COMP 140: Kevin Bacon Game (Writeup)

Rixner

Joel Abraham

1   Recipe

We use the `find_path` method to find a path between a specified starting node to an ending node in a given graph, and return it. The inputs of the `find_path` method are: a graph, *graph*, a node in the graph *start_person* (that represents a certain actor from which we want to find a path), and another node in the graph *end_person*, (also representing an actor to which we want to find a path), and a mapping *parents* from a set of nodes to a set of nodes that maps each node (except for the node given by *start_person*) to another node such that the two nodes are connected by an edge in the graph and the distance between the latter node and *start_person* is less than the distance between the distance between the former node and *start_person*. The output of this method is a sequence of

1: **procedure** FIND_PATH(*graph*, *start_person*, *end_person*, *parents*)
2:     *edge* ← ∅
3:     *path_tuple* ← 2-tuple with *end_person* at first position and *edge* at second position
4:     add *path_tuple* to the front of the sequence *path*
5:     **while** *start_person* ≠ *end_person* **do**
6:         *parent* ← the value that *end_person* maps to in the mapping *parents*
7:         **if** *parent* = NIL (if *end_person* unreachable from *start_person*) **then return** an empty sequence
8:         **end if**
9:         *edge* ← the attributes of the edge (represented as a set) between *parent* and *end_person* in *graph*
10:         *end_person* ← *parent*
11:         *path_tuple* ← 2-tuple with *parent* at first position and *edge* at second position
12:         add *path_tuple* to the front of the sequence *path*
13:     **end while**
14:     **return** *path*
15: **end procedure**

2   Discussion

Discuss the results of the experiments you ran in 4.B (be sure to include the results). In particular, discuss the following questions:

1. The diameter of a graph is defined as the maximum distance between any two nodes in the graph. Describe a strategy using what you have done in this assignment for computing the diameter of a graph.

2. In the Kevin Bacon Game, highly connected actors are "valuable". The most connected nodes in this case actors in a graph are called "hubs". Describe a strategy for computing the hubs in a graph.

3. Just as certain actors can be identified as hubs in a movie graph, and therefore be seen as important to the structure to the graph, certain movies can also be considered more important to the structure of the graph. What criteria would you use to classify movies as such?

4. The `run` function uses the `simpleplot` module to plot the output of `distance_histogram` on two actors, Kevin Bacon and Stephanie Fratus, in the 5000-node subgraph. Describe the differences between these two plots; what does this tell you about these two actors?

Solution

1. If a graph has n nodes, we can iterate over all $\binom{n}{2}$ combinations of nodes – for one of these combinations, the distance between the nodes will be equal to the diameter of the graph. Using a BFS, we get a

mapping of distances between the start node and every other node in the graph. Then, we consider the largest distance value from the starting node which is the maximum of all the values in the distances mapping. Thus, we can perform n breadth first searches, one starting from each node in the graph, and find the maximum of every largest distance value obtained from each BFS to yield the diameter of the graph.

2. Hubs can be considered as vertices in the graph with high degrees. Thus, the k largest hubs in a graph can be computed by considering the k nodes with the highest degrees in the graph. This can be done by iterating through each node in the graph and considering its degree which is equal to the number of neighbors it has (the length of a sequence of its neighbors). Then, the nodes can be sorted from highest degree to lowest degree and we obtain a sequence of actors. The actors at the beginning of this sequence can be considered hubs. If we're specifically trying to find the k largest hubs in a graph, we can just consider the first k nodes in this sequence.

3. Movies that connect actors that are highly connected ("hubs") would likely be considered important to the graph's structure. Additionally, movies that connect strongly connected components of the graph would be critical to structure, since it would be key to decreasing the distances between many pairs of actors. Finally, movies that form a unique edge between two actors would also be important (although not as important as the above descriptions). Unimportant movies would be those that can be removed without modifying the structure of the graph (i.e. no edges are lost), such as if a movie connects two actors that are already connected by a different movie.

4. In general, we observe that Kevin Bacon seems to be closer connected to actors/actresses than Stephanie Fratus. Specifically, both have only 1 actor/actress from whom they are a distance of 0 (themselves), but Bacon has more actors/actresses from who he is a distance of 1 than does Fratus, indicating that Bacon has starred in more movies with other actors/actresses than has Fratus. Furthermore, Bacon has a very high number of actors/actresses from who he is a distance of 2, fewer actors/actresses from who he is a distance of 3, and fewer still actors/actresses from who he is a distance of 4; Bacon has no actors/actresses from who he is a distance of 5 or greater. Fratus, on the other hand, has many actors/actresses from who she is a distance of 4, slightly fewer actors/actresses from who she is a distance of 3, and a handful of actors/actresses from who she is a distance of 2, 5, or 6. This indicates that Bacon has starred in movies with other highly connected actors/actresses unlike Fratus, since from any given actor/actress, Bacon is likely to have a lesser distance than Fratus. These data indicate that not only is Bacon is a more prolific actor than Fratus, but also he works with other prolific actors, so his relationship tree branches at a greater rate.