

CS 48002 – Special Topics in CS: Network Science Project Paper

Community Detection on a Collaborative Art Project

Muhammed Orhun Gale, Ali Eren Ak

June 28, 2022

Abstract

User collaboration is not an unusual way of creating content on the internet; in fact many websites such as Wikipedia and Medium rely on collaborative effort for their contents. However, output of user collaboration is published after evaluation of the content according to different measures of adequacy which indicates the actions that users take are not directly observable. r/place is a collaborative art project that was conducted on Reddit which lets users place a tile on a 4 million-pixel canvas for every 5 minutes. For the 2022 version of the r/place, over 10 million users attended and around 70 million tiles (2.5 million tiles per hour) were placed. Unlike websites where user interactions are under review, on r/place user actions are not restricted except the time limitation which makes it possible to analyze users and their actions not as independent actors but as communities that can take actions en masse. In this work, we analyze the r/place data in order to detect communities with specific intentions and their interactions with other communities in accordance with their intentions. In order to create an interaction network where communities would be detected, we consider users as nodes and tile placements of two users on a given area of the canvas at a specified time interval as edges. We identify and classify communities by focusing on interactions which occurred on time intervals where events such as demolition of artworks, resistance against sabotage attempts and reinstatement of the demolished artworks. It is expected that network analysis will reveal communities that aim to sabotage artworks, communities that try to rebuild those demolished and unification of small communities against bigger communities for different intentions.

1 Introduction

As the most comprehensive data and knowledge source, internet can be described as one of the most significant creation of human collaboration. From the first days of the internet to present, content presented by the internet created by users. Websites such as Wikipedia and Medium depends solely on user collaboration in order to present its main content by filtering the content in terms of various feasibility to publishing criteria which are. [RCWA18] For example in Wikipedia, a content is publishable if it is properly sourced, does not contain insults such as racist or sexist discourses and conforms to many other standards. [wik] On social media platforms such as 9GAG, Facebook and Reddit, again present the content created by the users however, with less strict post/content publishing criteria. In this sense about most visited websites, which are arguably the most influential ones, it can be argued that there exist certain expectations from the collaborative works created by individuals and communities to conform which means interactions between users, for instance their immediate reactions to opinions and actions of other users, are restricted. Therefore, existence of any restrictive criteria on the interaction between users restricts the analysis of collaboration patterns between users. [RCWA18] [Arm18]



Figure 1: Latest picture of the r/place April's Fools 2022 event

r/place is a collaborative art project that was conducted on Reddit which lets users place a tile on a 4 million-pixel canvas for every 5 minutes. For the 2022 version of r/place according to the Reddit, 10 million users attended and around 70 million tiles (2.5 million tiles per hour) were placed. [red] Unlike interactions in influential websites where user interactions are restricted, on r/place user actions are not restricted except the time limitation. Also, since users responsible for placing only one tile, what aim does this tile serve does not create any real responsibility on users. In this sense, it can be said that in r/place event users can choose to become evil or good. For this reason, in order to analyze this collaborative artwork and understand aims and reactions of users, it is a must to analyze users and their actions not as independent actors but as communities that can take actions en masse.

In this project, we identify and classify communities by focusing on interactions which occurred on time intervals where events such as demolition of artworks, resistance against sabotage attempts and reinstatement of the demolished artworks. We expect that network analysis will reveal communities that aim to sabotage artworks, communities that try to rebuild those demolished and unification of small communities against bigger communities for different intentions.

2 r/place Data and Analysis Methods

2.1 Data

Reddit shares the r/place data as 80 different files and tile placement information that has around 12 GB with 70 million rows and two columns which are "time_stampt" and "ser_id,pixel_color, coordinate". Since column has nested information and data is on separate files and not sorted as we like to analyze, we merged, reshaped and sorted the data by using Python and Pandas library. In order to execute above mentioned steps, we used Sabancı University's toSUN HPC cluster and we processed the data via employing the SLURM job to the single computation node with 96 processors and 50GB memory. Consequently we left with 10381163 different users whose 9611718 of them placed more than 50 tiles which corresponds to 87803612 tiles in total. Mean tile placement per user is found as 1.15 and standard deviation of it is 0.66.

2.2 Analyze Approach

Due to the fact that r/place data is all about the information of tiles that were placed and nothing more, in order to represent it as a network, analyze it and interpret it, we came up with an strategy

in accordance with our aims. Since we aim to detect hostile communities which destroy works, those whose artwork is destroyed and those of which try to rebuild what is demolished, we analyzed the r/place event's manually and detected some attack events which destroyed works which can be created with a large collaborative effort and be destroyed with many people as well. To analyze, we found a set of events took place on the canvas which are involved in big attacks. We take the time interval (event time frame) and rectangular area of the artwork of where the attack took place. With this strategy, we narrow down the number of tile placement by 0.05% of the original data for per event. Note that this process can be automatized via employing a technique that can detect conflict regions on a given time frame and via comparing the before/after conflict versions of the rectangular area may detect that an attack took place or not. [VLCY20]

After narrowing down the data to the scope that is desired to analyze, we designed an algorithm to create multiple networks via considering four parameters.

First parameter is deciding how would discrete events be modeled in the network. We came up with two options for this decision as "Concatenated" and "Iterative". Concatenated approach is basically combining the tile placement data of all discrete events and creating the network via considering each individual user in one context. On the other hand, iterative approach is based on iteratively constructing the network which allows us to appreciate each individual user in different contexts.

Second parameter is to decide whether use or not the information about the colors of the tiles that are placed by users are compatible with others or not. This parameter is based on our observation which indicates that users in same communities place tiles with same colors in a specific event since in order to create an artwork or organizing an attack collaboratively color selections must be compatible. Note that placed tiles that distorts the shapes and colors of the artwork are interpreted as an attack to the artwork.

Third parameter is to deciding whether graphs are going to be weighted or unweighted. We decided to take this variable into consideration since this weight of edges can be used in community detection algorithms for example in the modularity-based community detection that can be performed in Gephi. To assign weights to edges, we considered the total number of tiles placed by users that would have edge between in the given context.

Last parameter is deciding a threshold number of tile placements for users that would be included to analysis. This parameter is important to eliminate users that placed a tile to the given rectangular area and time_frame randomly.

Pseudo-code of the designed algorithm can be seen as follows:

- Select parameters = {Event_process_style, color_match, weighted, threshold}
- Initialize set of events = {location, time_frame}
- Create graph
- For all users that has number of tile placement greater than or equal to the threshold in given context, create nodes
- For all tile placement and for all other tile placement, create an unweighted edge from the user that places the tile to the other users if weighted is "False" and edge between given nodes does not exist. Create a weighted edge if weighted is "True" and edge does not exist. Else, increment the weight of the edge.

- If color_match is "True" check colors of two tile placement and if they are not same delete the edge
- If Event_process_style is Iterative. Iterate over the set of events and expand the network

2.3 Community Detection

In order to perform community detection, we used two methods namely modularity-based community detection and Louvain algorithm. Reason why we selected to use these algorithms is because they can be used to detect community structures in large networks fast and accurately. [New06][BGLL08]

3 Results

We applied our algorithm with different parameter combinations on following set of events:

- 1) On (300,319),(431,319),(300,398),(431,398) between 04/01/2022 22.28 UTC to 04/01/2022 22.52 UTC (Figure 4)



Figure 2: Before

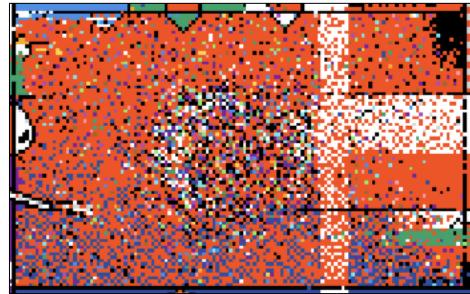


Figure 3: After

Figure 4: xQc community's first attack on Turkish flag

- 2) On (300,344),(479,344),(300,448),(479,448) between 04/03/2022 02.31 UTC to 04/03/2022 03.20 UTC (Figure 7)



Figure 5: Before

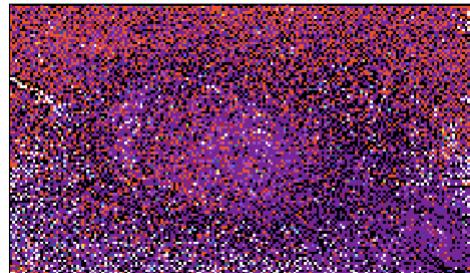


Figure 6: After

Figure 7: xQc community's second attack on Turkish flag

- 3) On (682,682),(772,682),(682,772),(772,772) 04/03/2022 03.31 UTC to 04/03/2022 04.24 UTC (Figure 10)

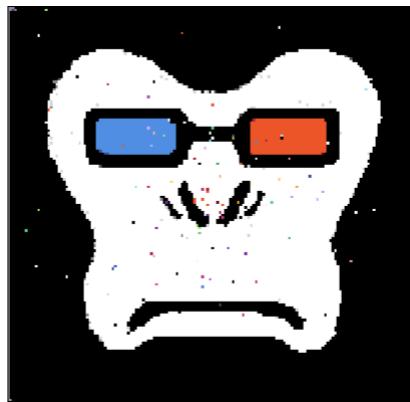


Figure 8: Before



Figure 9: After

Figure 10: Turkish community's attack as a response to xQc logo

4) On (1724,1173),(1882,1173),(1724,1328),(1882,1328) between 04/03/2022 21.09 UTC to 04/03/2022 22.15 UTC (Figure 13)



Figure 11: Before

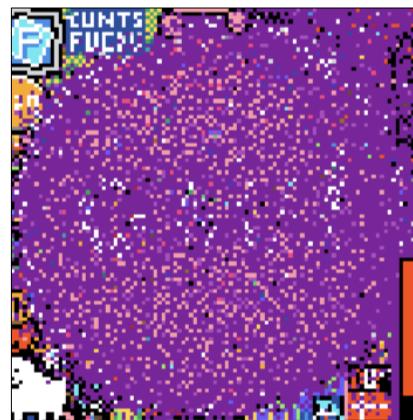


Figure 12: After

Figure 13: xQc community's attack on r/osuplace community

5) On (321,314),(331,314),(321,324),(331,324) between 04/01/2022 17.39 UTC to 04/01/2022 18.11 UTC (Figure 16)

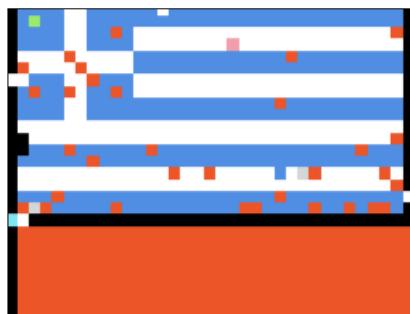


Figure 14: Before



Figure 15: After

Figure 16: Creation of hearth symbol between Turkish and Greece flags

6) On (1201,701),(1328,701),(1724,772),(1882,772) between 04/02/2022 17.39 UTC to 04/02/2022 19.21 UTC (Figure 19)

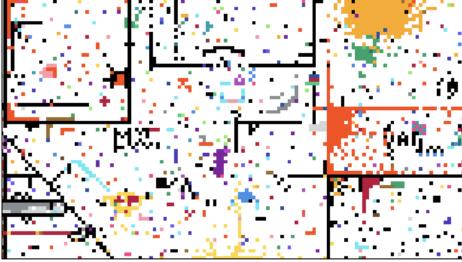


Figure 17: Before



Figure 18: After

Figure 19: A meme created collaboratively by Turkish and Greek communities

These 6 events took place between 4 different communities namely, Turkish community (mostly from the community of Twitch streamer Elraenn), community of Twitch streamer xQc, osu! community (r/osuplace) and Greek community. Specifically, first event is xQc community's first attack to the Turkish flag, second one is xQc community's second attack to the renewed Turkish flag, third one is Turkish community's attack to the xQc's logo in response, fourth event is xQc community's attack to the logo of r/osuplace, fifth event is creation of a heart symbol between Turkish and Greek flags and last event is the creation of a meme with collaboration of Turkish and Greek communities.

With above stated set of events, we aimed to detect 4 main communities, relationships between them and existence of communities that we cannot observe by eyes which may behave different in different contexts. In this sense, we initialized our algorithm with different combinations of parameters.

3.1 Number of Tile Placement Threshold

While number of tile placement threshold is a crucial parameter to eliminate users that placed a tile randomly, it is also important to decide main actors of communities. Moreover, it affects the computational complexity of the operations that we would perform on the network since it directly decides the number of nodes in the graph.

In order to decide it, we tried different values for this threshold via considering that users can place a tile for every 5 minutes and length of the time interval which events took place and consequently we decided that "3" is a reasonable choice to characterize significant interactions between communities in the network that we would create.

3.2 Concatenated - Color Match - Unweighted

In order to visualize the user interactions in the most abstract way, we initially tried to concatenate events and form single interaction context. To analyze it, we created a unweighted graph with color match criteria. Consequently we get a strongly connected graph that basically shows users that placed a tile with same color at least once. Eventually, we found modularity as 0.476, average clustering coefficient as 0.867, 1 connected component, network diameter as 3 and graph density as 0.241. Note that, community detection according to modularity and Louvain algorithm gave same communities.

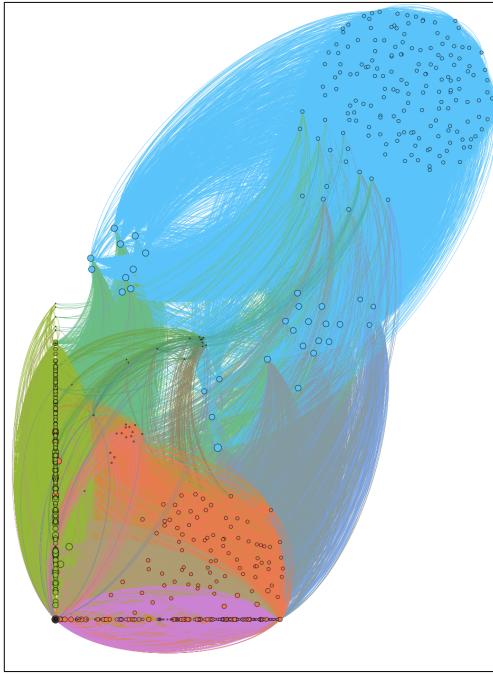


Figure 20: Concatenated - Color Match - Unweighted Network

3.3 Iterated - No Color Match - Weighted

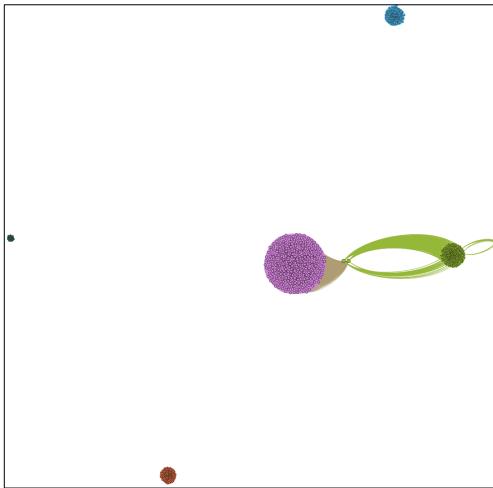


Figure 21: Iterated - No Color Match - Weighted Network

With "Iterated + No Color Match + Weighted" configuration, we aimed to reveal users who are interacted in different events (contexts) since we created edges between user nodes via iterating over events and no color match criteria ensures to capture all interactions. Also, we employed our weighting strategy to detect inter-community strength as well as revealing users that blends into different communities since edge weights are also computed in accordance with context. Consequently, we found modularity as 0.479, average clustering coefficient as 0.999, 4 connected components, network diameter as 3 and graph density as 0.283. Also, community detection according to modularity and Louvain algorithm gave same communities.

3.4 Iterated - Color Match - Unweighted/Weighted

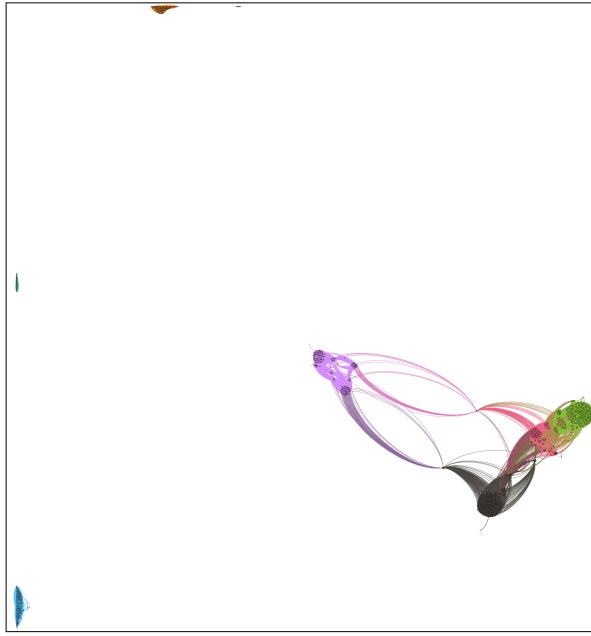


Figure 22: Iterated - Color Match - Unweighted Network

Eventually, in order to detect communities that collaboratively creating artworks or demolishing others', we created a network with "Iterated - Color Match - Unweighted" and another network with "Iterated - Color Match - Weighted" configurations. With these networks, we tried to capture communities mainly via color matching in different events via iteratively creating the network. With unweighted/weighted alterations, we tried to understand which model would represent the complex structure in a more modular way.

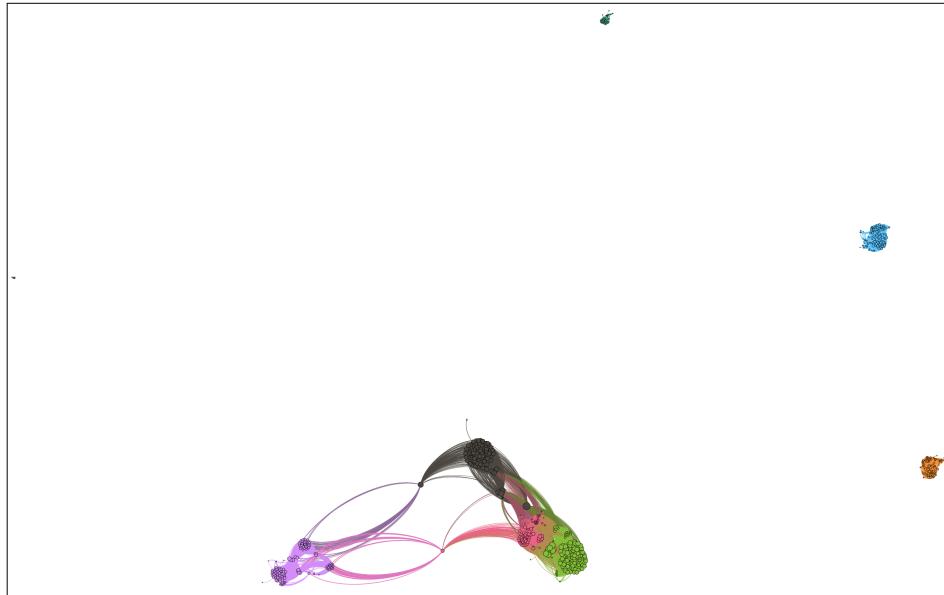


Figure 23: Iterated - Color Match - Weighted Network

For unweighted network, we found modularity as 0.737, average clustering coefficient as 0.888, 7 connected components, network diameter as 6 and graph density as 0.085. For weighted network, we

found modularity as 0.751, average clustering coefficient as 0.888, 7 connected components, network diameter as 6 and graph density as 0.085. Note that, modularities of these two network are significantly higher than the networks on sections 3.1 and 3.2. Also, community detection according to modularity and Louvain algorithm gave same number of communities (10 communities) with same labels on same nodes.

4 Discussion

First of all according to our findings, in order to analyze and visualize interactions between different communities over different events, creating a network via constructed iteratively would be more expressive since interactions that took place throughout the event would be more accurate.

Since we want to detect communities, we aimed to find a network with high modularity and low density. In this sense, with also interpreting these networks with our contextual knowledge of r/place 2022 event, we decided to use the network that is created with configurations "Iterated - Color Match - Weighted" (Figure 23). Also, visually, it represents communities better than others over given set of events. Resulting network can be interpreted contextually as follows however, note that these interpretations are only speculations since user information is hashed which means we cannot check the correct identity of users:

- Large network (Figure 24) represents the interactions between main communities that involved in given set of events which are, Turkish community, xQc community, r/osuplace community and Greek community. Communities that are too close to each other but had different colors probably sub-communities (for example Turkish community could be divided into many sub-communities which one and large of them is Twitch streamer Elraenn's community) or mutual supporting communities (Turkish - Greek communities).



Figure 24: Large Network

- Since the network is created via considering color match for edge placement, probably, these three communities (Figure 25) shows Turkish sub-communities and Greek community.

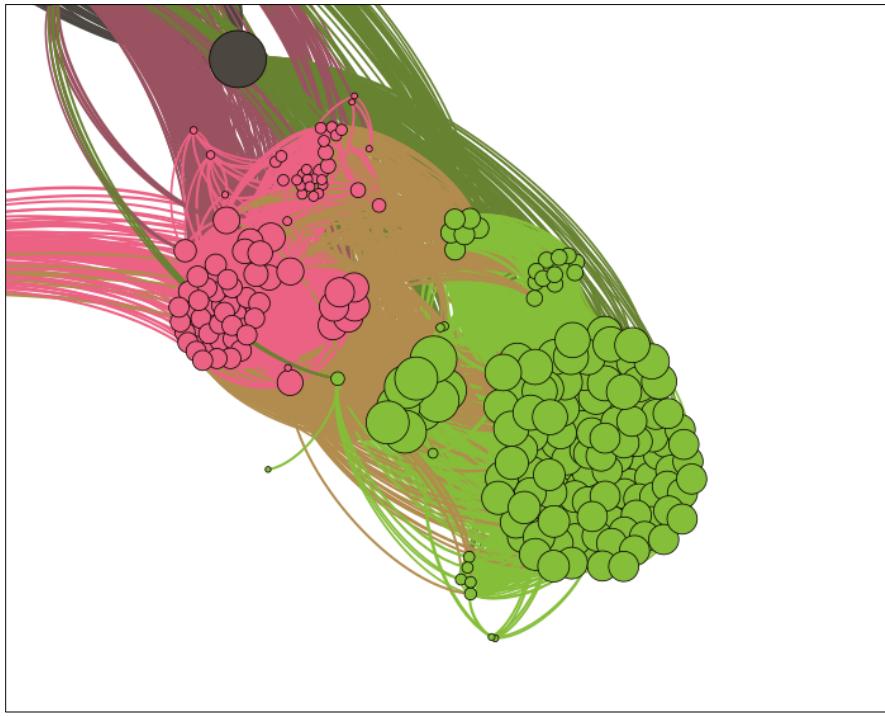


Figure 25: Turkish-Greek Community

- Since it seem standalone but had some interactions that have same colors with the Turkish-Greek community, this community (Figure 26) is probably the r/osuplace.

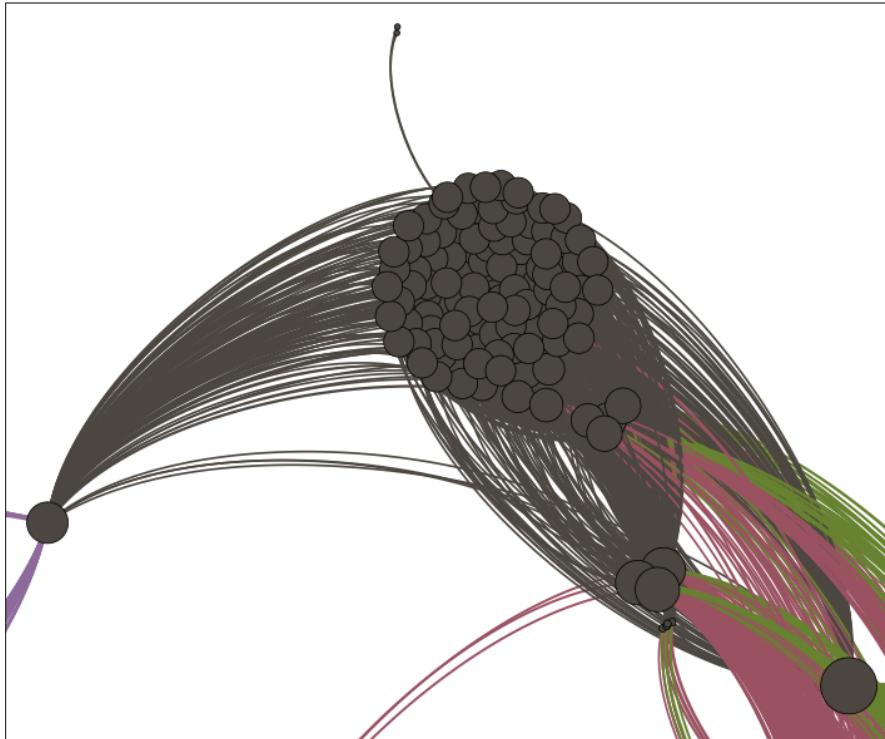


Figure 26: r/osuplace community

- With different hubs and having nearly 0 color match with other communities, this community

(Figure 27) is probably the xQc's community which attacks other communities. Hubs may indicate this community's internationality since users may do behave differently in different contexts.

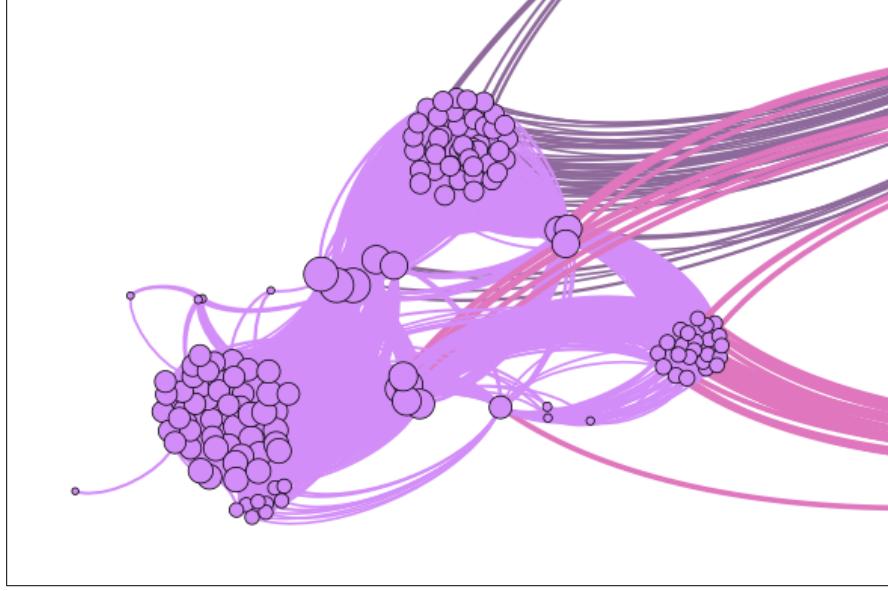


Figure 27: xQc community

- Following communities (Figure 28-29-30-31) are probably communities that are organized in small numbers and try to participate attack events as attackers or defenders.

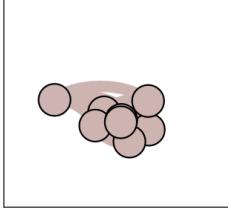


Figure 28:

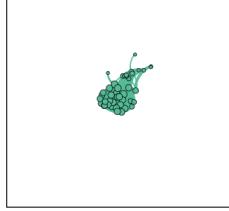


Figure 29:

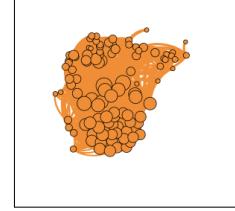


Figure 30:

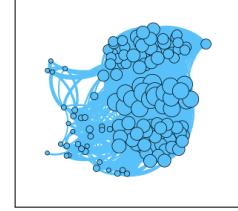


Figure 31:

In this sense, we may distinguish large attacking communities and defending communities via checking their interactions with other communities in following way:

- If one community has many edges (same colored tile placement) between another community, this indicates that these two community support each other in different events (contexts). On the other hand, if one community has less connections with other communities this community would be an attacking community.

However, note that these argument is valid for events that are classified as attacks, it is highly possible to be not hold for random events.

5 Conclusion

So far, we discussed a way to analyze community interactions over different events and identify roles of communities as attackers or defenders/supporters. We found that there exist communities and their sub-communities involved in attacks as well defences in different contexts. Therefore it can be

argued that communities have various motivations to manipulate different regions in canvas also they react to others' actions on them which supports that unrestricted user interactions should be further investigated in order to reveal various collaboration patterns.

In future works, one can try to analyze whole canvas and investigate various simultaneous interactions, rather than attacking/defending, to get better understanding on r/place data with a more powerful system. While investigating more events, one can try to use higher tile placement criteria. Also, one can try to detect attack events via employing conflict region technique that was mentioned in the "Analyze Approach" section (Section 2.2) to make event selection process more efficient.

References

- [Arm18] Benjamin Armstrong. Coordination in a peer production platform: A study of reddit's /r/place experiment. 2018.
- [BGLL08] Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008, 04 2008.
- [New06] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
- [RCWA18] Jérémie Rappaz, Michele Catasta, Robert West, and Karl Aberer. Latent structure in collaboration: the case of reddit r/place. 04 2018.
- [red] reddit.com. r/place datasets (april fools 2022).
- [VLCY20] Prateek Vachher, Zachary Levonian, Hao-Fei Cheng, and Svetlana Yarosh. Understanding community-level conflicts through reddit r/place. pages 401–405, 10 2020.
- [wik] wikipedia.org. Wikipedia:understanding wikipedia/'s content standards.