

千葉大学大学院工学研究科  
修士論文

力学モデルを用いた特定のノードを中心に  
据えるグラフ描画手法の提案

平成29年3月 提出

建築・都市科学専攻 都市環境システムコース

竹之内 洋樹

## ABSTRACT

In recent years, with the progress of information processing technology, analysis of networks generated from large-scale data has been actively conducted, networks of roads and transportation networks, networks between web sites to which the Internet is linked, networks of human relationships of organizations etc. Accordingly, there is a strong demand for a technique for modeling and expressing the structure and characteristics of the network in a visually easy-to-understand manner. One of these techniques is graph drawing.

This method is mainly oriented toward the layout of the entire graph, and there are many accumulations in research to satisfy them. On the other hand, many analytical methods targeting specific nodes and edges are known for network analysis. In order to deal with these analytical tasks, it is requested to change the layout based on specific nodes and edges.

One of the most popular graph drawing methods is force-directed model. This model constructed a virtual dynamic system in which every two nodes are connected by a spring, then obtained the state in which the total spring energy of the system is minimal. We propose that graph drawing method to laying into the center for a specific node by extending Kamada-Kawai, which is one of a force-directed method. Specifically, adjusting attracting force and repulsive force acting between two nodes by introduces a parameter to objective function used for Kamada-Kawai methods. In this paper, we confirmed the proposed method is an effective by drawing on graph with a simple structure and complicated.

## はじめに

近年、情報処理技術の進歩により、道路や交通網のネットワーク、インターネットが結びつけるWebサイト間のネットワーク、組織の人間関係のネットワークなど、大規模なデータから生成されたネットワークの分析が盛んに行われている。それに伴い、ネットワークの構造や特徴を視覚的に分かりやすくモデル化して表現する技術が強く要請されている。この技術の一つがグラフ描画である。

グラフ描画には、これまでに力学モデルと呼ばれる手法に関する多数の研究がなされている。力学モデルはノード同士が引力および斥力を及ぼし合う仮想的な物理系をグラフから構築し、その系の平衡状態を計算するものである。この手法は、グラフ全体のレイアウトを志向したものが主流であり、それらを満足するための研究には多くの蓄積がある。一方で、ネットワークの分析には特定のノードやエッジを対象とした分析手法が多数知られている。それらの分析課題に対応するには、特定のノードやエッジを基準にレイアウトを変化させることが要請される。

本研究では、既存の力学モデルによるグラフ描画手法を拡張し、特定のノードを中心に据えたレイアウトを実現するグラフ描画手法を提案する。その具体的な方法として、代表的な力学モデルであるkamada-kawai法に用いられている目的関数にパラメータを導入し、任意のノードに作用する引力および斥力を調整することでレイアウトを変化させた。

本論文は、提案した手法により、レイアウトが容易に想像できる簡単な構造のグラフや一般的な構造のグラフにおいて描画仮説を満足することを確かめるとともに、具体的なネットワーク分析課題への応用を目指し、その成果と課題について示したものである。第1章では、現在の情報処理技術課題におけるグラフ描画問題の研究背景と本研究の位置づけについて述べた。第2章では、本研究で扱うグラフ描画問題の位置づけについて述べるとともに、代表的な力学モデルについて紹介する。第3章では、具体的な問題設定と、その解法であるStress Majorizationについて詳述し、グラフを描画した結果を示した。第4章では、特定のノードを中心に据えるための描画仮説と描画仮説に基づいたKK法の拡張手法について提案するとともに、提案手法により実際にグラフを描画した結果、レイアウトの変化が描画仮説を満足することを確かめた。第5章では、具体的なネットワーク分析への応用を目指し、語と語の共起関係を示すネットワークの可視化に提案手法を適用し、分析への有用性について述べた。第6章では、本研究の成果と課題について整理した。

# 目次

第1章	序論	1
1.1	研究の背景	2
1.1.1	グラフの自動描画	2
1.1.2	ネットワーク分析	2
1.2	研究の目的	3
第2章	関連研究	4
2.1	本研究の位置づけ	5
2.1.1	描画対象のグラフ	5
2.1.2	グラフの自動描画問題の評価観点	5
2.2	力学モデルによるグラフ描画手法	6
2.2.1	Fruchterman-Reingoldの手法	6
2.2.2	Kamada-Kawaiの手法	7
2.2.3	FR法とKK法による描画結果	8
2.3	関連研究と本研究の位置づけ	9
2.3.1	計算の高速化を目指した研究	9
2.3.2	レイアウトを変化させる研究	9
第3章	グラフ描画手法の定式化と解法	11
3.1	問題設定と定式化	12
3.1.1	基礎知識	12
3.1.2	グラフ描画の考え方	13
3.1.3	最小化問題への定式化	13
3.2	最小化問題の解法	14
3.2.1	Stress Majorization	14

3.2.2	目的関数の収束性について . . . . .	16
3.3	描画結果と考察 . . . . .	16
3.3.1	対象グラフ . . . . .	16
3.3.2	描画結果 . . . . .	17
3.3.3	レイアウトのゆがみに関する考察 . . . . .	18
第4章	特定のノードを中心に据えたグラフ描画手法	19
4.1	描画仮説と拡張方法 . . . . .	20
4.1.1	描画仮説 . . . . .	20
4.1.2	拡張方法 . . . . .	21
4.1.3	予備実験 . . . . .	22
4.2	描画結果と考察 . . . . .	26
4.2.1	対象グラフ . . . . .	26
4.2.2	描画結果 . . . . .	26
4.2.3	ノードの中心性指標によるレイアウトへの影響に関する考察 . . . . .	29
第5章	ネットワーク分析への応用	31
5.1	テキストマイニングと共起ネットワーク . . . . .	32
5.1.1	テキストマイニング . . . . .	32
5.1.2	共起ネットワーク . . . . .	32
5.2	描画結果と考察 . . . . .	33
5.2.1	対象ネットワーク . . . . .	33
5.2.2	描画結果 . . . . .	33
5.2.3	分析への応用性に関する考察 . . . . .	34
第6章	結論	37
6.1	研究のまとめ . . . . .	38
6.2	課題と今後の展望 . . . . .	39
参考文献		40

# 第1章

## 序論

本章では，本研究の背景と目的を述べる．グラフの自動描画問題におけるテーマの経年的な変遷から，現在のグラフ描画問題の研究が，基礎的な理論の研究から分析への応用に関心を移していることと，近年，ネットワーク分析の対象領域が拡大し，ネットワークの構造を可視化する技術が要請されていることを述べる．また，背景を踏まえ，本研究がネットワーク分析への応用を目指した研究であることを述べる．

## 1.1 研究の背景

本研究の背景を、(1)グラフの自動描画、(2)ネットワーク分析、の二つのキーワードから述べる。

### 1.1.1 グラフの自動描画

「グラフ」とは、あるシステムを構成する要素を点（ノード）とし、それらの接続の様子を辺（エッジ）の集合で表した数学的概念である。会社の組織図や作業の工程図など、我々は日常生活の中でグラフを図にして分かりやすく表現することがある。それらのレイアウトはほとんどの場合、手動で行われているが、構成するデータが大量に存在すれば、人の手によってグラフの「良い」レイアウトを実現することは困難となる。そのため、計算機などを用いて可読性や美的基準といった人間の認知に関する指標を満足するグラフのレイアウトを表現する方法が要請されている。これをグラフの自動描画(Graph Drawing)と呼ぶ [10]。

グラフの自動描画を対象とした研究は、一つの学問的領域を形成しており、現在も様々な研究がなされている。初期のグラフの自動描画研究では、数学の分野としてのグラフに関する諸性質を基に、描画規則や美的基準を設定するなどの基礎的な研究が主流であったが、1980年代に入ると、コンピュータのグラフィック性能や計算機能が向上したため、実際にグラフを描画させる実証的な研究が盛んになされた。現在のグラフ描画の基礎にある「力学モデル」もこの時期に登場している [1, 2, 3]。現在では、一般家庭用のコンピュータの性能も非常に優れたものになり、グラフ描画アプリケーションも簡単に手に入るようになった [4, 5, 6]。また、現在では情報通信技術の発達により、大規模な容量のデータがインターネットを通じてリアルタイムに流通するようになった。そのため、近年のグラフ描画研究は、美的基準を満足するための描画制約の解消法などの基礎的な問題から、大量なグラフデータを高速に処理する手法や、動的に構造が変化するグラフをリアルタイムに描画するためのアルゴリズムなどの提案などに関心が移っている [7]。また、この背景にはネットワーク分析の対象領域の拡大が挙げられる。

### 1.1.2 ネットワーク分析

道路や交通網のネットワーク、インターネットが結びつけるWebサイト間のネットワーク、自然界における生物間の関係、人体中の遺伝子やたんぱく質の関係など、これらは

「ネットワーク」と呼ばれ、グラフ中のノードやエッジに、大きさや長さ、コストなどの数値を与えたものである。ネットワーク分析とは、これらの様々な対象の構造的な特徴を探る研究方法である。

ネットワーク分析はこれまで、人間関係や集団間の関係を扱う社会学，人類学，心理学などの人文社会科学，情報科学やオペレーションズ・リサーチなどの自然科学の両方において用いられている。また近年の情報処理技術の進歩により，これまで対象としてこなかった大規模な構造のネットワークを分析する手法が提案されている [8]。それに伴い，ネットワークの構造や特徴を視覚的に分析する，情報可視化の技術が強く要請されている。グラフの自動描画はその技術の一つであり，グラフ自動描画の技術的発展はネットワーク分析を助けるものとなる。

ネットワーク分析に寄与するグラフ描画手法として，ノードにラベルを付与して表示する，重要なノードを拡大して表示するなどの手法があり，既存のネットワーク分析アプリケーションにも搭載されている [4, 5, 6]。一方で，レイアウトを変化させるアプローチでは，全体のレイアウトを変化させてネットワークの構造の把握やクラスターの発見を行う分析に対応しているが，ネットワーク分析には，特定のノードやエッジを対象とした分析手法が多数知られている [9]。それらの分析課題に対応するには，特定のノードやエッジを基準にレイアウトを変化させることが要請される。

## 1.2 研究の目的

以上の背景を踏まえ，本研究では，ネットワーク分析への応用を目指したグラフ描画手法を研究する。

本研究では，グラフのレイアウトから特定のノードやエッジに着目する分析に貢献することを目指し，既存のグラフ描画手法を拡張する。対象としたグラフ描画手法は，力学モデルと呼ばれているオーソドックスな手法のうち，特に有名なKamada-Kawai法 [2]である。本研究では，Kamada-Kawai法 [2]に用いられている目的関数にパラメータを導入し，グラフレイアウトを変化させることで，特定のノードを中心に据えることを目的とする。



## 第2章

# 関連研究

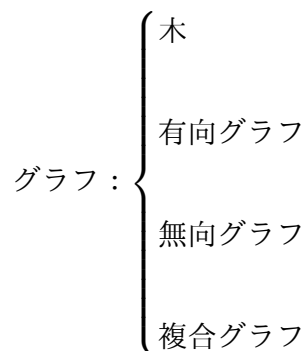
本章では、まず、本研究で扱うグラフ描画手法がグラフ描画問題の中でどのような位置づけにあるかについて述べる。次に、本研究の背景技術である、Fruchterman-Reingold法およびKamada-Kawai法について概説する。上に述べた2つの手法は、現在のネットワーク分析アプリケーションにも採用されている有名なグラフ描画手法である。最後に、力学モデルの応用研究のうち、計算の高速化を目指した関連研究と、レイアウトを変化させる関連研究について述べる。

## 2.1 本研究の位置づけ

まず，本研究で描画の対象とするグラフの構造的な規則について定義する．その後，グラフの種類ごとにレイアウトの評価観点を整理した文献から，本研究で扱うグラフ描画手法の背景技術がグラフ描画問題の中でどのような位置づけにあるのかについて述べる．

### 2.1.1 描画対象のグラフ

グラフは構造的な規則に基づき，いくつかのクラスに分類される．以下にグラフの大きな分類を示す．



本研究の背景技術である力学モデルが対象とするグラフは，エッジが向きを持たない「無向グラフ」であり，特に，多重辺や自己ループを持たない，「一般無向グラフ」と呼ばれるグラフである．

### 2.1.2 グラフの自動描画問題の評価観点

グラフの自動描画による可視化の有用性は，可視化されたグラフのレイアウトに強く依存している．すなわち，良好なレイアウトによるグラフを見れば，人はそのグラフが示す意味を正確に理解し，文書で記述されるよりも容易に記憶することができるが，悪いレイアウトによるグラフからは，人はそのグラフの示す意味を正確に理解することはできず，再現性も困難となりえる．そのため，グラフの自動描画問題では，人間の知覚的な基準を計算機に与えるアルゴリズムに反映させることが重要となる．

書籍『グラフ自動描画法とその応用』[10]では，グラフの「良い」レイアウトとは何かに関する共通で基本的な基準を「描画規則」と称し，整理している．描画規則は複数存在するが，一つのアルゴリズムでそれらすべてを満足することは困難であるとされ，一般には描画規則ごとに優先順位を与えたアルゴリズムを構築することが多い．優先順位は，

グラフのクラスに応じて付けられることがあり、例えば、「木構造」を持つグラフ(例えば、家系図や組織図)においては、

1. 根を最上段に置き、順次子孫を下方の平行線上に配置する(階層性)
2. 辺は交差させない(無交差)
3. 同一階層において隣のノードとは最小距離 $d$ 以上離す

などの描画規則が優先される。

本研究において描画対象とする一般無向グラフの自動描画手法では、以下の描画規則を優先して考慮するものとされる。

1. 隣接しているノード同士を近くに配置する(近接性)
2. エッジは近づけすぎないように配置する(最小分離)
3. エッジの長さを一定にする(辺長一定)
4. 対称性を顕示する(対称性)
5. 与えられた描画領域の中にノードを一様に分布させる(一様分布)
6. 描画枠に合わせる(枠への適合)
7. エッジの交差を最小にする(辺交差最小)

## 2.2 力学モデルによるグラフ描画手法

一般無向グラフを対象とするグラフ描画手法の分野では、「力学モデル」と呼ばれる手法が主に用いられている。力学モデルの基本的な考え方は、グラフ中のノードを鉄のリングに、エッジを力学系を構成するバネに置き換え、リングにかかるバネの力が最小のエネルギー状態を実現するような配置を求めるものである。この手法で念頭に置かれている描画規則は「近接性」および「対称性」である。

力学モデルに基づいたグラフ描画手法の中でも、Fruchterman-reingoldの手法 [3]とKamada-Kawaiの手法 [2]の2つは特に有名であり、多くのグラフ分析アプリケーションに採用されている。本稿では、既往研究として、以上の2つの手法について解説する。

### 2.2.1 Fruchterman-Reingoldの手法

Fruchterman-Reingoldの手法(以下、FR法と称す)では、グラフ中のノードにおいて、エッジで直接接続されているノード同士に引力が働き、そのほかのノード間には斥力が

働くようなバネの力学系を仮定する。各ノードは、バネの働きによって移動するが、その移動距離が描画領域を大きくはみ出さないように、以下のような制限を与える。

$$l = C \sqrt{\frac{\text{描画領域}}{\text{ノードの数}}}. \quad (2.1)$$

ここで、定数Cは実験的に見いだされる。

二つのノード間の距離を $d$ とすると、それらの間に作用する引力 $f_a$ と斥力 $f_r$ は次式で定義される。

$$f_a = \frac{d^2}{l}, \quad (2.2)$$

$$f_r = -\frac{l^2}{d}. \quad (2.3)$$

ここで、 $f_a + f_r = 0$ のとき、 $d$ は $l$ に等しくなる。そのため、 $l$ はノード間の理想距離と呼ばれる。

FR法では、はじめにランダムに配置された各ノードに作用する引力および斥力を計算し移動量を決め、ノードを再配置することを繰り返していくことで、理想的なグラフィックアウトを得る。

## 2.2.2 Kamada-Kawaiの手法

Kamada-Kawaiの手法（以下、KKと称す）では、グラフ中の全てのノードの組に、バネ定数を $k$ とした、仮想的なバネを付与したモデルを構築する。二つのノードの間にあるバネの自然長（バネに力が働かない長さ）を $l$ 、現実の長さを $d$ とすると、ノードに働く力 $f$ は、Hookeの法則より、

$$f = k(d - l), \quad (2.4)$$

である。すなわち、 $d$ がバネの自然長より長ければノード間には引力が働き、短ければノード間には斥力が働く。また、この力学系のスカラーポテンシャルは

$$E = \sum_i \frac{k}{2} (d_i - l_i)^2, \quad (2.5)$$

で表される。KK法では、このスカラーポテンシャルを減少させるために、 $E$ を各ノードの配置により偏微分し、各ノードの移動量を求め再配置する。ノードの再配置を十分繰り返すか、スカラーポテンシャルが十分小さくなったときのノードの配置から理想的なグラフィックアウトを得る。

### 2.2.3 FR法とKK法による描画結果

参考に、プログラミング言語Rのグラフ分析用パッケージ「igraph」 [4]に搭載されたFR法とKK法によって、ランダムに生成した同じ構造のグラフを描画させた。図 2.1はFR法によるレイアウト、図 2.2はKK法によるレイアウトである。両方のグラフレイアウトを概観すると、FR法によるレイアウトは、全ノードを満遍なく分布させ、KK法によるレイアウトはグラフの局所的な構造を強調したようなレイアウトになっている。

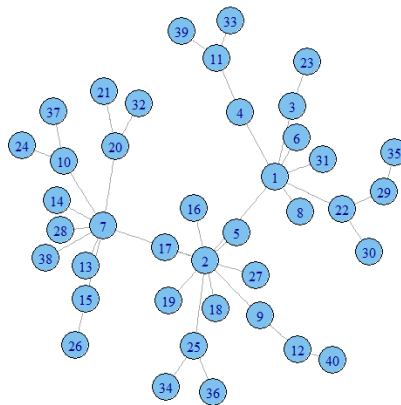


図2.1: FR法によるグラフレイアウト

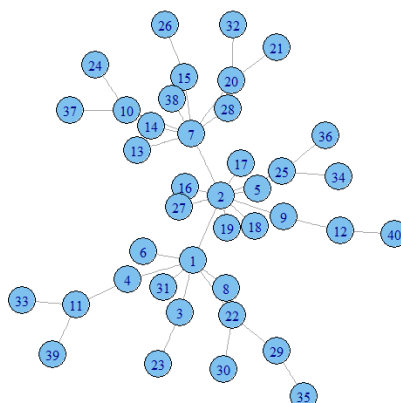


図2.2: KK法によるグラフレイアウト

## 2.3 関連研究と本研究の位置づけ

力学モデルの研究の方向性は様々だが、本稿では、計算の高速化を目指した研究とレイアウトを変化させる研究を、本研究との関連研究として述べる。

### 2.3.1 計算の高速化を目指した研究

計算の高速化は、近年の情報処理技術の進歩やインターネットの普及を背景に、得られるネットワークのサイズが巨大化していることから関心の高いテーマである。紹介した力学モデルの2つの手法は、グラフの全ノードの情報について計算するアルゴリズムを採用しているため、描画するグラフのサイズが大きくなることによる計算量の増大は深刻である。

KK法を高速化する代表的な手法として、多次元尺度構成法の分野で利用されている Stress Majorization が知られている [11]。これは、KK法で得られた力学系のスカラーポテンシャルが、多次元尺度構成法で知られる stress 関数と呼ばれる損失関数と考え方が同じであるという知見に基づいている。文献 [11] より、Stress Majorization は KK法で従来用いられてきた Newton法 [2] と比べ高速に解を求められることが確かめられている。本研究も KK法で用いられる目的関数の最小化に Stress Majorization を用いた。具体的な計算手法は第3章で述べる。

現在のグラフ描画コミュニティでは、Stress Majorization を高速に計算する手法 [12] や、数値最適化の新しい手法によって KK法、FR法を共に高速に解く手法など [13] の研究が進められている。

### 2.3.2 レイアウトを変化させる研究

力学モデルにおいてグラフレイアウトを変化させるには、ノードに作用する引力、斥力を変化させる手法が用いられる。

例えば、KK法の応用研究として、stress関数の形を変形させた研究がある [14]。この研究では、KK法で用いられているバネの自然長を等しくすると、円形のグラフレイアウトが得られることを応用し、ノード配置が均一なグラフレイアウトを得た。この結果は、KK法においてレイアウトを変化させるには、バネの自然長に操作を加える手法が有用であることを示唆しており、本研究もこの発想から KK法を拡張させた [15]。

上述した研究のグラフレイアウトの手法では、得られたレイアウトの全体的な balan

スを志向している。一方で、ネットワークの分析には、特定のノードやエッジの集合を対象とした分析が多数知られている。既存のネットワーク分析アプリケーションでは、これらの分析課題に対応するため、グラフ理論の観点から重要なノードを中心性指標などを基に計算し、対象のノードやエッジのサイズを拡大して表示する機能が追加されている [4]。本研究では、同じスキームから特定のノードやエッジを強調させるアプローチとして、レイアウトの局所的な変化を考える。提案手法と既存の機能を併用することにより、ネットワーク分析の簡便性の向上に寄与できることを期待している。

## 第3章

# グラフ描画手法の定式化と解法

本章では，グラフ描画の基本的な考え方について述べ，最小化問題への定式化を行う．定式化した最小化問題は，Stress Majorizationと呼ばれる手法によって数値的に解が得られることを示す．また，いくつかの簡単な構造のグラフと，実世界で生成される一般的な構造のグラフを対象に描画し，描画結果が良好であることを確かめる．最後に，得られたレイアウトのゆがみに関する考察を行う．



## 3.1 問題設定と定式化

本稿で用いるグラフに関する基礎的な知識について説明し，グラフ描画の考え方とそれに基づく問題設定を行う．

### 3.1.1 基礎知識

本研究で扱うグラフの諸要素について，数学用語により以下の様に定義する．

- i).  $n$  : グラフのノードの総数
- ii).  $m$  : グラフのエッジの総数
- iii).  $V = \{v_1, v_2, \dots, v_n\}$  :  $n$ 個のノードの集合
- iv).  $E = \{e_1, e_2, \dots, e_m\}$  :  $m$ 本のエッジの集合
- v).  $\mathbf{p} = \{p_1, p_2, \dots, p_n\}$  : 二次元平面上に描画されたグラフの各ノードの座標の集合
- vi).  $|p_i - p_j|$  :  $v_i$ と $v_j$ の間の二次元平面上におけるユークリッド距離
- vii).  $h_{ij}$  :  $v_i$ と $v_j$ の間のグラフ理論的距離（ダイクストラ法などで求められる，単純なホップ数）

グラフの構造を表現する概念として，距離行列と呼ばれるものがある．距離行列とは，グラフ中のノード間のグラフ理論的距離を行列の成分として表した行列である．表 3.1に距離行列の例を示す．

表3.1: グラフの距離行列

ノードid	A	B	C	D
A	0	1	2	3
B	1	0	1	2
C	2	1	0	1
D	3	2	1	0

### 3.1.2 グラフ描画の考え方

グラフのノードを二次元平面上に配置する際、ノード間のグラフ理論的距離が反映されていないと、人間の知覚的な直観と乖離が生じてしまう。

例えば、以下の図 3.1は、表 3.1の距離行列を二つの異なるレイアウトで描画したグラフであるが、両グラフの構造は同じでも、左側のレイアウトにおいては、ノードAとノードDのユークリッド距離はノードAとノードCのユークリッド距離よりも短くなってしまっているため好ましいレイアウトとはいえず、右側のレイアウトの方が、ノードAとノードDのユークリッド距離を正確に反映させているという観点から、好ましいレイアウトである。

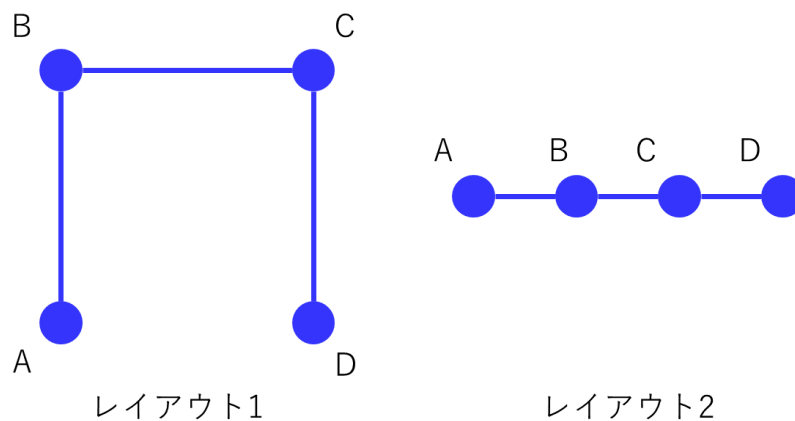


図3.1: レイアウトの異なる同じ構造のグラフ

### 3.1.3 最小化問題への定式化

3.1.2で述べた考え方に基づき、本研究では、問題設定として、「グラフ中の全ノード間のユークリッド距離とグラフ理論的距離の誤差がなるべく小さくなるようなノード座標を求める」ことを与える。そこで、ノード $i, j$ 間のグラフ理論的距離 $h_{ij}$ を理想的な距離 $d_{ij}$ として定義し、ノード座標 $\mathbf{p}$ について以下の目的関数を与える。

$$\epsilon(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n) = \sum_{i=1}^n \sum_{j \neq i} (|\mathbf{p}_i - \mathbf{p}_j| - d_{ij})^2. \quad (3.1)$$

ここで、式(3.1)の右辺はノード $i, j$ 間のユークリッド距離とグラフ理論的距離の差を2乗したものの総和である。一般的を失わず、グラフ中のノード $v_1$ の二次元平面上の座標を $p_1$ としたとき、 $p_1$ は描画したグラフの座標系の原点とし、 $p_1 = (0, 0)$ としてもよい。理想的なグラフィックレイアウトを得るような $\mathbf{p}$ は以下の最小化問題 (3.2) の解として与えられる。

$$\epsilon(p_2, \dots, p_n) = \arg \min_{(z_2, \dots, z_n)} \epsilon(z_2, \dots, z_n). \quad (3.2)$$

KK法で用いた目的関数の式 (2.5) において、ノード $i, j$ 間のバネの自然長をノード $i, j$ 間のグラフ理論的距離 $h_{i,j}$ として与えると、式 (3.1) の目的関数を最小化することは、ノード同士が引力および斥力を及ぼし合うバネからなる系の平衡状態を計算すること（スカラーポテンシャルを最小化すること）と同義であることがわかる。

## 3.2 最小化問題の解法

最小化問題の解法としてのStress Majorizationの具体的な手順について述べるとともに、得られる解の収束性について議論する。

### 3.2.1 Stress Majorization

式 (3.2) において、 $\epsilon(p_1, p_2, \dots, p_n)$ は、 $(p_1, p_2, \dots, p_n)$ の非線形な関数であるため、最適な座標 $\mathbf{p}$ を解析的な手法で求めることはできない。しかし、最急降下法などの数値計算によって局所的に最適な座標 $\mathbf{p}$ を求めることが出来る。本研究では、多次元尺度構成法の分野で知られるStress Majorizationと呼ばれる手法 [11]を用いて目的関数を局所最小化し、各ノードの座標を求める。以下、解法について述べる。はじめに、以下の式 (3.3) で定義される $\mathbf{g}(p_2, \dots, p_n, q_2, \dots, q_n)$ について考える。

$$\mathbf{g}(p_2, \dots, p_n, q_2, \dots, q_n) \stackrel{\text{def}}{=} \sum_{i,j;i \neq j} \left\{ |p_i - p_j|^2 + d_{ij}^2 - 2|p_i - p_j||q_i - q_j| \frac{d_{ij}}{|q_i - q_j|} \right\}. \quad (3.3)$$

このとき、 $\epsilon(p_2, \dots, p_n)$ について式 (3.4) が成り立つ。

$$\begin{aligned} \epsilon(p_2, \dots, p_n) &= \mathbf{g}(p_2, \dots, p_n, p_2, \dots, p_n) \\ &\leq \mathbf{g}(p_2, \dots, p_n, q_2, \dots, q_n). \end{aligned} \quad (3.4)$$

Stress Majorizationでは、最適なノードの座標を求めるために、再帰的な計算を行う。 $t$ 回目のステップで、ノードの座標 $(p_2^{(t)}, \dots, p_n^{(t)})$ が得られていると仮定する。 $t + 1$ 回目のステップで得られるノードの座標を $(p_2^{(t+1)}, \dots, p_n^{(t+1)})$ とすると、 $(p_2^{(t+1)}, \dots, p_n^{(t+1)})$ は

$\mathbf{g}(p_2, \dots, p_n, p_2^{(t)}, \dots, p_n^{(t)})$  を極小化した座標である。すなわち、 $(p_2^{(t+1)}, \dots, p_n^{(t+1)})$  は以下の式を満足する。

$$\begin{aligned} & \left[ \frac{\partial \mathbf{g}(p_2, \dots, p_n, p_2^{(t)}, \dots, p_n^{(t)})}{\partial p_i} \right]_{(p_2, \dots, p_n) = (p_2^{(t+1)}, \dots, p_n^{(t+1)})} \\ &= 2 \sum_{j=1; j \neq i}^n (p_i^{(t+1)} - p_j^{(t+1)}) - (p_i^{(t)} - p_j^{(t)}) \frac{d_{ij}}{|p_i^{(t)} - p_j^{(t)}|} \\ &= 0. \end{aligned} \quad (3.5)$$

ここで、 $p_1^{(t)} = p_1^{(t+1)} = 0$  である。式(3.5)から、以下の等式が得られる。

$$\begin{aligned} (n-1)p_1^{(t+1)} - \sum_{j=1; j \neq i}^n p_j^{(t+1)} \\ = x_i^{(t)} \sum_{j=1; j \neq i}^n \frac{d_{ij}}{|p_i^{(t)} - p_j^{(t)}|} - \sum_{j=1; j \neq i}^n \frac{d_{ij} p_j^{(t)}}{|p_i^{(t)} - p_j^{(t)}|}. \end{aligned}$$

上の式より、以下の漸化式 (3.6) が得られる。

$$\begin{aligned} (p_{2x}^{(t+1)}, \dots, p_{nx}^{(t+1)})^T &= A^{-1} (\beta_{2x}^{(t)}, \dots, \beta_{nx}^{(t)})^T \\ (p_{2y}^{(t+1)}, \dots, p_{ny}^{(t+1)})^T &= A^{-1} (\beta_{2y}^{(t)}, \dots, \beta_{ny}^{(t)})^T. \end{aligned} \quad (3.6)$$

ここで、

$$\begin{aligned} \beta_i^{(t)} &= x_i^{(t)} \sum_{j=1; j \neq i}^n \frac{d_{ij}}{|p_i^{(t)} - p_j^{(t)}|} - \sum_{j=1; j \neq i}^n \frac{d_{ij} p_j^{(t)}}{|p_i^{(t)} - p_j^{(t)}|}, \\ A &= \underbrace{\begin{pmatrix} n-1 & -1 & \dots & -1 \\ -1 & n-1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & n-1 \end{pmatrix}}_{n-1}. \end{aligned}$$

式 (3.6) において、 $p_{ix}^{(t+1)}$  と  $p_{iy}^{(t+1)}$  は、 $p_i^{(t+1)}$  の  $x$  座標と  $y$  座標であり、 $\beta_{ix}^{(t)}$  と  $\beta_{iy}^{(t)}$  は、 $\beta_i^{(t)}$  の  $x$  座標と  $y$  座標である。 $p_i^{(t)}$  の定義と式 (3.4) より、次のことがわかる。

$$\begin{aligned} \epsilon(p_2^{(t)}, \dots, p_n^{(t)}) &= \mathbf{g}(p_2^{(t)}, \dots, p_n^{(t)}, p_2^{(t)}, \dots, p_n^{(t)}) \\ &\geq \mathbf{g}(p_2^{(t+1)}, \dots, p_n^{(t+1)}, p_2^{(t)}, \dots, p_n^{(t)}) \\ &\geq \mathbf{g}(p_2^{(t+1)}, \dots, p_n^{(t+1)}, p_2^{(t+1)}, \dots, p_n^{(t+1)}) \\ &= \epsilon(p_2^{(t+1)}, \dots, p_n^{(t+1)}). \end{aligned}$$

したがって、式(3.6)を介して再帰的に得られるノードの座標は単調に目的関数を減少させ、いずれ局所最適解に収束する。Aの逆行列は以下の式で表すことが出来る。

$$A^{-1} = \frac{1}{n} \underbrace{\begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2 \end{pmatrix}}_{n-1}.$$

### 3.2.2 目的関数の収束性について

Stress Majorizationによって解を求めるには、はじめにランダムな初期座標を各ノードに与えなくてはならない。式(3.1)の目的関数はほとんどの場合において複数の局所最小値をもつため、得られたノードの座標は必ずしも最良のレイアウトを保証するものではない。しかしながら、本研究で用いたStress Majorizationによって得られる収束値は、グラフ中の全ノード間の理想的な距離があらかじめ与えられている場合、初期値依存性が非常に小さいことが知られている [16]。

したがって本研究では、異なる100パターンのノードの初期座標を与え、得られた収束値のうち最も小さい収束値を得たノードの座標を採用することで、良好なレイアウトを得られることが出来ると判断した。

## 3.3 描画結果と考察

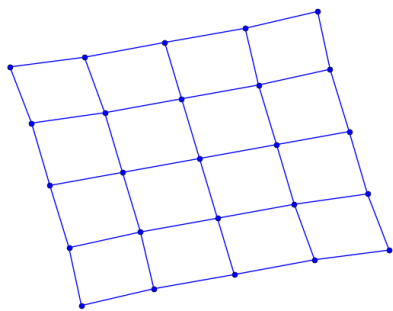
3.2で述べた手法により、いくつかの簡単な構造のグラフと、実世界で生成される一般的な構造のグラフを描画し、描画結果について考察する。

### 3.3.1 対象グラフ

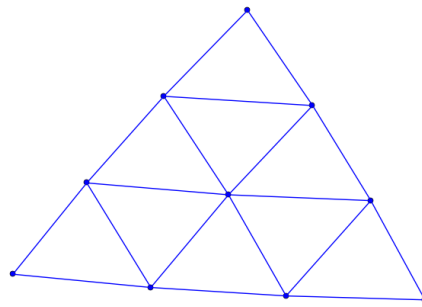
3.2で述べた手法により、実際にグラフを描画した。本章では、レイアウトが容易に想像できる簡単な構造の3つのグラフと、V.ユゴーの小説『Les Misérables』の登場人物の関係をグラフ化したもの [17]の描画を行った。それぞれのグラフの諸元を表 3.2に示す。

表3.2: グラフの諸元

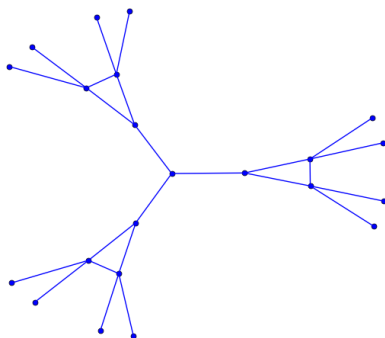
グラフの名前	ノード数	エッジ数
(a)四角格子状のグラフ	25	40
(b)三角格子状のグラフ	10	18
(c)放射状のグラフ	22	24
(d)Les Miserables [17]	77	254



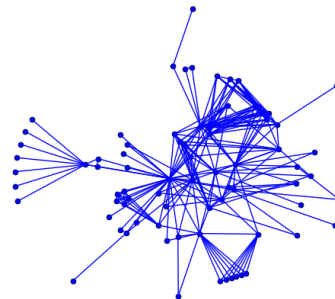
(a) 四角格子状のグラフ



(b) 三角格子状のグラフ



(c) 放射状のグラフ



(d) Les Miserables

図3.2: グラフの描画結果

### 3.3.2 描画結果

図 3.2は描画対象のグラフを描画した結果である。それぞれにレイアウトにおいて、近接性や対称性を顕示していることが確認でき、本章で用いた手法により良好なレイアウトを得られた。

### 3.3.3 レイアウトのゆがみに関する考察

図 3.2で得られたグラフレイアウトのうち、(a)四角格子状のグラフ、(b)三角格子状のグラフにおいて、格子を構成する正四角形や正三角形は、厳密にトポロジを再現できておらず、わずかにゆがみが認められる。また、(c)放射状のグラフでも、三角形を構成する辺の長さは等しく描画されていないことが確かめられる。これは、三角不等式を根拠にした説明ができる。例えば、以下の表 3.3の距離行列で表現されるようなグラフを描画するとき、図 3.3のように、三角不等式よりノードAとノードC、ノードDとの二次元平面上の距離はグラフ理論的距離よりも短く描画されてしまう。

KK法では、ノード間の全てのグラフ理論的距離を参照してノード配置を求めるため、四角形の対角にあたるノード間の理想距離を2として計算している。さらに、グラフが格子状に連なっている場合、格子の両端の対角の距離とグラフ理論的距離の誤差は格子の数に比例して大きくなるため、レイアウトをゆがめてしまうと考えられる。

表3.3: グラフの距離行列

ノードid	A	B	C	D
A	0	1	2	2
B	1	0	1	1
C	2	1	0	1
D	2	2	1	0

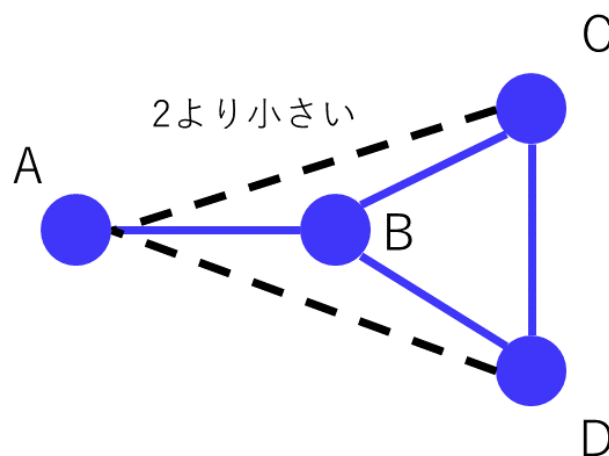


図3.3: 三角不等式からなる距離の誤差

## 第4章

# 特定のノードを中心に据えたグラフ 描画手法

本章では、特定のノードを中心に据えるグラフ描画手法を提案し、提案手法によって実際にグラフを描画した結果を示すとともに、描画結果の考察を述べる。まず、特定のノードを中心に据えたグラフィックレイアウトの描画仮説を説明するとともに、描画仮説に基づいた具体的なKK法の拡張方法について詳述する。提案した手法により、第3章で描画したグラフを再び描画し、レイアウトを変化させた結果が描画仮説を満足していることを確かめる。最後に、中心性の異なるノードを描画した際のレイアウトへの影響について考察した。



## 4.1 描画仮説と拡張方法

特定のノードを中心に据えたグラフレイアウトの描画仮説を説明し、描画仮説に基づいた具体的なKK法の拡張方法について詳述する。

### 4.1.1 描画仮説

特定のノードを中心に据えるための、描画仮説について述べる。特定のノードを中心としてグラフをレイアウトしたとき、特定のノードと関係の近いノード（グラフ理論的距離が短いノード）同士の距離は拡大して配置され、関係が遠いノード同士の距離は縮小して配置される。これは、家系図の特定の人物を中心としたグラフとした場合、両親や配偶者、子供などの1ホップでつながる人物との関係を示すエッジは長く描かれるが、何世代も離れた親戚の血縁関係を示すエッジは短く描かれることが一般的であることに基づく仮説である。仮説の概念図を図 4.1に示す。

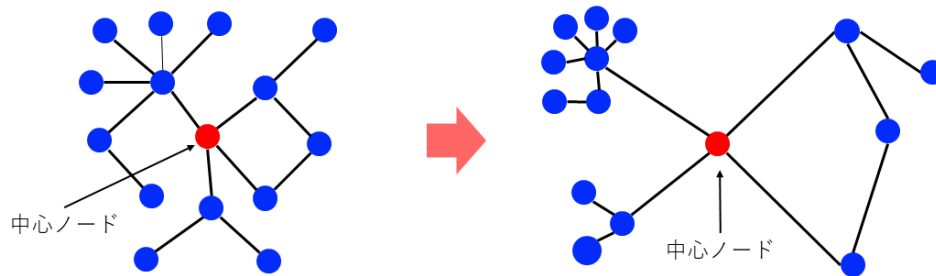


図4.1: 描画仮説の概念図

### 4.1.2 拡張方法

描画仮説に基づき，グラフ中のノード集合 $V$ から中心ノードとして選択したノード $v_c$ とグラフ理論的距離が遠いノード同士の間にはエッジが接続されていれば，エッジの長さを適当に短くするように工夫する．そこで，式(3.1)の説明変数である理想距離 $d_{ij}$ を，パラメータを導入したノード間のエッジの長さによって新たに定義しなおした．以下，具体的な手順について述べる．

手順1  $v_i$ と，中心に据えたいノードからのグラフ理論的距離（ $v_c$ からのホップ数）を $h_{ic}$ とする．

手順2  $v_i$ と $v_j$ がエッジで接続されているとき，エッジの長さ $e_{ij}$ を以下の式で与える．

$$e_{ij} = \alpha^{\max\{h_{ic}, h_{jc}\}-1} \quad (4.1)$$

ここで， $0 < \alpha \leq 1$ は距離に関するパラメータである．

手順3 式(4.1)で計算したエッジの長さを用いて，全てのノード間のグラフ理論的距離を（ダイクストラ法などで）計算し，それをノード間の理想的な距離 $d_{ij}$ とする．

手順4  $v_i$ と $v_j$ の間のホップ数 $h_{i,j}$ が

$$h_{ij} \geq h_{ic} + h_{jc}$$

を満たす場合は， $v_i$ と $v_j$ の理想的な距離 $d_{ij}$ を以下の式で置き換える．

$$d_{ij} = d_{ic} + d_{jc}$$

以上の手順によって計算した理想距離 $d_{ij}$ を用いて，Stress Majorizationによってノードの配置を得る．

式(4.1)により，ノード $i$ とノード $j$ と中心ノードがともに隣接しているとき(すなわち， $h_{ic}=h_{jc}=1$ のとき)，エッジの長さは本来の長さとは変わらず，中心ノードとのグラフ理論的距離が長いノードと，その他のノードの間に接続されているエッジの長さは，中心ノードとの距離に比例して短くなる．そのため，式(3.1)により得られるグラフィックレイアウトにおいて，中心ノードと隣接しているノードの間の距離はそのほかのノード間の距離よりも相対的に長く表示される．また，距離に関するパラメータ $\alpha$ は，0に近い値を設定するほど，距離を短くする補正は強く作用する．逆に1に近い値を設定するほど，補正は弱く作用し，1としたとき，エッジの長さはまったく変化せず，得られるレイアウトは3.3で得たレイアウトと変わらなくなる．

[手順4]による距離の修正は、例えば、図 4.2のような構造のグラフにおいて赤色で示したノードを中心に描画する際、式(4.1)を適用すると、中心ノードと1ホップでつながっているノードA, B, Cの間の理想距離は、間にあるエッジの長さが短縮されることで、極端に短くなってしまふことを考慮した修正である。2つのノードの間と、それぞれの中心ノードとの間のホップ数を参照して、中心ノードとの1ホップの距離関係を保持させる。

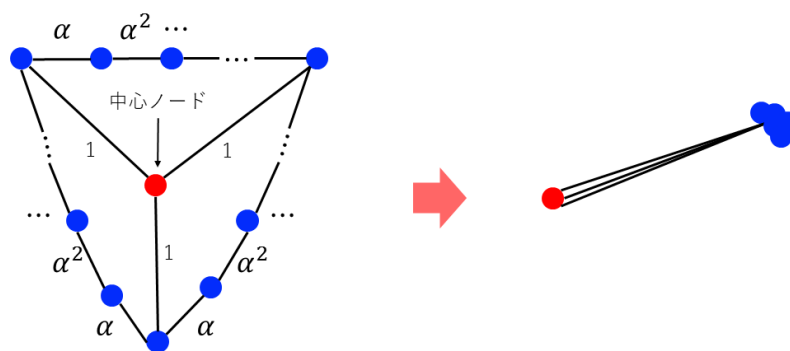


図4.2: 理想距離の極端な縮小

### 4.1.3 予備実験

提案手法がグラフの各エッジの長さによつてどのように影響を及ぼすのかを詳細に見ていくために、非常に単純な構造をしたグラフを対象に予備実験を行う。表 4.1の構造を持つグラフ(図 4.3)について、ノードidが3のノード(赤色で表す)を中心に描画することを目標に、提案手法を適用した場合、距離行列の各成分がどのように変化していくか、手順ごとに見ていく。

表4.1: グラフの距離行列

ノードid	0	1	2	3	4	5	6
0	0.0	1.0	1.0	2.0	3.0	3.0	4.0
1	1.0	0.0	1.0	2.0	3.0	3.0	4.0
2	1.0	1.0	0.0	1.0	2.0	2.0	3.0
3	2.0	2.0	1.0	0.0	1.0	1.0	2.0
4	3.0	3.0	2.0	1.0	0.0	2.0	1.0
5	3.0	3.0	2.0	1.0	2.0	0.0	1.0
6	4.0	4.0	3.0	2.0	1.0	1.0	0.0

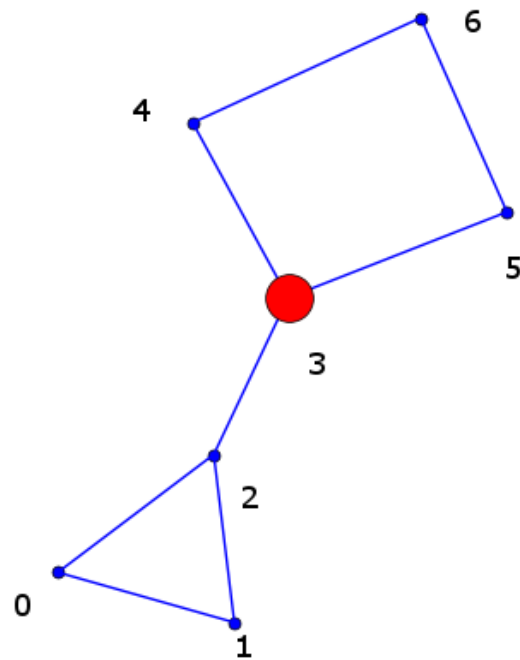


図4.3: 予備実験のグラフ(距離修正前)

パラメータ $\alpha$ を0.5としたとき, [手順1], [手順2]により, 距離行列の中で1の成分, すなわちエッジの長さは表 4.2のように変化する. また, そのほかの成分の長さは未知であるため, 無限大とする.

表 4.2の距離行列の長さが未知の成分は, ダイクストラ法などで計算できる. 計算した後の距離行列を表 4.3に示す.

表 4.1より, ノードid4とid5の間のホップ数は, ノードid3とノードid4の間のホップ数と, ノードid3とノードid5の間のホップ数と等しい. [手順4]の条件を満たすため, 計算した理想距離 $d_{45}$ を $d_{34} + d_{35}$ の値に置き換える. 表 4.4は, 最終的に得られた理想距離 $d_{ij}$ の距離行列である. [手順4]により更新した部分を赤色で着色する.

表4.2: [手順1][手順2]によって変化したグラフの距離行列

ノードid	0	1	2	3	4	5	6
0	0.0	0.5	0.5	inf	inf	inf	inf
1	0.5	0.0	0.5	inf	inf	inf	inf
2	0.5	0.5	0.0	1.0	inf	inf	inf
3	inf	inf	1.0	0.0	1.0	1.0	inf
4	inf	inf	inf	1.0	0.0	inf	0.5
5	inf	inf	inf	1.0	inf	0.0	0.5
6	inf	inf	inf	inf	0.5	0.5	0.0

表4.3: ダイクストラ法によって変化したグラフの距離行列

ノードid	0	1	2	3	4	5	6
0	0.0	0.5	0.5	1.5	2.5	2.5	3.0
1	0.5	0.0	0.5	1.5	2.5	2.5	3.0
2	0.5	0.5	0.0	1.0	2.0	2.0	2.5
3	1.5	1.5	1.0	0.0	1.0	1.0	1.5
4	2.5	2.5	2.0	1.0	0.0	1.0	0.5
5	2.5	2.5	2.0	1.0	1.0	0.0	0.5
6	3.0	3.0	2.5	1.5	0.5	0.5	0.0

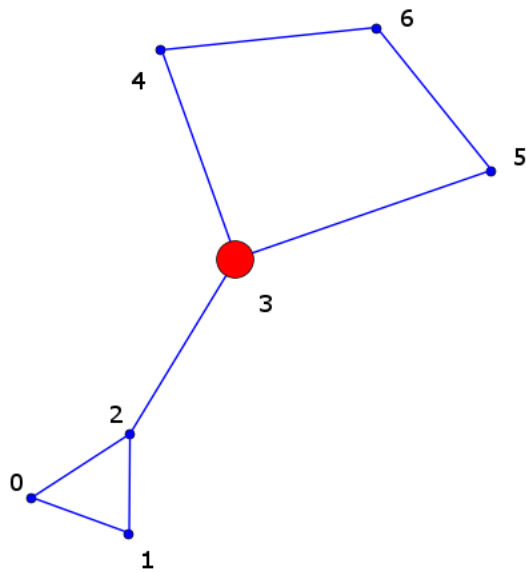


図4.4: 予備実験のグラフ (距離修正後)

表4.4: [手順4]によって変化したグラフの距離行列

ノードid	0	1	2	3	4	5	6
0	0.0	0.5	0.5	1.5	2.5	2.5	3.0
1	0.5	0.0	0.5	1.5	2.5	2.5	3.0
2	0.5	0.5	0.0	1.0	2.0	2.0	2.5
3	1.5	1.5	1.0	0.0	1.0	1.0	1.5
4	2.5	2.5	2.0	1.0	0.0	2.0	0.5
5	2.5	2.5	2.0	1.0	2.0	0.0	0.5
6	3.0	3.0	2.5	1.5	0.5	0.5	0.0

距離の変化したグラフレイアウトを図 4.4に示す。中心ノードであるid3の周辺の四角形が拡大され、id3と2ホップ離れている三角形が縮小してレイアウトされていることが確認でき、本手法が有効に機能したことがわかる。

## 4.2 描画結果と考察

提案した手法により，第3章にて描画したグラフを再び描画し，レイアウトの変化が描画仮説を満足することを確認する．また，得られた描画結果を評価するとともに，中心とするノードの選択がレイアウトの変化に及ぼす影響について考察する．

### 4.2.1 対象グラフ

提案手法により，第3章の表 3.2のグラフを再び描画する．

### 4.2.2 描画結果

図 4.5, 図 4.6, 図 4.7, 図 4.8はそれぞれのグラフでパラメータ $\alpha$ を1.0, 0.6, 0.4, 0.2, 0.05, 0.01としたレイアウトである．それぞれのグラフレイアウトで，中心ノードは星形で示してある．

レイアウトの変化について考察する．図 4.5, 図 4.6のグラフレイアウトの変化をみると， $\alpha$ の値が小さくなるにつれて周囲の格子が縮小し， $\alpha$ が0.2より小さいレイアウトでは，中心ノードと2ホップ以上離れているノード同士の関係性は強く縮小された．図 4.7のレイアウトの変化では，描画仮説の通りの縮小現象が確認された．一般的なグラフ構造であるLes Misérablesのレイアウトでは， $\alpha$ の値が小さくなるにつれて，中心ノードと1ホップで接続されているノードの所在が強調してレイアウトされるようになった．これは，通常のKK法によって得られるグラフレイアウト(提案手法における 4.8aのレイアウト)では視認することは難しい．Les Misérablesのグラフは小説中の人間関係をグラフ化したものなので，描画結果は特定の登場人物の視座からみた人間関係の主観的な世界観を表現したグラフレイアウトともいえる．

定性的ながら，提案手法は，それぞれの異なる構造のグラフにおいて概ね描画仮説を満足する結果を得たと評価できる．

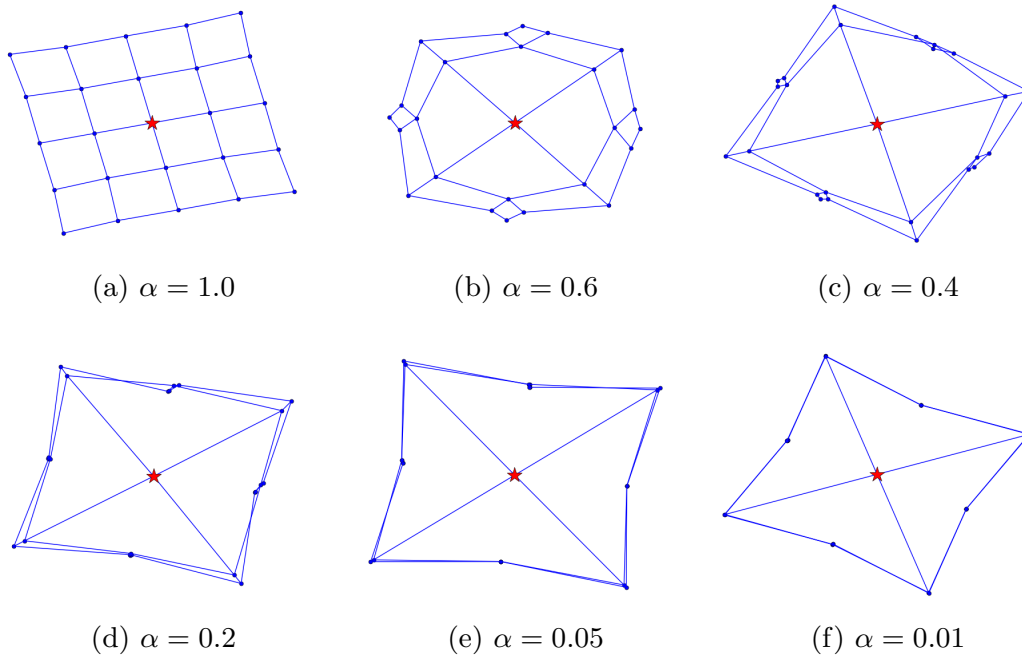


図4.5: 四角格子状のグラフレイアウトの変化

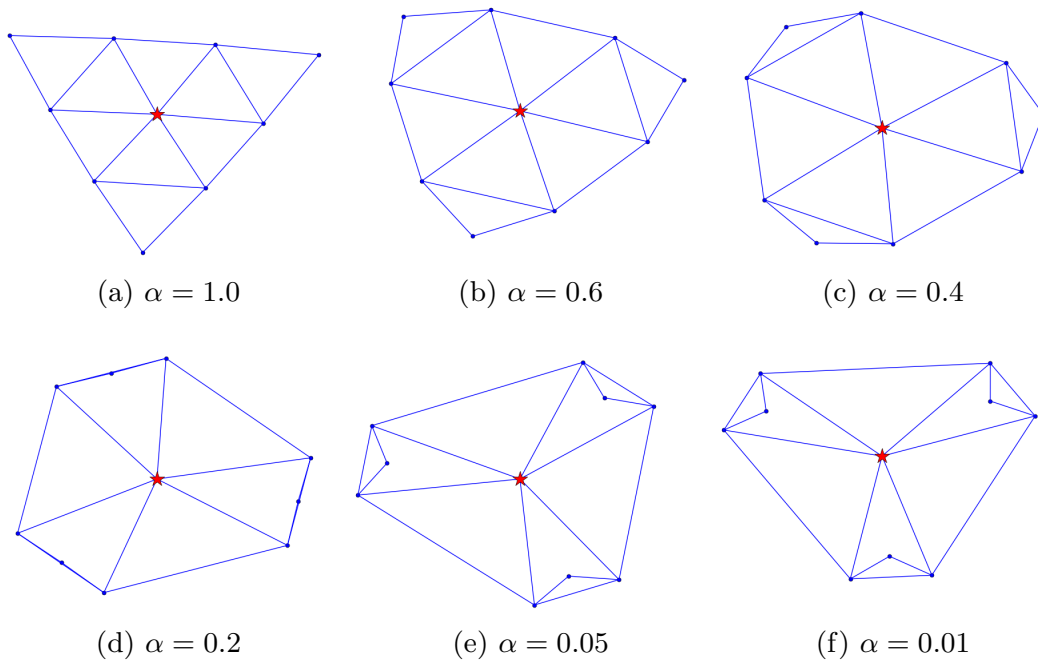


図4.6: 三角格子状のグラフレイアウトの変化



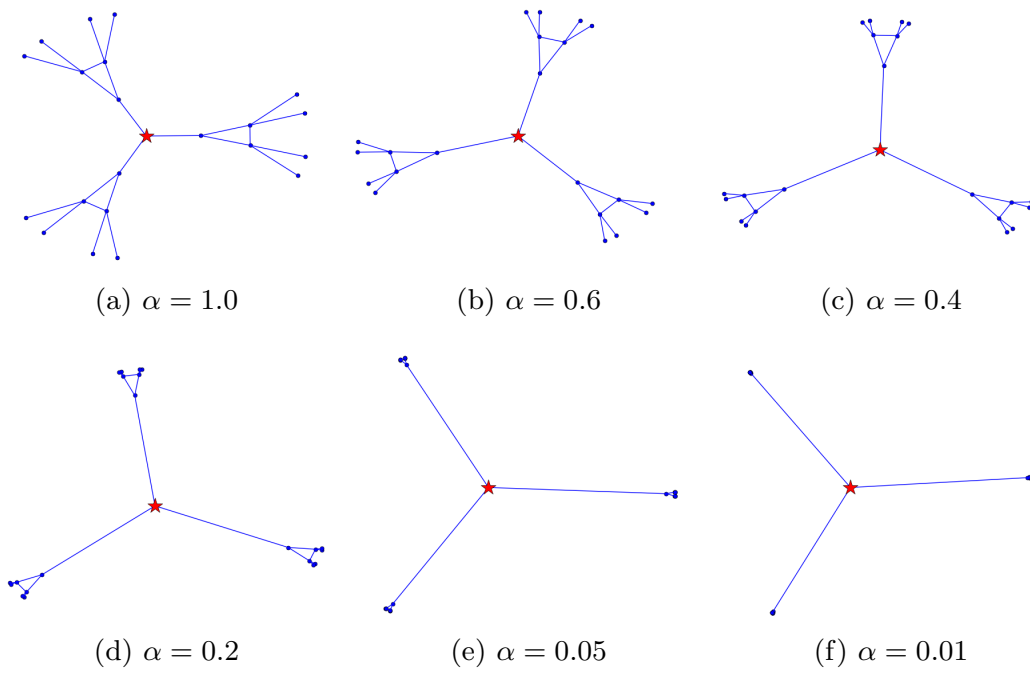


図4.7: 放射状のグラフレイアウトの変化

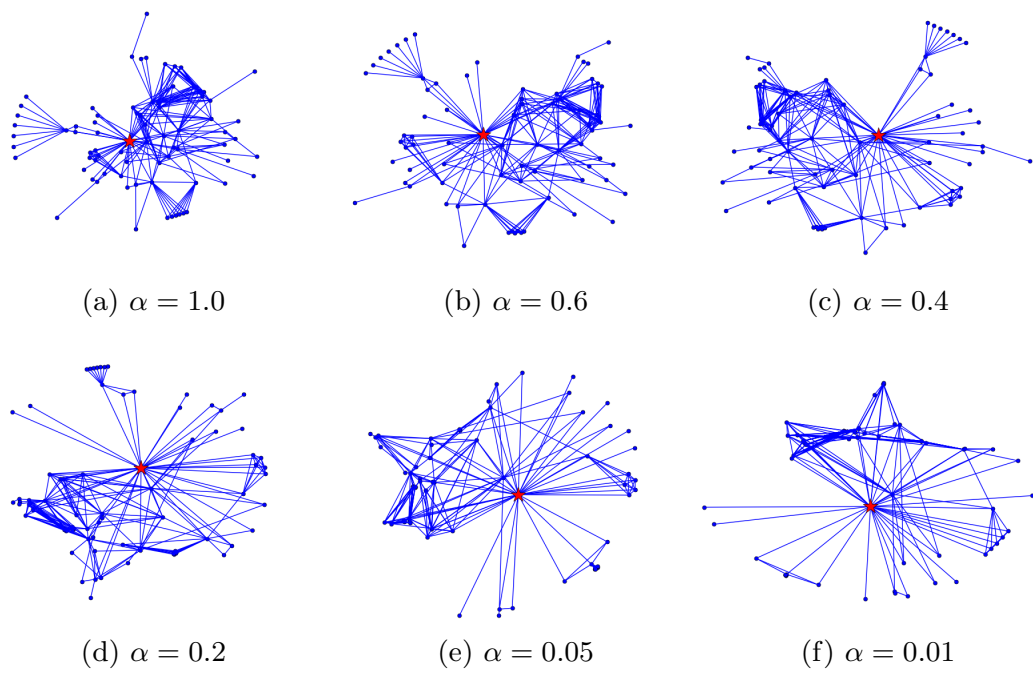


図4.8: Les Miserablesのグラフレイアウトの変化

### 4.2.3 ノードの中心性指標によるレイアウトへの影響に関する考察

ここまで、特定のノードを中心に据えたレイアウト手法が概ね有効であることを定性的に確かめた。本項では、中心に据えるノードの属性がレイアウトにどのような影響を及ぼすかを考察する。

特定のノードを対象としたネットワークの分析では、ネットワークの中で重要なノードはどれかを把握するために、「中心性」という概念が用いられている [9]。本研究では、中心に据えるノードの中心性がレイアウトにどのような影響を及ぼすかについて考察する。

ノードの中心性には多くの種類があるが、本研究では「媒介中心性」と呼ばれる指標を用いる。媒介中心性とは、「グラフ中のあらゆるノードの橋渡しをするノードはグラフの中心である」という観点に基づく。無向グラフを対象とした、ノード  $i$  の媒介中心性は以下の式 (4.2) で表される。

$$C_b(i) = \sum_{i \neq j \neq k} \frac{h_{jk}(i)}{h_{jk}} \quad (4.2)$$

ここで、 $h_{ik}$  はノード  $i$  とノード  $k$  の最短経路であり、 $h_{jk}(i)$  はノード  $j$  とノード  $k$  の間の最短経路のうちノード  $i$  を通るものである。本項では、Les Misérables のグラフ中の媒介中心性指数を計算し、媒介中心性の高いノードと低いノードをそれぞれ中心ノードとして、提案手法によりレイアウトを描画する。対称ノードの id と媒介中心性指数を表 4.5 に示す。

表4.5: ノードの媒介中心性

ノードid	媒介中心性指数 $C_b(i)$
0	1008.0
27	425.1
41	17.1
55	586.0
59	3.4
76	1.3

図 4.9 は、 $\alpha$  を 0.1 としたそれぞれのノードを中心に据えたグラフレイアウトである。中心ノードは赤色で着色してある。それぞれのレイアウトにおいて、中心ノードと遠いノード

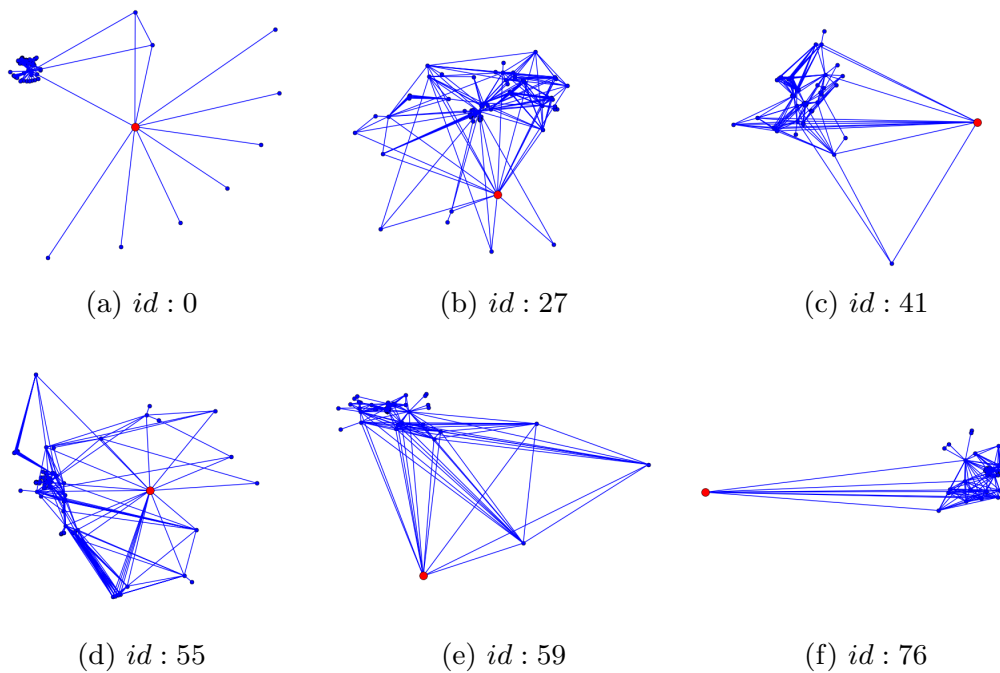


図4.9: 媒介中心性によるレイアウトの変化 (Les Misérables,  $\alpha = 0.1$ )

ド集合の距離関係が縮小して表現されている点は同じだが、中心ノードの描画領域における位置に違いがみられる。媒介中心性の高いノードを対象としたレイアウト(特に、4.9a, 4.9d)では中心ノードは描画領域の中心に存在するが、中心性の低いノードを対象としたレイアウトは、中心ノードは描画領域の端に配置されるようになった。グラフレイアウトを見ると、媒介中心性の高いノードは、ほかのノードと接続していないノード(次数中心性の低いノード)と単独で接続していることが視覚的に分かる。ネットワーク分析では、媒介中心性の高いノードと直接リンクしているノードに着目する分析も考えられるため、本手法がそのようなノード集合をレイアウトから抽出するなどの応用が期待できる。

## 第5章

# ネットワーク分析への応用

本章では，提案した手法による具体的なネットワーク分析への応用を目指した研究について詳述する．本研究では，データマイニングと自然言語処理の複合技術であるテキストマイニングに用いられている，語と語のつながりを表現する「共起ネットワーク」を構成するグラフを対象に，提案手法を用いて描画した．得られたレイアウトの変化が第4章での描画結果に関して述べた傾向と同じであることを確認するとともに，提案手法の有用性について，テキストマイニングの作業の簡便化の観点から議論する．

## 5.1 テキストマイニングと共起ネットワーク

提案手法の応用対象である共起ネットワークの概念と、その背景技術であるテキストマイニングについて簡単に解説する。

### 5.1.1 テキストマイニング

テキストマイニングとは、あるテキストデータを（形態素解析や構文解析などによって）データマイニングの手法で分析可能なデータ形式に整形するとともに、文章を単語や文節で区切り、語の出現頻度や語と語のつながりを定量的に分析する技術である [18].

近年、テキストマイニングが様々な研究課題に応用されている [19, 20]. また、近年のIoT (Internet of Things) の技術的発展や、TwitterなどのSNSの流行を背景に、大量のテキストデータを収集する技術が確立されてきた。そのため、今後テキストマイニングの技術はますますその重要性を強めていくと考えられる。

### 5.1.2 共起ネットワーク

テキストマイニングの分析手法は、文中に頻出する語の抽出、似たような文書の群を分類するクラスター分析など各種知られているが、その一つに、文中から文を特徴づける語の抽出を行い、抽出された語と語のつながり（共起関係）を可視化する「共起ネットワーク」がある。

語と語の共起関係は、「同じ文の中で対象の語がどれくらい出現するか」という基準で評価される。例えば自由記述のアンケート文書において「良い」、「楽しい」などの肯定的な語と特定のイベント名や人物名が共起関係を示していれば、そのイベントや人物について肯定的な文が頻出していることが定量的に評価できる。

文章中の語の共起関係を取得するための代表的な手法として「Jaccard係数」が知られている [21]. Jaccard係数とは、語 $\omega$ と語 $\omega'$ の共起関係を示し、大きいほど共起が強いことを示す。 $p(\omega \cap \omega')$ は語 $\omega$ と語 $\omega'$ の共起確率を表し、 $p(\omega \cup \omega')$ は語 $\omega$ と語 $\omega'$ のいずれかが出現する確率を示す。語 $\omega$ と語 $\omega'$ のJaccard係数は以下の式(5.1)で与えられる。

$$Jaccard(\omega, \omega') = \frac{p(\omega \cap \omega')}{p(\omega \cup \omega')} \quad (5.1)$$

式 (5.1) から、 $Jaccard(\omega, \omega')$ は0から1の値をとる。

文章の共起関係を分析する際には、語と語の個別の共起関係を一つ一つ抽出するより

も、文書全体の共起関係を可視化するほうが分析作業の利便性の観点から有用である。そこで、文章中のJaccard係数が強い共起関係を一定数取り出し、ネットワークとして可視化する手法がとられる。

共起ネットワークを可視化することは、語をノード、語と語の共起関係をエッジに置き換えたグラフを描画することと同義である。共起関係は向きを待たないことから共起ネットワークは無向グラフに分類される。また、既存のテキストマイニングソフトには共起ネットワークのレイアウトに力学モデルが用いられている [22]。したがって、共起ネットワークの可視化には本研究で提案した手法が適用できる。

## 5.2 描画結果と考察

第4章で提案した手法により、テキストマイニングソフトで生成した共起ネットワークを描画する。また、得られた描画結果から提案手法の分析作業への有用性について考察する。

### 5.2.1 対象ネットワーク

本研究では、フリーのテキストマイニングソフトであるKH coder [22]を用いて、共起ネットワークを構成するグラフを描画する。

KH coderのチュートリアルに用いられる、夏目漱石の『こころ』のテキストデータから、頻出する上位100語のうち、Jaccard係数が上位40の共起関係を表す共起ネットワークを生成した。生成したネットワークは、ノード(語)が21、エッジ(共起関係)が40の無向グラフとなる。また、得られたグラフは非連結な成分を持たない連結グラフである。

### 5.2.2 描画結果

本研究では、中心ノードを「K (『こころ』に登場する主要人物)」として、第4章の提案手法によりグラフを描画した。図 5.1, 図 5.2は、パラメータ $\alpha$ を1.0, 0.6, 0.4, 0.2とした共起ネットワークのレイアウトである。得られたネットワークの変化は、4.2.2で述べたことと同じような傾向を示している。

### 5.2.3 分析への応用性に関する考察

図 5.1, 図 5.2からわかるように, レイアウトの変化によって「K」と共起関係にある語(作中にて恋仲にある「お嬢さん」など)が強調された。また, 「K」と共起関係のない語同士の共起関係は縮小して表示され, これらは, 視覚的にはあまり情報を持たないノードの集合とみなすことが出来る。共起ネットワークを分析する際, ネットワークに含めるための共起条件を緩めてしまうと, あまり重要でない語の共起関係もネットワーク化してしまうため, エッジが多く分かりにくいレイアウトになってしまう。そのため, はじめに緩い条件でネットワークを可視化し, 徐々に条件を厳しくしていくことで重要な共起関係を可視化していく作業が行われる。提案手法を用いて, 分析対象としたい語と共起関係のない語の共起関係を縮小させてレイアウトすることで, 分析作業の簡便化に貢献することが期待される。

本研究では, 非連結成分のない連結グラフとして構成されている共起ネットワークを対象に描画した。共起ネットワークの出現条件は, 語と語の共起関係からのみ設定されるため, 一般的な共起ネットワークには独立した共起関係が構成要素になることがある。KH coderなどのテキストマイニングソフトは, 非連結なグラフを考慮してレイアウトすることができるため, 本手法を非連結なグラフに応用することが今後の課題となる。

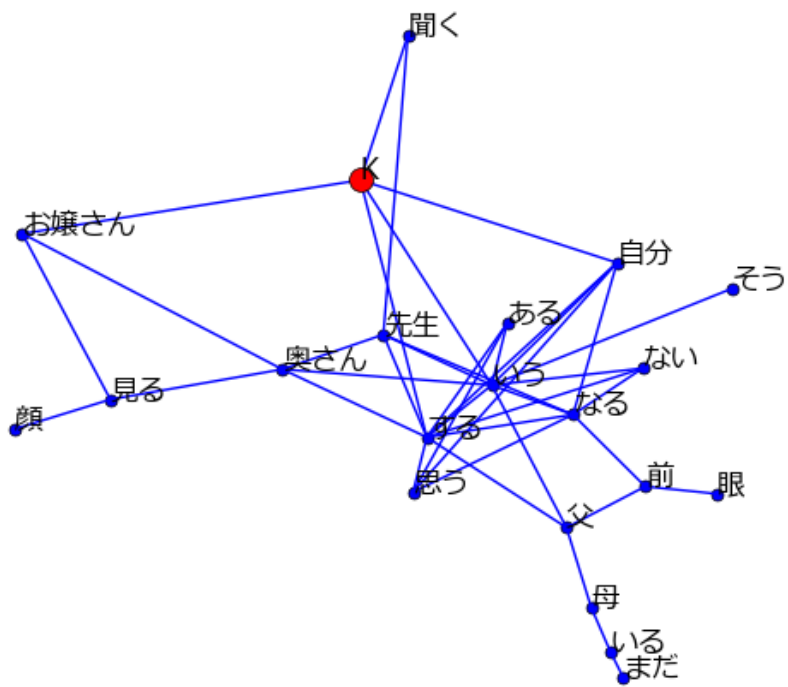
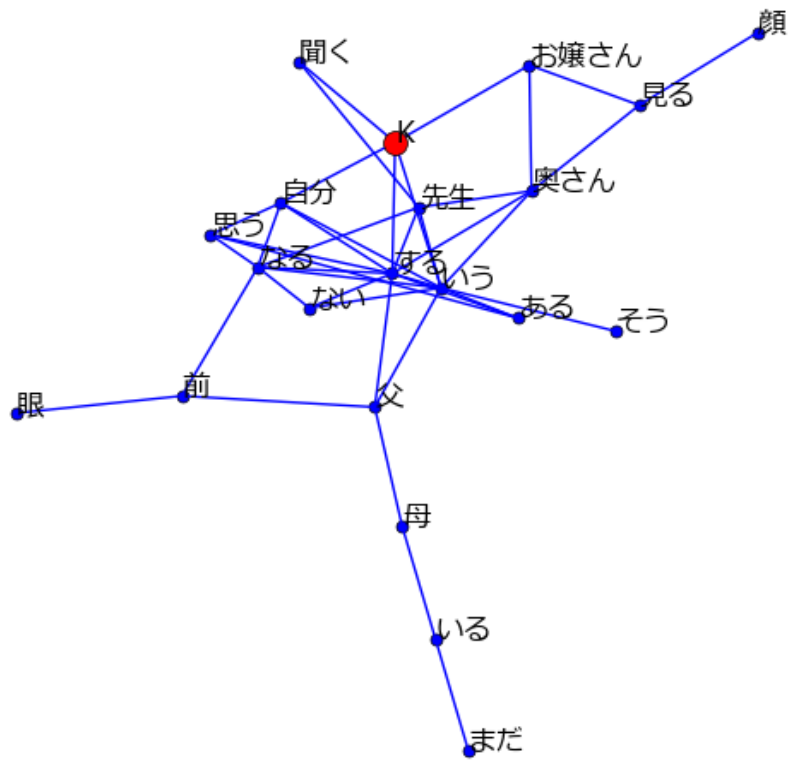
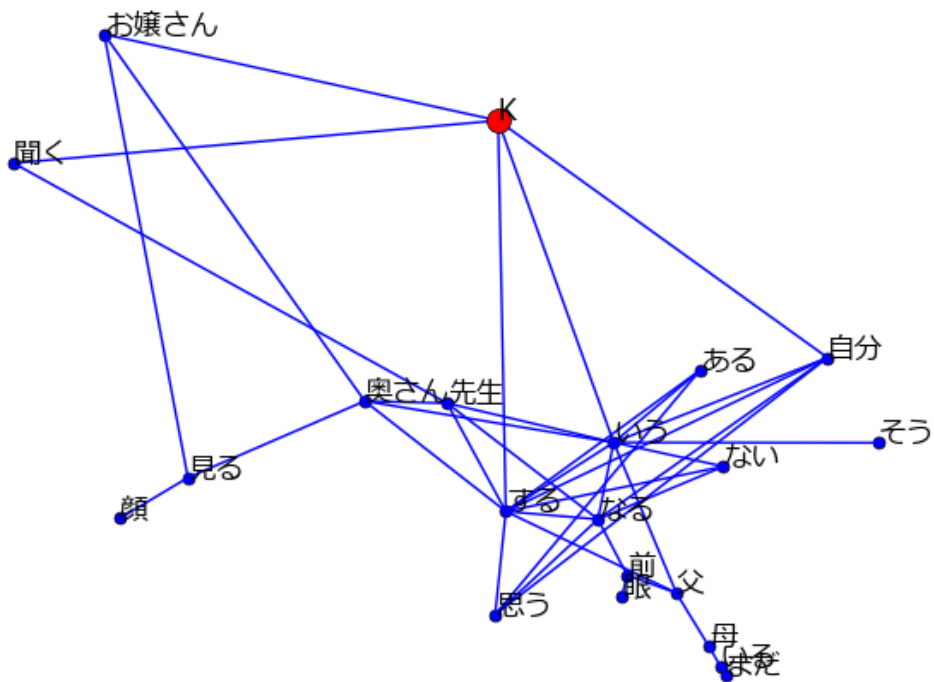
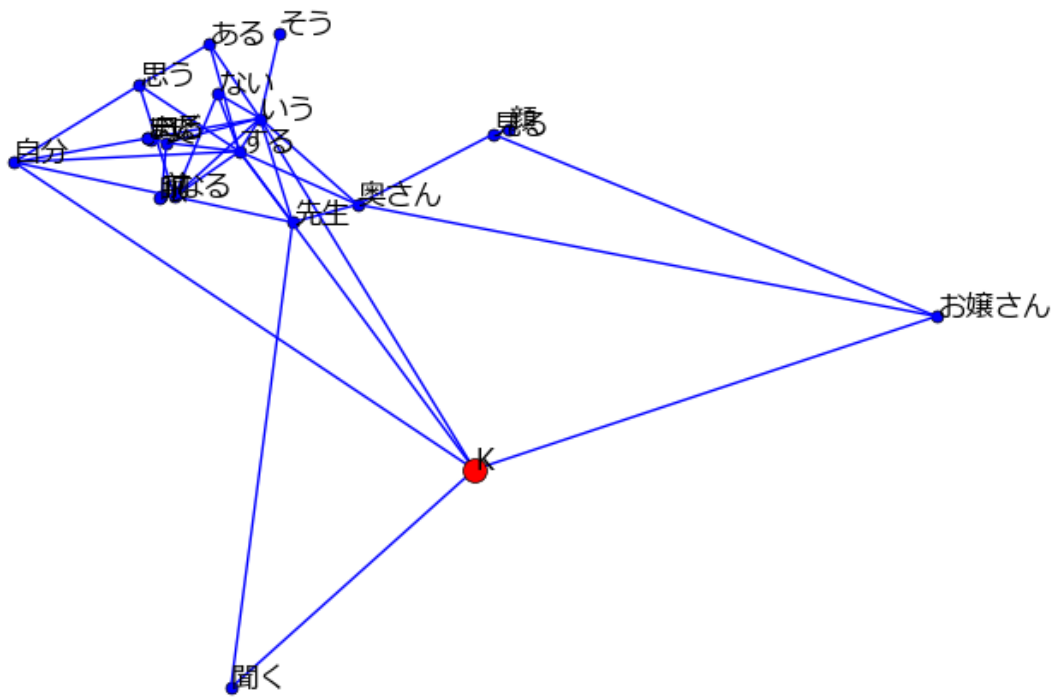


図5.1: 『こころ』の共起ネットワーク  $\alpha = 1.0, 0.6$ のレイアウト



(a)  $\alpha = 0.4$ (b)  $\alpha = 0.2$ 図5.2: 『こころ』の共起ネットワーク  $\alpha = 0.4, 0.2$ のレイアウト

## 第6章

# 結論

本章では、本論文の各章の総括として、本研究の動機、関連研究を踏まえた本研究の位置づけ、具体的な研究手法と各種研究成果についてまとめる。また、本研究では、提案手法から得られた結果の考察から、いくつかの課題を発見した。本章では、課題を踏まえた今後の展望についても述べる。

## 6.1 研究のまとめ

ネットワーク分析に用いられるグラフ描画手法は、ノードの拡大やラベル付けなどの応用がなされているが、レイアウトはグラフ全体のレイアウトを志向したものが主流である。一方、ネットワーク分析手法には、特定のエッジやエッジを対象としたものが知られており、特定のノードやエッジを強調させるためにレイアウトを変化させる手法が要請されているという背景から、本研究では、代表的なグラフ描画手法である力学モデルを基礎技術として、特定のノードを中心に据えるグラフ描画手法を提案した。具体的には、KK法に用いられている目的関数にパラメータを導入することで、ノード間に作用する引力および斥力を調節することでノードを移動させ、レイアウトを変化させた。この手法は、KK法の目的関数に導入されている仮想的なばねの自然長に操作を加えることでグラフレイアウトを変化させた既往研究 [14]からアイデアを得たものである。本研究では、中心に据えたいノードとの距離情報をバネの自然長に反映させるアプローチをとった。提案手法は、既往研究のようにグラフレイアウトを完全に変える手法ではなく、KK法の単純な拡張法であり、パラメータによって従来のKK法を用いたグラフ描画手法によるレイアウトを完全に保全することが出来る点において特徴がある。

はじめに基礎研究として、純粋なKK法と同じ手法でグラフを描画させた。KK法では、ノード同士が引力および斥力を及ぼし合うバネの力学系のスカラーポテンシャルをNewton法を用いて最小化することで、グラフレイアウトを求める手法をとっていたが、最近の研究により、KK法で定式化された力学系のスカラーポテンシャルの考え方が多次元尺度構成法の分野で用いられているstress関数という損失関数と同じであることが発見され、より高速に解を求められるStress Majorizationが採用されている。本研究では、グラフ描画の基本的な考え方を与え、ノード間のグラフ理論的距離と二次元平面上の距離の誤差を最小化させる問題として定式化し、これがKK法と同じ形の関数であることから、Stress Majorizationを用いて最小化問題を数値的に説いた。得られて解によって実際にいくつかの簡単な構造のグラフと、一般的な構造のグラフを描画した結果、近接性、対称性を顕示する良好なグラフレイアウトを得た。単純な構造のグラフレイアウトでは、想像されるレイアウトよりもゆがんだレイアウトが得られることについての考察を行った。

KK法によって得られたグラフレイアウトを踏まえて、特定のノードを中心に据えるための拡方法を提案した。特定のノードを中心に据えたグラフレイアウトでは、中心ノードと近いノード集合の関係性は拡大してレイアウトされ、遠いノード集合の関係性は縮小

してレイアウトされることを描画仮説として与え、KK法の目的関数に用いられているバネの自然長を、中心ノードとの距離により変化させた。提案手法により、KK法によって描画したグラフを対象に描画し、各レイアウトの変化を確認した。提案手法は、それぞれの異なる構造のグラフにおいて概ね描画仮説を満足する結果を得た。また、提案手法は、中心性の高いノードを対象に描画させることで、ネットワークの分析手法に有用なレイアウトを得ることを確認した。

提案手法が概ね有効に働くことを確認したことを踏まえ、本研究では具体的なネットワーク分析への応用を目指し、テキストマイニングに用いられている、語と語のつながりを示す「共起ネットワーク」のレイアウトを提案手法により変化させた。共起ネットワークは力学モデルを用いて描画されるため、提案手法が適用できる。描画実験により、特定のノード（語）との共起関係を強調して示すことができた。また、この結果について共起ネットワークの分析作業の観点から考察し、提案手法がネットワーク分析の手法に有用性を持つ一例を示した。

## 6.2 課題と今後の展望

提案手法による描画では、中心性の高いノードを中心に描画すると良好なレイアウトが得られたが、中心性の低いノードを中心ノードとするとレイアウトの中央にノードが配置されず、レイアウトの端部に配置されてしまった（図 4.9）。提案手法の普遍的な有用性を担保するために、ノードの様々な属性に合わせた距離の修正方法を考える必要がある。

また、第 5 章にて共起ネットワークの描画を行ったが、共起ネットワークを構成するグラフは、一般的には非連結である。本手法では、非連結なグラフは描画することが出来ないため、さらなる拡張が必要である。

最後に、本研究ではKK法の拡張のみを扱ったが、FR法についても特定のノードを中心に据えるためレイアウトを変化させる手法が考えられる。FR法では、隣接しているノードに引力が、それ以外のノードには斥力が働くモデルを採用しているが、引力および斥力によりノードが移動する距離を制限するために以下の式 6.1を用いている。

$$l = C \sqrt{\frac{\text{描画領域}}{\text{ノードの数}}} \quad (6.1)$$

本研究の提案手法の様に、この理想距離について、中心ノードからの距離の情報に基づいたパラメータを導入することが考えられる。

## 参考文献

- [1] P.Eades, A Heuristic for Graph Drawing, *Congressus Numerantium*, 42, pp.149-160, 1984.
- [2] T.Kamada and S.Kawai, An Algorithm for Drawing General Undirected , *Information on Processing*, 31-1, pp.7-15.
- [3] T.Fruchterman and E.Reingold, Graph Drawing by Force directed Placement, *Software Practice*, 21-11, pp.1129-1164, 1991.
- [4] " igraph", <http://igraph.org/redirect.html>(2017.01.30閲覧).
- [5] " Cytoscape", <http://www.cytoscape.org/>(2017.01.30閲覧).
- [6] " Gephi", <http://oss.infoscience.co.jp/gephi/gephi.org/>(2017.01.30 閲覧).
- [7] 三末和夫, ネットワークの可視化技術—大規模ネットワークと動的ネットワークへの挑戦—, *電子情報通信学会誌*, volume.92, No.2, pp.112-117, 2009.
- [8] 篠田孝裕, 大規模ネットワークデータの収集と分析, *理論と方法*, volume.27, No.1, pp.175-186, 2012.
- [9] 鈴木 努, Rで学ぶデータサイエンティスト8 ネットワーク分析, 共立出版, 2009.
- [10] 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.
- [11] E.Gansner, Y.Coren, and S.North, Graph Drawing by Stress Majorization, in *Graph Drawing*, ser.Lecture Note in Computer Science, pp.239-250, 2005.
- [12] Y.Wang and Z.Wang, A fast successive over-relaxation algorithm for force-directed network graph drawing, *Science Chine Information Sciences*, 55-3, pp677-688, 2012.
- [13] H.Hosobe, Numerical optimization-based graph drawing revisited *PacificVis*, 2012, IEEE, pp.81-88, 2012.
- [14] Y.Coren and A.Civril, The Binary Stress Model for Graph Drawing, *GD 2008*,

- Graph Drawing, volume.5417 of LNCS, pp.193-205, 2008.
- [15] 竹之内洋樹, 塩田茂雄, レイアウトを変化させるグラフ自動描画手法, 電子情報通信学会基礎・境界ソサイエティ/NOLTAソサイエティ大会講演論文集, A-1-13, p13, 2015.
  - [16] S.Shioda, K.Shimamura, Relative localization of sensors based on their responses to moving objects, IEEE Commun.Mag., volume.46, pp.112-120, 2008.
  - [17] ” Network data repository” , <http://networkdata.ics.uci.edu/index.php>(2017.01.30閲覧)
  - [18] 樋口耕一, 社会調査のための計量テキスト分析—内容分析の継承と発展を目指して, ナカニシヤ出版, 2014.
  - [19] 戸田浩之, 北川博之, 藤村考, 片岡良治, 奥雅博, グラフ分析を利用した文書集合からの話題構造マイニング, 電子情報通信学会論文誌D, Volume.J90-D, No.2, pp.292-310, 2007.
  - [20] 森田哲夫, 入澤覚, 長塩彩夏, 野村和弘, 塚田伸也, 大塚裕子, 杉田浩, 自由記述データを用いたテキストマイニングによる都市のイメージ分析, 土木学会論文集, volume.68, pp.315-323, 2012.
  - [21] 佐々木靖弘, 佐藤理史, 宇津呂武仁, 関連用語収集問題とその解法, 自然言語処理, volume.13, No.3, pp.150-175, 2006.
  - [22] ” KH Coder”, <http://khc.sourceforge.net/>(2017.01.30閲覧)

## 研究成果

- 竹之内洋樹, 塩田茂雄, ”レイアウトを変化させるグラフ自動描画手法,” 電子情報通信学会ソサエティ大会, 2015 (宮城).

## おわりに

本研究を進めるにあたり，ご多忙の中ご指導・ご鞭撻を頂きました，塩田茂雄教授に深く感謝を申し上げます。研究ゼミを通じて，研究の方向性や，提案手法の改善法について沢山のアドバイスを頂きました。また，塩田研究室の諸先輩後輩の方々，同輩達には，大変お世話になりました。

ありがとうございました。