

# アンチバイオグラムの菌株同定グラフからの同一菌株推定

長山 和樹<sup>1,a)</sup> 江藤 賢太<sup>1,b)</sup> 原口 尚也<sup>1,c)</sup> 平田 耕一<sup>2,d)</sup>

**概要:** 本論文では、菌に対する薬剤の感受性情報であるアンチバイオグラムから、菌株を同定するための菌株同定グラフを定式化する。そして、菌株同定グラフの独立集合を菌株の代表元の集合とみなすことで、同一菌株を推定する。そのために、まず、大阪府立急性期総合医療センターから提供された薬剤感受性検査データのアンチバイオグラムから、菌株同定グラフを構築する。そして、独立集合を求める近似アルゴリズムを実装して菌株同定グラフに適用し、同一菌株を近似推定する。さらに、最大独立集合を求める厳密指数時間アルゴリズムを実装し、実際の菌株同定グラフから同一菌株を厳密に推定するために必要な計算時間について議論する。

## Inferring Bacterial Culture from Identification Graph of Bacteria for Antibioqram

KAZUKI NAGAYAMA<sup>1,a)</sup> KENTA ETO<sup>1,b)</sup> NAOYA HARAGUCHI<sup>1,c)</sup> KOUICHI HIRATA<sup>2,d)</sup>

**Abstract:** In this paper, we formulate an identification graph of bacteria for antibiogram, and infer the bacterial culture by regarding an independent set of the identification graph as a set of representative bacterial cultures. First, we construct identification graphs of bacteria for antibiogram in bacterial culture data provided from Osaka Prefectural General Medical Center. Then, by applying a simple greedy algorithm to find an independent set to identification graphs, we infer the bacterial culture approximately. Furthermore, by implementing three exact exponential time algorithms to find the maximum independent set, we investigate the computational time of them.

### 1. はじめに

現在社会問題となっている院内感染の予防対策の一つとして、薬剤感受性検査データの利用が挙げられる。この中の薬剤の感受性情報の記録をアンチバイオグラム(antibiogram)という。このアンチバイオグラムは、通常は菌名ごとに記録されているが、その菌が同一菌株か否かは、DNA 検査などを経なければ一般には判断できない。一方で、院内感染

予防のための医療施設の科、病棟、病室などの特定の場所における同一菌の蔓延の判断や医療従事者による同一菌の伝搬の解析では、同一菌内の同一菌株を把握することで初めて意味が出てくる。したがって、同一菌株の同定は、院内感染予防のために非常に重要な問題である。

そこで、DNA 検査の代替手法として、本論文では、2次元キャリアマップ(2-dimensional carrier map, 2DCM) [2] で用いられている S(感受性)、I(中間)、R(耐性) の SIR 値からの同一菌株の同定に着目する。ここで、SIR 値の S と I、および、I と R の違いは、臨床検査では誤差範囲とされている値の違いであり、したがって、I は S の可能性と R の可能性がある。

このようなアンチバイオグラムの SIR 値に基づく菌株の同定について、例えば、図 1 の SIR 値を考える [2]。ここで、PIPC, GM, CPFY はそれぞれ薬剤の名称である。また、一番右の欄は、菌株のグループの可能性である。

<sup>1</sup> 九州工業大学情報工学部  
Undergraduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Kawazu 680-4, Iizuka 820-5502, Japan  
<sup>2</sup> 九州工業大学情報工学研究院  
Department of Artificial Intelligence, Kyushu Institute of Technology, Kawazu 680-4, Iizuka 820-5502, Japan  
a) nagayama@dumbo.ai.kyutech.ac.jp  
b) eto@dumbo.ai.kyutech.ac.jp  
c) haraguchi@dumbo.ai.kyutech.ac.jp  
d) hirata@ai.kyutech.ac.jp

表 1 アンチバイオグラムによる菌株の同定例 [2].

菌株	PIPC	GM	CPFX	
1	R	I	I	1
2	S	S	I	2,4
3	S	I	R	2,3
4	S	R	R	3
5	S	S	S	4

図 1 の場合, 菌株 1 は PIPC で R であり他菌株が S なので, 他株と別の菌株と考えることができる. 一方, 菌株 2 はグループ 2 として菌株 3 と同じかもしれない, 同時にグループ 3 として菌株 4 と同じかもしれない. 菌株 3 は同時にグループ 3 の菌株 4 と同じかもしれない. なお, 図 1 は 3 薬剤に対する菌株の同定例に過ぎないが, 実際のアンチバイオグラムからの菌株同定はこれに比べてより複雑となる.

この複雑さを回避するために, 菌株の同定では, I を独立したカテゴリとして扱ったり, 共通に検査されている薬剤だけで同定を行う方法が用いられてきた. 一方, 上述したように, S の可能性も R の可能性もある I を独立して扱うことは SIR 値の意味から不適切であり, また, 共通に検査される薬剤だけを用いる場合は 2 菌株で検査されなかった薬剤が S と R である可能性もある.

そこで 2DCM [2] では, 2 つの菌株の間で, もし 1 薬剤でも S と R の組合せがある場合は別の菌株とし, それ以外の場合で, 閾値数以上の薬剤に対して一致がある場合には「別の菌株とは言えず, 同じ菌株である可能性が高い」と判断する菌株の同定方法を採用している.

そこで本論文では, まず, それぞれの菌ごとにこの同定関係をグラフ  $G = (V, E)$  として表現した菌株同定グラフ (identification graph of bacteria) を導入する. つまり,  $V$  はある菌のアンチバイオグラムの集合であり, アンチバイオグラム  $u$  と  $v$  に対して, 一方が S でもう一方が R となる薬剤が存在しないとき,  $(u, v) \in E$  としたグラフである. 表 1 のアンチバイオグラムの菌株同定グラフは図 1 となる.

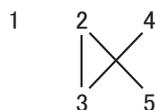


図 1 表 1 の菌株同定グラフ.

本論文では, 実際の薬剤感受性検査データとして, 2001 年から 2007 年の大阪府立急性期総合医療センターから提供された薬剤感受性検査データを用いる. このデータは 164713 レコードからなっており, 102 の薬剤に対するアンチバイオグラム (ただし, すべての薬剤に SIR 値が記録されているわけではなく, 空欄もある) が, 344 種類の菌と共に記録されている. そこで本論文では, まず, このアンチバイオグラムから菌株同定グラフを求める.

菌株同定グラフにおいて, 独立集合 (independent set) を菌株の代表元の集合とみなすことができる. したがって, 菌

株同定グラフから要素数が最大の最大独立集合 (maximum independent set) を求めることができれば, 菌株をすべて求めることができる. 例えば, 図 1 に対して,  $\{1\}$ ,  $\{2, 4\}$  などが独立集合であり,  $\{1, 2, 4\}$ ,  $\{1, 3, 5\}$ ,  $\{1, 4, 5\}$  が最大独立集合である. しかしながら, グラフから最大独立集合を求める問題は NP 困難である [4].

菌株同定グラフの独立集合を求めるために, 本論文では, まず, 独立集合を求める単純な貪欲近似アルゴリズム [1], [6] を実装する. そして, 実際のアンチバイオグラムに適用することで同一菌株を近似推定する.

次に, 菌株同定グラフの最大独立集合を求めるために, 3 種類の厳密指数時間アルゴリズム [3] を実装する. そして, それぞれのアルゴリズムによって実際の菌株同定グラフから同一菌株を厳密に推定するために必要な計算時間について議論する.

## 2. アンチバイオグラムの菌株同定グラフ

グラフ  $G = (V, E)$  に対して,  $v \in V$  に接続する辺の数を  $v$  の次数 (degree) といい  $d(v)$  と表す. また,  $\Delta(G) = \max\{d(v) \mid v \in V\}$  とし, これを  $G$  の次数 (degree) という. さらに,  $\bar{\Delta}(G) = \frac{1}{|V|} \sum_{v \in V} d(v)$  とし, これを  $G$  の平均次数 (average degree) という.

$v \in V$  に対して,  $v$  に隣接する頂点の集合を  $v$  の近傍 (neighbor) といい  $N(v)$  と表す. また,  $N[v] = \{v\} \cup N(v)$  を  $v$  の閉近傍 (closed neighbor) という.

グラフ  $G = (V, E)$  と  $S \subseteq V$  に対して,  $S$  に含まれる頂点, および,  $S$  に接続する辺をすべて除いて得られるグラフを  $G \setminus S$  と表す. すなわち,  $G \setminus S = (V \setminus S, E \setminus \{(u, v) \in E \mid u \in S\})$  である.

グラフ  $G = (V, E)$  と  $S \subseteq V$  に対して, 任意の  $u, v \in S$  に対して  $(u, v) \notin E$  となるとき,  $S$  を  $G$  の独立集合 (independent set) という. 要素数が最大となる  $G$  の独立集合を  $G$  の最大独立集合 (maximum independent set) といい, その要素数を  $\alpha(G)$  と表す. なお, グラフから最大独立集合を求める問題は NP 困難である [4].

本論文では, 菌  $A$  に対して,  $V$  を菌  $A$  のアンチバイオグラムの集合であり, アンチバイオグラム  $u$  と  $v$  に対して, 一方が S でもう一方が R となる薬剤が存在しないときに  $(u, v) \in E$  として得られるグラフ  $G = (V, E)$  を, 菌  $A$  の菌株同定グラフ (identification graph of an bacterium  $A$ ) として定式化する.

2001 年から 2007 年の大阪府立急性期総合医療センターから提供された薬剤感受性検査データの 344 種類の菌のうち, レコード数 (#) が 10 以上となる菌は 144 種類, 100 以上となる菌は 50 種類, 300 以上となる菌は表 2 の 22 種類である.

まず, 全く同一のアンチバイオグラムを同一視する. この結果を菌株同定グラフの頂点  $V$  とする. そして,  $u, v \in V$

表 2 レコード数 300 以上の菌に対する菌株同定グラフ

菌名	#	V	E	$\Delta$
Staphylococcus aureus	9778	892	169096	589
Pseudomonas aeruginosa	4052	1038	140773	868
Escherichia coli	2651	440	39157	342
Enterococcus faecalis	2330	303	15279	276
Klebsiella pneumoniae	1697	132	3587	113
Staphylococcus epidermidis	1629	528	70192	387
Enterobacter cloacae	1378	346	18697	310
Candida albicans	996	18	47	10
Staphylococcus coagulase	995	479	56863	313
Streptococcus pneumoniae	872	409	13527	405
Serratia marcescens	750	166	5853	137
Stenotrophomonas maltophilia	738	408	16283	388
Enterococcus faecium	700	195	3974	188
Streptococcus agalactiae	562	74	1391	72
Candida glabrata	497	12	25	8
Acinetobacter calcoaceticus	465	191	6455	181
Klebsiella oxytoca	428	75	646	70
Enterobacter aerogenes	339	139	3697	130
Proteus mirabilis	316	59	607	51
Haemophilus influenzae	312	105	1830	103
Streptococcus intermedius	304	130	881	129
Bacteroides fragilis	300	75	456	37

に対して,  $u$  と  $v$  のアンチバイオグラムを比較することで  $(u, v) \in E$  となるか否かを判定して辺  $E$  を求める. さらに, 得られた菌株同定グラフ  $G = (V, E)$  とその次数  $\Delta$  を表 2 にまとめている.

図 2 は, 表 2 で頂点数 20 以下となる *Candida albicans* と *Candida glabrata* 菌株同定グラフである. このグラフから, *Candida albicans* と *Candida glabrata* における各頂点の次数をまとめると表 3 のようになる.

表 3 *Candida albicans* と *Candida glabrata* の各頂点の次数

<i>Candida albicans</i>								
次数	2	3	4	5	6	7	8	10
頂点	214	99	1	378	15	297	42	38
		150	4	490	26		317	
		201	20	166				
			213	228				

<i>Candida glabrata</i>							
次数	1	2	3	4	5	6	8
頂点	492	3	18	1	4	6	53
			59	97	88	176	
			77				

また, 図 3 は, *Candida albicans* と *Candida glabrata* 以外で頂点数 100 以下, 辺数 1000 以下となる *Bacteroides fragilis*, *Proteus mirabilis*, *Klebsiella oxytoca* の菌株同定グラフの概形である. なお, 図 2 と 3 のグラフは, Graphviz [5] の書式に変換して描画している.

### 3. 同一菌株の近似推定

菌株同定グラフにおいて, 独立集合を菌株の代表元の集

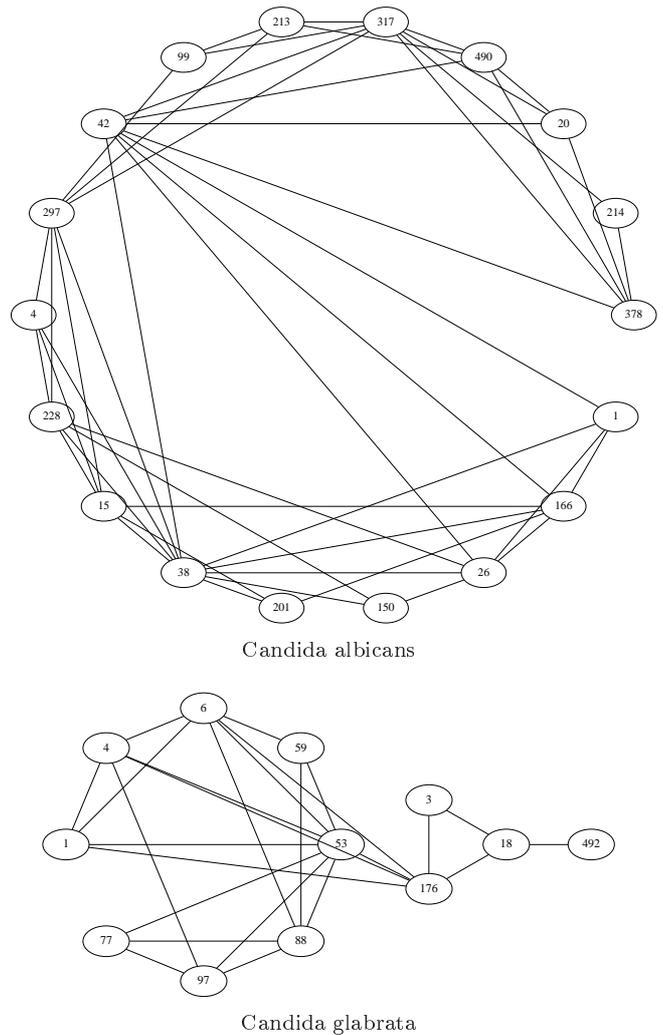


図 2 *Candida albicans* と *Candida glabrata* の菌株同定グラフ

合とみなすことができる. したがって, 菌株同定グラフから最大独立集合を求めることができれば, 菌株をすべて求めることができる. しかしながら, グラフから最大独立集合を求める問題は NP 困難である [4].

そこで本節では, グラフ  $G$  に対して  $|I| \geq |V|/(\bar{\Delta}(G)+1)$  となる独立集合  $I$  を出力する単純な貪欲アルゴリズム [1], [6] であるアルゴリズム 1 の Greedy を利用する.

```

procedure Greedy( $G$ )
    /*  $G = (V, E)$  : グラフ */
1    $I \leftarrow \emptyset$ ;
2   while  $G \neq \emptyset$  do
3       select 最小次数の  $v \in V$ ;
4        $I \leftarrow I \cup \{v\}$ ;  $G \leftarrow G \setminus N[v]$ ;
5   return  $I$ ;
    
```

アルゴリズム 1: Greedy

例えば, 図 2 の *Candida albicans* と *Candida glabrata* の菌株同定グラフからアルゴリズム Greedy によって得られ

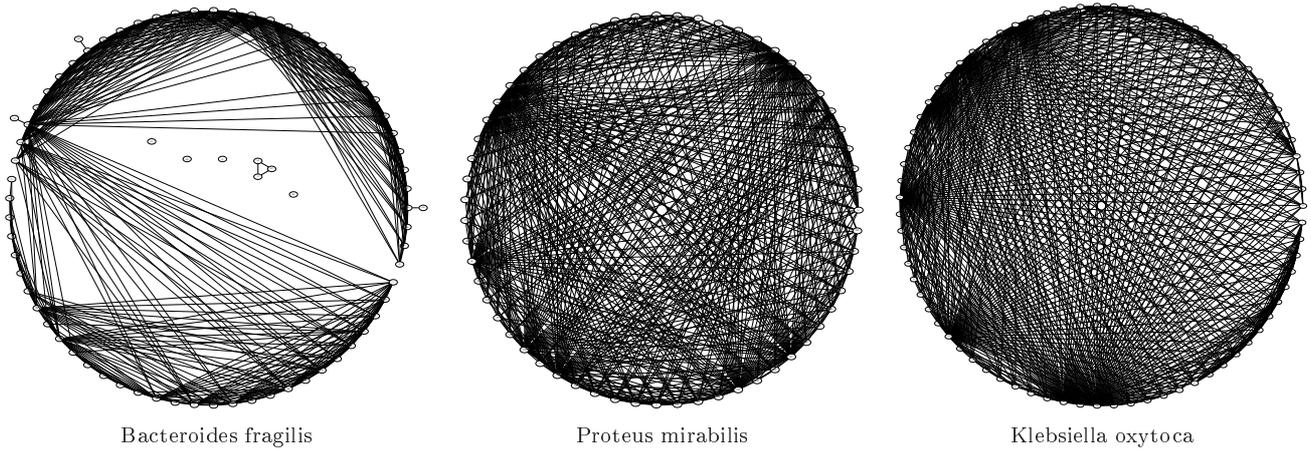


図 3 Bacteroides fragilis, Proteus mirabilis, Klebsiella oxytoca の菌株同定グラフの概形

る独立集合は、3 行目で選ばれる頂点の順に表 4 のようになる。

表 4 Candida albicans と Candida glabrata の独立集合

Candida albicans	214, 99, 150, 201, 1, 4, 20
Candida glabrata	492, 3, 59, 77, 1

表 5 は、菌株同定グラフの頂点数が 100 以上の菌とそのグラフからアルゴリズム Greedy によって得られる独立集合の要素数 ( $|I|$ ) とそのときの下限  $lb = |V|/(\Delta(G) + 1)$  である。なお、CPU は Intel(R) Core(TM) i7-3770 3.40GHz, RAM は 8GB, OS は Windows 7 Home Premium SP1 という計算機環境の下で、すべての菌株同定グラフから独立集合を求める計算時間は 1.750 秒である。

表 5 より、下限値よりも大きな要素数の独立集合が求まっていることが分かる。それぞれの菌は少なくとも  $|I|$  個以上の異なる菌株を含むことになる。

#### 4. 同一菌株の厳密推定

本節では、以下の 3 種類の最大独立集合の要素数を求める厳密指数時間アルゴリズム [3] を実装する。頂点数  $n$  に対して、アルゴリズム 2, 3, 4 は、それぞれ  $O(3^{n/3} \text{poly}(n))$  時間,  $O(1.2905^n)$  時間,  $O(1.3426^n)$  時間で  $G$  の最大独立集合を求めるアルゴリズムである。ここで、次数 2 のグラフ  $G$  に対する  $\alpha(G)$  はアルゴリズム Greedy を用いて多項式時間で求めることができる。

```

procedure mis1( $G$ )
  /*  $G = (V, E)$  : グラフ */
  1 if  $|V| = 0$  then
  2   | return 0;
  3 select 最小次数の  $v \in V$ ;
  4 return  $1 + \max\{\text{mis1}(G \setminus N[v]) \mid y \in N[v]\}$ ;

```

アルゴリズム 2: mis1

表 5 頂点数 100 以上の菌株同定グラフと独立集合の要素数

菌名	$ V $	$ E $	$ I $	$lb$
Pseudomonas aeruginosa	1038	140793	157	3.81
Staphylococcus aureus	892	169096	133	2.35
Staphylococcus epidermidis	528	70192	137	1.98
Staphylococcus coagulase (-)	479	56863	86	2.01
Escherichia coli	440	39157	113	2.46
Streptococcus pneumoniae	409	13527	113	6.09
Stenotrophomonas maltophilia	408	16283	76	5.05
Enterobacter cloacae	346	18697	51	3.17
Enterococcus faecalis	303	15279	56	2.97
Enterococcus faecium	195	3974	62	4.67
Acinetobacter calcoaceticus	191	6455	35	2.78
Serratia marcescens	166	5853	47	2.32
Enterobacter aerogenes	139	3697	19	2.56
Enterococcus avium	139	2690	49	3.50
Klebsiella pneumoniae	132	3587	37	2.38
Streptococcus intermedius	130	881	42	8.93
Haemophilus influenzae II	106	1285	29	4.20
Haemophilus influenzae	105	1830	32	2.93

```

procedure mis2( $G$ )
  /*  $G = (V, E)$  : グラフ */
  1 if  $\exists v$  s.t.  $d(v) = 0$  then
  2   | return  $1 + \text{mis2}(G \setminus \{v\})$ ;
  3 if  $\exists v$  s.t.  $d(v) = 1$  then
  4   | return  $1 + \text{mis2}(G \setminus N[v])$ ;
  5 if  $\Delta(G) \geq 3$  then
  6   | select 最大次数の  $v \in V$ ;
  7   | return  $\max\{1 + \text{mis2}(G \setminus N[v]), \text{mis2}(G \setminus \{v\})\}$ ;
  8 if  $\Delta(G) \leq 2$  then
  9   | return  $\alpha(G)$ ;

```

アルゴリズム 3: mis2

実験では、大阪府立急性期総合医療センターから提供された薬剤感受性検査データのうち、頂点数が 120 以下の

```
procedure mis3(G)
  /* G = (V, E) : グラフ */
1  if  $\Delta(G) \geq 3$  then
2    select 最大次数の  $v \in V$ ; return
   |  $\max\{1 + \text{mis3}(G \setminus N[v]), \text{mis3}(G \setminus \{v\})\}$ ;
3  if  $\Delta(G) \leq 2$  then
4    return  $\alpha(G)$ ;
```

アルゴリズム 4: mis3

データ 298 種類に着目する。図 4 は、そのような菌株同定グラフに対して、頂点数が 10 ごとに該当するグラフすべての最大独立集合を計算し、その計算時間を該当グラフの数で割って平均を求めたものである。縦軸は 100 ナノ秒の対数である。計算機環境は 3 節と同一である。

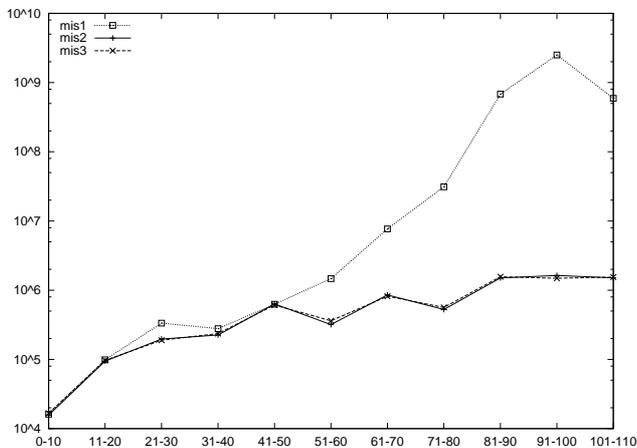


図 4 各アルゴリズムの実行時間

その結果、mis1 が最も遅く、mis2 と mis3 は同程度の速さで計算時間は指数的には緩やかに増大している。

## 5. まとめと今後の課題

本論文では、アンチバイオグラムの菌株同定グラフを定式化し、実際のアンチバイオグラムから菌株同定グラフを構築した。そして、菌株の代表元の集合を菌株同定グラフの独立集合とみなし、独立集合の近似アルゴリズムと最大独立集合の厳密指数時間アルゴリズム実装することで、独立集合として同一菌株を推定した。

本論文の厳密指数時間アルゴリズムでは、最大独立集合を見つけるために必要となる計算時間が膨大となるため、本論文では頂点数が多い菌株同定グラフについては最大独立集合を計算することができなかった。そこで、それらの菌株同定グラフからの最大独立集合の計算、特に、本論文の設定におけるヒューリスティックの導入や他の厳密アルゴリズムの利用などが今後の課題として挙げられる。また、それにより菌株同定グラフにおけるアルゴリズム Greedy の精度を解析し、その結果を基に、同一菌株推定におけるア

ルゴリズム Greedy での代用方法を考案することも今後の課題である。

最大独立集合は一般には一意ではなく、複数存在する。よって、最大独立集合の選び方によって、同一菌株の選び方も変わってくる。今後は、どのような(最大)独立集合が同一菌株の同定に必要なかという選択基準を考察する必要がある。特に、同一視したアンチバイオグラムの数によって頂点に重み付けした、重み付き独立集合の意義も考慮しながら、独立集合の選択基準を考察することは今後の課題である。

また、本論文の菌株同定グラフは重みなしグラフを用いたが、同一のアンチバイオグラムの数を辺の重みとすることで、重み付き菌株同定グラフを導入することもできる。この場合、S と R の同時出現がないアンチバイオグラム間の辺で重み 0 のものも存在する。このことを考慮しつつ、重み付き菌株同定グラフを利用した同一菌株推定も今後の課題である。

最後に、1 節で述べた通り、自発的には動かない菌に対して、医療施設の科、病棟、病室などの特定の場所における同一菌株の蔓延の判断や医療従事者を介した同一菌株の伝搬を求めることで、院内感染予防の原因を明らかにすることができる。したがって、本研究で取り組んだアンチバイオグラムからの菌株同定は、同一菌株の蔓延や伝搬において、まず最初に取り組むべき問題である。したがって、今後の重要な課題として、医療施設内の同一菌株の蔓延や伝搬の解析が挙げられる。特に、同一菌株の伝搬について、軌跡マイニング手法を利用することで解析する。

謝辞 本論文の動機となった菌株同定問題を提供いただいた、大阪府立急性期・総合医療センターの松岡喜美子氏とケーディーアイコンズ(株)の横山茂樹氏に深謝する。

## 参考文献

- [1] R. Boppana, M. M. Halldórsson: *Approximating maximum independent sets by excluding subgraphs*, BIT **32**, 180–196, 1992.
- [2] 藤本修平: アンチバイオグラムの自動分類と 2 次元キャリアマップ (2DCM) の概要, 2009, <https://www.nih-janis.jp/2dcm/2dcmwebtraining.html>.
- [3] F. V. Fomin, D. Kratsch: *Exact exponential algorithms*, Springer-Verlag, 2010.
- [4] M. R. Garey, D. S. Johnson: *Computers and intractability - A guide to the theory of NP-completeness*, W. H. Freeman & Company, 1979.
- [5] Graphviz, <http://www.graphviz.org>
- [6] M. M. Halldórsson, J. Radhakrishnan: *Greed is good: Approximating independent sets in sparse and bounded-degree graphs*, Algorithmica **18**, 145–163, 1997.