Experimental Analysis of Networks in the Age of Social Media

EAS 499 Senior Capstone Project and Thesis
April 29, 2020

Written by Katherine Li

Thesis Advisor: Aaron Roth

University of Pennsylvania
School of Engineering and Applied Science

Table of Contents

1. Abstract

Social networks exist across different scopes and can be used to explain trends in human behavior. This project seeks to analyze how recent developments in social media have changed social networks and subsequently changed human behavior. The first topic this paper seeks to cover is the small world phenomenon. While the small world phenomenon has been studied in the past, the results do not consider the wealth of social media platforms that are present today. To analyze how the small world phenomenon exists in today's society, we ran a small world experiment on Instagram and analyzed the results. To better contextualize this small world phenomena on Instagram as well as the landscape of social media in general, we further analyze Instagram's network as well as competitors such as LinkedIn and Twitter using graph theory concepts. Through our analysis and research, we will see that each platform's graph structure and user interaction greatly influence one another.

2. Prior Research on the Small World Phenomenon

One of the most useful properties of social networks is the power of communication. With every change in social media and technology, each person's individual network continues to expand. Especially given how widespread social media is today, we can look at how the shortest path between any two strangers has changed. However, in order to fully analyze the impact of social media on communication networks, we need to first understand what this network looked like before the introduction of social media.

2.1 The First Small World Experiment

Research on what communication looks like in social networks dates back to ***Milgram's famous small world experiment*** in 1969 [1]. The goal of the experiment was to determine the minimum number of intermediaries required to link two randomly selected individuals [1].

In his experiment, Milgram solicited 296 starting individuals and one target individual who was a stockbroker in Boston, Massachusetts [1]. The starting individuals were also split into three subpopulations based on their similarities with the target individual [1]. To assess whether geographical distance would have an impact on the chains formed, two subpopulations consisted of individuals living in Nebraska, while one group consisted of individuals living in Boston where the target individual also lived [1]. In addition to geography, Milgram also wanted to account for the factor of professional circles [1]. Of the two Nebraska subpopulations, one consisted of bluechip stockholders who probably had access to the investment business while the other group was randomly selected [1].

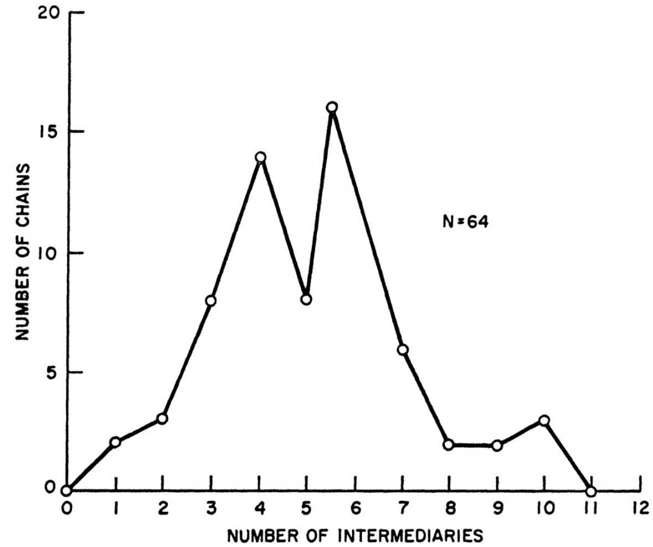

FIGURE 1

Lengths of Completed Chains

Figure 1. Distribution of the number of intermediaries in Milgram's experiment [1].

The starting individuals were asked to physically mail a letter to whomever they believed could bring the letter closer to the target individual in Boston [1]. The target stockbroker received 64 letters from successful chains (22% success rate), with the mean number of intermediaries being 5.2 people [1]. The median number of intermediaries was 6 people, which has become more commonly known as the ***"six degrees of separation"*** [2]. Looking deeper into the the subpopulations, Milgram discovered that geography did influence the distance of the paths [1]. The average number of intermediaries from the Boston subpopulation was 4.4 intermediaries, while the average of the Nebraska chains was 5.7 intermediaries [1]. Milgram stated the existence of a "small world" would suggest that "social networks are in some sense tightly woven, full of unexpected strands linking individuals seemingly far removed from one another in physical or social space" [1] and that is exactly what his experiments showed.

This experiment was monumental in shaping social network research due to the key conclusions drawn. The first major contribution of this experiment was the upper bound it set on the minimum number of intermediaries needed to connect any two people in the United States [1]. In this experiment, individuals were limited in the number of people they could send the document to [1]. However, in reality if someone was trying to reach an unknown individual, they would likely contact more than one person. Thus, six intermediaries serves as an upper bound to the number of intermediaries it should take to reach a target individual [1]. The second key conclusion from this study was the pattern of convergence uncovered by the experiment [1]. Milgram suggested that "the convergence of communication chains through common individuals is an important feature of small world nets, and it should be accounted for theoretically" [1].

2.2 Internet Based Search of the Small World

Milgram suggested that more work needed to be done in this field [1], and in 2003 researchers Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts answered the call [3]. From Milgram's experiment in 1969 to 2003, the world was transformed by the creation of the Internet. Although Milgram's experiment served as substantial evidence that the United States exhibited small world properties, not much documentation existed on whether the global population exhibited the same properties [3]. Dodds, Muhamad, and Watts conducted a ***global search experiment*** which asked over 60,000 email users to attempt to reach one of 18 target people in 13 countries by ***forwarding emails*** to acquaintances [3]. Their results showed that the world also exhibited small world properties as most searches were completed in up to seven steps [3].

In addition to looking at the number of intermediaries, they also asked participants to provide what their relationship was with the next person they forwarded the message to, as well as the "strength" of their relationship with the next person [3]. In *Figure 2* below, we can see a distribution of the types of relationships as well as the strengths of relationships that participants recorded [3]. We can see that friends were the most popular relationship, particularly those developed at work, and most people reported relationships that were "casual" to "very close" [3].

Table 1. Type, origin, and strength of social ties used to direct messages. Only the top five categories in the first two columns have been listed. The most useful category of social tie is medium-strength friendships that originate in the workplace.

Type of relationship	%	Origin of relationship	%	Strength of relationship	%
Friend	67	Work	25	Extremely close	18
Relatives	10	School/university	22	Very close	23
Co-worker	9	Family/relation	19	Fairly close	33
Sibling	5	Mutual friend	9	Casual	22
Significant other	3	Internet	6	Not close	4

Figure 2. Relationship types and strengths of ties between intermediaries [3].

However, the most interesting result to note is the distribution of relationship strengths among successful versus unsuccessful chains. It was discovered that in successful chains, "casual" relationships were chosen 15.7% more frequently than in unsuccessful chains [3]. In successful chains, "not close" relationships were also chosen 5.7% more often than in unsuccessful chains [3], supporting the claim that ***weak ties are disproportionately responsible for social connectivity*** [2].

2.3 The Strength of Weak Ties

The conclusions that Dodds, Muhamad, and Watts drew on the importance of weak ties is an important aspect of small world networks. This importance of weak ties in social contagion was studied in detail by researcher Mark S. Granovetter. Given two individuals *A* and *B*, and a set *S* of people with ties to both *A* and *B*, Granovetter claimed, “the stronger the tie between *A* and *B*, the larger the proportion of individuals in *S* to whom they will both be tied” [4]. The more important revelation is that a triad in which *A* and *B* have a strong tie, *A* and *C* have a strong tie, and *B* and *C* have no tie is extremely unlikely to occur [4].

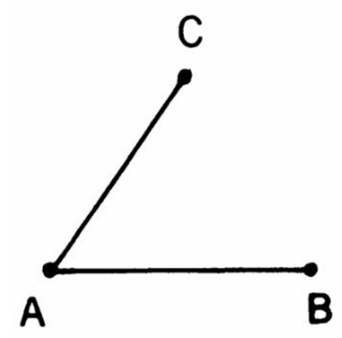

FIG. 1.—Forbidden triad

Figure 3. Example of the combination of strong ties and absent ties that cannot exist [4].

The importance of this triad can be seen through the analogy of a bridge, where we define a “bridge” as a connection in a network that serves as the only path between two points *A* and *B* [4]. Then this bridge is also the only path along which information can move from anyone connected (both directly and indirectly) to *A*, to anyone connected to *B* [4]. Thus, in terms of sending a message across a network, this bridge is very important in moving information to new contacts [4]. Such bridges must always be weak ties and not strong ties [4]. If we assume by contradiction that some bridge *A*-*B* is a strong tie, and *A* also has a strong tie to *C*, then we would expect a tie between *B* and *C* as well [4]. This would contradict the fact that *A*-*B* is a bridge, and thus this cannot be the case [4].

2.4 Facebook Connects the World in New Ways

With the introduction of the Internet also came the introduction of new websites focused on sharing digital media. One of the most notable developments was the introduction of Facebook in 2004 [5]. With a means for messaging and sharing with friends, Facebook has gained mass popularity since its founding. By September 2012, there were over 1 billion people using Facebook, and Mark Zuckerberg (Facebook CEO) was named “Person of the Year” by Time Magazine in 2010 [5].

With Facebook's meteoric rise, researchers became interested in the implications of digital messaging on the small world phenomenon. In 2012, researchers from Cornell, the University of Milan, and Facebook studied Facebook's graph to analyze whether the average distance between any two people had shrunk from 6 intermediaries [6]. Unlike Milgram's physical routing experiment, these researchers used algorithms to analyze Facebook's graph of approximately 721 million users at the time [7].

The researchers used a ***neighborhood function $N_G(t)$*** of a graph *G* that returns the number of pairs of nodes *(x,y)* such that node *y* is reachable from person *x* in at most *t* steps [7]. Using a diffusion based algorithm called ***HyperANF***, the researchers were able to approximate the neighborhood function for large graphs [7]. Some factors that the researchers wanted to consider were region and time [7]. Thus, they used HyperANF runs over the course of six years (2007-2012) on the global Facebook graph, as well as graphs within certain countries to analyze whether physical distance could have an impact [7]. The results of these runs revealed the following properties.

Average Degree and Density

As time went on, the degree of each node grew very quickly [7]. However, ***average density*** of the network ***decreased*** as time went on [7]. The table in *Figure 4* below shows a summary of the average degree of each node growing over time [7]. As shown in the plot of graph density to number of users in *Figure 5* below, the largest graph (the entire FB network) has the lowest density, and all the graphs show a downwards trend over time [7]. Thus, while the number of users has grown, each additional user appears to have less friends in their network.

	it	se	itse	us	fb
2007	1.31	3.90	1.50	119.61	99.50
2008	5.88	46.09	36.00	106.05	76.15
2009	50.82	69.60	55.91	111.78	88.68
2010	122.92	100.85	118.54	128.95	113.00
2011	198.20	140.55	187.48	188.30	169.03
current	226.03	154.54	213.30	213.76	190.44

Table 4: Average degree of the datasets.

Figure 4. Average degree of each node from 2007-2012 across all graphs [7].

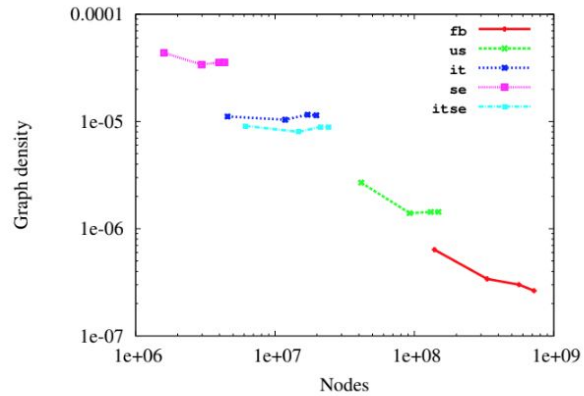

Figure 5. Density of graph in relation to number of users from 2009-2011[7].

Average Distance

At the crux of the small world phenomenon, we are interested in the average distance between any two users. Perhaps the biggest takeaway was that the ***average distance*** between any two users in the entire Facebook network was **4.74** [7]. Thus, the 2012 world of Facebook communication has made Milgram's small world even smaller. In *Figure 6*, we can see a visualization of the decrease in the average distance across each graph and subgraph analyzed from 2007 to 2012.

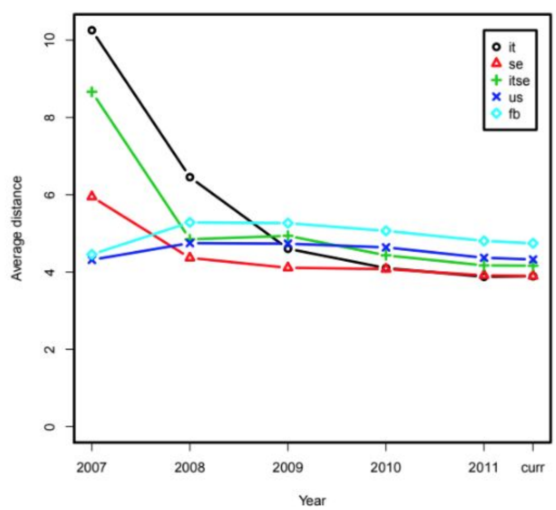

Figure 6. Average distance from 2007-2012 visualized in graph form [7].

	it	se	itse	us	fb
2007	10.25 (±0.17)	5.95 (±0.07)	8.66 (±0.14)	4.32 (±0.02)	4.46 (±0.04)
2008	6.45 (±0.03)	4.37 (±0.03)	4.85 (±0.05)	4.75 (±0.02)	5.28 (±0.03)
2009	4.60 (±0.02)	4.11 (±0.01)	4.94 (±0.02)	4.73 (±0.02)	5.26 (±0.03)
2010	4.10 (±0.02)	4.08 (±0.02)	4.43 (±0.03)	4.64 (±0.02)	5.06 (±0.01)
2011	3.88 (±0.01)	3.91 (±0.01)	4.17 (±0.02)	4.37 (±0.01)	4.81 (±0.04)
current	3.89 (±0.02)	3.90 (±0.04)	4.16 (±0.01)	4.32 (±0.01)	4.74 (±0.02)

Figure 7. Average distance from 2007-2012, including standard error [7].

Not only are users connected by an average distance of 4.74, researchers found that 92% of feasible pairs of individuals were connected by 5 edges or less as shown in *Figure 8* [7]. Looking at *Figure 6* above, we can see the small world phenomenon came into play from 2008 - 2012. The average distance actually increased from 2007 to 2008 which we belive can be explained by the structure of Facebook in its earliest days. Prior to September of 2006, Facebook was only open to students at select universities and for employees in certain businesses [5]. The average distance between users should be lower because most users in the network had some real affiliation with one another (same education or employer). Thus, when Facebook expanded to the general public, it is logical that in the first few years of expansion the average distance actually increased as its users came from much more diverse backgrounds. However, in later years it was clear that the small world phenomenon was present in Facebook's network. Even though the graph was becoming less dense, the average distance between any two users was still decreasing [7].

	it	se	itse	us	fb
2007	65% (11)	64% (6)	67% (9)	95% (5)	91% (5)
2008	77% (7)	93% (5)	77% (5)	83% (5)	91% (6)
2009	90% (5)	96% (5)	75% (5)	86% (5)	94% (6)
2010	98% (5)	97% (5)	91% (5)	91% (5)	97% (6)
2011	90% (4)	86% (4)	95% (5)	97% (5)	89% (5)
current	88% (4)	86% (4)	97% (5)	97% (5)	91% (5)

Figure 8. Percentage of pairs that can be reached within average distance [7].

Since the experiments from 2012, Facebook researchers have revisited their work on the small world phenomenon. In 2016, Facebook ran the algorithms again on their network of 1.6 billion users and found that the average distance between two individuals had shrunk again [6]. Researchers found that in 2016 the average distance between two individuals was **4.57** (3.57 intermediaries), and the average distance between two individuals in the U.S. graph was **4.46** (3.46 intermediaries) [6]. Although the average distance is not that much less than the distance found in the 2012 experiments (average distance of 4.74 over the entire Facebook network), the 2016 findings show that the small world phenomenon is still in effect. Especially given that Facebook's network in 2016 was more than double the size of its network in 2012, it is evident that individuals who appear to be far removed have surprisingly short paths between them.

3. Testing the Small World Phenomenon in an Era Dominated by New Social Media

Since the most recent installment of small world experiments in 2016, the world of social networks has changed dramatically. Facebook is not the only means of online communication, as other applications have become popular among select age groups. Most notably Snapchat and Instagram have both gained wide popularity among younger demographics. In 2019, 67% of people age 18-29 used Instagram, and 62% of people age 18-29 used Snapchat [8].

This poses the question of whether the average distance between two individuals could be even lower through some of these alternate social networks. We conduct an experiment below that seeks to explore this possibility through a small world experiment on Instagram. Out of the alternative social media networks from Facebook, Instagram was chosen as the platform for this experiment because it was founded later in 2010 and it is a popular platform for young adults [9].

3.1 Brief History of Instagram

Instagram was founded in 2010 by Kevin Systrom and Mike Krieger [9]. Systrom and Krieger realized that sharing mobile photos would become a mainstream activity in the near future and wanted to build a platform to capture this new market [10]. Instagram is now a photo-sharing mobile platform that reports over 400 million monthly active users who have shared over 40 billion pictures [11]. Users can upload photos, follow other users, like or comment on photos, and message other users. Unlike Facebook, Instagram connections are not necessarily bidirectional. Whereas on Facebook two users must mutually agree to be connected (be “Friends”), “following” a user on Instagram and being followed by that user are mutually exclusive activities. Users on Instagram also have the option of making their profile “public” (visible to the anyone), or “private” (visible to only their followers).

3.2 Designing the Experiment

Procedure
This experiment was set up to mimic to Milgram’s study of the small world problem. First an Instagram user was selected as the target user. Next, a group of Instagram users were selected to be starting users. Like the participants in Milgram’s study, they attempted to generate an “acquaintance chian” [1] from themselves to the target user. Each starting user was given a picture that described this experiment, listed the Instagram username of the target user, and asked them to send the picture on.

Starting Users
The pool of starting users was comprised of 25 Instagram users who volunteered to participate in the experiment. All of the starting users were individuals who did not know the target user in real life and did not follow the target user directly on Instagram. In Milgram’s original experiment, the starting population was split into two groups to test whether geographical distance had any impact on the results [1]. One subpopulation also lived in Boston with the target individual, and the other subpopulations lived in Nebraska [1]. Since the target user is a college student, we would consider their most recent place of residence to be on their college campus. Thus, the starting users of this experiment were

split into two subpopulations of users as well - those who currently do or have in the past attended the same university as the target user (6 users), and those who have not attended the same university (19 users).

Intermediaries

Participants were allowed to select the next intermediary from anyone in the list of accounts they follow on Instagram. The intermediaries in this experiment were selected by starting users or previous intermediaries. Their participation in the study was voluntary, as we did not contact or interfere with their choice to continue passing the picture along.

Target User

The target user was a volunteer who is currently a first-year Dental Student at the University of Pennsylvania and grew up in Vancouver. In Milgram's original experiment, prior to the creation of the Internet, the other participants were given a standard set of information on the target individual, such as the age, sex, and occupation [1]. However, this experiment is meant to capture how social distance has changed given the meteoric rise of social media. Thus, we wanted the users to be able to use the full power of Instagram. Instead of selecting information to give the other participants, we gave them the freedom to use any information they could find on Instagram about the target user.

The target user's Instagram account is private, so only those who follow his account have access to his full profile. However, those who do not follow the account can still see limited information upon searching for his username. From the perspective of the starting users who all did not follow the target user, the information the target chose to list publicly are the following: name, high school, high school graduation year, university (Penn), university graduation year, a picture of a Taiwanese Flag, a picture of a Canadaian flag, and a list of mutual followers (users who follow the person searching for the profile as well as the target user).

Passing The Picture

The starting users were all sent a picture with instructions on what they should do. All the pictures included a request to participate in the experiment and a description of this experiment. The picture also included the target user's Instagram username, and the instructions described above on how they can look for any supplemental information about the target user on Instagram. The participants were encouraged to use any information found to inform their decision on who the next logical participants should be.

The next part of the picture gave instructions on how they should continue to forward this image. In Milgram's experiment, each participant was only allowed to forward the image to one individual [1]. This is similar to conducting a Depth First Search (DFS) [12] from each starting individual to the target individual since we are pushing forward along one path until we find the target, or until the thread dies. In Easley and Kleingberg's analysis of Milgram's experiment, they state that "to really find the shortest path from a starting person to the target, one would have to instruct the starter to forward ta letter to all of his or her friends, who in turn should have forwarded the letter to all of their friends, and so forth" [2]. This “flooding” approach would be analogous to a Breadth First Search (BFS) of the network instead [2].

While it would be feasibly impossible for us to track a true Breadth First Search [12] from a starting user on Instagram due to the volume of people each person follows, we wanted to incorporate the factor of breadth into the experiment in some capacity for comparison. Thus, the starting users were given two variations of instructions on how to progress the experiment forward. 22 users were given the photo in *Figure 9*, instructing them to send the photo (*Figure 9*) to ***only one*** person that they follow on Instagram. These 22 users consist of 5 users who attended Penn, and 17 users who did not attend Penn. Three users (1 who attended Penn, 2 who did not) were given the photo in *Figure 10*, instructing them to send the photo (*Figure 10*) to ***five*** people they follow on Instagram.

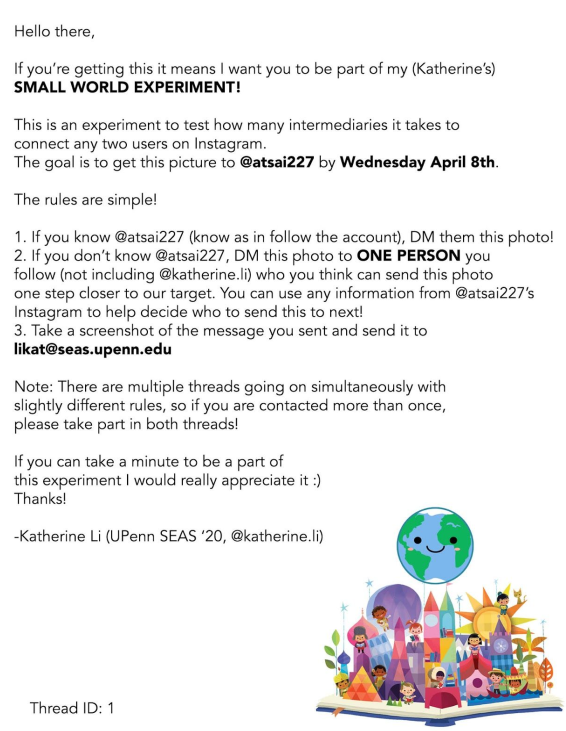

Figure 9. Version 1 of Picture Message.

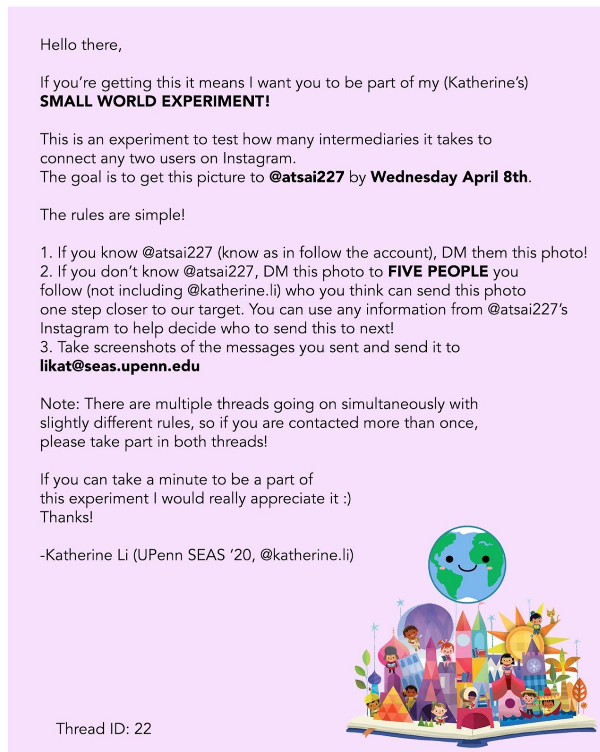

Figure 10. Version 2 of Picture Message

If the image was passed to someone who followed the target user directly, the participant was instructed to send the image to the target user and the thread would be considered complete.

Variables Measured

The primary variables this experiment seeked to measure were:

1. The number of intermediaries (and distance) of each thread.
2. The number of threads that reached completion.
3. For completed threads, the amount of time it took to reach completion.

4. The breadth of participants reached, particularly how many different penultimate intermediaries were used to reach the final target.

3.3 Limitations

Looking at previous research on the small world phenomenon, Milgram's approach and Facebook's approach are very different as one was conducted through experimentation and one was conducted through algorithms. Although this study looks at whether we have improved upon Facebook's results from four years ago, we do not have access to the entire Instagram network to run an algorithm that generates theoretical estimates. Thus, we chose to follow Milgram's approach and collect data empirically.

Because this is not a professional research experiment, the biggest limitation we faced was the selection of participants. The pool of individuals that we could pick from for our starting users and ending user were from our known network of acquaintances. In the experiment, it was specified that the author of this experiment could not be an intermediary, but inevitably this is a smaller network than that of the entire Instagram network. Thus, when looking at the results, we need to take into consideration the fact that all the initial participants do share a connection with the author of the experiment. However, there are no other restrictions on the intermediaries. We do not have access to as many starting users and our sample size was limited by the number of volunteers we could find for the experiment.

3.4 Running the Experiment

This experiment was run over the course of one week, and any threads that did not reach the target after a week were considered incomplete. Given that sending the picture to the next participant is instantaneous over Instagram, we believed that one week was an ample time frame for getting the image to the target.

In the instructions on the images, it was included that each participant should send a screenshot of the message they send, with the username of the next participant visible, to the the author's email address (likat@seas.upenn.edu). The screenshots allowed us to document where each image was sent next. In addition to collecting these screenshots, we also labeled each thread with an ID number. At the bottom of each picture sent to the starting users, there was an ID number listed from 1 to 25. Thus, in the screenshots we were sent by email, we could also see where the sender got the image from.

Whenever the target user received an image from a participant, he would send an email to us as well to keep track of if and when the threads reached completion.

3.5 Results

Number of Threads Completed

At the end of the experimentation period, 19 of the 25 threads reached the target user. Thus, there was a 76% success rate. In Milgram's experiment, only 29% of the chains made it to completion [1]. Without considering degree of separation yet, we can see that Instagram has made the world more connected as there are more feasible ways to reach the target. Of the 25 starting users, 20 users (80%) sent the photos to a first intermediary. In Milgram's experiment, 217 out of the initial 296 starting individuals (73%) sent the document to an acquaintance [1]. The percent of people who took the first step to starting the chain has increased but the difference is not as significant. We believe the similarity in the two percentages lies in the inherent apathy present in any selected population. However, because sending a message on Instagram is much more convenient than going to a post office to mail a letter, we can see that there is an increase in initial engagement.

Time to Completion

The average time to completion of successful threads was 1.78 days. The average time to completion within successful threads started by Penn affiliates was 1 day, while the average time for threads that did not start with Penn affiliates was 1.89 days. A full distribution of completion times is shown in *Figure 11* below. The graph is skewed to the right, with most threads being completed in either less than 1 day (labeled as 0.5 days), or after 1 day. This distribution reflects the power of social media platforms like Instagram to shorten communication time significantly.

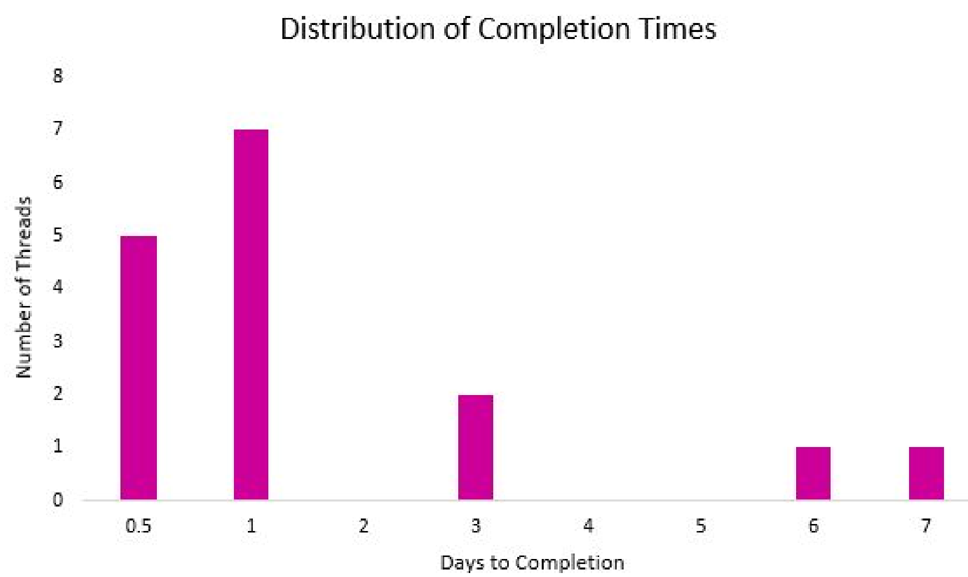

Figure 11. Distribution of times to completion for successful threads.

Average Distance Between Users

Overall, the average distance between the starting users and the target user using only Instagram as a means of communication was a distance of **2.32**, or 1.32 intermediaries. As mentioned in the Limitations, the starting users did not span as expansive of a network as Milgram's experiment. However, we believe this experiment is a good representation for the average distance of Instagram's most popular demographic of 18-29 year olds [8] since most of the intermediaries were users that were not in our known network.

The distribution of the number of intermediaries can be seen in *Figure 12*. Of the threads that started from a Penn student, the average distance was **2** (1 intermediary). Of the threads that started from someone who did not attend Penn, the average distance was **2.43** (1.43 intermediaries). From *Figure 12*, we can see that the results are right skewed, with the mode being 1 intermediary.

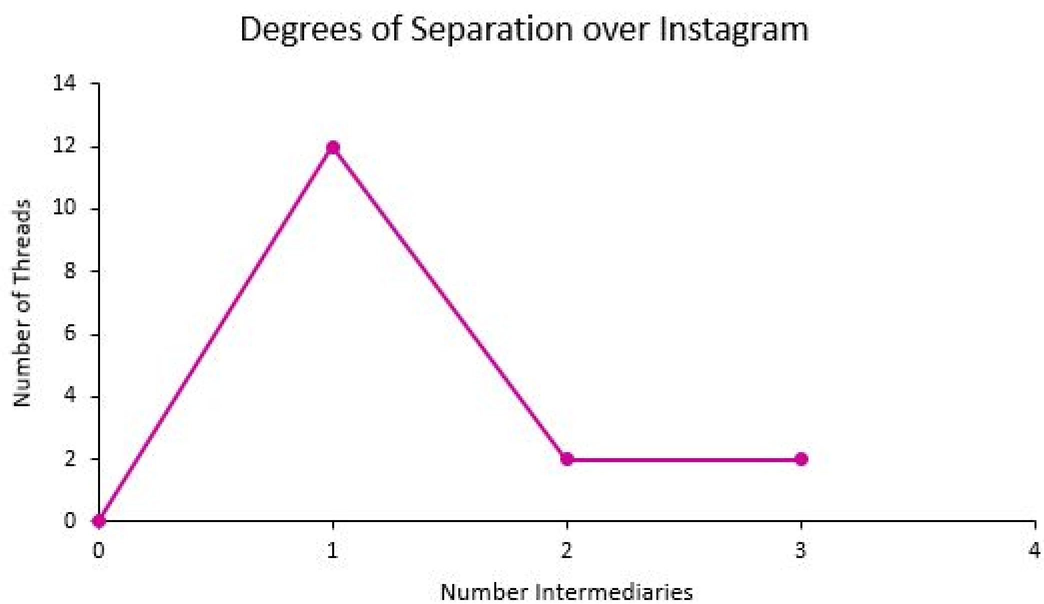

Figure 12. Number of intermediaries in successful threads.

Thus, social media in the current age has decreased the number of degrees of separation for Instagram users. We believe the biggest advantage that Instagram provided was the fact that Instagram users could see if they shared any mutual followers with the target user. Although this may seem like an obvious shortcut, this is reflective of the breadth of information that new social media platforms have provided in our lives. This functionality likely had a big impact on the high number of threads with only one intermediary.

Even compared to Facebook's most recent study in 2016 (average distance of 4.46) [6], the average expected distance on Instagram in 2020 is shorter. Part of the difference must be inevitably attributed to the limited population of users that we were able to select as starting users. We believe part of the difference can also be attributed to the fact that the number of users has grown significantly even from 2016 to 2020. Facebook's network grew from 1.6 billion users [6] to 2.5 billion as of their fourth quarter in 2019 [13]. Instagram has reported over 1 billion monthly active users as of January 2020 [14], while they reported about 600 million monthly active users in 2016 [15]. As discussed later in the analysis of Instagram's network, there are other aspects of Instagram's graph structure that attribute to a smaller degree of separation as well.

Geographical Impact

One of the more surprising results was the success rate within each subpopulation. Of the 6 potential chains that started from users who have attended Penn, only 33% of the chains (2 chains) were completed. Of the 19 potential chains that started with users who did not attend Penn, there was an 89% success rate of completion (17 completed chains). The low success rate among those who went to Penn may be attributed to greater apathy from the starting users, as 3 out of the 4 unsuccessful chains never made it past the starting user. This is likely because within the Penn community, we intentionally asked more distant acquaintances to be in the study because the starting users could not already be friends with the target user. Thus, they may not have been as invested in the experiment. Nevertheless, the high success rate outside of Penn suggests that with the use of Instagram, physical distance is not a barrier to success.

In *Figure 13* on the following page, we can see a visualization of how each thread spread from starting user to target user. The users that are represented as pink dots are people that are affiliated with Penn while the users that are represented as orange dots are those that are not Penn affiliated. As shown from the graph, many users were able to find a connection across subpopulations since many threads include participants from both populations.

Figure 13. Visualization of acquaintance chains from Instagram experiment.

Breadth First Search (BFS) Effects

There were three starting users who were given pictures that instructed them to pass the image along to five Instagram users they follow instead of one. This was implemented to simulate the benefits of amplification that a true Breadth First Search (BFS) would involve [12]. One of the threads was never moved forward by the starting user, who was a student at Penn. One of the starting users sent the thread to five users but none of those five intermediaries passed the message on so that chain was incomplete as well. The last user sent the image to five intermediaries and four of those five threads were successful in reaching the target user. All four of the successful threads went through one intermediary. Since the single path searches were so successful, it is hard to make a conclusion on whether increasing the breadth of the search improved the results. The first image out of the four successful threads was received by the target in under one day (0.5 days), which is less than the average time to completion of 1.78 days so it may have improved upon the search time. However, we believe more data would need to be collected to reach a conclusive result.

Penultimate Intermediaries

One of the most interesting aspects of Milgram's experiment was the pattern in penultimate intermediaries present in successful chains. In Milgram's experiment, he observed that there was a large amount of convergence [1]. The full convergence chain is shown in *Figure 14* which shows that the 64 letters that reached the target were sent by 26 penultimate intermediaries [1]. The distribution of letters among the 26 intermediaries was also heavily skewed [1]. Sixteen letters (25%) went through one of the target's neighbors (neighbor in real life), and another 10 went through one of the target's business contacts [1].

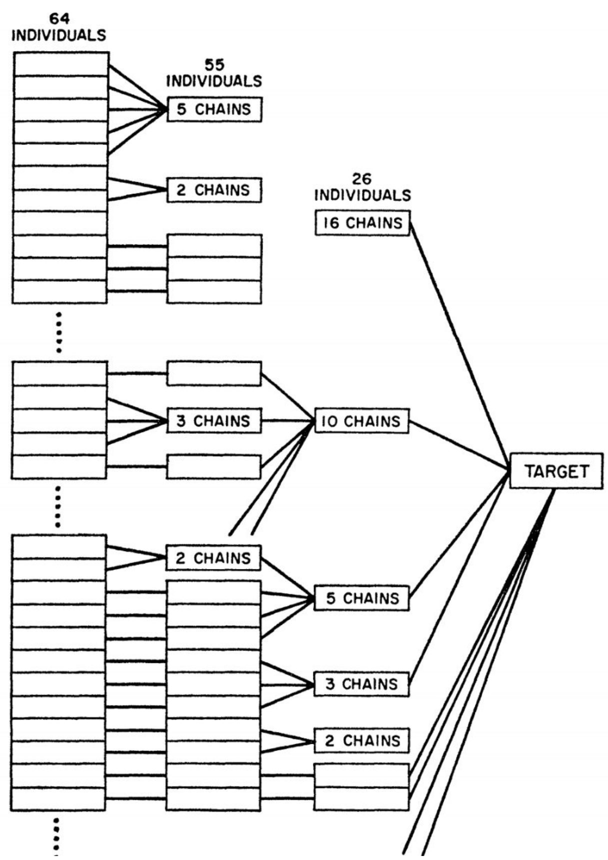

Figure 14. Penultimate contacts of successful chains in Milgram's experiment [1].

In our Instagram experiment, much ***less convergence*** was observed. Going back to *Figure 13*, it is apparent that the penultimate contact in most threads were different people. There were 15 different penultimate intermediaries, with only one intermediary who was part of more than one thread. Similar to Milgram's experiment, the intermediary that received multiple images did represent a significant proportion of successful chains though. The intermediary in the Instagram experiment received 4 images representing 21% of successful chains, which is similar to the 25% of successful chains that the most popular penultimate contact of Milgram's experiment was involved in [1].

4. Analyzing Instagram's Platform with Graph Theory

4.1 In-Degree and Out-Degree

Perhaps one of the biggest differences between Facebook and Instagram is the fact that Instagram can be represented as a directed graph, while Facebook is an undirected graph. On Facebook, two users must mutually agree to be "friends" so there is no directionality. On Instagram, there is directionality because User A's decision to follow User B is independent of User B's decision to follow User A. In a study run by Arizona State University on the profiles of one million Instagram users, the log ratio of followings to followers was observed and plotted in *Figure 15* [16]. In more technical terms, this shows us the relationship between in-degree and out-degree for each node in Instagram's graph.

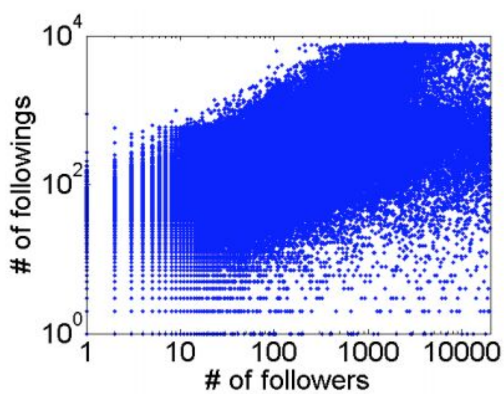

Figure 3: log-log plot for followings and followers

Figure 15. Plot of in-degree to out-degree for one million vertices on Instagram [16].

While there is a relatively linear relationship between the two, indicating that most people follow and are followed by a similar amount of people, the spread is not completely symmetrical. Overall the graph is skewed towards the bottom right corner, indicating that there are a good amount of accounts that have many followers (high in-degree) but do not follow many people (low out-degree). The most extreme of these cases can be attributed to celebrities and branded accounts (ex. Clothing brands, restaurants, etc.).

Looking back at our experiment, the slight inbalance in the in-degree and out-degree of some nodes may be a factor in the fewer degrees of separation present on Instagram versus Facebook. The greater the in-degree of a vertex, the more opportunities there are for that vertex to receive one of the pictures being passed around. The fact that most vertices have a higher in-degree than out-degree means they have more opportunities to receive the image than they would have if this were run on a undirected graph.

4.2 Reciprocity

Another factor to consider since Instagram is a directed graph is the degree of reciprocity present in the network. In the Arizona State University study on Instagram, it was discovered that only **14.9%** of relationships on the platform were reciprocal [16]. In other words, on average only 14.9% of the accounts that an individual follows are also following that individual. This number is likely heavily influenced by all the celebrities in that bottom right corner of the graph above, who have an extremely low percent of reciprocal relationships [16]. The median reciprocity would likely be higher than this number.

Nevertheless, this figure implies that Instagram has a high proportion of ***weak ties***. The “strength” of a tie would be based on the amount of time spent together, the emotional investment, and trust [4]. Thus, two people who are connected by a strong tie should have a mutual following on Instagram. The low percentage of reciprocity implies that many ties are not very strong. As mentioned earlier, weak ties are important because they connect people who otherwise have no connection [4].

The high proportion of weak ties on Instagram should decrease the average degrees of separation between two individuals because weak ties span large social distance. An image that would have gone through two or three individuals with medium strength ties can instead go through one person connected by a weak tie. The number of weak ties is also likely a factor in the increased number of threads that made were completed (76% success rate). Removing a weak tie would split one group into two connected components who have no feasible way of communicating. While a network such as Instagram probably has very few large connected components and no one user will disconnect these big components, at a local level weak ties still do have an impact on the feasibility of success for these acquaintances chains [4].

4.3 Clustering

Another structural feature of social networks is their ***clustering coefficient***. The clustering coefficient measures the extent to which a user’s friends are also friends of each other [16]. The clustering coefficient of each vertex C_i can be calculated as shown below [16].

$$C_i = \frac{|e_{jk} : v_j, v_k \in N_i, e_{jk} \in E|}{(k_i)*(k_i - 1)}$$ [16]

N_i is the neighborhood of the i'th node, and k_i is the total number of nodes present in the neighborhood N_i [16]. After each vertex's individual clustering coefficient was calculated, the average was taken to get the cluster coefficient for Instagram's network [16]. The clustering coefficient from Arizona State University's study of Instagram was 0.42 [16]. Instagram's coefficient is relatively high compared to other social media platforms like Twitter, implying that Instagram has a higher number of small cliques compared to other platforms [16].

4.4 Small World Properties

Given the results of the small world experiment we ran, and the structural properties of Instagram's graph, we can conclude that Instagram does exhibit the properties of a small world network. Small world networks have small distances between any two nodes and high clustering [17]. The average degrees of separation between any two individuals from the experiment was 2.32 which shows that individuals who seem "far removed" have become much more connected from the rise of Instagram. Instagram also has high clustering of small cliques as shown above, so it exhibits both properties of small world networks.

4.5 How Graph Theory has Shaped Instagram Applications

Aside from its function as a site for sharing photos with friends, Instagram has become a hub for marketing due to its directed graph structure. Because of the directed graph structure, users are able to follow celebrities and brands that they enjoy. Thus, it is a great platform for companies to advertise to a huge audience. Additionally, given the small world nature of Instagram demonstrated in our experiment, a picture posted can travel a long way as most people are connected by a short distance.

The structure of Instagram's graph has also informed the business model the company has come to adopt. Instagram has given rise to a new method of marketing through certain users known as "influencers". Influencers are users who have built a large and engaged following on Instagram and have lots of "influence" because their followers respect their opinions [18]. On the plot in *Figure 15* earlier, influencers represent the the points that skewed to the bottom right corner, but are not as extreme as celebrities.

However, influencers' biggest value add to companies on Instagram is not the number of people who follow their accounts. These influencers are so important because they play the roles of the ***weak ties*** between brands and the high number of small cliques.

Companies could pay for huge advertising slots directly but that would be very costly and likely ineffective. Essentially they would be appealing to tons of people they have no tie to, who consequently have no motivation to pay attention to the advertisement. Companies will enter partnerships with influencers to market their products more authentically to target audiences [18]. Because influencers are such a valuable role to Instagram's revenue model, they have carved themselves a hefty slice of Instagram's budget. As of 2020, the Instagram influencer marketing industry is estimated to be a $2 billion industry [18]. This tactic has also proved to be very effective. 72% of users reported having made a fashion or beauty related purchase after seeing a paid promotion (from an influencer) on Instagram [19]. The nature of this weak tie is the perfect tool for marketers - the tie is weak because people have no personal connection to the influencers they follow, but they usually do have a strong connection to the fashion sense or self-branding of these influencers.

5. Alternative Platform Analysis: LinkedIn

In addition to Instagram, a number of other social media sites have found their niche in today's market and host their own social network of users. Using concepts from graph theory, we can analyze how these other sites have been able to successfully capture their audience. One of the most popular and distinct networks is LinkedIn.

5.1 Brief History of Linked In

LinkedIn was founded in 2002 and launched in 2003 [20], just after the "dot-com bubble" burst in 2000 [20]. The founders Reid Hoffman, Allen Blue, Konstantin Guericke, Eric Ly, and Jean-Luc Vaillant sought to create a social network for professional development as opposed to social development [20]. A user on LinkedIn is able to create a profile that describes their professional accomplishments, connect with other users, and see job availabilities listed by companies. Companies can also create profiles on the platform that describe their business and any hiring opportunities. Two years after the platform launched, they were able to hit 1.7 million professionals [20]. Since its founding in 2002, LinkedIn's user base has grown tremendously. As of April 2020, LinkedIn has reported about 675 million users registered from over 200 countries [21].

Aside from its success in number of users, LinkedIn has been lauded for its unique business model. The company didn't attempt to make money until 2005 through premium membership services [22]. Today, the company has three main streams known as Talent Solutions, Premium Subscriptions, Marketing Solutions [23]. This is different from most platforms who rely primarily on advertisements for revenue. In this analysis, we will see how LinkedIn's business model successfully used concepts from graph theory to recognize what the platform's value proposition is.

5.2 Graph Structure and Density

LinkedIn can be described as an undirected graph, as two users can only be connected if there is mutual agreement on the connection. One of the most notable properties of LinkedIn's network is its ***high density*** compared to other platforms. Density is measured as the number of edges in a graph compared to the number of potential edges in the graph [24]. The maximum number of edges exists in the case in which every vertex in the graph is connected [24]. This would be calculated as **(|V| $\times$ |V| - 1) / 2** in an undirected graph [24].

The median number of connections on LinkedIn is between 500 and 999 connections but this includes profiles of companies as well as regular users [22]. However, even among regular users, the average number of connections is 400 per user, which is still relatively high [25].

|E| = 400 edges / user $\times$ 675 million users ÷ 2 = 135 trillion edges

Max Number Edges = (|V| $\times$ |V| - 1) / 2 = (675mm $\times$ (675mm - 1)) / 2 = 2.278 $\times$ 10^{17} edges

LinkedIn Density = |E| / Max Number Edges = **5.925 $\times$ 10^{-7}**

If we assume that each user has 400 connections and there are 675 million users, then the number of edges in LinkedIn's network would be 135 trillion edges. Because LinkedIn is an undirected graph, we need to divide by 2 to avoid double counting each edge (each connection). The number of vertices |V| is equal to the number of users which is 675 million. Thus, the maximum number of edges possible is 2.278 $\times$ 10^{17} and the density of the network is 5.925 $\times$ 10^{-7}.

For comparison, the average number of friends per Facebook user is 338 and the median number of friends is 200 [26]. As of Q4 2019, Facebook reported 2.498 billion Monthly Active Users in their latest earnings report [27]. Using the same metrics for calculating density, we can see that Facebook is less dense than LinkedIn.

|E| = 338 edges / user $\times$ 2.498 billion MAU ÷ 2 = 422 trillion edges

Max Number Edges = (|V| $\times$ |V| - 1) / 2 = (2.498bn $\times$ (2.498bn - 1)) / 2 = 3.12 $\times$ 10^{18} edges

Facebook Density = |E| / Max Number Edges = **1.35 $\times$ 10^{-7}**

The greater density in LinkedIn's network reflects the functionality of the platform. Whereas Facebook is meant to help people stay in touch with those in their personal network, LinkedIn is meant to help its users reach a larger network of people who allow them advance professionally. Thus, users are very inclined to connect with people outside their immediate friend circle which leads to a denser network. We would expect heavy overlap in the populations of people who use Facebook and LinkedIn, albeit for different purposes. If Facebook's network represents an individual's strong ties, then the greater density of LinkedIn's network also implies that the network is full of weak ties, which the network has recognized as extremely valuable.

5.3 The Value in Weak Ties

As analyzed earlier, weak ties in any network are important for diffusion of information because they "bridge" together two individuals and their entire networks of friends who would otherwise have no contact [4]. In LinkedIn's network, weak ties are especially important because of the network's heavy focus on transmitting information to targeted sources. As companies get inundated with applications, an introduction through an individuals as opposed to an application pool is important. 35.5 million users have reported getting a job through someone whom they were connected with on LinkedIn [22].

Weak ties are important in this process because they allow users to reach new companies and contacts. While strong ties are more loyal and have higher incentive to help a user get a job, they are likely to have the same contacts as the user. Weak ties help the user reach professionals quickly in industries or companies that the user does not have connections in but would like to work in. Studies have shown that longer chains of introductions tend to have a lower success rate. 85% of introduction requests were approved by the first intermediary, but less than 33% make it to the final target [28]. Thus, one weak connection between you and a professional you want to be connected with, is much more powerful than a few strong ties between you and the same professional.

This also relates to why LinkedIn has a higher density than other platforms. First degree connections are more effective in passing information along successfully so it is in users' best interests to make as many connections as possible.

5.4 The Value of Idle Users

Another curious aspect about LinkedIn's network is the amount of time spent on the platform. On average, LinkedIn users spend about 10-20 minutes on the site daily [22]. Compared to other platforms, this is a very small amount of engagement in time. In a study conducted by researchers Saleem Alhabash and Mengyan Ma, they found that users spend an average of 106.35 minutes/day on Facebook and 88.92 minutes/day on Instagram [11]. This is logical, as we can check on what our friends are doing every day but there's no reason why we would be looking for a new job on LinkedIn every day. Although LinkedIn has a functionality that allows users to make posts to their network in a similar fashion to Facebook and Instagram, only 1% of users post content weekly [22].

While low engagement may seem to indicate that a business is not prospering, these idle users are arguably the most valuable aspect of LinkedIn’s network. There are essentially two ways in which jobs are filled [28]. There are positions that are filled because an applicant is actively looking for a new job and finds a job posting [28]. There are also positions that are filled because the company discovers a qualified candidate and actively seeks them out to see if they would be interested in joining the firm [28]. The first scenario of hiring an actively seeking applicant would still occur without the help of LinkedIn [28]. Companies can advertise job openings on their website, in newspapers, or at information sessions.

LinkedIn’s value-add primarily comes from cases in which a firm finds a new hire who was not actively looking for a new job. If someone already has a job, they will not be actively looking for a new job unless they have a strong reason to. That does not mean however, that they are not *open* to taking a different job if it is a better opportunity. These passive potential applicants are equally as important to employers since they could be very qualified for the job, but without LinkedIn it is difficult for employers to scout out such people. LinkedIn’s huge network of professionals and their resumes makes these passive potential applicants accessible to numerous companies. From the employer’s perspective, LinkedIn is useful because it expands their hiring pool to include currently employed individuals who may be willing to switch jobs. From any employee’s perspective, passively recruiting offline can be difficult because they run the risk of upsetting their current employer. However, with LinkedIn it becomes socially acceptable to share their resumes publicly all the time where other employers can see it.

5.5 Building a Successful Business Model using Graph Theory

LinkedIn is a platform that has been able to use structural properties such as graph density and the presence of weak ties to their advantage. The company also recognizes the value of Milgram’s conclusion that people are more closely connected than we would expect. Although discontinued now, LinkedIn once even had a feature in 2014 based directly on the concept of the “Six Degrees of Separation” from Milgram’s experiment that allowed users to see their entire network of individuals up to six degrees away [29]. Today users can still see whether any user is within three degrees of separations away from themselves.

LinkedIn is also considered a ***multi-sided platform (MSPs)***, which is a product that creates value by enabling direct interaction between two or more groups [30]. Multi-sided platforms occupy privileged positions in their respective industries because they reduce search costs for both groups [30]. In the case of LinkedIn, the two groups are employers and potential employees. After two years of operation in 2005, LinkedIn began to think about how they should monetize their platform [20]. There were two basic models they could adapt [28]. They could introduce new premium services that offer additional insights into your network, or they could charge everyone a flat fee for using the services [28].

A business should adopt a platform that correctly incentivizes its users to generate the most revenue possible. In the case of LinkedIn's multi-sided platform position, the question lies in which group benefits from the greatest cost reduction, and which group provides the most value to LinkedIn as a service. Recognizing the value brought in by weak ties and idle users, LinkedIn decided that its network of potential employees was the most valuable aspect of their platform and must be incentivized to stay on the platform. Thus, they ultimately adopted three premium services that companies and employers would primarily bear the cost of.

LinkedIn's greatest source of revenue comes from their Talent Solutions service. In 2015, LinkedIn reported that Talent Solutions comprised of 62% of their revenue from the first quarter [23]. Talent Solutions provides services that alert recruiters when a candidate who is suitable for a particular opening at a company is recognized [31]. Once alerted, recruiters have the ability to "tap" a candidate or let them know that a company has viewed their profile and is interested [31]. This directly relates to the value derived from LinkedIn's dense network of idle users. LinkedIn does not charge idle users to incentivize them to stay on the network even when they are not actively recruiting. Instead, LinkedIn has chosen to target recruiters and employers since they are willing to pay for accessibility to idle or actively searching candidates.

LinkedIn's two other streams of revenue come from their Premium Subscriptions (19% of revenue Q1 2015) and Marketing Solutions (19% of revenue Q1 2015) [23]. Marketing solutions are the standard advertising revenue that most social media sites benefit from [23]. Premium Subscriptions charge users a monthly fee in exchange for special tools, and consist of different price tiers as well [32]. The base level premium subscription allows any user on LinkedIn to send a limited number of "InMail messages" to any other user without being connected, while normal users must be connected to send messages to one another [32]. A higher price-tiered subscription level also provides users with advanced search filters and more details on who has viewed their own profile [32]. At the highest tier, a user can view the full profile of anyone in their extended network, which includes people who are two, or three degrees of separation away [32].

This premium service recognizes the value of weak ties and monetizes its value. In different variations, each level of service is essentially offering more information about the weak ties in a user's network, or even allows a user to establish a weak tie with someone outside their network. These Premium Subscriptions range from $30/month to $120/month [32] which is higher than most premium level subscription services (ex. Spotify premium, Amazon prime). However, because weak ties are so important in a network focused on reaching professional contacts, LinkedIn's success has shown that users belive the hefty price is worth its value.

6. Alternative Platform Analysis: Twitter

Another popular network with an interesting graph structure is Twitter. Although Twitter faces high competition from other socially driven networks like Facebook and Instagram, Twitter has also found a niche for success among today's audience. In the following sections, we will analyze how Twitter's functionalities have informed the structure of its graph.

6.1 Brief History of Twitter

Twitter is a social media site founded in 2006 by Jack Dorsey and his associates that brought together new media sharing culture and dispatch enthusiasm [33]. As of 2020, Twitter has about 275 million users around the world [34]. Twitter is a platform that allows users to share posts, known as "tweets" and connect with other users. The dispatch enthusiasm describes the nature of posts that Twitter users make [33]. Until 2017, users were only allowed to include 140 characters in their posts, and from 2017 to present users are allowed to include 280 characters in their posts [35]. However, brevity is clearly at the core of Twitter's platform as only 1% of tweets hit the 280 character limit, and only 12% even surpass the 140 character limit [36]. With this radio dispatch quality, Twitter has become the platform people turn to for sharing quick thoughts, making announcements, and engaging in debates.

6.2 Vertex Degrees

Like Instagram, Twitter can be represented as a directed graph. On Twitter, a user may follow other users and has an audience of users who follow them. Following an account and having the account follow that user back are mutually exclusive events. Thus, when analyzing degrees we can look at both the in-degree and out-degree. In a study conducted by Martin Grandjean on the Twitter network of the humanities community, he found that 63.1% of users have an in-degree and out-degree of less than 100 people [37]. Among the other ~40% of users studied, the distribution of in-degree to out-degree is shown below in *Figure 16* [37].

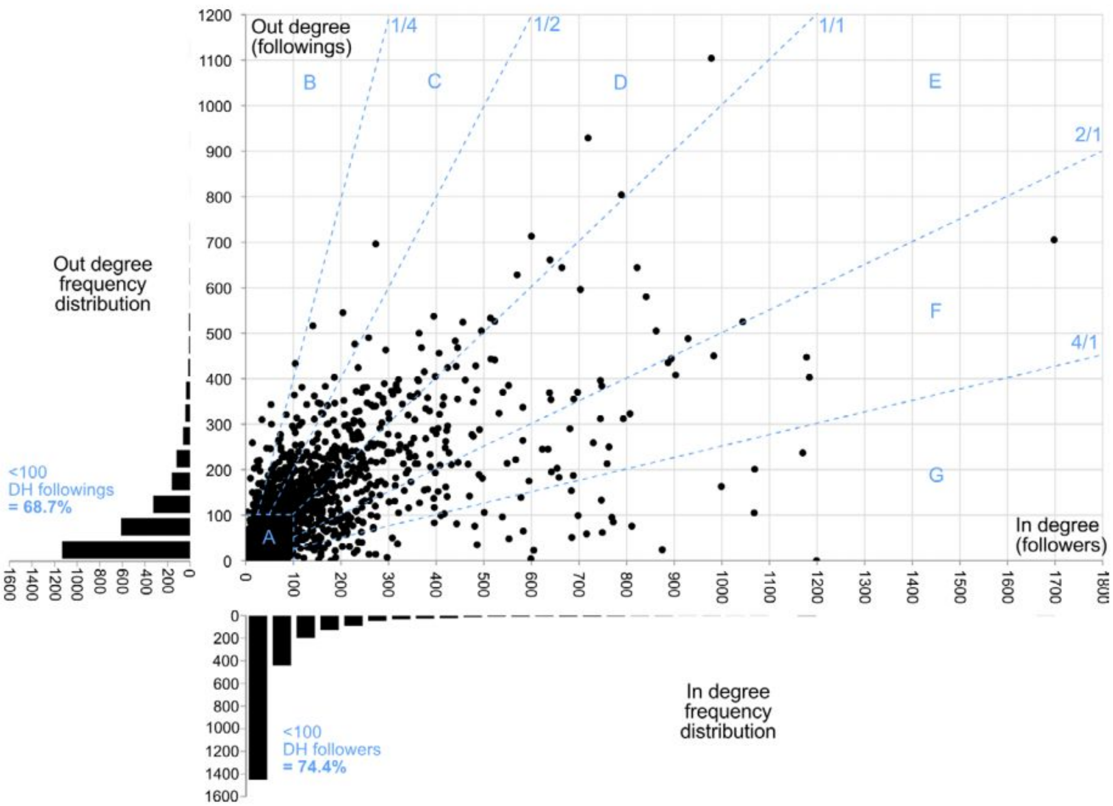

Figure 16. Distribution of in-degree to out-degree of Twitter users [37].

Categories B and C represent users who follow at least four times and at least two times as many users as they have followers respectively [37]. Category D represents people who follow up to two times as many users as they have followers, and is the second largest category outside of people who follow and are following under 100 users [37]. The users that fall into categories B, C, and D likely use Twitter as a "technological monitor", meaning they use Twitter to keep informed on news or topics they care about [37]. Thus, they are inclined to follow many accounts, but do not post content garnered towards increasing their own follower count. This population of users seems unique to Twitter's platform, as there was very little spread towards high out-degree and lower in-degree in Instagram's distribution (*Figure 15*).

Categories E, F, and G represent accounts that are followed by multiples more users than they follow themselves [37]. These accounts represent individuals who are distinguished in their fields and thus have a large following [37]. This distribution is similar to what we find on Instagram, with a small percentage of users able to garner this type of audience.

The distribution of in-degrees and out-degrees in Twitter's network also categorizes the network as a ***scale free network***. A scale free network is a network whose degree distribution follows a power law, where some nodes have a large degree but the majority of nodes have a small degree [38]. In the *Figure 17* and *Figure 18* below, we can see the power law distribution of both in-degrees and out-degrees that make this network scale free. Because Twitter is "scale free", as the network continues to scale, we can expect to observe the same distribution of degrees [38].

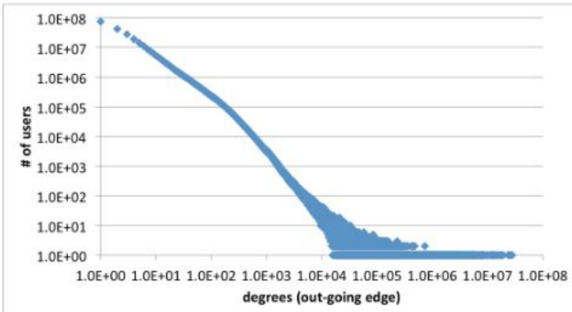

Figure 6. Degree Distribution of Followers

Figure 17. Distribution of in-degrees [38].

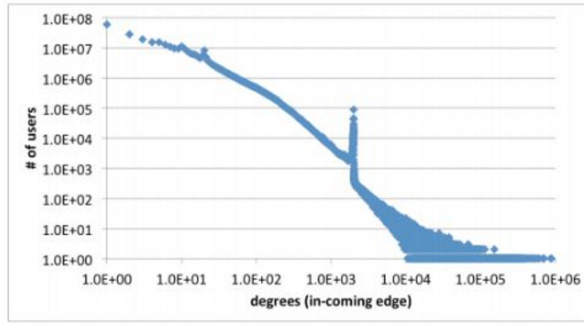

Figure 7. Degree Distribution of Friends

Figure 18. Distribution of out-degrees [38].

6.3 Reciprocity

In researchers Masaru Watanabe and Toyotaro Suzumura's study of Twitter's growing network, it was discovered that reciprocity in Twitter had decreased from 2009 to 2012 [38]. In 2009, Twitter's graph exhibited 22.1% reciprocity and in 2012 Twitter's graph exhibited 19.5% reciprocity [38]. One explanation for the decrease in reciprocity is that in those three years, Twitter's network expanded a lot internationally so greater gaps in interests, customs, or languages occured between users [38]. Given the nature of Twitter for sharing short announcements or thoughts, it should be expected that the reciprocity is not very high. Unlike Facebook, Twitter followings are based more on interest than on personal connection. Thus, as Twitter's network expands to include more diverse people, it is natural that there is a greater disparity in interests which results in lower reciprocity.

6.4 Small World Network Properties - Degrees of Separation and Diameter

Watanabe and Suzumura used the same HyperANF algorithm that the Facebook researchers used in their study to run a simulation of the small world experiment on Twitter's network [38]. The table below in *Figure 19* summarizes there findings, with the numbers on the left indicating which run the data came from (four runs total).

	Degree of separation		Diameter	
	2009	2012	2009	2012
1	4.39	4.48	25	70
2	4.46	4.65	26	71
3	4.53	4.54	25	70
4	4.62	4.71	25	71

Table 3. The results of degree of separation and diameter on each run.

Figure 19. HyperANF algorithm results on Twitter network [38].

This data is relatively consistent with the results from Facebook's experiment also run around the same time. However, one interesting factor is that the degrees of separation have actually increased from 2009 to 2012. In a true small world network, the degrees of separation should decrease when the number of users increases, as demonstrated in Facebook's experiment [7] and the results from our Instagram experiment. Thus, Twitter may not represent a small world network. This is likely due to ***lower clustering*** than normal small world networks. Because everyone has different interests and many interests, it is less likely that two neighbors of a user also follow each other. However, given that this research was conducted eight years ago, the results are not conclusive on what Twitter's network looks like today [38].

Another aspect that Watanabe and Suzumura measured was the diameter of the network from 2009 to 2012 [38]. The diameter of a graph is the maximum value of the shortest-path length of all pairs of users [38]. This measure can help estimate how expansive a network is. From the runs on 2009 and 2012 data, it is clear that the diameter of Twitter has increased significantly [38].

6.5 Limitations to the Power of Weak Ties

Based on the properties discovered above, it seems that Twitter's network is dominated by weak ties. Users follow accounts that they enjoy the content of, but are not necessarily emotionally close to. However, the ***power of weak ties*** in a network like Twitter with low clustering and low reciprocity is ***diminished*** compared to highly clustered networks like Facebooks or Instagram. Weak ties derive their value because instead of connecting just two people, they connect the clusters that these two individuals are a part of [4]. In a network like Twitter that is is not highly clustered, a weak tie does not bring as many people together. In this case, weak ties may even be a disadvantage in transmitting information because people connected by weak ties have less motivation to help pass a message along.

7. Conclusion

The world today is connected by numerous social media platforms that each occupy a different functionality. At its core, each platform is a graph of individuals and its **social properties** can be explained by the **properties of the graph** and vice versa. The most important aspect of social media examined in this paper was the **small world phenomenon** and whether it holds today. In Milgram's original experiment, he proved that without the help of the Internet, the maximum degrees of separation we could expect between any two people in the United States was six people [1]. Decades later in 2016, Facebook researchers concluded that users on Facebook's platform were now connected by 4.46 degrees of separation [6].

Given the rise of new social media platforms since Facebook was founded, we conducted an experiment to test whether the degrees of separation have decreased once again with these new platforms. As a very popular and newer platform, **Instagram** was selected as the platform of analysis. We conducted Milgram's **small world experiment** of passing a message to a target individual on Instagram and the results show that Instagram exhibits small world properties as well. The average degree of separation from our experiment was lower than Facebook's results from 2016 [6], and there was a much higher success rate of completed threads compared to Milgram's experiment. Thus, we can see that the world is becoming more connected as it's becoming more feasible to reach people and the average distance between two people is decreasing despite a growing user base.

This small world phenomenon is made largely possible by the existence of **weak ties** on social media platforms like Instagram. Because weak ties serve as a "bridge" between two unrelated groups, they allow messages to traverse large social gaps in these networks [4]. In addition to weak ties, the directed graph structure of Instagram is also a factor in the lower distance between two users. The greater in-degree of most vertices (users) compared to out-degree indicates that users have more opportunities to receive messages than they would have in an undirected graph.

To develop more context on Instagram's small world properties and graph structure, we conducted additional analysis on competitors like **LinkedIn**. The most notable property of LinkedIn's graph was the **high density** compared to other networks. Vertices on average had a higher degree [25], which is a result of LinkedIn's function as a professional social network. In a similar fashion to Instagram, LinkedIn's network was also strengthened by weak ties that introduce two clusters of users to one another [4].

The other competitor considered for comparison was **Twitter**, which also follows a directed graph structure like Instagram. On Twitter there was a greater spread of **in-degree to out-degree ratios**, as there were many users who followed more accounts than they have followers and vice versa [37]. This spread is a result of Twitter's role as a dispatch of short news and announcements which influences users to follow many sources of news [33]. While Twitter has an abundance of weak ties like Instagram, it does not benefit from the power of weak ties because Twitter is not highly clustered.

With the presence of all these social networks, the world really is becoming bigger and smaller at the same time. Despite the growing size of the population, the average distance between any two people over social media is astonishingly small. Not only are we becoming more connected by social media, but we now see a greater variety of platforms that have different graph structures influenced by their niche in the social media industry. As the industry continues to grow in users and diversity, we believe the next step in research on the small world phenomenon will be to analyze how this connectivity has influenced all sectors of industry in the world.

8. Works Cited

[1] Milgram, S., & Travers, J. (1969). An Experimental Study of the Small World Problem. *Sociometry, 32*(4), 425-443. https://www.jstor.org/stable/2786545

[2] Easley, D., & Kleingberg, J. (2010). The Small-World Phenomenon. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press.

[3] Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An Experimental Study of Search in Global Social Networks. *SCIENCE,* 301, 827-829. www.sciencemag.org

[4] Granovetter, M.S. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360-1380. https://www.jstor.org/stable/2776392?seq=1

[5] Brugger, N. (2015). A brief history of Facebook as a media text: The development of an empty structure. *First Monday, 20*(5). http://dx.doi.org/10.5210/fm.v20i5.5423

[6] Bhagat, S., Burke, M., Diuk, C., & Ismail O. F. (2016, February 4). *Three and a half degrees of separation*. Facebook Research. https://research.fb.com/blog/2016/02/three-and-a-half-degrees-of-separation/

[7] Backstrom, L., Boldi, P., Rosa, M., Ugander, J., & Vigna, S. (2012). Four Degrees of Separation. *ACM Web Science Conference (WebSci)*. https://research.fb.com/publications/four-degrees-of-separation/

[8] Anderson, M., & Perrin, A. (2019, April 10). *Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018*. Pew Research Center. https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

[9] New Audience Media (2018). *Brief History of Instagram.* New Audience Media. https://newaudiencemedia.com.au/brief-history-of-instagram/

[10] Peng, R., Sun, D., & Tsai, W. (2014). Success factors in Mobile Social Networking application development: Case study of instagram. *Proceedings of the ACM Symposium on Applied Computing,* 1072-1079. doi: 10.1145/2554850.2554902.

[11] Alhabash, S., & Ma, M. (2017). A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? *Social Media Society*, *3*(1). doi: 10.1177/2056305117691544

[12] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3rd ed.). Cambridge, MA: Massachusetts Institute of Technology.

[13] Clement, J. (2020, January 30). *Number of Facebook users worldwide 2008-2019.* Statista. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide

[14] Instagram for Business (n.d.). Facebook Business. https://www.facebook.com/business/marketing/instagram#

[15] Clement, J. (2019, December 3). *Number of monthly active Instagram users 2013-2018.* Statista. https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/

[16] Manikonda, L., Hu, Y., & Kambhampati, S. (2019). Analyzing User Activities, Demographics, Social Network Structure and User-Generated Content on Instagram. *Cornell University*.

[17] Nykamp, D. Q. (n.d.). *Small world networks*. Math Insights. https://mathinsight.org/small_world_network

[18] Cronin, N. (2020, February 10). *A Simple Guide To Instagram Influencer Marketing In 2020*. Hopper. https://www.hopperhq.com/blog/instagram-influencer-marketing-2020/

[19] Salpini, C. (2017, August 23). *Study: Instagram influences almost 75% of user purchase decisions*. Retail Dive. https://www.retaildive.com/news/study-instagram-influences-almost-75-of-user-purchase-decisions/503336/

[20] Yeung, K. (2013, May 5). *LinkedIn is 10 Years Old: Here's How it Changed the Way we Work.* The Next Web. https://thenextweb.com/insider/2013/05/05/linkedin-10-years-social-network/

[21] LinkedIn (n.d.). *About LinkedIn.* LinkedIn. https://news.linkedin.com/about-us#statistics

[22] Iqbal, M. (2020, April 24). *LinkedIn Usage and Revenue Statistics (2020)*. Business of Apps. https://www.businessofapps.com/data/linkedin-statistics/

[23] Niu, E. (2015, May 20). *This Company Has The Best Business Model in Social Media.* The Motley Fool. https://www.fool.com/investing/general/2015/05/20/this-company-has-the-best-business-model-in-social.aspx

[24] Saint Louis University Department of Computer Science (2020, January 16). *CSCI-1080 Intro To CS: World Wide Web.* Saint Louis University. https://cs.slu.edu/~esposito/teaching/1080/webscience/graphs.html

[25] Petrov, C. (2020, February 17). *63 LinkedIn Statistics That Matter in 2020.* Tech Jury. https://techjury.net/stats-about/linkedin/#gref

[26] Smith, K. (2019, June 1). *53 Incredible Facebook Statistics and Facts.* Brandwatch. https://www.brandwatch.com/blog/facebook-statistics/

[27] Facebook. (2019). 2019 4Q Earnings Presentation. Retrieved April 20, 2019 from https://investor.fb.com/financials/default.aspx

[28] Yildirim, P. (2019, February 1) *Social Strategies & LinkedIn* [Powerpoint slides]. Canvas. https://canvas.upenn.edu/courses/1440988/files/folder/Lecture%20Notes?preview=74569382

[29] Cornell University (2018, September 20). *Six Degrees of LinkedIn Separation.* Cornell University. https://blogs.cornell.edu/info2040/2018/09/20/six-degrees-of-linkedin-separation/

[30] Hagiu, A. (2013, December 19). *Strategic Decisions for Multisided Platforms.* MIT Sloan Management Review. https://sloanreview.mit.edu/article/strategic-decisions-for-multisided-platforms/

[31] Team, T. (2015, November 4). *Why Talent Solutions Is The Most Valuable Division For LinkedIn?* Forbes. https://www.forbes.com/sites/greatspeculations/2015/11/04/why-talent-solutions-is-the-most-valuable-division-for-linkedin/#7082e3df7ebc

[32] Florentine, S., & Kapko, M. (2019, April 23). *Why LinkedIn Premium is worth the money.* CIO. https://www.cio.com/article/2877153/why-linkedin-premium-is-worth-the-money.html

[33] Bruns, A., Burgess, J., Mahrt, M., Puschmann, C., & Weller, K. (2014). Twitter and society. *Digital Formations*, *89*. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47764-2

[34] Clement, J. (2019, July 23). *Twitter: number of users worldwide from 2014 to 2020.* Statista. https://www.statista.com/statistics/303681/twitter-users-worldwide/

[35] Perez, S. (2017, November 7). *Twitter officially expands its character count to 280 starting today.* TechCrunch. https://techcrunch.com/2017/11/07/twitter-officially-expands-its-character-count-to-280-starting-today/

[36] Perez, S. (2018, October 30). *Twitter's doubling of character count from 140 to 280 had little impact on length of tweets.* TechCrunch. https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/

[37] Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities.* ff10.1080/23311983.2016.1171458ff. ffhal-01517493f

[38] Suzumura, T., & Watanabe, M. (2013). How Social Network is Evolving? - A Preliminary Study on Billion-scale Twitter Network. *WWW'13 Companion: Proceedings of the 22nd International Conference on World Wide Web,* 531-534. https://doi.org/10.1145/2487788.2487988