

大規模階層型データ可視化手法を用いた遺伝子ネットワークの可視化

西山慧子 伊藤貴之

お茶の水女子大学 理学部情報科学科

1. 概要

情報可視化は世の中にある一般的な情報を可視化する研究分野であり、最近になって生物情報の可視化の研究が活発に進められている。現在ヒトゲノム解読は完了しているが、DNAを構成する塩基配列が解読されただけで、その遺伝子のふるまいなどは、はっきり分かっていない。そこで現在その遺伝子の振舞いについての研究が必要とされている。その中でもマイクロアレイデータ[1]からの遺伝子ネットワーク同定問題は、バイオインフォマティクス分野における重要なトピックのひとつであると言える。

遺伝子ネットワークとは、各遺伝子をノードとし、遺伝子間にエッジがあるようなグラフ構造で、ゲノム上での位置関係、代謝、制御パスウェイ上での隣接関係、転写時の共発現率、たんぱく質相互作用など、多くの性質を表現するために用いられる。またパスウェイなどの相互関係は、エッジが方向を持つ。これらを分析することで、遺伝子に変異したときに何が起こるか予測することができる。

このデータは数千の遺伝子群からなることもあり、複雑な連結成分をふくんでおり解釈や把握が困難であるので、何らかの方法でより興味深い遺伝子群を抽出し注目すべき対象を絞り込むことが必要である。しかしながら常に目的に叶った結果をきりだせていないのが現状である。

また遺伝子解析において、それぞれの相互関係を示す現象として、マルチドメイン(2種の遺伝子の機能を同時にもち遺伝子)等が挙げられる。

これらの遺伝子は、発現率によるクラスタリング結果と遺伝子ネットワークを組み合わせて可視化することにより、その存在が理解しやすくなると考えられる。

そこで本研究では、例えば複雑なネットワークの中から、遺伝子のクラスタリングの結果を同時にみせつつ、ネットワーク中の注目部分を抽出できる可視化手法を提案する。ここでいう注目部分とは、例えば検索結果である。この研究結果により、研究者は全体の中から特定の遺伝子の相互関係を知りたい時、混乱せず抽出でき利用できるのではないかと考えている。

2. 関連研究

ネットワークの可視化手法は、すでに多く発表されているが、遺伝子ネットワークの研究に向いているネットワーク可視化は、複雑な操作がいらず、簡単に注目部分だけを抽出でき、大規模なネットワークを扱える手法である。

注目部分を強調する為のネットワーク可視化の研究は、すでに多く発表されている。データ中の注目部分を、魚眼レンズでみたように拡大表示し、その周囲部分を画面の端によせる可視化手法が報告されている[2]。だが拡大表示部分の外にあるノードは、エッジで連結されているのに画面の端においやられてしまう。また、データ中の注目部分を3次元的に引き上げて、それとエッジで連結されているノードも浮き上がらせることで、データの注目部分の相互関係をわかりやすく表示する手法が報告されている[3]。だがこれも、複数のノードを同時に注目したい時の対応がなされていない。

3. 提案内容

本研究では、ネットワークのデータ中の注目部分の抽出に向けた情報可視化を実現するために、以下のような点を重視する。

- (1) 注目している部分を強調でき、相互作用関係を判別しやすいこと
- (2) できるだけ多くの遺伝子をクラスタ単位で表示できること

そこで(1)を満たすために、データ中の注目部分を、3次元的に引き上げることにより、そのノードとエッジでむすばれているノードの関連を強調させ、周囲との関連の表示を行う。

また(2)を満たす為、まずxy平面に、クラスタごとにノードを配置し、続いてそのノードのx座標値、y座標値を変更することなく相互関係を強調させるようにする。xy平面上のノード配置には、平安京ビュー[4]を拡張した手法を提案する。

本手法は、M個のマイクロアレイ上にそれぞれにあるN個の遺伝子、 $M \times N$ のマトリクス型データを対象とし、N個の遺伝子のうち発現率傾向の近いものをクラスタリングし、平安京ビューにおける階層型データに変換する。以下の図1に概要を記す。本研究は、3次元空間の底面に平安京ビューを用いて、遺伝子のクラスタリング結果を可視化する。このノード一つ一つが、一つの遺伝子に対応し、クリックブルになっており、その遺伝子の詳細情報を提示することができる。

Visualization of genome network using large-scale hierarchical data visualization technique

Keiko Nishiyama, Takayuki Itoh,
Ochanomizu University

{nishy, itot}@itolab.is.ocha.ac.jp

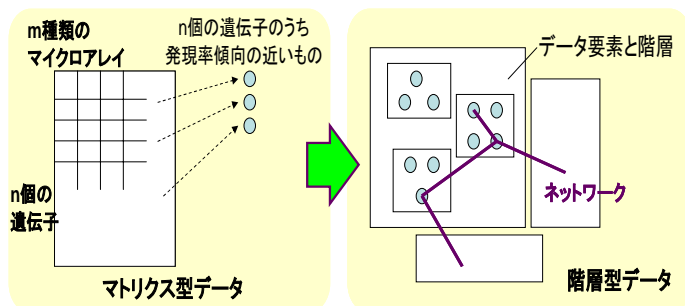


図 1：階層型データへの変換

またその上にネットワークを重ね、注目ノードをクリックするか、または検索エンジンのような GUI で指定すると、その注目ノード、およびそれをエッジで連結されたノードを、3 次元的に引き上げて表示する。

ネットワークは、2 個のノード（遺伝子）のペア、nodeA, nodeB とし m 個のマイクロアレイがあったとすると、

nodeA の発現率を $A=\{a_1, a_2, \dots, a_m\}$

nodeB の発現率を $B=\{b_1, b_2, \dots, b_m\}$ とする。

発現率どうしの相関性を算出する式は、

$1.0 - \text{distance}(A, B) / \text{maxDistance}$ とする。

なお $\text{distance}(A, B)$ は A, B 間のユークリッド距離の二乗であり、 maxDistance はその最大値である。この算出式は、クラスタリングに使用したソフトウェア Cluster3.0 にも用いられている。の値が一定値（自分で与えた値）より大きい時、この 2 つのノードを接続するエッジを表示する。

また、クラスタ間をまたぐエッジを多く持つ遺伝子を、あらかじめ、光らせるようにする。これにより、特殊な反応のありそうな遺伝子群を表示する。

本研究により、クラスタリングを行い平安京ビューで表示を行う為、整然とした表示可能になり、全体を一画面表示することにより遺伝子などの膨大なデータに対応できる。また、ネットワークを重ねることにより、相互関係が理解しやすく、クラスタリングだけでは表現しにくかった特性（マルチドメインなど）を表現できる。

4. 実行結果



図 2：提案手法を用いた、注視ノードが一つの実行例
本手法による可視化の例を図 2 に示す。ネットワー

クを表示させ、注視ノードを一つとし、そのノードと結ばれているエッジだけを表示している。また、クラスタ内のノードすべてと、注視ノードがつながっているクラスタを丸で囲んだ。

この実行例から、この注視ノードは、丸で囲んでいるクラスタ内すべてと、結ばれていることより、すべての属性をふくんでいるマルチドメインかもしれないという事が発見できた。また、このデータにおいて、注視ノードから出ているエッジは、先ほど光らせたノードとほとんど結ばれていることより、クラスタを多くまたぐエッジを多く持つ遺伝子同士が複雑に絡み合いネットワークを作っているという事も発見できた。

5. まとめと今後の課題

本論文では、複雑なネットワークの中から、遺伝子のクラスタリングの結果を同時にみせつつ、ネットワーク中の注目部分を、3 次元的に引き上げることにより抽出できる可視化手法を提案した。

本手法において、ネットワークとクラスタを同時表示することにより、マルチドメインなどの特性を表現でき、またクラスタをまたぐエッジを多く持つ遺伝子を、色をつけることにより、おもしろい反応のありそうな遺伝子の絞込みもできることが解った。

今後の課題として、発現率の代わりに、グラフに向きのあるパスウェイを用いた遺伝子ネットワークの可視化を試みたい。また、オントロジーなどの遺伝子同士の関係性などもくわえていきたいと考えている。また、可視化された特徴的な現象が、本当に遺伝子学的に興味深い特性なのか否かの検証も必要である。

謝辞

本報告のクラスタリングに関して、コロンビア大学 Mickael De Hoon 氏から、cluster3.0 のソフトを提供していただいた事を感謝いたします。

遺伝子ネットワークに関する議論に関して、東京大学宮野教授、中谷助教授、渋谷講師、瀬々助手、井本助手から貴重なご意見を賜ったことを感謝いたします。

参考文献

- [1]有田正規、遺伝子ネットワークと確率モデル、2001 年ベイジアンネットチュートリアル。
- [2]Manojit Sarcar, MarcH .Brown, Graphical Fisheyes Views of Graphs, Communication of the ACM, pp73-83, Volume 37, March 1994.
- [3]塩澤他、「納豆ビュー」の対話的な情報視覚化における位置付け、情報処理学会論文誌, 38, 11, pp. 2331-2342, 1997
- [4]伊藤、小山田。平安京ビュー～階層型データを基板状に配置する視覚化手法、可視化情報学会第 9 回ビジュアルリゼーションカンファレンス、2003.