

# オープンなグラフ処理を実現するインターデータネットワーク のためのコンテンツ探索機能の試作

岡部 貴之<sup>†</sup> 須加 純一<sup>†</sup> 関屋 元義<sup>†</sup> 松原 正純<sup>†</sup> 片桐 徹<sup>†</sup> 金子 晋丈<sup>‡</sup>  
富士通株式会社<sup>†</sup> 慶應義塾大学<sup>‡</sup>

## 1. はじめに

近年、文献や人を探す場面において、文献や人を表すコンテンツをノード、コンテンツ間の関係をエッジとするグラフを使った探索サービス(Connected Papers[1]や Linked Jazz[2])が登場している。これらのサービスは、論文や特定の分野に関する人のコミュニティといった限られた領域での探索であるのに対して、筆者らはインターネットのような多様な領域のコンテンツを対象とするグラフ探索を目指したインターデータネットワーク (Inter Data Network: 以降, IDN) [3]を提案している。IDNでは、ユーザの知識や行動に基づいて、コンテンツ間のグラフを成長させることで、多様なユーザ視点でのグラフによる探索を目指している。このような多様な意味を持つグラフに対し、ユーザが効果的な探索を行うには、ユーザのグラフ探索の操作や、操作結果による表示といったユーザインタフェースがより重要である。筆者らは、ユーザ向けに、グラフの参照とグラフの演算をインタラクティブに操作することで、所望のコンテンツを探索する方式を考案し、試作した。

## 2. IDNの概要と課題

IDNは、ユーザのコンテンツに対する思考や、コンテンツの利用履歴といったユーザ視点の知識や行動を、コンテンツ自身およびコンテンツの関係性のグラフに反映するとともに、グラフを成長させることで、その後のユーザが、グラフを活用して、目的に応じたコンテンツや新たな気付きとなるコンテンツを容易に発見することを目指したプラットフォームである。その実現に向けて、IDNでは、以下の3つのレイヤを構成する(図1)。

### (1) コンテンツ仮想化

コンテンツ仮想化は、ユーザが、コンテンツの利用に対し、その利用価値や利用目的、概要をメタデータとして扱うレイヤである。IDNでは、

本メタデータを Virtual File(以降, VF)と表現する。コンテンツを利用したユーザが、VFを生成・流通することで、その後のユーザが目的に合ったコンテンツの発見を容易にする。

### (2) カタログ管理

カタログ管理は、ユーザが、目的に対して利用した複数のコンテンツのVFをグループとして管理するレイヤである。IDNでは、複数のVFのグループを Catalogue と表現する。Catalogueによって、複数のVFを関連付けすることが可能となる。また、1つのVFは、複数のCatalogueに属することもあり、これによってVFをノード、Catalogueをエッジとした HyperGraph[4]を構成する。

### (3) グラフ演算

VFとCatalogueの関係によって構成されるグラフに対して、グラフ演算を提供するレイヤである。本レイヤでは、中心性指標や Page Rank, コミュニティ抽出を代表とする多様なグラフ演算を用意している。

IDNは、上記の3つのレイヤから構成され、各種グラフ演算処理を外部にオープンにすることで、開発者による、これらを活用した多様なコンテンツ探索機能の開発を期待している。

3章では、IDNで提供されるグラフ演算を活用したコンテンツ探索機能の一試作を提案する。

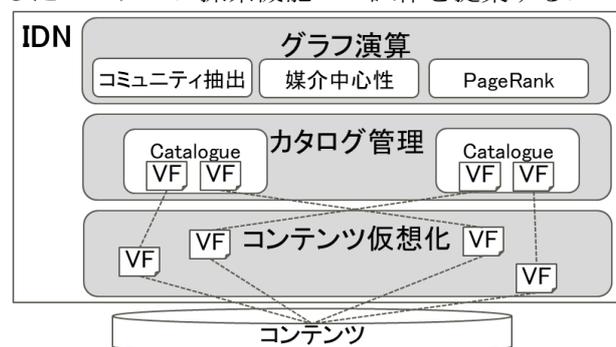


図1 IDNのアーキテクチャ

## 3. 提案方式

従来では、限られた領域でのコンテンツ探索であったため、ユーザにとって、所望のコンテンツを発見するのは比較的容易であった。一方、IDNでは、多様な領域でのコンテンツ探索を行う

A Prototype of Contents Explorer for Inter Data Network aiming Open Graph Processing

<sup>†</sup>Takayuki Okabe, Junichi Suga, Motoyoshi, Sekiya, Masazumi Matsubara, Toru Katagiri, Fujitsu Limited

<sup>‡</sup>Kunitake Kaneko, Keio University

ため、ユーザによる多様なグラフ演算の操作の実行と、その実行結果によるグラフの参照の繰り返しによって、徐々に探索範囲を絞り込み、最終的にユーザがコンテンツを発見する仕組みが必要となる。また、ユーザのコンテンツ探索をより効率的に行うには、ユーザが閲覧したコンテンツの履歴を探索の入力として活用することも重要である。

本コンテンツ探索機能では、ユーザに対し、グラフ演算の操作とグラフ演算結果の対象となるコンテンツ一覧を示すユーザインタフェース(a)と、コンテンツの閲覧履歴を基に、それらの周辺のコンテンツのノードを含むグラフを表示するユーザインタフェース(b)を提供し、ユーザは、これらをインタラクティブに操作することで、容易なコンテンツの発見を可能とする。

(a) グラフ演算操作・コンテンツ一覧表示

ユーザは、本ユーザインタフェースを介して、探索範囲とその範囲内のコンテンツを柔軟に絞り込める。ユーザは、グラフをエッジの密度に基づいて複数の部分グラフ(コミュニティ)に分割する「コミュニティ抽出」のグラフ演算を使い、探索範囲をグラフ全体から一部のコミュニティに絞りこむ。また、絞り込んだコミュニティに含まれるコンテンツ群を、複数の中心性指標に基づいた順序で表示する。

ユーザは探索範囲を、コミュニティ抽出におけるグラフの分割の粒度と探索の起点となるノードの制御によって変更できる。例えば、探索範囲が図2のコミュニティXで、ノードAを探索の起点とする場合、コミュニティ抽出の粒度を上げると、その起点を含めたより広いコミュニティYに探索範囲が広がる。ユーザがノードBに探索の起点を移して、コミュニティ抽出の粒度を下げると、探索範囲はコミュニティZになる。このようにしてユーザは、探索範囲を柔軟に変えることができる。

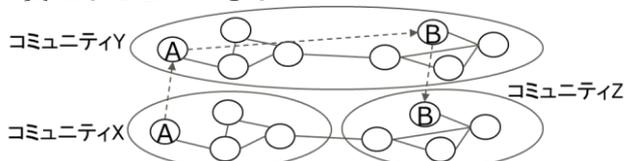


図2 探索範囲とするコミュニティの変更

また、ユーザは複数の中心性指標のそれぞれの目標値を設定でき、本探索機能は、各コンテンツについて各中心性指標を計算し、目標値と近い値であるコンテンツを優先的に表示する。例えば、度数中心性と媒介中心性の目標値を高く設定すれば、ただエッジが多いだけでなく複数のコミュニティを繋ぐ重要な橋渡しにもなっ

ているノードを特定しやすくなる。これによりユーザは、グラフの構造的な特徴に基づいてコンテンツを柔軟に絞り込むことができる。

(b) 閲覧履歴を基にしたグラフ表示

本ユーザインタフェースは、閲覧履歴をユーザが興味のあるコンテンツであると考えて、閲覧履歴に関する最小シュタイナー木グラフを描画する。これは、閲覧履歴を結ぶ長さ最小の木構造のグラフである。図3に最小シュタイナー木グラフの一例を示す。閲覧履歴を結ぶエッジは元のグラフの中から選ぶが、閲覧履歴を直接結ぶエッジがない場合は、その間の中継となるノード(シュタイナー点)とエッジを含める。シュタイナー点の中で、複数の閲覧履歴の近傍にあるものは、ユーザの興味に関連する可能性が特に高いと考えられる。ユーザは、(a)グラフ演算操作・コンテンツ一覧表示で、このシュタイナー点を起点に探索することで、自分の興味に関連するコンテンツを効率的に探索できる。

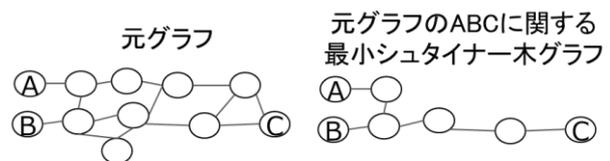


図3 最小シュタイナー木グラフの例

4. おわりに

本稿では、IDNにおいて、グラフの参照とグラフ演算をインタラクティブに操作することで、所望のコンテンツを探索する方式を提案し、その方式に基づいて試作したコンテンツ探索機能について述べた。今後は、コンテンツ探索機能の有効性を検証する。

参考文献

[1] Alex E, Eddie S, “Connected Papers”, <http://www.connectedpapers.com/>, (参照 2022-1-5)  
 [2] Matthew M, Jeff W, Cristina P, Visualizing linked Jazz: A web-based tool for social network analysis and exploration, Proceedings of the American Society for Information Science and Technology, vol. 49 pp1-3 2012  
 [3] 高野明彦, 嘉村哲郎 (2020). アートシーンを支える 勉誠出版  
 [4] M.M. Wolf, A.M. Klinvex, D.M. Dunlavy, “Advantages to modeling relational data using hypergraphs versus graphs,” 2016 IEEE High Performance Extreme Computing Conference (HPEC), pp. 1-7, 2016