

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

クラスタ情報の可視化による  
大規模2部グラフの可読性向上手法

佐藤 修治

(コンピュータサイエンス専攻)

指導教員 田中 二郎

2008年3月

## 概要

ネットワークの情報が容易に取得できるようになった近年、大規模ネットワークの描画手法の開発への要求は高まっている。従来手法ではネットワークにおける要素の数を減らすことで対応することが多かったが、この手法では曖昧な情報のみしか読み取ることができない。また、読み手により適切なグラフの規模は異なるということに対応することができない。

本研究では、2つの新しい描画手法を用いて2部グラフを対象とした大規模ネットワークの可読性向上を狙う。可読性向上には、「見易さ」と「読取りやすさ」の向上が不可欠である。

見易さを向上させる手法として「縮退描画手法」を開発した。クラスタリングによりノード間の類似度を求めておき、閾値を超えた類似度のノード集合を1つのクラスタとして描画する可視化手法である。閾値は読み手がスライダーにより変更でき、ネットワーク図の拡張を自由に行うことが可能である。

読取りやすさを向上させる手法として「等類似度線描画手法」を開発した。縮退描画手法同様クラスタリングの情報を扱い、閾値を超えた類似度のノード集合を閉曲線で囲み描画する可視化手法である。ノード間の類似度の高さを等高線のように表示することで関連性を直接可視化することにより、より理解しやすくさせるのが狙いである。

縮退描画手法と等類似度線描画手法の有効性の検証と評価をするため被験者実験を行った。結果として、両手法ともグラフの可読性を向上させることに成功しており、特にグラフからの情報の読取りやすさが向上することが確認された。

本描画手法は、読み手が求める適切な情報を表示することで、大規模ネットワークの可読性向上を達成している。従来手法では困難であった、大規模ネットワークの全体的な俯瞰と特定の注視要素の関係情報が一度に読取ることが可能になった。

# 目次

|                              |           |
|------------------------------|-----------|
| <b>第1章 序論</b>                | <b>1</b>  |
| 1.1 ネットワーク                   | 1         |
| 1.2 ネットワークの可視化               | 1         |
| 1.3 可視化の目的                   | 2         |
| 1.4 可視化方法の重要性                | 3         |
| 1.5 ネットワークの肥大化と可視化への影響       | 3         |
| 1.6 目的                       | 4         |
| 1.7 本研究の貢献                   | 4         |
| <b>第2章 ネットワーク描画と可読性</b>      | <b>5</b>  |
| 2.1 グラフ                      | 5         |
| 2.2 2部グラフ                    | 5         |
| 2.3 2部グラフの表現形式               | 6         |
| 2.3.1 数学的形式                  | 6         |
| 2.3.2 行列形式                   | 7         |
| 2.3.3 網図形式                   | 7         |
| 2.4 スプリングモデル                 | 7         |
| 2.5 アンカーマップ                  | 9         |
| 2.6 ネットワーク図における可読性           | 10        |
| <b>第3章 可読性を向上させるためのアプローチ</b> | <b>14</b> |
| 3.1 概要                       | 14        |
| 3.2 クラスタリング                  | 14        |
| 3.2.1 クラスタリング手法の選択           | 15        |
| 3.2.2 ノード間類似度の導出方法           | 15        |
| 3.2.3 クラスターの構築               | 17        |
| 3.2.4 本手法の特徴                 | 20        |
| 3.3 描画システム                   | 21        |
| <b>第4章 縮退描画手法</b>            | <b>24</b> |
| 4.1 概要                       | 24        |
| 4.2 手法                       | 24        |

|              |                        |           |
|--------------|------------------------|-----------|
| 4.2.1        | グラフ構造の変換 . . . . .     | 25        |
| 4.2.2        | 描画 . . . . .           | 26        |
| 4.3          | 実装 . . . . .           | 26        |
| <b>第 5 章</b> | <b>等類似度線描画手法</b>       | <b>30</b> |
| 5.1          | 概要 . . . . .           | 30        |
| 5.2          | 手法 . . . . .           | 31        |
| 5.2.1        | グラフ構造の変換 . . . . .     | 31        |
| 5.2.2        | 描画 . . . . .           | 32        |
| 5.3          | 実装 . . . . .           | 34        |
| <b>第 6 章</b> | <b>評価実験</b>            | <b>35</b> |
| 6.1          | 実験目的 . . . . .         | 35        |
| 6.2          | 実験方法 . . . . .         | 35        |
| 6.3          | 実験結果 . . . . .         | 36        |
| <b>第 7 章</b> | <b>適用例と考察</b>          | <b>40</b> |
| 7.1          | 購買データの可視化 . . . . .    | 40        |
| 7.1.1        | 小規模購買データ . . . . .     | 40        |
| 7.1.2        | 大規模購買データ . . . . .     | 42        |
| 7.2          | 論文共著者情報の可視化 . . . . .  | 46        |
| 7.3          | 考察 . . . . .           | 49        |
| 7.3.1        | 縮退描画手法の有効性 . . . . .   | 49        |
| 7.3.2        | 等類似度線描画の有効性 . . . . .  | 49        |
| 7.3.3        | グラフ読解の支援 . . . . .     | 49        |
| 7.3.4        | 本手法の問題点 . . . . .      | 49        |
| 7.4          | 今後の課題 . . . . .        | 50        |
| 7.4.1        | ラベリング . . . . .        | 50        |
| 7.4.2        | より大きなグラフへの対応 . . . . . | 50        |
| <b>第 8 章</b> | <b>関連研究</b>            | <b>51</b> |
| 8.1          | グラフの自動描画 . . . . .     | 51        |
| 8.2          | 知識の抽出 . . . . .        | 51        |
| 8.3          | 大規模グラフへの対応 . . . . .   | 51        |
| 8.4          | クラスタ構造の活用 . . . . .    | 52        |
| 8.5          | 可読性向上の追求 . . . . .     | 52        |
| <b>第 9 章</b> | <b>結論</b>              | <b>55</b> |
|              | <b>謝辞</b>              | <b>56</b> |



|      |    |
|------|----|
| 参考文献 | 57 |
| 後注   | 62 |

# 図目次

|     |                                    |    |
|-----|------------------------------------|----|
| 1.1 | ネットワークの例 . . . . .                 | 2  |
| 1.2 | グラフを図で表す例 . . . . .                | 2  |
| 2.1 | 2部グラフの表現形式 . . . . .               | 8  |
| 2.2 | スプリングモデルのイメージ . . . . .            | 8  |
| 2.3 | スプリングモデルによるレイアウト . . . . .         | 9  |
| 2.4 | アンカーマップ . . . . .                  | 10 |
| 3.1 | クラスタリング手法の分類 . . . . .             | 16 |
| 3.2 | デンドログラム . . . . .                  | 16 |
| 3.3 | 最短距離クラスタリング . . . . .              | 17 |
| 3.4 | クラスタリングの過程 . . . . .               | 19 |
| 3.5 | クラスタリング順による変化 . . . . .            | 21 |
| 3.6 | ClusteredAnchorVizの概観 . . . . .    | 22 |
| 3.7 | システム構成 . . . . .                   | 23 |
| 4.1 | スライド変更時の走査 . . . . .               | 26 |
| 4.2 | デンドログラム切断時の表示するノードとクラスタ . . . . .  | 27 |
| 4.3 | アプリケーション初期画面 . . . . .             | 28 |
| 4.4 | クラスタの展開の様子 . . . . .               | 29 |
| 5.1 | 等類似度描画手法を用いた表現 . . . . .           | 31 |
| 5.2 | アノテーションで囲み線を用いる例 . . . . .         | 32 |
| 5.3 | 凸包 . . . . .                       | 33 |
| 5.4 | 閉曲線を描画する方法 . . . . .               | 34 |
| 6.1 | 実験に用いたグラフをアンカーマップにより描画 . . . . .   | 36 |
| 6.2 | 実験に用いたグラフを縮退描画手法で描画した様子 . . . . .  | 38 |
| 6.3 | 実験に用いたグラフを等類似度線で描画した様子 . . . . .   | 39 |
| 7.1 | 購入時間帯-商品の関係のアンカーマップによる描画 . . . . . | 41 |
| 7.2 | 提案手法による購入時間帯-商品の関係の描画 . . . . .    | 42 |
| 7.3 | 購入者-商品のアンカーマップによる描画 . . . . .      | 44 |

|     |                                                     |    |
|-----|-----------------------------------------------------|----|
| 7.4 | 等類似度線描画手法 ( $t = 100\%$ ) による購入者-商品の関係の描画 . . . . . | 44 |
| 7.5 | 縮退描画手法による購入者-商品の関係の描画 . . . . .                     | 45 |
| 7.6 | 2 ホップ . . . . .                                     | 46 |
| 7.7 | アンカーマップによる論文共著者情報の可視化 . . . . .                     | 47 |
| 7.8 | 等類似度線描画手法 ( $t = 0\%$ ) による論文共著者情報の可視化 . . . . .    | 47 |
| 7.9 | 縮退描画手法による論文共著者情報の可視化 . . . . .                      | 48 |
| 8.1 | 本研究の位置付け . . . . .                                  | 54 |

# 第1章 序論

## 1.1 ネットワーク

ネットワークはこの世界の至る所に満ちあふれている。言葉こそインターネットの普及により身近になったばかりという感覚を受けるが、人間関係、鉄道路線、電話回路網など人間の生活には切り離せない場所に偏在している（図 1.1<sup>1)</sup>）。

近年 Web2.0 という言葉が叫ばれるようになり、一般のユーザが他のユーザに影響を与えるようなシステムが増えてきている。ソーシャル・ネットワーキング・サービス（SNS）はその典型であり、リアルな世界以外のヴァーチャルな友人関係が世界の様々な場所で構築され、その人間関係のネットワークは拡大の一途を辿っている。

人間関係のネットワークの研究において、しばしば「スモールワールド」もしくは「六次の隔たり（6-degree）」という単語が現れる。これは、スタンレー・ミルグラムが 1967 年に行ったスモールワールド実験 [1] に依るもので、友人関係のネットワークから 2 人を任意に抽出した場合、最高でも 6 人の人間の仲介で接続されるという仮説である。

ソーシャル・ネットワーキング・サービス mixi の例を示すと、2005 年の湯田ら、森ら、安田らによる分析結果によると、友人関係のネットワークはマイミク（友人）の平均数が 10.4、6 ホップで 96% をカバーでき、スモールワールドの特徴を備えていることがわかっている [2]。

このようなネットワークの場合、分析には友達同士の関係だけではなく、彼らが所属するコミュニティを読み解くことで、理解が深まるという言うことが知られている [3]。人と所属の関係は **2部グラフ（2部ネットワーク）** という形式で表現され、社会的なネットワークの特徴を捉える上で重要な役割を担う表現であると考えられている [4]。

## 1.2 ネットワークの可視化

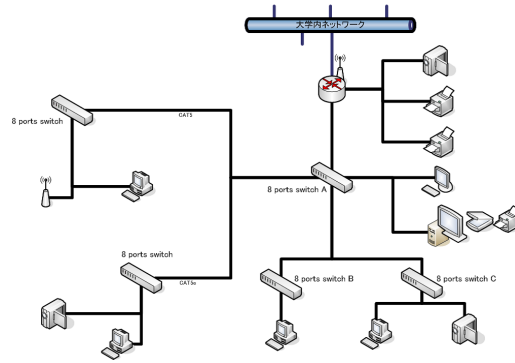
普段のネットワークを意識して生活している人々はそれほど多くはない。なぜなら、ネットワークは目に見えない抽象的なオブジェクトであるためである。

ネットワークは物（オブジェクト）とそのつながりにより構成される構造体の内、フロー（流れ）が発生するものを指す。この構造体はグラフとして数学的に表現することが可能である。言い換えると、何らかの意味を持ったグラフがネットワークである。

数学的表現のグラフは、人間が直感的に要素間のつながりを認知するには困難な表現手法である。そこで、我々は日常的グラフを認識する方法として用いているのが、「図」というイ



(a) 鉄道路線図



(b) LAN 回線図

図 1.1: ネットワークの例

$$\begin{aligned}
 G &= (V, E) \\
 V &= \{A, B, C, D, E, F\} \\
 E &= \{(A, B), (B, C), (B, D), (C, E), \\
 &\quad (D, E), (D, F), (E, F)\}
 \end{aligned}$$

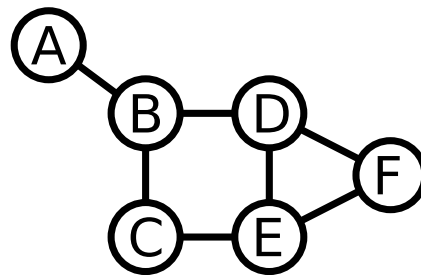


図 1.2: グラフを図で表す例

ンタフェースである．図 1.2 にグラフの数学的表記から図への変換例を示す．このように，抽象的な「データ」を，直接見ることが出来る表現へ変換することを，**可視化**（Visualization）と言う．

可視化の用途は，構造の認識や他人との意思疎通，データマイニング，問題発見などがあり，様々な分野で研究が盛んに行われている．

### 1.3 可視化の目的

人間は外界の情報を視覚，聴覚，触覚，味覚，嗅覚の五感を使い知覚している．この中でも視覚は全情報の 80 % もの情報を得ていると言われており [5]，視覚情報をわかりやすくすることがいかに理解につながるかが想像できる．

本研究で扱う情報可視化（Information Visualization）は科学的なデータを視覚化する科学的可視化（Scientific Visualization）の流れを汲むものである [6]．情報可視化の目的は知識の取得と発見である．多量のデータをコンピュータにより可視化により支援し，人間だけでは得

られない情報と知識の取得，発見を促進する．すなわち，グラフをコンピュータで処理をして結果を見せるのではなく，人がそれを情報として利用するための仕組みが目的の技術である．

そのため，結果を読み手に提示させるだけの一方向作用だけではなく，読み手が入力した操作を反映させるインタラクションシステムの環境が適していると言われている [7]．

## 1.4 可視化方法の重要性

ネットワークの可視化における表現手法には様々有るが，図は大きく分けて4つに分類することが出来る．

- 座標系: 折れ線グラフ，棒グラフなど
- 行列系: 表など
- 領域系: ベン図など
- 網図（連結）系: ネットワーク図，フローチャートなど

これらの表現手法の内，座標系や行列系は従来から研究が進んでおり，比較的表現方法がはっきりしている．それに対し領域系，網図系，およびその複合系は，ネットワーク構造の視覚的記法として重要で有るにもかかわらず，描画方法があまり発達していない．

領域系や網図系の図は日常的に利用され，それにより理解が容易になることは経験的に認識されている．ただし，その図の認識のしやすさ，されやすさは，図の書き方に強く依存する．図の美しさによりその図の価値が決定され，理解の支援にも混乱を引き起こす原因にも成り得るのである [8]．そのため，構造がわかりやすい等の美しい図を描くことは，可視化にとって最も重要な点の一つであり既存研究の大きな課題となっている．

## 1.5 ネットワークの肥大化と可視化への影響

情報技術の発展により，身の回りにあるネットワークが肥大化の一途を辿っていることは衆知の事実である．

先に述べた mixi は，2005 年時点でノード数 36 万個，エッジ数 190 万リンクであった．2008 年 1 月現在では既に 1000 万ノードを超えており，顕著な規模増加が起きている．また，mixi の友人関係は擬集性の高いネットワークであることが判明している．つまり，これを 2 次元平面上に可視化した場合，狭い空間にノードがひしめき合い，エッジの交差が多数存在する塊が随所に現れるような表現で描画されることが予想される．

ノード数とエッジ数の増加は，読み手によるネットワークの認識を難しくさせる [9]．その理由の一つに可視化結果の可読性の低下がある．表示するデバイスの解像度を上げることできめ細かい描画を行い，すべてのノードとエッジを表示することは可能であるが，人間の理

解度の向上には繋がりにくい。人間の空間認識能力がはある一定以上上がらないためである。つまり、読み手がネットワークを認識・理解するのに適当な大きさの領域に、簡潔に描画することが重要となる。

グラフ描画において可読性の問題は重要なトピックスの一つであり、可読性の向上に関する研究は盛んに行われている。

## 1.6 目的

本研究では、2部グラフで構成される大規模ネットワークの可視化における可読性を向上させることを目的とする。2部グラフという特殊な構造を持ったネットワークを、その特徴を利用した可視化を行うことで読み手の理解を高める。従来の手法は読むことが難しかった構造を読み手に把握し易くし、新たな知識の取得を支援する。

## 1.7 本研究の貢献

従来手法では注目されていなかったグラフの意味を読み取り易くするという観点を取り入れ、大規模2部グラフの可読性を向上させる手法を開発した。この手法を用いることで、グラフ描画のエキスパートだけではなく普段グラフを読まない人にも、効率よく知識を得られるような描画が可能となる。

## 第2章 ネットワーク描画と可読性

本章では、ネットワークの可視化およびその可読性についての考察を行う。まず、ネットワークの構造であるグラフの説明を行い、本研究で扱う2部グラフについて構造と描画手法の説明を行う。最後に、グラフを可視化したときの可読性とは何かについてを考察する。

### 2.1 グラフ

グラフとは、**ノード**（節点・頂点）の集合とノード間を接続する**エッジ**（辺・リンク）の集合により構成される数学的なデータ表現である。ノード集合を  $V$ 、エッジ集合を  $E$  としたとき以下の式 2.1 として表記される。

$$\begin{aligned} G &= (V, E) \\ E &\subseteq V \times V \end{aligned} \tag{2.1}$$

グラフはネットワークの特性により分類することが可能である。エッジには有向エッジと無向エッジの2種類があり、それを有するグラフのことはそれぞれ有向グラフと無向グラフと呼ばれる。一般に木構造と言われる構造も、閉路を持たず単連結である無向グラフのことを指す。グラフは抽象化された概念の1つの有力な表現形式であるので、様々な分野で基礎的なモデルとして広く利用されている [8]。

様々なグラフ構造のうち、本研究では2部グラフとして表現されるネットワークに焦点をあてる。

### 2.2 2部グラフ

ノードの集合を二つの排他的な集合  $V_1$  と  $V_2$  に分割することができ、エッジの集合  $E$  が  $V_1 \times V_2$  の部分集合であるようなグラフを2部グラフという。数学的には以下の式として表現される（式 (2.2)）。

$$\begin{aligned} G &= (V_1 \cup V_2, E) \\ V_1 \cap V_2 &= \phi \\ E &\subseteq V_1 \times V_2 \end{aligned} \tag{2.2}$$

2部グラフをもつ構造のネットワーク（二部ネットワーク）は実世界の様々な場面で現れている。一般的に所属関係のネットワークは必ず2部グラフの構造を持っている。これは、ソーシャルネットワークにおける「コミュニティ」と「メンバー」の関係であったり、芸能界に



における「出演作品」と「出演者」の関係、購買ネットワークにおける「顧客」と「購買商品」の関係などが例に挙げられる。amazon.comで買い物をする際に、「この商品をチェックした人はこんな商品もチェックしています」という項目があるが、これも「商品」と「購入者」の2部グラフの関係が形成されているものであり、日常的に利用される二部ネットワークの一例である。

研究分野でも2部グラフは注目されている。特に社会ネットワークにおいてその注目度は高い。ソーシャルネットワークに関しては友人関係だけを見るのではなく、彼らの属するコミュニティの関係も読み解くことで、理解が深まるのではないかという仮説があることは先に述べた通りであり、それに関する研究も多数行われている[10]。ウェブログについてもブロガーとコミュニティの2部ネットワークを解析することで、ネットワークが発展するとの報告がある[4]。論文の共著者関係のネットワークにおいても、その著者を結ぶ論文の存在との関係を見ることで、ネットワークの特徴の新たな発見を期待できる[11, 12, 13]。

## 2.3 2部グラフの表現形式

2部グラフは様々な形式で記述することが可能である。グラフは基本的に数学的な構造体であるが、先に述べた通り人間にとって数式表現のまま構造をとらえるのは難しい。そのため、人間に読みやすい形に視覚化を行う。本節では2部グラフの可視化表現のうち代表的な表現形式を述べる。商品と購買者から成る購買関係の2部グラフを例に用い、実際に図で表しながら説明を行う。

### 2.3.1 数学的形式

数学的表記はグラフ構造の表現形式としては基本である。2部グラフにおいてもそれは変わらず、グラフのすべての情報を表現出来るため、情報が欠落してはいけない伝達をする時やグラフを定義する時に用いられる。

$$G=(V_1 \cup V_2, E)$$

$$V_1=\{\text{石垣, 岩波, 斉藤, 坂崎, 鈴木, 中里}\}$$

$$V_2=\{\text{珈琲, 紅茶, 麦茶, 烏龍茶, ビール, コーラ, 栄養ドリンク}\}$$

$$E=\{(\text{石垣, 珈琲}), (\text{石垣, ビール}), (\text{石垣, コーラ}), (\text{岩波, 紅茶}),$$
$$(\text{斉藤, 珈琲}), (\text{斉藤, 紅茶}), (\text{斉藤, 烏龍茶}), (\text{斉藤, 栄養ドリンク}),$$
$$(\text{坂崎, 紅茶}), (\text{坂崎, 麦茶}), (\text{坂崎, コーラ}), (\text{鈴木, 紅茶}), (\text{鈴木, 烏龍茶}),$$
$$(\text{中里, 珈琲}), (\text{中里, 紅茶}), (\text{中里, 麦茶}),$$
$$(\text{中里, 烏龍茶}), (\text{中里, コーラ}), (\text{中里, 栄養ドリンク})\}$$

上記がグラフを数学的形式で表現した例である．この手法は無論グラフを読み解く際にはあまり向かない．

### 2.3.2 行列形式

グラフ構造は行列形式で記述することが可能である．エッジは2つの集合間にのみ存在するという2部グラフの特性から通常のグラフよりも簡潔に表現されるため，よく用いられる手法である．それぞれのノード集合を行と列に分解して表記する図2.1(a)の例では行に購入者，列に商品に対応させ，購入者が購入した商品の欄に数値を入れることで表現している．

行列表現の特徴としては，一つのノードを注目したときの，エッジの量がわかりやすいことから他のノードの関係を認識するのに良好であることがあげられる．

また，行列表現では情報を落とさない表現が可能であることも特徴の一つである．例えば，購買情報においては「購入者がどれだけその商品を買ったか」といった情報が存在する．グラフ表現において，この情報はエッジの「重み (weight)」として表現される．行列表現においてはこのエッジの重みを数値として明確に表現することが出来るため，正確な詳しい情報を知り得たいような状況では，特に有効な手法である．

但し，視覚的な工夫がなされている訳ではないため，購入者同士の関係構造などの把握にはあまり適さない．また，描画領域の空間効率化は他の手法と比較してあまり良くないため，大規模なグラフを表現するときには広い描画領域が必要となる．

### 2.3.3 網図形式

ネットワークの可視化にはよく網図形（連結系）の表現が用いられる．そのうち，2部グラフの表記に特化した形式では2層形式がある．二層形式はノードの2つの集合をそれぞれ1列で表現する形式である．ノードを点や丸，長方形などで表し，ノード間を直線や曲線で接続することでエッジを表し，グラフの関係を表現する．

図2.1(b)は，左側に購入者を，右側に購入商品を配置した二層形式の図である．一列に集合が整列されているおり，それぞれの集合に属するノードが理解しやすいことが特徴であるため，2部グラフの概要を説明する際に用いられることが多い．他の特徴としては，それぞれの集合に属するノードの数の把握がしやすいこと，エッジ数の多いノードが把握しやすいなどが挙げられる．

この手法の場合エッジの交差の問題の解決が難しく，完全2部グラフに近くなるほどどのノードとどのノードが接続しているかがわかりにくくなってしまう．また，エッジの重みを表現することが難しいため，グラフの解析を目的する時の可視化手法としてはあまり適さない．

## 2.4 スプリングモデル

グラフの可視化に用いられる網図系の代表的な描画法の1つとして，**スプリングモデル**（ばね埋込モデル，Spring Embedded Model）[14]がある．エッジを自然長をもつバネと仮定し，

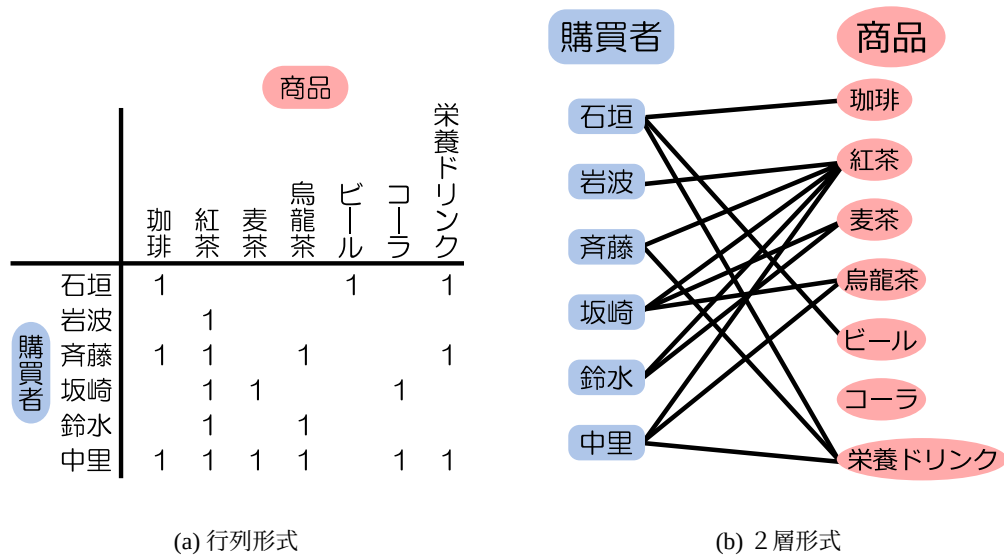


図 2.1: 2部グラフの表現形式

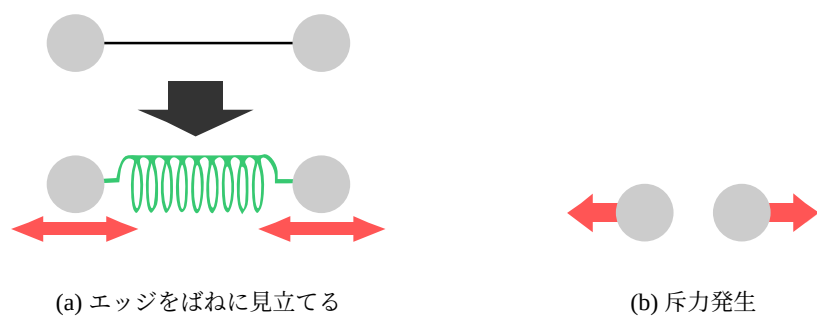


図 2.2: スプリングモデルのイメージ

エッジにより接続されないノード同士は斥力により反発力を発生させる（図 2.2）。バネと斥力によりノードを配置し、安定状態を計算することで最終的にノードのレイアウトを求める手法である。

ノードのレイアウトに関しては以下の流れで行う。

1. 初期配置を決定する（ランダム配置など）
2. 各々のノードに働く力の合計を求める
3. 求めた力に従い、ノードのレイアウトを変更する
4. 2に戻る。

スプリングモデルは2部グラフ専用の表現形式ではないが、二層形式と異なり重みのある

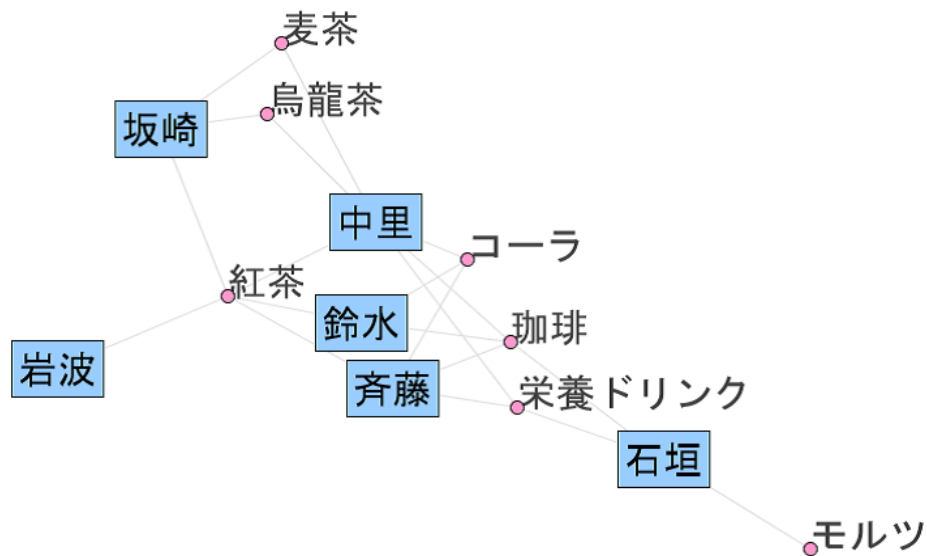


図 2.3: スプリングモデルによるレイアウト

エッジを表現することが可能であるため、ノードの関係をより良く知ることが出来る。そのため、グラフの自動描画の基本レイアウトとして多くの研究で用いられている。

図 2.3 は、スプリングモデルを適応した例である。長方形で表示したノードが購買者を、円で表示したノードが商品を表しており、エッジは直線の接続で表現されている。エッジに埋め込まれたばねの力で接続情報が似ているノードが近くにレイアウトされ、ノード同士の関連性が認知しやすくなっている。

## 2.5 アンカーマップ

**アンカーマップ** [15] は、スプリングモデルを発展させて 2 部グラフに特化させた描画手法である。2 部グラフの 2 種類のノードの集合のうち一方のノード集合の位置に制約を課した描画スタイルと定義される。

前節のスプリングモデルでは全体的な構造は把握しやすいが、2 つの集合（例では購買者と商品）が混在しているため、ネットワークの構造がわかりやすいとはいえない。二層形式ではノード集合同士は分割されているので把握はしやすいが、それぞれのノード集合内部同士の関連性の把握は難しい。

ネットワーク構造を効率よく把握するには観点の導入が有効であり、それによりネットワークの構造的特徴を認知するのが容易となる。アンカーマップは、片方のノード集合を固定し座標軸のような効果を与え、観点導入の役割を果たしている。2 種類のノードは位置に制約を課す方をアンカーと呼び、もう片方をフリーノードと呼ぶ。アンカーとフリーノードは交換可能で、どちらをアンカーにするかは自由に変更することが出来る。アンカーは円周上に等間隔配置（正多角形の頂点に配置）され、フリーノードはアンカーとの関係を表現する適

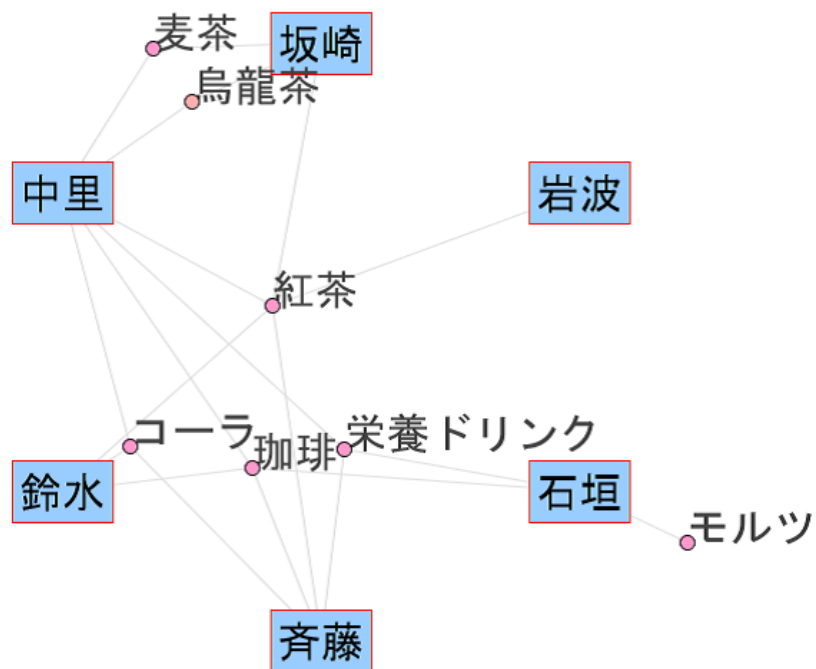


図 2.4: アンカーマップ

切な位置にスプリングモデルでレイアウトされる。エッジはアンカーとフリーノードを直線で接続する形式で表現される。

アンカーの配置（順番）はいくつかの美的基準を満たすように行われる。

- エッジの総線長を最も短くする
- エッジの交差数を最も少なくする
- 同じフリーノードに接続するアンカーを近くに配置する

図 2.4 は、購入者をアンカー、商品をフリーノードとしてアンカーマップで描画した図である。「紅茶はグラフの中心に配置されているため多くの購買者に買われている」といった、スプリングモデルからは発見が難しい情報を素早く認知することが出来る。

現在、数多の 2 部グラフの表現手法の内、アンカーマップは 2 部グラフを最も効率的に表現できる手法の一つと考えられる。

## 2.6 ネットワーク図における可読性

ネットワークを可視化した図は、良い描画をすることは重要であることは第 1 章で述べた。その描画の善し悪しはグラフの可読性に影響する。

先に述べたアンカーマップにおいても、アンカーの配置順により可読性の向上を狙っている [16]. しかし、ノードの数が増大していくと既存手法のみで対応することは難しい. 出力する画面というものは大きさが限られており、画面を埋め尽くすほどのノードを描画しなければならないときには、ノードのレイアウトを変更しても読み手は見易いと感じないと予想できる. また、必要以上に大きい画面や異常なまでに高精細な解像度で出力したところで、これもまたグラフの意味を掴むのは用意ではないだろう.

では、可読性とは如何なるものなのか.

従来のグラフィックレイアウト手法において可読性を高めるための指標として設定されているのが**美的基準** (Aesthetic Criteria) である. 人間の認知特性に関する様々な基準が定義されている.

- ノードの重なりが少ない
- エッジの重なりが少ない
- 隣接関係を持つノードが近くに配置される
- エッジの交差数が少ない
- エッジの総線長が短い
- 描画領域が小さい
- エッジの屈折点が少ない
- ノードが一様に分布されている
- 構造の階層性を表現する
- 構造の対称性を表現する

これらを満たすレイアウトを生成することが望ましいと考えられている. むろんこれらは排他的な性質を持つわけではなく、競合的な性質を持つ組み合わせや、逆に相乗効果をもたらす組み合わせも存在する.

私は可読性にはいくつかの側面があり、その中でも以下の二つが重要であると考えた.

- **空間構造的な可読性**
- **意味的な可読性**

「**空間構造的な可読性**」は、可視化したグラフの「**見易さ**」のことを指す. 見易さとは、人間が視覚的に見易いと感じることの指標であり、見た目に「すっきりしている」という抽象的な感覚のことをここでは定義する. 具体的にグラフ描画においては「構成要素 (ノード・エッジ) が少ない」であったり、「エッジがあまり交差していない」、「ノードが大きい」などのことで、空間構造的な可読性が向上すると考える. 既存手法のアンカーマップにおけるアン

表 2.1: 可読性向上手法とその効果

|          | (情報を保ったままの手法)<br>レイアウト変更など | (情報を削減する手法)<br>クラスタリングなど |
|----------|----------------------------|--------------------------|
| 可読性 (全体) | ○                          | ○                        |
| 構造的な可読性  | ○                          | ◎                        |
| 意味的な可読性  | ○                          | △ ↔ ○                    |

カーのレイアウトの美的基準に「エッジの交差数を最小にする」というものがあるが、これはこの可読性を直接的に向上させるものである。

この指標はグラフの自動描画の既存手法で、可読性において一番重要視されてきた点であると考えられる。なぜなら、「見易い」「見難い」はグラフのみならず、広範囲の分野で人間の感覚に直接効果を及ぼすものであるからである。全く同じ意味を持つ2つの対象があり、見易いものと見づらいものがあれば、見易いものが良いことは自明の理であり可視化の目的の1つはまさにそれである。

大規模グラフの空間構造的な可読性における先行研究のアプローチとしては、高精細描画とクラスタリングが主なものとしてあげられる。高精細描画は解像度の高い出力を行い、広大な出力空間を有効に用いてノードを配置することで可読性を向上させようという試みである。解像度を高くするため、グラフの規模が増大すれば一つ一つの要素は拡大しなければ目視できなくなる可能性もある手法である。クラスタリングとは集合分類手法のひとつで、3.2にて詳しく述べる。類似した情報を持つノードを1つのノードとして表示する手法をとることで、ノードとエッジの数を減少させてグラフを「すっきり」と描画する。空間構造的な可読性を向上させるに当たっては、単純であるが極めて効果的な方法であると考えられる。

「**意味的な可読性**」は、可視化したグラフの「**読み取り易さ**」のことを指す。読み取り易さとは、読み手がどの程度グラフの意味を掴みやすいと感じるかということである。こちらは、「見易さ」と違い直接視覚に訴えてくることはない。しかし、人間が可視化したグラフを見る目的はグラフの内容を知ることである。すなわち、内容を取りやすい可視化結果は可読性の高い可視化結果であると言い換えることも可能である。

先に述べたクラスタリングによるノードの数を減らす手法では、描画したグラフの構造とクラスタリング前の元のグラフの構造とが異なってしまう。すなわち、可読性を確保するため情報量を犠牲にしているのである。そのため、グラフの構造を正しく理解することが難しくなる危険性がある。

先行研究において、意味的な可読性はあまり重要視されていない。無論、この可読性を向上させる手法もあるが、多くが空間構造的な可読性を向上した結果の間接作用であり、直接的に効果を及ぼす既存手法は少ない。グラフの自動描画の可読性においては、見易さは一番重要なファクターだと考えられる。しかし、意味がとれないように見易くしても本末転倒である。

いかなる表現にも共通することだが、「**読み手**」により**解釈が変化する**という点も重要で

ある．いかに作り手が一番わかりやすい表現を作成したところで，全世界の人間が一番わかりやすいとは考えない．人間はそれぞれが異なった感覚を持ち認識能力も各々異なるためである．画面上に配置されているノードの数がどの程度で適切であると感じるかは人によって異なる．加えて，その時の目的により欲しい描画は異なる [17]．すなわち，可読性の向上を考える際には「読み手」が読みやすい可視化を考える必要性は常に付きまとう．

私は，「空間構造的な可読性」「意味的な可読性」という 2 つの可読性の向上が大規模グラフの可読性向上のキーであると考えた．加えて「読み手」各々に最適な出力結果が得られるような表現手法を取り入れることで，更なる可読性の向上が期待できる．次章よりその手法について説明を述べる．



## 第3章 可読性を向上させるためのアプローチ

本章では大規模なネットワークの可読性を向上させる手法を述べる。第2.6節で述べた2つの可読性、「空間構造的な可読性」「意味的な可読性」を向上させることを目的とし、そのアプローチとして、グラフの構造をコンピュータで解釈し、その情報が読み手によって効果的に反映させることが可能な表現手法で対応する。

### 3.1 概要

私は大規模な2部グラフの可読性を向上させる描画手法として、「縮退描画手法」と「等類似度描画手法」を考案した。

可読性向上のためには、第2.6節で述べた「空間構造的な可読性」と「意味的な可読性」の向上が必要であると考ええる。加えて、読み手の意図に対応した描画状態、つまり読み手の操作に対応して構造を変更・描画に反映させる手法が可読性向上に繋がると考える。

本描画手法を開発するにあたってのコンセプトは以下の二つである。

- コンピュータで構造を解析し、それを可視化結果に反映させること
- 読み手の操作に対応して可視化結果を変更するインタラクションシステムであること

まず、グラフの構造を捉えるためにクラスタリングを用いた。描画にはアンカーマップの拡張表現を用い、読み手に最適な描画状態を与える手法を加えた。

本章では、データの解析から描画までの仕組みを述べ、描画手法については次章以降で詳しく述べる。

### 3.2 クラスタリング

クラスタリングとは、異なる性質のもの同士が混在している集合の中から互いに類似したものを集めてクラスを形成することで、集合を分類する手法である [18]。

一般的にグラフ描画におけるクラスタリングは、論理的に同じ接続構造を持つノードを一つのノードとして描画する手法として扱われることが多い。しかし、この手法はグラフが大規模になると効果が低くなる。グラフ全体に対して同じ接続構造をもつノードの割合は減少すると考えられるためである。無論、対象とするグラフに依り効果に違いはあるが、すべてのグラフに対応することはできないことがわかる。

上記の手法を拡張し、階層型クラスタリングを用いて多階層のクラスタ構造を形成する手法がある。この手法ではスケーラビリティが高く、どんな大きいグラフでも小さくすることが出来るが、

- 小さくしたことにより情報が欠落する
- クラスタの集約・展開方法が洗練されていない

といった問題が残されている。

本研究で扱うクラスタリング手法は、ノード間のエッジの接続関係から類似度を計算し、階層構造のクラスタを生成するものである。ただし、クラスタが2部グラフの2種類の集合を跨いで形成されることはないようにして、2部グラフの構造を保持するよう注意する。

### 3.2.1 クラスタリング手法の選択

クラスタリングの手法は、まず大きく「階層型クラスタリング」と「分割型（区分型）クラスタリング」の二つに分けることが出来る（図3.1）。

分割型（区分型）クラスタリングとは、要素全体をいくつかのグループに分割する手法であり、自己組織化マップの生成などに利用される。代表的な手法に k-means 法などがあり、あらかじめ、分割するクラスタの数が分かっている場合には有効な手法である。

階層型クラスタリングとは階層構造を成すようにグループを生成するものである。入れ子を構成するようにクラスタを生成すると考えても良い。

階層型クラスタリングは「分岐型（分割型）」「凝集型（併合型）」に二分される。分岐型は、要素集合を1つの大きなクラスタとして考え、クラスタを分割しながら、階層構造を形成していく物である。逆に凝集型は1つの要素を1つのクラスタとする初期状態を与え、最も距離が近い二つのクラスタを逐次的に結合していき、すべての要素が1つのクラスタになるまで階層構造を形成する。両者ともクラスタリングの結果はデンドログラムで表現することが出来る（図3.2）。

一般的に階層型クラスタリングは集約型の方が多く用いられ、方法論的にも整理されている [19]。集約型の代表的な手法としてはウォード法などがある。

グラフ構造の解析と描画をする上で、スケーラビリティが高いのは階層型クラスタリングである。また、クラスタリングの情報を1つの一貫した軸で扱いたいため、集約型の最短距離法を用いて開発を行った。詳しい特徴については後述の3.2.4節で述べる。

### 3.2.2 ノード間類似度の導出方法

クラスタリングを行う際には、要素同士の関連性を表す何らかの指標が必要となる。一般にその指標は要素間の「距離」や「類似度」、「非類似度」などで定義される。グラフにおいてはノード同士の関連度からクラスタリングを行う。本論文ではこの関連度のことを、ノード間の「**類似度**」という言葉で定義する。本手法では、ノード間の類似度の指標としてジャックカード係数を採用した。

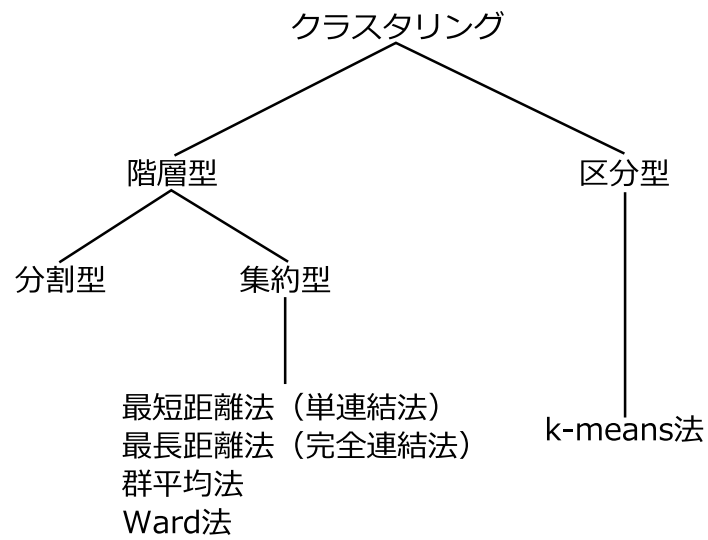


図 3.1: クラスタリング手法の分類

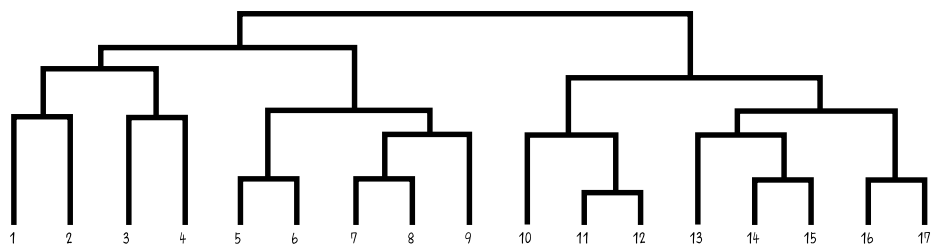


図 3.2: デンドログラム

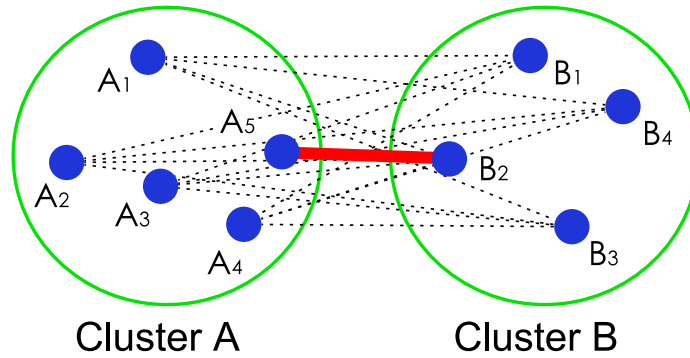


図 3.3: 最短距離クラスタリング

### ジャックカード係数（距離）

ジャックカード係数を用いて定義したノード間類似度  $S$  は、以下のように求められる。ノード  $x$  が接続するノードの集合を  $A(x)$  としたとき 2 つのノード  $x, y (x, y \in V)$  の類似度  $S(x, y)$  は式 (3.1) で表される。

$$S(x, y) = \frac{|A(x) \cap A(y)|}{|A(x) \cup A(y)|} \quad (3.1)$$

例として、 $A(x) = \{\alpha, \beta, \gamma\}$ 、 $A(y) = \{\alpha, \beta\}$  としたとき、 $S(x, y)$  は以下のように算出される。

$$\begin{aligned} S(x, y) &= \frac{|\{\alpha, \beta\}|}{|\{\alpha, \beta, \gamma\}|} \\ &= \frac{2}{3} \approx 0.66 \end{aligned}$$

### 3.2.3 クラスタの構築

最短距離クラスタリングとは、クラスタとクラスタの類似度（距離）を内包する要素でそれぞれ一番小さいものをクラスタの類似度とする考え方である。

図 3.3 の例に従って、最短距離クラスタリングを説明する。小さい青い円はノードを表しており、青い円を大きく囲んだ緑色の閉曲線はクラスタを表す。青い円の距離は類似度に比例しているとする。ここで、Cluster A と Cluster B の類似度を求めるとしたとき、Cluster A に含まれる要素と Cluster B に含まれる要素のペアのうち、一番近いペア（類似度が高いペア）の類似度を、クラスタ間の類似度にする。すなわち、図中の  $A_5$  と  $B_2$  の類似度が、Cluster A と Cluster B の間の類似度になる。

これを数学的に定義し、構築の方法を以下に述べる。

クラスタ  $C$  はノードとクラスタの集合であり、以下の式で定義する。

$$\begin{aligned} C &= \{c_1, \dots, c_n\} \quad (n \geq 2) \\ c_1, \dots, c_n &\in V \cup C \end{aligned}$$

クラスタにおける類似度  $S(C)$  を以下のように定義する.

$$S(C) = \max_{1 \leq i < j \leq n} S(c_i, c_j) \quad (c_i, c_j \in V \cup C)$$

式 (3.1) に加え, 類似度  $S(p, q)$  ( $p, q \in V \cup C$ ) を以下のように定義する.

$$S(p, q) = S(q, p) \quad \text{可換性}$$

$$S(p, q) = \begin{cases} \frac{|A(p) \cap A(q)|}{|A(p) \cup A(q)|} & (p, q \in V) \\ \max_{q' \in q} S(p, q') & (p \in V, q \in C) \\ \max_{p' \in p, q' \in q} S(p', q') & (p, q \in C) \end{cases}$$

類似度が上記の定義になるようなクラスタを構築する手順を以下に示す.

1.  $s = \max_{\forall x, \forall y \in V_2, x \neq y} S(x, y)$  となる  $s$  を求める.
2.  $S(x, y) = s$  となるノード  $x, y$  が存在する場合,  $x \in C_x, y \in C_y$  を満たすクラスタ  $C_x, C_y$  が存在するかを確認する.
  - (a)  $C_x$  と  $C_y$  が両方存在しない場合,  $x$  と  $y$  を含み, 類似度  $S(C) = s$  のクラスタ  $C$  を生成する (図 3.4(a)).
  - (b)  $C_x$  が存在し,  $C_y$  が存在しない場合,  $C = G(C_x)$  とを満たすクラスタ  $C$  を求める (図 3.4(b)).
    - i.  $S(C) = s$  の場合,  $C$  に  $y$  を加える
    - ii.  $S(C) > s$  の場合,  $y$  と  $c$  を含み, 類似度  $S(C') = s$  のクラスタ  $C'$  を生成する.
  - (c)  $C_y$  が存在し,  $C_x$  が存在しない場合, 2b の  $x$  と  $y$  を入れ替えて同様の動作を行う (図 3.4(c)).
  - (d)  $C_x$  と  $C_y$  が両方存在する場合,  $C'_x = G(C_x C_y)$  と  $C'_y = G(C_y)$  を求める (図 3.4(d)).
    - i.  $s = S(C'_x) = S(C'_y)$  の場合,  $C'_x$  と  $C'_y$  を結合する
    - ii.  $s = S(C'_x)$  かつ  $s < S(C'_y)$  の場合,  $C'_x$  に  $C'_y$  を加える.
    - iii.  $s = S(C'_y)$  かつ  $s < S(C'_x)$  の場合,  $C'_y$  に  $C'_x$  を加える.
    - iv. それ以外の場合は,  $C'_y$  と  $C'_x$  を含み, 類似度  $S(C) = s$  のクラスタ  $C$  を生成する.
3.  $s$  であるノードの組み合わせが残っていれば, 2 に戻る
4.  $s > 0$  の場合,  $s$  より小さい類似度の値中次に大きいものを  $s$  として, 2 に戻る

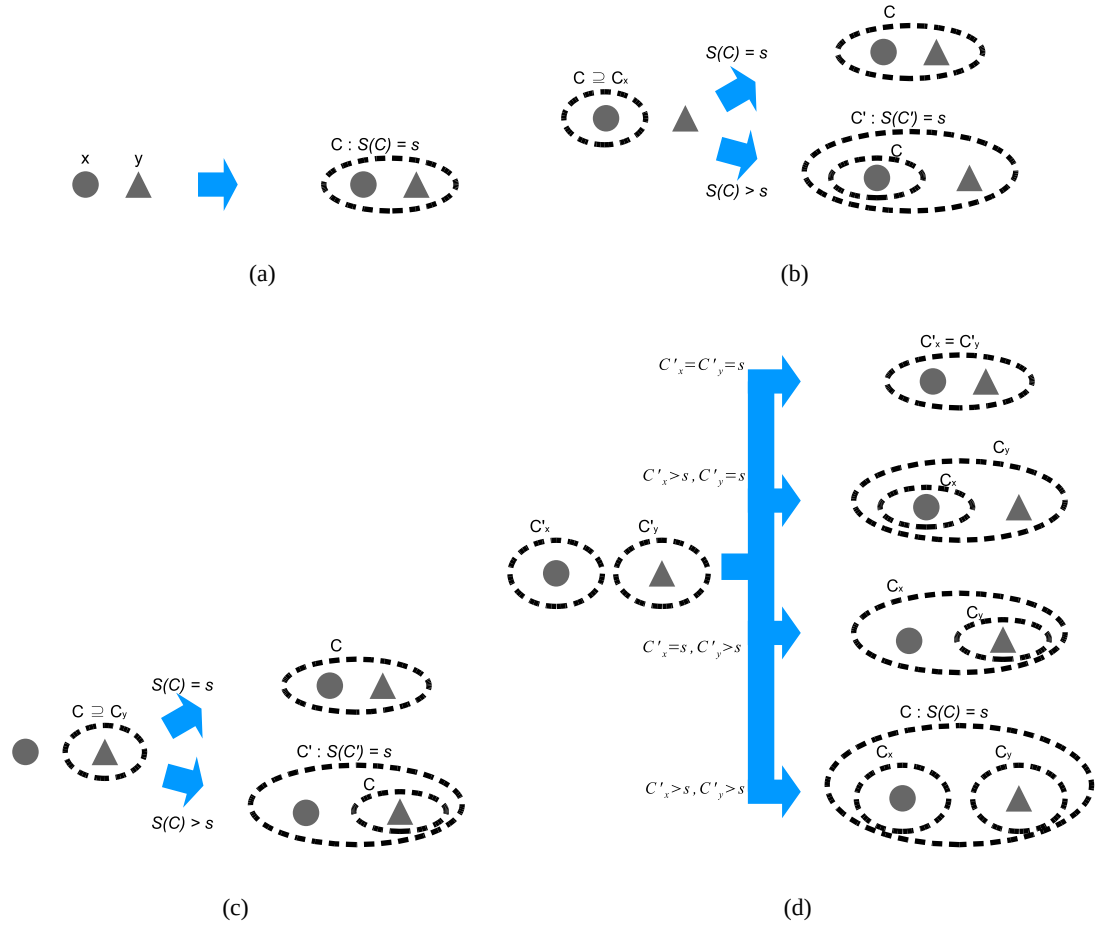


図 3.4: クラスタリングの過程

$G(C)$  は  $C$  の最上位の親クラスターで、以下の式で定義する。

When  $C' \supseteq C' \supseteq C$

$G(C) = C'$

satisfied  $S(C') = \min_{C' \in C'} S(C')$

以上を行うことにより、2つのノードの組み合わせの結果から全ノードのクラスターを木構造で構築していくことが出来る。エッジについては、それぞれの接続ノードを包含するクラスターにも同様に接続する。

|       | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|-------|-------|-------|-------|-------|
| $v_1$ | -     | 0.9   | 0.8   | 0.72  |
| $v_2$ |       | -     | 0.9   | 0.81  |
| $v_3$ |       |       | -     | 0.9   |
| $v_4$ |       |       |       | -     |

表 3.1: 類似度表

### 3.2.4 本手法の特徴

このクラスタリングの特徴として、通常の階層型クラスタリングよりも計算速度が速いことが挙げられる [20]. 通常の階層型クラスタリング手法は、

1. すべてのノードの組み合わせの内最も高い類似度の要素を 1 組探索し、クラスタとして結合させる
2. 残りのノードの集合と生成されたクラスタの類似度を再計算

を繰り返し、階層構造を形成していく. それと比較し、本手法は類似度の計算は 1 回で良いので、類似度計算自体のオーダは  $O(n^2)$  にとどめることが出来る. 加えて、類似度の情報が絶対値で与えられることも特徴である. 通常のクラスタリングは、類似度が高い要素ごとにクラスタリングを行うため、後にクラスタリングした要素の類似度がその前にクラスタリングした要素の類似度よりも高いか低いかは不明である. 後に述べる描画手法ではこの類似度の絶対性が有効に働く.

もう一つの特徴は、ノードの並び順（クラスタの結合順序）によりクラスタリング結果が変わらない一貫性である [19]. 例として、4 つのノード  $v_1, \dots, v_4$  の接続ノードの集合  $A(v_1), \dots, A(v_4)$  を

$$A(v_1) = \{2, 3, 4, \dots, 10\}$$

$$A(v_2) = \{1, 2, 3, \dots, 10\}$$

$$A(v_3) = \{1, 2, 3, \dots, 9\}$$

$$A(v_4) = \{0, 1, 2, \dots, 9\}$$

として、類似度を計算した結果を表 3.1 に示す.

ここで、類似度が 0.9 のものが 3 つあるが、結合する順番はノードの並び順に依る. すなわち、 $v_1-v_2$ ,  $v_2-v_3$ ,  $v_3-v_4$  のどれを始めにクラスタとするかである. それぞれを最初にクラスタリングを行い、全ノードをクラスタリングした場合の結果をデンドログラムとして表したのが、図 3.2.4 である.  $v_1-v_2$  と  $v_2-v_3$  は論理的に同じ構造であるが、 $v_3-v_4$  は他二つとは異なる構造になる.

本論文のクラスタリング手法でこれらノード集合のクラスタリングを行うと、どの順番で行っても、 $v_1 \sim v_4$  は類似度 0.9 のクラスタで結合され、結果の一貫性が保たれる. 通常のク

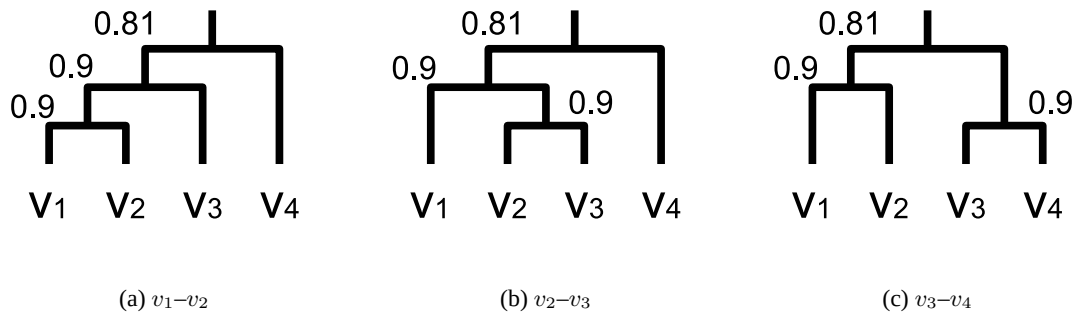


図 3.5: クラスタリング順による変化

ラスタリングよりも高い類似度で結合されるクラスタが多くはなるが、可読性は変わらないと考える。また、後から形成されたクラスタが、以前に形成したクラスタよりも類似度が高い状態になる「反転現象」とよばれる歪みも発生しない。順序依存性と反転現象の影響を受けない階層型クラスタリング手法は、最短距離法のみである。

### 3.3 描画システム

本研究ではクラスタリングを行ったデータを可視化するツールとして、「GTBC-GraphML Translator for Bipartite Clustering-」と「ClusterdAnchorViz」の開発を行った [21]。

GTBC はグラフのクラスタリングツールであり、2部グラフのデータにクラスタリングを付加する。GTBC は Java™ 6.0 (Java™ Platform, Standard Edition 6 Development Kit <sup>2</sup>) を用いて製作した。GraphML<sup>3</sup> 形式で記述された2部グラフのXMLデータファイルを入力として、クラスタリングのデータを付加したNestedGraphのGraphML形式のXMLを出力する。GraphMLとはグラフをXMLで記述する形式の一つであり、多くの可視化研究においても用いられている [22]。

ClusterdAnchorViz は、クラスタ情報を効率的に可視化する描画機能を備えた2部グラフ描画ツールである (図 3.6)。グラフの描画レイアウトには、アンカーマップ [16] のレイアウトを用いている。ClusterdAnchorViz も Java™ 6.0 を開発言語とし、Applet として作成している。

クラスタリングの情報を用いてのネットワーク描画スタイルは、従来手法ではノードとエッジをまとめて表示する手法が一般的である。加えて、そのインタラクションには該当ノードのクリックにより集約と展開を行う手法が多い。これらの手法では、大規模の場合読み手が要求するネットワーク図にたどり着くまでかなりの作業を要する、似た傾向をもつクラスタの発見などが困難であるなどの問題がある。

ClusterdAnchorViz には「縮退描画手法」と「等類似度線描画手法」という二つの描画機能が備わっている。縮退描画手法は、クラスタの類似度情報によって複数のノードを1つのノードとして表示させる手法である。等類似度線描画手法は、クラスタに属するノードを閉曲線描画するスタイルである。集約と展開とは異なる描画スタイルで、要素間の関連性の情報を



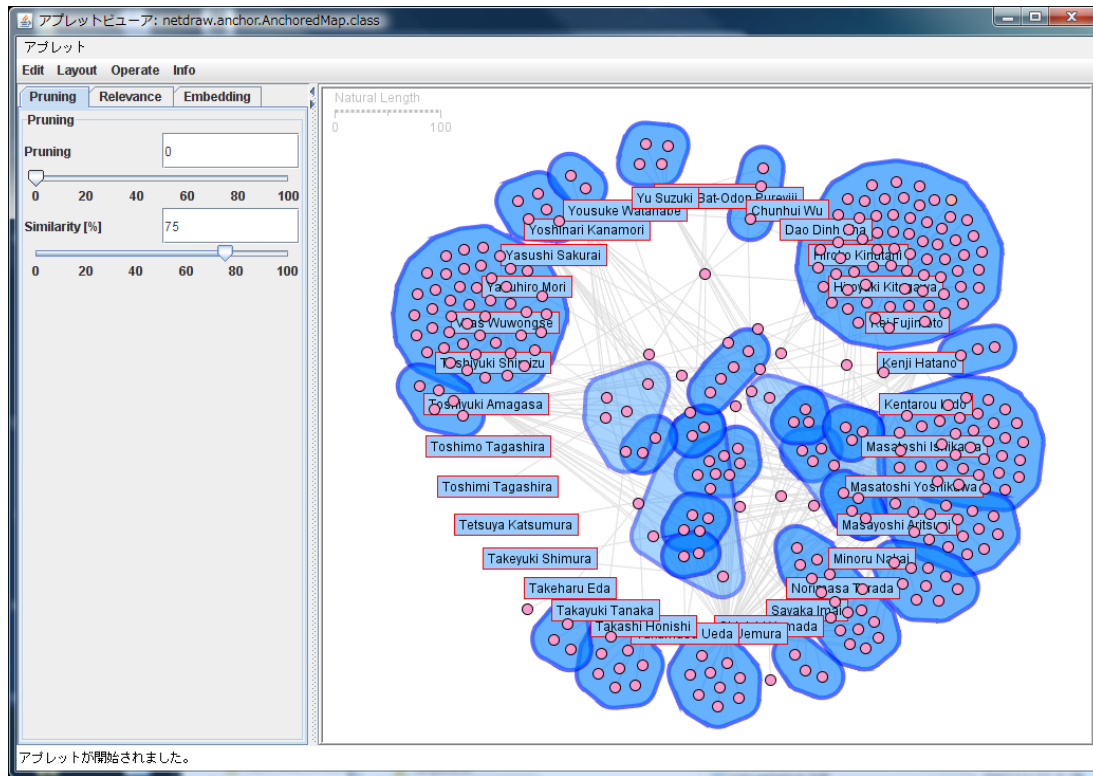


図 3.6: ClusteredAnchorViz の概観

読み手にわかりやすくする狙いがある。

いずれも、スライダーにより描画をインタラクティブに変更できる。これにより、各々の読み手に適応した描画が出来ると考えられる。

システムの構成を図 3.7 に示す。

まず、可視化する 2 部グラフの GraphML 形式の XML データを、GTBC を用いてクラスタリング情報付きの 2 部グラフデータに変換する。変換された GraphML データを ClusterdAnchorViz で読み込む。

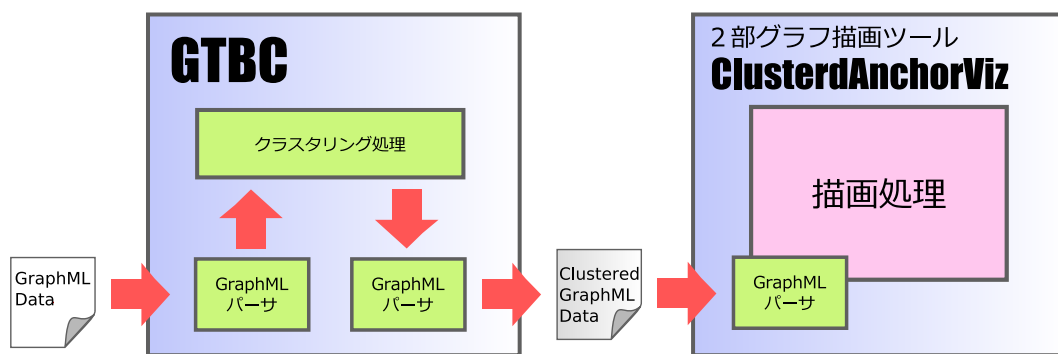


図 3.7: システム構成

## 第4章 縮退描画手法

### 4.1 概要

大規模2部グラフの可読性向上のための描画手法として、「縮退描画手法」の開発を行った。本手法は、クラスタの情報を利用し複数ノードを結合して表示する手法である。

大規模なグラフはノード数が多い。そのため、ノード数を削減することでグラフの関係を簡潔にすることができ、空間構造的な可読性の向上を狙うことが可能である。グラフ自動描画においてクラスタリングを利用する従来手法は、ノードを結合して表示する本手法のようなスタイルが通常である。

従来手法において、分割型クラスタリングを用いてノードを結合する描画手法の場合、生成された図は簡潔になっているが、クラスタリング前から情報が欠落してしまっているため、グラフから読み取れる内容が薄くなる危険性がある。また、k-means 法などに代表されるとおり、この手法はノードの集合を  $n$  個のクラスタに分割する手法である。分割する数を読み手により指定できたとしても、分割数を変更する度にクラスタリングをやり直さなければならない。大規模グラフの場合、クラスタリングの計算量はかなり大きくなるため、インタラクシオン性に問題が生じる可能性がある。

従来手法において階層型クラスタリングを用いた描画手法は、読み手の操作に対応してグラフが動的に変化する手法が多く存在する。ただし、クラスタの集約と展開を行うインタラクシオン方法としては、ノードやクラスタのクリックにより行う手法がほとんどである。このインタラクシオンの欠点がグラフが大規模化しノードの数が膨大になると、読み手が適当と感じるグラフを作り出す作業が大変になる、と言うことは先に述べたとおりである。また、クラスタリングにより生成されたデンドログラムはただの二分木の構造になっているため集約と展開を行う基準はデンドログラム上の分岐点でしか行うことが出来ない。

提案手法ではこれらを解決するために、クラスタリングの際に使用した類似度情報を用いる。グラフで一貫した類似度情報を持つことで、デンドログラムを垂直に裁断した情報に意味を生じさせることが出来る。すなわち、ノード間の関係に類似度という観点を加えることで、グラフの可読性を向上させるのである。

### 4.2 手法

ノードとクラスタは  $0[\%] - 100[\%]$  の類似度でクラスタリングされる。ある一定の閾値  $t$  を定め、 $t$  よりも類似度の高いクラスタのみを描画することを基本とする。

本手法は、グラフ構造の変換部分とグラフの描画部分の大きく二つに分けられる。上記の仕組みを構築する手法を次項より述べる。

#### 4.2.1 グラフ構造の変換

本手法のグラフはノード、エッジ、クラスタの集合の構造になっている。描画すべき要素を判断するため、要素はそれぞれ、現在グラフ要素に含まれているかどうかを表す属性値「**Active**」を持つようにした。Active は boolean 値を取り、True の時にその要素はグラフの構造に含まれる（描画される）ことを示す。逆に False の時はグラフに含まれない要素であることを示す。その他の属性として、クラスタは属性として自分が含むノード集合を持ち、エッジは属性として自分が接続するノード 2 つを持つ。

#### クラスタ探索

プログラムの起動時とスライダの値  $t$  が変更された場合、以下の順で各要素の Active 属性を切り替えていく。なお、初期値は Active = **False** とする。

1. 探索クラスタ  $c$  = (ルートクラスタ) とおく
2.  $S(c) < t$  であれば
  - (a)  $c$  の保持しているノードすべての Active 属性を True とする
  - (b)  $c$  の Active 属性を False とする
  - (c)  $c$  が子クラスタを持っていたら、 $c$  = 子クラスタとし 2 に戻る（子クラスタが複数あった場合はすべての子クラスタについて探索を行う）
3.  $S(c) \geq t$  であれば
  - (a)  $c$  の保持しているノードとクラスタのすべての Active 属性を False とする
  - (b)  $c$  の Active 属性を True とする

ノードとクラスタ走査が終了したら、Active 属性が True となっているノードとクラスタを接続するエッジを Active = True, 他のエッジを Active = False とする。最終的に、Active = True となっているノードとエッジ、クラスタでグラフを構成することで、要求を満たすグラフを可視化することが可能となる。

図 4.1 に、クラスタリングの構造をデンドログラム（樹状図）で示す。ここでは、分岐点を四角で表示してクラスタ要素であることをわかりやすく描いている。クラスタの探索の動きを矢印で、それにより決定された表示されるべきクラスタとノードのみを赤で塗りつぶして表示している。

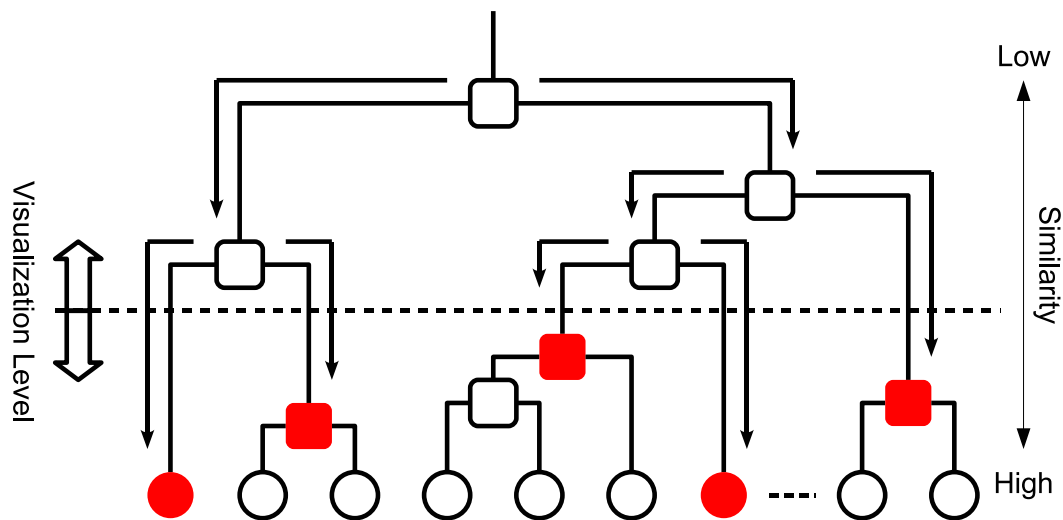


図 4.1: スライダー変更時の走査

#### 4.2.2 描画

デンドログラム上の底辺の要素はノードであり，すべての節点（分岐点）はクラスタに対応する．閾値  $t$  以上の類似度のクラスタを表示するということは，デンドログラムを水平に切断し，切断面のエッジに接続する底辺側のノードとクラスタのみを表示するのと同義である（図 4.2）．

この閾値  $t$  を読み手が可視化されたグラフを見ながら変更することで，動的クラスタリングを行っているような描画状態を与える．これにより読み手に適切な規模のグラフを可視化できると考えた．

### 4.3 実装

図 4.3 は，アプリケーションの初期画面である．右側のグラフ描画パネル，左側にコントロールパネルが設置している．コントロールパネル内には類似度の閾値を調整するためのスライダーがある．

スライダーの値は最大値 101[%]，最小値 0[%] を設定している．最大値が 101[%] であるのは，クラスタ化されないオリジナルのノードのみで構成されるグラフを描画するためである．ノード間類似度の定義域は 0[%]～100[%] であるので，便宜的に 101[%] という値を設定している．

初期値は 101[%] でこの状態では先行研究の AnchoredMap[16] と同様の振る舞いをする．また逆に最小値に位置しているときは，全ノードが一つのクラスタにまとめられ，画面上には 1 つのクラスタのみを表示する．

クラスタは，形状はフリーノードと同一であるが，アンカーやフリーノードとは異なる色で表示することでクラスタであることを表現している．また，ラベルも “cluster” と表示する

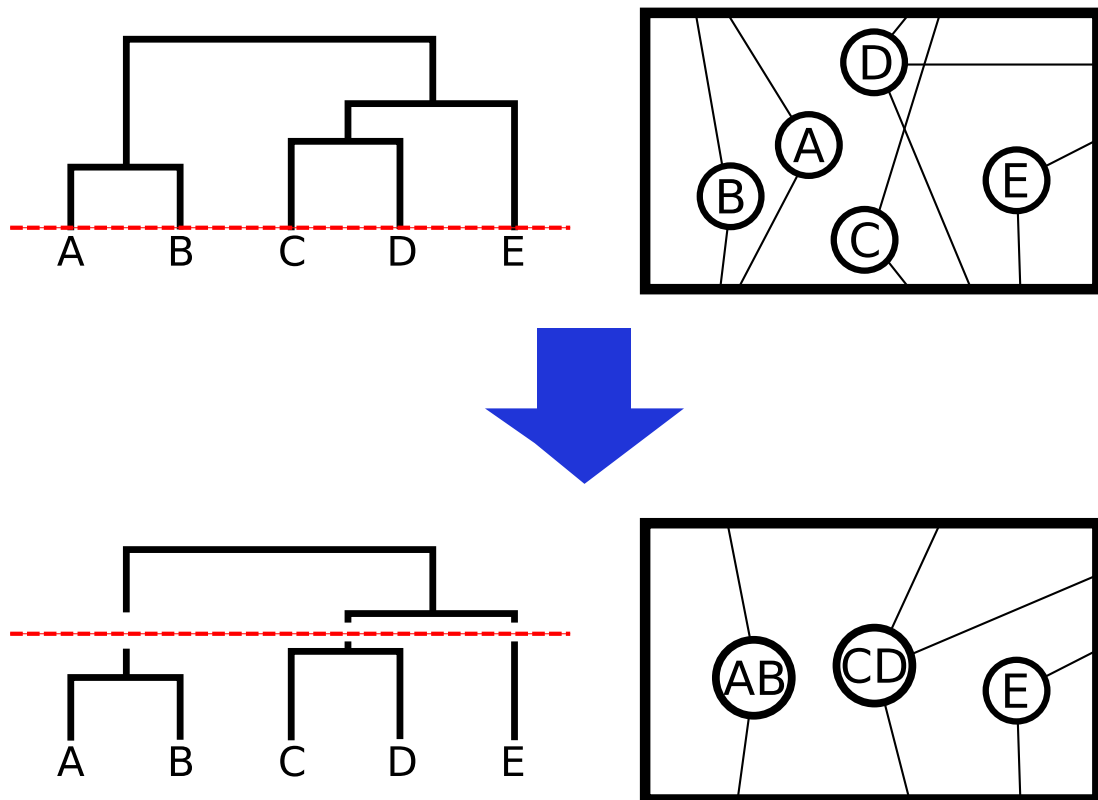


図 4.2: デンドログラム切断時の表示するノードとクラスタ

ことによりクラスタであることがわかるようにした。

### クラスタの内部表示

クラスタリングされ縮約されたグラフのある注視点を詳細に確認するのに有効な手法として、焦点になるノードのみの展開表示が挙げられる。そこで、指定したクラスタをダブルクリックにより、展開表示できるようにした。

可視化の目的の一つには注目したいノードの関係情報を読みとることが挙げられる。その際、クラスタリングされた内部のオリジナルノードの情報が重要となるが、内部構造を把握するにはクラスタリングを解除し、オリジナルのネットワーク構造にしなければならない。すると、注目する点に不必要なノードが多数現れてしまう。特定のクラスタの展開が出来ることで、注目するノードの論理的構造を注視することができ、利用者の認識を適切に補助することが可能となる。

属性 Active の変更は以下の通りである。

1.  $c$  = (指定クラスタ) とする

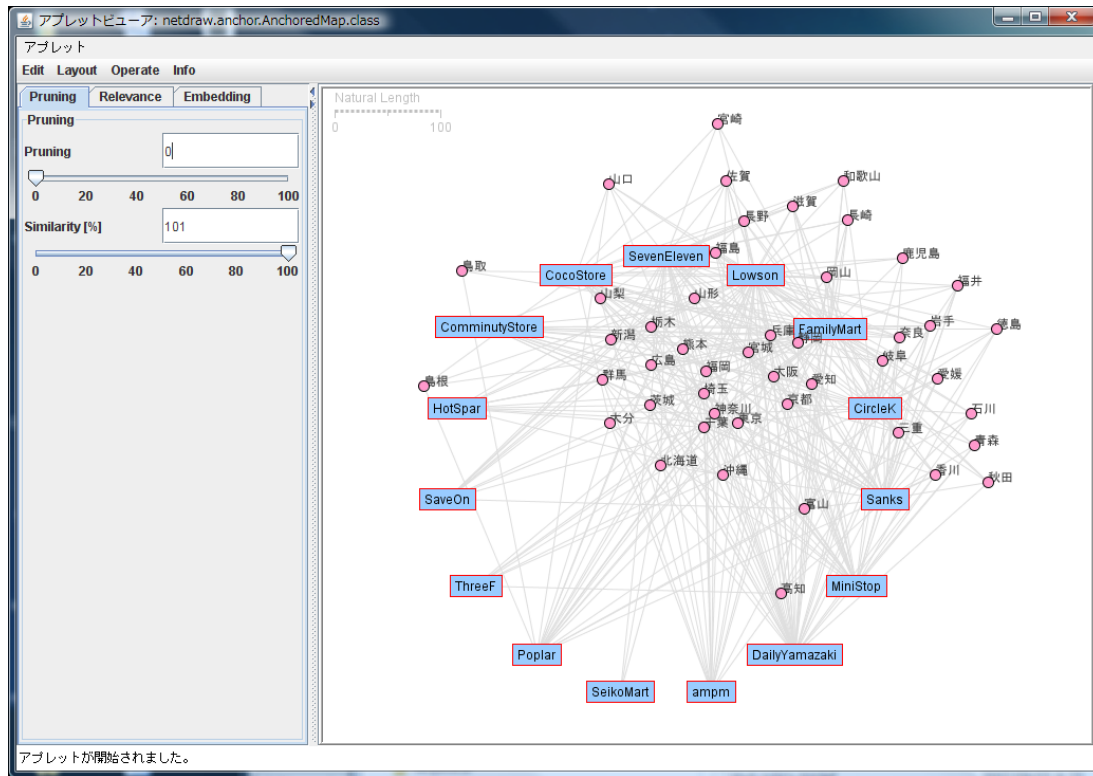


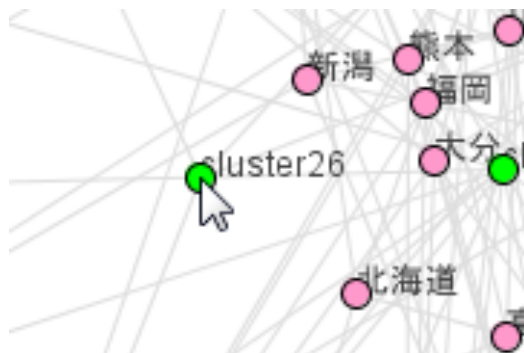
図 4.3: アプリケーション初期画面

2.  $c$  の保持しているすべてのノードとクラスタの Active 属性を True とする

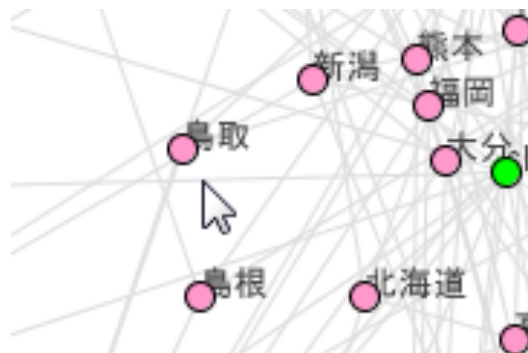
3.  $c$  の Active 属性を False とする

ノードとクラスタの処理後,  $c$  に接続するエッジの Active 属性を False とし,  $c$  の保持しているノードとクラスタを接続するエッジの Active 属性を True とする.

図 4.4 に展開の様子を示す. ラベル名「cluster26」となっているクラスタをダブルクリックすることで, 内部のノードである「鳥取」と「島根」が展開される.



(a) 展開前



(b) 展開後

図 4.4: クラスタの展開の様子



## 第5章 等類似度線描画手法

### 5.1 概要

大規模2部グラフの可読性向上のためのもう一つの描画手法として、「等類似度線描画手法」の開発を行った(図5.1)。本手法は、クラスタリングの情報を利用し、関連度の高いノード集合を閉曲線で囲んで表示する手法である。

グラフを縮約させる手法とは別の観点で可読性を向上させる手法として、読み手の視点を注視させる手法の導入が考えられる。意味的な可読性に焦点を当て、これをフルに向上させる手法である。

大規模なグラフはその構造が複雑になる傾向がある。グラフを通常の方法で描画し、そこに加える形で構造をわかりやすくアノテーションをつけることで、可読性を向上させようというものである。

人間の注視特性より、重要な部分を囲み線などであらわす方法が有効であることは良く知られている。日常生活でもメモのアノテーションとして使われたり、他人に説明を行う際にもこの行為はしばしば見受けられる(図5.2)。

本手法を用いない可視化結果、つまり既存のアンカーマップからもノードの集まりでクラスタの形成を読み取ることは出来るが、クラスタリングの結果を等高線のように表示することでノード間の関連性を能動的に注視させ、より理解しやすくさせるのが狙いである。また、アンカーマップから読み取れるクラスタとクラスタリングの解析結果の違いも読み取ることができ、そこから新たな知見が得られることが期待できる。

構造の認識には観点の導入が効果的であることは第2.5章で述べたとおりである。等類似度線描画はクラスタリングの結果をノードをまとめることに利用せず、同じ類似度を持つノードの周囲を地図の等高線のように描画することに利用することで、ノード間の関連性を能動的に注視させることが狙いである。

本手法も、読み手や場面の違いに対応するため、どの類似度以上の囲み線を表示するかを調節する必要があると考えた。インタフェースは縮退描画同様スライダの形式を取っている。スライダでは表示するクラスタの類似度の閾値を調整することができ、閾値 $t$ 以上のクラスタを示す囲み線を表示出来るようにした。

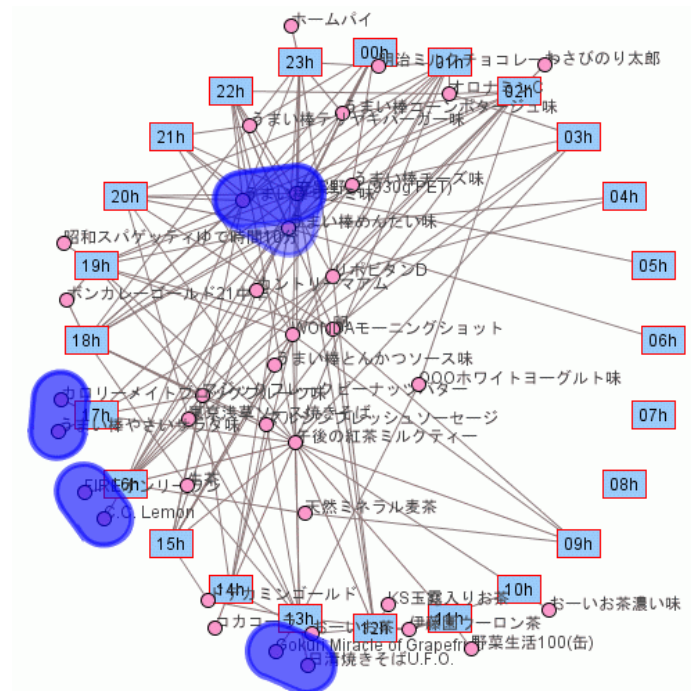


図 5.1: 等類似度描画手法を用いた表現

## 5.2 手法

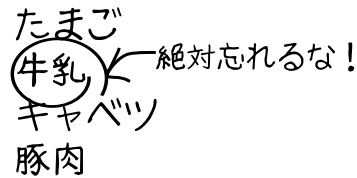
本研究で用いる階層型クラスタリングは木構造を持っている。そのため、網図系の木構造から領域系の入れ子構造に変換し表現することが可能である。本手法も、グラフ構造の変換部分とグラフの描画部分の二つに大きく分けられる。

### 5.2.1 グラフ構造の変換

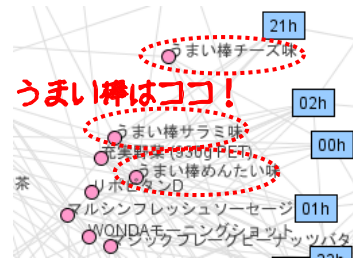
等類似度線描画手法も、縮退描画手法同様事前のクラスタリング情報を用いて描画を行う。こちらグラフは、ノード、エッジ、クラスタの構造を持っており、クラスタの描画方法が異なると考えれば良い。

等類似度線描画手法は、ノードとエッジは常にオリジナルのグラフの情報で描画される。閾値  $t$  の変更により変化させるのは、クラスタを描画するかしないかである。そのため、Active 属性の変更作業はクラスタのみを扱う。

スライダの位置している値の類似度以上のクラスタを表示するというのは、デンドログラムを水平に切断し、切断面のエッジに接続する底辺側のクラスタのみを表示するのと同義である。



(a) メモでのアノテーション例



(b) グラフ描画でのアノテーション例

図 5.2: アノテーションで囲み線を用いる例

## クラスタ探索

1. 探索クラスタ  $c$  = (ルートクラスタ) とおく.
2.  $S(c) < t$  であれば,  $c$  の Active 属性を False とする
3.  $S(c) \geq t$  であれば,  $c$  の Active 属性を True とする
4.  $c$  が子クラスタを持っていたら,  $c$  = 子クラスタとし 2 に戻る. (子クラスタが複数あった場合はすべての子クラスタについて探索を行う)

### 5.2.2 描画

クラスタは, そのクラスタが保持しているノード (子クラスタが保持している分も含め) を囲む閉曲線を描画することで表現する.

閉曲線描画は以下のステップで行う.

1. クラスタに含まれるノード集合を  $V_A$  とする (図 5.4(a))
2. Graham 走査 [23] によりノード集合  $V_A$  の最外周凸多角形を形成するノード集合  $V_B$  を探索する (凸包形成). (図 5.4(b))
3. ノード集合  $V_B$  の周囲を直線と円弧による閉曲線で囲む. (図 5.4(c), 5.4(d))

凸包は  $n$  個の点の集合  $P_n$  があつたとき, 任意の 2 点  $p_1, p_2 \in P_n$  を結ぶ線分を必ず含む閉曲線と定義される. 言い換えると, 点の集合の一番外側の点を結んだ多角形のことである. (図 5.3)

Graham 走査 (Graham Scan) は多点集合から凸包を求めるアルゴリズムの一つである. 最小の  $y$  座標を持つ点  $p_0$  とし点集合  $P_n$  を偏角順にソートする. その後  $p_0$  から順に線分を引いて行き, なす角が  $180^\circ$  以下になる点のみを選択することで凸包となる点を探索する. 単純な

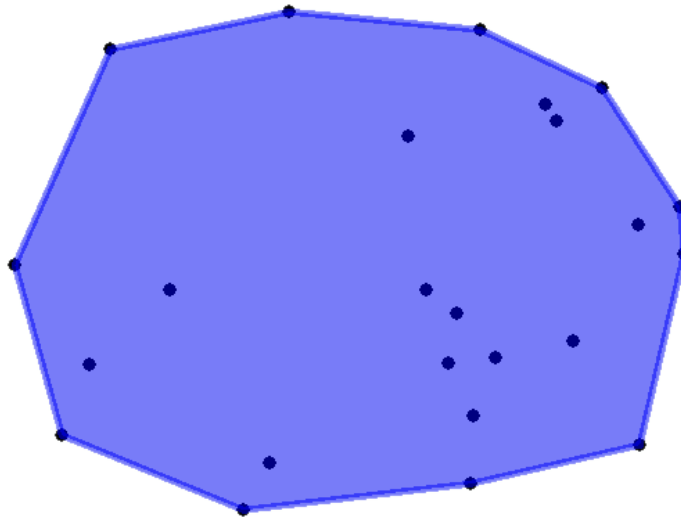


図 5.3: 凸包—黒い点をすべて囲み，凹みのない多角形

アルゴリズムだが，グラフがレイアウトされてしまえばソートする必要がなくなり計算量はあまり多くなるため，この手法を採用した．

囲み閉曲線は直線と円弧により描画を行った．凸包の点を  $p_0, p_1 \cdots p_n \in V_B$  としたとき，閉曲線は以下のステップで直線と円弧により描画を行う．

1.  $i = 0$
2. 点  $p_i$  を通り，点  $p_i$  と点  $p_{i+1}$  を結ぶ線分に垂直な直線上で，点  $p_i$  から距離が Offset の距離にある点を点  $\hat{p}_i$  とし，点  $p_{i+1}$  から距離が Offset の距離にある点を点  $\hat{p}_{i+1}$  とする．
3.  $i < n$  の時  $i = i + 1$  として 1 に戻る
4. 点  $p_n$  を通り，点  $p_n$  と点  $p_0$  を結ぶ線分に垂直な直線上で，点  $p_n$  から距離が Offset の距離にある点を点  $\hat{p}_n$  とし，点  $p_0$  から距離が Offset の距離にある点を点  $\hat{p}_0$  とする．
5.  $i = 0$
6. 点  $\hat{p}_i$  と点  $\hat{p}_{i+1}$  を線分で接続する
7. 点  $\hat{p}_{i+1}$  と点  $\hat{p}_{i+1}$  を， $\hat{p}_{i+1}$  を中心とした半径 Offset の円弧で接続する．
8.  $i < n$  の時  $i = i + 1$  として 5 に戻る
9. 点  $\hat{p}_n$  と点  $\hat{p}_0$  を線分で接続する
10. 点  $\hat{p}_0$  と点  $\hat{p}_0$  を， $\hat{p}_0$  を中心とした半径 Offset の円弧で接続する．

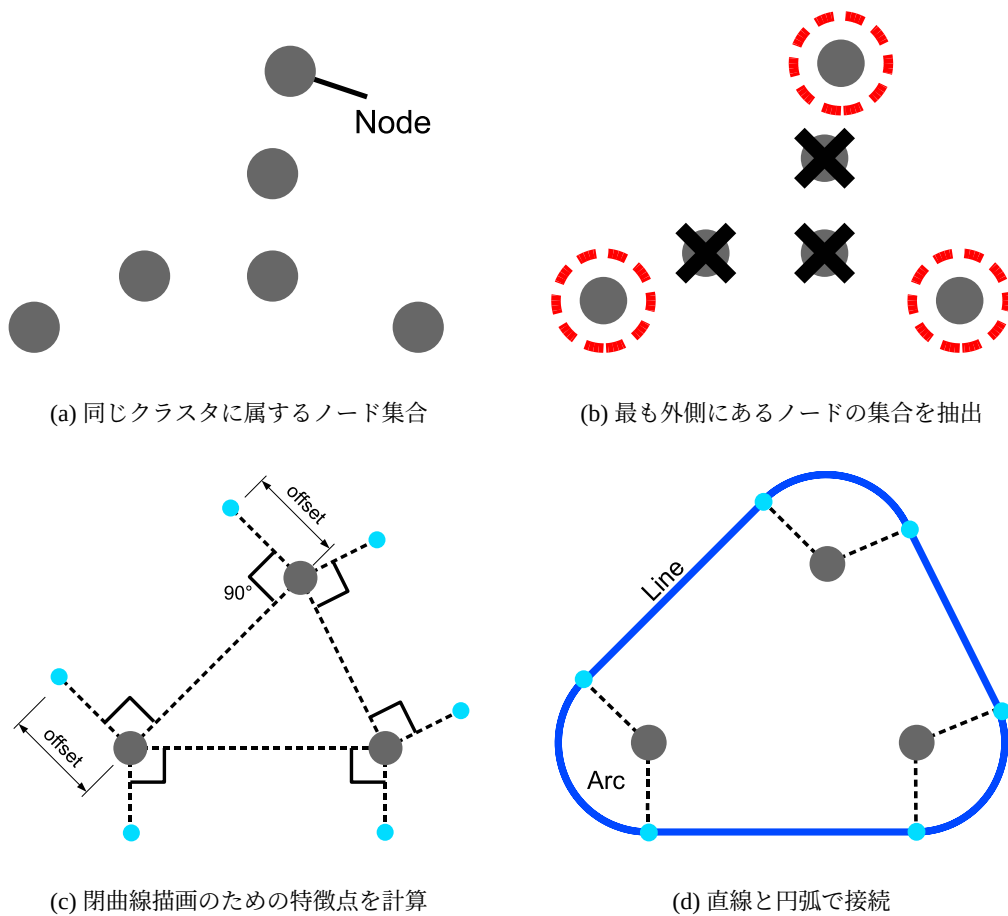


図 5.4: 閉曲線を描画する方法

### 5.3 実装

図 4.3 は、アプリケーションの初期画面である。初期画面は縮退描画手法と全く同じであり、スライダの定義域も同様に設定している。なお、ノードのレイアウトはアンカーマップのレイアウトを踏襲しており、スライダを動かすことによって位置が変化することはない。

等類似度線は透過を持たせた青色で塗りつぶした閉曲線で描画される。透過度はクラスタの類似度に比例しており、類似度が高いほど透過度を低くなるように設定している。透過色にすることでクラスタの重なりを表現でき、ノード間の関連度が高いところや、多くの階層を形成するクラスタが濃く表示される。

## 第6章 評価実験

提案手法による描画手法により、可読性が向上したかどうかを評価実験により検証する。先行研究のアンカーマップを比較対象とし、被験者のグラフから得られる知識が増加することを確認する。

### 6.1 実験目的

実験の目的は「提案手法は大規模グラフの可読性を向上させることが出来る」という事柄の検証である。

また、以下の仮説を検証することも実験の目的となっている。

**仮説1** 可読性は「空間構造的な可読性」と「意味的な可読性」の2つの要素を持っている

**仮説2** 縮退描画手法は空間構造的な可読性を向上させる。

**仮説3** 等類似度線描画手法は意味的な可読性を向上させる

**仮説4** 読み手により「可読性が高い」と感じる表現は異なる。

### 6.2 実験方法

実験方法は、筆者側で2部ネットワークを用意し、それを先行研究のアンカーマップの表現と縮退描画手法の表現、等類似度線描画手法の表現の3つで描画し、被験者に読ませる。被験者には、ノードの中から1つの注視点を選択させ、それぞれの手法でその注視点に関して得られた知識を記述形式で解答させる。これにより、得られる知識量の向上が見られた場合、意味的な可読性の向上が得られたとする。また、提案する2つの描画手法に対しては、「見易いか?」という2択の選択しを用意し、空間構造的な可読性の向上を確認する。最後に、操作に対しての忌憚のない意見も述べて貰う。

なお、どのようなグラフを選択し可視化するかは大きな問題である。グラフを可視化して読むという行為は日常生活上あまり遭遇しない行為であり、読み手の興味度によりグラフの読み取り能力がかなり変化する。そのため、選択するグラフによっては被験者に差が出る可能性が高い。今回用いたグラフは、コンビニ店舗数と都道府県の関係のグラフである。理由として、日本人であれば馴染みのあるコンビニと、被験者の出身地の情報が得られるということで、興味をもってもらえる可能性が高いと考えたためである。



表 6.1: 縮退描画手法実験結果

| 評価項目            | A   | B   | C   | D   | E   | F | G | H   |
|-----------------|-----|-----|-----|-----|-----|---|---|-----|
| 見易いかどうか         | ○   | ○   | ○   | ○   | ○   | × | × | ×   |
| 見易いと思った時の類似度の閾値 | 80% | 70% | 85% | 74% | 75% | — | — | 70% |
| 新たな知見が得られたか     | ○   | ×   | ○   | ○   | ○   | ○ | × | ○   |

表 6.2: 等類似度描画手法実験結果

| 評価項目                 | A | B   | C   | D | E    | F   | G | H   |
|----------------------|---|-----|-----|---|------|-----|---|-----|
| 見易いかどうか              | × | ○   | △   | × | ×    | ○   | × | ×   |
| 見易いと思った時の類似度の閾値      | — | 90% | 94% | — | 100% | 80% | — | 85% |
| ノードの関連性は読み取りやすいと感じたか | ○ | △   | ○   | ○ | ×    | ○   | — | ○   |
| 新たな知見が得られたか          | × | ○   | ○   | ○ | ○    | ○   | × | ○   |

まで大きくなかったため（フリーノードが47個）、縮退描画の効果が薄かったためではないかと考えられる。

等類似度線描画手法において、この描画手法が見易いと答えた者は2人で、空間構造的な可読性向上はあまり見込めないことが確認できる。このうち2人についても、「すっきりした」と感じたわけではなく、近くにたくさんノードが有るような状況で、似ているノードと似ていないノードとが区別出来るため、「見易い」と答えたと推測される回答であった。

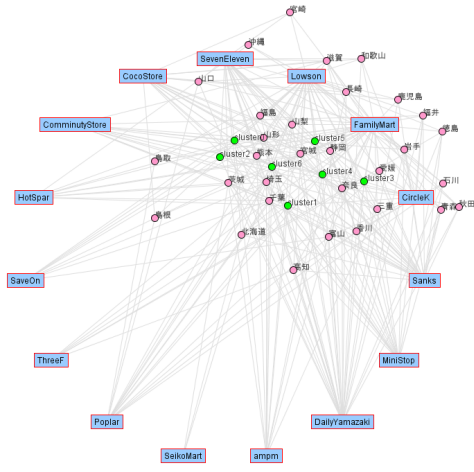
ノードの関連性の読み取りやすさは5人が向上したと回答し、また新しい知見も得られたと回答した者は6人に上った。このことから仮説3が示され等類似度線はグラフの構造の把握に有効に働くことが分かる。

注目すべき点は、「見難い」と答えているにもかかわらず、「ノードの関連性を取りやすい」と答えている者が多数存在することである。見易さと意味の取りやすさは別の問題であるという仮定を裏付ける結果であり、すなわち、仮説1が示されたことになる。

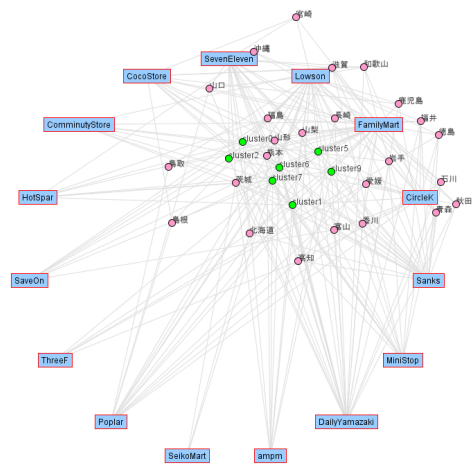
閾値の読み手によっての違いであるが、見易いと感じる閾値については被験者ごとにばらつきが生じており、読み手により見易いグラフの規模が異なることを示す結果が得られた。最低値の85%と70%は構造がかなり異なり、読み手自身がクラスタリングの強さを変更出来ることでの可読性への効果が表れているのではないかと考えられる（図 6.2(c), (e)）。最後に新たな知見が得られたと答える者は6人で、この手法により意味的な可読性向上にも貢献出来たことを示す結果が得られた。

縮退描画手法と同様に等類似度描画手法にもこの傾向が表れている。見易いと思う閾値は、最高の被験者 E が 100%、最低の被験者 H が 85% という結果になった。図 6.2(a), (c) はそのときの描画状態である。見た目にもかなり異なり、両者は異なる視点でグラフを観察していることが伺える。ここから、仮説4が示される結果となった。

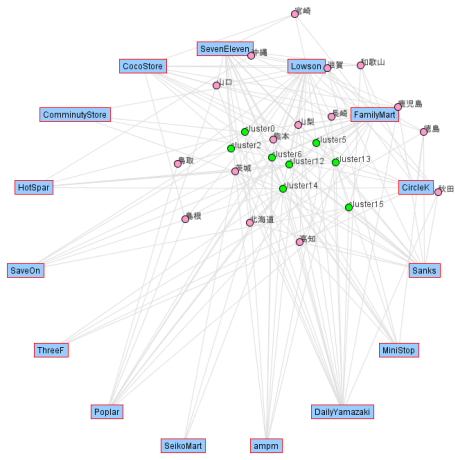




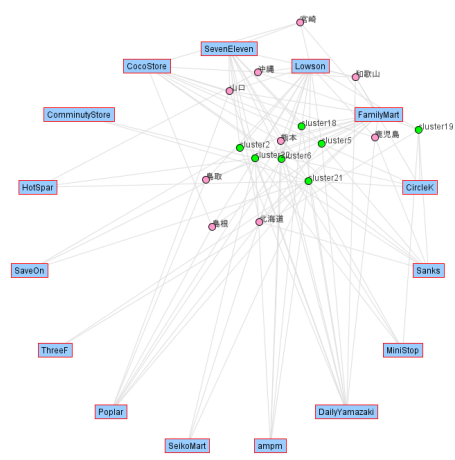
(a)  $t = 100$



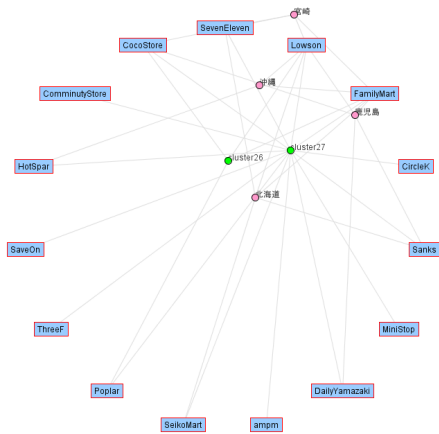
(b)  $t = 90$



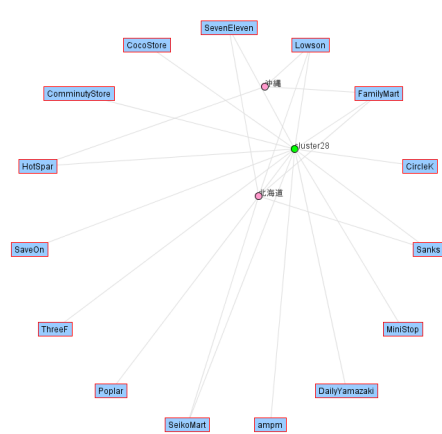
(c)  $t = 85$



(d)  $t = 80$

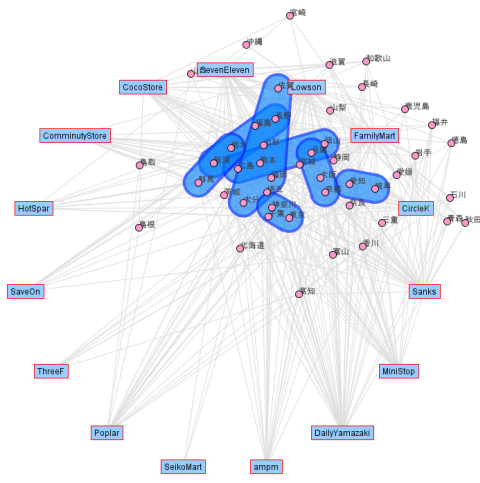


(e)  $t = 70$

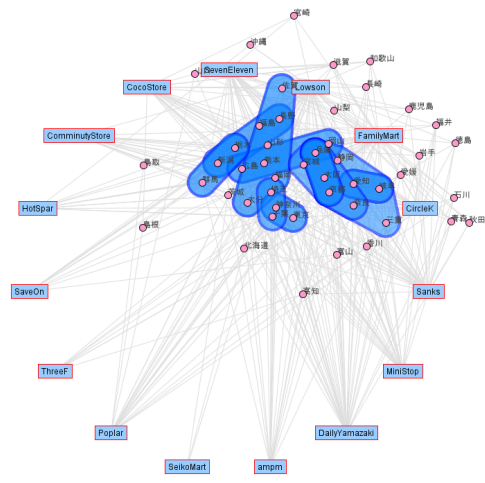


(f)  $t = 55$

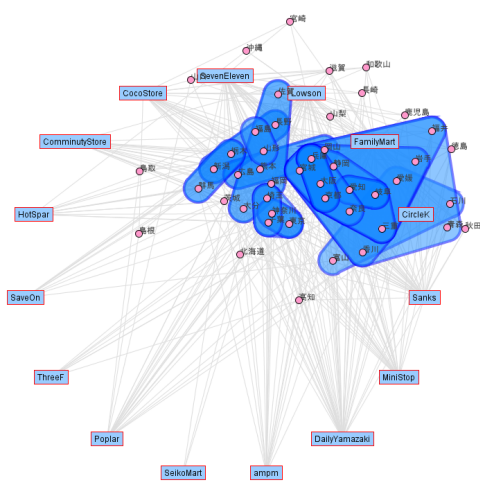
図 6.2: 実験に用いたグラフを縮退描画手法で描画した様子



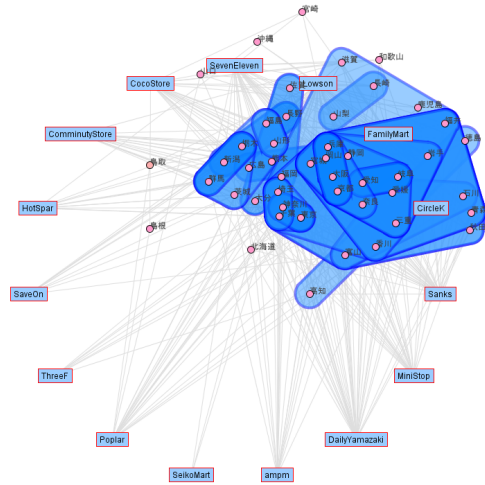
(a)  $t = 100$



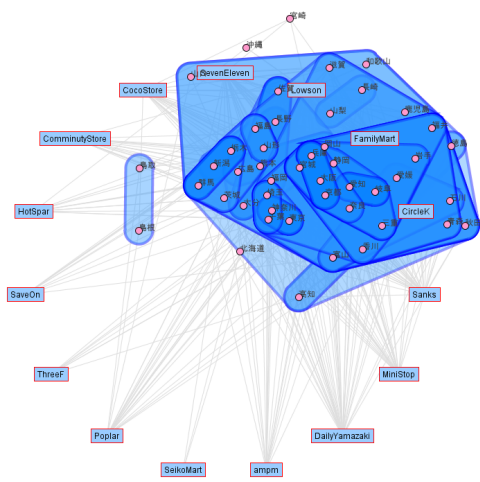
(b)  $t = 90$



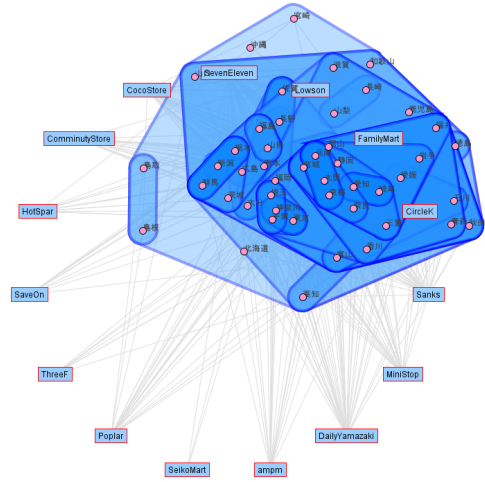
(c)  $t = 85$



(d)  $t = 80$



(e)  $t = 70$



(f)  $t = 55$

図 6.3: 実験に用いたグラフを等類似度線で描画した様子

## 第7章 適用例と考察

縮退描画手法と等類似度線描画手法により様々な2部グラフの可視化を行った。本研究の表現手法は、可視化するグラフによりその様相がかなり変化する手法である。

本章では、具体的な使用例を示しながら、得られた新たな知見とその議論を行う。

### 7.1 購買データの可視化

私が所属している研究室では「お買い物システム」という購買システムが運用されている。食料品が販売されており、教員と学生が購入した商品と購入した時間がデータベースに保存されている。

#### 7.1.1 小規模購買データ

商品とそれが売れた時間帯の関係のグラフを用いる。2006年6月から9月までの期間、比較的小規模のアンカーは商品が購入された時間帯とし「00h」～「23h」の24個のノードを、フリーノードは購入された37種類の商品とした。エッジは商品とそれが購入された時間帯に接続され、スプリングの強さは購入された商品数に比例して強くしている。

比較のためアンカーマップで描画した図を示す(図7.1)。アンカーマップのスタイルでは、ノードの配置から「午前中は買い物が少なく、夕方から夜にかけて頻繁に買い物がなされている」といったことを直観的に読み取ることが可能である。また、ラベルを読むと、「正午近くと18時から深夜にかけて食品が売れ、飲料製品は時間帯に関係なく購入されている」という事柄も把握することが出来る。

縮退描画手法で類似度の閾値  $t = 35\%$  を指定したときの例を図7.2(a)に示す。スライダにより閾値を下げていくとノードが集約されていく。このグラフ中には「うまい棒」と言う商品の味違いがいくつか存在する。コンポタージュ味、テリヤキバーガー味、サラミ味、めんたい味、チーズ味、トンカツソース味の6種類であるが、表示レベルを  $t = 35\%$  まで下げたときに、描画されるのはトンカツソース味のみである。すなわち、トンカツソース味は購入される時間の傾向が他の味とは異なることが分かる。加えて、15hに強く引かれていることから、他の味よりも3時のおやつに食べられる傾向が高いのではないかという推測が得られる。

このように、類似していないという情報だけではなく、それがなぜ類似していないのかの理由が分かるのは提案手法の特徴である。

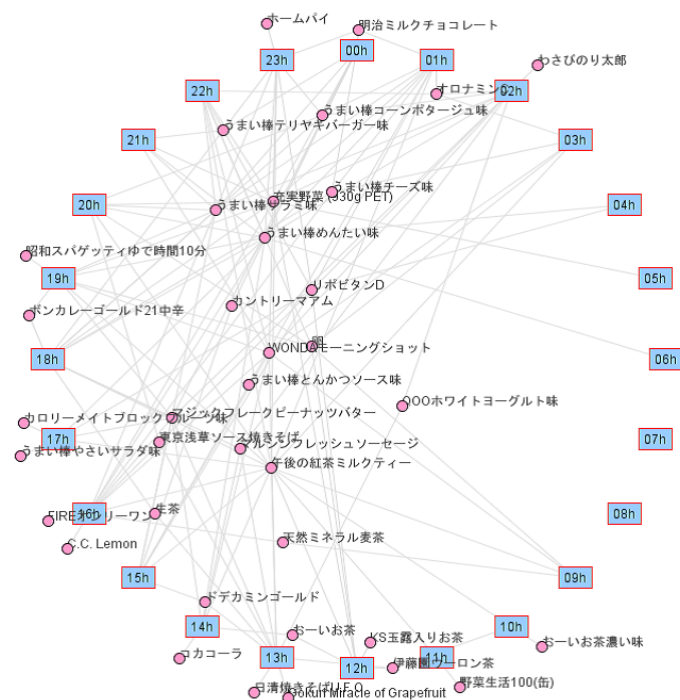


図 7.2(b) は等類似度線描画を閾値  $t = 50$  で描画した例である。中心左上に、ノードが集まっているクラスタが確認できる。アンカーマップの描画でも「うまい棒サラミ味、うまい棒めんたい味、充実野菜、うまい棒チーズ味は傾向が似ている」ということは認識できるが、本手法ではそれに加え、うまい棒と充実野菜はかなり傾向が似ており、次にうまい棒めんたい味の傾向が似ている、チーズ味は他 3 つに比較するとあまり似ていない、といった情報を読み取ることが出来る。

この程度の小さい図であればグラフの規模を小さくする必要がないため、空間構造的な可読性の向上は難しい。しかし、スライダによるグラフ構造の変更は意味的な可読性の向上に有効であることが確認できる。

ここで、注意しておきたいのが右上から左下まで大きく広がったクラスタである。これは 02h とのみ接続するわさびのり太郎と、02h と 13h に接続する QOO ホワイトヨーグルト味が 50% でクラスタリングされた結果であるが、それと同時に様々なノードがクラスタリングされ、巨大なクラスタを構成している。本手法で用いられる最短距離クラスタリングは空間濃縮という特性を持っており、この効果によりあまり似ていないと感じるノード同士で、巨大なクラスタが構成されてしまう。(空間濃縮とは、併合されてできた新しいクラスタは、以後の併合の対象として選ばれる可能性が加速度的に増加する現象である。) 本手法でグラフを読む際には、このことを念頭に置いて読むことが必要である。

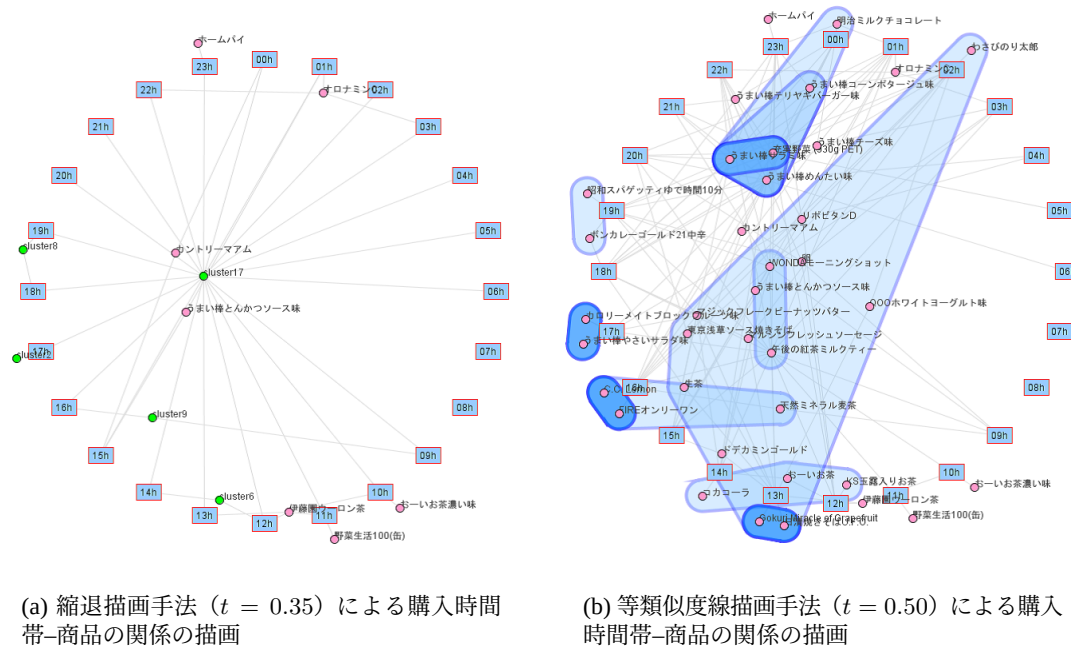


図 7.2: 提案手法による購入時間帯-商品の関係の描画

## 7.1.2 大規模購買データ

大規模データの可視化には、2006 年 6 月から 2007 年 5 月までの期間における商品と購入者の関係のデータを用いた。アンカーを 25 人の購入者、フリーノードを 377 種類の商品として、提案手法で描画する。エッジ数は 1382 本であり、スプリングの強さは購入商品数に比例して強くしている。

図 7.3 はアンカーマップによる描画である。アンカーの並びはエッジの総線長が短くなるように並び替えて、似た接続傾向を持つアンカーが近くなるように配置している。左下が密、右上が粗と大量消費者とそうでない者に分割され、レイアウト変更が有効に働いていることは確認できるが、ノード数がこのくらいの量になるとグラフの構造の把握が難しいことが分かる。

等類似度線描画を用いた描画を図 7.4 に示す。類似度の閾値は 100% に設定した。描画しているノード数は変化していないにもかかわらず、クラスタの分布が理解しやすくなっている。つまり、本手法により意味的な可読性が向上したといえる。特筆すべきは、左下の長方形のような形をした巨大なクラスタである。クラスタ内のノードのエッジを見てみると、1 つのアンカーとのみ接続している。すなわち、この購入者は他の購入者が買っていない商品をたくさん買っており、とてもユニークな購買傾向を持っていることが分かる。

図 7.5(a), 7.5(b) は、類似度の閾値をそれぞれ  $t = 100\%$ ,  $t = 50\%$  に設定した縮退描画を用いて描いたグラフである。 $t = 100\%$  では変化があまり見られないが、閾値を  $t = 50\%$  まで下げると、フリーノード数が 4 つ（内クラスタ 1 つ）となり、すっきりした構造が見える。ク

ラストをのぞいたノード3つは購買傾向が特殊であることがわかり，このうち2種類がコーヒーである．つまり，コーヒーは研究室内において万人に受ける商品ではないことが推測できる．2種類のコーヒーの接続先には3人の共通アンカーがあり，この3人はコーヒーが好き，もしくはコーヒーの銘柄に拘らないの人間ではないか，といったことが推測できる．また，先程のユニークな購入者は購入量も特に多いことが分かる．



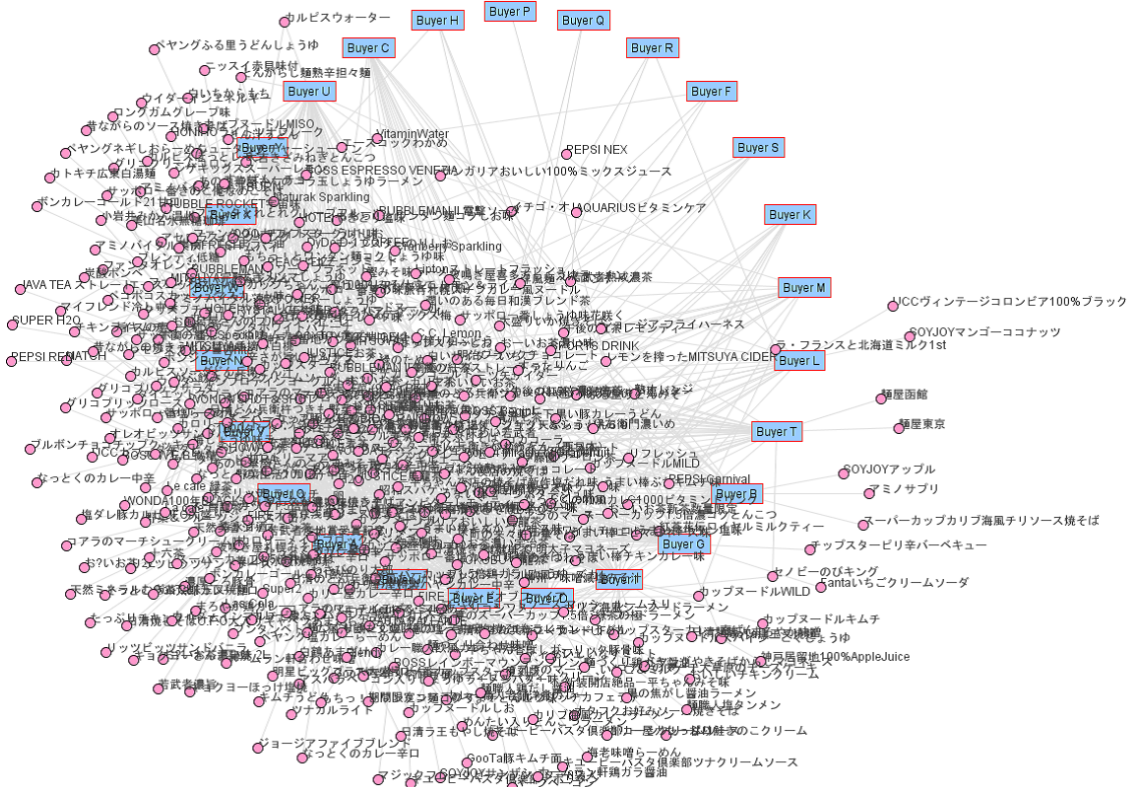


図 7.3: 購入者—商品のアンカーマップによる描画

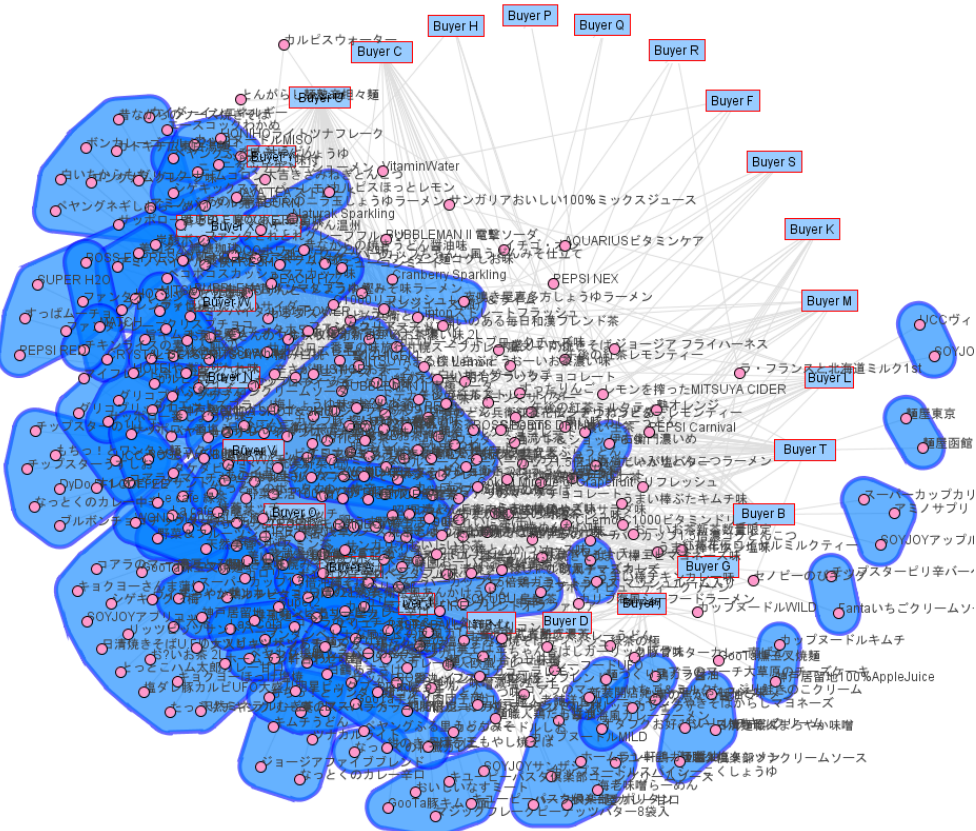
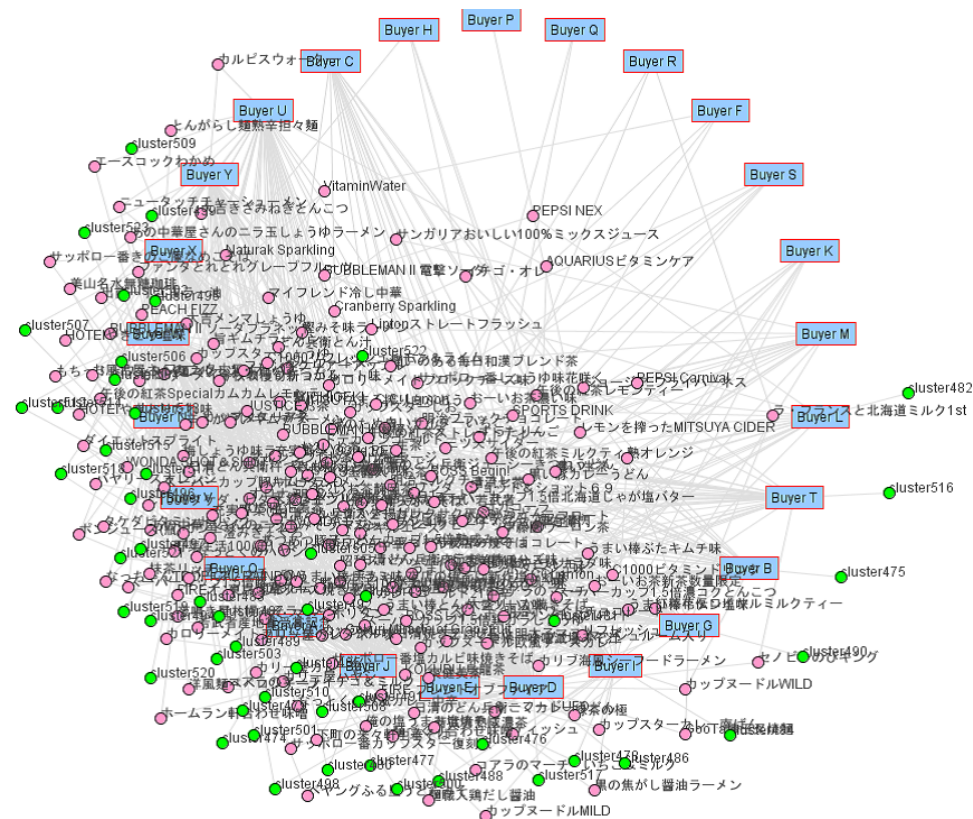
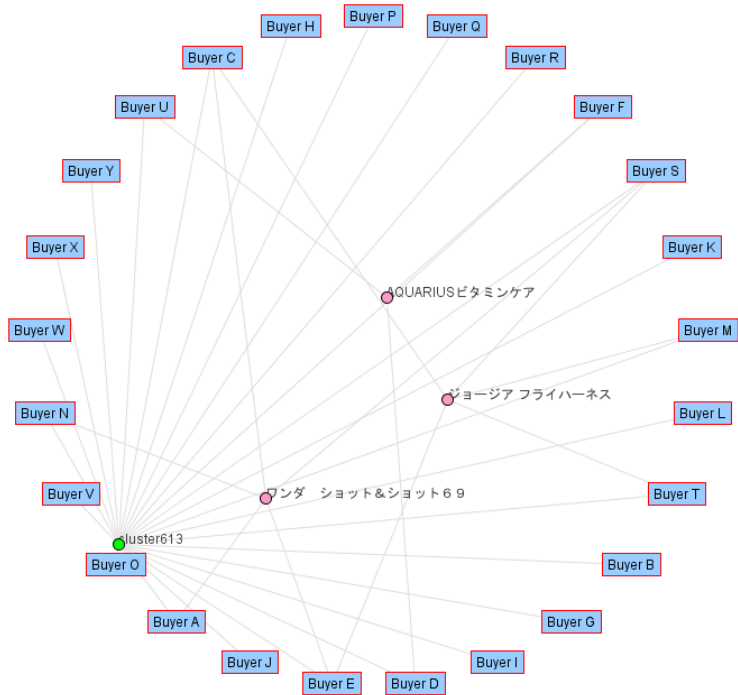


図 7.4: 等類似度線描画手法 ( $t = 100\%$ ) による購入者—商品の関係の描画



(a)  $t = 100\%$



(b)  $t = 47\%$

図 7.5: 縮退描画手法による購入者-商品の関係の描画



## 7.2 論文共著者情報の可視化

DBLP<sup>4</sup>より取得した論文の共著者関係の可視化を行った。取得したデータは、ある研究者から2ホップ離れた著者までと彼らの書いた論文の関係データである。ここでいう2ホップとは、著者間が2つの論文と一人の共著者で結ばれる距離のことを言う（図7.6）。

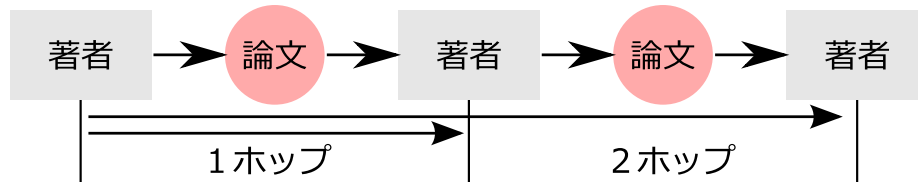


図 7.6: 2 ホップ

今回用いるグラフは、69 人の著者をアンカー、1891 個の論文をフリーノードとして、2222 本のエッジで接続を行った。また、このグラフもエッジの総線長を短くするように、アンカーの並べ替えを行っている。

ノードの量が多いため、通常のアଙ୍କーマップでは内容を把握することすら難しい（図7.7）。ここで、等類似度線描画を適用しノードラベルを消した状態で表示した（図7.8）ラベルを消してしまったため内容まで分からないが、グラフの構造としては購買関係の比較してとても単純なクラスタ構造をしていることを知見として得ることが出来る。

縮退描画表現手法で描画したのが図7.9である。スライダを101%から100%に操作した瞬間、ノード数が約150個に激減する。つまり、論文ノードは同じ接続関係を持っていることが多く、いつも同じメンバーで論文を書く傾向が高いことが分かる。また、スライダの値を50%位まで下げて行くと中心にいくつかの論文が残る。これらの論文は、いつもとは違うメンバーで書かれており、それぞれのコミュニティを接続する論文であると考えられる。

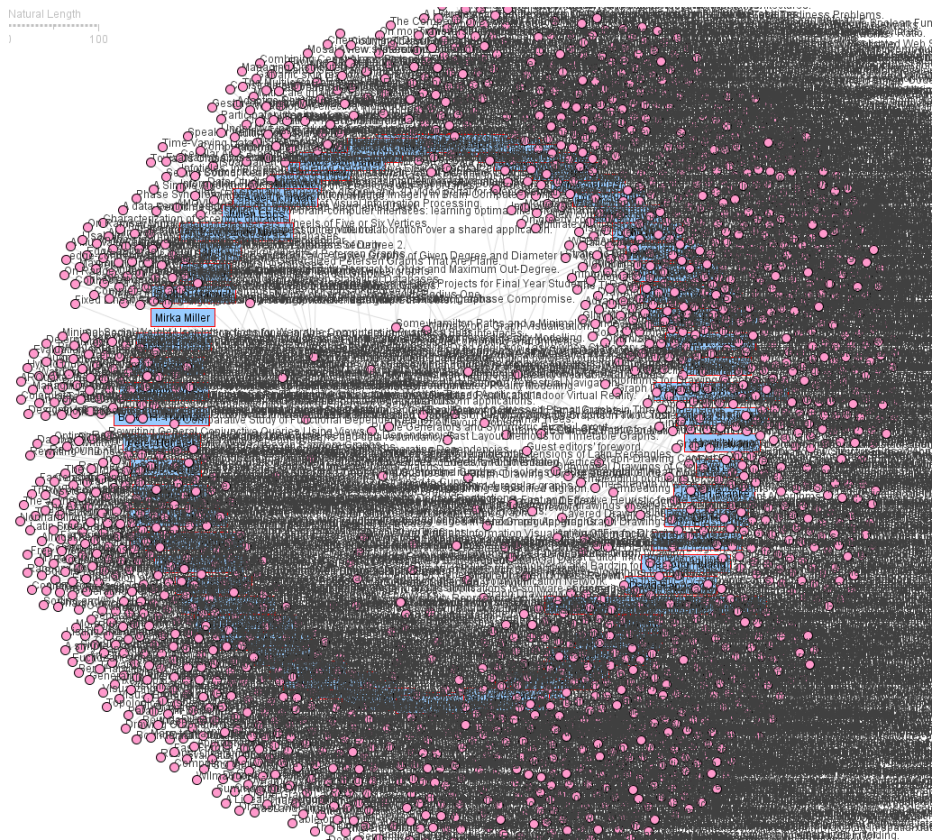


図 7.7: アンカーマップによる論文共著者情報の可視化

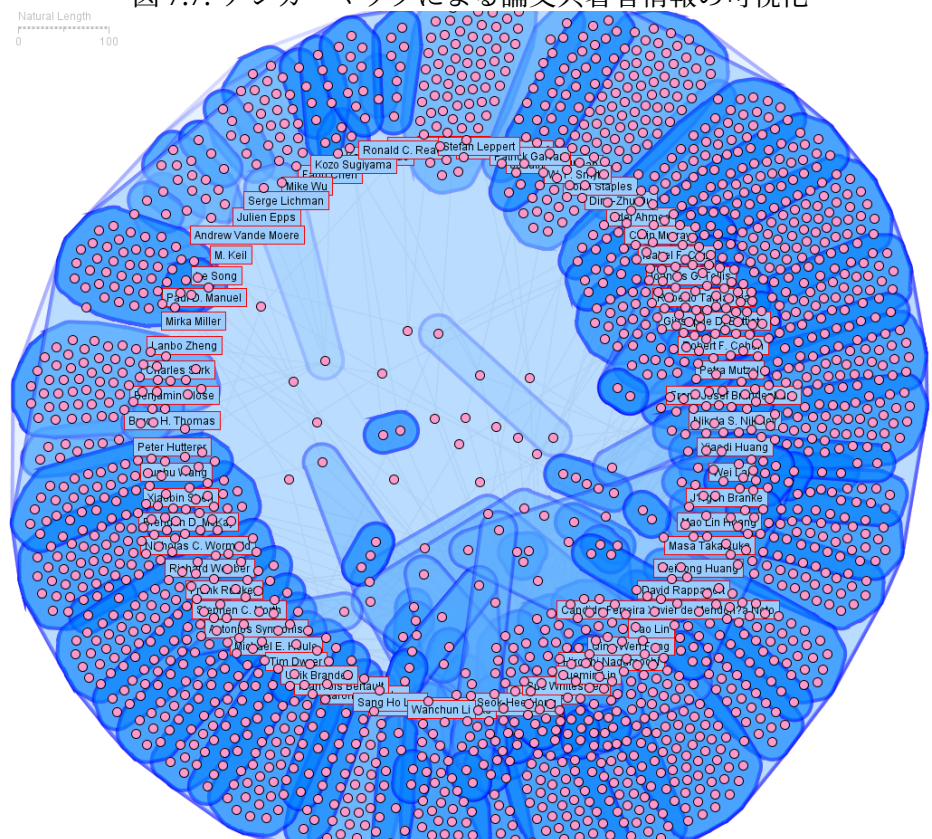
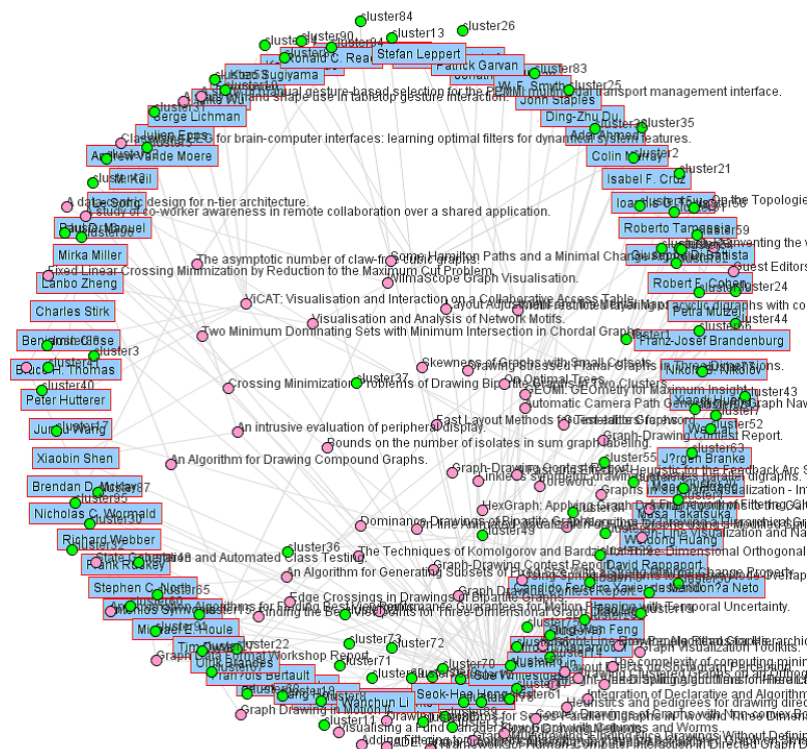
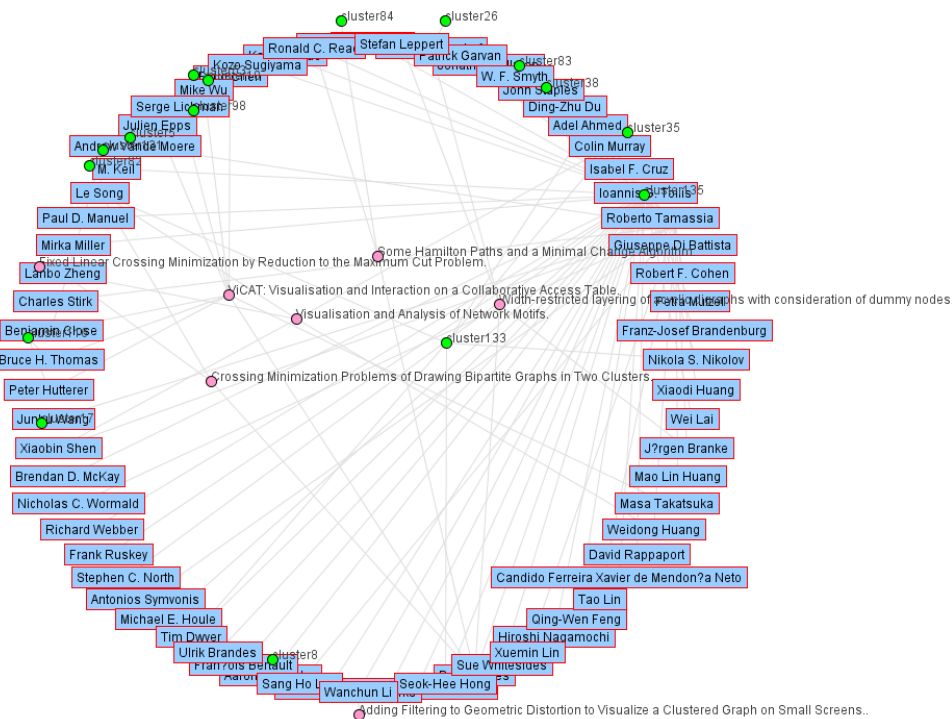


図 7.8: 等類似度線描画手法 ( $t = 0\%$ ) による論文共著者情報の可視化



(a)  $t = 100\%$



(b)  $t = 50\%$

図 7.9: 縮退描画手法による論文共著者情報の可視化



## 7.3 考察

### 7.3.1 縮退描画手法の有効性

縮退描画手法は主にネットワークの空間構造的な可読性を向上させる手法として導入したが、意味的な可読性を向上させる意味合いの方が高いことが、評価実験および実例への適用により明らかになった。ただし、ある閾値で縮退描画を行った断片的な画像からはこの効果は現れない。スライダを読み手が動かし、グラフを変化させる過程から意味的な可読性の向上が発現することが実証された。

また、この手法は共著者関係におけるコミュニティ間を接続する論文や、購買データにおける他と購買傾向が異なる商品などの発見に対して極めて有効であることがわかった。すなわち、特徴のあるノードの発見に効果を示す結果が実証された。

### 7.3.2 等類似度線描画の有効性

等類似度線描画は意味的な可読性の向上のために開発した手法であるが、狙い通り効果があることが評価実験により証明された。こちらは縮退描画手法と異なりノードが結合せず残っているため、ある一瞬の静止画でもノードの関連性が読みとることが可能であり、一覽性の性質が強い手法であると感じられた。

こちらの手法は主に高い閾値で効果が有効に働き、結びつきの強いノードを観察するのに適している。

### 7.3.3 グラフ読解の支援

グラフを可視化し読んで理解するという作業には、読み手の読解力がとても重要になる。しかし、日常的にグラフを読んでいる人はまれな存在であり、多くの人がグラフを読むことが不得手である。グラフの「どこを見ればいいのか分からない」ことが彼らにとっての問題であり、本手法はその部分に対応力が高いと感じた。

### 7.3.4 本手法の問題点

#### スライダ操作

スライダを左の向きに操作し結合するノードの類似度の閾値を下げていき、ある一定の値でノードが結合した際、スライダを右に一旦戻す動作をすると今集まったノードがどれかを確認することが可能となる。ただし、人間が能動的に確認作業をする必要があり、システム側がそれを可視化した方が良いとも考えられる。

また、スライダが掴みにくくホイール操作を推奨する意見もあった。類似度の閾値を変更するインタフェースとしてスライダの他に最適な方法があるかは検討する必要がある。

## 集約・展開の表現

同じ類似度で結合するクラスタが複数あった場合は、それぞれのクラスタの要素を確認することが難しい。例として、10 個のノードが2つのクラスタに結合した場合、片方のクラスタは10 個中どの要素が集約されたかは集約される前の位置により判断するしかない。そのため、クラスタの集約の過程の表現については検討の必要がある。

## 等類似度線内部のクラスタ外ノード

等類似度線描画において、ノードのレイアウトは変更していないため、等類似度線内にクラスタに含まれないノードが混入してしまう。アンカーマップのスプリングモデルのレイアウトを維持し、レイアウトにより発見できるクラスタと等類似度線により認知できるクラスタの違いを知りたいために現在の仕様にしたが、評価実験ではこの点に関して良い評価を得ることは出来なかった。

## 7.4 今後の課題

### 7.4.1 ラベリング

本手法では、クラスタのラベルは単純に通し番号を振りラベルとしてきた。しかし、この手法ではクラスタの中身が直接分からないのが難点である。かといって、クラスタ内部のノードすべてのラベルを羅列するのでは、ノードを集約して表示する意味が低下する。見易く意味も分かりやすいラベルは必要な機能と考える。

### 7.4.2 より大きなグラフへの対応

現在、クラスタリングツール GTBC が扱うことの出来るグラフの大きさの上限は8000 ノード程度であった。日常的に触れることの出来るネットワークの規模は日に日に大きくなっているため、より大きなグラフも扱えるようにする必要がある。

また、10 万ノードを超えるようになると、レイアウト計算処理が重くなりインタラクティブな操作に支障をきたすようになることが予想できる。レイアウトアルゴリズムの高速化も今後の課題である。

## 第8章 関連研究

### 8.1 グラフの自動描画

グラフの自動描画に関する研究は盛んに行われており、その多くが Eades のばねモデル (Spring Embedded Model) を利用している [10, 25, 24, 26]. Sun 社の Java 言語の Graph Layout プログラムもその普及に一躍買っており, Demo プログラムをウェブページからダウンロードすることが出来る.

### 8.2 知識の抽出

情報可視化の研究の主な目的は, 可視化したデータから知識を得ることである.

ManyEyes は可視化情報共有サイトについての研究であり, 可視化データの知識を共有することで, 集合知によりさらなる知識の抽出を支援するものである [27].

情報可視化のためのツールとしては, graphviz<sup>5</sup>や prefuse<sup>6</sup> などがある. prefuse はインタラクティブツール群で, 知識の取得と発見を支援するいくつかの表現手法を提供する [25]. これらのツールは可視化の導入を容易にするため, 利用して何らかの知識を得ようとする研究も多数存在する [28].

KeyGraph は文書内の主張を表すキーワードを抽出し, グラフに可視化することで新しい知見を得る試みである [24].

### 8.3 大規模グラフへの対応

大規模グラフを描画する手法としては, 3D 化や高解像度に関するアプローチが主流である.

Hyperbolic Tree は, 双曲空間上に樹形図を表現する focus + context 手法であり, 深い階層構造を持った樹形図を効率よく 1 画面に表現することができる表示方法である [29]. また HyperBolic Tree を 3D 化した手法として H3 があり, さらに多くの情報を 1 画面に納めることに成功している [30, 31].

Lehmann らの手法は, クラスタを形成している部分は近くに配置し, クラスタ同士は離してレイアウトする手法である. 加えて, 高精細高解像度で描画するのが本手法の特徴でもある [32].

Tominski らの手法は, 大規模グラフ描画を俯瞰する際にノードの集積度が高い部分は論理的な FishEyeView を用いて詳細情報を見られるようにする試みである [33].

Tseng らの手法は、それぞれのノードにランクをつけヒストグラム状にし、フォーカスを当てた山の部分だけを表示することで、読み手が見たい部分だけを描画する手法である [34].

伊藤らの手法は、アンカーマップを 3D 空間に拡張して表現領域を拡大し、多くの情報を 1 画面に表示できる手法である [35].

## 8.4 クラスタ構造の活用

クラスタの構造をネットワークの可視化に活用する手法としては、元々ネットワークがクラスタ構造を持っているグラフを可視化する手法 [37, 38, 44, 45, 46, 47, 48] とグラフにクラスタリングを施してそれを可視化に活用させる手法 [40, 41, 47, 54] に大きく分けられる.

階層型の木構造の表現の成功例としては Treemaps や Cone-Trees が挙げられる [36]. Cone-Tree は 2 次元では書ききれないようなデータを 3 次元の奥行き効果を利用することで表示領域を効果的に広げている [37]. Spiral Tree は Cone-Tree の拡張であり、同一階層内のノード同士の重なる減少に成功している [38]. FractalView は Cone-Tree のような 3 次元描画に、FishEye のような Focus+Context の手法を取り入れたものである [39].

クラスタリングを事前処理として行い可視化結果に反映させる手法として、小林らの研究がある [40]. Web における情報検索の可視化を行うため、クラスタリングを事前処理として行い可視化結果に反映させる手法を用いている.

Huang らの手法は、クラスタによりグラフをいくつか分割する手法をとっており、各々のグラフは小さく見易くなるため、構造の把握に有効であるとしている [41].

クラスタリングの情報を描画する手法として、井上らの研究がある [42]. クラスタリングの結果を可視化するため、樹状図を描画レベルと注視するノードを変更可能にしたネットワーク図で描画する手法を用いている.

「データ宝石箱」「平安京ビュー」は長方形の入れ子型表現手法を用いて空間効率の良い描画を提供する [43, 44].

Frishman らの研究は、クラスタ構造を持つグラフを二次元もしくは三次元平面上に描画し、グラフの情報の動的変化に対応した自動レイアウトを行うことが特徴である [45, 46].

H-BLOB はクラスタを球状の包含関係で描画する手法で、クラスタ内部に含まれるノードの若干の類似度の変化を「くびれ」によって表現することが特徴である [47].

表らの手法は、クラスタグラフを描画でき、なおかつインターセクションの扱いが可能なことが特徴である [48].

## 8.5 可読性向上の追求

ネットワークが動的に変化する「動的ネットワーク」に対応し、これの可読性を向上させる研究がある [26, 28, 49, 50, 51].

グラフに何らかの付加情報を与える研究として、石原らの手法は 2 つの 2 部グラフ間の差分変化に注目し、変化が発生したノードに三角印のアノテーションを与えることで可読性の

向上をねらっている [26].

複数の表現を並列に描画して閲覧することで、可読性を向上させる試みも行われている. MatLink, MatrixExplorer は行列表現を取り入れることで、クラスタの発見等に効果があるという報告を行っている [52, 53]. Abello らは Treemap を並列に描画し、階層構造をもったグラフの可読性向上を狙っている [54].

インタラクションにより可読性を向上させる試みに関する研究もなされている. 「納豆ビュー」は対話性に主眼をおいており、ノードの「持ち上げ操作」により関連性の有るノードをわかりやすくする手法を提供する [55].

ノードを減らすのではなく、エッジを効率的に減らすことで可読性を向上させる試みも存在する. Newbery らによる手法は、edge concentration node と呼ばれるエッジをまとめるノードをグラフ内に導入することで、エッジの交差数を減少させることに成功している [56]. Eppstein らによる手法は、複数エッジをなめらかな曲線によりまとめる手法であり、なめらかな交差ではないエッジ同士は交差、なめらかな交差をしているエッジは接続と見なし、エッジ数を減らしている [57].

Yamada らは、既存のスプリングモデルを改良し大規模化したときにノードを見易くレイアウトする手法を提案している [58].

本研究は大規模 2 部グラフを対象とし、クラスタ構造とインタラクティブな操作を活用した手法であることが、従来手法と異なる. また、読み取りやすさに焦点を当てているという点、敢えて要素を増やすことで可読性の向上を狙うという点も従来手法とは違う特徴である. 図 8.1 に本研究の位置付けを示す.



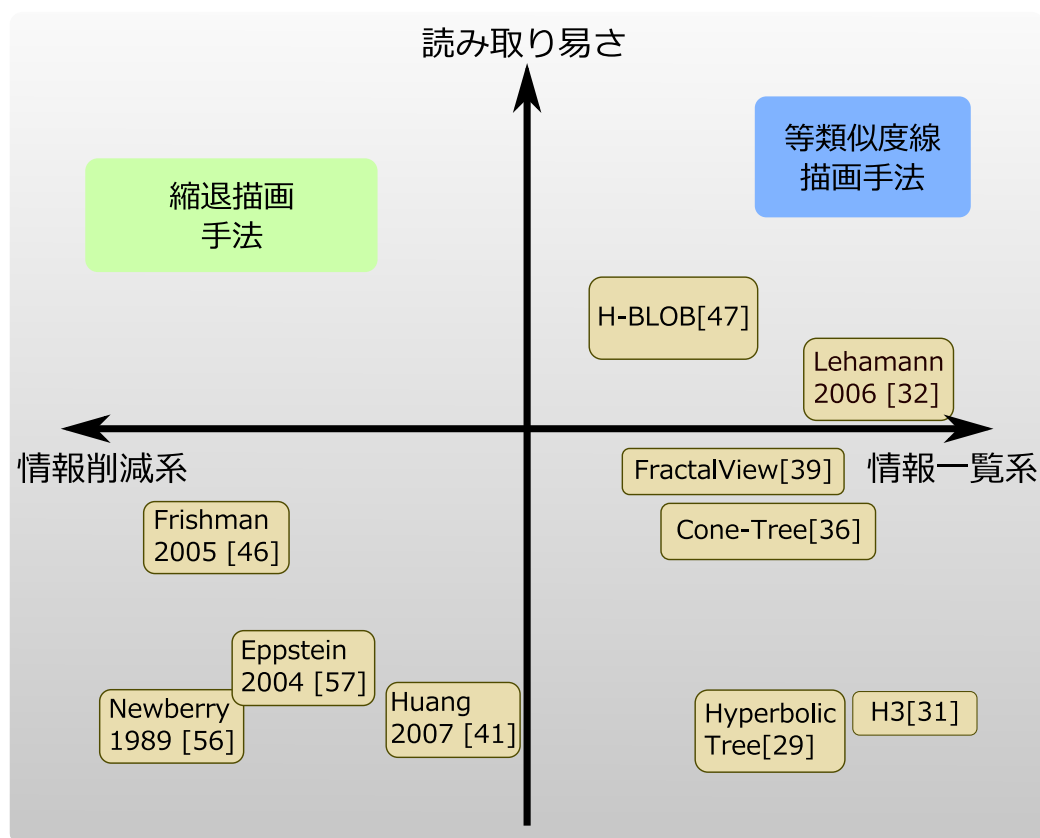


図 8.1: 本研究の位置付け

## 第9章 結論

大規模2部グラフの可読性を向上させるために、クラスタリングを取り入れた描画手法の開発を行い、その有効性の評価を行った。可読性の二つの重要な要素である「見易さ」と「読み取りやすさ」を向上させる手法として、「縮退描画手法」と「等類似度線描画手法」を開発した。クラスタリングによりグラフ構造をコンピュータ側で解析し、ノード間の関連度を視覚化することで読み手の理解を高める。

評価実験より、縮退描画手法と等類似度線描画手法は従来手法と比較して、読み手のグラフの把握を高める効果が示された。また、読み手の操作に対応したインタラクティブなグラフの推移により縮退描画手法が特徴的なノードを、等類似度線描画手法が特徴が似ているノードの集合を認知しやすくさせることが示された。

本手法を用いることで、従来手法では困難であった、大規模2部グラフの全体的な俯瞰と特定の注視要素の関係情報が一度に読み取ることが可能となった。最後に本手法を様々なグラフに適用し、本手法でしか読み取ることが難しい効能の詳細を示した。

## 謝辞

本研究にあたり、御指導を頂きました指導教員の田中二郎教授には心より感謝いたします。研究の心得から論文の執筆に至るまで、丁寧な御助言と御指導を頂きました。研究室内の各種設備の支援や学外の研究発表の機会を与えて下さり、博士前期課程の2年間は有意義で忘れられないものとなりました。最大限の感謝の意を添えて、ここに厚く御礼を申し上げます。

副指導教員の三末和男准教授には研究方法の初歩から、方向性、論文の執筆に至るまで懇切丁寧な御指導を頂き感謝いたします。的確な助言の御陰で、沢山の迷惑をかけながら迷いながらも出口に来ることが出来ました。心から深く感謝致します。

志築文太郎講師、高橋伸講師には、研究発表についてひとかたならぬ御指導をいただきました。心より感謝致します。

NAIS チームの同級生と情報可視化チームの面々、OBの方々には大変お世話になりました。酒井慎司君には研究内容についての相談にとどまらず、論文執筆等に必要な図の指導を頂きお世話になりました。鈴木優君には研究の組み立てやプログラムに関して丁寧な相談をして頂きました。中園長新君には研究室内の計算機等の整備について丁寧な仕事をしてくださりお世話になりました。小池諭君には研究についてのアイデアを多数頂きました。石原正樹氏には可視化研究の基礎と研究者としての在り方について丁寧な御指導を頂きました。ここに感謝いたします。

インタラクティブプログラミング研究室の方々には、研究活動のみならず私生活に至るまで大変お世話になりました。貴重な時間を割いて評価実験に快く参加していただいた皆様には、この場を借りて御礼を申し上げます。特に NAIS チームのメンバーは活発な議論や私生活の交流を経て心許せる大切な仲間となりました。皆様方のさらなるご活躍をお祈りするとともに、感謝の言葉を申し上げます。

また、この素晴らしい学生生活を送らせて頂き、特に資金面で支えてくださった家族には心よりの感謝の意を贈らせていただきます。

最後に、ここに書ききれなかった方も含め、学生生活並びに私生活の中でお世話になったすべての方々にあらためて深く感謝いたします。

## 参考文献

- [1] Stanley Milgram. “The Small World Problem,” *Psychology Today*, pp. 60–67, 1967.
- [2] 安田 雪, 松尾 豊. “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.
- [3] ダンカン・ワッツ, (訳: 辻竜平, 友和政樹). “スモールワールド・ネットワーク—世界を知るための新科学的思考法,” *阪急コミュニケーションズ*, 2004.
- [4] 石田和成. “潜在的 Weblog コミュニティ抽出のための二部グラフ分割アルゴリズム,” *人工知能学会研究会*, SIG-SWO-A404-01, 2004.
- [5] 岸野文郎, 大野義夫, 藤代一成, 北村喜文. 情報の可視化 (岩波講座マルチメディア情報学6), 岩波書店, 2001.
- [6] 土橋喜. “情報視覚化と問題発見支援—問題構造の可視化による仮説生成,” *あるむ*, 2000.
- [7] 田村 博. *ヒューマン・インタフェース*. コロナ社, 1987.
- [8] 杉山 公造. “グラフ自動描画法とその応用,” *計測自動制御学会*, 1993.
- [9] Mohammad Ghoniem, Jean-Daniel Fekete, Philippe Castagliola. “A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations,” In *Proceedings of IEEE Symposium on Information Visualization 2004(InfoVis2004)*, pp.17–24, 2004.
- [10] Jeffrey Heer, Danah Boyd. “Vizster: Visualizing Online Social Networks,” In *Proceedings of IEEE symposium on Information visualization (INFOVIS '05)*, pp. 5-12, 2005.
- [11] 市瀬龍太郎, 武田英明, 植山浩介. “コミュニティマイニングのための研究者情報の視覚化,” *信学技報*, Vol. 104, No. 587, pp. 1–6, 2005.
- [12] M. E. J. Newman. “Coauthorship networks and patterns of scientific collaboration,” In *Proceedings of the National Academy of Sciences of the USA*, Vol.101, No. suppl.1, pp. 5200–5205, 2004.
- [13] 芳鐘冬樹, 影浦峽. “共著ネットワークの分析: 視点と手法,” 第 49 回日本図書館情報学会研究大会発表要綱, p. 55-58, 2001.

- [14] Peter Eades. “A heuristic for graph drawing,” *Congressus Numerantium*, Vol.42, pp.149-160, 1984.
- [15] Kazuo Misue. “Drawing Bipartite Graphs as Anchored Maps,” In *Proceedings of Asia Pacific Symposium on Information Visualization 2006 (APVIS2006)*, pp.169–177, 2006.
- [16] Kazuo Misue. “Anchored Maps: Visualization Techniques for Drawing Bipartite Graphs,” In *Proceedings of 12th International Conference on Human-Computer Interaction (HCI International 2007)*, pp. 106-114, 2007.
- [17] Weidong Huang, Seok-Hee Hong<sup>1</sup>, Peter Eades. “Layout Effects on Sociogram Perception,” In *Proceedings of Graph Drawing '06*, pp.262-273, 2006.
- [18] John A. Hartigan. クラスター分析, マイクロソフトウェア, 1983.
- [19] 宮本貞明. クラスター分析入門, 森北出版, 1999.
- [20] 神嶌 敏弘. “データマイニング分野のクラスタリング手法 (1) – クラスタリングを使ってみよう！ –, ” 人工知能学会誌, vol.18, no.1, pp.59–65, 2003.
- [21] 佐藤修治, 三末和男, 田中二郎. “大規模2部グラフの可読性向上のためのクラスタ構造の動的描画,” 第15回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2007), 日本ソフトウェア科学会, pp.19-24, 2007.
- [22] Ulrik Brandes, Christian Pich. “GraphML Transformation.” In *Proceedings of 12th International Symposium Graph Drawing (GD '04)*, LNCS 3383, pp. 89-99, 2004.
- [23] Ronald L. Graham, “An efficient algorithm for determining the convex hull of a finite planar set,” *Information Processing Letters*, vol.1, 132–133, 1972.
- [24] 大澤幸生, 谷内田正彦. “キーワード抽出法 KeyGraph の転用による地震履歴データからの要注意活断層発見支援 (発見科学), ” 人工知能学会誌, Vol.15, No.4, pp.665–672, 2000.
- [25] Jeffrey Heer, Stuart K.Card, James A. Landay. “Prefuse: a toolkit for Interactive Information Visualization.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2005*, pp.421–430, 2005.
- [26] 石原正樹, “動的ネットワークの成長過程と差分の可視化手法, ” 筑波大学大学院コンピュータサイエンス専攻修士論文, 2007.
- [27] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, Matt McKeon. “Many Eyes: A Site for Visualization at Internet Scale,” *Transactions on Visualization and Computer Graphics*, vol.13, No.6, pp.1121–1128, 2007.

- [28] 鈴木 祐太, 古川園 智樹, 青山 希, 井庭 崇, “動的ネットワークの可視化ツールの構築,” MPS, 数理モデル化と問題解決研究報告, Vol.2006 No.29, pp.81–84, 2005.
- [29] John Lamping, Ramana Rao. “The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies,” *Journal of Visual Languages and Computing*, Volume 7, Issue 1, Pages 33-55, 1996.
- [30] Tamara Munzner. “H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space,” In *Proceedings of the 1997 IEEE Symposium on Information Visualization*, pp.2–10, 1997.
- [31] Tamara Munzner. “Drawing Large Graphs with H3Viewer and Site Manager,” In *Proceedings of Graph Drawing '98*, pp.384-393, 1998.
- [32] Katharina A. Lehmann and Stephan Kottler “Visualizing Large and Clustered Networks,” *Graph Drawing*, pp. 240-251, 2006.
- [33] Christian Tominski, James Abello, Frank van Ham, Heidrun Schumann. “Fisheye Tree Views and Lenses for Graph Visualization,” 10th International Conference Information Visualization (IV06), pp.17–24, 2006.
- [34] Belle L. Tseng, Junichi Tatemura, Yi Wu. “Tomographic Clustering To Visualize Blog Communities as Mountain Views,” The 14th International World Wide Web Conference (WWW2005), Workshop on the Weblogging Ecosystem, 2005.
- [35] 伊藤隆朗, 三末和男, 田中二郎. “Sphere Anchored Map : 大規模 2 部グラフの 3D 描画手法,” pp.159–160 ,2007.
- [36] Brian Johnson, Ben Shneiderman. “Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures,” In *Proceedings of the 2nd International IEEE Visualization Conference*, Pages 284-291, 1991.
- [37] George Robertson, Jock D. Mackinlay, Stuart K. Card. “Cone Trees: Animated 3D Visualizations of Hierarchical Information,” In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (SIGCHI'91)*, pp189–194, 1991.
- [38] 大隈 隆史, 竹村 治雄, 片山 喜章, 岩佐 英彦, 横矢 直和. “三次元視覚化手法 Spiral Tree による情報のブラウジング支援,” 第 40 回システム制御情報学会研究発表講演会講演論文集, No. 2011, 1996.
- [39] Hideki Koike, Hirotaka Yoshihara. “Fractal Approaches for Visualizing Huge Hierarchies”, In *Proceedings of the 1993 IEEE Symposium on Visual Languages (VL'93)*, pp.55-60, 1993.
- [40] Takumi Kobayashi, Kazuo Misue, Buntarou Shizuki, Jiro Tanaka. “Information Gathering Support Interface by the Overview Presentation of Web Search Results,” In *Proceedings of Asia Pacific Symposium on Information Visualization 2006 (APVIS2006)*, pp.103–108, 2006.

- [41] Mao Lin Huang and Quang Vinh Nguyen. “A Fast Algorithm for Balanced Graph Clustering,” 11th International Conference Information Visualization (IV07), pp.46–52, 2007.
- [42] 井上悦子, 吉廣卓哉, 中川優. “大規模クラスタリング結果のグラフによるインタラクティブな可視化手法,” 情報処理学会 情報学基礎研究会報告, 2006-FI-85-(4), pp. 21–28, 2006.
- [43] 山口裕美, 伊藤貴之, 梶永泰正, 池端裕子 “階層型データ視覚化手法「データ宝石箱」とウェブサイトの視覚化,” 画像電子学会論文誌 VisualComputing 特集号, Vol. 32, No. 4, pp. 407-417, 2003.
- [44] 伊藤貴之, 小山田耕二. “平安京ビュー ～階層型データを基盤状に配置する視覚化手法,” 可視化情報学会第9回ビジュアリゼーションカンファレンス, 2003.
- [45] Yaniv Frishman, Ayellet Tal. “Dynamic Drawing of Clustered Graphs,” In *Proceedings of IEEE Symposium on Information Visualization(InfoVis)*, pp.191–198, 2004.
- [46] Yaniv Frishman, Ayellet Tal. “Visualization of Mobile Object Environments,” In *Proceedings of the 2005 ACM symposium on Software visualization*, pp.145–154, 2005.
- [47] T. C. Sprenger, R. Brunella, M. H. Gross. “H-BLOB A Hierarchical Visual Clustering Method Using Implicit Surfaces,” 11th annual IEEE Visualization Conference(Vis2000), pp.61-68, 2000.
- [48] 表寛樹, “インターセクションの扱いが可能なクラスターグラフの力指向による自動描画法の研究,” 北陸先端科学技術大学知識科学研究科知識システム基礎学専攻修士論文, 2000.
- [49] 長谷川 幹根, 石川 佳治. “T-Scroll : 時間的トピックの推移をとらえる可視化システム,” 電子情報通信学会第18回データ工学ワークショップ (DEWS2007), C8-1, 2007
- [50] 豊田正史, 喜連川優. “WebRelievo: ウェブにおけるリンク構造の発展過程解析システム,” 第12回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2004), pp.89–94, 2004.
- [51] N. Nakazono, K. Misue, J. Tanaka, “NeL2: Network Drawing Tool for Handling Layered Structured Network Diagram,” In *Proceedings of Asia Pacific Symposium on Information Visualization 2006 (APVIS2006)*, pp.109-115, 2006.
- [52] Nathalie Henry, Jean-Daniel Fekete. “MatrixExplorer: a dual-representation system,” In *Proceedings IEEE Symposium on Information Visualization(InfoVis2006)*, pp.677-684, 2006.
- [53] Nathalie Henry, Jean-Daniel Fekete, Michael J. McGuffin. “NodeTrix: a Hybrid Visualization,” In *Proceedings IEEE Symposium on Information Visualization(InfoVis2007)*, pp.1302–1309, 2007.

- [54] James Abello, Stephen G. Kobourov, and Roman Yusuf. “Visualizing Large Graphs with Compound-Fisheye View and Treemaps,” In *Proceeding of Graph Drawing*, pp. 431-441, 2004.
- [55] 塩澤秀和, 西山晴彦, 相馬隆宏, 松下温. “情報の関連性と他人数アクセスに着目した WWW 空間の視覚化,” 情報処理学会グループウェア研究会研究報告 96-GW-18, pp.61-66, 1996.
- [56] Frances J. Newbery. “Edge Concentration: A Method for Clustering Directed Graphs.” In *Proceedings of 2nd International Workshop on Software configuration management*, pp.76–85, 1989.
- [57] David Eppstein, Michael T. Goodrich, Jeremy Yu Meng. “Confluent Layered Drawings,” In 12th International Symposium on GraphDrawing(GD’04), pp.184-194, 2004.
- [58] Takeshi Yamada, Kazumi Saito, Naonori Ueda. “Cross-Entropy Directed Embedding of Network Data.” In *Proceedings of International Conference on Machine Learning*, pp.832–839, 2003.



## 後注

<sup>1</sup>東京メトロ路線図. <http://www.tokyometro.jp/index.html>

<sup>2</sup>JAVA. <http://java.sun.com/>

<sup>3</sup>GraphML. <http://graphml.graphdrawing.org/>

<sup>4</sup>DBLP. <http://www.informatik.uni-trier.de/~ley/db/>

<sup>5</sup>graphviz. <http://www.graphviz.org/>

<sup>6</sup>prefuse. <http://prefuse.org/>