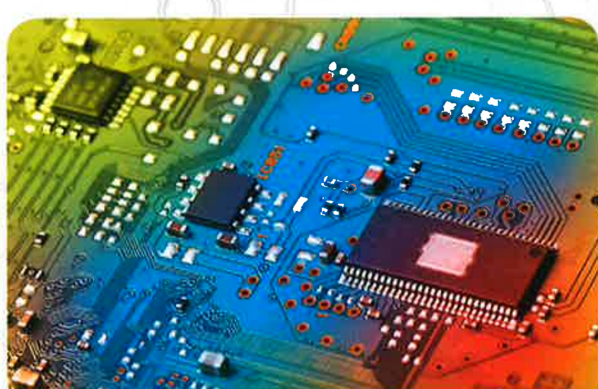
15 Optimizing complex networks: graph theory

Concepts
- Representation
- Systems

Microconcepts
- Simple, weighted and directed graphs
- Minimum spanning tree
- Adjacency and transition matrices
- Eulerian circuits and trails
- Hamiltonian paths and cycles
- Graphs: directed, connected, complete; degree of a vertex, weight of an edge
- Adjacency matrix, lengths of walks, connectivity of a graph, transition matrix, steady state probabilities
- Spanning trees, minimum spanning trees, Prim's and Kruskal's algorithms, Prim's algorithm from a table
- Eulerian circuits and trails, leading to the Chinese postman problem
- Tables of least distances
- Classical and practical travelling salesman problems

This chapter explores how diagrams (called graphs) can be used to model the connections between metro stations, components in electrical circuits and people within a social network. It also looks at how algorithms can be implemented to find optimal routes, for example for delivering parcels in a given area in the shortest possible time.

What is the quickest route between two metro stations?



How could an electrical engineer model the connections on a circuit board?

How can the driver of a snowplough ensure that they clear every road in a town after a blizzard as efficiently as possible?

How can a postman deliver letters to every house on their route while minimizing the distance they travel?

A national park has bicycle trails connecting seven viewpoints. The distances between pairs of viewpoints are given in the table below. The entrance to the park is at A.

Viewpoints	A	B	C	D	E	F	G
A	0	5	4.5				
B	5	0	4	3.5			
C	4.5	4	0	2		6	
D		3.5	2	0	6	5.5	
E				6	0	4.5	3
F			6	5.5	4.5	0	3.5
G					3	3.5	0

- What do you think that a blank cell represents?
- Name three pieces of information represented in the table.
- Draw a diagram to represent the information in the table. Compare with others.
- How could you use your diagram to find the shortest route around all the viewpoints?

Developing inquiry skills

Write down any similar inquiry questions you could ask and investigate for your local park. What information would you need to find?

Could you write similar inquiry questions about your local supermarket or shopping mall? Or any other place?

Think about the questions in this opening problem and answer any you can. As you work through the chapter, you will gain mathematical knowledge and skills that will help you to answer them all.

Before you start

You should know how to:

1 Raise a matrix to an integer power.

eg If $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ find A^4.

Using a GDC: $A^4 = \begin{pmatrix} 5 & -4 \\ 4 & -3 \end{pmatrix}$

2 Find the steady state vector for a transition matrix.

eg Find the steady state vector for the transition matrix $P = \begin{pmatrix} 0.1 & 0.6 \\ 0.9 & 0.4 \end{pmatrix}$.

Form equations $\begin{matrix} 0.1x + 0.6y = x \\ 0.9x + 0.4y = y \end{matrix}$

Solve either of these along with $x + y = 1$:

$x = 0.4, y = 0.6$

Skills check

Click here for help with this skills check

1 If $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & 1 \\ -1 & 0 & 2 \end{pmatrix}$ find

a A^2 **b** A^4.

2 Let $P = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}$ be a transition matrix in which the probability of moving from state i to state j is P_{ji}.

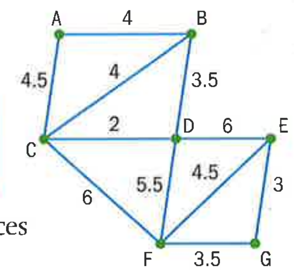
a By forming a system of linear equations find the steady state vector.

b Verify your answer is correct by calculating large powers of P.

c What is the long-term probability of being in state j?

15.1 Constructing graphs

For the opening problem you may have drawn a diagram like this. This is a **graph**.



> A graph is defined as set of **vertices** and a set of **edges.**
>
> A vertex represents an object. An edge joins two vertices.

In the graph of the national park the viewpoints are the vertices of the graph. The trails are the edges.

In mathematical graph theory, graphs do not have to be drawn to scale and the positions of the vertices do not need to relate to their positions in the real world. A graph will always show which vertices are directly connected.

The graph above also shows the distances between vertices. A graph that shows values like these (called **weights**) is called a **weighted graph**. The "weight" can be any quantity, such as cost, time or distance.

The map of the London Underground is a famous example of an unweighted graph. It gives no information regarding the distances between the stations, only the connections between them.

> **TOK**
>
> Have you heard of the Seven Bridges of Königsberg problem? Königsberg is now Kaliningrad in Russia. Do all mathematical problems have a solution?

> A **walk** is any sequence of vertices passed through when moving along the edges of the graph.

In the graph of the national park ABC and ABDBC are both walks from A to C.

The information in an unweighted graph can also be contained in an **adjacency table**. In an adjacency table the entries indicate the number of direct connections between two vertices.

Example 1

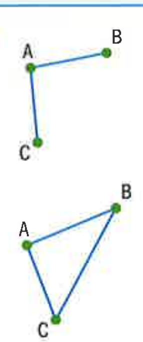
The table shows the connections between stations on a small mountain railway. An empty cell indicates no direct connection between the stations.

a Show this information on a graph.

b Write down two possible walks from A to D.

	A	B	C	D
A		1	1	
B	1		1	
C	1	1		1
D			1	

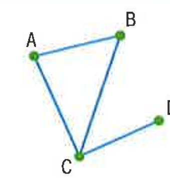
a



To draw the graph begin with vertex A.

From the first row in the table, A is connected to B and C, so add them to the graph.

From the second row, B is connected to A and C. The first of these is already included, so you only need to add the edge connecting B to C.



The third row shows that you need to add the edge CD.

The final row includes no extra edges. Use it to check your graph is drawn correctly.

b eg ABCD, ACD or ABACD

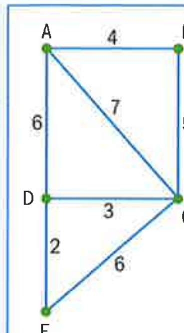
There are infinitely many answers to part **b** as you can pass through each vertex and along each edge as many times as you like.

Example 2

The table shows costs in dollars of travelling by bus between towns.

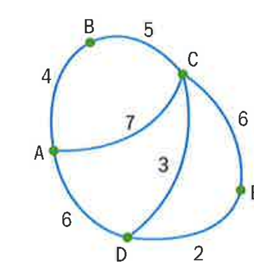
Show this information on a graph.

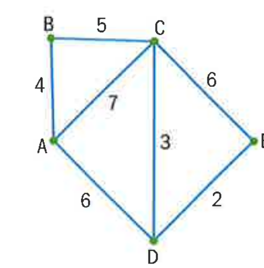
	A	B	C	D	E
A		4	7	6	
B	4		5		
C	7	5		3	6
D	6		3		2
E			6	2	

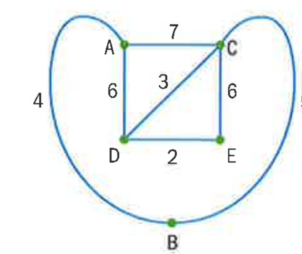


Because the costs are given in the table you must also show them on the graph.

Note that there are many different ways to draw the graph, eg:







Any depiction is fine so long as you show the correct connections.

> **Reflect** How do you construct a graph from information given in a table?
> What information can be readily obtained from a graph that cannot easily be seen in a table?

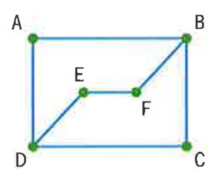
Geometry and trigonometry

Exercise 15A

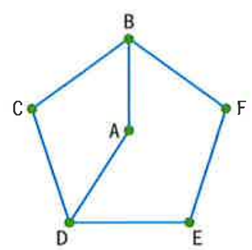
1 State which of the graphs represent the information given in the table.

	A	B	C	D	E	F
A		1		1		
B	1		1			1
C		1		1		
D	1		1		1	
E				1		1
F		1			1	

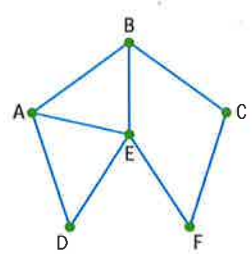
a



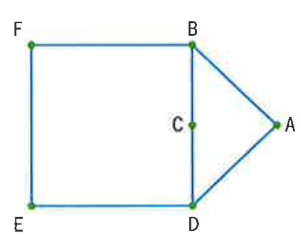
b



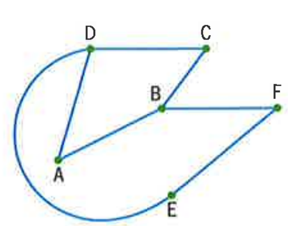
c



d

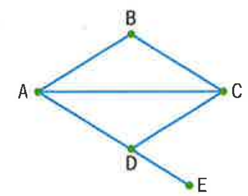


e

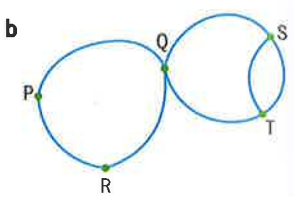


2 Draw a table to represent the number of direct connections between the vertices in the graphs below.

a



b



3 Draw the weighted graphs represented by the tables.

a

	A	B	C	D	E
A		15	12	10	
B	15			13	9
C	12				12
D	10	13			16
E		9	12	16	

b

	A	B	C	D	E
A		6	10		
B	6		7		
C	10	7			
D					5
E				5	

4 Draw weighted graphs represented by the tables below. Each table is read from row to column so the weight of the edge from A to B is 10. You will need to decide how to show the fact that A to B is not necessarily the same as B to A.

a

	A	B	C	D
A		10		7
B			8	
C		8		12
D				

b

	A	B	C	D
A		10		
B	10		6	
C		7		8
D	11			

5 Draw unweighted graphs to represent the tables below. A 1 in the intersection of row *i* and column *j* indicates that there is a connection from *i* to *j*.

a

	A	B	C	D
A		1		1
B	1			
C		1		1
D			1	

b

	A	B	C	D
A		1	1	1
B			1	
C		1		1
D			1	

6 A salesman needs to travel from Sheffield to Manchester to Nottingham and then back to Sheffield. A route finder algorithm has been programmed with the data in the table below, showing distances in miles between the major cities in England. Unfortunately the information on the distance from Manchester to Nottingham has not been included.

a Write down the distances from Sheffield to Manchester and from Nottingham to Sheffield.

In trying to give an estimate for the shortest distance the algorithm will find the shortest distance from Manchester to Nottingham via one other town. To do this efficiently it needs to reduce the table as much as possible.

b Explain why all those towns further than 77 miles from Manchester can be excluded from the search.

c Draw a graph showing all remaining routes from Manchester to Nottingham which go via one other town.

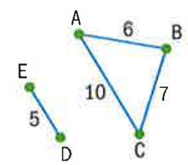
d Hence state an estimate for the shortest distance the salesman must travel to visit the two towns and return to Sheffield.

	Birmingham	Bristol	Derby	Exeter	Leeds	Liverpool	London	Manchester	Newcastle	Norwich	Nottingham	Oxford	Portsmouth	Sheffield	Southampton
York	129	217	88	292	25	97	194	65	82	181	80	174	252	53	239
Southampton	128	74	164	107	229	217	77	208	319	190	160	65	17	195	
Sheffield	76	164	37	240	34	72	160	38	128	146	39	130	208		
Portsmouth	141	94	175	124	242	231	71	222	332	184	173	78			
Oxford	63	70	90	141	163	153	56	144	254	141	95				
Nottingham	49	137	16	213	70	99	123		159	123					
Norwich	161	209	139	282	174	217	112	185	260						
Newcastle	204	288	161	364	94	155	274	131							
Manchester	81	162	59	238	41	34	184								
London	111	114	123	170	191	198									
Liverpool	90	161	81	237	73										
Leeds	110	196	70	271											
Exeter	164	75	203												
Derby	40	127													
Bristol	88														

In a **connected** graph it is possible to construct a walk between any two vertices.

A **subgraph** of a graph G consists entirely of vertices and edges that are also in G.

The graph from Exercise 15A question **3b** is an example of an **unconnected** graph.



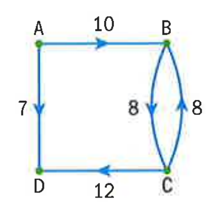
It is unconnected because no walk exists from vertex A to vertex D, for example.

It is, however, made up of two connected **subgraphs.**

In a **directed** graph all the edges are assigned a specific direction.

A directed graph is **connected** if a walk can be constructed in at least one direction between any two vertices. A directed graph is **strongly connected** if it is possible to construct a walk in either direction between any two vertices.

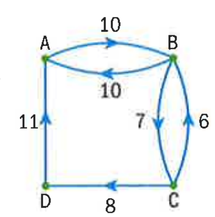
The graphs from Exercise 15A question **4** are directed graphs.



The graph drawn in question **4a** is a connected, directed graph; sometimes this is referred to as **weakly** connected.

It is possible to go from A to D, for example, but not from D to A.

The graph drawn in question **4b** is strongly connected. It is possible to go in both directions between any two vertices.



The context often makes it clear whether a directed or undirected graph is required. For example, a bus may go from town A to town B but not make the reverse journey.

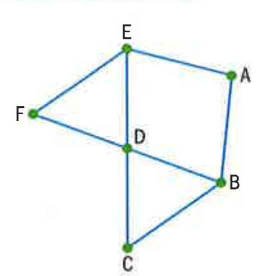
Reflect When is it appropriate to use directed or undirected graphs?

Except in those cases where extra clarity is needed, the word **graph** will be taken to mean an undirected graph. If the graph is directed it will be explicitly mentioned.

The **degree** (or order) of a vertex in a graph is the number of edges with that vertex as an end point.

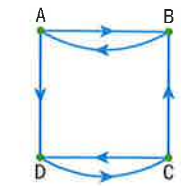
A vertex whose degree is an even number is called an **even vertex** or is said to have **even degree**.

The degree of each vertex in the graph shown is: A 2, B 3, C 2, D 4, E 3, F 2.



The **in-degree** of a vertex in a directed graph is the number of edges with that vertex as an end point. The **out-degree** is the number of edges with that vertex as a starting point.

In the graph shown A has an in-degree of 1 and an out-degree of 2.



Investigation 1

1 Five people shake hands before a meeting as indicated in the table.

	A	B	C	D	E
A		1		1	1
B	1		1		
C		1		1	
D	1		1		1
E	1			1	

a Show this information in a graph with each edge representing a handshake between two people.

b From the graph write down the total number of handshakes that take place before the meeting.

2 If possible, draw a graph consisting only of vertices with the degrees listed. If it is not possible, explain why not.

a 1, 2, 2, 2, 3 **b** 1, 2, 2, 2 **c** 2, 2, 3, 3 **d** 2, 2, 2, 4

3 From your graphs in questions **1** and **2**, can you find a link between the degrees of the vertices in a graph and the number of edges? If necessary draw more graphs and record the degrees of the vertices and the number of edges.

4 Using your result from question **3**, explain why it is not possible to construct a graph with a degree sequence of 3, 2, 4, 1, 5, 2, 3, 3.

5 The link conjectured in question **3** is called the handshaking theorem. Explain why with reference to your answer to question **1**.

6 Can you find a similar theorem linking the degrees of the vertices in a directed graph and the number of edges?

Exercise 15B

1 A group of people share several attributes which can be shown in a graph, with the five people as the vertices. The edges represent the connections between them. For the following connections discuss whether it would be best to draw a directed or an undirected graph.

An edge represents:

a lives in the same block as

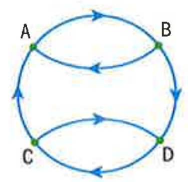
b is the sister of

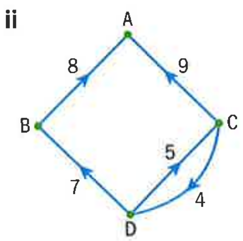
c is a friend of

d is a follower on Twitter of.

Geometry and trigonometry

2 a Draw a table to show the information contained in each of the following directed graphs.

i 

ii 

b State whether each of the graphs is connected, strongly connected or not connected.

3 a i Draw an unweighted directed graph containing the information given in the following table.

	A	B	C	D
A		1		1
B			1	
C	1	1		1
D	1		1	

ii For each vertex state the in-degree and the out-degree.

b i Draw a weighted directed graph containing the information represented by the following table.

	A	B	C	D
A		6		4
B	6			
C		5		7
D			7	

ii For each vertex state the in-degree and the out-degree.

4 Without drawing the graph, write down the in-degree and the out-degree of each vertex given by the table below.

	A	B	C	D	E
A		1		1	1
B			1		
C		1			1
D			1		1
E	1			1	

5 The table shows part of the time-table for the ferries on Lake Starnberg in Germany.

Stop (village)	1	2	3
Starnberg	9.35	12.00	14.30
Possenhofen	10.07		14.44
Tutzing	10.31		15.06
Ammerland		12.44	15.22
Bernried	10.49		
Ambach		13.01	15.39
Seeshaupt	11.08	13.23	15.58
Ambach	11.32		
Bernried		13.48	16.23
Ammerland	11.51		
Tutzing	12.08	14.08	16.43
Possenhofen		14.34	17.09
Starnberg	12.28	14.50	17.25

a Show the connections between the towns on a directed graph.

b State one piece of information that is easier to see on the graph than on the timetable.

c Give two pairs of towns that are directly connected in one direction only.

d From your graph give four possible routes from Starnberg to Ambach which do not call at any town more than once.

e Use the timetable to find how many of the routes in part **d** could be completed in one day.

Developing inquiry skills

In the opening problem, you drew a graph to represent the information in the table. Check that your graph is a weighted graph that contains all of the information in the table.

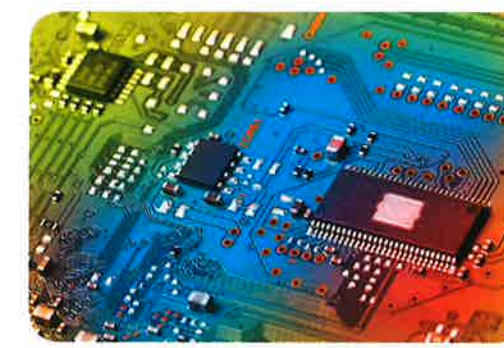
Is the graph connected?

State the degree of each vertex.

15.2 Graph theory for unweighted graphs

Often, we are interested only in the connections between vertices and not in the weights of the edges.

Examples of such situations might be a map of a metro system, connections on an electronic circuit board or in a large file-storage system, or a network of friends on social media.



This information will normally be shown in an unweighted graph.

Two vertices are said to be **adjacent** if they are directly connected by an edge. An element A_{ij} of an adjacency matrix A is the number of direct connections between vertex i and vertex j.

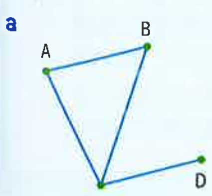
TOK

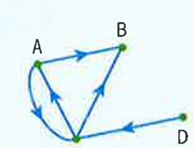
To what extent can shared knowledge be distorted and misleading?

Unlike in a table, when giving information about connections in a matrix, if there is no connection a 0 has to be put in as a place holder rather than leaving the entry empty.

Example 3

Find the adjacency matrices for the two graphs shown.

a 

b 

Continued on next page

Geometry and trigonometry

a

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

As the graph is undirected the matrix is symmetric in the diagonal from top left to bottom right (the leading diagonal).

b

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

An adjacency matrix indicates possible movement from row to column; this is the opposite to the convention for transition matrices.

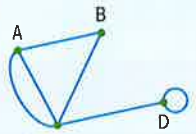
An edge from a vertex back to itself is called a **loop**.
If there are **multiple edges** between two vertices the graph is a **multigraph**.
A **simple graph** is one with no loops or multiple edges.

Reflect What can you deduce about the adjacency matrix of a simple graph from the above definition?

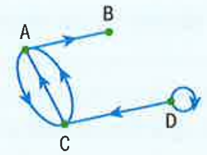
Example 4

Find the adjacency matrices for the two graphs shown.

a



b



a

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \end{array}$$

In an undirected graph a loop is shown as a 2 in an adjacency matrix (as going both ways along it is possible).

b

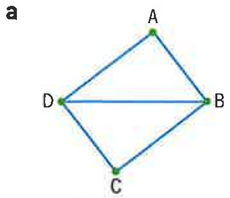
$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

In this graph the loop is shown as a 1 in the adjacency matrix as the edge is directed.

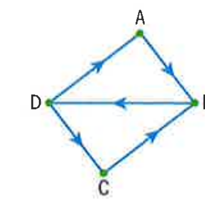
Exercise 15C

1 Find the adjacency matrices for the following graphs.

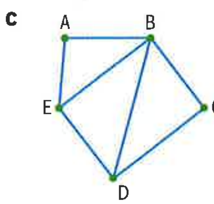
a



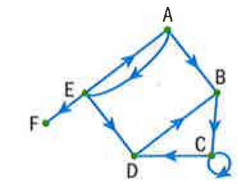
b



c



d



2 State the features of an adjacency matrix which tell you that a graph is simple.

3 Explain how you can use the adjacency matrix to find:

- **a** the degree of each vertex in an undirected graph
- **b** **i** the in-degree
  - **ii** the out-degree

  of each vertex in a directed graph.

4 The following adjacency matrix shows the relationships between a group of people where a 1 in the entry in row i and column j indicates that person i knows the name of person j.

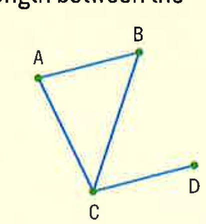
$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

- **a** **i** Explain why the matrix has a 1 in each entry on the leading diagonal.
  - **ii** Explain why the graph does not have to be symmetric.
- **b** Without drawing the graph find:
  - **i** how many names A knows
  - **ii** the number of people who know the name of C
  - **iii** who knows the most names
  - **iv** whose name is known by the largest number of people in the group.
- **c** State whether there is anyone who does not know A's name and also does not know anyone who knows A's name.

Investigation 2

1 For this graph find the number of walks of the given length between the two given vertices. Write down each of the walks.

- **a** length 2 between A and C
- **b** length 2 between C and C
- **c** length 3 between A and C
- **d** length 3 between D and D



2 **a** Find the adjacency matrix M for the graph shown.

- **b** Find M^2 and M^3.
- **c** Conjecture a link between the powers of an adjacency matrix and the lengths of walks between vertices.

Continued on next page

Geometry and trigonometry

3 Draw a **directed** graph of your own and verify that the conjecture in question **2c** still holds.

4 **Factual** How do you find the number of walks of length n between two vertices in a graph?

5 Explain how the matrices M, M^2 and M^3 can be combined to give a matrix which shows the number of walks of length 3 or less between the vertices of the graph.

6 **Factual** What can be calculated from the powers of an adjacency matrix?

7 **Factual** How can you use the powers of an adjacency matrix to find the **minimum** length of path between two vertices?

8 **Conceptual** How does the adjacency matrix allow you to analyse the paths between two points on a graph?

Let M be the adjacency matrix of a graph. The number of walks of length n from vertex i to vertex j is the entry in the ith row and the jth column of M^n.

The numbers of walks of length r or less between any two vertices are given by the matrix S_r where $S_r = M + M^2 + \ldots + M^r$

Exercise 15D

1 The adjacency matrices for the graphs in Exercise 15C question **1** are shown below.

a

	A	B	C	D
A	0	1	0	1
B	1	0	1	1
C	0	1	0	1
D	1	1	1	0

b

	A	B	C	D
A	0	1	0	0
B	0	0	0	1
C	0	1	0	0
D	1	0	1	0

c

	A	B	C	D	E
A	0	1	0	0	1
B	1	0	1	1	1
C	0	1	0	1	0
D	0	1	1	0	1
E	1	1	0	1	0

d

	A	B	C	D	E	F
A	0	1	0	0	1	0
B	0	0	1	0	0	0
C	0	0	1	1	0	0
D	0	1	0	0	0	0
E	1	0	0	1	0	1
F	0	0	0	0	0	0

Use these matrices to find the number of walks of length 3 between:

i A and A **ii** C and D.

In each case use the graphs in Exercise 15C question **1** to verify your result by listing all the possible walks.

2 By considering an appropriate power of a matrix, verify your answer to Exercise 15C question **4** part **c i**.

3 Use the adjacency matrices M to answer the following questions.

a

	A	B	C	D
A	0	1	1	0
B	1	0	1	0
C	1	1	0	1
D	0	0	1	0

b

	A	B	C	D	E
A	0	1	0	0	0
B	1	0	1	0	0
C	0	1	0	1	1
D	0	0	1	0	1
E	0	0	1	1	0

i Find M^2 and M^3.

ii Write down the number of walks of length 2 between C and A.

iii Find the number of walks of length 3 or less between C and A.

The diameter of a graph is the length of the maximum shortest walk between any two vertices (so all the vertices can be reached from any other vertex in a walk of length less than or equal to the diameter of the graph).

iv Use the powers of the adjacency matrix to find the diameters of each graph. Justify your answer.

4 The following adjacency matrix shows the connections between some ports in the Shetland Islands off the coast of Scotland. The ferries go in both directions between each pair of ports.

	Castlebay	Eigg	Lochboisdale	Mallaig	Rum	Tiree	Tobermory
Castlebay	0	0	1	0	0	1	1
Eigg	0	0	0	0	1	0	0
Lochboisdale	1	0	0	0	0	0	1
Mallaig	0	0	0	0	1	0	0
Rum	0	1	0	1	0	0	0
Tiree	1	0	0	0	0	0	1
Tobermory	1	0	1	0	0	1	0

a Find the number of routes you can travel to get from Lochboisdale to Castlebay in three ferry trips.

b By considering the powers of the adjacency matrix find which ports it is possible to get to from Tiree using the ferry system.

An extra ferry route is now added between Mallaig and Lochboisdale.

c Write down the new adjacency matrix and hence state the two ports that are furthest apart.

5 **a** Use the directed graph you drew for Exercise 15B question **5** to write down an adjacency matrix for the ferry connections on Lake Starnberg.

b By taking a suitable power of the adjacency matrix find all the pairs of towns that are not connected directly or with at most one stop, stating clearly the direction of the connection.

Reflect How might you work out the maximum number of steps necessary to move between any two vertices on a graph?

How can we use powers of the adjacency matrix to determine whether or not a graph is connected?

TOK

Matrices are used in computer graphics for three-dimensional modelling.

How can this be used in real-life situations in other areas of knowledge?

Transition matrix

A random walk on a graph is a walk in which the vertex moved to is chosen randomly from those available.

In a graph with a finite number of vertices a random walk is a finite Markov chain.

A transition matrix can be constructed for both directed and undirected graphs.

The probability of moving from one vertex to any of the adjacent vertices is defined to be the reciprocal of the degree of the vertex in an undirected graph and the reciprocal of the out-degree in a directed graph.

As discussed in earlier chapters, the probability of moving from vertex i to vertex j will be the entry in the ith **column** and jth **row** of the matrix.

HINT

See Chapter 9 for an introduction to Markov chains.

The direction of movement in a transition matrix is **opposite** to that in an adjacency matrix.

The steady state probabilities indicate the proportion of time that would be spent at each vertex if a random walk was undertaken for a long period. The transition time between the vertices is ignored.

The Google PageRank algorithm was developed by Larry Page and Segei Brin in 1996 while working at Stanford University. They wanted to be able to rank web pages in an Internet search so that the ones most likely to be useful come towards the top of the list.

The algorithm they developed considered links from web pages as the edges in a directed graph. The steady state probabilities calculated from the transition matrix indicate which site would be visited most often if someone were randomly clicking on links in web pages. This site would come top of the list in a search.

The justification for this ranking is that the sites most likely to be visited will either have lots of links to them, or be linked to by sites with lots of links going to them. In either case this is likely to reflect their relative importance.

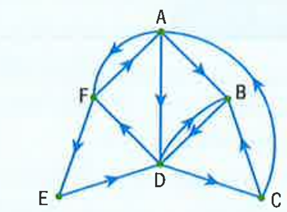
EXAM HINT

Within the IB syllabus questions will only be set in which the graph is connected. In the case of a directed graph, it will be strongly connected.

EXAM HINT

If a walk around a graph is equally likely to take any of the edges leading from a vertex, it is called a **random walk**.

Example 5



In this graph vertices A to F represent web pages and the edges indicate links between the pages.
For example, page A has links from both C and F and contains links to B, D and F.

a Construct the transition matrix for a random walk around this graph.

b Find the steady state probabilities for the network.

c Hence rank the vertices in order of importance.

a The transition matrix is

$$\begin{pmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 \end{pmatrix}$$

For example, A has three edges connecting it to vertices B, D and F, and so each of these has a probability of $\frac{1}{3}$, which is shown in the first column of the matrix.

Because the question does not specify which method to use, you can just look at high powers of the transition matrix using your GDC.

You need to consider powers large enough to ensure there is no change in the third significant figure. An alternative would be to solve a system of six linear equations or diagonalize the matrix.

b The steady state probabilities are given by the vector

$$\begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} \begin{pmatrix} 0.130 \\ 0.207 \\ 0.109 \\ 0.326 \\ 0.076 \\ 0.152 \end{pmatrix}$$

c The ranking would be D, B, F, A, C, E.

The vertex with the highest probability is placed at the top of the list.

Reflect How do you construct a transition matrix for a given graph?
Why might a random walk indicate the most important sites on a graph?
How can the steady state probabilities be used to rank lists?

Geometry and trigonometry

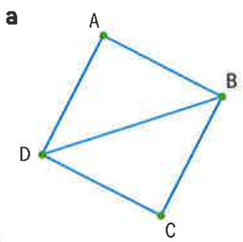
Exercise 15E

1 For each of the two graphs below find the transition matrix, $\boldsymbol{P}$, and hence the steady state probabilities by:

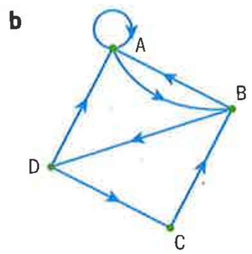
i considering high powers of $\boldsymbol{P}$

ii solving a system of linear equations.

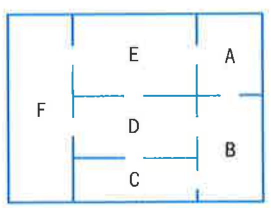
a



b



2 In an experiment on artificial intelligence a robot is put in a maze which has six rooms labelled A to F. Each of the rooms has gaps in the wall connecting it to adjacent rooms.



a Draw the maze as a graph, with the vertices representing the different rooms and the edges representing the gaps between rooms.

The robot moves randomly and once in a room is equally likely to leave through any of the gaps.

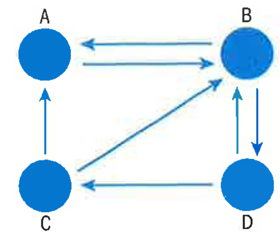
b Write down a transition matrix for the maze.

c If the robot begins in room A, what is the probability it is again in room A after passing through four of the gaps?

d **i** Find the two rooms which are most likely to be visited by the robot.

ii Determine the percentage of time that the robot will be in each of these rooms.

3 The following diagram shows four websites. The arrows indicate a link from one of the pages to another; for example, there is a link from website D to website C.



a Write down the transition matrix to represent a random walk around this graph.

b Show that it is not possible to link from B to C using exactly three links, and write down two other connections that are also not possible using exactly three links.

c By solving four linear equations find:

i the steady state probabilities

ii the proportion of time that a person following random links will be on site B.

d Based on this information write down the order in which the sites might be listed by a search engine, giving the one with the largest steady state probability first.

4 Within a large social network links can be created from one person's page to another. A small group of people share links according to the table below. For example, from Antoine's page it is possible to link directly to the pages of Belle, Charles and Emil.

Antoine	Belle	Charles	Dawn	Emil	Frances
Belle	Antoine	Antoine	Emil	Antoine	Dawn
Charles	Charles	Belle	Frances	Dawn	
Emil		Dawn			

George is also part of the network and friends with all those listed. He decides to visit the pages of each of them in the following way: having chosen a starting page he will then move to another page by choosing randomly from the list of possible links on his current page.

a Show the information from the table in a directed graph.

b From the graph explain why it is not possible for George to visit each of the friends' pages in the first six pages he looks at if he begins on Antoine's page.

c Determine whether it is possible to visit each of the friends' pages in the first six visits if he begins on another page. If so, give a possible order; if not, say why not.

d Construct a transition matrix for your graph.

George begins by visiting Antoine's page.

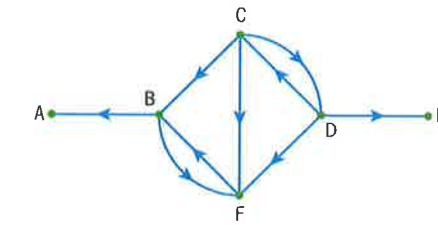
e Find the probability he is back on Antoine's page after visiting five further pages.

f Determine the probability he visited the pages in the order Antoine, Belle, Charles, Dawn, Emil, Antoine.

g George continues moving through the web pages in the same way for a long time. Find which page he is likely to visit:

i the most **ii** the least.

5 **a** Explain why the long-term probabilities for this graph will not be independent of the starting position.



b Write down the probability that a random walk would end at vertex A if it began at vertex B.

c If X is the number of steps it would take for the walk to reach A from B find $E(X)$.

d Find the transition matrix for the graph.

e By evaluating high powers of the transition matrix find the long-term probability of being at vertex E if you began at vertex C.

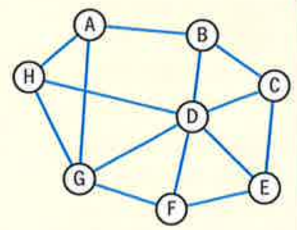
Developing inquiry skills

In the opening problem, what is the minimum number of trips needed to travel between any two of the viewpoints?

A walker got lost in the park and was randomly travelling along trails. After a long period of time a rescuer went out to try and find him. Given that the rescuer decides to wait at one of the viewpoints, which one should they choose if they want to maximize the chance that the walker will pass through that viewpoint first.

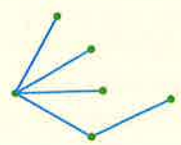
15.3 Graph theory for weighted graphs: the minimum spanning tree

A **cycle** is a walk that begins and ends at the same vertex and has no other repeated vertices.
In the graph shown ABDGHA is a cycle.



A **tree** is a connected graph which has no cycles, such as the one shown on the right.

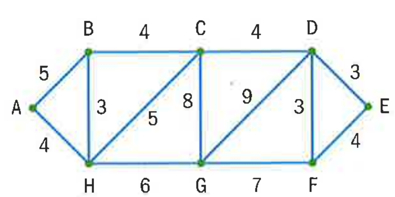
A **spanning tree** for a graph is a subgraph which is a tree and contains all the vertices in the graph.

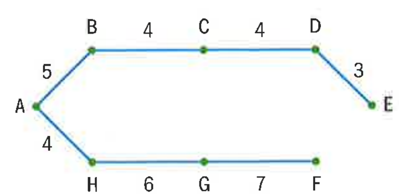


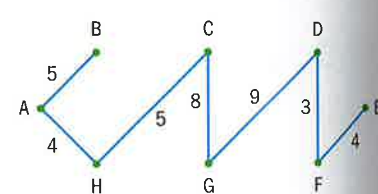
TOK

Is imagination more important than knowledge?

The two trees below are both spanning trees for the given graph.







Let the graph represent new houses built on an extensive estate and the weighted edges the costs (in €1000s) of connecting them to mains electricity.

If connected using the first tree the total cost would be €33 000, and with the second it would be €38 000.

The problem that needs to be solved is finding the spanning tree of least weight, which is normally referred to as the **minimum spanning tree**.

The minimum cost solution cannot contain any cycles. If there were a cycle the edge with greatest weight could be removed and the houses would still be connected.

Reflect What would be the least-weight way to connect all the vertices in a graph?

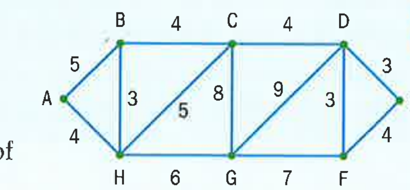
You need to learn two algorithms to find the minimum spanning tree, Kruskal's and Prim's algorithms.

Kruskal's algorithm

1 Find the edge of least weight anywhere in the graph. If there are two or more edges with the same weight any may be chosen.
2 Add the edge of least weight that has not already been selected and does not form a cycle with the previously selected edges.
3 Repeat the second stage until all the vertices are connected.

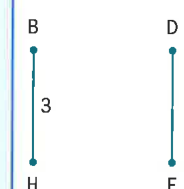
Example 6

Use Kruskal's algorithm to find the minimum cost of connecting all the houses on the estate. The costs of connecting each pair of houses are shown on the weighted graph. The weights are the costs in €1000 of connecting the houses.



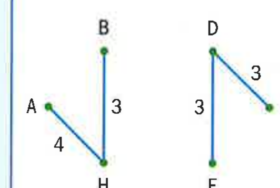
1 Find the edge of least weight anywhere in the graph. If there are multiple edges with the same weight any may be chosen.

In this example any of the edges of weight 3 can be selected.



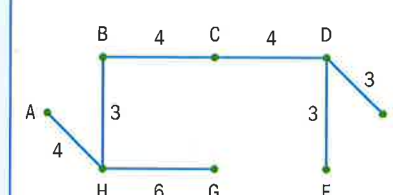
2 Add the edge of least weight that has not already been selected and does not form a cycle with the previously selected edge.

Add another of the edges of weight 3.



3 Repeat the second stage until all the vertices are connected.

The next edge added will be the third of the edges with weight 3. The fourth one cannot be EF as this would create a cycle and so AH, BC or CD is added.



All the edges of weight 4 are added next. Neither of the edges of weight 5 can be added as they both form cycles, so GH of weight 6 is added. All the vertices are now connected.

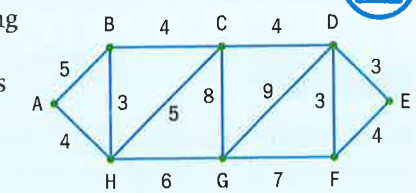
The minimum cost for connecting all the houses to electricity will be
$4 + 3 + 4 + 4 + 3 + 3 + 3 + 6 = 27$, ie €27 000

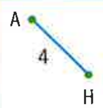
Prim's algorithm

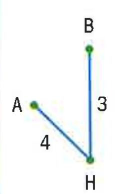
1 Select any vertex and add the edge of least weight adjacent to it.
2 Add the edge of least weight that is incident to the tree formed in first step and does not connect to a vertex already in the tree.
3 Repeat this process until all the vertices have been added.

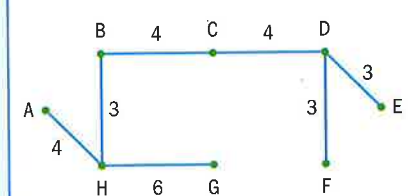
Example 7

Use Prim's algorithm to find the minimum cost of connecting all the houses on the estate. The costs of connecting each pair of houses is shown on the weighted graph. The weights are the costs in €1000 of connecting the houses.









The minimum cost for connecting all the houses to electricity will be

$4 + 3 + 4 + 4 + 3 + 3 + 6 = 27$, ie €27 000

1 Select any vertex and add the edge of least weight adjacent to it.

For example, if vertex A is selected it is connected to H.

2 Add the edge of least weight that is incident to the tree formed in the first step and does not connect to a vertex already in the tree.

If two vertices adjacent to the tree being formed have the same weight then either can be used.

In this example the next edge would be HB.

3 Repeat this process until all the vertices have been added.

The tree formed will be a minimum spanning tree for the graph.

The order in which the edges were added to the tree was AH, HB, BC, CD, DE, DF, HG.

All minimum spanning trees will have the same weight but may not consist of exactly the same edges.

For very large graphs Prim's algorithm is usually quicker than Kruskal's as there is no need to check whether adding an edge will form a cycle.

It is possible to use Prim's algorithm directly from a table without needing to draw the graph.

EXAM HINT

Unless told which algorithm to use, you can use either Prim's or Kruskal's algorithm in an exam.

Example 8

Find the minimum spanning tree for the graph represented by the table below.

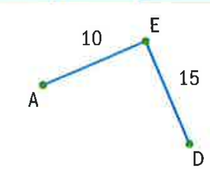
	A	B	C	D	E
A	0	40	45	25	10
B	40	0	25	15	30
C	45	25	0	20	35
D	25	15	20	0	15
E	10	30	35	15	0

Begin with vertex A.

Cross out the columns of vertices as they are added to the tree and indicate at the end of the rows the order in which they have been added. You can then look along these rows to find the entry with least weight. These entries should be circled.

		A	B	C	D	E
(1)	**A**	0	40	45	25	(10)
	B	~~40~~	0	25	15	~~30~~
	C	~~45~~	25	0	20	~~35~~
	D	~~25~~	15	20	0	~~15~~
(2)	**E**	~~10~~	30	35	15	~~0~~

		A	B	C	D	E
(1)	**A**	~~0~~	40	45	~~25~~	(10)
	B	~~40~~	0	25	~~15~~	~~30~~
	C	~~45~~	25	0	~~20~~	~~35~~
(3)	**D**	~~25~~	15	20	~~0~~	~~15~~
(2)	**E**	~~10~~	30	35	(15)	~~0~~



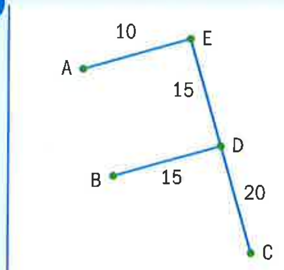
It is often easiest to draw the tree at each stage of the algorithm.

The sum of the circled entries will give the weight of the minimum spanning tree.

		A	B	C	D	E
(1)	**A**	~~0~~	~~40~~	45	~~25~~	(10)
(4)	**B**	~~40~~	~~0~~	25	~~15~~	~~30~~
	C	~~45~~	~~25~~	0	~~20~~	~~35~~
(2)	**D**	~~25~~	(15)	20	~~0~~	~~15~~
(3)	**E**	~~10~~	~~30~~	35	(15)	~~0~~

		A	B	C	D	E
(1)	**A**	~~0~~	~~40~~	45	~~25~~	(10)
(4)	**B**	~~40~~	~~0~~	25	~~15~~	~~30~~
(5)	**C**	~~45~~	~~25~~	0	~~20~~	~~35~~
(2)	**D**	~~25~~	(15)	(20)	~~0~~	~~15~~
(3)	**E**	~~10~~	~~30~~	35	(15)	~~0~~

Continued on next page



$0 + 15 + 15 + 20 = 60$

Reflect Why is Prim's a better algorithm than Kruskal's for a large graph or when the information is given in a table?

How should the algorithm be changed if extra restrictions are added, for example if there must be a direct connection between A and B?

Exercise 15F

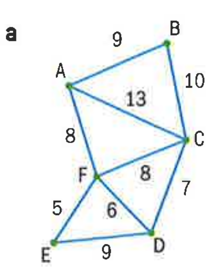
1 Use Prim's algorithm to find the minimum spanning tree for the following graphs, beginning with vertex A.

State clearly the order in which the edges are selected.

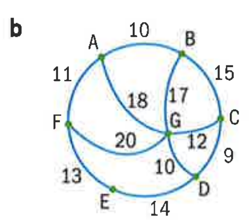
EXAM HINT

It is likely an exam question will require this as evidence you have performed the algorithm correctly.

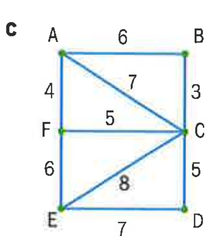
a



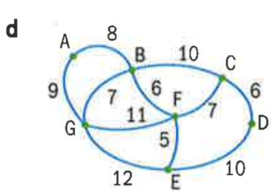
b



c



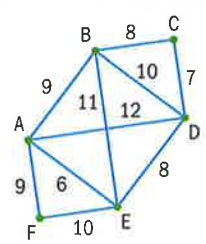
d



2 Use Kruskal's algorithm to find a minimum spanning tree for the graphs in question **1**. List the order in which the edges are selected.

3 Determine the number of edges there will be in a spanning tree containing v vertices Justify your answer through a consideration of the application of Prim's algorithm.

4 The following graph shows the offices of a finance company. The management wants to connect all the offices with extremely fast internet cables. The costs of doing so (in \$1000s) are indicated on the graph below.



a Use Kruskal's algorithm to find the minimum cost for connecting all the offices. List the order in which the edges are connected.

b **i** State how you would adapt the algorithm if it is essential for offices B and D to be connected.

ii Find the minimum cost in this situation.

5 Use Prim's algorithm to find the weight of the minimum spanning tree for the following tables.

a

	A	B	C	D	E
A	0	40	25	15	20
B	40	0	25	45	30
C	25	25	0	15	35
D	15	45	15	0	20
E	20	30	35	20	0

b

	A	B	C	D	E
A	0	4	7	9	12
B	4	0	8	5	3
C	7	8	0	10	5
D	9	5	10	0	6
E	12	3	5	6	0

c Comment on why it would be more difficult to use Kruskal's algorithm than Prim's algorithm when the information is given in a table.

6 The following table shows the distances between homes needing to be connected to a mains water source at A.

	A	B	C	D	E	F	G
A	0	20	17	23	11	10	15
B	20	0	9	16	21	15	10
C	17	9	0	22	16	10	12
D	23	16	22	0	19	13	18
E	11	21	16	19	0	16	12
F	10	15	10	13	16	0	25
G	15	10	12	18	12	25	0

a Find the minimum length of pipe needed to connect all the houses and draw the corresponding spanning tree.

b It is decided that A and B must be directly connected. Find the extra length of pipe that will be required.

7 **a** **i** Use Prim's algorithm to find the minimum spanning tree for the graph represented by the table shown.

ii State the weight of the spanning tree.

	A	B	C	D	E	F	G	H
A		10	7	12	9			
B	10		8	5				
C	7	8						
D	12	5			9		7	6
E	9			9		5	6	
F					5		3	
G				7	6	3		4
H				6			4	

b A ninth vertex, I, needs to be added to the tree. In the graph I is connected to vertex A and to vertex B. By consideration of your tree from part **a** find the weight of the new minimum spanning tree if:

i the weight of the edge connecting I to A is 9 and to B is 10

ii the weight of the edge connecting I to A is 5 and to B is 6.

Developing inquiry skills

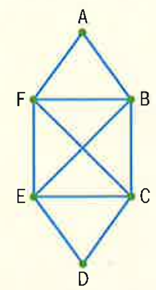
Look back to the opening question about the national park.

Extreme weather causes extensive damage to all of the trails. The park rangers decide to close some trails and repair the others to keep them open. All viewpoints still need to be accessible, and repair costs need to be kept to a minimum. Which trails should they keep open? What assumptions can you make about how the cost of the repairs relates to the length of a trail?

Geometry and trigonometry

15.4 Graph theory for weighted graphs: the Chinese postman problem

Investigation 3



1 Find four different ways to draw this graph without taking your pen off the paper. Begin from at least two different vertices.

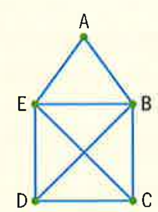
2 a What do you notice about the degree of each vertex?

b What do you notice about the starting vertex and finishing vertex?

c i **Conceptual** From the above results conjecture a sufficient condition for being able to traverse all the edges in a graph exactly once, ending at the vertex at which you began.

ii Verify your conjecture on a few graphs of your choosing.

iii Explain in your own words why your conjecture might also be a necessary condition.



3 Is it possible to draw this graph without taking your pen off the paper? If so, find at least two ways this can be done, starting from at least two different vertices.

4 a What do you notice about the starting and finishing vertices and the degrees of the vertices in the graph?

b i **Conceptual** Conjecture necessary conditions for being able to draw a graph without taking your pen off the paper, ending at a different vertex from the one at which you began.

ii Draw some graphs to verify your conjecture.

iii Justify your conjecture.

International-mindedness

The Chinese postman problem was first posed by the Chinese mathematician Kwan Mei-Ko in 1962.

A **trail** is a walk that repeats no edges. A **circuit** is a trail that starts and finishes at the same vertex.

An **Eulerian trail** is a trail that traverses all the edges in a graph, and an **Eulerian circuit** is a circuit that traverses all the edges in a graph.

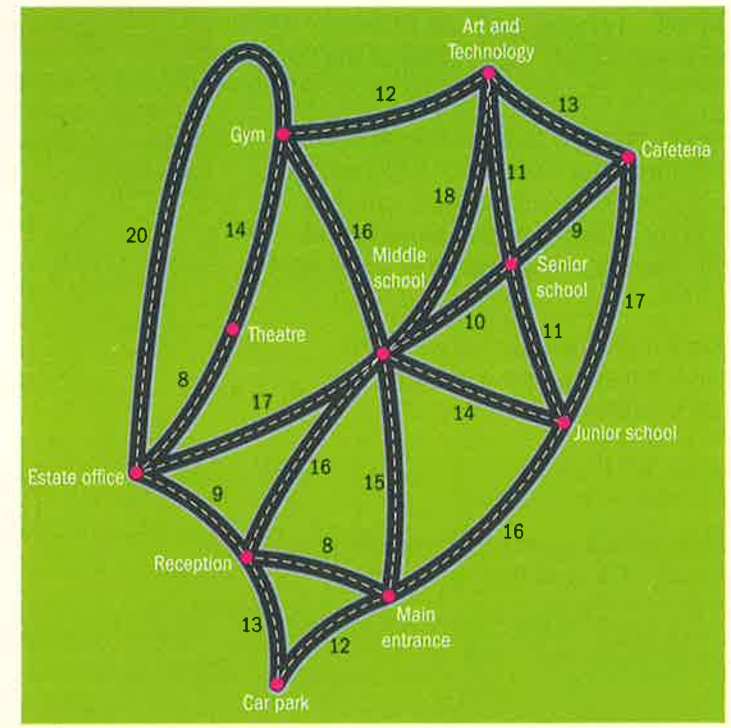
The results from Investigation 3 can therefore be summarized as follows:

- For a graph to have an Eulerian circuit all the vertices must have even degree.
- For a graph to have an Eulerian trail there must be exactly two odd vertices, and the trail must begin at one of these vertices and end at the other.

It can also be proved that every graph whose vertices are all even will have an Eulerian circuit and every graph with exactly two odd vertices will have an Eulerian trail.

Investigation 4

After a fall of snow the maintenance department in a school needs to clear all the paths before the students get to school. A graph of the school is shown. The weights on the graph show the time it would take to clear the paths. The time it takes to walk back along any cleared paths is one quarter of the time taken to clear them.



The maintenance department needs to begin and end their clearing at the estate office.

1 What would be the minimum time to clear all the paths if none of the paths needed to be repeated?

2 By considering the degrees of the vertices explain why it is not possible to clear all the paths without having to walk back along some of them.

3 The path between two buildings will need to be repeated; say which buildings these are.

4 In which order would you recommend the maintenance department clear the paths if they are to take the least possible time?

5 Given that there is an easy alternative to taking the path from the estate office to the gym (go via the theatre instead), it is decided that the direct path does not need to be cleared. How much time would be saved by not clearing this path?

The Chinese postman problem is to find the route of least weight around a weighted graph and to return to the starting vertex. The question above was an example of a Chinese postman problem, modified slightly by having a reduced time to walk along the cleared paths.

Another context for the problem is that of a postman seeking to find the shortest way to walk along all the streets on their round, repeating as little distance as possible.

Clearly if the graph consists of just even vertices, you can find an Eulerian circuit and the least weight will simply be the sum of the weights of all the edges.

Recall from the handshaking theorem that it is not possible to have an odd number of odd vertices, so the minimum number of odd vertices in a graph is two. If there are exactly two odd vertices, you can find an Eulerian trail from one vertex to the other that goes along all the edges. You then must return to the starting vertex using the route of least weight.

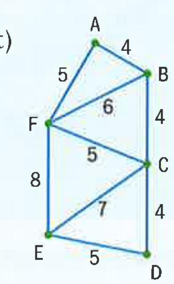
One way to think about this is to add a second edge between the vertices on the repeated route. This creates a graph with just even vertices, and hence there is an Eulerian circuit.

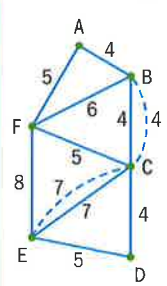
TOK

What is most important in becoming an intelligent human being: nature or nurture?

Example 9

Solve the Chinese postman problem (find a route of least weight) for this graph, beginning and ending at A, and find the total weight for this route.



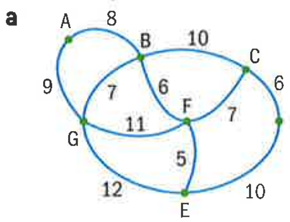
There are two odd vertices, B and E. 	Begin by considering the degrees of all the vertices. In this example there are two odd vertices. By inspection find the route of least weight between these two vertices. Here it is BCE. Add extra edges to your graph to show this route. Use this second graph to find an Eulerian circuit beginning and ending at A.
A possible solution to the Chinese postman problem is ABCDEFBCECFA.	There are many possible routes and it is normally relatively easy to find one by inspection.
The total weight is 48 + 11 = 59	The weight of this route will be the weight of all the edges in the original graph plus the weights of the repeated edges.

Reflect How do the conditions for the existence of an Eulerian circuit or trail help in solving the Chinese postman problem?

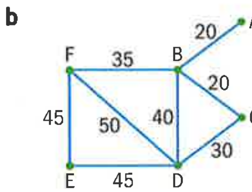
Exercise 15G

1 Solve the Chinese postman problem for the graphs below, starting at vertex A. State the edges that are repeated and the weight of the route taken.

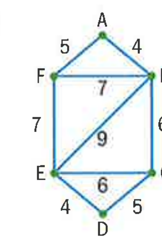
a



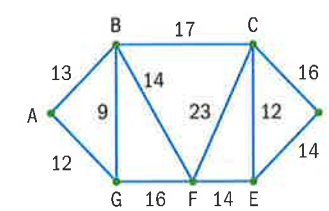
b



c



2 **a** Explain why the following graph has an Eulerian trail but not an Eulerian circuit.



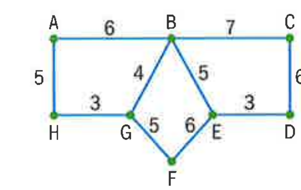
b The weights on the graph represent the times taken in minutes for a postman to walk along the streets on his round, which begins and ends at vertex A. If he walks down any street for a second time it takes one third of the time, as he does not need to deliver letters.

i Find the minimum length of time it could take for the postman to complete his round, and which streets would need to be walked along twice.

ii If the postman has to walk down both sides of every street explain why no extra streets will need to be covered.

Give a possible route he could take.

3 The following graph represents bus routes between towns, and the weights are the costs of the journeys in dollars. A tourist staying at town A wishes to travel along all the routes and return to A as cheaply as possible.



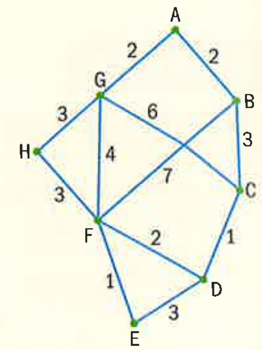
a Find the minimum possible total cost.

b A new direct bus route is added between towns G and E. Explain why the addition of this route will result in a lower cost for travelling all the routes and write down the new amount.

Investigation 5

1 The graph on the right has four odd vertices. List these vertices.

To solve the Chinese postman problem for this graph you will need to connect the odd vertices in pairs and find the route of least weight between them. You should then select the lowest of these as your solution.



2 List the three pairings of the four vertices and state the weight of the least-weight route connecting the vertices in each of the pairs.

3 Show, on a copy of the graph, the routes connecting the pairs which have the least total weight. Hence solve the Chinese postman problem, beginning and ending at vertex A.

4 Find the total weight of the route of least weight that will traverse all the edges in the graph at least once.

The requirement to start and finish at the same vertex is now removed.

5 Use the weights listed in question **2** to find where you should now choose to start and finish.

6 **Factual** How many extra possible routes do we need to consider when the graph has four vertices of odd degree?

7 **Conceptual** How does an understanding of an Eulerian trail help determine how the algorithm should be adapted if the trail does not have to begin and end at the same vertex?

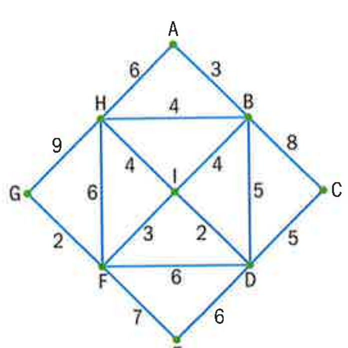
What does the solution to the Chinese postman problem represent?

Exercise 15H

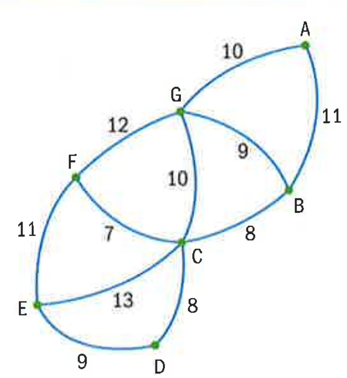
1 For each of the following graphs list the three pairings that connect two vertices of odd degree, and find the minimum weight of the walk between the vertices in each pair.

Hence find which edges need to be repeated in the solution to the Chinese postman problem.

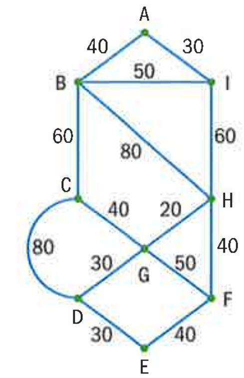
a



b



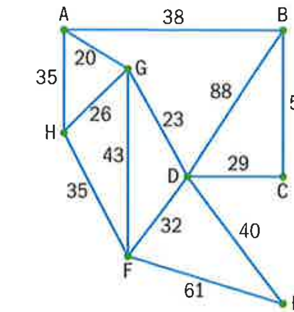
2 The graph below shows the lengths of connecting roads in a factory complex. Each evening a security guard walks along all of the roads and returns to his office at vertex A.



a Find a route which minimizes the length of his walk and state its length.

b A friend offers to drop him off at one of the vertices and to pick him up at another. State which two vertices should be chosen to minimize the distance the security guard has to walk.

3 The graph below shows the roads that need to be taken by a postman when delivering letters. The weights on the edges are the lengths of the roads in metres, and the postman needs to start and finish his deliveries at vertex A.



a Find a possible route he could take in order to minimize the distance he has to walk and state the length of this route. Fully justify your answer.

b A friend offers to pick him up from vertex H at the end of his round and take him back to A. Explain why this would not decrease the length he would have to walk.

c State where would you advise the postman to be collected in order to minimize the length he needs to walk, and find the total length of the repeated roads in this case.

Developing inquiry skills

Look back to the opening question about the national park.

What is the minimum distance you would need to cycle to travel along all the trails? Find one way of cycling along all the trails. Do you need to cycle along any trails more than once?

EXAM HINT

In an exam question there can only be 0, 2 or 4 vertices of odd degree.

Geometry and trigonometry

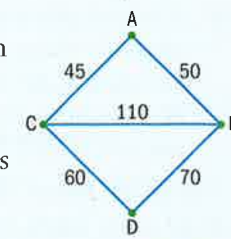
15.5 Graph theory for weighted graphs: the travelling salesman problem

A **table of least distances (weights)** shows the length of the shortest route between each pair of vertices in a graph.

Example 10

This graph shows the direct bus connections between four towns. The weights are the times of the journeys in minutes and the buses travel both ways along the routes.

Construct a table of least distances to show the shortest journey times between each pair of towns. Assume that there is no need to change buses on any of the journeys.



	A	B	C	D
A	0	50	45	105
B	50			
C	45			
D	105			

	A	B	C	D
A	0	50	45	105
B	50	0	95	70
C	45	95	0	60
D	105	70	60	0

From the graph we can see that the shortest routes from A to B and to C are just the direct connections, but from A to D the shortest route will be via C.

Because the graph is undirected both the first row and the first column can be completed.

Normally a direct route will be quicker than a route that passes through other vertices, but it is important to check. In this example it is shorter to go from B to C via A, which takes 95 minutes.

In the context it might be that the direct route is over poor roads or has more stops on the way.

A **path** is a walk which does not pass through any vertex more than once. A **cycle** is a walk that begins and ends at the same vertex, but otherwise does not pass through any vertex more than once.

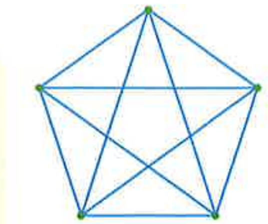
A **Hamiltonian** path or cycle is a path or cycle which passes through all the vertices in a graph.

A **complete graph** is one in which every vertex is directly connected to every other vertex.

TOK

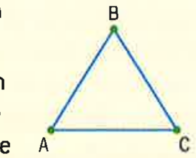
Hamiltonian paths and cycles are named after the 19th-century Irish mathematician William Rowan Hamilton.

The diagram on the right shows a complete graph with five vertices.



Investigation 6

There is just one Hamiltonian cycle in a complete graph with three vertices, ABC.



This is because a cycle which passes through the vertices in exactly the reverse order of another cycle is regarded as the same cycle. For example, CBA would be considered the same as ABC.

Any cycle with the vertices in the same order but beginning at a different vertex is also regarded as the same cycle. For example, BCA is the same as ABC.

Therefore, when counting cycles you can always start from the same vertex.

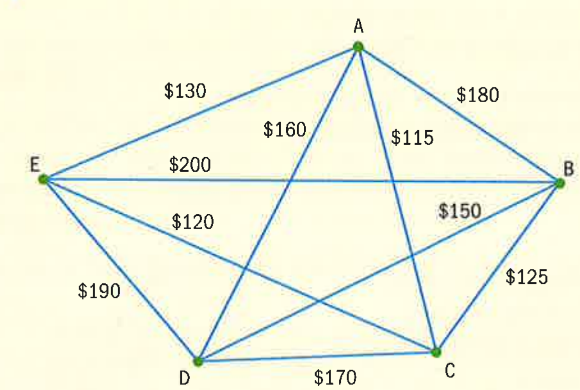
1 How many different Hamiltonian cycles are there in a complete graph with:
a 4 **b** 5 **c** 6 vertices?

2 **Factual** Write down a formula for the number of Hamiltonian cycles in a complete graph with n vertices.

3 There are about 10^{80} atoms in the universe. How many vertices would there need to be in a complete graph for the number of Hamiltonian cycles to exceed the number of atoms in the universe?

Before doing the calculation guess what you think the answer might be, and then compare this guess with the calculated value.

Investigation 7



In the diagram the vertices represent towns and the edges represent the costs of flying between those towns. A salesman wishes to visit all the towns and return to his starting point at A.

1 What is the minimum number of flights he would need to take?

2 Explain why the total cost of visiting all the towns would be greater than $640.

Continued on next page

Geometry and trigonometry

3 Find the weight of any Hamiltonian cycle beginning and ending at A.

4 Use your answers to questions **2** and **3** to give a lower and an upper bound for the cost of visiting all the cities.

5 In order to try and find a cheaper route the salesman decides that he will take the cheapest ongoing flight from each airport that takes him to a town he has not already been to and then back to A. Find the cost of the route if he follows this method beginning the algorithm at A.

6 Investigate the cost of visiting all the towns if he begins the algorithm at a different town.

7 Which of the routes you found would you recommend he take, beginning at town A?

8 Can you find a cheaper route? How likely do you think it is to be the cheapest? Justify your answer.

The **classical travelling salesman problem (TSP)** is to find the Hamiltonian cycle of least weight in a complete weighted graph.

The **practical TSP** is to find the walk of least weight that passes through each vertex of a graph, starting and finishing at the same vertex. In this case the graph need not be complete.

One context for the practical TSP would be a salesman needing to visit a number of towns and return to his starting point in the shortest possible time.

From the investigation you will have realized that finding the weight of each cycle individually to see which is shortest would take a considerable length of time.

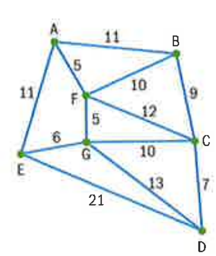
Unfortunately, it has been shown that there is no other simple way to guarantee that a route you have found is indeed the shortest.

Instead you need to find **upper and lower bounds** for the solution which will give a range of values in which the solution must lie.

One procedure for doing so is illustrated below.

Finding an upper bound

The graph shows the connections between seven cities. A driver needs to make deliveries to all of these cities, beginning and ending at city A. The weights are the driving times between the cities, measured in hours.



Any route that passes through all the vertices and returns to the starting point would be an upper bound for the solution to the TSP as the shortest route cannot have a greater weight than this route.

For example, by inspection ABFCDGEA is a cycle so its length would be an upper bound, namely $11 + 10 + 12 + 7 + 13 + 6 + 11 = 70$ hours. It is unlikely though that this is the best possible upper bound.

The **best upper bound** is the upper bound with the smallest value.

A method that is likely to give a better upper bound is the **nearest neighbour algorithm**.

The algorithm has two stages:

1 If the graph is not complete, create a table of least distances (weights) showing the shortest route between each pair of vertices.

2 Using your table (or the equivalent graph) choose a starting vertex and move around the graph, always going to the nearest vertex that has not already been included in the cycle. Once all the vertices have been visited, the shortest route back to the starting point is taken.

Completing a table of least distances converts a practical TSP into a classical TSP because it ensures that the graph is complete and so a Hamiltonian cycle exists.

The algorithm should only be used on a complete graph that satisfies the triangle inequality (which means the direct route between two adjacent vertices is always the shortest route). This will always be the case when working with a graph or table that shows the least distances.

EXAM HINT

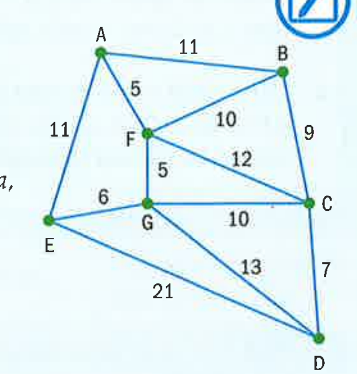
In exams if there is more than one route between vertices the route of least weight between them is found by inspection.

If there is a direct route between two vertices this is often the shortest route, but it does not have to be.

Geometry and trigonometry

Example 11

This graph shows the connections between seven cities. A driver needs to make deliveries to all of these cities, beginning and ending at city A. The weights are the driving times between the cities, measured in hours.



a The table of least distances is given below. Find the values of a, b and c.

	A	B	C	D	E	F	G
A	0	11	a	23	11	5	10
B	11	0	9	16	b	10	15
C	a	9	0	7	16	12	10
D	23	16	7	0	c	18	13
E	11	b	16	c	0	11	6
F	5	10	12	18	11	0	5
G	10	15	10	13	6	5	0

Continued on next page

b Use the nearest neighbour algorithm, beginning at town A, to find an upper bound for the length of the driver's journey.

c If the driver were to follow this route, which towns would he pass through more than once?

a $a = 17, b = 21, c = 19$

The values are taken by inspection from the graph. It is important to check alternative routes as the most obvious is not always the shortest. For example, the route from B to E is longer going along the two outside edges than it is along the three edges through the centre of the graph.

Note also that the shortest route from D to E is not the direct route.

b

	A	B	C	D	E	F	G
A	~~0~~	11	a	23	11	5	10
B	~~11~~	0	9	16	b	~~10~~	15
C	~~a~~	9	0	7	16	~~12~~	10
D	~~23~~	16	7	0	c	~~18~~	13
E	~~11~~	b	16	c	0	~~11~~	6
F	~~5~~	10	12	18	11	~~0~~	5
G	~~10~~	15	10	13	6	~~5~~	0

When using the algorithm on a table of least weights it is a good idea to cross off the columns of the vertices already visited and to use the rows to find the nearest neighbours.

The final solution is:

AFGECDBA with weight
$5 + 5 + 6 + 16 + 7 + 16 + 11 = 66$

So beginning at A the shortest distance is to F. These two columns are crossed out as shown and the F row is used to find the next edge of least weight, which is [FG].

Hence an upper bound for the length of time the driver would take is 66 hours.

After reaching the final vertex you must return from this vertex to A.

c The actual route would be AFGE**G**CD**C**BA and so towns G and C would be passed through twice.

Though no vertices are repeated in the table, some of the shortest routes between towns pass through other towns. These need to be included if the driver's route has to be given.

Reflect What algorithms can you use to find an upper bound for the TSP for a particular graph?

TOK

How long would it take a computer to test all of the Hamiltonian cycles in a complete weighted graph with just 30 vertices?

Exercise 15I

Keep your answers to this exercise available as they will be needed for Exercise 15J.

1 For each of the following graphs:

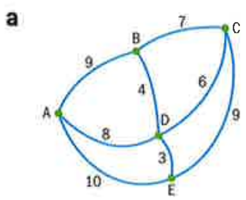
a Construct a table of least weights.

b Hence use the nearest neighbour algorithm to find a Hamiltonian cycle around the complete graph and also the route that would need to be taken around the original graph.

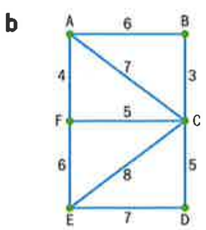
State the total weight of the route.

If more than one route is possible give the upper bound for both routes and select the "better" upper bound.

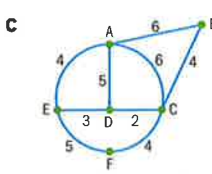
a



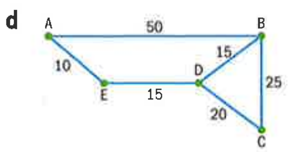
b



c

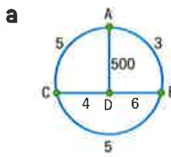


d

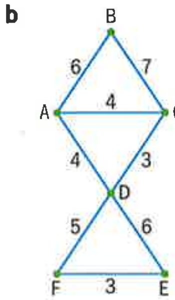


2 The following diagrams illustrate why the nearest neighbour algorithm should always be done on a complete graph which satisfies the triangle inequality. In each case apply the algorithm directly to the graph and comment on your findings.

a



b



Finding a lower bound

For a graph with v vertices a solution to the TSP will need to traverse at least v edges, so one lower bound would clearly be v multiplied by the weight of the edge with least weight or the sum of the v lowest weights. In most cases this is not likely to be a good lower bound.

An alternative is to use the **deleted vertex algorithm**.

The procedure is as follows.

1. Choose a vertex and remove it and all the edges incident to it from the graph.
2. Find the minimum spanning tree for the remaining subgraph.
3. A lower bound is the weight of the minimum spanning tree plus the combined weight of the two edges of least weight removed in step 1.
4. The process can then be repeated by removing a different initial vertex. The best lower bound (the one with the largest weight) is then taken.

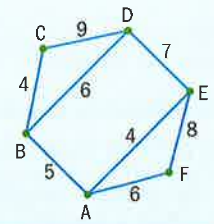
The example below includes an explanation of why this algorithm gives a lower bound.

HINT

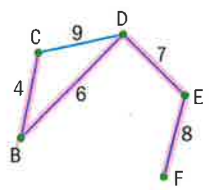
This algorithm should also be used only on a complete graph that satisfies the triangle inequality. This will always be the case when working with a graph or table showing least distances.

Example 12

Find a lower bound by deleting vertex A from the graph.

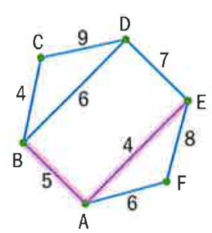


With vertex A deleted, the minimum spanning tree is



This has weight 25.

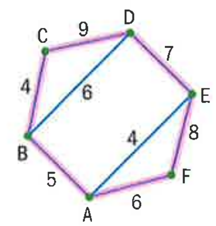
The two deleted edges of least weight have weights 5 and 4.



Lower bound is 25 + 5 + 4 = 34

Explanation of why this will give a lower bound:

The actual solution to the TSP is



When A is deleted the remaining part of the solution to the TSP will clearly be a tree, and so the weight of this part will be greater than or equal to the weight of the minimum spanning tree. In this example it is 28.

A is reconnected using the two edges of least weight. The combined weight of the two edges incident to A in the solution of the TSP will be greater than or equal to the combined weight of these two edges. In this example they add to 11.

The total weight of edges produced using the algorithm will therefore be less than or equal to the weight of the solution of the TSP, which in the example is 28 + 11 = 37.

Reflect How do the algorithms used in the travelling salesman problem help address the fact that it is often not possible to test all the possible routes?

How can you make sure that the nearest neighbour algorithm and the deleted vertex algorithm will give the upper and lower bounds for a practical TSP?

HINT

If the weights are given in a table then Prim's algorithm can be used to find the minimum spanning tree, having removed the rows and columns of the deleted vertex.

TOK

Could we ever reach a point where everything important in a mathematical sense is known?

What does it mean to say that the travelling salesman problem is "NP hard".

Is there any limit to our mathematical knowledge?

Exercise 15J

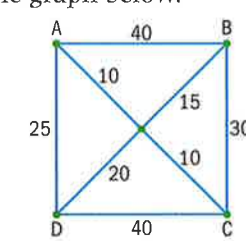
1 For each of the tables of least weights produced for the graphs given in Exercise 15I, question **1**:

i find a lower bound using the deleted vertex algorithm and deleting vertex A

ii find a higher lower bound by deleting a different vertex

iii use the upper bound calculated in Exercise 15I plus your answer to part **ii** to give bounds for the solution of the TSP.

2 Answer the following questions which relate to the graph below.



a By deleting vertex E on the graph, use the deleted vertex algorithm to find a lower bound for the TSP.

b Complete a table of least distances for this graph.

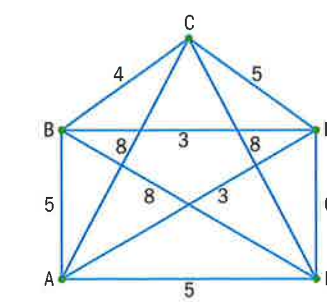
c Use the nearest neighbour algorithm to find an upper bound for the TSP.

d State what you notice and explain the reason for this.

3 In Example 10 an upper bound was found for the time it would take for a driver to travel between seven cities. The table of least distances is shown. Use the table and Prim's algorithm to find a lower bound for the above problem, by first deleting vertex A.

	A	B	C	D	E	F	G
A	0	11	17	23	11	5	10
B	11	0	9	16	21	10	15
C	17	9	0	7	16	12	10
D	23	16	7	0	19	18	13
E	11	21	16	19	0	11	6
F	5	10	12	18	11	0	5
G	10	15	10	13	6	5	0

4 The complete graph below shows the times in minutes to walk between five shops in a city.



a Use the nearest neighbour algorithm to find an upper bound for the length of time to walk to all of the shops, starting and ending at A.

b Use the deleted vertex algorithm to verify that 21 minutes is a lower bound for the time to walk to all the shops.

c By consideration of the edges adjacent to C and to E explain why it is not possible to find a cycle with length 22 or less.

d The shopper decides they still need to start and finish at shop A but their last visit before returning to A has to be to shop B. State how the nearest neighbour algorithm might be adapted to allow for this.

Geometry and trigonometry

Chapter summary

- A graph is a set of vertices connected by edges.
- When raised to the power n an adjacency matrix gives the number of walks of length n between two vertices.
- The transition matrix for a graph gives the probability of moving from one vertex to another if all edges from the vertex are equally likely to be taken.
- The steady states of a transition matrix give the proportion of time that would be spent at each vertex during a random walk around the graph.
- Kruskal's and Prim's algorithms are used to find the minimum spanning tree for a graph.
- The Chinese postman problem is to find the walk of least weight that goes along every edge at least once.
- In the Chinese postman problem for a graph with two odd vertices, the shortest route between the two odd vertices is found and indicated using multiple edges. The solution to the Chinese postman problem will be an Eulerian circuit around the graph formed.
- With four odd vertices the shortest total route between any two pairs of the odd vertices is found and indicated using multiple edges. The solution to the Chinese postman problem will be an Eulerian circuit around the graph formed.
- The classical travelling salesman problem is to find the Hamiltonian cycle of least weight in a weighted complete graph.
- The practical travelling salesman problem is to find the route of least weight around a graph which visits all the vertices at least once and returns to the starting vertex.
- A practical travelling salesman problem should be converted to the classical problem by completion of a table of least distances where necessary.
- The nearest neighbour algorithm is used to find an upper bound for the travelling salesman problem.
- The deleted vertex algorithm is used to find a lower bound for the travelling salesman problem.

Developing inquiry skills

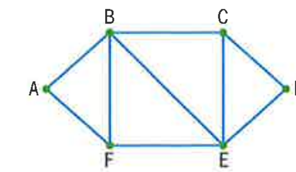
Look back to the opening questions about the national park. What would be the shortest route for a visitor wanting to visit all of the viewpoints and return to A. Approximately how long would their journey be?

Chapter review

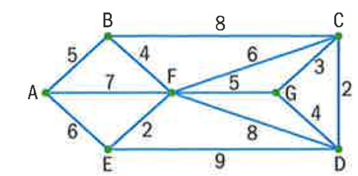
Click here for a mixed review exercise

1 **a** Explain why the graph shown has a Eulerian trail.

b Write down a possible Eulerian trail.

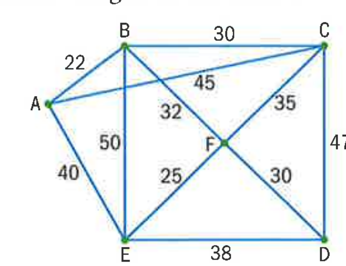


2 Use Kruskal's algorithm to find the minimum spanning tree for the graph shown. State the order in which the edges are added and the weight of the tree.



3 Solve the Chinese postman problem for the weighted graph shown, beginning at vertex A.

State which edges need to be repeated and the total weight of the route.



4 Consider the adjacency matrix M which represents the graph G.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 2 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

This question should be done **without** drawing the graph G.

a State whether G is:

i directed **ii** simple

iii complete **iv** a tree.

Justify your answers.

b State the degree of each of the vertices of G.

c Using your answer to part **b** state whether G has:

i a Hamiltonian cycle

ii an Eulerian circuit

iii an Eulerian trail.

Justify your answers.

d By considering powers of M state whether it is possible to move between all vertices of G in two or fewer steps.

Justify your answer.

5 **a** Use Prim's algorithm to find the minimum spanning tree for the graph represented by the table below.

State the weight of the minimum spanning tree.

	A	B	C	D	E	F	G
A	0	20	10	30	15	11	15
B	20	0	8	22	27	13	9
C	10	8	0	24	16	10	12
D	30	22	24	0	29	13	18
E	15	27	16	29	0	16	14
F	11	13	10	13	16	0	25
G	15	9	12	18	14	25	0

Another vertex H is now added to the graph. H is connected to A by an edge of weight 9, to B by an edge of weight 12 and to C by an edge of weight 11.

b Find a lower bound for the travelling salesman problem for this extended graph.

6 The weights of the edges of a graph with vertices A, B, C, D and E are given in the following table.

	A	B	C	D	E
A	0	16	10	25	15
B	16	0	8	22	27
C	10	8	0	24	16
D	25	22	24	0	19
E	15	27	16	19	0

Geometry and trigonometry

a **i** State whether or not the graph represented by the table is complete.

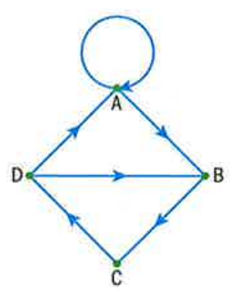
ii Draw the graph represented by the table.

iii Use the nearest neighbour algorithm, beginning at vertex A, to find an upper bound for the travelling salesman problem.

b **i** Use Kruskal's algorithm to find the minimum spanning tree for the subgraph obtained by removing vertex A from the graph and state the weight of this tree.

ii Hence find a lower bound for the travelling salesman problem.

7 The graph G is shown below.



a State whether G is:

i directed **ii** simple

iii strongly connected.

Justify your answers.

b For each vertex write down the in-degree and the out-degree.

c **i** State whether G has an Eulerian circuit.

ii Conjecture a necessary condition for a directed graph to have an Eulerian circuit.

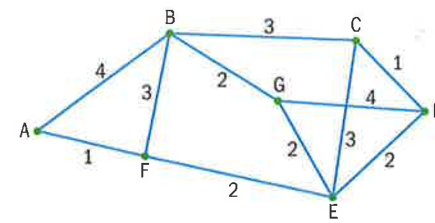
iii Does G have an Eulerian trail? If so write down a possible trail.

iv Conjecture a necessary condition for a directed graph to have an Eulerian trail.

d Construct a transition matrix for G.

e Find the steady state probabilities and comment on your results.

8 Consider the weighted graph below.



a Write down the four vertices of odd degree.

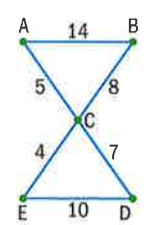
b Find a walk starting and ending at A, of minimum total weight, which includes every edge at least once, and state the weight of this walk.

Fully justify your answer.

c Find the minimum weight of a walk that includes every edge at least once if it is no longer necessary to start and finish at the same vertex.

State a possible starting and finishing vertex for this walk.

9 Let G be the weighted graph below.



a Explain why G does not have a Hamiltonian cycle.

b Give an example of a Hamiltonian path in G.

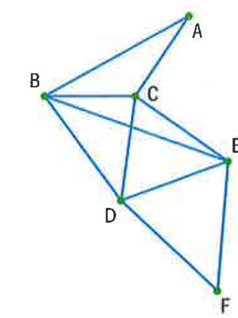
c Construct a table of least distances for G.

d Use the nearest neighbour algorithm to find a Hamiltonian cycle on the complete graph represented by the table of least distances which begins and ends at vertex A.

e If this cycle was taken around G, state the number of times the cycle passes through vertex C.

Exam-style questions

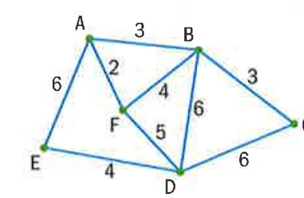
10 P1: Consider the graph below.



a State how you know a Eulerian circuit exists in the above graph. (1 mark)

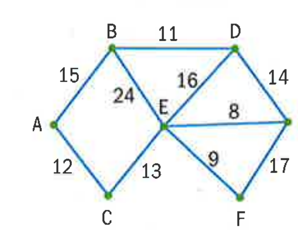
b Write down one such Eulerian circuit. (2 marks)

11 P1: Use Prim's algorithm (starting at A) to determine a minimum spanning tree for the graph. State clearly in which order you are adding the edges. (3 marks)



Determine the weight of the minimum spanning tree. (2 marks)

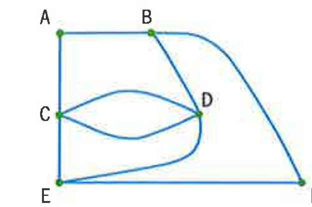
12 P1: The following graph shows seven points on a computer network, to be connected by electronic cable. The cost of connecting any two computers (in hundreds of pounds) is given by the number on each respective arc of the graph.



a By using Kruskal's algorithm, find the minimum spanning tree for the graph, stating the order in which you added the arcs. (2 marks)

b Draw the minimum spanning tree and hence find the least cost for connecting the computers. (4 marks)

13 P1: Consider the network.

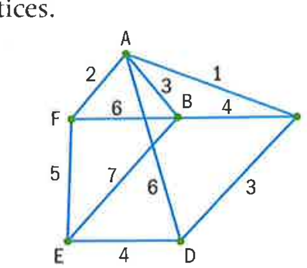


a Write down a Hamiltonian cycle, starting at the point F. (2 marks)

b Explain why it is not possible to construct a Eulerian circuit from this graph. (2 marks)

c Suggest which edge could be removed in order that it be impossible construct a Hamiltonian cycle from any point. Justify your answer. (2 marks)

14 P1: From the graph, write down all the odd vertices. (1 mark)



a Hence find a shortest route that starts and finishes at vertex A, travelling through every edge at least once. State also its weight. (6 marks)

15 P2: A network is represented by the following table.

	A	B	C	D	E
A	0	24	17	18	21
B	24	0	31	26	20
C	17	31	0	19	13
D	18	26	19	0	25
E	21	20	13	25	0

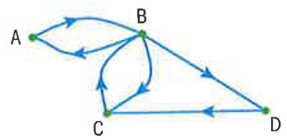
a Draw a weighted graph representing the network. (3 marks)

b Using Kruskal's algorithm, find a minimum spanning tree for the graph.

State clearly the order in which you add the edges. (2 marks)

c Hence find the weight of the minimum spanning tree. (2 marks)

16 P2: The following directed graph represents four islands served by a particular shipping company.



a Construct the adjacency matrix M for the shipping routes. (2 marks)

b **i** Find the matrix M^3.

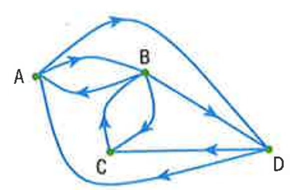
ii Hence determine the number of ways it is possible to travel from A → C in exactly three journeys. (3 marks)

c **i** Find the matrix S_3, where $S_3 = \sum_{i=1}^{3} M^i$.

ii Hence find the number of ways it is possible to travel from B → C in fewer than four journeys.

iii List all such possible ways. (8 marks)

17 P2: The shipping company from question 8 decides to introduce an express route from A → D and vice versa.



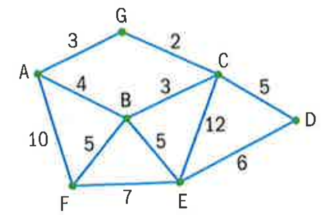
a Construct a transition matrix T for the new set of shipping routes. (3 marks)

b Show that a ship starting its journey from port B is just as likely to be stationed at any of the ports following three journeys. (3 marks)

c Find the steady state probabilities for the shipping network. (2 marks)

d Using your answer to part **c**, determine the best port for the shipping company to base its headquarters, justifying your answer. (2 marks)

18 P2: The graph shows seven schools, denoted by letters A, B, C, D, E, F, G.



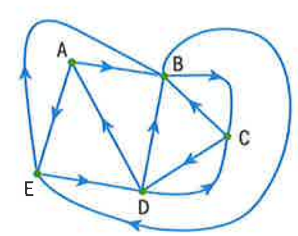
A school inspector starts from school A, and is required to visit each of the schools, returning to his starting point. The weights of each edge denote the travelling times for the inspector.

a Construct a table of least distances. (3 marks)

b By using the nearest neighbour algorithm, determine a best upper bound for the length of the journey. (3 marks)

c State the actual route taken by the inspector. (2 marks)

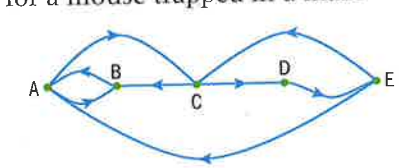
19 P2: Consider the graph below, where vertices A to E form a network of directed routes as shown.



a Construct the transition matrix for this graph. (3 marks)

b If a "random walk" was undertaken starting at vertex A, determine average time per hour (to the nearest minute) spent at each vertex. (4 marks)

20 P2: Consider the following network, illustrating the possible routes available for a mouse trapped in a maze.

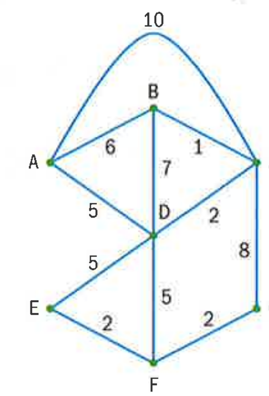


a Construct the adjacency matrix M for this network. (2 marks)

b **i** By evaluating the matrix M^7, show that there is only one journey in this network that can be done in exactly seven moves.

ii Describe the journey, listing (in order) the vertices visited. (6 marks)

21 P2: The diagram shows a network of cycle pathways joining seven points in a park, with the weights indicating the time of each journey (in minutes). Nasson aims to cycle along every pathway at least once in the shortest possible time. He starts from, and returns to, point A.



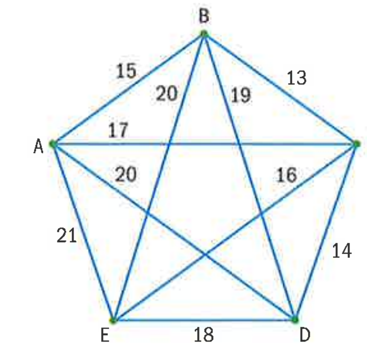
a State which pathways Nasson will need to cycle along twice, and find the total time that his journey will take. (7 marks)

b Hence, find a suitable route for Nasson to take. (2 marks)

c Instead of his original plan, Nasson decides to start at point F. He still needs to cycle every path in the least amount of time, but can finish at any other point.

Determine at which point Nasson should aim to finish, and justify your answer. (4 marks)

22 P2: Consider the following graph.



a By using the nearest neighbour algorithm, and starting at vertex A, determine an upper bound for the travelling salesman problem. State clearly the order in which you are taking the vertices. (3 marks)

b By deleting vertex A and using the deleted vertex algorithm, obtain a lower bound for the travelling salesman problem. (4 marks)

c Show that by deleting vertex B, the same lower bound will be found. (4 marks)

d Hence write down an inequality for the minimum length (l) of a tour. (2 marks)

e Write down a tour satisfying this inequality. (1 mark)

Click here for further exam practice

Modelling and investigation activity

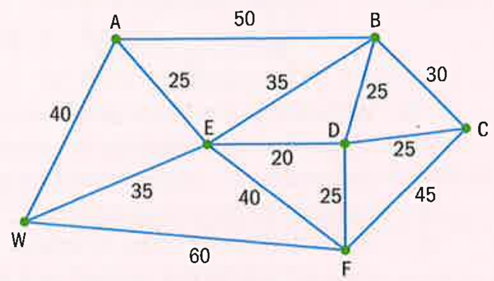
Choosing the right path

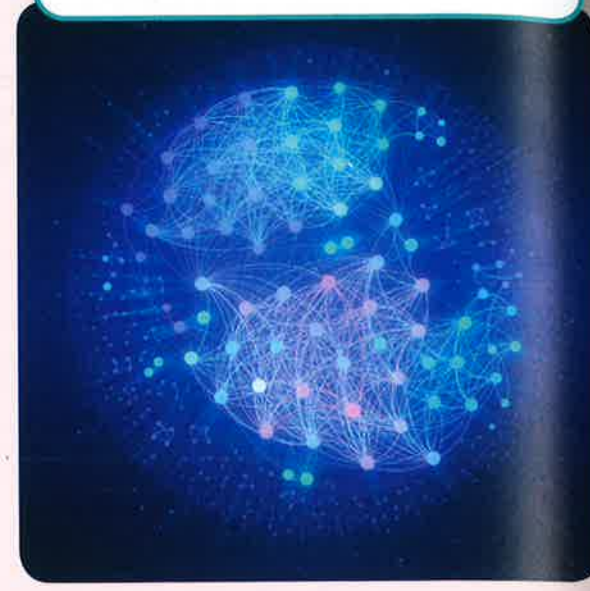
Approaches to learning: Communication, Research, Reflection, Creative thinking
Exploration criteria: Presentation (A), Mathematical communication (B), Personal engagement (C), Reflection (D)
IB topic: Graph theory

Minimum spanning trees

How many different spanning trees can you find for this graph?

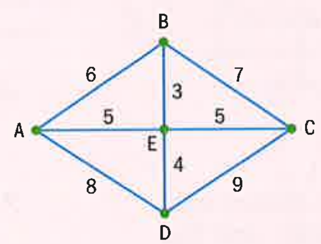
What is the minimum length spanning tree?





Hamiltonian paths and cycles (the travelling salesman)

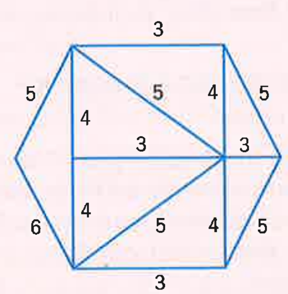
Find the minimum length Hamiltonian cycle on this graph:



Eulerian trails (The Chinese postman)

Find the minimum length Eulerian circuit on this graph.

What if you did not need to return to the starting vertex?



What real-life situations could you use these to solve?

In order to solve a real-life problem, you must:

- clearly define what it is that you want to find
- consider the method or algorithm you will follow to make this possible
- identify what the vertices and edges of the graph represent and find, record or collect the information required to put on this graph
- be able to draw a representation of the graph, perhaps using technology.

You might also:

- reflect on the accuracy and reliability of the values you have found and any assumptions or simplifications it has been necessary to make in order to collect them and finally answer the question
- consider what extensions you could solve based on what you have found out so far.

Your task

Choose one of these problems and think of a situation that is relevant to you that you could use this process to solve.

Based on this, try to give a brief answer to each of these questions:

- What real-life issue are you going to try to address?
- What is the **aim** of this exploration?
- What **personal** reason do you have for wanting to do this exploration?
- What **data** will you need to collect or find?
- What **research** will you need to do for this exploration?
- What **sources of information** will you use?
- What **definitions** will you need to give?
- What possible **representations** will you need to include?
- How will you **draw** these?
- What **technology** will you require?
- What **assumptions** or **simplifications** have you needed to make in this exploration?
- What possible **extensions** could there be to your exploration?
- Are there any **further information/comments** that may be relevant with regards to your exploration?

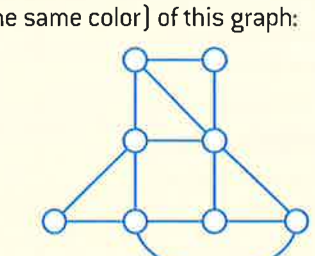
Extension

There are also other types of problems in Graph theory that have not been covered in this chapter. Here are two examples on vertex colouring and domination.

What real-life situations could you use these to solve?

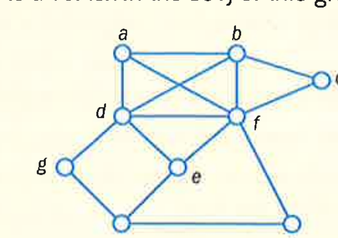
Vertex colouring:

Find the chromatic number (the smallest number of colours needed to colour the vertices so that no two adjacent vertices are the same color) of this graph:



Domination:

Find the domination number (the size of a smallest **dominating set** which is a set of vertices of the graph such that all vertices not in the set are adjacent to a vertex in the set) of this graph:



Modelling and investigation activity