**PageRank**

**Introduction**

PageRank is a model that was developed by Larry Page and Sergey Brin and was the first algorithm used by Google for ranking web pages within search engine results. It is based on the idea that the importance of a given page is determined by the number and quality of the links to that page. This is done by taking a directed graph of elements and converting it to an adjacency matrix. Keeping to the example of a google search, the nodes of the graph represent the individual pages on the search result page, and the edges represent the links between the pages, as shown in the example below.

**Explanation of the Algorithm**

A diagram of a diagram

Description automatically generatedA number grid with numbers

Description automatically generated

Figure 1: A Directed Graph and its Corresponding Adjacency Matrix

Given a network of pages linked to one another, that can be represented as a directed graph, PageRank solves the linear system of equations set by to address the importance of pages within the network. In the equation. A is the adjacency matrix, D is the diagonal out-degree matrix, I is the identity matrix, alpha is the damping factor, which was experimentally determined to be 0.85, n is the is the number of nodes in the graph, and e is a vector of all 1s. To derive D, the out-degree diagonal matrix, we sum the out-degrees across each row and assign it to the diagonal node. For example, the diagonal matrix for the graph in figure 1 would be represented as:

Figure 2: Corresponding Diagonal Out-Degree Matrix for the Directed Graph

As we can see from the D matrix, node A has two outward edges, node B has two outward edges, node C has one outward edge, and node D has one outward edge, which lines up with the diagram of the graph in Fig. 1.

The next step to solve this would be to take the transpose of the adjacency matrix and the inverse of the D matrix, and multiply.

Then, solving for x becomes simple using the provided equation and a solver.

From this, we can see that node C has the highest rank, followed by node A, which lines up with the graph, as Node C has an in-degree of 3, and node A has an in-degree of 2.

**Applications of PageRank**

Due to the dependence of PageRank on only a directed graph structured network, there are a variety of applications that it can be suited to. One such example is the multitude of streaming platforms that are available for video, music, and other media.

1. Personalization

PageRank’s ability to analyze and rank nodes – in this case, media, based on their connectedness can be leveraged to provide personalized recommendations based on similarity scores depending on how the directed graph is assembled. Creating a graph where nodes represent users and tracks, and edges represent user interactions, such as listens or likes, can be used to suggest new songs to a user based on other users’ preferences that also listen to a specific genre of music. This generates suggestions that are popular and are more likely to be closer to the user’s preferences.

1. Content Popularity and Trends  
   Expanding the network to the global scale can offer more insight into the current trends across the globe, informing users of what or who are the most influential or popular content creators, scoped to the global, regional, or local levels. For corporate usage, this can give companies insight into which content creators or influencers to partner with to spread brand awareness and popularity.

**Limitations of PageRank**

However, PageRank does not come without drawbacks. After all, there is a reason that Google kept innovating to upgrade its algorithms for better web searching.

1. Link structure Dependency  
   The PageRank model depends on a directed graph model, which means that there is a connected network of links between nodes. Based on the example from before, it is easy to see that nodes with more incoming links are more popular, simply for being more accessible. This leads to two issues – the rich-get-richer phenomenon, and artificial manipulation of links.   
   The rich-get-richer phenomenon occurs when a node accumulates more and more incoming links due to being around for a longer period. This can cause a feedback loop where the PageRank of older, already popular, content continues to dominate over newer or less-linked content – for example, this would be if Taylor Swift persisted on your recommendations page over newer artists.   
   Links can be artificially manipulated as well, through methods such as link farming and artificial link building. This artificially inflates the PageRank of certain content, and compromises the integrity of search results, as unpopular content is promoted over actually popular content.
2. Damping Factor Sensitivity  
   Furthermore, in the calculations of the PageRank, the damping factor alpha is a strong influence on the ranking of the nodes. The alpha value represents the probability that the user will continue to click through links rather than jumping to a new search altogether – for a practical example, a user would click through an article on data science to learn more about topics like data mining or data visualization, rather than jumping to a new article on baking. However, this is not the case for all networks. The accepted value of 0.85 can fluctuate between networks depending on the changing user behavior and can be difficult to determine computationally. Furthermore, the PageRank model depends on the assumption that the user follows the Random Surfer model, which the real user may not always follow. The deviation from the Random Surfer model can also impact the accuracy of page ranking, especially in highly dynamic and complex networks.

**Conclusion**

PageRank’s development as a search engine algorithm can be extrapolated to many industries, especially streaming platforms. PageRank operates to rank the quality and popularity of the links within a network, which can provide benefits to both media consumers and content producers. However, the model comes with limitations from within the model, as it depends on the user following the Random Surfer model, the links being free from external manipulation, and the damping factor being applied for the correct market.