

Reliance on Science データを活用した 特許・論文間引用分析

—ポストグラフェン二次元材料の二硫化モリブデン研究の構造—

Analysis of patent-paper and paper-paper citations using freely accessible “Reliance on Science” data. –
Research structure of molybdenum disulfide studies

国立研究開発法人科学技術振興機構

山下 泰弘

独立行政法人産業技術総合研究所技術と社会研究センター、山形大学評価分析室等を経て現職。計量書誌学分析に従事。

✉ yasuhito.yamashita@jst.go.jp

☎ 03-5214-7640

国立研究開発法人科学技術振興機構

吉田 秀紀

株式会社東芝研究開発センターを経て、現在 JST でファンディング業務や分析業務に従事。

1 はじめに

技術が細分化・複合化している現代においては、科学と技術の関わりもまた多様化が進んでいる。そのためファンディング現場においても、従来にも増して分野横断的な取り組みが求められている。特定技術分野でのみ利用可能なデータを整備することは、分野間の比較ができない点、さらに費用対効果、分析対象ごとの個別データ整備に時間を要する点からも効率的ではない。その点において、特許は幅広い技術領域をカバーする稀有な体系的データソースであり、特許による科学文献の引用は、科学研究成果の技術へのスピルオーバーの指標として期待されている。

国立研究開発法人科学技術振興機構においても、上記の視点から特許による科学論文の引用に注目してきたが（例えば、治部・松邑・斎藤 2012; 山下・治部 2017 など）、特許による科学論文引用のデータベースは、科学論文間の引用索引データベースほど汎用的なものではなく、導入コストも高額なため、定常的な利用には至っていなかった。しかしながら、近年 Digital Science 社の Dimensions が特許・論文間引用を含む広範なリンクデータを提供するようになった他、フリーで

利用できるデータとして、ボストン大学の Marx らが、Microsoft Academic Graph (MAG) のバルクデータを特許データと突き合せた特許・論文間引用 (PCS) リンクデータを、MAG の書誌データと合わせて公開し始めている（以下では MAG 書誌データを包含するデータセットを、Marx らの論文タイトル (“Reliance on Science”) にちなんで ROS データと呼称する）。

本稿では、上記の Marx らによる公開データを紹介した上で、我々が分析手法開発のテストフィールドとしているポストグラフェン二次元材料分野における主要な物質である二硫化モリブデン (MoS_2) の研究を対象とした試行的な分析を行う。

2 ROS データの概要と利用における留意点

2.1 概要

従来、PCS データは、古くは CHI リサーチ社が提供していた米国特許の引用データ、クラリベイト社の Derwent Innovations Index など限られたデータベースでのみ有償提供されており、論文データベースに加えてそれらを継続的に契約することはコスト面から困難であった。Marx と Fuegi は、フリーで入手可能な

論文データ (MAG) と特許データ (PatentsView が公開している米国特許バルクデータ) を突合することにより、PCS データを作成し、ウェブサイト (<http://relianceonscience.org>) で広く一般に公開した。彼らの取り組みで特徴的な点は、作成のプロセスやプログラムについても併せて公開していることである。それによって、彼らと同様の環境を構築すれば、最新のデータを自ら作成することも可能となっている。当該データの作成方法と内訳については、Marx and Fuegi (2020) において説明されている。筆者らは ver.23 をダウンロードして使用しているが、2020年8月31日現在の最新バージョンは、2020年8月7日に公開された ver.24 となっている。同日に複数のマイナーリビジョンが公開されているが、筆者らはその中でも2020年7月時点で最新と思われる DOI:10.5281/zenodo.3755799 を使用している。Marx and Fuegi (2020) によると、収録期間は、米国特許は1947～2018年、米国以外の特許は1782～2018年、論文は1800～2018年とされている。しかしながら、ダウンロードしたデータには2020年のデータも含まれており、バージョンアップの都度データも更新されているものと思われる。

論文書誌は MAG のバルクデータから必要事項を切り出すだけの処理だが、書誌マッチを行う PCS リンクテーブルにはバージョン間での精度の違いが顕著に表れ得る。ver.21 までは、米国特許のカバーページ由来の引用のみのテーブルと本文から抽出した引用も含むベータ版テーブルが提供されていたが、ver.22 および 23 では後者のみとなった。2020年8月31日現在で最新の ver.24 では、ベータ版ではなくなっている¹。

2.2 ROS データの利用における留意点

このデータは、研究者が自発的に公開しているものであり、論文やウェブ上のドキュメントでその諸元が公開されているとは言え、商用データと比べて不明な点も多い。筆者らも利用に当たっては、ある程度の検証が必要となった。本節では、後続の利用者の便を考慮し、これまでに筆者らが得た留意点を整理したい。現在筆者ら

が利用しているデータは、すでに最新ではなく、また PCS リンクデータがベータ版であるため、最新データに必ずしも当てはまらない可能性がある点はお含みおきたい。

MAG の論文データは、書誌事項や論文間引用データまで ROS データに含まれているが、特許については、PCS リンクテーブル (`_pcs_mag_bodytextbeta`) に特許番号が含まれているのみである。そのため、非特許文献を引用した特許をより詳細に把握するためには、別途特許データを入手する必要がある。

そこで、ROS の特許データのソースとなっている PatentsView のダウンロードサイト (<https://www.patentsview.org/download/>) から `patent.tsv` (データの最終更新日は2020年6月10日) をダウンロードして、特許データを引用リンクテーブル中の特許番号と確実に紐付けることができるかを検証した。また、PCS リンクテーブル (`_pcs_mag_bodytextbeta`) に付されたデータ項目にも若干の不明点があり、検証を進めている。それらについて、現時点で分かったことも報告したい。

筆者らが利用している PCS リンクテーブルは、34,959,193 件のレコードを含む。そのうち米国特許からの引用リンクは、表1において `uspto` 項目が“1”または空白となっている 30,711,637 件である。この米国特許番号を PatentsView よりダウンロードした米国特許データと紐付けた結果、30,457,015 件 (99.2%) をリンクさせることができた。`uspto` フラグが“1”の特許のみについての検証であるが、リンクされなかった 31,613 件から任意で数件抽出して、USPTO のサイトで検索したが、該当レコードはなかったため、取り下げ等の何らかの理由で削除されたものと思われる。

`uspto` フラグについて、“1”と空白の違いは公開されているドキュメントのみからはわからないが、特許文献中でのリファレンスの記載箇所の分布は明らかに異なる。リファレンスの記載箇所がフロントページのみの場合、すべて `uspto` フラグが“1”、本文のみの場合には空白である。ただし、フロントページと本文の両方に含まれる場合は、どちらのフラグもあり得るので、記載箇所のみに基づいて付与されているとは言えな

1 このバージョンアップにおける大きな変更点として、2つのテーブルに分かれていた MAG と Pubmed の PCS リンクデータが1テーブルに集約された。

表1 PCS リンクテーブルの内訳

	uspto 項目			Total
	1	(blank)	0	
wherefound 項目				
frontonly	14,327,594	0	0	14,327,594
both	2,787,411	1,102,829	0	3,890,240
bodyonly	0	12,493,803	0	12,493,803
(blank)	0	0	4,247,556	4,247,556
計 (a)	17,115,005	13,596,632	4,247,556	34,959,193
うち米国特許あり (b)	17,083,392	13,373,623		
b/a	99.8%	98.4%		

い²。Marx らは PCS リンクの確からしさ (confidence score) を 1 から 10 (ただし PCS リンクデータにはスコア 2 以下のリンクは含まれていない) の数値で表しているが、リファレンスの記載箇所ごとにその平均を取ると、「フロントページのみ」と「フロントページと本文の両方」のリンクは、それぞれ 9.4、9.6 であるのに対し、「本文のみ」のリンクは 6.8 と顕著に低い。実際にリファレンスの記載箇所ごとに、数件の PCS リンクを任意抽出し、USPTO のサイトで特許文献に当たって該当する論文が実際に特許文献中に出現するかを目視確認したところ、「フロントページのみ」と「フロントページと本文の両方」については 100% 該当論文が記載されていたが、「本文のみ」については 10 件中 6 件に止まった。

DOCDB に含まれる米国以外の特許からの引用も PCS リンクテーブルに含まれているが、それについては詳しいドキュメントがなく、スクリプトも公開されていない。米国特許以外についてのみ DOCDB パテントファミリーと紐づけられている。何故このような仕様になっているかは著者らには不明である³。

3 MoS₂ 研究の構造と技術との関連性の把握

前章の検証結果から、筆者らがダウンロードした ROS データの PCS リンクデータについては、精度の面から米国特許のフロントページ由来のものを抽出して

- 2 米国特許の処理についてはスクリプトが公開されているので、解析すれば付与ルールがわかる可能性はあるが、筆者らは現時点ではそこまで処理しきれていない。
- 3 この点はウェブページにある質問用メールアドレスに問い合わせたが、2020 年 8 月 31 日時点ではまだ回答をいただけていない。

使用することが好ましいという証左を得た。以下ではそれを踏まえて、代表的なポストグラフェン二次元材料である MoS₂ の研究構造と技術との関連性の把握を試みる。

3.1 二次元材料としての MoS₂

2004 年に初めて単離されたグラフェンは重要な二次元材料として注目されているが、バンドギャップが存在しないため、半導体としての応用は困難であった。半導体に応用可能なポストグラフェン二次元物質の一つとして、2011 年に Radisavljevic らは、MoS₂ の半導体特性を明らかにした (Radisavljevic et al. 2011。以下 R2011 と呼称する)。それを契機として、MoS₂ が属する遷移金属原子 1 つと酸素族元素 (カルコゲン) の原子 2 つの組み合わせからなる二次元物質群である遷移金属ダイカルコゲナイドの多様な物質の探索が進むとともに、MXene、シリセン、ゲルマネンなどの新たな二次元物質が創製された。

今日では、多様な二次元物質が、触媒、トポロジカル絶縁体、半導体 (トランジスタ、センサ、発光素子等) 等幅広い用途に向けて研究されている。MoS₂ は、自然界に存在する物質であり、強い毒性もなく、従来から潤滑剤として用いられるほどありふれているため入手が容易であるなど、実用上多くの利点があり、ポストグラフェン二次元物質の中でも、非常に幅広く研究されている物質である。本稿では、二次元物質としてのみならず、古くから継続されている潤滑剤としての MoS₂ の研究も含めて計量分析を行い、その相互の関連性についても明らかにしたい。

3.2 使用するデータ

前章で検証を行った ROS データに対し、以下のよう
に前処理して、特許データも含む分析データを構成した。

まず、論文については、ROS データの分野分類
(fieldidname テーブルの fieldname 項目)において、
“molybdenum disulfide” が付与されたものを MoS₂
研究論文として抽出した。ROS データには非常に雑多
な種類の文献が含まれているので、定期刊行物（ジャー
ナル ID が付与された媒体）に収録された文献に限定す
ることで文献の質を一様にした。通常計量書誌学分析
に用いられる商用データベース（Web of Science、
Scopus 等）に含まれるドキュメントタイプに相当す
るデータ項目が存在しないため、これ以上のふるい分け
は行っていない。本稿ではデータセット内での引用関係
を分析するため、データセット内の他の論文ないし特許
と引用関係を持たない孤立した論文は除外した。上記の
処理によって、3,352 件の論文データを得た。

PCS リンクデータ (_pcs_mag_bodytextbeta テー
ブル)は、米国特許のみを使用する。米国特許の抽出の際、
uspto フラグを無視し、特許番号で PatentsView より
入手した米国特許データにリンクされるデータを米国
特許と見做した。特許本文にのみ記載された引用文献は、
高い精度が期待できないことから除外し、フロントペー
ジに記載された引用文献のみ（wherefound 項目の値
が“frontonly”または“both”のレコード）を分析対
象とした。前述の MoS₂ 論文データとリンクさせるこ
とにより、505 ペアの PCS リンクデータを得た。

論文間引用については、ROS データに含まれるもの
を使用した。MoS₂ 論文間の引用リンク数は、18,528
ペアである。本稿では特許からの論文引用の発生までに
注目し、その後発生する特許間引用については簡略化
のため捨象した。

3.3 分析方法

次章では、MoS₂ 研究の態様に大きな変化をもたらした
と考えられる R2011 に注目し、その出現前後での
研究の構造変化を分析する。

研究の構造を可視化するために、本稿では直接引用
関係に基づいてクラスタリングを行う。柴田らは、直接
引用、共引用、書誌結合の3つの方法で構築されたネッ
トワークの比較を行い、リサーチフロントの識別におい

て直接引用が最も良好な結果を出すことを示している
(Shibata et al. 2008)。直接引用は、短期間では引
用リンクが十分に形成されないため、長期間の観測に適
している(Klavans and Boyack 2017)。直接引用は、
共引用、書誌結合と異なり情報が要約された形で表示さ
れないので、構造の視認性はやや低くなるが、引用リン
クにウェイト付けされないため、結果の恣意性は低いと
言える。

直接引用では、非常に大きなサイズのクラスタが形成
されることが多いので、コミュニティ分割により、サブ
クラスタに分割する。コミュニティ分割の近似計算法に
はスピングラスを用いる。

論文出版に対応した特許出願動向を把握するため、特
許については公開年ではなく出願年に基づいて集計を行
う。

4 結果

4.1 MoS₂ 研究の可視化

MoS₂ 論文とその引用特許の引用関係の俯瞰図を図
1に示す。青いノードが MoS₂ 論文、赤いノードが本
稿で特に注目する R2011、緑が特許である。R2011
における単層 MoS₂ の半導体特性の発見により大幅に
研究の潮流が変化したにも関わらず、若干の例外を除い
て、すべての論文が引用関係で結合されていることがわ
かる。特許による論文引用は、論文間引用と比較してか
なり少数であり、また R2011 を含む一部の論文に偏
在している。

最大連結成分（最もサイズが大きいクラスタ）は極め
てサイズが大きいので、研究トピックの構造的な把握を
行うため、コミュニティ分割を行って、19のサブクラ
スタ（以下研究領域と呼称する）に分割した。19の研究
領域のうち8領域は論文数10件未満と極めて小規模
なため捨象する。残りの11領域の内訳は表2の通りで
ある。特許数は各研究領域に含められる特許数であり、
必ずしも論文の被引用数とは合致しない。また、右列の
「分野」は、ROS データに付与されている“molybdenum
disulfide”以外の分野を、頻度上位10件まで表示し
ている。

潤滑／トライボロジー（#4、#13）、脱硫（#6）、
複合材料（#19）に関する研究領域は、相対的に古く、

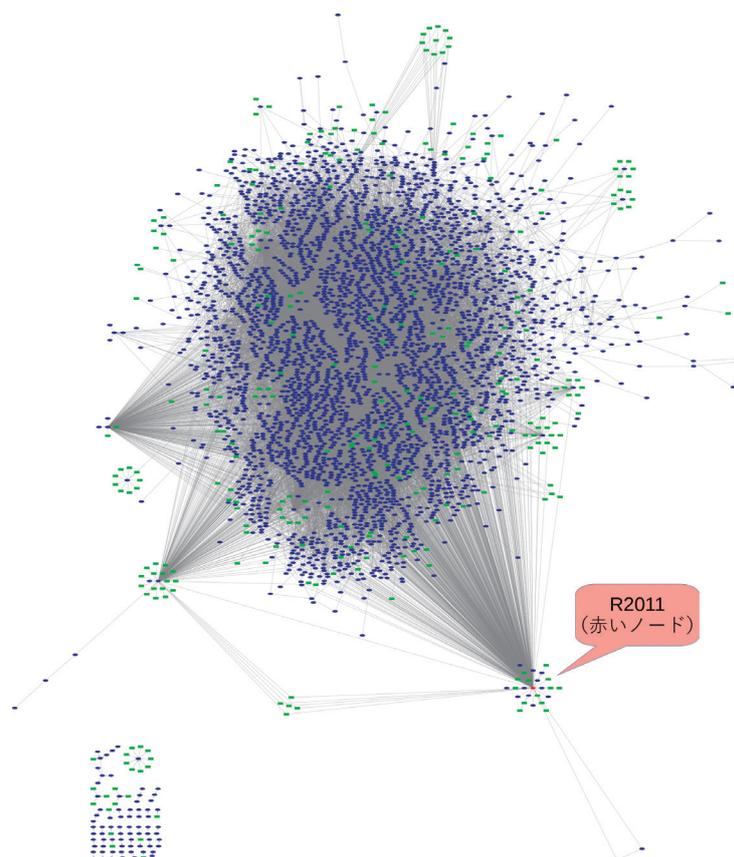


図1 MoS₂ 研究の直接引用ネットワーク

表2 論文数 10 以上の研究領域の内訳

#	論文数	特許数	平均年		分野
			論文出版	特許出願	
1	387	3	2015.6	2014.3	inorganic chemistry; materials science; chemistry; lithium; anode; electrochemistry; chemical engineering; nanotechnology; graphene; electrode
3	480	13	2016.7	2016.8	inorganic chemistry; chemistry; catalysis; materials science; tafel equation; overpotential; chemical engineering; electrocatalyst; nanotechnology; electrochemistry
4	41	2	2010.7	2010.0	materials science; composite material; metallurgy; tribology; lubricant; microstructure; powder metallurgy; composite number; sintering; dry lubricant
5	887	110	2016.1	2017.0	materials science; monolayer; optoelectronics; analytical chemistry; nanotechnology; physics; chemistry; chemical vapor deposition; condensed matter physics; graphene
6	75	9	2010.7	2011.8	chemistry; inorganic chemistry; catalysis; hydrodesulfurization; molybdenum; organic chemistry; heterogeneous catalysis; materials science; sulfur; dibenzothiophene
7	46	0	2016.3	-	saturable absorption; physics; optics; optoelectronics; fiber laser; laser; materials science; ultrashort pulse; pulse duration; nanotechnology
11	437	24	2016.7	2017.8	chemistry; inorganic chemistry; materials science; nanotechnology; analytical chemistry; detection limit; graphene; nanocomposite; exfoliation joint; biosensor
13	308	39	2007.0	2013.2	materials science; composite material; metallurgy; tribology; chemistry; lubricant; inorganic chemistry; lubrication; nanotechnology; analytical chemistry
17	86	6	2016.8	2011.0	materials science; chemistry; analytical chemistry; nanotechnology; inorganic chemistry; chemical engineering; nanocomposite; heterojunction; x ray photoelectron spectroscopy; graphene
18	140	25	2017.4	2018.2	chemistry; inorganic chemistry; adsorption; materials science; chemical engineering; membrane; aqueous solution; graphene; nanotechnology; analytical chemistry
19	368	48	2009.5	2009.3	materials science; inorganic chemistry; chemistry; composite material; nanocomposite; chemical engineering; intercalation; organic chemistry; nanoparticle; polymer

表3 特許からの被引用数上位10論文

順位	被引用数	第一著者	タイトル	ジャーナル	出版年	研究領域
1	42	Branimir Radisavljevic	Single layer MoS ₂ transistors	Nature Nanotechnology	2011	#5
2	33	Manishkumar Chhowalla	The chemistry of two dimensional layered transition metal dichalcogenide nanosheets	Nature Chemistry	2013	#11
3	31	Kengku Liu	Growth of large area and highly crystalline MoS ₂ thin layers on insulating substrates	Nano Letters	2012	#5
4	24	Ke Liu	Atomically thin molybdenum disulfide nanopores with high sensitivity for DNA translocation	ACS Nano	2014	#18
5	18	L Britnell	Field effect tunneling transistor based on vertical graphene heterostructures	Science	2012	#5
6	17	Yih sien Lee	Synthesis of large area MoS ₂ atomic layers with chemical vapor deposition	Advanced Materials	2012	#5
7	16	Yuda Zhao	Two dimensional material membranes an emerging platform for controllable mass transport applications	Small	2014	#18
8	15	Arpana Verma	Tribological behavior of deagglomerated active inorganic nanoparticles for advanced lubrication	Tribology Transactions	2008	#19
9	12	Xianguo Hu	On the size effect of molybdenum disulfide particles on tribological performance	Industrial Lubrication and Tribology	2005	#19
10	10	Goki Eda	Photoluminescence from chemically exfoliated MoS ₂	Nano Letters	2011	#11

いずれも平均出版年が R2011 の出版以前である。それ以外の研究領域の多くは、リチウムイオン電池 (#1)、単層/オプトエレクトロニクス/グラフェン (#5)、光子吸収/ファイバーレーザー (#7)、バイオセンサー (#11) など、二次元材料としての MoS₂ に期待されている研究であり、相対的に新しい。

個別の論文の特許からの被引用数に注目すると、上位論文は表3の通りである。トップは、本稿で注目している R2011 であり、MoS₂ 論文群が得ている 505 引用 の 8.5% を占めている。表中で R2011 が属する研究領域 #5 の論文が上位を占めており、技術応用への期

待の高さが伺われる。8位と9位にトライボロジーに関する論文が含まれているが、いずれも R2011 以前に発行されたものであり、二次元材料が近年の研究の主流となっていると考えられる。

4.2 R2011 が MoS₂ 研究に与えた影響

R2011 が MoS₂ 材料研究に対して非常に高いインパクトを示していることは明らかであるが、MoS₂ 研究の各領域の発展にどのような形で寄与したのであろうか。

図2に表1に示した MoS₂ 各研究領域からの



図2 各研究領域からの R2011 の引用状況

R2011 を引用する論文の数と割合を示す。R2011 は MoS₂ 分野において非常に高い引用を得ているが、属する研究領域（#5）において、その割合が顕著に高く、R2011 が引用される機会がない 2010 以前の論文まで含む集計にも関わらず、研究領域 #5 の論文の約 60% から引用を受けている。すなわち、この研究領域形成の原動力とも言えよう。一方、表 1 に挙げたすべての研究領域から引用されていることから、従来からの研究領域であるトライボロジーを含め、MoS₂ 研究全般に対して、R2011 はインパクトを与えている。

R2011 出版後、研究領域 #5 にはどのような変化が生じたのであろうか？図 3 に R2011 出版後 2 年間の研究領域 #5 の引用ネットワークの変化を示す。出版直後の 2011 年には R2011 が 4 件の引用を得ているが、研究領域内の引用は極めて疎であり、この時点では研究領域が形成されていないといえる。図中に引用の向きを示していないが、R2011 は図示された研究領域 #5 の論文を引用しておらず、引用している MoS₂ 論文は、SFM によって MoS₂ 分子層の計測を行った研究領域 #11 の論文 1 報のみである。翌 2012 年には 2 報の重要なフォロワーが出版され、R2011 とそれらが共引用される形で研究領域 #5 内の引用ネットワークが形成され始める。それらは大規模原子層薄膜合成に関する Lee らの研究と Liu らの研究である。より実用に近い研究が早期に進められたことで R2011 において創出された MoS₂ デバイス研究の関連研究が急速に発展したと思われる。2012 年になると R2011 と直接

の引用関係を持たなかった Britnell2012 が統合され、引用ネットワークが密になり始める。Britnell2012 は、R2011 とは直接の引用関係を持たないのみならず、特許からの引用が先行している点も特徴的である。グラフをベースに、窒化ホウ素または MoS₂ をトランスポート障壁として、複合したトンネル FET に関する研究であり、実用に近い性質のものと考えられる。

5 まとめ

本稿では、公開された PCS データを活用して、近年注目を集めている二次元材料研究の特許と論文の引用ネットワークの形成過程を分析した。ROS データの存在は、これまでデータがなくて研究が進まなかった特許・論文間の引用に関する包括的な分析を加速するものと思われる。同データを公開し、バージョンアップし続けている Marx らのグループに感謝したい。

我々のグループにおいても、ROS データは導入して間もなく、まだその性質を熟知しているとは言えない。本稿で紹介した分析は、今後同データに基づく分析を進めるための試行的なものであり、より深めていく必要がある。本稿では視認による解釈に止まったが、特徴的な研究論文のネットワーク指標とその変化についても分析していく必要がある。

試行的な分析ではあるが、本稿からは二次元材料研究についていくつかの示唆が得られた。1 つには R2011 が MoS₂ デバイス研究の研究領域を形成したこと、そ

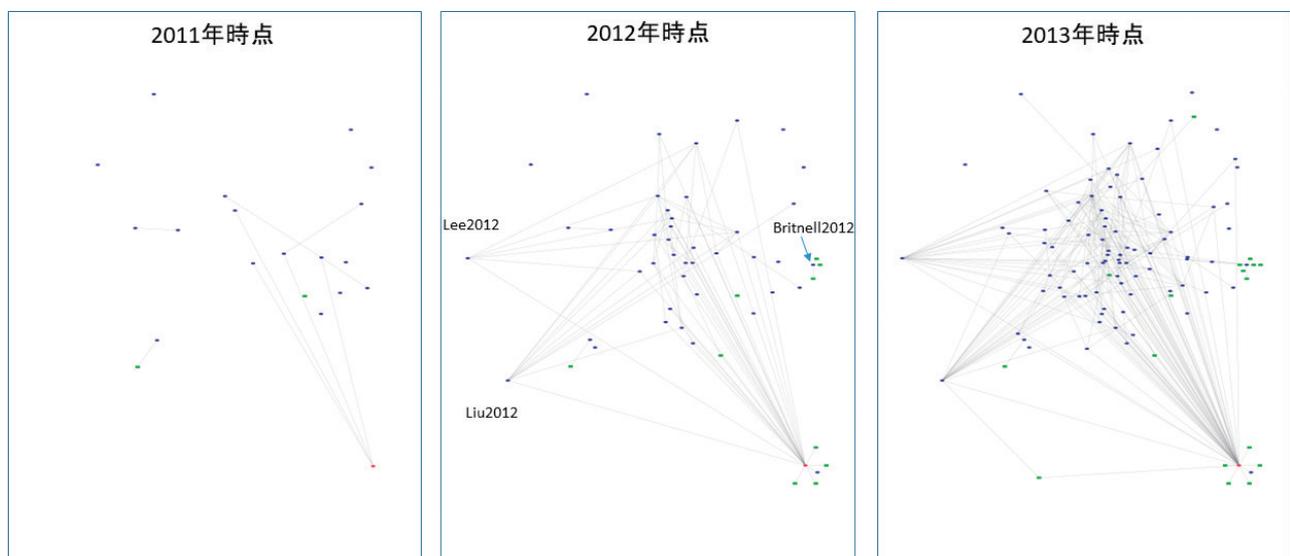


図3 領域 #5 の発展過程 (2011-2013)

れに際してそれをフォローする複数の重要な論文がサポートしたことがデータで示されたことである。また、R2011とは独立になされた Britnell らの研究が後続研究によって、引用ネットワークに取り込まれていく過程も観測された。今後は、R2011 以外のこのような特徴的な研究の役割についても、分析を進める必要がある。

参考文献

- [1] 治部真里, 松呂勝治, 斉藤隆行. (2012). 「J-GLOBAL foresight の構築について」. 情報管理. Vol.54, pp.639-651. <https://doi.org/10.1241/johokanri.54.639>.
- [2] Klavans, R. & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, 68: 984-998. <https://doi.org/10.1002/asi.23734>.
- [3] Marx, M. & Fuegi, A. (2020). Reliance on science: Worldwide front - page patent citations to scientific articles. *Strategic Management*. 41, 1572-1594. <https://doi.org/10.1002/smj.3145>.
- [4] Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. (2011). Single-layer MoS₂ transistors. *Nature Nanotechnology*. 6: 147-150. <https://doi.org/10.1038/nnano.2010.279>.
- [5] Shibata, N., Kajikawa, Y., Takeda, Y. & Matsushima, K. (2010). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, 60: 571-580. <https://doi.org/10.1002/asi.20994>.
- [6] 山下泰弘, 治部真里. (2017). 「特許からの論文引用に関する定量的評価指標の検討—日本の研究機関についての試行的分析—」. Japio YEAR BOOK 2017. pp.144-155.