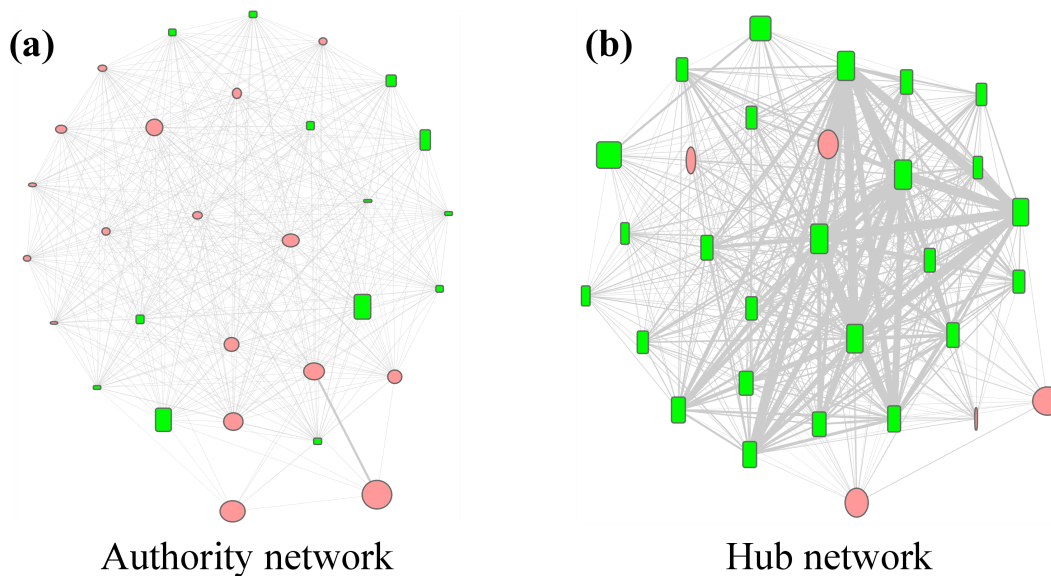


図 3.20 (a) オーソリティネットワークと (b) ハブネットワーク。それぞれ度数の高い上位 30 社から構成されており、建設業(緑)とその他の業種(赤)で区別している。またノードの幅は入次数の対数、高さは出次数の対数に比例させている。リンクの太さはそれぞれの隣接行列の重みに比例している。

HITS の計算結果において、ハブ度の高い企業は全て総合工事業、いわゆるゼネコン企業であった。そこで以降は建設業の企業に注目して解析を行う。オーソリティ度、ハブ度の高い上位 30 社から構成される、オーソリティネットワーク、ハブネットワークを図 3.20 に示す。図 3.20 においてリンクの太さは、対応する隣接行列の値に比例しており、建設業の企業を緑で、その他の業種を赤のノードとして示している。またノードの幅と高さは、入次数と出次数に比例させている。オーソリティネットワークはリンクの重みがほぼ一様で、その多くが中小企業から構成されるネットワークである。一方、ハブネットワークでは建設業の大きな企業同士が、太いリンクで結ばれている。これらの大きな建設企業は、オーソリティとなる多くの中小下請け企業を仲介して、強く結びついている。つまり図 3.18 の右側のような構造が、これらの大企業間では共通の下請け企業を仲介することで、よく観測される。

逆にオーソリティ度の高い企業は、こうした建設大企業と取引をすることでオーソリティ度を高めている。図 3.21 に全体のオーソリティ度と、ハブ度最大の企業と取引をしている企業のオーソリティ度を、順位プロットで示す。図 3.21 からハブ度最大企業の隣接ノードだけで、全体のオーソリティ度分布をほぼ再現できることがわかる。2リンク構造ネットワークでは建設の大企業が強く結びつき、それと同時に多数の中小企業がこれら大企業からお金をもらう取引をしている。この構造がハブ度のテイルの広い分布、オーソリティ度の指数分布の原因となっている。

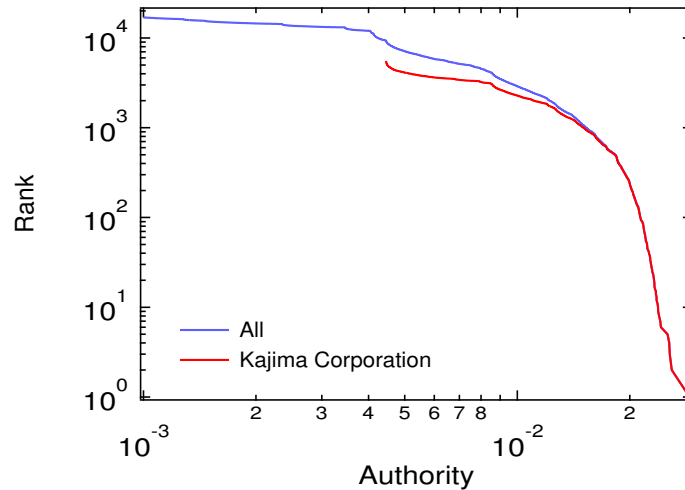


図 3.21 オーソリティ度の順位プロット。全体のオーソリティ度(青)とハブ度最大の企業の隣接企業のオーソリティ度(赤)について示す。

3.5.4 企業間取引ネットワークと成長モデルの改良

本項では前項で観測された性質を、成長モデルで再現することを試みる。表 3.1 で観測されたように、ハブ度の高い建設企業は入次数が少なく出次数が大きい。このことから新規下請け企業がリンクを張る際に、建設業の企業からお金を払ってもらう取引を行う確率は、他の業種よりも高いと推測される。そこで新たに優先的接続の確率を、下記の方法により建設業の企業に限定して測定する。

業種を限定した優先的接続の確率

業種を限定した優先的接続の確率の計算法

1. ネットワークのノードを、ネットワークに新しく追加されたノードと、それ以外のノードから構成されるネットワークの 2 つに分ける。
2. 新しいノードがネットワークに追加され、ある業種のノードとリンクを張る際に、接続先のノードの次数 k_{con} を測定する。
3. これを全ての新規ノードに対して行い、接続先の次数の確率分布 $Q_{con}(k)$ を測定する。
4. $Q_{con}(k)$ を次数 k のノード数 $N(k)$ で割り、優先的接続の確率 $\Pi_{con}(k_i)$ を得る。

建設業の出次数に対して、この業種別の優先的接続の確率、及び式 (3.4) による $\kappa(k)$ を測定する。この確率が意味するものは、新規ノードが建設業の企業からお金をもらう取引を開始する確率が、その出次数の何乗に比例するかということである。

図 3.22 に建設業の出次数について測定した優先的接続の確率を示す。企業間取引ネッ

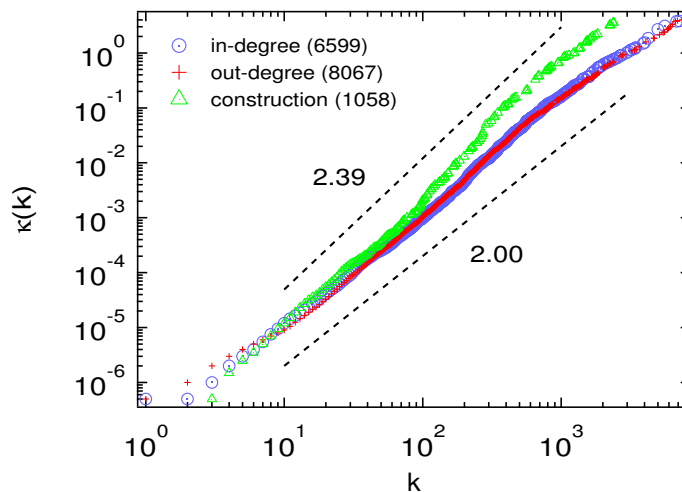


図 3.22 優先的接続の累積確率 $\kappa(k)$. 全体の入次数 (青), 全体の出次数 (赤), 建設業の出次数 (緑) について示す. 全体の入出次数の優先的接続の確率に関しては, 図 3.11 と同じである. カッコ内は確率分布 $Q(k)$ を構成する企業数である.

トワーク全体ではほぼ次数に線形比例する優先的接続の確率を示すのに対し, 建設業に限定すると出次数の 1.39 乗に比例した強い優先的接続を行なっていることがわかる. したがって建設業においては, 大企業が中小の新規下請け企業からリンクを張られるバイアスが通常よりも強い.

成長モデルによる 2 リンク構造ネットワークの再現

成長モデルのシミュレーションで, HITS アルゴリズムの計算結果を再現することを試みる. 建設業の大企業がハブネットワークにおいて, 中小の下請け企業を仲介して強く結びついているという点と, 建設業の出次数では優先的接続の確率の指数が全体のネットワークよりも大きいという点を考慮し, 2 章で構築した成長モデルを次のように改良する.

改良した成長モデルのアルゴリズム

2.(b) 確率 b でノードの追加を行う.

確率 p_{con} で入次数が $2(k_{in} = 2)$ の新規ノードを 1 つネットワークに追加し, 新規ノードに入るリンクの接続先を確率 $\Pi(k_{i,out}) = (k_{i,out}^{1.39} + 1) / \sum_j (k_{j,out}^{1.39} + 1)$ で優先的に選択する.

確率 $1 - p_{con}$ で入出次数が $1(k_{in} = k_{out} = 1)$ の新規ノードを 1 つネットワークに追加し, 新規ノードから出るリンクの接続先を確率 $\Pi(k_{i,in}) = (k_{i,in} + 1) / \sum_j (k_{j,in} + 1)$ で, 新規ノードに入るリンクの接続先を確率 $\Pi(k_{i,out}) = (k_{i,out} + 1) / \sum_j (k_{j,out} + 1)$ で優先的に選択する.

改良した成長モデルでは, 2 章で構築した成長モデルのノードの生成の事象を, 上記のように変更する. ここで確率 p_{con} で生成するノードは中小下請け企業を想定しており, こ

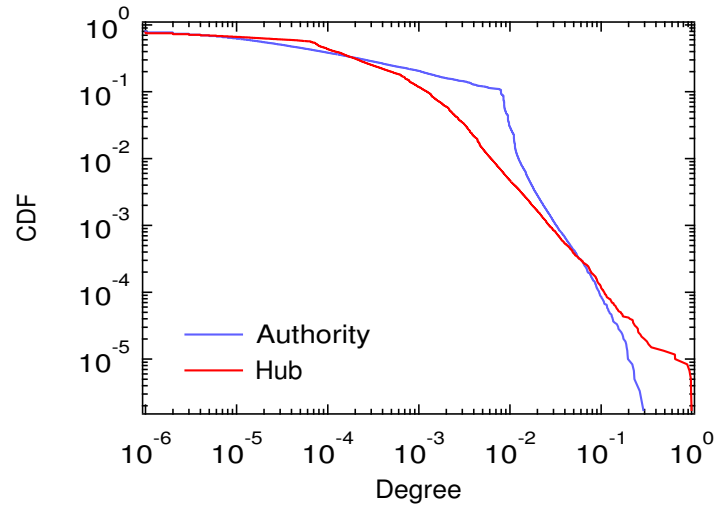


図 3.23 改良した成長モデルによるオーソリティ度 (青), ハブ度 (赤) の累積分布. シミュレーション 10 回のアンサンブル平均を示す.

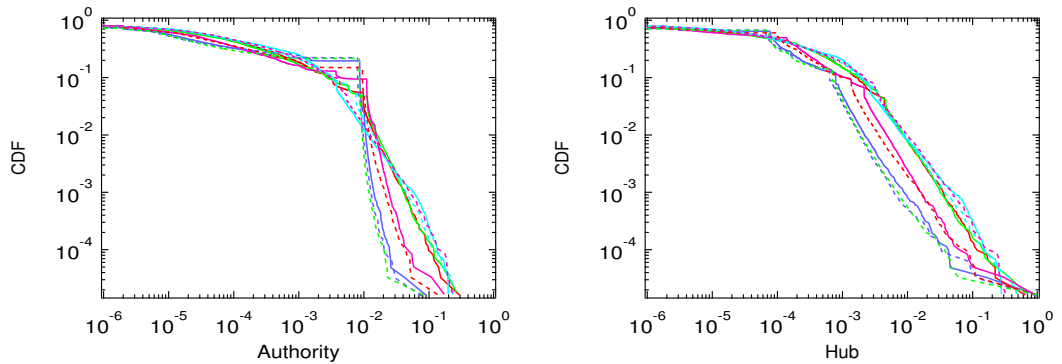


図 3.24 改良した成長モデルによるオーソリティ度 (青), ハブ度 (赤) の累積分布. 各回のシミュレーション結果を示す.

のノードが強い優先的接続によりハブとなるノードと接続するようにしている.

優先的接続の確率の測定で, 新規ノードがリンクを接続する次数の確率分布 $Q(k)$ を構成するノード数は, 全体の入次数が 6,599, 全体の出次数が 8,067, 建設業の出次数が 1,058 であった. そこでこの観測された企業数から, $p_{con} = \frac{1058}{6599+8067} = 0.0721, c = 0.3, N_0 = 10^5, T = 10^7$ として, 改良した成長モデルのシミュレーションを行う.

この改良モデルのシミュレーションから生成されたネットワークの最大強連結成分を抽出し, HITS アルゴリズムを計算する. 得られたオーソリティ度, ハブ度の分布を 10 回のアンサンブル平均として図 3.23 に示す. ハブ度の方がオーソリティ度よりもテイルの広い分布となり, 観測結果を再現できている. また各回のシミュレーション結果を図 3.24 に示す. 各回のシミュレーション結果を見ると, 必ずしも現実のネットワークのオーソリティ度, ハブ度の分布とは一致しない. これはいま確率 p_{con} で生成するノードを入次数が 2 と限定してしまっているため, 生成したネットワークには大きなハブ度を持つノード

が2つしか存在せず、確率 p_{con} で生成したノードはほぼ全てこの2つのハブノードに接続されているからである。

現実の企業間取引ネットワークには、ハブ度の高い企業が2つだけではなく多数存在する。シミュレーションではアンサンブル平均を取ることで、ハブ度の高いノードを多数作り出し現実の統計性を再現しているが、新規に生成する優先的接続の強いノードのリンク数を柔軟に変化させることで、アンサンブルではなく各試行から計算されるオーソリティ度、ハブ度の分布を再現できると思われる。

第4章

結論

本章では本論文のまとめ及び、更なる解析が必要な項目や、今後の研究の発展の方向性について簡潔に述べる。

4.1 本論文のまとめ

本論文では、既存の複雑ネットワークの成長モデルに新たにノードの凝集という概念を導入し、構築したモデルのシミュレーションと理論解析から、既存の成長モデルでは再現できない性質を観測した。また構築した成長モデルの応用例として企業間取引ネットワークに着目し、このネットワークの統計的性質を成長モデルを用いて説明することを試みた。

まず1章では先行研究について述べ、先行研究で考察されていない部分をいくつか挙げ、本研究の方向性を示した。次に2章では、先行研究で考慮されていなかった凝集現象の物理に着目し、複雑ネットワークの成長モデルに新たにノードの凝集を導入した。この新たに追加した効果の影響を、構築したモデルのシミュレーションを行い調べた。また同時に、1ステップでの次数の遷移確率を考慮してマスター方程式を導出し、これを理論的・数値的に解き、シミュレーション結果と一致することを確かめた。モデルのネットワークのサイズ依存性、シミュレーション時間依存性について調べ、またノードの消滅と凝集の割合を変化させることで、次数分布がどのように変わるかを調べた。ノードの消滅に対して凝集の割合が増加することで、次数分布が指数分布からべき分布へと変化する事を見出した。

3章では構築した成長モデルの、企業間取引ネットワークへの応用を行なった。日本企業約100万社から構成される企業間取引ネットワークを構築し、このネットワークの解析を行い統計的性質を示した。またモデルのシミュレーション結果と現実のネットワークの比較を行い、ミニマルな成長モデルで現実のネットワークの性質をおおよそ再現できることを確認した。さらにHITSアルゴリズムによるオーソリティ度、ハブ度の分布を示し、V字型構造に着目した2リンク構造ネットワークの解析を行なった。ハブの大企業同士の強いつながりが、中小の下請け企業を介して生成されていることを確認し、この性質を成

長モデルにわずかな改良を加えることで再現した。

4.2 今後の展望

本研究は凝集現象を考慮した成長モデルの構築を行なった。またこのモデルで企業間取引ネットワークに応用し、ネットワークのスケールフリー性を説明できることを示した。しかしモデルの素過程がノードの消滅、生成、凝集という最小限の要素しか含まないため、現実のネットワークを再現するには至らない点も多い。本研究を企業間取引ネットワークに応用するためには次のような研究が必要であると考えられる。

企業間取引ネットワークの更なる解析とモデルの複雑化

いま成長モデルの数理的な構造は理解できているが、モデルの応用性という点では不十分である。実際の企業活動では、倒産一つをとっても販売不振、放漫経営、赤字累積、過小資本、偶発的原因、在庫状況悪化、設備投資過大、他社倒産の余波、売掛金等回収難、金融機関・取引先の打ち切りなど様々な理由が存在する [83, 84]。いま本研究で構築した成長モデルでは、ノードを1つランダムに選択しこのノードと付随するリンクを全て取り除くとしているが、これでは現実のネットワークを再現するモデルとしては不十分なのは明らかである。創業からの年数や、サイズ、業種、地域など様々な要因が変数として存在し、またリンクも全て除去されとは限らない。

ノードの生成や凝集においても同じようなことが言える。特にノードの凝集時に、優先的接続の確率の指数を新規参入時と同じく1としているが、これはデータから観測された事実ではなく仮定である。東京商工リサーチ社のデータには合併などに値するデータが存在しないので、この合併時の優先的接続の確率の指数を確認することはできなかったが、帝国データバンク社の調査データなど別のデータを用いて確認していく必要がある。また本研究で構築した成長モデルが、これら別の企業データでも応用できるのか更なるデータの解析が必要である。

このような解析が十分になされ、成長モデルの改良の後には、次のような研究の発展が可能になると考えられる。

成長モデルの企業間取引ネットワークへの更なる応用

成長モデルによって企業間取引ネットワークの成長の再現性が増加することで、様々な企業間取引ネットワークの現象が説明できるようになる。一例として企業間取引ネットワークのストレステストが可能となる。ネットワークからどれだけ企業や取引を取り除くと、ネットワークが崩壊するかや、災害などによって企業が多数倒産した場合、元の企業数に戻るまでどれだけの時間がかかるかなどが推定できる。特にいま、成長モデルのシミュレーション時間と現実のネットワークの経過年数は理論的に対応付けられており、現実には則した年数で説明することが可能である。

ネットワークからノードやリンクを取り除いていき、ネットワークが崩壊する現象を調

べることは、物理ではパーコレーション [85] として知られている。次数分布のベキ指数を変化させたネットワークを多数成長モデルにより生成することで、こうしたパーコレーションの研究にも応用することができる。

成長モデルと輸送モデルの融合

本研究で構築した成長モデルは、ネットワークの生成過程を再現するモデルであり、ネットワーク上で起こる現象については説明できない。企業間取引ネットワーク上で起こる現象としては、渡邊ら [71] や田村ら [70] によるお金の輸送現象が挙げられる。企業間取引ネットワーク上では企業同士の取引によるお金の流れが存在する。彼らはこの流れを、リンク両端のノードのサイズに非線形比例することを見出し、輸送モデルを構築することで説明した。しかしこれらのモデルは静的なネットワークでのみ説明され、現実の企業間取引ネットワークのように常に構造が変化し続ける状況においては、何も理解が進んでいない。

本研究で構築した成長モデルとこれらの輸送モデルを組み合わせることで、ネットワークの成長が輸送に与える影響と、輸送がネットワークの成長に与える影響が同時に理解できるようになる。前者の成長が輸送に与える影響は、キャッシュを多く持つ企業と取引が開始されるような現象を表し、後者の輸送が成長に与える影響は収入の少ない企業は倒産してネットワークから消滅するような現象を表す。これらの成長モデルと輸送モデルを同時に考慮することにより、複雑ネットワークの理解が進むだけでなく、企業間取引ネットワークで起こる経済現象の理解も進むと考えられる。

謝辞

研究を終えるにあたり、まず高安美佐子先生に感謝したいと思います。先生にはどんな時でも研究の相談をしていただき、また研究だけにとどまらず、様々なご指導をこの5年間で受けることができました。研究が進まないような時でも、常に励ましていただき、そのポジティブさに何度救われたかわかりません。大変お世話になりました。また高安秀樹先生にも、美佐子先生と一緒に、研究の指導をしていただきました。秀樹先生のアイデアとクリティカルな助言により、研究が大きく進むということが多々あり、大変感謝しております。

本論文の審査をしていただいた、小野功先生、樺島祥介先生、出口弘先生、寺野隆雄先生にもお世話になりました。審査のために貴重な時間を割いていただき、様々な質問とコメントをいただきました。特に論文の構成について熱心に議論していただき、そのおかげで本論文の構成は多少なりとも改善することができたと思います。

秘書の町田理香さんにも感謝しております。研究を遂行する上で様々なサポートをいただいただけでなく、私生活の様々な相談にまで乗っていただき、快適な研究生活を送ることができました。研究室 OB, OG の山田健太さん、佐野幸恵さん、渡邊隼史さん、また東京大学の西立顕さんにもお世話になりました。様々な研究のアドバイスや、新しい知

識・技術を教えていただき、研究のモチベーションを保つことができました。研究室の後輩みんなにも感謝しております。この5年間、決して楽しいことばかりではありませんでしたが、それでも研究室生活を楽しく過ごすことができたのは、みんなと一緒に遊んだり他愛もない会話をすることができたからだと思います。特に田村光太郎君には本論文を書く間、本の様々な執筆作業を任せてしまい、大変ありがたく思っております。

この謝辞に書ききることができない多くの方々に、私はお世話になりました。様々な学会や研究会でお会いし、質問やコメントをいただき、研究とは一人でするものではなく、人と人とのつながりの中で育まれるものだと教わりました。そうした方々にもこの場を借りてお礼を申し上げたいと思います。本当にどうもありがとうございました。

参考文献

- [1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, January 2002.
- [2] S N Dorogovtsev and J F F Mendes. Evolution of networks. *Advances in Physics*, 51(4):1079–1187, June 2002.
- [3] M E J Newman. The Structure and Function of Complex Networks. *SIAM Rev.*, 45(2):167–256, January 2003.
- [4] Mark Newman. *Networks: An Introduction*. Oxford University Press, March 2010.
- [5] 増田 直紀 and 今野 紀雄. 複雑ネットワーク. 基礎から応用まで. 近代科学社, April 2010.
- [6] 矢久保 考介. 複雑ネットワークとその構造. 共立出版, February 2013.
- [7] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.
- [8] Romualdo Pastor-Satorras, Alexei Vázquez, and Alessandro Vespignani. Dynamical and Correlation Properties of the Internet. *Phys. Rev. Lett.*, 87(25):258701, November 2001.
- [9] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.
- [10] Hawoong Jeong, Sean P Mason, A L Barabási, and Zoltan N Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.
- [11] Fredrik Liljeros, Christofer R Edling, Luís A Nunes Amaral, H Eugene Stanley, and Yvonne Åberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, June 2001.
- [12] Albert-László Barabási, Hawoong Jeong, Zoltan Nédá, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3):590–614, 2002.
- [13] Roger Guimerà, S Mossa, A Turtshi, and L A N Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, May 2005.
- [14] José M Montoya, Stuart L Pimm, and Ricard V Solé. Ecological networks and their

- fragility. *Nature*, 442:259–264, July 2006.
- [15] J-P Onnela, J Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, K Kaski, J Kertész, and A L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.
- [16] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.
- [17] Réka Albert and Albert-László Barabási. Topology of evolving networks: local events and universality. *Phys. Rev. Lett.*, 85(24):5234–5237, December 2000.
- [18] Jon M Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S Tomkins. The web as a graph: Measurements, models, and methods. *Lecture Notes in Computer Science*, 1627:1–17, 1999.
- [19] P L Krapivsky, S Redner, and F Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85(21):4629, 2000.
- [20] Sergey N Dorogovtsev, José Fernando Ferreira Mendes, and Alexander N Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Lett.*, 85(21):4633, 2000.
- [21] Ginestra Bianconi and A L Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436, 2001.
- [22] P L Krapivsky and S Redner. Organization of growing random networks. *Phys. Rev. E*, 63(6):066123, May 2001.
- [23] M E J Newman, S H Strogatz, and D J Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64(2):026118, July 2001.
- [24] K I Goh, B Kahng, and D Kim. Universal Behavior of Load Distribution in Scale-Free Networks. *Phys. Rev. Lett.*, 87(27):278701, December 2001.
- [25] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65(2):026107, January 2002.
- [26] Konstantin Klemm and Víctor M Eguíluz. Growing scale-free networks with small-world behavior. *Phys. Rev. E*, 65(5):057102, May 2002.
- [27] Cristopher Moore, Gourab Ghoshal, and M E J Newman. Exact solutions for models of evolving networks with addition and deletion of nodes. *Phys. Rev. E*, 74(3):036121, September 2006.
- [28] Derek J De Solla Price. Networks of Scientific Papers. *Science, New Series*, 149(3683):510–515, July 1965.
- [29] S Redner. How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B*, 4(2):131–134, August 1998.
- [30] Hawoong Jeong, Zoltan Néda, and Albert-László Barabási. Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4):567, 2003.
- [31] Serguei Saavedra, Felix Reed-Tsochas, and Brian Uzzi. Asymmetric disassembly and

-
- robustness in declining networks. *Proceedings of the National Academy of Sciences*, 105(43):16466–16471, 2008.
- [32] George M Hidy and James Rush Brock. *Topics in current aerosol research*. Pergamon, 1971.
- [33] M H Ernst. Kinetics of Clustering in Irreversible Aggregation. *Fractals in Physics*, pages 289–302, 1986.
- [34] 早川 尚男. 凝集現象の数理. *数理科学*, 291:42–49, 1987.
- [35] David J Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.
- [36] F Leyvraz. Scaling theory and exactly solved models in the kinetics of irreversible aggregation. *Physics Reports*, 383(2-3):95–212, August 2003.
- [37] Sheldon Kay Friedlander. *Smoke, Dust, and Haze*. Fundamentals of Aerosol Dynamics. Oxford University Press, 2000.
- [38] M Lee. On the Validity of the Coagulation Equation and the Nature of Runaway Growth. *Icarus*, 143(1):74–86, January 2000.
- [39] Hans R Pruppacher and James D Klett. *Microphysics of Clouds and Precipitation*. Springer Netherlands, Dordrecht, 1978.
- [40] Walter H Stockmayer. Theory of molecular size distribution and gel formation in branched-chain polymers. *The Journal of Chemical Physics*, 11(2):45, 1943.
- [41] M V Smoluchowski. Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Zeitschrift fur Physik*, 17:557–585, 1916.
- [42] Robert M Ziff, E M Hendriks, and M H Ernst. Critical properties for gelation: A kinetic approach. *Phys. Rev. Lett.*, 49:593–595, 1982.
- [43] Warren H White. On the form of steady-state solutions to the coagulation equations. *Journal of Colloid and Interface Science*, 87(1):204–208, 1982.
- [44] E Ben-Naim and P L Krapivsky. Kinetics of aggregation-annihilation processes. *Phys. Rev. E*, 52(6):6066, 1995.
- [45] P L Krapivsky and S Redner. Transitional aggregation kinetics in dry and damp environments. *Phys. Rev. E*, 54(4):3553, 1996.
- [46] PGJ Van Dongen and M H Ernst. Dynamic scaling in the kinetics of clustering. *Phys. Rev. Lett.*, 54(13):1396, 1985.
- [47] Hisao Hayakawa. Irreversible kinetic coagulations in the presence of a source. *Journal of Physics A: Mathematical and General*, 20(12):L801, 1987.
- [48] Hideki Takayasu, Ikuko Nishikawa, and Hal Tasaki. Power-law mass distribution of aggregation systems with injection. *Physical Review A*, 37:3110–3117, 1988.
- [49] Hideki Takayasu. Steady-state distribution of generalized aggregation system with injection. *Phys. Rev. Lett.*, 63(23):2563–2565, 1989.

- [50] Hideki Takayasu, Misako Takayasu, Astero Provata, and Greg Huber. Statistical properties of aggregation with injection. *Journal of Statistical Physics*, 65(3-4):725–745, 1991.
- [51] Misako Takayasu. Universal power law observed in an exponentially growing particle system. *Physical Review A*, 46(2):782, 1992.
- [52] Misako Takayasu, Hideki Takayasu, and Y-H TAGUCHI. NON-GAUSSIAN DISTRIBUTION IN RANDOM TRANSPORT DYNAMICS. *Int. J. Mod. Phys. B*, 08(28):3887–3961, December 1994.
- [53] 三浦 航, 渡邊 隼 史, 水野 貴之, 高安 秀樹, and 高安 美佐子. 企業間取引に基づいたネットワーク成長モデルの構築. *素粒子論研究*, 117(5):E112–E113, December 2009.
- [54] 三浦 航, 渡邊 隼 史, 水野 貴之, 高安 秀樹, and 高安 美佐子. 企業間取引に基づいたネットワーク成長モデルの構築. *物性研究*, 93(5):702–703, February 2010.
- [55] 高安 美佐子, 三浦 航, and 田村 光太郎. 企業取引ネットワークの形成と特性. *システム制御情報: システム制御情報学会誌*, 56(10):517–522, October 2012.
- [56] Wataru Miura, Hideki Takayasu, and Misako Takayasu. Effect of Coagulation of Nodes in an Evolving Complex Network. *Phys. Rev. Lett.*, 108(16):168701, April 2012.
- [57] Colm Connaughton, R Rajesh, and Oleg Zaboronski. Stationary Kolmogorov solutions of the Smoluchowski aggregation equation with a source term. *Phys. Rev. E*, 69(6):061114, June 2004.
- [58] F Clementi, T Di Matteo, and M Gallegati. The power-law tail exponent of income distributions. *Physica A: Statistical Mechanics and its Applications*, 370(1):49–53, October 2006.
- [59] Aaron Clauset, Cosma Rohilla Shalizi, and M E J Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, 2009.
- [60] Yannick Malevergne, Vladilen Pisarenko, and Didier Sornette. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Phys. Rev. E*, 83(3):036111, 2011.
- [61] Shouji Fujimoto, Atushi Ishikawa, Takayuki Mizuno, and Tsutomu Watanabe. A New Method for Measuring Tail Exponents of Firm Size Distributions. *Economics E-Journal*, 2011.
- [62] T Di Matteo, T Aste, and M Gallegati. Innovation flow through social networks: productivity distribution in France and Italy. *Eur. Phys. J. B*, 47(3):459–466, October 2005.
- [63] Yukiko Umeno Saito, Tsutomu Watanabe, and Mitsuru Iwamura. Do larger firms have more interfirm relationships? *Physica A: Statistical Mechanics and its Applications*, 383(1):158–163, 2007.
- [64] Misako Takayasu, Sinjiro Sameshima, Takaaki Ohnishi, Yuichi Ikeda, Hideki Takayasu, and Kunihiro Watanabe. Massive Economics Data Analysis by Econophysics Methods—The case of companies’ network structure. *Annual Report of the Earth Simulator Center*

April, 2008, 2007.

- [65] C Guilmi, F Clementi, T Di Matteo, and M Gallegati. Social networks and labour productivity in Europe: an empirical investigation. *J Econ Interac Coord*, 3(1):43–57, March 2008.
- [66] Takaaki Ohnishi, Hideki Takayasu, and Misako Takayasu. Hubs and authorities on Japanese inter-firm network: characterization of nodes in very large directed networks. *Progress of Theoretical Physics Supplement*, 179:157–166, 2009.
- [67] Takaaki Ohnishi, Hideki Takayasu, and Misako Takayasu. Network motifs in an inter-firm network. *J Econ Interac Coord*, 5(2):171–180, June 2010.
- [68] T Iino, K Kamehama, H Iyetomi, Y Ikeda, T Ohnishi, H Takayasu, and M Takayasu. Community structure in a large-scale transaction network and visualization. *J. Phys.: Conf. Ser.*, 221:012012, June 2010.
- [69] Y Fujiwara and H Aoyama. Large-scale structure of a nation-wide production network. *Eur. Phys. J. B*, 77(4):565–580, September 2010.
- [70] Koutarou Tamura, Wataru Miura, Misako Takayasu, Hideki Takayasu, SATOSHI KITAJIMA, and HAYATO GOTO. ESTIMATION OF FLUX BETWEEN INTERACTING NODES ON HUGE INTER-FIRM NETWORKS. *Int. J. Mod. Phys. Conf. Ser.*, 16:93–104, January 2012.
- [71] Hayafumi Watanabe, Hideki Takayasu, and Misako Takayasu. Biased diffusion on the Japanese inter-firm trading network: estimation of sales from the network structure. *New J. Phys.*, 14(4):043034, April 2012.
- [72] Eduardo Viegas, Misako Takayasu, Wataru Miura, Koutarou Tamura, Takaaki Ohnishi, Hideki Takayasu, and Henrik Jeldtoft Jensen. Ecosystems perspective on financial networks: Diagnostic tools. *Complexity*, 18:34–48, July 2013.
- [73] Uri Alon. *An Introduction to Systems Biology*. Design Principles of Biological Circuits. CRC Press, July 2006.
- [74] Royi Itzhack, Yelena Mogilevski, and Yoram Louzoun. An optimal algorithm for counting network motifs. *Physica A: Statistical Mechanics and its Applications*, 381:482–490, July 2007.
- [75] Sergei Maslov and Kim Sneppen. Specificity and Stability in Topology of Protein Networks. *Science, New Series*, 296(5569):910–913, May 2002.
- [76] R L Axtell. Zipf Distribution of U.S. Firm Sizes. *Science*, 293(5536):1818–1820, September 2001.
- [77] Yoshi Fujiwara, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, and Wataru Souma. Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A: Statistical Mechanics and its Applications*, 335(1):197–216, 2004.
- [78] Linton C Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, January 1978.

- [79] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.
- [80] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.
- [81] Ayman Farahat, Thomas LoFaro, Joel C Miller, Gregory Rae, and Lesley A Ward. Authority Rankings from HITS, PageRank, and SALSA: Existence, Uniqueness, and Effect of Initialization. *SIAM J. Sci. Comput.*, 27(4):1181–1201, January 2006.
- [82] Robert Sedgewick. *Algorithms in C++ Part 5. Graph Algorithms*. Pearson Education, December 2001.
- [83] 独立行政法人中小企業基盤整備機構. 企業倒産調査年報. 調査報告書, pages 1–138, August 2006.
- [84] Yoshi Fujiwara. CHAIN OF FIRMS' BANKRUPTCY: A MACROSCOPIC STUDY OF LINK EFFECT IN A PRODUCTION NETWORK. *Advances in Complex Systems*, 11(05):703–717, 2008.
- [85] D Stauffer and A Aharony. *Introduction To PERCOLATION THEORY*. Taylor & Francis, 1994.