ネットワーク成長の連鎖的な構造変化に関する研究

2022年9月

稲福 和史

ネットワーク成長の連鎖的な構造変化に関する研究

筑波大学
人間総合科学学術院人間総合科学研究群
情報学学位プログラム

2022年9月

稲福 和史

ネットワーク成長の連鎖的な構造変化に関する研究
Study on cascading structural changes representing growth of complex networks

学籍番号：202030509
氏名：稲福 和史
Inafuku Kazufumi

ネットワークとはノード (点) とそれらの繋がりを表すエッジ (線) を基本要素とするデータ構造であり，現実世界のさまざまな繋がりを扱える．例えば人物をノード，二者間の関係をエッジとみなせば人間関係ネットワークが構築できる．また交差点をノード，道路をエッジとみなせば道路網ネットワークであるし，同様に駅と線路を対象とすれば鉄道網である．更には，ソーシャルネットワーキングサービス（SNS）のフォロー関係や Web ページのハイパーリンク，商品の共購買から論文の共著に至るまでその対象は多岐に渡る．近年，これらのネットワークを分析し，現実社会の課題を解決する研究に注目が集まっている．上述したネットワークを対象にした例では，友人コミュニティの中心的な人物や交通網における主要経路の検出，Web ページの評価や，オンラインショッピングサイトにおける商品推薦などが取り組まれている．

ここで，現実世界のネットワークの多くは時々刻々とノードとエッジが追加され，成長するネットワークである．例えば，SNS のフォローネットワークでは新たなフォロー関係が発生する度に，新たなエッジが追加される．新しくアカウントを作ったユーザーのはじめてのフォローであればノードも同時に出現する．また，交通網であれば道路の建設や新しい空路の就航，商品の共購買ネットワークであれば，新たな購買によってネットワークは成長する．このような常に変化を伴うネットワークを分析する際には，特定時刻のスナップショットを利用する，あるいは一定期間の変化を重畳するなど静的なネットワークへと変換する処理を行うことが多い．しかし，ネットワーク構造そのもの，ネットワーク特徴量，ノードの中心性指標などはネットワーク成長に伴い変化していく．このため，静的なネットワークへ変換するアプローチでは成長に伴う変化自体を捉えることはできない．

そこで，本研究ではネットワーク成長における構造変化そのものに着目する．そもそも，ネットワークは人物や駅などのノードと，ノード間の繋がりをエッジを用いて表現したものである．そして，ネットワーク構造の変化は，あるノードが他のノードへと何らかの影響を与え繋がりをもつこと，すなわちエッジが出現するということと同値である．（ここでは孤立ノードの出現は扱わず，新たなノードの出現は必ずエッジの出現を伴うものとする）．さらに出現したエッジの影響は，エッジ両端のノードだけに留まらない．影響を受けたノードは，影響を与えるノードとして他のノードへとその影響を伝播させ，新たな

エッジの出現を促す．この影響の伝播とエッジ出現の連鎖により，ネットワーク構造全体が変化していくのである．すなわち，ネットワーク成長は「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」の２つの要素から構成されるといえる．これら２つの要素を本研究における課題として捉え，それぞれに対する解決法を考究する．

第１の課題は，ネットワーク成長におけるエッジ影響力の定量化である．上述したように，ネットワークの成長過程では，１つのエッジの出現が他のノードに影響を与え，新たなエッジの出現を促す．この時，出現を促進したエッジの量を影響力として定量化する．高影響なエッジを抽出し，ネットワークの変化に貢献するエッジを明らかにすることは，ネットワーク成長を理解するために重要である．また，迅速な情報拡散，バイラルマーケティングの効果検証をはじめとして，様々な応用も期待できる．

第２の課題は，ネットワーク成長における構造推移の定量化である．ネットワークの成長過程では，同じ構造を持っていても，そこに至るまでの推移は多様であり，得られる知識や解釈もまた異なる．例えば，友人関係の構築過程において，顔の広い人物によって出会うのか，小さな仲良しグループが合流していく形を取るのか，といった違いが見られるだろう．逆に全く異なる構造を持っていても，そこに至るまでの推移，あるいは推移の一部には類似する特徴が見られるということもあるだろう．ネットワーク構造がどのように変化していくのか定量的に評価することは，異なる構造推移の類似度計算を始めとしてネットワーク成長の理解・議論に欠かせない．

本研究の目的は，以上述べてきた２つの課題を解くことでネットワーク成長全体を定量的に評価することである．

本論文の構成は次の通りである．まず**第１章**で，本研究の背景と課題設定について説明し，ネットワーク成長の理解のためには「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」の２点を明らかにする必要があることを説明する．また，それぞれの課題についてのアプローチとして，１点目の課題は情報拡散モデル及び情報カスケードの考え方を用いてエッジの影響力を定量化すること，２点目の課題はトライアド（３ノードからなる有向ネットワーク）の推移パターンを用いてネットワークの構造推移を定量化することを説明している．

第２章では，本研究の関連研究を示す．まず，複雑ネットワークの評価指標・中心性指標について整理する．静的ネットワークに対する基本的な手法について述べた後，大規模なネットワークや動的ネットワークにアプローチについて説明する．次に，情報拡散ネットワークに関連する研究について示す．大きく，情報拡散モデルに関する研究，影響力の定量化に関する研究，情報拡散の分析・予測に関する研究の３つに分けて整理する．続いて，トライアドに基づくネットワーク分析研究について整理する．まず，トライアドを始めとしたネット

ワークを低次元のベクトルとして扱う関連研究を示す．続いてトライアドの推移に着目した研究について述べる．最後に，情報拡散及びトライアドとは異なる観点からの動的ネットワーク分析を示す．

第 3 章では，ネットワークの成長過程におけるエッジ影響力を定量化する指標:STM(Stimulation Index) を提案する．ネットワーク成長の最小単位はエッジの出現である．同時にあるエッジの出現は接続先のノードに影響を与え，さらに新たなエッジの出現を促す．すなわち，ネットワークの成長とはエッジ出現の連鎖であるといえる．この時，エッジ単位の影響力を定量化できれば，ネットワーク成長に貢献する重要なエッジの検出が可能になる．

本研究では，情報カスケード及び情報拡散モデルの 1 種である線形閾値モデルの考え方を用いて，あるエッジ生成が契機となって生まれた後続するエッジの量をエッジの影響力とみなし定量化を試みる．SNS における情報伝達のネットワークを対象として，提案手法の有効性と限界を明らかにする．

第 4 章では，トライアド推移パターンに基づくネットワーク構造変化の分析手法を提案する．動的ネットワークにおいては最終的に類似する構造を持っていても，その過程は異なることが当然起こりうる．本研究では，トライアド（連結 3 ノードからなる最小の有向ネットワーク．13 種類存在する．）の推移を用いて，動的な構造変化を定量化する．ネットワーク中のトライアド 13 種の分布を調べることで，その構造的特徴を把握することができる．またネットワークが変化すると，当然トライアド自体も別種のトライアドへと推移する．このトライアドの推移パターンは 28 種存在する．これを数え上げることによってネットワークの構造推移を分析する．商品の購買順序関係を示すネットワーク:PHG（Purchase History Graph）を分析対象として，提案手法の有効性と限界を明らかにする．

第 5 章では，「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」の 2 つの課題に対する提案手法の効果について考察を行う．また，提案手法が有効に機能する対象や，提案手法の限界についても議論する．

最後に**第 6 章**で本研究についてまとめている．情報化社会の発展や計算機の性能向上により，近年ではますます大規模なデータが収集・利活用されるようになった．これに伴い，ネットワーク分析の分野においてもその動的な性質そのものを捉える手法が必要とされている．本研究では，ネットワーク成長における構造変化そのものを分析すべく，エッジ影響力の定量化指標として Stimulation Index 及びトライアド推移パターンに基づく構造変化の分析に取り組んだ．結果，前者によりネットワーク成長における高影響エッジの抽出が可能なことを明らかにした．また，後者では商品カテゴリ固有のトライアド推移パターンが抽出できたほか，逆に様々な推移が混在するカテゴリの存在などを明らかにし，提案手法の有効性を示した．

本研究の貢献は連鎖的に構造変化するネットワークにおいてネットワーク全体の一連の変化における個別の変化の影響力及び，ネットワーク全体がどのよう

に構造変化をしたのかを明らかにする手法の提案である．本研究で提案した2手法を使い分ける，あるいは併用することによって，ネットワーク成長における連鎖的な構造変化の分析において有用な結果を得られると期待できる今後の課題として，拡大・縮小の混在するネットワークや連続的な時間軸を持つネットワークに対する拡張が挙げられる．本研究では，離散的な時間軸かつ単調拡大するネットワークを対象としている．提案手法を拡張することで，より現実的なネットワーク成長の分析が可能になると期待できる．

Study on cascading structural changes representing growth of complex networks

Student id : 202030509
Name : Inafuku Kazufumi

A network is a data structure composed of nodes and edges that can represent various connections in the real world. For example, friendship networks are composed of people as nodes and friendships as edges. Similarly, railroads are networks with stations as nodes and tracks as edges. The targets of the networks are various, following on social networking services (SNS), hyperlinks on Web pages, co-purchases of products, and co-authorship of articles. In recent years, there is a lot of research on solving real-world problems by analyzing networks; detecting a key person in communities, evaluating web pages, and recommendation system of e-commerce sites. Many real-world networks are constantly growing, adding nodes and edges. In a follow network of SNS, a new edge appears every time a new follow occurs. When someone creates a new account, a new node appears. Transportation networks grow through the construction of roads and the opening of new air routes, and co-purchasing networks grow through new purchases. When analyzing, such dynamic networks are often transformed into static networks; using snapshots at specific times, accumulate growth over some time. However, network structure, network features, and node centrality change with network growth. Therefore, the approach of converting a dynamic network to a static network cannot deal with changes with network growth.

In this study, we focus on the structural change in network growth. The smallest unit of change in a network is the appearance of edges. Furthermore, new edges stimulate the appearance of new edges. In other words, the network grows through a chain of edge appearances.

This study sets two research tasks. The first task is to quantify edge influence in network growth. As mentioned above, in the process of network growth, the appearance of an edge influences other nodes and stimulates the appearance of new edges. The edge influence is quantified as the number of edges that stimulate the appearance of new edges. Identifying high influence edges that contribute to network change is important for understanding network growth. It is also expected to have a variety of applications, including rapid information diffusion and effectiveness verification of viral marketing.

The second task is the quantification of structural transitions in network growth. In the process of network growth, even if the structure is the same, there are various transitions leading up to that state. For example, there is a significant difference in the process of forming friendships, whether it is by a central person or a small group merging.
Our goal is to quantify network growth by solving these two research tasks. The paper is organized as follows. **Chapter 1** describes the background of this study and the problem set. We explain that to understand network growth, it is necessary to clarify two points: "the quantification of edge influence in network growth" and "the quantification of structural transitions in network growth.
Chapter 2 describes related works. First, we summarize evaluation methods and centrality metrics for complex networks. We describe basic methods for static networks and approaches for large and dynamic networks. Next, we divide information diffusion network research into three categories: information diffusion models, quantification of influence, and analysis and prediction of information diffusion. Next, we will introduce network analysis based on triads. Studies that treat the network as a low-dimensional vector and studies that focus on the triad transition are described. Finally, dynamic network analysis, such as information diffusion, is explained.
Chapter 3, we propose an index to quantify edge influence in the network growth process: the Stimulation Index (STM). The smallest unit of network growth is the appearance of edges. The appearance of an edge simultaneously influences the nodes to which it is connected, and stimulates the appearance of new edges. In other words, network growth is the result of a chain of edge appearances. By quantifying the influence of each edge, it is possible to detect important edges in the network growth. Based on the linear threshold model, which is a type of information cascade and information diffusion model, this study attempts to quantify the number of subsequent edges as influences. An evaluation experiment is conducted using a network of information transmission in SNS to show the validity and limitations of the proposed method.
Chapter 4, we propose a method for analyzing network structure change based on triad transition patterns. In dynamic networks, even if the structures are similar the last time, the formation processes are various. In this study, we propose a method to quantify dynamic structural changes using triad transitions. By counting the distribution of the 13 triads in the network, we can understand their structural features. When the network changes, the triad also transitions to a different type of triad, there are 28 different patterns of triad transition. By counting up these transition

patterns, we analyze the structural transition of the network. We analyze a network of purchase order relationships: PHG (Purchase History Graph) to clarify the validity and limitations of the proposed method.
Chapter 5, we discuss the results of the evaluation experiments in Chapters 3 and 4. We also discuss the networks on which the proposed method works effectively and the limitations of the proposed method.
Finally, **Chapter 6** summarizes this research. With the development of the information society and the improvement of computer performance, large-scale data has been collected and utilized more and more in recent years. In the field of network analysis, there is a need for a method to capture the dynamic features of the network. In this study, we proposed a method for analyzing structural changes in network growth using the Stimulation Index and the Triad Transition Pattern. We conducted evaluation experiments and extracted influential edges in network growth using STM. The validity of the proposed method was shown by extracting product category-specific growth trends based on triad transition patterns.

目次

図目次

表目次

第 1 章

序論

1.1 研究背景

近年，現実世界の様々な関係性・事象をネットワークで表現し，ノード関係や主要構造，ネットワークの変化，ネットワーク上での現象を分析する研究が盛んに取り組まれている．ネットワークはノード (点) とそれらの繋がりを表すエッジ (線) を基本要素とするデータ構造であり，現実世界のさまざまな繋がりを表現できる．社会関係を例にとると，人をノード，人同士の友人関係に従いノードをエッジで繋ぐことで，友人関係ネットワークを構築できる．これは現実の友人関係のみならず，Twitter や Facebook，Instagram などの SNS にも拡張可能である．アカウントをノード，アカウント同士のフォロー関係をエッジに反映することで，SNS 上での繋がりを表現するフォローネットワークとなる．また，道路網や線路網などの物理関係や食物連鎖や商品の共購買といった事象をはじめとして，様々な分野の関係性・事象をネットワークとして表現することが可能である．近年では計算機の性能向上により，大規模なネットワークを扱えるようになったことからそれら分析して様々な知見を得る研究が盛んである．

また，現実世界のネットワークの多くは，時々刻々とノードとエッジが追加される動的なネットワークである．例えば SNS のフォローネットワークにおいては，新しいアカウントが生まれる度にノードが，新しいフォローが発生する度にエッジが増える．上述した道路網であれば道路の建設，共購買ネットワークでは新商品の発売や新たな購買によってネットワークは成長する．そのため，現実のネットワークを分析する際には，特定時刻のスナップショットの取得や一定期間の変化を重畳するなどの処理により，着目している特徴量や特徴量の変化が顕在化するように動的ネットワークの時間軸を縮退して静的ネットワークとして扱うことが多い．しかし，このアプローチではネットワークの特徴が本来とは異なるものに見えたり，一部のノードの重要度が過大あるいは過小評価されることが起こりうる．よって，動的ネットワークを分析する上ではその構造変化を踏まえた手法が必要である．

動的ネットワーク研究では，ネットワークの特徴量やリンク構造の推移を分析するもの異常値検出や変化点検出などの大きな変化を検出するもの，将来のノード属性やエッジ構造などを予測するものの大きく 3 つが盛んに取り組まれている．リンク構造の推移を分析する研究として，中田らは友人関係ネットワークの成長過程に着目し，どのような順序で友人を作るとより多くの友人を得られるのかを明らかにしている [1]．変化点検出の研究では，Peel らがネットワーク中のコミュニティを対象に変化点検出を行っている [2]．ネットワーク中にはコミュニティが存在し，またそれらは階層的であることが知られている．ここでは hierarchical random graph model [3] を用いて，コミュニティの階層レベルが増減するタイミングの検出に取り組んでいる．また予測研究では，Nigam らによるノードの意見変化予測 [4] や Hanjun らによるレコメンデーションのためのエッジ出現予測 [5] などが知られている．近年ではネットワークが大規模になるにつれその変化の大きさ・速度も大幅に上がっていることから，上述したようなネットワークの動的な性質を捉える手法に注目が集まっている．

1.2　ネットワークの成長

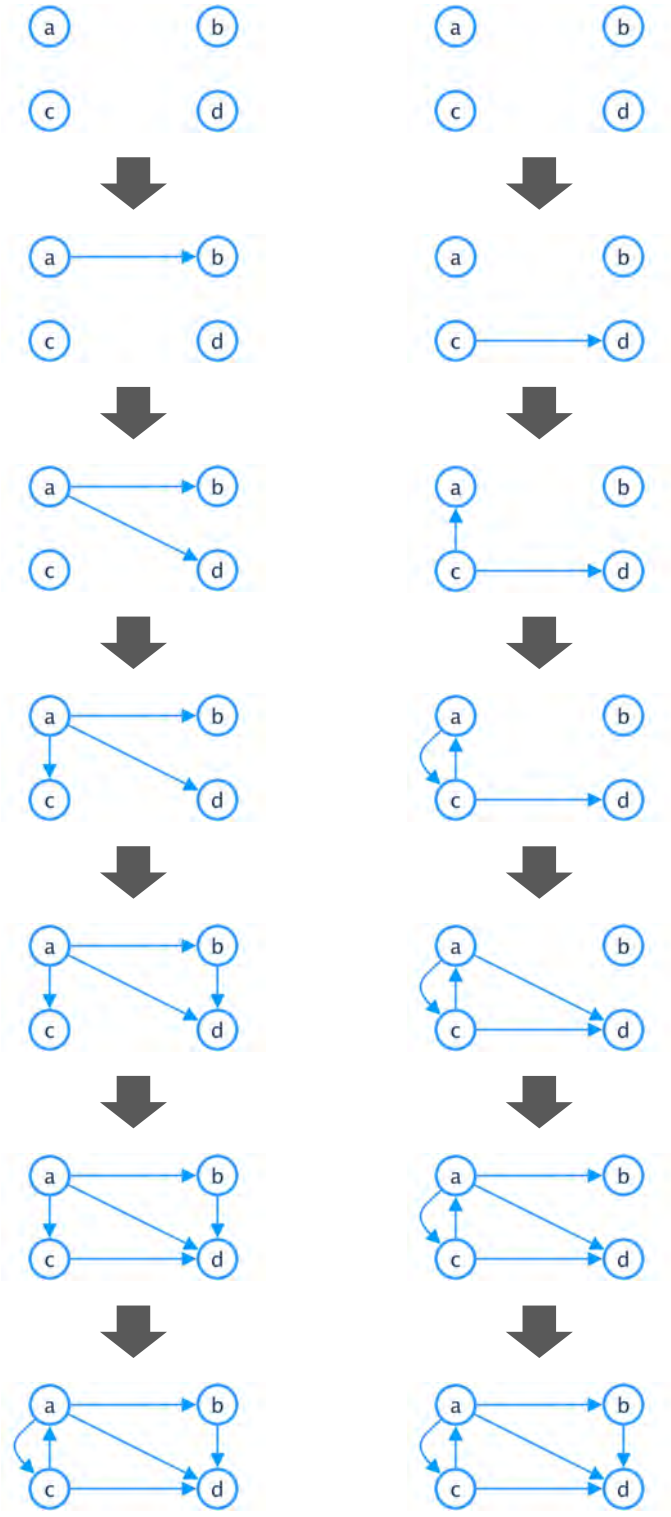
本研究で取り扱うネットワークの成長について説明する．現実のネットワークの多くはその構造を常に変化させ成長するネットワークである．このネットワーク成長の捉え方は，ネットワークの拡大・縮小の扱い方及び構造変化の時間軸により 4 つに大別される．前者について，まずネットワークにはエッジやノードが出現する「拡大」と逆に消失する「縮退」の 2 つが存在する．そして，拡大・縮退のいずれかが発生する単調モデルと，拡大・縮退のいずれもが発生する混合モデルの 2 つに分けられる．後者の構造変化の発生する時間軸について，ステップ単位の変化として離散的に扱うものと，実時間と同様に連続的に扱うものの 2 つに分けられる．例えば，SNS のフォローネットワークにおいて，新たなフォローの発生は「拡大」であり，逆にフォローの解除は「縮退」にあたる．フォローの発生だけに着目した場合は単調 (拡大) モデルであり，フォローの発生・解除を同時に扱う場合は混合モデルである．また時間軸についても，フォロー・アンフォローの発生順序だけに着目した場合は離散的，実際の発生時刻に着目した場合は連続的に扱うケースに当たる．本研究で扱うネットワーク成長は，これらのうち離散的な時間軸におけるネットワークの単調 (拡大) モデルとする．このため，ネットワーク成長における最小ステップは，エッジの出現及びこれに伴うノード出現とする．エッジ出現を伴わないノード，すなわち孤立ノードはその出現時点において，ネットワーク成長に寄与しないことからネットワークに含めないものとして扱う．孤立ノードは他のノードと接続した時刻においてネットワークに出現したものとする．

本研究で対象とするネットワーク成長は，「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワー

ク構造全体がどのように変化していくか」の 2 つの要素から構成される．これら 2 つの要素を本研究における課題として捉え，それぞれに対する解決法を提案した．第一の課題，ネットワーク構造全体の変化における個別の変化の影響力は，STM によって明らかにできる．第二の課題，ネットワークがどのように構造変化したのかはトライアド推移を用いて明らかにできる．両手法を使い分ける，あるいは併用することによって，ネットワーク成長における連鎖的な構造変化の分析において有用な結果を得られると期待できる．

上述したように，動的なネットワークは時々刻々とノードとエッジを増やしその構造を変化させる．よって特定の時刻に同じ構造であったとしても，そこに至るまでの成長過程が異なれば，得られる知識や読み取れる事柄も変わる．例として，**図 1.1** に成長過程の異なる 2 つの動的ネットワークを示す．初期状態はノードのみとし，逐次的に 1 本ずつエッジが出現するケースを考える．**図 1.1(a)** 及び**図 1.1(b)** において，初期状態及び最終ステップにおける構造は等しいが，その成長過程は大きく異なっている．これをユーザーをノード，ユーザー間の情報伝達をエッジとする情報拡散のネットワークとして捉えると，**図 1.1(a)** からは，ユーザー a が情報源であり，かつユーザー b, c, d の 3 名にダイレクトに情報を伝えていることから，ユーザー a による直接的な情報拡散であることが読み取れる．一方で，**図 1.1(b)** からは，情報源自体はユーザ c である一方で，直接伝達するのはユーザー a, d の 2 名にとどまっており，ユーザー a の中継によってユーザー b へ情報が伝わっていることが読み取れる．すなわち，こちらのパターンは**図 1.1(a)** に比べるとユーザー a を介した間接的な情報拡散であるといえる．このように読み取れる事柄が大きく異なることから，ネットワーク成長の過程を定量的に評価・分析することの意義は大きい．

ここで，そもそもネットワークは，何らかのオブジェクトに対応するノードと，そのノード間の繋がりを表現したものである．ネットワーク成長の最小単位とは，あるノードが他のノードへと影響を与えた結果繋がりを持つこと，つまりエッジの出現だといえる．そしてこの時，影響の範囲はエッジが出現した 2 ノード間に留まらない．影響を受けたノードは，さらに他のノードへとその影響を伝播させる．このステップを繰り返すことにより，ネットワーク全体の構造が変化していく．すなわち，ネットワーク成長は「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」の 2 つの要素から構成されるといえる．これら 2 つの要素を本研究における課題として捉え，それぞれに対する解決法を考究する．



(a) ネットワーク成長の過程:パターン 1 (b) ネットワーク成長の過程:パターン 1

図 1.1: ネットワーク成長

1.3　研究目的とアプローチ

第 1.2 節で述べたネットワーク成長を構成する 2 つの要素:「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」を踏まえて，本研究では大きく 2 つの課題に取り組む．

1 つ目はネットワークの成長過程におけるエッジの影響力の定量化である．動的ネットワークでは 1 つのエッジの出現が他のノードに影響を与え，新たなエッジの出現を促す．この影響力を定量化してネットワークの活性化に貢献するエッジ出現を明らかにすることで，ネットワーク成長を理解する．また，迅速な情報拡散やバイラルマーケティングの効果検証をはじめとして，様々な応用も期待できる．

2 つ目はネットワーク成長における構造推移の定量化である．**図 1.1** で示したように，同じ構造を持っていても，そこに至るまでの推移は多様であり，得られる知識や解釈もまた異なる．逆に全く異なるような構造を持っていても，その推移自体には類似する特徴が見られるということもあるだろう．ネットワーク構造がどのように変化していくのか定量的に評価することで，異なる構造推移の類似度計算を始めとしたネットワーク成長の理解・議論を深める．

本研究の目的は，以上述べてきた 2 つの課題を解くことでネットワーク成長全体を定量的に評価することである．具体的なアプローチとして，1 つ目の課題については，人伝てにうわさが広がっていく様子をモデル化した情報拡散モデルを利用し，ネットワークの成長過程におけるエッジの影響力の定量化を試みる．また 2 つ目の課題は，3 ノードからなる最小のネットワークであるトライアドの分布及びその推移に着目して，ネットワーク成長における構造推移を分析する手法を提案する．

1.4　本論文の構成

本論文の構成は次のとおりである．**第 1 章**の序論に続き，**第 2 章**では，本研究の関連研究を示す．まず，複雑ネットワークの評価指標・中心性指標について整理する．静的ネットワークに対する基本的な手法について述べた後，大規模なネットワークや動的ネットワークに適用するためのアプローチについて説明する．次に，情報拡散ネットワークに関連する研究について示す．大きく，情報拡散モデルに関する研究，影響力の定量化に関する研究，情報拡散の分析・予測に関する研究の 3 つに分けて整理する．続いて，トライアドに基づくネットワーク分析研究について整理する．まず，トライアドを始めとしたネットワークを低次元のベクトルとして扱う関連研究を示す．続いてトライアドの推移に

着目した研究について述べる．最後に，情報拡散及びトライアドとは異なる観点からの動的ネットワーク分析を示す．

第 3 章では，ネットワークの成長過程におけるエッジの影響力の定量化する Stimulation Index を提案する．ネットワーク成長の最小単位はエッジの出現である．同時にあるエッジの出現は接続先のノードに影響を与え，さらに新たなエッジの出現を促す．すなわち，ネットワークの成長とはエッジ出現の連鎖であるといえる．この時，エッジ単位の影響力を知ることで，一連のネットワーク変化における重要なエッジの検出が可能になる．本研究では，あるエッジが起点となって生まれた後続するエッジの量を，当該エッジが保持する影響力とみなして定量化を試みる．SNS における情報伝達を示すネットワークを対象として，提案手法の有効性と限界を明らかにする．

第 4 章では，トライアド推移パターンに基づくネットワーク成長における構造推移の定量化手法を提案する．ネットワーク成長においては最終的に類似する構造を持っていても，その過程はネットワークやノードの性質によって異なることが考えられる．本研究では，トライアド[*1]の推移を用いて，動的な構造変化を定量化する．ネットワーク中のトライアド 13 種の分布を調べることで，その構造的特徴を把握することができる．またネットワークが変化すると，当然トライアド自体も別種のトライアドへと推移する．このトライアドの推移パターンは 28 種存在する．これを数え上げることによってネットワークの構造推移を分析する．商品の購買順序関係を示すネットワーク PHG（Purchase History Graph）を分析対象として，提案手法の有効性を評価する．

第 5 章では，提案手法の到達点と限界について議論する．ネットワークの成長は，エッジやノードが出現する「拡大」と逆に消失する「縮退」の 2 つの変化に分けられる．また構造変化の発生する時間軸も，ステップ単位の変化として離散的に扱うものと，実時間と同様に連続的に扱うものの 2 つに分けられる．本研究においてはこのうち離散的な時間軸における拡大するネットワークの成長について扱っているので，今後の課題としてネットワークの「縮退」や連続的な実時間への拡張について議論する．

最後に**第 6 章**で本研究についてまとめる．

[*1] 連結 3 ノードからなる最小の有向ネットワーク．13 種類存在する．

第 2 章

関連研究

複雑ネットワークの連鎖的な成長に着目した本研究の関連研究について，複雑ネットワークの定量評価に関する研究，情報拡散ネットワークに関する研究，局所構造を用いたネットワーク変化に関する研究，動的ネットワーク分析に関する研究の 4 つに分けて整理する．最後に本研究の位置づけについて述べる．

2.1 複雑ネットワークの定量評価に関する研究

ネットワーク全体の性質を評価するため様々な特徴量が提案されている．代表的な特徴量としては，ネットワークのスモールワールド性の議論で知られる平均ノード間距離 [6, 7]，友人の友人は友人であるか，という 3 ノードの連結度合いに着目した平均クラスター係数 [8]，次数相関や次数分布のべき指数 [9]，モチーフパターン [10] などが挙げられる．また，ネットワークのノードをランキングして要となるノードを知ることも重要な課題である．ネットワーク構造に基づきノードの重要性を評価する様々な中心性指標が提案されている．接続しているノード数に基づく次数中心性や他のノードと平均的にどの程度近いかを表す近接中心性 [11, 12]，ノード間を橋渡しする役割の強さを表す媒介中心性 [12, 13]，中心に近い人もまた中心に近いとする固有ベクトル中心性や Katz 中心性 [14–16] などが知られている．インターネット中の重要度の高い Web ページを評価する PageRank [17, 18] や HITS [19] なども応用の幅広さからよく知られている．さらにこれらを近年の大規模なネットワークに適用すべく，高速化や近似解を求める手法も数多く提案されている [20–23].

また, 動的ネットワークに対しても上述した手法の拡張が提案されている [24–26]．林らはエッジの挿入と削除の両方を扱える媒介中心性の近似値計算に取り組んでいる [25]．媒介中心性のスコアを効率よく更新するためのデータ構造とアルゴリズムを提案し，実験によって高いスケーラビリティを持つことを示している．また，Magnien らは動的ネットワークにおける近接中心性の推移に着目している [27]．電子メール及び観光行動の 2 つのネットワークを用いて実験を行い，近接中心性が時間経過と共に大きく変化することを明らかにした．特

定のスナップショットだけでなく時間経過を考慮した手法の必要性を定量的に示した研究として知られている．

2.2　情報拡散ネットワークに関する研究

本研究において，動的ネットワークの構造推移における影響力に関する提案手法は，情報拡散現象がベースになっている．ここでは，情報拡散ネットワークに関する研究を，情報拡散モデル，影響力の定量化，情報拡散現象の予測・分析の 3 つの視点から整理する．

2.2.1　情報拡散モデルに関する研究

ある人物から発信された情報が社会全体へと拡がってゆく過程を明らかにすることは，情報拡散の性質を理解する上で重要な課題である．人から人へと情報が伝わる時，情報は二者の間でやり取りされるだけでなく，情報の受け手が新たな人へと伝達することでより広く拡散する．この情報伝達の連鎖現象は情報カスケード（Information Cascade）[28] と呼ばれ，情報拡散の最も基本的なプロセスである．人をノード，知人関係をエッジとする社会ネットワークを用いた情報拡散研究では，この情報カスケードに基づくモデルが広く利用されている．中でも独立カスケードモデル [28] と線形閾値モデル [29,30] が主要なモデルである．

独立カスケードモデルは送信者中心型のモデルである．まず，ネットワーク中のすべてのエッジについて，それぞれ情報伝達の成否確率を設定する．次に，情報源となるノードから隣接するノードに対して情報の伝達を試みる．この時，情報伝達の成否ははじめにエッジごとに設定した確率に従って，それぞれが独立に決定される．伝達に成功し情報を受け取ったノードは，次のステップにおいて送信者ノードとなり，同様に隣接するノードへと情報伝達を試みる．これを送信者ノードがいなくなるまで繰り返すことで，情報拡散をシミュレートするのが独立カスケードモデルである．

一方，線形閾値モデルは情報の受け手側のノードを中心に情報拡散をシミュレートする．まず，ネットワーク中の各ノードに重みの閾値を割り当てる．次に，情報源となるノードを選択し隣接するノードに対し情報の伝達を行う．この時，独立カスケードモデルと異なり情報伝達自体は必ず成功し，送信者ノードから受信者ノードへと一定の重みが渡される．受信者ノードは受け取った重みの和が閾値を超えた場合，次のステップにおいて送信者ノードへと変化し隣接するノードへ情報を送信する．こちらも送信者ノードがいなくなるまで繰り返すことで，情報拡散をシミュレートする．

近年ではこれらのモデルを用いた情報拡散の予測や大規模ネットワークへ

の適用を始めとした様々な拡張が盛んである．独立カスケードモデルについて，Saito らは EM アルゴリズムに基づく情報伝播確率の予測に取り組んでいる [31]．また，大規模なソーシャルネットワークにおける情報拡散分析も取り組まれている [32–34]．さらに Chen らは線形閾値モデルに基づく影響最大化問題を高速に解く手法を提案している [35]．有向非巡回グラフ（DAG）を用いてネットワークサイズに対して線形時間で計算可能であり，既存手法よりも高速に動作することを示した．また，Berbier らは情報の内容・トピックに着目し，独立カスケードモデル及び線形閾値モデルを拡張している [36]．ソーシャルネットワークにおける情報伝播はトピックに関連しているとし，ネットワーク上のトピック分布を示している．

2.2.2　影響力の定量化に関する研究

ソーシャルネットワーク研究において，情報拡散に対する影響力の定量化は重要な研究課題である．多くの研究では，オピニオンリーダーやインフルエンサーと呼ばれる影響力の高いユーザーの検出に焦点が当てられている．例えば，SNS などで有名ユーザーに自社商品を紹介してもらうプロモーションでは商品紹介を行うユーザーの影響力が広告効果に直結する．そのため，なるべく広く情報を拡散できるユーザーを発見したいという要求がある．このような問題は，ネットワーク中から情報拡散を最大化するノードを発見する影響最大化問題と呼ばれ盛んに研究されている [37–40]．また，実世界のネットワークはその構造を常に変化させている．SNS において新規ユーザーの出現は新しいノードの出現であり，新しくユーザーをフォローすれば新しいエッジの出現となる．このように常に変化するネットワークにおいて，影響最大化問題を解くための手法も提案されている [41, 42].

ネットワーク構造を用いてノードの影響力を定量化する試みは数多く取り組まれている [43–48]. その中でも素朴な手法としては中心性指標の利用が挙げられる．しかし，対象ノードとネットワーク全体の構造から算出される媒介中心性や近接中心性といったグローバルな指標は高影響ノードの検出に有効であるが計算量が多くナイーブに適用することは難しい．一方，対象ノードとそれを取り巻く局所的なネットワーク構造に基づく，次数中心性などのローカルな指標は計算量が少ないものの，影響力の大きいノードの検出には有効でない．この問題を解決するため，Gao らはグローバル指標とローカル指標の特性を組み合わせた local structural centrality を提案している [46]．伝染病の感染拡大・治癒状態を示す SIR モデルを用いてシミュレーションを行い，提案手法がノードの影響力を定量化する指標として有効であることを示している．

さらに実際の情報拡散過程を考慮すると，広範囲に情報を広げることができるかはもちろん，短い時間で情報を拡散できるかも重要である．Ullah らは，SNS のユーザインタラクションのネットワークを対象として，ネットワーク構

造に加え時系列の特徴量を考慮した情報拡散モデル及び高影響ノードの抽出手法を提案している [49]．より直近に交流したユーザから受ける影響が大きくなるように重み付けを行うことで，情報拡散の最大化と拡散時間の最小化を両立する高影響ノードを発見できるとしている．

また，ネットワーク構造を用いた手法の多くは，ネットワーク上の特定の構造に位置しているユーザーを高影響だとみなす手法である．しかし，実際にはユーザーの属性や役割，また同じ友人でも仲が良い人と疎遠な人というような関係性が存在することから，これらを考慮した手法が提案されている [50]．Amir らはユーザー間のインタラクション量からユーザーの役割を推定し影響力を決定する手法を提案している [51]．Twitter のデータを用いてシミュレーションを行い，SNS における高影響ユーザーの位置が従来手法よりも広範に分布することや，情報の拡散範囲，拡散速度の 2 点において優れていることを示した．

2.2.3　情報拡散の分析・予測にする研究

情報拡散の分析では，実際の SNS を対象とした研究が盛んである [45,52,53]．特に SNS ならではの特徴として，ネットワーク構造やインタラクション情報に加え，情報の内容に着目した研究がある [54–56]．Stieglitz らは情報に付随する感情に着目し，ソーシャルメディア上のコンテンツに付随する感情とユーザの情報共有行動の関係性を検証した [55]．Twitter における政治的なツイートを分析し，感情的なツイートは中立的なツイートに比べ，高頻度かつ短時間でリツイートされる，すなわち拡散されやすいことを示した．

また，ソーシャルネットワーク上では常に様々な種類の情報拡散が同時に発生している．その中でも災害情報や風評を始めとした，現実社会において重要度の高い情報拡散を早期に発見できれば，迅速な避難やデマ情報の訂正に活用できる．Kawamoto らは社会的に重要度の高い情報カスケードを定義しこれの早期検知に取り組んだ [56]．Twitter の実データを対象に，ツイートのテキスト特徴や情報拡散に参加したユーザー情報などを基に特徴量設計を行い，機械学習によって社会的影響力と情報拡散範囲を予測・分類している．

このような情報拡散の予測もまた，活発な分野である [57]．Facebook における写真シェアの規模予測 [58] や Twitter におけるリツイート規模の予測 [59] などが知られている．これらの研究では，予測精度を向上させるために，ネットワーク構造，時間的特徴量，情報の内容に基づく特徴量を用いた機械学習アプローチを採用している．

2.3　局所構造を用いたネットワーク変化に関する研究

本研究では動的ネットワークの構造推移をトライアドとその推移を用いて表現する．そのためここでは，局所構造を用いてネットワークをベクトルとして扱う研究，トライアドを用いた構造変化の分析の 2 つについて整理する．

2.3.1　ネットワークのベクトル表現

トライアド分布のように，ネットワーク構造を低次元のベクトルで表現する研究は多く存在する．大きく，ネットワーク内のサブグラフ構造の出現頻度により表現する手法 [60–62]，グラフスペクトルなどの固有ベクトルを用いる手法 [63]，グラフ間の距離からベクトルを学習する手法 [64] に分けられる．本研究の提案手法はネットワーク内のサブグラフ構造の出現頻度を用いて表現する手法に分類される．中でも最小単位であるトライアドに着目した研究として，Salihoglu は無向グラフを 3 ノード及び 4 ノードのモチーフパターン出現頻度ベクトルで表現することで，道路ネットワーク，自律ルーティングネットワーク，共購買ネットワークの構造はトライアドパターン出現頻度ベクトルに反映されることを明らかにしている [61].

2.3.2　トライアドの構造変化分析

トライアドを用いて実世界の動的ネットワークを分析した研究として，伏見らの研究がある [65]．彼らは，Cookpad におけるユーザ間のフォロー数関係，Enron 社における社員間の Email 送受信関係，Twitter における Reply と Retweet 関係の有向グラフに対して，トライアドの推移パターンを抽出し，コミュニケーションの性質上，「片方向リンクから相互リンク」，「片方向リンクのまま」というパターン推移が有意に見られた点で，ソーシャルメディアによって構造が異なっていることを確認した．また，コミュニケーションの構造変化を分析する研究として，中田らの研究がある [1]．トライアドリンク構造の状態遷移図を用いて，友人関係ネットワークの成長過程を分析している．Tuomo は，トライアドの推移に着目し，対象とするパターンが多く出現するような人工グラフの生成モデルを提案している [66].

2.4　動的ネットワークの分析

本節では，トライアドパターンや情報拡散とは異なる観点からの動的ネットワーク研究について説明する．Leskovec らは動的ネットワークは時間経過とともに高密度化し直径が減少することを報告している [67]．時間経過とともに

ノードの出次数が増加し，ノード間の距離が徐々に短くなる結果として，ネットワーク自体の直径も縮まることを確認している．また，これらの現象を反映したネットワーク成長モデルも提案している．また，ネットワークの構造変化の度合いを捉える定量指標の研究も行われている．Albert らは，動的ネットワークにおいてノードやエッジの削除に対する頑健性を評価した [68]．高次数ノードと接続するエッジを削除し，ネットワークの平均クラスタ係数や直径がどのように変化するかを議論している．

異常値検出や変化点検出など，動的ネットワークにおける大きな変化を検出する研究も盛んである．ネットワーク中にはコミュニティが存在し，またそれらは階層的であることが知られている．Peel らは hierarchical random graph model [3] を用いて，コミュニティの階層レベルが増減するタイミングの検出に取り組んでいる [2]．Fushimi らはネットワークにエッジが追加・削除されたときの各ノードの影響度を定量化し，構造変化に基づく影響度指標を提案している [69]．人工ネットワークと実データを用いた評価実験により，離れたノード間の接続やコミュニティ間のエッジ削除などの大きな変化を検出できることを示した．

2.5　本研究の位置づけ

本研究では，ネットワークの構造変化に対するエッジの影響力指標として Stimulation Index を提案している．従来，ネットワーク上の情報拡散 (構造変化) に対する影響力を評価する際は，より多くのノードに情報を伝播させられる高影響ノードを検出することが大きな目的であった．本研究では，一連の情報拡散現象はエッジ出現の繰り返しであることから，まず構造変化の最小単位であるエッジに着目している．その上でネットワークの構造変化にどの程度貢献したか，という観点から情報到達ノード数ではなく，連鎖的に出現させたエッジ数を影響力とみなしている点で異なる．

また，ネットワークの構造推移を分析する手法としてトライアド推移を用いた手法も提案している．上述したとおり，トライアド推移は伏見ら [65] や中田ら [1] の研究でも扱われているが，本研究では，特定のトライアド構造を維持した時間を表現する滞留度を導入している点で大きく異なる．また，後者とは重み付き有向グラフを対象としている点でも異なることから，適用可能な分野も多く，より現実世界を反映した分析を実現している．

第 3 章

ネットワーク成長における エッジの影響力

3.1 はじめに

ここでは**第 1 章**で設定した課題のうち，ネットワーク成長におけるエッジ影響力の定量化に取り組む．本章では，エッジの発生が後続エッジの発生に与える影響を定量化した指標である Stimulation Index を提案する．評価実験等は情報拡散研究の文脈で行いつつ，Stimulation Index 自体は動的ネットワークの形成過程におけるエッジ影響の定量化手法として汎用性を高めることを目指す．

Twitter や Facebook，Instagram などのソーシャルネットワーキングサービス（SNS）の普及により，ニュース，商品のレビューや評判，意見，噂を始めとしたさまざまな情報が人から人へと飛び交うようになった．SNS の利用は個人によるものに留まらず，芸能人や政治家，一般企業，行政機関などが公式アカウントを設置し情報発信を行っている．もはや，SNS は情報の社会インフラの役割を担っているといえるだろう．

近年の SNS 活用では，情報をより広く伝播させられる高影響なユーザーを発見することが重要な課題となっている．例えば，企業の広告手段として SNS のシェア機能を利用したバイラル（口コミ）マーケティングがある．まず，SNS に自社製品のプロモーション投稿を行う．それを閲覧したユーザーが「この商品はよいものだ」と思えば，そのプロモーション投稿を自身のフォロワーへとシェアする．シェアされたユーザーが同様にその投稿をシェアすることで，プロモーション投稿はより多くのユーザーの目に触れることとなる．この時，情報の発信源，すなわち広告投稿を行うアカウントによって広告効果は大きく異なる．自社の公式アカウントが十分な広告効果を上げられるほどの力を持っていればよいが，現実的には自社アカウントだけでは不足であったり，もっと広告効果を積み増したいというケースが想定される．この時，SNS 上の他のユーザーに自社製品のプロモーション投稿をしてもらうことで，より高い広告効果を上げることが期待できる．当然 SNS 上には膨大なユーザーが存在するので，

プロモーションを依頼するユーザーの選定がビジネスの成否を決めるといえる．これは，多くのユーザーに情報を伝えられる重要ユーザーの推定タスクとして，影響最大化問題と呼ばれており，盛んに研究されている．[37, 70].

これまでの影響最大化問題では，上述したように情報をより広く伝播させられる高影響なユーザー（ノード）の検出に重きが置かれてきた．しかし，情報拡散プロセス自体はユーザーが他のユーザーへと情報伝達することの繰り返しからなる．そのため，一連の情報拡散プロセスの中でも，それぞれの情報伝達の情報拡散全体に対する貢献度は異なる．誰から誰への情報伝達が情報拡散に寄与したのかを明らかにすることは，情報拡散プロセスの理解のために重要である．

また SNS において，ユーザー同士は使用言語や居住地，趣味など様々な要素により互いに繋がりあいコミュニティを形成する．ユーザーはコミュニティ内の他のユーザの投稿を閲覧し，影響を受け，また自身も情報を発信する．この情報発信が活発なコミュニティは，自ずと多くのユーザーを獲得しその規模を拡大させてゆく．コミュニティの規模拡大は，多様・多数な意見から刺激を受け，情報発信を促すのでさらなる規模拡大を生む．よって，コミュニティひいてはネットワーク全体が成長するためには，受信側のユーザが情報の発信者となって他の受信者ユーザーを刺激しさらなる情報発信を誘発する，情報発信の連鎖が重要である．このようなネットワーク成長の観点からも，どのような情報伝達が多くの連鎖を発生させ，成長に貢献するのかを明らかにすること有用である．

上述したプロセスは，情報伝達（エッジ）が連鎖的に発生する動的なネットワークと見なせる．本研究では，このような動的ネットワークから重要な情報伝達（エッジ）を検出することを目的として，エッジの重要度指標である Stimulation Index(STM) を提案している．STM は，情報が人から人へと連鎖的に伝播する情報カスケード現象をベースに，あるエッジの発生が後続エッジの発生に与える影響を定量化した指標である．

また，STM は一連のネットワーク成長を入力として，その成長過程におけるエッジの影響を定量化するものである．すなわち，既に発生した情報拡散やネットワーク成長に対する評価指標といえる．しかし実際の応用では，エッジの出現時点で影響力がどの程度あるのか，そのポテンシャルを測りたい．そこで，エッジ出現時点の特徴量を用いて，将来の影響力推定の可能性について検討する．具体的には，次数中心性を拡張した k-次数中心性と STM に一定の相関があることを示す．

本研究の貢献は次の通りである．1) 情報拡散ネットワークにおけるエッジ重要度を定量化するための Stimulation Index(STM) を定義した．2) エッジ出現時の特徴量を用いて，将来の STM の予測が可能であることを示した．

本研究の構成は次のとおりである．本章に続き，**第 3.2 節**で提案手法である

Stimulation Index と k 次数中心性について説明する．**第 3.3 節**で提案手法の妥当性を評価するための実験設計について説明し，**第 3.4 節**で結果について述べる．**第 3.5 節**ではエッジ出現時の k-次数中心性と将来の STM の相関について説明し**第 3.6 節**で実験結果について議論する．最後に**第 3.7 節**で本論文と今後の課題についてまとめる．

3.2　Stimulation Index の提案

ここでは，提案手法である Stimulation Index（STM）の基本的なアイデアとアルゴリズムについて説明する．なお，予めここで扱うネットワークや用語について整理する．まず，ここで扱うのはエッジに方向のある有向ネットワークである．また，フォロー関係について**図 3.1** のように，ユーザー A をユーザー B をフォローしているとき，ユーザー A はユーザー B のフォロワーであり，ユーザー B はユーザー A のフォロイーであるとする．

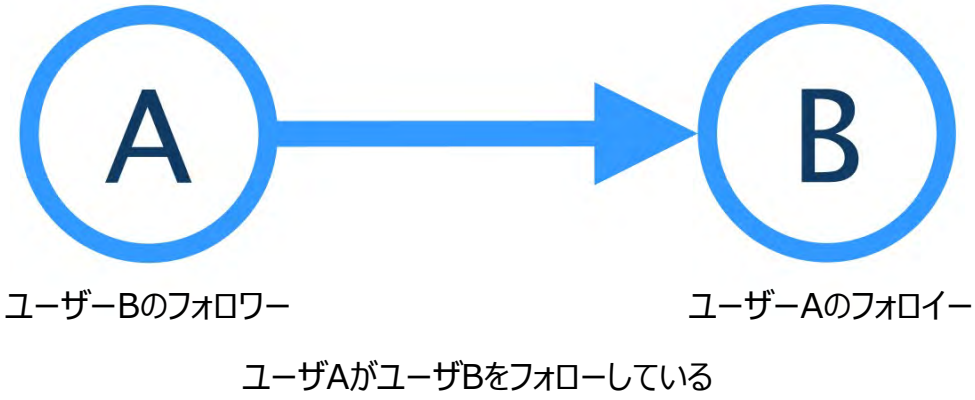


図 3.1: フォロワー・フォロイーの関係:ユーザー A をユーザー B をフォローしているとき，ユーザー A はユーザー B のフォロワーであり，ユーザー B はユーザー A のフォロイーである．

3.2.1　Stimulation Index の基本的なアイデア

先述したように，SNS では常に多くのユーザーによる情報発信が行われ，様々な情報が大量のユーザーへと拡散している．

はじめに，基本的な情報拡散の構造について説明する．まず情報源となるユーザーが情報を発信すると，そのフォロワーが情報を受け取り受信ユーザーとなる．受信ユーザーは受け取った情報を，更に自身のフォロワーへと発信する．この受信ユーザーが発信ユーザーとなる過程が繰り返されることで，情報は一人の発信ユーザーからネットワーク中の多くのユーザーへと拡散する．このように情報が連鎖的に伝播する現象を情報カスケード [28] と呼ぶ．

ここでユーザーをノード，ユーザー間の情報伝達を有向エッジとするネットワークを考えると，情報カスケードは，あるエッジの出現が後続するエッジ出現を誘発するといえる．そして，連鎖的にネットワークの規模を拡大していくネットワーク成長だといえる．

このとき最初に出現するエッジを除く全てのエッジは，先行して出現したエッジの影響により発生したと考えられる．これを踏まえ，本研究では動的ネットワークに出現するエッジが，将来的にどの程度エッジ出現を誘発するのかに着目する．情報拡散モデルの研究でよく用いられる線形閾値モデル [29] を応用し，ネットワーク成長におけるエッジの影響力を定量化した Stimulation Index(STM) を提案する．

STM の基本的なアイデアについて説明する前に，扱う記号を整理する．ネットワーク中のユーザから構成されるノード集合を $\mathcal{V}$，情報の発信ノードと受信ノードの順序付きペアで表される情報伝達のエッジ集合を $\mathcal{E}$ とする，有向ネットワーク $G = (\mathcal{V}, \mathcal{E})$ を考える．このとき，各エッジには出現順序に従ってインデックスを付与する．例えば，エッジ e_1 は最初に出現するエッジであり，エッジ e_2 はエッジ e_3 よりも先に出現する．

STM の基本的なアイデアについて，4 つのケースを例に説明する．**図 3.2** は最も基本的なエッジの誘発である．最初の情報発信ノードは v_1 である．まず，ノード v_2 が v_1 から情報を受け取り，それを隣接する v_3 へと送信している．この時 v_3 への情報伝達であるエッジ e_2 は，v_2 から v_1 への情報伝達エッジ e_1 によって誘発されたものとみなす．このように，あるエッジが誘発した後続するエッジ数を基本的な STM のスコア (STM スコア) とする．ここでは e_1 の STM スコアは $STM(e_1) = 1$ となる．また，e_2 は他のエッジに影響を与えていないので，STM スコアは $STM(e_2) = 0$ である．

図 3.3 は複数の先行エッジが協力して後続エッジを誘発するケースである．まず，ノード v_3 は，v_1 から情報を受け取った後に隣接するノード v_4 へと情報

を送信している（e_3）．そのため，e_1 の STM スコアを $STM(e_1) = 1$ としたいところであるが，エッジ e_3 が出現する前に，ノード v_3 は v_2 からも情報を受け取っている．この時，エッジ e_3 は e_1 と e_2 の協力によって出現したと捉えるのが自然であるので，2 本のエッジで STM スコアを分け合うこととする．よって，**図 3.3** における STM スコアは $STM(e_1) = 1/2 = 0.5$, $STM(e_2) = 1/2 = 0.5$, $STM(e_3) = 0$ となる．**図 3.2** 同様に，e_3 は他のエッジに影響を与えていないため STM スコアは $STM(e_3) = 0$ である．

続いて**図 3.4** のようなケースについて考える．これまで説明した方法で STM スコアを考えると，エッジ e_1 は単独で e_2 を誘発，さらに e_3 と協力して e_4 を誘発しているので STM スコアは $STM(e_1) = 1 + 1/2 = 1.5$，エッジ e_3 は e_1 と共に e_4 を誘発しているので STM スコアは $STM(e_3) = 1/2 = 0.5$ となる．しかしこの方法では，先に出現したエッジの影響力が残り続けるので，古いエッジほど STM スコアが大きくなってしまう．実際の情報拡散現象を鑑みても，ある情報伝達の影響力が永続して残り続けることは現実的ではない．そこで他者からの影響を受けて情報発信を行い，その後新たに情報伝達を受けた時，その情報伝達は異なる情報カスケードであるとみなす．具体的に説明する．まず，ノード v_2 は v_1 からの影響を受けて，ノード v_4 へと情報伝達を行う．その後，ノード v_2 は v_3 から異なる影響を受けて，今度はノード v_5 へと情報伝達を行う．この時，v_1 が v_2 与えた影響は残っていないものとする．すなわち，エッジ e_1 は単独で e_2 のみを，同様にエッジ e_3 は単独で e_4 のみを誘発したと考える．最終的に STM スコアは $STM(e_1) = 1$, $STM(e_2) = 0$, $STM(e_3) = 1$, $STM(e_4) = 0$ とする．

図 3.5 は間接的なエッジ誘発の例である．このケースにおいて，情報発信ノードは v_1 である．まずノード v_2 が v_1 から情報を受け取り，それを隣接する v_3 へと送信する．続けてノード v_3 が v_2 から影響を受け，v_4 へと影響を与える．ここで後方のエッジから順に考えると，e_3 は他に影響を与えていないので $STM(e_3) = 0$ である．続いて，エッジ e_2 は e_3 を誘発しているため，STM スコアは $STM(e_2) = 1$ である．エッジ e_1 については，先述したように情報拡散はユーザ同士の情報伝達が連鎖することによって発生する．e_1 から e_2，e_2 から e_3 と連鎖的に発生したエッジであり，e_1 がなければ発生しなかったエッジでもある．よってエッジ e_1 は直接的に e_2，間接的に e_3 を誘発しているとみなせるので，STM スコアは $STM(e_1) = 2$ とする．

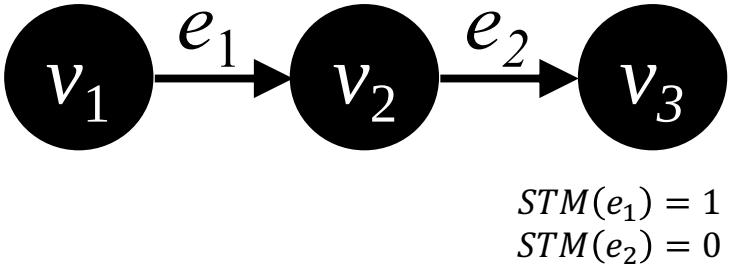


図 3.2: STM の例:基本的なエッジ誘発

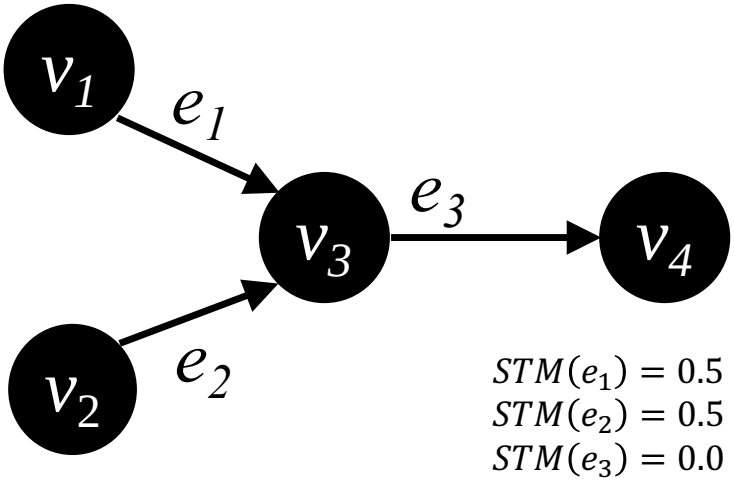


図 3.3: STM の例:複数の先行エッジによる後続エッジの誘発

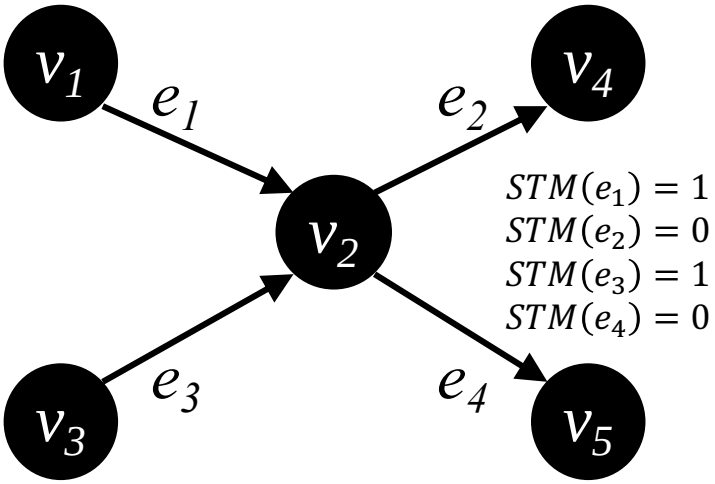


図 3.4: STM の例:異なる情報カスケードによるエッジ誘発

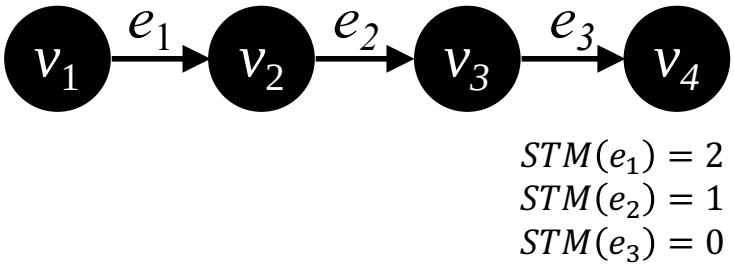


図 3.5: STM の例:間接的なエッジ誘発

3.2.2　Edge Relation Graph を用いた STM スコアの算出手法

本研究では STM スコアを算出するにあたり，ラングラフを拡張した Edge Releation Graph(ERG) を用いる．まず，ラングラフについて説明する．ラングラフは通常のグラフ G のエッジをノード，ノードをエッジとみなしたグラフである．リンク集合を $\ell \in \mathcal{L} \subset \mathcal{E} \times \mathcal{E}$ とすると，G のラングラフは $L(G) = (\mathcal{E}, \mathcal{L})$ と定義できる．例えば $L(G)$ において $e_s = (v_a, v_b) \in \mathcal{E}$ は $e_t = (v_c, v_d) \in \mathcal{E}$ に対して，$v_b = v_d$ の時に出リンクを張る．ここで $L(G)$ の $\mathcal{L}$ と G の $\mathcal{E}$ を区別するために，前者をリンク，後者をエッジと呼ぶ．$L(G)$ において，エッジはリンクで結ばれている．

なおネットワークが時間経過に伴い成長することを考慮し，e_s は e_t に先んじて出現するという制約を加えている．また**第 3.2.1 節**及び**図 3.4** でも説明したように，情報発信後に改めて情報を受信した場合，その刺激は異なる情報カスケードであると考える．すなわち，$e_s = (v_a, v_b) \in \mathcal{E}$ が $e_t = (v_c, v_d) \in \mathcal{E}$ に対して出リンクを持つのは，$v_b = v_c$ かつ $e_x = (v_b, *), e_y = (*, v_b) \notin \mathcal{E}$ かつ $s < x < y < t$ のときとする．これにより，古いエッジの STM スコアが極端に高くなることを防ぐ．本研究では，この制約を加えたグラフを従来のラングラフ $L(G)$ と区別するために，Edge Relation Graph(ERG) $ER(G)$ と呼ぶ．

図 3.6 にオリジナルグラフ G とその ERG $ER(G)$ の例を示す．まず**図 3.6(a)** での情報拡散の流れについて説明する．最初の情報発信ノードは v_1 である．まずノード v_2 が v_1 から情報を受け取り，さらにそれを隣接ノードへと送信する．すなわちノード v_3 及び $v4$ は v_2 から情報を受け取る．ここでノード v_3 は他のノードへと情報を送信していないことから，誰に対しても影響を与えていないといえる．一方ノード v_4 は v_2 に加え v_5 から影響を受け，さらに v_6, v_7 へと影響を与えている．また，v_6 は v_4 から影響を受け，v_8 へと情報を送信している．加えて別の情報カスケードとして v_7 から影響を受け，v_9 へと影響を与えている．**図 3.6(a)** を俯瞰すると，最終的に v_1 から v_2 への情報伝達 e_1 が，$\{e_2, e_3, e_5, e_6, e_7, e_8, e_9\}$ の情報伝達，すなわちエッジの出現を誘発していることが分かる．

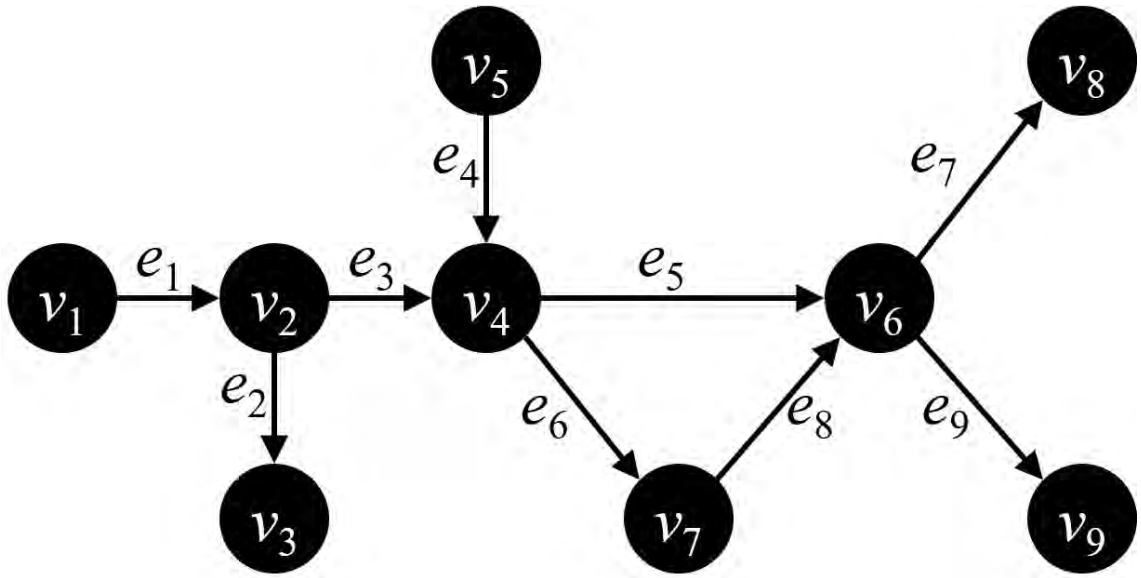
これを踏まえ，**図 3.6(b)** のように Edge relation graph $ER(G)$ を構築する．$e_1 = (v_1, v_2)$ と $e_2 = (v_2, v_3)$ は v_2 を介していることから，$ER(G)$ においては e_1 から e_2 へと有向エッジを引く．このようにラングラフの概念を導入することで，エッジの誘発・被誘発関係及びその連鎖関係を表現できる．

例えば，エッジ e_1 は e_2 と e_3 の 2 エッジの出現に影響を与えている．また，エッジ e_3 は e_5 及び e_6 に出リンクがあるため，2 エッジに影響を与えていることが分かる．同時にエッジ e_4 も e_5，e_6 に出リンクがあるので，やはり 2 エッジに影響を与えている．すなわち，エッジ e_3 と e_4 が協力して e_5 及び e_6 の 2 エッジ

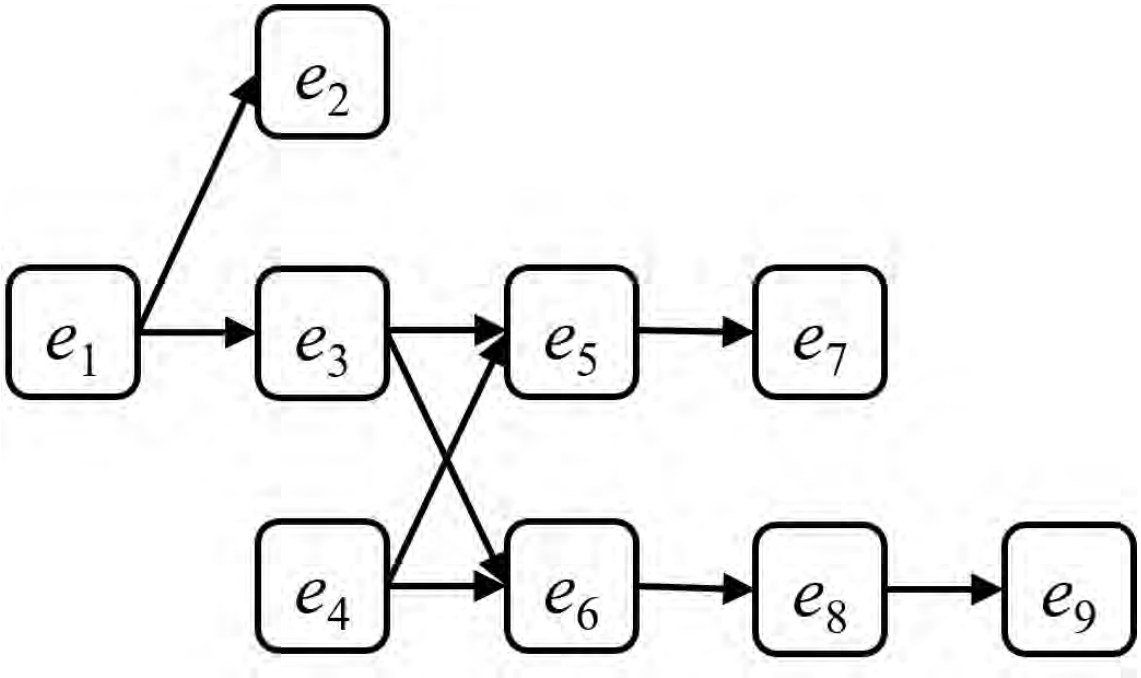
を誘発していることが分かる．そのため，3.3 でも述べたように，エッジ e_3 及び e_4 の誘発したエッジ数を分け合い，$1/2 + 1/2 = 1$ とする．

また，**図 3.6(a)** の e_9 に着目すると，一見 e_5 と e_8 の影響を受けているように見える．これは古いエッジの影響が延々と残り続けることを意味するが，これは現実的な事象ではない．そこで 3.4 でも述べたように，e_5 は e_7 にのみ影響し，その後 e_5 とは関係なく e_8 が e_9 に影響を与えているとみなす．これを**図 3.6(b)** に示す ERG で表現するとエッジ e_5 のリンク先は e_7, エッジ e_8 のリンク先は e_9 となる．よって，e_5 及び e_8 の STM スコアは $STM(e_5) = 1$, $STM(e_8) = 1$ となる．

加えて**図 3.5** でも説明したように，e_3 は $ER(G)$ の e_5 や e_6 のように直接繋がるエッジだけでなく，$\{e_7, e_8, e_9\}$ のような間接的に繋がるエッジについても刺激している．すなわち e_3 は $\{e_5, e_6, e_7, e_8, e_9\}$ の 5 エッジを直接・間接的に誘発しているといえる．これらのエッジは e_4 によっても刺激されていることから，影響エッジ数を分け合って，e_3 と e_4 の STM スコアは $STM(e_3) = 5/2 = 2.5$, $STM(e_4) = 5/2 = 2.5$ となる．



(a) オリジナルグラフ G



(b) Edge relation graph $ER(G)$.

図 3.6: 情報拡散のオリジナルグラフ及び Edge Relation Graph

3.2.3　Stimulation Index の計算アルゴリズム

ここでは，STM の計算アルゴリズムについて説明する．まず，与えられたネットワーク G を ERG $ER(G)$ に変換する方法について説明する．G の各エッジ $e \in \mathcal{E}$ について，$e_1, e_2, \ldots, e_T$ となるよう出現時間の昇順にエッジの順序をソートする．このとき，$T = |\mathcal{E}|$ である．2 つのエッジ e_s と e_t が，**第 3.2.1 節**で説明した制約を満たすとき，$\ell = (e_s, e_t)$ のようにエッジを有向リンクで接続する．以上の手順で，DAG 構造を持つ ERG $ER(G)$ を構築する．

Alg. 1 の 7 行目 30 行目では，各ノード $v \in \mathcal{V}$ に隣接するエッジ間の関係を求める．各ノード v に対して，out-neighbor(Forward 方向) の隣接ノード集合 $\mathcal{F}(v)$，in-neighbor(Backward 方向) の隣接ノード集合 $\mathcal{B}(v)$ を構築し，これらのエッジをキュー $\mathcal{Q}$ に格納する (9-14 行目)．続いて，各エッジのフラグを用いてそのエッジがノード v の流入エッジが流出エッジかを判断する（10，13 行目)．キューを出現時間の昇順にソートし（15 行目)，エッジ e_s をポップし（17 行目)，エッジ e_s とキューに格納されている他のエッジ e_t との関係を調べる（20〜27 行目)．エッジ e_s のフラグが $+1$，すなわち e_s が v の in-edge である場合，v の状態は情報を受信して保存した「stored」となる（18〜19 行目)．v の状態が「released」，かつ e_t のフラグが $+1$ の場合，これは v が既に情報を発信した上で e_t を介して情報を受け取ったことを意味する．すなわち，e_t と e_s は異なるカスケードとして扱われるため，ここでプロセスを終了する（21-23 行目)．

一方，エッジ e_t のフラグが-1 の場合には，e_s と e_t を同一のカスケードとして扱う．リンク集合 $\mathcal{L}$ にリンク (e_s, e_t) を追加し，v の状態を「released」に変更する．これにより v が受信した情報を放出したことになる（24-26 行目)．

次に，各エッジ $e \in \mathcal{E}$ に対する STM スコア $STM(e)$ の計算方法について説明する．ここで，$\mathcal{O}(e_s)$ と $\mathcal{I}(e_s)$ を，それぞれ $ER(G)$ における e_s の out-neighbor,in-neighbor の集合とすると，$e_s = (v_a, v_b)$, $\mathcal{O}(e_s) = \{e_t \in \mathcal{E}; \ell = (e_s, e_t) \in \mathcal{L}\}$ 及び $\mathcal{I}(e_s) = \{e_r \in \mathcal{E}; \ell = (e_r, e_s) \in \mathcal{L}\}$ となる．

また，直接接続しているエッジに対する STM スコアは，直接接続する out-neighbor に対する入次数の逆数の合計として，次のように表される．

$$\widetilde{STM}(e_s) = \sum_{e_t \in \mathcal{O}(e_s)} \frac{1}{|\mathcal{I}(e_t)|}.$$

Alg. 1 の 34 39 行目が，$\widetilde{STM}(e_s)$ の計算にあたり，e_s によって直接影響されたエッジの数を示している．また，間接的に接続しているエッジの STM スコア

$\widetilde{STM}(e)$ を再帰的に累積することで，最終的な STM スコアを算出できる．

$$STM(e_s) = \sum_{e_t \in \mathcal{O}(e_s)} \frac{\widetilde{STM}(e_t) + STM(e_t)}{|\mathcal{I}(e_t)|}. \tag{3.1}$$

この再帰関係においては，あるエッジの STM スコアを計算するためには，そのエッジの子孫のスコアを累積する必要がある．従って，STM 素顔を累積する際は幅優先探索とは逆の順序で累積していく．我々のアルゴリズムでは，エッジを発生時間の昇順にソートし，スタック $\mathcal{S}$ に格納する（41 行目）．これにより $ER(G)$ の DAG 構造の下流から，全てのエッジに対して順序よくアクセスできる（43-48 行目）．

このアルゴリズムは，Brandes が提案した幅優先でソースノードから他者への検索を行い，発見の逆順にスコアを再帰的に算出する媒介中心性のアルゴリズムと類似している．

3.3　データセット・比較手法

ここまで，STM のアイデア及びアルゴリズムについて説明してきた．ここでは，**第 3.4 節**及び**第 3.5 節**で Stimulation Index の評価を行うに当たり，データセット及び比較に用いる中心性指標について説明する．

3.3.1　人工的に生成したフォローネットワークデータ

第 3.4 節では，情報拡散のシミュレーションを行い，STM がネットワーク成長におけるエッジの重要度指標として機能するか評価する．本節では，情報拡散シミュレーションのために SNS のフォロー関係を模した人工フォローネットワークを作成する．フォローネットワークは，スケールフリー性[*1]があり，コミュニティ構造を持つことが知られている．この条件を満たすネットワーク生成モデルとして，Lancichinetti – Fortunato – Radicchi Benchmark [71] の LFR モデルを用いた．LFR モデルは，コミュニティ検出のタスクで知られており，ノードの次数とコミュニティサイズ (コミュニティ構成するノード数) がべき乗則に従うネットワークを生成できる．今回の実験では，LFR モデルのパラメータを**表 3.1** の設定とし，500 ノード，1138 エッジ，9 個のコミュニティからなる無向ネットワークを生成した．なお，ノード数以外のパラメータは文献 [71] の公開実装の初期値と同様とした．

ここで LFR モデルで生成できるのは無向ネットワークであるが，フォロー機能には有向･無向の 2 種類が存在し，SNS によって異なる．Twitter や Instagram

[*1] 次数分布がべき乗則に従う性質

Algorithm 1 Stimulation index

Input: $G = (\mathcal{V}, \mathcal{E})$
Output: $STM(e)\ \ \forall e \in \mathcal{E}$
// Constructing link set $\mathcal{L}$
$\mathcal{F}(v) = \{w; (v, w) \in \mathcal{E}\}\ \ \forall v \in \mathcal{V}$
$\mathcal{B}(v) = \{u; (u, v) \in \mathcal{E}\}\ \ \forall v \in \mathcal{V}$
$\mathcal{L} \leftarrow \varnothing$
for $v \in \mathcal{V}$ **do**
    $\mathcal{Q} \leftarrow$empty queue
    for $u \in \mathcal{F}(v)$ **do**
        $(u, v).flag \leftarrow +1;\ \ (u, v) \rightarrow \mathcal{Q}$
    end for
    for $w \in \mathcal{B}(v)$ **do**
        $(v, w).flag \leftarrow -1;\ \ (v, w) \rightarrow \mathcal{Q}$
    end for
    $Sort(\mathcal{Q})$
    while $\mathcal{Q}$ not empty **do**
        Pop $e_s \leftarrow \mathcal{Q}$
        if $e_s.flag = +1$ **then**
            $v.status \leftarrow$ "$stored$"
            for $e_t \in \mathcal{Q}$ **do**
                if $v.status =$ "$released$" **and** $e_t.flag = +1$ **then**
                    break
                end if
                if $e_t.flag = -1$ **then**
                    $\mathcal{L} \leftarrow \mathcal{L} \cup \{(e_s, e_t)\};\ \ v.status \leftarrow$ "$released$"
                end if
            end for
        end if
    end while
end for
// Calculating $\widetilde{STM}(e)$
$\mathcal{O}(e_s) = \{e_t; (e_s, e_t) \in \mathcal{L}\}\ \ \forall e_s \in \mathcal{E}$
$\mathcal{I}(e_s) = \{e_r; (e_r, e_s) \in \mathcal{L}\}\ \ \forall e_s \in \mathcal{E}$
for $e_s \in \mathcal{E}$ **do**
    $\widetilde{STM}(e_s) \leftarrow 0$
    for $e_t \in \mathcal{O}(e_s)$ **do**
        $\widetilde{STM}(e_s)$ += $\frac{1}{|\mathcal{I}(e_t)|}$
    end for
end for

// Calculating $STM(e)$
$\mathcal{S} \leftarrow sort(\mathcal{E})$
$STM(e) \leftarrow 0 \ \ \forall e \in \mathcal{E}$
while $\mathcal{S}$ not empty **do**
    Pop $e_t \leftarrow \mathcal{S}$
    for $e_s \in \mathcal{I}(e_t)$ **do**
        $STM(e_s)$ += $\frac{\widetilde{STM}(e_t)+STM(e_t)}{|\mathcal{I}(e_t)|}$
    end for
end while

におけるフォロー機能は，二者間でのフォローが別々に存在する有向フォローである．ユーザーは他者を自由にかつ一方的にフォローすることができる．[*2]そのため，ユーザー A がユーザー B をフォローしているからといって，ユーザー B がユーザー A をフォローしているとは限らない．この 2 つのフォロー関係は別の事象である．なお，A と B がお互いにフォローしている状態は，一般に相互フォローと呼ばれる．一方，Facebook におけるフォロー機能は無向フォローである．ユーザー A がユーザー B をフォローする際にはまず「友達申請」を送り，ユーザー B が申請を許諾すれば，両者の間に無向のフォロー関係が生じる．すなわち，ユーザー A がユーザー B をフォローしていることは，ユーザー B がユーザー A をフォローしていることと同値である．

本研究では，**第 3.5 節**で Twitter のネットワークを実験対象としていることや，他のドメインのネットワークへの適用も考慮し，有向フォローネットワークを構築したい．そこで，LFR モデルで生成した無向ネットワークを有向ネットワークに加工する．具体的にはランダムウォークによってエッジの方向を追加する．初期ノード u と隣接するノード集合のからランダムに 1 つのノード v を選び出し，移動することを H 回繰り返す．このとき，u から v への移動に応じて，有向エッジ $e = (u, v)$ を設定する．同じ 2 つのノードの組み合わせが複数回出てくる場合には，出現回数の多い方向を採用する．なお，出現回数が同じであれば双方向のエッジ（相互フォロー），ランダムウォークで通過しなかったエッジについては，$e = (u, v)$，$e = (v, u)$ 及び相互リンクの 3 種からランダムに方向を決定する．各コミュニティから 5 つのノードを初期ノードとして選び，移動回数を $H = 200$ とした．

[*2] なお，Twitter や Instagram におけるアカウントはパブリック状態とプライベート状態を任意に設定できる．自由なフォローはパブリックなアカウントに対してのみ行える．プライベートなアカウントについては，フォローする側が「フォローリクエスト」を送付し，フォロイーがそれを許諾することでフォロー関係が成立する．ただしこの時もパブリックアカウントへのフォロー同様，有向フォローである．

表 3.1: LFR モデルのパラメータ

パラメータ	値
ノード数	500
次数分布の指数	3.0
コミュニティサイズ分布の指数	2.0
平均次数	5.0
コミュニティサイズの下限ノード数	50
コミュニティ内のエッジ比率	0.95

3.3.2　Higgs Twitter Dataset

第 3.5 節では，エッジ出現時の特徴量から将来の STM スコアを予測できるか検討する．予測実験には，Twitter のインタラクションを収集した Higgs Twitter Dataset を用いているため，本節ではこれについて説明する．まず，Twitter 上で他のユーザーとやり取りするためには主に 3 種の方法:リプライ・メンション・リツイートが存在する．リプライは他者のツイートに対する返信を意味する．メンションは自分のツイート中で他者に対して言及することであり，ツイート中に相手のユーザー ID を挿入することで発生する．リツイートは他者のツイートを自分のフォロワーへとシェアする機能である．

Higgs Twitter Dataset は，2012 年 7 月に発見された Higgs 粒子に関するツイートのインタラクション (リプライ，メンション，リツイート) を収集したものである．具体的には，lhc, cern, boson, higgs などのキーワードを含むリプライ，メンション，リツイートについて，発信ユーザー，受信ユーザー，投稿時刻を収集している．収集期間は 2012 年 7 月 1 日から 2012 年 7 月 7 日である．

このデータセットから情報拡散のネットワークを構築する手順について説明する．時刻 t にユーザー u がユーザー v へリプライやメンションを送ると，u から v へ有向エッジ $e_t = (u, v, t)$ が生成される．これをインタラクションの発生順に繰り返すことで，情報拡散を表現する動的なネットワークが構築できる．本研究では，インタラクションの種類別に 3 つのネットワークを構築した．また，データセットの最後に出現するインタラクションの発生時刻 T における最大の弱連結成分を抽出し実験に用いた．

各ネットワークの最終時刻 T におけるノード数，エッジ数を**表 3.2** に示す．データセット中には同じノードペアであっても，異なる時刻に繰り返し出現する多重エッジが存在する．本研究ではこれをエッジの重みではなく，別々のエッジとして取り扱い，その累計エッジ数を動的エッジ数と表記している．静的エッジ数はこの同じノードペア間に出現したエッジをユニークに数え上げたエッジ数である．

表 3.2: Higgs Twitter Dataset のネットワーク規模

インタラクション	ネットワーク名	ノード数	静的エッジ数	動的エッジ数
リプライ	Reply-NW	1,233	1,622	1,971
メンション	Mention-NW	31,947	44,566	51,472
リツイート	Retweet-NW	45,804	62,817	74,380

3.3.3　比較に用いる中心性指標

本節では評価実験の比較手法として用いる中心性指標について説明する．本研究では，次数中心性，近接中心性，媒介中心性を比較手法として用いる．ただし，これら 3 つの中心性はノードの重要度指標である一方，STM はエッジの重要度指標である．そこで，ノードとエッジを反転させたライングラフ $L(G)$ を用いてエッジの中心性を測る．なお，時刻 s に出現するはエッジは e_s，同時刻のオリジナルグラフのスナップショットは G_s，ライングラフのスナップショットは $L(G_s) = (\mathcal{E}_s, \mathcal{L}_s)$ と表される．

次数中心性は，自身がどの程度，他者と直接繋がっているかを示す指標である．直接情報伝達できる人数が多いほど情報拡散への貢献度も高くなるものとして選定した．エッジ e_s の次数中心性 $DGC_s(e_s)$ は，e_s の次数をライングラフ $L(G_s)$ 全体のエッジ数のうち，接続するエッジ数の割合として次のように定義する．

$$DGC_s(e_s) = \frac{|\{e \in \mathcal{E}_s; (e_s, e) \in \mathcal{L}_s \vee (e, e_s) \in \mathcal{L}_s\}|}{|\mathcal{L}_s|} \tag{3.2}$$

近接中心性は，自身から他人まで平均的にどの程度近いかを示す指標である．他者に近いほど情報伝達，ひいては情報拡散が容易になるとして選定した．エッジ e_s の近接中心性 $CLC_s(e_s)$ は，ライングラフ $L(G_s)$ における e_s から他のエッジ $e \in \mathcal{E}_s \setminus \{e_s\}$ までの平均距離 $d_s(e, e_s)$ の逆数として定義する．

$$CLC_s(e_s) = \left(\frac{1}{s-1} \sum_{e \in \mathcal{E}_s \setminus \{e_s\}} d(e, e_s) \right)^{-1}. \tag{3.3}$$

媒介中心性は，ネットワーク上の行き来において橋渡しする役割の強さを示す指標である．切断されると接続先のコミュニティ全体に情報が行き渡らなくなることから，重要度の高いエッジを検出できるとして選定した．エッジ e_s の媒介中心性 $BWC_s(e_s)$ は，ライングラフ $L(G_s)$ において，e_r から e_t への最短経路 $\sigma_s(e_r \to e_t)$ のうち，e_s を経由する経路 $\sigma_s(e_r \to e_s \to e_t)$ の割合の和として定義する．

$$BWC_s(e_s) = \frac{1}{(s-1)(s-2)} \sum_{e_r \in \mathcal{E}_s \setminus \{e_s\}} \sum_{e_t \in \mathcal{E}_s \setminus \{e_s, e_r\}} \frac{\sigma_s(e_r \to e_s \to e_t)}{\sigma_s(e_r \to e_t)}. \tag{3.4}$$

3.3.4　本研究で扱うネットワーク

評価実験に用いるネットワークについて整理する．本研究では主にフォローネットワークと情報拡散ネットワークを扱っている．フォローネットワークはSNS におけるユーザのフォロー構造を表現したものである．本研究ではこれを静的かつ潜在的なネットワークとして扱う．

そしてフォローネットワーク上で観測される情報拡散系列を情報拡散ネットワークとする．すなわち情報拡散ネットワークとは動的に顕在化するネットワークである．提案手法の適用である「成長」するネットワークは情報拡散ネットワークのこと指すものとする．

3.4　Stimualtion Index の妥当性評価

3.4.1　シミュレーションによる情報拡散系列の作成

本節では，**第 3.3.1 節**で生成したフォローネットワークを用いて情報拡散のシミュレーションを行う．情報拡散モデルには LT モデル [29, 30] を，ノードの再アクティブ化を許可するように拡張した SIS-LT モデルを用いる．

第 2.2 節で説明したように，一般的な LT モデルの振る舞いは次のとおりである．まず，情報源となるノードを選択し隣接するノードに対し情報の伝達を行う．この時，送信者ノードから受信者ノードへと一定の重みが渡される．受信者ノードは受け取った重みの和が，予めノード毎に割り当てられたしきい値を超えた場合，次のステップにおいて送信者ノードへと変化（アクティブ化）し隣接するノードへ情報を送信する．送信者ノードがいなくなるまで繰り返すことで，情報拡散をシミュレートする．このときノードの状態遷移には，初期値である不活性状態（Susceptible），病原菌を受け取りアクティブ化した感染状態（Infected），抗体を獲得した回復状態（Recovered）の 3 状態を持つ病理感染モデルの SIR モデルが用いられる．すなわち，受け取った重みがしきい値を超えて，ノードがアクティブ (Infected) になり情報を送信すると，その後は情報拡散に関わらない (Recovered) ノードとなる．しかし実際の情報拡散では，感染症とは異なり，同じユーザーが繰り返し情報を発信することも当然起こりうる．そこでノードの状態遷移に，アクティブ状態で情報を送信した後，非アクティブ状態（Susceptible）に戻る SIS モデルを採用する．ただし SIS モデルだとネットワーク構造によっては情報伝達の無限ループが発生しうる．そこで，再アクティブ化する度にアクティブ化のためのしきい値を上げることで情報拡散を収束させる．すなわち，情報発信をする度に次回情報発信するためのしきい値が上げ，いわば「話題に飽きる」状況を再現する．具体的には，ノード v の初期しきい値を θ_v，再アクティブ化係数 $\theta_v(0)$ と表現する．そして，j 回目の再アク

ティブ化のしきい値を $\theta_v^{(j)} = (1 + \theta_v^{(0)})\theta_v^{(j-1)}$ と表す．

さらに現実の情報拡散は，ネットワーク全体に一様に拡散することは少ない．興味関心はユーザー毎に異なることから，拡散するトピックに興味のあるユーザーやコミュニティを中心に拡散していく．そこで，通常のシミュレーションに加えて，情報拡散の発生しやすさを偏らせたシミュレーションを行う．上述したように SIS-LT モデルでは，すべてのノードがアクティブ化のためのしきい値 θ_v を持つ．これに対し偏り係数 δ_v を乗じることで，情報拡散におけるユーザーの興味の偏りを表現する．情報拡散を発生しやすくしたいノード集合を $\mathcal{V}_\delta$ とするとき，ノード $v \in \mathcal{V}_\delta$ の偏り係数 δ_v を次のように設定する．

$$\delta_v = \begin{cases} \frac{|\mathcal{V}_\delta|}{|\mathcal{V}|} & v \in \mathcal{V}_\delta \\ \frac{|\mathcal{V}|+|\mathcal{V}_\delta|}{|\mathcal{V}|} & v \notin \mathcal{V}_\delta. \end{cases}$$

この偏り係数は対象のノードには情報が流れやすくなる一方，対象外のノードには情報が流れにくくなるように設定されてている．これは，全体に流れる情報量を通常のシミュレーションとなるべく揃えるためである．また偏り係数 δ_v の大小は，対象ノード集合 $\mathcal{V}_\delta$ のサイズによって決定される．$\mathcal{V}_\delta = \mathcal{V}$ のとき，$\delta_v = 1$ となり，全てのノードの偏り係数が等しい通常シミュレーションと同じ設定になる．この手順により，偏った情報拡散系列を通常の情報拡散系列とほぼ同じ規模の情報伝達数で作成することができる．

情報拡散系列のシミュレート方法について説明する．各ノード $v \in \mathcal{V}1$ つを情報源ノードとし，上述した SIS モデルにより情報拡散をシミュレートした．全てのノードの状態が不活性（susceptible）になると，情報拡散のシミュレーションを終了し，ノードのアクティブ化回数 $as(v;G)$ とアクティブ化したノード数 $ar(v;G)$ をカウントする．ここで，$as(v;G)$ は繰り返しアクティブ化したノードについてはアクティブ化した回数を数え上げること，$ar(v;G)$ はアクティブ化したノードのユニーク数を数え上げることに注意する．本実験では M 回のシミュレーションを行い，m 回目のシミュレーションで得られた上述の値をそれぞれ，$as(v;G)^{(m)}$ 及び $ar(v;G)^{(m)}$ と表記する．また，同様に全ノードの平均値を $\sigma_{as}(G) = 1/|V| \sum_{v\in V} \sigma_{as}(v;G)$，$\sigma_{ar}(G) = 1/|\mathcal{V}| \sum_{v\in\mathcal{V}} \sigma_{ar}(v;G)$ として算出した．

第 3.3.1 節において LFR モデルを用いて生成したフォローネットワークの可視化結果を**図 3.7** に示す．赤く示されているノードが興味の偏りを生じさせたノード集合 $\mathcal{V}_\delta$ であり，その規模は $|\mathcal{V}_\delta| = 51$ である．$\mathcal{V}_\delta$ は 9 つのコミュニティからランダムに 1 つのコミュニティを選択した．このネットワークを用いて，$|\mathcal{V}| = 500$ 個の各ノード一つずつを情報源とし，$M = 10$ 回のシミュレーションを行った．すなわちシミュレーションは合計 5000 施行であり，得られた情報拡散系列の総数も同様に 5000 である．また，再アクティブ化係数は $\theta(0) = 1.1$ とした．

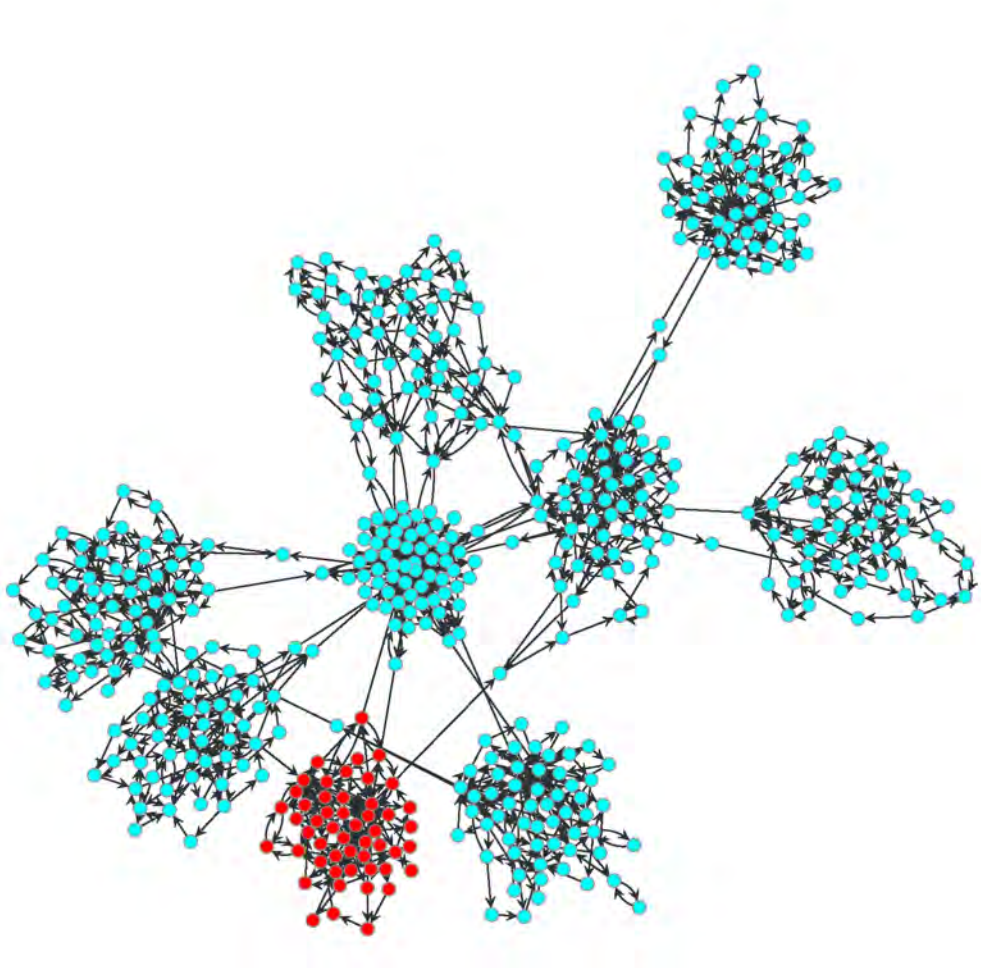


図 3.7: 人工ネットワークの可視化結果:赤いノードが情報拡散しやすいノード群である

3.4.2　エッジ切断による情報拡散の阻害

STM が情報拡散において重要度の高いエッジを検出できるか評価する．STM は各エッジが情報拡散，ひいてはネットワーク成長にどの程度貢献したかを定量化した指標である．そこで，今回は同じフォローネットワークからエッジを切断した場合の情報拡散の違いを以て，提案指標を評価する．具体的には，フォローネットワーク G からいくつかのエッジを切断しサブグラフ G' を抽出する．このサブグラフ G' を対象に，同様の設定で情報拡散のシミュレートを行う．エッジを切断するとその両端ノードを結ぶ動的エッジも発生しなくなるため，アクティブ化指標である $\sigma_{ar}(G')$ と $\sigma_{as}(G')$ は切断前に比べて少なからず低下する．

これらの低下が大きいほど，エッジ切断が情報拡散をより大きく阻害している，裏を返せば，情報拡散における重要なエッジだとみなせる．STM の高いエッジを切断することで，$\sigma_{as}(G')$ と $\sigma_{ar}(G')$ が比較手法に基づくエッジ切断よりも大きく減少することを確認し，提案手法の有効性を確かめた．

ただし，これまで説明してきたように提案手法は動的エッジの指標である一方，フォローネットワークは静的な構造である．そのため，$e_x = (u, v)$ と $e_y = (u, v)$ のように異なる時刻 t_x, t_y に，繰り返し同じノード間を結ぶエッジが出現することがある．このとき，STM スコアは e_x 及び e_y それぞれについて計算される．そこで，これらの STM スコアを静的エッジに集約し，STM スコアの期待値を求めることでこの問題に対処する．フォローネットワーク G 上の任意のノード v を情報源として，m 回目のシミュレーションを行った際の動的ネットワークを $G_{v,m} = (\mathcal{V}_{v,m}, \mathcal{E}_{v,m})$，時刻 t に出現したノード u から v への動的なエッジを $e_t = (u, v, t)$ と表す．このとき，静的エッジ $(u, v) \in \mathcal{E}$ の STM スコアの期待値 $stm((u, v))$ を次のように定義する．これにより，STM スコアが高くなりやすい静的エッジを抽出できる．

$$\mathcal{A}_{v,m} = \{e_t \in \mathcal{E}_{v,m} | e_t = (u, v, t)\}$$

$$STM((u, v)) = \frac{\sum_{v \in \mathcal{V}} \sum_{m=1}^{M} \sum_{e_t \in \mathcal{A}_{v,m}} STM(e_t)}{|\mathcal{V}| \times M}.$$

実験結果について説明する．**第 3.4.1 節**でシミュレートした情報拡散系列から静的エッジの STM スコアの期待値をランキングしたものを提案手法とする．また，比較手法として次数中心性，媒介中心性，近接中心性のランキングを用いた．比較手法は，いずれも代表的な中心性指標であることに加え，情報拡散という観点においても重要な役割を果たすと考えられることから選定した．

各指標のランキング順にエッジを切断したときの，アクティブ化回数 $\sigma_{as}(G')$，アクティブ化ノード数 $\sigma_{ar}(G')$ を**図 3.8** 及び**図 3.9** に示す．横軸はエッジ切断

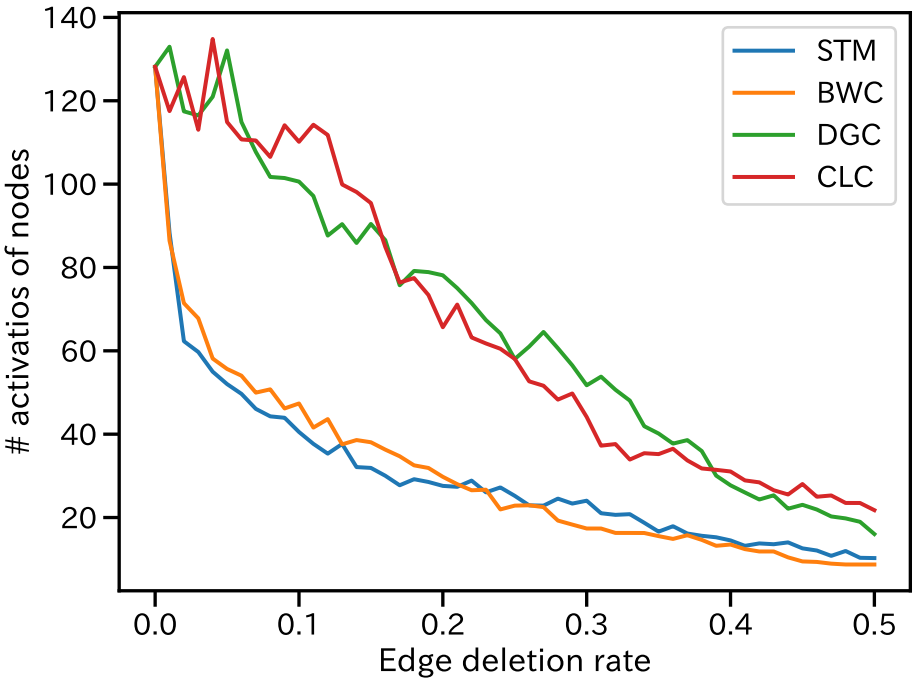
率 i，縦軸はメトリクス 2 種:$\sigma_{as}(G')$ 及び $\sigma_{ar}(G')$ である．なお，プロット線は手法に対応しており，STM スコアの期待値を STM，媒介中心性を BWC，次数中心性を DGC，近接中心性を CLC と表記している．例えばエッジ切断率 $i = 0.1$ は，各手法のスコアランキング上位 10% のエッジを切断した際の各メトリクスの値を示す．切断率が大きくなるほど情報拡散は阻害されるため，$\sigma_{as}(G')$ と $\sigma_{ar}(G')$ は小さくなりやすい．なお，エッジ切断割合 $i = 0.0$ は，エッジを切断しないことを意味するので，すべての手法で同じシミュレーション結果を用いた．よって，すべての手法で同じメトリクスの値が示されている．**第 3.4.1 節**で説明したように，SIS-LT モデルにおける情報拡散のしきい値 θ_v を一様にした場合と偏らせた場合で 2 種類のシミュレーションを行った．

まず，情報拡散のしきい値 θ_v が一様なシミュレーションの結果を**図 3.8** に示す．提案手法 (STM) を用いた場合，次数中心性 (DGC) と近接中心性 (CLC) に比べて，エッジ切断率が低いうちから $\sigma_{as}(G')$ と $\sigma_{ar}(G')$ が大きく減少している事がわかる．すなわち，僅かなエッジの切断が情報拡散が大きく阻害しているといえる．裏を返せば，この切断されたエッジは情報拡散において重要度の高いエッジであるといえる．またエッジ切断率が $i \leq 0.05$ 以降，提案手法の $\sigma_{as}(G')$ と $\sigma_{ar}(G')$ の減少がゆるやかになる．これは情報拡散の重要エッジを概ね切断し終わったためであると考えられる．一方，次数中心性 (DGC) と近接中心性 (CLC) は一定のペースで減少してゆく．すなわちランダムに切断していく場合とふるまいが大きく変わらず，情報拡散の重要度ランキングとしてはうまく機能していないと考えられる．なお，提案手法 (STM) と媒介中心性 (BWC) との間には大きな差は見られなかった．

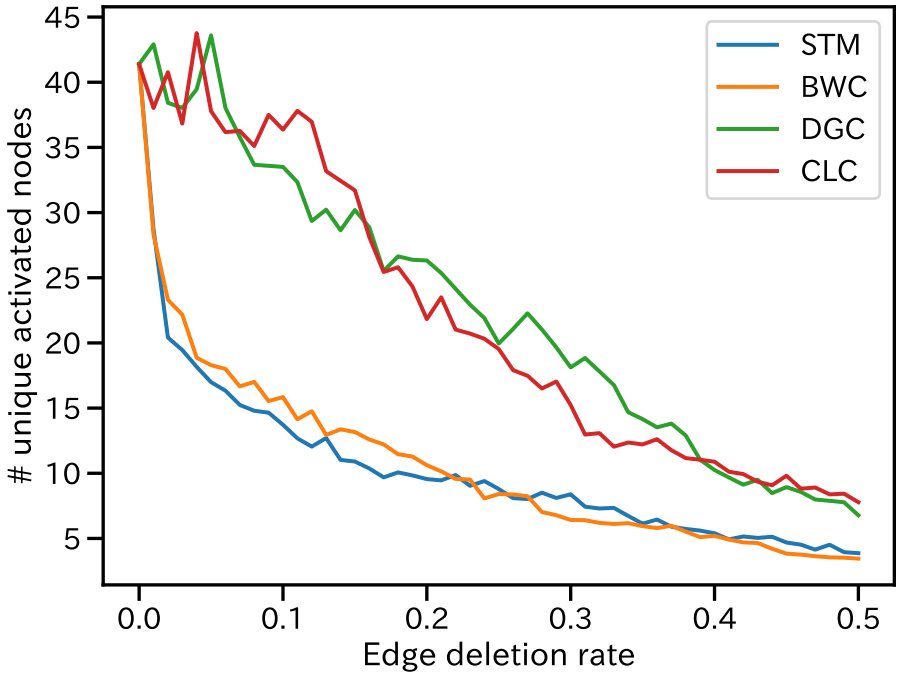
続いて，情報拡散のしきい値 θ_v が偏らせたシミュレーションの結果を**図 3.9** に示す．まず，**図 3.8** と同様に，STM 手法でエッジを切断した場合，次数中心性 (DGC) 及び近接中心性 (CLC) よりも $\sigma_{as}(G')$ と $\sigma_{ar}(G')$ が著しく減少することがわかる．また，**図 3.8** とは異なり，提案手法 (STM) と媒介中心性 (BWC) の間にも差が見られる．特にエッジ切断率 $i \leq 0.05$ 未満ではその差が特に顕著であり，提案手法 (STM) によるエッジ切断は，媒介中心性 (BWC) よりも大きく情報拡散が阻害できていることがわかる．また，この時切断されたエッジはもコミュニティ間を繋ぐエッジであると考えられる．加えて，**図 3.9(a)** のエッジ切断率 $i \leq 0.05$ 以降において，媒介中心性 (BWC) は一定ペースで減少していく一方で，提案手法の STM はゆるやかではあるが，下に凸のカーブとなっている．すなわち媒介中心性 (BWC) はコミュニティ間のエッジを切断しきった後は，情報拡散に貢献するエッジを検出できていないといえる．一方，STM はコミュニティ内における重要エッジを検出できていると考えられる．よって偏りのある情報拡散において，提案手法は重要エッジを上位にランキングできており，すべての比較手法に対してより有効に機能しているといえる．

以上の結果から，提案手法は情報拡散における重要エッジを適切にランキン

グでおり，動的なネットワークにおけるエッジ重要度の指標として妥当であるといえる．

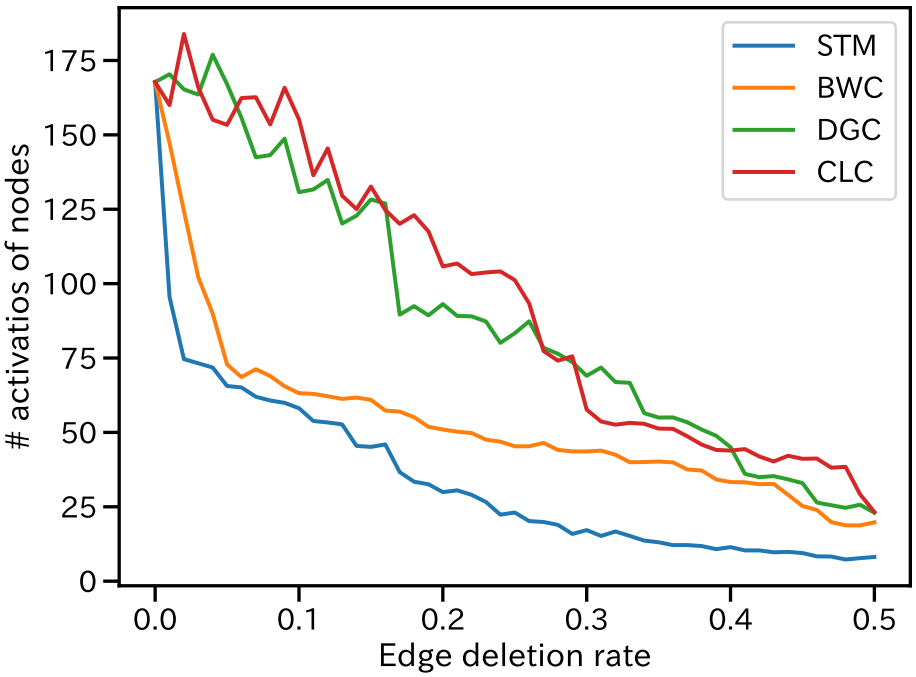


(a) アクティブ化回数 $\sigma_{as}(G')$

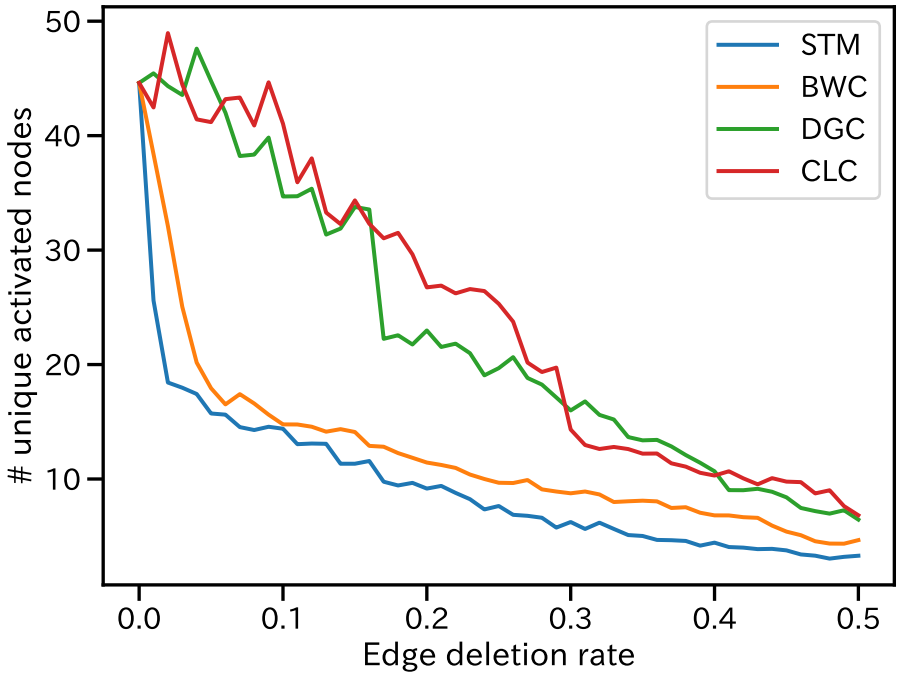


(b) アクティブ化ノード数回数 $\sigma_{ar}(G')$

図 3.8: 情報拡散のしきい値 θ_v が一様なシミュレーション



(a) アクティブ化回数 $\sigma_{as}(G')$



(b) アクティブ化ノード数回数 $\sigma_{ar}(G')$

図 3.9: 情報拡散のしきい値 θ_v を偏らせたシミュレーション

3.5　Stimulation Index の予測可能性

Stimulation Index は，情報拡散やネットワーク成長が発生した後に，その過程において各エッジがどの程度の影響力を持っていたのか，定量化する指標である．しかし応用面を考慮すると，エッジの出現時点においてそのエッジが将来的にどの程度影響力を持つのか，そのポテンシャルを測りたい．そこで本研究では，エッジ出現時点における特徴量を用いた将来における STM スコアの予測が可能か検討する．

3.5.1　エッジ出現時の特徴量

ここでは STM スコア予測モデルへの入力となる，エッジ出現時の特徴量について説明する．まず，**第 3.4 節**でも比較手法として利用した次数中心性，近接中心性，媒介中心性の 3 手法を採用する．また，対象が逐次変化する情報拡散ネットワークやネットワーク成長であることを考慮すると，高速に計算可能な特徴量を採用したい．しかし，**第 2.2.2 項**でも説明したように，次数中心性は計算コストが低いものの，自身とそれを取り巻く局所的なネットワーク構造しか参照しないことから情報量が少ない．一方で近接中心性と媒介中心性は，ネットワーク構造全体と自身の関係性を参照するので情報量が多いが，計算コストは高いという課題がある．そこで，計算コストの低い次数中心性について，参照範囲を拡張した k 次数中心性を用いる．

k 次数中心性について説明する．そもそも繰り返しの情報発信が発生する情報拡散のプロセスでは，既に情報伝達の実績を持っているノード間には将来的にも継続して情報伝達が発生する可能性が高い．そして，直接の情報伝達に加えて受信ノードを介して間接的に遠くのノードへと影響を与えることで，新たなエッジ出現を促し STM スコアは高くなる．そのため，STM スコアの高いエッジ (発信ノードと受信ノードの有向ペア) は，エッジの出現時点で受信ノードと直接リンクするノードはもちろん，2 ホップ，3 ホップ離れた情報伝達の実績があるノードが多く存在すると考えられる．次数中心性は自身が直接リンクするノード数の全ノード数に対する割合であるのに対し，k ホップ先まで繋がるノード数を参照したものが k 次数中心性であり，次のように定義する．

まず，時刻 s に出現するエッジ $e_s = (u, v)$ の k 次数中心性を求める時は，エッジが出現する直前，すなわち時刻 $s-1$ のネットワークスナップショット $G_{s-1} = (\mathcal{V}_{s-1}, \mathcal{E}_{s-1})$ を参照する．ノード u の in-neighbor 側に k ホップ先までのノード集合を $\mathcal{B}^{(k)}_{s-1}(u) = \{w \in \mathcal{V}_{s-1}; d_{s-1}(w, u) \leq k\}$，同様にノード v の out-neighbor 側の k ホップ先までのノード集合を $\mathcal{F}^{(k)}_{s-1}(v) = \{w \in \mathcal{V}_{s-1}; d_{s-1}(v, w) \leq k\}$，ノード u と v の最短距離長を $d_{s-1}(u, v)$ とする．この時，エッジ $e_s = (u, v)$ の k ホッ

プ先の次数中心性は $\mathrm{DGC}^{(k)}_{s-1}(e_s) = |\mathcal{B}^{(k)}_{s-1}(u) \cup \mathcal{F}^{(k)}_{s-1}(v)|$ で表される．k には任意の自然数が設定され，$k = 1$ のとき k 次数中心性は，通常の次数中心性と等しい．本評価実験では，この k 次数中心性を提案特徴量，次数中心性，近接中心性，媒介中心性を比較特徴量として STM スコアの予測を試みる．

3.5.2　STM スコアの回帰分析

ここでの目的は，高性能な STM スコア予測モデルの構築ではなく，予測可能性について検討し STM の実用性を確かめることである．そこで本研究では，解釈性の高さを重視して回帰分析を行う．エッジ出現時の特徴量を説明変数，最終時刻 T における STM スコアを目的変数とした回帰モデルを作成する．その後，回帰モデルの決定係数 R^2 を用いてその性能，すなわち STM スコアの予測可能性について検討する．また，偏回帰係数に基づき寄与度の高い変数を明らかにする他，回帰式による予測値と実測値の誤差分析も行う．

モデル構築

まず，モデル構築について説明する．提案特徴量 k 次数中心性 $(k = 1, \ldots, 10)$ のみを説明変数する単回帰モデル，比較特徴量（次数中心性，近接中心性，媒介中心性）のみを説明変数とする重回帰モデル提案・比較特徴量の両方を説明変数とする重回帰モデルの 3 つのモデルを構築した．いずれのモデルにおいても，目的変数は各エッジの最終時刻 T における STM スコアである．また，各変数のべき分布を考慮し，すべての変数を常用対数に変換している．実験で用いるネットワークは，**第 3.3.2 項**で構築した 3 つのネットワーク:Reply-NW・Mention-NW・Retweet-NW である．これらのネットワークは，出現時刻順に降順ソートされたエッジリスト $\mathcal{E}$ として与えられる．なお，エッジの出現時刻による STM スコアの差を除外するため，エッジリストの中央 50% 部分のみを回帰分析に用い，先頭と末尾のそれぞれ 25% ずつのエッジは分析対象から除外している．これは次のような不均衡:例えば，最終時刻 T に出現したエッジ e_T はデータセット中に後続するエッジが存在しないことから，STM スコアが常に 0 となってしまうことを考慮した処理である．

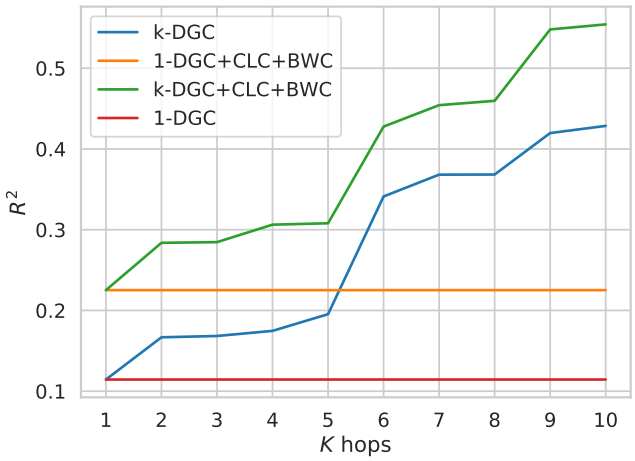
決定係数

回帰モデルの決定係数 R^2 を**図 3.10** に示す．横軸は k 次数中心性のホップ数 k，縦軸は決定係数 R^2，各プロット線がモデルに対応する．提案特徴量 k 次数中心性 $(k = 1, \ldots, 10)$ のみを説明変数とする単回帰モデルを k-DGC，比較特徴量（次数中心性，近接中心性，媒介中心性）のみを説明変数とする重回帰モデルを 1-DGC+CLC+BWC，提案特徴量と比較特徴量の両方を説明変数とする重回帰モデルを k-DGC+CLC+BWC と表記する．なお，次数中心性 (DGC)

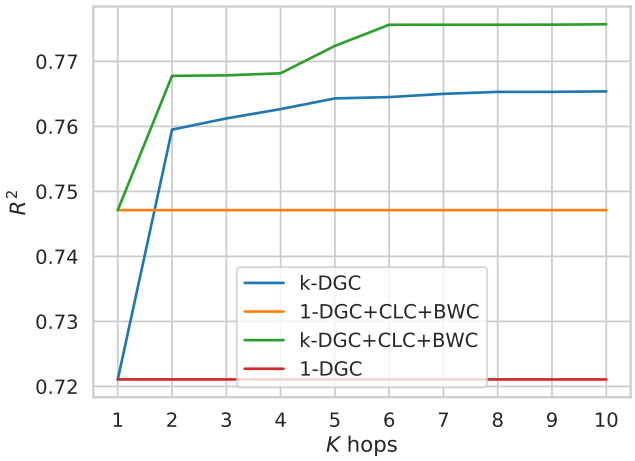
は $k = 1$ の k 次数中心性と等しいことから 1-DGC として別途記載している．

図 3.10(a) に Reply-NW の結果を示す．まず，提案特徴量を含む $k - DGC$ 及び k-DGC+CLC+BWC の 2 モデルは k の値が大きくなるほど決定係数 R^2 が大きくなる．比較特徴量のみの 1-DGC+CLC+BWC は k の値によって変動しない変数であるため直線である．提案特徴量である k-DGC モデルは $k = 1$ の時点での決定係数 R^2 は非常に低いものの，$k = 10$ では決定係数 R^2 が 0.4 程度まで上昇する．一方，比較特徴量である 1-DGC モデルは決定係数が 0.2 程度と低い値を示す．しかし，ここに提案特徴量を加えることで $k = 10$ において決定係数 R^2 を 0.55 程度まで大きく引き上げることができている．一般に決定係数 R^2 は 0.7 以上あれば説明力のあるモデルとされていることから，モデル自体の性能がよいとはいえないものの提案特徴量の k-DGC が性能向上に大きく貢献していることが分かる．

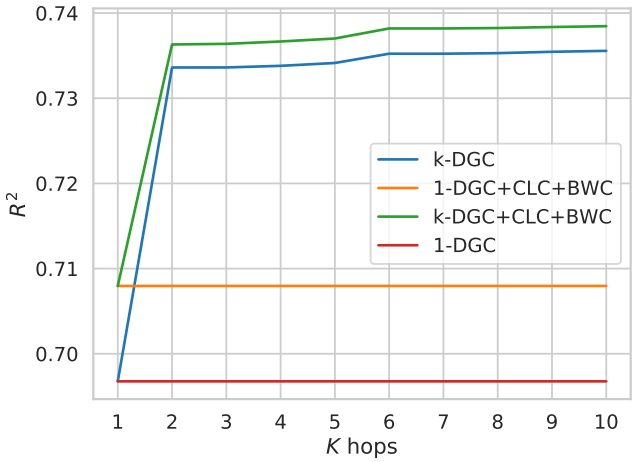
図 3.10(b) に Mention-NW の結果を示す．まず，$k = 1$ の時点でいずれのモデルも決定係数 R^2 が 0.7 を超えていることから，エッジ出現時の特徴量によって将来の STM スコアを予測できているといえる．また，k-DGC モデル及び k-DGC+CLC+BWC モデルでは，$k = 1$ から $k = 2$ 時点で決定係数 R^2 が大きく増加し，その後は緩やかなに増加するふるまいがみられる．このことから**図 3.10(a)** と同様に k 次数中心性がモデルの性能向上に貢献しているといえる．また，k が大きいほど計算量は大きくなることを踏まえると，Mention-NW においては $k = 2$ で十分な性能向上が見られることから k を比較的小さく設定することが有効だと考えられる．なお**図 3.10(a)** に示す Reply-NW は，Mention-NW と同様の振る舞いを示している．



(a) Reply-NW



(b) Mention-NW



(c) Retweet-NW

図 3.10: 決定係数 R^2 の推移:いずれのネットワークにおいても，K が大きくなるほど R^2 が大きくなることが分かる．特に Mention-NW と Retweet-NW では，$K = 2$ としたときの増加量が顕著である．

表 3.3: K-DGC+CLC+BWC を入力とした際の標準偏回帰係数．有意水準は 0.1 とし、t 検定でこの水準を下回った特徴量のみを記した。

	Reply-NW	Mention=NW	Retweet-NW
CLC	—	0.3571	—
BWC	-0.6997	—	-0.0659
1-DGC	-0.9469	-1.1364	-1.0152
2-DGC	—	0.1539	0.2448
3-DGC	—	—	0.0836
4-DGC	—	-0.2182	—
5-DGC	-10.7392	—	—
6-DGC	8.7608	—	—
7-DGC	—	—	—
8-DGC	33.3527	—	—
9-DGC	-39.8219	—	—
10-DGC	—	—	—

標準偏回帰係数

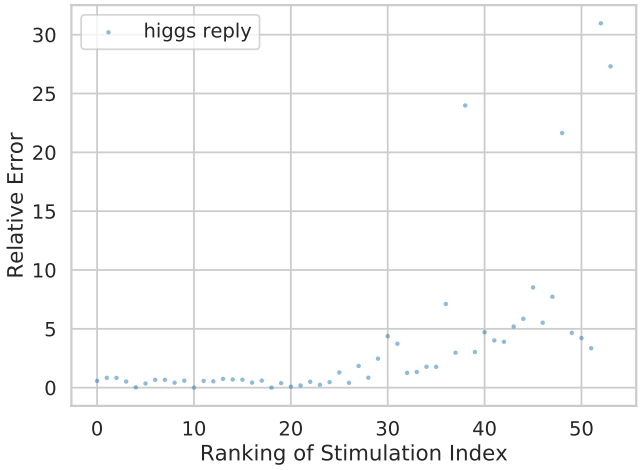
次に，どの説明変数が STM スコアの予測に寄与しているか議論する．最も説明変数の多い，10-DGC+CLC+BWC モデルの標準偏回帰係数を**表 3.3** に示す．これは変数の重要度を反映した数値であり，絶対値が大きいほどその変数が重要であるといえる．有意水準を 0.1 とし，t 検定で有意差のあった特徴量を記載する．

まず Reply-NW について，$k = 5, 6, 8, 9$ における k-次数中心性 (k-DGC) の標準偏回帰係数は次数中心性（1-DGC），近接中心性 (CLC)，媒介中心性（BWC）と比べて，その絶対値が大幅に大きいことがわかる．**図 3.10(a)** を見ても，k の増加と比例して決定係数 R^2 も高くなっていることから，Reply-NW では k が比較的大きい時の k 次数中心性が重要な変数だといえる．

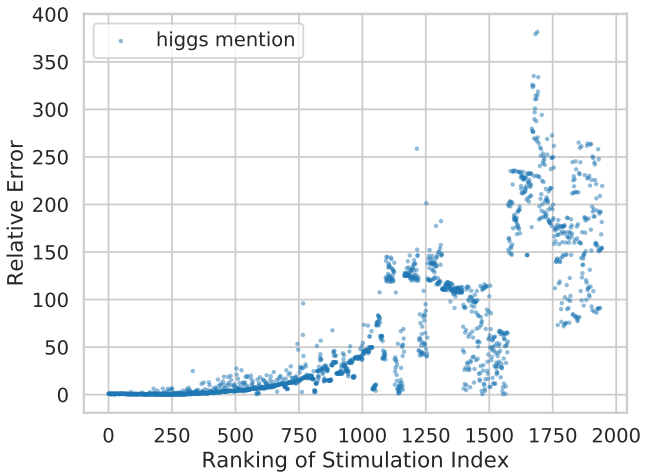
一方 Mention-NW と Retweet-NW では，1-DGC が最も重要な変数であることがわかる．加えて $k = 2, 3, 4$ における k-DGC も，近接中心性 (CLC) や媒介中心性 (BWC) に比べると高いスコアを示している．このことから Mention-NW と Retweet-NW においては，k の小さい k-DGC が重要な変数だといえる．**図 3.10(b) 図 3.10(c)** において $k = 3$ 以降の決定係数の増加は緩やかである．これは $k = 5$ 以降の変数が有意水準を満たさなかったことをとも符合しており，大きく予測に寄与しなかったのだと考えられる．このように，標準偏回帰係数の観点からも，提案する特徴量 k 次数中心性が STM スコア予測に有効であるといえる．

予測値の相対誤差

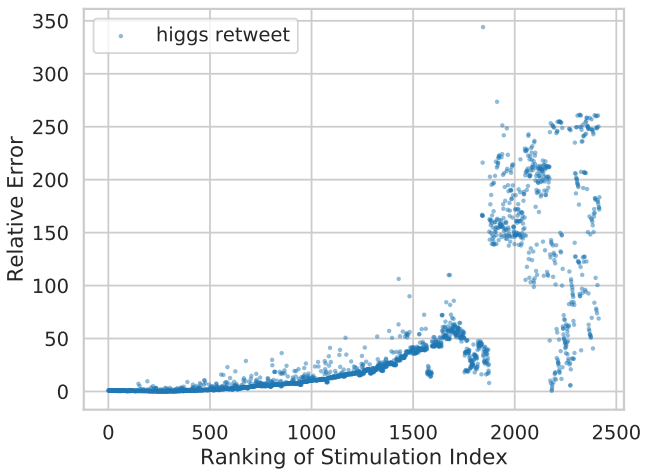
続けて，予測値の相対誤差を評価する．ここでは，出現時点で STM スコアが高くなると予測したエッジについて，どの程度の誤差があったのか議論する．実測値を y，予測値を y'，相対誤差を $|1-(y'/y)|$ と定義する．相対誤差が 0 であれば、全く誤差がないことを意味する．分析には最も決定係数 R^2 の高い 10-DGC+CLC+BWC モデルの結果を用いた．Reply-NW，Mention-NW，Retweet-NW それぞれの予測誤差について**図 3.11** に示す．横軸は STM スコアの実測値ランキング，縦軸は相対誤差を示しており，各プロット点がエッジに対応する．いずれのネットワークでも，STM スコアの実測値ランキング上位のエッジに関しては相対誤差が小さいことが分かる．一方，STM スコア実測値ランキング下位のエッジについては大きく外している事がわかる．このことから，エッジ出現時の特徴量を用いた回帰モデルにおいては，STM スコアの高い重要エッジについてはそのスコアやランキングを正しく推定できているといえる．しかし，STM スコアの低いエッジについては相対誤差が大きくなっていることから，本来重要ではないエッジを重要だと検出してしまうおそれがあるといえる．



(a) Reply-NW



(b) Mention-NW



(c) Retweet-NW

図 3.11: 1-DGC+BWC+CLC モデルにおける予測値の相対誤差。横軸は STM の順位，縦軸は相対誤差，プロット点がエッジを示す．どのネットワークでも，高い STM の重要なエッジほど相対誤差が小さい．すなわち提案特徴量を用いた重回帰モデルは，重要なエッジを正しく推定できるといえる．

3.5.3　計算時間

ここでは出現時のエッジ特徴量の計算コストについて扱う．まず，N ノード，T エッジからなるネットワークを考える．このとき，比較手法で用いた近接中心性 (CLC) の計算量は $O(T)$，媒介中心性 (BWC) の計算量は，$O(N^2 + NT)$ である．特に媒介中心性については，大規模なネットワークにおいて厳密な解を求めることすら困難である．もちろん，近似解を求める手法は多数提案されているが，動的に更新されるネットワークに対して逐次再計算を行うことは難しい．一方，提案特徴量である k-DGC は，対象エッジの両端ノードの隣接ノード数を数え上げるというシンプルな処理のため効率が良い．

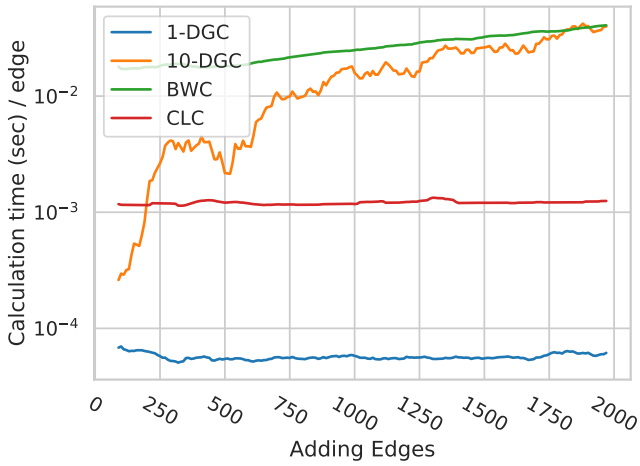
k-DGC($k = 10$)，次数中心性 (1-DGC)，近接中心性 (CLC)，媒介中心性 (BWC) の計算時間を**図 3.12** に示す．横軸は対象のエッジが何番目に追加されたか，すなわちエッジ追加時点における自身を含むネットワーク中のエッジ数を示す．縦軸は各エッジにおける特徴量の計算時間を示す．計算時間の差が大きいため対数目盛となっていることに留意されたい．なお，計算時間は平滑化のために過去 10 エッジの移動平均を用いた．実験環境に用いた CPU は Intel(R) Xeon(R) CPU E5-2680 v4@2.40 GHz x28 であり，RAM は 512 GB である．

Reply-NW における各特徴量の計算時間を**図 3.12(a)** に示す．先述したように，媒介中心性 (BWC) の計算時間はネットワークが大きくなるにつれ徐々に悪化していくことが分かる．これは自身とネットワーク構造全体との関係性を参照する手法であることから，ネットワークが拡張するほど探索範囲も増えていくためである．また，$k = 10$ の k-次数中心性 (10-DGC) も同様にネットワーク規模が拡大するほどエッジあたりの計算時間は悪化していく．最終的には媒介中心性 (BWC) との差は僅かなものとなる．これは Reply-NW 自体が小規模なネットワークであることから，10 ホップ先まで探索すると，概ねネットワーク構造の概ね全域を探索することに等しくなるためであると想定される．

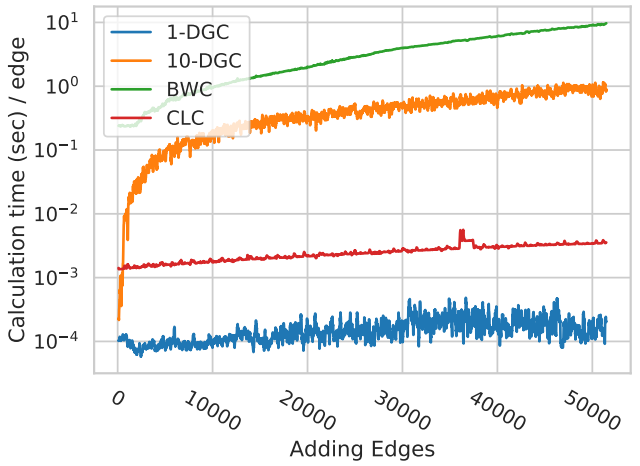
一方，**図 3.12(b)** に示す Mention-NW については振る舞いが異なる．媒介中心性 (BWC) の計算時間が最も長い点は Reply-NW と変わらない．一方，最終時刻において，提案特徴量の 10-次数中心性 (10-DGC) の計算時間は媒介中心性 (BWC) に対しおよそ 10 倍程度速い．また，近接中心性 (CLC) は徐々に計算時間が長くなるものの k-次数中心性 (10-DGC) より 100 倍程度高速に動作することが分かる．ただし，次数中心性 (1-DGC) は近接中心性 (CLC) より 10 倍程度速く，Mention-NW の予測モデルでは $k = 2$ で十分性能が出ていることを踏まえると提案特徴量の k 次数中心性は，k の値を適切に設定することで速度と性能を両立できると考えられる．また，**図 3.12(c)** に示す Retweet-NW も同様の傾向を示した．

以上の結果から，k-DGC は動的に更新されるネットワークに対して一定程度

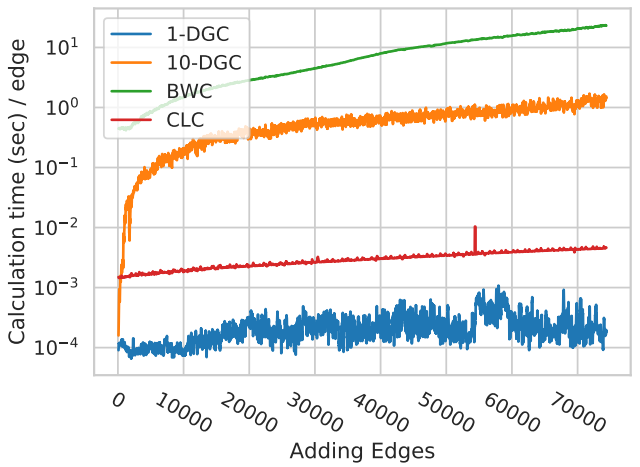
高速に計算可能であり，有効に機能するといえる．



(a) Reply-NW



(b) Mention-NW



(c) Retweet-NW

図 3.12: エッジ特徴量の計算時間．横軸は各特徴量を算出した時点のエッジ数，すなわちネットワークの規模を示す．縦軸は 1 エッジあたりの計算時間を示す．なお，過去 10 エッジの移動平均により平滑化を行っている．すべてのネットワークにおいて次数中心性が最も高速であること，及び 10-DGC が BWC よりも高速であることがわかる．

3.6　考察

3.6.1　ネットワーク別の STM スコア予測性能の差異

第 3.5 節において，エッジ出現時の特徴量を説明変数，最終時刻の STM スコアを目的変数とする回帰分析の結果，Mention-NW と Retweet-NW についてはモデルの決定係数が高く，良好な性能を得ることができた．すなわち，エッジ出現時の特徴量から将来的なエッジ誘発のポテンシャルを測ることができるといえる．一方，2 つのネットワークに比べると Reply-NW の決定係数は低く，エッジ出現時の特徴量と将来の STM スコアの間に強い関係を認めることができなかった．この理由は，STM スコアがネットワークの成長・拡大を前提とする情報拡散の評価指標であるためと考えられる．

Retweet-NW のデータソースとなったリツイート機能は，他者のツイートを自身のフォロワーへと一斉にシェアする機能である．広範囲へ情報を伝達する情報拡散現象であることから，STM の適用対象として適切であるため回帰モデルは期待通りの性能を発揮したといえる．

一方，Reply-NW のデータソースであるリプライ機能は情報伝達ではあるが，その性格は狭い範囲における名指しされたユーザー同士のインタラクションである．リプライの送信ユーザーと受信ユーザーの双方をフォローしているユーザーにしか表示されない[*3]こともあり，情報拡散の文脈には適合しにくいといえる．極端な例では 2 ノード間で繰り返しリプライのキャッチボールが繰り広げられる場合もあり，リツイートにように広範囲に拡散しにくい．この時，2 ノード間に動的エッジが重畳して発生し，誘発するエッジ数，すなわち STM スコアも上昇していく．一方で，ネットワーク構造そのものには変化が生まれないので，エッジ出現時の特徴量にも変化がない．このことが，説明変数であるエッジ出現時の特徴量と，目的変数である STM スコアの乖離を生み，回帰モデルの性能低下を招いたと想定される．

またメンション機能は自分のツイート中で他者に対して言及する機能である．リプライとは異なり，メンションツイートは自身のフォロワーのタイムラインに表示されるそのため，リプライとリツイートの中間の性質を持っているといえる．

このようなインタラクションの性質による違いは，**図 3.13** に示すエッジ多重度の分布からも観察できる．各ネットワークのエッジ数が異なるので，縦軸はエッジ数の累積比率とした．横軸はエッジの多重度，各プロット線が各ネットワークに対応している．多重度 1 のエッジの累積比率に着目すると，Reply-NW

[*3] リプライツイート自体は第三者からも閲覧可能であるが，送信ユーザー・受信ユーザーの双方をフォローしていないと，自身のタイムラインに表示されない．

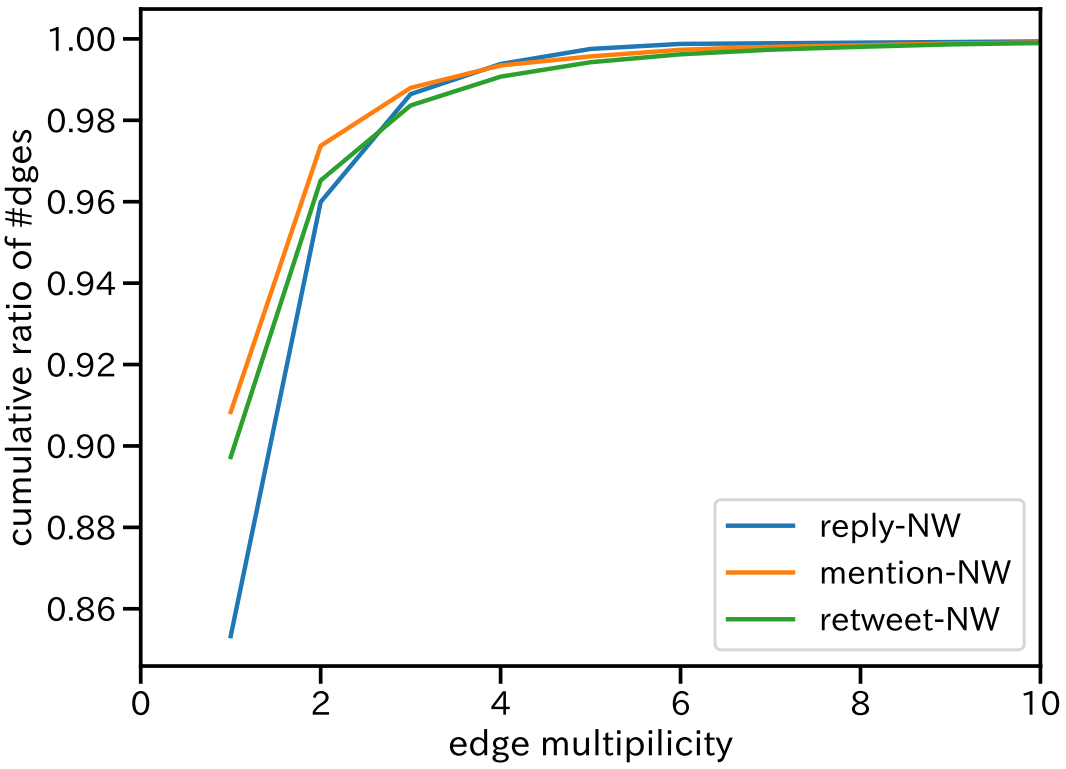


図 3.13: エッジ多重度の分布．横軸はエッジの多重度，縦軸はエッジ数の累積比率を示す．多重度 1 のエッジに着目すると，Reply-NW で 86% なのに対し，Mention-NW と Retweet-NW で 90% 程度である．Reply-NW の方が多重度の高いエッジの割合が大きいといえる．

が約 86%，Mention-NW と Retweet-NW がともに約 90% を示している．すなわち，Reply-NW は他の 2 つのネットワークに比べ，多重度の高いエッジの割合が高いことがわかる．これは，同じ 2 ユーザーの間でリプライの応酬が発生しているためだと考えられる．よって，Reply-NW の STM スコア回帰モデルの決定係数 R^2 が低いのは，そもそも STM の適用対象として適切ではなかったためだと考えられる．

3.7　まとめ

近年，Twitter や Facebook，Instagram などに代表される SNS は社会における情報インフラの役割を持ちつつある．SNS における情報拡散プロセスは，ユーザーが他のユーザーへと情報伝達することの繰り返しからなる．そのため，一連の情報拡散プロセスの中でも，それぞれの情報伝達の情報拡散全体に対する貢献度は異なる．誰から誰への情報伝達が情報拡散に寄与したのかを明らかにすることは，情報拡散プロセスの理解のために重要である．

本研究では，情報拡散のプロセスを情報伝達（エッジ）が連鎖的に発生する動的なネットワークとみなし，重要な情報伝達（エッジ）を検出するため，エッジの重要度指標である Stimulation Index(STM) を提案した．STM は人から人へと情報が伝播する情報カスケード現象をベースに，あるエッジの発生が後続エッジの発生に与える影響を定量化した指標である．評価実験において人工フォローネットワークを用いて情報拡散をシミュレートを行い，高い STM を持つエッジを削除することで，情報拡散が阻害されることを明らかにした．STM が情報拡散における重要度指標として有効に機能することを示した．

また，STM は動的ネットワークその成長過程におけるエッジの影響を定量化するものである．つまり既に発生した情報拡散やネットワーク成長に対する評価指標といえる．しかし応用面を考慮すると，エッジの出現時点で将来的な影響力がどの程度あるのか推定したい．そこで，Twitter のインタラクションデータから構築した情報拡散ネットワークを対象に，エッジ出現時の特徴量を説明変数，ネットワークの最終時刻における STM スコアを目的変数とする回帰モデルを作成した．決定係数 R^2，標準偏回帰係数，予測値の相対誤差を分析し，シンプルなモデルでも STM スコアが予測可能であることを示した．

本研究の貢献は次の通りである．1) 情報拡散ネットワークにおけるエッジ重要度を定量化するための Stimulation Index(STM) を提案した．2) エッジ出現時の特徴量を用いて，将来の STM の予測が可能であることを示した．

今後の課題には，拡散される情報の内容を用いることが挙げられる．STM はネットワーク構造のみに基づく手法であるが，加えてコンテンツベースの手法を相補的に用いることで，より情報拡散プロセスの理解を深めることが期待できる．また，STM スコアの予測モデルの性能向上にも繋がると考えられる．

本研究において，主に情報拡散の文脈で Stimulation Index の提案・評価を行ってきた．しかし，STM そのものは様々な他分野への応用を期待できるものである．例えば，引用ネットワークにおいて議論を活性化させたきっかけを評価する指標として用いることが考えられる．直接引用数だけでなく論文の間接的な影響力を評価することも可能である．EC サイトの共購買ネットワークであれば，連鎖的な購買を生み出すことのできる商品推薦へと応用が可能であろう．連鎖的に構造が変化していくネットワークであれば適用可能であることから，提案手法の汎用性は高いといえる．

第 4 章

トライアド推移に基づくネットワーク成長分析

4.1 はじめに

ここでは**第 1 章**で設定した課題のうち，第 2 の課題であるネットワーク成長における構造推移の定量化に取り組む．本章では，ネットワーク中のトライアド[*1]がどのように変化するかに着目したトライアド推移を用いた分析手法を提案している．Stimulation Index とは異なり，こちらはオンラインショッピングサイトにおける購買情報から構築する，商品の購買順序関係に着目した購買順序グラフ (PHG:Purchase History Graph) を対象としている．評価実験では，よく買われる商品の組み合わせや順序関係の形成過程に着目しつつ，ネットワークの成長過程を捉える手法として汎用性の高いものとすることを目指す．

近年，携帯型インターネット接続端末の普及や計算機の性能向上，ストレージの低廉化などを背景として，様々なサービスにおいて，大量のユーザ行動履歴データの取得・蓄積が可能となった．例えば，オンラインショッピングサイトにおける購買履歴，インターネットテレビにおける視聴履歴，観光地を巡る旅行者の移動履歴などが挙げられる．ユーザ行動履歴を分析することは，各種サービスにおけるマーケティングや施策決定のために重要であり，実際にこれらのデータを用いたユーザ行動の推定 [72, 73] やアイテム推薦 [74–76] の研究が盛んに行われている．

オンラインショッピングサイトの購買履歴データを対象とした分析では，多くのユーザの購買履歴データを用いることでアイテムがどのような順序で購入されるのか，というアイテム間の関係を抽出することができる．**図 4.1** にアイテム間の関係の例を示す．**図 4.1(a)** のスマートフォンとアクセサリーのような 1 つの中心的なアイテムを起点として複数の関連アイテムへと購買が発展する関係や，**図 4.1(b)** のコミックスのような連続して購入される関係など，アイテ

[*1] 3 ノードからなる最小の有向ネットワーク

ムのジャンルや特徴によって異なる関係が考えられる．そのため，商品の販売戦略やアイテム推薦などの場面では，「アイテムの買われる順序」を考慮することが有効だと考えられる．

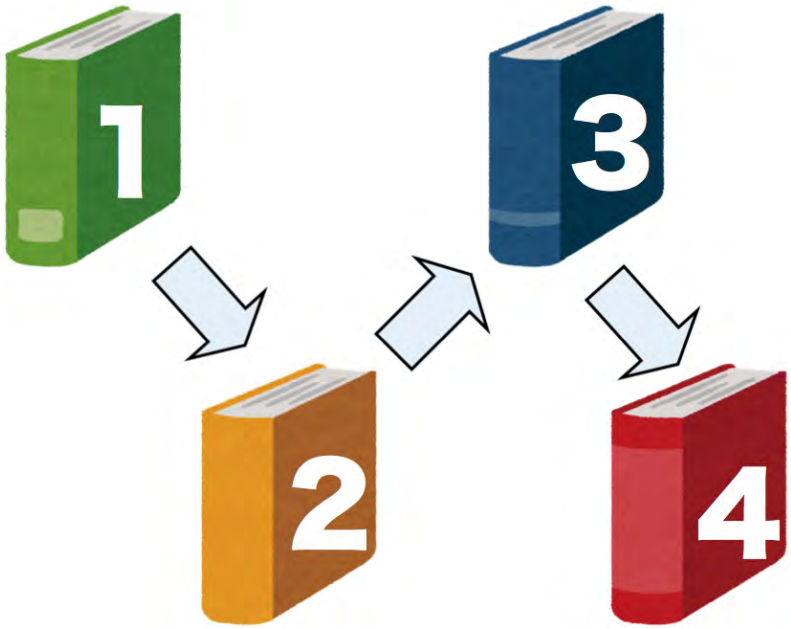
筆者らはこのようなアイテム間の順序関係を抽出するために，購買履歴からネットワークを構築し，各連結成分におけるトライアドパターンの分布に基づいて分析する手法を提案している [77]．その結果，アイテムの購買順序が明確に決まっているタイプや，2 商品同士の関係が長く連鎖するタイプ，中心的な商品と複数の関連商品からなるタイプが特に典型的な購買関係として抽出できることを明らかにしている．

これらの典型的な購買関係は数多くの購買行動によって構築されたものである．ここで，この購買関係が形成されるまでの過程に着目する．**第 1 章**でも述べたように，同じような購買関係であっても形成過程は異なることもあれば異なる購買関係の中にも共通する形成過程が見いだせることがあるだろう．典型的な購買関係を「売れやすい構造」でもあるので，その形成過程を理解することはマーケティングや広告戦略，商品推薦など幅広い分野において重要だといえる．そこで本研究では，この形成過程を分析するためにトライアド推移に着目した手法を提案する．

具体的には，ユーザによって連続購買されたアイテム間に有向エッジを張ることで重み付きネットワークを構築する．この時，購買時刻順にエッジを張ることでネットワーク成長を観察することができる．このネットワーク成長過程に対して，14 種類のトライアドパターンの推移傾向及び各構造における滞留度を算出する．これにより，商品カテゴリや推移パターンによる推移傾向や滞留度の違いを明らかにできると考えられる．成長傾向の差異を明らかにすることで，発売初期のタイミングや定番商品として定着したタイミングなど，各成長段階に合わせた商品推薦や販売戦略などに応用できると期待している．

本章の構成は次のとおりである．まず，**第 4.2 節**でトライアド推移を用いた提案手法について詳述する．**第 4.3 節**で実験設定について説明し，**第 4.4 節**で実データを用いた評価実験を行う．**第 4.5 節**で実験結果について議論し，最後に**第 4.6 節**で本章についてまとめる．

(a) スマートフォンとアクセサリ



(b) コミックス

図 4.1: アイテム間の購買順序の例

4.2　トライアド推移に基づくネットワーク成長の分析手法の提案

ここでは提案手法の大まかなステップについて説明し，その後各ステップの詳細について示す．本研究で分析対象とするのは，オンラインショッピングサイトにおけるユーザーの購買履歴から構築する購買履歴グラフ（PHG：Purchase History Graph）である．PHG はアイテム間の購買順序関係を示すネットワークであり，全ユーザーの購買履歴データに基づき，アイテムをノードとし，購買順序に従ってアイテム間に有向エッジを追加することで構築する．PHG についてアイテム間のエッジ方向を観察することで，アイテム同士の購買順序関係，すなわちどちらの商品が先に買われるのか，といった知見が得られる．

ここで，PHG は様々な商品が含まれたネットワークであることから，後述する提案手法をそのまま適用しても有効な示唆を得られない．そこで，購買順序関係が密なアイテム群を抽出するため，確率的ブロックモデル [78] によりいくつかのコミュニティに分割する．これにより，類似するカテゴリの商品やよく買われる組み合わせの商品ごとに部分ネットワークへと分解できる．なお，この部分ネットワークに対し**図 4.3** に示す 14 種のトライアドパターンの分布を算出することで，典型的な購買順序関係が得られる．これについては，著者らが文献 [77] で取り組んでいる．

ここで，購買履歴データには時刻が付帯していることから，PHG はその時刻順に逐次エッジが追加される動的なネットワークとみなすことができる．この時，エッジが追加される毎に，PHG 中のいくつかのトライアドがその構造を異なるトライアドパターンへと変化させる．すなわち PHG 中の任意のトライアドについて，トライアドパターンの推移系列が得られる．コミュニティ毎に推移系列の出現頻度を数え上げることで，それぞれのコミュニティの成長過程を表現する特徴量ベクトルが得られる．このコミュニティ毎に得られた特徴量ベクトルをクラスタリングすることで，類似の傾向を持つベクトルに分類する．さらに，各コミュニティやクラスタにおいてどのような推移系列が偏在しているか調べることで，典型的な成長過程を抽出する．

以下，提案手法の各ステップについて詳述する．

4.2.1　購買履歴グラフの構築

ユーザ集合を U，アイテム集合を I と定義する．ユーザ $u \in U$ が時刻 t にアイテム i を購入した場合，その購買行動を $r = (u, i, t)$ と表す．ここで，ユーザ u が時刻 t までに購買したアイテムの時刻順のリストを $I(u)^{(t)} = [i_1, i_2, \ldots]$ とし，$I(u)^{(t)}$ において連続するアイテムをペアにしたリストを

$$SI(u)^{(t)} = [(i_1, i_2), (i_2, i_3), \ldots] \subset I(u)^{(t)} \times I(u)^{(t)}$$

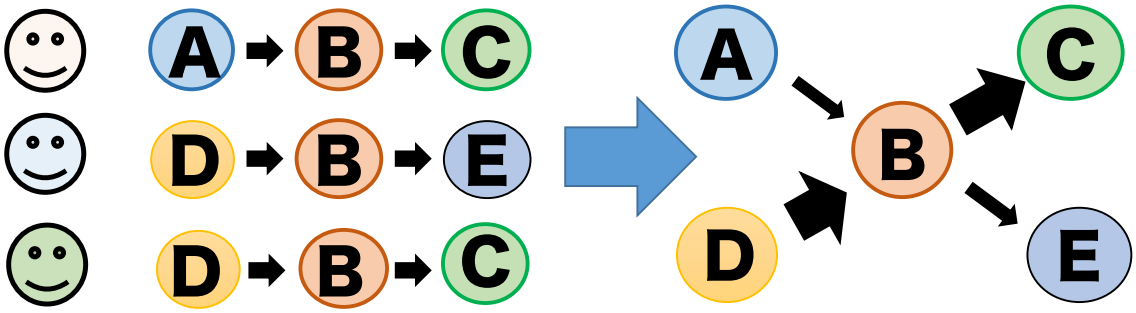


図 4.2: PHG の構築手法

とする．

いま，すべてのユーザーの連続購買アイテムペアリスト $SI(u)^{(t)}$ から，PHG $G^{(t)} = (V, E^{(t)})$ を構築することを考える．すなわち，PHG のノードはアイテムであるからノード集合は $V = I$ であり，エッジ集合は以下のように定義する：

$$E^{(t)} = \left\{ (i,j) \in \bigcup_{u \in U} SI(u)^{(t)} \right\}.$$

一般的に，各アイテムは異なる時刻に複数のユーザにより購買されることがあるため，アイテムペア (i,j) の出現回数をエッジ重み $w(i,j)$ として定義する．PHG を構築する際の模式図を**図 4.2** に示す．**図 4.2** では，3 人のユーザが A,B,C,D,E の 5 種のアイテムを購入した場合に構築される PHG の例を示している．アイテム (D,B) のペアは，複数のユーザに $D \to B$ の順序で購買されたため，PHG におけるエッジの重み（太さ）が大きくなっている．この $G^{(t)} = (V, E^{(t)})$ で表される有向グラフを時刻 t までの動的 PHG と呼ぶことにする．時刻 t が進むに連れ，新たなアイテムペア間にエッジが追加される，あるいは既存のエッジの重みが大きくなり，PHG は成長していく．

次に，購買履歴データにおける最終時刻を T としたとき，最終時刻 T における PHG $G^{(T)}$ を確率的ブロックモデル [78] により複数のサブグラフに分割する．なお，分割したサブグラフをコミュニティと呼び，m 番目のコミュニティを G_m と表記する．

4.2.2　トライアド推移の特徴量ベクトル化

トライアド，すなわち連結する 3 ノードからなる有向ネットワークの構造は，**図 4.3** に示すように 13 種類存在する [10]．本研究では，これらの 13 種類に 0 番のトライアドパターンを加えた 14 種をトライアドパターンとして扱う．以下，各パターンを TP0，TP1，...，TP13 と略記する．

第 4.2.1 項で説明したように，時刻 t が進むに連れてユーザーによる新たなア

イテムの購買が発生し，新たなエッジが追加されることから，PHG の構造は動的に変化する．この時，PHG に含まれるいくつかのトライアドはそのトライアドパターンを変化させる．例えば，TP1 を示すトライアドにエッジが 1 本追加された時，そのトライアドは TP3 か TP5 へと推移する．ただし既にエッジが存在するノード間へのエッジ追加，すなわち重複エッジが発生する場合には TP1 の状態に留まる．

このようなトライアドパターンの推移を**図 4.4** に示す．トライアド推移は，エッジ本数によって 5 層に分けられる．第 1 層はエッジ本数が 2 本の段階である．第 1 層のトライアドに 1 本エッジが追加された場合，第 2 層の TP3,5,7,9 に推移しうる．ただし，重複エッジが追加された場合には構造が変化しないため，第 1 層に留まる．同様に第 2 層にエッジが追加されれば，第 3 層の TP6,8,10,11 のいずれかに推移しうる．また，第 4 層は TP12 のみ，第 5 層は TP13 のみから構成される層である．

エッジが逐次的に追加されていくようなネットワークにおいて，すべてのトライアド推移は第 1 層の TP0,1,2,4 のいずれかを起点として始まる．例えば，TP4 を起点とした場合には，次の状態として TP7 もしくは TP9 に推移する．さらに TP7 及び TP9 から TP8,10,11 へと推移し TP12 を介して TP13 へ到達する．こうして形成される $TP4 \rightarrow TP7 \rightarrow TP10 \rightarrow TP12 \rightarrow TP13$ のようなトライアドパターンの推移系列を，本研究では TP 推移パターンと称する．この TP 推移パターンは最大で 28 通り存在する．

ここで各コミュニティに対して，時刻 t がすすむにつれて変化する TP 推移パターンを数え上げる．具体的には，コミュニティ $G^{(m)}$ に属するノード集合 $V^{(m)}$ から，隣接関係にある 3 ノードの組み合わせであるトライアド集合 $C^{(m)}$ を抽出する．全ての要素についてエッジの付与ごとにトライアドパターンを取得する．これにより，$G^{(m)}$ の TP 推移パターンの分布である $\mathbf{x}_m$ が算出できる．$\mathbf{x}_m$ は各 TP 推移パターンの出現頻度を表しており，28 次元のベクトルとなる．これは TP 推移パターンの出現頻度の累積度からより典型的な TP 推移パターンを抽出するため，出現頻度ベクトルを正規化 (L_1 norm) することで，TP 推移パターンの出現確率ベクトルとする．なお 4.4 で後述するが，本研究では第 3 層までの推移について扱う．

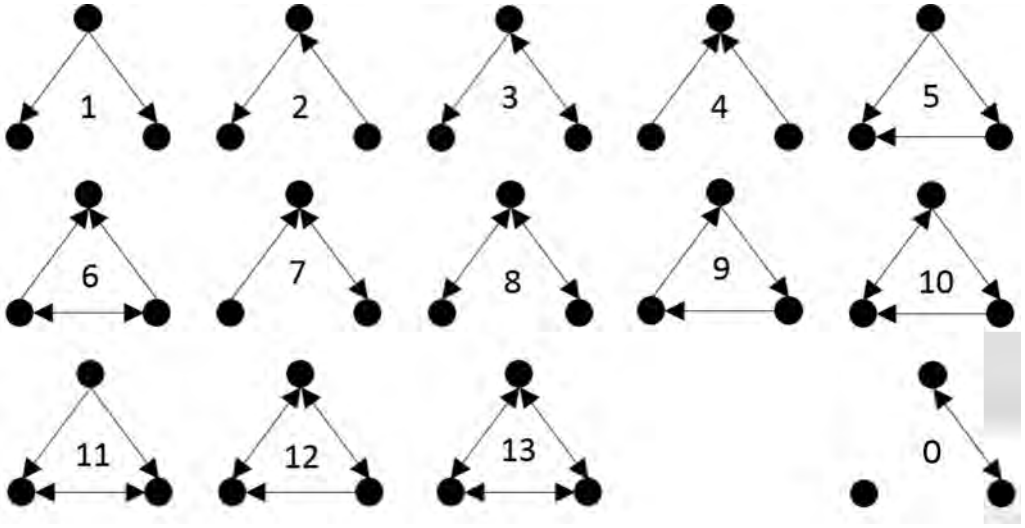


図 4.3: トライアドパターン (全 14 種)

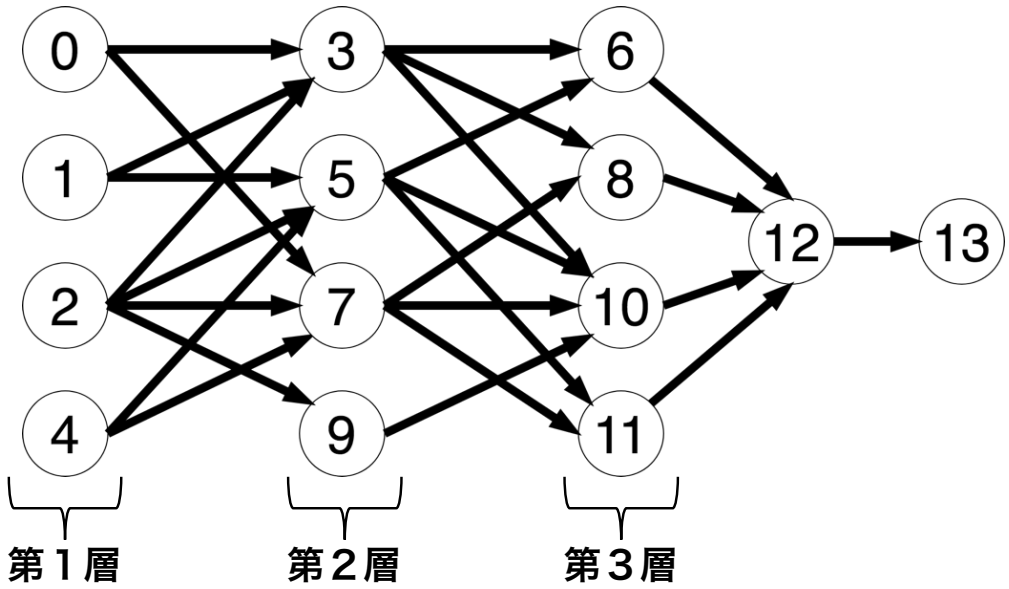


図 4.4: トライアド推移パターン (全 28 種)

4.2.3　トライアド推移における滞留度

第 4.2.2 項でも説明したように，重複エッジが追加される場合にはトライアドパターンの推移が起こらない．このとき，同じ TP 推移パターンであっても，なかなかトライアド推移が発生せず，最初の層に留まる時間が長い系列や逆に速やかに次層へ推移する系列があると考えられる．そこで，トライアドへのエッジ追加が発生しているにも関わらず，構造変化が発生せずに同一のトライアドに滞留する期間を定量化する．PHG は逐次的にエッジが追加されるネットワークであることから，滞留期間を同じトライアドパターンを維持している間のステップ数とする．

本研究では，コミュニティ内においては同じ TP 推移パターンは似たような滞留度を示すという仮定のもと，コミュニティ毎に TP 推移パターンの滞留度を集約する．コミュニティ m 内で観測された推移パターン $a \to b \to c$ における第 l 層の滞留数を $x_m^{(l)}(a \to b \to c)$ とし，全 3 層の滞留数の和を $x_m(a \to b \to c) = \sum_{l=1}^{3} x_m^{(l)}(a \to b \to c)$ と表す．

そして，各層の和に対する割合を $y_m^{(l)}(a \to b \to c) = x_m^{(l)}(a \to b \to c)/x_m(a \to b \to c)$ とし，第 l 層の滞留度と呼ぶ．各層の滞留度を要素とする 3 次元ベクトル $\mathbf{y}_m(a \to b \to c)$ をコミュニティ m の推移パターン $a \to b \to c$ の滞留度ベクトルとして定義する．1 つのコミュニティにおいて，最大で 28 種類の TP 推移パターンが出現しうるので，コミュニティごとに 28 種の 3 次元滞留度ベクトルが得られる．[*2]

4.2.4　トライアド推移の分析手法

ここでは，TP 推移パターン出現確率ベクトル及び滞留度ベクトルの分析手法について説明する．PHG から取得した 2 種類のベクトルを用いて明らかにしたい事柄は大きく分けて次の通りである．一点目は，コミュニティ毎に出現する TP 推移パターン及び滞留度の違いを確かめることである．二点目は，類似するコミュニティをグルーピングし，典型的な TP 推移パターン及び滞留度を抽出することである．

具体的な分析手順について説明する．まず，全 M コミュニティの滞留度ベクトルについて k-means 法により K 個のクラスタに分類する．ベクトル間の距離計算にはユークリッド距離を用いる．その上で，各クラスタにおける TP 推移パターンの偏在度を定量化する．

K 個のクラスタに分類した際，あるクラスタには TP 推移パターン $a \to b \to c$

[*2] 本研究では第 3 層までの TP 推移パターンについて扱うため，滞留度ベクトルも同様に第 3 層までについて扱う

が，別のクラスタには d から始まる TP 推移パターン $d \to Y \to Z$ が，また別のクラスタには e で終わる TP 推移パターン $X \to Y \to e$ が偏って分類されているかなどを定量化する．クラスタ C_k に M 個中 n 個のコミュニティの TP 推移パターン $h : a \to b \to c$ が割り振られた場合，TP 推移パターン h の，クラスタ C_k での出現頻度を $c_{h,k} = n$ とする．TP 推移パターン h の出現確率を $p_h = \sum_{k=1}^{K} c_{h,k}/L$，クラスタ C_k の割り当て確率を $q_k = \sum_{h=1}^{28} /L$ とする．ここで，$L = \sum_{k=1}^{K} \sum_{h'=1}^{28} c_{h',k}$ は全 TP 推移パターンの総出現回数を意味する．クラスタと TP 推移パターンの間に偏りがなく，各 TP 推移パターンが一様ランダムにクラスタに割り振られると仮定すると，周辺確率 p_h と q_k を用いて期待値 $L \cdot e_{h,k} = L \cdot p_h q_k$ を計算できる．実際の出現頻度と期待値との差，および，標準偏差により以下の Z スコアを計算する：

$$z_{h,k} = \frac{c_{h,k} - L \cdot e}{\sqrt{L \cdot e_{h,k}(1 - e_{h,k})}}. \tag{4.1}$$

Z スコア $z_{h,k}$ はランダムな場合と比較した際の偏り具合を表し，値が正で大きいほどクラスタ k に TP 推移パターン h が統計的有意に多く存在するといえる．すなわち Z スコアの値が大きい TP 推移パターンは，当該クラスタにおける典型的な TP 推移パターンといえる．これにより，「類似するコミュニティをグルーピングし，典型的な TP 推移パターンを抽出」できる．またクラスタをコミュニティに置き換えることで，各コミュニティにおける TP 推移パターンの偏在度を算出する．これにより「コミュニティ毎に出現する TP 推移パターンの違い」を確かめることができる．また，滞留度ベクトルをクラスタ及びコミュニティ毎に集約することで，それぞれの滞留傾向を知ることができる．

4.3　購買履歴グラフ

4.3.1　データセット

本研究では，商品の購買順序を表す PHG を構築するにあたり，楽天市場のレビューデータセット[*3]を利用する．楽天市場は楽天グループ株式会社によって運営されているマーケットプレイス型のオンラインショッピングサイトである．本サイトにおいて，ユーザーは商品の購入はもちろん，商品に対しテキスト及び 5 段階評価のレビュー評価を付与することができる．データセットはこのレビューを対象としたものであり，ユーザー ID，商品 ID，レビュー投稿時刻，レビューのテキスト，レビューの 5 段階評価，商品名，商品カテゴリ，購入フラグなどが収集されている．**表 4.1** に示すようにおよそ 6,500 万レビュー存在する．

データクレンジングについて説明する．レビューの際，ユーザーは自身のユーザー ID を匿名化することができる．また，基本的に購入した商品についてのみレビューを投稿できる仕組みとなっているが，一部の店舗については未購入の商品についてもレビューできる．本研究では，利用者の購買順序を求める必要があることから，投稿者が一意に判別できるレビューを抽出した．その際，購入した事が確認できない，あるいは投稿日時が欠落しているレビューを除外した．以上のデータクレンジングを行い，2,445,084 ユーザによる 17,794,337 レビューを評価データセットとした．

4.3.2　購買履歴グラフ

4.3.1 で抽出した約 1,780 万レビューを基に，**第 4.2.1 項**の手順に従い，商品をノード，連続して購買された商品ペア間にエッジを付与し，PHG $G^{(T)}$ を構築した．続いて，確率的ブロックモデルにより複数のコミュニティ (部分ネットワーク) に分割した．構築した PHG の規模を**表 4.2** に示す．また，コミュニティのサイズ（ノード数）分布を**図 4.5** に示す．縦軸はコミュニティサイズ（ノード数），プロット点はコミュニティと対応する．これを見ると，大半のコミュニティがおおよそ数百ノードから一千ノード程度で構成されていることが分かる．

[*3] http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html

表 4.1: 楽天市場データのレビュー数

	ユーザー数	レビュー数
全レビュー	-	65,221,470
ユーザーが一意に識別可能なレビュー	-	30,450,224
評価に用いるレビュー	2,445,084	17,794,337

表 4.2: PHG の規模

メトリクス	サイズ
ノード数	273,994
エッジ数	459,110
コミュニティ数	207

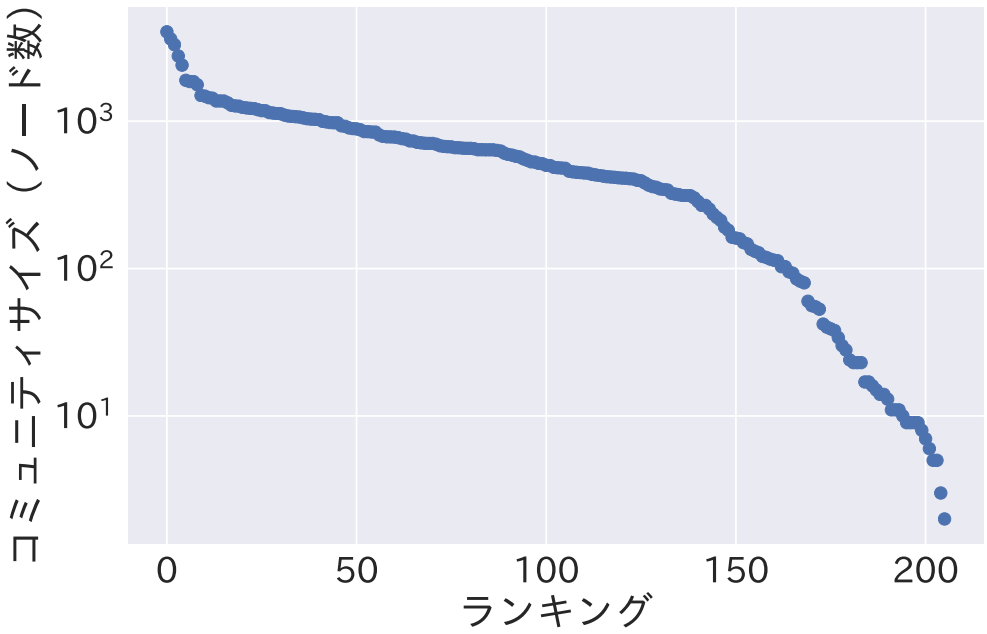


図 4.5: コミュニティサイズ分布

4.3.3　PHG のトライアドパターン分布

第 4.3.2 項で構築した静的な PHG のトライアドパターンの分布を**図 4.6** に示す．各トライアドパターンの出現頻度をトライアドの総和で除し，出現確率へと変換している．また，静的 PHG であるため 3 ノードが連結していない TP0 については除外している．これを見ると，TP3, 7, 8 が高頻度で出現していることが分かる．**図 4.4** に示すように，TP8 へは TP3 及び TP7 からしか推移できない．このことから，PHG の構造推移は TP3 及び TP7 から TP8 への TP 推移パターンが高頻度に出現すると考えられる．

4.3.4　PHG の TP 推移パターンの到達層分布

第 4.3.2 項で構築した PHG のトライアド集合から TP 推移パターンを取得した．これらの TP 推移パターンには，第 5 層である MP13 まで推移するもの第 3 層の MP8 までの推移に留まるものなどが存在し，到達する層はそれぞれ異なる．**図 4.7** に TP 推移パターンの到達層の分布を示す．縦軸が TP 推移パターン数，横軸が最終時刻おける到達層を示す．

まず，第 1 層，第 2 層の数が非常に多く支配的であり，逆に第 4 層の TP12，第 5 層の TP13 に到達しているトライアドはそれぞれ 250 組程度と少ないことが分かる．これは PHG 自体が最終時刻のスナップショットネットワークであり，成長途中のネットワークであるためである．第 3 層において，TP 推移パターンの終着点となりうるトライアドパターンは TP6,8,10,11 の 4 種類である．一方で，**図 4.6** に示すように，PHG において TP6, 10, 11 はほとんど出現しない．このことから，TP8 が第 3 層まで到達した TP 推移パターンの終着点のほとんどを占めていると考えられる．また同様に，第 2 層を構成するトライアドパターンは TP3,5,7,9 の 4 つであるが，PHG 全体のトライアドパターンをみると TP5,9 はほとんど出現しないことがわかる．加えて，**図 4.4** に示すように，TP3 及び TP7 は TP8 への経路にある．このことから第 2 層に到達したトライアドの多くは TP8 への成長過程にあると考えられる．

本研究では，ネットワーク構造の形成過程を分析することが目的である．第 4 層，第 5 層まで到達するトライアドが少なく，かつ第 4 層以降の TP 推移パターンは一通りである．また，上述したように第 1 層，第 2 層のトライアドは未だ第 3 層・TP8 への発展途上であると考えられる．形成過程を分析する上で，発展途上の部分ネットワークが混在するのは適切でない．以上を踏まえ，本研究では最終時刻までに第 3 層に到達した TP 推移パターンのみを分析対象とする．

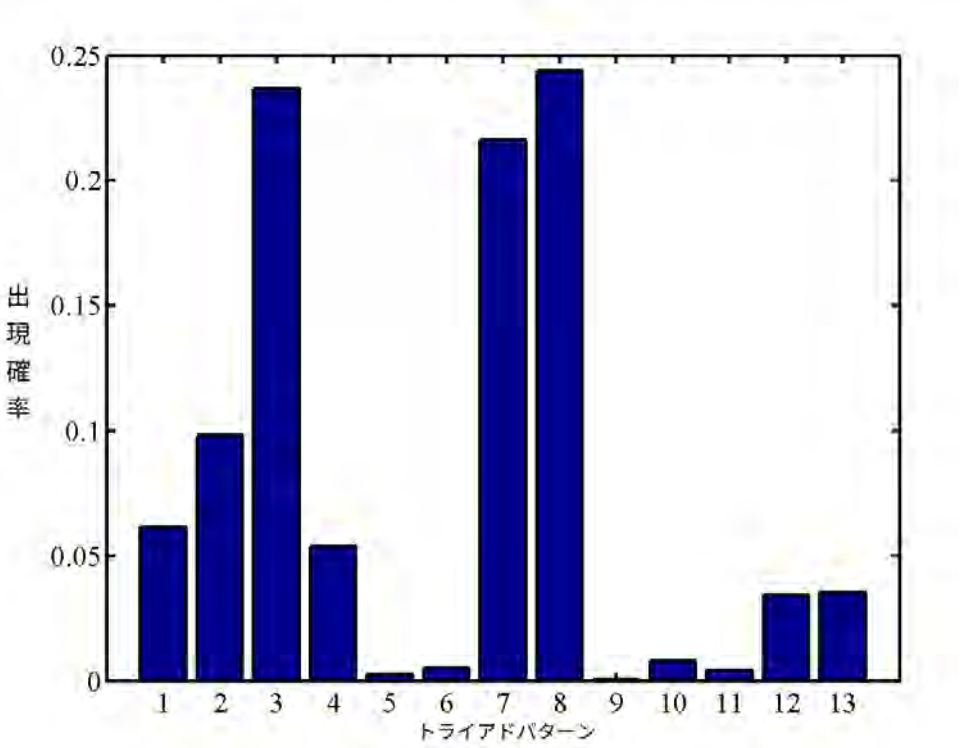


図 4.6: PHG のトライアドパターン分布

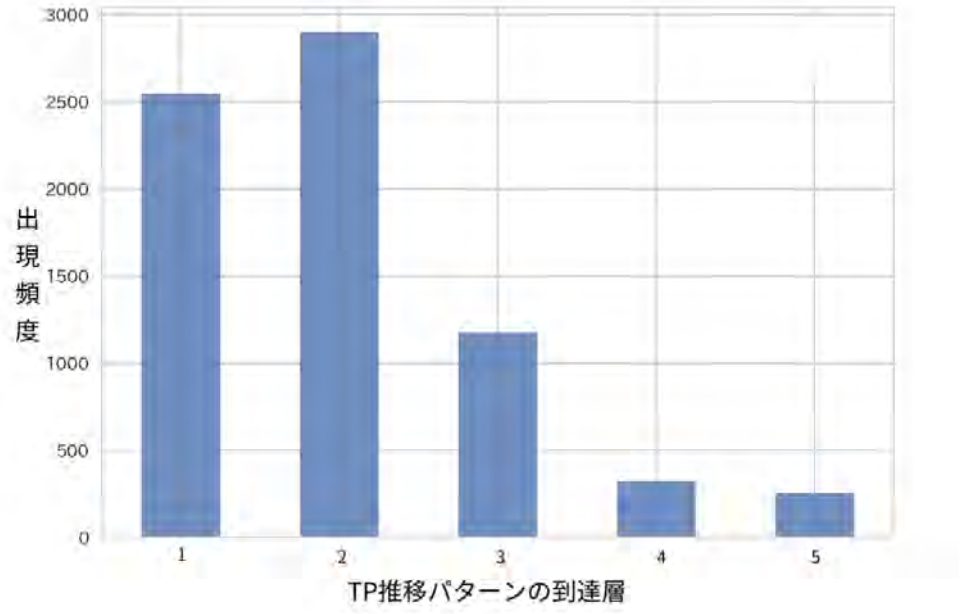


図 4.7: TP 推移パターンの到達層分布

4.4　評価実験

本節では実データを用いて構築した PHG に対し，提案手法を適用し，その有効性を確かめる．具体的には，各コミュニティにおける特徴的な TP 推移パターンが存在するかを偏在度を用いて確かめる．また，滞留度ベクトルを用いてクラスタリングを行い，TP 推移パターンにおける滞留傾向にどのような違いがあるか確かめる．また，滞留度ベクトルのクラスタリング結果と TP 推移パターンを照らし合わせることで，TP 推移パターン特有の滞留度が存在するか確かめる．

4.4.1　コミュニティにおける TP 推移パターンの偏在度

207 コミュニティそれぞれについて，**第 4.2.2 節**で説明した 28 種の TP 推移パターンがどのような分布で出現するか，Z スコアを用いて算出した．**図 4.8** にコミュニティ別の推移パターンの偏在度を示す．横軸が各推移パターン，縦軸が偏在度を表す Z スコア，各プロット線がコミュニティに対応している．なお，コミュニティ数が多いため，色の凡例については割愛する．一般に Z スコアの絶対値が 2 を超えると，全体の分布に対して出現数が偏っているといえる．207 コミュニティ中 194 コミュニティにおいて，いずれかの推移パターンの Z スコア絶対値が 2 を超えていることが確認された．このことから，コミュニティ毎に出現する TP 推移パターンには偏りがあり，ネットワークの形成過程はそれ

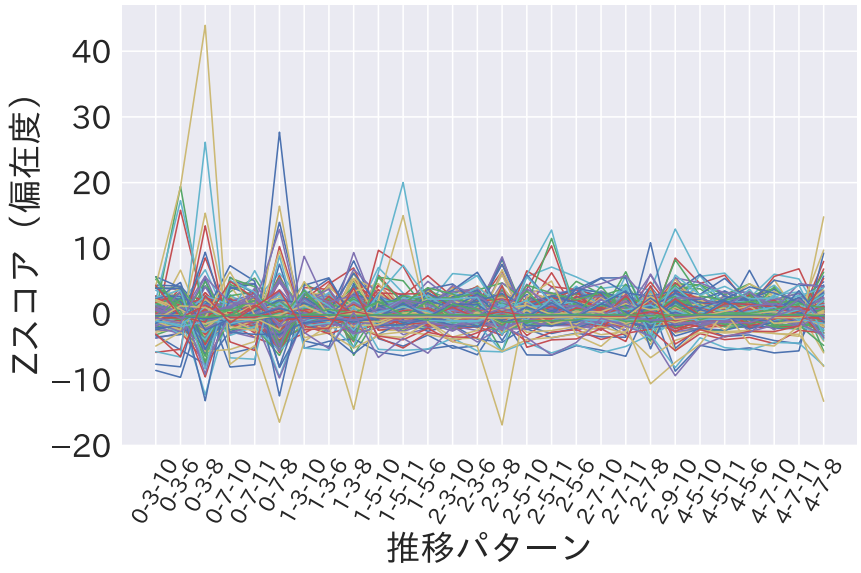


図 4.8: コミュニティにおける推移パターン偏在度

ぞれ異なるといえる．

4.4.2　滞留度ベクトルのクラスタリング

207 コミュニティそれぞれに対して，**第 4.2.3 節**で定義した 28 種の滞留度ベクトルを取得した．各コミュニティについてすべての TP 推移パターンが出現する場合，207 コミュニティ × 28 種 = 5,796 個のベクトルが取得できる．実際には，コミュニティによって出現しない TP 推移パターンがあるため取得できた滞留度ベクトルは 5,052 個であった．得られた 3 次元の滞留度ベクトルについて，k-means クラスタリングを行い，クラスタ数 K は，$K = 2$ から $K = 10$ までのシルエット分析を行い決定した．**図 4.9(a)** は，横軸をクラスタ数 K，縦軸を平均シルエット係数とした，平均シルエット係数の推移である．この図で，$K = 3$ の時に最も高い値 0.414 を示したことから，$K = 3$ 時のシルエット図を**図 4.9(b)** に示す．クラスタの大きさに偏りはあるものの，クラスタ間でのシルエット係数の差は小さい．これらの結果からクラスタ数は $K = 3$ とした．

クラスタ別の滞留度ベクトルを**図 4.10** に示す．横軸は各層，縦軸は各層における滞留度，プロット線は各滞留度ベクトルを表す．また，プロット線の色は推移パターンの種別に対応しているが，紙面の都合上，凡例は割愛する．これらを観察すると，いずれのクラスタも第 3 層の滞留度が最も大きい点が共通している[*4]．一方で，第 1 層，第 2 層の滞留度はクラスタ間である程度異なる．クラスタ 1 は比較的第 1 層の滞留度が大きく，初期構造に長く滞留する傾向にある．一方，クラスタ 2 は第 1 層，第 2 層にはほとんど滞留せず，すぐに第 3 層へと推移していることが分かる．クラスタ 3 は第 1 層より第 2 層に，第 2 層より第 3 層に長く留まることが分かる．

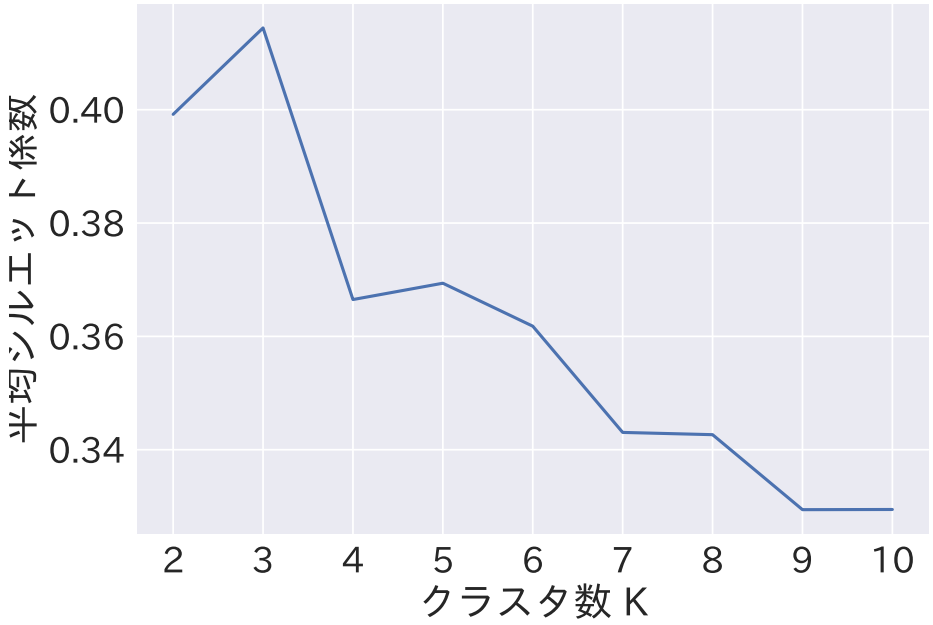
滞留度ベクトルに加えて，各クラスタにおける推移パターンの出現頻度分布を**図 4.11** に示す．横軸は各推移パターン，縦軸は推移パターンの出現数を示す．また，**図 4.12** にクラスタ別に，起点となる第 1 層のトライアドパターン，及び終点となる第 3 層のトライアドパターン毎の推移系列数を示す．分布の違いを明確にするため，合計が 1 になるように正規化を行った．横軸は各クラスタ，縦軸は推移系列の割合を示す．

図 4.12(a) から，クラスタ 1 はほぼ全て TP0 から始まる推移パターンであることが読み取れる．一方，クラスタ 2，3 に TP0 を起点とする推移パターンはごく僅かしか含まれていない．加えてクラスタ 1 は第 1 層の滞留度が大きい傾向にあった．以上のことから，TP0 を起点とする TP 推移パターンは第 1 層 (TP0) の滞留度が大きい傾向にあるといえる．

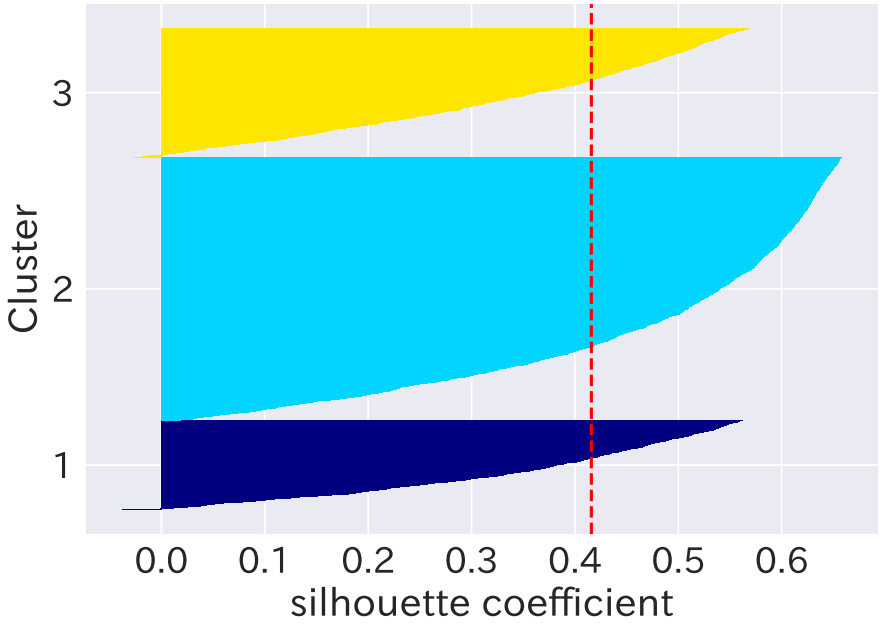
また**図 4.11(b)** 及び**図 4.11(c)** からは，クラスタ 2，3 の推移パター

[*4] 第 3 層の滞留度と全層の滞留数の和の関係を明らかにするために，両者の相関係数を求めたところ −0.008 であり，相関は見られなかった．

出現頻度の高低がちょうど逆であり，補完的な関係を示すことが読み取れる．**図 4.12(a)** においてクラスタ 2，3 間の分布はそれほど変わらない一方，**図 4.12(b)** においては TP8 に到達する TP 推移パターンの有無で差が生まれていることがわかる．

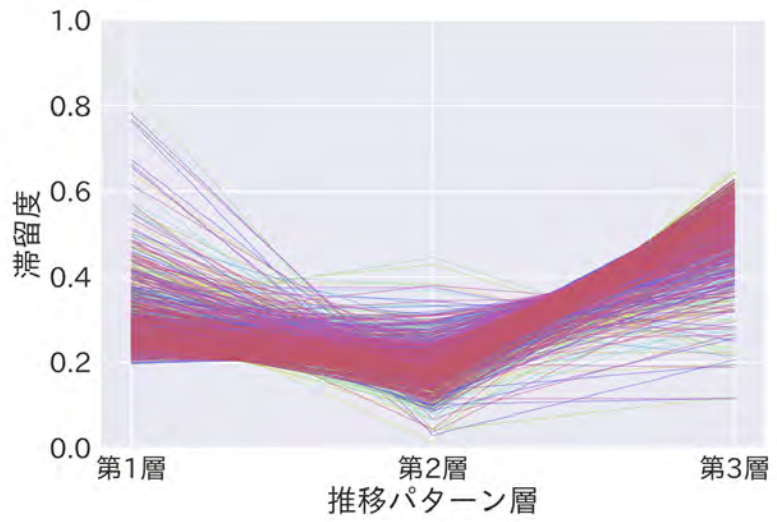


(a) 平均シルエット係数

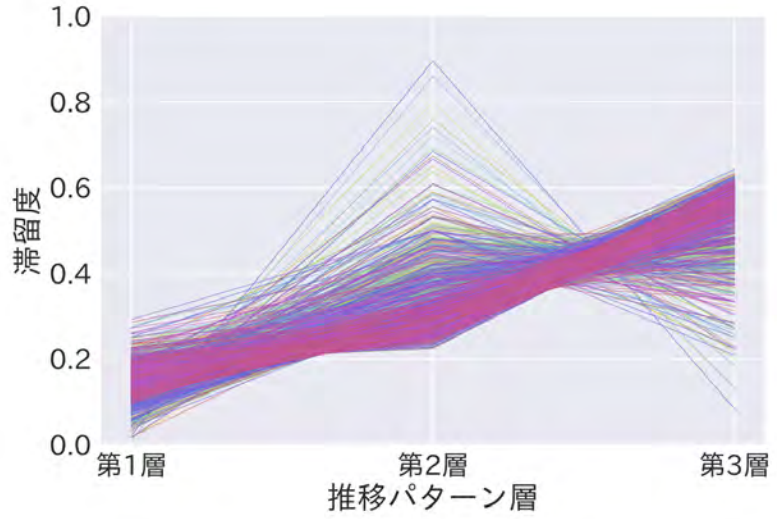


(b) シルエット図

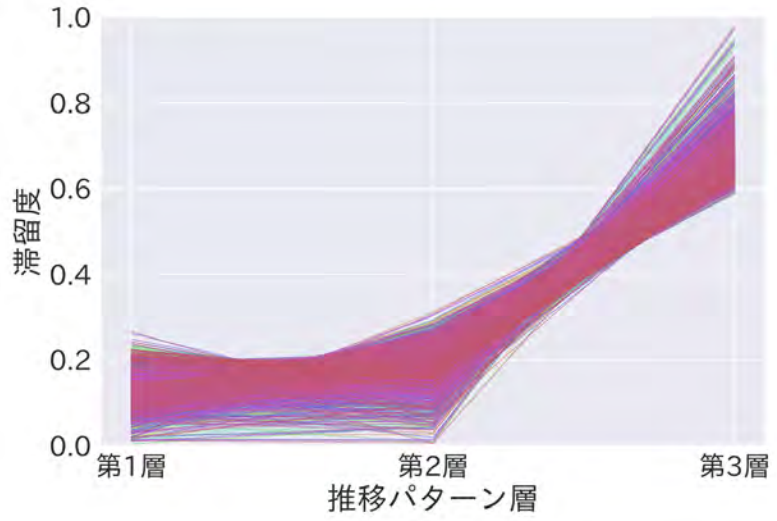
図 4.9: シルエット分析



(a) クラスタ 1

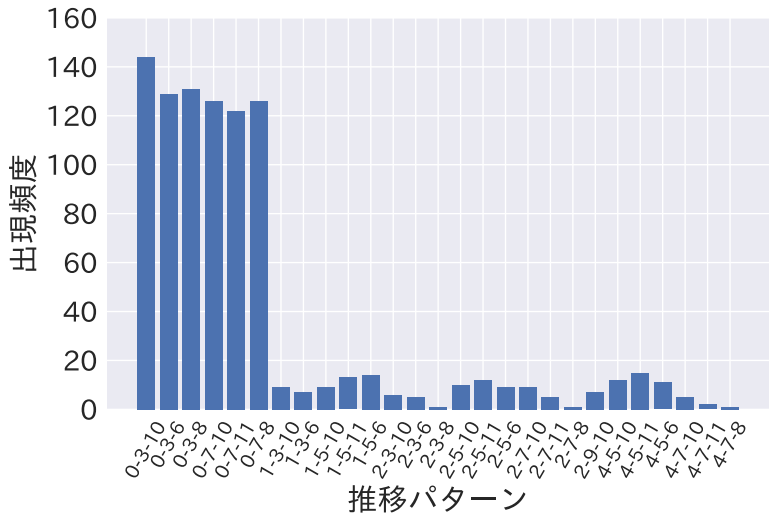


(b) クラスタ 2

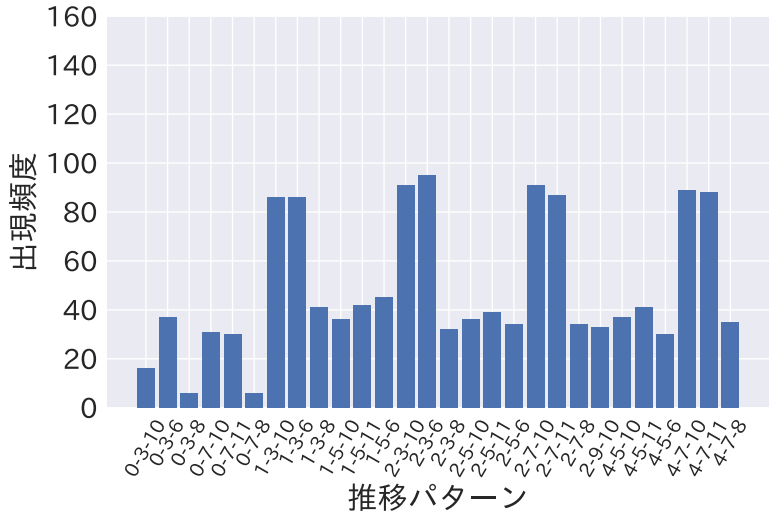


(c) クラスタ 3

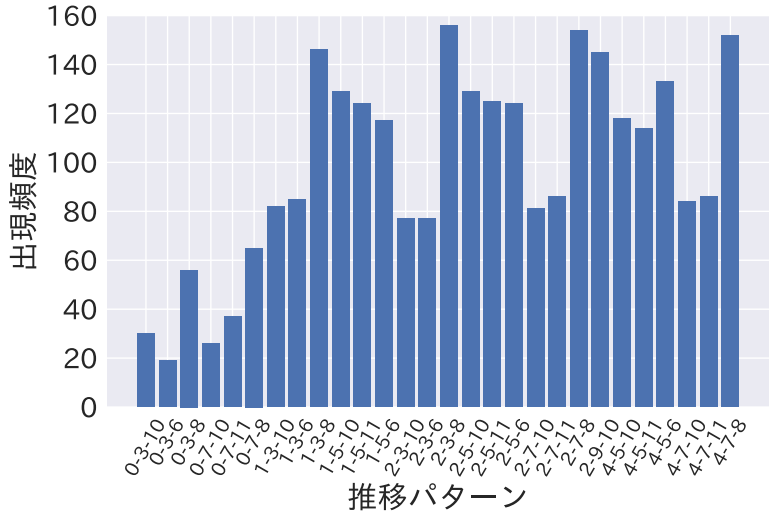
図 4.10: クラスタ毎の滞留度ベクトル



(a) クラスタ 1

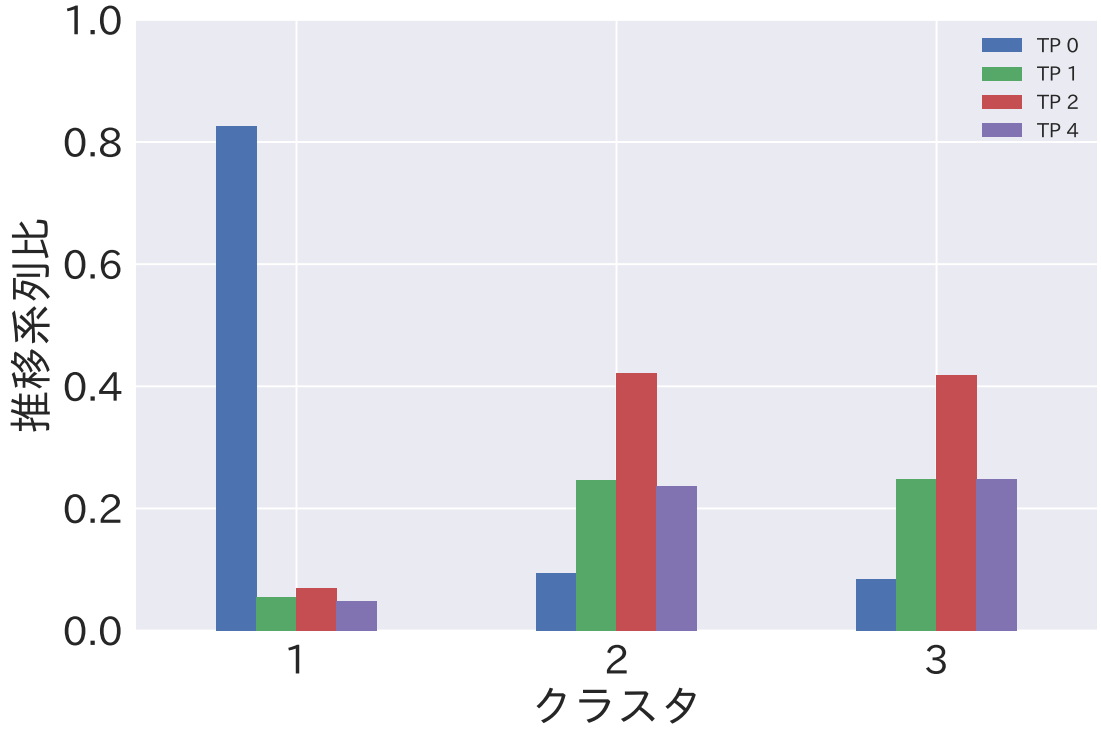


(b) クラスタ 2

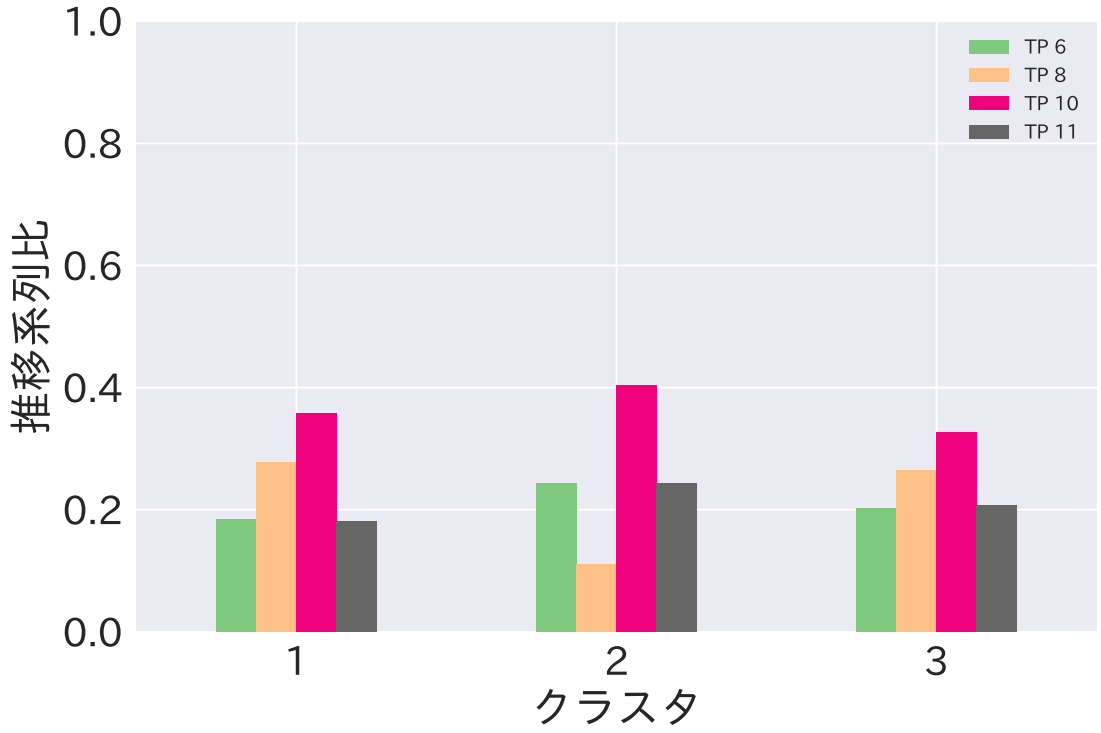


(c) クラスタ 3

図 4.11: TP 推移パターンの出現頻度



(a) 第 1 層のトライアドパターン



(b) 第 3 層のトライアドパターン

図 4.12: 第 1 層及び第 3 層のトライアドパターンの出現割合

4.4.3　クラスタにおける TP 推移パターンの偏在度

第 4.4.2 節で得られた 3 つのクラスタそれぞれにおける TP 推移パターンの偏在度を**図 4.13** に示す．横軸が推移パターン，縦軸が偏在度を表す Z スコア，各プロット線がクラスタに対応する．Z スコアが高ければ高いほど，全体の分布に対し，そのクラスタに多く出現（偏在）することを意味する．一般に Z スコアが 2 を超えると有意に多く出現しているといえる．これを見ると，**図 4.12(a)** で示した，TP0 を起点とする 6 つの TP 推移パターンがクラスタ 1 に多く出現していることが分かる．また，クラスタ 2 は TP8 に到達しないような TP 推移パターンが偏在していることがわかる．このように，異なる傾向の滞留度ベクトルから成るクラスタにおいて，頻出する推移パターンも有意に異なることから，TP 推移パターンによって滞留度には一定の傾向があることが定量的に示されたといえる．

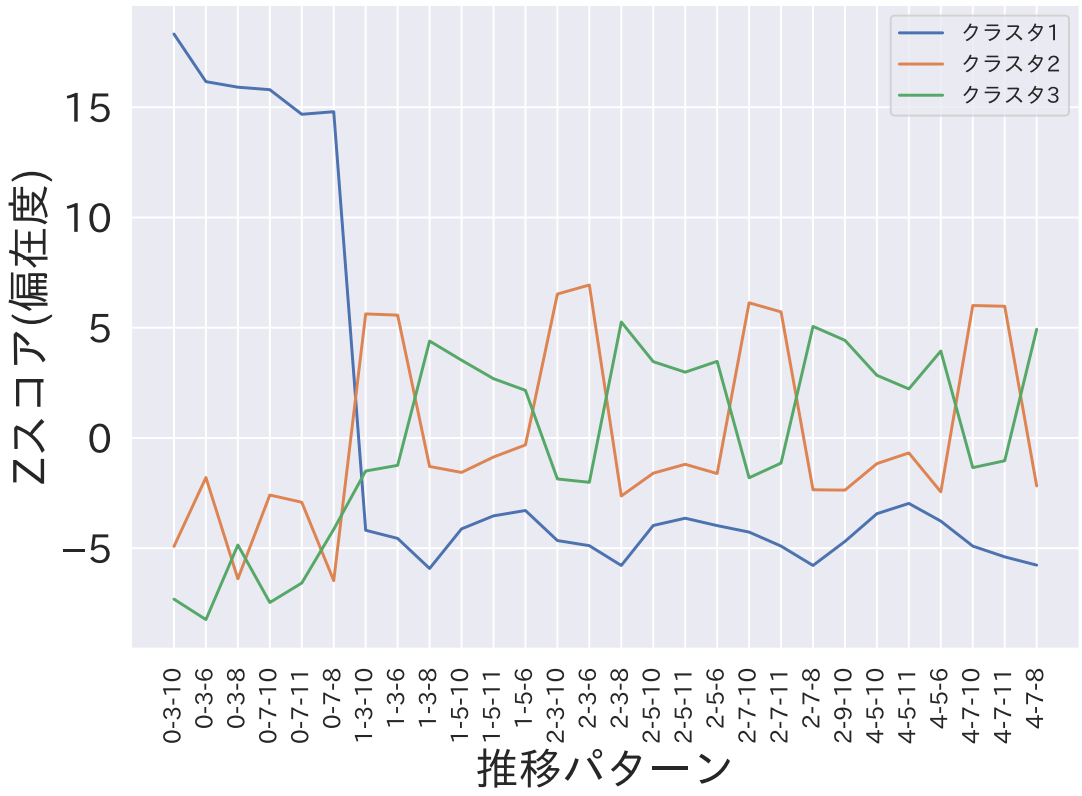


図 4.13: クラスタにおける推移パターン偏在度

4.5　考察

4.5.1　商品カテゴリ

ここでは，コミュニティ毎の TP 推移パターン及び滞留度の偏在度について，商品カテゴリを含めて議論する．

図 4.14 にコミュニティにおける推移パターンの偏在度に顕著な傾向が見られたものを一部抜粋して示す．横軸が各推移パターン，縦軸が偏在度を表す Z スコア，各プロット線がコミュニティに対応している．コミュニティに属するノード（アイテム）の属性からカテゴリを調査したところ，コミュニティ 23 及び 110 は主に食品カテゴリからなるコミュニティであった．

これらのコミュニティに出現する推移パターンは，TP6,10,11 を第 3 層とするが，途中経路となる第 1，2 層には大きな差異があった．例えば，コミュニティ 23 で最も高い偏在度を示す推移パターン 2-9-10 は，コミュニティ 110 においてマイナスの偏在度を示している．同じ食品カテゴリを主な構成要素とする 2 つのコミュニティが異なる偏在度を示した要因として，食品は価格帯が類似していること，また食品間の組み合わせも多様であることが，多様な購買行動の遷移を引き起こしたことが考えられる．

コミュニティ 54 と 66 はそれぞれ女性ファッションカテゴリと男性ファッションカテゴリからなるコミュニティである．**図 4.14** で両コミュニティの推移パターンの偏在度を観察すると，多くの推移パターンで類似した偏在度が観察される．しかし，コミュニティ 66(男性ファッション）においては，推移パターン 0-7-8 が極端に大きな値を示している．一方で，コミュニティ 54（女性ファッション）では，TP8 に至る 5 種の推移パターンが偏在度 5 近くを示す傾向にある．類似するファッションカテゴリでも，女性ファッションカテゴリに比べ，男性ファッションカテゴリのコミュニティには特徴的な成長傾向 (0-7-8) があると言える。

コミュニティ 139 及び 156 はインテリアカテゴリから成るコミュニティである．**図 4.14** で推移パターンの偏在度を見ると，コミュニティ 139 は推移パターン 0-3-6 の偏在度が高いことが分かる．一方，コミュニティ 156 では極端に高い偏在度を示す推移パターンはなく，コミュニティ 139 と 156 は大きく異なる推移パターンからなるといえる．

ここで，コミュニティの滞留度ベクトルに着目する．**第 4.4.2 節**で説明したように，滞留度は推移パターンの種別に大きく左右される．**図 4.15** にコミュニティ 139 及び 156 それぞれの滞留度ベクトルを示す．横軸が各層，縦軸が滞留度，プロット線の色は，TP0 起点かつ TP8 終点の推移パターンが緑，TP0 起点の推移パターンが青，TP8 終点の推移パターンが赤，その他の推移パターンが

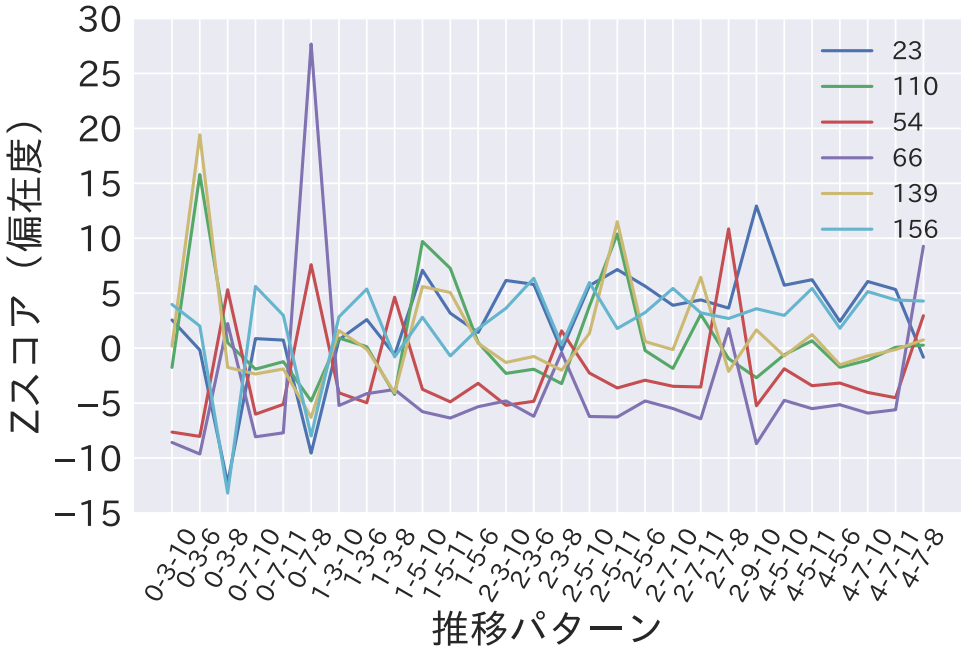
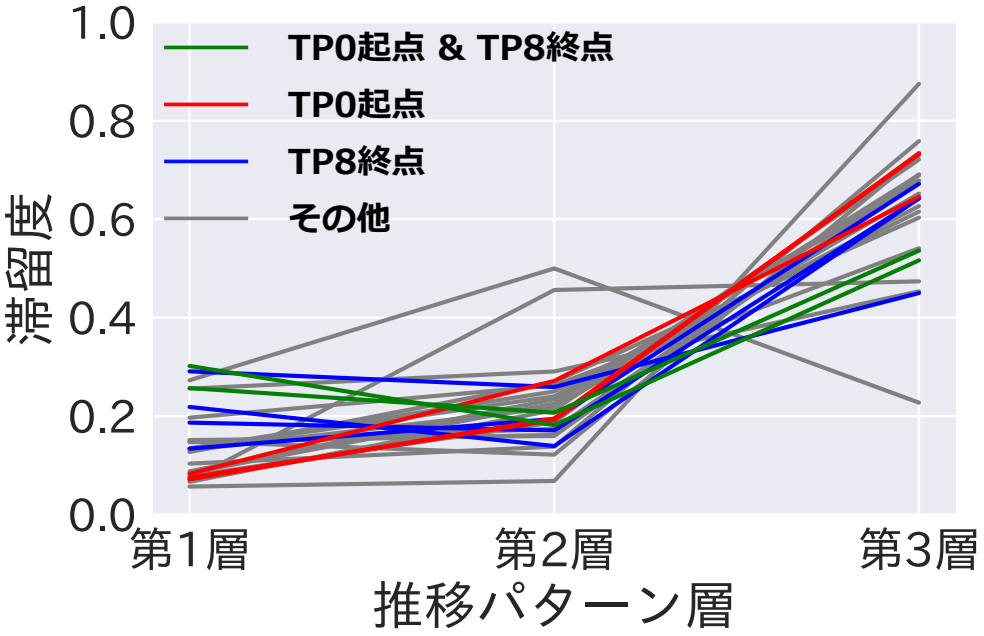
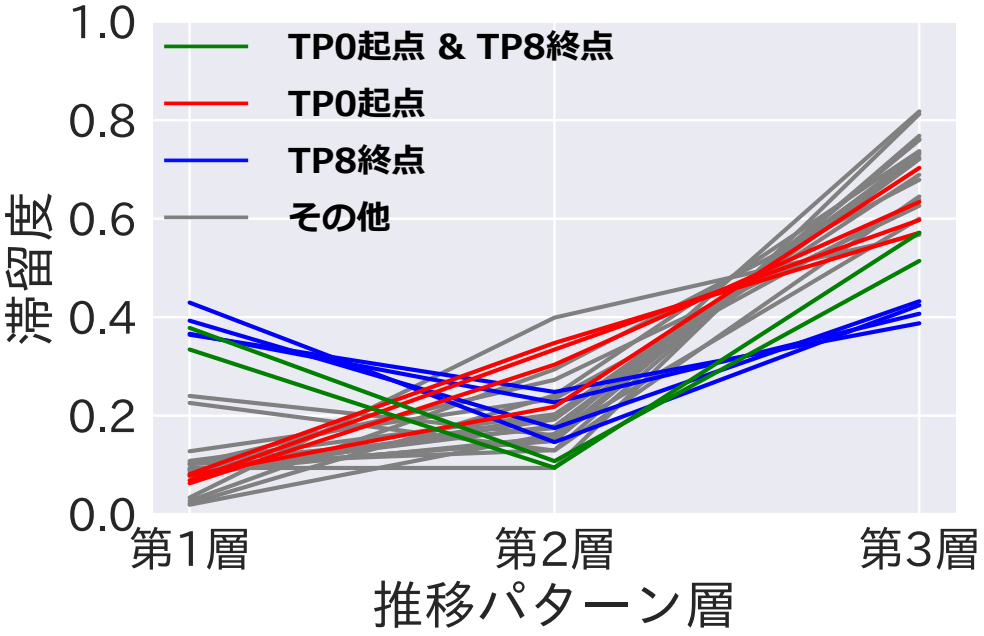


図 4.14: コミュニティにおける推移パターン偏在度（一部抜粋）

灰色に対応している．これを見ると，コミュニティ 139 に比べ，コミュニティ 156 の第 1 層への滞留度が大きい推移系列が多いことが分かる．これら 3 つのケースから，同じカテゴリのアイテムからなるコミュニティであっても，その成長傾向が異なる場合があることが示された．



(a) コミュニティ 139



(b) コミュニティ 156

図 4.15: コミュニティにおける滞留度ベクトル

4.5.2　グラフ構造の可視化

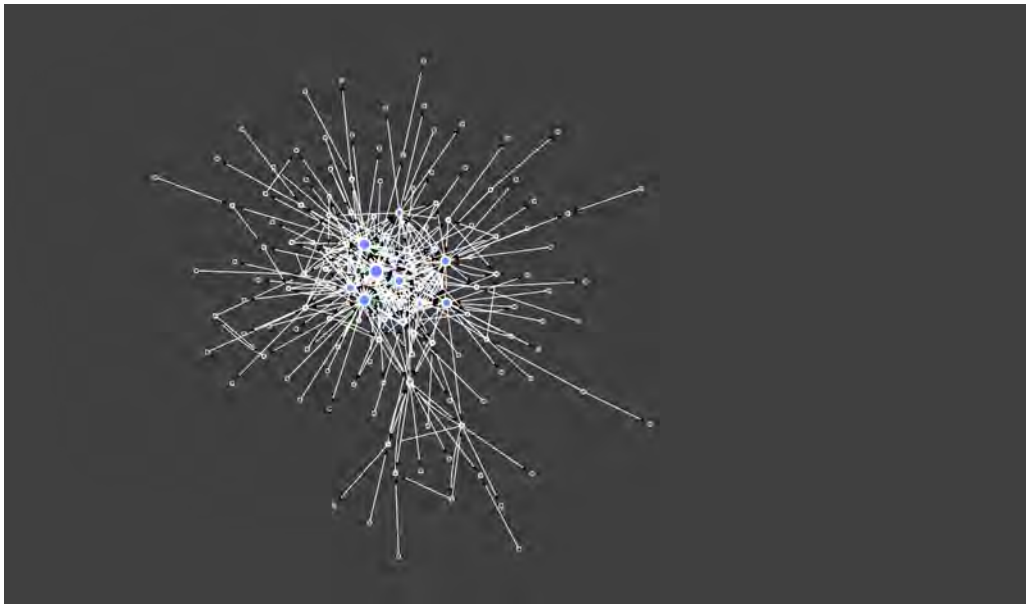
コミュニティ 139 及び 159 について，グラフの可視化を行った．一部抜粋した結果を**図 4.16** に示す．ノードの大きさは次数に対応している．コミュニティ 139 の可視化結果からは，**第 4.1 節**で述べた**図 4.1(a)** のように，1 つのアイテムを中心として複数のアイテムがリンクする構造が観察される．これは一つのメインアイテムに対し，ユーザの嗜好に合わせた複数のオプションアイテムが存在する構造である．このため，複数のユーザがメインアイテムを購入する際に，個々人の嗜好に合ったオプションアイテムを同時購入する機会が多いと考えられる．そのため早い段階で複数のノードとリンクする，すなわちトライアドパターンが推移するので，第 1 層への滞留度が小さいと考えられる．一方，コミュニティ 156 の可視化結果からは，**図 4.1(b)** のように，長いパスが複数観察される．これは，アイテム同士の購買関係が明確に決まっている場合に見られる構造であり，多くのユーザが特定の組み合わせや特定の順序でアイテムを購買する．この場合には，アイテム間のリンクが限定的となるため，新たなノードとリンクすることによるトライアド推移が発生しにくく，結果として第 1 層の滞留度が大きくなったと考えられる

これを推薦に応用すると，前者のようなアイテムは成長初期の段階から推薦するアイテムをなるべく多様にすることで，購買実績を重ね，さらなる購買を誘発できると考えられる．逆に後者のようなアイテムは，推薦するアイテムを限定的にし，既に購買実績のあるアイテムを中心とした推薦が効果的だと考えられる．特に，長いチェインを構成するようなアイテム群については，チェインの開始アイテムを購入したユーザに対する「まとめ買い」推薦なども有効であろう．チェインを構成するアイテムは，ユーザの嗜好に依らないアイテムであり，購買時の満足度も高いと期待できる．「まとめ買い」により購買実績が伸びることで前者同様さらなる購買を誘発できると考えられる．以上のことから，アイテム推薦や販売戦略策定等の場面において，提案手法が有効に機能することが期待できる

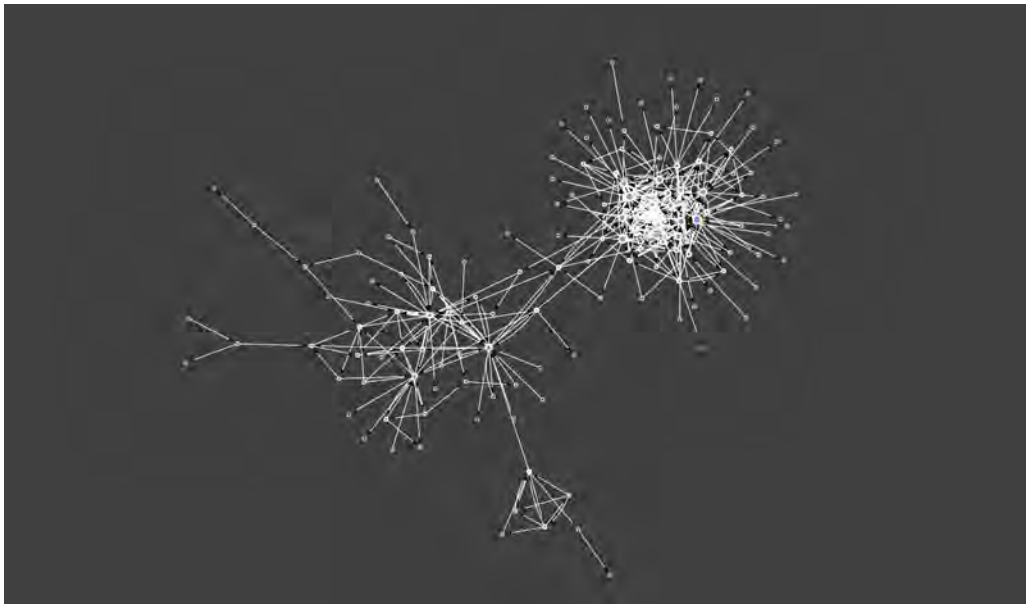
4.6　まとめ

提案手法では，動的に購買履歴グラフを生成し，推移パターンの偏在度の算出や滞留度ベクトルのクラスタリングを行っている．動的に成長する購買履歴データに適用し，類似するアイテムカテゴリでも異なる成長傾向を示すことを確認した．

オンラインショッピングサイトの購買履歴データを対象とした分析では，多くのユーザの購買履歴データを用いることでアイテムがどのような順序で購入



(a) コミュニティ 139



(b) コミュニティ 156

図 4.16: ネットワーク構造の可視化結果 (一部抜粋)

されるのか，というアイテム間の関係を抽出することができる．著者らはこれまで，購買履歴から構築する PHG を対象に「アイテムの買われる順序」に着目し，典型的な購買関係の抽出に取り組んできた．

しかし，同じような商品や同じ購買関係を示していても，それらが形成されるまでの過程が異なれば商品推薦や販売戦略も異なる．本研究では，この購買関係が形成されるまでの過程に着目し，トライアドの推移パターン及び滞留度に着目した購買行動の成長分析手法を提案している．ネットワーク成長過程に

対して，14 種類のトライアドパターンの推移傾向及び各構造における滞留度を算出する．これにより，商品カテゴリや TP 推移パターンによる推移傾向や滞留度の違いを明らかにした．

楽天市場のレビューデータから構築した PHG を対象に TP 推移パターンや滞留度の分析を行い，コミュニティ毎に特定の TP 推移パターンが顕著に表れることや，商品カテゴリと TP 推移パターン・滞留度の関係性について議論した．

本研究の貢献は次の通りである．1)PHG の形成過程を分析する手法として TP 推移パターンと滞留度を提案した．2) 提案手法を実データに適用し，頻出する TP 推移パターンの抽出等を行い，その有効性を示した．

また，提案手法はノード同士の関係性を分析する手法であることから，評価実験で用いた PHG 以外のネットワークにも適用可能である．具体的には SNS のフォローネットワークなどの友人関係や共著ネットワークなどが挙げられる．フォローネットワークであれば，フォロー関係やコミュニティの成立過程を分析し，例えば中心的な人物のもとに発達したコミュニティであるといった知見が得られるだろう．また同様に共著ネットワークを対象にすれば，共著関係がどのように成立していくのかを明らかにできる．

今後の課題としては，滞留度ベクトルの算出手法の改良が挙げられる．本研究ではトライアドの構造を変化させるエッジが 1 本でも追加された場合，トライアドが次層に推移したと定義して，各層の滞留度を算出している．3 つのノード a,b,c からなるトライアド $triad_{abc}$ を考える．今，$triad_{abc}$ には $a \to b$ と $a \to c$ の 2 種のエッジが付与されており，トライアド構造は第 1 層の TP1 である．ここで，新規に $b \to c$ のエッジが 1 本付与されたとすると，構造は第 2 層の TP5 へと推移する．さらに，$a \to c$ のエッジが続けて 5 回付与された場合，この 5 本のエッジは第 2 層 TP5 の滞留度として計上される．しかし，このケースでは TP1 に 5 回滞留したと計上する方が実際に即した結果になると考えられる．このように滞留度ベクトルの算出においてトライアド推移を判定する閾値を動的に設定することで，より現実に即した分析が可能になると期待できる．

第 5 章

考察

5.1 ネットワーク成長におけるエッジの影響力

第 3 章では，ネットワーク成長におけるエッジ影響力の定量化に取り組んだ．ネットワークの成長過程では，1 つのエッジの出現が他のノードに影響を与え，新たなエッジの出現を促す．ここでは，あるエッジが起点となって生まれる後続するエッジの量をエッジの影響力とみなし定量化する STM(Stimulation Index) を提案した．人工的に生成したフォローネットワークを用いて情報拡散シミュレーションを行い，高い STM を持つエッジを削除することで，情報拡散が阻害されることを示した．これにより STM が情報拡散における重要度指標として有効に機能することを確かめた．また，評価実験は情報拡散の文脈で行ったものの，STM はエッジの出現が新たなエッジ出現を促すような連鎖的に変化するネットワークに対して汎用性がある．提案手法が適用可能なネットワークとして**第 4 章**で扱った PHG が挙げられる．これはある商品の購入が新たな商品の購入を促すネットワークであることから，連鎖的な購買を生み出すことのできる商品推薦や販売戦略の策定に応用が可能である．また論文の引用ネットワークは新たな論文・引用関係が出現することで連鎖的に成長するネットワークであり，STM が適用可能である．具体的には，議論を活性化させたきっかけを評価する指標として用いることが考えられるほか，直接引用数だけでなく論文の間接的な影響力を評価することも可能である．

エッジ出現時の特徴量を用いたネットワーク成長後の STM スコアの推定に取り組んだ．ネットワークは常にその構造を変化させているため，比較的軽量に計算可能な特徴量として k-DGC を提案した．k-DGC 及び各種中心性を説明変数，STM スコアを目的変数とした重回帰分析を行ったところ，Mention-NW と Retweet-NW を対象とした重回帰モデルにおいて決定係数 R^2 は約 0.7 であり，STM スコアの推定が可能であることを示した．また，提案特徴量である k-DGC が推定に大きく貢献していることを確かめた．

今後の課題としては，STM スコアの推定モデルの性能向上が挙げられる．本

研究では，まずSTMの有効性を示すことに主眼をおき，STMスコアの推定可能性の検討を行った．その結果，シンプルな重回帰モデルによって一定の性能を出せている．説明変数として用いるエッジ特徴量や推定モデルの検討を深めることで，より精度の高い推定モデルの構築が可能だと考えられる．

5.2　ネットワーク成長における構造推移の定量化

第 4 章では，ネットワーク成長における構造推移の定量化に取り組んだ．ネットワークの成長過程では，同じ構造を持っていてもそこに至るまでの推移は異なることがある．逆に全く異なる構造を持っていても，そこに至るまでの推移，あるいは推移の一部には類似する特徴が見られることもある．このような構造推移を分析する手法として，本研究ではトライアド推移パターンを用いる手法を提案した．実際のオンラインショッピングサイトの購買履歴データから，PHGを構築し典型的な購買順序パターンの抽出を行った．その上で，それらの典型的なパターンや商品カテゴリについて，推移パターンの比較を行った．結果，商品カテゴリ固有のトライアド推移パターンが抽出できたほか，逆に様々な推移が混在するカテゴリの存在などを明らかにし提案手法の有効性を示した．また，提案手法はノード同士の関係性を分析する手法であることから，評価実験で用いたPHG以外のネットワークにも適用可能である．具体的にはSNSのフォローネットワークなどの友人関係や共著ネットワークなどが挙げられる．フォローネットワークであれば，フォロー関係やコミュニティの成立過程を分析し，例えば中心的な人物のもと発達したコミュニティであるといった知見が得られるだろう．また同様に共著ネットワークを対象にすれば，共著関係がどのように成立していくのかを明らかにできる．

5.3　提案手法の到達点

現実のネットワークの多くは，その構造を常に変化させ成長するネットワークである．このネットワーク成長は，エッジやノードが出現する「拡大」と逆に消失する「縮退」の 2 つの変化に分けられる．また構造変化の発生する時間軸も，ステップ単位の変化として離散的に扱うものと，実時間と同様に連続的に扱うものの 2 つに分けられる．SNSのフォローネットワークにおいて，新たなフォローの発生は「拡大」であり，逆にフォローの解除は「縮退」にあたる．また，時間軸についても，フォロー・アンフォローの発生順序だけに着目した場合は離散的，実際の発生時刻に着目した場合は連続的に扱うケースに当たる．

本研究においては，このネットワーク成長のうち，離散的な時間軸におけるネットワークの拡大について扱っている．本研究で対象とするネットワーク成長は，「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）を

どの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」の 2 つの要素から構成される．これら 2 つの要素を本研究における課題として捉え，それぞれに対する解決法を提案した．第一の課題，ネットワーク構造全体の変化における個別の変化の影響力は，STM によって明らかにできる．第二の課題，ネットワークがどのように構造変化したのかはトライアド推移を用いて明らかにできる．両手法を使い分ける，あるいは併用することによって，ネットワーク成長における連鎖的な構造変化の分析において有用な結果を得られると期待できる．

例えば，PHG におけるアニメーショングッズの購買を考える．ここでは，キャラクターのグッズやアニメーションの Blu-ray Disk，原作となるコミックスなど様々な商品が存在する．Blu-ray Disk やコミックスなどは購買順序が明確であるため，その形成過程は 1 巻チェインが伸びていく様としてトライアド推移に表れるだろう．一方，キャラクターのフィギュアや缶バッジなどは，多くのユーザーが好むメインキャラクターを中心とした少数の密なネットワークが成立したのち，ユーザーによって好みが分かれる多数のサブキャラクターへの購買が生まれる構造がトライアド推移により観測できると期待できる．また，このネットワーク成長において，コミックスのコミュニティでは 1 巻，2 巻の購買がその後の連鎖的な購買を大きく誘発しており，高い STM スコアを示すだろう．キャラクターグッズのコミュニティではメインキャラクター同士の購買関係が高い STM スコアを示すと考えられる．これは，メインキャラクターのグッズ購入が，ユーザーの購買意欲を掻き立て，サブキャラクターのグッズ購入を誘発していると考えられるためである．

本研究で提案する 2 手法;STM とトライアド推移により，ネットワーク成長における影響力の高いエッジの抽出，及び構造変化の理解を達成できる．これにより，一連のネットワーク成長について概観できたといえる．

5.4　提案手法の限界

第 5.3 節で説明したように，本研究では，離散的な時間軸におけるネットワークの拡大を対象としている．すなわち，ネットワークの縮退及び連続的な時間軸については検討できていない．しかしこの点についても提案手法の拡張により，適用可能になると考えている．

まずネットワークの縮退について，エッジの出現と同様に，エッジの切断も連鎖的に発生すると考えられる．例えば，友人関係において二者の関係が悪化することにより，コミュニティ自体が疎遠になってしまうということは実際に起きうる事象である．このとき，STM でエッジ消失の連鎖を定量化することで，コミュニティが疎遠になった要因として最初の二者の関係悪化が抽出できるだろう．また，トライアド推移についても，推移の向きを逆転させることで，

エッジ数の少ないトライアドパターンへとどのように推移していくのか，ネットワークの縮退過程が抽出できる．ただしネットワーク縮退を扱う上で課題となるのが，明示的な縮退と暗黙的な縮退の扱いである，SNS のフォローネットワークにおいて，新たなフォロー関係の誕生は明示的な「フォロー」によって観測される．同様に，フォロー関係の解消は明示的な「アンフォロー」によって観測される．しかし実際には，「アンフォロー」に加えて「フォロー関係にはあるが今はほとんど交流がない」というような，暗黙的なフォロー関係の解消があり得ると考えられる．このような暗黙的な縮退も含めて扱うことで，提案する 2 手法はネットワーク縮退に対して有効に機能すると考えられる．

次に，連続的な時間について説明する．STM は後続するエッジの発生をどの程度誘発したか，という観点で影響力を定量化したものである．本研究では，評価実験に用いたデータの取得期間が 1 週間程度と比較的短いこともあり，連続的な実時間を考慮していない．しかし実際には，影響力は強いほどエッジ出現を誘発するまでの時間が短く，影響力が弱くなると誘発するまでの時間も長くなると考えられる．また，トポロジ上では誘発判定にあっても，実時間を鑑みるとそう解釈できないこともあるだろう．そのため STM では，エッジの出現時間を考慮することでより現実に即した影響力の定量化が可能になると考えられる．また，トライアド推移における滞留度は，離散的なステップ数に基づいて算出している．同じトライアドパターンを維持した実時間を計測することで，連続的な実時間に基づいた特徴量にできる．このように提案手法自体は連続的な時間に対しても拡張可能な手法である．ネットワークの性質や分析の目的に合わせて，また離散的な時間と連続的な時間を切り替えることで，提案手法はより有効に機能すると考えられる．

第 6 章

結論

6.1　本研究のまとめ

情報化社会の発展や計算機の性能向上により，近年ではますます大規模なデータが収集・利活用されるようになった．これに伴い，ネットワーク分析の分野においてもその動的な性質そのものを捉える手法が必要とされている．ネットワークは人物や駅などのノード及びノード間の繋がりを，エッジによって表現したものである．そして，ネットワーク構造の変化は，あるノードが他のノードへと何らかの影響を与え繋がりをもつこと，すなわちエッジが出現するということと同値である．さらに出現したエッジの影響は，エッジ両端のノードだけに留まらない．影響を受けたノードは，影響を与えるノードとして他のノードへとその影響を伝播させ，新たなエッジの出現を促す．この影響の伝播とエッジ出現の連鎖により，ネットワーク構造全体が変化していくのである．すなわち，ネットワーク成長は「ネットワークの変化（エッジの出現）が新たな変化（エッジの出現）をどの程度引き起こすか」と「ネットワーク構造全体がどのように変化していくか」の 2 つの要素から構成されるといえる．これら 2 つの要素を本研究における課題として捉え，それぞれに対する解決法を考究した．

第 1 の課題は，ネットワーク成長におけるエッジ影響力の定量化である．上述したように，ネットワークの成長過程では，1 つのエッジの出現が他のノードに影響を与え，新たなエッジの出現を促す．本研究では，情報カスケード及び情報拡散モデルの 1 種である線形閾値モデルの考え方を用いて，あるエッジが起点となって生まれた後続するエッジの量をエッジの影響力とみなし定量化する Stimulation Index を提案した．高影響なエッジを抽出し，ネットワークの変化に貢献するエッジを明らかにすることは，ネットワーク成長を理解するために重要である．また，迅速な情報拡散，バイラルマーケティングの効果検証をはじめとして，様々な応用も期待できる．人工ネットワークを用いた情報拡散シミュレーション，及び実際の情報拡散データセットを用いて提案手法の有効性を確認した．

第 2 の課題について，ネットワーク成長における構造推移の定量化である．ネットワークの成長過程では，同じ構造を持っていても，そこに至るまでの推移は多様であり，得られる知識や解釈もまた異なる．例えば，友人関係の構築過程において，顔の広い人物によって出会うのか，小さな仲良しグループが合流していく形を取るのか，といった違いが見られるだろう．逆に全く異なる構造を持っていても，そこに至るまでの推移，あるいは推移の一部には類似する特徴が見られるということもあるだろう．ネットワーク構造がどのように変化していくのか定量的に評価することは，異なる構造推移の類似度計算を始めとしてネットワーク成長の理解・議論に欠かせない．本研究では，トライアド（連結 3 ノードからなる最小の有向ネットワーク．13 種類存在する．）の推移を用いて，動的な構造変化の定量化を試みた．ネットワーク中のトライアド 13 種の分布を調べることで，その構造的特徴を把握することができる．またネットワークが変化すると，当然トライアド自体も別種のトライアドへと推移する．このトライアドの推移パターンは 28 種存在する．これを数え上げることによってネットワークの構造推移を分析する．商品の購買順序関係を示すネットワーク:PHG（Purchase History Graph）を分析し，商品カテゴリ固有のトライアド推移パターンが抽出できたほか，逆に様々な推移が混在するカテゴリの存在などを明らかにし，提案手法の有効性を示した．

第一の課題，ネットワーク構造全体の変化における個別の変化の影響力は STM によって明らかにできる．第二の課題，ネットワークがどのように構造変化したのかはトライアド推移を用いて明らかにできる．本研究で提案した両手法を使い分ける，あるいは併用することによって，ネットワーク成長における連鎖的な構造変化の分析において有用な結果を得られると期待できる．

今後の課題として，ネットワークの縮退や連続的な時間軸を持つネットワークに対する拡張が挙げられる．本研究では，離散的な時間軸におけるネットワークの成長を対象としている．提案手法を拡張することで，より現実的なネットワーク成長の分析が可能になると期待できる．

謝辞

本研究の遂行及び論文の作成にあたり，筑波大学図書館情報メディア系・佐藤哲司教授には学部 3 年生の頃から主指導教員として懇切なるご指導とご鞭撻を賜りました．ここに深く感謝の意を表します．また，同系の芳鐘冬樹教授，鈴木伸崇教授には副指導教員としてご指導いただき，感謝申し上げます．さらに，同系の加藤誠准教授，筑波大学計算科学研究センターの天笠俊之教授には学位論文審査委員を快く引き受けていただき，多くの助言を賜りました．感謝申し上げます．

また，本論文の核となる原著論文の共著者でもある東京工科大学の伏見卓恭講師には日頃のゼミはもとより，実験から論文執筆，発表に至るまで様々な面でご支援いただきました．心よりお礼申し上げます．

また，7D140 研究室のメンバーとは，雑談から研究の議論まで多くのことを共に語らい，楽しい学生生活を送ることができました．ありがとうございました．

たくさんの方のご支援により，本論文をまとめることができました．ありがとうございました．

参考文献

[1] 中田豊久，加藤義彦，國藤進，“友人ネットワークの状態遷移図による分析.”情報処理学会論文誌数理モデル化と応用（TOM），vol. 2，no. 1，pp. 87–97，February 2009.

[2] L. Peel, and A. Clauset, “Detecting change points in the large-scale structure of evolving networks,” *CoRR*, vol. abs/1403.0989, pp. 2914–2920, 2014.

[3] A. Clauset, C. Moore, and M. E. J. Newman, “Hierarchical structure and the prediction of missing links in networks,” *Nature*, vol. 453, pp. 98–101, 2008.

[4] A. Nigam, K. Shin, A. Bahulkar, B. Hooi, D. Hachen, B. K. Szymanski, C. Faloutsos, and N. V. Chawla, “One-m: modeling the co-evolution of opinions and network connections,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 122–140.

[5] H. Dai, Y. Wang, R. Trivedi, and L. Song, “Deep coevolutionary network: Embedding user and item features for recommendation,” *arXiv preprint arXiv:1609.03675*, 2016.

[6] D. J. Watts, and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, January 1998.

[7] *The Small-World Phenomenon: An Algorithmic Perspective*, 2000.

[8] S. Wasserman, and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[9] A. Albert, and A. L. Barabási, “Statistical mechanics of complex networks,” *Rev. Mod. Phys*, pp. 47–97, 2002.

[10] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: simple building blocks of complex networks.” *Science (New York, N.Y.)*, vol. 298, no. 5594, pp. 824–827, October 2002.

[11] A. Bavelas, “Communication patterns in task-oriented groups,” *The journal of the acoustical society of America*, vol. 22, no. 6, pp. 725–730, 1950.

[12] L. Freeman, “Centrality in social networks: Conceptual clarification,” *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.

[13] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, pp. 35–41, 1977.

[14] P. Bonacich, “Power and Centrality: A Family of Measures,” *The American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, March 1987.

[15] M. E. J. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.

[16] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, pp. 39–43, 1953.

[17] S. Brin, and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer Networks and ISDN Systems*, vol. 30, pp. 107–117, 1998.

[18] A. N. Langville, and C. D. Meyer, “Deeper Inside PageRank,” *Internet Mathematics*, vol. 1, no. 3, pp. 335–380, 2004.

[19] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *J. ACM*, vol. 46, pp. 604–632, September 1999.

[20] T. Akiba, Y. Iwata, and Y. Yoshida, “Fast exact shortest-path distance queries on large networks by pruned landmark labeling,” in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 2013, pp. 349–360.

[21] U. Brandes, “A Faster Algorithm for Betweenness Centrality,” *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.

[22] U. Brandes, and C. Pich, “Centrality estimation in large networks,” *International Journal of Bifurcation and Chaos*, vol. 17, no. 07, pp. 2303–2318, 2007.

[23] K. Okamoto, W. Chen, and X.-Y. Li, “Ranking of closeness centrality for large-scale social networks,” in *International workshop on frontiers in algorithmics*. Springer, 2008, pp. 186–195.

[24] P. Crescenzi, C. Magnien, and A. Marino, “Finding top-k nodes for temporal closeness in large temporal graphs,” *Algorithms*, vol. 13, no. 9, p. 211, 2020.

[25] 林孝紀，秋葉拓哉，吉田悠一，“動的なネットワークにおける媒介中心性の高速計算手法”，in 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM2015)，2015.

[26] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, “Eigenvector-based centrality measures for temporal networks,” *Multiscale Modeling & Simulation*, vol. 15, no. 1, pp. 537–574, 2017.

[27] C. Magnien, and F. Tarissan, “Time evolution of the importance of nodes in dynamic networks,” in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2015, pp.

1200–1207.

[28] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, vol. 100, no. 5, pp. 992–1026, 1992.

[29] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.

[30] D. J. Watts, and P. S. Dodds, "Influentials, networks, and public opinion formation," *Journal of consumer research*, vol. 34, no. 4, pp. 441–458, 2007.

[31] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," in *International conference on knowledge-based and intelligent information and engineering systems*. Springer, 2008, pp. 67–75.

[32] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545–576, 2012.

[33] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," in *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012, pp. 918–923.

[34] S. Galhotra, A. Arora, S. Virinchi, and S. Roy, "Asim: A scalable algorithm for influence maximization under the independent cascade model," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 35–36.

[35] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *2010 IEEE international conference on data mining*. IEEE, 2010, pp. 88–97.

[36] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," *Knowledge and information systems*, vol. 37, no. 3, pp. 555–584, 2013.

[37] D. Kempe, J. Kleinberg, and É. 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*. ACM, 2003, pp. 137–146.

[38] M. Kimura, K. Saito, and R. Nakano, "Extracting influential nodes for information diffusion on a social network," in *AAAI*, vol. 7, 2007, pp. 1371–1376.

[39] M. Li, X. Wang, K. Gao, and S. Zhang, "A survey on information diffusion in online social networks: Models and methods," *Information (Switzer-*

land), vol. 8, 2017.

[40] 吉川友也，斉藤和巳，元田浩，大原剛三，木村昌弘，“情報拡散モデルに基づくソーシャルネットワーク上でのノードの期待影響度曲線推定法”，電子情報通信学会論文誌 D，vol. 94，no. 11，pp. 1899–1908，2011.

[41] T. Murata, and H. Koga, “Extended methods for influence maximization in dynamic networks,” *Computational Social Networks*, vol. 5, 2018.

[42] S. Osawa, and T. Murata, “Selecting seed nodes for influence maximization in dynamic networks,” *Studies in Computational Intelligence*, vol. 597, pp. 91–98, 2015.

[43] L. Yang, Y. Qiao, Z. Liu, J. Ma, and X. Li, “Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm,” *Soft Computing*, vol. 22, no. 2, pp. 453–464, 2018.

[44] L. Jain, R. Katarya, and S. Sachdeva, “Role of opinion leader for the diffusion of products using epidemic model in online social network,” in *2019 Twelfth International Conference on Contemporary Computing (IC3)*. IEEE, 2019, pp. 1–6.

[45] A. U. Rehman, A. Jiang, A. Rehman, A. Paul, M. T. Sadiq *et al.*, “Identification and role of opinion leaders in information diffusion for online discussion network,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2020.

[46] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, “Ranking the spreading ability of nodes in complex networks based on local structure,” *Physica A: Statistical Mechanics and its Applications*, vol. 403, pp. 130–147, 2014.

[47] V. Arnaboldi, M. Conti, A. Passarella, and R. I. Dunbar, “Online social networks and information diffusion: The role of ego networks,” *Online Social Networks and Media*, vol. 1, pp. 44–55, 2017.

[48] H. Huang, H. Shen, Z. Meng, H. Chang, and H. He, “Community-based influence maximization for viral marketing,” *Applied Intelligence*, vol. 49, no. 6, pp. 2137–2150, 2019.

[49] F. Ullah, and S. Lee, “Identification of influential nodes based on temporal-aware modeling of multi-hop neighbor interactions for influence spread maximization,” *Physica A: Statistical Mechanics and its Applications*, vol. 486, pp. 968–985, 2017.

[50] M. Li, X. Wang, K. Gao, and S. Zhang, “A survey on information diffusion in online social networks: Models and methods,” *Information*, vol. 8, no. 4, p. 118, 2017.

[51] A. Sheikhahmadi, M. A. Nematbakhsh, and A. Zareie, “Identification of influential users by neighbors in online social networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 486, pp. 517–534, 2017.

[52] X. Chen, F. Zhou, K. Zhang, G. Trajcevski, T. Zhong, and F. Zhang, "Information diffusion prediction via recurrent cascades convolution," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019, pp. 770–781.

[53] J. Yang, and S. Counts, "Predicting the speed, scale, and range of information diffusion in twitter," in *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

[54] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, "Quantifying echo chamber effects in information spreading over political communication networks," *EPJ Data Science*, vol. 8, no. 1, pp. 1–13, 2019.

[55] S. Stieglitz, and L. Dang-Xuan, "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior," *Journal of management information systems*, vol. 29, no. 4, pp. 217–248, 2013.

[56] 川本貴史，豊田正史，吉永直樹，"マイクロブログにおける社会的影響力を持つ情報カスケードの早期検知に向けて"，第 8 回 Web とデータベースに関するフォーラム論文集，vol. 2015，pp. 48–55，2015.

[57] D. Varshney, S. Kumar, and V. Gupta, "Predicting information diffusion probabilities in social networks: A bayesian networks based approach," *Knowledge-Based Systems*, vol. 133, pp. 66–76, 2017.

[58] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?" in *Proceedings of the 23rd international conference on World wide web*. ACM, 2014, pp. 925–936.

[59] T. B. N. Hoang, and J. Mothe, "Predicting information diffusion on twitter–analysis of predictive features," *Journal of computational science*, vol. 28, pp. 257–264, 2018.

[60] A. Inokuchi, T. Washio, and H. Motoda, "An apriori-based algorithm for mining frequent substructures from graph data," in *Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery*. London, UK, UK: Springer-Verlag, 2000, pp. 13–23.

[61] S. Salihoglu, "Networks as vectors of their motif frequencies and 2-norm distance as a measure of similarity," Stanford University, Tech. Rep., 2006.

[62] N. Sidere, P. Heroux, and J.-Y. Ramel, "A vectorial representation for the indexation of structural informations," in *Structural, Syntactic, and Statistical Pattern Recognition*, N. da Vitoria Lobo, T. Kasparis, F. Roli, J. T. Kwok, M. Georgiopoulos, G. C. Anagnostopoulos, and M. Loog, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 45–54.

[63] B. Luo, R. C. Wilson, and E. R. Hancock, "Spectral embedding of graphs," *Pattern Recognition*, vol. 36, no. 10, pp. 2213–2230, 2003.

[64] K. Riesen, and H. Bunke, *Graph Classification and Clustering Based on Vector Space Embedding*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2010.

[65] 伏見卓恭，佐藤哲司，“レシピ投稿サイトにおけるユーザ間コミュニケーションのネットワーク分析”，in 電子情報通信学会技術研究報告，vol. 115，no. 230，September 2015，pp. 71–76.

[66] T. Mäki-Marttunen, “An algorithm for motif-based network design,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 5, pp. 1181–1186, 2017.

[67] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graph evolution: Densification and shrinking diameters,” *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, March 2007.

[68] R. Albert, H. Jeong, and A. L. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, pp. 378–382, 2000.

[69] T. Fushimi, T. Satoh, K. Saito, and K. Kazama, “Comparison of influence measures on structural changes focused on node functions,” in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*. New York, NY, USA: ACM, 2015, pp. 16:1–16:10.

[70] J. Kim, J. Bae, and M. Hastak, “Emergency information diffusion on online social media during storm cindy in u.s.” *International Journal of Information Management*, vol. 40, pp. 153–165, 2018.

[71] A. Lancichinetti, S. Fortunato, and F. Radicchi, “Benchmark graphs for testing community detection algorithms,” *Physical review E*, vol. 78, no. 4, p. 046110, 2008.

[72] A. Hayashi, M. Kohjima, T. Matsubayashi, and H. Sawada, “Regularity measure and influence weight for analysis and visualization of consumer’s attitude,” in *2015 19th International Conference on Information Visualisation*, July 2015, pp. 290–299.

[73] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda, “Topic tracking model for analyzing consumer purchase behavior,” in *Proceedings of the 21st International Jont Conference on Artifical Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009, pp. 1427–1432.

[74] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in *Proceedings of the 10th International Conference on World Wide Web*. New York, NY, USA: ACM, 2001, pp. 285–295.

[75] F. Isinkaye, Y. Folajimi, and B. Ojokoh, “Recommendation systems: Principles, methods and evaluation,” *Egyptian Informatics Journal*,

vol. 16, no. 3, pp. 261–273, 2015.

[76] P. Symeonidis, E. Tiakas, and Y. Manolopoulos, “Product recommendation and rating prediction based on multi-modal social networks,” in *Proceedings of the Fifth ACM Conference on Recommender Systems*. New York, NY, USA: ACM, 2011, pp. 61–68.

[77] K. Inafuku, T. Fushimi, and T. Satoh, “Extraction method of typical purchase patterns based on motif analysis of directed graphs,” in *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*. ACM, 2016, pp. 86–95.

[78] K. Nowicki, and T. A. B. Snijders, “Estimation and prediction for stochastic blockstructures,” *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 1077–1087, 2001.

全研究業績のリスト

学術雑誌論文 (査読あり)

- Kazufumi Inafuku, Takayasu Fushimi and Tetsuji Satoh. Predicting stimulation index of information transmissions by local structural features in social networks, Social Network Analysis and Mining, Vol.12, Article 40, 2022, pp.1-14, https://doi.org/10.1007/s13278-022-00865-0.

- 稲福和史, 伏見卓恭, 佐藤哲司. トライアド推移に基づく購買行動の成長分析. 情報処理学会論文誌, vol.60, no.4, pp.1141-1150.【推薦論文】

国際会議論文 (査読あり)

- Kazufumi Inafuku, Takayasu Fushimi, Tetsuji Satoh. Stimulation Index of Cascading Transmission in Information Diffusion over Social Networks, Complex Networks & Their Applications IX. COMPLEX NETWORKS 2020 2020. Studies in Computational Intelligence, pp.469-481, ONLINE(Poster), Dec. 2020.【論文誌推薦】

- Kazufumi Inafuku, Takayasu Fushimi, Tetsuji Satoh. Structural Transition Analysis of Dynamic Network based on Roles of Adding Edges. Proceedings of the 21th International Conference on Information Integration and Web-based Applications & Services(iiWAS2019), pp.372-378, Munich, Germany, Dec. 2019.

- Kazufumi Inafuku, Takayasu Fushimi, Tetsuji Satoh. Growth Analysis of Purchase History Graph Based on Relative Value of Edge Multiplicity. Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services(iiWAS2018), pp.65-69, Yogyakarta, Indonesia, Nov. 2018.

- Kazufumi Inafuku, Takayasu Fushimi, Tetsuji Satoh. Growth Analysis of Purchase History Graph Based on Relative Value of Edge Multiplicity. Proceedings of 19th International Conference on Information Integration and Web-based Applications & Services(iiWAS2017), pp.358-365, Salzburg, Austria, Dec. 2017.

- Kazufumi Inafuku, Takayasu Fushimi, Tetsuji Satoh. Extraction Method of Typical Purchase Patterns Based on Motif Analysis of Directed Graphs. Proceedings of 18th International Conference on Information Integration and Web-based Applications & Services(iiWAS2016), pp.88-97, Singapore, Singapore, Nov. 2016.

- Mutsuki Yamazaki, Kazufumi Inafuku, Tetsuji Satoh. Tag Recommendation Method for Enhancing Web Novel Retrieval, 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), pp.43-48, ONLINE, Sep. 2020.

国内会議論文

- 稲福和史, 伏見卓恭, 佐藤哲司. ソーシャルネットワーク上での情報拡散におけるエッジ誘発度指標の提案. 知識ベースシステム研究会, 人工知能学会, C001-07, pp.33-38. オンライン, Jul. 2020.

- 稲福和史, 伏見卓恭, 佐藤哲司. ソーシャルネットワークにおける成長誘発エッジの早期検出手法. 電子情報通信学会, 2020 年電子情報通信学会総合大会, D-1-7, 広島県東広島市, Mar. 2020.

- 稲福和史, 伏見卓恭, 佐藤哲司. 複雑ネットワークにおける出現位置と役割に着目した効率的な成長誘発エッジ検出手法. 電子情報通信学会 データ工学研究専門委員会 他共催, 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2020), C3-3, 福島県郡山市, Mar. 2020.

- 稲福和史, 伏見卓恭, 佐藤哲司. 最終球への配球推移に基づくキャッチャー成績分析. 電子情報通信学会 データ工学研究専門委員会 他共催, 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019), I1-2, 長崎県佐世保市, Mar. 2019.

- 稲福和史, 伏見卓恭, 佐藤哲司. 近傍エッジとの関係に着目したグラフマイニング手法の提案と評価. 電子情報通信学会 データ工学研究専門委員会 他共催, 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), J2-5, 福井県あわら市, Mar. 2018.

- 稲福和史, 佐藤哲司. 最終球への配球推移に基づくキャッチャー成績分析, WebDB Forum 2017, COC-17(ポスター), 御茶ノ水女子大学, Sep. 2017.

- 稲福和史, 伏見卓恭, 佐藤哲司. EC サイトにおける購買行動の成長分析, 情報処理学会, マルチメディア，分散，協調とモバイル (DICOMO2017) シンポジウム論文集, pp. 1107-1113, 北海道札幌市, Jun. 2017.【最優秀論文賞，優秀プレゼンテーション賞，論文誌推薦】

- 稲福和史, 伏見卓恭, 佐藤哲司. レビュー順序グラフに基づく購買行動パターンの分析. 電子情報通信学会 データ工学研究専門委員会 他共催, 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017), A3-4, 岐阜県高山市, Mar. 2017.【学生プレゼンテーション賞】

- 稲福和史, 伏見卓恭, 佐藤哲司. レビュー順序グラフを用いた購買行動パターンの抽出手法の提案. 情報処理学会 データベースシステム研究会 他共催, WebDB Forum 2016, B10-3, 慶応大学日吉キャンパス, Sep. 2016.