

## Network Theory Final Project: Homeschooling, Social Connection, and Clustering

Coco Louie

### Overview and Background.

The proposed network theory investigates the role of social and community connection in the rise in homeschooling. Homeschooling's growth rate has exploded in recent years, beginning with the onset of the pandemic but sustaining through the most recent school year. A Washington Post investigation collected some of the most substantial and recent data on the topic reveals several interesting pieces of information. First, increases in homeschooling are consistent across demographics, with no one socioeconomic class or racial groups choosing to homeschool at a significantly higher or lower rate. Correlation with religion has significantly reduced from before the pandemic, and the divide between Republicans and Democrats is about even. Additionally, the Washington Post found no correlation between school district quality (measured by standardized test scores) and home-schooling growth.<sup>1</sup> The decision to homeschool, the Washington Post concludes, is "no longer driven by shared ideology and political goals but by circumstances specific to individual families."<sup>2</sup> Since demographic factors do not appear to have an impact of the likelihood of homeschooling, examining social network features as potential predictors is well motivated.

One concern with homeschooling is that it removes children from a common, often diverse (or at least more diverse than their household) social environment. A child who is attending public school and feels connected to their school community but then switches to homeschooling has higher potential to experience adverse social effects than a child more

---

<sup>1</sup> "Home Schooling's Rise," Washington Post.

<sup>2</sup> "Home Schooling Today," Washington Post.

isolated from the school community. This analysis agrees with survey results that list social concerns over school environment and bullying to be key decision-making factors. Given the assertion that higher community connection would serve as a barrier to switching to homeschooling, it follows that less central households would be more likely to homeschool.

A second possibility is that higher rates of homeschooling lead to decreased centrality for all households. Anecdotally, public schools (especially in smaller communities) function as a community gathering space and promote community cohesion. When large groups of people decide to homeschool, they remove themselves from a primary community space and source of new social connections. Furthermore, many homeschooling families join micro schools or co-ops, which tend to group along ideological lines, furthering in-group cohesion – this could mean that higher rates of homeschooling lead to more clustering. A related effect is seen in on the evolution of networks and political attitudes.<sup>3</sup> They find significant tendency for individuals to shift their political views towards the views of their associates, providing grounds for the theory that relatively isolated homeschooling groups might experience increased cohesion and conformity.

### **Theory.**

The hypothesis is that a household with lower eigenvector centrality would be more likely to switch to homeschooling and that, likewise, communities with higher overall rates of homeschooling would lower centrality per household and lower density. Links would be defined as a “significant social connection” outside of the immediate household. Significant, as specified as the network survey questions, would be the 5-8 people whom an individual most frequently

---

<sup>3</sup> Lazer, et al.

speaks to over the course of a typical week. Defining the nodes is surprisingly complex, as though the decision to homeschool is made as a household, each individual has the independent ability to make social connections. Defining nodes as individuals (with each household then being a complete network) captures the full network picture but creates complication when analyzing the outcome. To solve this, analysis is best done by averaging the clustering coefficients of all household members and assigning that value as the “household” clustering before performing outcome analysis.

As outlined above, there is potential for bidirectional correlation between network structure and homeschooling rates. Note that the proposed theory, data collection, and data analysis process does not seek to isolate the direction of causality. In communities where large public schools serve as a significant source of social connections for students and families (this is more likely to be the case smaller communities and for families with younger children), a household with lower centrality would be more isolated from the school system, reducing social and community barriers to homeschooling. If this is the case, social network ties would be a significant predictor of homeschooling when there is little evidence as to any consistent predictive factors.

Alternatively, the decision to homeschool removes a household from the school community, eliminating a significant source of new social links and potentially deprecating some existing links over time. This alone would have the effect of increasing clustering. An additional effect that could occur is that households spend more time with households to whom they have existing strong social links, increasing the links between individuals in the two households and therefore increasing clustering. This effect would only be compounded by ideological clustering. If homeschooling does result in higher clustering, the significance of the result lies in the

reduction in network level cohesion (which, though an imperfect measure, does correspond to cohesion)<sup>4</sup> and corresponding increase in in-group cohesion, especially at a time when polarization and community fracturing is on the rise.

### **Data Collection.**

Network data would be collected through a survey distributed to all members of households with school aged children and restricted by school district. Parents would be emailed and/or mailed with the survey, and children could be sent home with flyers to increase responses. Younger children (kindergarten through third grade) would be excluded from the survey for ease of distribution and data integrity. It is unlikely that a kindergartener would be able to identify their social contacts in a useful way without the assistance of an adult familiar with their social life, and introducing the assistance of such an adult, whether it be a parent, teacher, or other, biases the data towards the group that adult is most familiar with. Additionally, in the case of younger children, the answers of parents would be more representative of household connections than with older children. Districts track which families homeschool, so surveys would be mailed to homeschoolers, ideally wrapped into some of the existing mandated data collection to increase response rate. The survey would be formatted as such:

1. List 8 (minimum 5) people who you most frequently speak to outside of your immediate family in a week. Please provide first and last names.
2. List the 3 families/households (outside of your immediate household) who you most frequently interact with in the typical week. Please provide the full names of as many members of each family as you can.

---

<sup>4</sup> Moody, et al.

The following questions would be distributed only to parents.

3. Please list the ages and grades of your child(ren).
4. What are your education plans for your children next year?
  - a. I plan to place my child(ren) in traditional school.
  - b. I am very likely placing my child(ren) in traditional school.
  - c. I plan to homeschool my child(ren).
  - d. I am very likely to homeschool my child(ren).
  - e. Neutral/Undecided.
  - f. Other: \_\_\_\_\_

Question 2 would be used to check the success of the first question at accurately locating between-household connections. Questions 3 as a basic control and additional path of investigation regarding predictors and reasons for homeschooling.

The survey would be distributed 4 times across the course of 2 years, once each at the beginning and end of the school year. For outcome data, data on district level homeschooling rates is regularly collected by school districts. Data would be needed corresponding to each time the survey was conducted. Question 4 of the survey collects data on the homeschooling decisions of each separate household.

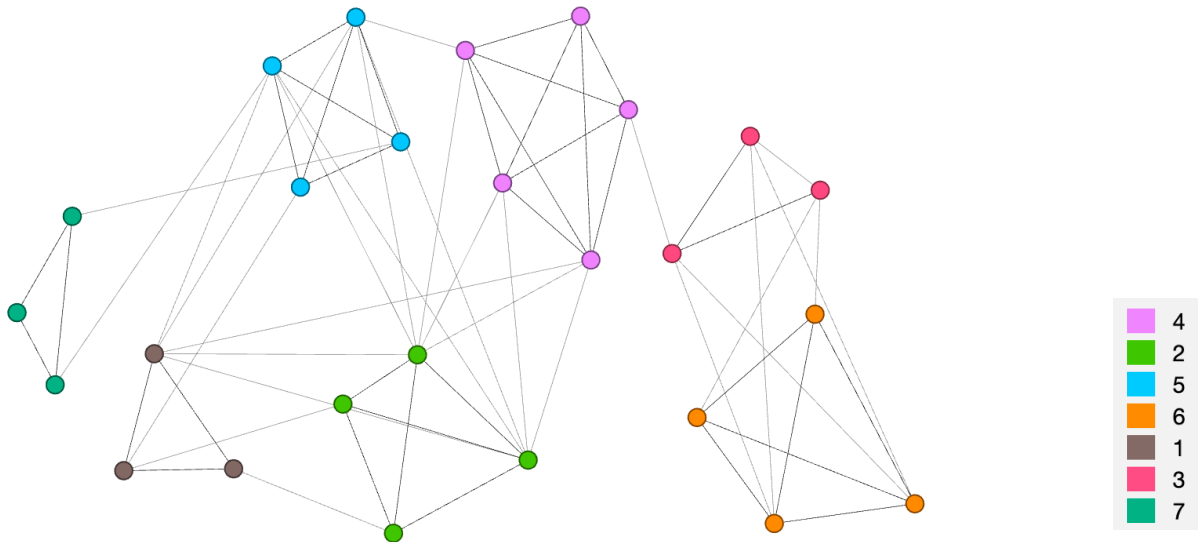
Optimally, data would be collected for as many districts as possible covering the complete range of demographics across the country. If that was not possible (as it would require a vast number of resources), data would be collected in 3+ school districts as close in demographics as possible. The most critical variables to control for would be several measures of size, including population density, proportion of families with school aged children, and number

of public schools vs. private schools. Other variables, like race, politics and quality of schools would be good to control for, but as the Washington Post data shows, not shown to have a significant effect on homeschooling – though they would likely have an impact on the network shape.

The relationship between clustering and homeschooling has the potential to be tested on both the node and network level. At the node level, under the hypothesis that households with lower centrality are more likely to homeschool, the eigenvector centrality of each household would need to be obtained. This approximated be done by averaging the eigenvector centrality of individuals in a household and combined with the data from survey question 4. From there, a simple regression analysis could be performed. A regression would also be performed including the average age of children in each household, as this could be a significant factor in the decision. At the network level the hypothesis is that higher rates of homeschooling lead to higher network clustering. To test this, several district level homeschooling rates and clustering coefficients would need to be compared.

### **Proof of Concept.**

All data is completely fabricated. Though district level data on homeschooling exists, much of the proposed analysis requires data on homeschooling choices at the individual level, so it is more easily representative to fabricate all the data.

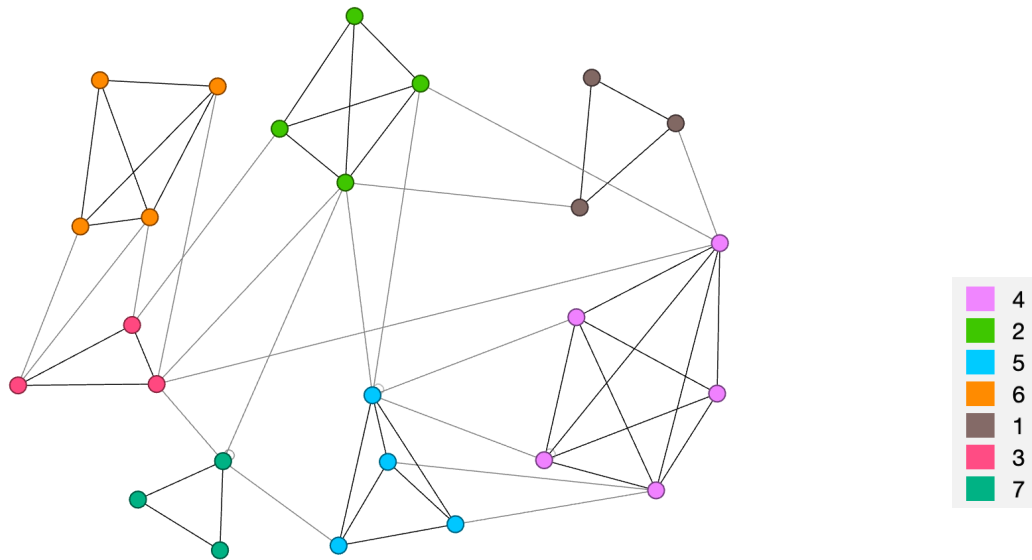


*Fig. 1.* Example of a network where lower eigenvector centrality corresponds to a higher likelihood of homeschooling (see table). Node colors correspond to household number (right).

<i>Household</i>	<i>Homeschool</i>	<i>Eigen_Avg</i>	<i>AgeChildren_Avg</i>
1	0	0.436787874	7
2	0	0.687508528	9
3	1	0.069240739	12
4	0	0.598794893	10.33333333
5	0	0.571893057	6
6	1	0.041759327	13.5
7	1	0.182946779	15

*Table 1.* Example of network data collected for regression analysis, extracted from the network shown in figure 1.

As is shown in the table, households with lower average eigenvector centrality are more likely to homeschool their children. Visually, these households are also more isolated from the main network. With additional data, a multivariable regression could be run to test the hypothesis more rigorously.



*Fig. 2.* Illustration of a district with lower clustering (0.4386792) and theoretical higher rates of homeschooling. Compare to figure 1 with clustering 0.5145631.

Beyond household eigenvector centrality and network clustering, node clustering is also worth investigating. Initially, the hypothesis was that higher clustering would correspond a higher likelihood of homeschooling due to more resilient inter-household social connections (that would therefore be less dependent on participation in the school community), but the hypothesized network (fig 1) shows that this is not the case. In the example network, individual node clustering has no clear correlation with homeschooling, although it would still be worth investigating with a larger sample size.

### **Discussion of Limitations.**

The primary social condition required by the theoretical explanation is that schools function as community hubs and a significant source of social connections for both children and parents. If this is not the case in a particular community, the theory would not apply. It could be worthwhile



to include a survey question directed at testing this assumption. Furthermore, if it is shown that school's do not serve as a social hub but clustering still correlates with homeschooling, it is unclear as to where the correlation may come from. The model may also deprecate in larger, more densely populated communities, as people typically have more diverse sources of interaction in these scenarios.

Additionally, geography could play a substantial role in determining the strength of inter-household connections. Consider the concept of a "family friend." Though this is often the result of a separate shared community, like a church or school, it is often also the result of residential proximity. It is easier to see a group of people frequently when they live next door. Furthermore, the choice of community spaces, like schools and churches, is often at least partially determined by geography. Though this does not necessarily diminish the validity of a social network analysis approach, it does suggest that a model that doesn't consider the impacts of geography might fail to capture a critical effect.

Finally, as briefly noted previously, though demographic factors have not been shown to impact homeschooling decisions, they may have an impact on network shape. A more robust analysis would collect demographics along with survey data or join the survey data with additional demographic information. This, however, is resource intensive, and controlling for demographics at the network level may be sufficient for network comparison.

## References

- “Home Schooling Today Is Less Religious and More Diverse.” The Washington Post, September 26, 2023. <https://www.washingtonpost.com/education/2023/09/26/home-schooling-vs-public-school-poll/>.
- “Home Schooling’s Rise from Fringe to Fastest-Growing Form of Education.” The Washington Post, October 31, 2024. <https://www.washingtonpost.com/education/interactive/2023/homeschooling-growth-data-by-district/>.
- Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz, and Michael Neblo. “The Coevolution of Networks and Political Attitudes.” *Political Communication* 27, no. 3 (August 4, 2010): 248–74. <https://doi.org/10.1080/10584609.2010.500187>.
- Moody, James, and Jonathan Coleman. “Clustering and Cohesion in Networks: Concepts and Measures.” *International Encyclopedia of the Social & Behavioral Sciences*, 2015, 906–12. <https://doi.org/10.1016/b978-0-08-097086-8.43112-0>.

## Supporting Information

Data, Gephi files, R code can be found here:

<https://github.com/clelouie/NetworksandPolitics/tree/main/finalProject>