

クリーク分割問題の定式化について

鮎川 矩義

東京理科大学 工学部 情報工学科

クラスタリングは、機械学習やデータマイニングで幅広く利用されている基本的なデータ解析手法である。クラスタリングをモデル化した最適化問題の一つとして、クリーク分割問題が知られている。この問題では、枝重み (正負あり) 付き無向グラフが与えられ、頂点集合を任意の個数のクラスターに分割し、クラスター内の枝重みの総和を最大化することが求められる。汎用性が非常に高い問題であり、実際、Régnier の問題、相関クラスタリング問題、モジュラリティ最大化問題、といった主要なクラスタリング問題がクリーク分割問題として記述される。

クリーク分割問題に対しては、Grötschel & Wakabayashi (1989) による 0-1 整数線形計画問題としての定式化が知られているが、頂点数の 3 乗に比例する膨大な数の制約式を含むため、数千頂点のグラフに対する適用は実用的には難しい。Miyachi & Sukegawa (2015) は、同定式化の制約式の冗長性を指摘し、新たな定式化を提案した。具体的には、枝重みの符号に基づく条件式によって最適解集合に無関係な制約式の一部を検出できることを示し、条件式を満たす制約式を事前に取り除くことを提案した。しかしこの定式化も、実験的には、従来の定式化と同程度の性能を示す場合も多く、その効果は限定的であると言える。

本発表では、実社会から生じるクリーク分割問題のインスタンスの性質に着目し、制約式の本数を大幅に減らした新たな定式化を提案する。提案する定式化は、数千頂点のグラフに対しても適用することができる。さらに、整数線形計画問題としての定式化に加えて、最大充足性問題としてのエンコーディングについても議論する。本発表の内容は、宮内敦史氏 (理化学研究所)、藪部知大氏 (国立情報学研究所) との共同研究に基づいている。