Spatial pattern mining of tech clusters of dynamics and industry mix based on quantitative methods in England area, UK

Zeqiang Fang

CASA0012, MSc Spatial Data Science and Visualisation Dissertation

Supervisor: Dr Max Nathan

Repository: https://fang-zeqiang.github.io/CASA0012-Dissertation/

This dissertation is submitted in part requirement for the
MSc (Or MRes) in the Centre for Advanced Spatial Analysis,
Bartlett Faculty of the Built Environment, UCL

Word count: 8,000

2021-08-08

Abstract

Some abstract text

Declaration

I, Zeqiang Fang, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is xxx words in length

Acknowledgements

I would like to thank blah blah

Contents

List of Figures

List of Tables

Abbreviations

SIC07 type	...2	5-digit SIC07 code	SIC07 heading
Class	...	26110	Manufacture of electronic components
Class	...	58210	Publishing of computer games
Class	...	62090	Other information technology and computed service
Class	...	63110	Data processing, hosting and related activities
Class	...	27510	Manufacture of electric domestic appliances
Class	...	86230	Dental practice activities

Chapter 1

Introduction

1.1 Background

1. tech cluster development
2. dynamics cause better performance
3. industry clustering pattern and economics performances

1.2 Research Question and Objectives

How does tech clusters' dynamics pattern change in UK from 1998 to 2018? / What factors can affect tech clusters' dynamics pattern change in UK?

To what extent will dynamic change affect tech clusters'performance

1.3 Report Structure

1. data clean
2. tech cluster recognition
3. dynamics index generation
4. hypothesis (OLS estimation)

5. regression
6. residual analysis
7. result interpretation

Chapter 2

Literature Review

2.1 Industry Cluster & Tech Cluster

Tech clusters like Silicon Valley play a central role for modern innovation, business competitiveness, and economic performance. This paper reviews what constitutes a tech cluster, how they function internally, and the degree to which policy makers can purposefully foster them. We describe the growing influence of advanced technologies for businesses outside of traditional tech fields, the strains and backlash that tech clusters are experiencing, and emerging research questions for theory and empirical work.

2.2 Cluster Dynamics

Industrial dynamics and clusters: a survey, regional research. This article reviews clusters and their impact on the entry, exit, and growth of firms, as well as the literature supporting the evolutionary dynamics of cluster formation. This extensive review shows strong evidence that clusters promote the entry of manufacturers, but the evidence that clusters can promote the growth and survival of firms is rather weak. From a number of open-ended questions, this

research extracts various future research paths that emphasize the importance of manufacturer heterogeneity and the exact mechanism that supports the localized economy.

2.3 Location Quotient

On average, companies in large cities are more productive. There are two main explanations: corporate choice (big cities strengthen competition and only allow the most productive people to survive) and agglomeration economies (big cities promote interaction and increase productivity), which may be strengthened by the natural advantages of localization. In order to distinguish them, we nested a general version of the easy-to-handle company selection model and a standard agglomeration model. Stronger choices in large cities cut the distribution of productivity to the left, while stronger gatherings move to the right and expand the distribution. Using this forecast, French firm-level data, and new quantile methods, we show that firm choices cannot explain differences in spatial productivity. The results are applicable to various departments, city size thresholds, institutional samples and regional definitions.

2.4 How location affect entry pattern in UK/Global

2.5 How time affect entry pattern in UK/Global

2.6 Other factor can affect dynamics pattern in UK

Chapter 3

Methodology

3.1 Research Framework

1. Data Clean & Select
2. Identifying Tech Cluster
3. Measuring the Dynamics & Industry Mix
4. Quantitative Method Research
5. Temporal Qualitative Analysis

In this study, a data set containing all companies in UK will be cleaned and attribute selected, and technology companies will be identified and screened according to the classification and definition of technology companies on the official website of the British government. Before the quantitative study, this study is based on time and The spatial dimension counts the number of technology companies, and calculates the dynamic indicators of enterprise clusters and industrial combination indicators in a specific year and a specific region. Then this study conducts multiple regressions, univariate and bivariate variables Moran index testing to conduct spatial quantitative research, and finally combines Qualitative spatial pattern trend research on the spatial changes of indicators in three different time periods

3.2 Data Source and Processing

This raw dataset is collected from the core company data from Open Corporates master company database (Open Corporates, 2018). And the size of dataset accounts for 15 GB which is handled with `read_stata` and `get_chunk` function to read large data file in chunks, then increasing the reading speed. The "primary uk sic 2007" identification field is the basis of industry finding and the "birth year" is the key to measure dynamics variables . All rows whose these two values are empty are removed 17% incorperate date is missing and sic code is complete).

3.3 Identifying Tech Cluster

For the identification of science and technology companies, this study introduces the main 2007 sic code table to judge the science and technology industry, referring to the classification method of the Science and Technology Classification data set on the ons.gov.uk website; in order to better identify science and technology companies for the UK The economic contribution is officially based on the 2007 British Standard Economic Activity Classification, combined with different data sources, to classify and label science and technology companies (Office for National Statistics, 2015).

Table 1 2007 sic code for science and technology industry (part)

This research refers to the science and technology classification table provided by the government. The technology indicator is used to position the technology industry of all industries, and a total of 168 sic codes for the technology industry in 2007 were obtained, accounting for about 16% of all industry categories in the UK, including 5 industry categories such as Digital Technologies, Life Sciences & Healthcare, Publishing & Broadcasting , Other scientific/technological

SIC07 type	...2	5-digit SIC07 code	SIC07 heading
Class	...	26110	Manufacture of electronic components
Class	...	58210	Publishing of computer games
Class	...	62090	Other information technology and computed service
Class	...	63110	Data processing, hosting and related activities
Class	...	27510	Manufacture of electric domestic appliances
Class	...	86230	Dental practice activities

manufacture and Other scientific/technological services, details of the classification form will be attached in the attachment. There are almost 20% firms in the raw data belonging to tech firms according to the method of category as mentioned above.

3.4 Quatitative Analysis and Methods

3.4.1 Time Range Selection

3.4.2 Tech Cluster Identifying

Why I choose ttwa

Why ttwa can work as a firms' cluster

3.4.3 Dynamics Measuring Index

$$Entry\ Rate_{i,t} = \frac{Incorporating\ Firms_{i,t}}{Total\ Firms_i}$$

Where i means location(travel to work area), t means year

3.4.4 Dynamics Analysis

3.4.5 Herfindahl Index

$$Herfindahl\ Index_{i,t,k} = \frac{Tech\ Firms_{i,t}}{Total\ Firms_i}$$

Where `k` represents industry

— [1] (Herfindahl-Hirschman Index HHI

50 (50) ( ) HHI

3.5 Limitations

1. missing data `diss_year` have 99% missing value
2. ttwa

- ttwa might not be a data-driven method
- ttwa 2015
- ttwa (Ozkul,2014)

Ozkul, B., 2014. Changing home-to-work travel in England and Wales. Regional Studies, Regional Science, 1(1), pp.32-39.

3.6 1. Herfindahl-Hirschman Index

3.7 Ethical Statement

The data for this project comes from OpenCorporates, a firm which aggregates company-level data from around the world [link to their website]. In this case, OpenCorporates have taken data from the UK Companies House register [link

to Companies House]. As detailed by Nathan and Rosso (2015), all limited companies in the UK need to registers with Companies House when they are set up, and provide annual returns and financial statements. These include details of directors and company secretary, registered office address, shares and shareholders, as well as company type and principal business activity. Thus, all the data used here is already in the public domain.

The research objectives are tech firms in the UK for this project and the individual data will not be collected and measured in this project. For issues of deanonymisation or privacy, traceable information such as the real companies name and ID will not be utilised in the research. The raw data will be cleaned and filtered by several key variables include industries instead of the company's name or other sensitive information before doing the research. Through data cleaning, pre-processing, desensitisation or other processing methods, the risks of damage to company interests (such as social reputation, economic benefits and etc.) will be mitigated to an as low as possible level in the research process.

Besides, this project will not cause discrimination of industries or job categories. The final analysis results, such as the different industry concentration in each region, will not deepen some people's stereotypes and prejudices about the region (This content will be fully discussed in the project discussion section). It is necessary to point out and declare the objectivity of the analysis and the non-absoluteness of the results in the disclaimer. Consider the feelings of people and governments in different parts of the UK, this research will prevent the influence of personal preferences and subjective emotions.

The leakage of companies' name and information will be protected. For example, in the reflection of the results section of academic research, the name and related information of the companies that moved may be revealed. Although this information may be open to the public you need to know that this information may be used by people with other ulterior motives. This project will desensitise the company name and information at the stage of chart presentation, such as

using A, B, and C to replace them to achieve this purpose.

Chapter 4

Results

4.1 Visualisation and Analysis of Tech Cluster

4.1.1 Distribution

4.1.2 Descriptive Analysis

4.2 Visualisation and Analysis of Dynamics

4.2.1 Regression

Chapter 5

Discussion

Short introduction to the chapter, reviewing the previous chapter and detailing what this one aims to achieve and build upon.

To be done

5.1 Research significance

5.1.1 Global development goals

5.1.2 Local policy

5.1.3 Academic research

5.2 Limitations

To be done

5.3 Transferability

To be done

Chapter 6

Conclusion

Short introduction to the chapter, reviewing the previous chapter and detailing what this one aims to achieve and build upon.

To be done

References

a b c d e f g h i j k l m n o p q r s t u v w x y z

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (2012),The Productivity Advantages of Large Cities: Distinguishing AgglomerationFrom Firm Selection. Econometrica, 80: 2543-2594.https://doi.org/10.3982/ECTA84427

Frenken, K., Cefis, E., & Stam, E. (2015). Industrial Dynamicsand Clusters: A Survey. Regional Studies, 49(1), 10-27.doi:10.1080/00343404.2014.904505

Kerr, William R., and Frederic Robert-Nicoud. 2020. "Tech Clusters."Journal of Economic Perspectives, 34 (3): 50-76.https://www.aeaweb.org/articles?id=10.1257/jep.34.3.50

Open Corporates. (2018), The core company data from OpenCorporates master company database, electronic dataset, OpenCorporates Data Dictionary, viewed 12 June 2021, https://opencorporates.com/.

Office for National Statistics, 2015. Identifying Science and Technology Businesses in Official Statistics. [online] London, UK: Office for National Statistics, pp.10-14. Available at: https://webarchive.nationalarchives.gov.uk/20160105170025/http://www.ons.gov.uk/ons/site-information/using-the-website/rss-news-feeds/index.html [Accessed 28 July 2021].

Appendix A Classification Form

Science and Technology Classification

Appendix B Proposal