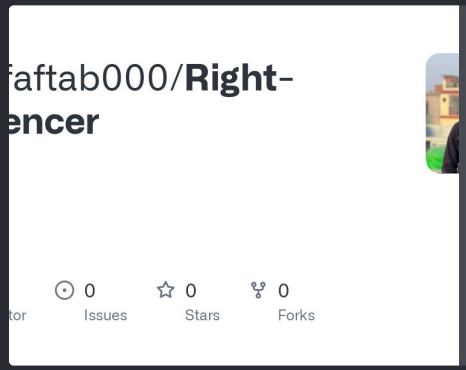
Choosing the Right Influencer: A Social Network Analysis

Based on the case study 'Who is the Right Influencer?'

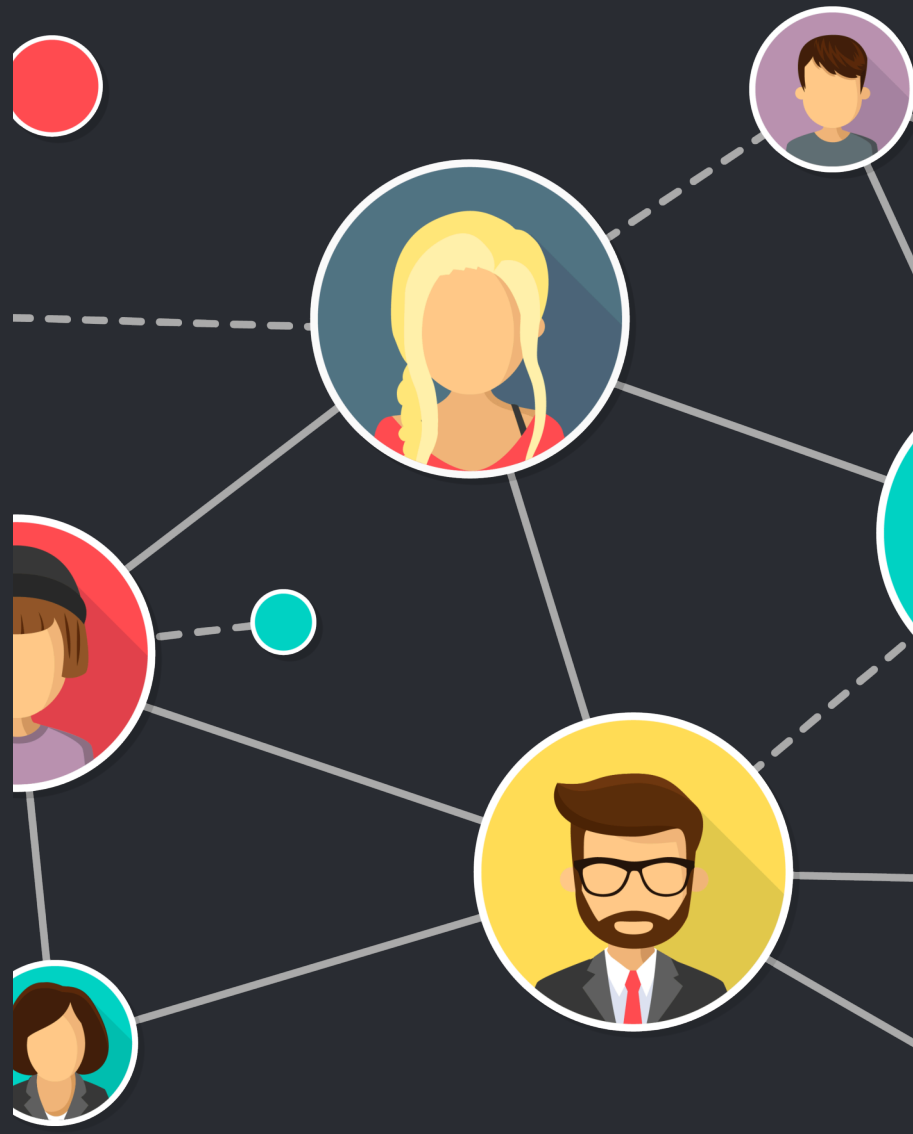
Alkaif(22504) , Anant(22505) , Labhanshu(22529) , Manam(22530)



GitHub

GitHub - alkaifaftab000/Right-Influencer

Contribute to alkaifaftab000/Right-Influencer development by creating an account on...



Problem Statement

Who's Involved?

Tina Lohn, CMO of Space Games Inc. (SGI), is leading the launch campaign for "Thrive in Space 5."

Goal

Generate buzz around the game launch through collaborations with Twitch influencers, targeting gamers aged 16-25.

Challenge

Deciding on the right influencers: follower count or Social Network Analysis (SNA)?

Importance

Justify the cost and demonstrate a strong ROI; a failed campaign could lead to negative publicity.

Objectives of Case Study are:

- Three influencers with unique channel
- SNA Metrics used to select potential influencer
- Strategies used to complement this selection
- Characteristics of dataset provided
- Expected ROI from selected Influencer

Key SNA Metrics for Influencer Selection (Part 1)

Degree Centrality

Measures the number of connections an influencer has.

In-degree & Out-degree distribution (number of followers vs. following)

High degree centrality indicates well-connected influencers who can reach a large audience.

Betweenness Centrality

Measures how often an influencer lies on the shortest path between other nodes.

High betweenness centrality indicates brokers or gatekeepers who control information flow.

Key SNA Metrics for Influencer Selection (Part 2)

Closeness Centrality

Measures how close a node is to all other nodes in the network.

High closeness centrality indicates influencers who can quickly spread information across the network.

Eigenvector Centrality

Measures the influence of a node based on the influence of its connections.

High eigenvector centrality indicates influencers connected to other influential nodes.

Clustering Coefficient

Measures how tightly knit a node's connections are.

High clustering coefficient indicates influencers who are part of close-knit communities.

Dataset Overview

Dataset Source

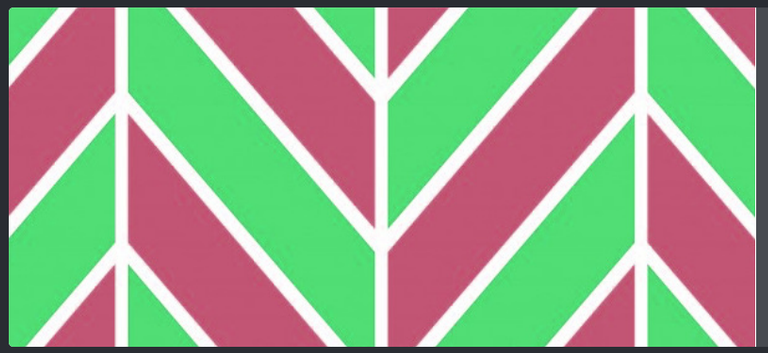
The data was obtained from Twitch, adapted from Rozemberczki, Allen, and Sarkar (2021).

Dataset Size

Contains information on 7,000+ Twitch streamers who stream in English.

Dataset File

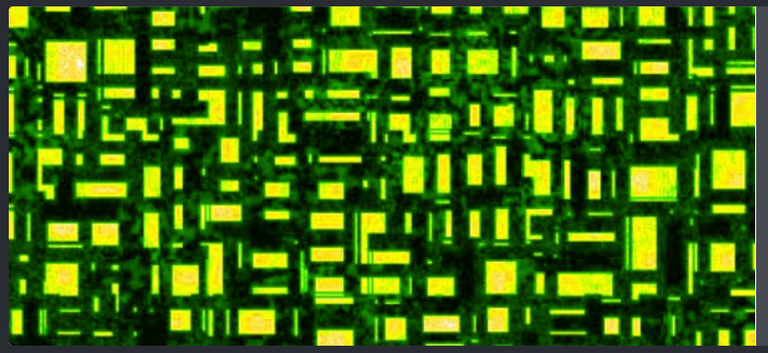
- **Twitch_nodes.csv**: Contains attributes of streamers (e.g., ID, days active, views, mature content, affiliate status).
- **Twitch_edges.csv**: Describes relationships between streamers (who follows whom).



www.kaggle.com

TwitchEdges

Kaggle is the world’s largest data science community with powerful tools and resources to help you achieve your data science goals.



www.kaggle.com

twitchNode

Kaggle is the world’s largest data science community with powerful tools and resources to help you achieve your data science goals.

Data Characteristics

Directed

The relationships are directional (e.g., Streamer A follows Streamer B, but not necessarily vice versa)

Not Weighted

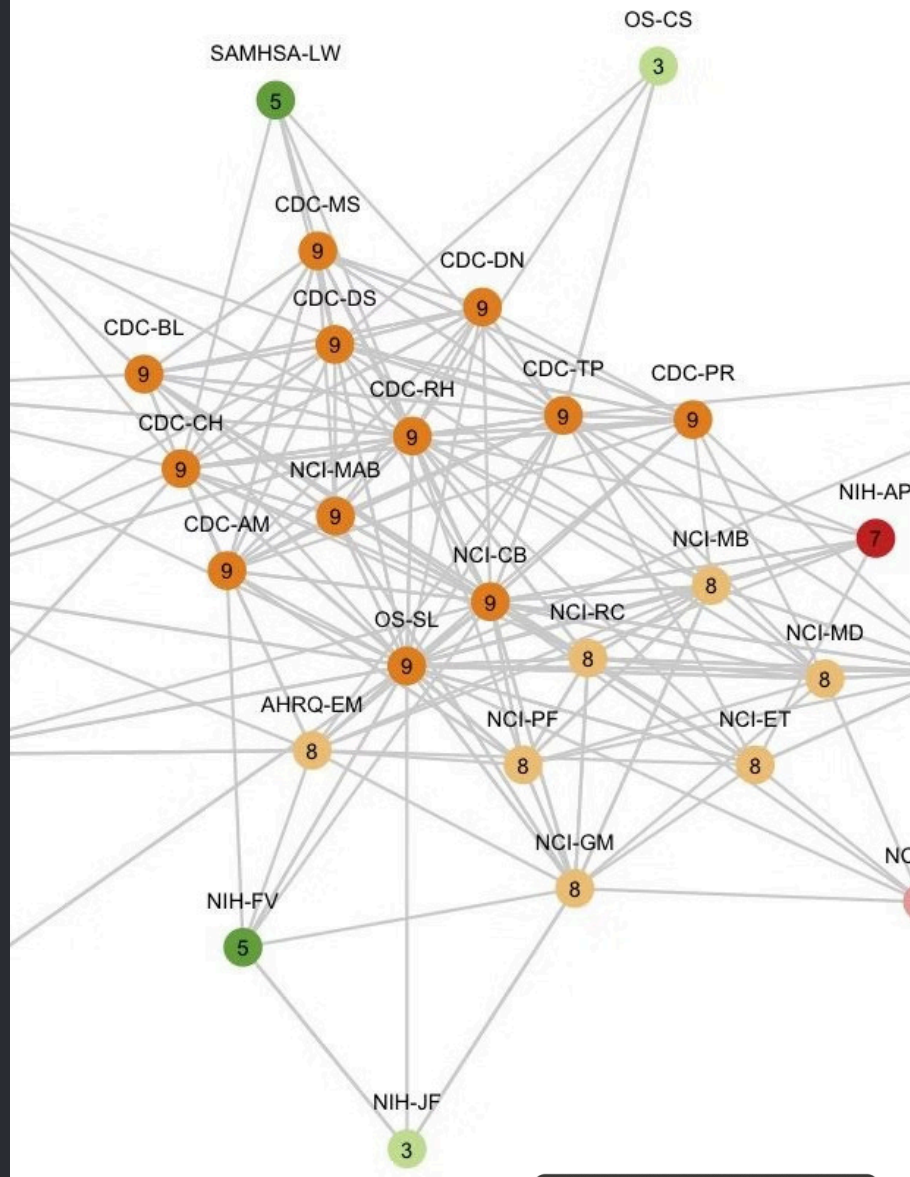
The edges do not have weights (e.g., no strength of connection is measured).

Not Dynamic

The data represents a snapshot in time (not a time-series or evolving network).

Network Statistics:

1. **Total Nodes (Streamers):** 7,126
2. **Total Edges (Connections):** 35,324
3. **Average Degree:** ~4.91 (Each streamer, on average, has ~5 connections)
4. **Network Density:** ~0.001 (Very sparse network; most streamers are not directly connected)
5. **Network Diameter:** 14 (Max. Distance to reach from one streamer to farthest node.)
6. **Major Communities:** 13 communities calculated through



Process of Analysis

Loading the Data

The data was loaded into Gephi using the Twitch_nodes.csv and Twitch_edges.csv files. The edges file was imported as an edges table and the nodes file as a nodes table.

Calculating SNA Metrics

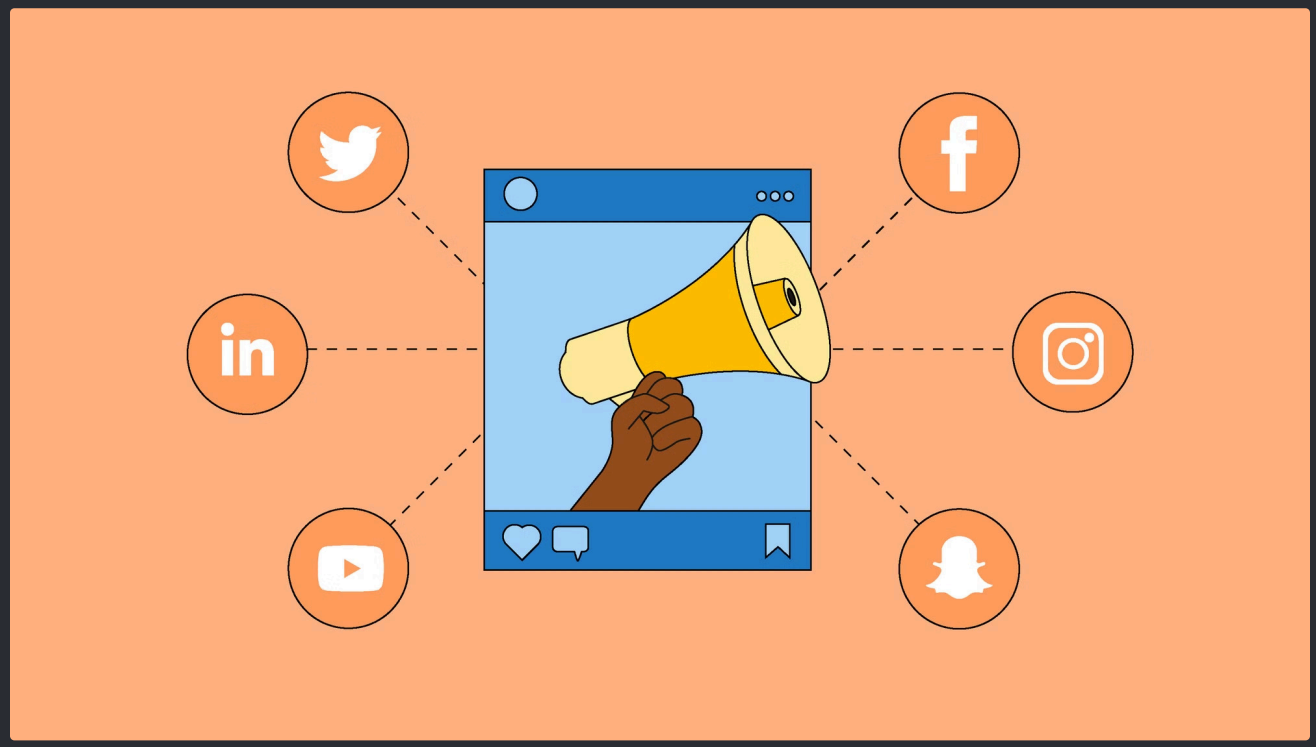
Centrality metrics (degree, betweenness, closeness, eigenvector) were calculated using Gephi's Statistics tool. Filters were applied to focus on key variables (e.g., views, days active, affiliate status).

Visualizing the Network

The ForceAtlas 2 layout in Gephi was used to visualize the network. Node size and color were adjusted based on centrality metrics (e.g., larger nodes for higher degree centrality).

Identifying Influencers

Filters were applied to narrow down the dataset (e.g., streamers with high views, active days, and centrality scores). Top influencers were selected based on a combination of metrics and alignment with SGI's target audience.



Top Influencers Selection

- Picked **top 10 influencers** for each centrality metric.

Finding Common Influencers

- Found **influencers common in all three metrics**.
- Found **influencers appearing in at least two metrics**.

Comparing Influencer and Other Factors

Views vs. Centrality

Explored the link between streamer views and their network centrality metrics.

Followers vs. Centrality

Assessed the relationship between follower count and influence within the network.

Partnered Status

Investigated how partnership status affects a streamer's centrality scores.

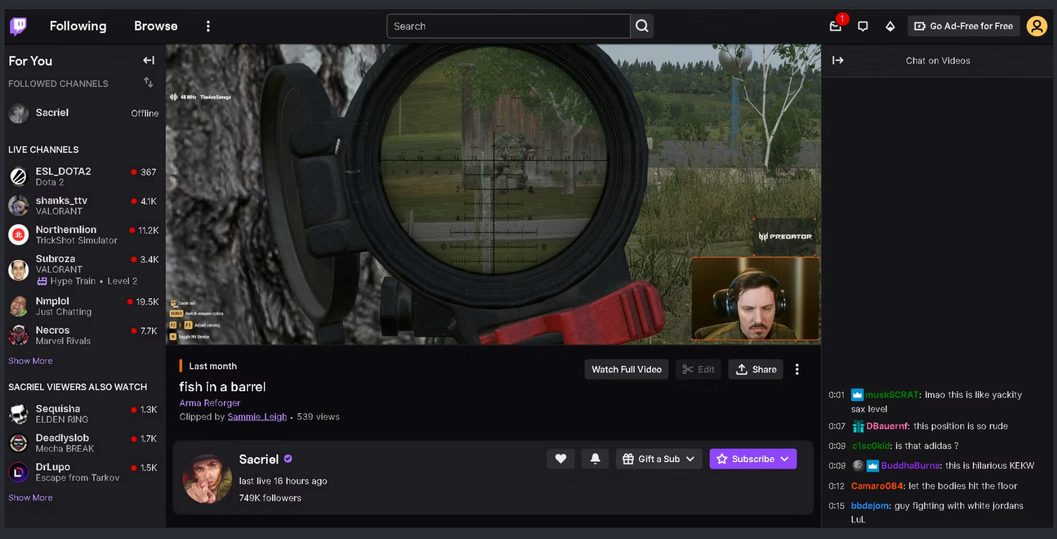
Top 3 Influencer

Sacriel

Click here

Channel Id - 23735582

Id - 4949

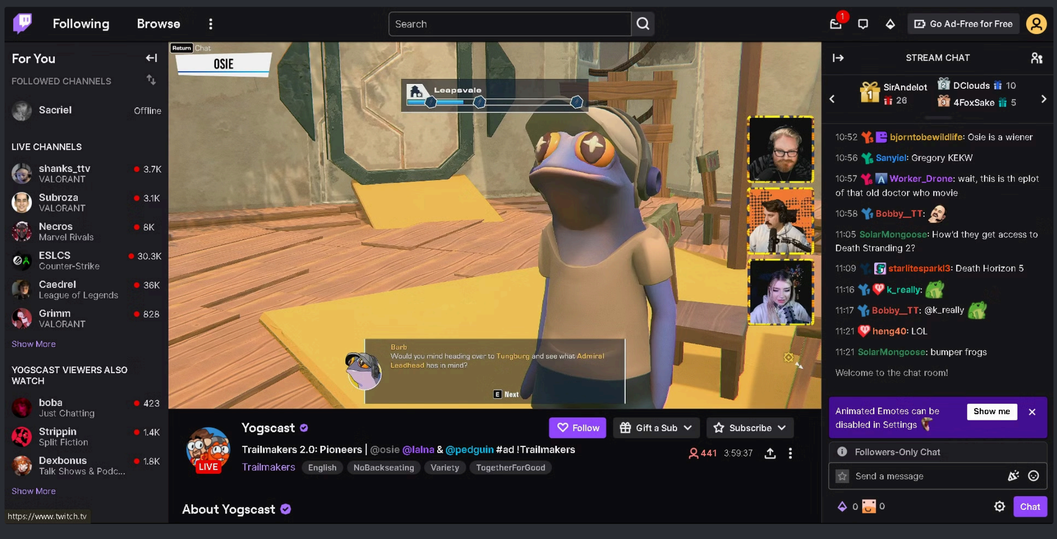


Yogscast

Click here

Channel Id - 20786541

Id - 1773

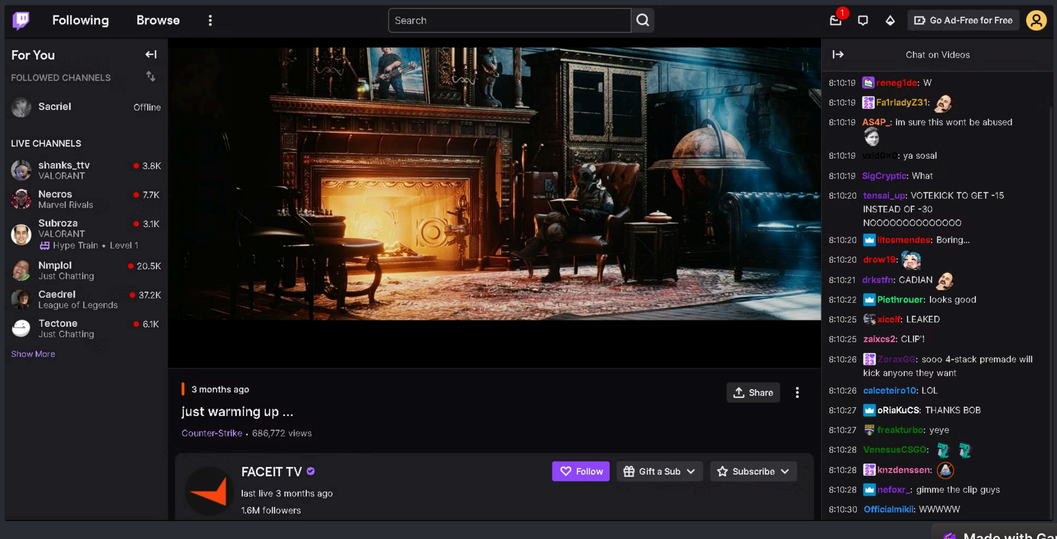


FACEIT

Click here

Channel Id - 27942990

Id - 6136



Conclusion

Influencers with high betweenness centrality act as bridges in the network, connecting different communities and ensuring wide reach for SGI's campaign.

> In the age of data, decisions are only as good as the insights behind them.