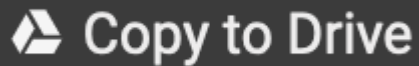


## Instructions

Please make a copy of this file before continuing. You can click on the "Copy to Drive" button above. It looks like this:



### ▼ 1 Setup

We first need to install `networkprocessing`. This is the same file that was included in the zip folder for week 8 (in case you want to load this in your local Jupyter Notebook rather than on Google Colab).

```
!pip install -q git+https://github.com/geil002/networkprocessing.git
```

```
Preparing metadata (setup.py) ... done
Preparing metadata (setup.py) ... done
Building wheel for networkprocessing (setup.py) ... done
Building wheel for statistics (setup.py) ... done
```

### ▼ 2 Loading data

Fetch the sample data. This is the list of people involved in the creative roles of the Harry Potter films. The following instructions will create a Harry Potter folder.

```
!wget -q https://github.com/geil002/sampleddata/raw/main/harrypotter.zip
!unzip -q harrypotter.zip
```

### ▼ 3 Loading modules

```
import pandas as pd
import seaborn as sns
import networkprocessing as np
```

### ▼ 4 Exercise instructions

This is the same as what you saw in the videos for week 8.

### ▼ 4.1



Load the Network Information spreadsheet into a dataframe called df and display it (there are only two rows, so you don't need to use head(), you can merely write the name of your dataframe df).

```
np.make_network("harrypotter")
df = pd.read_excel("harrypotter_analysis/networkInfo.xlsx", index_col=0)
df
```

```
Processing harrypotter/recent.txt
Processing harrypotter/older.txt
```

	nodes	edges	avg path length	avg degree	avg weighted degree	diameter	radius	d
network_id								
recent	6	13	1.133333	4.333333	12.000000	2	1	0.1

## 4.2

Write a few sentences describing the differences between the older and newer networks in the Harry Potter films.

The older network has a lower avg weighted degree and density as compared to the recent network but has a higher avg path length

## 4.3

Load the node information spreadsheet into a new dataframe called df2. You will use this dataframe for instructions 4-9.

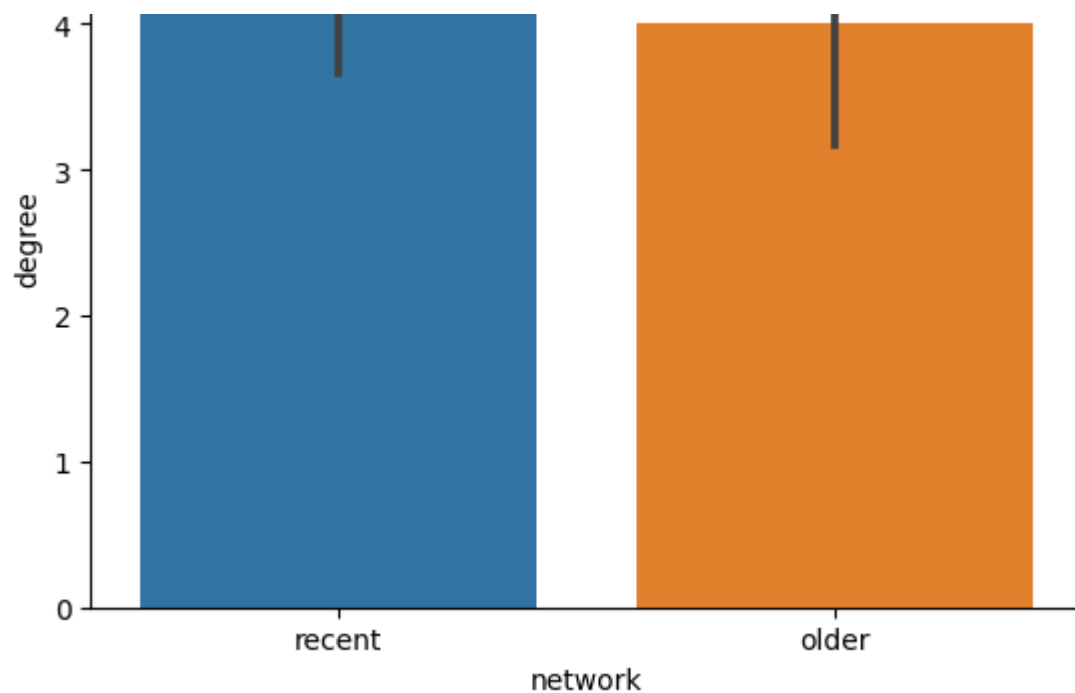
```
df2 = pd.read_excel("harrypotter_analysis/nodeInfo.xlsx", index_col=0)
```

## 4.4

Draw a barplot that compares the "degree" for the two networks (older and newer).

```
ax = sns.barplot(data=df2, x="network", y="degree")
```





#### 4.5

Which node has the lowest “closeness”? Use `sort_values()` to find the answer. Write your answer in a new cell using markdown (the answer should be only one node).

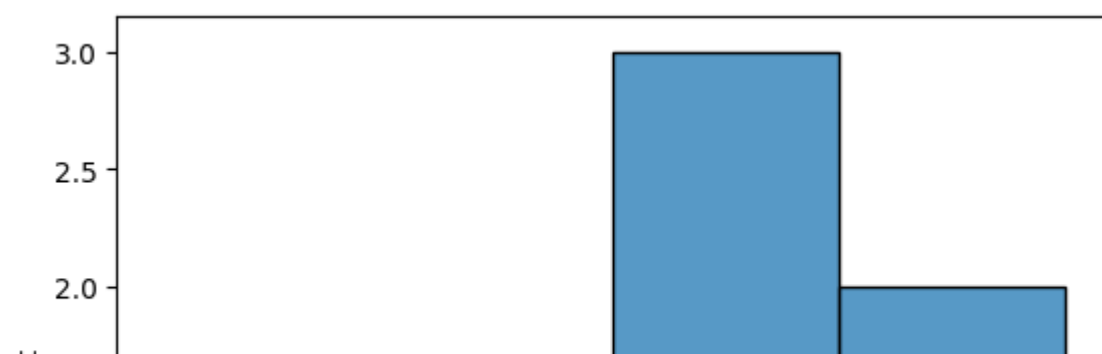
```
df2.sort_values(by="closeness",ascending=True).head(1)
```

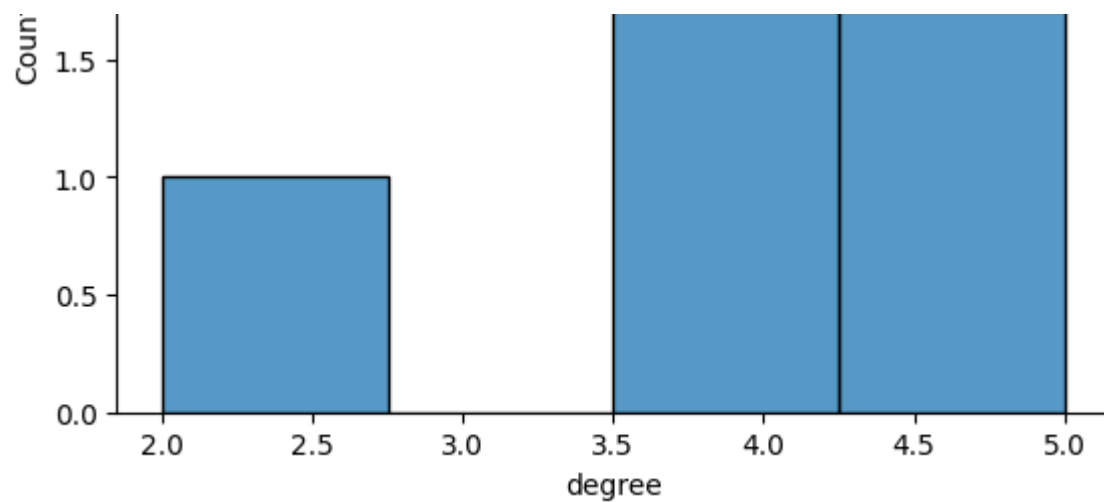
	label	degree	weighted_degree	betweenness	normalized_betweenness	ec
<b>11</b>	Mike Newell	2	2	0.0		0.0

#### 4.6

Draw a histogram for the degrees of the “older” network. Hint: you need to select only a portion of `df2`.

```
older = df2["network"] == "older"
ax = sns.histplot(data=df2.loc[older],x="degree")
```





#### 4.7

Upload your completed Jupyter Notebook (with the .ipynb extension) it to Canvas.

#### 4.8

Choose one of the networks ("older" or "newer") and use Gephi to change the look and feel of the network.

#### 4.9

Save an image of your completed network and upload it to Canvas.

#### 4.10

You have until the end of week 10 to submit these files.

#### 4.11

If you are done and have no questions, you can leave.

