MATH20902: Discrete Maths, Coursework
Due by 2:00 PM, Friday 18 March 2016

Please write out answers to all of the problems that appear below and hand them in, accompanied by a standard Cover Sheet, at the reception desk of the Alan Turing Building.

Interval Graphs

This year's coursework involves *interval graphs*, which are undirected graphs whose vertex set consists of a finite collection of distinct open intervals $\{I_1, \ldots, I_n\}$ and whose edge set includes the pair (I_j, I_k) if and only if the corresponding intervals are distinct and have a non-empty intersection. That is, $(I_j, I_k) \in E$ if and only if $j \neq k$ and $I_j \cap I_k \neq \emptyset$:. Figure 1 provides an example.

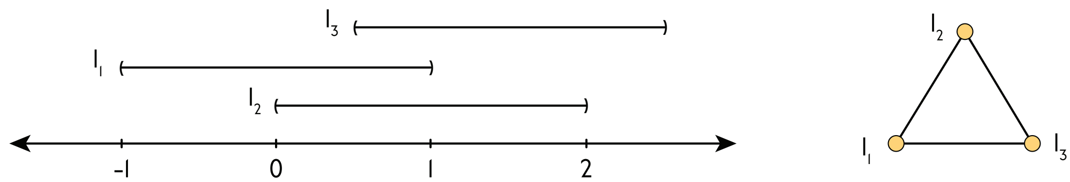

Figure 1: *Three intervals,* $I_1 = (-1, 1)$*,* $I_2 = (0, 2)$ *and* $I_3 = (1/2, 5/2)$*, and the associated interval graph, which is isomorphic to* K_3*.*

Given a collection of intervals, it's easy to draw the corresponding interval graph, but it's harder to solve the inverse problem—to start with a graph and decide whether it can be represented as an interval graph. The coursework is mainly devoted to this second problem.

Motivation: the fine structure of genes

The mathematical theory of interval graphs was developed in the 1960s, partly in response to a brilliant paper[1] by the geneticist Seymour Benzer[2] that described an experiment designed to probe the internal organisation of genes. At the time Benzer wrote, Watson and Crick's celebrated paper about the double helix[3] had already appeared, so biologists knew that genetic information was carried by DNA, but it was not yet understood *how* genetic information was represented or organised. Benzer established that the internal parts of genes are arranged like subintervals of a line segment.

---

[1]S. Benzer (1959), On the Topology of the Genetic Fine Structure, *Proceedings of the National Academy of Sciences of the USA*, **45**(11):1607–1620.

[2]Benzer, who died in 2007 at the age of 86, was one of the great scientists of the 20th century. After completing the work sketched here he went on to make fundamental contributions to behavioural genetics, identifying mutations that cause distinct, heritable changes in the behaviour of fruit flies. He was also a much-loved teacher and is the subject of a prize-winning biography: J. Wiener (1999), *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior*, Knopf, New York.

[3]J.D. Watson and F.H.C. Crick (1953), Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid, *Nature*, **171**:737–738.

He worked with mutant strains of a virus called λ-phage that infects bacteria and he was especially interested in a process through which two damaged, mutant strains of phage can—when both happen simultaneously to infect the same bacterial cell—recombine their genetic material to produce undamaged offspring. The following passage, which I've abridged from the original paper, explains the key idea:

> Consider now a perfect tape recording of a piece of music. Such a structure can be rendered unacceptable by a blemish—one false note, perhaps, or a blank interval (due to a jump of the tape, say). Given two independent "mutant" versions, it may be possible to fabricate a standard one, but only if the two blemishes do not overlap. Now if various mutant versions are tried two-by-two, in each case noting only whether or not successful recombination is possible, a new sort of result may be found. A blemish in the recording can involve a segment of the structure. It may therefore occur that one mutation intersects two others that do not themselves intersect. Given enough defective versions, the yes-or-no results of recombination experiments would enable one to construct a linear map showing the various defects in their relative positions within the standard structure.

The matrix at the left of Figure 2 illustrates a small part of his data[4]. The rows and columns are labelled by strains of the virus and the matrix has a *zero* if the two corresponding strains were able to recombine to produce undamaged offspring and a *one* otherwise. Thus, for example, strains 215 and 221 recombined to produce undamaged viruses, while strains 184 and 215 did not. This may seem an awkward convention (in-

	184	215	221	250	C33
184	1	1	0	1	1
215	1	1	0	0	0
221	0	0	1	0	1
250	1	0	0	1	1
C33	1	0	1	1	1

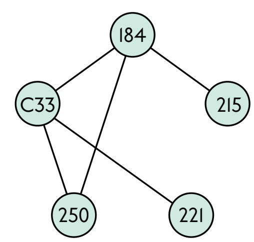

Figure 2: *A matrix representing a subset of Benzer's data and the corresponding graph, whose vertices are viral strains and whose edges connect strains that* couldn't *recombine to produce undamaged offspring.*

deed, it's the opposite of the one Benzer used), but it's natural if one is thinking about interval graphs: if the parts of a gene really are arranged like subintervals along a line, then successful recombination is only possible between two strains whose mutations *don't* overlap. Benzer's main question: *Could the components of genes be organised linearly?* thus becomes: *Is the matrix above the adjacency matrix of some interval graph?*[5]

He addressed this question by finding a way to permute the rows and columns of the matrix in Figure 2 that converts it into the form illustrated in the right panel of Figure 3, where only the nonzero entries are shown so as to highlight the structure. If one examines the entries below a diagonal entry (or, by symmetry, to the right of it) they consist of an initial sequence of consecutive ones, followed by zeroes. Such a matrix can be represented as an interval graph, as is illustrated in the right panel of Figure 3.

[4]The 1959 paper involved 145 mutant strains, though he eventually mapped out more that 2400.

[5]When answering this question we should, of course, ignore the 1's on the diagonal of the matrix, as

Figure 3: *The left panel above shows the matrix from Figure 2 with its rows and columns permuted, its nonzero entries highlighted and its zero entries suppressed. The panel at right shows one possible pattern of mutations (there are many others) consistent with the permuted matrix. If one thinks of the bars as intervals, then the matrix at left is—except for the 1's along the diagonal—the adjacency matrix of the corresponding interval graph.*

Formalising Benzer's approach

The following definitions and propositions—which are mine rather than Benzer's—are meant to clarify the relationship between adjacency matrices and interval graphs.

Definition. *A symmetric, n-by-n matrix A with entries $A_{ij} \in \{0, 1\}$ is a* **Benzer matrix** *if, for each $1 \leq j \leq n$, the list of elements $\{A_{j,j}, \ldots, A_{n,j}\}$ on and below the the j-th diagonal entry has the property that $A_{k,j} \geq A_{k+1,j}$ for all $j \leq k < n$.*

As the matrices in question contain only zeroes and ones, this means that the subdiagonal part of each column in a Benzer matrix consists of a sequence of zero or more 1's, followed by 0's. Thus the matrix on the left in Figure 3 is an example of a Benzer matrix, while the one at the left of Figure 2 is not. The next definition allows us to place lists of intervals in order.

Definition. *If $I_1 = (l_1, r_1)$ and $I_2 = (l_2, r_2)$ are a pair of intervals we will write that $I_1 \prec I_2$ if either one of the following conditions holds:*

(a) $l_1 < l_2$;

(b) $l_1 = l_2$ *and* $r_1 < r_2$.

We will refer to this relation as the **lexicographic order** *on intervals. If a collection of intervals $\mathcal{C} = \{I_1, \ldots, I_n\}$ has the property that $j < k \Rightarrow I_j \prec I_k$ then we will say that the collection $\mathcal{C}$ is* **in lexicographic order**.

The lexicographic order on intervals, along with the following proposition, explains why Benzer was interested in Benzer matrices:

Proposition 1. *Suppose $\mathcal{C} = \{I_1, \ldots, I_n\}$ is a collection of non-empty intervals, no two of which are identical, and define a matrix A by*

$$A_{ij} = \begin{cases} 1 & \text{if } I_i \cap I_j \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

If the collection $\mathcal{C}$ is in lexicographic order, then A is a Benzer matrix.

interval graphs don't have loops that connect vertices to themselves.

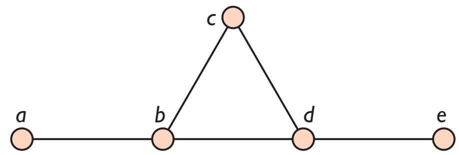

Figure 4: *The sequence* $(v_1, v_2, v_3, v_4, v_5) = (a, c, b, d, e)$ *is a perfect elimination scheme for the graph above. That is,* a *is simplical in the original graph* G*,* c *is simplical in* $G \backslash a$*, and so on.*

Testing whether $G(V, E)$ is an interval graph

Although Benzer's work encouraged a lot of mathematicians to think about interval graphs, some had already been working on them for a couple of years, as the Hungarian mathematician György Hajós had posed the problem of deciding when a graph could be represented by collection of intervals[6] in 1957. All this work eventually produced a number of elegant characterisations[7] one of which I'll introduce below.

Definition. *A vertex* v *in a graph* $G(V, E)$ *is said to be* **simplicial** *if the subgraph induced by* $\{v\} \cup A_v$ *is isomorphic to a complete graph. Here* A_v *is the adjacency list of* v*.*

Definition. *A* **perfect elimination scheme** *for a graph with* n *vertices is an ordering of the vertex set as a sequence* $(v_1, \ldots, v_n)$ *with the property that* v_j *is simplicial in the subgraph induced by the set* $\{v_j, \ldots, v_n\}$*.*

Figure 4 provides an example.

If a graph has one or more prefect elimination schemes, it turns out to be easy to find one of them: the algorithm below provides an explicit recipe, while Figure 5 provides a worked example.

Algorithm (Perfect elimination)**.**
Given a graph $G(V, E)$*, either produce a perfect elimination scheme* $S = (v_1, v_2, \ldots, v_n)$ *or establish that no such scheme exists.*

(1) Set a counter $j \leftarrow 1$*, define* $G_1 = G$ *and initialise* S *to be the empty sequence:* $S \leftarrow ()$*.*

(2) Find a simplicial vertex w *in* G_j*. If* G_j *contains no simplicial vertices, then stop: the original graph does not have a perfect elimination scheme. Otherwise, go to the next step.*

(3) Set $v_j \leftarrow w$ *and add* v_j *to the right end of the sequence, so*

$$S \leftarrow (v_1, \ldots, v_j).$$

Set $G_{j+1} \leftarrow G_j \backslash w$
Set $j \leftarrow j + 1$ *and, if* $j \leq n$*, go to step (2). If* $j > n$*, then stop:* $S = (v_1, \ldots, v_n)$ *is the desired sequence.*

[6] G. Hajós (1957), Über eine Art von Graphen, *Internationale Mathematische Nachrichten*, **11**.

[7] The standard reference is M.C. Golumbic (2004), *Algorithmic Graph Theory and Perfect Graphs*, 2nd edition, Elsevier, Amsterdam.

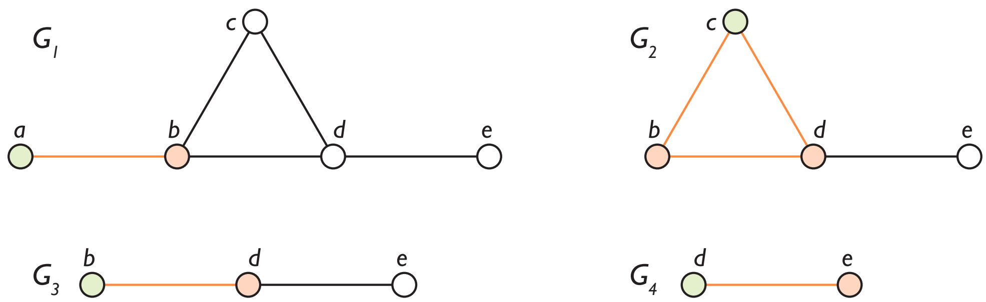

Figure 5: *The sequence of graphs $G_1 \ldots G_4$ produced when one applies the perfect elimination algorithm to the graph in Figure 4 and generates the perfect elimination scheme $(v_1, v_2, v_3, v_4, v_5) = (a, c, b, d, e)$. In each of the G_j the simplicial vertex w is highlighetd in green, while its neighbours and the edges of the subgraph induced by $\{w\} \cup A_w$ are shown in orange.*

The point of all this is the following result

Proposition 2. *If $G(V, E)$ is an interval graph, then it has a perfect elimination scheme.*

This result, via its contrapositive, provides a quick way to establish that a graph *isn't* an interval graph. One applies the algorithm above and if it fails to produce a perfect elimination scheme, then G can't be an interval graph.

Problems

---

1 (Constructing interval graphs).
Find a set of intervals whose corresponding interval graph is:

(a) the complete graph K_n for any $n \geq 1$;

(b) the path graph P_n for any $n \geq 2$;

(c) a tree consisting of a single vertex of degree k and k vertices of degree 1, with $k \geq 3$.

2 (Most cycle graphs are not interval graphs).
Prove that the cycle graph C_n is not an interval graph unless $n = 3$.

3 (Two "Benzer problems").
The matrices below are data from simulated experiments.

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

For each one, draw the corresponding graph G, then determine whether G is an interval graph. Your answers should include either an explanation for why G cannot be an interval graph or a collection $\mathcal{C}$ of lexicographically ordered intervals whose interval graph is isomorphic to G.

4 (A mystery).
The mathematician Claude Berge was a founding member of the French literary group OuLiPo and once wrote a mystery story whose resolution depends on properties of interval graphs. The Duke of Densmore was murdered by an explosion that completely destroyed his castle. The main suspects are his many ex-wives who had, in the weeks before his death, been visiting him. It seems that one of them must have visited twice, for the forensic evidence suggests that the bomb was very skillfully crafted to fit precisely into one of the Duke's wardrobes and so must have been made to some previously-gathered measurements.

But each of the ex-wives claims to have visited the Duke just once, staying for a few days and then leaving. Police interviews established that their visits overlapped as follows:

- Alice met Barbara, Carol, Emma and Fran during her visit
- Barbara met Alice, Carol, Diane, Emma and Gemma
- Carol met Alice, Barbara and Diane
- Diane met Barbara, Carol and Emma
- Emma met Alice, Barbara, Diane and Gemma
- Fran met Alice and Gemma
- Gemma met Barbara, Emma and Fran

Is it possible that all 7 women are telling the truth? If not, whom do you suspect, and why?

5 (Benzer matrices and lexicographic order).
Prove Proposition 1, the one that relates collections of intervals $\mathcal{C} = \{I_1, \ldots, I_n\}$ in lexicographic order to Benzer matrices.

6 (The converse of Prop. 2).
Prove the following or find a counterexample:

If $G(V, E)$ has a perfect elimination scheme, then it is an interval graph.

Warning Against Plagiarism
Plagiarism is the presentation of other people's work as your own. It is both dishonest and unfair to other students. The University takes plagiarism very seriously and so will I: if your coursework submission contains material copied without proper attribution from textbooks (including Jungnickel's), Wikipedia or other web sites or any other source except the lecture notes, problem sets and solutions from this module, I will refer you to the School's Academic Malpractice Panel. Standard penalties include a mark of zero *for the entire coursework*. If you are unsure whether or how to acknowledge a source that you have used, please consult the School's Undergraduate Handbook[8] or check with me well before the due date.

[8]See the `Handbook page on academic malpractice`.