芸術科学会論文誌 Vol. 6 No. 3 pp. 106 - 116

「平安京ビュー」を用いた階層型遺伝子ネットワークの可視化

西山 慧子　　伊藤 貴之
お茶の水女子大学大学院
E-mail : {nishy, itot}@itolab.is.ocha.ac.jp

概要

遺伝子ネットワークとは，各遺伝子をノードとし，遺伝子間をエッジで接続して構築されるデータである．数千，数万といった大量の遺伝子群で構成される遺伝子ネットワークには，複雑な連結成分を含むことが多く，その解釈や把握が困難な場合も多い．

本論文では，遺伝子群にクラスタリングとネットワーク化を同時に適用して構築される，階層型ネットワークデータを対象とした可視化手法を提案する．提案手法では，各々の遺伝子は数種類の発現率を持つと仮定し，その発現率の相関性の高さによりクラスタリングを行う．それと同時に提案手法では，発現率の相関性が高い遺伝子間をエッジで連結することにより，ネットワークデータも同時に生成する．提案手法ではこの遺伝子ネットワークデータを，大規模階層型データ可視化手法「平安京ビュー」の拡張手法を用いて可視化する．提案手法を用いることにより，遺伝子学の研究者は，膨大な遺伝子群の中から，特定の遺伝子の相互関係を分析，あるいは興味深い特徴を持つ遺伝子の発見，などが容易になるものと考えられる．

なお本論文は遺伝子ネットワークの可視化を試みるものであるが，提案手法における階層型ネットワークデータの可視化手法は，拡大性とランダム性の高い複雑ネットワークと呼ばれるネットワーク全般に適用可能な，応用範囲の広い可視化手法である．

Visualization of Hierarchical Gene Network Using HeiankyoView

Keiko Nishiyama　　Takayuki Itoh
Graduate School of Humanities and Sciences, Ochanomizu University

Abstract

Gene network is a network that denotes each gene as node and connects pairs of genes by edges. It composes of a large amount of gene cluster, and therefore they contain complex connected elements, which are often difficult to interpret and grasp.
This report presents a technique for visualizing hierarchical gene network data, constructed by clustering and networking techniques. The technique assumes that each gene has multiple expression rate values, and they are clustered and networked according to the correlation of the expression rate values. We visualize the data by using the enhanced "HeiankyoView", which has been originally presented as a large-scale hierarchical data visualization technique. Our technique makes easier to analyze specific genes and discover genes which has interesting features.

1. はじめに

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[1]である．その応用範囲は非常に広いが，最近では特に生物情報の可視化の研究が活発に進められている．生物情報の中でも特に急速に研究が進んでいる分野に，遺伝子（ゲノム）解析があげられる．現在既に，ヒトゲノム解読は完了しているといわれているが，これは DNA を構成する塩基配列が解読されたというだけであり，その遺伝子の振る舞いなどは，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[2]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ッ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ネットワーク構造で，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，蛋白質相互作用など，多くの性質を表現するために用いられる．遺伝子ネットワークは多くの場合において無向グラフとして扱われるが，パスウェイなどの遷移関係を表す場合に限って有向グラフとして扱われる．このようなネッ

トワーク構造を分析することで，遺伝子が変異したときに何が起こるか予測することができる．

遺伝子ネットワークは大変膨大なものであり，複雑な連結成分を含むため，そのままでは解釈や把握が困難である．よって，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得る事ができていないというのが現状である．情報可視化はこのような目的において非常に有効であると考えられる．

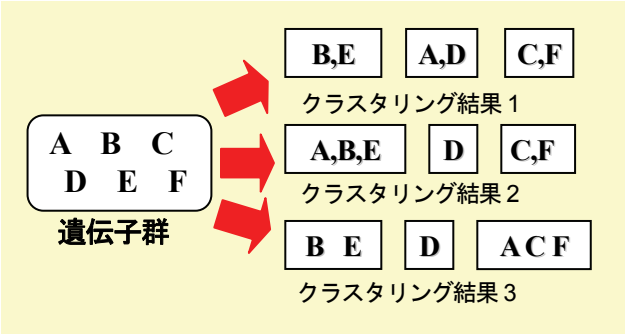


図 1. クラスタ生成の一例

本論文では，数万から数十万の遺伝子発現を一度に調べることが可能である，マイクロアレイデータを参照して各遺伝子に数種類の発現率を仮定し，この相関性の高さで遺伝子をクラスタリングし，さらに相関性の高い遺伝子同士をエッジで接続した階層型ネットワークデータを想定する．このとき図 1 で示すように，クラスタリングの方法や実行条件によって，クラスタリング結果はさまざまに変化する．図 1 より，Aは{B,E},{D},{C,F}の 3 組の遺伝子と同一のクラスタに属する可能性があるといえる．このことより，Aは複数の遺伝子の機能を同時に持つ遺伝子かもしれない，と予測できる．このように，2 種以上の遺伝子の機能を同時にもつ遺伝子は，マルチドメインと呼ばれ，この発見は遺伝子分析の中でも興味深い問題である．しかし 1 つのクラスタリング結果だけを可視化しても，このような特性は発見しにくい．このような現象は，遺伝子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．

本論文では，遺伝子群に対してクラスタリングとネットワーク化を同時に適用して生成される，階層型ネットワークデータの可視化手法を提案する．提案手法は図 2 に示すような，異なるクラスタ間をまたいで相関性を有する遺伝子間をエッジで表現することで，マルチドメインに代表される遺伝子の興味深い現象の発見に貢献するものである．提案手法では情報可視化手法「平安京ビュー」[3]を用いてクラスタリング結果を階層型データとして可視化し，それにエッジを重ねて描くことにより階層型ネットワークデータを表現する．

なお本論文が提案する階層型ネットワークデータ可視化手法は，3.4 節にて後述するとおり，大規模かつランダム性の高い複雑ネットワーク全般に適用できる，きわめて適用範囲の広い手法である．

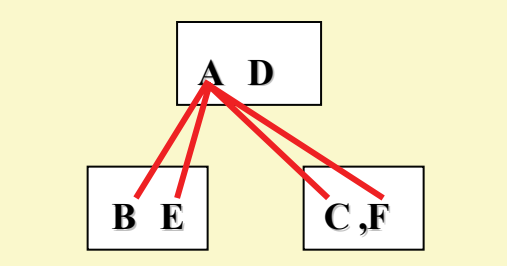


図 2. クラスタにネットワークを重ねた一例

2. 関連研究

2.1 遺伝子情報の可視化

遺伝子情報は大規模かつ複雑な性質をもつことが多い．そのため，その全貌を理解するために情報可視化技術は有用であると考えられる．特に，3 章で後述するマイクロアレイから得られる遺伝子情報の可視化は，近年活発に議論されており，その諸手法を比較する論文も発表されている[4].

マイクロアレイから得られる遺伝子情報は，一般的には遺伝子名および実験方法を行と列にする表形式データとして与えられる．このような表形式データを可視化する最も単純な方法は，表の構造をそのまま画面上に表現する技術である．遺伝子分析の分野で最も有名な TreeView[5]というオープンソースの可視化技術は，まさに遺伝子名と実験方法を行と列にした表形式の可視化を実現している．

しかし遺伝子情報に限らず，表形式データの中には，情報が非常に大規模かつ疎であるものも多い．そのため，このようなデータをそのまま表として表示することは，画面空間の有効利用の点で必ずしも合理的であるとは限らない．これを改善する一案として，表形式データからクラスタリングによって形成される階層型データ，あるいはゼロでない値をもつ行と列を連結して形成されるネットワークデータに変換してから表示する，という試みが多く行われている[6]．本論文の提案手法は，この考え方に基づき，表形式データとして与えられる遺伝子情報を，階層型ネットワークデータに変換して可視化するものである．近年では，表形式データを表のまま可視化する手法と，木構造やグラフに変換して可視化する手法との比較に関する研究も発表されている[7].

2.2 ネットワークデータの可視化

ネットワークデータの可視化手法は，すでに多様な観点か

らの研究が進んでいる．ネットワークデータのノード位置の算出のために力学モデルを用いた手法[8]や，大規模ネットワークの部分拡大表示[9]やインクリメンタルな表示[10,11]を実現した手法などは，この研究分野を活性化した代表的な研究成果といえる．またウェブのリンク構造の可視化[12]をはじめとして，ネットワークデータの可視化の応用分野の開拓も活発に進んでいる．

複雑に絡むネットワークデータ中の注目部分をわかりやすく表示するための代表的な手法として，3 次元的な引き上げ操作により，注目ノード，および注目ノードとエッジで連結されているノードも連鎖的に引き上げて表示する「納豆ビュー」という手法が報告されている[13]．本論文の提案手法は，納豆ビューに類似した考え方で，ネットワークデータ中の注目部分を 3 次元表示するものである．

2.3 階層型データの可視化

階層型データの可視化手法の著名な手法の中には，階層構造を木構造として表現する手法と，階層構造の末端にあたる葉ノードを 2 次元的に画面空間に展開する手法がある．前者の中で有名な手法には，Hyperbolic Tree[14]や Cone Tree[15]があげられる．後者の中で有名な手法には，画面空間の 2 次元的分割により葉ノードを一括表示する TreeMaps[16]があげられる．本論文の提案手法が用いる階層型データ可視化手法「平安京ビュー」も，後者に属する手法である．本論文では大量の遺伝子情報を一画面に展開して一括表示することを目的としているため，後者のような階層型データ可視化手法のほうが適切であると考えられる．

階層構造とネットワーク構造の両者を併せ持つ可視化技術は，近年になって活性化している研究分野である．旧来の研究の例として，3 次元的に階層構造を表現するネットワークデータ可視化手法[17]や，クラスタごとにズーム値を変えた 2 次元的なネットワークデータ可視化手法[18]などが知られている．また近年では，Cone Tree や TreeMaps などの既存の階層型データ可視化手法にネットワークデータを付加する形式の可視化手法[19,20]が発表されると同時に，その画面上の混雑を回避するためのネットワークデータの曲線化に関する手法[6]も発表されている．本論文の提案手法では[6,19,20]と同様に，階層型データ可視化手法にネットワークデータを付加する形で，階層構造とネットワーク構造の両者をあわせもつ情報を可視化するものである．

2.4 階層型データ可視化手法「平安京ビュー」

本論文の提案手法では，「平安京ビュー」[21]を用いて遺伝子情報の階層構造を可視化する．「平安京ビュー」は，階層型データの葉ノードを長方形のアイコンで，枝ノードを長方形の枠で表現し，階層構造を 2 次元の長方形群の入れ子構造で表現し，これらをできるだけ小さい画面空間に配置することで，階層型データ全体を一画面に表示する．

この手法は 2.2 節でも論じたように，階層型データ中の葉ノードと枝ノードの親子関係よりも，階層型データ全体に分布する葉ノード群を全て一画面に表現することに主眼をおいた視覚化手法である．

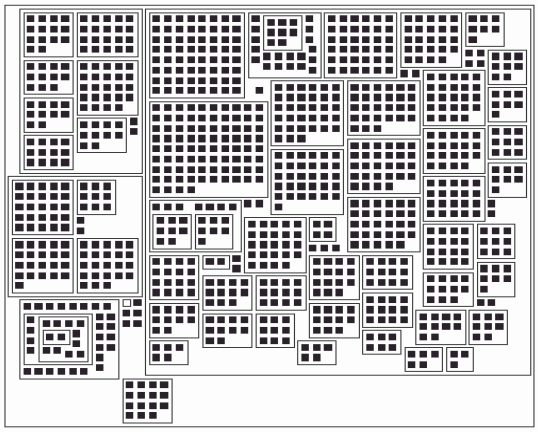


図 3. 「平安京ビュー」による大規模階層型データの可視化の例

平安京ビューの特徴の一つに，階層型データを構成する全ての葉ノードを，同じ大きさ・同じ形状で，かつ画面上で全く重なり合わないように表示する点がある．提案手法において遺伝子情報を画面上で探索する際に，全ての遺伝子が平等に同じ大きさで，かつ画面上で重ならないように表示されることは重要である．同じような特徴を有する階層型データ可視化手法に，TreeMaps[15]から派生した Quantum Treemap[21]という手法がある．Quantum Treemap と「平安京ビュー」の実行結果は文献[22,23]にて数値比較されている．この比較結果によると「平安京ビュー」は，部分領域のアスペクト比，類似データ間の可視化結果の類似度, の 2 点において Quantum Treemap よりも大幅に良好な結果をあげている．これらの利点もまた，階層化された遺伝子情報の可視化に有用であると考えられる．

3. 提案内容

3.1 本論文が対象とするデータ

DNA マイクロアレイとは, スライドガラスやシリコンなどの基板上に，数千数万のスポットが配置され，各スポットに一種類ずつ DNA や遺伝子を固定し，整列配置（アレイ化）したものの総称したものである．このスライドガラスに，化学反応実験を施すと，反応するスポットだけが蛍光する．各

スポットの蛍光強度をスキャナで読み取ることにより，発現率傾向を採取し，数千から数万の遺伝子発現情報を一度の実験で採取可能になっている．

また遺伝子は，複数の塩基により構成されており，遺伝子の発現率とは，一般的に遺伝子を構成する塩基群の中の反応した塩基の確率を指す．

本論文で扱う遺伝子ネットワークとは，このDNAマイクロアレイから採取される遺伝子の発現率を元に，構築されるネットワークを示す．遺伝子ネットワークの構築手法には，グラフィカル・ガウシアンモデルを用いた手法[24]等が知られている．また推定された遺伝子ネットワークの実用例として，薬剤ターゲット遺伝子の同定[25,26]等も発表されている．

3.2 階層型遺伝子ネットワークデータの構築

提案手法は，m 個のマイクロアレイ上に n 個の遺伝子があり，その各々の発現率が実数値として与えられているとする．提案手法では，この実数値から m×n の表形式データを構築し，n 個の遺伝子の発現率を m 次元ベクタとして扱うとする．そしてクラスタリングによって，発現率傾向の近い遺伝子が同一のクラスタに属するような階層構造を構築し，この構造を平安京ビューで表示可能な階層型データに変換する．さらに，この階層型データにネットワークデータを重ねるように表示することで，階層型ネットワークデータを可視化する．提案手法におけるクラスタリングおよびネットワーク化の手順の概要を図 4 に示す．

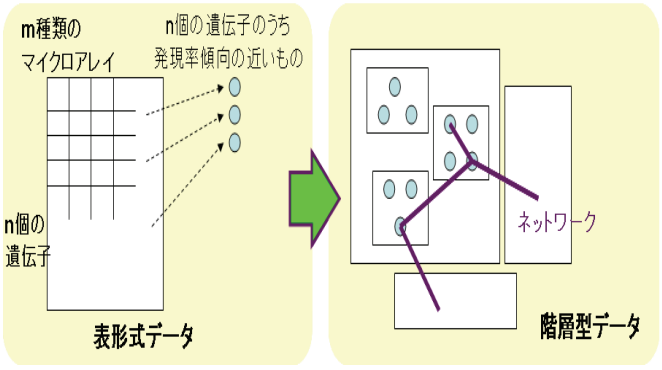
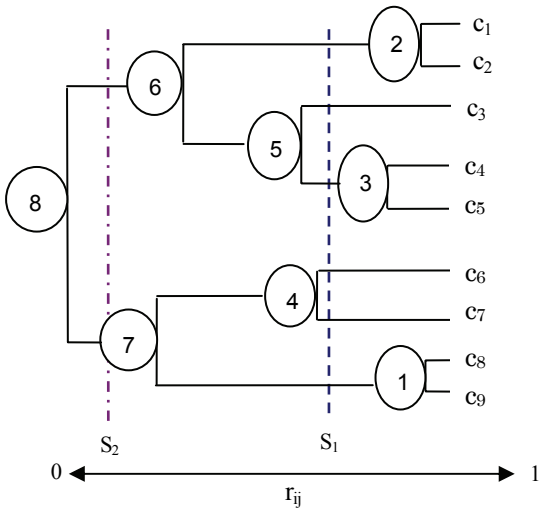


図 4. 階層型データへの変換

我々の実装では，Cluster 3.0 [27]というクラスタリングソフトウェアに実装されている階層的クラスタリングアルゴリズムを適用して，階層型データを構築する．図 5(上)において，クラスタをc_1～c_9とすると，提案手法では距離が近いクラスタに対して併合処理を反復することで，デンドログラムを作成し，階層的クラスタリングを実現する．このときクラスタ間距離に複数の閾値を設け，この閾値より距離の小さいクラスタを一階層に収める，という処理を反復することで階層型データを構築する．仮に図 5(上)に示すS_1,S_2の 2 つの閾値を設けたとすると，平安京ビューによる階層型データ可視化結果は図 5(下)のようになる．



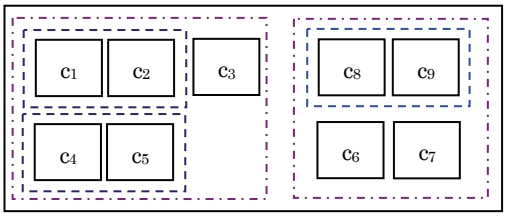


図 5. (上) 階層的なクラスタリング
(下) 平安京ビューにより表示されたクラスタ

続いてネットワーク化の手順について説明する．任意の 2 個のノード（遺伝子）を nodeA, nodeB とし，m 種類のマイクロアレイに対する発現率が与えられているとする．さらに，

nodeA の発現率を $A=\{a_1,a_2...,a_m\}$
nodeB の発現率を $B=\{b_1,b_2...,b_m\}$

とする．このとき nodeA と nodeB の発現率同士の相関性 r_{ab} を，以下の式で算出する．

$$r_{ab}=1.0-\frac{d_{ab}}{D_{\max}} \quad (1)$$

ただし d_{ab} は A,B 間のユークリッド距離の 2 乗で，

$$d_{ab}=\sum_{i=1}^{m}(a_i-b_i)^2 \quad (2)$$

で示される．$D_{\max}$ は，すべてのノードの組み合わせにおける d_{ab} の最大値である．提案手法では，r_{ab} 値が一定値より大きい時，この 2 つのノードを接続するエッジを表示する．

以上の算出式は，クラスタリングに使用したソフトウェア Cluster3.0 に導入された算出式である．クラスタリングとネットワーク生成の結果に一貫性を持たせるため，提案手法でも Cluster3.0 と同様に，ユークリッド距離空間を用いてネットワークを生成した．原理的にはユークリッド距離空間以外の距離空間（例えばマンハッタン距離空間）も採用可能であるが，その有効性について我々はまだ検証していない．

以上の処理により遺伝子データは，以下の要素から構成される階層型ネットワークに変換される．

- 遺伝子を表現するノード
  $N = \{n_1,...,n_n\}$
- ノード 2 個を連結する p 本のエッジ
  $e = \{n_i, n_j), E = \{e_1,...,e_p\}$
- 1 個以上のノードで構成される q 個のクラスタ
  $c = \{n_i,...,n_j), C = \{c_1,...,c_q\}$
- 階層型ネットワークデータ $D = \{N, E, C\}$

3.3 階層型ネットワークデータの可視化

本論文では 1 章にて述べたとおり，遺伝子群にクラスタリングとネットワーク化の両方を適用した階層型ネットワークデータの可視化手法を提案する．提案手法は以下の機能性を重視した手法である．

(1) できるだけ多くの遺伝子を一画面に, クラスタ単位で表示できること
(2) 注目したい任意の遺伝子を強調でき, その遺伝子と相関性の高い遺伝子の分布を理解しやすいこと

まず(1)を満たすために, 提案手法では 2.4 節で紹介した「平安京ビュー」を用いて，遺伝子を表すノード群 N を均一な大きさの正方形で表現し，クラスタ群 C を長方形の枠で表現する．これらのノードは画面上でクリッカブルな状態で表示されている．このため，クリック操作によって遺伝子の詳細情報などを提示するような GUI を構築することも可能である．

続いて(2)を満たすために提案手法では，特定の遺伝子を表すノード（以下, 注目ノードと称する）をユーザに指定させ，エッジ群 E の中から注目ノードに連結されているエッジだけを表示する．さらに提案手法では，注目ノード，および注目ノードから直接エッジで連結されているノード（以下，連結ノードと称する）を 3 次元的に表示する．ここで x,y,z の 3 軸で構成される直交座標系を仮定し，「平安京ビュー」による N および C の画面配置結果を平面 $z=0$ 上に表示するとする．提案手法では，注目ノード・連結ノード以外のノードは $z=0$ 上の正方形として平面的に描画するが，注目ノード・連結ノードはこの正方形を底面とする角柱として立体的に描画する．

このような立体的な描画を適用することを，本論文では以下「z 軸方向に沿って引き上げる」と称する．この引き上げる操作により，提案手法では注目ノードと連結ノードの接続性を強調表現することが可能になる．我々の実装では，平安京ビューの画面上で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キーボード入力による GUI で注目ノードを指定すると，その注目ノードおよび連結ノードを，z 軸方向に引き上げて表示する．

一般的に，膨大な遺伝子群の中から，注目すべき興味深い遺伝子を視覚的に発見することは容易ではない．ここで本研究の目的において，クラスタ間をまたぐエッジを多く持つ遺伝子は，マルチドメインなどの興味深い現象をもつ遺伝子である可能性が高い．そこで提案手法では，クラスタ間をまたぐエッジを一定以上有するノードを，あらかじめ所定の色で表示する．これにより，特殊な反応のありそうな遺伝子群を発見しやすくできる．

3.4 提案手法の応用例

世の中には，様々なネットワークデータが存在する．本論文では，遺伝子をノードとし，相関性の高い遺伝子をエッジで連結するネットワークを対象としているが，このネットワークは近年注目されている「複雑ネットワーク」の一種であると考えられる．

複雑ネットワークとは，際限ない拡大性を有し，ランダム度の高いリンク構造を有するネットワークの総称である．特に近年では情報技術の発達により，多くの分野において複雑ネットワークが見られる．例えば以下のようなネットワークは，複雑ネットワークの一種であると考えられる．

- 文書データベースに出現するキーワード間の相関性から構築したネットワーク．
- 計算機のアクセス履歴，コンピュータウィルスの感染経路などのログから構築したネットワーク．
- ニューロンやタンパク質の情報伝達経路から構築したネットワーク．
- 会社や社会の人間関係における様々な人間関係のネットワーク．
- ウェブのリンク構造のネットワーク．

本論文の提案手法は，遺伝子ネットワークに限らず，上記のような複雑ネットワーク全般に適用可能な，応用範囲の広い手法であると考えられる．

4. 実行結果

我々は提案手法を Java SDK 1.5 で実装し，COMPAQ EvoD510 CMT (CPU 2.8GHz, RAM 1GB) 上で実行した．オペレーティングシステムには Windows XP を用い，ディスプレ

イ解像度は 1024x768 画素に設定した. また遺伝子データとして, 以下の URL に公開されているイースト遺伝子発現率データを用いた.

http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/demo.txt

以下に結果画像を示し，結果画像に対する考察を論じる.

4.1 結果画像

まず我々は，クラスタリングにおいて作成したデンドログラムにおけるクラスタ間距離に 2 つの閾値（図 5 におけるS_1およびS_2）を設け，遺伝子データから 2 階層のクラスタを構築した．著者らの実験ではS_1=0.9, S_2=0.8 であった.

続いてS_1,S_2とは別に，もう一つの閾値を設け，式(2)におけるr_{ab}値が閾値以上である 2 ノード間にエッジを生成した．著者らの実験では閾値は 0.98 であった.

このようにして生成された階層型遺伝子ネットワークファイルを読み込み，提案手法を用いて表示した．提案手法によって各ノードの画面上の位置を算出するのに，0.24 秒を要した．なお「平安京ビュー」の処理時間や画面配置結果は，ノード数や階層の深さには単純に比例しない．「平安京ビュー」の処理時間や配置結果を悪化させる方向に影響を及す変数は，上位クラスタの配下に属する下位クラスタの個数である．よって提案手法のために複数の閾値を用いてクラスタリングを行う場合には，特定の上位クラスタの配下に属する下位クラスタの個数が大きくなりすぎないように，という点に留意して閾値を決定する必要がある．この閾値を適切に自動算出する手法の確立は，今後の課題のひとつといえる.

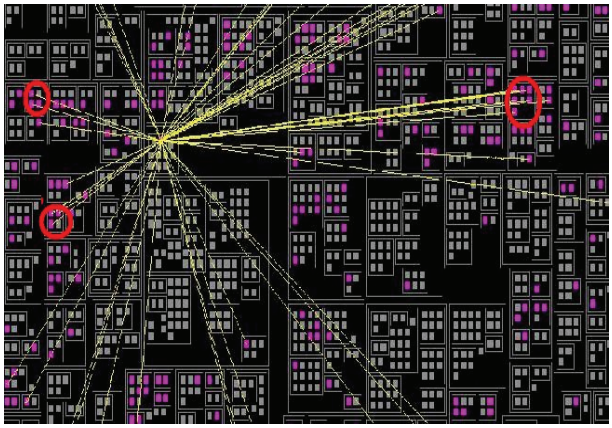


図 6. 提案手法を用いた，注視ノードが一つの実行例.

以上の処理による可視化の例を図 6 に示す．この可視化結果では，クリック操作によって単一のノードを注目ノードとして指定し，その注目ノードと連結ノードとの関係を黄色いエッジで表示している．クリック操作等に伴う再描画の処理時間は，画面のズーム率やエッジ表示数に大きく依存するので一概には言えないが，おおむね 0.2～0.5 秒程度であった．筆者らの実装では，ソフトウェアの可搬性の高さの観点から，GPU などのグラフィックス高速表示装置を全く用いず，また OpenGL や DirectX などの 3 次元グラフィックスライブラリを全く用いていない．これらを用いるように実装しなおすことで，描画時間は大きく向上すると考えられる.

一般的に，マイクロアレイにて一度に測定する遺伝子情報は，数百～数千個である場合が多い．一方で我々は経験上，1024x768 画素程度の画面解像度において「平安京ビュー」を用いる場合，ノード個数が 3000～5000 個程度であれば，階層型データ全体をクリッカブルな状態で一画面に表示できることを観察している．この性能は，提案手法においてクラスタリング結果から得られる階層型データにおいても同等であると考えられる．よって提案手法を用いることで，一般的なマイクロアレイ実験結果から得られる遺伝子情報を，一画面上に観察可能になると考えられる.

図 6 では，クラスタ内のノードすべてと注目ノードが連結している，というクラスタを丸で囲んだ．この赤丸で示すようなクラスタが存在するということは，注目ノードは現在属するクラスタの他に，丸で示すクラスタに属していてもおかしくない，ということを示している．つまり，この注目ノードが示す遺伝子はマルチドメインかもしれない，ということが推測できる.

また，図 6 を詳しく調べてみると，所定の色（紫）で表示されたノードを両端とするエッジが多く存在していることが解る．また他のノードを注目ノードに指定した場合にも，同様の結果が観察された.

このことより，図 6 に示す遺伝子ネットワークは，マルチドメインの可能性のある遺伝子同士が複雑に絡み合ったネットワークである，といえる.

図 7,8 は，提案手法により，注目ノードと連結ノードを z 軸方向に引き上げた結果画像である.

図 8(左)の注目ノードを引き上げていない画像では，どのノードが注目ノードとエッジで連結されているのか，一目には理解しにくい．それに対して図 8(右)では，注目ノードを引き上げることにより，注目ノードと連結ノードを一目瞭然に発見できることがわかる.

これらの結果画像より，ネットワークの注視部分を z 軸方向に引き上げることにより，ノード間の連結関係が理解しやすくなると言える.

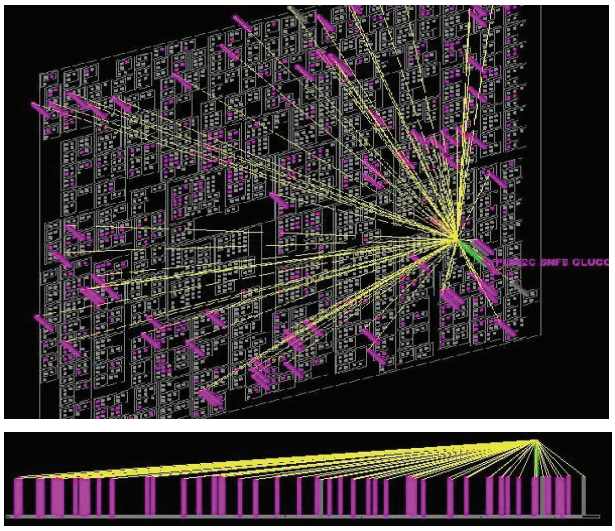


図7. 注視ノードを１段階引き上げた表示画像

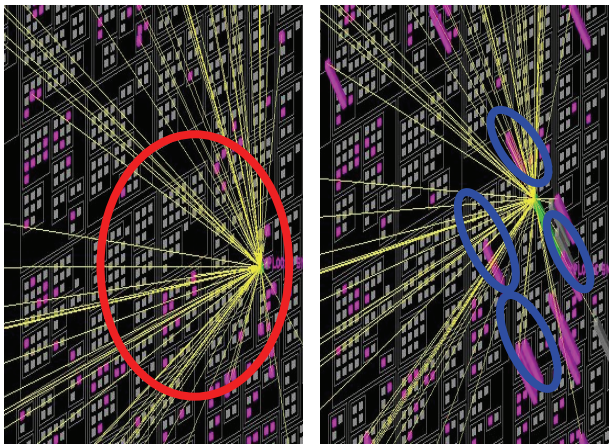


図 8. （左）注視ノードをひきあげてない結果画像
（右）注視ノードを引き上げた結果画像

4.2 結果画像の考察

4.2.1 遺伝子情報分析の観点からの考察

我々は提案手法により得られた結果画像を，遺伝子情報分析を専攻する研究者に提示し，結果画像が遺伝子ネットワークを効果的に表現できているか尋ねた．

4.1 節にて結果画像を示したように，提案手法ではクラスタをまたぐエッジを多く持つ遺伝子に色をつけている．結果画像では，この色がついた遺伝子同士が複雑に絡み合いネットワークを構成していることが可視化できている．この可視化結果画像を遺伝子情報分析の研究者に提示したところ，興味深い遺伝子群の絞込みに効果的である，という評価を頂いた．しかし，これらの遺伝子がマルチドメインなのか否かを検証するためには，遺伝子実験にまで遡る必要があるかもしれない，とのことであった．4.1 節に示したように，我々はまだインターネット上に公開された遺伝子情報だけを有し，また遺伝子情報分析を専攻する情報処理の専門家との議論を経ただけであり，遺伝子実験の専門家と共同で研究を進める体制を持っていない．この問題については，今後の課題として検討していきたい．

なお提案手法を用いることで，1 章で示したマルチドメインの特性以外にも，以下のような遺伝子特性も表現できると考えられる．

[オーソログ遺伝子：] 複数の生物種間で存在する遺伝子で，共通の祖先種では同一の遺伝子であり，現在の機能も同一の遺伝子群．複数の生物種の遺伝子情報から構成される階層型遺伝子ネットワークデータにおいて，同じクラスタに複数の生物種の遺伝子が属する場合，その遺伝子群はオーソログ遺伝子の可能性が高いと判断できる．

[パラログ遺伝子：] ある生物種に存在する 2 つの遺伝子が，祖先種では同一の遺伝子であるような遺伝子群．提案手法による可視化技術では，エッジで連結された遺伝子を 3 次元的に表現することから，非常に強い相互関係を持つパラログの理解にも向いていると考えられる．

また，現在ノードの色を，クラスタをまたぐエッジを多く持つノードに特定する為に用いている．しかしノードの色を使用するのについては，事例ごとに指定する事も可能である．例えばオーソログ遺伝子の事例の場合，複数の生物種それぞれに色をつける等で，より特性を発見することが容易になると考えられる．

4.2.2 可視化技術に対する主観評価

続いて，提案手法による可視化技術が遺伝子ネットワークを効果的に可視化できているか否かについて，可視化に知識のある被験者 11 人よりアンケートを採取することで検証した．アンケートでの質問項目は，以下の 2 点である．

項目 1：クラスタをまたぐエッジを多く持つ遺伝子に色をつけた結果(図 9(左))と，色をつけない結果(図 9(右))を比較し，探索対象となる遺伝子の絞り込みやすさを採点する．

項目 2：以下の 4 つの結果画像を比較し，注目したい任意の遺伝子，およびその遺伝子と相関性の高い遺伝子の分布を理解しやすいか否かを採点する．

- どのノードも引き上げなかった結果 (図 10(左上))
- 注目ノードだけを引き上げた結果 (図 10(右上))
- 注目ノードと連結ノードを，同じ高さまで引き上げた結果 (図 10(左下))
- 注目ノードと連結ノードの両方を引き上げ，しかも注目ノードを連結ノードよりも高く引き上げた結果

(図 10(右下))

また項目 2 においては, 上記 4 つの結果画像それぞれに対し, 以下の 4 つの観点においてそれぞれ評価して頂いた.

質問 1 ：注目ノードの把握

質問 2 : エッジにより連結されたノードの把握

質問 3 : 注目ノードや，連結されたノードの把握

質問 4 : 連結されたノードの分布においての把握

以上の計 6 画像について被験者から，1 から 5 までの 5 段階で数値評価を頂いた．この 5 段階評価は，5 が最高点，1 が最低点であるとする.

表 1 : 項目 1 の評価

	評価の平均
図 9(左)	4.2
図 9(右)	1.6

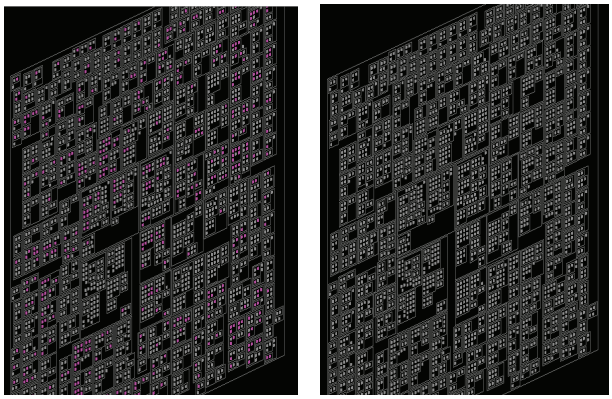


図 9. （左）クラスタをまたぐエッジを多く持つ遺伝子に色を付けた結果. （右）色を付けない結果.

表 2. 項目 2 の評価

	質問 1	質問 2	質問 3	質問 4	総合評価
図 10 (左上)	2.60	3.10	4.00	3.40	3.27
図 10 (右上)	4.50	3.30	3.60	3.40	3.70
図 10 (左下)	4.30	4.40	3.70	3.80	3.97
図 10 (右下)	4.60	4.50	4.20	3.80	4.27

まず項目 1 について検証する．クラスタをまたぐエッジを多く持つ遺伝子に，色を付けた結果(図 9(左))と付けない結果(図 9(右))について，被験者による数値評価の平均値を表 1 に示す．この結果より，遺伝子を表現するノードに色を付けることが，探索対象とする遺伝子を絞り込む目的において有用であることが検証された.

続いて項目 2 について検証する. 図 10 に示した 4 枚の可視化結果について，被験者による数値評価の平均値を表 2 に示す.

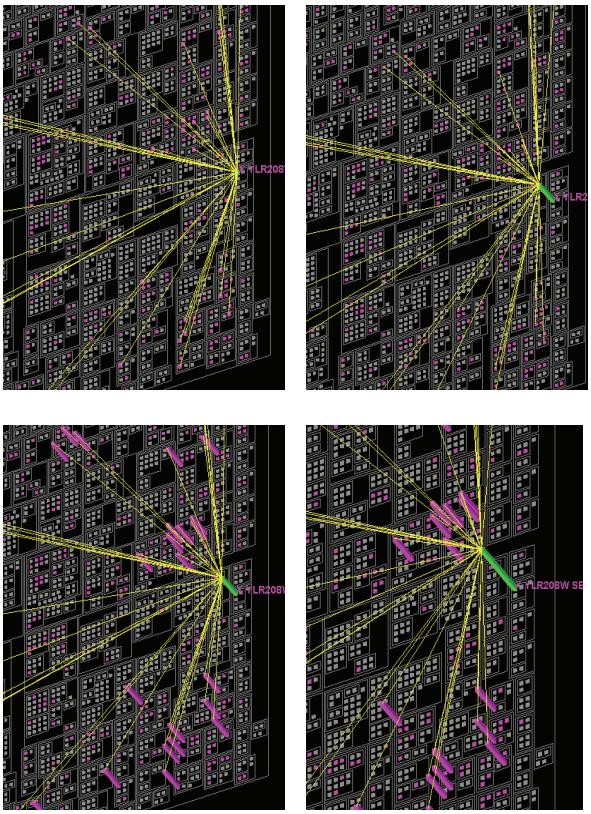


図 10. （左上）どのノードも引き上げなかった結果（左下）注視ノードとエッジで連結されているノードを，同じ高さまで引き上げた結果（右上）注視ノードだけを引き上げた結果（右下）注視ノードが，エッジで連結されているノードより，一段階高く引き上げた結果

表 2 に示す結果について考察する．まず質問 1,2 より，それぞれの結果を比較し，注目ノードを引き上げることの効果が検証できたことがわかった．質問 3 において，各図において評価に大きく差はでなかった．この問題に関しては，エッジを引いていることにより，対象となるノードを確認できることが, 理由と考えられる. 質問 4 においては, 図 1 と図 2, 図 3 と図 4 が，それぞれ同値となった．また図 3 および図 4 の平均値が，図 1 および図 2 に比べ，高いことより，全体分布の把握に関しても，ノードを引き上げることが効果的に作用

していると考察できた．最後に総合評価からも，ノードを引き上げることで，また注目ノードと連結ノードの引き上げ方に差をつけることにより，より効果的に提案手法の結果を示すことができると考えられる．しかし，注目ノードの位置が変わったり，連結ノードがより多くなるにつれて，この数値評価結果が変わりうることも想定される．また現段階では，注目ノードと連結ノードだけをエッジで連結しているが，今後の課題として，連結ノードとエッジで結ばれているノード等も引き上げることによって，さらに遺伝子分析に貢献できないか検討するべきと考えている．

4.3 既存の遺伝子情報可視化ソフトウェアとの比較

マイクロアレイデータから得られる遺伝子発現率情報の可視化ソフトウェアの中の多くは，ノード間の相互関係をエッジで結ぶ古典的なネットワーク 2 次元可視化手法[28]や，TreeView[5]と呼ばれるクラスタリング結果の可視化手法を搭載しており，遺伝子分析に携わる多くの研究者がこれらを利用している．以下，これらの手法に対する提案手法の優位性について論じる．

まず前者の方法では，発現率の相関性の高いノードをエッジで結んで表示することから，遺伝子間の関連性は一目瞭然である．しかし，一画面に表示するノード数は数十～数百程度にとどまっている．またクラスタリング結果を同時に表示してはいない．それに対して提案手法には，

- クラスタ単位で，整然と構造化された形で遺伝子群を表示する．
- 数千，数万といった膨大な量の遺伝子の分布の全貌を，一画面に一括表示できる．

といった点で利点があると考えられる．

続いて後者の TreeView は，N 個の遺伝子に関する M 種類の発現率情報を，N 行 M 列の表形式データとして表現する．この手法は全てのノードの組み合わせに対する相関性を網羅的に表現できる利点がある．しかし，その組み合わせの多くは相関性が低いものであり，必ずしも画面空間を有効に利用した可視化結果を提示しているとは限らない，という問題がある．また，クラスタを単位とした概略的な傾向をつかみにくい，という問題もある．それに対して提案手法には，

- 入れ子構造による階層型データ表示により，遺伝子群をクラスタ単位で概略的に可視化できる．
- 相関性の高い 2 ノード間のみをエッジで表現することにより，相関性の高いノードにのみ注視した可視化を実現できる．

といった点で利点があると考えられる．

5. まとめと今後の課題

本論文では，遺伝子発現率情報に対してクラスタリングとネットワーク化の両方を適用して得られる階層型ネットワークデータの可視化手法を提案した．

提案手法はネットワークとクラスタを同時表示することにより，遺伝子学的に興味深いマルチドメインなどの特性の発見に貢献できると考えられる．また，クラスタをまたぐエッジを多く持つノードに特定の色をつけることにより，興味深い遺伝子の早期発見に貢献できると考えられる．

今後の課題として，以下の点を議論したいと考えている．

- 結果画像から発見された現象が，本当に遺伝子学的に興味深い特性なのか否か，遺伝子実験の専門家を交えての検証．
- オーソログ遺伝子やパラログ遺伝子を含めて，より多くの遺伝子特性を意識した可視化結果の考察．
- 複数の注目ノードを z 軸方向に引き上げた時，あるいは注目ノードだけでなく連結ノードに連結されたノードまで含めて多段階にわたってノードを引き上げた時，の効果的なネットワークの表現手法の確立．
- 有向グラフを構成する遺伝子ネットワークの可視化．
- オントロジーなどの情報を加味した，より遺伝子の研究に貢献できる可視化ソフトウェアとしての開発．
- 各クラスタの画面上の位置の最適化．
- クラスタリングの適切な閾値（図 5 のS_1,S_2に相当する変数値）の発見方法に関する考察．
- 遺伝子ネットワークに限らず，複雑ネットワーク全般に応用できる階層型ネットワーク可視化手法の確立，および遺伝子ネットワーク以外の階層型ネットワークデータでの検証．

謝辞

ソフトウェア Cluster 3.0 の開発者であるコロンビア大学 Michael De Hoon 氏には，クラスタリング技術に関して貴重なご助言を賜ったことを感謝いたします．

遺伝子ネットワークに関する議論に関して，東京大学宮野悟教授，中谷明弘助教授，渋谷哲朗講師，井本清哉助手，お茶の水女子大学瀬々潤准教授から貴重なご意見を賜ったことを感謝いたします．

本研究の一部は，日本学術振興会科学研究費補助金の助成に関するものです．

参考文献

[1] Card s. k., Mackinlay J. D., Shneiderman B., Reading in Information Visualization: Using Vision to Think, Morgan

Kaufmann, ISBN1-55860-533-9, XVII, pp. 686-712, 1998.

[2] 有田, 遺伝子ネットワークと確率モデル Genetic Networks and Probabilistic Models, 2001 年ベイジアンネットチュートリアル, pp. 50-53, 2001.

[3] Itoh T., Takakura H., Sawada A., Koyamada K., Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, IEEE Computer Graphics and Applications, Vol. 26, No. 2, pp. 40-47, 2006.

[4] Saraiya P., North C., Duca K., An Evaluation of Microarray Visualization Tools for Biological Insight, IEEE Information Visualization 2004, pp. 1-8, 2004.

[5] TreeView, http://www.gmod.org/node/91

[6] Holten D., Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, IEEE Information Visualization 2006, pp. 741-748, 2006.

[7] Ghoniem M., Fekete J., Castagiloia P., A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations, IEEE Information Visualization 2004 , pp. 17-24, 2004.

[8] Eades, P., "A Heuristic for Graph Drawing," Congressus Numerantium, Vol. 42, pp. 149-160, 1984.

[9] Sarcar M. , Brown M. H., Graphical Fisheyes Views of Graphs, Communication of the ACM, Vol. 37, pp. 73-83, March 1994.

[10] Huang M. L., et al., WebOFDAV−Navigatingand Visualizing the Web On-Line with Animated Context Swapping, 7th WWW Conf, pp. 636-638, 1998.

[11] North S., Incremental Layout in DynaDAG, Graph Drawing ' 95, pp. 409-418, 1995.

[12] Mukherjea, S., Foley J. and Hudson S., Visualizing Complex Hypermedia Networks through Multiple Hierarchical Views, Proceedings of ACM SIGCHI '95, Denver, Colorado, pp. 331−337, May 1995.

[13] 塩澤, 西山, 松下, 「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, Vol. 38, No. 11, pp. 2331-2342, 1997.

[14] Lamping, J. and Rao, R., "The Hyperbolic Browser: A Focus + Context Technique for Visualizing Large Hierarchies," Journal of Visual Languages and Computing, Vol. 7, No. 1, pp. 33-55, 1996.

[15] Carrire J. and Kazman R., "Research Report: Interacting with Huge Hierarchies: Beyond Cone Trees," Proceedings of the IEEE Conference on Information Visualization '95, IEEE CS Press, pp. 74-81, 1995.

[16] Johnson B., et al., Tree-Maps: A Space-Filing Approach to the Visualization of Hierarchical Information Space, IEEE Visualization ' 91, pp. 275-282, 1991.

[17] Eades P., et al., Multilevel Visualization of Clustered Graphs, Graph Drawing ' 96, pp. 101-112, 1996.

[18] Schaffer D., et al., Navigating Hierarchically Clustered Networks through Fisheye and Full-Zoom Methods, ACM Trans. Computer-Human Interaction, Vol. 3, No. 2, pp. 162-188, 1996.

[19] 我妻, 藤代, 堀井, 階層的因果関係の対話的可視化, 第 10 回ビジュアリゼーションカンファレンス, 2004.

[20] Fekete J.-D., Wang D., Dang N., Plaisant C., Overlaying Graph Links on Treemaps, IEEE Information Visualization 2003 Poster Conpendium, pp. 82-83, 2003.

[21] Bederson B., Schneiderman B., Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, ACM Transactions on Graphics, Vol. 21, No. 4, pp. 833-854, 2002.

[22] Itoh T., Yamaguchi Y., Ikehata Y., Kajinaga Y., Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm, IEEE Transactions on Visualization and Computer Graphics, Vol. 10, No. 3, pp. 302-313, 2004.

[23] 伊藤, 山口, 小山田, 長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善, 可視化情報学会論文集, Vol. 26, No. 6, pp. 51-61, 2006.

[24] De Hoon, M.J.L., Imoto, S. Kobayashi, K., Ogasawara, N. & Miyano, S., Inferring gene regulatory networks from time-ordered gene expression data of Bacillus subtilis using differential equations, Pac. Symp. Biocomput., 8, pp. 17-28, 2003.

[25] Savoie, C.J. Aburatani, S. Watanabe, S. Eguchi, Y. Muta, S. Miyano, S., Imoto, S., Kuhara, S. & Tashiro, K., Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades, DNA Research, No.10, pp.19-25, 2003.

[26] Imoto, S., Savoie, C.J., Aburatani, S., Kim, S., Tashiro, K., Kuhara, S. & Miyano, S., Use of gene networks for identifying and validating drug targets, J. Bioinform, Comput. Biol., No.1, pp. 459-474, 2003.

[27] Open Source Clustering Software (Cluster 3.0), http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/

[28] Open Source gene network Software (IPA), http://www.digital-biology.co.jp/japanese/ingenuity/index.html

[29] 西山, 伊藤, 「平安京ビュー」を用いた階層型遺伝子ネットワークの可視化, 第 22 回 NICOGRAPH 論文コンテスト, 2006.

著者紹介

西山　慧子

2006 年お茶の水女子大学理学部情報科学科卒業．現在お茶の水女子大学大学院人間文化研究科数理・情報科学専攻在学中．情報処理学会会員．

伊藤 貴之

1990 年早稲田大学理工学部電子通信学科卒業. 1992 年早稲田大学大学院理工学研究科電気工学専攻修士課程修了．同年日本アイ・ビー・エム(株)入社．1997 年博士(工学)．2000 年米国カーネギーメロン大学客員研究員．2003 年から 2005 年まで京都大学大学院情報学研究科 COE 研究員(客員助教授相当)．2005 年日本アイ・ビー・エム(株)退職，2005 年よりお茶の水女子大学理学部情報科学科助教授．ACM, IEEE Computer Society, 情報処理学会，芸術科学会，画像電子学会，他会員．