

博士論文

構造が創発する相互作用

～網羅的ネットワークデータセット構築手法の提案～

臼井翔平

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	背景 . . . . .	1
1.1.1	複雑ネットワーク上で観測される現象 . . . . .	1
1.1.2	実世界に存在する様々な複雑ネットワークの構造 . . . . .	5
1.1.3	ネットワーク構造と現象の関係 . . . . .	8
1.2	本研究の目的と位置づけ . . . . .	10
1.3	本稿の構成 . . . . .	11
<b>第 2 章</b>	<b>既存研究とその限界</b>	<b>14</b>
2.1	はじめに . . . . .	14
2.2	複雑ネットワークの構造の基礎 . . . . .	14
2.2.1	スモールワールド . . . . .	14
2.2.2	スケールフリー . . . . .	15
2.2.3	ネットワーク生成モデル . . . . .	17
2.2.4	ネットワークの構造特徴 . . . . .	18
2.3	従来手法とその限界 . . . . .	20
2.3.1	情報拡散に関する比較研究 . . . . .	20
2.3.2	協調行動に関する比較研究 . . . . .	23
2.4	既存の分析の限界について . . . . .	25
2.4.1	本論文で用いる既存ネットワーク生成モデル . . . . .	25
2.4.2	既存研究で用いられているネットワーク生成モデルの性質 . . . . .	34
<b>第 3 章</b>	<b>任意の構造を含むネットワークデータセットの構築手法の提案</b>	<b>40</b>
3.1	はじめに . . . . .	40
3.2	Arbitrary Network Generation Algorithm . . . . .	41

3.2.1	概要 . . . . .	41
3.2.2	候補ネットワークの生成 . . . . .	42
3.2.3	最適ノードの選択 . . . . .	45
3.2.4	ANGA の評価実験 . . . . .	45
3.3	網羅的ネットワークデータセット . . . . .	56
3.3.1	データセットの構築手法 . . . . .	57
3.3.2	データセットの性能評価 . . . . .	59
3.4	結論 . . . . .	70
<b>第 4 章</b>	<b>情報拡散の統計的分析</b>	<b>72</b>
4.1	はじめに . . . . .	72
4.2	情報拡散モデル . . . . .	73
4.2.1	独立カスケードモデル (Independent Cascade model:IC model) . . . . .	74
4.2.2	線形閾値モデル (Linear Threshold model:LT model) . . . . .	74
4.2.3	ネットワークの拡散率 $\delta$ . . . . .	75
4.3	拡散現象の統計的分析 . . . . .	75
4.3.1	既存研究の検証 . . . . .	75
4.3.2	拡散率 $\delta$ に影響を与える特徴の分析 . . . . .	79
4.3.3	拡散率 $\delta$ の推定実験 . . . . .	86
4.3.4	決定木分析 . . . . .	90
4.3.5	ネットワークサンプル . . . . .	94
4.3.6	IC モデルと LT モデルの拡散率の関係 . . . . .	95
4.4	結論 . . . . .	97
<b>第 5 章</b>	<b>協調行動の統計的分析</b>	<b>100</b>
5.1	はじめに . . . . .	100
5.2	協調行動モデル . . . . .	101
5.2.1	囚人のジレンマゲーム (Prisoner ' s Dilemma game : PD game) . . . . .	101
5.2.2	空間的囚人のジレンマゲーム (Spatial Prisoner ' s Dilemma : SPD) . . . . .	102
5.3	空間的囚人のジレンマゲームの統計的分析 . . . . .	102
5.3.1	既存研究の検証 . . . . .	102
5.3.2	相関関係による分析 . . . . .	108
5.3.3	協調戦略エージェントの割合 $P_c(1.2)$ の推定実験 . . . . .	110

5.3.4	ネットワークサンプル . . . . .	113
5.3.5	協調戦略に対する裏切り戦略が得る利得 $b = 1.2$ と $b = 1.5$ の時 の協調戦略エージェントの割合の関係 . . . . .	114
5.4	結論 . . . . .	118
<b>第 6 章</b>	<b>結論</b>	<b>119</b>
6.1	総括 . . . . .	119
6.2	今後の課題 . . . . .	122
	<b>参考文献</b>	<b>127</b>

# 第 1 章

## 序論

### 1.1 背景

#### 1.1.1 複雑ネットワーク上で観測される現象

我々の暮らす社会は，個人個人が相互に関係する事で形成されている．この個人間の相互関係は複雑ネットワークとみる事が出来る．他にも，交差点を道路で結んだ道路網，企業を取引関係で結んだネットワーク等，世の中には多種多様なネットワークを定義する事が出来る．

この複雑ネットワーク上では多くの現象が観測される．例えば，友人関係ネットワーク上では情報の拡散や意見形成等，企業間の取引関係では連鎖倒産等様々な現象が観測される．これらの現象は社会に対して強い影響を与える．

ソーシャルメディア上に形成される人間関係ネットワークであるオンラインソーシャルネットワーク (Online Social Network : OSN) 上では情報拡散が現在高い注目を集めている．**図 1.1** に，世界中で最も用いられている facebook <sup>\*1</sup> の Monthly Active User 数を示す．facebook のユーザ数は年々増加し，アクティブユーザ数は 15 億人を超えている．すなわち，現在では多くの人がこの OSN に所属している．また，facebook では一日約 47 億件の投稿行われており，ユーザは月に平均 21 時間をそこで過ごしている．facebook 上には非常に多くの情報がシェアされており，特に若年層では facebook を情報源として用いているユーザが多く存在する．facebook における情報の拡散は，テレビ等の既存のメディアとは異なり，ネットワークを介して人から人へと伝わっていく．このような OSN の情報の拡散能力には非常に期待が集まっている．

---

<sup>\*1</sup> <https://www.facebook.com/>

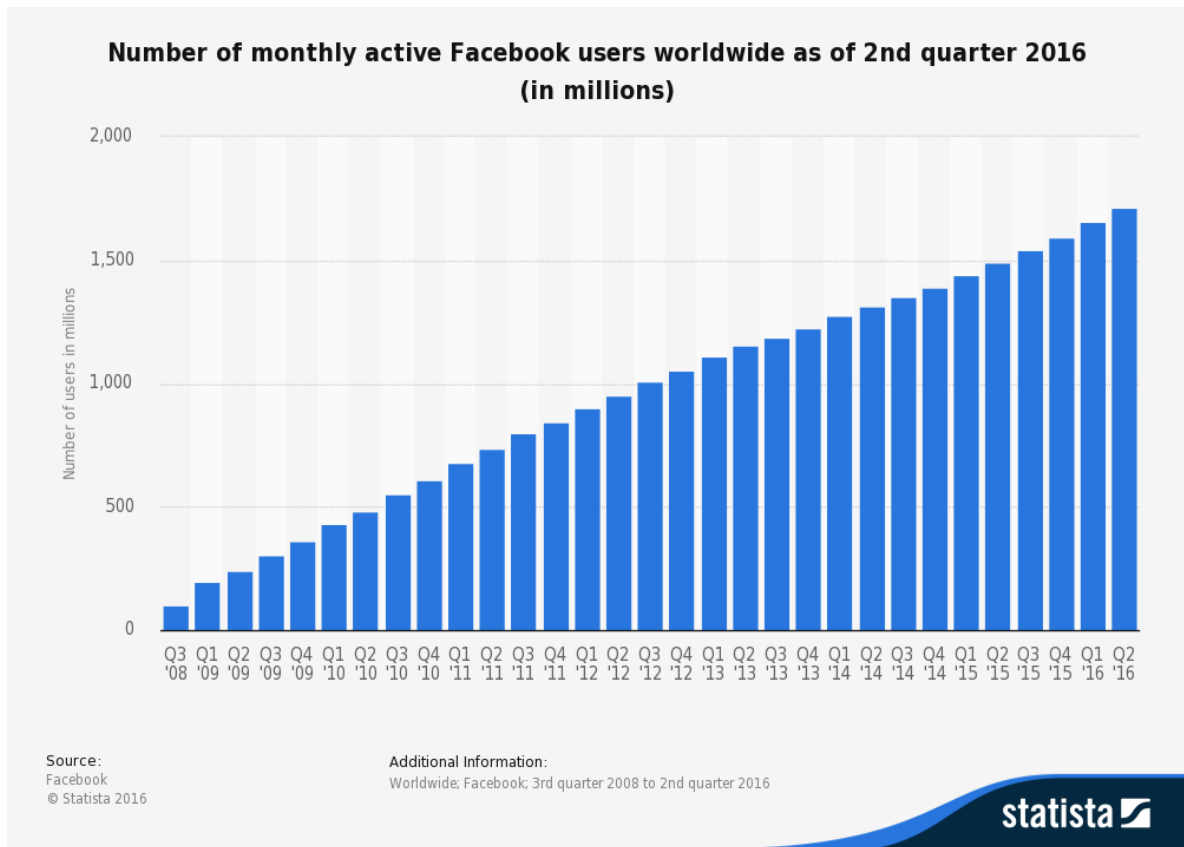


図 1.1: facebook : Monthly Active User 数

国内でも多くの OSN ユーザが、OSN によって情報を収集・発信を行っている [1]。これに伴って日本では、2013 年 4 月 19 日に改正公選挙法によって、これまで禁止していたインターネットを用いた選挙活動、所謂'インターネット選挙運動'の解禁を定めた。現在まだ少数ではあるが、インターネットを利用した選挙運動も行われている [2]。このような動きは政治活動に限らず、OSN 上の情報拡散をマーケティングに利用しようとする動きも数多く存在している [3]。また、ソーシャルメディアの情報拡散能力は、災害のような非常事態時での活躍が期待されている。実際に 2011 年 3 月 11 日に日本で発生した東日本大震災時には Twitter<sup>\*2</sup>が情報拡散に貢献したと報告されている<sup>\*3</sup>。拡散された情報は、避難情報や安否情報に留まらず、災害支援に関する情報に至るまで多岐に渡っていた。

しかし、OSN 上での情報拡散はいい面だけではない。同震災時には、誤情報も多く拡

<sup>\*2</sup> <https://twitter.com/>

<sup>\*3</sup> 総務省：平成 23 年度情報通信白書、  
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/pdf/index.html>

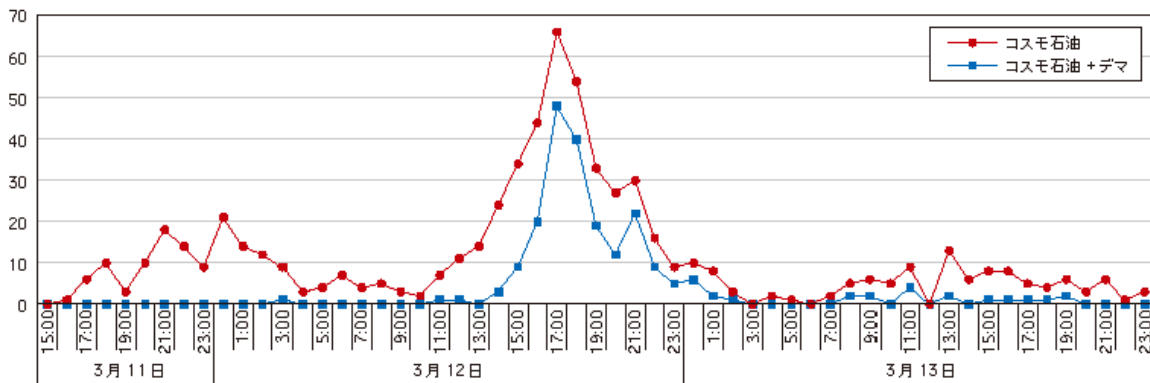


図 1.2: 誤情報の拡散 (出典：平成 23 年度情報通信白書)

散されていた事が報告されている。東日本大震災時には、以下のような誤情報が拡散されている\*4。

「コスモ石油の爆発により有害物質が雲などに付着し、雨などといっしょに降るので外出の際は『傘』か『カッパ』などを持ち歩き、身体が雨に接触しないようにして下さい！！」

この情報は 2011 年 3 月 11 日に拡散され、同 3 月 12 日には、コスモ石油株式会社から公式に「そのような事実はない」とする発表が行われている。図 1.2 には、このデマ情報に関連する tweet が 1 分間に投稿された回数を示している。このように、震災時には誤情報に惑わされたユーザも多く存在する事が示されている。

この誤情報には、悪意のある場合もあるが、悪意がない場合も多い。OSN での情報拡散では、一般のユーザが情報源となる事が多く、多くの場合事実確認は行わない。この事実確認を行わない事が、迅速な情報拡散につながるメリットともなるが、このような誤情報の拡散を許してしまう危険性もはらんでいる。すなわち、ユーザにどれだけ注意喚起しても誤情報の拡散を止める事は難しい。このデマの拡散を防ぐには、ネットワーク構造等のシステムとして防ぐ方法を考案する必要がある。このデマの拡散に関する研究は多く行われており、デマの拡散を防ぐ方法が検討されている [4, 5, 6]。このように OSN 上の情報拡散は良い悪いにせよ、社会に大きな影響を与えていると言える。

OSN 上での情報拡散の特徴は個々のユーザが情報を発信可能であることである。これは、個人からの情報の発信が低コストで可能であるためである。このような自由な環境で

\*4 出典：平成 23 年度情報通信白書

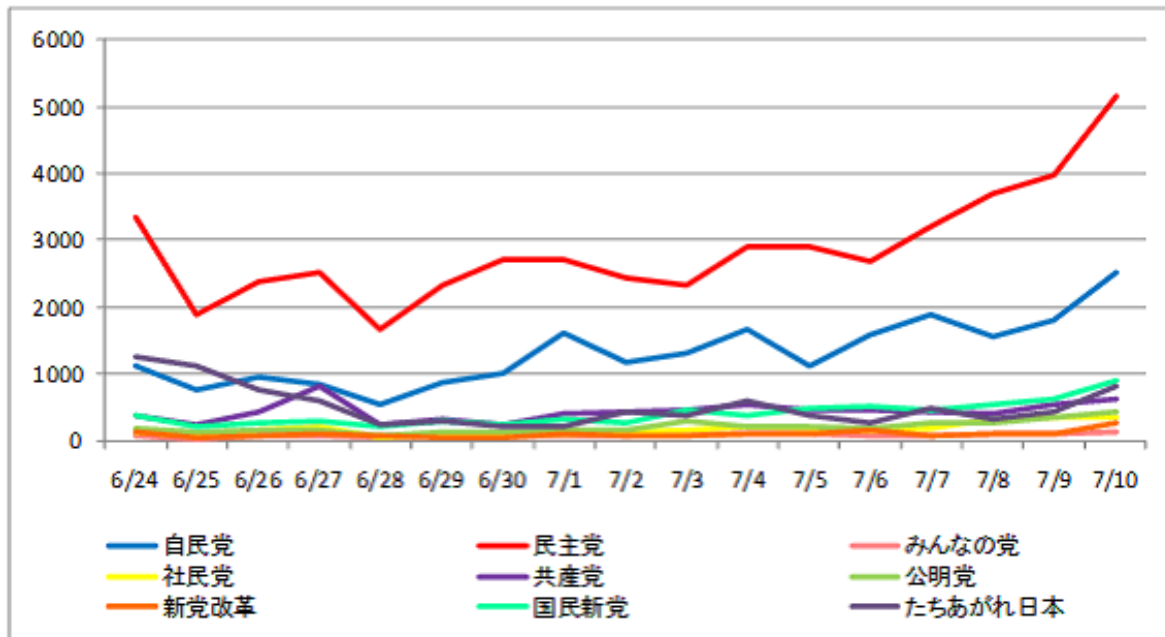


図 1.3: Twitter 上での各政党に関する話題調べ (出典：クチコミ@参院選 2010 Twitter 分析レポート)

あれば、政治的、思想的な議論が行われる事も必然と言える。OSN 上では、毎日のように様々な議論が展開されている。現在、国内で人気のソーシャルメディアである Twitter では、政治に関する単語で検索をかけると、多くの情報が拾える。図 1.3 は、2010 年の参議院選挙の際に Twitter 上に投稿された、各政党に関する tweet 数である<sup>\*5</sup>。Twitter 上で、政治に関する話題が多く扱われていた事がよくわかる。

しかし、この OSN 上の議論に関しても、メリットばかりではない。一つの問題点として挙げられるのが、意見の二極化である。この原因として、ネットワーク構造が挙げられている。現実社会のネットワークでは、意見の違う人とも半ば強制的に関係を結ぶ(ネットワーク的にはリンクを張る)必要がある。しかし、OSN 上では、意見の違う人と関係を結ぶ必要はほぼ皆無である。すなわち、同じ意見の人同士がリンクを結ぶネットワークが構築される。このようなネットワーク上では、自分の周りにはほぼ全員が同じ意見の持ち主となる。この環境で議論を行えば、当然意見は増長される傾向を見せる。こうして、OSN 上では、意見が増長され、極端な意見へと二極化するのである。このような意見の二極化は危険な思想を生み出したり、炎上のような他者への強い攻撃性へとつながるため非常に

\*5 出典：クチコミ@参院選 2010 Twitter 分析レポート



危険な現象として、現在も研究が進んでいる [7, 8, 9].

OSN のような社会ネットワーク以外で発生する現象についてもここで触れよう。近年、問題となっている連鎖倒産という言葉がある。現在の金融システムでは、多数の金融機関が様々な金融商品を介して複雑なネットワークを形成している。このようなネットワークは金融ネットワーク (Financial network) と呼ばれ、一部の研究者から注目を集める話題となっている。この金融ネットワーク上に、外的なショックが加わる。この時、外的ショックが加わった金融機関がデフォルト (債務不履行) を起こしたとする。この不履行となった債権を直接的に保有する金融機関に損失が発生する。このショックが非常に大きい時、このショックは金融ネットワーク上を波及し、連鎖的にデフォルトが発生する。これを連鎖倒産 (Chain bankruptcies) と呼ぶ。

連鎖倒産の危険性を語る上で、最も重要な例は、2007 年に発生したリーマンショック (The financial crisis) であろう。2007 年に発生したサブプライムローン問題を契機に、アメリカの投資銀行であるリーマン・ブラザーズ (Lehman Brothers) が破綻した。このリーマン・ブラザーズの金融破綻が取引先の企業等へ波及する事に対する恐れから、アメリカ経済への不安が広がり、世界的な金融危機に陥った。図 1.4 にリーマンショック前後の各国の平均株価を示す\*<sup>6</sup>。リーマンショックを機に、各国の平均株価が暴落している事がよくわかる。この世界的恐慌以降、様々な研究者がこの連鎖倒産の危険性について訴え、研究を進めている [10, 11, 12]。この連鎖倒産は、ネットワークの構造によって起こる、システムック・リスクとされ、どのようにネットワークを構築すればこのようなリスクが回避されるのかは、現在も研究が進んでいる [13]。

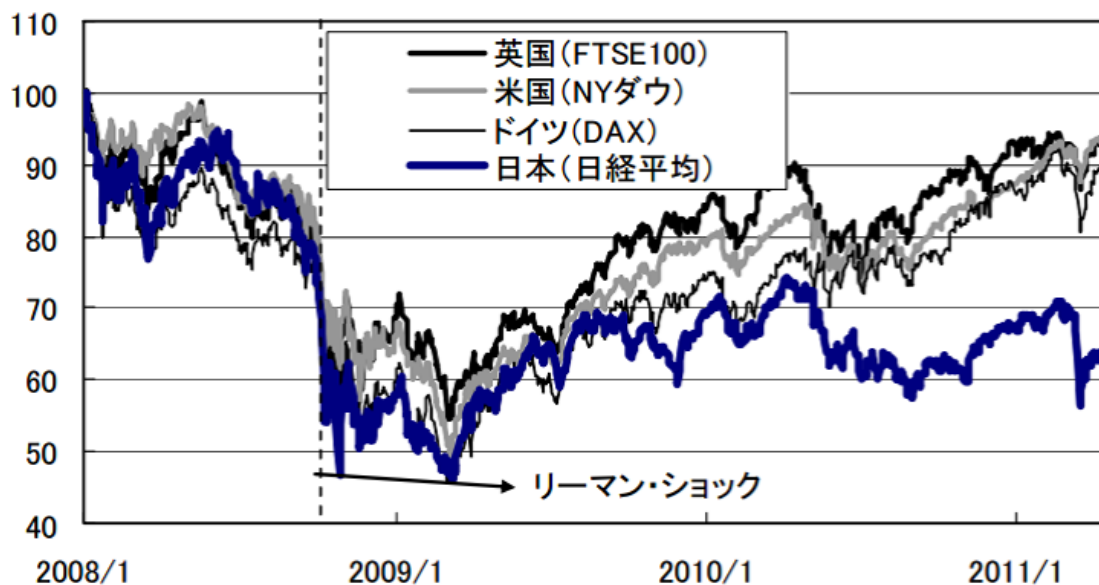
このように、ネットワークの上では様々な現象が発生している。これらの現象を理解する事は社会の安定化に重要な役割を果たす。これらのネットワーク上で起きる現象は、ネットワークの構造から強く影響を受ける。従って、これらの現象を理解するためには、ネットワークの構造を理解する必要があると言える。

### 1.1.2 実世界に存在する様々な複雑ネットワークの構造

複雑ネットワークとは一般に点 (ノード:Node) と線 (リンク:Link) で表記される。様々なものとその関係を点と線として抽象化する事によって、多種多様な関係性を表記できる。このように抽象化する事による大きなメリットの一つは、複雑に見える問題を抽象化し、単純なネットワークという形に書き表す事によって、分析を容易にできる事である。

---

\*<sup>6</sup> 財務省レポート：リーマン・ショック後の経済金融危機における財政投融资の対応



(出所)ブルームバーグ

図 1.4: 出典：リーマンショック前後の各国の平均株価 (財務省レポート：リーマン・ショック後の経済金融危機における財政投融资の対応)

また、ネットワークという統一された形式で表記する事によって、これまで蓄積され、体系化された様々な知識が利用可能になる。多種多様なネットワークはそれぞれ独自のネットワークの構造を取っており、同じ構造のものは一つとしてないと言っても過言ではない。複雑ネットワークがどのような構造を持っているかを理解する事は、ネットワーク研究における大きなテーマである。

図 1.5 は、アメリカの航空網を表したネットワークである\*7。このネットワークはノードを各空港とし、航空機の空路をリンクとしたネットワークである。このような、地理的特性を持つネットワークには、実距離が考慮される事が多い。実距離が近いノード同士はリンクが張られやすく、遠い程にリンクは張られにくい。従って、中心に位置するような空港がハブ (次数を集めやすいノード) になりやすいという特徴を持つ。

図 1.6 は食物連鎖のネットワークである。このネットワークはノードを動物や植物とし、食物連鎖の関係をリンクとした食物網である\*8。食物網では、木構造に近いネットワークとなる。すなわち、図の下から上に向かって捕食関係のリンクが伸びており、逆向

\*7 <http://tokyoexpress.info/>

\*8 <http://joshmaskillbiomeproject.weebly.com/food-web-interaction-with-explanation.html>

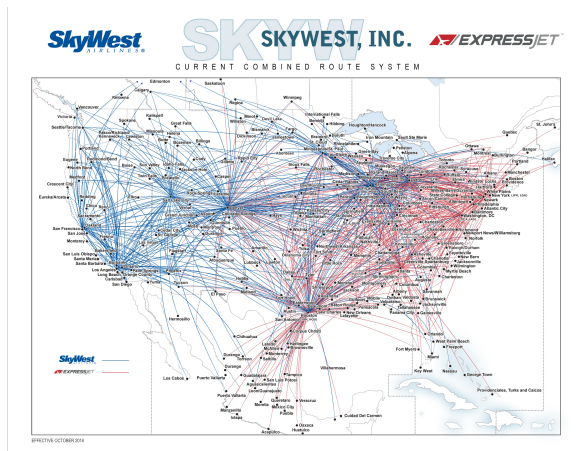


図 1.5: アメリカの航空網 (引用 : Tokyo Express)

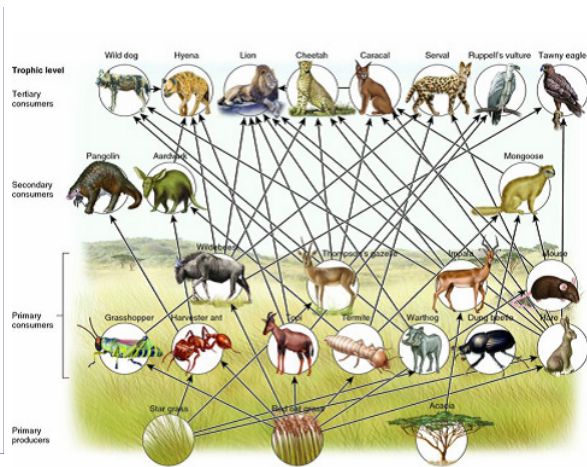


図 1.6: 食物網 (引用 : Savannah plains)

きのリンクはほとんどないといった特徴を持つ。

また、他のネットワークの例として、ウェブ上に形成されたネットワークがある。インターネット上には多くのウェブサイトが存在し、それらは互いにリンクを結んでいる。これらのウェブサイトをノードとしてネットワークが形成される。図 1.7 はウェブ上のサイト間ネットワークである<sup>\*9</sup>。このウェブネットワークは蝶ネクタイ型 (Bow-Tie) 構造を取る事が知られている [14]。蝶ネクタイ型構造とは、図 1.9 のように、リンクを一方的に出すノード群、リンクを一方的に受けるノード群、それ以外の強連結成分の 3 種類に分かれるという構造である。

また、図 1.8 は、amazon<sup>\*10</sup>における、商品をノードとした商品推薦ネットワークである<sup>\*11</sup>。商品推薦ネットワークでは、商品がカテゴリ化されているため、カテゴリ毎に密な連結成分 (サブネットワーク) が形成され、カテゴリ間が疎にリンクされるという構造を持っている。

ここまで例示してきたように、実世界には様々な複雑ネットワークが存在している。これらの複雑ネットワークはそれぞれ特徴のある構造をしており、様々なネットワーク構造が存在する。これらのネットワークは、ノード間の相互関係によって構築される複雑システムである。これらのシステム上では様々な現象が発生する。例えば、航空網の上では感

<sup>\*9</sup> <http://www.opte.org/>

<sup>\*10</sup> <https://www.amazon.co.jp>

<sup>\*11</sup> <http://www.annehelmond.nl/2012/02/22/visualization-amazon-recommendation-network-for-software-studies-a-lexicon/>

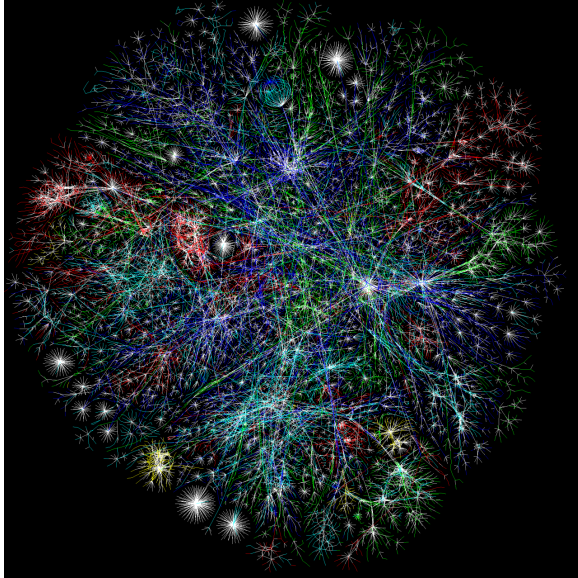


図 1.7: Web ネットワーク構造 (引用 : opte project)

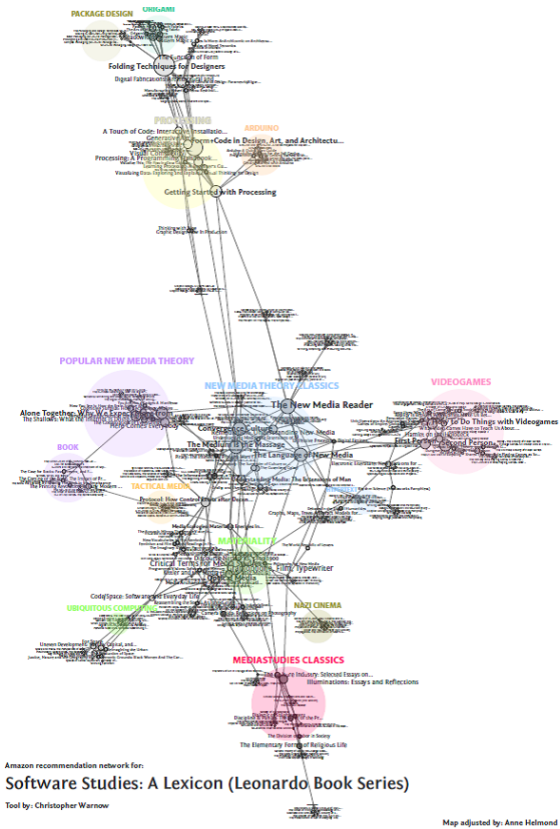


図 1.8: amazon の商品ネットワーク (引用 : Anne Helmond)

感染症の伝播 [15], 食物網の上では種の進化 [16] 等の現象が起こる. ネットワーク構造と現象の間には密接な関係があり, これらを結び付けて考える事は, 現象を理解するために必須であると言える.

### 1.1.3 ネットワーク構造と現象の関係

ネットワーク上で起きる現象は, 個々のノードの相互関係によって起きるため, その多くはネットワーク構造の影響を強く受ける. 例えば, 情報拡散では, 拡散が行われるネットワークの構造によって, 大きく拡散される事もあれば, 拡散されにくい事もある. どのようなネットワーク構造であれば拡散されやすいのかは, 多くの研究者の興味を引いている [17, 18, 19]. ネットワーク構造によってどのような拡散の仕方をするのかが明らかになれば, 情報が拡散しやすいネットワークや, デマを止めるためのネットワークを構築する指針になり得る.

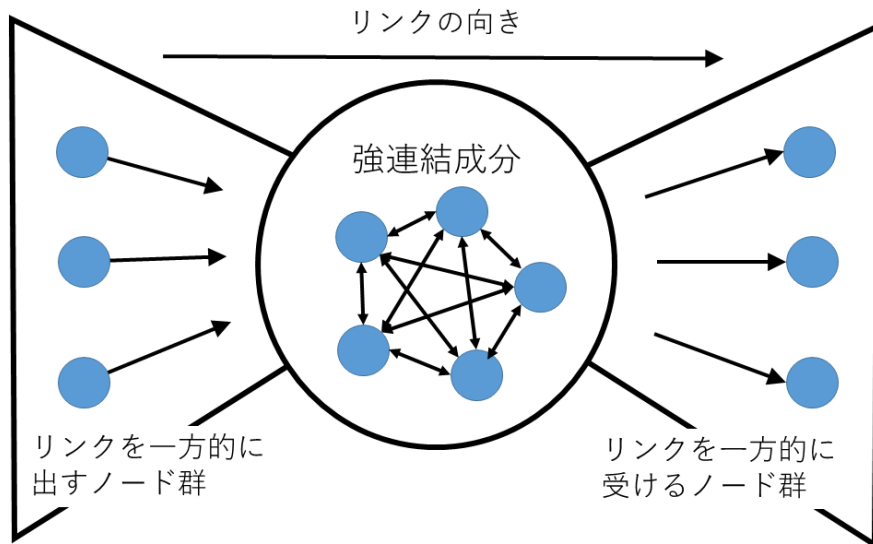


図 1.9: 蝶ネクタイ型構造

また、Nowak らは、エージェント集団にネットワークが存在する事で協調行動が維持される事を示している [20]. この Nowak らの発見から、協調行動とネットワーク構造の分析も多くの研究者の興味を引いている [21, 22]. これによって、企業間の協調が維持される関係の構築方法がわかれば、より大きな利得を生み出す事が期待される。

情報拡散と協調行動について、現象とネットワーク構造の関係の分析は歴史が古く、数多くの研究が存在している. その他にも、サービスネットワークの構造と耐衝撃性、耐攻撃性の関係についての分析 [23, 24, 25] や、企業間取引関係の構造と連鎖倒産の関係についての分析 [10, 11, 12] 等が扱われている. サービスネットワークの構造と耐衝撃性、耐攻撃性の関係についての分析は、ネットワークが障害に陥り、ノードやリンクが欠損した時に、ロバストな構造や、レジリエントな構造を探すものである. また、企業間取引関係の構造と連鎖倒産の関係についての分析は、リーマンショック以降特に危惧される連鎖倒産のシステミック・リスクを減らすためのものである.

これらの分析では、ネットワーク生成モデルを用いて生成したネットワークにより比較を行っている事が多い. 例えば、Erdős-Rényi(ER) グラフ [26] や Watts-Strogatz(Ws) モデル [27], Barabási-Albert(BA) モデル [28] のような基本的なネットワーク生成モデルが用いられている. これらのネットワーク生成モデルは、特定の構造を持つネットワークを

生成する．そのためネットワークが取り得る構造全体から考えれば極一部のネットワークによる比較であり，統一的な議論の展開は望めない．また，いくつかの研究では，ネットワーク構造から特徴量を抽出し，それらの特徴がどのような影響を与えているかを分析している．例えば，Yamir et al. [29] は次数相関 [30] に着目し，拡散現象と構造特徴の関係の分析を行っている．しかし，彼らの研究も一部のネットワークを用いているに過ぎず，ネットワーク構造が与える現象への影響は明らかになっていない．

現実には存在するネットワークを用いている研究も多く存在する．Shina et al. [31] は，投票結果におけるソーシャルネットワークの持つ影響力について調査するために，簡単なモデルを提案した．彼らは，ソーシャルネットワークのデータセットとして Facebook におけるネットワークを用いている．しかし，データとして収集できるネットワークは限られており，全てを網羅する事は非常に困難である．また，多くのデータを収集したとしても，それに含まれていない構造を持つネットワークが現実には存在しない事は証明できない．従って，同様にネットワーク構造全体から見れば一部の構造であり，現象の本質を捉える事は難しい．

このように現在までの研究では，ネットワーク構造が与える現象への影響は数多く分析されているものの，一部のネットワーク構造しか用いられていないため，信頼に足る結論には至っていない．

## 1.2 本研究の目的と位置づけ

本研究では，ネットワーク上で発生する現象とネットワーク構造の関係を統計的な分析によって明確にする事を目的とする．統計的に正確な分析を行うためには，一部のネットワーク構造のみのデータセットではなく，あらゆるネットワーク構造を網羅するようなデータセットを用いる必要がある．同一ノード数，リンク数から成るあらゆるネットワーク構造を含むネットワークデータセットを用意するためには，網羅的にネットワークを生成する必要がある．しかし，現在提案されている多くのネットワーク生成モデルは，生成したネットワークが一定の構造傾向を持つように提案されているため，ネットワーク構造を網羅的に生成する事はできない．

図 1.10 に，既存の複数のネットワーク生成モデルによってネットワークを生成し，ネットワークから抽出した特徴に基づき 3 次元特徴空間上にプロットした図を示す．この図は，各軸にネットワークの構造特徴を取り，各点であるネットワークがネットワーク生成モデル毎に色分けされている．この図を見ると，既存モデルはネットワーク構造特徴空間上の一部のネットワークしか生成できない事がわかる．本研究では，ネットワーク構造と

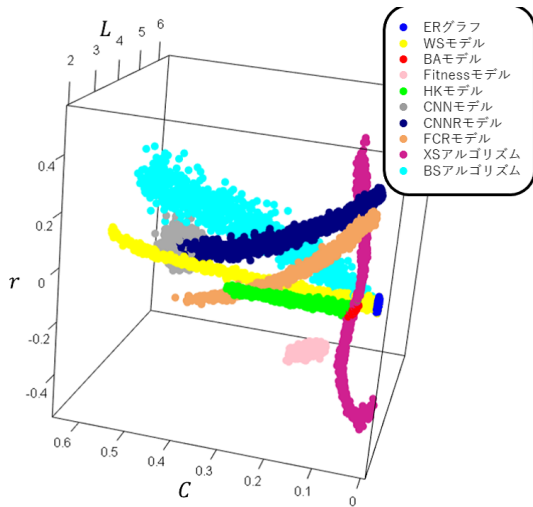


図 1.10: 既存ネットワーク生成モデルによるネットワーク群

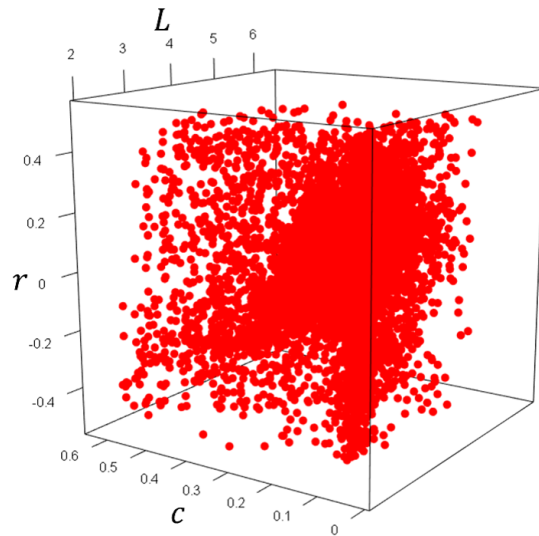


図 1.11: 提案手法によるネットワーク群

現象の関係性を分析するため、**図 1.11** のように、構造特徴空間を網羅するようなデータセットの構築手法の提案を行う。

このためには、様々な構造を持つネットワークを作り得るネットワーク生成手法が必要である。そこで、この構造特徴空間上の任意の特徴点を取り、その特徴点のネットワーク構造特徴から逆にネットワークを生成する手法を考案する。我々の手法は、ネットワーク構造特徴を自由に決め、貪欲法を用いてネットワークをその特徴を持つネットワークに成長させるアルゴリズムである。本研究では、このネットワーク生成アルゴリズムを用いて、様々なネットワークを含むネットワークデータセットを構築する。さらに、データセット上で現象を統計的に分析する事によって、ネットワーク構造と現象の関係を明らかにする。

本研究では任意の構造を持つネットワークを生成する手法を提案し、その手法を用いて網羅的に生成したネットワーク上で様々な現象の統計的な分析を行うことである。これにより、現実に存在するネットワーク構造を網羅したネットワーク構造全体での分析が可能となり、現象の本質を理解できる事を期待する。

### 1.3 本稿の構成

本稿の構成として、第 2 章において既存研究で行われている構造と現象の関係についての研究を紹介し、それらの研究の限界について述べる。前述のように、既存研究におい

て、数多くのネットワーク生成モデルが提案されている。しかし、この中でも、構造と現象の関係についての研究で用いられているネットワーク生成モデルはほんの一部である。ここでは、この研究で用いられているネットワーク生成モデルは十分であるのかを検討する。また、現在提案されている数多くのモデルを用いて、あらゆる構造を生成可能であるかについても同時に検討する。

第3章では、網羅的ネットワークデータセットの構築手法の提案について述べる。本手法では、まず、構造特徴空間を定義する。そして、この構造特徴空間を埋めるようにデータセットを構築する事を目標とする。この時、既存ネットワーク生成モデルを用いてこの目標を実現する事は出来ない。そこで、ここでは、構造特徴空間上の任意の点にネットワークを生成する方法を提案する。この任意の特徴点をランダムに設定し、そこにネットワークを生成することによって、様々な構造を持つネットワークを生成する事が出来る。これらのネットワーク群を一つのデータセットにまとめて、本研究では包括的データセットとする。ここでは、このデータセットが既存のモデルと比較して、どれだけ高い性能を持っているかを議論する。

以降では、このデータセットを用いて、現象とネットワーク構造の関係性について分析を行う。本研究では、現象の例として、情報拡散と、協調行動を取り上げる。

第4章では、現象の例としての情報拡散とネットワーク構造の関係性について分析する。ここでの情報拡散は、情報が個人間の相互作用によって、個人から個人へと伝播していくプロセスとする [32]。情報拡散は、背景に存在するネットワークの構造に強く依存する現象の一つである。情報拡散と構造の関係は様々に議論されているが一定の見解には至っていない。本章では、統計的に分析を行う事によって、これを明らかにする。さらに、既存研究における見解と比較していく。

第5章では、現象の例としての協調行動とネットワーク構造の関係性について分析する。一般に、利己的なエージェントで囚人のジレンマゲーム (Prisoner's Dilemma game : PD game) を行えば、協調戦略を行うエージェントは淘汰され、裏切りが支配的な戦略となることが知られている。協調行動を創発する一つの要因として、ネットワーク互恵 [20] がよく知られている。ネットワーク互恵とは、エージェント集団に空間的關係の偏り、つまりネットワークを仮定すると、エージェントの戦略が協調もしくは裏切りのみという単純な枠組みであっても、協調が維持されるというものである。このようなエージェント集団の間にネットワークを仮定した囚人のジレンマは空間型囚人のジレンマ (Spatial Prisoner's Dilemma; SPD) ゲームと呼ばれる。SPD ゲームにおいて、ネットワーク構造は重要な要素である。本研究では、この SPD ゲームとネットワーク構造の関係性について統計的な分析を行う。この時、情報拡散の場合と同様に既存研究の見解と比較しながら



分析を行う。

最後に第 6 章で，本紙のまとめと今後の課題についてまとめる。

## 第2章

# 既存研究とその限界

### 2.1 はじめに

これまでに、複雑ネットワークの構造と現象の関係は様々に分析されてきた。扱われる現象として、情報拡散 [18, 33, 34]、協調行動 [20, 21, 22] や、金融機関の連鎖倒産 [10, 11, 12] や、サービスネットワーク等の耐障害性、耐攻撃性に関する分析 [23, 24, 25] 等が行われている。これらの研究では、ネットワーク構造を対象とし、様々なネットワーク構造の上でシミュレーションを行う事によって、ネットワーク構造が現象に与える影響を明らかにしようと務めてきた。

ここでは、この中から本論文で取り扱う現象である情報拡散と協調行動についての関連研究を挙げる。これらの既存研究では、一部のネットワーク構造のみを対象として分析しているため、構造特徴と現象の関係に関する全貌は見えてこない。本章では、各既存研究が用いているネットワーク生成モデルを挙げ、それらのネットワークがネットワーク構造の一部しか表現できない事を示し、かつ、これらを用いた研究の限界について議論する。

本章では、まず本論文で議論する上で重要となる、複雑ネットワークの構造に関する研究の基礎について触れる。その後、ネットワーク構造と現象に関する先行研究を挙げ、それらの研究の限界について述べる。

### 2.2 複雑ネットワークの構造の基礎

#### 2.2.1 スモールワールド

まず、複雑ネットワークの構造に関する重要な発見に「スモールワールド性」がある。スモールワールド性が体系化されたのはダンカン・ワッツ (Duncan J. Watts) という物理

学者による [27]. 彼らはこの論文中で以下のように記述している.

We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs.

すなわち彼らは, 現実のネットワークは以下の二つの性質を持つとし, これをスモールワールド性 (Small worldness) とした.

- 任意の 2 ノードが短い経路長で結ばれる
- 高度にクラスター化されている

ここで 2 ノード間の経路長とは, あるノードからあるノードまで最短で辿り着くまでに通ったリンクの数と定義される. 次に, クラスター化について解説する. 自分の友人を 100 人集めてきたとして, その中の互いに友人であるペアを考える. おそらく相当数が互いに友人であるはずである. もし, 日本人 1 億 2694 万人 (平成 28 年現在) の中から無作為に友人を選んでいるならこうはならない. すなわち, 友人の友人は友人である確率が高いと言える. この性質をクラスター性と呼ぶ. クラスター性を持つネットワークは友人関係ネットワークに限らない. web 上のネットワークを初めとして様々なネットワークがこのクラスタ性を持っている.

ワッツらはこの 2 つの性質を持つネットワークを構築するための WS モデルを提案した. このモデルは, 図 2.1 左に示すような格子ネットワークから, 張り替え確率  $p$  に従ってリンクの張り替えを行う. このリンクの張り替えはランダムに行われ, このリンクが増える事で, 任意の 2 ノードの間の経路は短くなる. この張り替えられたリンクをショートカットと呼ぶ. ショートカットが比較的少ない時, ネットワークは図 2.1 中のような, 任意の 2 ノード間の距離は短く, 高度にクラスター化された, スモールワールドネットワークとなる.

## 2.2.2 スケールフリー

複雑ネットワークの構造に関するもう一つの大きな発見に 1999 年に発見された「スケールフリー (Scale-free) 性」がある. スケールフリー性とはネットワークのリンク数 (次数)  $k$  の分布 (次数分布)  $p(k)$  がべき則に従うというものである. この次数分布  $p(k)$  は次数  $k$  のノードの存在確率を表している. すなわち, ネットワークからランダムに一つノードを取り出したときに, そのノードの次数が  $k$  である確率が  $p(k)$  である. スケールフリー

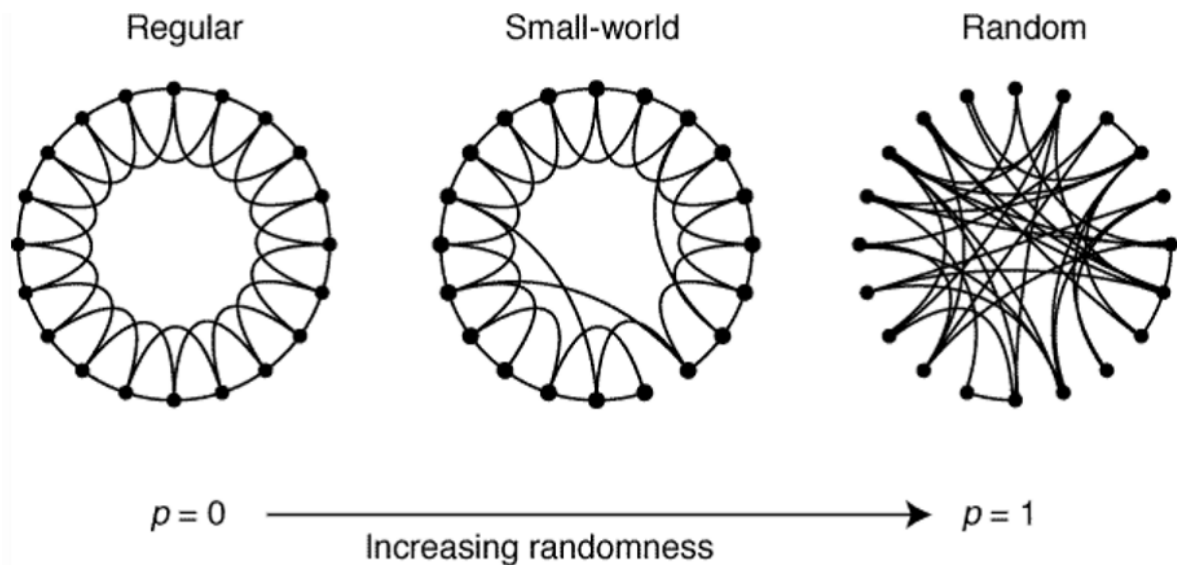


図 2.1: WS モデルによるネットワーク [27]

性は以下の式で表される.

$$p(k) \propto k^{-\gamma} \quad (2.1)$$

なお,  $\gamma$  はべき指数と呼ばれる. WS モデルは格子ネットワークからランダムで張り替えを行うため, その次数分布  $p(k)$  はポアソン分布に従う. しかし, 現実の多くのネットワークは, このスケールフリー性を持っているという. この発見をしたのは, バラバシ・アルバート=ラズロ (Barabási Albert László) らである.

彼らはネットワークがこのスケールフリー性を持つ要因として, ” Rich become richer.” と説明している. すなわち, 次数を多く持つノードはより多くの次数を勝ち取りやすい. 例えば, 友人の多い人は, さらに友人が出来る機会に恵まれるために, 友人の少ない人よりも友人が増えやすい. また, 多くのリンクを張られているウェブサイトは他の人から閲覧される機会が増えるため, さらにリンクを獲得しやすい.

さらに彼らは, この性質をモデル化し, Barabási-Albert(BA) モデルを提案した. BA モデルの特徴は以下の 2 点にある.

- ネットワーク成長モデル
- 優先的選択

このネットワーク成長モデルでは, 小さいサイズのネットワークをノードやリンクを追加していく事によって成長させていくモデルである. このモデルでは, ネットワークに入るタイミングに差があるため, 次数に差がしやすい. さらに, リンクを追加する際に,

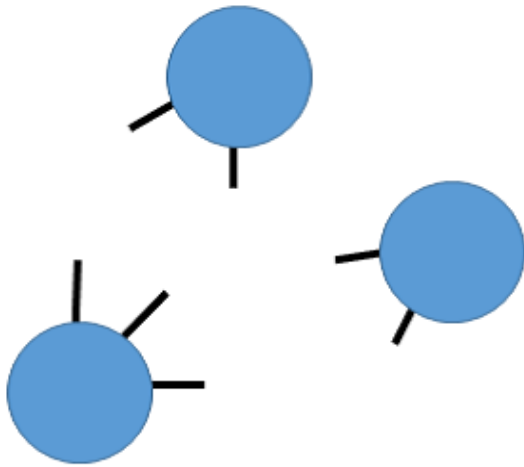


図 2.2: コンフィグレーションモデルの生成過程

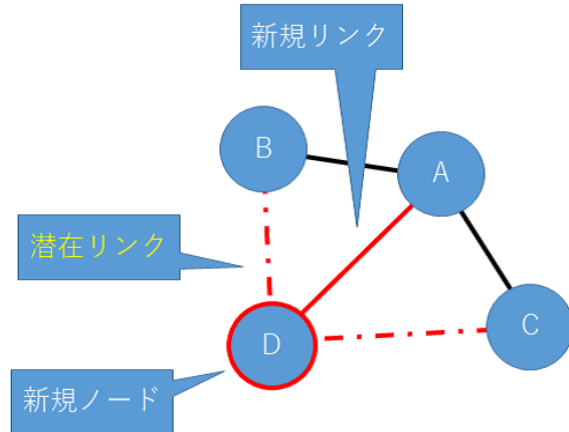


図 2.3: CNN モデルの生成過程

ノード  $i$  の選択される確率  $\pi_i$  を，ランダムではなくノード  $i$  の次数  $k_i$  に比例した確率で選択されるようにした．

$$\pi_i = \frac{k_i}{\sum_j k_j} \quad (2.2)$$

これによって，スケールフリーネットワークの生成に成功している．

### 2.2.3 ネットワーク生成モデル

ワッツによるスモールワールド性，バラバシによるスケールフリー性の発見後，15年あまりで複雑ネットワーク学は加速度的に進歩してきた．ワッツやバラバシが体系化したネットワーク生成モデルを基にして，現実のネットワークの持つ性質を表すために数多くのモデルが提案された．

一つは，一般化ランダムグラフと呼ばれるモデル [35, 36, 37] である．例えば，Newman が提案したコンフィグレーションモデル [38] は様々な分析に用いられている [39]．ランダムグラフでは，次数分布はポアソン分布に従うが，コンフィグレーションモデルでは，任意の次数分布を作る事が出来る．そういう意味で「一般化」されたランダムグラフである．コンフィグレーションモデルでは，初めに各ノードの次数を決定する．この次数に従ってランダムにリンクを配置していく．従って，最初に任意の次数分布を定義しておくことが可能である．すなわち，次数分布のべき指数  $\gamma$  を自由に変更する事が出来る．具体的には図 2.2 に示されるように，各ノードは手を持っており，その手をランダムに結ん

でいくことによってネットワークを生成していく。

もう一つは、BAモデルの流れを組む、ネットワーク成長モデルである。BAモデルは、その生成方法から、クラスター性を持つネットワークを生成する事ができない。それを解決するためにいくつかのモデルが提案されている。例えば、Vazquezら[40]はConecting Nearest Neighbor(CNN)モデルを提案した。このモデルは、友人の友人は友人になる確率が高いという考えに基づいたモデルである。ネットワークが成長する過程で、新規のノードを追加する際に、リンク先のノードの隣接ノードとの間に潜在リンクを生成する。すなわち、図2.3のように潜在リンクが生成される。この潜在リンクを確率的に実リンク化する事で、ネットワークを形成する。この手法は、クラスター性を高める事に直結する。もう一つの有名なモデルとして、Holme-Kim(HK)モデル[41]がある。HKモデルはトライアドフォーメーションという手法を採用している。トライアドフォーメーションとは、優先的選択によってリンクを生成した時に、確率的にリンク先のノードの隣接ノードにもリンクを結ぶという手法である。これもCNN同様にクラスター性を高めるのに直結しており、この確率が高い程クラスター性が高くなる。

CNNモデルやHKモデルのような、BAモデルの拡張は非常に多くのバリエーションがある[42, 43, 44, 45]。このように、現在に至るまで数多くのネットワーク生成モデルが提案されている。これは、現実世界に無数に存在するネットワーク構造に対応して、ネットワーク生成モデルを提案してきた結果である。しかし、ここまで様々なネットワーク生成モデルが存在していても、現実のネットワークの全ての構造を表せるわけではない。本研究では、構造特徴からネットワークを生成する手法を考案する事によって、あらゆる構造を持つネットワークを生成することを目指す。

## 2.2.4 ネットワークの構造特徴

ネットワークの構造を定量的に議論するために、ネットワークから構造特徴を抽出する。構造特徴はマクロスケール構造特徴、メソスケール構造特徴、ミクロスケール構造特徴に分けられる。

マクロスケール構造特徴とは、ネットワークの構造を俯瞰して見た時の、ネットワーク全体としての特徴量である。ここまでに紹介してきた経路長の平均やクラスター性[27]等はマクロスケール構造特徴に分類される。この他には次数相関[30]という構造特徴も有名なマクロスケール構造特徴である。次数相関は、リンクで結ばれているノード間の次数の差が大きいかどうかをピアソンの相関係数を用いて定義したものである。これらの特徴に関する詳細は第3章に示す。マクロスケール構造特徴は、ネットワークそのものがど

のような構造をしているかを定義づけたり、ネットワーク同士の構造の比較等に用いられる事が多い。

ミクروسケール構造特徴は、逆に一つ一つのノードに着目した構造特徴である。この特徴量には、ノードの中心性等の特徴量が分類される。ノードの中心性とは、ネットワーク構造上において、ノードの重要性について議論した特徴量である。この中心性には、用途に応じて様々なものが提案されている。例えば、近接中心性、媒介中心性 [46]、ページランク [47] といったものがこれに当たる。

近接中心性とは、自分から他人まで平均的にどのくらい近いかを表す特徴量である。すなわち、あるノード  $v_i$  の近接中心性はノード  $v_i$  から各ノードまでの最短経路長の逆数である。

$$c_i = \frac{1}{L_i} = \frac{N-1}{\sum_{j=1}^N dis(v_i, v_j)} \quad (2.3)$$

ここで、 $N$  はネットワークのノード数、 $dis(v_i, v_j)$  はノード  $v_i$  からノード  $v_j$  までの最短経路長である。

媒介中心性 [46] とは、ノード  $v_i$  が異なる集団を繋ぐ役割をする度合いである。ノード  $v_i$  の媒介中心性  $b_i$  は以下のように定義される。

$$b_i \equiv \frac{\sum_{s,d(s<d, s \neq i, d \neq i)} \frac{l(i;s,d)}{l(s,d)}}{(N-1)(N-2)/2} \quad (2.4)$$

ここで、 $l(s,d)$  はノード  $s, d$  間の最短経路長の数、 $l(i;s,d)$  はその内でノード  $v_i$  を含むものの数である。

ページランク [47] とは、重要ノードからの多くのリンクを集めているノードが重要であるとする特徴である。ノード  $v_i$  のページランクは以下の漸化式を解くことによって定義する。

$$\xi_i = \frac{1-d}{N} + d \sum_{v_j \in V(v_i)} \frac{\xi_j}{k_j} \quad (2.5)$$

ここで、 $V(v_i)$  は  $v_i$  と隣接するノード集合、 $d \leq 1$  はダンピングファクターである。ダンピングファクター  $d$  は収束時間やページランクの値に影響を与える。これらのミクروسケール構造特徴はネットワークの中で重要な部分(サブネットワーク)を発見する時に用いられる事が多く、様々なネットワークにおいて用いられている。

メソスケール構造特徴は、マクروسケール構造特徴とミクروسケール構造特徴の間に位置する構造特徴である。メソスケール構造特徴にはコミュニティ構造が分類される。コミュニティ構造は、モジュール構造、クラスタ構造等の呼ばれ方をされるが、ネット

ワークを複数の塊に分割したものである。ネットワークをコミュニティに分割する方法は数多く提案されている [48, 49, 50]。コミュニティ分割の重要なものの一つにニューマン (Mark Newman) の提案したニューマン法がある [51]。ニューマンがモジュラリティ (modularity) という概念を導入した事によって、コミュニティ研究にブレイクスルーを迎えた。モジュラリティ  $Q$  は以下の式によって定義される。

$$Q \equiv \frac{1}{2M} \left[ \sum_l \sum_{i,j=1; v_i, v_j \in g_l} (A_{ij} - \frac{k_i k_j}{2M}) \right] \quad (2.6)$$

ここで、 $M$  はネットワークのリンク数、 $g_l$  は 1 つのコミュニティを表す。ニューマン法はこのモジュラリティを最大化するようにコミュニティ分割を行う。このモジュラリティを最大化する事によって、コミュニティ分割を行う方法がいくつか提案されている [52, 53]。これらの手法を用いて分割されたコミュニティ構造はメソスケール特徴量であり、これらの特徴はネットワークから意味のある集合を抽出する際に用いられることが多い。

このように、既存研究では様々なネットワーク構造特徴が提案され、ネットワークの構造を規定してきた。本研究では、ネットワーク構造全体をとらえるために、マクロスケール構造特徴を用いる。複数のマクロスケール特徴量を用いてベクトルとする事でネットワーク構造間の距離を定義する。すなわち、このマクロスケール特徴量が同じネットワークは類似した特徴を持つとする。

## 2.3 従来手法とその限界

### 2.3.1 情報拡散に関する比較研究

ここでは、情報拡散と構造の関係についての関連研究に関して詳細に述べていく。情報拡散は、ネットワーク構造の影響を強く受ける事が知られており、構造との関係は古くから分析されている。情報拡散の分析は、感染症モデルを応用して行われる事が多い。情報が伝わっていく様子を、感染症が伝染していく様子として捉える事が可能であるためである。そのため、情報拡散をより広くとらえ、感染症モデルへのネットワーク構造の影響を分析している論文について紹介する。

#### 単純モデルを用いた比較研究

感染症モデルへのネットワーク構造の影響は非常に多くの研究者によって分析されている。Zhou ら [18] は、SIR モデル [54] を用いて、ネットワーク構造が噂の伝搬にもたらす影響について分析した。SIR モデルとは、感受性保持者 (Susceptible:S)、感染者



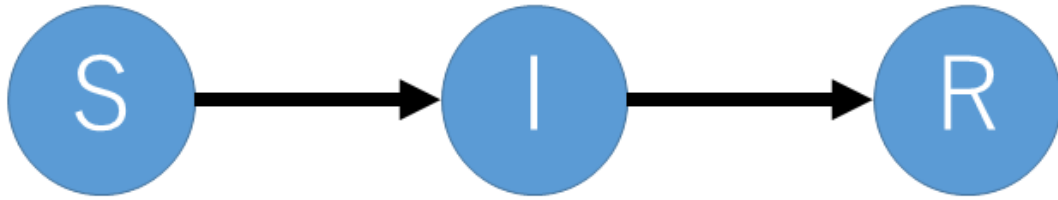


図 2.4: SIR モデルにおける状態遷移図

(Infected:I), 免疫保持者 (Recovered:R) のそれぞれの頭文字を取ったモデルである。図 2.4 に SIR モデルにおける 3 つの状態の状態遷移図を示す。このモデルの特徴は、状態が不可逆である事である。すなわち、一度感染し、回復したノードは免疫を持ち、二度と感染はしない。このモデルはデマ情報の拡散モデルとしても利用されている。デマ情報に感染した後、それがデマだと認識したノードは二度とそのデマを信じる事はないためである。Ikeda らは、SIR モデルを用いて、2011 年 3 月の東日本大震災時に流れたデマについての分析を行っている [5]。

この SIR モデルとネットワーク構造との関係について、過去に非常に多くの研究がなされてきた。これらの研究のほとんどは複数のネットワーク生成モデルを用いてネットワークを生成し、その上での挙動について分析するというものである。Zanette ら [19] は、WS モデルを用いて SIR モデルとネットワーク構造との関係について分析を行っている。彼らは WS モデルにおけるショートカットが出来る確率  $\epsilon$  と、情報の拡散率には深い関係性がある事を示している。Zhang ら [34] は SIR モデルを拡張して噂の伝搬モデルを提案した。提案モデルは、次数の多いノードは隣接ノードとの接触も多いはずであるという仮定を基に拡張したものである。このモデルを WS モデル、BA モデル、ER グラフの上で挙動確認を行い、拡散率がネットワーク構造の影響を強く受ける事を示している。

SIR モデル以外にも様々な感染モデルとネットワーク構造との関連について調べられている。Zhang ら [33] は、SIS モデルを用いて、感染症の流行における予防接種の影響の分析を行った。SIS モデルとは、SIR モデルを可逆にした感染モデルである。すなわち、感染状態から回復しても、もう一度感染する可能性があるモデルである。彼らは、この予防接種の効果について、BA モデル、ER グラフの上で比較を行い、BA モデルの方が感染症の流行を抑え込むことが容易である事を示した。

また、Nagata ら [55] は、Ising model を用いて、拡散現象に与えるネットワーク構造の

影響について分析をしている。彼らは、CNN モデル、WS モデル、Klemm Eguiluz(KE) モデル [56] 上で比較を行った。彼らは、それぞれのモデルの生成ルールに焦点を当て、それぞれのネットワーク生成モデルのパラメータと情報拡散の関係について明確化した。

このように、多くの研究では、単純なネットワーク生成モデルが利用されている。そのため、これらの研究での知見は、使用されたモデル上でのみの知見である可能性が高い。例えば、BA モデルと ER グラフでの比較は、そのネットワーク自体の特性の比較であり、スケールフリー性の有無による比較にはなっていない可能性がある。また、WS モデルのパラメータを変化させた時の情報拡散の影響は、WS モデル上でしか言えず、実際のネットワーク上でどのような意味を持つのかは定かではない。そこで次に、ネットワークから抽出される構造特徴に着目した研究について紹介する。

### ネットワーク構造特徴を用いた比較研究

Yamir ら [29] は次数相関に着目し、拡散現象と構造特徴の関係の分析を行った。彼らは次数相関の高いネットワークと次数相関のないネットワークを用意し、それぞれのネットワーク上で SIR モデルによる拡散シミュレーションを行った。このように次数相関の異なるネットワーク上でのシミュレーションによって、次数相関の高いネットワークよりも、次数相関のないネットワークの方が、大きく拡散される事を示している。

Yinzuo ら [57] は、SIR モデルを用いて、クラスタ係数が拡散現象に与える影響について分析している。彼らは、HK モデルを用いて、クラスタ係数の違うネットワークを生成してシミュレーションを行った。HK モデルは、パラメータ  $p$  ( $0 \leq p \leq 1.0$ ) によりクラスタ係数を自由に変化させる事が出来る。彼らは、このようなクラスタ係数の異なるネットワーク上でのシミュレーションによって、クラスタ係数の増加が拡散率の低下につながる事を示している。

これらの研究では、次数相関及びクラスタ係数の与える影響についてみているが、複合した要因については検討していない。Badham ら [58] は、クラスタ係数と次数相関の複合的な影響について、SIR モデルを用いて検証した。彼らは、2次元特徴空間において、Grid 状にネットワークを生成し、その上でシミュレーションを行い分析を行った。その結果、クラスタ係数も次数相関も、共に低い方が拡散率が高い事と主張している。

これらの分析の主張は一貫した主張となっているが、着目した以外の特徴量に関しては全く触れていない。例えば、平均経路長が情報拡散に影響を与える事は自明であるが、この値がどのようになっているかについて触れていない。

### 2.3.2 協調行動に関する比較研究

協調行動は古くからネットワーク構造との関係を分析されている現象の一つである。協調行動の分析は多くの場合、囚人のジレンマゲーム (the Prisoners's Dilemma game:PD game) が用いられる。囚人のジレンマゲームは現実社会の社会ジレンマ下における相互作用を比喩するものであり、これの応用によって、様々な現象に対しての知見が得られている [59, 60, 61]。囚人のジレンマゲームの詳細は第 5 章に示す。一般に、利己的なエージェントの the well-mixed group(完全ネットワーク) で行われる囚人のジレンマゲームは、裏切りが支配的な戦略となることが知られている。したがって、社会においては何等かの要因が働いて協調が創発されているはずである。この要因としては様々な説が挙げられ、研究が進められている。例えば、囚人のジレンマゲームを行う各エージェントが学習をすることによって協調が創発されるという研究が行われている [62, 63, 64]。また、協調行動はネットワーク構造によっても創発される。Nowak らは、エージェント集団に空間的關係の偏り、つまりネットワークを仮定すると、エージェントの戦略が協調もしくは裏切りのみという単純な枠組みであっても、協調が維持されると発表した [20]。これはネットワーク互惠と呼ばれる。

#### 空間的囚人のジレンマの研究

Nowak らはこのネットワーク互惠を、空間的囚人のジレンマゲーム (Spatial Prisoner's Dilemma game; SPD game) として体系化した。Nowak の発表以来、ネットワーク構造は協調的な社会が実現されるための重要な要素として考えられている。

協調行動が進化する要因の一つとして、ネットワークが挙げられてから [20]、近年に至るまで、空間的囚人のジレンマに関する研究は数多く行われている。Juan ら [65] は、Coalitions(エージェントのグループ) に着目した SPD モデルを提案し、正方格子上で分析を行った。本モデルでは、エージェントが Coalitions に参加する事ができ、Coalition 内の戦略をリーダーエージェントが決定するというものである。このモデルで、古典的な確率的しつぺ返し戦略 (probability Tit-for-Tat) が独立学習 (isolated learning) より効果的協調戦略を増加させると主張している。Bijan ら [66] は、各エージェントがそれぞれ個別の利得行列を持った continuous dynamical model を提案した。彼らは、個別の利得行列を使った SPD ゲームにおける新たな分析方法を提案し、BA モデル及び格子ネットワークを用いて分析を行った

## 空間的囚人のジレンマとネットワーク構造に関する研究

空間的囚人のジレンマに関する研究の中でも、ネットワーク構造と協調行動の関係性についての研究には以下のようなものがある。Tomochi ら [67] は格子ネットワークにランダム性を加える事で、ランダム性が高い事が裏切り戦略エージェントの原因であると主張し、クラスタ係数が協調達成に大きく影響する事を示している。Abramson ら [68] は、ネットワークの構造と SPD ゲームの関係を理解するために、WS モデル [27] 上でシミュレーションを行った。WS モデルのパラメータ  $\epsilon$  ( $\epsilon = 0$  で格子ネットワーク,  $\epsilon = 1$  でランダムネットワーク) を段階的に変化させて生成したネットワーク上で SPD ゲームを行い、 $\epsilon$  の変化によって裏切り戦略の割合にどのような変化が生じるかを分析した。その結果、 $\epsilon = 0.1$  にピークが存在する事を突き止めた。Gómez-Gardeñes ら [69] は、ER グラフ [26] と BA モデル [28] の 2 種のネットワークにおいて、協調戦略の増え方が異なる事を指摘した。ER グラフでは協調戦略のクラスターが複数できるが、BA モデルでは、1 つのコアを中心として、協調戦略の塊ができる事を示した。

これらの研究ではネットワーク構造の差異によって、協調戦略の割合に差が生じる事を示しているが、これらの研究で扱っている WS モデルや ER グラフ、BA モデルは、ネットワーク構造全体から見れば極一部の構造であり、総括的な議論は望めない。

## ネットワーク構造特徴を用いた比較研究

また、ネットワークの構造特徴に着目した研究が行われている。Rong ら [21] は次数相関に着目した。Xulvi-Sokolov(XS) アルゴリズムを用いて、次数相関  $r = 0.0 - 0.3$  のスケールフリーネットワークを生成し、その上で SPD ゲームを行う事で、次数相関が協調戦略の進化に与える影響を観察した。その結果として、協調戦略は次数相関のないネットワークの方が残存しやすい事を示した。また、彼らは、Dorogovtsev Mendes-Samukhin(DMS) モデル [70] を用いて、スケールフリーネットワークの次数分布  $P(k) \propto k^{-\gamma}$  を  $\gamma$  を変化させながら生成し、次数分布が協調戦略の進化に与える影響について分析した [71]。彼らは、 $\gamma$  が大きくなるほど、裏切った時の利得  $b$  を増やしたときの協調戦略の低下が激しい事を示した。Assenza ら [22] は、クラスタ係数に着目した。Holme-kim(HK) モデル [41] により、様々なクラスタ係数を持つスケールフリーネットワークを生成し、分析を行った。彼らは、裏切った時の利得  $b$  がクラスタ係数の高いネットワークで協調戦略は生き残りやすいが、 $b$  が高くなると急激に減少し、 $b$  が大きくなると逆に協調戦略が生き残りにくくなる事を示した。これらの研究では、個々の構造特徴からの影響を分析しているが、他の構造特徴に関しては考慮されていない。したがって、偽相関である可能性や、構造特

徴の複合的な作用に関しては無視されている。

## 2.4 既存の分析の限界について

ここまで見てきたように、ネットワーク構造と現象の関係を分析している既存研究の多くは、特定のネットワーク生成モデルを用いた分析を行っている。このネットワーク生成モデルが生成するネットワークは、その生成方法に依存した構造の傾向を持っている。ここでは、それぞれのモデルを用いてネットワークを生成しその特徴を確認する。ER グラフ、BA モデルのように常に一定の構造を持つネットワークを生成するモデルや、WS モデルのような制御パラメータを変える事によって様々な構造のネットワークを生成できるモデル、HK モデルや XS モデルのように特定の構造特徴を変化させる事が出来るモデルの表現領域についてそれぞれ述べる事で、これらのネットワークが極一部の構造特徴のみである根拠を挙げる。

### 2.4.1 本論文で用いる既存ネットワーク生成モデル

#### Erdős-Rényi(ER) グラフ

Erdős-Rényi(ER) グラフ [26] はエルデシュらによって提案されたランダムネットワークである。ノードを  $N$  個用意し、 $N \times N$  の確率行列を用意する。各要素には  $[0,1]$  の乱数を入れる。この確率行列の内  $p$  を超えるものを全てリンクとする。 $p$  は ER グラフのパラメータである。このパラメータの値によってリンク数  $M$  は制御される。リンク数  $M$  のネットワークを求める時、 $p$  の値は以下ようになる。

$$p = \frac{M}{N C_2} \quad (2.7)$$

図 2.5 に ER グラフを用いて生成したネットワークを示す。

#### Watts-Strogatz(WS) モデル

Watts-Strogatz(WS) モデルは Watts らによって提案された。WS モデルはレギュラーネットワークから確率  $p$  に従ってリンクの張り替えを行う事によって、スモールワールドネットワークを生成するモデルである。なお、 $p$  はモデルのつなぎ変え確率である。WS モデルのアルゴリズムをアルゴリズム 1 に示す。また、図 2.6 に WS モデルを用いて生成したネットワークを示す。WS モデルで生成されたネットワークは、図 2.6 のような高度にクラスタ化された構造を持っている。

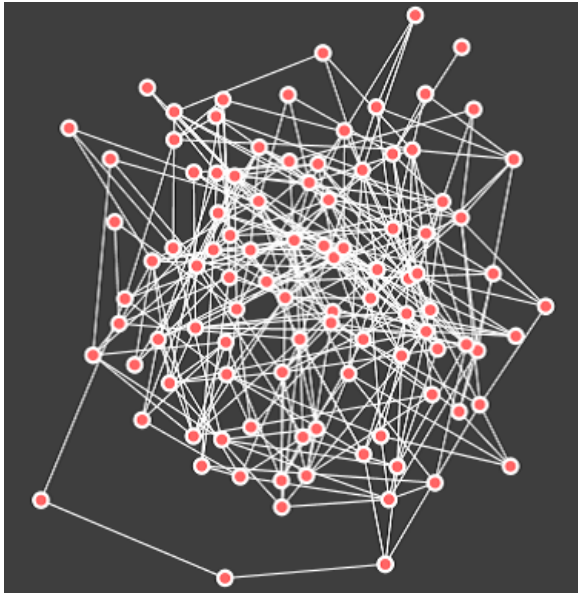


図 2.5: ER グラフのネットワーク

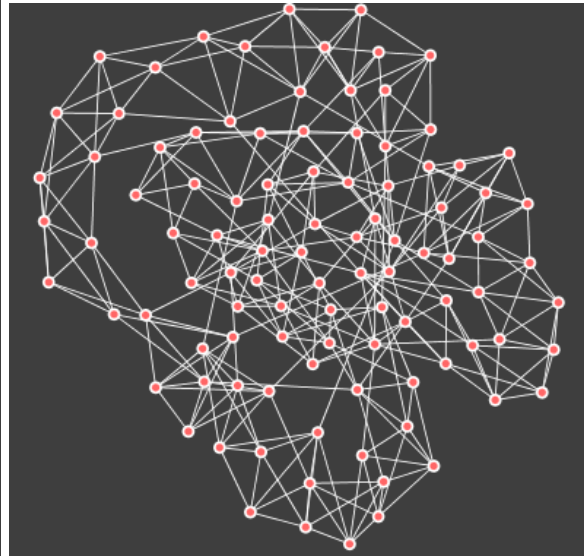


図 2.6: WS モデルのネットワーク

---

#### Algorithm 1 WS モデル

---

平均次数  $m$

$N$  ノードを輪状に置き, 各ノードを左右  $\bar{k}/2$  個まで隣接させたレギュラーネットワークを用意する

**for** 各リンク **do**

**if**  $p$  **then**

    片方の端点とはつないだままにして, もう片方の端点から切り離し, ランダムに選んだ新たなノードに接続する.

**end if**

**end for**

---

#### Barabási-Albert(BA) モデル

Barabási-Albert(BA) モデルは Barabási らによってスケールフリーネットワークを表現するために提案されたネットワーク生成モデルである. ネットワークを生成する際に以下の2つを用いる.

- ネットワーク成長モデル  
ネットワークが小さい状態から開始し, 少しずつ大きくしていく手法.

---

**Algorithm 2** BA モデル

---

平均次数  $m$

**Ensure:**  $m$  ノードの完全ネットワークを用意する.

**while** ネットワークが求めるノード数となるまで **do**

**for**  $m$  **do**

    ノードを追加する

    優先的選択で選んだノード  $v$  と新規ノード間にリンクを生成

**end for**

**end while**

---

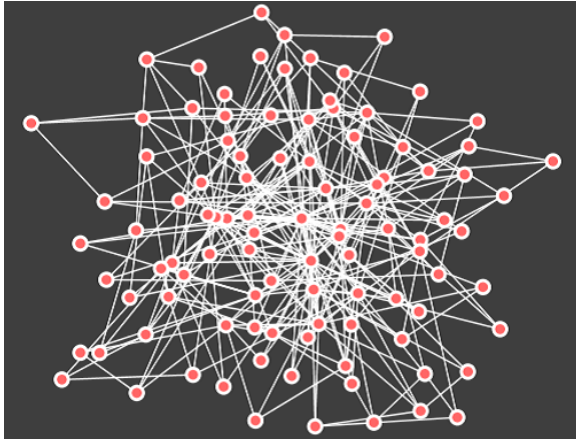


図 2.7: BA モデルのネットワーク

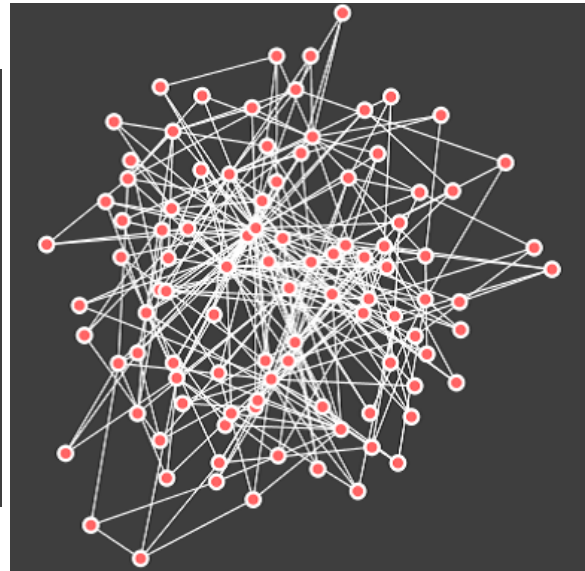


図 2.8: Fitness モデルのネットワーク

- 優先的選択

ノード  $v_i$  が選択される確率  $\pi_i$  を以下とする.

$$\pi_i = \frac{k_i}{\sum_j k_j} \quad (2.8)$$

アルゴリズム 2 に BA モデルのアルゴリズムを示す. また, 図 2.7 に BA モデルで生成したネットワークの例を示す. BA モデルで生成されたネットワークには, 図 2.7 のように次数の集中するハブノードが存在する.

## Fitness モデル

Fitness モデル [42, 43] は, BA モデルの優先的選択を拡張し, 後から追加されたノードでもリンクを獲得出来るようにしたモデルである. Fitness モデルの優先的選択において, 次数  $k_i$  のノード  $i$  が選択される確率  $\pi_i$  は以下のように算出される.

$$\pi_i = \frac{k_i \eta_i}{\sum_j k_j \eta_j} \quad (2.9)$$

ここで,  $\eta_i$  は「適応度 (fitness)」である. 適応度  $\eta_i$  は, 他者と次数を競う能力を表し, 実社会においては, 他者への影響度や魅力等と考える事が出来る. 図 2.8 に Fitness モデルで生成したネットワークの例を示す. Fitness モデルで生成したネットワークでは, 図 2.8 に示すように, BA モデルと非常に近い構造を持っている. しかし, Fitness モデルで生成したネットワークでは BA モデルで生成したネットワークと比較して, 複数のハブノードの存在が確認できる.

## Holme-kim(HK) モデル

Holme-kim(HK) モデルは, BA モデルのクラスタ係数の小ささを補うために拡張したネットワーク生成モデルである [41]. HK モデルはトライアドフォーメーションを用いてクラスタ係数の高いスケールフリーネットワークを生成する. トライアドフォーメーションでは, 新規ノードが優先的選択によってリンクを生成した時に, リンク先のノードの隣接ノードにもモデルパラメータ  $p$  の確率でリンクを結ぶという手法である. これは, ネットワークに参加する時に, 何人かの塊を形成してネットワークに参加する事に対応している. 従って,  $p$  の値を変更する事によって, クラスタ係数を調整する事が出来る. HK モデルの詳細をアルゴリズム 3 に示す. また, 図 2.9 に HK モデルで生成したネットワークの例を示す. HK モデルのネットワークには, BA モデルのネットワークでは見られない, ノード同士が三角形に結びついている関係が多くみられる.

## Conecting Nearest Neighbor(CNN) モデル

Conecting Nearest Neighbor(CNN) モデルは, 「友人の友人は, 友人になる」という考えに基づいたモデルである [40]. Davidsen, Ebel, Bornholdt [72] らにより DEB モデルと呼ばれる原案が提案され, Vazquez によって「潜在リンク」を実リンク化するというプロセスの整理を経てネットワーク成長モデルへと発展した. ここで「潜在リンク」は, 「友人の友人」という関係を表す. このモデルは, ノードとリンクを繰り返し行う事によって, ネットワークを少しずつ成長させていくモデルである. ノードを追加する際には, 新規



---

**Algorithm 3** Holme-Kim model

---

**Ensure:** ノード数  $m$  の完全ネットワークを用意する

平均次数  $m$

**while** ノード数が規定のノード数になるまで **do**

ノードを追加する

優先的選択で選んだノード  $v$  と新規ノード間にリンクを生成

**for**  $m-1$  **do**

**if**  $p$  **then**

ノード  $v$  の隣接ノード  $u$  と新規ノードの間にリンクを生成

**else**

優先的選択により選んだノード  $w$  と新規ノード間にリンクを生成

**end if**

**end for**

**end while**

---

ノード  $v$  とネットワークからランダムに選択したノード  $u$  との間にリンクを結ぶ。この際、ノード  $u$  の全ての隣接ノードと、新規ノード  $v$  との間に潜在リンクを結ぶ。リンクを追加する際は、この潜在リンクを実リンク化する事によって、「友人の友人は、友人になる」確率を上げたモデルである。なお、ノードを追加する確率はパラメータ  $p$  によって制御され、その値は生成するネットワークのノード数及びリンク数によって算出される。CNN モデルの詳細をアルゴリズム 4 に示す。また、図 2.10 に CNN モデルで生成したネットワークの例を示す。CNN モデルのネットワークでは、図 2.10 のように、ノード同士が三角形に結びついている事がわかる。

### CNNR モデル

CNNR モデルは、Yuta らによって CNN モデル、ランダム接続の手法を組み合わせで提案されたネットワーク生成モデルである [45]。CNNR モデルは、モデルパラメータ  $p$  を調整する事によって、CNN による接続の割合を変更し、様々なネットワークを生成する。Yuta らは、この CNNR モデルによって、mixi<sup>\*1</sup> に形成されるユーザネットワークに性質の近いネットワークを生成する事に成功している。CNNR モデルの詳細をアルゴリズム 5 に示す。また、図 2.11 に CNNR モデルで生成したネットワークの例を示す。CNNR モ

---

\*1 <https://mixi.jp/>

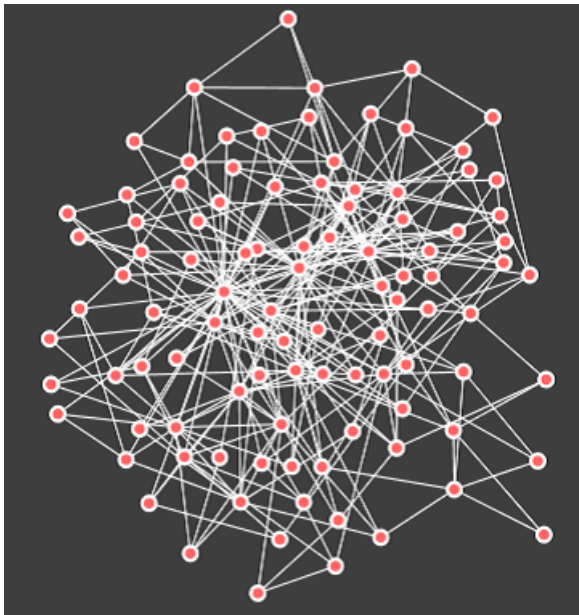


図 2.9: HK モデルのネットワーク

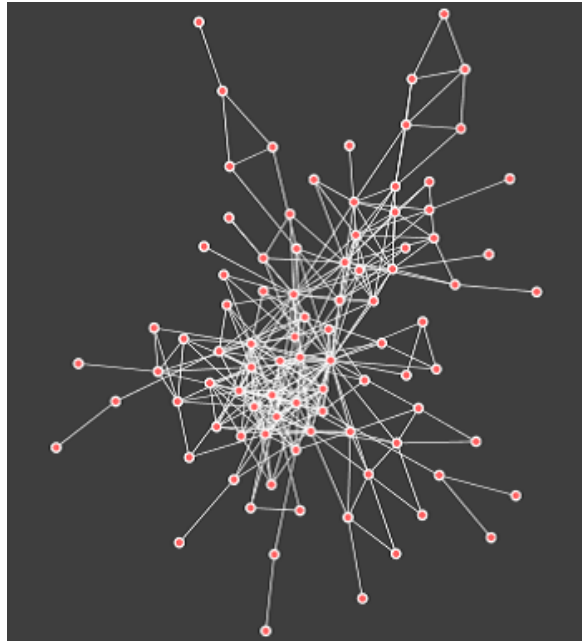


図 2.10: CNN モデルのネットワーク

---

**Algorithm 4** CNN model

---

**Ensure:** 1 ノードから成るネットワークを用意する.

**for** M **do**

**if**  $p$  **then**

    新規ノード  $v$  を追加する.

    ネットワークからランダムに選択したノード  $u$  と、新規ノードとの間にリンクを生成する.

    ノード  $v$  とノード  $u$  の隣接ノード群  $B(u)$  の全てに対して潜在リンクを生成する.

**else**

    潜在リンクを 1 つ選択し、それを実リンクにする.

**end if**

**end for**

---

デルのネットワークは、CNN モデルのネットワークと非常に類似しているが、ランダム性が大きく、三角形の関係が少ない事がみられる.

---

**Algorithm 5** CNNR model

---

**Ensure:** 1 ノードから成るネットワークを用意する.

**for**  $M$  **do**

**if**  $p$  **then**

    新規ノード  $v$  を追加する.

    ネットワークからランダムに選択したノード  $u$  と, 新規ノードとの間にリンクを生成する.

    ノード  $v$  とノード  $u$  の隣接ノード群  $B(u)$  の全てに対して潜在リンクを生成する.

**else**

**if**  $\rho$  **then**

      ネットワークからランダムに選んだ 2 ノードの間にリンクを生成する.

**else**

      潜在リンクを 1 つ選択し, それを実リンクにする.

**end if**

**end if**

**end for**

---

## FCR モデル

FCR モデルは, 鳥海らによって Fitness モデル [42, 43], CNN モデル [40, 72], ランダム接続の手法を組み合わせる提案されたネットワーク成長モデルである [44]. このモデルは既存の手法を確率的に選択する事によって, 様々な構造のネットワークを生成可能なモデルである. 鳥海らは, 様々な SNS 上に構築されるユーザネットワークに近い性質を持つネットワークをこのモデル一つによって, 生成する事に成功した. FCR モデルの詳細をアルゴリズム 6 に示す. また, 図 2.12 に FCR モデルで生成したネットワークの例を示す. FCR モデルのネットワークでは, CNNR モデルのネットワークと比較して, ハブノードが多く存在する事がわかる.

## Xulvi-Sokolov アルゴリズム

Xulvi-Sokolov(XS) アルゴリズム [73] は任意の次数相関を持つネットワークを生成するために用いられる手法である. XS アルゴリズムでは, あらかじめ設定されたネットワークから, 次数相関が高く (または低く) なるように張り替えを行う手法である. アルゴリ

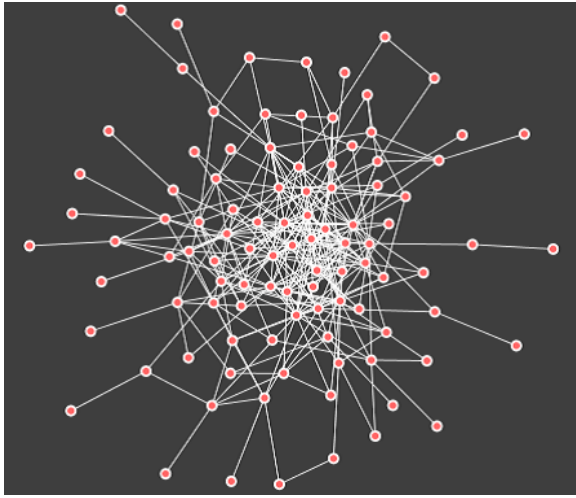


図 2.11: CNNR モデルのネットワーク

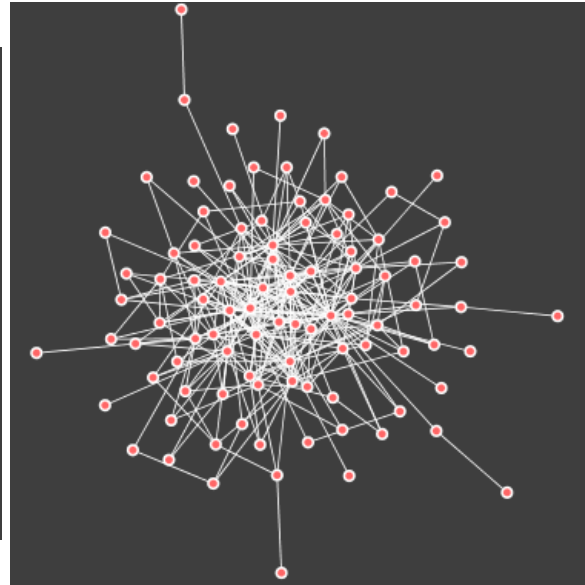


図 2.12: FCR モデルのネットワーク

---

**Algorithm 6** FCR model

---

**Ensure:** 1 ノードから成るネットワークを用意する.

**for**  $M$  **do**

**if**  $p$  **then**

    新規ノード  $v$  を追加する.

    ネットワークから適応度優先的選択により選択したノード  $u$  と, 新規ノードとの間にリンクを生成する.

    ノード  $v$  とノード  $u$  の隣接ノード群  $B(u)$  の全てに対して潜在リンクを生成する.

**else**

**if**  $\rho$  **then**

      ネットワークからランダムに選んだ 2 ノードの間にリンクを生成する.

**else**

      潜在リンクを 1 つ選択し, それを実リンクにする.

**end if**

**end if**

**end for**

---

---

**Algorithm 7** Xulvi-Sokolov アルゴリズム

---

**Ensure:** 初期ネットワークを生成する.

目標次数相関  $r_o$ , ネットワークの次数相関  $r_g$

**while**  $|r_o - r_a| > 0.001$  **do**

    リンクを 2 つ, すなわち 4 つのノードを選択する.

    選択したリンクを切断する.

**if**  $p$  **then**

**if**  $r_o > r_a$  **then**

            次数の大きい 2 つのノードと小さい 2 つのノードの間にリンクを追加

**else**

            次数の最も大きいノードと最も小さいノード, それ以外の 2 つのノードの間に  
            リンクを追加

**end if**

**else**

        ランダムに 2 つずつノードを選び, リンクを追加

**end if**

**end while**

---

ズムの詳細をアルゴリズム 7 に示す. また, 図 2.13 に XS アルゴリズムで生成したネットワークの例を示す. 図 2.13 の左は正の次数相関を持つように生成し, 右は負の次数相関を持つように生成した. 正の次数相関を持つように生成したネットワークでは, 次数の近いノード同士が結びつき, 負の次数相関を持つように生成したネットワークでは, 次数の高いノードと低いノードが結びついている事がみられる.

### Badhama-Stocker アルゴリズム

Badhama-Stocker(BS) アルゴリズム [73] は任意の次数相関とクラスタ係数を持つネットワークを生成するために用いられる手法である. BS アルゴリズムは, あらかじめ決められた次数を持つノードが円形に配置され, 近くのノード程リンクが結ばれやすいアルゴリズムである. この時, 近くのノードにリンクが結ばれる程, 格子ネットワークのような構造となり, クラスタ係数が高くなる. また, 円形に並べるノードの順番を並べ替える事によって, 次数相関を変更する事が可能である. アルゴリズムの詳細をアルゴリズム 8 に示す. また, 図 2.14 に BS アルゴリズムで生成したネットワークの例を示す. 図 2.14 の

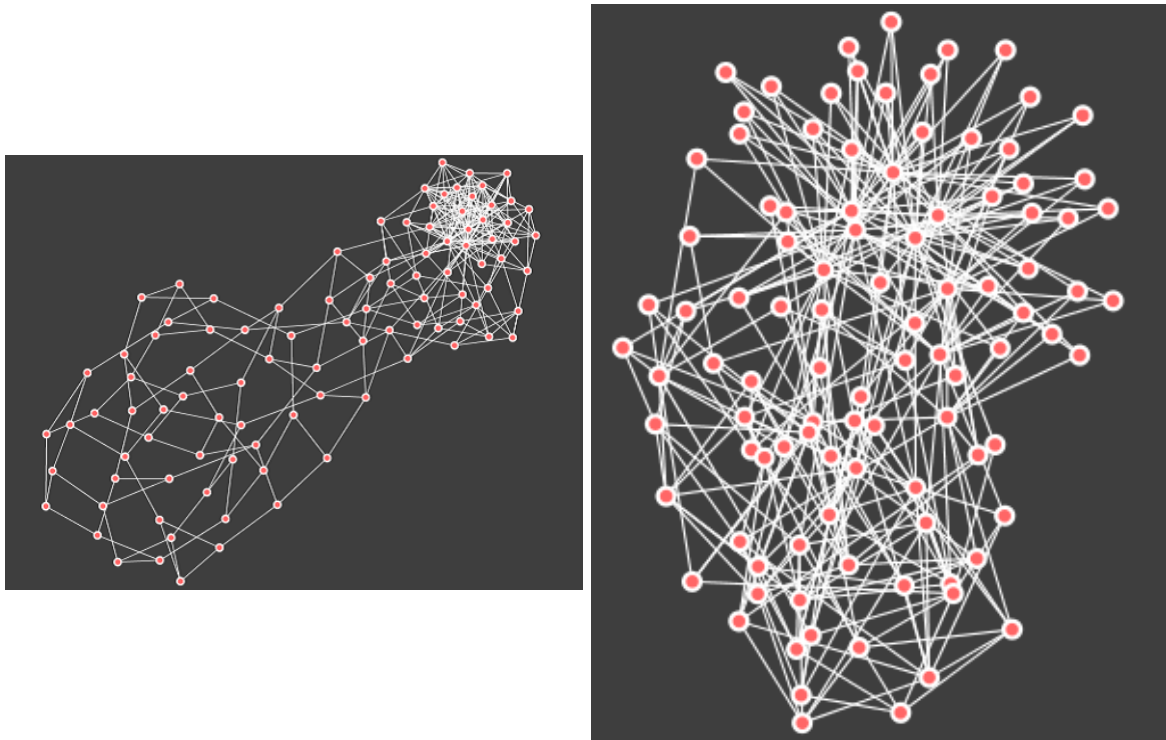


図 2.13: XS アルゴリズムのネットワーク (左: 正の次数相関, 右: 負の次数相関)

左は高いクラスタ係数を持つように生成し, 右は低いクラスタ係数を持つように生成した. ネットワークは円環構造を持っている事がわかる.

## 2.4.2 既存研究で用いられているネットワーク生成モデルの性質

### ネットワーク生成設定

前節で解説した 10 個のネットワーク生成モデルを用いて, ネットワークを生成し, その性質を見る. それぞれのネットワーク生成モデルで 1000 個ずつネットワークを生成し, 生成されたネットワークから各特徴を抽出する. 各ネットワークのノード数は 1000 とし, 平均次数  $\bar{k}$  は  $19 \leq \bar{k} \leq 20$  となるように生成した. また, 各モデルパラメータはランダムに変化させネットワークを生成する. また, XS アルゴリズムにおけるあらかじめ設定したネットワークとして, BA model で生成したネットワークを用いる.

### 生成ネットワークの性質

以下の 3 つのネットワーク特徴量に関して, これらのモデルが取り得る値の範囲を示す.

---

**Algorithm 8** Badhama-Stocker アルゴリズム

---

**Ensure:** あらかじめ決められた次数を持つノードを円形に配置する.

**Ensure:** クラスタ係数  $C$  及び次数相関  $r$  を指定する.

**while** 設定された次数相関を超えるまで **do**

**for** ノードの交換回数 **do**

    ランダムに一对のノードを選択する.

    選択されたノードの空間的隣接ノードの平均次数を算出する.

    次数の高い方のノードの隣接ノードの平均次数が、もう一方のノードの隣接ノードの平均次数より低ければ、2つのノードの位置を交換する.

**end for**

**for** 各ノード  $i$ (円形に並べたノードの  $i$  番目) **do**

**while** ノードの次数があらかじめ決めた次数になるまで **do**

**for**  $G=1$  **do**

**if**  $2 \cdot C$  **then**

        ノード  $i$  とノード  $i+G$  の間にリンクを結ぶ

**end if**

**if**  $2 \cdot C$  **then**

        ノード  $i$  とノード  $i-G$  の間にリンクを結ぶ

**end if**

$G = G + 1$

**end for**

**end while**

**end for**

**end while**

---

- 平均経路長  $L$  ( $1 \leq L$ )
- クラスタ係数  $C$  ( $0 \leq C$ )
- 次数相関  $r$  ( $-1 \leq r \leq 1$ )

上記の 10 個のネットワーク生成モデルで生成したネットワークの  $L$ ,  $C$ , 及び  $r$  の関係を図 2.15 に示す. この図では, それぞれのモデルの表現可能な領域を示している. また, 図 2.15 には同時に実ネットワークも示している. 実ネットワークは, 4つのカテゴリ (14 social, 12 biological, 36 infrastructure, 6 information) にまたがるネットワーク 68

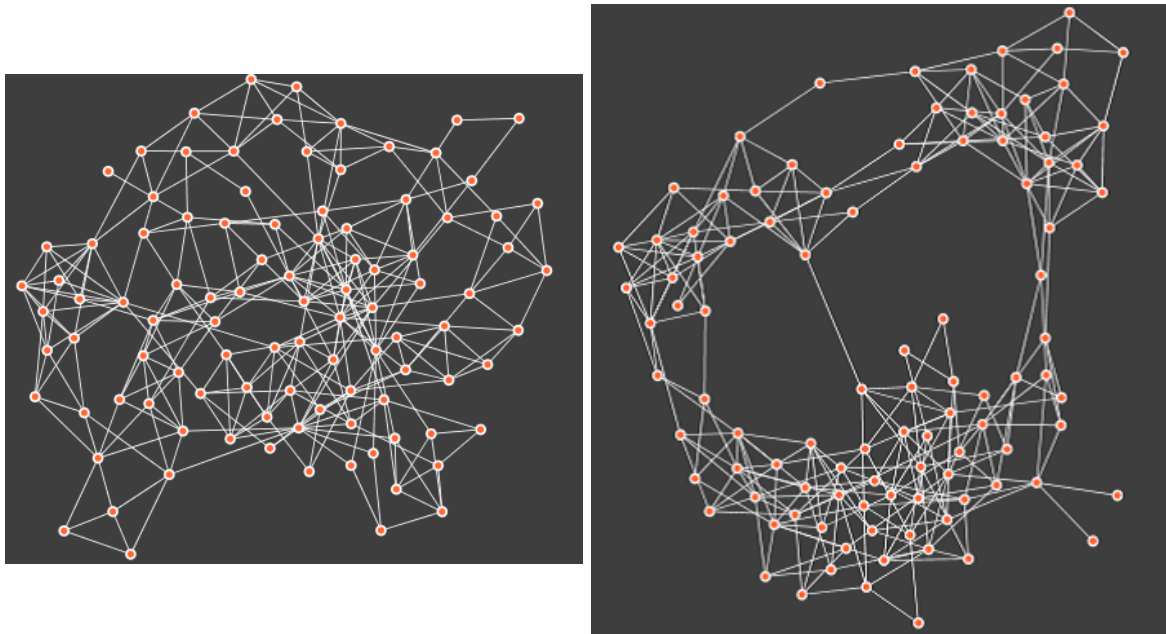


図 2.14: XS アルゴリズムのネットワーク (左：正の次数相関, 右：負の次数相関)

個を用いた\*2.

まず、ネットワーク生成モデルは確率モデルであり、ネットワークを生成する度に異なる構造のネットワークが生成される。しかし、あるモデルで作られた全てのネットワークには、その生成方法に依存した一定の傾向が表れる。したがって、これらのモデルで生成できるネットワークの性質は非常に限られており、一部の構造を表現するためのものが多い。実際に、ER グラフ (図 2.15 の青い点) と BA モデル (図 2.15 の赤い点) のような制御パラメータのないネットワーク生成モデルでは、図 2.15 に示すようにわずかな領域のネットワークしか生成できない。ER グラフでは平均経路長  $L$  は短く、クラスタ係数  $C$  と次数相関  $r$  はほぼ 0 となる。BA モデルでは、平均経路長  $L$  は ER グラフよりさらに短く、クラスタ係数  $C$  と次数相関  $r$  はほぼ 0 となる。また、図 2.15 では見られないが、ER グラフは次数分布がポアソン分布となり、BA モデルは次数分布がべき分布となる。このように、これらのモデルが生成するネットワークは常に一定のネットワークである。すなわち、これらのネットワーク上で分析を行っても、たかだか 2 つのサンプルネットワークでの比較となり、これは統計的には意味を持たない。

CNN モデルと Fitness モデルは BA モデルの拡張モデルであり、ER グラフや BA モデル同様、制御パラメータを持たない。これらのモデルも同様に確率モデルではあるが、そ

\*2 <https://github.com/basimbaig/robust14>



の生成方法に依存した一定の傾向が表れる。実際に、CNN モデル (図 2.15 の灰色の点) と Fitness モデル (図 2.15 のピンクの点) では、図 2.15 に示すようにわずかな領域のネットワークしか生成できない。CNN モデルでは、クラスタ係数が高く、BA モデルよりは平均経路長が長い。また、Fitness モデルでは、平均経路長が BA モデルよりもさらに短く、次数相関が負の相関を取っている。

次に WS モデル、HK モデルについてみる。WS モデルのネットワークは張り替え確率  $\epsilon$  の値を変化させる事によって、ネットワークの平均経路長とクラスタ係数に変化を与えている。すなわち、ショートカットの作成は平均経路長とクラスタ係数の値を下げる事であると言い換える事も可能である。しかし、このショートカットの作成は平均経路長とクラスタ係数を同時に下げるので、この 2 つの特徴を独立に動かす事は出来ない。従って、平均経路長とクラスタ係数は相関係数 0.665 で相関している。WS モデルを用いた分析の場合、ショートカットが増える事による影響はみられるが、それが現実において何を意味するのかは分からない。また、平均経路長とクラスタ係数のどちらの性質の影響なのかも定かではない。

HK モデルも同様に、平均経路長とクラスタ係数が相関係数 0.913 で相関している。従って、この HK モデルを用いた分析でも平均経路長とクラスタ係数のどちらの性質の影響なのかは分からない。また、これらの 2 つのモデルでは次数相関が 0 であるため、次数相関が 0 である場合の関係しか見る事が出来ない。

また、CNNR モデルと FCR モデルに関しても同様にパラメータの変化によって複数の特徴が同時に動いている事がわかる。CNNR モデルと FCR モデルでは、クラスタ係数と次数相関が共に強く相関しており、CNNR モデルでは相関係数  $-0.969$  で相関し、FCR モデルでは相関係数  $-0.975$  で相関している。従って、これらのモデルでも、構造特徴空間にプロットしてみれば、ほんの一部の領域しか表現できていない事は明確である。

XS アルゴリズムについてみる。XS アルゴリズムのネットワークは様々な次数相関の値を取っているが、次数相関と共に平均経路長も変化している事がわかる。すなわち、次数相関が 0 付近にある時に、平均経路長が最も短いネットワークが得られ、次数相関が高く (または低く) なるにつれて平均経路長が長くなっていく事は明らかである。したがって、XS アルゴリズムを用いた分析では、次数相関が影響しているのか、平均経路長が影響しているのかを判断するのは難しい。また、XS アルゴリズムで生成したネットワークはクラスタ係数がほぼ 0 であるため、クラスタ係数が 0 である場合の関係しか見る事が出来ない。

最後に BS アルゴリズムについてみる。BS アルゴリズムのネットワークは様々な次数相関とクラスタ係数の値を取るが、平均経路長も共に変化している。すなわち、次数相関

やクラスタ係数が小さい程，平均経路長も小さい値を取っている事がわかる．また，このアルゴリズムもネットワーク構造特徴空間の一部しか表現できていない事がわかる．

**XS** アルゴリズム，**BS** アルゴリズム及び**HK** モデルの 3 つのネットワーク生成モデルは，任意の特徴値を持つネットワークを生成する事を目的としている．しかし，ここまででこれらの 3 特徴のみで考えても，極めて狭い領域しか表現する事が出来ない事を明らかにしてきた．これらのネットワークを用いて分析を行っても，偏りのある結果しか出てこないだろう．

また，既存のネットワーク生成モデルでは実ネットワークの性質も捉えていないものが多い．ここで，平均経路長  $L$  が 1 以上離れておらずクラスタ係数  $C$  及び次数相関  $r$  が 0.1 以上離れていないものは 68 個のネットワーク中，わずか 10 個である．それ以外の 58 個のネットワークに関しては，平均経路長  $L$ ，クラスタ係数  $C$ ，次数相関  $r$  の少なくともどれか一つは大きく離れている．すなわち，既存ネットワーク生成モデルでは，実ネットワークの複数の特徴を同時に表現する事が出来ない事がわかった．

ここまで 10 個の手法を挙げてきたが，3 次元特徴空間において，全ての領域を表現できるわけではない．また，既存のネットワーク生成モデルでは実ネットワークの性質も捉えていないものが多い．このように，ネットワーク生成モデルが生成するネットワークには一定の傾向が存在する．これらのネットワーク上で分析を行っても，この特徴空間全体からすれば一部の構造であり，統一的な議論は行えない．ネットワーク構造と社会現象の間の統一的な関係性を示すためには，このような 3 次元構造特徴空間を全て埋めるようなネットワークデータセットが望ましい．そこで，本研究では，与えられた任意の特徴値を持つネットワークを生成する手法を提案し，このような構造特徴空間を全て満たすようなデータセットを構築し，現象の分析を行う．

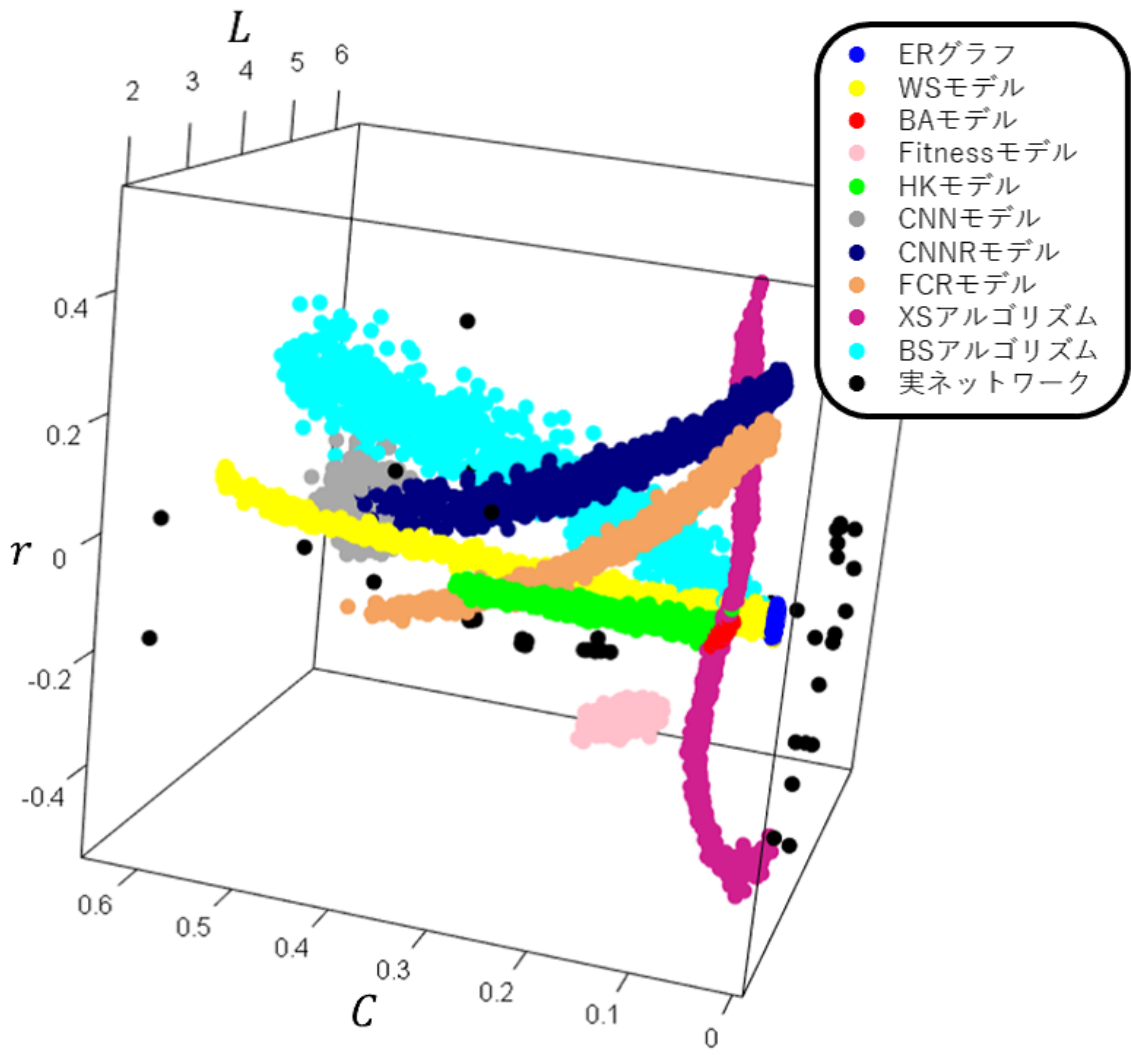


図 2.15: 既存モデルの表現可能領域

## 第 3 章

# 任意の構造を含むネットワークデータセットの構築手法の提案

### 3.1 はじめに

本研究では、様々なネットワーク構造を含むデータセットの構築手法の提案を行う。ネットワーク構造は無数に存在する。例えば、ノード数 1000 のネットワークにおけるリンクは  $\frac{1000 \times 999}{2} = 499500$  通り存在する。つまり、リンク数 10000 のネットワークは  $499500 C_{10000}$  通り存在する。これを網羅する事は事実上不可能であり、また、その必要もない。なぜなら、この  $499500 C_{10000}$  通りのネットワークの中には類似した構造のネットワークが大量に存在するためである。

ネットワーク構造は一般に複数のマクロスケール構造特徴によって表現される。複数のマクロスケール構造特徴が近いネットワークは、類似した特徴を持つネットワークである。本研究では、類似した構造特徴を持つネットワークは類似したネットワークであるとする。ここで、ネットワーク構造特徴を軸とした、ネットワーク構造特徴空間を定義する。類似した特徴を持つネットワークは、ネットワーク構造特徴空間上で近い点を取る。本章では、このネットワーク構造特徴空間を網羅する事で、類似していないネットワーク構造を網羅する事を目指す。このようなデータセットである事を目指すのは、統計的に議論をするに際して、網羅的に議論する事が可能であるためである。

このデータセットを構築するために、我々はこのネットワーク構造特徴空間上の任意の点にネットワークを生成する事を考える。これができれば、ネットワーク構造特徴空間に散らばるようにネットワークを構築する事によって、ネットワーク構造特徴空間を網羅するネットワークデータセットを構築する事が出来る。現在まで提案されている既存のネッ

トワーク生成モデルや、アルゴリズムで生成されたネットワークは、その生成過程に対応した構造特徴を持つため、任意の点にネットワークを構築する事はできない。

本研究では、ネットワーク構造特徴空間上の任意の点にネットワークを生成する手法として、Arbitrary Network Generation Algorithm(ANGA)を提案する。ネットワークの構造はいくつかの構造特徴によって表現されるが、ANGAでは逆に構造特徴からネットワークを生成する手法である。すなわち、あらかじめネットワーク構造特徴を任意に設定し、それを基にネットワークを生成する。ANGAは貪欲な手法を用いる事によって、設定された特徴点に向かってネットワークを成長させる手法である。

本章では、第3.2節にて、提案するANGAの詳細について説明し、ANGAの性能の実証を行う。次に、第3.3節にて、ANGAを用いてネットワーク構造特徴空間を網羅するようなネットワークデータセットを構築し、その性質について検証を行う。

## 3.2 Arbitrary Network Generation Algorithm

### 3.2.1 概要

本節では、Arbitrary Network Generation Algorithm(ANGA)について説明する。ANGAはネットワーク構造特徴空間上の任意に設定した点にネットワークを生成するアルゴリズムである。ANGAはConnecting Nearest Neighbor(CNN)モデル [74] に用いられているネットワーク成長モデルを基にした手法となる。各ステップにおいて、以下の2つのいずれかを確率  $p$  によって選択する。

**確率  $p$  ノードの追加**

新規ノードを一つ追加し、現ステップネットワークから選択されたノードとの間にリンクを一つ追加する。

**確率  $1 - p$  リンクの追加**

現ステップネットワークから2つのノードを選択し、その間にリンクを追加する。

この時、生成するネットワークのノード数  $N$  とリンク数  $M$  から確率  $p$  は以下のように算出される。

各ステップにおいて、ノードの追加では、ノードとリンクが1つずつ増え、リンクの追加ではリンクが1つ増える。すなわち、 $i$  ステップ目のネットワークのノード数  $n_i$  とリンク数  $m_i$  は以下のように表される。

$$n_i = i \cdot p \quad (3.1)$$

$$m_i = i \quad (3.2)$$

$$n_i = m_i \cdot p \quad (3.3)$$

あるステップにおいて、ネットワークはノード数  $N$ 、リンク数  $M$  となる必要がある。すなわち、 $n_i = N$ 、 $m_i = M$  となる  $i$  が存在するように  $p$  を設定する必要がある。従って、確率  $p$  は以下の様に計算できる。

$$N = M \cdot p \quad (3.4)$$

$$p = \frac{N}{M} \quad (3.5)$$

この確率  $p$  に従う事によって、ネットワークはノード数  $N$ 、リンク数  $M$  に収束する。

ノードの追加とリンクの追加の際に、**ANGA** では複数のネットワーク候補を生成する。各ステップにおいて、この複数のネットワーク候補の中から最適なネットワークを選出する事によって、ネットワークはネットワーク構造特徴空間上の指定した点に向けて成長する。なお、**アルゴリズム 9** に **ANGA** のアルゴリズムの概要を示す。

以下では、ネットワーク候補の生成方法と最適ネットワークの選択方法について述べる。なお、本手法において重要であるのはこの枠組みであり、この候補の生成方法や最適ネットワークの選択方法に関しては、よりよい方法と容易に入れ替え、拡張する事が可能である。すなわち、拡張性の高い手法となっている。

### 3.2.2 候補ネットワークの生成

候補ネットワークの生成においては、ノードの追加かリンクの追加のどちらかを行い、現ステップのネットワークからネットワーク候補を生成する。

#### 候補の生成

ノードの追加では、現ステップネットワークに新規ノードを追加し、ネットワークから選択したノードとの間にリンクを追加する。これを一つの候補ネットワークとして、 $c$  個の候補ネットワークを生成する。

リンクの追加では、現ステップネットワークから 2 つのノードを選択し、選択されたノード間にリンクを追加する。これを一つの候補ネットワークとして、 $c$  個の候補ネットワークを生成する。ノードの選択方法に関しては後述する。

---

**Algorithm 9** Arbitrary Network Generation algorithm

---

**Ensure:** 空のネットワークを用意する.

**while** 現ステップネットワークのリンク数が生成するネットワークのリンク数  $M$  になるまで繰り返す **do**

(ネットワーク候補の生成)

**for** 候補ネットワークの生成数 **do**

**if**  $p = \frac{N}{M}$  **then**

ノードの追加

新規ノードを一つ追加し, 現ステップネットワークから選択されたノードとの間にリンクを一つ追加した候補ネットワークを作成する.

**else**

リンクの追加

現ステップネットワークから 2 つのノードを選択し, その間にリンクを追加した候補ネットワークを作成する.

**end if**

**end for**

(最適ネットワークの選択)

候補ネットワークの中から最適なネットワークを選択して次ステップネットワークとする.

**end while**

---

### ノードの選択方法

本手法はノードの追加とリンクの追加において, 以下の 3 つのノードの選択方法を用いる.

- 優先的選択

ノードを選択する確率が, そのノードの持つ次数に比例する選択方法であり, 次数  $k_i$  を持つノードが選択される確率  $\pi_i$  を以下の式で表す.

$$\pi_i = \frac{k_i}{\sum_{j \in N} k_j} \quad (3.6)$$

- Connectin nearest neighbor(CNN)

CNN は一つのノードが選択された際に, そのノードから距離が 2 であるノードの

表 3.1: ノードの選択手法の組み合わせによる type 分け

type-R	ランダム選択
type-P	優先的選択
type-RC	ランダム選択, CNN
type-RP	ランダム選択, 優先的選択
type-PC	優先的選択, CNN
type-RPC	ランダム選択, 優先的選択, CNN

中からランダムに選択する選択方法である。したがって、この手法はリンクの追加のみで有効である。

- ランダム選択

ランダム選択は、ネットワークの中から一様確率で一つのノードを選択する。

これらの選択方法の組み合わせによって、7 個の type に分ける。

表 3.1 に type 毎の選択手法の組み合わせを示す。以下では例として、type-RPC の場合を示す。type-RPC では、ノードの追加の際には以下の 2 つの手法のいずれかを等確率で用いてノードを選択する。

- ランダム選択
- 優先的選択

リンク追加の際には、1 つ目のノードをランダムで選択し、2 つ目のノードを以下の 3 つの手法のいずれかを等確率で用いてノードを選択する。

- ランダム選択
- 優先的選択
- CNN

図 3.1 に ANGA の type-RPC の概要を示す。その他の type では、用いられないノードの選択手法が除かれた形になる。



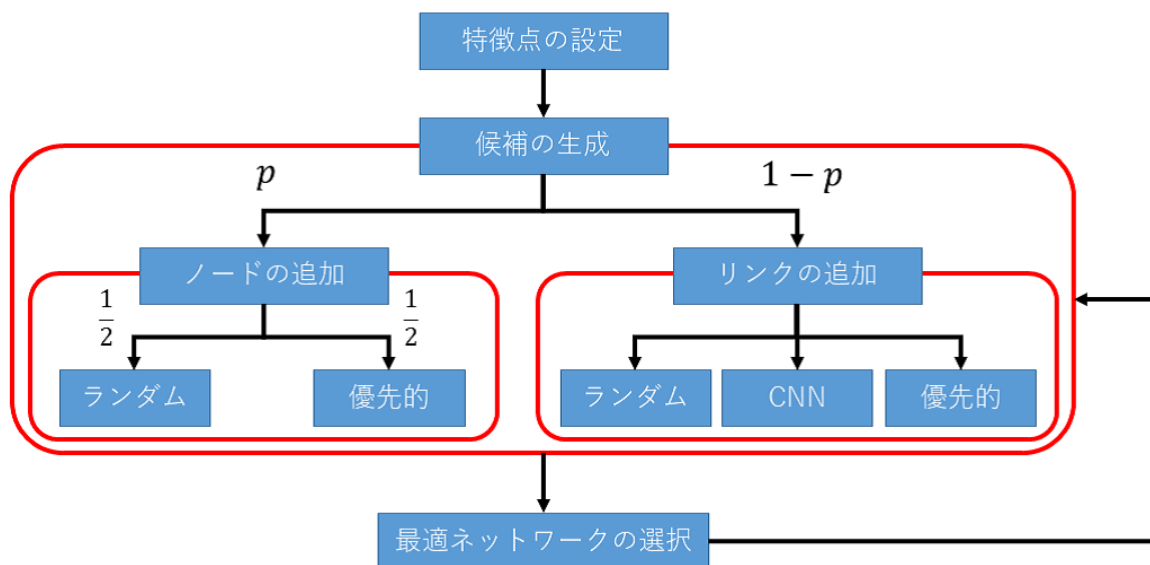


図 3.1: ANGA type-RPC の概要

### 3.2.3 最適ノードの選択

ANGA では、上記によって生成された複数の候補ネットワークの中から最適なネットワークを選択する。選択には、特徴点間距離を用い、候補ネットワークの中から目標とする特徴点に最も近いネットワーク候補を選択する。特徴点間距離としては、ユークリッド距離  $D_{\mathbf{F}}$  を採用する。

$$D_{\mathbf{F}} = \frac{1}{|\mathbf{F}|} \sum_{f \in \mathbf{F}} \left( \frac{f_a - f_b}{\sigma_f} \right)^2 \quad (3.7)$$

なお、 $\mathbf{F}$  は特徴集合とし、 $f_i$  は特徴点  $i$  の特徴  $f$ 、 $\sigma_f$  は特徴  $f$  に関する分散とする。この際、 $\mathbf{F}$  として用いる特徴集合は任意の特徴を用いる事が可能である。

### 3.2.4 ANGA の評価実験

本節では、ANGA が任意の特徴点にネットワークを生成可能であることを示す。ここでは、2つの評価実験を行う。一つは、実ネットワークから抽出した特徴点にネットワークを生成可能であることを示す。もう一つは、既存のネットワーク生成モデルを用いて生成した特徴点にネットワークを生成可能であることを示す。これら二つを示す事によって、ANGA が様々な特徴点にネットワークが生成可能であることを示す。

本節ではまず、評価実験に用いる特徴量の説明を行う。次に、それぞれの評価実験の設定と結果について解説する。

## ネットワーク構造特徴

本研究で用いるネットワーク構造特徴  $\mathbf{F}$  は以下の 4 つである。

- 平均経路長  $L$  [27],
- クラスタ係数  $C$  [27],
- 次数相関  $r$  [30],
- 次数分布  $p_k$ .

すなわち、 $\mathbf{F} = \{L, C, r, p(k)\}$  となる。それぞれについて以下で解説する。

### ■平均経路長 $L$

平均経路長  $L$  は、ネットワーク上に存在する任意の 2 ノード間の最短経路長の平均である。2 ノード間の最短経路長は、あるノードからあるノードまで最短で辿り着くまでに通ったリンクの数と定義される。ノード数  $N$  のネットワークにおける平均経路長  $L$  は以下の式で定義される。

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j} l_{ij} \quad (3.8)$$

なお、 $l_{ij}$  はノード  $v_i$  とノード  $v_j$  の間の最短経路長である。平均経路長はネットワークの長さを定義する指標である。

### ■クラスタ係数 $C$

クラスタ係数  $C$  は、あるノードの隣接ノード間にリンクが存在する確率を表し、以下の式で定義される。

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (3.9)$$

$$C = \frac{1}{N} \sum_i C_i \quad (3.10)$$

ここで、 $k_i$  はノード  $v_i$  の次数を表し、 $E_i$  はノード  $v_i$  の隣接ノード間に存在するリンクの数を表す。実社会に存在するネットワーク（人間関係、論文の引用関係など）の多くは、ノード数  $N$  が大きい場合でも、平均経路長  $L$  が小さく、クラスタ係数  $C$  が比較的大きいという性質を持つことが知られている。このような性質はスモールワールド性と呼ばれ、平均経路長が短く、クラスタリング係数が高いとき、スモールワールド性が高いという。

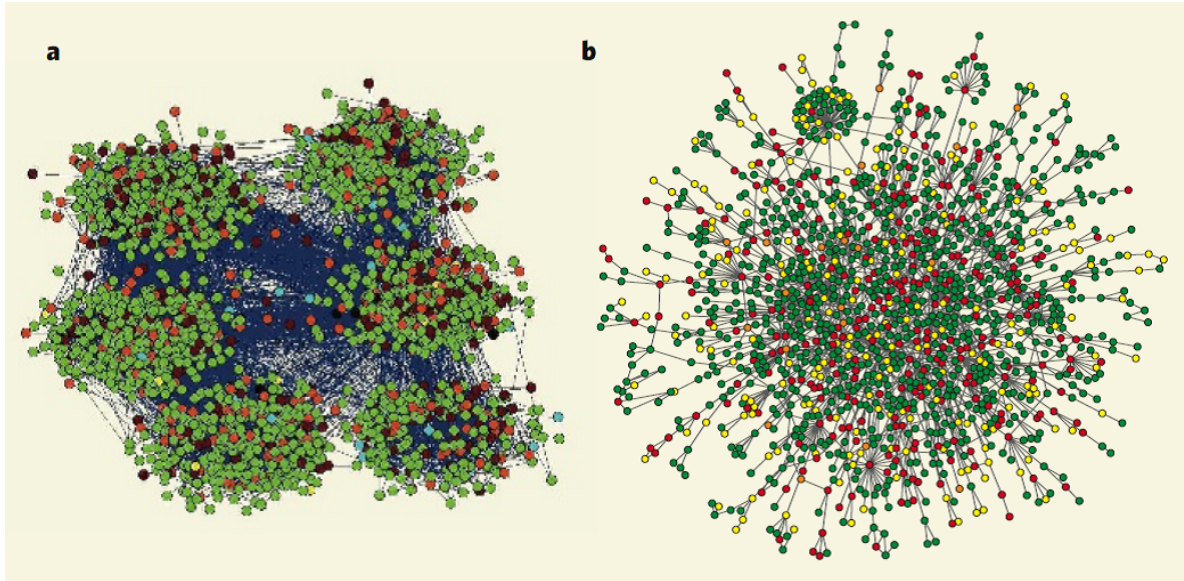


図 3.2: (a). 正の次数相関を持つネットワーク (b). 負の次数相関を持つネットワーク (出典 : Teasing out the missing links)

### ■次数相関 $r$

次数相関  $r$  は、各リンクの両端にあるノードの次数間の相関を指標である。次数相関  $r$  は以下の式で定義される。

$$r = \frac{M^{-1} \sum_m k_m l_m - [M^{-1} \sum_m \frac{1}{2}(k_m + l_m)]^2}{\frac{1}{2} M^{-1} \sum_m (k_m^2 + l_m^2) - [M^{-1} \sum_m \frac{1}{2}(k_m + l_m)]^2} \quad (3.11)$$

ここで、 $M$  は全リンク数を表し、 $k_m, l_m$  はリンク  $m$  の両端にある 2 ノードの次数を表す。なお、次数相関  $r$  は  $-1 \leq r \leq 1$  の値をとる。次数相関  $r$  が正の時は、次数が高いノード同士、低いノード同士が接続されており、負の時は、次数が高いノードと低いノードが接続されている傾向にある事を示している。図 3.2(a) に次数相関が正のネットワーク、図 3.2(b) に次数相関が負のネットワークの例を示す。一般に、共演・共著関係といった人間関係のネットワークでは  $r > 0$  となり、遺伝子ネットワーク、タンパク質ネットワーク、神経回路など自然界のネットワークでは  $r < 0$  となることが知られている。また、電力供給網やインターネットのような工学的なネットワークでは比較的小さいながらも次数相関が負の相関を示すことが知られている。

### ■次数分布 $p(k)$

ネットワークの次数分布は、次数  $k$  を持つノードの存在確率分布  $p(k)$  として定義され

る。すなわち、 $p(k)$  はネットワークからランダムにノードを選択した場合に、そのノードの次数が  $k$  である確率を表す。本研究では、ネットワークが持つ  $p(k)$  からベータ分布  $q(k)$  を構築する。

$$q(k) = \frac{x^\alpha(1-x)^\beta}{B(\alpha, \beta)}, \quad (3.12)$$

$$B(\alpha, \beta) = \int t^{x-1}(1-t)^{y-1}, \quad (3.13)$$

なお、 $\alpha$  及び  $\beta$  はフィッティングパラメータである。ここで、次数分布における  $x$  は以下の式で表される。

$$x = \frac{k}{k_{max} + 1}. \quad (3.14)$$

また、ベータ分布を確率分布とする (すなわち、 $\sum_k^{k_{max}} q(k) = 1$ ) ために、以下の処理を行う。

$$q'(k) = \frac{q(k)}{\sum_{k'}^{k'_{max}} q(k')}. \quad (3.15)$$

ベータ分布は  $\alpha$  及び  $\beta$  を調整する事によって、様々な分布の形状を表現する事が可能である。 $\alpha = \beta$  の時、 $x = 0.5$  が平均値となる正規分布のような形となり、 $\alpha < \beta$  の時、グラフのピークは  $x < 0.5$  となり、逆に  $\alpha > \beta$  の時、グラフのピークは  $x > 0.5$  となる。また、 $\alpha$  及び  $\beta$  の値が大きくなるほど、グラフの尖度が増加していく。

目標特徴点をネットワークから作る場合、そのネットワークからベータ分布を構築し、特徴とする。ネットワークの次数分布に対して  $\alpha$  及び  $\beta$  を、焼きなまし法 (Simulated Annealing:SA) [81] によって最適化する。最適化の際の目的関数として、ネットワークの次数分布  $p(k)$  とベータ分布  $q(k)$  の分布間二乗距離を、この特徴に関する距離として定義する。

ここで、実ネットワークに対して、それぞれ  $\alpha$  及び  $\beta$  を最適化し、分布間二乗距離を算出する。実データとして、SNAP<sup>\*1</sup>によって提供されているデータセットを用いた。これらのデータセットの中から Facebook の ego-network 10 個を用いる [75]。Ego-network とは、被験者として選択した ego-node とリンクしているノードのみによって形成されるネットワークである。つまり、ego-node の友人間に形成される友人ネットワークである。

分布間二乗距離の平均は 0.00639 であり、非常に小さい値となった。従って、ネットワークの次数分布をベータ分布によって表現する事が出来ている事が示された。図 3.3 に

---

\*1 <http://snap.stanford.edu/data/>

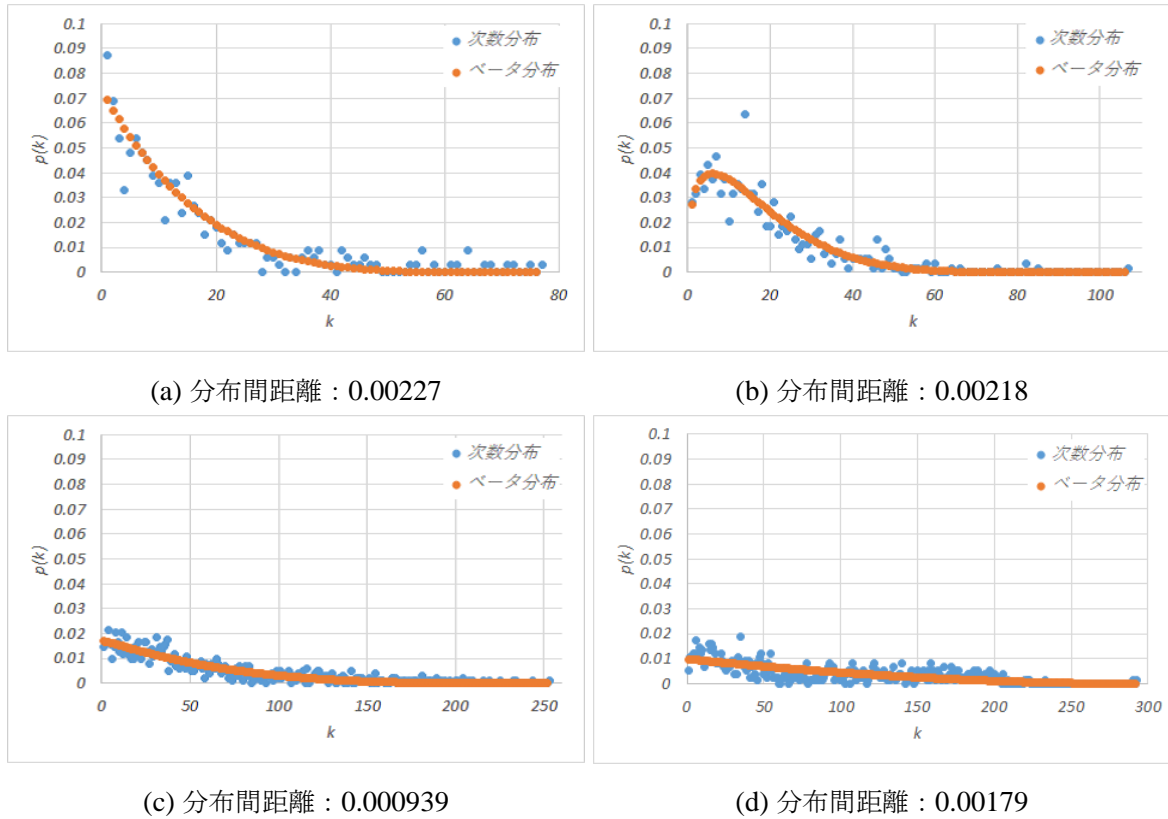


図 3.3: 実ネットワークと最適化したベータ分布

実ネットワークの次数分布と、それに対して最適化したベータ分布を示す。図 3.3 から、様々な次数分布をよく表現できている事がわかる。

### 実ネットワークを特徴点として用いた評価実験

続いて、実データを用いて ANGA がどの程度任意の特徴点を生成可能かを評価する。実ネットワークとしては、Facebook の ego-network を用いる。表 3.2 に Facebook の ego-network のそれぞれの特徴を示す。

まず初めに、候補生成数  $c$  について述べる。使用する type は type-RPC を用いる。図 3.4 に候補生成数  $c$  を  $c = 10, 20, 30, 40, 50$  と変化させた時の、type-RPC における生成精度を示す。横軸は距離  $D_F$  であり、縦軸は距離  $D_F$  が横軸の値以下で生成できたネットワークの割合を示している。すなわち、 $D_F < 0.1$  で  $c = 10$  では 8 個のネットワークを、その他では 10 個全てを生成できている事を示している。この図を見てみると、候補生成数  $c$  が多い程精度が上がっていくのがよくわかる。しかし、候補生成数  $c = 10$  から  $c = 30$  までは大きく精度が上昇しているが、それより大きいところでは、そこまで大きな

表 3.2: Facebook の ego-network の指標

Number	$N$	$M$	$L$	$C$	$r$	$\log\alpha$	$\log\beta$
1	40	220	1.95	0.725	0.0125	0.00	0.424
2	44	138	2.57	0.455	0.0530	0.285	0.703
3	148	1692	2.69	0.679	0.304	0.120	0.299
4	168	1656	2.43	0.534	0.0841	0.0520	0.569
5	224	3192	2.52	0.544	0.223	0.00	0.432
6	324	2514	3.75	0.522	0.233	0.00	0.738
7	532	4812	3.45	0.546	0.222	0.139	0.864
8	744	30023	2.56	0.638	0.503	0.00	0.461
9	775	14006	3.04	0.471	0.327	0.0254	0.518
10	1034	26749	2.95	0.526	0.431	0.00	0.653

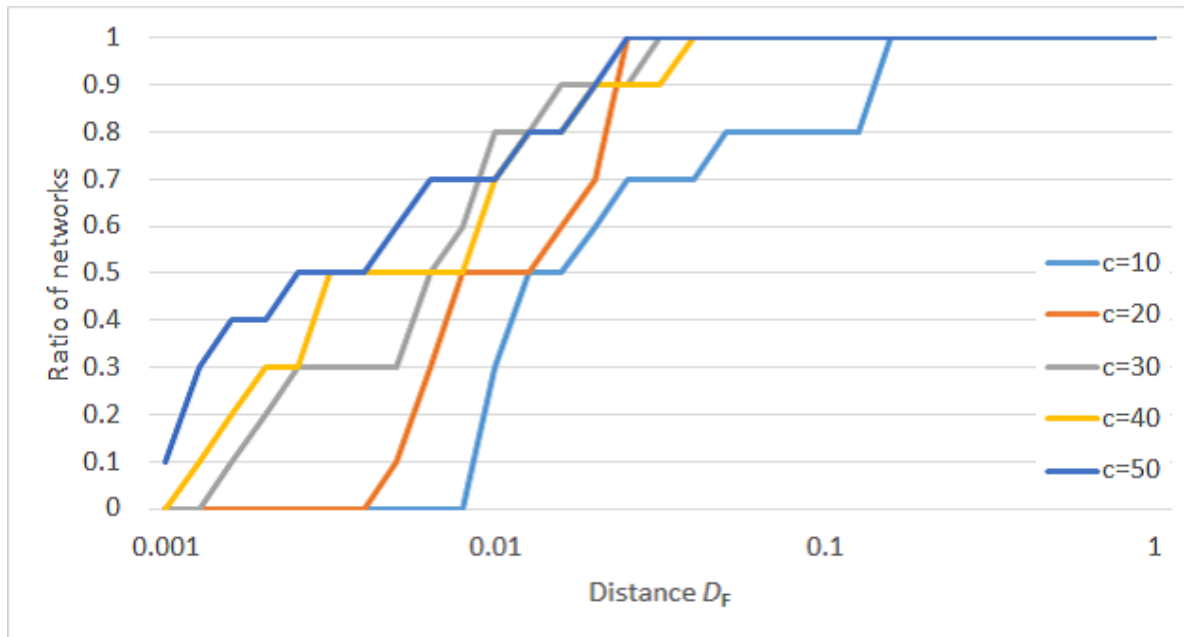


図 3.4: 候補生成数  $c$  を変化させた時の, type-RPC における生成精度

精度の向上は見込まれない。当然、この生成精度と生成時間はトレードオフの関係になっており、候補生成数を増やせば精度の向上は見込めるが、大量の計算時間を必要とする。従って、候補生成数  $c = 30$  より増やしても大きな精度向上が見込まれない以上、候補生成数  $c = 30$  で十分であると言える。以降では、候補生成数  $c = 30$  として、ANGA を運

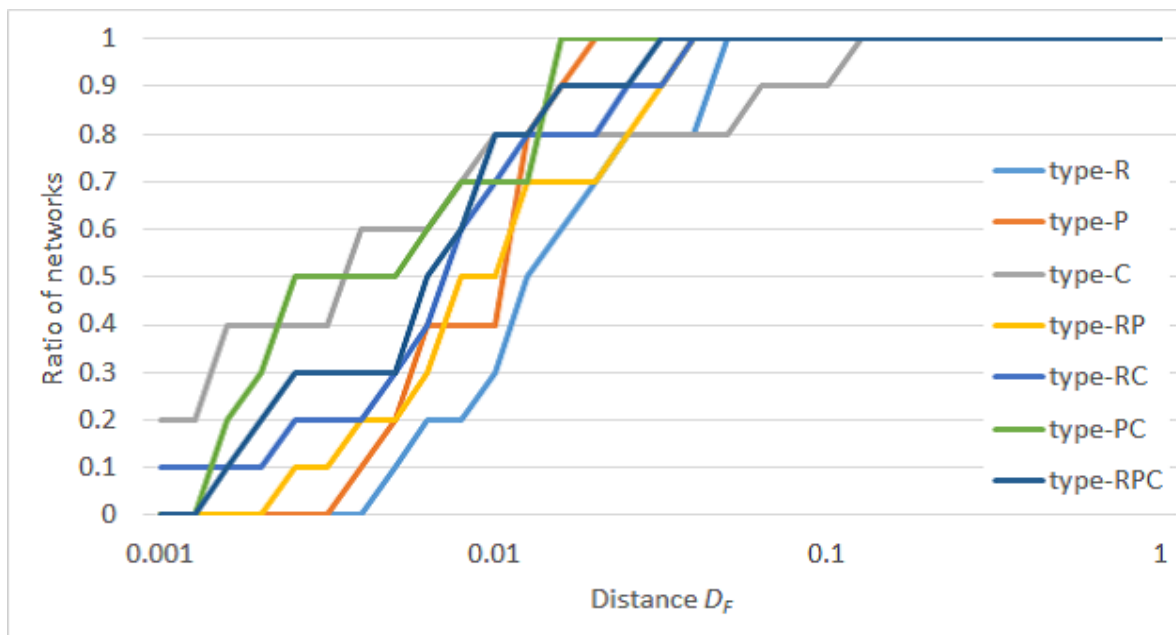


図 3.5: 実ネットワークに対する type 別のネットワーク間距離

表 3.3: 実データに対する各 type における平均値  $\overline{D_F}$  (昇順)

	type-PC	type-RPC	type-P	type-RC	type-RP	type-R	type-C
$\overline{D_F}$	0.00622	0.00832	0.0109	0.0131	0.00960	0.0187	0.0189

用する。

図 3.6 に各 type において、10 個のネットワークの内、ある距離以下で生成できたネットワークの割合を示す。横軸にネットワーク間距離を取り、縦軸には横軸の距離  $D_F$  の値以下の距離で生成できたネットワークの割合を示す。つまり、type-RP と type-C では 8 個のネットワークを距離 0.01 以内で生成できている事が示されている。しかし、type-C では残り 2 つのネットワークに対する精度が高くない。これは type-C がクラスタ係数が非常に高い時のみ精度よくネットワークを生成できるが、それ以外の時の精度は低いためである。10 個のネットワーク全てが最も小さい距離で生成できているのは type-PC である。

表 3.3 に各 type における平均値  $\overline{D_F}$  を示す。この平均値からは、精度よく生成できている type は type-PC 及び type-RPC であり、最も精度が悪いのは type-C であることがわかる。図 3.5 の type-RPC を見ると、10 個全てを生成できる距離は、type-PC、type-P に続く 3 番目である。しかし、type-P と比較すると、1 つを除く 9 個のネットワークの精度は type-RPC の方が高い。従って、type-PC 及び type-RPC が精度よく特徴点にネット

表 3.4: 実データに対する各 type における平均値  $\overline{D_F}$

Network ID	$N$	$M$	$\overline{D}$
1	40	220	$2.65 \times 10^{-2}$
2	44	138	$3.83 \times 10^{-2}$
3	148	1692	$1.33 \times 10^{-2}$
4	168	1656	$8.66 \times 10^{-3}$
5	224	3192	$8.89 \times 10^{-3}$
6	324	2514	$3.48 \times 10^{-3}$
7	532	4812	$5.88 \times 10^{-3}$
8	744	30023	$8.41 \times 10^{-3}$
9	775	14006	$4.66 \times 10^{-3}$
10	1034	26749	$4.66 \times 10^{-3}$

ワークを生成できている事が示された。

次に、ネットワーク毎の精度の差を見てみる事にする。表 3.4 に各ネットワーク毎の平均値  $\overline{D_F}$  を示す。明らかに 1 番, 2 番のネットワークの生成精度が低い。このネットワークは、ノード数  $N$ , リンク数  $M$  が共に小さいネットワークである。ノード数が小さいネットワークであると、貪欲な手法により特徴点に近づく前に処理が終了してしまう。また、ネットワークのサイズが大きくなる程精度が向上していく事がわかる。これより、ANGA はある程度の大きさがなければ、任意の特徴点にネットワークを生成する事は困難である事がわかる。

次に、最も大きなネットワークである 10 番のネットワークに絞って、各 type でどのようなネットワークが生成されているのかを見てみる。表 3.5 に各 type におけるネットワーク構造特徴量の例を示す。全ての type で、小さい平均経路長, 高いクラスタ係数, 高い次数相関をそれぞれ再現できている事がわかる。

### 既存ネットワーク生成モデルを用いた評価実験

ここでは既存のネットワーク生成モデルを用いて、ANGA がどの程度任意の特徴点を生成可能かを評価する。既存ネットワーク生成モデルを用いて、ネットワークを生成し、それらから抽出した特徴点を対象として、ネットワークの生成を行う。既存ネットワーク生成モデルとしては前章にて解説した以下のものを用いる。



表 3.5: 各 type によるネットワークサンプル

type	$L$	$C$	$r$	$\log\alpha$	$\log\beta$	$D_{\mathbf{F}}$
Real data	2.95	0.526	0.431	0.00	0.653	-
type-R	2.96	0.521	0.428	0.00	1.03	0.00481
type-P	2.95	0.525	0.425	0.00	0.957	0.00357
type-C	2.95	0.525	0.431	0.00	0.905	0.00135
type-RP	2.95	0.524	0.422	0.00	0.802	0.00359
type-RC	2.94	0.522	0.424	0.00	0.768	0.00437
type-PC	2.95	0.525	0.428	0.00	0.810	0.00162
type-RPC	2.93	0.523	0.420	0.00	0.928	0.0133

- ER グラフ
- BA モデル
- WS モデル
- Fitness モデル
- HK モデル
- CNN モデル
- CNNR モデル
- FCR モデル
- XS アルゴリズム
- BS アルゴリズム

それぞれのネットワークモデルで 10 個ずつ、すなわち 100 個のネットワークを生成し、生成されたネットワークから特徴点を抽出する。各ネットワークのノード数は 1000 とし、平均次数  $\bar{k}$  は  $19 \leq \bar{k} \leq 20$  となるように生成した。また、各モデルパラメータはランダムに変化させネットワークを生成する。また、XS アルゴリズムにおけるあらかじめ設定したネットワークとして、BA model で生成したネットワークを用いる。

図 3.6 に 90 個のネットワークの内、ある距離以下で生成できたネットワークの割合を示す。横軸に距離  $D_{\mathbf{F}}$  を取り、縦軸には横軸の距離  $D_{\mathbf{F}}$  の値以下の距離で生成できたネットワークの割合を示す。すなわち、type-P, type-RP, type-PC, type-RPC は、80% 以上のネットワークを距離 0.01 以内で生成できている事が示されている。

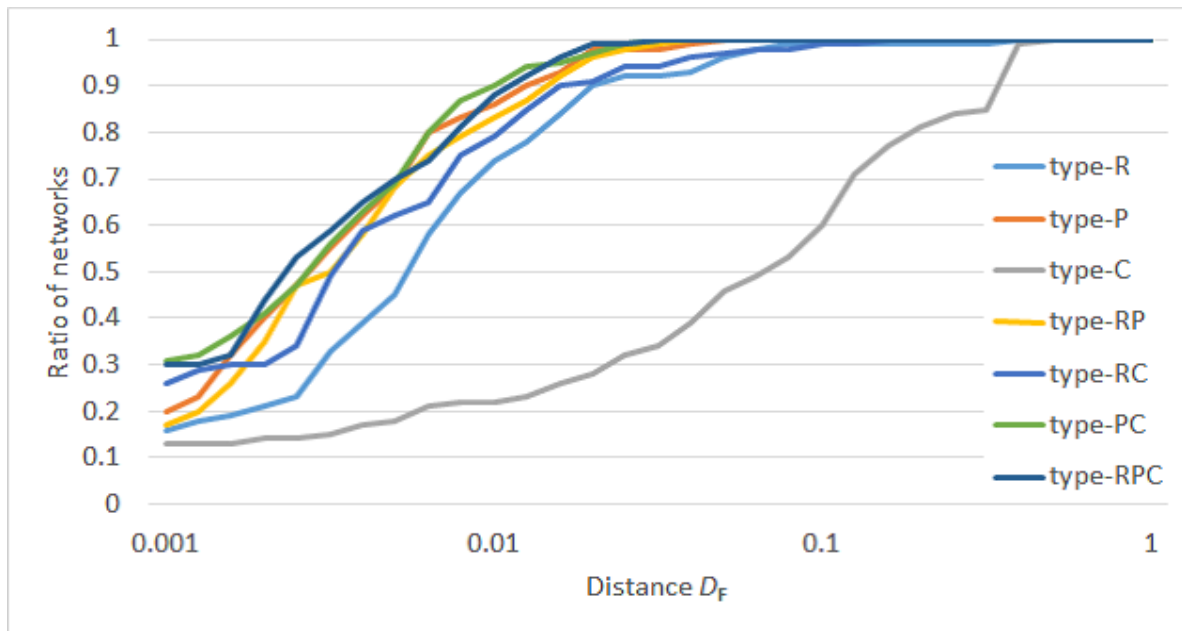


図 3.6: 既存モデルで生成したネットワークに対する type 別のネットワーク間距離

表 3.6: 既存モデルによるネットワークに対する各 type における平均値  $\overline{D_F}$  (昇順)

	type-PC	type-RPC	type-P	type-RP	type-RC	type-R	type-C
$\overline{D_F}$	0.00428	0.00435	0.00502	0.00540	0.00851	0.0129	0.110

まず, type-C は明らかに精度が低い事が見て取れる. type-C はリンクを形成する時に, 三角形を作る様にしかリンクが出来ない. 従って, 必ずクラスタ係数の非常に高いネットワークとなるため, クラスタ係数の低いネットワークに対して精度が低い. 次に, type-R 及び type-RC がわずかに生成精度が低い. 精度が低い type は全て優先的選択を用いない type である. すなわち優先的選択が重要な要素となっている事がわかる.

表 3.6 に各 type における平均値  $\overline{D_F}$  を示す. この平均値からも type-C が非常に精度が低い事がわかる. また, type-P, type-PC, 及び type-RPC は同等に精度のよい結果が出ている.

次に, 各ノードの選択方法について分析する. 優先的選択を含む type-PC の平均距離 0.00428 は type-C の平均距離 0.110 より小さく, type-RP の平均距離 0.00540 は type-R の平均距離 0.0129 より小さく, type-RPC の平均距離 0.00435 は type-RC の平均距離 0.00851 より小さい. 従って, 優先的選択を含まない type よりも優先的選択を含む typeの方が高精度で特徴点にネットワークを生成できている. また, CNN 選択を含む

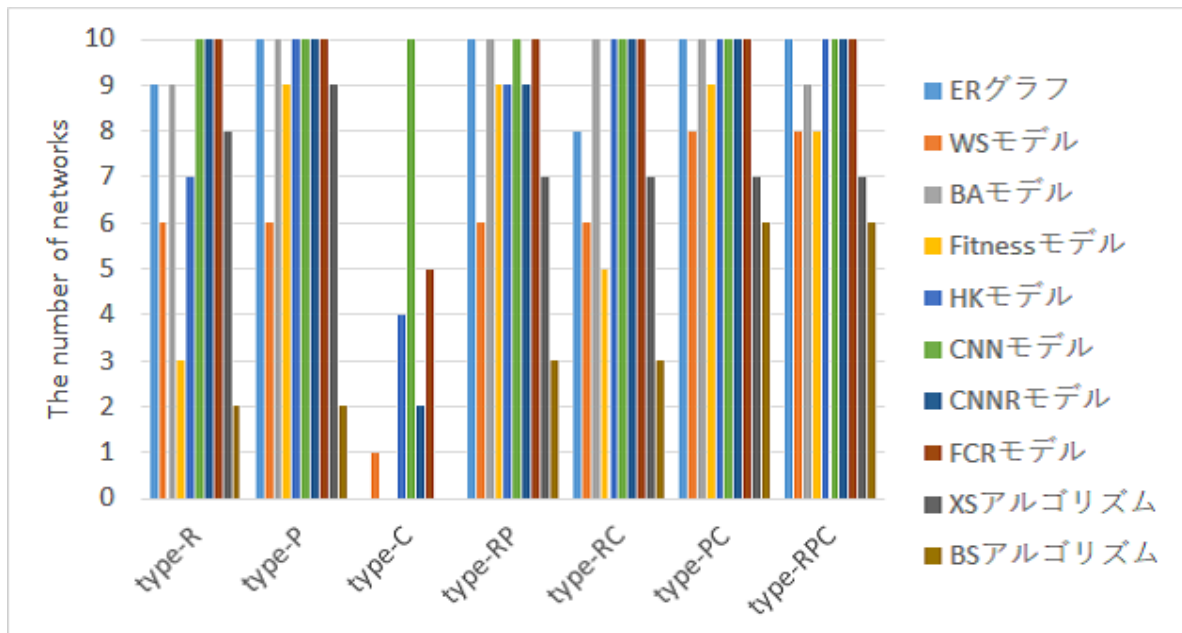


図 3.7: 各モデル毎に距離  $D < 0.01$  で生成できたネットワークの割合

type-PC の平均距離 0.00428 は type-P の平均距離 0.00502 より小さく，type-RC の平均距離 0.00851 は type-R の平均距離 0.0129 より小さく，type-RPC の平均距離 0.00435 は type-RP の平均距離 0.00540 より小さい．従って，CNN 選択を含まない type よりも CNN 選択を含む type の方が高精度で特徴点にネットワークを生成できている．すなわち，優先的選択と CNN 選択が重要である事が示された．一方，ランダム選択を含む type-RC の平均距離 0.00851 は type-C の平均距離 0.110 より小さいが，type-RP の平均距離 0.00540 は type-P の平均距離 0.00502 より大きく，type-RPC の平均距離 0.00435 は type-PC の平均距離 0.00428 よりも大きい．すなわち，ランダム選択は重要な役割を果たしていない事が示された．

次に， $D < 0.01$  における，既存モデル毎の生成精度を確認する．図 3.7 に各モデル毎に，距離  $D < 0.01$  で生成できたネットワークの割合を示す．まず，type-C では，クラスタ係数の非常に高くなる，CNN モデル以外のネットワークは生成精度が非常に低い事がわかる．ここから type-C がクラスタ係数の低いネットワークを生成できない事が裏付けられる．次に優先的選択に着目する．優先的選択を含む type-RPC，PC，RP，P では，その他の type よりも Fitness モデルから抽出した特徴点に対する精度が高い事がわかる．つまり，Fitness モデルのような，ハブノードを複数持つようなネットワークは優先的選択を用いなければ生成できない事が示唆される．次に，CNN 選択に着目する．CNN 選択を含

表 3.7: type-PC 及び type-RPC で生成したネットワークサンプル

type	$L$	$C$	$r$	$\log\alpha$	$\log\beta$	$D_F$
CNNR モデル	2.88	0.101	0.261	0.00526	0.831	-
type-PC	2.88	0.102	0.261	0.00526	0.830	0.000107
type-RPC	2.88	0.102	0.262	0.00526	0.831	0.000303
FCR モデル	2.74	0.0958	0.197	0.122	0.980	-
type-PC	2.74	0.0961	0.197	0.122	0.980	0.000195
type-RPC	2.75	0.0959	0.197	0.122	0.980	0.000238

まない type-RP, P, R でも, CNN モデルのネットワークから抽出した特徴点に高精度で生成出来ている事がわかる. また, XS アルゴリズムのネットワークから抽出した特徴点に対する精度は type-P 以外, やや低い精度となっている. 精度の低かった XS アルゴリズムのネットワークを見てみると, 非常に強い負の次数相関を持っていた. 従って, 非常に強い負の次数相関を持つネットワークを生成するためには, 優先的選択が重要である事がわかる. また, 全体として BS アルゴリズムと WS モデルのネットワークに対する精度が低い事がわかる. ANGA の用いるネットワーク成長モデルでは, この2つの構造に共通する円環構造を持っているネットワークの構造は生成できない. このような特徴点により高精度にネットワークを生成する事は今後の課題とする.

最後に type-PC 及び type-RPC を用いて生成したネットワークの例を表 3.7 に示す. 各特徴の値を見れば, 基にしたネットワークの特徴と生成したネットワークの特徴の間にはほとんど差がない. 従って, 高精度で生成されたネットワークは基になったネットワークと同等の特徴を持っている事が示された.

以上より, type-PC 及び, typeRPC によって, 任意の特徴点に対して高精度でネットワークを生成できる事を示した. 実データによる分析でも type-PC が最も良い精度となったため, type-PC が最も有効であるとする. 今後の分析には type-PC を用いて分析を行っていく.

### 3.3 網羅的ネットワークデータセット

本節では, 前節で提案した ANGA を用いて, ネットワーク構造特徴空間を埋めるようなネットワークデータセットを構築する. このデータセットを構築する事によって, 様々

な複雑ネットワーク上で起こる現象について分析を行い、理解を深める事に寄与できる事が期待される。まず、データセットの構築方法について解説を行い、次にこのデータセットの性能について既存モデルと比較しながら検証していく。本節では、以下の4つの検証手法を用いて検証を行う。

- 相関関係
- 既存ネットワークに対する包括性
- 実ネットワークに対する包括性
- ネットワーク構造特徴空間に対する充填率

### 3.3.1 データセットの構築手法

網羅的なデータセットを構築するために、以下の二つの工程を行う。

- ネットワーク構造特徴空間上の点を無作為に抽出する。
- 抽出された点を基に、**ANGA** を用いてネットワークを生成する。

この工程が成功すれば、ネットワーク構造特徴空間上のある点にネットワークを生成する事が出来る。これを繰り返し行う事によって、ネットワーク構造特徴空間上を網羅するようなネットワークデータセットを構築する事が出来る。

本稿では、無作為に抽出する点として、以下の特徴の範囲内から抽出する事とする。

- 平均経路長  $L$  ( $2 \leq L \leq 6$ ),
- クラスタ係数  $C$  ( $0 \leq C \leq 0.6$ ),
- 次数相関  $r$  ( $-0.5 \leq r \leq 0.5$ ),
- ベータ分布のパラメータ  $\log\alpha$  ( $0 \leq \log\alpha \leq 3$ )
- ベータ分布のパラメータ  $\log\beta$  ( $0 \leq \log\beta \leq 4$ )

なお、ベータ分布のパラメータ  $\alpha$  及び  $\beta$  によって構築したベータ分布を特徴量として用いる。

抽出された点の中には、生成するのが困難な点が存在する。例えば、 $\log\alpha = \log\beta$  となった時は、ノード間に次数の差が小さくなる。この時、次数相関は次数の差がなければ負にならないため、次数相関は必ず正となる。このように、ネットワーク構造として取り得ない点が設定される事がある。その場合、**ANGA** を用いても抽出された点に近いネットワークを生成する事が出来ない。**ANGA** により、生成されたネットワークと抽出され

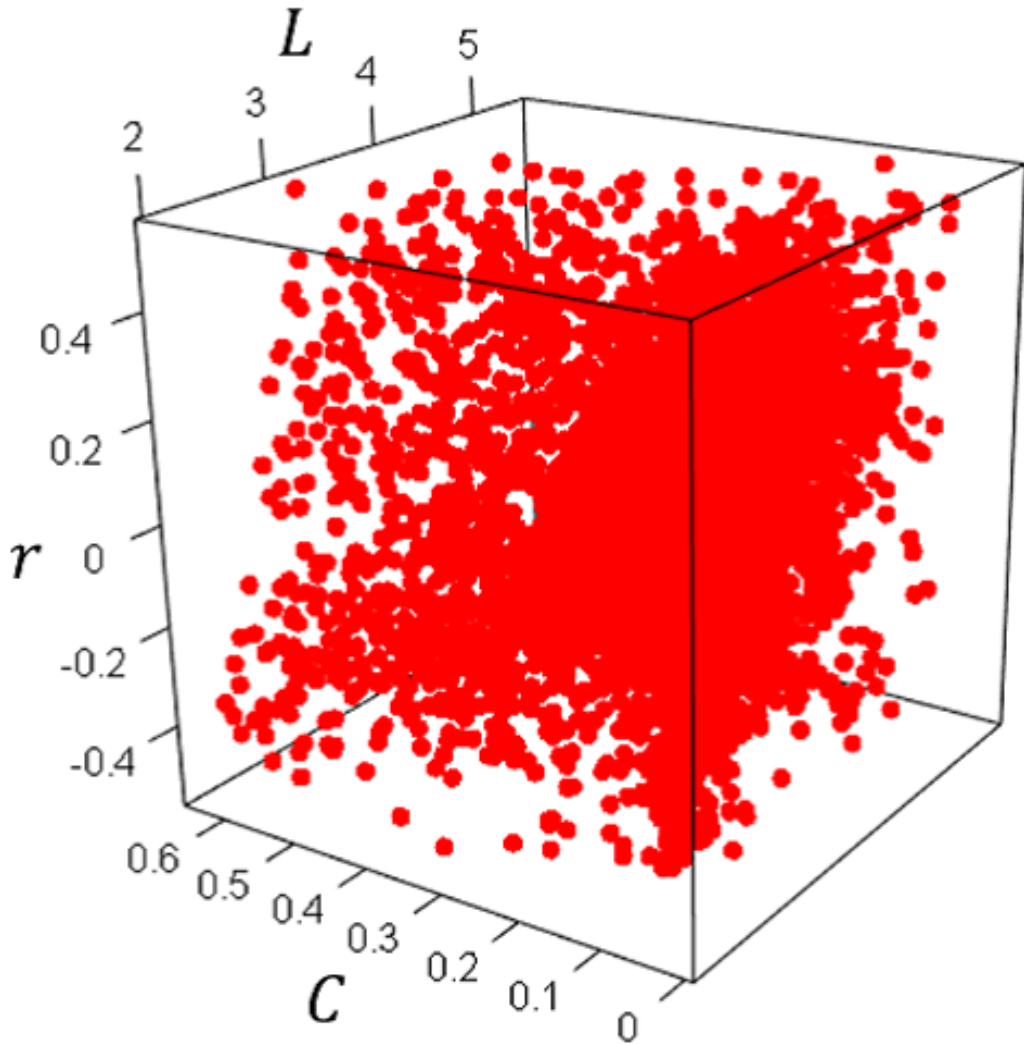


図 3.8: Expressive capacity of ANGA

た特徴点の距離が  $D \geq 0.1$  である時、そのネットワークは破棄するものとする。

本研究では、ノード数 1000、リンク数 10000 のネットワークを 10000 個生成し、生成したネットワークに関して、全ての特徴量を計算し直した。さらに、類似したネットワーク構造を除外するために、半径  $D_F = 0.1$  の多次元球内にネットワークが高々 1 つとなるように無作為抽出した。その結果として、7028 個のネットワークを含むデータセット  $\mathbb{G}_n$  を得た。図 3.8 に 3 次元特徴空間  $(L, C, r)$  への射影した図を示す。

### 3.3.2 データセットの性能評価

#### 比較検証のためのデータセット

提案手法によるデータセット  $\mathbb{G}_n$  の性能を評価するために、比較対象データセット  $\mathbb{G}_e$  を用意する。比較対象データセット  $\mathbb{G}_e$  は、既存のネットワークモデルを用いて構築する。既存モデルは以下の 10 個のネットワーク生成モデル及びネットワーク生成アルゴリズムを用いる。

- ER グラフ
- WS モデル
- BA モデル
- Fitness モデル
- CNN モデル
- HK モデル
- CNNR モデル
- FCR モデル
- XS アルゴリズム
- BS アルゴリズム

これらのモデル及びアルゴリズムを用いて、それぞれノード数 1000, 平均次数  $\bar{k}$  ( $19 \leq \bar{k} \leq 20$ ) となるようにネットワークを 1000 個ずつネットワークを生成した。この際、各モデルのパラメータは毎回ランダムに設定して生成を行う。XS モデルの初期ネットワークとして、BA モデルのネットワークを用いる。このようにして生成された 10000 個のネットワークを比較対象データセットとする。

#### 相関関係

■各特徴間の相関係数 表 3.8 に提案手法によるデータセット  $\mathbb{G}_n$  における各特徴間の相関係数を示す。表 3.8 は行と列の相関が示されている。本データセットにおける、強く相関が出ている上位 3 つを以下に示す。

1.  $\log\alpha$  と  $\log\beta$  : 0.809
2. 平均経路長  $L$  とクラスタ係数  $C$  : 0.260
3. 平均経路長  $L$  次数相関  $r$  : 0.140

表 3.8: 提案手法によるデータセットにおける相関係数

	$C$	$r$	$\log\alpha$	$\log\beta$
$L$	0.260	0.140	-0.0304	0.0775
$C$	-	-0.00720	-0.00594	0.00912
$r$	-	-	0.0903	-0.0776
$\log\alpha$	-	-	-	0.809

表 3.9: WS モデルで生成したネットワークにおける相関係数

	$C$	$r$	$\log\alpha$	$\log\beta$
$L$	0.665	0.127	0.700	0.118
$C$	-	0.169	0.957	0.510
$r$	-	-	0.166	0.0873
$\log\alpha$	-	-	-	0.658

表 3.10: HK モデルで生成したネットワークにおける相関係数

	$C$	$r$	$\log\alpha$	$\log\beta$
$L$	0.913	-0.0442	0.000	0.121
$C$	-	-0.372	0.0780	0.312
$r$	-	-	-0.248	-0.483
$\log\alpha$	-	-	-	0.646

$\log\alpha$  と  $\log\beta$  の間に強い相関が見られる。次に高い相関は平均経路長  $L$  とクラスタ係数  $C$  の間である。これらについて着目して分析を行う。

比較対象として、表 3.9 に WS モデルで生成したネットワークにおける相関係数を、表 3.10 に HK モデルで生成したネットワークにおける相関係数をそれぞれ示す。WS モデルで生成したネットワークにおける、強く相関が出ている上位 3 つを以下に示す。

1. クラスタ係数  $C$  と次数分布のパラメータ  $\log\alpha$  : 0.957
2. 平均経路長  $L$  と次数分布のパラメータ  $\log\alpha$  : 0.700
3. 平均経路長  $L$  とクラスタ係数  $C$  : 0.665



最も高い相関である、クラスタ係数  $C$  と次数分布のパラメータ  $\log\alpha$  の間の相関は非常に強く 1.0 に近い。また、それ以外の 2 つにも非常に高い相関が示されている。

次に、HK モデルで生成したネットワークにおける、強く相関が出ている上位 3 つを以下に示す。

1. 平均経路長  $L$  とクラスタ係数  $C$  : 0.913
2. 次数相関  $r$  と次数分布のパラメータ  $\log\beta$  : -0.483
3. クラスタ係数  $C$  と次数相関  $r$  : -0.372

HK モデルで生成したネットワークも、平均経路長  $L$  とクラスタ係数  $C$  に 1.0 に近い非常に強い相関が示されている。また、その他にも強い相関が示されている事がわかる。

■ $\log\alpha$  と  $\log\beta$  の間以外の相関について まず、 $\log\alpha$  と  $\log\beta$  の間の相関以外に着目する。提案手法によるデータセット  $\mathbb{G}_n$  において最も強く相関が出ているのは、平均経路長  $L$  とクラスタ係数  $C$  の間の相関であり、これ以外はほとんど相関が見られない。平均経路長  $L$  とクラスタ係数  $C$  の間の相関係数は 0.260 と弱い相関のみが認められる。この相関は、WS モデル、HK モデルで生成したネットワークと比較して、非常に小さい。従って、提案手法によるデータセットは既存ネットワーク生成モデルで生成したネットワークと比較して、各特徴が互いに相関していないデータセットであると言える。

各特徴が互いに相関しているネットワーク群を用いて分析を行ったとしても、ネットワーク構造特徴が与える影響を見る事は困難である。また、HK モデルで生成したネットワークにおける相関係数を見てみると、独立に動く特徴は一つも存在していない事が明確である。特に、HK モデルの着目特徴であるクラスタ係数  $C$  は平均経路長  $L$  や次数相関  $r$  と強く相関している。また、WS モデルで生成したネットワークにおいては、次数相関が独立であるように見えるが、これは WS モデルで生成するネットワークにおける次数相関  $r$  は常にほぼ 0 であるためである。特に、WS モデルで大きく変化する平均経路長  $L$  とクラスタ係数  $C$  は強く相関している。このように独立に動く特徴が存在しないために、このようなネットワーク群を分析に用いた場合、因果関係が曖昧になる。本データセットではこのような相関を取り除くことに成功している。

■ $\log\alpha$  と  $\log\beta$  の間の相関について 次に、 $\log\alpha$  と  $\log\beta$  の間の相関に着目する。提案手法によるデータセットにおいて、次数分布のパラメータ  $\log\alpha$  と  $\log\beta$  の間には、相関係数 0.806 の相関がある。また、WS モデルや HK モデルでも次数分布のパラメータ  $\log\alpha$  と  $\log\beta$  の間には強い相関 (WS モデル : 0.658, HK モデル : 0.646) が確認できる。従って、次数分布のパラメータ  $\log\alpha$  と  $\log\beta$  の間の相関はネットワーク構造という制約上やむを

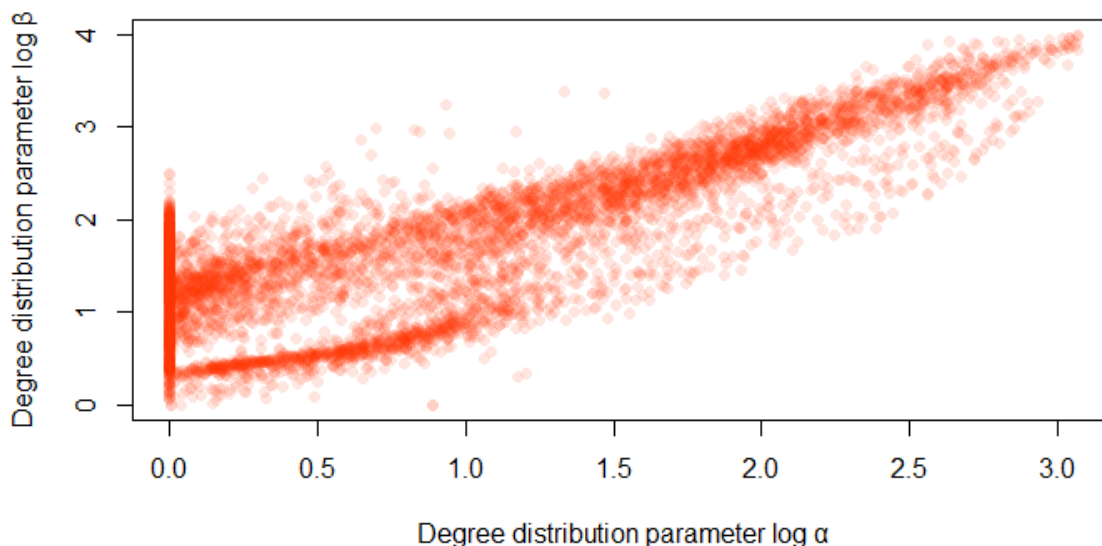


図 3.9: 提案手法によるデータセットにおける次数分布のパラメータ  $\log\alpha$  と  $\log\beta$  の関係

得ない可能性がうかがえる。

図 3.9 に提案手法によるデータセットにおける次数分布のパラメータ  $\log\alpha$  と  $\log\beta$  の関係を示す。図 3.9 は縦軸に  $\log\beta$ 、横軸に  $\log\alpha$  を示す。 $\log\alpha$  の増加と共に  $\log\beta$  も増加している事が見て取れる。逆に言えば、 $\log\alpha$  が非常に小さく  $\log\beta$  が大きいネットワーク、及び  $\log\alpha$  が大きく  $\log\beta$  が小さいネットワークが存在していない。すなわち、 $\alpha \ll \beta$  及び  $\beta < \alpha$  のネットワークが存在していない。

図 3.10 に  $\log\alpha$  及び  $\log\beta$  を変化させた時のベータ分布の形状を示す。図 3.10 において、 $\log\alpha = \log\beta$  の時、次数分布の頂点がちょうどグラフの中心に位置する。 $\log\alpha < \log\beta$  となるように  $\log\alpha$  を増加させていけば、頂点は中心より右に移動する。すなわち、グラフの頂点が平均次数  $\bar{k}$  よりも高くなる。全てのノードが  $\bar{k}$  を取れば、 $k_{max} = \bar{k}$  となるため、 $\log\alpha > \log\beta$  となるネットワークも存在する事は可能であるが、このようなネットワークを ANGA のような確率的手法によって構築する事は極めて困難である。従って、ANGA を用いて構築したデータセットには、 $\log\beta < \log\alpha$  であるネットワークが存在しない事が示唆される。この空間を少しでも減らす事は、今後の課題とする。

一方、 $\log\alpha < \log\beta$  となるように  $\log\alpha$  を減少させていけば、頂点は中心より左に移動する。 $\log\alpha$  が非常に小さい、すなわち  $\alpha \ll \beta$  の時、次数分布の山は左端にあり、次数

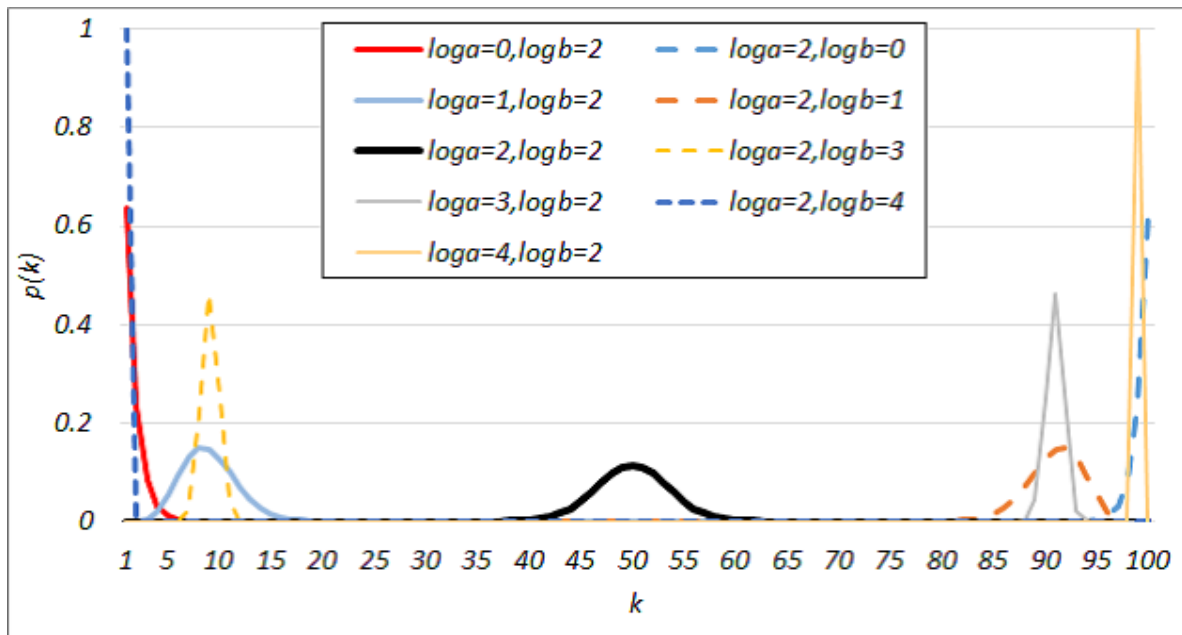


図 3.10:  $\log \alpha$  及び  $\log \beta$  の値の変化による次数分布の形状変化

分布はべき分布に近い形状を取る． $\log \beta$  が  $\log \alpha$  よりある程度大きくなった時，低次数ノードの存在確率が非常に高くなる．例えば，図 3.10 の  $\log \alpha = 0, \log \beta = 2$  の時は，次数  $k = 1$  のノードから次数  $k = 5$  までのノードの存在確率を足すと  $\sum_{k=1}^5 p(k) \simeq 1$  となる．ネットワークの平均次数は  $\bar{k} = 20$  であるから，このネットワークとしての構造を取りにくい．従って， $\alpha < \beta$  の時は取れない構造が多く存在する．

以上のように， $\log \beta < \log \alpha$  であるネットワークは ANGA では生成する事が困難であり， $\log \alpha \ll \log \beta$  であるネットワークは存在する事ができない．故に，図 3.9 における，ネットワークが存在しない領域が存在し，相関が生まれている．

### 既存ネットワークに対する包括性

ここでは，提案手法によるデータセット  $\mathbb{G}_n$  が比較対象データセット  $\mathbb{G}_e$  を包括している事を示す．すなわち，比較対象データセット  $\mathbb{G}_e$  で提供できる全てのネットワークを提案手法によるデータセット  $\mathbb{G}_n$  でも提供できる事を示す．比較対象データセット  $\mathbb{G}_e$  に含まれる各ネットワークに対して，最も小さい  $D_F$  を持つネットワークを提案手法によるデータセット  $\mathbb{G}_n$  から抽出する．比較対象データセット  $\mathbb{G}_e$  中のあるネットワーク  $g$  に対

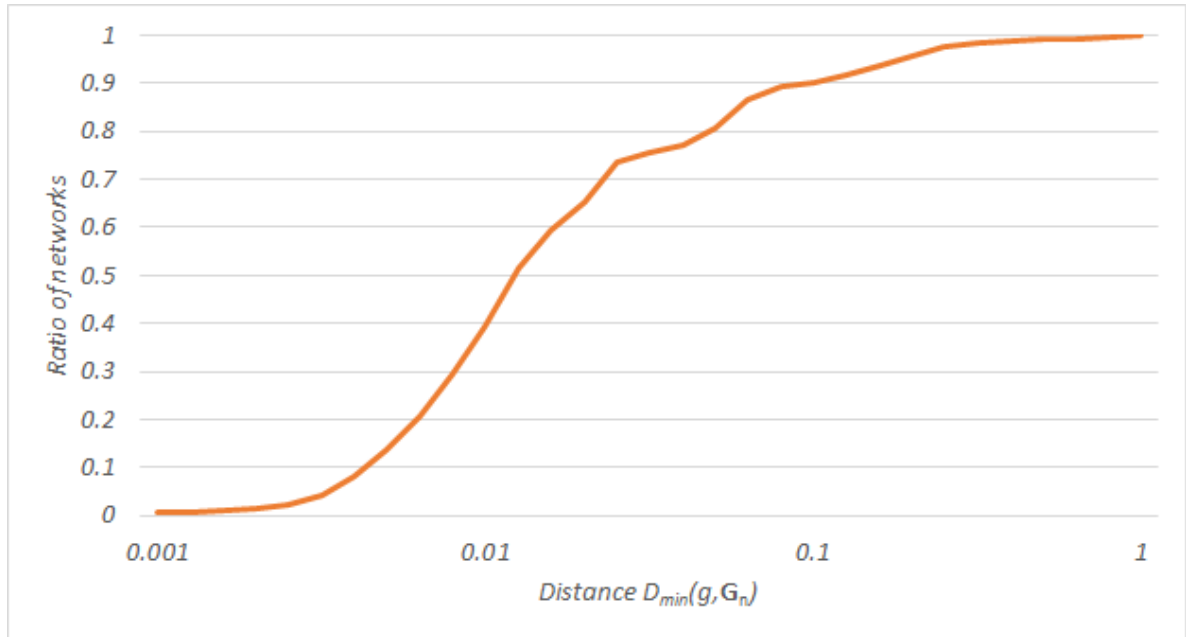


図 3.11:  $\mathbb{G}_e$  に含まれる各ネットワーク  $g$  に関する最近傍距離  $D_{min}(g, \mathbb{G}_n) | g \in \mathbb{G}_e$  の累積割合

するこの距離  $D_{\mathbf{F}}$  をネットワーク  $g$  に対する最近傍距離  $D_{min}(g, \mathbb{G}_n)$  とする.

$$D_{min}(g, \mathbb{G}_n) = \min \left\{ \frac{1}{|\mathbf{F}|} \sum_{f \in \mathbf{F}} \left( \frac{f_i - f_g}{\sigma_f} \right)^2 \mid i \in \mathbb{G}_n \right\} \quad (3.16)$$

図 3.11 に、 $\mathbb{G}_e$  に含まれる各ネットワーク  $g$  に関する最近傍距離  $D_{min}(g, \mathbb{G}_n) | g \in \mathbb{G}_e$  の累積割合を示す. 横軸は最近傍距離  $D_{min}(g, \mathbb{G}_n) | g \in \mathbb{G}_e$  であり, 縦軸は, 横軸の距離  $D_{min}(g, \mathbb{G}_n)$  の値よりも近いネットワークの割合である. 横軸の値が 0.1 である時, 90.1% のネットワークを  $D_{min}(g, \mathbb{G}_n) < 0.1$  で提供できることを意味している. なお, 提案手法によるデータセットは, 半径  $D_{\mathbf{F}} = 0.1$  の多次元球内にネットワークが高々 1 つとなるようにデータセットを構築している.  $D_{min}(g, \mathbb{G}_n) < 0.1$  であれば, いずれかのネットワークを中心とする多次元球内に所属していると言える. 従って, 提案手法によるデータセット  $\mathbb{G}_n$  内のどのネットワークの多次元球にも所属していない, 比較対象データセット  $\mathbb{G}_e$  のネットワークは全体の 9.9% である. すなわち, 比較対象データセット  $\mathbb{G}_e$  に所属するほとんどのネットワークは, 近くに提案手法によるデータセット  $\mathbb{G}_n$  に含まれるネットワークが存在すると言える. この距離は十分に小さく, ほとんどの既存モデルによるネットワークに対して類似したネットワークを提供できる事を意味している.

表 3.11: 既存モデルによるネットワークサンプルと、それに対する提案手法によるデータセット中の最近傍ネットワーク

モデル	データセット	$L$	$C$	$r$	$\log\alpha$	$\log\beta$	$D$
CNN モデル	$\mathbb{G}_e$	3.54	0.485	0.00982	0.00	1.39	$9.29 \times 10^{-5}$
	$\mathbb{G}_n$	3.54	0.483	0.00424	0.00	1.39	
FCR モデル	$\mathbb{G}_e$	2.72	0.239	0.00	0.0746	1.16	$3.24 \times 10^{-4}$
	$\mathbb{G}_n$	2.72	0.233	0.00	0.0809	1.16	

最後に、表 3.11 に  $D_{\min}(g, \mathbb{G}_n)$  が小さかった、既存モデルによるネットワークとそれに対する提案手法によるデータセット中の最近傍ネットワークのペアの各特徴量の例を 2 つ示す。表 3.11 の一つ目は、CNN モデルのネットワークに対して、ほぼ同じ特徴を持つネットワークを提供する事が出来ている。CNN モデルで生成したネットワークの少し高い平均経路長  $L$  の値や、高いクラスタ係数  $C$  値等、同等の特徴を持つネットワークが提供できる。次に、FCR モデルのネットワークに関しても、低い平均経路長  $L$  や、やや高いクラスタ係数  $C$  等、非常に類似した特徴を持つネットワークを提供する事が出来る。

以上より、比較対象のデータセット  $\mathbb{G}_e$  によって提供される多くのネットワークは、提案手法によるデータセットによって類似したネットワークを提供可能である事が示された。

### 実ネットワークに対する包括性

次に、提案手法によるネットワークデータセット  $\mathbb{G}_n$  が実世界に存在するネットワーク  $\mathbb{G}_r$  を包括している事を示す。実ネットワーク  $\mathbb{G}_r$  としては、以下の 4 つの分野のネットワークを含む 68 個のネットワークデータセットを用いる<sup>\*2</sup>。

- Social : 14 個
- Biological : 12 個
- Infrastructure : 36 個
- Information : 6 個

前節と同様に  $g \in \mathbb{G}_r$  であるネットワーク  $g$  に対して  $D_{\min}(g, \mathbb{G}_n) | g \in \mathbb{G}_r$  を求める。また、比較対象データセット  $\mathbb{G}_e$  と比較するため、 $D_{\min}(g, \mathbb{G}_e) | g \in \mathbb{G}_r$  も共に計

<sup>\*2</sup> <https://github.com/basimbaig/robust14>

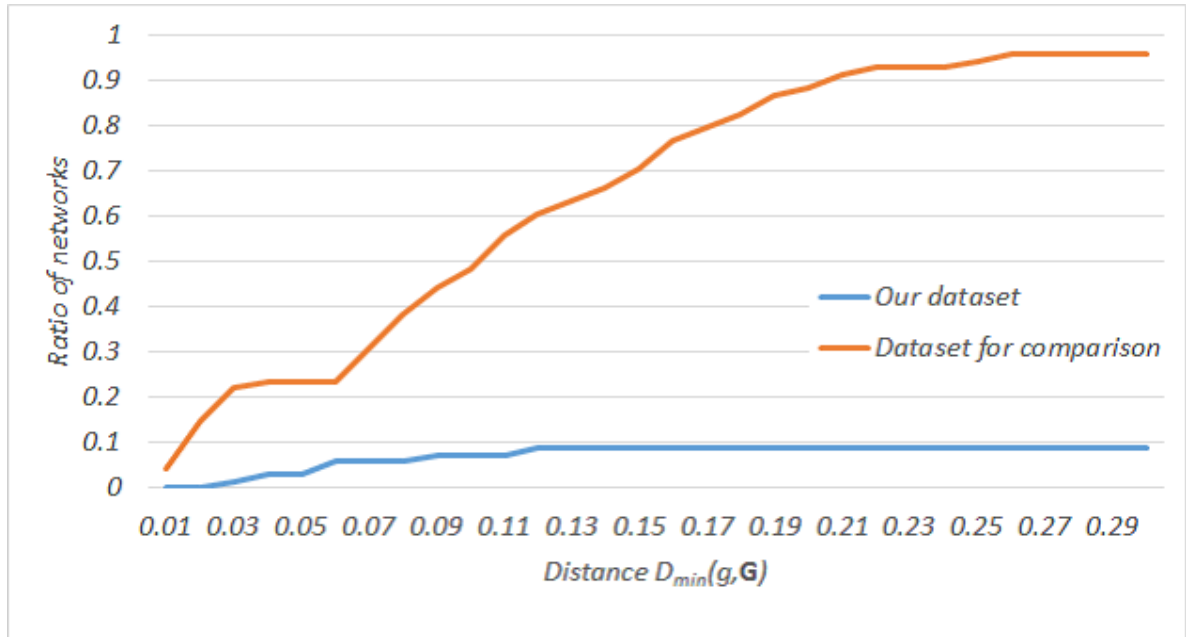


図 3.12:  $\mathbb{G}_r$  に含まれる各ネットワーク  $g$  に関する最近傍距離  $D_{min}(g, \mathbb{G}_n) | g \in \mathbb{G}_r$  及び,  $D_{min}(g, \mathbb{G}_e) | g \in \mathbb{G}_r$  の累積割合

算する. 図 3.12 に  $D_{min}(g, \mathbb{G}_n) | g \in \mathbb{G}_r$  の累積割合を示す. 横軸に  $D_{min}(g, \mathbb{G}_n)$  及び  $D_{min}(g, \mathbb{G}_e)$  を取り, 縦軸に横軸の値以下で提供できたネットワークの割合を示している. すなわち, 実ネットワーク  $\mathbb{G}_r$  に含まれるネットワークの 90% のネットワークを距離  $D_{min}(g, \mathbb{G}_n) < 0.21$  で提供できている事が示されている. 提案手法によるデータセット  $\mathbb{G}_n$  は, 比較対象データセット  $\mathbb{G}_e$  より実ネットワーク  $\mathbb{G}_r$  に近いネットワークを提供できている事が示されている. また, 68 個の実ネットワークの全てのネットワークに関して比較対象データセット  $\mathbb{G}_e$  よりも, 提案手法によるデータセット  $\mathbb{G}_n$  の方が近いネットワークを提供できている. さらに, 最近傍距離  $D_{min}(g, \mathbb{G}_n)$  及び  $D_{min}(g, \mathbb{G}_e)$  の平均値,  $D_{min}(\mathbb{G}_r, \mathbb{G}_n)$  及び  $D_{min}(\mathbb{G}_r, \mathbb{G}_e)$  を求める.

$$D_{min}(\mathbb{G}_r, \mathbb{G}_n) = \frac{1}{|\mathbb{G}_r|} \sum_{g \in \mathbb{G}_r} D_{min}(g, \mathbb{G}_n) \quad (3.17)$$

$$D_{min}(\mathbb{G}_r, \mathbb{G}_e) = \frac{1}{|\mathbb{G}_r|} \sum_{g \in \mathbb{G}_r} D_{min}(g, \mathbb{G}_e) \quad (3.18)$$

表 3.12 にそれぞれの平均値を示す. 平均値に対しても,  $D_{min}(\mathbb{G}_r, \mathbb{G}_e) = 0.820$  に対して,  $D_{min}(\mathbb{G}_r, \mathbb{G}_n) = 0.262$  は小さい値を取っている. 以上の結果から, 提案手法による

表 3.12: 最近傍距離  $D_{min}(g, \mathbb{G}_n)$  及び  $D_{min}(g, \mathbb{G}_e)$  の平均値

$D_{min}(\mathbb{G}_r, \mathbb{G}_n)$	$D_{min}(\mathbb{G}_r, \mathbb{G}_e)$
0.262	0.820

表 3.13: 比較対象ネットワークサンプルと，それに対する提案手法によるデータセット中の最近傍ネットワーク

サンプル ID	データセット	$L$	$C$	$r$	$\log\alpha$	$\log\beta$	$D$
サンプル 1	$\mathbb{G}_r$	4.64	0.0109	0.0355	0.00	1.41	
	$\mathbb{G}_n$	4.67	0.0274	0.0866	0.00	1.35	0.008304
	$\mathbb{G}_e$	4.21	0.121	0.304	0.997	2.06	0.688
サンプル 2	$\mathbb{G}_r$	3.04	0.421	0.112	0.201	1.41	
	$\mathbb{G}_n$	2.94	0.445	0.135	0.113	1.40	0.00995
	$\mathbb{G}_e$	3.21	0.413	0.0790	0.00	1.26	0.0289

データセット  $\mathbb{G}_n$  は，比較対象データセット  $\mathbb{G}_e$  より実ネットワーク  $\mathbb{G}_r$  に近いネットワークを提供できていると言える。

最後に，表 3.13 にネットワークの提供例を示す．サンプル 1 では，全ての特徴において，比較対象データセット  $\mathbb{G}_e$  のネットワークよりも，提案手法によるデータセット  $\mathbb{G}_n$  のネットワークの方が実ネットワークに近い特徴を示している．また，サンプル 2 では，クラスタ係数はわずかに比較対象データセット  $\mathbb{G}_e$  のネットワークの方が，提案手法によるデータセット  $\mathbb{G}_n$  のネットワークよりも実ネットワークのクラスタ係数  $C$  に近いものの，その他の特徴は比較対象データセット  $\mathbb{G}_e$  のネットワークよりも，提案手法によるデータセット  $\mathbb{G}_n$  のネットワークの方が実ネットワークに近い特徴を持っている。

### ネットワーク構造特徴空間に対する充填率

最後に，ネットワーク構造特徴空間に対する充填率を求める．この充填率は，ネットワーク構造特徴空間をどの程度満たしているかを表す指標として定義し，ネットワーク構造特徴空間をより埋めている方が，あらゆる構造のネットワークを含んでいる事を意味する．あらゆる構造のネットワークを含んでいるデータセットを構築することが出来れば，複雑ネットワーク上の現象の分析に寄与できると考えられる。

まず，各特徴領域 ( $2 \leq L \leq 6$  等) を 10 分割する．すなわち，5 次元特徴空間を  $10^5$  個

表 3.14: The filling ratio of each dataset.

	$a(\mathbb{G}_n)$	$a(\mathbb{G}_e)$
The filling ratio	$3.40 \times 10^{-2}$	$3.38 \times 10^{-3}$

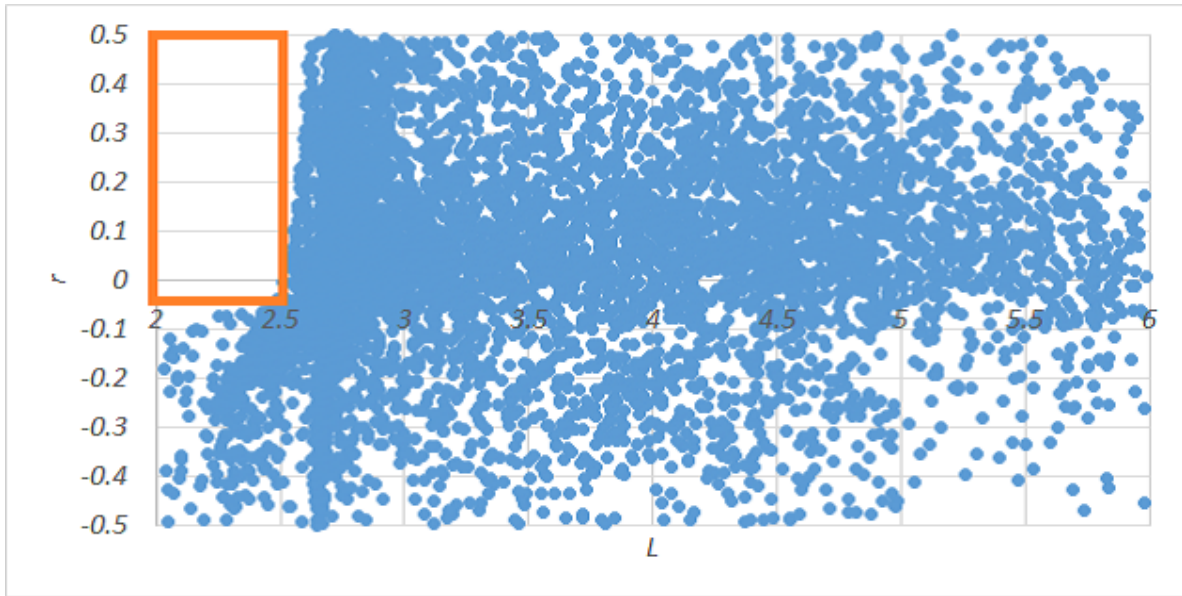


図 3.13: 提案手法によるデータセットにおける平均経路長  $L$  と次数相関  $r$  の間の関係

の多次元立方体に分割する．この多次元立方体の中で，提案手法によるデータセット  $\mathbb{G}_n$  に含まれるネットワークを含んでいるものの割合を充填率  $a(\mathbb{G}_n)$  として定義する．この時，同時に比較対象データセットに対する充填率  $a(\mathbb{G}_e)$  も求める．

**表 3.14** にそれぞれ， $a(\mathbb{G}_n)$  及び  $a(\mathbb{G}_e)$  を示す．これより， $a(\mathbb{G}_n)$  は  $a(\mathbb{G}_e)$  の 10 倍以上大きい値であることがわかる．従って，提案手法によるデータセット  $\mathbb{G}_n$  は比較対象データセット  $\mathbb{G}_e$  よりも充填率が高い．また，提案手法に対する充填率  $a(\mathbb{G}_n)$  の最大値は，全てのネットワークが別々の多次元立方体に含まれた時であり，高々  $7028 \div 10^5 = 7.03 \times 10^{-2}$  である．本データセットの  $a(\mathbb{G}_n) = 3.40 \times 10^{-2}$  は最大値の 48.3% であり，最大値に対してもかなり高い値となっている事がうかがえる．従って，提案手法によるデータセット  $\mathbb{G}_n$  は様々な構造のネットワークを含んでいる事が示された．

しかし，提案手法によるデータセット  $\mathbb{G}_n$  においても，ネットワークが存在しない領域が存在する．例えば，**図 3.13** に提案手法によるデータセット  $\mathbb{G}_n$  における平均経路長  $L$  と次数相関  $r$  の間の関係を示す．**図 3.13** における点は提案手法によるデータセット  $\mathbb{G}_n$



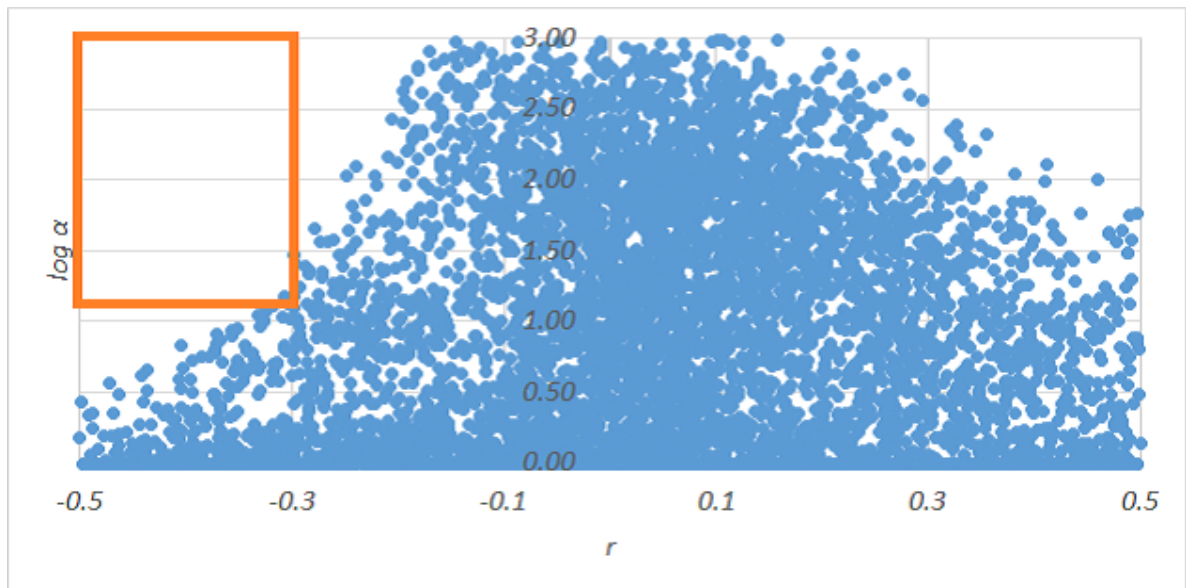


図 3.14: 提案手法によるデータセットにおける次数相関  $r$  と次数分布のパラメータ  $\log \alpha$  の間の関係

の各ネットワークである．このオレンジの四角の中にはネットワークが存在しない．これは，ANGA がこの領域にネットワークを生成できない事を示している．しかし，この領域がネットワーク構造として取り得る領域かどうかは数学的に証明する必要があると考える．

また，図 3.14 に提案手法によるデータセットにおける次数相関  $r$  と次数分布のパラメータ  $\log \alpha$  の間の関係を示す．同様に点は提案手法によるデータセット  $\mathbb{G}_n$  の各ネットワークであり，オレンジの四角の中にはネットワークが存在しない．次数分布のパラメータ  $\log \alpha$  が大きい時，各ノードの次数の偏りがなくなる．このようなネットワークでは，次数相関  $r$  が負になる事はあり得ないので，この領域にはネットワークが存在しえない．

ところで，平均経路長  $L$ ，クラスタ係数  $C$ ，次数相関  $r$  の 3 次元の特徴空間に対する充填率は 0.768 であり，非常に高い値である．従って，平均経路長  $L$ ，クラスタ係数  $C$ ，次数相関  $r$  の 3 次元特徴空間ではかなりの領域を埋める事が出来ている．逆に言えば，次数分布のパラメータがネットワークの構造を大きく制約している事がうかがえる．従って，次数分布に関して，より任意の次数分布を持つネットワークを生成する事を今後の課題として挙げる．

## 3.4 結論

本章では、ネットワーク構造特徴空間を埋めるようなネットワークデータセットを構築を行った。このネットワークデータセットを構築するために、様々な構造のネットワークを集める必要がある。しかし、第2章より、既存のネットワークモデルを用いても、ネットワーク構造特徴空間を埋める事は出来ない。そこで、本研究では様々な構造のネットワークを生成可能な手法として **Arbitrary Network Generation Algorithm(ANGA)** を提案した。

**ANGA** は、ネットワーク構造特徴からネットワークを生成する手法である。このネットワーク構造特徴として、ネットワーク構造特徴空間上の任意の特徴点を与える事によって、ネットワーク構造特徴空間を網羅するようなネットワークデータセットを構築する事が可能である。**ANGA** は以下の2つの手法を用いて提案される。

- ネットワーク成長モデル
- 貪欲法

この2つの手法を用いる事で、ネットワークを指定した任意の特徴点に向かって成長させる。

**ANGA** が任意の特徴点を生成可能かどうかの検証を行った。実験には実ネットワークによる検証と、比較対象ネットワークによる検証の2つを行った。その結果、実ネットワークから抽出した特徴点も、比較対象ネットワークから抽出した特徴点も、共に高精度で生成できる事を示した。

本研究で提案した **ANGA** は拡張性の高い手法であり、ノードの選択手法や特徴間距離を変更する事によって、拡張する事が可能である。現在の **ANGA** では、より複雑な条件でのネットワーク生成をする事は困難である。例えば、特徴にモジュール化の度合いを表すモジュラリティを採用したとしても、様々なモジュラリティを持つネットワークを構築する事は難しい。様々なモジュラリティを持つネットワークを構築するためには、ノードの選択手法を検討し、モジュール構造を変化させる事が出来るようなノードの選択手法を採用する必要がある。また、本研究では、次数分布を表す特徴として、ベータ分布を用いた。これによって、べき分布を利用するよりも、様々な構造のネットワークを生成する事が可能となっている。しかし、ベータ分布で表現できない次数分布も様々な存在する。この利用する分布を検討する事によって、より広い範囲で次数分布を表現する事が可能であると考えられる。**ANGA** では、このように生成できないネットワークが様々な存在する。

より様々なネットワークを構築するためには、ANGA を拡張していく必要がある。

続いて、ANGA を用いて、ネットワークデータセットの構築を行った。ランダムに設定した特徴点に ANGA を用いてネットワークを生成する事によって、ネットワーク構造特徴空間を埋めるようなネットワークデータセットを構築した。本データセットが様々なネットワークを含んでいるかどうかを検証するために以下の 4 つの検証方法で検証を行った。

- 相関係数
- 既存ネットワークに対する包括性
- 実ネットワークに対する包括性
- ネットワーク構造特徴空間に対する充填率

この検証の結果、提案手法によるデータセットは、既存モデルによって構築したデータセットよりも、遥かに様々な構造を含み、偏りのないデータセットとなっている事を示した。

本データセットを用いる事によって、ネットワーク構造と現象の正確な関係が明らかになる事が期待される。

## 第4章

# 情報拡散の統計的分析

### 4.1 はじめに

本章では、ネットワーク上で起こる現象の例として、情報拡散を取り上げる。ここでの情報拡散は、情報が個人間の相互作用によって、個人から個人へと伝播されていくプロセスとする [32]。ある個人が得た情報は個人から他者へと受け渡され、全体へと拡散されていく。この時、全く知らない誰かに情報を伝搬させるとは考えにくく、相互に関係する他者へと情報を伝える事になる。すなわち情報の拡散は、個人が形成する人間関係のネットワーク上で行われる。

近年では、ソーシャルメディアの発展に伴い、人間関係ネットワークが目に見える形で具現化され、ネットワーク上での情報の伝達が明確になった。ソーシャルメディア上の多くのユーザは、ソーシャルメディアによって情報収集・発信をしているため、このソーシャルメディア上の情報拡散が注目を集めている。この情報収集・発信は、個人的な趣味趣向に関する情報だけに留まらず、政治や経済といった情報にまで至る。ソーシャルメディア上の情報拡散は、マーケティング [76]、政治的意見形成 \*<sup>1</sup>等、様々な影響を現実の社会現象に与えている。

ソーシャルメディアによる情報拡散は、災害のような非常事態時での活躍が期待されている。一般に、非常事態時には、電話等が使用不可能となり、被災地との連絡が取れない事態が起こると想定されている。その際、ソーシャルメディアを介する事でインターネット回線に大きな負担をかけることなく情報拡散が可能であると考えられている。実際に、2010年の the 2010 Chile earthquake [77] や 2011年の the Great East Japan Earthquake [78]

---

\*<sup>1</sup> 出典：クチコミ@参院選 2010 Twitter 分析レポート

では、Twitter を使った情報拡散が有用であった事が報告されている。

情報拡散の拡散率を左右する最も大きな要因としてネットワークの構造が挙げられる [17]。ネットワークの構造によって、情報が大きく拡散される事も、拡散されない事も考えられる。このネットワーク構造と情報拡散の関係については、多くの研究者が分析を重ねている [17, 18, 19]。しかし、これらの分析で用いられているネットワークは第 2 章で示したように、ネットワーク全体の構造からみれば一部のネットワークでしかない。このように一部のネットワークのみを用いても統計的に正しい結果は得られない。

これに対して、本研究では第 3 章で提案した手法によるデータセットを用いる事によって、一部の構造のネットワークではなく、様々な構造を持つネットワークを用いた分析を行う。本章では、提案手法によるデータセットに含まれる各ネットワークの上で情報拡散モデルによる情報拡散シミュレーションを行う。これにより、各ネットワークの情報拡散能力を測定し、情報拡散能力とネットワークの構造特徴との関係について分析を行う。この分析により、情報拡散とネットワーク構造の関係を明確にする。

## 4.2 情報拡散モデル

情報拡散モデルとは、情報の発生源からエージェントの間を情報が伝搬されていくプロセスをモデル化したものである。情報は、エージェント同士の相互作用によって伝搬し、エージェントはそれぞれ個別に情報を伝えるか受けとるかを判断する。多くの場合、エージェントは自分の周囲の局所的情報のみによって情報の送受信を行う。これは一つの複雑システムであり、相互作用の連鎖によって一つの現象が生み出されている。複雑システムでの現象はモデルを立てて分析される事が多い。

情報拡散をモデル化した情報拡散モデルは、感染モデルを応用して定義される事が多い。すなわち、情報を感染症と捉える事によって、情報を取得した状態を感染状態として取り扱ったものである。例えば、SIR(Susceptible-Infected-Recovered) モデル [54] や SIS(Susceptible-Infected-Susceptible) モデル [?] 等の感染モデルを用いて分析される事が多い。本研究ではその中から、対となる 2 つの情報拡散モデルを用いる。一方は、発信者主体の独立カスケードモデル (Independent Cascade model:IC model) であり、他方は受信者主体の線形閾値モデル (Linear Threshold model:LT model) である [79]。それぞれについて以下で解説する。

### 4.2.1 独立カスケードモデル (Independent Cascade model:IC model)

IC model は情報の送り手を主体とした情報拡散モデルである。IC model において、リンクで結ばれたノード  $v, u$  は情報拡散率  $s_{vu} (0 \leq s_{vu} \leq 1)$  をもつ。

ある時刻  $t$  において、アクティブになったノード  $v$  は、リンクで結ばれた非アクティブなノード  $u$  をアクティブにする機会を一度だけ与えられる。すなわち、時刻  $t$  においてアクティブになったノード  $v$  は時刻  $t+1$  に一度だけ情報の発信を行う。その際、ノード  $v$  が情報を発信する確率、つまり、ノード  $v$  の隣接ノード  $u$  がアクティブになる確率は  $s_{vu}$  となる。その実行が成功した場合、ノード  $u$  は時刻  $t+1$  においてアクティブとなる。この情報拡散過程は、アクティブなノードが増加しなくなった時点で終了する。

### 4.2.2 線形閾値モデル (Linear Threshold model:LT model)

LT モデルは情報の受け手を主体とした情報拡散モデルである。LT モデルにおいて、リンクで結ばれたノード  $v, u$  は影響力  $w_{vu}$  をもつ。 $w_{vu}$  はノード  $v$  の隣接ノード集合  $B(v)$  に対して以下の式を満たす。

$$\sum_{u \in B(v)} w_{uv} \leq 1 \quad (4.1)$$

すなわち、ノード  $v$  に対する隣接ノードからの影響の総和は 1 を超えない。また、各ノード  $v$  は閾値  $\theta_v$  を持つ。

各時刻  $t$  において、非アクティブノード  $v$  は、以下の式が満たされた時、アクティブとなる。

$$\sum_{u \in B_t(v)} w_{uv} \geq \theta_v \quad (4.2)$$

なお、 $B_t(v)$  は時刻  $t$  において、ノード  $v$  の隣接ノードの内、アクティブなノードの集合である。すなわち、隣接するアクティブノードからの影響の総和が  $\theta_v$  を超えた時、ノード  $v$  はアクティブとなる。これを、新しくアクティブになるノードがいなくなるまで続ける。

### 4.2.3 ネットワークの拡散率 $\delta$

各モデルの  $t = 0$  において，初期アクティブノード  $v_s$  を用意する．この時，収束後のアクティブノードの割合を  $\delta(v_s)$  とし，拡散率  $\delta$  を以下のように定義する．

$$\delta = \frac{1}{N} \sum \delta(v_s) \quad (4.3)$$

なお， $N$  はネットワークのノード数とする．各ネットワークで 10 回拡散率  $\delta$  を算出し，平均をそれぞれのネットワークのモデルにおける拡散率  $\delta$  とする．

本章では，まず IC モデルによって算出した拡散率を用いて分析を行い，その後，IC モデルの拡散率と LT モデルの拡散率の比較を行う．

## 4.3 拡散現象の統計的分析

### 4.3.1 既存研究の検証

ここでは，まず既存のネットワーク構造と情報拡散との関係に関する知見について，本データセットにおける分析においても同様の事が言えるかどうかを検証する．まず，クラスタ係数についての知見に着目し，次に次数相関についての知見に着目する．それぞれ，既存の研究と一致する事を検証し，一致しなければ，なぜ矛盾が起きたのかについて考察を行う．

#### クラスタ係数に関する既存研究との比較

まず，クラスタ係数に関する知見と比較を行う．Yinzuo ら [57] は，SIR モデルを用いて，クラスタ係数が拡散現象に与える影響について分析している．彼らは，第 2.4.1 節で解説した HK モデルを拡張し，動的な要素を取り入れた．HK モデルには制御パラメータ  $q$  が存在し， $q$  を大きくするとトライアドフォーメーションをする確率が高くなり，クラスタ係数の増加に直結する．彼らは，HK モデルのパラメータ  $q$  を変動させる事によって，様々なクラスタ係数の値を持つネットワークを生成し，その上で情報拡散の分析を行った．この時，情報拡散の最中にリンクが確率  $p$  によって張り替えられるようにした．**図 4.1** に Yinzuo らの論文から引用したクラスタ係数  $C$  と拡散率  $\delta$  の関係を示す．**図 4.1** は赤い丸がシミュレーション結果であり，黒い実線が理論値となっている．これにより， $p$  の値に関わらずクラスタ係数  $C$  の増加が拡散率の低下につながる事を示している．なお，本結果において  $p = 0$  の時は，HK モデルを用いた静的な結果となる．

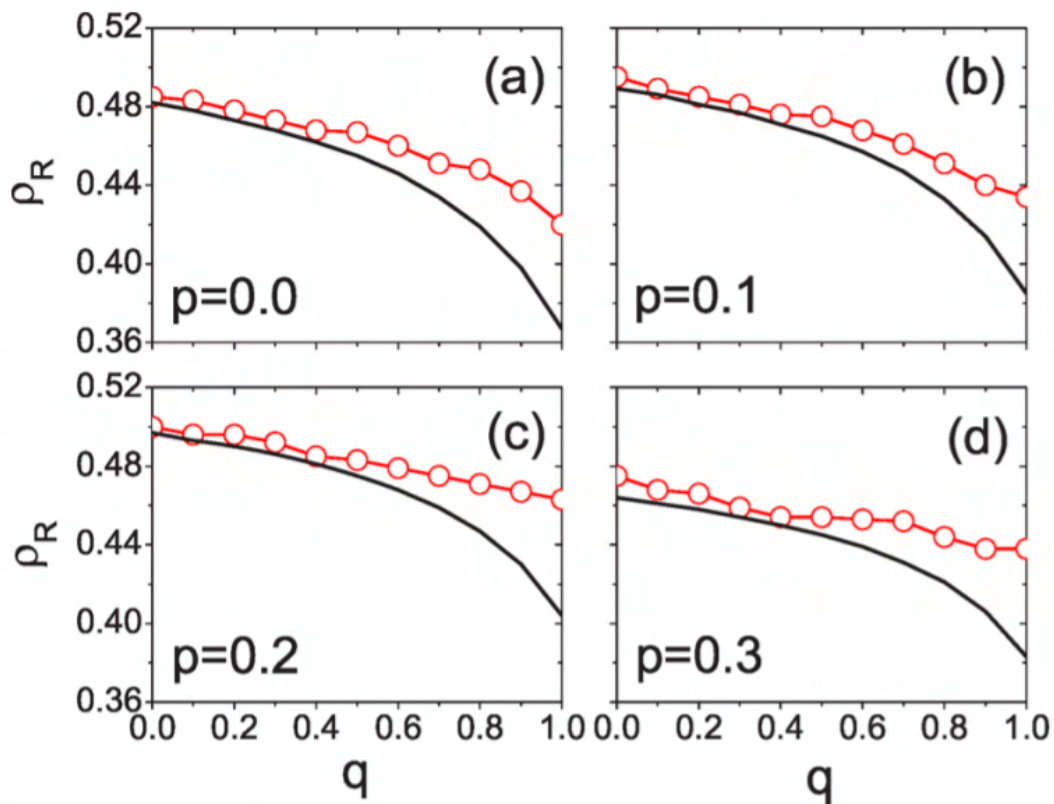


図 4.1: Yinzuo らが示したクラスタ係数  $C$  と拡散率  $\delta$  の関係 (赤：シミュレーション結果, 黒：理論値) [57]

本研究での結果を分析し、彼らの主張と比較する。図 4.2 に、本データセットにおける、拡散率  $\delta$  とクラスタ係数  $C$  の関係を示す。縦軸に拡散率  $\delta$ 、横軸にクラスタ係数  $C$  をとり、各点はデータセットに含まれるネットワークである。クラスタ係数  $C$  が低い構造の時に、多くのネットワークにおいて、高い拡散率  $\delta$  を獲得している事がみられる。この相関係数は  $-0.511$  と非常に強い負の相関が示されている。すなわち、クラスタ係数が低い程、ネットワークの拡散率は高い事が示された。これは、Yinzuo らの知見と一致する。

#### 次数相関 $r$ に関する既存研究との比較

次に、次数相関  $r$  に関する知見との比較を行う。Badham ら [58] は次数相関  $r$  が低いネットワークの方が拡散率が高い事を示している。Badham らは BS アルゴリズムを用いる事によって、様々なクラスタ係数  $C$  と次数相関  $r$  を持つネットワークを生成し、その



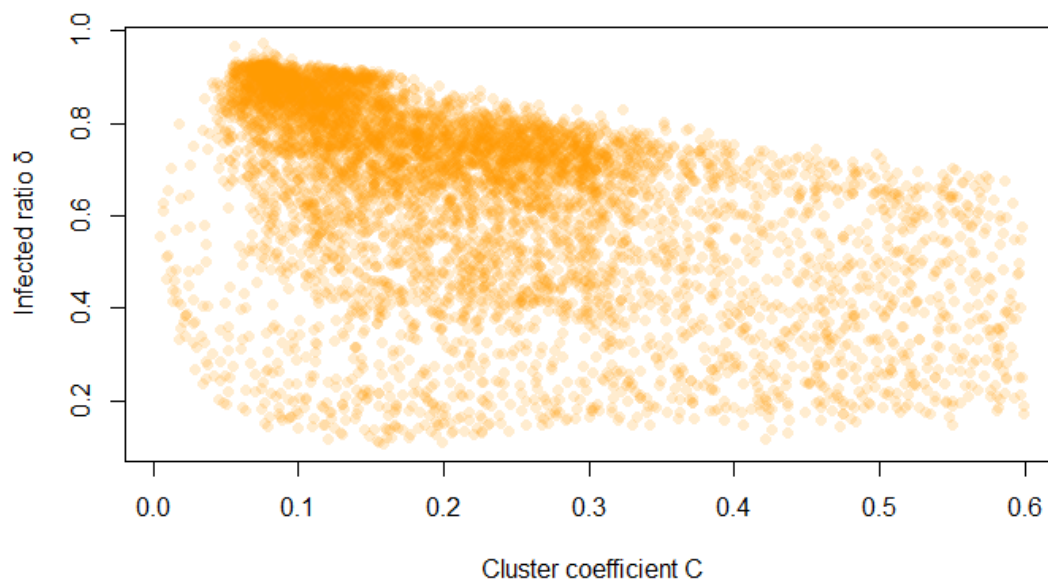


図 4.2: Infected degree vs Cluster coefficient

$r$	0.0	0.1	0.2	0.3	0.4
0.0	0.80	0.79	0.79	0.77	0.73
0.1	0.79	0.79	0.78	0.76	0.71
0.2		0.76	0.74	0.68	0.69
0.3		0.73	0.70	0.64	0.59

図 4.3: Badham らが示したクラスタ係数  $C$ ，次数相関  $r$  と拡散率  $\delta$  の関係 [58]

上で実験を行った。図 4.3 に Badham らの論文から引用した，クラスタ係数  $C$ ，次数相関  $r$  と拡散率  $\delta$  の関係 [58] を示す。次数相関  $r$  を列，クラスタ係数  $C$  を行としている。この結果より，次数相関  $r$  が低く，かつクラスタ係数  $C$  が低い方が拡散率が高い事を示している。クラスタ係数  $C$  に関しては，本研究の分析と一致している事が確認できる。

図 5.2 に拡散率  $\delta$  と次数相関  $r$  の関係を示す。縦軸に拡散率  $\delta$ ，横軸に次数相関  $r$  をとり，各点はデータセットに含まれるネットワークである。次数相関  $r$  が高い時も低い時も同様に，拡散率  $\delta$  が高いネットワークと低いネットワークが共存している事がみられる。

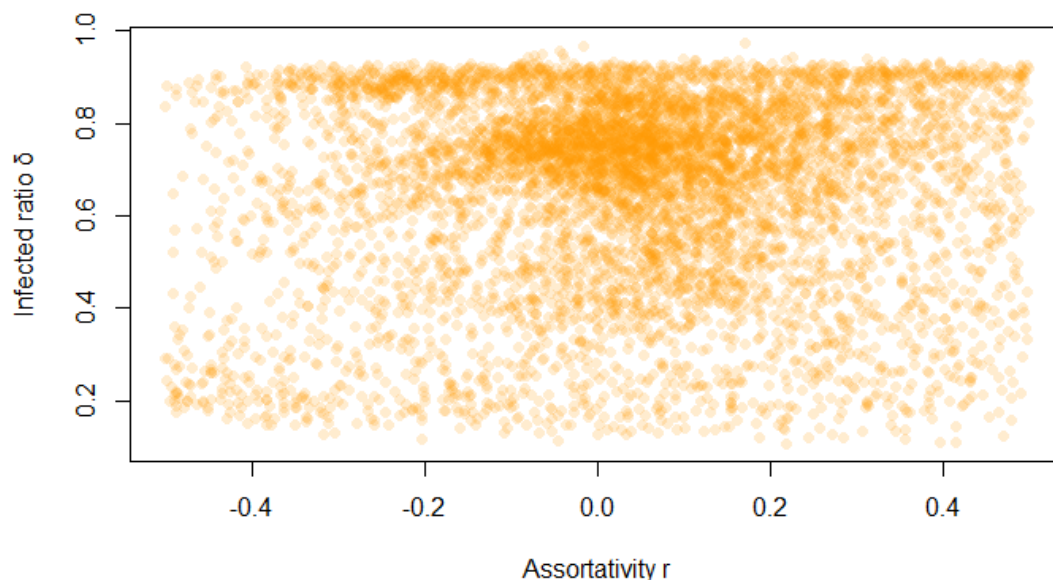


図 4.4: Infected degree vs Assortativity

この相関係数は 0.0650 と非常に低く、相関は全くない事がわかる。この結果は、Badham らの主張と矛盾している。

なぜこのような違いが生じたのかについて考察する。彼らが用いた BS アルゴリズムでは、次数相関と平均経路長が相関係数 0.723 で相関している。すなわち、次数相関がない時に平均経路長が最も短くなる。ところで、平均経路長と拡散率  $\delta$  が相関するのは自明である。経路が長ければその分ネットワークに存在する全てのノードに情報が届く確率は低くなるためである。すなわち、平均経路長が短い程拡散率が高くなる。Badham らの主張は次数相関がない時の方が拡散率が高いという知見であるが、次数相関がない時は平均経路長が短いため拡散率が高いと考えれば、我々の知見と合致する。

Badham らは分析する際に平均経路長の値については言及していない。従って、彼らの主張する、拡散率  $\delta$  と次数相関  $r$  の関係は偽相関である可能性が高い。ここで、Badham らが用いた BS アルゴリズムの各特微量と拡散率  $\delta$  との間の相関係数を表 4.1 に示す。ここから、平均経路長  $L$ 、クラスタ係数  $C$ 、次数相関  $r$  と拡散率  $\delta$  に高い負の相関が得られた。拡散率  $\delta$  と次数相関  $r$  の関係から平均経路長  $L$  の関係を取り除いた偏相関係数  $r_{r\delta \cdot L}$

表 4.1: BS モデルの各特微量と拡散率  $\delta$  との間の相関係数

	$L$	$C$	$r$	$\log\alpha$	$\log\beta$
$P_c(1.2)$	-0.800	-0.735	-0.634	0.100	0.243

を求める.

$$r_{r\delta \cdot L} = \frac{r_{r\delta} - r_{r\delta}r_{rL}}{\sqrt{1 - r_{r\delta}^2}\sqrt{1 - r_{rL}^2}} \quad (4.4)$$

偏相関係数は,  $r_{C\delta \cdot L} = -0.119$  となり, 次数相関  $r$  と拡散率  $\delta$  の間はわずかにしか相関がない事がわかる. 従って, この相関は偽相関である事が示された.

### 4.3.2 拡散率 $\delta$ に影響を与える特徴の分析

#### 拡散率 $\delta$ と各特徴間の相関係数

ここでは一つ一つの構造特徴と拡散率  $\delta$  の関係を見るために, 今回採用した 5 つの特徴と拡散率  $\delta$  の相関を分析する.

表 4.2 に, 各特徴と拡散率  $\delta$  との相関係数を示す. 表 4.2 より, 平均経路長  $L$ , クラスタ係数  $C$  に負の相関があり, 次数分布のパラメータ  $\log\alpha$  に正の相関がある事が明確となった. また, 次数分布のパラメータ  $\log\beta$  と次数相関  $r$  は大きな相関がない事がわかる. 以降では, 相関のあった特徴の内まだ分析していない平均経路長  $L$  及び次数分布のパラメータ  $\log\alpha$  についてそれぞれ見ていく事とする.

#### 拡散率 $\delta$ と平均経路長 $L$ の関係についての分析

図 4.5 に平均経路長  $L$  と拡散率  $\delta$  の関係を示す. 縦軸に拡散率  $\delta$ , 横軸に平均経路長  $L$  をとり, 各点はデータセットに含まれるネットワークである. まず, このグラフは右肩下がりになっている事から, 平均経路長  $L$  が高い程, 拡散率  $\delta$  は小さい傾向にある事がわかる. この相関係数は  $-0.502$  である.

次にグラフの中央付近から右側にほとんど点が存在していない領域がある事に着目する. この領域の上下で 2 つのグループに分かれるように直線を引く. この直線は, 直線

表 4.2: IC model の拡散率  $\delta$  と各特長の相関係数

$L$	$C$	$r$	$\log\alpha$	$\log\beta$
-0.502	-0.511	0.0650	0.398	-0.0637

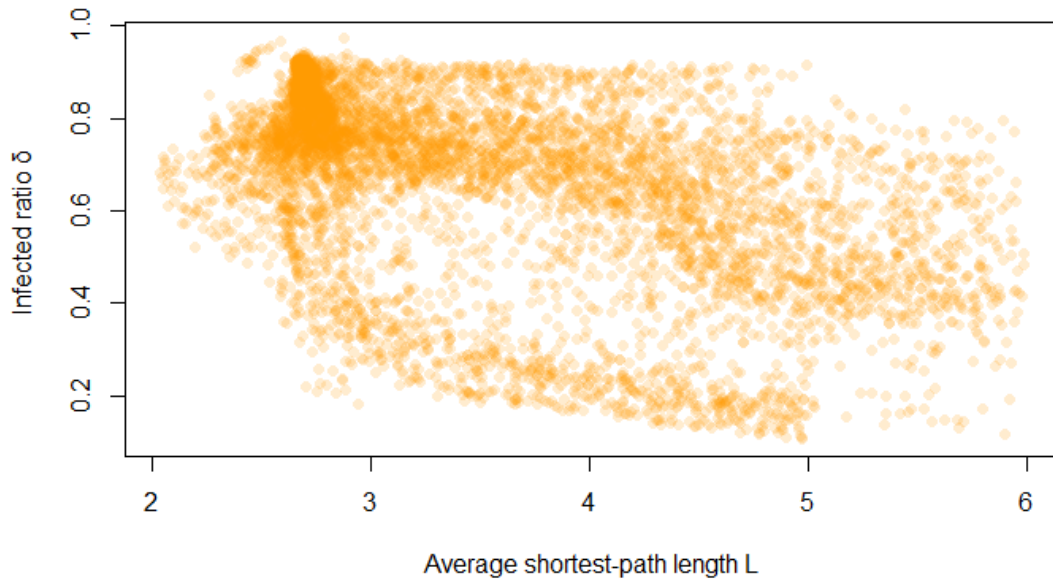


図 4.5: 平均経路長  $L$  と拡散率  $\delta$  の関係

からの正規化距離が 0.05 より小さい点の数が最も少なくなるように引いた。図 4.6 に図 4.5 を 2 つのグループに分けた図を示す。グラフの下部、拡散率の低いグループ 2 では、平均経路長  $L$  と拡散率  $\delta$  は線形に相関しており、相関係数は  $-0.806$  である。この 2 つのグループに対して決定木を構築し、図 4.7 に示す。この決定木は、それぞれの条件におけるグループ 1 とグループ 2 のネットワークの割合をそれぞれ示している。これより、 $\log\alpha \simeq 0$  かつ  $\log\beta > 1.18$  であるネットワークの 87.5% がグループ 2 に所属している事がわかる。図 4.8 に、図 4.5 を  $\log\alpha \simeq 0$  かつ  $\log\beta > 1.18$  であるネットワークを青で、それ以外を赤で色分けした図を示す。図 4.8 より、青い点はグループ 2 に所属するか、グループ 2 とグループ 1 の境界付近にプロットされている事がわかる。ところで、 $\log\alpha$  が非常に小さく  $\log\beta$  が大きい時、次数分布はべき分布に近い形状を取る。従って、次数分布がべき分布になる時、ネットワークはグループ 2 に所属し、拡散率  $\delta$  は小さくなる。

最後に、図 4.6 におけるグループ 1 に着目する。グループ 1 では、平均経路長が  $L = 4.4$  を超えた辺りで拡散率  $\delta$  が急激に下がっているのがわかる。すなわち、 $L > 4.4$  を閾値として情報が急激に広まらなくなっていく事が示されている。具体的な数値に関しては、第 4.3.4 節にて分析する。

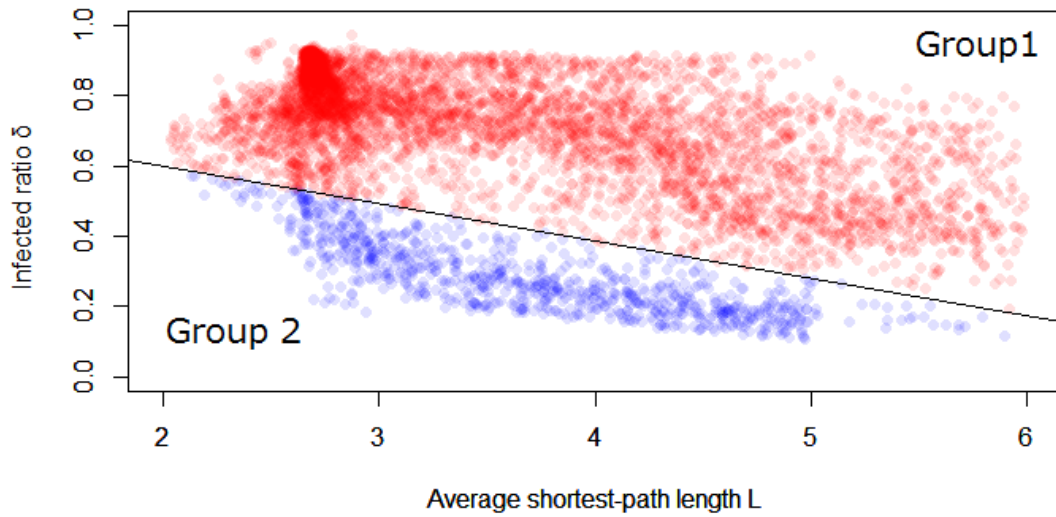


図 4.6: 平均経路長  $L$  と拡散率  $\delta$  の関係によるグループ分け

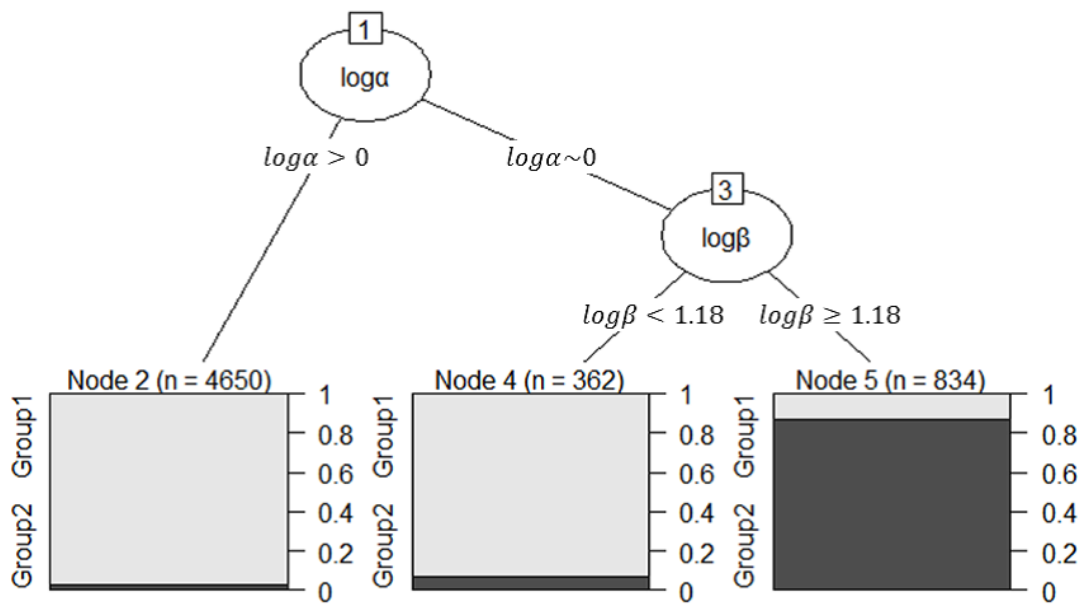


図 4.7: グループに関する決定木 (平均経路長  $L$  に関して)

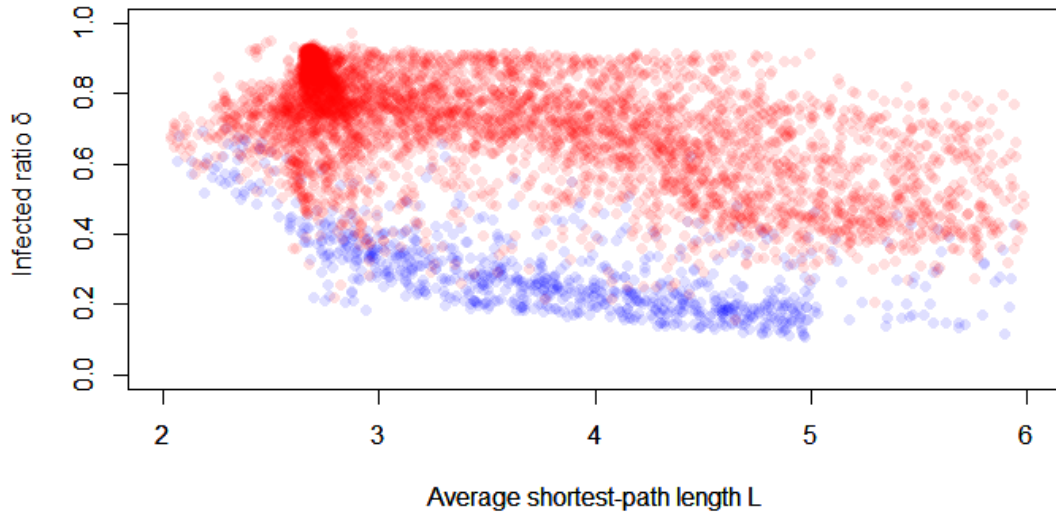


図 4.8: 平均経路長  $L$  と拡散率  $\delta$  の関係 (青 :  $\log\alpha \simeq 0$  かつ  $\log\beta > 1.18$ , 赤 :  $\log\alpha > 0$  または  $\log\beta \leq 1.18$ )

#### 拡散率 $\delta$ と次数分布のパラメータ $\log\alpha$ の関係についての分析

図 4.9 に拡散率  $\delta$  と次数分布のパラメータ  $\log\alpha$  の関係を示す。縦軸に拡散率  $\delta$ 、横軸に次数分布のパラメータ  $\log\alpha$  をとり、各点はデータセットに含まれるネットワークである。これより、 $\log\alpha \simeq 0$  でなければ、拡散率  $\delta$  は小さな値を取らない事がわかる。実際に、表 4.2 より、次数分布のパラメータ  $\log\alpha$  と拡散率  $\delta$  の間には相関係数 0.398 の強い正の相関が確認されている。 $\log\alpha$  が非常に小さい時、次数分布はべき分布に近い形状を取る。すなわち、次数分布がべき分布に近い形状を取った時は情報が拡散されにくくなると言える。

図 4.9 について、点の密度が濃い部分が 2 箇所見られる。密度の濃い部分だけを見ると、 $\log\alpha$  が高くなると急激に上昇し、 $\log\alpha = 0.5$  付近から徐々に下がっていく様子が見える。これを分析するために、以下の式を用いて分析を行う。

$$\delta = \begin{cases} \frac{a}{1+b \cdot \exp(c \cdot \log\alpha)} & (\log\alpha \leq d \text{ の時}) \\ e \cdot \log\alpha + \frac{a}{1+b \cdot \exp(c \cdot d)} - e \cdot d & (\log\alpha > d \text{ の時}) \end{cases} \quad (4.5)$$

この時、 $a, b, c, d, e, f$  はフィッティングパラメータであり、これらの値を、この曲線から正

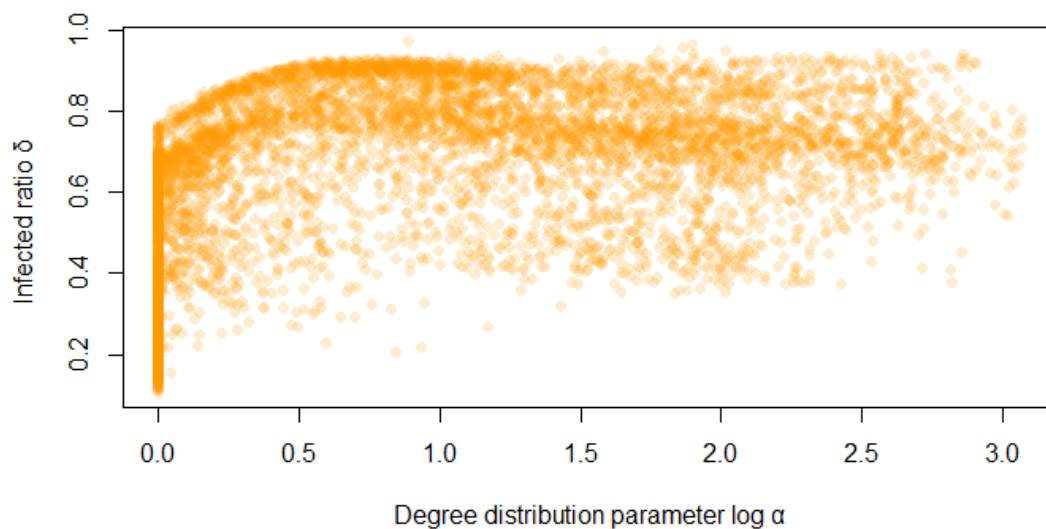


図 4.9: 次数分布のパラメータ  $\log \alpha$  と拡散率  $\delta$  の関係

規化距離 0.05 の範囲にデータセットが多く存在するように設定した．これにより，2 本の曲線を引き，各ネットワークをどちらの曲線に近いかによって，2 つのグループに分けた．

図 4.10 に図 4.9 を 2 つのグループに分けた図を示す．この 2 つのグループに対して決定木を構築し，図 4.11 に示す．この決定木は，それぞれの条件におけるグループ 1 とグループ 2 のネットワークの割合をそれぞれ示している．これより， $C \geq 0.151$  であるネットワークの 94.9% が拡散率の低いグループ 2 に所属している事がわかる．一方， $C < 0.151$  かつ  $L < 2.82$  であるネットワークの 82.1% が拡散率が高いグループ 1 に所属している．

図 4.12 に，図 4.9 を  $C < 1.51$  を緑， $C \geq 1.51$  かつ  $L < 2.82$  を青，それ以外を赤で色分けした図を示す．図 4.12 より，緑の点はグループ 1 に所属し，青い点はグループ 2 に所属しているのがわかる．また，赤い点は様々な値の拡散率  $\delta$  を取っている．これより， $C < 1.51$  の時は，拡散率  $\delta$  が高いグループ 1 に所属し， $C \geq 1.51$  かつ  $L < 2.82$  の時は拡散率  $\delta$  の低いグループ 2 に所属する事が確認された．

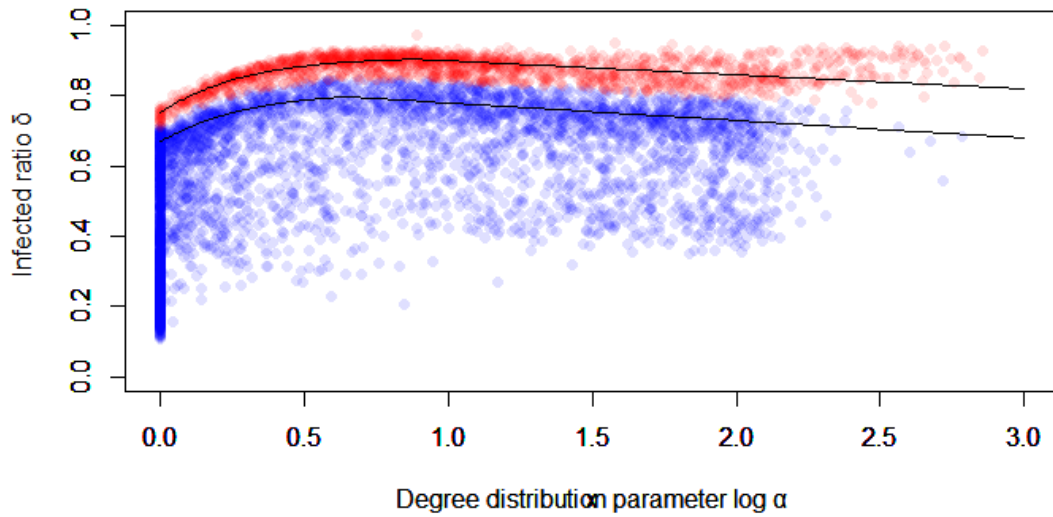


図 4.10: 次数分布のパラメータ  $\log \alpha$  と拡散率  $\delta$  の関係によるグループ分け

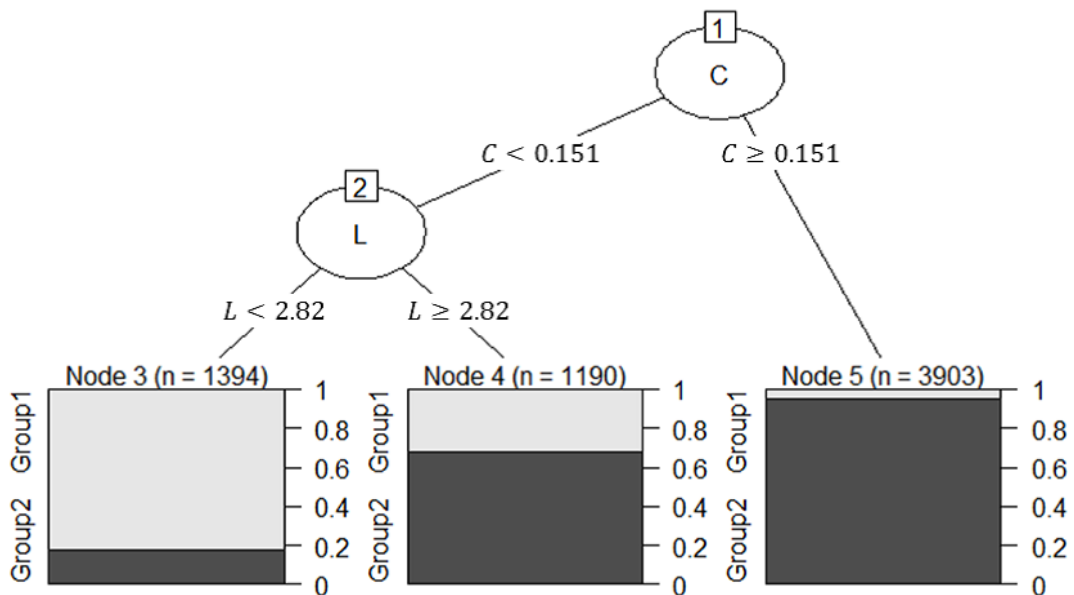


図 4.11: グループに関する決定木 (次数分布のパラメータ  $\log \alpha$  に関して)



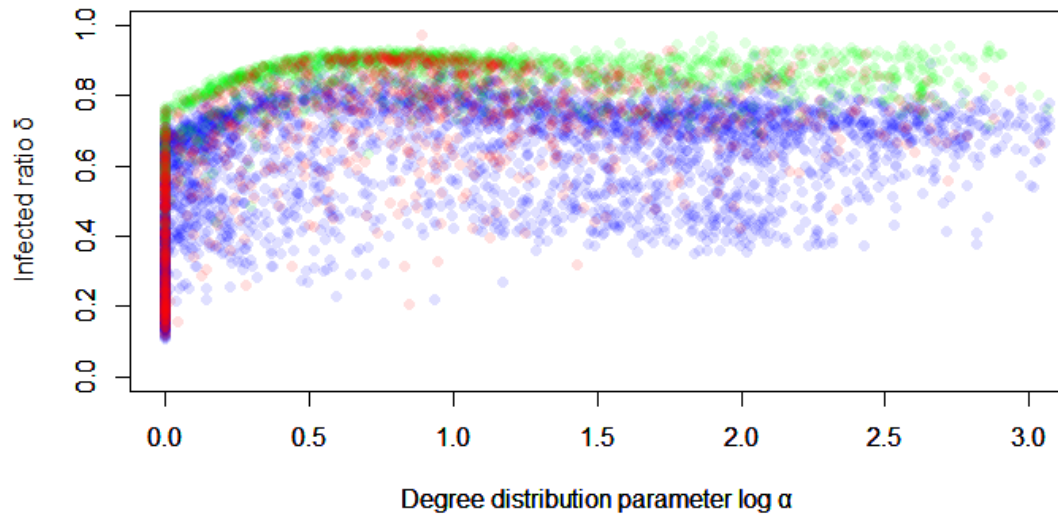


図 4.12: 次数分布のパラメータ  $\log \alpha$  と拡散率  $\delta$  の関係 (緑 :  $C < 1.51$ , 青 :  $C \geq 1.51$  かつ  $L < 2.82$ , 赤 : その他)

### 3次元構造特徴空間

ここまでの結論において情報拡散率の高いネットワークの特徴として以下のような特徴が得られた。

- 平均経路長が小さい
- 次数分布のパラメータ  $\log \alpha > 0$
- クラスタ係数が小さい

これを念頭に置きつつ、最後に相関の大きかった平均経路長  $L$ 、クラスタ係数  $C$ 、次数分布のパラメータ  $\log \alpha$  の3次元特徴空間における拡散率  $\delta$  の分布をみしてみる。

図 4.13 に3次元の特徴空間に各ネットワークをプロットした図を示す。赤い点は拡散率  $\delta$  が大きいネットワークであり、青い点が拡散率  $\delta$  が小さいネットワークである。 $L, C$  が小さく、図 4.13 で言えば手前に赤い点が集中している事がわかる。また、 $\log \alpha < 0.5$  まではわずかに小さい拡散率を取っているが、 $\log \alpha > 0.5$  では赤い点が集中している。従って、これらの3つの特徴が拡散率  $\delta$  に影響している事は明らかである。

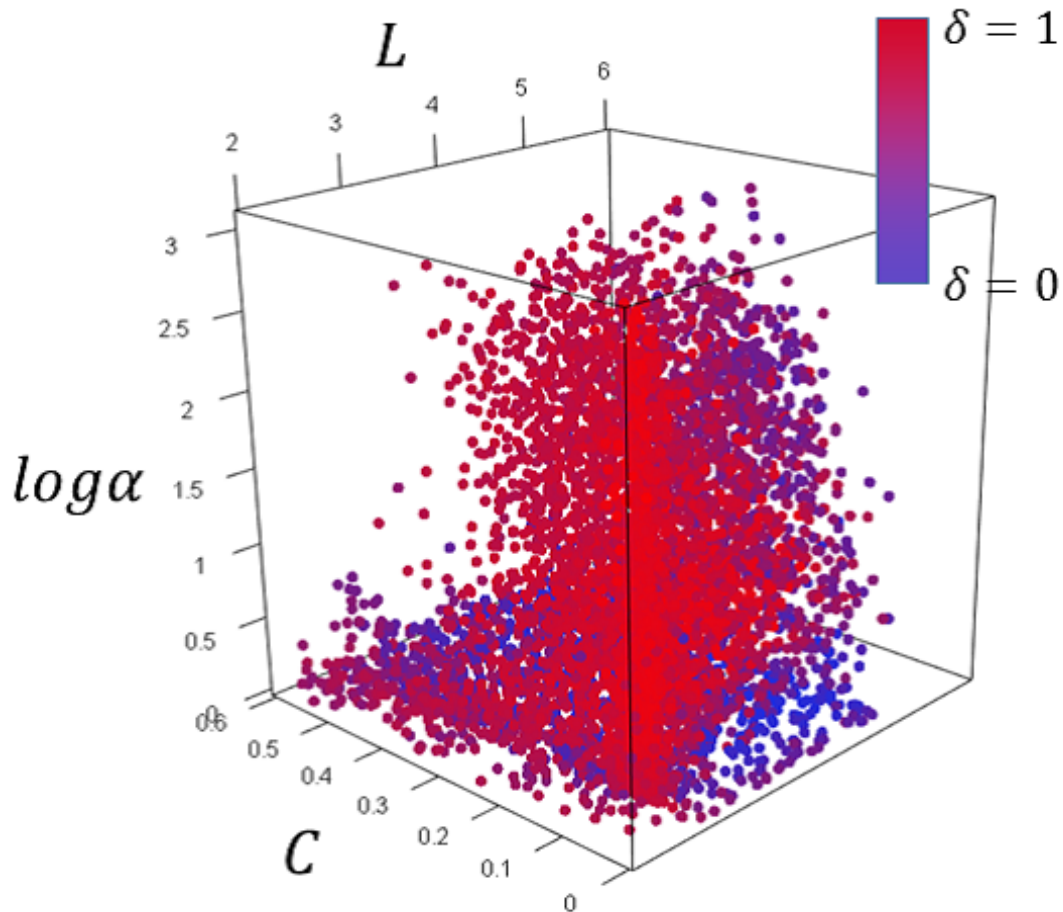


図 4.13:  $L, C, \log \alpha$  と IC mode における Infected degree  $\delta$  の関係

### 4.3.3 拡散率 $\delta$ の推定実験

前節では、4つの特徴が拡散率  $\delta$  に影響している事を示した。ここでは、これらの4特徴が拡散率  $\delta$  の決定にどの程度寄与しているのかを分析する。本研究では、5特徴を用いて分析を行っているが、そもそもこの5特徴で拡散率  $\delta$  が決定されない可能性もある。すなわち、5特徴以外に拡散率  $\delta$  に影響している特徴が存在する可能性である。これを検証するために、本研究では、構造特徴を用いて拡散率  $\delta$  を推定する事が可能かどうかを検証する。この推定が可能であれば、推定に用いた構造特徴を用いて拡散率  $\delta$  を決定する事が出来る事を意味する。なお、回帰分析に当たり、各特徴の値は標準化した。

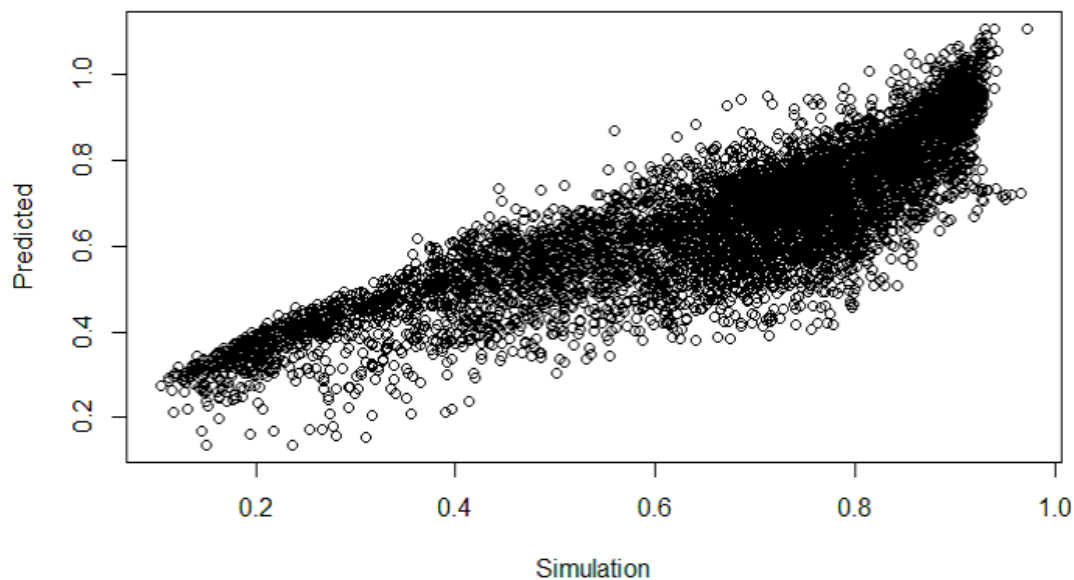


図 4.14: 重回帰分析の結果の推定値と実測値の関係

#### 拡散率 $\delta$ の重回帰分析

本研究ではまず、各特徴の拡散率  $\delta$  への複合的な影響を分析するために、重回帰分析を行う。重回帰分析の結果を図 4.14 に示す。図 4.14 は、縦軸に推定値を取り、横軸に実測値を取ったグラフで、各点はネットワークを表している。この推定値と実測値の関係の決定係数  $R^2$  は 0.691 であり、低い決定係数が得られた。この結果より、重回帰分析では拡散率を高精度に推定できていない事がわかる。特に、拡散率  $\delta$  の実測値が 0.4 より大きいところでは、推定値のばらつきが非常に大きく、推定できていない事がうかがえる。これは、前節で確認したように、各特徴と拡散率  $\delta$  の関係が線形相関ではない事に起因する。非線形での推定が必要になると考えられる。

また、表 4.3 に各特徴に関する偏回帰係数を示す。偏回帰係数は、次数分布のパラメータ  $\log\alpha$  及び  $\log\beta$  に関して非常に強く出ている事がわかる。 $\log\alpha$  は大きい程、 $\log\beta$  が小さい程、拡散率  $\delta$  が大きい事が示されている。 $\log\alpha$  は小さく、 $\log\beta$  が大きい時は、ネットワークはべき分布に近い形状の次数分布を取る。したがって、べき分布の時に拡散率  $\delta$  が低いという前節の知見と一致する。

さらに、平均経路長  $L$  及びクラスタ係数  $C$  にもわずかに負の値が出ている事がわかる。平均経路長  $L$  及びクラスタ係数  $C$  が小さい程拡散率  $\delta$  が小さいという前節の知見と一致する。最後に、次数相関が最も影響を与えていないという結果も前節の知見と一致する。

### 拡散率 $\delta$ の非線形な推定実験

■**実験設定** 非線形回帰分析として、サポートベクター回帰 (Support Vector Regression:SVR) [80] を用いる。SVR は、サポートベクターマシン (Support Vector Machine:SVM) を回帰分析のために拡張した非線形回帰分析手法である。SVR による回帰式は以下のような式で与えられる。

$$y = f(x) = \sum_{k=1}^n (\alpha_k - \hat{\alpha}_k) K(x, x_k) + b \quad (4.6)$$

なお、 $\alpha_{i=1}^n, \hat{\alpha}_{i=1}^n$  及び、 $b$  は学習の結果として得られるパラメータである。また、 $K(x, x')$  はカーネル関数である。カーネル関数としては、ガウシアンカーネル (Radial Basis Function:RBF) $K(x, x')$  を用いる。

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4.7)$$

この際、 $\gamma$  はガウシアンカーネルのハイパーパラメータである。

また、ガウシアンカーネルを用いた SVR には以下の 2 つのハイパーパラメータが存在する。

- コストパラメータ  $C$
- ガウシアンカーネルのパラメータ  $\gamma$

ここでは、この二つのハイパーパラメータを焼きなまし法 (Simulated Annealing:SA) [81] によって最適化する。さらに、評価関数として、推定値と実測値の決定係数  $R^2$  を与える。評価には 8 分割交差検定 (8-fold cross validation) を用いる。

■**推定結果** まず、本研究で用いた 5 つの構造特徴を用いて、拡散率  $\delta$  を推定可能であるかを検証する。推定結果の決定係数  $R^2$  は 0.906 であった。図 4.15 に推定値とシミュ

表 4.3: 重回帰分析の偏回帰係数

特徴	$L$	$C$	$r$	$\log\alpha$	$\log\beta$
偏回帰係数	-0.0668	-0.0172	0.0124	0.213	-0.183

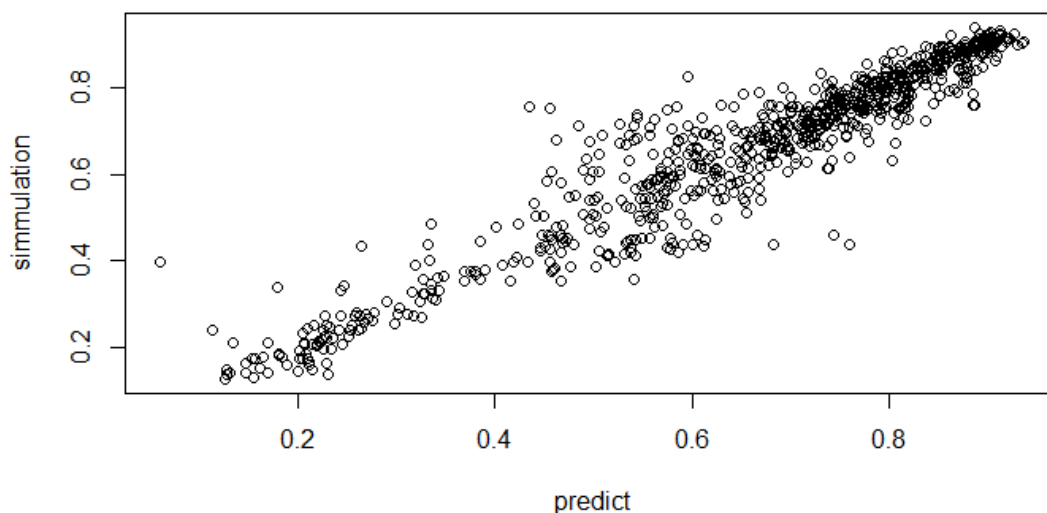


図 4.15: 拡散率  $\delta$  の推定結果

表 4.4: 説明変数毎の Infected degree  $\delta$  の推定結果

	$L, C, r, \log\alpha$	$L, C, r, \log\beta$	$L, C, \log\alpha, \log\beta$
$R^2$	0.780	0.628	0.897
	$L, r, \log\alpha, \log\beta$	$C, r, \log\alpha, \log\beta$	
$R^2$	0.856	0.790	

レーション値の関係を示す。縦軸にシミュレーション値を取り、横軸に推定値をとっている。本結果より、5つの構造特徴から拡散率  $\delta$  を十分推定できているといえる。従って、この5つの構造特徴を用いれば、拡散率  $\delta$  を決定できる。

次にそれぞれの構造特徴の寄与度を推定する。各特徴をそれぞれ抜いた4特徴から拡散率  $\delta$  を推定し、その時の下がり幅が大きい程、その特徴が与えている影響は大きい。表 4.4 に各4特徴からの拡散率の推定結果の決定係数  $R^2$  を示す。表 4.4 には、使用した4特徴とそれらの特徴からの推定結果の決定係数  $R^2$  が示されている。例えば、 $L, C, r, \log\alpha$  は平均経路長  $L$ 、クラスタ係数  $C$ 、次数相関  $r$  及び次数分布のパラメータ  $\log\alpha$  を用いて推定実験を行った事を意味する。

まず、 $L, C, \log\alpha, \log\beta$  での推定結果に着目する。推定結果の決定係数  $R^2$  は0.897を示

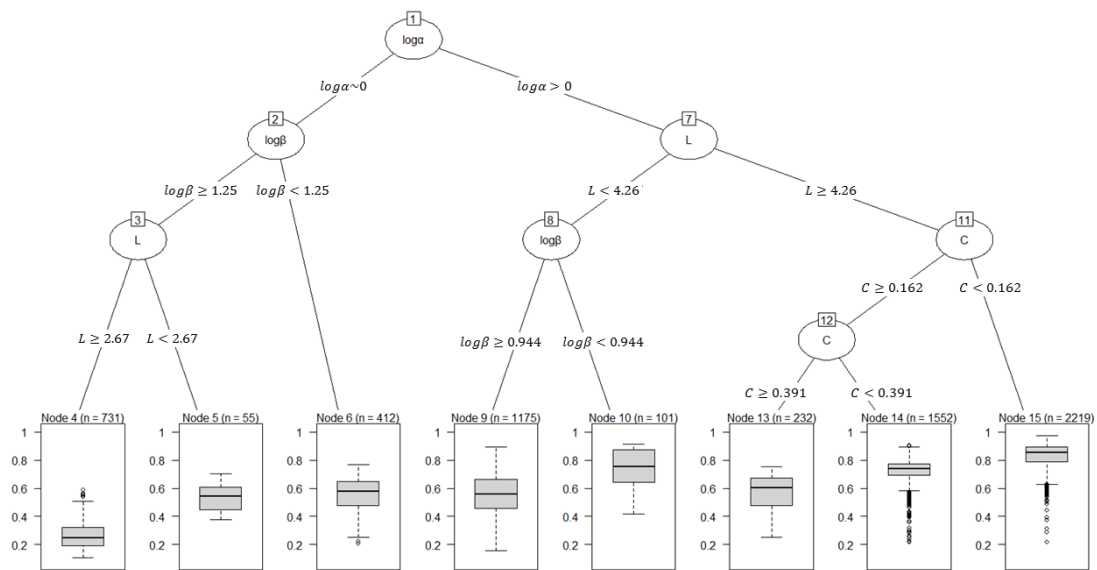


図 4.16: 決定木

し、これは 5 特徴での推定結果とほぼ変わらない。この推定結果から、次数相関  $r$  は拡散率  $\delta$  に寄与していない事を示している。この結果は、前節の結果を裏付ける内容である。

続いて、 $L, C, r, \log \alpha$  での推定結果に着目する。推定結果の決定係数  $R^2$  は 0.780 となった。この推定結果より、次数分布のパラメータ  $\log \beta$  は拡散率  $\delta$  の推定に強く寄与している事が示されている。

最後に、 $L, C, r, \log \beta$  での推定結果に着目する。推定結果の決定係数  $R^2$  は 0.628 を示し、最小の結果となった。この推定結果より、 $\log \alpha$  が最も推定結果に寄与している事がわかった。

### 4.3.4 決定木分析

ここまでの分析で、ある特徴と拡散率  $\delta$  の関係が、他の特徴によってク変わってくる場合がある事を確認した。すなわち、単独の特徴ではなく、複数の特徴を同時に見る必要があると考えられる。ここでは、本研究において用いている 5 つの構造特徴を説明変数とし、拡散率  $\delta$  を目的変数として決定木を構築した。なお、決定木の複雑度パラメータ (Complexity:  $cp$ ) は  $cp = 0.01$  とし、10 分割交差検定の誤差は 0.241 であり、十分に精度の高い決定木であるといえる。図 4.16 に決定木を示す。

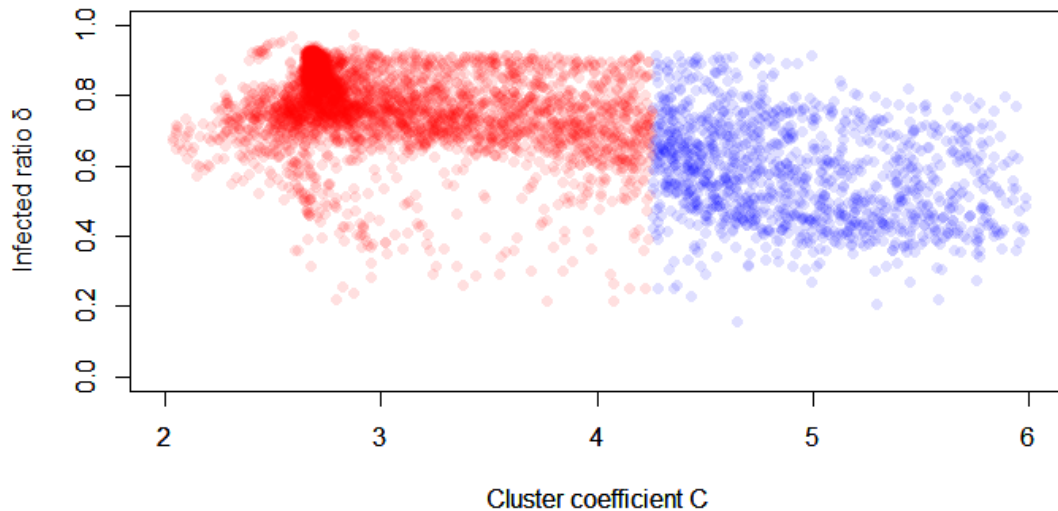


図 4.17:  $\log\alpha > 0$  における平均経路長  $L$  と拡散率  $\delta$  の関係 (赤 :  $L < 4.26$ , 青:  $L > 4.26$ )

まず、本決定木において着目する点は、次数相関による分岐が存在しない点である。すなわち、情報拡散  $\delta$  に対して、次数相関は影響を及ぼさない事が示されている。これはここまでの結果を裏付ける。

次に着目する点は、次数分布のパラメータ  $\log\alpha$  である。次数分布のパラメータ  $\log\alpha$  は一番根に近いところで分岐があり、 $\log\alpha > 0$  であるかどうかで分岐している。4.3.1 節において我々は、 $\log\alpha \simeq 0$  でグループ分けできる事を示した。本決定木で得られた結果はこの示唆を裏付ける。

さらに、右の枝を辿っていくと、平均経路長  $L$  についての分岐があり、 $L < 4.26$  であるかどうかで分岐している。図 4.17 にこの分岐後、つまり、 $\log\alpha > 0$  である時の平均経路長  $L$  と拡散率  $\delta$  を示す。 $L < 4.26$  である部分を赤い点、 $L > 4.26$  である部分を青い点として表す。この図を見れば、青い点のところで、急激に拡散率  $\delta$  が低下している様子がわかる。従って、拡散率は  $L = 4.26$  を超えると急激に減少する事がわかる。4.3.1 節において我々は、 $L = 4.4$  付近で急激な変化を起こしている事を示唆した。本決定木で得られた結果はこの示唆を裏付けている。

さらに、右の枝をたどっていけば、クラスタ係数  $C$  についての分岐があり、 $C > 0.162$  であるかどうかで分岐している。図 4.18 にこの分岐後におけるクラスタ係数  $C$  と拡散率

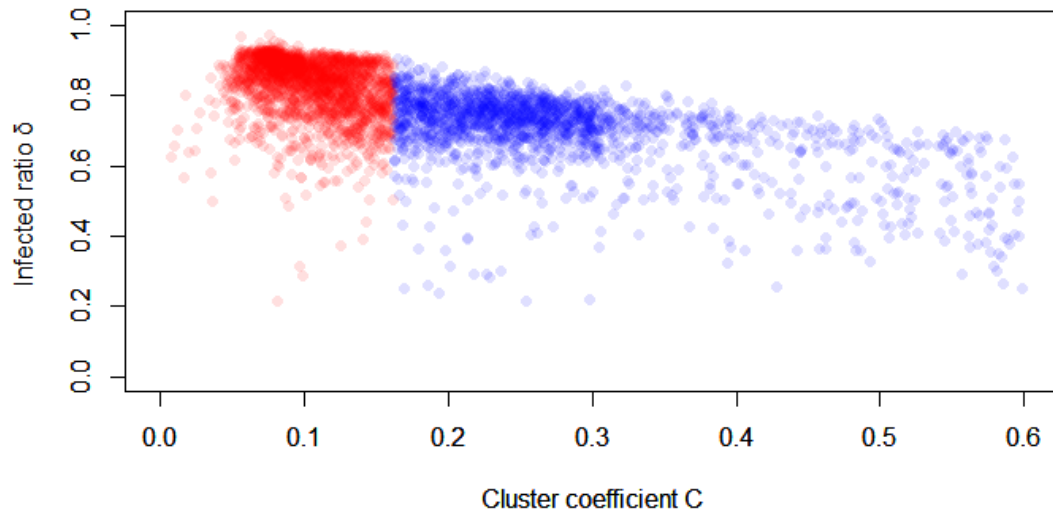


図 4.18:  $\log \alpha > 0$  かつ  $L < 4.26$  におけるクラスタ係数  $C$  と拡散率  $\delta$  の関係 (赤:  $C < 0.162$ , 青:  $C > 0.162$ )

$\delta$  の関係, つまり,  $\log \alpha > 0$  かつ  $L < 4.26$  である時のクラスタ係数  $C$  と拡散率  $\delta$  を示す.  $C < 0.162$  である部分を赤い点,  $C > 0.162$  である部分を青い点として表す. この図から, 青い点のところと赤い点のところでは明らかな差があるのがわかる. 4.3.1 節において我々は,  $C = 0.15$  付近で変化を起こしている事を示唆した. 本決定木で得られた結果はこの示唆を支持している.

最後に, 次数分布のパラメータ  $\log \beta$  について触れよう. 本決定木において, いくつかの次数分布のパラメータ  $\log \beta$  による分岐が存在する. 4.3.1 節での相関による分析では, 次数分布のパラメータ  $\log \beta$  の影響は見られなかった. しかし, 本決定木においては, 次数分布のパラメータ  $\log \beta$  は拡散率  $\delta$  に影響している事が示された. すなわち, 相関分析では見えてこなかった関係を見つける事ができた. つまりこれは,  $\log \beta$  は単独の特徴のみでは影響を与えていないが, 複合的な影響を与えている.

まず,  $\log \alpha \simeq 0$  における次数分布のパラメータ  $\log \beta$  と拡散率  $\delta$  の関係について分析する. 図 4.19 に  $\log \alpha \simeq 0$  の時の次数分布のパラメータ  $\log \beta$  と拡散率  $\delta$  の関係を示す. 縦軸に拡散率  $\delta$ , 横軸に次数分布のパラメータ  $\log \beta$  を取り,  $\log \beta < 1.25$  であるネットワークを赤い点,  $\log \beta > 1.25$  であるネットワークを青い点として表示している. これより,



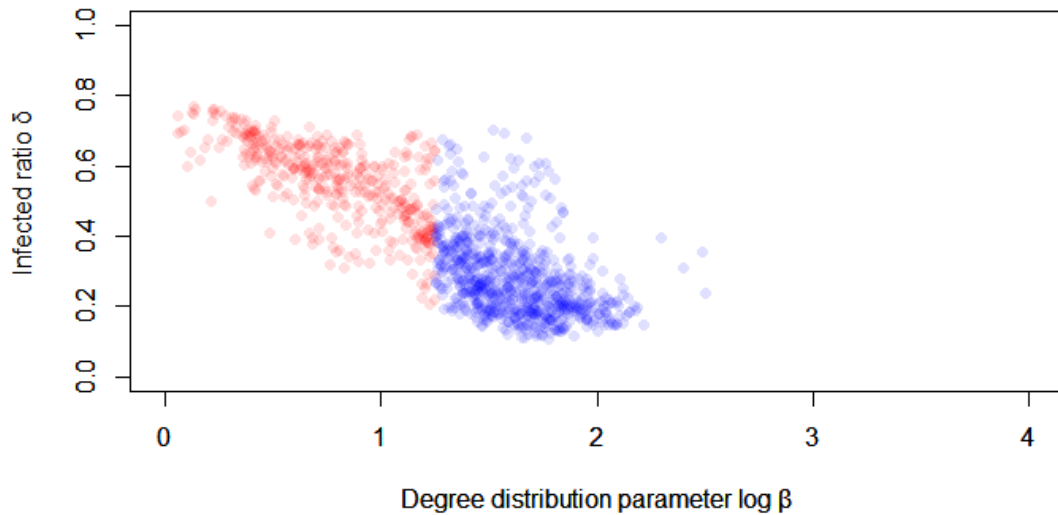


図 4.19:  $\log\alpha > 0$  における次数分布のパラメータ  $\log\beta$  と拡散率  $\delta$  の関係 (赤:  $\log\beta < 1.25$ , 青:  $\log\beta > 1.25$ )

明らかに、次数分布のパラメータ  $\log\beta$  と拡散率  $\delta$  は相関している。相関係数は  $-0.819$  であり、非常に強い負の相関を示している。従って、次数分布のパラメータ  $\log\alpha \simeq 0$  である時、次数分布のパラメータ  $\log\beta$  と拡散率  $\delta$  は強く相関する事がわかった。

ところで、 $\beta$  が大きい時のベータ分布の形状についてみる。図 4.20 に  $\log\alpha = 0$  の時のベータ分布の形状変化を示す。  $\beta$  が大きくなっていくと、次数分布の傾きが大きくなっていく。すなわち、大きなハブノードができる事となる。これは、べき指数を増やしていった時と同じである。従って、べき乗が強くなるほど、情報が伝わりにくくなる事が示唆された。

次に、 $\log\alpha > 0$  かつ  $L < 4.26$  の時の次数分布のパラメータ  $\log\beta$  と拡散率  $\delta$  の関係について見る。図 4.21 に  $\log\alpha > 0$  かつ  $L < 4.26$  の時の次数分布のパラメータ  $\log\beta$  と拡散率  $\delta$  の関係を示す。縦軸に拡散率  $\delta$ 、横軸に次数分布のパラメータ  $\log\beta$  を取り、 $\log\beta < 0.944$  であるネットワークを赤い点、 $\log\beta > 0.944$  であるネットワークを青い点として表示している。赤い点の方が高い値を取っている事がわかる。

本節では決定木を構築して拡散率と各特徴に関して分析を行い、特徴同士の複合的な影響を見る事が出来た。

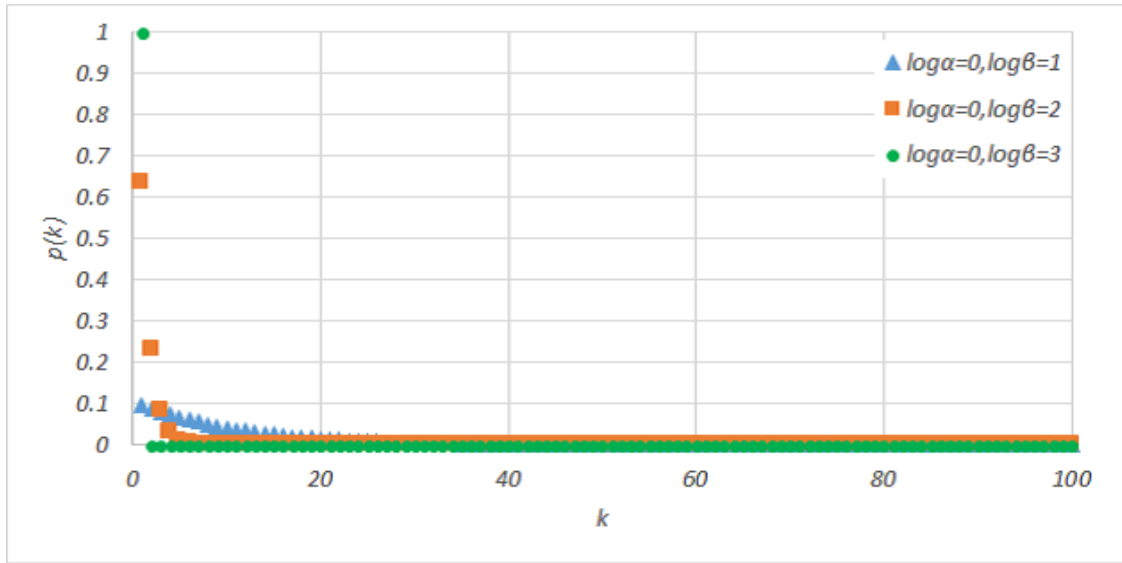


図 4.20:  $\log \alpha = 0$  の時のベータ分布の形状変化

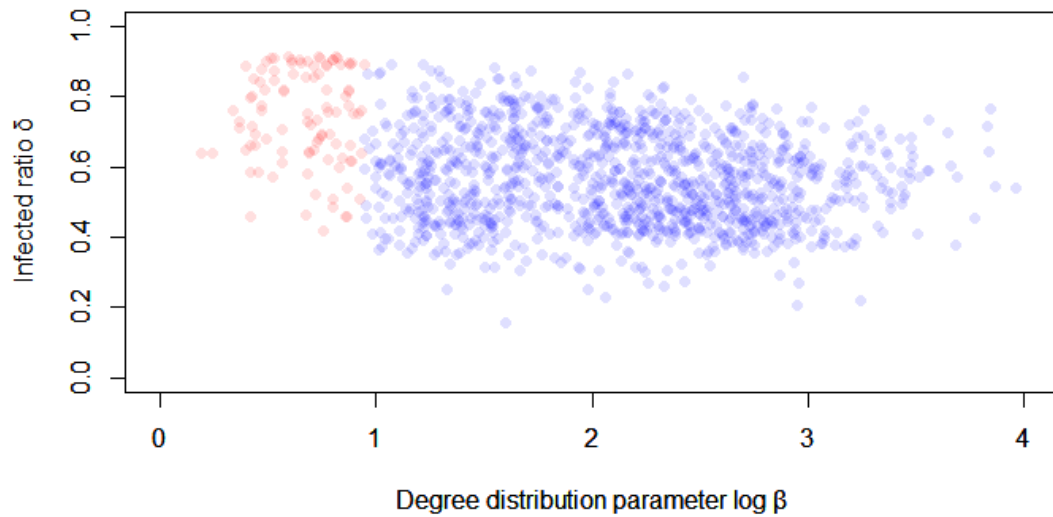


図 4.21:  $\log \alpha > 0$  かつ  $L < 4.26$  の時の次数分布のパラメータ  $\log \beta$  と拡散率  $\delta$  の関係 (赤 :  $\log \beta < 0.944$ , 青:  $\log \beta > 0.944$ )

### 4.3.5 ネットワークサンプル

最後に本データセットにおいて、最大の拡散率  $\delta$  と最小の拡散率  $\delta$  を持つネットワーク (a)(b) 及び、ネットワーク (c)(d) の各特微量と拡散率  $\delta$  を表 4.5 に示す。これらのネット

表 4.5: ネットワークサンプルの特徴量

	$L$	$C$	$r$	$\log\alpha$	$\log\beta$	$\delta$
network(a)	2.54	0.0787	-0.0433	1.86	3.01	0.956
network(b)	2.58	0.0560	-0.0154	1.90	3.03	0.967
network(c)	4.97	0.159	0.218	0.00	1.77	0.112
network(d)	4.89	0.155	0.397	0.00	1.67	0.106

ワークは、4.3.2 節の推定実験において高い推定結果を得ており、この特徴は本分析結果として高い拡散率  $\delta$  を持つと判定されるものである。

最大の拡散率  $\delta$  を持つネットワーク (a) 及び (b) は図 4.16 の決定木においては最も右の条件に合致する。すなわち、 $\log\alpha > 0$  かつ  $L < 4.26$  かつ  $C < 0.162$  である。最大の拡散率  $\delta$  を持つネットワークは、ネットワーク (a)(b) 共にランダムネットワークに近いネットワークとなっている。逆に、最小の拡散率  $\delta$  を持つネットワーク (c) 及び (d) は、図 4.16 の決定木においては最も右の条件に合致する。すなわち、 $\log\alpha \simeq 0$  かつ  $\log\beta > 1.25$  かつ  $C > 2.67$  である。最小の拡散率  $\delta$  を持つネットワークは、ネットワーク (c)(d) 共にスケールフリー性の強いネットワークとなっているのがわかる。

#### 4.3.6 IC モデルと LT モデルの拡散率の関係

最後に、IC モデルにおける拡散率と LT モデルにおける拡散率がどの程度一致するかを検証する。IC モデルと LT モデルは全く違った拡散過程である。IC モデルは送信者主体のモデルであり、LT モデルは受信者主体のモデルである。しかし、情報拡散である事は一致しているため、共通している部分も多い。例えば、ノードからノードに伝わる事や、リンクに確率が付随している事等である。この 2 つの値がどの程度似ているのかをここでは確認する。

図 4.22 に IC モデルにおける拡散率と LT モデルにおける拡散率の関係を示す。縦軸に LT モデルの拡散率を示し、横軸に IC モデルの拡散率を示している。この 2 つの関係には、グラフ左下、グラフ右上にそれぞれ点が密集した領域を確認する事が出来る。これらを分割するために、凝集型クラスタリングを行い 2 つのクラスタに分割する。凝集型クラスタリングは、まず全てのノードを別のクラスタとし、距離の近いクラスター同士を結合していくクラスタリング手法である。ここで得られた 2 つのクラスタに対して、それぞれ回帰直線を引く。さらに各ネットワークを、どの回帰直線に近いかによって、分割しなお

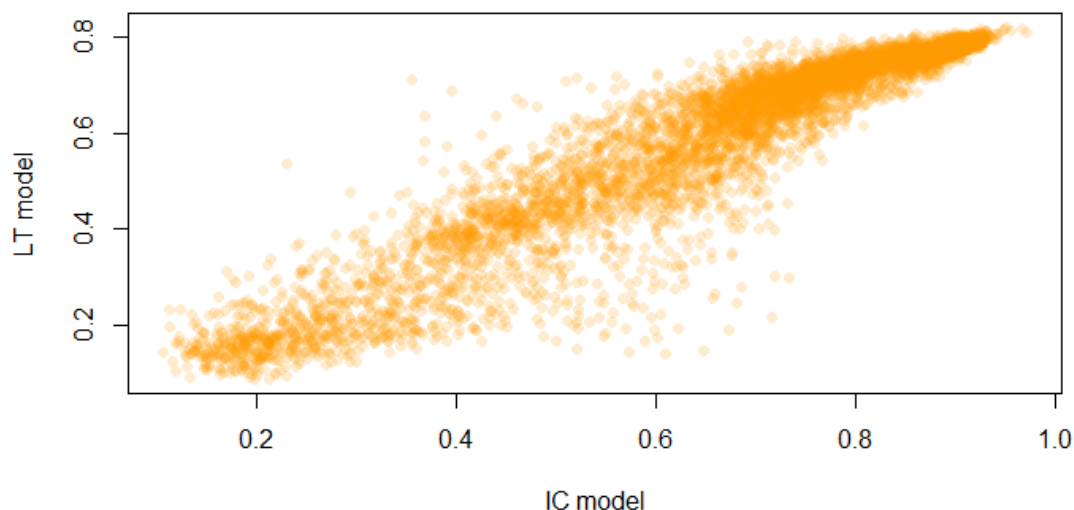


図 4.22: IC モデルにおける拡散率と LT モデルにおける拡散率の関係

す。図 4.23 に分割した結果を示す。図 4.23 における青のクラスタをクラスタ 1、赤いクラスタをクラスタ 2 とする。この図より、赤のクラスタ 2 よりも青のクラスタ 1 の方が回帰直線の傾きが大きい事がわかる。従って、青のクラスタ 1 は IC モデルの拡散率に対して LT モデルの拡散率が高く、赤のクラスタ 2 は IC モデルの拡散率に対して LT モデルの拡散率が低い。ここでは、ネットワークの構造がどのような時にクラスタ 1 及びクラスタ 2 になるかを分析する。

図 4.24 に、上記のクラスタを目的変数として構築した決定木を示す。まず、最左に位置する葉 ( $\log\alpha \simeq 0$  かつ  $\log\beta > 1.29$ ) に属するネットワークの内、88.5% のネットワークがクラスタ 2 に属している。一方、最右に位置する葉 ( $\log\alpha > 0$ ) に属するネットワークの 95.8% と、真ん中の葉 ( $\log\alpha \simeq 0$  かつ  $\log\beta \leq 1.29$ ) に属するネットワークの 88.5% がクラスタ 1 に属している。ここで、 $\log\alpha \simeq 0$  かつ  $\log\beta > 1.29$  の条件のネットワークはべき分布に近い次数分布を持っている。すなわち、次数分布がべき分布に近いネットワークがクラスタ 2 に属し、それ以外のネットワークはクラスタ 1 に属している。従って、べき分布を持つネットワークは IC モデルの拡散率に対して LT モデルの拡散率が低い事がわかる。

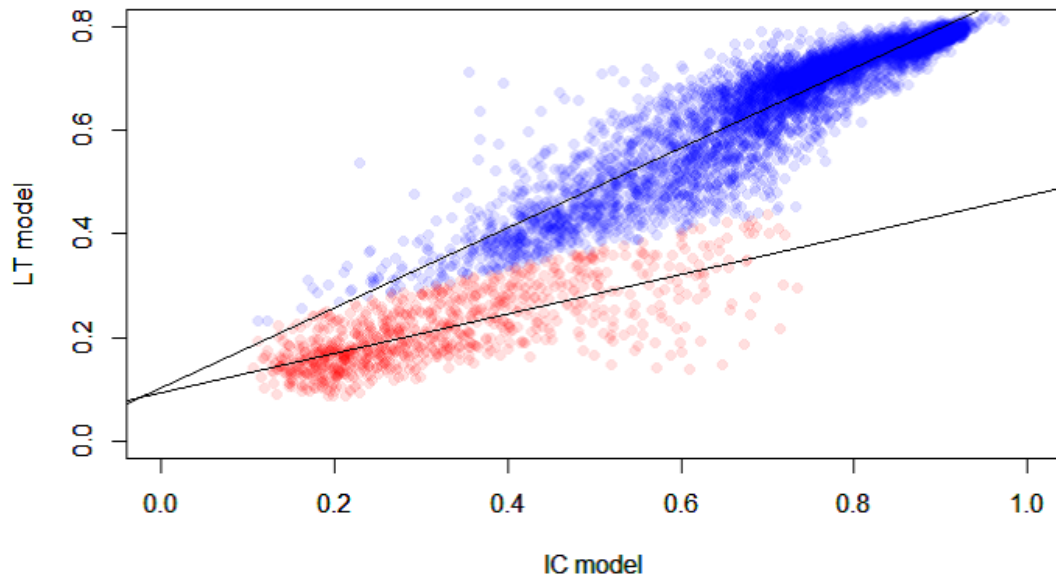


図 4.23: IC モデルにおける拡散率と LT モデルにおける拡散率の関係に関するグループ分け

## 4.4 結論

本章では、前章にて提案したネットワーク構造特徴空間を網羅するようなネットワークデータセットを用いて、情報拡散とネットワーク構造の関係について分析を行った。分析によって 5 つの構造特徴の内、情報拡散に影響を与える構造特徴とそうでない構造特徴を分類した。

分析には相関分析、決定木分析、回帰分析の 3 つの手法を用いた。その結果、平均経路長  $L$ 、クラスタ係数  $C$ 、次数分布のパラメータ  $\log\alpha$  及び  $\log\beta$  が情報拡散に影響を与え、次数相関  $r$  は影響を与えていない事を明らかにした。拡散しやすいネットワークの条件は以下の通りである。

- 平均経路長  $L$  が小さい
- クラスタ係数  $C$  が小さい
- 次数分布のパラメータ  $\log\alpha$  が大きい

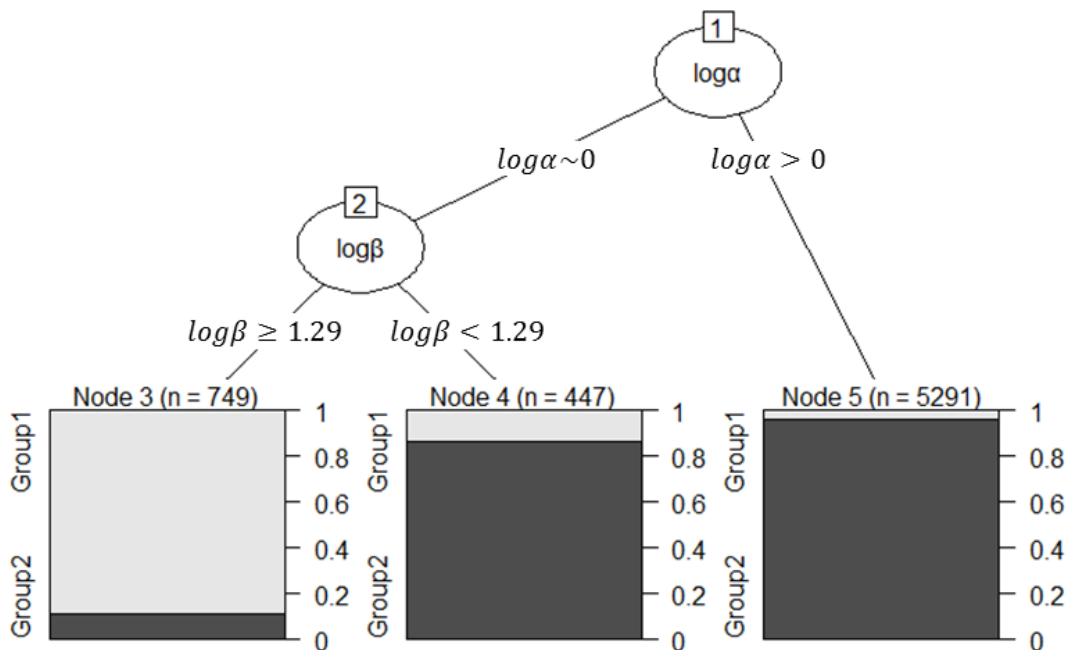


図 4.24: IC モデルにおける拡散率と LT モデルにおける拡散率の関係に関するグループ分けに関する決定木

- 次数分布のパラメータ  $\log\beta$  が小さい

この結果は、3つの分析で一貫しており、様々な角度から支持する結果を導いた。

また、分析の中で、以下に示すような急激に拡散率が変化するような点が存在した。

- 平均経路長  $L = 4.26$
- クラスタ係数  $C = 0.162$
- 次数分布のパラメータ  $\log\alpha \simeq 0$

これらの点においての変化は、今回用いていない未知の特徴の変化によって引き起こされている可能性が高い。この未知の特徴を発見し、その特徴をデータセットに組み込む事は今後の課題とする。

また、最後に別の情報拡散モデルである LT モデルの拡散率と IC モデルの拡散率を比較し、べき分布を持つネットワークは IC モデルの拡散率に対して LT モデルの拡散率が低い事が示された。

このようにして、本章ではネットワーク構造が情報拡散に与える影響を明らかにする事

に成功した.

## 第5章

# 協調行動の統計的分析

### 5.1 はじめに

人間社会はもとより自然界においても、多くの社会的ジレンマ下における集団的協調行動は観測される。協調行動は社会学や生物学、統計物理学等の多くの分野で興味を持たれ、多くの研究が行われている。多くの研究者はこの現象を扱うために囚人のジレンマゲーム (the Prisoners's Dilemma game:PD game) を扱っている。囚人のジレンマゲームは、現実社会の社会ジレンマ下における相互作用を比喩するものであり、各プレイヤーが利己的な戦略を取ると、全体としての利得が下がるというジレンマを扱う。例えば、競合関係にある企業間の価格競争や、環境問題等がこの囚人のジレンマゲームの代表として挙げられる。この囚人のジレンマゲームの応用によって、様々な現象に対しての知見が得られている [59, 60, 61].

また、これらの協調行動がどのようにして創発されるかを解明するための研究も多く行われている。一般的に、利己的なエージェントの the well-mixed group(完全ネットワーク) で囚人のジレンマゲームを行えば、協調戦略を行うエージェントは淘汰され、裏切りが支配的な戦略となることが知られている。したがって、社会においては何等かの要因が働いて協調が創発されているはずである。この要因としては様々な説が挙げられ、研究が進められている。例えば、囚人のジレンマゲームを行う各エージェントが学習をすることによって協調が創発されるという研究が行われている [62, 63, 64]. この協調行動を創発する一つの要因として、ネットワーク互惠 [20] がある。ネットワーク互惠とは、エージェント集団に空間的關係の偏り、すなわちネットワークを仮定すると、エージェントの戦略が協調もしくは裏切りのみという単純な枠組みであっても、協調が維持されるというものである。この空間型囚人のジレンマ (Spatial Prisoner's Dilemma; SPD) ゲームにお



いて、ネットワーク構造は協調的な社会が実現されるための重要な要素となる。

多くの研究者が空間的囚人のジレンマゲームにおける協調行動の進化とネットワークの構造の関係について分析をし、様々な知見が得られている。例えば、Abramson ら [68] は、ネットワークの構造と SPD ゲームの関係を理解するために、WS モデル [27] 上でシミュレーションを行っている。多くの研究においては、ER グラフ [?] や WS モデル [27] や BA モデル [28] が用いられる。いずれのネットワークも単純なアルゴリズムで作られたネットワークである。これらのネットワークは第 2 章で示したように、ネットワーク全体の構造からみれば一部のネットワークでしかない。このように一部のネットワークのみを用いても統計的に正しい結果は得られない。

また、ネットワークの構造特徴に着目した研究も行われている。一部の研究者は、次数相関やクラスタ係数を変化させることによって、それぞれの指標の違いが協調の進化にどのような影響を与えるのかを分析した [21, 22, 82, 83, 84]。これらの研究では、個々の構造特徴からの影響を分析しているが、他の構造特徴に関しては考慮されていない。したがって、偽相関である可能性や、構造特徴の複合的な作用に関しては無視されている。

このように、空間的囚人のジレンマゲームの分析は限定的な条件の下で行われており、一般的な条件が明らかになったとは言えない。十分な分析を行うためには、同一ノード数、リンク数から成るあらゆるネットワーク構造を含むネットワークデータセットを用意する必要がある。そこで、本研究では第 3 章で提案した手法によるデータセットを用いる事によって、一部の構造のネットワークではなく、様々な構造を持つネットワークを用いた分析を行う。これによって、ネットワーク構造と協調進化の関係の本質を捉える事を目的とする。

## 5.2 協調行動モデル

### 5.2.1 囚人のジレンマゲーム (Prisoner ' s Dilemma game : PD game)

囚人のジレンマゲームは 2 人で行うゲームである。プレイヤーは協調  $C$  か裏切り  $D$  のいずれかの戦略を持ち、表 5.1 に示す利得行列に従って、利得を得る。すなわち、自分が得られる利得は以下のように決まる。

- 共に協調なら  $R$
- 自分が協調、相手が裏切りなら  $S$
- 自分が裏切り、相手が協調なら  $T$
- 共に裏切りなら  $P$

表 5.1: 利得行列

	C	D
C	R/R	S/T
D	T/S	P/P

これらの利得は  $T > R > P > S$  を満たす. 多くの場合,  $T, R, P, S$  は  $T = b, R = 1, P = \epsilon, S = 0$  とされる. この時,  $\epsilon$  は極小値であり,  $b$  は協調戦略に対する裏切り戦略が得る利得である. ここでは,  $b = 1.2$  場合を分析する.

## 5.2.2 空間的囚人のジレンマゲーム (Spatial Prisoner ' s Dilemma : SPD)

本研究で用いる空間的囚人のジレンマゲーム (SPD game) をアルゴリズム 10 に示す. 空間的囚人のジレンマゲームはゲームステップと進化ステップの 2 つに分かれる. ゲームステップでは, 各リンクにつき両端のノードをプレイヤーとして, 1 回ずつ囚人のジレンマゲームを行う. それぞれのゲームで獲得した利得の総和をそのノード  $i$  の利得  $P_i$  とする. 進化ステップでは各ノードにつき, 隣接ノードからランダムに 1 つノード  $j$  を選択する.  $P_i$  よりも  $P_j$  の方が大きければ, 以下の確率  $\prod_{ij}$  でノード  $i$  はノード  $j$  の戦略をコピーする.

$$\prod_{ij} = (\max\{k_i, k_j\}b)^{-1}(P_i - P_j) \quad (5.1)$$

これを 30000 ステップ行い, 20000 ステップ経過後の 10000 ステップの協調戦略エージェントの割合の平均を算出し, ネットワークにおける協調戦略エージェントの残存率  $P_c(b)$  とする.

## 5.3 空間的囚人のジレンマゲームの統計的分析

### 5.3.1 既存研究の検証

ここでは, まず既存のネットワーク構造と協調行動との関係に関する知見について, 本データセットにおける分析においても同様の事が言えるかどうかを分析する. 最初に次数相関についての知見に着目し, 次にクラスタ係数についての知見に着目する. それぞれ, 既存の研究と一致する事を検証し, 一致しなければ, なぜ矛盾が起きたのかについて考察

---

**Algorithm 10** Spatial Prisoner's Dilemma

---

**Require:**  $s_i(t)$ :  $t$  世代におけるノード  $i$  の戦略

**Ensure:**  $s_i(0)$  はランダムに設定

```
for  $step = 0$  to 30000 do
  ゲームステップ
  for Links do
    囚人のジレンマゲーム
  end for
  進化ステップ
  for Node  $i$  do
    隣接ノードからランダムに一つノード  $j$  を選択
    if  $(P_j > P_i)$  AND  $(\max\{k_i, k_j\}b)^{-1}(P_i - P_j)$  then
       $s_i(t + 1) = s_j(t)$ 
    end if
  end for
end for
```

---

を行う。

### 次数相関に関する既存研究との比較

まず、次数相関に関する知見との比較を行う。Rong ら [21] は次数相関に着目した。Xulvi-Sokolov(XS) アルゴリズムを用いて、次数相関  $r = 0.0 - 0.3$  のスケールフリーネットワークを生成し、その上で SPD ゲームを行う事で、次数相関が協調戦略の進化に与える影響を観察した。図 5.1 に Rong らの論文から引用した次数相関  $r$  と協調戦略エージェントの割合の関係 [21] を示す。この図では、縦軸が協調戦略エージェントの割合であり、横軸は協調戦略に対する裏切り戦略が得る利得  $b$  である。また、各線はそれぞれの次数相関  $r$  の値を持つネットワークである。その結果として、協調戦略は次数相関が 0 に近いネットワークの方が残存しやすい事を示している。

図 5.2 に本データセットでの、協調戦略エージェントの割合  $P_c(1.2)$  と次数相関  $r$  の関係を示す。縦軸に協調戦略エージェントの割合  $P_c(1.2)$ 、横軸に次数相関  $r$  をとり、各点はデータセットに含まれるネットワークである。次数相関  $r$  は  $r = 0$  の時に協調戦略エージェントの割合  $P_c(1.2)$  は最も高く、正負に高くなるにつれて、低くなっているのがわか

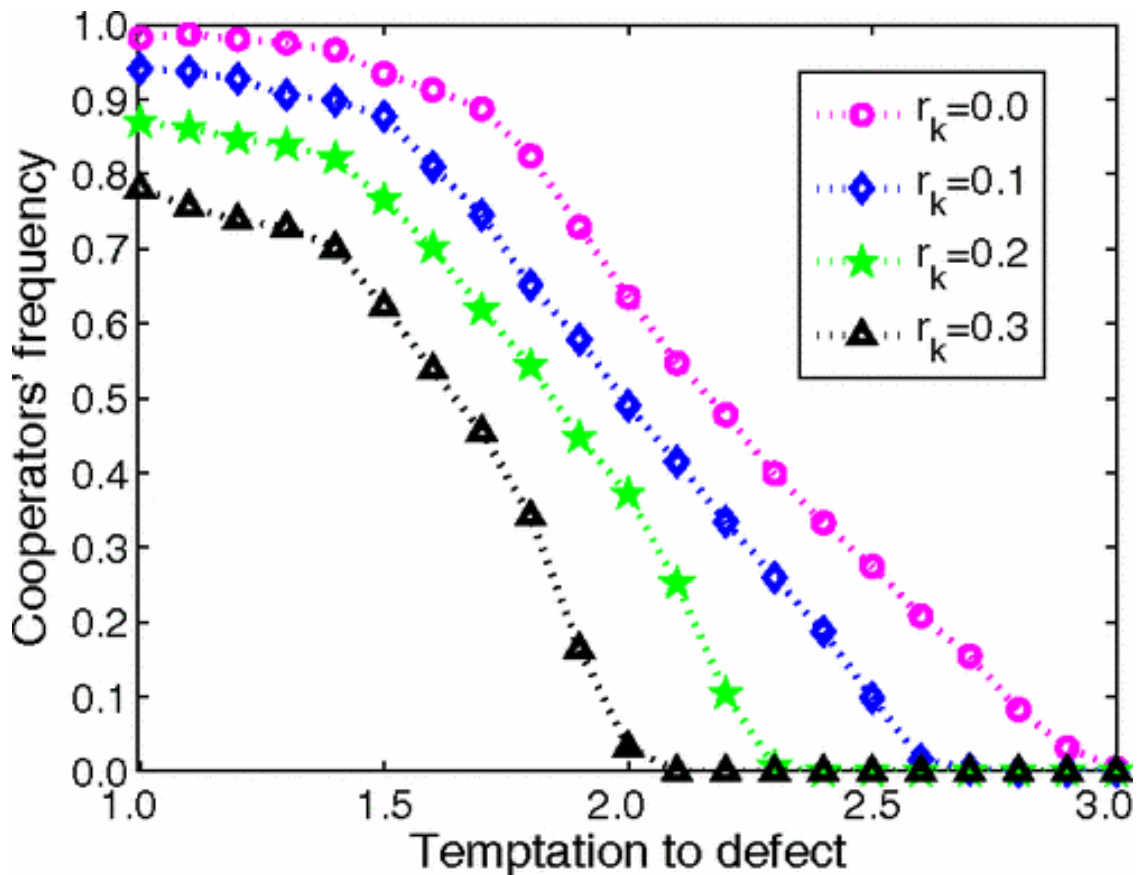


図 5.1: Rong らが示した次数相関  $r$  と協調戦略エージェントの割合の関係 [21]

る。この結果は Rong らの主張と一致している。しかし、Rong らは正の次数相関しか実験していない。本研究では、負の次数相関での実験も行い、次数相関が  $r = 0$  の時に協調戦略エージェントの割合  $P_c(1.2)$  が最大となる事を示した。

次数相関が強い負の相関を示すネットワークでは、次数の高いノードと次数の低いノードが結びついている。すなわち、ハブノード同士はリンクしていない事が多い。ハブノードの周囲は同じ戦略の集団となり、ハブノード同士がリンクしていなければ、ハブノードの戦略もほとんど変わらない。それぞれのハブノードが協調戦略をとるか、裏切り戦略をとるかはランダムに決まる。なお、ハブノードに隣接していないノードに関しては、裏切り戦略をとりやすいため、協調達成率は低い値となると考えられる..

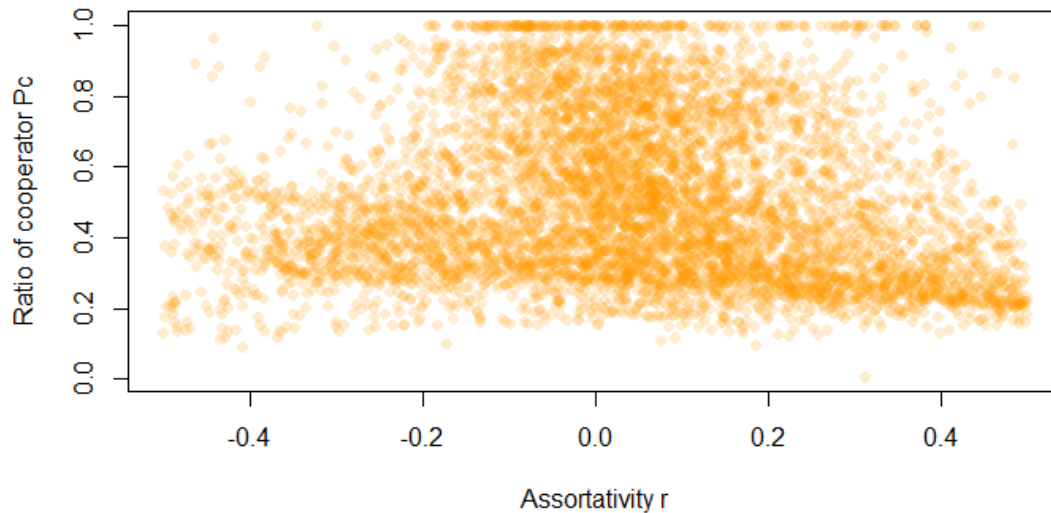


図 5.2: 次数相関  $r$  と協調戦略エージェントの割合  $P_c(1.2)$  の関係

#### クラスタ係数に関する既存研究との比較

次に、クラスタ係数に関する知見と比較を行う。Assenza ら [22] は、Holme-kim(HK)モデル [41] により、様々なクラスタ係数を持つスケールフリーネットワークを生成し、分析を行った。図 5.3 に Assenza らの論文から引用したクラスタ係数  $C$  と協調戦略エージェントの割合の関係 [21] を示す。この図では、縦軸が協調戦略エージェントの割合であり、横軸は協調戦略に対する裏切り戦略が得る利得  $b$  である。また、各線はそれぞれのクラスタ係数  $C$  の値を持つネットワークである。この結果、彼らは協調戦略に対する裏切り戦略が得る利得  $b$  が小さい時には、クラスタ係数の高いネットワークで協調戦略は生き残りやすく、この関係は協調戦略に対する裏切り戦略が得る利得  $b$  が高くなると逆転すると主張している。本研究での結果を分析し、彼らの主張と比較する。

図 5.4 に、本データセットにおける、協調戦略エージェントの割合  $P_c(1.2)$  とクラスタ係数  $C$  の関係を示す。縦軸に協調戦略エージェントの割合  $P_c(1.2)$ 、横軸にクラスタ係数  $C$  をとり、各点はデータセットに含まれるネットワークである。クラスタ係数  $C$  が低い時には、クラスタ係数の値に関わらず協調戦略エージェントの割合  $P_c(1.2)$  は高い値から低い値まで様々な値を取っている。従って、クラスタ係数  $C$  が低い時にはクラスタ係数

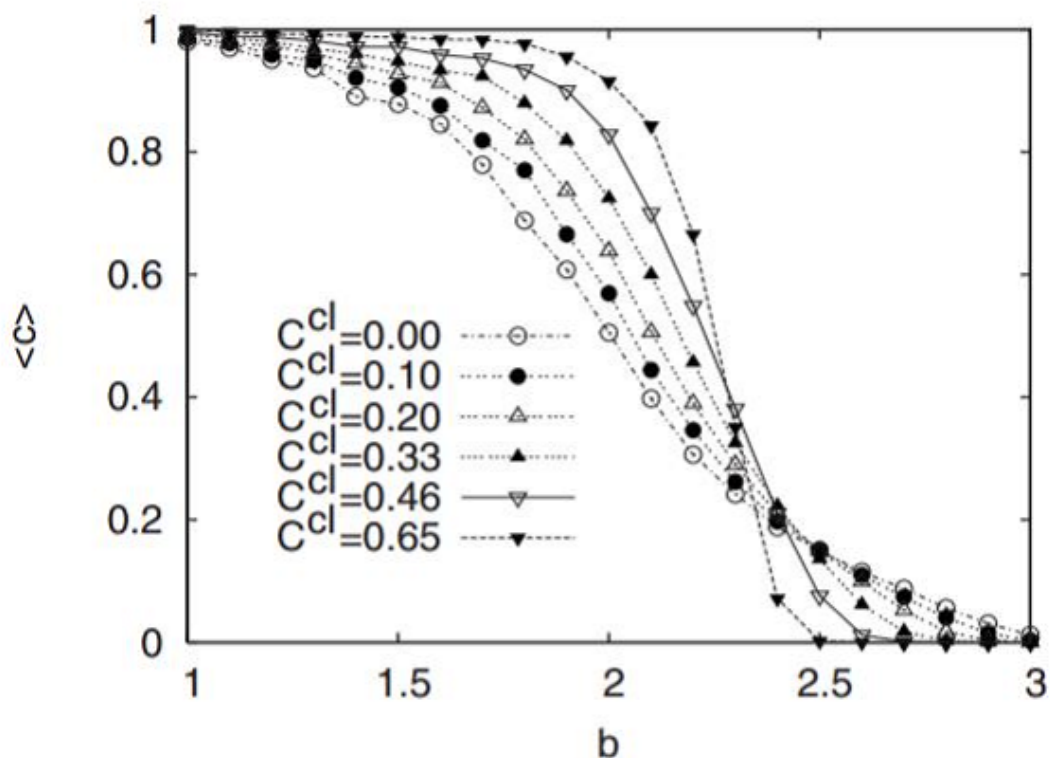


図 5.3: Assenza らが示したクラスタ係数  $C$  と協調戦略エージェントの割合の関係 [22]

表 5.2: HK モデルの各特徴量と協調戦略エージェントの割合  $P_c(1.2)$  との間の相関係数

	$L$	$C$	$r$	$\log\alpha$	$\log\beta$
$P_c(1.2)$	-0.628	-0.566	0.0988	0.0129	0.0193

$C$  と協調戦略エージェントの割合  $P_c(1.2)$  との間に相関はない。クラスタ係数  $C$  が高い時は、クラスタ係数  $C$  が高くなるにつれて、協調戦略エージェントの割合  $P_c(1.2)$  が減少していくのが見える。

この結果は Assenza らの主張と食い違っている。では、なぜこのような違いが起こったのかを考察する。まず、HK モデルの各特徴量と協調戦略エージェントの割合  $P_c(1.2)$  との間の相関係数を表 5.2 に示す。ここから、平均経路長  $L$ 、クラスタ係数  $C$  と協調戦略エージェントの割合  $P_c(1.2)$  の相関が高い事がわかる。また、彼らが用いた HK モデルの特徴間の相関係数は表 3.10 に示した通りであり、平均経路長  $L$  とクラスタ係数  $C$  は相関係数 0.913 で非常に強く相関している。

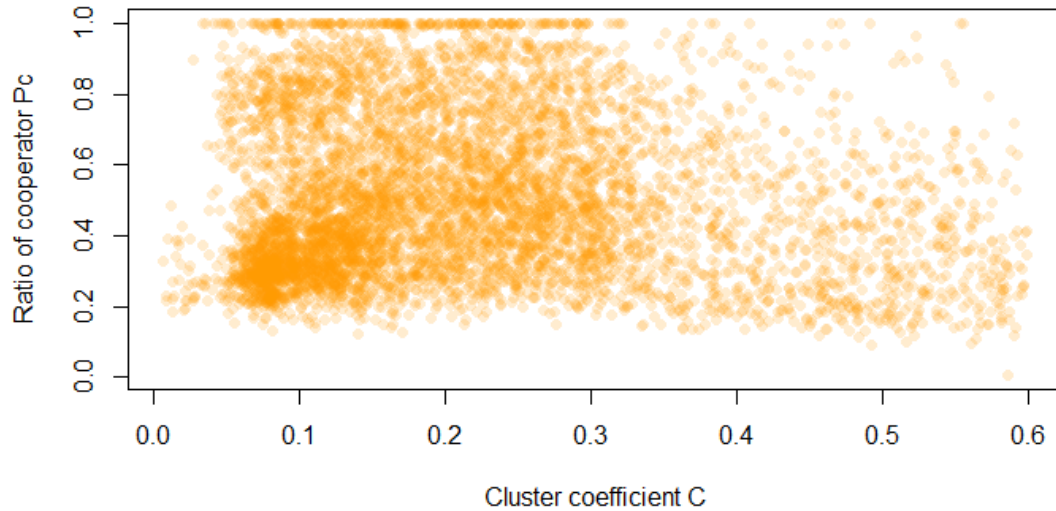


図 5.4: 協調戦略エージェントの割合  $P_c(1.2)$  とクラスタ係数  $C$  の関係

ここで、図 5.5 に、我々のデータセットにおける協調戦略エージェントの割合  $P_c(1.2)$  と平均経路長  $L$  の関係を示す。縦軸に協調戦略エージェントの割合  $P_c(1.2)$ 、横軸に平均経路長  $L$  をとり、各点はデータセットに含まれるネットワークである。これを見ると、平均経路長  $L$  が短い時は非常に高い協調戦略エージェントの割合  $P_c(1.2)$  の値をとっているが、平均経路長  $L$  が大きくなるにつれて高い協調戦略エージェントの割合  $P_c(1.2)$  をとらなくなる。相関係数は  $-0.198$  とさほど強い相関ではないが、相関している事がわかる。

従って、Assenza らの示したクラスタ係数  $C$  と協調戦略エージェントの割合  $P_c(1.2)$  の相関は、平均経路長  $L$  の偽相関である事が考えられる。ここで、協調戦略エージェントの割合  $P_c(1.2)$  とクラスタ係数  $C$  の関係から平均経路長  $L$  の関係を取り除いた偏相関係数  $r_{CP_c(1.2) \cdot L}$  を求める。

$$r_{CP_c(1.2) \cdot L} = \frac{r_{CP_c(1.2)} - r_{CP_c(1.2)}r_{CL}}{\sqrt{1 - r_{CP_c(1.2)}^2} \sqrt{1 - r_{CL}^2}} \quad (5.2)$$

偏相関係数は、 $r_{CP_c(1.2) \cdot L} = 0.0240$  となり、クラスタ係数  $C$  と協調戦略エージェントの割合  $P_c(1.2)$  の間の相関は偽相関である事が示された。

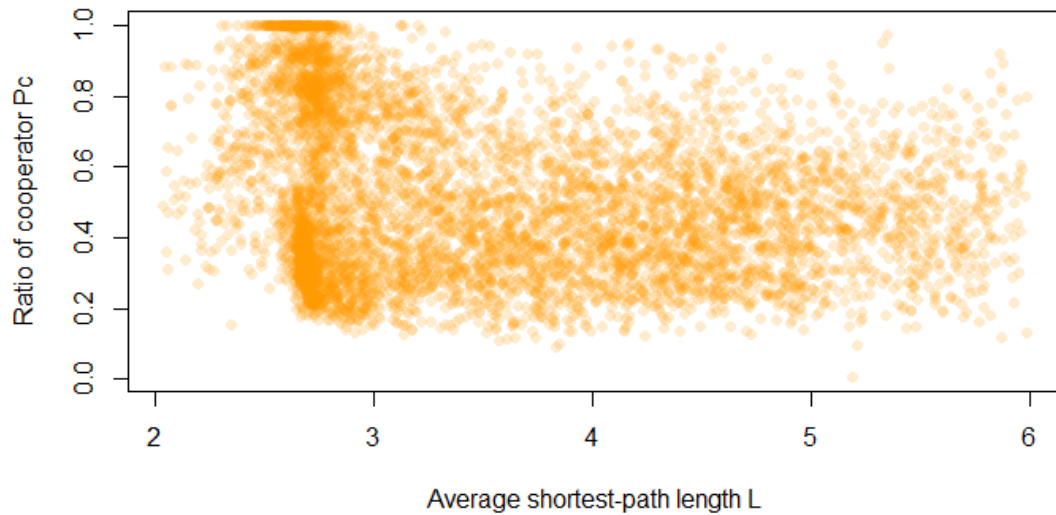


図 5.5: 協調戦略エージェントの割合  $P_c(1.2)$  と平均経路長  $L$  の関係

表 5.3: 協調戦略エージェントの割合  $P_c(1.2)$  と各特長の相関係数

$L$	$C$	$r$	$\log\alpha$	$\log\beta$
-0.198	-0.0515	-0.114	0.274	0.372

### 5.3.2 相関関係による分析

#### 拡散率 $\delta$ と各特徴間の相関係数

最初に、今回採用した 5 つの特徴と協調戦略エージェントの割合  $P_c(1.2)$  の相関を分析する。表 5.3 に、各特徴と拡散率  $\delta$  との相関を示す。表 5.3 より、平均経路長  $L$  に負の相関があり、次数分布のパラメータ  $\log\alpha$  及び、 $\log\beta$  に正の相関がある事が明確となった。また、クラスタ係数  $C$  と協調戦略エージェントの割合  $P_c(1.2)$  の間には相関はなかった。これは前節の考察と一致している。なお、次数相関  $r$  と協調戦略エージェントの割合  $P_c(1.2)$  の相関が小さいのは、線形に相関していないためである。平均経路長  $L$  と協調戦略エージェントの割合  $P_c(1.2)$  との関係はすでに触れた通りである。以下では、まだ触れていない次数分布のパラメータ  $\log\alpha$  及び、 $\log\beta$  についてみていく。



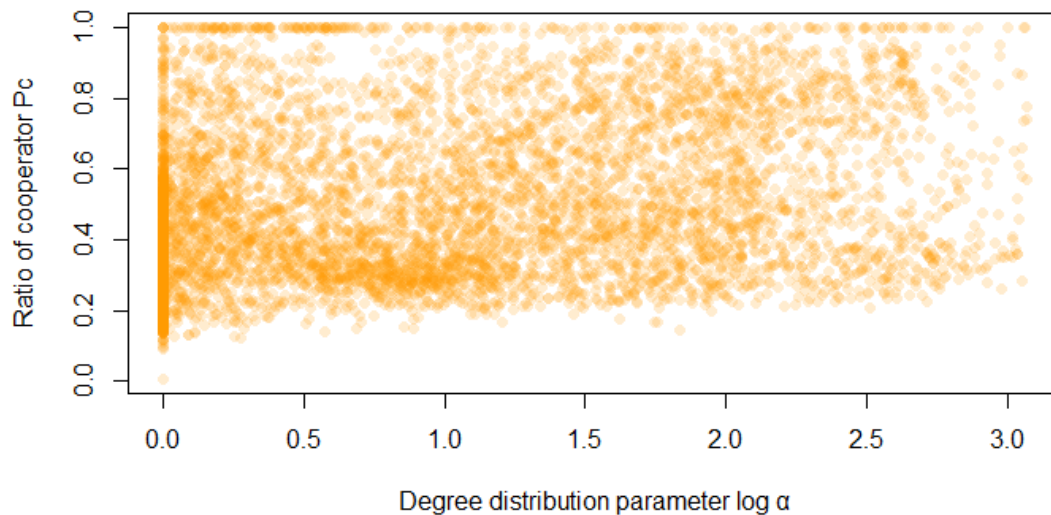


図 5.6: 協調戦略エージェントの割合  $P_c(1.2)$  と次数分布のパラメータ  $\log \alpha$  の関係

図 5.6 に、協調戦略エージェントの割合  $P_c(1.2)$  と次数分布のパラメータ  $\log \alpha$  の関係を示す。縦軸に協調戦略エージェントの割合  $P_c(1.2)$ 、横軸に次数分布のパラメータ  $\log \alpha$  をとり、各点はデータセットに含まれるネットワークである。これを見ると、次数分布のパラメータ  $\log \alpha$  に対して、高低様々な協調戦略エージェントの割合  $P_c(1.2)$  が取られている。すなわち、ほとんど関係がない事が考えられる。

図 5.7 に、協調戦略エージェントの割合  $P_c(1.2)$  と次数分布のパラメータ  $\log \beta$  の関係を示す。縦軸に協調戦略エージェントの割合  $P_c(1.2)$ 、横軸に次数分布のパラメータ  $\log \beta$  をとり、各点はデータセットに含まれるネットワークである。これを見ると、次数分布のパラメータ  $\log \beta$  が大きい程、協調戦略エージェントの割合  $P_c(1.2)$  が高い事がわかる。相関係数は 0.372 であり、強い相関がある。

次数分布のパラメータ  $\log \beta$  が大きい時、ネットワークの次数分布はべき分布に近い形状を取り、ネットワークにはハブノードが存在する。次数が大きなハブノードは利得の総和が大きくなり、戦略を変更する事が少なくなる。さらにハブノードの隣接ノードはハブノードの戦略をコピーする。したがって、協調戦略のハブノードの周りには協調戦略ノードの集団が、裏切り戦略のハブノードの周りには裏切り戦略の集団が形成される事となる。この時、協調戦略のハブノードの利得は裏切り戦略のハブノードの利得を上回る。し

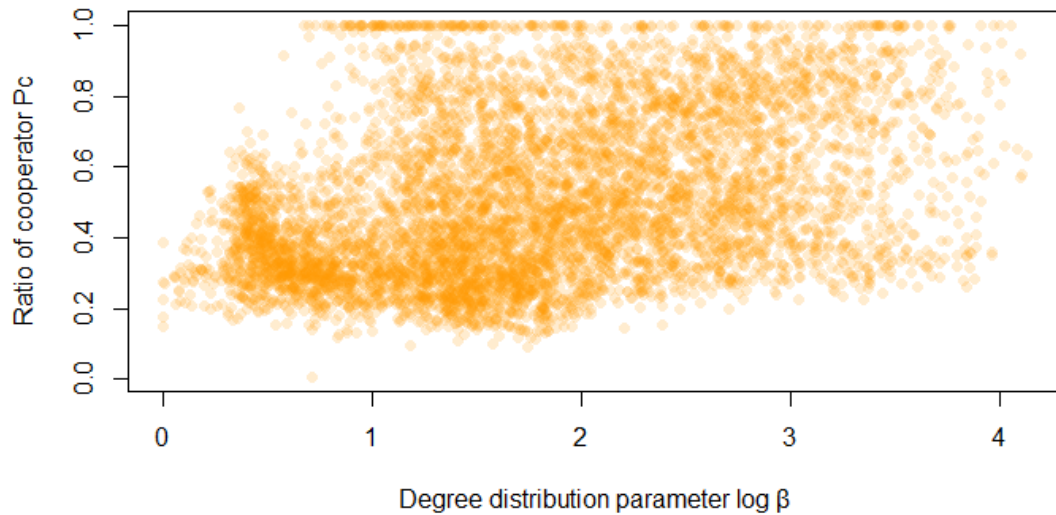


図 5.7: 協調戦略エージェントの割合  $P_c(1.2)$  と次数分布のパラメータ  $\log \beta$  の関係

たがって、協調戦略が波及し、協調達成率が上昇する。これは、Santos らの知見と一致する [88].

従って、相関係数からの分析では、平均経路長  $L$ 、次数相関  $r$ 、次数分布のパラメータ  $\log \beta$  が協調行動に影響を与えている事がわかった。

### 5.3.3 協調戦略エージェントの割合 $P_c(1.2)$ の推定実験

前節まででは、平均経路長  $L$ 、次数相関  $r$ 、次数分布のパラメータ  $\log \beta$  の 3 つの特徴が協調戦略エージェントの割合  $P_c(1.2)$  に影響している事を示した。ここでは、各特徴が協調戦略エージェントの割合  $P_c(1.2)$  の決定にどの程度寄与しているのかを分析する。本研究では、5 特徴を用いて分析を行っているが、そもそもこの 5 特徴で協調戦略エージェントの割合  $P_c(1.2)$  が決定されない可能性もある。すなわち、5 特徴以外に協調戦略エージェントの割合  $P_c(1.2)$  に影響している特徴が存在する可能性である。これを検証するために、本研究では、構造特徴を用いて協調戦略エージェントの割合  $P_c(1.2)$  を推定する事が可能かどうかを検証する。この推定が可能であれば、推定に用いた構造特徴を用いて協調戦略エージェントの割合  $P_c(1.2)$  を決定する事が出来る事を意味する。

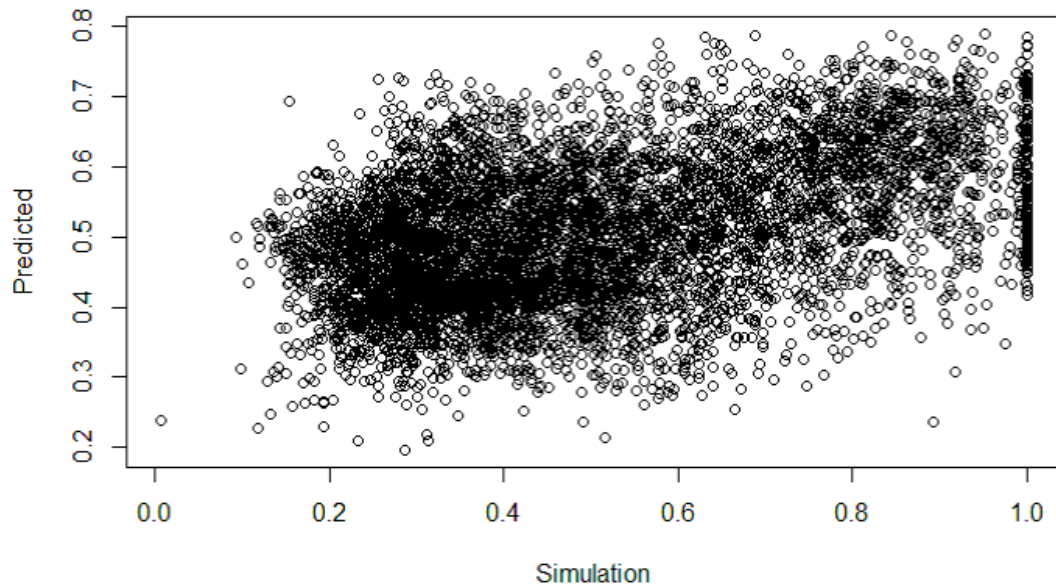


図 5.8: 重回帰分析による協調戦略エージェントの割合  $P_c(1.2)$  の推定結果の推定値と実測値の関係

#### 協調戦略エージェントの割合 $P_c(1.2)$ の重回帰分析

ここではまず、各特徴の協調戦略エージェントの割合  $P_c(1.2)$  への複合的な影響を分析するために、重回帰分析を行う。重回帰分析の結果を図 5.8 に示す。図 5.8 は、縦軸に推定値を取り、横軸に実測値を取ったグラフで、各点はネットワークを表している。この推定値と実測値の関係の決定係数  $R^2$  は 0.205 であり、低い決定係数が得られた。この結果より、重回帰分析では拡散率を高精度に推定できていない事がわかる。これは、前節で確認したように、各特徴と協調戦略エージェントの割合  $P_c(1.2)$  の関係が線形相関ではない事に起因すると考えられる。非線形での推定が必要になると考えられる。

また、表 5.4 に各特徴に関する偏回帰係数を示す。これより、次数分布のパラメータ  $\log\beta$  にのみ比較的高い正の偏相関係数が出ている事がうかがえる。これは、次数分布のパラメータ  $\log\beta$  が高い程、協調が達成しやすい事を示しており、この結果は前節の知見と一致する。その他の偏回帰係数に関しては、重回帰分析の精度が低いために、信頼できないと考えられる。

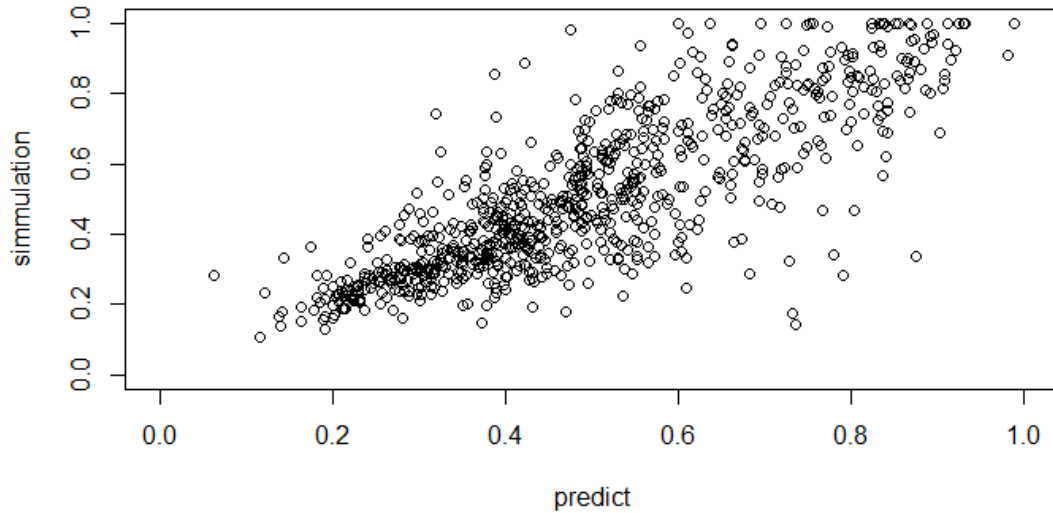


図 5.9: 協調戦略エージェントの割合  $P_c(1.2)$  の推定結果

#### 協調戦略エージェントの割合 $P_c(1.2)$ の非線形な推定実験

推定には、第 4.3.3 節と同様の設定で SVR [80] を用いる。まず、本研究で用いた 5 つの構造特徴を用いて、協調戦略エージェントの割合  $P_c(1.2)$  を推定可能であるかを検証する。推定結果の決定係数  $R^2$  は 0.651 であり、重回帰分析の結果よりはるかに高精度で推定できた。図 5.9 に推定値とシミュレーション値の関係を示す。縦軸にシミュレーション値を取り、横軸に推定値をとっている。本結果より、本研究で用いた 5 つの構造特徴を用いても、協調戦略エージェントの割合  $P_c(1.2)$  を推定しきれなかった。従って、この 5 つの構造特徴以外にも協調戦略エージェントの割合  $P_c(1.2)$  に影響を与えている特徴があると考えられる。この特徴を検討し、分析する事は今後の課題とする。

次にそれぞれの構造特徴の寄与度を推定する。各特徴をそれぞれ抜いた 4 特徴から協調

表 5.4: 協調戦略エージェントの割合  $P_c(1.2)$  に関する重回帰分析の偏回帰係数

特徴	$L$	$C$	$r$	$\log\alpha$	$\log\beta$
偏回帰係数	-0.0485	-0.0237	-0.00759	-0.0446	0.125

表 5.5: 各 4 特徴による協調戦略エージェントの割合  $P_c(1.2)$  の推定結果

	$L, C, r, \log\alpha$	$L, C, r, \log\beta$	$L, C, \log\alpha, \log\beta$
$R^2$	0.471	0.487	0.590
	$L, r, \log\alpha, \log\beta$	$C, r, \log\alpha, \log\beta$	
$R^2$	0.614	0.553	

戦略エージェントの割合  $P_c(1.2)$  を推定し、その時の下がり幅が大きい程、その特徴が与えている影響は大きい。図 5.5 に各 4 特徴からの協調戦略エージェントの割合  $P_c(1.2)$  の推定結果の決定係数  $R^2$  を示す。

まず、 $L, r, \log\alpha, \log\beta$  での推定結果に着目する。推定結果の決定係数  $R^2$  は 0.614 を示し、これは 5 特徴での推定結果をほとんど下げていない。この推定結果から、クラスタ係数  $C$  は協調戦略エージェントの割合  $P_c(1.2)$  にあまり寄与していない事を示している。この結果は、前節の結果を裏付ける内容である。

次に、 $L, C, \log\alpha, \log\beta$  での推定結果に着目する。推定結果の決定係数  $R^2$  は 0.590 を示し、これは 5 特徴での推定結果よりわずかに下がっている。この推定結果より、次数相関  $r$  は協調戦略エージェントの割合  $P_c(1.2)$  の推定に大きくは寄与している事がわかる。この結果も前節の分析結果を支持している。

次数分布のパラメータ  $\log\alpha$ ,  $\log\beta$  及び平均経路長  $L$  を除いて、行った推定実験はそれぞれ、非常に精度が下がった。特に、次数分布のパラメータ  $\log\alpha$ ,  $\log\beta$  を除いた時の推定精度は 0.471, 0.487 と大きく減少した。従って、協調戦略エージェントの割合  $P_c(1.2)$  の推定に最も強く寄与するのは次数分布であると言える。

### 5.3.4 ネットワークサンプル

最後に本データセットにおいて、最大の協調戦略エージェントの割合  $P_c(1.2)$  を持つネットワークの各特徴と協調戦略エージェントの割合  $P_c(1.2)$  を表 5.6 に示す。これらのネットワークは、前節の推定実験において高い推定結果を得ており、この特徴は本分析結果として高い協調戦略エージェントの割合  $P_c(1.2)$  を持つと判定されるものである。これらのネットワークは共に低い平均経路長を  $L$  を示し、次数相関  $r$  は  $r \simeq 0$  であり、また、 $\log\beta$  は共に 1 を超えている。これは、前節の分析と一致している。

表 5.6: ネットワークサンプルの特徴量

	$L$	$C$	$r$	$\log\alpha$	$\log\beta$	$P_c(1.2)$
network(a)	2.60	0.195	0.0537	0.613	1.40	0.987
network(b)	2.63	0.130	-0.0184	2.71	3.52	0.931

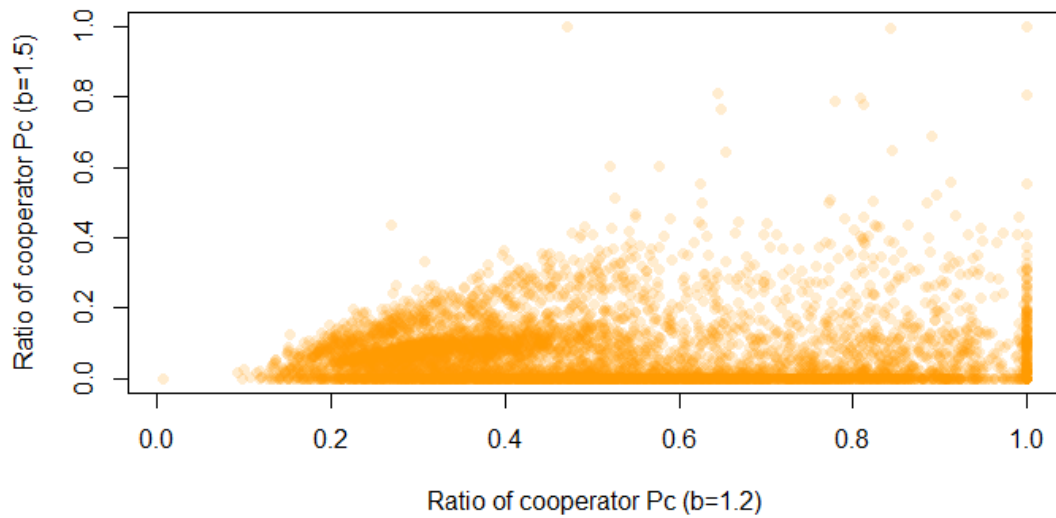


図 5.10: 協調戦略に対する裏切り戦略が得る利得  $b$  の変化による協調戦略エージェントの割合  $P_c(b)$  の変化

### 5.3.5 協調戦略に対する裏切り戦略が得る利得 $b = 1.2$ と $b = 1.5$ の時の協調戦略エージェントの割合の関係

本章の最後に、図 5.10 に協調戦略に対する裏切り戦略が得る利得  $b = 1.2$  と  $b = 1.5$  の時の協調戦略エージェントの割合の関係を示す。縦軸に協調戦略に対する裏切り戦略が得る利得  $b = 1.5$  の時の協調戦略エージェントの割合  $P_c(1.5)$  を示し、横軸に  $b = 1.2$  の時の協調戦略エージェントの割合  $P_c(1.2)$  を示している。 $b$  が大きくなれば、裏切った時の利得が大きくなるため、当然協調戦略エージェントの割合  $P_c(b)$  は減少する。全てのネットワークにおいて同じように下がれば、この二つの値は線形相関するはずである。しか

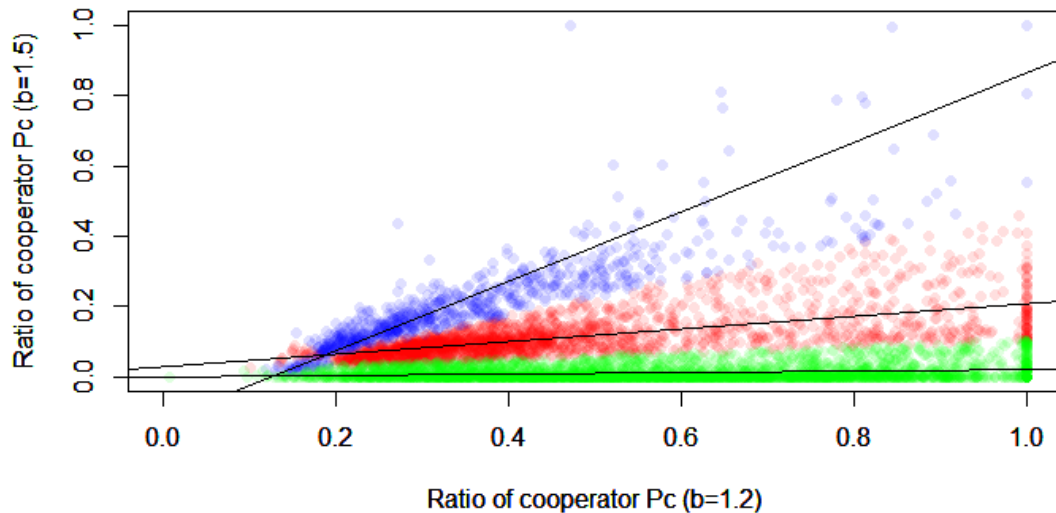


図 5.11: 3 本の回帰直線に対してのクラスタリング結果

し、図 5.10 では一つの線形相関では表せない。従って、ネットワーク構造によって、この下がり幅は違う事が示唆される。

図 5.10 では点の密度が濃くなっている部分が 3 つある。この点が密集している点を凝集型クラスタリングによって抽出し、回帰直線を引く。さらに、各ネットワークに関して、どの回帰直線に最も近いかによって、クラスタリングを行う。図 5.11 は、図 5.10 に対して 3 本の回帰直線を引き、それぞれのデータをクラスタリングした結果を示す。上から、青いグループを  $g_1$ 、赤いグループを  $g_2$ 、緑のグループを  $g_3$  とする。青いグループ  $g_1$  に含まれるネットワークは、協調戦略に対する裏切り戦略が得る利得を  $b = 1.2$  から 1.5 に変えても、協調戦略エージェントの割合  $P_c(b)$  はほとんど減少していない。逆に緑のグループ  $g_3$  に含まれるネットワークは、協調戦略に対する裏切り戦略が得る利得を  $b = 1.2$  から 1.5 に変えると協調戦略エージェントの割合  $P_c(b)$  が大きく減少する。

ここで、このグループ  $g_1$  に含まれるネットワークは協調戦略に対する裏切り戦略が得る利得が上昇しても、協調戦略エージェントの割合  $P_c(b)$  が減少しないネットワークであるとし、このグループ  $g_1$  について分析する。図 5.12 に決定木を示す。決定木の Yes はグループ  $g_1$  である割合、No はグループ  $g_1$  ではない割合をそれぞれ示している。まず、 $\log \alpha > 0$  であるネットワークの内、96.8% のネットワークがグループ  $g_1$  ではない。従っ

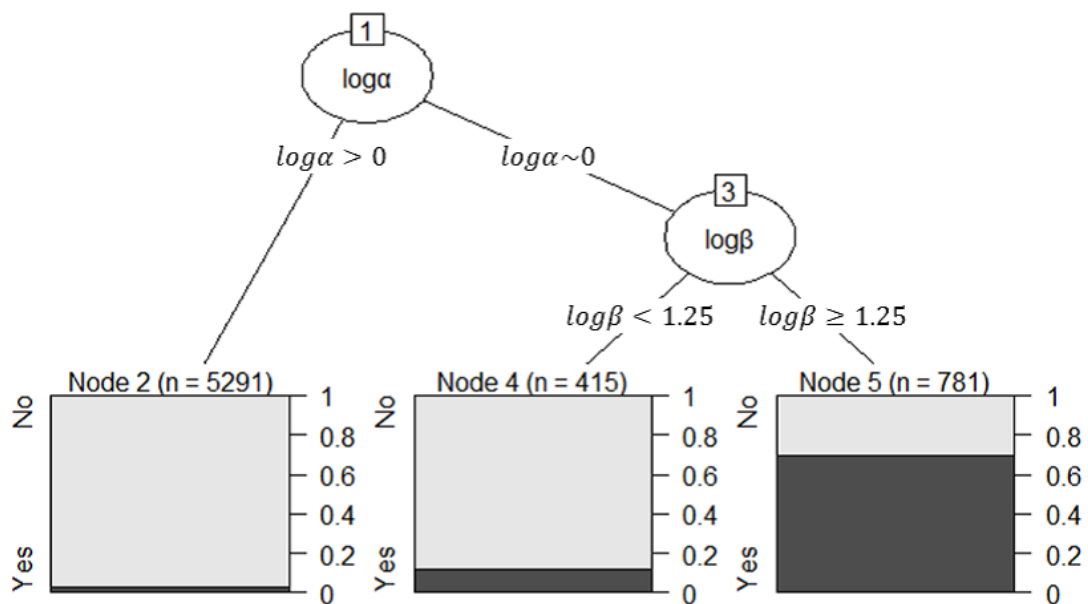


図 5.12: グループに関する決定木

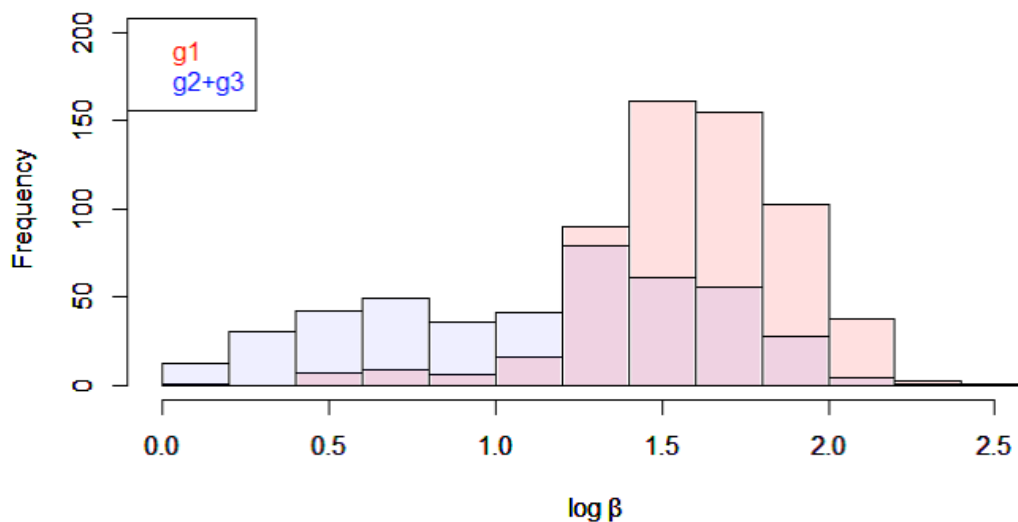


図 5.13:  $\log\alpha \sim 0$  である時の, グループ  $g1$  とそれ以外の  $\log\beta$

て, グループ  $g1$  となるための条件は  $\log\alpha \sim 0$  であることが挙げられる. また,  $\log\alpha \sim 0$  かつ  $\log\beta > 1.25$  であるネットワークの内 69.8% がグループ  $g1$  であった. 図 5.13 に,  $\log\alpha \sim 0$  である時の, グループ  $g1$  とそれ以外の  $\log\beta$  に関するヒストグラムである. 明



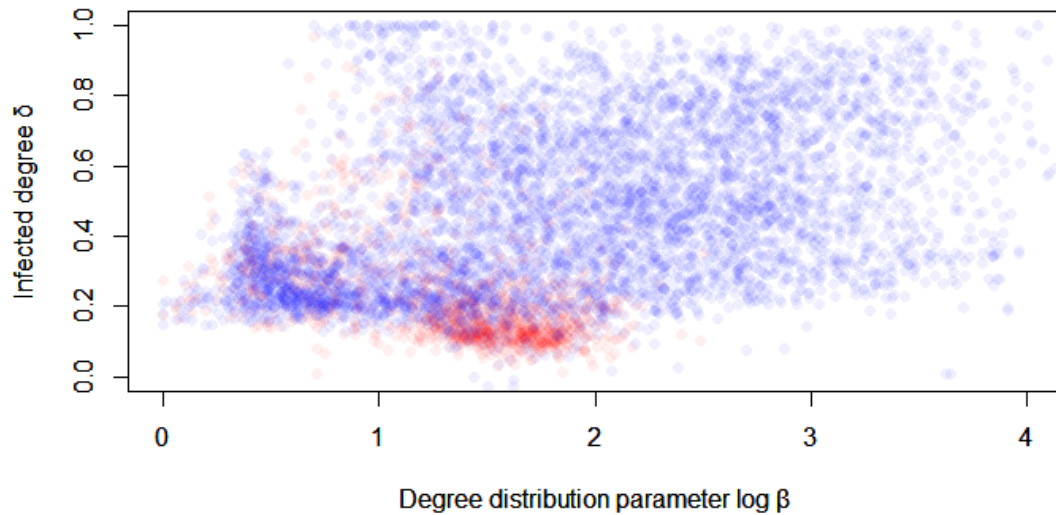


図 5.14: 協調戦略エージェントの割合  $P_c(1.2)$  と協調戦略エージェントの割合  $P_c(1.5)$  の差分 ( $P_c(1.2) - P_c(1.5)$ ) (赤  $\log\alpha \simeq 0$ , 青  $\log\alpha > 0$ )

らかに、グループ  $g1$  はその他のグループより  $\log\beta$  が大きい。従って、ネットワークの次数分布がべき分布に近いネットワークが  $g1$  に含まれやすい。すなわち、協調戦略に対する裏切り戦略が得る利得を  $b = 1.2$  から  $1.5$  に変えた時、スケールフリーネットワークにおいては協調戦略エージェントの割合  $P_c(b)$  はほとんど減少しない事が示唆された。

最後に、図 5.14 に協調戦略エージェントの割合  $P_c(1.2)$  と協調戦略エージェントの割合  $P_c(1.5)$  の差分 ( $P_c(1.2) - P_c(1.5)$ ) を示す。縦軸に  $P_c(1.2) - P_c(1.5)$  を取り、横軸に次数分布のパラメータ  $\log\beta$  をとる。赤い点は  $\log\alpha \simeq 0$  であり、青い点は  $\log\alpha > 0$  である。つまり、 $\log\beta$  が高い赤い点はべき分布を持つネットワークとなる。この図より、 $\log\beta$  が  $1 \leq \log\beta \leq 2$  の区間の  $\log\alpha \simeq 0$  のネットワークが差分が小さい事がわかる。これより、べき分布を持つネットワークは協調戦略エージェントの割合  $P_c(1.2)$  と協調戦略エージェントの割合  $P_c(1.5)$  の差分が小さい事がわかる。

## 5.4 結論

本章では、第3章にて提案したネットワーク構造特徴空間を網羅するようなネットワークデータセットを用いて、協調行動とネットワーク構造の関係について分析を行った。分析によって5つの構造特徴の内、協調行動に影響を与える構造特徴とそうでない構造特徴を分類した。

本研究では、ネットワーク上での協調行動のモデルとして空間的囚人のジレンマを用いる。空間的囚人のジレンマには協調戦略に対する裏切り戦略が得る利得  $b$  が存在する。協調戦略エージェントの割合  $P_c(b = 1.2)$  に対して各特徴が与える影響を調べた。

分析には相関分析、回帰分析の2つの手法を用いた。その結果、平均経路長  $L$ 、次数相関  $r$ 、次数分布のパラメータ  $\log\alpha$  及び  $\log\beta$  が協調行動に影響を与え、クラスタ係数  $C$  は影響を与えていない事を明らかにした。協調戦略エージェントが多く残存しやすいネットワークの条件は以下の通りである。

- 平均経路長  $L$  が小さい
- 次数相関  $r$  が0に近い
- 次数分布がべき分布

この結果は、2つの分析で一貫しており、様々な角度から支持する結果を導いた。

また、協調戦略エージェントの割合  $P_c(b = 1.2)$  と  $P_c(b = 1.5)$  において同様の結果が得られる事がわかった。協調戦略に対する裏切り戦略が得る利得  $b$  の値が大きくなると  $P_c(b)$  の値は減少していく。この差分はネットワーク構造によって大きく異なる。本章の分析によって、次数分布がべき分布に近い時、 $P_c(b)$  の減少幅が小さい事を突き止めた。

このようにして、本章ではネットワーク構造が協調行動に与える影響を明らかにする事に成功した。

## 第6章

# 結論

### 6.1 総括

本研究では、複雑ネットワーク上の様々な現象とネットワーク構造の関係を明らかにする事を目的とした。既存研究では、ネットワーク生成モデルで生成したネットワーク上での分析が多い。このネットワーク生成モデルによって生成されるネットワークはその生成過程に依存した構造の傾向を持つ。従って、ネットワーク構造特徴空間上のわずかな領域のネットワークを用いている事になる。このようなネットワークデータセットを用いて分析を行っても、統計的に正しい傾向を見る事は困難である。

本研究では、既存のネットワーク生成モデルやネットワーク生成アルゴリズムを用いて1000個ずつネットワークを生成し、これらのネットワークの平均経路長  $L$ 、クラスタ係数  $C$ 、次数相関  $r$  の値を分析した。これによって、既存のネットワーク生成モデル及びアルゴリズムがネットワーク構造特徴空間上の極一部の領域しか表現する事が出来ない事を示した。これらのモデルやアルゴリズムの中には、着目した特徴を様々な値にするものが含まれるが、それらのモデルですら、一部の領域のネットワークしか生成できない。これより、既存のネットワーク生成モデル及びアルゴリズムによって生成されたネットワークでは、統一的な議論を行うのは難しい事を示した。

そこで本研究では、このネットワーク構造特徴空間を満たすようなネットワークデータセットの構築手法を提案する。このネットワークデータセットを構築するために、様々な構造のネットワークを集める必要がある。しかし、既存のネットワーク生成モデルを用いて、ネットワーク構造特徴空間を満たす事は出来ない。そこで、本研究では様々な構造のネットワークを生成可能な手法として **Arbitrary Network Generation Algorithm(ANGA)** を提案した。ANGA は、ネットワーク構造特徴空間上の任意の特徴点を与える事によ

て、特徴点にネットワークを生成する手法である。生成過程では、各ステップにおいて  $c$  個のネットワーク候補を生成し、その中から特徴点に最も近い候補を選択する。また、ANGA はノードの選択方法によって、7つの type に分類した。

ANGA の評価を、実ネットワークを用いて検証した。まず、候補生成数  $c$  を増やす事による精度の向上を分析し、30 以上大きくしても精度は上がらない事がわかった。さらに、各 type を比較する事によって、優先的選択と CNN 選択が重要である事を示した。また、優先的選択と CNN 選択を用いる type-PC の精度が最も高くその平均精度は 0.00622 であり、高精度で生成する事が出来る事を示した。次に、既存のネットワーク生成モデルを用いて行った。10 個の既存ネットワーク生成モデルを用いて 10 個ずつネットワークを生成し、各ネットワークから特徴点を抽出して用いた。本実験においても type-PC が最も精度が高く、その平均精度は 0.00428 であり、高精度で生成する事が出来る事を示した。

この ANGA を用いて、ネットワークデータセットの構築を行った。さらに、提案手法を用いて構築したデータセットを評価した。まず、既存モデルによって構築したデータセットよりも、相関がない事を示した。次に、提案手法を用いて構築したデータセットは、既存モデルによって構築したデータセットに含まれる 90.1% のネットワークと類似したネットワークを含んでいる事を示した。また、実ネットワークに対しては、既存モデルによって構築したデータセットよりも類似したネットワークを含んでいる事を示した。最後に、構造特徴空間に対する充填率を求めた。提案手法を用いて構築したデータセットの充填率は  $3.40 \times 10^{-2}$  であり、既存モデルによって構築したデータセットの充填率  $3.38 \times 10^{-3}$  の 10 倍以上であり、提案手法を用いて構築したデータセットはより、様々な構造を含み、空間を埋めている事が示された。本データセットを用いる事によって、ネットワーク構造と現象の正確な関係が明らかになる事が期待される。

このネットワークデータセットを用いて、ネットワーク構造と現象の関係について分析を行う。本研究では、現象の例として、よく取り上げられる情報拡散と協調行動の 2 つの例を挙げて分析を行う。この分析を通して、本研究で提案したデータセットを用いる事によって、現象とネットワーク構造の関係性を分析可能である事を示す。

まず、本ネットワークデータセットを用いて、情報拡散とネットワーク構造の関係について分析を行った。分析によって 5 つの構造特徴の内、情報拡散に影響を与える構造特徴とそうでない構造特徴を分類した。分析には相関分析、決定木分析、回帰分析の 3 つの手法を用いた。

まず、各特徴と拡散率  $\delta$  の間の相関を分析した。拡散率  $\delta$  と平均経路長  $L$  との相関  $-0.502$ 、クラスタ係数との相関  $-0.511$ 、次数分布のパラメータ  $\log \alpha$  との相関  $0.398$  が示された。

次に、決定木による分析を行い、それぞれの特徴において急激に変化が起こる点を示した。平均経路長  $L = 4.26$ 、クラスタ係数  $C = 0.162$ 、 $\log\alpha \simeq 0$  において、急激な拡散率  $\delta$  の変化が起こる事を示した。さらに、単独では相関していなかった  $\log\beta$  も複合的に見れば影響を与えている事を示した。

さらに、本研究で用いた 5 つの特徴から拡散率  $\delta$  の推定を行った。推定には SVR を用い、シミュレーション結果と推定値の決定係数  $R^2$  は 0.906 であり、この 5 つの特徴によって、拡散率が決定される事を示した。また、各特徴を除いた 4 特徴からの推定結果を分析する事によって、各特徴の拡散率  $\delta$  の推定に対する寄与度を示した。推定結果から、次数分布のパラメータ  $\log\alpha$  が最も寄与度が高く、次数相関  $r$  はほとんど寄与していない事が示された。なおこの結果は、ここまでの結果を裏付けるものであった。

また、最後に別の情報拡散モデルである LT モデルの拡散率と IC モデルの拡散率を比較した。スケールフリーネットワークは IC モデルの拡散率に対して LT モデルの拡散率が低い事が示された。このようにして、本章ではネットワーク構造が情報拡散に与える影響を明らかにする事に成功した。

次に、本ネットワークデータセットを用いて、協調行動とネットワーク構造の関係について分析を行った。分析によって 5 つの構造特徴の内、協調行動に影響を与える構造特徴とそうでない構造特徴を分類した。分析には相関分析、回帰分析の 2 つの手法を用いた。

まず、各特徴と協調戦略エージェントの割合  $P_c(1.2)$  の間の相関を分析した。協調戦略エージェントの割合  $P_c(1.2)$  と平均経路長  $L$  との相関  $-0.198$ 、次数分布のパラメータ  $\log\alpha$  との相関  $0.274$ 、次数分布のパラメータ  $\log\alpha$  との相関  $0.372$  が示された。

さらに、本研究で用いた 5 つの特徴から協調戦略エージェントの割合  $P_c(1.2)$  の推定を行った。シミュレーション結果と推定値の決定係数  $R^2$  は 0.651 であり、この 5 つの特徴によって協調戦略エージェントの割合  $P_c(1.2)$  は推定しきれない事がわかった。従って、協調行動に影響を与えるネットワーク構造特徴がこの 5 つのネットワーク構造特徴以外に存在する事が示された。また、各特徴を除いた 4 特徴からの推定結果を分析する事によって、各特徴に関して協調戦略エージェントの割合  $P_c(1.2)$  の推定に対する寄与度を示した。推定結果から、次数分布のパラメータ  $\log\alpha$  及び  $\log\beta$  が最も寄与度が高く、クラスタ係数  $C$  はほとんど寄与していない事が示された。

最後に、協調戦略に対する裏切り戦略が得る利得  $b = 1.2$  と  $b = 1.5$  の時の協調戦略エージェントの割合を比較した。協調戦略に対する裏切り戦略が得る利得を  $b = 1.2$  から  $1.5$  に変えた時、スケールフリーネットワークにおいては協調戦略エージェントの割合  $P_c(b)$  はほとんど減少しない事が示唆された。

以上のように、本研究において、2 つの現象を例として挙げる事で、提案手法によって

構築されたネットワークデータセットが、現象とネットワーク構造の関係を分析するために有用である事を示した。本データセットは、ネットワーク上のあらゆる現象の分析に貢献できる事が期待される。

## 6.2 今後の課題

本研究では、あらゆるネットワーク構造を含むネットワークデータセットを構築する事を目指した。このネットワークデータセットを構築するために任意の特徴点にネットワークを生成する事が可能なネットワーク生成アルゴリズム、**ANGA** を提案した。本研究では **ANGA** を評価し、様々な特徴点にネットワークを生成可能である事を示した。

しかし、本研究では、**ANGA** が生成可能な特徴の範囲は示されていない。すなわち、**ANGA** の限界を分析する事を今後の課題として挙げる。また、この限界を広げる事も今後の課題とする。

その一つの例として、次数分布を挙げる。本研究では次数分布を表す分布として、ベータ分布を用いた。ベータ分布は様々な形状を取る事ができ、これを用いる事で様々な形状の次数分布を生成できた。しかし、表現できない次数分布も多く存在する。より、表現力の高い分布を用いて、任意の次数分布を持つネットワークを生成する事を今後の課題として挙げる。

また、本研究ではネットワーク構造特徴として、マクロスケール構造特徴のみを用い、ネットワークの全体としての特徴のみを扱った。しかし、ネットワークのより細かい部分に目を向け、ミクロスケール構造特徴を用いた分析も今後の課題として挙げる。

本データセットにもネットワークが存在しない領域がある事を示した。この領域は、**ANGA** によって生成できない領域であるか、ネットワーク構造として取り得ない領域であるかはわからない。また、ネットワーク構造として取り得ない領域の数学的な証明は既存研究においてもなされていない。従って、**ANGA** の拡張によって、生成できない領域にネットワークを生成する、または、その領域にネットワークが存在しない数学的証明を今後の課題としてあげる。

さらに、本ネットワークデータセットはノード数 1000、リンク数 10000 のネットワークを含んでいる。このサイズは十分に大きいサイズとは言えない。**ANGA** では、ノード数の大きいネットワークの生成を行うには膨大な計算時間を必要とする。この計算時間を削減し、ノード数を増やしたデータセットを構築する事も今後の課題の一つとする。

次に、本研究ではネットワークデータセットを用いて、現象と構造の分析を行った。本研究では例として2つの現象、情報拡散と協調行動を取り扱った。情報拡散モデルとし

て、IC モデルと LT モデルを取り扱い、協調行動として、囚人のジレンマを取り扱った。しかし、情報拡散や協調行動にも本研究で取り扱ったものとは別のモデルを用いた分析も必要とされる。情報拡散では、SIR モデルによるデマ情報 [4, 5, 6] の拡散等がある。協調行動では、複数プレイヤーによる囚人のジレンマゲーム [85] や、公共財ゲーム [86] 等が存在する。これらのモデルに対してもネットワーク構造と現象の関係について分析が必要であると考えられる。

また、本研究では、情報の拡散率や、協調の達成率といった現象の収束後の結果について分析を行った。しかし、情報拡散や協調行動を定義づける指標はこれらだけではない。例えば情報拡散であれば、拡散の速度や、ハブノードからの拡散率等が挙げられる。このような様々な指標に対しての分析も今後の課題とする。

分析をするにあたって、本紙で用いた 5 つの構造特徴では不十分な可能性がある。実際、協調行動の分析では、5 つの構造特徴以外にも協調行動の影響する特徴が存在する事が確認された。この 5 つの構造特徴以外の要素を発見し、これを含めたネットワーク構造特徴空間を埋めるようなネットワークデータセットへと拡張する必要がある。例えば、ネットワークのコミュニティ構造がある。コミュニティ構造は協調の達成に重要な役割を果たすと言われている [87, 89]。従って、任意のコミュニティ構造を持つデータセットが分析に必要となる。それに伴い、ANGA の拡張も必要になると考えられる。この ANGA の拡張と、ネットワークデータセットの修正、それを使った現象の分析を今後の課題とする。

特に、情報拡散の分析において、ある構造特徴が閾値を超えると急激に拡散率が低下する点があった。この急激な現象は今回扱っていない未知の構造特徴による変化である可能性が高い。この未知の構造特徴を発見し、分析する必要がある。

本研究では、情報拡散と協調行動を取り扱ったが、現実世界には様々な現象がネットワーク上で起こっている。本ネットワークデータセットが有用であるのはこの 2 つの現象の分析に限らない。本ネットワークデータセットは様々なネットワーク上で起こる現象を分析するために用いる事ができ、様々な現象の分析に貢献出来る事が期待される。例えば、意見形成、連鎖倒産、障害に対する頑健性等の分析に有用である。従って、本データセットを様々な現象に適用する事によって、現象とネットワーク構造に関する様々な知見を得る事を今後の課題とする。

本研究では静的なネットワークを使った分析を取り扱った。しかし、実際には、ネットワークは現象が発生している最中にも刻一刻と変化していく。例えば、情報拡散では、情報が発信されてから収束するまでに数日かかる。この数日の間にネットワーク構造は変化する事が考えられる。このような動的なネットワーク構造を扱う研究も多く存在してい

る [57, 90]. 本研究では, このような研究に対しても有用なデータセットを提供する事を今後の課題として挙げる.

最後に, 本研究で分析された最適なネットワーク構造を現実のネットワーク上に構築する手法の提案を今後の課題として挙げる. この方法としては, リンク推薦や, ユーザの行動の誘導等の方法が考えられる. どのようにリンクを推薦すれば最適なネットワークとなるかを分析する事を今後の課題とする.



# 謝辞

本研究は東京大学大学院工学系研究科システム創成学専攻における博士論文研究として行ったものです。論文作成にあたりましては、先生方をはじめ様々な方のご協力を賜りました。指導教員の鳥海先生には、学部4年生の卒論時から含めて6年間にわたってご指導賜りました。修士の2年間は遠隔指導という形で、遠方から熱心に指導していただきました。研究の仕方、発表の仕方、論文の書き方から私生活に至るまで、様々な姿勢や考え方の指導を手厚くしていただきました。先生と出会った当初、何もできなかった自分が、博士論文の執筆に至る事が出来たのは、先生の熱心な指導の賜物だと思います。先生と出会っていなければ博士課程への進学はなかったと断言できます。鳥海先生には深く感謝申し上げます。大橋先生には、博士課程の3年間に渡ってご指導いただきました。ミーティングの際に、常に鋭い視点から意見をいただき、自分の研究に関する思慮を深めるきっかけとなりました。先生の各研究に対する意見は、自分にはない角度の考え方でとても参考になりました。また、大橋先生の研究室にいたことは、私生活について考える良い機会となりました。大橋先生には深く感謝いたします。和泉先生、白山先生、栗原先生には、論文の審査にあたり大変お世話になりました。様々な議論をする事によって、論文を深める事につながりました。大変ご多忙にもかかわらず論文の副査を担当していただき、本当に有難うございます。また、修士2年間の指導を賜った名古屋大学情報科学研究科の間瀬先生、平山先生には、博士になってからも気にかけて頂き、学会等で声をかけていただいた事は大変感謝しております..

大橋・鳥海研究室に所属していた方々には大変お世話になりました。浅谷さんは博士課程で入った時に、先輩として様々なアドバイスをいただきました。浅谷さんとは、研究に関して多くの議論をし、研究を進めました。卒業した後も部屋に来て議論や、相談に乗っていただき本当に助かりました。浅谷さんには深く感謝しております。博士として他大学から入学してきた私を温かく迎え入れ、様々な議論をして同じ時間を過ごした、卒業生の小村君、福井さん、松澤君、間部君、小山君、馬場君、梶原君、本田君、山本君には深く感

謝いたします。研究室の備品管理等，様々な雑事を一緒にこなしてくれた河津君，常に丁寧で早い論文のチェックをしていただいた上子さん，研究や論文執筆の息抜きに文句一つ言わずに付き合ってくれた服部君，平野君，川崎君，山崎君，垣内君，園田君にも深く感謝いたします。また，研究室の **Fan** 君，前田さん，王さん，村上君，丸山君，村山さん，福馬君，謝君，山辺君，広瀬さんにも感謝を申し上げます。

最後に，研究生生活を支えていただいた，温かい友人達，家族にも深く感謝を申し上げます。

## 参考文献

- [1] 榎剛 史, 松尾 豊. ソーシャルセンサとしての twitter : ソーシャルセンサは物理センサを凌駕するか?(<特集>twitter とソーシャルメディア). 人工知能学会誌, Vol. 27, No. 1, pp. 67–74, 2012.
- [2] 杉山 あかし. 「第 1 回ネット選挙」-2013 年参議院選挙-の実証的研究. 比較社会文化 : 九州大学大学院比較社会文化学府紀要, Vol. 20, pp. 11–29, 2014.
- [3] 池田 和史, 服部 元, 松本 一則, 小野 智弘, 東野 輝夫. マーケット分析のための twitter 投稿者プロフィール推定手法. 情報処理学会論文誌コンシューマ・デバイス&システム (CDS) , Vol. 2, No. 1, pp. 82–93, 2012.
- [4] Laijun Zhao, Xiaoyan Qiu, Xiaoli Wang, and Jiajia Wang. Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks. *Physica A: Statistical Mechanics and its Applications*, Vol. 392, No. 4, pp. 987 – 994, 2013.
- [5] Yoshiyuki Okada, Keisuke Ikeda, Kousuke Shinoda, Fujio Toriumi, Takeshi Sakaki, Kazuhiro Kazama, Masayuki Numao, Itsuki Noda, and Satoshi Kurihara. Sir-extended information diffusion model of false rumor and its prevention strategy for twitter. *JACIII*, Vol. 18, No. 4, pp. 598–607, 2014.
- [6] Ke Wu, Song Yang, and Kenny Q. Zhu. False rumors detection on sina weibo by propagation structures. In *2015 IEEE 31st International Conference on Data Engineering*, pp. 651–662, 2015.
- [7] Frédéric Amblard and Guillaume Deffuant. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, Vol. 343, pp. 725 – 738, 2004.
- [8] 山口 真一. 実証分析による炎上の実態と炎上加担者属性の検証. 情報通信学会誌, Vol. 33, No. 2, pp. 53–65, 2015.
- [9] 平井 智尚. なぜウェブで炎上が発生するのか : 日本のウェブ文化を手がかりとして.

- 情報通信学会誌, Vol. 29, No. 4, pp. 61–71, 2012.
- [10] Erlend Nier, Jing Yang, Tanju Yorulmazer, and Amadeo Alentorn. Network models and financial stability. *Journal of Economic Dynamics and Control*, Vol. 31, No. 6, pp. 2033–2060, 2007.
- [11] Stefano Battiston, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald, and Joseph E. Stiglitz. Default cascades: When does risk diversification increase stability? *Journal of Financial Stability*, Vol. 8, No. 3, pp. 138 – 149, 2012.
- [12] Robert M. May and Nimalan Arinaminpathy. Systemic risk: the dynamics of model banking systems. *Journal of The Royal Society Interface*, Vol. 7, No. 46, pp. 823–838, 2010.
- [13] Patrick Wan-Hin Luk, Ga Ching Lui, and Kwok Yip Szeto. Optimization of systemic stability of directed network using genetic algorithm. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6, 2016.
- [14] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Comput. Netw.*, Vol. 33, No. 1-6, pp. 309–320, 2000.
- [15] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 103, No. 7, pp. 2015–2020, 2006.
- [16] Wilfred M. Post and Stuart L. Pimm. Community assembly and food web stability. *Mathematical Biosciences*, Vol. 64, No. 2, pp. 169 – 192, 1983.
- [17] Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. Why rumors spread so quickly in social networks. *Commun. ACM*, Vol. 55, No. 6, pp. 70–75, 2012.
- [18] Jie Zhou, Zonghua Liu, and Baowen Li. Influence of network structure on rumor propagation. *Physics Letters A*, Vol. 368, No. 6, pp. 458 – 463, 2007.
- [19] Damián H. Zanette. Critical behavior of propagation on small-world networks. *Phys. Rev. E*, Vol. 64, p. 050901, 2001.
- [20] Martin A. Nowak and Robert M. May. Evolutionary games and spatial chaos. *Nature*, Vol. 359, p. 826, 1992.
- [21] Zhihai Rong, Xiang Li, and Xiaofan Wang. Roles of mixing patterns in cooperation on a scale-free networked game. *Phys. Rev. E*, Vol. 76, p. 027101, 2007.
- [22] Salvatore Assenza, Jesús Gómez-Gardeñes, and Vito Latora. Enhancement of coopera-

- tion in highly clustered scale-free networks. *Phys. Rev. E*, Vol. 78, p. 017101, 2008.
- [23] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, Vol. 406, No. 6794, pp. 378–382, 2000.
- [24] Paolo Crucitti, Vito Latora, Massimo Marchiori, and Andrea Rapisarda. Error and attack tolerance of complex networks. *Physica A: Statistical Mechanics and its Applications*, Vol. 340, No. 1, pp. 388–394, 2004.
- [25] Mahdi Jalili. Error and attack tolerance of small-worldness in complex networks. *Journal of Informetrics*, Vol. 5, No. 3, pp. 422 – 430, 2011.
- [26] Pál Erdős and Alfréd Rényi. On random graphs. I *Publ. Math. Debrecen*, Vol. 6, No. 2, pp. 290–297, 1959.
- [27] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, No. 393, pp. 440–442, 1998.
- [28] Albert-Laszlo Barabási and Reka Albert. Emergence of scaling in random networks. *Science*, Vol. 286, No. 5439, pp. 509–512, 1999.
- [29] Yamir Moreno, Javier B. Gómez, and Amalio F. Pacheco. Epidemic incidence in correlated complex networks. *Phys. Rev. E*, Vol. 68, p. 035103, 2003.
- [30] Mark E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, Vol. 67, No. 2, p. 026126, 2003.
- [31] Sigal Sina, Noam Hazon, Avinatan Hassidim, and Sarit Kraus. Adapting the social network to affect elections. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS '15*, pp. 705–713, Richland, SC, 2015.
- [32] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. *Social Media Mining: An Introduction*. Cambridge University Press, New York, NY, USA, 2014.
- [33] Haifeng Zhang, Jie Zhang, Changsong Zhou, Michael Small, and Binghong Wang. Hub nodes inhibit the outbreak of epidemic under voluntary vaccination. *New Journal of Physics*, Vol. 12, No. 2, p. 023015, 2010.
- [34] Zi-Ke Zhang, Chu-Xu Zhang, Xiao-Pu Han, and Chuang Liu. Emergence of blind areas in information spreading. *PLoS ONE*, Vol. 9, No. 4, pp. 1–7, 2014.
- [35] Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics probability and computing*, Vol. 7, No. 3, pp. 295–305, 1998.
- [36] Fan Chung and Linyuan Lu. The average distances in random graphs with given ex-

- pected degrees. *Proceedings of the National Academy of Sciences*, Vol. 99, No. 25, pp. 15879–15882, 2002.
- [37] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Phys. Rev. Lett.*, Vol. 88, p. 128701, 2002.
- [38] Mark E. J. Newman, Steven. H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, Vol. 64, No. 2, p. 026118, 2001.
- [39] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, Vol. 87, pp. 925–979, 2015.
- [40] Alexei Vázquez. Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations. *Physical Review E*, Vol. 67, No. 15, pp. 056104, 2003.
- [41] Petter Holme and Beom J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, Vol. 65, No. 2, p. 026107, 2002.
- [42] Ginestra Bianconi and Albert-László Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, Vol. 54, pp. 436–442, 2001.
- [43] Ginestra Bianconi and Albert-László Barabási. Bose-einstein condensation in complex networks. *Phys. Rev. Lett.*, Vol. 86, pp. 5632–5635, 2001.
- [44] 鳥海 不二夫, 石田 健, 石井 健一郎. SNS におけるネットワーク成長モデルの提案 (コミュニティ形成支援,<特集>インターネット技術とその応用論文). 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 7, pp. 1135–1143, 2010.
- [45] Kikuo Yuta, Naoaki Ono, and Yoshi Fujiwara. A gap in the community-size distribution of a large-scale social networking site, 2007.
- [46] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, Vol. 40, No. 1, pp. 35–41, 1977.
- [47] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, Vol. 30, No. 1-7, pp. 107–117, 1998.
- [48] Mark E. J. Newman. Communities, modules and large-scale structure in networks, 2012.
- [49] Andrea Lancichinetti, Santo Fortunato, and Janos Kertesz. Detecting the overlapping and hierarchical community structure of complex networks, 2008.

- [50] Yong-Yeol Ahn, James P. Bagrow, and Sune Lehmann. Collective dynamics of 'small-world' networks. *Nature*, No. 466, pp. 761–764, 2010.
- [51] Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, Vol. 69, p. 026113, 2004.
- [52] Aaron Clauset, Mark E. J. Newman, and Christopher Moore. Finding community structure in very large networks. *Phys. Rev. E*, Vol. 70, p. 066111, 2004.
- [53] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, p. P10008, 2008.
- [54] William Ogilvy Kermack and Anderson Gray McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 115, No. 772, pp. 700–721, 1927.
- [55] Katsuya Nagata and Susumu Shirayama. Method of analyzing the influence of network structure on information diffusion. *Physica A: Statistical Mechanics and its Applications*, Vol. 391, No. 14, pp. 3783 – 3791, 2012.
- [56] Konstantin Klemm and Victor M. Eguiluz. Growing scale-free networks with small-world behavior. *Physical Review E*, Vol. 65, No. 5, p. 057102, 2002.
- [57] Zhou Yinzuo and Zhou Jie. Epidemic spreading in partial dynamic networks. In *Control Conference (CCC), 2012 31st Chinese*, pp. 1190–1194, 2012.
- [58] Jennifer Badham and Rob Stocker. The impact of network clustering and assortativity on epidemic behaviour. *Theoretical Population Biology*, Vol. 77, No. 1, pp. 71 – 75, 2010.
- [59] Monica Babes, Enrique Munoz de Cote, and Michael L. Littman. Social reward shaping in the prisoner's dilemma. In Lin Padgham, David C. Parkes, Jürgen P. Müller, and Simon Parsons, editors, *AAMAS (3)*, pp. 1389–1392. IFAAMAS, 2008.
- [60] Han The Anh, Luis Moniz Pereira, Francisco C. Santos, and Tom Lenaerts. Why is it so hard to say sorry? evolution of apology with commitment in the iterated prisoner's dilemma. In *in Proc. of 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, AAAI Press, in press, 2013.
- [61] Christian Hilbe, Arne Traulsen, and Karl Sigmund. Partners or rivals? strategies for the iterated prisoner's dilemma. *Games and Economic Behavior*, Vol. 92, pp. 41 – 52, 2015.
- [62] Koichi Moriyama. Utility based q-learning to facilitate cooperation in prisoner's

- dilemma games. *Web Intelligence and Agent Systems*, Vol. 7, No. 3, pp. 233–242, 2009.
- [63] Koichi Moriyama, Satoshi Kurihara, and Masayuki Numao. Evolving subjective utilities: Prisoner’s dilemma game examples. In Liz Sonenberg, Peter Stone, Kagan Tumer, and Pinar Yolum, editors, *AAMAS*, pp. 233–240. IFAAMAS, 2011.
- [64] Koichi Moriyama, Satoshi Kurihara, and Masayuki Numao. Cooperation-eliciting prisoner’s dilemma payoffs for reinforcement learning agents. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS ’14, pp. 1619–1620, Richland, SC, 2014.
- [65] Juan C. Burguillo-Rial. A memetic framework for describing and simulating spatial prisoner’s dilemma with coalition formation. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS ’09, pp. 441–448, Richland, SC, 2009.
- [66] Bijan Ranjbar-Sahraei, Haitham Bou Ammar, Daan Bloembergen, Karl Tuyls, and Gerhard Weiss. Evolution of cooperation in arbitrary complex networks. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS ’14, pp. 677–684, Richland, SC, 2014.
- [67] Masaki Tomochi. Defectors’ niches: prisoner’s dilemma game on disordered networks. *Social Networks*, Vol. 26, No. 4, pp. 309 – 321, 2004.
- [68] Guillermo Abramson and Marcelo Kuperman. Social games in a social network. *Phys. Rev. E*, Vol. 63, p. 030901, 2001.
- [69] Jesus Gómez-Gardeñes, Michel Campillo, Luis Mario Floría, and Yamir Moreno. Dynamical organization of cooperation in complex topologies. *Phys. Rev. Lett.*, Vol. 98, p. 108103, 2007.
- [70] Sergey N Dorogovtsev, José Fernando F Mendes, and Alexander N Samukhin. Structure of growing networks with preferential linking. *Physical review letters*, Vol. 85, No. 21, p. 4633, 2000.
- [71] Zhi Hai Rong and Xiang Li. The emergence of stable cooperators in heterogeneous networked systems. In *ISCAS*, pp. 2534–2537. IEEE, 2008.
- [72] Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, Vol. 88, pp. 128701+, 2002.
- [73] Ramon Xulvi-Brunet and Igor M. Sokolov. Reshuffling scale-free networks: From random to assortative. *Phys. Rev. E*, Vol. 70, p. 066102, 2004.



- [74] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E*, Vol. 67, p. 056104, 2003.
- [75] Julian J. McAuley and Jure Leskovec. Discovering social circles in ego networks. *CoRR*, Vol. abs/1210.8182, 2012.
- [76] 鶴見 裕之, 増田 純也, 中山 厚穂. 商品に関する twitter 上のコミュニケーションと販売実績の関連性分析 (<特集>声・質的データを経営・マーケティングに活かす). *オペレーションズ・リサーチ：経営の科学*, Vol. 58, No. 8, pp. 436–441, 2013.
- [77] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. Twitter under crisis: can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics, SOMA '10*, pp. 71–79, New York, NY, USA, 2010. ACM.
- [78] Mai Miyabe, Asako Miura, and Eiji Aramaki. Use trend analysis of twitter after the great east japan earthquake. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion, CSCW '12*, pp. 175–178, New York, NY, USA, 2012. ACM.
- [79] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '03*, pp. 137–146, New York, NY, USA, 2003. ACM.
- [80] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, Vol. 14, No. 3, pp. 199–222, 2004.
- [81] Scott Kirkpatrick, Daniel C. Gelatt, and Mario P. Vecchi. Optimization by simulated annealing. *SCIENCE*, Vol. 220, No. 4598, pp. 671–680, 1983.
- [82] Mark E. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, Vol. 89, No. 20, p. 208701, 2002.
- [83] Annett Nold. Heterogeneity in disease-transmission modeling. *Mathematical Biosciences*, Vol. 52, No. 3, pp. 227 – 240, 1980.
- [84] Marián Boguñá and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Phys. Rev. E*, Vol. 66, p. 047104, 2002.
- [85] Henry Hamburger. N-person prisoner’s dilemma. *The Journal of Mathematical Sociology*, Vol. 3, No. 1, pp. 27–48, 1973.
- [86] Chuang Lei, Te Wu, Jian-Yuan Jia, Rui Cong, and Long Wang. Heterogeneity of allocation promotes cooperation in public goods games. *Physica A: Statistical Mechanics and its Applications*, Vol. 389, No. 21, pp. 4708 – 4714, 2010.

- [87] Zhihai Rong, Zhi-Xi Wu, and Chi Kong Tse. Community structure promotes the emergence of persistence behavior in social networks. In *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*, pp. 898–901, 2015.
- [88] Xiaojie Chen, Feng Fu, and Long Wang. Prisoner’s dilemma on community networks. *Physica A: Statistical Mechanics and its Applications*, Vol. 378, No. 2, pp. 512 – 518, 2007.
- [89] Francisco C. Santos, Jorge M. Pacheco and Tom Lenaerts, Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 103, No.0, pp. 3490-3494, 2006.
- [90] 大澤 翔吾, 村田 剛志. 動的ネットワークにおける影響最大化. 人工知能学会論文誌. 論文特集「ネットワークが創発する知能」, Vol. 30, No. 6, pp. 693–702, 2015.