

E. coli の統合的生体ネットワークにおける数理的モデル化 と解析 (2014年2月7日版)

高橋 勝人^{†1,a)} ホセ ナチエル^{†2}

概要: 細胞内において、遺伝子制御関係から代謝やタンパク質間相互作用まで様々な生命分子が存在し、様々な要素の関係から成る複雑ネットワークが存在している。この複雑性によって生物機能が発現されていると考えられている。これら細胞内ネットワークはネットワーク同士互に影響を及ぼし合っているが、従来の研究ではそれぞれ独立したネットワークである様に扱われていた。そこで、*E. coli* のデータに対し、統合的ネットワークとして扱う手法を考案した。本研究では遺伝子制御関係と代謝化合物関係の2つのネットワークをそれぞれ層として使い、2層ネットワークを構成し、ネットワークとしての性質と機能予測への応用に関して研究を行った。この *E. coli* ネットワークを用いて、ランダムウォークベースアルゴリズムをこのネットワークへ適用し、遺伝子と生物機能の関係における予測能が従来の手法より優れることを示すことも明らかにし、ネットワークを複数層化する事の優位性を示す事ができた。次に、複雑ネットワークにおける代表的な統計的解析を行い、統計的に特徴のある振る舞いを示す事を明らかにした。最後に、この振る舞いを再現する論理モデル及びその振る舞いを説明する数理モデルの提案を行い、モデルが *E. coli* の2層ネットワークの振る舞いを説明しうる事を確認した。

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, 数理モデル, 遺伝子機能予測

Analysis and modeling of interacting networks for *E. coli* and its predictions (version 2014/2/7)

MASATO TAKAHASHI^{†1,a)} JOSE C. NACHER^{†2}

Abstract: In a cell, life molecules are organised in a variety of complex networks, from metabolic pathways to protein-protein interaction and gene regulatory networks, to perform biological functions. However, none of these networks are independent, instead the cellular networks are coupled to each other, and define a complex interaction web of macromolecules that jointly participate in functional tasks. Here, we first integrate experimental data for *E. coli* organism corresponding to metabolic pathways and gene regulatory network, where links between genes and enzymes are determined on whether a gene expresses a given enzyme. Then, the bipartite enzymatic-chemical compound network is projected onto the chemical space, which allows us to define relationships between regulatory genes and chemical compounds. The integrated structure of the real metabolic-gene regulatory network of *E. coli* is then statistically analysed using network science tools. In particular, by using a diffusion-based algorithm that navigate on the integrated network, we predict genes associated to a particular biological process. Finally, a theoretical framework supporting the experimental data analysis is also presented.

Keywords: Complex Network, Bioinformatics, Multiplex Networks, Mathematical Model, Gene function prediction

1. はじめに

複雑な生物ネットワーク、技術的ネットワークそして社会ネットワークなどは、全て要素を表すノードとその関係を示すエッジを用いて、グラフあるいは複雑ネットワークとして表現する事が可能である [1]。近年の計測技術の発達により、大量の臨床データが蓄積され、この情報によってシステムバイオロジーやバイオインフォマティクスといった研究手法が発展してきた。我々はこれらデータの蓄積によって、生命をシステム科学の視点から捉えるようになってきた。代謝反応やタンパク質相互作用といった生体システムは微分方程式系として表現し、この精密なフレームワークで解析されてきている [2], [3], [4]。一方、複雑ネットワークの様に大胆な抽象化を行っても、対象のシステムを解明するには有用であることが明らかとなり、生体ネットワークとしての研究も行われている [5], [6], [7], [8]。

細胞内での相互作用ネットワークは総称してインタラクトームと呼ばれている。特に解析が進んでいる部分は、代謝経路、タンパク質相互作用そして遺伝子制御ネットワークである。これらネットワークは相互に影響しあって複雑な機能を創発している。

近年では統計物理と計算科学的手法によってネットワーク全般における解析とモデル化が進んできた [5], [9], [10], [11], [12], [13]。これまでの研究では、様々なネットワーク情報が蓄積されているにもかかわらず、遺伝子相互作用やタンパク質相互作用といった対象はそれぞれ独立のネットワークとして研究が行われていた。更に、ネットワーク成長モデルの研究においても単層構造の1部グラフで考えられてきた。

そこで、本研究ではネットワークを複数層化を提案し、検証を行った。*E. coli* をモデル生物とし、代謝経路と遺伝子制御の情報を EcoCyc から取得し、2層ネットワークを作成した。この *E. coli* の2層ネットワークと遺伝子制御ネットワークの単層ネットワークを用い、ランダムウォークベースアルゴリズムを適用して遺伝子機能予測能の評価を行った。この実験から複数層ネットワークとして扱う事で単層ネットワークより機能予測能力が高いことを確認した。

次に、複数層ネットワークに対してネットワーク科学で用いられてきた解析手法を応用するため、ノード、エッジ、次数そしてクラスタ構造の定義を拡張した。この定義を用いて *E. coli* のネットワークにおいて統計的に特徴的な振る舞いを示す事を明らかにした。

最後に、*E. coli* の振る舞いを説明しうるモデルを導入

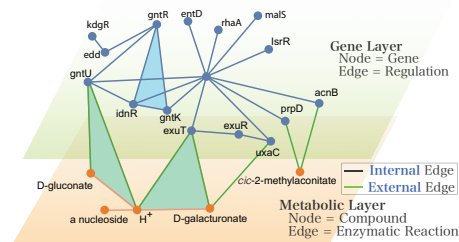


図 1 *E. coli* の複数層ネットワークの部分グラフ
Fig. 1 A sub-graph of *E. coli* multi-layer network

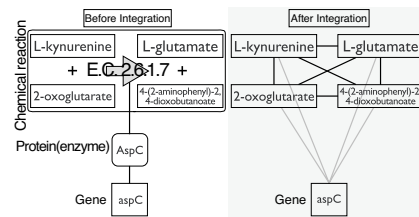


図 2 遺伝子と代謝化合物のデータ統合例
Fig. 2 An example of the integrating of biological data

し、この特徴的振る舞いを再現することを確認した。

2. データの取得と統合

2.1 データセット

我々は、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*) をモデル生物にして研究を行った。*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)[14] 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。この情報から我々は次節に示す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。統合したネットワークは図1に示した様になる。

本研究では遺伝子オントロジー情報 (Gene Ontology)(2013年9月25日取得)[15] を用いて、EcoCyc から得たネットワークを用いて遺伝子機能の予測能力を検証した。GO情報は生物学的プロセス、細胞の構成要素、分子機能の3つの大分類され、これらの分類をさらに細分化したサブカテゴリに分けられている。我々は、生物学的プロセスに着目し、*E. coli* のネットワークに存在する遺伝子を抽出した。この遺伝子群の中で遺伝子が100個以上存在するGO情報のクラスタを10個抽出し、遺伝子機能予測のテスト情報として用いた。

2.2 ネットワークの統合

E. coli の2層ネットワークは次に示す手順で構成した。まず、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

^{†1} 現在、公立はこだて未来大学大学院 システム情報科学専攻
Presently with Future University Hakodate

^{†2} 現在、東邦大学 情報科学科
Presently with Toho University

a) g2112024@fun.ac.jp

物の情報を取得し、EC番号と化合物の2部グラフを作成する。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。図2は遺伝子 aspC における例である。

3. 遺伝子機能予測とその評価方法

我々は、*E. coli* の2層ネットワークを遺伝子機能予測に応用し、その性能を複数層にすることで向上できるか評価した。機能予測には、遺伝子相互作用ネットワークにランダムウォークベースアルゴリズム (RWR) を応用した先行研究 [16] と、RWR アルゴリズムを遺伝子と表現型の2層ネットワークへ適用した先行研究 [17] (RWRH) の2つを使用した。RWRH アルゴリズムでは先行研究にて最も良い予測能を示したとされるパラメータ組を用いて計算を行った。RWR アルゴリズムでは RWRH アルゴリズムと重複する、ランダムウォーカーが初期ノードへ戻る確率パラメータを同じ値として同様に計算した。これら両アルゴリズムにて、定常状態はある時点 t とその直前の時点 $t-1$ での確率ベクトルの L_1 ノルムを計算し、 10^{-6} 以下になった時点とした。この計算には初期ベクトルに既知の機能クラスを指定する必要があるため、前節で紹介した GO 情報を用いて初期ベクトルを設定した。

機能予測結果の検定方法には leave-one-out 交差検証 (LOOCV) を用い、GO 情報のクラスにそれぞれ LOOCV を行った。

最後に、LOOCV の結果に ROC 曲線解析を行い、AUC 値で予測能の比較を行った。ROC 曲線とは、陽性を陽性と判断する割合 (Sensitivity) と、陰性を陰性と判断する割合 (Specificity) の2つの指標をプロットし、その曲線から予測能を評価する手法である。AUC とは ROC 曲線以下の面積の事であり、この値を比較することで機能予測能を比較した。

4. 定義の拡張

本研究では、一般的にグラフ理論で使われる次数とクラスタ構造の定義を拡張し、層の内部と外部を区別して統計的な特徴を調べた。この章では、次数とクラスタ構造の定義の拡張に関して説明する。

4.1 次数

グラフ理論において次数とは、あるノードが他のノードと接続しているエッジの本数の事を指す。本研究では属している層によってノードとエッジを区別する。層の内部を

繋ぐエッジを内エッジとし、各ノードの内エッジの本数のみを数え上げた次数を内次数と定義する。同様に、ある層から別の層を繋ぐエッジを外エッジとし、各ノードの外エッジの本数のみを数え上げた次数を外次数と定義する。

A 層 B 層と区別した場合でこれらの表記の仕方を説明する。A 層の内次数は k^A とし、この内次数分布は $p^A(k^A)$ と表記する。A 層に属するノードに関して、B 層と繋がっている外エッジを数え上げた外次数は l^A となり、この外次数分布は $p^{AB}(l^A)$ と表記する。同様に B 層の外次数は l^B 、外次数分布は $p^{BA}(l^B)$ となる。

4.2 クラスタ係数

クラスタ係数 (狭義では局所クラスタ係数) とは、ある着目ノードがどれだけクラスタ構造に寄与しているかを示す指標である [18]。ここでのクラスタ構造とは、ネットワーク内で3つのノードと3本のエッジで閉じた三角形構造の部分を示す (図1着色した三角形の面)。

本研究では、同じ層のノードだけで構成されるクラスタ構造と、複数の層に属するノードで構成されるクラスタ構造を区別して扱う。従って、A 層または B 層に属するノードだけで構成されているクラスタ構造を内クラスタと定義し、A 層と B 層の両方のノードから構成されたクラスタ構造を外クラスタと定義する。

次数分布と同様に A 層 B 層の2層で説明する。内クラスタをにおけるクラスタ係数を内クラスタ係数と呼ぶ。あるノード i に関して A 層の内クラスタ係数は C_i^A 、B 層内は C_i^B と表す。この内クラスタ係数は、単層構造で扱っていた一般的にグラフ理論で扱われるクラスタ係数の定義と等しい。

外クラスタ係数の場合、A 層に1つのノードを、B 層に残り2つのノードを有する外クラスタを C_i^{AB} 、逆に B 層に1つ、A 層に2つノードが属する外クラスタの場合は C_i^{BA} と表記する。

以上の様に次数とクラスタ構造の定義を複数層ネットワークに向けて拡張した、これらの定義を用いて *E. coli* のネットワーク上で統計的振る舞いが得られるか計算した。

5. *E. coli* ネットワークのモデル化

EcoCyc から得たネットワークに関し、外次数分布及び外クラスタ係数分布を計算した結果、統計的に特徴のある振る舞いを示す事が明らかとなった。そこで我々はこの振る舞いを説明する論理モデルと、そこから得られる数理モデルを作成した。

5.1 ネットワーク成長モデル

我々は2層ネットワークにおける *E. coli* の次数分布、クラスタ係数分布の特徴的振る舞いを考慮した論理モデルを、実際に得られた結果を説明するため導入した。ネット

ワークの成長規則は AB 層を用いて次の通りである。

A 層側の成長規則.

- (1) 各タイムステップにおいて, 1 個 A 層のノードを追加する
- (2) 既にネットワークに存在する m_f 個の A 層の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 m_f 個のノードと接続しているノードから, m_s 個の A 層のノードと m_c 個の B 層の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 層のノードを追加する
- (2) ノードを追加した場合, m_a 個の A 層のノードと m_b 個の B ノードをランダムにそれぞれ選択する
- (3) ランダムに選択したノードを新しく追加した B 層のノードと接続する

以上の成長規則は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される。

5.2 数理モデル

E. coli に関する論理モデルから, 我々は数理モデルを導入した. この節では導入した数理モデルの説明をする.

5.2.1 内次数分布

まず, k_i^A を第 i 番目の A ノードの内次数であるとする. 導入した成長モデルから, この次数は次の微分方程式として表現することができる.

$$\frac{dk_i^A}{dt} = m_f m_s \frac{k_i^A}{\sum_i k_i^A} + \frac{m_f}{N^A(t)} \quad (1)$$

ここで, $N^A(t)$ とは時刻 t における A ノードの数であり, モデルから $N^A(t) = t$ となる. 第 i ノードがネットワークへ追加された時点 t_i にて, 初期条件を $k_i^A(t_i) = m_f(1 + m_s)$ と置いてこの微分方程式を解くと, 解は次の様になる.

$$k_i^A = m_f(1 + m_s) \left(1 + \frac{2}{m_s}\right) \left(\frac{t}{t_i}\right)^{\frac{m_s}{2(1+m_s)}} - \frac{2m_f(1 + m_s)}{m_s} \quad (2)$$

この解から, 次数分布は

$$P^A(k^A) = \left. -\frac{1}{t} \frac{dt^A}{dk_i^A} \right|_{k_i^A=k^A} \propto (k^A + C)^{-3 - \frac{2}{m_s}} \quad (3)$$

として得られる. ここで C は任意定数である.

5.2.2 外次数分布

次に, 本研究で導入した外次数の数理モデルを導入する. 成長モデルより, A 層の外次数は次の微分方程式を用いて

表すことができる.

$$\frac{dl_i^A}{dt} = \frac{m_a}{N^A(t)} \quad (4)$$

初期条件を $l_i^A(t_i) = m_f m_c$ とすると, この微分方程式は

$$l_i^A(t) = m_a \log \frac{t}{t_i} + m_f m_c \quad (5)$$

として解くことができ, 外次数分布は内次数分布と同様な手続きで

$$P^{AB}(l^A) \propto \exp\left(-\frac{l^A - m_f m_c}{m_a}\right) \quad (6)$$

として得られる.

B 層の外次数分布に関しても, 成長モデルから微分方程式を立てると

$$\frac{dl_i^B}{dt} = m_f m_c \frac{l_i^B}{\sum_i l_i^B} \quad (7)$$

として表す事ができる. ここで, $\sum_i l_i^B = m_f m_c t$ となる.

初期条件を $l_i^B(t_i) = \beta m_a$ とすると, この微分方程式は

$$l_i^B(t) = \beta m_a \left(\frac{t}{t_i}\right) \quad (8)$$

となり, 次数分布はこの解から

$$P^{BA}(l^B) \propto (l^B)^{-2} \quad (9)$$

として得られる.

5.2.3 クラスタ係数分布

まず内クラスタ係数分布を考える. A 層の内クラスタ係数分布の数理モデルを考えるため, E_i^A を i 番目のノードが寄与しているクラスタ構造の数とする. 成長モデルより, このクラスタの数は

$$\frac{dE_i^A}{dt} = m_f m_s \frac{k_i^A}{\sum_i k_i^A} + \frac{m_f m_s}{t} \quad (10)$$

と表すことが出来る. この微分方程式の解は C, C', C'' をそれぞれ任意定数として

$$E_i^A(t) = k_i^A + m_f(m_s - 1) \frac{1}{C} \log(k_i^A + C') + C'' \quad (11)$$

となり, クラスタ係数分布は

$$C_i^A \propto (k_i^A)^{-1} \quad (12)$$

となる.

次に, 外クラスタ係数分布の数理モデルを導入する. 外クラスタ構造の数は成長モデルから

$$\frac{dE_i^{BA}}{dt} = m_f m_s \frac{l_i^B}{\sum_i l_i^B} \quad (13)$$

となり, この微分方程式の解は

$$E_i^{BA} = l_i^B - \beta m_a \quad (14)$$

となる. よって, 外クラスタ係数分布は

$$C_i^{BA} \propto (l_i^B)^{-1} \quad (15)$$

となる.

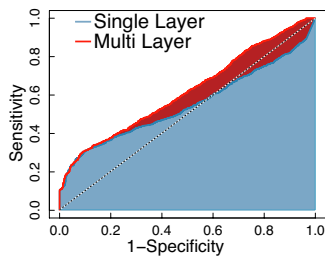


図 3 GO:0008152 を用いた ROC 曲線の結果

Fig. 3 A result of ROC curve analysis with using GO:0008152

表 1 AUC 値一覧

Table 1 The table of AUC values

GO term	#Gene	AUC	
		Single	Multi
GO:0008152	955	0.558	0.614
GO:0005975	135	0.673	0.691
GO:0006807	544	0.553	0.626
GO:0006082	176	0.640	0.681
GO:0006725	430	0.529	0.596
GO:0006793	223	0.597	0.656
GO:0018130	198	0.536	0.550
GO:0034641	441	0.519	0.571
GO:0044248	183	0.607	0.603
GO:0090304	253	0.529	0.520

6. *E. coli* ネットワークを用いた結果

6.1 遺伝子機能分離能

図 3 は GO:0008152 の遺伝子セットを用いて LOOCV を行い、ROC 解析を適用した結果である。従来の単層ネットワークで遺伝子機能予測を行った結果が青い実線部分、複数層ネットワークの結果が赤い実線部分である。複数層ネットワークを用いたほうが予測精度が高いことが示された。表 1 は、10 種類の GO 情報クラスタを用いて ROC 解析を行い、それぞれ AUC を求めた結果である。AUC 値は IEEE 754 に従い小数第 4 位で丸めを行った。10 セット中 8 セットにおいて、複数層ネットワークで解析を行う方が精度の良い結果であることが示された。AUC 値で劣った 2 セットの結果においても、従来の方法とほぼ同じ精度での予測性能であった。

6.2 度数分布

図 4 は EcoCyc から得た *E. coli* ネットワークの外度数分布、インセットは累積外度数分布である。実線部分はそれぞれ回帰分析を行って得られた結果である。図 4(b) では冪乗則を示すことがわかったため、Aaron 等による Power-Law fitting[19] を用いて冪指数と冪乗則が適用される最小次数を求めた。グラフ中の k_{min} はスケールを開始する最小次数、 σ は標準誤差を示す。これら 2 つの値は Power-Law fitting を適用時に得られる。

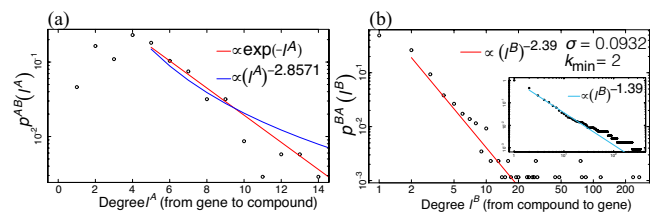


図 4 EcoCyc のデータを用いた実ネットワークの度数分布

Fig. 4 Degree distributions using EcoCyc dataset

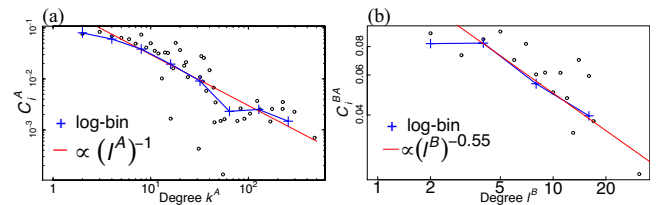


図 5 EcoCyc のデータから得られたクラスタ係数分布

Fig. 5 Clustering coefficient distributions using EcoCyc dataset

E. coli の 2 層ネットワークでは、外度数分布において非対称かつ非ランダムな度数分布を示すことが明らかとなった。遺伝子ノードでの外度数分布では指数分布を、対して、化合物ノードでは冪指数分布を示した。遺伝子ノードにおける外度数分布を評価するため非線形二乗法を用いた結果、指数関数へデータを回帰した結果、 p 値は $7.11 \cdot 10^{-5}$ となった。

6.3 クラスタ係数分布

クラスタ係数分布は図 5 に示す通りの結果となった。内クラスタ係数分布は既知の通り冪乗則を示すことを確認した。外クラスタ係数分布においても、内クラスタ係数分布と同様に冪乗則を示すことが明らかとなった。

なお、クラスタ係数分布において回帰分析を行うにはノイズが多かったため、log-bin を行いデータを平滑化した後のデータ (図 5 青ドット部) に非線形二乗法で回帰を行った。回帰での p 値は C_i^A では $6.35 \cdot 10^{-5}$ 、 C_i^{BA} では $4.85 \cdot 10^{-4}$ となった。

7. モデル化による結果

7.1 度数分布

図 6 は論理モデルから計算して得られた度数分布の結果である。パラメータ組は $(m_a, m_b, m_c, m_f, m_s) = (1, 1, 1, 1, 2)$ とした。ネットワークの大きさは $N = 10^3, 10^4, 10^5$ とて、それぞれ 100 個ネットワークを生成し、各ネットワークから得られた度数分布を平均したて得られた結果である。実線部分は数理モデルから予測されるモデルの振る舞いを示す。我々の導入したモデルは、内外両方の度数分布に関して *E. coli* の振る舞いをよく説明する振る舞いである事が確認できた。

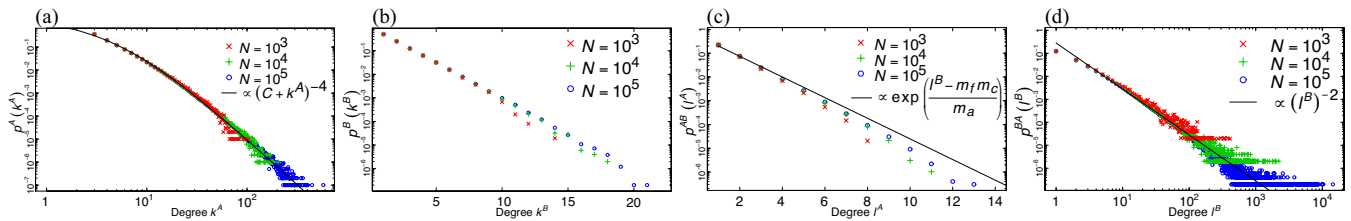


図 6 論理モデルから得られた次数分布

Fig. 6 Degree distributions obtained from theoretical model

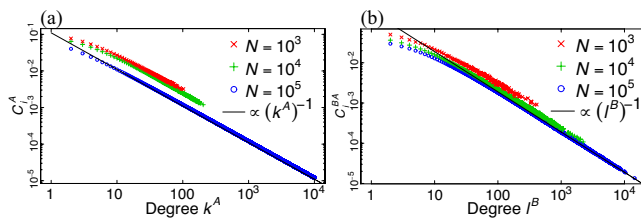


図 7 論理モデルから得られたクラスタ係数分布

Fig. 7 Clustering coefficient distributions obtained from theoretical model

7.2 クラスタ係数分布

クラスタ係数分布の結果は図 7 に示した通りである。パラメータは次数分布の計算と同じである。クラスタ係数分布においても数理モデル通りの振る舞いを示すことが確認された。

以上の論理モデルのシミュレーション結果から、我々の提案した論理モデルは内外次数分布やクラスタ係数分布で *E. coli* の結果と同様の特徴を示すモデルが提示できたと主張できる。

8. まとめと今後の課題

従来の研究では、遺伝子制御ネットワークや代謝物ネットワークはそれぞれ独立なネットワークとして扱われていた。しかし、実際の生体ネットワークは各ネットワークが相互に影響し合い機能を発現している。そこで我々はこの点を考慮して、複数層ネットワークとして扱う手法を提案した。本研究では *E. coli* をモデル生物とし、遺伝子制御ネットワークと代謝物ネットワークをそれぞれ層として捉え、2層ネットワークとして機能予測の性能とネットワークとしての統計的解析を行った。

まず、*E. coli* の実データから得たトポロジを用いて遺伝子機能予測の性能評価を行った。従来の遺伝子制御ネットワークの層だけで予測を行うより、複数層ネットワークの方が高い予測能を示すことを確認した。

次に、従来のネットワーク理論で定義されていない、層構造を考慮したノードとエッジそして次数やクラスタ構造の定義を行い、ネットワークの数理的表現方法を提供した。この拡張した次数とクラスタの概念を *E. coli* のネットワークへ適用し、外次数分布に関して非対称かつ非均一な構造の振る舞いを示すことが明らかとなった。クラスタ

係数分布の計算を行った結果、内外どちらのクラスタ係数分布においても冪乗則を示すことが確認できた。

最後に、これら実データから得られた結果を説明しうる数理モデルを導入し、シミュレーションを行った。モデルの振る舞いは *E. coli* の内外の次数分布とクラスタ係数分布をどちらも良く説明しうる振る舞いを示すことが確認できた。

本研究では、*E. coli* を次数分布とクラスタ係数分布において、説明しうる数理モデルが導入できた。しかし、より実ネットワークを正確に記述できるモデルの導入が可能ではないか考える必要がある。次に、現在、他の有機体のコネクトーム情報が蓄積されているため、本手法を適用し、統計的に特徴的な振る舞いをするか研究をする必要がある。

最後に、本研究では次数とクラスタ係数に関して定義の拡張と適用を行ったが、他の複雑ネットワークにおける定石的手法の応用はできなかった。特にネットワークにおける制御とレジリエンスの評価は、生体システムの制御や新薬創世や遺伝子治療といった応用といった対象には有用である。従って、これらのフレームワークを複数層ネットワーク上で応用できるよう研究を行う必要がある。

参考文献

- [1] Albert, R. and Barabási, A.-L.: Statistical mechanics of complex networks, *Reviews of Modern Physics*, Vol. 74, No. 1, pp. 47–97 (2002).
- [2] Krauss, M., Schaller, S., Borchers, S., Findeisen, R., Lippert, J. and Kuepfer, L.: Integrating Cellular Metabolism into a Multiscale Whole-Body Model, *PLoS computational biology*, Vol. 8, No. 10, p. e1002750 (2012).
- [3] Bellouquid, A. and Delitala, M.: *Mathematical modeling of complex biological systems*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Boston Inc., Boston, MA (2006).
- [4] Deutsch, A., Bruschi, L., Byrne, H., de Vries, G. and Herzog, H.: *Mathematical Modeling of Biological Systems, Volume I*, Cellular Biophysics, Regulatory Networks, Development, Biomedicine, and Data Analysis, Vol. I, Springer (2007).
- [5] 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).
- [6] Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M. and Barabási, A.-L.: The human disease network., *PNAS*, Vol. 104, No. 21, pp. 8685–8690 (2007).

- [7] Yldrm, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L. and Vidal, M.: Drug—target network, *Nature biotechnology*, Vol. 25, No. 10, pp. 1119–1126 (2007).
- [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] Pastor-Satorras, R. and Vespignani, A.: Epidemic Spreading in Scale-Free Networks, *Physical review letters*, Vol. 86, No. 14, pp. 3200–3203 (2001).
- [10] Cohen, R., Erez, K., ben Avraham, D. and Havlin, S.: Resilience of the internet to random breakdowns, *Physical review letters*, Vol. 85, No. 21, pp. 4626–4628 (2000).
- [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] 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).
- [13] 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).
- [14] 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).
- [15] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium., *Nature genetics*, Vol. 25, No. 1, pp. 25–29 (2000).
- [16] Köhler, S., Bauer, S., Horn, D. and Robinson, P. N.: Walking the interactome for prioritization of candidate disease genes., *American journal of human genetics*, Vol. 82, No. 4, pp. 949–958 (2008).
- [17] Li, Y. and Li, J.: Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data., *BMC genomics*, Vol. 13 Suppl 7, p. S27 (2012).
- [18] Burt, R. S.: *STRUCTURAL HOLES: The Social Structure of Competition*, Harvard University Press (1995).
- [19] 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).