North Carolina Agricultural and Technical State University

Project Progress Report

Rethinking NBA Player Rankings

March 2025

By

Elijah Shepherd

Mentor: Dr. Laurie Zack

North Carolina Agricultural & Technical State University

Contents

Abstract

NBA player rankings often rely on statistical analysis, media polls, and awards. This study introduces a new ranking system for these players using the modified PageRank algorithm, accounting for both direct and indirect player contributions. We investigate whether this system effectively predicts top performers and reveals underrated players. Adapting the PageRank algorithm, originally for web pages, we model player interactions as a network. Players are nodes, and statistical interactions (assists, on-court impact, win contributions, offensive/defensive rating, net rating, box plus-minus) define connections. The algorithm distributes ranking scores based on these interactions, rewarding players who contribute to winning teams or perform well against top competition. We pulled our data from an NBA data site known as 'Basketball Reference'. We hypothesized that our system would identify both top players (validated against All-NBA selections) and overlooked contributors. Our system was able to correctly identify All-NBA team members 85 percent of the time based on the produced ranking. This study demonstrates PageRank's potential for more objective evaluation of NBA players. Future work could refine metrics and incorporate factors such as player roles and team dynamics. This data-driven system has implications for player evaluation, awards, and player contracts.

1 Introducion

The NBA, which stands for National Basketball Association, is filled with the best basketball players from around the world. Players from all different parts of Europe, Africa, Asia, and many more make up the NBA's diverse population. Due to this vast pool of talent, the ranking of these players has long been a topic of discussion among fans, former players, and media personalities.

In the past, the way players were ranked was through a combination of traditional and advanced statistics, media-driven narratives, and expert opinions. All of these converging opinions combined to produce All-NBA teams at the end of each season, as well as MVP votes. Traditional statistics look at things that can be tracked from game to game. "How many points does a player average for the season?" "How many assists, rebounds, steals, etc..". These things can be tracked over the course of a game, and require a basic computation of averages. Advanced statistics, like the below image, takes traditional statistics and tries to apply different weights or formulas to them for more insight. These statistics are more complex, and often require a combination of traditional statistics in order to be computed.

Figure 1: Advanced Statistics

1.1 Problem Statement

Why is this a problem? To put it simply, a lot of people don't understand how to apply context to the numbers that they are looking at. Fans and T.V media personalities often try to push narratives or opinions about certain players or teams. Citing different statistics or reasoning for their rankings, without providing why those statistics are significant. Take the following image for example.

Figure 2: MVP Odds

This was an official graphic used by ESPN, a sports network channel. They were discussing the battle for the MVP award, and decided to showcase the Vegas Super-book to show the odds of certain players to win MVP. This graphic has no other source, no other statistics to back it up, or any reasoning behind it other than these were the players that people were placing the most bets on to win MVP. This highlights how often these rewards are influenced by the public (betting books). This use of graphics is very common, with networks showing a ranking of players from an analyst or even a poll, and explaining their reasoning for about 30-45 seconds. This easily influences public vote for MVP, All Star selections, and All-NBA votes.

1.2 Comparisons

Traditionally, when ranking players fans tend to look at them individually. Often citing numbers and statistics of player A to compare to player B. However, these numbers are brought up separately. For example, let's say that someone comes up to you and says that Kyrie Irving is a better player thank Damian Lillard. Often, people will bring up Kyrie's own statistics, and compare them to Damian's Stats.

Why is this a problem? Comparing players in this way often takes out the actual comparison part. There is nothing in place that allows for people to see how players actually played against each other. When discussing the best of the best, they should be compared to each other. While looking at individual statistics is important, it may not tell the full story as to how some players play against the best.

2 Focus and Research Questions

The focus of this research was to develop a PageRank algorithm based on player's statistical interactions with each other. Factoring in how highly ranked players performed against other highly ranked players. From this, patterns could be identified such as is there a specific type of player that is rated low in the algorithm but high in by traditional NBA ranks. Creating a more objective and non-bias ranking system based on player performance against one another is the main goal. From this, applying this ranking system to multiple positions and using multiple metrics will really round out this study.

2.1 Research Questions

Some of the research questions from this project are as follows. Are there players that are underrated or fly under the radar as far as voting and awards go? Are there certain metrics that contribute most to a player's ranking? Did the algorithm do a good job at ranking players compared to their actual averages for that year? Which players saw award based success that did not rank highly in the algorithm? Will this algorithm be able to accurately rank players in the future based on the captured results?

3 Method

In this research the goal was to highlight a potential new and more effective way to rank players. To achieve this, a modified version of the PageRank theory algorithm was utilized. The goal is to add context to traditional and advanced statistics in order to provide a more objective ranking of players.

3.1 The PageRank Algorithm

The PageRank algorithm gives each page a rating of its importance, which is recursively defined measure whereby a page becomes important if important pages link to it. One way to think about PageRank is to imagine a random surfer on the web, following links from page to page [Wills, 2006]. The page rank of any page is roughly the probability that the random surfer will land on a particular page. Since more links go to the important pages, the surfer is more likely to end up there. Let's look at this visual representation of how PageRank works.

The first four players in the points matrix were used to create a visual example. So, Trae Young (T.Y), Jrue Holiday (J.H), Terry Rozier (T.R), and Spencer Dinwiddie (S.D) were selected. The goal is to showcase how these players statistics, in this case points, interact with each other within our algorithm. The way players are linked is shown below.

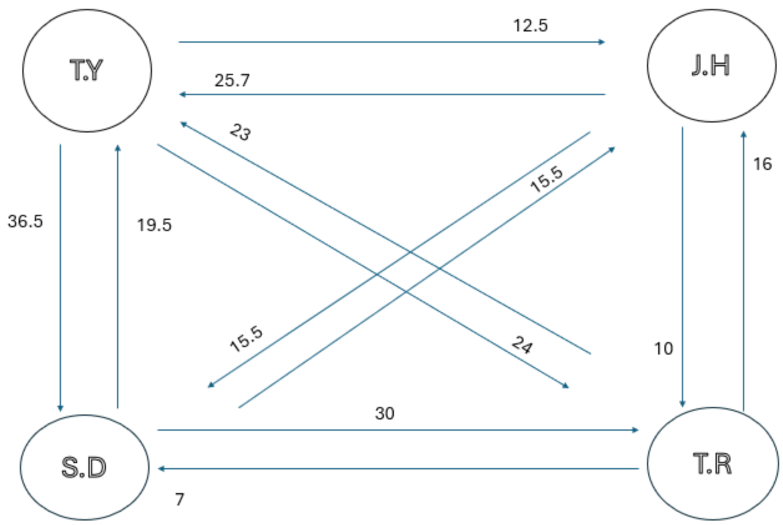

Figure 3: Visual Representation

The arrows represent the link between the given players. That "link" represents the amount of games played against each other. The numbers that you see on top of or next to the arrows represent the average number of points a player had against another. An arrow pointing towards a player indicates the value that player had when playing against the player that the arrow originated from. For instance, Trae Young averaged 25.7 points against Jrue Holiday and Jrue averaged 12.5 points against Trae. So, the arrow that points from Trae to Jrue has 12.5 on top of it, and the arrow pointing from Jrue to Trae has the number 25.7 on top of it. Between two players, the one with the higher average is weighted higher in our algorithm. In this case, Trae's average was higher so this positively attributes to his rank.

3.2 How to Find PageRank

Let's say that you are given three web-pages A, B, and C. How would one go about obtaining a PageRank for each of these pages? Consider that figure 3, is converted to a hyperlink matrix P. The hyperlink matrix P for three web pages (A, B, and C) is given by:

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

A row stochastic matrix S is formed from P and is then used to model the random web surfer with the equation: $G = \alpha S + (1-\alpha)yv$ where α is the damping factor, y is a column vector of ones, and v is called the personalization vector. The vector v is a probability distribution vector, and is currently unknown. However, during the development of this algorithm by Sergey Brin and Lawrence Page [Brin and Page 1998; Page et al. 1999], v was defined as:

$$v = \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix}$$

The damping factor models the random web surfer's ability to move to a different webpage by means other than following a link, with probability $(1-\alpha)$. The damping factor used was $\alpha = 0.85$. In most research done since 1998, values of α range between 0.85 and 0.99 [Wills 2006]. For this example and throughout the paper, $\alpha = 0.85$ will be used, and because there are three webpages in this example, the personalization vector v is given by:

$$v = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Using the equation $G = 0.85S + 0.15yv$ we obtain the Google matrix G:

$$G = \begin{bmatrix} \frac{3}{15} & \frac{7}{15} & \frac{5}{15} \\ \frac{5}{15} & \frac{3}{15} & \frac{7}{15} \\ \frac{7}{15} & \frac{5}{15} & \frac{3}{15} \end{bmatrix}.$$

The PageRank vector π is then found by computing the corresponding left eigenvector satisfying: $\pi G = \pi$ [Zack Lamb, and Ball 2012]. Since G is row stochastic, 1 is the dominant eigenvalue, which means π can always be computed [Brian and Leise 2006]. To compute this, the power method was used. The power method is a iterative algorithm that can be used to determine the largest eigenvalue and it's corresponding eigenvector [Brian and Leise 2006]. It does so by starting with a random initial vector, and then multiplying the original matrix to the vector. Then, you normalize the results, and repeat the process over and over to get improved approximations for the eigenvector associated with the largest eigenvalue. For this particular matrix, the PageRank vector is approximately: $\pi \approx (0.370, 0.330, 0.300)$. Therefore, the webpage ranking listed from most important to least important is: A, B, C.

3.3 Relating to the project

Using the established idea of RankTheory, the web pages are replaced with the NBA players themselves. The connection between the players are going to be represented by the statistical connections in their head to head match-ups. This allows for a better insight into the statistical impact of players in relation to one another. For instance, comparing how many points per game player A averaged against player B in the 2023-2024 season. If player A averaged more points against player B, then the connection arrow would point to him in his favor, boosting his rank score. A player's rank score would improve if they performed well against other highly ranked players.

Do certain players with specific strengths go undervalued? Are certain players possibly overrated by the media and fans? Do players benefit from the team that they play on, causing a potential boost in their numbers? Do some players struggle against a specific player or players? All questions and more are hoped to be answered with this study.

3.4 Data Collection

The dataset for this study was manually compiled to analyze the performance of NBA starting point guards against each other during the 2023-2024 NBA season. The primary data source was "Land of Basketball", or visit: `https://www.landofbasketball.com/games_between.html`, a reputable database that provides comprehensive statistics on NBA players and teams, as well as "StatMuse" or visit `https://www.statmuse.com/nba`, and "Basketball Reference or visit `https://www.basketball-reference.com/`, which both work a similar function. To facilitate a structured comparison, a 29×29 matrix was created, where each row and column represented one of the 29 starting point guards in the league. This matrix was used to input the needed averages for players.

Each cell in the matrix contains the individual statistical averages of a given guard when matched up against another starting guard. These statistics were manually extracted and inputted into the dataset, ensuring precision in data recording. The collected metrics include key performance indicators such as points per game (denoted as PPG), and assists per game (denoted as ASPG), both a simple but vital indicator of how good a point guard is. "Points" are defined as made field goals during live or dead play. Points can be worth 1 for free throws, 2 points for a basket made inside the three point line, and 3 points for a basket made behind the three point ine. "Assists" are defined in basketball as a pass that leads directly to a field goal. These matrices will be compared by utilizing the PageRank method established in section 0.2.1. This will produce a new matrix, ranking players by their offensive impact. Players would have to perform well in both scoring and passing for a higher rank, as only performing well in one may not cause a high ranking.

By structuring the dataset in this way, we enable direct player-to-player comparisons. Offering insights into matchup-specific performance trends rather than season averages. This provides a valuable foundation for further statistical analysis, including predictive modeling and matchup-based performance evaluations. When players did not play each other, it was originally logged by "DNP". This was then replaced by that players season averages. This was done to handle the missing data within the dataset without removing it.

4 Results and Discussions

Two 29 x 29 matrices were created, one showcasing players points per game in head-to-head matchups, and one showcasing a players assists per game. Currently, the process of extracting PageRank scores when weighting the points and assists matrices against each other is still in progress. The points and assists matrices were transformed into Row Stochastic matrices. From there, using the method from Section 0.2.1, matrix G was created. From there, the matrix of G was transposed, and the "power.method" function within R was used. This function allows to input the function we want iterated (the transpose of matrix G), the maximum number of iterations (maxiter = 100), and the how close you want your approximation to be to the original matrix before it stops (tol = 1e-10).

Top 10 Point Guards by Points Per Game

Rank	Player	PageRank Value
1	Luka Doncic	0.05769
2	Shai Gilgeous-Alexander	0.04994
3	Jalen Brunson	0.04683
4	De'Aaron Fox	0.0447
5	Stephen Curry	0.04454
6	Trae Young	0.04444
7	Devin Booker	0.0433
8	Tyrese Maxey	0.04292
9	Damian Lillard	0.04063
10	Cade Cunningham	0.03888

Figure 4: A look into the top 10

The points per game matrix was looked at first. After normalizing the results from applying the power.method, the PageRank values ended up producing the above ranking. Compared to the top ten scoring leaders list at the point guard position, the list looks as follows:

1. Luka Doncic
2. Shai-Gilgeous Alexander
3. Jalen Brunson

4. Devin Booker
5. De'Aaoron fox
6. Stephen Curry
7. Tyrese Maxey
8. Trae Young
9. Damian Lillard
10. Cade Cunningham

The top 10's are close, which shows that our PageRank does a good job at ranking the best scores in the leauge. However, it is interesting to note that Devin Booker was apart of the top 4 scorers at the point guard position during this year. However, within our rank he was ranked below the top 5. Even when changing the weights he still never cracked the top 5 in the rank. This could be a sign that during the regular season he might not have performed that well against the highly ranked players. Also, Trae Young is two spots higher in the PageRank than he is on the top scorers list. It is possible that he played slightly better against good competition.

Top 10 Point Guards by Assists Per Game

Rank	Player	PageRank Value
1	Trae Young	0.05549
2	Tyrese Haliburton	0.0526
3	Luka Doncic	0.05
4	James Harden	0.04493
5	Fred VanVleet	0.04314
6	Tyus Jones	0.04249
7	Damian Lillard	0.03845
8	Cade Cunningham	0.03814
9	D'Angelo Russell	0.03267
10	Jalen Brunson	0.03205

Figure 5: A look into the top 10

Now we can move on to the assist matrix. In this matrix our top ten ranking came out as seen above. The top assist leaders for th year were as follows:

1. Tyrese Haliburton
2. Trae Young
3. James Harden

4. Luka Doncic
5. Fred Vanvleet
6. Cade Cunningham
7. Tyus Jones
8. Damian Lillard
9. Jalen Brunson
10. Jamal Murray

Now these results were a little more interesting. For one, Trae Young and Tyrese Haliburton were flipped in our PageRank list. From the numbers, Trae did have a great year when it offensively against highly ranked players. Tyrese on the other hand did not see that same consistent success, so this could have attributed to their rankings. Everyone else seemed to be one or two spots off in the Page ranking.

4.1 Combined Ranking

Top 10 Point Guards by Points and Assists

Rank	Player	PageRank Value
1	Luka Doncic	0.05246
2	Trae Young	0.0521
3	Tyrese Haliburton	0.04734
4	James Harden	0.04425
5	Fred VanVleet	0.03849
6	Damian Lillard	0.0395
7	Jalen Brunson	0.0393
8	Cade Cunningham	0.03829
9	Shai Gilgeous-Alexander	0.03786
10	Devin Booker	0.03746

Figure 6: A look into the top 10

In our combined ranking we ended up with the above results. For one, Trae Young ended up ranking as the second best point guard in the leauge by our metrics. However, he was nowhere to be found on the All-NBA teams at the end of the year. Statistically speaking, Trae had one of the better years for a point guard, and played well against other competition. The ranking that was found highlights this, and shines light on the underrated year that he had. On the flip side, Luka Doncic ranked number 1 in the ranking and this is right in line with NBA rankings. Luka arguably had one of his best seasons this year, making the All-Star team and being named to All-NBA first team. He also ended up taking the Dallas Mavericks to the NBA finals, looking every bit like the best point guard in the league with the numbers to back it up.

4.2 Anomalies

An interesting player that is worth mentioning was Stephen Curry. Within the raking, Stephen did not make the top 10 of best point guards in the combined ranking. In the points ranking he ranked 8th, and in the assists ranking he ranked 17th. However, he was still named to both the All-Star team and received Third team All-NBA honors. One could say that team success maybe attributed to this but the Golden State Warriors did not make the playoffs last year, losing to the Sacramento Kings in the first Play-in tournament. If you don't know, Steph has consistently been one of the leagues best players since 2015. Winning MVP's, championships, being nominated to multiple All stars, and having more accolades than one could name. This was an obvious down year for him though, seeing a dip in his numbers across the board and an obvious dip in team success.

Trae Young experienced almost the opposite. Trae was in his 6th season, and although he had received an all star appearance he was not named to All-NBA team. Furthermore, he only appeared in the All-Star game because the original pick (Julius Randle) was injured. By the metrics, he was one of the better point guards in the league. Ranking top 5 in both assists and points rankings as well as number 2 in the combined rank. It was a shock to see that his name was not brought up amongst the best. His team even performed better than the Steph Curry's Warriors. Trae seems to have fell victim to a lack of media coverage, a lack of sustained excellence, or a lack of team success. By the measures though, he was one of the top point guards in the league.

Fred Vanvleet also cannot go unmentioned. He was not selected to any all star teams, nor All-NBA teams. However, he was ranked as high as 5th in the PageRank algorithm. He was also top 5 in the assist ranking, so his high overall ranking is not too much of a shock. Fred is a good player for the Houston Rockets, often asked to carry a good amount offensively for them. He had a potentially underrated year and the algorithm highlights this.

Shai-Gilgeous Alexander had to be the biggest surprise. Shai had a breakout year, averaging 30.1 points per game, and leading the Oklahoma City Thunder to the third best record in the league. Shai was an All-Star starter, made an All-NBA first team, and was second in league wide MVP voting(!!!). He was very low in the ranking, showing that there are other metrics to account for when trying to determine who's the best point guard in the league. Going forward, adding more metrics to our algorithm will be key.

4.3 Final Thoughts

The PageRank algorithm was able to highlight lesser-known players who are integral to team success but might not have standout individual stats. For example, a player like Tyus Jones, whose stats might not be as flashy but is known for his scoring and playmaking, received a relatively high PageRank score(0.03718 which would have ranked him 11th)due to his importance in team dynamics. Some players may not be the best on the team, but may have to do a decent amount of scoring and playmaking when the best players are not playing well or are injured. For instance, take Terry Rozier for example. Terry is not the Miami Heat's main scoring option, however he is a important of keeping the offense moving when things get stagnant.

This PageRank model did a relatively decent job at highlighting potentially underrated players. Guys like Trae Young, Terry Rozier, or Cade Cunningham were shown to have high PageRank scores highlighted by their role within their respective offenses. On the other hand, someone like Stephen Curry maybe saw a dip in in his rank due to a potential down year. The PageRank algorithm created lined up well with end of year ranks. All of the top 10 was either an All Star or an All-NBA selection.

5 Conclusion and Future Work

Some players were ranked a little lower than the general public would probably agree with and this is due to a few reasons. For one, this study was conducted only looking at two metrics, PPG and ASPG. Players efficiency when scoring the basketball was something that was not accounted for. A stat like a player's True shooting percentage, or their effective field goal percentage could give a better look into how efficient they were in scoring. Looking at players ability to effectively distribute the ball is also important. For instance, looking at how many turnovers did they average in head to head matchups, what was there turnover to assist ratio, etc. Looking at these efficiency related metrics would further improve the accuracy of the rankings. In the future, adding more advanced metrics to look at one's efficiency would be key.

In the future, ranking players of any position by any metric is an exciting reality. This will allow teams to look into rankings for all types of playstyles for any position that they might need. Let's use an example, say the Boston Celtics need a new Center. They want their Center to be able to rebound and play defense really well. This means that the following metrics could be incorporated into our algorithm; blocks, steals, defensive win shares, rebounds, and any other stat involving rebounding or defense. Following the steps to obtain the PageRank score one could easily find out who the best Centers were that fit the criteria. Then, the Celtics could try and go after the best players through a trade or free agency.

All in all, this is a good first step in providing a more concrete approach to ranking players. Allowing for fans and media members to refer back to. If fleshed out, it would be a very valuable addition for team's front offices in regards to decision making based on players. Factoring how all of this statistics impacted winning as well would allow for using this ranking em to determine which players they might want to target in free agency that could go under the radar.

6 References

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine Computer Networks and ISDN Systems, 30(1-7), 107–117. https://doi.org/10.1016/s0169-7552(98)00110-x

Bryan, K., & Leise, T. (2006). The $25,000,000,000 Eigenvector: The Linear Algebra behind Google. SIAM Review, 48(3), 569–581. https://doi.org/10.1137/050623280

[Page et al. 1999] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," tech report, Stanford InfoLab, 1999.

[Wills 2006] R. S. Wills, "Google's PageRank: The math behind the search engine," *SIAM Rev.* 48:3 (2006), 569–581.

[Zack, Lamb, and Ball 2012] L. Zack, R. Lamb, S. Ball, "An application of Google's PageRank to NFL rankings,“ *Involve* No.4, 463-471 (2012)

Basketball Reference. (n.d.). Basketball Statistics and History — Basketball-Reference.com. Basketball-Reference.com. https://www.basketball-reference.com/

"A look into player's advanced statistics.“

NBA Players Head-to-Head: Games Played Between Two NBA Stars. (2025). Landofbasketball.com. https://www.landofbasketball.com/games_between.htm

"Head to head matchups for the 2023-2024 season.“

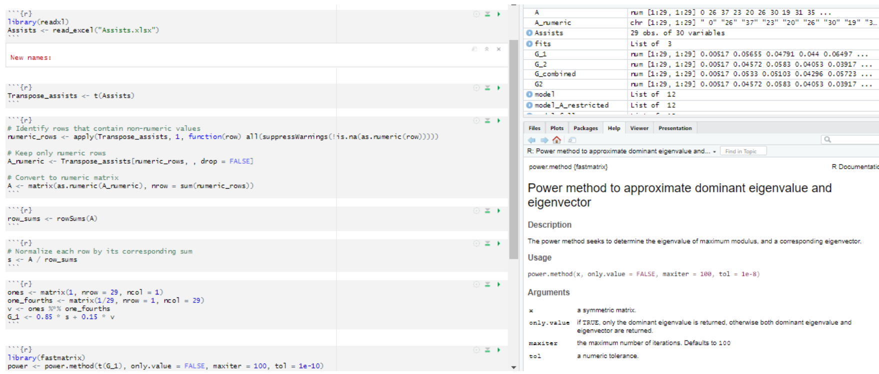

Figure 7: Full Code

A Code Scripts

Here is some images showcasing how my code flowed.

```
```{r}
library(readxl)
Assists <- read_excel("Assists.xlsx")
```
```

Figure 8: Loading Data

First we start by loading the 29 x 29 matrix that was created

```
```{r}
Transpose_assists <- t(Assists)
```
```

```
```{r}
# Identify rows that contain non-numeric values
numeric_rows <- apply(Transpose_assists, 1, function(row)
all(suppressWarnings(!is.na(as.numeric(row)))))
```

Figure 9: Transpose

Due to an initial error in entering the data, transposing the initial data is necessary.

```
# Keep only numeric rows
A_numeric <- Transpose_assists[numeric_rows, , drop = FALSE]

# Convert to numeric matrix
A <- matrix(as.numeric(A_numeric), nrow = sum(numeric_rows))
```
```

```
```{r}
row_sums <- rowSums(A)
```
```

Figure 10: Row Sums

After transposing the initial matrix, converting it into a fully numeric matrix was necessary. After that, the row sums of each row were found.

```{r}
# Normalize each row by its corresponding sum
s <- A / row_sums
```

```{r}
ones <- matrix(1, nrow = 29, ncol = 1)
one_fourths <- matrix(1/29, nrow = 1, ncol = 29)
v <- ones %*% one_fourths
G_1 <- 0.85 * s + 0.15 * v
```

Figure 11: G-Formula

First, we normalize the S matrix by dividing the A matrix by the row sums. Then, create the matrix of all 1's and the matrix of all 1/4's. Multiply the two together to get "V". Plug V into the formula for G.

```{r}
library(fastmatrix)
power <- power.method(t(G_1), only.value = FALSE, maxiter = 100, tol = 1e-10)
```

Figure 12: Power Method

Here, the power function within R was used. This function allows for an input of your original function. "only.value" if TRUE, only the dominant eigenvalue is returned, otherwise both dominant eigenvalue and eigenvector are returned. "Maxiter" is the number of iterations that you want to multiply your original matrix to the random vector. "tol" is the tolerance, signifying how close the results need to be from the previous result in order to stop.

```{r}
library(ohenery)
pagerank_2 <- normalize(power$vector)
pagerank_2
```

```
Warning: package 'ohenery' was built under R version 4.2.3 [1] 0.05549597 0.02815756
0.02658112 0.03161234 0.02881470 0.03424416 0.05000414
 [8] 0.03272244 0.03818139 0.02847489 0.04319116 0.05280750 0.04493243 0.03627013
[15] 0.02424689 0.03895369 0.03061419 0.02723742 0.03605018 0.03276212 0.01820311
[22] 0.03286751 0.03514096 0.02873901 0.02936439 0.03436345 0.03187385 0.02568788
[29] 0.04240543
```

Figure 13: Results

Here, the ohenery package in R was loaded so that the normalization of the power vector could be performed. Afterwards, the final outputted PageRank values appear as seen.