Fake News and Social Networks:

How Network Contacts Affect the Belief of Disinformation within Social Groups

A Research Design Paper

Kendall Goodland

PSCI 7381.002

Dr. Lauren Santoro

Introduction

Disinformation campaigns have greatly interrupted the flow of reliable, trustworthy information in both global and domestic politics. The consequences of these campaigns have already reared their heads – Donald Trump's election fraud rhetoric regarding the 2020 presidential election largely influenced political violence against the United States capitol building in 2021. Disinformation campaigns against Rohingya Muslims in Myanmar resulted in hundreds of thousands of people fleeing the country to escape ethnic genocide. Despite what is known about the association between disinformation and political unrest, not much is understood about the spread of disinformation, or "fake news". In the interest of understanding how disinformation spreads, it is important to analyze the intricacies of political conversation and action among various groups of people. One study that utilized content analysis among multiple sources from the year 2010 to 2021 found that disinformation campaigns followed consistent empirical themes (Broda & Strömbäck, 2024). One of those themes was strategic use, which is the focus of a large chunk of "fake news"-related research (Broda & Strömbäck, 2024). This validates sentiments around bad actors using disinformation for malicious intent, but there is another side to this phenomenon that requires focus as well – individual behavior. Network analysis has the potential to reveal much about how disinformation spreads, who is susceptible to it, and how its effects can be negated.

This is the purpose of the social network analysis approach – the more that is understood about social networks and their dynamics, the more conclusions can be drawn about how disinformation is spread within groups. However, it is important to know that these analyses are highly nuanced – it is very difficult to truly capture and understand network behavior, even when dealing with the most perfect of data. This research is not without its limitations, but there is still much to be learned from social networks in the context of political disinformation. This research has the potential to hold strong implications about how disinformation manifests, grows, and changes depending on the individual relationships among people in social networks. Public policy falls far behind when it comes to disinformation campaigns and their consequences – the findings of this research could prove essential in policymakers' decisions.

Research Question & Hypotheses

When it comes to social network analysis, there are multiple variables in play. In the context of this research, the goal is to understand how the strength of social ties within a network affects the spread of disinformation within those networks. Theory is a foundational part of social network analysis, and opens a lot of doors, but execution is difficult and must be thought about carefully. The scope of this research will remain small, as the necessary data is hard and time consuming to acquire. This study will focus on one main research question:

How does the number of known contacts that believe a certain piece of disinformation influence an individual's belief in that disinformation?

The analysis portion of this research will focus on two hypotheses:

H1: Individuals who report higher exposure to those who believe election fraud-related disinformation are more likely to believe that piece of disinformation.

H2: Individuals with politically homophilous networks are more likely to believe election fraud-related disinformation shared by those within their network.

Background & Theory

There are many theories and concepts that support social network research. Theory acts as a foundation in research, and as such, is an important pillar of design. One theory of focus is homophily. Homophily, in the context of social network research, purports that ties form between individuals or other types of nodes based on shared attributes – i.e., birds of a feather will flock together (Robins, 2015). This phenomenon coincides with the idea of "echo chambers" and political tribalism. For instance, a network study analyzing political twitter data during a 2017 Norwegian election found that retweet networks were largely polarized and reflected homophily (Enjolras & Salway, 2022). This is an interesting finding, as it supports general assumptions about how certain social networks interpret certain pieces of information – people tend to identify with new information that they agree with. More importantly, people tend to cycle information within their own circles, and this information tends to be less diverse. Homophily is equally relevant under this scope of research – there is a lot of resentment among scholars regarding tribalism and "echo chambers", but nonetheless, they are necessary considerations. However, it is important to note that homophily as a theory is only a starting point – one cannot produce grand conclusions with just assumptions.

Social contagion is also a relevant theory in this study, as it is one of the most foundational theories in network analysis. Social contagion is a phenomenon in which information, beliefs, or behaviors are influenced by your social ties or connections. According to one study, social contagion can be difficult to nail down, but there can be certain individual attributes that contribute to general social contagion processes, such as employment preferences, happiness, and tendencies to collaborate (Parker et al., 2022). Social contagion is the foundation of this study's first hypothesis, which predicts that those who report higher exposure to disinformation will be more likely to believe it.

Methods/Design

This study will largely rely on egocentric network data. For the purposes of this research, 2020 election fraud-related disinformation pieces will be the focus. There are many different avenues one could take to acquire these pieces, but the method this study will largely rely on is databases. For example, a 2022 paper including a "ElectionMisinfo2020" dataset compiled by scholars at the University of Washington contains millions of tweets involving over 400 different disinformation narratives about the 2020 U.S. presidential election (Kennedy et al., 2022). Tweets are often shared outside of Twitter/X itself, meaning individuals who do not interact with information on Twitter are still susceptible to online instances of disinformation. Databases like this are extremely useful, because they will make it easier to isolate specific narratives and put them to the test.

Web scraping could also be useful for gathering data, but there are limitations. First, many social media platforms are opting to paywall their data, making it less accessible and more expensive to acquire. Second, data gathered this way can be less than ideal when it comes to reliability. Web scraping can elicit sampling bias – the results may only reflect the loudest voices on any platform, at least when it comes to social media. As such, disinformation data collection will be done using online databases like ElectionMisinfo2020 and the American National Election Study (ANES). The ANES contains data that is used to assess political polarization's effect on changes in political beliefs (Joo & Fletcher, 2020). However, the most important data will come from original sources related to the research.

For the purposes of this study, cross-sectional egocentric network data using survey design will be utilized to collect measurable data. This data will be used to examine any influence of interpersonal exposure to election fraud-related disinformation on the belief in false claims. Egocentric data will help draw links between individual beliefs and the beliefs and ideological profiles of their named contacts or alters. Egocentric modeling in survey design treats disinformation as a relational outcome that captures influence from one's immediate social environment. This research design reflects that of a 2020 study conducted by Won-tak Joo and Jason Fletcher, in which they utilized data from the General Social Survey and the American National Election Study to assess the relationship between individual political beliefs and their social contacts. Their sample included thousands of respondents, some of whom completed post-election interviews (Joo & Fletcher, 2020). Using a well-representative sample, they utilized a survey instrument in which respondents were asked about their political beliefs, their social networks, and their opinions on their own connectedness, or lack thereof (Joo & Fletcher, 2020). This model will reflect theirs, but with a few alterations.

For the purposes of this study, a sample size of around 500-1,000 respondents (adjusted depending on the sufficiency of statistical power/power calculations) will be used. This sample will be stratified for ideology, geographical location, race, age, and educational attainment. These respondents will be recruited using an online survey platform, such as Qualtrics, in which the researcher can control the diversity in their sample. The egocentric network data collection series will involve two phases: a name generator (for ego network elicitation) and name interpreter questions. In the first phase, respondents will be asked to identify six people with whom they regularly discussed politics related to the 2020 U.S. presidential election in the past year. Respondents' prompts may look something like this:

"Thinking about the people in your social circles with whom you discuss politics, please list up to six individuals with whom you have discussed the 2020 U.S. presidential election in the past 12 months. These people can include friends, family, coworkers, or other close contacts.

Please use initials, nicknames, or simple descriptions to identify each of them (i.e., 'mother', 'best friend', 'Selina from work')."

This model is commonly used in these types of analyses, such as those in McPherson et al., 2001 and Bearman & Paragi, 2004. The next phase is name interpreter questions, during which each respondent will have the opportunity to answer multiple questions containing different themes:

1. Potential belief in disinformation
    a. "Based on your prior knowledge, does [insert name here] believe that the 2020 presidential election involved voter fraud, and was stolen from candidate Donald Trump?"
    b. Answers: Yes, No, or Unsure.
2. Political Ideology
    a. "To the best of your knowledge, what is [insert name here]'s political affiliation?"
    b. Scale: Very liberal, somewhat liberal, moderate, somewhat conservative, very conservative.
3. Strength of contact ties
    a. "How close do you believe you are to [insert name here]?"
        i. Scale: Close, somewhat close, very close
    b. "How much do you trust [insert name here]'s thoughts and opinions about U.S. politics?
        i. Scale: Not at all, a little, somewhat, a lot
    c. "How often do you discuss U.S. politics with [insert name here]?"
        i. Scale: Rarely, monthly, weekly, daily, more than once daily

Once respondents have had the chance to report their answers to these questions, they will next be prompted to complete a self-report of their own. This will include belief in disinformation (i.e., “do you believe that the 2020 U.S. presidential election was stolen from Donald Trump?”). their political affiliation/party identification (ideological self-placement), news media exposure (i.e., “which news sources do you receive your political news from?”), and their demographic characteristics (race, age, gender, etc.). The overall design might look something like this:

DV: Ego’s belief in election fraud; Measurement: 5-point Likert scale

IV: Number of contacts sharing the ego’s ideology; Measurement: Count (0-6)

Moderating variable: The percentage of contacts sharing the ego’s ideology; Measurement: Proportion or binary (100 = max homophily)

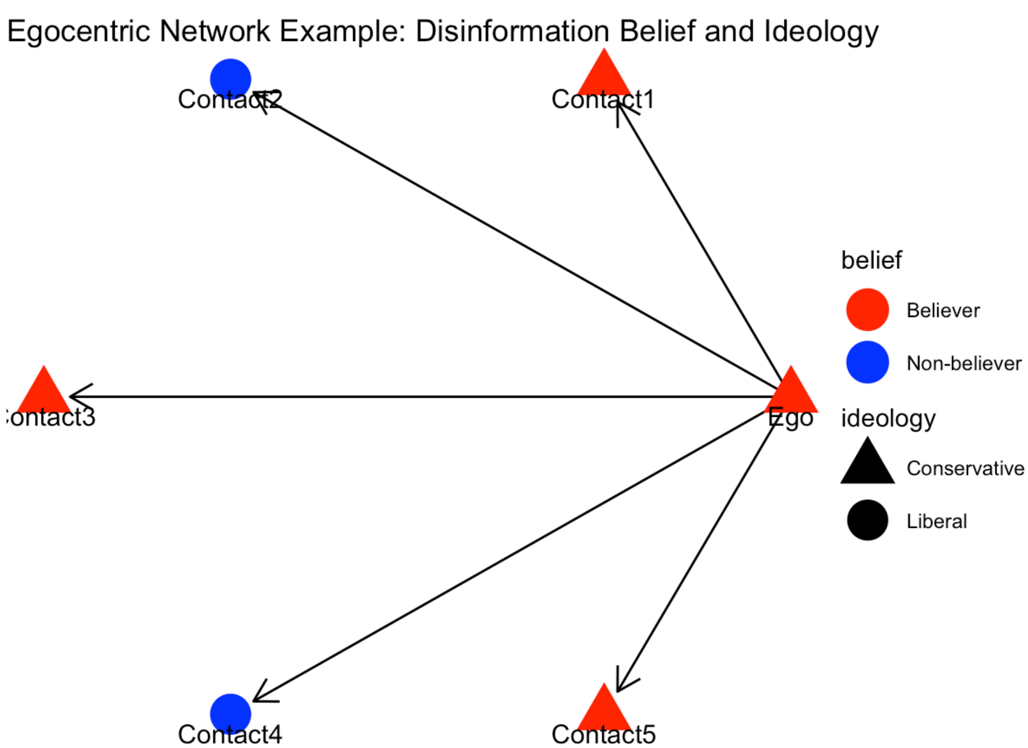
Control variables: Ego ideology, political partisanship, related demographics, usage of media; Measurement: Standardized items, may include recodes

Tie strength evaluation: The average closeness/trust of contacts; Measurement: Mean across all contact ratings

Using R, the researcher will first evaluate summary statistics regarding the distribution of belief in disinformation, number of contacts believing disinformation, political composition of networks, and ego-contact ideological lineup. Next, using logistic regression modeling, there will be four phases: the first model (exposure only; number of contacts believing disinformation), the second model (adding in homophily measure), the third model (interaction terms; exposure x homophily), and the fourth model (robustness checker; adding controls for tie strength, closeness/trust, and media sources/use). The relevant code may look something like this:

```r
# building models
model1 <- glm(belief_ego ~ num_believing_contacts, data = network_data, family = binomial)
summary(model1)
model2 <- glm(belief_ego ~ num_believing_contacts + prop_homophily, data = network_data,
              family = binomial)
summary(model2)
model3 <- glm(belief_ego ~ num_believing_contacts * prop_homophily, data =
                network_data, family = binomial)
summary(model3)
model4 <- glm(belief_ego ~ num_believing_contacts * prop_homophily +
                ideology_ego + media_trust + age + education + gender + race,
              data = network_data, family = binomial)
summary(model4)

# plotting
library(interactions)
interact_plot(model4, pred = num_believing_alters, modx = prop_homophily,
              plot.points = TRUE, interval = TRUE)

# optional diagnostics
library(ResourceSelection)
hoslem.test(network_data$belief_ego, fitted(model4))  # Hosmer-Lemeshow

library(car)
vif(model4)  # multicollinearity check

```

<u>Network Visualization & Analysis</u>



Using R, a preliminary example of what network visualization might look like was created. This visualization consists of ego nodes surrounded by 5 contact nodes. The nodes are color-coded according to respective belief in disinformation (red color = believes, blue color = doesn't believe). The nodes are shaped according to ideology, where triangles represent conservatives and circles represent liberals. The ties represent political discussions. What is important to note is that the network structure should reflect the concept of "echo-chambers", which are an expected structure in this study. Doing this will make evaluating homophily and exposure much easier. The logistic regression model will be used to evaluate H1 and H2 – the interaction term will evaluate whether the effect of exposure depends on there being similar ideologies within the network. The optional diagnostics will include checks such as the Hosmer-Lemeshow goodness-of-fit test and multicollinearity (Hosmer et al., 2013).

Conclusion

Political disinformation campaigns studied in the context of social network analysis can produce many favorable outcomes. Findings from studies like this may have implications that will allow researchers to better understand who is most vulnerable to disinformation and why. This is important because this knowledge can be used to thwart the efforts of future disinformation campaigns and target vulnerable consumers with the necessary tools and skills to spot disinformation. These tools are already known – however, targeting the necessary populations is a difficult task. One must know the intricacies of how disinformation spreads within social groups and which characteristics are most prominent in those groups. Focusing on name generator egocentric network design, one can bridge the gaps within methodological and theoretical approaches to understanding how disinformation is spread from person to person.

Works Cited

Bearman, P. S., & Parigi, P. (2004). Cloning headless frogs and other important matters: Conversation topics and network structure. Social Forces, 83(2), 535–557. https://doi.org/10.1353/sof.2005.0001

Broda, E., Strömbäck, J. (2024). Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review. *Annals of the International Communication Association, Volume 48, Issue 2*, 139–166, https://doi.org/10.1080/23808985.2024.2323736

Enjolras, B., & Salway, A. (2022). Homophily and polarization on political twitter during the 2017 Norwegian election. *Social Network Analysis and Mining*, 13(1), 10-. https://doi.org/10.1007/s13278-022-01018-z

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). *Wiley*. https://doi.org/10.1002/9781118548387

Joo, W. T., & Fletcher, J. (2020). Out of Sync, Out of Society: Political Beliefs and Social Networks. *Network science (Cambridge University Press),* 8(3), 445–468. https://doi.org/10.1017/nws.2020.2

Kennedy, I., Wack, M., Beers, A., Schafer, J. S., Garcia-Camargo, I., Spiro, E. S., & Starbird, K. (2022). Repeat Spreaders and Election Delegitimization: A Comprehensive Dataset of Misinformation Tweets from the 2020 U.S. Election. *Journal of Quantitative Description: Digital Media*, 2. https://doi.org/10.51685/jqd.2022.013

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Parker, A., Pallotti, F., & Lomi, A. (2022). New Network Models for the Analysis of Social Contagion in Organizations: An Introduction to Autologistic Actor Attribute Models. Organizational Research Methods, 25(3), 513–540. https://doi.org/10.1177/10944281211005167

Robins, G. (2015). Doing social network research: network-based research design for social scientists. *Los Angeles: Sage Publications Ltd*.

Appendix

Network Visualization Plot Information/Code

```
# packages
install.packages("igraph")
install.packages("ggraph")
install.packages("tidygraph")
install.packages("ggplot2")
library(igraph)
library(ggraph)
library(tidygraph)
library(ggplot2)
```

```
# create nodes
nodes <- data.frame(
  name = c("Ego", paste0("Contact", 1:5)),
  belief = c("Believer", "Believer", "Non-believer", "Believer", "Non-believer", "Believer"),
  ideology = c("Conservative", "Conservative", "Liberal", "Conservative",
               "Liberal", "Conservative")
)

# create edges (from ego to contacts)
edges <- data.frame(
  from = rep("Ego", 5),
  to = paste0("Contact", 1:5)
)

# igraph object
g <- graph_from_data_frame(d = edges, vertices = nodes, directed = TRUE)

# convert to tidygraph
g_tbl <- as_tbl_graph(g)

# plot
ggraph(g_tbl, layout = 'circle') +
  geom_edge_link(arrow = arrow(length = unit(4, 'mm')), end_cap = circle(4, 'mm')) +
  geom_node_point(aes(
    color = belief,
    shape = ideology
  ), size = 8) +
  geom_node_text(aes(label = name), vjust = 1.5, size = 4) +
  scale_color_manual(values = c("Believer" = "red", "Non-believer" = "blue")) +
  scale_shape_manual(values = c("Conservative" = 17, "Liberal" = 16)) +
  # triangle = 17, circle = 16
  theme_void() +
  ggtitle("Egocentric Network Example: Disinformation Belief and Ideology")
```