

遺伝子機能の関係を明示するグラフ操作手法

伊澤 亜紀子 (指導教員：瀬々 潤)

1 はじめに

遺伝子間の繋がりを表す遺伝子ネットワーク、Web、ソーシャルネットワークなど、グラフ構造を持つデータが増え、これらの解析が重要性を増している。近年、様々なグラフの解析の研究が行われているが [1]、既存研究で扱われたグラフと、近年のグラフで決定的に異なる点は遺伝子には機能、Web ページにはカテゴリ、SNS にはコミュニティなど、頂点に多数の属性情報が付与されている点に特徴がある。このように属性を付与されたグラフデータは、グラフ構造と属性が複雑に組み合わさっているためグラフ構造だけあるいは属性だけの解析ではユーザに必要な情報を提供できず、両者を同時に扱った多彩な視点の解析と可視化が求められている。

このような属性付きグラフを解析した先行研究 [2, 3, 4] はいずれも結果を静的に返すものであり、ユーザのニーズに対して解析、可視化手法が合致していれば良い結果となるが、少し異なった視点から見たい場合には融通が利かない困難さがある。また、パスウェイ(遺伝子ネットワークに類似した生体内ネットワーク)の例として KEGG [5] がある。ここで提供されている静的なネットワーク図は生物学的知識を活かして配置しているため機能情報ごとに分類され一見綺麗に配置されている。しかし、多様な生命活動の理解には着目する機能が近くに配置された、異なった配置を求める場合も多く、静的レイアウトの限界を示している。

このような問題点に対し、本研究では属性情報に着目したユーザビリティの高いレイアウト手法を提案する。本手法を用いることで、1,000 個の頂点、約 1 万本の辺を有する大きな遺伝子ネットワークからでも容易に遺伝子ネットワークのレイアウトを変更できることを示す。

2 提案手法

本研究では、頂点に属性が付与されたグラフから、有用な知識を可視化するため、属性に着目してユーザが配置を決めることが出来るインタフェースを提供する。

提案手法の概観を図 1 に示す。グラフ $G = (V, E, A)$ を与えられたグラフとする。ここで V は頂点集合、 E は辺集合、 A は頂点に付与された属性の集合を表す。

図 1(A) に与えられたグラフ例を示した。頂点数 $|V| = 13$ 、辺の数 $|E| = 15$ 、頂点の属性数 $|A| = 3$ のグラフである。図では頂点の属性を色で表しており、遺伝子ネットワークでは、遺伝子の持つ機能や属するパスウェイを示している。図 1(A) に示した図からは、各属性の関係性はわかりにくい。

提案手法は、同一の属性を持った頂点を同時に操作することができれば、属性付きグラフの操作が簡単になるのではないかと、という点に着目している。このため、属性ごとの関係性を示したグラフを作成する。まず、同一の属性を持つ頂点を一つのグループとし、頂点間にはグループ間に跨がる辺の数に応じた太さの辺を張った新たなグラフを用意する(図 1(C))。各グルー

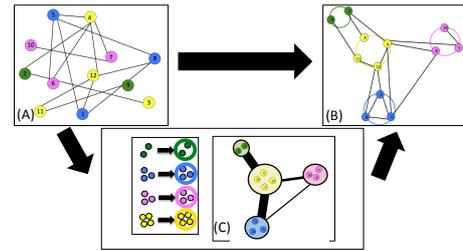


図 1: 提案手法の概観

プの大きさは属する頂点の数に比例した大きさとする(図 1(C))。本図で示されるグラフをグループグラフと呼ぶ。

グループグラフ描画の際にグループ間の配置手法が問題となるが、本提案ではグループの配置を手動で移動する事も可能なように設計する。これは組み込まれている自動のレイアウト(後述)を用いても構わないし、それを元に手でレイアウトを変更しても良い。グループの配置を可変にすることで、利用者の要望に応じた配置が可能となる。グループ数は一般に頂点数より遙かに小さいので、一般には操作しやすいグラフである。

グラフ G の再配置はグループグラフのレイアウト指定後、利用者の合図により行う。まず、各属性について指定された位置周辺に属する点を配置し、その後元の辺に沿うように辺を配置する。図 1(B) が再配置後の図である。このように再配置されたグラフ G を見ると、グループ間の関係が改めて明確になることが分かる。またグラフ G に対して操作すべき頂点数も 4 つであり、少ない数の移動でレイアウト変更が可能である。

3 実行結果

3.1 Cytoscape

本研究では実装に遺伝子ネットワークの研究で頻繁に用いられるグラフ描画ソフトである Cytoscape [6] を利用した。Cytoscape は Java 等の言語によるプラグインを作成することで、機能拡張を簡単に行えるため、本手法の実装にも適している。

Cytoscape では複数の異なるグラフを表示できるため、同時に複数のグラフを操作することが可能である。本実装では、一つの窓に表示された指定したグラフに対し、プラグインをメニューから実行すると新しい窓が作られ、その窓に図 1(C) に相当する図が表示されるようになっている。

3.2 実行画像

実際の酵母から採取されたデータ [4] を解析して得られるネットワークに本可視化手法を適用し、有用性を検証する。利用するネットワークは、ノード 1,184、エッジ 9,807、属性 65 を有するネットワークである。

図 2(A) は処理前のネットワークであり、複雑に辺が張り巡らされている。この状態で、本手法を適用すると図 2(C) に示すグループネットワークが作成される。

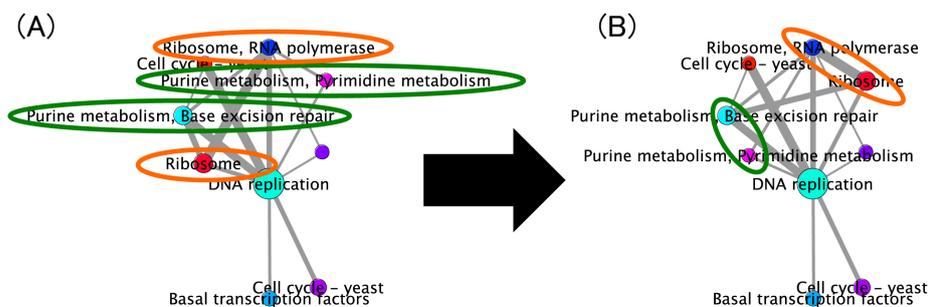


図 3: 酵母データから生成されたグループネットワーク

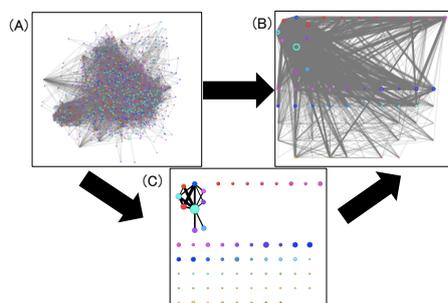


図 2: 実行画像

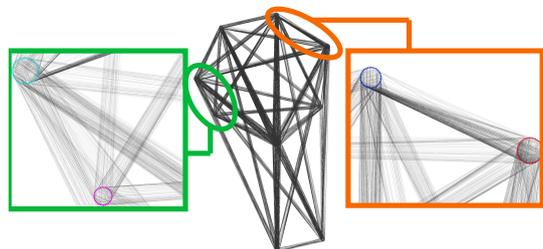


図 4: 機能が近いグループ同士を近くに配置したネットワーク

この図では各ノードは同一の属性を有する頂点（遺伝子）の集合を表しており、任意の位置に点を移動することが可能である。また、ノードの移動は Cytoscape に実装されているノードの自動レイアウト機能を用いて移動することも可能であるため、全てをユーザが指定する必要は無い。自動レイアウト後ユーザが頂点の位置を移動する事も可能である。移動後再度プラグインの実行を行うと、配置したノードの位置に従って、図 2(B) に示す図の様にレイアウトされる。

図 3 に解析例を示した。図 2(B) の左上部分の連結したネットワークを拡大した物である。初期のレイアウト（図 3(A)）では、グラフ構造のみを元にネットワークが描かれているため、緑で囲われた塩基代謝に関わる同一の機能群が離れた位置に配置され、またオレンジで囲われたリボソームに関連する遺伝子群も離れた位置に配置されている。この配置から、塩基代謝とリボソームに関連する機能を知るために、互いに近い位置に配置した物が図 3(B) である。この図から元のグラフの再配置を行う事が可能である。

緑の塩基代謝には 83 遺伝子が、オレンジのリボソームには 86 遺伝子が存在するため、提案するグループグラフによるレイアウト機能が無い場合、169 遺伝子

を選択して移動する膨大な作業を行わなければならないが、本提案手法を使えば 4 ノードの移動だけで事足りる。更に、グループグラフでは属性間の辺の多さも見る事が可能である。例えば、オレンジのリボソームのグループ群は互いに密に辺が張られている事がわかるが、右の塩基代謝に関わるグループは互いに辺がほとんど張られていない事が、グループグラフの図から判定できる。このように、機能間の関係を見る目的においても、グループグラフは有用である。

4 考察と今後の課題

本研究の実行結果より、グラフを利用者の要望に応じた属性ごとの配置にすることができた。属性情報間の配置に関しては、利用者に任意の方法で配置してもらうことで、ユーザの要望通りの配置が可能となった。属性数が多く人手でのレイアウトが困難な場合は、Cytoscape の自動レイアウトと併用することで、属性ごとの配置が可能となっている。よって本可視化手法を用いることで、自動レイアウトを各々のノードの配置に対してだけでなく、集合単位の自動レイアウトとしても利用できると考えられる。

頂点や辺の数が増えると、グループに属する頂点数やグループ間に跨る辺の本数も多くなる。それにより、グラフの頂点をグループごとに配置することができても、グループ間のエッジが雑然と並び、明確さが失われてしまうことがある。

今後は、頂点数や辺の本数に関わらずグループ間の関係を明確に保持することが課題である。

参考文献

- [1] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, vol. 99, 12, 7821–6, 2002.
- [2] M. Seki and J. Sese. Identification of active biological networks and common expression conditions. *BIBE2008*, 1-6, 2008.
- [3] T. Itoh, C. Muelder, K. Ma, and J. Sese. A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs. *IEEE Pacific Visualization Symposium 2009*, pp.121-128, 2009.
- [4] Aika Terada and Jun Sese. Discovering large network motifs from a complex biological network. *Journal of Physics: Conference Series*. 012011, vol. 197, 2009.
- [5] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Research*. Vol. 28, No. 1, pp. 27–30, 2000.
- [6] C.S. Melissa, *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols*. vol. 2, pp. 2366–2382, 2007.