The VAST Challenge 2008 – The Paraiso Manifesto

- Ajay Shaan Shanmugam, Siddharth Chandrasekaran

Link to Video Demo:

https://www.useloom.com/share/e6fa9a50d758452987ca0b33ec180804

Mini Challenge – 1:

Problem Statement:

We need to infer the list of malicious / counter-productive wiki edits on the Paraiso Manifesto wiki page, and we find out if violent activities were a part of the paraiso.

Dataset Provided:

Text file containing Wiki edits.

Approach:

We created a social graph comprising of all the wiki editors. Edges denote interaction between the users through edits/undo/collaborate. This helps us understand the users who most frequently collaborate/disagree. We used all comments used by users of interest identified from the previous graph and plotted the words in a word embedding plot where similar words are clustered together to look for violent activities. Since there is no direct support to aesthetically plot network graphs in bokeh, we had to write a framework which uses the python louvian module which aesthetically orients the nodes and then we rendered using bokeh's glyphs.
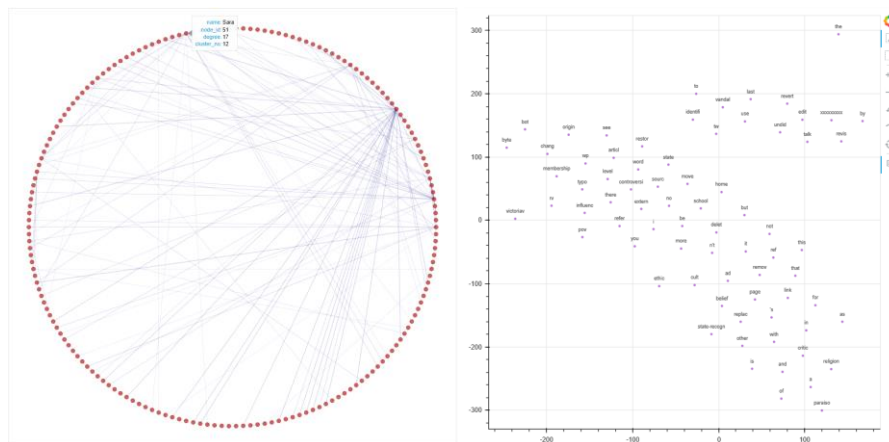
Results:

We identified that Victoria, Savannah and RyogaNica are staunch supporters of paraiso. We idenfified that Rm99, Dove and Squeakbox are against the movement. Edemir, Sierra, Adriano, Socorro, Honoria, Gustava and Sara, having unbiased views, are determined to be the neutral editors. We also identified several mentions to violence involved in the paraiso moment by analyzing edits containing the negative words clustered together in our word embedding graph.

Mini Challenge – 2:

Problem Statement:

Following the crackdown on Paraiso fanatics by the Isla Del Sueno government, a mass illegal immigration of its residents to US was noticed. Records of interdictions and successful landings were collated and provided to us and we're tasked with using visual analytics to find patterns and answer some very interesting questions related to this.

Dataset Provided:

Data contains records for fields such as interdiction encounter/landing and launch coordinates, number of passengers, types of vessels used by the immigrants and coast guards, number of deaths during voyage etc.

Approach:

1. How would you characterize the choice of landing sites over the three years?
   We used a GMap plot to furnish information on choice of landing sites over the three years. The colors of glyphs represent the number of deaths in a certain landing site and upon hover, one can learn the exact coordinates, number of passengers in that successful voyage and the names of passengers if available in the dataset.
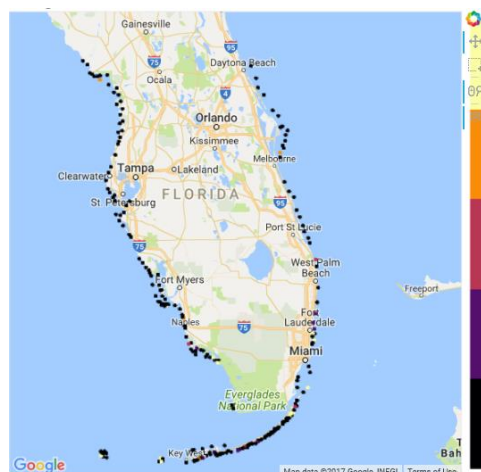
2. How would you geographically characterize the interdiction patterns over the three years?

   We used a GMap plot to furnish information on interdiction patterns over the three years. The colors of glyphs represent the number of deaths in a certain site of interdiction and upon hover, one can learn the exact coordinates, number of passengers caught in the interdiction and the names of passengers if available in the dataset. We initially plotted glyphs such that theirs sizes were based on the number of passengers. But it obstructed clear viewing of the plotted points and we had to incorporate that measure in the hover tooltip and used number of deaths as a measure for the color mapper.
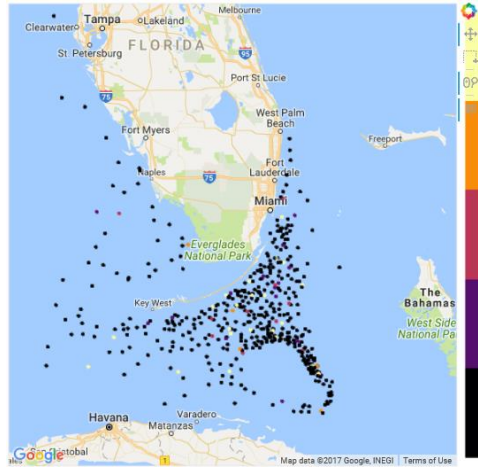
3. What are the vessels of choice used by the migrants over the three years?

   A stacked categorical bar chart was created to represent the choice of vessels used by the migrations over the period of time.
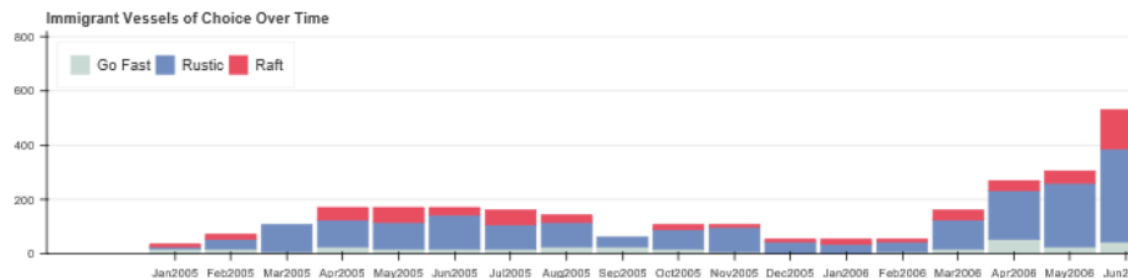

4. What is the successful landing rate over the time period?

   A categorical bar chart was created to observe the rate of successful landings over the three years. Upon interacting with a drop-down list box, the user can toggle between this graph and the one displaying the death rate over time.

5. What is the death rate on voyages?

   A categorical bar chart was created to observe the rate of death in these perilous voyages over the three years. Upon interacting with a drop-down list box, the user can toggle between this graph and the one displaying the successful landing rate over time.

Landing Rate Over Time

Rate_over_Time Successful Landing Rate
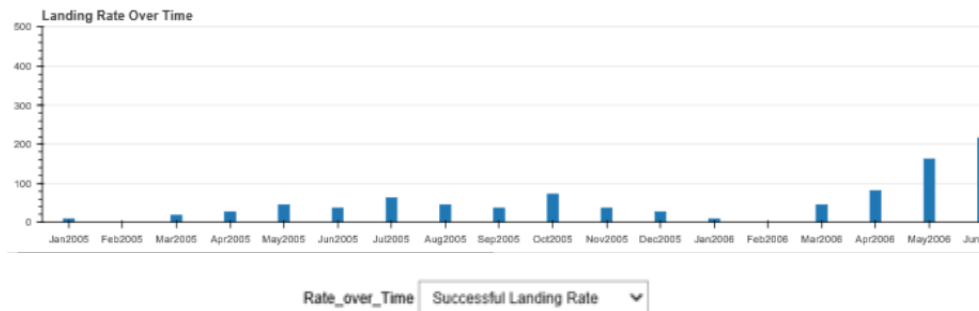
6. Are there any of the characters of interest in this data?
   Our GMap plot used for question 1 provides a list of passengers who made it to US soil, giving coast guards the slip. They are our characters of interest.

Results:

- Migration pattern kept changing over the years. Intially the immigrants preferred landing at the part of the US soil closest to the island. Over the second year, as the coastal guards caught on and tightened security, the choice of landing site started moving towards north east part of Florida. Over the third it moved further north and focussed on the north western part. This can be explained by analyzing the interdictions and landing patterns. Immigrants kept avoiding the paths known for a history of interdictions.
- It looks like the rustic boat was the most preferred choice of vessel for them. Gofast boats had more success helping the immigrants land in US soil though, and rafts had the most death rates.
- They seemed to have made the most successful landings in the months between May and September. This indicates that either the immigrants aren't used to cold weather or that the waters aren't as navigable in the colder months of the year or both.

Mini Challenge - 3:

Problem Statement:

Using phone call records over a ten-day time period, we're to use visual analytics to tag potential big-shots in the Catalano network who are suspected to be at the heart of terrorizing Paraiso activities.
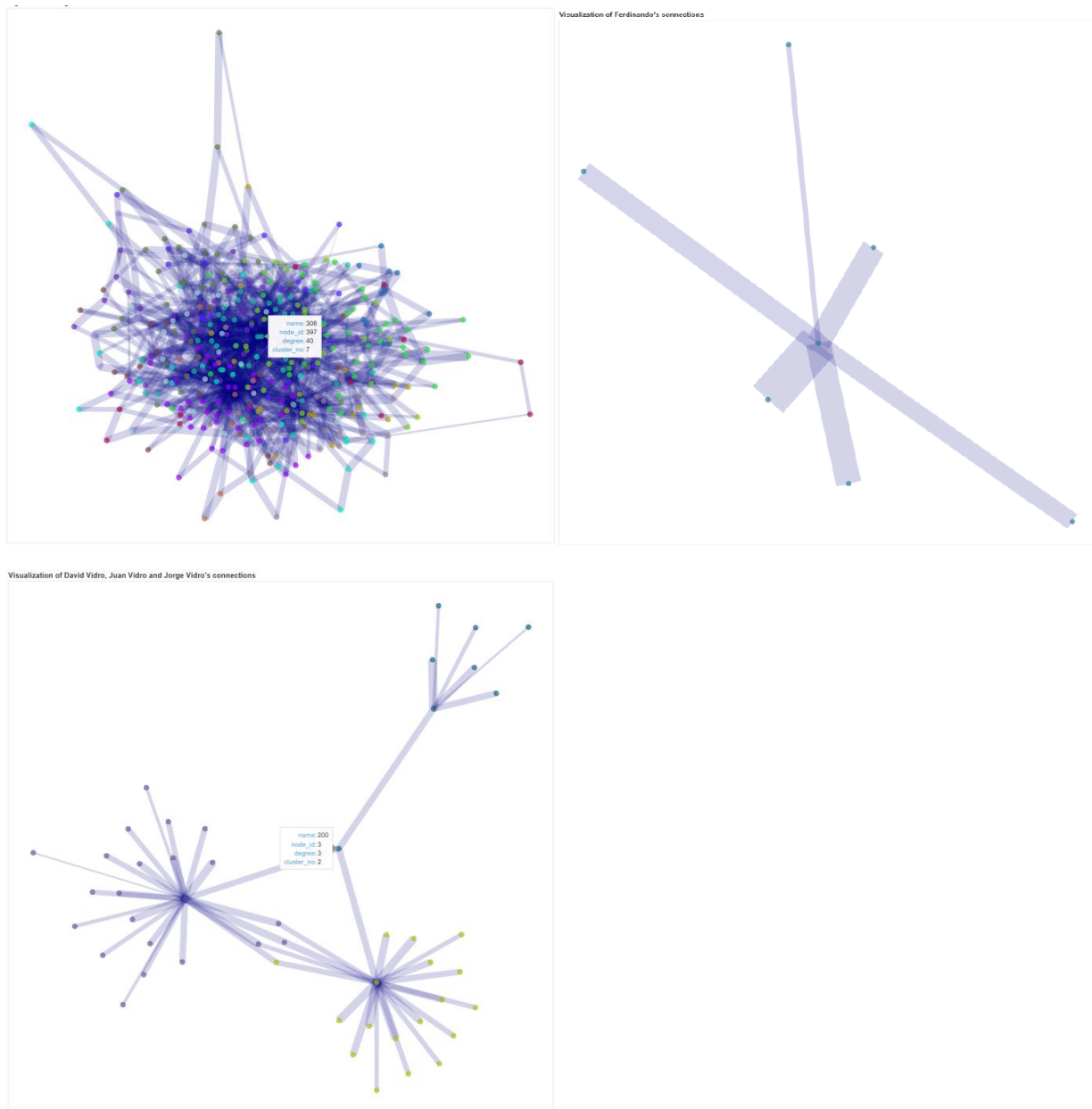
Dataset provided:

Cellphone Call records containing records for fields such as caller, receiver, timestamp, call duration and cell tower.

Approach:

Since we were provided with information on identity of a prime suspect, we plotted the entire CDR social network with lengths of edges representing a measure of frequency and length of communication. We

then clustered this data using KMeans to identify the closed networks and color coded the nodes using cluster id. We were able to identify the suspects using hints provided in the question and were able to sketch the social network of the paraiso fanatics. We had to determine a good measure of magnitude of communication between individuals in the graph, and we came up with the idea of using a measure that multiplies frequency of calls between two callers and the duration of calls to calculate the thickness of the edges in the graph.

Visualization of Ferdinando's connections

name: 306
node_id: 397
degree: 40
cluster_no: 7

Visualization of David Vidro, Juan Vidro and Jorge Vidro's connections

name: 200
node_id: 3
degree: 3
cluster_no: 2

Results:

Given that identifier 200 is Ferdinando and based on the provided hint that Ferdinando talks to his brother Estaban most frequently, we identify node 5 (The node connected to node 200 by the thickest edge) to be Estaban.

Another clue that's been provided to us is that close relatives and associates that Ferdinando would be calling include David Vidro, Juan Vidro and Jorge Vidro, in addition to his brother Estaban. Therefore, out of the other 5 nodes, the nodes connected to Ferdinando by the next 3 thickest edges would be David Vidro, Juan Vidro and Jorge Vidro.

The final hint that we've been given is that David organizes most of the Paraiso activities. This gives us a strong reason to believe that he must have the most number of connections among the three. Therefore, we plotted a new graph outlining all connections of David, Juan and Jorge. From this, we inferred that node 2 has the highest degree (the most number of connections) and therefore we identify him as the mastermind behind the Paraiso activities, David!

As for distinguishing between Juan and Jorge, we don't have enough information in the problem statement or in the dataset to make a judgement on their identities.

Another interesting inference from the new graph is that nodes 0, 1, 34, 27 are mutual connections of David and Juan/Jorge. Since they're confirmed to be David's associates, we can infer that Juan/Jorge, David, Estaban, and the aforementioned nodes are definitely a part of the Paraiso activities.

# Mini Challenge - 4:

Problem Statement:

An explosion occurred at the DOH building in Miami, rumored to be caused by Paraiso supporters. Our objective is to use all provided information and apply visual analytics to determine suspects, where the bomb was planted (after considering possibility of a suicide bomb attack) etc.

Dataset Provided:

Text file containing data on Building sketch, DOH occupants' RFID assignments and RFID records prior to and after the explosion.

Approach:

We plotted the coordinates of the building and the rfid locations of all occupants of the building. Each person is identified by a red dot and text label with their rfid. Upon hover their name gets displayed. We also have a slider, play and pause buttons which lets us visualize the entire course of happenings on the day of bombing. We were able to identify the suspects, casualities, time and exact location of the bombing. When implementing the mechanism for time traversal using a slider bar, the updates made during interaction weren't clearing out the data points previously in the plot before plotting them anew, and we had to inject javascript code through bokeh to make sure the updates happen as expected.
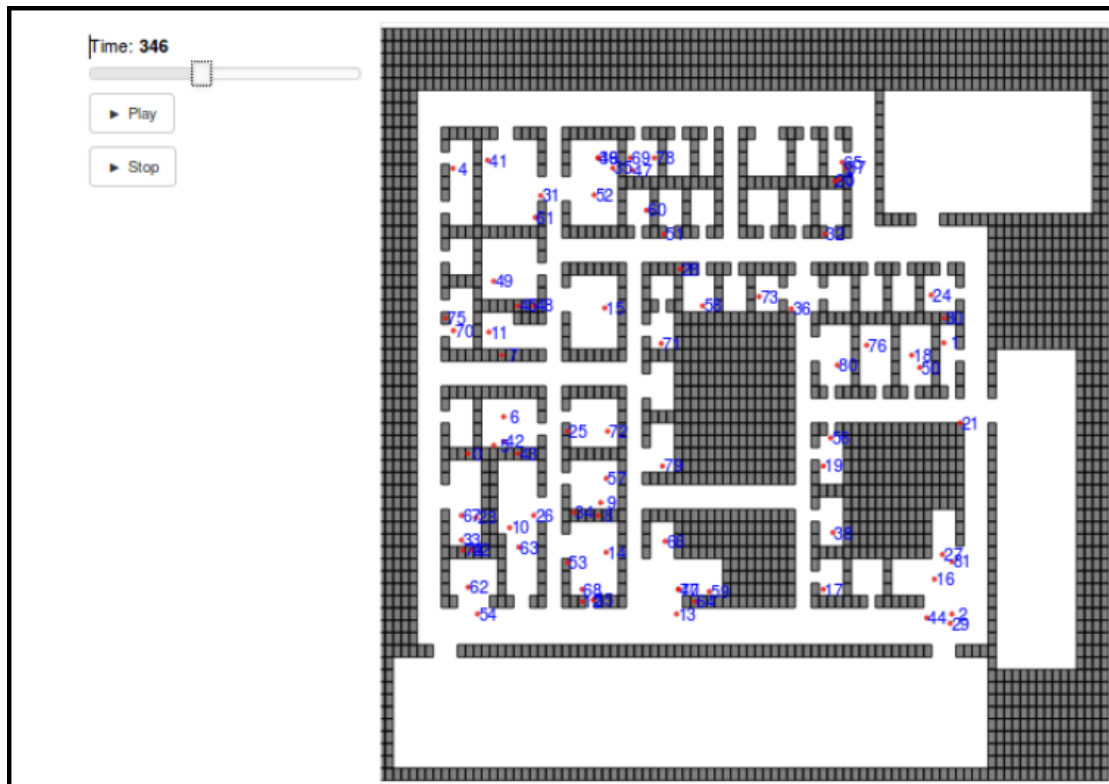
Results:

We think Ramon Katalanow whose RFID tag was 21 placed the small explosive device in the building. At the time tick of 374 the small explosive device was set off. We can see that Ramon walks away and meets someone shortly before the bombing. He meets with RFID number 1. This could be a coincidence but definitely peaks our interest. We believe that the bomb was at 61 x 33 to 67 x 29. The casualities are those whose RFID were 18 (Gale Welsh), 19 (Max Valdez), 50 (Lottie Staley), 56 (Cleveland Jimenez) and 76 (Fawn Sparks). Folks whose RFIDs were 39 (Phil Marin), 47 (Rosario Oakley), 59 (Olive Palmer), 60 (Lavon Lockhart), 65 (Dian Crum), 69 (Morton Kilgore), 78 (Cleveland Hutchison) got injured and passed out or died during their escape, as they stopped and never moved again.

Grand Challenge:

From the above inferences, we could make the following strong affirmations.

- We could conclude that the Paraiso members were certainly involved in the bombing at the health building in August 2007. The unusually high number of immigrants who were successfully able to land in the US in the months leading up to that suggests that they were aided by the Paraiso community. They weren't your usual illegal immigrants.
- Judging by their antics, the leaders of the Paraiso are likely to be planning to strike back and seek vengeance on the Isla del Sueno government after mass propagation of their beliefs in the US and rallying them into a rebellion.