Assignment 2 and 3

Colm O'Riordan

October 23, 2023

1 Assignment 2: Indexing

The data available is in the following format. Each match played is stored in one file. Each file comprises tuples which are in temporal order; i.e. the timestamp is in increasing order.

Each tuple comprises the following:

timestamp, $\{< p\_id, x, y >\}, < x, y, z >$

where the timestamp is just an integer value indicating the tuple's position in the sequence; there are a set of p_id, x,y triples indicating the x,y location of the players, and finally a triple giving the location of the football.

1. The analysts wish to be able to create heatmaps (visualisation describing where a player is over the course of the game). In order to do this efficiently, suggest an indexing approach and an algorithm to caulculate the amount of time a player spends at a certain location. You may break the pitch into a grid of rectangles/squares at some level of granularity.

2. How could you parallelise the above approach?

3. The coaches wish to generate queries of the following type. Consider subdividing the pitch into 9 rectangles (3 by 3) grid. The coaches wish to find the times (Which could then be used to generate a heatmap) of when any given player is in a specified rectangle. Specify a suitable indexing approach. Outline pseudocode to calculate the values for the heatmap.

4. Building on the previous query, they wish to find when two players (from either team) are in the same grid location ? Outline indexing approach and pseudocode.

2 Assignment 3: Graphs

Recent research work has considered analysing team sports as a dynamic graph and then analysing the graphs generated.

Consider viewing the team at any time as a graph. Each node represents a player and the edge between the nodes measures some notion of the distance between them. Note for any games, we may generate the graph every few minutes (taking a snapshot or an average of several snapshots)

1. Suggest a way of representing this graph in a relational database.
2. Suggest a suitable means to represent the data in a data structure.
3. Suggest an algorithm to measure the similarity of two graphs.
4. Consider the following constraint; if the distance between two players is greater than k, keep the edge, else discard the edge.
    - Given sample data, write code/pseudo code to calculate the degree of each node for any snapshot (a given timestamp).
    - Given sample data, write code/pseudo code to determine which node(s) is on the most paths?