

遺伝子ネットワークと Gene Ontology の統合可視化の一手法

中澤 里奈[†] 伊藤 貴之[†] 瀬々 潤[‡] 寺田 愛花[‡]

[†]お茶の水女子大学 〒112-8610 東京都文京区大塚 2-1-1

[‡]東京工業大学 〒152-8550 東京都目黒区大岡山 2-12-1

E-mail: [†]{leena, itot}@itolab.is.ocha.ac.jp,

[‡]sesejun@cs.titech.ac.jp, terada.a.ab@m.titech.ac.jp,

あらまし 未知の遺伝子機能や他の遺伝子との関係性を解明するために、各遺伝子をノード、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を構築することがある。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、遺伝子機能情報である GO(Gene Ontology)と遺伝子ネットワークとの統合可視化手法を提案する。本手法では、遺伝子間の相互作用をエッジで表現するとともに、GO が遺伝子に割り当てる機能の種類をアイテム情報として付加し、ノードの色で表現している。また、視認性をあげるためにアイテム情報毎にエッジを束ねるというエッジバンドリングを施して煩雑さを低減させている。

キーワード 可視化, 遺伝子ネットワーク, 遺伝子オントロジー, エッジバンドリング

A Visualization Technique for Gene Networks and Gene Ontology

Rina NAKAZAWA[†] Takayuki ITOH[†] Jun SESE[‡] and Aika TERADA[‡]

[†] Department of Information Sciences, Ochanomizu University 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

[‡] Department of Computer Science, Tokyo Institute of Technology 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550 Japan

E-mail: [†]{leena, itot}@itolab.is.ocha.ac.jp,

[‡]sesejun@cs.titech.ac.jp, terada.a.ab@m.titech.ac.jp,

Abstract A gene network is constructed with genes as nodes, and interactions between genes as edges so as to reveal unknown gene functions and relationship. However, nodes and edges of 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. This paper presents our technique on visualization of gene networks and gene ontology (GO), which summarizes gene functions and attributes. The technique represents the functions defined by GO as colors of nodes, and bundles edges depending on the gene functions to ease visual complication of the network.

Keyword Visualization, Gene Network, Gene Ontology, Edge Bundling

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、各遺伝子をノード、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を構築することがある。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 FRUITSN[1]を用いて、遺伝子機能情報である GO (Gene Ontology) [2]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、生物学の研究分野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、共通の語彙をしようして定義し、有効活用できるようにした概念体系のことである。計算機上での GO データは DAG(Directed Acyclic Graph:無閉路有向グラフ)構造によって階層的に表現されている。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。

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[3]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[4]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。Breitkreutz[5]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 [1]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[1]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。

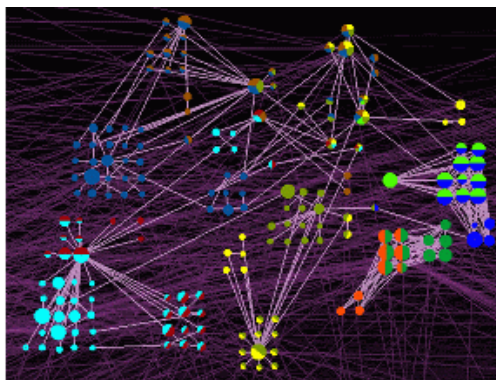


図1 FRUITSNetの可視化例

3. 提案手法

3.1 問題点

本手法では FRUITSNet[1]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，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。

FRUITSNet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。

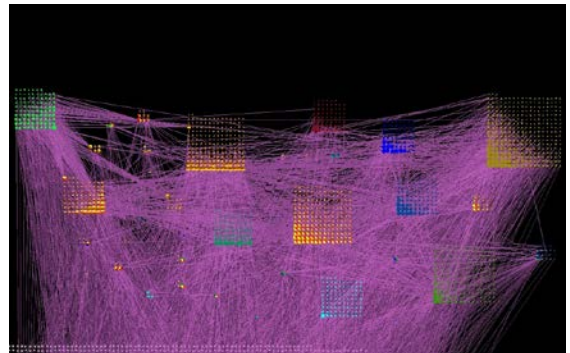


図2 可視化例

対象とするデータが文献[1]よりも大規模になっているため，可視化結果が雑然としてしまい，関係性の明確さが失われていると考えられる。この問題を解決するために，以下の処理を適用する。

3.2 GO クラスタリング

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 ノードのクラスタリングと画面配置

さらに、FRUITSNet[1]を使用する前処理段階として、ノードクラスタリングを実施する。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リングをする。そして、クラスタを単位として FRUITSNet 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。

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、Holten[6]による階層型データのバンドリング手法や、Gansnerら[7]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、方向性だけに基づ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、一つのクラスタに注目し、そのクラスタと他のクラスタ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クラスタ間に対して実施する。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、束を折れ線で描画し、ノードのクラスタとの重なりを回避する。

4. 実行例

4.1 使用データ

我々は iRefIndex[8]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 に示す。

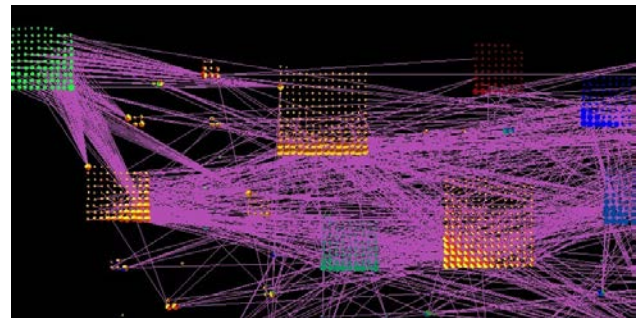


図 3 バンドリング処理前

図の左上部では機能間の関係性などを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 4 では、図 3 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。



図 4 バンドリング処理後

さらに、この可視化結果の左上部の黄緑色と黄色の二つのクラスタの関係性に着目する。図 5(左)がエッジバンドリング処理前の図、図 5(右)がバンドリング処理後の図である。図 5(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。一つの関係性は黄色のクラスタと黄緑色のクラスタのノードが一対複数対応であるのに対し、もう一つの関係性がどのようになっているかまでは図 5(左)からは読み取

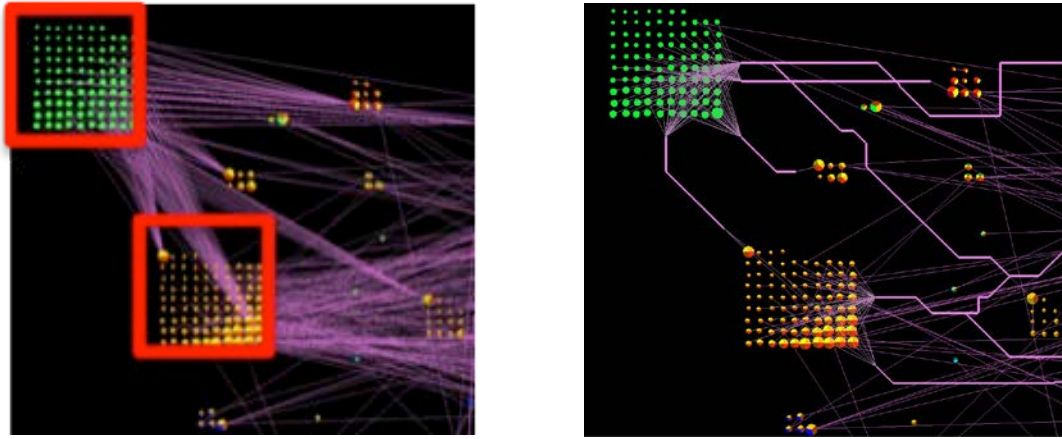


図5 可視化結果拡大図 (左)バンドリング前 (右)バンドリング後

れない。一方で、図5(右)にて同じクラスタ間の接続に着目する。図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黄色のクラスタと黄緑色のクラスタのノードが少数対複数対応であることがわかる。ここでバンドリング前の図から黄色のクラスタ内左上にあるノードが黄緑色のノードに対して一対複数対応だったことを考慮すると、他のノードも同様に一対複数対応関係にあるのではないかと考えられる。これらの黄色クラスタ内の右下にあるノードは黄緑色のクラスタの以外とも関係を持っている一方で、黄色クラスタ内左上のノードは黄緑色のクラスタとしか関係性を持っていないことから、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、その他のGOがこの共通した関係性に影響を与えているといった推測ができ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、遺伝子ネットワークと Gene Ontology を、可視化手法 FRUITSNET を用いて統合可視化する手法を提案し、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

なってしまふことがあるので、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。
- 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。

謝辞

本研究を進めるにあたり多数のご助言をくださいました、お茶の水女子大学の由良敬教授に感謝の意を表します。

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, pp.121-128, 2009.
- [2] The Gene Ontology, <http://www.geneontology.org/>
- [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, Vol. 6, No. 3, pp. 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, 4:R22.
- [6] D. Holten, Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, IEEE Transactions On Visualization And Computer Graphics, Vol. 12, No. 5, 2006.
- [7] E. R. Gansner, Y. Hu, S. North, C. Scheidegger, Multilevel Agglomerative Edge Bundling for Visualizing Large Graphs, IEEE Pacific Visualization Symposium, pp187-194, 2011.
- [8] iRefIndex, <http://irefindex.uio.no/wiki/iRefIndex>