CASA MSc Spatial Data Science and Visualisation

Dissertation: Current Results and Ideas

Student: Xudong Zhang

Date: 17th, March 2021

content

1. **Dataset I got**
2. **Initial research questions for dissertation**
3. **Current research results**
4. **How to respond these questions**
5. **Other Questions**

Dataset I got:

- The **website of China Patent Information Center** (https://www.cnpat.com.cn) released a patent dataset to public in February, 2020.
- China Patent Information Center and another institute, jointly developed the sharing platform of epidemic prevention patent information, and **free to the public**. The platform published this dataset is here: https://ncp.patentstar.cn/
- It is an **open dataset** and I downloaded it by using the method of **web content scraping**.

Dataset I got:

- **7000** pieces of China and foreign patents related to COVID-19 epidemic prevention in total were selected from the database.
- The dataset was sorted according to technical relevance and importance, and subdivided into **9 first-level branches, 35 second-level branches and 78 third-level branches**.
- It covers therapeutic drug use, preventive drug use, detection and diagnostic reagents, medical devices, protective products, medical disinfection, medical waste treatment, wastewater treatment, artificial intelligence and big data application and other technical fields.

Dataset I got:

- **This patent dataset mainly contains these columns:**

- **Publication number**: the unique id of one patent authorized by relevant agency.
- **Abstract**: descriptions of function of patent
- **Disclosure Date**: the date patent published
- **IPC**: International Patent Classification, it shows the technologies used in one patent.This is defined by WIPO (Worldwide Intellectual Property Office). Each patent may contain one or more IPC number.
- **Countries/provinces/cities**: the location of company which applicants belong to.
- **Patent citation**: the patents' that were cited
- **Current statues**: valid or invalid information.This shows whether the technologies this patent claims are under protection of Intellectual Property law or not.
- **Type of patent**: invention, utility, appearance design.
- **Applicants**: most of them are organisation,company, university.
- **Inventors**: those who have made contributions to corresponding patents.
- **Main Classification**: the main IPC labelled to the patent which could reflect the main technology used in the patent, such as C07D309/10.
- the applicant(s) and address,

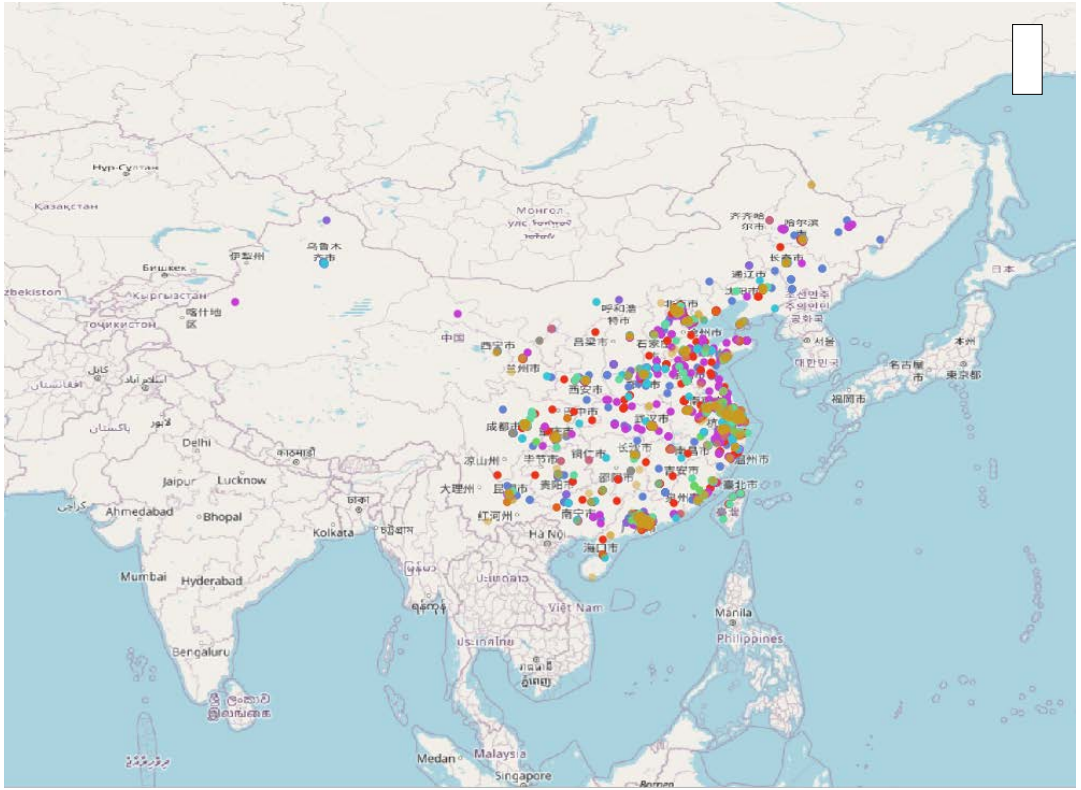
Dataset I got:

- The address of patent applicants could be down to street level.

- Geographical distribution of patent applicants

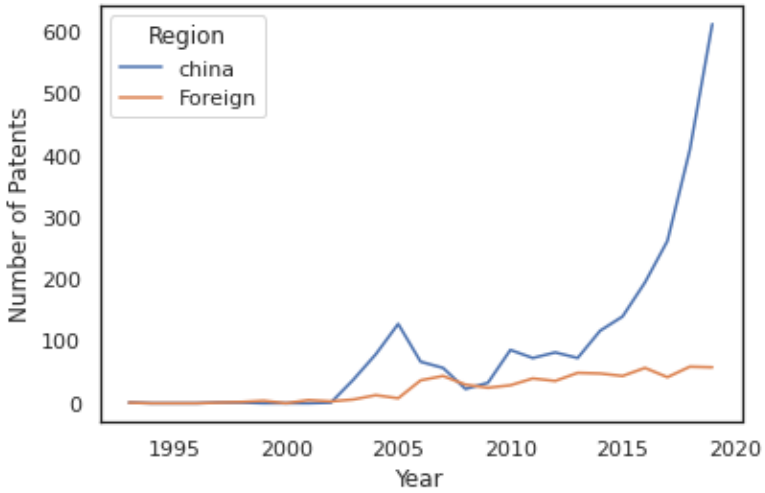
Dataset I got:

- Care Products
- Detection and diagnostic kits
- Drug use for treatment
- Medical apparatus and instruments
- Medical Sterilization
- Waste disposal methods
- Method for wastewater treatment
- Prophylactic
- Other

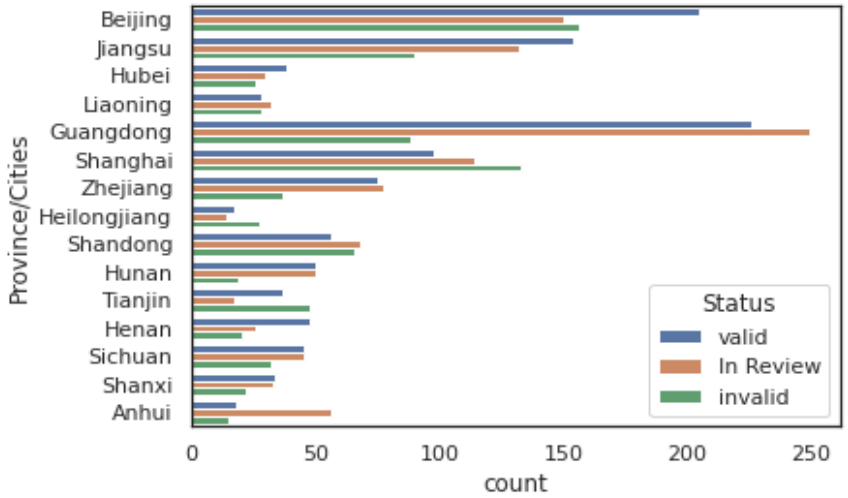


Geographical spatial distribution of patents of different types of epidemic prevention products

Dataset I got:



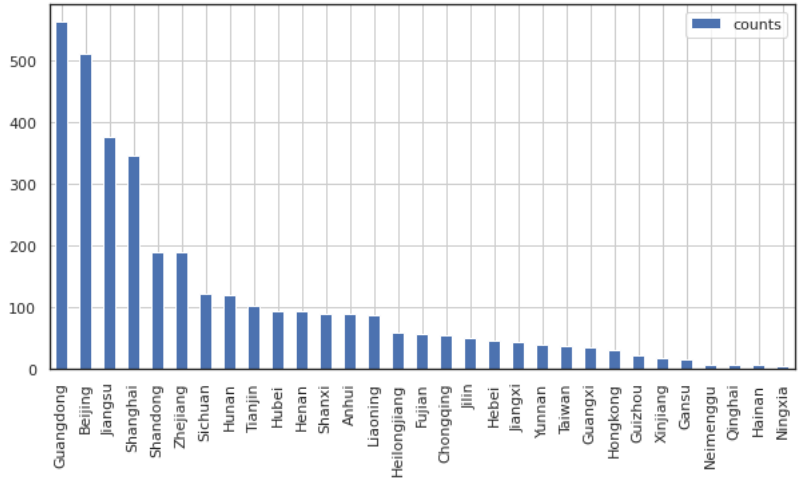
Trends of number of patents

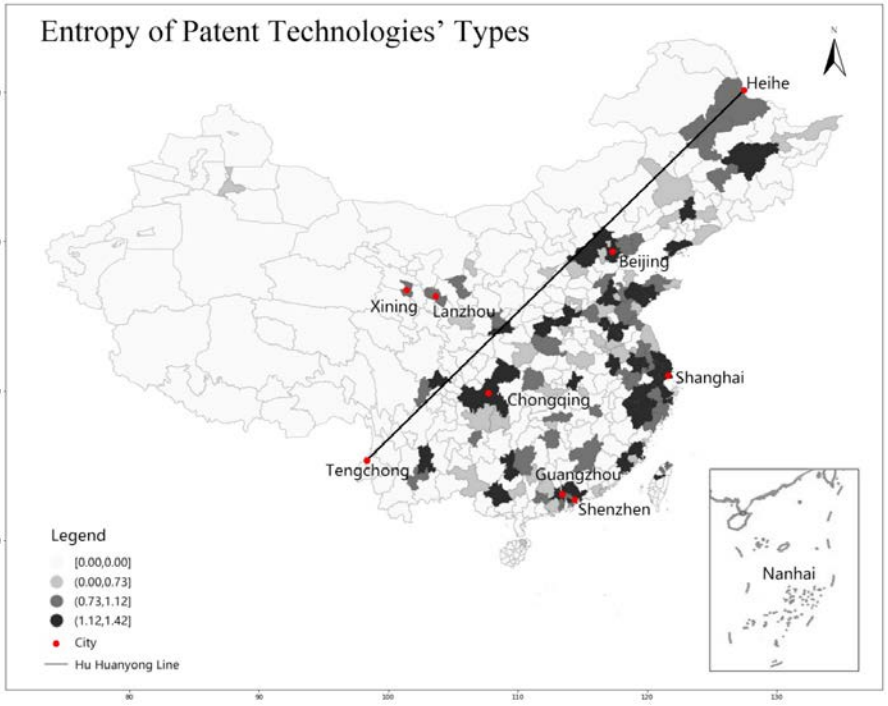


Validity of patents in Top 15 provinces

Current research results

- 1. Entropy of patent technologies' types (right) Patent number distribution in top 15 provinces (below)





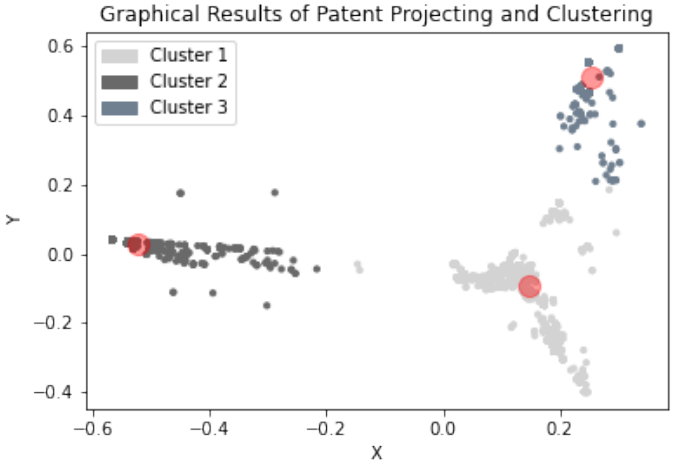
Current research results

- 2. Detect the main technologies used in this patent dataset:

  Method I used here: K-means, Multidimensional Scaling

Table 1. Top 10 IPCs for each cluster. (This result was formed by using a set of methods from one paper)

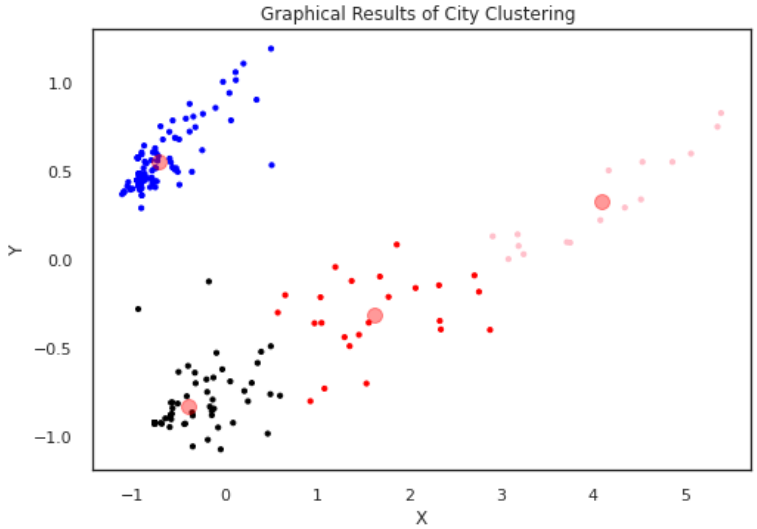
Rank	Patent Data Set		
	Cluster 1	Cluster 2	Cluster 3
1	A61H35/02(246) *	A61P31/16 (374)	A61B5/01(200)
2	G06F19/00(136)	A61P11/00(343)	A61B6/03(85)
3	C12Q1/68(126)	A61P31/14(290)	A61B5/00(39)
4	B25J11/00(118)	A61P31/12(184)	A61M16/00(33)
5	C12Q1/70(114)	A61K39/215(82)	A61B6/00(22)

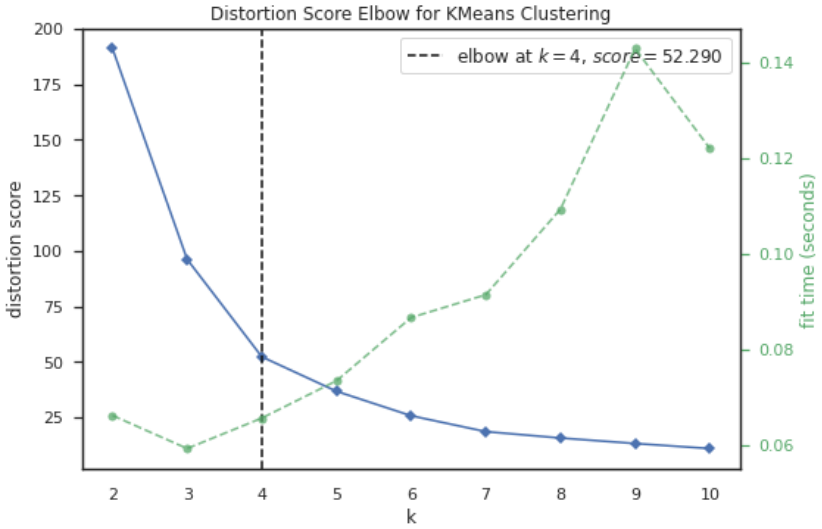


Current research results

- 3. Detect city clusters in the context of patent technologies

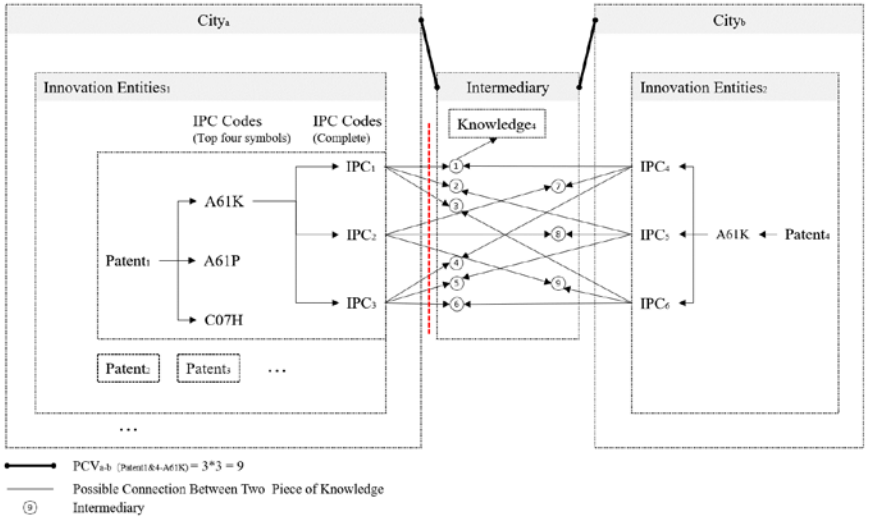
Method I used here: K-means, Multidimensional Scaling

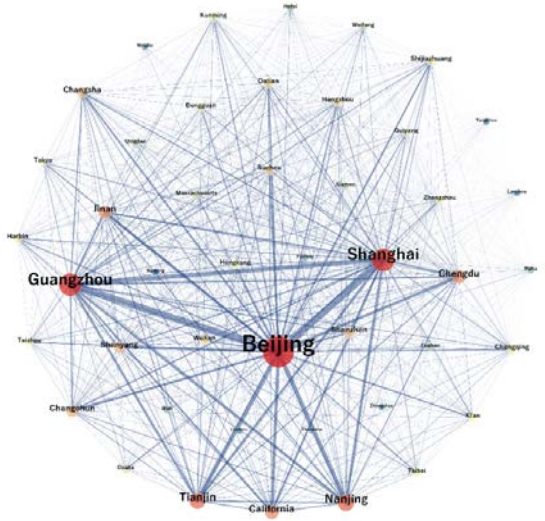




Current research results

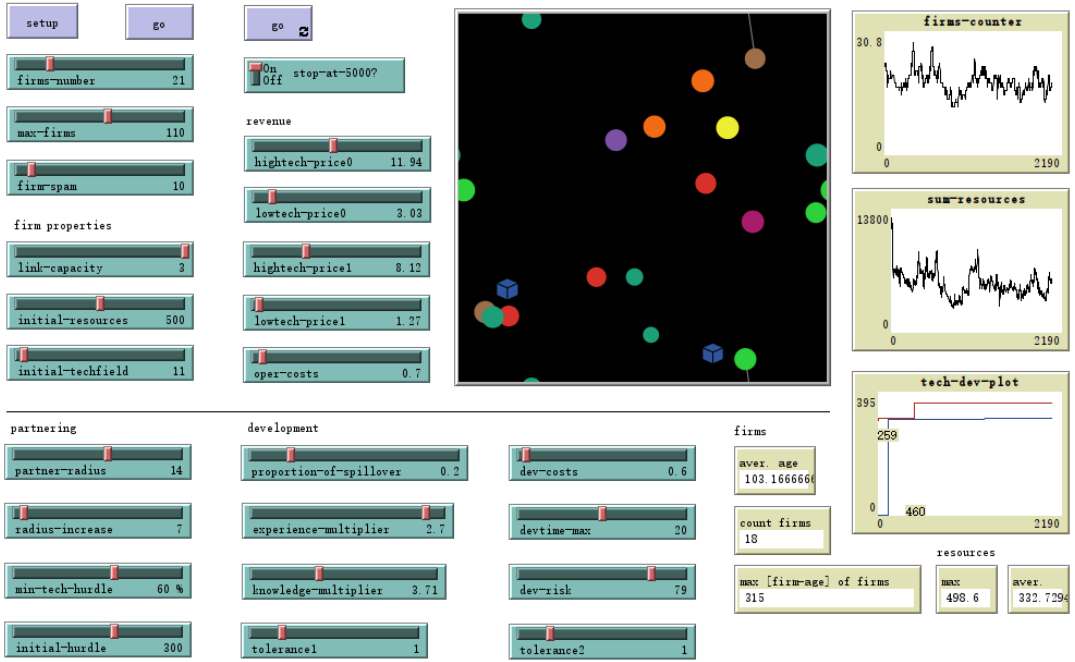
- 4. Defining cities' potential knowledge connection network by using the analysing framework: Global City Network (Taylor, 2001).





Current research results

- 5. Modelling to simulate the process of innovation.



Initial research questions for dissertation:

- If I define the technologies that used in patents as a kind of **capital**, how could these technologies **cluster, flow and combine** with each other to make a new product?
- √ Do these technologies used in EPP patents have **different influences**?
- √ What kind of **cities' network** emerges when these technologies connect with each other?
- What are the relations between this network and **city network based on patent citing**?
- How do these patents cluster on city level? How could I measure the characteristics of agglomeration and diffusion of spatial innovation?
- How could I **simulate the process of firms creating new technologies** based on patent dataset?

How to accomplish these tasks

- **1. Revising my current work results to make this network more concisely.**

- **Aims:** I try to make this part of work achieving the level of publication.

- Procedures:

1.Classified network into 3 parts based on the clusters I got from previous work

2.Adding the weights of IPCs into networks

3.Analysing the characteristics of these networks, such as degree distribution, centrality, community.

How to accomplish these tasks

- **2. Using methods to measure the characteristics of agglomeration and diffusion of spatial innovation**

- Methods:
- location quotient, Horizontal clustering location quotient,
- Locational Gini coefficient, The Herfindall-Hirschmann index
- Ellison-Glaeser geographical clustering index, Geographic co-occurrence index
- Moran’s I

How to accomplish these tasks

- **3. Modelling to simulate the process of innovation.**
- This will be initially finished by doing ABM coursework, which I try to simulate **how firms create new technologies based on patent dataset**.

  Firstly, I will try to build up one model which was shown in some academic papers.

  And then for dissertation, I will focus on how to modify this model to adapt my context.

Other Questions

- Do you have any recommendations for literature in this area? (patent analysis when considering the spatial factors and simulating the process of innovation)
- Do you have any suggestions how I could look deep into patent analysis when considering the spatial elements?

CASA MSc Spatial Data Science and Visualisation

Dissertation: Current Results and Ideas

Student: Xudong Zhang

Date: 17th, March 2021