

グラフマイニングに基づく学際性を考慮した学術貢献度評価指数 CRA-index

大槻 明[†] 川村雅義[‡]
Akira Otsuki Masayoshi Kawamura

1. はじめに

2021年2月現在においても COVID-19 が全世界で猛威を振るっている。BBC News[1]によると、世界中でワクチン接種が始まっており状況の改善も見られるが、COVID-19の根本的な対策は未だ確率していない。

NHK[2]は、この知見導出のアプローチとして全論文解説を行っており、特に影響力が大きい世界トップクラスの研究者を明らかにした。しかし、このアプローチでは、論文の引用・被引用数や単語頻度から世界トップクラスの研究者を抽出するシンプルなモデルである。他方で、BOYACKら[3]は、共引用分析、書誌結合分析、直接引用分析の比較検証を通じて、リサーチフロントを最も正確に表す手法について明らかにしているが、これは研究領域(クラスター)を抽出するというマクロな観点の手法であり、重要な科学者や論文を抽出するというミクロな観点の手法ではない。このミクロな手法の一つに、科学的貢献度を評価する指数があり、その代表的なものに h-index などがあるが、しかし、Otsuki[4]らが指摘しているように、共同研究者が多い科学者においては、お互いの論文で引用し合うことで h-index スコアを上げることができてしまうという問題点がある。そこで、Otsukiらは共著者に含まれる科学者同士が引用しあっている状況の特徴量として算出し、その特徴量を元に h-index から減算することで求められる h^{MC}-index を提案した。さらに、COVID-19 の根本的な対策を検討するためには、医学系の知見だけにとどまらず、自然科学、社会科学及び人文科学という学際的な見地からも COVID-19 の対策に資する知見を明らかにすることは極めて意義があると考えられる。

ゆえに本研究では、COVID-19 に関する重要な論文を抽出するための一つのアプローチとして、Otsukiらが提案した h^{MC}-index に学際性の観点から重み付けすることで算出される新たな科学的貢献度指数である CRA-index (to Consider Research Area - index)について提案する。

2. 提案アプローチ

2.1 論文データベース

東京大学医学図書館では、医学に関する定番データベース[5]が公開されている。このうち、International Journal を対象とし、つまり国内論文誌のみを対象とするデータベースは除外したうえで、筆者がアクセス可能で、かつ引用・被引用情報が取得できるデータベースを表 1 にまとめる。

表 1. 医学系論文データベース

学問分野										
医学	感染学	生物学	薬学	看護学	衛生学	生命科学	物質科学	自然科学	社会科学	人文科学

Web of Science ¹									○	○	○
PubMed ²	○		○	○	○	○					
SciFinder ³									○		
Scopus	○									○	○
Google scholar	○	○	○	○	○	○	○	○	○	○	○

Scopus は、2021年2月現在において「covid-19」のキーワードで検索した結果、5,712 の論文しかヒットしなかったため対象外とした。

PubMed は、「covid-19」のキーワードで 104,113 件の論文がヒットしたが、検索結果が 1 万を超える場合は、PubMed 側が設けている制限により最初の 1 万件しか取得することができないため対象外とした。

Google scholar は、Howlandら[6]が指摘しているように、検索する範囲とその結果の正確さに対して不信感があるため対象外とした。Howlandらは、図書館が契約するデータベースよりも Google scholar の方が学術的であると結論づけているものの、上述の不正確さについては否定できておらず、Google scholar の検索結果が学術的かどうかという検証についてはより包括的で客観的な研究が必要であると指摘している。

以上から、また第 1 章でも述べたように、COVID-19 の根本的な対策を検討するためには医学系の知見だけにとどまらず、自然科学、社会科学及び人文科学といった学際的な見地からも COVID-19 の対策に資する知見を明らかにすることは極めて意義があると考えられるため、本研究では Web of Science の論文を対象として学術俯瞰分析を行うこととした。

2.2 学術俯瞰グラフネットワークの構築

本研究が提案する、グラフ理論に基づく学術俯瞰ネットワークを構築するまでの流れを下記①～③に示す。

① 論文データの取得

Web of Science で「COVID-19」をキーワードとしてトピック検索した結果(検索日は 2021/2/22, トピック検索の検索対象フィールドは、「タイトル, 抄録, 著者キーワード及び Keywords Plus[®]」である), 81,968 件の論文がヒットしたため、このうち欠損レコードを除いた論文を対象に学術俯瞰分析を行う。

② 学術俯瞰グラフネットワークの構築

取得した論文を元に、直接引用法を用いて学術俯瞰グラフネットワーク(有向グラフ)を構築する。ここで、共引用分析、書誌結合分析、直接引用分析の中からも直接引用法を用いたのかという理由は、第 1 章でも述べた通り、クラスター内コヒーレンスで評価するような研究領域の抽出には書誌結合法が望ましいと報告されているが[3]、図 1 に示すように、共引用は同じ論文 C から引用されている論文 A, B 間にリ

¹ https://ja.wikipedia.org/wiki/Web_of_Science

² <https://ja.wikipedia.org/wiki/MEDLINE#%E5%AE%88%E5%82%99%E7%AF%64%E5%9B%B2>

³ <https://www.jaici.or.jp/SCIFINDER/>

リンクを張る手法であり、書誌結合は同じ論文Cを引用している論文D、E間にリンクを張る手法である。しかし、本研究の目的は重要な研究領域の抽出ではなく、重要な科学者(≒論文)の抽出であるため、共引用分析や書誌結合分析では肝心な論文Cにリンクを張ることができない。ゆえに、本研究においては直接引用法が最も相応しいと考えられる。

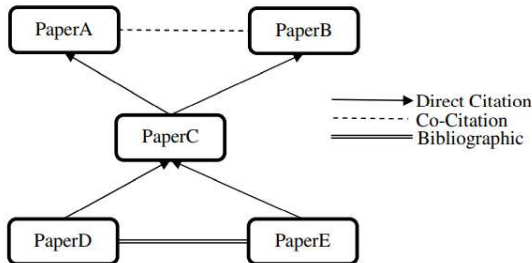


図1. Three analysis techniques in bibliometrics[4]

そして、構築した学術俯瞰グラフネットワークを対象に、CRA-index (2.3 節) を算出する。もしネットワークマップとして可視化する場合はこのスコアをノードのサイズに設定し、ノードの色分けは WOS の SC (研究分野) を元に行う。

③ TF-IDF 分析

さらに、CRA-index のスコアが高い上位 5 の論文を抽出し、これらの論文を対象に TF-IDF 分析 (2.4 節) を行うことで各論文の特徴を明らかにする。

2.3 CRA-index (提案手法)

本稿で提案する CRA-index は次式のように表される。

$$CRA - index = \frac{RA}{\max(RA)} \quad (1)$$

ここで、RA (Research Area) の計算モデルは下記①~④に示す通りである。

- ① ある論文 PaperA を仮定して、PaperA の全被引用論文の研究分野を調査する。
- ② 表 2 に示すように WOS (Web of Science) の研究分野はカテゴリ (大項目) とサブカテゴリ (中項目) で構成されているため、図 2 の左に示すように、PaperA の全被引用論文に対して、サブカテゴリの研究分野まで一致していれば「3」、カテゴリのみの一致であれば「2」、カテゴリすら一致していない場合は「1」とそれぞれ重みづける。

表 2. Web of Science の研究分野 (一例)

Arts & Humanities	Life Sciences & Biomedicine	Physical Sciences	Social Sciences	Technology
Art	Agriculture	Chemistry	Geography	Computer science
History	Cell Biology	Mathematics	Communication	Engineering

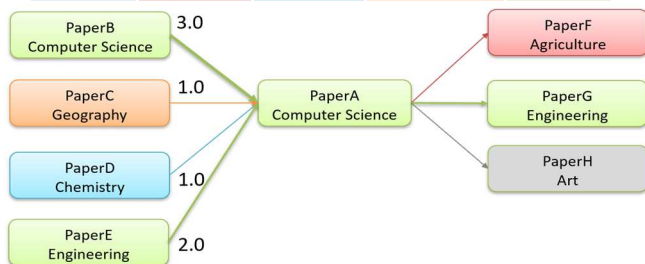


図 2. RA (Research Area) の計算モデル

- ③ RA (Research Area) を計算する。分子は 1 (1 論文) で固定し、分母については②の重みを適用する。つまり、研究分野の大項目が一致していれば 2、研究分野の中項目まで一致していれば 3 として設定して計算し、全被引用論文のスコアを合計することで RA を算出する。図 2 の例では次のように計算される。

$$RA \text{ of PaperA} = 1/3(B) + 1/1(C) + 1/1(D) + 1/2(E) \doteq 2.83 \quad (2)$$

なお、下記の例 1、例 2 に例示するように、図 2 の PaperA~E の各研究分野は、WOS によって複数の研究分野が登録されているが、計算負荷の問題と、著者と一番関係がある研究分野が一番最初に登録されているであろうという仮説から、本研究では一番最初に登録されている研究分野を使用することとした。つまり、下記例 1 では「Chemistry」を、例 2 では「Computer Science」をその論文の研究分野としてそれぞれ使用する。

【例 1】 Chemistry; Engineering; Materials Science; Physics
【例 2】 Computer Science; Environmental Sciences & Ecology; Microbiology; Environmental Sciences & Ecology

- ④ 最後に、式 1 により CRA-index を計算する。CRA-index は、他分野からの引用の場合は 1 本、同じ分野からの引用の場合は 0.5 本及び 0.33...本として計算するアルゴリズムであるため、CRA-index は、学際性の高さを表すと同時に被引用数の多さも表すモデルである。

2.4 TF-IDF 分析

TF (Term Frequency) は、単語が 1 つの文書中にどれだけ出現するかを表すものであり、IDF (Inverted Document Frequency) は、全文書の中でその単語が出現する文書がどれだけあるかを表すものである。IDF は式 3 にて求められる。N は文書数、df は当該単語を含む文書の数、log は 2 を底とした対数を求める関数であることを意味する。

$$IDF = \log \frac{N}{df} \quad (3)$$

そして、これらを掛け合わせたものが Seki[15] の論文でも示されている TF-IDF である。例えば、似たような文書がいくつかあった場合に、TF-IDF を用いることによって、それらの文書の違いが何なのかを大まかに把握することが可能となる。本研究では、CRA-index を用いて抽出した複数の論文の特徴把握を行うために TF-IDF を行う。

3. 評価実験及び考察

3.1 はじめに

科学者の科学的貢献度を評価する指数として、代表的なものに h-index (3.2.1 節) があるが、上述したように共同研究者が多い科学者がお互いの論文を引用し合うことで h-index スコアを上げることができてしまうという問題点があるため、この問題点を改善した h^{MC}-index (3.2.1 節) を比較対象とする。さらに、直接引用法を用いて構築した学術俯瞰グラフネットワークに中心性分析 (3.2 節) を適用することで、本研究と同様に COVID-19 に関する重要な論文を抽出することが可能である。

ゆえに、本評価実験では、 h^{MC} -index 及び媒介中心性分析 (Betweenness, 3.3.4 節) と CRA-index を比較検証することで CRA-index の有効性を確認する。具体的には、下記①～④の流れで比較検証を行う。

【比較検証の準備】

- ① CRA-index, Betweenness 及び h^{MC} -index のスコアの上位 100 論文をそれぞれピックアップし、
- ② ①の上位 100 論文それぞれの被引用論文の研究分野を調査する。

【比較検証】

- ③ 比較検証は 2 通り行うが、まず 1 つ目は、CRA-index, Betweenness 及び h^{MC} -index のスコアと上記②の引用論文の研究分野数との相関関係について比較検証する。
- ④ さらに、②の 100 論文の研究分野数の分散や標準偏差を算出することで比較検証する。

3.2 h -index 及び h^{MC} -index

3.2.1 h -index

J. E. Hirsch[7]が提案した指数であり、例えば表 3 が PaperA の被引用論文リスト (全 12 論文) を表している場合に、この 12 論文の被引用数 (Cited 列) を降順ソートし、PaperA の各被引用論文の被引用数 (Cited 列) が、累積被引用論文数 (No 列) 以上を満たす最下位の累積被引用論文数が h -index となる。

表 3. h -index

No	Cited	Paper Name
1	1,998	Paper1
2	133	Paper2
3	76	Paper3
4	48	Paper4
5	37	Paper5
6	34	Paper6
7	27	Paper7
8	26	Paper8
9	20	Paper9
10	17	Paper10
11	6	Paper11
12	2	Paper12

h -index : 10 →

3.2.2 h^{MC} -index

Otsuki[4]らが提案した指数であり、共同研究者が多い科学者においては、表 4 に示すようにお互いの論文で引用し合うことで h -index スコアを上げることができてしまうという問題点がある。そこで、Otsuki らは、共著者に含まれる科学者同士が引用しあっている状況の特徴量として算出し、その特徴量を h -index から減算することで求められる h^{MC} -index を提案した (表 5)。

表 4. 共同研究者等で引用しあっているイメージ

		Cited Paper
		Paper1
Paper1	←	Paper2
		Paper3
Paper3	←	Paper4
Paper2, 4	←	Paper5
Paper1, 4	←	Paper6
Paper1, 2, 4	←	Paper7
Paper1, 2	←	Paper8
Paper1, 2, 3	←	Paper9

Paper2	←	Paper10
--------	---	---------

表 5. h^{MC} -index の計算イメージ

Substraction			Cited Paper	
				Paper1
0.50	←	1/2	Paper1	← Paper2
				Paper3
0.50	←	1/2	Paper3	← Paper4
0.33	←	1/3	Paper2, 4	← Paper5
0.33	←	1/3	Paper1, 4	← Paper6
0.25	←	1/4	Paper1, 2, 4	← Paper7
0.33	←	1/3	Paper1, 2	← Paper8
0.25	←	1/4	Paper1, 2, 3	← Paper9
0.50	←	1/2	Paper2	← Paper10
2.99	←	h^{MC}-index		

3.3 中心性分析

3.3.1 はじめに

本研究における中心性分析とは、膨大なグラフネットワークの中から、中心的な役割を担っているノードを抽出する手法のことをいう。中心性分析を用いることで、2.2 節②で構築した COVID-19 関連論文の学術俯瞰グラフネットワークから主要な論文を抽出することができる。なお、次節以降に示すように、「何に重きを置くか」という視点の違いによって様々な中心性分析が存在する。Bonacich[7]によると、社会ネットワーク分析の分野では、次節以降に示す 4 つの中心性分析が代表的な尺度とされているが、本研究では学際性を重視しているため、度数に重きを置く度数中心性やページランク、さらには他のノードと距離を重視する近接中心性ではなく、他の研究分野の論文とのハブの役割を担っている論文を抽出できる媒介中心性を比較対象に用いることとした。

3.3.2 度数中心性 (Degree Centrality)

Proctor ら[8]が考案した手法であり、リンクが集まるノードほど中心性が高いと考えるシンプルな手法である。サンプルネットワーク (図 3) を用いて説明すると、ノード 1, 6, 11 が最も多くのリンクを (それぞれ 6 本ずつ) 持っているため、中心性が高いと評価される。

3.3.3 近接中心性 (Closeness Centrality)

Beauchamp[9]が考案した手法であり、他のノードと距離が近いほど中心性が高いとする手法である。津川[10]によるとノード i の近接中心性は次式で表される。なお、 d_{ij} はノード i から j への最短経路長を表す。

$$C_i(v) = \frac{1}{\sum_j d_{ij}} \tag{4}$$

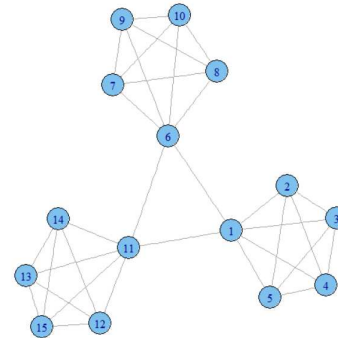


図 3. サンプルネットワーク

3.3.4 媒介中心性 (Betweenness Centrality)

Freeman[11]が 1977 に提案した手法であり、ハブとなっているノードほど中心性が高いと評価するモデルである。林ら[12]によると、 σ_{st} で s, t 間の最短経路の個数を表し、 $\sigma_{st}(v)$ で s, t 間の最短経路のうち v を通る経路の個数を表す。ただし、 $s = v$ または $t = v$ のとき $\sigma_{st}(v) = 0$ とする。このとき、グラフ G 上の頂点 v の媒介中心性は次式のように表される。

$$C_B(v) = \sum_{v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (5)$$

3.3.5 ページランク中心性

Brin ら[13]が考案した手法であり、重要なノードから多くのリンクを受け取るノードほど中心性が高いと考える手法である。ここで、重要なノードとは多くのインリンクを持つノードのことである。篠田ら[14]によると、ノード i の PageRank C_i^P は次式により求められる。

$$C_i^P = \sum_{j \in \{1..N | a_{ij}=1\}} \frac{(1-\alpha)C_j^P}{k_j} \quad (6)$$

a_{ij} は、ノード i とノード j がリンクで接続されているかどうかを示す値、 α はリンクを取らない訪問を想定した定数である。

3.4 関連手法との比較検証の結果

3.4.1 比較検証の結果

表 6 に、CRA-index のスコア上位 100 件 (途中は省略) の①論文タイトル (title)、②CRA-index のスコア、③研究分野数 (No of cited areas) をそれぞれ示す。表 7~8 には同様に Betweenness 及び h^{MC} -index の上位 100 論文のデータをそれぞれ示す。また、図 4~6 には、表 6~8 の③研究分野数 (No of cited areas) の Bar chart を、図 7~9 には表 6~8 の②と③の Scatter diagram をそれぞれ示す。

図 7~9 の相関係数 (R^2) の比較から、CRA-index の②と③には強い正の相関関係が認められたのに対し、Betweenness 及び h^{MC} -index では殆ど相関関係がみられなかった。これは図 4~6 に示す通り、CRA-index は CRA-index のスコアが高くなるにつれ被引用論文の研究分野数が多くなっているのに対し、Betweenness 及び h^{MC} -index は、スコアに関係なく被引用論文の研究分野数は横軸に散在していたためであると考えられる。

表 6. CRA-index のスコア上位 100 論文

	①title	②CRA-index	③No of cited_areas
1	Clinical Characteristics of Coronavirus Disease 2019 in China	1.000	6340
2	Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study	0.864	5504
	(省略)		
99	Cardiovascular comorbidity and its impact on patients with COVID-19	0.081	492
100	Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing	0.080	495

表 7. Betweenness のスコア上位 100 論文

	①title	②Betweenness	③No of cited_areas
1	COVID-19: Some clinical questions after the first 4 months	75020180.393	100
2	Uncertainty about the Efficacy of Remdesivir on COVID-19	50024914.529	36
	(省略)		
99	Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications	2695626.611	83
100	A proposal for staging COVID-19 coagulopathy	2685823.361	16

表 8. h^{MC} -index のスコア上位 100 論文

	①title	② h^{MC} -index	③No of cited_areas
1	Mask use during COVID-19: A risk adjusted strategy	1.000	12
2	Multiple-centre clinical evaluation of an ultrafast single-tube assay for SARS-CoV-2 RNA	1.000	4
	(省略)		
99	Covid-19: Portugal closes all medical schools after 31 cases confirmed in the country	0.850	3
100	Covid-19: 90% of cases will hit NHS over nine week period, chief medical officer warns	0.850	3

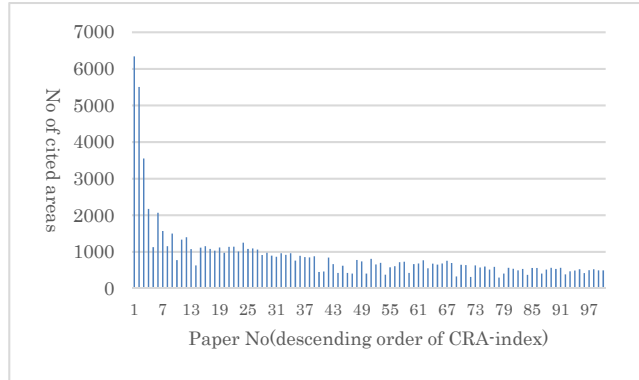


図 4. Bar chart of CRA-index

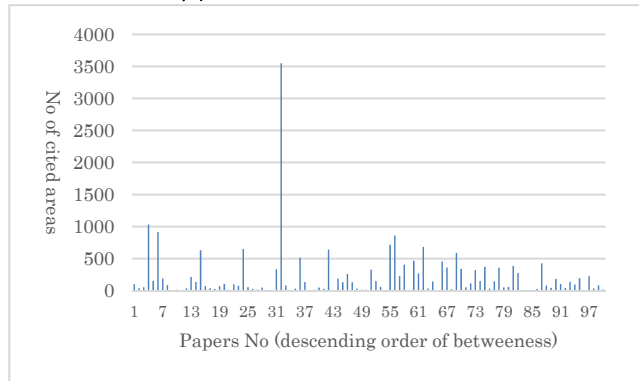


図 5. Bar chart of Betweenness

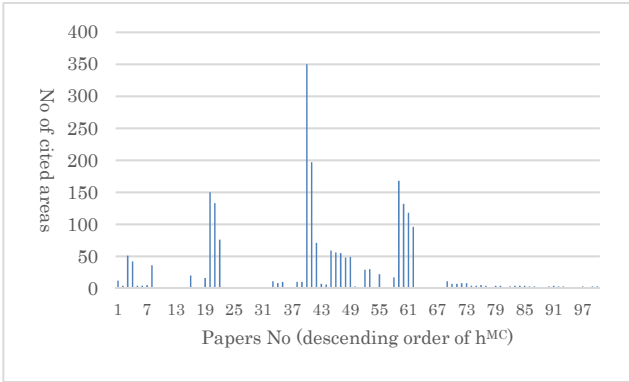


図 6. Bar chart of h^{MC} -index

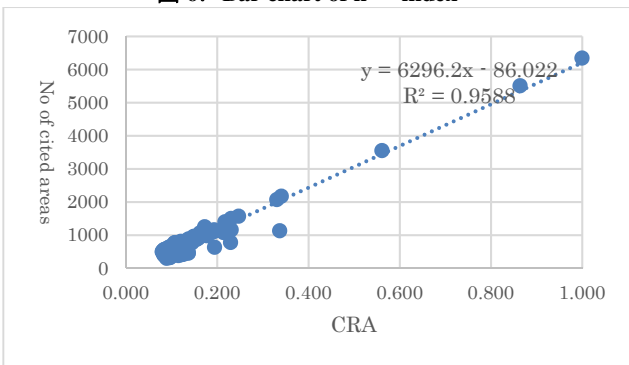


図 7. Scatter diagram of CRA

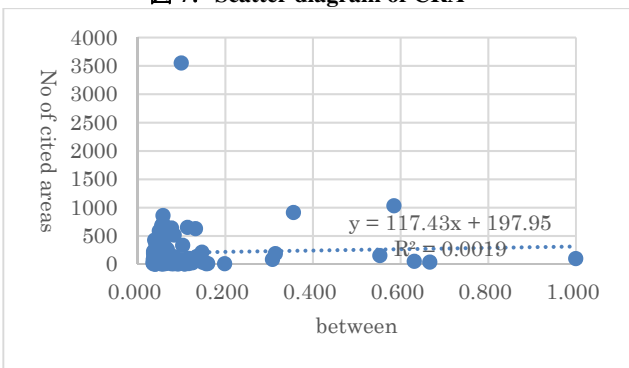


図 8. Scatter diagram of Betweenness

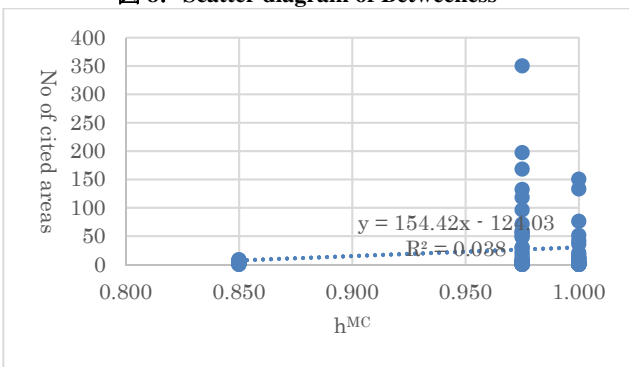


図 9. Scatter diagram of h^{MC} -index

3.5 比較検証を受けての考察

本節では、前節の比較検証を受けての考察を行う。前節の検証結果の原因として考えられることの一つは、CRA-index のスコアに影響するのはレベル1の被引用論文数のみであるが、Betweenness は、レベル1の被引用論文数だけで

なく、レベル2 (レベル1の被引用論文を引用している被引用論文)以降の被引用論文数もスコアに影響するためであると考えられる。さらに、 h^{MC} -index の元は h -index であるため、「その著者の全論文の数」がスコア算出の基準になるために、つまり、論文の被引用数は直接スコアとは関係が無い前節のような結果になったのだと考えられる。

本研究では、直接引用法を用いて構築したグラフネットワークを対象とした指標の作成を目指しているため、レベル1の被引用論文を重視して評価する指標が望ましいと考えられる。さらに、被引用論文の研究分野の多様性を評価することが重要であるため、前節の結果は本研究のコンセプトに対して有効であると考えられる。

3.6 TF-IDF 分析による論文解釈

CRA-index スコアの TOP5 の論文リストを表9に示す。本節では TF-IDF 分析を用いてこれら5論文の簡単な特徴解釈を行う。今回は名詞に限定して TF-IDF 分析を行う。さらに、Covid-19 や単位以外の単語に主眼を置いて分析を行いたいため、分析結果から Covid-19 (Covid や 19 単体も) や単位に関する単語を除外することとした。なお、本研究のコンセプトは、前節までで考察を行った CRA-index を提案することであるため、CRA-index で抽出した論文の有効性検証は今後の課題としたい。

表 9. CRA-index TOP5 の論文リスト

No.	CRA-index	Title
1	1.000	Clinical Characteristics of Coronavirus Disease 2019 in China, 2020/2/28.
2	0.864	Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, 2020/3/28.
3	0.561	Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention, 2020/2/24.
4	0.340	COVID-19: consider cytokine storm syndromes and immunosuppression, 2020/3/28.
5	0.337	Virological assessment of hospitalized patients with COVID-2019, 2020/4/1.

表10に、5論文の TF-IDF スコア top5 の名詞を示す。なお、表10の「No.n」は表9の「No.n」に対応しており、表10は No.1 の TF-IDF スコアで降順ソートして表示している。

表 10. 5論文の TF-IDF スコア top5 の名詞リスト

Noun	No.1	No.2	No.3	No.4	No.5
Patients 患者	61.0	63.0	5.0	6.0	41.0
Nonsevere	39.9	0.0	0.0	0.0	0.0
Disease 疾患	31.0	22.0	5.0	2.0	6.0
Composite 複合	26.6	0.0	0.0	0.0	0.0
Radiologic 放射 性物質	26.6	0.0	0.0	0.0	0.0
Survivors 生存者	0.0	92.9	0.0	4.6	0.0
Cardiac 心臓	0.0	66.4	0.0	0.0	0.0
SOFA	0.0	56.5	0.0	0.0	0.0
Score スコア	0.0	49.9	0.0	0.0	0.0
Shedding 排出	0.0	46.4	0.0	0.0	18.6
CFR	0.0	0.0	33.2	0.0	0.0
Outbreak	6.9	6.9	26.1	0.0	0.0
Deaths 死亡者数	1.7	1.7	22.6	0.0	0.0
Response	0.0	0.0	20.9	0.0	7.0

HLH 血球貪食性リンパ組織球症	0.0	0.0	0.0	16.6	0.0
Hyperinflammation	0.0	0.0	0.0	16.6	0.0
DFM 組織名	0.0	0.0	0.0	13.3	0.0
Immunosuppression 免疫抑制	0.0	0.0	0.0	13.3	0.0
Cytokine	0.0	2.3	0.0	11.6	0.0
Samples	0.0	6.9	3.5	0.0	86.8
copies	0.0	0.0	0.0	0.0	59.8
SARS	21.2	31.7	17.2	0.0	51.6
RNA	1.7	24.3	0.0	0.0	50.4
Replication 複製	0.0	4.6	0.0	0.0	48.8

次に、表10の単語を元にNo.1~5の元論文を調査して特徴把握を行った結果について述べる。下線が引かれている単語は5論文のTF-IDFスコアtop5の名詞をそれぞれ表す。

まず、No.1は2019/12~2020/1/に中国本土の30の州、自治区、および自治体の552の病院の検査で確認されたCovid-19の患者(1099人)に関する「臨床的な特徴」に関して調査した論文であった。具体的には、患者の年齢の中央値は47歳であり、患者の41.9%が女性で、主要な複合エンドポイント¹⁾は、ICUへの入院(5.0%)、人工呼吸器を付けた人(2.3%)、死亡した人(1.4%)を含む67人の患者(6.1%)で発生したと報告されている。

No.2は、中国武漢における「COVID-19の成人入院患者の臨床経過と死亡の危険因子について特定」した論文であり、具体的には次のように述べられている。

- 入院時のSOFA(重要臓器の障害度を数値化した指数スコア)が高いほど、院内死亡のオッズが高くなる。
- 血中IL-6、高感度心臓トロポニンI、乳酸デヒドロゲナーゼ及びリンパ球減少症のレベルの上昇は、重度のCOVID-19疾患で死亡した人の半数以上で見られた。
- 1 μ g/mLを超えるd-ダイマー²⁾がCOVID-19の致命的な結果と関連していることを発見した。

No.3は、中国疾病管理予防センターが発表した、中国本土におけるCOVID-19の症例シリーズ(2020/2/11までの72314症例)を要約し、COVID-19の流行に関する新たな発見と教訓について議論した論文であった。具体的に、患者の殆どは30~79歳(87%)で、多くの症例は湖北省で診断され(75%)、武漢関連の暴露(86%;武漢の居住者または訪問者、または武漢の居住者または訪問者との濃厚接触者)であったと報告されている。また、全体の致死率(CFR)は2.3%であり(44672人の確定患者のうち1023人が死亡)。9歳以下の死亡例はなく70歳から79歳の患者のCFRは8.0%、80歳以上の患者のCFRは14.8%であった。また、SARSやMERSとのアウトブレイク比較を通じて今後の展望について論じていた。

No.4は、重度のCOVID-19患者がサイトカインストーム症候群³⁾を持っている可能性があることを示唆した論文であった。平野[17]の論文でも同様の指摘がなされており、COVID-19に見られる急性呼吸促進症候群(ARDS)はサイトカインストームにより生ずると考えられ、CAR-T治療における致命的な副作用であるサイトカインストームに対して抗IL-6受容体抗体トシリズマブは有効であると述べられている。

最後に、No.5はCOVID-19とSARSの比較検証を通じてCOVID-19の特徴や対策について論じている論文であった。例えば、香港のSARS患者のRT-PCRで陽性と判定された鼻または鼻咽頭スワブ⁴⁾のサンプル98件のうち38件だけが陽性であったのに対し、COVID-19では、調査対象の患者のうち、1人を除いてすべての症例において、症状がまだ軽度であるか、前駆期(流行しているパンデミック疾患の一般的な認識が得られた後、殆どの患者が発症する期間)の段階で単純な鼻咽頭スワブで十分な感度が得られたと報告している。また、喉への向性拡張の可能性を示唆する最も興味深い仮説の1つとして、SARSには存在しないSARS-CoV-2スパイクタンパク質のS1-S2接合部に多塩基性プリン型切断部位が存在していたと報告している。SARS-CoVのS1-S2領域に多塩基性切断部位を挿入すると、中程度ではあるが識別可能な融合獲得活性がもたらされることが以前に示されている。つまり、高濃度のウイルスRNAとサブゲノムmRNAを含む便中に時折検出される細胞の組み合わせは消化管での活発な複製を示している。ゆえに、本論文では、ウイルスの拡散を封じ込めるための対策として、媒介物ではなく飛沫に基づく感染対策を目指すべきだと述べている。

4. おわりに

本研究では、COVID-19対策に資する知見導出を目的として、引用・被引用情報を元にした学術俯瞰ネットワークマップを考案し、このネットワークマップを対象に学術横断的な見地からハブとなっている論文ノードを抽出する指標(CRA-index)を提案した。今後の課題としては、本アプローチで抽出した論文の有用性について専門家の評価を通じて確認していきたい。

参考文献

- [1] BBC News: 「新型コロナウイルスのワクチン 世界の接種状況は」、<https://www.bbc.com/japanese/features-and-analysis-56051716> (2021/2/16閲覧済み)。
- [2] NHK: 「新型コロナ 全論文解説~AIで迫るいま知りたいこと~」、https://www3.nhk.or.jp/news/special/coronavirus/nhksspecial_1108/?fbclid=IwAR1LSvFbKXakna6au3opBUEPqWFFrUitkKjE0z7S5U5_Cb8HX5BFe8 (2021/2/16閲覧済み)。
- [3] BOYACK Kevin W, KLAUVANS Richard: Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately?, *J Am Soc Inf Sci Technol*, Vol.61, No.12, pp.2389-2404, 2010.
- [4] Akira Otsuki, Masayoshi Kawamura: Research Evaluation Contribution Index "h^{MC}-Index", *Global Advanced Research Journal of Engineering, Technology and Innovation*, 3(9), pp.217-238, 2014.
- [5] 東京大学医学図書館定番データベース: <https://www.lib.m.u-tokyo.ac.jp/db/kaisetsu1.html> (2021/2/16閲覧済み)。
- [6] Howland, Jared L.; Wright, Thomas C.; Boughan, Rebecca A.; Roberts, Braiam C. How Scholarly Is Google Scholar?, *A Comparison to Library Databases*. College & Research Libraries, vol. 70, No. 3, p. pp.227-234, 2009.
- [7] J. E. Hirsch: An index to quantify an individual's scientific research output, *Proc Natl Acad Sci U S A*, 2005 Nov 15; 102(46): 16569-16572, 2005.
- [8] Bonacich, P.: Factoring and weighting approaches to status scores and clique identification, *The Journal of Mathematical Sociology*, Volume 2, pp.113-120, 1972.
- [9] Proctor CH, Loomis CP: Analysis of sociometric data. *Res Methods Soc Relat* 2:561-85, 1951.
- [10] M. A. Beauchamp. An improved index of centrality. *Behavioral Science*, 10:161-163, 1965.
- [11] 津川翔: 情報メディア実験 M28: ソーシャルネットワーク分析, <http://www.mibel.cs.tsukuba.ac.jp/~s-tugawa/jikken/centrality.pdf> (2021/3/20閲覧済み)。
- [12] Freeman, Linton: "A set of measures of centrality based on betweenness". *Sociometry*. 40 (1): 35-41, 1977.
- [13] 林孝紀, 秋葉拓哉, 吉田悠一: 動的なネットワークにおける媒介中心性の高速計算手法, *DEIM Forum 2015 E3-4*.
- [14] Sergey Brin and Lawrence Page.: The Anatomy of a Large-scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, pp.107-117, 1998.
- [15] 篠田孝祐, 松尾豊, 中島秀之: 複数の中心性尺度によるネットワーク生成モデル, *知能と情報/20* 巻, 3号, pp.410-422, 2008.
- [16] Seki Yohei: Sentence Extraction by tf/idf and Position Weighting from Newspaper Articles, *Proceedings of the Third NTCIR Workshop 2002*.
- [17] 平野俊夫: 新型コロナウイルス感染症(COVID-19)はサイトカインストーム症候群である, <https://www.covid19-jma-medical-expert-meeting.jp/topic/4565> (2021/6/7閲覧済み)。

¹ <https://e-kangosyoku.jimdofree.com/2015/11/10/%E6%96%B0%E3%81%97%E3%81%84%E3%83%96%E3%83%AD%E3%82%B0%E3%82%92%E6%9B%B8%E3%81%8F/>
² <https://www.acute-care.jp/ja-jp/learning/glossary/immunoassay/ddimer>

³ <https://www.cas.org/ja/resource/blog/covid-19-cytokine-storms>

⁴ <https://www.bdj.co.jp/pdf/48-kentai.pdf>