# An Analysis of Google's PageRank Algorithm : How Linear Algebra Influences Google Search

Joey Kaplan U1364550

October 2024

## 1 Introduction

This project seeks to analyze the algorithm behind Google Search and what makes this algorithm efficient. This project is based on a Cornell lecture, `Lecture #3:PageRank Algorithm - The Mathematics of Google Search`

## 2 History of the Algorithm

In the 1990's, search engines commonly used *text based ranking systems* to decide which pages would be most relevant for a given query. While I will not touch on the exact logic or algorithmic approach of this text based ranking, these algorithms provided notable problems. First, these algorithms simply scanned for keyword count, meaning that a website with the same word repeated millions of times would appear at the top of the search results. For instance, searching "Linear Algebra" could possibly return a website with these words repeated over and over at the top of the search results.

Obviously, this was a problem since users wanted useful results and not web pages with large amounts of the keyword.

A notable algorithm that largely made these text based ranking systems obsolete was the Page Rank algorithm used the Google search engine. Created by Larry Page and Sergey Brin, who were graduate students at Stanford, this algorithm brought forth the idea that web pages could be sorted by importance by using links to that webpage and other web pages.

As such, it can be useful to visualize the Internet as a directed graph, with nodes representing web pages and edges representing the links between these web pages.

## 3 Notable Example

Put in slides

## 4 Dynamical Systems Point of View

Matrix A represents an adjacency matrix and vector v represents the $\frac{1}{n}$ probabilities associated with n-nodes. Multiply them to get new vector and keep doing A times Av until reach equilibrium vector.

## 5 Dangling Nodes and Disconnected Components

Dangling nodes need to have a column with $\frac{1}{n}$ in each entry to ensure rank does not approach 0. Disconnected Components need to use damping to "teleport" to other parts of the "system".

# 6 The Solution of Page and Brin

Essentially, multiply (1 - damping) with A and damping with v and add them to get M.

# 7 Notable Theorems

Ensures that probabilities are actually valid. Maybe just say something about how probabilites must be positive and this theorem ensures that via eigenvector?

## 7.1 Power Method Convergence Theorem

Convergence to equilibrium vector is necessary and sequence does that due to $\frac{1}{n}$ probabilities.

# 8 Solving Problem #4

Solve via slide

# 9 References

`https://pi.math.cornell.edu/ mec/Winter2009/RalucaRemus/Lecture3/lecture3.html`

`https://youtu.be/meonLcN7LD4?si=JiwCDwsxEHcMoonS`