Challenge 2: Answers to Review Questions

good work! i only have a couple of questions for further clarification:

1. you say that "The importance of each term is calculated, the top terms are selected and the edges are created. Thereafter, the graph is visualized." It is clear what the nodes signify, but i can't understand how the edges are created and what do they signify. Can you please clarify?

Answer: The edges and nodes were created using Corpus, Document term matrix (DTM) and a Matrix. Here is a brief explanation:

A corpus is a collection of text documents. The function “Corpus(vectorSource())” creates a corpus from the most occurring words in the learning journal.

Then from the corpus, a document term matrix (DTM) was created, in the matrix, each row corresponds to a document in the corpus and each column corresponds to a unique term or word across all documents.

The presence or absence of a term(word) in a document is represented by binary values (1 for presence, 0 for absence) or by word frequencies. This process essentially creates the edges as a binary one represents a connection between two terms(words).

These operations were abstracted away by the functions used; I would have added this explanation to the submitted work.

1. the visualization shows a directed graph. What does the direction show?

Answer: The directionality in the visualization represented the influence and or the proximity between the two terms(words). For example, we can see pairs like (ggplot, chart), (model, analysis), (bar, counts)… This makes sense as the word pairs are related to another, and we can say for example that “counts” is close to and somehow influences most “bar” charts in the study.

1. are there any graph metrics (e.g. degree centrality?) that you could further use to characterize the concepts shown in the graph?

Answer- Yes there are, I can apply the centrality measure to my graph and choose from the various type of centrality available like the degree centrality, closeness centrality, betweenness centrality, eigenvector centrality amongst others. Centrality measures the importance or influence of nodes within a graph.

Thank you