

# Steam Network Visualization

Guillermo Rodriguez

University of Colorado at Boulder

## ABSTRACT

*The Steam Network visualization was intended to allow users to visualize their steam friends within the context of a network visualization. This method allows users to understand how their friends, and communities are connected. With the use of the Steam Web API, this project queries a root user and then grabs all their steam friends. Once the friends list has been gathered, the tool will then begin to query the friends list of all the root user's friends and call an additional API to grab all their information which is then stored in a graph format for additional editing. When the finished visualization was completed there was several notable clusters which represented communities both in real life and virtual.*

**Keywords:** Steam, friends, network, visualization, graph

## 1 INTRODUCTION

Steam when it was released was intended to serve as a digital distribution platform for Valve. It started from humble beginnings but soon grew to the biggest gaming distribution platform in the world. Apart from distribution games, Steam has built communities and connected gamers with the use of their friends list. This social feature allows us to explore and visualize the connections between users by using a network visualization to demonstrate connections between gamers. This visualization would allow us to find patterns within our friends list as a network visualization can be arranged in many ways (based on the algorithm) which allow us to quickly group and determine connections between friends. This information is important as it allows us to see how communities are formed within Steam.

## 2 RELATED WORK

There has been many usages of networks and neural network graphs throughout the past several years. Neural networks have been used consistently within machine learning and other applications. They have also been used within social network visualizations such as in the case of Facebook. During my research into this subject I came upon several journal articles that went into several aspects of networks.

In a paper called "Exploring network structure, dynamic, and functions using NetworkX", the paper detailed the usage of a python library called NetworkX and demonstrated how to use the library to create network graphs. The authors used NetworkX to study "synchronization of coupled oscillators" (Hagberg, Swart and Schult 2008) during their research project. This paper highlights the flexibility of NetworkX as the nodes within this library can be anything hashable "such as strings, numbers, files, functions, and more" (Hagberg, Swart and Schult 2008) which greatly extends the usage of the library in many fields.

With the previous paper, I knew that I could create a graph but I wanted to see if I could manipulate it further. I came across a paper called "Gephi: An Open Source Software for Exploring and Manipulating Networks". The paper discusses the use cases for the program and makes a case for the advantages of using the software. Gephi allows for a visual algorithmic approach to manipulating a network in real-time by leveraging "the computer graphic card, as video games" (Bastian, Heymann and Jacomy 2009) do to manipulate thousands of nodes and edges. These two papers established my foundation of tools and libraries I would leverage to conduct my own Steam visualization.

My exploration of tools did however lead to another paper called “Cytoscape.js: a graph theory library for visualisation and analysis” which explored the use of Cytoscape. This JavaScript library supported graph visualization within HTML, however I decided against the use of this library as I wanted something more hands on such as Gephi.

With the tools selected, I researched the topic of social network visualization to get a better understanding of network graphs. In “Social Network Visualizations: Can We Go beyond a Graph?” the authors used two visualizations within their research, “a traditional network graph with email contacts as nodes” (Viégas and Donath n.d.) and a second visualization which “depicts the temporal rhythms of interactions in dyadic relationships between ego and individual contacts” (Viégas and Donath n.d.). These two visualizations allowed participants to get a better understanding of their connections within email as the additional layer allowed them to get a better understanding of their data the clustering of users within the context of relationship. I used a similar approach in my second visualization which grouped steam users into three groups: online community member, real life friends, and other.

In “Network visualization and problem-solving support: A cognitive fit study”, I explored additional visualization methods that could be leveraged to visualize networks. This research paper explored four methods to visualize a network: a table, node-link, matrix, and textual representation of the data. The researchers found that the node-link representation and matrix had a higher response accuracy to other methods presented to their research participants.

I wanted to also see how researches use network graphs in a different context than social networks. In the paper “Representing, Analyzing, and Visualizing Scholarly Data in Support of Research Management”, researched explored using network visualizations to represent the research process. The used networks to tie together different researchers such as PHD students, faculty, and external collaborators and used the size to

represent the amount of publications and color to indicate the degree of each node. This work allowed me to extend my own visualization by also adding both color and size.

After researching network visualizations, I wanted to explore how Steam data is represented. I found a paper titled “Condensing Steam: Distilling the Diversity of Gamer Behavior” which conducted a large-scale data mine of the Steam API to gather information on more than 108.7 million user accounts, game data, and community groups. With this data in hand the team did an analysis of “gamer behavior across the dimensions of social connectivity, playtime, game ownership, genre affinity, and monetary expenditure” (O'Neill, et al. n.d.). This paper helped me understand the data that the Steam API could provide as they listed a breakdown of each table within their dataset, as well as an in-depth analysis of several factors within the Steam ecosystem.

The final paper I reviewed was “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software” which detailed the use of the algorithm ForceAtlas. This paper demonstrated the several use cases for the algorithm and helped me determine that it would best be used in my visualization as I wanted to group nodes to a central node or hub. ForceAtlas provided the ability to do this as it uses a force-directed layout to position nodes and edges. This design choice allowed for my ideal representation of my graph as opposed to other algorithms provided by Gephi.

### 3 DESCRIPTION

The Steam Network Friendship graph was constructed using Python and was intended to demonstrate how a steam user is connected to friends. Originally my initial approach to this project was using the data captured by Brigham Young University however the database file provided was too large to handle as the uncompressed SQL file was around 170GB which taxed all of my local development builds. I explored using cloud based solutions (AWS and

Google Big Compute) to query the big data, however these were too costly to be feasible for this research project. Due to this setback I had to refactor my original project to attack a smaller dataset and use the same APIs used to datamine for the BYU dataset.

Using the Steam API, specifically the user and friend endpoints, I could construct a python script that dynamically gathered data on the user and all their friends to put together my network graph. When completing the first graph, I learned that the built-in representation provided by NetworkX was not enough to provide a clear representation of this data. NetworkX however provides a method of exporting a graph file which can be imported into several tools. This is when I used Gephi to import the graph file and begin manipulating the nodes.

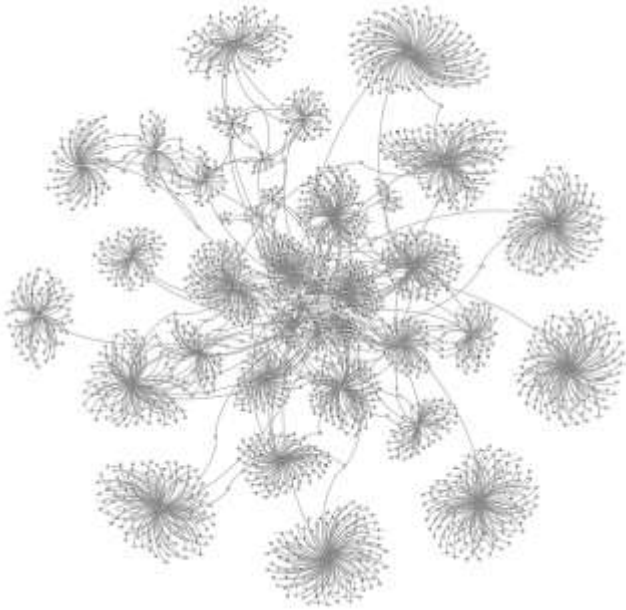


Figure 1 The Steam Network of my Steam friends

After using several of the built-in algorithms provided by Gephi to position the nodes, I settled on ForceAtlas as this provided the best representation of hubs due to the way the algorithm applies gravity and force to each node. The hubs that started to appear once the algorithm began to shape demonstrate groups of friends linked to a single person all of which are connected to the root user.

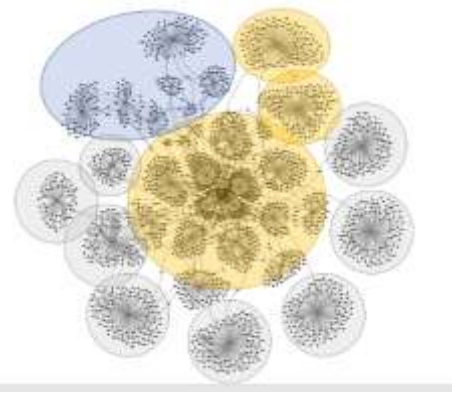


Figure 2 This graph groups nodes into online community members (yellow), real life friends (blue), and other users (grey)

## 4 DISCUSSION

The completed graphs provide several insights which I had never thought of until being able to visualize my friends through this method. This is the case with several sectors within the graph, as they line up with different types of relationships.

The central cluster consisted on friends my account made during my time managing internet communities. This central cluster (grouped as yellow) consists of 8 different friends which share a very large number of connections with each other due to this community. In the top region (grouped in blue), the connections to my root account are a little sparser however, this region lines up with real life friends and co-workers. The other hubs surround the central cluster (grouped as grey) are one off friends which are not connected to other friends within the graph.

If I had additional time I would like to expand my dataset deeper by gathering data of friends of friends, this would provide additional challenges however as even at 2 levels down, there is already over 2000 nodes. This is based off 51 friends in my account. Adding this additional dataset would exponential increase the number of nodes in my dataset.

## 5 REFERENCES

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. *Gephi : An Open Source Software for Exploring and Manipulating*

*Networks*. Association for the Advancement of Artificial Intelligence.

2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS One*.

Franz, Max, Christian Lopes, Gerardo Huck, Yue Dong, Onur Sumer, and Gary Bader. 2016. "Cytoscape.js: a graph theory library for visualisation and analysis." *Bioinformatics* 309-311.

Hagberg, Aric, Pieter Swart, and Daniel Schult. 2008. "Exploring Network Structure, Dynamics, and Function using NetworkX." *Proceedings / Talks*.

Neiryneck, Thomas, and Katy Börner. 2007. "Representing, Analyzing, and Visualizing Scholarly Data in Support of Research Management." *Conference Proceedings of 11th Annual Information Visualization International Conference* 124-129.

O'Neill, Mark, Elham Vaziripour, Justin Wu, and Daniel Zappala. n.d. *Condensing Steam: Distilling the Diversity of Gaming Behavior*. Santa Monica: IMC.

Viégas, Fernanda, and Judith Donath. n.d. *Social Network Visualization: Can We Go Beyond the Graph?* Cambridge: MIT Media Lab.

Welles, Brooke, and Weiai Xu. 2018. "Network visualization and problem-solving support: A cognitive fit study." *Social Networks* 162-167.