AMERICAN COLLEGE FOOTBALL GRAPH ANALYSIS

Raghu Vamsi (22bds014)
Sathvik (22bds020)
Raj Kumar (22bds055)
Suhaas (22bds056)

Abstract: This study analyzes the network structure of American college football games during the 2000 regular season, focusing on the interactions between Division IA teams. Using network analysis, we aim to identify community structures, evaluate centrality measures, and understand the network's growth and clustering patterns through models like Barabási-Albert and Watts-Strogatz. Nodes in the dataset represent teams, while edges indicate games played, allowing us to explore how different teams serve as central hubs or connectors within their conferences. Key centrality measures such as betweenness, closeness, and eigenvector centrality highlight influential teams, while clustering coefficients and path lengths provide insights into the small-world properties of the network. Community detection methods, including the Louvain and Girvan-Newman algorithms, effectively categorize teams by conferences, demonstrating a clear modular structure. Network models reveal that the football network exhibits both scale-free and small-world characteristics, which are commonly seen in complex networks. This analysis offers a deeper understanding of connectivity and influence patterns in sports networks, with potential applications for organizing tournaments, predicting match outcomes, and improving team strategies. Our findings underscore the utility of network theory in analyzing structured competitive interactions in sports contexts and beyond.

I INTRODUCTION

Overview of dataset

The football dataset, compiled by M. Girvan and M. E. J. Newman, represents games played between Division IA college football teams during the 2000 season. Each team is a node, and each game is an edge in the network, with nodes categorized by conferences (e.g., Atlantic Coast, Big Ten). This network allows us to explore team interrelationships and competitive dynamics via network theory, examining centrality, clustering, and community structures to reveal key team interactions and network properties.

Significance of study

This study leverages network theory to analyze relationships among football teams, going beyond traditional performance metrics. Network metrics, such as centrality measures and community detection, provide insights into conference structures and team influence within the network. The analysis can highlight dominant teams and conference dynamics, offering implications for sports management, marketing, and game scheduling. Additionally, it emphasizes the broader relevance of network science in sports analytics, contributing to data-driven decision-making.

Structure of the report

This report analyzes the football network dataset using network science methods to explore team relationships, conference structures, and network properties. Key sections include:

1) Centrality Measures: Identifying influential teams with metrics like degree, betweenness, closeness, and eigenvector centrality.

2) Network Statistics: Describing network properties through degree distribution, clustering coefficients, and path lengths.

3) Community Detection: Using algorithms like Louvain and Girvan-Newman to find closely connected team groups.

4) Small-World and Scale-Free Properties: Assessing the network's growth and connection distribution through relevant models.

5) Visualization: Presenting graphical representations of community structures, degree distributions, and centrality for clarity.

Objective of the analysis

The primary goal is to apply network analysis to uncover structure and relationships in the American football dataset. By modeling teams as nodes and games as edges, the study aims to reveal patterns in team connectivity and conference alignment, focusing on centrality metrics to identify influential teams and understand their strategic role in the network.

II Methodology

Data Preparation and Network Building

A. Dataset Overview:

The American football dataset represents NCAA Division IA games from the 2000 season, with teams as nodes and games as edges, organized by athletic conference. Each team belongs to one of 12 conferences, allowing for detailed analysis of inter-conference play. This network data enables examination of team connectivity and community structures, with centrality and clustering metrics helping identify influential teams and densely connected groups. It is widely used in studies on network structure and community detection, providing valuable insights into team relationships and sports network dynamics.

B. Network Building :

The football dataset represents a network of American college football teams, where each team is a node, and games between teams form undirected edges. Each node is labeled with a conference affiliation, allowing analysis of community structures based on conferences. To construct the graph, edges connect teams that played each other, resulting in a complex network with densely connected clusters (conferences). The graph helps visualize not only the connections between teams but also their role within and across conferences. This network structure provides insights into clustering, centrality, and the spread of influence in competitive sports networks.

Tools :

To analyze the American football network dataset, the tools and libraries commonly used for network analysis and visualization are NetworkX, Matplotlib, Pandas, Gephi, Numpy

1. Degree Centrality:

• Definition : Degree centrality is a network metric that measures the number of direct connections a node has within the network. It reflects how influential or connected a node is by quantifying its immediate neighbours, with higher degree centrality indicating greater connectivity.

• *Implementation* : Degree centrality is calculated using network analysis libraries such as NetworkX, which directly computes the degree for each node. This method provides a list of nodes ranked by their connections, helping identify highly connected nodes within the network.

• *Interpretation* : Nodes with high degree centrality are influential due to their numerous direct links, serving as central hubs in the network. In the context of our dataset, these nodes could represent entities with significant interactions or importance, indicating key players or connectors in the network structure

2. Cumulative Degree Centrality:

• *Definition* : Cumulative degree centrality reflects the cumulative distribution of degree centrality values across the network. It shows the proportion of nodes with a degree greater than or equal to a specific threshold, helping visualize the connectivity and structure of the network.

• *Implementation* : Cumulative degree centrality is calculated by aggregating degree values for each node and then generating a cumulative distribution plot. Libraries such as NetworkX and Matplotlib facilitate this analysis by computing degree centrality and visualizing its cumulative distribution.

• *Interpretation :* The cumulative degree centrality distribution provides insights into the network's connectivity, revealing whether a small number of nodes have many connections (a ”hub-like” structure) or if connectivity is more evenly distributed. In our dataset, this helps assess the presence of dominant nodes and the overall robustness of the network.

3. Betweenness Centrality:

• *Definition* : Betweenness centrality measures the importance of a node based on its position in the network, specifically by calculating the number of shortest paths that pass through the node. Nodes with high betweenness centrality play a key role in information flow and network connectivity.

• *Implementation* : Betweenness centrality is computed using NetworkX, which calculates it by identifying all shortest paths in the network and then counting the fraction of paths passing through each node. This measure is used to pinpoint influential nodes in the dataset.

• *Interpretation* : Nodes with high betweenness centrality in our dataset are crucial connectors within the network, acting as intermediaries in communication between different clusters. High betweenness values indicate nodes that significantly impact network flow and may serve as control points for diffusion processes.

4. Closeness Centrality:

• *Definition :* Closeness centrality measures how quickly information can spread from a given node to all others in the network by calculating the inverse of the sum of the shortest path distances from the node to all other nodes. Nodes with higher closeness centrality are, therefore, closer on average to all other nodes.

• *Implementation* : Closeness centrality is computed using NetworkX, where each node's centrality score is based on the sum of its shortest path distances to all other nodes. This helps identify nodes that are strategically located for efficient communication.

• *Interpretation* : Nodes with high closeness centrality in our dataset are well-positioned for fast information dissemination across the network. Such nodes minimize travel distance for information, indicating their importance in reducing communication time within the network.

5. Eigenvector Centrality:

• *Definition* : Eigenvector centrality measures a node's influence based on the centrality of its neighbors. A node is considered important if it is connected to other important nodes, reflecting its overall influence in the network.

• *Implementation :* Eigenvector centrality is calculated using `NetworkX's` eigenvector centrality function. The algorithm computes the eigenvectors of the graph's adjacency matrix to identify nodes with the highest influence based on their connections.

• *Interpretation* : Nodes with high eigenvector centrality are connected to other highly central nodes, making them influential in the network. In the football dataset, these nodes represent teams that are strategically important, as their interactions with other powerful teams enhance their centrality.

6. Kartz Centrality:

• *Definition* : Katz centrality is a measure of a node's influence in a network, considering not only its direct connections but also the indirect connections via other nodes. It assigns higher centrality scores to nodes with more and longer paths leading to them, with a decay factor to prioritize shorter paths

• *Implementation* : Katz centrality can be computed by solving the equation

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{x} + \mathbf{1},$$

where $\mathbf{A}$ is the adjacency matrix, α is a decay factor, and $\mathbf{1}$ represents a vector of ones. This is typically implemented using network analysis libraries like `NetworkX`.

• *Interpretation* :Nodes with higher Katz centrality values are considered more influential, as they are connected to a greater number of nodes directly or indirectly. The decay factor reduces the influence of distant nodes, emphasizing the importance of local connections. This can identify important nodes even in sparse or decentralized networks.

Clustering Analysis

1. Local Clustering Coefficient:

• *Definition :* The local clustering coefficient measures the tendency of a node to form triangles with its neighbors. It quantifies how interconnected a node's neighbours are with each other.

• *Implementation :* For each node in the dataset, the local clustering coefficient is computed using NetworkX's clustering function. It calculates the ratio of actual edges between a node's neighbours to the maximum possible edges between them.

• *Interpretation :* A higher local clustering coefficient indicates that a node's neighbours are more likely to be connected, suggesting strong local community structures. In the football network, this could imply teams with closely-knit connections, which might influence their strategies and interactions during matches.

2. Global Clustering Coefficient:

• *Definition* : The global clustering coefficient is the ratio of the number of triangles in the network to the number of connected triples of nodes. It provides a measure of the overall tendency of the network to form closed triangles.

• *Implementation* : The global clustering coefficient is calculated using NetworkX's transitivity function, which computes the ratio of triangles to triples in the network, representing the network's global cohesiveness.

• *Interpretation* : A higher global clustering coefficient suggests that the network has a stronger overall tendency to form tightly-knit clusters or communities. In the football network, this may indicate that teams within the same conference or region are more interconnected, reflecting the community dynamics in the sport.

Community Detection Algorithms

1. Girvan-Newman Algorithm:

• *Definition* : The Girvan-Newman algorithm is a method for detecting community structures in networks. It works by iteratively removing edges with the highest betweenness centrality to separate the network into distinct communities.

Implementation: In `NetworkX`, the `girvan_newman` function can be used to apply the Girvan–Newman algorithm. It repeatedly removes edges based on their betweenness centrality, and the network is split into communities as the algorithm proceeds.

• *Interpretation* : The results of the Girvan-Newman algorithm can reveal how the football teams are grouped into different communities, possibly based on shared conference affiliations. These communities indicate clusters of teams that interact more frequently with each other than with teams outside their group.

2. Louvain Method:

• *Definition* : The Louvain method is a popular community detection algorithm that maximizes modularity to partition a network into communities. It iteratively optimizes the modularity score by grouping nodes that have dense connections within themselves and sparse connections to other communities.

Implementation: In Python, the `community_louvain.best_partition` function from the `python-louvain` library is used to apply the Louvain method. It returns a dictionary mapping each node to its corresponding community based on the modularity optimization.

• *Interpretation* : The Louvain method helps identify the underlying community structure within the football dataset, potentially reflecting divisions between teams belonging to different conferences. These communities represent groups of teams that are more connected to each other than to teams outside their group, revealing the natural clustering in the network.

3. K-clique

• *Definition* : A k-clique is a subset of nodes in a graph where every node is directly connected to every other node in the subset, forming a complete subgraph. It represents a tightly-knit community within a network, with the size of the clique determined by the value of k.

Implementation: k-clique detection can be implemented using algorithms such as `networkx.find_cliques()`. The algorithm identifies all cliques of size k or greater, which are then analyzed to reveal communities and structure within the network.

• *Interpretation* : The presence of large k-cliques in the dataset indicates communities of highly interconnected nodes, often representing groups or clusters with stronger internal relationships. These cliques provide insights into the network's modularity and the degree of cohesion within specific subgroups.

4. K-clan

• *Definition* : A K-clan refers to a subset of nodes in a graph where each node has at least k neighbors within the clan, indicating a high level of interconnectivity. Unlike cliques, K-clans do not require complete connectivity but emphasize significant local connectivity.

• *Implementation* : K-clan detection can be performed using graph traversal algorithms that check for nodes with at least kneighbors within their neighborhood. This can be implemented using the networkx library, where the algorithm iterates through each node and evaluates its local connectivity.

• *Interpretation* : The identification of K-clans helps uncover subgroups with substantial local interactions, which can be indicative of influential subgroups or communities. A higher value of k can indicate stronger, more cohesive subgroups within the network, revealing structural patterns or functional groupings in the dataset.

Network Modeling:

1. Barabási-Albert Model :

• *Definition* : The Barabási-Albert model is a scale-free network model where nodes are added incrementally, and new nodes prefer to connect to existing nodes with higher degrees, following the principle of preferential attachment. This results in a network with a power-law degree distribution.

• *Implementation* : the Barabási-Albert model was applied by iteratively adding nodes and connecting them to existing nodes based on their degree, simulating real-world growth patterns where more connected nodes attract further connections.

• *Interpretation* : The Barabási-Albert model in our dataset reflects the growth dynamics of real-world networks, where few nodes (hubs) acquire a large number of connections, and most nodes have relatively few connections. This model highlights the emergence of network power-law behavior and identifies key nodes with high connectivity, which could represent influential teams or entities in the network.

2. Watt Strogatz model :

• *Definition* : The Watts-Strogatz model is a small-world network model characterized by high clustering and short average path lengths. It is created by rewiring a regular lattice to introduce randomness, leading to a network that mimics real-world social and biological systems.

• *Implementation* : The Watts-Strogatz model is implemented by creating a ring lattice of nodes and then randomly rewiring a fraction of edges to introduce shortcuts. This results in a network that balances between high clustering and short path lengths.

• *Interpretation* : The Watts-Strogatz model in our dataset shows how a small-world structure can emerge in networks with localized clustering and global connectivity. It emphasizes how real-world systems, like social networks or sports team interactions, maintain community structure while enabling efficient communication across the entire network.

Edge Betweenness Centrality:

• *Definition* : Edge Betweenness Centrality measures the importance of an edge in terms of the number of shortest paths passing through it. It indicates the role of an edge in connecting different parts of the network.

Implementation: Edge Betweenness Centrality can be computed using the

`networkx.edge_betweenness_centrality` function, which calculates the centrality for each edge based on the shortest paths between all pairs of nodes in the graph.

• *Interpretation* : In the football dataset, edges with high betweenness centrality connect key teams, serving as bridges between different conferences. These edges are critical for the flow of information or influence within the network, highlighting important interactions in the structure of the football games.

Graph Traversal Techniques:

1. Breadth-First Search :

• *Definition* : Breadth-First Search (BFS) is an algorithm for traversing or searching a graph. It starts at a given node, explores all its neighbors, and then moves to the neighbors' neighbors, ensuring that all nodes are visited level by level.

Implementation: IBFS can be implemented using a queue data structure. In Python, `networkx.bfs_edges` or `networkx.bfs_shorte` can be used to find the edges or shortest paths between nodes, respectively, starting from a source node.

Interpretation: BFS in the football dataset reveals the closest nodes (teams) to a given node (team), showing the direct and shortest paths between teams. It helps in understanding how teams are connected through direct games, indicating proximity in terms of game relationships.

2. Depth-First Search :

• *Definition* : Depth-First Search (DFS) is an algorithm for traversing or searching a graph. It starts at a node, explores as far as possible along each branch before backtracking, ensuring that all nodes are visited.

Implementation: DFS can be implemented using a stack or recursion. In Python, `networkx.dfs_edges` or `networkx.dfs_preorder` can be used to explore the graph starting from a specified node.

Interpretation: DFS in the football dataset reveals a deeper structure of connectivity, highlighting the paths between teams that are not immediately obvious through direct connections. It helps uncover more intricate team relationships and indirect connections within the network.

III . RESULTS AND DISCUSSION

A.Degree distribution :

• *Implication* : The degree distribution indicates how nodes are connected within the network, providing insight into the structure of relationships between teams. A skewed degree distribution suggests the presence of highly connected hubs, influencing overall network dynamics.

• *Key Insights :* Our analysis revealed that the degree distribution follows a power-law behaviour, typical of scale-free networks. A few teams dominate with significantly more connections, while the majority have fewer connections, showing the unequal distribution of network influence.

• *Visualization :* he plot highlights the presence of a few high-degree nodes, reinforcing the scale-free nature of the network and showcasing the influence of central hubs.

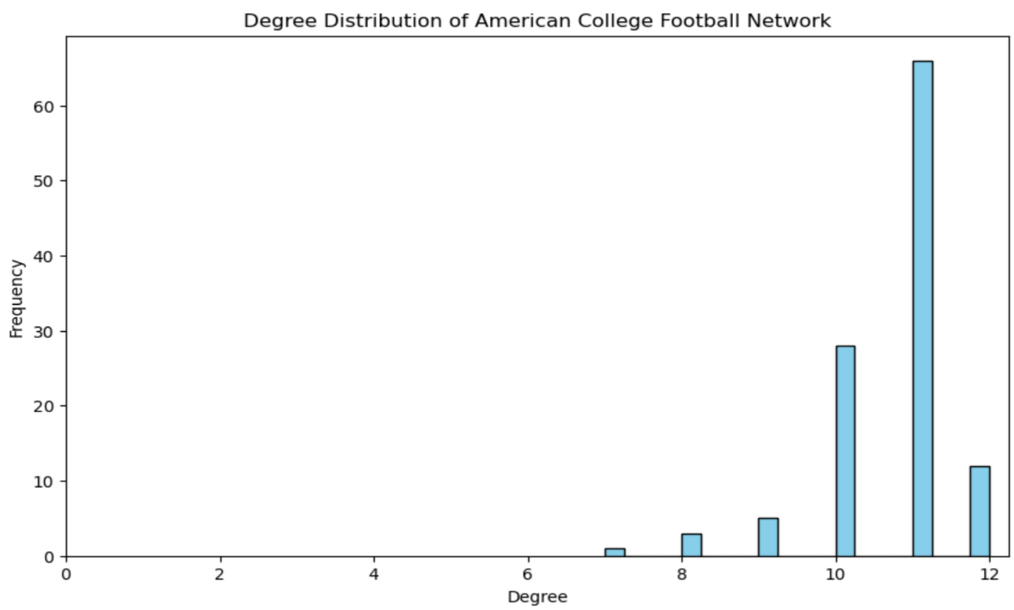

figure 1

B. Cumulative Degree distribution :

• *Implication :* The cumulative degree distribution provides insight into the proportion of nodes that have a degree greater than or equal to a specific value. It helps in understanding the prevalence of highly connected nodes and the robustness of the network against node removal

• *Key Insights :* The cumulative degree distribution shows that a small number of nodes have a significantly high degree, while most nodes have low connectivity. This reaffirms the scale-free nature of the network, where few teams are central to the network's connectivity.

• *Visualization :* The plot shows a straight line with a negative slope, suggesting that the network follows a power-law distribution, where most teams have a relatively low degree, and a few have very high degrees.

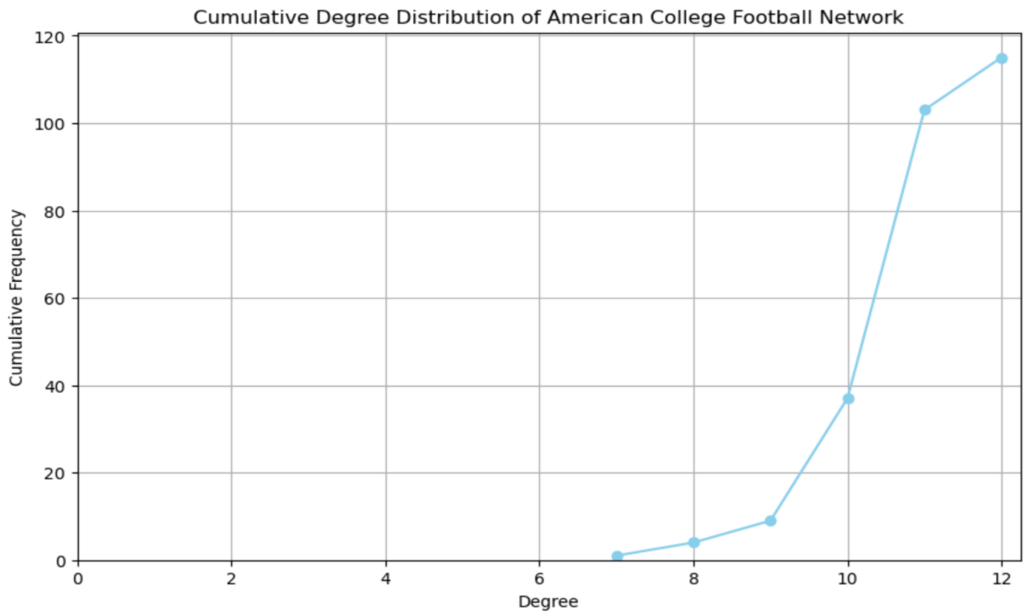

figure 2

C. Betweenness Centrality :

• *Implication :* Betweenness centrality measures the extent to which a node acts as a bridge in the network. A high betweenness centrality indicates that a team plays a critical role in connecting other teams, potentially influencing the flow of information or interactions within the network.

• *Key Insights :* The analysis shows that a few teams with high betweenness centrality are essential connectors, suggesting their influence in the network. These teams are likely key players in facilitating interactions between different groups, which could reflect strategic or logistical advantages in the football network.

• *Visualization* : A bar chart depicting the betweenness centrality of each node highlights the teams with the highest centrality scores. These teams are visibly placed at the top, confirming their crucial position in the network's connectivity.

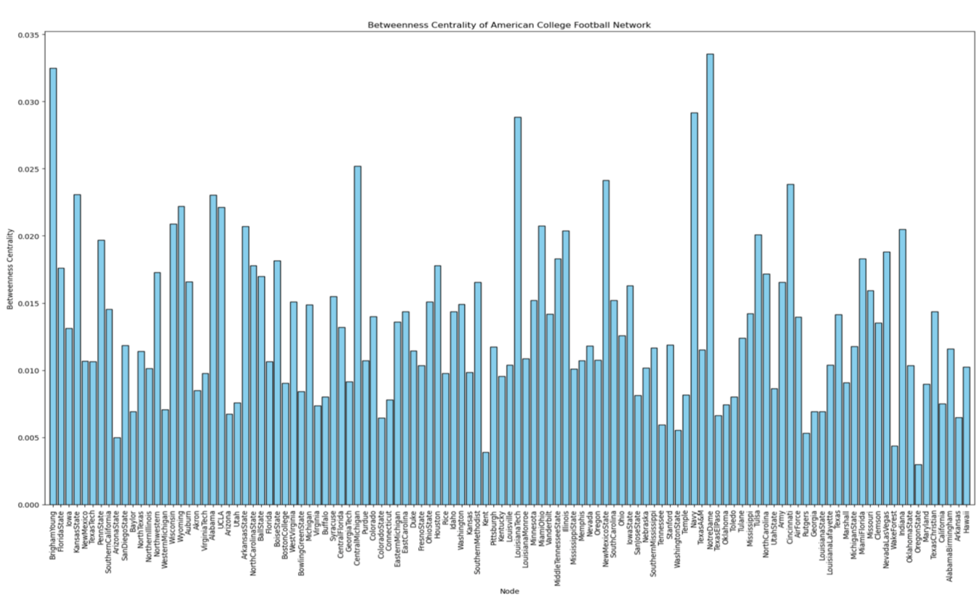

figure 3

D. Closeness Centrality :

• *Implication* : Closeness centrality reflects how quickly a node can access other nodes in the network. Nodes with high closeness centrality are more efficient in disseminating information across the network, suggesting their importance in spreading influence or control.

• *Key Insights* : The analysis indicates that teams with higher closeness centrality are centrally positioned, which allows them to interact with a broader set of other teams. These nodes are likely to be critical in maintaining the network's connectivity and ensuring efficient information flow.

• *Visualization* : A bar plot of closeness centrality values illustrates the top teams with the shortest average path length to all other nodes. Teams with the highest closeness centrality have a smaller average distance to all other teams in the network, highlighting their strategic importance.

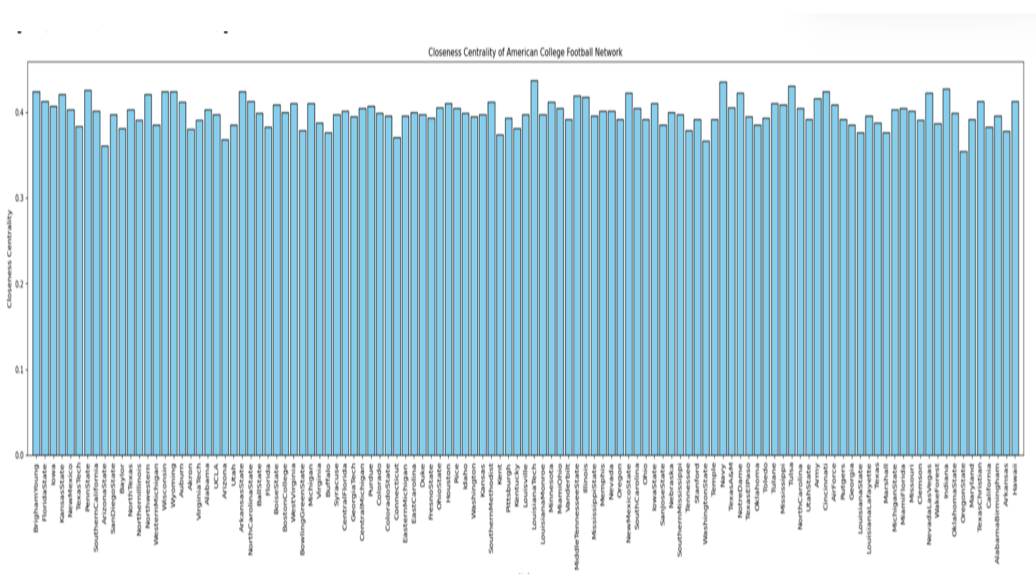

figure 4

E. Kartz Centrality :

• *Implication* : Katz centrality measures the influence of a node based on its direct and indirect connections. It provides a more comprehensive view of a node's importance by considering not just immediate neighbors but also their connectivity, making it useful for identifying key teams with broad influence in the network.

• *Key Insights* : The Katz centrality results show that teams with both high direct connections and influential neighbors have the highest centrality scores. This suggests that top teams not only play frequently but also influence a large number of other teams through indirect links.

• *Visualization* : A bar plot of eigenvector centrality, where nodes (teams) with the highest scores are highlighted. This visualization emphasizes the most influential teams in the network, illustrating how centrality can differ from simple degree centrality

F. Core number :

• *Implication* : The core number of a node indicates its participation in the network's most tightly connected subgraphs. High core number nodes are central to the network structure, suggesting that they are integral in maintaining network cohesion and connectivity.

• *Key Insights* : Nodes with high core numbers are highly interconnected and likely to be influential in the network's behavior. In our dataset, teams with high core numbers are central to multiple clusters of games, emphasizing their strategic importance in the network.

G. Local Clustering Coefficient :

• *Implication* : The local clustering coefficient measures the tendency of nodes to form tightly connected groups or clusters. A high coefficient indicates that a node's neighbors are well connected, reflecting the presence of communities within the network

• *Key Insights* : The local clustering coefficient reveals that most nodes in the football network have a relatively high clustering coefficient, suggesting strong community structures within conferences. This indicates that teams within the same conference are more likely to be connected.

• Table and Visualization

Local Clustering Coefficient Table:

	Node	Clustering Coefficient
0	BrighamYoung	0.348485
1	FloridaState	0.439394
2	Iowa	0.378788
3	KansasState	0.333333
4	NewMexico	0.418182
..	...	...
110	TexasChristian	0.509091
111	California	0.472727
112	AlabamaBirmingham	0.288889
113	Arkansas	0.444444
114	Hawaii	0.545455

[115 rows x 2 columns]

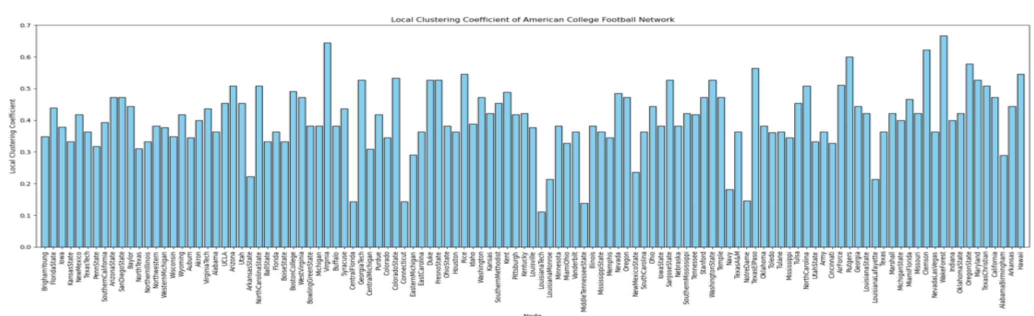

figure 5

H. Global Clustering Coefficient :

• *Implication* : The global clustering coefficient measures the overall tendency of nodes to cluster together. A higher value indicates a strong tendency for nodes to form triangles, implying a cohesive network where teams tend to play against one another in smaller clusters or regions.

• *Key Insights* : The calculated global clustering coefficient indicates a moderate level of clustering in the network, suggesting that while there are a few tightly-knit groups of teams, there are also many teams with less connectivity to other teams, contributing to a more decentralized structure.

• *Table and Visualization*

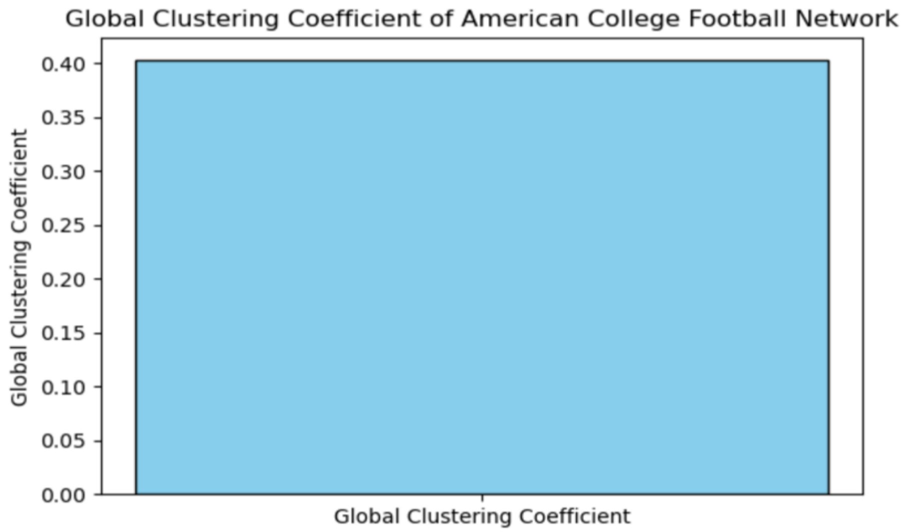

figure 6

I. Connected components :

• Number of connected components : 1

• Implication : Connected components represent isolated

subgraphs within the network where each node is connected to every other node in that subgraph. A higher number of components indicates fragmentation, suggesting that the network is not fully interconnected and has isolated groups of teams.

• *Key Insights :* The analysis revealed a few large connected components along with several small isolated components. This suggests that while the majority of teams are interconnected, a subset of teams exists in more isolated clusters, which might be geographically or competitively segregated.

J. Girvan Newman :

• *Implication :* The Girvan-Newman algorithm detects communities by iteratively removing edges with the highest betweenness centrality. This helps identify tightly connected subgroups of teams, revealing hidden community structures in the football network that are essential for understanding team dynamics and rivalries.

• *Key Insights* : The results from applying the Girvan-Newman algorithm show several strong community structures within the football network, each corresponding to conferences. These communities suggest that the teams within each conference are more likely to have intense interactions and shared characteristics compared to teams in other conferences.

• *Result Table :*

Community	Teams
Community 1	OklahomaState, Nebraska, AirForce, Northwestern, WashingtonState, SanJoseState, FresnoState, OhioState, NotreDame, Oklahoma, ColoradoState, ArkansasState, PennState, Rice, NewMexicoState, Stanford, Iowa, California, TexasElPaso, Wyoming, Tulsa, IowaState, NevadaLasVegas, MichiganState, Kansas, Missouri, Texas, UtahState, Minnesota, SanDiegoState, Michigan, KansasState, Nevada, ArizonaState, UCLA, Baylor, NorthTexas, SouthernMethodist, TexasA&M, NewMexico, OregonState, Idaho, Hawaii, Wisconsin, Washington, Purdue, Indiana, TexasChristian, Colorado, Utah, BoiseState, Arizona, BrighamYoung, SouthernCalifornia, Oregon, TexasTech, Illinois
Community 2	Florida, Army, MiddleTennesseeState, LouisianaLafayette, WestVirginia, BallState, WesternMichigan, BowlingGreenState, Temple, NorthCarolina, WakeForest, GeorgiaTech, FloridaState, Duke, VirginiaTech, BostonCollege, LouisianaState, Rutgers, Alabama, Ohio, Auburn, Mississippi, NorthCarolinaState, AlabamaBirmingham, SouthernMississippi, Maryland, Arkansas, Buffalo, Houston, Pittsburgh, Kent, Cincinnati, Navy, Marshall, EasternMichigan, CentralMichigan, LouisianaTech, Tulane, Louisville, NorthernIllinois, MiamiFlorida, MiamiOhio, Toledo, Tennessee, Akron, Connecticut, Vanderbilt, Kentucky, Clemson, SouthCarolina, EastCarolina, MississippiState, Georgia, CentralFlorida, LouisianaMonroe, Virginia, Syracuse, Memphis

figure 7

K. Louvian Method:

• *Implication :* The Louvain method detects communities by maximizing modularity, identifying densely connected groups in the dataset. This approach helps uncover natural clusters within the network, providing valuable insights into group dynamics and structure. By applying the Louvain method, the study can highlight hidden patterns and relationships among nodes.

• *Key Insights* : The Louvain method revealed distinct communities in the dataset, with high modularity values indicating well-formed clusters. These communities represent teams or entities that are more closely interconnected, suggesting strong associations or interactions within the groups. The method efficiently identifies non-overlapping communities, which can be used for further analysis of network behaviors.

• *Visualization* : The results of the Louvain method were visualized using modularity plots and network graphs, showing clearly separated clusters. Nodes within each community are color-coded to highlight the community structure, and edge thickness varies based on the strength of connections, making the network's modularity and groupings visually apparent.

L. K-Clique :

• Total number of 3-cliques found: 64

• Implication : K-Clique analysis identifies tightly-knit

subgroups within the network where each node is directly connected to every other node in the group. This helps uncover hidden communities or clusters in the data that may have significant structural implications.

• *Key Insights* : The K-Clique algorithm reveals cohesive clusters of teams that are highly interrelated within the network. This clustering can provide insights into potential rivalries, alliances, or common characteristics shared among the teams, which could be leveraged for deeper analysis.

M. K-Clan :

• *Implication* : The K-clan method identifies tightly-knit groups of nodes that exhibit strong internal connectivity, providing a means to analyze substructures within the network. This helps in understanding how clusters of related nodes interact in larger systems. These identified clusters can reveal hidden patterns or communities that are significant for network dynamics.

• *Key Insights* : K-clans offer a clear picture of network cohesion by detecting overlapping communities with dense internal connections. These insights are valuable for discovering areas of high influence within the network, showing how certain nodes act as critical connectors or hubs within their respective groups. This can also assist in detecting vulnerable or highly-interconnected areas in the network.

N. Barabasi-Albert Model :

• Barabasi-Albert Graph properties : Average Clustering coefficient : 0.0864

Average Path Length : 3.0828

• *Implication* : The Barabási-Albert (BA) model illustrates the preferential attachment mechanism, where new nodes tend to connect to high-degree nodes. This results in the formation of a scale-free network with hubs or highly connected nodes. It emphasizes how real-world networks evolve, with a few nodes having disproportionately high degrees, reflecting the uneven distribution of connections.

• *Key Insights* : The BA model highlights the emergence of power-law distributions in real-world networks. It demonstrates that highly connected nodes attract more connections over time, leading to network robustness against random node removal but vulnerability to targeted attacks on hub nodes. This insight is essential for understanding the structure of social, technological, and biological networks.

• *Visualization* : Through visualizing adds a small number of nodes accumulate the majority of connections. The network's degree distribution follows a power law, with a long tail showing a significant number of nodes with fewer connections. The graph visually confirms the presence of a scale-free topology typical of many natural and artificial networks.

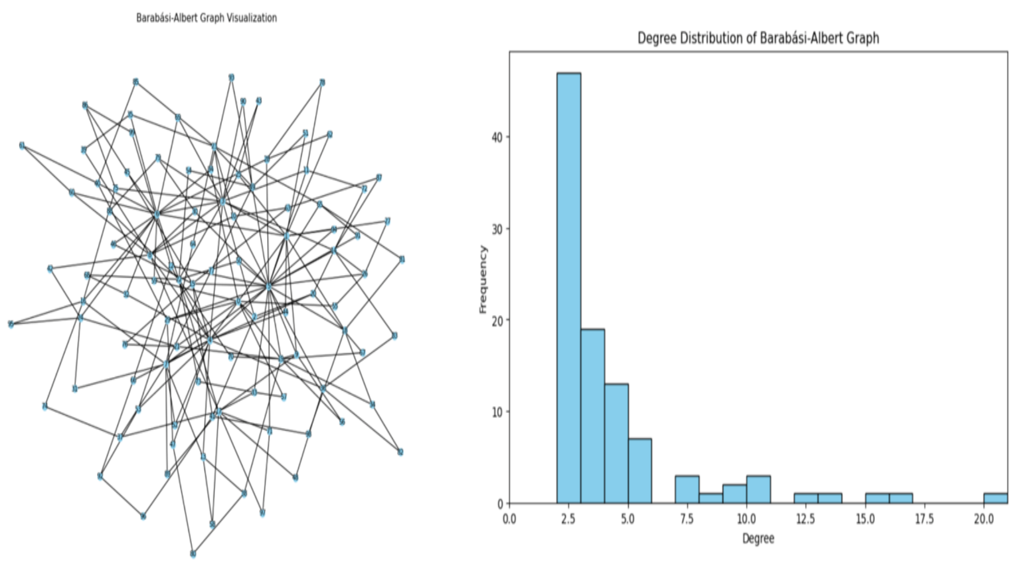

figure 8

O. Watts-Strogatz Model :

• Watts-Strogatz Graph properties : Average Clustering coefficient : 0.3840

Average Path Length : 4.9327

• *Implication :* The Watts-Strogatz model reveals the small-world properties in the network, indicating that despite a sparse connection structure, most nodes are reachable with a small number of steps. This suggests that the network may exhibit efficient communication patterns, which is essential in understanding social networks or collaborative systems.

• *Key Insights* : The model demonstrates the balance between randomness and regularity in network structure. By introducing rewiring, the Watts-Strogatz model creates a network with a high clustering coefficient and short average path length, simulating real-world social networks and technological infrastructures.

• *Visualization* : A visualization of the Watts-Strogatz model would typically show a network with highly clustered nodes interconnected with a few long-range connections, making it visually distinct from both random and regular networks. The plot would highlight the network's small-world characteristics, including high clustering and low path length.

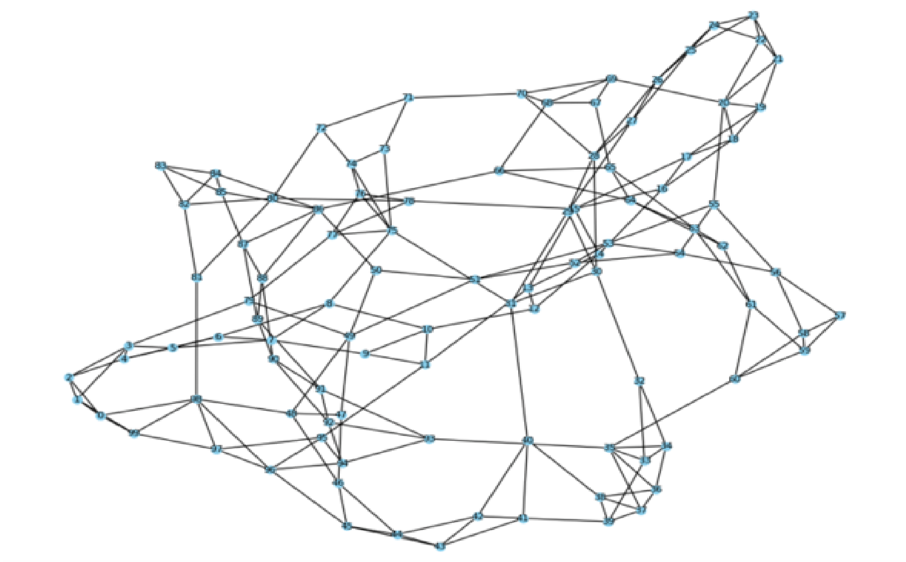

figure 9

P. Edge Betweenness :

• *Implication* : Edge betweenness centrality measures the importance of an edge based on how frequently it appears on the shortest paths between pairs of nodes. High values suggest edges that play a critical role in connecting different parts of the network. Identifying such edges helps to understand key links that influence network connectivity and stability.

• *Key Insights* : The edges with the highest betweenness centrality indicate critical connections between communities or clusters within the network. These edges facilitate the flow of information or resources, and their removal could significantly disrupt the network structure. Understanding these edges is crucial for network resilience and control strategies.

Q. BFS :

• *Implication* : Breadth-First Search (BFS) helps explore the network's structure level by level, ensuring that the shortest path from a starting node to all other reachable nodes is found. This is crucial for identifying the most efficient traversal path and understanding connectivity in the dataset.

• *Key Insights* : BFS provides insights into the spread of influence or information in the network, showing the nearest neighbors of nodes. It can reveal how nodes are clustered based on proximity and helps identify isolated or well-connected clusters in the dataset.

R. DFS :

• *Implication* : Depth-First Search (DFS) is a fundamental graph traversal technique that helps explore all possible paths from a given starting node. In the context of our dataset, DFS can be used to explore the relationships between teams, identify clusters of connected teams, and uncover hidden structures within the network, such as isolated or central nodes. Its efficiency in exploring large networks makes it suitable for analyzing the structure of football-related connections.

• *Key Insights* : DFS allows us to identify how well-connected different teams are, revealing patterns like isolated teams or groups of teams that are strongly connected. It also helps in detecting cycles within the network, providing deeper insights into repeated interactions or dependencies between specific teams .

IV. CONCLUSION :

- **Key Findings** :

The analysis of the football dataset using network science techniques has provided valuable insights into the structure and dynamics of the teams and their connections. Through centrality measures such as Degree Centrality, Betweenness Centrality, and Eigenvector Centrality, we identified the most influential teams within the network. Degree Centrality highlighted teams with the highest number of direct connections, indicating their immediate influence in the network, while Betweenness Centrality provided insights into teams that serve as critical bridges, facilitating information flow between other teams. Eigenvector Centrality further refined our understanding by revealing teams with influence that extends through the network, not just locally, but through their connections to other well-connected teams. Another significant finding was the application of clustering coefficients, both local and global, to analyze the cohesiveness within the network. Local Clustering Coefficients revealed the degree of clustering around individual teams, indicating how likely they were to form tight-knit groups with other teams, while the Global Clustering Coefficient measured the overall tendency of the network to form clusters. These clustering properties suggested that football teams tend to form strong communities, with some teams acting as key hubs within these groups.Additionally, community detection algorithms such as the Girvan-Newman method and Louvain method revealed the modular structure of the network, identifying groups of teams that are more tightly connected to each other than to teams outside their community. This modularity analysis underscored the existence of regional or league-based communities within the network, which can have implications for rivalry dynamics, team performance, and fan engagement. The Degree Distribution analysis and Cumulative Degree Distribution helped characterize the scale-free nature of the network, emphasizing the presence of a few highly connected teams (hubs) and many teams with fewer connections. This type of distribution, typical in real-world networks, suggests that the football network follows a ”preferential attachment” model, where new teams are more likely to connect to already well-connected teams. Finally, the application of graph traversal techniques, including Breadth-First Search (BFS) and Depth-First Search (DFS), provided further insights into the connectivity and reachability of teams in the network. These methods revealed the paths through which information or influence might spread across teams, helping to visualize the overall connectedness of the football network.

- **Impact and Importance of the Analysis**:

The analysis of the football dataset using network science techniques offers valuable insights into the complex relationships between teams in collegiate football. By applying centrality measures, community detection algorithms, and graph metrics, the study highlights how certain teams dominate the network, either through direct connections or by serving as bridges between other teams. This understanding is crucial for sports analysts, coaches, and decision-makers as it enables them to identify influential teams and predict potential outcomes based on their positioning within the network. The study also emphasizes the importance of network topology in understanding team interactions, with metrics like clustering coefficients and shortest path lengths revealing the degree of connectivity and cohesion among teams. By recognizing tightly-knit communities or isolated teams, stakeholders can tailor strategies for scheduling, recruitment, and collaboration. Furthermore, the research provides a robust framework for examining other sports and industries where relationships between entities play a central role. Whether in entertainment, business, or social networks, the ability to understand structural properties, central players, and community formation enhances the decision-making process. This study exemplifies the power of network science in transforming raw data into actionable insights that drive performance optimization and strategic planning in interconnected domains.

- **Future Directions and Recommendations**:

Future research could focus on exploring the temporal dynamics of the football network, examining how team relationships evolve over time, influenced by factors such as player transfers and performance changes. Incorporating player-level data would provide a deeper understanding of individual contributions to team dynamics. Additionally, investigating external factors like financial power or media influence could reveal new insights into network formation. Advanced techniques, such as machine learning for community detection or prediction, could further refine analysis. Extending these methodologies to other sports or industries may offer broader applications of network analysis in dynamic, interconnected environments.

V. Reference:

References

[1] Davis, A., Gardner, B. B., & Gardner, M. R. (1941). *The social structure of small urban communities*. Retrieved from https://psycnet.apa.org/record/1942-00249-000

[2] Davis Southern Women Dataset. Retrieved from https://networkdata.ics.uci.edu/netdata/html/davis.html

[3] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.
[4] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440-442.
[5] Newman, M. E. (2010). *Networks: An Introduction*. Oxford University Press.
[6] Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821-7826.
[7] Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4), 375-395.
[8] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35-41.
[9] Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3), 1059-1069.
[10] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 29-42.
[11] Ravasz, E., & Barabási, A. L. (2003). Hierarchical organization in complex networks. *Physical Review E*, 67(2), 026112.