

Quantifying cross-disciplinary knowledge flow from the perspective of content: Introducing an approach based on knowledge memes

コンテンツの観点からの学際的な知識の流れの定量化：知識ミームに基づくアプローチの紹介

<https://www.sciencedirect.com/science/article/pii/S1751157720301516>

Highlight

- 特別なタイプの知識ユニットである知識ミームが識別されます。
- 拡散カスケードは、知識ミームの拡散プロセスをモデル化するために適用されます。
- 知識ミームによって測定される分野間の知識関係は、引用によって測定されるものとは異なります。
- 優先的なアタッチメントは、学際的な知識のミームの拡散に効果を発揮します。
- 学際的な知識ミームのパターンが開示されています。

概要

分野間の知識の流れは、通常、出版物間の引用を通じて測定されます。本研究では、特殊な知識ユニットである知識ミームを導入することにより、コンテンツの斬新な視点から学際的な知識の拡散を定量化します。拡散カスケードは、知識ミームの拡散プロセスをモデル化するために提案されています。医療情報学（MI）を模範的な学際的分野として採用することにより、それと4つの関連分野との間の知識関係を測定します。学際的なミームの拡散パターンは、拡散カスケードのネットワーク構造を分析することによっても識別されます。結果は、引用によって測定されたものとは異なる、知識ミームによって測定された分野間の知識関係を示しています。優先的なアタッチメントは、学際的な知識ミームの拡散に効果があることが示されています。さらに、学際的な知識のミームは、一般的に、MIのミームよりも早く発生し、影響が大きくなります。この研究は、分野間の知識関係を定量化するための新しいアプローチへの洞察を提供し、測定可能な知識ユニットを通じてコンテンツ拡散メカニズムの理解を促進します。

イントロダクション

知識の普及は、参加者が車輪の再発明を回避し、元の作業に集中するのに役立つため、産業界と学界の両方で重要な役割を果たします。それは非常に遍在しているので、複数の表現を持つ人間の知識は、多様な放送メディアを介してさまざまなエージェント間で普遍的に転送されます。産業界（Rogers、2010年）、学界（Liu、Shan、およびYu、2011年、Rinia、Van Leeuwen、Bruins、Van Vuren、およびVan Raan、2002年、Yan、2016年）、およびそれらの間（Narin、Hamilton、およびOlivastro、1997；Sun&Ding、2018）。論文や特許間の引用を代理として扱い、これまでの研究では、出版、ジャーナル、分野など、さまざまな集約レベルで知識の流れを調査してきました。ただし、引用の背後にある動機が異なる可能性があるため、すべての引用が知識キャリアとして機能するわけではありません（Brooks、1986）。したがって、引用を通じて伝達された知識の実際の内容を分析することが不可欠です。これにより、知識の拡散をよりきめ細かく、より正確に測定できます。この目的のために、この研究では、知識ミームと呼ばれる特殊なタイプの知識ユニットを導入して、学際的な知識の流れの拡散パターンを定量化します。

もともと、ミームはコミュニティや文化に広がる最も単純な文化的単位として定義されており、継承の機能が似ているため、生物学の遺伝子の文化的類似物として扱うことができます（Dawkins、1976）。学術的な文脈では、知識ミームは、引用された論文から論文を引用することによって継承および複製される単語やフ

レーズのような短いテキスト単位として具体化されます。たとえば、「VOSViewer」、「CiteSpace」などの単語やフレーズによって運ばれる科学マッピングの知識は、図書館学や情報科学からさまざまな分野（Lou、Zhang、Li、He、2020）。最近、Kuhn、Perc、and Helbing（2014）は、統計的特性に基づいて科学出版物から論文を引用することによって継承された知識ミームを識別する方法を提案しました。知識ミームに基づいて、予備調査では、分野内（Liang、Mao、Cao、およびLi、2019）および科学と技術の間（Sun & Ding、2018）の知識普及のパターンを調査しました。比較として、引用ベースのアプローチは、学術的知識の普及を調査する上で依然として支配的です。具体的には、他の分野のさまざまな理論、概念、方法を統合する学際的研究の特徴を特定するために、分野間の知識の関係が広く研究されてきました（Stember、1991、Wagner、2011）。学術出版物間の引用は、科学的知識の移転を象徴しており、これは研究者の間で一般的に受け入れられています（Choi、1988、Rinia et al、2002）。引用のカウンティングの恩恵を受けて、多くの指標は、分野/分野間または分野内の知識関係を測定するために引用に大きく依存しています。関連するアプローチは、関心のある関連分野を特定し（Van Leeuwen & Tijssen、2000）、分野の知識構造を分解し（Choi、1988）、学際的な知識の流れを定量化し（Rinia、2002）、測定するために適用されています。フィールドの学際性（Porter & Chubin、1985、Stirling、2007）。ただし、引用ベースの方法には、知識の流れを調査する際に認識できる欠陥がいくつかあります。出版物を引用する動機は自然に多面的であるため、すべての引用が移転された知識を持っているわけではないと主張されています（Brooks、1986、Garfield、1965）。極端な場合として、他人の研究を非難し、以前の発見を批判する現象は、多くの研究記事で一般的です。引用分析において、そのような負のクレジットを持つ引用を正のクレジットを持つ引用と同一視することは不公平です。引用の重要性を区別することで、知識の流れを測定する研究が改善されます（Hassan、Safder、Akram、およびKamiran、2018年）。この目的を達成するための実行可能なアプローチの1つは、引用を通じて広まった知識の内容を調査することです。たとえば、Hassan et al（2018）は、引用コンテキストの機能を使用して、機械学習手法によって重要な引用と重要でない引用を区別します。別のルートとして、知識の内容を観察することによって知識の流れを調査することで、科学的知識の広がりについてさまざまな洞察を得ることができます。一つには、知識コンテンツの拡散パターンは、引用のパターンとは異なる特性を持っている可能性があります。もう1つは、知識の内容によって測定される分野間の知識の関係が異なる場合があります。これにより、この調査では、知識コンテンツを分野間で広めることに焦点を当てることができます。

知識コンテンツの拡散を定量化するには、まず、本、記事、会議論文などのソースから、数えられて入手しやすい知識単位を特定する必要があります（Liu、Rafols、およびRousseau、2012年）。操作性のために、出版物のテキストから簡単に識別できるキーワードは、多くの研究で知識の測定可能な代理であると主張されています（Sarafoglou & Paelinck、2008）。キーワードを表現として使用することにより、特定の種類の知識が具体的かつ知覚可能になります。その後、特定の知識ユニットの拡散プロセスを定量化できます。このパラダイムに続いて、最近のいくつかの研究では、研究ソフトウェア（Pan、Yan、Cui、およびHua、2018）、アルゴリズム（Wang & Zhang、2018）などの普及を定量的に調査しています。ただし、彼らは通常、知識ユニットの大規模なサンプルを調査するのではなく、ケースとしていくつかの選択された知識ユニットに焦点を合わせます。さらに、分野間の知識の関係は、知識ユニットを通じてまだ調査されていません。本研究は、知識ミームを観察することにより、学際的分野における学際的知識の流れの拡散パターンを定量化することを目的としています。基本的に、知識ミームは、引用に沿って拡散した主要な知識単位を反映しています。多数の知識ミームを特定することにより、関連分野から学際的分野への学際的知識の流れを定量的に分析することができます。ケーススタディでは、新しい学際的なドメイン、つまり医療情報学（MI）を選択し、それに関連する4つのソース分野を特定します。ネットワークを使用して、論文間の引用に基づいて知識ミームの拡散をモデル化し、ネットワーク分析を適用して、学際的な知識の流れの拡散パターンを調査します。私たちの研究では、次の研究質問に答えることを目指しています。

1. MIとそれに関連する分野との関係は、知識ミームによって測定され、引用によって測定される関係とは異なりますか。
2. 学際的な知識ミームの拡散パターンは何ですか？

3. MIと関連分野の知識ミームの拡散パターンの違いは何ですか？

この研究は模範的な領域で実施されていますが、この論文で採用されている分析方法論は、一般に、ソース分野とデスティネーション分野の間の知識関係を調査するために適用できます。知識ミームを通じて実用的なアプローチを提供する私たちの研究は、知識の代用（引用）ではなく、知識の内容の観点から学際的な知識の流れの新しい測定に貢献します。また、知識ミームを通じて伝達された知識の詳細を提供することにより、学際的研究における知識統合の理解を促進することができます。さらに、私たちの分析の結果は、特に学際的な文脈で、知識ユニットの拡散メカニズムへの貴重な洞察を提供します。

関連研究

引用による学際的な知識の拡散の測定

科学的コミュニケーションで形成され使用されている広く受け入れられている規範として、引用は、学術的前任者の関連する作品を認める簡単な方法と見なされてきました。したがって、引用は知識キャリアと見なされ、共引用分析や書誌結合分析など、科学計量学の多くの分析方法の基礎を形成します。引用の恩恵を受けて、分野間の知識関係は識別可能で定量化可能になります。知識の拡散を測定するために、以前の研究では、数えられる参照に基づいて、分野間で広がる知識の量を定量化するためのいくつかの有用な指標が提案されました（例：Citation OutCategories）（Porter&Chubin、1985）。指標は主に、あるフィールドが別のフィールドに与える影響を反映しています。特定の分野や分野に関する多くの実証的研究が同時に現れました。ある分野の場合、引用パターンを通じて複数のサブ分野間の知識関係を示すことで、その分野の構造についての洞察を得ることができます（Choi、1988）。このアプローチは、関心のある分野、特に磁性ナノ粒子研究などの新しい分野を調べて理解するためによく使用されます（Liu、Chen、Liu、およびXie、2016年）。ホリスティックサイエンスの場合、知識の境界と分野間の関係も引用を通じて調査できます（Rinia et al.、2002）。分野間の相互関係に基づいて知識ネットワークを構築することにより、ネットワーク分析を使用して、地域または世界の科学の特徴を説明することができます（Liu et al.、2011; Yan、2016）。最近の焦点は、複数の分野からの知識を含む学際的な研究です。経験的研究は、いくつかの分野が他の分野との高度な知識の相互作用を持っていることを示しています。これは、分野の学際性に反映されます。学際性を測定するために、他の分野への参照の分布に基づくいくつかの指標が提案されています。たとえば、Rao-Stirling（Stirling、2007）。基本的に、学際性指標は、ターゲット分野と他の関連分野（統合、バランス、多様性など）との間の知識関係の多次元特性を調査します（Leydesdorff、Wagner、およびBornmann、2019; Wang&Schneider、2019）。ただし、上記の研究では、引用の頻度のみをカウントし、引用によって運ばれる内容は考慮していません。科学的コミュニケーションで広く受け入れられている規範として、引用は、学術的前任者の関連する作品を認める簡単な方法と見なされてきました。したがって、引用は知識キャリアと見なされ、共引用分析や書誌結合分析など、科学計量学の多くの分析方法の基礎を形成します。引用の恩恵を受けて、分野間の知識関係は識別可能で定量化可能になります。知識の拡散を測定するために、以前の研究では、数えられる参照に基づいて、分野間で広がる知識の量を定量化するためのいくつかの有用な指標が提案されました（例：Citation OutCategories）（Porter&Chubin、1985）。指標は主に、あるフィールドが別のフィールドに与える影響を反映しています。特定の分野や分野に関する多くの実証的研究が同時に現れました。ある分野の場合、引用パターンを通じて複数のサブ分野間の知識関係を示すことで、その分野の構造についての洞察を得ることができます（Choi、1988）。このアプローチは、関心のある分野、特に磁性ナノ粒子研究などの新しい分野を調べて理解するためによく使用されます（Liu、Chen、Liu、およびXie、2016年）。ホリスティックサイエンスの場合、知識の境界と分野間の関係も引用を通じて調査できます（Rinia et al.、2002）。分野間の相互関係に基づいて知識ネットワークを構築することにより、ネットワーク分析を使用して、地域または世界の科学の特徴を説明することができます（Liu et al.、2011; Yan、2016）。最近の焦点は、複数の分野からの知識を含む学際的な研究です。経験的研究は、いくつかの分野が他の分野との高度な知識の相互作用を持っていることを示しています。これは、分野の学際性に反映されます。学際性を測定するために、他の分野への参照の分布に基づくいくつかの指標が提案されています。たと

えば、Rao-Stirling (Stirling, 2007)。基本的に、学際性指標は、ターゲット分野と他の関連分野（統合、バランス、多様性など）との間の知識関係の多次元特性を調査します (Leydesdorff, Wagner, および Bornmann, 2019; Wang & Schneider, 2019)。ただし、上記の研究では、引用の頻度のみをカウントし、引用によって運ばれる内容は考慮していません。

コンテンツアイテムの知識の拡散

最近のいくつかの研究では、普及した知識のコンテンツを説明することにより、学際的な知識の拡散を分析しようとしています。課題は、スプレッドコンテンツをどのように表現するか、および他のフィールドから継承されたコンテンツを正確に識別する方法です。最も観察されたコンテンツ項目は、研究分野の主要なホットスポットを明らかにする用語とトピックです。用語とトピックを取得するために、いくつかの最先端のテキスト分析方法が、フィールド内のドキュメントのタイトル、要約、および全文に適用されています。フィールドの主要なテーマを表す用語は、用語の頻度 (Xu, Guo, Yue, Ru, および Fang, 2016年) または共語分析 (Ba, Cao, Mao, および Li, 2019年, Hinze, 1994年) に基づいて識別できます。もう1つの広く使用されているアプローチは、トピックモデリング (潜在的ディリクレ割り当てなど) です。これは、ドキュメントテキストから顕著なトピックを抽出できます (Nichols, 2014; Song & Kim, 2013)。上記のアプローチによって明らかにされた内容は、表現の用語を通じて簡単に解釈できます。学際的なテーマを反映するために、ドキュメントの共引用クラスターも使用されました (Fang, 2015年)。ただし、識別されたドキュメントクラスターの解釈には、人間の専門家がドキュメントの内容を決定する必要があります。一般的なトピックに加えて、調査ソフトウェア (Pan et al., 2018)、調査方法 (Wang & Zhang, 2018)、イノベーション (Zhai, Ding, & Wang, 2018)。トピックを取得した後、専門知識を適用して、Baetal. のように他の分野から継承されたトピックを特定できます (2019)。しかし、人間の主観を避け、専門家の負担を軽減するために、文献で客観的な証拠を見つけることに、より多くの注意が払われてきました。最も適用される方法は、トピックが表示される記事からの推論です。幸いなことに、多くの学術文献データベースは、出版物の主題カテゴリを示すメタデータを提供します。メタデータを使用して、Xu et al. (2016) 用語の学際性を測定するために、さまざまな分野での用語の分布に基づくメトリックを提案しました。学術データベースによって提供されるサブジェクトカテゴリのメタデータは不正確である可能性があるとして主張されています (Franceschini, Maisano, および Mastrogiacomio, 2016年)。別の方法は、手動でフィルタリングする場合でも、慎重に設計された検索クエリを使用して、さまざまなドメインの記事を収集することです。たとえば、Xu et al. (2018) この戦略を使用して、学際的なドメインとその3つの関連ドメインのドキュメントを収集しました。学際的なドメインの形成を説明するために、3つのドメインで単独で発生したキーワードを分析しました。ただし、用語が属する分野を正確に特定することは困難です。用語のセマンティクスは、異なる分野で、または同じ分野内でさえ異なる可能性があります (Chen, Ding, および Ma, 2018)。コンテンツからの知識の拡散を調査することは、ミクロレベルの観点から分野の学際性を理解するためのアプローチを提供します。Chakraborty (2018) は、分野の学際性を測定するためにキーワードの多様性を調査し、コンピューターサイエンスの成長を説明するためにメトリックを適用しました。同様に、Xu et al. (2016) 図書館情報学の学際的なトピックを特定するために彼らの測定基準を適用しました。ある分野から別の分野に広がるコンテンツを観察することで、分野間の知識の関係についての鋭い洞察を得ることができます。Engerer (2017) は、情報検索の出版物が言語概念をどのように使用しているかを観察することにより、情報検索の開発における言語学の役割を調査しました。Baetal. (2019) コンピュータサイエンスからのキーワードの医療情報学の共同単語ネットワークへの埋め込みを定量化しました。これらの研究における分野間の関係は、分野の出版物における用語の出現のみを考慮し、知識の流れの明確な証拠に依存していません。私たちの研究では、引用の出典と対象の出版物に同じ用語が含まれている場合、その引用はその用語の普及の証言であると主張します。分野を超えたコンテンツアイテムの広がりを定量化した研究はほんのわずかです。パンら。 (2018) 3種類の計量書誌学的マッピングソフトウェアを使用した多くの分野の出版物を数えました。ソフトウェアの普及と同様に、Zhai et al. (2018) 潜在的ディリクレ配分の境界を越えた普及のタイムラインを作成しました。これは、革新の普及の事例と見なされていました。これらの研究は、特定のコンテンツ項目 (つまり、ソフトウェアと革新的なコンピューティングモデル) に焦点を当て、知識ユニ

ットの関連する言及のみをカウントしました。2つの研究ギャップを特定することができます。コンテンツアイテムに関しては、一般的なパターンを特定するために、それらの拡散プロセスの調査が非常に求められています。分野に関しては、それらの関係はコンテンツレベルで定量化されていません。

知識の拡散のモデル化

観察可能な項目を通じて分野間の知識の拡散を経験的に調査する研究とは別に、学者は知識の拡散をモデル化するために多くのアプローチを適用してきました。初期の研究では、知識の普及の結果である知識の成長を数学的に評価しようとしていました。たとえば、成長は知識単位の観点からロジスティック曲線としてモデル化できます (Meyer, 1994; Modis, 2007)。より最近の研究では、人口動態と科学的知識の普及のエピデミックプロセスを明らかにする根本的なメカニズムをさらに調査しました (Vitanov & Ausloos, 2012年)。理論的根拠は、科学的知識の普及は人間によって、そして人間のネットワークの間で達成されるということです。疫学モデル (Bettencourt, Cintrón-Arias, Kaiser, および Castillo-Chávez, 2006) は、知識の拡散プロセスをシミュレートするこれらのモデルの初期のタイプの1つです。一部のモデルは、より明らかになる要因も調査しました。たとえば、タイムラグのロトカ・ヴォルテラモデルは、知識の普及とその採用の間の時間差を考慮します (Karmeshu, 1982)。さらに、引用ネットワーク (Chen & Hicks, 2004) や共著者ネットワーク (Singh, 2005) などのネットワークモデルが、科学的知識の拡散プロセスを追跡するために適用されています。経験的データから構築されたこれらのネットワークは、拡散プロセスの詳細な説明を可能にします。しかし、私たちの知る限り、経験的データに基づいてコンテンツアイテムの拡散プロセスをモデル化する研究はありません。この研究では、コンテンツアイテムの引用ネットワークを紹介して、特に学際的なコンテキストでの拡散パターンを調査します。

データと方法論

ドメインの選択とデータの取得

この研究の研究の質問に対処するために、学際的な分野と、知識を学際的な分野に伝達する関連分野の書誌データを取得しました。医療情報学 (MI) は、関連分野からのさまざまな知識を統合した若い分野横断的な分野として成長してきました (Haux, 2010)。この分野は過去数十年間に多数の記事を発表しており、その結果、大量の参照と知識ミームが生まれています。このようなデータを使用して、フィールドでの学際的な知識の拡散の特性を分析することができます。したがって、ケーススタディではMIを選択しました。2018 Journal Citation Report (JCR) によると、医療情報学の主題カテゴリにある26のジャーナルの46,253の記事が、1964年から2018年までのWeb of Science (WoS) から取得されました。参照は、フィールドへの知識の流れを調査するためによく使用されます。知識提供者と見なすことができる他の分野から (Wu, Hill, および Yan, 2017年)。その主要な知識提供者を特定するために、MIの参照が解析され、ジャーナルおよびJCRサブジェクトカテゴリによってグループ化されました。上位に引用されたカテゴリから、分析のために4つの分野を選択しました。統計と確率 (統計); 医学、一般および内科 (中); とコンピュータサイエンス、人工知能 (CS)。次に、MI記事からの参照頻度に従って、4つの分野のジャーナルがランク付けされました。引用数が少ない一部のジャーナルは除外され、MIに属するジャーナルも考慮されませんでした。最終的に、HC、Stat、Med、およびCSについて、それぞれ29、25、25、および24のジャーナルを取得しました。次に、WoSからこれらのジャーナルに、2018年までの発行年の記事のメタデータを収集しました。MI内、およびMIと他の4つのフィールド間の記事の引用は、筆頭著者、省略されたジャーナルタイトル、巻番号、開始ページ、およびデジタルオブジェクト識別子 (DOI) を照合することによって取得されました。これらの知識提供者からMIへの知識普及のパターンは、この研究の主な焦点です。

知識ミームの識別

知識ミームは科学出版物のテキスト単位であり、分解したり変更したりすることなく、多くの引用出版物に配布されています (Kuhn, 2014)。出版物から知識ミームを識別するために、タイトル、要約、キーワー

ド、およびキーワードプラス（Garfield、1990）から抽出されたテキストは、字句解析、ストップワードの削除、およびトークン化によって前処理されました。キーワードの用語とKeywordPlus、およびタイトルと要約のn-gram（ $n = 1, 2, 3$ ）は、知識ミームの候補として扱われました。このアプローチ（Kuhn et al.、2014）に従って、次の式に従って候補項 m のミームスコア M_m を計算しました。

$$M_m = f_m \cdot P_m$$

$$P_m = (d_m \rightarrow m / (d_m \rightarrow m + \sigma)) / (d_m \rightarrow m / \sigma + d \rightarrow m / \sigma)$$

ここで、 f_m はテキストフィールドで候補用語 m に言及している論文の割合、 P_m は引用ネットワークでのミームの出現の継続性を評価する伝播スコア、 $d_m \rightarrow m$ は少なくとも1つに言及して引用している論文の数を示します。同じ用語を含む紙。 $d \rightarrow m$ は、少なくとも1つの用語を含む紙を引用するすべての紙（ m を含むかどうかに関係なく）の数です。同様に、 m について言及しているが、そのタイプの論文のいずれも引用していない論文の数です。は、参考文献に用語を含む論文が含まれていないすべての論文の数です。最後に、分母がゼロになるように修正し、まれにしか発生しない候補が高いミームスコアを受け取らないようにするために、Kuhn et al. の実践に従って、平滑化係数 σ が導入され、3に設定されます。（2014）。つまり、ミームスコアは、同じ用語を引用している論文を引用している論文で頻繁に発生するが、用語を含む論文を引用していない論文ではめったに出現しない用語で高いです。MIの最終的な知識ミームを決定するために、次の基準を満たさない候補を除外しました。

1. ミームスコアが0を超えています。
2. 名詞または名詞句であるか、語彙外（OOV）の単語で終わっています。
3. フリーライディング用語ではありません。短い用語は、それらを含む長い用語よりも高いスコアを取得する可能性があります。これは、フリーライディングと呼ばれます（Kuhn et al.、2014）。これは、n-gramセグメンテーションが原因です。
4. 5つの分野すべてで一般的に使用される一般的な用語ではありません。
5. ミームスコアが最も高い残りの上位10%の候補者は、MIの知識ミームとして保持されました。

MIおよびナレッジプロバイダーの各出版物でのミームの言及は、タイトルと要約のキーワード、Keyword Plus用語、およびn-gramと照合することによっても取得されました。

知識ミーム拡散カスケード

引用ネットワークは、紙レベルで観察することによって知識の流れを調査するために広く適用されてきました（Chen&Hicks、2004）。基本的に、引用関係には、引用された記事から引用された記事に拡散する科学的知識が含まれています。この研究では、知識ミームを使用して、知識ミームが両方の当事者の共有用語を示しているという点で、転送された知識をキャプチャします。従来の論文の引用ネットワークと同様に、有向ネットワークを使用して、知識ミームの拡散、つまり知識ミームのカスケードをモデル化します。MIの出版物と、選択された4つの分野への参照との間の引用ネットワークが最初に構築されます。MIの出版物間の相互引用も、引用ネットワークに含まれています。正式には、引用ネットワークは次のように表すことができます。 $G = (V, E)$ ここで、 V はノードのセット（つまり、出版物）であり、 E はノード間の引用関係のセットです。引用ネットワーク G は有向グラフであり、直接エッジが引用論文と被引用論文を接続します。MI内、およびMIと4つの知識プロバイダー間の知識の流れは、引用ネットワークを介して調査できます。出版物の引用ネットワークに基づいて、知識ミームの拡散カスケードを構築しました。各MIミームについて、カスケード内の各ノードがミームを運ぶことを要求することにより、拡散カスケードを G のサブグラフとして誘導できます。本質的に、ミームのカスケードも有向グラフです。図1は、1つのミームのカスケードの例を示しています。ノードP1、P2、およびP6は、ナレッジプロバイダー、ヘルスケアおよびコンピューターサイエンスの論文であり、他のノードはMIの出版物です。図1は、MIの出版物が他の分野から知識ミームを継承していることを示しています。一部のミームは複数の分野から継承されていることがわかり、一部はMIにの

み表示され、他の4つの分野には表示されないことに注意してください。他のノードに付随しない孤立したノードは、拡散パスにないという点で除外されます。知識ミームのカスケードに依存して、4つの知識プロバイダーから学際的な分野への知識の拡散をネットワーク分析を通じて定量化することができます。この研究では、知識ミームカスケードのいくつかの基本的および高度なプロパティを適用しました。

基本的なネットワークプロパティ

ノード数。ノードの数はカスケードネットワークのサイズを示しており、ミームの拡散パスにある出版物の数を示唆しています。エッジカウント。エッジの数は、カスケードネットワークのサイズに関する補足的な測定値であり、拡散プロセスのパスの数を示します。コンポーネント数。コンポーネントはネットワーク内のサブグラフであり、2つのノードが接続されており、サブグラフ以外のネットワーク内の頂点に接続することはできません。コンポーネントを使用して、知識の普及中に個別のコミュニティの数を測定できます。コンポーネント数を計算するとき、エッジの方向は無視されます。直径。最大のコンポーネントの直径は、拡散プロセスの距離または深さを測定するために使用されます。コンポーネントの直径は、ノード間のすべての最短パスの中で最も長いものです。

高度なプロパティライフタイム

知識ミームカスケードの存続期間は、カスケード内の最も古いMI出版物と最も若いMI出版物の間の期間です（年数プラス1）。さまざまな分野にさまざまな長さの開発履歴があるため、MIで送信されるミームの活力を正確に評価できるように、他の分野の論文を寿命の計算から除外します。たとえば、図1では、最も古いMI出版物はP3であり、最も若い出版物はP9とP10です。したがって、カスケードの寿命は6です。速度。知識ミームの拡散速度は、ノード数を寿命で割ったものとして定義されます。これは、ミームが1年に出現する出版物の平均数を測定します。短期的な影響。カスケード内のアウトディグリーがゼロのノードは、知識の拡散プロセスを開始します。したがって、知識ミームの短期的な影響は、それらのノードの1ホップ（直接接続された）ネイバーの数として測定されます。たとえば、図1のノードP1、P2、およびP6のワンホップネイバーは、ノードP3、P4、P5、およびP8です。したがって、図1のミームの短期的な影響は4です。さらに、短期的な影響は、ノードのサブジェクトプロパティに従って、分野レベルで集計できます。図1のHCとCSの短期的な影響は、それぞれ4と1です。長期的な影響。知識ミームの長期的な影響は、アウトディグリーがゼロのノードの2ホップネイバーの数として定義されます。たとえば、図1のノードP1、P2、およびP6の間接的な隣接ノードは、P7、P9、およびP10です。したがって、示されているミームの長期的な影響は3であり、短期的な影響よりもわずかに少なくなります。分野レベルでは、HCとCSの長期的な影響は3と1です。

知識ミームの規律

知識ミームに基づいて学際的な知識の流れを定量化するには、学際的な知識ミームの特定が必要です。ただし、ミームを特定の分野に割り当て、それが別の分野に属する可能性を否定するための完璧なアプローチはありません。簡単にするために、学際的な知識ミームを、MI以外の分野（つまり、4つの知識プロバイダー）からの少なくとも1つの出版物が拡散カスケードにあるミームとして定義します。学際的なミームに加えて、残りのミームはMIミームと見なされます。2種類のミームの拡散パターンの違いを調べます。さらに、学際的な知識ミームは、拡散カスケードに知識プロバイダーが1つしかない場合、ドメイン固有のミームとして認識されます。たとえば、図1のナレッジミームには2つのナレッジプロバイダー（HCとCS）があります。したがって、それは学際的なミームですが、ドメイン固有のミームではありません。拡散カスケードには知識提供者の出版物が存在しないため、この識別方法はMIミームには適用できません。別の方法として、MI固有のミームを抽出するためにChen and Xiao（2016）によって提案されたアプローチを採用します。この方法は、国の競争分野を特定する比較優位指数に基づいています。キーワードアクティビティインデックス（KAI）が定義され、用語頻度（TF-KAIと呼ばれる）と組み合わせられて、高いドメイン特異性を示す用語が選択されます。用語のTF-KAIスコアは次のように計算されます。

$$TF-KAI = n(m, d) \times n(m, d) / n(m, all) \times n(d) / n(all)$$

ここで、 $n(m, d)$ は、分野 d で知識ミーム m を掲載している出版物の数であり、 $n(m, all)$ は、分野に関係なく、知識ミーム m を掲載している出版物の総数を示し、 $n(d)$ および $n(all)$ それぞれ、分野 d およびすべての出版物からの出版物の数です。このアプローチは、与えられた分野によって強調された知識ミームにより高いスコアを割り当てます。

結果と議論

知識ミームの結果

ミーム識別を行った結果、合計で7839個のミームが得られた。表1は、ミームが含まれる論文の数を示している。ミームの数が最も多いのはMI分野の論文であり、次いで医学分野、計算機科学分野、統計学分野の論文である。ただし、全ての論文にミームが含まれているわけではない。また、ミームを持つ論文の割合が最も高いのは当然ながら情報科学であり、他の知識提供者よりも情報科学の方が多い。また、統計学の論文には平均して5.99個のミームが含まれている。この結果は、知識ミームの絶対数と論文数の観点から、HCとMedが2大知識提供者であり、CSはミームを持つ論文の割合と論文あたりの平均ミーム数で際立っていることを示している。合計すると、7839個中3518個（44.9%）の知識ミームが分野横断的なミームとして同定される。TF-KAIスコアが最も高い上位500個のミームは、最終的にMIに特化したドメインとして同定される。さらに、HC、Stat、Med、CSの領域特異的な知識ミームとして、それぞれ461、441、383、615個の知識ミームが同定された。

表2は、知識ミームを定性的に理解するために、ミームのスコアが高い上位10個のミームを示したものである。ミームのスコアが高いほど、引用による継承に成功していることを意味し、そのミームがその分野においてより高い生命力を持つことを意味する。これらのミームは、健康状態を向上させるために関連情報を利用する能力（ehealth literacy, numeracy）、医学研究方法（continual reassessment method, network meta-analysis）、医療情報システム（遠隔医療情報システム、患者ポータル）に関するミームであった。

短期的な規律インパクトが最も高いナレッジミームを4つの知識提供者について表3に示す。情報通信と医療は、最も影響力のある10個のミームのうち5個を共有している。これらは、電子カルテや費用対効果分析など、患者の情報を管理し治療効果を測定するための情報技術やアプローチに関するものである。CSでは、診断に大きな影響を与える新しいトピックである医用画像の自動分割・分類に関するミームが上位を占めた。一方、統計学のインパクトの大きいミームには、シミュレーション研究、最尤推定、混合モデルなど、多様な手法が見受けられる。また、MIでは、知見の要約や仮説の検証のために、記述的、初等的、多変量的、その他多くの統計手法がしばしば用いられる（Scotch, Duggal, Brandt, Lin, & Shiffman, 2010）。

一つの知識ミームが、複数の知識提供者から様々な影響を受ける可能性がある。各主体が提供する知識をより理解するために、表4に示すように、短期的な影響が最も大きいドメイン固有の知識ミームも調査している。異なる主体のミームは重複していない。表3に示した知識ミームと比較すると、各主体のドメイン固有ミームは、特定の研究手法で使用するデータの種類や測定基準など、より細かい概念を反映しており、より具体的であることがわかる。HCでは、Health Utilities Indexという質問票から得られる効用スコアなど、患者の健康状態評価に関するミームが多く観察された。一方、Medでは、胃がんに関係するヘリコバクター・ピロリ（胃に生息する細菌の一種）など、疾病に関連するミームが上位を占めた。CSとStatのトップミームは、画像登録、ウェーブレット変換、主成分層別化など、複数の画像を同時に正確に分析し、因果関係を解析することができる特定のデータ処理技術を含んでいる。

上位の知識ミームの質的分析に基づくと、異なる知識提供者の知識が織り込まれ、概念、理論、方法論を提供することでMI研究を後押ししていることがわかった。また、この結果は、MIが学際的な分野であり、他の分野の知識がMI研究に統合されていることを反映している。例えば、一連の機械学習モデルは、コンピュータ支援診断やウェルビーイング評価を含むMIに関する多くの研究において、既に基本的な役割を担ってい

る。多くの機械学習モデルは、まずCSによって発明され、その後MI研究者によって採用され、研究を促進するために修正されたものである。

MIと4つの知識プロバイダーとの関係

MI知識の源についてより正確に理解するために、図2に示すように、MIと4つの知識提供者の関係を探る。従来、引用による知識の拡散経路の分析では、内容項目が考慮されていませんでした。知識ミームを観察することで、分野を超えて広がる知識コンテンツの量を定量的に把握することができます。

従来の引用分析によれば、医療情報学は統計学やコンピュータサイエンスよりも医学・医療と密接な関係にある。MIはMedの論文を最も多く引用し、HC、Stat、CSと続く（図2の青い棒）。知識のミームという観点からは、知識の流れに含まれるミームについても同様の結果が得られる（図2のオレンジ色の棒グラフ）。知識提供者からMIへの引用によって運ばれるミームの頻度をカウントしている。

さらに、4つの分野からの引用あたりの平均ミームを計算する。この値が高いほど、1回の引用でより多くのミームが引用論文に継承されていることを示す。CSは4分野の中で最も引用あたりのミーム数が多く（図2の緑線）、ミームフロー数を引用数で割ることで算出される。平均すると、1回の引用で1.01ミームがCSからMIに拡散している。Medは引用数、ミーム数ともにMIと最も近い関係を示しているが、MIへの知識拡散は引用1件あたりのミーム数で見ると最も効率が悪い。1回の引用でMedからMIに転送されたミームはわずか0.71個であった。

この結果は、引用における知識の密度、すなわち引用あたりのミームの数に基づいて、分野間の関係を明らかにするものである。もちろん、すべての引用が知識提供者から知識ミームを伝達するわけではなく、1つの引用によって複数の知識ミームが伝達されることもある。CSの出版物の引用は、CSの引用数が少ないにもかかわらず、より知識集約的である。しかし、Medは被引用数は多いが知識ミームの密度は低い。このように、内容面から捉えた分野間の関係は、従来の引用分析とは異なるものである。このように、内容から捉えた分野間の関係は、従来の引用分析とは異なり、引用の下にある実際の知識伝達の間関係を明らかにするものである。

分野横断的な知識ミームの拡散パターン

知識ミームのうち、4つの知識提供者からMIへの拡散経路が観測された3518個が分野横断的なミームとして特定された。分野横断的な知識ミームの拡散パターンを調べるために、それらの知識ミームに対して拡散カスケードを構築した。これらのカスケードについて、ネットワークの特性を表5に示す。いくつかの特性の分布は、中央値だけでなく、第1、第3四分位値で示されるように、非正規で右斜めになっている。基本特性では、半数以上の拡散カスケードが論文数、引用数ともに18件以下と規模が小さい。また、直径も小さく、これらの知識ミームが一般に広まっていないことがわかる。しかし、ノード数、エッジ数、成分数の平均値は、極端な事例があるため、第3四分位値よりも有意に大きくなっている。例えば、電子カルテの普及カスケードは、ノード数2120、エッジ数5789、直径11で、23のコンポーネントから構成されています。このような大規模なカスケードは稀であるが、相当量の普及リソース（すなわち、論文や引用）を消費する。高度な性質については、寿命と速度の平均値はその中央値に近い。半数以上の知識ミームの寿命は15年以上であり、年間1.39件の論文が引用される。また、ミームの短期的なインパクトは長期的なインパクトよりも一般的に高い。実際、より多くの分野横断的な知識が、MI出版による伝達よりも、知識提供者のオリジナル出版物から直接的に継承されていることがわかる。

図3は、基本特性の分布を示したものである。カスケードのノード数（図3a）、エッジ数（図3b）、構成要素数（図3c）に大規模な異質性が観察される。80%のカスケードがノード数58、エッジ数78、構成要素数7以下である一方、ごく一部の知識ミームはネットワークサイズの点で他よりもはるかに成功しています。この結果は、ソーシャルメディアミームの人気に不均衡を見出した先行研究（Gleeson, Ward, O'Sullivan, & Lee, 2014; Weng, Flammini, Vespignani, & Menczer, 2012）と一貫している。

ノード数、エッジ数、コンポーネント数の分布は、指数関数的にカットオフされたべき乗分布に従う。数学的には、ある量が確率分布から描かれる場合、べき乗則に従う。ここで、指数として知られるパラメータは、Clauset, Shalizi, & Newman, 2009である。適合する曲線はほぼ直線で、対数スケールで下降するテールを持つ (Fig.) 分布のテールは予想よりも早く減衰しているが、これは出版物のサイズが有限であることや観測されたタイムスパンが有限であることが原因と考えられ、成功した知識ミームのさらなる普及を制限している (Alstott, Bullmore, & Plenz, 2014; 山崎ら, 2006)。のすべての値について厳密にべき乗分布に適合する実世界のデータは稀であり、べき乗分布は多くの場合、～の間のデータにしか適用できないことが示されている (Broido & Clauset, 2019; Clauset et al.) ノード数、エッジ数、成分数のデータに最尤推定法 (Clauset et al., 2009) でべき乗分布を当てはめた。推定されたパラメータは、ノード数、エッジ数、成分数でそれぞれ1.91、1.89、2.07である。また、パワーローフィットの下限と上限をFig.3に示す。

Kolmogorov-Smirnov検定による適合度のp値は0.15、0.58、0.31と0.1より大きく、経験データに対してべき乗分布が有効であることが示唆される。指数パラメータが2以下の場合、分布は有限の平均と分散を持ち、システムサイズが大きくなっても発散しない (Newman, 2005)。つまり、カスケードにおけるノード数とエッジ数の推定平均と分散は、出版物セットが成長し続けても無限にスケールアップすることはないが、コンポーネント数についてはその限りではない。成功した知識ミームはより多くの注目を集め、より多くの論文に普及する可能性があるが、新しい知識ミームやカスケードが小さいミームは引用を集めるのが難しい可能性がある。この結果は、知識ミームの普及においても、preferential attachment (Barabási & Albert, 1999)やMatthew Effect (Merton, 1968)が起こっていることを示唆するものである。

直径の分布は、上記の特性とは異なる。それは、半対数スケールでほぼ直線上に位置する (Fig.3d)。このデータに対して、指数分布を当てはめた。の推定値は0.27で、p値は0.23であった。指数分布は、特に右尾のデータ点では、べき乗分布よりもはるかに速く減衰する。このことは、カスケードネットワークの直径は、これまでの特性 (ノード数、エッジ数、成分数) よりも極値まで成長する確率が低いことを示している。本研究では、巨大成分の直径をカスケードの直径とみなした。その結果、カスケードネットワークの規模が拡大し続けても、巨大成分の直径は比較的安定していることがわかった。ミームが拡散する過程で、新たに出現した成分が既存の成分と結合せず、巨大成分の成長が阻害されることがある。例えば、有害事象報告のカスケードには2つの主要な構成要素があり、これは米国食品医薬品局の有害事象報告システムでデータマイニングを実行するための2つの異なるアプローチを表している。この分野では、統計学が初期の主流であり、対応する論文が大きな構成要素を形成している (Huang, Zalkikar, & Tiwari, 2011)。しかし、注目が高まっている機械学習の採用により、新たな構成要素が形成された (Baer et al., 2016)。

学際的ミームとMIミームの相違点

分野横断的な知識フローのパターンは、MIの他のミームと比較することで顕在化することができる。表6は、3518個の学際的ミームと他の4321個のMIミームとの間のいくつかの特性の比較を示している。学際的ミームは、MIミームと比較して、より多くのノード、エッジ、コンポーネントを持つより大きなサイズのカスケードネットワークを持ち、また、より大きな直径を持っていることが一般的である。さらに、分野横断的なミームは、寿命、速度、短期および長期の影響という点で、MIミームより優れている。Mann-Whitney検定により、すべての差が統計的に有意であることが示された。

興味深い結果は、学際的ミームとMI知識ミームの最初の出版年数の差に見られる。図4に示すように、学際的ミームは、一般に、MIミーム (中央値: 2009年、Q1: 2001年~Q3: 2014年) よりも早い時期に発生する (Mann-Whitney $U = 5.076E + 06$, $p < 0.001$)。先行研究では、学際領域は一般的に関連分野から知識を吸収してから独自の知識を生産・輸出することが、引用に基づく観点から証明されている (Chakraborty, 2018; Xu et al.) 我々の実証結果は、知識ミームの観点からこの知見を確認するものである。初期のMI出版物は、分野横断的な知識の拡散を通じて、知識提供者から知識を受け継いだ。学問が徐々に成熟するにつれ、MIは独自の研究テーマを開発し、革新的な知識を生み出すことができたが、拡散カスケードの大きさと

いう点では依然としてインパクトが劣っている。例えば、eHealthはMIにおいて最も有望な分野の一つであり、遠隔医療、在宅医療、臨床情報システムに関する研究が含まれている（Haux, 2010）。

さらに、分野横断的な知識拡散の分野別バリエーションに光を当てるため、分野別の知識ミームの違いを調査しています。MI、HC、Stat、Med、CSの比較では、それぞれ500、461、441、383、615のドメイン固有知識ミームが含まれる（Table 1）。表7は、5つの分野のドメイン固有知識ミームのカスケードネットワークの特性を示している。MI特有のミームはエッジ数が最も多く、寿命もやや長い。CSのミームはノード数、直径、速度、短期的影響力の点で第一位である。Statについては、そのミームの寿命が他より劇的に長く、MIのミームがそれに続く。HCとMedは、ノード数、エッジ数、寿命、両タイプのインパクトで他の被写体に劣る。しかし、成分数と直径は全分野でかなり似通っている。

結論

現在のほとんどの研究は、引用を通じた知識の流れを測定しています。本研究では、内容の観点から知識ミームによる学際的知識拡散を定量化する。知識ミームは、出版物における科学的知識の継承を反映する特別なタイプの知識単位として定義されます。知識ミームの拡散カスケードを導入して、それらの拡散プロセスを観察しました。さらに、学際的な分野と関連分野との間の知識の関係を定量化し、拡散カスケードのネットワーク構造を分析することにより、学際的なミームの拡散パターンを決定しました。医療情報学は、この研究の模範的な学際的な分野として使用され、その参考文献の4つの代表的な分野が知識提供者として選択されました。MIとその4つの知識プロバイダー間の知識の流れは、特定された知識ミームに基づいて分析されました。

主な調査結果と影響

この研究では、いくつかの興味深い結果が見つかりました。知識ミームに基づいて、学際的な知識の流れの特定の内容を識別することができます。知識の拡散の密度は、引用ごとに転送された知識ミームの平均数によっても測定できます。MIの場合、私たちの結果は、引用数の少ない分野（コンピュータサイエンスなど）への引用は、引用数の多い分野（医学など）への引用よりも知識集約的である可能性があることを示唆しています。

最も注目すべきことに、私たちの知る限りでは、これは知識ミームを介した学際的な知識の流れのパターンを定量化した最初の研究です。カスケードプロパティの分布を分析することにより、基本的なネットワークプロパティの分布に極端な不均衡が見られました。拡散カスケードのノード数、エッジ数、およびコンポーネント数は、指数分布に従うべき乗則分布に従い、直径は指数分布に従いました。これは、MIの知識ミームのごく一部だけが非常に成功しているのに対し、ほとんどの知識ミームの影響は限定的であることを示唆しています。当てはめられたべき乗則の分布は、学際的な知識ミームの拡散における優先的アタッチメントまたはマタイ効果の効果を示しています。一方、新しい切断されたコンポーネントは、既存のコンポーネントの成長を遅らせる可能性があり、その結果、拡散カスケードの直径が極端に達する可能性が低くなります。

医療情報学のミームと比較して、学際的な知識のミームは、より大きく、より深い拡散カスケードを持ち、より長い時間普及し、より大きな影響を示すことがわかりました。以前の研究（Chakraborty, 2018; Xu et al., 2018）と一致して、学際的な知識のミームは、一般に、学際的な分野のミーム、つまりこの研究のMIよりも早く発生するようです。MIは、学際的な分野として、初期の段階で他の分野から積極的に学びました。ドメイン固有の知識ミームが異なれば、拡散パターンも異なります。コンピュータサイエンスからの知識はより速いペースでより大きな影響を与えて広められましたが、統計学の知識は普及プロセスでより長い時間維持されました。ヘルスケアと医学は多くの特性で類似性を示しましたが、拡散カスケードのサイズと影響の点で他の2つより劣っていました。

この研究は、引用ベースのアプローチとは異なる視点である、知識ミームを介した学際的な知識の流れを測定することにより、知識の拡散に関する貴重な洞察を提供します。方法論的には、知識ミームを使用する

と、知識ミーム識別アルゴリズムを採用することにより、コンテンツベースの知識拡散の定量分析が可能になります。理論的には、引用はそれらが運ぶミームに従って区別することができます。ミームを明示的に転送しないノイズの多い引用は、ナレッジフロー分析で除外できます。したがって、知識ミームによって測定される分野間の知識拡散関係は、引用ベースの関係とは異なります。さらに、提案された知識ミームの拡散カスケードは、ネットワーク分析を適用することにより、ミーム拡散パターンの定量的調査を容易にします。

制限と将来の研究

確かに、この研究にはいくつかの制限があります。まず、知識ミームは、さまざまな出版物でさまざまな用語で表現される可能性があります。私たちのテキスト処理ステップでは、引用および引用された論文からミームのすべてのバリエーションを特定できなかった可能性があります。第二に、特定の分野に用語を割り当てることは困難でした。拡散カスケードの構造に基づいて、学際的かつドメイン固有の知識ミームを特定しました。MIドメイン固有のミームは、TF-KAIアプローチによって具体的に選択されました。この方法では、知識のセマンティクスは使用されません。将来の作業では、コーパス統計と機械学習技術を組み合わせて、用語の学際性を理解する可能性があります。さらに、この作業では形式知の流れに焦点を当てました。知識のミームを共有した出版物間の引用のみが、有効な知識の移転であると見なされました。私たちの観察によれば、多くの出版物は同じ用語を共有していますが、引用関係はありません。このような暗黙知の流れも重要であり、今後、その拡散メカニズムを探るのはさらに難しいでしょう。