

質的産業連関表の粗視化とネットワーク指標の頑健性

筑波大学図書館情報メディア系 田村肇

Hajime Tamura
School of Library and Information Media
University of Tsukuba

1. はじめに

本報告は、数年前から情報処理学会や人工知能学会の研究会や全国大会で報告してきた研究に基づいている。

産業連関表は、5年ごとに作成されるが、その間をつなぐものとして、延長表が作成されている。しかし、作成のたびに産業分類などに変更があるため、これまではあまり長期間接続された産業連関表は存在しなかった。そのため、産業連関表を用いて、経済の中長期的なダイナミックスを明らかにすることは難しかった。

産業連関表は、5年ごとに作成されるが、その間をつなぐものとして、延長表が作成されている。しかし、作成のたびに産業分類などに変更があるため、これまではあまり長期間接続された産業連関表は存在しなかった。そのため、産業連関表を用いて、経済の中長期的なダイナミックスを明らかにすることは難しかった。

最近になって、20年間に渡る（1980年～2000年：5年間隔）長期接続産業連関表が作成され、公表されたため、利用が可能になった。（独立行政法人経済産業研究所（RIETI）による）

これまでは、せいぜい、3年分程度しか連続した産業連関表が入手できなかったため、産業連関表が表している経済構造のダイナミズムを明らかにすることは難しく、単年度での分析が中心であった。その意味では、このようはデータが入手できるようになったことは画期的なことである。しかし、視点を変えると、新たな困難に直面する。

産業連関表は、通常、500前後の部門から構成される。したがって、中間取引の部分のみに注目しても500×500のマトリックスになり、非常に高次元のデータとなる。単年度の産業連関表を、ある確率的データ生成過程から得られた一つの標本とみなせば、データの次元が非常に高い割には標本数は過小となる。

2. 高次元小標本データとしての産業連関表

このような高い次元のマトリックスで表現されるようなデータを用いて経済構造をモデル化するためには、多くの変数が必要になる。そのため、少ない標本数では、統計的に全ての変数の係数の推定が行えない。

500×500のマトリックスを一つのデータとみなせば、このデータは500×500次元空間上の点(X_{11} , X_{12} , ..., X_{500} , 500)で表せる。このような点データからなる標本空間をモデル式で表すことにする。いま、回帰モデルを採用するとすれば、例えば

$$X_{11} = \alpha_{00} + \alpha_{12}X_{12} + \alpha_{13}X_{13} + \dots + \alpha_{500, 500}X_{500, 500} + U$$

と表せる。ここでUは、攪乱項。このモデル式のパラメータ α を推定しようとするれば、500×500個より多くの標本データが必要になるのは自明。

このように、産業連関表データのダイナミズムの分析（時系列分析）を行おうとすれば、統計的に異常な事態に直面することになる。

単年の産業連関表を一つの高次元データとみなしたところ、標本数が少ないため、推定しようとする産業構造の多変量モデルのパラメータのごく一部しか推定できないことになり、モデルの構造が明らかにできない。これは、これまでは単年度の産業連関表（1サンプル）しか利用できないため、気づかなかっただけに過ぎない。

このような問題を解決する方法の一つは、次元を圧縮することである。そのための一番単純な方法は、部門の統合である。つまり、似たような産業部門は集計して一まとめにする。このように集計することで部門数を圧縮した産業連関表はこれまでも普通に公表されてきた。ただし、このようにして圧縮された産業連関表は、元の産業連関表が表していた経済構造を正確には保存しないことが以前から指摘されている。このように、産業連関表の次元圧縮の問題は、まだまだ研究の余地があると考えられる。

2 ネットワーク分析の観点からみると

産業連関表の中間取引の部分のみに注目する。そうするとこのマトリックスは正方行列になるが、産業間の取引ネットワークの関係を表していることになる。したがって、この行列を隣接行列とみなせば、そのままネットワーク分析の手法が適用できる。隣接行列とは、グラフを表現するために用いる行列である。

ある頂点 v と w の間の枝の本数を行列の (v, w) 成分に割り当てる。単純グラフであれば、枝があるとき (v, w) を 1 に、枝がないとき (v, w) を 0 にする。有向グラフの場合、 v から w に向かう枝があるときのみ (v, w) を 1 に、そうでないとき (v, w) を 0 にする。また、枝に重みがついているグラフの場合は、 (v, w) に重みを代入する。

同じネットワークの異時点間の比較を考える。ネットワークの構造に変化が起きていないかどうかを明らかにするためには、そのネットワークのダイアド、トライアド、クリークなどを順に比較して、それにどれだけの変化があるかを明らかにする必要がある。このとき、比較するネットワークが大きい場合（次元が高い場合）、現実的な時間で問題が解けない可能性がある。そこで、データの次元を下げて比較を行う必要がある。

産業連関表は、そのまま用いれば、情報量が多すぎると考え、産業連関表を 0-1 のマトリックスに変えて分析を行う方法も過去から行われてきた。このような分析は、旧来の産業連関分析の枠組みでは、質的産業連関表の分析と呼ばれ、グラフ理論を用いた研究が有名である。

0-1 マトリックスに変換する方法は、基本的にはある特定の域値を超えたセルは 1 とし、それ以下のセルは 0 とすることで行われる。この域値の設定の仕方には、幾つかの方法が考案されているが、方法によって得られる 0-1 マトリックスも異なってくるので、まだまだ検討の余地があると思われる。このように、産業連関表の中間取引マトリックスを 0-1 マトリックスに変換し、これを隣接行列とみなし、重みづけのないネットワークのダイナミズムとして考えるのが、出発点としては一番妥当かもしれない。

また、0-1 化によって次元の縮小が行えるのは、先の例の多変量回帰モデルにおいて 0 のセルに対応する変数が落ちることより明らかであろう。0 のセルが増えるほど、データの持つ次元は低くなる。

本研究では、質的産業連関表を情報の粗視化のための手法として用いることを考える。ネットワーク分析を行うにあたって、第一次接近として、各産業部門をノード、取引関係をリンクと考えた「つながり方」のみに着目するため、産業連関表を質的産業連関表に変換して用いる。

質的産業連関分析では、産業間の取引を重要なものとそうでないものの2値変数に変換するため、なんらかの方法で重要な取引を抽出する必要がある。本稿では、Aroche-Reyes(1996)の定式化にもとづき、次式を用いて重要な産業を特定化する。

$$r_{ij} = 1 / [a_{ij}(b_{ji} + (b_{ii} / x_i) x_j)]$$

ここで、 a は投入係数、 b はレオンチェフ逆行列の係数、 x は生産高である。この式では、投入係数 a_{ij} の値が大きくなるにつれて r_{ij} の値は小さくなる。すなわち、第 i 産業と第 j 産業間の取引が大きく、その取引が重要であるほど r_{ij} の値は小さくなる。したがって、特定の閾値 r_{ij}^* を定めて、その値よりも小さな r_{ij} の値を返す投入係数 a_{ij} に対応する取引を重要な取引として抽出することができる。

3. 本研究における問題設定

異なる時点での産業連関表を比較して、構造変化の検出を行うことにする。標本数が少なすぎるため、前章と同様の理由により、完全な比較は不可能なため、何らかの要約尺度の比較で行いたい。ここでは、中心性の尺度を用いることにする。

ネットワーク分析における中心性の尺度は色々存在するが、いずれの中心性尺度を用いるにしても、次のような問題が生じる。すなわち、異なる時点で二つの質的産業連関表を比較する場合、閾値 r_{ij} の値の変化に応じて中心性の尺度で測った産業のランキングにも変化がある場合、このような比較は無意味になる。つまり、二つの質的産業連関表の相違は、閾値 r_{ij} の値によって異なってくるからである。

本研究では、閾値 r_{ij} の変化に対して頑強なネットワーク中心尺度はどれかを明らかにする。具体的には、平成 20 年度実質情報通信産業連関表を用いて、Aroche-Reyes と同じ方法で、いくつかの粗視化の度合いの異なる質的情報通信産業連関表を作成する。

今例えば、粗視化の度合いの異なる質的産業連関表を表 A、表 B、表 C とする。産業連関表はこの順に稠密なるものとする。このとき、それぞれの表の中心性尺度を求めこれをそれぞれ A 尺度、B 尺度、C 尺度とする。二つの尺度ベクトルの順位相関係数を、相関 (B 尺度、C 尺度) のように表すならば、相関 (A 尺度、B 尺度) の値が十分大きく、かつ、相関 (A 尺度、B 尺度) と相関 (A 尺度、C 尺度) の値にあまり差が無いならば、この中心性尺度は頑健ということになる。なぜならば、このことが意味するのは、閾値 r_{ij} の変化に対して中心性尺度の順位はあまり変わっていないことを意味しているからである。このような方法で、複数のネットワーク中心性尺度から頑健なものを選び出す。

相関（A 尺度、B 尺度）の値が十分大きい必要があるのは、この相関が低い場合、相関（A 尺度、B 尺度）と相関（A 尺度、C 尺度）の値にあまり差が無くても、B 尺度と C 尺度の並びは相当異なっている場合があるからである。

本研究で用いるネットワーク中心性尺度とは、ネットワーク分析のなかでも、最もよく用いられる尺度の一つである。点中心性 (node centrality) は、ネットワークを構成する各頂点が、ネットワークにおいてどのくらい「中心的」であるかを示す尺度であり、一般に中心性といえばこれを指すことが多い。中心性は、ネットワークにおける各頂点の重要性を評価したり、比較するための尺度である。ここで注意が必要なのは、中心性はあくまでもネットワークの構造で決まるのであって、頂点が表している「何か」の属性で決まるのではないということである。例えば、社会ネットワークにおいて個人の性格や肩書きは中心性とはまったく関係がない。

本研究で用いる中心性尺度は、近接中心性、次数中心性、PageRank、媒介中心性の 4 種類である。

4. 中心性の指標の頑健性について

さて、今回は、表 A の閾値として $r < 100$ 、表 B の閾値として $r < 1000$ 、表 C の閾値として $r < 10000$ とする。順位相関を計算した結果は、表 1 にまとめられている。順位相関は、スピアマンの順位相関とケンドールの順位相関を記載している。

この表より、 $r < 1000$ と $r < 10000$ においては、どの中心性尺度も順位相関は高く、閾値として 1000 以上を設定した場合、どの中心性尺度も頑健性を有している。次に、 $r < 100$ と $r < 10000$ の順位相関と $r < 1000$ と $r < 10000$ の順位相関にあまり変化が見られない中心性尺度は、次数中心性、と PageRank であり、これに近接中心性（別の頂点からある頂点へ）が続く。近接中心性（ある頂点から別の頂点へ）と媒介中心性はかなり相違がある。

近接中心性 (IN) スピアマン		近接中心性 (OUT) スピアマン	
<100/<10000	<1000/<10000	<100/<10000	<1000/<10000
0.7763	0.9420	0.4024	0.9129
ケンドール		ケンドール	
0.6264	0.8092	0.3420	0.7803

次数中心性	スピアマン	Pagerank	スピアマン
<100/<10000	<1000/<10000	<100/<10000	<1000/<10000
0.8030	0.9475	0.8068	0.9087

ケンドール	ケンドール		
0.6046	0.8221	0.6319	0.7575

媒介中心性	スピアマン
<100/<10000	<1000/<10000
0.5930	0.9030
ケンドール	
0.4540	0.7506

表 1 順位相関

参考文献

- 1) Acoche-Reyes, F., "Important Coefficients and Structural change: A Multi-layer Approach," Economic Systems Research, 8(3), 235-246, 1996.