Investigating the Twitter Behavior of Tax Experts Networks in Social Network Sites

Authors: Regina Ruane & TL Hill
Presenter: Regina Ruane
Presenter's email: regina.ruane@temple.edu

Fox School of Business
TEMPLE UNIVERSITY®

IARIA

Presenter Biosketch

- Regina Ruane is an assistant professor of research in the Department of Statistical Science and executive director of Temple University's Data Science Institute. Ruane also directs pedagogical research with the Fox Translational Research Center (TRC).
- Ruane joined the Department of Statistical Science after having been a research fellow for Harvard University and Temple's Data Science Institute, where she conducted research in data science and digital analytics with a focus on networks, information systems and statistics.
- Prior to her work at the Fox School, Ruane founded and directed the First-Year Exploratory Studies program at Drexel University, which became the fastest-growing program at the university with a 900 % enrollment increase within the first two years of offering. Additionally, she has been a strong leader, conducting visionary strategic planning, innovative program development, and relationship and partnership development.
- Her work has appeared in journals across Information Systems and Education, including Communications of the Association for Information Systems and Online Learning, where she received the Best Paper Award.
- Ruane earned her PhD from Drexel University, her MA from Columbia University and a BA from Villanova University.

Overview

- ***With the pervasiveness of social network sites, determining and quantifying how content develops on these platforms and subsequently impacts or affects other areas is becoming increasingly important.***
- ***In this paper, we adopt a network analysis method that enables the estimation of the coevolution of the users' social network structure.***
  - ***Specifically, we offer a method to analyze social media data, typically recorded in social media ecosystems, in the presence of network effects.***
  - ***The proposed method can help disentangle network effects of interest from feedback effects on the network.***

Introduction

- Social network sites (SNS) provide a unique value with the engaging interactions they offer between the two possible roles of users, i.e. "content producers" and "content consumers".
- Content producers actively post, comment, and share content with their friends and content consumers view and react to the information posted by content producers.
- Content producers add considerable value by generating and sharing content through the network .
- SNS user behavior is influenced in part by individual-level factors, such as demographics, traits, among others and partially by online social network dynamics, such as the extent of network clustering, network betweenness, and other social network factors.

Introduction

- Earlier studies have demonstrated that networks form with advantages constituted from social capital.
- Previous research has also shown that people cluster into groups as a result of interaction opportunities defined by the places where people meet [21].
- Other studies have postulated that communication is more frequent and influential within than between groups as people in the same group develop similar views, demonstrating the influences of homophily and social influence [14].
- Education, occupation, and occupational prestige have been shown to have strong levels of effect on homophily in network studies [14].

Introduction

- Future research surrounding expanded consideration of multiplexity of both networks and foci and the need for dynamic data on changes over time in networks are necessary to determine:
  - the ways in which networks evolve over time through cumulative processes of tie creation and
  - dissolution as they are embedded in a changing community of multiplex relations spawned by multiple organizational [14].

Introduction

- To address this gap, we have made the following inferences about the nature and extent of peer influence &
- homophilous peer selection in content production in tax conversations on the SNS of Twitter.
    1. we provide evidence for homophily based on similarity in the individual-level covariate of organization affiliation
    2. we report the existence of peer influence,
        1. finding different stages of friendship formation among independent and organizational-related users.
        2. Individuals befriend others who are similar in content production behavior
    3. we provide evidence that the strength of social influence varies as a function of the specific level of the behavior.

Literature review

- SNSs provide a unique value with the engaging interactions they offer between the two possible roles of users, i.e. "content producers" and "content consumers".
- Content producers actively post, comment, and share content with their friends
  - add considerable value by generating and sharing content through the network
- Content consumers view and react to the information posted by content producers.
- SNS user behavior is influenced in part by individual-level factors, such as demographics, traits,
  - & by online social network dynamics, such as the extent of network clustering, network betweenness, and other social network factors.

Literature review

- Other studies have postulated that communication is more frequent and influential within than between groups as people in the same group develop similar views,
  - demonstrating the influences of homophily and social influence
- Education, occupation, and occupational prestige have been shown to have strong levels of effect on homophily in network studies [14].
- Future research surrounding expanded consideration of multiplexity of both networks and foci are necessary to determine
  - the ways in which networks evolve over time through cumulative processes of tie creation and dissolution as they are embedded in a changing community

Research

- To address this gap, we have made the following inferences about the nature and extent of peer influence as well as homophilous peer selection in content production in tax conversations on the SNS of Twitter.
  1. providing evidence for homophily based on similarity in the individual-level covariate of organization affiliation.
  2. reporting the existence of peer influence, finding different stages of friendship formation among independent and organizational-related users. Individuals befriend others who are similar in content production behavior.
  3. providing evidence that the strength of social influence varies as a function of the specific level of the behavior
  4. We show that low content posters are more susceptible than heavy content posters to have higher network measures, e.g indegree and outdegree centrality.

Research

- The findings are robust to the presence of significant differences in understanding the tax-related conversations due to the significant levels of separate clusters.
- While there are levels of understanding occurring between people, the levels in separate clusters are much greater.
- These findings indicate high levels of nonredundant information that are more additive to the conversations and advancement of discussion.

Method

- For this analysis, we collected 60,000 tweets sent between 1 January and 31 October 2020 by United States tax policy actors.
- The data collected included Twitter handles (names), hashtags, mentions, text, and time sent.
- Such data is useful because it reveals the *actual* content and sharing patterns of user-generated content, not individuals' self-reported perceptions and intentions.

Data

- The data were scraped from the Twitter API
- the collection process was seeded by a list of key tax policy actors identified by tax policy and social media experts.
- The resulting dataset is the population of tweets, not a sample, and
- fits the theoretical and methodological assumptions of network analysis.

Method

- The method employed in this study has the following components:

i. construction of a sample of tax-related Twitter data from Twitter's public data API
ii. organization, analysis, and interpretation of the network clusters.
iii. The resulting data contains metadata for the tax-related discussions from the aggregate activity on Twitter.

-

Network Analysis

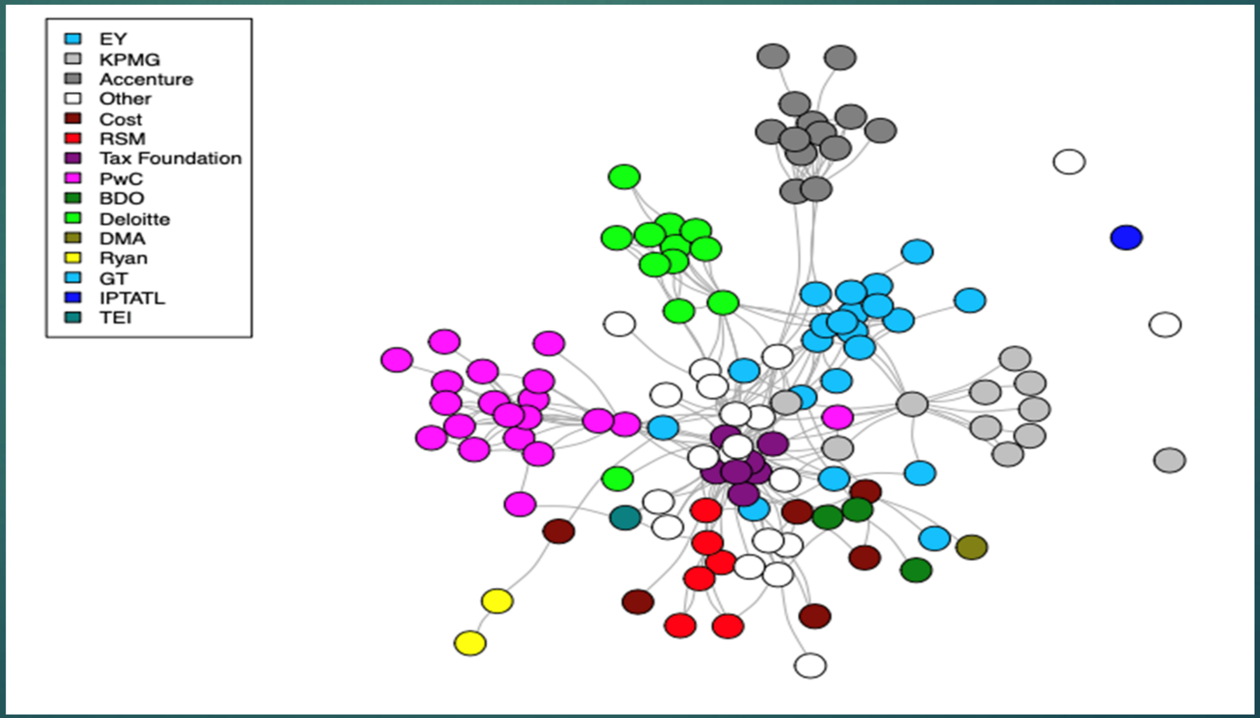
- The network analysis of Twitter that we conducted is based on the structure induced by the communications surrounding tax.
- The communications surrounding tax were determined by directed links between Twitter users,
  - i.e., a user has a (directed) edge to another user for each directed communication the user sends or receives.
- With the extremely large dataset and the large number of edges the interactions of the users produced, the sociogram demonstrated large amounts of activity.

The Tax Network

- To examine connectivity of users, we used a threshold of a minimum of two or more direct communications and
- filtered the network eliminating connections stemming from only one direct communication.
- As part of the analysis, we examined the Twitter handles that were connected based on directed communications.
  - determined the level of closeness, i.e. whether handles were in the same organization or not.
  - looked at the volume and type of communication among the twitter handles and retweets.
  - investigated the betweenness centrality measure and the influential aspects

The Tax Network

- Figure 1 shows the filtered network. This allowed us to keep the network within a size that we could process.



Clustering

- To visualize the clusters of the Twitter network, we used R to represent the users and their directed interactions in sociogram form.
  - We added attribute data regarding organization affiliation for users.
  - User organization affiliation is denoted by color in the sociograms.
  - The organization affiliation distinctions allowed us to determine the tweet clustering for users and the larger signification(s) for the network.
  - The unconnected nodes denote users who did not have any directed communications (tweets) with other users.
  - These users may have tweeted broadcast messages to others or may have had one directed communication with other users.

Clustering

- Using the edge betweenness clustering algorithm in igraph, we performed a clustering detection for the filtered tax network.
- The edge betweenness clustering algorithm calculates the edge betweenness of the graph,
  - removes the edge with the highest edge betweenness score,
  - then recalculates edge betweenness of the edges and
  - again removes the one with the highest score.
  - The algorithm performs a hierarchical clustering in which a node is assigned to a cluster if doing so maximizes the modularity of the network, continuing until either a single node remains or modularity cannot be increased further

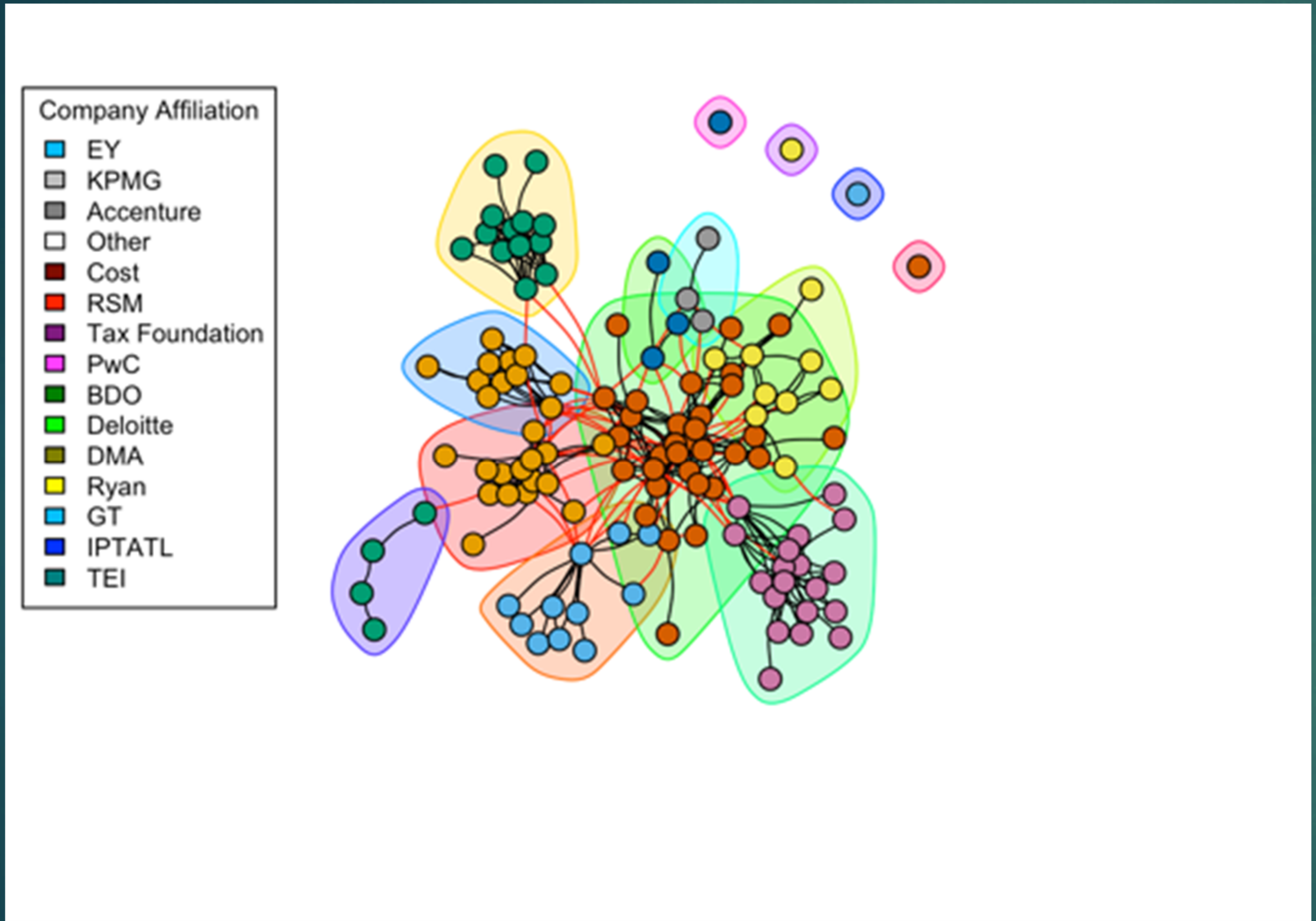
Social Network Analysis

- The social network analysis includes the creation of sociograms to view the network from a macro perspective, clustering and betweenness centrality measures.
- Typically, in social network analysis, the structure of a network is defined as G=(N, Y),
  - in which N is a set of nodes that represent individual entities, and Y is a set of edges that represent the connectivity pattern among the nodes.
- The social network analysis results show the coevolution of network structure and user behavior in online social networks.
- We show that user-specific factors, like organizational affiliation, play crucial roles in shaping users' varying reactions to the policy change.

Social Network Analysis

- The network analysis provides understanding of the nature of peer effects on social network evolution on social network platforms and has clear practical implications.

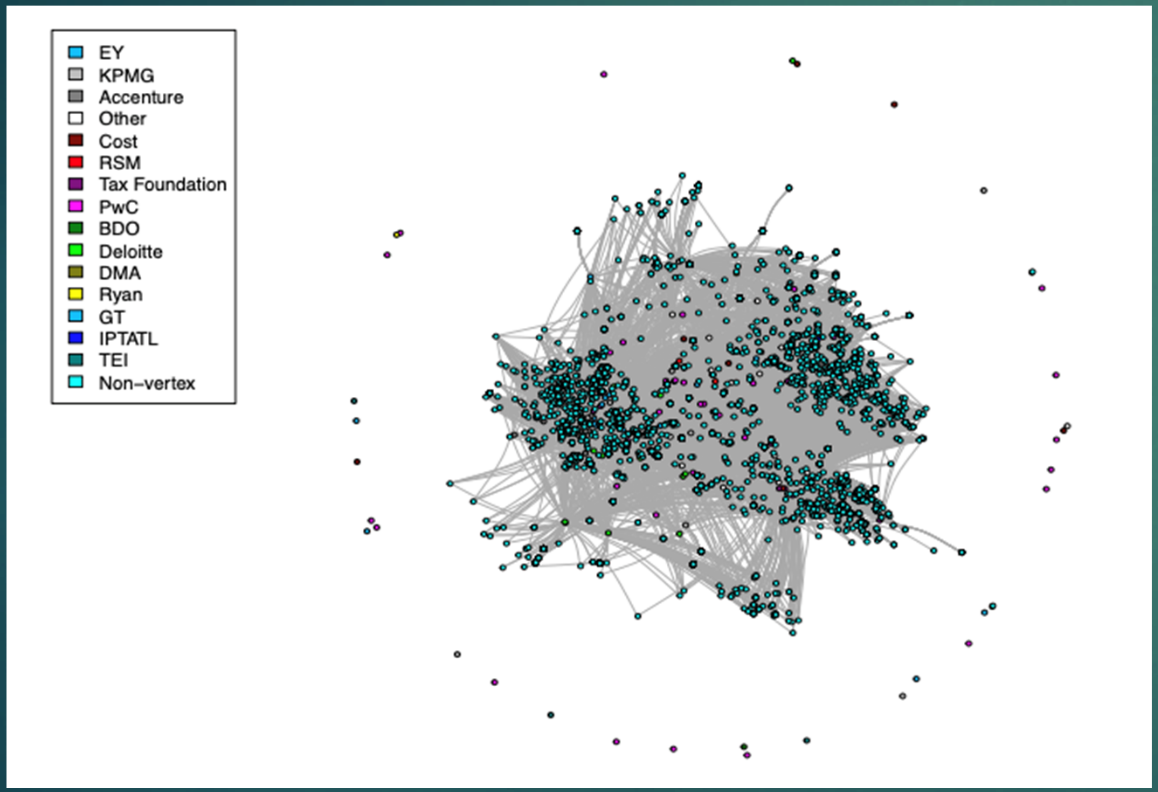
1. we offer a framework within which online user contributions can be studied to identify a larger and more involved network.
2. our results provide intelligence to identify and better target powerful and well-connected users on social network platforms.
3. our analysis offers insights into posting behaviors on these platforms, especially with the inclusion of attribute data that allows for the incorporation of larger network patterns,
    - such that managers and researchers can effectively determine structural patterns and forecast the diffusion of this content through social networks.



Social Network Analysis

- This figure depicts the filtered network following the clustering algorithm application.
-

Social Network Analysis



- Fig. 3 demonstrates a very high-level feel for the directionality of the tweets.
- This network is a highly clustered, ordered network.
- The cumulative arrows in gray
- The Diffusion/flow implications show that the central core both receives and is thereby thought of as powerful, demonstrating influence through outward communications the many actors outside of the central core.

Social Network Analysis

	IN DEGREE	OUT DEGREE	BETWEENNESS	CLUSTERING
Min	0.0	0.0	0.00	0.00
25%	1.0	0.0	0.00	0.17
Median	1.0	0.0	0.00	1.00
Mean	1.5	1.5	17484.96	0.63
75%	1.0	0.0	0.00	1.00
Max	38.0	1079.0	11215427.66	1.00

- The measures listed are all positive and statistically significant.
- The mean indicates that the actors in average score are equal in terms of prominence and influence.
- The maximum scores indicate that influence is a much more significant aspect of this network than prominence with such a sharp difference in the range of maximum scores.
- The minimum scores indicate inactivity levels among users in terms of direct communications in this network.
- The outdegree activity to other actors denotes the pushing out of information about tax policy to others across the entire network (not just the core).

Conclusion

- Understanding the nature of network evolution on SNSs has clear and practical implications for social scientists, social media platforms, and their users.
- We offer a framework within which online user contributions can be studied as a function of the underlying network.
- Our analysis has shown that Twitter has a strongly-connected tax network which is nonetheless differentiated through users' organizational affiliations and by directed communications by users.
- Our analyses demonstrated the clusters that emerged through the directed exchanges among users.
- We observed that the clusters mostly developed based on organizational affiliations, but that some users exchanged

Conclusion

- This research represents the foundation for future social media related applications targeting a multitude of possible applications with specialized foci, such as cryptocurrency and economic issues.
- The large dataset of 60,000 tweets indicates the tax-related tweets comprise substantial activity from more than a hundred users, many of whom have reputable organizational affiliations in the tax sector.
- Our network of Twitter users, though large, is only a partial depiction of the tax-related Twitter discussion and its evolution.
- The majority of the Twitter data that we scraped and analyzed was during the COVID-19 pandemic.

Future Research

- Further research could study a longer time period to extend our work and determine the patterns and network structure of tax-related discussion on Twitter.
- As a network study of Twitter data, there are larger factors that may have influenced the twitter discussions, e.g. the American presidential election results, among others.
  - Future research could explore these influences

Conclusion

- This study serves to illustrate the network structure and patterns that emerge from tax-related discussions in a social media platform.
- As an initial attempt to model and analyze the coevolution of network structure and user behavior in the online social network, Twitter, this study is offers opportunities for future research.
- This paper focuses on providing a sound method to uncover the dynamic nature of tax-related discussions in a social network.
- Additional analyses are required to further separate the specific rationale behind why individuals show the observed clustering effects.