Problem Set #2 Question 3

MACS 30000, Dr. Evans
Nan Ge

We like to hang out with friends because we have something in common to talk about, but why do we make friends with our like in the first place? This relates to the classic inquiry about the source of homophily. Two mechanism have been proposed, namely, choice homophily and induced homophily. Choice homophily corresponds to an individualistic view of the world, suggesting that our friend circle is a result of individual preference. We choose to become friends with those who share more similarities, because we are more comfortable with them. Induced homophily, however, states that our choice of friendship is merely the result of structural proximity. The environment constrains our choice set; shared environment leads to similar plausible choice sets. In the end, population of a shared group would display higher level of homogeneity. In their 2009 paper, Kossinets, Gueorgi and Duncan J.Watts asked the question: what is the dominant source of homophily, individual preference or structural proximity?

In most cases, individual preference and structural proximity act together. To disentangle these two forces, this paper exploits a longitudinal dataset covering 30,396 undergraduate and graduate students, faculty, and staff in a large U.S. university. The research data is merged from three sub-datasets: (1) the logs of e-mail interactions within the university over one academic year, (2) a database of individual attributes (status, gender, age, department, number of years in the community, etc.), and (3) records of course registration, in which courses were recorded separately for each semester. There are 30396 individuals. Each is paired with 17 variables consisting of personal characteristics, organizational affiliations, course-related variables, and e-mail related variables. There are 7156162 e-mails exchanged over 270 days of observation. A detailed definition of all variables could be found in Appendix A. Table 2 and Appendix C give a statistical description.

A lot of tough decisions have to be made to deal with such a sophisticated dataset. One of the decisions is that the authors narrow their dataset to those sent to recipients inside the university. Therefore, the evidence presented by the data could only exhibit relationship formation pattern inside the university. As the authors also point out in the discussion part, population in the university is much more homogeneous than other groups. Therefore, the processing method could strengthen the power of structural proximity.

To disentangle the forces of individual preference and structural proximity, we need to measure similarity and shared foci between a pair. The authors construct a variable *similarity* that is defined as the number of matching items between two individuals out of the following characteristics: *gender, status, field, age, year, and from U.S.* To measure the impact of structural proximity, the authors specify explicit and implicit social foci. An explicit social focus is measured as sharing at least one class in the dataset. However, explicit social foci might be biased towards students, and neglecting other potential shared environments. An implicit social focus describes situations where two or more individuals are connected through other focal activities, such as “social groups, sporting and cultural organizations, shared housing, and so forth”. The problem is dataset doesn’t provide any straightforward measurement for implicit foci, and leaves the impact of potential shared environment indistinguishable. The authors address this problem by mining the available bulk email (where one email has multiple recipients) data. They quantify the “strength of shared membership” for every pair of individuals i and j as the number of times they appear together among the recipients of a bulk message in one semester. Then they set a threshold g*=140, and define that a pair has “strong” implicit foci if the pair’s bulk email exchange exceeds g*.

The paper explores further about the formation of relationship in a triadic closure, the dissolution of ties. They also looked into the compounding effect of choice homophily and induced homophily after generations, which might have more propounding policy implications.