

Social network analysis



Ping Yu
Min Hu
Nayeoung Kim

Table of Content:

1. Motivation	3
2. Related Work	3
3. Methods	4
3.1. Collecting Data for Ten Groups for Network Analysis and Power Law distribution.....	4
3.2 Gathering Data for Hundred and Ten Communities for interest distribution.....	4
3.3 Emails and Interviews.....	5
4. Data Analysis	5
4.1 The Number of Videos Uploaded V.S. the Number of Subscribers.....	5
4.1.1 Discussions.....	9
4.2 Power Law Distributions on YouTube.....	9
4.2.1 Discussions.....	11
4.3 Network Analysis.....	11
4.3.1 Friend Networks and Subscriber Network within Groups.....	11
4.3.2 Users to Friend Networks and Users to Subscriber Networks.....	13
4.3.3 Comparison with a Random Network.....	15
4.3.4 Combined Network.....	16
4.3.5 Discussions.....	16
4. 2 Communities and Interests in YouTube.....	17
4.4.1 1st Phase- Methodology.....	17
4.4.2 1st Phase-Findings.....	17
4.4.3 2nd Phase – Methodology.....	19
4.4.4 2nd Phase – Findings.....	20
5. Conclusions	21
References	21
Acknowledgement	22
Appendix:	22

1. Motivation

YouTube is a video sharing website, where users can upload, watch and share videos with others. YouTube, created in 2005, is a relatively new website that has not been studied intensively. Therefore, this project is going to examine the distribution of contribution from users, network structure, as well as how diverse or similar users' interests are.

Four main questions are addressed in this paper:

- We want to examine if the number of videos a user uploads is correlated to the number of people who subscribe to them.
- We like to see whether the distribution of numbers of videos, numbers of subscribers, and numbers of friends follow power law.
- We like to know if users are connected and form networks on YouTube through subscriptions and friends.
- We want to explore if users have diverse or similar interests in the YouTube community.

2. Related Work

Mislove et al (2007) presents a large-scale (11.3 million users, 328 million links) measurement study and analysis of the structure of four popular online social networks: Flickr, YouTube, LiveJournal, and Orkut. They gather data from multiple sites to identify common structural properties of online social networks.

The result showed that the group sizes of these social network sites follow a power-law distribution, in which the vast majorities have only a few users each. However, we are more interested in distribution of the number of videos uploaded, to verify free-riders issues in YouTube. Interestingly, they found all of the networks with the exception of YouTube show that high-degree nodes tend to connect to other high-degree nodes to form a "core" of the network. For our project we may not be able to examine the YouTube community as a whole, but we may look at the structure of sub-community such as friend networks and subscriber networks on YouTube.

Cheng et al. (2007) looks at YouTube.com and the characteristics of its videos. The authors understand that YouTube has millions of videos and try to point out the problems that it's causing like network traffic cost per bandwidth. This paper also looks at small world properties YouTube creates of its users and videos.

For our project, we were able to find some interesting points from this paper that may be helpful. First of all, this paper provides a lot of background information about YouTube and its videos. It briefly mentions that about 58% of users do not have friends. This fact is likely for us to cross while trying to identify networks in YouTube. Also, this paper presents a network with small world properties in terms of videos and their related videos. This might reveal some information on how users find each other and get connected. Also, this paper looks at the data across multiple

time points. However, to keep our project within scope, we are going to look at the data as a whole.

Lerman (2007) studied Digg, a social networking site, to see the role of social networks in filtering and found that users digg stories their friends submit. Also, users do use friends interface to find new interesting stories. Similarly, Lerman (2007) also found that users on Flickr tend to view images produced by good photographers, the views and favorites they receive correlate most strongly with the number of reverse contacts the photographer has. These findings showed that most of people may read similar stories or view images took by certain “good photographers” through social browsing or collaborative filtering. Hence, it may imply that people may share some common interests with others. Our study use slightly different approaches to examine the question. We would like to know when users subscribe to a video if it is more likely to be coincidental or they do share common interests.

3. Methods

3.1. Collecting Data for Ten Groups for Network Analysis and Power Law distribution

Ten out of the twelve categories in the YouTube community, including comedy, people & blogs, pet & animals, entertainment, autos & vehicles, news & politics, music, travel & events, sports, and animation are chosen.

After that, we randomly chose one group from these ten categories¹, and wrote a Perl script to get: 1) Number of videos, friends, and subscribers each member has in each community to perform data analysis 2) Members in each group, and members’ friends and their subscribers to construct friend networks and subscriber networks.

3.2 Gathering Data for Hundred and Ten Communities for interest distribution.

In order to examine how diverse or similar each user is, we initially selected 20 communities, 2 communities each for 1 category. We picked communities that had about 100 to 500 users. We then used a Perl script to crawl all of the users who had favorite videos, the category of their favorite videos and the number of categories for the favorite videos for each category for each user. We then created a gdf file that would show the network of users from 10 communities² connected to each category, to see what users from communities would be interested in. And then, we gathered data from 8 more communities for each category plus 10 more from the Howto & Style category to look at the number one favorite category for each user and the overall percentages from each community.

¹ See Appendix

² See Appendix

3.3 Emails and Interviews

In order to explore reasons why a user subscribes to people or makes friends with others in depth, we sent an email to si.open.all@umich.edu, and several messages to YouTube users we choose to conduct surveys. Also, we talked with two YouTube users in SI and try to understand how users interact with others in the YouTube community.

4. Data Analysis

4.1 The Number of Videos Uploaded V.S. the Number of Subscribers

We obtained the number of videos and the number of subscriber each member has in each community that was chosen from 12 categories. Later on we also got the number of friends they have. After we get these numbers for users in all the ten communities, we aggregated all the data and calculated their correlation coefficient and p-value. Here is the output in excel for regression analysis performed on number of videos and number of subscribers:

Figure: Regression Analysis for Number of Videos and Subscribers

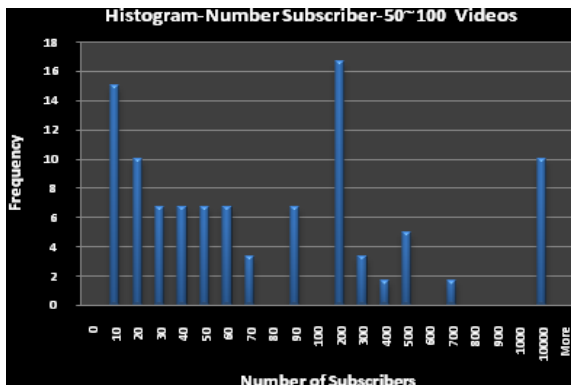
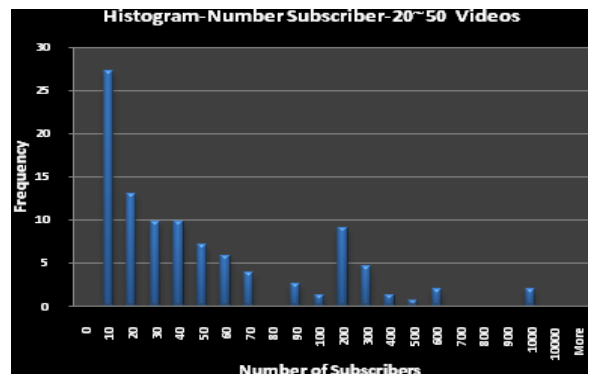
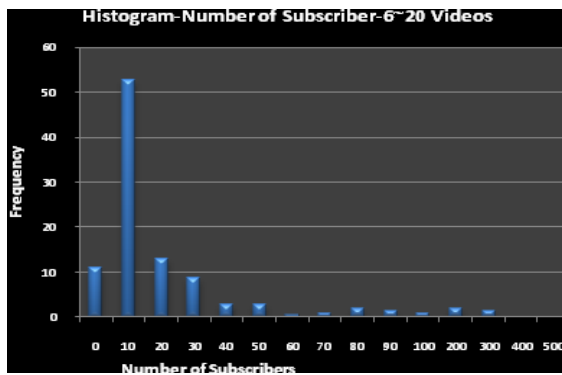
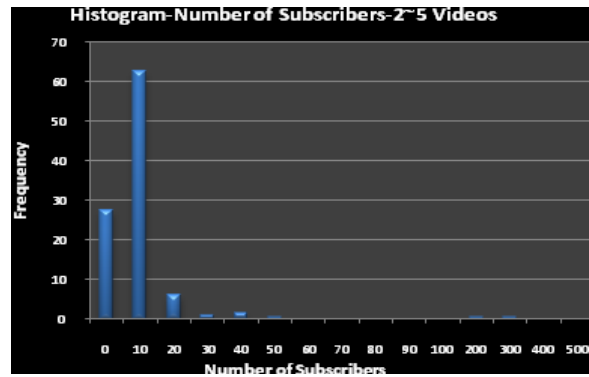
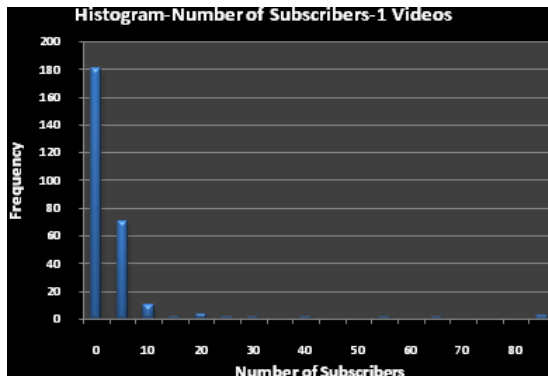
Regression Statistics	
Multiple R	0.312451
R Square	0.097626
Adjusted R Square	0.097271
Standard Error	110.355
Observations	2548

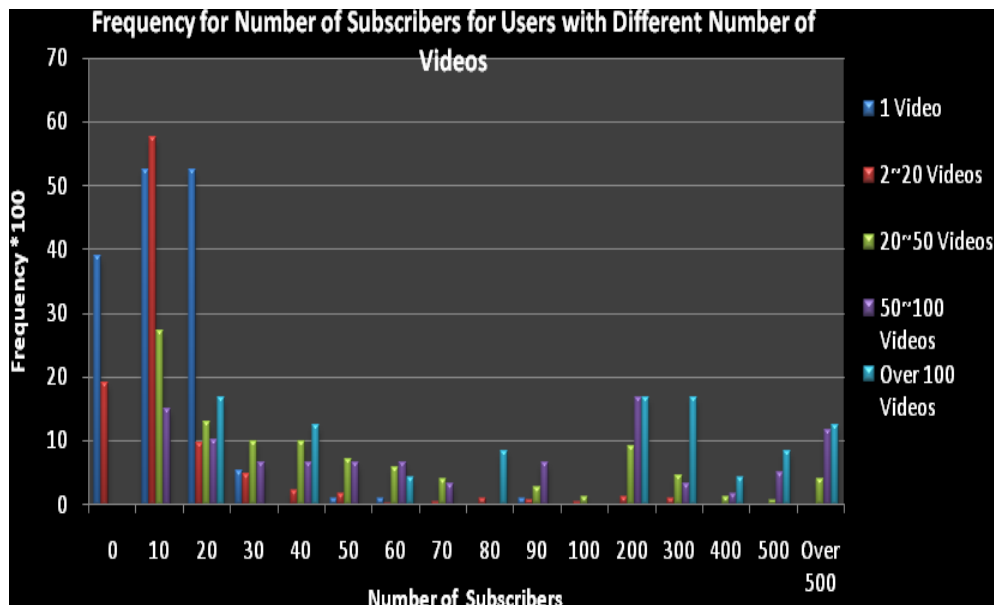
	df	SS	MS	F	Significance F
Regression	1	3354435	3354435	275.445	8.13E-59
Residual	2546	31005769	12178.2		
Total	2547	34360203			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	9.43233	2.275772	4.14467	3.51E-05	4.969777	13.89488	4.969777	13.89488
X Variable 1	1.410459	0.084985	16.5965	8.13E-59	1.243812	1.577106	1.243812	1.577106

The correlation coefficient is 0.312, which means there is a weak correlation. P value is really low. This means we can be confident about the correlation. Thus, there is a significantly weak correlation between number of videos uploaded and number of subscriber a user get.

Besides doing statistical analysis on the dataset, we also looked at distribution of number of subscribers. The following graph shows histograms of number of subscribers based on the number of videos a user have.





Here is the output for regression analysis performed on the number of subscribers and the number of users in excel:

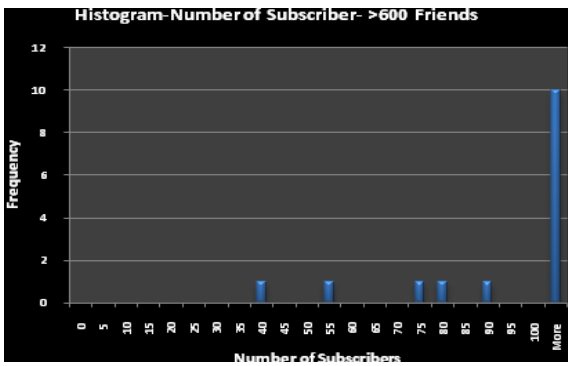
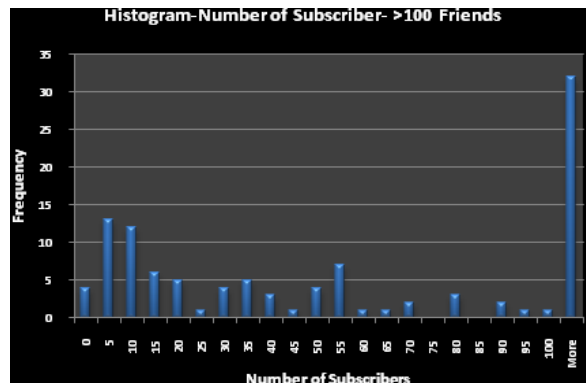
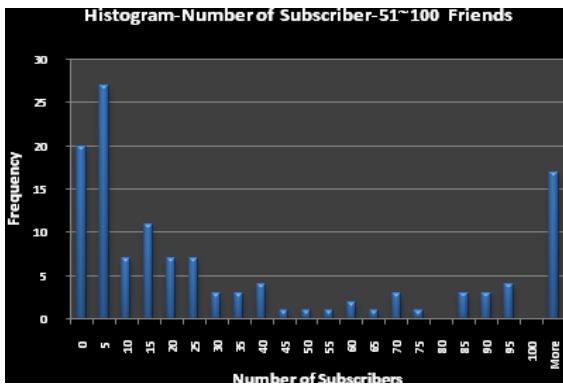
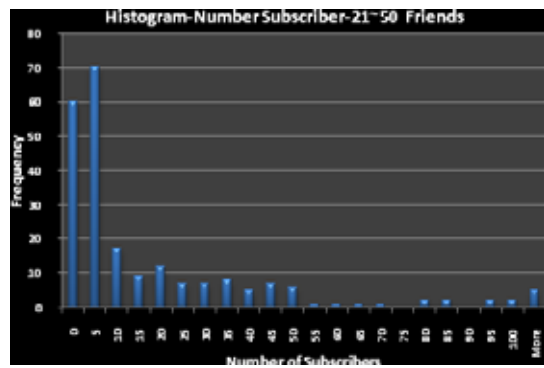
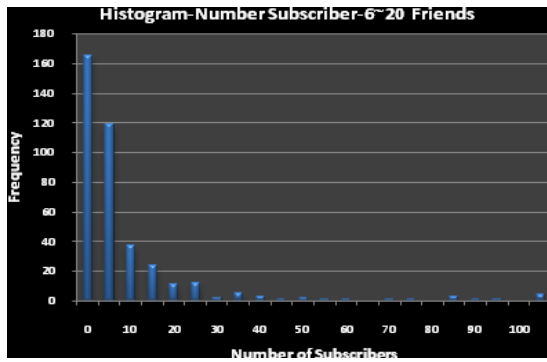
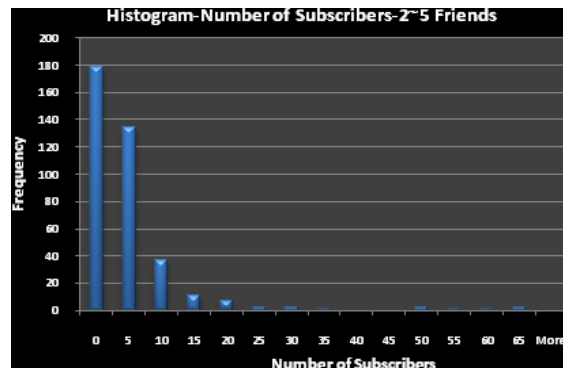
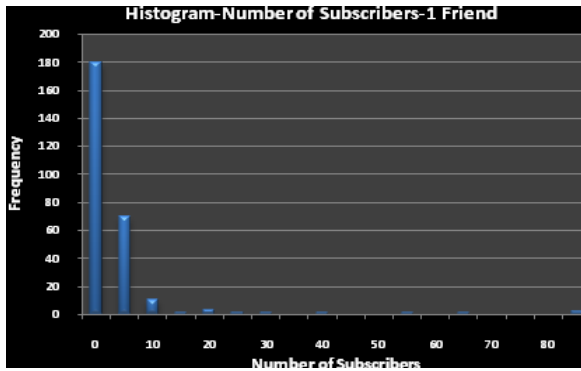
Figure: Regression Analysis for Number of Subscribers and Users

Regression Statistics	
Multiple R	0.550174
R Square	0.302691
Adjusted R Square	0.302231
Standard Error	154.7616
Observations	1518

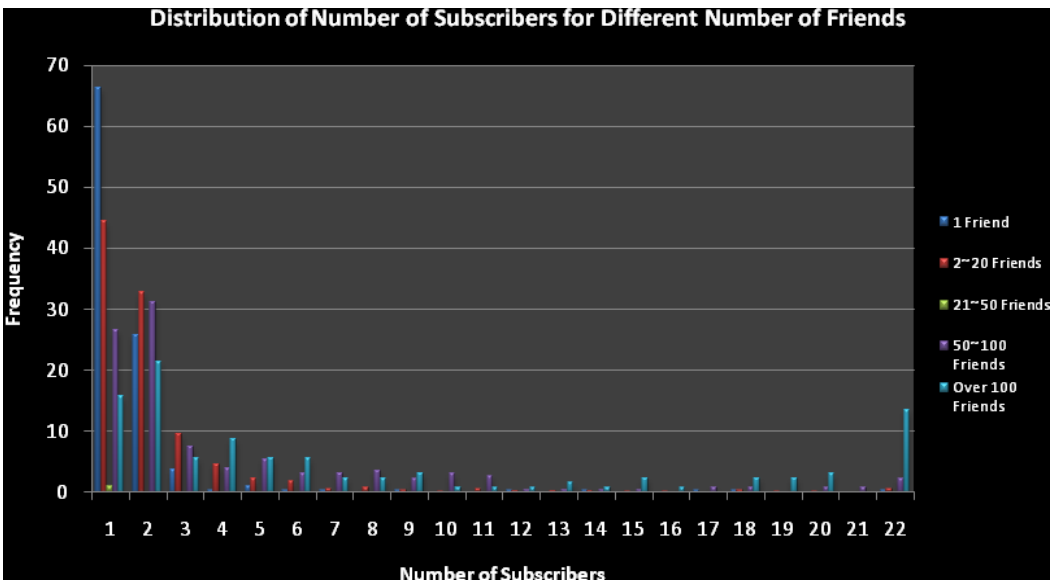
ANOVA					
	df	SS	MS	F	Significance F
Regression	1	15761580	15761580	658.0723	7.70E-121
Residual	1516	36309925	23951.14		
Total	1517	52071505			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	21.01325786	4.064503	5.169945	2.65E-07	13.04061	28.9859	13.04061	28.9859
X Variable 1	0.975425971	0.038024	25.65292	7.70E-121	0.900841	1.050011	0.900841	1.050011

The correlation coefficient here is 0.55, which means there is a weak correlation. The p value here is pretty low. This means there is a significantly weak correlation between these two variables.



Combined data:



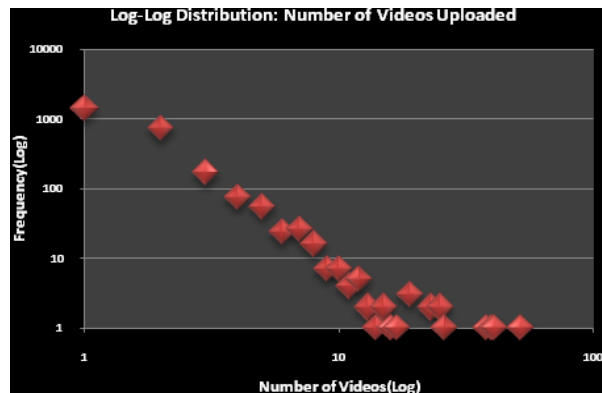
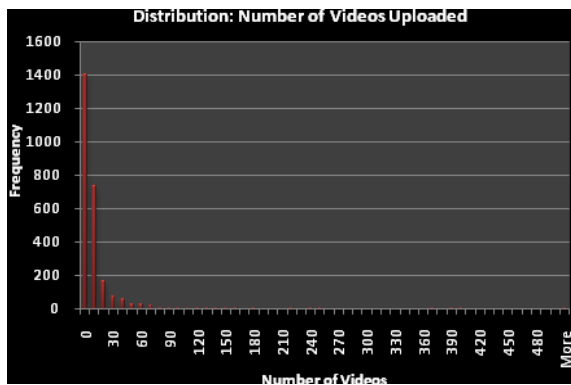
We can observe the same trend as above. The more friends a user has, the larger the probability that he/she would have of receiving a higher number of subscribers.

4.1.1 Discussions

Based on above discussion, there is weak correlation between number of videos and number of subscribers a user has. There is also a correlation between number of friends and number of subscribers a user has. As a user uploads more videos, he/she tends to get more subscribers and more friends. When a user uploads a new video, there is a better chance that his/her subscribers and friends will be watching these videos. Thus, those who get more friends and subscribers tend to have more influence on popularity of the videos. This implies that as a user uploads more videos, he/she tends to be more influential on YouTube.

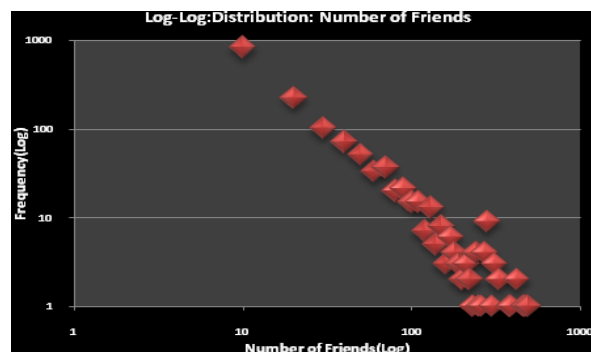
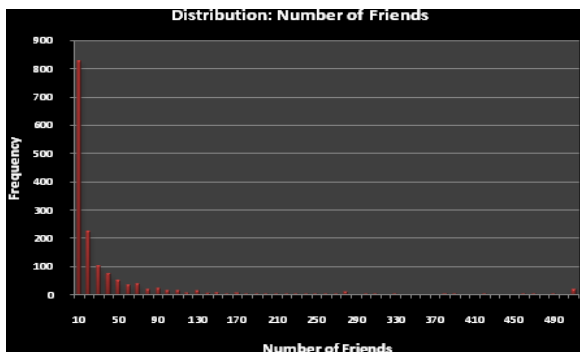
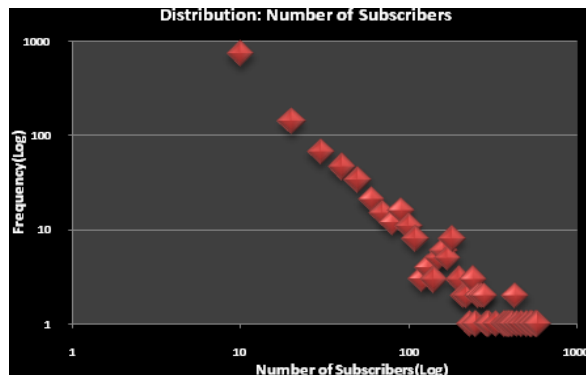
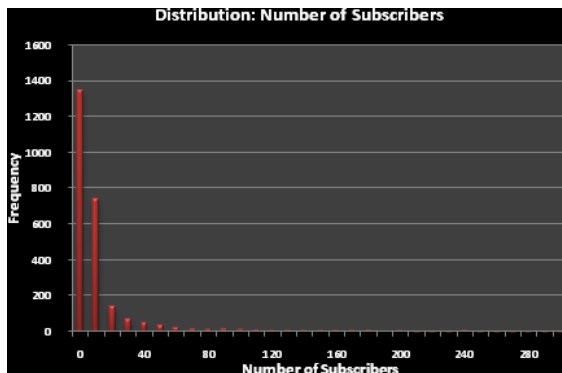
4.2 Power Law Distributions on YouTube

We generated a histogram of distribution in terms of number of videos a user upload. Here is the histogram and the log-log plot:



Based on the above graphs, there is power law distribution on number of videos users upload. Although on the log-log plot, the data is skewed at the tail. However, those numbers represents less than tenth of order of users who have that amount of videos, which is not really representative.

We also tried to fit number of subscribers and friends and it seems those numbers fit to power law, too:



4.2.1 Discussions

From the above graphs, we can find that distribution of numbers of videos, numbers of subscribers and numbers of friends all follow power law. This tells us that the huge amount of videos is actually mostly contributed by a small number of users. Also, most of users are subscribing and making friends with small portion of users. Based on our previous conclusion, that user with more videos tends to get more subscribers and friends. We can probably guess that this small portion of users probably overlap with the users who are contributing most of the videos. And those users tend to be influential. Based on all these observations, it seems that YouTube is actually largely “controlled” by the small portion of users who uploads large amount of videos.

4.3 Network Analysis

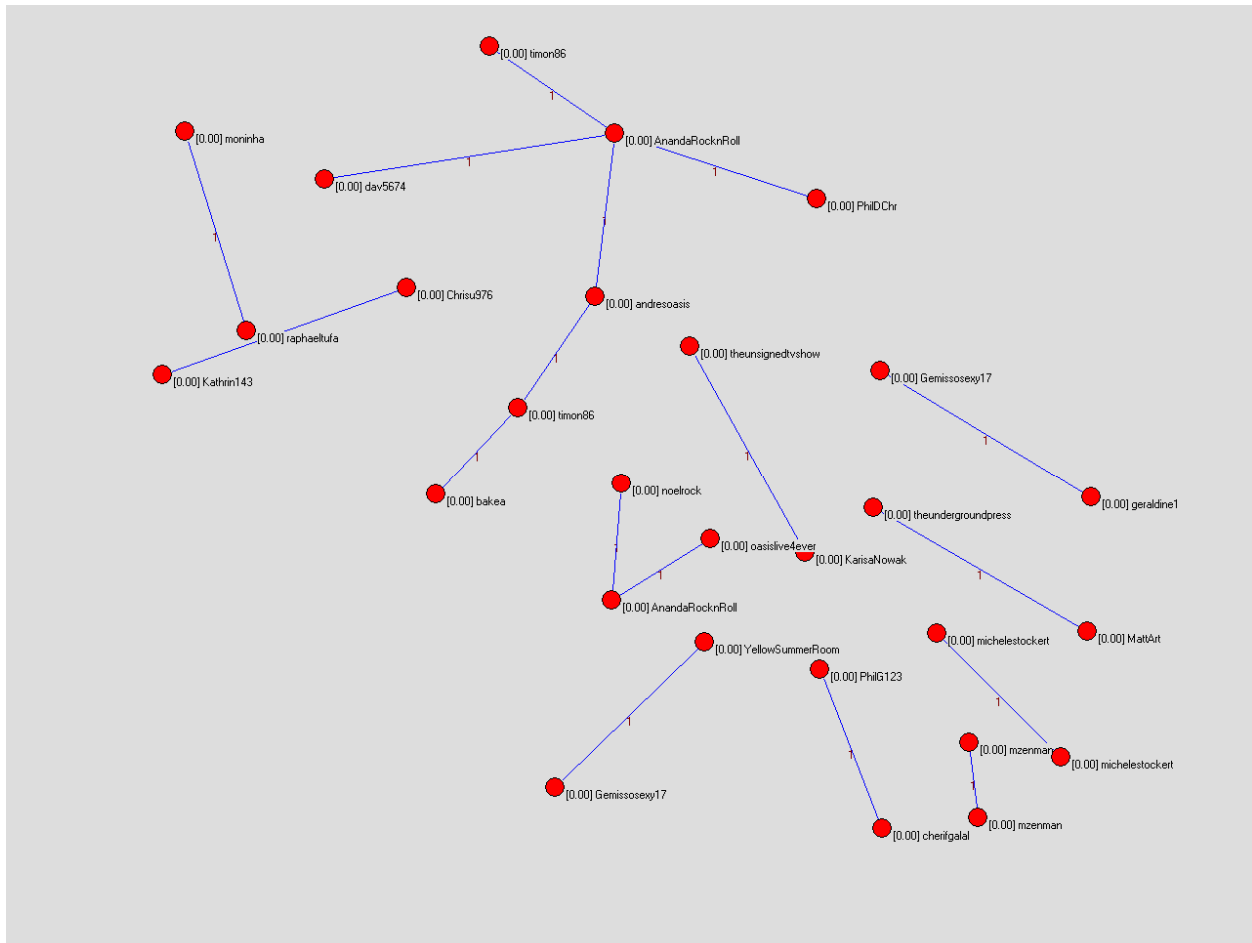
Ten out of the twelve categories in the YouTube community, including comedy, people & blogs, pet & animals, entertainment, autos & vehicles, news & politics, music, travel & events, sports, and animation are chosen. We collected data about members in each group, and members’ friends and their subscribers to construct friend networks and subscriber networks of ten groups we chose. Network analyses were performed to exam the structure of these networks.

4.3.1 Friend Networks and Subscriber Network within Groups

At first, we tried to construct friend networks and subscriber networks within one group. The assumption was that users in one group should be linked tighter than linkages between users in the YouTube community as a whole. However, the results surprisingly showed that members in one group are not neither friends, nor subscribers to each other. We tried to establish group networks for four groups, including animation, comedy, entertainment and music, but obtained similar results, that is, members are not well-connected through friendship or subscription.

As an example, we constructed friendships among music group members. We identified each group members and to see if they make friends with other members in the same group. However, we found that the Clustering Coefficient of the friend network for music group is 0.

Graph: Friend Network within Music Group



From the graph above we can see that although some people in the music group do know each other and make friends within groups, there are no three users who are mutual friends. There are many friend pairs as presented above, but we observed no cliques in this community.

Our result contradicts to earlier research done by Mislove et al. (2007). They found that the average Clustering Coefficient of YouTube groups are 0.34. The contradiction to earlier result drove us to explore why a user subscribe to others, why he/she makes friends with others, and the differences between these two activities. However, after we emailed a survey through si.open.all@umich.edu, we only received one response from SI. The user said that she adds users as friends if they know each other in person or have some interactions, such as commenting, through YouTube. She also subscribes to a user if she thinks their videos are interesting.

In order to find whether members within one group connected to each other through other methods, we keep finding members' subscribers and their friends and establish users to friend network and users to subscriber network.

4.3.2 Users to Friend Networks and Users to Subscriber Networks

To see if members in one group are more connected to each other through their subscribers and friends, we collected data about members' friends and their subscribers, and then construct users to friend networks and users to subscriber network.

4.3.2.1 Users to Subscriber Networks

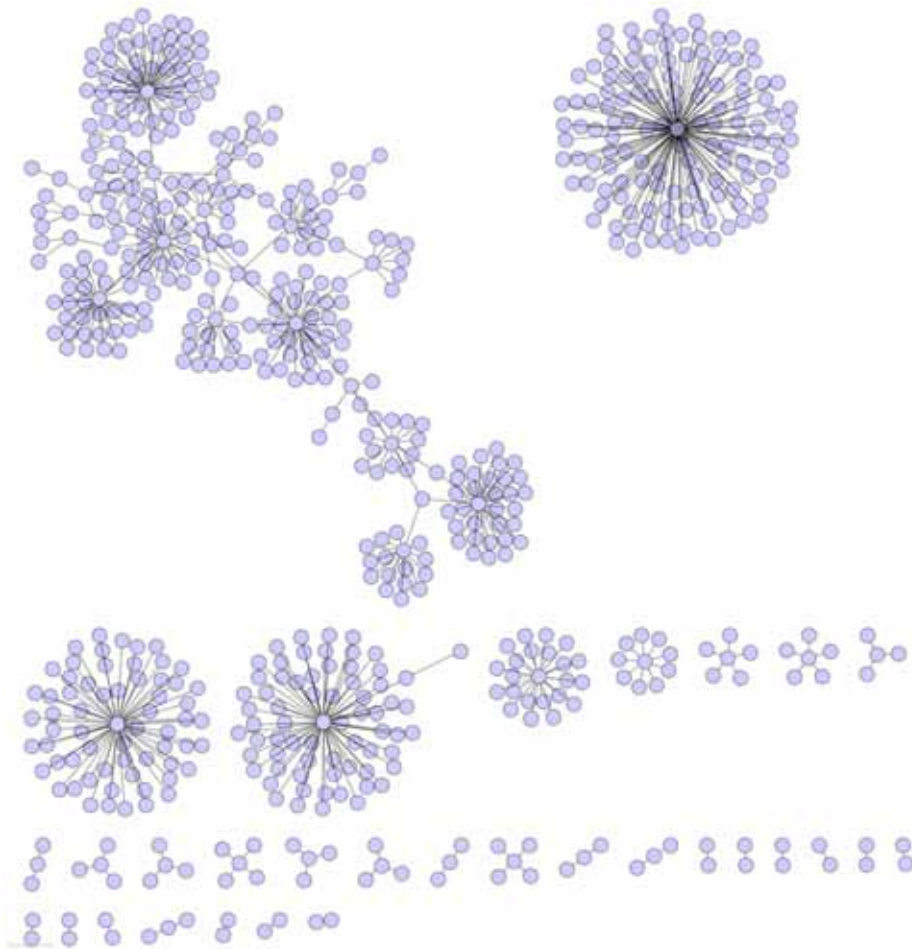
The statistics of users to subscriber networks showed that vertices in these networks are not well-connected. The average betweenness, one of the centrality measures, of users to subscriber networks is 0.31, which seems not too bad. The average value of average shortest path of users to subscriber networks is 4.43, but the number of unreachable pairs is large. Also, the Clustering Coefficient is very low, which is 0.01.

Figure: Statistics of Users to Subscriber Networks

Name	# of Vertices	Avg. shortest path (# of unreachable pairs)	CC	Betweenness
Comedy	2378	4.66(925046)	0.00	0.46
People	2741	4.20(271030)	0.00	0.30
Pet	1141	4.39(96120)	0.00	0.36
Entertainment	7467	4.50(4444546)	0.01	0.19
Auto	658	5.52(257768)	0.00	0.22
News	7153	3.93(1151220)	0.02	0.24
Music	5588	4.78(3767352)	0.00	0.32
Travel	3516	4.38(3762088)	0.01	0.41
Sports	159	3.57(3678)	0.06	0.23
Animation	4172	4.35(627774)	0.01	0.34
Avg.	3497.3	4.43 (1530622)	0.01	0.31

All these data and the graph showed below suggested that there are several central nodes that have high betweenness and are linked by many vertices. Therefore vertices can through them to communicate with some of other vertices in the network and it reduce the average shortest path. However, both the Clustering Coefficient and the number of unreachable pairs suggested that the network is not well-connected. The graph below also demonstrated that vertices in one cluster tend not to connect to each other. Also, there are very few links between clusters.

Graph: Subscribers Network for Auto Group



4.3.2.2 Users to Friend Networks

Vertices in users to friend networks are not well-connected as well. As observed in the subscriber network, the betweenness, one of the centrality measures, of users to friend network is 0.36, which seems not too bad. The average value of average shortest path of users to subscriber networks is 3.94, but there are many unreachable pairs. Also, the Clustering Coefficient is very low, which is 0.02. Again, both the Clustering Coefficient and the number of unreachable pairs suggested that vertices in these networks are not well-connected.

Name	# of Vertices	Avg. shortest path (# of unreachable pairs)	CC	Betweenness
Comedy	3937	4.25(715114)	0.01	0.36
People	2112	4.19(167936)	0.00	0.30
Pet	734	3.49(97452)	0.00	0.48
Entertainment	11234	3.94(2035948)	0.04	0.13
Auto	723	3.24(183314)	0.00	0.61
News	4165	3.88(984776)	0.02	0.22
Music	4355	4.18(1450782)	0.00	0.30
Travel	2190	4.18(1174692)	0.00	0.38
Sports	150	3.54(3678)	0.09	0.51
Animation	5499	4.52(1239596)	0.02	0.27
Avg.	3509.9	3.94 (805329)	0.02	0.36

We also noticed that there are no major differences in the friend network versus the subscriber network. Although in reality the subscriber network is directed, as a user can subscribe to whoever he likes. And friend network is undirected; the user needs confirmation from the user he wants to connect to before they become friends. It takes more effort to get a friend than getting a subscription. However, we do not find many differences between the two networks. Part of the reason might be that since both networks are not well-connected, even there is difference; it would be hard to observe through the parameters. Another guess is that people do not have uniform pattern of obtaining friends versus obtaining a subscription. The decision is on an individual base rather than a social activity. In other words, YouTube does not clearly differentiate the role of a friend versus a subscriber, which leads to random decisions by users.

4.3.3 Comparison with a Random Network

In order to examine how these users to friend and subscriber networks perform, we compared them with random networks. One large group with over 7,400 vertices and a small group with 658 vertices were chosen to do the comparison.

After comparing with a random network, the results demonstrated the real networks have shorter average shortest path and higher betweenness than random networks. However, the Clustering Coefficient of real networks is similar to random networks, which is almost 0(see Figure). The same thing also happens to friend network. The real friend networks have higher betweenness and average shortest path than random network, but the Clustering Coefficient is the same as a random network. This confirms previous discussion that although some people do know each other through subscription and friends, Youtube groups are poorly connected.

Figure: Comparisons between Real Subscriber and Random Networks

Entertainment	# of Vertices	Avg. shortest path (# of unreachable pairs)	CC	Betweenness
Real Network	7467	4.50	0.01	0.19
Random Network	7467	8.29	0.00	0.07
Auto	# of Vertices	Avg. shortest path (# of unreachable pairs)	CC	Betweenness
Real Network	658	5.52	0.00	0.22
Random Network	658	8.60	0.00	0.10

4.3.4 Combined Network

Since vertices in users to subscriber and friends networks are not well-connected to each other, we tried different methods to connect vertices and make these network denser. In the process of collecting data, we found that some vertices have common subscribers, and wonder what will happen if these vertices that have common subscribers are linked. A friend network for music group was selected for the experiment. After vertices were connected if they have the same subscribers, the Clustering Coefficient increases dramatically from 0.07 to 0.76. However, it is unclear if it is reasonable to connect vertices that have common subscribers in a friend network and further study is needed.

Based on our previous discussion that user might not have strict differentiation between friends and subscribers. It is possible that if we combine the friend network and subscriber network, we can get a more well-connected network. Since we only obtained users' friends and subscribers, not their friends' friends or subscribers. Further data collection and research is needed to prove this.

4.3.5 Discussions

After the network analysis, compared to Facebook, MySpace, and Livejournal etc., YouTube is not really a "social networking" site per say. Members in groups do not socialize with other members through friendships and subscriptions as much as users in other sites do. And they don't have a lot of common connectors. Users make connections without really knowing each other, but may base on common interest.

This leads us to start thinking that instead of socializing with people, users are more interested in the video content and they are linked through videos. After two in-depth short interviews with YouTube users, we found that both of them enjoy commenting on videos and read comments about videos made by other users. It may indicate that users are linked through videos, not friendships or subscriptions. This can be studied by looking at users' comments or viewing history of specific videos.

Secondly, as discussed above, a combined network of friends and subscribers might increase the

network performance. If we further explore the network of users, friends, friends' friends, subscribers and subscribers' subscribers. We might be able to find a better connected network or prove our guess that people do not make networks through people, but more through common interest in videos.

4. 2 Communities and Interests in YouTube

As mentioned in the motivation, one of the questions we had in general was whether or not people have a collective interest (in this case interest in different categories of media), or are people's interests indeed too diverse these days. From this YouTube project we wanted to see if this holds true for users in communities in YouTube.

4.4.1 1st Phase- Methodology

In the case that users have similar interests with other users in the same community, we wanted to see specifically what those interests might be. We also wanted to see what would some of the popular interests might be for communities as a whole. Each user in YouTube may or may not have videos labeled as the user's favorite video. Each video is tagged as categories (Auto & Vehicles, Comedy, Entertainment, Film & Animation, Howto & Style, Music, News & Politics, People & Blogs, Pets & Animals, Sports and Travel & Events) which are given by YouTube. We decided to define each user's 'interest' as category of their favorite videos.

To accomplish all of these tasks, we initially selected 20 communities, 2 communities each for 1 category. We picked communities that had about 100 to 500 users. We then got all of the users who had favorite videos, the category of their favorite videos and the number of categories for the favorite videos for each category for each user. We then created a gdf file that would show the network of users from 10 communities connected to each category, to see what users from communities would be interested in. Initially when we gathered data from Howto & Style, we didn't see a strong interest in any one of the categories so we did not collect any more data from that community. However, we did not exclude the category Howto & Style from other communities' user data, to make sure that Howto & Style is not worth looking at.

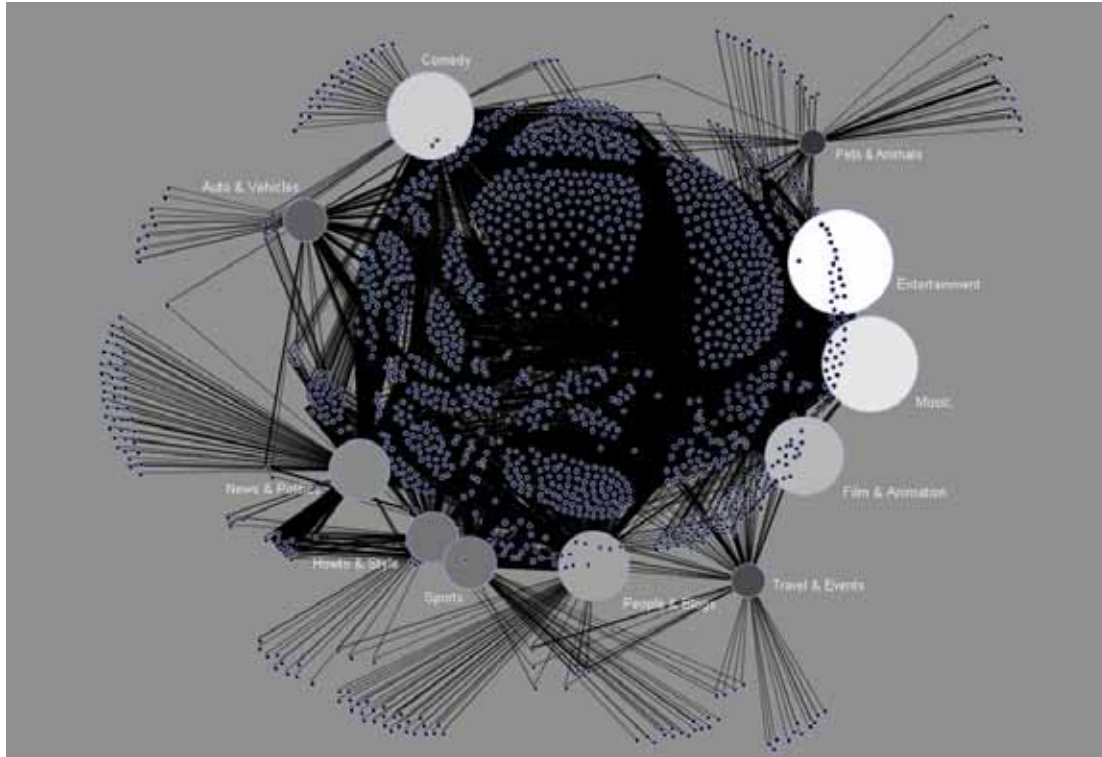
We put all our data in a whole network consists of nodes of users and categories. With each user pointing to its most highly rated interest. After we took a look at the network created in GUESS, and looked at the percentages of users from each category interested in music and Entertainment, we noticed dominate interest in these two categories in almost every community we observed. As it became obvious that music and entertainment are the most popular categories on YouTube, we thought it would be interesting to see how the percentages of interests for each community would shift once we take out Music and Entertainment data from each community. We still left Music for the 2 communities in the category Music, and leave Entertainment data for the 2 communities in the category Entertainment. Below we will talk about our findings.

4.4.2 1st Phase-Findings

To take a look at the overall interests from each of the 20 communities, we created a network

linking users to categories. Below is the figure which illustrates this point.

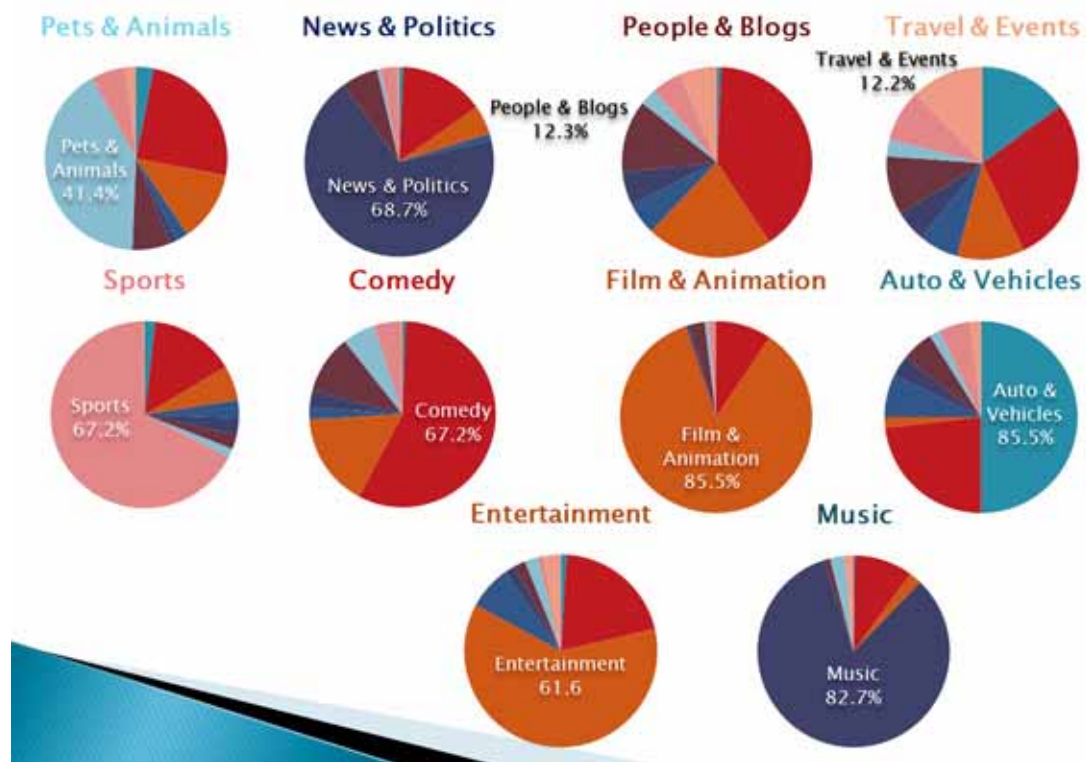
Figure: Users to Interests from 10 communities



In this network, the smallest nodes represent the users. The larger nodes in varying sizes represent the categories. The sizes and the color of the categories vary according to their degree. You can see easily from the network that Entertainment and Music are some of the bigger categories. Comedy and Film & Animation are second biggest categories. All other categories seemed to be receiving less attention from users.

Related to the history that how YouTube became popular. It became really obvious why music and entertainment are the two biggest categories in terms of interest. YouTube is a community that has mostly videos that grew up based on videos of music and entertainment. The interesting thing from this network is that the category Howto&Style, which we didn't see much interesting results from, seemed to receive more attention than some categories like Auto & Vehicles and Pets&Animals.

While gathering the data, we also noticed that for many of the communities, their own community category is at least in the top third most popular category. Thus, based on the network, we decided to take out Music and Entertainment data from all communities except for the Music and Entertainment Communities. Then for each community, we looked at the percentages of users and their number one interest. Since we had two communities for each category, we averaged the data for each category. The below pie charts illustrate our point.

Figure: Distribution of Interest per Category

Except for the communities from People & Blogs and Travel & Events, without the data from Music and Entertainment, the most popular category for each both communities combined for each category of Interest are the community categories themselves. It's important to acknowledge the large percentage (mostly over 60%) of users considering the categories as their number one interests, since each community is divided by 11 different categories of interests.

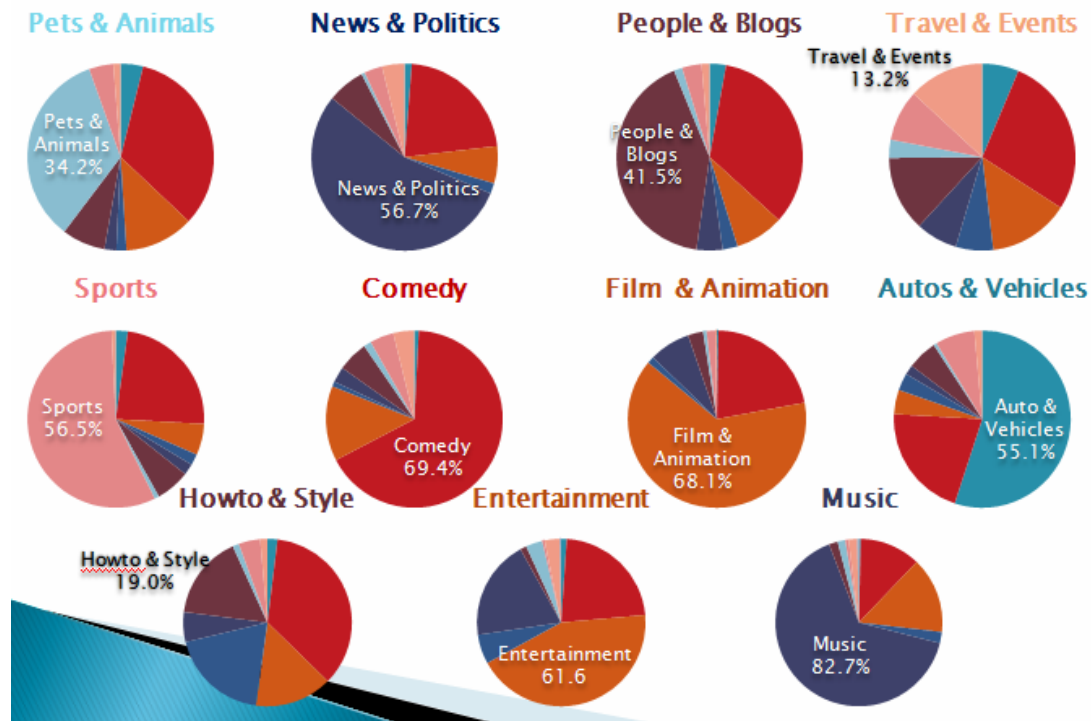
Now we can safely say that overall, the most popular categories for users from all of the communities are Entertainment and Music. However, we believe that it is necessary to look at more communities to look at each community individually. While there are more than 100 users in each community, we weren't sure if two communities are enough to represent all communities in each category. Thus we move on to phase two of this section's methodology and findings.

4.4.3 2nd Phase – Methodology

Now that we can safely say that overall the main interests for all communities are Music and Entertainment, and we can hypothesize that for each community, most of the users are most interested in their community's category. After reaching this hypothesis, we decided to gather some more data to see if this is indeed true. Thus, like we did for phase one of data gathering, we gathered data from 8 more communities for each category. As we did before, for each new community we looked at the number one favorite category for each user. Then we looked at the overall percentages from each community. As you can see below, we represented the aggregated findings in a pie chart as we did before.

4.4.4 2nd Phase – Findings

Figure 3 Distribution of Interest per Category V. 2



As you can see from the above pie charts, after gathering 8 more communities for each group and 10 for Howto & Style community, overall, without the category music and entertainment included, the most popular category for each community is the community category themselves. This holds true for all communities except for communities in Howto & Style and Travel & Events. When you compare to the pie charts in phase 1, most of the percentages for favorite category has gone down, except for People & Blogs and Travel & Events. The percentage after gather 8 more communities have gone up substantially. For Travel & Events, the percentage has gone up and is now the number two most popular category for People & Blogs communities, but it's still not the communities' most favorite category. However, by this we can conclude for now that if we exclude Music and Entertainment the most popular category for each community is the community's own category.

Thus in conclusion to these findings perhaps interests of certain communities still remain somewhat unified. At least this seems to hold true for YouTube communities. It's nice to see that users in YouTube can relate to each other in this way (Since YouTube users don't seem to be very social with each other in other aspects). For further research, it would be interesting to see whether there is a pattern of affinity of interests. For example, a community that's interested in Travel & Events may also tend to have more users that are also interested in Autos & Vehicles. It would be an interesting spin-off from what we looked at for this project.

5. Conclusions

Based on four main questions, we gained some interesting insights into the YouTube community from the data and network analysis. Also these findings may be good for future research and study.

- **More movies you upload, more influential you are.** When a user uploads a new video, there is a better chance that his subscribers and friends will be watching these videos. Thus, those who get more friends and subscribers tend to have more influence on the popularity of the videos. And those who upload more videos tend to get more friends and subscribers. This implies that as a user uploads more videos, he tends to be more influential on YouTube.
- **Power Law Dominates YouTube.** We find that the distribution of numbers of videos, numbers of subscribers and numbers of friends all follow power law. This tells us that the huge amount of videos is actually mostly contributed by a small number of users. Also, most of users are subscribing and making friends with a small portion of users, which tends to be users who upload a lot of videos.
- **YouTube is not “Social” in terms of friendships and subscriptions.** Members in one group tend not to make friends with other members and not to subscribe to others in the same group. Moreover, members’ subscribers and their friends tend not to link to each other. But through interviews, we found that users enjoy commenting on videos and reading comments about videos made by other users. It may indicate that users are linked through videos, not friendships or subscriptions. This can be studied by looking at users’ comments or viewing the history of specific videos.
- **A combined network of friends and subscribers might increase the connection between users.** If we further explore the network of users, friends, friends’ friends, subscribers and subscribers’ subscribers. We might be able to find a better connected network or prove our guess that people do not make networks through people, but more through common interest in videos.
- **Interests of communities in YouTube tend to be unified.** According to the data we gathered, Entertainment and Music are some of the bigger categories. Comedy and Film & Animation are the second biggest categories. If we exclude Music and Entertainment, the most popular category for each community is the community’s own category. For further research, it would be interesting to see whether there is a pattern of affinity of interests.

References

1. Kristina Lerman, “Social Networks and Social Information Filtering on Digg”, ICWSM’

- 2007 Boulder, Colorado, USA
2. Kristina Lerman, Laurie A. Jones, “Social Browsing on Flickr” ICWSM’ 2007 Boulder, Colorado, USA
 3. Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi Peter Druschel, Bobby Bhattacharjee. “Measurement and Analysis of Online Social Networks” IMC’07, October 24-26, 2007, San Diego, California, USA.
 4. Xu Cheng, Cameron Dale and Jiangchuan Liu. “Understanding the Characteristics of Internet Short Video Sharing: YouTube as a Case Study.” 2007

Acknowledgement

We would like to thank Professor Lada Adamic for providing scripts and invaluable suggestions to our project.

Appendix:

Researched Communities

pets: <http://youtube.com/group/acquaPhyles>
 music: <http://youtube.com/group/Oasisvideos>
 entertainment: <http://youtube.com/group/tvb>
 sports: <http://youtube.com/group/skateboardingdenmark>
 news: <http://youtube.com/group/ronpaul4president>
 auto: <http://youtube.com/group/name=FJCruiser>
 Comedy: <http://youtube.com/group/name=buffycaz>
 People: <http://youtube.com/group/NewYorkCityscapes>
 travel: <http://youtube.com/group/scuba>
 animation: <http://youtube.com/group/digimon>

Interest Groups

Autos & vehicles
<http://youtube.com/group/autoweek>
<http://youtube.com/group/f1history>
<http://youtube.com/group/flight>
<http://youtube.com/group/newmini>
<http://youtube.com/group/REVVolution>
<http://youtube.com/group/scooterclub>
<http://youtube.com/group/speedfreaks>
<http://youtube.com/group/topgearnoshit>
<http://youtube.com/group/5xracingaautofilms>
<http://youtube.com/group/FJCruiser>

Comedy

<http://youtube.com/group/pazankachok>
<http://youtube.com/group/senseofhumor>
<http://youtube.com/group/someting>
<http://youtube.com/group/southparkWOW>
<http://youtube.com/group/standuptoday>
<http://youtube.com/group/buffycaz>
<http://youtube.com/group/conan>
<http://youtube.com/group/kalkofe>
<http://youtube.com/group/MTMOAM>
<http://youtube.com/group/Smosh>

Entertainment

<http://youtube.com/group/CHOCOLATEOBSESSION>
<http://youtube.com/group/director>
<http://youtube.com/group/friendsofsuna>
<http://youtube.com/group/magiczone>
<http://youtube.com/group/misshorrorfest2007>
<http://youtube.com/group/morepizzahut>
<http://youtube.com/group/potter16>
<http://youtube.com/group/adgoodness>
<http://youtube.com/group/tvb>
<http://youtube.com/group/originalshortmovie>

Film&Animation

<http://youtube.com/group/PokeFansgroup>
<http://youtube.com/group/Cardcaptor>
<http://youtube.com/group/digimon>
<http://youtube.com/group/If>
<http://youtube.com/group/indiefilmakers>
<http://youtube.com/group/inuyasha911>
<http://youtube.com/group/matzoball101>
<http://youtube.com/group/motions>
<http://youtube.com/group/sauskaku>
<http://youtube.com/group/bakin14group1>

Howto&Style

<http://youtube.com/group/cgsbioblog>
<http://youtube.com/group/evolving>
<http://youtube.com/group/jlhc>
<http://youtube.com/group/LoveofMakeup>
<http://youtube.com/group/nextmillionare>
<http://youtube.com/group/slashdotreview>
<http://youtube.com/group/TheArtistStudios>
<http://youtube.com/group/VintageSesameStreet>
<http://youtube.com/group/flockstars>
<http://youtube.com/group/atheistscumunited>

Music

<http://youtube.com/group/80sMetal>
<http://youtube.com/group/hiphopblogs>
<http://youtube.com/group/hod>
<http://youtube.com/group/musiclove>
<http://youtube.com/group/musicvideoindustry>
<http://youtube.com/group/plantpage>
<http://youtube.com/group/poptastic>
<http://youtube.com/group/progrock>
<http://youtube.com/group/Oasisvideos>
<http://youtube.com/group/livemusic>

News&Politics

<http://youtube.com/group/GNN>
<http://youtube.com/group/ltradio>
<http://youtube.com/group/oetiraq>
<http://youtube.com/group/rechwing>
<http://youtube.com/group/antiwar>
<http://youtube.com/group/ConservativeCorner>
<http://youtube.com/group/europeanresistance>
<http://youtube.com/group/ronpaul4president>
<http://youtube.com/group/youtubedemocrats>
<http://youtube.com/group/YTAIA>

People & Blogs

<http://youtube.com/group/Rainclips>
<http://youtube.com/group/ashleyssummerfun>
<http://youtube.com/group/khalifah>
<http://youtube.com/group/makeoverhaircut>
<http://youtube.com/group/MatureWivesofKansas>
<http://youtube.com/group/nornnafans>
<http://youtube.com/group/Prettybrasandpanties>
<http://youtube.com/group/LIBRAREO>
<http://youtube.com/group/UnitedMuslims>
<http://youtube.com/group/Stephenfry>

Pets & Animals

<http://youtube.com/group/aqualife>
<http://youtube.com/group/cat>
<http://youtube.com/group/catslove>
<http://youtube.com/group/horses>
<http://youtube.com/group/petsrus>
<http://youtube.com/group/ratclub>
<http://youtube.com/group/ridinggirl>
<http://youtube.com/group/snowleopardcontest>

<http://youtube.com/group/acquaPhyles>
<http://youtube.com/group/bunnyaddict>

Sports

<http://youtube.com/group/bikestuff>
<http://youtube.com/group/hincharealmadrid>
<http://youtube.com/group/surf>
<http://youtube.com/group/TallWomenVidoes>
<http://youtube.com/group/WrestlingClub>
<http://youtube.com/group/golfinteractive>
<http://youtube.com/group/pumas>
<http://youtube.com/group/skateboardingdenmark>
<http://youtube.com/group/sportcouk>
<http://youtube.com/group/martialartist>

Travel & Events

<http://youtube.com/group/Airlines>
<http://youtube.com/group/Australia1>
<http://youtube.com/group/Bangkok>
<http://youtube.com/group/BeautyOfNature>
<http://youtube.com/group/Colorado>
<http://youtube.com/group/England>
<http://youtube.com/group/GrupulRomanilor>
<http://youtube.com/group/indiatube>
<http://youtube.com/group/jonluc>
<http://youtube.com/group/railfans>