

(続紙 1)

京都大学	博士 (情報学)	氏名	井上 雅章
論文題目	Modeling and Statistical Inference of Preferential Attachment in Complex Networks: Underlying Formation of Local Community Structures (複雑ネットワークにおける優先的選択のモデリングと統計的推測: 局所的コミュニティ構造の形成)		
(論文内容の要旨)			
<p>実世界のグラフ構造を持つデータを大規模に計算機で解析し、その相互作用のダイナミクスや構造的な機能を明らかにしようとする研究が、複雑ネットワーク分野において行われている。古典的な成長機構である優先的選択は、ノードの次数が大きいほど新たなエッジを獲得しやすいとするメカニズムであり、ネットワークのスケールフリー性を説明する。しかしながら、ノードの次数という一次構造しか用いない優先的選択の成長過程では、クラスターの形成、三角構造の多さなど実際のネットワークデータに共通して見られる局所的なコミュニティ構造が形成されることの説明ができない。複雑ネットワークにおけるコミュニティ構造の形成過程を明らかにするためには、従来の優先的選択を超えた成長モデルと統計的推定手法の開発が重要である。</p> <p>本学位論文は、複雑ネットワークの代表的な成長機構である優先的選択をベースとして、局所的なコミュニティ構造の形成過程を反映させた成長モデルとその成長関数の推定法を提案し、シミュレーションによるモデルの妥当性の検証結果と実データの解析結果を述べるものである。</p> <p>本学位論文の第1章は序章となっており、研究目的と扱う問題、そして本学位論文の主要な貢献が述べられている。</p> <p>第2章では、第3章から第6章までで提案するグラフ成長モデル及びハイパーグラフ成長モデルの準備として、ネットワーク成長機構の背景知識を概説している。はじめに、複雑ネットワーク分野で研究対象となっている代表的なグラフ成長機構として優先的選択や推移性についてモデルとその背景を述べている。次に、従来はグラフで考えられてきた複雑ネットワークの成長モデルをハイパーグラフへ拡張する準備として、ハイパーグラフ成長の問題設定について導入を行なっている。</p> <p>以降は、本学位論文の方法論や解析結果とそれらに関する議論からなる主要な貢献部分であり第1部から第3部までで構成されている。</p> <p>第1部は第3章と第4章からなる。第3章ではグラフの成長機構である優先的選択と推移性を同時に考慮した成長モデルと、それらの成長機構をそれぞれ決定づける関数のノンパラメトリック同時推定法を提案している。そして第4章では実データの解析結果を示している。ここで優先的選択は、次数が多いほどそのノードは新たなエッジを獲得しやすい、という各ノード周りの一次構造を反映した成長機構である。一方で推移性は、2つのノードが共通の隣接ノードを多くもつほどそのノード間に新たなエッジが獲得されやすい、というノード周りの二次構造を反映した成長機構である。提案する優先的選択と推移性のノンパラメトリック同時推定法は、従来の推定法で必要とした関数形の仮定あるいは2つのメカニズムを個別に推定するという制限を必要としない。そのため、提案法は両方の成長機構が同時に存在する場合に従来法で生じていた推定値の偏りを軽減することができ、また、優先的選択だけでは再現しきれなかった実際のネットワークに含まれる密な三角構造などのコミュニティ構造を説明できる</p>			

ようになった。第3章ではさらに、提案法が用いている最尤推定法の効率化のために反復最適化法の一つであるMMアルゴリズムの更新式を導出し、提案法の推定精度についてのシミュレーション結果を示した。そして現実のネットワーク成長過程において優先的選択と推移性のどちらがより支配的であるかについて、その影響力を定量的に評価する方法についても提案している。第4章では、実データ解析として、2分野の論文共著者ネットワークを用いて成長シミュレーションを行い、提案モデルは次数や共通隣接ノード数の分布の再現性に優れており、また、実験で用いたデータの成長過程では優先的選択よりも推移性の方が優位であることを示している。

第2部は第5章と第6章からなり、第5章ではハイパーグラフの成長における優先的選択のモデルを考え、その成長を決定づける関数の最尤推定法を提案している。この推定法は第1部と同様にノンパラメトリックで関数形の仮定が不要である。また、第6章では現実にはハイパーグラフ構造をもつ13種のデータの解析を行なった結果が述べられている。第5章で提案されるハイパーグラフ成長モデルでは、従来はグラフで考えられてきた優先的選択をハイパーグラフとして考えることで、グラフのエッジでは表現できない3以上の個体を含む相互作用がハイパーエッジによって表現可能となる。実際のネットワークに大きなクリーク、あるいは、多くのクリークが存在することについて、グラフ上の優先的選択モデルよりもハイパーグラフ上の優先的選択モデルの方がこれらの構造の再現性が良いことをシミュレーションで示している。第5章ではさらに、Newton's Identitiesにもとづく再帰関係式の導出によって尤度関数の計算量を大幅に削減する方法、及び、推定精度の向上のためハイパーグラフとともに登場する新たなノードの選択バイアスを軽減する条件付き確率の導出を行なっている。第6章では、13種のハイパーグラフ構造を持つ実データを用いたシミュレーションにより、優先的選択によるグラフ成長と優先的選択によるハイパーグラフ成長のモデルを比較し、提案法はグラフの一次構造、二次構造の双方の観点で再現性が良いことを示している。

第7章を含む第3部は、学位論文全体の結論となっている。これまでに得られた研究成果の概要のまとめが述べられるとともに、今後の研究に関する展望について議論されている。

(論文審査の結果の要旨)

グラフ構造を持つ実世界の個体間の相互作用を大規模に解析する研究が複雑ネットワーク分野において行われており、グラフの成長プロセスに関する動的な性質が注目されている。グラフのノードが個体、リンクが個体間の関係を表すとき、次数の大きいノードがより多くのリンクを獲得するという優先的選択 (Preferential Attachment, PA) は成長機構の中でも従来から研究されている。PAは実データがもつ次数分布のスケールフリー性を説明するが、各個体の次数にのみ依存した機構であるため、複数の個体が協力行動によりコミュニティを形成することの説明ができない。複雑ネットワークのコミュニティ形成のプロセスを明らかにするためには、従来のPAを超える成長機構のモデリングと、それをデータから統計的に推測する手法が重要である。

本論文は、コミュニティ構造を示す複雑ネットワークを対象とした成長機構のモデリングと統計的推測について論じている。成長過程における局所的なコミュニティ形成をモデリングし、そのメカニズムを決定づける成長関数を統計的に推定する。グラフの三角構造の形成、クリークの形成等に注目して、計算機シミュレーションによる性能評価や実データ解析によるモデルの妥当性の検証を実施している。具体的には、以下に示す研究成果を得ている。

1. 2つのノードが共通して接続しているノード数が大きいほど、その2つのノード間に新たなリンクが獲得されやすいという性質を推移性 (transitivity) という。本論文では、グラフ成長機構としてPAと推移性を同時に考え、それらの成長関数を関数形を仮定することなくノンパラメトリックに同時推定する手法を提案した。従来手法は、一方のメカニズムを単独で推定するか、あるいは何らかの関数形を仮定して両方を同時推定するかのどちらかであった。推定量をより効率的に求めるために、反復最適化法的一种であるMMアルゴリズムの更新式を導出した。PAと推移性が同時に存在している場合に、提案手法が関数形を仮定することなく両方の成長関数を高い精度で推定できることを確認した。また、ある分野の科学論文共著ネットワークを解析して2つの成長機構の貢献度を調べたところ、推移性がPAを上回っていることが示された。

2. 通常のグラフは2個のノード間の関係性しか扱えない。そこで本論文ではハイパーグラフを用いることで、3個以上のノード間の関係性をハイパーエッジで表現し、その成長機構としてハイパーグラフのPAを考えた。そして、PAの成長関数を関数形を仮定することなく最尤推定する手法を提案した。PAの計算コストを大幅に削減する再帰式の導出によって実用的な計算時間とすることに成功し、さらにハイパーグラフ成長における誕生ノードの選択バイアスを軽減して推定精度を向上させた。13種の実データを解析したところ、ハイパーグラフのPAを用いた成長機構が従来のグラフのPAよりもグラフの1次構造と2次構造の両方の観点から当てはまりが良いことがわかった。

このように、本論文ではグラフの成長機構として従来のPAを超えるモデルとその統計的推定法を提案し、その有効性を示しているものであり、博士 (情報学) の学位論文として価値あるものとして認める。また、令和4年2月21日に論文内容とそれに関連した事項について口頭試問を行った結果、合格と認めた。本論文のインターネットでの全文公表についても支障がないことを確認した。