GRAPHS AND NETWORKS

Euler's Rule

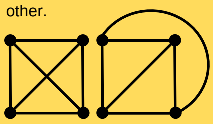
If a graph is planar, then:

$$V - E + F = 2$$

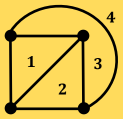
V: Number of vertices
E: Number of edges
F: Number of faces/regions

Planar Graphs

A graph is planar if it can be redrawn in such a way that no edges cross over each other.



Note: When counting faces in a planar graph, outside region counts as 1 face.



Including the outside face, total number of faces is 4. From Euler's rule, $4 - 6 + 4 = 2$, hence graph is planar.

Graph and Network Terminology

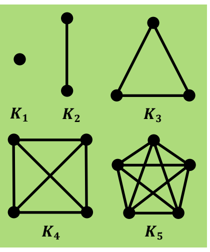
Loop: an edge that joins a vertex to itself (e.g. A and F).
Multiple Edges: two or more edges that have the same start and end vertices (e.g. B to C have two edges between them).
Isolated Vertex: a disconnected vertex in the graph (e.g. G).
Bridge: connects parts of a graph that would otherwise result in an isolated vertex or vertices (e.g. A to B and E to F).
Degree: Number of edges connected to a vertex. Loops are counted twice (e.g. $deg\ A = 3$, $deg\ C = 4$, $deg\ G = 0$).
Subgraph: A graph that has vertices and edges that are a subset of a larger graph (e.g. $D - C - E$ is a subgraph).
Simple Graph: a graph that has no loops or multiple edges.
Directed Graph (Digraph): a graph where all edges are directed from one vertex to another (shown by an arrow).
Weighted Graph: a graph whose edges has been assigned weights (weights are listed next to each edge in the graph).
Connected Graph: a graph with a path between every vertex.

Complete Graph (K_n)

A graph with n vertices in which every vertex is connected to every other vertex by one edge.

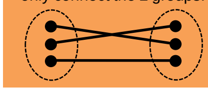
$$\#\ Edges\ in\ K_n = n(n-1)/2$$

Number of edges follow the triangular number sequence: 0, 1, 3, 6, 10, 15, 21, 28… (i.e. difference between each number is 1, 2, 3, 4, 5, 6, 7…)



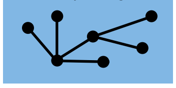
Bipartite Graph

A graph that has 2 sets of vertices. Any edges can only connect the 2 groups.



Tree

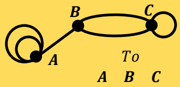
A connected graph that has no cycles or multiple edges.



Adjacency Matrix

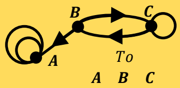
Matrix that shows how many times each vertex is connected (adjacent) to another vertex by a single path. Loops count as 1 edge in an adjacency matrix.

In an undirected graph, the matrix is symmetrical along the diagonal:



$$\begin{array}{c|ccc} & \multicolumn{3}{c}{To} \\ & A & B & C \\ \hline From\ A & 2 & 1 & 0 \\ B & 1 & 0 & 2 \\ C & 0 & 2 & 1 \end{array}$$

In a directed graph, the matrix is *not* symmetrical along the diagonal:



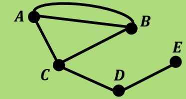
$$\begin{array}{c|ccc} & \multicolumn{3}{c}{To} \\ & A & B & C \\ \hline From\ A & 2 & 0 & 0 \\ B & 1 & 0 & 1 \\ C & 0 & 1 & 1 \end{array}$$

Tips to Find Shortest Path Between Two Vertices

Tip 1: Find all possible paths between each of the two vertices and test each path individually.
Tip 2: Where there are multiple edges, ignore the higher weighted edges.
Tip 3: Sometimes the shortest path doesn't mean the least amount of edges used; check all options

Transition Matrix

Any entries of 0 in an adjacency matrix to the n^{th} power means it's impossible to move to and from the points corresponding to that entry in n steps.



1 Step: adjacency matrix M

$$M = \begin{bmatrix} 0 & 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

It is *impossible* to go from A to A, B to B, C to C, D to A, D to B, D to D, E to A, E to B, E to C or E to E in 1 step.

An example of a 1 step option is that it is *possible* to go from A to B in 1 step in 2 different ways.

2 Step: adjacency matrix M^2

$$M^2 = \begin{bmatrix} 5 & 1 & 2 & 1 & 0 \\ 1 & 5 & 2 & 1 & 0 \\ 2 & 2 & 3 & 0 & 1 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

It is *impossible* to go from D to C, E to A, E to B or E to D in 2 steps.

An example of a 2 step option is that it is *possible* to go from A to A in 2 steps in 5 different ways.

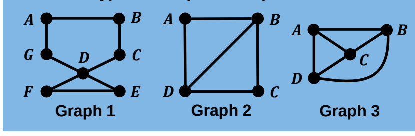
1 or 2 Step: adjacency matrix $M + M^2$

$$M + M^2 = \begin{bmatrix} 5 & 3 & 3 & 1 & 0 \\ 3 & 5 & 3 & 1 & 0 \\ 3 & 3 & 3 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

It is *impossible* to go from E to A or E to B in 1 or 2 steps.

An example of a 1 or 2 step options is that it is *possible* to go from A to C in 1 or 2 steps in 3 different ways.

Different Types of Graphs Examples



Using Graph 1, give an example of a:
- Eulerian Circuit: ABCDEFDGA

Using Graph 2, give an example of a:
- Semi-Eulerian Trail: BDABCD

Using Graph 3, give an example of a:
- Hamiltonian Cycle: ABCDA
- Semi-Hamiltonian Path: ABDC

Walks (Open and Closed)

Walk: alternating sequence of vertices/edges in a graph. Length of a walk is the number of edges used.
Open Walk: a walk that starts/ends on two different vertices. Closed Walk: a walk that starts/ends on the same vertex.

Different Types of Walks and Graphs

Rule	Open/Closed	Name	Also Called	Edges	Vertices
Uses only some vertices or edges in a graph	Open	Path	-	Can't Repeat	Can't Repeat
	Closed	Path	Cycle*	Can't Repeat	Can't Repeat
	Open	Trail	-	Can't Repeat	May Repeat
	Closed	Trail	Circuit	Can't Repeat	May Repeat
Rule	**Type**	**Name**	**Open/Closed**	**Edges**	**Vertices**
Uses all edges or vertices in a graph	Eulerian**	Circuit	Closed	All Edges in Graph Once	May Repeat
	Semi-Eulerian***	Trail	Open		
	Hamiltonian	Cycle	Closed	May Repeat	All Vertices in Graph Once
	Semi-Hamiltonian	Path	Open		

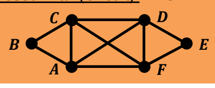
*Cycle: there is an exception; the starting and finishing vertex is allowed to repeat whilst all other vertices cannot.

**Eulerian: a graph is eulerian if every vertex has an even degree (the circuit starts and ends on any vertex).

***Semi-Eulerian: a graph is semi-eulerian if it has one pair of vertices with an odd degree (the trail starts and ends on either one of these two vertices).

Different Types of Walks Examples

Using the graph, give an example of a:
- Open Walk: BACAFDE
- Closed Walk: DCFACD
- Open Path: BADEF
- Closed Path (Cycle): BCDEFAB
- Open Trail: BCDFCA
- Closed Trail (Circuit): FACFDEF



BIVARIATE DATA

Types of Variables

- **Numerical or Quantitative** → Discrete, Continuous
- **Categorical or Qualitative** → Ordinal, Nominal

Numerical or Quantitative: have values that describe a measurable quantity as a number, like 'how many' or 'how much'.
Discrete: can take whole values (e.g. number of children or number of cars).
Continuous: can take any value (e.g. height, time and temperature).
Categorical or Qualitative: have values that describe a 'quality' or 'characteristic' of data.
Ordinal: observations that can logically ordered or ranked (e.g. academic grades such as A, B, C, D or clothing sizes such as small, medium, large).
Nominal: observations that cannot be ordered logically (e.g. eye colour, brand, gender, religion).

Response and Explanatory Variables

Response Variable	Explanatory Variable
Dependent Variable	Independent Variable
Vertical Axis (y-axis)	Horizontal Axis (x-axis)

Explanatory Variable causes the Response Variable (e.g. being immunized causes resistance to disease or number of books read causes reading speed to be low, medium or high).

Pearson's Correlation Coefficient (r)

The value r such that $-1 \leq r \leq 1$ measures the direction and strength of a linear relationship between two variables.

Coefficient of Determination (r^2)

The value r^2 such that $0 \leq r^2 \leq 1$ shows the percentage of the variation in the response variable with the variation in the explanatory variable. It shows what percent of the data that is the closest to the line of best fit (i.e. if $r^2 = 0.85$, then 85% of the data is close to the line of best fit). Also, r^2 is equal to Pearson's Correlation Coefficient squared.

Least-Squares Line/Line of Best Fit ($y = ax + b$)

A linear equation that summarises the relationship between two variables where a is the gradient of the line (calculated by $a = rise/run$) and b is the y-intercept.

Describing a Scatterplot

Form: The type of pattern that the points follow (i.e. linear or non-linear).
Direction: What direction the points tend towards (i.e. positive or negative).
Strength: How closely the points follow a linear pattern (i.e. perfect, strong, moderate, weak or no relationship).

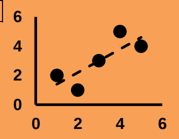
Value of r	Form	Direction	Strength
$\mathbf{r = 1}$	Linear	Positive	Perfect
$\mathbf{0.75 \leq r < 1}$	Linear	Positive	Strong
$\mathbf{0.5 \leq r < 0.75}$	Linear	Positive	Moderate
$\mathbf{0.25 \leq r < 0.5}$	Linear	Positive	Weak
$\mathbf{-0.25 \leq r < 0.25}$	None	None	None
$\mathbf{-0.5 \leq r < -0.25}$	Linear	Negative	Weak
$\mathbf{-0.75 \leq r < -0.5}$	Linear	Negative	Moderate
$\mathbf{-1 < r < -0.75}$	Linear	Negative	Strong
$\mathbf{r = -1}$	Linear	Negative	Perfect

Residuals

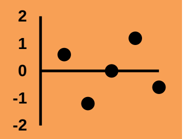
Residual formula: $e = y - \hat{y}$

e: is the residual
y: is the observed value (y co-ordinate from the data)
$\hat{y}$: is the predicted value (substitute x co-ordinate into line of best fit equation)

Scatterplot



Residual Plot

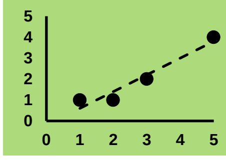


x	y	$\hat{y}$	e
1	2	1.4	0.6
2	1	2.2	-1.2
3	3	3	0
4	5	3.8	1.2
5	4	4.6	-0.6

Step 1: Create a scatterplot and determine the correlation coefficient and the line of best fit (i.e. line of best fit is $y = 0.8x + 0.6$ and $r = 0.8$).
Step 2: determine the residual using the formula $e = y - \hat{y}$ and create a residual plot.
Step 3: analyse residual plot (i.e. random pattern indicates linear model is a good fit).

Interpolation

Using the line of best fit to predict values that lie within the range of the data.



Line of best fit: $y = 0.8x - 0.2$

Estimating y when $x = 4$ can be determined by substituting $x = 4$ the line of best fit. This is considered interpolation as $x = 4$ is within the range of x values (0 – 5).

$y = 0.8(4) - 0.2 = 3$

$\therefore (4, 3)$ is the interpolated point.

Extrapolation

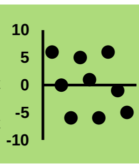
Using the line of best fit to predict values that lie outside the range of the original data. Not recommended as the nature of the data beyond what was recorded is unknown (especially if the correlation coefficient is weak). Using the graph on the left, estimating the value of y when $x = 10$ is considered extrapolation as $x = 10$ lies outside the range of x values (0 – 5).

$y = 0.8(10) - 0.2 = 7.8$

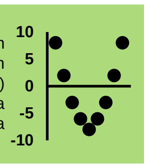
$\therefore (10, 7.8)$ is the extrapolated point.

Residual Plots

Random Pattern:
A random pattern in a residual plot indicates that the data is a good fit for a linear model.



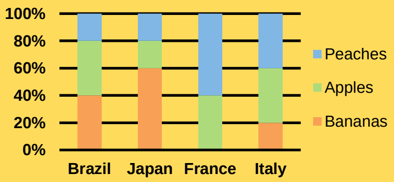
Non-Random Pattern:
A non-random pattern in a residual plot (such as a U-shaped pattern) indicates that the data is not a good fit for a linear model.



Stacked Column Graph

Data "stacked" on top of each other and totals to 100% on the y-axis. Below are the results of a survey taken from 4 countries that shows their preferred fruit.

To determine what percentage of each fruit that a country likes, find how wide each coloured column is by comparing it to the y-axis.



Breakdown of Percentages by Country:

Country	Peaches	Apples	Bananas
Brazil	20%	40%	40%
Japan	20%	20%	60%
France	60%	40%	0%
Italy	40%	40%	20%

Two-Way Table

Displays data between two variables. Below is a two-way table showing the popularity of apples, bananas and peaches among males and females:

Fruit	Male	Female	Total
Apple	20	40	60
Banana	90	110	200
Peach	50	70	120
Total	160	220	380

What % of apples are liked by males?

$$\frac{total\ likes\ of\ apples\ by\ males}{total\ apples} = \frac{20}{60} = 33.33\%$$

What % of males or females don't like peaches?

$$\frac{total\ likes\ of\ bananas\ and\ apples}{total\ males\ and\ females} = \frac{260}{380} = 68.42\%$$

Construct a table of percentages:

Fruit	Male	Female	Total
Apple	5.26%	10.53%	15.79%
Banana	23.68%	28.95%	52.63%
Peach	13.16%	18.42%	31.58%
Total	42.11%	57.89%	100%

Correlation Does Not Imply Causation

If two variables have a strong correlation between them, it does not necessarily mean that one variable causes the other variable in reality (e.g. if the variables *ice cream sales* and *number of deaths due to drowning* have a strong positive correlation coefficient of 0.9, it doesn't mean the two variables have a strong observable relationship in real life).

Causes of Incorrect Calculations of Pearson's Correlation Coefficient

- Coincidence: it could be a coincidence that data collected has a strong correlation (i.e. there is always the possibility that the data collected showed a strong correlation by random chance). To reduce the chance of a coincidence occurring, more data needs to be collected (at least 25 results).
- Confounding: a third variable that was failed to be taken into account had an influence between the two variables being tested (i.e. *ice cream sales* are impacted by another variable; the time of year, which will have an effect on the *number of deaths due to drowning* in the summer months).

It also works in reverse; just because two variables have a weak correlation, due to coincidence and confounding, the two variables may in fact have a strong observable relationship in reality.

SEQUENCES

Arithmetic and Geometric Sequences Formulae

Type	Explicit	Recursive	Sum of Series	Sum to Infinity
Arithmetic (+ or −)	$T_n = a + (n-1)d$	$T_{n+1} = T_n + d$, $T_1 = a$	$S_n = \frac{n}{2}(2a + (n-1)d)$	$S_\infty = \infty$ or $-\infty$
Geometric (× or ÷)	$T_n = ar^{n-1}$	$T_{n+1} = T_n \times r$, $T_1 = a$	$S_n = \frac{a(1-r^n)}{1-r}$	$S_\infty = \frac{a}{1-r}$

T_n: n^{th} term in the sequence
a: first term in the sequence (i.e. T_1)
r: common ratio between terms
d: common difference between terms
S_n: sum of the first n terms in the sequence
S_∞: sum of all possible terms in the sequence

Growth or Decay Sequences Formulae

Type	Explicit	Recursive
Growth (+)	$P_t = a(1+r)^t$	$P_{t+1} = (1+r)P_t$, $P_1 = a$
Decay (−)	$P_t = a(1-r)^t$	$P_{t+1} = (1-r)P_t$, $P_1 = a$

r: rate of growth or decay (as a decimal)
P_t: population at time t
a: initial amount (i.e. P_1)
t: time in years

Arithmetic Sequence Examples

Some values of an arithmetic sequence are <u>shown in the table below:</u>

n	4	5	6	7
T_n	21.5	24.2	26.9	29.6

<u>Find the explicit rule for the n^{th} term.</u>

Need to determine a and d:
Calculating a: $a = 21.5 - (3\times2.7) = 13.4$
Calculating d: $d = 24.2 - 21.5 = 2.7$
Substitute values into $T_n = a + (n-1)d$
Hence, $T_n = 13.4 + (n-1)\times2.7$

<u>Find the recursive rule for the $(n+1)^{th}$ term.</u>

From above, $a = 13.4$ and $r = 2.7$
Substitute values into $T_{n+1} = T_n + d,\ T_1 = a$
Hence, $T_{n+1} = T_n + 2.7,\ \ T_1 = 13.4$

Geometric Sequence Examples

Some values of a geometric sequence are <u>shown in the table below:</u>

n	3	4	5	6
T_n	0.5	2	8	32

<u>Find the explicit rule for the n^{th} term.</u>

$T_3 = ar^{3-1} = \frac{1}{2} \dots Equation\ 1$
$T_4 = ar^{4-1} = 2 \dots Equation\ 2$
Solve for a and r: $a = 0.03125$ and $r = 4$
Substitute into $T_n = ar^{n-1}$
Hence, $T_n = 0.03125\times4^{n-1}$

<u>Find the recursive rule for the $(n+1)^{th}$ term.</u>

From above, $a = 0.3125$ and $r = 4$
Substitute values into $T_{n+1} = T_n \times r,\ \ T_1 = a$
Hence, $T_{n+1} = 4T_n,\ \ T_1 = 0.03125$

Recurrence Relation Example

A recurrence relation is defined as:
$T_{n+1} = aT_n + b$ <u>for some value of a and b.</u>
<u>Find the recurrence relation of a sequence where the first three terms are 3, 4 and 7.</u>

3 —($\times a + b$)→ 4 —($\times a + b$)→ 7

From diagram above, create two equations that links T_1 with T_2 and T_2 with T_3.
$T_2 = aT_1 + b \rightarrow 4 = 3a + b \dots Equation\ 1$
$T_3 = aT_2 + b \rightarrow 7 = 4a + b \dots Equation\ 2$

Using ClassPad, solve $Equation\ 1$ and $Equation\ 2$ to find a and b: $a = 3$ and $b = -5$
Substitute into $T_{n+1} = aT_n + b,\ \ T_1 = 3$
Hence $T_{n+1} = 3T_n - 5,\ \ T_1 = 3$

Bouncing Ball Formulae

Below are shortcut formulae for the geometric sequence that models a ball dropped from an initial height a bouncing at r% efficiency.

Ball height after n^{th} bounce:
$Height = ar^n$

Total vertical distance travelled (S_∞):
$Distance = a\left(\frac{1+r}{1-r}\right)$

Vertical distance travelled up to n^{th} bounce:
$Distance = a\left(\frac{1+r-2r^n}{1-r}\right)$

r: bounce common ratio (as a decimal)
a: drop height
n: number of bounces

Long Term Steady State Solution

Two methods to find steady state solution:

- Substitute T_{n+1} and T_n with T and solve for T.
- Using ClassPad Sequences App, find a term for a large value of n (e.g. T_{50}) and look for a consistency.

Find the long term steady state solution for the sequence $T_{n+1} = 0.8T_n + 24,\ T_1 = 196$

- Solving: $T = 0.8T + 24$ gives $T = 120$
- ClassPad Sequences App: $T_{30} = 120.0941$ and $T_{50} = 120.0011$ which approaches 120

Simple Interest Formulae

$I = PRT$ $A = I + P$

A: amount (principal plus interest)
P: principal (starting amount)
I: total amount of interest
R: interest rate (as a decimal)
T: time in years

Simple Interest Example

Noah purchased an iPhone worth $600 using his credit card that charges <u>19.8% p.a.</u> simple interest on the <u>30th of March</u>. He paid the account on the <u>11th of April.</u>

<u>What is that total interest that was charged?</u>
$I = PRT = 600\times0.198\times\frac{13}{365} = \$4.23$

<u>What is the total amount Noah paid for the iPhone?</u>
$A = I + P = 4.23 + 600 = \$604.23$

Compound Interest Formulae

$A = P\left(1 + \frac{r}{n}\right)^{nt}$ $I = A - P$

A: amount (principal plus interest)
P: principal (starting amount)
I: total amount of interest
r: annual interest rate (as a decimal)
n: number of times interest is compounded per year
t: time in years

Compound Interest Recurrence Relation

$A_{n+1} = \left(1 + \frac{i}{n}\right)A_n + r,\ \ A_0 = P$

i: interest rate (as a decimal)
n: number of times interest is compounded per year
r: regular payments (for investments, r is positive and for loans and annuities, r is negative)
P: principal (initial amount)

Compound vs. Simple Interest

Simple interest has a linear pattern (meaning that interest is constant overtime).

Compound interest has an exponential pattern (meaning that interest increases overtime).

Compound Interest Example

Oliver borrowed $50,000 and makes monthly repayments of $1,120 to pay off the loan. Interest is 12% p.a. compounding monthly. Find the recurrence relation that shows amount owing.

$A_{n+1} = \left(1 + \frac{0.12}{12}\right)A_n - 1120,\ A_0 = 50000$

<u>How much does Oliver still owe after two years?</u>
$A_{24} = \$33{,}276.45$

<u>How much interest is charged during this period?</u> To calculate total interest, use formula: $I = A - P$ Total paid off loan $= 50000 - 33276.45 = \$16{,}723.55$ Total repayments $= 1120\times24 = \$26{,}880$
Total interest $= 26880 - 16723.55 = \$10{,}156.45$

EXTRA NOTES AND EXAMPLES