

Title	知的生産を行う集団のネットワーク構造分析
Author(s)	井上, 寛康
Citation	年次学術大会講演要旨集, 24: 671-676
Issue Date	2009-10-24
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/8719
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

知的生産を行う集団のネットワーク構造分析

井上寛康（大阪産業大学）

1. はじめに

技術的イノベーションが経済発展を支えていることは広く認知されており、それは科学的知見に依拠しているといわれている [1]。しかしながら、そのプロセス、すなわち科学的知見がいかに技術的イノベーションに発展しているかの過程は大きなギャップがあり、いまもなお盛んに研究が行われている。

現代において、技術的イノベーションのプロセス自体に変化が見られていることは、広く認識されているところである。その1つに従来クローズドであったプロセスがオープンに行われるようになった [2] ということがある。ここでクローズドであるとは、研究開発、製品販売、売上げの獲得というサイクルを1つの企業（あるいは企業グループ）内で行うということである。これは日本の伝統的なイノベーションのプロセスであり、我が国が得意とする形であった。これに対してオープンであるとは、上記のプロセスにおける研究開発の部分に、外部の研究開発の成果を自在に取り込むということである。すべての企業にとって、オープンイノベーションを目指す必要があるかどうかは議論の余地があるとして、いったいそのようなプロセスを実現するにはどのような組織を構築すべきかどうかをここでは議論する。

本研究では、技術的イノベーションにおいて重要なプレーヤーである、科学者、技術者の間でどのような協業が行われているのかに焦点をあて、その相互の影響の解明に接近する。これにより、上述のような現代的なイノベーションプロセスの理解に寄与できると考える。本研究では、データとして論文と特許を用いる。そして、科学的知見の発見者たる論文著者のネットワーク、およびその具現者である発明者のネットワークを構築する。その上で、それらネットワークで共通するノードを基準に多重ネットワークを構築する。この多重ネットワークにおいてどのような部分構造が有意に現れるかどうかを尤度に基づいて判断する。

2. 多重ネットワークデータ

ネットワークはノードと呼ばれる点とそれをつなぐリンクからなる。本論文が対象とした多重ネットワークデータは2つのネットワークからなり、それぞれ論文著者ネットワークと発明者ネットワークである。重要な点は、これらネットワークのノードは、ある個人に対応しており、これら2つのネットワーク間で共通していることである。

イノベーションのプロセスは、論文により科学的知見が発見され、特許によりそれが具現化されるという順になるが、ここではデータの作成の都合から逆の順、すなわち発明者のネットワーク、論文著者のネットワークの順に説明する。

発明者ネットワークの元となるデータは、日本の公開特許公報において1993年1月から2002年12月の10年間に記載された4,998,464件の特許データであり、TamadaDatabase[3]を利用する。発明者ネットワークにおいて、ノードは発明者である（ただし論文著者ネットワークと共通である）。また、リンクは発明者間で共同で1つでも特許が出願されていれば形成される。

図1は、発明者ネットワーク構築の様子である。黒丸は発明者であり、ノードである。それらをつなぐリンクは、それらノードの間で1つでも特許が申請されていた場合に引かれる。したがって、リンク

については多重度を考慮せず，リンクが存在するかしないかの2値である．国内の発明者のネットワークはすでに構築済みであるが，論文著者ネットワークのデータについては，すべてを保有していないため，このネットワークの一部を用いる．これについては後述する．

次に，論文著者ネットワークの元となるデータは，国立情報学研究所が提供する CiNii において，2009/02/15 時点での論文データであり，12,039,089 件の論文が登録されている．論文著者ネットワークにおいて，ノードは論文著者である（ただし発明者ネットワークと共通である．）また，リンクは論文著者の間で共同で1つでも論文が執筆されていれば形成される．

ここまで論文著者ネットワークおよび発明者ネットワークをそれぞれ述べたが，次に，どのようにして発明者ネットワークと論文著者ネットワークのノードを一致させるのかについて述べる．手順は次のようになる（1）発明者ネットワークからある部分のノードを取り出す（2）それらの間のすべての2者の組み合わせを生成する（3）2者の氏名が同一の論文の著者となっていれば，それらのノードの間に論文著者ネットワークのリンクがあるとする．ここで1つの問題は，同じ氏名の人物がいるという可能性である．確かに同じ氏名の人物は大量に存在するが，同じ氏名を持つ2組の2者が論文を作成する可能性は，きわめて低いことから無視することができる．

本研究では論文著者ネットワークおよび発明者ネットワークの全体を対象とせず，科学的知見を元に発明に至っている，数十人程度の部分ネットワークを2つ抽出した．

1つは同志社大学教授山口栄一氏を中心とした73名からなるネットワークである．同氏は固体物理学が専門であり（株）パウデック，Algan（株）などを起業した論文著者・発明者である．同氏から3ステップで到達できるノードは発明者ネットワークにおいて72であり，これら73ノードの間の論文著者ネットワークを取得した．これは半導体分野を代表する1つのネットワークといえる．以後このデータを Yamaguchi-3 とする．

もう1つは大阪大学客員教授森下竜一氏を中心として127名からなるネットワークである．同氏は臨床遺伝子治療が専門であり，アンジェス MG（株）などを起業した論文著者・発明者である．同氏から2ステップで到達できるノードは発明者ネットワークにおいて126であり，これら127ノードの間の論文著者ネットワークを取得した．これは遺伝子治療薬分野を代表する1つのネットワークといえる．以後このデータを Morishita-2 とする．

3. 多重ネットワーク依存モデル

本研究で知りたいことは，論文著者と発明者という異なるネットワークの間にどのような関係性が形成されているかである．そのための分析手法として，Random Graph Models for Multiple Relation[4]（多重ネットワーク依存モデル）を用いる．これは p^* に基づくネットワーク分析 [5] を多重ネットワーク分析に拡張したものである．

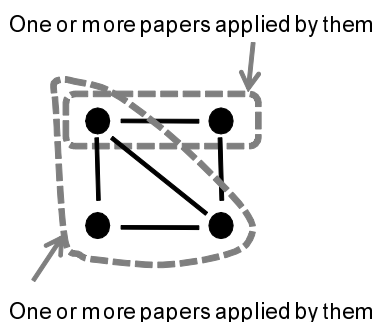
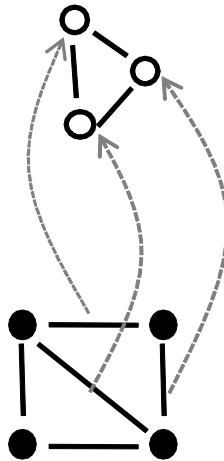


図 1: 発明者ネットワーク構築の様子

Dependence network (model)



Actual (authors' or inventors') network

図 2: 依存ネットワーク

多重ネットワーク依存モデルでは、ネットワーク x は、他のリンクに依存して決まるような確率分布に従って出現したとし、次のように表す。

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

ここで $\kappa = \sum_x \exp(\sum_{A \subseteq N_D} \lambda_A z_A(x))$ は正規化定数、 A はノードの部分集合、 N_D は依存ネットワーク (後述) のノード、 $z_A(x) = \prod_{x_{ijm} \in A} x_{ijm}$ であり、 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。ネットワーク m というのは、ネットワークの種類を表しており、この場合は論文著者ネットワークと発明者ネットワークを表している。

依存ネットワークとはネットワークのリンクをノードとするグラフである。図 2 は依存ネットワークの例である。下の黒丸のネットワークが実在のネットワークであり、上の白丸のネットワークが依存ネットワークである。依存ネットワークは実在のネットワークのリンクをノードとしている。依存ネットワークのリンクは、それらの間で相関があることを意味する。これはモデルであり、この依存ネットワークによって実際の多重ネットワークをどれくらい説明できるか求めることが、この分析の目的である。これは式 (1) における λ を求めることと同じである。

このままでは λ の数が多すぎ、計算ができない。そこで、同位性を用いてその数を減らす。すなわち、依存ネットワークにおける (たとえばトライアングルなどの) 同じ形はすべて同じ効果とみなすことになる。本研究で扱う部分構造は Choice, Multiplexity, Role interlocking, Transitivity とする。これを表したものが図 3 である。これらを選んだのは、先行研究 [4] における基本的な有向グラフのモデルに合わせたためであり、本研究は無向グラフを扱うため、修正を施してある。これ以上に複雑なモデルはいくらでも考えられるが、本研究ではまずこの簡単な構造について分析を行う。

一般的に式 (1) を同位性を考慮した形に直すと、

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

となる。この式を用いて実在の多重ネットワークが尤度最大になるように λ を求めればよい。この尤度を求める方法として、擬似尤度法 (Pseudolikelihood function) [9] があり、本研究ではこれを用いる。

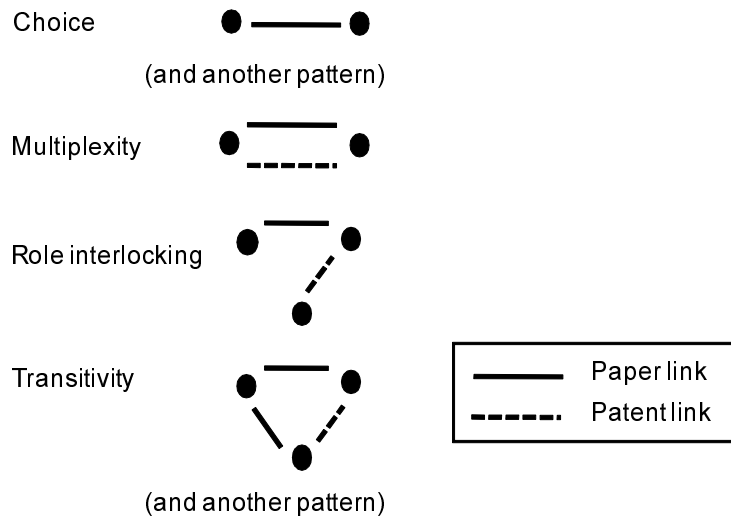


図 3: ネットワーク依存パタンの種類

表 1: 多重ネットワークモデル回帰結果

Yamaguchi-3			
Model	G_{PL}^2	# of variables	
1. Patent choice, paper choice	3006.7	2	
2. Patent choice = paper choice	3353.7	1	
3. 1 + Multiplexity	2932.5	3	
4. 3 + Role interlocking	2826.0	7	
5. 4 + Transitivity	2328.6	11	

Morishita-2			
Model	G_{PL}^2	# of variables	
1. Patent choice, paper choice	9984.7	2	
2. Patent choice = paper choice	10048.9	1	
3. 1 + Multiplexity	8629.1	3	
4. 3 + Role interlocking	7268.7	7	
5. 4 + Transitivity	5856.1	11	

4. 分析結果および議論

擬似尤度法はロジスティック回帰分析により行えるが、この分析には SPSS 17.0 を用いた。表 1 は分析結果を表している。Yamaguchi-3 および Morishita-2 のデータに対して、多重ネットワーク依存モデルの各コンフィギュレーションに対する（擬似）-2 対数尤度 (G_{PL}^2) およびそのときの説明変数の数を示している。一般的に回帰モデルは説明変数が増えるほど G_{PL}^2 が小さくなるので、その有意な差を決める必要がある。そのような値として用いられるのは $-2n(n-1)r \log(1-\delta)$ である。ここで n はノードの数、 r はネットワークの数、 δ は定数であり、一般的に 0.001 や 0.005 が用いられる。この値はそれぞれのデータについて、105.4, 320.8 である。

表 1 を見ると、どちらのデータの場合も Patent choice, paper choiceの方が Patent choice=paper choice より優れている。すなわち特許と論文のリンクについて、発生する確率は別々と捉えるほうが、

モデルの精度がよい（-2 対数尤度は正の数であり、小さいほうがよいモデルである。）これはそれぞれのリンクの発生確率が異なることに単に起因する。重要であるのは、Model1 と 2 の-2 対数尤度の差の大きさである。Yamaguchi-3 については前述した目安の閾値を超えているが、Morishita-2 については、超えていない（有意な差とはいえない。）これは、1つの解釈としては、半導体分野と遺伝子治療薬分野を比較すると、論文と特許の生産を両方向うグループの存在は後者の方がより可能性が高いということになる。

このモデル1に対して、Multiplexity を足したのがモデル3であるが、ここでも結果は異なる。Yamaguchi-3 の方はモデル1 と 3 が有意な差でないのに対して、Morishita-2 の方は有意な差になっている。これは、論文と特許の生産を両方向うグループの存在が高いという前述の内容を補強している。このモデル3 が有意であるということは、1つのネットワークでのリンクの存在が、他方のネットワークでの同じノード間のリンクの存在を刺激することを示している。モデル1, 2, 3 のこの結果は、技術分野によって協業の傾向が異なることを明確に表している。

つづくモデル4の Role interlock およびモデル5の Transitivity はどちらのデータにおいても有意という結果である。モデル4を解釈すれば、ある人物がいて、論文か特許を出すパートナーがいる場合、それと異なる生産（論文なら特許、特許なら論文）を行う別のパートナーがいる確率が高いということになる。すなわち、周囲に2つの集団（集団か1人かはわからない）が異なる役割を果たしており、適宜参加する集団を替えているともいえる。

さらに、モデル5を解釈すれば、3人が集団を構築する際、ある人を中心に論文あるいは発明が行われるが、それをともに行った別の2人が、それと異なる生産を行う可能性が高いということになる。モデル5はモデル4との対比で考えるとわかりやすい。モデル4では論文と特許の両方を生産する（リーダー的な）人がいる形であるが、モデル5では論文か特許かのどちらかを専門的に生産する（エキスパート的な）人がいて全体として両方が生産される形である。

最後に本研究の結果の意義について述べる。複数のネットワークの関係性を扱ったこれまでの研究は、主に Multiplexity の分析のみであった。この Multiplexity は、ネットワークの重複度合いの観察と言い換えることができる。本研究の結果からわかるように、Multiplexity は分野によってその様子が異なる。そのため、ある集団の結果に対して、この重複度合いだけを観察するような単純な分析をした場合、科学的発見と発明を行う集団は協業する / しないという誤った結論を導くことになる。これに加えて、本研究での重要な示唆は、集団における2人の関係を単純に観察しただけでは、協業の実際を見ることはできないということである。本論文で見たように、3人が関係するような関係（コンフィギュレーション）が科学的発見と発明での協業において有意に現れていることが、本研究において初めて明確にされた。

また、近年ネットワーク分析は急速に発展しており、他分野への応用が積極的に展開されている。多重ネットワーク分析はネットワーク分析の中でもカッティングエッジにあり、その分析手法が研究開発プロセスの分析に適用可能であることが示された。

5. 結論

技術的イノベーションが経済発展を支えていることは広く認知されており、それは科学的知見に依拠しているといわれている。しかしながら、そのプロセス、すなわち科学的知見がいかに技術的イノベーションに発展しているかの過程はいまだによくわかっていない。本研究では、科学者、技術者の間でどのような協業が行われているのかに焦点をあて、その相互の影響の解明に接近した。具体的には、科学的知見の発見者たる論文執筆者とその具現者である発明者において、その協業ネットワークを抽出し、これらがどのように影響しあうかについて多重ネットワーク分析を行った。

分析では、半導体分野と遺伝子治療薬分野を比較した。論文と特許の生産を両方向うグループの存在は、後者の方がより可能性が高かった。同様に Multiplexity は遺伝子治療薬分野が高かった。このように技術分野によって協業の傾向が異なることを明確に示した。

Role interlock および Transitivity によって，論文と特許の両方を生産する（リーダー的な）人がいる形，および論文か特許かのどちらかを専門的に生産する（エキスパート的な）人がいて全体として両方が生産される形が有意に現れることがわかった．

これまでの研究は Multiplexity の分析のみであった．本研究の結果からわかるように，Multiplexity は分野によってその様子が異なるため，そのような分析は十分ではないとわかった．また，3人が関係するようなコンフィギュレーションが科学的発見と発明での協業において有意に現れていることが，本研究において初めて明確にされた．

謝辞

本研究は科研費（20730268）の助成を受けたものである。

参考文献

- [1] E. Mansfield. Academic research and industrial innovation. *Research Policy*, 20(1), 1991.
- [2] H. W. Chesbrough. *Open innovation*. Harvard Business School, 2003.
- [3] S. Tamada, Y. Naitou, F. Kodama, K. Gemba, and J. Suzuki. Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, 68(2):289–302, 2006.
- [4] L. M. Koehly and P. Pattison. Random graph models for social networks: Multiple relations or multiple raters. In P.J. Carrington, John Scott, and Stanley Wasserman, editors, *Models and Methods in Social Network Analysis*, chapter 9. Cambridge University Press, New York, 2005.
- [5] 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.
- [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, John Scott, and Stanley 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.