



## エッジの束化を用いた遺伝子ネットワークの可視化\*

中澤 里奈<sup>1)</sup>, 伊藤 貴之<sup>2)</sup>, 瀬々 潤<sup>3)</sup>, 寺田 愛花<sup>3)</sup>

### A Visualization Technique for Gene Networks with Edge Bundling

Rina Nakazawa, Takayuki Itoh, Jun Sese and Aika Terada

#### ABSTRACT

Gene networks have been constructed with genes as nodes and interactions between genes as edges. Because edges whose both ends have different functions may indicate important functional connections, visualization of the functional connection in the network helps us to understand cellular mechanisms. However, very numerous nodes and edges in the gene networks prevent the exploration of the relations from the visualization, if it is visualized by traditional techniques. In this paper, we introduce a simultaneous visualization of a large gene network and gene functions provided by gene ontology (GO) consortium. By bundling edges between groups whose genes have similar function, complication connections and many crossing edges between groups are represented as essential thick edges, resulting easily understanding of connections between gene functions.

**Keywords:** Gene Network, Gene Ontology, Edge Bundling,

#### 1. はじめに

近年、様々な遺伝子関連情報が解読され、明らかになってきたが、未だに遺伝子の機能や他の遺伝子との関係性がわからないものも多く存在している。その解明のための情報共有手段として、以下がよく用いられている。

**GO(Gene Ontology)**<sup>2)</sup>: 遺伝子の機能や概念を共通語で記述するための用語体系およびその策定プロジェクト。機能が解明された遺伝子には GO が規定する 1 個以上の用語 (GO ターム) が割り当てられている。

**遺伝子ネットワーク**: 遺伝子をノード、2 遺伝子間の関係をエッジとする構造。

ネットワークは情報可視化の主要な対象のひとつであり、遺伝子ネットワークの可視化もこれまでに多く試みられている。我々は以下を目標として遺伝子ネットワーク可視化の新しい手法を研究開発している。遺伝子群を機能でクラスタリングした際に、多数の相互作用が見られる密接な関係のクラスタを発見する。それとは逆に、クラスタ内のごく少数の遺伝子に見られる「外れ値」のような相互作用を発見する。

この目標を満たすための手法として本論文では、GO で定義された遺伝子機能の属性をノードに付与した遺伝子ネットワークを可視化する一手法を提案する。この GO によって定義された機能情報を色に対応づけて、遺伝子ネットワーク上にて機能情報と接続構造を同時に可視化することで、未知の遺伝子機能や関係性を推測しやすくする。具体的には、多数の相互作用で密接に接続されたクラスタを観察することで、当該クラスタに属する遺伝子について未発見の相互作用を予測できる場合があると考えられる。逆に外れ値のような相互作用を発見することで、遺伝子の機能や相互作用に関する新たな知見や誤認識の発見につながると考えられる。このような可視化の工程で得られる知見は、生物学の研究における実験計画の支援や、医学分野の疾患原因究明に繋がると考えられる。

前述の通り、遺伝子ネットワークの可視化もこれまでに多く試みられているが、既存の遺伝子ネットワーク可視化手法には主に以下のような問題点が残っている。

**[問題 1]** ノード数やエッジ数が非常に多いため、その中でどこが注目すべきか一目で理解するのが難しい。これを解決するためにはネットワークの構成を明快に表示することが重要である。

**[問題 2]** エッジの接続構造が複雑であるため、ノードやエッジの画面上での重なりが多く生じる。これが機能と相互作用の視認性を妨げることがある。

前者の問題を解決するために本手法では、可視化手法 FRUITS Net<sup>1)</sup>を画面配置アルゴリズムとして適用するこ

\* 原稿受付 2012年12月11日

1) 非会員 お茶の水女子大学大学院 人間文化創成科学研究科(〒112-8610 東京都文京区大塚 2-1-1, E-mail : leena@itolab.is.ocha.ac.jp)

2) 正会員 お茶の水女子大学大学院 人間文化創成科学研究科

3) 非会員 東京工業大学大学院 情報理工学研究科

とで、エッジで連結された遺伝子クラスタ、あるいは共通の機能を有する遺伝子クラスタを、画面上でできるだけ近くに配置する。これによってネットワークの全体的構成を明快に表現する。ネットワークのエッジには有向エッジと無向エッジがあるが、本論文では無向エッジのネットワークを対象とする。

後者の問題を解決するために本手法では、エッジの束化に関する新しい手法を提案する。本手法では同一の2クラスタ間に接続されたエッジが一定本数以上を有する場合に、これを束におきかえる。このとき各々の束を、ノードとの重なりを回避するような折れ線で表現する。これによってノードやエッジの画面上での重なりを低減し、機能と相互作用の関係性に関する視認性を向上する。

## 2. 関連研究

### 2.1 遺伝子ネットワークの可視化手法

遺伝子ネットワークの可視化手法は既に多数発表されている。西山ら<sup>3)</sup>は、大規模階層型データ可視化手法「平安京ビュー」を用いて遺伝子発現量と遺伝子ネットワークを可視化している。伊澤ら<sup>4)</sup>は、遺伝子の機能情報に着目し、その情報毎にユーザが位置を決め、遺伝子ネットワークの再配置を行う可視化手法を提案している。Breitkreut<sup>5)</sup>らは、遺伝子の機能情報と遺伝子ネットワークを、ユーザがネットワークのレイアウトを選択肢の中から選択可能なかたちで可視化している。しかし、どの手法もある程度大規模なデータになってきた時のノードやエッジの重なりを回避する処理にはあまり着目していない。また、Dinkla<sup>6)</sup>らはマトリックスを用いて遺伝子制御ネットワークを可視化している。この手法は部分ネットワークも表現できている、といった利点が挙げられる一方で、ネットワークの規模が大きくなると横または縦に広がってしまうため、画面占有面積を抑えることをあまり考慮していないという問題点も存在している。

### 2.2 FRUITS Net

伊藤らはFRUITS (Framework and User Interface for Tangled Segments) Net<sup>1)</sup>というネットワーク可視化手法を提案し、遺伝子ネットワークの可視化に適用している。この手法は、ノードに付与された1個以上のアイテム情報を色で表現し、同一アイテムを共有するノードをできるだけ近くに配置する。これにより、ノードに与えられたアイテム情報とネットワーク構造の関係性を可視化することができるので、本報告の目的に向いていると考えられる。Fig. 1にその可視化の一例を示す。文献<sup>1)</sup>では遺伝子の発現量とネットワーク構造の可視化を試みているが、発現量は実験などから得た数値情報であり、誤差などが反映されていることもある。さらに、GOタームも論文をもとに人の手によってつけられているため、遺伝子につけられているGOタームが適切なものであるかを検証する目的でも、

発現量ではなく、GOタームを用いることは有用なのではないかと考えられる。また、GOタームはキュレーターと呼ばれる専門研究者によって付与されており、その労力削減、及び人為的なミスを防ぐためにも有用だと考えられる。そこで、本研究では発現量の代わりにGOによって定義される機能情報を可視化するものである。

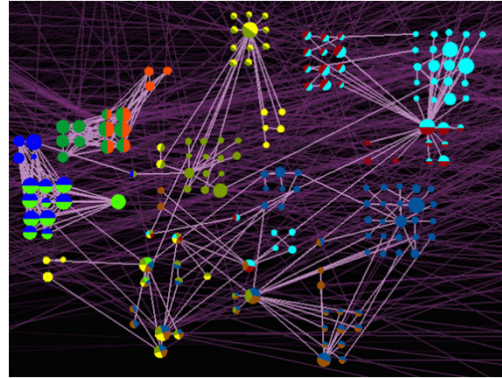


Fig. 1 FRUITS Net の可視化例

### 2.3 エッジの束化

エッジの束化は階層型グラフを対象としたHoltenの手法<sup>7)</sup>で脚光を浴びるようになり、階層型データ可視化手法TreemapをベースにしたMuelderらの手法<sup>8)</sup>などにも同じ考えが継承された。しかしこれらの手法はノードの上を束が通過することがしばしばあり、これが接続構造の可読性を低下させる場合があった。一方でエッジの最適な束化のためにエネルギーや力学の概念を導入した手法<sup>9) 10) 11)</sup>が多数発表されたが、これらの手法には計算時間が常に問題となる。例えばGansner<sup>9)</sup>らの手法は地理情報に基づくものであり、各ノードの位置は緯度と経度によって固定されている。このことから、束の生成時には元のエッジ位置からの移動量を小さくする必要があり、この制約を満たす束化の実現のために大きな計算時間を要する。いずれの手法にせよ、既存のエッジの束化手法では、膨大な数のエッジの主要経路を効果的に描画しているが、ノードと束との重なり回避処理を適用していないという問題点がある。

## 3. 提案手法

### 3.1 問題点

本手法ではFRUITS Net<sup>1)</sup>を用い、遺伝子をノード、遺伝子間の相互作用をエッジとして遺伝子ネットワークを可視化する。ここでGOタームをアイテム情報としてノードに付与し、ノードを色付けする。本手法で発現量の代わりにGOタームを付与する理由は前述のとおり、発現量は実験から得た数値であるので誤差などが反映される可能性もあるからである。さらに、既に遺伝子につけられているGOタームの適切さを検証する目的としても、GOタームをアイテム情報として付与するのは有用ではないかと

考えられる。

FRUITS Net を用いて遺伝子ネットワークと GO タームを可視化した一例が Fig. 2 である。使用したデータは、4.1 節にて後述するデータと同一である。Fig. 2 では遺伝子をノードとし、遺伝子間の相互作用をエッジとし、無作為に選んだ 10 種類の GO のタームでノードを色付けた。この可視化結果の主要な問題として、以下の二点が挙げられる。

- (1) 人が区別できる色の数には限りがある。そのため使用データにおける GO タームが数百種類あるにもかかわらず、たかだか 10 種類程度の GO タームしか表現できない。
- (2) エッジとノードの数が非常に多いために、可視化結果が非常に煩雑である。

これらの問題点を解決するために本手法では以下のアプローチをとる。まず(1)を解決するために本手法では、意味の近い GO ターム群をクラスタ (以後 GO タームクラスタと称する) にまとめて、GO タームクラスタを単位としてノードを色付けする。GO タームクラスタを生成するために本手法では、3.2 節で述べる通り、用語体系の DAG (Directed Acyclic Graph: 無閉路有向グラフ) 構造に対してクラスタリングを適用する。DAG 構造であるため、GO タームには 1 つ以上の親の GO タームを持つものが存在する。続いて(2)を解決するために、本手法では FRUITS Net の画面配置アルゴリズムを適用し、さらにエッジの束化を適用する。本手法では 3.3 節で述べる通り、まず GO タームクラスタに基づいて遺伝子群のクラスタ (以後遺伝子クラスタと称する) を生成し、遺伝子クラスタの画面位置を決定する。このとき本手法では、多数のエッジで連結された遺伝子クラスタ、あるいは共通の GO タームクラスタを有する遺伝子クラスタが、画面上で近くに配置されるように画面配置を決定する。続いて 3.4 節で述べる通り、同一の遺伝子クラスタ間を連結する多数のエッジを束化することで、エッジの可読性を向上する。

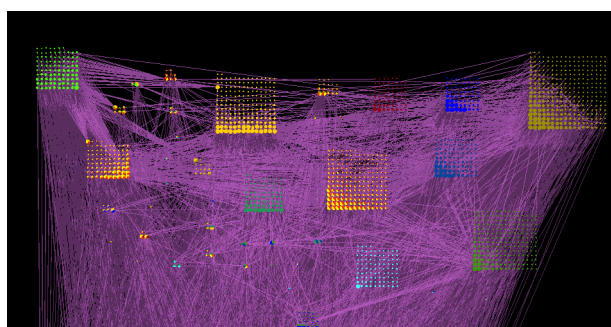


Fig. 2 遺伝子ネットワークの可視化に関する問題点

### 3.2 GO クラスタリング

本手法では、より多くの GO タームから得られる知見を可視化結果に反映するために、もともと DAG 構造をとっている GO タームを、データ構造上の距離をもとにクラ

スタリングする。現時点での実装では、以下のようにして GO タームのクラスタを生成している。

まず GO から 1 個の GO タームを選び、それと直接接続される子の GO タームを全て抽出する。これらの GO タームが割り当てられている遺伝子数の合計が一定値以上であるものを、1 個のクラスタとする。

このようにして生成されたクラスタ群を、最上位の親ノードによってグループ化し、1 個のグループに属するクラスタ群のうち 10~15 種類の GO タームクラスタをアイテム情報としてノードに付与し、可視化に用いている。以上により、ユーザが識別できる程度の色数で GO 情報を表現する。

### 3.3 遺伝子クラスタリングと画面配置

さらに、FRUITS Net を使用する前処理段階として、遺伝子にクラスタリングを適用する。まず各遺伝子に割り当てられた GO タームクラスタの組み合わせによって遺伝子群を分類し、さらに各分類に対してエッジの密度にもとづいて遺伝子群をクラスタリングすることで、遺伝子クラスタを形成する。現時点での実装では、以下のようにして遺伝子のクラスタを生成している。

まず各遺伝子に割り当てられた GO ターム群を参照して、それらが属する GO タームクラスタを各遺伝子に割り当てる。続いて、各遺伝子を GO タームクラスタの組み合わせで分類し、全く同一の GO のタームクラスタ群を有する遺伝子群をまとめて 1 個のクラスタとする。さらに必要に応じて、エッジの密度にもとづいて各クラスタ内の遺伝子群をさらに複数のクラスタに分類する。我々の実装では、エッジの密度に基づくクラスタリングに Fast Modularity Community Structure Inference Algorithm<sup>12)</sup>を採用している。

そして、クラスタを単位として FRUITS Net の配置アルゴリズムを用いることで、遺伝子に対応する各ノードの配置を決定する。FRUITS Net の配置アルゴリズムは力学モデルと空間充填モデルの 2 ステップで構成される。まず遺伝子クラスタをノードとするグラフ構造を形成し、これに力学モデルの反復計算を適用することで、遺伝子クラスタの仮の位置を決定する。これによって、エッジで連結された遺伝子クラスタや、同一の GO タームクラスタを共有する遺伝子クラスタを、画面上で近くに配置する。続いて、この仮の配置結果を参照しながら空間充填モデルを適用することで、遺伝子クラスタの画面上での重なりを回避しつつ、配置画面占有面積を減らす。

### 3.4 エッジの束化

続いて、エッジの視認性を向上するために束化を施す。本研究では遺伝子の接続関係にかかる位相的な情報だけが重要であり、ノード配置結果の幾何的な情報は重要ではない。そのことに注目して本手法では、既存のエッジ束化手法よりも簡易で高速な手法を採用する。

エッジの束化の手順としては、まず一つの遺伝子クラスタに注目し、他の遺伝子クラスタとの間に接続しているエッジの本数を集計する。遺伝子クラスタ間のエッジが所定の本数以上ならば、クラスタ間に太線を束として描き、接続していたエッジとその束を結ぶ。さもなければ、束化せずにエッジをそのまま描く。以上の処理を全ての遺伝子クラスタペアに対して実施する。ここでクラスタ間のエッジの本数  $k$  の関数 (式(1)) を用いて描画する束の太さ  $W$  を調整する。我々の実装では  $a=0.7$ ,  $b=2.5$  とし、 $W$  の単位を画素数として扱う。

$$W = a \log k + b \quad (1)$$

この時、束がノードを横切ってしまうと視認性がさがるので、Fig. 3に見られるように束を線分ではなく折れ線で描画する。まず 2 つの遺伝子クラスタを接続する束を、Fig. 3(a)に示すように 1 本の線分として生成する。これが画面上で遺伝子クラスタと交差するようであれば、遺伝子クラスタを表現する長方形内部にある 1 点を長方形外部に移動し、束をその 1 点を通過する折れ線に変換する。同じ処理を再帰的に進めることで、Fig. 3(b)に示すように遺伝子クラスタを回りこむ折れ線を生成する。以上の処理によって、束と遺伝子クラスタとの重なりを回避する。

しかしこの方法を適用しただけでは、遺伝子クラスタを表現する長方形の稜線に沿って、束どうしが一直線上に重なってしまい、束の接続関係が視覚的にわからなくなる場合がある。そこで束どうしが重なる場合には、Fig. 3(c)に示すように束の位置をずらすことで、この問題を回避する。しかし遺伝子クラスタが密集している部位では、束の重なりを回避することが不可能な場合がある。このような場合にのみ、束が遺伝子クラスタと交差するのはやむを得ないとして、遺伝子クラスタ間を線分で結ぶことで描画している。このエッジの束化の切り換えやクラスタ間のエッジの本数の閾値  $k$  に関しては次節に述べる GUI 部品を用いることでユーザが対話的に制御することが可能である。

### 3.5 GUI 機能

我々による本手法の実装では、1 個のウィンドウを描画

領域と操作パネルに分割しており、操作パネル上に各種の GUI 部品が搭載されている。操作パネルには可視化結果の拡大縮小や平行移動の切り替えボタン、ノード強調表示を適用するための GO ターム選択ボタン、エッジ束化の有効化ボタンが搭載されている。また、束化を適用するエッジの最低本数を設定するスライダや、束化されなかったエッジを控えめに描画するための透過率を設定するスライダが搭載されている。さらに、描画領域上でノードをクリックすると、対応する遺伝子の ID を表示し、その遺伝子に割り当てられた GO タームクラスタに関する詳細情報を表示する。それと同時に、クリックされたノードに接続するエッジをハイライトすることで、当該ノードの詳細な接続関係を提示する。このエッジのハイライトは 2 つのノードについて同時に適用可能であり、これによって両ノードそれぞれの接続関係を比較できる。

## 4. 実行例

### 4.1 使用データ

我々は iRefIndex<sup>13)</sup>に公開されているショウジョウバエの遺伝子ネットワークを適用し、各遺伝子をノード、タンパク質間相互作用をエッジとし、GO によって定義される遺伝子機能情報をノードに付与して可視化を試みた。iRefIndex では、異なる論文や異なる実験で報告されているエッジもそれぞれの論文、あるいは実験毎にエッジを保存している。このため、同じエッジが複数回現れることがあるが、今回はこれらの重複を除いて 1 本のエッジとして扱っている。また、ノードに相当するタンパク質情報として NCBI の Entrez Gene<sup>14)</sup>データベースを用いているが、iRefIndex のノード全てが Entrez Gene と対応がとれているわけではないため、エッジの両端のノードが Entrez Gene と対応できるもののみを使用した。ノード数は 8945、エッジ数は 32703、既知の GO ターム数は 259、今回アイテム情報としてノードに付与した GO タームクラスタは 12 個であった。

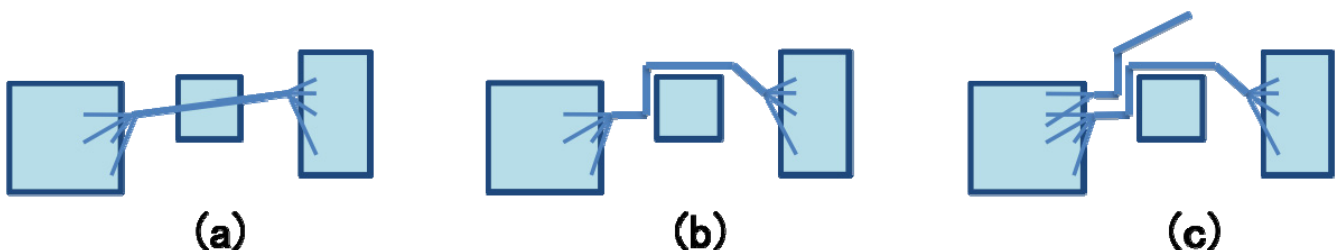


Fig. 3 エッジの束化手順 (a)線分生成 (b)折れ線生成 (c)重なり回避

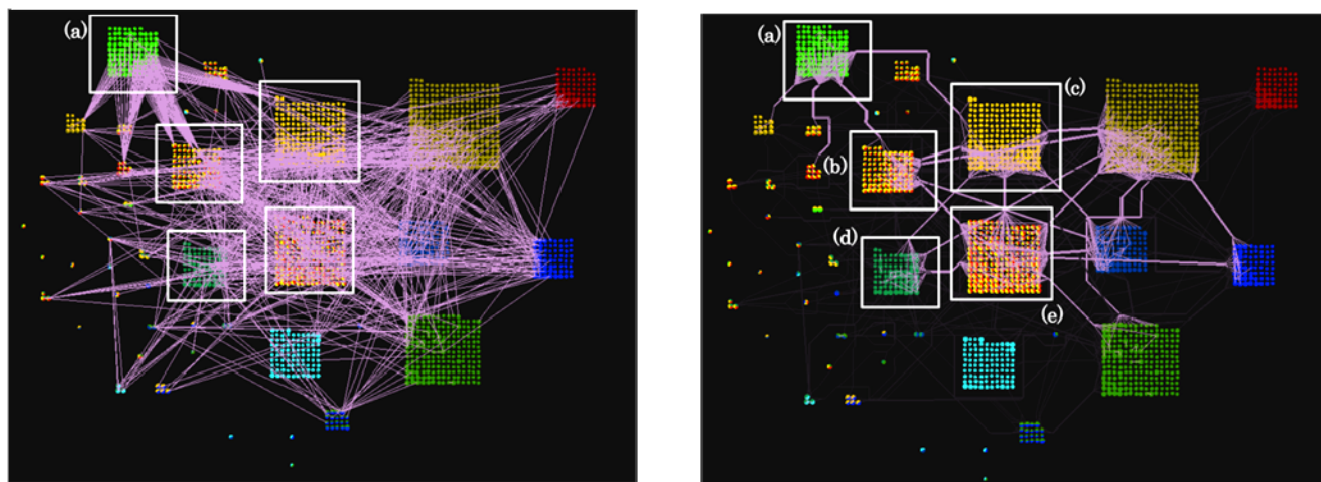


Fig. 4 可視化結果 (左) エッジの束化前 (右) エッジの束化後

## 4.2 実行結果

エッジ束化処理を施す前の可視化結果を Fig. 4(左)に、エッジ束化処理を施した後の可視化結果を Fig. 4(右)に示す。Fig. 4(左)の左上部では、例えば(a) と他機能との相互作用などは比較的理解しやすいが、中央部から右部にかけてエッジの密度が高く、遺伝子間の関係性どころか機能間の主要な関係性さえも読み取りづらい。一方で Fig. 4(右)では図の中央部から右部にかけての機能間の主要な関係性が、Fig. 4(左)と比べて視認しやすくなっていることが確認できる。ここでは(a)と(b)、(a)と(c)、(b)と(e)、(d)と(e)の各遺伝子クラスタ間において、より多くの相互作用が見られることがわかった。

さらに、この可視化結果の左上部の二つのクラスタの関係性に着目する。Fig. 5(左)がエッジの束化適用前、Fig. 5(右)がエッジの束化適用後の図である。Fig. 5(左)を見ると、二つの遺伝子クラスタ間の接続にも AB1 と AB2 の二種類の関係性があることがわかる。一つの関係性は遺伝子クラスタ A と遺伝子クラスタ B のノードが一对多数対応であるのに対し、もう一つの関係性がどのようなになっているかまでは Fig. 5(左)からは読み取れない。一方で、Fig. 5(右)にて同じクラスタ間の接続に着目する。Fig. 5(左)では二種類あった関係性が一つにまとめられてしまっているが、束に接続するエッジから遺伝子クラスタ B と遺伝子クラスタ A のノードが少数対多数対応であることがわかる。ここでエッジの束化前の Fig. 5(左)から AB1 の接続関係を見ると、遺伝子クラスタ B の内左上にあるノードが遺伝子クラスタ A のノードに対して一对多数対応だったことを考慮すると、AB2 における他のノードも同様に一对多数対応関係にあるのではないかと考えられる。そこで、これらの注目すべきノードをクリックし、エッジをハイライトさせると、実際に遺伝子クラスタ A

と接続している遺伝子クラスタ B のノードは、遺伝子クラスタ A と一对多数対応であるということが確認できる (Fig. 6)。これらの遺伝子クラスタ B 内部の右下にあるノードは、遺伝子クラスタ A 以外のノードとも接続している一方で、遺伝子クラスタ B 内部の左上のノードは遺伝子クラスタ A のノードとの間にしかエッジが接続されていない。このことをふまえると、この左上のノードも同様の関係性を持っているか、右下のノードとは異なる GO タームを持っているか、これらの遺伝子クラスタ B の遺伝子が今回ノードの色に割り当てなかった共通の GO タームを他に持っていて、それがこの共通した関係性に影響を与えているのではないかと推測ができるのではないかと考えられる。

今回の実行結果を従来手法と比較する。2.3 節でも述べたとおり、従来手法の多くはエッジの束とノードとの重なりを回避していない。そのためエッジの束が重なって描画されてしまったノードの関係性が視認しにくいことが想像される。本手法ではノードの重なりを回避しながらエッジの束を描くため、そのような問題点は生じにくいと考えられる。一方で、本章で紹介した結果から改善すべき点も見られた。それはノードとの重なりを回避するためにエッジの束を折れ線で表現していることに起因する。Fig. 5(右)では、左上の緑色のクラスタ A に対して、AB1 の左のノード 6 個のクラスタ、B の左のノード 8 個のクラスタ、の各々とバンドリングされたエッジを持っている。しかし現時点での結果では両者が重なってしまい、束が 2 本あることを見落としてしまう危険性がある。既存手法の中には束を曲線表示する手法も多くあり、そのほうが、束が完全に重なる可能性は低い。よって今後の課題として、既存手法と同様に束を曲線化することが考えられる。

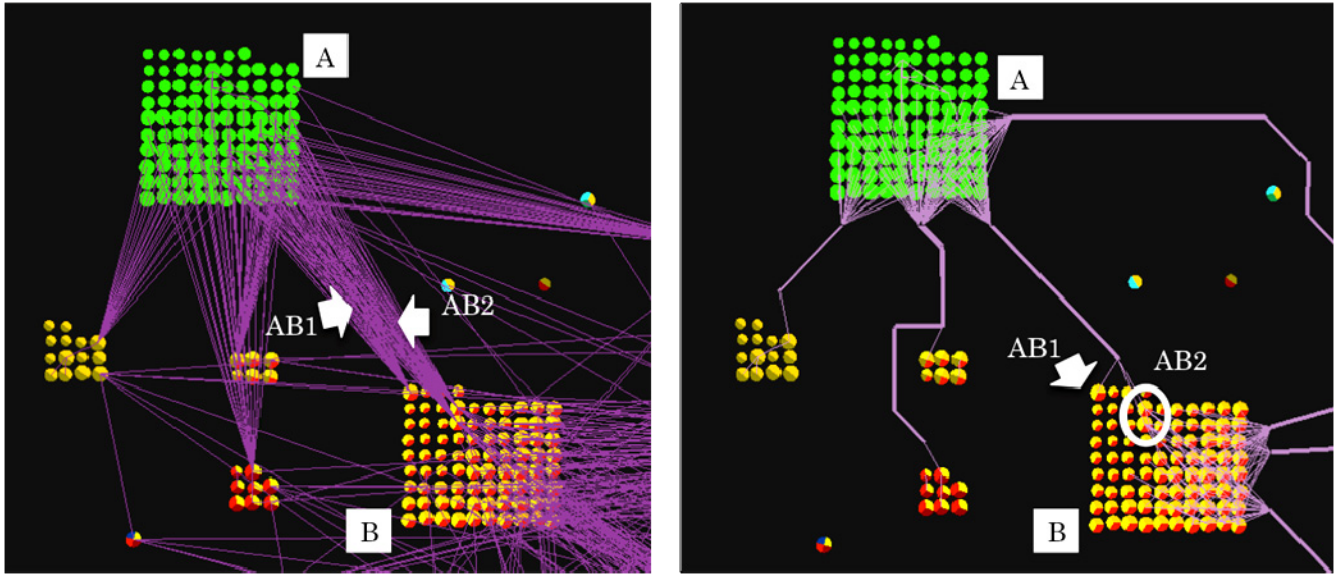


Fig. 5 可視化結果詳細図 (左) エッジの束化前 (右) エッジの束化後

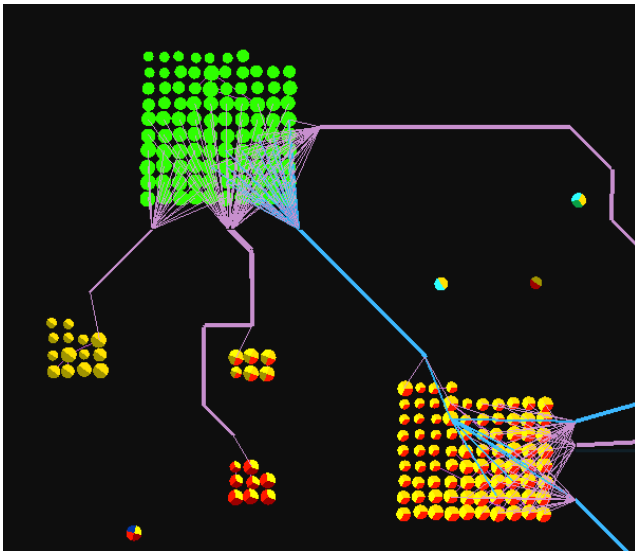


Fig. 6 ノードをクリック時

また、本手法を適用して得られた別の可視化結果 Fig.7 において生物学的知見との一致が見られるかどうか考察を行った。Fig.7 (左) ではノードとエッジの重なりが著しいが、Fig.7 (右) では(a)が他の機能に対してハブの役割を果たしていることや、(c)内部における関係性が見てとれる。なお(a)~(h)に割り当てた機能情報は以下の通りである。

- (a) regulation of biological quality
- (b) determination of adult life span
- (c) cell cycle process
- (d) embryonic development
- (e) cell proliferation
- (f) cell division
- (g) cell adhesion
- (h) cell death

(a)は制御因子であるので、(a)がハブであるという可視

化結果は、(a)に所属する遺伝子が(b)~(h)に所属する遺伝子に対して何らかの制御を行っているために関係性が多く見られるという生物学的知見に一致する。

このように、エッジの束化を適用することで、新たな機能や関係性を確認できそうな特異点を見つけることができ、その上エッジの束化を適用していなかった時にはエッジが煩雑で見えなかった同じ遺伝子クラスタ内のノード同士の接続性を把握することができた。従って、エッジの束化を適用することで視認性が向上し、より効果的に遺伝子ネットワークの関係性を把握し、機能の推測ができるようになるのではないかと考えられる。

## 5. まとめと今後の課題

本論文では、エッジの束化を施したネットワーク可視化手法 FRUITS Net を用いて、遺伝子ネットワークと Gene Ontology を統合可視化する手法を提案し、その実行例を示した。本手法では GO クラスタリングによって得られた GO タームクラスタを遺伝子に付与した上で、遺伝子をクラスタリングする。そして、力学モデルと空間充填モデルを併用した FRUITS Net の画面配置アルゴリズムにより、ノードの画面上の位置を決定する。力学モデルの採用によって、エッジで接続されたノードや同一の GO タームクラスタを共有するノードが画面上で近くに配置されるようにノードを配置し、さらに空間充填モデルの採用によって画面占有面積の低減を図る。さらに、エッジの本数が膨大であり、かつエッジとノードの重なりが多い、ということに起因する視認性の問題を解決するために、エッジの束化によって視認性を向上する。描画する束はできる限りノードと重ならないように、さらに束どうしが重ならないように折れ線化される。

本手法を iRefIndex に公開されているショウジョウバ

エの遺伝子ネットワークに適用し、実行結果として、Fig. 4~7の結果画像を得た。実行結果から、各ノードに遺伝子機能情報を付与し、重なりを回避しつつ線情報を要約したことによって、遺伝子機能間の関係性をより把握しやすくなったことが確認できた。さらに、新たな機能や関係性を確認できそうな特異点を見つけることができた。また、今回の実行結果から、機能間の繋がりに関して既に明らかになっている生物学的な知見との一致が見られた。この知見の一致から、本手法は例えば、たくさんの知見を一画面に全部示すことで利用者間での同意を手早く得たい場面にも有効であると言える。

今後の課題としては、以下のようなことを考えている。まず我々自身が本手法を使用して見た感想として、以下のような機能拡張や改良が望ましいと考えている。

- 束化時に別々に束ねたものの、束の辿る経路が同じ故に束同士の重なりが生じることがある。これを軽減するため束の曲線化を行い、視認性の向上をはかる。
- アイテム（本論文ではGOタームで定義された遺伝子の機能）の表現に色を用いているが、同系色の組み合わせのアイテムは識別が難しいという問題があるため、色を用いたアイテム表現を改善する。
- 現状の遺伝子データには機能情報の欠落が見られるが、それゆえに機能情報が既知である遺伝子と欠落している遺伝子との比較が重要な場合がある。そこで機能情報が既知である遺伝子と欠落している遺伝子との関係性を強調表示する機能等を追加する。
- 現時点の実装ではクラスタ内部の各ノードを格子状に配置しているが、クラスタ内部のレイアウトにも力学モデルを適用することでクラスタ内部の視認性を向上できるか検討する。
- 現在の可視化結果にアンチエイリアシングなどを施して結果画像を改善する。

また、本論文の実行結果では比較的上位層のGOタームを中心にノードに割り当てたことで、遺伝子ネットワーク上の機能情報に関する概略的な知見を可視化結果から確認できた。一方で、遺伝子間のより詳細な差異を観察するためには、もっと下位層に属するGOタームを適用しての可視化が重要になる。そこで今後の課題として、まず上位階層のGOタームで全体像を描き、続いてユーザが選択した局所について下位層のGOタームを適用することで詳細な機能情報を可視化する、といったような対話的かつ段階的な操作も含めた可視化手法を考えていきたい。

これらの機能拡張を試みた上で、以下の観点から、本手法による可視化結果についてさらに検証を進めたい。

- GOタームおよび遺伝子に関する現時点でのクラスタリング結果の妥当性を検証し、問題がある点があれば改善する。
- 可視化結果から遺伝子学的に興味深い関係性や機能が見いだせるか、遺伝学の専門家を交えてさらなる検証を続ける。  
そして最終的には、本手法の実用性を高めるために、以下のような点に着手したい。
- 生物学の研究でよく用いられるCytoscape<sup>15)</sup>といった既存のツールとの連携、もしくは既存のプラットフォーム上でのアドオンやプラグインとしての実装を検討する。
- 遺伝子ネットワークに限らず、属性情報を有する大規模なネットワークへの本手法を適用し、その有効性を検証する。

## 謝 辞

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) T. Itoh, C. Muelder, K. Ma, J. Sese: A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs, IEEE Pacific Visualization Symposium, 121-128, 2009.
- 2) D. Botstein, J. M. Cherry, M. Ashburner, C. A. Ball, J. A. Blake, H. Butler, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig: Gene Ontology: tool for the unification of biology. Nat Genet 25: 25-29, 2000.
- 3) 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, 芸術科学会論文誌, 6(3), 106-116, 2007.
- 4) 伊澤, 瀬々: 機能間の関係を明示する遺伝子ネットワークの可視化, DEIM Forum, E4-5, 2010.
- 5) B. Breitkreutz, C. Stark, and M. Tyers: Osprey: a network visualization system, Genome Biology, 2003.
- 6) K. Dinkla, M. A. Westenberg, and Jarke J. van Wijk: Compressed Adjacency Matrices: Untangling Gene Regulatory Networks, IEEE Transactions on Visualization and Computer Graphics, 18(12), 2457-2466, 2012.
- 7) D. Holten: Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, IEEE Transactions On Visualization And Computer Graphics, 12(5), 741-748, 2006.
- 8) C. Muelder, K.-L. Ma: A Treemap Based Method for Rapid Layout of Large Graphs, IEEE Pacific Visualization Symposium, 231-238, 2008.
- 9) E. R. Gansner, Y. Hu, S. North, C. Scheidegger: Multilevel Agglomerative Edge Bundling for Visualizing Large Graphs, IEEE Pacific Visualization Symposium, 187-194, 2011.
- 10) D. Holten, J. van Wijk: Force-Directed Edge Bundling for Graph Visualization, Computer Graphics Forum, 28(3), 983-990, 2009.
- 11) H. Zhou, X. Yuan, W. Cui, H. Qu, B. Chen: Energy-Based Hierarchical Edge Clustering of Graphs, IEEE Pacific Visualization Symposium, 55-61, 2008.
- 12) <http://www.cs.unm.edu/~aaron/research/fastmodularity.htm>
- 13) iRefIndex, <http://irefindex.uio.no/wiki/iRefIndex>
- 14) NCBI Entrez Gene, <http://www.ncbi.nlm.nih.gov/gene>
- 15) Cytoscape, <http://www.cytoscape.org/>

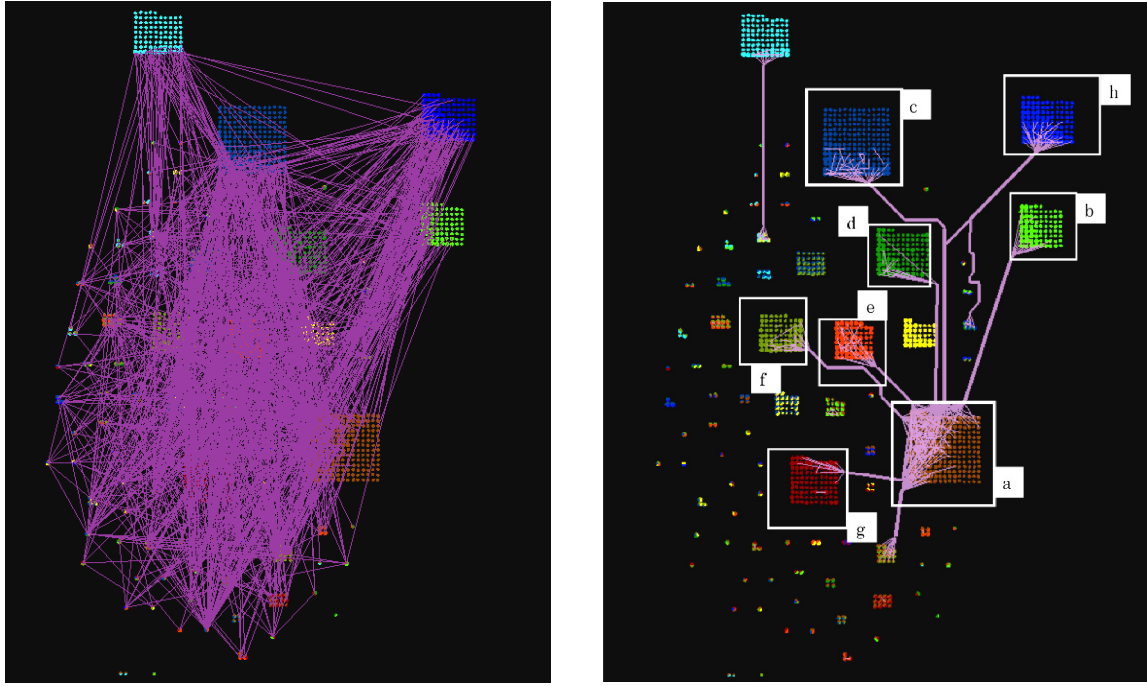


Fig. 7 可視化結果 2 (左) エッジの束化前 (右) エッジの束化後