# 複数層ネットワークを用いた *E. coli*の生体ネットワークに おけるロバストネス解析手法の提案 (2014年2月7日版)

# 高橋 勝人<sup>†1,a)</sup>

概要:生体ネットワークの研究において,複雑ネットワークを応用した機能予測やネットワークの制御の 研究が行われている.従来の研究では遺伝子や化合物といった対象を単一層のネットワークとして研究し ていた.しかし,実際の生体ネットワークでは複雑にネットワーク同士でも相互作用を及ぼし合っている. そこで本研究では生体ネットワークを,複数層構造を持つ統合ネットワークとして扱う方法を導入した. モデル生物に E. coli を用い,遺伝子制御ネットワークと代謝ネットワークの2つのネットワーク情報を 統合し,パーコレーション過程の計算実験を行た.また,この E. coli ネットワークを説明しうる論理モデ ルを導入し,同様にパーコレーション過程の計算を行った.この結果,層構造を考慮したパーコレーショ ン過程の計算を行った結果, E. coli のネットワークでは頑健な構造を有している事が明らかとなり,従来 の方法では捉えられない構造的特徴を明らかにすることができた.

**キーワード**:複雑ネットワーク,バイオインフォマティクス,複数層ネットワーク,パーコレーション, モデル化

Masato Takahashi<sup>†1,a)</sup>

# 1. はじめに

生体ネットワークから社会ネットワークに至るまで,全 ての相互関係は複雑ネットワークのフレームワークを用い て説明することができる [1], [2], [3], [4]. 複雑ネットワー クは生体ネットワークの解析と理解において,その大胆な 抽象化にもかかわらず,有用な情報を提示してきた.細胞 内での相互作用ネットワークは,特にインタラクトームと 呼ばれ,遺伝子制御ネットワークから遺伝子疾患ネット ワークに至るまで様々なレベルにおいて研究が行われてい る [2], [5], [6], [7], [8]. また,複雑ネットワーク上でのパー コレーション過程から,レジリアンスという視点で生体 ネットワークを理解する試みもある [9].しかし,先行研究 において,本来ならば遺伝子相互作用や代謝経路等といっ た生体ネットワークは,相互に影響を及ぼし合っているに も関わらず,各ネットワークはそれぞれ独立に研究が行わ れてきた.

 $^{a)}$  g2112024@fun.ac.jp

そこで、本研究では生体ネットワークを複数層の構造を 持つ1つの統合ネットワークとして扱う事を提案する.検 証のため、*E. coli*をモデル生物にパーコレーション過程を 適用して研究を行った.

## 2. 研究手法

# 2.1 データセット

本研究では、遺伝情報と代謝経路に関して多くの情報が 得られている *Escherichia coli K-12 MG1655* (*E. coli*)を モデル生物とした. *E. coli* のネットワーク情報は EcoCyc データベース (Version 17.1 2013 年 7 月 11 日取得)[10] の 情報から構成した. EcoCyc の登録データから 4625 の遺 伝子と 2421 の代謝化合物が得られた. この情報から次の 方法で酵素の存在を媒介とした 2 層ネットワークを作成し た. なお,統合したネットワークは図 1 に示した様になる.

先ず,ある遺伝子が発現している酵素の情報を得る.次 にその酵素がどの酵素反応に関わっているか,EC番号を もとに参照する.EC番号で表された酵素反応に関わる化 合物の情報を取得し,EC番号と化合物の2部グラフを作

<sup>&</sup>lt;sup>†1</sup> 現在,公立はこだて未来大学大学院 Presently with Future University Hakodate



図 1 E. coli の複数層ネットワークの部分グラフ Fig. 1 A sub-graph of E. coli multi-layer network

成する. この2部グラフを化合物の1部グラフへ写像す る. 着目していた遺伝子と酵素反応で関係付けられた化合 物を,結びつける. ここまでの手順を各遺伝子に行い,最 後に,化合物のノードに重複が存在した状態になるため, 重複した化合物ノードをそれぞれ1つのノードにまとめ る.以上の手順で *E. coli* にの遺伝子と代謝物の2つの要 素を持つ複数層ネットワークを構成した.

#### 2.2 論理モデル

EcoCyc から得た E. coli のネットワークから,内次数分 布は冪乗則,外次数分布では指数分布と冪乗則をそれぞれ 示すことを確認している.ここで,内次数分布とはある層 の内部のノード同士を繋ぐエッジの本数,外次数分布とは ある層から別の層を繋ぐエッジの本数を示す.この E. coli の複数層ネットワークにおける振る舞いを説明しうるネッ トワークの成長モデルを導入した.成長規則は A 層 B 層 の 2 層を用い,次の様に構成する.

A層側の成長規則

```
2)既にネットワークに存在するmf個のAノードをファーストコンタクト
としてランダムに選択する
```

- 3)ファーストコンタクトとして選んだmf個のノードと接続しているノード からms個のAノードとmc個のBノードをランダムに選ぶ
- 4)1で追加したノードとランダムに選ばれたノードを結ぶ

B層側の成長規則

1)各タイムステップにおいて、確率βで新しくB層へノードを追加する 2)ノードを追加した場合、ma個のA層ノードとmb個のB層ノードをランダムに選択 3)ランダムに選択したノードを新しく追加したB層ノードを接続する

以上の成長規則をネットワーク全体のノード数が指定したノード数Nに達するまで 繰り返し適用してネットワークを成長させる

## 2.3 マスター・スレーブパーコレーション

パーコレーションとは、ノードまたはエッジが除去され た割合と最大クラスタの大きさの変化の事である.ここで のクラスタとは、ネットワークの連結部分を指す.本研究 では、ノードの除去によるサイトパーコレーションを適用 し、ネットワークのレジリアンスを検証した.

先ず,2層ネットワーク全体として比較するため,両方の 層のノードのうちどちらか一方を各時点で1個ずつ除去し, ネットワーク全体の最大クラスタを計算した.単一層ネッ トワークとの比較のため,Barabási-Albert モデル (BA モ デル)[11],ランダムモデル (ER モデル)を用いて同様の計 算を行った. なお, *E. coli* ネットワーク以外は EcoCyc か ら得たネットワークと同じノードとエッジの数になるよう パラメータを設定した. ノードの除去する順番は,一様ラ ンダムと次数の高い順の2つの場合で行った. なお,ラン ダムなノード除去は10回試行し,その平均をとった.

次に,複数層ネットワークにおける主従関係を考慮した パーコレーション過程の計算を行った.本研究ではこの パーコレーションを区別してマスター・スレーブパーコ レーションと呼称する.

遺伝子制御や外的要因によって,ある酵素反応に関わる 遺伝子発現が全て無くなった場合,その酵素反応は停止す ると考え,酵素反応は遺伝子発現に従属していると捉える. そこで,ネットワークにおいて従属側と考えられる層をス レーブレイヤー,もう一方をマスターレイヤーと定義する. 本研究では,遺伝子制御ネットワーク側をマスターレイ ヤー,代謝ネットワーク側をスレーブレイヤーと考え,層 構造ネットワークにおける主従関係を考慮したパーコレー ションを提案する.

マスター・スレーブパーコレーションでは,遺伝子制御 ネットワークのレイヤーでのみノードの除去する.この過 程で代謝ネットワークノードの外次数が0となった時,そ の代謝化合物ノードを酵素反応ネットワークから除去する. これによって,遺伝子制御ネットワークのノードの残存率 と,代謝関係ネットワークのノードの残存率の関係を考え る.論理モデルでは,A層を遺伝子制御ネットワーク,B 層を代謝関係ネットワークとして同様の計算を行った.こ のマスター・スレーブパーコレーションに関しても,一様 ランダムと次数順で計算を行った.なお,ここでの次数は 内外次数の合計を用いた.

### 3. 結果

図2ネットワーク全体に対するパーコレーションの計算 結果である.縦軸はネットワークの最大クラスタの大きさ *S*,横軸はネットワーク全体の残存ノード数  $\phi$  をそれぞれ 標準化した値である.

ランダムなノードの除去に対しては全てのモデルで同様 の振る舞いを示した.対して,次数の高い順にノードを除 去した場合, *E. coli* と論理モデルではネットワークから ノードを 25%除去した段階で最大クラスタがほぼ消えるこ とが明らかになった.この振る舞いは BA モデルに似てお り,導入した論理モデルが *E. coli* のネットワークの特徴 に似た結果を示した.

マスター・スレーブパーコレーション結果は図3に示す 通りとなった.縦軸にはスレーブ層の最大クラスタの大き さ*S'*,横軸にはマスター層の残存ノード数 ψ をそれぞれ 標準化した値である.ランダムにノードを除去した場合, *E. coli* ネットワークと論理モデルは似た振る舞いを示し, 全体のパーコレーションの様子に似た頑健な振る舞いを示

<sup>1)</sup>各タイムステップにおいて、1個Aノードを追加する



図 2 パーコレーションの計算結果 Fig. 2 Results of percolation process



図 3 マスター・スレーブパーコレーションの計算結果 Fig. 3 Results of master-slave percolation

した.対して,次数順にノードを除去した場合では E. coli ネットワークは図2で得られた振る舞いと異なり,緩やか なネットワークの崩壊を示した.

# 4. まとめと今後の課題

本研究では,独立したネットワークとして扱われていた 複数の実ネットワークを複数層構造のネットワークとして 統合する手法を提唱した.検証のため E. coli のデータを 用いてパーコレーションを用いてネットワークのレジリア ンスを評価し,複数層ネットワーク全体のレジリアンスは, スケールフリーネットワークである BA モデルと同様であ る事を確認した.また,この振る舞いは本研究で導入した 論理モデルでも同様であった.

次に,層構造を考慮したマスター・スレーブパーコレー ションの過程を計算した結果では,*E. coli*のネットワーク は次数に基づく優先的ノードの削除に対して頑健である事 が伺えた.これに対して,我々の導入したモデルではその ような特徴を示さなかった.以上の事から,*E. coli*ネット ワークには2層ネットワークとしてランダム,選択的両障 害に対して頑健性がある事が明らかとなった.

本研究は、複数層ネットワークとして表現した実ネット ワークにおける構造的特徴の一端を明らかにした.構造的 特徴の他にも、パーコレーションに関わる研究として、モ ジュール構造の発見アルゴリズムは重要な課題の1つであ る[12],[13].他に、ネットワーク上での情報拡散やコント ロールといった課題も存在する[14],[15].本手法では複数 のネットワークにまたがる構造的特徴を明らかにできた. よって,これら課題に対しても複数層ネットワークとして のアプローチは有用であり,応用する価値があると考えら れる.

## 参考文献

- Palla, G., Barabási, A.-L. and Vicsek, T.: Quantifying social group evolution., *Nature*, Vol. 446, No. 7136, pp. 664–667 (2007).
- [2] Barabási, A.-L. and Oltvai, Z. N.: Network biology: understanding the cell's functional organization., *Nature reviews. Genetics*, Vol. 5, No. 2, pp. 101–113 (2004).
- [3] Toivonen, R., Onnela, J.-p., Saramäki, J., Hyvönen, J. and Kaski, K.: A model for social networks, *Physica A: Statistical Mechanics and its Applications*, Vol. 371, No. 2, pp. 851–860 (2006).
- [4] Clauset, A., Shalizi, C. R. and Newman, M. E. J.: Power-Law Distributions in Empirical Data, *SIAM Review*, Vol. 51, No. 4, pp. 661–703 (2009).
- [5] Barzel, B. and Barabási, A.-L.: Network link prediction by global silencing of indirect correlations., *Nature biotechnology*, Vol. 31, No. 8, pp. 720–725 (2013).
- [6] Manichaikul, A., Ghamsari, L., Hom, E. F. Y., Lin, C., Murray, R. R., Chang, R. L., Balaji, S., Hao, T., Shen, Y., Chavali, A. K., Thiele, I., Yang, X., Fan, C., Mello, E., Hill, D. E., Vidal, M., Salehi-Ashtiani, K. and Papin, J. A.: Metabolic network analysis integrated with transcript verification for sequenced genomes., *Nature Methods*, Vol. 6, No. 8, pp. 589–592 (2009).
- [7] Vidal, M., Cusick, M. E. and Barabási, A.-L.: Interactome networks and human disease., *Cell*, Vol. 144, No. 6, pp. 986–998 (2011).
- [8] Barabási, A.-L., Gulbahce, N. and Loscalzo, J.: Network medicine: a network-based approach to human disease., *Nature reviews. Genetics*, Vol. 12, No. 1, pp. 56–68 (2011).
- [9] Rozenfeld, H. D., Gallos, L. K. and Makse, H. a.: Explosive percolation in the human protein homology network, *The European Physical Journal B - Condensed Matter*, Vol. 75, No. 3, pp. 305–310 (2010).
- [10] Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñiz-Rascado, L., Bonavides-Martínez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P. and Karp, P. D.: EcoCyc: a comprehensive database of Escherichia coli biology., *Nucleic acids research*, Vol. 39, No. Database issue, pp. D583–90 (2011).
- [11] Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science (New York, N.Y.)*, Vol. 286, No. 5439, pp. 509–512 (1999).
- [12] Fortunato, S.: Community detection in graphs, *Physics Reports*, Vol. 486, No. 3-5, pp. 75–174 (2010).
- [13] Newman, M. E. J.: Communities, modules and largescale structure in networks, *Nature Physics*, Vol. 8, No. 1, pp. 25–31 (2011).
- [14] He, G., Yang, Q., Fu, F. and Kwak, K. S.: Percolation theory aided data diffusion for mobile wireless networks, *ICT Convergence (ICTC)*, 2012 International Conference on, pp. 61–65 (2012).
- [15] Zhang, Z.-K., Zhang, C.-X., Han, X.-P. and Liu, C.: Emergence of Blind Areas in Information Spreading, *arXiv*, p. 4707 (2013).