

# The Interplay Of Information: Exploring References Across Media Platforms

By Sagnik Chand, M.Sc. Data Analytics & Computational Social Science



University of  
Massachusetts  
Amherst

## Abstract

In today's information-saturated landscape, comprehending communication networks across diverse media platforms is imperative. Our study employs network analysis techniques and statistical methods to investigate interaction patterns within information dissemination networks. Utilizing a robust dataset from Kaggle's network analysis collection, we scrutinize the relationships among TV channels, newspapers, and online articles.

The dataset comprises two files: an edge list and a node list. It encompasses 17 distinct information networks, serving as nodes, alongside attributes such as media type, audience size, and edge weights. Our analysis entails exploring centrality measures, community detection algorithms, and descriptive statistics to identify influential nodes and clusters within the network.

Through visualizations and hypothesis testing, we unravel the intricate dynamics between media types and interaction patterns. While affirming the emergence of distinct clusters based on shared properties, our findings challenge assumptions regarding the correlation between audience size and network centrality.

Ultimately, our study contributes to a nuanced understanding of information flow across media platforms, with implications for refining information dissemination strategies and policymaking.

## Introduction

In the contemporary era characterized by information saturation, comprehending the intricacies of communication networks among diverse information sources holds significant importance. Our research endeavors to decode these dynamics by leveraging a comprehensive dataset encompassing TV, Newspapers, and online articles. Through the lens of network analysis, we aim to unravel the underlying mechanisms shaping interactions and references across these media platforms.

### Research Question:

What are the communication patterns among various information sources, and how do these interactions influence the flow of information across different media platforms?

To address this question, we delve into the rich tapestry of data to analyze patterns and structures emerging from interactions and references.

### Hypothesis:

- H1:** Information sources with a larger audience size are likely to have more connections or references with other media platforms.
- H2:** Distinct clusters of information sources may emerge based on the type of media and the nature of their interactions (e.g., hyperlinks or mentions).

Through our investigation, we seek to gain valuable insights into the dynamics of information flow and collaboration across media landscapes, offering implications for understanding and navigating the complex network of information dissemination in the digital age.

## Dataset

Our datasets, downloaded from Kaggle as part of a rich network analysis collection, comprise:

- Edge List:** This dataset details relationships between nodes, including source and target node IDs, connection frequency (weight), and link type (hyperlink or mention).
- Node List:** Providing node-specific information such as unique IDs, media representation (e.g., NY Times, Washington Post), media type (Newspaper, TV, Online), media type label, and audience size.

	from <chr>	to <chr>	weight <dbl>	type <chr>
1	s01	s02	10	hyperlink
2	s01	s02	12	hyperlink
3	s01	s03	22	hyperlink
4	s01	s04	21	hyperlink
5	s04	s11	22	mention
6	s05	s15	21	mention

Fig. Edgelist

id <chr>	media <chr>	media.type <chr>	type.label <chr>	audience.size <dbl>
1	s01	NY Times	1 Newspaper	20
2	s02	Washington Post	1 Newspaper	25
3	s03	Wall Street Journal	1 Newspaper	30
4	s04	USA Today	1 Newspaper	32
5	s05	LA Times	1 Newspaper	20
6	s06	New York Post	1 Newspaper	50

Fig. Nodelist

Facing diverse challenges, we prioritize the task of identifying meaningful communication patterns. Our objective is to extract actionable insights from the intricate web of inter-media interactions, unveiling significant communication patterns amidst complexity.

## Methodology

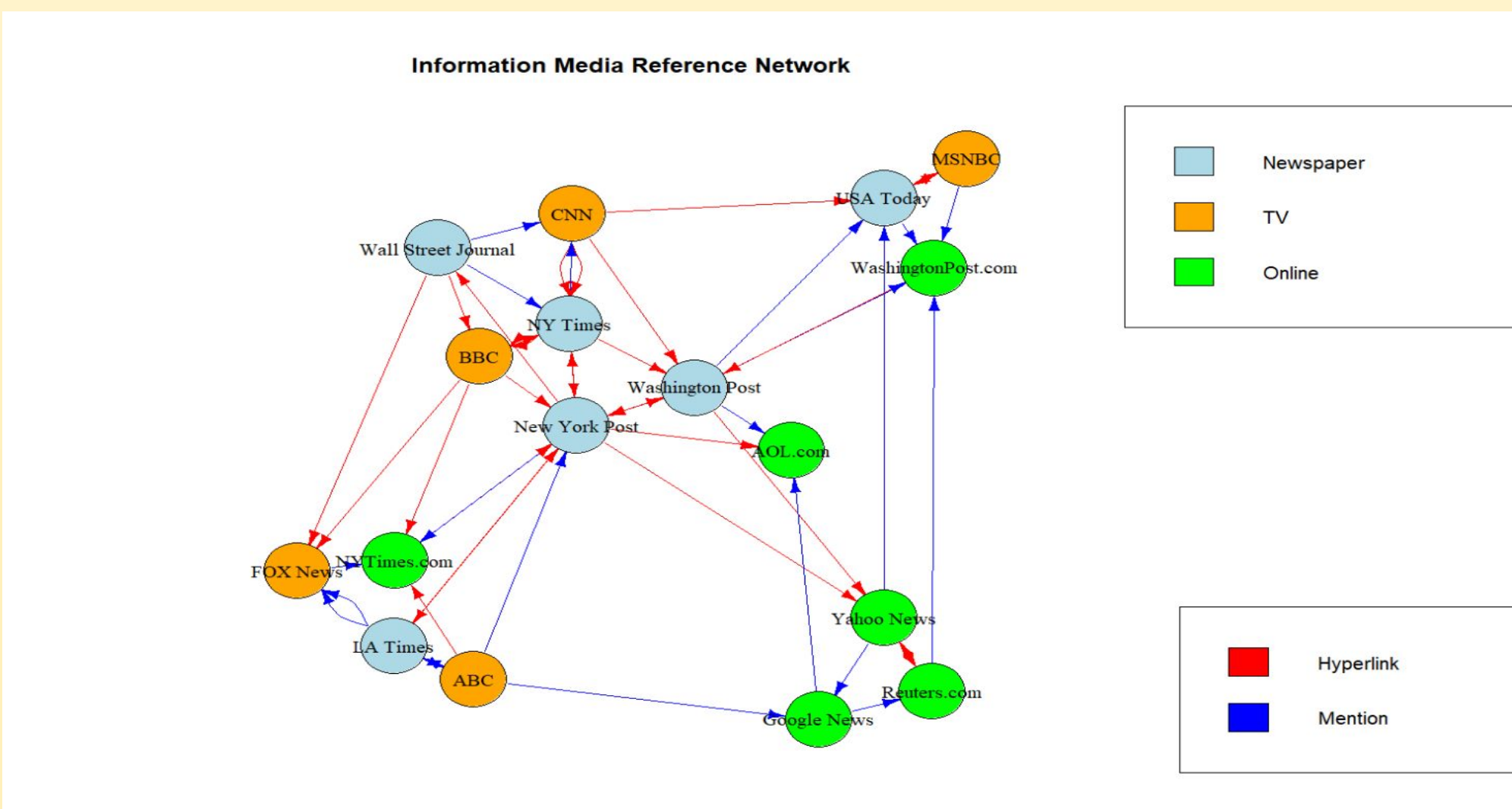
In our analysis, we utilize a blend of network analysis techniques and statistical methods. Specifically:

- Network Analysis:** We employ centrality measures and community detection algorithms to pinpoint influential information sources and unveil clusters of media platforms sharing communication patterns.
- Graph Description Techniques:** Various techniques are applied at both node and graph levels to comprehend our network object thoroughly.
- Tools and Packages:** We leverage igraph and statnet packages, alongside tools like tidyverse, to facilitate our analysis.
- Methodological Approaches:**
  - Node Centrality and Degree:** These metrics aid in exploring our first hypothesis by identifying influential information sources within the network.
  - Community Detection:** This approach helps us investigate our second hypothesis by revealing distinct clusters of media platforms with similar communication patterns.
- Advanced Visualization:** We employ sophisticated visualization techniques to interpret and effectively communicate our findings.

## Results

### Network Description:

- The dataset comprises 17 nodes representing various media platforms, connected by 51 directed ties or edges.
- The dataset includes two edge attributes: weights and types of ties.
- The global transitivity score stands at 0.37, suggesting 37.5% presence of all three triad ties among nodes.



### Network Degree & Centrality Analysis:

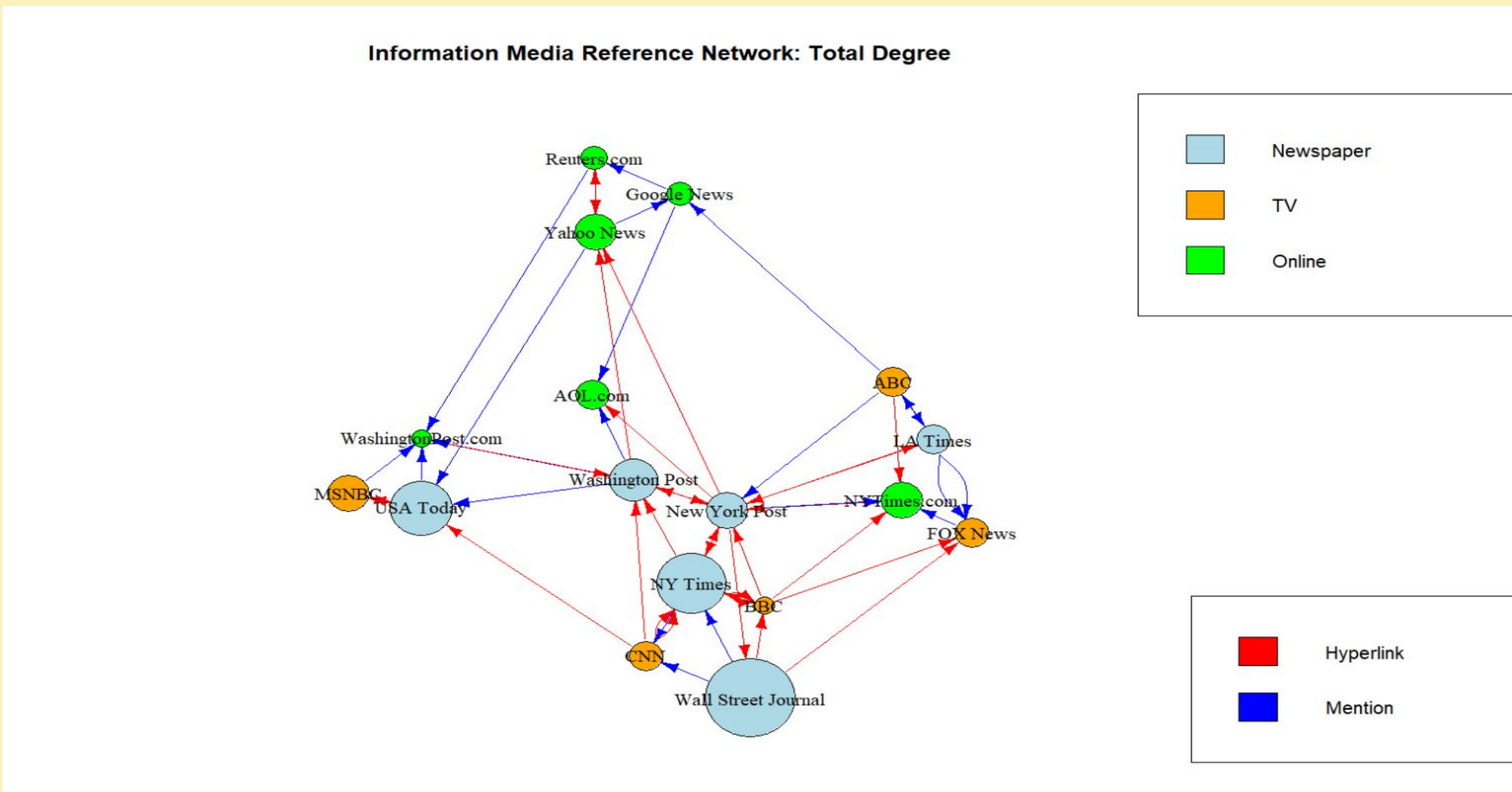
Centrality metrics serve as key indicators, shedding light on influential nodes driving network dynamics:

- Degree Analysis:**
  - Total Degree:** Wall Street Journal, NY Times, and USA Today lead with highest total degrees, signifying their influence.
  - Out-Degree:** These newspapers exhibit strong referencing activity, with Wall Street Journal topping with 7 links.
  - In-Degree:** Wall Street Journal and NY Times are most referenced, suggesting their pivotal roles in information dissemination.

Degree & Centrality analysis unveils nuanced network dynamics, where influence transcends audience size. Platforms like Wall Street Journal exemplify potent influence.

Name <chr>	Degree <dbl>	In-Degree <dbl>	Out-Degree <dbl>
1 Wall Street Journal	13	6	7
2 NY Times	10	5	5
3 USA Today	9	4	5
4 Washington Post	7	3	4
5 New York Post	6	4	2
6 MSNBC	6	2	4

Fig. Node Degree Dataframe

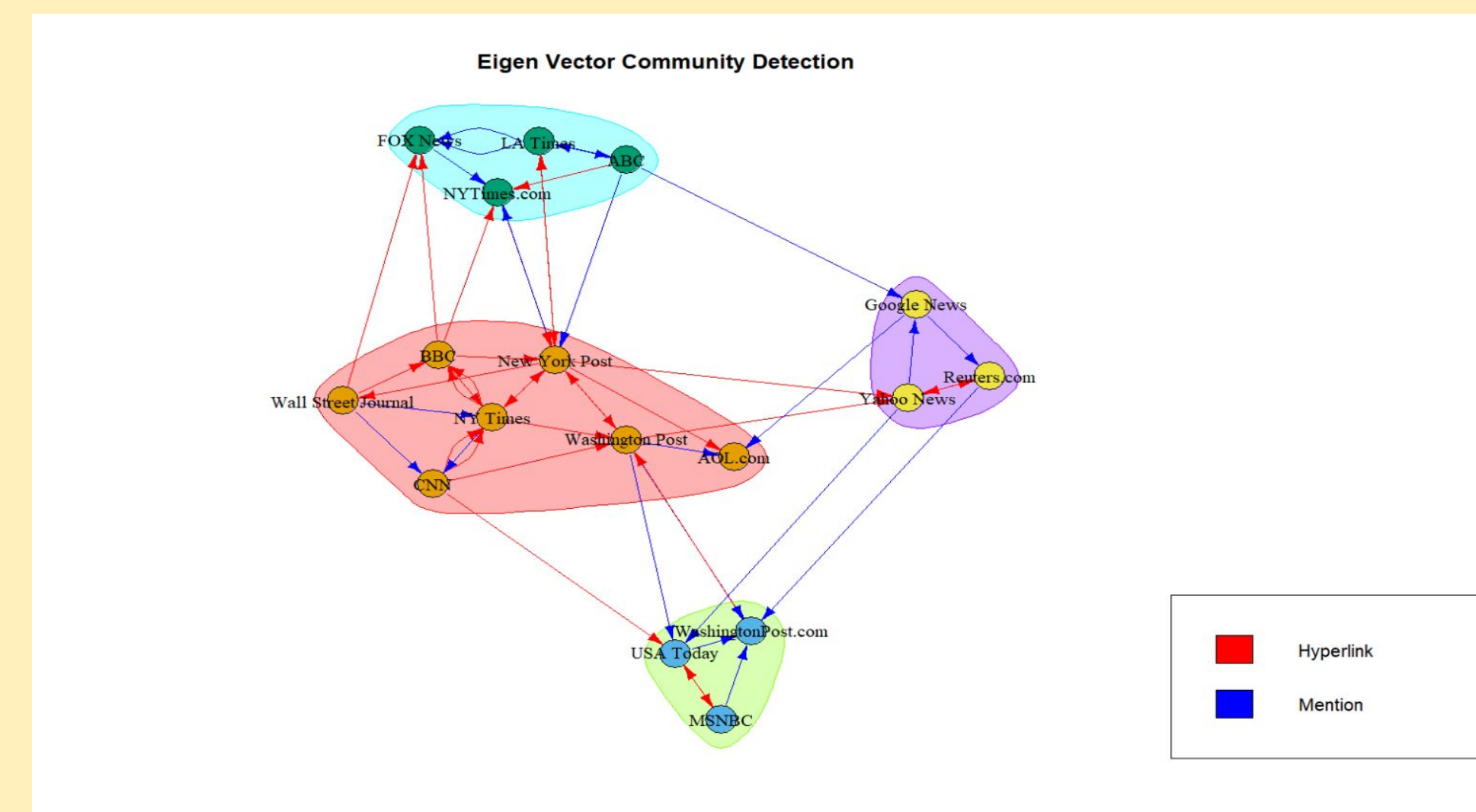


### Community Detection Analysis:

Community detection algorithms are instrumental in uncovering clusters or communities within networks, revealing nodes with similar attributes or behaviors. We employed three different algorithms.

- Walktrap Algorithm:**
  - Detects dense subgraphs via random walks, considering edge weights for weighted networks.
  - Identified a diverse community comprising newspapers, television channels, and an online platform, challenging initial assumptions about media type clustering.
- Edge Betweenness Algorithm:**
  - Emphasizes sparse connections between communities by eliminating high-betweenness nodes.
  - Grouped entities across various media platforms, showcasing distinct clusters within the network.
- Leading Eigenvector Algorithm:**
  - Relies on modularity matrix eigenvectors to identify communities.
  - Produced similar results to the Walktrap algorithm, highlighting consistent clustering patterns within the network.

Our Evaluation from community detection algorithm is that While clusters aren't solely determined by media type, there's evidence of distinctive groups formed by shared properties such as connectivity and centrality measures.



## Conclusion

Our social network analysis project delved into the complex dynamics of information dissemination networks, employing a diverse range of techniques.

- Clustering algorithms identified distinct network clusters, validating our hypothesis regarding cluster emergence based on shared properties.
- Our initial hypothesis correlating audience size with network centrality was refuted.

While our analysis yielded valuable insights, unraveling complex network structures demands ongoing exploration and resources. Further research is essential to deepen our understanding of information media platform networks.