

# Application of Google's PageRank to Pro Boxers

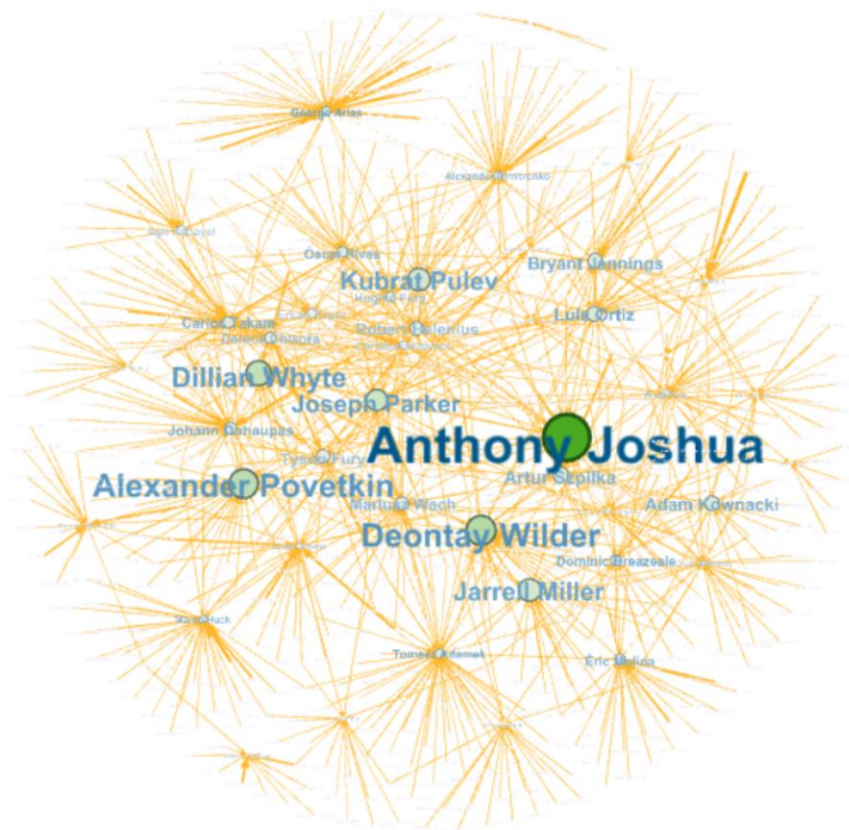
Tommy Lang

3/15/19

## 1 Introduction

In any competitive sport there is a natural desire to know who the best is. However, in boxing the rankings have been criticized as criminally suspect. Spoiled by monetary incentives, boxing promoters often create mismatched bouts to inflate their boxers' records. Also, there are 5 sanctioning organizations that award titles in trivialized weight classes resulting in the potential of over a 100 boxers named as "world champion" at any one time.

In this document I describe an objective method to rank boxers and to determine who is the best current heavyweight boxer. I structure boxers into a network and use Google's PageRank algorithm to quantify which boxer is the most important in the network. The graph below shows the network of boxers where the links between them represent a boxing match outcome between the fighters and the size of a boxer's name is proportional to his rank.



## 2 Data Collection

The win / loss record of the top 50 current heavyweight boxers as ranked by boxrec.com was extracted from Wikipedia.com. These boxers were chosen mainly for convenience. Ideally, I would have liked to have a chained sample with the top 50 boxers as seeds but that proved difficult to obtain. With the convenience sample, boxers in the network won't get as much credit for defeating some of their opponents but that's an acceptable compromise because a boxer that isn't ranked in the top 50 are likely not very good. Only current active heavyweight boxers were sampled because boxers from different weight classes and eras are unlikely to be linked together.

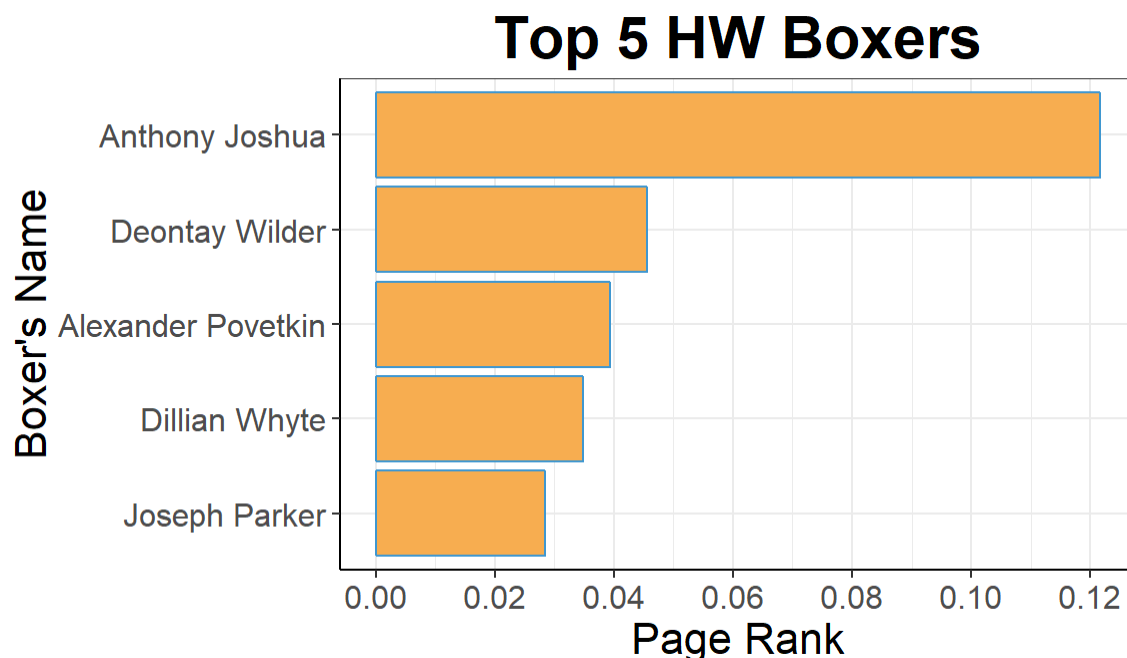
## 3 Network Definition

The network contains 783 nodes and 1075 links / edges. Each node represents one of the top 50 heavyweight boxers or a boxer that they defeated. The directed links represent a boxing match that resulted in a win going from the losing boxer to the winning boxer. The weights of the links are based on the type of resulting win. If the win was more definite (e.g. knockout) it was weighted higher than a win that's more open to interpretation (e.g. split decision).

## 4 Boxer's Rankings

Google's PageRank algorithm was used to rank the boxers in the network. The algorithm is the most well-known method that Google's search engine used to rank webpages in search results. It was designed so that webpages that received a lot of links from other webpages that received a lot of links would be ranked higher. It's appropriate for the network of boxers because boxers that defeated other quality boxers should be ranked higher. The damping factor of the algorithm was set to 0.999 so that boxers with more quality wins "absorb" more of the importance in the network and are ranked higher.

### 4.1 Analysis of Rankings



The top 5 boxers in the network is displayed in the bar chart above. The results are reasonable except for the notable exclusion of Tyson Fury. He is rightfully in anyone's top 5 list but is ranked number 22 by the algorithm. The main reason for his lower ranking is because the network doesn't account for his defeat of Wladimir Klitschko. Klitschko was dominate in his era and including him in the sample would have advanced Tyson Fury to be ranked second.

Also, the algorithm assigns too much importance to boxers with a lot of wins against low quality boxers. For example, Tomasz Adamek is ranked number 10 because he has 51 links / wins, but those links are to boxers that don't have incoming links to them. Eric Molina knocked out Adamek but Molina is ranked lower at 17 because he only has 24 links. I believe some these issues would be solved if a larger sample was used which would result in a network that's more interconnected (current network density: 0.0018).

## 5 Conclusion

An objective ranking of boxers creates more credibility and interest for the sport. I applied Google's PageRank algorithm to rank current heavyweight boxers and the results were mostly consistent with intuition. A larger sample of boxers and some modifications to the algorithm could provide more reasonable rankings.