

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 (指導教員: 伊藤貴之)

1. 概要

近年、様々な遺伝子関連情報が明らかになってきたが、未だに機能のわからない遺伝子も多く存在している。この機能解明に向けて、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 FRUITSNet[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、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。

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。Breitkreutz[4]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。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。

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 1 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、ただだか 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。

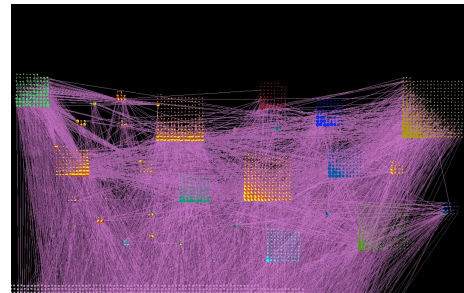


図 1 可視化例

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。

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。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、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[5]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

実行結果を図 2 に示す。

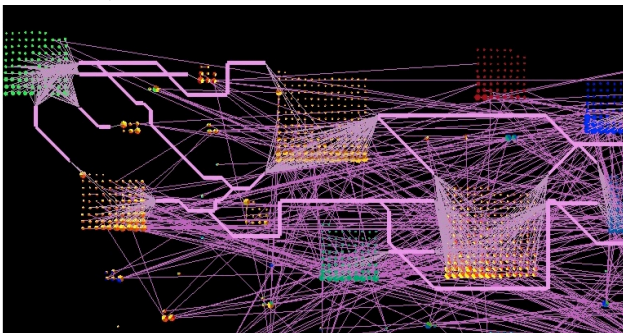


図 2 実行結果

図 2 では、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。

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 3(左)がエッジバンドリング処理前の図、図 3(右)がバンドリング処理後の図である。図 3(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になっているかまでは図 3(左)からは読み取れない。一方で、図 3(右)にて同じクラスタ間の接続に着目する。図 3(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 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对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 B クラスタ内の右下にあるノードは A のクラスタの以外とも関係を持っている一方で、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。

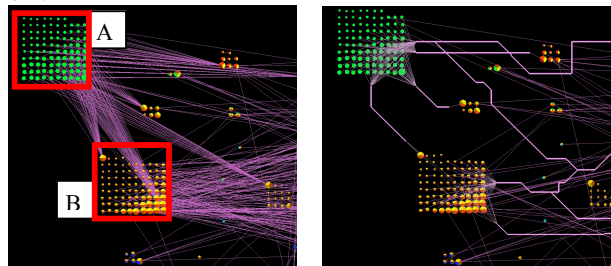


図 3 可視化結果拡大図
(左)バンドリング前 (右)バンドリング後

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 FRUITSNNet を用いて統合可視化する手法を提案し、実行例を示した。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。
- 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いただきました。東京工業大学の瀬々潤准教授、寺田愛花氏、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] B. Breitkreutz, C. Stark, and M. Tyers, Osprey: a network visualization system, Genome Biology 2003, 4:R22.
- [5] iRefIndex, <http://irefindex.uio.no/wiki/iRefIndex>