ENG1005 S2 2024 Workshop 4
Webpage Ranking

24 marks total

This problem set is intended for you to apply the mathematical skills you are learning. It is also designed to practice communicating your work clearly.

It is expected that you will use the workshop to develop (rough) solutions. During the workshop, you should discuss the problems with your peers and the academic staff who are there to assist you. In particular, if you are uncertain about what the problems are asking or you are stuck on a particular point, this is the time to get assistance. The time between the end of the workshop and when the solutions are due is only meant to be for writing up your solutions and for this you should not need more than an hour or two at most.

General submission information:

1. Electronic submission of your solutions is due on Moodle by **11:55 pm on Sunday of the same week**.
2. **Your solutions should include a description/explanation of what you are doing at each step and relevant working.** Without these you will receive limited marks. The description should be in complete English sentences. All mathematics should be appropriately laid out and with appropriate notation. Your writing should be similar in style to the worked solutions from the Applied Class problem sheets, not the annotations from the videos. For more information and advice, please read the "Guidelines for writing in mathematics" document posted under the "Additional information and resources" section of the ENG1005 Moodle page.
3. Your solutions may be typed or handwritten and scanned (the latter is encouraged). The **final document should be submitted as a single pdf file that is clearly and easily legible**. If the marker is unable to read it (or any part of it) you may lose marks.

Academic integrity:

You can (and should!) discuss your solutions with the other students, but **you must write up your solutions by yourself**. Copying solutions is serious academic misconduct and will be penalised according to Monash University guidelines. Other examples of academic misconduct include asking a personal tutor to do any of your assessments and posting your assessments to a "homework" website. Please refer back to your Academic Integrity module if you are in any doubt about what constitutes academic misconduct. **Your integrity is an important part of who you are. It is much more important than any grade you could receive.**

Webpage Ranking

Eigenvalues and eigenvectors of matrices are extremely important tools with applications in every field of engineering. In this workshop we will study the Google page rank algorithm which revolutionised the search engine, leading to Google becoming one of the largest companies in the world and almost monopolising the internet search market.

You are given the task of analysing internet traffic between Google, Intel, and Microsoft's webpages. After some data analysis, you find that for people visiting Google's page, after each hour, 20% of them would stay on Google's page, 40% would follow a link and go to Intel's page, and 40% to Microsoft's page. For those visiting Intel's page, 50% would stay, 10% would go to Google's page, and 40% to Microsoft's page. For those visiting Microsoft's page, 40% would stay, 30% would go to Google's page, and 30% to Intel's page.

1. Suppose initially there are 1000 visitors to each company's webpage. Assuming there are no additional visitors, how many people are on each company's webpage after 1 hour? **[1 mark]**

2. Let G_n, I_n, M_n denote the number of people on Google, Intel and Microsoft's webpage after n hours respectively, and $\mathbf{x}_n = \begin{bmatrix} G_n \\ I_n \\ M_n \end{bmatrix}$. Write down a matrix $\mathbf{A}$ such that $\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n$. This is called the transition matrix. **[2 marks]**

3. Using the transition matrix $\mathbf{A}$ and the help of Matlab or CAS, find out how many people are on each company's webpage after 3 hours. What about after 24 hours? Round your answer to the nearest integer. **[1 marks]**

4. Now suppose initially that Google's webpage receives 2000 visitors, and that each of Intel and Microsoft's webpages receives 500 visitors. How many people are on each company's webpage after 3 and 24 hours respectively? Round your answer to the nearest integer. **[1 marks]**

You notice that the numbers of visitors to the company's webpages seem to stabilise as time passes, regardless of where the 3000 visitors start initially! Let's investigate this further.

5. Show that if the number of visitors stabilises to some time independent constant $\mathbf{x} = \begin{bmatrix} G \\ I \\ M \end{bmatrix}$, then $\mathbf{x}$ is in fact an eigenvector of $\mathbf{A}$ with eigenvalue 1. **[2 marks]**

6. Write down the characteristic equation and find the eigenvalues of the transition matrix $\mathbf{A}$. **[2 marks]**

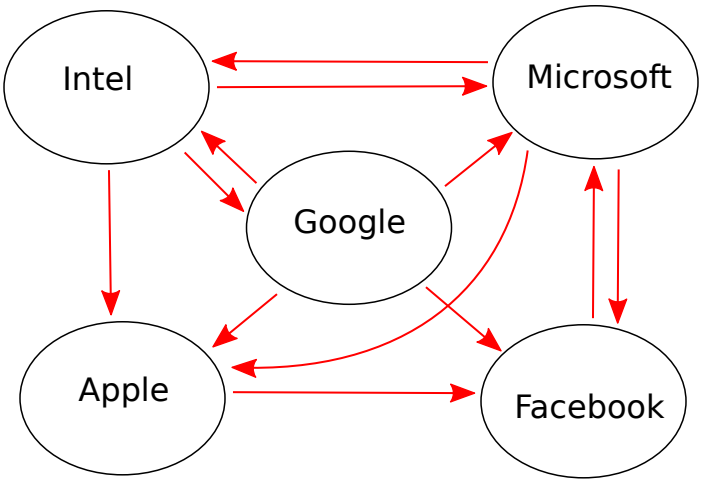
7. For each eigenvalue of $\mathbf{A}$, find the corresponding eigenvector(s). **[4 marks]**

8. Diagonalise the matrix $\mathbf{A}$, i.e., find a diagonal matrix $\mathbf{D}$ and an invertible matrix $\mathbf{V}$ such that $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$. **[2 marks]**

9. Show that $\mathbf{A}^n = \mathbf{V}\mathbf{D}^n\mathbf{V}^{-1}$ in terms of $\mathbf{D}$ and $\mathbf{V}$. **[1 mark]**

10. For the particular $\mathbf{D}$ you have, what happens to $\mathbf{D}^n$ as n becomes large? Use your observation and the previous part to explain why the numbers of visitors to the company's webpages stabilise as time passes, regardless of where the 3000 visitors start initially. **[3 marks]**

You can now analyse a more complicated network of webpages. The diagram below shows the webpages of Apple, Facebook, Google, Intel and Microsoft. A directed arrow indicates that there is a link on the starting webpage to the ending webpage. Suppose that after each hour, a person visiting a webpage has an equal chance of staying on the webpage or following any one of its links. For example, in this diagram, anyone visiting Google has a 20% chance of being on Apple, Facebook, Google, Intel and Microsoft's webpage after an hour.



11. Write down the transition matrix **A** for this network of webpages, ordering the companies alphabetically for your variables. **[1 mark]**

12. For the same reason as in Question (10), the number of visitors to each page will stabilise regardless of initial distribution. With the help of Matlab or CAS, rank the webpages by the stabilised number of visitors in descending order. **[2 marks]**

This is essentially Google's page ranking algorithm. It uses the links between webpages to determine which pages are likely be the most important and of the most interest to users. Of course, the actual algorithm is more sophisticated and remains proprietary, but the basic idea remain the same: it's simply eigenvalues and eigenvectors!

There is also 1 additional mark given for the quality of the English and 1 additional mark for correct mathematical notation. These marks are easy to obtain but the markers will be instructed to be strict in awarding these marks.