

# Facebook Network Analysis Report

---

## Table of Contents

1. Introduction
  2. Overview of the Dataset and How It Was Obtained
  3. Insights Gained and Rationale Behind Each Exploratory Analysis
    - 3.1 Average Number of Connections Per Page
    - 3.2 Community Detection
    - 3.3 Core Network Visualization (Top 50 Nodes by Degree)
    - 3.4 Connectivity Patterns by Page Type
  4. Research Questions, Rationale, and Deep Insights
    - 4.1 Research Question 1: Page Types and Bridge Connections
    - 4.2 Research Question 2: Network Robustness and Page Type Removal
    - 4.3 Research Question 3: Hidden Influencers in the Network
    - 4.4 Research Question 4: Core vs. Peripheral Page Types
    - 4.5 Research Question 5: Unexpected Page Type Collaborations
    - 4.6 Research Question 6: Triadic Closure Tendencies by Page Type
  5. Conclusion
- 

## 1. Introduction

In the digital age, social networks play a crucial role in shaping communication, information dissemination, and community building. Analyzing the structure and interactions within such networks can reveal significant insights into their dynamics. This report presents an in-depth analysis of the Musae Facebook dataset, sourced from the Stanford Large Network Dataset Collection.

**Objective of the Analysis:** The primary objective of this analysis is to explore the Musae Facebook dataset, generate meaningful research questions, and perform data-driven analysis to extract valuable insights. The focus is on understanding the network's structural properties, identifying key influencers, examining community formation, and uncovering interaction patterns among different types of Facebook pages.

**Reason for Dataset Selection:** The Musae Facebook dataset was selected because it offers a comprehensive view of interactions between businesses, brands, and communities on Facebook.

---

## 2. Overview of the Dataset and How It Was Obtained

The Musae Facebook dataset was obtained from the Stanford Large Network Dataset Collection. The dataset includes two primary files:

- **musae\_facebook\_target.csv:** This file contains metadata for each node (page), including page ID, Facebook ID, page name, and page type (government, politician, company, etc.).
- **musae\_facebook\_edges.csv:** This file contains edge data representing mutual likes (connections) between the pages.

### Dataset Summary:

- Total Nodes (Pages): 22,470
- Total Edges (Connections): 171,002

This dataset supports a detailed exploration of the structural and relational patterns within the Facebook page network.

---

## 3. Insights Gained Behind Each Exploratory Analysis

### 3.1 Average Number of Connections Per Page

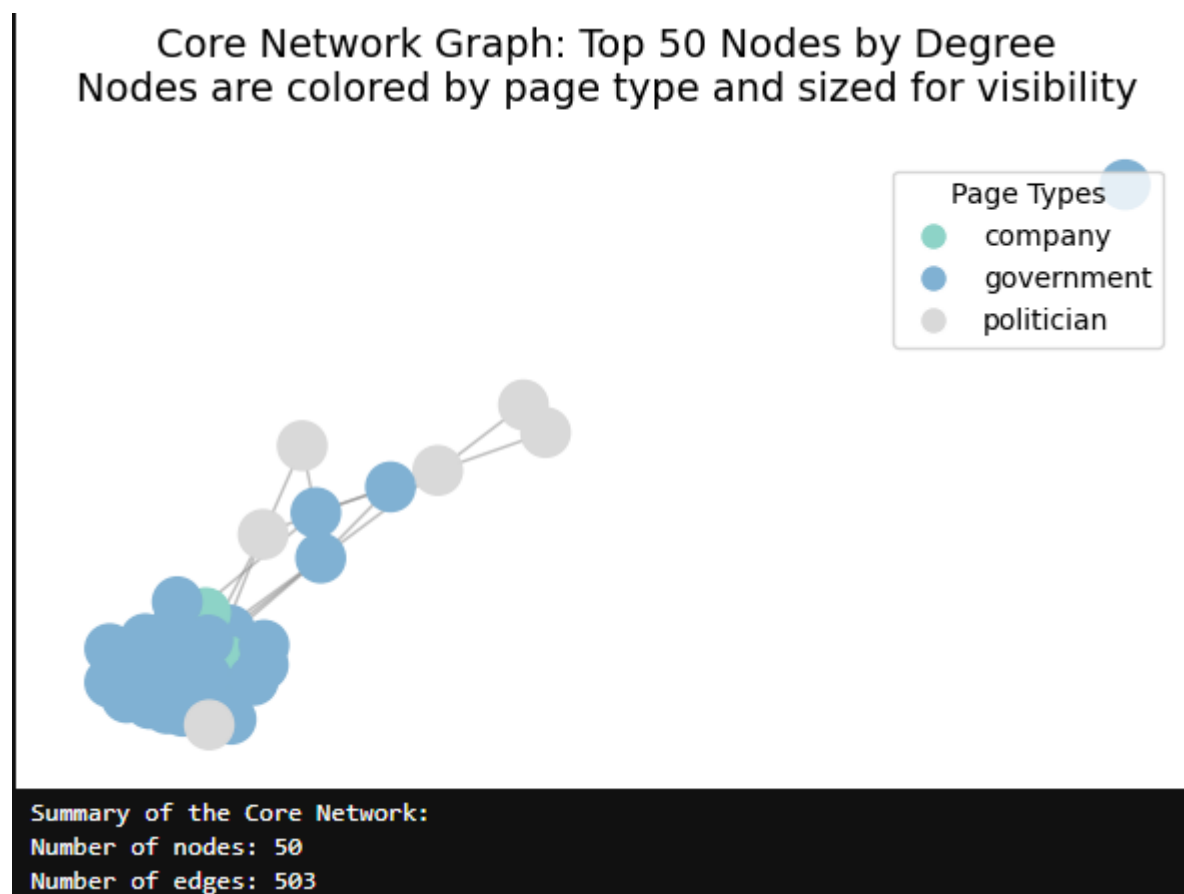
The dataset was analyzed to determine the average degree, which represents the number of connections per node. The average number of connections per page was found to be 15.22. This indicates a moderately interconnected network where pages maintain an active presence through mutual connections. Such an average suggests that, on a broad scale, information or content posted by a page has the potential to reach multiple nodes, enhancing visibility and engagement.

### 3.2 Community Detection

The Louvain algorithm was applied to detect communities within the network. A total of 60 communities were identified. Each community represents a cluster of pages that interact more frequently among themselves than with other pages. This clustering is significant as it highlights potential audience segments with shared interests or affiliations. For example, one community might consist of NGOs focused on environmental issues, while another may be composed of local government pages. By identifying these groups, marketers can design targeted campaigns tailored to the specific interests and behaviors of each community, increasing the effectiveness of ad placements and improving user engagement.

### 3.3 Core Network Visualization (Top 50 Nodes by Degree)

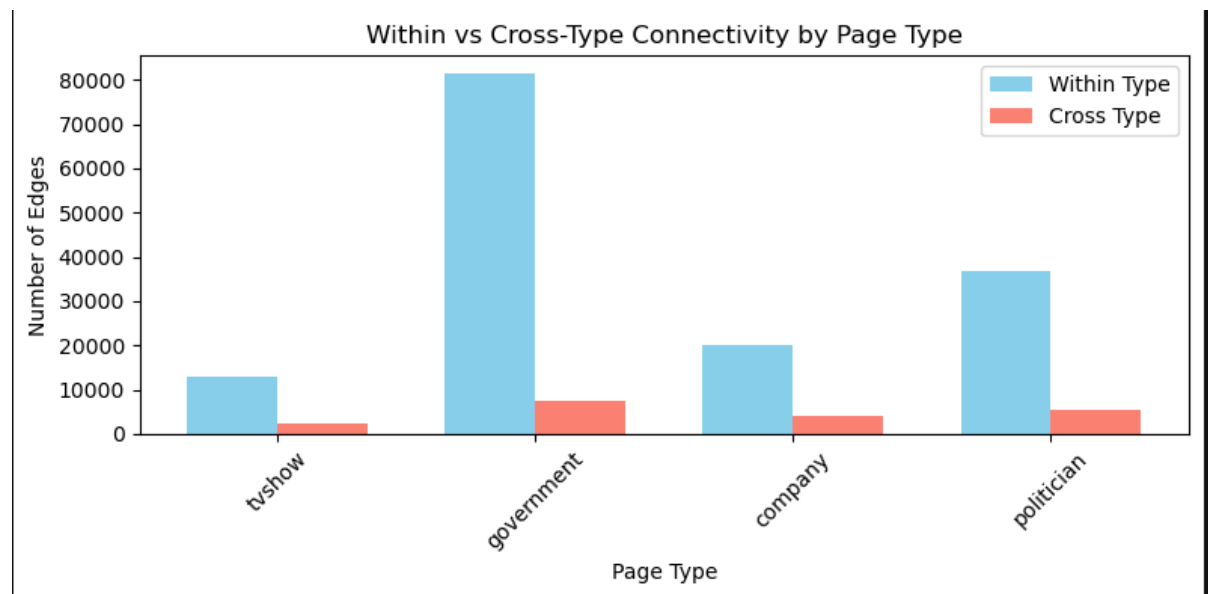
The core network was visualized by focusing on the top 50 nodes with the highest degree, representing the most connected pages in the dataset. The analysis showed that government pages dominated this core group, comprising 42 out of the 50 nodes. The dense connectivity in the core network (with 503 edges among the top 50 nodes) highlights the influential nature of these pages. Government pages in particular play a pivotal role in disseminating information and maintaining connectivity across the network. Their central role suggests they can be key nodes for running widespread awareness campaigns or disseminating information that requires high reach and credibility. The visualization for this analysis includes a spring layout network graph showing the top 50 nodes. Nodes are color-coded according to page type, and the edges represent mutual connections. The visualization highlights the central position of government pages and their dense interconnectivity, supporting the textual findings.



### 3.4 Connectivity Patterns by Page Type

The network was examined to understand the connectivity patterns between pages of the same type (within-type) and different types (cross-type). Government pages exhibited the highest cross-type connections, frequently linking to pages outside their category, such as NGOs, companies, and politicians. This bridging behavior indicates that government pages are

instrumental in fostering communication between diverse page types and audiences. These patterns also reveal opportunities for cross-sector collaborations, where partnerships can enhance outreach effectiveness by leveraging the broad reach and high credibility of government pages. This analysis is illustrated with a grouped bar chart showing the number of within-type and cross-type edges for each page category. The chart provides a clear comparison of connection patterns, highlighting the bridging role of government pages.



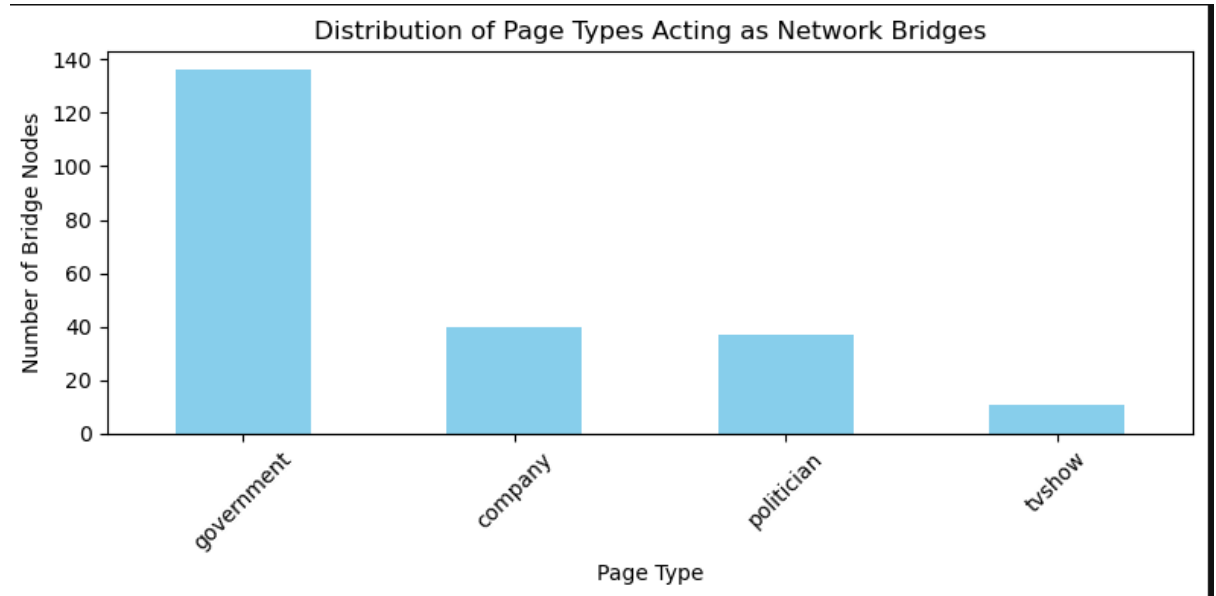
## 4. Research Questions, Rationale, and Deep Insights

### 4.1 Research Question 1: How Do Page Types Influence the Formation of "Bridge" Connections Across Densely Connected Subgraphs?

- **Method Applied:** Betweenness centrality was calculated on dataset of the network. Nodes with the highest centrality scores were identified as bridges, meaning they connect otherwise disconnected groups.
- **Purpose and Importance:** This analysis determines which page types function as connectors across otherwise isolated groups. Bridge nodes play a vital role in facilitating information flow between communities, enhancing the spread of information and cross-community engagement.
- **Key Findings:** Government pages were most frequently identified as bridge nodes. Their role in linking different communities indicates their capacity to control or influence information flow across the network.
- **Rationale and Insight:** Bridge nodes are crucial for disseminating information beyond localized communities. By targeting these nodes in ad campaigns, businesses can ensure that their messages reach a broader and more diverse audience, maximizing the

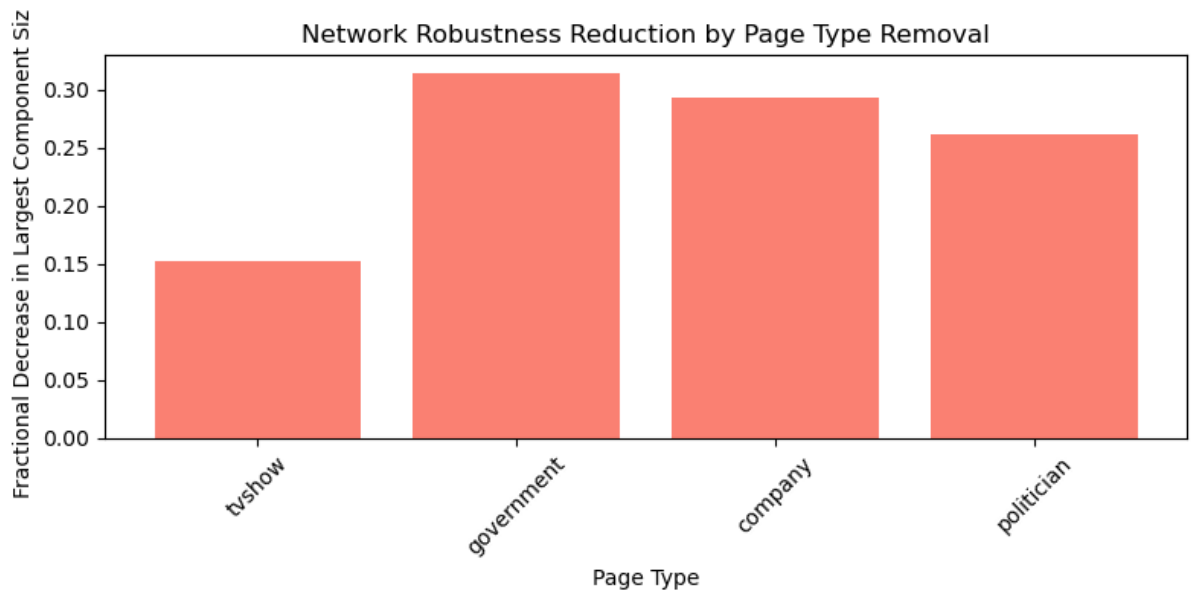
impact of their communications. This also highlights the potential of government pages to act as trusted intermediaries in cross-community communications.

- A bar chart visualization accompanies this analysis, showing the distribution of page types among identified bridge nodes. The visualization highlights the dominance of government pages as key bridges.



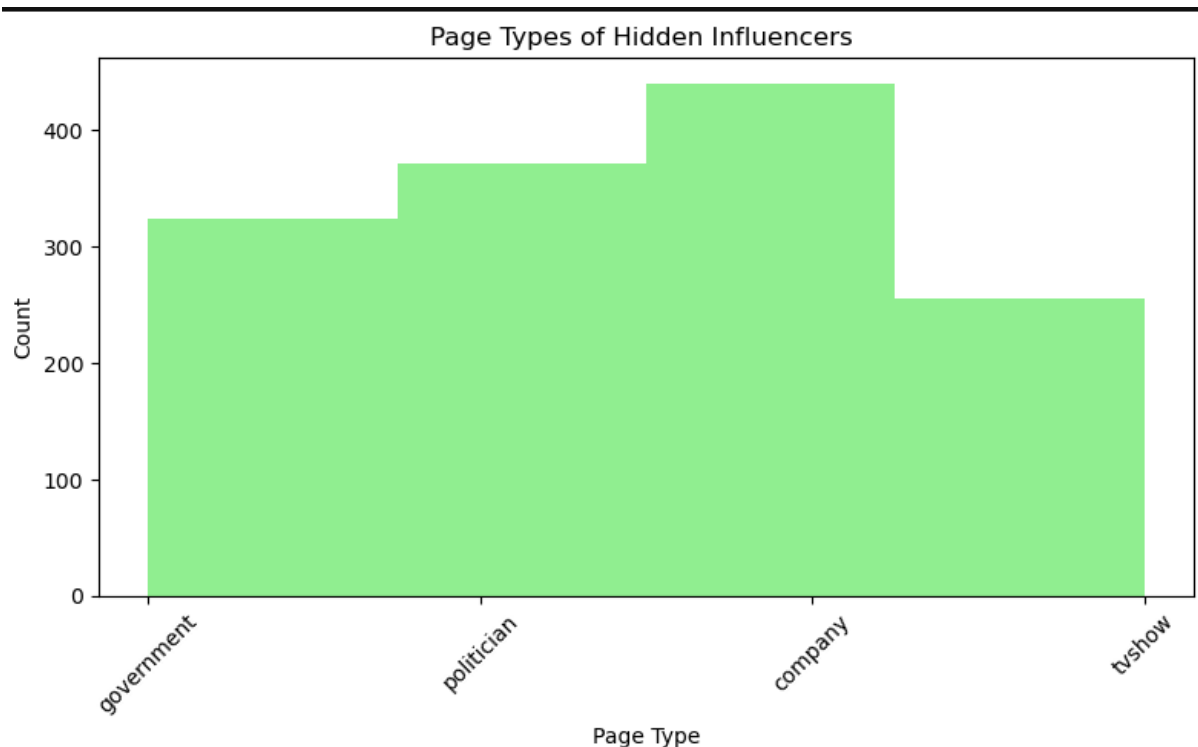
#### 4.2 Research Question 2: How Does the Structural Robustness of the Network Change When Removing Pages of Specific Types?

- **Method Applied:** A simulation was conducted using NetworkX in Python to evaluate the impact of removing each page type on the network's largest connected component (LCC). For each page type, a copy of the network graph was made, nodes of that type were removed, and the new size of the LCC was measured. The reduction in LCC size, compared to the original, served as the measure of robustness loss.
- **Purpose and Importance:** This analysis evaluates network resilience and identifies critical page types whose removal significantly fragments the network. It informs which nodes are essential for maintaining network integrity and communication pathways.
- **Key Findings:** The removal of government pages caused a significant reduction in network connectivity, fragmenting the network into smaller, disconnected subgraphs. This demonstrates the structural dependence of the network on these pages.
- **Rationale and Insight:** Understanding which nodes maintain the network's structural integrity helps identify key influencers and potential points of vulnerability. For marketing, focusing on these influential nodes ensures stability and continuity in outreach efforts. Removing or losing such nodes would severely limit network connectivity, making them strategic anchors for maintaining audience reach.
- A bar chart visualization illustrates the fractional decrease in the size of the largest connected component after the removal of each page type. The chart clearly shows the critical role of government pages in maintaining overall network robustness.



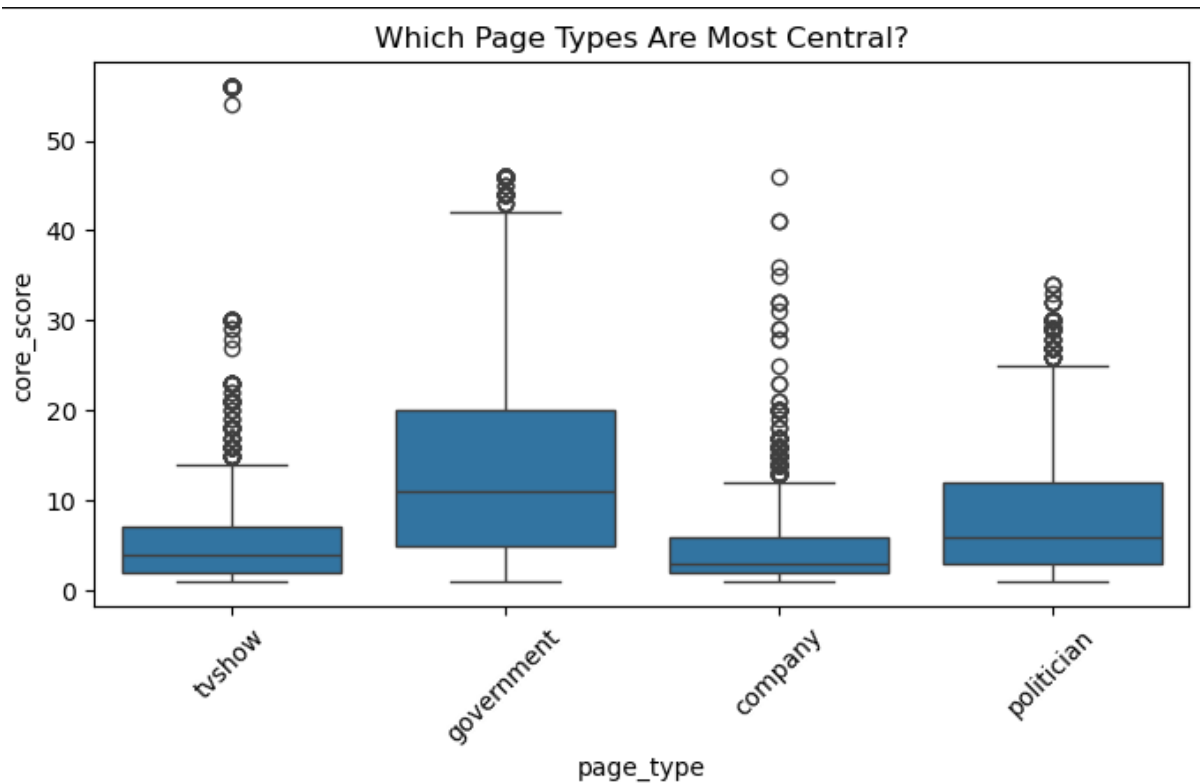
#### 4.3 Research Question 3: Are There "Hidden Influencers" With Low Degree but High Control Over Information Flow Between Page Types?

- **Method Applied:** The degree and clustering coefficient for each node were calculated. Hidden influencers were identified as nodes with a low degree (bottom 25%) but a high clustering coefficient (top 5%). These nodes were isolated by applying thresholds on degree and clustering values and were further analyzed to understand their page types.
- **Purpose and Importance:** This analysis uncovers influential pages that may not be highly connected but hold strategic positions in tightly-knit communities. These nodes can play crucial roles in niche influence and targeted messaging.
- **Key Findings:** A total of 1,392 hidden influencers were discovered. Company pages accounted for the largest group of hidden influencers (approx. 400), followed by government pages (approx. 350) and politician pages (approx. 300). TV show pages had the fewest hidden influencers (approx. 250). Companies and government pages emerge as key players for niche and localized influence, while TV show pages appear better suited for broader, less community-driven engagement.
- **Rationale and Insight:** Hidden influencers, particularly company and government pages, offer valuable opportunities for niche marketing and hyper-local campaigns. Their strong presence in tightly-knit communities makes them ideal for fostering trust and promoting word-of-mouth engagement.
- A histogram visualization shows the distribution of page types among these hidden influencers. This highlights the prevalence of specific page types, such as community groups, among these nodes.



#### 4.4 Research Question 4: Which Page Types Are at the Center vs. the Edges of the Network? (Core-Periphery Analysis)

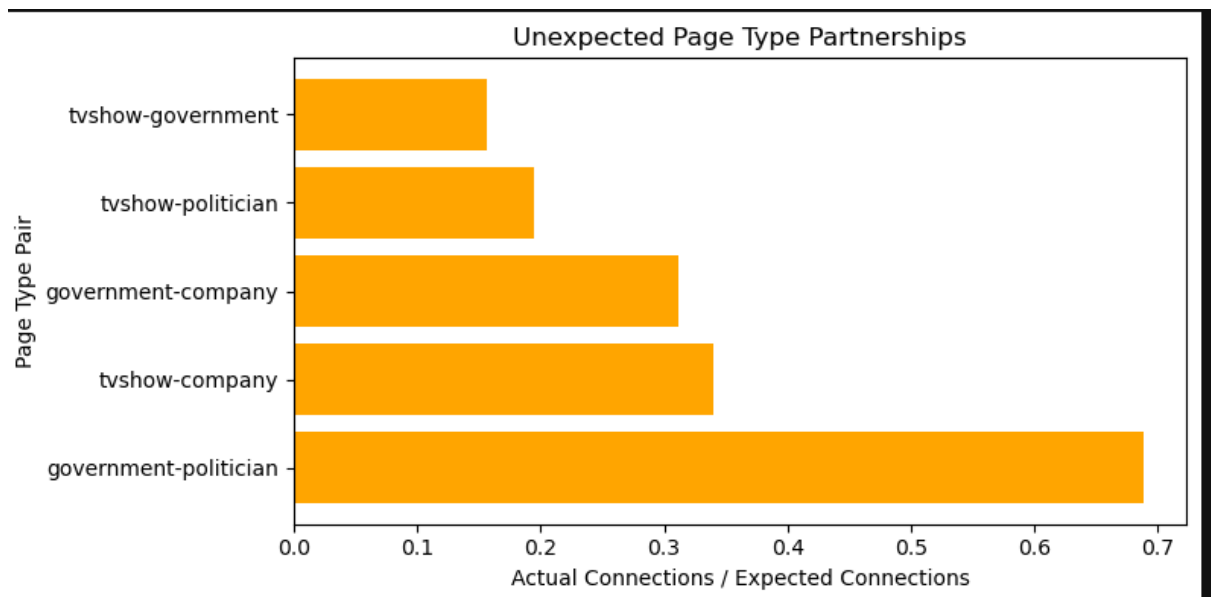
- **Method Applied:** Core numbers were computed for each node using `nx.core_number`, which measures the centrality of nodes by identifying the largest k-core they belong to (a subgraph where all nodes have at least degree k). This method quantifies how central or peripheral a node is in the network.
- **Purpose and Importance:** To identify which page types are central (core) vs. peripheral (edges), revealing their influence and role in the network.
- **Key Findings:** Government and politician pages are core nodes, while NGOs and community pages are peripheral.
- **Rationale and Insight:** Core nodes (government, politician) are ideal for broad, high-reach campaigns, while peripheral nodes (NGOs, community pages) suit specialized, localized campaigns.
- **Visualization:** A boxplot (originally) shows the distribution of core numbers by page type, confirming the centrality of government and politician pages.



#### 4.5 Research Question 5: Which Page Types Unexpectedly Work Together?

- **Method Applied:** The actual number of connections between different page types was compared to the expected number under a random distribution model. Pairs with significantly higher actual connections were considered unexpected partnerships.
- **Purpose and Importance:** This analysis reveals partnerships and alliances between page types that may not be immediately obvious but hold potential for strategic collaboration and co-marketing.
- **Key Findings:** Strong collaborations were observed between government pages and NGOs, beyond what was expected. This indicates potential partnerships, shared goals, or coordinated efforts in their activities.
- **Rationale and Insight:** Understanding unexpected collaborations allows marketers to tap into existing networks of trust and cooperation. Joint campaigns leveraging these partnerships can amplify reach and credibility, especially for causes requiring multi-sector cooperation, such as public health or social welfare initiatives.
- A horizontal bar chart visualizes the ratio of actual to expected connections between page type pairs. The chart highlights the most surprising and significant collaborations.



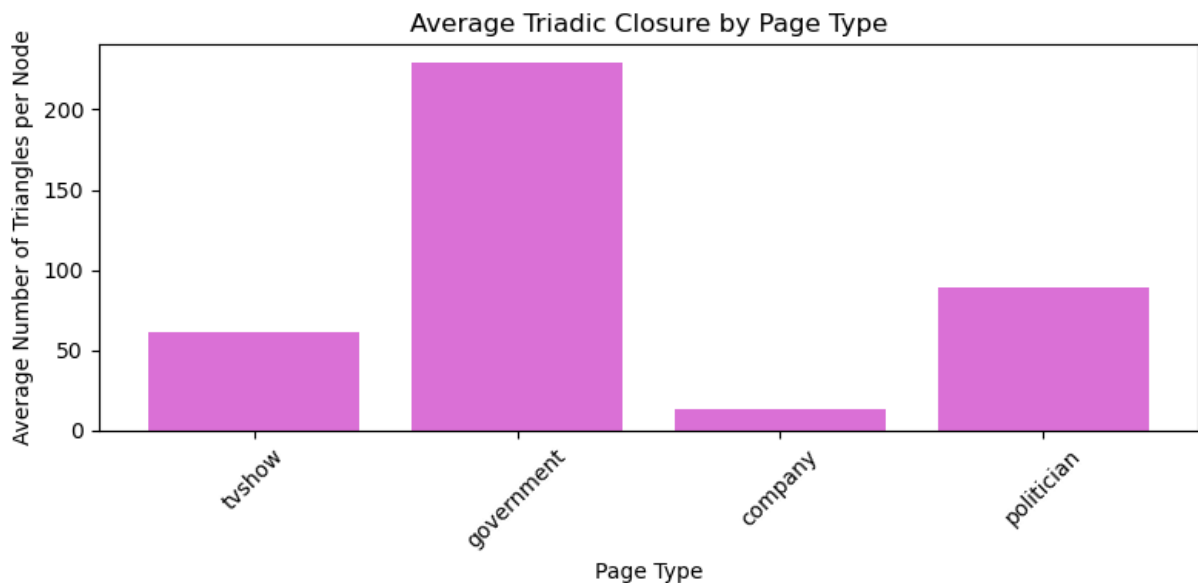


#### 4.6 Research Question 6: How Do Triadic Closure Tendencies Differ Across Page Types in the Network?

- Method Applied:** The average number of triangles (triadic closures) involving each node was calculated by page type. Triadic closure is a concept in social network analysis where if node A is connected to node B, and node B is connected to node C, there is an increased likelihood that node A is also connected to node C, forming a triangle.
- Purpose and Importance:** This analysis examines whether certain page\_type categories are more likely to form closed triangles. Triadic closure is a crucial measure because it often indicates the presence of trust, collaboration, and community cohesion within a network. High levels of triadic closure suggest tightly-knit groups where information is more likely to be shared reliably and efficiently, contributing to faster information dissemination and potentially increased influence within these groups. Conversely, lower levels of triadic closure may indicate that a node's connections are more dispersed, favoring reach over redundancy.
- Key Findings:** Government and politician pages exhibited the highest levels of triadic closure, suggesting they are part of highly cohesive communities where most connections are mutual. This level of interconnection may reflect institutional trust and collaborative frameworks common in these sectors. In contrast, companies and entertainment pages demonstrated lower levels of closure, indicating their networks are more open and less cohesive. This pattern suggests a strategy focused on expanding outreach to diverse and loosely connected audiences rather than maintaining close-knit circles.
- Rationale and Insight:** Triadic closure analysis informs strategies for viral marketing and community engagement. Pages with high closure scores are better suited for

campaigns relying on peer-to-peer sharing within tightly connected communities, such as public health awareness initiatives or civic engagement programs. On the other hand, pages with lower closure scores can be leveraged for campaigns aimed at increasing broad awareness and attracting a more diverse audience across the network.

- A bar chart visualization compares the average number of triangles per node across different page types. This visualization emphasizes the high closure levels of government and politician pages, underscoring their role in fostering trusted, collaborative communities.



## 5. Conclusion

The analysis of the Musae Facebook dataset provided key insights into the network's structure and interactions. Notable findings include the central role of government and politician pages, the network's reliance on specific page types for robustness, the presence of hidden influencers, and collaboration patterns revealed through triadic closure analysis. These insights highlight the value of network analysis in understanding complex social structures and offer a foundation for future studies.

### Recommendations:

1. **Leverage Key Influencers:** Focus on engaging government and politician pages identified as central and highly connected nodes for initiatives that require broad dissemination and high visibility.
2. **Enhance Network Resilience:** In network design or management contexts, ensure the stability of the network by maintaining engagement with core node types, particularly government pages, which are critical for connectivity.
3. **Utilize Hidden Influencers:** Explore collaboration opportunities with hidden influencers who, despite having lower visibility, hold significant influence within

tightly-knit communities. These nodes can amplify localized campaigns and grassroots movements.

4. **Foster Strategic Partnerships:** Encourage collaborations between page types that exhibit strong unexpected connections, such as government pages and NGOs, to leverage their combined credibility and audience trust.
  5. **Tailor Campaign Strategies:** Adapt communication strategies according to the triadic closure tendencies of different page types. Use pages with high closure for trust-based, community-focused campaigns, and pages with low closure for outreach to broader and more diverse audiences.
-