

複雑ネットワーク上で観測される非線形輸送現象と そのパラメータ推定

三浦航¹, 田村光太郎¹, 北島聡², 後藤隼人², 高安秀樹^{3,4}, 高安美佐子¹

¹ 東京工業大学 総合理工学研究科

² 株式会社帝国データバンク 産業調査部

³ ソニーコンピュータサイエンス研究所

⁴ 明治大学 先端数理科学インスティテュート

概要

本研究では、日本企業約4万社間の取引金額を用いて、企業の売上高から企業間の取引金額を推定するモデルを構築した。このモデルを日本企業約60万社の大規模ネットワークに適用することで、ネットワークのリンク構造のみから、企業の規模を代表する売上高の値を再現することができる。このモデルはあるパラメータでページランクを再現し、散逸、注入を含む一般的な非線形輸送現象のモデルとなっている。我々は、このモデルのパラメータを推定することで、実際の売上高との相関も高い値が得られることを示す。

Non-linear transport phenomena observed on a complex network with the parameter estimation

Wataru Miura¹, Koutarou Tamura¹, Satoshi Kitajima², Hayato Goto², Hideki Takayasu^{3,4},
and Misako Takayasu¹

¹ Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

² TEIKOKU DATABANK, LTD. Industrial Survey Department

³ Sony Computer Science Laboratories

⁴ Meiji Institute of Advanced Study of Mathematical Sciences, Meiji University

Abstract

We constructed a model that estimates the amount of transaction between 40,000 business firms in Japan from the total sales value of each firm. We apply this model to a business firm network comprised of about 600,000 firms in Japan to reproduce the sales values that represent the scale of the firms from only the link structures in the network. This model is a general model of non-linear transport phenomena including the effect of dissipation and injection, which reproduces PageRank in some parameter set. Here we show the parameter estimation method of the model and find a high correlation value between simulation results and real sales values.

1 はじめに

近年、素子同士の複雑な相互作用を記述する数的手法として、複雑ネットワークが注目を集めている。複雑ネットワークはインターネット [1] や WWW [2], 食物網 [3], タンパク質の相互作用 [4] など、自然科学の様々な分野で存在しており、その多くはリンク数の分布がべき分布に従う、スケールフリーネットワークであることが知られている。このスケールフリー性を再現するモデルとしては、BA モデル [5] などが知られており、本研究で扱う企業間取引ネットワークでは、三浦らによるモデル [6] が知られている。

複雑ネットワークの生成メカニズムの解明も重要な問題だが、複雑ネットワーク上で起こる現象についても同様に注目が集まっている。その中の一つに輸送現象 [7] が挙げられる。複雑ネットワーク上の輸送現象としては、ネットワークの中心性の問題と関連して、ページランク [8] がよく知られている。ページランクはあるノードが持つ値を、リンク数に応じて周りのノードに等分配するという、線形輸送のモデルである。しかし、ページランクは値がノードの入次数に比例する [9] ことが知られており、売上高のように企業の規模を代表するような値とは一致しない。

本研究では帝国データバンク社が保有する財務データと取引データを利用して、企業間取引ネットワークについての解析を行なった。財務データからは売上高や従業員数など、企業の規模を代表する値がわかり、取引データからはある企業がどの企業と取引を行なっているかや、その取引金額がわかる。我々はこの取引金額のデータを解析し、ページランクでは再現できない、お金の非線形な輸送現象を観測した。この観測結果からお金の非線形輸送モデルを構築し、モデルのシミュレーションから、実際の企業の売上高を高い精度で統計的に再現した。

次の章では企業データと、データから構成される企業間取引ネットワークについて説明し、3章では取引金額データから観測されるお金の非線形な輸送現象および、そのモデルについて説明する。4章ではモデルのパラメータ推定について、5章ではモデルのシミュレーション結果について説明し、最終章でまとめと今後の展望について述べる。

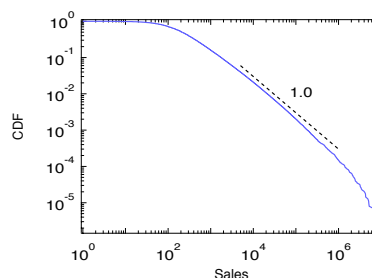


図 1: 売上高の累積分布 (Cumulative Distribution Function)

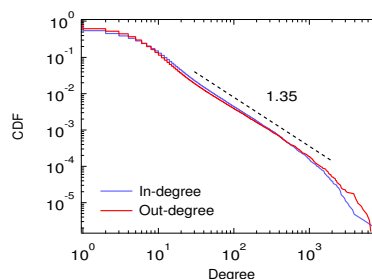


図 2: 次数の累積分布 (Cumulative Distribution Function)

2 企業間取引ネットワーク

本研究で扱うのは、帝国データバンク社が保有する 2011 年 7 月の、日本企業 2,331,272 社の財務データ (COSMOS2) と、4,016,066 取引関係の取引データ (CCR) [10] である。このうち 41,652 取引が取引関係に加え、取引金額も含む。我々はこのデータから企業がノード、発注社から受注社へのお金の流れがリンクとなる有向ネットワークを構築した。この企業間取引ネットワークは、日本の主要な企業だけでなく中小企業までほぼ全てを網羅し、ノード数 627,262, リンク数 3,844,684 の大規模ネットワークとなる。このネットワークの売上高分布 (図 1) と、次数分布 (図 2) について見ると、前者は累積で指数が 1.0 の、後者は入次数、出次数ともに累積で指数が 1.4 のべき分布となり、企業間取引ネットワークはスケールフリー性を示すことがわかる。それと同時に売上高と次数分布の指数の値が異なることから、ページランクではネットワークの構造を表す特徴量として適切でないということがわかる。

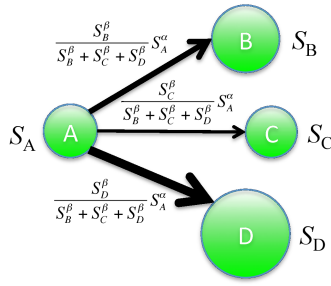


図 3: 非線形輸送モデルの概念図

3 お金の非線形輸送モデル

次に実際の売上高から、企業間の取引金額を推定する方法について考える。ネットワーク上での非線形な隣接相互作用から、ノード A, B がそれぞれスカラー値 S_A, S_B を持つとき、ノード A からノード B への輸送量 f_{AB} を、

$$f_{AB} = \frac{S_A^\alpha S_B^\beta}{\sum_{j=1}^N A_{Aj} S_j^\beta} \quad (1)$$

として表す。ここで A_{ij} はネットワークのリンク構造を表す隣接行列であり、ノード i からノード j へリンクが存在するとき値が 1, 存在しなければ値が 0 となる。この輸送モデルの概念図を図 3 に示す。これは取引ネットワークでは、売上高 S_A, S_B からその企業間の取引金額 f_{AB} を推定することに対応している。 f_{AB} がノード B に依存せず、ノード A の次数で等分配する場合 ($\alpha = 1, \beta = 0$)、この非線形輸送モデルはページランクと一致する。

ネットワーク内での企業同士の取引だけでなく、ネットワーク外の個人との取引や、人件費などのコストのように散逸する量なども考慮すると、より一般的にこの非線形輸送モデルは次のように拡張することができる。

$$\frac{dS_M}{dt} = \sum_{i=1}^N f_{iM} - (1 + \nu) S_M^\alpha + F_M \quad (2)$$

左辺はノード M の売上高の時間変化を表し、それが周囲のノードからの流入量 $\sum_i f_{iM}$ と、ノード M からの流出量 S_M^α , またネットワーク外とのやり取りとして、自己への注入項 F_M と散逸項 νS_M^α で記述されている。

4 モデルのパラメータ推定

いま取引金額も記載されている、40,251 取引データを用いて α, β を実際に推定する。現実の取引金額を f^* とし、対数値に対して全ての取引金額の誤差 σ を評価する。すなわち、

$$\sigma = \sqrt{\frac{1}{M} \sum_{l=1}^M (\ln f_l^* - \ln f_l)^2} \quad (M = 40251) \quad (3)$$

とし、 α, β を変化させて誤差 σ を計算する。この結果を図 4 に示す。図 4 より、 β よりも α を変化させることで、誤差 σ が急激に変化することがわかる。誤差 σ が最小となる $(\alpha, \beta) = (0.89, 0.26)$ において、ノードへのお金の流入量と実際の取引金額の関係を図示すると、図 5 のようになり、ノードへの流入量と実際の取引金額がほぼ同じ値となっている。

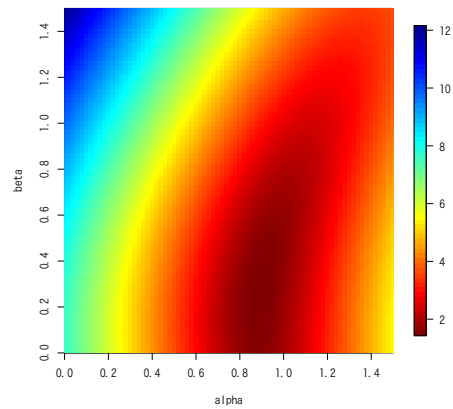


図 4: 各 α, β に対する誤差 σ の値

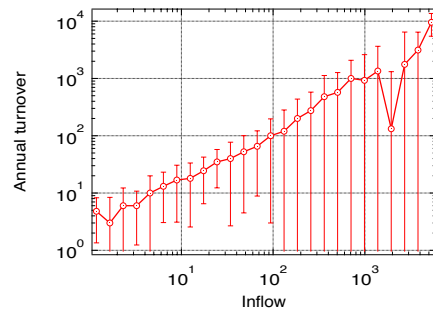


図 5: $(\alpha, \beta) = (0.89, 0.26)$ でのお金の流入量と取引金額の関係

5 モデルのシミュレーション結果

前章で推定したパラメータの値は、一部の取引金額のみに最適化した結果なので、実際には誤差を含む。また、ネットワーク外の取引を表す散逸項、注入項のパラメータ ν, F_M は分布の全体の形状には影響を与えない。そこで簡単のために、注入項のパラメータ $F_M = F$ で全てのノードに対して一定の値とする。いまパラメータを調整し、 $(\alpha, \beta, \nu, F) = (0.89, 0.31, 0.10, 95)$ として、非線形輸送モデルのシミュレーションを式 1, 2 にもとづいて行う。この結果を図 6 に示す。モデルのシミュレーション結果は統計的に高い精度で、実際の売上高分布を再現できていることが分かる。また分布だけでなく、売上高とシミュレーション結果を比較し、対数値のピアソン相関係数を計算した。この結果を表 1 に示す。製造業や卸・小売業など企業間の取引が主要な産業では相関係数が高い一方で、サービス業など個人との商取引が主要な産業では相関係数が低い値となっていることがわかる。

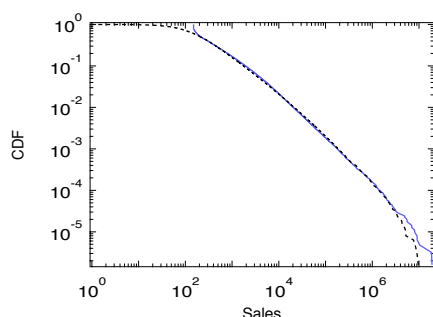


図 6: 非線形輸送モデルのシミュレーション結果 (実線) と実際の売上高分布 (点線)

6 まとめと今後の展望

本研究において我々は、企業間の取引ネットワークのリンク構造のみから、企業の売上高を再現するモデルを構築した。このモデルはノード間の非線形な相互作用から輸送量を決定する、ページランクを含む一般的なモデルである。実際の取引金額を用いてモデルのパラメータを推定し、シミュレーションを行うことで、企業の売上高を高い精度で統計的に再現できることを示した。特に企業同士の取引が盛んな産業において、モデルは高い相関係数で現実のお金の輸送現象を再現できることを示した。

今後の展望としては、ネットワーク外との取引す

業種 (ノード数)	ピアソン相関係数
全業種 (627,262)	0.415
農業 (3,097)	0.388
林業, 狩猟業 (324)	0.228
漁業 (752)	0.357
鉱業 (1,167)	0.607
建設業 (150,262)	0.320
製造業 (117,895)	0.546
卸・小売業, 飲食店 (197,876)	0.471
不動産業 (19,342)	0.193
運輸・通信業 (29,870)	0.409
電気ガス水道 (420)	0.655
サービス業 (106,257)	0.126

表 1: 業種別のピアソン相関係数

なわち、散逸・注入項の詳細化が挙げられる。企業間の取引データを解析するだけでなく、企業の費用や個人との商取引のデータを解析し、モデルの精度を高めることが重要だと考えられる。

参考文献

- [1] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [2] B. A. Huberman and L. A. Adamic, *Nature* **401**, 131 (1999).
- [3] J. M. Montoya, S. L. Pimm, and R. V. Solé, *Nature* **442**, 259 (2006).
- [4] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, *Nature* **411**, 41 (2001).
- [5] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [6] W. Miura, H. Takayasu, and M. Takayasu, *Phys. Rev. Lett.* **108**, 168701 (2012).
- [7] H. Watanabe, H. Takayasu, and M. Takayasu, *New J. Phys.* **14**, 043034 (2012).
- [8] S. Brin and L. Page, *Comput. Networks and ISDN Syst.* **30**, 107 (1998).
- [9] T. Ohnishi, H. Takayasu, and M. Takayasu, *Prog. Theor. Phys. Suppl.* **179**, 157 (2009).
- [10] <http://www.tdb.co.jp/lineup/ccr/>