

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

ソーシャルネットワーク分析のための
視覚的インタフェースの構築

高 杰

(コンピュータサイエンス専攻)

指導教員 田中二郎

2009年3月

概要

ソーシャルネットワークとは人や会社などの社会的単位（アクターと呼ぶ）およびそれらの間の関係によって構成されるネットワークである。ソーシャルネットワーク分析とは、そのようなネットワークを対象とし、そこに含まれる特性を解明する作業である。分析に際してはネットワーク図として視覚的に表現することで、人間の視覚を活かして、ネットワークをインタラクティブに探索する手法がしばしば利用されている。

しかしながら、これまでの手法には、アクターや関係に関する情報をただ1つのネットワーク図に描くものが多かった。そのため、ネットワークを探索する際に異なる視点から柔軟な操作が行いにくく、ネットワークの規模が大きくなるに連れ視覚的負担も大きくなるといった問題があった。

本研究では、そのような問題を解決するために、ソーシャルネットワークの構成要素に着目し、それぞれの構成要素に対応した視覚表現と各表現の連携を用いる手法を提案する。その際、アクター、つながり、コミュニティがソーシャルネットワークにおいて、最も重要な構成要素だと考えた。

提案手法に基づいて、視覚的ツール“MixVis”を構築した。MixVisでは、それぞれの構成要素に対応して、タグクラウド、ネットワーク図及びリストを用いる。また、各表現を連携するために、アクター間のつながりと集合（コミュニティ）の両方を同時に描画する複合グラフ描画手法を開発し利用した。それらの表現の連携によって、中心人物やソーシャルグループなどの情報の把握を支援する。

MixVisを用いてソーシャルネットワーク分析のタスクを達成する評価実験を実施した。実験結果により、MixVisはすべてのタスクを支援することができた。また、構成要素に完全には対応していない他のツールに比べて、MixVisの方がより使いやすことが示された。

目次

第1章 序論	1
1.1 ソーシャルネットワーク	1
1.2 ソーシャルネットワーク分析 (SNA)	1
1.3 ソーシャルネットワークの表現方法	2
1.4 本研究の目的	3
1.5 本研究の貢献	3
1.6 論文の構成	4
第2章 SNA において支援すべきタスクと必要となる技術	5
2.1 ソーシャルネットワーク分析のタスク	5
2.1.1 中心人物の特定と把握	5
2.1.2 特定要素の把握	6
2.1.3 パターンや傾向の発見	6
2.2 可視化手法の利用	6
2.2.1 情報可視化	7
2.2.2 ソーシャルネットワーク分析への応用	7
第3章 構成要素に対応した視覚表現の利用	8
3.1 概要	8
3.2 構成要素に対応した表現	8
3.2.1 タグクラウド	8
3.2.2 ネットワーク図	9
3.2.3 リスト	10
3.3 Linking & Brushing に基づく各表現の連携	11
第4章 複合グラフの描画手法	13
4.1 概要	13
4.2 複合グラフの定義	14
4.3 美的基準	15
4.4 空間効率の良い複合グラフ描画手法	16
4.4.1 論理構造の変換	16
4.4.2 レイアウト	16

4.4.3	配置エラーの軽減	17
4.4.4	レンダリング	18
4.5	描画結果	19
第 5 章	MixVis: ソーシャルネットワーク分析のためのツール	20
5.1	インタフェースの概要	20
5.2	インタフェースの設計	21
5.2.1	ネットワーク図の設計	21
5.2.2	タグクラウドの設計	23
5.2.3	リストの設計	23
5.2.4	各表現の連携	24
5.3	ツールの実装	26
5.4	システム構成	27
第 6 章	適用例	29
6.1	データの収集	29
6.2	データの編成	29
6.3	タスクの達成	29
第 7 章	評価実験	31
7.1	実験の目的	31
7.2	実験の概要	31
7.3	実験の内容	32
7.4	実験結果と考察	33
第 8 章	関連研究	36
8.1	SNA を支援するツール	36
8.2	複合グラフの描画に関する技術	37
第 9 章	結論と今後の課題	39
9.1	結論	39
9.2	今後の課題	39
	謝辞	40
	参考文献	41

図目次

1.1	グラフ表現	3
3.1	タグクラウドによるアクターの表現	9
3.2	ネットワーク図	10
3.3	リストによるコミュニティの表現	11
3.4	Linking & Brushing	12
4.1	隣接グラフ、包含グラフ、複合グラフの表現	13
4.2	クラスタを四角形、円と自由曲線で表現	14
4.3	クラスタのオーバーラップ	14
4.4	配置エラーの削除	17
4.5	ノード間の反発力の制御	18
5.1	ツールの概観	20
5.2	アクターの3種類の表現スタイル	22
5.3	アクターの選択とハイライト	22
5.4	タグクラウドのソーティング機能	23
5.5	中心性を示す吹き出し	23
5.6	リストにカラーエンコーディング	24
5.7	ネットワーク図から各ビューへの連携	25
5.8	タグクラウドから各ビューへの連携	26
5.9	リストからの各ビューの連携	27
5.10	MixVis のシステム構成	28
6.1	ソーシャルグループの発見	30
7.1	評価結果	34
7.2	コミュニティのオーバーラップ	35

表 目 次

1.1	ソーシャルネットワークの行列表現	3
4.1	隣接エッジの種類と自然長	17
4.2	階層エッジの種類と自然長	17
7.1	t 検定の結果	34

第1章 序論

1.1 ソーシャルネットワーク

ソーシャルネットワークとは、人や会社などの社会的単位(アクターと呼ぶ)をノードとし、人間と人間や会社と会社などの関係を表現したネットワークである。たとえば、友人関係や取引関係などによる様々なソーシャルネットワークを例として挙げることができる。

ネットワークは、形式的にはグラフ $G = (V, E)$ として表現される。ここで、 V はノードの集合でありアクターの集合を表す。また E はエッジの集合であり、アクター間の関係を表す。

近年、web2.0 の代表的なサービスとして、ソーシャルネットワーキングサービスが流行っている。それに伴って、ソーシャルネットワークはバーチャル世界にも広がっている。例えば、日本最大の SNS と言われる mixi¹ は、2008 年 7 月 13 日の時点では 1500 万人の会員数を突破したと言われている (mixi, Inc. のホームページより)。2004 年 1 月にアメリカで開始された MySpace² は、登録会員数 2 億人以上を有し、日本を含む 20 を超える国と地域でサービスを展開しているという。SNS を通じて、人と人の新しいつながりが増加し、ソーシャルネットワークがさらに拡大している。

1.2 ソーシャルネットワーク分析 (SNA)

ソーシャルネットワーク分析は、個人や企業の関係により構成されたネットワークに含まれる特性を解明する。例えば、ソーシャルネットワークの中で、重要な役割を果たす中心人物や親密な関係を持つアクターからなるグループ (ソーシャルグループ) を特定することである。

ソーシャルネットワーク分析にとって、グラフ理論は不可欠である。グラフ理論における中心性 (Centrality) や最短経路 (Shortest Path) などの指標を用いて、中心人物やソーシャルグループの特定などを行う。

1970 年代以後、計算機技術の発展に伴い、大規模なネットワークの分析が可能となり、ソーシャルネットワーク分析の研究が活発に行われている。人間関係、企業の取引関係、疾病の伝染関係、国の貿易関係による様々なネットワークを対象とした研究、研究が盛んである。代表的な例として、社会心理学学者スタンレー・ミルグラムが 1967 年に「スモールワールド実験」を行い、「世間は狭い」といった仮説を検証した。いわゆる「6 次隔たり」(人は自分の

¹<http://mixi.jp>

²<http://www.myspace.com>

知り合い 6 人以上を介すと世界中の誰にも間接的な知り合いになれる」という言葉で知られている [26]。

最近では、ソーシャルネットワーク分析は、企業の中で、社員と社員のインタラクションと企業の業績（業務遂行能力、仕事で得られた満足感、新しいアイデアの創造など）との関係を解明するための重要な手段となっている [14]。「親密な関係持つ社員が誰か」、「情報の交換やアイデアの創造がどのようにして行われるか」、「会社におけるリーダーのような存在を持つキーパーソンは誰か」「どのような小さなグループに形成したか」を解明することで、企業の生産性、効率の向上に大きな役割を果たしている。

また、政治活動の分析、テロリストネットワークの解析、組織分析、ナレッジマネジメント調査などを研究対象として、ソーシャルネットワークは、幅広く注目を集めている。

1.3 ソーシャルネットワークの表現方法

ソーシャルネットワークを表現するための方法の 1 つとして、2 次元の関係を表す隣接行列が用いられる。表 1.1 に示されるように、行列の 1 行目と 1 列目にアクターを並べるような $N \times N$ 行列である。アクター（ノード） v と w の間に何らかの関係（エッジ）が存在する場合、 v 行 w 列の交差する所を 1 とする。そうではなければ、その値は 0 とする。アクターはそれ自身との間に関係が築けないため、対角成分は 0 とする。また、関係の方向性を考えるか考えないかによって、値が変わる。方向がなければ、つまり $v \rightarrow w$ と $w \rightarrow v$ が等しいとき、 (v, w) と (w, v) は同値となる。逆に、方向を考えれば、 (v, w) と (w, v) の値が異なる場合もある。このような行列を統計的に処理することで、ソーシャルネットワークの特性が明らかにされている [23, 38, 5]。

しかし、隣接行列で表現したソーシャルネットワークは、直観的な方法とは言えない。なぜならば、数字や表ではソーシャルネットワークの特性を「見る」ことはできないからである。

ソーシャルネットワークに含まれた特性などを人間が直接見ることができる「図」で表現することは、ソーシャルネットワーク分析の一般的な手法と思われる。以前から、図 (visual images) がソーシャルネットワーク分析の分野で重要な役割を担っている [16]。ソーシャルネットワークの図表現は、ネットワーク構造に関する新たな知見を提供し、その知見をほかの人が使うときにも役に立つ。

表 1.1 で示されたネットワークを図で描くと、図 1.1 のようになる。ここで、円はノードを表し、ノードを結びつける直線はエッジを表す。このような図を「ネットワーク図」と呼ぶことにする。

ネットワーク図を見ることで、ネットワークの概観を把握することができる。たとえば、図 1.1 を見ることによって、このネットワークが大まかに 2 つのグループに分けられることが一目で分かる。1 つはノード A, B, C, D, E, F によって構成され、もう 1 つは G, H, I, J, K によって構成される。さらに、ノード C と G によって、この 2 つのグループが連結されることが分かる。

表 1.1: ソーシャルネットワークの行列表現

	A	B	C	D	E	F	G	H	I	J	K
A	0	1	1	1	1	1	0	0	0	0	0
B	1	0	0	0	1	0	0	0	0	0	0
C	1	0	0	1	0	1	1	0	0	0	0
D	1	0	1	0	0	0	0	0	0	0	0
E	1	1	0	0	0	0	0	0	0	0	0
F	1	0	1	0	0	0	0	0	0	0	0
G	0	0	1	0	0	0	0	1	0	1	0
H	0	0	0	0	0	0	1	0	1	1	0
I	0	0	0	0	0	0	0	1	0	1	1
J	0	0	0	0	0	0	1	1	1	0	1
K	0	0	0	0	0	0	0	0	1	1	0

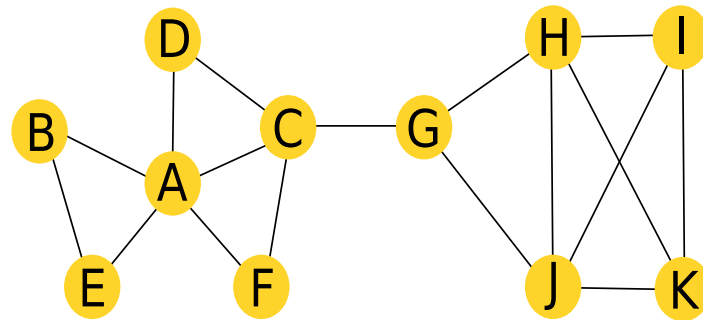


図 1.1: グラフ表現

1.4 本研究の目的

本研究では、ソーシャルネットワーク分析のための視覚的インタフェースの構築を目的とする。ソーシャルネットワークの構成要素に対応した視覚的表現を用い、それらの表現を連携することによって、ソーシャルネットワークのより柔軟な探索を可能にする。

1.5 本研究の貢献

ソーシャルネットワークの構成要素に着目し、各要素に対応した視覚表現と表現の連携によって、ソーシャルネットワーク分析を容易にする手法を開発した。この手法を用いることによって、分析者が注目する要素からソーシャルネットワークの分析ができ、効率よくタスクを成し遂げることができる。

1.6 論文の構成

続く第2章では、ソーシャルネットワーク分析において、支援すべきタスクと必要となる技術について述べる。第3章で、本研究の提案手法として、構成要素に対応した視覚表現を用いる手法を記述する。第4章で、アクター間のつながりとコミュニティを両方描画できる複合グラフの描画手法について述べる。第5章では、提案手法を元に開発した、ソーシャルネットワーク分析のためのツールについて詳しく説明する。第6章では、本ツールの適用例として、ソーシャルネットワーキングサービスから抽出したソーシャルネットワークの分析への利用について述べる。第7章では、提案手法の有効性と使いやすさを検証するための評価実験について述べる。第8章で本研究の関連研究を述べる。最後に、第9章では、本研究の結論と今後の課題について述べる。

第2章 SNAにおいて支援すべきタスクと必要となる技術

本章では、ソーシャルネットワークを分析するときに、支援すべきタスクと必要となる可視化技術について述べる。

2.1 ソーシャルネットワーク分析のタスク

ソーシャルネットワーク分析が達成しようとするタスクについて、Freeman は次のように述べている [8]。

For the most part, they (SNA) seek to uncover two kinds of patterns: (1) those that reveal subsets of actors that are organized into cohesive social groups, and (2) those that reveal subsets of actors that occupy equivalent social positions, or roles.

上述のものは多くのソーシャルネットワーク分析ツールの方針となってきた。そして、類似したコンセプトが Scott や Wasserman らによって主張された。[34, 35]。

以下、SNA のタスクについて具体的に説明する。

2.1.1 中心人物の特定と把握

我々が日常生活でよく耳にする「中心人物」や「キーパーソン」という言葉は、使う場面によって、それらの意味が異なるかもしれない。ソーシャルネットワーク分析においては、「中心性 (centrality)」という指標があり、それぞれのアクターがネットワークの中でどの程度中心的であり、どの程度末端な存在なのかを厳密に計測する。代表的な指標は、次数中心性、媒介中心性、近接中心性、ボナチッチ (固有ベクトル) 中心性が挙げられる。その中で、本研究は最も利用されている次数中心性と媒介中心性に注目する。ネットワークの中で、ハブのように、多くの他のアクターにつながっているアクター、あるいは、アクターとアクターの橋渡しとなるアクター（コネクタと呼ばれる）の特定は非常に重要だと思われる。このようなアクターを特定するために、アクターの中心性 (centrality) を測るという手段が用いられる。

- 次数中心性 (Degree Centrality)

それぞれのアクターがネットワークの他のアクターとのつながっている数を計算し、その数が多いほど次数中心性が高いとする。本研究で研究対象とするネットワークは無向

グラフのため、アクターに直接接続しているほかのアクターの数アクターの次数中心性となる。現実世界では、多くの友達を持つ外向的な人が中心人物になると思えば良いだろう。図 1.1 に示されたネットワークにおいて、アクター A は最も多い 5 つのつながりを持つので、次数中心性が一番高いとみなされる。

- 媒介中心性 (Betweenness Centrality)

任意のアクターにおいて、すべてのペアのアクター間の最短パスがそのアクターを通る回数が多いほど中心性が高いとする。本指標は、ネットワーク内のつながりを維持するために不可欠なアクターを表す指標とも言える。何故なら、媒介中心性の高いアクターがなくなれば、多くのつながりが切れてしまうからである。図 1.1 のグラフでは、すべてのアクターが、2 つ集団に分かれているように見える。その中、1 つの集団 A 、 B 、 C 、 D 、 E 、 F から、もう 1 つの集団 G 、 H 、 I 、 J 、 K へ最短パスで辿るときは、アクター C と G を通らなければいけない。すなわち、 C 、 G は 2 つの集団の橋渡しを果たす重要な存在である。そのようなアクターの媒介中心性が高いとする。

2.1.2 特定要素の把握

特定の要素の把握は、分析者がネットワークを探索するとき、気になったアクター、つながり、またはコミュニティについてさらに詳細を調べるタスクである。このタスクは、フィルタリングとも呼ばれ、多くのつながりとアクターから特定の要素に関連する情報の獲得を行う。意味のあるネットワークの一部の作成と提示によって、ソーシャルネットワークを分析する。

2.1.3 パターンや傾向の発見

ソーシャルネットワークから隠れたパターンや潜在的な傾向を発見することは SNA のもうひとつのタスクである。社会ネットワーク分析の対象としてのデータは、関係データ (relational data) と呼び、個々のアクターの属性だけではなく、多くのアクターのシステムとしての性質でもある。その性質はさらに特定の社会現象や社会的出来事を反映する。例えば、松尾らは、「mixi には、中心的ユーザ（次数中心性が高いユーザ）が形成する、大きな 2 つのクリーク構造が存在することは興味深い事実である」と指摘している [40]。Henry らは、可視化ツール NodeTrix を用いることで、国際会議 InfoVis の著者関係を分析した結果、著者らに cross pattern と block pattern の 2 つの共著パターンを発見した。ソーシャルネットワーク分析は、何らかの興味深い社会現象の解明、または検証のために働く。

2.2 可視化手法の利用

本研究は、ソーシャルネットワークを分析するために、統計的な数値を用いるよりも、視覚的に分析者が作業することに着目した。いわゆる可視化技術を用いて、ソーシャルネットワークを探索し、SNA の各タスクを成し遂げるのを支援することを目指す。

2.2.1 情報可視化

情報可視化とは、情報を図形など分かりやすい形で表現することである。ネットワークとグラフは、いずれも抽象的なものであるため、人間が直観的に理解することは簡単ではない。しかしながら、我々が日常的に見慣れてきた「図」による表現は、ネットワークの構造やノードの関係を分かりやすくすることができる。

2.2.2 ソーシャルネットワーク分析への応用

現在のソーシャルネットワーク分析は、視覚表現とそれに基づいたインタラクション技術を用いるアプローチが非常に多い。具体的には、まずネットワークの概観の提示から始まり、次に、ネットワーク表現へのインタラクティブな操作（フィルタリングまたはクラスタリング）によって、ネットワークの一部の図を作成する [21, 1]。それにより、ネットワークの概観の把握とともに、部分ネットワークの詳細などの更なる分析も可能になる。

第3章 構成要素に対応した視覚表現の利用

本章では、ソーシャルネットワーク分析のために、本研究で提案する「構成要素に対応した視覚表現を用いた手法」を紹介する。まず、提案手法の概要について説明する。そして、構成要素に対応したタグクラウド、ネットワーク図、リストについて詳しく述べる。最後に、これらの表現を連携する Linking & Brushing 技術の利用について述べる。

3.1 概要

ソーシャルネットワークを分析するとき、分析者はそれぞれの構成要素に視点を置き、中心人物などの探索を行うと考えられる。それぞれの構成要素は独自の特徴を持ち、異なる情報を提供する。分析者側は、必ずネットワークの全体か構成要素の一つに着目し、分析を行うと考えられる。一般的には、ネットワーク全体から俯瞰するとき、ネットワーク図を見るが、気になったノードやコミュニティの詳細をチェックするときは、ノードとコミュニティからの柔軟な探索を支援するインタフェースが必要となる。

また、これまでの研究の多くは、ノードの属性（たとえば中心性）とコミュニティ情報をネットワーク図にによって表現する傾向がある。小規模のネットワークの場合は、ノードとコミュニティの詳細の把握は容易であるが、リンクとノードの数が増えるにつれ、ノードの特定と詳細の把握は難しくなる。従って、より大規模なネットワークに対して視覚的に探索しやすいインタフェースが求められる。本研究では、ソーシャルネットワークの構成要素に対応した視覚表現を用いた、ソーシャルネットワークの分析ツールを構築した。

3.2 構成要素に対応した表現

ソーシャルネットワークにおいては、アクター、つながり、コミュニティが最も重要な要素であると考え、それらの要素に対応したタグクラウド、ネットワーク図とリストを採用した。

3.2.1 タグクラウド

SNA の各タスクを見ると、アクターの最も重要な特性は中心性である。分析者は中心性の高いアクターの特定と把握を望む。そのため、たくさんのアクターからこれらのアクターを迅速に見つける機能が必要となる。また、ハブとコネクタの特定、名前による検索などの支援が求められると考えられる。

アクターを表現するのに、本研究ではタグクラウド (tagcloud) を用いた。タグクラウドは一連の単語を並べた視覚表現である。通常は、タグ (本質的にはテキスト) のサイズ、色、重要度は、関連する単語の特徴を表す (図 3.1 参照)。現在のタグクラウドは、ソーシャルブックマークや画像共有サービスといったソーシャルタギングを行っているサービスにおいて多く採用されている。たとえば、写真を共有するサイトとして知られる flickr¹は、タグクラウドを利用し、タグの大小で人気のあるタグを一目で分かるようにしている。

タグクラウドを提示することで、アクターの検索、ブラウジングをより容易に行うことができる。さらに、タグクラウドの最大の利点は「印象形成」に及ぼす効果があると言われている。タグクラウドを見ることで、中心性の高いアクターをすぐに特定できる。タグクラウドの代替案として、リストでアクターを表現する手法もあるが、タグクラウドの方が多くのアクターを一覧できるといった面では優れていると考えた。今回実装したタグクラウドを図 3.1 に示す。

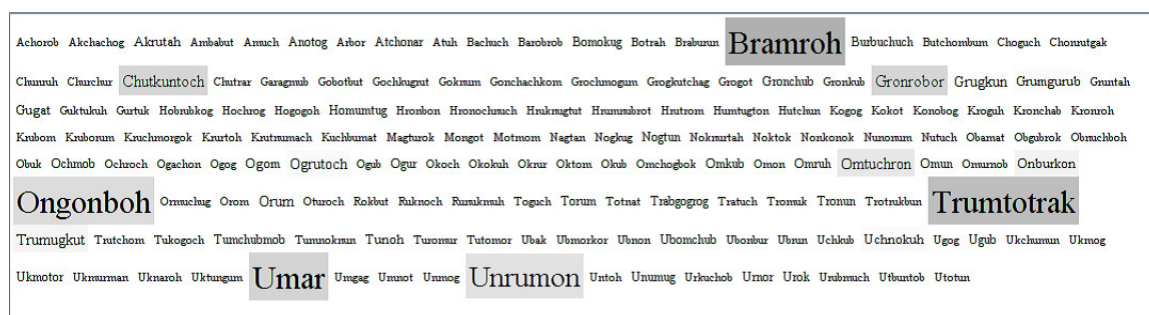


図 3.1: タグクラウドによるアクターの表現

3.2.2 ネットワーク図

つながりはネットワーク図に含まれるので、ネットワーク図とともに説明を行う。ソーシャルネットワークとグラフの対応関係は第 1.3 節で述べた通り、アクターとつながりはグラフ理論におけるノードとエッジに対応しており、ネットワーク図の描画は、グラフの描画技術の観点で考えた。

図 3.2 では、本研究で開発したツールでネットワークを描いた図の一例である。この図は、一定の美的基準に従って、力指向描画法によって描画された。単純な力指向描画法では、ノードとリンクを均等に配置することができるが、ノードの集合 (コミュニティ) を同一空間上に提示するためには、更なる工夫が必要とされた。

そのため、多くの研究は、Newman[28] が提案したグループ検出アルゴリズムを用いて、アクターのグループを求め、グループのすべてのメンバーを囲む凸包領域で表現する [31, 20]。または、コミュニティの構成を自由に組んで、行列表現で表す [22]。

¹<http://www.flickr.com>

これらの手法では、コミュニティのインタセクション、つまり複数のコミュニティへの所属を表現するのはできない。この問題を踏まえ、筆者はアクターとアクターのつながりかつアクターの集合の両方を描画する複合グラフの描画手法を開発した [18]。複合グラフの描画手法については、論文の第 4 章にて詳しく説明する。

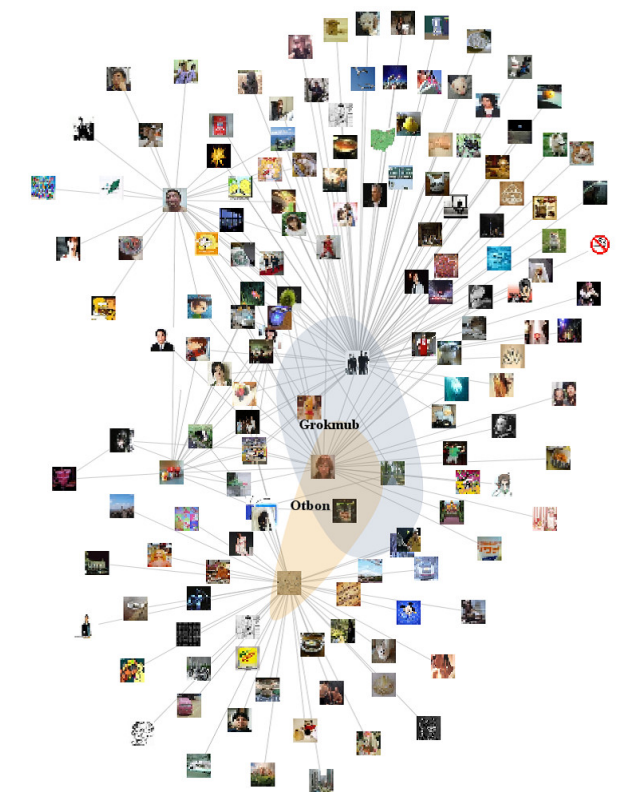


図 3.2: ネットワーク図

3.2.3 リスト

コミュニティの数を考慮したうえ、一般的に用いられるリストを使うことで、コミュニティを一覧できるようにする。

リストは、いくつかの情報をエンコードすることができる。最も知られているものは、アイテムの順番である。通常は、何らかのランキングに使われているが、その用途はランキングだけには限らない。

また、アイテムの背景色も情報の表現に利用できる。本研究では、リストを用いて、すべてのコミュニティを提示するうえ、アイテムの順番と背景色によって、コミュニティに関する情報を表す。図 3.3 では、本研究におけるリスト表現を示す。

	Achokog
	Agochoh
	Bonkah
	Chubkukmor
	Chukrorob
	Humom
	Kobkotot
	Krobug
	Kubchuh
	Kunob
	Marum
	Obkogran
	Obugun
	Ochok
	Omgoh
	Rurot
	Trachtuk
	Trutbokgab
	Ukburor

図 3.3: リストによるコミュニティの表現

SocialAction[31, 33, 32] にも、中心性によりソーティングしたアクターリストを搭載しているが、本研究は、数が多いアクターをタグクラウドで表現し、比較的少ないコミュニティをリストで表現する。

3.3 Linking & Brushing に基づく各表現の連携

Linking & Brushing とは、同じデータの複数のビューを連携させることで、1つのビューの変化を他のビューでも反映させるインタラクション手法である。Linking はデータを異なるビューで人間か分かる見せ方で提示することを指す。良く利用されるのは、同じデータを同じ色でハイライトする手法である。Brushing は Linking のコンセプトを更に拡張する。Brushing を用いることによって、任意のビューへのインタラクティブな選択操作とリアルタイムでのビューの更新を実現する。

Linking & Brushing の利点として、まず、異なるビューからデータの側面を把握することができる。また、複数のビューから気になったデータやデータ間の関係からデータの分析が可能となる。

Linking & Brushing の利点を生かして、ソーシャルネットワークの各要素に対応した表現を連携させる。タグクラウド、リンク、ネットワーク図に Linking & Brushing を適用することで、各要素についての詳細情報を獲得することができると期待している。

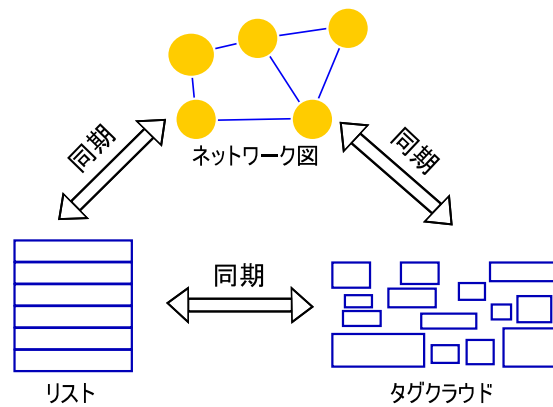


図 3.4: Linking & Brushing

上述の Linking & Brushing の定義の通り、Linking & Brushing は任意のビューへの操作（例えば、選択操作）に強い関係がある。本研究は、図 3.4 のように、各ビューの同期によって、リアルタイムな連動を実現する。そして、ノード、タグ、アイテムの選択・選択の取り消しによる各ビューの状態を同一にする。

本研究で用いる Linking & Brushing を実現するために、複合グラフの描画手法が必要とされる。なぜならば、ネットワーク図の上にコミュニティを描くために、ノード間のリンクとノードの集合の両方を提示する必要があるからである。

第4章 複合グラフの描画手法

ネットワーク図上に複数のアクターを含むコミュニティを提示するために、複合グラフの描画手法を用いた。本章では、まず複合グラフの描画手法の概要を述べる。次に複合グラフの数学的定義、描画が従う基準と描画手法について説明する。

4.1 概要

アクターとアクターの関係は2種類に分けられる。1つは、友人関係や取引などの2つのアクター間の関係である。それらはnode-link図で表現できる。もう1つは、複数のアクターからなるコミュニティ（アクターの集合）と考えられる。コミュニティを表示するためには、視覚的に「メンバーを包含する」ように見える図が良いと考えられる。そこで、我々がよく知っている、集合関係を表すベン図を本研究で用いる。これらの2つの関係を描画するために、複合グラフの描画手法を開発した。

本研究で用いる複合グラフは、隣接関係と包含関係の2種類の関係を表すグラフ構造と定義した。複合グラフは隣接グラフと包含グラフを統合し、それらエッジを異なる表現で表す。

隣接エッジは直線で表現され、包含エッジは包含領域で表現される。隣接グラフと包含グラフの1つの表現をそれぞれ、図4.1のaとbに示す。その2つの表現を組み合わせた表現、すなわち、複合グラフ表現を図4.1のcに示す。

他のノードを包含する（包含エッジでつながる）ノードはクラスタと呼ぶ。今回開発した複合グラフの描画手法では、クラスタのインタセクションを許し、インタセクションを包含領域の共通領域として表現する。

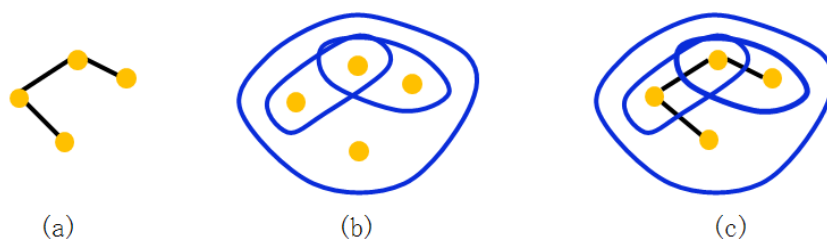


図 4.1: 隣接グラフ、包含グラフ、複合グラフの表現

複合グラフの描画手法は、[37, 36, 30, 13]により議論された。それらの研究では、クラスタを四角形や円のようなレギュラーな形で描画する (図 4.2a,b 参照)。しかしながら、それらの

手法で描いた図の空間効率はよくない。そこで、複合グラフ表現の空間効率を向上するために、クラスタを自由曲線で描画する手法を開発した [18]。提案手法は次のアイデアを基に展開した。

- 複数のノードの位置に合わせた曲線は、円や四角形より少ない空間を占める。

図 4.2c で示されたように、自由曲線はゴムバンドのようにノードを包み込むことに対して、四角形や丸は無駄な空間を空ける。従って、自由曲線の方が空間効率を向上させることが期待できる。

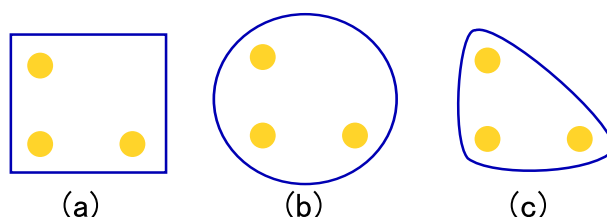


図 4.2: クラスタを四角形、円と自由曲線で表現

- 曲線で描くことによって、クラスタとクラスタをさらに近づけることができる。

円や四角形がより多くの空間を占めるため、クラスタのクラスタのオーバーラップが発生しやすい。図 4.3 に示した通り、同じ場所に配置した 2 つのクラスタを、円と四角形で描くと、赤く色づけした部分でクラスタがオーバーラップしてしまう。しかしながら、自由曲線の場合は、オーバーラップを避けることが可能となる。オーバーラップの回避は、グラフ描画において、基本的な規則である [36]。

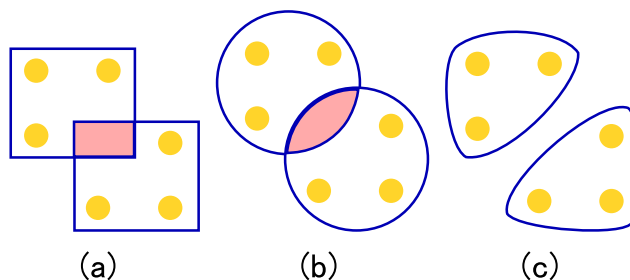


図 4.3: クラスタのオーバーラップ

4.2 複合グラフの定義

複合グラフ G は包含グラフ $G_c = (V, F)$ と隣接グラフ $G_a = (V, A)$ の組み合わせとして定義する。ただし、 V はノードの集合、 F は包含関係を表すエッジの集合、 A は隣接関係を表す

エッジの集合である。複合グラフの数学的定義は次の通りである。

$$\begin{aligned} G &= (V, F, A), \\ V &= \{v_1, v_2, v_3, \dots, v_n\} \\ F &\subset \{(m, n) \mid m, n \in V, m \neq n\} \\ A &\subseteq \{\{u, v\} \mid u, v \in V, u \neq v\} \end{aligned}$$

本描画手法において、包含グラフにおいて一つ以上のノードを含むノードをクラスタと呼び、包含されたノードをクラスタの子と呼ぶ。ほかのノードを含まないノードをリーフと呼ぶ。

包含エッジはクラスタから子に向く有向 (directed) エッジであり、子を囲む領域で表現する。その一方で、隣接エッジは無向エッジとし、ノードをつなぐ直線分で表す。

4.3 美的基準

図の視覚的な「良さ」とは抽象的なものであるため、それを具体化するために「美的基準」を設ける必要がある。本研究で提案する手法は空間効率の良いかつ人間にやさしい図を目指す。その目標を向けて、描画規約と描画規則を設計した。

描画規約は、グラフを描画するときに、必ず満たさなければならない約束を指す。本手法で従う規約は、次の通りである。

- リーフは円か四角形で表す。
- クラスタは自由曲線で表す。
- 隣接エッジは直線で表す。
- 包含エッジはクラスタの包含関係で表す。
- クラスタは自身が含むすべてのノードを囲むように描画される。
- 複数のクラスタに含まれるノードはそれらの交わりの領域に配置される。

描画規則はできるだけ満足すべき条件を意味する。本手法で設定した規則を優先順に並べると次の通りである。

- i 空間効率を最大化する
- ii ノード間の不要なオーバーラップを避ける
- iii クラスタはリーフと適当な距離を保持する

4.4 空間効率の良い複合グラフ描画手法

複合グラフの構造を持つデータを入力した後、提案手法では、まずその論理的構造を変換する。変換の目的は、複合グラフをスプリングモデル [11] にあてはめることである。スプリングモデルは単純グラフのノードを平面上に自動的に描画する手法であり、多くのグラフ描画で採用されている。次に、すべてのリーフの位置がスプリングモデルによって決められる。最後に、リーフが円または四角形で、エッジは直線で、クラスタ自由曲線で描かれる。

4.4.1 論理構造の変換

包含エッジを隣接エッジとみなして、包含エッジの集合 F と隣接エッジの集合 A を 1 つのエッジ集合 E に結合することによって、複合グラフを単純グラフに変換する。変換の結果となる単純グラフの式は次の通りである。

$$G' = (V, E) \quad (4.1)$$

$$E = A \cup \{(u, v) | (u, v) \in F\} \quad (4.2)$$

新しいグラフ G' はスプリングモデルによって自動的に配置することができる。

4.4.2 レイアウト

スプリングモデルでは、エッジは接続するノードを反発または引き寄せるばねとみなされる。エッジのばねの力は式 4.3 の通りである。

$$f_s = C_s \times (l - d) \times \frac{1}{d} \quad (4.3)$$

ここで、 d はノード間の距離であり、 l はばねの自然長とする。 C_s はばねの強さをコントロールする定数である。

また、ノードは電荷とみなされる。ノード間が互いに反発する力が働くようにする。電荷の反発力式 4.4 のように定めた。

$$f_r = \begin{cases} C_r \times \frac{1}{d} & d \leq L_0 \\ 0 & d > L_0 \end{cases} \quad (4.4)$$

ここで、 C_r は定数であり、 L_0 は反発力が働く最短距離とする。

スプリングモデルを利用する際に、ノード間の適切な反発力とばねの適切な自然長を設けることは、グラフのレイアウトに影響を及ぼす重要なパラメータと考えられる。本研究では、複合グラフにおけるノードの種類（クラスタかリーフ）によって、それらを結びつけるばねを分類した。また、エッジの種類によって、異なる自然長を与えた。表 4.1、4.2 はエッジの種類と自然長を示す。

表 4.1: 隣接エッジの種類と自然長

	cluster	leaf	
cluster	l_1	l_2	
leaf	l_2	sister	otherwise
		l_3	l_4

表 4.2: 階層エッジの種類と自然長

	leaf	cluster
cluster	l_5	l_6

ここで、これらのエッジの自然長は経験的に決定されるものである。我々は、各種類のエッジの自然長の最適な組み合わせを探した。勿論、最適な組み合わせは、できるだけ前述の美的基準を満たすものである。ここでは、それらの自然長の大まかな関係を式 4.5 に示す。

$$(l_1, l_2, l_6) > (l_4, l_5) > l_3 \quad (4.5)$$

4.4.3 配置エラーの軽減

クラスタに包含されないノードがクラスタを表現する領域内に配置されることを配置エラーと呼ぶ。エッジを分類し、適切な値に割り当てることによって、ある程度配置エラーを避けることができるが、それを更に減少させるための方法を考案した。

配置エラーを取り除くために、配置エラーを検出し、配置エラーが起こったノードにかかわる斥力を調整する。具体的に配置エラーが起こったノードとそのクラスタのすべてのノードの間に、標準より長い距離を保持するように斥力を作用させることで、そのノードをクラスタから押し出す。

式 4.4 が示した通り、反発力はノード間の距離 d の線形関数である。しかし、この力は、 $d \leq L_0$ の条件を満たさなければ働かない。グラフを再配置するときに、図 4.4 の星形のように、本来クラスに属していないノードが領域内に配置されたら、再配置における配置エラーの検出回数 N と一定の距離 ΔL の積で、最短距離が増え続ける。

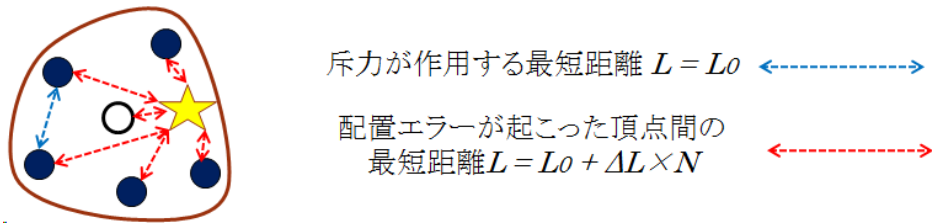


図 4.4: 配置エラーの削除

しかしながら、ノードをランダムな位置から配置する初期状態では、この反発力の調整は行わない。その理由は、スプリングモデルは、力の物理的なバランスを取るために、ある程度美的基準を満たす位置で、ノードを配置するためである。

アクター間の反発力 f_r のコントロールを図 4.5 に示す。複合グラフの描画処理は、配置フェーズ、調整フェーズ、バランスフェーズの3つのフェーズに分けた。フェーズの遷移は、スプリングモデルにおいて安定状態を求めるためにノードを少しずつ移動させる各ステップにおけるすべてのノードの移動量によって決められる。最初は、ノードがランダムに画面上に配置される。その時点では、力のバランスが取れていないため、移動が非常に激しく、配置フェーズに入る。少し時間が経つと、ノードの移動量が減少し、微調整フェーズに入る。そこで、配置エラーの有無を検出する。もし配置エラーがあれば、それを取り除くようにプログラムが働く。力の調整が原因で、一時的配置フェーズに戻る可能性もある。ノードの移動量が設定した閾値より少なければ、バランスフェーズに入る。

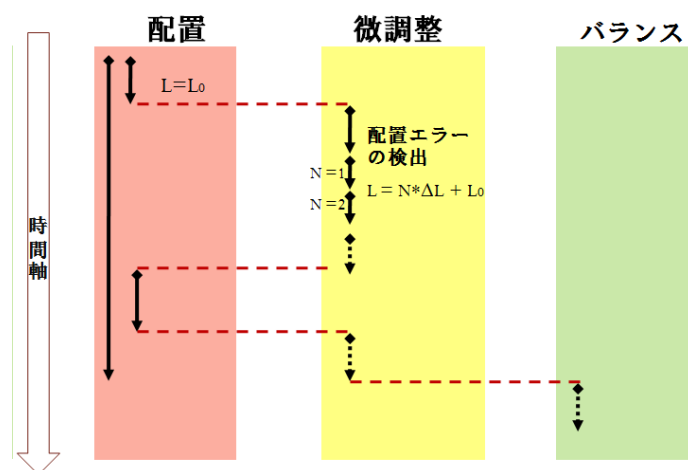


図 4.5: ノード間の反発力の制御

4.4.4 レンダリング

ノードは描画規約で定めた通りに、円か四角形で描かれる。エッジはノードをつなぎ直線分で描画される。

クラスタをレンダリングする時に、まず、クラスタのすべての子の座標を計算する。次に、Graham Scan[19]を用いて、それらの座標を含む最少の多角形(いわゆる凸包)を求める。次に、凸包の中心を求めて、中心から多角形の頂点までの線を適当な距離で延ばして、新しい多角形を得る。拡大した多角形は元の凸包の相似形となり、ノードとの間に適切な隙間ができる。最後に、拡大した多角形を基に、閉じたカーディナルスプラインを描く。メンバーの位置が変わったら、すべてのメンバーを包含するように、その曲線も動的に変化する。

4.5 描画結果

本描画手法で描いたグラフを図 3.2 に示す。描いた複合グラフは、ノード数 150、隣接エッジ数 200、包含エッジ数 8 とされる。ノードが四角形 (図) で示され、エッジが四角形をつながる直線分で表現される。境界線が滑らかな曲線である陰影領域はコミュニティを示す。

第5章 MixVis:ソーシャルネットワーク分析のためのツール

本章では、提案手法を元に開発した、ソーシャルネットワーク分析のためのツール“MixVis”について述べる。まず、MixVis の概要の説明を行う。その後、インターフェースの設計について詳しく述べる。最後に本ツールを実現するためのシステム構成を紹介する。

5.1 インタフェースの概要

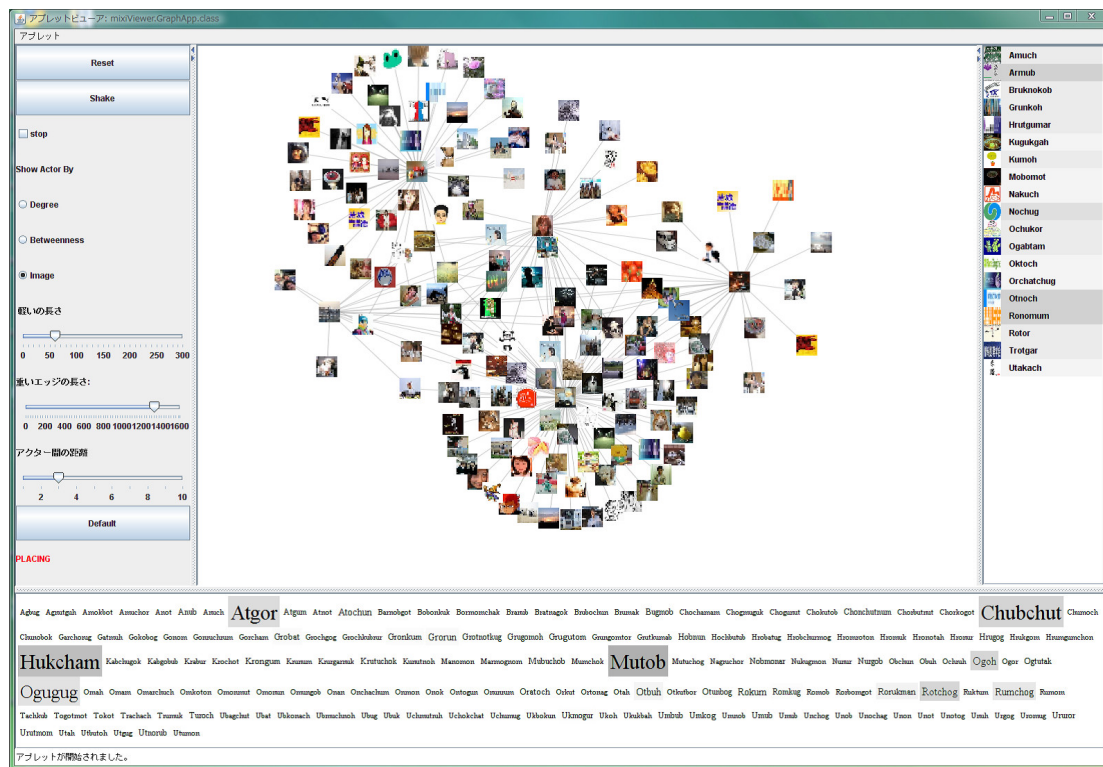


図 5.1: ツールの概観

提案手法に基づいて、ソーシャルネットワーク分析のための視覚的ツール MixVis を開発した。図 5.1 のように、MixVis は主に 3 つのビューで構成される。ソーシャルネットワークに関

するほとんどの情報を持つネットワーク図はインタフェースの真ん中に配置される。ネットワーク図の左側のパネルにはネットワーク図のレイアウトを調整するコントロールパネルが搭載される。右側のリストビューではアクターが所属するコミュニティのリストを示す。下部のタグクラウドはすべてのアクターをタグ形式で表す。これらのビューの設計は次節で詳しく述べる。

5.2 インタフェースの設計

ソーシャルネットワーク分析に向けて、柔軟性かつ良い操作性が備えるインタフェースの構築を目標にして、MixVis を設計した。

5.2.1 ネットワーク図の設計

ネットワーク図を描くときに、複合グラフの描画手法を用いる。ソーシャルネットワークのアクター、つながり、そして、コミュニティを複合グラフのノード、エッジ、そして、クラスに適用させる。

アクターは、次数中心性、媒介中心性、そして、写真情報のそれぞれを表す、3つの表現を持たせる。そのために、予めアクターの次数・媒介中心性を計算した。次数中心性は、アクターのつながりの本数を数えた。媒介中心性は Brandes のアルゴリズム [7] によって算出する。アクターには以下の3種類の表現スタイルを用意した。

- 1つは、図 5.2 の中央の図に示したように、アクターが名前を囲む四角形で表す、背景色でアクターの次数中心性を表すスタイルである。次数中心性の高いアクターは、薄い黄色の背景とする。このスタイルは、アクターの次数が注目される時に、ネットワークの概観を見ながら、アクターの大体の次数中心性の把握ができるように設計した。しかし、正確な次数（媒介）中心性の高さは、この図だけでは、明確には分からない場合もある。
- 図 5.2 の右側の図では、アクターが名前を囲む四角形で表され、背景色をアクターの媒介中心性とするスタイルである。濃い赤のアクターは媒介中心性が高いとした。
- それらに加え、アクターの写真を用いて、アクターを示すスタイルを備える (図 5.2 の左側の図参照)。アクターの写真は、名前よりも印象が残りやすいため、ネットワークでアクターの特定をより行いやすくなると考えた。

それらのスタイルを切り替えるための機能をコントロールパネルに入れた。

アクターが誰とつながっているかを知りたいときは、そのアクターをクリックすると、そのアクターにつながっているアクターが図 5.3 のようにハイライトされる。また、他の直接接続していないアクターを半透明することによって、調べたいアクターをより見やすくしている。さらに、アクターを Drag & Drop することで、アクターの位置を調整することができる。

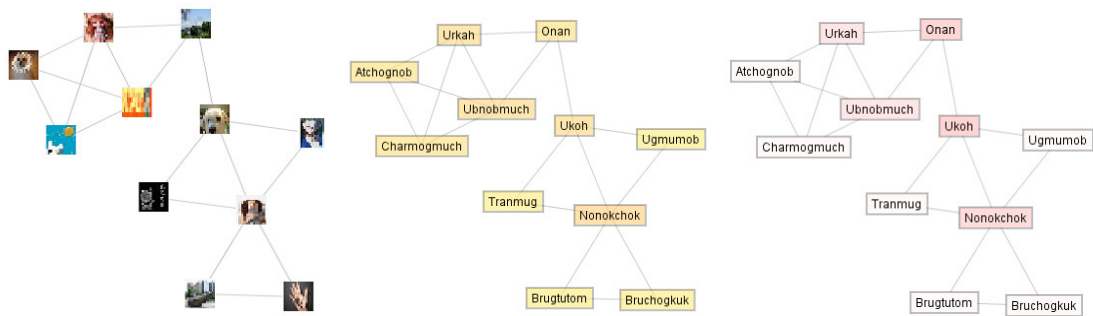


図 5.2: アクターの3種類の表現スタイル

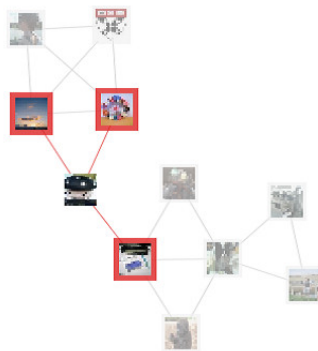


図 5.3: アクターの選択とハイライト

ネットワーク図のレイアウトを調節するための他の機能は次の通りである。

- アクターの位置を初期化して、再配置するときは、Reset ボタンを押す。
- アクターの重なりがあったとき、Shake ボタンをクリックすることで、アクターの位置を微調整する。
- 「Stop」ボックスをチェックすると、アクターの揺れを止める。
- 直線につながっているアクターとアクターの距離を大きくしたり、小さくしたりするときは、「エッジの長さ」スライドで調整する。また、その直線の長さを調節することで、ネットワークの「概観」と「詳細」を把握することができる。
- すべてのアクターの距離を調整するときは、「アクター間の距離」スライダを動かす。
- 「Default」ボタンを押すと、エッジの自然長とアクター間の反発力が初期値に戻る。

5.2.2 タグクラウドの設計

タグのテキストではネットワークにあるすべてのアクターの名前を示す。タグのサイズはアクターの次数中心性を表す。大きいタグはアクターの次数中心性が高いとする。タグの明るさはアクターの媒介中心性を表す。背景の暗いタグはアクターの媒介中心性が高いとする。タグをクリックすることによって、そのタグを選択する。タグ以外の場所で右クリックすると、ポップアップメニューが出てくる。メニューにある Sort by alphabet、Sort by Degree Centrality、Sort by Betweenness Centrality から 1 つのアイテムを選べると、すべてのタグはそのアイテムによってソートされる (図 5.4 参照)。

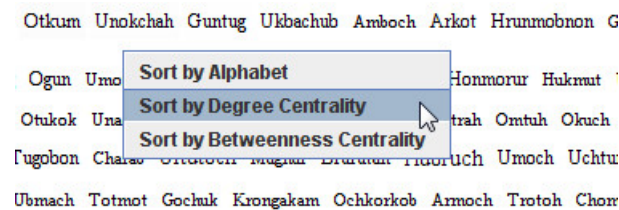


図 5.4: タグクラウドのソーティング機能

アクターの次数中心性と媒介中心性の正確な比較を行うときに、その二つの指標の実際の値が必要となる。そのため、次数中心性と媒介中心性を数値で表示する機能を搭載した。図 5.5 の通りに、マウスがタグの上に移動したら、タグの下方に緑色の吹き出しが表示され、アクターの次数中心性と媒介中心性が示される。

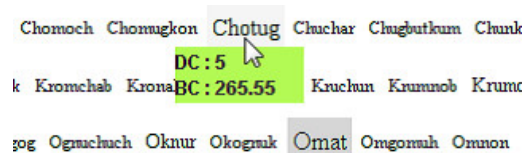


図 5.5: 中心性を示す吹き出し

5.2.3 リストの設計

リストはアクターが所属するコミュニティを示す。リストのアイテムの明るさはコミュニティのメンバー数を表す。暗いアイテムはコミュニティのメンバー数が多いとする。リストのアイテムをクリックすることで、コミュニティを選択することができる。Ctrl+Click と Shift+Click を使うことで、複数のコミュニティの選択も可能である。

コミュニティを選択するとき、区別のために、それぞれのコミュニティに異なる色を付けた。具体的は、HSV モデルにおける色相、彩度、明度の三つの成分をコントロールすること

によって、色相を選択する。リスト内のすべてのコミュニティを選択すると、図 5.6 のようになる。

任意のコミュニティをダブルクリックすると、非選択状態に戻る。



図 5.6: リストにカラーエンコーディング

5.2.4 各表現の連携

これらの3つのビュー連携は、通常1回の操作で終わるものではない。様々な視点から、様々な操作を繰り返すことによって、ネットワークの理解がどんどんと深まる。

- ネットワーク図から各ビューへの連携

ネットワーク図のアクターをクリックすると、そのアクターを表すタグがタグクラウドでハイライトされる（赤いボックスまたは青いボックス）。そして、そのアクターが所属しているコミュニティが異なる色でハイライトされる。図 5.7 は、ネットワーク図のアクターをクリックした後の、各ビューのフィードバックを示す。

分析者がネットワークの概観、またはアクター間のつながりに着目して探索を行うとき、ネットワーク図から、気になったアクターを特定し、そのアクターの情報を取得する。図 5.7 のように、各ビューを覗くだけで、気になったアクターの次数中心度（または媒介中心度）が把握でき、アクターの所属を簡単に理解することができる。

- タグクラウドからの各ビューの連携

タグクラウドでタグをクリックすると、そのタグが表すアクターはネットワーク図でもハイライトされ、アクターが属しているコミュニティもハイライトされる。

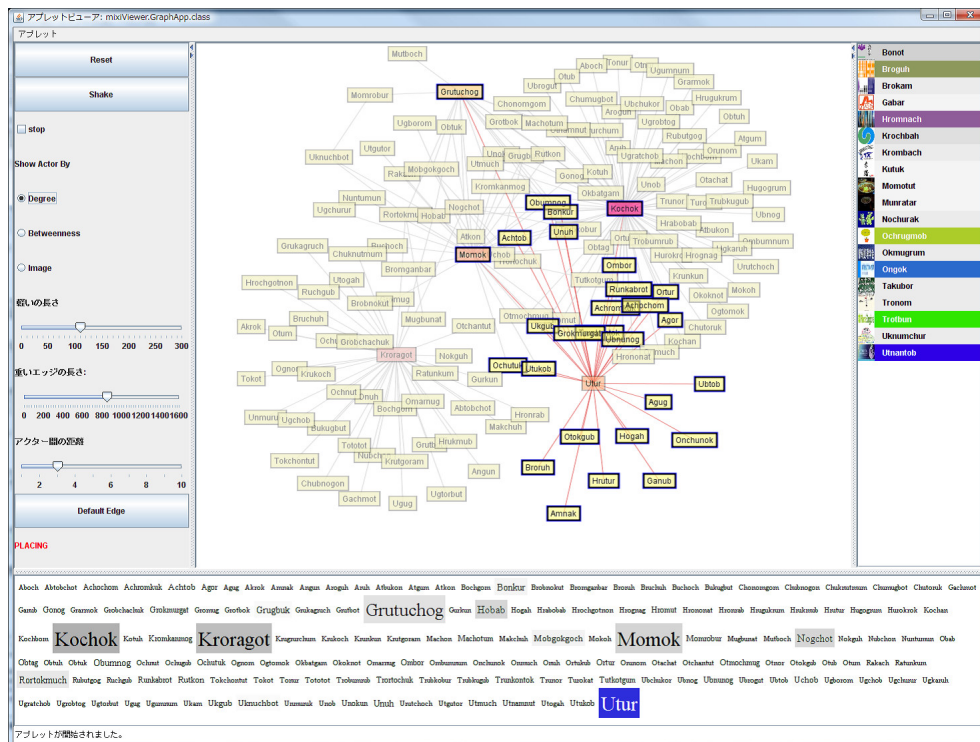


図 5.7: ネットワーク図から各ビューへの連携

図 5.8 は、ひとつのタグをクリックした後の MixVis のスナップショットである。分析者が特定のアクター（たとえば、名前が「筑波太郎」や次数中心性上位 3 人や媒介中心性最も低いアクターなど）に興味を持ち、そのアクターがどのようなアクターにつながっているか、どのようなコミュニティに属しているかを調べるときに、この図のように、MixVis は分析者の作業を支援する。

- リストから各ビューの連携

リストのアイテムが選択されると、そのコミュニティのメンバーを囲む領域がアイテムと同じ色でハイライトされる。メンバーを表すタグもその色でハイライトされる。複数のコミュニティに属しているアクターのタグは、それらのコミュニティを混ぜ合わせた色で表示される。図 5.9 は 2 つのコミュニティを選択した後の各ビューの様子を示している。

リストでは、選択した 2 つのコミュニティの背景色が茶色と赤色となった。ネットワーク図で対応したコミュニティが、同じ色で透過性のある領域で描かれた。ここで、複合グラフの描画手法におけるコミュニティの描画によって、コミュニティのメンバーがその領域に囲まれ、共有されるメンバーが領域のインタセクションに配置される。そして、コミュニティに入っていないアクターは、できるだけ領域から取り除かれるよう

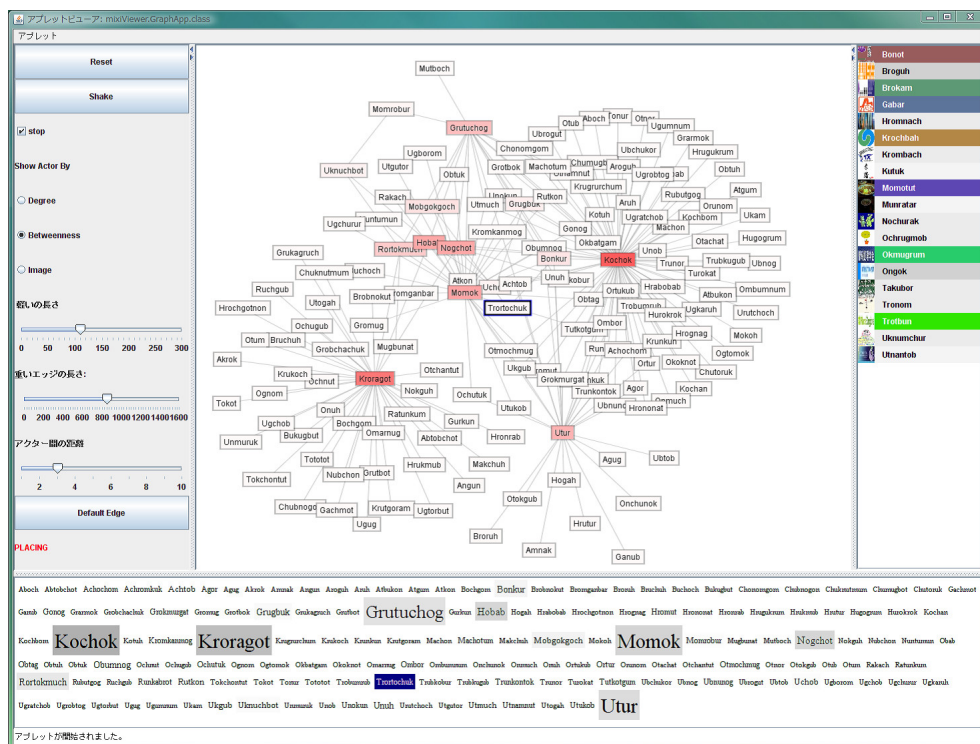


図 5.8: タグクラウドから各ビューへの連携

にプログラムが働いている。この図を見ると、コミュニティの構成や共有されたメンバーなどが容易に把握できる。タグクラウドからコミュニティの構成が分かる。加えて、アクターの中心性もはっきり見える。図 5.9 のタグクラウドを見ると、選択した 2 つのコミュニティは、次数中心性の高いアクターと低いアクターの両方が入っていることが分かる。そして、次数中心性の最も高いアクターもいることが分かる。

5.3 ツールの実装

本研究では、ソーシャルネットワーク分析のための視覚的ツール MixVis を開発した。実装の言語としては、Java™Platform, Standard Edition 6 Development Kit¹を用い、MixVis は Java™Applet として開発した。

システムの入力データには GraphML データを用いた。GraphML は、グラフを記述する XML に基づくファイル形式であり、多くのグラフ描画分野の研究において採用されている [6]。

¹<http://java.sun.com/>

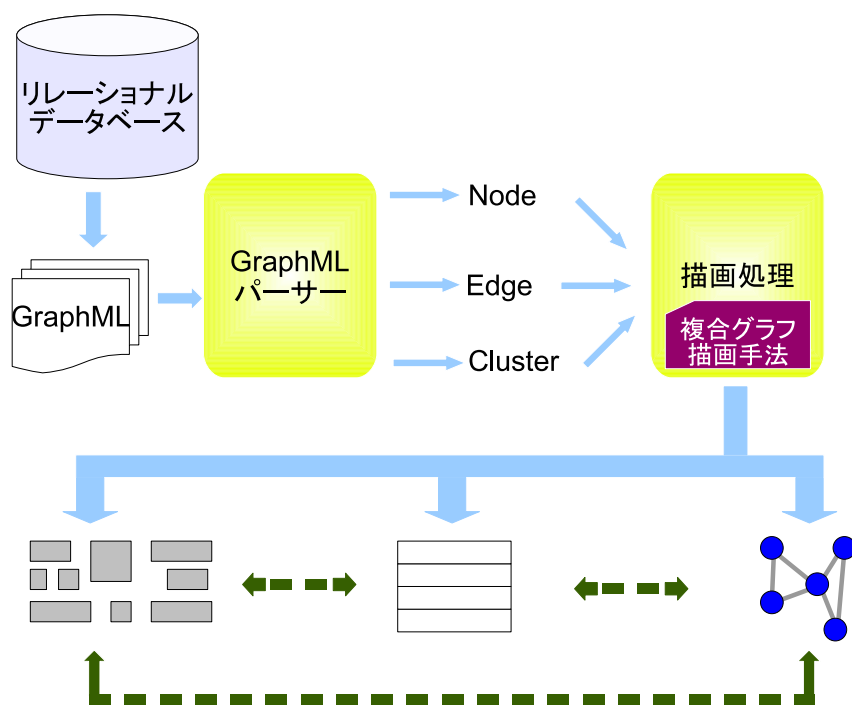


図 5.10: MixVis のシステム構成

第6章 適用例

本章では、MixVis の適用例として、実際の SNS から抽出したソーシャルネットワーク分析について述べる。

6.1 データの収集

利用例に用いたデータは日本国内のあるソーシャルネットワーキングサービス (SNS) から収集した。その SNS では、ユーザが登録した後、自分の知り合いや友達とバーチャルな世界の関係を作っていく。興味や学校などをテーマとした機能を備え、それによって、ユーザが自由にコミュニティを作ったり、参加したりすることができる。

インタフェースの説明と実験を行うために、その SNS から、ユーザ数約 6 万人、コミュニティ数 3 万個を抽出した。

6.2 データの編成

抽出したデータはそれぞれレコードとして、データベースに入れた。そのデータから、さらに、複合グラフの論理構造を持ちソーシャルネットワークを表す GraphML データを編成した。

今回用いたネットワークの規模は、150 人の法則 [10] に沿って設定した。「150 人の法則」とは、現実の社会的ネットワークにおけるメンバーは 150 程度の人数に限定される、というものである。

6.3 タスクの達成

- 重要人物の特定と把握

ソーシャルネットワークの重要人物を特定するときに、ネットワーク図の次数・媒介中心性モードを切り替えて、中心性の高い人を利用者の目で発見する方法がある。

もし中心性の高い複数のアクターがあるとする、中心性の大小を正確に目で判断することは難しい。しかしながら、タグクラウドのソーティングとラベル表示機能が、このタスクを簡単にする。

重要人物の特定だけではなく、重要人物の重要性を分析者にわかりやすく示すことも視覚的ツールの役目である。従って、特定した後、各ビューの連携によって、利用者

の視線がネットワーク図へ移り、そのアクターのネットワークでの位置を視覚的に理解する。

- 特定要素の把握 (フィルタリング)

このタスクでは、利用者が注目するアクター、リンク、またはコミュニティから、図 5.8、5.7、5.9 で示したような三つのビューの連携によって、各要素に関連する情報が獲得できる。そして、柔軟な操作が、ネットワークに含まれたパターンや傾向の発見を助ける。

- パターンの発見の支援

まず最初に、Freeman が述べたソーシャルグループの発見を MixVis でも支援することを可能にした。ネットワーク図のレイアウトを調整することによって、図 6.1 のような、ネットワークの概観を示す図を得た。その図を見ると、いくつかの塊が明らかになった。

赤い矢印が指している塊は、お互いに親密の関係を持つアクターから成る、いわゆるソーシャルグループである。そして、それらの塊をつなげる橋渡しのような存在であり、重要人物と言われるアクターは、この図から一目で特定することができる。

青い矢印がソーシャルグループのコネクタを果たすアクターを示している。これらのアクターは、塊と塊の間に配置され、どちらにもつながっているという特徴がある。

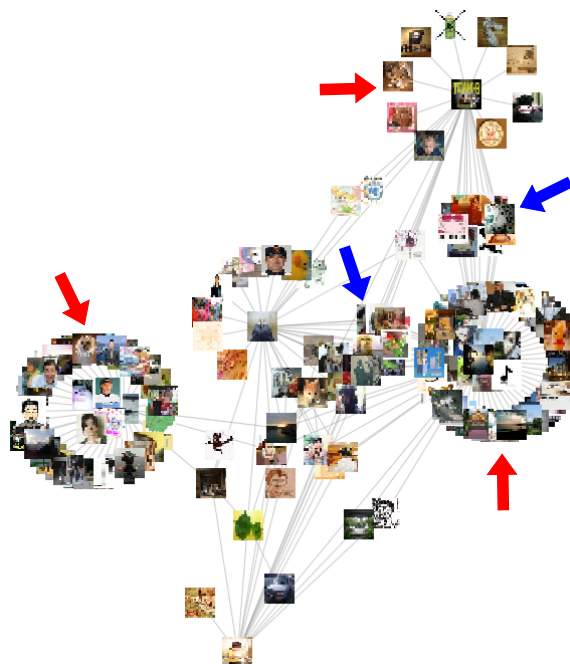


図 6.1: ソーシャルグループの発見

第7章 評価実験

7.1 実験の目的

今回の評価実験では、主に以下のことを目的とする。

- 本ツールを用いることによって、前述した SNA のタスクを達成できるかどうかを検証する
- 提案手法として、構成要素に対応した表現及び各表現の連携の使いやすさを評価する

7.2 実験の概要

ほとんどの情報をネットワーク図だけで表現する手法と比較するために、本インタフェースの一部の機能を外した視覚的ツール ToolA と ToolB を作った。ToolA と ToolB の具体的な説明は次の通りである。

- ToolA:

MixVis のタグクラウドとリストを除外し、タグクラウドで表すアクターの中心性をネットワーク上のノードの色で提示するツールである。それで、写真で表現したアクターをテキスト(名前)だけで表し、テキストの背景色を利用して、重要人物などの特定を行う。そして、コミュニティリストを利用せずに、コミュニティ情報もネットワークにかぶせる。

ToolA はこれまでの多くの SNA ツールの代表で、ソーシャルネットワークに関するあらゆる情報をネットワーク図だけで表現するものである。

- ToolB:

ToolB は、MixVis のリストをはずして、ネットワーク図とタグクラウドだけを利用し、コミュニティ情報をネットワークで示すツールである。

従来手法で利用されたもうひとつの手段として、リストなどを使ってノードの関連情報を独立のビューで示し、異なるビュー間の連携を活用したツールも少なくない。ToolB は、そのようなツールの代表である。

ToolA と ToolB では、コミュニティの表示と非表示を切り替えるチェックボックスをコントロールパネルに入れた。

ツールの名前が評価実験に影響を及ぼさないように、MixVis に仮名 “ToolC” を付ける。

ToolA、ToolB、ToolC を同時に起動して、それぞれのツールを利用して、設定したタスクを実行してもらう。

3つのツールで利用するソーシャルネットワークのデータは同じ構造を持つが、アクターとコミュニティの名前は全部違うものとした。ただし、この情報を被験者には伝えずに、実験した。

3つのツールを使用する順番は定めず、被験者に自由にツールを使ってもらった。

評価実験には20代の大学生および大学院生に協力してもらった。被験者は全員が情報系の学生であり、視覚的インタフェースについてはある程度分かるが、ソーシャルネットワーク分析についての知識はほとんど持っていない。

7.3 実験の内容

実験用データは第6章の適用例に用いたデータから抽出した153人によって構成されたソーシャルネットワークを用いた。153人のメンバーからなるグラフは、リンク数202、コミュニティ数19の規模となった。

まず、以下の内容について、著者が被験者に10分程度簡単に説明した。

- ソーシャルネットワークの定義
- ソーシャルネットワーク分析の目的とタスク
- 実験に使うデータについて
- ツールの機能と使い方

次に、ToolA、ToolB、ToolC(MixVis)のそれぞれを使って、ソーシャルネットワーク分析を想定した7項目のタスクを実行してもらった。

- タスク1: ハブのような役割を果たすアクターを把握する。

まず、次数中心性の最も高いアクターを特定してもらう。そのアクターの役割を理解したかどうかを検証するために、それが持つつながりの数を見積もってもらい、ハブ、コネクタ、あるいは、わからないという3つ選択肢から、アクターの役割を選んでもらう。

- タスク2: グループを連結するコネクタを把握する。

まず、図6.1のような概観図を作成してもらう。次に、その図から複数のソーシャルグループをつなぐ役割のアクターを特定してもらい、それらの役割を正しく理解したかどうかを確かめる。

- タスク 3: 名前によるアクターの特定制とそのアクターのつながりやコミュニティ情報を把握する。

具体的には、与えた名前によってアクターを特定してもらい、そのアクターが所属するコミュニティとそのアクターにつながっているアクターの名前を書いてもらう。

- タスク 4: つながりによるアクターの特定制はそのアクターに関する情報を把握する。

このタスクでは、多くの中心性の高いアクターにつながるアクターを特定してもらい、そのアクターの中心性と所属するコミュニティを書いてもらう。

- タスク 5: コミュニティの特定制とコミュニティの構成を把握する。

このタスクでは、多くの中心人物が参加しているコミュニティとこれらのコミュニティの共有メンバーを特定してもらう。

- タスク 6: ソーシャルグループを発見する。

アクターとアクターの関係の近さによって推定されたソーシャルグループを読み取れるかどうかを検証する。

- タスク 7: すでに発見したパターンを検証する。

このタスクでは、「次数中心性の高いアクターが多くのコミュニティに参加している」という傾向を検証する。パターンの発見ではなく、検証するタスクを設計する理由はパターンの発見はいろいろな要素にかかわっているためである。まず、実験用のデータにパターンや傾向が必ずあるとは言えない。また、パターンの発見は被験者によって、不確定な面があるので、本ツールを使うことで、必ずパターンを発見することができるとは言えない。そもそも提案手法の役割は、柔軟な探索手法を提供することで、パターンの発見を容易にすることである。そのため、評価実験では、筆者が知っている傾向を被験者に教えておいて、各ツールを使ってもらい、その傾向を読み取れるかどうかを答えてもらう。

7.4 実験結果と考察

実験を進めながら、アンケート調査を行った。

アンケート用紙に記入された被験者の答えによると、ほとんどのタスクは、3つのどのツールでも、達成することができることがわかった。

各タスクを行った後、3つのツールの使いやすさについて点数(1~5点)を付けてもらった。5点が一番良いとした。

10人が付けた点数の平均値を図7.1にまとめた。すべてのタスクで MixVis による操作の方が、他の2つのツールより良い結果が得られた。

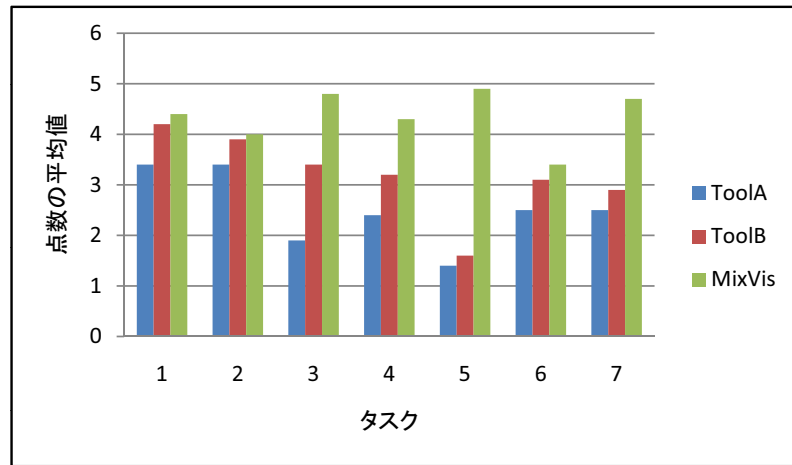


図 7.1: 評価結果

表 7.1: t 検定の結果

	ToolA と MixVis の t 検定		ToolB と MixVis の t 検定	
	$ t $	有意差	$ t $	有意差
タスク 1	3.87	○	1.50	×
タスク 2	1.96	×	0.56	×
タスク 3	12.43	○	4.58	○
タスク 4	10.58	○	4.71	○
タスク 5	15.65	○	12.68	○
タスク 6	2.86	○	1.41	×
タスク 7	11.00	○	7.22	○

3つのツールにつけた点数の平均値に有意な差があるかどうかを検証するための t 検定を行った。有意水準として、5%を定めた。両側検定における t の境界値は 2.26 となる (自由度 9)。検定の結果を表 7.1 に示す。

タスク 1 とタスク 2 では、一つだけの検定に有意な差があった。その理由は、中心人物の特定は、タグクラウドのソーティングを使わずに、ネットワーク図のアクターの背景色だけで、行えたからだと考えられる。

タスク 3、タスク 4 及びタスク 6 では、狙い通りに、被験者は ToolB と MixVis の各ビューでの操作と連携を用いて、タスクをこなした。従って、連携のない ToolA、2つのビューの連携を持つ ToolB、3つのビューの連携ができる MixVis に有意な差が出た。タスク 6 における ToolB と MixVis の検定で、有意な差がない理由が、タスクを実行する際に、リストを使わずに、タグクラウドのソーティング機能を利用したと考えられる。

タスク 5 とタスク 7 においては、MixVis の平均値の差が非常に大きい。 t 検定の結果によっ

て、有意な差があると示された。この2つのタスクでは、コミュニティの表示が必要とされることが原因のようである。ToolA と ToolB では、コミュニティを一斉に表示すると、コミュニティの重なりが多く、コミュニティの名前と構成を特定し難いと考えられる。図7.2はToolA と ToolB でコミュニティを描いたネットワーク図である。一方で、MixVis はコミュニティを表示するリストを提供することによって、興味あるコミュニティだけを表示可能である。これらのタスクの達成が容易であったと考えられる。

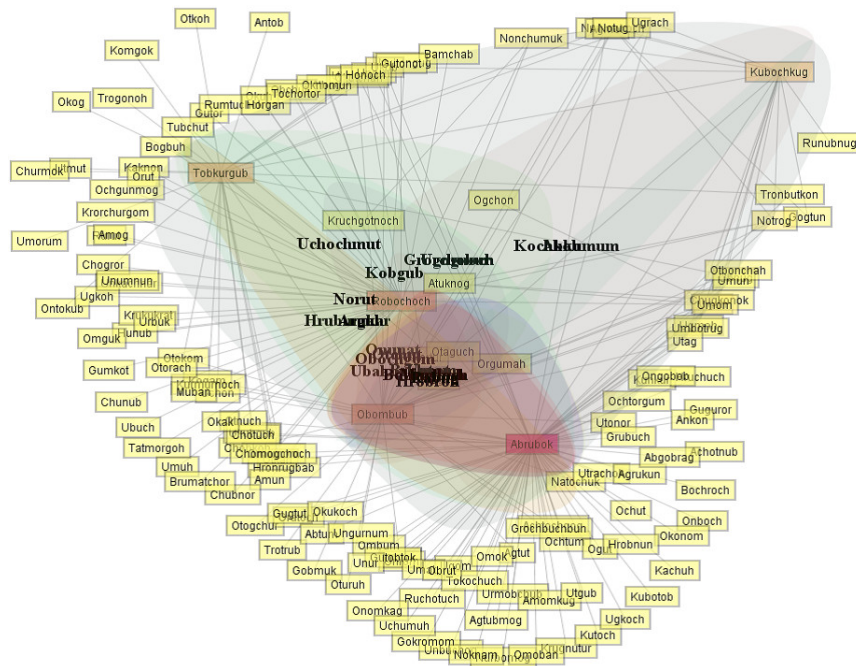


図 7.2: コミュニティのオーバーラップ

アンケートの結果から、タグクラウドとリストの有用性及び各ビューの連携の便利さがわかった。タグクラウドとリストの有用性について、すべての被験者は「あった方が良い」か「なくてはいけない」という選択肢を選んだ。ビューの連動性が便利か不便かの質問に対して、すべての被験者が「とても便利」か「便利」と答えた。

さらに、MixVis を操作することによって、実験に用いられたデータから、新たな知見を得ることができた。今回のタスクを実行するときに、データの意義を考慮したうえ、気になったことと読み取ったことを書いてもらった。例えば、「あらゆるコミュニティの中に、一つの次数中心性の高いメンバーが入っている」、「次数中心性の高いアクターは媒介中心性も高い傾向がある」などが発見された。

第8章 関連研究

8.1 SNAを支援するツール

ソーシャルネットワーク分析を支援するツールは主に統計的分析を中心にしたツールと視覚的表現を中心にしたツールの2種類に分けられる。UCINET[5]とPajek[9]に代表される分析ソフトウェアは主に統計的な分析手段を用いて、メニューに基づくツールを提供することで、ネットワークに関する様々な情報を提供する。

UCINETは、行列データを入力として処理を行い、多次元スケーリングやクラスタ分析などの機能を備える。それらを使うことで、ソーシャルネットワークのほとんどの特徴を数値化することができる。

Pajek[3]は大規模ネットワークの分析と視覚化のためのツールである。Pajekでは、クラス係数、ネットワーク密度、中心性など、ネットワークの特徴を表す様々な値を計算することができる。また、可視化機能を用いて、ネットワークを効果的に分析することができる。

StOCNET[4]は、6種類の統計モデルに基づく統計的手法を使って、ソーシャルネットワークを分析する。さらに、新しいルーティンをStOCNET06に組み込むことも可能である。

以上のツールは、統計的な分析と視覚的表現の両方を提供するが、比較的統計的なモデルを重視し、基本的なnode-link図しか提供しない。

NetDrawはUCINETを元に組み込まれた分析ツールである。PajekやStOCNETとは異なり、ソーシャルネットワーク図に焦点を合わせて分析を行う。具体的には、ドット単位でノードの位置調整やフルカラーに対応する色表現が行え、Pajekよりも豊富な描画機能を有している。

ソーシャルネットワークを表現する図形に着目して、その図形をどのように賢く描くか、図とユーザ間にどのような効率的なインタラクションを設計するかを焦点に合わせる研究が最近盛んである。

Heerらによって開発されたVizter[20]は、オンラインソーシャルネットワークのためのインタラクティブな視覚的ツールである。それを用いることによって、friendster¹、tribe²、orkut³などのSNSの社会構造を探索することができる。Vizsterは、ネットワークのアクター間のリンクを探るのに、インタラクティブなソシオグラムを提供する。それに加え、つながりによるコミュニティの特定機能とコミュニティの異なる表現スタイルを備える。

PivotGraphは、「多変量」を持つグラフ(例えば、ノードの複数の属性)の描画手法に着目し、ネットワーク構造の分析を行うツールである[39]。グローバルなグラフトポロジーの視

¹<http://www.friendster.com>

²<http://www.tribe.net>

³<http://www.orkut.com>

覚化ではなく、PivotGraph は、ノードの属性とつながりに焦点を合わせ、グリッドベースのアプローチを使用する。

Henry らは、Node-link 図と行列の 2 つの表現を用いる、ソーシャルネットワークを探索するためのシステム MatrixExplorer を開発した [21]。そのシステムは、フィルタリングとクラスタリング機能の他に、行列を再配置することで、異なるレイアウトとクラスタの比較による知見の発見を支援する。また、ユーザによく知られている Node-link 図を提供することで、探索結果の伝達を支援する。

NodeTrix は、隣接行列と Node-link 図を複合したネットワーク表現を用いる、社会ネットワークの可視化システムである [22]。Node-link 図がネットワークのグローバルなネットワークの構造を示すことに対して、隣接行列はネットワークの任意の部分を表す。この研究のもう 1 つの重要な貢献として、NodeTrix への柔軟な操作によって、2 種類の表現の切替・変形という機能を提供する。

SocialAction は、Perer らによって開発されたソーシャルネットワーク分析のためのツールである [31, 33, 32]。SocialAction は、可視化技術と統計データを結合することによって、分析プロセスを改善する。

ここまで紹介したツールは node-link に基づくネットワーク図を主な表現とするが、node-link 以外の視覚表現を用いるツールも挙げられる。

NetLens[24] は複数の順序付きリストとヒストグラムの協調したビューを用いて、情報のコンテンツ・アクターモデルを表す。コンテンツ・アクターの例として、科学出版物とその著者の関係、電子メールと受信者・送信者の関係などが挙げられた。NetLens は、コンテンツとアクターのネットワークのペアをリンクしたビューで示し、利用者が繰り返して 1 つのビューから、もう 1 つのビューへフィルターしたデータを転送することで、クエリを洗練することを可能にした。

TreePlus[25] は、ツリースタイルのレイアウトに基づいた、インタラクティブなグラフ可視化システムである。TreePlus ではグラフが木構造とされ、可視化とインタラクション技術によって、失われたグラフ構造を明らかにされる。

8.2 複合グラフの描画に関する技術

複合グラフの描画に関する研究は、今まで多く行われている。

Misue & Sugiyama[27, 37] は、ノードを四角形で表し、子を含む四角形で表す描画手法を提案した。

Omote & Sugiyama[29, 30] は、力指向描画法に基づく Intersecting Compound Mixed Graph(ICMG)の描画法を開発した。

Frishman らは、グラフのクラスタ構造を維持しながら、変化するグラフを動的に描画するアルゴリズムを提案した [17]。

Eades ら [12] は、orthogonal-grid rectangular cluster drawings として知られている規約を踏まえ、複合グラフの平面描画 (Planar drawings) を生成するアルゴリズムを開発した。Feng[15]

はその技術にある問題を解決し、複合グラフのマルチレベル表現を提案した。

Balzer ら [2]、2D と 3D における大規模で複雑な複合グラフを分かりやすくするための level-of-detail 可視化技術を提案した。本研究では、クラスタを単純化した表現として、陰関数曲面を用いる。そして、エッジのバンドルを用いることで、エッジの表現を単純化する。グラフの概要から詳細までのビューを独特な遷移手法によって提供する。

第9章 結論と今後の課題

9.1 結論

ソーシャルネットワーク分析を支援するために、ソーシャルネットワークの構成要素のそれぞれに対応した視覚表現を用いる手法を開発した。そして、各表現を連携するために、アクターのつながりと集合の両方を同時に表せる複合グラフの描画手法を開発した。これらの技術に基づいて、ソーシャルネットワーク分析のための視覚的ツール“MixVis”を開発した。

評価実験により、MixVisを用いることで、ソーシャルネットワーク分析の代表的なタスクを実行できることが示された。また、MixVisは、構成要素に対応する表現を備えていないツールよりも使いやすいことを示すことができた。

9.2 今後の課題

本研究で開発した MixVis では、ソーシャルネットワーク分析において参考にするべき指標のうち、次数中心性、媒介中心性、クリーク (ソーシャルグループ) を利用した。それら以外にも参考にするべき指標はある。例えば、近接中心性、推移性、エゴネットワークなどが挙げられる。これらの指標の計算機能及び視覚的な表現機能を導入することでルールを更に便なものにできると考えられる。

また、MixVis に用いた各表現についても更なる改善の余地がある。例えば、タグクラウドとリストのカラーコーディングで、タグ (アイテム) の明るさとハイライト色に異なる意味を与えたため、たまに間違えるケースがあった。また、評価実験により、3つ以上のコミュニティをネットワーク図で表示した場合、コミュニティを特定し難くなるという指摘を受けた。そのような問題を解決するために、コミュニティの配置とラベルの提示法を改善する必要がある。

謝辞

私に日本で留学・研究する大変貴重なチャンスを与えて頂き、また、この研究の遂行にあたり終始適切な助言を下さった田中二郎教授に感謝いたします。

副指導教員である三末和男准教授には、日頃から研究の進み具合を気にかけていただき、優しい言葉で私を励まして頂きました。ありがとうございました。

専攻の研究発表にあたり、志築文太郎講師と高橋伸講師に的確なアドバイスを頂きまして、ここに感謝いたします。

そして、筑波大学システム情報工学研究科コンピュータサイエンス専攻インタラクティブプログラミング研究室 (IPLAB) のメンバーには、研究活動や日常生活など到大変お世話になりました。

最後に、経済面や精神面で、多大なサポートをしてくれた家族、そして友人に心より感謝を申し上げます。ありがとうございました。

参考文献

- [1] Eytan Adar. Guess: a language and interface for graph exploration. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 791–800, New York, NY, USA, 2006. ACM.
- [2] M. Balzer and O. Deussen. Level-of-detail visualization of clustered graph layouts. In *Asia-Pacific Symposium on Visualisation 2007*, 2007.
- [3] Vladimir Batagelj and Andrej Mrvar. Pajek - program for large network analysis. *Connections*, Vol. 21, pp. 47–57, 1998.
- [4] Peter Boer, Mark Huisman, Tom A.B. Snijders, Christian E. G. Steglich, Lotte H.Y. Wichers, and Evelien P.H. Zeggelink. Stocnet: An open software system for the advanced statistical analysis of social networks, version 1.7. *Groningen:ICS/Science Plus*, 2006.
- [5] S. P. Borgatti, M. G. Everett, and L. C. Freeman. Unicet for windows: Software for social network analysis. *Analytic Technologies, Inc.*, 2002.
- [6] U. Brandes and C. Pich. GraphML Transformation. *GD*, pp. 89–99, 2004.
- [7] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, Vol. 25, pp. 163–177, 2001.
- [8] P.J. Carrington, J. Scott, and S. Wasserman. *Models and Methods in Social Network Analysis*. Cambridge University Press, 2005.
- [9] Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj. *Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences)*. Cambridge University Press, January 2005.
- [10] Robin Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard Univ Pr, 10 1998.
- [11] Peter Eades. A heuristic for graph drawing. *Congressus Numerantium*, Vol. 41, pp. 149–160, 1984.
- [12] Peter Eades and Qing-Wen Feng. Drawing clustered graphs on an orthogonal grid. In *Graph Drawing*, pp. 146–157, 1997.

- [13] Peter Eades, Qing-Wen Feng, and Xuemin Lin. Straight-line drawing algorithms for hierarchical graphs and clustered graphs. In *GD '96: Proceedings of the Symposium on Graph Drawing*, pp. 113–128, London, UK, 1997. Springer-Verlag.
- [14] K. Ehrlich and I. Carboni. Inside social network analysis. Technical report, IBM Watson Research Center, 2005.
- [15] Qingwen Feng. *Algorithms for Drawing Clustered Graphs*. PhD thesis, 1997.
- [16] Linton C. Freeman. Visualizing social networks. *Journal of Social Structure*, Vol. 1, pp. 132–133, 2000.
- [17] Yaniv Frishman and Ayellet Tal. Dynamic drawing of clustered graphs. In *IEEE Symposium on Information Visualization (INFOVIS'04)*, pp. 83–91, 2004.
- [18] Jie Gao, Kazuo Misue, and Jiro Tanaka. Drawings of compound graph using free-form curves. In *Proceedings of the 70th National Convention of IPSJ*, 2008.
- [19] Ronald Graham. An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters*, Vol. 1, pp. 132–133, 1972.
- [20] Jeffrey Heer and Danah Boyd. Vizster: Visualizing online social networks. In *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, Washington, DC, USA, 2005. IEEE Computer Society.
- [21] Nathalie Henry and 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.
- [22] Nathalie Henry, Jean-Daniel Fekete, and Michael J. McGuffin. Nodetrix: A hybrid visualization of social networks. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, Vol. 13, No. 6, pp. 1302–1309, 2007.
- [23] Paul W. Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, Vol. 76, No. 373, pp. 33–50, 1981.
- [24] Hyunmo Kang, Catherine Plaisant, Bongshin Lee, and Benjamin B. Bederson. Netlens: Iterative exploration of content-actor network data. Technical report, October 2006.
- [25] Bongshin Lee, C. S. Parr, C. Plaisant, B. B. Bederson, V. D. Veksler, W. D. Gray, and C. Kotfila. Treeplus: Interactive exploration of networks with enhanced tree layouts. *Transactions on Visualization and Computer Graphics*, Vol. 12, No. 6, pp. 1414–1426, 2006.

- [26] Stanley Milgram. The small world problem. *Psychology Today*, Vol. 2, pp. 60–67, 1967.
- [27] Kazuo Misue and Kozo Sugiyama. On automatic drawing of compound graphs for computer aided diagrammatical thinking. *Transactions of Information Processing Society of Japan*, Vol. 30, No. 10, pp. 1324–1334, 1989.
- [28] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Vol. 69, No. 6, 2004.
- [29] Hiroki Omote and Kozo Sugiyama. Method for drawing intersecting clustered graphs and its application to web ontology language. In *Asia-Pacific Symposium on Information Visualization 2006 (APVIS2006)*, pp. 85–92, 2006.
- [30] Hiroki Omote and Kozo Sugiyama. Development of drawing method for kj-diagram. In *The 4th Conference on the Support Systems for Knowledge Creation, 2007*, pp. 76–83, 2007.
- [31] Adam Perer and 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. Student Member-Adam Perer and Senior Member-Ben Shneiderman.
- [32] Adam Perer and Ben Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 265–274, New York, NY, USA, 2008. ACM.
- [33] Adam Perer and Ben Shneiderman. Systematic yet flexible discovery: guiding domain experts through exploratory data analysis. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pp. 109–118, New York, NY, USA, 2008. ACM.
- [34] John Scott. *Social Network Analysis: A Handbook*. Sage Pubns Ltd, 2000.
- [35] Katherine Faust Stanley Wasserman. *Social Network Analysis*. Cambridge University Press, 1994.
- [36] Kozo Sugiyama. *Graph Drawing and Applications for Software and Knowledge Engineers*. World Scientific Publishing Co.Pte.Ltd, 2004.
- [37] Kozo Sugiyama and Kazuo Misue. Visualization of structural information: Automatic drawing of compound digraphs. *IEEE Transaction on systems, man, and cybernetics*, Vol. 21, No. 4, pp. 876–892, 1991.
- [38] Stanley Wasserman and Philippa Pattison. Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, Vol. 61, No. 3, pp. 401–425, 1996.

- [39] Martin Wattenberg. Visual exploration of multivariate graphs. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 811–819, New York, NY, USA, 2006. ACM.
- [40] 松尾豊, 安田雪. SNS における関係形成原理–mixi のデータ分析–. 人工知能学会論文誌, Vol. 22, No. 5, 2007.