

高速鉄道が知識移転に与える効果

—日本の特許引用を用いた実証分析—*1

田村 龍一*2

要 約

1997年10月の北陸（長野）新幹線の開通を用いた自然実験フレームワークを採用し、知識移転の発現を日本の特許引用データから得ることで、高速鉄道が知識移転に与える因果効果の発現を検証する。特許出願に関わった発明人の実際の地理分布に基づき、北陸新幹線開通の影響を受ける地域を処置群に、どの新幹線からも離れていて新幹線以外の特許引用に影響を与える要因をコントロールする地域を参照群に設定する。各群の特許引用パターンに差分の差分分析を実行した結果、高速鉄道は知識移転を促進し、移転の地理範囲を拡大することを実証した。

キーワード：知識移転，特許引用，スピルオーバー，高速鉄道，自然実験

JEL Classification：D62, C99, L91

I. はじめに

イノベーションにつながる新知識はどのように普及していくのか。古くは Arrow (1962) が洞察していたように、知識や情報の拡散はエンジニアや研究者といった人的資本の移動によってもたらされるのであり、本研究では情報ソース（人）へのアクセスが行われ別の人へ伝達される現象を知識移転（Knowledge transfer）と呼ぶことにする。知識移転は「送り手」「受け手」「移転のメカニズム」という3つの要素から構成されると考えることができる。本研究では、知識のスピルオーバーが発生するような移転の

メカニズムに考察を集中する。なぜなら経済学で理論及び実証の両側面から強調されているように知識資本ストックのスピルオーバー効果は経済成長に不可欠な要素となるからである。

本研究ではスピルオーバーによる知識移転を考える視点として、送り手と受け手の地理的位置関係に着目する。経済学では知識は公共財として扱われるが、後述のようにその性質が及ぶ範囲は地理的に限定されていることが多くの実証研究により明らかになっている。送り手と受け手の間で知識移転に関する何らかの相互作用

*1 本研究は JST/RISTEX プロジェクトの助成を受けたものである。本研究のトピックに関しては 2013 年に楡井誠氏、Mark Schankerman 氏との議論から派生したものである。しかしながら、本研究における誤りなどについては全て筆者の責任である。また、東京大学空間情報セミナー及び日本大学 ARISH セミナーの参加者からは貴重なコメントをいただいたことに感謝申し上げます。

*2 一橋大学イノベーション研究センター

の存在を仮定すると、このための費用を大きく低減する社会基盤の存在は、地理的により広範囲の知識移転を実現できる可能性がある。本研究で着目するそのような社会基盤は新幹線（高速鉄道）である。新幹線による目的地への旅行時間の大幅な短縮は人の流動を促進し得る。例えば本研究が事例として用いる北陸新幹線では、開通によって長野駅から東京駅までの所要時間は在来線を用いた場合の2時間30分から最短1時間19分まで短縮した。これによって首都圏といった多くのイノベーションが発生する地域から、北信越および上信地域という相対的にイノベーションの少ない地域への知識移転が容易になる可能性が示唆される。本研究では、日本の特許出願をイノベーションの指標、特許引用を知識移転の痕跡と捉え、新幹線は知識移転を促進する手段として効果的であるのかどうかを、これらの特許関連データを用いた実証分析によって明らかにする。

知識移転におけるスピルオーバー効果の発現に関する実証研究の系譜は、Gliriches (1979) が提案した「知識生産関数」アプローチによる研究開発投資のスピルオーバー効果の計測から始まっていると考えられる。このアプローチを採用した例として、Jaffe (1986, 1989) は民間企業や大学の研究開発投資額と産業技術的に近い関係にある近隣地域の他の民間企業の生産性の間に正の相関を見出し、1組織の研究開発投資は他者にスピルオーバーすることを示し、Acs, et. al (1994) は類似の分析を企業規模別に実行し、スピルオーバー効果は大企業よりも中小企業の方が多く享受できることを示した。これらの分析は国家あるいは地域単位で集計された経済変数の間の関係を解明することに主眼があり、知識移転の構成要素である「受け手」「送り手」が具体的に誰かを把握することはできない。この把握を「特許引用記録」という文書によって追跡する手法を提案したのがJaffe, et. al (1993) である。彼らの手法では、特許内容で表される新知識をもつ被引用側の出願人から、それを享受する引用側の出願人に至る知識

のスピルオーバーの経路が明確となる。

時期を同じくして、米国特許商標管理局 (USPTO) の特許データが広く利用できるようになる。特許データには特許（新知識）の創成に関与した出願人（組織）と発明人の名称と住所、特許内容や技術分野に関する大量かつ詳細な情報が利用できる。このため、Jaffe, et. al (1993) 以降、特許データを用いた知識移転、または特許を契機とした知識共有のフォームに関する実証研究が経済学、経営学、あるいは情報学といった様々な分野で盛んになった。

知識移転の具体的なメカニズムに関する実証研究は3つに大別することができる。ひとつにはイノベーションにつながる知識を備えた研究者が転職や移動を行う中で、所属先の同僚がスピルオーバーを享受するという知識移転のメカニズムである (Nakajima, et. al (2010), Breschi and Lissoni (2009), Agrawal, et. al (2006), また日本の特許データを用いた分析として、Saito and Yamauchi (2015) がある)、2つには過去の共著関係、人種や国籍といった文化・社会的結びつきといった社会ネットワークが知識の移転または共有を促進するという実証及び事例研究 (Saxenian (1999), Singh (2005), Agrawal, et. al (2008), 3つには、Jaffe, et. al (1993) と同じ手法を用いた知識のスピルオーバーの地理的な局所化の測定に関する実証研究である (Thompson and Fox-Kean 2005, Murata, et. al (2014))。

3つ目の研究ラインから得られる知見とは、知識移転をスピルオーバーという形で享受するためには、情報ソースへの近接性が重要であるという命題が肯定的に支持されることである。本研究ではこの点に関して次の問題提起を行う。近接性のみがスピルオーバーを享受するための要件なのか。本研究と同様のデータを用いて筆者が2013年に行った試算によると、我が国でのスピルオーバーの範囲は約60km程度である。しかしながら、たとえ情報ソースから60kmより離れた地域にいるエンジニアや研究者でも、もし情報へのアクセス費用を大きく減

じることが可能な手段やメカニズムがあれば、その地域にいながら遠方で生まれた知識からのスピルオーバーを享受することも可能であると考えられる。また、実証研究では特許という公開文書とその引用記録が知識移転の結果として分析に活用されている。ここでは受け手と送り手は明確であるものの、どのようなメカニズムで知識移転が実現されているのかについてはJaffe, et. al (1993) 以来、「ブラックボックス」である。問題提起の2点目は、イノベーションにつながる外部情報を取得する費用を大きく減じる社会資本の存在が、このブラックボックスに含まれているのではないかということである。本研究では知識移転を促進させる社会資本として高速鉄道、特に新幹線の存在をとりあげる。新幹線は遠隔地への（一般化された）旅行費用を減少させるという形で沿線における人の流動を促進し、地域イノベーターが知識移転の受け手である場合、新知識へのアクセス範囲が広がることを示唆する。これら2つの問題提起は、日本の地域経済にてイノベーションを起こそうとする際に、その地域に存在しない新知識へのアクセスを促進させる社会資本が実際に存在することを確認するという点で意義があると考えられる。

I-1. 分析手法

新幹線の開通が地域の外部知識取得活動に与える影響を実証分析によって考察するために、本研究では外部知識取得費用の低下に起因する効果は知識移転の頻度と地理的広がりにも現れると想定し、これを次の因果関係に関する3つの仮説として設定した。「知識移転の頻度が増加する」（仮説1）、「知識移転の範囲はより遠方へ拡大する」（仮説2）、「その結果「知識移転は相対的に地域の内から外へ移動する」（仮説3）。これらの因果効果を捉えるために、1998年の長野オリンピックの開催に合わせて1997年10月に開通した北陸（長野）新幹線を外生イベントとした「自然実験（Natural experiment）」による分析を行う。沿線地域で活動する企業や

発明人（これを処置群 [treatment group] と設定する）の外部知識取得活動にとっては、北陸新幹線の各駅へのアクセス（道路長）は外生的な割り当てである。沿線にはそれまでまったく遠方への知識アクセスを行ったことのない企業、外部知識取得範囲は周辺地域が大半である企業、親会社や共同研究を行う別企業が東京や大阪を初めとする遠方にありそれらの研究開発成果を頻繁に参照する企業など、企業固有の事情により知識移転の頻度や地理的広がりに関しては企業ごとに差異があると考えられる。彼らに対する北陸新幹線駅へのアクセスの割り当てはランダムに行われるため、地域で集計し平均するとこれらの差異は互いに相殺されると想定する。よって開通前後に知識移転の頻度や地理的範囲に違いがあるならば、それは新幹線の影響（平均処置効果：Average treatment effect）によるものであると考えることができる。

以上が自然実験によって新幹線開通の因果性を捕捉する基本的なアイデアであるが、同じ時期に知識移転費用の大きな低下をもたらす新幹線以外の社会資本の拡充、例えばその他の交通手段の発達やインターネットの普及に留意する必要がある。これらが全国規模で知識移転に影響を与えているのであれば、この効果は北陸新幹線を含むどの新幹線の駅からも遠く離れていて、よって新幹線以外の手段を用いる地域の知識移転の状況を把握すればよい。本研究では、これを参照群（Control group）として設定し、全国規模での新幹線以外の要因の通時的変化を把握する。すると、北陸新幹線開通が知識移転に与える因果効果とは処置群サンプルの開通前後の差から、参照群の開通前後の差をひいたもの、すなわち「差分の差（difference in difference）」として識別することができる。

本研究では我が国の特許庁に提出された特許出願をデータとして用いる。特許出願は全てのイノベーション創成に関わる知識移転をカバーしているわけではないが、我が国のイノベーション活動を把握するための代理指標とみなすことができる（Acs, et. al 1994）。本研究で分析に活

用する特許出願データの特色は、住所情報から得られた地理座標を使い、新幹線ネットワークからの近接性の判断が可能となっている点である。実際、分析サンプルの構築にあたっては新幹線駅への道路長を近接性の指標として導入し、それと閾値距離（R キロメートル）の大小関係に基づいて処置群と参照群の選定を行う。

I-2. 論文構成

本研究は以下の構成である。第II節では主要

なデータである特許書誌情報の整形と地理属性の追加といった作業を説明し、処置群を決定するための実験パラメータ R の設定の方法を実際のデータから導入する方法を示す。つづいて自然実験が機能するような分析サンプルの構築について詳述したあと第III節では上記の3つの命題の成立に関する因果分析を実行する。第IV節では結論と残された課題について言及する。

II. データ

特許出願に関するデータソースは IIP パテントデータベース（2015年版）である。収録されているレコードから、本州・四国・九州に所在地がある組織（企業、公的機関、大学など）と発明人を分析の対象とする。IIP パテントデータベースは特許庁に提出された特許に関する包括的なデータであるが書誌情報は「ありのまま」記載されているため、電子化に際してイレギュラーな入力や固有名称・住所に関する一貫性のない表記が多数存在している。本分析のためにはこれらを一貫性のある形にクリーンアップする作業が不可欠であり、その概要を補論に示した。これによって、1976年から2012年までの約990万の特許出願データを基礎とし、識別された約50万の出願人（個人を含む）と217万人の発明人に関するデータセットを準備した。

外部知識取得活動の痕跡の直接的な指標である特許引用については、Tamada (2006) によって作成された特許全文中に現れる特許引用記録をテキスト解析によって抽出したデータベース「td5」を用いる¹⁾。IIP パテントデータベースに収録されている特許引用データは審査官が審

査拒絶の際の根拠として提示する先行文献のコレクションであり、当該特許が参照した情報源とみなすことは難しい。一方、td5 データベースでは出願人が本文中に直接記載したものが収集されているので、知識のフローを示す指標としてより適切であると考えられる。この特許引用記録の中には発明人や出願人がこれまでに出願した記録を指すものがある。これらは自己引用（self citation）と呼ばれ、別の組織または発明人の間の知識移転をもたらす知識のスピルオーバーの観点からは含めないほうがよいと思われる。全特許引用のうち約30%がそのような自己引用であり、サンプルから除外してある。

出願人および発明人の住所は新幹線駅との地理的近接性の判断に用いるため、ジオコーディング作業によってできる限り細かい地理単位での緯度経度情報を得ている²⁾。新幹線を含む交通施設データは国土交通省が配布する「国土数値情報ダウンロードサービス」から得られるシェイプファイルを用い、道路網ネットワーク及び道路経路長の計算は日本の Open Street Map データを備えた Open Street Routing Machine (OSRM) を活用する。

1) td5 データベースは一橋大学イノベーション研究センター所蔵のバージョンを利用した。

第Ⅲ節では引用距離の面的広がりを分析するために、引用距離を計算している。引用距離は引用特許と被引用特許の発明人のすべてのペアの道路長を計算し、この中央値として定義している。

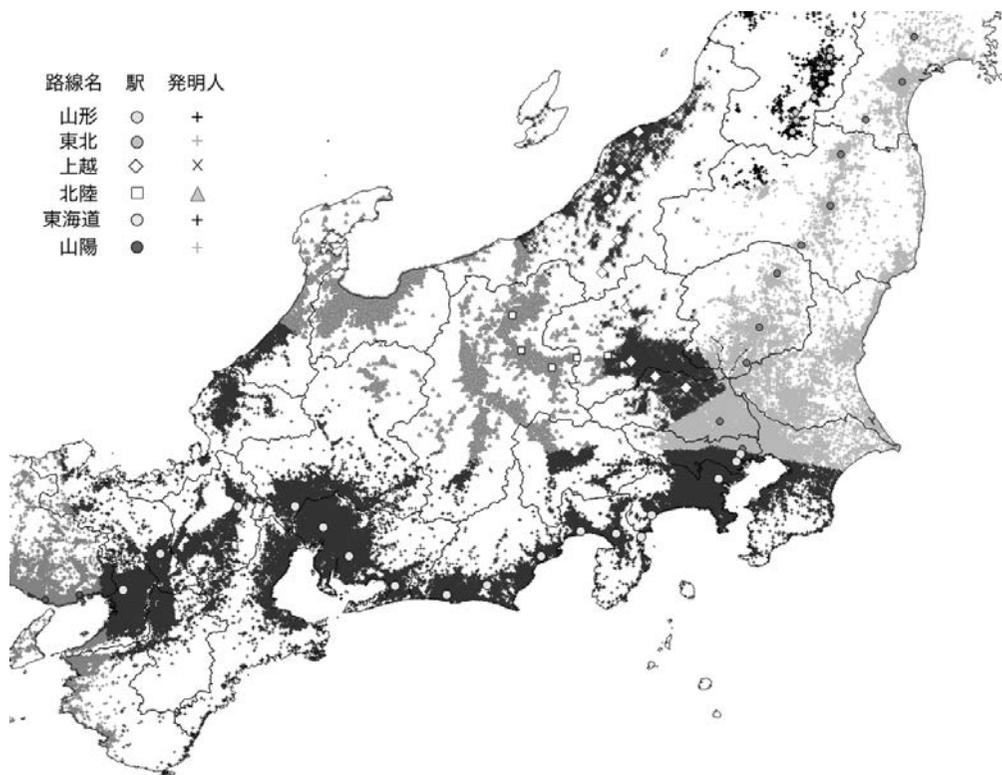
Ⅱ-1. 特許出願の位置と「沿線」「圏外」特許の定義

本研究で基礎となる情報とは、特許出願の内容がどの地域から生まれたもので、それが新幹

線駅との近接性である。これらについてジオコードされた特許書誌情報の住所をもとに決定する。書誌情報には出願人住所と発明人住所が含まれるが、出願人住所は本社や本所であることがほとんどである一方、発明人住所は出願組織の本所と異なる住所（支所や研究所など）が記載されていることが多い。この点から、本研究では出願につながる特許出願位置の特定には、発明人住所を用いることにする。

特許出願の位置と新幹線への近接性を把握す

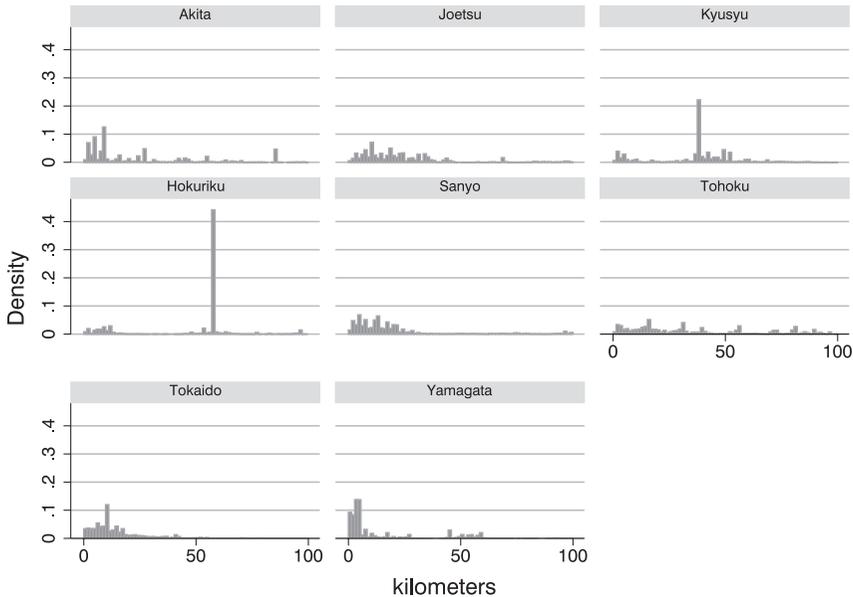
図1 「最寄駅マッチ」の例（1998年から2012年まで）



(注) 発明人の位置と最寄り新幹線駅の対応は図中の凡例を参照

2) ジオコーディングでは東京大学空間情報センターが開発する「dams」というソフトウェアによって住所の緯度経度変換を試みた。その際、国土交通省街区レベル位置参照情報 (<http://nlftp.mlit.go.jp/isj/>) を dams に追加することにより、一部の自治体についてはより高い精度（号レベル）での座標が与えられている。dams によるジオコーディングに失敗する住所の多くは町名変更や地区町村合併の影響で現存しない旧住所である。これらを最新の情報に変換し、座標を求めるために、Microsoft BingMap API サービスを用いた。

図2 最寄駅までの距離分布（新幹線路線別）



(注) 横軸に最寄り駅までの距離を km 単位で表示し、縦軸に各距離に位置する発明人の相対頻度（密度）を示している

るために、最初に発明者住所を新幹線ネットワークによって区分する作業を行う。具体的には所与の発明人の住所に対して最短時間で到達することのできる、出願日の時点で営業中の新幹線駅を最寄駅と定義し、この最寄駅と発明人住所の間のルート長を計算するのである³⁾。図1に北陸新幹線開通後の「最寄駅マッチ」の結果を示した。図の中央部にある長野県及び隣接県に広がるドットは、1998年から2012年までに申請され北陸新幹線各駅を最寄りとする発明人の位置である。

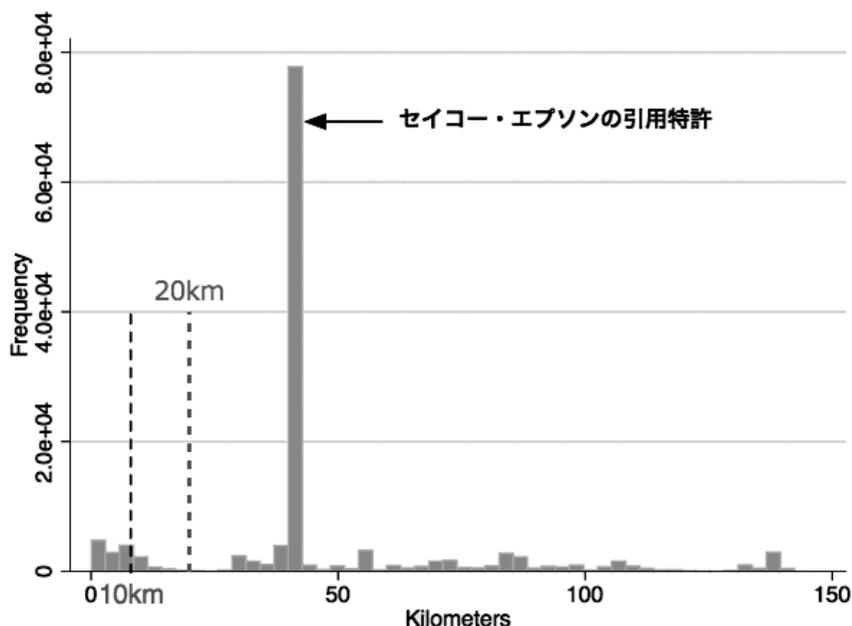
出願時点で営業中の新幹線駅に対する近接性を考えるために、図2に発明人住所の最寄駅までの距離の分布をヒストグラムで示した。この図を観察すると全ての路線について約20-30kmまでに距離分布の最初の峰が位置していることが分かり、外部知識の取得に新幹線を利用する

可能性が高い距離に対して指針を得ることができる。本研究では駅までの距離が最初の峰に含まれるとき「近い」それ以外るとき「遠い」という近接性の判断を行うことにする。これをサンプル構築に反映させるために、対応する「沿線特許」「圏外特許」という2つのタイプを定義する。ある新幹線の「沿線特許」とは、発明人全員が最寄り新幹線駅から（道路長で）R km以内に位置している特許出願と定義する。「圏外特許」は発明人全員が最寄り新幹線駅から少なくともR km離れた場所に位置している特許出願と定義する。

この閾値Rは本研究では北陸新幹線の最寄駅までの距離分布（図3）を観察してR=10kmと定義する。図4には全ての新幹線に対する沿線および圏外特許の同定内容を示している。全ての発明人が濃いドットのどれかに位置

3) 新幹線駅への車移動による所要時間とルートの計算には Open source routing machine (OSRM) を利用した。最短時間を与えるルートはダイクストラ法を用いて求めている。

図3 沿線・圏外判断の閾値の決定



(注) 北陸新幹線駅を「最寄り駅」とする特許出願について、その発明人から最寄り駅までの距離の分布を示している。40km 近辺に位置する度数のピークはセイコーエプソン及び関連会社からの特許出願である。

表1 沿線・圏外特許の判定結果 (R=10km, Total: 出願総数, On: 沿線, Off: 圏外)

Line Name	Start Year	End Year	Total	On	Off	(On+Off)/Total
Akita	1998	2012	4,779	2,154	2,141	89.87%
Joetsu	1983	2012	193,987	106,562	75,488	93.85%
Kyusyu	2005	2012	10,145	1,140	8,732	97.31%
Sanyo	1976	2012	790,509	399,128	326,605	91.81%
Nagano	1998	2012	220,787	30,660	182,768	96.67%
Tohoku	1983	2012	1,001,586	422,842	514,978	93.63%
Tokaido	1976	2012	7,431,086	5,541,755	1,421,684	93.71%
Yamagata	1998	2012	9,350	6,676	1,846	91.14%
All Lines	1976	2012	9,662,229	6,510,917	2,534,242	93.61%

していればその特許は沿線特許、全ての発明人が薄いドットに位置しているならば、その特許は圏外特許と同定するのである。

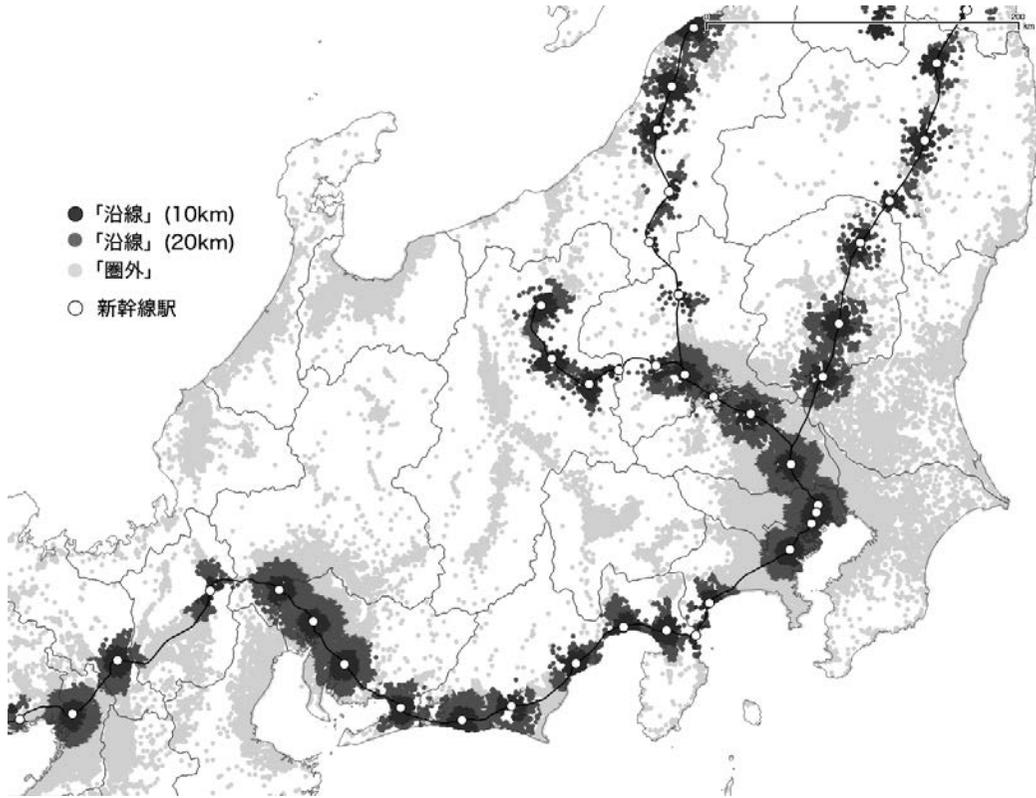
この方針によると、複数発明人が新幹線駅からの距離が閾値前後である住所に位置する場合「圏外」「沿線」のどの定義にも該当しない。しかしながらそのような特許出願は表1の最右

欄にしめすようにほとんどの場合10%未満である。

II-2. 沿線引用・圏外引用の定義

沿線特許、圏外特許の定義に基づいて、第Ⅲ節の分析にて行われるサンプルの構築を行う。沿線引用とは引用・被引用側がともに沿線特許

図4 「沿線特許」と「圏外特許」の同定のための発明人位置の分類



(注) 発明人位置を新幹線駅からの距離の観点から3段階のドットによって示す。薄いドットと濃いドットはそれぞれ新幹線駅から20km、10kmの場所に位置する発明人を表す。最も薄いドットはどの新幹線駅からも20km以上離れた「圏外」に位置する特許である。図中右上のスケールの長さは200kmである。

表2 分析サンプル

	開通前 (Pre-period)	開通後 (Post-period)
長野新幹線沿線引用	565	4658
圏外引用	327658	603510

である引用、圏外引用とは引用・被引用側がともに圏外特許となる引用である。沿線引用サンプルが北陸新幹線開通の因果効果推定における処置群、圏外引用サンプルが対応する参照群である。なお、第I節でも述べたように、処置群は開通前後を通して北陸新幹線駅地域に存在する発明人に限定している。よって、新幹線の開

通によって「圏外」から「沿線」に移動する出願人・発明人はこの引用サンプルには含まれていないことに留意されたい。このようにして構築された分析サンプルを用い、第III節にて引用の時間的推移、引用の距離と方向に関する命題を検証する。

Ⅲ. 分析

Ⅲ-1では最初の検証仮説である「開通後に特許引用件数は増加する」をグラフによる目視によって把握する。その引用先が圏外特許であるとすると新幹線を使わずとも取得できた知識である可能性が高い。Ⅲ-2では北陸新幹線発引用の行き先が同様に沿線特許であるか、それとも圏外特許であるかを調べるためにグラフによって開通前後で時間的推移を観察する。Ⅲ-3およびⅢ-4では、2, 3番目の検証仮説である「開通後は引用距離は伸び、新幹線沿線地域の内から外への引用が増加する」ことについて統計的な検証を行う。

Ⅲ-1. 沿線引用の件数の増加に関する分析結果

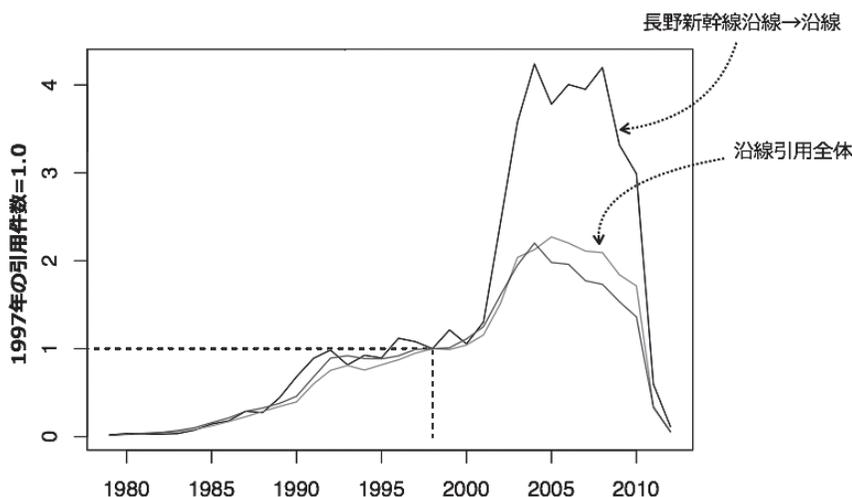
図5に沿線引用の時間的推移を示した。黒実線で開通前後の北陸新幹線駅周辺における特許出願からの引用件数の推移を、2つの灰実線で「圏外引用」及び「他の新幹線路線の沿線引用

全体」の推移を表している。これらの推移から分かるように、開通後約4年の時間的ラグの後北陸新幹線沿線引用は大きく増加していることが分かる。通常研究プロジェクトは複数年かかることを考えると、この結果は新幹線開通の効果であるとみなすことができると思われる。

Ⅲ-2. 沿線引用の行先

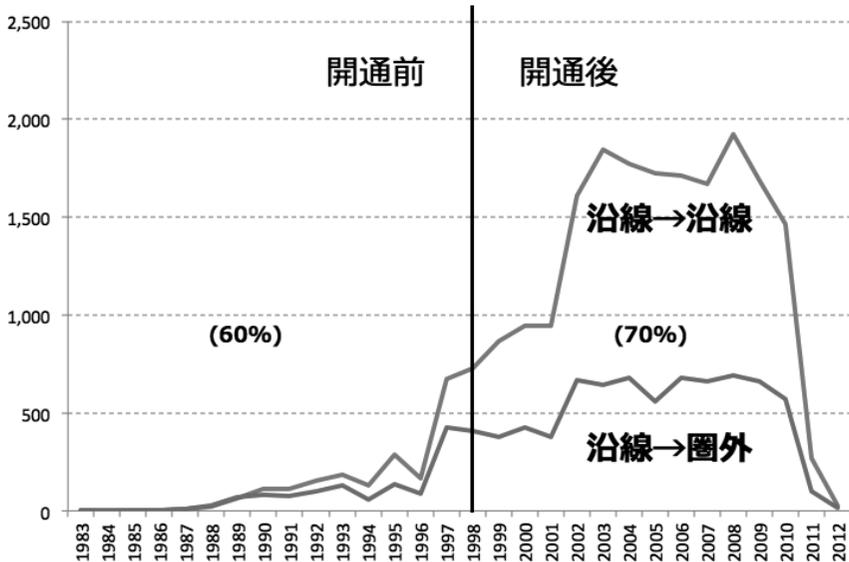
北陸新幹線の沿線引用はどこを引用しているのか、これを確認するために、引用サンプルについて行き先を沿線または圏外に分類し、その件数の推移をグラフ化したものが図6である。これから分かるように、開通前の引用先は沿線と圏外がほぼ同数であるのに対し、開通後は沿線から沿線への引用件数が大きく増加している。沿線間引用の比率でみると、開通前後のそれぞれの平均をとった結果約60%から70%に増加していることが分かる。

図5 北陸新幹線沿線引用、沿線引用全体、及び圏外引用の時間的推移



(注) 黒線で北陸新幹線沿線引用を、薄い灰色線で圏外引用を、最も薄い灰色線で全ての新幹線路線についての沿線引用を表す。

図6 沿線引用の行き先



(注) 横軸に出願年を、縦軸に引用件数を示している。

表3 「差分の差」検定の結果

	開通前			開通後			差分の差分
	圏外引用	沿線引用	差分	圏外引用	沿線引用	差分	
引用距離	261.61	191.11	-70.51	255.43	219.97	-35.46	35.04
標準誤差	0.39	5.42	5.44	0.29	1.81	1.84	5.74

Ⅲ-3. 引用距離に関する結果

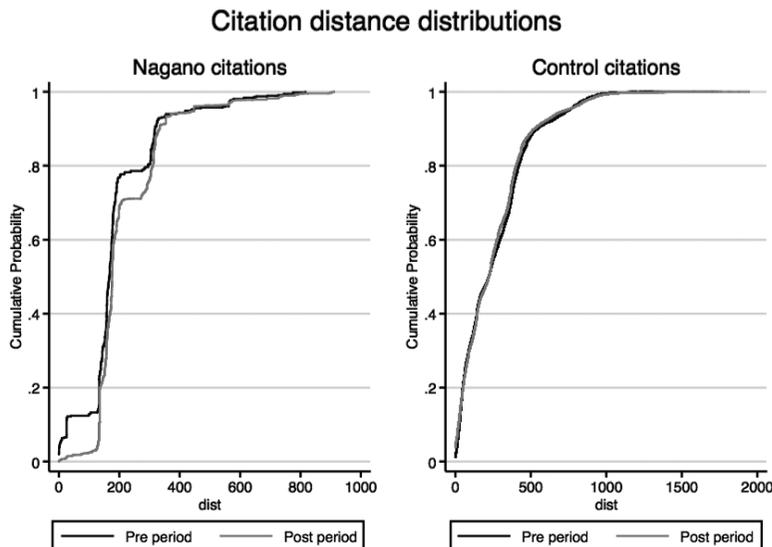
次に開通前後で引用距離の変化について「差分の差分法」を使った統計的分析を実行する。もし北陸新幹線の開通が特許引用距離の地理的拡大に影響を及ぼさないのであれば、開通前後での北陸新幹線沿線引用の距離の差分と圏外引用の差には有意な違いが見られないはずである。このことを調べるための引用サンプルとして北陸新幹線の沿線引用を処置群 (treatment group)、開通前後でどの新幹線からも離れている圏外引用を参照群として、差分の差検定を行った結果が、下の表3である。差分の差の推定値は正かつ有意である。よって、新幹線開通によって引用範囲は地理的に拡大し、距離の伸びは平均35kmである。

引用距離の伸びをさらに調べるために、北陸

新幹線開通前後でどの引用が距離的に大きな影響を受けたのかを図で示したのが図7である。引用距離に関する累積密度関数の形状から、参照群 (右側) については開通前後の影響をほとんどうけてない一方処置群 (左側) は特定の距離で大きな変化をおこしていることが分かる。具体的には100kmまでの近距離と、200-400kmまでの距離である。

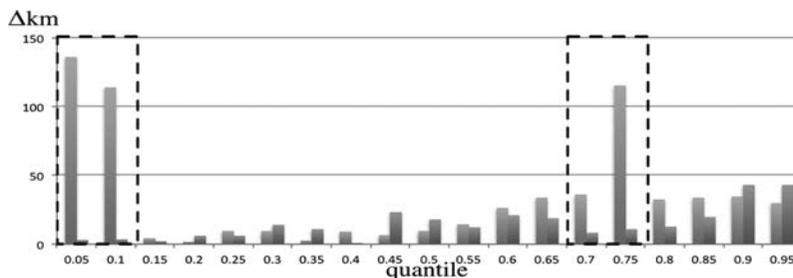
開通の影響が一樣ではないことを統計的に検討するために、分位点毎に差分の差法を適用したのが図8である。図中の引用距離の各分位点において、推定値及び標準誤差をそれぞれ左右の棒の高さで表現している。この推定結果から、開通前後で引用距離分布の分位点は5%、10%点で100kmから150kmほど遠方にシフトし、70%で40km、75%点で120kmほど遠方にシ

図7 開通前後での引用距離の累積密度分布（Pre-period: 1976-1997, Post-period: 1998-2012）



（注）北陸新幹線開通前（1976-1997）及び開通後（1998-2013）の2期間について、左図では北陸新幹線から他の新幹線路線駅までの引用（沿線引用）の距離分布の変化を、右図ではどの新幹線路線駅からも10km以上離れた引用（圏外引用）の距離分布の変化を示している

図8 分位点ごとの差分の差推定の実行結果



（注）横軸には引用距離分布の5%分位点から95%分位点まで5%刻みに表示している。各分位点で左側の棒は推定値の大きさを、右側の棒は標準誤差の大きさを表す

フトしている。これらの結果からいえるのは、開通後には近距離の引用（開通前の距離分布で50km程度）が相対的に大きく減少しているということである。開通前後の5%、10%距離を実現する沿線引用は、北陸新幹線駅（開設予定）近辺地域への引用から、東海道新幹線駅（特に新横浜駅）近辺への引用へと変化している。遠距離に関する有意な結果には東海道、山陽、東

北新幹線への引用が含まれており、明瞭な解釈は難しい。なぜこれらの分位点にて距離の伸長が起きたのか、引用被引用側の発明人の関係を調べるなどして個別に考察を加える必要がある。しかしながら、少なくとも近距離に関する結果より、新幹線開通によって沿線地域内の引用は相対的に減少していることは明らかになった。すなわち、知識のフローは新幹線の開通に

よって地域内から 100km 以上離れた外の地域へと伸びている。

IV. おわりに

本研究では、特許引用によって捕捉される知識移転に高速鉄道が与える役割を考察するために、1997年の北陸新幹線開通というイベントを用いた自然実験による実証分析を行った。新幹線駅から10km以内に位置する発明人による特許出願を処置群に設定し、どの新幹線駅からも少なくとも10km離れた特許を参照群として差分の差(DID)による分析を行った結果、グラフによる時間的推移の観察から処置群の引用頻度は開通後に増加し、相対的に新幹線沿線地域からの知識移転が増えていることが分かった。DID推定によって知識移転範囲の地理的变化については、開通後には平均して30km程度拡大していることが判明した。分位点によるDID推定によって近距離の知識移転は相対的

に減少していることも示され、開通によって知識移転の先は沿線地域内から外へ向いていることが明らかになった。北陸新幹線は沿線地域のイノベーション創成活動における知識移転を促進し、移転範囲を広げる効果があるとまとめることができる。

本研究の処置群及び参照群の構成にあたっては、時間を通して参照群から処置群へ、またはその逆方向に移動する企業や発明人を排除し、自然実験の仮定を崩さないようにサンプルを選択しているが、これによって知識移転が促進される地域に企業や発明人が自己選択によって移動するという興味深いケースは考察していない。このケースに関する分析手法を確立し、考察を行うことが今後の課題である。

参 考 文 献

- 後藤晃，元橋一之（2005）「特許データベースの開発とイノベーション研究」知財研フォーラム 63号，p. 43
- 中村健太『IIP パテントデータベースユーザーマニュアル』2015年7月
- Acs, Zoltan J., David B. Audretsch, and Maryann P. Feldman (1994) "R&D spillovers and recipient firm size," *The Review of Economics and Statistics*, vol. 76, no. 2, pp. 336-340.
- Acs, Zoltan, Luc Anselin, and Attila Varga (2002) "Patents and innovation counts as measures of regional production of new knowledge," *Research Policy*, vol. 31, pp. 1069-85.
- Agrawal, Ajay, Cockburn, Iain, McHale, John (2006), "Gone but not forgotten: Labor flows, knowledge spillovers, and enduring social capital," *Journal of Economic Geography*, vol. 6, no. 5, pp 571-591,
- Agrawal, Ajay, Devesh Kapur, and John McHale (2008), "How do spatial and social proximity influence knowledge flows? Evidence from patent data," *Journal of Urban Economics*, vol. 64, pp. 258-69,
- Arrow, Kenneth, J (1962) "Economic welfare and the allocation of resources for invention," in *The rate and direction of inventive activity*, ed. by R. Nelson, pp. 609-626. Princeton University Press, Princeton,

- NJ..
- Griliches, Zvi (1979) "Issues in Assessing the Contribution of R&D to Productivity Growth," *Bell Journal of Economics* 10, pp. 92-116.
- Jaffe, Adam B. (1986) "Technological Opportunity and spillovers of R&D," *American Economic Review*, vol. 76, pp. 984-1001..
- Jaffe, A.B., M. Trajtenberg, and R. Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, vol. 108, no. 3, pp. 577-598..
- Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura (2014), "Localized knowledge spillovers and patent citations: A distance-based approach", *Review of Economics and Statistics* vol. 96, pp. 967-985.
- Nakajima, Ryo, Ryuichi Tamura, and Nobuyuki Hanaki (2010) "The effect of col- laboration network on inventors' job match, productivity and tenure." *Labour Economics*, vol. 17, pp. 723-734..
- Saito, Yukiko Umeno, and Isamu Yamauchi "Inventors' Mobility and Organizations' Productivity: Evidence from Japanese rare name inventors," *RIETI Discussion Paper Series* 15-E-128.
- Saxenian, AnnaLee, *Silicon Valley's new immigrant entrepreneurs*, Public Policy Institute of California, 1999.
- Schumpeter Tamada, Yusuke Naito, Kiminori Gemba, Fumio Kodama, and Jun Suzuki (2006), "Significant Difference of Dependence upon Scientific Knowledge among Different Technologies," *Scientometrics*, Vol. 68, No. 2, pp. 289-324..
- Singh, Jasjit (2005) "Collaborative networks as determinants of knowledge diffusion patterns," *Management Science*, vol. 51, no. 5, pp. 756-770, .
- Stefano Breschi, and Francesco Lissoni (2009) "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows," *Journal of Economic Geography* vol. 9, pp. 439-468.
- Thompson, Peter, and Melanie Fox-Kean (2005). "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review*, 95: 450-460..
- Trajtenberg, Manuel, Gil Shiff, and Ran Melamed. (2006) "The NAMES GAME: Harnessing inventors' patent data for economic research." *NBER Working Paper* #12479..

補 論

発明者名の名寄せ法について

本研究では自己引用の検出などに一意に識別された出願人及び発明者の情報を用いている。「誰がどの特許出願を行ったか」一方、IIP パテントデータベースに収録されている特許書誌情報では、特許出願記録について出願人が記入したものがそのまま記載しており、その記入方法は自由形式である。そのため電子化されるまでに表記の揺れが発生する可能性がある。この表記の揺れは目視によって容易に検知することができるが、約 1000 万件を越える書誌情報に対して研究者が手作業で必要な修正を施すことは実質不可能であり、コンピュータープログラムを用いて自動化処理を実行することが必然的な選択であり、本補論ではプログラムを用いた名寄せ作業の方法を示す。

I. 名寄せ法の概略

「どの日本の民間企業がどの特許出願を行ったか」を明らかにする出願人名の名寄せについては、文部科学省科学技術・学術政策研究所による「NISTEP 辞書」が存在しており、約 90% の特許出願について出願番号それぞれについて出願人の名称が「識別番号」を介して紐づけられている。一方、「どの日本在住の発明者がどの特許出願を行ったか」を明らかにする発明者の名寄せについてはこれまで公開されている識別情報は存在していない⁴⁾。ここでは日本在住の発明者を一意に識別し、「どの発明者がどの特許出願を行ったか」について確定的な情報を得るための手法を述べる。

名寄せ法要点とは、表記の一貫性のなさに起因する同一性判定の誤りをコントロールし、これらの誤りをできるだけ減らすことである。

(A) 同一人物であるのに、誤って別人物と判

断してしまう誤り

(B) 別人物であるのに、誤って同一人物であると判断してしまう誤り

Trajtenberg, et. al (2006) によって提案された Computerized Matching Procedure (以下『CMP』と表記) は、発明者の氏名と住所、及びこの発明者が関わった特許出願の書誌情報を用いてこれらの 2 つの判断の誤謬をコントロールする。Nakajima, Tamura, and Hanaki (2010) では米国在住の民間企業発明者の識別に用いられている。

[第 1 段階]

CMP の第 1 段階では (A) の同一人物であるのに別人物と判断してしまう要因を列挙する。これらは 2 つに大別できる：表記に一貫性がないため、本来同一である発明者住所及び出願人名称・住所が異なると判断してしまい、よって同一人物性を判定する際に別人物と判断してしまう可能性、及び発明者の姓と名が同一であるにも関わらず、別人物と判断してしまう可能性である。これらの要因を第 1 段階ではできるだけ少なくさせるために、各出願の書誌情報全てに以下の処理を施す。

(1) 発明人・出願人の氏名と住所の表記の変動を可能な限り除去する。住所表記については具体的には、漢数字及びアラビア数字、全角半角表記が混在する番地表記を半角表記に統一する、都道府県名が省略されている場合にはそれを補う、地名レベル以上の住所文字列の変更、市町村合併を反映させて最新 (2014 年度) の住所表記に統一することなどが含まれる。また、氏名と名称については拗音を対応するカナに変更する、日本語入力ソ

4) なお、IIP パテントデータベースでは 2014 年度版までは発明者について「row」という項目があり記載された「氏名」と「住所」のペアで発明者を識別している。ただし、これらの表記の揺れを吸収したものではないために、同一人物を別人物として判断する誤りを犯す可能性は依然として存在する。

フトウエアでは何種類も存在するハイフンに類似した記号を全角ハイフンに変更するという処理、そして出願人名については特に官公庁及び大学からの出願で「代表者の人名」が組織名の後に続いている場合があるため、これらを削除する。

(2) 姓名が同一である場合には、この段階で「同一人物」とであるとみなし第2段階にうつる。

(1) の作業を名称及び住所の標準化と呼ぶ。そして、発明人の全レコードを(2)の基準によって同姓同名のグループに分割する。これをTrajtenberg, et. al (2006) にならい「PSet」と呼ぶ。「PSet」の要素は氏名が同一の発明人が関わった全ての特許出願であり、このうち発明人及び出願人の住所と氏名は(1)によって標準化されている。PSetの総数は165万5千件程度である。

[第2段階]

第1段階で作成されたPSetには同姓同名の別人物による特許出願が含まれている。これを異なる人物に分割する作業を実行する。分割には「スコアリング」という方法をとる。PSetの要素である全ての特許出願情報に関してペアマッチを行い、特許書誌情報の各項目が一致した場合にはあらかじめ定義してある「スコア」を加点し、総スコアがある閾値を超えた場合に、このペアは「マッチした」すなわち「同一人物

のものである」と判断する。具体的には「住所」「過去の共著者」「IPC技術分類」「引用出願」といった書誌情報の項目である。

II. 推移性の強制

このようなペアマッチ法でサンプルを分割する際問題となるのが、推移性の成立しない場合と呼ばれる現象である。具体的には特許出願「A, B, C」があるとき、「AとBがマッチ」して「BとCがマッチ」したとき「CとA」はマッチしていると期待されるが、実際には「CとAはマッチしていない」場合である。Trajtenberg, et. al (2006) ではこのケースでは推移律を強制して「CとAはマッチする」とみなすことを提案しており、本研究でもこれに従っている。推移律のみたされないマッチ結果の対処方法については以下の例を参照されたい。

〈例〉

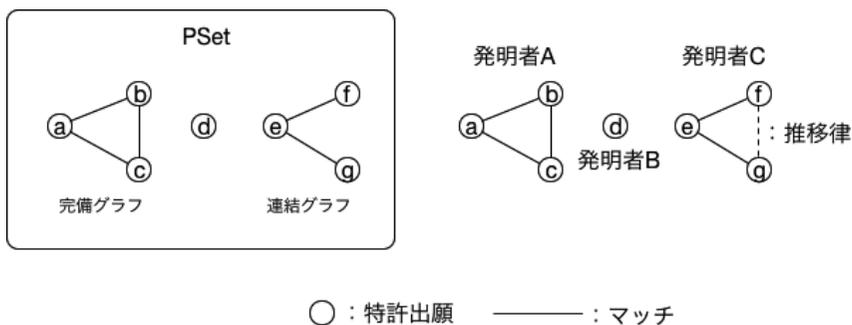
姓名が共通のPSetから発明者A, B, Cの3人を分割する作業を考える。このPSetの要素は{a, b, c, d, e, f, g}という7つの特許出願とする。第2段階のペアマッチによって、aとbとcが互いにマッチし、dはどれともマッチせず、eはfとgにマッチし、fとgはマッチしないとすると、下三角行列で図9のように表記することができる。

この結果はさらに、図10左図に示すような7つの特許出願をノード、マッチ成功の可否をエッジとする無向グラフによって表現すること

図9 マッチ結果行列

	a	b	c	d	e	f	g
a							
b	1						
c	1	1					
d	0	0	0				
e	0	0	0	0			
f	0	0	0	0	1		
g	0	0	0	0	1	0	

図 10 PSet から発明者への分割



ができる。ここでグラフの形状から $\{a, b, c\}$ をノードとするサブグラフは完備グラフ、 $\{d\}$ は単一ノード、 $\{e, f, g\}$ は完備ではないが一方のノードから他方のノードへエッジを介して繋がっている連結グラフである。上述のように、ノード f と g には推移律を課してマッチしたとみなす。完備グラフならば連結グラフであることに注意すると、PSet の出願を発明人単

位に分割する作業とは、このグラフ群から連結グラフを発見することと等価である⁵⁾。結果、この PSet は3つの「連結グラフ」があり、それぞれを構成する特許出願 $\{a, b, c\}$ 、 $\{d\}$ 、 $\{e, f, g\}$ が同姓同名ながら別の発明者 A, B, C によるものであるというように分割されたことになる。

5) PC 上の多くのグラフパッケージが連結グラフの発見機能を備えている。ここでは単一ノードも連結グラフとみなす *igraph* (<http://igraph.sf.net/>) を利用した。