DEIM Forum 2010 E4-5

機能間の関係を明示する遺伝子ネットワークの可視化

伊澤亜紀子† 瀬々 潤†

† お茶の水女子大学理学部情報科学科 〒112–8610 東京都文京区大塚 2–1–1
E-mail: †{izawa.akiko, sesejun}@sel.is.ocha.ac.jp

あらまし 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解析が近年重要性を増している．特に遺伝子には機能，Web ページにはカテゴリ，SNS にはコミュニティなど，頂点に属性情報が付与されたネットワークが多い．このようにネットワークと属性の入り交じったデータは，ユーザによって異なる視点から解析をしたい場合が多いため多様な可視化が求められている．本研究では，特に属性情報の関係性に着目し，ユーザが属性情報を任意にレイアウトすることで，全頂点の属性毎の関係を簡単に可視化できる手法を提案する．また，本研究ではグラフ描画ツールである Cytoscape のプラグインとして実装した上で 1,184 ノード，9,807 エッジ，65 属性を持つ生物学の大規模ネットワークに適用し，遺伝子機能間の関係が簡単に可視化できることを示す．
キーワード グラフ，可視化，バイオインフォマティクス

Making of gene networks that specify the relations between functions visible

Akiko IZAWA† and Jun SESE†

† Dept. of Computer Science, Ochanomizu Univ. Otsuka 2–1–1, Bunkyo-ku, Tokyo, 112–8610 Japan
E-mail: †{izawa.akiko, sesejun}@sel.is.ocha.ac.jp

Abstract Recently, large graph structures such as gene network describing relations between genes, the Web and social netowkrs (SNS) have been studied. One of the common points of the networks is that the nodes of the networks have some attirbutes. For example, genes have their functions, and persons in SNS belong to communities. The complex dataset has been currently analyzed by various points of views. In this study, we focus on the relationships betwenn attributes on vertives, and we provide a simple visualization interface with which user can determine own group layout of attirbutes. We implimented the user interface as a plugin of Cytoscape, a visualization tool of large graphs, and applied a biological network containing 1,184 nodes, 9,807 edges and 65 attributes. A demonstration shows that our method allows us to understand the relationships between gene finctions easily.
Key words Graph, Visualization, Bioinformatics

1. は じ め に

遺伝子ネットワーク，Web に見られる様な大規模グラフ構造情報への注目が高まり，このグラフを解析する手法の研究が盛んに行われている [1], [2]．グラフ構造の解析では，データマイニングによる解析アプローチも数多く行われているが，ユーザによって興味ある場所が異なるため，複雑なグラフを可視化することによって，目視により関係性を知る事も重要である．

この目的に対しグラフ構造を可視化するツールが開発されている [3]．しかし，これらいずれのツールも頂点と辺から成るグラフ構造の可視化に着目していることから，木構造の様な可視化により状態を容易に把握できるネットワークを除き，一般的な大規模かつ複雑なネットワークを可視化すると，雑然としたネットワークが描かれがちである．実際，単細胞生物の酵母のネットワークでさえ複雑な形状となり生命現象の理解には程遠い (図 3(A) に図示の例を示す)．

それに対し，特定の用途に限って人により綺麗にレイアウトされた図が存在する．一例として，Kyoto Encyclopedia of Genes and Genomes (KEGG) [4] の代謝マップは遺伝子とその関係が生命科学的知識に基づき綺麗にレイアウトされている．しかしこのマップ自身も本質は変わらず，存在が確かな辺を中心に書かれているため疎なグラフとなっているが，よく見ると遠く離れた点に対して辺が張られている場合がある．

このような代謝の図は，一般に作成者の意図によって描かれている. そのため，興味ある現象の異なる研究者が見た場合，本当は近くにあって欲しいパスウエイが遠く離れた場所に配置さ

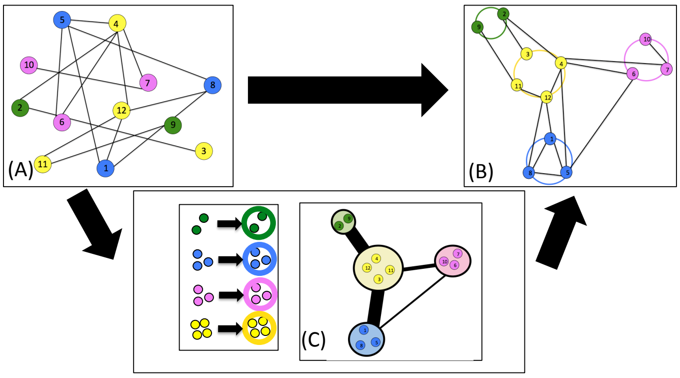

図 1　提案手法の概観

れ，この分裂したグラフを往復しながら，関係を見なくてはいけないことが起こる．

この問題を解消するため，インタラクティブに頂点の編集が可能なグラフ描画ソフトでは，利用者が頂点の位置を移動させれば問題を解決できると考えるかもしれない．しかし，代謝のパスウエイや SNS のコミュニティの場合，移動したいのは単独の頂点ではなく，同一グループに属する数十の頂点であり，これらを一つずつ選択して移動するのは困難である．本研究では，このように近年のグラフ構造では頂点に属性が付与されている事が多い事に着目し，ラベル毎にユーザが欲するレイアウトに移動できるインタフェースを提供する．

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，属性毎に利用者が位置を決めることが出来るインタフェースを提供する．

2.1 作成手順

本節では提案手法の概観を示し，以下の章で実装の詳細について説明する．

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．また，$a(v)(v \in V)$ で頂点 v の属性を表す．

図 1(A) に与えられたグラフ例を示した．頂点が $0, \ldots, 12$ の 13 点，辺が 15 本のグラフである．頂点の属性は色で表してあり，ここでは青，緑，紫，黄の 3 つの属性が存在する．属性は，遺伝子ネットワークに関して言えば，遺伝子の持つ機能や代謝パスウエイを示している．図 1(A) で示したグラフでは，各属性の関係性はわかりにくい．

次に属性毎の関係性を示したグラフを作成する．同一の属性を持つ頂点を一つのグループとする．各グルー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(B))．更に，グループ間に跨がる辺の数に応じてグループ間に辺を張る (図 1(C))．本図で示されるグラフをグループグラフとここでは呼ぶ．

ここで，グループグラフにおいてグループ間の配置手法について問題となるが，本提案では任意の方法で行う．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，

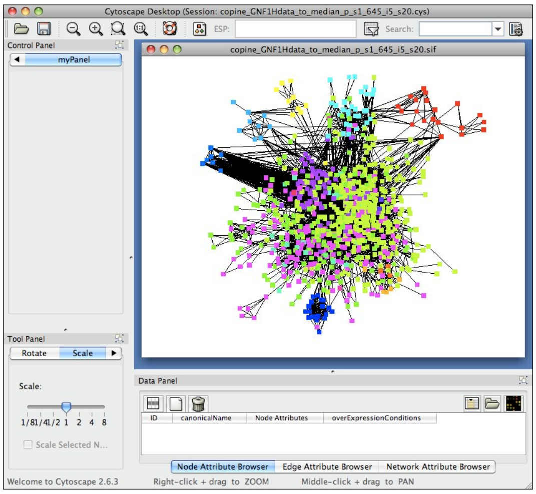

図 2　Cytoscape の画面

それを元に人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．

元のグラフ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(C) が再配置後の図である．このように再配置された元のグラフを見ると，グループ間の関係が改めて明確になることが分かる．

3. 実行結果

本節では提案手法の実装とその結果について示す．

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・解析ソフトである Cytoscape [3] 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 の表示画面を図 2 に示す．Cytoscape では窓毎に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(B) 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，寺田ら [5] によって解析されたネットワークであり，遺伝子数（頂点数）1,184，辺の本数 9,807 本，属性数 65 個を有している．

図 3(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，本研究で開発したプラグインを実行すると図 3(B) に示す図が作成される．この図では各ノードは一つの属性に属するノードの集合を表しており，任意の位

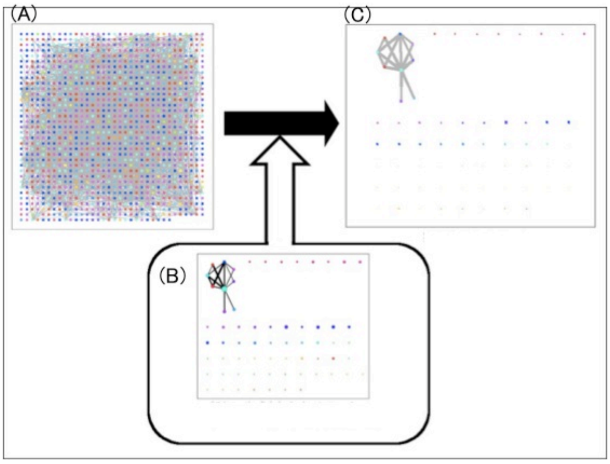

図 3　実 行 画 像

置に点を移動することが可能である．また，Cytoscape にはデフォルトでノードの自動レイアウト機能が組み込まれている．バネモデルや円環上への配置など，主要なレイアウトが可能である．このレイアウト機能を用い，任意の位置にノード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 3(C) に示す図の様にレイアウトされる．

図 4 に解析例を示した．図 3(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 4(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．

機能が近いグループ同士の関係を知りたいため，近くに配置した物が図 4(B) 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ことが可能である．たとえば，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置いても，グループグラフは有用である．

3.3　考　　察

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. しかし，属性ごとで配置が可能となった一方で，ノードやエッジの数が膨大になるほど，属性間のエッジが雑然と並び，関係の明確さが失われる傾向がある. このことから，どのように属性間の関係の明確さを保持するか，考慮する必要がある.

4.　関 連 研 究

本研究は遺伝子につけられた機能情報に着目し，機能情報ごとの配置をおこなっている. これはすなわち，クラスタごとに配置をすることと同等である. 公開されている Cytoscape のプラグインの中でも MCODE [6] がある. MCODE はクラスタ検索，サブネットワークの抽出，クラスタ毎の解析が可能である.

伊藤等によりアイテムが付与されたネットワークの可視化手法が提案されている [7]．しかし，この手法は計算機がレイアウトを決定するため，ユーザの望まないレイアウトが構成される可能性もあるため，未知な情報の多い生物データには，本手法の方が利用価値が高い.

5.　まとめと今後の課題

本研究では頂点に属性が付与されたネットワークについて，属性情報の関係性が明確になるような再配置をおこなう可視化手法を提案した. 第一にグループ毎のレイアウトをユーザが独自に決められるインタフェースを簡易グラフを描くことで提供し，ユーザがレイアウト後，そのレイアウトを参考に元のグラフを書き直す方法を提供した.

また，簡易グラフは同一のグループに属する頂点の数やグループ間に跨がる辺の数に比例して，頂点の大きさ，エッジの太さを決定しているため，元のグラフよりもグループ間の関係性がわかりやすく，関係の強さも明確である. これにより利用者はグラフを，グループ間の関係性を把握しながら配置できるようになった. しかし，頂点や辺の数が多くなると，グループに属する頂点数やグループ間に跨が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して，明確さが失われてしまうことがある. また，本可視化手法ではグループごとに配置をおこなったが，グループ中の頂点の配置に関しては関係が明確ではない.

今後は，超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，そしてグループ間だけでなく，グループ中の頂点同士の関係性も明確にするような配置をおこなうことが課題である.

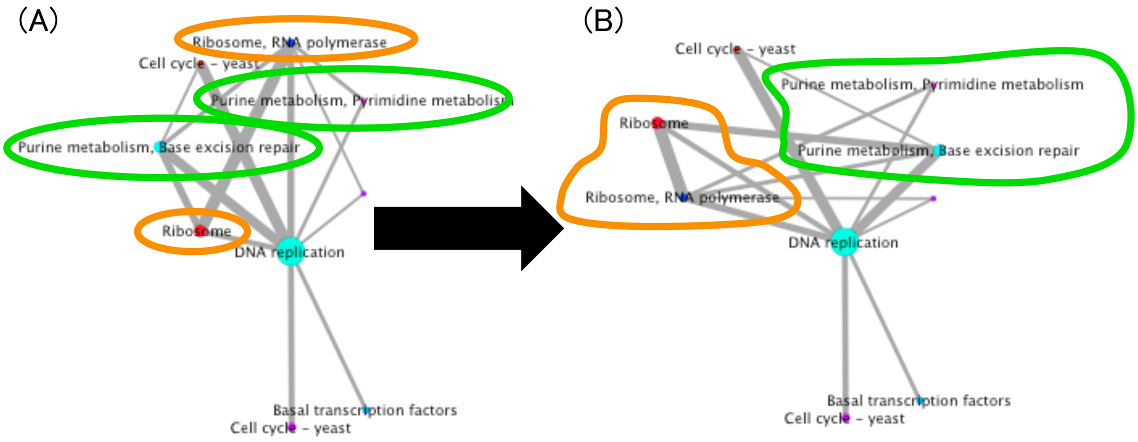

図 4　酵母データから生成されたグループネットワーク

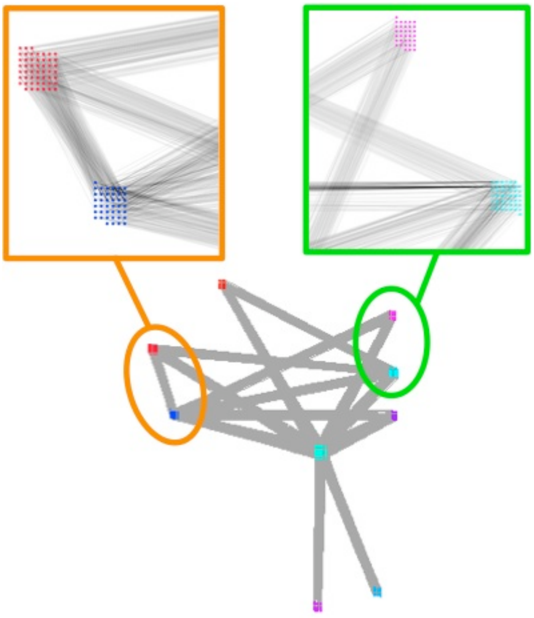

図 5　機能が近いグループ同士を近くに配置したネットワーク

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] A. L. Barabási and Z. N. Oltvai. Network biology: understanding the cell's functional organization. Nature Reviews Genetics. Vol. 5, No. 2, pp.101–113, 2004.

[3] C.S. Melissa, *et al.* Integration of biological networks and gene expression data using Cytoscape. Nature Protocols. vol. 2, pp. 2366–2382, 2007.

[4] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Research. Vol. 28, No. 1, pp. 27–30, 2000.

[5] Aika Terada and Jun Sese. Discovering large network motifs from a complex biological network. Journal of Physics: Conference Series. 012011, vol. 197, 2009.

[6] Gary D Bader and Christopher WV Hogue. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. vol. 4, issue 2, 2003.

[7] 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.