

エッジの束化を用いた遺伝子ネットワークの可視化

A Visualization Technique for Gene Networks with Edge Bundling

中澤 里奈^{*1} 伊藤 貴之^{*1} 瀬々 潤^{*2} 寺田 愛花^{*2}
 Rina Nakazawa Takayuki Itoh Jun Sese Aika Terada

^{*1} お茶の水女子大学大学院人間文化創成科学研究科 ^{*2} 東京工業大学大学院情報理工学研究科
 Ochanomizu University Tokyo Institute of Technology

Gene networks have been constructed with genes as nodes and interactions between genes as edges so as to reveal gene functions and relationship. However, nodes and edges of the gene networks are usually very numerous. Because of that, it may be difficult to understand relations between genomic functions and gene-gene interactions, if it is visualized by traditional techniques. In this paper, we introduce a simultaneous visualization of a large gene network and gene ontology (GO), which summarizes gene functions and attributes. The technique represents the functions provided by GO as colors of nodes, and bundles edges depending on the gene functions to ease visual complication of the network.

1. 概要

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。この関係性を解明するために、遺伝子をノード、遺伝子間の相互作用をエッジとする遺伝子ネットワークが構築され、遺伝子機能発見に役立てられている。しかし、遺伝子ネットワークのノードとエッジの数は非常に膨大であり、そのまま可視化したのでは機能と相互作用の関係性を視認しづらい。

そこで本報告では、可視化手法 FRUITS Net[T.Itoh 2009]を用いて、遺伝子機能情報である GO (Gene Ontology) [GO Consortium]と遺伝子ネットワークの統合可視化手法を提案する。GO とは、種や現象ごとにこれまで独立につけられていた遺伝子情報の名称や定義を、共通の語彙を使用して定義し、有効活用できるようにした概念体系のことである。GO によって定義された用語は GO タームと呼ばれる。この GO によって定義された機能情報を、遺伝子ネットワークと同時に可視化することで、遺伝子の未発見の機能や関係性を推測しやすくし、生物学の研究分野で実験を行う方針をたてる支援や、遺伝病などの原因究明に繋がるようになるのではないかと考えられる。さらに、大量の線情報を束ねるエッジバンドリングを用いることで視認性を向上させる。本手法により、遺伝子ネットワークの関係把握や機能推測がより容易になるのではないかと考えられる。

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 2007]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 2010]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。[B.Breitkreutz 2003]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べる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処理にはあまり着目していない。

2.2 FRUITS Net

伊藤らは FRUITS (Framework and User Interface for Tangled Segments) Net [T.Itoh 2009]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。図 1 にその可視化の一例を示す。文献[T.Itoh 2009]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、本報告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。

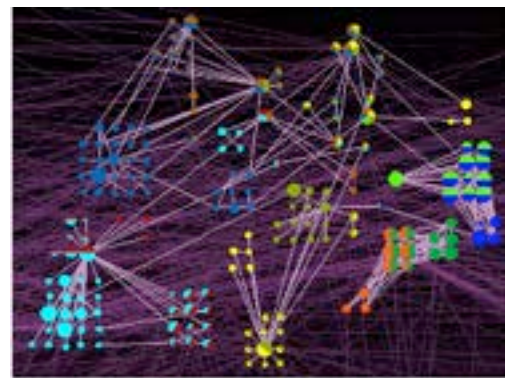


図 1 FRUITS Net の可視化例

3. 提案手法

3.1 問題点

本手法では FRUITS Net[T.Itoh 2009]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、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、

連絡先: 中澤 里奈, 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, leena@itolab.is.ocha.ac.jp

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図 2 である。使用したデータは、4.1 節で述べているものと同様のデータである。問題点として以下の二点が挙げられる。

- 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。
- 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。

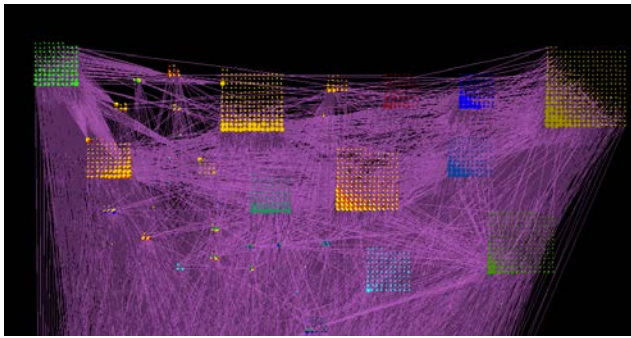


図 2 遺伝子ネットワークの可視化に関する問題点

3.2 GO クラスターリング

より多くの GO 情報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ーリングする。その後、特定の条件を満たす 10~15 種類の GO タームクラスターをアイテム情報としてノードに付与する。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 ノードのクラスターリングと画面配置

さらに、FRUITS Net を使用する前処理段階として、ノードクラスターリングを実施する。ノードのアイテム情報である GO タームの組み合わせと、エッジの密度の両者を考慮してクラスターリングをする。そして、クラスターを単位として FRUITS Net の配置アルゴリズムを用いることでノードの配置を決定する。力学モデルによって共通のアイテム情報を持つクラスターを近くに配置し、空間充填モデルによりノードやクラスターの重なりを回避しつつ、配置画面占有面積を減らすことができる。この時、ノードの大きさはそのノードが有しているエッジの本数に比例して大きく描いている。

3.4 エッジバンドリング

続いて、線情報の視認性を向上するためにエッジバンドリングを施す。よく知られた手法として、[D.Holten 2006]による階層型データのバンドリング手法や、[E.R.Gansner 2011]による経路情報のバンドリング手法がある。後者は経路情報であるがゆえにエッジ自体の位置情報も重要である。そのためバンドリング結果である束に対して元のエッジ位置からの移動量を小さくする必要があり、大きな計算時間を要する。一方で提案手法では、接続された遺伝子ペアという位相的な情報だけが重要なので、もっと簡易な手法により高速にバンドリングを実現する。また、方向性だけに基いてエッジをバンドリングしてしまうと、異なる機能を有するノードに接続されたエッジ群をひとつの束にしてしま

う恐れがある。そこで本手法では、同じ機能の組み合わせを持つクラスターを単位としてエッジバンドリングを適用する。

バンドリングの手順としては、まず、一つのクラスターに注目し、そのクラスターと他のクラスターとの間に接続しているエッジの本数をそれぞれ集計する。クラスター間のエッジがユーザの指定した本数以上ならば、クラスター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。エッジの本数が満たない場合は、バンドリングを行わず、エッジをそのまま描く。以上の処理を全ての 2 クラスター間に対して実施する。ここでクラスター間のエッジの本数 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、束を直線ではなく折れ線で描画し、ノードのクラスターを回り込むようにしてノードとの重なりを回避する。

3.5 GUI 機能

操作パネル上では可視化結果の拡大縮小や移動を行うことができる以外に、ノード強調表示を適用するための GO タームを選択することができる。また、エッジバンドリングを行うかどうかの切り替えを行い、バンドリングを適用するエッジの最低本数を操作パネル上のスライダーによってユーザが指定することができる。さらに、可視化したノードをクリックすることで、その遺伝子の ID とそれが可視化結果上で所属しているクラスターの GO タームを詳細情報として表示するとともに、そのノードに接続するエッジをハイライトする。エッジのハイライトを行うことで、より着目したいノードの詳細な接続関係を見ることができる。このエッジのハイライトをノード二つに対してまで行うことで、二つのノードそれぞれの接続関係を比較することができる。

4. 実行例

4.1 使用データ

我々は iRefIndex[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。ノード数は 8945、エッジ数は 32703、既知の GO ターム数は 259、アイテム情報としてノードに付与した GO クラスターは 12 個であった。

4.2 実行結果

エッジバンドリング処理を施す前の可視化結果を図 3(上)に示す。

図 3(上)の左上部では、例えば(a)のクラスターの機能間の関係性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、エッジバンドリング処理を行った後の可視化結果である図 3(下)では、図 3(上)では読み取りにくい、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)や(d)と(e)のクラスター間において、より強い関係性をもっているということがわかった。

なお、(a)から(e)までのクラスターに付与されている GO タームは次の通りである。

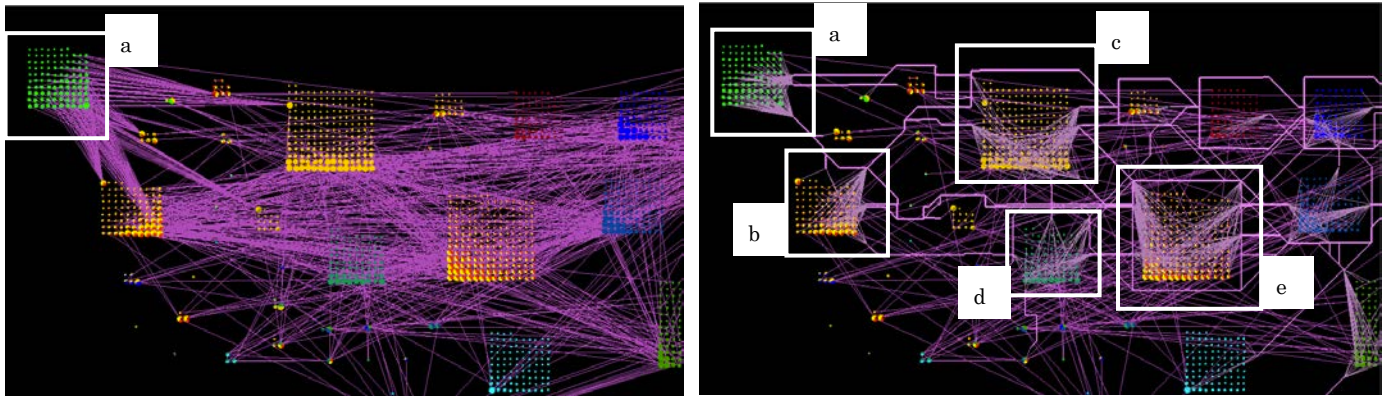


図3 可視化結果 (左)バンドリング前 (右)バンドリング後

- (a): protein-DNA complex (GO:0032993)
 (b): intracellular organelle part (GO:0044446),
 non_membrane_bounded organelle (GO:0043228),
 organelle part (GO:0044422)
 (c): non_membrane_bounded organelle (GO:0043228)
 (d): cell projection (GO:0042995)
 (e): intracellular organelle part (GO:0044446),
 organelle part (GO:0044422)

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。図 4(左)がエッジバンドリング適用前、図 4(右)がエッジバンドリング適用後の図である。図 4(左)を見ると、二つのクラスタ間の接続にも二種類の関係性があることがわかる。一つの関係性は B のクラスタと A のクラスタ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にになっているかまでは図 4(左)からは読み取れない。

一方で、図 4(右)にて同じクラスタ間の接続に着目する。図 4(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 B のクラスタと A のクラスタのノードが少数対多数対応であることがわかる。ここでエッジバンドリング前の図 4(左)から B のクラスタ内左上にあるノードが 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、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 A と接続している B のノードは、A と一对多数対応であるということが確認できる。

これらの B クラスタ内の右下にあるノードは A のクラスタ以外のノードとも接続している一方で、B クラスタ内左上のノードは A のクラスタのノードとの間にしかエッジが接続されていないことをふまえると、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 を持っているか、これらの 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、といった推測ができるのではないかと考えられる。

次に、クラスタ C に着目する。エッジバンドリング前の図 4(左)においては、クラスタ C のノードはそれぞれが多数のエッジを保有しているように見ることができる。しかし、エッジバンドリング後の図 4(右)を見ると、クラスタ C に属するノードのうちエッジが接続されているノードはごく一部だけであるということがわかる。ま

た、エッジバンドリング前の図 4(左)ではエッジが煩雑で視認できないが、エッジバンドリング後の図 4(右)ではクラスタ C 間のノード間にもエッジが接続されていることが見てとれる。

このように、エッジバンドリングを適用することで、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、その上エッジバンドリング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、エッジバンドリング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

本報告では、エッジバンドリング機能を追加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、遺伝子ノードをクラスタリングする。そして、エッジで接続されたノードや同一アイテムを共有するノードが画面上で近くに配置されるように、ノードの画面上の位置を決定する。それでも、そのまま可視化したのでは視認性に問題があるので、エッジバンドリングによって視認性を向上する。

今後の課題としては、以下の点を議論したいと考えている。

- 現時点のエッジバンドリング処理の実装では、同じ経路をたどる束が重なるので、その重なりを回避する。
- 束とエッジを接続した時に同方向のエッジが重なってしまうことがあるので、描画を工夫する。
- 現在のクラスタリング結果の妥当性の検証を行う。
- 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の検証を行う。
- 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する。

謝辞

本研究を進めるにあたり多数のご助言をいただきました、お茶の水女子大学の由良 敬教授に感謝の意を表します。

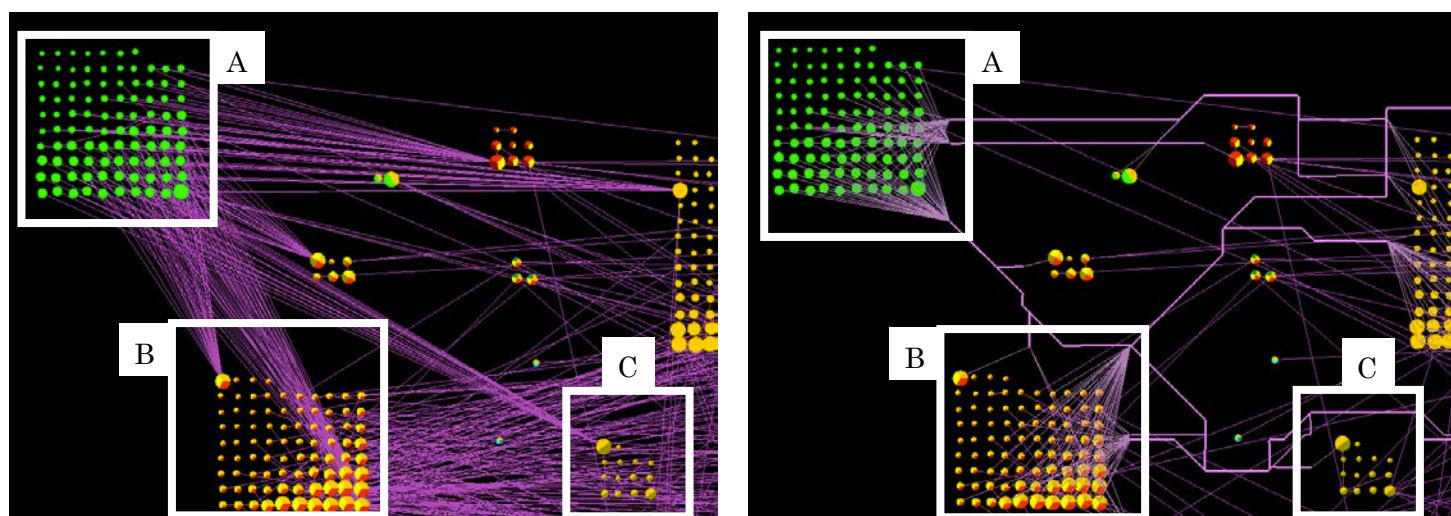


図4 可視化結果の拡大図 (左)エッジバンドリング前 (右)エッジバンドリング後

参考文献

- [T.Itoh 2009] 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, pp.121-128, 2009.
- [GO Consortium] GO Consortium: The Gene Ontology, <http://www.geneontology.org/>
- [西山 2007] 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, Vol. 6, No. 3, pp. 106-116, 2007.
- [伊澤 2010] 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.
- [B.Breitkreutz 2003] B. Breitkreutz, C. Stark, and M. Tyers: Osprey: a network visualization system, Genome Biology, 2003.
- [D.Holten 2006] D. Holten: Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, IEEE Transactions On Visualization And Computer Graphics, Vol. 12, No. 5, 2006.
- [E. R. Gansner 2011] E. R. Gansner, Y. Hu, S. North, C. Scheidegger: Multilevel Agglomerative Edge Bundling for Visualizing Large Graphs, IEEE Pacific Visualization Symposium, pp.187-194, 2011.
- [iRefIndex] iRefIndex, <http://irefindex.uio.no/wiki/iRefIndex>