

Social Network Analysis: Midterm

Name: Shreyans Kothari | **UNI:** SK4819

Part 1

Researcher is interested in understanding what affects one's social circle's religious diversity. Used 1985-2004 GSS modules.

Variables:

- **Pct_None** : percentage of one's social circle with no religion. Ranges from 0% to 100%. Considered a number of factors, including: a series of indicator variables for what religion ego is (Protestant=reference, Catholic, Jewish, None, and Other).
- **Educ** : how educated ego is (ranging from 0 to 20 years of schooling)
- **Age** : how old ego is (ranging from 18 to 89 or older)
- **Married** : whether ego is married (=1) or not (=0)
- **Bio_Both** : whether ego lived with both biological parents when aged 16 (=1) or not (=0)
- **2004** : what year the survey was being done (1985 is the reference vs. 2004)
- **Nungiven** : size of ego's network, ranging from 0 up to 6 people

Used the Herfindahl-Hirschman Index to predict the diversity of one's social circle. The H-H Index makes use of the proportions of each religion (whether Protestant, Catholic, Jewish, None, and Other) in someone's network.

1) In Model 1, the single biggest predictor of what percentage of ego's network will be made up of people with no religion is ... whether ego's religion is None ($B=35.11^{*}$). Tell me specifically, what that coefficient means and how it is properly interpreted in the context of multiple regression. Is it surprising that this variable is the most powerful predictor? Why? Why not?**

In Model 1, the dependent/outcome variable **None** is an indicator for whether Ego's religion is none, i.e. whether the Ego does not follow any religion.

Interpretation: Holding all other variables constant, on an average, the Ego not following any religion is associated with a 35.11 percentage points increase in the percentage of the Ego's social circle not following any religion.

None being the strongest predictor of what percentage of ego's network will be made up of people with no religion is not surprising at all. As the proverb goes, "birds of a feather flock together." People often choose to surround themselves with other people who have similar interests and ideologies. This is commonly studied in Sociology as the in-group vs out-group phenomenon: people identify more with other people who belong to the same religion (or lack thereof), nationality, sexuality, etc.

2) In Model 2, the outcome variable is the H-H Index. In general, how is this Index calculated? What does this Index tell us, compared to what the Pct_None variable tells us? What does it mean if the H-H Index is at ≈ 0 ? What if it is at 1? Give me one reason why we might prefer this measure of diversity over the Pct_None one used above? Tell me specifically, what the coefficient on None ($B = -0.101^{**}$) in Model 2 means and how it is interpreted in this case.

The Herfindahl-Hirschman Index is very commonly used by the government to evaluate how a potential merger between two companies might affect market concentration in a given industry. It is measured as the sum of the market share percentage of all the firms in the industry. Higher H-H Index is associated with a higher concentration in the market- meaning less competition, and an index closer to 0 indicates perfect competition.¹

In the context of this research, the H-H Index is measuring the religious diversity of an Ego's social circle. It is calculated as the sum of the proportion of each religion (whether Protestant, Catholic, Jewish, None, and Other) in the social network. A H-H Index closer to 0 implies that the Ego's social circle is very diverse in terms of religious identity. A larger H-H index (closer to 1) indicates more similarity (less diversity) in the social network in terms of religious identity.

Pct_None is simply the percentage of one's social circle with no religion; it ranges from 0% to 100%. The H-H Index, on the other hand, is a better measure for calculating religious diversity since it incorporates the proportion of all the religions (all in the scope of this study means Protestant, Catholic, Jewish, None, or Other) in a network. The H-H Index measure captures a lot more nuance in the relationship between Ego's religious identity and preferences, and that of their social circle. We would prefer to use H-H Index since it actually captures religious diversity in a social circle, whereas Pct_None only captures those people in the network who don't follow any religion and groups all other categories (Protestant, Catholic, Jewish, and Other) together.

Interpretation: Holding all other variables constant, on an average, the Ego not following any religion is associated with a 0.101 points decrease in the H-H Index, i.e., the Ego not following any religion is associated with the Ego's Social Circle being more religiously diverse by 0.101 points.

¹ <https://www.investopedia.com/terms/h/hhi.asp>

3) Why is it important that the researcher control for ego's degree, via Numgiven, in the models? In this case, why might the researcher think that Numgiven and Pct_None, or Numgiven and the H-H Index would be related? Provide one explanation.

It is important for the researcher to control for ego's degree/size because the size could very directly influence the religious diversity of one's social circle: more people you know, more likely you are to know people who follow different religions than you, and less people you know, less likely are you to know people with religious preferences different from yours.

In both Model 1 and Model 2, the researchers are right to include Numgiven because this allows them to exclude the possibility that the religious diversity or the proportion of people not following any religion in a social circle is being influenced by the size of that network. The properties of regression allow us to hold numgiven constant as we analyze the effects of other variables (like None, or Educ, etc.) on the outcome variable. If we didn't include Numgiven, we would have failed to account for how the relationship between the outcome and the independent variables could differ for Egos with different degrees/sizes of social networks, solely because of having more/less people in the network. We can see from the summary statistics that the size of the network varies from 0 to 6: there are some people who don't have anyone in their network, and there are others who have up to 6 people listed in their network.

Part 2

Researcher interested to test whether structural holes matter to scientists. Data from an early experiment which used computer mediated communication- 31 academics allowed to contact each other using an EIES. Acquaintance information sociomatrix code:

- 1 = close personal friend or friend
- 0 = a person I've met, or a person I've heard of but not met, or a person unknown to me
- A tie was considered present if either of the two people nominated each other as a friend of some kind; the tie did not need to be reciprocated

Variables:

- **No. of Cites** : total number of citations each academic received on their papers. This is the outcome variable.
 - **Constraint** : Based on Burt's formula
 - **Density** : Alternative predictor. Ego-network density.
 - **Sociology** : Control variable. Indicator for academic discipline- Sociology. Reference.
 - **Anthropology** : Control variable. Indicator for academic discipline- Anthropology
 - **Statistics** : Control variable. Indicator for academic discipline- Statistics
 - **Other** : Control variable. Indicator for academic discipline- Other
 - **Degree** : Ego-networks degree. Ranges from 0 to 31
-

4) Explain to me what structural holes are. Why does Burt think they are so important? Why is his measure of "constraint" a good proxy for what structural holes are meant to be? In the context of sending messages and getting papers cited, do you think having lots of structural holes would be useful to scientists here? Why? Why not?

Structural holes are non-redundant ties that connect disparate social clusters. In social networks, structural holes exist when there is an absence of a direct tie between two nodes/entities. The theory behind structural holes was developed by Ronald Burt in the context of good ideas and within-firm competition. According to Burt, people/nodes within a group are more likely to think homogeneously in the presence of structural holes since there

would be a restriction on free flow of information between different groups. In tightly clustered networks, there would be redundancy in terms of ideas and information. Following this logic, people connected across groups are in a beneficial position to get access to alternative ways of thinking. An individual who is able to form a brokerage between different clustered groups could facilitate the transfer of information that seems mundane to one group but innovative to another.¹

Burt's study concluded that Managers that operated near (bridged/brokered) structural hole(s) had the greatest chance of having "good ideas". Managers who had a lot of contacts outside their immediate work groups were the ones whose ideas were ranked highest. These managers were well positioned in their company because of the structural holes in their networks that restricted flow of ideas between groups, except for through them. The manager existing near a structural hole, that is, a manager well-connected with people outside of their group also have less duplication of information because of their connection to unique (non-redundant) nodes.

The constraint of a network can be calculated as the sum of each connection's constraints. It measures the extent to which your network time and energy is concentrated in a single group. Constraints are a good measure of structural holes because they directly capture how much an ego can negotiate/exploit potential structural holes in their network for their own benefit. An ego who acts as a bridge (in the case of less constraints and more structural holes) will be better off than other people in their network because of informational benefits and *tertius gaudens*; more people would depend on the ego for connections to other groups and different (read: new) ideas.

In the context of sending messages and getting papers cited, I think having lots of structural holes and less constraints would be very useful to scientists since the ideas that a well-connected scientist would get access to from one group would be "unique" for other groups. I suspect scientists to have clustered networks- scientists from an academic field would be more likely to be connected closely with other scientists from that same academic field. In the presence of structural holes, the network would be less-redundant (and sparsely connected), which means less people would be sharing ideas across academic disciplines/departments. Any scientist that is in a beneficial position of "bridging/brokering" between unconnected groups would get access to innovative/new ideas that they could publish articles on. This would get those well-connected / brokering scientists more citations since there would be limited journal articles on that new topic in their department. However, the structural holes would ONLY be useful for those scientists that act as the bridge between the unconnected groups.

¹<https://medium.com/@agreenmoment/introduction-to-structural-holes-theory-124c51c3ae31>

5) Look at the coefficient on Constraint in Model 1. Does the researcher find support for the structural holes argument here? Explain your answer, and please interpret that coefficient fully as well.

In Model 1, the coefficient on Constraint seems to be statistically significant with p-value

less than 0.05. The coefficient is negatively associated with No. of Cites; less constraints are associated with more citations.

Interpretation: Holding all other variables constant, on an average, an increase in the constraint measure in a scientist's network is associated with a decrease in the total citations the scientist receives, by 41.44 citations. For every additional unit of constraint in the scientist's network, the total number of citations they receive on their papers decrease by 41.44 citations.

Yes, the researcher does indeed find support for the structural holes argument here. The constraints in a scientist's network are negatively associated with the total number of citations they receive on their papers. If a scientist has a less constrained network, they are able to get access to different academic fields and obtain unique ideas that are not commonly discussed in their own field. Writing papers on these unique ideas allows them to garner more citations on these innovative journal articles.

6. Look at the coefficient on Density in Model 2. What is Density? Explain how is it calculated in ego networks that are symmetric and binary, like this one? Would we think, given Burt's logic, that Density should produce opposite effects on the No. of Cites, as Constraint does, or the same effect? Explain your answer, and please interpret that coefficient fully as well.

Density in an ego-network is the proportion of how well-connected the network. In the context of this study, the researcher considers a tie to exist between two scientists if any scientist nominates another scientist as a friend of some kind (=1), even if the tie is not reciprocated. In this case, the density would be calculated as:

$$\text{Ego-Network density} = \text{Actual ties} / \text{All possible ties}$$

Given Burt's logic, I would expect density to be negatively correlated with No. of Cites: as the network density increases, I would expect scientists to have less total citations on their papers. In the situation of high density, more of the scientist's peers would be connected to each other, which would facilitate increased exchange of ideas between different departments/groups and the scientist in question (ego) would lose their beneficial position of bridging / brokering structural holes (demonopolize their access to new and unique ideas). More connected alters would inflict a higher constraint on the ego.

Following this logic, constraint and density should have a similar effect on No. of Cites. This is evident from Model 2 as well: the coefficient for density is negatively related with the outcome variable No. of Cites.

Interpretation: Holding all other variables constant, on an average, every unit increase in a scientist's network density from being closer to 0 to closer to 1 (i.e., alters more connected to each other), the total number of citations that the scientist receives would reduce by 24.62 citations. As explained above, more connections between the alters would result in the ego not being able to take advantage of any structural holes in the network. A more dense network would facilitate free flow of ideas and information between nodes from different

departments/disciplines/groups. No one node would be able to act as a bridge/broker between nodes and utilize their connections to get access to ideas that others their groups in a sparse network would not have had access to.