

project

March 14, 2019

1 MTH5001: Introduction to Computer Programming 2018/19

1.1 Final Report Project: "Networks"

1.1.1 Instructions:

First, please type your name and student number into the Markdown cell below:

Name:

Student number:

You must write your answers in this Jupyter Notebook, using either Markdown or Python code as appropriate. (You should create new code and/or Markdown cells in the appropriate places, so that your answers are clearly visible.)

Your code must be well documented. As a rough guide, you should aim to include one line of comments for each line of code (but you may include more or fewer comments depending on the situation). You should also use sensible variable names, so that your code is as clear as possible. If your code works but is unduly difficult to read, then you may lose marks.

For this project, you will need to use the Python package `NetworkX` extensively. However, to test your coding skills, in certain questions you will be restricted to using only specific functions. These restrictions are made clear below (see questions 4 and 8).

1.1.2 Submission deadline:

You must submit your work via QMPlus (to the "Final Report Project" assignment in the "Final Report Project" section).

The submission deadline is **11:55pm on Monday 29 April, 2019**. Late submissions will be penalised according to the School's [guidelines](#).

Your lecturers will respond to project-related emails until 5:00pm on Friday 26 April, 2019, only. You should aim to have your project finished by this time.

1.1.3 Marking:

The project is worth 70% of your final mark for this module.

The total number of marks available for the project is 100.

Attempt all parts of all questions.

When writing up your project, good writing style is even more important than in written exams. According to the [advice in the student handbook](#),

To get full marks in any assessed work (tests or exams) you must normally not only give the **right answers** but also **explain** your working clearly and give reasons for your answers by writing legible and grammatically correct English sentences. Mathematics is about logic and reasoned arguments and the only way to present a reasoned and logical argument is by writing about it clearly. Your writing may include numbers and other mathematical symbols, but they are not enough on their own. **You should copy the writing style used in good mathematical textbooks, such as those recommended for your modules.** **You can expect to lose marks for poor writing (incorrect grammar and spelling) as well as for poor mathematics (incorrect or unclear logic).**

1.1.4 Plagiarism warning:

Your work will be tested for plagiarism, which is an assessment offence, according to the [School's policy on Plagiarism](#). In particular, while only academic staff will make a judgement on whether plagiarism has occurred in a piece of work, we will use the plagiarism detection software "Turnitin" to help us assess how much of work matches other sources. You will have the opportunity to upload your work, see the Turnitin result, and edit your work accordingly before finalising your submission.

You may summarise relevant parts of books, online notes, or other resources, as you see fit. However, you must use your own words as far as possible (within reason, e.g. you would not be expected to change the wording of a well-known theorem), and you **must** [reference](#) any sources that you use. Similarly, if you decide to work with other students on parts of the project, then you **must** write up your work individually. You should also note that most of the questions are personalised in the sense that you will need to import and manipulate data that will be unique to you (i.e. no other student will have the same data).

1.2 Background information

In this project you will learn about **a field of mathematics called graph theory**. A **graph** (or **network**) is simply a collection of **nodes** (or **vertices**), which may or may not be joined by **edges**. (Note that this is not the same as the 'graph' of a function.)

Graphs can represent all sorts of real-world (and, indeed, mathematical) objects, e.g.

- social networks (nodes represent people, edges represent 'friendship'),
- molecules in chemistry/physics (nodes represent atoms, edges represent bonds),
- communications networks, e.g. the internet (nodes represent computers/devices, edges represent connections).

In this project we will only consider **undirected** graphs (see the above Wikipedia link for a definition).

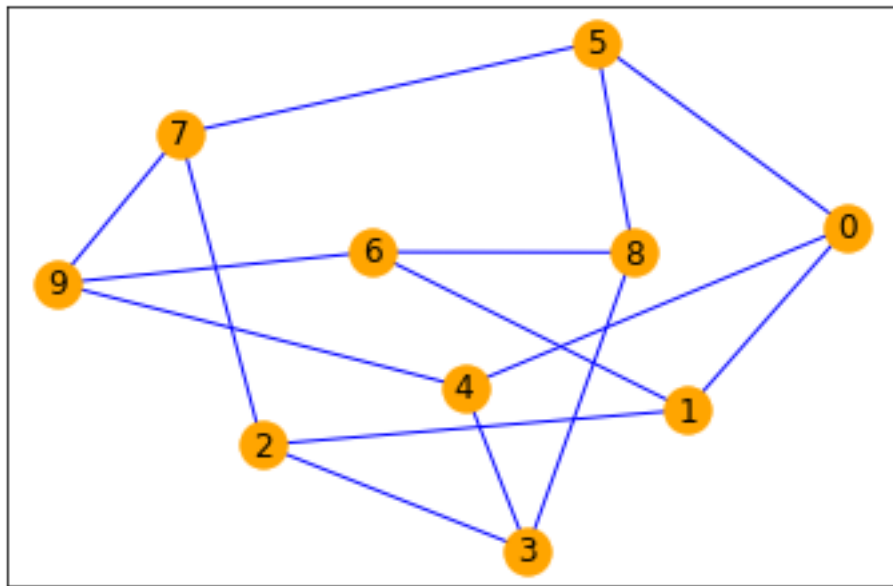
Conveniently, **Python has a package, called NetworkX, for constructing and analysing graphs**. Let's look at an example. Below we create the famous [Petersen graph](#) and use some basic NetworkX functions to learn a bit about it.

```
In [1]: # import NetworkX and other useful packages
import numpy as np
import numpy.linalg as la
import matplotlib.pyplot as plt
import networkx as nx
```

```
# create the Petersen graph, storing it in a variable called "PG"  
PG = nx.petersen_graph()
```

Before we doing anything else, it would make sense to draw the graph, to get an idea of what it looks like. We can do this using the NetworkX function `draw_networkx` (together with our old favourite, `matplotlib`).

```
In [2]: nx.draw_networkx(PG, node_color = 'orange', edge_color = 'blue', with_labels=True)  
plt.xticks([])  
plt.yticks([])  
plt.show()
```



We can see that the graph has 10 nodes, labelled by the integers $0, 1, \dots, 9$. It is also possible to label nodes with other data types, most commonly strings, but we won't do that in this project. The nodes of a graph can be accessed via the method `nodes()`:

```
In [3]: PG.nodes()  
Out[3]: NodeView((0, 1, 2, 3, 4, 5, 6, 7, 8, 9))
```

You can convert this to a Python list if you need to:

```
In [4]: list(PG.nodes())  
Out[4]: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

This (hopefully) makes it clear that the node labels do in fact have type `int`, at least in our example. You can also see from the picture that the graph has 15 edges. These can be accessed using the method `edges()`:

```
In [5]: PG.edges()
```

```
Out[5]: EdgeView([(0, 1), (0, 4), (0, 5), (1, 2), (1, 6), (2, 3), (2, 7), (3, 4), (3, 8), (4, 9)])
```

Again, you can convert this to a list if you need to (try it), and you will see that the elements of the list are tuples. In either case, if you compare the output with the picture, it should become clear what it means, i.e. two nodes labelled i and j are joined by an edge if and only if the pair (i, j) appears in `PG.edges()`.

So far we haven't said much about how graphs are related to **mathematics**. It turns out that a graph can be completely defined by its **adjacency matrix**. This is simply a matrix A defined as follows:

- A has size $n \times n$, where n is the number of nodes in the graph;
- if the nodes labelled i and j form an edge, then the (i, j) -entry of A is 1; if they don't form an edge, the (i, j) -entry of A is 0.

This idea is the foundation of **algebraic graph theory**, a field of mathematics used to study graphs by analysing certain matrices.

Not surprisingly, you can compute the adjacency matrix of a graph using an appropriate `NetworkX` function. Let's do this for the Petersen graph:

```
In [6]: A = nx.adjacency_matrix(PG)
```

Note that if you print this 'adjacency matrix' (try it), it doesn't actually look much like a matrix. This is because it doesn't have type `numpy.ndarray` like the matrices/arrays we've worked with in class:

```
In [7]: type(nx.adjacency_matrix(PG))
```

```
Out[7]: scipy.sparse.csr.csr_matrix
```

However, you can convert it to a `numpy.ndarray` by calling the method `toarray()`:

```
In [8]: A = A.toarray()
```

```
In [9]: type(A)
```

```
Out[9]: numpy.ndarray
```

After doing this, the adjacency matrix looks like you would expect, so you can use all the usual `numpy.linalg` functions on it:

```
In [10]: print(A)
```

```
[[0 1 0 0 1 1 0 0 0 0]
 [1 0 1 0 0 0 1 0 0 0]
 [0 1 0 1 0 0 0 1 0 0]
 [0 0 1 0 1 0 0 0 1 0]
 [1 0 0 1 0 0 0 0 0 1]
 [1 0 0 0 0 0 0 1 1 0]
 [0 1 0 0 0 0 0 0 1 1]
 [0 0 1 0 0 1 0 0 0 1]
 [0 0 0 1 0 1 1 0 0 0]
 [0 0 0 0 1 0 1 1 0 0]]
```

Make sure that you understand what all these 0's and 1's mean: the (i, j) -entry of the adjacency matrix is 1 if and only if the edges labelled i and j form an edge in the graph; otherwise, it is 0. For example (remembering that Python starts counting from 0, not from 1): the $(0, 4)$ entry is 1, and in the picture above we see that nodes 0 and 4 form an edge; on the other hand, the $(1, 7)$ entry is 0, and accordingly nodes 1 and 7 don't form an edge.

You will be working with matrices related to graphs quite a lot in this project, so before you begin you should make sure that you understand what the code we've given you above is doing. You may also like to work through the official [NetworkX tutorial](#) before attempting the project, bearing in mind that not everything in the tutorial is relevant to the project. (Alternatively, Google for another tutorial if you don't like that one.)

A final remark before we get to the project itself:

You can rest assured that the graphs we consider in this project all have the following nice properties:

- They are **connected**. This means that for every pair of nodes i and j , there is a 'path' of edges joining i to j . For example, the Petersen graph is connected, e.g. the nodes labelled 6 and 7 do not form an edge, but we can still reach node 7 from node 6 via the edges $(6, 9)$ and $(9, 7)$.
- They are **simple**. This means that there is never an edge from a node to itself.

You may come across these terms when you are researching the relevant mathematics for various parts of the project, so you should know what they mean.

1.3 The project

As we have already mentioned, in this project you will make extensive use of the Python package [NetworkX](#), which allows you to create and analyse graphs. You are expected to read the relevant parts of the NetworkX documentation, or otherwise learn how to use whatever Python functions you need to complete the project. However, the mini-tutorial which we have provided above should be enough to get you started.

You will also need to research and summarise some mathematics related to graphs, and to use your findings to write certain pieces of code 'from scratch', instead of using NetworkX functions. In these cases (questions 4 and 8), it is **strongly recommended** that you use NetworkX to check your answers.

You should structure your report as follows:

1.3.1 Part I: Data import and preliminary investigation [10 marks]

You have been provided with a Python file called "[data.py](#)" on QMPlus, which you should save in the same directory as this Jupyter notebook. This file contains a function `create_graph` which constructs a random graph that you will be analysing throughout the project. By following the instructions in question 1 (below), you will create a graph that is unique to you, i.e. no two students will have the same graph.

1. [5 marks] Execute the following code cell to create your graph, storing it in a variable called `G` (you can change the variable name if you like, but we recommend leaving it as it is). You **must** replace the number "123456789" with your 9-digit student number.

Important note: If you do not do this correctly, you will score 0 for this question, and if you are found to have used the same input as another student (rather than your individual student number), then your submission will be reviewed for plagiarism.

```
In [ ]: from data import create_graph

# Replace "123456789" below with your 9-digit student number

G = create_graph(123456789)

# Replace "123456789" above with your 9-digit student number
```

2. [5 marks] Draw your graph, and calculate how many nodes and edges it has.

1.3.2 Part II: Distance matrices and shortest paths [30 marks]

Many properties of graphs can be analysed by using matrices/linear algebra. The rest of your report will involve researching/summarising some of the relevant mathematics, and writing/using Python code to analyse certain properties of your graph from Part I. As explained above, you are allowed to summarise information from books and/or web pages, but you must use your own words and clearly reference any sources you use.

3. [10 marks] Explain what a "path" between two nodes in a graph is, and what the "distance" between two nodes is. Explain also what the "distance matrix" of a graph is, and how it can be computed using the adjacency matrix. Here you should discuss arbitrary (undirected, simple, connected) graphs, not your specific graph from Part I.

Note: You do **not** need to give any proofs, but you must reference any material you use, as explained in the plagiarism warning above.

4. [10 marks] Write a function `distance_matrix` which computes the distance matrix of a graph. Your function should return a matrix, represented as an array of type `numpy.ndarray`, of the same shape as the adjacency matrix of the graph. You may use the NetworkX function `adjacency_matrix` to compute the adjacency matrix of the input graph, but you **must not use any other NetworkX functions**.

5. [5 marks] Using your function from Question 4, find a pair of nodes (i, j) in your graph from Part I with the property that the distance from i to j is maximal amongst all pairs of nodes in the graph.

Note: This means that for every *other* pair of nodes (i', j') , the distance from i' to j' is less than or equal to the distance from i to j .

6. [5 marks] Find a shortest path between your nodes from Question 5, i.e. one with the shortest possible length, and re-draw your graph so that this path is clearly visible. You should use one colour for the nodes and edges in the path, and a different colour for all other nodes and edges.

Hint: You should be able to find a NetworkX function that computes a shortest path.

1.3.3 Part III: Laplacian matrices and spanning trees [30 marks]

So far you have learned about two matrices associated with a graph: the adjacency matrix, and the distance matrix. Now you will study a third matrix: the Laplacian.

7. [10 marks] Explain what the "degree" of a node in a graph is, what the "Laplacian matrix" of a graph is, what a "spanning tree" of a graph is, and how the Laplacian matrix can be used to calculate the number of spanning trees in a graph. Here, again, you should discuss arbitrary (undirected, simple, connected) graphs, not your specific graph from Part I.

Note: You do **not** need to give any proofs, but you must reference any material you use, as explained in the plagiarism warning above.

8. [10 marks] Write a function `number_of_spanning_trees` which takes as input a graph G and returns the number of spanning trees in G . You may use the NetworkX function `adjacency_matrix` to compute the adjacency matrix of the input graph, but you **may not use any other NetworkX functions**.

Note: You will probably need to compute the determinant of a certain matrix somewhere in your code. If you use the function `numpy.linalg.det` then your determinant will only be computed approximately, i.e. to a certain numerical precision. This is fine; you will not lose any marks if your code is otherwise correct.

9 [5 marks] Use your function from Question 8 to calculate the number of spanning trees in your graph from Part I.

10 [5 marks] Find a minimal spanning tree of your graph from Part I, i.e. one with the smallest possible number of edges. Re-draw your graph in such a way that this spanning tree is clearly visible. You should use one colour for the edges in the spanning tree, and a different colour for all other edges.

Hint: You should be able to find a NetworkX function that computes a minimal spanning tree.

1.3.4 Part IV: Eigenvalues and triangles [30 marks]

By now you have seen that certain matrices associated with a graph can tell us a lot about the structure of the graph. In this final part of the project, you will investigate this further, by learning how eigenvalues can be used to reveal even more information about graphs.

11. [5 marks] Explain what a "triangle" in a graph is, and quote a formula for calculating the number of triangles in a graph from the eigenvalues of the adjacency matrix.

Note: You do **not** need to give any proofs, but you must reference any material you use, as explained in the plagiarism warning above.

12. [5 marks] Calculate the number of triangles in your graph from Part I, using the formula discussed in question 11. Your answer **must** be an integer.

Hint: What is the "trace" of a matrix and how is it related to the eigenvalues?

13. [10 marks] Write a function `all_triangles` which finds all of the triangles in a graph. Use your function to count the number of triangles in your graph, and compare with your answer to question 12. (The two answers should, of course, be the same.)

Note: You will need to use your function in the next question, so you should think carefully about what kind of data structure you want it to output.

14. [10 marks] Re-draw your graph from Part I once more, so that all of its triangles are clearly visible. You should use one colour for the edges that appear in at least one triangle, and a different colour for all other edges.