What is the Most Loving Gaming Community?

CS 6630

Basic Information

Team Information

Name	uID	Email
Leo Kell	u1266030	u1266030@utah.edu
Michael Northrup	u0992000	u0992000@utah.edu
Paul Wissler	u0791342	u0791342@utah.edu

Project Repository

The project repository may be found here:

https://github.com/lkell/Gaming-Community-Toxicity.

Background and Information

We enjoy playing video games and have a diverse set of gaming interests. It is popular to find gaming communities on Reddit due to the site's popularity as well as its "subreddits" that make it easy to form niche communities. In our experience, gaming communities tend to have varying levels of toxicity, though. We are interested if their online communities, particularly how they interact with other communities, indicate general community toxicity.

Although measuring toxicity is a complex topic, we have chosen to use Sentiment Analysis (SA) data because members of our group are interested in exploring this kind of data. This project serves as a method for learning how to visualize and derive insight from SA data.

Project Objectives

The primary objective of this project is to create a series of interactive visualizations to explore the differences, connections and overall trends of post sentiment between different gaming subreddits. We break this into three different questions/objectives:

- How do different subreddits interact?

- On average, how positive/negative are subreddits?
- Which subreddits are the most loving/toxic?

The benefits include:

- Understanding the behavior of subreddits over time
- Identifying which subreddits are more loving and which are more toxic

Data

We are currently using data from the following location:
https://snap.stanford.edu/data/soc-RedditHyperlinks.html

For the time being we are using the 'body' and 'title' TSVs.

Each row represents a single post from a source subreddit that contains a hyperlink to a target subreddit. One of the included attributes is a calculated VADER sentiment score of the post. The key assumption we make is that large-scale sentiment averages model the overall "like" or "dislike" between subreddits.

We also created an API to mine comments from the top 100 hottest posts for our curated subreddits. This data set also includes a VADER sentiment analysis, as well as their Automated Readability Index.

Data Processing

For the Stanford Link Sentimetn data set, we filtered out a significant portion of the data, i.e. most of it. Data processing was handled in Python, and we converted the TSVs to JSON files.

Major processing steps include:

- Selecting a subset of data based on 'source' subreddits
- Indexing the data by Post ID and Source Subreddit
- Breaking out the PROPERTIES column of CSV data into their own columns

Sentiment scores are already calculated in the dataset, so we currently see no need for derived values. We originally considered aggregating the data first, but realized this was

unnecessary due to how much the data set shrank upon filtering it. We also decided to let D3 handle data aggregation because we wish to visualize underlying distributions.

For the comment data set, we mined data via the Reddit API, grabbing the top 100 hottest posts and performing VADER sentiment analysis and calculating the Automated Readability Index. The general structure was intentionally designed to be similar to the Stanford Links data set.

Visualization Design

Summary of Design Process

The first alternative design is seen in Figures 2,3. Second in Figure 4. Third design in Figure 6. The final design incorporates Figures 5,7,8.

We began the design process by having a meeting where we began an overall design discussion, ending up with the idea of making a node diagram sit at the center of our primary visualization (Figure 1). Strictly speaking, this node diagram would not be a good visualization by itself for analyzing the sentiment analysis data, but makes for a good controller to dig into different portions of the data, which is why we decided to include it.

After this initial meeting, we all made our own alternative visualization designs. The first of these can be seen in Figures 2,3, the second in Figure 4, and the third in Figure 6. We then brought these together in Figures 5, 7, 8, splitting the visualization design into three separate views.

In the brain storming stage, several views were considered:

Node/Network View - Must Have

The core of the design includes a network visualization of all of the connections between the different subreddits, showing how each of the subreddits are connected, allowing it to serve as a good controller view. It also naturally leads into one of our optional features, that being a graph view that provides a larger picture view of how all the subreddits we chose interact with each other.

Summary View - Must Have

This view would have summary charts detailing the overall behavior/sentiment of the selected target view. The design sketches can be seen in Figures 4,5. We plan on generating three charts, all looking at complete data for the selected source subreddit. We will show a violin plot for the distribution of love/hate on a diverging scale using sentiment analysis, a line chart for how the number of cross-subreddit posts have changed over time, and a bar chart breaking out underlying sentiment scores, colored according to whether we judge they fall into love or hate.

Overall Ranking View - Must Have

A view that shows ranked summary visualizations for all of the subreddits. Can see the initial sketch in Figure 3 and a more final version in Figure 9. It shows time series, distribution, and aggregate data for love/hate metrics for selected subreddits over time.

Relationship View - Optional 1

A view of the relationship between two explicitly selected subreddits. Seen in Figure 4 tooltip. This would generally be a filtered version of the Summary View, looking only at how the source subreddit interacts with the target subreddit.

Graph View - Optional 2

This view should offer a "zoomed out" overview for all the connections between all subreddits, which includes the outgoing and incoming sentiment averages. This view shows overall landscape and should provide surrounding context for the more "zoomed in" Node/Network view. In fact, ideally this view would transition to the Node/Network view when an individual node is selected. There are several ways of visualizing the overall network, including as a force-directed network or adjacency matrix; see Figure 8 for several candidate sketches.

Implementation

The final implementation was split between two pages: the Home page and the Ranked page.
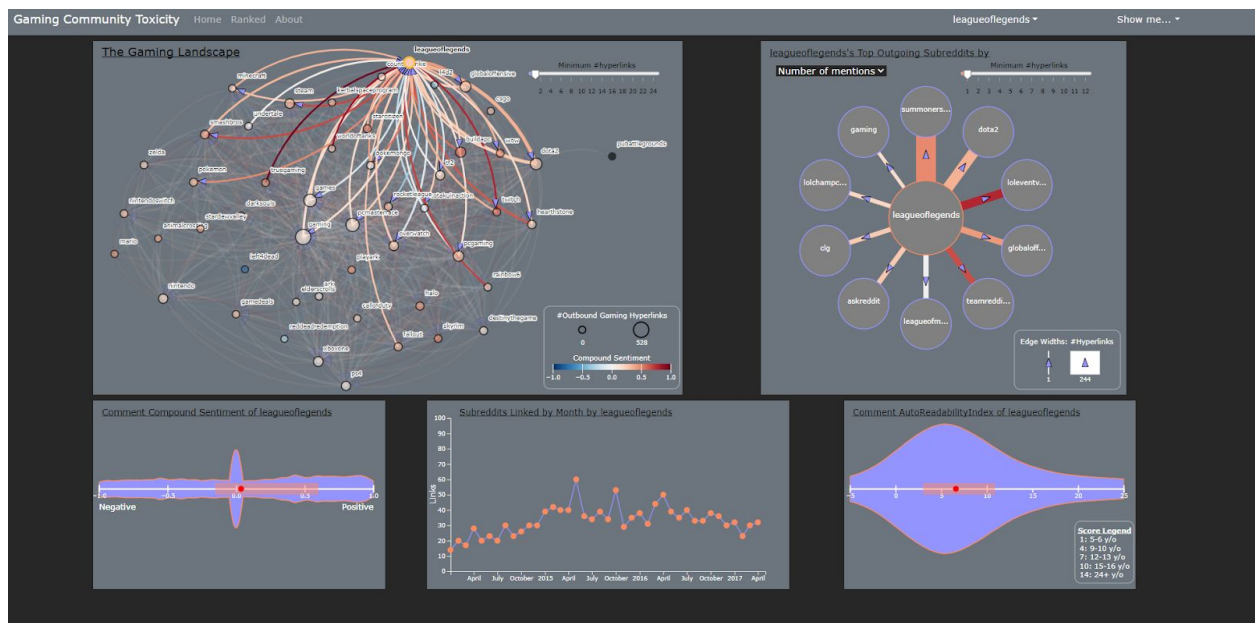
Figure Imp. 1: The Home Page. Top left graph gives node connections between curated gaming subreddits. Top right chart gives the most referenced subreddit by selected subreddit. Bottom left chart gives the compound sentiment distribution of the selected subreddit. Bottom center chart gives the links to other subreddits by source subreddit over time, by month. Bottom right chart gives distribution of the automated readability index.
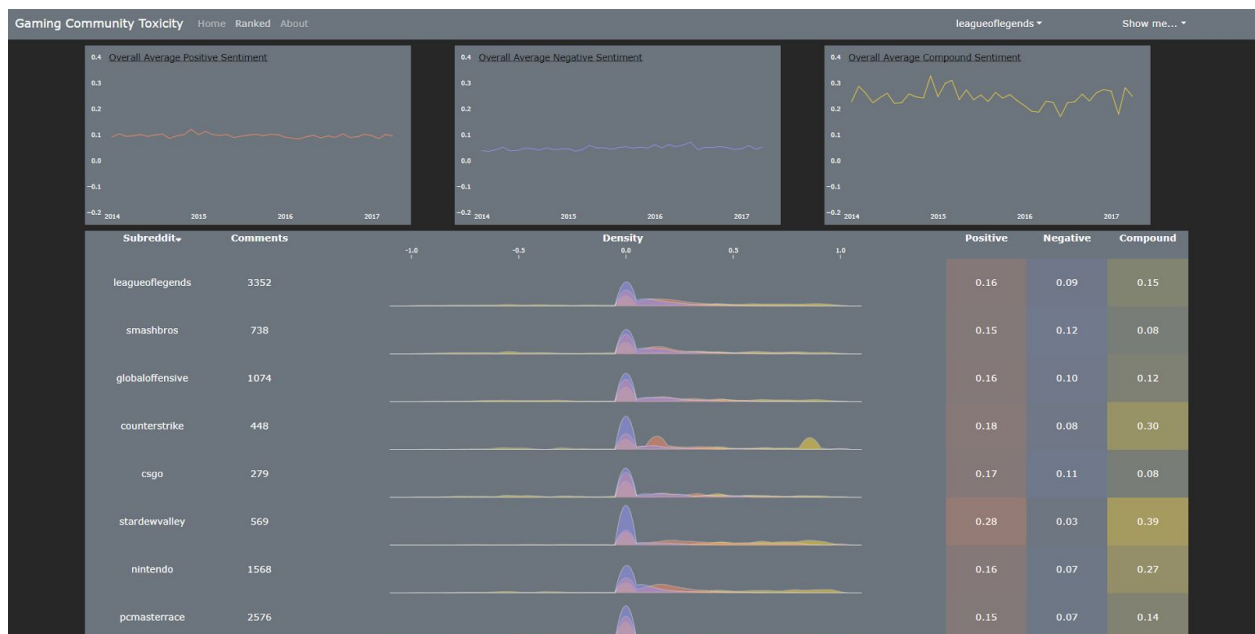

Figure Imp. 2: The Ranked Page. Top right chart gives the positive sentiment of all curated subreddits over time for link data. Top center chart gives the negative sentiment of all

curated subreddits over time for link data. Top right chart gives the compound sentiment for all curated subreddits over time for link data. Bottom view gives a breakdown of each subreddit's sentiment densities, averages, and links to other subreddits, with sentiment data computed from comment data instead of link data.
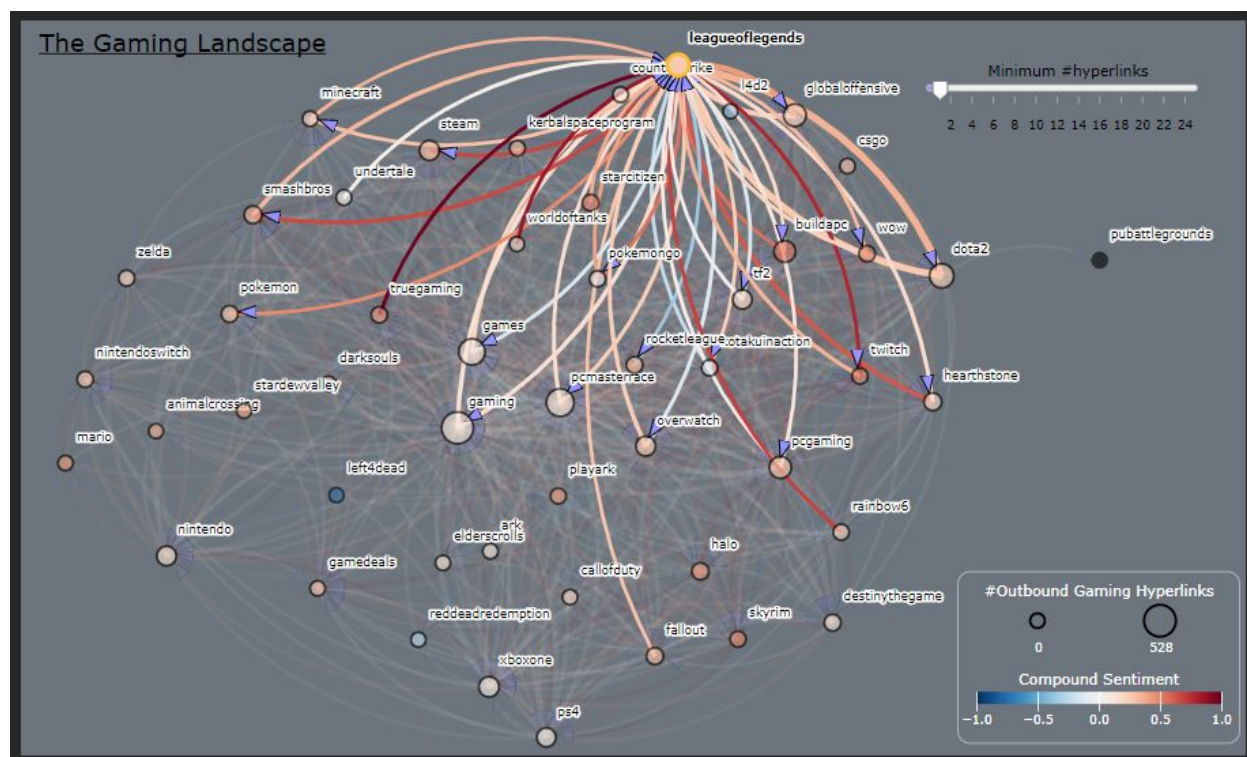

Figure Imp. 3: Force-directed network plot of gaming subreddit interactions. Node size and edge widths encode the number of subreddit-to-subreddit hyperlinks contained in reddit posts, while node and edge colors encode average sentiments of the associated posts. Hovering over a node highlights the node and surrounding edges, while clicking on it selects the associated gaming subreddit for all charts in the Home Page. The slider at the top right removes all edges containing fewer hyperlinks than the specified threshold.
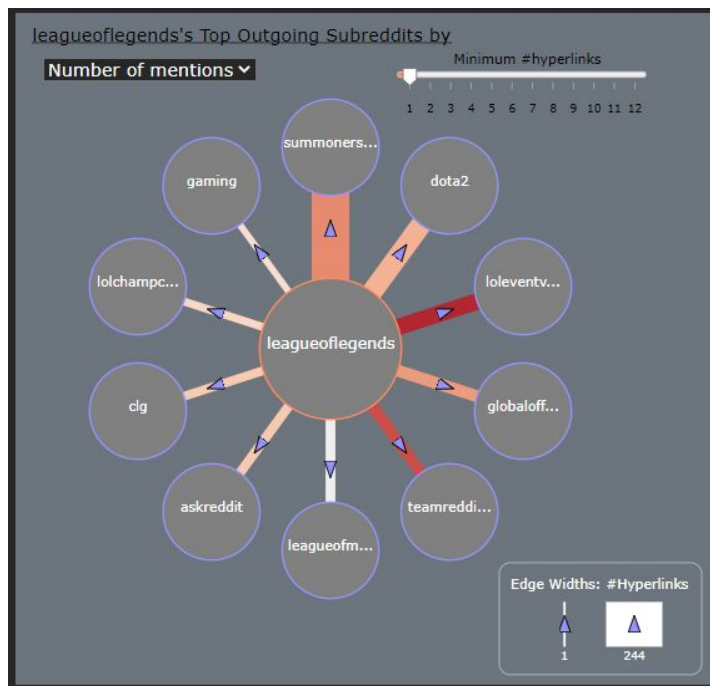
Figure Imp. 4: The node chart displays the top-linked subreddits from the selected gaming subreddit. The criterion for determining the top-linked subreddits can be interactively determined with the dropdown. Similar to the network plot, this chart also has a "Minimum #hyperlinks" slider. In addition, hovering over a node renders a tooltip which details the number of hyperlinks and average between the selected subreddit and target subreddit.
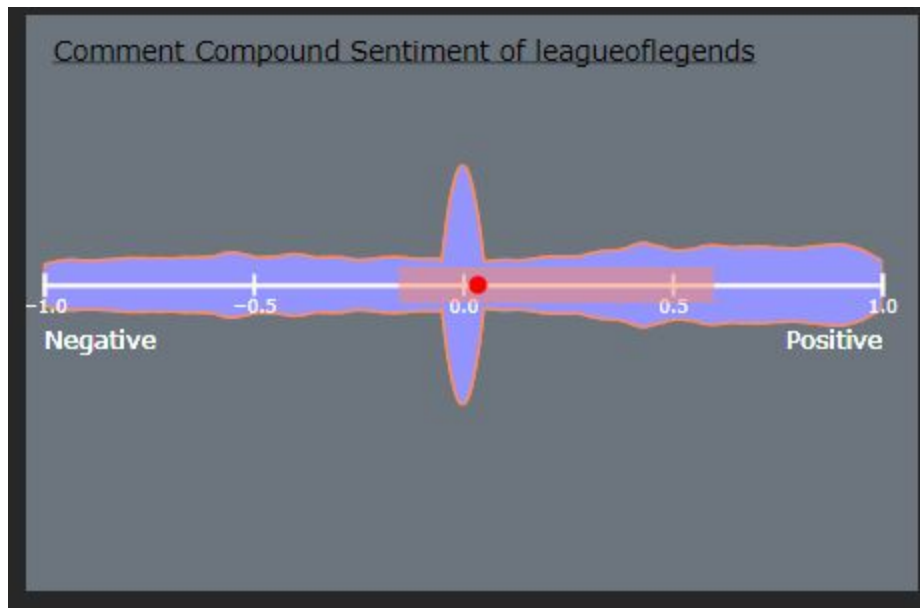

Figure Imp. 5: Violin plot of comment compound sentiment, with interquartile box in orange and median as a red dot. Hovering over the median renders a tooltip with the

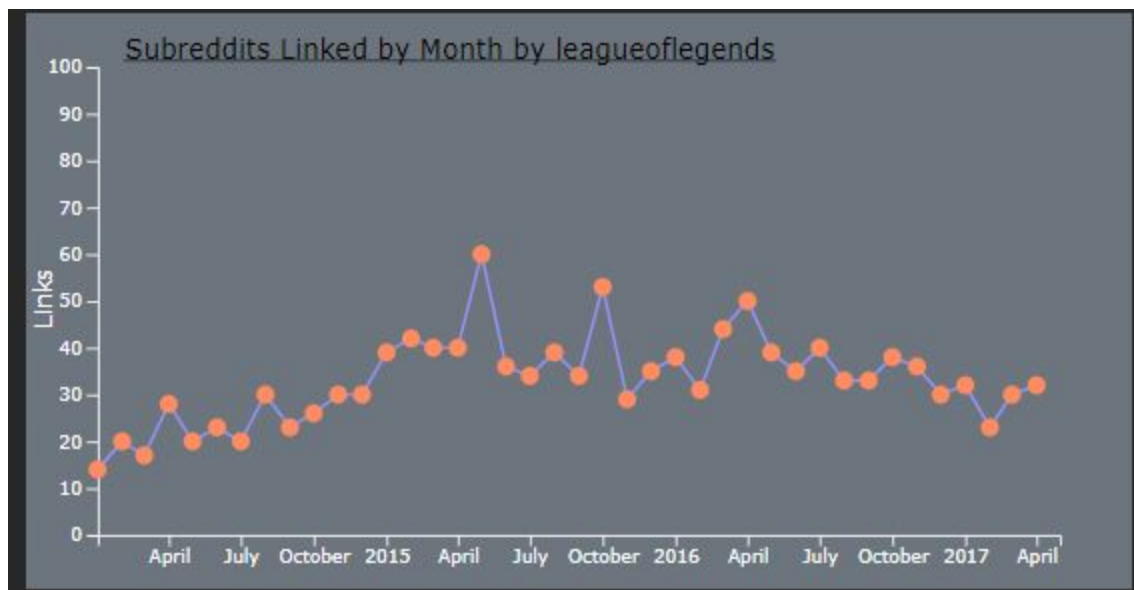median value, and hovering over the interquartile box renders a tooltip with other interquartile metrics.

Figure Imp. 6: Time-series plot of links from selected subreddit to other subreddits over time. Hovering over the orange dots renders a tooltip giving the total number of links in a given month.
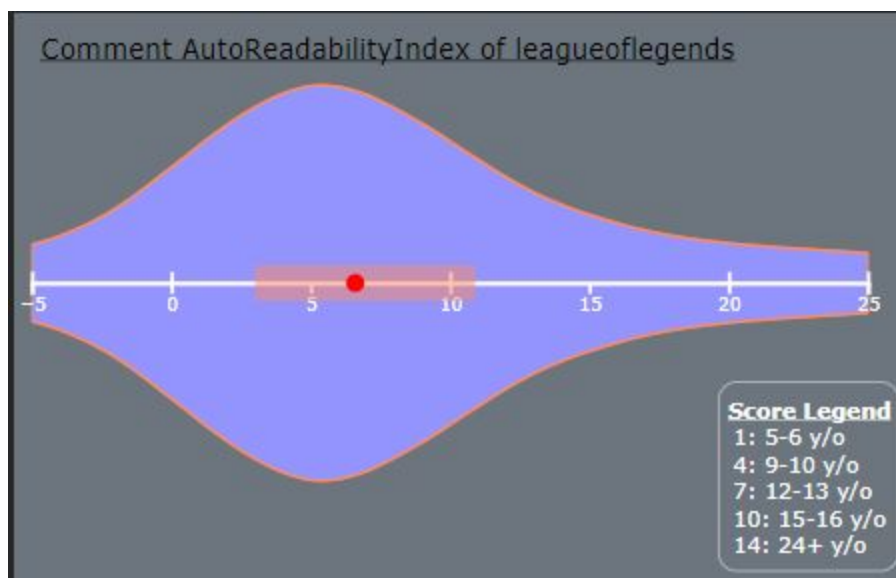

Figure Imp. 7: Violin plot of comment automated readability index, with interquartile box in orange and median as a red dot. Hovering over the median renders a tooltip with the median value, and hovering over the interquartile box renders a tooltip with other interquartile metrics.

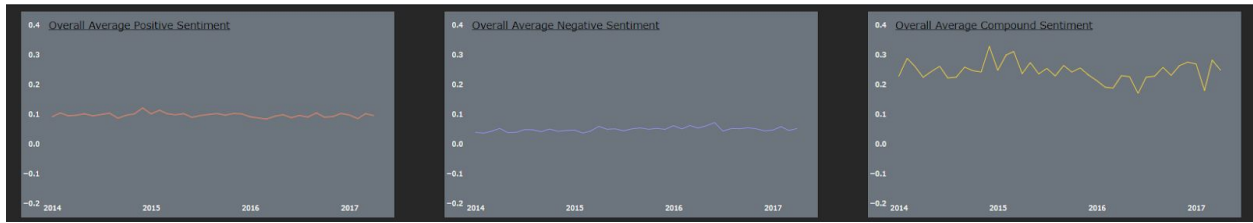**CS 6630 Gaming Community Toxicity**

Figure Imp. 8: Sentiment Link Charts. These charts show the weekly average for the positive, negative and compound sentiment for all subreddits using the links dataset. There is no further interactivity.
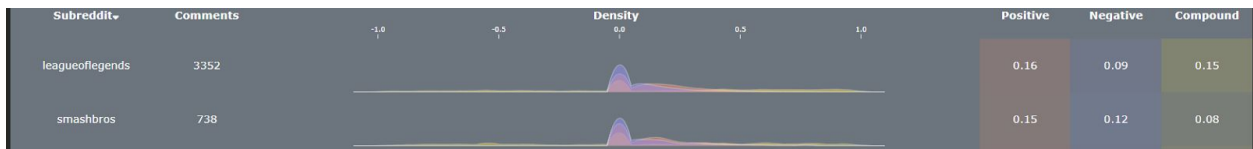

Figure Imp. 9: Ranked Table. The ranked table includes 6 sortable columns including the Subreddit name, the number of comments related to that subreddit, a density plot including the overlaid densities plots for the positive, negative and compound sentiment as well as heatmap columns with values for each. The columns can all be sorted and when a row is clicked the subreddit is selected in the main view.

Figure Imp. 10: Story-telling box rendered over graph view, giving some information on the selected subreddit from the "Show me . . ." drop-down, also with the correct subreddit selected accordingly.

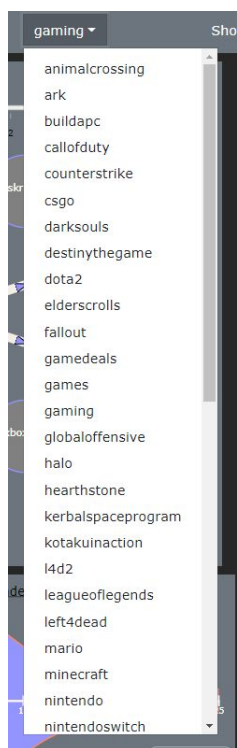**CS 6630 Gaming Community Toxicity**

Figure Imp. 11: Filter drop down allows individual gaming subreddits to be selected, which then updates the charts in the Home Page to display the selected gaming community.
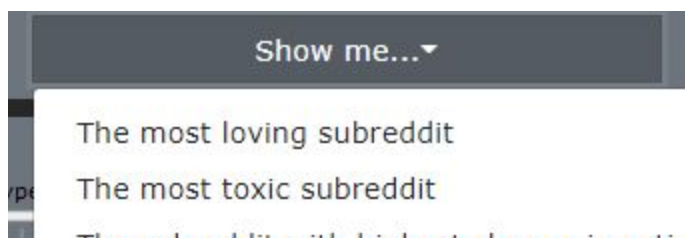

Figure Imp. 12: Drop-down to do story-telling that answers questions like what is the most loving subreddit. Once selected, selects the appropriate subreddit and displays a box over the graph view as seen in Figure Imp. 10.

Evaluation

We learned that the original links data set was too limited. Firstly because it had too few records for many subreddits, as the dataset was limited to only those posts that link to other subreddits, and secondly because there seems to be a bias towards positivity within the data set. Although we do not know why, exactly, we suspect it is because generally people will post links to things that they like, rather than to things that they dislike.

To overcome these limitations when analyzing toxicity and lovingness, we created a new data set using the Reddit API and NLP processes that looks at comments on the hottest posts in each subreddit. Through this, we were able to get a clearer idea of the distribution of sentiments of each subreddit.

Overall, our visualizations proved effective at visualizing the toxicity of different subreddits, both through seeing how they behaved toward other subreddits, as well as with how the subreddits behaved within themselves (via comment analysis). However, there were some limitations with our approach.

First, it is not entirely intuitive how the Automated Readability Index Violin Plot fits into the narrative. Initially, it might indicate a general education level, perhaps, or overall sophistication of the user base, but because the scale for the Automated Readability Index is not intuitive, it makes it difficult to interpret unless one has read up on it.

Second, the Links time series plot on the Home page only provides a relative level of activity. Although it scaled alright with links (aside from in one case with the PokemonGo subreddit), we suspect that if we switched to another method of measuring subreddit activity (e.g. number of posts per month or number of comments per month), the scale may be too disparate between low activity and high activity subreddits. This could perhaps be handled by switching to a logarithmic scale, but logarithmic scales are not intuitive.

Third, the overall sentiment time series plots on the Ranked page are so steady that they do not reveal any particularly interesting information over time. Perhaps if we switched to a comments data set, and added the ability to filter the time series by subreddit, we may find some more interesting information, but that would introduce a new problem, that being computing power. We only have so much memory and processing power for the website, and the larger our data set, the slower the website will be.

Appendix

Color: red for negative, blue for positive (gradient), arrows going in both directions

Size: Number of Hyperlinks increases size of target ndoe

Limit linked nodes to 10 for readability

Dropdown/search filter to select nodes?

What views should be linked to this viz?

Use logos for node representation (stretch goal)

Tooltips for more info (full name)

Legend for icons and names?

Measure "unrequited love"?

Limit source nodes to gaming, but allow outside targets
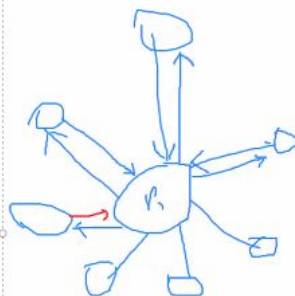
Holistic Love Score (divergent scale?)

Figure 1: *Initial sketch drawn up in Zoom call, brainstorming different possibilities. This is where the team also decided on the Primary Node View.*
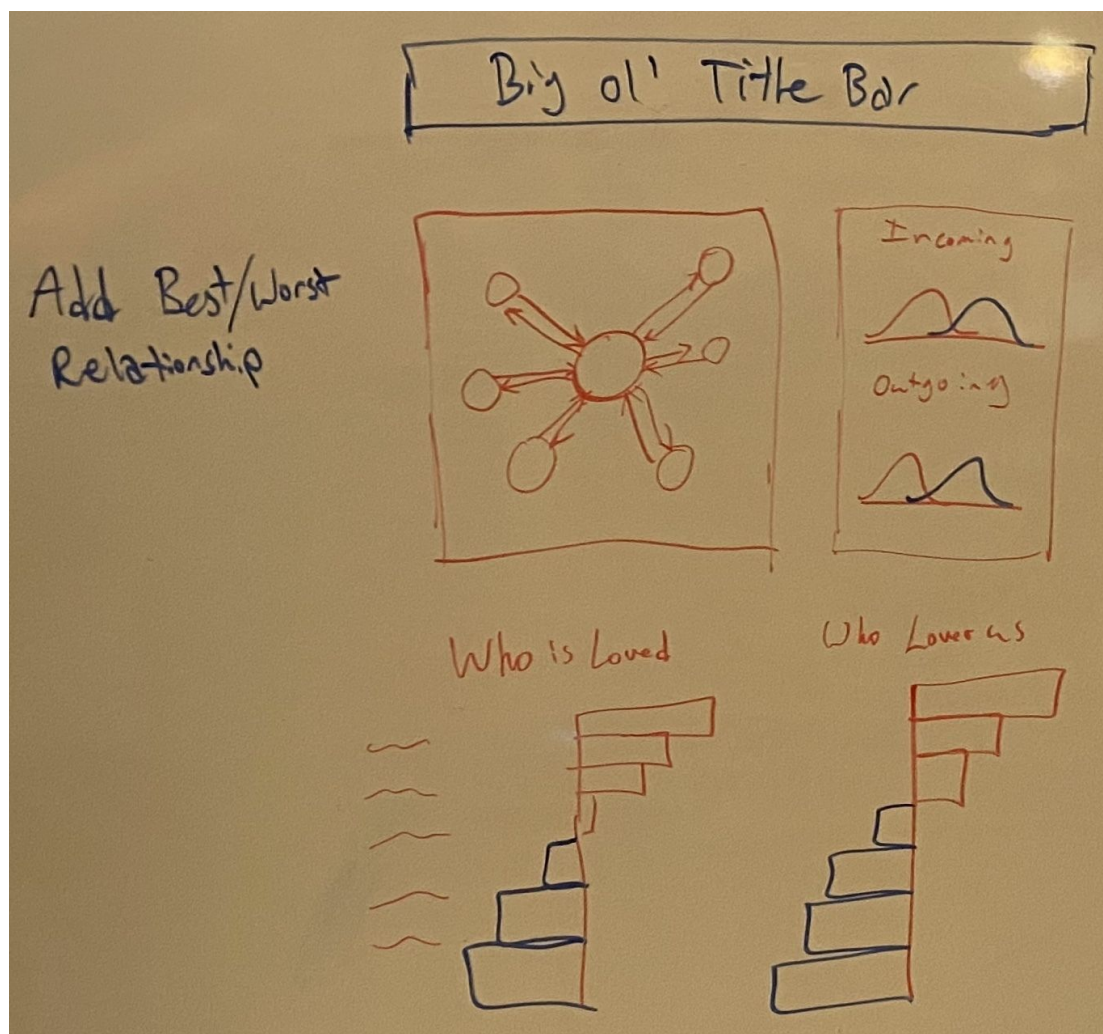
Figure 2: *Node view with double density plot to the right and summary below. One of three initial mockups.*
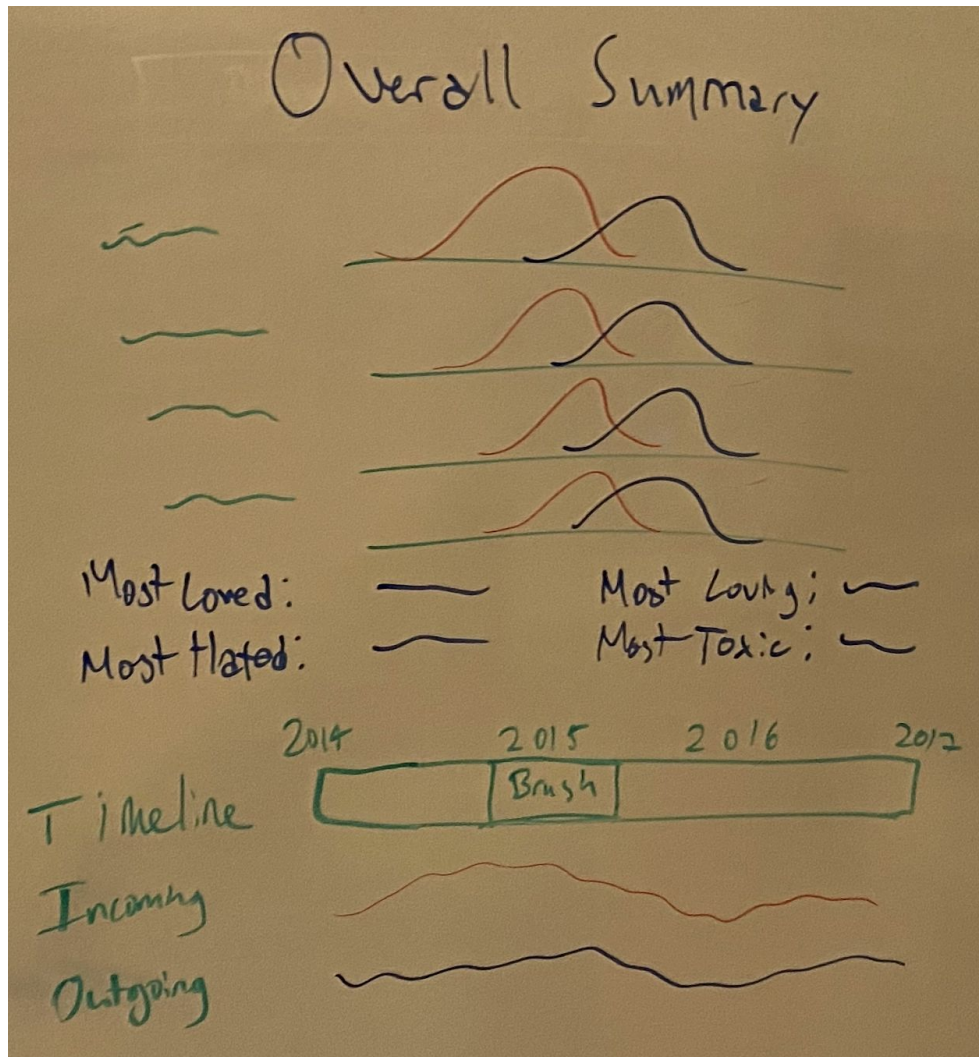
Figure 3: An additional view that would be a popup, an extension of figure 2. Includes overall double density plot for each subreddit and a number of time series charts including average values for different categories.

CS 6630 Gaming Community Toxicity

Figure 4: First attempt at a combined Node/Summary View. Also where Drilldown View was initially proposed. Can see Relationship View mock up in tooltip, as well as Summary View on bottom and Node View in the center.

CS 6630 Gaming Community Toxicity
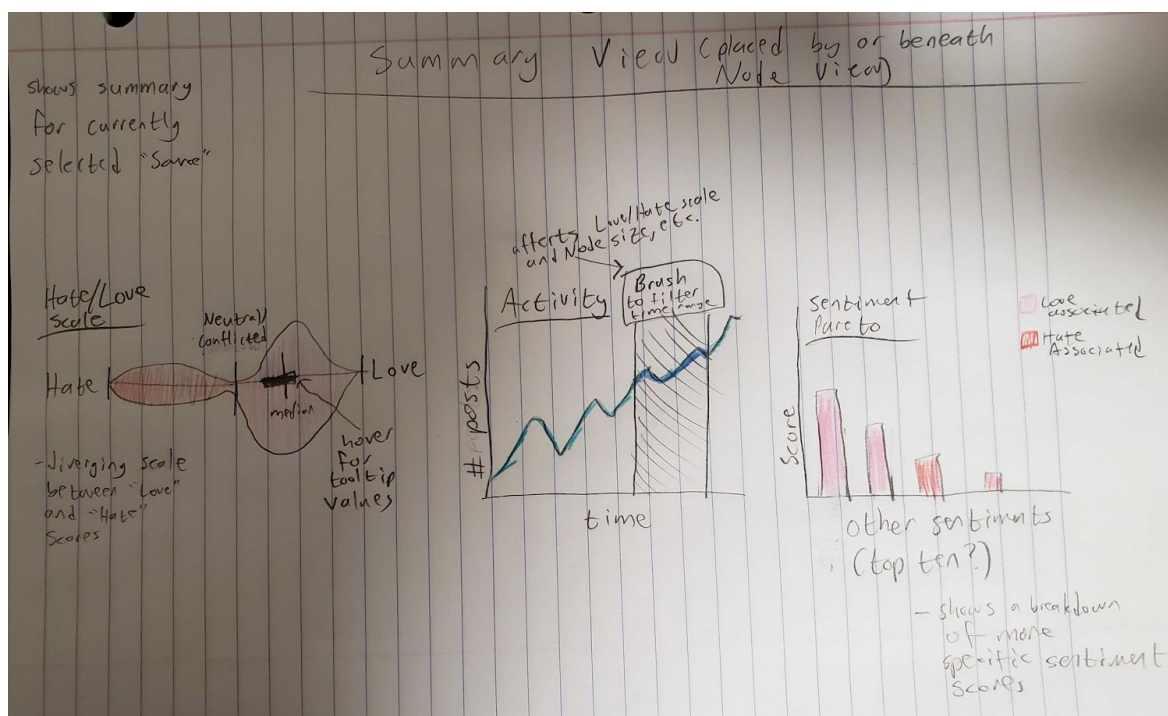
Figure 5: Second revision of Summary View, focusing on the three specific charts that would be useful for summarizing the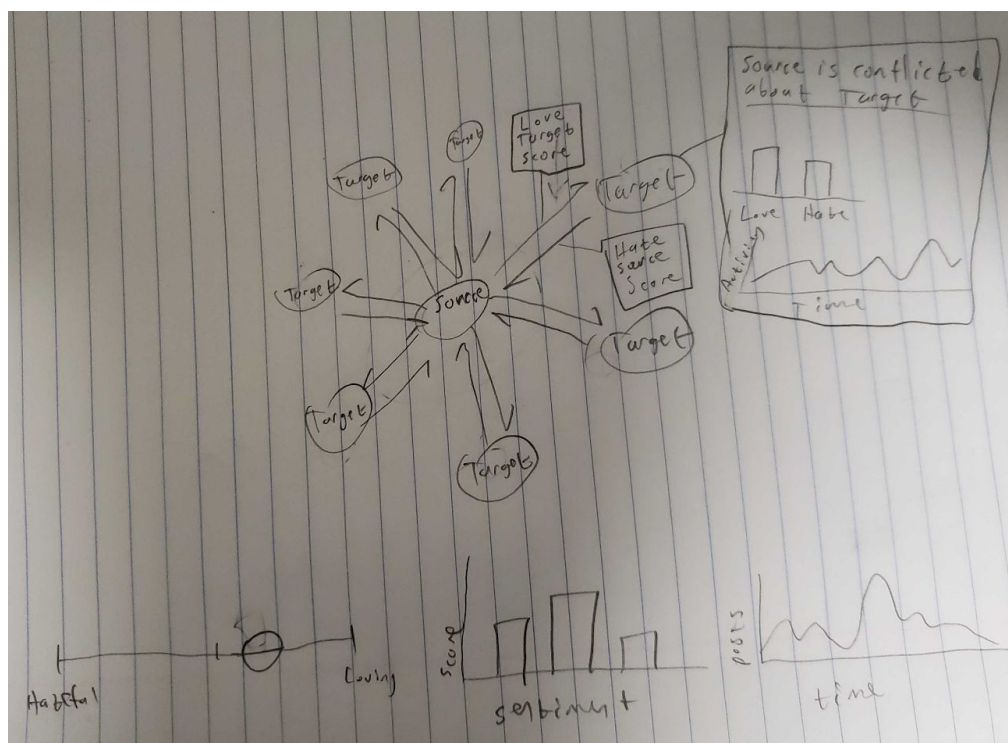 statistics of the selected "Source" node. Uses a Violin Plot on a diverging Love/Hate scale to measure general toxicity/lovingness of a community. Tracks the number of posts over time and may have use a brush in the Activity chart to filter data to a given time period. FInal chart is a bar chart breaking out the different kinds of sentiment and coloring them based on whether they align with Love or Hate. Also introduces a blue/red/pink color scale.
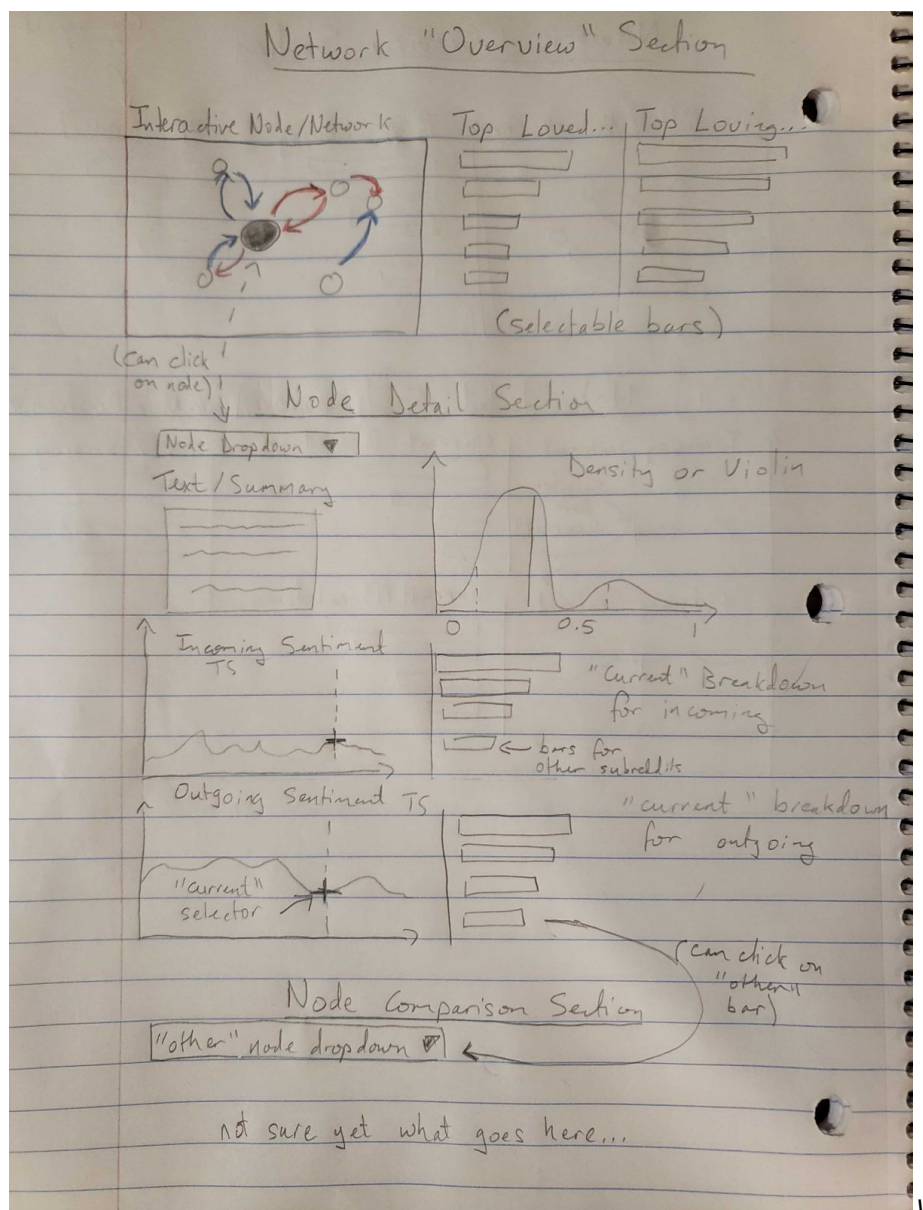
Figure 6: Candidate mock up of the visualization dashboard, also showcasing a hierarchical view layout. The "Overview" section displays the high level information and topological structure of the Reddit gaming landscape, whereas the "Node Detail" displays data centered around a single selected subreddit. A key feature of the "detail" section is the synchronized time series and snapshot bar charts. A third optional section, which is not filled in here, is the "Comparison" section.
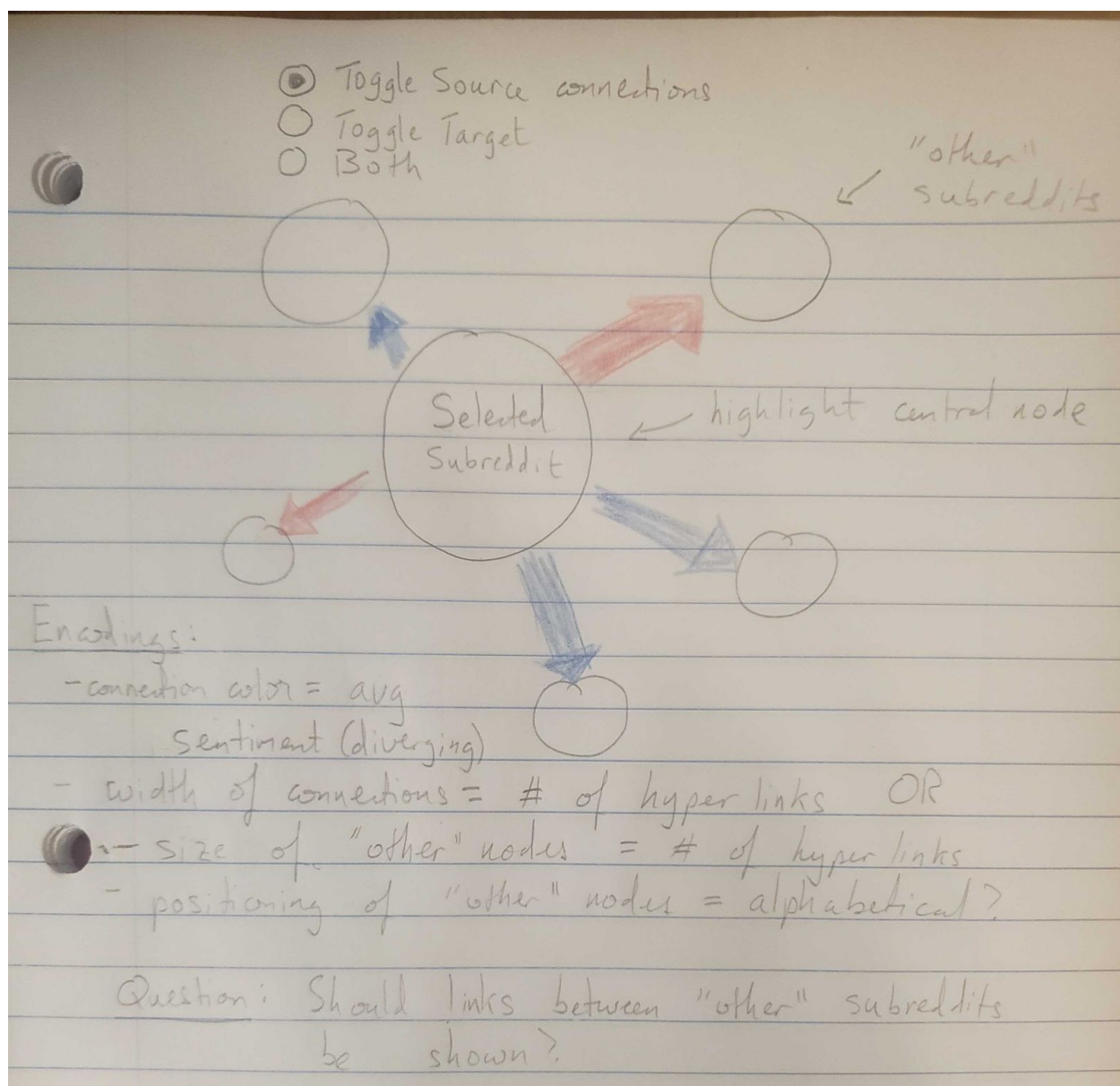
Figure 7: Sketch of a centralized Node/Network view. Outgoing or incoming sentiment connections are shown by arrows colored by average positivity or negativity. Clicking on "other" nodes would shift focus to them. Some of the encodings used are arrow color, arrow diameter, node size, and node placement.

CS 6630 Gaming Community Toxicity
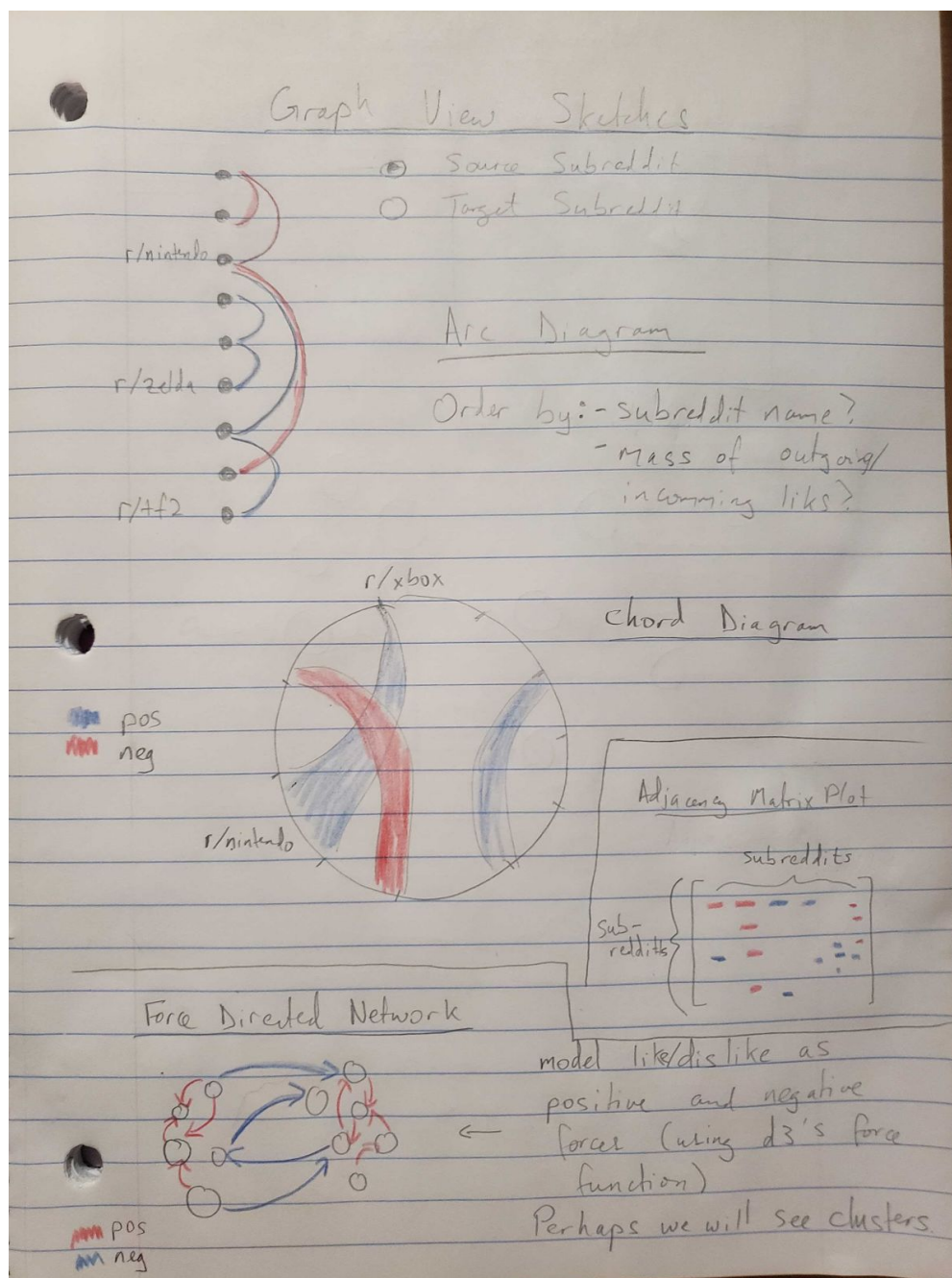
Figure 8: Graph View sketch. Four possibilities are shown: Arc Diagram, Chord Diagram, Force Directed Network, and Adjacency Matrix. Each of these would be interactive, letting the user select "source" or "target" view.
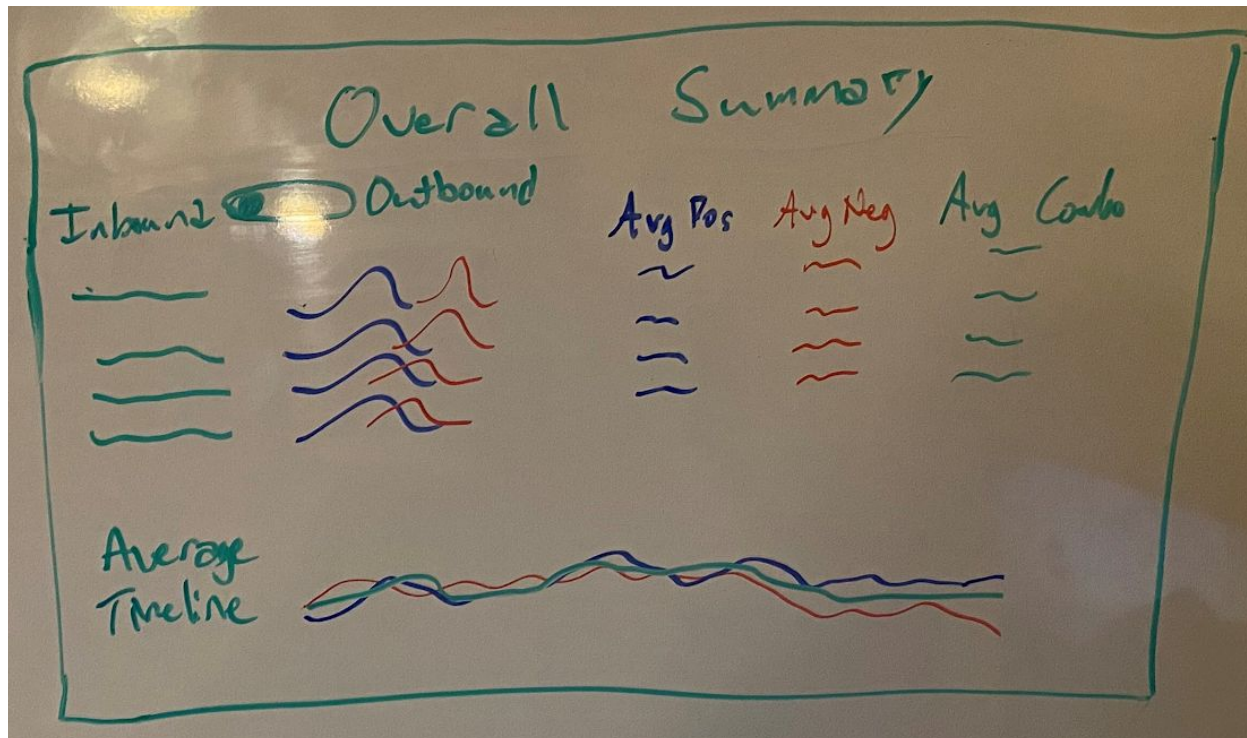
Figure 9: *Summary popup view. A summary button on the main page would cause this to come up. Includes a slider to show either inbound comments or outbound comments. Shows a double density plot of the positive/negative values for each subreddit, as well as a table with the average positive/negative and combined value. A time series with the average for each value below.*

Process Book

Detailing how the design changed over time

CS 6630 Gaming Community Toxicity

Data Processing - PJW - 10/27

SUMMARY

Focused on data processing using Python. Made an initial commit to GitHub repo.

Goals

- To create an initial data process prototype

Achievements

- Loading data from TSV
- Exporting data to JSON
- Began process of filtering data to appropriate subreddits, began discussion of inclusion/exclusion criteria on Slack
- Create and read list of subreddits from a YAML configuration file
- Create and read column mapping for properties from a YAML configuration file
- Broke out the PROPERTIES values into separate columns
  - Current column count is now at 92
- Got data_processing files into GitHub

NOTES

- Data for most sentiment analysis is split into comma separated values in the PROPERTIES column, need to spread these out and give them appropriate names
- Collating different source subreddits from Michael and Leo
- **MAKE ABSOLUTELY CERTAIN NOT TO COMMIT THE TSV!!!!!**

ACTION ITEMS

- Meet with Michael and Leo to discuss inclusion/exclusion criteria
- Add aggregate columns (post counts, average sentiment scores, etc.)

- Consider whether to use body or title TSV data or both

FIGURES

Average Sentiment from Top Linking Subreddits
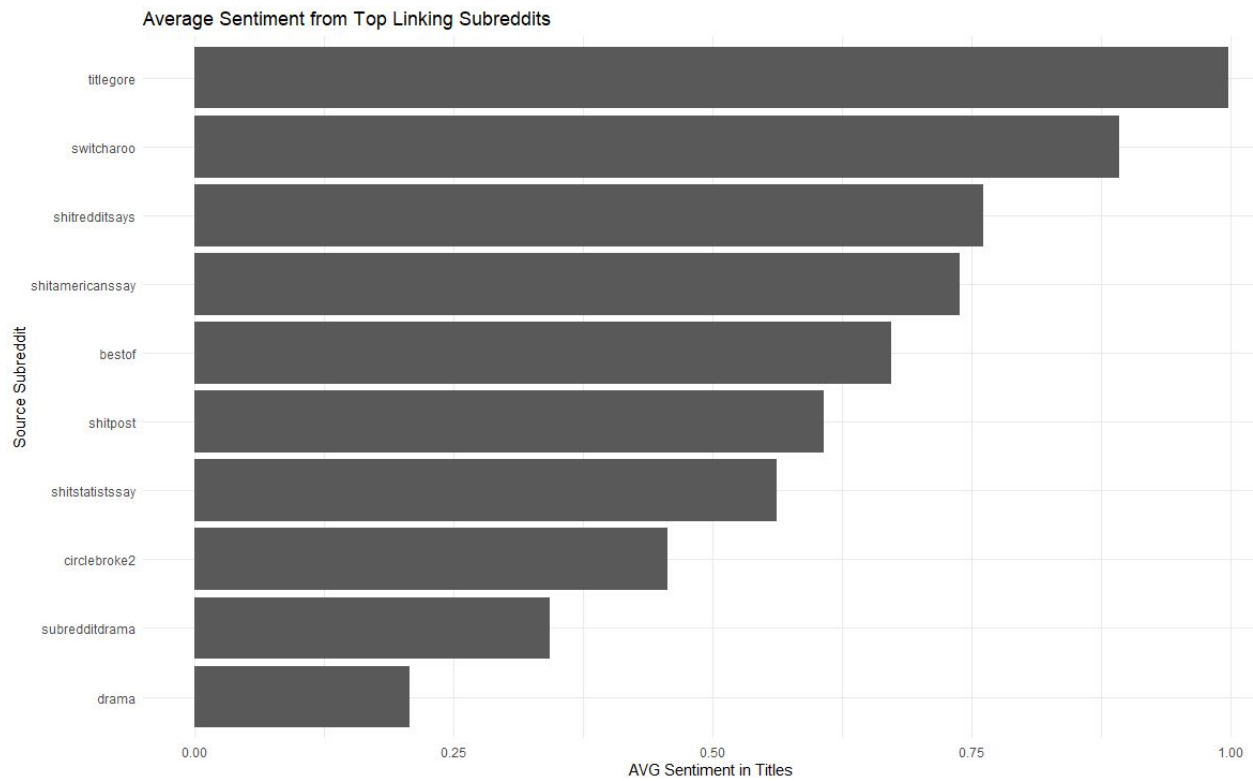
Chart of average sentiment made by Leo Kell.

CS 6630 Gaming Community Toxicity

Data Processing - PJW - 10/28

SUMMARY

Adding aggregate columns to the data set.

Goals

- Add aggregate columns to the data set.
    - GOAL CHANGED, SEE NOTES

Achievements

- Aggregation seems to be perhaps a bad idea at this stage, since
- Created multi-index using SOURCE_SUBREDDIT and POST_ID columns

NOTES

- Aggregation seems to be perhaps a bad idea at this stage, since some visualizations currently proposed want to chart underlying data distribution, so changed plan to instead simply add multi-index

- Had to add drop_duplicates call on POST_ID and SOURCE_SUBREDDIT multi-index, currently record loss is minimal (69 of 2370), but this should be revisited later

ACTION ITEMS

- Review drop_duplicates risk (see NOTES)

Proposal Meeting - All- 10/28

SUMMARY

Review the current version of proposal and assign action items to have it complete by the due date of 10/30.

Goals

- Add remaining sketches to proposal

- Review proposal

    - Compare to instructions

- Assign action items for completing proposal

NOTES

- Decided on traditional file structure with boilerplate code and the like seen in the hw

ACTION ITEMS

- Paul will initiate the dashboard code directory
- Leo will get a code formatter
- Leo will find where to put in assumptions about toxicity
- Leo will make another Node View Sketch
- Paul will submit the proposal and do final grammar checks
- Michael will finish his sections in the proposal by Friday morning

Project Setup - PJW- 10/30

SUMMARY

Created basic program structure.

Goals

- Simply wanted to create basic file structure

Achievements

- Created structure and uploaded new files to Repo

NOTES

- N/A

ACTION ITEMS

- Begin creating Summary View as outlined in Proposal

Init. Summary View (SV) - PJW-10/31

SUMMARY

Begin creating the Summary View.

Goals

- Create necessary files and JS Objects

- Get some ink on the canvas

NOTES

- Split charts/tooltips into at least three different views with appropriate subdirectories

- Failed to get ink on canvas

ACTION ITEMS

- Get ink on canvas

Post Count SV - PJW - 11/01

SUMMARY

Get ink on canvas.

Goals

- Create a basic line chart for time series number of posts.

Achievements

- Removed PROPERTIES column from exported data
- Typed columns on data load
- Drew axes for first chart

NOTES

- Current typing approach for data is brute forced, other methods should be considered

ACTION ITEMS

- Do roll up or group by to get post count (possibly cumulative) for selected data
- Draw line for selected data

FIGURES

Figure: Drew scales for Post Line Chart in Summary View

Expanding Data Set - PJW - 11/02

SUMMARY

Expanding data set with new subreddits and title data set.

Goals

- Get more data into our data set.

Achievements

- Doubled number of subreddits
  - However, many of these do not seem to be represented in the data set
- Appended title data to body data and fixed subreddit names to be lowercase, increasing dataset size by ~900% to ~18k items
- Switched dev branch to pjw, so now main can be primary branch for project and we can do pull requests

NOTES

- Took extra subreddits from https://www.reddit.com/r/gaming/comments/abp756/top_20_game_titles_on_reddit_sorted_by_subscribers/ and https://www.elecspo.com/games/top-30-best-gaming-subreddits-with-most-reddit-subscribers/

ACTION ITEMS

- Go through data processing code in meeting tomorrow
- Should a column be included to specify in JSON data whether data came from title or body?

Review - All - 11/03

SUMMARY

First review of project status and code.

Goals

- Review data processing code
- Review proposal for peer review

Achievements

- Reviewed code and proposal

NOTES

- Use pull requests to update main branch (give special attention to CSS)
- Keep latest post, assuming duplicates are edited posts

ACTION ITEMS

- Block out HTML file in group meeting
- Michael will set general stylings
- Leo and Michael need to make branches
- Leo and Michael will initialize prototype views
- Discuss interactions between different views

Create branch/Styling/Class stubs- MLN- 11/04

SUMMARY

I branched the project for myself, added overall styling to the CSS based on what is on Reddit and created the directory and files for the view I'm going to work on.

Goals

- Branch the project

- Add to css

- Create files and directory for view

Achievements

- Branched the project

- Added to Reddit fonts to css

- Created files and directory for view

NOTES

- Searched for reddit font-family since it seemed like a fun idea to use.

ACTION ITEMS

- Plan additional general font styling such as default font-size, weight, etc.

Began Development on Node View- LJK- 11/04

SUMMARY

I created a personal branch and started work on a NodePlot JS class.

Goals

- Branch the project

- Create NodeView div in html

- Create NodePlot JS class

- Create methods to dynamically process dataset to calculate links and other attributes for view

Achievements

- Branch the project

- Create NodeView div in html

- Create NodePlot JS class

- Create methods to dynamically process dataset to calculate links and other attributes for view

ACTION ITEMS

- Continue working on view: create scales and set up prototype SVG plot, static for first pass

Peer Feedback - All - 11/05

SUMMARY

Mandatory peer with David Miller, Qianlang Chen (Qian) and Jiawen Song (Kevin).

NOTES

- David – we should make sure that we're not including too much data for JS to handle
- David is a big fan of graphs, especially for online communities, so he thinks the graph view would be really neat

- David suggests that the graph view could have different levels of semantic meaning (e.g. start with global/overall, zoom to neighborhoods and individual subreddits)

- Other group members stress that we should make sure to have a consistent color scale across all views

- Kevin frequents a lot of gaming subreddits and finds our project potentially very interesting

- David suggests we can incorporate storytelling with larger digraph view

- Qian things we have a lot of potential for good interaction, but is interested in seeing a holistic mockup

Formatted Data for Force-Directed Networks LJK- 11/05

SUMMARY

I learned how to use d3's nesting() and rollup() to create grouped and summarized data, and used this to dynamically create a "nodes" and "links" dataset. Both of these datasets will be used to create node and link positions using d3's "forceSimulation".

Aggregating Data - PJW - 11/06

SUMMARY

Need to aggregate data to count of posts per month for Summary View.

Goals

- Aggregate post data using .group and .rollup

- Display line chart for posts in Summary View

Achievements

- Aggregated time-based data, which was harder than expected for next part
- Displayed line chart data of number of posts per month

NOTES

- Refer to "Summary View" as "Current Subreddit Summary View"
- Displaying time based data was more frustrating than initially expected, but hopefully the lessons learned today will aid with future work

ACTION ITEMS

- Revisit topic of making SOURCE_SUBREDDIT data arrays instead of Objects indexed by POST_ID
- Formalize names for each of the large-scale views (with full group)
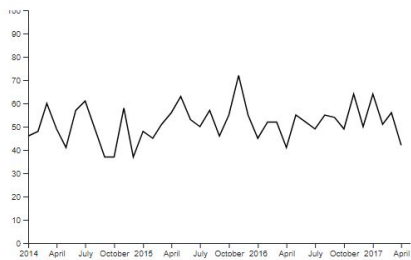
FIGURES

Figure: Current view with Post Count, using aggregated data (count)and
time-based data

Init. Violin Plot - PJW - 11/07

SUMMARY

Initialize Violin Plot.

Goals

- Create prototype Violin Plot using 'gaming' subreddit data

Achievements

- Found KDE that group can use for distribution calculations
- Created prototype violin plot sans axes

NOTES

- Found excellent violin plot code here:
  https://github.com/asielen/D3_Reusable_Charts/blob/master/distro_chart/distrochart.js
- The 'gaming' subreddit data seems like a perfect normal distribution, assuming my plotting is correct
- Plot seems like it is getting cut off

ACTION ITEMS

- Add axes
- Compare d3 plot to altair plot for verification

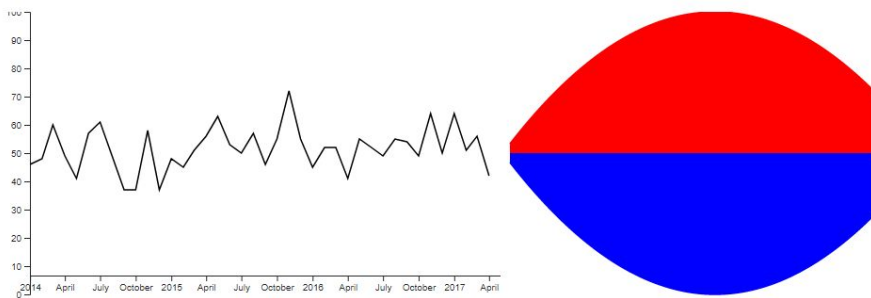## FIGURES

Gaming Community Toxicity

Figure: Added Violin Plot. Looks bad right now, especially because top and bottom halves are colored differently for debugging purposes. Data seems like a perfect normal distribution, but I want to verify against an altair plot of CompoundSentiment data. Also need axes. Note that plot is heavily derived from the link seen in the Notes section.

Page Layout Idea - MLN - 11/07

SUMMARY

Spent time looking through the "hall of fame" projects for ideas on how to layout our page. Came up with the idea of adding a Navbar at the top to toggle between different views ie the Node/Stats View, Ranked View an About page with additional info and potentially a story telling mechanism.

Goals

- Resolve how to switch between the different views and integrate potential optional features.

Achievements

- Came up with an idea for how to organize the views.

NOTES

- Inspiration taken from the following projects:

    - https://psshyu.github.io/dataviscourse-GeochemOilandGas/

    - https://uvril.github.io/VisProject/#

ACTION ITEMS

- Make navbar mockups to show team

Sentiment Bar and Improved Violin Plot - PJW - 11/08

SUMMARY

Create prototype sentiment bar charts and investigate strange normality of violin plot data.

Goals

- Fix Violin Plot or verify its content

- Create prototype sentiment bar chart

Achievements

- Violin Plot distribution fixed
- Created box plot prototype

NOTES

- Violin plot was showing strangely normal data due to overly large bandwidths and insufficient resolution. Was originally using a bandwidth of 20, but this needs to be scaled down to <1 to be informative (went with 0.1 for now). Also scaled up resolution to 1000.
- Found a github repo talking about VADER: https://github.com/cjhutto/vaderSentiment
- Saw a guideline for compound sentiment interpretation:
  - Positive: >0.05 -> Color blue or pink?
  - Negative: <-0.05 -> Color red?
  - Neutral otherwise -> Color grey?
- Found a presentation on LIWC and other sentiment lexicon stuff from Stanford: https://web.stanford.edu/~jurafsky/slp3/slides/21_SentLex.pdf
- Filtered sentiments of interest for sentiment breakout to: Posemo, Negemo, Anx, Anger, and Sad

○ Wish there were more positive breakout emotions to work with,
    though
- Changing sentiment breakout to a boxplot, as that should more accurately
  display the data of interest

ACTION ITEMS

- Color Violin Plot appropriately
- Add axis to Violin Plot
- Add quantile/metric info to violin plot
- Fix box plot
    ○ Fix outliers
    ○ Fix rendering
        ■ Center boxes appropriately
        ■ Add lines between boxes and whiskers.
        ■ Add line at median

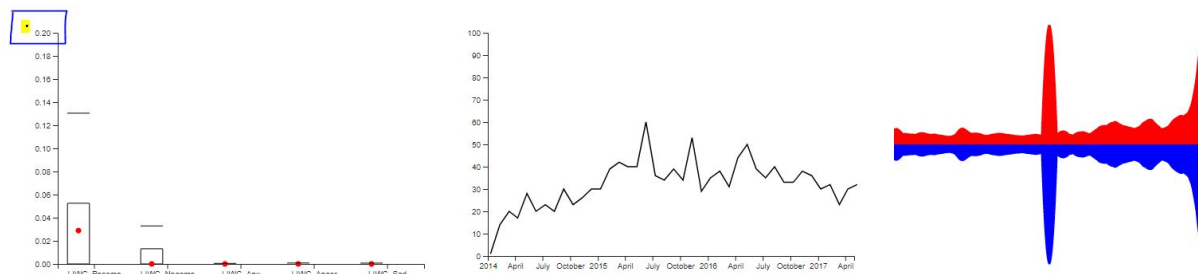## FIGURES

Gaming Community Toxicity

Figure: Box plot prototype created. Blue rectangle and highlighted region
indicates where outliers are being mapped. Obviously, this is a bug.

Graph View Prototype - LJK- 11/10

SUMMARY

Spent time looking through the "hall of fame" projects for ideas on how to layout our page. Came up with the idea of adding a Navbar at the top to toggle between different views ie the Node/Stats View, Ranked View an About page with additional info and potentially a story telling mechanism.

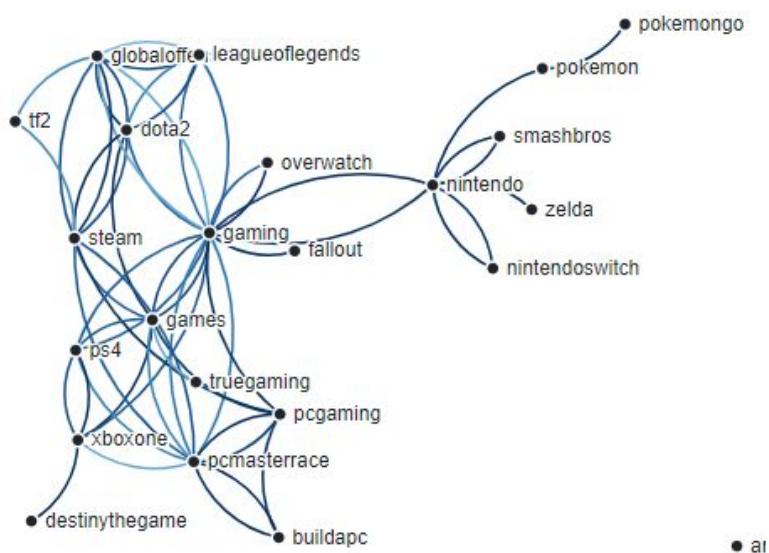I read through the D3 "forceSimulation" documentation and created a prototype "graph" plot.

Goals

- Learn how to and create a force-directed graph

Achievements

- Created a prototype force-directed graph

Figures

NOTES

- Much of my code was taken and tweaked from this tutorial:
https://observablehq.com/@d3/mobile-patent-suits?collection=@d3/d3-force

ACTION ITEMS

Continue working on prototype, use a filter option to turn into zoomed-in
"Node" view

CS 6630, Gaming Community Toxicity/Lovingness

Code Review - All - 11/10

SUMMARY

Design meeting to discuss action items.

Goals

- Check in

- Merge code

Achievements

- Merged codebases

NOTES
- If Leo can get node and graph views done, will move on to creating distribution charts at graph view
- Will use downvote blue for negative (#9494FF) and upvote orange for positive (#FF8b60)
- Leo will reach out to Devin about node and graph views

ACTION ITEMS
- Leo will create a Pull Request for his branch by 11/13
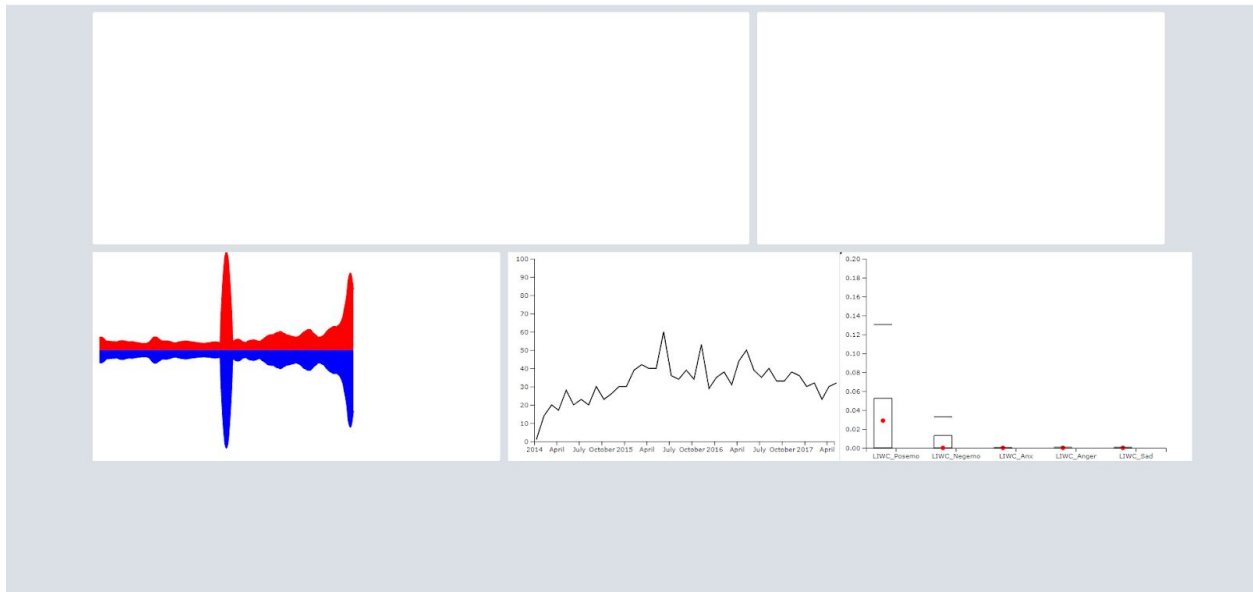
FIGURES

Figure: After merging pjw with mln

Ranked Layout/Tab Stuff- MLN-11/13

SUMMARY

Set up the layout of the ranked view, started a table for charts and data, added title and favicon to the tab.

Goals

- Layout the ranked view

- Pretty up tab

Achievements

- Created ranked view layout
- Prettied up tab

ACTION ITEMS

- Add charts and data to tables

CS 6630, Gaming Community Toxicity/Lovingness

Action Item Cleanup - PJW - 11/13

SUMMARY

Prototype Summary View is created but there are several action items that need to be enacted (see goals).

Goals

- Color Violin Plot appropriately – DONE
- Add axis to Violin Plot – DONE
- Add quantile/metric info to violin plot – DONE
- Fix box plot
    - Fix outliers
    - Fix rendering
        - Center boxes appropriately
        - Add lines between boxes and whiskers.
- Add line at median

NOTES

- For Violin Plot axis, decided to add an axisTop and axisBottom to make ticks render appropriately
- Alter SVG and Card widths to fix rendering
- Sentiment Box Plot seems largely uninformative; perhaps should take other metrics, like composition metadata (readability index, number of long words, special characters, etc.) to provide some supplementary information about general "refinement" of the subreddit

ACTION ITEMS

- Add tooltip to Violin Plot for metrics
- Add tooltip to Post plot for time and number of posts
- Change Sentiment Box Plot?
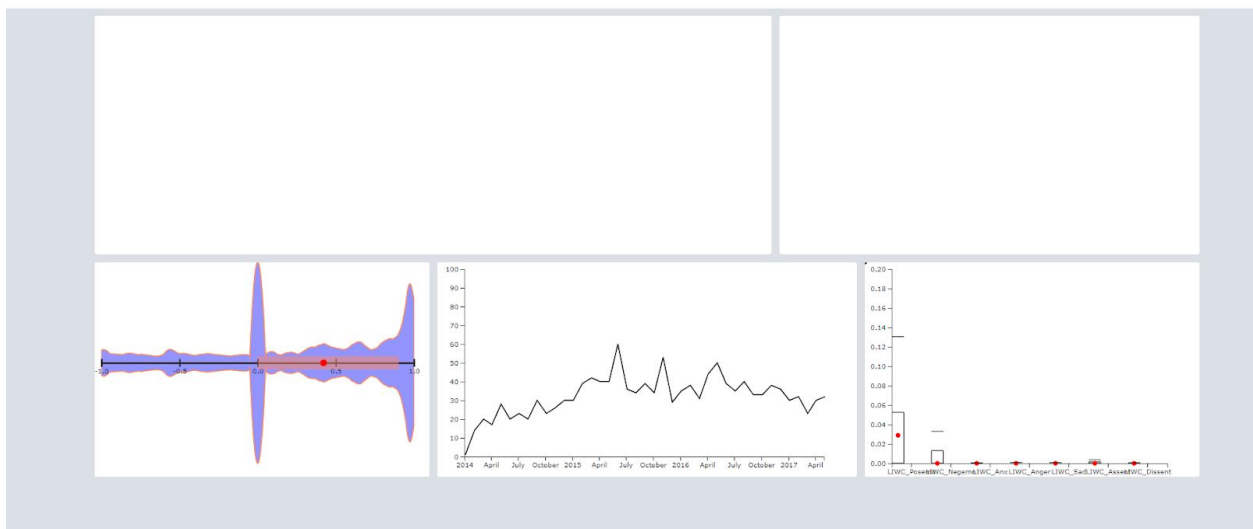   - Discuss with Leo and Michael

FIGURES

Figure: Updated Summary View. Changed violin plot coloring and added median as red circle and orange quartile box, as well as an axis depicting sentiment. Also added more values to the sentiment box plot, though this may change.

Summ. Vw. Tooltips - PJW - 11/14

SUMMARY

Adding tooltips to the Violin Plot and Post Plot.

Goals

- Add metrics tooltip to Violin Plot

- Add Number of Posts tooltip to Posts Plot

Achievements

- Achieved Goals for the day
- Added circle points to posts chart and colored appropriately

NOTES

- There are several TODO items in the Summary View Code; these should probably be addressed tomorrow
- Also need to figure out how to add transitions
  - Wait until Leo's Node View stuff is done?
- Don't forget: need to add brush to Post Plot that updates Violin Plot and Sentiment (now Meta?) box plot

ACTION ITEMS

- Work with team to get tooltip stylings
- Complete TODO items in code
- Add transitions
- Need to clarify what "Posts" are in Posts Plot because it isn't what one would expect (only shows posts that link to other subreddits; this only gives us an idea of the level of activity relative to how they interact with other subreddits)
- Posts Tooltips seem to have unreliable rendering; investigate why this might be

FIGURES

Figure: Latest version of Summary View with new coloring on posts chart. Note that the time scale does not capture the earliest points; this needs to be addressed.

Figure: Metrics Tooltip rendering

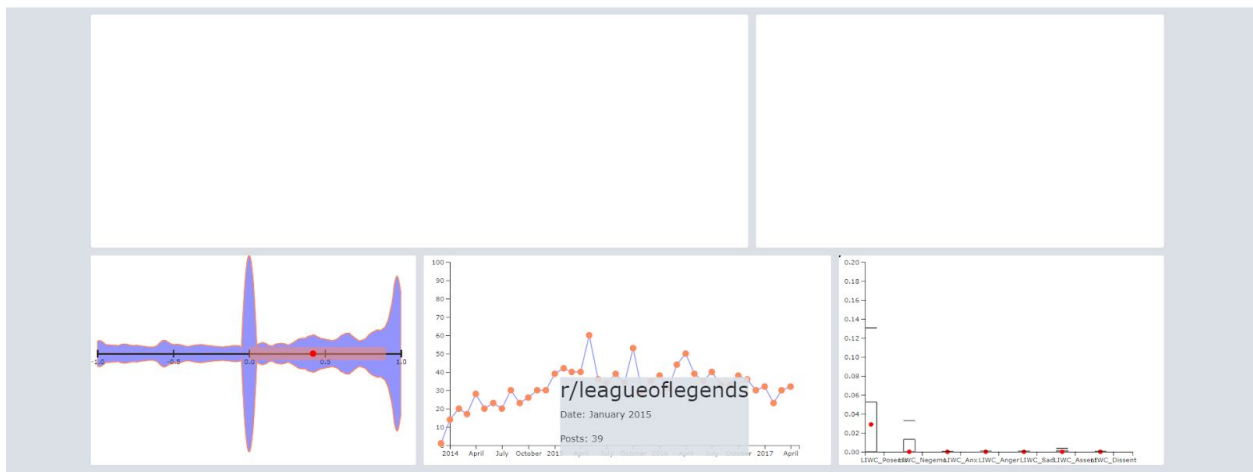Figure: Median Tooltip rendering

Figure: Posts Tooltip rendering

Milestone, Review - All - 11/15

SUMMARY

Need to commit a "feedback_exercise" file to our repository (click here for instructions). Also review current status of project and assign action items.

Goals

- Create feedback_exercise file and commit to repo

- Review code and design of project

- Review goals for the week

ACTION ITEMS

- Paul will do a Pull Request for the feedback_exercise.md today (hard deadline)
- Paul needs to fix transitions for Summary View now that Leo has Subreddit Selection partially implemented
- Leo need to implement scrollable divs for the Subreddit Dropdown and the Graph View
- Need to display current subreddit (Leo will take care of this)
- Leo will change color scheme in graph and node view
- Leo will change Node View to only look at how SOURCE feels about TARGETS, and will no longer limit TARGETs by whether or not they are gaming subreddits
- General: add chart titles
- General: add hover legends where appropriate
- General: add axis titles
- Paul will change sentiment box plot to look at meta textual analysis (e.g. % long sentences)
- Leo will encode node sentiment by color
- Leo will encode node activity by size
- Michael will add assets for styling

- Michael will add comparison filters to overall view
- Michael will add # of posts column to overall view
  - Encode low vs high number of posts with different colors
- General: change "posts" to "links" (more accurate)
- General: make CSS grid layout play nice with SVGs and Divs

FIGURES

Figure: whiteboard action items

Action Item Cleanup - PJW - 11/16

SUMMARY

There are plenty of hanging action items for the summary view that need to be addressed. Not all goals may be addressed today.

Goals

- Work with team to get tooltip stylings
  - Leo has done this
- Complete TODO items in code
  - Still have more to do
- Add transitions
  - Not done yet
- Need to clarify what "Posts" are in Posts Plot because it isn't what one would expect (only shows posts that link to other subreddits; this only gives us an idea of the level of activity relative to how they interact with other subreddits)
  - Will rename to "Links"
- Posts Tooltips seem to have unreliable rendering; investigate why this might be
  - Still needs to be investigated
- Violin Plot tooltip on-hover changes margin/padding?
  - Still need to do this
- Fix metric box and median
  - Completed

Mentor Feedback - All - 11/16

SUMMARY

Feedback with Devin Lange about the project.

Goals

- Get Feedback from Devin

    - Find flaws with design

    - Get impression of how far along project is

        - On track to completion?

NOTES

- Leo presented from his machine (has most up to date version)
- Devin recommended hover legends (hover over Card or button in Card to see Legend)
- Need to provide future work, caveats about data in About Tab

Action Item Cleanup - PJW - 11/18

SUMMARY

There are plenty of hanging action items for the summary view that need to be addressed. Not all goals may be addressed today.

Goals

- Complete TODO items in code
  - Still have more to do
- Add transitions
  - Complete
- Need to clarify what "Posts" are in Posts Plot because it isn't what one would expect (only shows posts that link to other subreddits; this only gives us an idea of the level of activity relative to how they interact with other subreddits)
  - Will rename to "Links"
    - Complete
- Posts Tooltips seem to have unreliable rendering; investigate why this might be
  - Still needs to be investigated
- Violin Plot tooltip on-hover changes margin/padding?
  - Still need to do this
- Change Sentiment Breakout Chart to AutomatedReadability Violin Plot
  - Done

CS 6630, Gaming Community Toxicity/Lovingness

New Data Investigation - PJW-11/20

SUMMARY

One major concern with our [data set](#) is that because it only analyzes links, we lose a great deal of data. If we are to examine the general toxicity of a given subreddit, it would be nice to find or create a data set that mines comments as well.

Goals

- Investigate new data sources to perhaps add on top of our current model

NOTES

- Found an interesting [source](#) that creates a data set including VADER analysis

- Created prototype data set, will continue playing with it

  - Gets top N hottest posts and then grabs top comments (implementation details are a bit fuzzy because code came from outside source; still need to study it to make certain I understand what it's doing with the API calls)

ACTION ITEMS

- Meet with Michael and Leo ASAP to talk about new data source
- Review new data source code with Leo and Michael and get them to sign off on it

New Data and Presentation - PJW - 11/21

SUMMARY

We have about a week and a half to conclude this project (due 2020-12-02). There are still some major action items that need to be completed in that time. We need to begin preparing our final submission. This will include a PDF of our Process Book, a GitHub release, a GitHub page (website), our data, and a complete README that will point to our screencast of the project among other things. We will also likely need to prepare a venv for our Python code and include a spec-file that the teaching staff can use to replicate our results.

Goals

- Figure out GitHub Pages

- Rough draft of README

- Find and prioritize action items for Summary View

Achievements

- Worked with Leo to get GitHub Page set up
- Rough draft and structure for README put together
- Created a set of action items, primarily focused on Summary View, but with extra info.

NOTES

- We will drop the brush feature on the time series links plot

    ○ Will be too hard to integrate with our new data source, and will not add too much to the data; will also likely add considerable computational overhead for limited gain

- Should we add a toggle to use hot, controversial, or top comment sentiment data?
- [Website for GitHub Pages](#)

ACTION ITEMS

- Limit links data to 2014 and after (get rid of weird scaling issues)
- Add titles/descriptions to charts
- Add appropriate axis titles
- Consider adding y scale to violin plots
- Talk with Leo about using new data to measure general toxicity of websites for graph view, but keep links data set for relationship measurements
- Fill out About section

Network and Node Plots in Stable State - LJK- 11/21

SUMMARY

The network and node plots are in a stable state - things need to be tweaked and added from here on out. They are synced up so that the Network plot allows individual nodes to be selected that will dynamically update the other plots showing up in the page.

In addition, dropdowns have been added (not functional yet) allow subreddits to be selected and facilitate visual storytelling.

Github pages was also enabled for the repository.

Goals

● Set up Github pages with Paul's help

● Get stable state of Network and Node plots

ACTION ITEMS

● Add scales, and descriptions to Network and Node plots
● Add functionality to the Node plot dropdown, and sort outgoing nodes properly (clockwise)
● If time permits, add a slider so the user can determine how many "mentions" is required for a link to be shown in the Network plot; this will allow different levels of topological structure to be shown
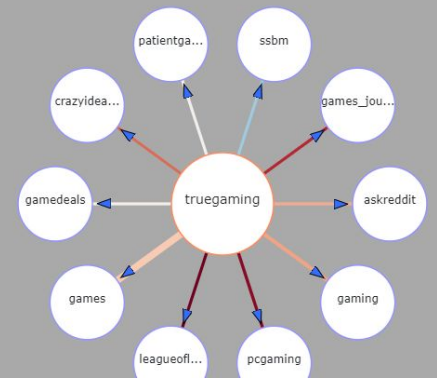● If time permits, begin adding storytelling components

FIGURES

The Gaming Landscape
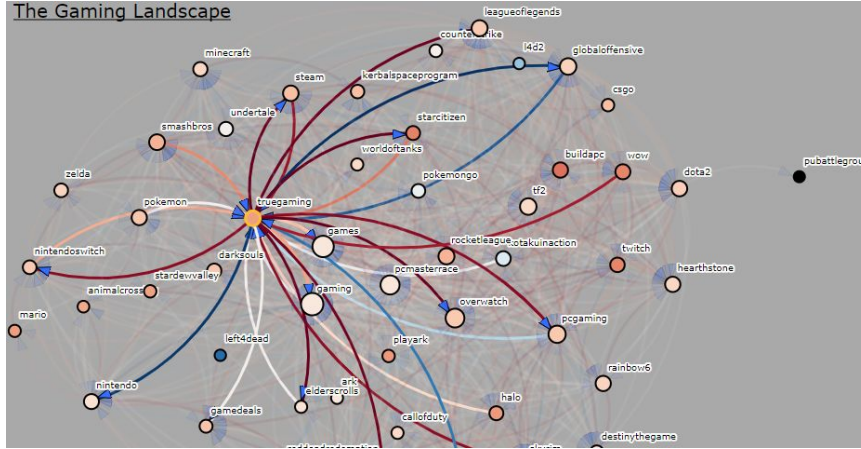
truegaming's Top Outgoing Subreddits by

Number of mentions ▾

CS 6630 Gaming Community Toxicity

Action Item Cleanup - PJW - 11/22

SUMMARY

We're getting down to the wire. Fortunately, we don't have to get a formal presentation put together (will likely just do a screencast). However, there are several important hanging items that need to get taken care of. The Goals section details important items which I have identified as having to be completed, at least for the Summary View, before we can call this project complete.

Goals

- Limit links data to 2014 and after (get rid of weird scaling issues)
    - hyperlinks data set updated
- Add titles/descriptions to charts
    - Done
- Add appropriate axis titles
    - Done
- Consider adding y scale to violin plots
    - What exactly are the units of KDEs?
    - Units are relative; will not add a y axis
- Talk with Leo about using new data to measure general toxicity of websites for graph view, but keep links data set for relationship measurements
    - Have talked with Leo; will discuss further in review
    - Resolved; will keep Leo's color scheme, coloring by links
- Fill out About section
- Create python venv, including instructions in README
    - Created requirements.txt

ACTION ITEMS

- Fix time series link plot for globaloffensive subreddit (see Figures)
- Resolve which data sets to use and how in each scenario

- Add information about what each plot represents
  - Make as obvious as possible, especially when the data is different from other charts
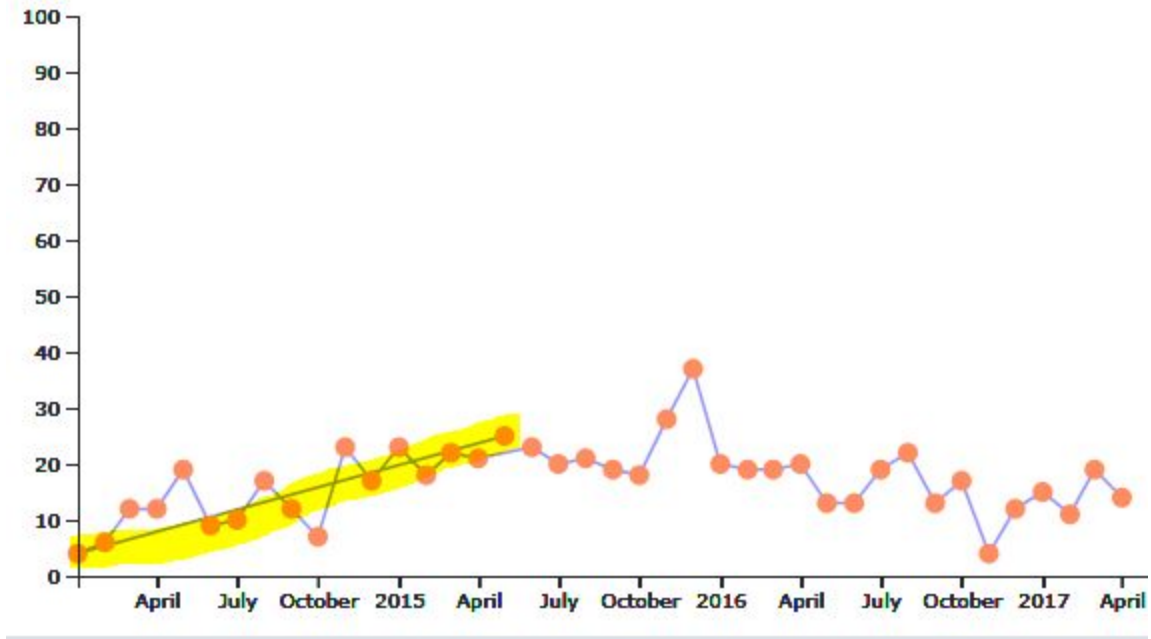
FIGURES

Figure: Bug in time-series links plot when globaloffensive is selected.

Review - All - 11/22

SUMMARY

Identify critical action items for the week, and determine what is necessary to complete the project and prepare submission.

Goals

- Talk about color scheme, especially for graph view

- Identify hanging features and prioritize completion

ACTION ITEMS

- Michael will try using comment data set in Ranked View
- Michael will update color scheme to dark mode
- Add new "axis" class to make formatting consistent (Michael and Paul)
- Leo will remove scrollable div (doesn't add much)
- Michael will find a way to integrate drop downs into the Ranked View
- Leo will create working legends for his graph and node views
- Leo will apply links filter to node plot
- Leo will made dropdown work with node view
- Leo will add summary info to central node of node plot
- Set up meeting with Devin to go over current project status
- Michael will investigate using checkbox filter in Ranked View
- All will create 30 second script for screencast on own components

Addressed Action Items - LJK-11/23

SUMMARY

I addressed most of my action items assigned at the 11/22 review.

Goals

- Complete action items assigned at the review meeting.

Achievements

- Added sentiment color scale to legend

- Added minimum hyperlinks filter to the Node plot

- Fixed the subreddit dropdown to work with Node view

- In addition to the action items listed, I also fixed a bug that was preventing the outgoing subreddits in the Node plot to be sorted by angle

FIGURES - Current Working State of Node View

Visual Updates - MLN- 11/24/20

SUMMARY

Worked on various site theme updates. Worked on changing the color scheme from a light color scheme to a dark color scheme. Added a loading gif. Also worked on all ranked view charts, added heatmap.

Goals

- Get most of the theme ready.

- Make the loading time less bothersome

- Add density plots

- Add heatmap

- Update color scheme for new theme

Achievements

- Got most of the theme ready.

- Made the loading time less bothersome

- Added density plots

- Added heatmap

- Updated color scheme for new theme

NOTES

- The theme changes caused some issue with the chart fonts and colors.

ACTION ITEMS

- Fix individual charts color schemes to fit new theme

Fixing Styling Issues - PJW - 11/24

SUMMARY

The new theme didn't merge well with some text and tooltips. Fixed stylings so that they will render appropriately.

Meeting 2 w/ Devin - All - 11/25

SUMMARY

Project is at ~95% completion. We want to get Devin's feedback to make sure we are on track to get a good grade for the project and that we don't have any massive blindspots.

Goals

- Get Devin's feedback

NOTES

- Devin says to make sure we have at least one storytelling example
    - Probably use ranked view code?
- When changing min. # of hyperlinks slider, keep edges highlighted
- Change scale of auto readability index violin plot (has negative values)
- Devin says we can leave color scheme as is
    - We were concerned that using orange-red for positive and periwinkle-blue for negative might be confusing
- Devin recommends using just years on x axis for time series chart in ranked view
    - Possibly change in summary view as well?
- Keep the selected subreddit at top of the ranked view for consistency? Add styling? Use hover to display subreddit?
- Label negative and positive sentiment axes in comment violin plot
- Use the compound color scheme in comment sentiment violin plot?
- Add width legend to node plot
- Add x labels to readability index violin plot?
- Could add info icons for more detailed explanations
- Add circle legend to graph view
- For time series links plot, check max links and redraw? Kinda sneaky…
    - Label as broke scale (kinda cheeky; Devin's favorite)
    - https://vdl.sci.utah.edu/publications/2019_infovis_clipped_graphs/

New Features - PJW - 11/25

SUMMARY

Add hanging features onto charts.

Goals

- Fix scales for readability chart
- Add axis labels to contextualize scale values

NOTES

- Used this link to determine auto readability value meanings:
  https://en.wikipedia.org/wiki/Automated_readability_index
- Set auto readability scale to read from -5 to 25
  - Does not capture full range of data, but makes it easier to tell what is happening as distribution moves
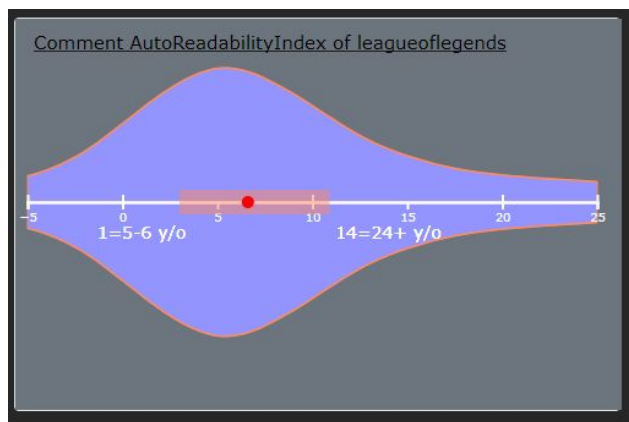   - Also, 5-6 is Kindergarten level while 24+ is Professor level reading

FIGURES

Figure: New scaling of AutoReadabilityIndex Violin Plot, with captions indicating what the scale means.
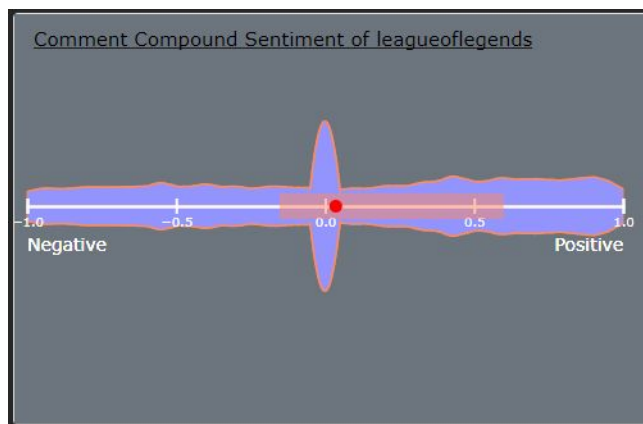
Figure: New Comment Compound Sentiment Violin Plot with axis labels indicating what the scale means.

Quick Bug Fix - PJW - 11/26

SUMMARY

Fixed bug where time series summary view link chart was not sorting data properly. Bug was noted when subreddit was set to globaloffensive.
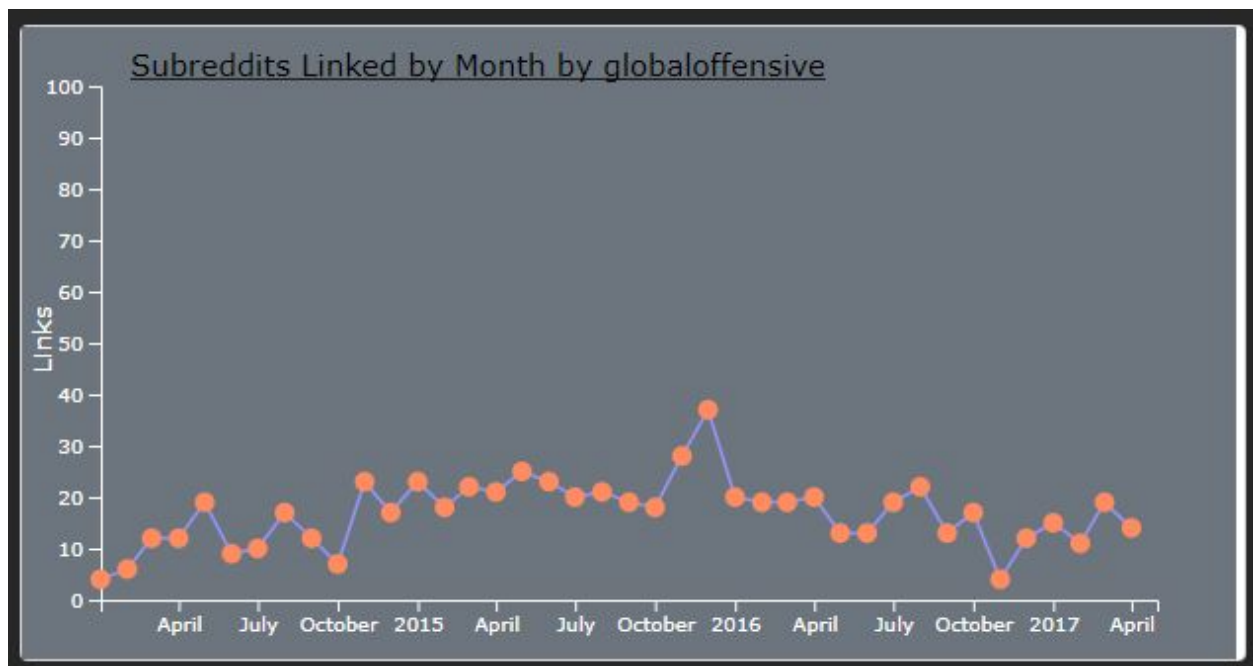
FIGURES

Figure: Fixed bug with globaloffensive subreddit.

Clean Up and Script - PJW - 11/27

SUMMARY

Do some clean up of the summary view.

Goals

- Format tooltip rounding

- Make rough draft of script for summary view

- Finish README

- Add StoryTeller class

NOTES

- NOTE: Only have 2 minutes for full presentation. Needs to be split up into four parts: Intro, Graph/Node View, Summary View, and Ranked View
  - Gives me 30 seconds
- Need Leo and Michael to fill out README sections and review
- Can't get the story teller div to stop displaying on click; fix it!
  - Also has to change currently selected subreddit

SUMMARY VIEW SCRIPT DRAFT 1

The bottom three charts break out information about the selected subreddit. The leftmost violin plot gives the distribution of comment sentiment on the hottest posts at a given point in time to measure general community toxicity, including a quartile box and median with tooltips. In the middle, is a line plot breaking out the number of links by month to give a relative indication of how active each subreddit is over time, also helping to contextualize the measured link sentiment in the graph view above. The rightmost violin plot gives a distribution of the reading level of the top comments on the hottest posts at a snapshot in time as calculated with the Automated Readability Index, where a score of 1 corresponds to a

Kindergartener's reading level (so 5-6 years old), and 14 corresponds to a Professor's (so 24 years old and above). This readability violin plot also gives the interquartile box and median.
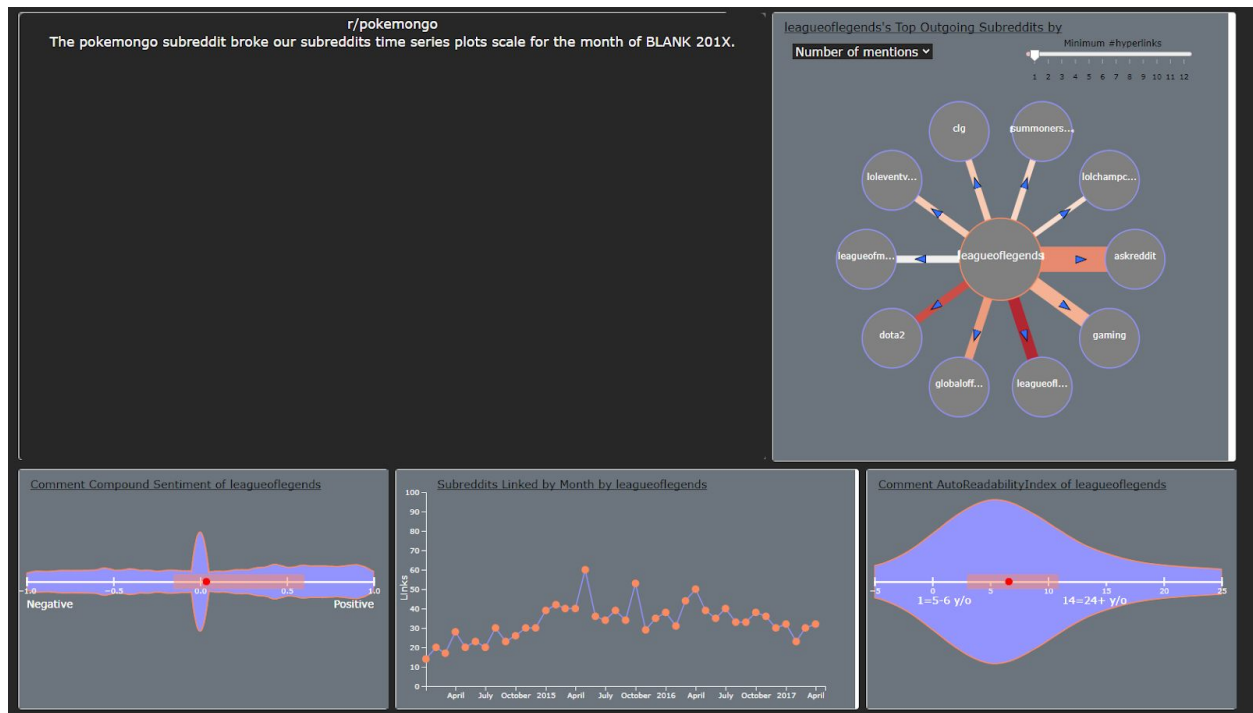
FIGURES

Figure: Story-telling div rendering.

Clean Up Storytelling - PJW - 11/28

SUMMARY

Fixing final parts of the story-telling div.

Goals

- Get story-telling div to disappear on click anywhere

- Have storyteller change currently selected subreddit appropriately

NOTES

- Done; now need to review with Michael and Leo

Storytelling - All - 11/28

SUMMARY

Meeting with all team members to discuss final implementation of story-telling view.

Goals

- Set timeline to wrap up story-telling view

- Resolve pull request conflicts

Final Review - All - 11/29

SUMMARY

Final review of project.

Goals

- Finalize script

- Take one final look at project

- Finalize README

- Review Process Book requirements

- Delete unnecessary files from repo

- Draft new "proposal" that includes Evaluation of project

ACTION ITEMS

- Add Auto Readability Index legend
- Remove r/ prefix in tooltips

Script - All - 11/29

SUMMARY

Script for screencast, to be delivered by MLN.

SCRIPT

This is the website investigating what gaming community is the most loving on Reddit. Driving our visualizations are two datasets: one from Stanford giving the Links between subreddits from January 2014 to April 2017, the other being the top level comments for the top 100 hottest posts at one point in time for each subreddit analyzed. Both of these datasets use VADER sentiment analysis and Automated Readability Index.

The first chart plots provides an overview of the sentiment topology for the selectlist of gaming subreddits. This is a directed, weighted graph, where each node represents a particular subreddit and each edge represents one or more hyperlinks between two subreddits.The hyperlink slider allows a minimum threshold to be set. Clicking on a node dynamically updates other charts on the page. The zoomed in node chart on the top right details the relationship of the selected subreddit with its top-linking related subreddits. The top ten list can be defined with the dropdown according to number of links, overall positive or negative sentiment

The bottom three charts display information for the selected subreddit. The leftmost violin plot gives the distribution of comment sentiment, the middle is a line plot indicating the number of links to other subreddits per month, and the rightmost violin plot is a distribution of the reading level of comments as calculated with the Automated Readability Index, where a score of 1 corresponds to a Kindergartener's reading level, and 14 corresponds to a Professor's.

RANKED VIEW PARAGRAPH

The ranked view page is designed to display an overview of all of the subreddits. This includes a set of line charts which display average sentiment overtime for all subreddits combined and a sortable table which includes each subreddits links, average positive, negative and compound sentiment, and a density plot with all three sensitments overlaid.

STORYTELLING PARAGRAPH

NEW SCRIPT GOALS

Answer three questions:

1. What community is active?
2. How does the community interact with other communities?
3. How loving/toxic is that community based on comment sentiment?

Kotakuinaction (see in ranked view is most negative)? Buildapc (see in ranked view is most positive)? Truegaming (justifies showing underlying distribution)?

NEW SCRIPT

The "Gaming Community Toxicity" dashboard examines gaming subreddit toxicity both in how subreddits interact with other communities, and within themselves using VADER sentiment analysis.

The first two charts on the home page explore how different gaming communities interact with each other.

The bottom three charts detail how the currently selected community behaves, both through NLP analysis of comments and subreddit activity as measured by links to other subreddits.

Users can use the "ranked" page to sort and compare different subreddits according to sentiment scores and link activity.

Suppose we want to find out which is the most negative gaming community. Using the sorting capabilities of the ranked view, we see that kotakuinaction is the most negative.

Selecting this subreddit and returning to the main page, we notice a few interesting things. Apparently, kotakuinaction seems to hate the drama and nottheonion subreddits. It also seems quite toxic according to the Comment Sentiment Violin plot which shows that it's first quartile sits at about -.42.

Personally, I enjoy PC gaming, so suppose I want to make my own PC, but I am unsure what that community is like. By selecting the buildapc subreddit from the dropdown, I immediately notice that it is a very active and very positive community. The node chart clearly shows that this subreddit interacts very often with buildapcforme in a very positive way, so that subreddit may be worth checking out as well. The Compound Sentiment violin plot also shows that the comment sections tend to be very positive based on the quartile metrics.

As a final example, let's examine an atypical subreddit. If we select truegaming from the network plot, we see it has almost no neutral disposition compared to others, but two fairly large tails, implying that this is a very polarized subreddit.

These are just a few examples of what insights this tool can provide about gaming communities. We hope that you will check it out, and make some discoveries of your own!

Final Touches - LJK- 11/29

SUMMARY

Over the past week I finished the final touches in the network and node plots. Some of these were minor action items that resulted from the meeting with Devin. In addition, I made some contributions to the storytelling aspect.

Goals

- Final touches to node view and contribute to storytelling dropdown.

Achievements

- Completed legend on network plot, including the addition of a "node size" legend

- Add an edge width legend to the node plot

- Fix tooltip on central node in the node plot

- Add tweaks to the storytelling dropdown (i.e. display the appropriate storytelling labels) and implement storytelling behavior for "The Subreddit with the most links"

FIGURES - Current Working State of Node View (note the legends on the bottom right corners of both plots)
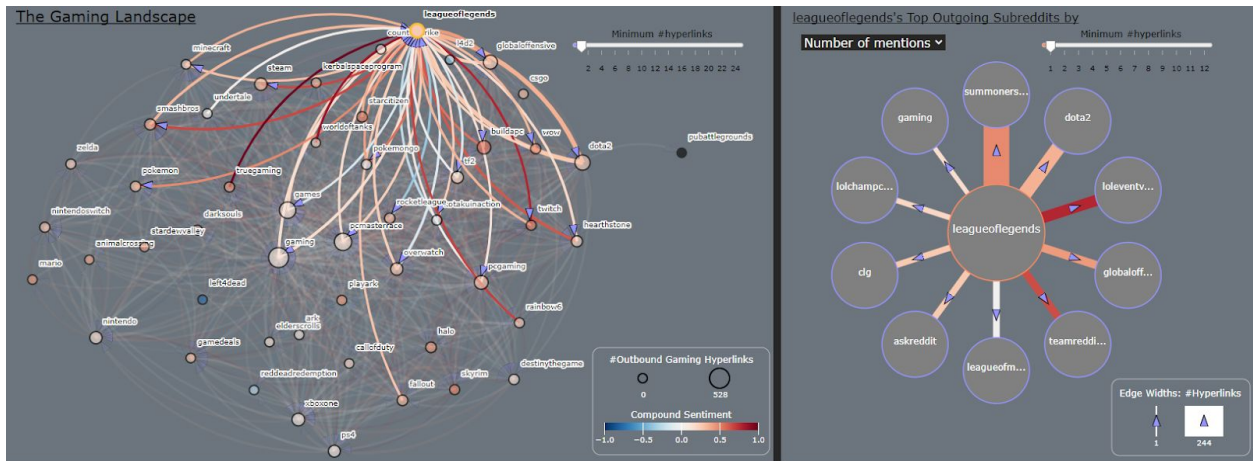
The subreddit with the most links storytelling
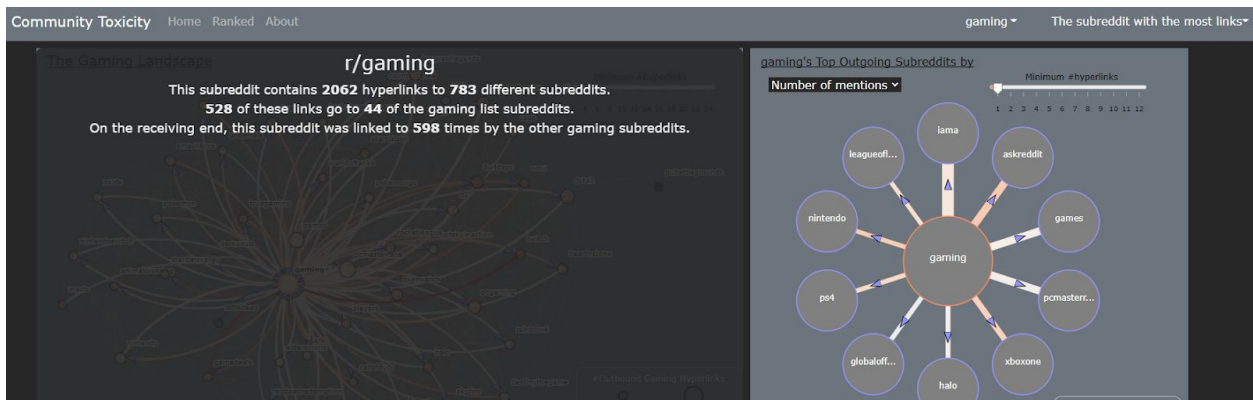
Action Items

- Finish draft of screencast script