TU München, Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme
Prof. Dr. Thomas Neumann

Exercises for *Foundations in Data Engineering*, WiSe 19/20

Timo Kersten (kersten@in.tum.de)
http://db.in.tum.de/teaching/

Sheet Nr. bonus3

Bonus Project 3

Task In this project, we revisit the calculation of the Bacon-Number. Only this time, we do not only care for the distance of Kevin Bacon to everyone else, but for the shortest chain length between any two given actors.

Write a program that uses a given file `playedin` of csv entries with these columns:

1. actor id (integer)
2. movie id (integer)

The integer values form a dense range starting from 0 to the number of elements enumerated.

Based on this graph and two actor ids, your program must compute the length of the shortest coworking chain between two actors. Chains are built of actors that acted in the same movie. For example if Kevin Bacon and Dustin Hoffman worked on the same movie, the chain length between them is 1. If Dustin Hoffman then worked with Barbra Streisand, but Barbra Streisand did not work with Kevin Bacon, the shortest chain length is 2. This extends to arbitrary chain lengths. If no chain can be found, your program should return -1. The chain length from an actor to himself is defined to be 0.

How to submit

Clone the project at https://gitlab.db.in.tum.de/kersten/fde19-bonusproject-3.

Put your implementation into DistCalculator.cpp. The constructor of DistCalculator receives the path to the playedin.csv file described above. The function DistCalculator::dist(Node a, Node b) receives two actor ids and must return the distance.

Use git commit and git push in order to have your submission evaluated for the *leaderboard*. All subissions which complete the workload in less than 20 seconds are accepted for the bonus.