Introduction:

Background in Graph Theory and the importance of Eigen Values

Graph Theory is a branch of mathematics that deals with the study of Graphs, which are abstract representations of pairwise relations between objects. A Graph consists of vertices (nodes) and edges(links) connecting these vertices.

Graph theory is widely applicable in many fields including computer science, biology, physics, and social sciences.

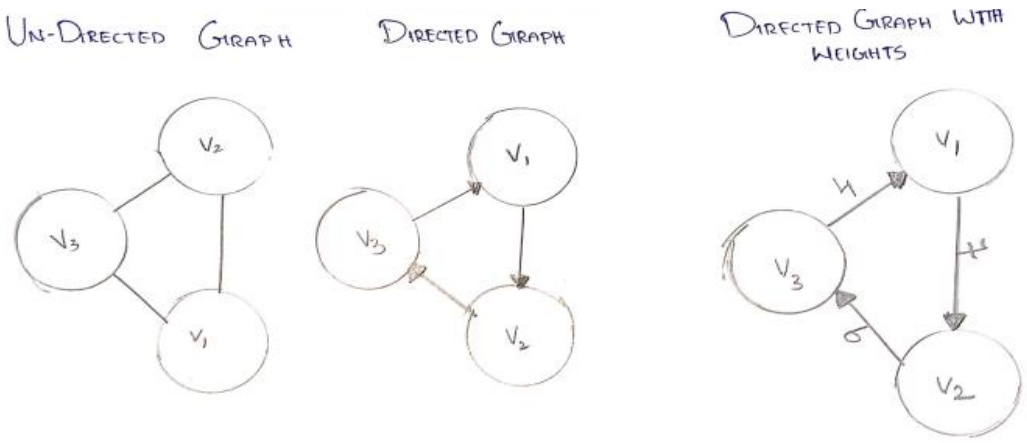
In the context of web, graph theory is particularly useful for representing the complex network of interconnected web pages. Each webpage is considered as a vertex, and the hyperlinks between the pages act as directed edges, forming a directed graph. The structure of this graph reveals important relationship between web pages, such as how frequently a page is linked to by others, which can indicate its relevance or importance.

Key Concepts:

- Vertex: A fundamental unit in graph.
- Edge: A connection between two vertices.
- Degree of a Vertex: The number of edges connected to a Vertex.
- Path: A sequence of edges that connects two vertices.
- Cycle: A path that starts and ends at the same vertex.

Types of Graphs:

- Directed Graphs: Edges have a direction (e.g., social networks)
- Undirected Graphs: Edges have no direction (e.g., road networks)
- Weighted Graphs: Edges have associated weights (e.g., transportation networks)



One of the most powerful tools in analysing the structure of a graph is the adjacency matrix. For a graph with vertices, the adjacency matrix is an matrix where the entry represents the number of edges from vertex to vertex. The adjacency matrix captures the connectivity of the graph, and by analysing this matrix, we can extract valuable insights into the nature of the graph.

Applications of Graph Theory:

- Social Network Analysis
- Computer Science
- Biology
- Transportation…etc

In linear algebra, eigenvalues and eigenvectors are key concepts when dealing with square matrices, such as the adjacency matrix. An eigenvalue of a matrix is a scalar such that there exists a non-zero vector (the eigenvector) that satisfies the equation:

$$Av=Kv$$

In many real-world applications, understanding the dominant eigenvector of a graph's adjacency matrix is essential for ranking or prioritizing nodes. For example, in social networks, the dominant eigenvector can help identify influential individuals, while in citation networks, it can highlight the most cited academic papers. This same principle underlies the PageRank algorithm, where the dominant eigenvector is used to rank web pages by their importance.

Key Applications:

- Spectral Clustering: Grouping vertices based on their similarity.
- Centrality Measures: Identifying important nodes in a graph.
- Graph Isomorphism: Determining if two graphs are structurally equivalent.
- Graph Partitioning: Dividing a graph into smaller connected components.

Some Basic Mathematical Definitions and PageRank Algorithm:

- **Matrix**: A rectangular array of numbers arranged in rows and columns. In PageRank, matrices are used to represent the web graph and perform calculations on it.
- **Square Matrix**: A matrix with the same number of rows and columns, which is the case for the adjacency matrix of a graph with pages.
- **Stochastic Matrix**: A square matrix used to represent a Markov process, where each column's elements sum to 1. In PageRank, the normalized adjacency matrix becomes a stochastic matrix representing the probabilities of transitioning between pages.
- **Markov Chain**: A random process that undergoes transitions from one state to another, with the probability of each state depending only on the previous state. The web's link structure is modelled as a Markov chain in PageRank.
- **Normalization**: Adjusting values so they sum to 1. In the PageRank context, each column of the adjacency matrix is normalized to represent the probability of transitioning between pages.
- **Eigenvalue**: For a square matrix, a number is an eigenvalue if there exists a non-zero vector (called an eigenvector) such that. In PageRank, the largest eigenvalue (dominant eigenvalue) is crucial for finding the importance of pages.
- **Eigenvector**: A non-zero vector that, when multiplied by a matrix, results in a scalar multiple of itself. The dominant eigenvector of the web's adjacency matrix represents the relative importance of each web page.
- **Dominant Eigenvector**: The eigenvector corresponding to the largest eigenvalue of a matrix. In PageRank, the dominant eigenvector's components indicate the relative importance of web pages.
- **Power Iteration**: An iterative method to find the dominant eigenvector of a matrix. Starting with an arbitrary vector, it repeatedly multiplies by the matrix and normalizes the result until it converges. PageRank uses this method to calculate page rankings.
- **Steady-State Distribution**: A probability distribution that remains unchanged as the Markov process evolves. The dominant eigenvector in PageRank represents the steady-state distribution of the web pages.

Vector Scaling and Normalization

- **Scaling**: Multiplying a vector by a scalar. In PageRank, the final dominant eigenvector is scaled so that the sum of its components is 1, representing the relative importance of each page.
- **Normalization of Vectors**: Adjusting the components of a vector so their sum or magnitude equals a specific value (often 1). This step is crucial in ensuring that the probabilities in the PageRank vector are meaningful.

Overview of PageRank and Its Relevance to Web Search

The PageRank algorithm, developed by Larry Page and Sergey Brin in the late 1990s, is one of the most well-known applications of graph theory and eigenvalue analysis. PageRank was created to solve the problem of ranking web pages in an objective and scalable way, based on the link structure of the web. The algorithm has since become a cornerstone of web search engines, including Google, for determining the relevance and importance of web pages.

The central idea behind PageRank is that a web page should be considered important if it is linked to by other important pages. In other words, not all links are equal—being linked to by a highly reputable page is more valuable than being linked to by a less-known or less-reputable one. This recursive principle is naturally captured through the framework of eigenvectors: the PageRank of a page is proportional to the sum of the PageRank scores of the pages linking to it.

Key Concepts:

- Web Graph: A directed Graph where nodes represent webpages and edges represent hyperlinks.
- Rank Score: A numerical value assigned to each web page indicating their importance.
- Iterative Calculation: PageRank is calculated iteratively, considering the rank scores of linked pages.
- Damping Factor: A parameter that controls the influence of random surfing on the rank scores

To implement PageRank, the web is represented as a directed graph, where each node corresponds to a web page, and each directed edge represents a hyperlink from one page to another. The connectivity of the web graph is encoded in an adjacency matrix, where each entry indicates the presence or absence of a link between pages.

The goal of the PageRank algorithm is to compute the dominant eigenvector of this adjacency matrix. The components of this eigenvector represent the steady-state probabilities of a random surfer model, where a hypothetical user continuously clicks on hyperlinks, randomly traversing the web. Over time, the random surfer tends to visit more

important pages more frequently, and the dominant eigenvector reflects the relative importance of each page in the network.

However, in the real web, issues such as dangling nodes (pages with no outgoing links) and rank sinks (pages that disproportionately attract all PageRank scores) must be addressed. To resolve these issues, the adjacency matrix is slightly modified to form the Google matrix, which introduces a teleportation factor. This teleportation allows the random surfer to occasionally jump to a random page, ensuring that the algorithm converges to a unique solution and prevents certain pages from dominating the rankings unfairly.

The iterative computation of the dominant eigenvector is typically done using the power iteration method, where the adjacency matrix is repeatedly applied to an initial vector until convergence is reached. The resulting eigenvector is then normalized to ensure that its components sum to 1, with each component representing the relative importance (or PageRank) of a web page.

Relevance of PageRank to Web Search

The relevance of PageRank to web search cannot be overstated. Before the development of PageRank, search engines primarily ranked pages based on keyword matching, which often failed to capture the true relevance of a page to a user's query. PageRank introduced a paradigm shift by incorporating the link structure of the web into the ranking process. By considering both the number and quality of incoming links, PageRank provided a more accurate measure of a page's importance and relevance.

PageRank continues to be a vital component of search engines today, though it is often combined with other signals, such as content relevance, user behaviour, and personalization, to create more sophisticated ranking algorithms. Its success has also inspired numerous applications in other fields, such as social network analysis, citation

ranking, and recommender systems, making PageRank a widely used tool for ranking and analysing large, complex networks.

Relevance to Web Search:

- Quality Assessment: PageRank helps identify high-quality web pages that are likely to be relevant to user queries.
- Link Analysis: It considers the structure of the web graph to determine the importance of pages.
- User Experience: PageRank contributes to a better user experience by presenting relevant and informative search results.

//Page rank more details, and visual example.