

複雑ネットワーク情報空間の没入型 インタラクティブ可視化システム

山崎正晃^{*1} 熊野雅仁^{*2} 小野景子^{*2}
木村昌弘^{*2} 斉藤和巳^{*3}
大原剛三^{*4} 元田浩^{*5}

効果的な可視化や直感的な操作は複雑ネットワークから知識を発見する上で重要なことである。我々は複雑ネットワーク情報空間の没入型インタラクティブ可視化システムを提案する。本システムは、複雑ネットワークを情報拡散過程に基づき 3D 空間に可視化する。そして、情報空間を探検するため没入できる環境を提供し、また、直感的に複雑ネットワークを操作し、ノードの情報を調べることができる。我々は、実データをを用いることで、実験を通じて提案システムの有用性を示した。

Constructing a 3D Immersive Information Space for Interacting with Complex Networks

M. YAMAZAKI,^{*1} M. KUMANO,^{*2} K. ONO,^{*2} M. KIMURA,^{*2}
K. SAITO,^{*3} G. OOHARA^{*4} and H. MOTODA^{*5}

Effective visualization and intuitive manipulation are vital for knowledge discovery from complex networks. We propose such a system that realizes a 3D immersive information space for interacting with complex networks. The system visualizes a complex network in 3D space so as to exhibit its information diffusion function. It provides an immersive environment to explore the information space, and enables us to intuitively manipulate the complex network and investigate the information of its nodes. Using real data, we experimentally demonstrate the effectiveness of the proposed system.

1. はじめに

古来より、情報は、実空間、人の脳内記憶や外化された記録物体として存在してきたが、20 世紀末からインターネットが爆発的に普及したことにより、情報は統一的に電子化され、アクセス可能な膨大な情報が情報空間に流入し続け、その価値が高まり続けている。しかし、アクセス可能であっても、膨大で混沌とした情報群の中から、必要なだけ有益な情報を獲得し、理解するためには、情報群から選択的に情報を抽出する技術が必要となる。

このような情報群の要素は、複雑な既存のリンク構造や潜在的な関係で相互に結ばれており、既存の実リンクで結ばれた情報だけでなく、必ずしも実リンクで結ばれていない有用な情報を効率良く抽出する必要性が生じ、第一の検索システムが登場した。第一の検索システムは、キーワードを入力することで、そのキーワードに関連する情報を抽出してランキング形式で出力する。しかし、同じキーワードを誰が入力しても同じ結果になるという問題が指摘されるようになった。この問題を克服する第二の検索システムとして、ユーザの嗜好に適応した情報群を自動抽出するリコメンデーションシステムが登場し、注目されている。

ただし、第一および第二の検索システムには、第一の検索システムから引き継がれた潜在的な問題や、第一から第二の検索システムへ進展するに従い、指摘されるようになった潜在的な問題がある。例えば、図書館や書店などの実空間で目的の本を探す場合、その空間内の移動時に、偶然目にした情報を契機に閃きを得ることがある。このような情報の探索中に、探している対象とは別の価値あるものを見つける能力はセレンディピティ (serendipity) と言われるが、第二の検索システムにおける個人の嗜好に過適応するほど、目的の情報以外は提示されず、セレンディピティを発揮する機会が損なわれるという指摘である。また、第一、第二の検索システムで得られる情報のリストは、そのリスト要素相互の関係の有無が捉えにくいという問題や、情報群の俯瞰視による、より全体的な関係を直感的に把握したいと

*1 現在、龍谷大学 大学院理工学研究所 電子情報学専攻

Presently with Division of Electronics and Informatics, Ryukoku University

*2 現在、龍谷大学 理工学部 電子情報学科

Presently with Department of Electronics and Informatics, Ryukoku University

*3 現在、静岡県立大学 経営情報学部 経営情報学科

Presently with School of Administration and Informatics, University of Shizuoka

*4 現在、青山学院大学理工学部情報テクノロジー学科

Presently with Department of Industrial and Systems Engineering, Aoyama Gakuin University

*5 現在、大阪大学 産業科学研究所

Presently with Institute of Scientific and Industrial Research, Osaka University

いう要求には向いていないという問題もある。

また、情報に触るといふ意味のタンジブルビット¹⁾の概念を掲げる石井は、これまでの情報システムが、ブラックボックスを増やし、情報が持つ価値の氷山の一角しか提示しなくなったという問題や、実空間での物事の理解の仕方として、仕組みや関係を追跡できる追跡可能性 (traceability)、視覚的な可読性 (legibility) の減退を招いた問題を指摘している。

このような観点を重視し、それらを克服するための、第三の検索システムにつながる視点が考えられる。それは、これまでの検索システムがユーザーの代行をしてきた情報自動抽出処理への依存度を減らし、ユーザが自ら情報空間内に没入し、移動し、ユーザの情報への選択的注視、情報の抽出という操作権を復権させ、情報の全体もしくは部分の関係把握を支援する、情報空間への没入型情報探索システムであり、様々な操作を支援するインタラクティブ性が必要となる。この情報空間への没入とオブジェクト化された情報のインタラクティブな操作により、serendipity, traceability, legibility への効果も期待できる。この第三の検索システムにつながる複雑ネットワークを可視化する研究としては、ノード群の大規模化への対応、処理の高速化などや、主にノードのレイアウト法が盛んに研究されている²⁾。その一方、可視化するだけでなく、可視化を通じて情報を得るための情報可視化の研究³⁾や、仮想空間を通じて視覚的に情報をマイニングする 3DVDM(3D Visual Data Mining)⁴⁾ など、第三の検索システムの実現に応用できる研究も盛んに研究されるようになった。

本研究では、没入型情報空間インタラクティブ可視化システムを構築する上で、複雑ネットワークの可視化法として、Saito ら⁵⁾の情報拡散に基づく可視化法を利用する。Saito らは、複雑ネットワークにおける情報拡散という動的な情報の流れを、その系統ごとに影響度の順序性に基づく静的なレイアウトに反映させ、特に情報群の俯瞰視で効果が発揮される直感的に把握し易い可視化法を提案した。この可視化法は、情報拡散上、近い関係のノード群が近くに密集する傾向がある。この場合、ノード群から具体的に有益な情報を得るために、俯瞰視だけでなく、細部の関係を見る必要があり、没入型インタラクティブ可視化システムの能力を評価する上で、親和性が高い可視化法であると見なした。また、可視化する複雑ネットワークのリンクを選択的に表示することで、ネットワークノード間の類似度の関係と、情報拡散に基づく影響度の順序性と近さを同時に検証することができる。本研究で可視化するデータとしては、潜在的な関係を持つ、実リンクのない電子化された情報群にリンクを与え、有益な情報の探検による発見を目指すタスクに取り組むため、人物の共起関係からネットワークを形成したウィキペディア人物ネットワークと、TFIDF に基づく文書間類似度⁶⁾に基づきネットワーク形成した、毎日新聞記事ネットワークという実データを用いる。

以上から、本研究では、情報空間を移動・探検しながら情報空間に可視化された情報群とリンクをインタラクティブに操作して俯瞰の視点、細部の視点から情報間の関係を見ることのできる、没入型インタラクティブ可視化システムのプロトタイプについて、移動・操作・探検による情報の獲得という観点からの評価を報告する。2章では、複雑ネットワーク可視化法について、3章では没入型インタラクティブ可視化システムについて、4章では提案システムの詳細について、5章では実験と評価を行い、6章でまとめを行う。

2. 複雑ネットワーク可視化法

我々は、複雑ネットワーク $G = (V, E)$ を、その情報拡散機能という動的な観点から理解するために、Saito らの手法⁵⁾を用いて 3次元空間内にレイアウトする。ここに、 V はノード全体の集合、 $E \subset V \times V$ はリンク全体の集合を表す。

本ネットワーク可視化法では、ネットワーク G 上の情報拡散モデルとして、*Independent Cascade(IC)* モデルと *Linear Threshold(LT)* モデルを考える。まず、これら情報拡散モデルに対して、サイズ K の影響最大化問題の解 $U_K = \{u_k; k = 1, \dots, K\}$ を、Kimura らの手法⁸⁾を用いて求める。ノード $u_k \in U_K$ はターゲットノードと呼ばれ、ノード $v_n \in V \setminus U_K$ は非ターゲットノードと呼ばれる。次に、各ターゲットノード u_k を情報源としたとき、非ターゲットノード v_n にその情報が拡散する条件付き確率を計算する。そして、ネットワーク内のターゲットノードと非ターゲットノードの全てのペアの条件付き確率の関係を持するように、ネットワーク内のすべてのノードを 3次元空間内に埋め込む。また、2つのノードラベリング法により、 G 上での情報拡散過程を表現する。以下に、本ネットワーク可視化法を概説する。

2.1 情報拡散モデル

本ネットワーク可視化法では、複雑ネットワーク上での情報拡散モデルとして、IC モデルと LT モデルが用いられる。以下に、Kempe らの研究⁷⁾にしたがって、これら情報拡散モデルを説明する。

まず、IC モデルおよび LT モデルでは、次が仮定されている。

- ノードは“アクティブ”か“非アクティブ”のどちらかの状態しかとらない。
- その情報が伝わったノードをアクティブノードとし、そうでないノードを非アクティブノードとする。
- ノードは非アクティブからアクティブには変化するが、その逆は変化しない。
- ネットワーク上でのその情報の広がり(拡散)は、アクティブノードの広がり(拡散)

として表現する．

- アクティブノードの初期集合 X が与えられたとき、 X に属するノードは時刻 0 で初めてアクティブになったとし、その他のノードは非アクティブとする．そして、アクティブノード拡散過程は離散時間 $t \geq 0$ で展開していく．

2.1.1 Independent Cascade モデル

IC モデルを定義する．本モデルでは、各リンク $(u, v) \in E$ に対して、実数値 $\beta_{u,v} \in (0, 1)$ を事前に指定する．ここに、 $\beta_{u,v}$ はリンク (u, v) を通しての拡散確率と呼ばれる．

初期アクティブノード集合 X が与えられたとき、IC モデルの拡散過程は次のように進んでいく．ノード v が時刻 t で初めてアクティブになったとき、 v は、非アクティブであるその各子ノード w をアクティブにする試行を時刻 t で行い、その試行は確率 $\beta_{v,w}$ で成功する．もし、 w の複数の親ノードが時刻 t で初めてアクティブになった場合は、それら親ノードが w をアクティブにする試行は任意の順序で独立に順々に行われることになるが、これらの試行はすべて時刻 t で行われる．そして、 w をアクティブにする試行のうち、少なくとも一つの試行が成功したとき、 w は時刻 $t+1$ においてアクティブとなる．ところで、 v が時刻 t で w をアクティブにするのに成功したか失敗したかにかかわらず、時刻 $t+1$ 以降では、 v はもはや w をアクティブにする試行を行うことはできない．新たにアクティブとなるノードが存在しなくなったとき、本拡散過程は終了する．

初期アクティブ集合 X に対して、IC モデルでのランダム過程における最終アクティブノード数の期待値を、 $\sigma(X)$ と定義する． $\sigma(X)$ はターゲット集合 X の“影響度”と呼ばれる．

2.1.2 Linear Threshold モデル

LT モデルを定義する．本モデルでは、任意のリンク $(u, v) \in E$ に対して、“重み”と呼ばれる正数 $w_{u,v}$ を

$$\sum_{u \in \Gamma(v)} w_{u,v} \leq 1$$

となるように前もって指定する．ここに、 $\Gamma(v)$ はノード v の親ノード全体の集合、 $\Gamma(v) = \{u \in V; (u, v) \in E\}$ である．

アクティブノードの初期集合が与えられたとき、本モデルにおけるアクティブノード拡散過程は、次のように進んでいく．まず、各ノード v に対して、“閾値”と呼ばれる実数 θ_v を区間 $(0, 1)$ から一様ランダムに選ぶ．ノード v を時刻 t での非アクティブノードとする．このとき、 v は、時刻 t でアクティブな親ノード u から、重み $w_{u,v}$ に従って影響を受ける．

$\Gamma_t(v)$ を、時刻 t でアクティブである v の親ノード全体の集合とする．もし、アクティブな親ノードからの重みの合計が閾値 θ_v 以上であれば、すなわち、

$$\sum_{u \in \Gamma_t(v)} w_{u,v} \geq \theta_v$$

であれば、 v は時刻 $t+1$ でアクティブとなる．新たにアクティブとなるノードが存在しなくなったとき、本アクティブノード拡散過程は終了する．

X をアクティブノードの初期集合とし、LT モデルの下での X からのランダム過程において、最終アクティブノード数の期待値を $\sigma(X)$ と定義する． $\sigma(X)$ はターゲット集合 X の“影響度”と呼ばれる．

2.2 影響最大化問題

IC および LT モデルに関して、 $|V|$ 未満の正整数 K に対する G 上でのサイズ K の“影響最大化問題”とは、「 G の要素数が K の V の部分集合のうち、影響度の最大値を実現するもの、すなわち、

$$U_K = \arg \max_{X \subset V, s.t. |X|=K} \sigma(X)$$

を求めよ」という問題である．Kempe らは、IC および LT モデルに関する影響最大化問題においては、貪欲アルゴリズムに基づく解法が有効であることを示した⁷⁾．しかしながら、大規模ネットワーク上で貪欲解を求めるには多大な計算コストが必要であった．Kimura らは貪欲解をボンドパーコレーションとグラフ理論に基づいて効率的に計算する手法を提案した⁸⁾．本ネットワーク可視化法では、IC モデルおよび LT モデルに対して、ネットワーク $G = (V, E)$ 上でのサイズ K の影響最大化問題の解 $U_K = \{u_k; k = 1, \dots, K\}$ を、Kimura らの手法を用いて求める．

2.3 ノード埋め込み

さて、複雑ネットワークを、IC モデルおよび LT モデルに基づいて、3次元空間内にレイアウトする手法をを説明する．

まず、1つのターゲットノード u_k のみが初期情報源であるとき、非ターゲットノード v_n にその情報が拡散する条件付き確率 $p_{k,n} = p(v_n|u_k)$ を考える．そして、ネットワーク内のターゲットノードと非ターゲットノードの全てのペアの条件付き確率の関係を保持するように、ネットワーク内のすべてのノードを3次元空間内に埋め込む．より正確には、 x_k ($k = 1, \dots, K$) と y_n ($n = 1, \dots, N$) をそれぞれ、ターゲットノード u_k ($k = 1, \dots, K$) と非ターゲットノード v_n ($n = 1, \dots, N$) の3次元空間内への埋め込み点とし、それらを次の

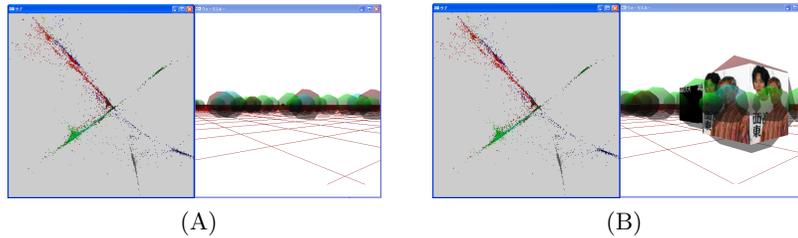


図 1 システム上における可視化法の表示

Fig.1 Scenery of actually operating the system

目的関数 \mathcal{E} の最小化問題として求める .

$$\mathcal{E} = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^K p_{k,n} d_{k,n} - \sum_{n=1}^N \sum_{k=1}^K (1 - p_{k,n}) \log(1 - \rho(d_{k,n})).$$

ここに, $d_{k,n}$ は x_k と y_n の 3 次元ユークリッド距離であり, $\rho(s)$ は $\rho(s) \in (0, 1]$ なる減少関数である .

2.4 ノードラベリング

次の 2 つのノードラベリング法を用いて外観を決めることにより, ネットワーク上での IC モデルおよび LT モデルに基づいた情報拡散過程を表現する . 1 つ目は,

$$l_1(v_n) = \arg \max_{1 \leq k \leq K} p_{k,n}$$

により, 各非ターゲットノード v_n にターゲットノードのラベル $l_1(v_n)$ を割り当てる . 2 つ目は, b を対数の底として,

$$l_2(v_n) = \left(-\log_b \max_{1 \leq k \leq K} p_{k,n} \right) + 1$$

により, 各非ターゲットノード v_n にターゲットノードからの情報の到達の度合いを反映したラベル $l_2(v_n)$ を割り当てる . ただし, $Z = \{v_n \in V \setminus U_K; \max_{1 \leq k \leq K} p_{k,n} = 0\}$ に属するノードのラベルには, Z に属さないノードのラベルの最大数を割り当てる .

3. 情報空間没入型インタラクティブ可視化システム

3.1 情報空間の視覚的提示法とインタラクティブ操作

Shneiderman¹⁰⁾ は, 情報可視化手法によって大量情報の視覚的可視化が可能になったとき, ユーザはどのような操作手順で情報可視化手法を使えばいいのかという情報可視化システムに求められる主要な操作機能として「まず全貌を眺め (overview), 重要な箇所に注

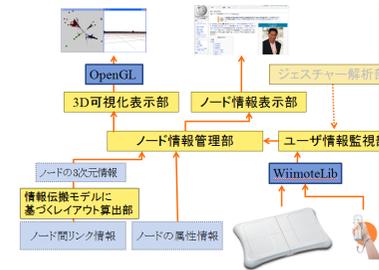


図 2 可視化システムの処理過程
Fig.2 Process of the system



図 3 実際のシステム操作
Fig.3 Scenery of actually operating the system

視し (zoom), 不要な局所は割愛し (filter), それに必要なに応じて詳細情報を探索しなさい (details on demand)」というマントラを提唱した . これらは, 情報可視化に求められる基本操作ではあるものの, 手先に偏った操作となる傾向があると思われる .

図 1(A) は, 情報拡散モデルに基づく複雑ネットワークの可視化結果を図 2 の処理過程を用いて表示したインターフェースの概観であり, 二つの窓で構成されている . 図 1(A) 左側の窓は, データ群の全体像を俯瞰視した例であり, この (A) 左窓で拡大・縮小や情報空間上の現在位置を確認することができる . この点で, overview と zoom の一例を実現していることになる . 図 1(A) 右側の窓は, 情報空間内のユーザ視点を表したものであり, 現在位置から近くのノードを見た様子である . このノードは, ユーザが近くに寄ると, その行為をノードの内部情報を知りたい意図として捉え, ノードの詳細情報 (ノード固有の文章や画像) を提示する (図 1(B) 右) . この点で, details on demand の一例を実現していることになる . また, 情報拡散モデルに基づいた可視化法では, 主要ノードに関連するノード群が同じ色で表現されるが, 同じノード群に属する任意のノードを選択し, そのノード群だけを残すことができる . また, 各ノードと他のノードのリンクの表示・非表示をノード単位, グループ単位で操作できる . これは, filter の一例を実現したことになる .

本研究では, これらのノードやリンクへの操作について, 大画面の利用を考慮し, キーボード操作だけでなく, 任天堂の Wii リモコンによって離れた位置から操作できるインタラクティブ操作法も採用した .

3.2 情報空間没入・探検のための移動操作

また, 本研究では, システムを構築するにあたり, 以上の Shneiderman の提唱した基本

操作概念を取り入れながら、情報空間に没入して探検するという概念も取り入れるため、移動という操作を手ではなく、身体性を復権させるインタフェースとして導入する。

石井のタンジブルピット¹⁾においては、身体動作により触り動かすことで対象を理解するという、その理解の過程における追跡可能性など、身体性を通じた情報理解を重視し、主に視覚に偏った情報システムに、身体性を復権させ、視覚と身体感覚を同時に用いる、同時・多感覚性を重視することも特徴となっている。また、近年、任天堂の Wii やマイクロソフトの Kinect など、ゲーム機器の分野から身体性を復権させるインタフェース技術が生み出される動向が発生しており、情報空間への没入型インタラクティブ可視化システムと操作・移動の面で親和性が高いことを考慮して、足の力点・体重移動で Wii ボードを操作する移動インタフェースの導入を行った。図 3 は、その様子を示したものである。これにより、机や椅子から離れた位置でノード操作と視覚的提示法の切り替え、移動操作を同時に操作でき、視覚と身体性による同時・多感覚的操作を支援することができる。

空間内を探検して情報を得るという行為は、人間本来の情報獲得法と通じるため、デジタルデバイス、情報格差、文化格差を乗り越えた情報探索システム実現への足がかりとなる可能性も期待できる。本研究では、そのような視点に考慮しつつ、3D 空間内をへと没入し、サーフィンの感覚で移動しながら情報探索行為と移動行為を繰り返す、人間本来の感覚に近づけたインタフェースを構築した。この情報空間への没入とオブジェクト化された情報のインタラクティブな操作により、serendipity, traceability, legibility への効果について有用性を検証する。

4. 提案システム

4.1 システム概要

本研究における可視化システムの概略図を図 2 に示す。システムとしては、ノード情報を読み取り、それらの情報を OpenGL を用い可視化、さらに移動情報を入力することで、3D 空間内を移動できるようになっている。また、ユーザビリティを考慮した機能も実装している。以下に主な機能を説明する。

4.2 複雑ネットワーク情報空間の没入型探索機能

3D 空間内の移動操作を可能にするため、安価で入手が容易なバランス Wii ボードを利用してシステムを構築した(図 3)。ユーザがバランス Wii ボード上で体重をかける重心移動の動作で、前進、方向転換などが行える様、インタラクションを作成した。また、足踏みによる歩行動作を検知するモードも構築したため、ウォークスルーの機能も備えている。

4.3 インタラクション機能

- 球など幾何学的な形状のみで 3 次元空間内のノードを表現するだけでは、内容を直感的に理解できない。最初から情報が過剰すぎることも避けることが望ましいと思われ、必要最小限のノードの情報を提示する機能が望まれる。そこで、あるノードに接近することがそのノードの最小限の情報を提示する意図として捉えられる直感的意図と合致し易い提示機能を実装した。これにより、視覚情報だけでノードの最低限の情報を把握でき、より詳細に内部情報を参照するかを検討させる余地を与えることができる。
- ノードをクリックすれば対象ノードのホームページ内容を別の画面に提示する機能を実装した。これにより、情報の段階的詳細提示が可能となった。また、可視化におけるノード間リンクの表示は、見難くなる場合が多い。そこで、注目するノードのリンク情報を任意に表示できる機能を実装した。
- 情報拡散モデルに基づくレイアウト法は、ノード間の重なりを配慮するが、ノードが込み入った領域は少なからず存在する。そのような箇所は見難く、関係が理解しにくい。そこで、込み入ったノード群の各ノードの位置を整理できるよう、選択した任意のノードを移動させるエディット機能を実装した。

5. 実験

提案システムを使用し、人物ネットワーク、新聞記事ネットワークの 3D 空間での配置、及びインタラクション機能を使用し、ネットワークに没入することで、ネットワークの細部を探索する。具体的には影響最大化問題によって選ばれたノード群(主要ノード)から第一近傍のノードを探索していく。その上で発見したことを実験結果とする。

5.1 人物ネットワーク

ネットワークとして、人物ネットワークを構成する。まず、ウィキペディア内の同一ページにおいて、2 人の人物が共起しているページ数をカウントしていき、一定ページ以上の共起がある場合、双方向リンクとみなす。これにより、ネットワーク内での人物リストを構成する。

5.1.1 実験データ

可視化の対象として、日本のウィキペディアの「人物リスト」から構成したネットワークを使用した。6 ページ以上の共起がある場合、双方向リンクとした。ノード総数は 9,481、有向リンクは 245,044 であった。また、クラスタ係数値は 0.39 であり、ウィキペディアネットワークはソーシャルネットワークと同様な特性を持つと言える。主要ノード数は 10 とする。

5.1.2 実験結果

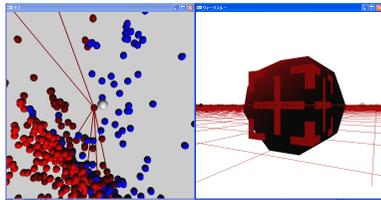


図 4 特徴的な位置にあるノード
Fig. 4 A node at special position



図 5 手塚治虫と水木しげるの位置
Fig. 5 The position of Osamu Tezuka
and Shigeru Mizuki

基本的な探索方法としては、ノードから表示されたリンクを辿ることで、あるノードに到達し、そのノードや周囲のノードを調べていく。この方法で探索していくと「六角 定頼」(図 4) というノードを発見した。「六角 定頼」というノードは主要ノード「織田信長」の色属性(茶色)を持っているが、配置的には「豊臣秀吉」のグループ(青色)近くにある。調べてみると以下の様な事がわかった。

- (1) 足利家は「織田信長」と因縁がある
 - (2) 「六角 定頼」は足利に仕え、楽市楽座を創設した人物とされている
 - (3) 楽市楽座は「豊臣秀吉」が発展させたとされている
- (1) により、「織田信長」は足利家と多くのリンクを持っているので、同様な色属性であると考えられる。(2) より「六角 定頼」は足利家と同様な色属性を持つというのは自然である。また(3) で「豊臣秀吉」からも影響を受けているため、このような特徴的な位置に配置されているのではないかと考えられる。これらのことから配置位置と色情報でそのノードの特性がある程度把握できることが分かった。

また、主要ノード「手塚治虫」の第一近傍を探索すると、「水木しげる(ゲゲゲの鬼太郎の作者)」であった。「水木しげる」と「手塚治虫」は同じグループであり、共に有名人である。この二人は、第一近傍で結ばれる関係であるため、本来ならば近くに位置してもおかしくは

ない。しかしながら、離れた位置にいる(図 5)。「水木しげる」の第一近傍ノードは、位置的にも近い位置にある。これらのノードからの情報拡散が反映されるため、「手塚治虫」から離れた場所に位置していると考えられる。この場合、「水木しげる」は主要ノードになり得た可能性があると考えられる。こうした知見は、選択的にリンクを可視化し、色のグループとの関係を同時に検証することで理解できる。また、こうした関係を他の場所で見つけることができれば、影響度が高いにも関わらず埋もれた主要ノードの候補を発見する手掛かりになる可能性がある。

5.2 新聞記事ネットワーク

もう一つの複雑ネットワークとして、文書間類似度に基づいて新聞記事ネットワークを構成する。TFIDF によって抽出された単語を要素とする単語ベクトル γ によって COS 類似度に基づく文書間距離 S を計算する。また、この文書間類似度に重みとして時間フィルタを導入する。時間フィルタは文書 m と文書 n の生成時間の差が a 日以内なら、時間情報面での類似度を \max に、 b 日以上なら 0 とした T を導入する。

$$S_{m,n} = 1 - \cos(\gamma_m \cdot \gamma_n) \cdot T \quad (1)$$

S の値がしきい値以下であれば文書 m 、文書 n には双方向リンクがあるとみなす。

5.2.1 実験データ

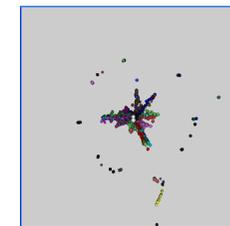


図 6 新聞記事ネットワークの可視化結果
Fig. 6 Result of Newspaper Network Visualization

実験データとして 1994 年の 1 月から 6 月に記載された毎日新聞の国際面記事データを用いた。文書間類似度のしきい値を 0.75 とした。総文書数は 2695、有向リンク 30,266 であった。また、クラスタ係数値は 0.46 であり、毎日新聞記事ネットワークはソーシャルネットワークと同様な特性を持つと言える。主要ノードは 30 とする。また、毎日新聞記事には人によって分けられた正解データが存在し、518 記事が 22 個のクラスタに分けられている。

5.2.2 実験結果

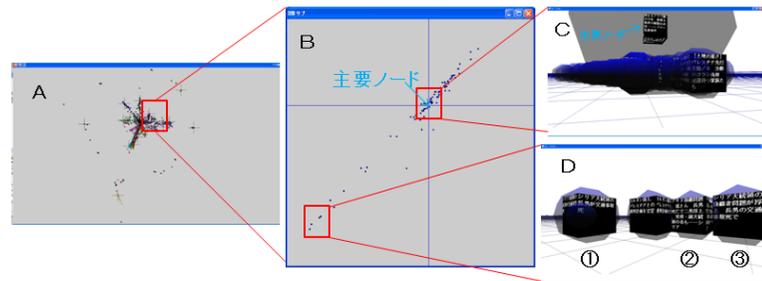


図 7 3次元没入型情報空間における情報探索の例

Fig. 7 An example of information exploration in the 3D immersive information space

図 7 は、主要ノードから同グループを探索した一例である。まず、俯瞰視図 7A において、一部を拡大したものが 7B である。この図 7B の中央上部にある縦横の直線の交点に、あるグループの主要ノードがあり、その周囲にある多数のノード群をユーザー視点で見たものが図 7C である。図 7C の空中に浮かんだノードが主要ノードである。主要ノードは青色であり、記事タイトルは「深まるアラブ側の不振、原理主義者の報復は必至へブロン乱射事件」であった。主要ノードの周辺ノードの付近を探検すると、パレスチナ問題を取り上げたノードが数多く存在した。次に、主要ノードから離れた位置にあるノードを調べたところ、図 7D の様な、関連性の高いノード群が見つかった。図 7D のノードは、「シリア大統領の長男が死亡」「後継者問題が浮上」「後継者に二男が浮上」といった内容の記事であった。地理的には、パレスチナの一部にシリアは含まれているが、パレスチナ問題と、この記事内容は直接的な関わりは薄いと考えられる。このことから、位置的に遠くあるノードほど、関係性が薄くなるように配置されていることがわかり、可視化空間の情報拡散に基づく影響度の関係がうまく反映できている一例と言える。

また、知識発見の例として、探検によってソビエトが崩壊した後のヨーロッパにおける各国の関係の変化に関する時代背景が発見できた例を図 8 に示す。図 8 左は、ノード (c) から出発して、主要ノード (a), (b) とのリンクを持つが、主要ノード (c) にはリンクを持たない。また、ノード (c') の色属性は (c) と同じである。(c) に対しリンクを持たないノード (c') を調べると以下の様なことが分かった。

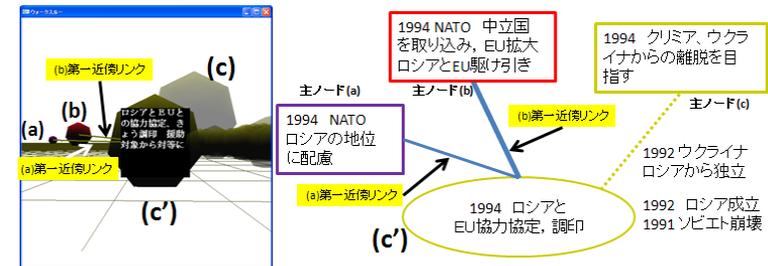


図 8 知識発見の例

Fig. 8 An example of knowledge discovery

- 1991 年 ソビエト崩壊
- 1992 年 ウクライナがロシアから独立
- 1994 年 NATO が中立国を取り込む (b)
- 1994 年 ロシアと EU 協力協定、調印 (c')
- 1994 年 クリミア、ウクライナからの離脱を目指す (c)
- 1994 年 NATO, ロシアの地位に配慮 (a)

また、クリミアはウクライナの一部であるので、クリミアを調べていくと、過去、ロシア領の一部であり、何度も独立を試みているという内容に辿りついた。ロシアが EU との協力協定に調印したことから、クリミアで再度独立の運動が出たとしてもおかしくはない。この様に、一つの記事から探索することで、歴史的な流れや、時代背景の一部を知ることが可能であった。

6. まとめ

効果的な可視化や直感的な操作は複雑ネットワークから知識を発見する上で重要なことである。我々は複雑ネットワーク情報空間の没入型インタラクティブ可視化システムを提案した。本システムでは、複雑ネットワークを情報拡散過程に基づき 3D 空間に可視化する。さらに、情報空間を探検するため没入できる環境を提供し、直感的に複雑ネットワークを操作して、ノードの情報を調べることができる。人物ネットワークデータと新聞記事ネットワークデータを用いて、提案システムの有効性を検証した。没入型インタラクティブ探索機能を用いて可視化されたネットワークを探索することで、可視化だけでは見えてこない知識を発見することができた。

参 考 文 献

- 1) 石井裕:タンジブル・ビット:情報と物理世界を融合する新しいユーザ・インタフェース・デザイン (〈特集〉仮想と現実の融合), 情報処理, Vol.43, No.3, pp.222-229 (2002)
- 2) 三末和夫, ネットワーク可視化技術 大規模ネットワークと動的ネットワークへの挑戦, 電子情報通信学会誌, Vol.92, No.2, pp.112-117 (2009)
- 3) Stuart K. Card, Jock D. Mackinlay, Ben Shneiderman: Readings in information visualization: using vision to think, Morgan Kaufmann (1999)
- 4) Henrik R. Nagel, Erik Granum, Soren Bovbjerg and Michael Vittrup: Immersive Visual Data Mining: The 3DVDM Approach, Visual data mining: theory, techniques and tools for visual analytics, Springer, pp.281-309 (2008)
- 5) K.Saito, M.Kimura and H.Motoda: Effective visualization of information diffusion process over complex networks, In W. Daelemans, B. Goethals and K. Morik (Eds.), *Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2008)*, LNAI 5212, pp.326-341, 2008.
- 6) 水田昌孝, 熊野雅仁, 木村昌弘: LDA と時間フィルタを用いた文書ストリームからのバースト潜在トピック抽出, 人工知能学会第 87 回知識ベースシステム研究会, pp.25-30, 2010.
- 7) D.Kempe, J.Kleinberg and E.Tardos: Maximizing the spread of influence through a social network, *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2003)*, pp.137-146 (2003) 2003
- 8) M.Kimura, K.Saito, R.Nakano and H.Motoda: Extracting influential nodes on a social network for information diffusion, *Data Mining and Knowledge Discovery*, Vol.20, pp.70-97, 2010.
- 9) D.M.Blei, A.Y.Ng and M.I.Jordan: Latent Dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp.993-1002, 2003.
- 10) Shneiderman B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization, *Proceedings of IEEE Symposium on Visual Languages ' 96*, pp.336-343 (1996)