

修 士 論 文

題 目 L^AT_EXを用いた修論・卒論の執筆

指導教員 電気 太郎 教授

京都大学大学院工学研究科 電気工学専攻

氏 名 電気 次郎

平成30年2月1日

Usage of The L^AT_EX Style File for KUEE

Jiro Denki

Abstract

This document briefly explains the usage of the L^AT_EX style file for KUEE bachelor thesis and master thesis.

目 次

第 1 章 導入	1
第 2 章 関連研究	6
2.1 生物学データの可視化	6
2.2 Graph-drawing Method for Group Structure	6
謝 辞	8
参 考 文 献	9

第1章 導入

生命は生化学現象を基盤とし、遺伝子や mRNA、タンパク質、表現型特徴などの異なるスケールで制御される複雑系からなる。この生命現象を理解するためには、様々な現象を観測し、そのダイナミクスや異なるスケール感の相互作用を細かに観察する必要がある。観測したデータから新たな知見を導き出すために、データ駆動型科学が行われている。これは事前の仮設なしにデータからないと言えるのかを考え、科学的手法における仮設生成をデータに基づいて行うというアプローチである。データ駆動型科学においては、対象となるデータをいかに可視化することが重要となる。効果的な可視化を用いれば、データ得られる仮設の信頼性は高くなり、データ解析の効率も向上すると考える。

近年では顕微鏡により計測された細胞の時空間胴体を自動的に観測するバイオイメージ・インフォマティクス技術が発達したことで、大量かつ複雑な生命科学データが取得されている。こうしたデータを有効に解析することができれば、新たな生物学的知見を得ることができるだろう。しかし、増え続けるデータに従来の手法で効果的に可視化するのは難しく、データに対応するため新たな可視化手法の開発も進んでいる。バイオイメージ・インフォマティクスの進歩に合わせ、可視化技術も進歩させることが有効なデータ駆動型科学には不可欠だと考える。

線虫 (*C. elegans*) は線形動物門双腺綱に属する体長 1mm ほどの糸状の生物である。実験材料として優れた性質をもつことから、モデル生物として広く利用されている他、多細胞生物として最初に全ゲノム配列が解読された生物でもある。我々の共同研究者である理化学研究所の生命科学者達は、*C. elegans* を用いて個体発生を制御する遺伝子ネットワークを明らかにしようとしている。比較的単純な細胞構造を持つ *C. elegans* の発生メカニズムが明らかになれば、人間のようなより複雑な多細胞生物の発生メカニズムの解明につながり、ひいてはガン研究や再生医療に大きな貢献ができるであろう。

理化学研究所の研究員達は、現在 *C. elegans* の表現型特徴ネットワークから生物学的な因果関係を示唆する関係性を抽出することを目指している。表現型特徴とはある生物の持つ遺伝子型を形質として示したものである。複数の表現型特徴間の相関や因果関係を探ることで、発生のメカニズムを解き明かすことという研究

である。発生中の胚の細胞の胴体を 4 次元的に計測した顕微鏡画像を取得し、その動画データに対して画像認識プログラムを適用して核の大きさ等の胚の表現型特徴を動態を計測する。計測された表現型特徴間の相関を計算することで、表現型特徴間の関係性から図 2.2 のような表現型特徴ネットワークが構築される。ここで、ノードはそれぞれの表現型特徴を表し、エッジは表現型特徴の相関を表す。このネットワークから表現型特徴間の繋がりを抽出したり、重要な表現型特徴を繋ぐパスを探索することで仮設生成を行っている。

図 2.2 は各表現型特徴をその発現時期と発現する細胞を基にグループ分けしている。一番左に 1 細胞期を配置し、その右は 1 細胞期から 2 細胞期への変移期、その右は 2 細胞期と横方向には表現型特徴の発現時期順にグループを配置している。縦方向は細胞ごとにグループを分けており、4 細胞期には細胞ごとに 4 つのグループが存在する。このレイアウトは各グループを階層的に配置しているために、全体のグループ構造を把握するには適しているが、実際のネットワーク解析では有効ではないと考えられる。というのも、このレイアウトにはノード間を繋ぐエッジの情報が介在していないからである。グラフ描画の分野では、ネットワークの可読性に影響する様々な要素が報告されている [1, 2, 3, 4]。具体的には、グループを表現する箱のアスペクト比やエッジの交差数、エッジの長さなどがネットワークの可読性を左右する。より良い解析を行うためにはより効果的なレイアウトによってネットワークを可視化する必要があり、またドメインからのそうしたニーズが存在する。

表現型特徴ネットワークに見られるノードのグループはコミュニティ構造と呼ばれ、この特徴を持つネットワークを複雑ネットワークと呼ぶ [5, 6, 7]。グラフ描画の分野では複雑ネットワークを可視化するため様々なネットワーク可視化手法が提案されてきた [8, 9, 10]。こうした手法はそれぞれ異なる目的のもと提案されている。例えば隣接行列はネットワークデータを可視化する手法の一つであり、ノードを適当に並べ替えられる、ノード間にエッジがあるかなどの判定が容易であるというメリットを持つが、一方で複数ノード間の繋がりが理解しにくい、ノードの配置に位置情報を反映できないなどのデメリットがある。このように各可視化手法には向き不向きがあることから、表現型特徴ネットワークにもそれに適した可視化手法を用いることが必要である。

Vehlow らは複雑ネットワークを対象にした可視化手法に関する調査論文を発表している [8]。Vehlow らは現行の手法をいくつかのカテゴリに分けて紹介してい

る。この中で、表現型特徴ネットワークは Disjoint flat のグループに入る。これはノードが複数のグループに属さず、またグループ間に階層構造が存在しないというものである。Disjoint flat なネットワークを可視化する手法にもいくつかの種類がある。ノードの色で属するグループを表現するもの、グループ構造をネットワークと別に表すもの、グループ内のつながりを遷移行列で表すもの、グループ毎にノードを円や四角形で囲うものなどがある。このうち、ノードの色でグループ構造を表すものはグループやノードの数が増えた際にカラーマップが難しくなるという欠点がある。遷移行列を用いるものはグループ間の繋がりや複数ノード間の接続が理解しにくく表現型特徴ネットワークに適さない。グループ毎に属するノードを図形で囲うという方法は、グループ構造のみならずエッジの情報など様々な要因を考慮してレイアウトができるため有効だと考えられ、Vehlow らも関連する研究を多数紹介している。

我々は、表現型特徴ネットワークを可視化するのには同一グループに属するノードが四角形で囲まれているレイアウトが最良と考える。他の図形に比べ四角形で囲まれたグループはその配置を目的に合わせコンピューターによる計算で調整できる他、スクリーンの使用効率が高いからである。こうした可視化手法は Group-In-a-Box レイアウト (GIB) と呼ばれ、そのうち 4 種類を図 2.2 に示す。GIB は各グループに属するノードがノード数に比例した大きさの箱に囲まれており、グループ内のネットワーク、グループ間のネットワーク、グループの大きさを同時に表現できる。図 2.2 も各グループが箱に入っているが、箱の大きさがグループに属するノード数に対応しておらず、厳密には GIB レイアウトではない。しかし、理化学研究所の研究者らが元来グループを箱として表現していたことから、GIB は研究目的に沿うものだと判断した。

GIB は表現型特徴ネットワークを表すのに適したレイアウトであるが、種類が多いためどのレイアウトが一番有効であるか分かっていない。いくつかの研究 [11, 12] で、計算的実験により GIB が評価されているが、我々の知る限りユーザー実験により GIB を評価した研究は未だ存在しない。上述したようにいくつかの計算可能な指標がグラフの可読性に影響することが分かっている [1, 2, 3, 4] が、グラフの理解には人間の認知の観点からも様々な未解明の要因が影響するため、ユーザー実験を通じた評価の必要性は高い。

以上のことから、我々は 2 種 <https://ja.overleaf.com/project/5b990a8b21b7c041ba8d08ba> 類のユーザー実験を行い、4 種の GIB のうちどのレイアウトが最もデータ駆動型

科学において表現型ネットワーク解析に有効かを検証する。対象となる GIB は図 2.2 に示す 4 種類であり、順に ST-GIB (Squarified-Treemap GIB), CD-GIB (Croissant-and-Doughnut GIB), FD-GIB (Force-Directed GIB), TR-GIB (Tree-Reordering GIB) である。ST-GIB は箱の配置に squarified treemap アルゴリズム [13] を使用している。CD-GIB はグループにエッジで繋がっている他のグループの数に基づきグループを配置するものである、エッジの長さを短くして視覚的乱雑さを軽減する目的がある [12]。FD-GIB は各グループをノードとして捉え、グループの配置に force-directed アルゴリズムを用いたものである [12]。TR-GIB は最適化問題を解いてエッジの長さの和が最小になるように ST-GIB 内で箱を並び替えたものである。

一つ目の実験では、4 種類のタスクを用いて様々な側面から 4 種類の GIB を評価することを目的とする。二つ目の実験では、一つ目の実験で好成績を残した 2 つのレイアウトに着目し、より実際の解析作業に近いタスクとデータを用いることでどちらのレイアウトが最良であるかを明らかにする。

実験では、正答率と完了時間を計測することでレイアウト間の優劣を比較した。これに加え、視線追跡システムを利用し実験中の被験者の視線データを記録した。視線追跡システムは、ヒューマンコンピュータインタラクションの分野でよく用いられる。実験中の視線データを解析することで、実験結果の優劣がなぜ生まれたかについて議論することができる [14, 15, 16]。具体的には、ネットワークグラフにおけるどの視覚特徴が被験者のパフォーマンスに影響したかを特定することを目指す。

本研究の主な有効性は 4 種の GIB を評価することでどのレイアウトが表現型特徴ネットワークの解析に最も有効かを明らかにすることである。我々は GIB はその視覚的特徴の違いから、ユーザー実験の結果にも差が出ると考えている。最良のレイアウトが明らかになれば、それを用いることで表現型特徴ネットワークの解析の効率化が測れるだろう。有効性としてもう一点、視線追跡データを解析することにはより、ネットワーク可視化において可読性に影響し得る視覚的特徴を特定することが挙げられる。こうした特徴を発見できれば、GIB や他のグラフ描画手法の可読性を向上させるアイデアを得られ、グラフ可視化手法の改善に貢献することができる。

また、本研究は表現型特徴ネットワークの解析効率化に動機づけられたものであるが、GIB は他の複雑ネットワークにも利用可能な手法である。上述した Twitter のような SNS ネットワークの他、企業間取引ネットワークや交通ネットワーク

にも応用が可能である。本研究で明らかになった最良の GIB はこうしたデータでも有効であると考えられる。

第2章 関連研究

本章では関連研究について記述する。本研究は生物学やグラフ描画、可視化の評価など様々な研究が関連するため、3節に分けてこれを述べる。2.1節では生物学データの可視化について述べる。特に、本研究は表現型特徴ネットワークの可視化に焦点を当てており、現在までの研究とその問題点について論述する。2.2節ではグループ構造を持つネットワークの可視化手法について述べる。こうした可視化手法にはいくつか種類があるが、その中でのGIBの立ち位置と、なぜGIBを研究対象にしたかを詳細に説明する。??節では視線追跡システムを用いた可視化の評価研究について記述する。可視化の評価という分野は近年その重要性を高めているが、その中でも視線追跡システムを用いたものは結果をより定量的に述べられる上更なる考察を与えられるがために学術的な価値が高い。一方で、高性能な視線追跡システムが一般に使用されるようになってから日が浅く、そのデータの膨大さから解析方法もあまり整っていない。関連研究を詳細に述べることで、本研究の目的に即した視線データ解析を目指す。

2.1 生物学データの可視化

顕微鏡により計測された細胞の動態を自動的に計測するバイオインメージ・インフォマティクス技術の発達により、大量の生命科学データが記録されている。こうしたデータを解析するために、ImageJ[17]に代表されるような取得された画像データを定量化し可視化する汎用的なツールも盛んに開発されている [18, 19, 20]

2.2 Graph-drawing Method for Group Structure

データの重要性が広く認知されている近年では、頻繁なデータ収集や計測機器の進化によりネットワークデータはその量も複雑性も増している。ネットワークの膨大さと複雑さに対応するため、グラフ描画の分野では様々な手法が提案されてきた。Eadesらはforce-directed layoutと呼ばれる手法を提案した[21]。この手法では、ノード間の斥力とエッジ間の引力によって各ノードが配置される。エッジが

引力を持つことから、エッジにより繋がるノード同士が近くに配置され、エッジの長さが短くなる。その計算的指標の優位性と見た目の美しさから、force-directed レイアウトは広く利用されいて [22]、かつこの手法にはにはいくつかの計算アルゴリズムが存在する [1, 2, 3]。

謝 辭

参 考 文 献

- [1] David Harel and Yehuda Koren, A fast multi-scale method for drawing large graphs, *International symposium on graph drawing*, (Springer, 2000), pp. 183–196.
- [2] Yehuda Koren, Liran Carmel, and David Harel, Drawing huge graphs by algebraic multigrid optimization, *Multiscale Modeling & Simulation*, Vol. 1, No. 4, (2003), pp. 645–673.
- [3] Stefan Hachul and Michael Jünger, Drawing large graphs with a potential-field-based multilevel algorithm, *International Symposium on Graph Drawing*, (Springer, 2004), pp. 285–295.
- [4] Emilio Di Giacomo, Walter Didimo, Giuseppe Liotta, and Fabrizio Montecchiani, Area requirement of graph drawings with few crossings per edge, Vol. 46, (2013), pp. 909–916.
- [5] Mark Newman, *Networks: An Introduction*, (Oxford University Press, Inc., New York, NY, USA, 2010).
- [6] Michelle Girvan and Mark EJ Newman, Community structure in social and biological networks, *Proceedings of the national academy of sciences*, Vol. 99, No. 12, (2002), pp. 7821–7826.
- [7] Mark EJ Newman, Detecting community structure in networks, *The European Physical Journal B*, Vol. 38, No. 2, (2004), pp. 321–330.
- [8] Corinna Vehlow, Fabian Beck, and Daniel Weiskopf, Visualizing group structures in graphs: A survey, *Comput. Graph. Forum*, Vol. 36, (2017), pp. 201–225.
- [9] Helen Gibson, Joe Faith, and Paul Vickers, A survey of two-dimensional graph layout techniques for information visualisation, *Information Visualization*, Vol. 12, No. 3-4, (2013), pp. 324–357.

- [10] Bahador Saket, Paolo Simonetto, and Stephen Kobourov, Group-level graph visualization taxonomy, *arXiv preprint arXiv:1403.7421*, (2014).
- [11] Yosuke Onoue and Koji Koyamada, Optimal tree reordering for group-in-a-box graph layouts, *SIGGRAPH Asia 2017 Symposium on Visualization*, (ACM, 2017), p. 13.
- [12] Snigdha Chaturvedi, Cody Dunne, Zahra Ashktorab, R Zachariah, and Ben Shneiderman, Group-in-a-box meta-layouts for topological clusters and attribute-based groups: Space-efficient visualizations of network communities and their ties, *Computer Graphics Forum*, Vol. 33, (Wiley Online Library, 2014), pp. 52–68.
- [13] Mark Bruls, Kees Huizing, and Jarke J Van Wijk, Squarified treemaps, *Data visualization 2000*, (Springer, 2000), pp. 33–42.
- [14] Gennady Andrienko, Natalia Andrienko, Michael Burch, and Daniel Weiskopf, Visual analytics methodology for eye movement studies, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, No. 12, (2012), pp. 2889–2898.
- [15] Andrew T Duchowski, Eye tracking methodology, *Theory and practice*, Vol. 328, (2007).
- [16] Kuno Kurzhals, Brian Fisher, Michael Burch, and Daniel Weiskopf, Evaluating visual analytics with eye tracking, *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, (ACM, 2014), pp. 61–69.
- [17] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri, Nih image to imagej: 25 years of image analysis, *Nature methods*, Vol. 9, No. 7, (2012), p. 671.
- [18] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al., Fiji: an open-source platform for biological-image analysis, *Nature methods*, Vol. 9, No. 7, (2012), p. 676.

- [19] Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A Guertin, Joo Han Chang, Robert A Lindquist, Jason Moffat, et al., Cellprofiler: image analysis software for identifying and quantifying cell phenotypes, *Genome biology*, Vol. 7, No. 10, (2006), p. R100.
- [20] Jianxu Chen, Liya Ding, Matheus P. Viana, Melissa C. Hendershott, Ruian Yang, Irina A. Mueller, and Susanne M. Rafelski, The allen cell structure segmenter: a new open source toolkit for segmenting 3d intracellular structures in fluorescence microscopy images, *bioRxiv*, (2018).
- [21] P. Eades, A heuristic for graph drawing, *Congressus Numerantium*, Vol. 42, (1984), pp. 149–160.
- [22] Stephen G. Kobourov, Force-directed drawing algorithms, *Handbook of Graph Drawing and Visualization*, (2013), pp. 383–408.

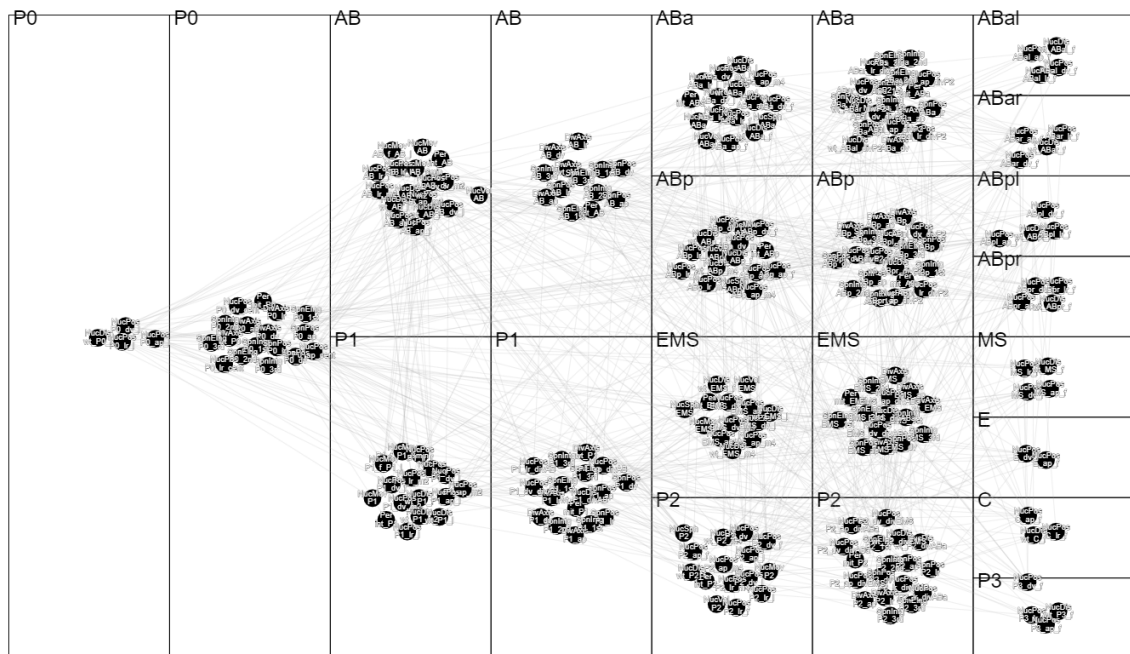


図 1.1: 表現型特徴ネットワークの例

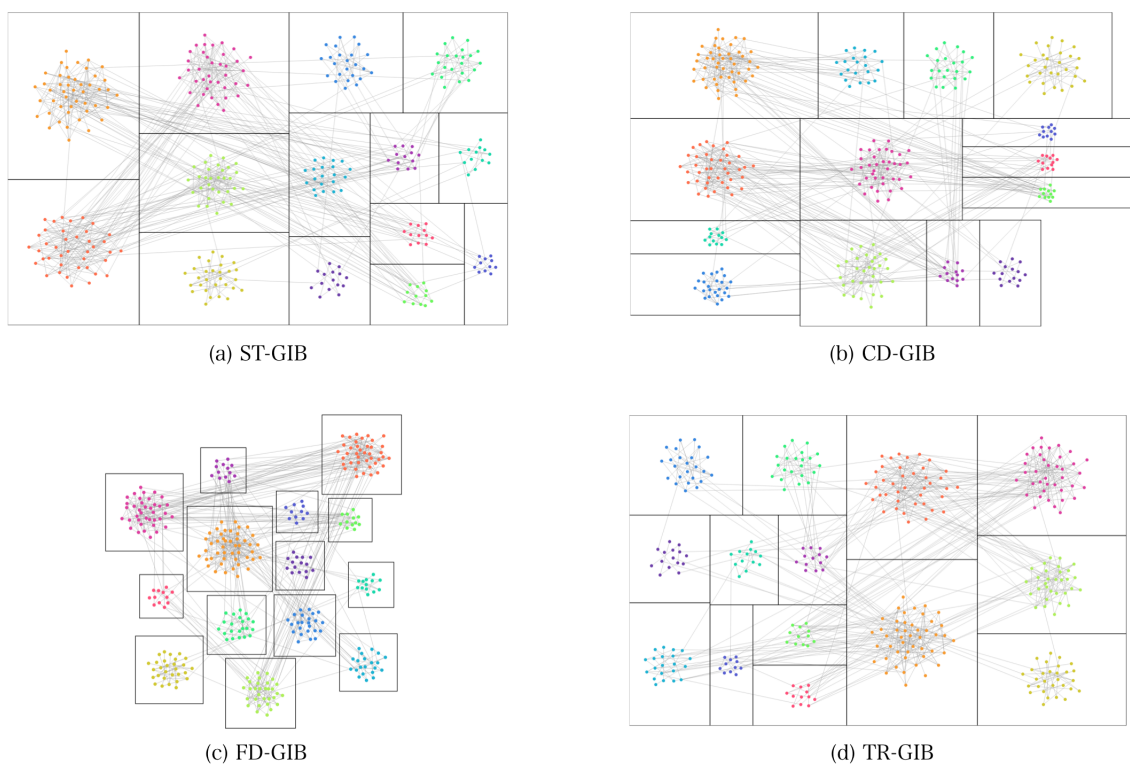


図 1.2: GIB レイアウトの種類。(a)ST-GIB, (b)CD-GIB, (c)FD-GIB, (d)TR-GIB