Understanding the Competitive Landscape of News Providers on Social Media

Devipsita Bhattacharya, Sudha Ram
dbhattac@email.arizona.edu, ram@eller.arizona.edu
Department of Management Information Systems
Eller College of Management, University of Arizona
Tucson, AZ, 85721

ABSTRACT
Social media has emerged as a mechanism for online news propagation. This in turn has changed the competitive landscape of news providers, a landscape that was previously partitioned based on the traditional channels of news dispersion. The channels of news distribution refer to - *television*, *newspaper*, *magazine*, *radio*, *news agency* and *online only*. In this paper, we examine similarities and differences in news propagation patterns on social media based on the primary channel of a news provider. We collected news article propagation activity data from Twitter for 32 news providers over a three-week period and analyzed their propagation networks. Our analysis shows that the structural properties of the propagation networks are statistically different based on the type of primary channel. Our study has useful implications for understanding the competition between news providers in an online environment.

General Terms
Measurement, Performance.

Keywords
Twitter; Micro-blogging; News Propagation; Article Propagation.

1. INTRODUCTION
Use of web and electronic presentation technologies has enabled news providers to create content rich webpages to provide news in a detailed and an engaging manner. For anyone with an internet connection, news is now an on-demand commodity. Furthermore, with the social recommendation and content sharing features of social media websites, news articles are being distributed and consumed at an unprecedented scale.

This phenomenon of news circulation has changed the landscape of news providers. That is, news providers that previously *competed* with other providers primarily based on the channel of news distribution, now find themselves competing with a whole new set of participants. For instance, before the Internet, newspapers such *New York Times* and *Washington Post* were competitors for subscriptions and advertising space in the printed media. Similarly, network news companies such as *CNBC* and *CNN* were competing with each other for audience engagement during prime time news hour. However, with each of these news providers now having news websites, the competition is no longer limited to their rivals in their primary channel. News providers now also contend to attract advertisers and readers for their article webpages. Moreover, the competition for content popularity and audience engagement also extends to the social media space. On Twitter for example, posts containing links to articles published on news websites maybe retweeted by a large number of users leading to article popularity and additional traffic to news website [11].

In this study we compare the patterns of news article propagation on Twitter based on the primary channel of news providers. We use network analysis methodology to extract user-user networks of selected news providers and analyze their structural properties. Such an analysis has important implications for understanding the competitive landscape of news providers on social media.

WWW 2016 Companion, April 11-15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2890097

2. BACKGROUND

2.1 Definitions and Methodology
Previous studies have examined the introduction of online sites and the competition between online and traditional channels of news distribution [9; 10; 15]. These studies however are largely based on surveys. Moreover, many of these studies are limited to specific countries or have compared only a few news providers. None of these studies have examined how news providers compete on social media. Our study is different from extant studies in that, it uses a network analysis methodology to analyze competition among news providers on social media. The term "news provider" refers to companies that are in the business of collecting and publishing news.

News providers on the Internet can be grouped into different categories based on their primary ***channel*** of distribution. We use Wikipedia provided news company information to examine six major operating news channels and classification listed below [1].

- *Television* (Network News) – Televised news companies. E.g. CNBC.
- *Radio* –Radio news broadcasting companies. E.g. NPR.
- *Newspapers* – Print Newspapers (Usually *dailies*). E.g. New York Times.
- *Magazines* – Periodical print publications. E.g. Forbes, Wired.
- *Online Only* – News providers that only use the online channel to provide news. This includes editorial news websites. E.g. Slate.com, Arstechnica.com.
- *News Agency* – News providers whose primary purpose is to generate news and deliver it to other subscribing news providers. These include news curator organizations such as Reuters and Associated Press.

We compare the competition among these news channel categories on social media by extracting Twitter based news propagation patterns for a selected set of news providers and analyzing similarities and differences among the networks. We examine the values of structural properties such as diameter, degree centrality and density of the propagation networks.

Our Twitter based propagation network is a user-user network defined for a single news provider [6]. It is a network of aggregated

propagation activity (i.e. across multiple articles) observed over a period of time. The network G = [N, E, W] is defined as follows:

- **Node** – Each node is this network represents a user that tweets (or retweets) a news article of a given news provider. Replies to tweets are not considered.
- **Edge** – An edge is a directed relationship between two Twitter users established as a result of retweeting activity. The direction of the edge is outwards from the source user to the target user.
- **Weight** – Since we are modeling aggregate propagation activity over time, we associate two kinds of weights with network entities.
    - ***Node (or User) Weight*** – For a given user, it is the number of tweets from a user.
    - ***Edge Weight*** – For a user-user edge, it is the number of times a target user retweets the source user.

We analyze the competition between a set of news providers (and news channels) by comparing their respective user-user propagation networks. Here, we make the assumption that users tweeting (or retweeting) article(s) of a given news provider form an implicit community of the news provider's consumers. Current research on news related user communities in social media have identified groups of users based on features such as news category [3; 17], statement bias [17], selection bias [17], news brand [5; 7; 8], sharing behavior [2] and voting patterns [12] to name a few. Our study on the other hand focuses on news distribution channel. We analyze the structural features of each of the user-user propagation networks to compare news providers.

2.2 Data Collection and Processing

We collected a dataset of 24 million tweets from Twitter for 32 news providers shown in Figure 1. This dataset was for a 3-week period, i.e. September 1- September 22, 2015.

Magazine	News Agency	Newspaper	Online	Television	Radio
• The Atlantic • Forbes • Wired	• Associated Press • Bloomberg • Reuters	• Boston Globe • Economist • Financial Times • Guardian • Independent • Los Angeles Time • New York Times • The Telegraph • USA Today • Wall Street Journal • Washington Post	• Arstechnica • Daily Beast • Engadget • Global Post • Mashable • Slashdot • Slate • Tech Crunch	• ABC News • Al-Jazeera • BBC • CNN • NBC • PBS	• NPR

Figure 1: List of News Providers and their Primary News Channels

The filtering criteria used to collect the tweets were as follows:

- For each organization, we identified a list of Twitter handles (or official Twitter accounts). We collected all the tweets (and retweets) of these Twitter accounts.
- For each news provider, we generated a list of common keywords such as "NYTimes" for New York Times. We collected all tweets containing these keywords.

These criteria allowed us to collect a comprehensive dataset containing the propagation activity for all articles of each news provider. We then constructed a propagation network for each of the 32 news providers. We used a combination of technologies including *Apache Pig*, *MySQL* and *Gephi* for processing, storing, visualizing and analyzing our networks [4].

A graphical visualization of ABC News' propagation network is shown in Figure 2. Here, the nodes (and consequently the edges) are colored based on *Modularity* clustering algorithm to outline different clusters of ABC News article readers formed over the time period of analysis i.e. 3 weeks .

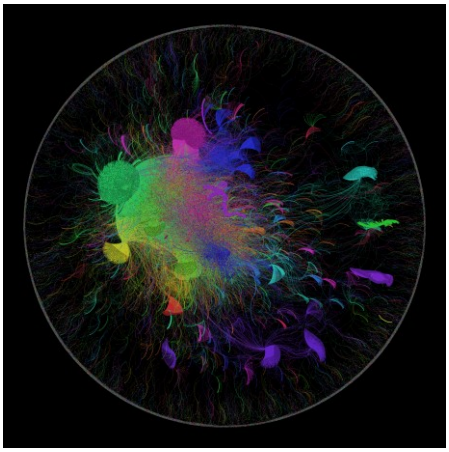

Figure 2: Visualization of User-User Network of ABC News with 261,062 nodes and 255,566 edges

Figure 3 and Figure 4 show the nodes and edges count for each of the news propagation networks. These propagation networks are very large (high node and edge counts) and thus provide an extensive representation of the news propagation.

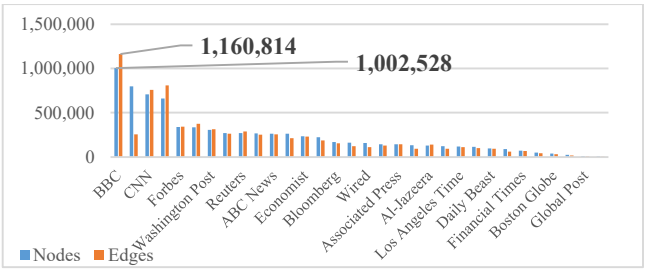

Figure 3: Node and Edge Counts of the User-User Networks

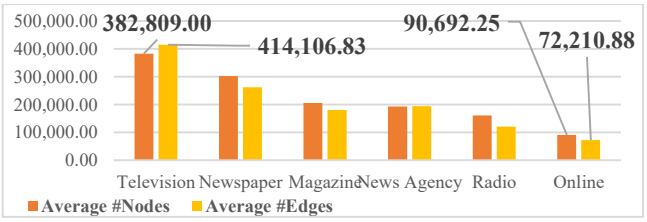

Figure 4: Average Number of Nodes and Edges for Propagation Networks Based on Primary News Channel

As previously stated, the goal of this paper to examine whether the propagation patterns associated with the primary channels of news distribution (and by extension the news providers) are similar or different from each other on social media (such as Twitter). We analyze the news providers' propagation networks and compare their structural properties.

3. ANALYSIS

We computed structural properties of every news provider's propagation network and compared them. Additionally, we grouped and analyzed the propagation networks of news providers according to their *primary news channel*.

3.1 Network Density and % Disconnected Nodes

Based on our initial analysis (Figure 3 and Figure 4) of the propagation networks we found that the *television* (network news) organizations have the largest propagation networks (highest number of nodes and edges). BBC in particular had the largest

network with > 1 million nodes and edges. PBS, the niche television station, with its *NewsHour,* had the smallest of the propagation networks for television channel with approx. 50,000 nodes and 40,000 edges. "*Online only*" news providers tend to have much smaller networks with an average of 90,000 nodes.

All these propagation networks are sparse with very low density values of the order of 10^{-5}. Also, interestingly, "*online only*" news providers had relatively higher density values when compared to rest of the news agencies (Figure 5).

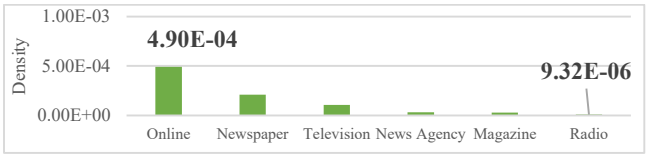

Figure 5: Average Density Values of News Providers for Primary News Channels

We further analyzed the network sparsity by calculating for each of the networks, the percentage of total nodes that were *disconnected*, i.e. they have zero incoming edges. We found that except for *New York Times* with a very high value of 70%, the values of *% disconnected nodes* range from 15% (Economist) to 45% (Slashdot) (Figure 6).

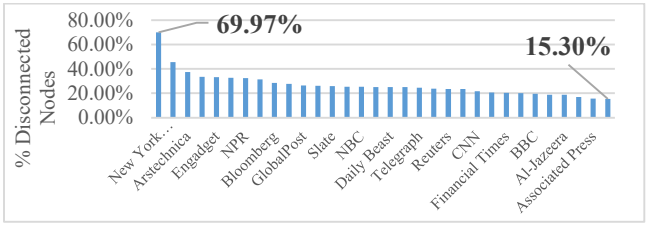

Figure 6: % of Disconnected Nodes for News Provider Networks

Moreover, we saw that the *radio* channel has the highest average value of % disconnected nodes (Figure 7). This was in accordance with our earlier examination of density values for *radio* channel (in Figure 5). Conversely, we did not find a significant correlation between *% of disconnected nodes* and *density*.

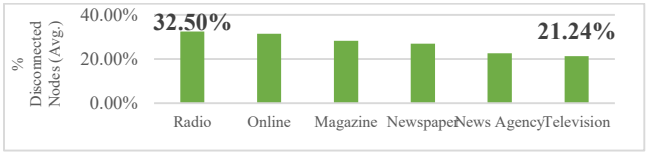

Figure 7: % of Disconnected Nodes for Primary News Channels

3.2 Eccentricity and Diameter

Next, we compared the propagation networks using eccentricity**.** *Eccentricity*, a network property defined for every node, is the longest geodesic distance from a node to the farthest node in the network. It provides information about the distances between the nodes in the network. We analyzed eccentricity values for the connected nodes in each news provider's propagation network. *Guardian* had the highest average eccentricity value with 4.29 user-user hops (Figure 8). This implies on an average, the connected nodes in *Guardian's* propagation network are at most four hops away from each other. This is in contrast to *Slashdot's* propagation network where the connected nodes on an average had a direct edge connecting them (average eccentricity is approximately 1). This indicates that on an average, *Guardian's* readers are more likely to participate in longer cascade chains.

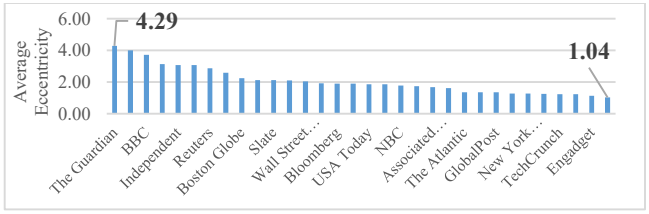

Figure 8: Average Eccentricity Values for News Provider Networks

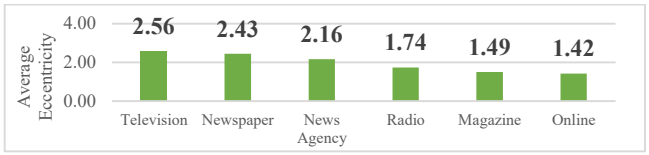

Figure 9: Average Eccentricity Values for Primary News Channels

On analyzing the average eccentricity values for news channels (Figure 9), we saw that *television* and *newspaper* channel tend to have higher average eccentricity values (2.56 and 2.43 respectively). This is interesting given the fact that the propagation networks for these channels have high node and edge counts.

Using eccentricity values we further calculated the diameter of each of the news provider propagation networks. *Diameter* is the longest geodesic distance observed in a network. For a network such as ours that captures the aggregate cascade activity for a news provider, *diameter* represents the longest cascade chain established between a set of users over a period of time [5].

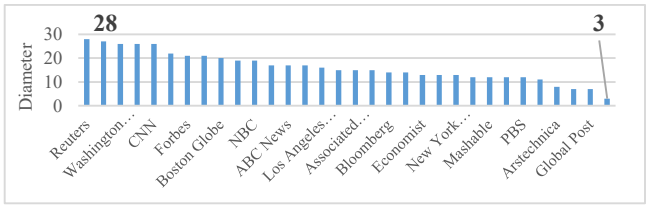

Figure 10: Diameter of the User-User News Provider Networks

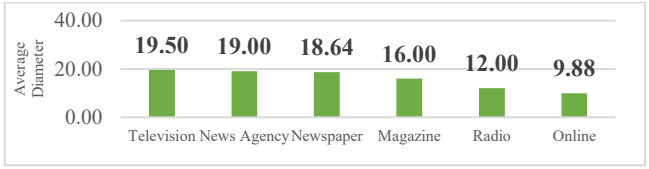

Figure 11: Average Diameter Values of News Providers for Primary News Channels

The diameter values (Figure 10 and Figure 11) vary from 28 user-user hops (*Reuters*) to 3 user-user hops (*Slashdot*). "*Online only*" news providers owing to their smaller networks also have shorter diameters. We did not identify any diameter specific patterns related to a primary news channel. However, we found that for a propagation network, average eccentricity value is inversely correlated with diameter (sig. at 0.01 level).

3.3 Degree Centrality and Clustering Coefficient

Degree centrality (and *weighted degree centrality* in particular) is a network metric that calibrates the "presence" of a user in the propagation network. Weighted Degree is calculated as the sum of the weights of the edges incident on a node. A higher average

weighted degree indicates that the nodes participate more in the news cascade.

Figure 12, Figure 13 and Figure 14 show the results of the degree centrality analysis of the **connected**[1] user nodes. We analyzed both the weighted and the non-weighted degree values of these nodes.

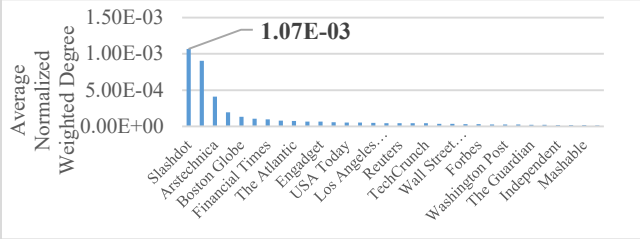

Figure 12: Average Normalized Weighted Degree for News Provider Networks

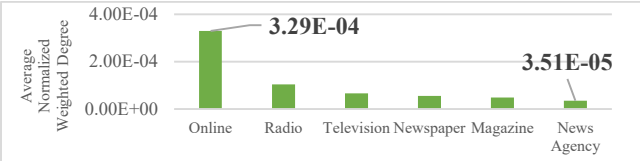

Figure 13: Average Normalized Weighted Degree for Primary News Channels

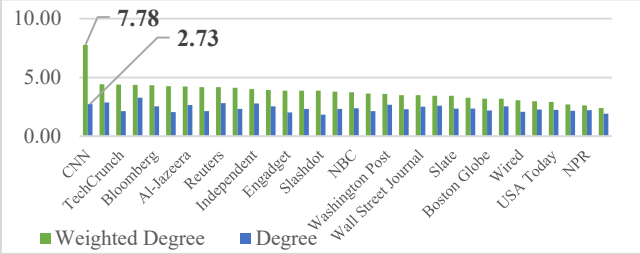

Figure 14: Average Degree and Average Weighted Degree for News Provider Networks

Television channel based news providers continued to perform better than rest of the news channels when considering average degree and weighted average degree values. However, after averaging normalized weighted degree (using highest weighted degree for a news provider's user-user network) we observed that "online only" news providers had the highest values.

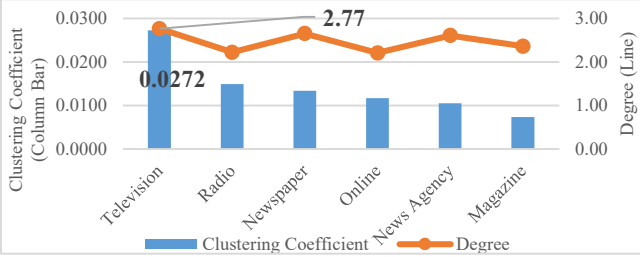

Figure 15: Clustering Coefficient and Weighted Degree Centrality Comparison of News Channels

We also analyzed the clustering coefficient of the **connected** nodes (Figure 15). We found that *television* news providers had the highest average clustering coefficient and the highest average degree amongst other news channels. With 3 connections per node (degree = 2.77), *television* network nodes have approximately double the average probability of forming triangles with its neighbors.

3.4 User Weight

We also compared the node (*user*) *weights* (number of tweets from a user) across propagation networks. However, as the tweeting activities of news providers' Twitter accounts (**NPro**) are likely to differ greatly from that of other Twitter users (**TwU**), we analyzed the two sets of users separately.

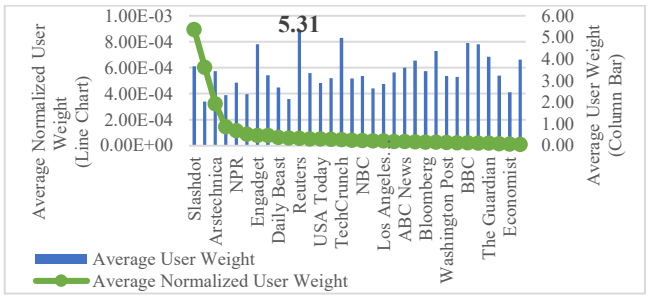

Figure 16: Average User Weight and Average Normalized User Weight for TwU for each News Provider's Network

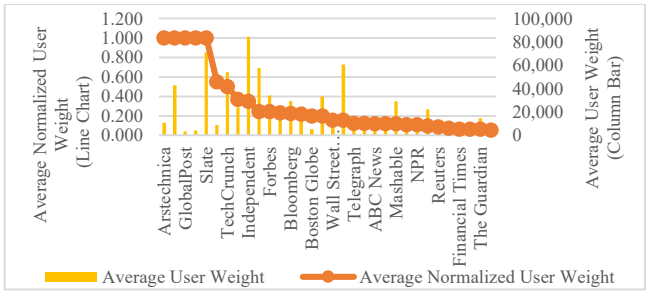

Figure 17: Average User Weight and Average Normalized User Weight for NPro for each News Provider Network

Figure 16 and Figure 17 show the user weight and the normalized weight (normalized using highest user weight in the respective user-user network) for Twitter Users (TwU) and News Provider's Twitter Accounts (NPro) respectively. We found that the user weights of *NPro* accounts have values greater than 3000, with *The Independent* having the highest tweeting activity (combined activity across all official Twitter accounts) of 84,570. On the other hand, Twitter users (TwU) had an average user weight of 3.42. This indicates that users (TwU) on an average have tweeted at least three news articles during the observation period of three weeks. For a news provider such as Reuters, the user weight was as high as 5, indicating high user participation on an average. Also, it's noteworthy to mention that we found a significant correlation between the average user weight of *NPro* nodes and the average user weight of *TwU* nodes (Significant at 0.05 level). This implies that a high tweeting activity of news provider accounts is associated with high retweeting (or tweeting) activity of its Twitter readers.

We further examined user weight values for each news channel and compared them (Figure 18). We did not observe any major differences in absolute values of average user weight between news

[1] Due to a highly skewed distribution of degree centrality and a high count of disconnected nodes, we present degree centrality analysis of connected nodes only.

channels. Based on average normalized user weight however, we saw that "online only" news providers had the highest value.

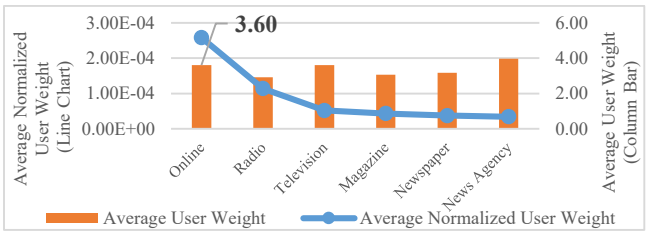

Figure 18: Average User Weight and Average Normalized User Weight for TwU for each News Channel

This is an interesting observation since for "*online only"* networks, with low edge and node counts, we had observed high average normalized degree and average normalized user weight values when compared to other news channels.

3.5 % Nodes in Largest Cluster

In addition to examining, eccentricity, degree, disconnected nodes and user weights of the propagation networks, we also analyzed how the nodes in these networks cluster together. Specifically, we looked into the largest cluster identified in each propagation network using *Modularity* algorithm [14].

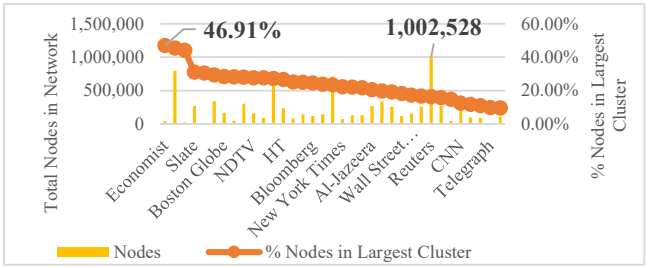

Figure 19: % Nodes in Largest Cluster compared to #Nodes for Each News Provider

Figure 19 shows the number of nodes in each propagation network (left Y-axis and column bars) and the % total nodes contained in the largest cluster (right Y-axis and dotted line). We saw that in some cases, the largest cluster had approximately 47% of the nodes in the network (*Economist*), whereas in case of *NPR* and *Telegraph*, the largest cluster comprised merely 10% of the nodes in the network. Also, we found no significant correlation between the number of nodes and the % of total nodes contained in the largest cluster.

Furthermore, we also compared number of nodes with the % nodes in the largest cluster for each news channel (Figure 20). Although, we did not observe any distinguishing pattern of node participation, it was interesting to observe that the average values of *online news* channel is high (28%) given that their networks have low node and edge counts.

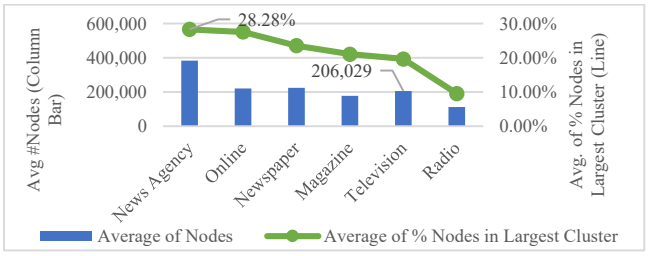

Figure 20: % Nodes in Largest Cluster compared to #Nodes for Each News Channel

3.6 Statistical Difference of Structural Properties

Finally, we conducted an *Independent Samples Kruskal-Wallis* Test for a set of structural properties to test for statistical difference between the news channels. The results are summarized in Table 1.

Table 1: Summary of Independent Samples Kruskal-Wallis Tests

Null Hypothesis – H_0	**Sig.**	**Result**
The distribution of *Eccentricity* is the same across all categories of news channel.	p<0.00001 (99%)	**Reject the Null Hypothesis**
The distribution of *Betweenness Centrality* is the same across all categories of news channel.		
The distribution of *Closeness Centrality* is the same across all categories of news channel.		
The distribution of *Degree* is the same across all categories of news channel.		
The distribution of *Weighted Degree* is the same across all categories of news channel.		
The distribution of *Clustering Coefficient* is the same across all categories of news channel.		

Note: Structural properties' values were considered for all the nodes in the networks.

Our analysis indicated that the structural properties of the propagation networks are statistically different for each news channel. However, this distinctiveness might not necessarily be a function of news distribution channel alone. There might be differences based on the brand of the news provider. Thus it would be interesting to investigate the combined effect of news channel and news agency brand on news propagation in social media.

4. DISCUSSION AND IMPLICATIONS

Generally, Porter's Five Forces model is used for strategic analysis of organizations in an industry [16]. However, in our study, we have used a set of structural properties to analyze the differences (or similarities) between the propagation networks of news providers. This analysis gave us insights into the *competition* among news providers on social media. We determined that "online only" news providers have the smallest networks (number of nodes and edges) with relatively higher density values. But, even with high density, their networks were found to have a higher concentration of disconnected nodes. This is expected since "online only" news providers have emerged only recently when compared to other news providers in our sample. For other news channels, we had mixed inferences when examining structural properties of their propagation networks. But, we were able to establish a statistically significant difference between the news channels based on *six* node-based structural properties.

Our analysis of the news channels using a network based methodology makes several contributions. *First*, it allows news providers to benchmark their social media based propagation performance against other competitors in the same or in a different primary news channel. This is particularly important since on social media, even traditional suppliers of news (e.g. News agencies such as Reuters, Associated Press) are considered direct competitors for any news provider hosting an online news website. *Second*, we were able to establish that our networks are indeed representative of the news article propagation activity since the values of structural properties are typical of information networks [13]. However, we also identified features unique to our Twitter-based

aggregate user-user networks. Primary of these was the presence of multiple disconnected clusters of nodes. On an average, we found that a news provider's propagation network had at least 4,000 disconnected clusters containing two or more nodes. This highlights the importance of news article tweeting activities independent of those originating from news provider Twitter accounts. This also emphasizes the need to consider not just the giant connected component of a news propagation network, but to also include the disconnected sections while analyzing values of relevant structural properties of propagation networks. *Finally*, we demonstrated that audience engagement can be measured in different ways such as user weight (tweeting frequency based participation), degree (retweeting frequency based participation), eccentricity (audience distance approximation) and % nodes in largest cluster (audience cohesiveness). These structural properties add new dimensions to competitive analysis which generally considers participation volume (number of users) to measure engagement. For instance, we observed that "news agency" (Reuters, Associated Press etc.) propagation networks had lower average counts of nodes and edges when compared to those of "television" (ABC News, CNN etc.) news networks (by a margin of 100,000). By considering these differences in values, television news providers emerge as "winners" in audience participation on social media when compared to "news agency" networks. However, we also ascertained that *television* and *news agency* networks had approximately equal values of network diameter (19.5 and 19 respectively). While on an average television based news agencies networks show higher tweeting and retweeting activity from their Twitter users, their audience's ability to connect amongst each other to form the longest cascade chain over time is the same as that of "news agency" networks having lower average Twitter user participation count. Thus, the analysis of competition among news providers on social media needs a comprehensive consideration of various facets associated with user participation.

5. CONCLUSION AND FUTURE WORK

In this research, we used a network analysis methodology to examine the competition among news providers on social media. Our analysis compared propagation networks extracted from a dataset of news article tweets collected from Twitter. Based on our analysis, we were able to gauge the competition between the news providers based on their *primary distribution channel*. Additionally, via statistical analysis, we were also able to establish that the structural properties of each of the news propagation networks is statistically distinct based on news channel. Our study provides a comprehensive account of the state of competition existing between news channels (and news providers) on social media. Our network-based methodology for competition measurement by extracting propagation patterns from social media can be used to understand the changing landscape of news distribution on social media.

We will examine the competition between the news providers' propagation networks by identifying overlaps among them. Such an analysis will be used to derive implications about brand loyalty of news readers on social media.

6. REFERENCES

[1] 2016. News media. Wikimedia Foundation, Inc., Wikipedia, the free encyclopedia.

[2] An, J., Quercia, D., and Crowcroft, J., 2014. Partisan sharing: facebook evidence and societal consequences. In *Proceedings of the ACM conference on Online social networks* (Dublin, Ireland, October 2014), ACM, 2660469, 13-24. DOI= http://dx.doi.org/10.1145/2660460.2660469.

[3] Bandari, R., Asur, S., and Huberman, B.A., 2012. The pulse of news in social media: Forecasting popularity. *arXiv preprint arXiv:1202.0332*.

[4] Bastian, M., Heymann, S., and Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of International Conference on Weblogs and Social Media* (San Jose, California, USA, May 17 – 20, 2009), *8*, 361-362.

[5] Bhattacharya, D. and Ram, S., 2012. News article propagation on Twitter based on network measures - An exploratory analysis. In *Proceedings of the Workshop on Information Technology and Systems* (Orlando, Florida, USA, December 2012).

[6] Bhattacharya, D. and Ram, S., 2012. Sharing News Articles Using 140 Characters: A Diffusion Analysis on Twitter. In *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (Istanbul, Turkey, August 2012*)*,966-971. DOI= http://dx.doi.org/10.1109/ASONAM.2012.170.

[7] Bhattacharya, D. and Ram, S., 2013. Community Analysis of News Article Sharing on Twitter. In *Proceedings of the Workshop on Information Technology and Systems* (Milan, Italy, December 2013).

[8] Bhattacharya, D. and Ram, S., 2015. RT@News: An Analysis of News Agency Ego Networks in a Microblogging Environment. *ACM Trans. Manage. Inf. Syst. 6*, 3, 1-25. DOI= http://dx.doi.org/10.1145/2811270.

[9] Chan, J.K.-C. and Leung, L., 2005. Lifestyles, reliance on traditional news media and online news adoption. *New Media & Society 7*, 3, 357-382.

[10] Dimmick, J., Chen, Y., and Li, Z., 2004. Competition between the Internet and traditional news media: The gratification-opportunities niche dimension. *The Journal of Media Economics 17*, 1, 19-33.

[11] Filloux, F., 2012. News providers should not like Facebook all the time Guardian Media Group, The Guardian.

[12] Lerman, K. and Ghosh, R., 2010. Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks. In *Proceedings of International Conference on Weblogs and Social Media* (Washington, D.C., USA, May 23 – 26, 2010), 90-97.

[13] Newman, M.E., 2003. The structure and function of complex networks. *SIAM review 45*, 2, 167-256.

[14] Newman, M.E., 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences 103*, 23, 8577-8582.

[15] Nguyen, A. and Western, M., 2005. The complementary relationship between the Internet and traditional mass media: the case of online news and information. *Information Research 11*, 3, 8.

[16] Porter, M.E., 1979. How competitive forces shape strategy. *Harvard Business Review*, 21-38.

[17] Saez-Trumper, D., Castillo, C., and Lalmas, M., 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the ACM International Conference on Information & Knowledge Management* (San Francisco, USA, October 2013) ACM, 1679-1684.