Differing emoji utilization across communities final report

Our central research question was looking at how emoji use changes over time, within different communities or as part of various short trends and memes. We hypothesized that there are a “core” group of emoji that have consistent use over time and the remainder being relatively unused or having peaks of popularity. We also hypothesize that the non-core emoji will be mostly limited to a small number of groups. This would be a useful tool for researchers looking into how ideas enter and spread through social networks, or just for curious individuals who use a lot of emojis.

Our initial plan was to show the emoji usage across groups in a basic bubble enclosure diagram. Those emojis would be surrounded by circles which represent accounts across the different platforms. Those circles could then be categorized inside of bigger circles to represent communities, or trends. The size of the circle indicated the frequency of that emoji or totals of emojis inside that circle. We planned to represent temporal data by allowing the user to click on an emoji, which would bring up a line graph in the corner of the visualization showing usage over time. Up to 3 emoji would be displayed on the line graph and each will have a color red, green, or blue. When the user hovered over the line graph with their cursor, the main circle graph would be colored by relative frequency of the emoji in the circle. This would help us prove or disprove our hypotheses by looking at the color of the main visualization over time. We could also look at the individual line graphs to classify emoji as core or non-core. To make sure this visualization was effective, we would ensure that labels for circles are visible and that emoji appear large enough on-screen. Since we wanted to show the whole visualization on one screen, with a huge amount of data points, we also planned to have a click-and-zoom feature.

When it actually became time to implement our data, we encountered a lot of issues. When we tried to implement the temporal data via the line graph, we found some large gaps in our data. This was probably caused by the fact that twitch had to be recorded continuously in a live stream of data, and any power/internet losses meant a lack of data for that, but it meant our temporal representation had jagged jumps for when we lost data. We considered smoothing the data to make it look nice, but we felt that would misrepresent the data even more.

Another issue we had was with the pure size of the data. With thousands of svg elements the browser would always become unresponsive as shown in figure 1. To fix this, we had to reduce our data considerably. We cut down the number of users, and emojis displayed by pruning our data for only the top elements of each category. We were able to fiddle with the our maximum set numbers for this to balance performance and visualization accuracy.

In addition to this, the packing algorithm we were using, tried to efficiently use space by having differing parent circles encompass the same sub circle. When the visualization reached the size that we demanded, that algorithm failed. It displayed circles wrong, as well as their labels. In figure one you can see that some grey circles have other circles inside of them. This should never have happened. While the sharing of sub circles for parents seemed like a good idea, in the end it hurt us. We had to change our visualization to a more basic packing algorithm, and update how the labeling was handled.

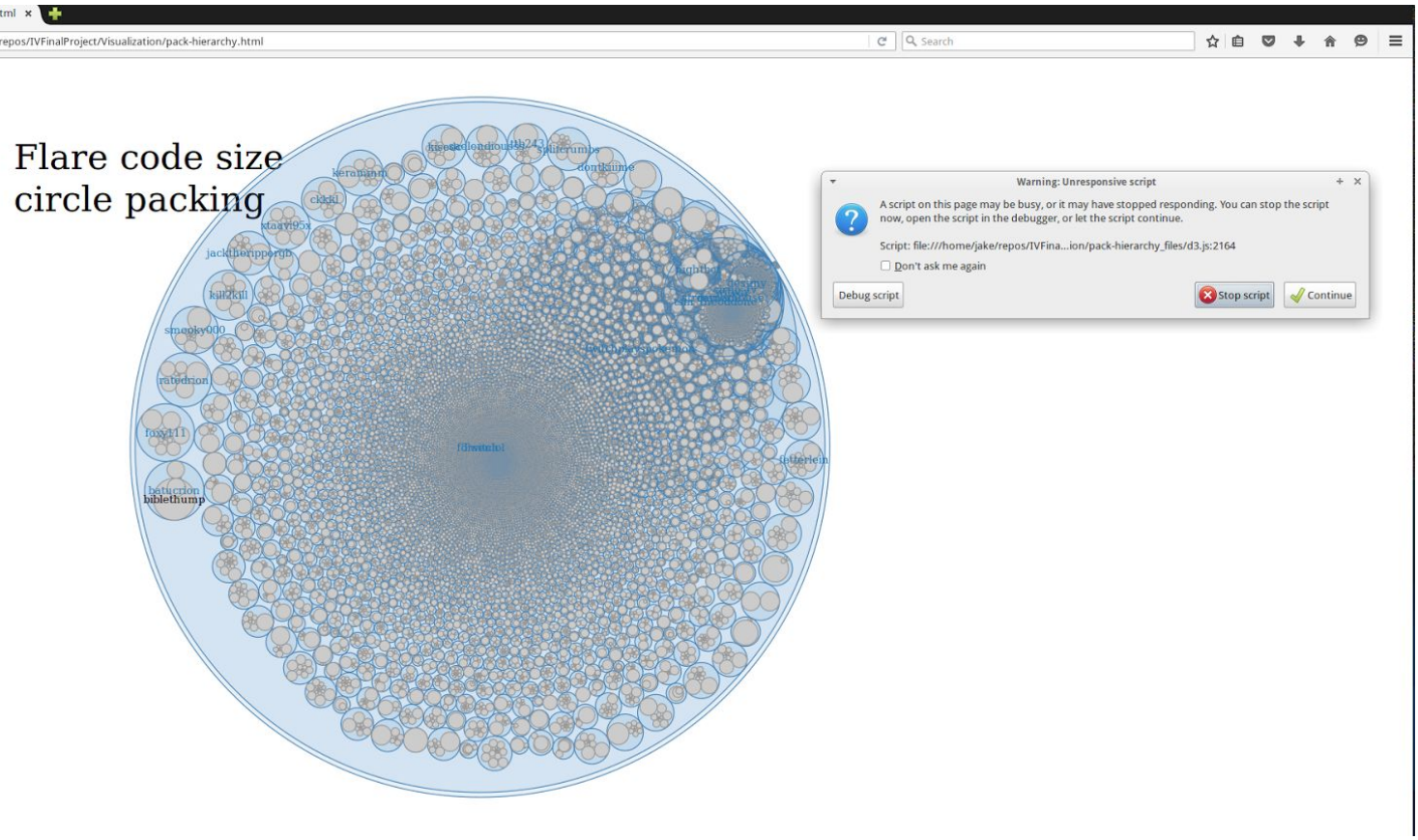

Figure one: Shows incorrect circle packing and the browser becoming unresponsive.

While using our visualization, we also saw the same emojis coming up time and time again. In this way we didn't see that much difference between the different groups. We considered removing the top used emojis to help users identify the frequent emojis that were unique to a specific group, but this would skew the data and portray an inaccurate representation. To counter this we added a table to the side of the page that showed the top emoji's and their counts. This way analysts could identify and subsequently ignore the emoji's when looking for emoji usage that was specific to individual channels.

The last major issue we encountered which were unfortunately unable to fix was the low resolution of the images we used. Because the images were only meant to be showed at a low resolution in chats on screen, the low resolution was amplified when users zoom in our