

グラフの特性量のデータベース (Graphical Characteristics Data Base : GCDB)の開発 —Multi-Layered Cyclic Fence Graph と特性量—

林 佳代子^{a,c}, 福田 詩織里^{a,c}, 長嶋 雲兵^{b,d*}, 細矢 治夫^a

^a お茶の水女子大学理学部情報科学科 〒112-8610 文京区大塚 2-1-1

^b 物質工学工業技術研究所基礎部理論化学研究室 〒305-8565 つくば市東 1-1

*e-mail: umpei@nair.go.jp

^c 現所属: NTT 株式会社

^d 現所属: 産業技術融合領域研究所

(Received: February 18, 1999 ; Accepted for publication: June 30, 1999 ; Published on Web: August 13, 1999)

グラフの特性量を集めたデータベース (Graphical Characteristics Data Base : GCDB) を開発した。GCDB にはグラフの名前、頂点数、頂点の次数、辺の数、隣接行列、トポロジカルインデックス、ケクレ数、非隣接数、 Z' 、マッチング多項式、距離多項式、 Sz 、ウィーナー数、 n 歩数、固有値という 16 種類の特性量とコメントおよびグラフの図が、蓄えられている。ユーザーインターフェイスは、WWW を用いる。

現在 Multi-Layered Cyclic Fence Graph 約 330 とその特性量をデータベース化し、お茶の水女子大学において仮公開を開始した。

(<http://www.sap.is.ocha.ac.jp/~hayasi/cgi-bin/gcdb.cgi>)

キーワード: Graphical Characteristics, Data Base, Multi-Layered Cyclic Fence Graph

1 はじめに

化学の構造式に代表されるように、分子の化学的性質の多くはその分子構造に関係付けられている。もともと化学はトポロジカルな情報の記述を基本的表現形式として発展してきたといえる。構造式を一つの数学的表現として考えると、それはグラフであり、一つのグラフは、さまざまな数学的的特性量を持つことが知られている [1]。

グラフとは、いくつかの点と、それらの間を結ぶいくつかの辺とで構成される図形、またはそのような図形で表示される対象物のことであり、LadderGraph, Cyclic Fence Graph (CFG)[2], Multi-Layered Cyclic Fence Graph (MLCFG)[3] などいろいろな系列がある。しかし、有限個の点と辺をもつグラフの種類も「組合せ的爆発」によって膨大な数に膨れ上がっていく。グラ

フ理論や情報科学の分野では、このような沢山のグラフを識別するために、グラフの形状を比較するだけではなく、各グラフの特徴ともいえる特性量を計算して比較することが多い。これまで、特性量の比較によって、グラフ間の相関関係が数多く見つけられてきたが、現状において特性量に関するデータは各グラフごとに別々のファイルに保存されており、その比較は全て手作業で行なわれている。よって、グラフの様々な種類の特性量を集めたデータベースの作成は、多くのグラフを機械的に体系付けることが出来る点において大きな意義をもつ。

本研究で取り上げた MLCFG は、CFG を拡張したものとして、近年定義されたグラフの系列である。六角形を敷き詰めたグラフと同相であることから、グラフアイト格子とも関連付けられると考えられる等の化学的な特徴だけでなく、特性多項式やマッチング多項式において漸化関係が見られる。また固有値が全て整数である整数グラフとなるものもあり、グラフ理論的にも興味深い特徴が挙げられ、現在研究がすすめられている。MLCFG には、沢山の種類があるが、その中でも代表的な $E_{m,n}, F_{m,n}$ は、 m と n の組合せによって非常に多くのグラフがあり、互いに同相なものが複雑に絡み合っている。

そこでわれわれはグラフの特性量を蓄積し公開するデータベース GCDB(Graphical Characteristics Data Base) を作成した。今回公開する GCDB では、特に MLCFG : $E_{m,n}, F_{m,n}$ とそれに関連するグラフの間の関係を系統的に調べるために、327 種類のグラフの特性量をデータベース化した。現在お茶の水女子大学において、試験公開を開始したので概要を報告する。
(<http://www.sap.is.ocha.ac.jp/~hayasi/cgi-bin/gcdb.cgi>)

2 Multi-layered Cyclic Fence Graph(MLCFG)

MLCFG : $*_{m,n}$ は、 $2n$ 個の頂点からなる cycle graph C_{2n} を縦に m 個並べて cycle graph の層を m 個作り、その層の間にある規則で結ぶことによって得られる。($*$ には cycle graph の層の結び方によって E、F などのアルファベットが入る。) MLCFG としてはいろいろな種類が考えられるが、本研究では、主にその中でも代表的なものである $E_{m,n}$ 、 $F_{m,n}$ と $G_{m,n}$ を取り扱う。Figure 1 はその例である。

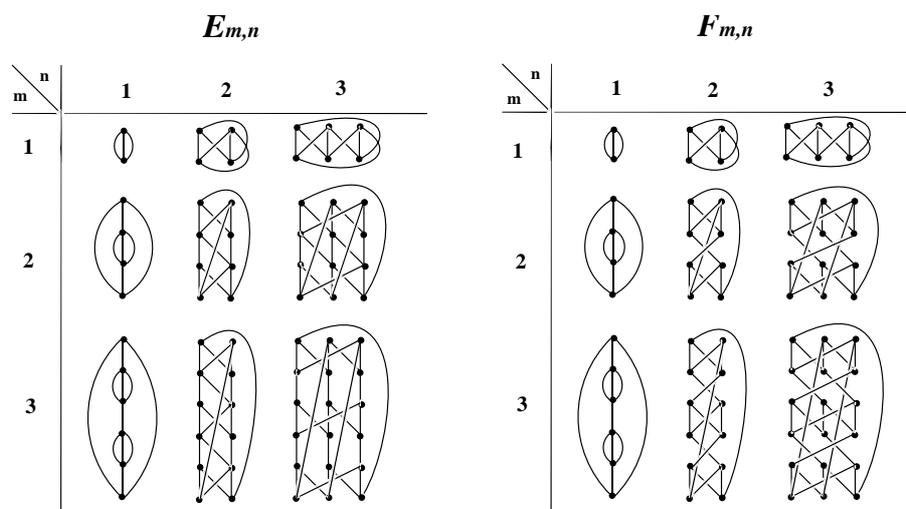


Figure 1. Multi-layered Cyclic Fence Graphs: $E_{m,n}, F_{m,n}$

MLCFG の特徴として、

1. 各頂点の次数が 3、つまり cubic である。
2. 2 部グラフである。
3. どの辺も交わずにドーナツの表面を覆うことができるトーラスグラフである。
4. 対称性が非常に高い。
5. 特性多項式や距離多項式において漸化関係がみられる。
6. 固有値が全て整数の整数グラフとなるものもある。
7. 六角形を敷き詰めたグラフと同相であり、グラフアイト格子と関連付けられる。

などが挙げられ、現在研究がすすめられている。

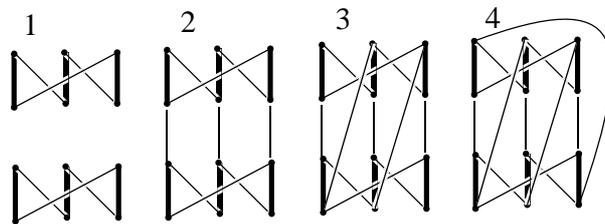


Figure 2. $E_{2,3}$ を得る手順。太線は、グラフ構造を識別しやすくするために用いており、細線との違いはない。

$E_{m,n}$

$E_{m,n}$ は、次のような手順で得られる。Figure 2 において、 $E_{2,3}$ を例に挙げて手順を紹介する。

1. cycle graph C_{2n} (図では C_6 を、Figure 2 の様に縦に m 個 (図では 2 個) 並べて層を作る。
2. 各層の下端の頂点と、その一つ下層の cycle graph の上端の頂点とを結ぶ。
3. 最下層の下端の頂点と、その右側に隣接する列の上端の頂点とを結ぶ。
4. 最上層の最左列の上端の頂点と最下層の最右列の下端の頂点とを結ぶ。

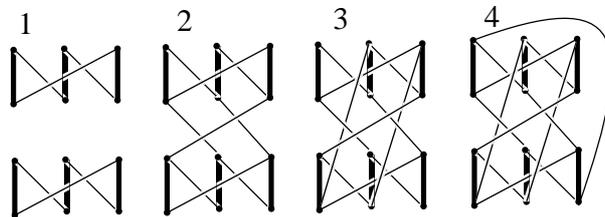


Figure 3. $F_{2,3}$ を得る手順。太線は、グラフ構造を識別しやすくするために用いており、細線との違いはない。

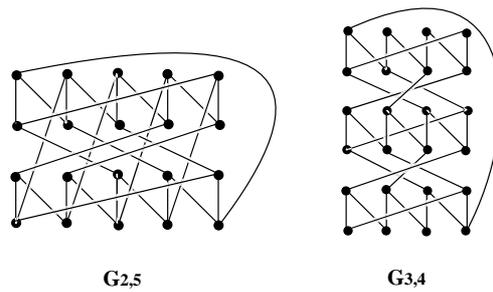


Figure 4. $G_{m,n}$ の例

$F_{m,n}$

$F_{m,n}$ の求め方は以下のとおりである。(Figure 3) $E_{m,n}$ を得る際の手順 2 において、各層の下端の頂点とその一つ下層の右側に隣接する列の上端の頂点とを結ぶ。次に、最左列の下端の頂点と最右列の上端の頂点とを結ぶ。他の手順は、 $E_{m,n}$ と同様にして得られる。

$G_{m,n}$

$G_{m,n}$ は $F_{m,n}$ の手順を更に一つ右側にずらしたものの、即ちで層のつなぎ方を 2 つずつずらして描いたグラフである。Figure 4 にその例を示す。

他に MLCFG としては、 $D_{m,n}$ 、 $X_{m,n}$ [3] が知られているが、ここではふれない。

3 グラフの特性量

グラフの相関関係を系統的に解析する手段として、グラフの特性量 [4] は非常に有効である。GCDB で取り上げた特性量はグラフの名前、頂点数、頂点の次数、辺の数、隣接行列、トポロジカルインデックス、ケクレ数、非隣接数、 Z' 、マッチング多項式、距離多項式、 Sz 、ウィーナー数、 n 歩数、固有値という 16 種類である。以下に GCDB で取り上げたグラフの特性量について簡単に説明する。また、ここで N はグラフの頂点数を表す。

隣接行列： A グラフ G に対して隣接行列は、次のような要素を持つ正方行列として定義される。

$$a_{ij} = \begin{cases} 1 & (\text{頂点 } i \text{ と } j \text{ が辺でつながれているとき}) \\ 0 & (\text{それ以外するとき}) \end{cases}$$

例えば、Figure 1 に示した $E_{1,3}$ の隣接行列は、

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

となる。

非隣接数： $p(G,k)$ グラフ G に対して、非隣接数は、どの辺も互いに隣接しないように k 本の辺を選ぶ選び方の数と定義する。 $k=0,1,\dots,N/2$ であり、 $p(G,0)=1$ と定義される。

例えば、 $E_{1,3}$ の非隣接数は、

$$p(G,0) = 1, p(G,1) = 9, p(G,2) = 18, p(G,3) = 6$$

となる。

ケクレ数： K ケクレ数は、グラフ G における N 個全ての頂点を、互いに隣接しない $N/2$ 本の線で覆い尽くす方法の数を指す。非隣接数における $p(G,N/2)$ と同値である。

例えば、 $E_{1,3}$ のケクレ数は、 $K=p(G,3)=6$ となる。

トポロジカルインデックス： Z 非隣接数の総和をトポロジカルインデックスとする。定義は以下のとおりである。

$$Z_G = \sum_{k=0}^{N/2} p(G,k)$$

特性多項式： $P_G(X)$ グラフ G の特性多項式は、次のように定義される。これは、隣接行列 A の対角項に $-X$ を代入した行列の行列式である。

$$P_G(X) = (-1)^N \det(A - XE) = \sum_{k=0}^N a_k X^{N-k} \quad (E : N \times N \text{ の単位行列})$$

Z' Z' は、特性多項式の係数列 $a_k (k=0,\dots,N)$ を用いて次のように定義される。

$$Z' = \sum_{k=0}^{N/2} (-1)^k a_{2k}$$

距離多項式： $S_G(X)$ グラフ G の距離多項式は距離行列 D ($[D]_{ij} = i$ から j までの最短距離の辺の数) を用いて、次のように定義される。

$$S_G(X) = (-1)^N \det(D - XE) = \sum_{k=0}^N b_k X^{N-k} \quad (E : N \times N \text{ の単位行列})$$

S_z 距離多項式の係数列 $b_k (k=0,\dots,N)$ を用いて、次のように定義される。

$$S_z = \sum_{k=2}^N |b_k|$$

ウィーナー数： W ウィーナー数は、距離行列 D の非対角要素の総和の半分であり、次のように定義される。

$$W = \sum_{i < j}^N D_{ij} = \sum_{k=1}^d k p_k$$

(p_k : k 本の辺によって隔てられた頂点の対の数, d : p_k の最大値)

固有値： X 固有値は、特性多項式 $P_G(X) = 0$ の一連の解を指す。

$$\{X\} = \{X \mid P_G(X) = 0\}$$

以上が GCDB の蓄積されたグラフの特性量である。次節で GCDB の概要と使用例について説明する。

4 Graphical Characteristics Data Base (GCDB) の概要と使用例

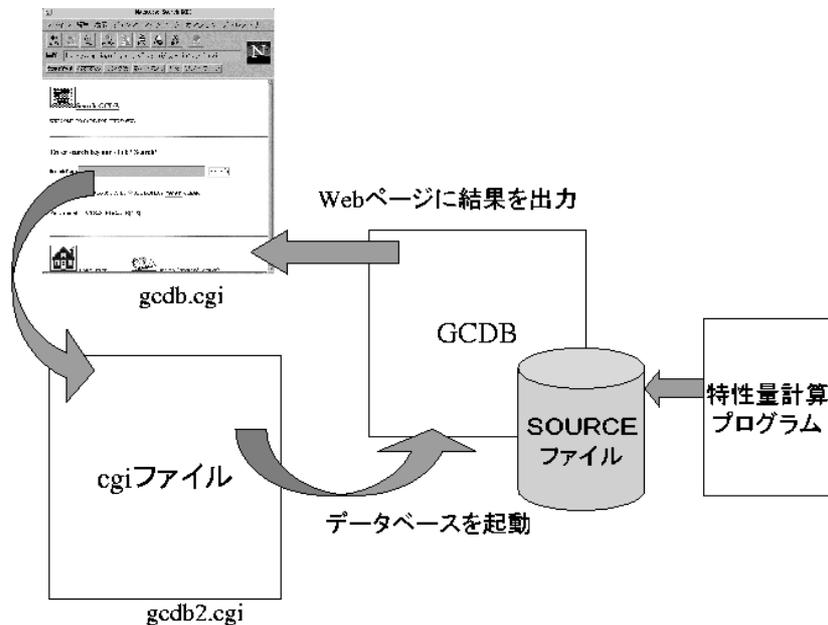


Figure 5. GCDB のシステム構成

4.1 GCDB の構成

GCDB のシステム構成を Figure 5 に示す。GCDB は、WWW ブラウザとそれから GCDB 本体にデータを引き渡す CGI スクリプトと GCDB 本体からなる。データは基本的に SOURCE ファイルに蓄積されるが、GCDB の setup コマンドによりインデックスファイル等高速検索のためのファイルが自動生成される。SOURCE ファイルは、特性量計算プログラムが計算したグラフの特性量がある程度蓄積されると、そのデータを取り込みデータベースの再構築を行なう。現在この部分は人手を介せねばならず、自動化されていない。

GCDB は、データベースの構造設計段階でのフィールド長の設定が不要で、表示の際にデータ長やデータの個数に合わせてフィールド長を自動的に調節する。表示は表形式ではなく、データとデータをカンマ(,)で区切って表示するため、無駄なスペースをとることを避けることができる。

4.2 データ項目

今回公開する GCDB の対象となるグラフは、MLCFG とそれに関連するグラフに限っている。ただし、GCDB 自体の構造は、他のグラフの特性量に関するデータも管理することができる構造をとっている。GCDB は以下の 16 項目のデータを管理している。

1. グラフの名前	(NAME)	9. マッチング多項式	(P(PR,CF))
2. 頂点数	(VTX)	10. Z-dash	(Z')
3. 辺の数	(EDG)	11. ウィーナー数	(W)
4. 各頂点の次数	(ODR)	12. n 歩数	(pn)
5. 隣接行列	(A_MTRX)	13. 距離多項式	(S(PR,CF))
6. 非隣接数	(k,p(G,k))	14. Sz	(Sz)
7. ケクレ数	(K)	15. 固有値	(X)
8. Topological Index	(Z)	16. コメント	(CMNT)

カッコ内の表示は GCDB で実際に使われている項目名である。このうち、1 ~ 5 は各グラフを決定するものであり、6 ~ 15 は各グラフの特性量である。コメントは、各グラフに関する特徴、同値なグラフなどの特記事項を挙げるための項目である。また、データのとれていない項目については、NULL を認める。

4.3 検索コマンド

Table 1. GCDB の検索コマンド一覧

コマンド	コマンドの内容
SEARCH・FIND	条件を満たすデータの検索
DISPLAY	検索データの表示
LIST	使用可能なパラメータを調べる
LOOK	あるフィールドの指定キーワード登録を調べ、インデックス表示
SET	パラメータ値の設定
REMINDE	現在の検索履歴の表示
END・QUIT	検索を終了
WHAT	各コマンドの説明
HOW	各コマンドの使用法の説明

GCDB で利用できる検索コマンドを Table 1 に示す。検索コマンドの詳細は WHAT または HOW を検索するか、後述する How to keyword search 等をクリックすることで得ることができる。

```

\begin{center}
E[1_3]/6/9/[1__3],[2__3],[3__3],[4__3],[5__3],[6__3]/[1__6],[1__2],[1__4
*],[2__3],[2__5],[3__4],[3__6],[4__5],[5__6]/[0__1],[1__9],[2__18],[3__6
*]/6/34/[6__1.0],[5__0.0],[4__-9.0],[3__0.0],[2__0.0],[1__0.0],[0__0.0]/
*10.0/21/p1=9,p2=6,p3=0,p4=0,p5=0,p6=0/[6__1.0],[5__0.0],[4__-33.0],[3__
*-104.0],[2__-72.0],[1__96.0],[0__112.0]/3.000000,0.000000,0.000000,0.0
*00000,0.000000,-3.000000/=F[1_3]=H[6_1_-1]=H[6_3_-3],INT_GRAPH\
E[1_4]/8/12/[1__3],[2__3],[3__3],[4__3],[5__3],[6__3],[7__3],[8__3]/[1__
*8],[1__2],[1__4],[2__3],[2__7],[3__4],[3__6],[4__5],[5__6],[5__8],[6__7
*],[7__8],/ .....
\end{center}

```

Figure 6. SOURCE ファイルの入力例

4.4 SOURCE ファイル

SOURCE ファイルでは、全てのグラフに関するデータを管理する。SOURCE ファイルにおいて、データ以外に区切り記号として、カンマ(,) スラッシュ(/) バックスラッシュ(\) @、アスタリスク(*)を用いる。カンマは、同一フィールド内での各データを、スラッシュは一つのグラフにおける各フィールドを区切る役割をする。バックスラッシュは一つのグラフに関するデータ列の最後に入れる。@は全てのデータ列の最後、つまり SOURCE ファイルの最後に入れる。

また、SOURCE ファイルには、一行に半角 80 字で入力をする。グラフ名に関するデータは必ず行頭から入力する。それに続くデータは 80 字ずつ改行し、行頭には*を置く。Figure 6 に実際の入力例(抜粋)を紹介する。

このように GCDB の基本的なデータは SOURCE ファイルにテキスト形式で蓄積されているため、少量のデータ修正はエディタで行なうことができる。

データの追加は、今回登録したデータ生成プログラムから自動的に SOURCE に追加することができる。

大規模なグラフの特性量の計算には、莫大な計算時間を必要とするので、並列分散処理を有効に利用した特性量計算システムの開発が必要である。

4.5 GCDB の検索例

Figure 7 に GCDB のフロントページ(検索待ち受け画面)と検索例を示した。

Search key 欄に F[2.3] などの Key word を入力し、Search ボタンをクリックすると、HIT COUNT が表示される。充分検索件数を絞り込んだ後に、ALL DISPLAY ボタンをクリックし、同様な key を入力して Search ボタンをクリックすると検索された情報全てが表示される。この時、名前(この例では、F[2.3])の所をクリックするとグラフが表示される。(Figure 8)

GCDB 自体の説明は、画面の Search GCDB 欄をクリックすると表示され、使用法等は画面の How to keyword Search 欄をクリックすることで得られる。画面をクリアするには、reset ボタンをクリックすればよい。

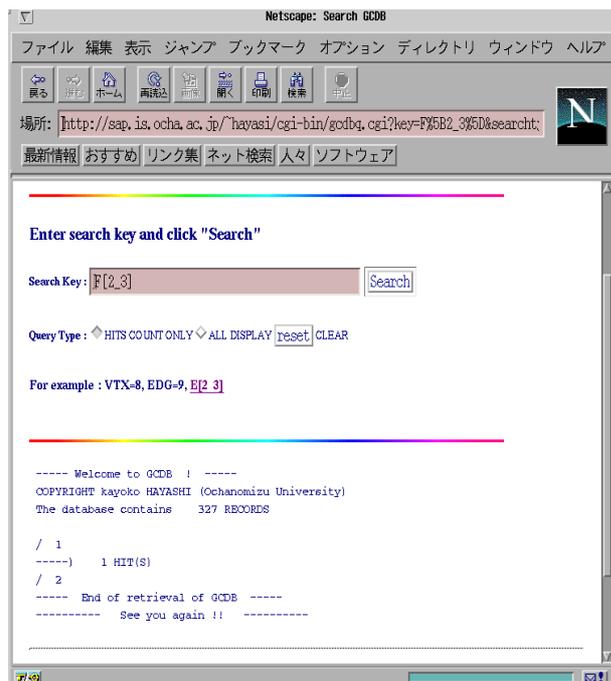


Figure 7. ヒットカウント表示

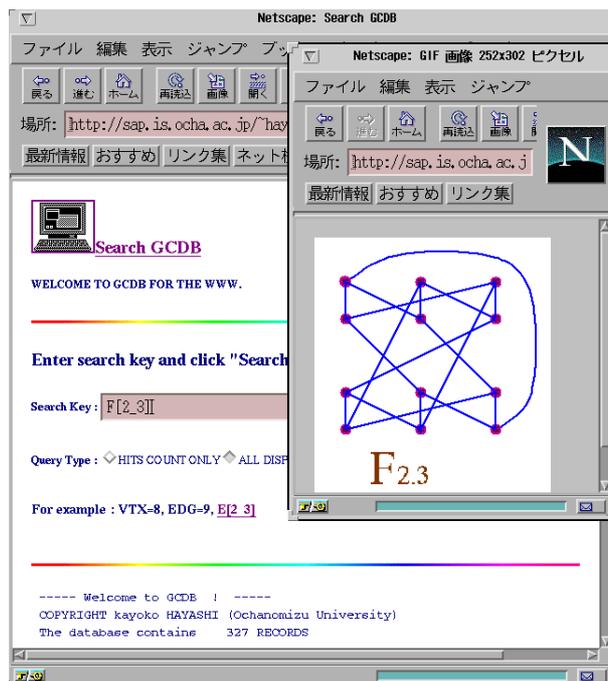


Figure 8. グラフの図形表示

5 まとめと今後の課題

MLCFGのグラフの特性量を集めたデータベース GCDB を開発し試験的公開を開始した。グラフの特性量は、グラフが決まるとプログラムを用いて計算することができるが、その計算時間はグラフのサイズの増加に伴い組合せ的に増大する。そのためそれらをデータベースとして蓄積・公開することは意義が大きい。

今後は、更に多様なグラフの特性量を登録していく必要がある。今回公開する GCDB では、ヒット件数が 0 の場合は、そのまま 0 を報告するが、今後はそのような場合、自動的に特性量計算プログラムを起動して、結果を蓄積するといった自己増殖的なデータベースとして発展させていく必要がある。そのためにはグラフの特性量の自動生成法と計算された特性量の自動登録法の開発が必須である。

また、グラフの同相関係探索ツールなど GCDB を用いた応用ツールの開発も行なっていく。

参考文献

- [1] 細矢 治夫, ラフ理論の化学物理への応用, 固体物理, Vol.32, No.7, 659-556(1997), Vol.32, No.8, 629-639(1997), Vol.32, No.9, 721-728(1997), Vol.32, No.10, 801-810(1997), Vol.32, No.11, 884-893(1997), Vol.32, No.12, 965-973(1997), Vol.33, No.1, 10-18(1998), Vol.33, No.3, 181-189(1998), Vol.33, No.6, 523-531(1998), Vol.33, No.7, 603-607(1998), Vol.33, No.9, 757-764(1998), Vol.33, No.11, 901-908(1998).
- [2] H. Hosoya, and F. Harary, On the Matching Properties of Three Fence Graphs, *J. Math. Chem.*, **12**, 211-218 (1993).

- [3] H. Hosoya, Y. Okuma, Y. Tsukano, and K. Nakada, Multi-layered Cyclic Fence Graphs: Novel Cubic Graphs Related to the Graphite Network, *J. Chem. Inf. Comput. Sci.*, **35**, 351-356 (1995).
- [4] M. Omote, M. Wakamatsu, S. Hyugaji, U. Nagashima, and H. Hosoya, Tables of Topological Characteristics of Polyhedral Graphs with Eight Vertices, *Natl. Sci. Rept. Ochanomizu Univ*, **45** (1994).
- [5] H. Hosoya, Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Japan*, **44**, 2332-2339 (1971).

Development of Graphical Characteristics Data Base : GCDB –Graphical Characteristics of Multi-Layered Cyclic Fence Graph–

Kayoko HAYASHI^{a,c}, Shiori FUKUDA^{a,c}, Umpei NAGASHIMA^{b*} and Haruo HOSOYA^a

^aDepartment of Information Sciences, Ochanomizu University
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, JAPAN

^bTheoretical Chemistry Lab., Department of Physical Chemistry, National Institute of Materials and Chemical Research
1-1 Higashi, Tukuba, Ibaraki 305-8565, JAPAN

*e-mail: umpei@nair.go.jp

^cPresent address: NTT Co.Ltd

We have developed and opened Graphical Characteristics Data Base (GCDB) at Ochanomizu University (<http://www.sap.is.ocha.ac.jp/~hayasi/cgi-bin/gcdb.cgi>). 16 kinds of graphical characteristics of 330 multi-layered cyclic fence graphs are included in this version of GCDB.

Details of graphical characteristics, system configuration and some examples are described in this paper.

Keywords: Graphical Characteristics, Data Base, Multi-Layered Cyclic Fence Graph