Based on Twitch Gamer: A Multidimensional Network Analysis

Xiaosheng Li (U6519742)

Abstract— **This paper presents a comprehensive study of the Twitch Gamers Social Network through network analysis methodologies, with a particular focus on the application of Graph Neural Networks (GNNs) for link prediction. By exploring the 'Twitch Gamers Social Network Dataset', which was collected via the public API in Spring 2018, we dissect the intricate social dynamics of Twitch users. Utilizing network models such as the Barabasi-Albert (BA) and Watts-Strogatz (WS) models, we reveal the underlying structures and behaviors that govern the interactions within the Twitch community. The contributions of our work are twofold: we provide insights into the influential factors that promote connections within Twitch's social network, and we demonstrate the efficacy of GNNs in predicting potential connections between users. Our results underscore the significance of user engagement, measured via views and life time on Twitch, as well as language compatibility, in the formation of these digital ties.**

I. INTRODUCTION

As the internet has evolved into a ubiquitous force, it has given birth to a plethora of new media platforms, revolutionizing the way we consume content [1]. The surge of live streaming websites exemplifies this transformative era, where video broadcasting becomes an interactive experience. Globally recognized platforms like Twitch and YouTube Live, regional giants such as Douyu TV and Huya TV in Chinese-speaking locales, and the novel rise of TikTok, which has harnessed short-form videos to propel its streaming capabilities – all depict the vibrant diversity of live streaming.

Among these, Twitch has emerged as a particularly influential force by riding the crest of the gaming wave. Gaming streams constitute the core of Twitch's appeal, and its global popularity has attracted an enormous community of gamers. Twitch functions as not just a streaming website but as a social live-stream channel, driven by the streaming community and subdivided into micro-communities, tethered together by viewers who watch, follow, and subscribe [2].

Twitch's unique place at the crossroads of entertainment and social interaction is a chance for detailed study. Analyzing Twitch can reveal how digital groups form, what their content habits are, and how they connect.

We are analyzing for several reasons. We want to better understand Twitch's user interactions, community development, and how streamers can engage their audience. By examining these, we aim to help create a unified and active streaming environment that benefits Twitch users and streamers and sets an example for future online platforms.

In this research, we use network analysis to uncover details of Twitch's active online environment – exploring the people who game, watch, and interact to make Twitch a successful streaming platform[3].

II. FUNDAMENTALS OF NETWORK ANALYSIS

Studying networks has evolved from a social science tool to one used across various fields such as computer science and biology. This approach aims to simplify complex systems by presenting them as networks made up of nodes linked by connections. It provides a clearer structure to analyze the relationships and patterns that may not be immediately visible in the initial complex systems. This network-centric view often unearths systemic patterns and latent architectures concealed within the original arrays of data.

A. Models

This investigation will incorporate theoretical frameworks central to network analytics, two of which are pivotal: the Barabasi-Albert (BA) model [5] and the Watts-Strogatz (WS) model [6].

Each model encapsulates vital attributes recognized within empirical network configurations. The BA model is instrumental in manifesting networks mirroring scale-free topologies. These networks are distinguished by their nodal degree distribution adhering to a power law; this translates to an asymmetry where a minority of nodes are substantially interconnected hubs amidst a plethora of sparsely linked nodes. Such a hierarchy resembles the inherent social web where select individuals command expansive networks contrasted by the majority's fewer connections.

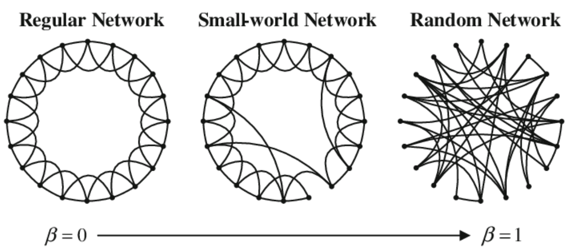

Fig. 1. An example of Watts-Strogatz small world network model [9].

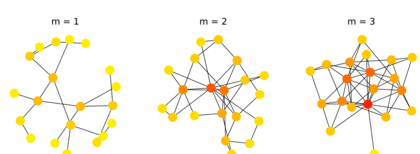

Fig. 2. An example of Barabasi-Albert (BA) model.

Conversely, the WS model's forte lies in its production of networks displaying both intense local connectedness (akin to that of structured lattices) and reduced path distances (reminiscent of random graphs). The resulting 'small-world networks' reflect these dual properties and are emblematic of numerous natural and artificial networks, spanning the gamut from societal interaction webs to neurological and utility frameworks.

In the context of network analysis, metrics like nodes' centrality (quantified by degree, betweenness, and closeness) and the clustering coefficient stand as foundational. Centrality parameters aid in pinpointing pivotal nodes, whilst the clustering coefficient gauges the nodes' propensity to form closely-knit groups [7]. Electing the BA and WS constructs for our exploration aligns with our objective to dissect and recreate the Twitch user network's authentic convolution. These models potentiate the investigation of network characteristics, for instance, the prevalence of hubs (eminent Twitch streamers with vast followings) and close-knit clusters (cohorts of Twitch users with mutual interests).

III. Dataset Overview

In conducting this research, we engaged the "Twitch Gamers Social Network Dataset," meticulously assembled through the Twitch public API in the spring season of 2018 [4]. Twitch commands an international audience, firmly establishing itself as a titan in the arena of video game broadcasting. The collected dataset offers a reflection of the complex social interactivity ingrained in the platform's environment.

The composition of the dataset encapsulates a graph-based model, portraying the Twitch participants, or 'vertices,' linked via 'connections,' indicative of the reciprocal following activity. With figures reaching 168,114 vertices for Twitch participants and 6,797,557 connections pertaining to the said following activity, the dataset reveals itself as a singular, profoundly unified construct, boasting a 0.0005 density and 0.0184 clustering coefficient. Such metrics indicate a high degree of interconnectivity and communal overlap amongst the Twitch participant network.

The dataset itself is underpinned by two principal data structures:The file named 'large twitch edges' delineates the list of connections by indicating the mutual following amongst the platform's users. Within this file, columns labeled 'numeric id 1' and 'numeric id 2' correspond to the individual Twitch users who are linked by a bilateral connection.Correspondingly, the file 'large twitch features' comprehensively details the attributes associated with Twitch users. It is a repository of crucial metrics, including 'views', signifying the viewership count of a user's content; 'mature', delineating whether content is marked appropriate for mature audiences; 'language', identifying the primary linguistic use by the user; 'affiliate status', denoting partnership with Twitch; and 'life time', representing the span of a user's activity on the platform.

Notably, the graphic representation is inherently bidirectional, emphasizing an equal footing in the follower relations — signifying reciprocity that if User A subscribes to User B, User B also reciprocates. While the dataset might not capture intricate traits specific to nodes and connections, it does encompass node labeling, facilitating a categorical analysis of user types.

IV. Network Analysis Methodology

To unravel the intrinsic complexities of Twitch gamer network, our study applies two pivotal network models to chart the landscape effectively: the Watts-Strogatz (WS) model for its capacity to frame 'small-world' networks, and the Barabasi-Albert (BA) model, which simulates the scale-free nature of real-world networks. Our discourse delves into the intricacies of these models, spotlighting the code that delineates their structure and functionality.

A. Forging the Small-World Network

The WS model serves as the scaffolding for our analytical expedition, designed to crystallize networks that harmonize high-clustered nodes with unabbreviated path spans. To execute this model, our code proceeds through stages:

1) Initiation of the WS Network::

- We commence with a pristine ring lattice base.
- Introducing stochasticity, we adjust the links with a probability p, refining network entropy.

In our script, we emplace a definable function that forges a WS configuration given our inputs - nodes, mean degrees, and rewiring odds.

2) Probing Graph Traits::

- Tallying nodes and connections discerns the graph magnitude.
- Encoding functions to distill the arithmetic mean of node connections, clustering index, and path spans across the network.

The codeset will then roll out an anthology of network indices, including counts of both nodes and connections, accompanied by measures of mean connection degree, clustering ratios, and average linkage span for our WS network.

B. Emulating the Scale-Free Universe

The BA model, on the other hand, imparts networks featuring nodal bonds that obey power-law distributions, underscoring 'preferential attachment'. To fabricate such a network, the methodology encompasses the next steps:

1) Construction of the BA Network::

- Initiating with a nucleus of densely knitted nodes.
- Iteratively annexing novices preferentially to the existing populous by degree.

Here, the code invents a BA network, pondering the original node count, the volume of connections per newcomer, and the formulated strategy for network evolution.

2) Assessment of Graph Features:: A snippet from our code repository extracts and displays the average clustering indices and linkage lengths characterizing the BA graphical design.

- Aggregating towards an average of both clustering coefficients and path lengths.

C. Visualize Network Structure

To enhance our analysis, it is important to visualize the architecture of both models. The WS model's graph, Fig.3, hints at an archetype akin to social constructs – vertices conceive individuals whilst the edges craft the interpersonal connections. Upon applying the spectral layout, the graph reveals community clusters, telling a tale of tight alliances within a larger social quilting. The labels conglomerating to the graph's left suggest a well-knit cluster akin to a vibrant local community. In contrast, peripherally-placed vertices, namely 571 and 586, denote potentially marginal figures within the network's broader scope, while a quintet of vertices—20, 22, 211, 378, 473—orbit just shy of the nucleus, potentially a niche ensemble with ties to but distinct from the prime cluster. Moreover, nodes 455 and 454 linger near the core aggregation, hinting at liaison roles bridging wider network segments.

The BA model's grapg, Fig.4, mainly uses spring layout.In the BA model's portrayal, a circular confederation emerges, highlighting a hierarchy of connectivity wherein the center nests a hub-like cluster – a testament to some vertices' highly interconnected status. Meanwhile, a peripheral band

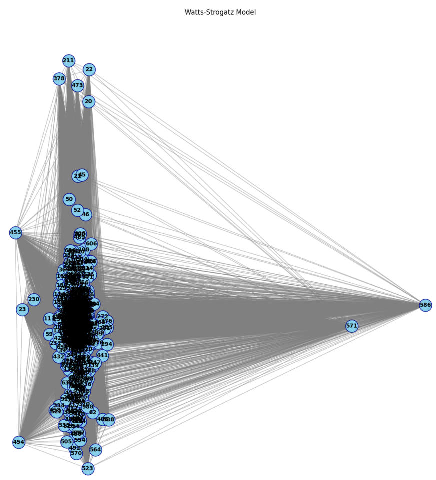

Fig. 3. Watts Strogatz model graph.

suggests a constellation of secondary assemblies, numerically less connected yet socially active. A singular label, 2656, skirts the fringes, disbursed with scant links to the core, emblematic of a fringe entity within this gamified cosmos. This scenography of Twitch's network recoils into a narrative of intricate communal relations, where our analytical foray seeks to render discernible the matrix's core, its emergent subcultures, and the individual members who diversify this digital ecosystem. On the broad digital expanse that Twitch encompasses, these representational graphs illuminate the propagation of ideas and the symbiotic interplay that fortifies the platform's thriving collective.

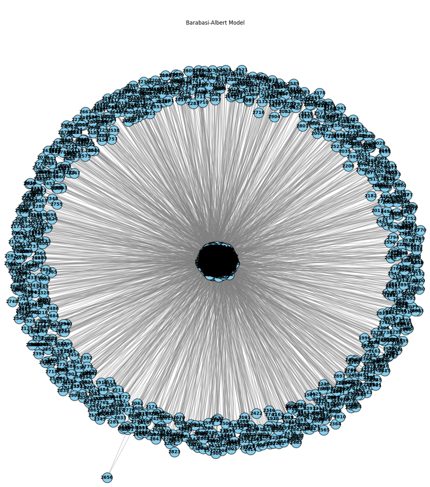

Fig. 4. Barabasi Albert model graph.

V. Graph Neural Networks Methodology

Graph Neural Networks (GNNs)Evolving our network analytical prowess, we integrate Graph Neural Networks (GNNs), poised at the cutting edge of deep learning techniques [8]. GNNs excel in analyzing and processing data configured in graph formats, adeptly mapping out complicated relationships within a network. GNNs function through a recursive mechanism dubbed ”message passing”, wherein each network node transmits and reciprocates information with its neighbors. Successive rounds of these exchanges infuse each node with a holistic comprehension of the network, allowing GNNs to develop comprehensive representations of each node, connection, and the overarching graphscape.

By embedding GNNs within our analytical toolset, we amplify our interrogation of the complexity within the network of Twitch users. Harnessing advanced pattern recognition and relational intricacies, GNNs promise to unearth groundbreaking perspectives on the interplay and communal fabric of Twitch.

A. *GNN Conceptual Overview*

As we move from traditional neural networks to Graph Neural Networks (GNNs), we start working with graphs to detail data relationships and patterns. Fig.5 in our research shows this graph setup, which connects points called nodes with links referred to as edges. This format brings out relationship details that typical non-graph data setups often miss.

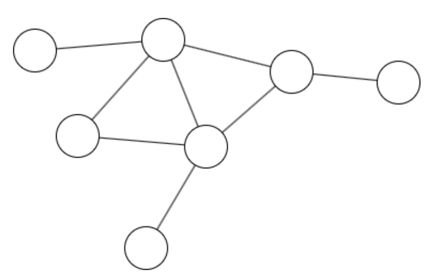

Fig. 5. Graph Data.

Adding to this graph-based data format is the neural network part, displayed in Fig.6. The neural network can use different designs, from straightforward perceptrons to more elaborate multi-layered or repeated patterns. The choice of design depends on the needs of the specific task.

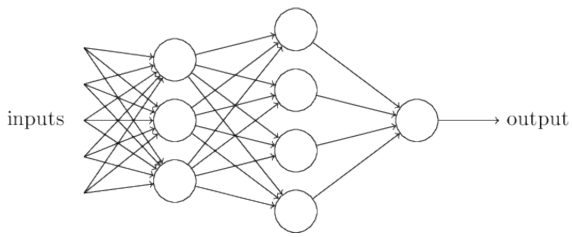

Fig. 6. General Neural Network Model.

Combining these parts gives us our GNN, shown in Fig.7. While it resembles traditional neural networks with its layers and adjustable parameters, the GNN has unique features. It works on the graph format using convolutions that make use of the connections and node features. It also uses functions that introduce non-linearity and can apply methods to avoid overfitting.

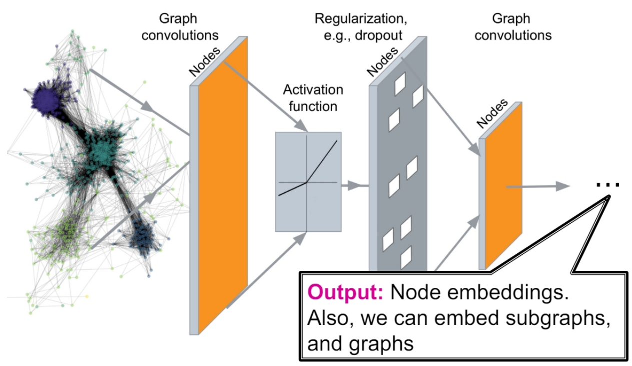

Fig. 7. Graph Neural Networks Model[11].

B. GNN Formalization

A typical GNN layer can be formalized in a relatively standard equation which abstracts its operation [10]:

$$H^{(l+1)} = f(H^{(l)}, A)$$

where:

- $H^{(l)}$ represents the node features (or embeddings) at layer l, with $H^{(0)}$ being the initial node features.
- A stands for the adjacency matrix of the graph which encodes the edge connections.
- f is a differentiable function that embodies the graph convolution operation. It updates the node features based on their own features and their neighbors' features as defined by A.

To delve further into the intricacies of GNN operations, let us expand the function f:

$$f(H^{(l)}, A) = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$

In which:

- $\widetilde{A} = A + I$, the adjacency matrix with added self connections through the identity matrix I.
- $\widetilde{D}$ is the degree matrix of $\widetilde{A}$.
- $W^{(l)}$ is the weight matrix for layer l that needs to be learned during training.
- σ denotes the activation function, such as ReLU or Sigmoid.

This formula makes graph convolution easier to understand as used in a Graph Convolutional Network (GCN). This structure is one of the basic Graph Neural Network architectures. It shows clearly how each node refreshes its features by combining details from its neighbors (and itself). After that, it adds changes using the weight matrix and applies the non-linear function.

In understanding and detailing GNNs, one must also account for how to perform pooling (to aggregate node features at different levels) and how the final output for tasks such as classification, regression, or link prediction is generated from the node representations learned by the network.

C. GNN Implementation for Twitch Gamer Network Analysis

1) Model Architecture: Our model's architecture, shown in Fig.8, begins with node input features. These features serve as the raw data that will be processed by our GNN. They contain key details about Twitch gamers, showing their activity and how they interact.

After taking in these node features, the data goes through two layers of GraphSAGE, a GNN design good at pulling in data from a node's nearby network. Using two layers of GraphSAGE means the resulting node profiles include not just close neighbors but also those further away, offering a wider context. Once it goes through the GraphSAGE layers, the resulting node profiles form a detailed mathematical snapshot of each node's details and their place in the network structure.

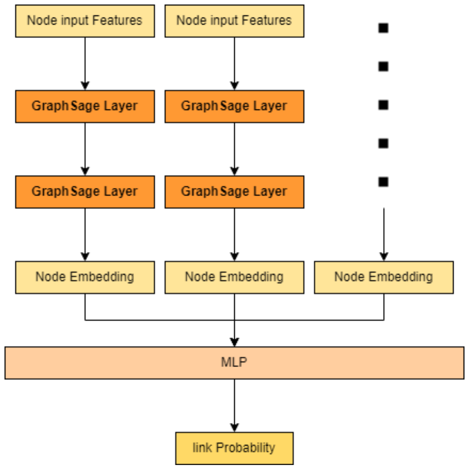

Fig. 8. Model Architecture.

These profiles then go into a multi-layer perceptron (MLP), a more standard neural network, for more changes. At the end of this process, the system outputs a link probability. This last part changes what we've learned from the profiles into clear predictions, specifically, how likely it is that users in the Twitch network will connect.

2) Link Prediction: The specific task we tackle with our GNN revolves around link prediction within the Twitch gamer network. Link prediction is a fundamental task in network science, aiming to determine whether two nodes are likely to create a connection. It has a host of practical applications, including friend recommendation in social networks or predicting protein-protein interactions in biological networks. We employ link prediction to estimate the likelihood of connections between Twitch users, which might signify potential collaborations or community formations that are yet to be observed.

3) Evaluation Dataset and Result: We begin to create our evaluation plan with the 'large twitch features' and 'large witch edges' datasets mentioned in Section 3. Our method relies on a balanced dataset with both positive examples, which are real connections, and negative examples, which show no links, to test the predictive ability of our GNN.

The steps for the evaluation are:

- Picking Positive Edge Samples: We take a subset from the 'large twitch edges' data where edges between pairs of gamers confirm connections.
- Choosing Negative Edge Samples: We make an equal number of pairs where no edge exists between nodes in the same dataset, to see if the model can tell apart existing from non-existing links.

With this careful evaluation plan, aware of actual connections shown in the Twitch data and considering made-up negative cases, we set up a solid way to compare how our GNN performs. The aim is not just to check known links but to find a pattern that could predict the creating of new nodes or links in the quickly changing Twitch social world. The outcomes from this model could give important knowledge about how communities grow and the way people interact on digital platforms like Twitch.

D. GNN Results

Before looking at feature analysis, it's important to consider the GNN model's operational performance. Over 100 epochs, we carefully monitored and recorded the model's performance with a loss curve that the accompanying Fig.9 shows.

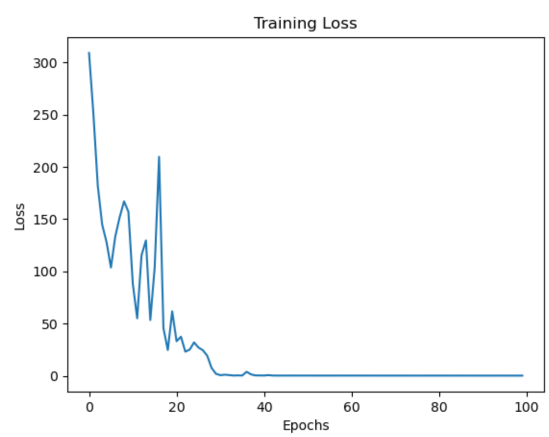

Fig. 9. Training loss of GNN.

Initially, the model's loss was quite high, over 300, signaling the difference between the model's predictions and actual data on connections. But as training moved forward, the loss decreased quickly. By about epoch 40, the loss leveled off, indicating the model's predictions were lining up better with the actual network structure. After finishing 100 epochs, the loss became stable and was close to zero, showing the model's effectiveness in learning from the data. This result confirms the GNN model's skill in identifying the detailed patterns within the Twitch gamer network that affect the chance of users connecting. The decreasing loss function not only proves the model's predictive power but also highlights the strength of the network analysis being done.

Analyzing the output from the GNN code yields a comprehensive understanding of the features that contribute to the likelihood of forming connections within the Twitch gamer network. Let us examine the detailed numeric values and key features to explain why certain users are most likely, and others are least likely, to establish a connection with the target user.

1) Target User Features:

- **Views:** 15
- **Mature Content:** No (0)
- **Lifetime Days on Twitch:** 69
- **Language:** English (EN)
- **Affiliation:** No (0)
- **Account Status (Dead Account):** No (0)

Numeric ID	Views	Mature	Life Time	Created At
24516 (Target User)	15	0	69	2018-05-12
121168	42,863,330	0	2289	2012-07-06
24140	10,458,706	0	1900	2013-07-30
21405	63,285,497	0	2339	2012-05-17
8227	11,553,880	1	3209	2009-12-29
120668	11,764,956	0	3209	2009-12-29

Updated At	Dead Account	Language	Affiliate
2018-07-20	0	EN	0
2018-10-12	0	TR	0
2018-10-12	0	KO	0
2018-10-12	0	ES	0
2018-10-12	0	EN	0
2018-10-12	0	DE	0

TABLE I

TOP 5 USERS MOST LIKELY TO CONNECT TO TARGET USER (24516) ON TWITCH NETWORK

2) Analysis on Top 5 Most Likely to Connect: The top 5 users most likely to connect all have significantly higher view counts compared to the target user. High viewership could correlate with higher visibility and influence within the community, naturally elevating these users' potential for connection.

- **Views:** The users with the highest connection potential have a large number of views, between 10,458,706 to 63,285,497. This indicates that the model sees users with a lot of engagement as having a greater chance to make new connections because of their presence and status in the Twitch community.
- **Language:** The language among the top 5 likely connections includes Turkish (TR), Korean (KO), Spanish (ES), English (EN), and German (DE). Despite the variety, the inclusion of English-speaking users suggests that while language may play a role, the model points to other, stronger common traits that lead to a high likelihood of connecting.
- **Life Time:**The time spent on Twitch by users most likely to connect is much longer, all exceeding 1,900 days. This shows that a prominent, long-term presence on the platform can make forming connections more likely, indicating an established place within the community.

Numeric ID	Views	Mature	Life Time	Created At
24516 (Target User)	15	0	69	2018-05-12
91955	2,288	0	1310	2015-03-08
65522	413	0	1846	2013-08-22
66601	1,252	0	1642	2014-03-15
135295	721	0	2134	2012-12-03
82946	617	0	1943	2013-05-06

Updated At	Dead Account	Language	Affiliate
2018-07-20	0	EN	0
2018-10-08	0	RU	0
2018-09-11	0	RU	0
2018-09-12	0	RU	0
2018-10-07	0	NO	0
2018-08-31	0	EN	0

TABLE II

TOP 5 USERS LEAST LIKELY TO CONNECT TO TARGET USER (24516) ON TWITCH NETWORK

3) Analysis on Top 5 Least Likely to Connect: With the Twitch network, an account's lifetime may indicate the degree of establishment within the community and potential networking.

- **Views:** Users with the lowest probability to connect have fewer views, ranging from 2,288 to 413. This matches the target user's low view count, suggesting that the model sees less chance of connection among users with lower activity.
- **Language:** Among those least likely to connect, languages such as Russian (RU) and Norwegian (NO) appear next to English (EN). The presence of non-English speakers, especially with low engagement, seems to lessen the chance of these users connecting with the English-speaking target user.
- **Life Time:**The time on Twitch for the least likely to connect varies from 1,310 to 2,134 days. This is less than those most likely to connect but still indicates an established presence, implying that time spent is important but not the only factor in predicting connections.

4) Other Parameters: Little or No EffectBeyond views, language, and time spent on Twitch, the GNN looked at other factors. But these other factors seem to matter less or not at all in guessing the chances of connections in the Twitch network. These are:

- **Mature Content:** Since the 'mature' attribute is uniformly set to 0 for both the top and bottom users for likelihood to connect, as well as the target user, it seems that mature content doesn't affect the GNN's predictions on the likelihood of making connections.
- **Creation and Update Dates:** Even with the variation in creation and last updated dates for accounts, there isn't a noticeable pattern or effect of these dates on the GNN's predictions about connections. Across the board, from the target user to other users in both the most and least likely to connect, there's no connection between the probability of forming a connection and these timelines.
- **Numeric ID:** The numeric ID is just an assigned number and doesn't contain any information about relationships or behavior. So, it's thought to not affect the analysis and serves only as a unique identifier for each account in the dataset.
- **Dead Account:** In this dataset, all users in both the most and least likely to connect have the 'dead account' feature set to 0, just like the target user. This uniformity suggests that in this case, the GNN doesn't use this feature to tell potential connections apart.
- **Affiliate Status:** In the same way, with the affiliation status being 0 for every user, including the target, this indicates that this detail doesn't add anything useful to how the GNN predicts in this analysis.

The trends suggest that high view counts and a lengthy time on Twitch are leading indicators of a higher likelihood to connect, according to the GNN. Even with language diversity in the group most likely to connect, this shows the GNN does not view language match as the key factor for these predictions. Rather, engagement level and a long-standing presence carry more weight in the model, which sees active and established users as key to the community's network, despite language differences. On the other hand, those least likely to connect are characterized by fewer views and shorter durations on Twitch, which, even with a language match with the target user, points to a complex insight from the GNN on what influences connections on Twitch.

VI. Conclusion

This study conducted a comprehensive analysis of the Twitch Gamers Social Network, using network models and Graph Neural Networks (GNNs) to understand Twitch's complex social network.

The methodology explored the network with Barabasi-Albert (BA) and Watts-Strogatz (WS) models. These helped deepen knowledge of the network's properties. The study used to practical GNN applications to decode Twitch interactions.Findings from a GNN trained over 100 epochs pointed to user activity and language as key predictors of connections. These insights not only confirmed the GNN's predictive accuracy but also highlighted the layers of social connectivity on Twitch that extend beyond content viewing.

In summary, the research combined data processing, network analysis, and neural networks. It provided a detailed perspective on the Twitch network. Every research phase, from data exploration to GNN analysis, built a clearer understanding of Twitch's social landscape.

VII. Link & Contribution

GITHUB link

Xiaosheng Li: All the work.

Appendix

Ethical Considerations and Societal Impact: This paper presents work focused on analyzing the Twitch Gamers Social Network using network analysis methodologies and Graph Neural Networks (GNNs) for link prediction. Our research aims to enhance the understanding of digital social interactions within the Twitch community. This work may bring various societal impacts. From an ethical perspective, it is essential to consider the implications of analyzing user data, even when publicly available, to ensure privacy and confidentiality are maintained. Additionally, while this research can improve user experience on platforms like Twitch, it is crucial to be mindful of the potential for such technologies to be used in ways that could exploit user behavior or contribute to issues such as online addiction. Future societal consequences might include the adoption of similar network analysis techniques in other social platforms, leading to more personalized user interactions but also raising concerns about data privacy and the ethical use of AI in digital ecosystems.

References

[1] McMullan, J. (2020). A new understanding of 'New Media': Online platforms as digital mediums. Convergence, 26(2), 287-301.

[2] Henten, A., Tadayoni, & R. (2008). The impact of the internet on media technology, platforms and innovation. The Internet and the mass media, 45-64.

[3] Napoli, P. M. (2011). Audience evolution: New technologies and the transformation of media audiences. Columbia University Press.

[4] Rozemberczki, B., & Sarkar, R. (2021). Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings. arXiv preprint arXiv:2101.03091.

[5] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.

[6] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature, 393(6684), 440-442.

[7] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social networks, 1(3), 215-239.

[8] Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57-81.

[9] Zhang, Z., & Wang, K. (2013). A trust model for multimedia social networks. Social Network Analysis and Mining, 3, 969-979.

[10] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

[11] https://web.stanford.edu/class/cs224w/slides/03-GNN1.pdf