INFO-623-001 – Social Network Analysis

Social Network Analysis of Meetup Network

By

Suresh Athanti (Sa3663@drexel.edu)

Abstract:

Social Network refers to the articulation of a social relationship, achieved, among individuals, families, households, villages, communities, regions, and so on. Each of them can play dual roles acting both as a node of a social network as well as a social actor. There are lot of platforms for people organizing and attending events and social functions. meetup.com is a website for people organizing and attending regular or semi-regular events (meetups). The relationships amongst users, who goes to what meetups are a social network. A community in a social network refers to a group of people who are more tightly interconnected than the overall network. Detecting tight communities is one of the main objectives of the social network. In this project we are going to analyze, social gatherings and most influential meetups. We are using network measures and other techniques in the networks to find the communities and influential network in the meetup network.

Introduction:

Meetup.com is a service used to organize online/offline event groups that host in-person events for people with similar interests. Meetup users have to physically engage with others to participate in the events and interact among themselves. Social links among the users are formed through repetitive event participation in real life. Users are connected by a social link in a group if they are participated in some common events in that group.

Nashville meetups are one such meetups organized through the meetup.com. Nashville meetups are a platform for different groups (tech, social, sports, arts, educational, network, music and writing etc.,) come together to organize meetings and help people to grow in the community. These Meetup exists to help people grow into the people they want to be and live happy and fulfilling lives by finding the communities that will help them thrive.

We have lot of groups in the Nashville community which forms network of people to analyze and find the top groups and communities in the network groups formed by the Nashville. Below I the link to get the Nashville meetup groups: https://www.meetup.com/cities/us/tn/nashville/

Related Work:

Meetup.com is an online platform which facilitates the organizing meetups for different interests, groups etc., Meetup concentrates on specific topic like sports, technology, music, art or language. This paper concentrates on constructing membership network for a given location. Here, Dublin’s meetup communities and groups are analyzed using social network methodologies.

Most interesting topic in the social network analysis is the community detection. This Paper deals with Louvain-based dynamic community detection algorithm which is derived from the previous network evaluation concepts. This algorithm develops a compressed graph in which the nodes represent the detected communities. This algorithm not only constructs the graph but also detects the communities through the integration of Louvain algorithm.

statistical properties of networked systems such as social networks are the focus in the recent years. small-world property, power-law degree distributions, and network transitivity are the some of the common properties in the network analysis. This paper deals with another such property community structure, in which the network nodes are tightly joined in a knit group with some loose connections. To detect such communities using the centrality to find the community boundaries Girvan and Newman proposed algorithm.

The work on detecting communities over large networks is a popular interest in recent times. One such work on detecting large communities over network of items for sales of a retail store is analyzed by using hierarchical agglomeration algorithm for detecting community structure. This algorithm is applied on the website of a large online retailer. If the buyer purchases the same item frequently then they are linked to form a network. The work in this paper shows that the algorithm can extract meaningful communities from this network, revealing large-scale patterns present in the purchasing habits of customers.

Problem Statement/Research Questions:

The project is focused on analyzing the meetup network data and find the most influential network groups and finding communities. We work through the basics of graph theory and extract meaningful insights about the meetup groups. Here, the aim is to understand the analysis and get answers to the following research questions:

- who are the most influenced people in the network? People influencing the transfer of information in the network?
- Find out the most central groups/members in the datasets?
- Which people are having best performance in the network communities?
- Detecting top performance groups and tight communities in the Network.
- Analyze and detect the communities based on the largest group in the group data using Louvain algorithm.

Dataset:

The source of dataset for my analysis is from Kaggle. This dataset was primarily generated for the network analysis with Networkx. The dataset contains below information. I have pulled out required datasets for the necessary analysis.

We have two datasets which contains their network data and meta-data for social analysis.

Member edges - Edge list for constructing a member-to-member graph. Weights represent shared group membership.
It contains:
Member1 – first person in the network
Member2 – second person in the network
Weight – Weights represent shared group membership

Number of nodes is: 11372
Number of edges is: 1176368

Group edges - Edge list for constructing a group-to-group graph. Weights represent shared members between groups.
It contains:
Group1 and group2 forms edge list for group to group graph.
Number of nodes is: 456
Number of edges is: 6692

meta- members - Information for each member, including name and location. member_id serves as index.
It contains: (member_id, name, hometown, city and state)
The number of people in the dataset are: 24591

meta-groups - Information for each group, including name and category. group_id serves as index.
It contains: (group_id, group name, category name, group_urlname, num-numbers)

num-numbers: It represents number of people in each group.
The number of groups in the dataset are: 602

Link for the Dataset: https://www.kaggle.com/stkbailey/nashville-meetup

Dataset Analysis:

Here, for the dataset analysis we will use all the network files listed in the datasets part. For analyzing the dataset, we have used different components of the network. Further analysis is given as follows:
Member Dataset Analysis:

Network connectivity:

Network connectivity can be determined by two factors; The direct/un-directed connection form one actor to another actor can be known from adjacency. Reachability of a network determines if two actors are connected or not by direct/indirect paths.

The connectivity of the network is tested using is_connected(G) networkx function.
nx.is_connected(G) = True (member dataset)
nx.is_connected(G) = True (Group dataset)

```
G = nx.path_graph(member_edges)
print(nx.is_connected(G))
```

```
True
```

Network Visualization:

As the dataset very huge, we have taken a subset of the dataset data and presented the visualization. The subset of the dataset contains 8000 edges and 4866 nodes.

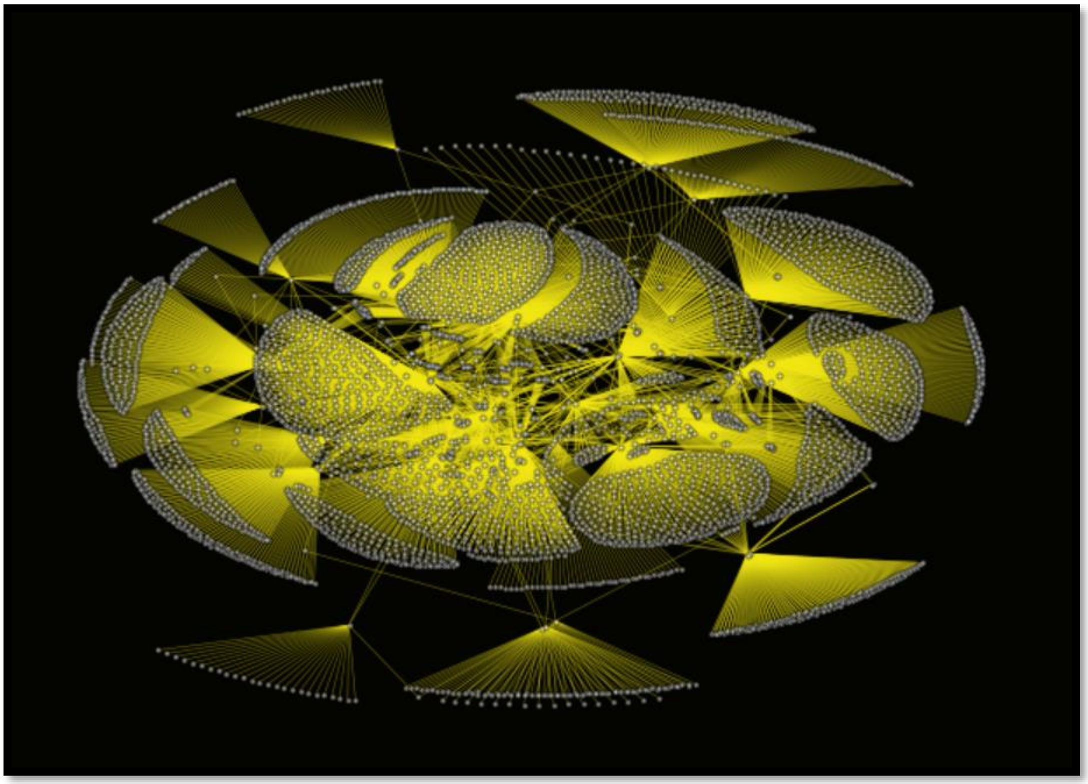

Figure 1. Network visualization

Centrality Measures:

In social network, Centrality indicates the most important vertices/node within a graph. We have different centrality measures. Each measure has its own definition of importance.

Degree Centrality:

Degree centrality is the simplest measure of node connectivity. we consider people with many connections to be important in the network. Degree centrality uses the same notion into a measure of network analysis. Degree centrality measure determines nodes with more connections should have high centrality.

The degree centrality C_d for node v_i in an undirected graph is

$$\mathbf{C_d(v_i) = d_i}$$

In directed graphs, we can either use the in-degree, the out-degree, or the combination as the degree centrality value:

$$C_d(v_i) = d^{in}_i + d^{out}_i$$

[195657825, 204669023, 6160486, 205193250, 25034832, 209453152, 234684445, 226754592, 115965992, 182190122]

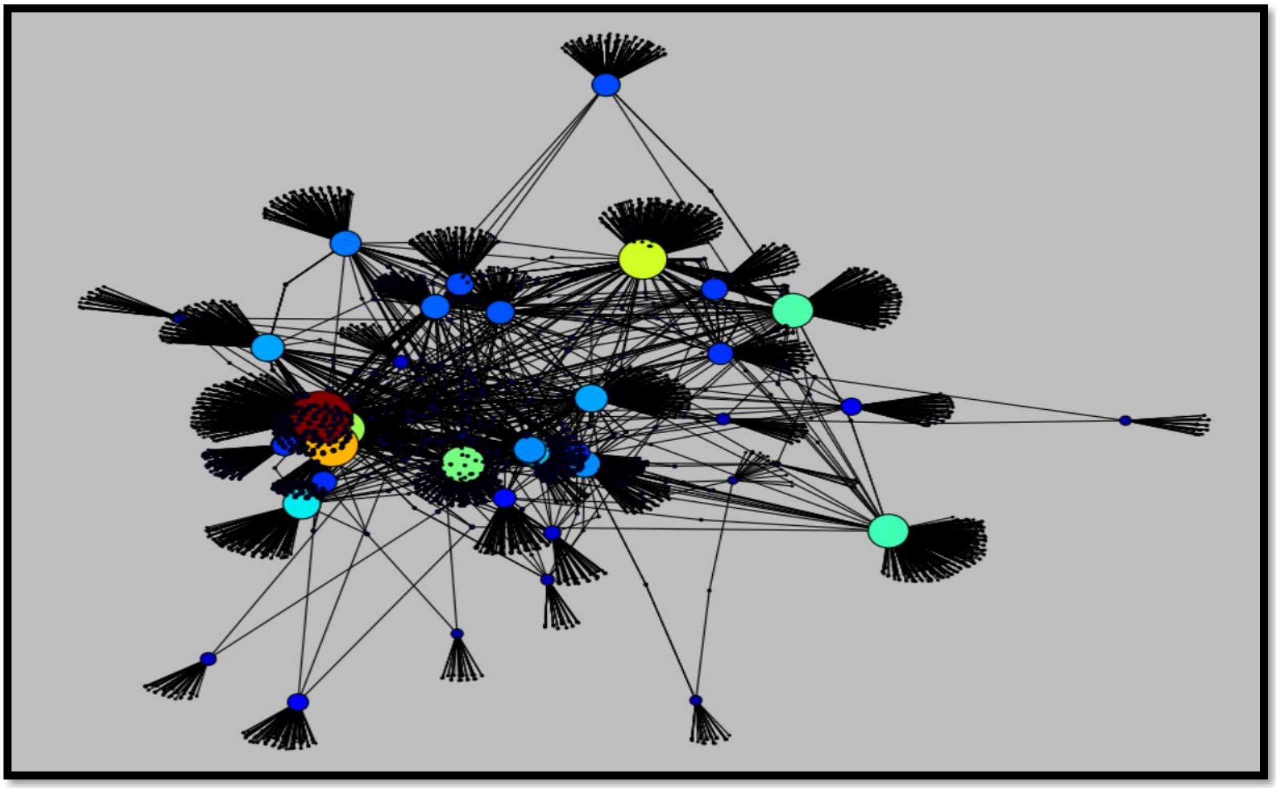

Figure 2. Network Based on Degree Centrality

The most centered people in the network based on degree_centrality are:

	Actor	degree_centrality
0	195657825	0.165673
1	204669023	0.118602
2	6160486	0.100103
3	205193250	0.090647
4	25034832	0.082014

Table 1. Degree Centrality

To find most connected people/individuals in the network we can use degree centrality measure. It is also used to find people who hold most information. If the degree centrality is higher for a person or individual, he can connect with wider network.

In the above network visualization (Figure 2), the size of the node represents the degree centrality. The list of actors above the visualization shows top ten actors in the network who hold most information. Color of the node is based on the nodal degree. Table 1 shows the top five actors with their degree centrality and their values.

Betweenness Centrality:

Betweenness centrality is a measure which tells us that the number shortest paths between other nodes that pass through a node. To compute betweenness for a node v_i, is to compute the number of shortest paths between other nodes that pass through v_i,

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}},$$

Where, σst is the number of shortest paths from node s to t and σst (vi) is the number of shortest paths from s to t that pass through vi.

The most centered people in the network based on betweenness_centrality are:

	Actor	betweenness_centrality
0	195657825	0.247442
2	6160486	0.185155
6	234684445	0.133982
7	226754592	0.131425
1	204669023	0.116682

Table 2. Betweenness centrality

Betweenness Centrality Measure can detect individuals who influence the transfer of information. If these individuals do not exist in the network, then the information cannot flow on both sides of the network.

Top ten actors based on the Betweenness centrality:
[195657825, 6160486, 234684445, 226754592, 204669023, 198574124, 101056562, 217350178, 25034832, 209453152]

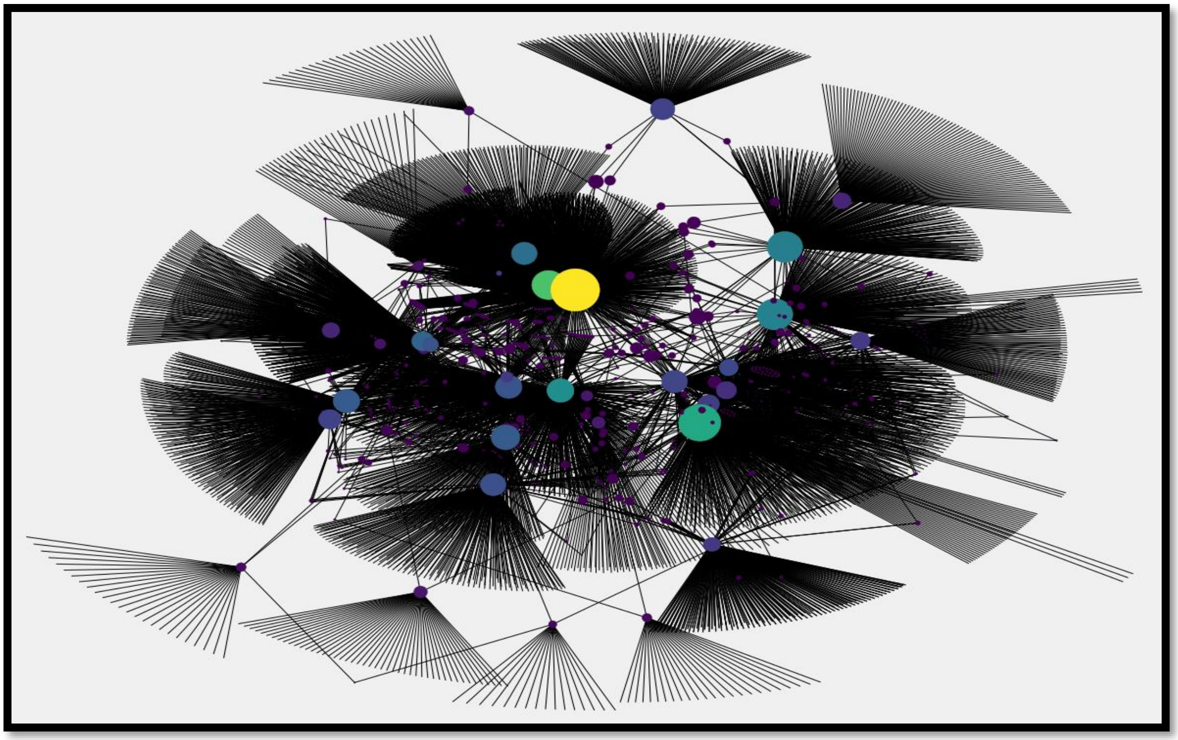

Figure 3. Network Based on Betweenness Centrality

In the above network visualization (Figure 3), the size of the node represents the Betweenness centrality. The list of actors above the visualization shows top ten actors in the network who influence the transfer of information. Color of the node is based on the nodal degree. Table 2 shows the top five actors with their Betweenness centrality and their values.

Closeness Centrality:

Closeness Centrality shows how close the actor is to the whole network and if the actor central then it can quickly interact with other nodes. If the nodes are more central, then they can quickly reach to other nodes. The centrality of the node is decided by the average shortest path length, the smaller the average shortest path length the higher the centrality.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)},$$

where $d(v,u)$ is the shortest-path distance between v and u, and n is the number of nodes in the graph.

The most centered people in the network based on closeness_centrality are:

	Actor	closeness_centrality
0	195657825	0.365460
1	204669023	0.363114
3	205193250	0.352843
2	6160486	0.337379
18	187334729	0.330750

Table 3. Closeness Centrality

Top ten centered nodes based on close ness centrality are:
[195657825, 204669023, 205193250, 6160486, 187334729, 25034832, 209453152, 217350178, 115965992, 182190122]

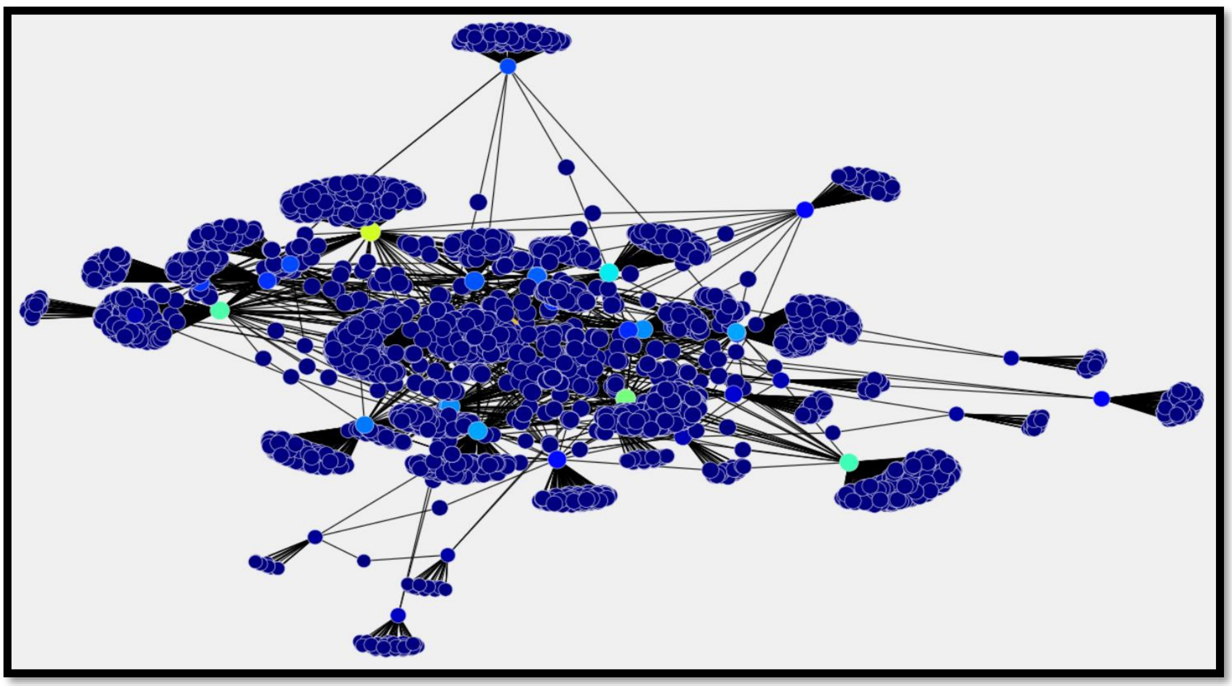

Figure 4. Network based on Closeness Centrality

In the above network visualization (Figure 4), the size of the node represents the Closeness centrality. The list of actors above the visualization shows top ten actors in the network who can quickly reach to other nodes. Color of the node is based on the nodal degree. Table 3 shows the top five actors with their Closeness centrality and their values.

Eigenvector Centrality:

Having more connected nodes does not mean that node is more important, instead having more important node provides a stronger signal. Eigenvector centrality generalizes the degree centrality by combining the importance of neighbors. We can calculate the eigenvector centrality for both directed and undirected networks.

$$\boxed{Ax = \lambda x}$$

where A is the adjacency matrix of the graph G with eigenvalue λ.

The most centered people in the network based on eigenvector_centrality are:

	Actor	eigenvector_centrality
0	195657825	0.460764
1	204669023	0.402211
3	205193250	0.356499
8	115965992	0.065005
23	34340942	0.057143

Table 4. Eigenvector Centrality

The nodes/actors having a high spectral centralized are the nodes who have the most relation in the network, they are central and have influence in a general way on the network.

The top ten nodes based on the eigenvector centrality are:
[195657825, 204669023, 205193250, 115965992, 34340942, 3380276, 59892862, 175300482, 189528179, 188671296]

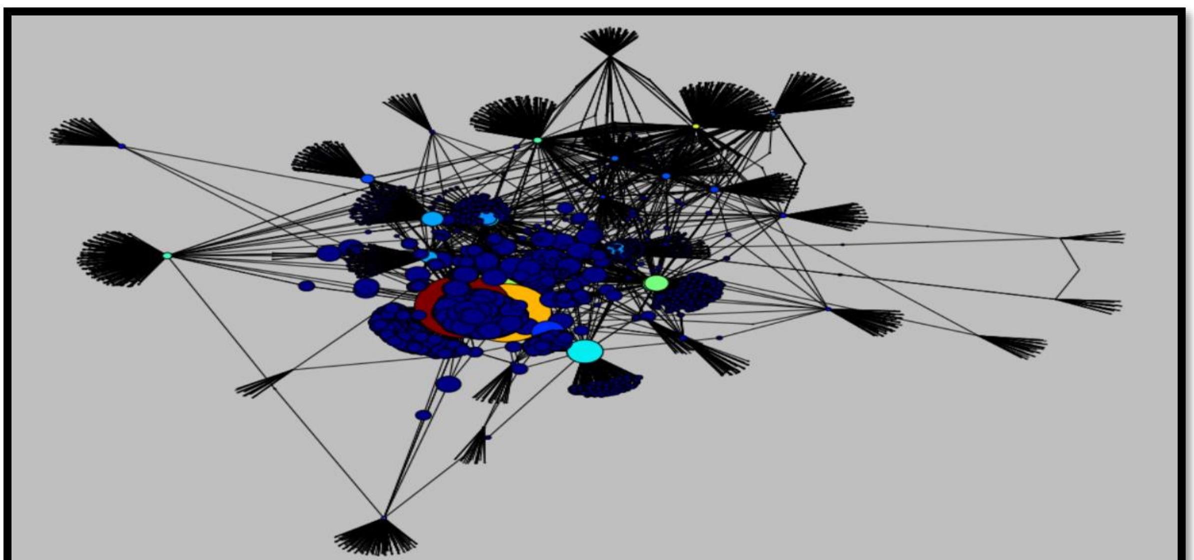

Figure 5. Network Based on Eigenvector Centrality

In the above network visualization (Figure 4), the size of the node represents the Eigenvector centrality. The list of actors above the visualization shows top ten actors in the network who have influence in a general way on the network. Color of the node is based on the nodal degree. Table 3 shows the top five actors with their Eigenvector centrality and their values.

Member-Member Centrality Measures:

	Actor	degree	betweenness_centrality	closeness_centrality	eigenvector_centrality	degree_centrality
0	195657825	806	0.247442	0.365460	0.460764	0.165673
1	204669023	577	0.116682	0.363114	0.402211	0.118602
2	6160486	487	0.185155	0.337379	0.002880	0.100103
3	205193250	441	0.045403	0.352843	0.356499	0.090647
4	25034832	399	0.073867	0.329094	0.030225	0.082014

Table 5. Centrality Measures of Member to Member data

Group Dataset Analysis:

The group dataset contains group1 and group2 which forms edge list for group to group graph.
Number of nodes is: 456
Number of edges is: 6692

Network Connectivity:

The network connectivity of the group dataset is given as follows:

```
G1 = nx.path_graph(group_edges)
print(nx.is_connected(G1))

True
```

The group dataset has group_id, group name, number of members in the group, category_id, category name. Using all the attributes in the data we can find the largest groups, most clustered groups, most central groups in the transferring the information and we can find communities within groups.

Network Visualization of group Dataset:

We have a total of 602 groups in the Nashville group-group data. After matching the group edges data with meta information we have around 456 groups. We have done the visualization of the group data.

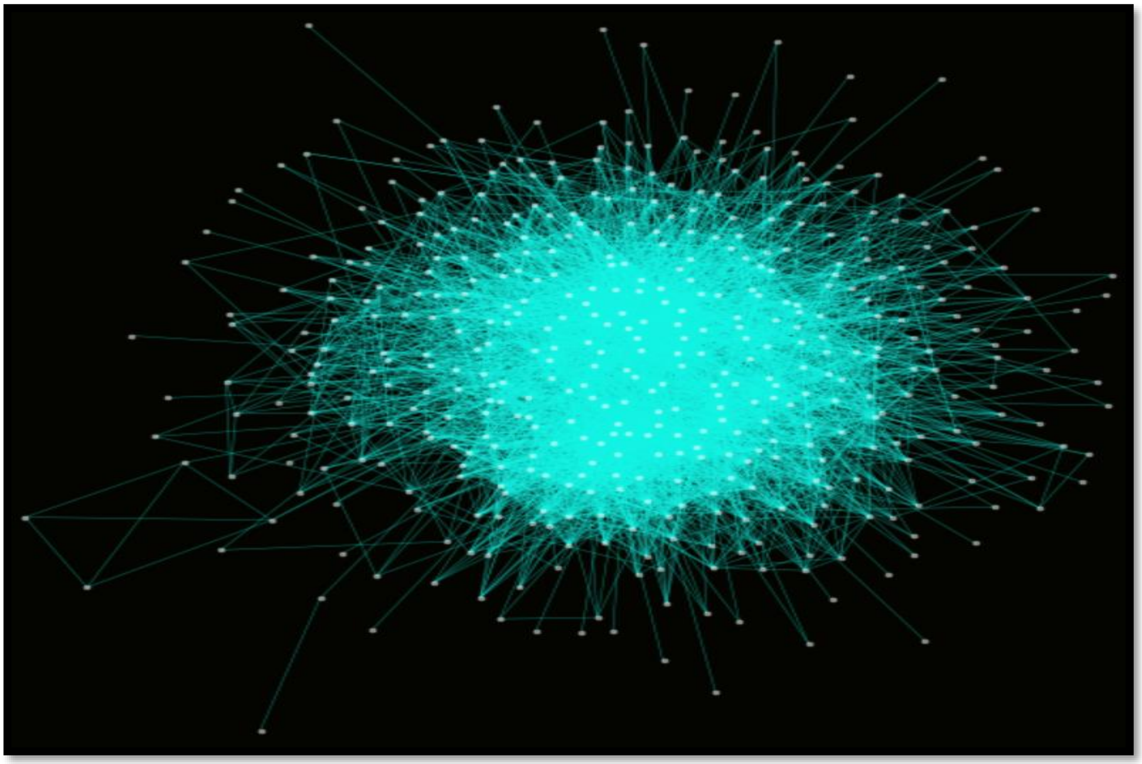

Figure 6. Network Visualization of Group Dataset

Finding the Centrality measures among the groups:

We have used different centrality measures degree centrality, Clustering and Betweenness Centrality. By using all these centrality measures we found top groups.

Using group dataset, we will work to analyze group-group of Nashville meetup data. In this, each node is an independent meetup and each edge is a number of shared members between each group. If tech group has 1000 members and 300 of them are also members of hook-ups group. By using this information, we will find out top groups among the Nashville meetups.

Degree Centrality of Group:

We have found the degree centrality of the group using networkx formulas. Below are the top ten groups based on the group centrality.

group_id indicates the group of the respective meetup and the group name is the category of the group in the Nashville meetup.

```
The ten most central groups in the network based on Degree centrality are:
group_id
19728145                     Stepping Out Social Dance Meetup
18955830                                        Eat Love Nash
1187715                                         What the Pho!
18506072                                    20s in Nashville
339011                               Nashville Hiking Meetup
4126912                        Nashville Online Entrepreneurs
18243826                                Middle TN 40+ singles
1776274      Nashville SEO & Internet Marketing, Over 1,600...
11077852                            Sunday Assembly Nashville
16487812                                   Code for Nashville
Name: group_name, dtype: object
```

Table 6. Group degree Centrality top groups

Network Visualization of Group with Degree Centrality:

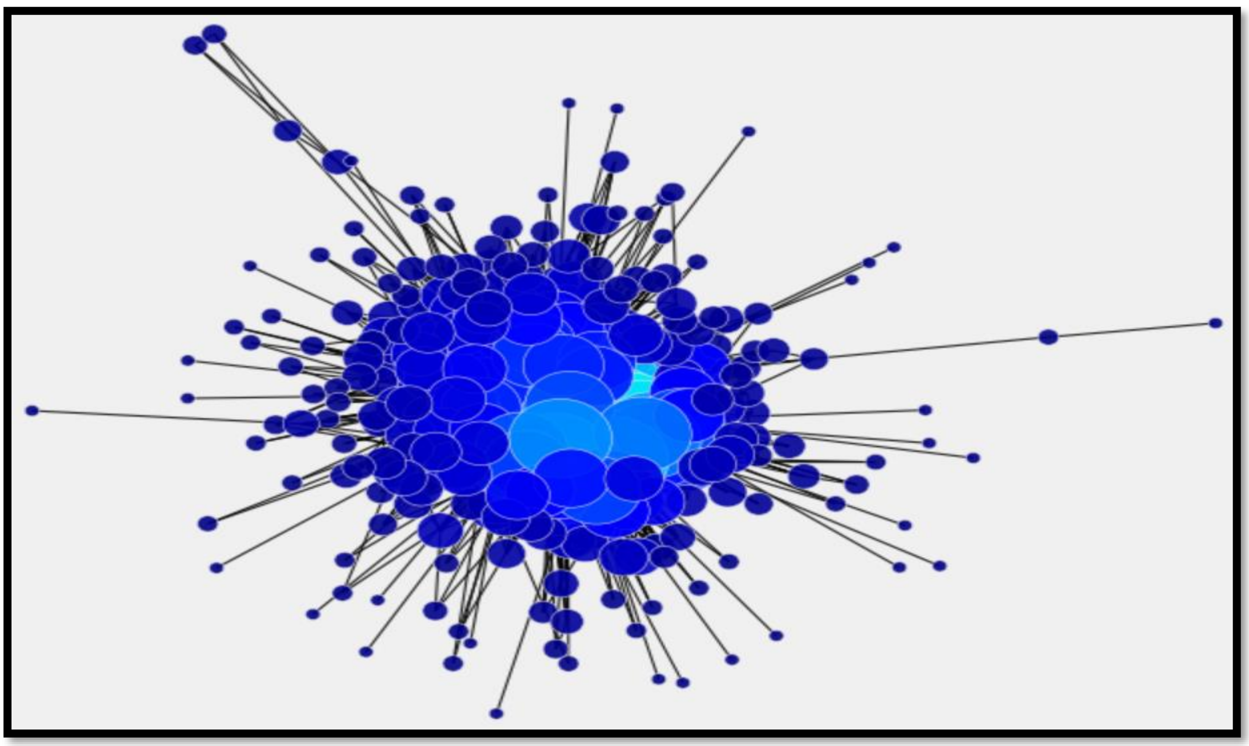

Figure 7. Network Visualization of Group Degree Centrality

In the above network visualization (Figure 7), the size of the node represents the degree centrality of the group. The list of actors above the table shows top ten groups in the network who hold most information. Color of the node is based on the nodal degree. Table 6 shows the top ten groups with their degree centrality and their group id’s.

Clustering of Group:

The clustering coefficient analyzes transitivity in an undirected graph. We have found the clustering of the group data using network measures.

```
The ten most clustered groups in the network based on clustering are:
group_id
25718653                      Nashville Hookah Bar Hook-ups!
22468329                                        Man Cave 615
16321082                       Nashville Cloudera User Group
24811380    Marigolds Author Group + Travel Club for Chris...
14775182                             Nashville Travel Meetup
24619540                               East Nashville YA Lit!
18665368            Second Saturday Divorce Workshop for Women
21262880               Amway Business Opportunity Info Session
20780059                                   Movies with Heide
24744100                          Nashville Instagram Meetup
Name: group_name, dtype: object
```

Table 7. Group Clustering Centrality top groups

The above table shows the top clustering groups in the group data.

Network Visualization of Group with Clustering Centrality:

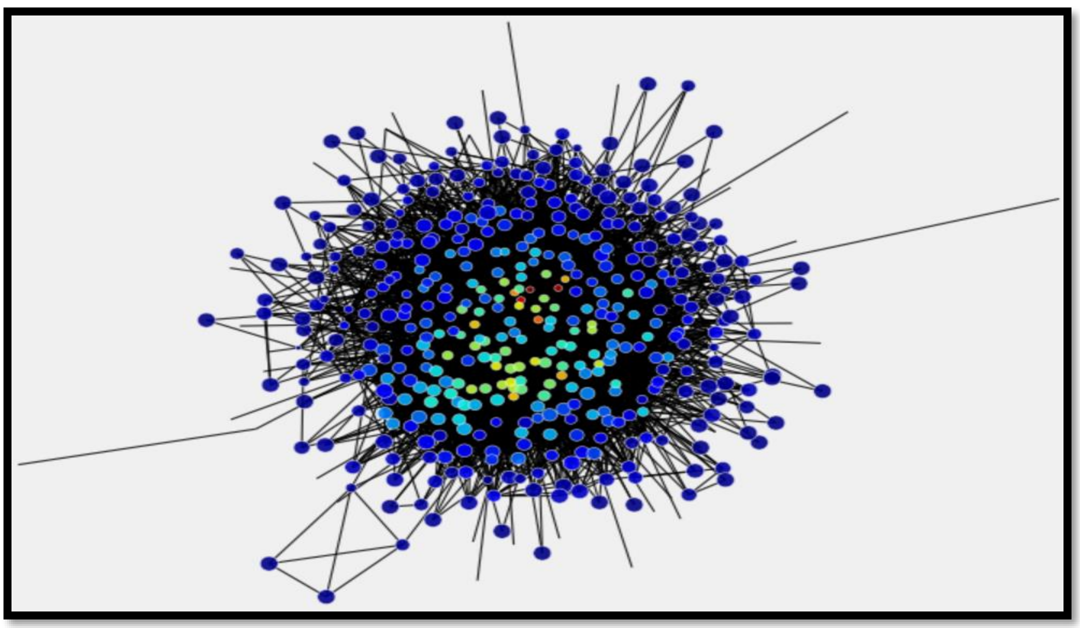

Figure 8. Network Visualization of Group Clustering Centrality

In the above network visualization (Figure 8), the size of the node represents the Clustering centrality of the group. The list of actors in the above table shows top ten groups in the network based on clustering. Color of the node is based on the nodal degree. Table 7 shows the top ten groups with their Clustering centrality and their group id’s.

Betweenness Centrality of Group:

Betweenness centrality is a measure which tells us that the number shortest paths between other nodes that pass through a node. Below

```
The ten most central groups in the network based on centrality are:
group_id
19728145          Stepping Out Social Dance Meetup
1187715                              What the Pho!
18243826                     Middle TN 40+ singles
18955830                             Eat Love Nash
19266390    Nashville Networking Business Luncheon
339011                     Nashville Hiking Meetup
18506072                         20s in Nashville
11077852                   Sunday Assembly Nashville
4126912               Nashville Online Entrepreneurs
1585196                      Tennessee Hiking Group
Name: group_name, dtype: object
```

Table 8. Betweenness Centrality top groups

Network Visualization of Group with Betweenness Centrality:

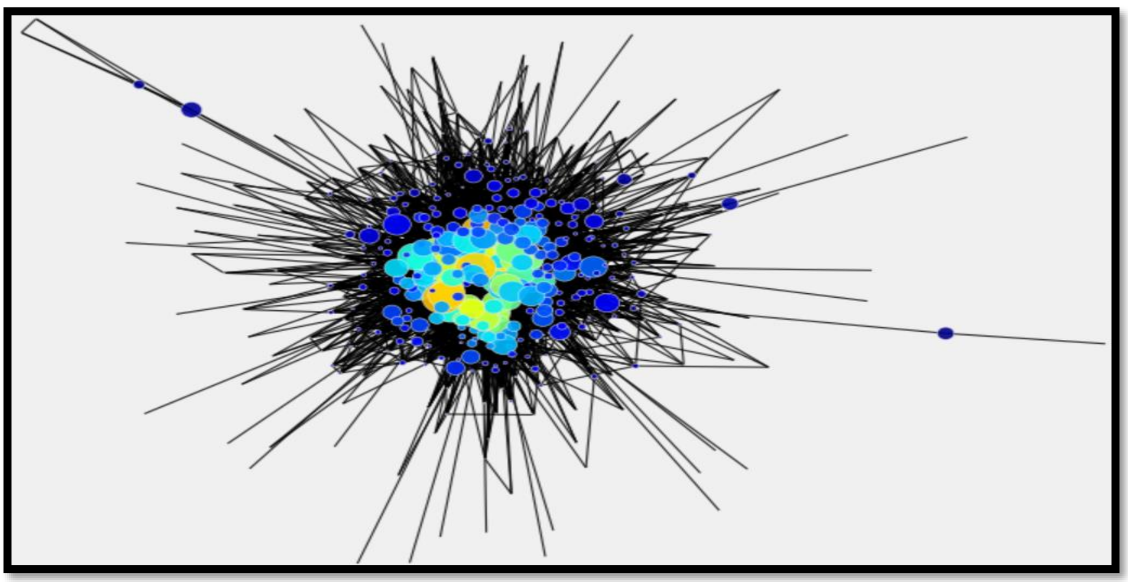

Figure 9. Network Visualization of Group Betweenness Centrality

In the above network visualization (Figure 3), the size of the node represents the Betweenness centrality. The list of actors in the above the visualization shows top ten actors in the network who influence the transfer of information in the group. Color of the node is based on the nodal degree. Table 8 shows the top ten actors with their Betweenness centrality and their group names.

Tech group Analysis in the Group Data:

We have 31 unique groups in the group-group data. By analyzing the data, we could see that the Tech group is the largest group with highest members in the group.

Deriving the technical groups from the data, different centrality measures are used to find the top centrality groups in the tech groups.

We found Degree, Clustering and Betweenness Centrality for the groups for the tech groups.

group_id	group_name	centrality
10016242	NashJS	0.065180
1728035	WordPress Nashville	0.045770
20947040	Nashville Blockchain Meetup	0.041613
6707902	Data Science Nashville	0.038525
13560402	Nashville Modern Excel & Power BI User Group	0.033072
16487812	Code for Nashville	0.031580
18494105	The Iron Yard - Nashville	0.028580
23353167	Nashville Docker Meetup	0.027664
10016162	Nashville PowerShell User Group (NashPUG)	0.026610
18616278	Franklin Developer Lunch & Learn	0.026427

Table 9. top ten tech groups based on centrality

Network Visualizations of tech groups based on centrality measures:

Visualizations are drawn based on the different centrality measures. Nodes sizes are based on the respective centrality measure.

Degree Centrality:

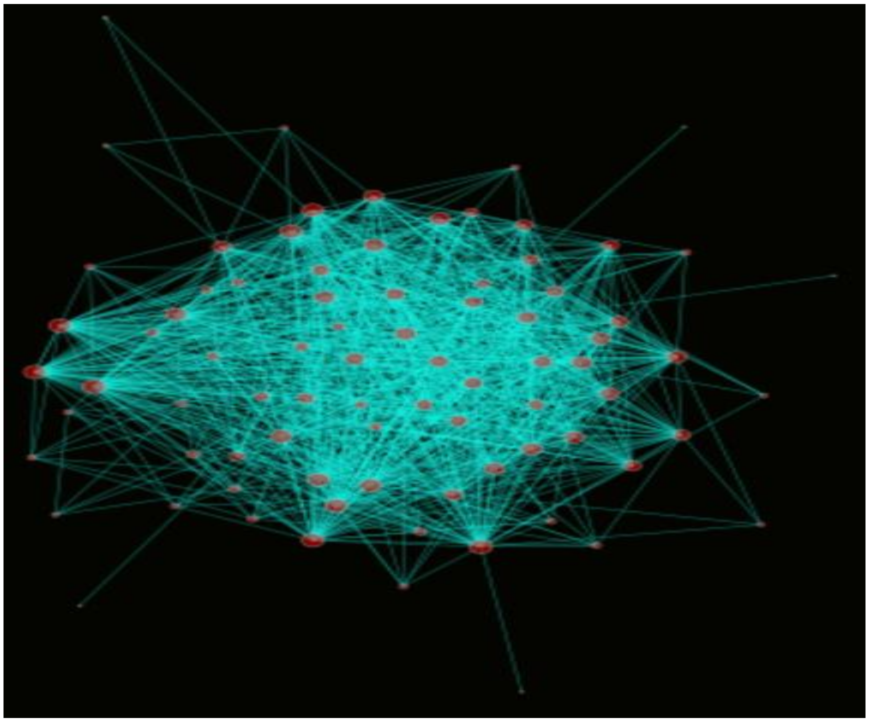

Figure 10. Degree Centrality for Tech group

Betweenness Centrality:

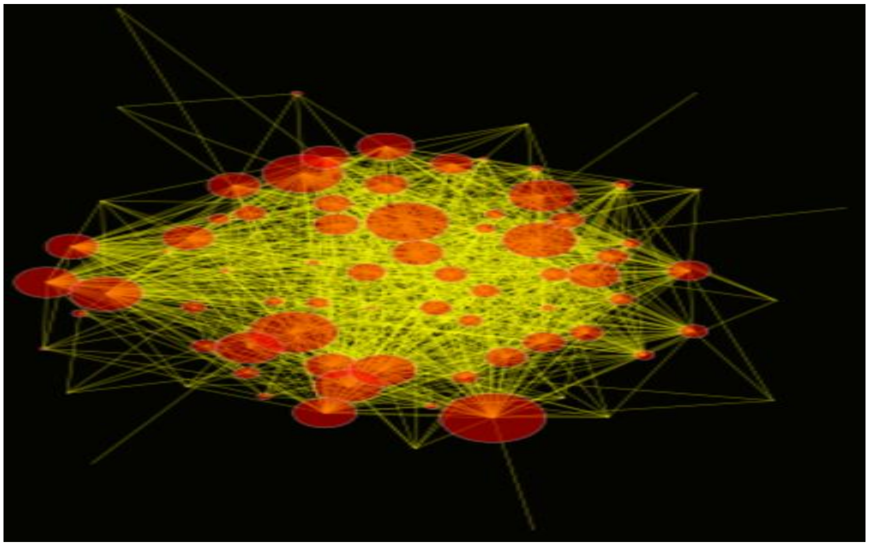

Figure 11. Betweenness Centrality of tech group

Community Detection in the network:

In community detection, data points represent actors in social media, and similarity between these actors is often defined based on the interests these users share. In community detection, individuals are connected to others via a network of links, whereas in clustering, data points are not embedded in a network.

Social media forms a dynamic and evolving environment. Like real-world friendships, social media interactions evolve over time. People join or leave groups; groups expand, shrink, dissolve, or split over time. Studying the temporal behavior of communities is necessary for a deep understanding of communities in social media.

In our dataset which is collected from Nashville meetup, we have many groups and communities which are organized. By detecting communities in the Nashville meetup, we can send/suggest recommendations for the people to join in the meetups or groups based on the interest of the individual.

Community Detection Algorithms:

There are number of community detection algorithms. We have mainly two types of community detection techniques in social network analysis.

1) Member based community detection
2) Group based community detection

Member-based community detection uses community detection algorithms that group members based on attributes or measures such as similarity, degree, or reachability. In group-based community detection, we are interested in finding communities that are modular, balanced, dense, robust, or hierarchical.

We apply community finding to the Nashville meetup network to organize it into a smaller number of communities of related meetups that are easier to interpret, as opposed to manually inspecting many meetups individually.

Louvain Community detection:

The Louvain method is a simple and easy method to implement for identifying communities in large networks. Louvain method can used to find hierarchies of communities which can find sub-communities and this method can be used for large networks.

This method can be used for optimizing modularity of the partition network. the optimization can be done in two steps. In step1 communities are detected using local modularity. In step2 new communities are formed by combining the nodes that belong to the same community. These two steps are repeated iteratively until the maximum modularity and sub-communities are found.

We have used the source code from the networkx module to find the partition network and found the modularity of the network using the modularity function. The community detection function which is used in the tech community detection is given in the coding file.

Communities visualizations:

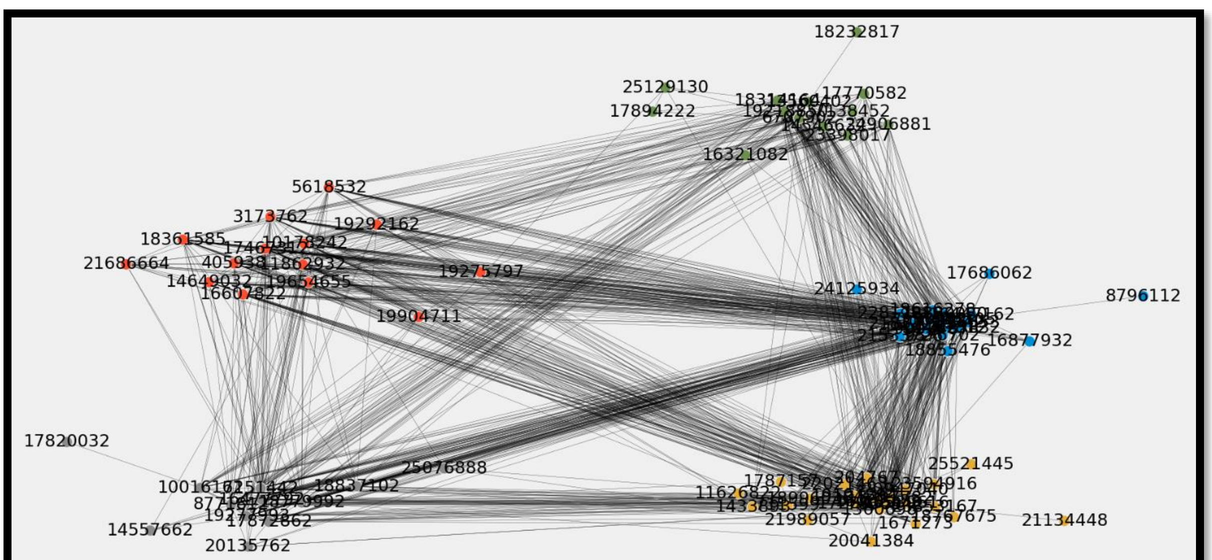

Figure 12. Tech group communities

Most central groups in each community are given as follows:

```
Most central groups in Community 0...
        NashJS
        Code for Nashville
        The Iron Yard - Nashville
        Franklin Developer Lunch & Learn
        PyNash
Most central groups in Community 1...
        Nashville Mobile Developers
        NashMicro - The Nashville Microcontroller Makers Group
        Nashville Java Users' Group
        Nashville Linux Users Group
        Nashville Virtual Reality Meetup
Most central groups in Community 2...
        WordPress Nashville
        Nashville Blockchain Meetup
        Nashville Docker Meetup
        Nashville UX
        Nashville PHP User Group
Most central groups in Community 3...
        Data Science Nashville
        Nashville Modern Excel & Power BI User Group
        NashBI
        Nashville Machine Learning Meetup
        Greater Nashville Healthcare Analytics
Most central groups in Community 4...
        Nashville PowerShell User Group (NashPUG)
        Nashville DevOps Meetup
        The Nashville Microsoft Azure Users Group
        Nashville .NET User Group
        Nashville Amazon Web Services User Group
```

Table 10. Most central groups in the community

Exploration of communities in the Nashville meetup:

In Nashville meetup group data graph, each meetup is group is represented by a node and the edges indicates the weight of the connection between pairs of meetups. The group dataset consists of 456 nodes and 6692 edges.

The betweenness centrality measure identifies strategic bridges in a network. Nodes that occur on many shortest paths between other nodes in the network have high centrality. In the weighted variant of betweenness, the weighted distances between nodes are considered.

Most central groups based on the centrality measures are listed in the tables 6, 7and 8 of the group data analysis. The meetups in Table 8 shows that betweenness centrality measures the ability of nodes in a network to connect disparate parts of that network.

We further analyzed the Nashville meetup by selecting the largest group in the dataset which is tech group. It has NashJS which has 1975 members, which has highest degree with this community. The above table 10 shows the top five communities in the tech group with different groups in the network.

Conclusions:

The above analysis shows the use of network analysis techniques which are used in exploring the datasets from meetup.com from Nashville. The analysis which was made on the Nashville meetups revealed the most important nodes based on centrality. We have used the meta data from the datasets to find the names of the communities which are associated with the nodes. By applying the centrality measures we have found the communities in the network. As of now the work is only on one city which is Nashville. The future work can be applied other cities in finding the most powerful communities. As future work we can develop an application for sending the recommendations for individuals for the meetups.

References:

1. R. Zafarani, M. A. Abbasi, and H. Liu, Social Media Mining: An Introduction, Cambridge University Press.
2. Stanley Wasserman and Katherine Faust, Social Network Analysis: Methods and Applications, Cambridge University Press.
3. Derek Greene, MeetupNet Dublin: Discovering Communities in Dublin's Meetup Network. arXiv:1810.03046
4. C-Blondel: An Efficient Louvain-Based Dynamic Community Detection Algorithm. IEEE Transactions on Computational Social Systems (Volume: 7 , Issue: 2 , April 2020)
5. Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. Physical review E 70(6), 066111 (2004)
6. M. Girvan, M. NewmanCommunity structure in social and biological networks. Proc. Nat. Acad. Sci., 99 (12) (2002), p. 7821