

WHO TO FOLLOW ON TWITTER

Group 7: POSTER 2

WHO TO FOLLOW ON TWITTER

Group 7: POSTER 2

The experiments

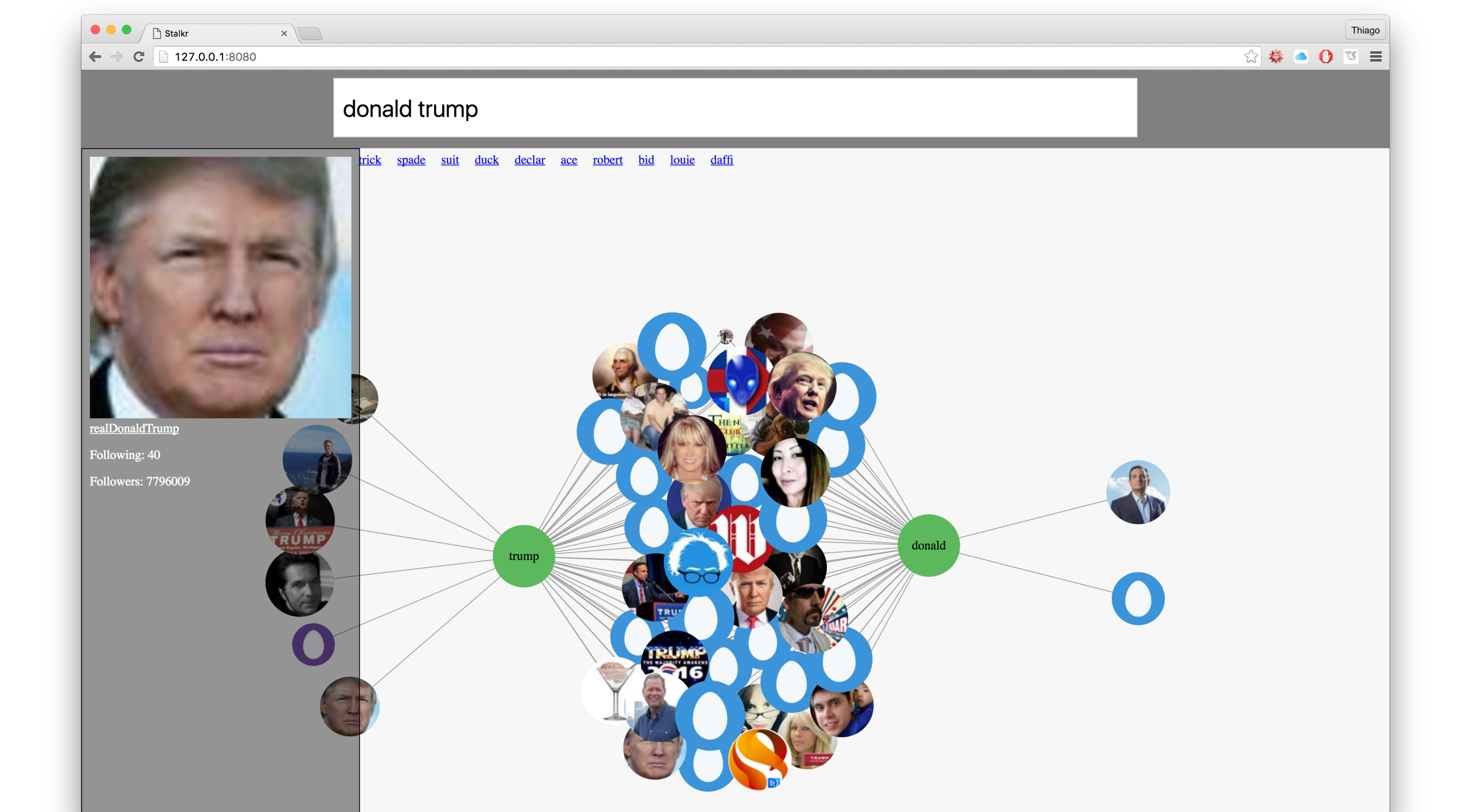
In both of our experiments, a toy query, "usa election president politics " was used. We first tried using only tf-idf and then a combined retrieval with standard PageRank, both of which were compared to combined retrieval with our novel PageRank variant, which considers how many followers a user has. We graded the top ten results of each retrieval as relevant or not and plotted a precision vs recall graph for each of them. Whether a user is relevant or not was decided subjectively by averaging the votes of each member of the group.

Conclusion

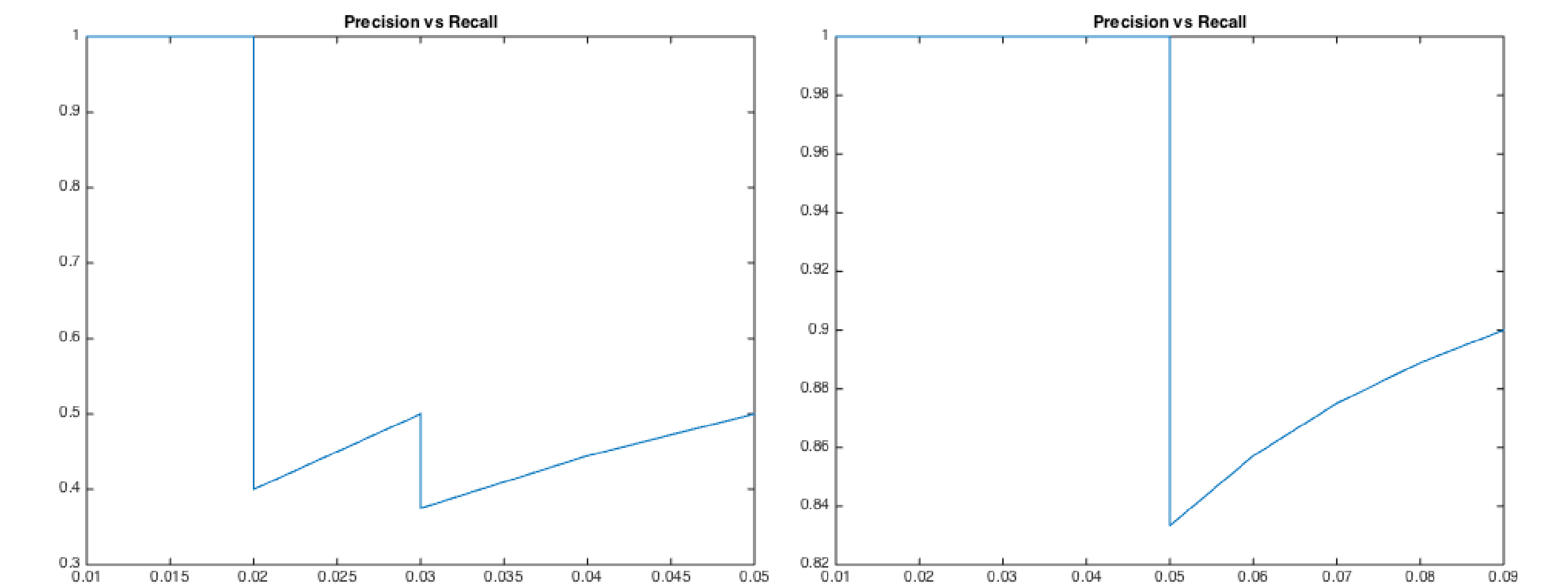
It is feasible to create a Twitter user recommender engine using tf-idf and PageRank. Twitter data is suitable to be represented as a graph. Looking at the results, using appropriate parameters for weighing tf-idf and PageRank yields a result that is satisfying. Measuring the extent of the satisfaction is subjective, but for the group of this project it is deemed that the result is satisfactory enough for the given task.

Due to the limited amount of data and the fact that the data is collected using a limited set of correlated hashtags, it is hard to predict if the results are applicable to a larger more representative dataset. It would perhaps be better to use a larger dataset with tweets from several different topics. However, considering the limited amount of time and resources, the project showed that the task is feasible.

Screenshot 1

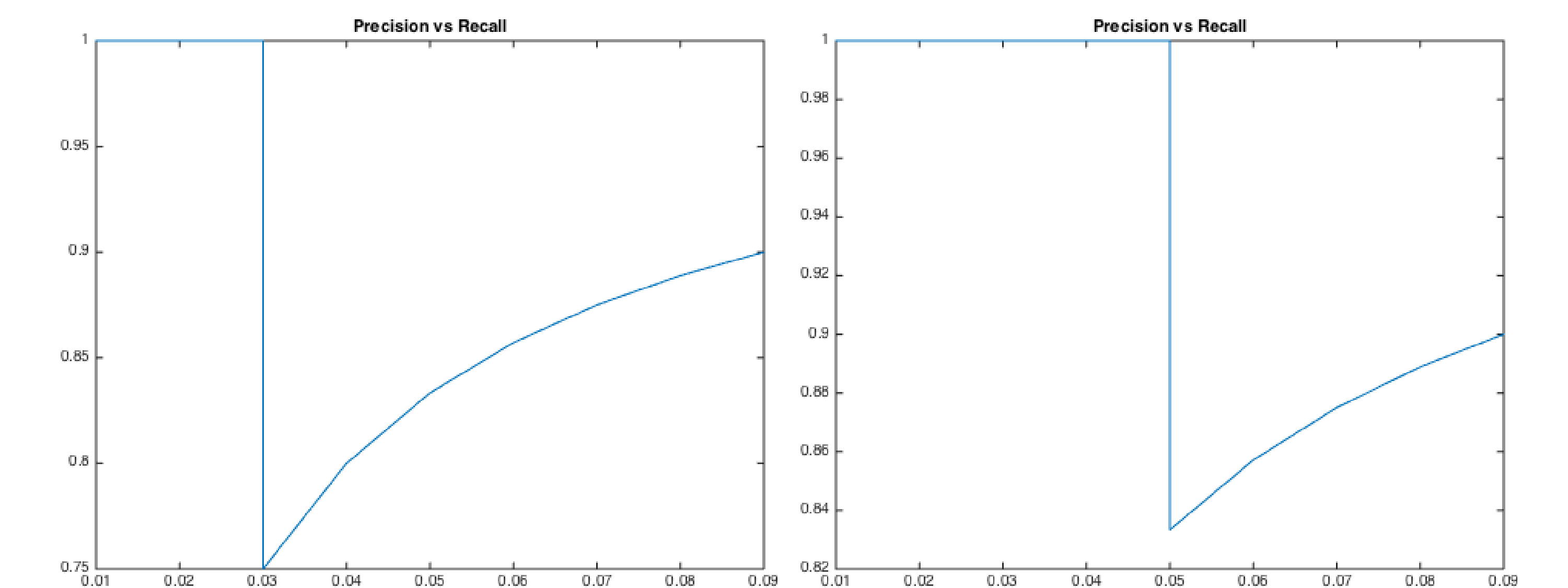


Result: tf-idf vs combination (tf-idf and PageRank weighted by #followers)



tf-idf results in more irrelevant users (documents) because it only considers words, whereas the combination with PageRank also takes into account the structure of the data (user authority).

Result: PageRank simple vs PageRank weighted by #followers



Not only can we consider the links between user nodes, but we can also weigh the authority by the number of followers of each user. This yields better results as shown above.

Screenshot 2

