

# The Oxford Summer School in Economic Networks

Sheet 2 — June 24, 2025

## Social networks tutorial

The paper on *The Diffusion of Microfinance* by Bannerjee et. al studies how centrality measures of village leaders (those who were initially seeded with information about microfinance by Bharatha Swamukti Samsthe (BSS)) are related to the adoption of microfinance. Here we recreate some of their analyses.

1. From `adj_allVillageRelationships_HH_vilno_1.csv`, read in the adjacency matrix of village 1. This is the household network that tells us how households are connected in village 1. Explore the network structure:
  - (a) What are the average number of connections in the network? How closely connected are the nodes (or households) in the network? Can you think of some other ways to explore the structure of the network?
  - (b) What do these measures look like across all 75 villages? Do you observe any assortative mixing in the network?
2. From `householdcharacteristics.csv`, we add demographic information at node (or household) level. Explore the centrality measures of nodes based on their demographic characteristics. For instance, do households with leaders (given by `leader status = 1`) have different centrality measures than non-leader nodes?
  - (a) Do you observe any assortative mixing based on node degrees?
  - (b) **BONUS:** Using node attributes, can you find out whether the network is assortative in some other ways (e.g., people of similar socioeconomic backgrounds are likelier to be connected in the network)?
3. We now define a simple diffusion model. Here diffusion of microfinance takes place in two steps. Step (1) is the participation (or adoption) step —where newly informed nodes decide whether they want to participate in the microfinance programme or not. The probability of participation is measured based on their own demographic characteristics  $P_C$  and the total number of participating neighbours  $N_P$ , using a logistic model:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 P_C + \beta_2 N_P)}}. \quad (1)$$

In the first instance, we remove any influence of demographic characteristics on participation decision,  $\beta_0 = \beta_1 = 0$  and  $\beta_2 = 0.1$ .

Step (2) is the transmission process where participant nodes and informed non-participant nodes transmit information to their neighbours with probabilities  $q_P = 0.05$  and  $q_N = 0.01$  respectively.

You can start the diffusion process in the network by informing (or seeding) 5% of the network. What strategies are useful in selecting the initial set of informed nodes (i.e. seed set)? How do these compare to a randomly selected seed set? Which of these strategies perform the best in maximising participation in the network (i.e. share of participants in the network)?