Project Proposal: The dynamics of innovation network

CLZZ Group

April 22, 2019

1 Group Information

- **Group Name**
  CLZZ
- **Member**
  Anhua Chen (anhua@uchicago.edu)
  David Liu (dliu5@uchicago.edu)
  Xiuyuan Zhang (xiuyuan1028@gmail.com)
  Xiang Zhang (zhangxiang0822@gmail.com)

2 Data

We are using the US patent citation dataset from United States Patent and Trademark Office (USPTO), to construct an innovation network, and then to analyze the dynamics of this innovation network. Rows of this data represent each citing patent, it will provide information on patent identification, date of citation, the patent this citing patent cites, category/industry of the patent etc.

The size of the dataset is around 8.5 GB, and the number of observation is more than 2 million.

2.1 Attributes of the dataset

There are mainly 9 attributes of the data. (Details listed below)

Data Element Name	Definition
uuid	unique id
patent$_i d$	patent number
citation$_i d$	identifying number of patent to which select patent cites
date	first day of the month the cited patent (citation$_i d$)$was granted$
name	name of cited record
kind	WIPO document kind codes
country	country cited patent was granted (always US)
category	who cited the patent (examiner, applicant, other etc)
sequence	order in which this reference is cited by select patent

3 Proposed Analysis

3.1 Dynamics of patent citation

- What are the time-serial dynamics of patent citations?
- What are the cross-sectional variation in terms of both patent stock and patent flow across various sectors?

3.2 Mapping out the network

- Is there a hierarchical structure across sectors in this innovation network?
- Can we identify the "upstream" and "downstream" industries in this patent network? How does it compare with the traditional identification of "upstream" and "downstream" sectors based on production network?
- Is it possible to identify endogenously-formed clusters of inventors/firms according to patent citations?
- What does the centrality measures look like across sectors in patent citation network? Namely, are there certain "key sectors" that pose to be especially influetial to other sectors?
- What does the connected measures look like across sectors? For example, is there any sectors relatively separated from others in patent network?
- Is innovation network symmetric?

3.3 Dynamics of network structure

- Across time, are citations happening more within sectors or between sectors?
- Across time, what sectors are gaining citations and who are losing them?

3.4 Prediction power of innovation network

- How much can the structure of innovation network (centrality/connectedness) explain the differential output growth and productivity growth across sectors? (Need to pair with NBER-CES Industry economic database)
- Does structure measure of innovation network have predicting power for sectoral growth in short/medium/long run?

4 Algorithm Design

References