# Addressing User Activity Bias in Bipartite Graph Ranking Using an Enhanced BiRank Algorithm

Layan Kaushik
Department of Computer Science, Duke University

**Abstract**

This research introduces an enhancement to the BiRank algorithm, specifically designed to improve fairness in bipartite graph rankings. The BiRank algorithm, primarily used in recommendation systems and search engines, is instrumental in ranking nodes on bipartite graphs, which consist of two distinct sets of entities, such as users and items. Our enhancement focuses on mitigating user activity bias, a common issue in these systems where user engagement disproportionately influences rankings. We propose a regularization extension to the original BiRank algorithm using user degree information. Our study reveals that while the original BiRank algorithm effectively minimizes user activity biases in graphs with a larger number of user nodes, our proposed regularization extension has a notable impact on graphs with fewer user nodes. Utilizing a newly developed fairness metric, we empirically demonstrate that our enhanced algorithm can, in some instances, achieve fairer results than the original BiRank. Our findings underscore the potential of our approach in reducing bias and enhancing fairness in systems that rely on bipartite graph rankings.

## 1 Introduction and Background

Bipartite graphs are fundamental structures in a wide array of applications, including recommendation systems, social networks, and collaborative filtering. These graphs uniquely consist of two distinct sets of vertices, with edges exclusively interconnecting these sets, creating a diverse and complex web of relationships. Their unique structure presents specific challenges in graph-based algorithms, particularly in ranking nodes, due to the absence of direct links within the same vertex set.

### 1.1 Bipartite Graphs

In bipartite graphs, let $G = (U \cup P, E)$, vertices are divided into two disjoint sets, $U$ and $V$, such that every edge connects a vertex from $U$ to one in $V$. This separation is represented by a weighted adjacency matrix $W = [w_{ij}]$, where $w_{ij}$ denotes the strength or weight of the edge between vertex $u_i \in U$ and $v_j \in V$. The degree $d_i$ of a vertex $u_i$, indicating the sum of weights of its connected edges, is captured in a diagonal matrix $D_U$, providing a measure of connectivity for each node.

### 1.2 BiRank Algorithm

The BiRank algorithm, as detailed by He et al. (2017) [1], offers a novel approach to ranking nodes in bipartite graphs, distinct from traditional ranking algorithms like PageRank. It operates by iteratively computing rank vectors for nodes in sets $U$ and $V$, employing the following update rules:

$$u^{(k+1)} = \alpha W' v^{(k)} + (1-\alpha) q_U$$
$$v^{(k+1)} = \alpha (W')^T u^{(k)} + (1-\alpha) q_V$$

Here, $W' = D_U^{-1/2} W D_V^{-1/2}$ is the normalized adjacency matrix, ensuring a balanced influence of nodes with varying degrees. The parameter $\alpha$, known as the damping factor, controls the balance between structure-driven and query-driven ranking, while $q_U$ and $q_V$ are query vectors representing inherent node importance. This tailored approach to bipartite graphs allows BiRank to effectively capture the nuances of these networks, differing significantly from algorithms designed for unipartite graphs.

## 1.3 User Activity Biases

A critical challenge in bipartite graph ranking, including in the BiRank algorithm, is the user activity bias. This bias arises when nodes (particularly users in $U$) with higher activity levels disproportionately influence the rankings, leading to skewed results. Such biases are especially pronounced in graphs with uneven activity distributions among nodes. Liao et al. (2019) [2] delve into temporal biases in bipartite graph rankings, highlighting the need for bias mitigation strategies. While the BiRank algorithm effectively ranks nodes, it does not inherently account for user activity bias. Our enhancement to BiRank addresses user activity biases, particularly in graphs adhering to random and power law distributions. We introduce a regularization mechanism, employing a formula $r_i = \gamma \frac{1}{\text{user\_activity}[i]+\epsilon}$ to balance node representation in the rankings and improve overall fairness and accuracy. This regularization approach effectively scales down the influence of highly active users, thereby yielding a more representative and fair ranking across the network. Figure 1 illustrates how a simple regularization mechanism adjusts product ranking in a bipartite graph by moderating the influence of high-activity user nodes.

# 2 Methodology

In this section, we outline the methodology adopted for refining the BiRank algorithm, with the dual objectives of enhancing its fairness and developing a novel fairness metric. Our approach is specifically tailored to address and mitigate biases stemming from user activity in the ranking of nodes within bipartite graphs. Crucially, the problem formulation underlying our methodology remains consistent with that presented in the original BiRank paper [1]. This ensures that our enhancements are built upon a well-established foundation, allowing for a direct and meaningful comparison between the original BiRank algorithm and our modified version.

## 2.1 BiRank Algorithm Extension

In this study, we extend the BiRank algorithm with a regularization mechanism to effectively address the bias caused by varying levels of user activity. While the original BiRank algorithm adeptly ranks nodes in bipartite graphs, it does not sufficiently account for disparities in user activity, potentially skewing the representation of less active users.

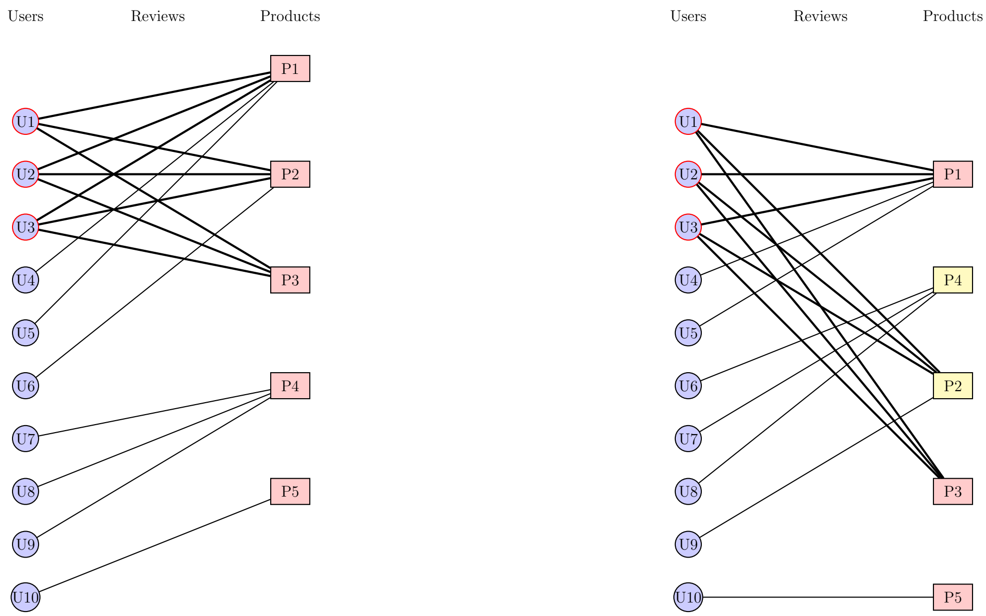


Figure 1: A simple illustration of Product Ranking Adjustment through Uniform Regularization of User Degrees in a Bipartite Graph.

Our proposed modification involves the integration of a regularization vector $r$ into the ranking process, enhancing the algorithm's ability to equitably represent user influence. This regularization vector is defined mathematically as:

$$r_i = \gamma \frac{1}{\text{user\_activity}[i] + \epsilon} \tag{1}$$

Here, $\gamma$ denotes the regularization strength factor, and $\epsilon$ is a small constant introduced to prevent division by zero. The metric *user_activity* for a user $i$ is computed as the degree of the user $i$, normalized by the total sum of degrees across all users. This normalization ensures a balanced representation of user influence, expressed as:

$$\text{user\_activity}[i] = \frac{\text{degree of user } i}{\sum_j^{|U|} \text{degree of user } j} \tag{2}$$

Here, $|U|$ represents the cardinality of the user set, i.e., the total number of users. Incorporating this regularization vector, the updated ranking formula for nodes in set $U$ becomes:

$$u^{(k+1)} = \left(\alpha W' v^{(k)} + (1-\alpha) q_U\right) \odot r \tag{3}$$

where $\odot$ denotes element-wise multiplication (dot product). For nodes in set $V$, the ranking formula remains unchanged from the original BiRank algorithm:

$$v^{(k+1)} = \alpha (W')^T u^{(k+1)} + (1-\alpha) q_V \tag{4}$$

The primary aim of this modification is to mitigate the overrepresentation of hyperactive users, fostering a more balanced and equitable network representation. By adjusting the influence of user activity levels, our extension of the BiRank algorithm promises a fairer and more representative ranking of nodes within bipartite graphs.

The above method is summarized in Algorithm 1.

**Algorithm 1** The Enhanced Iterative BiRank Algorithm

**Require:** Weight matrix $W$, query vectors $q_U$ and $q_V$, regularization vector $r$, hyperparameters $\alpha$; $\gamma$;

**Ensure:** Ranking vectors $u$; $v$;

1: Symmetrically normalize $W$: $S = D_U^{-\frac{1}{2}} W D_V^{-\frac{1}{2}}$;
2: Randomly initialize $u$ and $v$;
3: **while** Stopping criteria is not met **do**
4: $u \leftarrow \alpha S^\top v + (1-\alpha) q_U$;
5: Element-wise multiply $u$ with $r$: $u \leftarrow u \odot r$;
6: $v \leftarrow \alpha S u + (1-\alpha) q_V$;
7: **end while**
8: **return** $u$ and $v$;

## 2.2 Fairness Metric

To evaluate the effectiveness of our extended BiRank algorithm, we introduce a new fairness metric. This metric is designed to balance the minimization of disparity in activity ranks while maintaining a fair Gini index. The fairness metric is formulated as:

$$\text{Fairness Metric} = \frac{1}{\left(\frac{1-w}{|\text{Disparity Ratio}-1|+\epsilon} + \frac{w}{\text{Gini Coefficient}+\epsilon}\right)} \tag{5}$$

Here, the *Disparity Ratio* is calculated as the ratio of the average rank of high-activity users to that of low-activity users. Meanwhile, the *Gini Coefficient* is utilized to measure the inequality in the distribution of user scores. These components are defined as:

$$\text{Disparity Ratio} = \frac{\textit{High Activity Average Rank}}{\textit{Low Activity Average Rank} + \epsilon} \tag{6}$$

and

$$\text{Gini Coefficient} = \frac{\sum_{i=1}^{n}(2i-n-1)\cdot x_i}{n \cdot \sum_{i=1}^{n} x_i} \tag{7}$$

In the context of the Gini Coefficient, $n$ represents the number of scores, $x_i$ is the $i^{th}$ score in the sorted list, and $i$ indicates the rank of each score.

A lower value of the fairness metric indicates greater fairness. The metric is designed to penalize deviations from ideal disparity ratios and high Gini coefficients, aiming for a more equitable ranking system.

# 3 Experiment and Results

This section describes the comprehensive approach undertaken in our experimentation, detailing the datasets used, the configuration of our experimental framework, and the insights obtained from applying the enhanced BiRank algorithm.

| Graph Type / (#Users, #Products) | BiRank | BiRank with Reg |
|---|---|---|
| Random Graph (100,50) | 0.508 | 0.857 |
| Random Graph (1000,100) | 0.404 | 0.701 |
| Random Graph (10000,500) | 0.421 | 0.711 |
| Random Graph (100,500) | 0.722 | 1.211 |
| Power-Law Graph (100,50) | 0.906 | 0.604 |
| Power-Law Graph (1000,100) | 0.747 | 0.711 |
| Power-Law Graph (10000,500) | 0.701 | 0.724 |
| Power-Law Graph (100,500) | 1.241 | 1.259 |

Table 1: Average Fairness Metric values for BiRank and BiRank with Regularization

## 3.1 Dataset Description

For our experimental analysis, we employed synthetic datasets, to emulate real-world bipartite graph scenarios. These datasets encompass two primary types: random bipartite graphs and those adhering to a power-law distribution. We tailored the number of user and product nodes in each dataset, as detailed in Table 1. The random graphs were generated with a focus on achieving a uniform distribution of edges across the graph, thereby ensuring an equitable representation of connectivity. Conversely, the power-law graphs were constructed to replicate scenarios frequently observed in real-world networks, where a select few nodes exhibit a significantly higher degree of connectivity compared to others.

## 3.2 Experimental Setup

The foundation of our experimental setup was the application of both the original BiRank algorithm and our enhanced version, BiRank with regularization, to these synthetic datasets and estimating the proposed fairness metric. We meticulously adjusted key parameters, including the regularization strength factor $\gamma$ and the weighting factor $w$ in our proposed fairness metric. This parameter variation was instrumental in scrutinizing the influence these factors have on the resultant rankings. The principal objective of this experimental exercise was to rigorously evaluate the performance of our enhanced algorithm in terms of fairness and accuracy, especially in comparison to the traditional BiRank algorithm.

## 3.3 Results

The experimental outcomes, as summarized in Table 1, provide insightful revelations about the performance of our enhanced BiRank algorithm with regularization. It is observed that, in general, the modified algorithm yields rankings that are not only comparable but in some cases superior to those produced by the original BiRank, as gauged by the developed fairness metric. For instance, Figure 2 visually demonstrates these comparative results in a Random Graph environment. A particularly notable improvement is seen in power-law graphs, especially those with a smaller number of user nodes, as depicted in Figure 3. This improvement underscores the efficacy of the regularization approach in diminishing the bias associated with user activity in bipartite graph ranking. The effectiveness of regularization in adjusting product rankings across different graph types is further illustrated in Figure 4. These findings corroborate our hypothesis that the

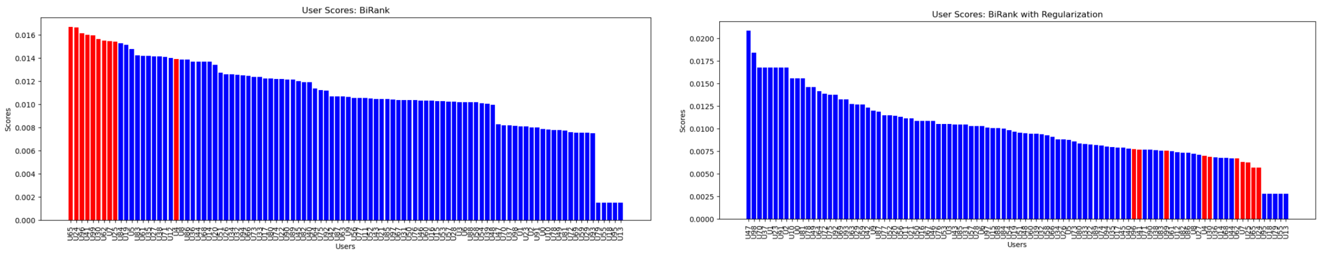


Figure 2: User Rankings in a Random Graph Environment: On the left, the BiRank algorithm's results, where red bars identify the most active user nodes, showcasing the initial ranking order. On the right, results from the BiRank algorithm with regularization, similarly highlighting the most active user nodes in red, display a modified ranking distribution. The implementation of regularization effectively moderates the influence of high-activity users.

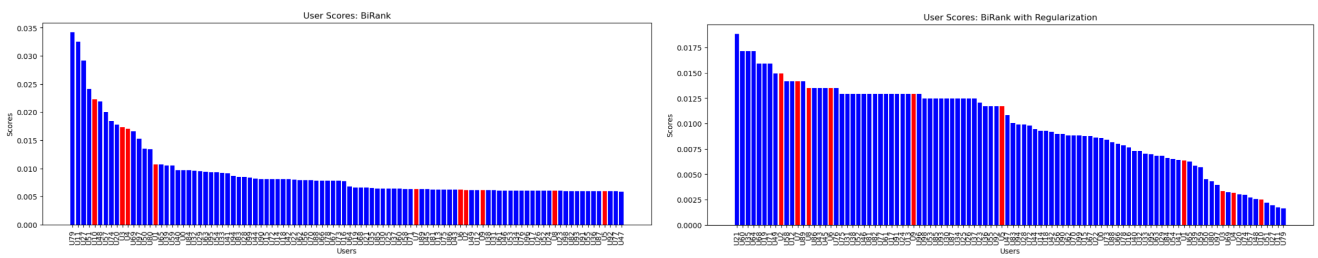


Figure 3: User Rankings in a Power Law Graph Environment: On the left, the BiRank algorithm's results, with red bars marking the most active user nodes, depict the original ranking structure. On the right, results from the BiRank algorithm with regularization, where the most active user nodes are also marked in red, show a more linear distribution of user scores, contrasting with the exponential scoring pattern observed with the standard BiRank.

integration of regularization into the BiRank algorithm can enhance fairness in ranking outcomes across some types of graph structures.

# 4 Discussion

The empirical results from our experiments with the enhanced BiRank algorithm yield significant insights into the dynamics of bipartite graph ranking. Notably, the incorporation of regularization has demonstrated marked improvements in fairness metrics, particularly within the context of power-law distributed graphs and those characterized by a smaller user base. This improvement suggests that our regularization mechanism is adept at counteracting the user activity bias prevalent in the standard BiRank algorithm.

In bipartite graphs following a power-law distribution, our findings indicate that the application of regularization can result in a more equitable ranking, especially in cases with fewer user nodes. This observation aligns with our hypothesis that user activity bias is more pronounced in such graph structures, and therefore, the benefits of regularization are more conspicuous. However, it is crucial to note that the effectiveness of this regularization is not uniform across all graph types and sizes. Our analysis reveals that the performance of the regularization in enhancing fairness varies, suggesting that its impact is context-dependent.

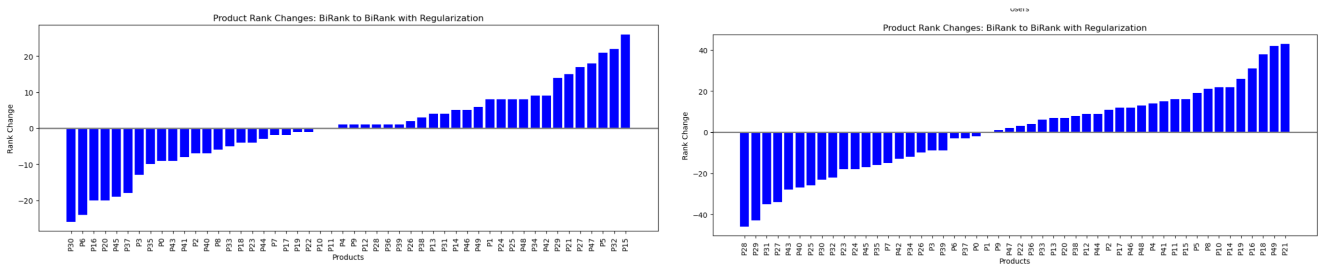


Figure 4: Comparative Analysis of Product Rank Adjustments: This figure showcases the changes in product rankings when transitioning from the standard BiRank algorithm to BiRank with regularization. The left panel illustrates these changes in a Random Graph setting, while the right panel displays the alterations in a Power-Law Graph context. This side-by-side comparison highlights the impact of regularization on product rankings in different graph types.

Moreover, the endeavor to improve fairness in rankings often brings to the fore the intricate balance that must be maintained with the accuracy of the rankings. This is a critical consideration in the design and implementation of bias mitigation strategies. While our enhanced BiRank algorithm has shown promise in terms of fairness, the trade-offs between fairness and accuracy present an ongoing challenge and an area for future exploration.

Furthermore, it is important to recognize that various methods exist for mitigating bias in ranking algorithms. The choice of method can significantly influence the outcomes, and as such, should be carefully considered in relation to the specific characteristics and requirements of the bipartite graph at hand. Our research contributes to this broader discourse by providing a tangible example of how algorithmic modifications, such as the regularization introduced in BiRank, can yield improvements in fairness while navigating the complexities of accuracy and bias mitigation.

# 5 Conclusion and Further Work

This study has enhanced the BiRank algorithm to better address user activity biases in bipartite graph ranking, particularly for graphs with power law distributions. Our regularization approach has shown promising improvements in fairness, evaluated through a newly developed fairness metric.

Future research directions include optimizing hyperparameters in the enhanced BiRank algorithm and exploring the integration of graph regularization with matrix factorization techniques. These steps could further refine the balance between fairness and accuracy in ranking systems, extending the applicability to more complex scenarios in recommendation systems and beyond. This research lays the foundation for more equitable and efficient bipartite graph ranking solutions in real-world applications.

## Resources

The source code, extension of [3], and materials used to generate the results of this paper are available on our GitHub repository. For full access to the codebase, including scripts and datasets, please visit our GitHub Page.

## References


[1] X. He, M. Gao, M.-Y. Kan, and D. Wang. Birank: Towards ranking on bipartite graphs. *IEEE Transactions on Knowledge and Data Engineering*, 29(1):57–71, 2017.

[2] Hao Liao, Jiao Wu, Mingyang Zhou, and Alexandre Vidmer. Addressing time bias in bipartite graph ranking for important node identification. *Journal or Conference Name*, 2019. Submitted on 28 November 2019.

[3] K.C. Yang, B. Aronson, and Y.Y. Ahn. Birank: Fast and flexible ranking on bipartite networks with r and python. *J Open Source Softw*, 5(51):2315, 2020.

# Appendix

## Setup for BiRank with User Activity Regularization

Given a bipartite graph with user nodes $U$ and product nodes $P$, we analyze the BiRank algorithm with and without user activity regularization. The setup is as follows:

- Users: $U1, U2$
- Products: $P1, P2, P3, P4, P5$
- Regularization formula: $r_i = \gamma \frac{1}{\text{user\_activity}[i]+\epsilon}$
- User activity calculation: $\text{user\_activity}[i] = \frac{\text{degree of user } i}{\sum_j \text{degree of user } j}$
- Regularization parameter: $\gamma = 0.1$

## Adjacency Matrix (W)

We construct the adjacency matrix $W$ to represent the connections between users and products.

$$W = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

## Degree Matrices (D_U and D_V)

The degree matrices for users and products are calculated as:

- User Degrees: $D_U = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$
- Product Degrees: $D_V = \begin{bmatrix} 2 & 2 & 1 & 1 & 1 \end{bmatrix}$

## Normalization of W (W')

The normalized adjacency matrix $W'$ is obtained as:

$$W' = D_U^{-1/2} W D_V^{-1/2}$$

## Regularization Vector (r)

The regularization vector $r$ is calculated based on the user activity:

$$\begin{aligned} \text{user\_activity}[U1] &= \frac{5}{5+2} = \frac{5}{7} \\ \text{user\_activity}[U2] &= \frac{2}{7} \\ r &= \left[ \gamma \frac{1}{\frac{5}{7}+\epsilon}, \gamma \frac{1}{\frac{2}{7}+\epsilon} \right] \end{aligned}$$

## Initial Rank Vectors

The initial rank vectors for users and products are set as:

$$
\begin{aligned}
u^{(0)} &= [1/2, 1/2] \\
v^{(0)} &= [1/5, 1/5, 1/5, 1/5, 1/5]
\end{aligned}
$$

## Iteration Process

The iterative process of BiRank with and without regularization is applied. We present the first two iterations as examples.

### Iteration 1

### Without Regularization

$$
\begin{aligned}
u^{(1)}_{\text{no-reg}} &= W' \times v^{(0)} \\
v^{(1)}_{\text{no-reg}} &= (W')^T \times u^{(1)}_{\text{no-reg}}
\end{aligned}
$$

### With Regularization

$$
\begin{aligned}
u^{(1)}_{\text{reg}} &= (W' \times v^{(0)}) \odot r \\
v^{(1)}_{\text{reg}} &= (W')^T \times u^{(1)}_{\text{reg}}
\end{aligned}
$$

Continue the process for subsequent iterations to achieve convergence.

## Analysis and Conclusion

Our analysis shows that without regularization, user U1 exerts significant influence on the ranking of all products, leading to user activity bias. With regularization, the influence of U1 is scaled down, resulting in a more balanced and equitable ranking of products. This demonstrates the effectiveness of user activity-based regularization in addressing user activity bias in the BiRank algorithm.

# Proof of Convergence for Enhanced BiRank Algorithm

**Theorem:** The enhanced BiRank algorithm, which incorporates a regularization term, converges under standard conditions.

**Proof:** We base our proof on the convergence properties of the original BiRank algorithm and extend it to include the regularization term.

**1. Original BiRank Convergence:** The original BiRank algorithm's convergence, as established, depends on the hyperparameters $a$ and $b$, both of which are in the range [0, 1]. The convergence analysis involves three scenarios:

*a) Boundary Cases:* When $a = 0$ or $b = 0$, the vectors $p$ and $u$ remain unchanged after the first iteration.

*b) Complete Graph Dependence:* When $a = 1$ and $b = 1$, the ranking depends solely on the graph structure. The iterative process then converges to the principal eigenvector of the matrices $STS$ and $SST$, as per standard linear algebra.

*c) Normal Cases:* For $a, b$ in (0, 1), the convergence to a stationary solution is assured. The algorithm converges to the equation $p = (I - abSTS)^{-1}[a(1-b)STu_0 + (1-a)p_0]$, and a similar form for $u$.

**2. Incorporating Regularization:** In the modified BiRank algorithm, we introduce a regularization vector $r$, defined as $r_i = \gamma/(user\_activity[i] + \epsilon)$, into the ranking process.

*a) Adjusted Update Rules:* The update rules for $p$ and $u$ are modified to include the regularization vector $r$. Specifically, the ranking vector $u$ is updated as $u^{(k+1)} = \left(\alpha W' v^{(k)} + (1-\alpha)q_U\right) \odot r$, where $\odot$ denotes element-wise multiplication. This modification aims to scale down the influence of highly active users, ensuring a more balanced representation in the ranking process.

*b) Convergence with Regularization:* To establish the convergence of this modified algorithm, we need to consider the impact of the regularization vector on the iterative process. The introduction of $r$ affects the magnitude of the ranking scores but preserves the fundamental properties of the iterative update rules. Since the regularization vector $r$ is fixed across iterations and the original BiRank algorithm is proven to converge, the modified algorithm, with its adjusted update rules, also converges to a stationary solution.

*c) Eigenvalue Analysis:* As in the original BiRank algorithm, the eigenvalues of the matrices involved in the iterative process are crucial for convergence. The regularization term, being a scalar multiplication in the update rule, does not alter the eigenvalue bounds required for convergence. Therefore, under the assumption that the eigenvalues of the matrix $STS$ (and similarly $SST$) are within the required bounds, the modified algorithm converges.

In conclusion, the modified BiRank algorithm, with the inclusion of a regularization term, preserves the convergence properties of the original algorithm. Under standard conditions, namely the eigenvalues of the involved matrices being within appropriate bounds and the hyperparameters $a$ and $b$ being in the range $[0, 1]$, the algorithm converges to a stationary and unique solution. This convergence, coupled with the intended impact of the regularization term, ensures the algorithm's effectiveness in providing a balanced and fair ranking in bipartite graphs.

# Intuitive Proof of Speed of Convergence for Enhanced BiRank Algorithm

In the enhanced BiRank algorithm, the convergence speed is closely tied to the graph's structure, the second largest eigenvalue of the matrix $STS$, and the chosen hyperparameters $\alpha$ and $\beta$. The lower magnitude of the second largest eigenvalue generally leads to quicker convergence. However, larger values of $\alpha$ and $\beta$ may slow down this process. In real-world applications involving sparse, high-dimensional data, the matrix $S$ typically has low-rank properties, aiding in faster convergence. The addition of the regularization term in the enhanced BiRank algorithm, however, introduces additional complexity, which can slightly slow down the convergence compared to the original BiRank algorithm. This change necessitates a careful balance in hyperparameter selection to maintain efficient convergence rates while achieving the desired regularization effect.