

# Observing & Measuring Social Interactions & Peer Effects

Rogayeh Dastranj Tabrizi

email: rda18@sfu.ca

Office: WMC 3607

Office Hours: Thursdays 12pm-2pm

Department of Economics

Simon Fraser University

30 March 2015

# Introduction

So far we focused on:

- What we know about social and economic networks,
- How we model and analyze social networks and their impact.

In this lecture we study **measurements of social networks** as well as **inference** from them.

- A challenge in working with social networks is that the structure of the network is generally **endogenous**,
- And it is related to many characteristics of agents involved.

# Identification

To establish that a certain behaviour is influenced by social network structure:

- Properly account for many other related characteristics which could be the driver of the behaviour,
- Properly account for the fact that there can be feedback between the social structure, behavior, and the background characteristics
- Properly specify a model to be able to sort out various effects
- Also, somewhat related question of when it is that we can deduce that social structure **cause** behaviour and not the other way around.

# Identification

- Be sure that these different characteristics are **Identifiable** and can really be sorted out given the data.
- **Identification**: isolating the impact that social networks have on various behaviours

## Identification in Peer Effects

It can be difficult to sort out whether individuals are behaving in a certain way because of:

- The direct influence of their neighbours
- Some influences are common to them and their neighbours
- Other factors that are related to their network position

# Specification & Omitted Variables

- Suppose that the degree in a network is correlated with age: **Older people have more connections**
- Suppose one **hypothesize** that **People with higher degree behave differently from people with lower degree.**
- Estimate the following equation:

$$Y_i = \alpha_d + \beta_d d_i + \epsilon_i \quad (1)$$

- $\alpha_d$  is a base degree
- $Y_i$  is a measure of the behaviour that we are interested in explaining,
- $d_i$  is individual  $i$ 's degree,
- $\epsilon_i$  is an error term that captures unobserved idiosyncratic factors.

# Omitted Variables

- Now suppose that degree varies with age, so that:

$$d_i = \delta + \gamma age_i + \nu_i \quad (2)$$

- $\nu_i$  is an idiosyncratic term.
- But if the **True Relationship** is:

$$Y_i = \alpha_a + \beta_a age_i + \nu_i \quad (3)$$

- Then by estimating eqn (1), we end up finding an estimate for the coefficient on degree,  $\beta_d$ , that looks **significant**, but is really just a **proxy for the effect of age**, which is filtered through degree through eqn (2).

# Omitted Variables

Now suppose that **True Relationship** is

$$Y_i = \alpha_d + \beta_d d_i + \beta_a age_i + \epsilon_i \quad (4)$$

So that both degree and age affect  $Y_i$ . For example:

- Suppose that we are examining a model of social capital,
- $Y_i$  is some measure of wealth or power accumulation
- We are testing for how this is influenced by a measure of social connectedness

Now if we only use equation (1), including degree but omitting age, then given that degree and age are related according to (2), we will end up with a **biased estimate of  $\beta_d$** , as it proxies for some of the effect of age.

# Omitted Variables

While such **specification** and **omitted variable** problems are not unique to analysis of social networks, they are particularly acute in social network settings **because of homophily**:

- People tend to associate with others who are similar when examined from a wide variety of perspectives
- So, if we see that behaviour of individuals tends to match patterns of the social network:
  - We **cannot attribute that behaviour directly to network influence**
  - Without being sure to properly **account for all of the other factors** that might be related to whom interacts with whom.



# Instrumental Variables

- It is difficult to know all of the potential factors that might covary with both social relationships and behaviour
- One way to view the bias is that the omission or misspecification leads the variables to be correlated with the error term.
- A standard approach to dealing with such problems is to work with instrumental variables.

## Instrumental Variables

An instrumental variable is one that is correlated with the regressor of interest (in this case degree) but uncorrelated with the error term. In this case, uncorrelated with the omitted variable age.

# Instrumental Variables

- Derive estimates based on the instruments using two-stage least squares estimation.
  - First regress degree on the instrumental variable, to develop estimated values of degree based on the instrument
  - Then in the place of degree in the original regression we use these estimated values of degree as predicted by the instrument
- These estimates are independent of the error terms as they are conditional on the instrument.

# Endogeneity

- Another problem is that social structure is often **endogenous to the things that it impacts**.
- In estimating the impact of social connectedness on wealth, it could also be that wealth impacts connectedness:
  - By providing additional access or opportunities
- This could also lead degree to be correlated with the error term:
  - Lead to biased and inconsistent estimation
- Instrumental variables can help in sorting out endogeneity issues.
  - We want an instrumental variable that is related to degree, but will not be related to the error terms.

# Endogeneity

For example:

- A government comes in and randomly chooses some villages to provide subsidized communication to (in the form of internet or telephone)
- This might enhance degree in some villages and not others
- Examine how these changes in degree for exogenous reasons influence wealth.
- Measure the induced degree differences that we attribute to the government program
- Then see the extent to which it influences wealth, using two-stage least squares, or other methods of working with instrumental variables.

# Problems with IV

- Finding **good** instruments, that are related to the problematic variable(s), but are not endogenous and not related to omitted variables and error terms, can be problematic.
- Never really be sure of this as we never directly observe the **true error term**
- Even if we find instruments that we are reasonably confident in, it can still be difficult to find ones that have a **strong enough relationship with the problematic variables** to be very useful in producing powerful estimates.

# Peer Effects & The Reflection Problem

**Identification:** meaning that the parameters of the model are **uniquely determined** by a data set.

- Manski describes an identification problem that is faced when examining social influences on individual behaviour and distinguishes between:
  - *exogenous (or contextual) effects*: the influence of exogenous peer characteristics,
  - *endogenous effects*: the influence of peer outcomes,
  - *correlated effects*: individuals in the same reference group tend to behave similarly because they are alike or face a common environment.

# Peer Effects & The Reflection Problem

To see the problem, imagine a very simple case:

$$Y_i = a + bE[Y_i|x_i] + \epsilon \quad (5)$$

- $Y_i$  is  $i$ 's behaviour,
- $x_i$  is  $i$ 's characteristics and that  $i$ 's cohort is other people who have the same characteristics.
- $\epsilon_i$  is a mean-zero error term.
- **Question: What will happen if we estimate this relationship?**
- Take expectation of both sides of this equation conditional on  $x_i$ :

$$E[Y_i|x_i] = a + bE[Y_i|x_i] \quad (6)$$

# The Reflection Problem

By substituting eqn (6) in (5) we have:

$$Y_i = E[Y_i|x_i] + \epsilon \quad (7)$$

Which is a trivial statement:

- This is what Manski refers to as the **reflection problem**:  $i$ 's behaviour is a function of the expectation of the peers' behaviours which is just the expectation of  $i$ 's behaviour which reflects  $i$ 's behaviour.



# The Reflection Problem

Even if we enrich the specification, to allow  $Y_i$  also depend on  $i$ 's characteristics, we still have an identification problem:

$$Y_i = a + b_1 E[Y_i|x_i] + b_2 x_i + \epsilon \quad (8)$$

After taking the expectation:

$$E[Y_i|x_i] = a + b_1 E[Y_i|x_i] + b_2 x_i$$

$$E[Y_i|x_i] = \frac{a}{1 - b_1} + \frac{b_2}{1 - b_1} x_i \quad (9)$$

$$\Rightarrow Y_i = \frac{a}{1 - b_1} + \frac{b_2}{1 - b_1} x_i + \epsilon_i \quad (10)$$

- We can estimate the two composite parameters if  $x_i$  is not constant.
- This does not identify the parameters  $a, b_1, b_2$ , as there are many different values of these which lead to the same composites.

# The Reflection Problem

- One basic assumption in the linear-in-means model: individuals interact in groups.
  - The population is partitioned in groups, and that individuals are affected by all others in their group and by none outside of it.
- Instead, assume each individual has his own specific reference group, defined by the individuals whose mean outcome and characteristics influence his own outcome.
  - Interactions are thus structured through a directed social network.
  - Relaxing the assumption of group interactions generally permits to separate endogenous and exogenous effects.

# Lab and Field Experiments

- Many of the challenges in empirical analysis of social interaction are in separating out effects to of different factors and behaviours
- Various forms of experiments on behaviour, where the **particular values of some variables are carefully controlled and varied**, allow one to track which conditions have been altered and thus uncover their impact.
- Given the power of experiments, they are a rapidly emerging tool of network analysis.

# Lab and Field Experiments

Experiments can be used in various ways:

- First, an experimenter can examine how social structure affects behaviour, and how an agent's position within a network influences his or her behaviour.
- Comparing behaviour across network position, and tracking changes in behaviour as one changes overall network structure, allows one to explicitly track the impact of social network structure on behaviour

# Lab and Field Experiments

Experiments can be used in various ways:

- Second, one can test alternative theories of behaviour in network contexts.
- Here one does not necessarily need to use variations in treatments within an experiment; instead one can simply find a setting where the predictions of different theories lead to different outcomes.
- For example, by changing the payoffs to network formation the experimenter should observe different networks form if agents are farsighted in their link formation rather than myopic.

# Lab and Field Experiments

Experiments can be used in various ways:

- Third, one does not necessarily have to run a horse race of different theories against each other.
- But one can simply test whether a given theory's predictions hold.
- For example, one can examine variations on undirected or directed connections model and then see whether pairwise stable or Nash stable networks form