

「共同利用機関における 意思決定のための指標の活用」

2019/05/22

総合地球環境学研究所

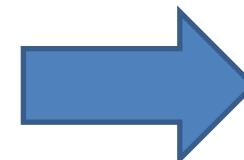
地球研・統数研共同研究キックオフシンポジウム
「研究力評価にむけた様々な指標つくり：人文学指標・学際研究指標・超学際研究指標・共同利用/共同研究指標

濱田ひろか 武井 美緒 本多啓介
統計数理研究所

統数研の異分野融合指標の開発

プロジェクトの計画

異分野融合の進展や効果を公正かつ適切に評価するための指標について、統計数理を活用した研究を平成28年度から実施し、平成31年度以降、大学等のIR機能強化に資するため研究の成果、特に得られた新指標を公開するとともに、公募型共同利用・共同研究の重点テーマの設定に活用する。



- ①指標を作る
- ②指標を使う
- ③指標を使ってもらう

公募型共同利用重点型研究「研究IR」

「学術文献データ分析の新たな統計科学的アプローチ」
「IRのための学術文献データ分析と統計的モデル研究の深化」
H.28年度～ 累計で50件、30以上の機関・大学から参画

計算資源（クラウド、グラフDB）の利用支援、開発整備
国際研究集会の企画、開催

①指標を作る

多様性指標の開発

②指標使う

共同利用重点テーマ選定

③指標を使ってもらう

地球研とのMOU

研究力強化を支援する

意思決定のための指標の活用

研究の状態を可視化する指標やツールを活用して機関の意思決定に資するような情報を探索する手法

指標を活用した意思決定支援

研究力 = 予算 × 個々のPJの構成

共同利用機関のミッション
「異分野融合」

指標
「いかに異分野型の研究をしているかを見る指標？」

指標：見たいものをよく説明するものさし

モデルを最大化する

なければ独自に作る

多様性指標

SABRmetrics、セイバーメトリクス

様々な野球統計（指標）を駆使してチームの状況や戦略を分析する手法

セイバーメトリクスを活用した戦略

チームの総合力 = 予算 × 選手の編成

定義（モデル）
「27回アウトを取られるまで終わらない競技」

指標
「出塁率」

指標：見たいものをよく説明するものさし

定義（モデル）を最大化する

モデルを最大化するための組み合わせを見つけるために指標を探して使う

指標の開発

異分野融合の進展や効果を
公正かつ適切に評価するための指標？



学術分野？

定義（モデル）

「論文は自分の研究に関係の強い（近い）論文を引用する」

複合科学（Multidisciplinary）のジャーナル?
共著者の所属学部、学科、横断？国際共著率？

↓ 進展や効果は
論文の属性や中身では測れない

研究の波及 → 「引用」構造

↓ (使い勝手)

いかに異分野に影響を与えたか → 引用の散らばり
→ 引用-被引用関係の学術分野間の相対的な距離

作るべき指標

「論文の引用関係（ネットワーク構造）から関係の近いものはより近く、遠いものはより遠く並べかえたグループを作り（クラスタリング）、その相対的な距離をスコア化したもの」

引用関係のみで分類されたクラスタ = 「潜在的学術分野」

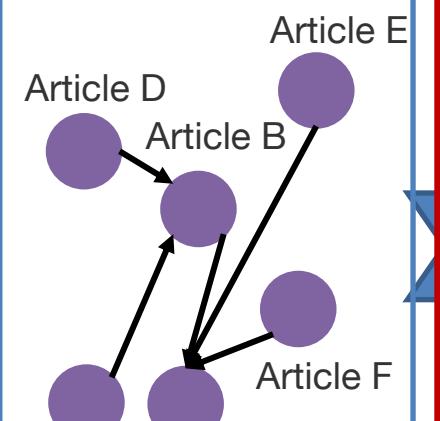
多様性指標の実装



- データ：Web of Science
- クラスタリングの手法：確率的ブロックモデル
- 分野間の偏りの補正：自己相互情報量
- 異分野融合の進展を測る：「多様性指標」

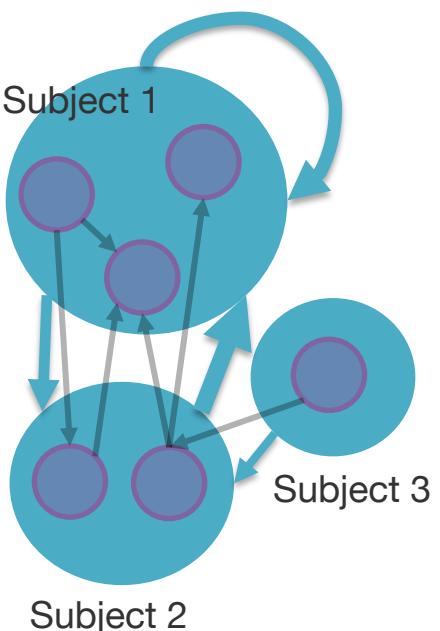
多様性指標開発フロー－概要

分析データ



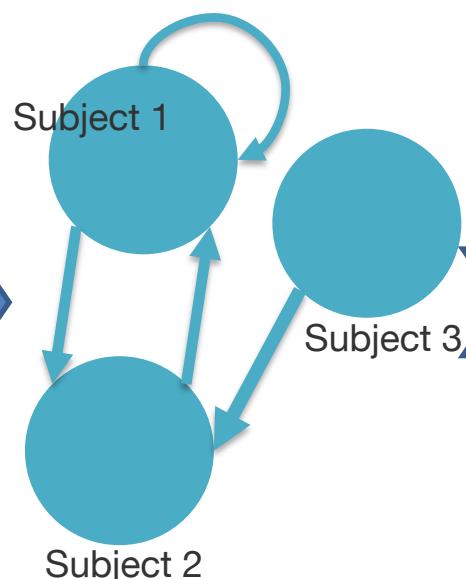
データ集計

using WoS Subject



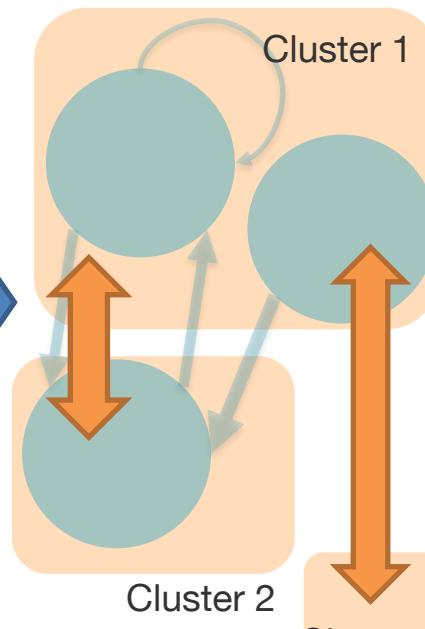
正規化

using PMI
(Pointwise Mutual Information)

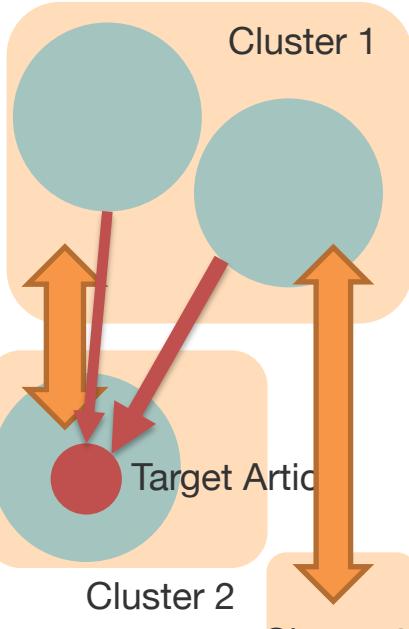


クラスタリング

using SBM
(Stochastic Block Model)



スコアの定義



WoSCC, Neo4j

PGX, Green-Marl

Python, Edward
GPUs

多様性指標開発における課題と採用手法

大規模なデータの処理

4577万のノードをクラスタリングすることになる。
ネットワークはさらに拡大し続けるため継続的な運用が困難。

WoSカテゴリの組合せ

論文数の偏りを正規化

データを集計すると集団ごとに論文数の偏りが発生する。
論文数が大きく異なる集団間のエッジ数を補正したい。

Pointwise
Mutual
Information

[Gerlof Bouma (2009)]

ネットワークのクラスタリング

引用-被引用関係ネットワークを基にクラスタリングすることで
データに基づいた「分野」を定義したい。

Stochastic Block
Model

[Krzysztof Nowicki & Tom A. B Snijders (2001)]

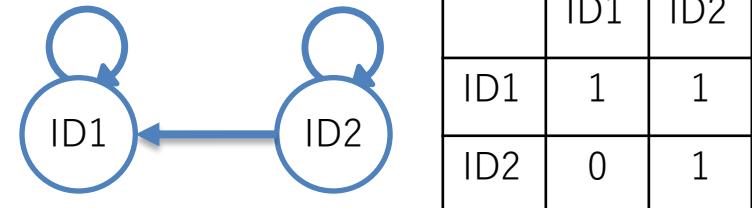
データ集計

大規模なデータの処理

4577万のノードをクラスタリングすることになる。
ネットワークはさらに拡大し続けるため継続的な運用が困難。

WoSカテゴリの組合せ

- WoSカテゴリの組合せをID (Subject ID) とし、Subject IDごとに被引用回数を集計。
- Subject IDは3120個。
- 3120×3120 の接続行列でネットワークを表現する。



	WoSカテゴリ	ESIカテゴリ
Official name	Web of Science Research Areas	Essential Science Indicators Research Areas
Num. of Category	275	22 (23)
Journal to Category	1 対 多	1 対 1

正規化

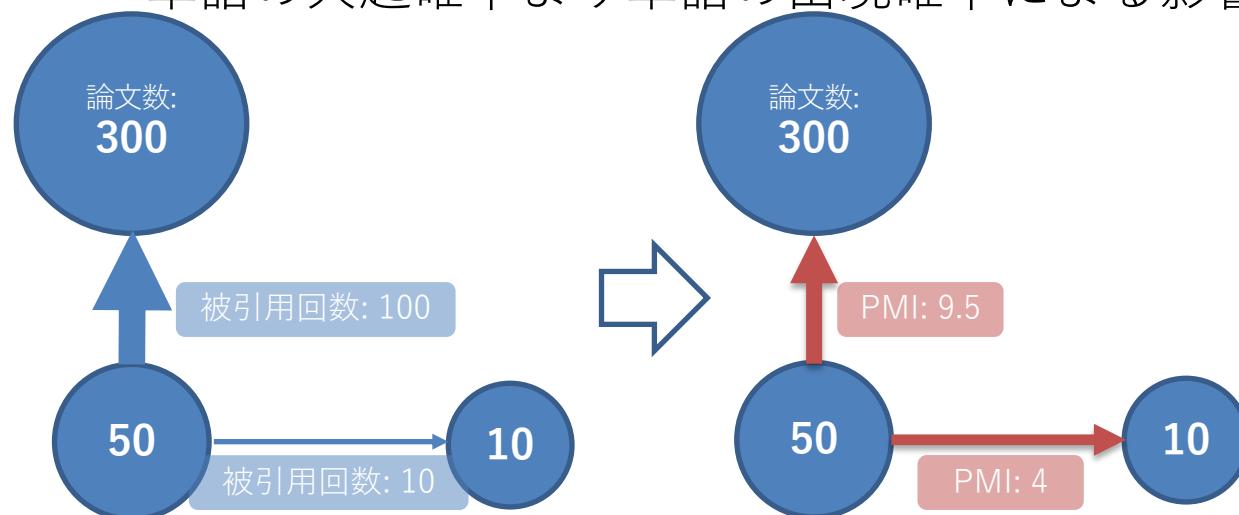
論文数の偏りを正規化

データを集計すると集団ごとに論文数の偏りが発生する。
論文数が大きく異なる集団間のエッジ数を補正したい。

Pointwise
Mutual
Information

Pointwise mutual information (PMI) [Gerlof Bouma (2009)]

- PMIは自然言語処理の分野で単語の共起確率を求めるために比較的よく用いられる手法。
単語の共起確率より単語の出現確率による影響を除外した共起確率を得ることができる。



$$PMI(x \rightarrow y) = \log_2 \left(\frac{C(x \rightarrow y) N}{C(x) C(y)} \right)$$

$PMI(x, y) > 0$	x と y は共起しやすい傾向にある
$PMI(x, y) = 0$	x と y は独立である
$PMI(x, y) < 0$	x と y は共起しにくい傾向にある

正規化結果

$$PMI(x \rightarrow y) = \log_2 \left(\frac{C(x \rightarrow y) N}{C(x) C(y)} \right)$$

$$\mu = \frac{1}{C(x)} \sum C(x \rightarrow y)$$

$$PMI'(x \rightarrow y) = \log_2 \left(\frac{C(x \rightarrow y) N}{C(x) C(y) \mu} \right)$$

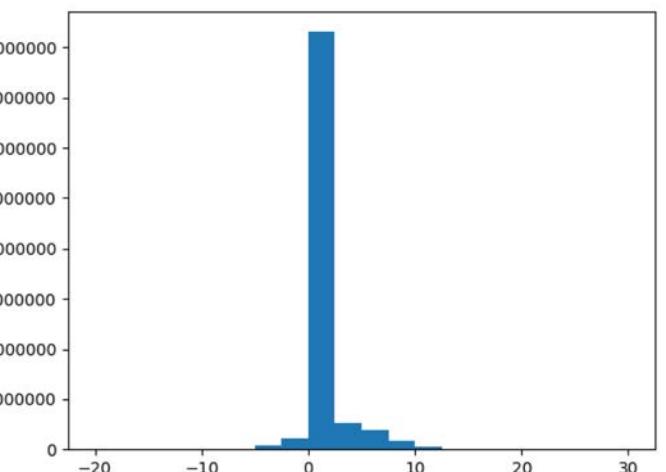
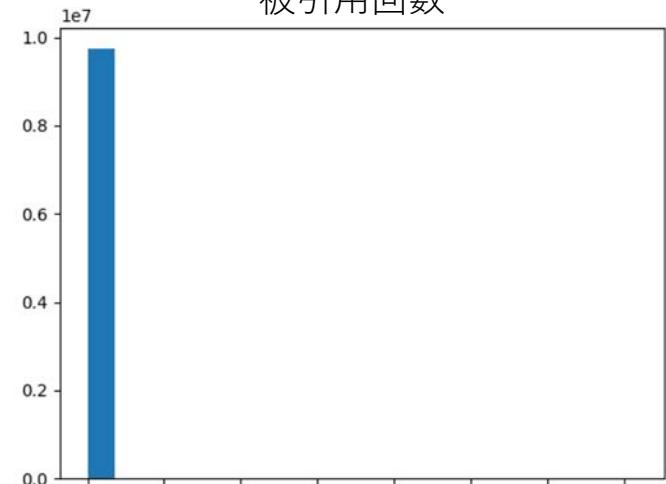
被引用回数の頻度分布

Class interval	Frequency
0 - 358151	9734253
358152 - 716303	86
716304 - 1074455	23
1074456 - 1432607	11
1432608 - 1790759	8
1790760 - 2148911	8
2148912 - 2507063	5
2507064 - 2865215	0
2865216 - 3223367	1
3223368 - 3581519	1
3581520 - 3939671	1
3939672 - 4297823	1
4297824 - 4655975	1
4655976 - 5014127	0
5014128 - 5372279	0
5372280 - 5730431	0
5730432 - 6088583	0
6088584 - 6446735	0
6446736 - 6804887	0
6804888 - 7163039	1

PMIの頻度分布

Class interval	Frequency
-20 - -17.6	0
-17.5 - -15.1	0
-15 - -12.6	0
-12.5 - -10.1	3
-10 - -7.6	455
-7.5 - -5.1	10239
-5 - -2.6	71681
-2.5 - -0.1	211054
0 - 2.4	8312340
2.5 - 4.9	514963
5 - 7.4	387606
7.5 - 9.9	170387
10 - 12.4	45179
12.5 - 14.9	8045
15 - 17.4	1700
17.5 - 19.9	587
20 - 22.4	130
22.5 - 24.9	25
25 - 27.4	6
27.5 - 29.9	0

被引用回数



クラスタリング

ネットワークのクラスタリング

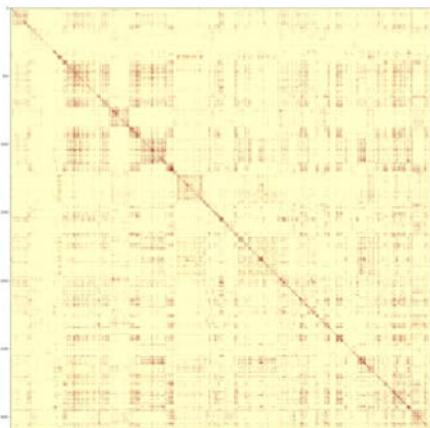
引用-被引用関係ネットワークを基にクラスタリングすることでデータに基づいた「分野」を定義したい。

Stochastic Block Model

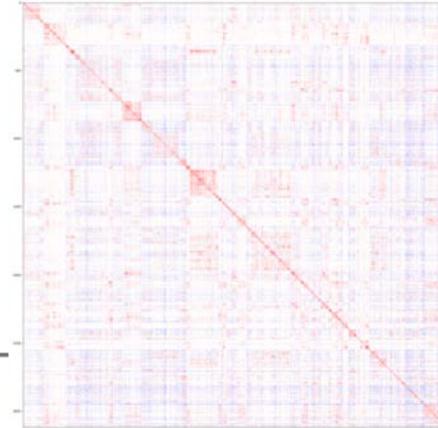
Stochastic Block Model (SBM) [Krzysztof Nowicki & Tom A. B Snijders (2001)]

- ネットワークのクラスタリング手法。
- ノードは必ずいずれか1つのクラスタに所属する、ノードの接続確率は所属するクラスタ依存して決まる、と仮定する。

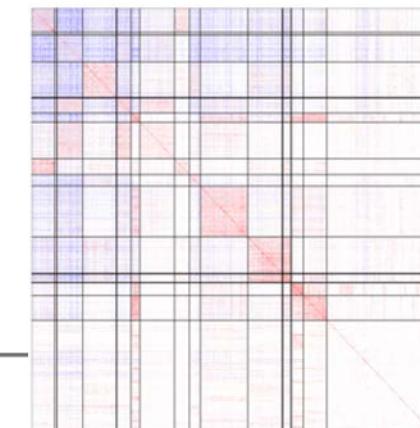
データ集計



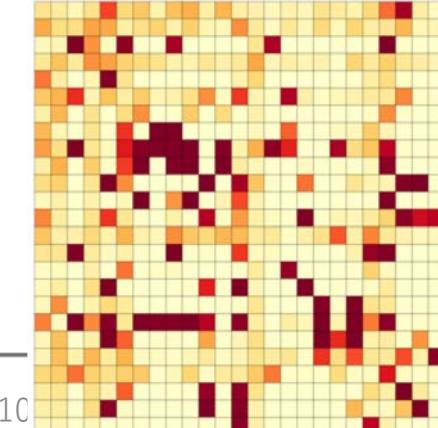
正規化



クラスタリング



クラスタ間の接続確率



C0	MATERIALS SCIENCE	C8	Engineering & Environmental	C16	Economics & Geoscience
	Chemistry, Nanoscience, Materials Science, Polymer Science		Engineering, Materials Science(Ceramics), Environmental Sciences, Water Resources		Business & Economics, Area Studies, Planning & Development, Geography, Cultural Studies
C1	Neurosciences & Popular medical science	C9	Engineering & Cybernetics, Ergonomics	C17	Human's Protein
	Neurosciences, Anesthesiology, Psychiatry, Infectious Diseases, Rheumatology		Engineering, Transportation, Ergonomics, Education, Scientific Disciplines, Cybernetics		Molecular Biology, Immunology, Andrology, Cell Biology, Virology, Cell & Tissue Engineering
C2	Neurosciences & Restricted medical science	C10	COMPUTER SCIENCE	C18	Materials & Geo science
	Neurosciences, Critical Care Medicine, Orthopedics, Pediatrics, Emergency Medicine		Computer Science, Cybernetics, Mathematics, Computer Science(Theory & Methods)		Chemistry, Materials Science(Paper & Wood), Mineralogy, Soil Science, Physics
C3	Ergonomics, Psychology, Health Care	C11	Physics & Astronomy	C19	Mathematics & Statistics & Computer Science
	Psychology, Ergonomics, Business & Economics, Health Care Sciences & Services		Chemistry, Physics, Engineering, Dance, Poetry, Astronomy & Astrophysics, Computer Science		Mathematics, Computational Biology, Statistics & Probability, Medical Informatics, Computer Science
C4	Surgery & Peripheral Vascular Disease	C12	BIOLOGY & BIOCHEMISTRY	C21	Integrative Medicine
	Cell Biology, Peripheral Vascular Disease, Pharmacology, Surgery, Transplantation		Biochemistry, Molecular Biology, Life Sciences & Biomedicine, Biology		Biochemistry & Molecular Biology, Parasitology, Legal Medicine, Integrative & Complementary Medicine
C5	ECONOMICS & BUSINESS	C13	AGRICULTURAL SCIENCES	C22	Software Engineering
	Economics, Business, Management, Hospitality, Environmental Sciences, Public Administration		Agriculture, Agricultural Engineering, Agronomy, Horticulture, Spectroscopy		Computer Science, Engineering, Physics, Mathematical, Acoustics, Software Engineering
C6	PSYCHIATRY/PSYCHOLOGY	C14	human behavior & Psychology	C23	Engineering & Marine
	Psychology, Family Studies, Education & Educational Research		Business, Psychology, Hospitality, Leisure, Sport & Tourism, Criminology & Penology		Engineering, Engineering, Marine, Mathematics, Engineering, Ocean, Robotics
C7	Mathematics & Biochemistry	C15	Literature & Science	C24	Neurosciences & Health
	Mathematical, Computational Biology, Agriculture, Pharmacology, Nanoscience		Engineering, Literature, Chemistry		Neurosciences, Primary Health Care, Sport Sciences, Ophthalmology, Neuroimaging

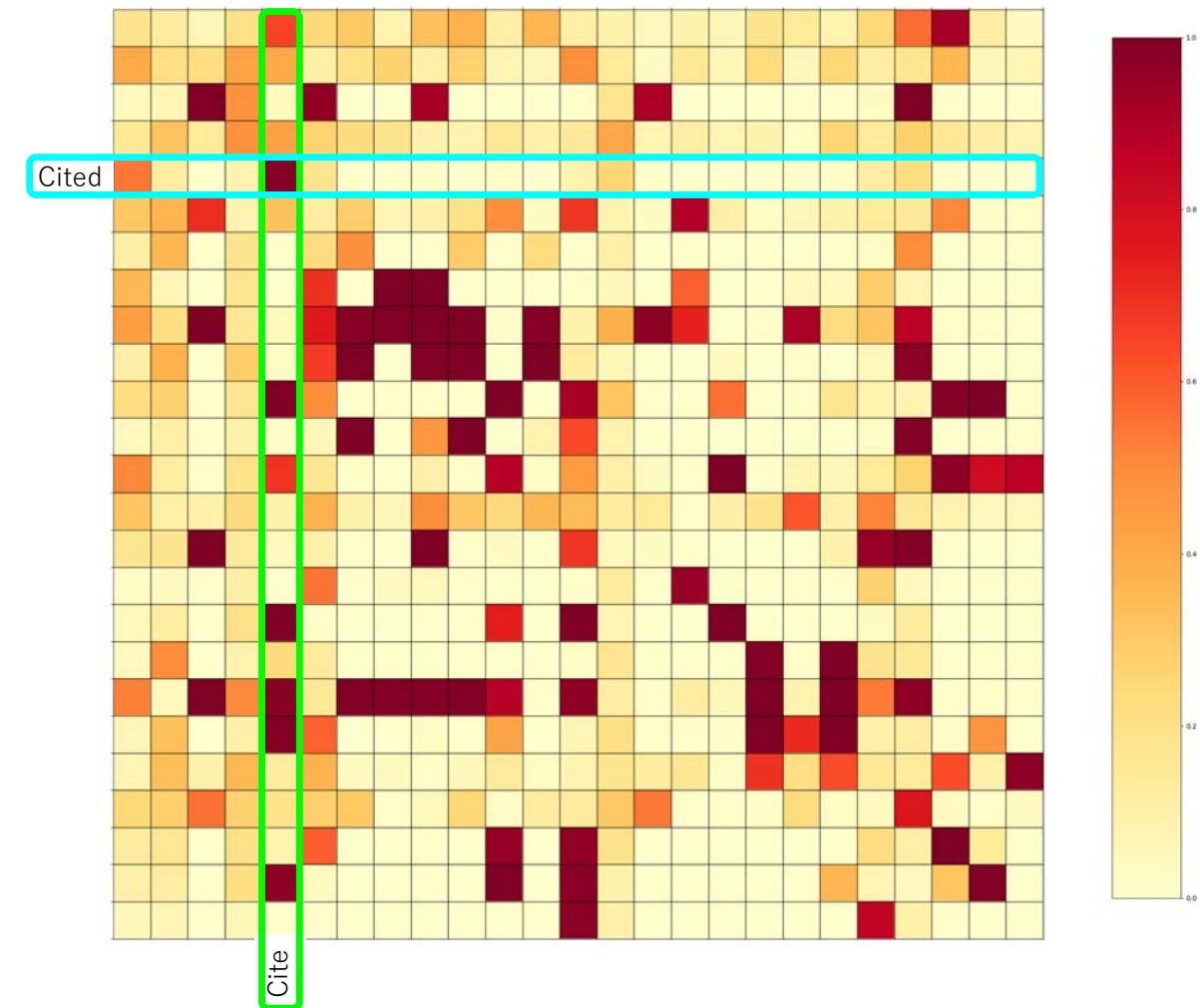
クラスタリング結果：接続確率のヒートマップ

[軸]

- X軸 = 被引用クラスタ
- Y軸 = 引用クラスタ

[色]

- 赤 = 近いクラスタ
- 橙 = やや遠いクラスタ
- 黄 = 遠いクラスタ



多様性指標のスコアリング

論文単位で多様性指標のスコアを計算

- ① 事前に定義した25分野のうち、評価対象論文・引用論文がどの分野かを特定
- ② 計算済みの接続確率を基にスコアを算出

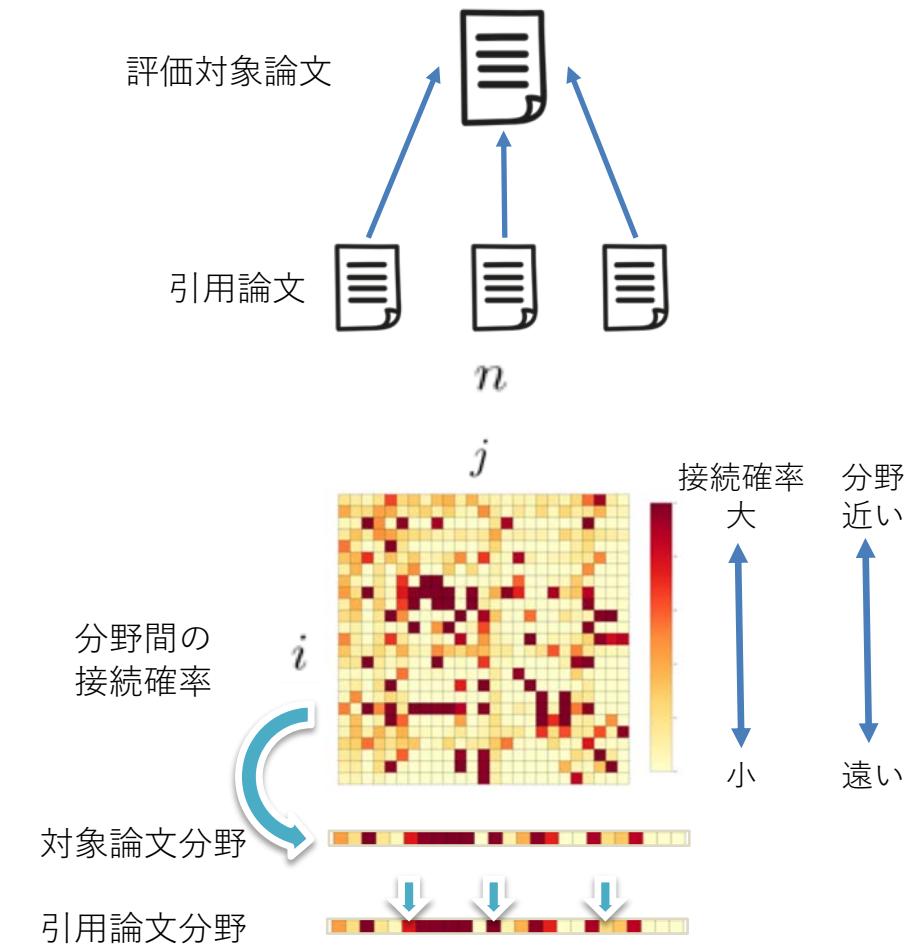
$$\text{REDi} = \frac{100}{n} \sum |1 - A_{(i \rightarrow j)}|$$

n = Cited count

i = Cluster of cited

j = Cluster of cite

$A_{(i \rightarrow j)}$ = Edge probability from i to j



「異分野融合の進展や効果を公正かつ適切に評価するための指標」の開発と実例

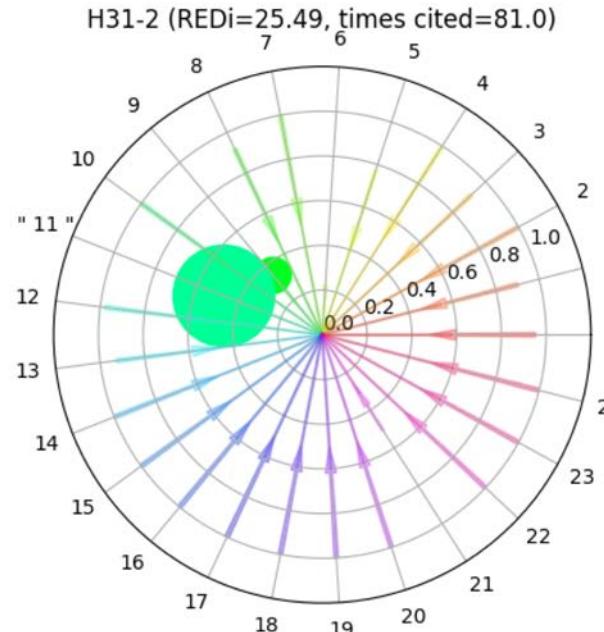
多様性指標

Research Diversity Index

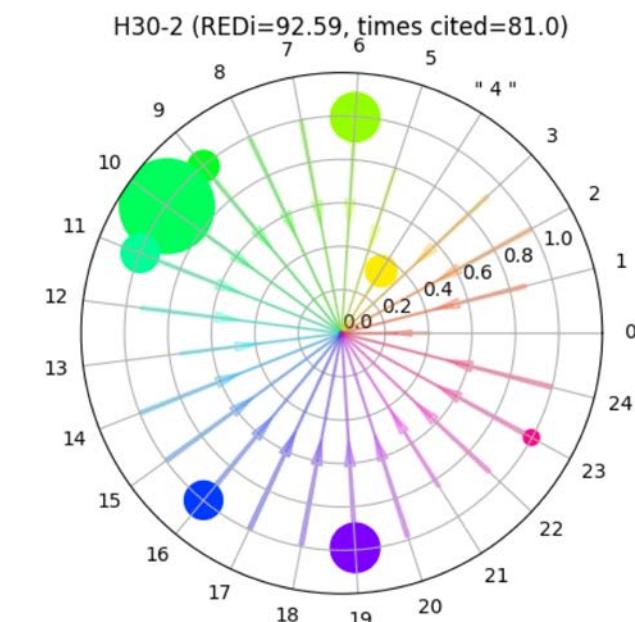
共同利用の成果や異分野融合の進展状況を可視化する方法として、書誌データのネットワーク構造に着目した新たな指標を開発しました。書誌の引用-被引用ネットワーク構造に対して確率的ブロックモデルというクラスタリング手法を適用、この結果を研究活動の波及度合いの関係を基にした「潜在的学術分野」として定義し、各クラスタ（ノード）間の接続確率を距離として、離れた分野から多く引用された論文のスコアをより高くなるように計算したものです。分野間の相対比較ができるように論文数を補正しています。

今回分析に用いたテーマごとの代表的論文に対する多様性指標の計算結果の例

引用数81 多様性指標25.5 *1



引用数81 多様性指標92.6 *2



各軸はクラスタリング結果である潜在的学術分野を表し、色分けされた円は各クラスごとの引用論文数を表しています。中心からの長さは自分の論文が属するクラスタとの距離を表し長い距離の論文（クラスタ）からの引用が多いほど多様性指標は大きくなります。矢印は相手論文からこの論文への引用（cited count）を示しています。

*1) Branch-duct intraductal papillary mucinous neoplasms of the pancreas: to operate or not to operate? GUT 2007

*2) Misunderstandings between experimentalists and observation lists about causal inference JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY 2008

多様性指標

IR機能の強化と異分野融合指標の開発

統数研では平成28年度以降、大学・研究機関における研究戦略への情報提供を目的とする「研究IR (Institutional Research)」の方法論について統計学的見地からの確立を目指し、公募型共同利用の重点テーマを中心に多角的に取り組んできました。特に「従来指標では捉えられない異分野融合研究の進展状況を可視化する方法の開発及び活用」を目標とし、「異分野融合の進展や効果を公正かつ適切に評価するための新指標の研究開発」を行ってきました。

多様性指標 = 多様性 + 引用情報



多様性指標：
Research Diversity
Index (REDi)



多様性
(Diversity)

イノベーションの源泉は、
個々の多様な属性・才能・知識 (≒多様性)

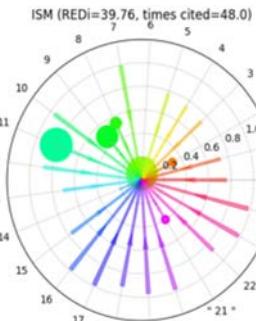


被引用回数

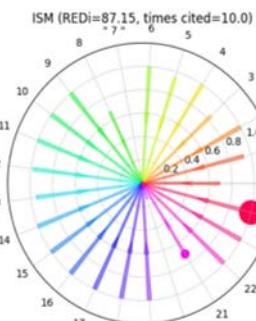
論文の定量的かつ客観的な指標としてよく利用される
被引用回数

- 引用数：「研究の質」を測る代表的な指標
- 異分野度：「ある学術分野」の多方面、多角的な発展を目指し、ボトムアップ型・萌芽型研究を支援するための公募型共同利用事業を運用する共同利用機関では特に重要な観点

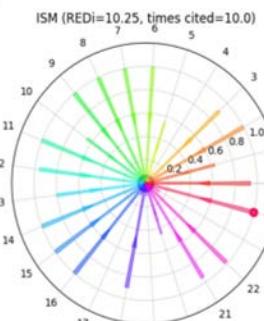
No. 1



No. 2

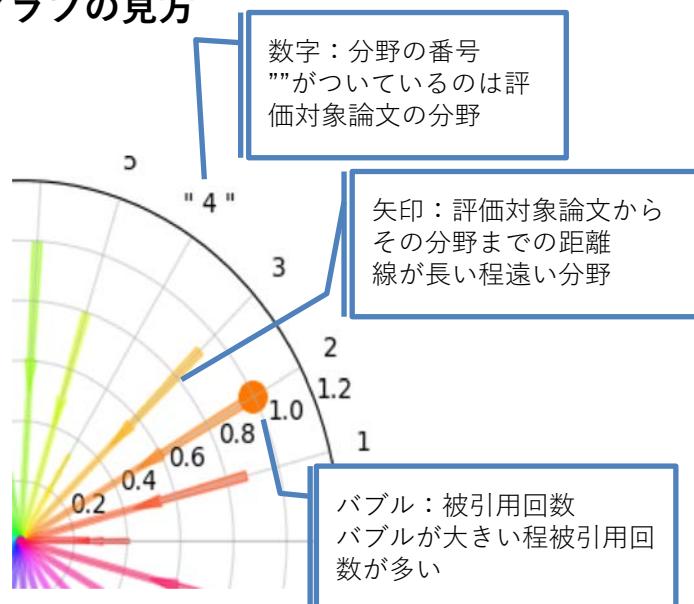


No. 3



3編の論文について多様性指標を計算し、
グラフ化した例

グラフの見方



多様性指標の設計と開発

研究グループは「異分野融合の進展や効果を公正かつ適切に評価するための指標」を開発するにあたりその指標が備えるべき要件を以下のように設定しました。

- 「論文単位で」かつ「書誌情報だけで」算出できること
- 分野間の論文数の偏りを適切に補正できること（分野間比較）
- 中長期的な影響を測定できること

これらの要求を満たすように、以下の要素技術を使って指標を構築しました。

- 確率的ブロックモデル（引用ネットワークのクラスタリング）
- 自己相互情報量（分野間の論文数の偏りの補正）
- 大規模書誌グラフデータベースの整備

多様性指標を活用した意思決定支援（1）

「平成31年度重点テーマ分析に関する報告」

・平成30年9月1日提出

特任研究員 濱田ひろか 特任技術専門員 武井美緒URA 本多啓介

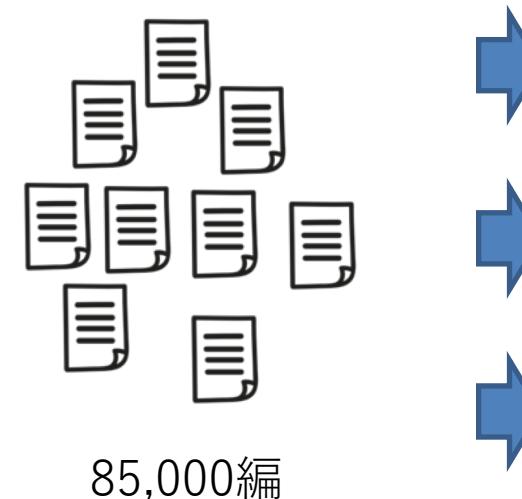
重点型研究次年度テーマ選定にあたり、各共同利用委員から提出された候補7件と現行3件の全10件について、それぞれのテーマと統計科学の両方に関係する論文を直近10年の書誌データから抽出し、その論文群について「多様性指標」を計算した結果をまとめた

分析手法の流れ

1. 学術文献データからトピック抽出
2. テーマとトピックのマッチング
3. 評価対象論文の抽出
4. 多様性指標のスコアを算出
5. 評価レポートの作成

1. 学術文献データからトピック抽出

- トピックモデルを利用して学術文献データからトピック抽出
- 利用論文数（2007-2016年のデータをサンプリング）
：85,000
- トピック数（グループ数）
：2,500
- 単語種類
：116,187（前処理後）
- 単語総数
：7,484,722（前処理後）
- 利用手法
：latent Dirichlet allocation (LDA)



Topic # 1	
school	0.22762163
student	0.1311173
educ	0.04603004
⋮	⋮

Topic # 2500	
court	0.08219621
case	0.05437009
judg	0.03039682
⋮	⋮

トピックモデルとは

- ドキュメントの集合内で発生する抽象的な「トピック」を発見するための統計モデルの一種
- 主なアウトプットとして、以下が得られる
 - ドキュメント毎のトピックの出現確率
 - トピック毎の単語の出現確率

“Arts”	“Budgets”	“Children”	“Education”
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Blei, D. M., Ng, A. Y., Jordan, and M. I., 2003. Latent dirichlet allocation. Journal of Machine Learning Research 3, 357 993-1022.

2. テーマとトピックのマッチング

- 各テーマと学術文献データから抽出したトピックをマッチング
- ① 各テーマからキーワードを抽出
各テーマのキーワード一覧

No	テーマ名	キーワード
	統計関連*	statistics, regression, inference, hypothesis
H30-1	データサイエンス人材育成メソッドの新展開	data science, talent training
H30-2	IR のための学術文献データ分析と統計的モデル研究の深化	scientific literature, Higher Education
H30-3	統計的機械学習の新展開	machine learning, artificial intelligence
H31-1	A	policy planning
H31-2	B	sports
H31-3	C	HPC
H31-4	D	medical issues, health care issues
H31-5	E	Environmental Assessment
H31-6	F	big data in marketing, marketing
H31-7	G	emerging IT environment, emerging computing environment

* 統数研の共同利用のテーマのため、統計に関連する内容でかつ各テーマに関連する論文を抽出する必要があり、統計に関連するトピックを別途定義し、キーワードを作成

2. テーマとトピックのマッチング

- ② 抽出したキーワードから各テーマがどのトピックに対応するかを、トピックモデルの結果を基に決定

例

テーマ名	キーワード
統計的機械学習の新展開	machine learning



Topic #
1

machine + learning

Topic # 1	
learn	0.45678
knowledg	0.02683
experi	0.02320
:	:
machin	0.00363
foster	0.00345
computer-bas	0.00345
:	:

$$0.00363 + 0.45678 = 0.46041$$

Topic # 2	
machin	0.11136
cut	0.05679
tool	0.05269
:	:
learn	0.00616
steel	0.00650
depth	0.00650
:	:

$$0.11136 + 0.00616 = 0.11752$$

基本的に値が大きいものを
そのキーワードのトピック
として採用

3. 評価対象論文の抽出

- 各テーマと統計関連トピック内でマッチ度が高い論文50編を評価対象論文として抽出

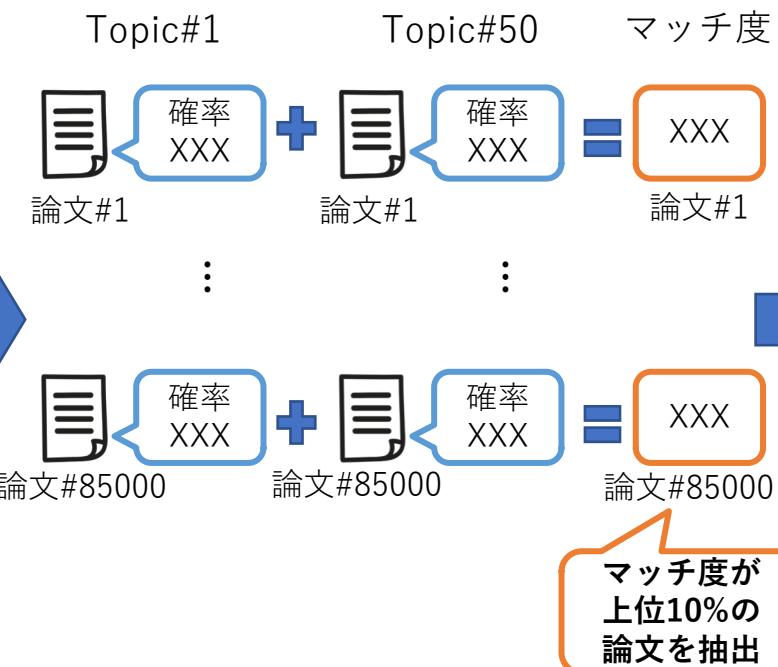
テーマキーワードのトピック番号

テーマ名	キーワード	Topic #
統計的機械学習の新展開	machine learning	1
	artificial intelligence	50

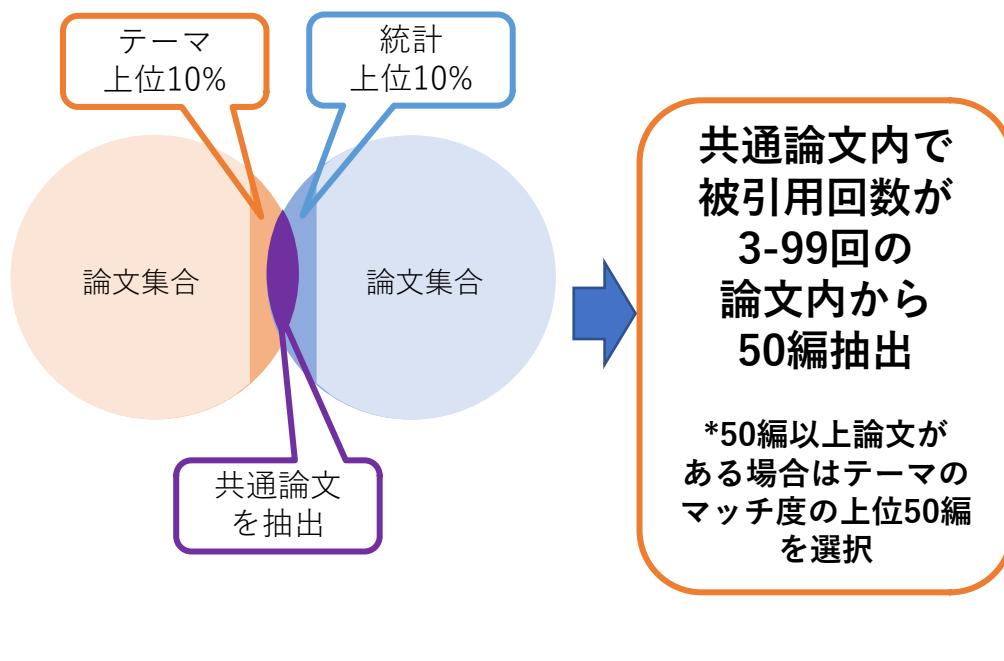
統計キーワードのトピック番号

テーマ名	キーワード	Topic #
統計関連	statistics	100
	regression	150

- ①テーマと統計関連のトピック内で
マッチ度上位10%の論文を抽出



- ②それぞれの上位10%論文から
共通論文を抽出



- ③被引用数を基に
50編抽出

共通論文内で
被引用回数が
3-99回の
論文内から
50編抽出

*50編以上論文がある場合はテーマの
マッチ度の上位50編を選択



4. 多様性指標のスコアを算出

- 抽出した論文ごとに多様性指標のスコアを計算

- 事前に定義した25分野のうち、評価対象論文・引用論文がどの分野かを特定
- 計算済みの接続確率を基にスコアを算出

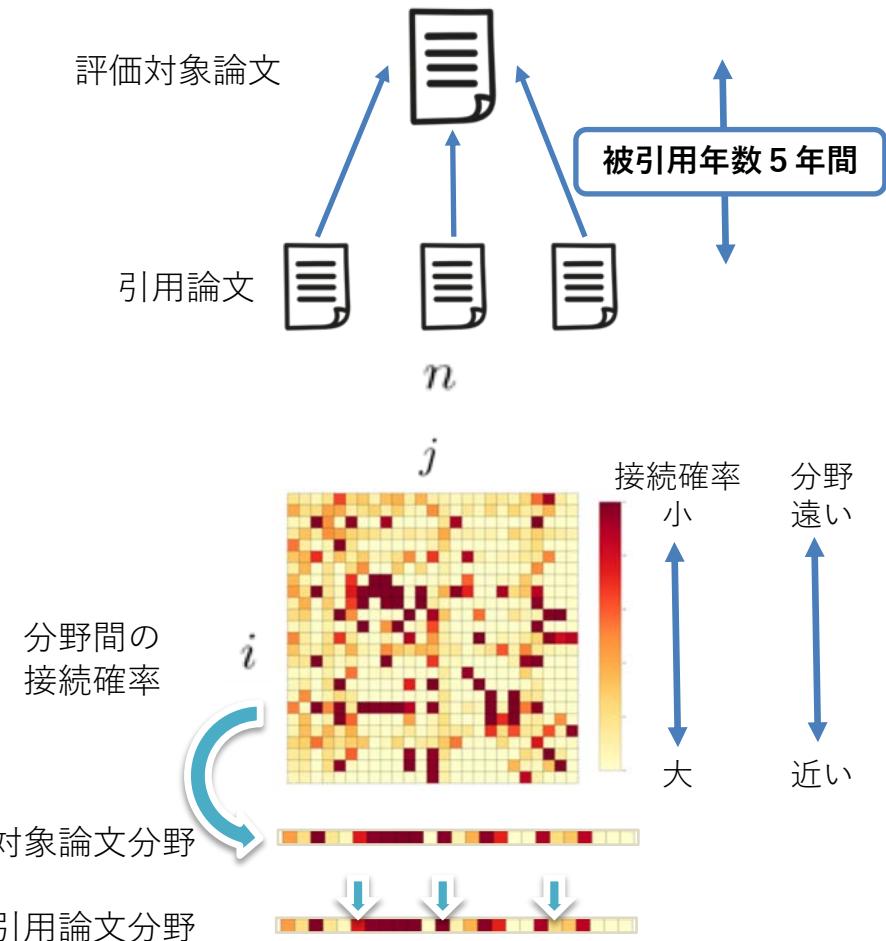
$$\text{REDi} = \frac{100}{n} \sum |1 - A_{(i \rightarrow j)}|$$

n = Cited count

i = Cluster of cited

j = Cluster of cite

$A_{(i \rightarrow j)}$ = Edge probability from i to j



5. 評価レポートの作成

- 評価対象論文50編について評価のためのレポートを作成

5. 評価レポートの作成

評価対象論文 詳細

H30-2：IRのための学術文献データ分析と統計的モデル研究の深化

No	論文のタイトル	論文の掲載ジャーナル	発行年	被引用数	多様性指標
1	Transdisciplinary Research, Transformative Learning, and Transformative Science	BIOSCIENCE	2013	11	33.08
2	Two Paths to Inequality in Educational Outcomes: Family Background and Educational Selection in the United States and Norway	SOCIOLOGY OF EDUCATION	2011	4	0.17
3	Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation	FRONTIERS IN COMPUTATIONAL NEUROSCIENCE	2012	15	66.06
4	Mapping a research agenda for the science of team science	RESEARCH EVALUATION	2011	19	46.72
5	Agent-Based Models in Empirical Social Research	SOCIOLOGICAL METHODS & RESEARCH	2015	6	96.95
6	Ethnic Inequality in Choice-driven Education Systems: A Longitudinal Study of Performance and Choice in England and Sweden	SOCIOLOGY OF EDUCATION	2012	9	11.28
7	The Impact of Prestige Orientation on Shadow Education in South Korea	SOCIOLOGY OF EDUCATION	2011	10	10.16
8	High Participation Systems of Higher Education	JOURNAL OF HIGHER EDUCATION	2016	3	0.07
9	It's never too late for 'us' to meet 'them': Prior intergroup friendships moderate the impact of later intergroup friendships in educational settings	BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY	2013	8	12.2
10	Fitting Science Into Legal Contexts: Assessing Effects of Causes or Causes of Effects?	SOCIOLOGICAL METHODS & RESEARCH	2014	4	49.81

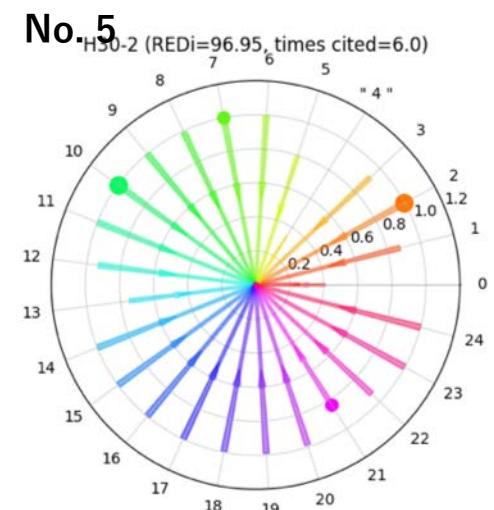
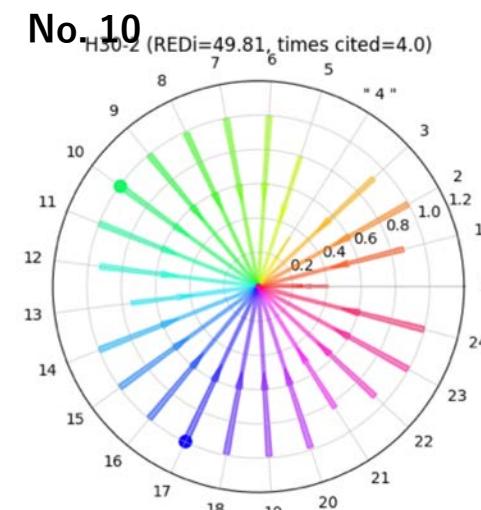
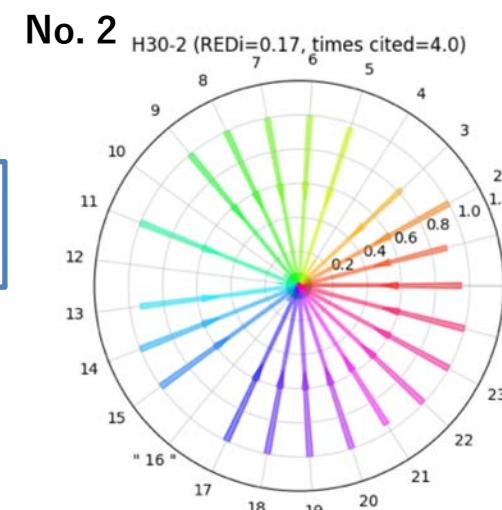
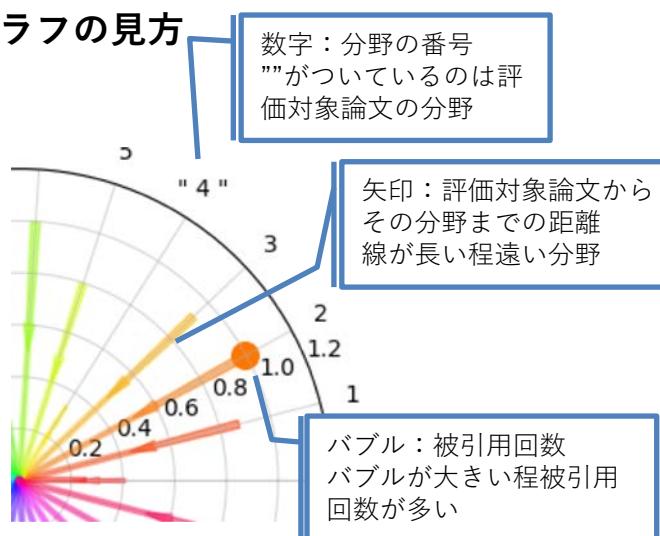
5. 評価レポートの作成

評価対象論文 多様性指標 可視化

H30-2：IRのための学術文献データ分析と統計的モデル研究の深化

No	論文のタイトル	論文の掲載ジャーナル	発行年	被引用数	多様性指標
2	Two Paths to Inequality in Educational Outcomes: Family Background and Educational Selection in the United States and Norway	SOCIOLOGY OF EDUCATION	2011	4	0.17
10	Fitting Science Into Legal Contexts: Assessing Effects of Causes or Causes of Effects?	SOCIOLOGICAL METHODS & RESEARCH	2014	4	49.81
5	Agent-Based Models in Empirical Social Research	SOCIOLOGICAL METHODS & RESEARCH	2015	6	96.95

グラフの見方



5. 評価レポートの作成

多様性指標スコア・被引用回数 基本統計量

評価対象論文の多様性指標のスコア

	H30-1	H30-2	H30-3	H31-1	H31-2	H31-3	H31-4	H31-5	H31-6	H31-7
count	50	50	50	50	50	50	50	50	50	50
mean	39.603	33.7084	35.2822	28.2066	30.605	29.5878	32.4524	35.1496	33.9868	32.4404
std	31.3622068	31.985609	32.2774758	31.9671141	31.2774097	29.1011305	28.8130084	31.5183033	35.4586785	35.2686554
min	0	0	0.02	0	0	0	0	0.01	0	0
25%	15.5025	4.625	4.39	0.2925	2.6175	3.3325	10.6025	11.475	0.405	6.4775
50%	32.99	25.585	24.485	19.715	24.02	21.67	27.76	26.35	21.36	15.835
75%	66.6675	50.3825	65.3225	40.44	43.6725	47.46	44.695	50.0175	64.3475	51.79
max	99.99	99.99	99.92	99.92	99.86	100	100	99.93	99.8	99.88

評価対象論文の被引用回数

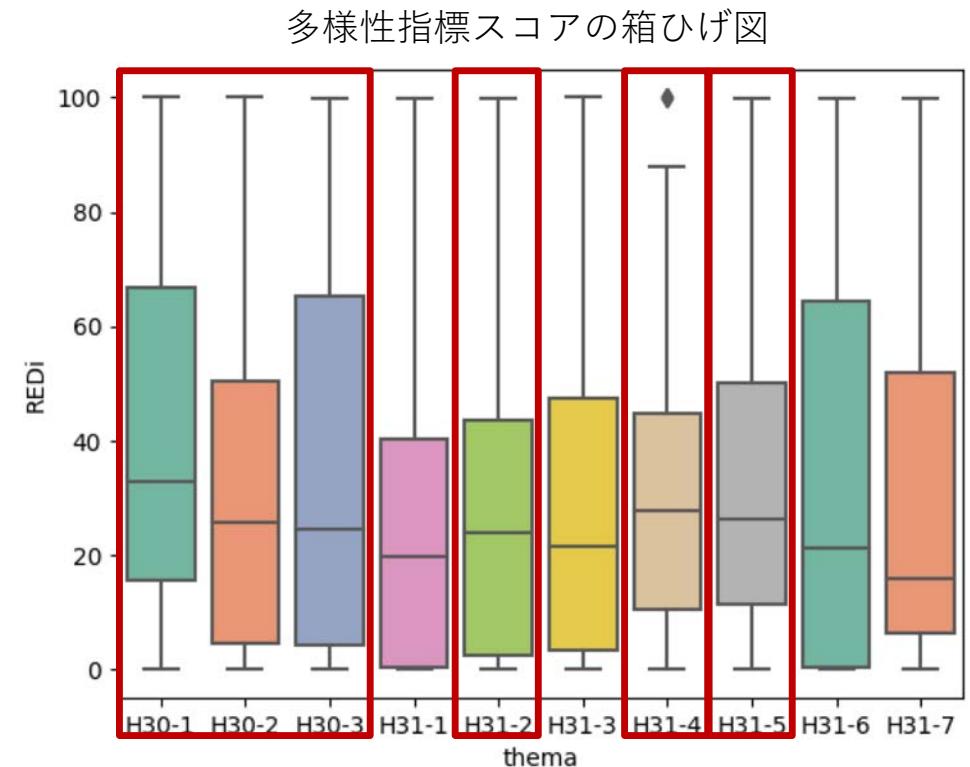
	H30-1	H30-2	H30-3	H31-1	H31-2	H31-3	H31-4	H31-5	H31-6	H31-7
count	50	50	50	50	50	50	50	50	50	50
mean	8.92	11.04	8.1	9.8	9.1	11.6	9.72	10.8	9.32	9.56
std	8.6963281	12.1082869	6.08863109	10.222025	12.9665976	14.9761715	8.89402459	11.8580034	8.60811198	6.43066633
min	3	3	3	3	3	3	3	3	3	3
25%	4	5	4	4	4	4	4	4	4	4
50%	7	8	6	6	5	6.5	7	6.5	6	7.5
75%	10.75	13.25	10	11.75	8	13	12.75	11.75	10	14.5
max	58	81	30	57	81	81	52	56	41	25

5. 評価レポートの作成

多様性指標スコア 評価

- 多様性指標の中央値から評価した場合
 - H31年度の課題TOP3
 - H31-2, H31-4, H31-5
 - H30年度の課題
 - H31年度の上位の値と同等のスコアとなり相応しい課題であったと考えられる

No	テーマ名
H30-1	データサイエンス人材育成メソッドの新展開
H30-2	IR のための学術文献データ分析と統計的モデル研究の深化
H30-3	統計的機械学習の新展開
H31-1	A
H31-2	B
H31-3	C
H31-4	D
H31-5	E
H31-6	F
H31-7	G



多様性指標の活用（2）

地球研との協働

H.31 2019年1月10日（木）

地球研-統数研 研究IRとオープンサイエンスに関する
ミーティング

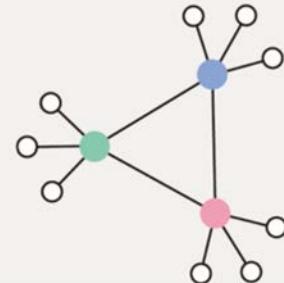
共著ネットワークと多様性指標の組み合わせによる分析

● > ●, ●

緑の研究者群と他の研究者
群の間の異分野度が高い

機関内外の共著ネットワーク分析手法の例

共著ネットワークの例。丸（ノード）が研究者、線（エッジ）が研究者間に共著論文があることを表している。色付きの3人の研究者はこのネットワーク上の中心的な位置にある。



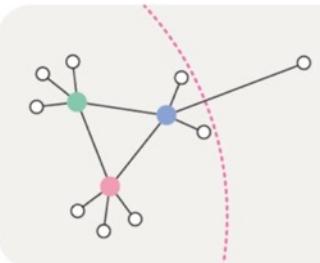
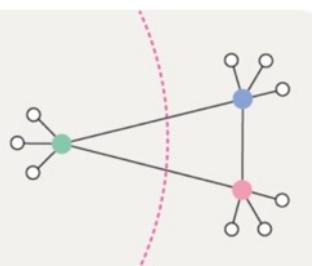
このような中心性をスコア化する指標として媒介中心性がある。ネットワーク上のハブ的なノード（この場合は研究者）を探索する場合に用いることができる。

上図では線の長さが同じであると仮定しているので
● ● ● の研究者の媒介中心性の値に差はない。

そこで**論文の異分野度**を線の長さとして導入した
重み付き共著ネットワークを考える。



（重み付き）媒介中心性に差が現れる



● > ●, ●

青の研究者が離れた分野との共同研究を実施している

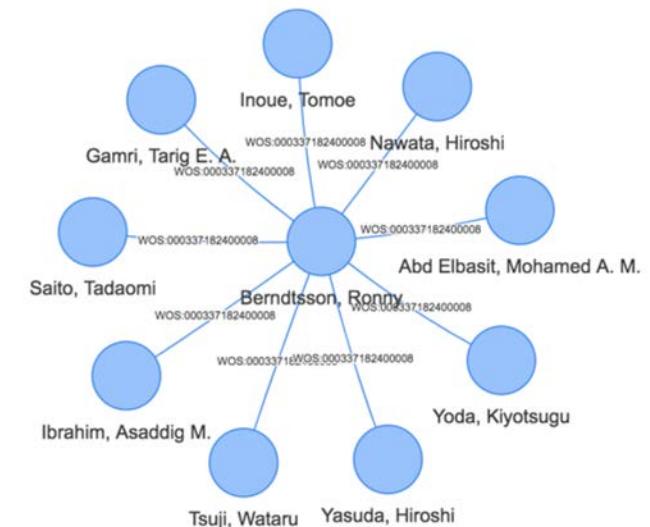
媒介中心性を用いた分析

- 論文の共著情報を用いて、グラフ分析手法の媒介中心性を算出
- 媒介中心性とは
 - あるノードが他のノード間の最短経路上に位置する程度を指標としたもの
 - 媒介中心性の高いノードは、そのグラフ内のノードの中で影響力が高いものとして考え

$$\begin{aligned} \text{betweenness}(i) &= \sum_{j < k} \frac{g_{jk}(i)}{g_{jk}} \\ &= \sum_{j < k} \frac{\text{点 } i \text{ を含む点 } j \text{ 点 } k \text{ 間の測地線の数}}{\text{点 } j \text{ 点 } k \text{ 間の測地線の数}} \end{aligned}$$

本分析のグラフの構造

- ノードプロパティ
 - full_name : 著者名
 - num_paper : 論文数
 - count_seqno : seq_numのsum
(論文数と加味して参考情報として利用)
- エッジプロパティ
 - uid : WoS の uid
- エッジは共著者に全部張る、無向グラフ



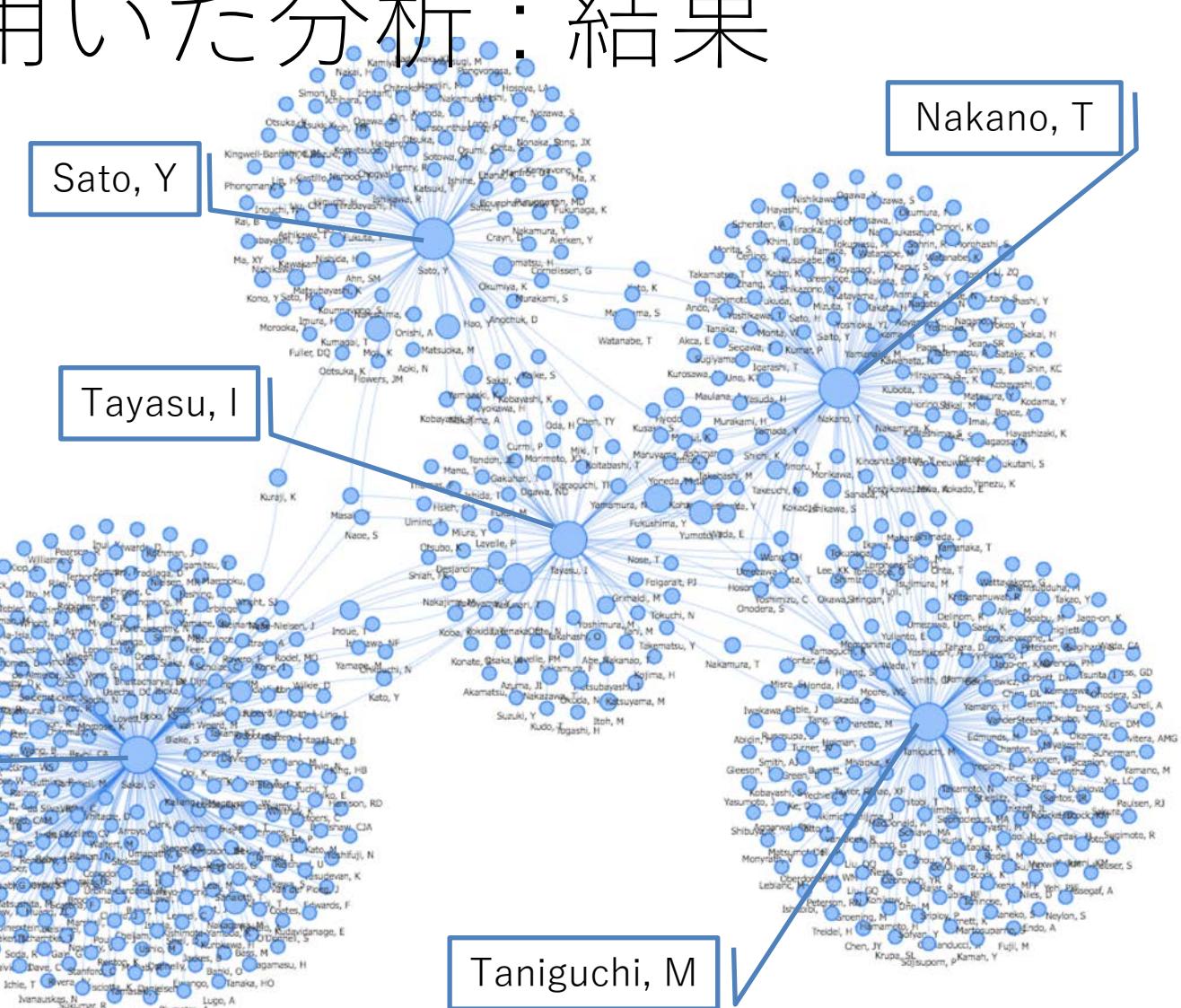
媒介中心性を用いた分析：結果

媒介中心性 TOP10

n. betweenness	id(n)	n. author_name	n. num_paper	n. count_seqno
781389.2292181909	1361	Nakano, T	46	159
751095.0652927489	1716	Sato, Y	26	105
711708.042568815	1988	Taniguchi, M	68	258
645261.4657163301	1677	Sakai, S	19	252
605490.0207185157	1998	Tayasu, I	25	93
541155.4836537292	383	Fukushima, Y	25	103
325580.6383957769	2309	Yasunari, T	16	65
315536.2155089928	1272	Moji, K	29	182
312837.6893439574	2325	Yokoyama, Y	8	31
311641.2833814308	1179	Matsuoka, M	13	57

媒介中心性TOP5 の共著者のネットワーク
(ノードの大きさ：媒介中心性の大きさ)

Sakai, S



Nakano, T

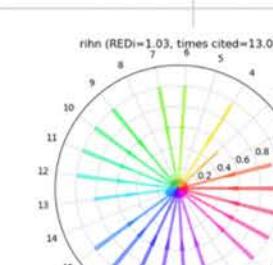
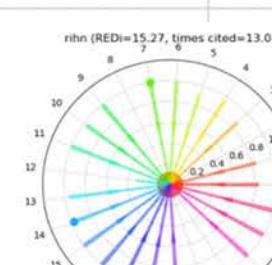
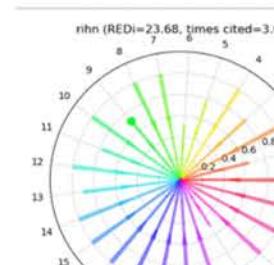
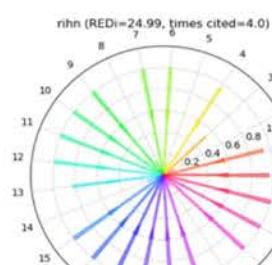
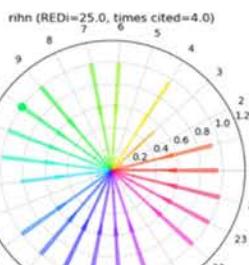
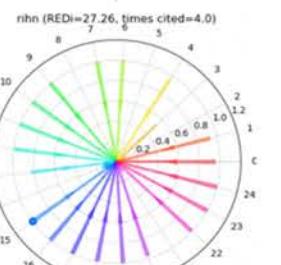
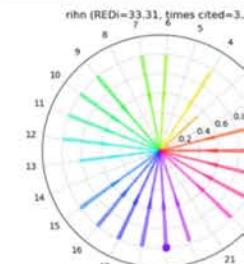
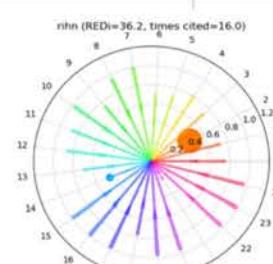
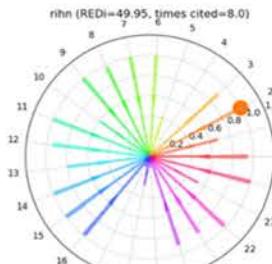
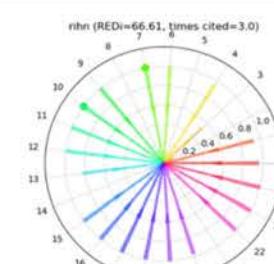
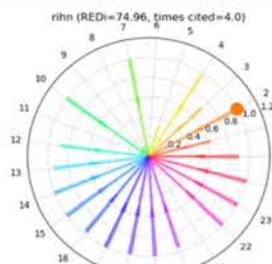
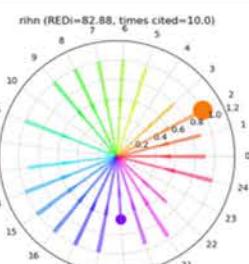
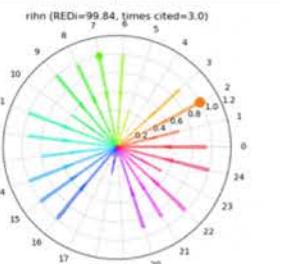
Sato, Y

Tayasu, I

Taniguchi, M

多様性指標（抜粋）

1 WOS:000367758200088	Is there an Environmental Kuznets Curve for SO ₂ emissions? A semi-parametric panel data analysis for China	Yuan Wang Rong Han Jumpei Kubota	RENEWABLE & SUSTAINABLE ENERGY REVIEWS	2016	3	99.84	19, 2, 7, 2
2 WOS:000347754700003	Methodological framework of sustainability assessment in City Sustainability Index (CSI): A concept of constraint and maximisation indicators	Koichi Mori & Tsuguta Yamashita bc	HABITAT INTERNATIONAL	2015	10	82.88	23, 2, 2, 2
3 WOS:000352716800002	Ecohealth research in Southeast Asia: past, present and the way forward	Nguyen-Viet, Doria S Tung DX, Mallee H, Wilcox BA, Grace D.	INFECTIOUS DISEASES OF POVERTY	2015	4	74.96	9, 2, 2, 2
4 WOS:000368206800011	The relationship between urbanization, energy use and carbon emissions: evidence from a panel of Association of Southeast Asian Nations (ASEAN) countries	Yuan Wang Lili Chen Jumpei Kubota	JOURNAL OF CLEANER PRODUCTION	2016	3	66.61	2, 10, 7
5 WOS:000356646200054	A semi-parametric panel data analysis on the urbanization-carbon emissions nexus for OECD countries	Yuan Wang Xiang Zhang Jumpei Kubota Xiaodong Zhu Genfa Lu	RENEWABLE & SUSTAINABLE ENERGY REVIEWS	2015	8	49.95	19, 2, 2, 19
6 WOS:000346907200109	The Release Rate of Environmental DNA from Juvenile and Adult Fish	Atsushi Maruyama, Keisuke Nakamura, Hiroki Yamanaka, Michio Kondoh, Toshifumi Minamoto	PLOS ONE	2014	16	36.2	21, 2, 14, 2,
7 WOS:000333776000021	A New Metric for Long-Range Transport Potential of Chemicals	Toru Kawai, Karolina Jagiełło, Anita Sosnowska, Katarzyna Odziomek, Agnieszka Gajewicz, Itsuki C. Handa	ENVIRONMENTAL SCIENCE & TECHNOLOGY	2014	3	33.31	2, 2, 19
8 WOS:000336654900004	Different isotopic evolutionary trends of delta S-34 and delta O-18 compositions of dissolved sulfate in an anaerobic deltaic aquifer system	Hosono, Takahiro & Lorphensriand, Oranuj & Onodera, Shin-Ichi & Okawa, Hirokazu & Nakano, Takanori & Applied GEOCHEMISTRY	2014	4	27.26	2, 2, 14, 15	
9 WOS:000333525400006	Initial design process of the sustainability science ontology for knowledge-sharing to support co-deliberation	Terukazu Kumazawa Email author Kouji Kozaki Takanori Matsui Osamu Saito Mamoru Ohta Keishiro Hara Michi	SUSTAINABILITY SCIENCE	2014	4	25	2, 10, 2
10 WOS:000329946000004	Extratropical forcing of tropical wave disturbances along the Indian Ocean ITCZ	Fukutomi, Yoshiki & Yasunari, Tetsuzo	JOURNAL OF GEOPHYSICAL RESEARCH-ATMOSPHERES	2014	4	24.99	2
11 WOS:000333088700009	Prevalence and risk factors of soil-transmitted helminth infection in Nepal	Parajuli RP, Fujiwara T, Umezaki M, Konishi S, Takami E, Maharjan M, Tachibana K, Jiang HW, Pahari K, \ Transactions of the Royal Society of Tropical	2014	3	23.68	6	
12 WOS:000334671600002	Estimating the water balance of a thermokarst lake in the middle of the Lena River basin, eastern Siberia	A. N. Fedorov P. P. Gavrilov P. Y. Konstantinov T. Hiroma Y. Iijima G. Iwahana	ECOHYDROLOGY	2014	13	15.27	14, 2, 7, 2, 2
13 WOS:000358322000003	Western Disturbances: A review	A. P. Dimri D. Niogi A. P. Barros J. Ridley U. C. Mohanty T. Yasunari D. R. Sikka	REVIEWS OF GEOPHYSICS	2015	10	11.06	2, 2, 17, 2,



まとめ

- 指標の開発、利用、利用推進
- Research metrics = 「研究IRの手法」の研究
 - データの問題
 - 分野の問題 地球研との協働
- 評価のフェーズとの関わり -> 次の発表
 - 内部評価/外部評価

Research metrics workshop

Research Metrics Workshop 2017

Date: Dec. 04-05, 2017

Place: Seminar Room 5, The Institute of Statistical Mathematics, Tachikawa, Tokyo

Admission Free, No booking Necessary

Research Metrics Workshop at ISM 2019

Date: March 14, 2019

Place: Seminar Room 2, The Institute of Statistical Mathematics, Tachikawa, Tokyo

[U.PORTO](#) Workshop on Research Metrics [Home](#) [Committee](#) [Speakers](#) [Abstracts](#) [Presentations](#) [その他](#) [検索](#)



Research Metrics Workshop at NUS 2019

14(Thursday Friday) - 15(Friday) November

Joint meeting with NLP group

Host: National University of Singapore: Department of Mathematics

E.O.F