



RIETI Discussion Paper Series 11-J-024

取引と特許共同出願の関係について

井上 寛康
大阪産業大学

玉田 俊平太
経済産業研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<http://www.rieti.go.jp/jp/>

取引と特許共同出願の関係について¹井上寛康² (大阪産業大学)

玉田俊平太 (経済産業研究所・関西学院大学)

要 旨

リーマンショック以降も経済の混乱は続いているが、とりわけ日本の経済は他国よりも回復が遅れている。日本では、選択と集中を合言葉に、基礎研究や多角化した事業を削ることによって厳しい経営環境に対応してきた。しかしそれは同時に、大企業単独でのイノベーションが困難になってきたことも意味する。そのため、今や複数の企業、あるいは大学・公的研究機関を巻き込んだオープンなイノベーション・ネットワークの構築が不可欠なものとなってきている。

本研究では、企業間の取引（ニーズ）と特許の発生（シーズ）に着目し、現在の日本の企業間イノベーションシステムがどのように構築されているかを分析した。データとして100万の企業をノードとし、それらにおける取引と特許の共同出願によるリンクからなる多重ネットワークを分析した。分析に、産業連関表、ERGモデル、ベイジアンネットワークの3つを用いた。

産業連関表の分析から、取引金額よりも取引件数の方が共同出願により影響を与えているであろうことが推測された。続いてERGモデルに基づく分析から、企業間の取引は双方向になり、取引と特許の共同出願は同時に発生しやすいことがわかった。またこれより複雑な関係はあまり有意に発生しないこともわかった。最後にベイジアンネットワークに基づく分析から、企業間の取引関係が判明すれば、特許の共同出願と産業の種類は独立になることがわかった。

キーワード：イノベーション、取引、共同出願、ネットワーク

JEL classification: O31, O32, O34

RIETI ディスカッション・ペーパーは、専門論文の形式でまとめられた研究成果を公開し、活発な議論を喚起することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、(独)経済産業研究所としての見解を示すものではありません。

¹ 本稿は、経済産業研究所における「多重ネットワーク分析指標を用いた新たな経済指標の検討」プロジェクトの研究成果の一部である。本稿の作成にあたっては、及川耕造理事長、藤田昌久所長、長岡貞男研究主幹、森川正之副所長、富田秀昭研究コーディネーターほかDP検討会参加者から貴重なコメントをいただいた。本研究は科研費(20330060)の研究助成を受けた。ここに篤く感謝申し上げる。なお、残された誤りは筆者達の責任に帰する。

² inoue@dis.osaka-sandai.ac.jp

1 はじめに

現代的な企業は国際的競争と不安定な経済状況に備え、従前のように多くの事業を取りそろえてシナジー効果を目指す経営ではなく、より事業を絞って専門性を指向した経営にシフトしている。結果として、多くの研究開発工程が企業単体で行われるのではなく、それぞれの企業が持つ強みを合わせることで遂行されるようになった。

直感的には、企業が何か技術革新を起こして、以前より物あるいはサービス（以降ではこれらを単に物と呼ぶ）が売れるという因果は理解しやすいであろう。しかし一方で、企業がどういった技術革新をすべきであるのかについての情報は、その対象となる物が先に顧客に使われていないとつかめないう因果もある。これらの因果における技術革新と取引の順序は全く逆であり、これらのどちらが先であるのかというのは古くからある問いである。上述でふれたように、いかに戦略的に共同研究開発を行うかが大事ということから考えても、この問いは再び重要性を増しているといえる。

そこで本論文では、企業間で行われる共同研究開発と取引の関係性について分析する。ここで用いられるデータは、数十万の日本企業を含んでおり、主要な企業活動はほぼすべて網羅されている。本論文は、まずこの企業間のネットワークの構造を分析し、つづいて産業連関表、ERGモデル、ベイジアンネットワークの3つの分析手法に基づいた分析について報告する。

第2節では本論文で扱うデータについて述べる。第3節では産業連関表、第4節ではERGモデル、第5節ではベイジアンネットワークの分析結果について述べる。最後に第6節では結論を述べる。

2 データ

本論文では2つの異なるデータが用いられている。まず1つ目は企業間の取引に関するデータである。このデータは、東京商工リサーチにより採取された、2005年における961,363企業間の7,808,760件の取引関係である。このデータからネットワークを作ることができる。このネットワークは企業をノードとし、取引関係を有向リンクとする。リンクの向きは金の流れである。図1は次数分布である。ここで次数とは、あるノードに入ってくる、あるいは出てくるリンクの両方を数えたものである。横軸は次数を表し、縦軸はランクを表している。両軸とも対数表記である。この図1は大まかには直線でフィットできそうに見える。したがって正規分布のような典型的なピークを持つ分布ではないと理解できる。

もう一方のデータは特許の共同出願データである。このデータは1993年から2006年までの公開特許公報に含まれる5,570,786特許を用いている。本論文の目的は共同研究開発と取引の関係を知ることであるが、本論文ではこの共同出願特許によって共同研究開発があるとする。すべての共同研究開発が共同出願特許にはならないが、日本の企業全体の傾向として考えるのであれば問題ないと考えられる。特許に含まれる出願人は個人やさまざまな法人が含まれるが、ここでは法人名に「株式会社」を含む企業のみを抽出した。

このデータを用いてネットワークを作成した。先ほどと同じく、企業がノードであり、共同出願を無向リンクとするネットワークである。もし3つ以上の企業が1つの特許を共同出願している場合は、それらの企業を完全グラフの要領でつないである。このネットワークは、54,197ノード、154,205リンクを含んでいる。このネットワークの次数分布は図2のようになる。グラフの意味は図1と同じである。

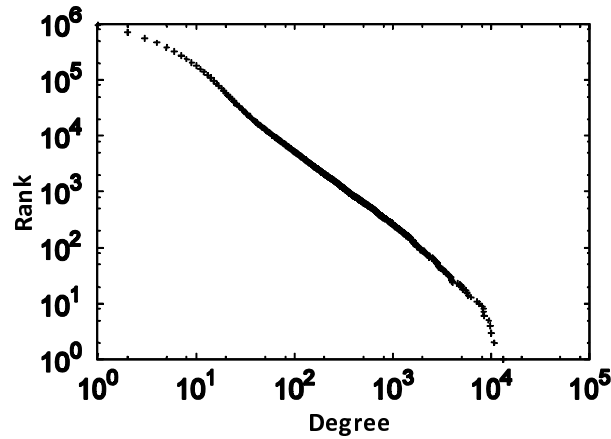


図 1: Degree distribution by rank in transaction network. Horizontal axis indicates degree and vertical axis indicates rank counted from highest degree.

Inoue らはこの後者のネットワークについてすでに分析している [1] . この分析ではこの次数分布がべき分布であることを述べている .

べき分布の分析には単純な最小二乗法を用いられることが多く , その傾きだけが議論されている . しかしながら , この方法には 2 つの欠陥がある . 1 つはどのような分布でもべき分布としての傾きを得ることができ , べき分布であるのかどうかに何の知見ももたらさない . もう 1 つは , どの部分がべき分布であるかを議論すべきであるのに , 分布全体にわたって検証するだけである . この欠陥を補ったのが Clauset らの方法 [2] であり , goodness-of-fit テストとコルモゴロフ-スミルノフ検定を組み合わせた方法である .

ある確率分布がべき分布であるとする と , 離散の場合は次のような式になる .

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})},$$

ここで

$$\zeta(\alpha, x_{\min}) = \sum_{n=0}^{\infty} (n + x_{\min})^{-\alpha},$$

である . また x_{\min} は下限であり , α は傾きを表すパラメータである . 下限が必要なのは , べき分布が $x = 0$ で発散するためである . この式に対応する累積確率分布は

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{\min})}.$$

となる . Clauset らの方法 [2] を用いた結果として , この共同出願のネットワークはべき分布であり , その時のパラメータは α が 2.03 であり , x_{\min} が 7 であった .

ここまで示した 2 つのネットワークは異なるネットワークであるが , どちらも企業をノードとする点で共通している . そこでこれらのネットワークを重ね合わせて 1 つのネットワークとする . このとき , 同一の企業は 1 つのノードとした . この判別は , 企業名と住所を用いた . 企業名だけではどちらのネットワークにおいても複数の企業が存在するが , 住所と組み合わせた場合はこの問題が発生しなかった . リンクについては区別したままであるので , 1 種類のノードと 2 種類のリンクが存在する多重ネットワークとなる . この多重ネットワークは 975,607 ノードを含んでいる .

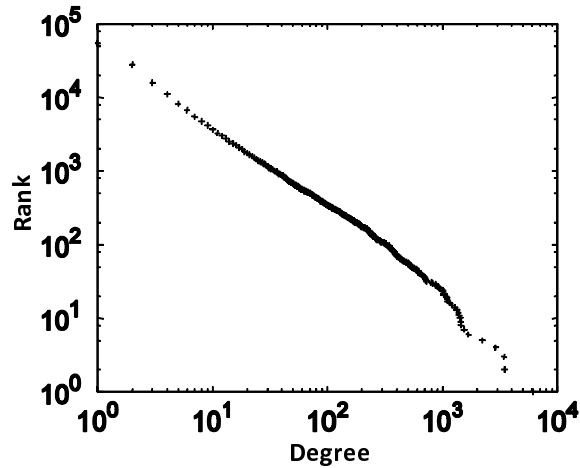


図 2: Degree distribution by rank in joint-patent application network. Horizontal axis indicates degree and the vertical axis indicates rank counted from highest degree.

3 産業連関表

本論文ではまずこの多重ネットワークを産業連関表として分析する。産業連関表は Leontief[3] によって開発され、産業間の資金の流れを行列で示したものである。ある産業での需要の発生が他の産業にどのように波及するかを予測するのに強力なツールであり、広く普及している。

本論文の多重ネットワークは、オリジナルの産業連関表が扱う金の流れの総額とは異なり、金の流れのリンクの数、共同出願のリンクの数が把握できる。したがって、本来の産業連関表とは異なる別の 2 つの行列を作ることができる。これら 3 つの行列を比較することによって、共同出願と関係の深い経済活動は、取引の総額であるのか、取引の件数であるのかなどを分析することができる。

本論文の目的は、企業間の関係性について議論することであるが、この産業連関表の単位は産業である。ここで企業の集まりである産業を扱うのは企業間関係の大局的な傾向を把握するためである。

表 1 は簡単な産業連関表の例である。この表は電気機械と情報・通信機器の産業の間の取引の関係を示している。行方向は取引において売った金額を、列方向は買った金額をそれぞれ示している。したがって、たとえば電気機械から情報・通信機器に 354,773 百万円の取引があると読み取ることができる。

本来の産業連関表の行列においては、行の産業から列の産業に流れた金額が記録されている。この行列を M_a とする。統計局で公開されている産業連関表には多くの種類があるが、ここでは、34 産業に分けられた産業大分類を用いる。

本論文で新たに作成する 2 つの行列は、34 産業間の金の流れのリンク数の行列と、共同出願のリンク数の行列である。取引の金額に基づいた産業連関表との違いを示すために図 3 を用いる。左が産業連関表の例である。行列の各要素はこのように産業間の関係の大きさを示している。同様に図の右上は産業間の取引の件数を表している。これに対応する行列を M_t とする。図の右下は産業間の特許共同出願の件数を表している。これに対応する行列を M_p とする。

図 4 は M_a と M_p の散布図である。行と列の同じ産業間成分 (i, j) すべてについて、

$$(x, y)_{(i, j)} = (M_{a_{ij}}, M_{p_{ij}}).$$

表 1: Pedagogic example of input-output tables. Numbers are flows for total amount of money. Money flows from industries in rows to ones in columns.

	Electrical machinery	Information and communication machinery
Electrical machinery	1,938,510 million yen	354,773 million yen
Information and communication machinery	2,394 million yen	422,096 million yen

としたとき、この (x, y) の値をそれぞれ、散布図の横軸と縦軸に対応させている。同様にして、図 5 は M_t と M_p の散布図である。そして Fig. 6 は M_t と M_a の散布図である。

各プロットに対応して、スピアマンの順位相関係数を計算したところ、0.27 (M_a と M_p)、0.66 (M_t と M_p)、0.31 (M_t と M_a) となった。また、ピアソンの積率相関係数を計算したところ、0.31 (M_a と M_p)、0.45 (M_t と M_p)、0.05 (M_t と M_a) となった。

M_a と M_p 、および M_t と M_a における順位相関係数と積率相関係数は値の順序が変わってしまっている。また、 M_t と M_a の積率相関係数においては、有意水準 5% で母相関係数が 0 であることは棄却されないが、順位相関係数では棄却される。これらのことから正規分布を仮定した積率相関係数よりも、順位相関係数を議論する方が安全である。

順位相関係数において M_a と M_p は 0.27、 M_t と M_p は 0.66 であり、有意水準 5% でそれぞれに対する母相関係数が同じであることは棄却される。

この結果からわかるように産業間の共同出願は、取引の総額ではなく取引の件数とより深い関係が

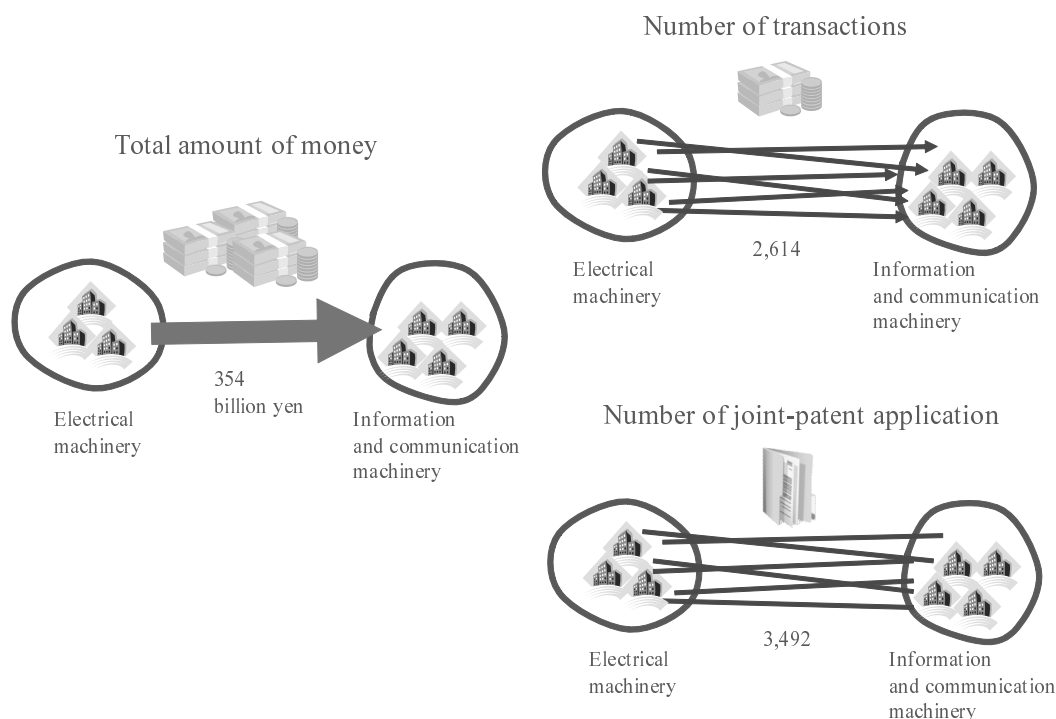
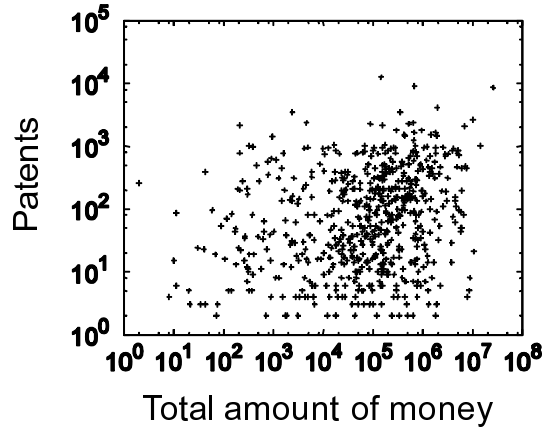
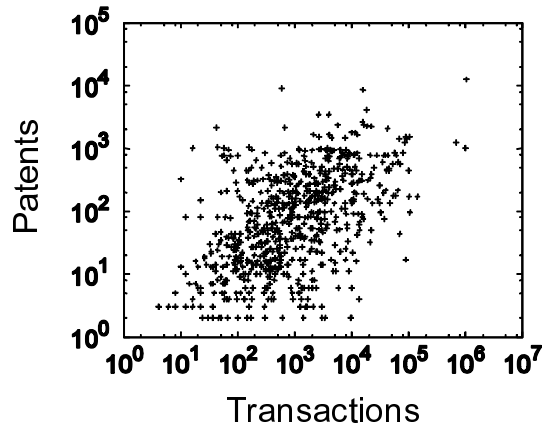


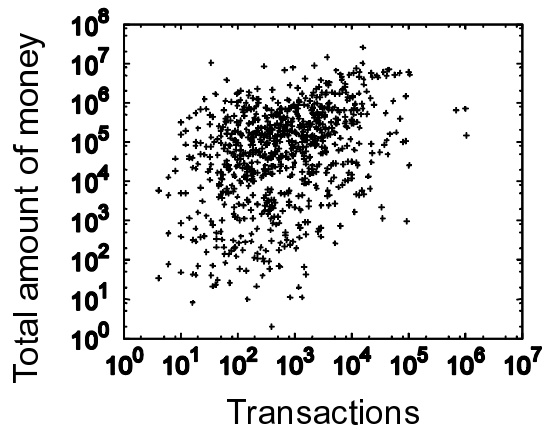
図 3: Difference between three matrices. These figures are examples and indicate how value of each matrix element is determined.



⊠ 4: Scatter plot for M_a and M_p . Each plot corresponds to elements that have same position in M_a and M_p . Spearman rank correlation coefficient is 0.27. Pearson product-moment correlation coefficient is 0.31.



⊠ 5: Scatter plot for M_t and M_p . Spearman rank correlation coefficient is 0.66. Pearson product-moment correlation coefficient is 0.45.



⊠ 6: Scatter plot for M_t and M_a . Spearman rank correlation coefficient is 0.31. Pearson product-moment correlation coefficient is 0.05.

あることがわかる。したがって、取引の大きさによって技術の知識が流れているというよりも、取引のパイプの太さによって知識が流れているという方がより正しい理解である。

なお、 M_p だけ対称行列であることから、上記の相関係数の議論は注意を要する。比較される2つの行列がともに非対称か対称であれば問題ない。しかしここでは M_p だけが対称行列であるので、この行列の成分と M_t (あるいは M_a) の2つの異なる行列成分を対応させることになる。これは説明変数のある値に対して被説明変数の異なる値が2つ存在する状況とも考えられる。この問題については逆のこと、すなわち対称行列の成分は被説明変数、非対称行列の成分は説明変数の方であるとすれば、解釈上は問題ない。また計算においても説明変数と被説明変数に扱いの違いはないので問題はない。

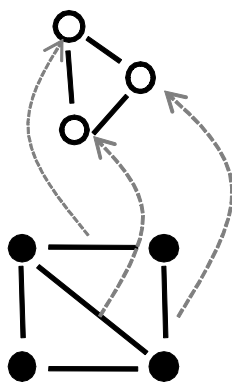
4 ERG モデル

多重ネットワークに対する第2の分析として、ERG モデルによる分析を行った [4]。ERG モデルとは、ネットワークにおいて有意に現れる部分構造を明らかにする手法である。

前節では産業間のレベルで共同出願と取引件数の関係が深いことが把握されたので、本節では企業間のレベルで共同出願と取引の関係は重複しているのか、あるいはもっと複雑な構造を持っているのか調べる。

ERG モデルは依存グラフにより説明できる [5]。依存グラフはあるネットワークのメタネットワークと考えるとわかりやすい。依存グラフは、あるネットワークのリンクをノードとするネットワークである。図7は依存グラフの例である。黒い点の(この例では無向)ネットワークは分析対象のネットワークであり、白い点のネットワークが依存グラフである。図のように黒い点のネットワークのリンクに対して、白い点のネットワークのノードが対応している。依存グラフは常に完全グラフである。そして依存グラフは、分析対象のネットワークの部分構造を表している。この図7では分析対象のネットワークのリンクは無向であるが、有向である場合は区別される。

Dependence network



Existing network

図 7: Example of dependence graph. Node in dependence graph corresponds to link in actual networks.

ERG モデルにおいては、分析対象のネットワークは、単に確率的にリンクが発現したネットワークの1つにすぎないを考える。それに対応する表記として、あるネットワークを x と表す。これを用い

であるネットワークは

$$P(x) = \kappa^{-1} \exp\left(\sum_{A \subseteq N_D} \lambda_A z_A(x)\right), \quad (1)$$

と表される．このとき， $\kappa = \sum_x \exp(\sum_{A \subseteq N_D} \lambda_A z_A(x))$ であり，正規化定数である．また， A はノードの部分集合， N_D は依存グラフのノード， x_{ijm} はネットワーク m におけるノード i から j へのリンクである．さらに， $z_A(x) = \prod_{x_{ijm} \in A} x_{ijm}$ である．ネットワーク m とはネットワークの種類（ここでは共同出願と取引）である．また， λ_A は係数である．ERG モデルのエッセンスは，実在するネットワーク x が発現する確率 $P(x)$ を最大化するような λ_A を見つけることである．

ERG モデルにおけるこの式 (1) は，大変冗長である．なぜならあらゆる部分構造（依存グラフ）を含んでいるためである．そのため，このままでは計算量が膨大であるとともに，そもそも一般的には 1 つしかないネットワークのデータに対して λ_A を見つけることは意味がない．そこで，同形のものと同じ係数 $\lambda_{[A]}$ を見つけることによって，この問題を回避する．

上記を反映すると式 (1) は次のように修正される．

$$P(x) = \kappa^{-1} \exp\left(\sum_{[A]} \lambda_{[A]} z_{[A]}(x)\right). \quad (2)$$

この式に対して $P(x)$ を最大化する λ_A を見つけることになるが，その方法にはいくつかある．1 つは maximum likelihood estimation [6, 7] であり，もう 1 つは Markov chain Monte Carlo methods [8] である．これらは非常によく研究されているが，計算が複雑であったり計算時間・計算空間が大きいという問題がある．そこで疑似尤度を用いた方法 [9, 10, 11, 12] があり，精度は悪くなく，計算も簡単のため，本論文ではこれを用いる．

疑似尤度による計算の説明には以下の表記が必要となる．

$$\begin{aligned} \mathbf{X}_{ijm}^C &= \{\mathbf{X}_{klh} : \mathbf{X}_{klh} \in N_D \text{ for all} \\ &\quad (k, l, h) \neq (i, j, m) \text{ and} \\ &\quad \mathbf{X}_{ijm} \text{ is undefined}\}, \\ \mathbf{x}_{ijm}^+ &= \{x_{klh}^+ : x_{klh}^+ = x_{klh} \text{ for all} \\ &\quad (k, l, h) \neq (i, j, m) \text{ and} \\ &\quad x_{ijm} = 1\}, \text{ and} \\ \mathbf{x}_{ijm}^- &= \{x_{klh}^- : x_{klh}^- = x_{klh} \text{ for all} \\ &\quad (k, l, h) \neq (i, j, m) \text{ and} \\ &\quad x_{ijm} = 0\}. \end{aligned}$$

これらを用いるとネットワーク m におけるノード i から j のリンクの条件付き確率は

$$\begin{aligned} &P(X_{ijm} = 1 \mid \mathbf{X}_{ijm}^C) \\ &= \frac{P(\mathbf{X} = \mathbf{x}_{ijm}^+)}{P(\mathbf{X} = \mathbf{x}_{ijm}^+) + P(\mathbf{X} = \mathbf{x}_{ijm}^-)} \\ &= \frac{\exp \sum_{[A]} \lambda_{[A]} z_{[A]}(\mathbf{x}_{ijm}^+)}{\exp \sum_{[A]} \lambda_{[A]} z_{[A]}(\mathbf{x}_{ijm}^+) + \exp \sum_{[A]} \lambda_{[A]} z_{[A]}(\mathbf{x}_{ijm}^-)} \end{aligned}$$

となる．
疑似尤度は

$$PL(\lambda) = \prod_{i \neq j} \prod_{m=1}^T P(X_{ijm} = 1 \mid \mathbf{x}_{ijm}^C)^{x_{ijm}} P(X_{ijm} = 0 \mid \mathbf{x}_{ijm}^C)^{1-x_{ijm}}$$

となる．疑似尤度を最大化する係数 λ はロジスティック回帰分析により求まることが示されている [11] ．

デビアンスはモデルのあてはまりの良さを表すのに使われる．ここでは疑似尤度 (Pseudo Likelihood) を用いることから，デビアンスを G_{PL}^2 と表す．説明変数を増やすことでデビアンスが小さくなることは明らかであるから，どの程度 G_{PL}^2 が小さくなったときに，その説明変数が有意なのかを決める必要がある．ここでは先行研究にならって，その大きさを $2n(n-1)r \log(1-\delta)$ とする [5]．ここで n はノードの数， r はネットワーク層の数， δ はパラメータである．前述の先行研究では δ は 0.001 か 0.005 がよいとされているので，ここでは δ を 0.001 とする．

本論文では ERG モデルにおいて調べる部分構造を，次の 5 つの部分構造とする．それらは，Choice，Multiplicity，Reciprocity，Multi-reciprocity と Transitivity である．ただし，Choice は比較の他の部分構造すべてと組み合わせ，有意性の確認に用いられる．これらは図 8 に示してある．これらよりもっと複雑な構造も考えられるが，それらについては調べる必要はないと分析の結果わかった．この理由については後で述べる．先行研究 [5] では多重ネットワークのリンクはいずれも有向リンクであったが，ここでは無向リンクであるので同じ部分構造でも取りうるパターンは少ない．

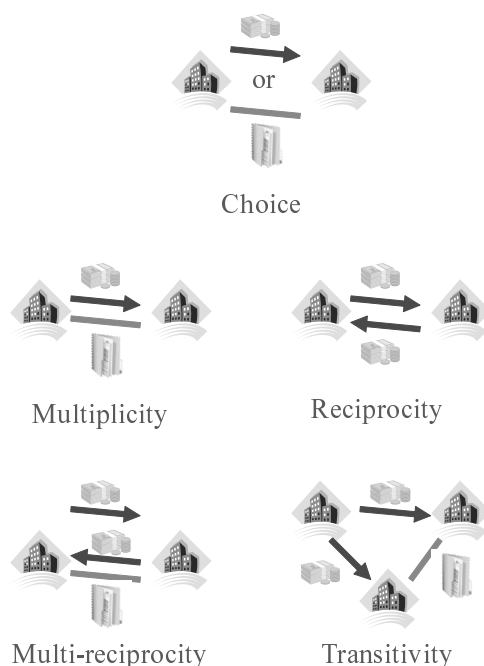


図 8: Basic and four other configurations. Possibility of these configurations is investigated in the analysis.

Choice は 2 つのノードの間にリンクが存在することを表す．したがって，この部分構造には 2 つの λ が必要である．このモデルのデビアンスを基準として，他のモデルが α を越える場合にそのモデルが有意であると判断する．

Multiplicity は 2 つのノードの間に両方のタイプすなわち，取引と特許共同出願のリンクが存在することを表す．したがって λ は 1 つである．Reciprocity は双方向の取引のリンクが存在することを表す．したがってこれも λ は 1 つである．Multi-reciprocity は Reciprocity に加えて，特許共同出願のリンクが存在することを表す．したがってこれも λ は 1 つである．

Transitivity は上記と違って，3 つのノード間での部分構造である．図 8 のように，ある 1 つのノードから取引のリンクが出ており，出た先の 2 つのノードの間で特許共同出願のリンクが存在すること

を表す．他にも多くの3つのノード間の部分構造があるが，それは調べる必要がなかった．それについては後で述べる．

ERGモデルを使ってネットワーク全体を分析することは計算時間・空間的に不可能であるが，それ以上に，ある構造が有意に現れたとしてもその結果を解釈することが難しい．そこでここでは，前節で用いた産業大分類を用いて34個のネットワークに分ける．そうすることで，産業間による違いを議論することができる．産業大分類にネットワークを分けるときに，産業間のネットワークはすべて消去した．また，それでもノード数が多いため，共同出願のリンクにおいて最大次数のノードから到達可能なノードを1ステップずつ増やし，初めて1,000ノードを越えたときにそれらのノードとそれらのノード間のリンクだけを抜き出した．

表 2: G_{PL}^2 acquired with the p^* model (First 20 industries)

Industry	Nodes	α	Choice	Multiplicity	Reciprocity	Multi-reciprocity	Transitivity
(01) Agriculture, forestry and fishery	5	-	-	-	-	-	-
(02) Mining	72	8.9	2,282.8	2,230.6*	2,212.6*	2,269.3*	2,281.1
(03) Foods	355	218.4	20,054.6	18,453.7*	19,209.3*	19,839.7	19,928.6
(04) Textile products	381	251.6	2,2823.9	21,923.0*	22,137.5*	22,744.5	22,731.4
(05) Pulp, paper and wooden products	242	101.4	10,872.0	10,133.7*	10,547.3*	10,799.2	10,784.2
(06) Chemical products	1,104	2,116.4	112,351.0	107,759.2*	108,230.3*	11,562.7	110,612.2
(07) Petroleum and coal products	119	24.4	5,077.1	4,847.1*	4,981.8*	5,057.3	4,987.4*
(08) Ceramic, stone and clay products	711	877.4	58,411.0	55,990.4*	56,533.3*	58,220.1	57,566.6
(09) Iron and steel	636	701.9	52,940.0	50,680.0*	51,835.0*	52,926.0	52,747.2
(10) Non-ferrous metals	521	470.9	43,581.2	42,079.3*	42,356.9*	43,390.6	43,101.6*
(11) Metal products	875	1,329.1	72,567.8	67,557.4*	70,268.63*	71,690.4	70,830.1*
(12) General machinery	1,554	4,194.5	166,446.0	158,330.2*	162,082.9*	165,777.5	163,636.7
(13) Electrical machinery	1,228	2,618.8	128,806.6	122,003.8*	124,843.3*	128,618.9	127,615.6
(14) Information and communication machinery	499	431.9	39,865.0	37,677.1*	38,530.7*	39,763.2	39,586.67
(15) Electrical equipment	503	438.9	36,666.1	34,513.2*	35,566.3*	36,604.7	36,253.8
(16) Transportation equipment	1,109	2,135.7	127,329.5	117,616.6*	123,611.1*	126,817.6	125,270.5
(17) Precision instruments	352	214.7	21,416.5	20,427.3*	20,659.3*	21,144.1*	21,120.2*
(18) Miscellaneous manufacturing products	1,096	2,085.9	93,997.5	89,052.7*	90,924.2*	93,542.2	92,539.2
(19) Construction	1,021	1,810.0	129,410.3	123,297.5*	127,415.4*	129,206.0	127,446.1*
(20) Electricity, gas and heat supply	384	269.1	33,457.3	31,895.1*	32,743.0*	33,293.9	32,659.4*

分析の結果は表 2 および表 3 のようになった。いくつかの産業はノードが足りなかったので省かれている。基準として 10 ノード以下の産業は省いた。

すべての産業において、Multiplicity と Reciprocity は Choice と比較して有意であるとわかった。Multiplicity は取引と共同出願のリンクが同時に現れることを表している。また、Reciprocity は取引は 2 つのノードの間で相互に行われることを表している。

一方で、Multi-reciprocity と Transitivity はいくつかの産業でのみ有意となった。Multi-reciprocity は Multiplicity と Reciprocity が混ざった形である。Multi-reciprocity は、鋳業、精密機械、水道・廃棄物処理、金融・保険、情報通信、教育・研究で有意となった。Transitivity は、石油・石炭製品、非

表 3: G_{PL}^2 acquired with the p^* model (Last 14 industries)

Industry	Nodes	α	Choice	Multiplicity	Reciprocity	Multi-reciprocity	Transitivity
(21) Water supply and waste management services	17	0.47	263.7	260.6 *	261.5 *	257.4 *	257.9 *
(22) Commerce	1,185	2,438.6	113,923.8	107,206.5 *	109,978.5 *	112,926.2	111,329.7 *
(23) Financial and insurance	257	114.3	13,985.6	13,766.6 *	13,619.7 *	13,852.2 *	13,890.5
(24) Real estate	136	31.9	5,686.5	5,564.9 *	5,572.9 *	5,678.1	5,677.9
(25) Transport	190	62.4	9,721.0	9,353.9 *	9,539.0 *	9,683.7	9,450.1 *
(26) Communication and broadcasting	316	173.0	20,061.1	18,471.19 *	19,515.4 *	19,745.45 *	19,146.87 *
(27) Public administration	0	-	-	- *	- *	-	-
(28) Education and research	55	5.2	1,827.5	1,821.6 *	1,718.7 *	1,674.1 *	1,813.7 *
(29) Medical service, health and social security and nursing care	4	-	-	-	-	-	-
(30) Other public services	0	-	-	-	-	-	-
(31) Business services	639	708.6	45,513.8	43,030.3 *	44,346.8 *	45,182.0	44,674.8 *
(32) Personal services	5	-	-	-	-	-	-
(33) Office supplies	0	-	-	-	-	-	-
(34) Activities not elsewhere classified	0	-	-	-	-	-	-

鉄金属，金属製品，輸送機械，建設，電力・ガス・熱供給業，商業，運輸，情報通信，教育・研究，対事業所サービスで有意となった．元の多重ネットワークを産業大分類で分割したのは，産業間で有意な部分構造が現れることを期待したためであった．しかしながらこの結果を見る限り，産業間に共通性を見出すことは難しい．たとえば一次，二次，三次といった産業で分けることはできない．あるいはインフラ・サービスと製造業といった区別もできない．したがってこの分析では，特定の産業に特有の部分構造があるという仮説を裏付ける検証に至らなかった．前述した，3つのノードにおいてこれ以上複雑な構造を調べる必要がない，というのはこのような理由である．

一方で，Multiplicityの結果は有用である．前述の通りすべての産業で有意であると示されたが，取引と共同出願は同時に現れやすい，すなわち相関がやはりあるということである．前節での産業間での結果をより強化している．この関係をより深く分析するため，次節ではその因果について分析を行う．

5 ベイジアンネットワーク

ベイジアンネットワークは変数間の因果関係に基づく推測をする分析手法である．ベイジアンネットワークは Pearl[13] によって提案された．本節では前の2つの節の結果を受けて，企業間の取引関係と共同出願関係はどちらが先行するのかということ調べる．注意すべき点として，ベイジアンネットワークで扱うネットワークと本論文でここまで扱ってきた多重ネットワークは全く関係のない別物である．

ベイジアンネットワークはグラフィカルモデリングであり，変数間の同時確率分布を表現する．アプリケーションについてはたとえば Lauritzen[14] に詳しい．ベイジアンネットワークは次の3つの用途に用いることができる (1) 観測されていない事象の確率を推定する (2) 確率分布を推定する (3) 変数間の構造を推定する．ここでは (3) の目的でベイジアンネットワークを用いる．

構造の推定方法は Rebane and Pearl[15] によって提案されている．Deal はその推定方法に基づいて Böttcher and Dethlefsen[16] によって作られた R の関数である．

図9はベイジアンネットワークの構造の推定に用いられた変数を表している．Industry A と B は企業が属する産業 (1 から 34) であり，Patent link は企業の間にも共同出願があるか (0/1) ，Transaction link は企業の間にも取引があるか (0/1) である．他の変数を入れることもできるが，このようにシンプルにすることで結果の解釈もしやすい．

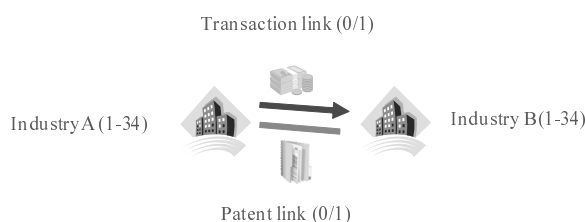


図 9: Parameters for structure learning of Bayesian network.

多重ネットワークは 975,607 ノードあるので，2つの企業の総当たりは $975,607 \times (975,607 - 1) / 2$ であり，計算するのに多すぎる．そこで次の4ステップで数を減らした (1) 1つのノードをランダムに選ぶ (2) そのノードに取引か共同出願でつながったノードを選ぶ (3) (1) と (2) のすべてのノード間の組み合わせをデータとする (4) (1) から (3) を必要なデータ数が得られるまで行う．この作業が必要なのは，ランダムに選んだ2つのノードが結び付く可能性は極めて0に近いためである．あらかじめリンクでつながれたノードを選ぶことが，直感的には恣意的であると思われるが，ベイジアンネットワークの性質を考えればこの選択は問題ない．この方法で 109,641 件のデータを得た．

図10が構造推定の結果である．この図から直感的に Industry A と B が Transaction link の原因であり，Transaction link が Patent link の原因であるように思われる．しかしながら，これは正しくない．構造推定ではこのような構造の場合，ちょうどリンクの向きをすべて逆にした場合と区別がつかない．したがってこの結果からわかることは次の3点である (1) Transaction link と Patent link は相関がある (2) Transaction link と Industry A と B は相関がある (3) もし Transaction link が判明していれば，Patent link と Industry A と B は独立である．残念ながら (1) のように Transaction link と Patent link の間に明確な因果関係を見つけることはできなかった．しかしながら，多くの構造がありうる中で，この構造が推定されたことは価値がある．たとえば，上記の (3) のように共同出願がある企業間に出るかどうかなどというのは，取引の有無に依存するのであり，それら企業がどのような産業であるのかということよりも強い，あるいは産業を無視できるということである．

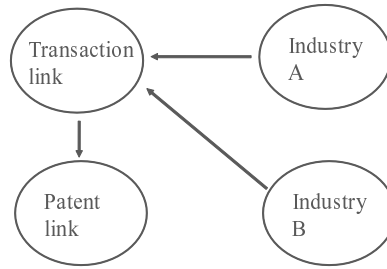


図 10: Bayesian network acquired by structure learning.

6 結論

本論文は、日本の企業間で起きた取引と共同出願の「関係」を調べた。取引に関するデータには 961,363 の企業と 7,808,760 の取引が含まれた。共同出願に関するデータには 54,197 の企業と 154,205 の共同出願が含まれた。これらの 2 つのネットワークを 1 つにまとめ、多重ネットワークとした。そのときノード数は 975,607 であった。

本論文では、この多重ネットワークの大まかな構造を調べるため、度数分布について調べた。つづいて、産業連関表とそれを模した新たな 2 つの表を作成し、比較した。スピアマンの順位相関係数を計算したところ、 $0.27 (M_a \text{ と } M_p)$ 、 $0.66 (M_t \text{ と } M_p)$ 、 $0.31 (M_t \text{ と } M_a)$ となった。取引金額よりも取引件数の方が共同出願により影響を与えているであろうことが推測された。したがって、単純に従来の金額による産業連関表から知のスピルオーバーが起きていることを推測するよりも、取引件数に基づいて推測する方がより望ましいことがわかった。

つづいて、特定の部分構造が有意に現れるかどうかを検証するために、ERG モデルに基づく分析を行った。それらの部分構造は Choice, Multiplicity, Reciprocity, Multi-reciprocity と Transitivity である。この分析では多重ネットワークを産業大分類で分割した。Multiplicity と Reciprocity はすべての産業で確認できたが、Multi-reciprocity と Transitivity は一部の産業にだけ確認され、それら産業間での共通性は確認できなかった。

最後に、取引と共同出願はどちらが先に現れるのかということ进行分析するため、ベイジアンネットワークを用いた。変数として、Industry A と B, Patent link, Transaction link を用いた。計算量を減らすため、データの件数を 109,641 件に減らした。得られた構造から次のことがわかった (1) Transaction link と Patent link は相関がある。(2) Transaction link と Industry A と B は相関がある。(3) もし Transaction link が判明していれば、Patent link と Industry A と B は独立である。

7 政策上の含意

本研究で明らかとなったことの一つは、取引金額の多寡よりも、取引件数の多寡の方が、共同研究開発の件数と相関が高いという事実である。この結果は、ERG モデルを用いた分析でも支持されている。取引関係と特許共同出願関係の両方が起きる確率は、ランダムな場合の確率より有意に高かったのである。この分析結果からは、何度も取引を重ねて信頼関係が醸成された企業同士が共同で研究開発を開始したり、共同研究開発の遂行を通じて信頼関係が構築された企業同士で取引が生じたりするという、企業間の信頼関係の醸成に伴うトランザクションコストの低下、それを通じた企業間関係の

進展という関係が推測される。

ベイジアンネットワーク分析の結果では、企業間の業種の組み合わせという変数が加わってもなお、取引と特許共願の有無という変数の間の結び付きは崩れなかった。イノベーション促進策として異業種のマッチングが一般的に推進されるが、そのように業種の組み合わせに注目することよりも、すでに醸成された取引関係の方が、共願関係の新たな生成に強い裏付けを与えるといえる。企業の間関係には無数の事情があり、他のどのような変数が取引や特許共願の関係に影響を与えるかは今後も議論の余地があるが、少なくとも上記の結果から、取引と特許共願の関係の正の相関性について強い裏付けが得られたといえる。

謝辞

本稿は、経済産業研究所における「多重ネットワーク分析指標を用いた新たな経済指標の検討」プロジェクトの研究成果の一部である。本稿の作成にあたっては、及川耕造理事長、藤田昌久所長、長岡貞男研究主幹、森川正之副所長、富田秀昭研究コーディネーターほか DP 検討会参加者から貴重なコメントをいただいた。本研究は科研費（20330060）の研究助成を受けた。ここに篤く感謝申し上げます。なお、残された誤りは筆者達の責任に帰する。

参考文献

- [1] H. Inoue, W. Souma, and S. Tamada. Analysis of cooperative research and development networks on Japanese patents. *Informetrics*, 4:89–96, 2010.
- [2] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [3] W.W. Leontief. *Input-Output Economics*. Oxford University Press, New York, second edition, 1986.
- [4] S. Wasserman and G. Robins. An introduction to random graphs, dependence graphs, and p^* . In P. J. Carrington, J. Scott, and S. Wasserman, editors, *Models and Methods in Social Network Analysis*, pages 148–161. Cambridge University Press, 2005.
- [5] L. M. Koehly and P. Pattison. Random graph models for social networks: Multiple relations or multiple raters. In P.J. Carrington, J. Scott, and S. Wasserman, editors, *Models and Methods in Social Network Analysis*, chapter 9. Cambridge University Press, New York, 2005.
- [6] S. Wasserman. Conformity of two sociometric relations. *Psychometrika*, 52:3–18, 1987.
- [7] O. Frank and K. Nowicki. Exploratory statistical analysis of networks. In J. Gimbel, J.W. Kennedy, and L.V. Quintas, editors, *Quo Vadis Graph Theory? A Source Book for Challenges and Directions*. Amsterdam, 1993.
- [8] P. J. Carrington, J. Scott, and S. Wasserman, editors. *Models and Methods in Social Network Analysis*. Cambridge university press, 2005.

- [9] J. E. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.
- [10] J. E. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Association*, 47:77–92, 1977.
- [11] D. Strauss and M. Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85:204–212, 1990.
- [12] P. Pattison and S. Wasserman. Logit models and logistic regressions for social networks: II. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52:169–193, 1999.
- [13] J. Pearl. *Causality*. Cambridge University Press, New York, second edition, 2009.
- [14] S.L. Lauritzen. Some modern applications of graphical models. In P.J. Green, N.L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*. Oxford University Press, 2003.
- [15] G. Rebane and J. Pearl. The recovery of causal poly-trees from statistical data. In *Proceedings of 3rd workshop on uncertainty in AI*, pages 222–228, 1987.
- [16] S.G. Bøttcher and C. Dethlefsen. deal: A package for learning bayesian networks. *Journal of statistical software*, 8(Issue 20), 2003.