

令和 5 年 6 月 13 日現在

機関番号：12608

研究種目：研究活動スタート支援

研究期間：2021～2022

課題番号：21K21282

研究課題名(和文) 合意モデルに基づく分散コミュニティ検出アルゴリズムの開発と理論的解析

研究課題名(英文) Community Detection Algorithm from Voting Process

研究代表者

清水 伸高 (Shimizu, Nobutaka)

東京工業大学・工学院・助教

研究者番号：10910127

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：グラフ上の合意モデルと呼ばれる確率過程をコミュニティ検出に応用できないかを理論的に研究した。具体的には、コミュニティ検出のベンチマークとしてしばしば用いられる確率的ブロックモデルと呼ばれるランダムグラフ上でのk-Majorityと呼ばれるプロトコルの振る舞いを調べた。合意モデルの文脈では完全グラフという非常に単純なグラフ上での振る舞いの解析がほとんどであるが、それよりも複雑なグラフでの解析が部分的に進展した。本研究では、特定の状況下でのk-Majorityの性能を示した。また、コミュニティ検出と深い関わりがある埋め込みクリーク問題と呼ばれる問題に対して計算量下界の改善を得た。

研究成果の学術的意義や社会的意義

本研究はグラフ上の確率過程の解析とランダムグラフの計算複雑性という二つの成果を得た。前者の成果は、これまで完全グラフ上での振る舞いしかよく分かっていなかった合意モデルの振る舞いを、より複雑な確率的ブロックモデル上で解析することを可能にした。後者の成果はコミュニティ検出に関する埋め込みクリーク問題に対して、困難性の増幅と呼ばれる結果を示すことに成功した。この結果を得るにあたって新たに提案した枠組みはより幅広いクラスの問題に対して適用することが可能であり、理論計算機科学において非常に重要な意義を持つ。

研究成果の概要(英文)：In this project, I studied the possibility of applying a stochastic process called the voting process on a graph to community detection from a theoretical point of view. Specifically, I investigated the behavior of a protocol called k-Majority on a random graph known as the stochastic block model, which is often used as a benchmark for community detection. The performance of the protocol varies depending on the value of parameter k, and I demonstrated that in certain situations, a larger value of k can improve the performance of community detection. Furthermore, I obtained improvements in computational lower bounds for the embedding clique problem, which is closely related to community detection.

研究分野：理論計算機科学

キーワード：確率過程 ランダムグラフ 平均時計算量

1. 研究開始当初の背景

ランダムウォークなどをはじめとしたグラフ上の確率過程やそれに基づくアルゴリズムは、分散計算やネットワーク解析などにしばしば応用される実用的な研究対象である。特にその理論解析はアルゴリズムの性能保証に直結する重要な課題となる。本研究の目的は、合意モデルという確率過程を用いてコミュニティ検出を行うアルゴリズムを与えられないかを明らかにすることである。

合意モデルとは各頂点が意見と呼ばれる離散的な値(例えば 0 と 1)を保持したグラフを考え、各頂点は何かしらのプロトコルに基づいて隣接頂点の意見に応じて自身の意見を確率的に更新していき、最終的に全頂点が同じ意見を持つ(合意)状態に至ることを目標とする。考えるプロトコルに応じて合意に至る早さ(合意時間)が決まる。合意モデルに関する既存研究のほとんどは、完全グラフといった、単一の密なコミュニティからなるグラフ上での合意モデルの振る舞いを解析し、その合意時間が高速であることを証明していた。しかしながらより複雑な形状を持ったグラフ上での合意モデルの振る舞いはほとんど分かっていなかった。

複数のコミュニティを持つグラフ上での合意モデルを考えると、直感的には各コミュニティ内で頂点がそれぞれ同じ意見を持つ。したがってその意見がコミュニティ間で異なっていれば、そのグラフのコミュニティ構造を抽出できることが期待される。

2. 研究の目的

本研究は合意モデルに基づいたコミュニティ検出アルゴリズムの設計を目的としている。コミュニティ検出はネットワーク解析の文脈でモジュラリティ最大化など様々なタイプのアルゴリズムが知られており、その中で合意モデルは非常に単純なプロトコルに基づくものなので、分散コンピュータ上で簡単に実装でき、しかも高速であるという利点がある。しかしその単純さゆえに得られるクラスタリングは良いものとは限らない。そのため、コミュニティ検出の観点で合意モデルの性能やその限界を理論的な観点から明らかにすることが本研究の内容である。

合意モデルは分散計算の合意問題に対する有効なアプローチの一つとして知られており、その文脈で研究されているが、本研究ではそれとは全く異なるコミュニティ検出への応用を見据えている。これを通じて合意モデルという確率過程の新たな応用を開拓していくことによって、コミュニティ検出の分野のみならず合意モデルを研究する確率過程の分野にも刺激を与えるという意味で本研究の意義は重要である。

3. 研究の方法

コミュニティ検出に関する理論的な既存研究のほとんどはアルゴリズムへの入力として特定のランダムグラフモデルを考えており、そのランダムグラフ上でのアルゴリズムの振る舞いを解析している。本研究では確率的ブロックモデルというランダムグラフ上での k -Majority(または Best-of- k と呼ばれる)の振る舞いを解析した。このプロトコルでは、各頂点が 0 または 1 の二種類の意見を持ち、一様ランダムに選んだ k 個の隣接点の意見を見てその多数決の意見に自身の意見を更新するというものである。

4. 研究成果

本研究の成果は、合意モデルの解析とコミュニティ検出の計算複雑性の解析の二つに大別される。

確率的ブロックモデル上の合意モデルの研究は 2-Majority と 3-Majority のみが知られており、コミュニティ内外の辺密度の比に応じて相転移現象が発生し、具体的にはこの比がある閾値より上か下かによってコミュニティ検出できるかどうかに分かれることが知られている。本研究ではこれをより一般の k -Majority で解析した。特に k が大きくなると漸的にこの閾値が 1 に近づくことを示した。本成果は LA シンポジウム 2021 にて発表した。

コミュニティ検出の理論的解析は確率的ブロックモデルなどのランダムグラフ上でのアルゴリズムの振る舞いを見ることによってなされる。これはランダムな入力上での計算複雑性を解析していることになり、理論計算機科学では平均時計算量(average-case complexity)の分野で研究されている。

特に平均時計算量では統計的なアルゴリズムと計算量的なアルゴリズムの性能ギャップ (statistical-computational gap) が研究されている。例えば確率的ブロックモデル上のコミュニティ検出はパラメータが適切な条件を満たせば全探索に基づいた最尤推定によって厳密に解けることが知られている。これを統計的なアルゴリズムという。しかしながら全探索は指数時間アルゴリズムであり、現実的ではない。これを効率的に解けるかどうか重要な問題であり、そのようなアルゴリズムをここでは計算量的なアルゴリズムと呼ばれている。

コミュニティ検出のほかにも埋め込みクリーク問題や主成分分析 (PCA) などでも同様に統計的なアルゴリズムと計算量的なアルゴリズムにギャップがあるだろうと予想されており、特に埋め込みクリーク問題が効率的に解ければある種のコミュニティ検出の問題も効率的に解けることが知られている。すなわち、埋め込みクリーク問題の平均時計算量を理解することはコミュニティ検出を含む様々なランダムグラフ上の問題の平均時計算量を理解することに繋がるため、非常に重要である。

本研究では埋め込みクリーク問題に対して**困難性の増幅**と呼ばれる結果を証明した。平均時計算量では問題の困難性を効率的に解ける入力個数によって評価し、そのような入力が少ないほど困難な問題であるとされる。困難性の増幅とは、「1%の割合の入力上で問題 P1 を正しく解くアルゴリズムが存在するならば、99%の割合の入力上で問題 P2 を正しく解くアルゴリズムが存在する」ことを主張する結果である(ここで考える問題 P1 と P2 は必ずしも同じ問題であるとは限らない)。対偶をとると、99%の入力上で P2 を解くことが困難ならば、1%の入力上で P1 を解くことすら困難であるという主張になり、これは P2 の弱い困難性に依拠して P1 の強い困難性が示せるという結果である。本研究では問題 P1 と P2 をどちらも埋め込みクリーク問題としたときにその困難性の増幅を初めて証明した。

本成果では非常に汎用的な枠組みを新たに提案し、それに基づいて困難性の増幅を証明している。したがってこの枠組みを用いてほかの問題に対しても困難性の増幅が示せることが期待される。本成果は理論計算機科学のトップ会議 Symposium on Theory of Computing (STOC23) に採択された。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Nobutaka Shimizu and Takeharu Shiraga	4. 巻 -
2. 論文標題 Reversible Random Walks on Dynamic Graphs	5. 発行年 2023年
3. 雑誌名 Random Structures and Algorithms	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 笛木 正雄
2. 発表標題 合意モデルのコミュニティ検出への応用
3. 学会等名 2021年度 冬のLAシンポジウム
4. 発表年 2022年

1. 発表者名 清水 伸高
2. 発表標題 Hardness Self-Amplification: Simplified, Optimized, and Unified
3. 学会等名 Symposium on Theory of Computing
4. 発表年 2023年

1. 発表者名 清水 伸高
2. 発表標題 Hardness Self-Amplification from Feasible Hard-Core Sets
3. 学会等名 Foundations of Computer Science
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------