筑波大学大学院博士課程

システム情報工学研究科修士論文

書誌情報ネットワークのビュー操作に着目した論文データ分析支援ツール

小池諭

(コンピュータサイエンス専攻)

指導教員　　三末和男

２００９ 年　３ 月

概要

研究分野のサーベイなど、論文データから有用な情報を得るための分析作業を行う場面がしばしばある。そのような作業を支援するための研究が数多く行われているが、それらは「論文データの一部の関係情報にしか着目していない」「分析作業に付随する情報探索を支援できていない」といった理由から、必ずしも満足いくものではなかった。

本研究では、論文データ分析における情報探索の特徴を分析し、論文データ分析に適したネットワークビューの構築及び操作体系を開発した。その上で、論文データに含まれる複数の関係に着目し、分析のための情報探索を支援するツール「STANICOV」を開発し、その有用性の評価を行った。STANICOV は、論文データの書誌情報が為すネットワークのノードをその属性ごとに領域を分割して可視化し、ユーザの操作によってインタラクティブにネットワークのビューを構築していくものである。論文データから分析者が得たい情報に合わせたビューの提供を可能にすると共に、分析における情報探索をノード操作という直感的な方法で実現する。

STANICOV を用いて論文データの分析を行う実験を行った結果、STANICOV は論文データ分析における情報探索に適したビュー操作体系を有し、分析に有用であることが示された。

本研究の新規性は、情報探索のためのネットワーク操作体系を構築したこと及びその操作体系を直感的な方法で実現するインタフェースを開発したことである。複数の関係からなるネットワークは実世界に数多く存在し、その分析に対する必要性も高いことから、本研究の成果はそれらネットワークの分析支援の基礎技術となる可能性がある。

目次

第 1 章　序論 ……1
1.1　論文データとその分析 ……1
1.2　論文データに含まれる関係情報とネットワーク ……1
1.2.1　論文データに含まれる関係情報 ……1
1.2.2　論文データが為すネットワーク ……1
1.3　従来研究 ……2
1.4　従来手法の問題点 ……3
1.5　本研究の目的 ……3
1.6　本研究の貢献 ……3
1.7　本論文の構成 ……4
第 2 章　論文データ分析を支援するための要件 ……5
2.1　論文データ分析における情報探索 ……5
2.1.1　分析において得たい情報とその分類 ……5
2.1.2　分析における情報探索の特徴 ……5
2.2　分析支援のための要件 ……6
第 3 章　インタラクティブなネットワーク構築 ……7
3.1　書誌情報ネットワークの可視化 ……7
3.2　得たい情報に応じたネットワークビューの構築 ……8
第 4 章　書誌情報ネットワーク ……9
4.1　論文データに含まれる関係情報 ……9
4.2　書誌情報ネットワークとは ……10
4.3　書誌情報ネットワークの分類 ……10
4.3.1　把握すべき関係情報の分類 ……10
4.3.2　ネットワークとの対応付け ……11
4.4　ネットワークを構築するための操作体系 ……11
4.4.1　局所的な情報を得るためのネットワークの構築 ……11
4.4.2　概観的な情報を得るためのネットワークの構築 ……12
4.5　書誌情報ネットワークの図表現 ……13
第 5 章　論文データ分析支援ツール「STANICOV」 ……15
5.1　ツールの設計方針 ……15
5.2　ツール概観 ……17
5.3　ツールの機能 ……17
5.3.1　ノードの選択及び移動 ……17
5.3.2　ネットワークのレイアウト及び操作の補助機能 ……21
5.3.3　情報探索の支援 ……23
5.4　ツール実装 ……25
5.4.1　システム構成 ……25
5.4.2　実装言語と入力データ形式 ……25
5.4.3　ネットワークのレイアウトアルゴリズム ……26

5.5　ツール利用例 ……27
第 6 章　評価実験と考察 ……31
6.1　評価実験 ……31
6.1.1　実験の目的 ……31
6.1.2　実験概要 ……31
6.1.3　被験者 ……31
6.1.4　手順 ……31
6.1.5　結果 ……32
6.2　考察 ……34
6.2.1　ツールの有用性 ……34
6.2.2　各機能の使用状況及び有用性 ……34
6.2.3　改善すべき点 ……34
6.3　今後の展望 ……35
6.3.1　他のデータへの適用可能性 ……35
6.3.2　ツールとしての発展案 ……35
第 7 章　関連研究と本研究の位置付け ……37
7.1　複数の関係に着目したネットワーク把握支援 ……37
7.2　ネットワーク可視化におけるインタラクション ……38
7.3　本研究の位置付け ……39
第 8 章　結論 ……41
謝辞 ……42
参考文献 ……43

図目次

図 3.1　友人関係のネットワークを網図表現で可視化した例 ……7
図 3.2　インタラクティブなネットワーク構築による情報探索支援 ……8
図 4.1　ある書誌情報ネットワークの全体構造 ……13
図 4.2　分類したネットワークの可視化例 ……14
図 5.1　フィールドの分割によるネットワーク構築 ……16
図 5.2　ノード属性の組み合わせとそれぞれの関係 ……16
図 5.3　STANICOVの初期画面 ……17
図 5.4　ノードの移動 ……18
図 5.5　隣接関係によるノードの選択 ……19
図 5.6　領域によるノードの選択 ……20
図 5.7　フィールド全体の選択 ……20
図 5.8　属性による移動ノードのフィルタリング ……21
図 5.9　同属性ノード間のエッジを隠した例 ……22
図 5.10　操作パネル……23
図 5.11　キーワード「input」で検索した結果 ……24
図 5.12　STANICOVのシステム構成及びデータフロー ……25
図 5.13　キーワード検索の結果……27
図 5.14　ノードを著者のフィールドに移動した結果 ……28
図 5.15　「information visualization」に関する著者のネットワーク ……28
図 5.16　ノードをキーワードのフィールドに移動した結果 ……29
図 5.17　著者とキーワードからなるネットワーク ……29
図 5.18「information visualization」と「interaction」に関するネットワーク ‥30
図 6.1　各機能のノード操作に占める割合 ……33
図 6.2　Mindmapの例 ……36
図 7.1　本研究の位置付け ……40

表目次

表 4.1 論文データの例 9
表 4.2 関係情報の分類例 10
表 5.1 操作体系とツールでの操作の対応表 15
表 5.2 注目度への指標の割り当て 24
表 6.1 各被験者のアンケートへの回答 32
表 6.2 各機能に対する評価 33

第1章　序論

1.1　論文データとその分析

情報技術の発展により、論文は電子化され、電子データとして保存することができるようになった。現在、ディジタルライブラリなどで多くの論文が電子的に蓄積されている。論文は、本文以外に「論文を執筆した著者の名前」や「論文が発表された会議の名前」、「論文検索用のキーワード」などいわゆる書誌情報を含む。ディジタルライブラリなどで論文が蓄積される場合、それら書誌情報も共に蓄積され、論文検索などに利用されている。本論文では、蓄積された論文の書誌情報を「論文データ」と呼ぶ。

研究分野のサーベイをする場面などにおいて、「著名な研究者が執筆した論文」や「興味のある研究トピックに関連する論文やその著者」などを知りたい場面がある。また、それらを総合して「研究分野の主要な研究トピックに対し、誰がどんなアプローチで研究をしているのか」といったことを知りたい場面がある。それら有益な情報の多くは論文の書誌情報に基づくものであるため、論文データから情報を得ることができる。本論文では、論文データから有益な情報を得るための知的活動を「論文データの分析」と呼ぶ。

1.2　論文データに含まれる関係情報とネットワーク

1.2.1　論文データに含まれる関係情報

論文の書誌情報はそのそれぞれの間に関係を持っている。例えば、以下のような関係がある。

- 論文とその著者との間には「その著者に執筆された／その論文を執筆した」という関係がある
- 論文が複数の著者によって執筆されたものである場合、その著者らの間には「論文を共著した」という関係がある
- 論文が他の論文を引用している場合、その論文の間には「引用した／引用された」という関係がある
- 論文と論文中のキーワードとの間には「そのキーワードを使用した／その論文に登場した」という関係がある

これら関係は論文データに文字や数値として明示的に記録されているわけではないが、情報としては含まれていると言える。1.1 節で挙げたような有益な情報の多くは論文データに含まれる関係情報を把握することで得られる。

1.2.2　論文データが為すネットワーク

論文データに含まれる関係情報はネットワークを為す。ネットワークの構造を把握することで関係情報を把握しやすくなり、論文データに含まれる有益な情報を発見しやすくなる。

本研究で扱うネットワークとは「オブジェクト(人や物)とオブジェクトとの関係による構造体」である。ネットワークはグラフとして表現することができる。すなわち、オブジェクトをノード(頂点)に対応させ、関係をエッジ(辺)に対応させて表現できる。ノードやエッジの付加的な情報(名前、ラベル、カテゴリなど)は属性として表現される。

ネットワークデータは数値や文字のデータであり、人間がその構造を把握することは難しい。そのため、ネットワーク把握支援のアプローチとして「可視化」がしばしば用いられる。ネットワークの要素を点で表し、要素同士のつながりを線で表現するといった視覚表現を用いることで、人間にとって理解しやすいものとなる。

1.3 従来研究

論文データが為す書誌情報ネットワークの構造を可視化することで有用な情報の発見を支援する研究はこれまで数多く行われてきた。Newman[4][5]や LaRowe ら[7]は、論文の共著関係が為すネットワークに着目し、ネットワークのパターンを分析することで研究者の研究活動の推移や研究者間のつながりなどを把握できることを示した。また、Ke ら[6]や Borner ら[18]は、論文の引用関係が為すネットワークから研究分野に影響を与えた論文などを発見できることを示した。

書誌情報ネットワークの構造把握を支援するための可視化手法やツールも数多く開発されている。

Thiel らは、論文中のキーワードと時間との関係が為すネットワークに着目し、多次元スケーリングアルゴリズム（MDS）を用いて可視化する手法を開発した[17]。これは、論文中の単語の特徴量を多次元ベクトルで表現し、MDS を用いて 2 次元平面状に配置するもので、単語同士の距離関係から同じ時間に使われることの多い単語群を知ることができる。

Huang らは、論文の共著関係が為すネットワークに着目し、それぞれの著者の時間毎の共著関係の推移を可視化する手法「InterRing Visualizer」を開発した[8]。時間を放射軸に置く同心円状の円グラフで著者それぞれの共著関係を表すことで、密接に関わりを持つ著者の推移を知ることができる。

Henry らは、論文の共著関係が為すネットワークをソーシャルネットワークと捉え、ネットワーク図と行列表現を組み合わせることで複雑なネットワークを表現する手法を開発している。ネットワーク図と行列表現を並列し、それぞれの表現に対してユーザが行った操作へのフィードバックを互いに同期させて複雑なネットワークの構造把握を支援するツール「MatrixExplorer」を開発した[16]。また、ネットワーク図と行列表現をハイブリッドで用いることでネットワークの概観構造と局所構造両方の把握を支援する表現手法「NodeTrix」を開発した[12]。

Ichise らは、論文中で同じキーワードを使っている著者同士の関係が為すネットワークに着目し、newsmap、折れ線グラフ、ネットワーク図を連携してネットワークの把握を支援するツールを開発した[2]。キーワードの出現頻度や著者同士の関係など複数の指標に対して適切な表現を提供することで著者同士がどのようなキーワードでつながっているのかを把握しやすくしている。

1.4　従来手法の問題点

1.2.1 節で挙げたように、論文データの書誌情報間の関係には複数の種類がある。そのため、論文データが為すネットワークは複数の関係から構成されている。論文データに含まれるそれぞれの関係に着目してネットワークの構造を把握することで書誌情報間の関係を俯瞰的に知ることができ、論文データ分析において多くの情報を獲得できるようになる。しかしながら、従来手法では以下の問題により十分に支援できていないのが現状である。

ネットワークを構成する関係に複数の種類があることを考慮していない

論文データ分析を目的とした従来研究は主に「共著関係」や「引用関係」など単一の関係からなるネットワークの可視化手法についての研究であり、関係に複数の種類があることを考慮したネットワーク把握支援についてはあまり研究されてこなかった。

論文データ分析における情報探索を支援できていない

論文データ分析においては書誌情報間の局所的な関係を把握しながら情報獲得のための探索をしていくことが重要である。しかしながら、従来は主にネットワークの概観から情報探索をしていく手法について研究されており、論文データ分析の特徴を考慮した情報探索支援技術についてはあまり研究されてこなかった。

本研究では、これら 2 点を解決するネットワーク探索手法を提案、開発する。

1.5　本研究の目的

本研究では、論文データに含まれる有益な情報の獲得を支援することを目的とし、論文データが為すネットワークの把握を支援するツールを開発する。論文データのネットワークが複数の関係からなることを考慮したネットワーク把握支援を行うと共に、論文データ分析の特徴を考慮した情報探索の支援を行う。

1.6　本研究の貢献

本研究の貢献は、複数の関係からなるネットワークの分析を支援するためのネットワーク探索手法を開発したことである。

従来、論文データの分析支援を目的とした研究では、論文データが持つ特定の関係に着目し、その把握を支援する研究が多かった。本研究は、論文データ分析を行う上で書誌情報間の関係に複数の種類があることを考慮し、それら関係が為すネットワークの把握支援を行った。このことによって書誌情報間の関係を俯瞰的に把握することができるようになり、論文データからより多くの情報を獲得することが可能となる。

また、従来のネットワークの把握支援技術の多くはネットワークの概観を把握することに主眼を置いており、ネットワークの探索技術についても概観から得られる情報をきっかけに情報探索をしていく手法が多かった。本研究では、論文データ分析の特徴を考慮し、ネットワークの局所的な構造からの情報探索が可能なネットワーク探索手法を開発した。このことによって、より論文データ分析に適したネットワーク探索を行うことができ、論

文データ分析を行いやすくなる。

1.7 本論文の構成

第 2 章では、論文データ分析の特徴を分析した上で論文データ分析を支援するための要件を述べる。第 3 章では、第 2 章の要件に基づいて、論文データ分析支援において得たい情報に応じたネットワークビューを提供することを提案する。第 4 章では、論文データから情報を獲得するために把握すべき関係情報とそれが為すネットワークについて述べ、ネットワークを得るための操作体系について述べる。第 5 章では、第 4 章で分類したネットワークの操作体系を搭載した論文データ分析支援ツール「STANICOV」について述べる。第 6 章では、「STANICOV」の有用性を評価し、それに対する考察を述べる。第 7 章で本研究に関連する研究及び本研究の位置付けについて述べ、第 8 章でまとめる。

第2章　論文データ分析を支援するための要件

本章では、論文データ分析における情報探索の特徴を分析した上で、分析を支援するための要件を挙げる。

2.1　論文データ分析における情報探索

2.1.1　分析において得たい情報とその分類

論文データ分析の目的は様々であり、それら目的によって得たい情報は異なる。論文データ分析の例として以下のようなものが挙げられる。

- 興味のある研究トピックについて知るために、研究トピックを表すキーワードを使っている論文や著者を調べる
- 研究分野の近年の動向について知るために、近年発表された論文に使われているキーワードを調べる
- 研究分野で行われてきた研究の流れについて知るために、研究分野の主要なトピックに関連する論文の引用関係をさかのぼって調べる
- 研究分野における研究者間のつながりについて調べるために、論文の共著関係を調べる

それぞれの目的によって得たい情報は様々であるが、これらは局所的な情報と概観的な情報の 2 つに分類できる。局所的な情報とは「特定の書誌情報に関連する書誌情報」であり、例えば「研究者Aが発表した論文」や「キーワードBを使っている論文や著者」などがそれにあたる。概観的な情報とはこれら局所的な情報が集まることでわかる大局的な情報であり、「研究分野の近年の動向」や「研究者間のつながり」などがこれにあたる。この概観的な情報の獲得が論文データ分析の目的であることも多い。

2.1.2　分析における情報探索の特徴

論文データ分析を行う上で、分析者は論文データから有益な情報を得るための情報探索を行う。2.1.1 節で分類した 2 種類の情報によって情報探索の方法は異なり、それぞれの特徴として以下の 2 点が挙げられる。

局所的な情報の探索：分析者は関係の意味に着目した情報探索を行う

局所的な情報を得るためには論文データが持つ関係情報を把握することが必要であるが、その際書誌情報間の関係に複数の種類があることが重要である。例えば、「論文が使っているキーワード」や「著者が執筆した論文」などを把握することで得たい情報を得られる場面は多い。このような場面において、分析者は「キーワードの使用関係」や「論文の執筆関係」など、書誌情報間の関係の意味に着目して情報獲得を行っていると言える。

概観的な情報の探索：分析者は連鎖的な情報探索を行う

概観的な情報を得るために局所的な情報の探索を繰り返す場面がしばしばある。例えば、「興味のある研究トピックに関連する論文を調べ、次にそれら論文の著者を調べ、さらにそれら著者が論文の中でどのようなキーワードを使っているかを調べる」といった情報獲得を繰り返すことで、「トピックに関して誰がどんな技術を用いて研究を行っているのか」といったことを知ることができる。このような場面において、分析者は書誌情報間の異なる関係に着目しながら連鎖的な情報探索を行っている。

2.2 分析支援のための要件

本研究では論文データ分析支援のためのツールを開発するにあたり、ツールの要件として以下の 2 点を挙げる。

- 関係の意味に着目した情報探索の支援
- 連鎖的な情報探索の支援

論文データが持つ局所的な情報の探索を支援するために、それぞれの書誌情報間の関係の有無だけでなく、関係の意味（どのような関係があるのか）を明確に把握できるようにする必要がある。また、概観的な情報の探索を支援するために、局所的な情報を連鎖的に探索できるようにする必要がある。

第3章　インタラクティブなネットワーク構築

本章では、第 2 章で挙げた要件を基に本研究のアプローチについて述べる。3.1 節では本研究で採用する論文データ表現手法について述べ、3.2 節では論文データ分析における情報探索を支援するためのアイデアについて述べる。

3.1　書誌情報ネットワークの可視化

第 1 章で述べたように、論文データに含まれる関係情報はネットワークを為し、ネットワークを可視化することで有益な情報を発見しやすくなる。本研究では、論文データが為すネットワークを可視化することで論文データ分析を支援する。

論文データが為すネットワークをどのように可視化するかであるが、本研究ではネットワークを網図表現によって表現する。網図表現とは、ネットワークの要素を点で、要素間のつながりを線で表すといった表現形式である[15](図 3.1)。

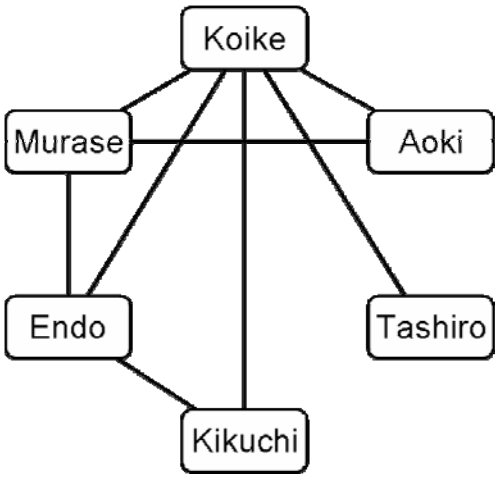


図 3.1　友人関係のネットワークを網図表現で可視化した例

論文データが為すネットワークを網図表現によって可視化することの利点は 2 つある。

1 つ目は、論文データが持つ関係情報を視覚的に捉えることができるという点である。論文データが為すネットワークの網図表現によって、書誌情報間の関係が線のつながりとして視覚的に理解できるようになる。

2 つ目は、論文データ分析における情報探索を行いやすい表現形式であるという点である。論文データ分析における情報探索とは、書誌情報間の関係構造すなわちネットワークを探索することと言える。網図表現はネットワークのパスを辿るのに適した表現形式であることから[29]、論文データ分析における情報探索にも適していると考えられる。

3.2　得たい情報に応じたネットワークビューの構築

論文データ分析において得たい情報によって把握すべき関係は異なるため、ネットワークビューは関係の意味が明確にわかるものである必要がある。また、分析者の認知的な負荷を軽減するためにも、得たい情報それぞれに対して過不足のない情報量でビューを提供することが好ましい。そこで筆者は、得たい情報に応じて把握すべき関係や情報量を決め、それに応じたネットワークビューを分析者の手で構築できればよいのではないかと考えた。

情報可視化において、可視化された図とのインタラクションを導入することで図から効果的に情報を獲得できるようになる。これは、見る者が図を操作することで能動的に情報獲得を行っているためである。この効果は書誌情報ネットワークの可視化においても有効であると考えられる。分析者による情報獲得の意図を持った操作によって関係情報を把握しやすいネットワークビューをインタラクティブに構築することで、能動的な情報探索が行えるようになると考えられる。また、概観的な情報を得るための連鎖的な情報探索を行う場合にもインタラクティブなネットワーク構築は効果的であると考えられる。得たい情報に対して適切なネットワークビューを閲覧し、その中から情報探索のきっかけを得て、それを基に新たなネットワークビューを構築するといった手順を繰り返すことで、書誌情報間の関係を大局的に知ることができると考えられる(図 3.2)。

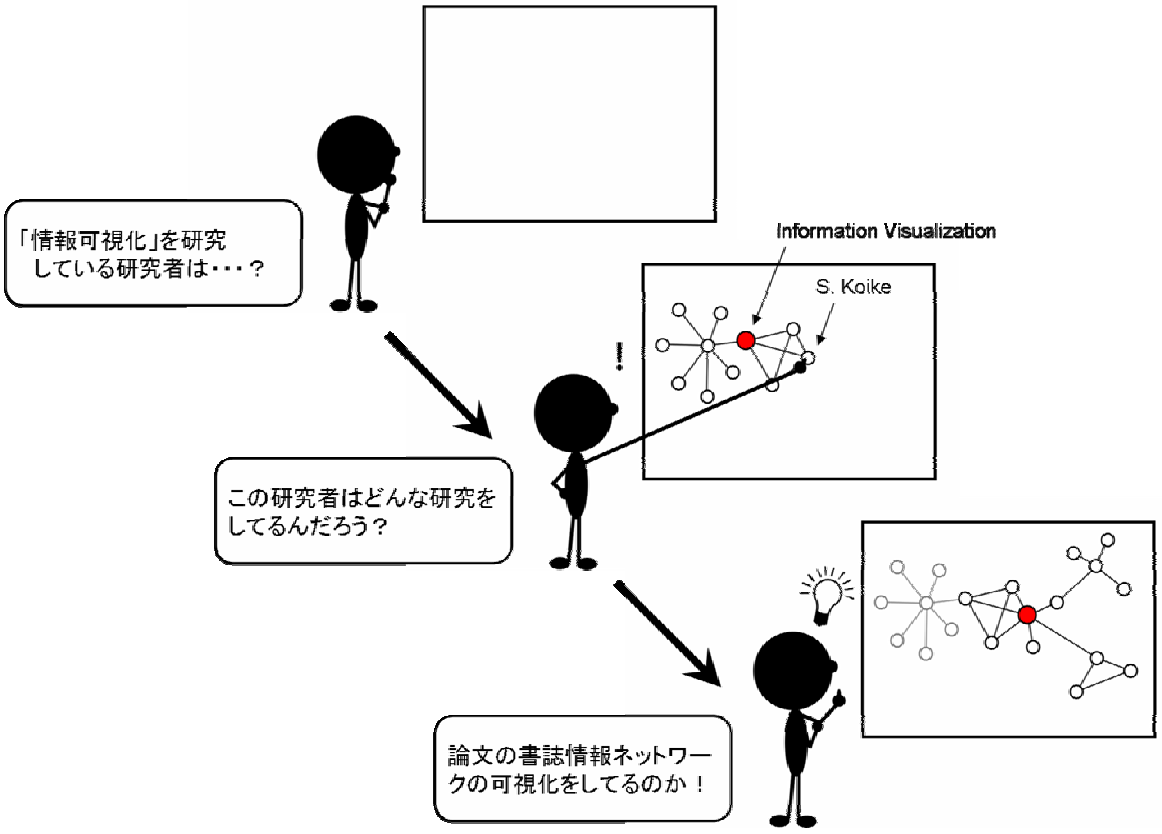


図 3.2　インタラクティブなネットワーク構築による情報探索支援

第4章　書誌情報ネットワーク

本章では、本研究で可視化対象とする書誌情報ネットワークとその分類について述べる。まず、論文データに含まれる関係情報について述べる。次に、書誌情報が為すネットワークを定式化する。最後に、得たい情報をネットワークで表現するために書誌情報ネットワークを分類し、それぞれのネットワークを得るために必要な操作について述べる。

4.1　論文データに含まれる関係情報

論文データを表形式で表現すると、例えば表 4.1 のようになる。

表 4.1 論文データの例[1]

発表年	セッション	タイトル	著者	キーワード
2007	Tree and Graph Visualization	NodeTrix: a Hybrid Visualization of Social Networks	N. Henry, M.J.McGuffin	Network visualization, Matrix visualization
2006	Graph Exploration	MatrixExplorer: a Dual-Representation System to Explore Social Networks	N. Henry, J. D. Fekete	Network visualization, interactive clustering
2005	Interactive information visualization	prefuse: a toolkit for interactive information visualization	J.Heer, S.K.Card, J.A.Landay	information visualization, user interfaces, navigation
・・・	・・・	・・・	・・・	・・・

ひとつの論文の書誌情報は「発表された会議」や「タイトル」、「著者」など意味的な属性を持つ各要素からなり、そのそれぞれの間に意味の違う関係がある。表 4.1 の例では、例えば 2 人の著者「N. Henry」と「M. J. McGuffin」との間には「論文を共著した」という関係があり、論文「MatrixExplorer: a Hybrid Visualization of Social Networks」とキーワード「Matrix visualization」との間には「論文にキーワードが使われた（論文の中に登場した）」という関係がある。また、著者「N. Henry」やキーワード「Network visualization」のように複数の論文で登場する要素もあり、それらは複数の論文にわたってその書誌情報の要素との間に関係を持つ。

[1] このデータは[12][16][31]の書誌情報を抜粋したものである。

4.2 書誌情報ネットワークとは

4.1 節で述べた書誌情報間の関係はネットワークを構成する。本研究では、論文データにおける書誌情報の各要素が為す関係構造を「書誌情報ネットワーク」と呼ぶ。

書誌情報ネットワークは、書誌情報の各要素をノード、要素間の関係をエッジとしたグラフとして表現することができる（式 4.1)。式中のGはグラフ、Vはノード集合、Eはエッジ集合である。ノードは「著者」「キーワード」など書誌情報が持つ意味的な属性ごとにn個の排他的な集合に分けられる。nは書誌情報が持つ属性の数である。

$$
\begin{aligned}
&G = (V, E) \\
&V = V_1 \cup V_2 \cup \cdots \cup V_n \\
&E \subseteq \left\{ \{v, w\} \middle| v, w \in V \right\}
\end{aligned}
$$

(式 4.1)

4.3 書誌情報ネットワークの分類

4.3.1 把握すべき関係情報の分類

第 2 章で述べたように、論文データ分析で得たい情報には局所的な情報と概観的な情報の 2 種類がある。局所的な情報を得るために把握すべき関係情報は「書誌情報の特定の要素に関連する書誌情報及びそれらの間の関係の意味」であり、概観的な情報を得るために把握すべき関係情報はそれら局所的な関係情報の集まりであると言える。

関係情報の「関係の意味」と「それが局所的か概観的か」に着目したとき、論文データ分析において把握すべき関係情報は「着目する関係の数(単数か複数か)」と「関係を把握する範囲(書誌情報の特定の要素に関わる範囲か、複数の要素間にわたる範囲か)」という観点で分類できる。この観点で分類した例を表 4.2 に示す。

表 4.2 関係情報の分類例

		着目する関係の数	
		単数	複数
関係を把握する範囲	特定の要素に関わる範囲	**研究者 A が執筆した論文** 「論文の執筆関係」を 「研究者 A に関わる範囲」で把握	**キーワード B に関する論文や研究者** 「キーワードの使用関係」と 「キーワードの登場関係」を 「キーワード B に関わる範囲」で把握
	複数の要素にわたる範囲	**研究者間のつながり** 「論文の共著関係」を 「複数の研究者にわたる範囲」で把握	**研究分野の動向** 「執筆関係」や「ｷｰﾜｰﾄﾞの登場関係」などを「複数の要素にわたる範囲」で把握

4.3.2 ネットワークとの対応付け

論文データ分析を書誌情報ネットワークの観点から見ると、分析において「共著関係」や「執筆関係」など特定の関係に着目するとき、分析者は「著者と著者との関係」や「著者と論文との関係」など、ノードの属性の組み合わせに着目していると言える。また、書誌情報の中の特定の要素に関わる範囲で関係情報を把握するとき、分析者はネットワークの中の特定のノードに直接つながる(隣接する)ノードに着目していると言える。

このことから、論文データが持つ関係情報を「着目する関係の数」と「関係を把握する範囲」という観点で分類したとき、それぞれの関係情報を表すネットワークは以下のように分類できる。

A) ネットワークを構成する関係の数による分類
  A-1) 単一の関係からなるネットワーク
  A-2) 複数の関係からなるネットワーク
B) ネットワークの構成範囲による分類
  B-1) 特定のノードに隣接するノードからなるネットワーク
  B-2) 複数のノードの隣接関係からなるネットワーク

4.4 ネットワークを構築するための操作体系

分析において得たい情報に応じたネットワークを構築するには、分析者が「着目する関係」や「ネットワークの構成範囲」を定める必要がある。4.3.2 節より、ノード属性の組み合わせを定めることで着目する関係が決まり、着目すべき特定のノードを定めることでネットワークの構成範囲が決まる。このことから、ネットワーク構築のために分析者が行うべき入力は、「ノード属性の組み合わせ」と「ノード」である。

4.4.1 局所的な情報を得るためのネットワークの構築

論文データ分析における局所的な情報を探索する時に把握すべきネットワークは、分析者が着目する属性の組み合わせやノードを入力として、以下に示すネットワークで表される(式 4.2、式 4.3)。式中、入力であるvは着目するノード、λは着目するノード属性、Λは着目するノード属性の集合である。なお、式 4.2 ではノードvが持つ属性と入力である属性λの組み合わせによって、式 4.3 ではノードvが持つ属性と入力である属性集合Λの要素の組み合わせによって着目する関係を決定している。

[A-1×B-1 のネットワーク] 入力： v,λ

$$
\begin{aligned}
G_{A1,B1} &= \left(\{v\} \cup V_{v,\lambda}, E_{v,\lambda}\right) \\
V_{v,\lambda} &= \left\{w \in V_{\lambda} \middle| \{v,w\} \in E\right\} \\
E_{v,\lambda} &= \left\{\{v,w\} \in E \middle| w \in V_{v,\lambda}\right\}
\end{aligned}
$$

(式 4.2)

[A-2×B-1 のネットワーク]　入力：　$v, \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_m\}$

$$G_{A2,B1} = \left(\{v\} \cup V_{v,\Lambda}, E_{v,\Lambda}\right)$$
$$V_{v,\mathrm{A}} = \left\{ w \in \bigcup_{\lambda \in \Lambda} V_\lambda \middle| \{v, w\} \in E \right\}$$
$$E_{v,\Lambda} = \left\{ \{v, w\} \in E \middle| w \in V_{v,\Lambda} \right\}$$

(式 4.3)

4.4.2　　概観的な情報を得るためのネットワークの構築

式 4.2 のネットワークや式 4.3 のネットワークの構築を繰り返すことによって概観的な情報を表すネットワークを構築することができる。概観的な情報を得るために把握すべきネットワークは、分析者が着目する属性の組み合わせやノード集合を入力として、以下に示すネットワークで表される(式 4.4、式 4.5)。式中、A は着目するノード集合、λ は着目するノード属性、Λ は着目するノード属性の集合である。なお、式 4.4 でのノード集合 A は同じ属性を持つノードからなり、その属性と入力である属性 λ の組み合わせによって着目する関係を決定している。また、式 4.5 ではノード集合 A の要素が持つ属性と属性集合 Λ の要素の組み合わせによって着目する関係を決定している。

[A-1×B-2 のネットワーク]　入力：　$A \subseteq V_a, \lambda$

$$G_{A1,B2} = \left(A \cup V_{A,\lambda}, E_{A,\lambda}\right)$$
$$V_{A,\lambda} = \left\{ w \in V_\lambda \middle| \{v, w\} \in E, v \in A \right\}$$
$$E_{A,\lambda} = \left\{ \{v, w\} \in E \middle| w \in V_{A,\lambda} \right\}$$

(式 4.4)

[A-2×B-2 のネットワーク]　入力：　A, Λ

$$G_{A2,B2} = \left(A \cup V_{A,\Lambda}, E_{A,\Lambda}\right)$$
$$V_{A,\Lambda} = \left\{ w \in \bigcup_{\lambda \in \Lambda} V_\lambda \middle| \{v, w\} \in E, v \in A \right\}$$
$$E_{A,\Lambda} = \left\{ \{v, w\} \in E \middle| w \in V_{A,\Lambda} \right\}$$

(式 4.5)

4.5　　書誌情報ネットワークの図表現

本研究では、書誌情報ネットワークをネットワーク図によって可視化する。ネットワークのノードを円で表現し、エッジを線分で表現する。ここで、4.3.2 節で分類したネットワークの可視化例を挙げる。ある論文データの書誌情報が為すネットワークの全体構造を図 4.1 に示し、それを分類した例を図 4.2 に示す。図中の水色のノードは論文、薄赤色のノードは著者、黄色のノードはキーワードを表す。なお、このネットワークは可視化例を説明するために抜粋されたごく一部のデータからなるものである。

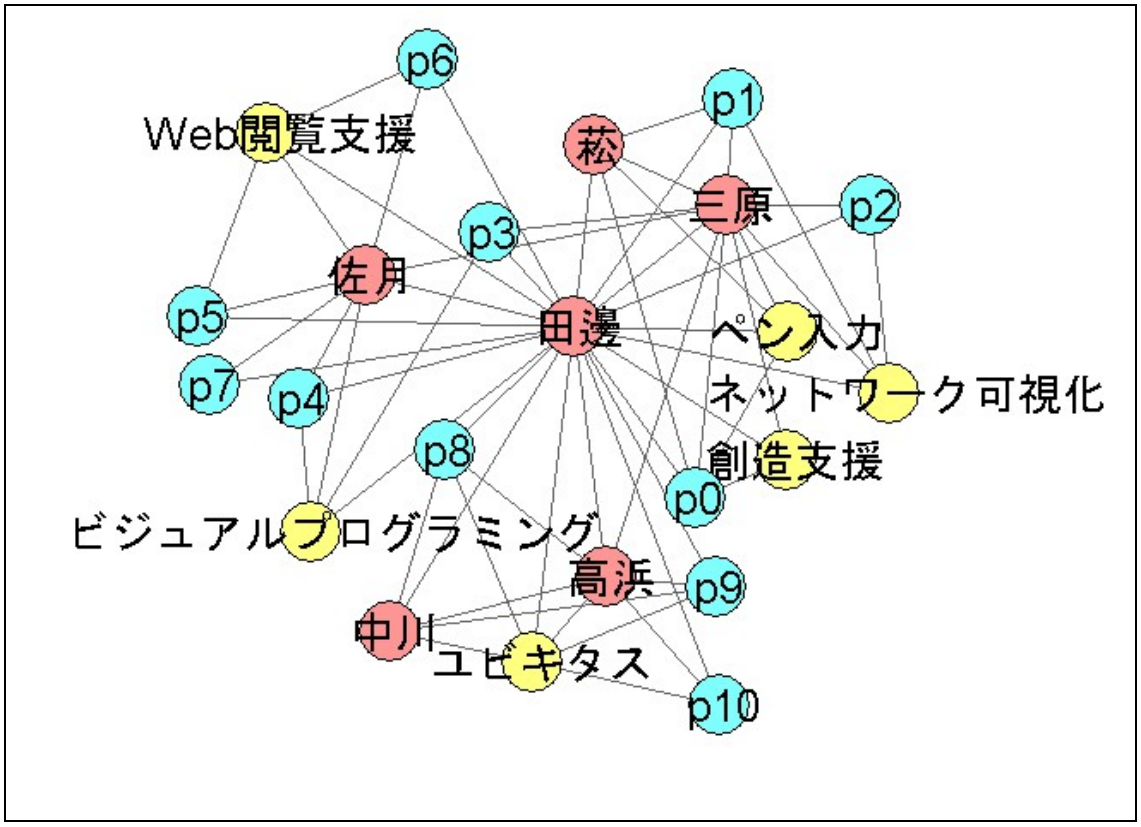
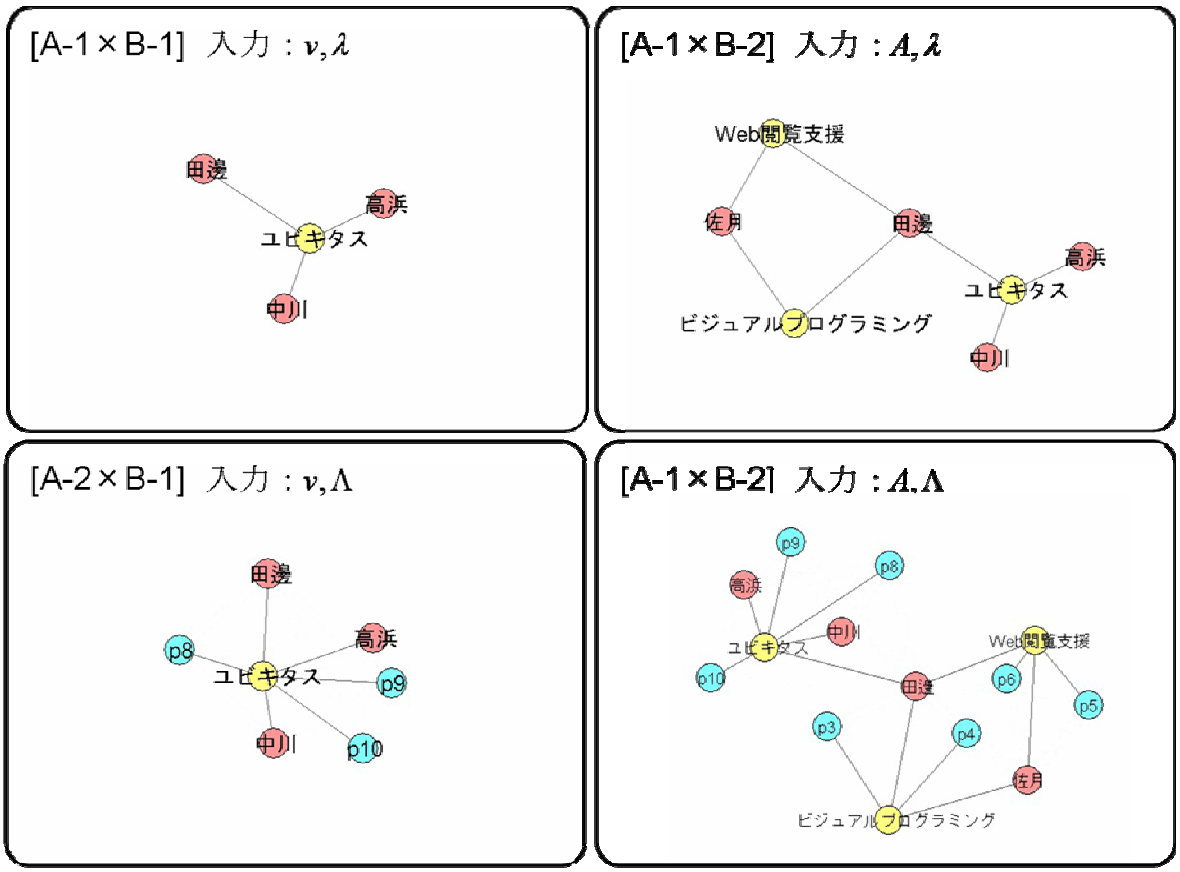


図 4.1　ある書誌情報ネットワークの全体構造



v＝ユビキタス, A＝{ユビキタス, Ｗｅｂ閲覧支援, ビジュアルプログラミング}
λ＝著者, Λ＝{著者, 論文}

図 4.2　分類したネットワークの可視化例

第5章　論文データ分析支援ツール「STANICOV」

本章では、第 4 章で述べたグラフ操作体系を搭載した論文データ分析支援ツール「STANICOV[1]」について述べる。まず、第 4 章で述べた操作体系への対応付けを含めたツール設計方針について説明する。次に、対話的にネットワークビューを構成するための機能について説明する。最後に、ツールのシステム構成やツールで使用したグラフレイアウトのアルゴリズム、実装言語、入力データ形式について述べる。

5.1　　ツールの設計方針

STANICOV では、ネットワークを構成するノード(書誌情報)をそれぞれの属性に従ってフィールドを分割して配置し、ノードをフィールド間で移動させることによってネットワークを構築する。その際、構築したネットワークを閲覧、保持するためのワークスペースのようなフィールドを用意しておく(図 5.1)。このとき、第 4 章で述べた選択項目は表 5.1 のように対応する。

表 5.1 操作体系とツールでの操作の対応表

操作体系における入力	ツールでの操作
ノード属性の組み合わせ(λ, Λ)	移動するフィールドの選択
ネットワークを構成するノード(v, A)	移動するノードの選択

ノード属性の組み合わせの選択を「ノードをどのフィールドに移動するか」という選択に対応させる(図 5.2)。このことにより、ノード属性の組み合わせをユーザの意思決定のみで入力できるようになる。また、ノードの選択をネットワークの直接操作によって実現し、直感的にネットワークを構築できるようになる。

ツールのユーザは、得たい情報に応じて各フィールドからノードを集めてワークスペースにネットワークを構築し、構築したネットワークを基にさらにノードを集めることによって情報探索を行う。

[1] **S**urvey **T**ool of **A**cademic-paper **N**etwork with **I**nteractive **C**omposing / **O**perating **V**iew

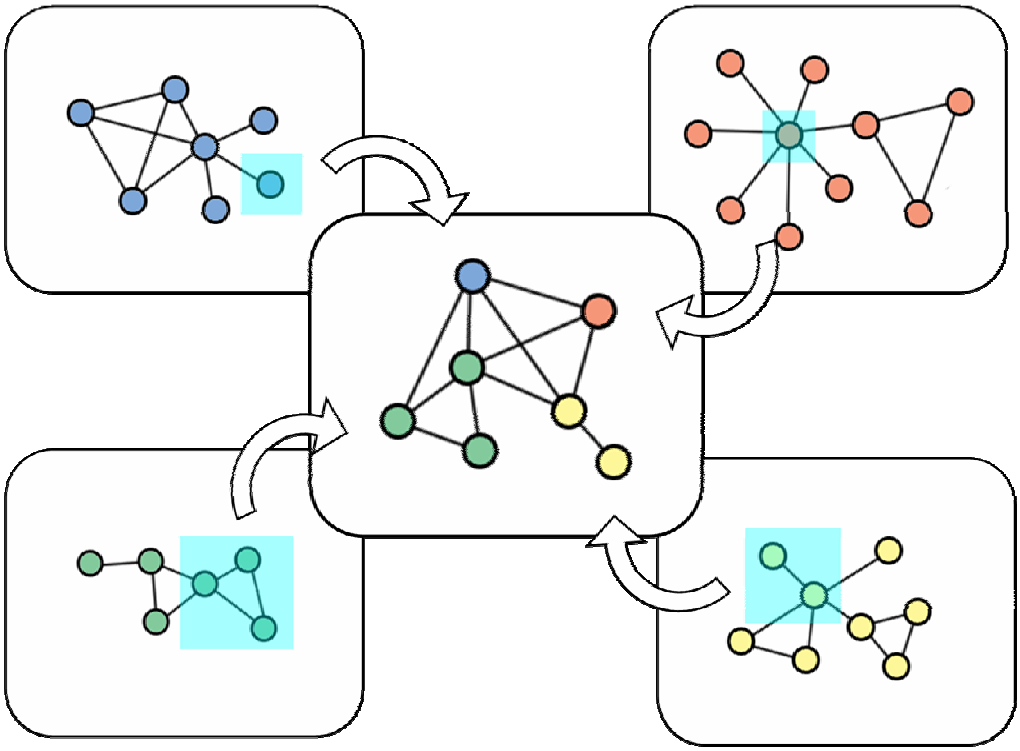


図 5.1　フィールドの分割によるネットワーク構築

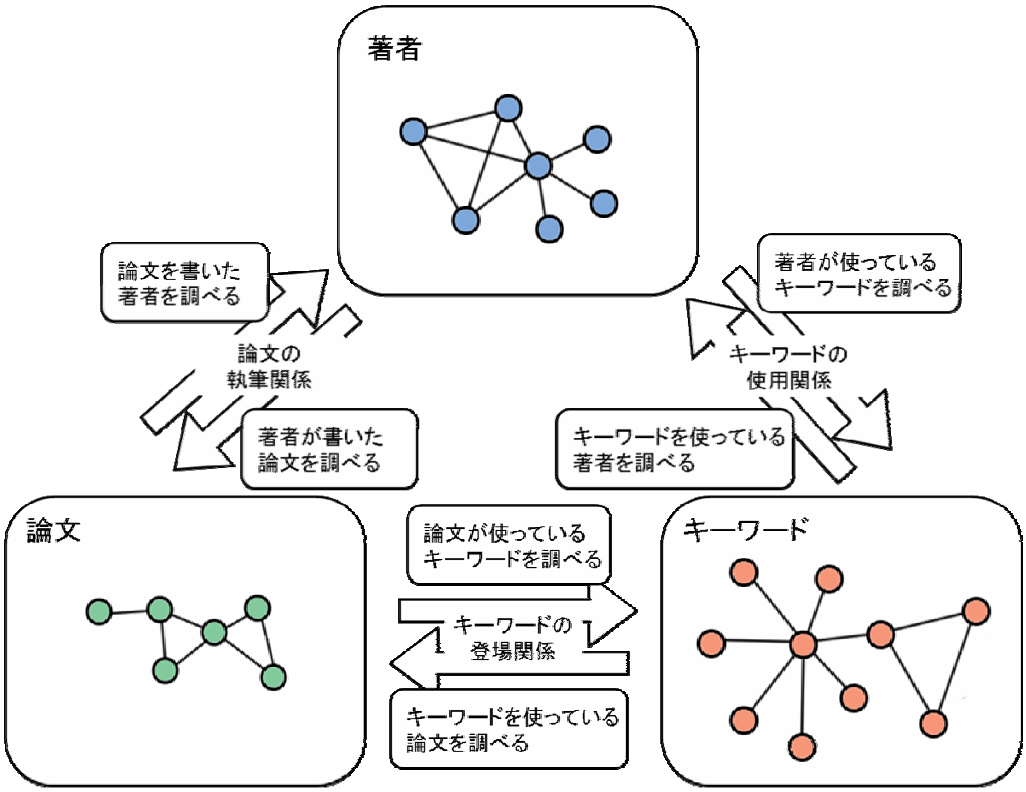


図 5.2　ノード属性の組み合わせとそれぞれの関係

5.2　　ツール概観

図 5.3 は STANICOV の初期画面である。ツールは画面左側の操作パネルと画面中央のネットワーク図表示部分からなる。ネットワーク図表示部分では、論文データのネットワークがネットワーク図で描画されている。ネットワーク図表示部分はノードの属性ごとにフィールドに分かれており、フィールドの中にあるノード間にのみエッジがつながるようになっている。フィールド及びその中のノードは属性ごとに色分けされている。これは、ネットワークを構築した際にそれぞれのノードがどの属性のノードかをわかりやすくするためである。ノードは関連の強いもの同士が近くに配置される力指向モデルによってその位置を変化させるが、自発的にフィールドの外に出ることはない。ネットワーク図表示部分中央にある「何も描かれてないフィールド」はワークスペースである。

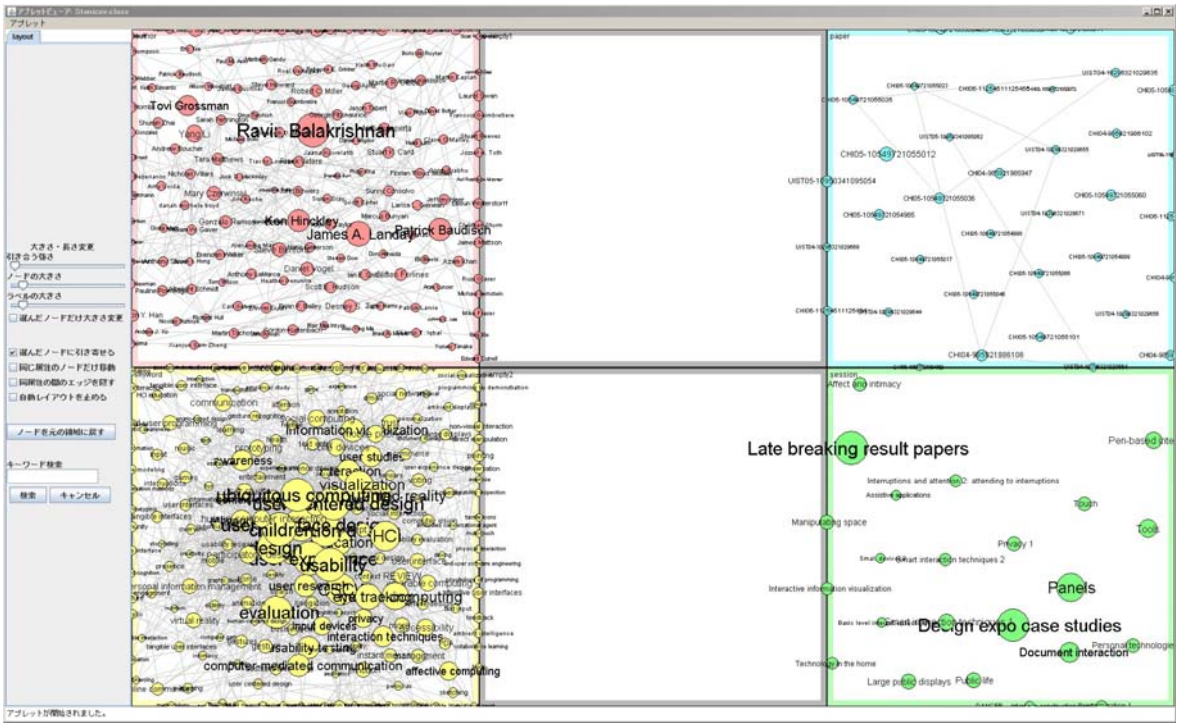


図 5.3　STANICOV の初期画面

5.3　　ツールの機能

5.3.1　　ノードの選択及び移動

本ツールにおける主な操作はノードの選択及び移動である。STANICOV では、移動するノードの選択に関するいくつかの機能を提供し、ノード選択をスムーズに行えるようにしている。

ノード移動の基本機能

ノードの移動はノードのドラッグ&ドロップによって行う。ノードを他のフィールドにドラッグ&ドロップすると、移動した先の領域の中にあるノードの中で移動したノードと関連があるものの間にエッジがつながる。ノードを移動する際、ドラッグしているノードは赤くハイライトされている。

図 5.4 は、著者のフィールドにあるノード「Ravin Balakrishnan」をワークスペースに移動する前後の図である。ノードをワークスペースに移動すると、ワークスペース内にあるノードとエッジがつながっていることがわかる。

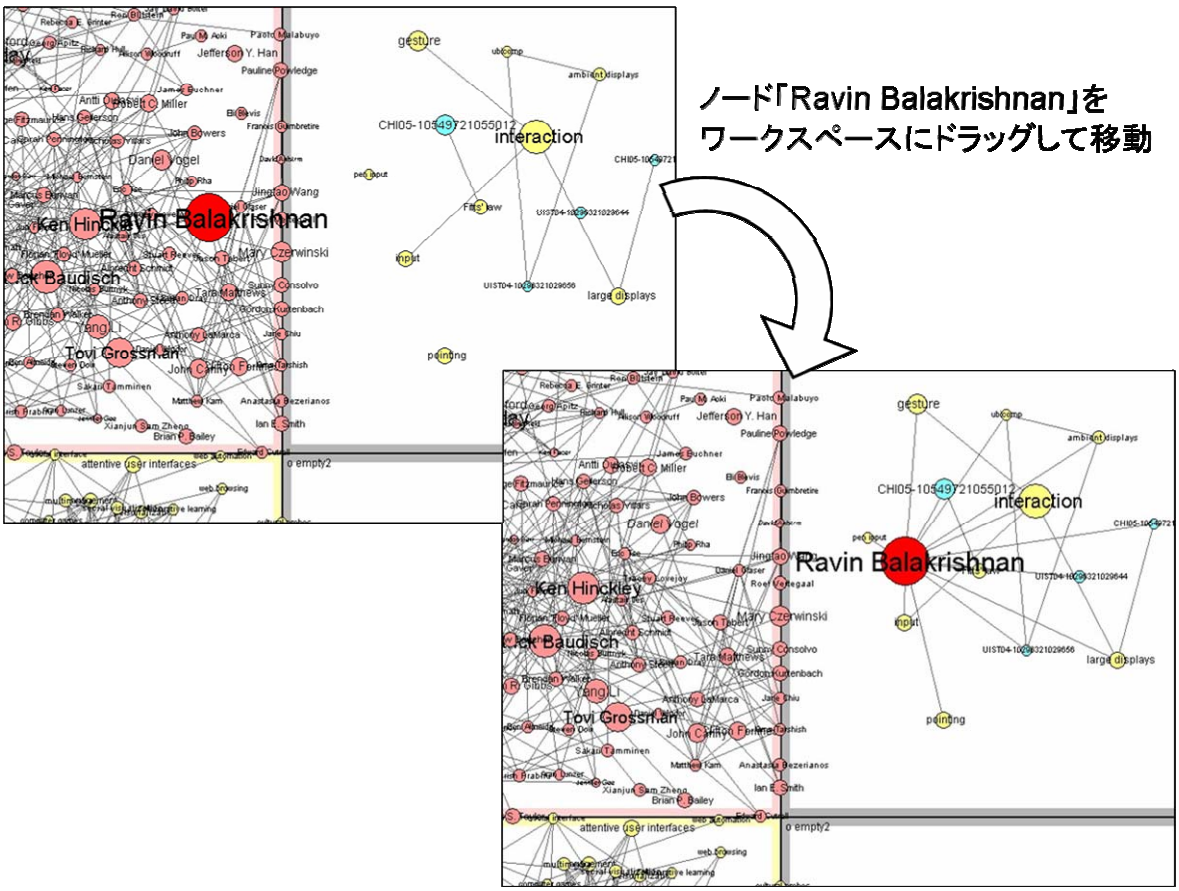


図 5.4　ノードの移動

隣接関係によるノードの選択

第 4 章で述べたように、論文データ分析における局所的な情報はノード間の隣接関係を把握することで得られる。本ツールでは、選択したノードに隣接するノードを選択したノードと共に選択、移動することができる。この機能によって、ネットワークから特定のノードとそれに直接関係するノードでネットワークを新たに構築することができる。

隣接するノードの選択及び移動は、ノードをクリックして選択し、そのノードをドラッグ&ドロップすることで行う。その際、選択されたノードに隣接するノードは紫色にハイライトされ、ドラッグ中は選択されたノードとの相対位置を保持したまま移動する。

図 5.5 は、ノード「Patrick Baudisch」に隣接するノードをまとめて選択している図である。ノード「Patrick Baudisch」に隣接するノードが紫色にハイライトされ、ノード「Patrick Baudisch」を移動させると共に移動する。

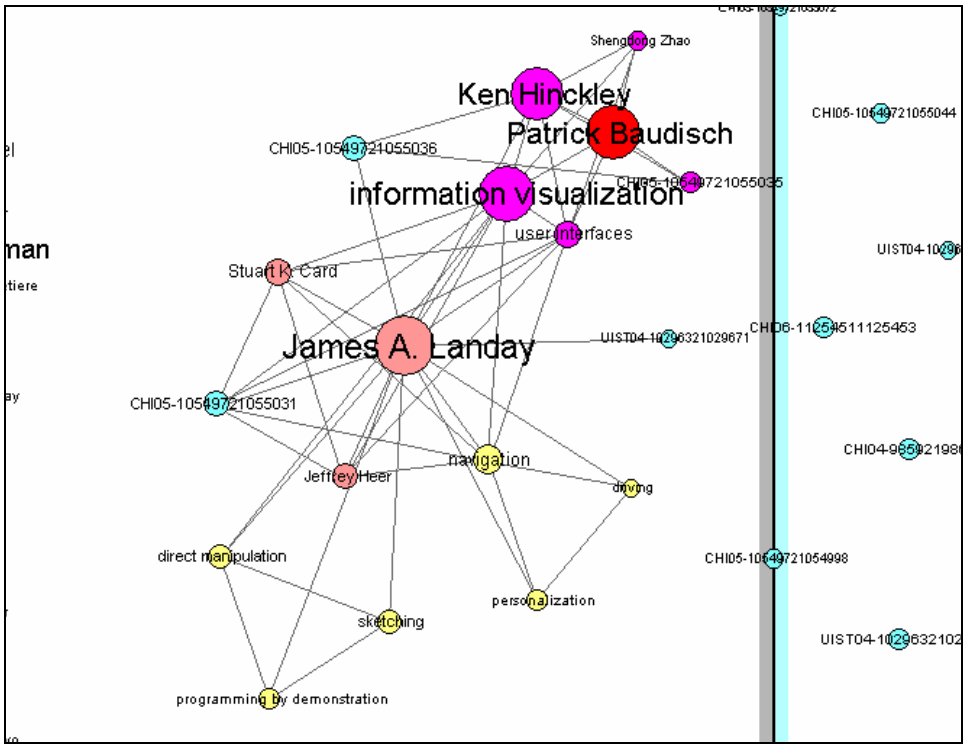


図 5.5　隣接関係によるノードの選択

領域によるノードの選択

フィールドの任意の場所に矩形領域を作り、矩形領域に含まれるノードをまとめて選択、移動することができる。この機能によって、ノードを移動することで構築したネットワークや密接にエッジがつながっているノード集合などをまとめて移動することができる。

領域によるノードの選択及び移動は、マウスドラッグによって矩形領域を作り、その矩形領域をドラッグ&ドロップすることで行う。その際、矩形領域は灰色に着色され、領域内にあるノードは紫色にハイライトされる。領域のドラッグ中、領域内にあるノードは矩形領域との相対位置を保持したまま移動する。また、フィールド全体に矩形領域を作ることもできる。フィールドによるノードの選択及び移動は、フィールドの淵にある色のついた枠をクリックしてフィールド全体に矩形領域を作り、選択領域をドラッグ&ドロップすることによって行う。移動中のノードの色及び位置は領域によるノード選択時と同様である。また、フィールド全体の移動を補助するために、選択領域が移動先のフィールドに重なる位置に近くなったとき、領域をフィールドに重なるようにスナップする機能を付加している。

図 5.6 は領域によるノード選択の例である。マウスドラッグによって作成した灰色の矩形に含まれるノードが紫色にハイライトされ、矩形を移動させると共に移動する。

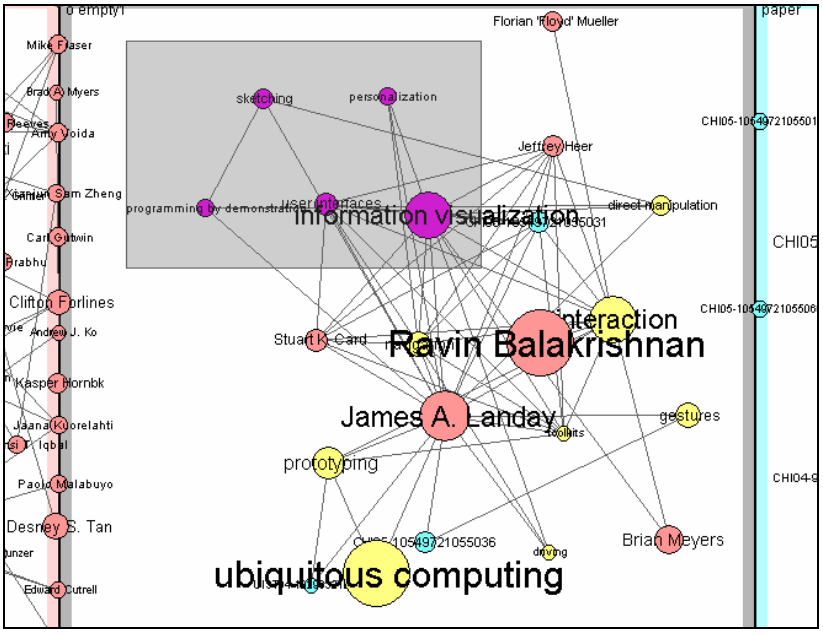


図 5.6　領域によるノードの選択

図 5.7 はフィールド全体に矩形領域を作った例である。フィールドの淵にある枠をクリックするとフィールド全体に矩形ができてフィールド内のノードが紫色にハイライトされ、矩形を移動すると共に移動する。

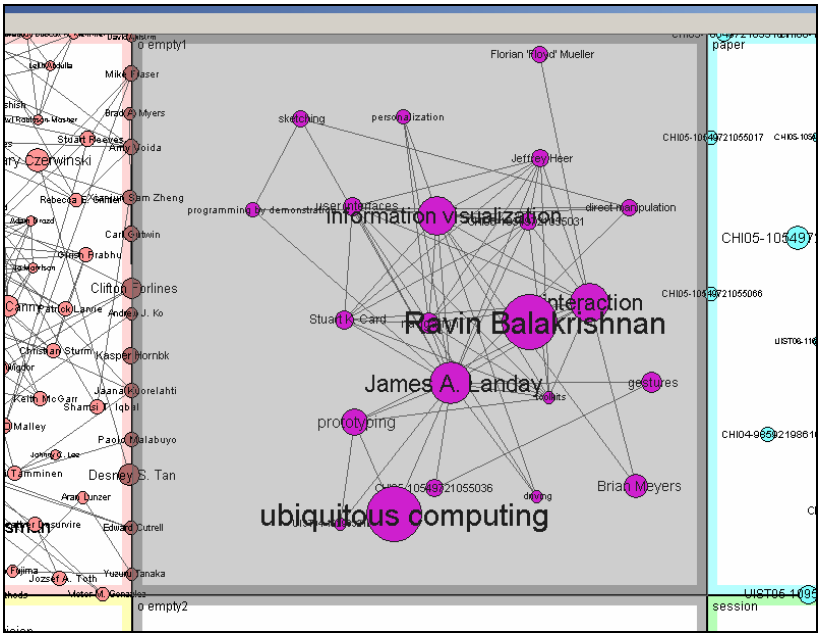


図 5.7　フィールド全体の選択

5.3.2　　ネットワークのレイアウト及び操作の補助機能

隣接ノードの引き寄せ

5.3.1 節で述べたように､選択ノードに隣接するノードを共に移動する際はノード同士の相対位置を保ったまま移動を行う。その際、ノード間の距離が大きく開いているとノード移動を行いにくい。この問題を解決するために、本ツールは選択ノードに隣接するノードを選択ノードに引き寄せる機能を搭載している。ノードが選択状態になるとそれに隣接するノードが近くに寄り、隣接ノードの移動を行いやすくなる。隣接ノードを引き寄せるか否かの切り替えは、画面左の操作パネルの中にあるチェックボックス「選んだノードに引き寄せる」で行う(図 5.10)。

属性による移動ノードのフィルタリング

構築したネットワークから共著関係だけを抽出して見たい場合など、同じ属性を持ったノードのみを移動したい場合がある。本ツールはフィールド内で同じ属性を持つノードのみを移動する機能を搭載している。画面左の操作パネルの中にあるチェックボックス「同じ属性を持つノードだけ移動」をオンにすることで、矩形選択で移動するノードを指定したノードと同じ属性を持つノードに限定することができる(図 5.10)。チェックボックスをオンにした状態で矩形を作り、矩形内のノードをドラッグすることで、そのノードと同じ属性を持つノードを共に移動することができる。

図 5.8 はフィールド内で「著者」の属性を持つノードのみを移動している図である。矩形内のノード「Ravin Balakrishnan」をドラッグすると、このノードと同じ「著者」の属性を持つノードのみがハイライトされ、共に移動する。

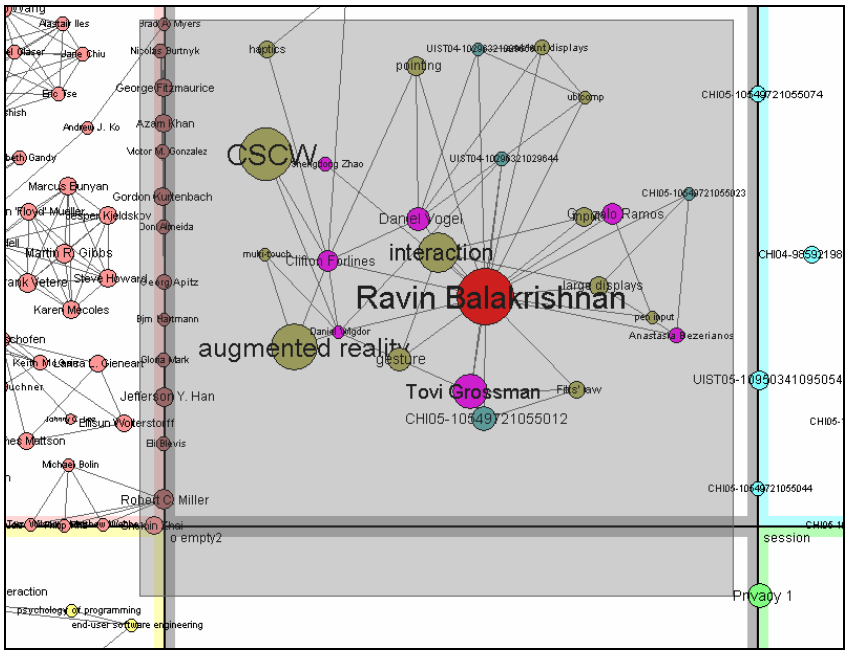


図 5.8　属性による移動ノードのフィルタリング

同属性ノード間のエッジの表示切り替え

著者とキーワードとの関係だけを見たいときなど、属性の違うノード同士のエッジだけに着目したい場合がある。本ツールは同じ属性を持つノード同士のエッジの表示切り替え機能を搭載している。画面左の操作パネルの中にあるチェックボックス「同属性の間のエッジを隠す」をオンにすることで、異なる属性のノード間のエッジのみを表示することができる(図 5.10)。

図 5.9 は著者とキーワードからなるネットワークから著者間、キーワード間のエッジを隠した例である。エッジを隠すことによって、著さとキーワードとの間の関係に着目してネットワークを閲覧することができる。

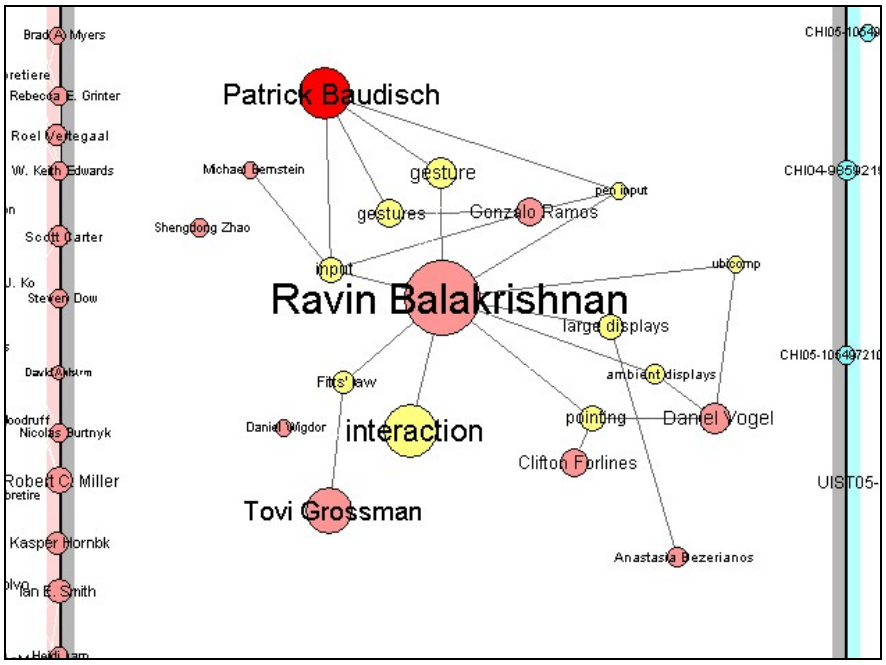


図 5.9　同属性ノード間のエッジを隠した例

ネットワークの自動レイアウトと手動レイアウトの切り替え

本ツールは力指向モデルによってネットワークの自動レイアウトを行うが、自動レイアウトを停止することもできる。自動レイアウトを停止することで、構築したネットワークのレイアウトをユーザ自身が決められるようになる。なお、自動レイアウト停止中もノード移動を行うことができる。自動レイアウトと手動レイアウトの切り替えは画面左の操作パネルの中にあるチェックボックス「自動レイアウトを止める」で行う(図 5.10)。

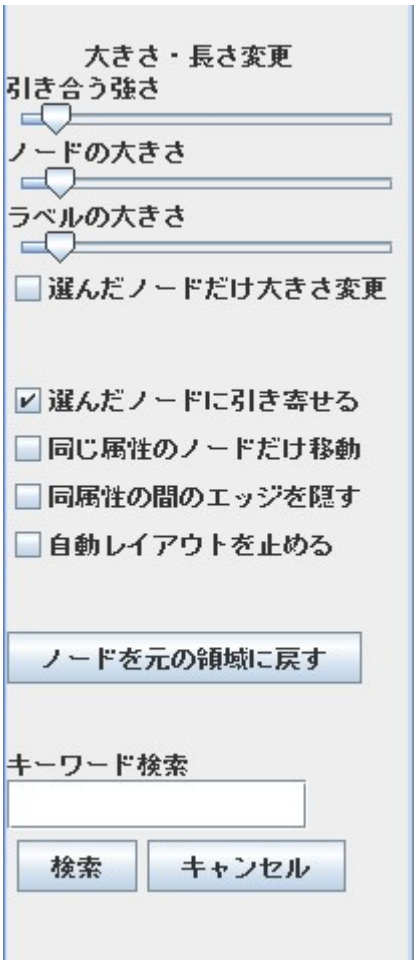


図 5.10　操作パネル

5.3.3　　情報探索の支援

論文データ分析における情報探索において、探索のきっかけを得ることは重要である。また、探索のきっかけがわかっている場合、それをネットワークの中から発見することも重要である。本ツールでは、探索のきっかけの獲得と発見の 2 つの観点から情報探索を支援する機能を搭載している。

きっかけの獲得：注目度の提示

「著名な研究者」や「注目されている研究トピック」などが探索のきっかけになることは多い。本ツールでは、それぞれの書誌情報の「著名である」「注目されている」といった情報を書誌情報の「注目度」とし、ノードの大きさに割り当てている。このことにより、多く引用された論文やその著者、多くの論文が使っているキーワードなどを発見しやすくなる。

論文や著者、キーワードなど書誌情報の注目度に割り当てる指標について、本研究では表 5.2 のように指標を割り当てた。

表 5.2 注目度への指標の割り当て

書誌情報の属性	注目度とする数値指標
論文	論文の被引用数
著者	書いた論文の被引用数の合計
その他(キーワードなど)	論文に登場した頻度

多くの論文に引用される論文及びその著者は著名な論文、著名な著者であると言える[21][22]。また、多くの論文に使われているキーワードはその研究分野において注目されているキーワードであると言える。

なお、ノードとラベルの大きさは画面左の操作パネルのスライダを移動することによって変更することができる(図 5.10)。

きっかけの発見：キーワード検索

自分が興味を持っている論文や研究トピックが探索のきっかけになることは多い。本ツールはユーザが探したいノードを検索するためのキーワード検索機能を搭載している。

操作パネルのテキストボックスにキーワードを入力して「検索」ボタンを押すと、ラベルにキーワードを含むノードが青色にハイライトされる(図 5.11)。

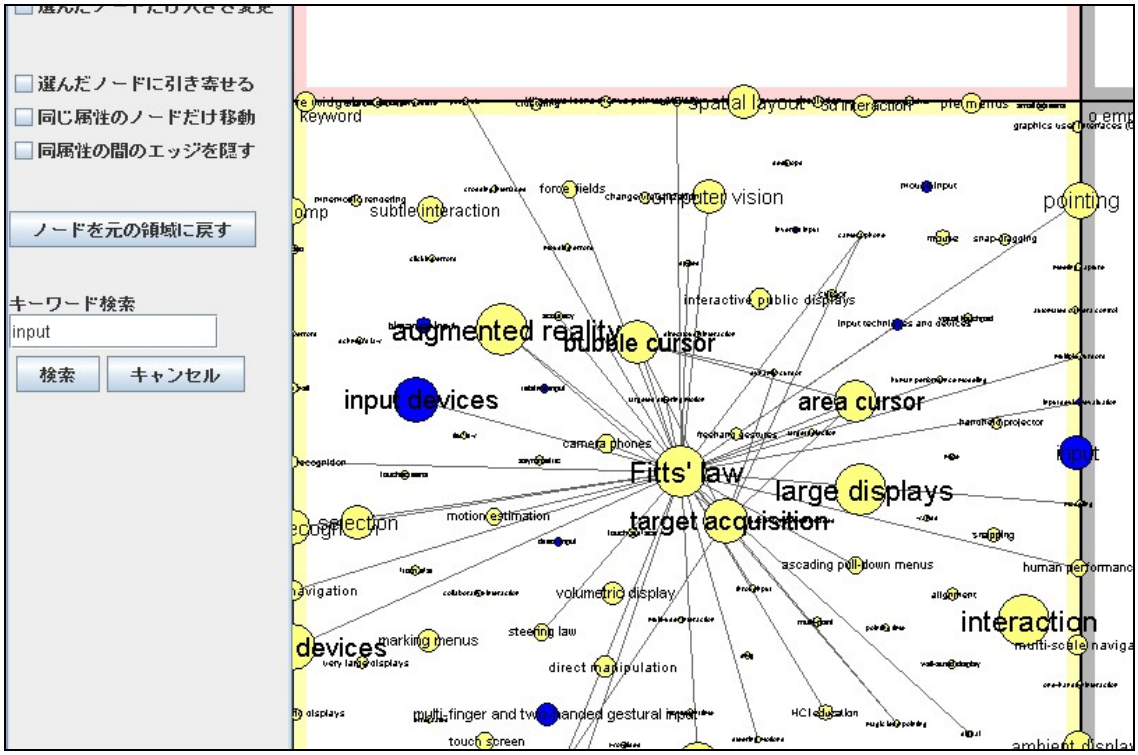


図 5.11　キーワード「input」で検索した結果

5.4　　ツール実装

5.4.1　　システム構成

図 5.12 は本ツールを利用するためのシステム構成である。システムは論文データを収集、解析するためのコンポーネントとインタフェースコンポーネントからなる。論文データ収集・解析コンポーネントは、論文情報 Web サイトから Web ページを取得し、論文情報データベースに格納する。次に、収集した Web ページの中から論文の書誌情報を抽出し、関係情報データベースに格納する。インタフェースコンポーネントは関係情報データベースからデータを読み込み、ユーザにネットワークビューを提供する。ユーザはネットワークビューを操作し、インタフェースからフィードバックを得る。

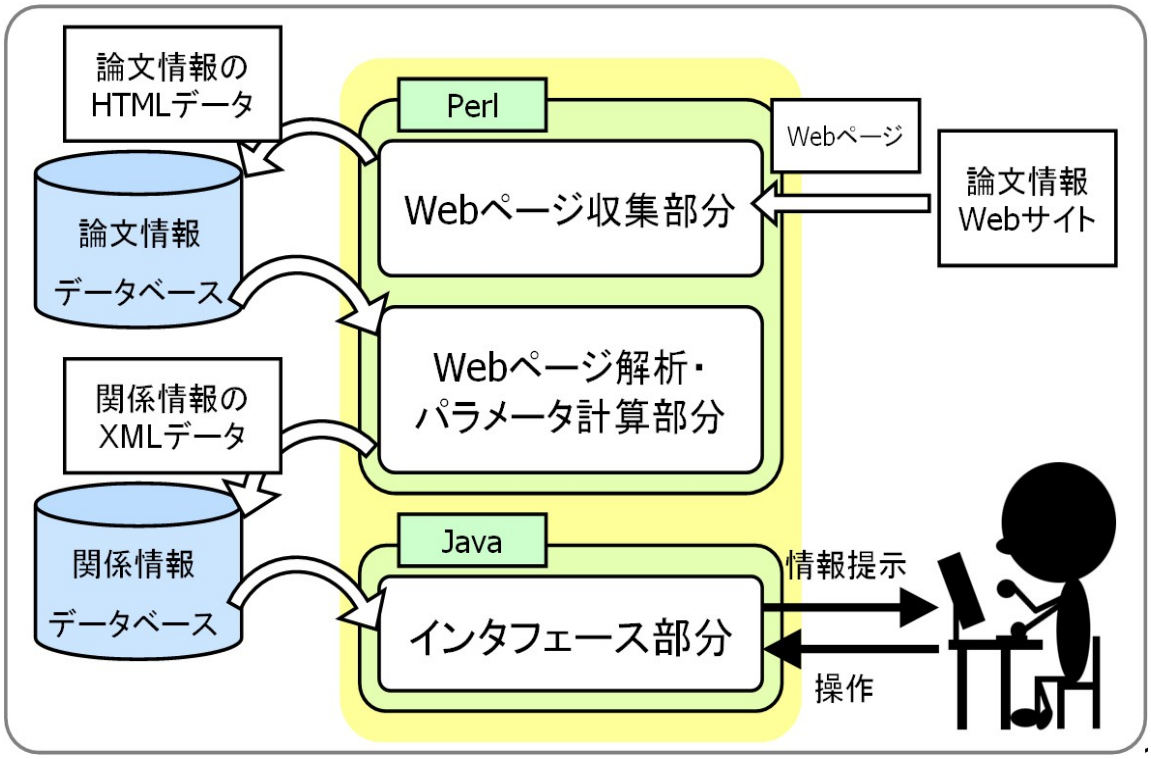


図 5.12　　STANICOV のシステム構成及びデータフロー

5.4.2　　実装言語と入力データ形式

論文データ収集・解析コンポーネントはPerl 5.10.0 を用いて実装した。インタフェースコンポーネントはJava™6.0(Java™ Platform, Standard Edition 6 Development Kit [2])を用いて実装した。インタフェースコンポーネントが読み込むデータの形式はGraphMLとした。GraphMLとはグラフ構造をXMLで記述する拡張形式の一つであり、多くのネットワーク可視化研究で用いられている[9]。

5.4.3 ネットワークのレイアウトアルゴリズム

本ツールは、ネットワークのレイアウトアルゴリズムとしてグラフの可視化に用いられる網図形の代表的な描画法の 1 つであるスプリングモデル[10]を用いた。スプリングモデルとは、エッジを自然長のあるばねと仮定し、エッジで接続されていないノード同士には斥力を発生させるもので、ばねの力と斥力によりノードを配置して安定状態を計算することでノードのレイアウトを求める手法である。

スプリングモデルにおいてノードのレイアウトは以下のように行う。

1. ランダム配置などにより初期位置を決定する
2. それぞれのノードにはたらくばねの力と斥力から合力を求める
3. 求めた合力に従ってノードのレイアウトを変更する
4. 2 に戻る

ばねモデルは、エッジに重みを負荷することで関係の強さを視覚的に表現することが可能である。関係の強いノード同士のばねの力を強くすることで、関係の強いノード同士が近くに配置されるようにレイアウトすることができる。本研究では、書誌情報間の関係の強さを「同じ論文の中で登場した頻度」に対応付けてネットワークを描画している。

5.5　　ツール利用例

ここでは、ヒューマンインタフェースの論文データから「情報可視化」という研究トピックに興味を持ってサーベイを行う場面を想定してツールの利用例を挙げる。ここでは、論文データから「著者」「論文」「キーワード」「論文が発表されたセッションの名前」の 4 種類の書誌情報を抽出したデータを用いる。

まず分析者は「information visualization」をクエリとしてキーワード検索を行う。入力の結果、キーワードのフィールドにあるノード「information visualization」と、論文が発表されたセッションのフィールドにあるノード「Interactive Information Visualization」が青くハイライトされる(図 5.13)。

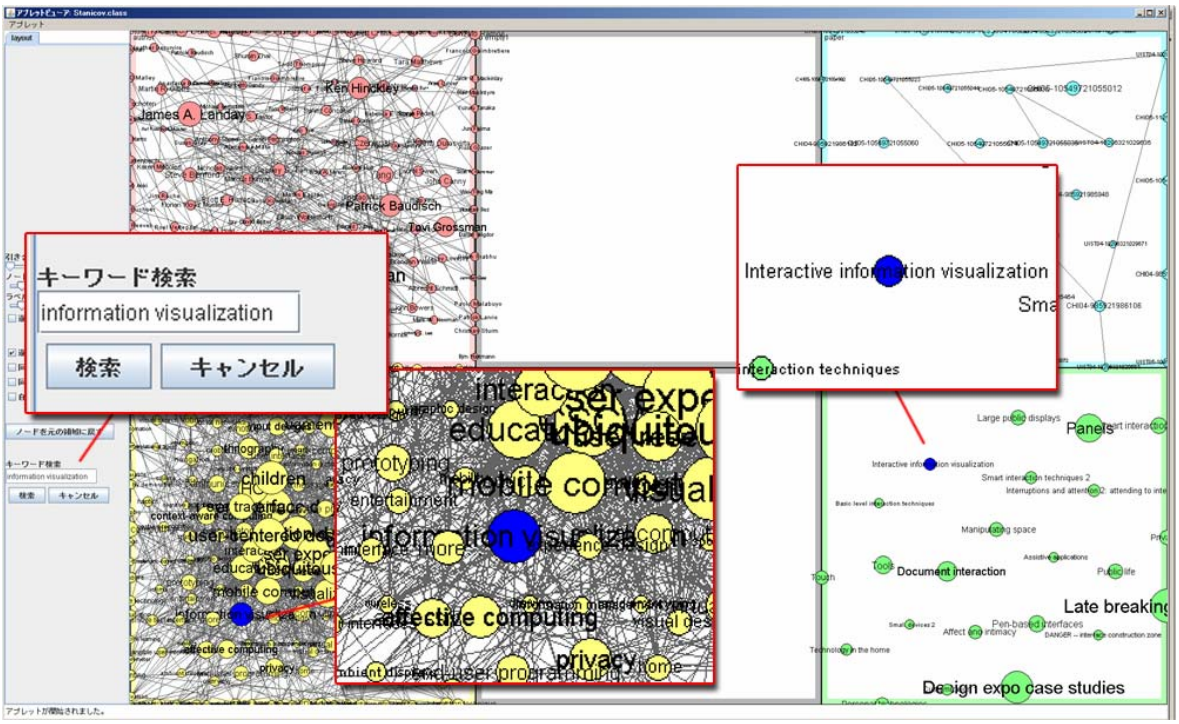


図 5.13　　キーワード検索の結果

検索にヒットしたノードをワークスペースに移動し、次に「information visualization」というキーワードを使っている研究者を調べるために、ノード「information visualization」を著者のフィールドに移動し、「information visualization」に関連のある研究者を集める(図 5.14)。

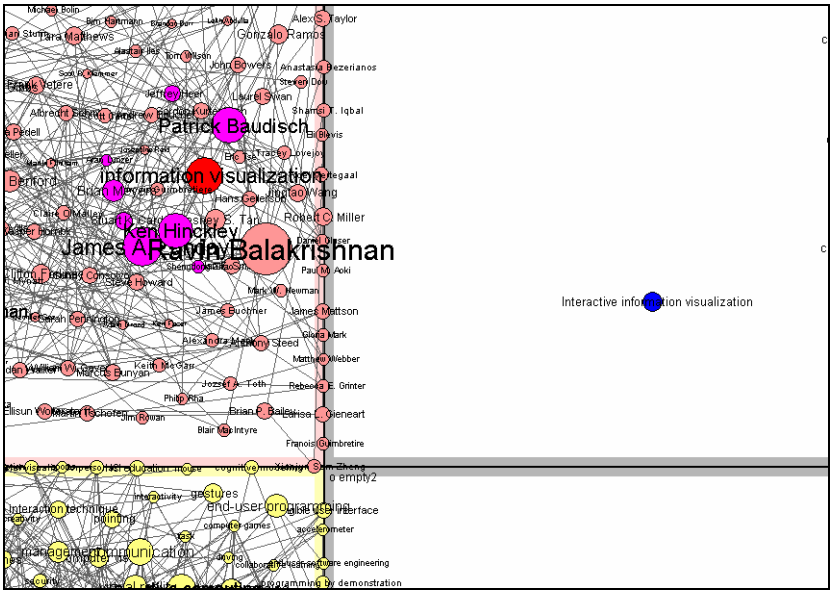


図 5.14　ノードを著者のフィールドに移動した結果

その結果、キーワード「information visualization」に関連する著者で構成されるネットワークが構築できる(図 5.15)。

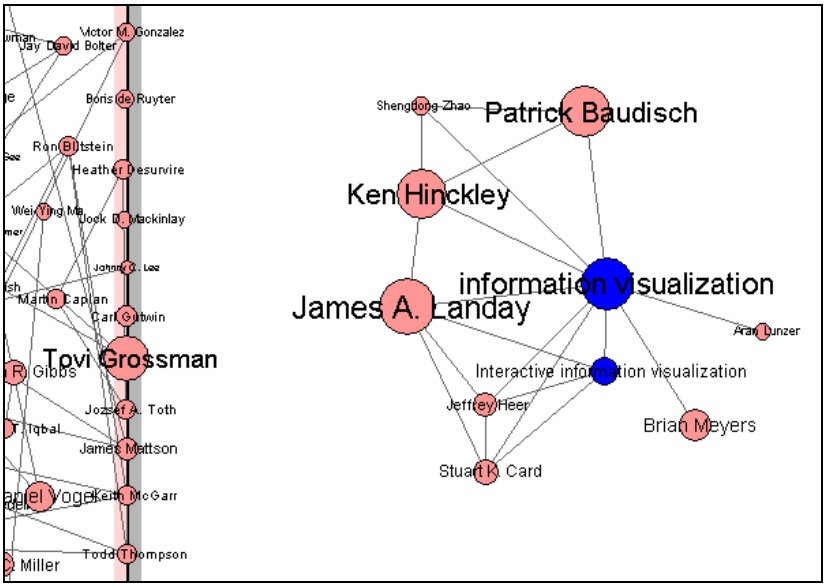


図 5.15　「information visualization」に関する著者のネットワーク

次に、information visualization に関連する著者がどのようなトピックについて研究を行っているかを調べる。ノード「Jeffrey Heer」をキーワードのフィールドに移動し、「Jeffrey Heer」と関連のあるキーワードを調べる(図 5.16)。

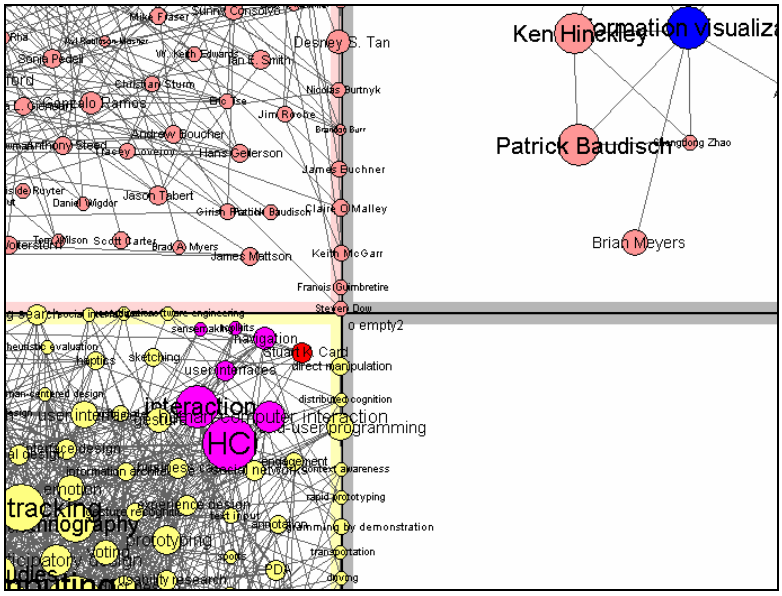


図 5.16　ノードをキーワードのフィールドに移動した結果

何人かの著者について同様の操作を行った結果、著者とそのそれぞれが使っているキーワードから構成されるネットワークを構築することができる。このネットワークから、「information visualization」に関連する著者がどのようなキーワードで研究を行っているのかがわかる(図 5.17)。

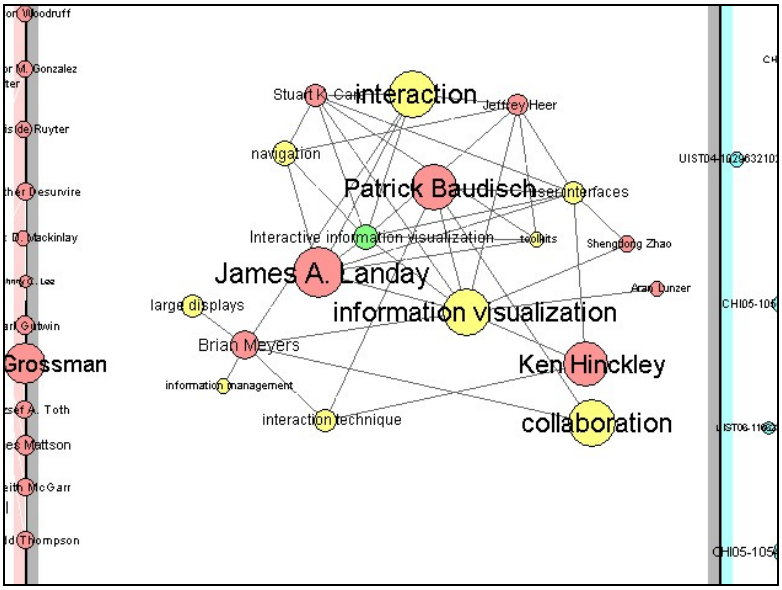


図 5.17　著者とキーワードからなるネットワーク

図 5.17 のネットワークを見ると、何人かの著者がキーワード「interaction」と関連があることがわかることから、彼らは情報可視化にインタラクションを組み合わせる研究を行っていると推測できる。そこで、このネットワークの中で「information visualization」と「interaction」に関連する情報を集めることにする。それぞれのノードをキーワードや論文のフィールドに移動してノードを集め、新たにネットワークを構築する(図 5.18)。

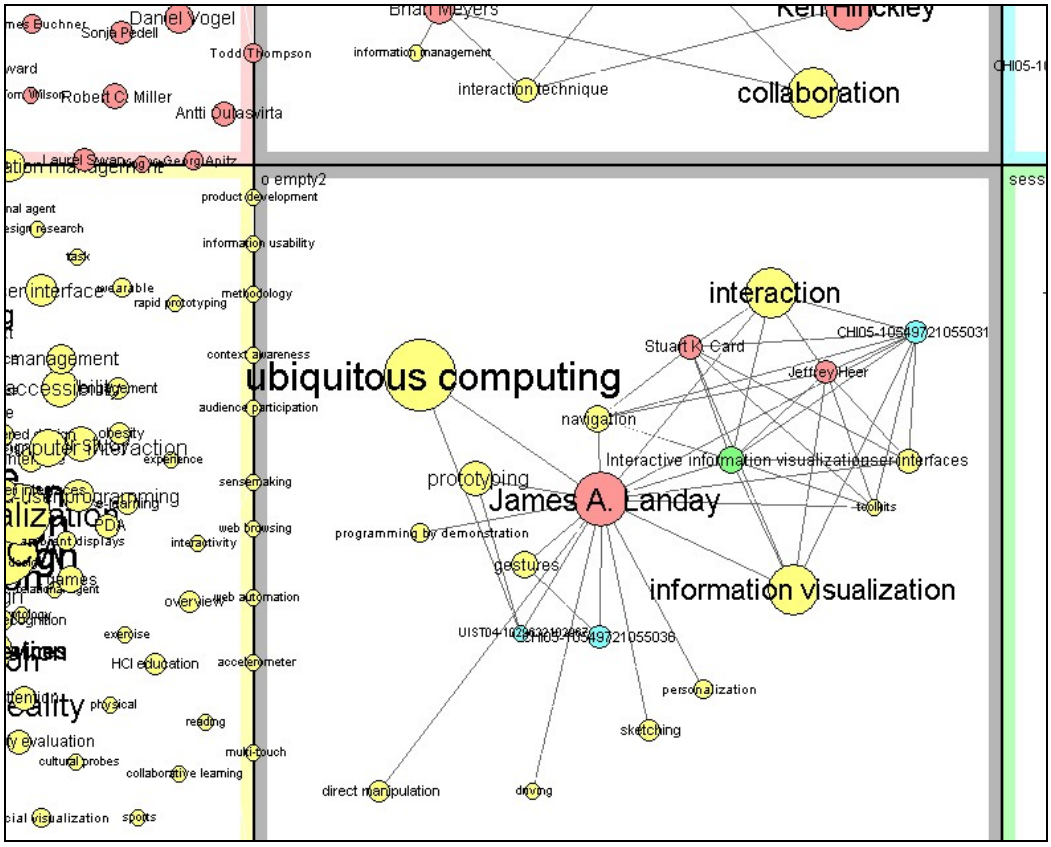


図 5.18 「information visualization」と「interaction」に関するネットワーク

その結果、情報可視化とインタラクションに関する研究者やそれぞれが行っている研究がおおよそどのようなものかが推測できる。例えば「Jeffrey Heer[1]」という研究者は「interaction」の他に「user interface」や「navigation」などのキーワードとつながっていることから、情報可視化のためのナビゲーション手法やそれを搭載したインタフェースの開発を行っているのではないかと推測できる。

[1] ネットワーク可視化のためのツールキットprefuse[31]の開発者である。

第6章　評価実験と考察

本章では、本研究で行った評価実験とその結果に基づく考察を述べる。また、考察を受けての本研究の今後の展望について述べる。

6.1　評価実験

6.1.1　実験の目的

実験の目的は以下の通りである。

- 「本研究で構築した操作体系及びツールが論文データの分析を行う上で有用である」という事柄の検証
- 「ツールに搭載されたそれぞれの機能の有用性と改善点」についての調査

6.1.2　実験概要

被験者には、研究分野のサーベイをする場面を想定し、STANICOV を利用して与えられたタスクをこなしてもらう。その後、各機能についてタスクを行う上での有用性や使用感についてアンケート調査を行う。また、各被験者についてツールの操作ログをとり、各被験者が行った操作を記録しておく。

6.1.3　被験者

コンピュータサイエンスを専攻する大学生及び大学院生 6 名を対象に実験を行った。被験者はいずれもヒューマンインタフェースに関する研究室に所属する学生であり、本研究で扱うネットワークの概念と表現方法についてある程度の知識を有する者である。

6.1.4　手順

まず、STANICOVの使用目的と使い方について 10 分程度で説明する。次に、被験者にSTANICOVを実際に使ってもらい、STANICOVの操作に慣れてもらう。試用の時間は各被験者に任せた。その後、ヒューマンインタフェースの研究分野のサーベイを模した問題に対し、STANICOVを使って調べ、答えてもらう。なお、データとしてヒューマンインタフェースに関する 2 つの国際会議CHI[1]とUIST[2]の 5 年分のデータを抜粋して用いた。

タスクは次項始めに示す 3 つについて行った。各タスクは、実際の論文データ分析でもしばしば見られるタスクであり、かつ連鎖的な情報の探索が必要なタスクである。

[1] Conference on Human Factors in Computing Systems, http://www.chi2009.org/
[2] Symposium on User Interface Software and Technology, http://www.acm.org/uist/uist2009/

1. あるキーワードに関連する論文とそれぞれの著者を調べる
2. ある研究者が行っている研究をその研究者に関連するキーワードから調べ、同じ研究トピックについて調べている他の研究者を調べる
3. タスク作成者が挙げたいくつかのキーワードから興味を持ったものを 1 つ選び、それに関連する書誌情報を自由に調べる

全タスク終了後、以下の項目についてアンケート調査を行う。回答はそれぞれ 5 段階評価と自由記述によって行ってもらう。

- タスク 1 とタスク 2 について、十分に調査できたと思うか？
- タスク 3 について、有益な情報を得ることはできたか？
- ノードを動かしてネットワークを作るという操作はタスクをこなす上で有用であったか？
- 領域をノード属性ごとに分割して配置するというレイアウトはタスクをこなす上で有用であったか？
- スプリングモデルによるネットワークの自動描画はタスクをこなす上で有用であったか？
- ツールに搭載された各機能はタスクをこなす上で有用であったか？

タスクの達成度に関する質問はタスク 1, 2 とタスク 3 で質問を変えた。これは、タスク 1, 2 がタスク作成者側で指定した書誌情報を調べるタスクなのに対し、タスク 3 は自由に書誌情報を調べるタスクであることから、「十分に調査ができたと感じるか」という点ではなく「調べた情報は有益であったか」という点で達成度を評価できると考えたためである。

6.1.5 結果

表 6.1、表 6.2 は各項目に対するそれぞれの被験者(A～F)の回答をまとめたものである。表 6.2 について、アンケート結果の平均値はその機能を使用した被験者のみの平均である。

表 6.1 各被験者のアンケートへの回答

評価項目	A	B	C	D	E	F	平均
タスク 1 とタスク 2 の達成度	5	5	4	3	5	4	**4.33**
タスク 3 の達成度	5	4	5	4	4	5	**4.50**
「ノード属性ごとに領域を分割して配置する」というレイアウトの有用性	5	4	4	5	5	5	**4.66**
「ノードを動かしてネットワークを作る」という操作の有用性	5	4	4	2	4	5	**4.00**
スプリングモデルによるネットワークの自動描画の有用性	4	2	4	4	3	4	**3.50**

1: とても低い　2: やや低い　3:どちらでもない　4: まあまあ高い　5: とても高い

表 6.2 各機能に対する評価

評価項目	A	B	C	D	E	F	使用者平均
隣接ノードの移動	5	5	5	3	5	5	**4.66**
矩形選択による移動	2	4	3	2	5	5	**3.50**
属性による移動ノードのフィルタリング	0	4	4	3	0	5	**4.00**
自動レイアウトの停止	0	5	5	5	4	5	**4.80**
同属性エッジの可視・非可視の切り替え	0	4	4	5	0	5	**4.50**
ノード・ラベルの大きさ変更	5	4	4	5	5	5	**4.66**
キーワード検索	5	3	5	5	5	5	**4.66**

0: 使わなかった　1: 全く有用でない　2: それほど有用でない
3:どちらでもない　4: まあまあ有用　5: とても有用

図 6.1 は、各被験者が用いたノード移動機能のノード操作の回数に占める割合である。図中、「single」は単一ノードの移動、「neighbor」は隣接ノードの移動、「area」は矩形選択による移動である。

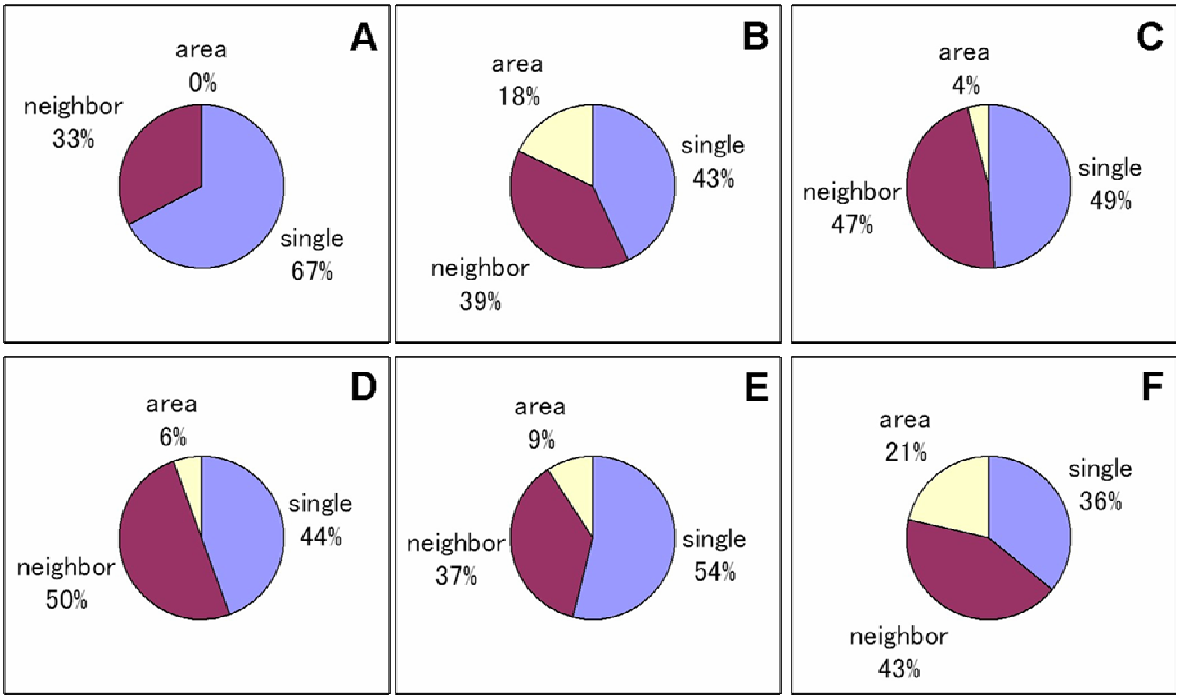


図 6.1　各機能のノード操作に占める割合

6.2 考察

6.2.1 ツールの有用性

「ノード属性ごとに領域を分割して配置する」ことや「ノードを移動することによってネットワークを構築する」ことの有用性についてはそれぞれ平均 4.0、平均 4.66 と高い評価を得た。また、各タスクの達成度はいずれも 4 以上と高い評価を得た。このことから、本ツールの特徴である「ノード属性によるノード配置領域の分割」と「ノード移動によるネットワーク構築」は論文データ分析に有用であり、論文データのタスクを十分に達成できるものであると言える。

6.2.2 各機能の使用状況及び有用性

ツールに搭載された各機能に対するアンケート結果から、ツールの主機能であるノードの移動に関する機能について、また補助機能に対する結果の中で注目すべきものについて述べる。

ノードの移動に関する機能については、「隣接ノードの移動」が平均 4.66 と高い評価を得ている。また、図 6.2 よりノード操作全体のうちおよそ 80%以上を単一ノードと隣接ノードの移動が占めている。このことから、論文データ分析における情報探索には書誌情報ネットワークの隣接関係の連鎖的な把握が重要であり、そのための隣接ノードの移動機能は有用であると言える。

ネットワークの自動レイアウトを停止する機能が平均 4.80 と高い評価を得た。また、タスク実行中に自動レイアウトを止め、ノードの属性や大きさなどによってその配置を分け、手動でレイアウトを行っている場面が見られた。このことから、分析者は構築したネットワークを自分なりに意味を持たせてレイアウトしたいという欲求があるのではないかと考えられる。ネットワークを分析者自らがレイアウトすることで、ネットワークが表す情報が整理され、情報を得やすくなる可能性がある。

6.2.3 改善すべき点

被験者 D が「ノード移動によるネットワーク構築の有用性」に対して「やや低い」と回答している。また、被験者 D は「単一ノードの移動と隣接ノードの移動の切り替えが煩わしく感じた」とコメントしている。本ツールでは単一ノードの移動をノードのドラッグ&ドロップによって行い、隣接ノードの移動をノードのクリックによる選択の後、ノードをドラッグ&ドロップすることによって行った。被験者 D が挙げた問題はノードの移動モードの切り替えにノードのクリックが必要であったことが原因であると考えられる。ノード移動機能のマウス操作への割り当てについては改善の余地がある。

また、注目すべきコメントとして、「単一のノードだけではなく、複数のノードに全て関連するノードを集められると良いと思った」というものがあった。本ツールで論文や研究者がどんなトピックに関連があるのかを調べる場合は、その論文や著者と関係のあるキーワードをいくつか集めることでトピックを推測することができる。そのトピックに関連する他の情報を集める際、単一のキーワードだけではなく、トピックを表すキーワード群に関連する情報を集めたいという欲求があると考えられる。すなわち、本研究では複数のノードの隣接関係からなるネットワークの構築を単一のノードの隣接関係を繰り返し把握することで行ったが、いくつかのノードの隣接関係を一度に把握することでネットワークを

構築したいという欲求があると考えられる。これを実現する案としては、いくつかのノードをノード群としてまとめる機能などを付加することで、この欲求を満足できると考えられる。

6.3 今後の展望

6.3.1 他のデータへの適用可能性

複数の関係情報を含むデータは実世界に多く見られ、その分析を行う必要性も高い。例えば商品の購買データでは、商品やそれが売れた時間、商品を購入した人物の年代や性別などがそれぞれ関係を持っており、それら関係情報を把握することで商品の購買傾向などを知ることができる。本研究で開発したツールはこのようなデータにも適用が可能であり、有益な情報の獲得を支援できると考えられる。

6.3.2 ツールとしての発展案

本研究では論文データが為すネットワークを構築するためのビュー操作の体系を作ることを主眼に置いて研究を行った。本研究で提案された操作体系は複数の関係からなるネットワークの把握支援を行う上での基盤技術となる可能性がある。今後の研究の方向性として、構築したネットワークに基づいて情報を整理する手法の開発や構築した操作体系を基盤にした可視化手法の開発が挙げられる。

6.2.2 節でも述べたように、構築したネットワークビューを分析者自身がレイアウトすることでネットワークが表す情報を整理できる可能性がある。情報整理を支援する手法として、Buzanらによる図解表現手法「Mindmap[1]」が参考になるのではないかと考えている。Mindmapは、言葉やイメージで表現したテーマから思いつくことをその意味や種類によって色分けした枝の先に書き、放射状にキーワードなどをつなげていくことで複雑な概念をコンパクトにまとめることができる手法である(図 6.2)[14]。論文データ分析における連鎖的な情報探索の手順はMindmapの作成手順に良く似ている。このことから、Mindmapにおける枝のように、意味や種類によってノードをまとめるような機能を搭載することで、ネットワークが表す情報を整理できると考えられる。例えば、構築したネットワークの中で興味のあるノードを中心に据え、それぞれのノードをそれぞれの関係の意味に従って放射状に配置するといったレイアウトを簡単に行えるようにする機能などが挙げられる。

[1] 「マインドマップ」および「Mindmap」は英国Buzan Organisation Ltd. の登録商標である。

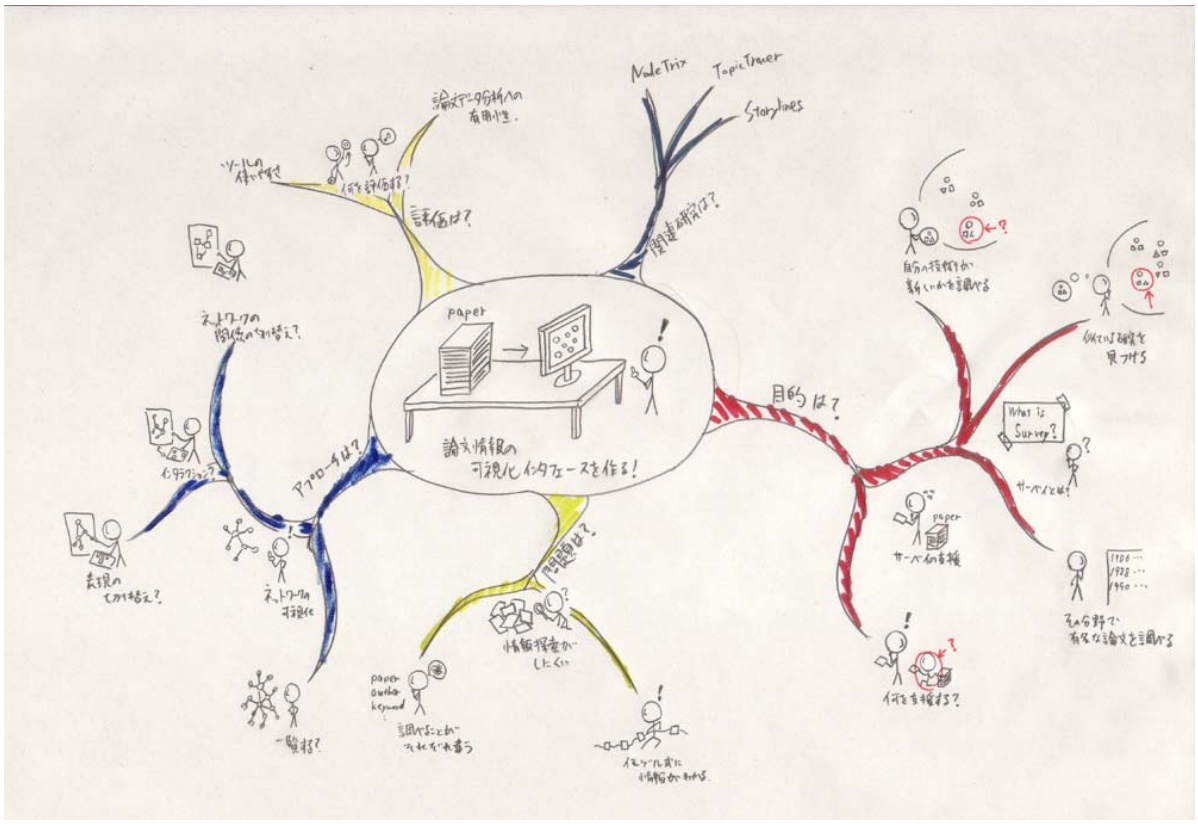


図 6.2　Mindmap の例

第7章　関連研究と本研究の位置付け

本章では、本研究に関連する研究について紹介し、本研究の位置付けを明らかにする。論文データ分析のためにネットワーク可視化を用いている研究については第 1 章を参照されたい。7.1 節ではネットワークの複数の関係に着目して把握支援を行っている研究について、7.2 節ではネットワーク可視化にインタラクションを導入している研究について、それぞれ関連する研究を挙げ、本研究との差分を述べる。最後に 7.3 節で本研究の位置付けについて述べる。

7.1　複数の関係に着目したネットワーク把握支援

実世界のネットワークの多くは複数の関係からなるネットワークであるが、ネットワーク把握支援に関する研究の中でネットワークが複数の関係からなることを考慮してその把握支援を行っている研究は多くない。複数の関係を持つネットワークはそのノードが属性を持つことから、ここではネットワークのノードが属性を持つことを考慮してその把握支援を行っている研究をいくつか挙げる。

Batagelj は大規模グラフを可視化するためのツールキット「Pajek」を開発し、その中でノード属性を色と階層配置によってわかりやすくしている[20][1]。また、Auberらは、大規模グラフ可視化の枠組み「Tulip」の中で、色によってノード属性を区別している[24][2]。

Wattenberg は、ノードの格子配置によって複数のノード属性を持つネットワークの把握を支援するツール「PivotGraph」を開発した[25]。2 次元空間上に平行軸をとり、格子の縦軸にノード属性を対応させ、横軸はノードを均等に配置するために用いることで、ノードの属性を明確に表し、同属性、他属性のノード間のエッジを見やすくしている。

Shneiderman らは､領域分割とノードの並列配置によって複数の関係からなるネットワークの把握を支援するツール NVSS(Network Visualization by Semantic Substrates)を開発した[27]。ネットワークのノードをその属性毎に領域に分けて領域内でノードを並列配置し、異なる属性のノード間の関係をエッジでつなぐことによって、ノードやエッジの重なりをなくしている。

Faisal らは、複数の関係がそれぞれに為すネットワークに対して違うビューを提供することで複数の関係からなるネットワークの把握を支援するツール「LKDVis」を開発した[33]。それぞれのネットワークをネットワーク図、領域図、座標軸などのビューによって可視化することで、それぞれの関係に適したネットワーク表現が可能となる。

これら研究はいずれもネットワークの概観の把握を支援するためにネットワークの要素を一覧できるように可視化するものであり、ネットワークの探索を支援するという点に着目しているものではない。これに対して本研究では、データ分析における情報獲得を支援するためにインタラクションを導入し、データから情報を獲得するためのネットワーク探索を支援している。このことにより､「意図を持った操作によってほしい情報を探す」とい

[1] http://www.labri.fr/perso/auber/projects/tulip/biblioTulip-abstracts.php
[2] http://vlado.fmf.uni-lj.si/pub/networks/pajek/

う能動的な探索を行えるようになると考えられる。また、これら研究はいずれもノードのレイアウトが決められており、観察者(ユーザ) が自由に配置を決めることができない。これに対して本研究では、ユーザ自身がネットワークビューを構築することでネットワークの探索とそれによる情報獲得を支援している。このことにより、情報獲得に過不足のないビューを得ることが可能になる。

7.2 ネットワーク可視化におけるインタラクション

情報可視化におけるインタラクションの有効性は Dix ら[30]をはじめとして情報可視化にインタラクションを導入した多くの研究によって示されている。それらインタラクション手法は Yi ら、Bederson らによって分類、体系化され、情報可視化におけるインタラクションのデザイン指針なども示されている[11][19]。

ネットワーク可視化においてもインタラクションの重要性が注目されるようになっている。Wills、Pretorius らによってネットワーク可視化におけるインタラクションの有効性が示された[23][26]。

ネットワークの可視化ツールにもインタラクション機能が導入されるようになってきている。また、Heerらは、ズーミングやフィッシュアイビュー、ノード配置のアニメーション表現などネットワーク表現のためのインタラクション機能を集めたツールキット「Prefuse」を開発した[31][1]。

ネットワーク可視化にインタラクションを導入している研究について、インタラクションを導入している目的別にいくつかの例を挙げる。

複数のネットワーク表現の間の対応を見るためにインタラクションを用いている研究もある。Henry らが開発したネットワーク図と行列表現の並列によるソーシャルネットワーク把握支援ツール「MatrixExplorer」[16]では、一方の表現に対するユーザの操作を他方の表現にも同期させることで 2 つの表現間の関連や対応を見やすくしている。また、Peter らはソーシャルネットワーク分析のための数値指標をランキングで表現し、ネットワーク図と同期させて分析を支援するツールを開発した[28]。ネットワーク図と表でハイライトなどの操作を同期させることでネットワークの各ノードが持つ数値をわかりやすくしている。また、Collins らはネットワーク表現する複数の手法を 2.5 次元空間に並列配置するツール「VisLink」を開発した[13]。2.5 次元空間に配置されたそれぞれの手法の間を関係線で結び、視点を切り替えることで手法間の対応を見やすくしている。また、Zhu らは文書データの単語が為す意味空間の特徴量を折れ線グラフとネットワーク図で表現することで話の筋を読み取るツール「Storyline」を開発した[1]。特徴量の多い意味空間上の単語群をネットワーク図で表現する際、折れ線グラフとネットワーク図を同期させることで対応をわかりやすくしている。

ネットワークの表現方法を切り替えたるためにインタラクションを用いている研究もある。Henry らによるネットワーク図と行列表現のハイブリッド方式によるネットワーク表現手法「NodeTrix」[12]では、ネットワーク図を部分的に行列表現へ切り替えるのに囲み選択とアニメーションを用いることで閲覧者の理解を助けている。また、Ichise らによるツール[2]では、書誌情報ネットワークを表現する 3 種類の手法をユーザの操作で切り替え

[1] http://prefuse.org/

ることでネットワークの把握を支援している。

従来、ネットワーク可視化におけるインタラクションの目的の多くはネットワーク及びその負荷情報を表現する複数の手法を連携することであり、ネットワークの探索を目的としたインタラクションについてはズーミングやフィッシュアイビューなどネットワークの概観から詳細な情報を提示するなどに留まった。これに対して本研究では、論文データ分析における情報探索の特徴を考慮し、局所的な情報探索のためのインタラクションを導入することによって概観情報を把握する手法を開発した。このことにより、論文データ分析の手順に沿った形で情報探索を行うことが可能となる。

7.3 本研究の位置付け

本研究の位置づけを図示すると図 7.1 のようになる。本研究は、論文データ分析支援を目的とし、書誌情報ネットワークの複数の関係に着目してその把握支援を行っている。また、ネットワークの把握支援にインタラクションを導入しており、インタラクションによってネットワークの局所的な関係構造からネットワークを探索する手法を提案、開発している。

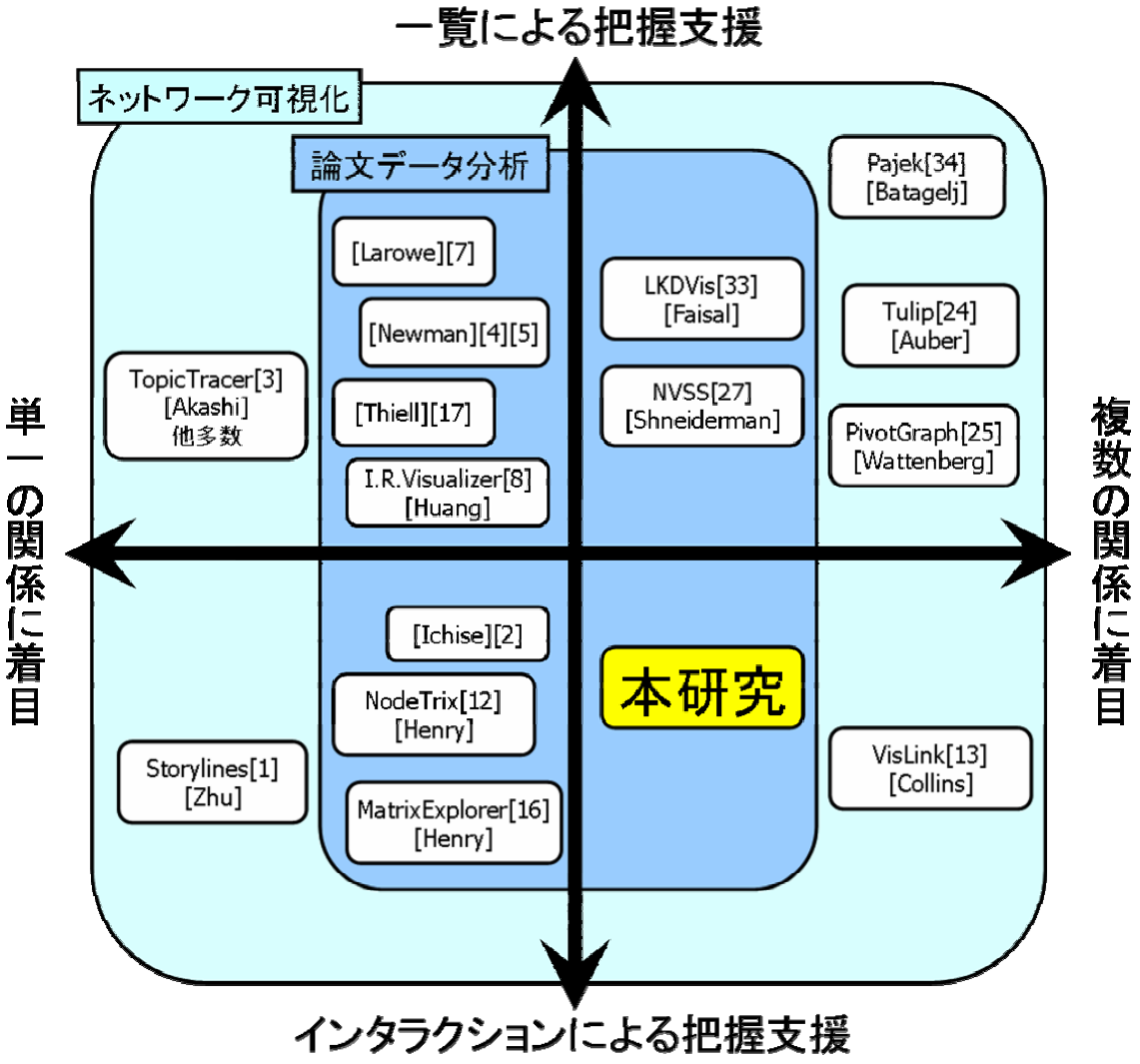


図 7.1　本研究の位置付け

第8章　結論

本研究では、論文データに含まれる有益な情報の獲得を支援することを目的とし、従来手法の問題点として書誌情報ネットワークが複数の関係からなることを考慮していない点と論文データ分析における情報探索を支援できていない点を挙げた。これらを解決するアプローチとして、分析者が得たい情報に応じてインタラクティブにネットワークビューを構築することを提案した。提案を実現するために、論文データ分析において把握すべき関係情報を分析した上で書誌情報ネットワークを分類し、分析者がネットワークビューを構築するための操作を体系化した。その上で体系化した操作を搭載した論文データ分析支援ツール「STANICOV」を開発した。STANICOV ではネットワーク可視化領域がノードの属性によって分割されており、分割したフィールド間でノードを移動させることによってネットワークを構築することができる。また、ネットワークをスムーズに構築するために、移動するノードに関連するノードを共に移動するなどの補助機能を搭載している。評価実験により、インタラクティブなネットワークの構築及びそのための操作体系は論文データ分析に有用であることが示された。

本研究により、論文データの書誌情報間の関係を総合的に把握することができるようになり、論文データからより多くの情報を獲得することが可能となった。また、論文データ分析により適したネットワーク探索を行えるようになった。実世界のデータは論文データと同様に複数の関係を持つデータが多い。本研究の提案手法は論文データだけでなく他のデータへも適用できると考えられることから、本手法はそれらデータを分析するための基礎技術となる可能性がある。

謝辞

本研究を行うにあたり、指導教員である三末和男准教授には、研究の進め方から論文執筆に至るまで、大変貴重なご助言、ご指導を頂きました。また、研究に取り組む姿勢についても親身にご指導いただき、その中で多くの事を学ばせていただきました。この修士論文を書き上げることができたのはもちろん、人間的にも大きく成長したと実感できるのは、ひとえに先生のご指導のおかげであると思っています。この場を借りて厚く御礼を申し上げます。

副指導教員である田中二郎教授には、研究内容などについて終始丁寧なご指導を頂きました。また、研究設備などの面で多大なる支援をしていただき、学外の研究発表の機会も多く与えていただきました。先生のおかげで大変有意義で充実した研究生活を送ることができました。心より感謝いたします。

志築文太郎講師、高橋伸講師には、研究発表などの場で大変貴重な議論の機会を与えていただきました。先生方に頂いたご助言は大変有意義で、研究を大きく進歩させるものでした。深く感謝の意を申し上げます。

インタラクティブプログラミング研究室の皆様には公私共に大変お世話になりました。特に、NAIS チームの皆様には感謝してもしきれません。石原正樹先輩、岩村憲一先輩、鈴木優先輩、酒井慎司先輩、中園長新先輩、佐藤修治先輩、宮崎裕久先輩の後ろ姿に研究に取り組む姿勢を学ばせていただきました。同期である成川新吾君、高杰君、後輩である伊藤隆朗君、島村祐介君、白石宏亮君、謝英双君、益子渉君、結城崇君、武田嵩太朗君、菱田哲史君、ラヴィンモホッティ君には暖かい励ましを頂きました。また、チームの皆様に終始的確なアドバイスを頂きました。皆様のおかげで 3 年間の研究室生活は有意義で忘れられないものになりました。深く感謝いたします。

学類時代からの友人には大変お世話になりました。特に、青木善友君、浅海純子さん、伊藤愛さん、海野雪絵さん、小田雅子さん、川井絵理子さん、栗林肇君、佐藤俊輔君、鈴木意織君、時田大介君、西山優香さん、松岡健太郎君、宮地健太君、村瀬香緒里さん、安岡宏俊君、渡辺友浩君には、授業や研究で行き詰まった時、論文執筆に疲れた時などに的確なアドバイスと心温まる激励を頂きました。皆さんと過ごした 6 年間はとても楽しかったです。ありがとう。

何よりも家族の支えなくしてこのように充実した研究生活を送ることはできませんでした。精神面、金銭面で多大な支えをしてくれた両親、いつも暖かい励ましをくれた兄弟に心から感謝します。

最後に、様々な面でお世話になった多くの友人、大学生活でお世話になった全ての方々に心より感謝申し上げます。本当にありがとうございました。

参考文献

[1] Weizhong Zhu, Chaomei Chen, "Visual Analytics: Storylines: Visual exploration and analysis in latent semantic spaces", Journal of Computers & Graphics, Vol. 31, pp.338-349, 2007

[2] Ryutaro Ichise, Hideaki Takeda, Taichi Muraki, "Research Community Mining with Topic Identification", Proceedings of the International Conference on Information Visualization(IV '06), pp. 276-281, 2006

[3] Mina Akashi, Koichi Hori, Ken Satoh, "Topic Tracer: a Visualization Tool for Quick Reference of Stories Embedded in Document Set", Proceedings of the International Conference on Information Visualization(IV '06), pp.101-106, 2006

[4] M. E. J. Newman, "The structure of scientific collaboration networks", Proceedings of the National Academy of Sciences (PNAS '01), vol. 98, no. 2, pp. 404-409, 2001

[5] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration", Proceedings of the National Academy of Sciences (PNAS '04), vol. 101, no. 1, pp. 5200-5205, 2004

[6] Weimao Ke, Katy Borrner, Lalitha Viswanath, "Major Information Visualization Authors, Papers and Topics in the ACM Library", Proceedings of IEEE Symposium on Information Visualization(InfoVis '04), pp.r1-r2, 2004

[7] Gavin LaRowe, Ryutaro Ichise, "Analysis of Japanese Information Systems Co-authorship Data", Proceedings of the International Conference on Information Visualization(IV '07), pp. 459-464, 2007

[8] Tze-Haw Huang, Mao Lin Huang , "Analysis and Visualization of Co-authorship Networks for Understanding Academic Collaboration and Knowledge Domain of Individual Researchers", Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation, pp. 18-23, 2006

[9] U.Brandes, M.Eiglsperger, I.Herman, M.Himsolt, M.Scott, "GraphML Progress Report", Proceedings of Symposium on Graph Drawing 2001 pp. 501-512, 2001

[10] Peter Eades, “A Heuristic for Graph Drawing,” Congressus Numerantium, 42, pp.149-160, 1984

[11] Ji Soo Yi, Youn ah Kang, John T. Stasko, Julie A. Jacko, "Toward a Deeper Understanding of the Role of Interaction in Information Visualization", IEEE Transaction on Visualization and Computer Graphics, Vol. 13, No. 6, pp. 1224-1231, 2007

[12] Nathalie Henry, Jean-Daniel Felete, Michael J. McGuffin, "NodeTrix: A Hybrid Visualization of Social Networks", IEEE Transaction on Visualization and Computer Graphics, Vol. 13, No. 6, pp. 1302-1309, 2007

[13] Christopher Collins, Sheelagh Carpendale, "VisLink: Revealing Relationships Amongst Visualizations", IEEE Transaction on Visualization and Computer Graphics, Vol. 13, No. 6, pp. 1192-1199, 2007

[14] Tony Buzan, Barry Buzan, "The Mind Map Book", BBC Books, 1993

[15] 杉山公造, "グラフ自動描画法とその応用 ―ビジュアルヒューマンインタフェース―", 計測自動制御学会, 1993

[16] Nathalie Henry, Jean-Daniel Fekete, "MatrixExplorer: a Dual-Representation System to Explore Social Networks", IEEE Transactions on Visualization and Computer Graphics, Vol.12, No.5, pp. 677-684, 2006

[17] Kilian Thiel, Fabian Dill, Tobias Kotter, Michael R. Berthold, "Towards Visual Exploratioin of Topic Shifts" , IEEE International Conference on Systems, Man and Cybernetics, ISIC, pp. 522-527, 2007

[18] Katy Borner, Luca Dall'Asta, Weimao Ke, Alessandro Vespignani, "Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams", Journal of Complexity, Vol. 10, No. 4, pp. 57-67, 2005

[19] Benjamin B. Bederson, Jesse Grosjean, Jon Meyer, "Toolkit design for interactive structured graphics", IEEE Transactions on Software Engineering, Vol. 30, No. 8, pp. 535-546, 2004

[20] Vladimir Batagelj, Andrej Mrvar, "Pajek - Program for Large Network Analysis", Journal of Connections, Vol. 21, pp. 47-57, 1998

[21] Chaomei Chen, "Searching for intellectual turning points: Progressive Knowledge Domain Visualization", Proceedings of the National Academy of Sciences of the United States of America (PNAS '04), Vol. 101, No. 1, pp. 5303-5310, 2004

[22] Chaomei Chen, “CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature”, Journal of the American Society for Information Science and Technology, Vol. 57, No. 3, pp. 359-377, 2006

[23] Graham Wills, "NicheWorks - Interactive Visualization of Very Large Graphs", Journal of Computational and Graphical Statistics, Vol. 8, No. 2, pp. 190-212, 1999

[24] David Auber, "Tulip - A Huge Graphs Visualizatioin Framework", Graph Drawing Softwares, Mathematics and Visualization, pp. 105-126, Springer-Verlag, 2003

[25] Martin Wattenberg. "Visual Exploration of Multivariate Graphs", Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 811-819, 2006

[26] A. Johannes Pretorius, Jarke J. van Wijk, "Multidimensional Visualization of Transition Systems", Proceedings of the International Conference on Information Visualization(IV '05), pp. 323-328, 2005

[27] Ben Shneiderman, Senior Member, IEEE, Aleks Aris, "Network Visualization by Semantic Substrates", IEEE Transactions on Visualization and Computer Graphics, Vol.12, No.5, pp. 733-740, 2006

[28] Adam Perer, Ben Shneiderman, "Balancing Systematic and Flexible Exploration of Social Networks", IEEE Transactions on Visualization and Computer Graphics, vol. 12, no. 5, pp. 693-700, 2006

[29] Mohammad Ghoniem, Jean-Daniel Fekete, Philippe Castagliola, "A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations", Proceedings of the IEEE Symposium on Information Visualization(InfoVis '04), pp.17-24, 2004

[30] Alan Dix, Geoffrey Ellis, "Starting simple - adding value to static visualisation through simple interaction", Proceedings of the working conference on Advanced visual interfaces(AVI '98), pp. 124-134, 1998

[31] Jeffrey Heer, Stuart K Card, James A Landay, "prefuse: a toolkit for interactive information visualization", Proceedings of IEEE Symposium on Information Visualization(InfoVis '05), pp. 421-430, 2005

[32] Sarah Faisal, Paul Cairns, Ann Blandford, “Building for Users not for Experts: Designing a Visualization of the Literature Domain”, Proceedings of the International Conference on Information Visualization(IV '07), pp. 707-712, 2007