

Data Structures and Algorithms in Python

Assignment 17 – Graph Algorithms and Six Degrees of Kevin Bacon

In this assignment, you will explore the small world phenomenon through the lens of the Hollywood film industry, specifically examining the "Six Degrees of Kevin Bacon" concept. Using a dataset of movies and actors, you will construct a graph where nodes represent actors, and edges represent co-appearances in films. This network will then be used to analyze the connectivity and "closeness" of actors in Hollywood.

Assignment components

- *Dataset*

We will use TMDb 5000 movie dataset, freely available on Kaggle

<https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata>

- *Graph Construction*

Create a graph using NetworkX where each node represents an actor. Draw an edge between two actors if they have appeared in at least one common film. Ensure the graph is undirected as the relationship "co-acted with" is bidirectional.

- *Graph Analysis*

Calculate basic properties of the network: number of nodes (actors), number of edges (co-acting relationships), average degree (average number of co-actors per actor), etc.

Identify the largest connected component in the graph (the largest set of actors who are connected through some path).

Browse the NetworkX documentation for graph algorithms and explore additional properties of the graph. <https://networkx.org/documentation/stable/reference/algorithms/index.html>

- *Kevin Bacon Number*

Implement functionality to calculate the "Bacon number" for any actor in the largest connected component, which is the number of edges in the shortest path from Kevin Bacon to the specified actor.

Develop a simple text-based interface that allows users to input an actor's name and outputs their path to Kevin Bacon (or another famous actor of your choice). Optionally, create a graphical user interface using libraries like Tkinter or PyQt.

Investigate the distribution of path lengths: Plot a histogram showing the frequency of Bacon numbers (or Actor X numbers) in the network.

Calculate and interpret the average path length and the diameter of the largest connected component.

- *Written report*

Prepare a one-to-two-page report in which you

- explain the background of how the graph was constructed, define key vocabulary terms, and cite the data source
- document the graph's properties discussed above and any additional interesting findings regarding the small world phenomenon in Hollywood
- explain how the properties of the Hollywood actors' graph compare with typical small world networks
- include an image of your histogram and examples of actor searches with their Bacon numbers
- discuss any challenges faced during the implementation and how they were resolved

- *Submission*

Submit your Python file(s) and a PDF of your written report via CREEKnet.

Ensure the code is well-commented and organized, and the report is clear and well-written, following the guidelines provided.