Facebook Game Network Analysis

Suhaas Prasad & Divye Khilnani

Introduction

Online social gaming has become increasingly popular, with applications on Facebook growing at viral rates unheard of in the gaming industry. With more than 200 million people playing social games every month, this industry is one of the fastest growing tech sectors, having only come into being fairly recently. In order to better understand the growth of these games and the interactions that occur within them, we analyze the data generated by the What To Wear Facebook application. The What to Wear application is an online social game in which players create virtual outfits to submit to daily competitions that need to match a certain theme. These outfits are voted on by the contestants who interact with each other by commenting on and judging each other's outfits. The items for these outfits are bought through real-life money or in-game credits, which are collected by participating in competitions and judging other outfits. The game has an active membership of approximately 30,000 players each month who interact with each other and participate in the daily contests.

The data given by Turiya Media that we analyze includes player attributes, competition histories, logs of game events, information about the virtual items, transactions with real money and credits, and various other relations between the aspects of the game. However, we would like to mention that data was sparse for many of the attributes thus limiting our analysis. We studied the network between players and how the aspects of the game affect their interactions in various ways. The data can be represented as a graph by creating nodes for each player and an edge between players when one interacts with another. Interaction in this case refers to a player judging another player's outfit. We then define communities within the network to be densely connected subgraphs of the larger network. (something on community detection)

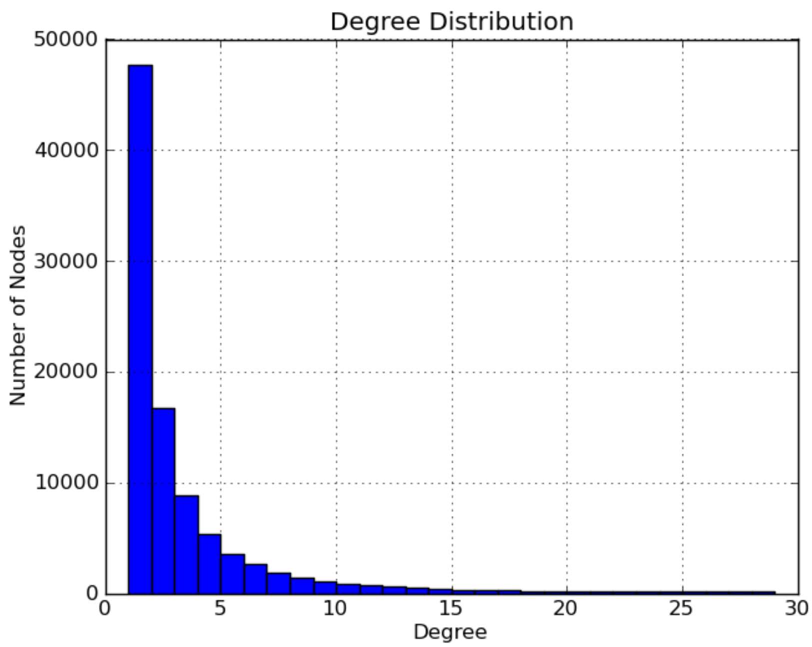
Network Generation and Structure Analysis

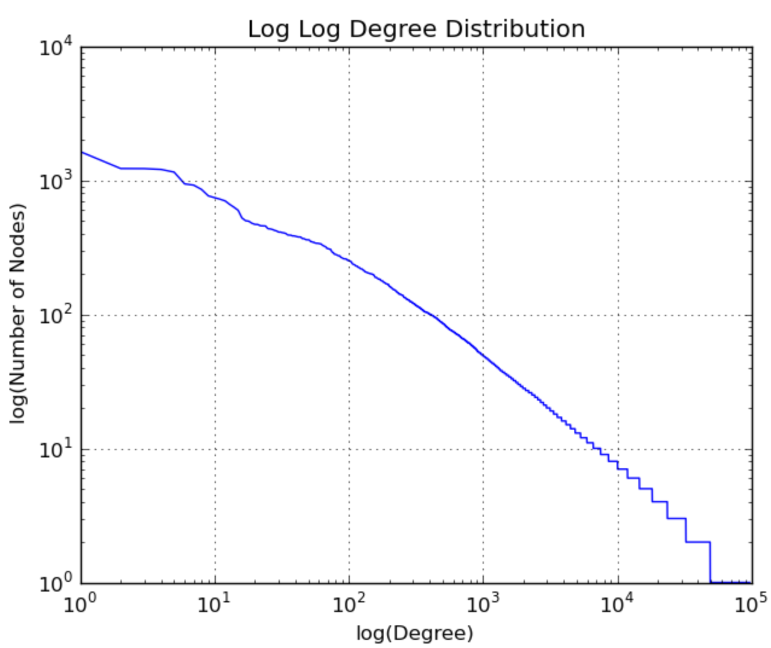
We used the w2w_outfit_review table for generating a network of the player interactions. Each row entry in this table corresponds to a review provided by one player for another player's outfit. A review can comprise of a rating (given in stars) and/or comments. For the purpose of graph generation we considered all interactions that have occurred till date. This resulted in a network of around 90000 nodes and nearly 1.25 million edges (confirm numbers)

Node Degree Distribution:

One of the key network structure attributes we analyzed was the degree distribution. Most social networks have a power law distribution in terms of the interactions between the entities of the network. This occurs primarily because they are always some nodes that are exponentially more active than the average node of the network. As anticipated, when we built an undirected graph of outfit reviews, we could observe the same power law distribution.

We also constructed a directed graph from the above mentioned table such that if player A reviewed player B's outfit there was a directed edge from node A to node B. We were able to observe an identical power law distribution for both the in-degrees as well as the out-degrees. Nodes with high in-degrees are indicative of those players who participate in several



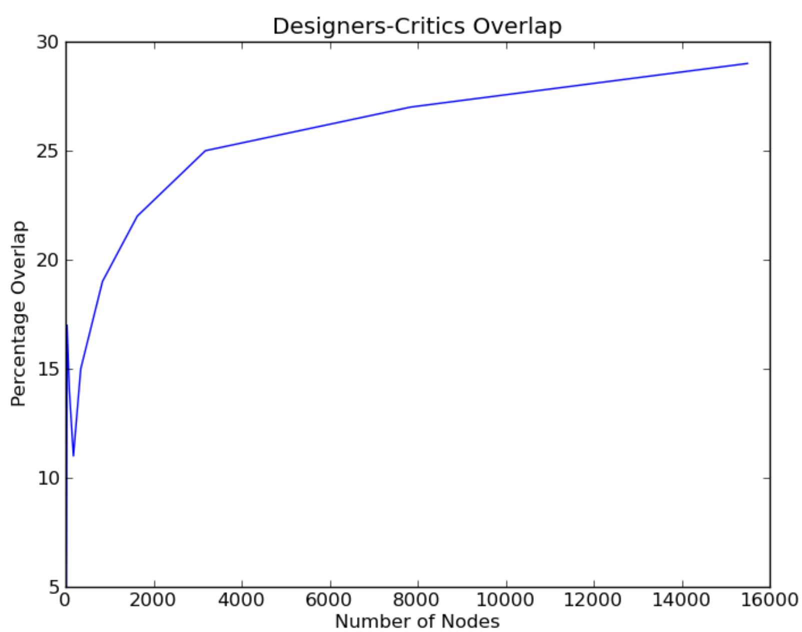


competitions or have popular avatars and hence receive several reviews. We classify such nodes as 'designers' since they mainly participate in the game by developing their outfits. Nodes with high out-degrees correspond to those players who spend a substantial amount of their time on the game reviewing other player's outfits. We classify such nodes as 'Critics'.

Designers are more conscious of the outfits they are making and hence are more likely to spend credits in order to acquire new things to decorate their avatars with. Critics on the other hand will be more inclined on sharing their opinions and hence can be implicitly used as advertisers for new items available in the 'gift shop'. However if might be possible that players are both 'Designers' as well as 'Critics'. Hence it is necessary to understand the extent of overlap between these two groups. In order to study this aspect of the network we created two separate lists of players, such that in one the players were sorted in descending order of in-degree and in the other the players were sorted in descending order of out-degree. Then for a given set of top k players in each list we determined the overlap percentage as follows:

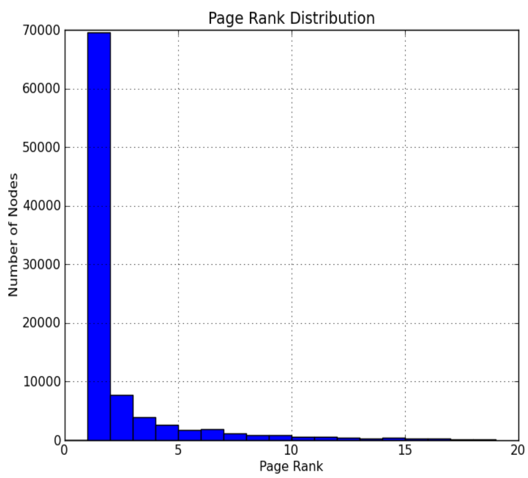
$$\text{overlap percentage} = \text{designersk} \cap \text{criticskdesignersk} \cup \text{criticsk} * 100$$

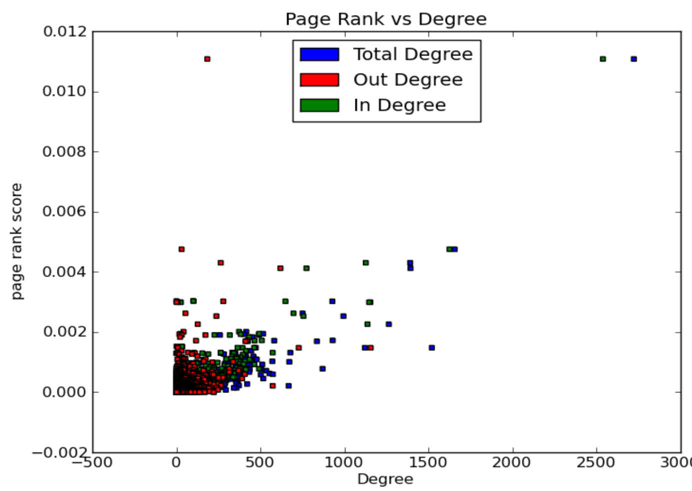
It was interesting to find the overlap percentage was low and the two groups are relatively distinct. Infact when we considered the top ten designers and the top 10 critics, the intersection of these two sets resulted in only one player. Moreover, given that the in-degree and out-degree distribution obey the power-law, suggests that a large portion of reviews are generated by a small number of players who are predominantly critics. Likewise a large number of outfits are generated by a small group of players who are predominantly designers. Hence for maximizing revenue we need to target the 'designers' but for maximizing virality we need to target the 'critics'.



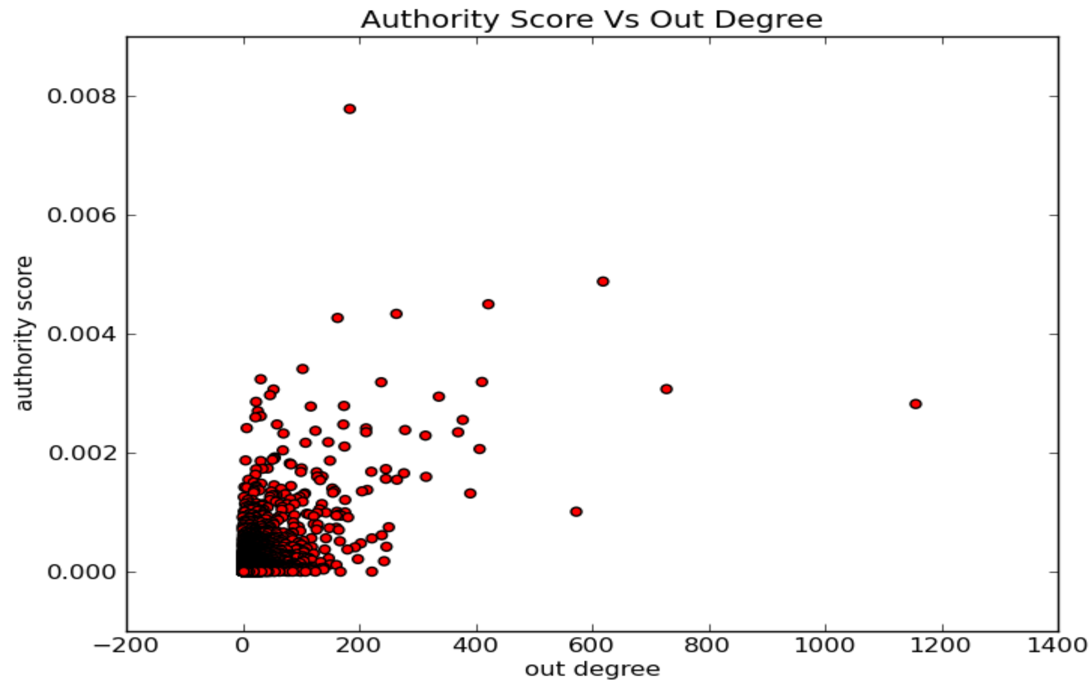
Page Rank & Hits

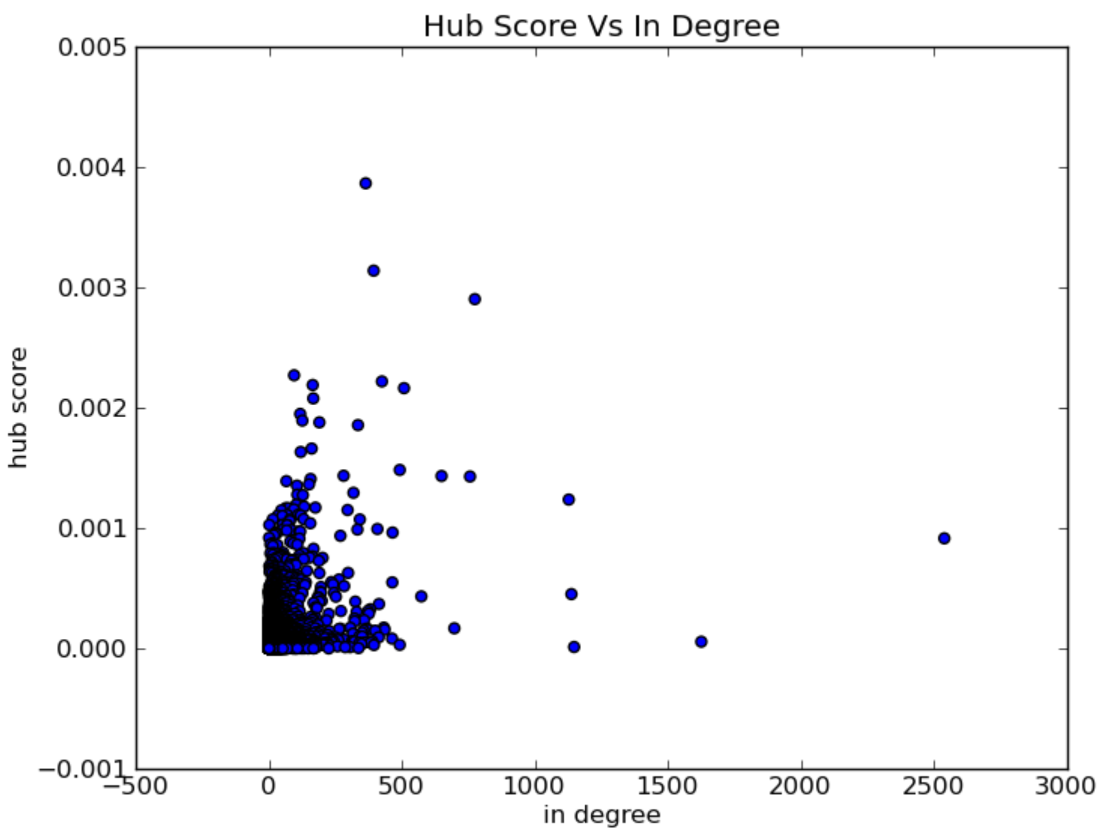
The Page Rank algorithm gives higher weights to pages that have more incoming links from other web pages that themselves have high page-ranks. This concept could easily be extended to this graph. By applying the page-rank algorithm we identify those players in the graphs that receive more reviews. Moreover, because the page-rank algorithm also considers the weight of the node itself, players receiving reviews from other players who themselves receive large number of reviews are automatically given higher page-ranks than players who get reviews from random players but have large number of them. Thus page-rank intrinsically gives a higher qualitative score to edges between players that actively participate in the game itself. The page rank scores of the players also follow a power-law distribution. Moreover as seen in the graph below, page-ranks of nodes correspond more closely with their in-degree rather than their out-degree. This validates the hypothesis that page-rank helps us identify those players whose outfits receive more reviews.





As established in the earlier discussion, the players of what to wear can be classified as 'designers' or 'critics'. This fact allows us to extend the Hits algorithm to the W2W network. Thus critics can now be considered as authorities since they review several outfits and hence have many outgoing edges while designers can be considered as hubs since their outfits are reviewed and hence they have many incoming edges. The hits algorithm up-weights those hubs that have incoming edges from better authorities and it also up-weights those authorities that have out going edges to better hubs. Thus the hub and authority scores are a more qualitative indicator of the 'designer' and 'critic' property respectively than just the in-degrees and out-degrees


Authority Score Vs Out Degree
authority score
out degree
0.008
0.006
0.004
0.002
0.000
−200
0
200
400
600
800
1000
1200
1400

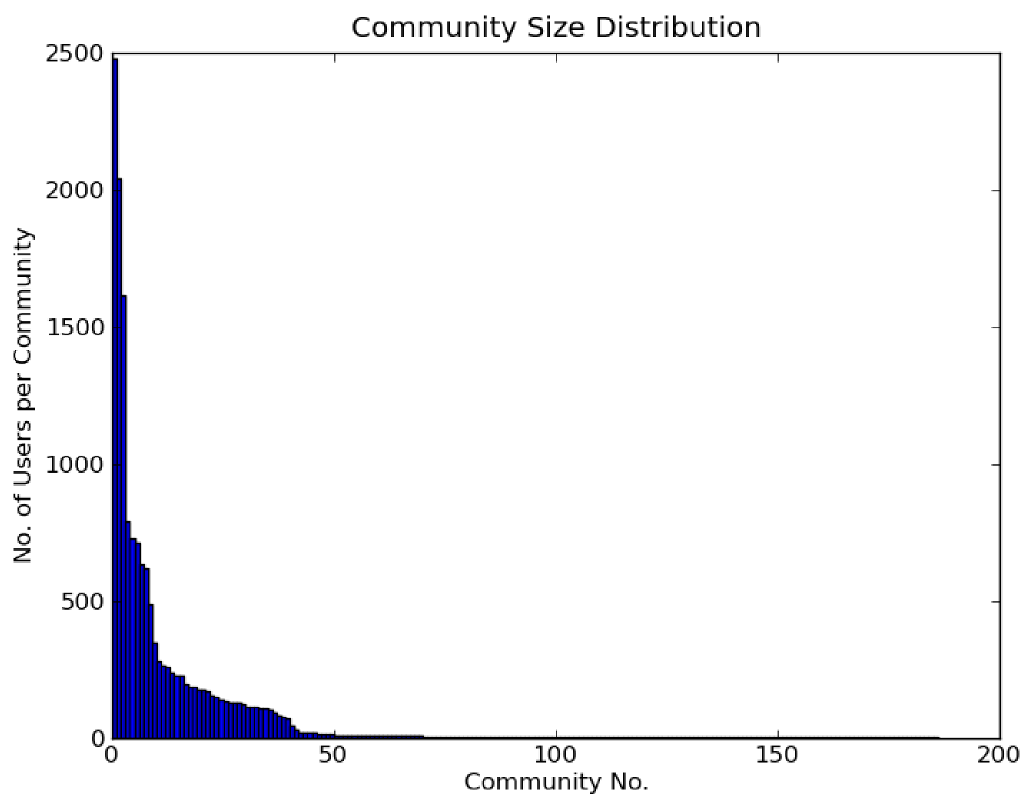

Hub Score Vs In Degree
hub score
in degree
0.005
0.004
0.003
0.002
0.001
0.000
−0.001
−500
0
500
1000
1500
2000
2500
3000

Community Structure

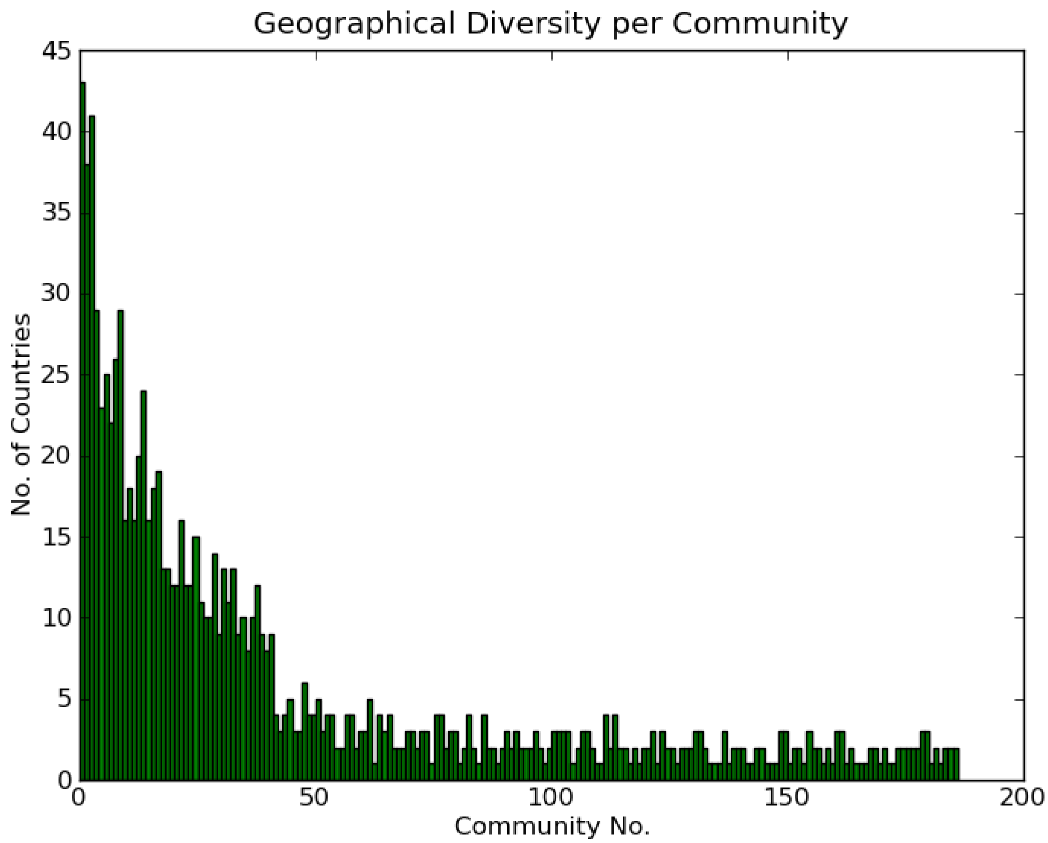
We implemented an algorithm for detecting community structure as described in Clauset, Newman and Moore's "Finding community structure in very large networks." The algorithm tries to gather nodes into groups so that there is a higher density of edges within groups than between them. It works by greedily attempting to optimize the modularity of the partitions of the graph, where modularity is a measure of how good a division of the network is as defined here, where k is the degree of the node, m is the number of edges in the graph, c is the community, and A is the adjacency matrix:

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w)$$

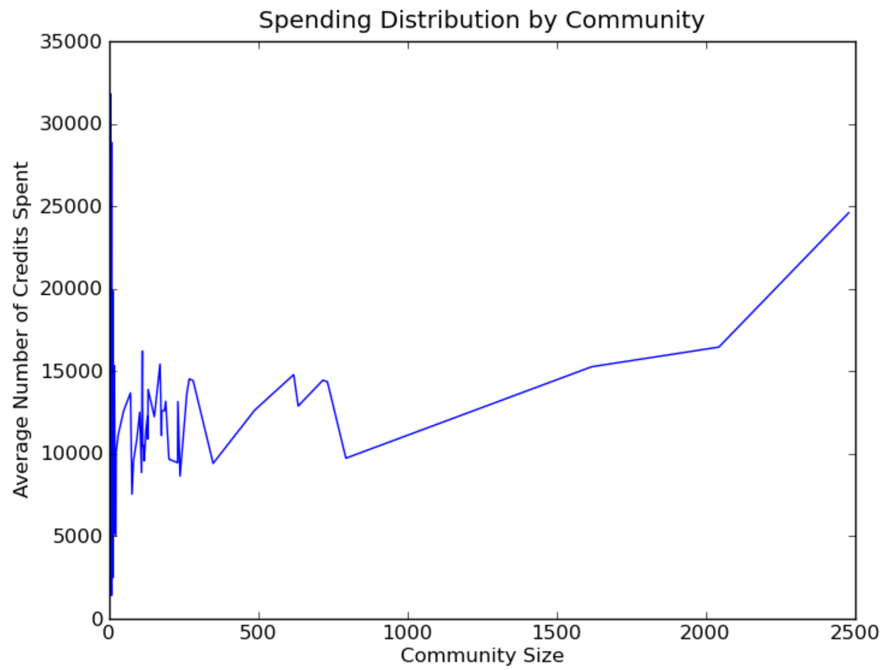
Higher values of Q, the modularity, indicate better divisions of the network into communities, but finding the division with the highest modularity is a hard problem. Therefore this algorithm uses a greedy optimization to provide a reasonable solution. The straightforward method involves starting each node in a single community and repeatedly merging communities that provide the highest increase (or slightest decrease) in Q. The division with the highest modularity is then used as our community partitions. After applying the algorithm to our network of active users over a period of three months, we divided the network into 186 communities with the following distribution:



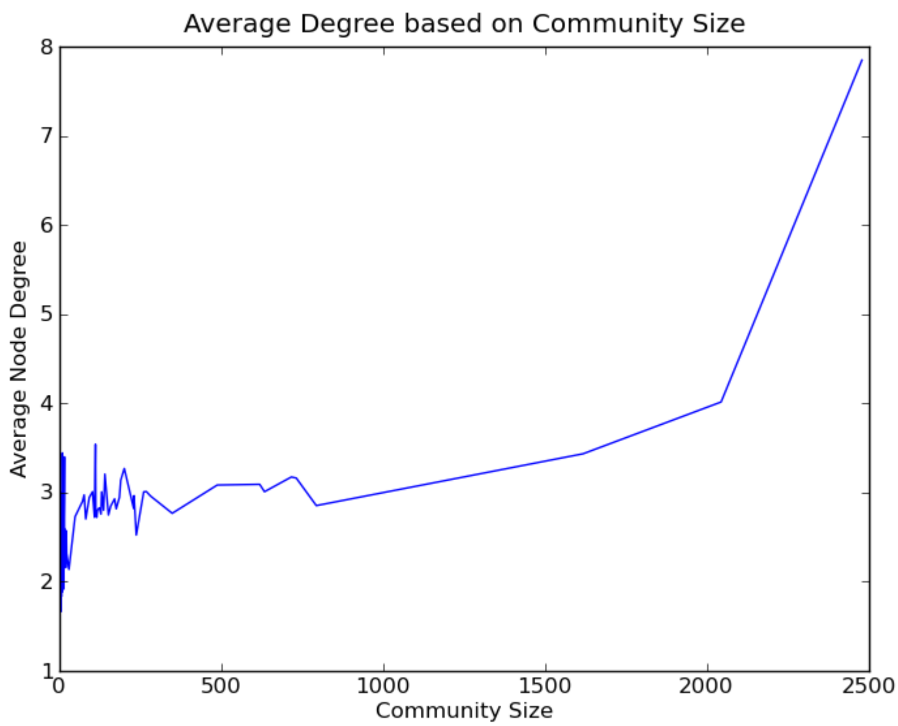
The distribution follows a power law distribution, which makes sense intuitively because preferential attachment results in this distribution since a new user is more likely to join a larger community than a smaller one. We can also see that this network has a strong community structure, unlike the network for a Facebook game described in Nazir, Raza, and Chuah's "Unveiling Facebook: A Measurement Study of Social Network Based Applications," which seems to indicate that different games will have different community structures. Furthermore, we measured the number of countries the users of each community.



We can see that there is a strong correlation between community size and geographical diversity; however, this also means that communities are not separated by country unless their size is very small. This indicates that communities consist of users in many diverse regions. In addition to geographical diversity, we examined other properties including the number of credits spent by users in each community.



In the game, credits are used to buy new outfits and items, and can be earned by rating other outfits or participating in contests, or by paying for them. We found that the average number of credits spent, which can be an indication of user participation and involvement, correlates with the size of the community. This may indicate that as users become part of a larger community, they interact more with each other, or vice versa. This seems to also be evidenced by the average degree of a node when compared with community size.



The degree distribution shows that users in larger communities tend to interact with other users more frequently than those in smaller ones. Users in smaller communities seem to interact with only a few friends, which in turn provides less incentive to participate further by spending more credits.

Conclusion/Future Work

In our analysis of the network structure of the What to Wear application, we examined several properties of the network, including node degree distribution, page rank scores, and community distribution. We were able to observe from the directed graph that active users tend to either spend time on creating outfits or reviewing others' outfits. The page rank analysis adds to this notion by correlating hubs and authorities with in and out degrees. By clustering nodes into communities, we saw that larger communities are geographically diverse and that users from larger communities tend to be more active and spend more credits. Unfortunately, much of the user statistics, including time played, credits purchased, and user activity was sparse or not stored at all, so we were not able to make any further analysis. Our work can be extended further by analyzing these properties as well as by building predictive models to determine how user activity and retention might be affected by certain events. Much of our analysis can be extended to other social applications to determine how users might be spending their time and how they behave.

References

A. Clauset, M. E. J. Newman, and C. Moore, Finding community structure in very large networks. Phys. Rev. E 70, 066111 (2004).

M. E. J. Newman, Fast algorithm for detecting com- munity structure in networks. Phys. Rev. E 69, 066133 (2004).