Task 1,3:

Subreddit: wallstreetbets, it was the top 1 at 24/4/2022

Data collected at 5/5/2022

For each post, we collected all the subcomments of the first at most 3000 level 0 comments

red top 1 post of that month

https://www.reddit.com/r/wallstreetbets/comments/txma40/wallstreetbets\_predictions\_tournament\_for\_april/?utm\_source=share&utm\_medium=web2x&context=3

blue top 2 post of that month

https://www.reddit.com/r/wallstreetbets/comments/ufc3eu/alright\_boys\_time\_for\_some\_classic\_wsb\_degeneracy/?utm\_source=share&utm\_medium=web2x&context=3

green top 3 post of that month

<https://www.reddit.com/r/wallstreetbets/comments/twk0qn/our_savior/?utm_source=share&utm_medium=web2x&context=3>

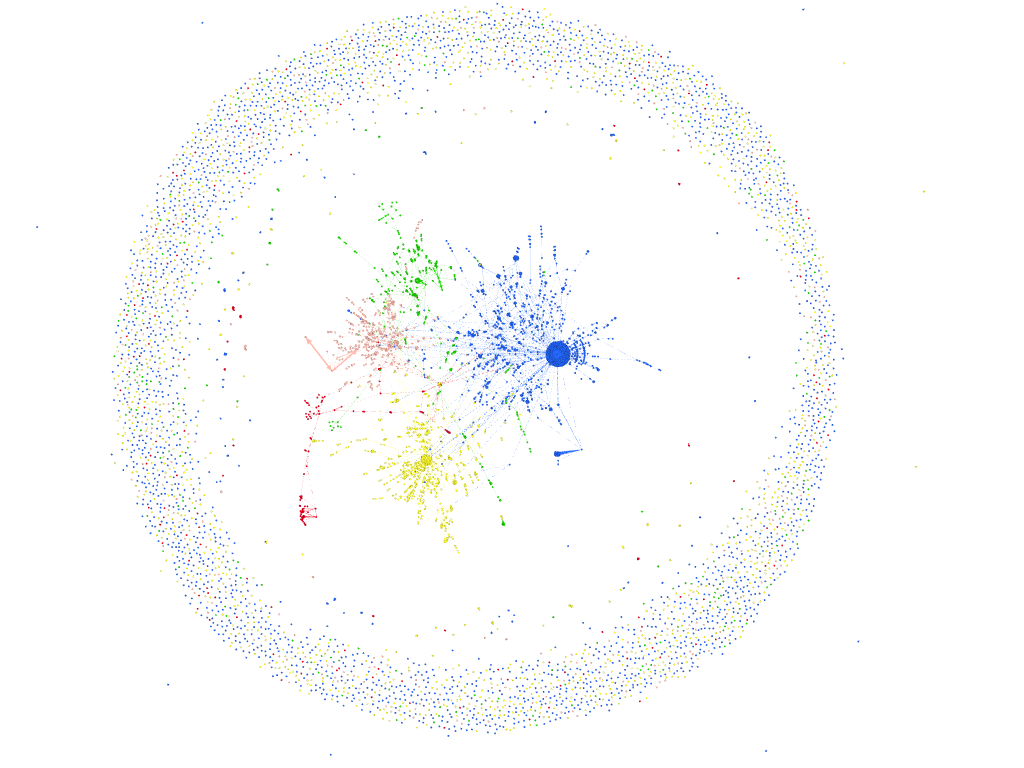
yellow top 4 post of that month

https://www.reddit.com/r/wallstreetbets/comments/u7wtfh/5k\_to\_100k\_overnight\_nflx\_put/?utm\_source=share&utm\_medium=web2x&context=3

pink top 5 post of that month

<https://www.reddit.com/r/wallstreetbets/comments/ug4351/suffering/?utm_source=share&utm_medium=web2x&context=3>

Whole graph



Software used: Gephi

Some important people:

Chart

Description automatically generated

Red arrows: the comments that user did.

Blue arrows: the comments he got

Chart, radar chart

Description automatically generated

A picture containing text, sky, outdoor, several

Description automatically generated

Chart, radar chart

Description automatically generated

A picture containing text, sky, document

Description automatically generated

Chart

Description automatically generated

Diagram

Description automatically generated with low confidence

A picture containing sky, text, map

Description automatically generated

A map of a city

Description automatically generated with low confidence

Chart

Description automatically generated with medium confidence

Some big arrows represent big weights. Weight = how many times someone commented in the same person

Chart

Description automatically generated

Chart, radar chart

Description automatically generated

Some other graph photos

A picture containing sky, map, several

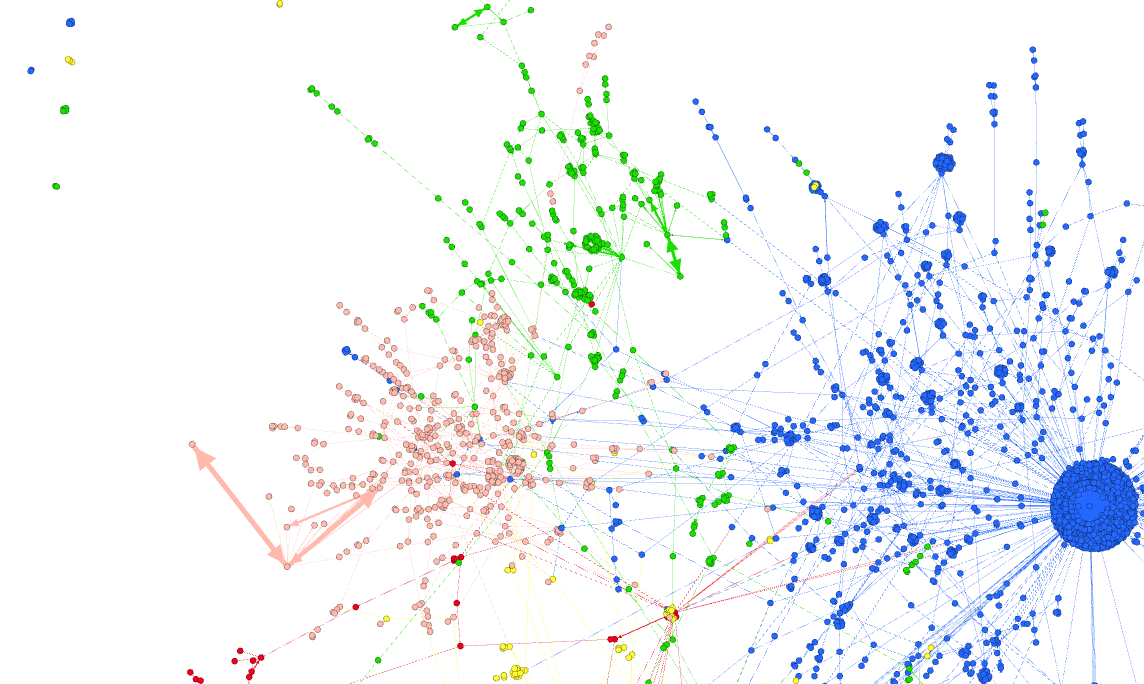
Description automatically generated

Map

Description automatically generated with medium confidence

A picture containing chart

Description automatically generated



A picture containing red, indoor

Description automatically generated

Chart, line chart

Description automatically generated

A picture containing ball, hitting

Description automatically generated

Chart, line chart

Description automatically generated

Text, chat or text message

Description automatically generated with medium confidence

Graphical user interface, text, application, chat or text message

Description automatically generated

Task 2

COMMENTS PER DAY.

THE FIRST DATE IS THE CREATION DATE OF THE POST

THE LAST DATE IS THE DATE OF THE LATEST COMMENT CAPTURED AT 5/5/2022

FOR THE TOTAL COMMENTS, FIRST DATE IS THE DATE OF THE OLDEST CREATED POST AMONG THESE 5 AND LAST DATE IS THE LATEST COMMENT DATE CAPTURED AMONG ALL THE COMMENTS

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

Task 4

Data extracted from Gephi/Python:

Variances can be observed here:

Note that we used a different version of the graph(the one that occurs if we remove nodes without neighbors and nodes with only selfloops) in order to get more specific results about the part of the graph we are interested on.

A picture containing background pattern

Description automatically generated

Considering the directions, the diameter is 23(by GePhi). That means that the maximum depth of continuously replies on a comment could be 23. However we don’t exactly know what is the maximum depth.

For example, we could have a case of a lvl 0 comment of a user x with a reply by a user y, with user x replying in a different lvl 0 comment z.

(y)->(x)->(z) could be:

Comment x Comment z

-reply comment y OR -reply comment x

Comment z -reply comment y

-reply comment x

2 depth 2 comments 1 depth 3 comment

If we don’t consider directions, diameter Is 21 which means 2 users can be connected via (at most) 19 middle users who made a comment or received a reply, starting from one of these 2 users

**From now on, we consider graph undirected:**

Average path length by Gephi: 5.40834432276679, Which means that the most of the pairs of 2 users are being connected through approximately 3 middle users in the same way described above

(slides 19-21 Lecture 3)

Clustering coefficient of the graph is a way to tell how much transitivity can be seen in the whole graph. Definition of transitivity:

If vertex u is connected by an edge with v, and v is connected  
by an edge with w, then u is connected by an edge with w

A high clustering coefficient is translated to a network with high transitivity (People belong to tight groups)

Without considering directions(people are connected if the first person made a comment to the second or the other way):

average\_clustering = sum of all node clustering/nodes

For average\_clustering, each node u clustering(normalized) is calculated with:

Graphical user interface, text

Description automatically generated

https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.cluster.clustering.html#networkx.algorithms.cluster.clustering

For betweenness centrality(not normalized) we used:

Text

Description automatically generated

<https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.betweenness_centrality.html#networkx.algorithms.centrality.betweenness_centralit>

(Slide 13-14)

Betweeness centrality captures how much a vertex falls between others.

Another way of looking at centrality is by considering  
how important nodes are in connecting other nodes

For closeness centrality(normalized) we used:

Text, letter

Description automatically generated

<https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.closeness_centrality.html#networkx.algorithms.centrality.closeness_centrality>

(slide 12)  
Indicates the closeness of a given node to all other nodes in  
terms of shortest path  
Typically, the more central a node is, the closer it is to all other  
nodes

size of giant component: A subgraph which indicates the maximum amount of people that can be connected via a path that consists of making/receiving replies.

in-degree centrality and out-  
degree centrality(not normalized) for each node: Calculated as the sum of in/out degrees. In degree centrality Indicates how many replies a user got and out degree centrality how many replies he did to other users