15/2/2024

Mapping the Digital Influence: A Social Network Analysis of the Bioneer Youtube Channel

Course:

Social Network Analysis

7th Semester

Supervising Lecturer:

Dimitris Pournarakis

Spanakis Panagiotis-Alexios (8200158)

Contents

Introduction 2
Data Overview 3
1. Graphical Representation of the Network 4
2. Basic Topological Properties 7
3. Component Measures 8
4. Degree Measures 10
4.1 Average Degree 10
4.2 Degree Distribution 11
4.3 In-Degree Distribution 14
4.4 Out-Degree Distribution 16
4.5 In-Degree vs Out-Degree 18
4.6 Weighted Degree Distribution 18
5. Centrality Measures 19
5.1 Degree Centrality 19
5.2 Betweenness Centrality 21
5.3 Closeness Centrality 23
5.4 Eigenvector Centrality 25
6. Clustering Effects 28
6.1 Average Clustering Coefficient 28
6.2 Clustering Coefficient Distribution 28
6.3 Triangles 30
6.4 Triadic Closure Phenomenon 32
7. Bridges and Local Bridges 34
7.1 Bridges 34
7.2 Local Bridges 36
8. Gender 38
9. Homophily 39
9.1 Homophily based on Country 39
9.2 Homophily based on Subscriber Count 41
9.3 Homophily based on Video Count and View Count 42
9.4 Homophily based on degree assortativities 43
10. Graph Density 44
11. Community Structure 45
11.1 Bodybuilding Themed channels 46
11.2 Calisthenics Themed Channels 46

11.3 Science Themed Channels 47
11.4 Digital Entertainment Based Channels 48
11.5 Tech Based Channels 48
12. PageRank 50
Conclusion 52
Appendix 53

Introduction

In the expansive universe of YouTube content, certain channels emerge as pivotal nodes, bridging disparate communities through their diverse content. The Bioneer channel, known for its unique blend of fitness, science, and technology content, appears to play such a role. This analysis aims to understand the Bioneer's integrative position within its network and examine its influence in connecting varied YouTube communities. By analyzing the network's topology, community structures, and the Bioneer's centrality, we seek to elucidate the channel's role in fostering content diversity and inter-community engagement on YouTube. The study will provide insights into how single nodes can serve as gateways to a broader range of content, thereby enriching the viewer's experience and contributing to the cohesion of the YouTube ecosystem, focusing specifically on the Bioneers connections at a subscription depth of 1, to uncover the intricate web of relationships and influences that shape its digital ecosystem.

The decision to focus on the Bioneer channel and at that particular depth of analysis was informed by an initial examination of various fitness channels' networks. These preliminary investigations revealed that many of these channels' networks were either too small or too expansive for effective analysis at the available subscription depths offered from the toolkit used for the generation of the data. In contrast, Bioneer's network presented a balanced and manageable size, offering a comprehensive yet intricate network, ideal for detailed study. By analyzing the network at a subscription depth of one, we aim to gain a clear understanding of Bioneer's immediate relational sphere, shedding light on its direct connections and broader position within the YouTube ecosystem.

To conduct this network analysis, Bernhard Rieder's YouTube Data Tools modules were utilized, a toolkit adept at extracting detailed network data from YouTube. This ensures that our dataset is both comprehensive and relevant. For the analysis and visualization, a combination of Python with the NetworkX library and the Gephi Software Tool was utilized. This selection was based on their robust capabilities in network analysis and their flexibility in handling complex network data. Python and NetworkX facilitated data manipulation and advanced metric analysis, while Gephi provided advanced visualization capabilities, enabling us to interpret the network's structure and key players effectively.

The primary goal of this analysis is to identify the critical channels within the Bioneer network and what role does the Bioneer hold in it. This encompasses uncovering the most influential channels directly connected to Bioneer, understanding the nature of these connections, exploring the community structure within the network, and examining centrality measures to pinpoint channels that serve as crucial nodes. This analysis is not only expected to provide valuable insights into the Bioneer channel's network but also to contribute to a broader understanding of content dynamics and interaction patterns on YouTube.

Data Overview

The dataset for the Bioneer YouTube channel network comprises several key attributes for each node, which represents an individual YouTube channel within the network:

- **Label**: This attribute represents the name of the YouTube channel, serving as a primary identifier within the network.
- **Is Seed**: A binary indicator denoting whether the channel was used as a seed in the network extraction process. A seed channel is typically the starting point for crawling the network.
- **Seed Rank**: If the channel was used as a seed, this numerical value indicates its rank or order in the seeding process, which may affect the network structure as it unfolds from this point.
- **Subscriber Count**: Reflecting the popularity and reach of the channel, this figure denotes the number of subscribers the channel has acquired.
- **Video Count**: This count represents the total number of videos that the channel has published, offering insight into the channel's content volume.
- **View Count (100s)**: The aggregate number of views across all videos of the channel, presented in hundreds for ease of data processing and analysis.
- **Country**: The registered country of the channel, which can influence the channel's audience demographic and content reach.
- **Published At**: The date when the channel was created, providing context for its age and potentially its growth trajectory.
- **Days Active**: A count of the number of days the channel has been active, from its creation to the present day or the last day of data collection.

These attributes offer a foundational understanding of each channel's presence, activity, and influence within the network, providing a basis for both quantitative and qualitative analysis of the Bioneer channel's ecosystem on YouTube. For our analysis, the columns **Label**, **Subscriber Count, View Count, Video Count** and **Country** were used.

1. Graphical Representation of the Network

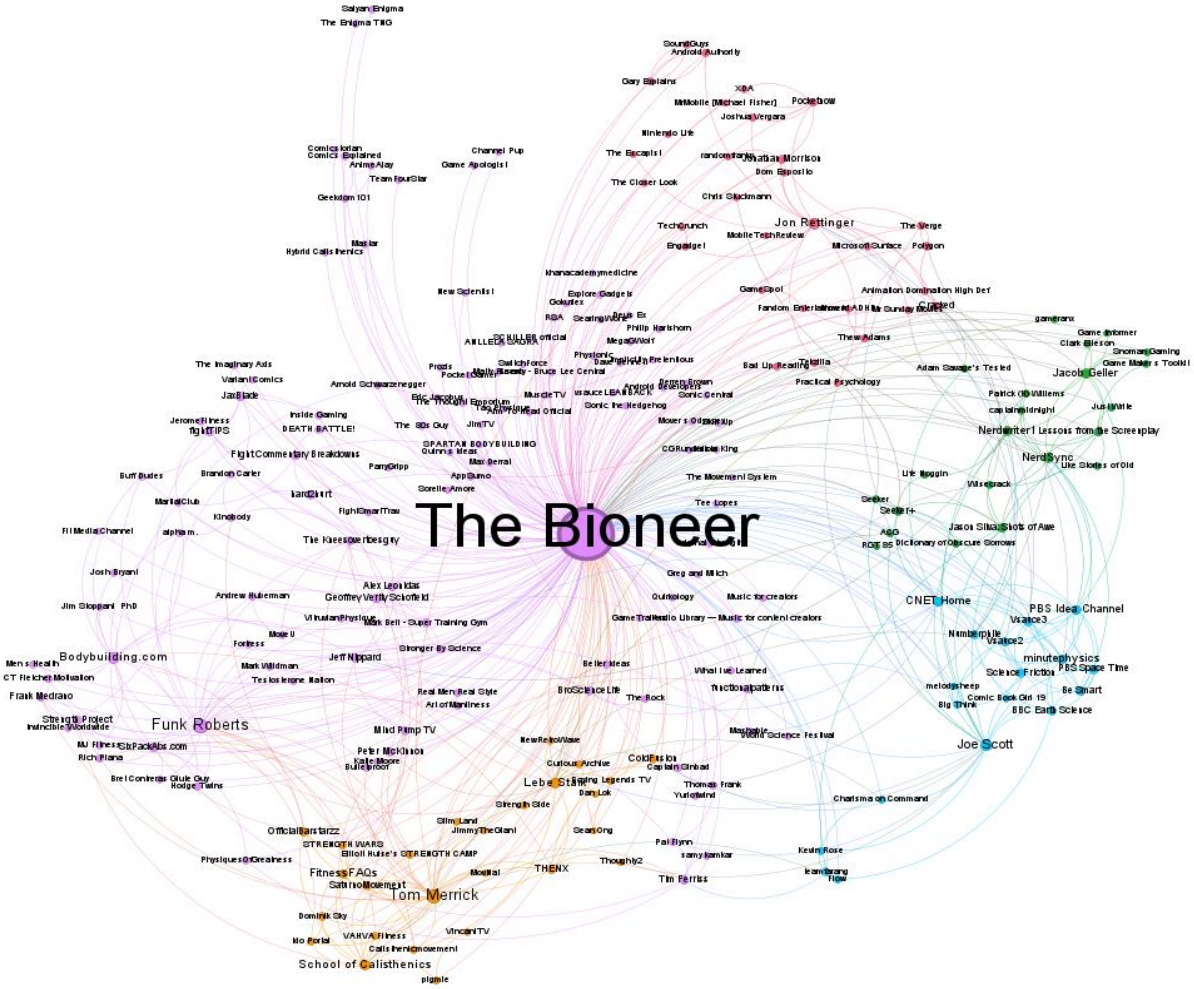
Utilizing the powerful visualization capabilities of Gephi, the Bioneer YouTube channel network has been rendered with precision and aesthetic clarity. This graphical representation is the result of a process where the node (each representing a unique YouTube channel) have been deliberately filtered to include only those with a subscriber count ranging from 50.000 to 8.318.210 (numbers chose from Z-Score outlier analysis). This filtering criterion ensures that the network's portrayal is not skewed by the overwhelming presence of colossal channels. Instead, it provides a balanced view that highlights the medium to highly influential channels without overgeneralization of the communities. With this filtering we end up with a network of 221 nodes with 604 edges.

The ForceAtlas2 layout algorithm, a feature of Gephi's suite of tools, was employed to spatially distribute the nodes, simulating a force-directed placement that naturally emphasizes the network's underlying structure. The algorithm's utility is evident in the way it magnifies the relational dynamics at play, organizing the channels in a way that intuitively showcases the interconnections at a subscription depth of 1.

Each connection or edge within this network, uniformly assigned a weight of one, signifies a relationship of equivalence, allowing us to observe the network without bias towards the number of shared subscribers. The nodes are imbued with colors based on their modularity class, calculated from Gephi using the modularity function, revealing the presence of distinct communities within the network. These vibrant clusters represent groups of channels that are more densely connected to one another than to the rest of the network, likely reflecting shared content themes or audience crossover. Also, the size of the nodes is equivalent to their degree, which reflects the number of connections each node has to other nodes.

The central nodes emerge as hubs of activity and influence, indicative of their prominent role within their modularity classes. These are the channels that, despite not being the largest on the platform by subscriber count, hold significant sway in their domains. The peripheral nodes, on the other hand, represent the more specialized interests within the Bioneer network—channels that cater to particular niches or are on the trajectory to greater connectivity and influence.

Through Gephi's advanced rendering and the strategic filtering of channels based on the degree measures, this visualization transcends mere aesthetic representation. It becomes a tool for insight, allowing us to understand the Bioneer channel's position within a broader context, pinpoint other critical influencers, and discern the subtle yet powerful forces that govern the digital ecosystem of YouTube in general.



The provided visual representation clearly positions the Bioneer channel as a central node within its network, indicated by the numerous connections radiating from it. This centrality is of considerable significance for several reasons:

1. **Influence and Reach:** The Bioneer channel's multitude of connections denotes a broad reach and major influence within the network. Channels with many connections often act as influencers or hubs within their community, suggesting that Bioneer plays a pivotal role in disseminating information, trends, or innovations within its sphere.
2. **Information Flow:** Bioneer's central position implies it serves as a primary source of content or as a crucial intermediary for information flow between other channels. The channel's content can quickly disseminate through its network, influencing and being influenced by other connected nodes.
3. **Community Engagement:** The extent of Bioneer's connections suggests strong community engagement. It can be inferred that the channel does not exist in

isolation but is a part of a vibrant community of content creators and consumers who interact with and respond to its content.

4. **Collaborative Potential:** The central node status of the Bioneer channel also highlights its potential as a collaborative partner. For other channels within the network, establishing a connection with Bioneer might mean gaining access to a wider or a different audience or even tapping into the channel's content creation expertise.
5. **Network Health and Sustainability:** The health of a network can often be gauged by the robustness of its central nodes. Bioneer's numerous connections indicate a healthy and sustainable network capable of withstanding fluctuations in individual channel performances.
6. **Content Strategy Insights:** Analyzing the Bioneer channel's connections provides insights into its content strategy, particularly how it positions itself among various content themes and communities within YouTube.

2. Basic Topological Properties

Now having inserted our data into Gephi, we observe 221 nodes, each signifying a unique YouTube channel. The nodes are interlinked by 604 edges, indicating a web of connections where each edge represents a relationship or interaction between channels, such as shared audiences or collaborative content.

Gephi's computational capabilities reveal that the network diameter is 9. This metric is pivotal as it indicates the longest shortest path within the network, providing insight into the extent of the network's spread. It suggests that even the most distant channels in the network are separated by no more than nine steps, highlighting the interconnected nature of this digital ecosystem.

Moreover, the average path length across the network is 4.274. This number offers a snapshot of the network's compactness, suggesting that, on average, any channel is separated from another by just over four steps. This relatively (compared to real world networks) short path length underscores a dense network topology, hinting at a closely knit community where information can flow swiftly and efficiently from one channel to another.

Calculating the reciprocity of the network using networkX, we get a result of around 0.083. Reciprocity in a network context refers to the tendency of node pairs to form mutual connections - in other words, the likelihood that if node A is connected to node B, node B is also connected to node A. In our case this means that if one channel (node) follows or references another channel, there is only about an 8.3% chance that the second channel also follows or references the first one. The low reciprocity influences the community structure within the network. It indicates that while there are communities or groups centered around certain nodes, the inter-community interactions are limited, reinforcing the dominance of certain nodes within their respective communities.

Lastly, calculating the degree assortativity coefficient (**Out-In,** which is the default option) for our network using networkX, we get a score of approximately -0.447, that reflects the dynamics of influencers within the network. Specifically, it indicates that nodes (channels in this case) tend to connect with others that have a different number of connections, and this heterogeneity means that well-connected channels often link to less-connected ones. A high negative assortativity coefficient suggests, also, a non-uniform network structure. Rather than a cohesive, evenly connected topology, the network is likely characterized by a few highly connected nodes (hubs) and many nodes with fewer connections. This pattern is typical of scale-free networks where some nodes act as central aggregators or disseminators of information.

These metrics serve not only as descriptors of the network's size and scope but also as indicators of the Bioneer channel's potential reach and influence within its niche on the YouTube platform and to other niches as well. They provide a framework for understanding how the channel fits within the larger narrative of the YouTube content creation landscape.

3. Component Measures

The structural integrity and cohesiveness of the Bioneer YouTube channel network are illuminated by its component measures, which have been calculated using both Gephi and NetworkX.

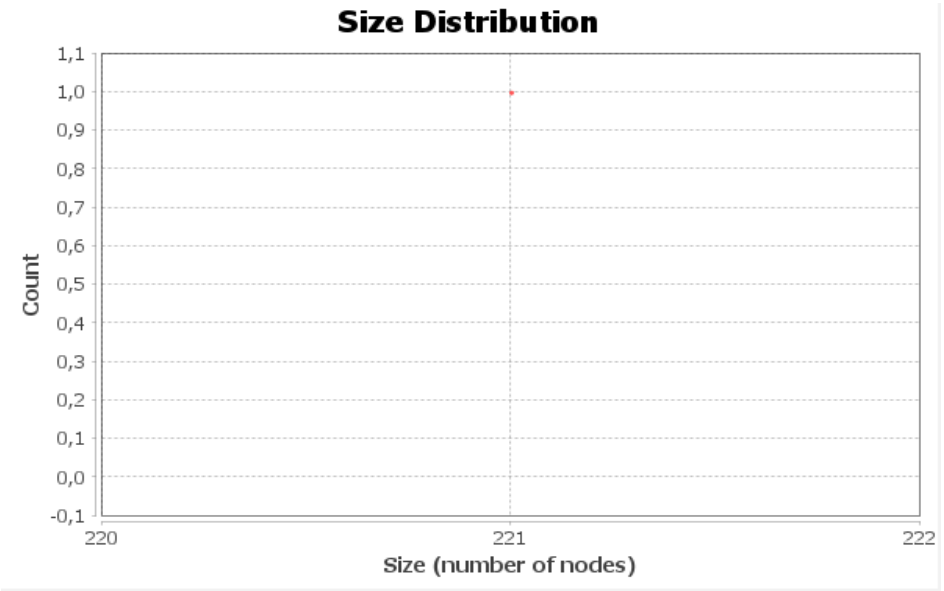
```
# Find the number of strongly and weakly connected components
num_strongly_connected_components = nx.number_strongly_connected_components(G.to_directed())
num_weakly_connected_components = nx.number_weakly_connected_components(G.to_directed())
print("Number of strongly connected components:", num_strongly_connected_components)
print("Number of weakly connected components:", num_weakly_connected_components)
```

```
Number of strongly connected components: 183
Number of weakly connected components: 1
```

The network is characterized by a single connected component, as evidenced by the fact that there is only one weakly connected component. This suggests a high degree of interconnectedness within the network, with no isolated clusters of channels standing apart from the rest.

There are 183 strongly connected components and only one weakly connected component. Strongly connected components are subnetworks where there is a path in each direction between each pair of nodes. The presence of numerous strongly connected components within a single weakly connected component indicates that while every channel can be reached from every other channel (hence the singular weakly connected component), the network has many subsets of channels that have more robust two-way interaction.

NetworkX computations shed light on the size of the giant component, which encompasses all 221 nodes in the network. This dominance of one giant component is indicative of a 'small world' network, where the majority of nodes are not just connected but are part of a single, extensive web of connections. The size distribution chart from Gephi, which spikes at count 221, resonates with this, showing that all nodes in the network are part of this giant component, and there are no smaller, isolated components.



Here we can see how the calculations were done using the Networkx library through python:

```
# Calculate component measures
# Get the number of connected components
num_connected_components = nx.number_connected_components(G.to_undirected())

# Identify the giant component
giant_component = max(nx.connected_components(G.to_undirected()), key=len)
giant_component_size = len(giant_component)

# Get the size distribution of the components
component_sizes = [len(c) for c in sorted(nx.connected_components(G.to_undirected()), key=len, reverse=True)]

print("Number of connected components:", num_connected_components)
print("Size of giant component:", giant_component_size)
print("Size distribution of components:", component_sizes)
```

```
Number of connected components: 1
Size of giant component: 221
Size distribution of components: [221]
```

The existence of a single, all-encompassing giant component is a testament to the network's robustness. It is reflective of a vibrant, interconnected community where information, influence, and content can disseminate freely and efficiently across the entire network. This structural characteristic is particularly beneficial for the central channel, Bioneer, as it implies a potential for wide reach and significant impact within the network. The network's architecture, thus, not only facilitates but also possibly amplifies the channel's visibility and engagement across the YouTube platform.

4. Degree Measures

4.1 Average Degree

The degree measures within the Bioneer YouTube channel network provide a granular understanding of individual nodes' connectivity. An average degree calculated as 2.733, using Gephi, reflects the average number of connections each node has to other nodes. This relatively modest average degree suggests that while some channels may have numerous connections, many channels maintain a more selective number of relationships within the network, potentially fostering stronger, more meaningful interactions. Calculating the average degree, using networkX, we get a result of 5.466, but we must keep in mind the fact that the average degree calculated by Gephi is approximately half of the average degree of interest, since Gephi defines ”average degree” as the average of average in-and out-degrees (source: https://web.stanford.edu/~jacksonm/Gephi-instruction-1-updatedApr2015.pdf , page 5).

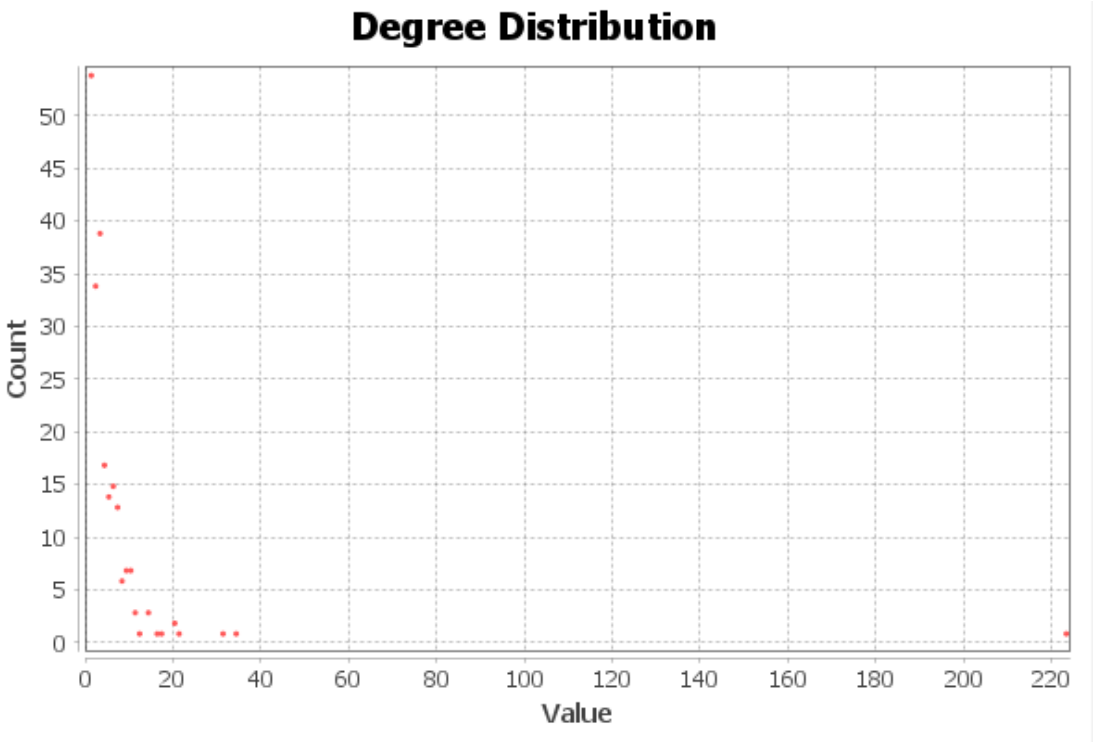
The lower average degree of 2.733 points to a network where channels exhibit selective interaction, with most having two or three connections to other channels. This selectivity is conducive to forming niche communities or clusters characterized by intense interaction among a smaller set of channels, indicative of close content alignment or audience overlap.

Furthermore, the implications of a lower average degree on the network's resilience are notable. Networks with a lower average degree can be more vulnerable to the loss of highly connected nodes, which could disrupt the network's structure and the flow of information. However, this also suggests that influence may be more democratically distributed among a larger number of channels, leading to a potentially more equitable network structure.

The average degree also informs us about the rate at which information spreads through the network, which in this case, might be slower but more effective due to the focused nature of the connections. Additionally, the network does not overly rely on a small number of hubs to maintain its structure, hinting at a certain level of robustness against targeted disruptions.

4.2 Degree Distribution

From the degree distribution graph, we can deduce that the minimum node degree in the network is very low, indicating the presence of channels with minimal connections, which may be newer or more niche content creators.



In contrast, the maximum node degree significantly exceeds two hundred, highlighting a highly connected hub within the network. This central hub, due to its extensive connections, plays a pivotal role in content dissemination and audience engagement.

We can also check more accurately using the networkX library the distribution of the degrees from the provided calculations below, where we can clearly see the distribution along with the **minimum** (1) and **maximum** (221) degree values:

```
# Calculate degree distribution for the filtered graph
filtered_degrees = [degree for node, degree in G.degree()]
filtered_degree_distribution = {degree: filtered_degrees.count(degree) for degree in set(filtered_degrees)}

# Calculate maximum and minimum degree for the filtered graph
filtered_max_degree = max(filtered_degrees)
filtered_min_degree = min(filtered_degrees)

filtered_degree_distribution, filtered_max_degree, filtered_min_degree
```

```
({1: 54,
  2: 34,
  3: 39,
  4: 17,
  5: 14,
  6: 15,
  7: 13,
  8: 6,
  9: 7,
  10: 7,
  11: 3,
  12: 1,
  14: 3,
  16: 1,
  17: 1,
  20: 2,
  21: 1,
  31: 1,
  34: 1,
  223: 1},
 223,
 1)
```

We can also see which channels hold these values:

```
# Extracting channel labels along with their degree value
channel_labels_with_degree = [(G.nodes[node]['label'], degree) for node, degree in G.degree()]

# Displaying the entries to check the data
channel_labels_with_degree.sort(key=lambda x: x[1], reverse=True)
```

```
channel_labels_with_degree
```

```
[('The Bioneer', 223),
 ('Funk Roberts', 34),
 ('Tom Merrick', 31),
 ('Joe Scott', 21),
 ('Bodybuilding.com', 20),
 ('Jon Rettinger', 20),
 ('Lebe Stark', 17),
 ('School of Calisthenics', 16),
 ('PBS Idea Channel', 14),
 ('CNET Home', 14),
 ('Jacob Geller', 14),
 ('NerdSync', 12),
 ('FitnessFAQs', 11),
 ('Nerdwriter1', 11),
 ('minutephysics', 11),
 ('Strength Project', 10),
 ('Be Smart', 10),
```

We can clearly see from these results that the Bioneer channel itself commands the network with an impressive 223 connections, solidifying its position as the most prominent hub. This high degree of connectivity signifies that the Bioneer is not just a passive participant but a pivotal conduit through which content and information flow, connecting a vast array of channels across the network.

Funk Roberts and Tom Merrick follow as notable entities with 34 and 31 connections, respectively. Their considerable number of connections suggests that they are key influencers within their niches, sharing a robust audience engagement and content collaboration with other channels.

Channels such as Joe Scott and Bodybuilding.com, each with around twenty connections, also play substantial roles in the network. Their relatively high degree indicates they are well-integrated within the network, acting as important nodes through which information and influence can disseminate.

As we move towards channels with a lower degree of connectivity, like School of Calisthenics, PBS Idea Channel, and Jacob Geller, each with connections ranging from 14 to 16, we see a shift towards channels that might be considered more specialized or focused on their content and audience interaction. Despite having fewer connections, these channels are integral to the network's diversity, offering unique content and perspectives.

The presence of channels such as NerdSync, FitnessFAQs, and Nerdwriter1, each with a degree around 11 to 13, further illustrates the network's reach into different content areas, from fitness to pop culture and beyond.

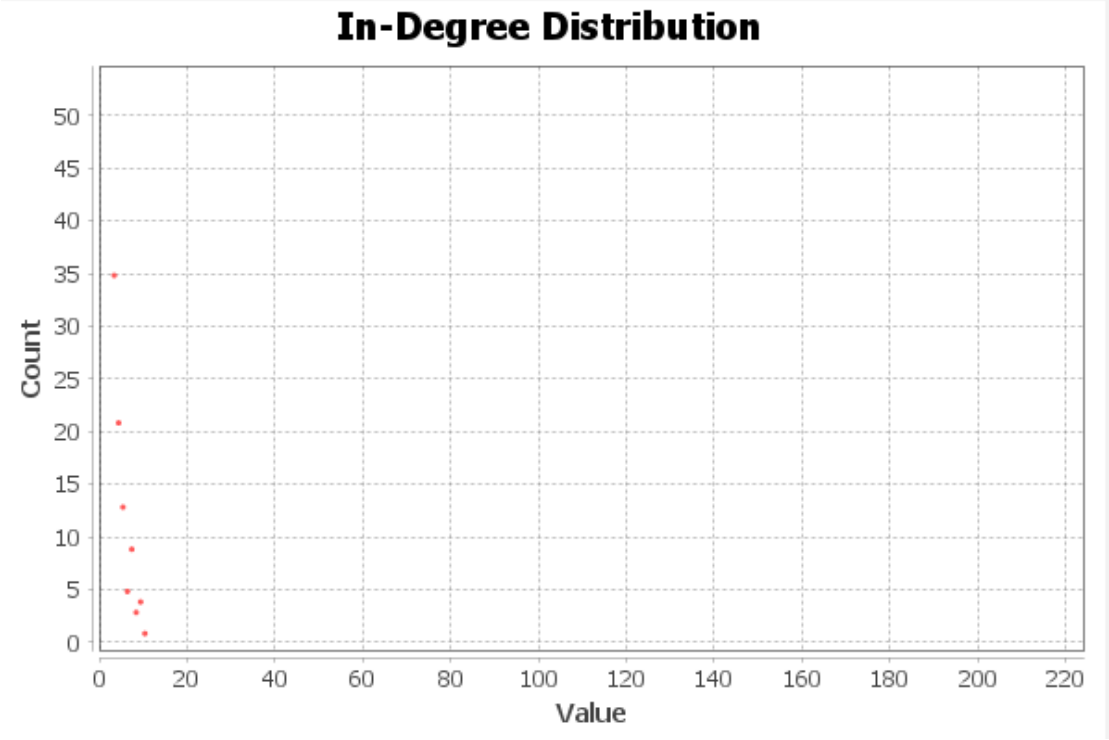
Channels at the lower end of the connectivity spectrum, like Jeff Nippard and PBS Space Time, although less connected with nine connections each, contribute to the network's depth. Their specialized content attracts a dedicated but possibly more niche audience.

In summary, the degree measures in the Bioneer network highlight the potential influence of certain central channels and the collective engagement across the network. The range from the minimum to the maximum degree emphasizes the network's diversity and the various roles different channels play within this digital ecosystem.

4.3 In-Degree Distribution

The in-degree distribution graph of the Bioneer YouTube channel network offers insights into the nature of incoming connections to various nodes within the network. In-degree in a network graph indicates the number of edges directed towards a node, which in the context of YouTube network could represent the number of channels that directly reference or are influenced by a given channel.

The distribution displayed in the graph suggests that most channels have a relatively low number of incoming connections, with a significant concentration of nodes having in-degrees close to zero. This pattern is indicative of a network where many channels serve as content consumers or peripheral participants rather than content disseminators or central hubs.



As the in-degree value increases, the frequency of channels with that degree decreases, which is typical of networks with a hierarchical structure. In such networks, there are fewer channels that attract many connections. These channels can be considered authorities or influencers within the network due to their apparent importance to others within the same community.

The graph does not show a smooth decline but rather a steep drop-off from the mode, emphasizing that highly influential channels are quite rare within this network. We can examine closely the channels and their in-degree values using networkX:

```
# Extracting channel labels along with their in-degree value
channel_labels_with_in_degree = [(G.nodes[node]['label'], in_degree)
                                 for node, in_degree in G.in_degree()]
channel_labels_with_in_degree.sort(key=lambda x: x[1], reverse=True)
```

```
channel_labels_with_in_degree
```

```
[('minutephysics', 10),
 ('FitnessFAQs', 9),
 ('Strength Project', 9),
 ('Jeff Nippard', 9),
 ('Vsauce2', 9),
 ('The Kneesovertoesguy', 8),
 ('Numberphile', 8),
 ('Vsauce3', 8),
 ('Peter McKinnon', 7),
 ('PBS Idea Channel', 7),
 ('ColdFusion', 7),
 ('Jonathan Morrison', 7),
 ('Be Smart', 7),
 ('Nerdwriter1', 7),
 ('BroScienceLife', 7),
 ('OfficialBarstarzz', 7),
 ('fightTIPS', 7),
```

In examining the in-degree distribution, we see channels like minutephysics, FitnessFAQs, Strength Project, Jeff Nippard, and Vsauce, with relatively high in-degrees between 9 and 10. These channels are pivotal influencers within the network, serving as key reference points and inspirations for a multitude of other channels. Their high in-degree values underscore their notable influence over the network's content direction and thematic focus, marking them as central hubs in the community.

Slightly lower in the hierarchy, yet still influential, are channels such as The Kneesovertoesguy, Numberphile and Peter McKinnon, with in-degrees ranging from 7 to 8. These channels represent specialized influencers, commanding respect and attention within their specific niches or topics. Their content and expertise resonate within their domains, contributing to the diversity and richness of the network.

The network also encompasses a wide array of channels with in-degrees from 1 to 6, including Elliott Hulse's STRENGTH CAMP, Wisecrack, Bodybuilding.com, and Calisthenicmovement. These channels contribute to more specific areas or themes within the network. While they may not wield broad influence across the entire network, they hold significant importance within their segments, offering unique perspectives and specialized content.

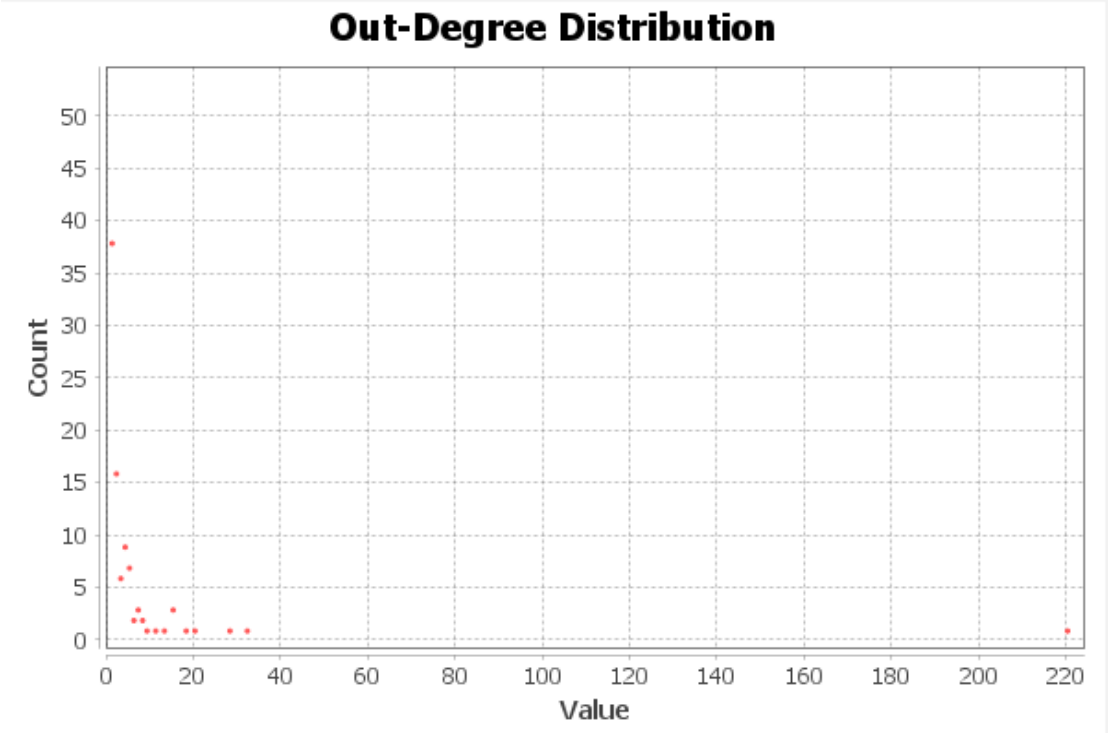
An interesting note to add is that the Bioneer, with an in-degree of three, occupies a unique position. It does not rank among the highest in terms of direct connections, yet its role in forming significant triangles and its specialized content focus make it a notable node. The Bioneer likely (as we will confirm later in chapter 7) acts as a bridge within the network, linking various themes and communities through its diverse content.

At the lower end of the spectrum are channels with minimal in-degrees, often just one. These channels are peripheral in the network's interaction dynamics, focusing more on content consumption or representing newer or more specialized entities with emerging influence.

To sum up, the overall in-degree distribution indicates a network where influence and content dissemination are not dominated by a single channel but are instead distributed among a variety of key players.

4.4 Out-Degree Distribution

The out-degree distribution graph provides a visual representation of the number of outbound connections each node has. The out-degree in a network graph indicates how many edges are directed from a node, which, in the context of YouTube, could represent the number of other channels each channel references or is influenced by.



The graph displays a skewed distribution with most nodes clustered at the lower end of the out-degree spectrum. This clustering suggests that most channels in the network tend to reference or connect with only a small number of other channels. Such a pattern is indicative of a network where content creators focus their interactions and references on a select group of other channels, due to similarities in content, audience, or collaborative relationships.

A few nodes exhibit a significantly higher out-degree, evidenced by the outliers on the right-hand side of the distribution graph. These nodes act as broadcasters within the network, likely disseminating content to, or influencing, a wide array of other channels. The presence of such nodes is crucial for the spread of information and ideas

across the network, as they can potentially connect disparate parts of the network through their extensive outbound connections.

Now, we will get the out-degree measures for each channel to analyze them in a descending order:

```
# Extracting channel labels along with their in-degree value
channel_labels_with_out_degree = [(G.nodes[node]['label'], out_degree)
                                  for node, out_degree in G.out_degree()]
channel_labels_with_out_degree.sort(key=lambda x: x[1], reverse=True)
```

```
channel_labels_with_out_degree
```

```
[('The Bioneer', 220),
 ('Funk Roberts', 32),
 ('Tom Merrick', 28),
 ('Joe Scott', 20),
 ('Jon Rettinger', 18),
 ('Bodybuilding.com', 15),
 ('School of Calisthenics', 15),
 ('Lebe Stark', 15),
 ('CNET Home', 13),
 ('NerdSync', 11),
 ('Jacob Geller', 9),
 ('JaxBlade', 8),
 ('SixPackAbs.com', 8),
 ('PBS Idea Channel', 7),
 ('Geoffrey Verity Schofield', 7),
 ('Cracked', 7),
 ('Science Friction', 6),
```

The out-degree distribution of the Bioneer YouTube channel network illustrates the broadcasting reach of each channel within the network. The Bioneer, with an out-degree of 220, clearly dominates as the principal broadcaster, indicative of a channel that actively references or is influenced by a large swath of the network, thereby shaping the network's narrative landscape.

Channels like Funk Roberts and Tom Merrick, with out-degrees of 32 and 28 respectively, also stand out as significant broadcasters within their niches. Their high out-degree values suggest they are not only creating content that others in the network find valuable but are also actively engaging with a diverse set of channels, bridging various sub-communities within the broader network.

Joe Scott and Jon Rettinger, with 20 and 18 connections respectively, along with others like Lebe Stark and Bodybuilding.com, show a strong presence within the network, acting as important nodes that contribute to the dissemination of ideas and content across different audience segments.

Notably, the channels with a single outbound connection signify a more focused approach to content sharing, indicating specialized or highly targeted content strategies. These channels may play unique roles within the network such as Geoffrey Verity

Schofield in the natural bodybuilding space, influencing specific subsets of the community or providing niche content that complements the broader content ecosystem.

In summary, the out-degree distribution highlights the Bioneer's role as the predominant broadcaster within the network, with a few other channels also playing key roles in content distribution. The presence of many channels with lower out-degrees illustrates a network where most content creators are selective in their engagements, leading to a diverse yet interconnected community where each channel contributes to the network's richness and complexity.

4.5 In-Degree vs Out-Degree

When comparing the in-degree and out-degree distributions within the Bioneer YouTube channel network, we observe essential characteristics that define the nature of interactions among the channels.

Comparative Insights:

- Channels with high in-degrees but lower out-degrees are considered content authorities or influencers, attracting attention from various nodes within the network without necessarily reciprocating the interaction.
- Conversely, channels with high out-degrees but lower in-degrees, such as the Bioneer, are actively engaging with many others, in an attempt to grow their influence or share diverse content.
- The presence of channels with both low in-degree and out-degree suggests a more niche positioning within the network, serving specific content areas or audiences.

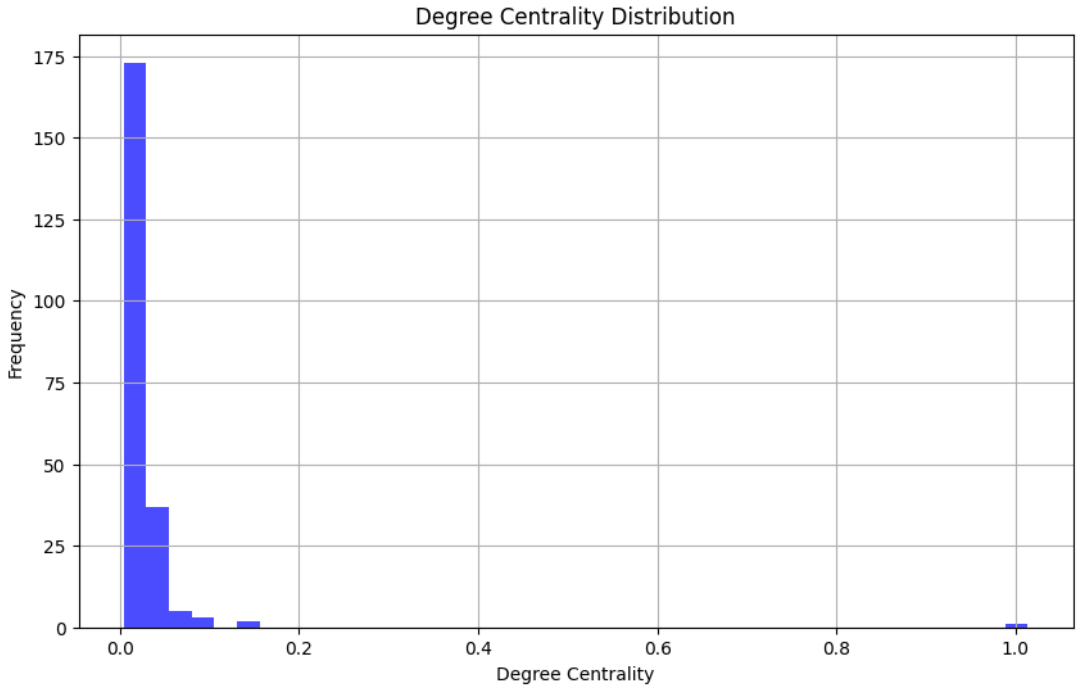
4.6 Weighted Degree Distribution

The weighted degree distribution in the Bioneer YouTube channel network could provide a nuanced perspective on the connectivity of nodes, accounting for the strength of connections. However, in this network, the weighted degree distribution mirrors the standard degree distribution because the weights of all edges are equal to one. This uniformity means that each connection contributes equally to a node's weighted degree, making it identical to the unweighted degree—a simple count of connections. Therefore, the insights gleaned from the weighted degree distribution do not differ from those obtained from the normal degree distribution. Channels with a higher degree, weighted or unweighted, are central to the network's structure and content flow, while those with lower degrees occupy more specialized niches or have a more selective reach within the network.

5. Centrality Measures

5.1 Degree Centrality

Degree centrality is a fundamental measure in network analysis that quantifies the importance of a node based on the number of connections it has to other nodes. In the context of the Bioneer network, degree centrality indicates how many different channels a particular channel is directly connected to. Knowing which channels have high degree centrality can be valuable for strategic networking, collaborations, and marketing efforts. These channels can be targeted for partnerships or promotions due to their extensive reach within the network. Since, Gephi does not have the functionality for the calculation of the degree centrality distribution we will utilize matplotlib and networkX through python to create a histogram of the distribution.



From the histogram, we can see a pronounced skew towards the left, indicating that a majority of the channels have a low degree centrality. This means that most channels are connected to only a few other channels within the network. Such a distribution is typical for networks where a lot of channels have specialized content, or a narrower audience reach.

We can see the degree centrality of each channel separately using again networkX:

```python
# Get degree centrality for each node with their labels
degree_centrality = [(G.nodes[node]['label'], degree) for node, degree in nx.degree_centrality(G).items()]

# Sort the nodes by their degree centrality
degree_centrality.sort(key=lambda x: x[1], reverse=True)
degree_centrality
```

```
[('The Bioneer', 1.0136363636363637),
 ('Funk Roberts', 0.15454545454545454),
 ('Tom Merrick', 0.1409090909090909),
 ('Joe Scott', 0.09545454545454544),
 ('Bodybuilding.com', 0.09090909090909091),
 ('Jon Rettinger', 0.09090909090909091),
 ('Lebe Stark', 0.07727272727272727),
 ('School of Calisthenics', 0.07272727272727272),
 ('PBS Idea Channel', 0.06363636363636363),
 ('CNET Home', 0.06363636363636363),
 ('Jacob Geller', 0.06363636363636363),
 ('NerdSync', 0.05454545454545454),
 ('FitnessFAQs', 0.049999999999999996),
 ('Nerdwriter1', 0.049999999999999996),
 ('minutephysics', 0.049999999999999996),
 ('Strength Project', 0.045454545454545456),
 ('Be Smart', 0.045454545454545456),
```

With a degree centrality of over one, the Bioneer stands out as a highly central node. This unusually high centrality score suggests that he is connected to every other node in the network and even has multiple types of relationships with some nodes. It indicates a dominant, pivotal role in the network, acting as a major hub or connector.

Channels like Funk Roberts, Tom Merrick, and Joe Scott, with centrality scores ranging from around 0.09 to 0.15, hold moderate influence within the network. They are well-connected but do not have the same level of connectivity as the Bioneer. and act as secondary hubs within their respective sub-communities.

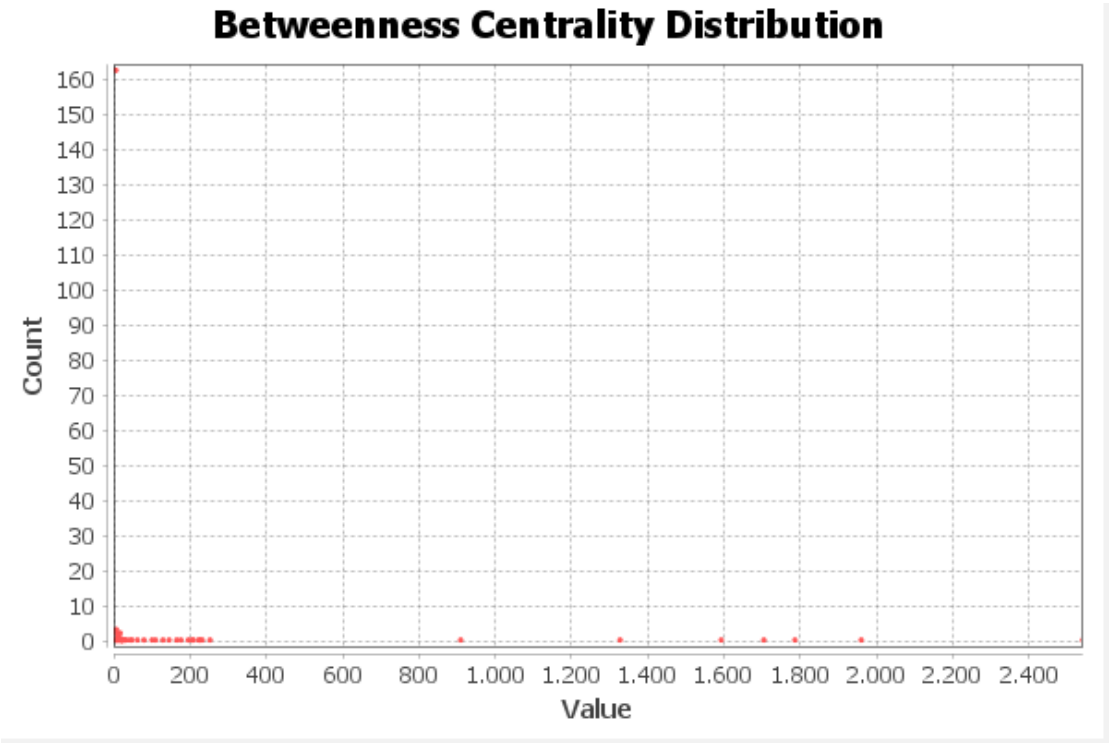
Channels such as Bodybuilding.com, Jon Rettinger, and Lebe Stark have lower centrality scores, indicating they are more niche or specialized. While they are part of the network, their role is less central compared to the Bioneer or even the moderately central channels. They cater to specific audiences or topics within the broader fitness and bodybuilding theme.

Lastly, a considerable number of channels, particularly those with centrality scores around 0.0045, like Sonic the Hedgehog, CGRundertow, and Matty Fusaro, are peripheral in the network. They have fewer connections and represent specialized niches or smaller audience segments within the larger fitness and bodybuilding community.

To conclude, the skewness of the distribution points to a network that is centralized around the Bioneer. The presence of some channels with moderate degree centrality suggests that while there are some nodes that function as secondary hubs, there is a clear hierarchy in terms of connectivity and influence.

5.2 Betweenness Centrality

Betweenness centrality is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex.



The betweenness centrality distribution graph from Gephi shows that most nodes in the Bioneer network have a low betweenness centrality, with values clustered close to zero. This suggests that most channels are not frequently positioned on the shortest paths between other channels, implying they have a more peripheral role in the network in terms of control over the flow of information.

However, there are some channels with a much higher betweenness centrality. These are the channels that act as important bridges within the network. They have significant control over the information flow, connecting various parts of the network and facilitating interactions between channels that might not be directly connected. Channels with high betweenness centrality can influence the spread and flow of content across the network and can be critical for the cohesion and dissemination of information within the network.

The few nodes with high betweenness centrality may indicate that while the network has well-defined clusters or communities, these key nodes provide the essential links between these groups. Their role is crucial in maintaining the network's connectivity and can be considered as strategic points of influence within the YouTube channel network.

Let us now get the betweenness centrality for each channel normalized from networkX:

```
# Calculating betweenness centrality for the directed graph
betweenness_centrality = nx.betweenness_centrality(G, normalized=True)

# Extracting channel labels along with their betweenness centrality
channel_labels_with_betweenness = [(G.nodes[node]['label'], betweenness_centrality[node])
                                   for node in G.nodes()]

channel_labels_with_betweenness.sort(key=lambda x: x[1], reverse=True)
```

```
channel_labels_with_betweenness
```

```
[('The Bioneer', 0.05264632627646327),
 ('Fight Commentary Breakdowns', 0.040587380655873806),
 ('hard2hurt', 0.03697592361975924),
 ('Tom Merrick', 0.0352843503528435),
 ('Kinobody', 0.03295973432959735),
 ('SaturnoMovement', 0.027452608274526086),
 ('FitnessFAQs', 0.018776809187768095),
 ('STRENGTH WARS', 0.005126608551266086),
 ('teamfarang', 0.004701120797011208),
 ('Flow', 0.0045765877957658786),
 ('BBC Earth Science', 0.004514321295143213),
 ('Funk Roberts', 0.004223744292237443),
 ('PBS Idea Channel', 0.004137263041372631),
 ('Bodybuilding.com', 0.003947004289470043),
 ('PBS Space Time', 0.003542272035422721),
 ('Jason Silva: Shots of Awe', 0.003313961533139615),
 ('Nerdwriter1', 0.0028988515289885153),
```

The Bioneer channel itself shows a normalized betweenness centrality of 0.0526, indicating it plays a significant role in mediating the flow of information within the network. It acts as a gatekeeper, connecting disparate parts of the network and facilitating the spread of content across various communities.

Channels such as Fight Commentary Breakdowns, hard2hurt, and Tom Merrick also have high betweenness centralities, with values like 0.04058 and 0.03697, and 0.03528, respectively. These channels are crucial connectors within their niches, enabling communication and content sharing between clusters that might otherwise be disconnected.

Kinobody and SaturnoMovement also emerge as significant intermediaries within the network, with betweenness centralities of 0.033 and 0.027. Their roles in the network involve bridging fitness enthusiasts with different focuses, suggesting that they are influential in shaping the network's broader content strategy and this is evident, as they are channels that do not specialize on a specific specter of fitness.

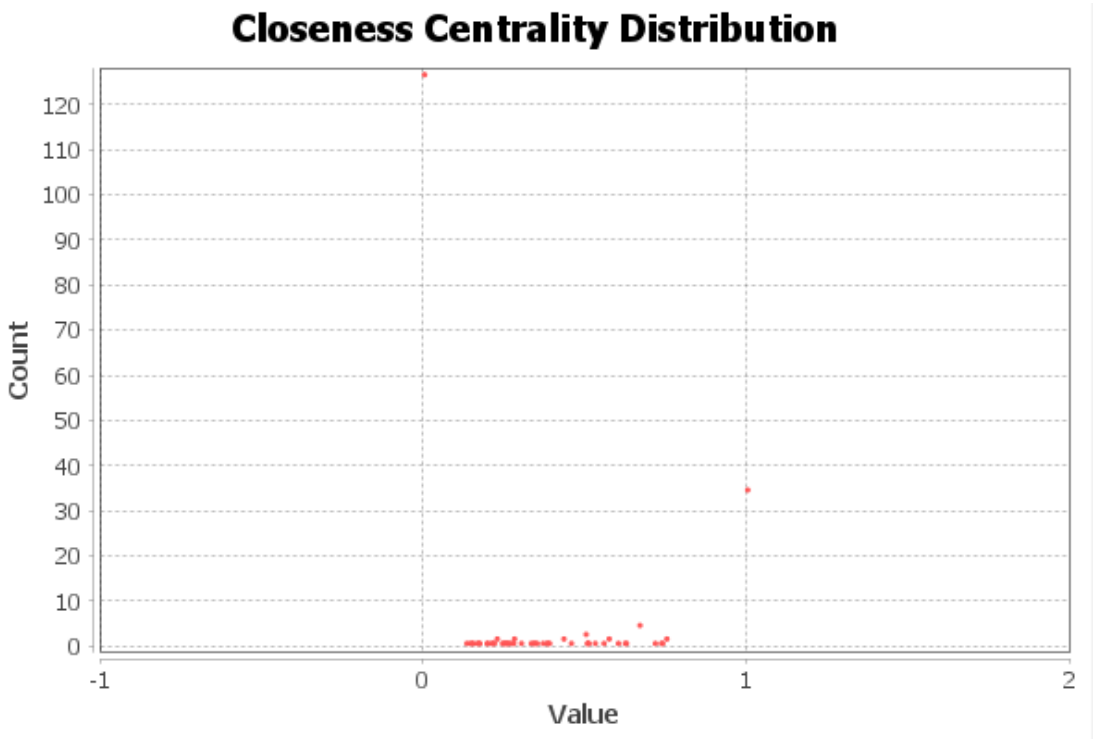
On the lower end, channels like FitnessFAQs, STRENGTH WARS, and teamfarang display smaller yet still notable betweenness centralities, such as 0.01877 and 0.00512. Even with these lower values, these channels contribute to the network's connectivity by linking together various smaller communities within the larger ecosystem.

Interestingly, a substantial number of channels have a betweenness centrality of zero, indicating that these channels do not lie on the shortest path between any other pair of channels. They may still hold influence within their immediate connections, but they do not serve as broader connectors within the network's architecture.

Overall, the betweenness centrality distribution underscores the multi-layered nature of the Bioneer network. It reveals not only which channels are central connectors but also the existence of many channels that, while not central, contribute to the network's diversity and the richness of content available within the YouTube fitness and lifestyle community.

5.3 Closeness Centrality

Closeness centrality is a measure of how close a node is to all other nodes in a network. It is calculated as the reciprocal of the sum of the shortest path distances from a node to all other nodes, giving us an idea of how quickly a node can interact with others in the network.



The closeness centrality distribution graph provided from Gephi suggests that most nodes in the Bioneer YouTube channel network are not centrally located, as indicated by the concentration of nodes with a lower closeness centrality value near zero. This implies that the average distance from these nodes to all others in the network is high, which can be characteristic of nodes that operate on the periphery of the network or within tightly knit communities with few outside connections.

However, a few channels show higher closeness centrality values, signifying that they are more central in the network. These channels are positioned in such a way that they can spread information efficiently or access content from others with relative

ease, making them potentially strategic for the flow of new ideas and trends within the network.

To get a more accurate picture of the closeness distribution, we will utilize the network library, more specifically we will use the Wasserman and Faust improved formula to calculate the closeness of each channel.

```
# Calculating closeness centrality for the directed graph
closeness_centrality = nx.closeness_centrality(G, wf_improved=True)

# Extracting channel labels along with their closeness centrality
channel_labels_with_closeness = [(G.nodes[node]['label'], closeness_centrality[node])
                                 for node in G.nodes()]

channel_labels_with_closeness.sort(key=lambda x: x[1], reverse=True)
channel_labels_with_closeness
```

```
[('Vsauce2', 0.07565256525652567),
 ('Jeff Nippard', 0.06888971684053652),
 ('Vsauce3', 0.06644268774703557),
 ('Numberphile', 0.05717377860235003),
 ('Jonathan Morrison', 0.05475935828877005),
 ('Wisecrack', 0.05419084461637653),
 ('Hodge Twins', 0.05171065493646138),
 ('ColdFusion', 0.051212121212121216),
 ('Frank Medrano', 0.05097902097902098),
 ('Strength Project', 0.05020661157024794),
 ('minutephysics', 0.04898236092265943),
 ('CT Fletcher Motivation', 0.04872994652406417),
 ('Tim Ferriss', 0.048484848484848485),
 ('The Kneesovertoesguy', 0.047603305785123964),
 ('Dictionary of Obscure Sorrows', 0.047141873278236913),
 ('PBS Idea Channel', 0.04566596194503171),
 ('BroScienceLife', 0.045454545454545456),
```

Closeness centrality, using the Wasserman and Faust improved formula, offers a refined view of a node's centrality by measuring its average distance to all other nodes in the network. This measure accounts for global positioning, rather than just local connectivity. The original formula for closeness centrality calculates it as the reciprocal of the sum of the shortest path distances from a node to all other nodes in the network. The Wasserman and Faust improved formula adjusts this by scaling the centrality value by the fraction of nodes reachable.

Vsauce2 emerges with the highest normalized closeness centrality of 0.07565, implying it can quickly interact with other channels across the network. This suggests Vsauce2's content or influence has the potential to rapidly reach various network segments and that makes sense, because it is the second channel of Vsauce, a very popular science and experiments-based channel.

Jeff Nippard and Vsauce3 also showcase high closeness centrality scores, indicating their central positioning in the network. These channels, along with Numberphile and Jonathan Morrison, appear well-placed to spread content efficiently due to their shorter path lengths to other nodes.

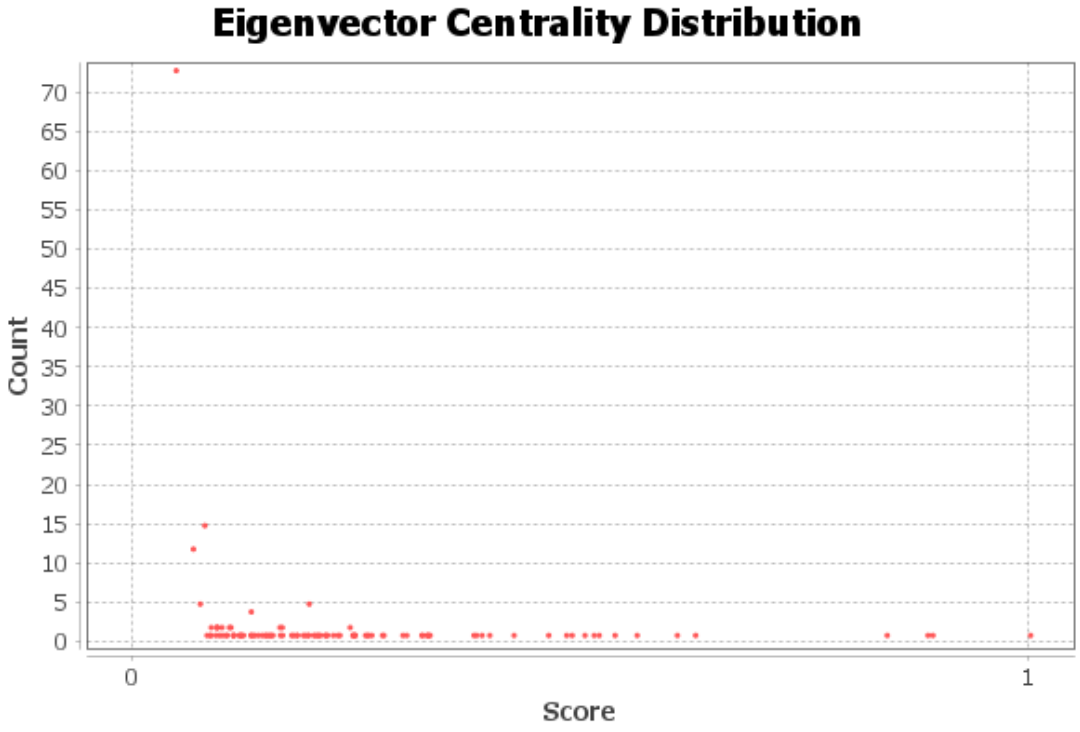
Channels such as Wisecrack and Hodge Twins hold intermediate positions in terms of closeness centrality, suggesting that while they may not be the quickest to reach every other channel, they still maintain a central position within their sub-networks.

Interestingly, the Bioneer, despite its high degree centrality, has a lower closeness centrality value. This implies that while the Bioneer has many direct connections, its average distance to all other nodes is greater, potentially due to its position relative to more peripheral or cluster-specific nodes.

The distribution of closeness centrality values across the network indicates a variety of roles. Some channels serve as central hubs, capable of quickly disseminating information, while others may be influencers within more localized community clusters. This diversity in node centrality underscores the complex fabric of the YouTube content creation ecosystem, where various channels serve distinct functions, from content dissemination to community building.

5.4 Eigenvector Centrality

Eigenvector centrality is a measure that extends beyond a node's immediate connections to consider the importance of its neighbors within the network. Unlike degree centrality, which only counts direct ties, eigenvector centrality reflects the notion that connections to high-scoring nodes contribute more to a node's score than equal connections to low-scoring nodes. Essentially, it measures a node's influence based on the principle that connections to influential nodes contribute more to the node's score.



The eigenvector centrality distribution graph from Gephi, calculated with many iterations (10.000), suggests a few nodes have significantly higher eigenvector centrality scores. This indicates that these nodes are not only well-connected but also connected to other well-connected nodes, enhancing their influence within the network. The majority of nodes, however, appear to have low eigenvector centrality, falling close to zero on the graph. These nodes may still be connected within the network, but either they or their neighbors are not central within the overall network structure.

Using networkX we get the following results about the eigenvector distribution:

```
# Calculating eigenvector centrality for the filtered graph
eigenvector_centrality = nx.eigenvector_centrality(G, max_iter=10000)

# Extracting channel labels along with their eigenvector centrality
channel_labels_with_eigenvector = [(G.nodes[node]['label'], eigenvector_centrality[node])
                                   for node in G.nodes()]

# Sorting to find the top eigenvector centrality values
sorted_channel_labels_with_eigenvector = sorted(channel_labels_with_eigenvector,
                                                key=lambda x: x[1], reverse=True)

# Displaying the top few entries to check the data
sorted_channel_labels_with_eigenvector
```

```
[('Vsauce2', 0.3142064765778091),
 ('Vsauce3', 0.28083863317329216),
 ('minutephysics', 0.2754207655568413),
 ('Numberphile', 0.26017056375687775),
 ('Jeff Nippard', 0.1881222956583481),
 ('Be Smart', 0.18771400408837208),
 ('Frank Medrano', 0.17676571737038302),
 ('CT Fletcher Motivation', 0.1640065994682905),
 ('FitnessFAQs', 0.16187136578015626),
 ('Strength Project', 0.15823481600919515),
 ('OfficialBarstarzz', 0.15469386289867096),
 ('PBS Space Time', 0.15171361842182393),
 ('PBS Idea Channel', 0.14570685627850752),
 ('The Kneesovertoesguy', 0.14086437926680465),
 ('Calisthenicmovement', 0.12796627195401178),
 ('Dictionary of Obscure Sorrows', 0.11870264650586548),
 ('Hodge Twins', 0.11852114347134651),
```

Vsauce2 leads the eigenvector centrality with a score of 0.3142, implying not only that it is well-connected but also that its connections are themselves influential within the network. This resonates with the already known idea that the Vsauce channel is a hub within a hub, an influential node that is part of a broader influential community.

Vsauce3 and minutephysics also display high eigenvector centrality, scoring 0.2808 and 0.2754, respectively. These channels, like Vsauce2, more particularly the minutephysic channel, are central in a community of channels that are themselves quite central, creating a core group of interconnected influencers.

Notably, while Jeff Nippard and Be Smart have lower eigenvector centralities than the Vsauce channels and minutephysics, their scores of 0.188 and 0.1877 still denote a significant level of influence within the network. They may not be at the core

of the most central community but are influential within their sub-communities or clusters.

Noteworthy to mention is that Vsauce, Numberphile and Jeff Nippard strike a balance between reach, influence, and strategic positioning. Their prominence in closeness and eigenvector centrality, despite medium degree and betweenness centrality measures, suggests they have optimized their network position not just for maximal direct influence, but for quality and strategic connectivity that enhances their overall standing in the network. They are important players, but not the sole or primary conduits of information and influence, which provides a more distributed structure to the network's connectivity and influence landscape.

The Bioneer channel, despite its vital role as indicated by the first two centrality measures, has a lower eigenvector centrality score of 0.03536. This suggests that while the Bioneer is highly connected, its direct connections may not be to other highly central nodes within the network, which influences its eigenvector score.

In summary, the eigenvector centrality scores reflect not just the presence of connections but the quality of those connections in terms of network influence. Channels with high eigenvector centrality are crucial in shaping the network's information flow, and their influence is amplified by the centrality of the channels they are connected to. These scores highlight a tiered structure within the network, with a core of highly influential channels surrounded by nodes with varying levels of influence.

6. Clustering Effects

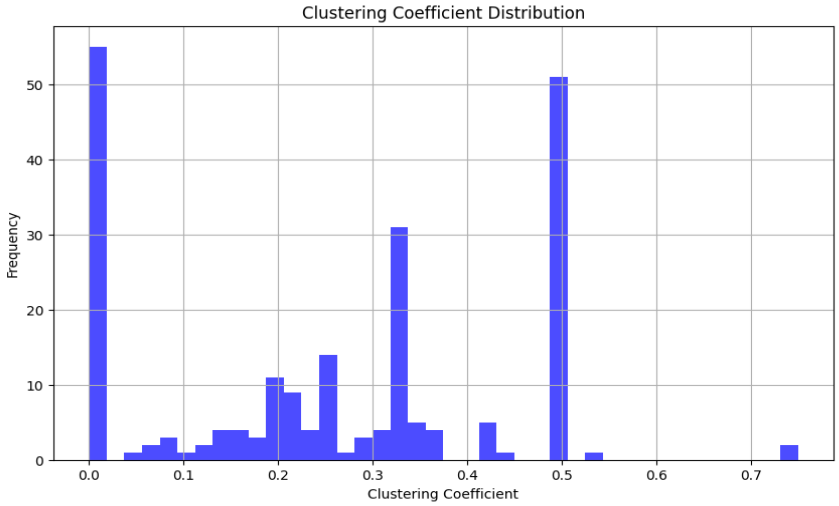
6.1 Average Clustering Coefficient

The clustering coefficient for a node is calculated based on directed triangles that the node participates in. These are connections where two other nodes that a given node points to also have a directed connection to each other. The average clustering coefficient according to the Watts–Strogatz model is the mean value of the local clustering coefficients of all the nodes in the network and provides insight into the local interconnectedness of nodes, revealing potential substructures and indicating the degree to which the network fosters reciprocal relationships among its constituents.

For the Bioneer YouTube channel network, an average clustering coefficient of 0.256 calculated from Gephi and from networkX (0.258), suggests a moderate level of clustering. This value indicates that while there is some tendency for channels to form tightly knit groups, where one channel subscribed to two others could mean those two are likely subscribed to each other, it is not an overwhelmingly prevalent feature of the network. It is indicative of a network that balances between a web of interconnected communities and more radial structures where a central node connects otherwise unconnected nodes. It also means that some subgroups within the network are interacting more closely, which can lead to more efficient communication and collaboration within those groups. This can be beneficial for targeted content strategies or advertising and for content creators and network strategists, as a clustering coefficient of this level suggests that there is room to build stronger community ties.

6.2 Clustering Coefficient Distribution

The clustering coefficient distribution of the Bioneer YouTube channel network, depicted in the provided histogram generated using matplotlib, offers insight into the local connectivity of the network. The use of python in this step was necessary, because Gephi does not provide a proper function to display this distribution graphically. The distribution of clustering coefficients among the nodes presents a multi-modal distribution, with spikes at 0 and a notable peak around 0.5, and then less frequent occurrences of higher clustering coefficients trailing off towards 0.7.



A large number of nodes with a clustering coefficient of zero indicates that there are numerous channels which do not form triangles. This means that while a channel might be subscribed to two other channels, those two are not subscribed to each other, resulting in a lack of closed triads.

The peak around the 0.5 mark suggests that there is also a considerable number of nodes that are part of highly interlinked communities. In these communities, if a channel is subscribed to two others, there is about a fifty percent chance that those two are also subscribed to each other, indicating a strong tendency towards group formation.

The presence of nodes with higher clustering coefficients, although less frequent, indicates the existence of clusters or cliques within the network, where most of the nodes are interconnected.

It would be useful to also examine more carefully the clustering coefficients of the channels:

```
clustering_coefficients = nx.clustering(G)
# Extracting channel labels along with their clustering coefficient
channel_labels_with_clustering = [(G.nodes[node]['label'], clustering_coefficients[node])
                                  for node in G.nodes()]
# Displaying the data
channel_labels_with_clustering.sort(key=lambda x: x[1], reverse=True)
channel_labels_with_clustering
```

```
[('SoundGuys', 0.75),
 ('Gary Explains', 0.75),
 ('Alex Leonidas', 0.5263157894736842),
 ('How to ADHD', 0.5),
 ('Clark Elieson', 0.5),
 ('Channel Pup', 0.5),
 ('Variant Comics', 0.5),
 ('samy kamkar', 0.5),
 ('Inside Gaming', 0.5),
 ('Testosterone Nation', 0.5),
 ('VitruvianPhysique', 0.5),
 ('Siim Land', 0.5),
 ('TechCrunch', 0.5),
 ('Chris Stuckmann', 0.5),
 ('NewRetroWave', 0.5),
 ('Dom Esposito', 0.5),
 ('The Enigma TNG', 0.5),
```

Notably, several channels exhibit a higher clustering coefficient, with SoundGuys and Gary Explains at the peak of 0.75. This suggests that their neighbors are highly interconnected, forming strong communities or subgroups within the larger network. Such high clustering coefficients are indicative of cohesive subgroups where the connected channels share similar content or audiences.

Channels like Alex Leonidas and How to ADHD have clustering coefficients above 0.5, again pointing towards significant local clustering. This could mean that subscribers of these channels are likely to subscribe to each other, strengthening community bonds and leading to more isolated communication patterns within these clusters.

The presence of channels with lower clustering coefficients, such as the Bioneer at 0.0081 and those with a coefficient of zero, highlights different things. Firstly, it highlights channels that serve more as bridges between different communities within the network rather than being embedded in a close-knit cluster, (which is the case for the Bioneer). Secondly, it shows the channels which operate in a more unique or specialized niche, different from its immediate peers. Lastly, it shows channels with a diverse range of connections across various niches or communities, prime example of which is the Bioneer.

6.3 Triangles

The concept of triangles plays a crucial role in understanding the interconnectedness and community structure within the network. A triangle in a network is formed when three nodes (in this case, channels) are mutually connected, indicating a higher level of interaction and potential content collaboration or thematic similarity.

Considering the Bioneer network is a directed graph, the 'simple_cycles' function from NetworkX was employed to accurately identify triangles, specifically by setting the cycle length bound to three. This approach is particularly suitable for directed graphs like the Bioneer network, as it effectively captures the directional nature of relationships between nodes (channels).

Using the 'simple_cycles' function, we found a total of eight triangles, which is a more accurate representation compared to the 491 triangles suggested by the standard triangles function in NetworkX, which only works if we convert the graph to be undirected. The latter is not as effective for directed networks, where the direction of connections significantly impacts the structure and interpretation of triangles.

The calculation of the triangles can be seen here:

```
import pandas as pd
# Finding triangles and sorting nodes within each cycle
triangles = list(nx.simple_cycles(G.to_directed(), length_bound=3))
triangles = [sorted(cycle, key=lambda x: G.nodes[x]['label']) for cycle in triangles if len(cycle) == 3]

# Remove duplicate triangles
triangles = list(set(tuple(cycle) for cycle in triangles))

# Create a dataframe to display the triangles with node labels
triangles_df = pd.DataFrame([tuple(G.nodes[node]['label'] for node in triangle) for triangle in triangles], columns=['Node 1', 'Node 2', 'Node 3'])
triangles_df
```

	Node 1	Node 2	Node 3
0	Bodybuilding.com	OfficialBarstarzz	THENX
1	BBC Earth Science	Be Smart	PBS Space Time
2	Android Authority	Gary Explains	SoundGuys
3	Alex Leonidas	Fight Commentary Breakdowns	The Bioneer
4	Be Smart	PBS Idea Channel	PBS Space Time
5	Fight Commentary Breakdowns	The Bioneer	hard2hurt
6	FitnessFAQs	SaturnoMovement	Tom Merrick
7	Alex Leonidas	Geoffrey Verity Schofield	The Bioneer

The identified triangles and their implications are as follows:

Triangle 1 - Bodybuilding.com, OfficialBarstarzz, THENX:

This triangle represents a strong interconnectedness within the fitness community, specifically focusing on bodybuilding and calisthenics. Bodybuilding.com is known for its comprehensive content on bodybuilding, OfficialBarstarzz is renowned for street workout and calisthenics, and THENX specializes in calisthenics and bodyweight training. Their mutual connection suggests a shared audience interested in diverse yet related aspects of physical fitness.

Triangle 2 - BBC Earth Science, Be Smart, PBS Space Time:

This grouping brings together channels focused on science and education. BBC Earth Science provides content on natural science and earth-related topics, Be Smart explores various scientific concepts in an accessible manner, and PBS Space Time delves into astrophysics and space science. The mutual connections among these channels indicate a shared interest in promoting scientific literacy and exploration of the universe.

Triangle 3 - Android Authority, Gary Explains, SoundGuys:

In this triangle, we see a combination of technology-focused channels. Android Authority covers news and reviews related to Android devices and software, Gary Explains offers in-depth technical explanations on a variety of topics, including technology, and SoundGuys deals with audio equipment and technology. This triangle could reflect a community interested in detailed technological insights, especially in mobile technology and audio gear.

Triangle 4 - Alex Leonidas, Fight Commentary Breakdowns, The Bioneer:

This triangle represents a focus on martial arts, fitness, and self-improvement. Alex Leonidas' channel offers content on physical training, Fight Commentary Breakdowns provides analyses of various fighting styles and matches, and The Bioneer integrates fitness with a variety of disciplines for overall self-improvement including martial arts at moments. The mutual connections suggest a shared audience interested in martial arts, physical fitness, and personal development.

Triangle 5 - Be Smart, PBS Idea Channel, PBS Space Time:

Here, we have another set of educational channels. As we have mentioned before, Be Smart offers accessible science content, PBS Idea Channel discusses the connections between pop culture and art to larger philosophical concepts, and PBS Space Time, as mentioned before, focuses on astrophysics and space. This triangle reflects a shared viewership interested in science, philosophy, and the exploration of complex ideas.

Triangle 6 - Fight Commentary Breakdowns, The Bioneer, hard2hurt:

This triangle seems to focus on martial arts, self-defense, and fitness. Fight Commentary Breakdowns and the Bioneer, as previously described, are complemented here by hard2hurt, a channel that focuses on self-defense and fitness. This grouping suggests a community interested in practical self-defense techniques, martial arts analysis, and fitness-oriented training.

Triangle 7 - FitnessFAQs, SaturnoMovement, Tom Merrick:

This group is centered around bodyweight fitness and calisthenics. FitnessFAQs is known for its calisthenics and bodyweight training content, SaturnoMovement focuses on movement and flexibility, and Tom Merrick specializes in bodyweight fitness with a focus on mobility and flexibility. The mutual connections among these channels indicate a strong community interest in calisthenics, holistic body training, and functional fitness.

Triangle 8 - Alex Leonidas, Geoffrey Verity Schofield, The Bioneer:

This triangle, featuring Alex Leonidas and the Bioneer once more, along with Geoffrey Verity Schofield, suggests a focus on fitness and bodybuilding, self-improvement, and at the same time lifestyle optimization, which stems from the fact that all three channels are discussing often about nutrition and healthy habits to live by. Geoffrey Verity Schofield's content includes mostly bodybuilding advice along with nutrition advice, complementing the other two channels. This triangle indicates a shared audience interested in personal development, fitness, and lifestyle optimization strategies.

Each of these triangles illustrates the interconnected nature of content themes and audience interests within the YouTube community, particularly in the realms of fitness, science, technology and self-improvement. These connections facilitate content collaboration, cross-promotion, and a shared viewer base that appreciates diverse yet related content.

Interesting to see is that the Bioneer is part of several triangles: with Alex Leonidas and Fight Commentary Breakdowns; Fight Commentary Breakdowns and hard2hurt; and Alex Leonidas and Geoffrey Verity Schofield. These triangles indicate that the Bioneer is at a pivotal intersection of channels focused on fitness, martial arts, and self-improvement. The mutual connections suggest a shared thematic interest or audience overlap. It points to the Bioneer's role in forming a cohesive community within these niches.

6.4 Triadic Closure Phenomenon

The concept of triadic closure is a key element in understanding social networks and their dynamics. It refers to the tendency for people who share connections in a social network to become connected themselves. In the context of the Bioneer YouTube

channel network, the presence of the eight identified triangles is a clear manifestation of this phenomenon.

As we have seen from the previous calculations, the presence of triangles and the clustering coefficients indicate that triadic closure does occur in our network, but not in a significant degree. To see if the forming of triangles is local phenomena or not, the transitivity of the network needs to also be taken into account, which in our case is 0.01069 (calculated from networkX). Transitivity, in network analysis, is the ratio of all triangles to all possible triangles in a network. It is a global measure of the degree to which nodes in a graph tend to cluster together.

Given the relatively low transitivity score in the Bioneer network, it suggests that while there are instances of triadic closure (as evidenced by the eight triangles), the network as a whole is not highly clustered. This could mean that while there are specific pockets or groups within the network where channels are highly interconnected (like those in the identified triangles), these are not widespread phenomena across the entire network.

Interesting to note is that the involvement of the Bioneer in specific triangles, despite its low overall clustering coefficient, highlights the channel's potential role in facilitating triadic closures within its immediate network. These closures are essential for creating new connections and strengthening existing relationships within the network. They can lead to collaborative opportunities, content cross-promotion, and shared audience growth among the channels involved in the triangles.

Additionally, the low clustering coefficient might also imply that the Bioneer's content is diverse enough that it does not lead to a high level of interconnectedness with other channels and serves as a bridge (as we will confirm in the next chapter) between different communities and thematic areas within the broader network. This can be a strength, as it allows the channel to link various otherwise disconnected parts of the network. For instance, the Bioneer could be introducing topics or perspectives that are not widely covered by other channels in the network, thereby attracting a diverse range of viewers and collaborators.

7. Bridges and Local Bridges

7.1 Bridges

Bridges in a network are edges that, if removed, would divide the network into separate components. In the context of our network, the presence of bridges after converting the graph to undirected indicates points of vulnerability and significant nodes for connectivity.

```
# Function to get the label of a node
def get_label(G, node_id):
    return G.nodes[node_id]['label']

# Finding bridges and replacing node IDs with labels
bridges = [(get_label(G, u), get_label(G, v)) for u, v in nx.bridges(G.to_undirected())]
```

```
bridges
```

```
[('The Bioneer', 'Aim To Head Official'),
 ('The Bioneer', 'Quinn s Ideas'),
 ('The Bioneer', 'Tee Lopes'),
 ('The Bioneer', 'MuscleTV'),
 ('The Bioneer', 'The  80s Guy'),
 ('The Bioneer', 'SwitchForce'),
 ('The Bioneer', 'RGT 85'),
 ('The Bioneer', 'Original Strength'),
 ('The Bioneer', 'Sonic the Hedgehog'),
 ('The Bioneer', 'CGRundertow'),
 ('The Bioneer', 'Matty Fusaro'),
 ('The Bioneer', 'The Closer Look'),
 ('The Bioneer', 'Tao Physique'),
 ('The Bioneer', 'Pocket Gamer'),
 ('The Bioneer', 'khanacademymedicine'),
 ('The Bioneer', 'GameTrailers'),
 ('The Bioneer', 'Explore Gadgets'),
 ('The Bioneer', 'vsauceLEANBACK'),
 ('The Bioneer', 'ParryGripp'),
 ('The Bioneer', 'Yuriofwind'),
 ('The Bioneer', 'The Movement System'),
 ('The Bioneer', 'Mastar'),
 ('The Bioneer', 'Eric Jacobus'),
 ('The Bioneer', 'AppSumo'),
 ('The Bioneer', 'World Science Festival'),
 ('The Bioneer', 'JimTV'),
 ('The Bioneer', 'Dave Bennett'),
 ('The Bioneer', 'Krista King'),
 ('The Bioneer', 'MegaGWolf'),
 ('The Bioneer', 'SPARTAN BODYBUILDING'),
 ('The Bioneer', 'The Thought Emporium'),
 ('The Bioneer', 'Android Developers'),
 ('The Bioneer', 'SearingWolfe'),
 ('The Bioneer', 'Skill Up'),
 ('The Bioneer', 'Derren Brown'),
 ('The Bioneer', 'Gokuflex'),
```

The list of bridges, with length of 54, all provided between the Bioneer and other channels suggests that these channels do not have any alternate path connecting them to the rest of the network. This indicates that Bioneer plays a crucial role in keeping these channels connected to the larger network. If he were to be removed, these channels would become isolated components.

The extensive number of bridges associated with him also points to a network that has a star-like structure, where the channel acts as a central hub for many peripheral nodes. These nodes may have been connected to the network primarily through him and not through interactions with other channels.

This can have noteworthy implications:

- **Information Spread**: Bioneer is critical for the dissemination of information to these peripheral nodes.
- **Network Robustness**: The network's robustness is compromised since removing him would fragment the network.
- **Influence**: Bioneer has a strong influence over these nodes, as it acts as their main or sole connector to the broader network.

In social networks, bridges are often seen as opportunities. For the Bioneer, these bridges represent potential areas to solidify or increase influence. However, they also represent risks since the connectivity and flow between him and these channels are dependent on a single connection point, making the network's cohesion sensitive to changes in Bioneer's status within the network.

7.2 Local Bridges

Local bridges in a network are edges that connect two nodes that would have a long "*span*" or distance between them if that edge were not present. In other words, the span of a local bridge is the shortest path length between the endpoints if the local bridge is removed. Let us examine the list of local bridges provided in a similar way as the bridges from networkX:

```
for local1,local2,span in nx.local_bridges(G.to_undirected()):
    print(f"({get_label(G,local1)}, {get_label(G,local2)}) , span={span}")
```

```
(The Bioneer, Aim To Head Official) , span=inf
(The Bioneer, Quinn s Ideas) , span=inf
(The Bioneer, Tee Lopes) , span=inf
(The Bioneer, MuscleTV) , span=inf
(The Bioneer, The  80s Guy) , span=inf
(The Bioneer, SwitchForce) , span=inf
(The Bioneer, RGT 85) , span=inf
(The Bioneer, Original Strength) , span=inf
(The Bioneer, Sonic the Hedgehog) , span=inf
(The Bioneer, CGRundertow) , span=inf
(The Bioneer, Matty Fusaro) , span=inf
(The Bioneer, The Closer Look) , span=inf
(The Bioneer, Tao Physique) , span=inf
(The Bioneer, Pocket Gamer) , span=inf
(The Bioneer, khanacademymedicine) , span=inf
(The Bioneer, GameTrailers) , span=inf
(The Bioneer, Explore Gadgets) , span=inf
(The Bioneer, vsauceLEANBACK) , span=inf
(The Bioneer, ParryGripp) , span=inf
(The Bioneer, Yuriofwind) , span=inf
(The Bioneer, The Movement System) , span=inf
(The Bioneer, Mastar) , span=inf
(The Bioneer, Eric Jacobus) , span=inf
(The Bioneer, AppSumo) , span=inf
(The Bioneer, World Science Festival) , span=inf
(The Bioneer, JimTV) , span=inf
(The Bioneer, Dave Bennett) , span=inf
(The Bioneer, Krista King) , span=inf
(The Bioneer, MegaGWolf) , span=inf
(The Bioneer, SPARTAN BODYBUILDING) , span=inf
(The Bioneer, The Thought Emporium) , span=inf
(The Bioneer, Android Developers) , span=inf
(The Bioneer, SearingWolfe) , span=inf
(The Bioneer, Skill Up) , span=inf
(The Bioneer, Derren Brown) , span=inf
(The Bioneer, Gokuflex) , span=inf
(The Bioneer, Hybrid Calisthenics) , span=inf
(The Bioneer, Implicitly Pretentious) , span=inf
(The Bioneer, Sonic Central) , span=inf
(The Bioneer, Physionic) , span=inf
(The Bioneer, Max Derrat) , span=inf
(The Bioneer, Sorelle Amore) , span=inf
```

The span being infinite (**`span=inf`**) on all local bridges suggests that there is no alternative path between the two nodes that the bridge connects, which means these are unique connections in the network.

The list of local bridges with the Bioneer channel, along with the fact that the span for those bridges is always infinite, implies that it forms exclusive connections with a variety of channels that do not have other direct or indirect paths to it. These unique bridges underscore its role as a gatekeeper of sorts, providing a vital link to these channels into the broader network.

Very interesting to mention is the fact that the local bridges are identical to the bridges (both having the same length and same contents), when the network is considered undirected, reinforces the point that these connections are not just rare but singular in nature. For the channels connected to Bioneer by these bridges, the Bioneer represents an irreplaceable conduit to the rest of the network.

The implications for network strategy are significant:

- **Dependency**: These channels depend on it for their network connectivity.
- **Influence Potential**: Bioneer has the potential to significantly influence these channels due to the exclusive nature of the connections.
- **Risk**: The network faces a risk of fragmentation if the channel ceases to function as a connector.

The variety of channels listed as having local bridges with the Bioneer, spanning various content areas and themes, indicates the Bioneer's diverse range of interactions and connections and suggests that the Bioneer is engaging with a wide array of topics or niches, serving as a bridge between various content areas that do not typically intersect. This diversity enhances the channel's ability to reach different audience segments and to cross-pollinate ideas across different content domains.

For the Bioneer, these local bridges represent opportunities for maintaining and strengthening its position within the network by being the sole connector for these channels. However, from a network resilience perspective, it would be beneficial to develop additional connections among these channels to reduce dependency on a single node and thus increase the robustness of the network against disconnection or loss of the central node.

8. Gender

Gender analysis in network contexts, particularly on platforms like YouTube, can be challenging due to certain factors, notably due to:

- **Anonymity and Pseudonymity**: A lot of YouTube channels operate under pseudonyms or brand names without a clear gender identifier.
- **Multiple Contributors**: A single channel might feature contributions from multiple individuals of different genders.
- **Content Variety**: The content itself may not always give clear indicators of the gender identity of the channel operators.
- **Privacy Concerns**: Directly identifying and labeling the gender of channel operators may raise privacy concerns, and it is not always possible to accurately infer gender without explicit confirmation from the individuals involved.

Given these challenges, gender analysis for YouTube channels often requires manual review and confirmation, which can be a labor-intensive process. Additionally, the task must be approached with sensitivity to avoid misgendering or making assumptions based on stereotypes.

In the network under consideration, empirical observations suggest a predominance of male-identified individuals or groups. This trend aligns with broader industry patterns where male dominance in certain digital content creation sectors, particularly within specific niches like technology, gaming, fitness, and science, is well-documented. This male prevalence in the network reflects societal trends and the historic underrepresentation of women in certain fields, which also translates into the digital sphere.

The implications of such a gender imbalance are multifaceted. It influences the types of content that are produced and shared, the nature of collaboration and interaction within the network, and the perspectives that are amplified within this digital ecosystem. The predominance of male-led channels also shapes the network's dynamics in terms of audience engagement, with potential impacts on the visibility and accessibility of diverse voices.

Moreover, while male dominance is noted, it is crucial to acknowledge that it does not account for the entire picture. There are indeed female content creators who contribute significantly to the richness and diversity of our network and of the YouTube fitness, tech, and science channel networks in general. Although their effect on the network can be subtle, their presence and impact serve as a counterbalance, offering alternative viewpoints and fostering a more inclusive environment.

9. Homophily

9.1 Homophily based on Country

Homophily is a principle in network theory that describes the tendency of individuals to connect with others who are similar to themselves. This similarity can be based on various attributes such as interests, social status, demographics, or, in the case of our YouTube channel network, the country.

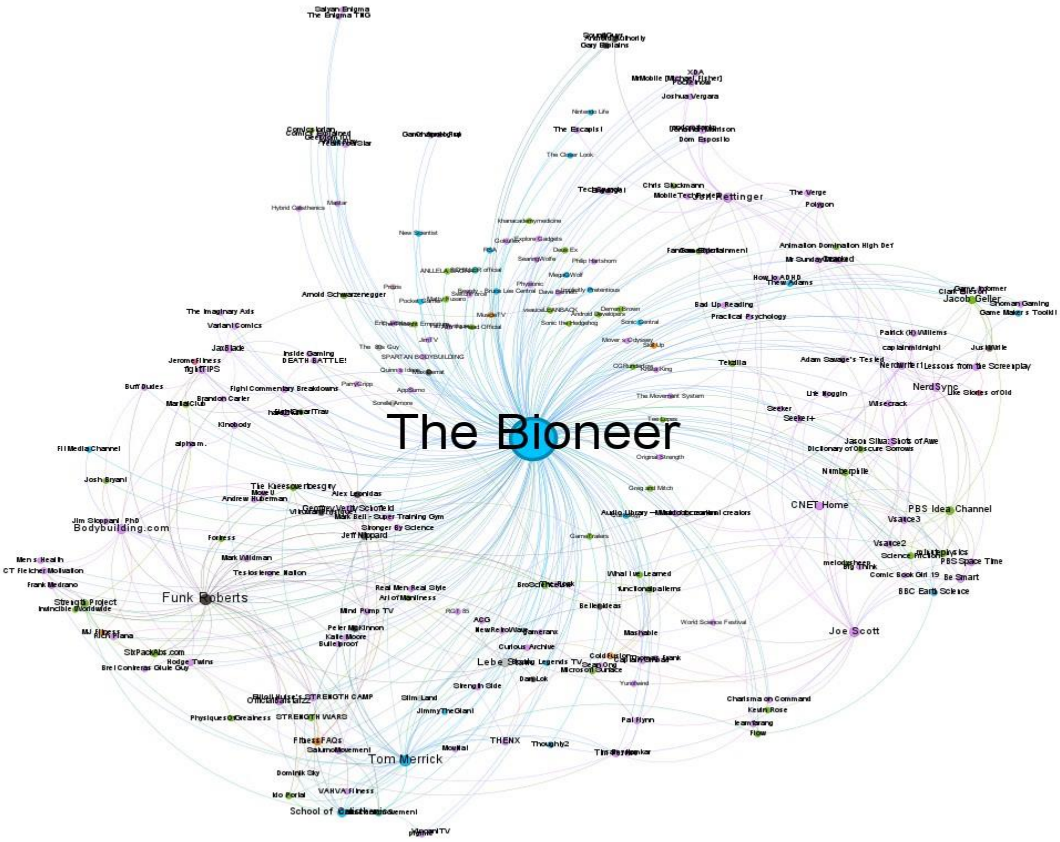
Mathematically, homophily in networks can be quantified using measures such as the assortativity coefficient. Assortativity is a correlation coefficient that represents the likelihood of nodes in the network connecting with others that share similar attributes. It ranges from -1 to +1.

```
# Calculate the assortativity coefficient for country
country_assortativity_coefficient = nx.attribute_assortativity_coefficient(G, 'country')
country_assortativity_coefficient
```

```
0.039256571978993265
```

In our network, the assortativity coefficient for the country is around 0.04, indicating that channels have a slight preference for connecting with others from the same country, but the effect is not strong. This level of homophily can be influenced by various factors, such as the language of the content, cultural specifics that resonate more strongly with a domestic audience, or the geographic focus of the channel's content. Channels naturally gravitate towards others in the same country due to shared cultural contexts, ease of collaboration, or similar time zones, which can facilitate interaction. However, the low magnitude of the coefficient suggests that while there is a hint of homophily based on country, the network is relatively open to cross-country connections, reflecting a more globally integrated community of YouTube channels.

We can get a clear view of the interaction of the channels based on the country from the graph below, where we can clearly see that the US (marked as purple), is the country where most of the channel have base at a percentage of 57.47%, Great Britain (marked as light blue) with a percentage of 9.5%, Canada (marked as black) with a percentage of 5.43% and Australia (marked as orange) with a percentage of 2.26%. Important to note is that a considerable proportion of the channels (20.36%) have not stated the country that they are based in, and these channels are colored as light green which also means that the assortativity that was calculated might not be entirely accurate about the state of the network in terms of country homophily.



Noteworthy to mention is the fact that we can also measure homophily based on the country using the number of cross-edges. Firstly, let us take the US channels (57.47%), which is the region that has the most youtubers. All the channels from other countries (including 'not-set') occupy 44.53% of the network. In that case, a channel has a probability of $p = 57.47\%$ of being in the US and $q = 1 - p = 44.53\%$ of being elsewhere.

In the scenario that there was no homophily in our network, the cross-country edges should be equal to $2pq = 0.511$ or 51.11% of the total edges (604), which makes up to 310 edges. Using networkX, however, to calculate the intra edges, we get a result of 173 edges. This result is less than we initially expected, which means that there is indeed evidence for the existence of homophily for the Bioneer network and that channels from the US tend to form more connections with channels that are also from the US rather than with other countries.

9.2 Homophily based on Subscriber Count

The assortativity coefficient for ‘subscriber count’ in the network, with a value of around -0.006 suggests a very mild tendency towards disassortative mixing based on subscriber count. In practical terms, this means there is a very small preference for channels to connect with others that have different subscriber counts, but the effect is minimal. However, the assortativity coefficient does not work well on continuous numerical data and surely does not provide a clear view for the state of our network.

In analyzing the structural dynamics of a YouTube channel network, particularly in terms of how channels of different sizes interconnect, the E-I (External-Internal) Index becomes a valuable tool (Krackhardt_E/I_Ratio).

The E-I Index, or External-Internal Index, is a measure used in network analysis to quantify the extent of a group's (or node's) external connections compared to its internal connections. It helps in understanding whether the interactions within a network are more homophilous (inward-focused) or heterophilous (outward-focused).

The E-I (External-Internal) Index calculated for the network, based on the subscriber count attribute with a threshold of three million, is approximately -0.39. A negative E-I Index in this context suggests that there are more connections (edges) within groups of channels that either both exceed, or both fall below the 3.000.000 subscriber threshold, compared to connections between one channel above the threshold and another below it. In simpler terms, channels tend to connect with others that are similar in terms of subscriber count. We can see the function that was used for the calculation below:

```
def calculate_e_i_index_by_subscriber_count(G, threshold):
    internal_edges = 0
    external_edges = 0

    for u, v in G.edges():
        # Check if 'subscribercount' attribute exists for both nodes
        if 'subscribercount' in G.nodes[u] and 'subscribercount' in G.nodes[v]:
            # Determine if both nodes are above or below the threshold
            u_above_threshold = G.nodes[u]['subscribercount'] >= threshold
            v_above_threshold = G.nodes[v]['subscribercount'] >= threshold

            if u_above_threshold == v_above_threshold:
                internal_edges += 1  # Both nodes are either above or below the threshold
            else:
                external_edges += 1  # One node is above the threshold and the other is below

    total_edges = internal_edges + external_edges
    if total_edges > 0:
        e_i_index = (external_edges - internal_edges) / total_edges
    else:
        e_i_index = None  # E-I Index is not defined for graphs without edges

    return e_i_index


# Define a subscriber count threshold
subscriber_threshold = 3000000  # example threshold

# Calculate the E-I Index
e_i_index = calculate_e_i_index_by_subscriber_count(G, subscriber_threshold)
print("E-I Index:", e_i_index)
```

```
E-I Index: -0.38741721854304634
```

The calculation method for this E-I Index involves assessing each edge (connection) in the network:

1. The function first checks if both nodes (channels) in an edge have the subscribercount attribute.
2. It then determines whether each of these nodes has a subscriber count above or below the specified threshold (**three million** in our case which was selected judging from the subscribercount distribution in the network).
3. If both nodes on either side of an edge are either above or below this threshold, it is counted as an internal edge. If one node is above and the other is below the threshold, it is counted as an external edge.
4. The E-I Index is then calculated by taking the difference between the number of external and internal edges and dividing by the total number of edges.

The negative value of the E-I Index indicates a tendency towards homophily based on subscriber count. Channels within similar subscriber count ranges (either both above or both below 3 million) are more likely to form connections. This reflects similarities in content type, production values, or audience engagement strategies that are typical of channels within certain size ranges. It also suggests that channels of similar sizes face comparable challenges and opportunities, leading them to collaborate or interact more frequently.

9.3 Homophily based on Video Count and View Count

The assortativity coefficients based on ‘video count’ and ‘view count’ in our network, with values of -0.005 and -0.0047 respectively, could provide insights into the homophily patterns related to these attributes. The near-zero values of these coefficients imply that there is no strong preference or trend for channels to connect with others that have a similar number of videos (video count) or a similar level of popularity (view count). Channels with many videos or high view counts are just as likely to connect with channels with fewer videos or views, and vice versa. However, in a real youtube network that is rarely the case.

Let us look at the E-I Indexes of these attributes, to get a better understanding of the state of the network. The E-I Index values for video count and view count, calculated with the same way as the subscriber’s count E-I Index, in the network, with thresholds of one hundred videos and ten million views respectively, provide significant insights into the network's connectivity patterns with respect to these attributes.

1. **E-I Index for Video Count (-0.788)**:
    - With a threshold of one hundred videos, an E-I Index of approximately -0.788 indicates a strong internal connectivity preference based on video count. This value suggests that channels with fewer than one hundred videos tend to connect more with each other, as do channels with more than one hundred videos, but there is significantly less connectivity between these two groups. This high negative value reflects a substantial level of segregation in the network based on the number of videos a channel has. It implies that the network's content creation frequency (as

measured by video count) plays a crucial role in how channels form connections.

2. **E-I Index for View Count (-0.868)**:
    - The E-I Index for view count, with a threshold of ten million views, is even more negative, at approximately -0.868. This suggests an even stronger internal connectivity preference based on the number of views. Channels with view counts below the threshold are much more likely to connect with each other, and the same is true for channels above the threshold. The significant segregation based on view count implies that audience size or popularity heavily influences network connections. Channels with similar levels of viewership tend to form a more cohesive subgroup within the network, potentially reflecting similarities in audience reach, content impact, or stage of channel development.

9.4 Homophily based on degree assortativities

The degree assortativity coefficients for different types of connections in our network reveal nuanced aspects of its structure. Let us explore what each of these metrics suggests about the network:

1. **Degree Assortativity Coefficient (In-In): 0.03**
    - This positive coefficient indicates a slight preference for channels to connect with other channels that have a similar number of incoming connections. Although the effect is minimal, it suggests a tendency towards homophily in terms of how channels are referenced or linked to by others. This reflects a pattern where channels of similar popularity or content types tend to connect with or reference each other.
2. **Degree Assortativity Coefficient (Out-Out): -0.056**
    - The negative coefficient here, though relatively small, suggests a slight tendency for channels to connect outwards to channels with dissimilar numbers of outgoing connections. This indicates that more active channels (in terms of referencing or connecting to others) are slightly more likely to link to less active channels, rather than to equally active ones.
3. **Degree Assortativity Coefficient (In-Out): -0.0258**
    - This negative value suggests a mild propensity for channels with more incoming connections to connect to channels with fewer outgoing connections, and vice versa. It reflects a slight level of cross-linkage between different types of channels, indicating that popular channels are referenced by a wider variety of channels, not just those that are similarly popular.
4. **Degree Assortativity Coefficient (Out-In): -0.446**
    - The significantly negative coefficient in this category is quite notable. It implies a strong tendency for channels that connect outwards a lot (those that reference many others) to connect to channels that have fewer incoming connections. This indicates a network dynamic where certain channels act as information or content disseminators, linking to a broad array of less referenced channels.

10. Graph Density

The graph density of our directed network, calculated as 0.012, from both Gephi and networkX, offers valuable insight into the overall connectivity of the network. Graph density is a measure that quantifies how close the network is to being fully connected. It is determined by the ratio of the actual number of edges in the network to the maximum possible number of edges.

In a directed network like the one in our case, this maximum possible number of edges is determined by the formula $\frac{|E|}{|V|*(|V|-1)}$, where E is the number of edges and V is the number of vertices in the graph (source: https://en.wikipedia.org/wiki/Dense_graph). Given the network has a density of 0.012, this suggests that it is relatively sparse. In practical terms, this means that out of all the possible directed connections that could exist in the network, only about 1.2% are actually present.

A low density like this is common in large networks, especially in contexts like social media platforms or YouTube channels, where not every channel is directly connected to every other channel. This sparsity reflects the selective nature of connections – channels choose which other channels to connect with, be it through collaborations, references, or shared audiences, rather than connecting indiscriminately.

The implications of a low-density network like ours are multifold:

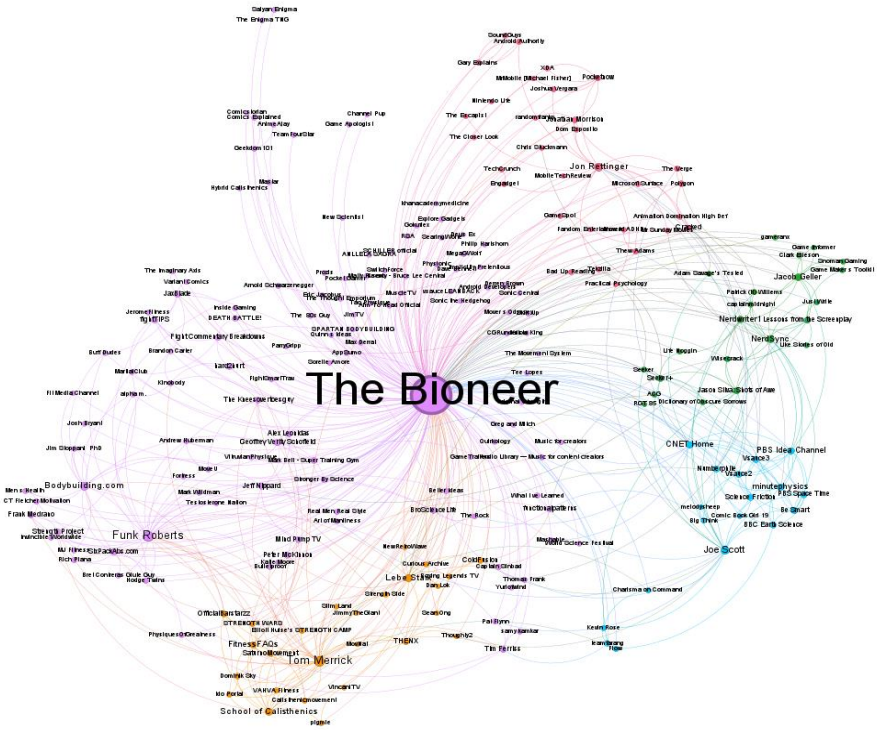
1. **Selective Interactions**: Channels form connections based on specific interests, content similarities, or mutual benefit, rather than connecting broadly and randomly.
2. **Room for Growth**: A sparse network indicates many opportunities for new connections and collaborations, suggesting potential for network growth and evolution.
3. **Navigability**: While lower density can mean less overall cohesion, it can also make the network easier to navigate, as there are fewer connections to consider when exploring the network.
4. **Community Structure**: Sparse networks like the Bioneer network have distinct community structures, with clusters of more densely connected nodes, (as we will find out in the next chapter). This can be beneficial for targeted marketing or community building strategies within the major hubs.

11. Community Structure

The community structure of the Bioneer network, as revealed by an analysis utilizing the modularity measure in Gephi, presents a fascinating insight into the network's composition and interaction patterns. With a modularity of 0.355 and an enhanced modularity with a resolution of 1.15 reaching 0.567, the analysis has identified five distinct communities within the network. The choice to proceed with the communities identified using the modularity with that particular resolution (1.15) is deliberate. In the context of our network, a higher resolution like 1.15 aligns better with the actual distribution of interactions and relationships within the network, compared to other resolutions that were, providing a more accurate and contextually relevant understanding of its community structure and, in our case, leads to the formation of communities that are more distinct and interpretable.

Modularity is a statistic that quantifies the degree to which the network is subdivided into such clearly defined communities. A higher modularity score, like the one observed with a resolution of 1.15, indicates a strong presence of communities or modules within the network. In the context of the Bioneer network, this suggests that there are significant groupings of channels that interact more densely within themselves than with channels outside their group.

Below is an image of the network, where the nodes are colored based on their modularity class:



In the network visualization provided, we can discern five distinct communities, differentiated by color, which correspond to various modularity classes. These communities represent groupings of YouTube channels that are more densely connected to each other than to nodes outside their community, suggesting shared attributes or types of content.

11.1 Bodybuilding Themed channels

The purple nodes represent the largest community (modularity class 0) comprising 39.82% of the network. This sizable proportion indicates a substantial cluster of interconnected channels, highlighting a core thematic or topical focus within the network. These nodes represent a community primarily focused on fitness, with channels that cater to a mainstream audience, as indicated by their larger subscriber counts. The presence of high-profile individuals like The Rock and Arnold Schwarzenegger (although in our network their effect is small as we have seen from previous chapters), both renowned for their impressive physiques—Arnold being a legendary figure in professional bodybuilding—underscores the community's lean towards fitness. Channels like BroScienceLife offer a humorous take on gym culture and bodybuilding, while platforms such as BodyBuilding.com and Men's Health provide a wealth of general fitness advice to a broad audience.

Within this same community, there are channels like GVS (Geoffrey Verity Schofield), Alex Leonidas, Vitruvian Physique, Buff Dudes, and Jeff Nippard, which, although having a smaller audience generally under one million subscribers, are lauded for delivering quality, unbiased fitness information. These channels have garnered respect for maintaining their integrity, often avoiding mainstream commercial sponsorships that could conflict with their message. In this spectrum lies also The Bioneer, known for integrating fitness with a scientific approach to body optimization.

Furthermore, there are channels with lower subscriber counts and general low influence, like Spartan Bodybuilding and Muscle TV, which delve deeper into the specific niche of bodybuilding. Despite the size of their following, these channels contribute to the community with content that is heavily influenced by the bodybuilding discipline.

In essence, while the scope and audience size vary among the channels within this purple community, they all share a common thread of focusing on **bodybuilding**, either as a core theme of their content or as a significant influence on the fitness information they provide. This balance between high-subscriber-count channels and those with specialized, authoritative content exemplifies the community's diversity and richness in the fitness and bodybuilding sphere.

11.2 Calisthenics Themed Channels

Following in size is the community represented by light orange nodes, belonging to modularity class 3, accounting for 25.79% of the network. This substantial community size suggests another major grouping with common characteristics or audience overlap.

The light orange nodes denote a community dedicated to calisthenics, a form of fitness that emphasizes bodyweight training and agility, transcending subscriber count as the common bond among them. Channels like FitnessFAQs, THENX, OfficialBarstarzz, and Calisthenicmovement are prominent within this group, focusing on calisthenics with an underlying goal of bodybuilding. They provide content that guides viewers through exercises and routines to build muscle and strength without the use of traditional gym equipment.

In this same community, there are channels such as Tom Merrick, Lebe Stark, and Saturno Movement, which not only cover calisthenics but also prioritize functional movement. This approach to fitness emphasizes the practical application of strength and flexibility in movements that replicate everyday activities or enhance athletic performance.

An interesting observation within the community's structure is the proximity of bodybuilding-oriented calisthenic channels to the purple nodes, indicating a thematic crossover with the more mainstream fitness community. In contrast, channels that focus purely on the art and technique of calisthenics, such as the 'School of Calisthenics,' are more centrally located within their own community, signaling a more specialized focus.

To summarize, while the community is diverse in terms of channel size and specific content focus, they all share a dedication to **calisthenics**. Whether their content is angled towards bodybuilding enhancement or the mastery of calisthenics itself, these channels collectively enrich the community with a multifaceted view of bodyweight fitness.

11.3 Science Themed Channels

The light blue nodes, which correspond to modularity class 4, make up 11.25% of the network. As a smaller yet significant community, this class represents a set of channels with unique or specialized content compared to the larger classes.

They signify a community predominantly comprised of channels that specialize in the realm of science and education. This community presents a cluster of channels that are not only unified by their content theme but also exhibit similar subscriber counts, suggesting a shared level of reach and influence within the YouTube ecosystem.

Notable among these channels are Joe Scott, known for explaining complex scientific concepts in plain language; Vsauce, which delves into scientific, psychological, mathematical, and philosophical topics; BBC Earth Science, which offers a window into the natural world and its scientific underpinnings; and Numberphile, dedicated to the wonders and puzzles of mathematics.

This community, while varied in the specific science topics covered, is bound by a common educational intent. The channels work to distill and disseminate scientific knowledge, making it accessible and engaging to a broad audience. From the exploration of abstract mathematical theories to the intricacies of Earth's ecosystems, these channels contribute to a collective understanding and appreciation of science in its many forms. To summarize, this community thrives on the dissemination of **science**,

with each channel contributing its unique flavor of specialized or general scientific content.

11.4 Digital Entertainment Based Channels

The dark green nodes, which correspond to modularity class 2 and make up to 10.92% of the network, represent a community dedicated to the digital entertainment sphere, particularly video games and movies. This group features channels with similar levels of reach and audience engagement, highlighting a shared interest among viewers.

In this cluster, gameranx and Game Informer stand out as channels that delve into gaming news, reviews, and features, while Seeker and Game Maker's Toolkit focus on the intricacies of game development and the scientific aspects of gaming. On the other hand, channels like NerdSync and Wisecrack explore the cultural and philosophical dimensions of movies and popular media, offering deeper analyses and discussions. Adam Savage's Tested adds a unique twist by merging movie insights with practical, hands-on experimentation and myth-busting.

It is also observed that channels with a primary focus on gaming content appear lower on the y-axis, indicating a possible sub-structuring within the community based on content focus. However, whether they lean towards gaming or cultural analysis, these channels collectively enrich the digital entertainment narrative by providing comprehensive coverage and thoughtful content that appeals to enthusiasts of virtual and cinematic worlds. In essence, this community encapsulates the diverse facets of **digital entertainment**, from the joy of gameplay to the artistry and critique of movies and popular culture.

11.5 Tech Based Channels

Lastly, the light red nodes belong in modularity class 1 and represent a smaller community, with 12.22% of the network.

The light red nodes signify a community centered around technology, encompassing a broad spectrum from general tech to specialized niches. Channels such as Jon Rettinger, Engadget, The Verge, and Microsoft Surface provide comprehensive coverage on computer technology, offering insights into the latest developments, reviews, and trends that appeal to a wide tech-savvy audience.

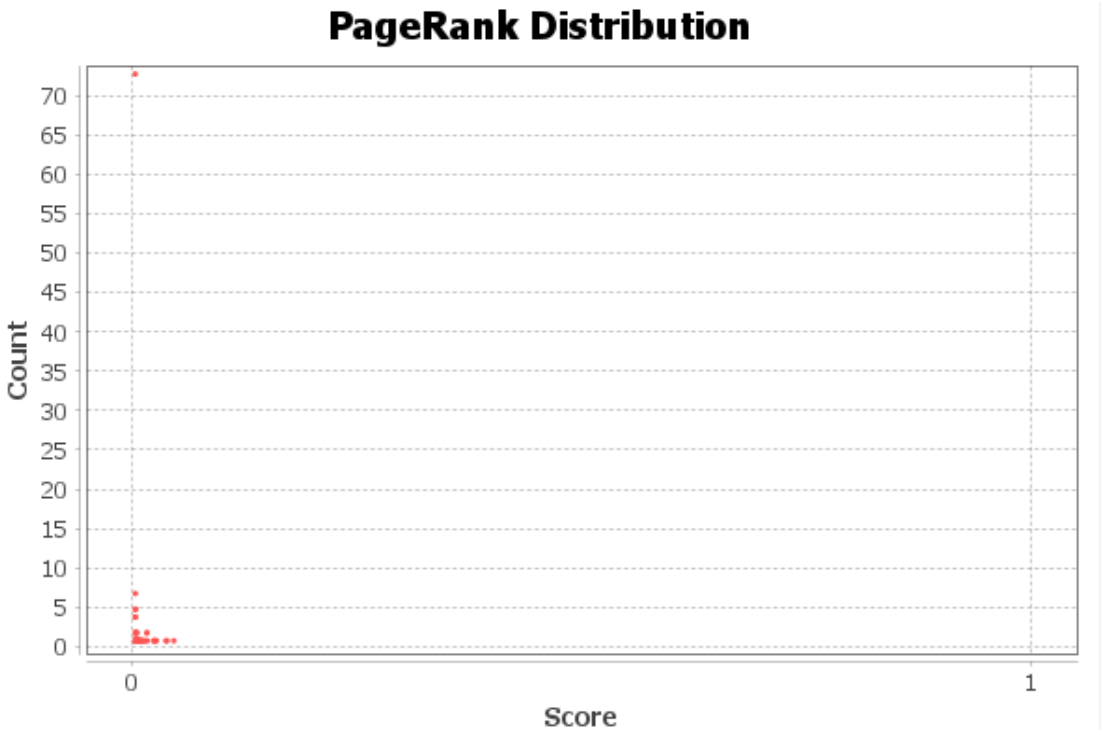
On the more specialized end of the spectrum, channels like Android Authority, MrMobile, and XDA are dedicated to mobile technology. They delve into smartphones, mobile applications, and the intricacies of mobile operating systems, catering to an audience deeply interested in handheld devices.

Interestingly, there is a discernible pattern in the community structure where channels with a narrower focus on mobile technology are positioned higher on the y-axis compared to those covering a broader range of tech topics. This suggests a layered community dynamic where content specialization influences the positioning within the network.

In summary, while the scope and content may vary, the common thread uniting these channels is their focus on informing and engaging with audiences on the latest in **technology**, from the newest smartphone releases to impactful tech innovations.

12. PageRank

The PageRank distribution for the Bioneer network, as seen in the provided graph, illustrates the relative importance or influence of each channel within the network. PageRank is a way of measuring the prominence of nodes in a network, based on the concept that connections to high-scoring nodes contribute more to the score of the node in question. (source: https://en.wikipedia.org/wiki/PageRank)



In this graph we can see that most nodes have a low PageRank score, indicating that a lot of channels have a similar level of influence within this network. However, there are a few nodes with higher scores, which stand out and can be considered as more influential or central within the network. These might be channels that function as hubs within the community, having many connections or particularly influential ones.

The calculation parameters used here, with an Epsilon of 0.001, indicate the precision of the algorithm — the calculation iterates until the change in PageRank values between iterations is less than 0.001. The Probability of 0.85, often referred to as the damping factor, in our case represents the likelihood at any step that a person randomly clicking on links (checking on connections between channels) will continue doing so instead of jumping to a random page (jumping in a random channel). A damping factor of 0.85 is quite standard and suggests that the PageRank values are leaning more towards the structure of the graph rather than random chance.

This distribution provides insights into how attention and authority are distributed across the network, with a few nodes being central influencers, while the vast majority have a more modest presence.

We can get a more accurate view of the distribution from the calculations provided using networkX:

```python
# Calculate PageRank
pagerank = nx.pagerank(G, max_iter=10000, alpha=0.85)
# Extracting channel labels along with their PageRank
channel_labels_with_pagerank = [(G.nodes[node]['label'], pagerank[node]) for node in G.nodes()]

# Sort the channels by their PageRank for better understanding
sorted_channel_labels_with_pagerank = sorted(channel_labels_with_pagerank, key=lambda x: x[1], reverse=True)
sorted_channel_labels_with_pagerank
```

```
[('Frank Medrano', 0.045199234397657356),
 ('Vsauce2', 0.03744961687264593),
 ('Vsauce3', 0.03650797118723785),
 ('Strength Project', 0.025888273024193848),
 ('CT Fletcher Motivation', 0.024411288869881993),
 ('fightTIPS', 0.023778017456458005),
 ('JeromeFitness', 0.022460645613480164),
 ('Android Authority', 0.015457353134957594),
 ('SoundGuys', 0.015366198652473122),
 ('Gary Explains', 0.015366198652473122),
 ('The Enigma TNG', 0.015108293578202134),
 ('Saiyan Enigma', 0.015108293578202134),
 ('Numberphile', 0.013628630522469386),
 ('Jonathan Morrison', 0.011531791522297684),
 ('Tim Ferriss', 0.01115719078800834),
 ('The Verge', 0.009151796646482191),
 ('Bodybuilding.com', 0.008857101323770246),
```

Channels like Frank Medrano and Vsauce (we are referring the secondary channels of Vsauce) lead with the highest PageRank scores, indicating their noteworthy influence and centrality within this network. These channels, due to the number and strength of their connections, play a pivotal role in the flow of information through the network, especially Vsauce, as we have seen from the previous chapters.

The PageRank score of the Bioneer channel, sitting at 0.00373, indicates a modest level of influence within the network when compared to the highest-ranking nodes. This score places the Bioneer in a unique position: while it is not among the most influential in terms of this specific metric, it does hold a certain level of centrality that could be pivotal for the network's connectivity.

In comparison to top-ranking channels like Frank Medrano and Vsauce2, which have PageRank scores more than ten times higher, the Bioneer does not serve as a primary hub in the network. However, it is important to note that PageRank is a relative measure. The Bioneer's score does not negate its importance or the quality of its content; rather, it reflects its role within the web of this network’s connections.

The Bioneer's PageRank suggests it serves as a bridge or connector between different clusters or communities within the network, rather than being a dominant influencer. This could be an advantageous position, as it allows the channel to bring together diverse ideas and audiences, fostering a rich exchange of information and potentially influencing multiple communities indirectly.

On the other end, a lot of channels have a smaller PageRank value, suggesting they occupy fewer central roles. This is not to say these channels are unimportant, but in the context of this specific network's topology and link distribution, they exert a smaller influence on the network's overall connectivity.

Conclusion

The comprehensive analysis of the Bioneer YouTube channel network has illuminated the intricate web of connections that define its digital ecosystem. The network, while vast, highlights a star-like structure with the Bioneer channel as a central hub, forming extensive connections that suggest its pivotal role in content dissemination and community engagement. Notably, the network's resilience is underlined by its diversity, with channels from various content domains forming a rich tapestry of interconnectedness, yet it also reveals points of vulnerability in its reliance on the Bioneer for maintaining structural integrity, as it was proved by the calculations that were shown in the previous chapters.

The presence of five distinct communities, each with a shared thematic or topical focus, demonstrates the network's segmentation into specialized clusters. These range from the largest community, dominated by mainstream and niche fitness channels with a bent towards bodybuilding, to communities focused on calisthenics, science, digital entertainment, and technology. Each community, while diverse within itself, contributes uniquely to the network's collective narrative.

Gender analysis has highlighted a predominance of male-led channels, reflecting broader societal trends and emphasizing the need for more inclusive representation within the network. The analysis of homophily based on country and subscriber count, as well as video and view count, has shed light on the network's inclination towards forming connections based on shared national backgrounds, audience sizes and similar video and view counts, while also maintaining some cross-boundary interactions essential for diversity and idea exchange.

Graph density metrics indicate a selective and sparsely connected network, indicative of room for growth and potential for more robust community structures.

The PageRank analysis highlighted the Bioneer's substantial influence, despite not having the highest score, indicating a strong engagement with its audience and the network at large. The community detection further revealed the channel's role in linking specialized communities, suggesting that the Bioneer effectively transcends the boundaries of content categories, fostering cross-community interactions. Overall, the Bioneer might not be the most influential channel, but it is an irreplaceable bridge between channels with various interests.

This study underscores the multifaceted impact a single channel can have in the context of content consumption on YouTube. The Bioneer exemplifies how content diversity and a strategic position within the network can facilitate the amalgamation of audiences with varied interests and backgrounds, contributing to the richness of the platform's content ecosystem. It showcases the potential for content creators to act as unifying entities in an increasingly segmented digital landscape.

To conclude, the Bioneer network is a microcosm of the YouTube platform at large, with its own unique dynamics and a constellation of channels that are interconnected yet distinct, each playing a role in the content landscape. The network is characterized by its central influencers, the selective nature of its connections, and

the presence of specialized communities that together create a dynamic and resilient ecosystem.

Appendix

All calculations and how they are done can be found in this repository: [panos-span/sna_youtube]