

第53回産総研人工知能セミナー（2022年1月12日）

「科学技術を発展させるAI技術のフロンティア」

テクノロジー・インフォマティクスによる 科学技術の未来予測

坂田一郎

東京大学 工学系研究科教授/総長特別参与
/テクノロジー・インフォマティクス社会連携講座代表
isakata@ipr-ctr.t.u-Tokyo.ac.jp

※学術論文データベースScopus全7千万論文の可視化

アブストラクト：

今日、科学技術に関する情報・知識が急激に増加しており、人間の情報処理能力に依存した従来型のエキスパートアプローチでは、それらを活かしきれない（「情報の海に溺れる」）状況にある。そうした事態に対応して、コンピュータショナルなアプローチを導入する動きが世界中で進んでいる。

我々は、自然言語処理、ネットワーク科学、機械学習等の手法を組み合わせ、科学技術ビッグデータを解析し、政府の研究開発戦略、企業の技術経営や個々の研究者の研究計画の立案などに対し有用な知見を提供する手法を「テクノロジー・インフォマティクス」と呼んでいる。

この手法とWeb of ScienceやScopusに収録されている大規模な論文データ及び産業技術総合研究所のABCIを用いた知識の構造化、技術の未来予測、バンドワゴン効果の把握、ネットワークを通じた科学技術ミームの伝播等の分析事例やそうした分析機能の一部を実装したウェブシステム「学術俯瞰・未来予測システム」を紹介する。

©Ichiro Sakata



自己紹介 ^{さかた} ^{いちろう} 坂田 一郎

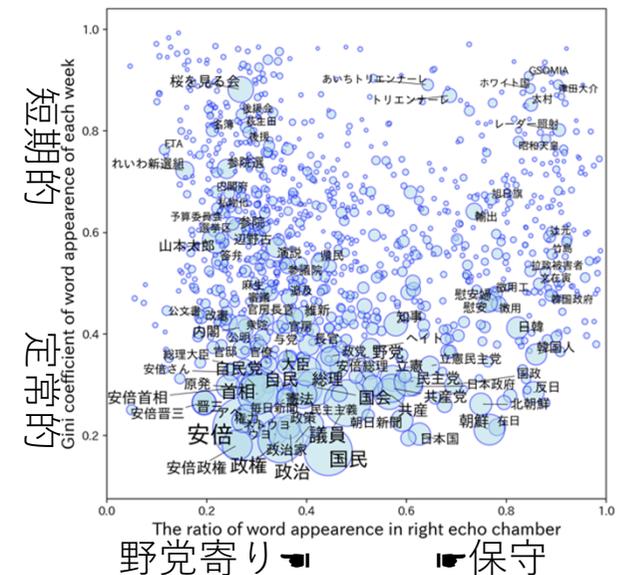
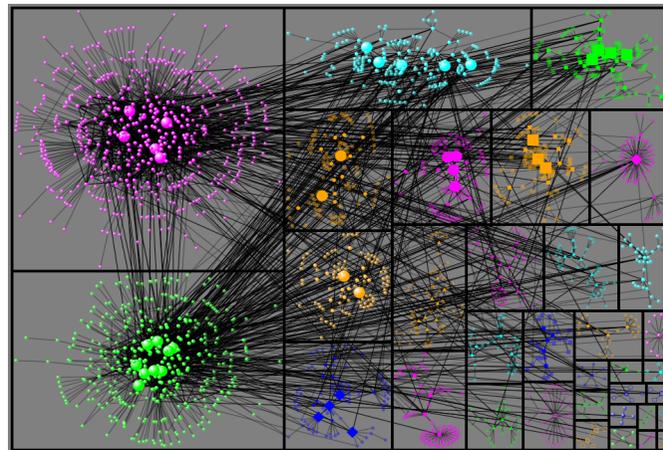
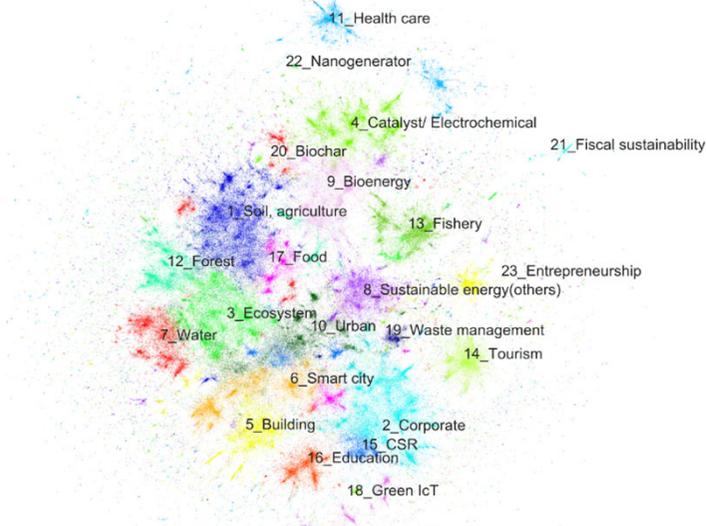
機械学習、計算自然言語処理、複雑系ネットワーク科学の手法（データサイエンス）を組み合わせ、人の判断によって行われる様々な活動から知見を引き出す研究を行う。主な対象としては、学術研究、企業の取引やイノベーション活動、ソーシャルメディア内での会話、鉄道を利用した人の移動。それら研究成果を意思決定支援ツールとして社会応用。学務としては、総長特別参与、FSI本部ビジョン形成分科会長、IFI副センター長等。

「テクノロジー・インフォマティクス」で政策判断や経営のDXに貢献

知識の構造化

地域の産業構造の新指標

ソーシャルメディア社会研究



サステナビリティ

被災地の企業ネットワーク

- ✓ 世界の30万件の知識マップ
- ✓ 「学術俯瞰・未来予測システム」特許化
- ✓ グローバル企業にライセンスし、研究戦略の意思決定支援
- ✓ 無から有を生み出す基礎研究のプロセス解明

- ✓ 「コネクター・ハブ」（地域の中核企業）
- ✓ 東北復興 ネットワーク補助金
NHK「震災ビックデータII」
- ✓ 「地域経済情報分析システム(RESAS)」
- ✓ 経産省「地域未来牽引企業」選定

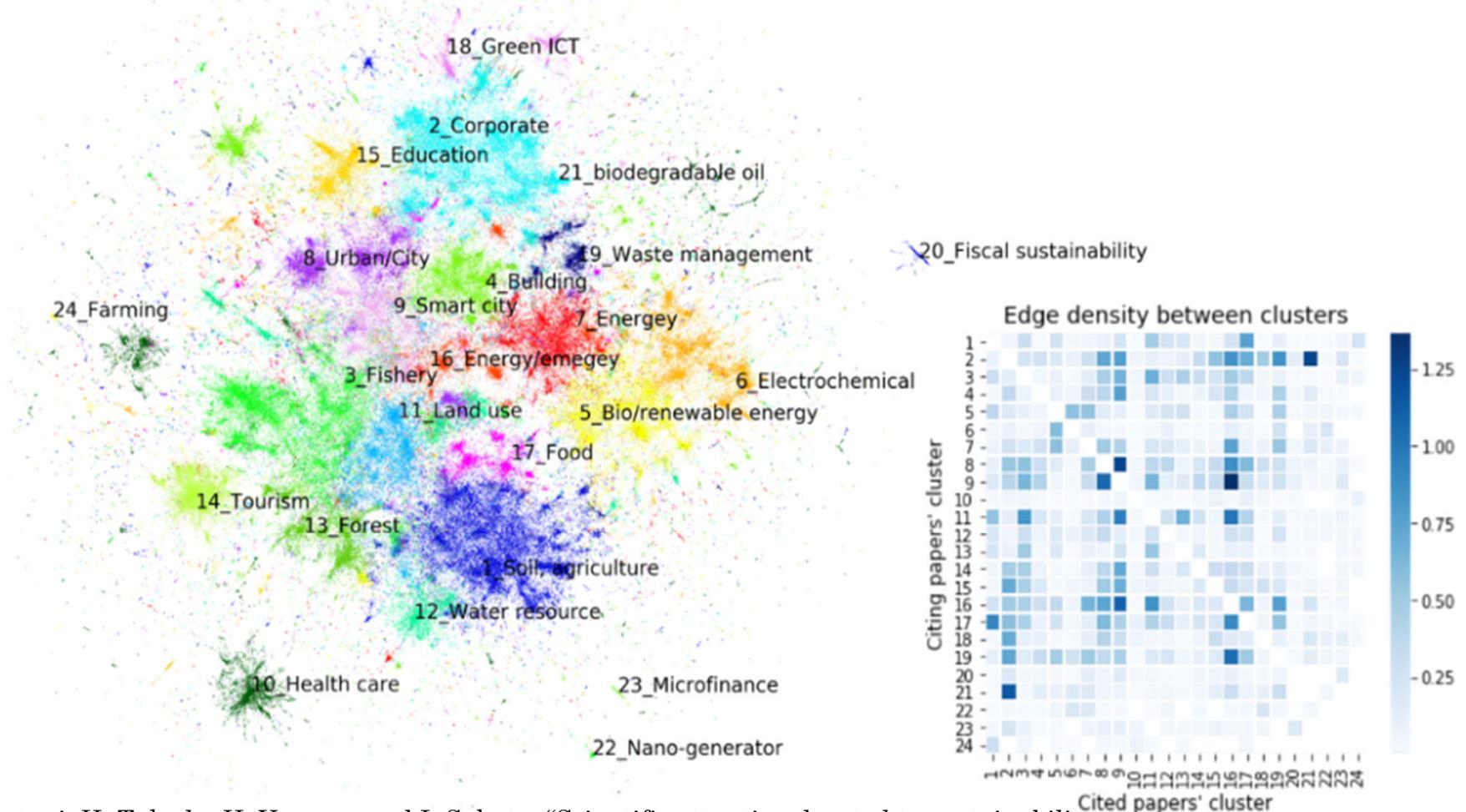
- ✓ 社会課題の早期特定システム
- ✓ 「エコーチェンバー現象」（意見が異なるグループに分離し固定化）のメカニズム解明
- ✓ 「沈黙の螺旋」（中間な意見を持つ層が発言を控える現象）の特定

[例1]

膨大な論文の山の中から知見(宝)を探す

サステナビリティに関する30万本の学術論文を引用関係を元に分類し、知識を構造化
手法: クラスタリング(グループ分け)や特徴語のコサイン類似度によりデータの意味を知る

👉 世界のサステナビリティに関する「知識の構造化」や内部の議論の把握

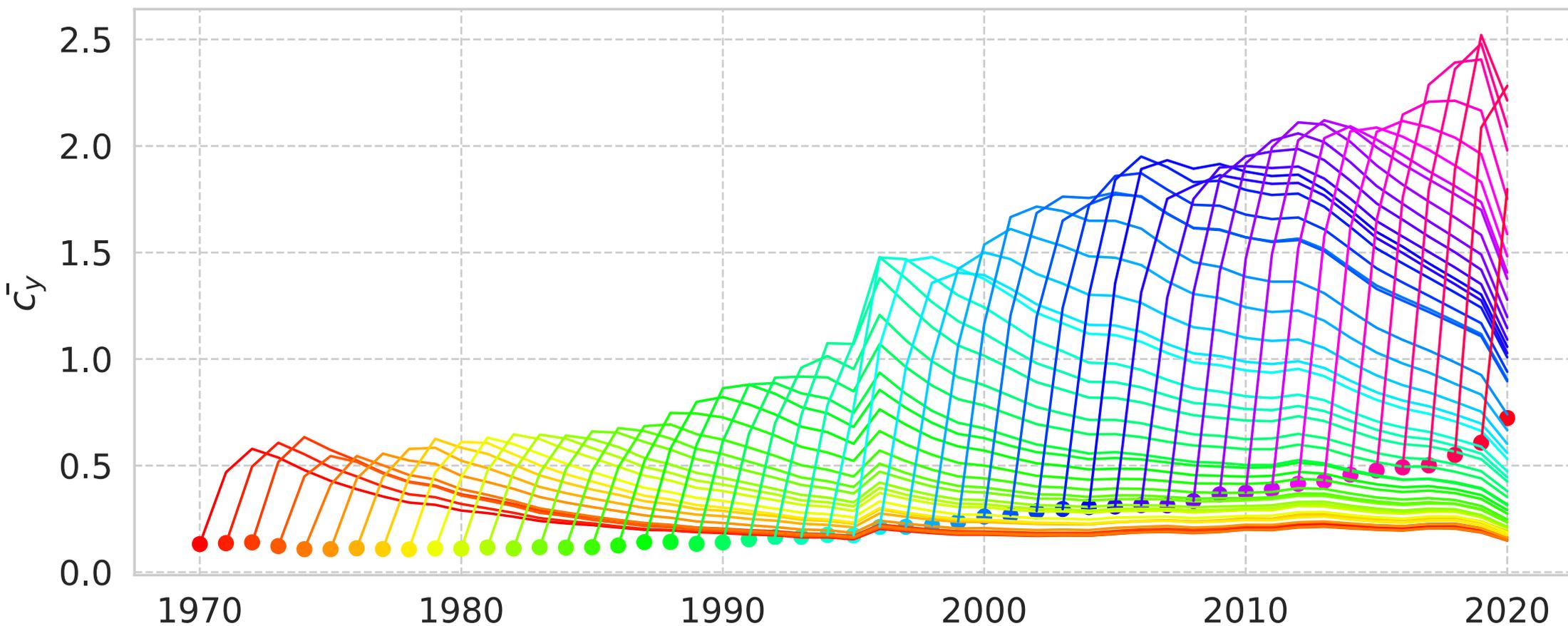


(出典) K. Asatani, H. Takeda, H. Yamano and I. Sakata, "Scientific attention devoted to sustainability and SDGs: Meta-analysis of academic papers", **Energies** vol.13(4) (2020),

[例2]

時系列でみる知の伝播の構造

引用＝著者による内容の関連性に関する判断と捉えて、引用関係やそれによる引用関係ネットワークを、知識の構造化だけでなく、多くの研究で活用している。



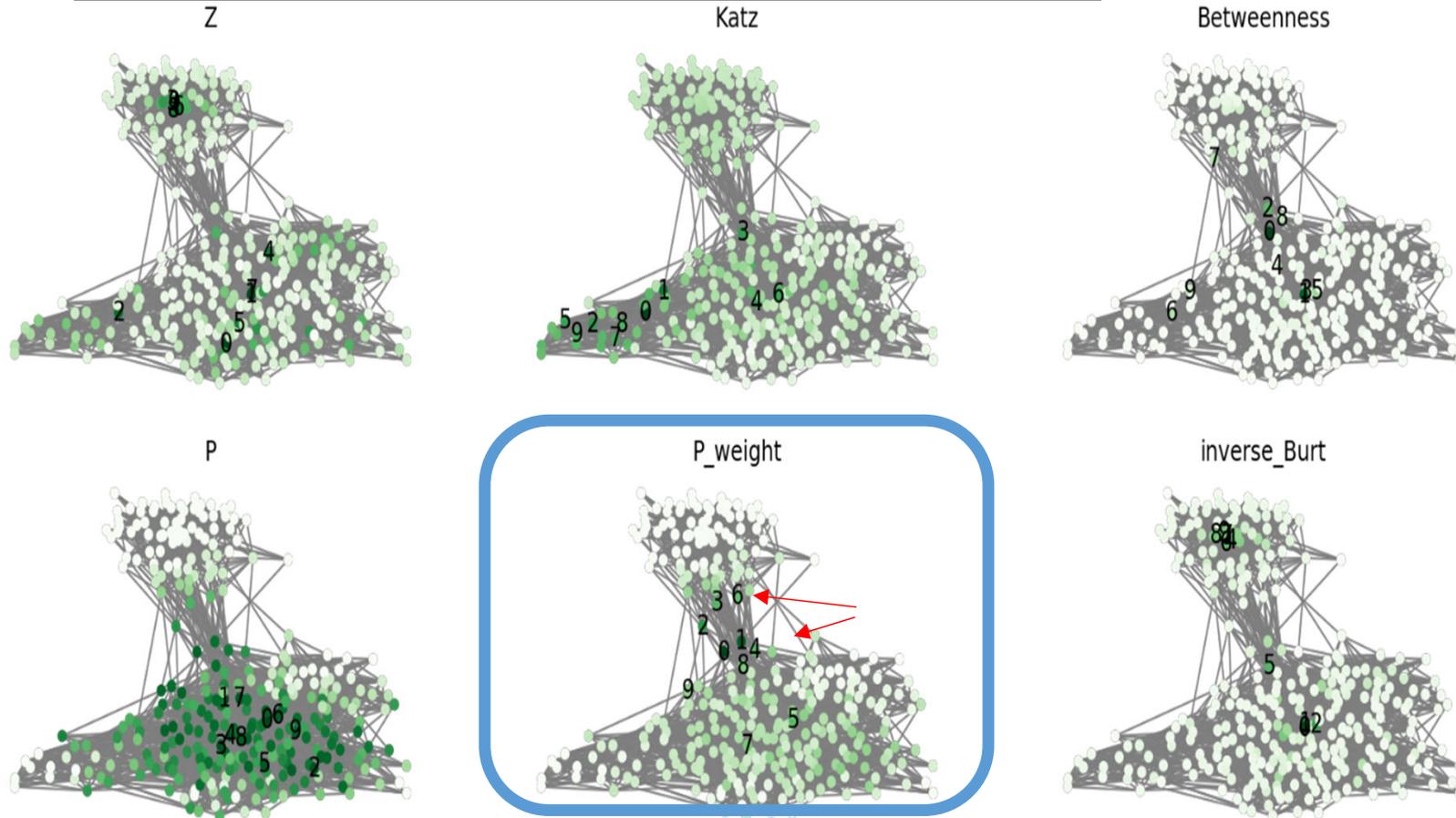
エルゼビア社Scopusの全データを用いた引用関係の時系列展開

(出典) T. Mira, K. Asatani, I. Sakata, "Large-scale analysis of delayed recognition using sleeping beauty and prince", *Applied Network Science* 6:48 (2021)

[例3]

多様性重視したネットワーク科学の新指標PW

👉 地域における企業間の取引関係ネットワークの可視化(下記は、東北地方の例)
稀少なパスを持ち、コミュニティの情報多様性に貢献するブローカー企業を特定
地域的な産業集積政策、地域未来牽引企業の選定に活用



知的対流に重要な企業の抽出指標

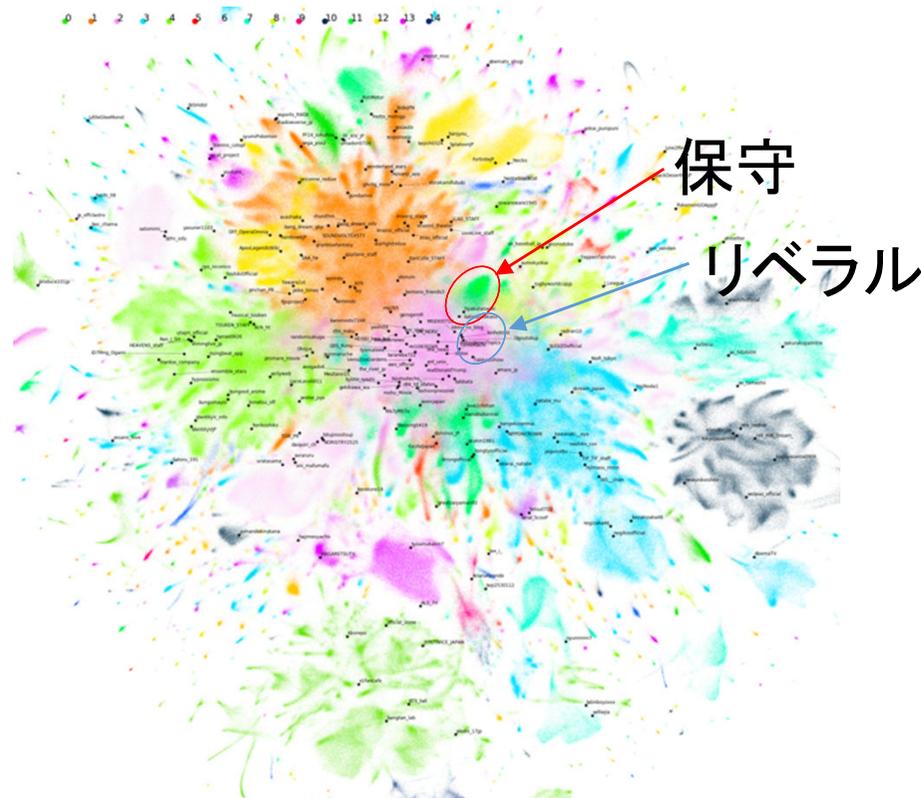
(出典) H. Yamano, K. Asatani, I. Sakata, "Evaluating Nodes of Latent Mediators in Heterogeneous Communities", *Scientific Reports* 10 (2020), 8456.

[例4]

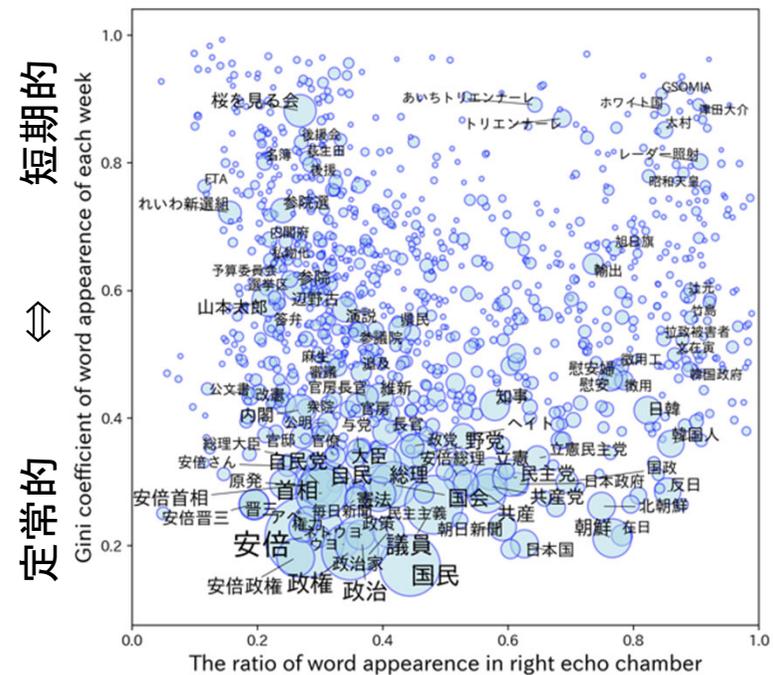
Twitter内の情報伝播の構造を特定

ツイッター4200万人のユーザー(396日、25億ツイート)の分類を行い、政策に関して濃度の高い議論を行っているクラスタ(ユーザーのグループ)を特定し、詳細に分析

👉 エコーチェンバー内のユーザーのCore-peripheryの構造、情報の流れの構造を説明、
市民的な政策課題の抽出



ツイッター4200万人のマップ



野党より ⇔ 与党寄り
2クラスタ内の特徴的な話題

(出典) K. Asatani, T. Sakaki, H. Yamano, I. Sakata, "Dense and influential core promotes daily viral information spreading in political echo chambers", *Scientific Reports* 11 (2021), 7491.

[例5]

SNSの世界の対話の特徴と世論の把握

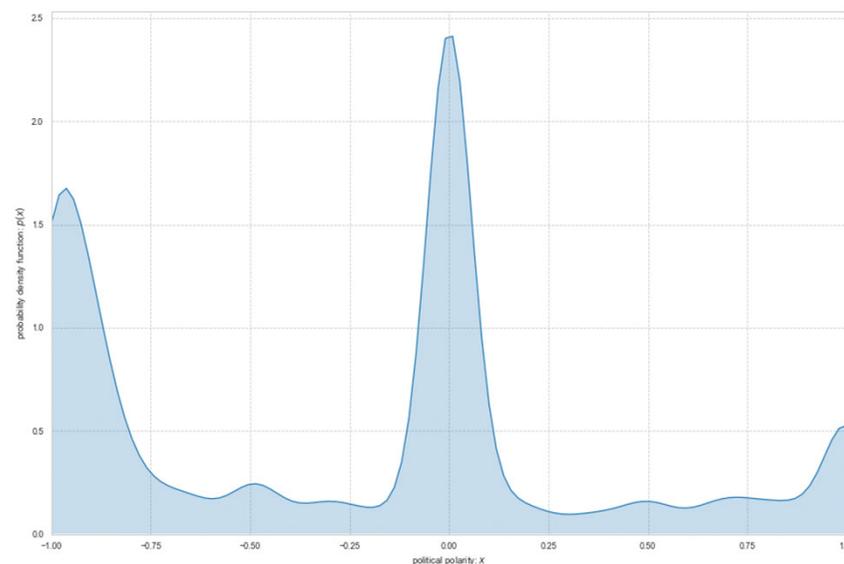
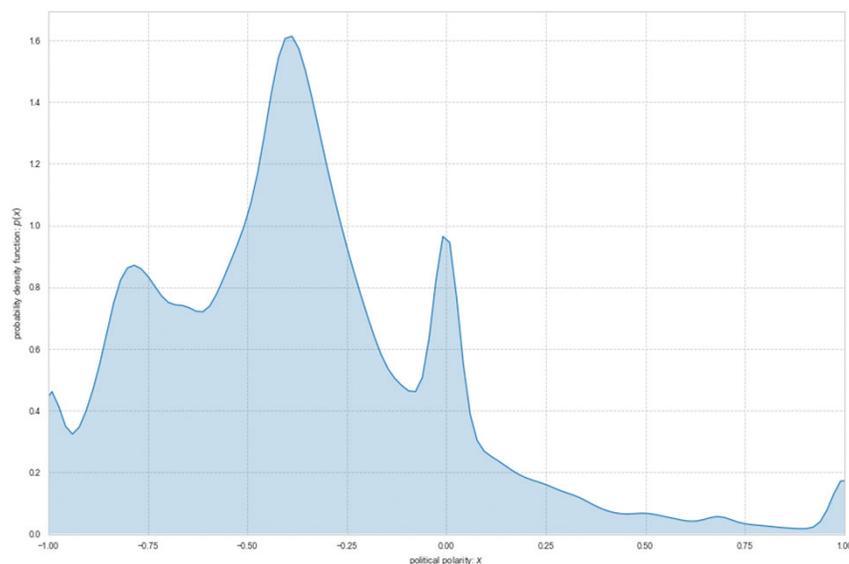
政治的な話題における「分極化」の発生

👉 構造を把握することによる世論の正確な把握

全ての話題を含む

2019年の参議院選挙のみ

ツイートの頻度



←
リベラル

→
保守

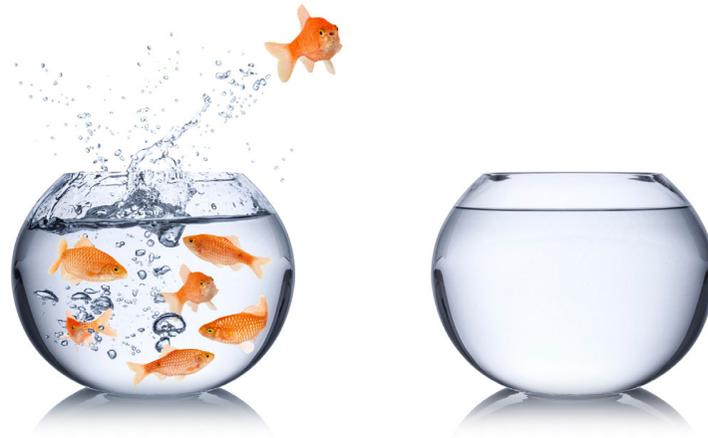
←
リベラル

→
保守

(出典)N. Sato, K. Asatani, T. Sakaki, and I. Sakata, "Measuring Polarization in Topic-independent Network on Twitter", NetSci 2020, The Network Science Society, delivered online (September 21-25, 2020).

テクノロジー・インフォーマティクス

(実用化事例: 学術俯瞰・未来予測システム)



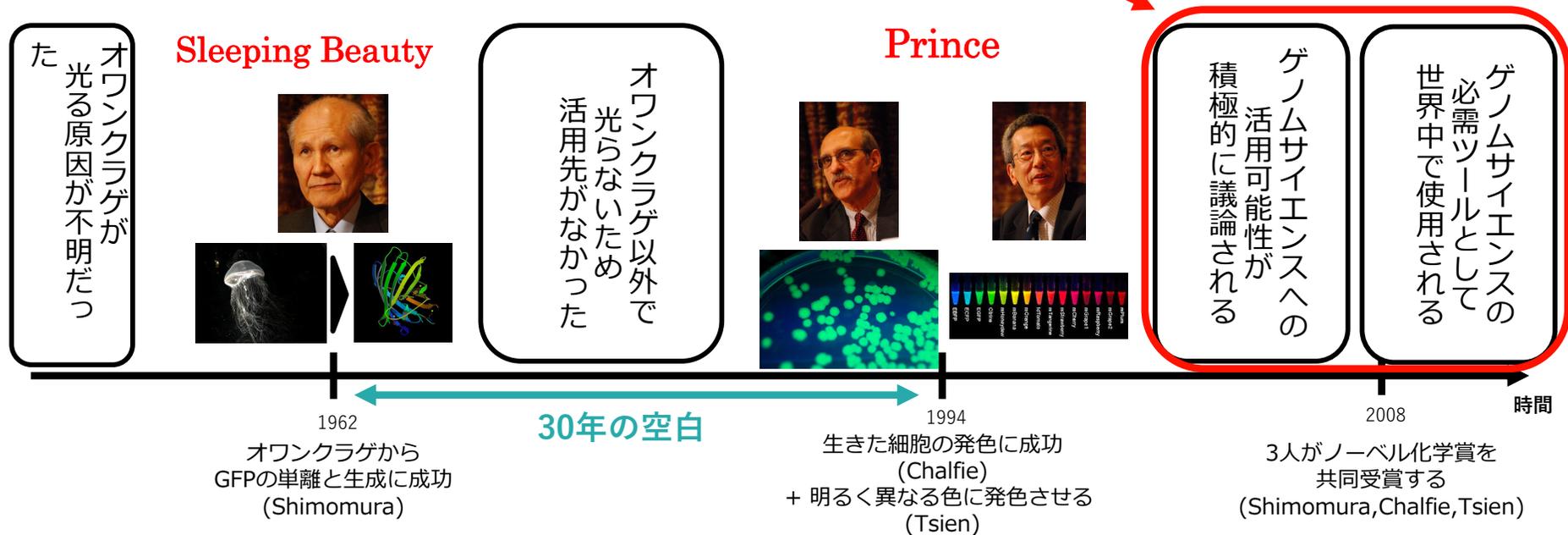
サイエンスリンケージの上昇

例: 緑色蛍光タンパク質(GFP)の発見

- 2008年にノーベル化学賞を受賞した発見
- 分野の流行とは全く異なる形で行われ、後に新しい分野を形成
- SBとPRの双方がノーベル賞を受賞
- 我々の3.7万件のSBとPRの組の分析では平均の睡眠期間は26.7年

その他の例

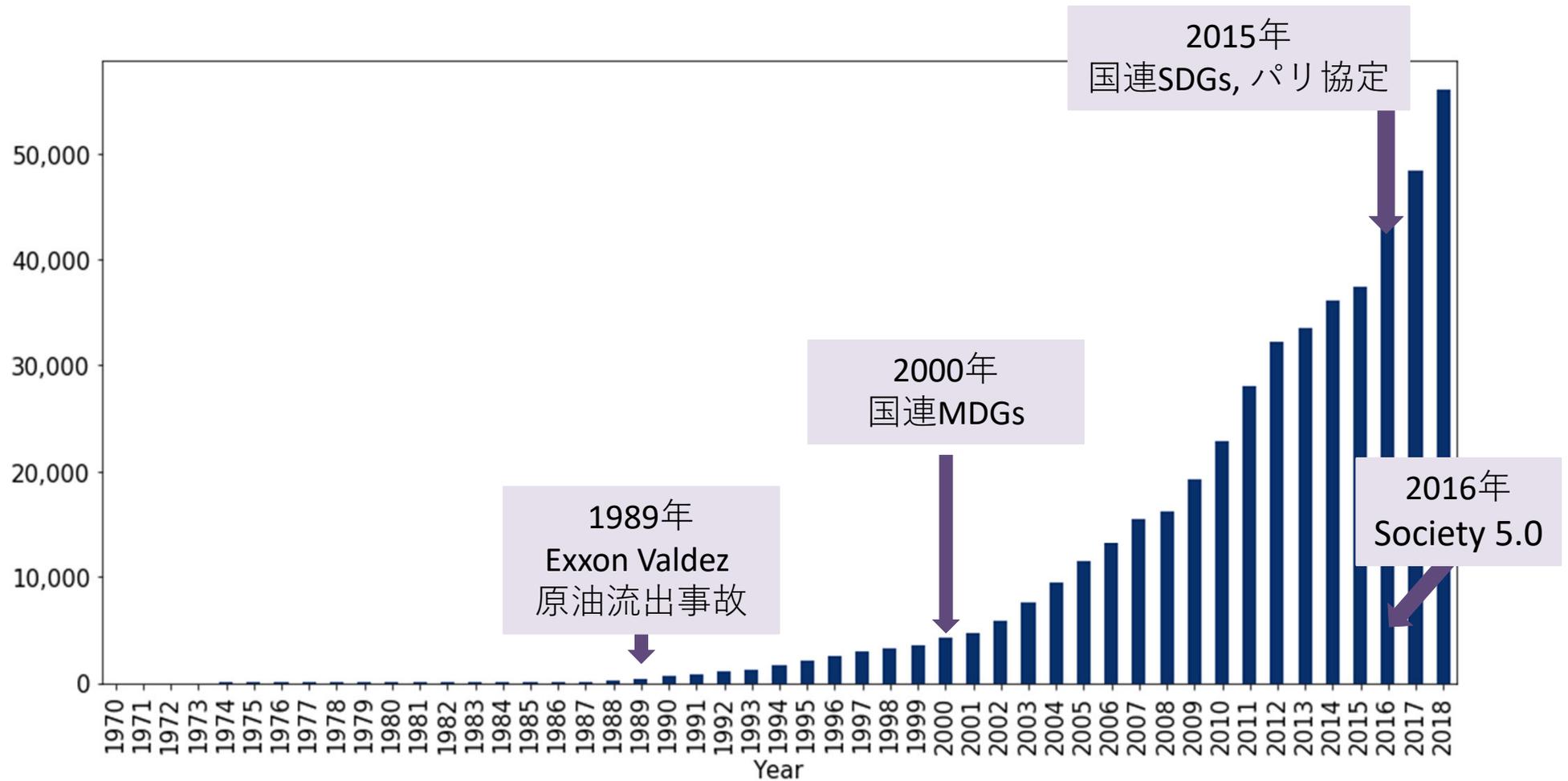
- **カテーテル**
当時禁忌とされた人体実験の結果がのちに証明される
- **水星の近日点移動**
惑星の観測結果が万有引力の発見により説明可能に
- **Deep Learning**
AI冬の時代を超えて、2010年から大きな発展を遂げる



(参考) GFP研究の歴史を紐解く ~下村.Chalfie,Tsien博士の偉業~ 宮脇敦史/サイエンスネット 第35号 2009年5月

(出典) T. Mira, K. Asatani, I. Sakata, "Large-scale analysis of delayed recognition using sleeping beauty and prince", *Applied Network Science* (2021)6:28.

知の爆発(サステナビリティ領域)

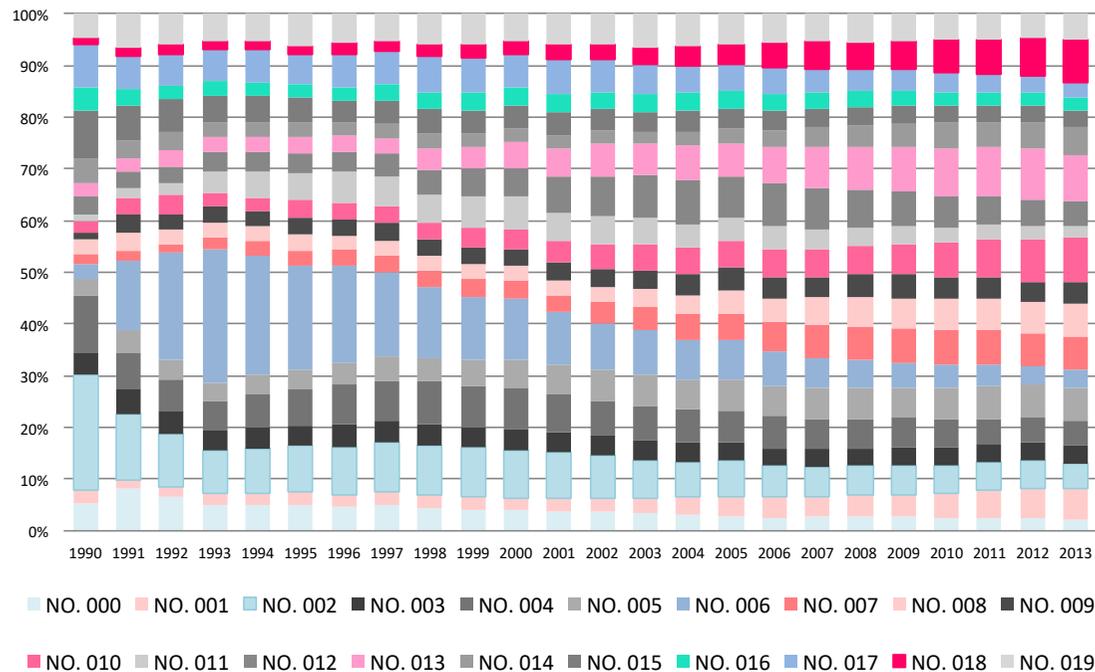


(備考) Scopusよりsustainability関連の論文(50万件弱)を取得し分析

知の構造変化の加速

ケース: Nano- carbon

30万件のナノカーボン領域の論文群について、ダイナミックピックモデリングを用いて分析したところ、上位 20トピックに、バッテリー、バイオセンサー、ナノ構造材、吸着剤などの応用研究が多く見られるようになってきている。



① Diminishing Topics

0、2、6、16、17

② Growing Topics

1、7、8、10、13、18、

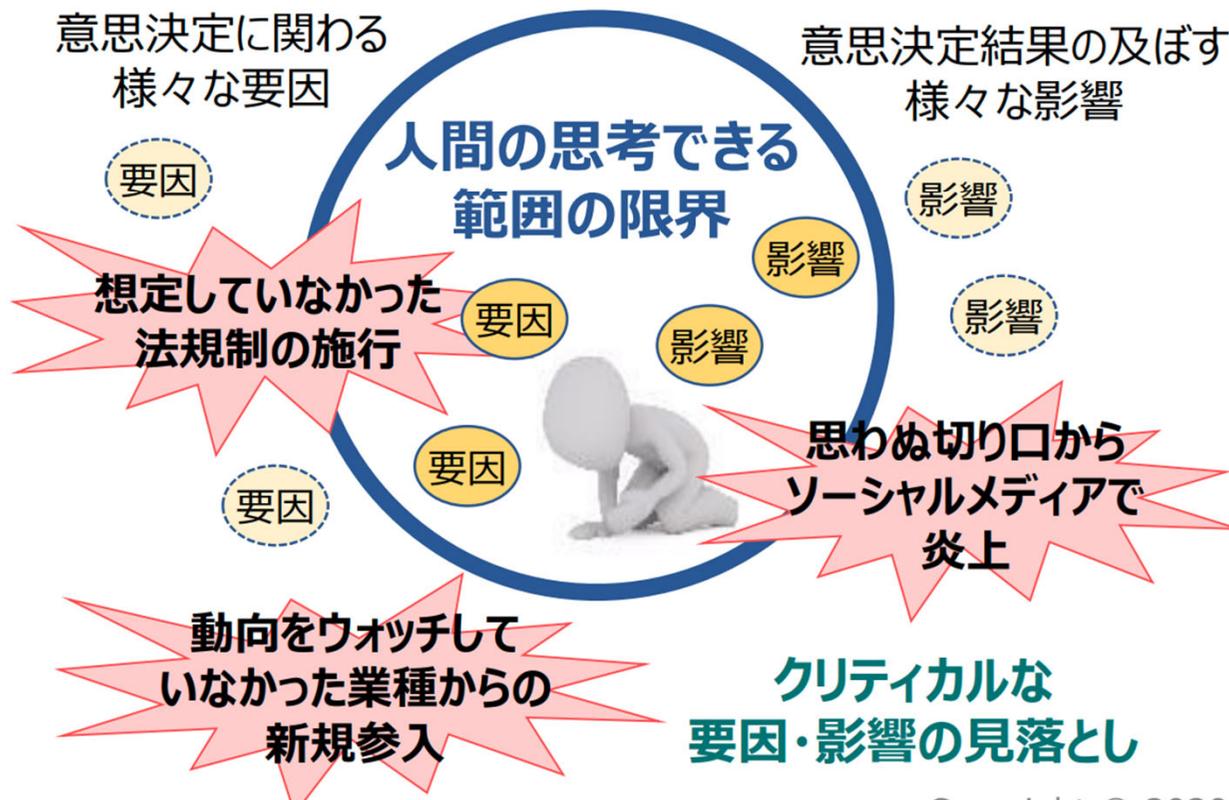
Topic No.	Topic Label
No.0	Biosensor
No.1	Lithium Ion Battery
No.2	Environmental Assessment (Eco System)
No.3	Environmental Assessment (Composites Analysis)
No.4	Bio Medicine
No.5	Adsorption
No.6	Fullerene
No.7	Nano Composites
No.8	Biosensor
No.9	Chemical Application
No.10	Nanoparticle
No.11	Solar Cell with Fullerene
No.12	Growth Mechanism of Carbon Fiber
No.13	Graphene
No.14	Drug Delivery
No.15	Methodology of making Nano Carbon
No.16	Basic Properties
No.17	Evaluation of Basic Properties
No.18	Catalysts and Battery
No.19	Alloys with Carbon Nanotube

(出典) Y Nakashio, et al., Detecting structural changes of nano carbon domain based on time distribution of text information of academic papers, Portland International Conference on Management Engineering and Technology 2016 (PICMET'16), in Hawaii (4-8, Sep, 2016)

「限定合理性」問題の顕在化

限定合理性の問題が深刻化

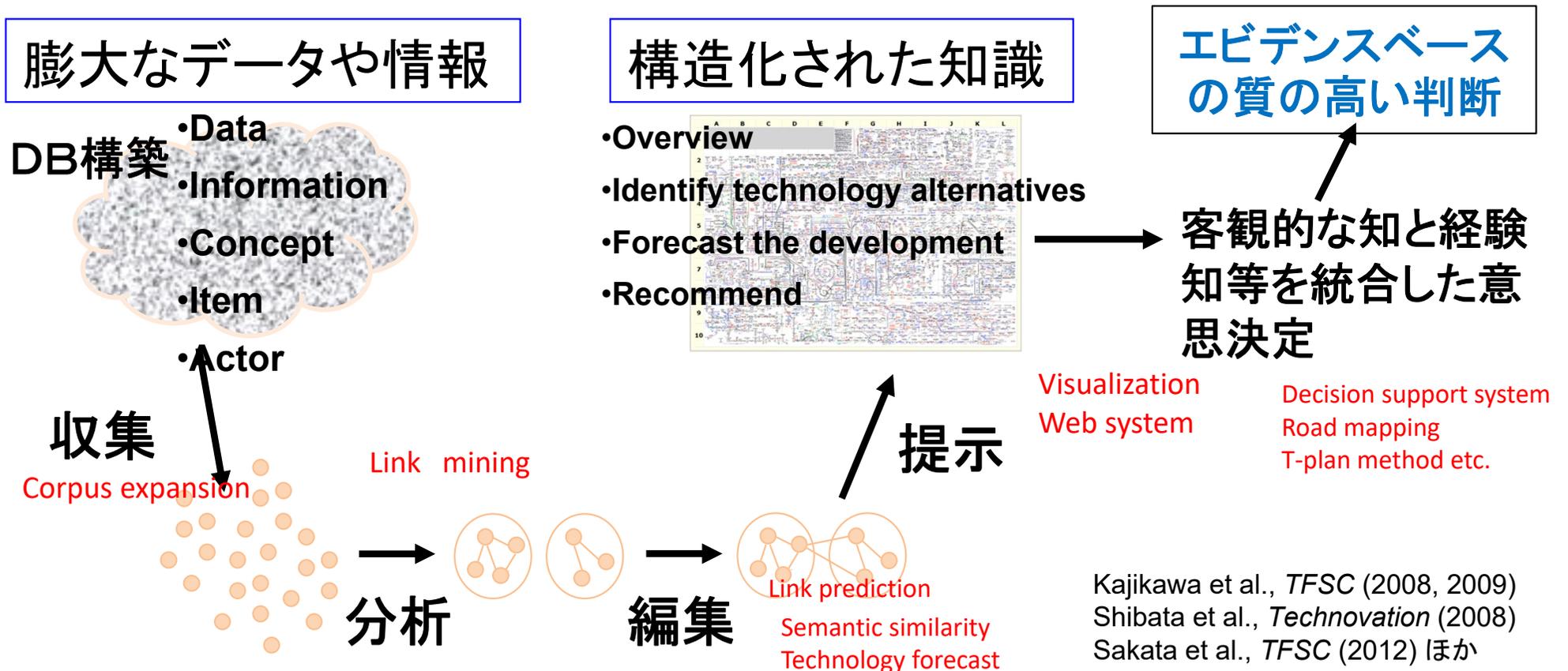
- 情報爆発やボーダレス化によって、意思決定に関わる要因・影響の可能性が膨大に
- クリティカルな要因・影響の見落としが深刻化



科学技術ビッグデータと意思決定

知識の構造化と意思決定支援

論文・特許、ウェブデータ、形式知化された経験知などの膨大な情報を収集・分析し、専門的知見による解釈を通じて、知識の森を理解可能にする手法を開拓。人間の認知能力の限界から来る「埋没知」の壁を超え、科学技術ビッグデータを有効活用し、意志決定の質の向上、早期化や知識の社会的有効利用を実現。



科学技術ビッグデータ活用の諸課題

収集

- 情報の散在、フォーマットの不統一、マッシュアップの困難さ
- 情報の申告コストの高さ(アメリカではPIの時間の42%が管理的業務)
- 情報の申告に関するインセンティブ不足(対極 ref. Research Gate)

分析

- コストと時間を要する専門家WSへの過度な依存
- 知の細分化による人間系での統合の困難さ
- 情報/ウェブ工学的手法の活用の不足
- ニーズや課題を知る現場で利用可能なインターフェース不足(対応例 RESAS)

橋渡し

- インプットとアウトプット(GDP、社会課題の解決への貢献)との接続の難しさ
- 情報取得期間とアウトプットを生む期間との差異
- 原情報発生とそれが利用可能となる時期とのタイムラグ(即時性の不足)
- 組織的な文化との不整合(途中でのスクリーニングの困難さ、専門家との溝等)

科学技術ビッグデータ活用の諸課題

収集

- 情報の散在、フォーマットの不統一、マッシュアップの困難さ
- 情報の申告コストの高さ(アメリカではPIの時間の42%が管理的業務)
- 情報の申告に関するインセンティブ不足(対極 ref. Research Gate)

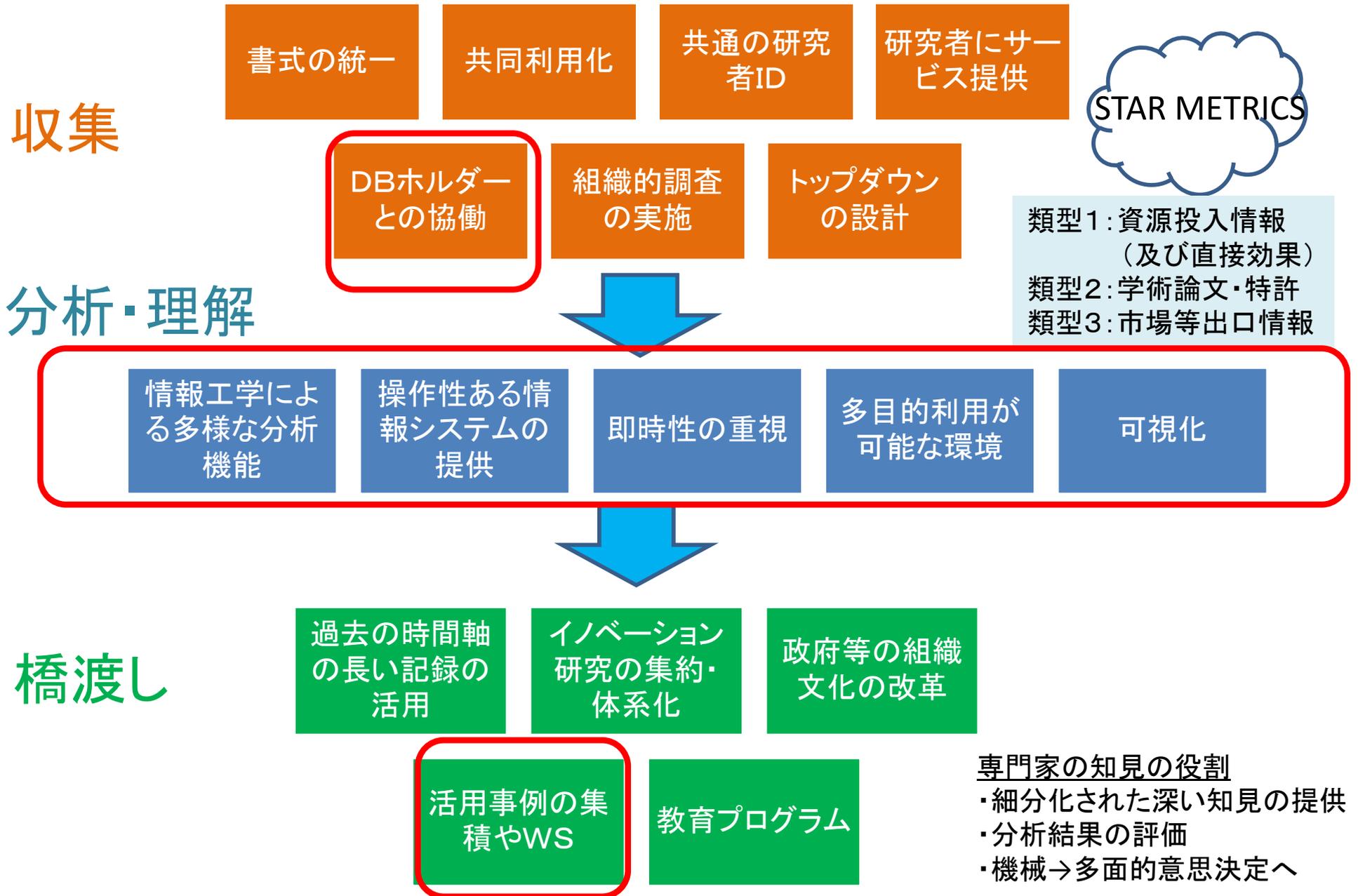
分析

- コストと時間を要する専門家WSへの過度な依存
- 知の細分化による人間系での統合の困難さ
- 情報/ウェブ工学的手法の活用の不足
- ニーズや課題を知る現場で利用可能なインターフェース不足(対応例 RESAS)

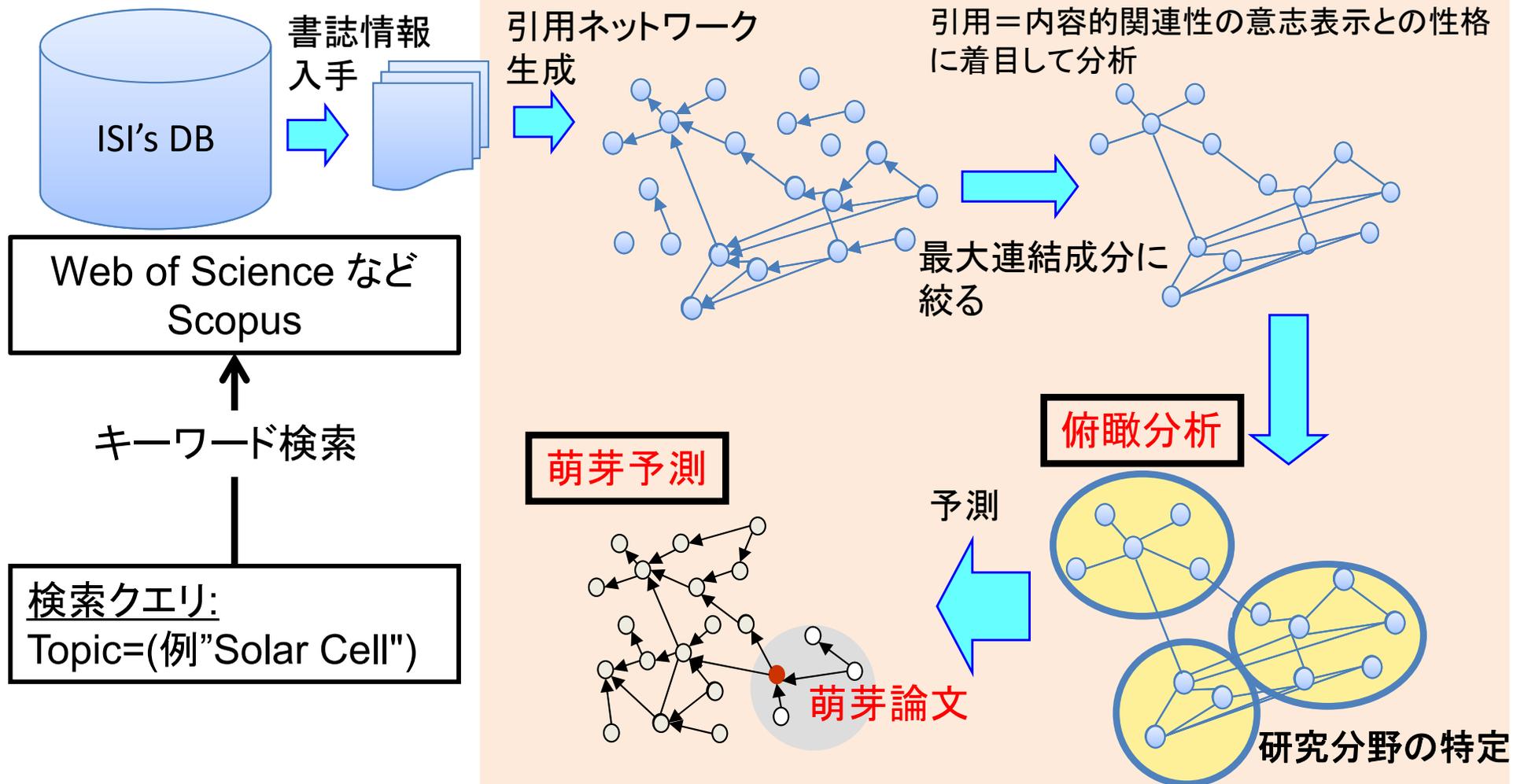
橋渡し

- インプットとアウトプット(GDP、社会課題の解決への貢献)との接続の難しさ
- 情報取得期間とアウトプットを生む期間との差異
- 原情報発生とそれが利用可能となる時期とのタイムラグ(即時性の不足)
- 組織的な文化との不整合(途中でのスクリーニングの困難さ、専門家との溝等)

3つの段階における課題の解決策



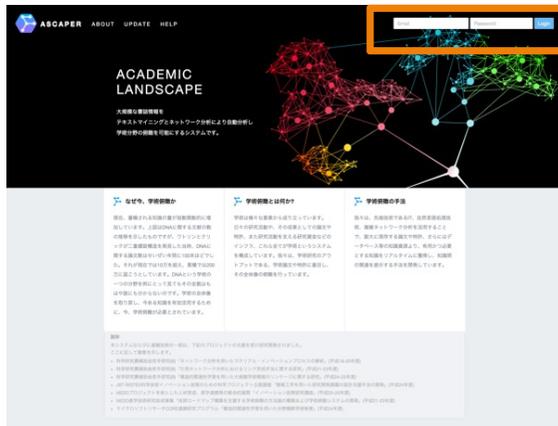
意思決定支援システム構想



1. Web of Science から取得した論文情報から、引用ネットワークを作成
2. ネットワーククラスタリングを実行し、内容的に近い領域(クラスター)に自動分類
3. クラスターの可視化(知識領域)、重要情報の抽出、ネットワーク指標の導出
4. これらの情報をもとに将来数多く引用される萌芽論文を特定

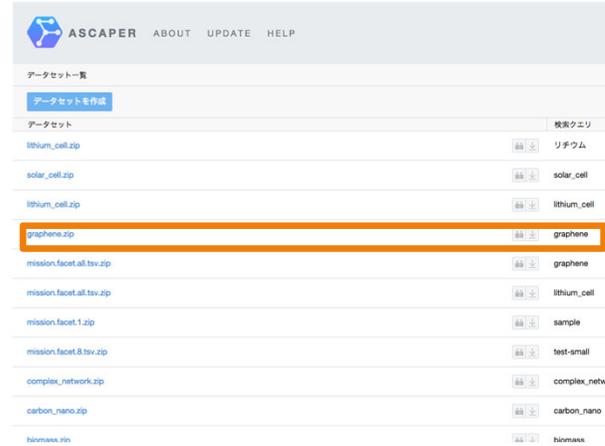
学術知識を構造化するウェブシステム

①システムへのログイン



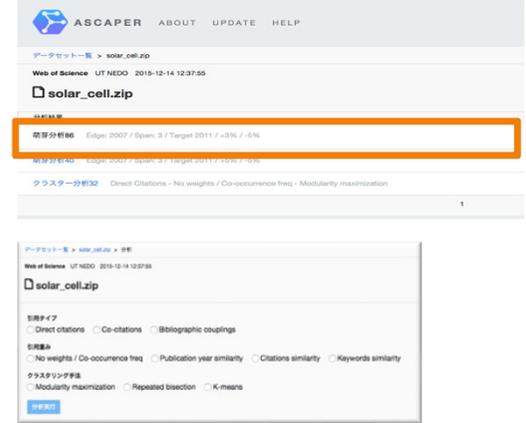
アカウント情報の入力

②データセットの作成/選択



書誌情報データのアップロード / データセットを選択

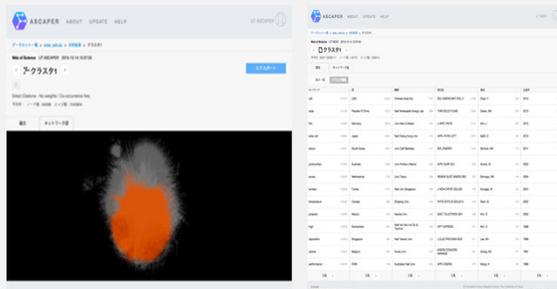
③分析の実行



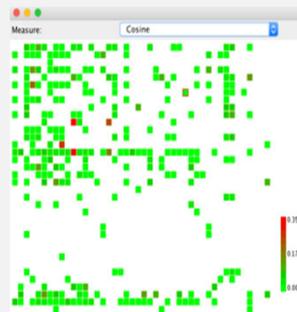
分析方法の選択
パラメータの入力

分析結果の表示

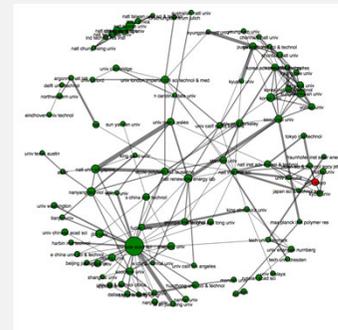
④俯瞰分析



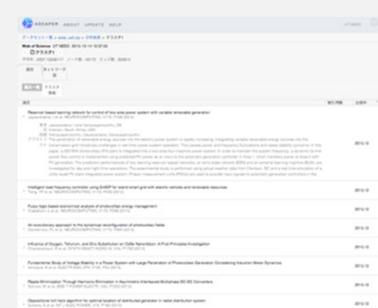
⑤ヒートマップ



⑥共著ネットワーク



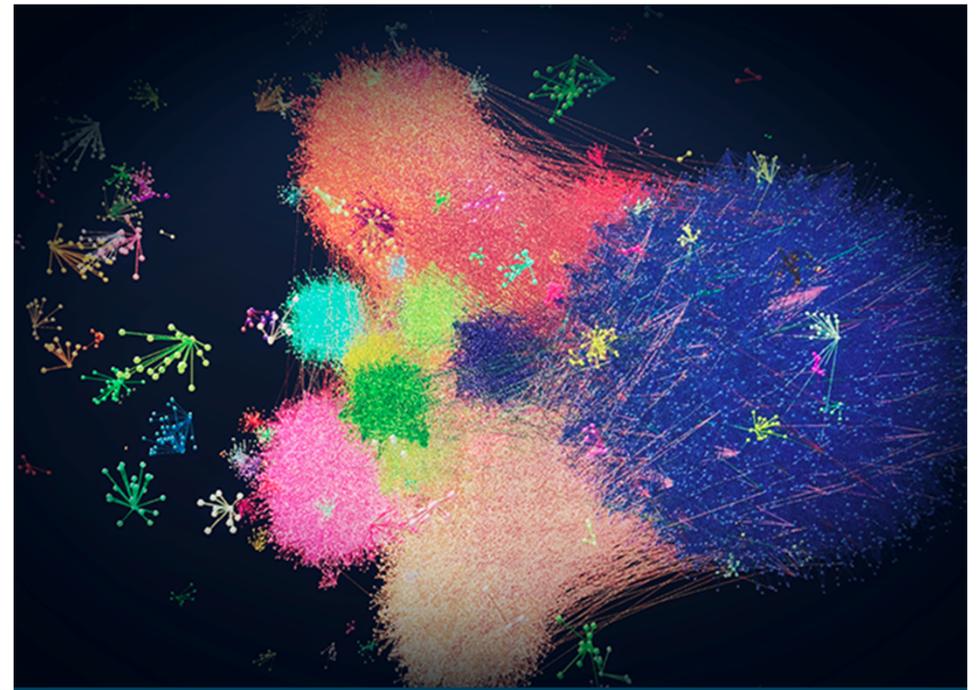
⑦萌芽論文特定



④ 学術知識の構造化（俯瞰分析）

2018年3月3日放送の「NHKスペシャル AIに聞いてみた どうすんのよ!? ニッポン 第2回 働き方」の中で、「働き方」に関する世界の学術研究を分析し、課題・論点を抽出するために、「学術俯瞰」技法が活用されている。
（同じ色が付いている論文群は、引用関係が密なグループ）

performance work-engagement
leadership creativity
job satisfaction work-environment
employee turnover innovation
burnout abusive supervision



<https://www.nhk.or.jp/special/askai/network.html>

④ 詳細な俯瞰分析(太陽電池)

- 3層クラスタリングによりエマージングな6分野を特定
- それぞれに対して、国内における大型の取り組むをマッピング

第1階層	第2階層	第3階層	論文数	平均出版年	国内論文数の 界ランク	論文数 第一位(国)
シリコン型	シリコン薄膜	パッシベーション	5051	2008.2	4	ドイツ
		ナノワイヤ	3843	2012		アメリカ
		アモ不ファス	3580	2004.4	1	日本
		ZNO	1437	2010.4		韓国
	システム	GA AS	3992	2008.8	3	アメリカ
		Module	3932	2009.7		アメリカ
		Power inverter	3345	2009.6		アメリカ
	CIGS	CdTe太陽電池	4855	1999		アメリカ
		Cu In太陽電池	3791	2008	3	ドイツ
		CZTS太陽電池	2303	2012	4	中国
色素増感型	TiO2	NanoTube	5348	2011.7	4	中国
		Quantum dot	3475	2012.2	5	中国
		ZnO太陽電池	3210	2011.6	5	中国
	Dye	TiO2太陽電池	4747	2010.2	3	中国
		electrolyte	2423	2010	3	中国
		perovskite	2018	2012.5	5	中国
	Dye Sensitized	3262	2011.5	2	中国	
ポリマー型	organic	organic	4087	2009.4	4	アメリカ
		polymer blend	3209	2012		アメリカ
		small molecule	1327	2011.9		中国
	film	Nanostructure	3076	2012.1	4	アメリカ
		hybrid-polymer	2568	2010.5		中国
		実装技術	2137	2012.3		中国
	polymer	fullerene	2381	2009.6	3	中国
		conjugated polymer	2074	2011.9	5	中国
		Copolymer / Organic	1291	2012.7	3	中国

※平均出版年2012年以降

ナノワイヤ太陽電池
文科省「革新的エネルギー
研究開発拠点形成事業」
“FUTURE-PV Innovation”

理論的には75%の変換効率
(東大先端研 岡田研究室)

ペロブスカイトを用いており、
容易に大型化可能
日本発(桐蔭横浜大学 宮坂)
NIMS, NEDOなどが取り組み

ナノ結晶
2005年にNEDOが着目
(NEDO海外レポート NO.966)
阪大(Prof. Fukuzumi)

実用化に向けたポリマー型
太陽電池の安定性向上

Π電子系骨格による高い変換効率
論文数4位(理化学研究所 瀧宮)

エアコンメーカーの特許分析

Daikin Patents (1980-2016)
Average age : 2006.4
Total # of patents 15,527
Node(patents) : 10,978
Edge(citation) : 198,359

C8 446 2008.8

adsorption
humidity
adsorption heat
adsorption heat exchanger
humidity control

C10 371 2008.4

scroll
chamber
piston
compression
cylinder

C1 808 2009.2

refrigerant
circuit
réfrigérant
heat exchanger
exchanger

C4 489 2006.9

composition
fluororubber
caoutchouc
elastomer
rubber

C7 457 2001.7

hydrogen fluoride
hfc
pentafluoropropane
hydrogen
fluoride

C3 555 2009.1

dust
indoor unit
air
indoor
coanda

C9 401 2002.4

polymer
fluorine
fluorine containing
layer
polyamide

C6 478 2005.9

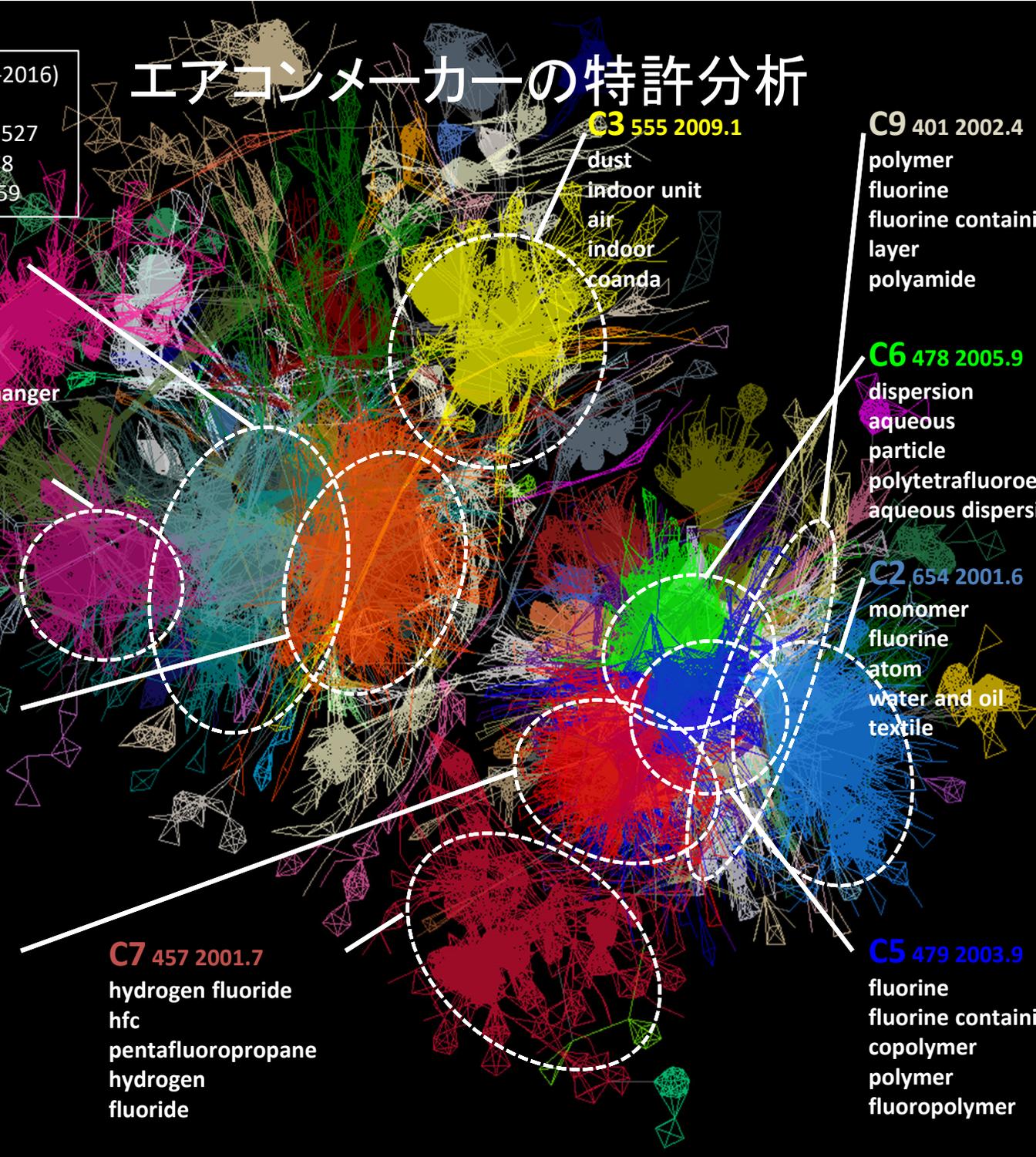
dispersion
aqueous
particle
polytetrafluoroethylene
aqueous dispersion

C2 654 2001.6

monomer
fluorine
atom
water and oil
textile

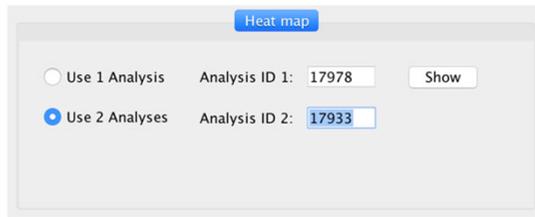
C5 479 2003.9

fluorine
fluorine containing
copolymer
polymer
fluoropolymer



⑤ヒートマップ分析(内容的な関連性)

ヒートマップ分析



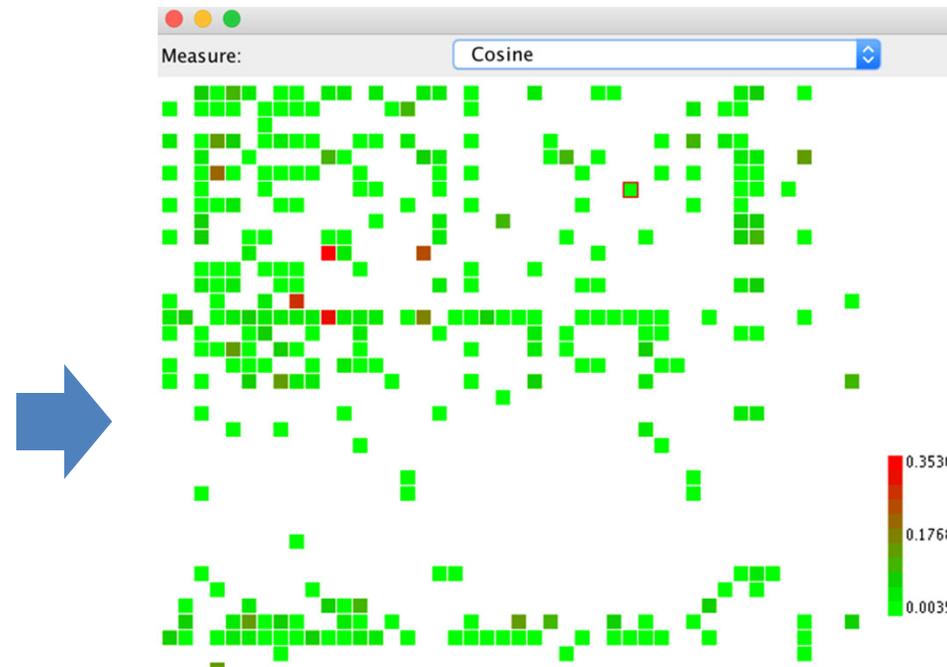
1. 比較したいデータセットのIDを入力
(同一データセット内での比較も可能)

2. showボタンを押す

例1)ソーラー電池のデータセットの各クラスタと、スマートシティの各クラスタを比較

例2)ソーラー電池のデータセットの各クラスタ同士を比較し、クラスタ間の関連性の強さを分析

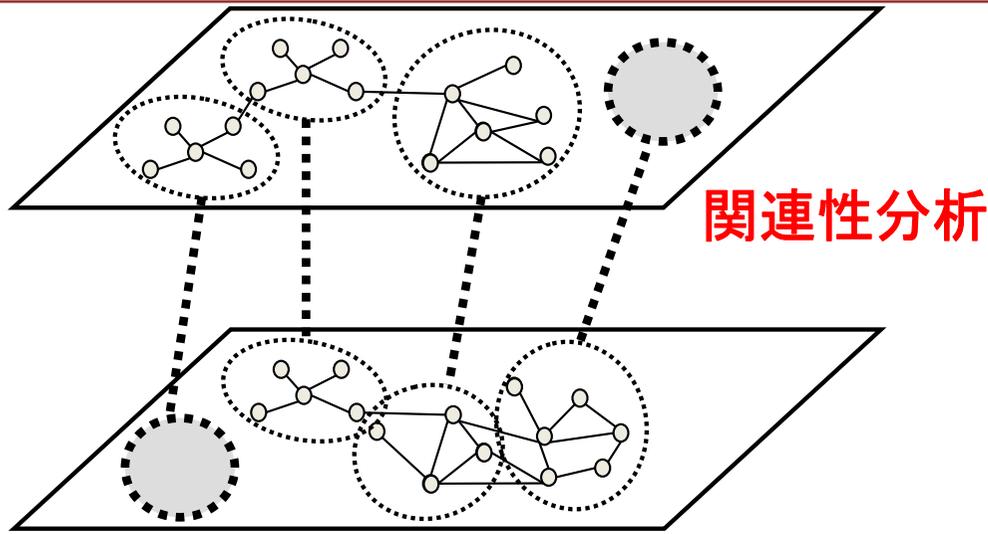
ヒートマップ分析結果



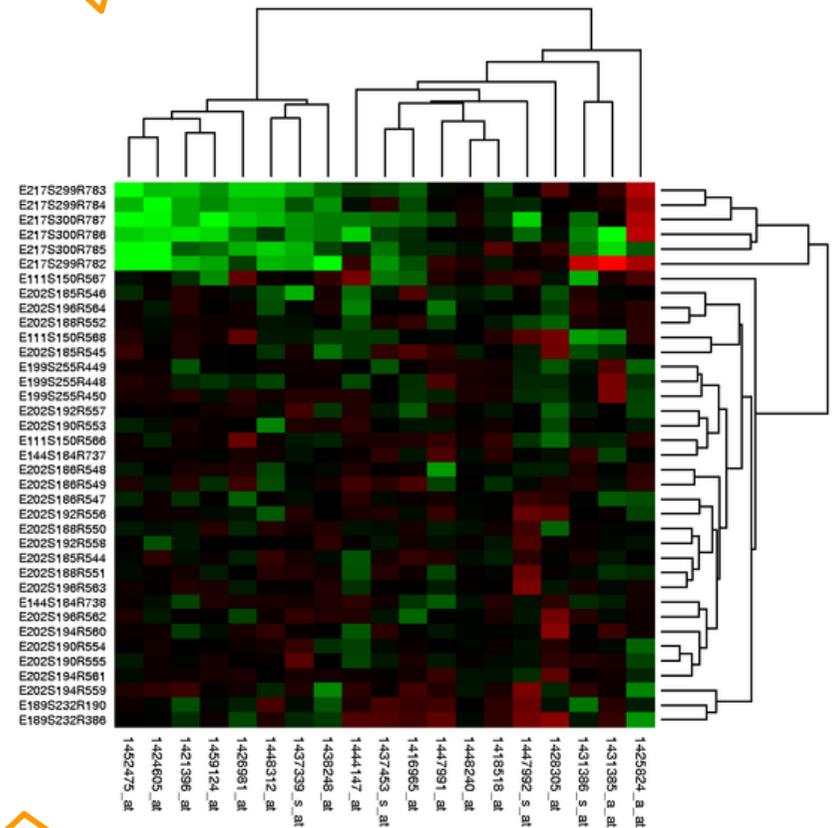
縦横にクラスタ番号、色が言語的関係性の強さを表す

関係の強いクラスタの抽出、将来的な技術の応用先の探索

⑤異なるレイヤーの知識を接続する手法



マトリックス表示



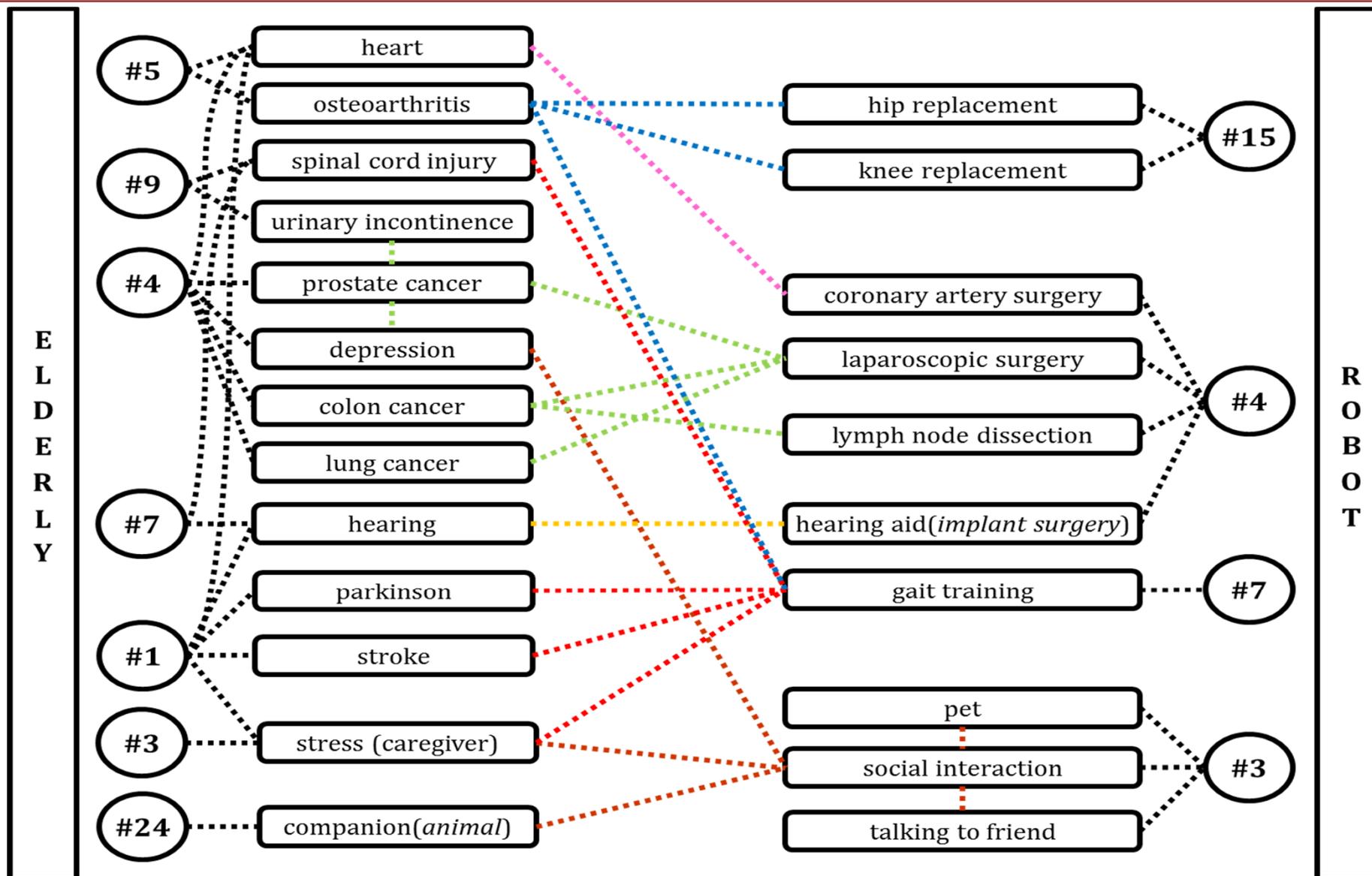
例) 特許 有り 無し

	有り	無し
論文 有り	A	B
無し	C	D

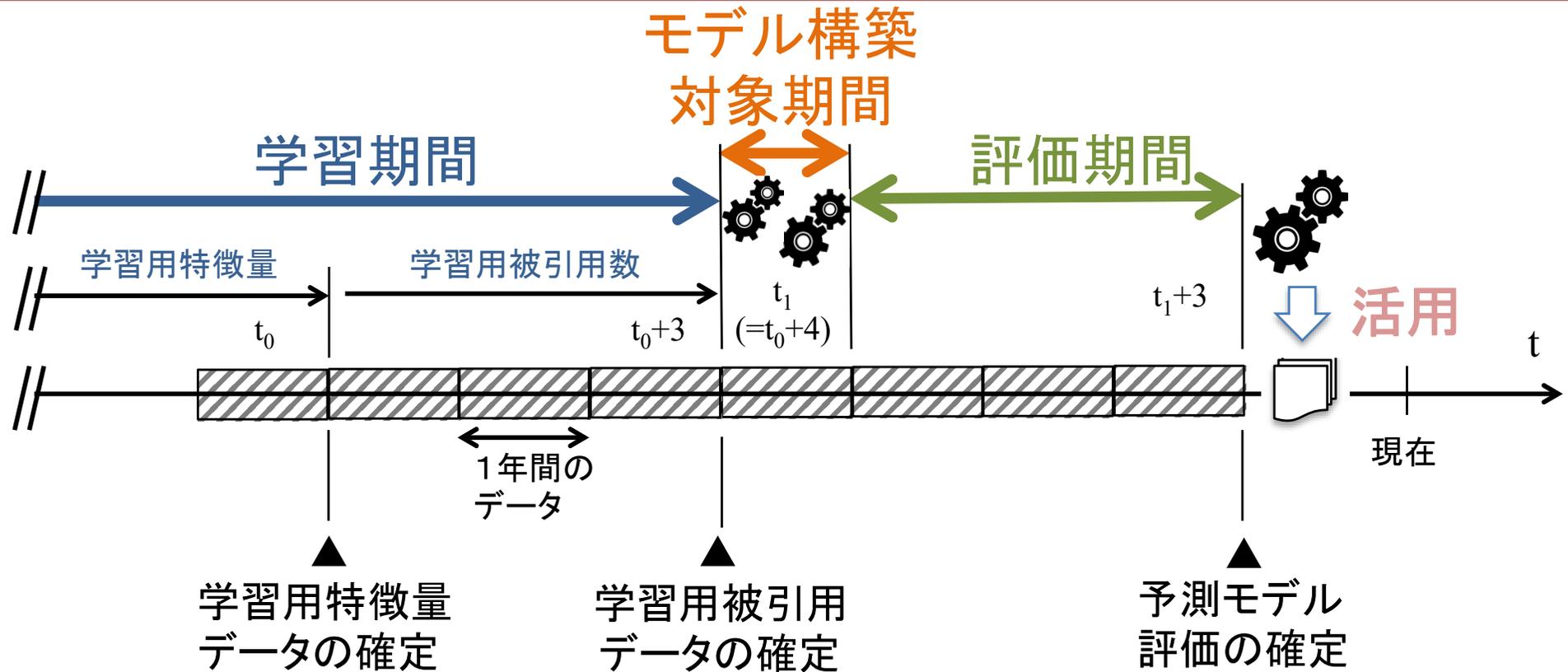


結果の解釈

⑤ 高齢化の課題とロボット技術の関係



⑦ 萌芽論文の予測モデル



「モデル構築対象期間」を中心にした前後のデータ
「過去のデータによる学習」と「3年後※の実際の被引用数による検証」
によってモデルを構築。 ※:3年は可変パラメータ。目的や対象分野によって適切な数字を設計する。

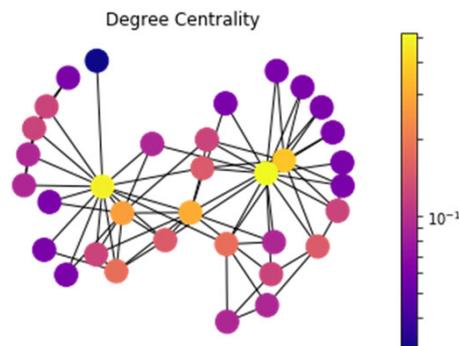
構築したモデル  を現在手元にある最新の論文データ  に活用

⑦ 萌芽予測の学習用特徴量

ノード(論文)の中心性、ノードが属するクラスタの情報、近隣ノードの情報を表す指標群を網羅的に計算し特徴量として学習に用いる

1. 中心性を表す指標群

ノードがネットワーク上の中心に位置しているかを示す



例: ページランク

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

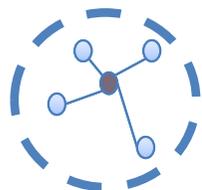
次数中心性
媒介中心性
近接中心性
固有ベクトル中心性
クラスタリング係数
HITS
ページランク

2. ノードが所属するクラスタの指標群

ノードが属しているクラスタの大きさや内部密集度や独立性を示す

全ノード数
全エッジ数
モジュラリティ
(独立性を表す)
クラスタリング係数
(エッジの密集度)

3. ノードの周囲のノードの特徴量の平均/最大/最小



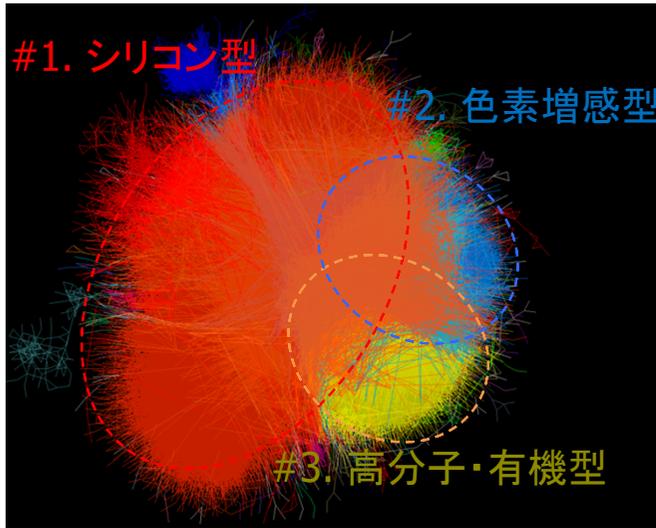
ノードの周辺の情報を集約し、ノードが属している近接分野の重要性を定式化



1,2,3合わせて63個の特徴量でノードの引用ネットワーク上での役割/位置を表現

ソーラーセル領域(2012年発表論文) 萌芽予測と評価結果(上位10件)

高効率かつ安定したポリマー型太陽光発電における界面材料



色素増感系太陽光発電における、光物性、電気化学特性

最も研究が盛んなクラスタ「#1シリコン型」から、「#2.色素増感型」や「#3高分子・有機型」へ注目が集まっていること示唆される。

クラスタ	論文名	雑誌	現被引用数
3	Recent advances in solution-processed interfacial materials for efficient and stable polymer solar cells	ENERGY & ENVIRONMENTAL SCIENCE	320
3	Morphology characterization in organic and hybrid solar cells	ENERGY & ENVIRONMENTAL SCIENCE	180
3	Recent progress and future aspects of organic solar cells	PROGRESS IN PHOTOVOLTAICS	68
3	Solid-state dye-sensitized and bulk heterojunction solar cells using TiO ₂ and ZnO nanostructures: recent progress and new concepts at the borderline	POLYMER INTERNATIONAL	50
3	Rational Design of High Performance Conjugated Polymers for Organic Solar Cells	MACROMOLECULES	584
2	Photophysical and Electrochemical Properties, and Molecular Structures of Organic Dyes for Dye-Sensitized Solar Cells	CHEMPHYSICHEM	104
3	Small Molecule Organic Semiconductors on the Move: Promises for Future Solar Energy Technology	ANGEWANDTE CHEMIE-INTERNATIONAL EDITION	529
2	Characterisation of electron transport and charge recombination using temporally resolved and frequency-domain techniques for dye-sensitized solar cells	INTERNATIONAL REVIEWS IN PHYSICAL CHEMISTRY	35
3	Tandem polymer solar cells featuring a spectrally matched low-bandgap polymer	NATURE PHOTONICS	865
2	The Electrochemistry of Nanostructured Titanium Dioxide Electrodes	CHEMPHYSICHEM	36

データセットにおける上位5%内論文(正例)

各対象分野における萌芽論文予測精度 (評価5セット [予測対象: '06~'10]の平均)

萌芽予測の予測結果

分野	適合率	再現率	F-scores	AUC
ナノカーボン	79.8%	88.8%	0.840 (0.021)	0.908 (0.018)
ソーラーフォト	84.5%	87.5%	0.859 (0.013)	0.919 (0.012)
ガリウムナイトライド*	75.3%	66.3%	0.706 (0.047)	0.784 (0.033)
複雑ネットワーク	73.6%	92.0%	0.818 (0.012)	0.888 (0.011)
ソーシャルシステム	76.4%	90.9%	0.828 (0.011)	0.892 (0.011)
リアルタイムロボット*	73.4%	78.9%	0.758 (0.034)	0.802 (0.041)
サイバーフィジカル	79.7%	75.6%	0.776 (0.026)	0.831 (0.018)

(括弧内は標準偏差)

● 各分野において平均で F-scores > 70, AUC > 0.7 を達成

萌芽論文の判定に寄与する特徴量

特徴量	寄与度
ネットワーク中心性(ページランク)	1.21
所属クラスタのノード(論文)数	0.99
雑誌名・著者名(「Science」など7件)	0.60~0.91
引用先論文の所属クラスタのノード(論文)数	0.58

非萌芽論文の除外に寄与する特徴量

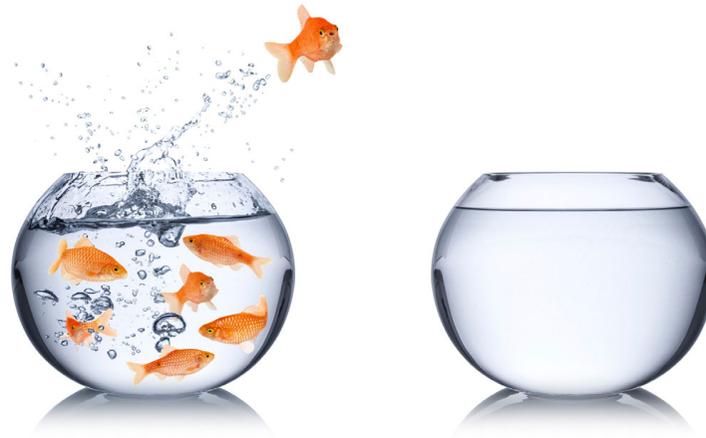
特徴量	寄与度
引用先論文の所属クラスタのモジュラリティ	-1.92
所属クラスタのモジュラリティ	-1.54
ネットワーク中心性(ネットワーク制約)	-0.71

(出典) 森純一郎, 原忠義, 梶川裕矢, 坂田一郎 (2016). 大規模論文データからの異種ネットワーク組み合わせによる萌芽論文の推定, 人工知能学会全国大会.

テクノロジー・インフォーマティクス

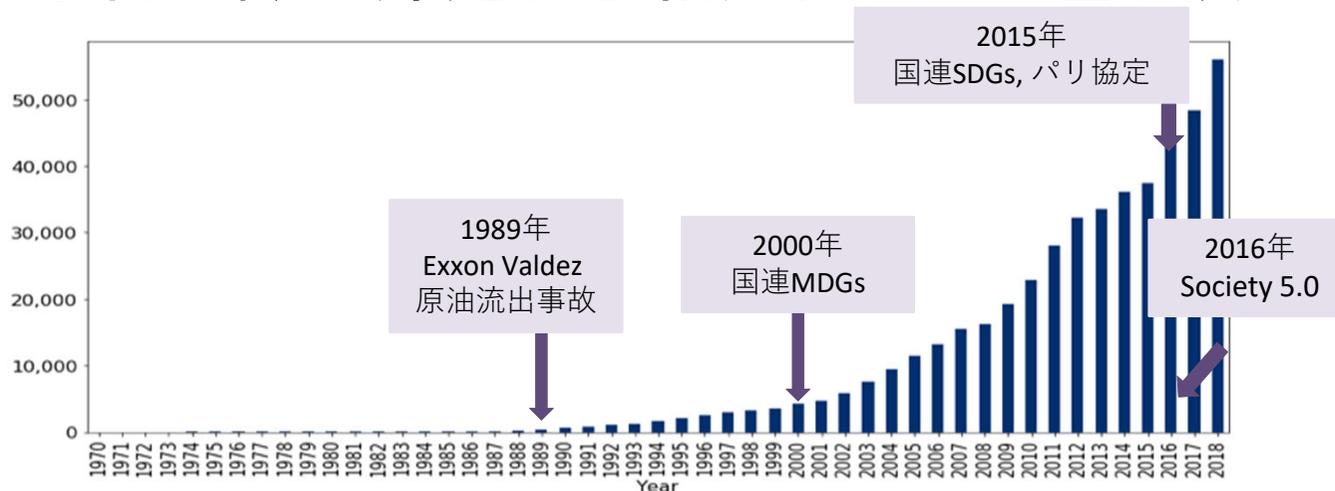
最新の開発課題：

ネットワーク科学と計算言語処理を組み合わせた人間の
意思決定を支援できる高精度な予測モデル等の開発



様々なタスクへの展開

• 学術知識の爆発及び細分化と知の埋没、分野融合



(左図) Scopusにおける sustainability関連の約50万件の論文の年次推移

⇒ 量的拡大と分野の細分化から専門家でも全体を把握することは困難

• これまでの取り組み

ホライゾン・スキャンニング支援



Ref, N. Shibata et al., TFSC vol.77, 2010
I. Sakata et al., TFSC vol. 80, 2013
(PMDAのプロジェクトに提供中)

Start-up Readiness の測定



Ref. T. Goji et al., Heliyon vol.6, 2020

技術コンバージェンス支援



Ref. V. Ittipanuvat et al, JETM vol.32, 2014
H. Nakamura et al., TFSC vol.94, 2015
T. Kose et al., TFSC vol.146, 2019

科学技術政策への知見提供



Ref. M. Sugiyama et al., Nature Vol.531, 2016
K. Asatani et al., Energies vol.13(4), 2020
H. Yamano et al., Scientific Reports 10, 2020

予測と開発戦略の策定支援

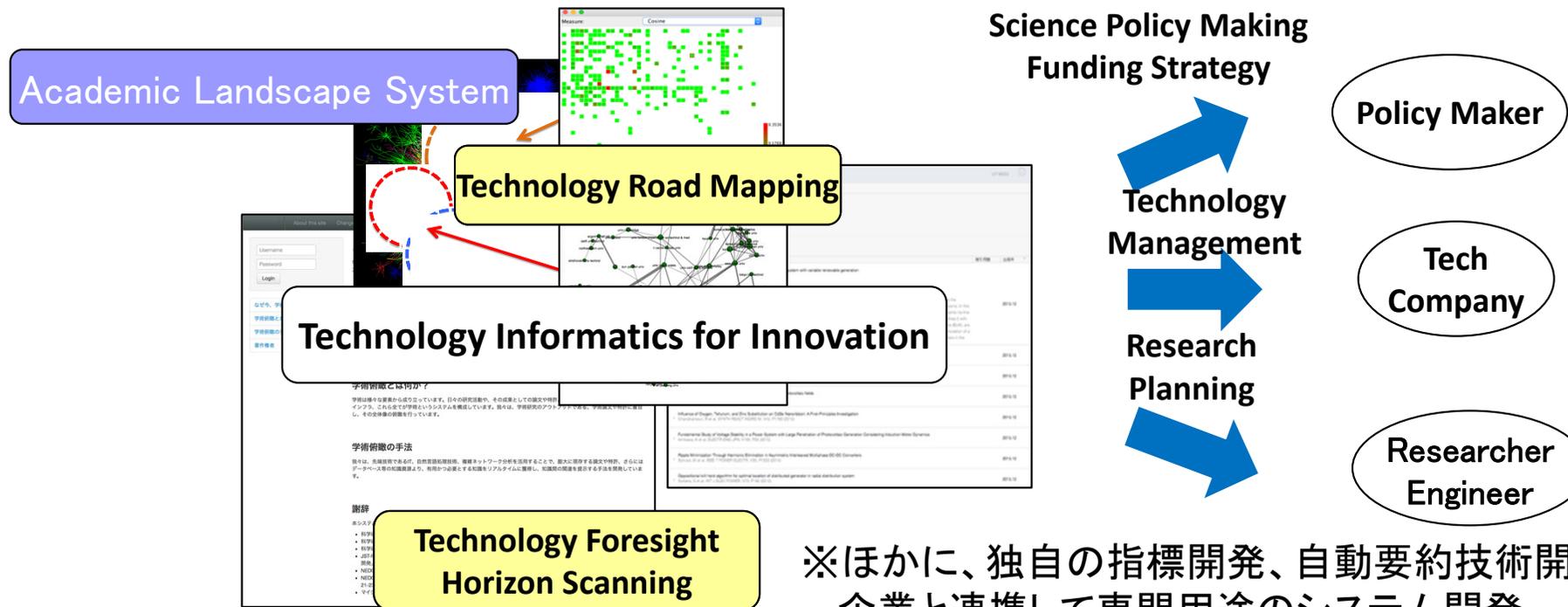
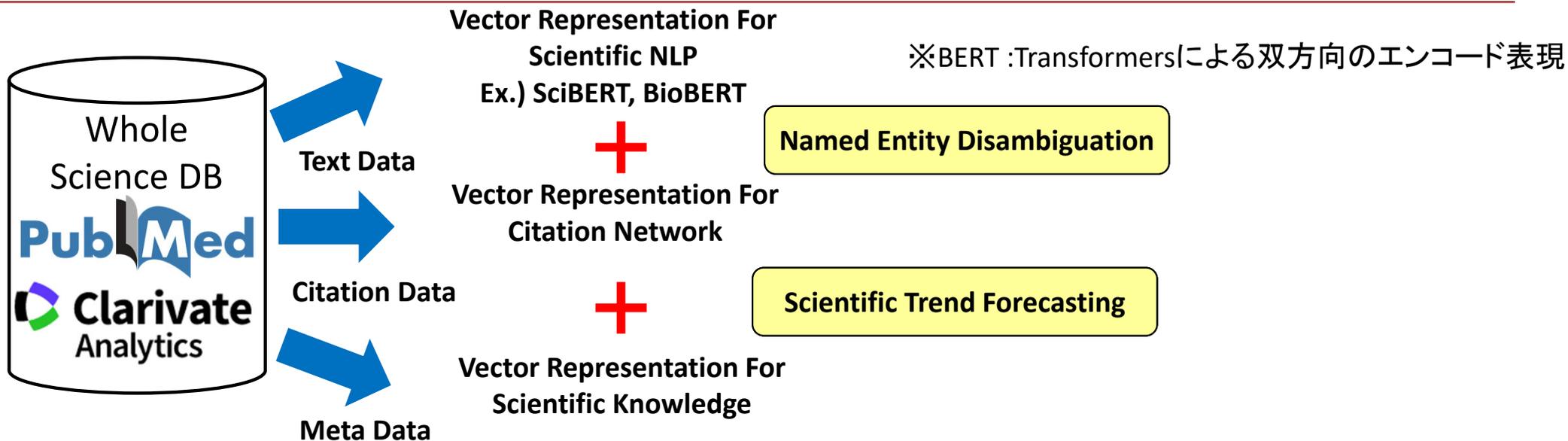


Ref, N. Shibata et al., TFSC vol.78, 2011
H. Sasaki et al., JSDEWES vol.4(4), 2016
K. Asatani et al., PLOS One vol.13, 2018
H.Sasaki et al., Technovation vol.100, 2021
(NEDO戦略センター等で試行実験中)

計算言語処理等の基礎研究

Ref. N. Shibata et al., JASIST vol.63, 2012
T. Isonuma et al., TAFL vol.6, 2021

展開中の研究(3種のデータ統合)

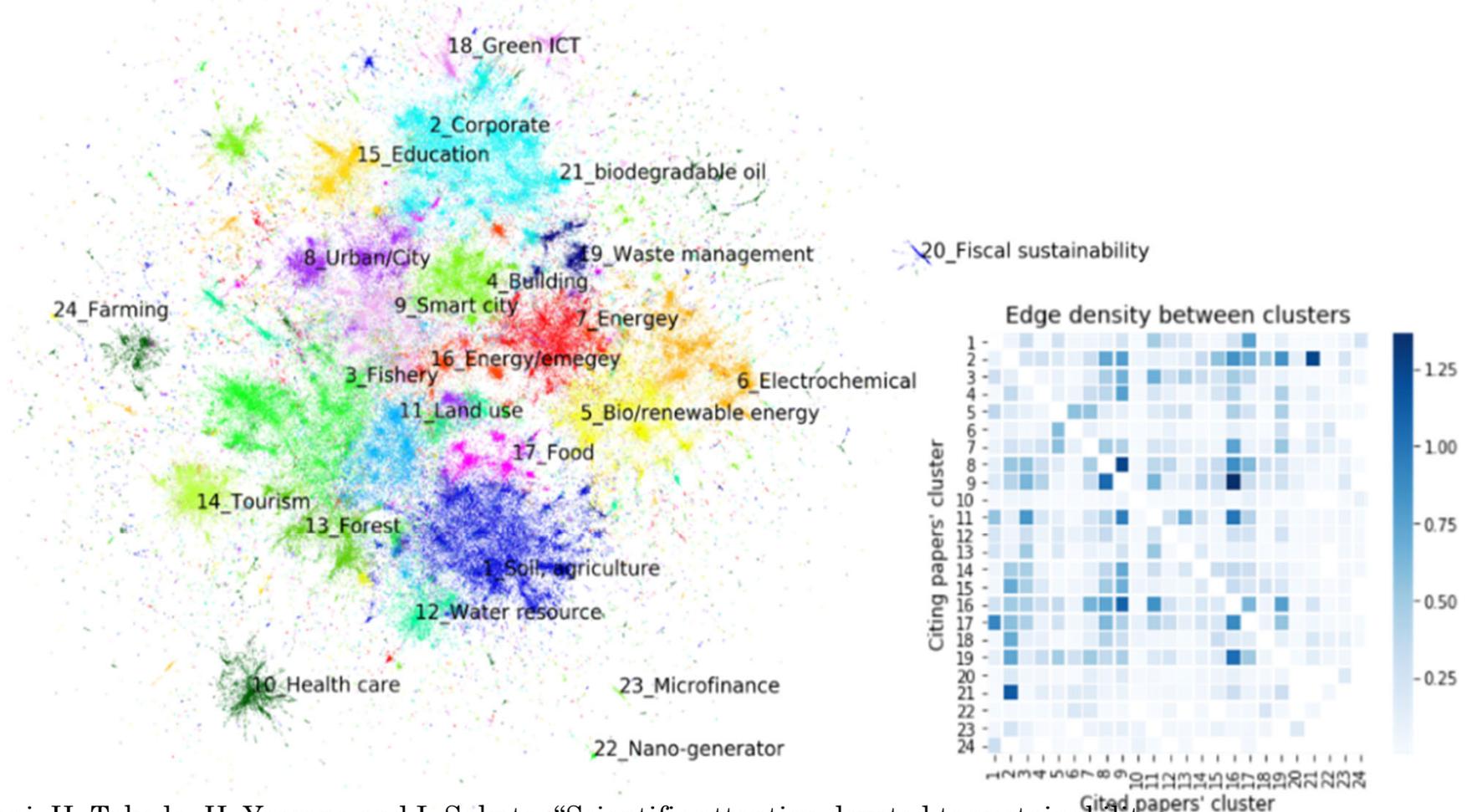


※ほかに、独自の指標開発、自動要約技術開発、企業と連携して専門用途のシステム開発

膨大な論文の山の中から知見(宝)を探す

サステナビリティに関する30万本の学術論文を引用関係を元に分類し、知識を構造化
手法: クラスタリング(グループ分け)や特徴語のコサイン類似度によりデータの意味を知る

👉 世界のサステナビリティに関する知識の構造化や内部の議論の把握



(出典) K. Asatani, H. Takeda, H. Yamano and I. Sakata, "Scientific attention devoted to sustainability and SDGs: Meta-analysis of academic papers", **Energies** vol.13(4) (2020),

サステナビリティの俯瞰

SDGsの17 Goalsと学術研究の関連性

UNの17のSDGs Goalsの文章と各クラスタ内の論文のタイトル・アブストラクトとの言語的類似度を観察することで、SDGs Goalに関する議論が活発な学術領域を抽出。

UN Sustainable Development Goals

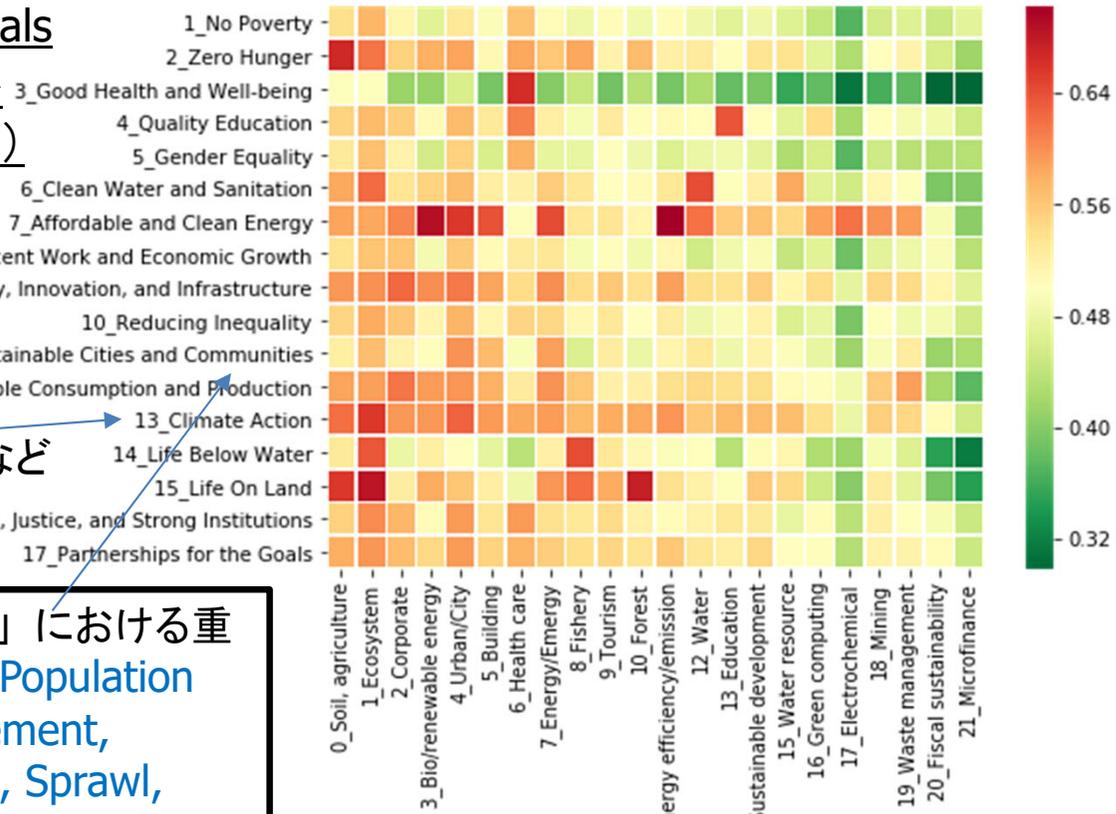


17のゴール
(国連文書)

Energy access, Biogas,
Electrification

Adaptive capacity など

Inclusivity, Total value など



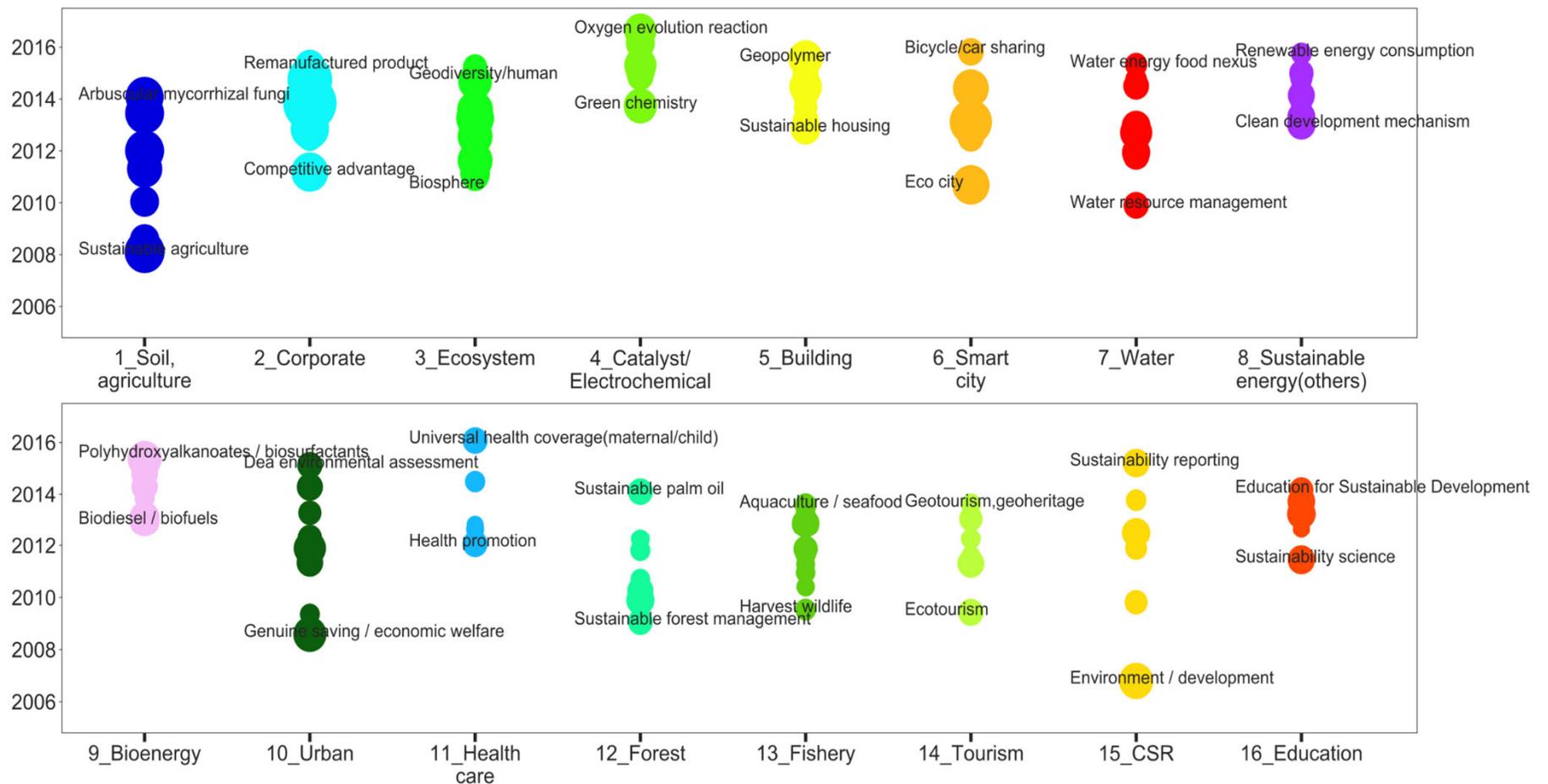
「11 Sustainable Cities and Communities」における重要学術領域: Sustainable urbanization, Population growth rate, Incubator, Informal settlement, Municipal solid, Disaster risk reduction, Sprawl, Affordable housing, Road safety, Sustainable City, Public transport, Waste management など

サステナに関する学術論文のグループ

各目標に関して学術界で集中的に論じられている単語を抽出し、世界の関心の最先端を把握⇒研究構想、応用先用途や規制リスク等の検討に活用

サステナビリティのホットスポット

サステナビリティ又はSDGsにアテンションしている学術論文30万件を内容別のクラスタに分割し、その中から、新しい 이슈を抽出。例えば、水・エネルギー・食料問題の交差点、ジオポリマー、ユニバーサルヘルスカバレッジ。

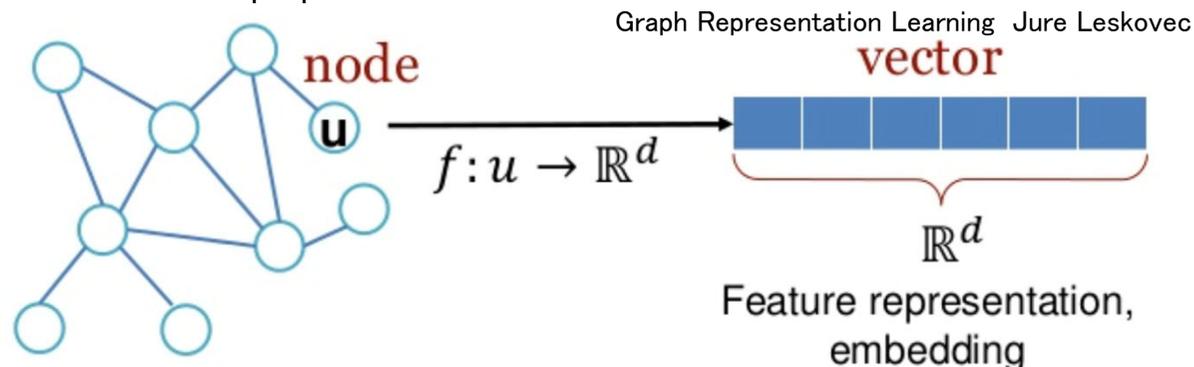


(出典) K. Asatani, H. Takeda, H. Yamano and I. Sakata, “Scientific attention devoted to sustainability and SDGs: Meta-analysis of academic papers”, **Energies** 13, no.4 :1-21. (2020)

ネットワークの表現学習の導入 (Network representation Learning)

ネットワーク構造よりノードの潜在的な表現を学習

Representation learning of large scale graphs in realistic time is proposed in DeepWalk (2014), and many subsequent methods have been proposed.



Network (sparse adjacency matrix)
コンピュータで処理するのが難しい

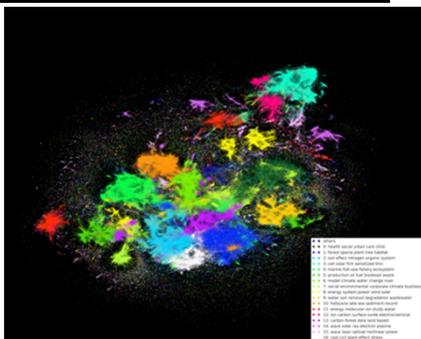
Vector representation

既存の機械学習のアルゴリズムにインプット可能



ネットワーク表現学習によるタスク

Large graph visualization



2.5 million papers by Largivis (2016)

クラスタリング・コミュニティ抽出

・ノードの分類やコミュニティをネットワーク情報のみから抽出

ラベル推定、教師あり学習

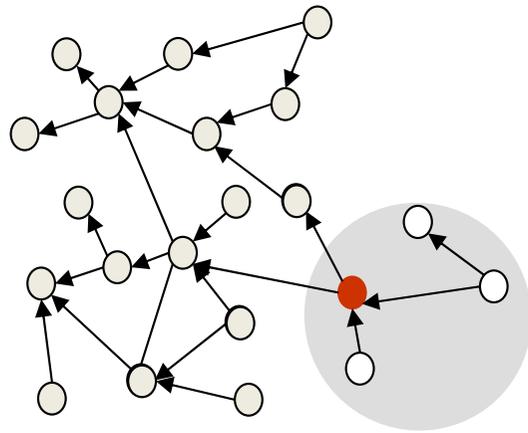
・ノードのラベルやノードの性質などに関して、一部のノードの正解データより全体を学習

リンク推定

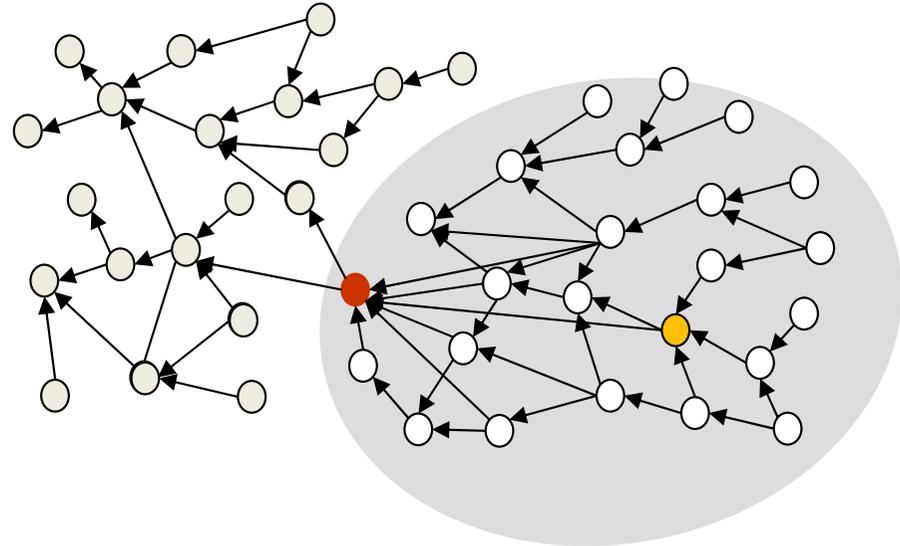
・将来的にもしくは潜在的に存在するリンクを推定

「時間」の情報を潜在的に含むネットワーク

Modeling of past cases
citation network at $t=T$



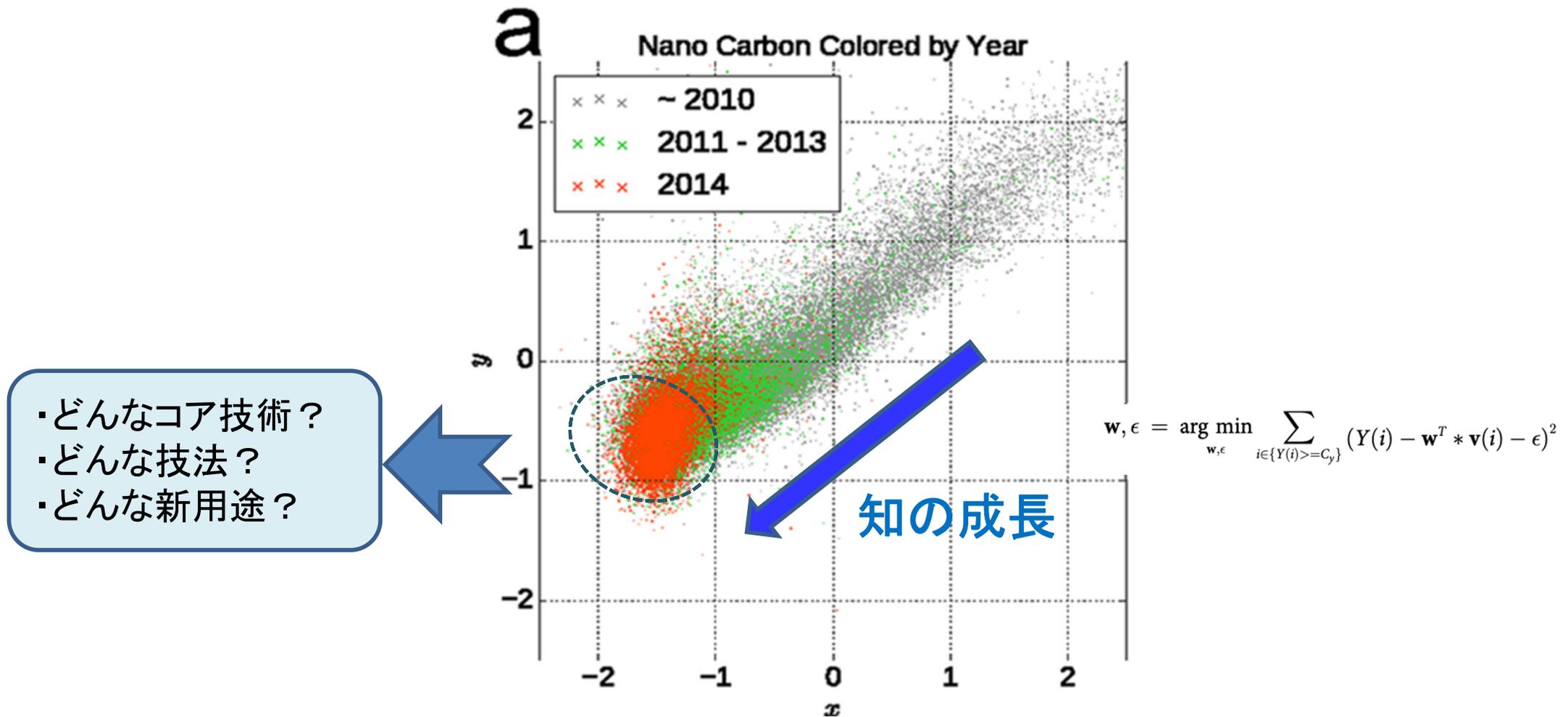
-> Prediction of the current case
citation network at $t=T+M$



“If you want to know a certain person, see his friends !”

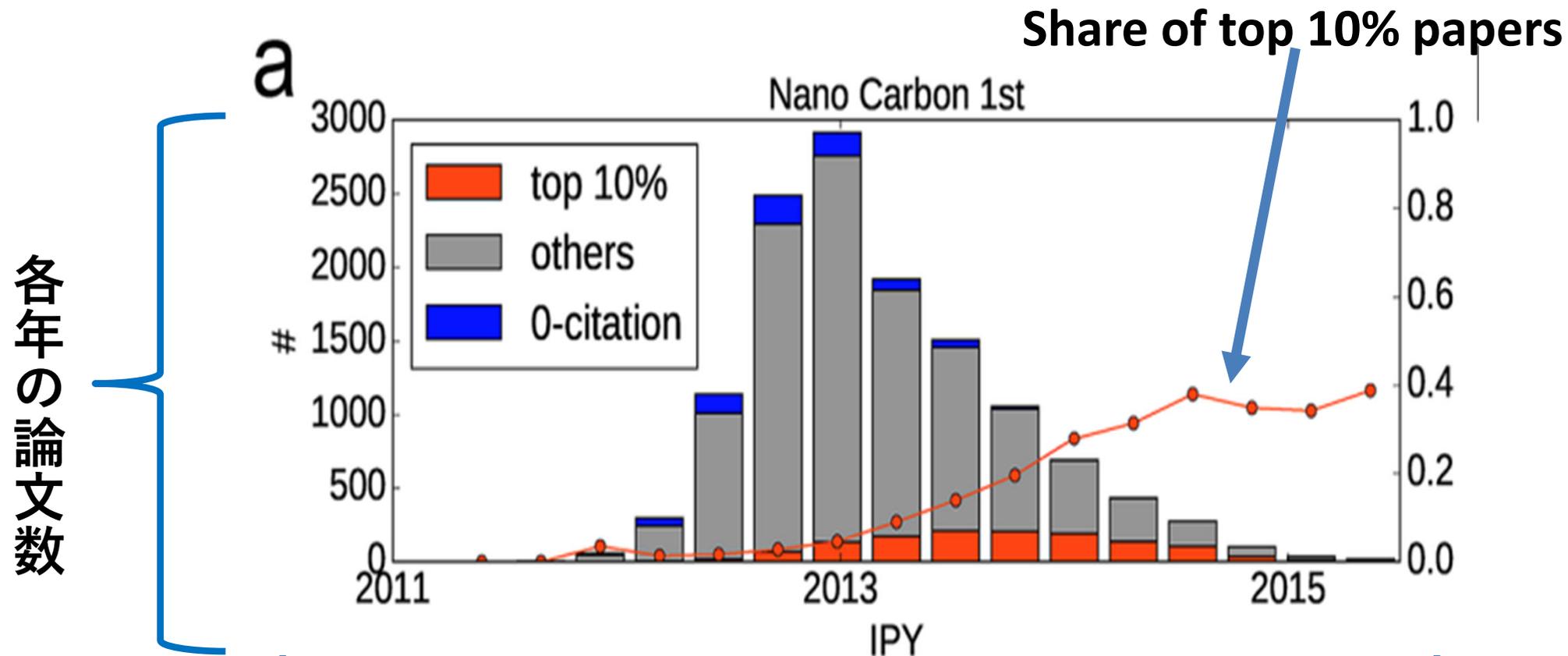
“We can estimate the nature of an element from the surrounding relationship !”

「知」の成長の方向性を知る



(出典)K. Asatani, J. Mori, M. Ochi, I. Sakata, "Detecting trends in academic research from a citation network using network representation learning" PLoS ONE 13(5) e0197260. (2018)

仮想的な出版年と論文引用の関係



各年の論文数

$$IPY(i) = \underline{\mathbf{w}}^T * \underline{\mathbf{v}}(i) + \epsilon$$

Growth Direction

[Degree of trend-following]

トレンドの先端にあるペーパーとトピック

Data: Paper of Nanocarbon published in October 2013 – December 2013

Predicted cutting edge paper and future citations

TITLE	JOURNAL	AUTHOR	NR	IPY
Nanocarbon Networks for Advanced Rechargeable Lithium Batteries	ACCOUNTS OF CHEMICAL RESEARCH	[Xin, S, Guo, YG, Wan, LJ]	299	2014.65
Semiconducting Monolayer Materials as a Tunable Platform for Excitonic Solar Cells	ACS NANO	[Bernardi, M, Palummo, M, Grossman, JC]	26	2014.33
Direct Optical Imaging of Graphene In Vitro by Nonlinear Femtosecond Laser Spectral Reshaping	NANO LETTERS	[Li, BL, Cheng, YW, Liu, J, Yi, CW, Brown, AS,	10	2014.27
Flexible graphene-based lithium ion batteries with ultrafast charge and discharge rates	PROCEEDINGS OF THE NATIONAL ACADEMY OF	[Li, N, Chen, ZP, Ren, WC, Li, F, Cheng,	315	2014.17
Crumpled Nitrogen-Doped Graphene Nanosheets with Ultrahigh Pore Volume for High-Performance Supercapacitor	ADVANCED MATERIALS	[Wen, ZH, Wang, XC, Mao, S, Bo, Z, Kim,	408	2014.17
Fingerprinting photoluminescence of functional groups in graphene oxide	JOURNAL OF MATERIALS CHEMISTRY	[Li, M, Cushing, SK, Zhou, XJ, Guo, SW, Wu,	72	2014.1
Quaternary Nanocomposites Consisting of Graphene, Fe ₃ O ₄ @Fe Core@Shell, and ZnO Nanoparticles: Synthesis and Excellent Electromagnetic Absorption Properties	ACS APPLIED MATERIALS & INTERFACES	[Ren, YL, Wu, HY, Lu, MM, Chen, YJ, Zhu,	150	2014.09
A high performance PZT ribbon-based nanogenerator using graphene transparent electrodes	ENERGY & ENVIRONMENTAL SCIENCE	[Kwon, J, Seung, W, Sharma, BK,	61	2014.03
Covalent Hybridization of Thiolated Graphene Sheet and Platinum Nanoparticles for Electrocatalytic Oxygen Reduction Reaction	JOURNAL OF NANOSCIENCE AND NANOTECHNOLOGY	[Ahmed, MS, Kim, D, Han, HS, Jeong, H,	32	2014.02
Nitrogen Doped Graphene-Rich Catalysts Derived from Heteroatom Polymers for Oxygen Reduction in Nonaqueous Lithium-O ₂ Battery Cathodes	ACS NANO	[Wu, G, Mack, NH, Gao, W, Ma, SG, Zhong,	236	2013.89
One-pot, solvothermal synthesis of TiO ₂ -graphene composite nanosheets	JOURNAL OF COLLOID AND INTERFACE SCIENCE	[Yang, WS, Zou, XX, Xu, FG, Wang, XD,	15	2013.89
Synthesis of porous Ni@rGO nanocomposite and its synergistic effect on hydrogen sorption properties of MgH ₂	JOURNAL OF MATERIALS CHEMISTRY	[Liu, G, Wang, YJ, Qiu, FY, Li, L, Jiao, LF, Yuan, HT]	30	2013.89
Renewing Functionalized Graphene as Electrodes for High-Performance Supercapacitors	ADVANCED MATERIALS	[Zhang, T, Luo, B, Jia, YY, Li, XL, Wang, B, Song, Q, Kang,	152	2013.88
Interfacial polymerized polyaniline/graphite oxide nanocomposites toward electrochemical energy storage	POLYMER	[Zhang, YH, Chen, MJ, Qu, HL, Zhang, X, Wei, HG, Luo, ZP,	88	2013.85
Graphene with Electrochemical Ionic-Liquid Assistance Synthesis to Explore the Metal Adsorption Ability and the High Electrical Conductivity in Graphene/Epoxy Composite	JAPANESE JOURNAL OF APPLIED PHYSICS	[Chang, CF, Chiang, YF, Chen, JR]	0	2013.82

Material and technology:

Nanocarbon network
 Graphene nano sheet
 Fingerprinting photoluminescence
 Graphene electrodes
 Graphene-Rich catalysts
 graphite oxide nanocomposites

Target and application

Lithium-ion batteries
 Supercapacitor
 Electromagnetic absorption
 Lithium-O₂ Battery Cathodes
 Energy storage

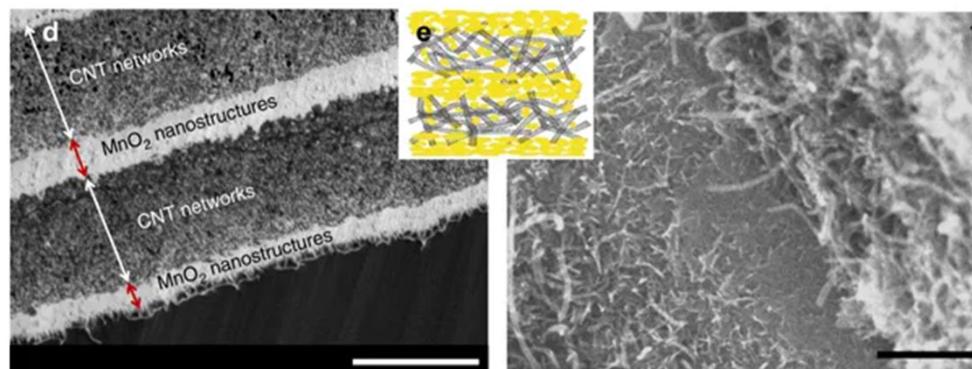
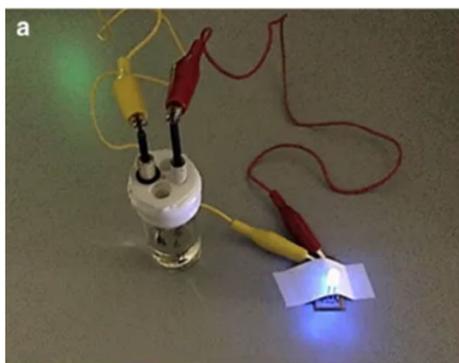
2013年に予測されていたトレンドと友人との研究

W. Gong, B. Fugetsu, Z. Wang ,I. Sakata, Mingda Li et al., “Carbon nanotubes and manganese oxide hybrid nanostructures as high performance fiber supercapacitors”, *Communications Chemistry* 1 (2018): s42004-018-0017-Z

W. Gong, B. Fugetsu, Z. Wang ,I. Sakata, Mingda Li et al., “Thicker carbon-nanotube/manganese-oxide hybridized nanostructure as electrodes for fiber-shaped high-energy-density supercapacitors”, *Carbon* (2019): vol.154, pp.169-177.



“Nanocarbon (CNT) Network” “Supercapacitor”



被引用数上位論文は予測できるか(内容or引用)?

- 目的: 高いインパクトを持つ研究(萌芽的研究)の早期発見, 文献データの持つ多様な情報を直接扱えるモデルへの期待.
- 課題: ディープラーニング技術を用いて科学研究のインパクト指標を予測する可能性について十分に検討されていない.
 - 文献データの「言語情報(文章)」と「引用情報」のどちらの予測力が高いのか?
- 解決策: 言語情報と引用情報を用いたモデルでのTop Tier論文の特定可能性分析を行う

データ

Scopus “Solar Cell”に関連する論文データのうち2006~2009に出版された論文57,935件を対象,

検証

2010~2012に100回以上引用された2009年出版の論文66件(Top Tiers)の分布の偏りを分析

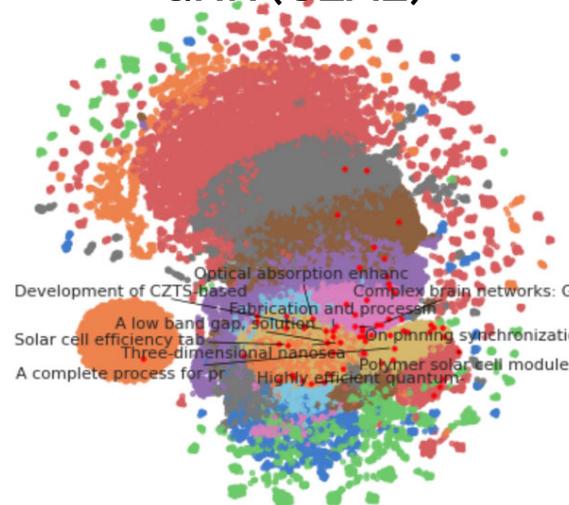
結論

引用情報を用いた方が”Solar Cell”に関しては特定可能性が高いとの結果

(出典)大知正直ほか,人工知能学会第35回全国大会(2021)

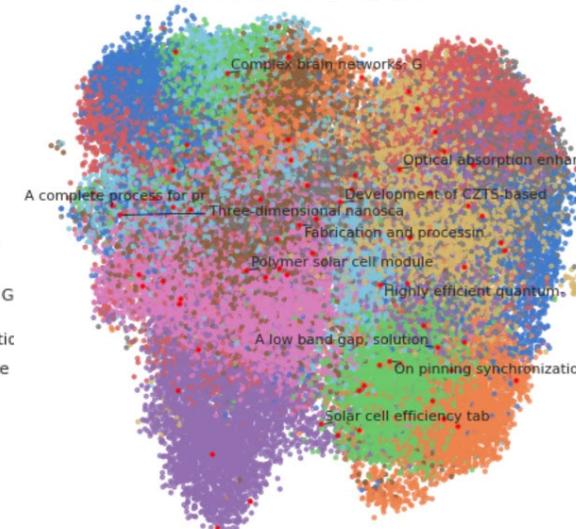
M.Ochi et al., WEBIST 2021

GNN(SEAL)



Entropy=1.742
引用情報による結果

Sentence-BERT



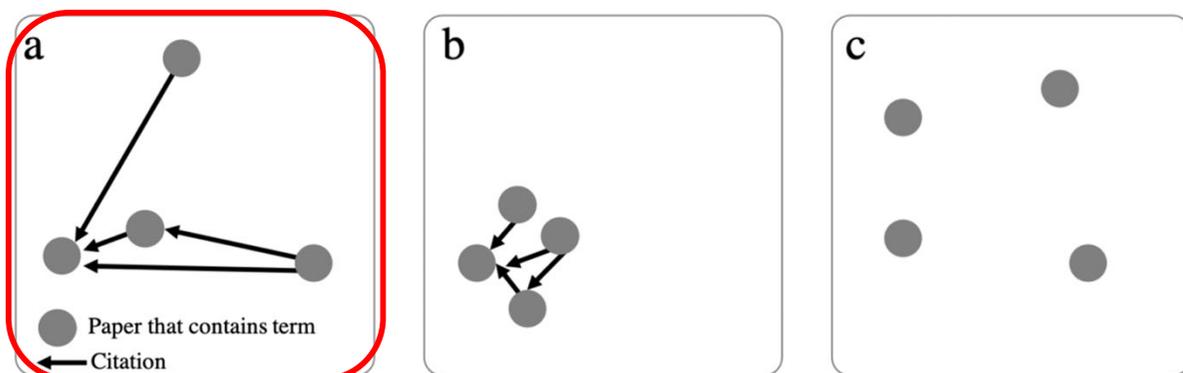
Entropy=2.900
言語情報による結果

※ 図中の文字はTopTier論文群を示しており, 引用情報の方が狭い範囲に集中していることが一見してわかる

学術コミュニティ内の重要な概念の抽出

引用ネットワーク上で伝わる単語をミームとして検出・スコア化することで、医学系論文からノーベル賞関連単語やWikipedia掲載単語などの学術界や社会にとって幅広く重要な単語を検出。

引用ネットワークを伝わる単語の累計



Meme score: high
Diffusion meme score(proposed): high

Meme score: high
Diffusion meme score(proposed): low

Meme score: low
Diffusion meme score(proposed): low

拡散ミームスコア

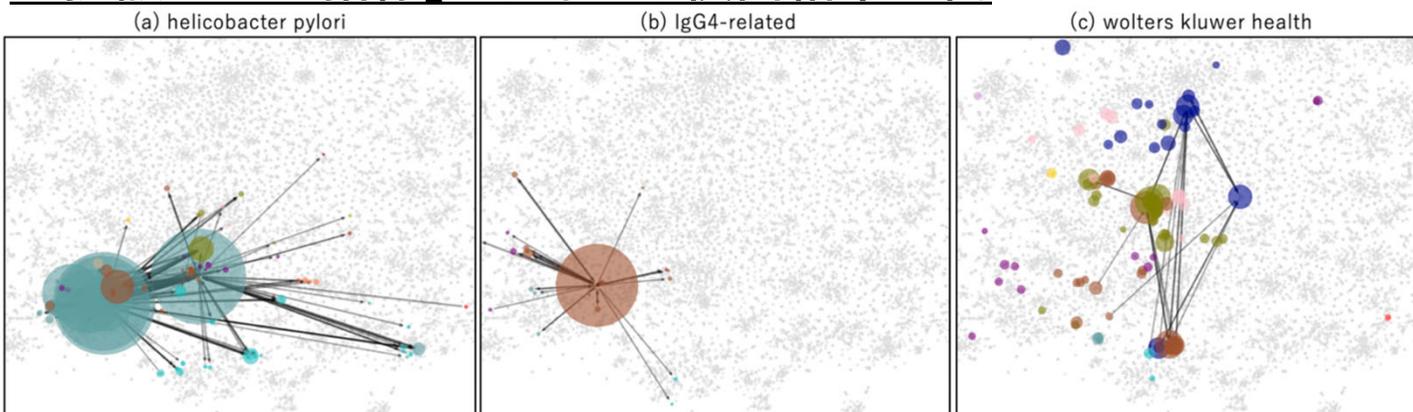
$$D = \ln\left(\sum_{e_s, e_t} d(e_s, e_t) + 1\right)$$

$$d(e_s, e_t) = \cos(\vec{e}_s, \vec{e}_t)$$

s, t 引用関係がある単語を含む論文のペア
 e_s, e_t : それぞれの所属クラスター
 d : ネットワーク表現学習で写像した空間上での距離

(a)のような幅広く伝わる単語を広く拡散したミームとして捉える

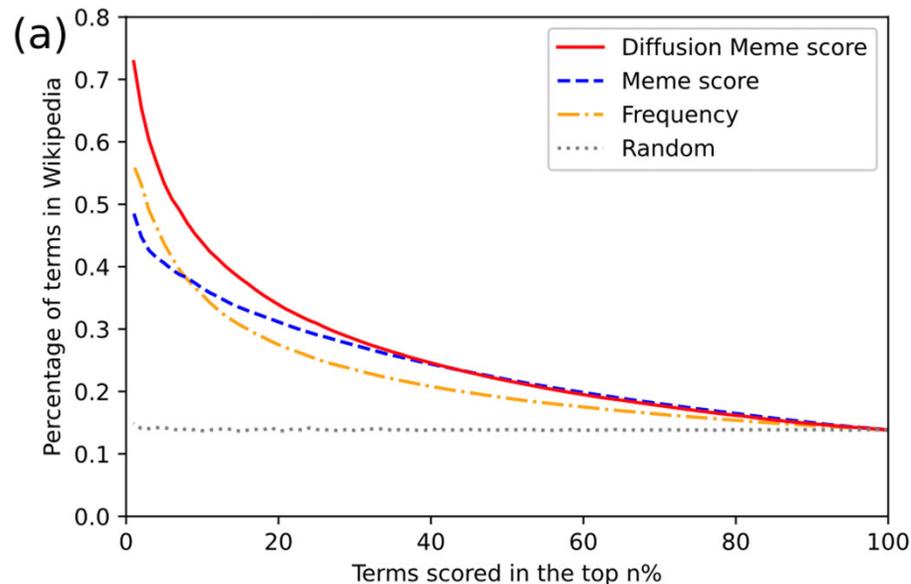
「拡散ミーム指標」を用いた検出結果の例



- ・Helicobacter Pylori(2005年ノーベル賞)などの重要概念を高く評価
- ・一方で、マイナーな疾患や出版社名などは高く評価しない

拡散ミームスコアを用いた抽出結果

各手法のトップ単語のWikipedia掲載率



- ・拡散ミームスコアが高い単語が他の手法に比べてWikipedia掲載率が高い
- ・その一方で疾患辞書の単語の予測精度は既存のミームスコアのほうが上(図省略)

拡散ミームスコア上位の単語

top terms rated high in D	D
helicobacter pylori+	10.870
biomed central ltd	10.576
elsevier ireland	10.554
hepatitis c virus+	10.530
toll-like+	10.509
bevacizumab	10.424
vascular endothelial growth factor*	10.344
IL-12	10.248
micrornas*	10.155
hiv-1+	10.061
chemokine	10.039
cox-2	10.016
nitric oxide synthase+	10.006
leptin*	9.903
infliximab	9.902
rituximab	9.901
adiponectin	9.802
bcl-2	9.671
imatinib*	9.662
foxp3*	9.657

+がノーベル賞関連単語

*がClarivate Citation Laureates

Top15単語はノーベル賞関連単語を数多く含む

単語の引用ネットワーク上の伝播距離の指標として定義し、ノーベル賞関連単語やWikipedia掲載単語などの学術界や社会にとって幅広く重要な単語を特定。「拡散距離」の合計という考え方は他のデータセットにも適用可能。

謝辞

本発表には、下記のプロジェクトによる研究成果を含みます。
ご支援に感謝致します。

NEDO/産業技術総合研究所 委託事業

「ネットワーク分析と言語処理の融合による大規模文献データからの技術の未来予測
プラットフォームの研究開発」

NEDO/産業技術総合研究所 委託事業

「事前学習言語モデルとネットワークモデルの融合に基づく科学技術トレンド予測」