Analysis of Google Playstore

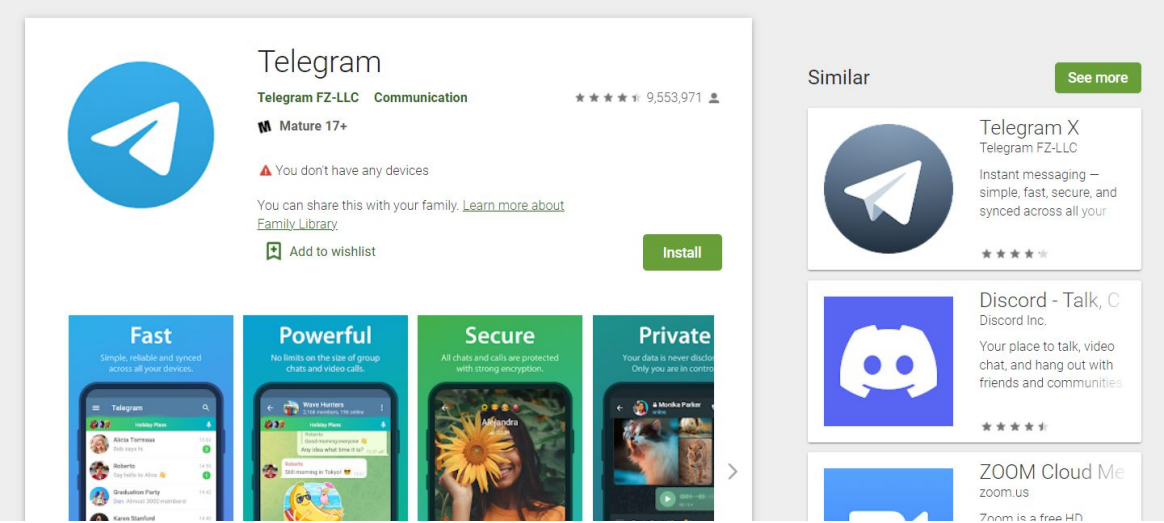
Apps

Scraped 780 apps from play.google.com

Contains a total of 3 million apps.

View of the Page

Since Appstore app pages provide links to similar apps.



Crawling method

Wrote a recursive crawler that takes the number of hops for crawling

Reason for this is to reverse engineer and capture the similarity information from the playstore and eventually to derive a graph with possible connections between nodes.

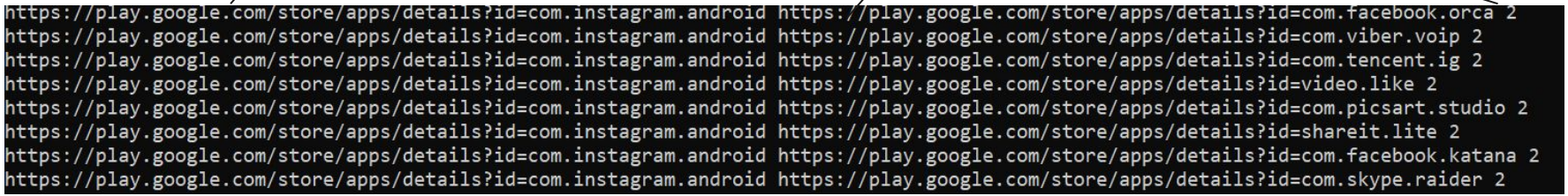
Try to obtain all the apps that are similar to a given app and also try to go a certain depth in that direction.

Results of this crawl

Destination App

Src App

Hop



Results continued

Crawled all the unique URLS derived from the set of (source , destination) to create a node set.

Used the information from the source and destination app at with depth set 2 to generate Edge pairs for Apps.

Derived 780 Nodes and 2300 edge connections between them.

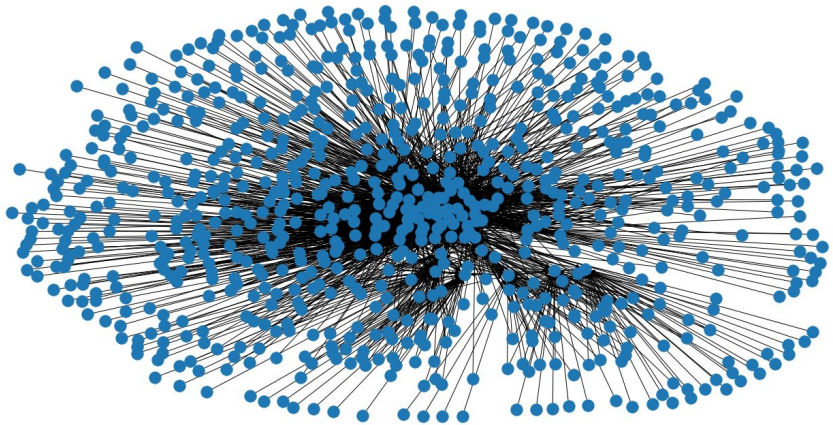
All these edges are assumed to be Directed edges

Node features captured for these 780 Apps

nodeid	Title	Description	App_tags	Publisher_info	Class	Ratings	Agg_Ratings	installs
0	Wikipedia	The best Wikipedia experience on your Mobile d...	Books & Reference	Wikimedia Foundation	Everyone	667888	45	50000000
1	Music player	also called app is one of the best media pla...	Music & Audio	gspro	Everyone	70273	48	1000000
2	WhatSmiley: stickers	integrate directly into your WhatsApp sticke...	Social	SiA - Smileys	Teen	152136	46	10000000
3	4shared	Meet our NEW version Better and fasterFree mob...	Entertainment	New IT Solutions	Everyone	796800	45	50000000

Constructing the Graph

Constructed a Digraph and looks like below with all connections.



Analysing the graph

Found the following properties of the graph for starters

Density of the graph.

Centrality measures (Closeness and betweenness centrality)

Found neighbours for each Node (App) in-order to investigate centrality measures

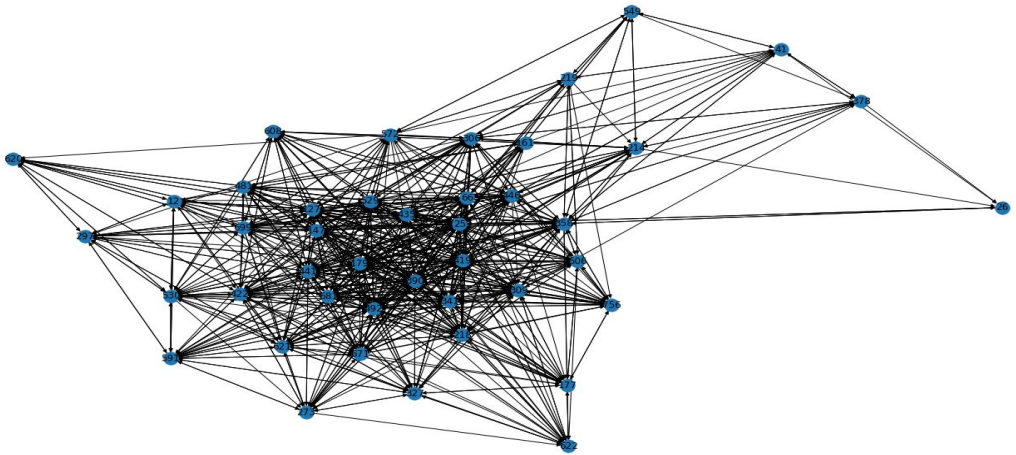
Betweeness centrality of nodes

Found betweenness centrality of all nodes.

Ranked these nodes based on betweeness centrality

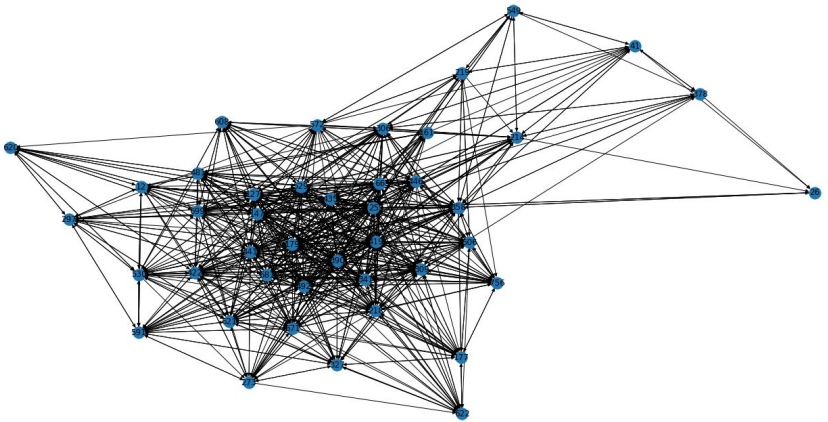
And investigated subgraphs stemming for this node. Example node: Whatsapp: nodeid: 25

Subgraph of Whatsapp and its neighbours.



This cluster of apps including whatsapp showed a strong Triadic relationship

Found the number of triangles that include whatsapp(node 25) as one of the vertices.



Whatsapp and its neighbours based on betweenness Centrality

Reason for this is to find whatsapps position in the graph structure with directly connected neighbours at 1 hop that are important in information discharge in the network.

Is whatsapp connected to other nodes that exhibit a high betweenness centrality

How do we evaluate the importance of some individuals in a network?

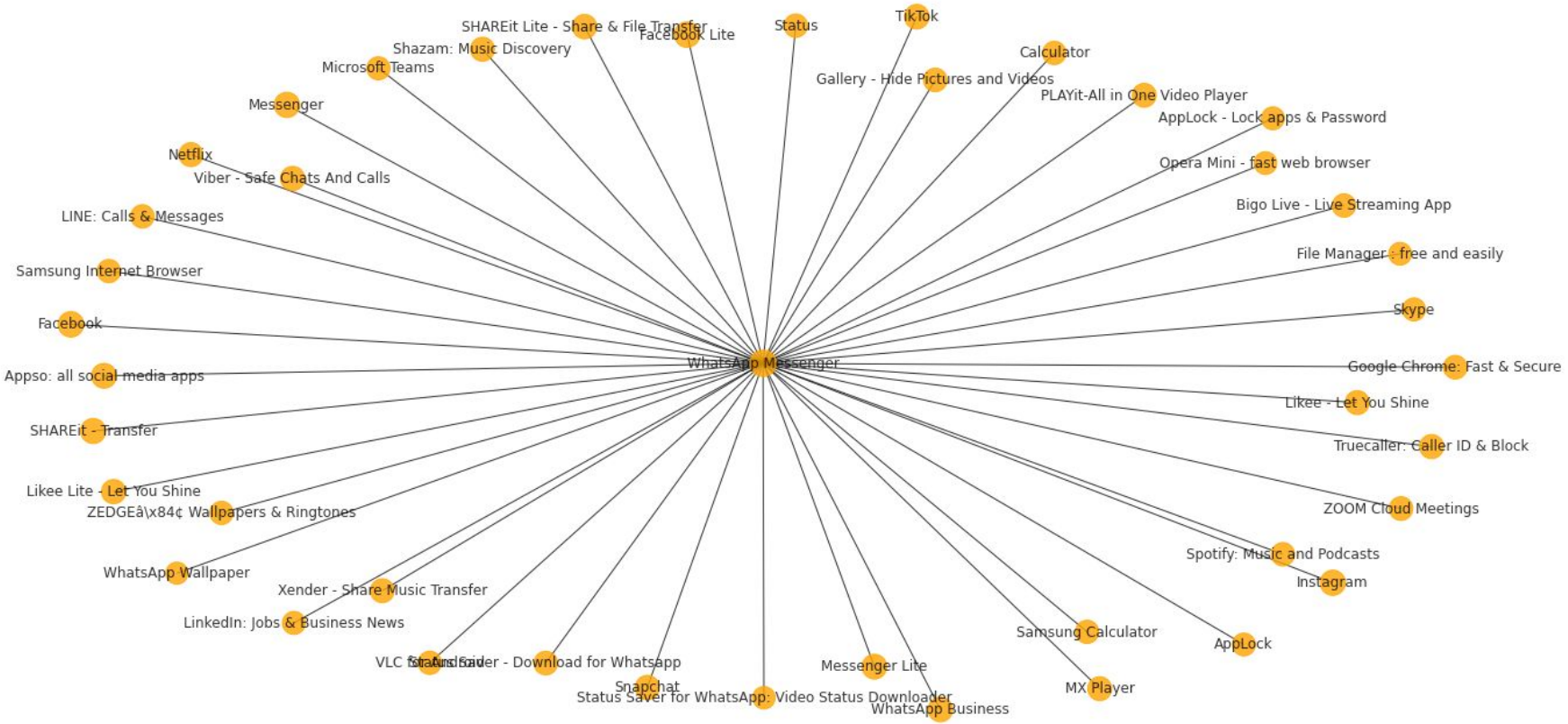
Is whatsapp a node who is hyper-connected with many more nodes that are important in that category?.


WhatsApp Messenger
Status Saver - Download for Whatsapp
Xender - Share Music Transfer
TikTok
MX Player
Facebook
SHAREit - Transfer
Likee Lite - Let You Shine
Messenger
WhatsApp Business
Viber - Safe Chats And Calls
AppLock
Bigo Live - Live Streaming
Calculator
Status
Messenger Lite
SHAREit Lite - Share & File Trans
Skype
Facebook Lite
Calls & Messages
Snapchat

Plotting closeness centrality measures for Whatsapp and its neighbours

To Derive how close whatsapp is to other nodes in the network.

Results in next slide—>

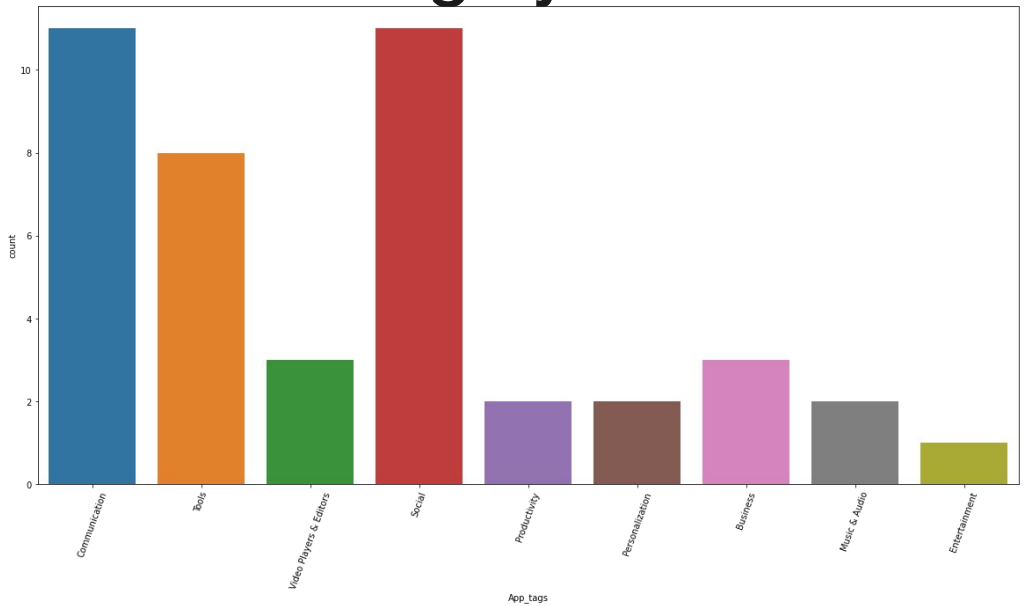

WhatsApp Messenger
SHAREit Lite - Share & File Transfer
Facebook Lite
Status
TikTok
Shazam: Music Discovery
Microsoft Teams
Calculator
Gallery - Hide Pictures and Videos
Messenger
PLAYit-All in One Video Player
AppLock - Lock apps & Password
Netflix
Viber - Safe Chats And Calls
Opera Mini - fast web browser
LINE: Calls & Messages
Bigo Live - Live Streaming App
File Manager : free and easily
Samsung Internet Browser
Skype
Facebook
Google Chrome: Fast & Secure
Appso: all social media apps
Likee - Let You Shine
SHAREit - Transfer
Truecaller: Caller ID & Block
Likee Lite - Let You Shine
ZEDGEâ\x84¢ Wallpapers & Ringtones
ZOOM Cloud Meetings
Spotify: Music and Podcasts
WhatsApp Wallpaper
Instagram
Xender - Share Music Transfer
LinkedIn: Jobs & Business News
Samsung Calculator
AppLock
VLC for Android
Messenger Lite
Snapchat
MX Player
Status Saver for WhatsApp: Video Status Downloader
WhatsApp Business

Plotting the categories that whatsapp is most connected to

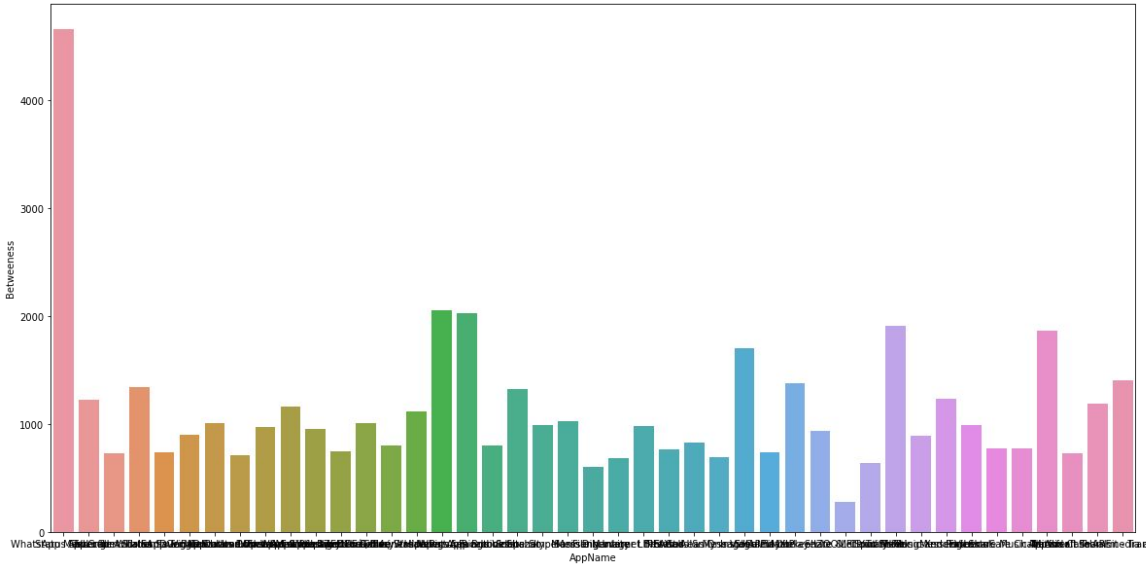
To check if whatsapp is connected to categories that are similar in function(Messaging, Social etc)

Plot in next page—>

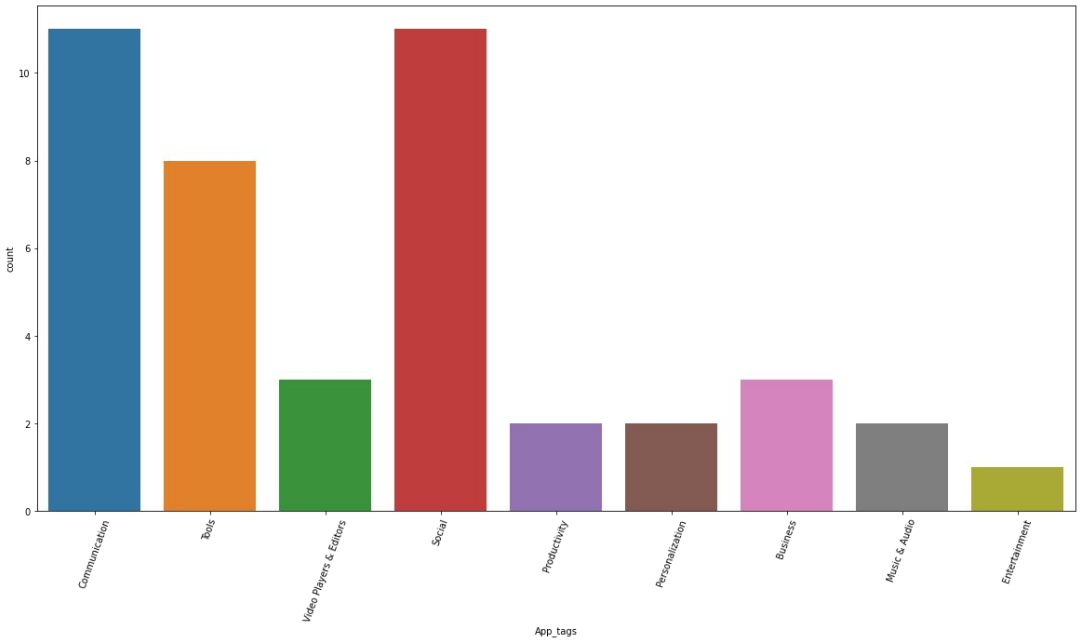
Whatsapp has most connections in Communication, Social and tools category



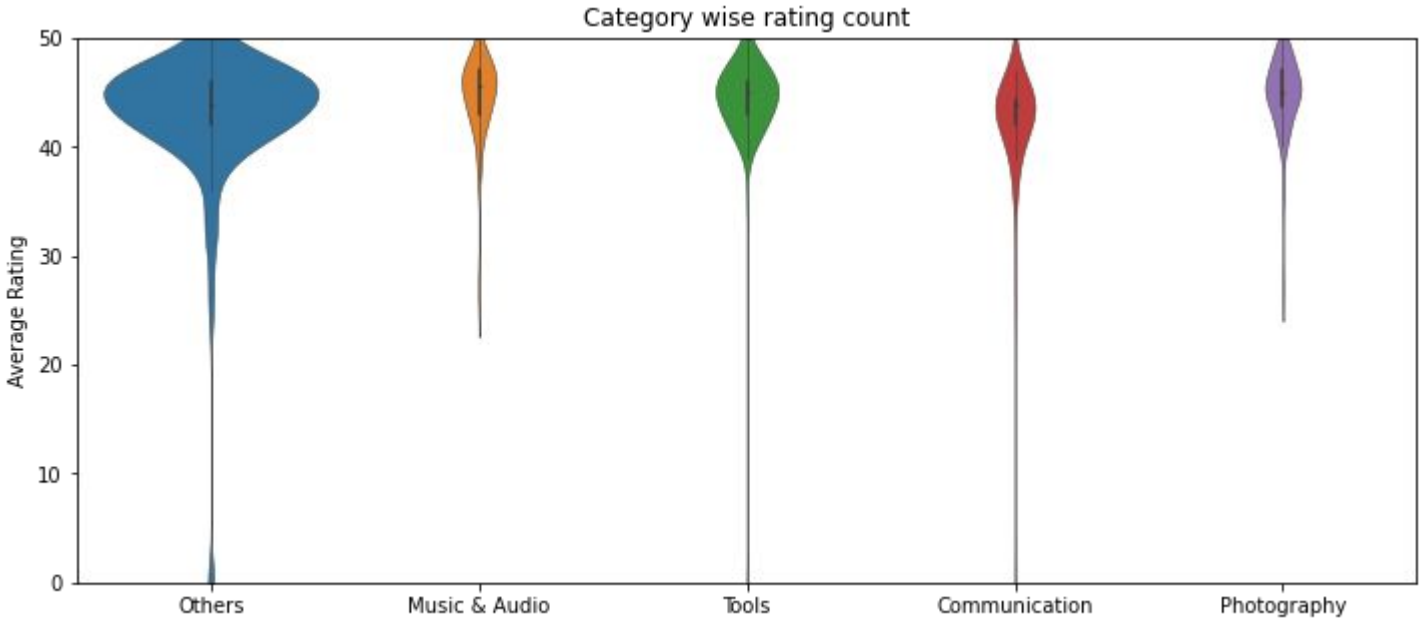
Y-Axis betweenness centrality, X-axis nodes connected to whatsapp



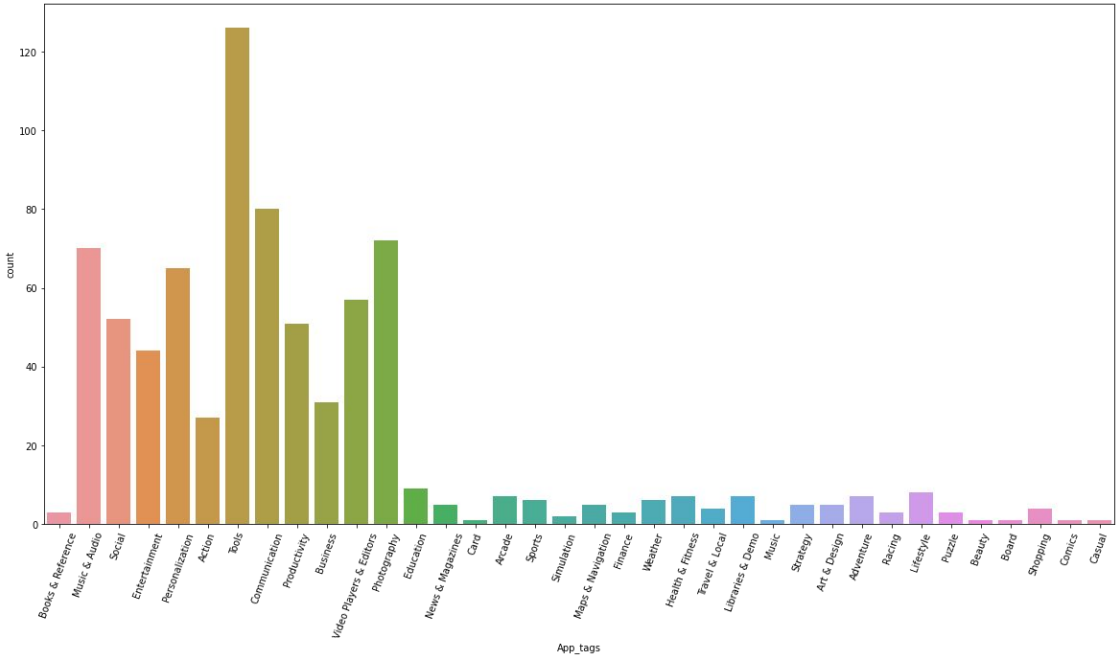
Categories and their density in the dataset



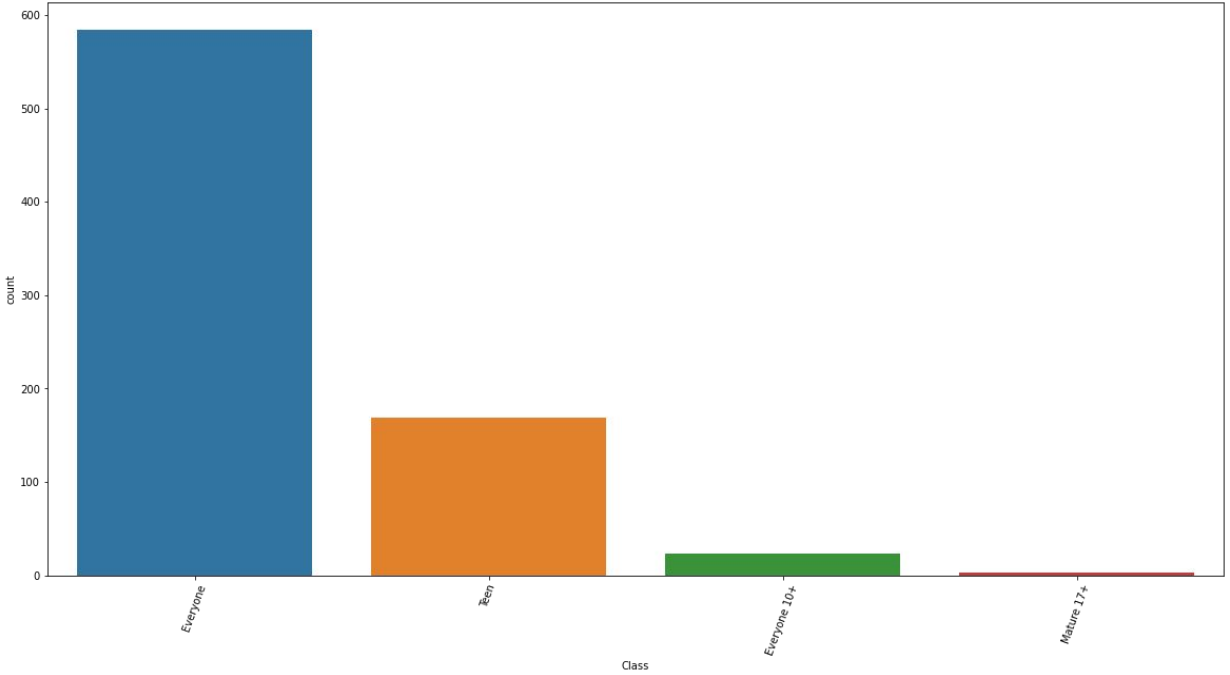
Violin plots for Average Rating vs Categories



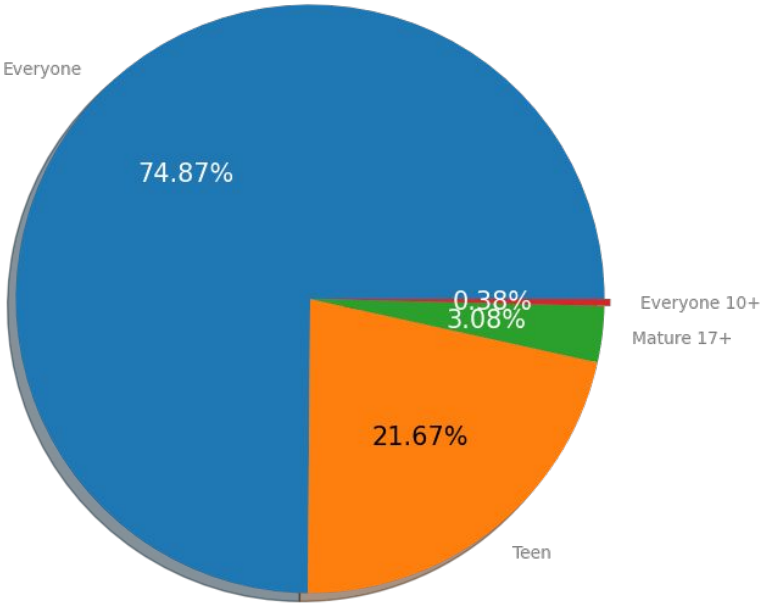
Counts of Apps vs their categories



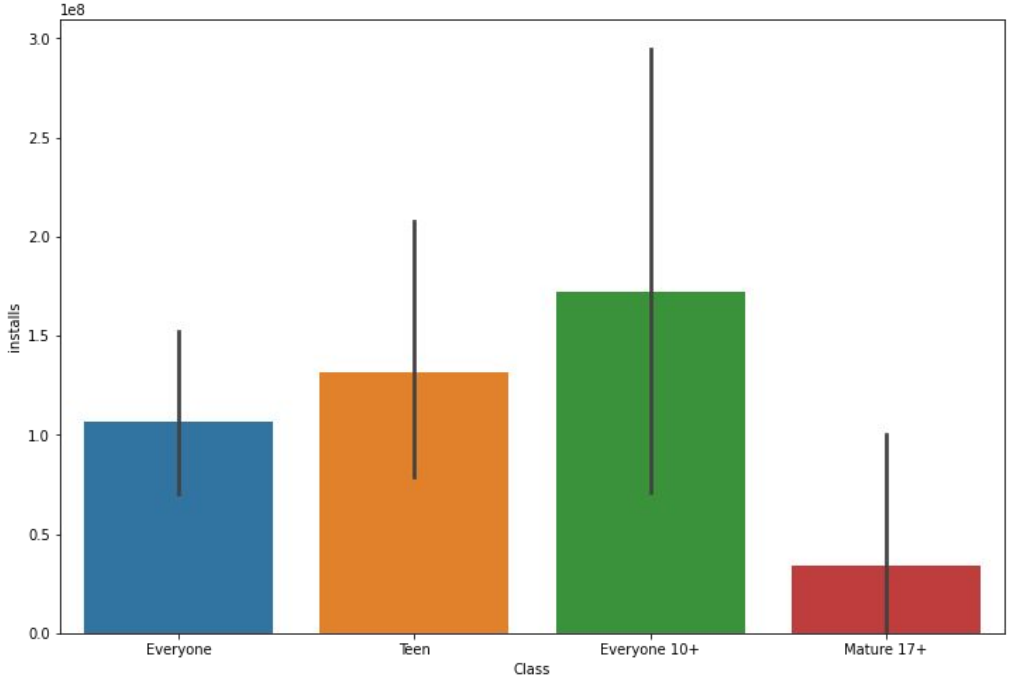
App user category vs counts



Continued



Apps user Category vs installs



Next steps

Finding correlation between these features.

Capturing the notion of popularity as a function of features(ratings,installs,category,Tags).

Converting the problem to a Supervised classification problem

Can we correctly classify the apps to their right Category (App category)

Model this problem into to use Graph neural network for classification and prediction activities and use centrality measures as node features to reinforce importance.

Generate embeddings for these nodes from graph classification and apply dimensionality reduction to study cluster separation.

Regression: Predicting ratings for the App

Possible rating to indicate its popularity.

Counterfeiting problem

Eventually try to derive similarity measures for counterfeit Apps through how similar the embeddings are to a popular app and derive a rating (TODO).

Questions?