

学 位 論 文 題 名

Extracting Many-to-many Related Communities in
Bipartite Graph

(二部グラフにおける多対多に関係するコミュニティの抽出)

学位論文内容の要旨

現実世界には、二部グラフで表すことの出来る関係が多数存在する。これらの中に、コミュニティと呼ばれる、その外部と比較して、内部では隣接関係が大きく類似しているノード集合が存在している。コミュニティを発見することが、システムの関係構造を理解するのに貢献している。従来のコミュニティ抽出方法では、非常に高速にコミュニティ構造が得られる一方、二部グラフの一部グラフ変換やコミュニティ外部の無視など、精度に問題点がある。そこで、本論文では、一部グラフ変換せず、多対多のコミュニティ対応関係を考慮することで、二部グラフにおけるより精度の高いコミュニティ分割を得る方法を提案した。まず、二部グラフのコミュニティ分割問題を、ノードにおける巡回セールスマン問題に置き換え、ノード間コスト関数と局所クラスタリング組織化法による巡回経路コスト最適化を用いて、コミュニティ間コストによるコミュニティ分割方法を提案した。さらに、二部グラフにおける複数の生成モデルに対して提案方法の有効性を検証した。次に、二部グラフにおける多対多のコミュニティ対応関係を考慮したモジュラリティを提案した。提案指標を様々な二部グラフとコミュニティ分割に対して適用し、有効性を確認した。与えられた二部グラフに対して、提案指標が大きな値を得るコミュニティ構造を実質的な計算時間で発見する方法を提案した。数千ノード、一万リンク程度の実二部グラフをコミュニティ分割して、有効性を示した。応用として、有向グラフを出部と入部の二部グラフとみなしてコミュニティ分割する方法を示した。また得られたコミュニティ構造から、コミュニティの機能を同定する方法を示した。現実の有向グラフとして *C.elegans* の神経細胞網を解析し、そのコミュニティ機能を明らかにした。

第1章では、研究背景として、複雑ネットワーク科学の始まり、複雑ネットワークとしての二部グラフ、コミュニティ分割における関連研究、さらに研究目的について述べた。二部グラフにおける多対多に関係しあうコミュニティ構造を抽出する方法を確立し、現実の二部グラフのコミュニティ間関係を明らかにすることを研究目的と述べた。

第2章では、隣接ノード関係を用いたノード間コスト関数と巡回経路コスト最適化法に基づくコミュニティ分割方法を提案した。ノード間コスト関数にはコサイン類似度、巡回経路コスト最適化法に局所クラスタリング組織化法を採用した、二部穴居人モデル及びにより二部 CNN モデルにより生成した二部グラフに対して提案方法を適用した結果、ノード間コストの頻度分布が多峰的となり、コミュニティ内コストとコミュニティ境界が生じ、コストのしきい値を設定することでコミュニティ分割できることを明らかにした。

第3章では、一部グラフで用いられるコミュニティ内リンク密度に基づくコミュニティ構造評価指標であるモジュラリティを二部グラフ用に改良した。モジュラリティはコミュニティ間のリンク密度を考慮していないため、多対多の関係を考慮できない。そのため、コミュニティ間対応度を導入し、全コミュニティ間リンク密度を評価に組み込んだ指標を提案した。多対多のコミュニティ関係を持つ最もシンプルなモデルである W 型二部グラフに対して様々なコミュニティ構造における提案

指標の値を計測した結果, 多対多のコミュニティ関係を表している構造に対しては最大値をとることを明らかにし, 提案指標の有効性を示した.

第4章では, 提案した指標を最大化させるコミュニティ構造を高速に得る方法を凝集的貪欲法に基づき提案した. まず, 多対多のコミュニティ対応関係を持つ小さな二部グラフにおける, コミュニティ数と提案指標の値の関係を解析し, 各コミュニティ数における最大値が全体の最大値に向かって山状になる一方, コミュニティ数に関わらず広い範囲で指標値が分散することを明らかにした. その結果に支持されて, 凝集的貪欲法を提案し, その計算量を結合ペアの限定と差分計算により削減する方法を提案した.

第5章では, 有向グラフを出部集合と出部集合で構成される二部グラフに変換することで, 類似した受信コミュニティにリンクが存在する送信コミュニティ, またその逆を得ることが可能となる方法を提案した. また, 受信コミュニティと送信コミュニティとの対応関係からコミュニティの機能を同定する方法も提案した. *C.elegans* の神経細胞網に対して提案方法を適用した結果, 収集, 拡散, 蓄積, 伝達の機能を抽出し, かつ蓄積機能を持つコミュニティは同時に収集や拡散の機能を持ち合わせている場合が多いことを明らかにした.

第6章では, ここまでの結論を取りまとめた.

学位論文審査の要旨

主査	教授	古川正志
副査	教授	栗原正仁
副査	教授	鈴木恵二
副査	教授	小野哲雄
副査	准教授	山本雅人

学位論文題名

Extracting Many-to-many Related Communities in Bipartite Graph

(二部グラフにおける多対多に関係するコミュニティの抽出)

本学位論文では、一部グラフ変換せず、多対多のコミュニティ対応関係を考慮することで、二部グラフにおけるより精度の高いコミュニティ分割を得る方法を提案し、その有効性を検証することを目的としている。

異なる二つの集合間の対応をグラフで表現したものは二部グラフと呼ばれる。グラフで表現したそれぞれの集合には、コミュニティと呼ばれる、別の集合の部分集合との対応関係が類似した部分集合が存在している。このようなコミュニティの発見は、システムの内在する関係構造を理解するのに貢献している。従来の二部グラフにおけるコミュニティ抽出方法は、非常に高速にコミュニティ構造を得ることができたが、一方で二部グラフの一部グラフ変換やコミュニティ外部の無視など、精度に問題があることが指摘されていた。そこで本論文では、まず、二部グラフのコミュニティ分割問題を、ノードにおける巡回セールスマン問題に置き換え、ノード間コスト関数と局所クラスタリング組織化法による巡回経路コスト最適化を用いて、コミュニティ間コストによるコミュニティ分割方法を提案している。さらに、二部グラフにおける複数の生成モデルに対して提案方法の有効性を検証している。次に、二部グラフにおける多対多のコミュニティ対応関係を考慮したモジュラリティの指標を新たに提案している。また、提案指標を様々な二部グラフとコミュニティ分割に対して適用することで、その有効性を確認している。さらに、与えられた二部グラフに対して提案指標が極大となるコミュニティ構造を実質的な計算時間で発見する方法を提案している。これを用いて数千ノード、一万リンク程度の現実二部グラフをコミュニティ分割して、その有効性を明らかにしている。応用としては、有向グラフを出力と入力の二部グラフとみなしてコミュニティ分割する方法を提案している。またこの方法から得られたコミュニティ構造から、コミュニティの機能を同定する方法を示している。具体的な適用として現実に有向グラフとして知られる *C.elegans* の神経細胞網を解析し、そのコミュニティ機能を明確にしている。

第1章では、研究背景として、複雑ネットワーク科学の始まり、複雑ネットワークとしての二部グラフ、コミュニティ分割における関連研究、さらに研究目的について述べた。研究目的として二部グラフにおける多対多に関係しあうコミュニティ構造を抽出する方法を確立し、現実の二部グラフのコミュニティ間関係を明らかにすることを述べている。

第2章では、隣接ノード関係を用いたノード間コスト関数と巡回経路コスト最適化法に基づくコ

コミュニティ分割方法を提案している。ノード間コスト関数にはコサイン類似度、巡回経路コスト最適化法に局所クラスタリンク組織化法を採用した、二部穴居人モデル及びにより二部 CNN モデルにより生成した二部グラフに対して提案方法を適用した結果、ノード間コストの頻度分布が多峰的となり、コミュニティ内コストとコミュニティ境界が生じ、コストのしきい値を設定することでコミュニティ分割可能なことを明らかにしている。

第3章では、一部グラフで用いられるコミュニティ内リンク密度に基づくコミュニティ構造評価指標であるモジュラリティを二部グラフ用に改良した。モジュラリティはコミュニティ間のリンク密度を考慮していないため、多対多の関係を考慮できない。そのため、コミュニティ間対応度を導入し、全コミュニティ間リンク密度を評価に組み込んだ指標を新たに提案している。多対多のコミュニティ関係を持つ最もシンプルなモデルである W 型二部グラフに対して様々なコミュニティ構造における提案指標の値を計測した結果、多対多のコミュニティ関係を表している構造に対して提案指標は極大値をとることを明らかにし、提案指標の有効性を示している。

第4章では、提案した指標を最大化させるコミュニティ構造を高速に得る方法を凝集的貪欲法に基づいて提案している。まず、多対多のコミュニティ対応関係を持つ小さな二部グラフにおける、コミュニティ数と提案指標の値の関係を解析して、各コミュニティ数における最大値が全体の最大値に向かって山状に分布するが、一方でコミュニティ数に関わらず広い範囲で指標値が分散することを明らかにしている。その結果を実現するような凝集的貪欲法を提案し、その計算量を結合ペアの限定と差分計算により削減する方法を提案している。

第5章では、有向グラフを出部集合と入部集合で構成される二部グラフに変換することで、類似した受信コミュニティにリンクが存在する送信コミュニティ、またその逆を得ることが可能となる方法を提案している。また、受信コミュニティと送信コミュニティとの対応関係からコミュニティの機能を同定する方法も提案している。これらの方法を *C.elegans* の神経細胞網に対して適用した結果、収集、拡散、蓄積、伝達の機能が抽出でき、かつ蓄積機能を持つコミュニティは同時に収集や拡散の機能を持ち合わせている場合が多いことを明確にしている。

第6章では、本論文のまとめと今後の展望が述べられている。

これを要するに、本論文では、従来は一部グラフに変換しなければ行えなかったコミュニティ分割に対し、数理最適化の方法と二部グラフのための新規なモジュラリティ指標が提案され、それらの有効性が検証されている。このことは大規模な二部グラフの高速な解析とその応用を可能とするため、複雑ネットワークの理論および複雑系工学に寄与するところ大であり、博士(情報科学)の学位を授与するに値するものと認める。