グラフ理論

柳本凌

　グラフ理論は生物学的データの分析に応用され、それは以下の４つの領域に分類される。

1. 生分子ネットワークのモデリング

１－１、転写調節ネットワークについて

　Vertex(node)は個々の遺伝子を、Edgeは遺伝子間の正負の調節を表し、方向性がある。

さまざまな応用例が存在するが、制限もある。それは、生化学ネットワークは動的であり、nodeとedgeは時間によって変化するからである。ただし、グラフがあらわす静的なつながりはネットワークの骨格を表すことができる。

１－２、タンパク質相互作用(PPIs)ネットワーク

　PPIsネットワークは主に非方向性のグラフであり、nodeはタンパク質を、edgeはPPIsを表す。PPIsに関するデータは膨大であり、それを十分に説明するためには複雑なモデルが必要であり、nodeやedgeに重要度を割り当て、スコア機構を作成することで達成される。

１－３、代謝ネットワーク

　代謝ネットワークは複雑であり、２つに分かれたグラフなどを利用することでより簡潔に表す。そのグラフでは、反応と要素をそれぞれ表すために２種類のnodeを使う。また、edgeは異なる種類のnodeのつながりである基質や生成物の関係を表す。

1. 生分子ネットワークの中心的役割や重要度の測定

　大規模で複雑なネットワークにおける最も重要なnodeを発見することはコンピュータ生物学において基本であり重要なことである。相互作用ネットワークにおける構造的に重要な遺伝子やタンパク質を発見するのにグラフ理論が応用された例は数多くある。

1. 生物学ネットワークにおけるmotifと機能的moduleの発見

　生物学ネットワークと工学ネットワークはモジュール性や回路要素の繰り返しなどの構造的原理を共有しており、よく比較される。Moduleは強い相互作用や共通の機能を有するnodeのセットである。基本的な機能を有するmoduleは頻繁に繰り返される。また繰り返される回路要素の小さなセット（３，４個のnodeセット）はmotifと呼ばれ、より頻繁に起こる。

1. 生分子ネットワークから新しい経路の発見

　研究室などで行われる実験によって、異なる生体での分子経路を発見し、その有効性を認めるためには多大な時間と労力が必要になる。グラフ理論を用いてコンピュータで経路を予測して自動化することで、生体システムの複雑なプロセスを明かすのに役立つ。