Application of PageRank in Social Media

Introduction

With the rapid development of social media platforms such as Facebook, Instagram, and Twitter, they have not only become central to social life but have also created significant economic impacts. Despite the vast user base of these platforms, the challenge remains in how to effectively leverage this user base for sustainable revenue. To maintain user activity and loyalty, operators need more precise advertising strategies and user retention programs. Identifying and utilizing "key users" is essential to these strategies.

Key users refer to individuals in social networks who have high connectivity and frequent interactions. These users can influence a large number of others and often display high levels of loyalty. They are critical targets for advertising campaigns and efforts to boost user engagement. As a result, identifying these key users is crucial for developing effective strategies. In this context, the PageRank algorithm has been widely applied.

PageRank Algorithm

PageRank was initially used for ranking web pages by analyzing the link relationships between them to assess their authority. The same principle can be applied to social networks by evaluating each user's influence through their interaction connections. In the paper *Identifying Key Users in Online Social Networks: A PageRank Based Approach*, the authors propose two methods based on PageRank for identifying key users.

The first method involves the **Weighted Activity Graph**. In this approach, user interactions are represented as edges in a graph, with the edge weights reflecting the strength of the interactions (e.g., frequency or importance of the interactions). This method goes beyond a simple "social graph" that merely shows friend connections, as it is built on actual interactions between users. For example, edges only exist between users who have real interactions. Then, the PageRank algorithm is used to compute the centrality score of each user, which represents their influence in the network. The PageRank "voting" mechanism indicates that a user's influence depends not only on the number of their direct connections but also on the influence of those they are connected to.

The traditional PageRank algorithm assumes all links are equal, but in social networks, the intensity of interactions between users varies. By assigning weights to edges, users with more frequent interactions have a greater influence when calculating centrality. The formula is as follows:

$$S(i) = \frac{1-d}{N} + d \sum_{j \in F(i)} \frac{w_{ij} S(j)}{\sum_{k \in F(j)} w_{jk}}$$

Where $S(i)$ is the centrality score of user i, d is the damping factor (commonly set to 0.85), and w_{ij} represents the weighted interaction strength between user i and j. By iterating this formula, the centrality score for each user can be determined, allowing for the identification of the most influential users.

Limitations of Traditional PageRank

However, traditional PageRank has certain limitations. It relies solely on direct connections between users (first-order relationships), neglecting the more complex interactions among multiple users. To address this, the paper *Ranking Users in Social Networks with Motif-Based PageRank* introduces a modified version of PageRank, known as Motif-Based PageRank, which captures higher-order relationships between users. By introducing a **Motif-Based Adjacency Matrix**, the algorithm captures the number of times two nodes (users) appear together in specific motifs (interaction patterns). This matrix represents more complex relationships between users compared to the traditional edge-based adjacency matrix. The authors propose two methods for combining traditional and motif-based matrices: linear and nonlinear combination methods.

- **Linear combination method**:

$$H_{Mk} = \alpha \cdot W + (1-\alpha) \cdot W_{Mk}$$

- **Nonlinear combination method**:

$$H_{Mk} = W^{\alpha} \cdot W_{Mk}^{1-\alpha}$$

Where W is the edge-based adjacency matrix, W_{Mk} is the motif-based adjacency matrix, and α is a balance coefficient between 0 and 1.

Application on Twitter

Twitter, as the largest microblogging platform globally, presents an example of how social network analysis can identify key users for specific topics. In the study *Social Network Analysis of Twitter to Identify Issuer of Topic using PageRank*, the research team used the Twitter API to gather tweet data related to specific topics. The example topic used in the paper was “Jokowi Capres.” For each tweet, data such as the timestamp, username, retweet status, and original poster were collected. The PageRank algorithm was then applied to calculate the influence score of each user, with the highest-scoring users identified as the primary initiators of the topic. The formula used was:

$$S(N_i) = (1 - d) + d \cdot \sum_{j \in In(N_i)} \frac{S(N_j)}{|Out(N_j)|}$$

Comparison with Other Centrality Measures

The study also compared PageRank scores with other centrality measures such as closeness centrality and betweenness centrality. The results showed that PageRank scores had a strong correlation with closeness centrality and indegree but a weaker correlation with betweenness centrality and follower count. This suggests that PageRank is more focused on a user's influence within a specific topic, while follower count alone does not directly reflect influence on a particular issue.

Conclusion

Through these applications and studies, the value of PageRank in social media has been demonstrated. It can be used not only to identify key users but also to help develop more effective advertising and user retention strategies. As social networks become more complex, further refining the PageRank algorithm to account for multi-level social interactions will be an important area of future research.