Elemente der Mathematik 80 (2025), 26–31
DOI 10.4171/EM/520

Short note A graph called *Harmony*

Carol T. Zamfirescu

1 Introduction

We present a fortunate coincidence between language, mathematics, and computer science. The mathematical objects we shall discuss are graphs: a *graph* is a pair of sets, one containing entities, the other establishing which of these entities, considered in pairs, are in some sense related. An intuitive example might be a group of people at a party and the relationship "be friends with." In their mathematical abstraction, the entities are typically called *vertices* and each of the related pairs of vertices is called an *edge*. For a thorough introduction to graph theory, we refer to Diestel's textbook [2].

Brendan McKay's software package "nauty" is a widely used computer program to generate sets of graphs, usually constrained for certain structural properties—in the absence of constraints, their number grows unwieldy very quickly. For example, there are 165 091 172 592 graphs on 12 vertices; see [6] (this is one of many illustrations of the phenomenon known as the *combinatorial explosion*). In fact, various sets of graphs have already been generated and are available in nauty's `graph6` format, downloadable for instance from the "House of Graphs" [1], a database of notable graphs accessible at https://houseofgraphs.org/. It is this format that we are here interested in.

McKay explains his `graph6` format on his website [3] (for details on "nauty" and the graph isomorphism problem from the practical point of view, we refer to [4,5]), but for the reader's convenience, in Section 2, we briefly explain how it works. The `graph6` format is for storing graphs in a compact manner, using only printable ASCII characters—these include (but are not limited to) the letters in the English alphabet, in both lower and upper case. For instance, the question mark "?" is represented in the (decimal) ASCII encoding by 63, "@" by 64, the upper case letters "A" to "Z" by 65 to 90, and the lower case letters "a" to "z" by 97 to 122.

While the vast majority of graphs have `graph6` strings such as `IheA@GUAo` (this yields *Petersen's graph*, a graph famous among combinatorialists), there exist graphs whose `graph6` strings are identical to words in the English language. The purpose of this piece is *not* to describe a computer program generating the intersection of the set of all `graph6` strings and a list of all words in the English language; such a program itself would be rather dull. However, its output, which is non-empty, might be of a certain appeal—and that is what we here focus on: first a non-trivial sample from the aforementioned intersection is presented, after which a characterisation of *all* words that appear as `graph6` strings of graphs is given (making the above program redundant).

2 The `graph6` format

We now briefly describe the `graph6` format, used to encode simple (i.e., containing neither loops nor multiple edges), undirected graphs with at most $2^{36} - 1 = 68\,719\,476\,735$ vertices. We shall here see bit vectors as finite strings of zeros and ones; we will sometimes read (always from left to right) these as binary numbers, so the bit vector 001101 may be read as the binary number 1101. The number of elements in a bit vector will be called its *length*, so 001101 has length 6. For a bit vector v, we will require the following three functions. We define $R(v)$ as the bit vector obtained after adding zeros to the right of v until the length of v is a multiple of 6, and $L_{18}(v)$ ($L_{36}(v)$) is the bit vector obtained after adding zeros to the left of v until the length of v is 18 (the length of v is 36).

For an integer n in decimal form, let $b(n)$ be its binary form. Consider an integer n with $0 \leq n \leq 2^{36} - 1$. Let $N(n)$ be the string of integers, each between 63 and 126, defined as follows.

- If $n \leq 62$, set $N(n) := n + 63$.
- If $63 \leq n \leq 258\,047$, split $L_{18}(b(n))$ into contiguous substrings of length 6 and convert each substring into decimal form. Add 63 to each number. Concatenate from the left with 126 to obtain $N(n)$.
- Else, split $L_{36}(b(n))$ into contiguous substrings of length 6 and convert each substring into decimal form. Add 63 to each number. Concatenate from the left with 126 126 to obtain $N(n)$.

For instance, for Euler's year of birth 1707, we have $b(1707) = 11010101011$, and thus $L_{18}(b(1707)) = 000000011010101011$. Converting contiguous substrings of length 6 back to decimal form, we obtain 0 26 43. Adding 63 to each, we obtain 63 89 106. Executing the final step yields $N(1707) = 126\ 63\ 89\ 106$.

Let G be a graph on n vertices with $0 \leq n \leq 2^{36} - 1$. For $n = 0$, the `graph6` representation of G shall be the symbol `?`. Henceforth, let $n > 0$. We label the vertices of G by $0, \dots, n-1$ and let $A(G) = (a_{ij})_{0 \leq i,\, j \leq n-1}$ be the adjacency matrix of G, i.e., the $n \times n$ matrix such that its element a_{ij} is 1 when there is an edge between i and j, and 0 when there is no edge. We are interested in the lower left part of $A(G)$, more precisely all elements a_{ij} with $i \in \{1, \dots, n-1\}$ and $j \in \{0, \dots, n-2\}$ satisfying $i \geq j + 1$, which we write in the following order as a bit vector $A^{\Delta}(G)$ of length $n(n-1)/2$:

$$a_{1,0}a_{2,0}a_{2,1}a_{3,0}a_{3,1}a_{3,2} \cdots a_{i,0}a_{i,1} \cdots a_{i,i-1} \cdots a_{n-1,0}a_{n-1,1} \cdots a_{n-1,n-2}.$$

Split $R(A^{\Delta}(G))$ into contiguous substrings of length 6 and convert each substring into decimal form. Add 63 to each number. We denote the result by $S(A(G))$.

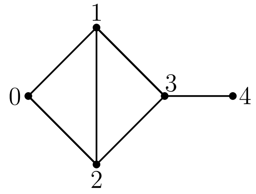

Figure 1

For a string s of integers, each between 63 and 126, we denote by $C(s)$ the string of corresponding ASCII characters. We refer to https://www.asciitable.com for a table giving this correspondence. For certain upcoming arguments, the reader may find it useful to run the free open-source mathematics software "SageMath," available at https://www.sagemath.org, in parallel to reading this note. It permits to execute code and quickly visualise some of the arguments. For instance, the aforementioned conversion is performed by the SageMath function `chr`. Considering our example from above, 126 63 89 106 becomes `~?Yj`.

Finally, the `graph6` representation of the graph G is the concatenation of $C(N(n))$ and $C(S(A(G)))$—the former encodes the order (i.e., the number of vertices) of G, while the latter encodes the edges of G. Note that the latter depends on the graph's adjacency matrix (or, equivalently, on how the graph's vertices are labelled with labels $0, \dots, n-1$), of which there may be many distinct ones, all of the same (unlabelled) graph—we shall come back to this observation in Section 4, but first give an example illustrating the definitions we have seen, and present the titular graph.

Example. Let G be the graph with vertices 0, 1, 2, 3, 4 and edges $\{0,1\}$, $\{0,2\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{3,4\}$; see Figure 1.

The adjacency matrix of G is

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ \mathbf{1} & 0 & 1 & 1 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} & 0 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & 0 \end{pmatrix},$$

which yields, as described above, the string $A^{\Delta}(G) = 1110110001$. The order of G is $5 \leq 62$, and so we have $N(5) = 68$. Moreover,

$$R(A^{\Delta}(G)) = R(1110110001) = 111011000100,$$

which yields 122 67. Why 122 67? We pad $A^{\Delta}(G)$ on the right-hand side with zeros until its bit vector is a multiple of 6 (in this case 12). Converting 111011 into decimal form yields $2^5 + 2^4 + 2^3 + 2^1 + 2^0 = 59$. Plus 63 gives 122. In the same way, 000100 yields 67. So this graph is 68 122 67. Finally, in ASCII, we have that 68 corresponds to `D`, 122 corresponds to `z`, and 67 corresponds to `C`; this gives the `graph6` string `DzC`.

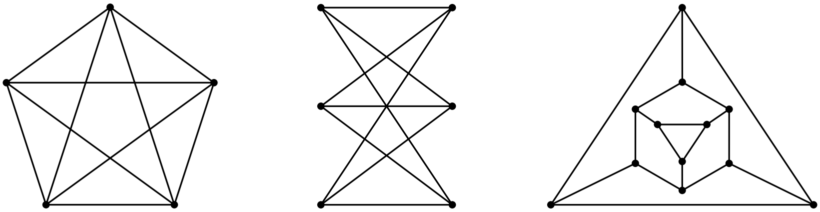

Figure 2. From left to right: K_5, the complete graph on five vertices; $K_{3,3}$, the balanced complete bipartite graph on six vertices; the Dürer graph.

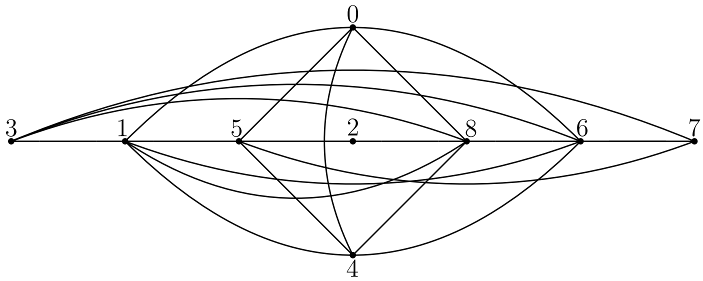

Figure 3. A depiction of `Harmony`.

3 Harmony

The vast majority of graphs have unappealing names—for a human at least—in `graph6`. We illustrate this with three examples for which we need a few definitions. The graph on n vertices in which any two vertices are adjacent is called *complete* and denoted by K_n. The graph obtained by taking two independent sets of vertices, A and B, and joining every vertex in A to every vertex in B is called a *complete bipartite graph*. It is denoted by $K_{|A|,|B|}$. If $|A| = |B|$, we say that this complete bipartite graph is *balanced*. The *Dürer graph* is named after German Renaissance artist Albrecht Dürer. His 1514 engraving "Melencolia I" (partially) shows a polyhedron, Dürer's solid, which has the Dürer graph as its 1-skeleton. The three examples we now give are depicted in Figure 2: the Kuratowski graphs K_5 and $K_{3,3}$ are `D~{` and `EFz_`, respectively, while the Dürer graph is `KhEKA?aCOT?i`. (This is somewhat imprecise, as there may exist distinct `graph6` strings yielding the same graph; we explain this in Section 4.)

But there are exceptions, of which we now present one, to this common occurrence. In Figure 3, we depict a diagram showing the graph (and its labelling) whose `graph6` string is `Harmony`.

Two graphs are *isomorphic* if there exists an *isomorphism* between them, i.e., a bijection between their vertex sets preserving adjacency. An *automorphism* is an isomorphism from a graph to itself. The set of all automorphisms of a graph G forms a group under composition called the *automorphism group* $\mathrm{Aut}(G)$. Unexpectedly (at least to the author), `Harmony`'s automorphism group has size 2, which alludes to the yin and yang of Ancient Chinese philosophy. We leave further such interpretations to the reader.

4 The rule

In this section, we assume that n is a non-negative integer with $n \leq 2^{36} - 1$. Given a `graph6` string, its corresponding graph is unique. However, there may be many pairwise distinct `graph6` strings yielding isomorphic graphs (ignoring vertex labellings). We now establish their precise number. To this end, we point out that, in the definition of the function R in Section 2, it is in fact irrelevant for the outcome whether we pad the bit vector exclusively with zeros or any other string of zeros and ones (but we do have to satisfy the length requirement); note that this is not true for L_{18} and L_{36}. For a given bit vector v of length ℓ, there are $2^{\ell \pmod 6}$ variations of $R(v)$. For any given n-vertex graph G, the string $A^\Delta(G)$ (as defined in Section 2) has length $n(n-1)/2$. Moreover, there are $n!/|\text{Aut}(G)|$ pairwise distinct adjacency matrices (as there are $n!$ different vertex labellings of G), all of which produce the same graph G. Therefore, an n-vertex graph G has

$$\frac{n!}{|\text{Aut}(G)|} \cdot 2^{\frac{n(n-1)}{2} \pmod 6}$$

pairwise different `graph6` strings; for instance any of the eight 2-symbol strings

`Bw   Bx   By   Bz   B{   B|   B}   B~`

will yield K_3.

Ultimately, if we want a word to be a `graph6` string, the only things we need to care of are (i) its first letter, including whether it is capitalised or not, and (ii) its length, i.e., the number of letters it contains. The first symbol in the `graph6` string indicates the order of the graph: it is ~ (which corresponds to 126) if and only if the graph's order is greater than 62; we are not interested here in such strings, so we focus on strings which do not start with ~. Let $A : \{\text{a}, \text{b}, \ldots, \text{z}, \text{A}, \text{B}, \ldots, \text{Z}\}$ be the function associating to a letter its decimal ASCII value. If, for a letter α, we denote its position in the alphabet by $p(\alpha)$, then we have

$$A(\alpha) = \begin{cases} p(\alpha) + 96 & \text{if } \alpha \text{ is a lower case letter,} \\ p(\alpha) + 64 & \text{if } \alpha \text{ is an upper case letter.} \end{cases}$$

Let $n \leq 62$ be a non-negative integer. For an n-vertex graph, we have that $N(n) = n + 63 = A(\alpha)$, so $n = A(\alpha) - 63$. Hence, the string $A^\Delta(G)$ has length $(A(\alpha) - 63)(A(\alpha) - 64)/2$. We then split $R(A^\Delta(G))$ into contiguous substrings of length 6, so the length of a word starting with the letter α must be

$$\ell(\alpha) := 1 + \left\lceil \frac{(A(\alpha) - 63)(A(\alpha) - 64)}{12} \right\rceil,$$

the "1 + " coming from the first symbol which encodes the graph's order. Therefore, a word that is not artificially constructed and starting with a lower case letter will be impossible to achieve—even if its first letter is an "a", we would need a 95-letter word (letters further down in the alphabet yield even longer words). So our word ought to start with a capital letter. These range from length 2 if we start with an "A" to length 60 if we start with a "Z", the latter once more being unreasonable. We point out that *if* the length of a word Σ starting with the upper case letter α is $\ell(\alpha)$, then it is guaranteed that there exists a graph whose `graph6` string is Σ because any symmetric square matrix containing only zeros and ones, and in which all entries of the main diagonal are zero, is the adjacency matrix of *some* graph, as ugly as it may be.

In conclusion, in order to assist in quenching the reader's burning desire to find further graphs whose `graph6` strings are words, in the table below, we present, for each upper case letter $\alpha \in \{A, \ldots, R\}$, the length a word Σ starting with α must have so that Σ is the `graph6` string of a graph:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2	2	2	3	4	5	6	7	9	11	12	14	17	19	21	24	27	30

Examples include `Overcommercialisation`, which gives a graph on 16 vertices, as well as `Katzenjammer`. The latter can be visualised with SageMath by executing the code

```
Graph('Katzenjammer')
```

There are various websites listing words starting with a specified letter and of a certain length, for instance 7-letter words starting with an "h" can be accessed at https://www.thefreedictionary.com/words-that-start-with-h#w7.

References

[1] K. Coolsaet, S. D'hondt and J. Goedgebeur, House of graphs 2.0: A database of interesting graphs and more, *Discrete Appl. Math.* **325** (2023), 97–107; available at https://houseofgraphs.org, visited on 26 January 2024.

[2] R. Diestel, *Graph theory*, 5th edn., Grad. Texts Math. 173, Springer, New York, 2016.

[3] B. D. McKay, Description of graph6, sparse6 and digraph6 encodings, http://users.cecs.anu.edu.au/~bdm/data/formats.txt, Last access date: 30 October 2023.

[4] B. D. McKay, Practical graph isomorphism, *Congr. Numer.* **30** (1981), 45–87.

[5] B. D. McKay and A. Piperno, Practical graph isomorphism, II, *J. Symbolic Comput.* **60** (2014), 94–112.

[6] N. J. A. Sloane, The on-line encyclopedia of integer sequences, https://oeis.org, Sequence A000088, visited on 26 January 2024.

Carol T. Zamfirescu
Department of Applied Mathematics,
Computer Science and Statistics, Ghent University
Krijgslaan 281 – S9, 9000 Ghent, Belgium
and Department of Mathematics, Babeş-Bolyai University
Cluj-Napoca, Roumania
czamfirescu@gmail.com