A Quick Dive Into the World of Celebrity Networks

Cassandra Chang
clchang01@email.wm.edu

Gergana Gospodinova
gmgospodinova@email.wm.edu

Russell Wolz
rwwolz@email.wm.edu

June 10, 2020

Abstract

Instagram is a popular image-sharing social media platform, hosting over one billion users each month. Due to this popularity, there are many celebrities with a vast number of followers, providing many possible networks to study. We set out to gain an understanding of the network of the top celebrities on Instagram and to determine which of these celebrities were the most well connected. We found that the clustering coefficient of the network is similar to that of other social media networks and that the in-degree distribution closely follows a power law distribution. To determine popularity, we considered the ego networks of the users with varying categories such as country and occupation. The application of the "Six Degrees of Kevin Bacon" game to the Kardashian family in this network showed that they are extremely well connected to other celebrities.

Keywords: *celebrity network; social network; Instagram; social circles; ego networks; analysis.*

1 Introduction and background

The word celebrity is defined as a "famous person" or "the state of being well- known." The definition of the word has not changed in the past few years, but the means of reaching this status have drastically changed. Being a celebrity was once an exclusive club reserved for big Hollywood actors, supermodels, and star athletes, but the inception of social media has increased the number of pathways to fame and made this initially exclusive club more accessible. There are now swarms of influencers with huge followings that are simply celebrities for being active on social media, promoting their hobbies or lifestyles to the rest of the world.

When first considering the Instagram social network, there were many questions about its structure and features. As another social media account, it could be similar to Twitter or Reddit networks, which have been studied as social networks in more detail(Myers, Sharma, Gupta, and Lin (2014), Olson and Neal (2015)). However, its focus on aesthetic images and the tendency for users to edit and perfect their posts differs from the more conversational and story-like posts on Twitter and Reddit. The users' need to curate their feed and profiles can lead to different characteristics in the ego networks that other social media platforms may not capture.

Social media also offers us a way to study the behavior of famous people. Celebrities must follow many social rules, carefully curating all of their moves. Who they interact with and who they have associated with plays a significant role in maintaining their fame and reputation. A great way to gain some insight into their world is by studying their ego networks on Instagram. Studying the ego networks of celebrities can show us if there are any other patterns related to who celebrities choose to follow. Our goal is to determine if celebrities tend to follow users from their own niche.

Ego networks can also reveal differences between celebrities from different countries. Since celebrities from the United States generally dominate the global market and set the standards for celebrities from other countries, we would expect them to be an authority within the celebrity community. We will also use a few measures of centrality to compare similarities and differences between networks as well as different metrics of the network, including where and how many hubs there are, connectedness, and clustering coefficient of the network.

Furthermore, we would like to recreate the Six Degrees of Kevin Bacon experiment, but with a twist. Instead of Kevin Bacon, we will use one of the most famous people on Instagram that owes her success to social media — Kylie Jenner. We would like to see if Kylie Jenner connects not just celebrities to each other, but also different niches to each other, widening the scope of the network and potentially making it more global.

2 Dataset

To answer these questions, we gathered data for the top 500 most-followed Instagram accounts, including the number of followers they have, a list of users they follow, a country they are based in, and their occupation, also referred to as their niche. We obtained the data by manually collecting usernames from Social Blade and location and occupation from Google. To get the number of followers and the list of users followed by each of the top Instagram users, we used an Instagram scraper written in Python (**Link**).

Based on who celebrities follow, we can build a graph $G = (V, E)$, where V is the set of Instagram accounts included in our research and the accounts that they follow, and E is the set of edges between them. G is a directed graph because of the follower-followee relationship on Instagram. There is a directed edge from one user to another if one follows the other.

To make the problem easier, celebrities known for multiple talents will only be categorized with the one they are most well-known for. Additionally, we exclude accounts labeled as a "company" in the analysis because the goal of this project is to examine individuals and their interactions rather than advertising efforts.

2.1 Basic Graph Characteristics

2.1.1 Graph Size

The graph itself has 187,331 nodes and 328,929 edges. These nodes are the celebrities selected by their popularity and their followers. The edges run between followers and followees regardless of status.

2.1.2 Degree Distribution

For this directed graph, both the in-degrees and the out-degrees are considered. The in-degrees correspond to the number of followers that a user has, and the out-degrees correspond to the number of accounts a user is following. Celebrity accounts will tend to have many followers and less following (i.e., high in-degree and low out-degree) while the opposite is true for spam accounts. In a sense, the in-degrees show how many other users are interested in a particular user. In contrast, the out-degrees show how many other users a specific user is interested in. Celebrities may tend to follow only each other, keeping a tight inner circle of similar interests and social groups.

Similarly, parody and humor accounts may tend to follow each other for content ideas. It is the average user or spam account whose interests are broader that could connect these circles. An average user could have interests in certain celebrities, as well as influencers in different topics, such as cooking, fitness, or pets. A spam user may randomly follow as many accounts as it can in efforts to scam as many people as possible.

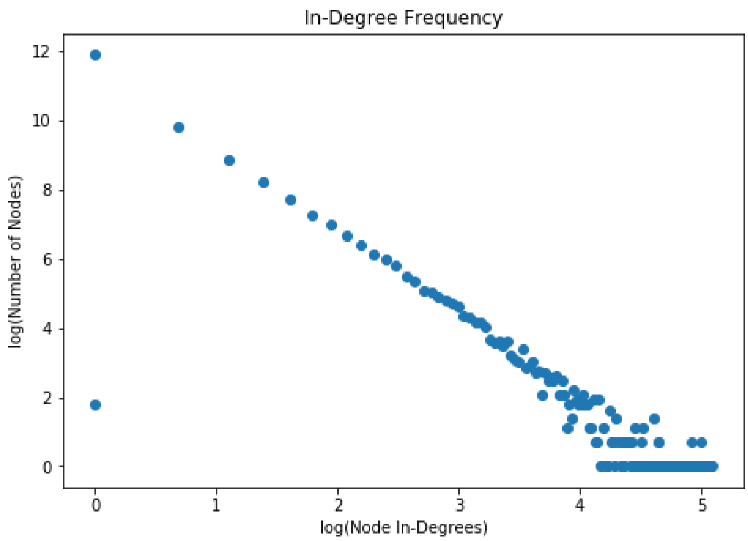

Figure 1: The in-degree distribution for the Instagram network. The axes are on a logarithmic scale. The points tend to fall linearly, pointing towards a power law distribution.

Figure 1 shows that the in-degree distribution graph generally resembles an empirical power-law distribution. There are generally more nodes with low in-degrees and few nodes with high in-degrees. Even among the most famous celebrities, there is a resemblance to the degree distribution of the small-world network. There are fewer nodes with no followers since an account typically will get at least one follow back from an account that they follow, and there are some accounts who are just followers of the celebrities, without all of their followers and following data included.

The out-degree distribution is also quite interesting and harder to pinpoint. As seen in Figure 2, it appears that most of the nodes have few to no out-degrees, with one or two nodes at each of the out-degree points. This reflects that there are a lot of users who don't follow many people, which is typical of celebrities. The ratio of followers to following is extremely high, due to the disproportionately large number of followers from fame.

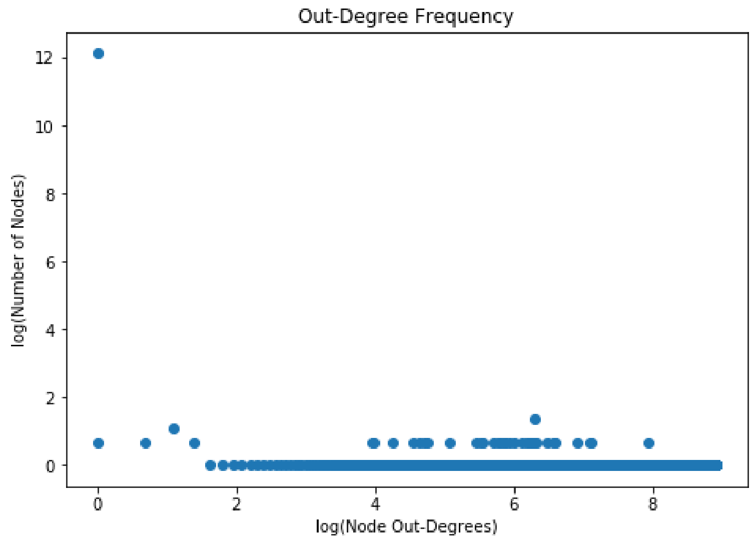

Figure 2: The out-degree distribution for the Instagram network. The axes are on a logarithmic scale. These points tend to have low numbers of nodes at each degree level, with an outlier at 0 degrees.

2.1.3 Clustering Analysis

Clustering in directed graphs is generally a difficult problem — in an undirected graph, all of the edges are treated similarly. In a directed graph, not all edges are equal. For example, stating that nodes A and B are connected by an edge is not the same as saying that node A leads to node B. In their paper on community detection in directed networks, Malliaros and Vazirgiannis discuss various methods for determining clustering in directed graphs. Their suggestions were based on transforming the directed graph into an undirected form and then manipulating them to represent the directionality of the edges (Malliaros and Vazirgiannis (2013)). One of the methods, introduced by Wang, Lou, Tang, and Hopcroft, used a greedy algorithm to determine clusters. The algorithm incrementally grew the cluster or subset by adding the next node with the most connections to the nodes in the subset (Wang, Lou, Tang, and Hopcroft (2011)). A re-creation of the greedy algorithm was attempted but required more processing power than was available. Due to this, the naive approach of treating the graph as an undirected graph was used. Although this loses some of the information on how users are connected, it may still reveal some base connections between Instagram users.

In the undirected Instagram network, there were two clusters found. These clusters were extremely different in size, with one cluster containing only eight nodes. All of the accounts in this cluster appeared to be some sort of advertising in Arabic, targeted at Middle Eastern countries. Upon further inspection, it was found that one particular user, a preacher from Egypt, was the connection to the other nodes. This node had seven out-degrees, and each of the other seven nodes had one in-degree. This preacher was the core of this cluster, and likely the connection from this niche set of accounts to the larger world-wide Instagram network. All of the other nodes were part of the larger cluster. Neither of these clusters were strongly connected, although this may also be due to the limitations of the dataset.

The clustering coefficient of the undirected Instagram network is approximately 0.14. Overall, the subset of the Instagram network that is included in our data set tends to be fairly connected, though it does not satisfy the accepted definition of being strongly connected. This value is comparable to the clustering coefficient of the Facebook social network, which is also undirected (Ugander, Karrer, Backstrom, & Marlow 2011). Since Facebook owns Instagram, and many people are on both social media platforms, it makes sense that the networks would cluster similarly.

2.2 Ego Networks

In addition to this base Instagram network, we can create ego networks where V is the set of Instagram accounts that include the ego and their alters, and E is the set of edges between them.

In Figure 4 we can see that the top countries for our celebrity network are USA, Brazil, and Indonesia.

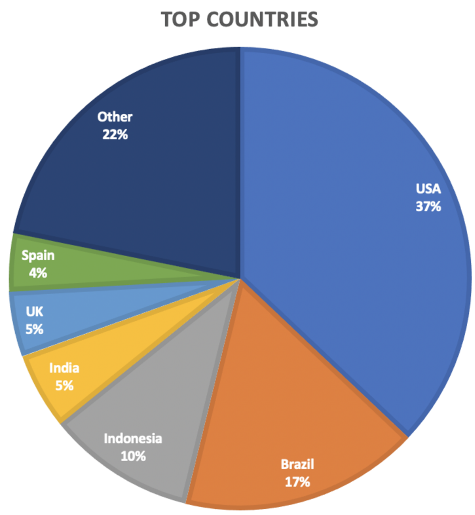

Figure 3: Top countries represented by the top Instagram users.

We looked at what each country is famous for, or what are the top categories of celebrities for each country. Results for some of the top countries can be seen in Figure 4.

For many countries including the US, Brazil, and Indonesia, actors and singers make up a majority of the Instagram stars. However, in Spain, the soccer stars are much more prevalent than the music and film stars. Other countries with a high prevalence of star soccer athletes include France and the UK.

In addition to these more conventional famous occupations, YouTube stars are increasingly popularity, as they fall into the top categories for some countries such as the US and Brazil, two of the top most represented countries in the network.

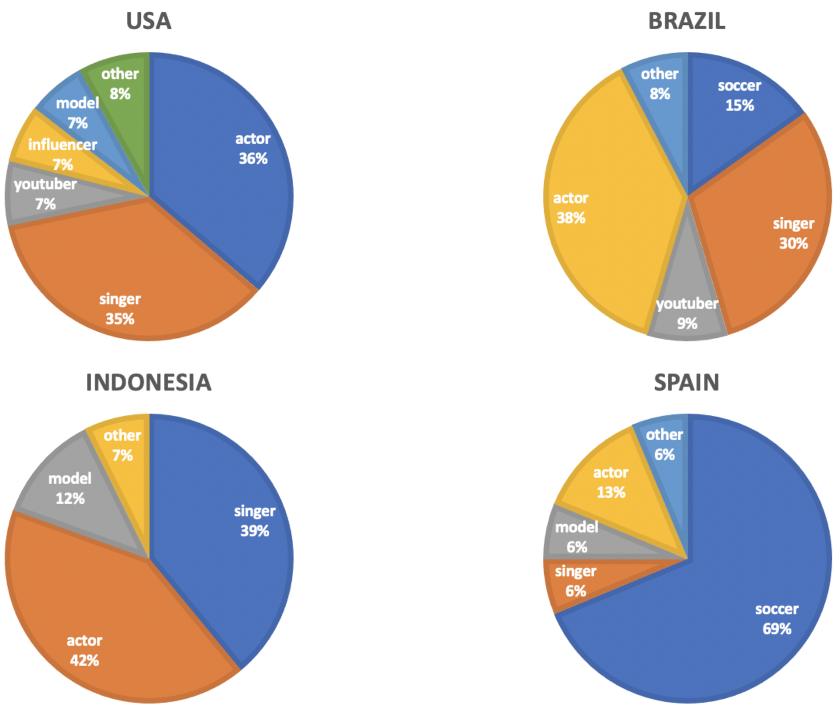

Figure 4: Top categories for top countries

3 Results

3.1 Six degrees of Kylie Jenner

Back in the 1990s, Kevin Bacon was at the peak of his career. Many would say that "he had worked with everybody in Hollywood or someone who's worked with them." A group of four students invented the game"Six Degrees of Kevin Bacon" after watching one of his movies (*The Oracle of Bacon* (n.d.)). They went as far as contacting a famous TV host and claiming that Kevin Bacon is the center of the entertainment universe. They made the game popular after appearing on a couple of famous talk show.

Our idea was to recreate this game with a big celebrity of our generation that acts as the center of the social media universe —- Kylie Jenner. Kylie is known for being an influencer, a reality TV star, and the owner of her own line of cosmetics. Other than her appearance on the reality show about her family, *Keeping Up with the Kardashians*, she is not known for acting, so the links between celebrities must be something other than movies as it was in the original game. Instead, we recreate the game by finding the shortest path between Kylie and all other celebrities on our list. After doing that, we find that instead of calling this game the "Six degrees of Kylie Jenner", we should call it "The 1.66 degrees of Kylie Jenner" (Figure 5), as 1.66 is the average degree of separation between Kylie Jenner and all other celebrities. This means that most other celebrities either follow Kylie and are directly connected to her, or that they are connected to someone who is connected to her. Only a few celebrities are three or more degrees from her.

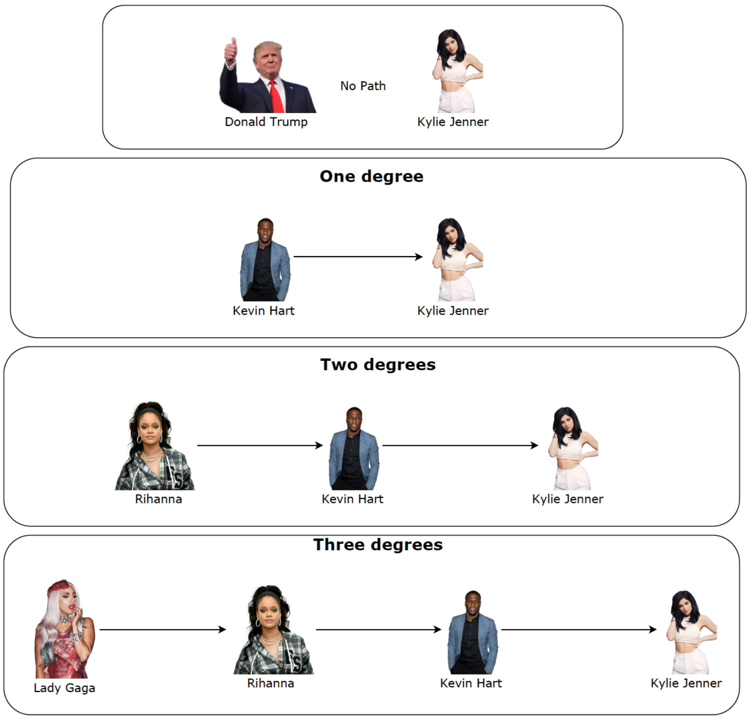

Figure 5: The 1.66 degrees of Kylie Jenner

There is an ongoing rivalry between the members of the Kardashian/Jenner family, as they are all trying to become more famous than the others. We attempted to quantify this by comparing the average degrees of separation for each of the Kardashian and Jenner girls, with the hope that this might indicate which one is the most famous and end their rivalry once and for all.

After computing the separation degrees, we find that Kim Kardashian is slightly better connected to everyone else with an average degree of separation of 1.65, as compared to 1.66 for Kylie(Figure 6). Even though this difference is quite small, we can still say that Kim Kardashian is more famous than Kylie Jenner. Kylie is then followed by her older sisters Kendall and Khloe, with degrees of separation 1.69 and 1.93, respectively. Not surprisingly, the least connected Kardashian is their mom, Kris. This makes intuitive sense, as Kris only had 33.4 million followers at the time of the experiments while Kim, Kylie, Kendall, and Khloe each had more than 100 million followers. Interestingly, Kim has the lowest degree of separation and the highest number of followers (168.6 million), and this pattern continues exactly all the way to Kris, who has the highest degree of separation and the lowest number of followers.

After finding that our assumption about Kylie being the center of the social media world was proven wrong, we wanted to conduct the same experiment for the whole celebrity network and discover who the true center of the Instagram world is. It turns out that Rihanna is the main celebrity in the network with an 1.57 average degree of separation. This was an interesting discovery, and prompted us to continue our research. The next question we looked to answer is which niche was the most well-connected.

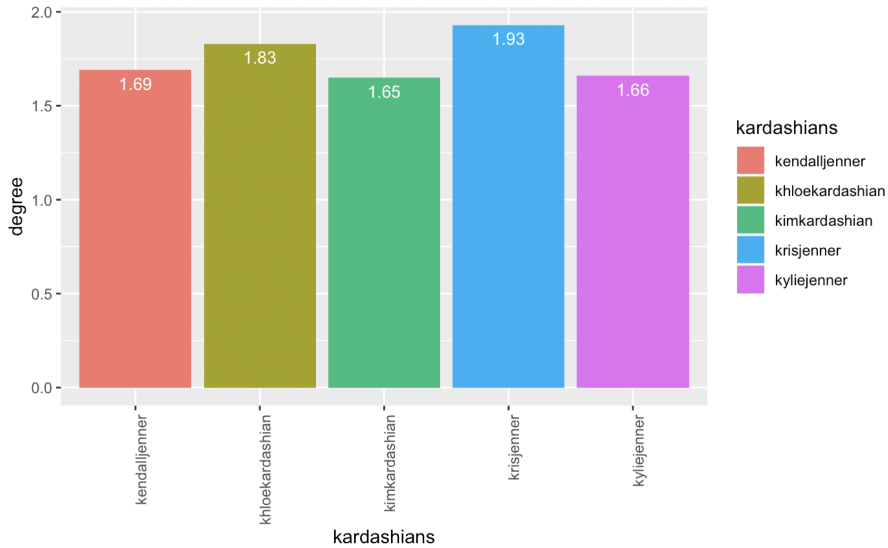

Figure 6: Kardashian/Jenner family degrees of separation

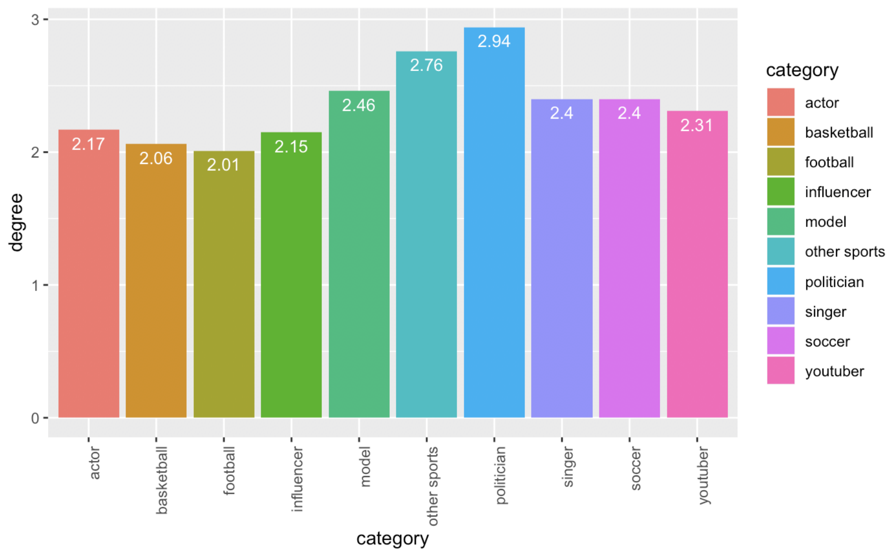

Figure 7: Degrees of separation per celebrity category

The category with the lowest degree of separation is football(Figure 7). Since a significant percentage of the celebrities on our list are from the US, it is not surprising that football players are very well connected. They are constantly interacting with each other and generally stay on good terms regardless of the teams that they have played on. The least well-connected stars are politicians — not so surprising result since some of the politicians on the list are not from the US, so not as many of the celebrities are likely to be directly connected to them. Another possible reason is that celebrities might not want to directly associate with politicians to maintain a particular social image, either because they do not want to share their political views, or because they do not support the politicians.

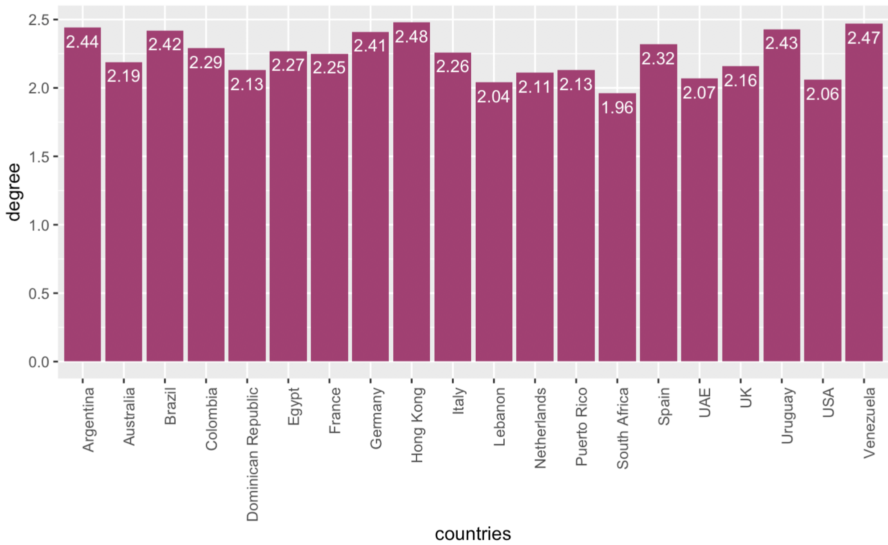

Figure 8: Degrees of separation per country(below average)

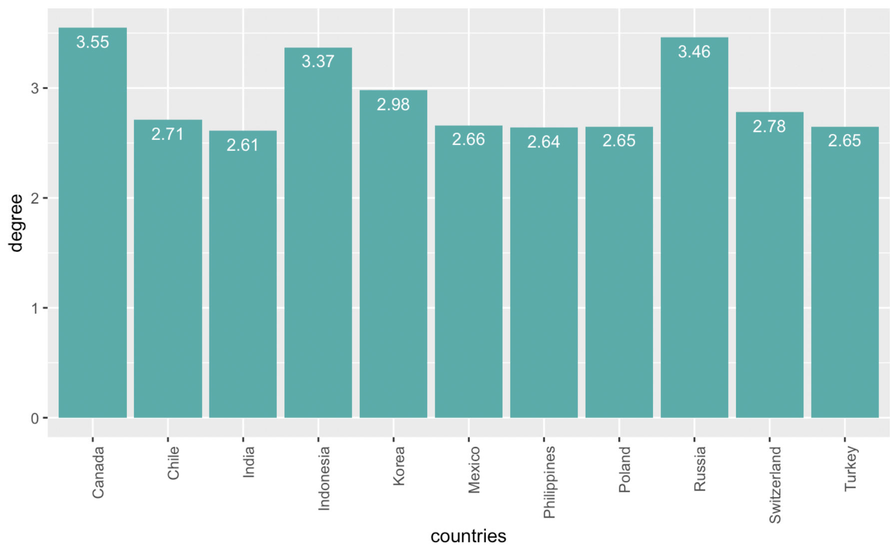

Figure 9: Degrees of separation per country(above average)

We also looked at how well-connected celebrities from different countries are. The average degrees of separation for all countries is 2.48. To make the visualization of the results for each country easier, we have divided the countries into two groups: countries that have an average degree of separation below the average(Figure 8) and above the average(Figure 9).

The top three most well-connected countries are South Africa, the USA, and Lebanon(Figure 8).

The results for South Africa and Lebanon can be explained by the fact that there are only a few celebrities from those two countries in the list of most popular accounts, so they have a few well-connected stars. The US, on the other, has many well-connected celebrities. Even though there are many instances of American celebrities on our list which would have a higher opportunity for the degree to go up, the US was still one of the best connected.

On the other hand, the three least well-connected countries are Canada, Russia, and Indonesia(Figure 9). The high degree for Canada and Russia can be explained by the fact that there are few celebrities from these countries on our list. Indonesia, on the other hand, has many stars who do not get as much influence from the US. Indonesian celebrities are well-connected to each other but not to other celebrities.

3.2 Social Circles in Ego Networks

When considering the ego networks from the Instagram data and, more specifically, social circles in the ego networks, there were a few limitations on what exactly we could achieve. Since we do not have any information about the actual interactions between the celebrities in our network, we cannot build a model that identifies social circles. However, we can still build a model using the data that we do have.

Since we have already identified the celebrities' niches and their country of origin, we can think of each niche or each country as one social circle. Doing this can help us answer two of our questions: if celebrities tend to follow other stars in their niche and if celebrities tend to follow celebrities from the same country as them.

3.2.1 Categories

We first considered the different niches of celebrities. We wanted to identify what kind of accounts each group tends to follow. In other words, we attempted to identify the social circles in the ego networks of celebrities in the same niche.

The results for some celebrity niches can be seen in Figure 10. We included six niches, and for each niche, we looked at the top three categories followed by the niche.

We first noticed that celebrities from every category follow singers and actors. This result is not very surprising since some of the most influential people in the world are, in fact, singers and actors. We can also see that models follow actors and singers closely as the third most influential category.

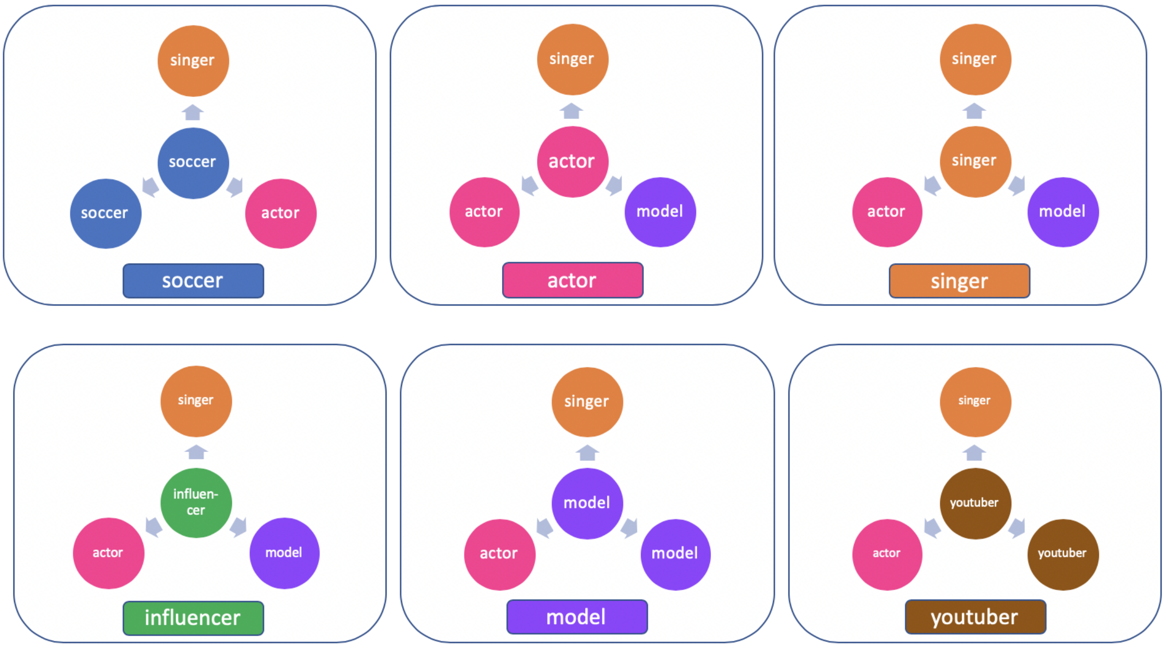

Figure 10: Ego network analysis - categories top 3

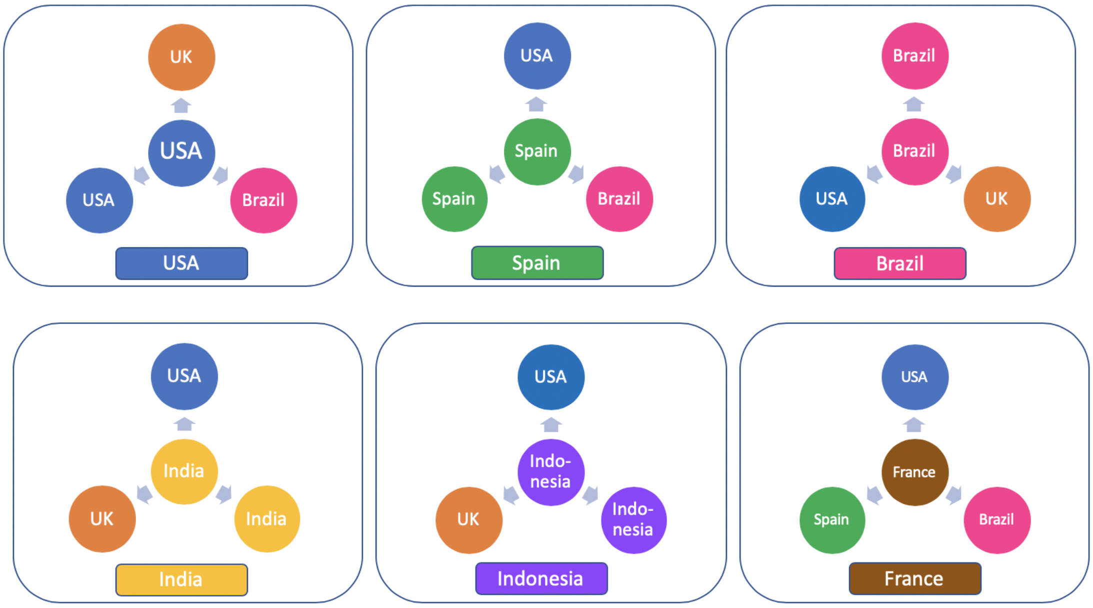

Figure 11: Ego network analysis - countries top 3

3.2.2 Countries

Next, we compare celebrities from different countries. We wanted to determine if celebrities follow other stars from the same country and if there is a country that dominates the global celebrity market and if that country is the US.

The results for some of the top countries can be seen in Figure 11. We included six nations, and for each country, we looked at the top three countries that each country follows.

We can see that it is common for most countries to follow a lot of celebrities from the same countries, but that is not always the case. French stars, for example, follow mostly the US, Brazil, and Spain-based celebrities. This result can be explained by the fact that most French celebrities are soccer players, and soccer players, as we saw from Figure 10, tend to follow other soccer players, singers, and actors. Spain and Brazil have many good soccer players, and the US has many famous singers and actors.

The US appears in the top three for every country, confirming that there is a country that dominates the global celebrity market and that it is, in fact, the US. We know for a fact that some of the most well-known celebrities worldwide come from the US. In some countries, US celebrities dominate more than in others. For example, one of the few countries where celebrities mostly follow other stars from the same country and not from the US is Indonesia. Other than Indonesia, we noticed the same thing for two other countries not included in Figure 11 - Switzerland and Egypt.

4 Conclusions and Future Directions

We looked at the network of top 500 Instagram stars in an attempt to learn about the differences based on users, countries, and occupations. We also compared the networks of the Kardashians in an attempt to recreate the Six Degrees of Kevin Bacon and determine who was the most well connected in the family. We found that the in-degrees of our Instagram network closely followed a power law distribution, while the out-degree distribution did not. This is likely due to the imbalanced nature of the celebrities we chose, who are followed by millions of users and follow far less. We found that the clustering coefficient of the Instagram network was .14, which is similar to the coefficient found for the Facebook social network. We could also notice differences between the categories of stars from different countries.

When comparing the Kardashians, we found that Kim Kardashian was the most well connected with 1.65 degrees of separation on average, just beating out Kylie Jenner, while Kris Jenner had the highest degree in the family. We also compared the degrees of separation by country and celebrity niche, finding that South Africans and football stars were the most well connected.

There are many possible next steps for this project. First, we would like to figure out a clean way to display and visualize our network, as it is too large to simply plot and visually seeing the network could help identify some patterns we may otherwise miss. Furthermore, We would like to compare a larger portion of the Instagram network to see if our results hold for a larger part of the platform's network. This could be done by taking random samples of nodes, finding the results for that network, and repeating the process: this could help provide more thorough statistics for the network analysis. Another direction would be to compare the results from Instagram to those of other social media networks, such as Twitter or Facebook. This could help understanding of the similarities and differences in the types of networks these platforms help create. Additionally, looking at the ego networks for the top percent of users per country could be a useful way to compare and contrast the use of Instagram in different countries.

References

Malliaros, F. D., & Vazirgiannis, M. (2013). Clustering and community detection in directed networks: A survey. *CoRR*, *abs/1308.0971*. Retrieved from http://arxiv.org/abs/1308.0971

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd international conference on world wide web* (p. 493–498). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2567948.2576939 doi: 10.1145/2567948.2576939

Olson, R. S., & Neal, Z. P. (2015). Navigating the massive world of reddit: using backbone networks to map user interests in social media. *PeerJ Computer Science*, *1*(4). doi: https://doi.org/10.7717/peerj-cs.4

The oracle of bacon. (n.d.). Retrieved from http://oracleofbacon.org/

Wang, L., Lou, T., Tang, J., & Hopcroft, J. E. (2011). Detecting community kernels in large social networks. In *2011 ieee 11th international conference on data mining* (p. 784-793).