

Title	計量書誌学分析によるホットトピック抽出の試み：ポストグラフィエン研究を事例として
Author(s)	山下, 泰弘; 吉田, 秀紀
Citation	年次学術大会講演要旨集, 34: 148-151
Issue Date	2019-10-26
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/16487
Rights	本著作物は研究・イノベーション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 Research Policy and Innovation Management.
Description	一般講演要旨

1 E 0 3

計量書誌学分析によるホットトピック抽出の試みーポストグラフェン研究を事例としてー

○山下泰弘, 吉田秀紀 (JST)

E-mail yasuihiro.yamashita@jst.go.jp

1. はじめに

新興領域に適時適切にファンディングを行うためには、領域が勃興する兆しを早期に捉える必要がある。ヒアリング等の質的な調査はその重要な手段であるが、俯瞰性・客観性・効率性を担保するためには、定量的なデータ分析によって質的調査を補完することが好ましい。そのような観点から、JST プログラム戦略推進部では、計量書誌学的手法を援用し、ホットトピック抽出と、その特性の把握を試みている。

分析手法の開発等を効率的に進めるため、現在は材料科学を事例的に取り上げ、分析とその手法開発を進めている。本稿では、1 E 0 2 で別途報告する田中他「材料科学リサーチフロントの体系化と著者所属国割合の比較」で網羅的に整理した材料科学関連 330 リサーチフロント(RF)のうち、特に近年論文数が顕著な増加を示しているポストグラフェン系 2 次元材料に関連する 30 リサーチフロント(RF)を事例とした、ホットトピック抽出について、現在進行中の取り組みを報告する。

2. データと分析方法

2.1. 使用するデータ

本研究では、データソースとして、クラリベイトアナリティクス社の Research Front の 2018 年版を使用する。これは、2013~2018 年に出版された論文のうち、被引用数上位 1%に入るものを共引用によりクラスタリングしたものである。論文数は、27,614 報 (2,443RF) である。本稿で扱う 30 リサーチフロントのコアペーパー数は 343 報である。

Research Front の論文は、論文の共引用クラスタリングにより作成されているが、製品には論文間の関係についての情報は含まれていない。論文間引用マップによる 30RF の構造の直観的な把握と、次節の分析②の論文間引用に基づく指標算出のため、Web of Science における論文間の引用関係のデータを使用した。また、後述する特許・論文間引用指標の NR-PCP 算出のために、DII(Derwent Innovation Index) における特許・論文間の引用関係を使用した

2.2. 分析方法

本研究では、まず、①対象 RF のコアペーパーを書誌結合によりクラスタリングし、RF 群の全体構造を俯瞰的に分析する。それを踏まえて、②特徴的な RF について、下記の二つの観点から試行的にその特性を分析する。

i. 先行研究からの革新的変化

先行研究に対する革新的変化の指標として、Wu、Wang 及び Evans (Wu, Wang, & Evans, 2019) によって disruption 指標が提案されている。disruption は $D=(n_r n_j)/(n_r+n_j+n_k)$ で表される。ここで、 n_r は対象論文を引用する論文のうち、対象論文とリファレンスを共有しない論文の数、逆に n_j は共有する論文の数、 n_k は対象論文の出版後に発行された、対象論文を引用せずそのリファレンスのみを引用する論文の数である。対象論文を引用する論文が、対象論文とリファレンスを共有しないほど大きくなる。取得データの制約から、本稿では n_k を除外し、 $D=(n_r n_j)/(n_r+n_j)$ を disruption の指標として使用する。従って、Wu らの定義では対象論文の発行によってそのリファレンス論文が参照されなくなる度合を指すと考えられるが、本研究では新たな潮流が派生された度合を示す。この定義では被引用数が 0 のケースは指標値を算出できないが、本研究で扱う論文は高被引用論文であるため問題は生じない。

ii. 産業技術との潜在的関連性

特許からの論文引用は科学知識のスピルオーバーの重要な指標であるが、引用の観測に時間を要するため、新しい論文に対しては適用が困難である。そこで、その代替として、新しい論文の技術との潜在的な近さの指標として、筆者の 1 人 (Yamashita, 2019) が提案した NR-PCP (特許に引

用された参照論文の数)を用いる。

3. 分析結果

3.1. 概観

2次元材料関連 30RF の書誌結合クラスタリング結果を図1 (全体) 及び図2 (部分拡大) に示す。RF ごとに着色し、外縁に分布する RF についてはコアペーパーのタイトルに基づき任意でキーワードを付与した(物質名が明記されない場合はアブストラクトも参照した)。図1を見ると、中央やや左下の凝集した部分に励起等の光物性物理に関連する RF、図の周縁部には個別的(物質別・用途別)の RF が布置している。密集部には多様な RF の論文が混在しており、後続研究での利用(共引用)の面で複数のクラスタに分かれる研究であっても、依拠する研究の共通性(書誌結合)は比較的高い。左側には光物性に関係する RF (#13075、#16171 等)が布置し、右側に水素発生に関係する複数の RF (#16927、#13094 等)が布置している。また、右下にがん治療に関する RF (#17895)が存在する点も特徴的と言える。物質別では、シリセン(#19402)、MXene(#17523)等が個別のクラスタで重点的に研究されているのに対し、MoS₂は図の下半分の幅広い RF で研究されており、研究の主流をなしていると考えられる。

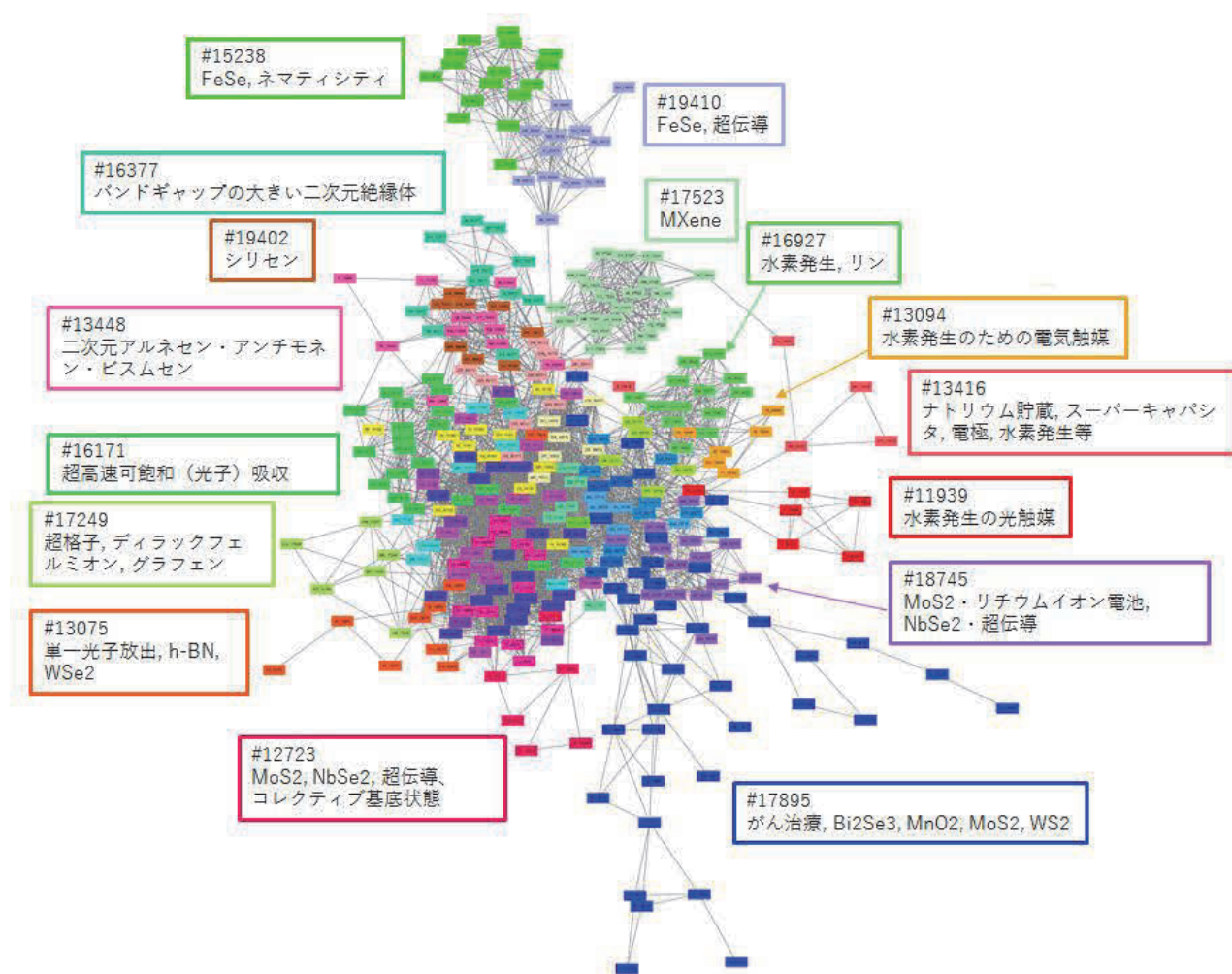


図1 2次元材料関連 30RF を構成するコアペーパーの書誌結合マップ
(2013-2018, cosine>=0.1)

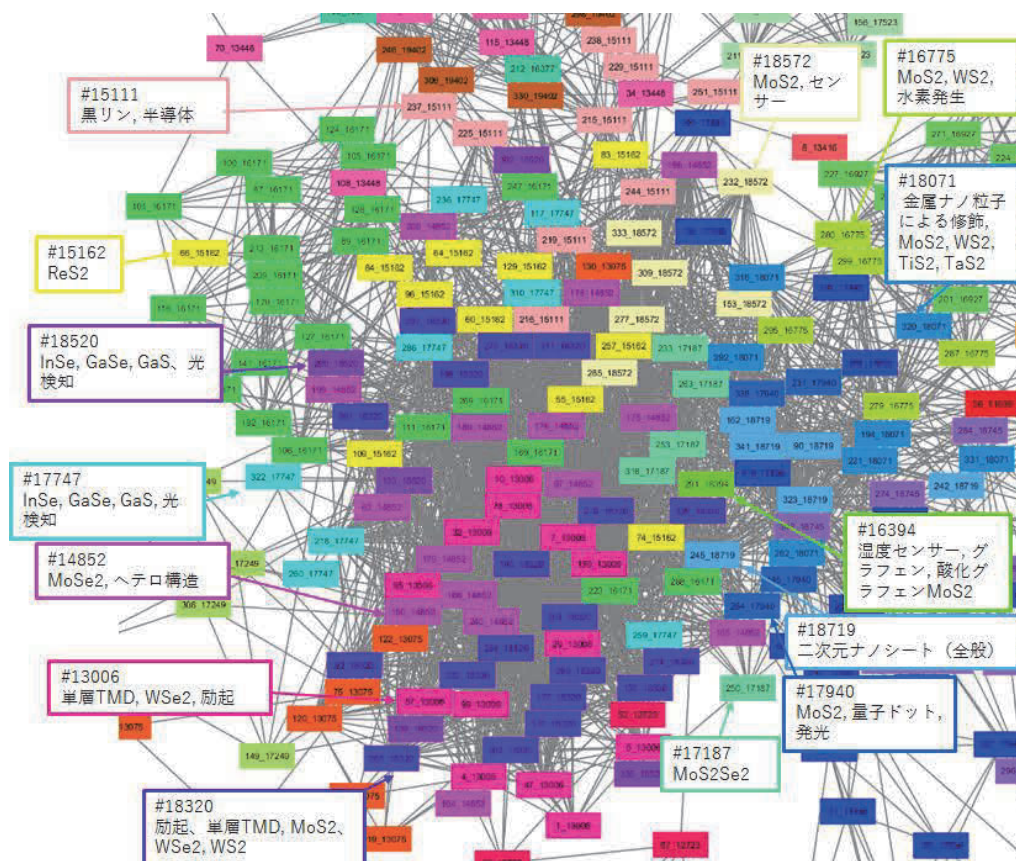


図 2 2次元材料関連 30RF を構成するコアペーパーの書誌結合マップ(部分拡大)
(2013-2018, cosine>=0.1)

3.2. 個別 RF の革新性と産業技術との関連性

次に、RF と個別の論文の性質について科学的な革新性と潜在的な技術関連性の 2 点について深掘りする。革新性の指標 (disruption) の算出には、コアペーパーを引用する各論文のリファレンスについての情報を要し、現時点で網羅的な分析が困難なため、3つの視点から表 1 の 6RF を選定し、試行的に分析を行った。なお、NR-PCP の算出に 2017 年までの論文データを使用したため、2018 年の論文 (1 編) については分析対象外とした。

表 1 分析対象 RF

RF_ID	略称	論文数 (2013-17)	備考
#11939	水素発生 1	7	明確な応用目的のための多様な観点からの複数RFがクラスターを形成しており、TMD研究の中でも重要課題と考えられる。
#13094	水素発生 2	8	
#16775	水素発生 3	5	
#16927	水素発生 4	13	
#17523	MXene	21	独立性の高い大規模RFで、全論文が高いNR-PCP値を示しており、潜在的にトピックの技術関連性が高いと考えられる。
#13006	励起	12	他のRFと密にリンクした基礎的なトピックのRF。他の応用的RFとの比較のために抽出した。

光触媒材料や燃料電池電極触媒などへの応用が明確な水素発生関連 4 RF (#11939、#13094、#16775、#16927)、後ほどデータを示すが、全論文の NR-PCP 値が高く潜在的に産業技術と相対的に関連性が強いと考えられる MXene 関連の#17523、逆に産業応用と距離があると考えられ、かつ RF クラスタの中核をなす光物性物理系の#13006 を対象とする。ここで、潜在的な技術関連性の指標として用いた

NR-PCP はリファレンスの出版年の影響を受け易いため、各論文の条件をそろえるため、出版年別に図示する。

革新性の面では 2017 年の#13094 の論文 1 報を除きすべての論文がマイナスの値を取っている。すなわち、後続の論文がコアペーパーの引用論文も引き続き引用する傾向があるということであり、ここで挙げたコアペーパーは過去からの安定した潮流の中に位置する質の高い論文であると言える。Wu ら (2019)によるとレビュー論文は平均的に disruption がマイナス、オリジナル論文はプラスの値を取るとされており、RF が両者を含むことを勘案しても、分析対象とした RF は相対的に disruption 値が低いと考えられる。一方、潜在的技術関連性 (NR-PCP) を見ると、近年になるほど平均的に各論文の値は低下しているが、RF ごとに見るとほぼ安定している。



図 3 各年のコアペーパーの disruption 及び NR-PCP (2013-17)

RF ごとに 2 軸で類型化すると、革新性が低く、潜在的技術関連性も低い左下領域には、#13006 (光物性) が布置している。この RF は、相対的に新しいため、後続特許および論文の引用により布置が変わる可能性があるが、2015 年以降のすべての論文が左下に位置している。一方、革新性が低く、潜在的技術関連性が高い RF としては、#17523 (MXene)、#16775 (水素発生 3) が顕著である。革新性が相対的に高く、潜在的技術関連性が低い RF には、#13094 (水素発生 2)、革新性・潜在的技術関連性が比較的高い RF としては、#11939 (水素発生 1) が挙げられる。ただし、#11939 (水素発生 1) は新しい RF であり、disruption、NR-PCP とともに論文ごとに大きなばらつきがある。#13094 のコアペーパーのうち 2017 年の 1 報のみ例外的に disruption が正値を取っているが、実際の被引用数を見ても、 n_i (disruptive な引用) が 37 回と比較的多く、この布置は論文の新しさに起因するものではないと考えられる。また、同年の#11939 (水素発生 1) の論文にも disruption が -0.15 と相対的に高い論文があり、こちらは NR-PCP 値も高いためトピックの潜在的な技術関連性も高いと考えられる。これらの論文は新たなトピックの創出に寄与している可能性もあるため、これらのサイティングペーパーのトピックについて別途分析を進める必要がある。

4. まとめ

本研究では、材料科学におけるホットな研究領域である、2 次元材料研究の構造を俯瞰し、特徴的な RF について、科学及び技術の 2 観点からの指標を適用して、特性の分析を行った。進行中の取り組みであり、今後も多くの知見が得られることが期待される。それについては口頭発表において紹介したい。

新しい論文に適用可能な指標については、引き続き検討が必要である。NR-PCP については、出版年が長期にわたる論文集合に一元的に適用可能であるか、検証する必要がある。disruption については、全般的に低い値となったが、Research Front が共引用に基づいて作成されていることも一因と考えられる。共引用・書誌結合ネットワーク形状と指標値の関連性や適正な閾値についての検討が必要である。

参考文献

Yamashita, Y. (2019). Identification of technologically relevant papers based on their references. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *Proceedings of 17th International Conference on Scientometrics & Informetrics (ISSI2019)* (pp.2145-2156). Rome: Edizioni Efestio.

Wu, L., Wang, D. and Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, 566, 378-382.