Estimating the effect of inventor mobility on invention productivity:
Term paper for *Methods for Causal Inference* course

Ashwin Iyenggar (1521001)
ashwin.iyenggar15@iimb.ernet.in

March 3, 2017

Abstract

I line up empirical evidence to demonstrate the heterogeneity in the geographical distribution of knowledge spillovers across various regions. I then explore three potential mechanisms that may help explain this heterogeneity: endogenous aspects of the regions themselves, complexity of work, and the intellectual property rights environment of the location. While the empirical results are yet inconclusive and incomplete, the current work extends prior work on geographic spillovers of knowledge by integrating three hitherto alternative explanations.

Keywords: Knowledge Spillovers, Endogeneity, Complexity, IPR

1 Introduction

There has been a long and illustrious scholarly tradition highlighting the agglomeration characteristics of economic regions, going back at least as far as

Marshall (2009), whose original work was published in 1890. More recently, scholars over the last three decades have demonstrated the paper trail of these knowledge spillovers through the study of patent citations (e.g., Almeida and Kogut (1999); Jaffe et al. (1993)). This tradition of scholarship has further shaped our theoretical understanding of knowledge spillovers through mechanisms such as the effects of inventor mobility (e.g., Almeida and Kogut (1999)), differential Intellectual Property Rights environments across locations (e.g., Zhao (2006)) and of the role of international geography (e.g., Singh (2007)). The nature and extent of the geographical distribution of knowledge spillovers observed in practice is so highly heterogenous across locations, firms and legal environments, that the understanding of the causal mechanisms leading to knowledge spillovers continues to intrigue the best of scholars. While this is, in no way dismissive of the enormous theoretical strides so far, the question is assumes greater significance in the environment surrounding the second machine age as some scholars have begun to highlight (McAfee and Brynjolfsson, 2014)

Motivated by empirical evidence surrounding the heterogeneity in the nature of knowledge flows across the various regions, I intend to explore the three mechanisms ostensibly influencing knowledge spillovers. Complexity of patents invented as a potential mechanism influencing the extent of local knowledge spillovers. This approach is not to be construed as yet another mechanical departure from the current theory on spillovers. I argue so with the following reasons. First, from a human capital perspective, it is valuable to understand the impact of MNCs that dominate much of the cutting- and bleeding-edge innovation in emerging markets on the development of the talent pool in the host country. Does a significant group of local inventors develop? Is this affected by the strength of the IPR regime in the host country? Second, a specific flavor of this question is the investigation of the spillover effects of the innovation process in emerging countries, or those known to have weaker IPR regimes. Specifically, do multinational firms that develop patentable technologies in emerging (or weaker IPR) countries create spillover effects in the host country talent pool? Or do the benefits remain localized to within multinational companies (MNCs) and their home country employees? Finally, the wide disparity in the extent of knowledge spillovers across locations, across firms and across IPR regimes is intriguing to the researcher and calls attention toward a creative response. a researcher to find the mechanisms that may lie behind such a phenomenon. Patents data allows us to ask these questions and to have them answered as has been in the tradition of Jaffe et al. (1993).

The choice of the three explanatory mechanisms is not arbitrary. Indeed, there has been a tradition of scholarly work in each of them[1]. First, several studies including Almeida and Kogut (1999) have conclusively demonstrated that the kind and extent of knowledge flows between firms in Silicon Valley is unparalleled in the rest of the world. Indeed, our analysis on a select chosen locations demonstrates this adequately in Figure **??** where the lines in purple and red stand out (Silicon Valley, or the broader San Francisco Bay Area is classified into two Metropolitan Statistical Areas (MSAs) by the United States census). Second, as evidenced by the respective scholarly traditions of Baldwin and Henkel (2015), Ethiraj and Levinthal (2004) and Yayavaram and Ahuja (2008), the complexity, intellectual property and organizational implications have been addressed by scholars in the context of patenting. In the spirit of Ethiraj and Levinthal (2004), I propose a definition of complexity that is rooted in the question of knowledge spillovers. Specifically, I suggest that complexity may be seen as either an attribute of usage, or as an attribute of invention. A patent that is used (cited) by several patents belonging to distinct and different patent technology classes maybe seen as modular by virtue of it being able to be plugged into multiple, diverse applications. Alternatively, a patent that is constructed with few dependencies may also be seen as being modular by virtue of its capacity to be developed standalone, or with minimal intervention from other modules. For the purposes of this study, I use a definition of Complexity that captures both the effects above. Finally, the scholarly tradition in the

[1] I am however, unable to go in much of those details in the current article

international business area has extensively analyzed the relationship between economic geography Singh (2007), intellectual property environments (Zhao, 2006) and political geography (Singh and Marx, 2013).

The current work is placed at the confluence of these three traditions, with the focus on implications for beneficial knowledge spillovers. A second objective of the current work is to understand the local impact of inventing activity by multinationals in emerging nations. I attempt to answer the following questions. First, how does the nature of the geographic distribution of citations made by inventions from a region affect the quantum of citations received. Second, how do complexity of inventions and cross border differences in intellectual property environments affect the previous relationship.

The benefits of understanding geographic and multinational collaboration in invention is that we may seek to inform both managers and firms about the potential opportunities of tapping into or creating spillover effects in the host country talent pool. Does a significant group of local inventors who develop due to spillovers? Do they then move around to cross-pollinate to other firms? How do domestic firms integrate and appropriate rents in this context. These are some of the many interesting and valuable directions spawned by the current approach.

Table 1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**N**
Moved Region (MR)	0.06	0.238	4368062
Moved Country (MC)	0.021	0.143	4368062
Productivity	1.79	2.647	4368062
Prior Patents of Inventor (PPI)	4.716	16.718	4368062
Prior Patents of Team (PPT)	24.557	112.7	3721064

The rest of this article is organized as follows. In the following section, I present the motivation for empirically designing this study from a selected sampling study. The section following that proposes the hypotheses I intend to test empirically. The approach I take is to assume prior scholarly results about spillovers as a given, and nuance new insights building on top of these giants. The following section on research design presents the constructs created for computing complexity, as well as a discussion on the methodology of the research. My initial, incomplete results are then presented, and a conclusion is drawn of this article as very much a work in progress.

2 Motivation

I motivate this study with a These disparate spillover behaviors across locations have been attempted to be explained by scholars (e.g., Singh (2007); Zhao (2006)) by slicing the problem in a specific context (e.g., of an MNC Parent - MNC Subsidiary). This rather wide disparity between a selection of inventing locations, provides us with the context to dive into the understanding the mechanisms that underlie this divergence in knowledge spillover patters across regions in the aggregate without making simplifying assumptions.

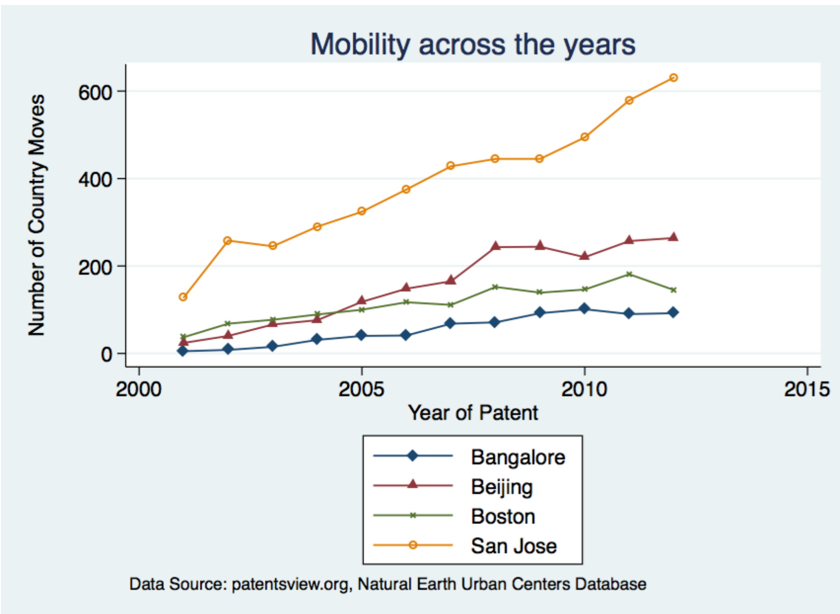

Figure 1: Country moves by year

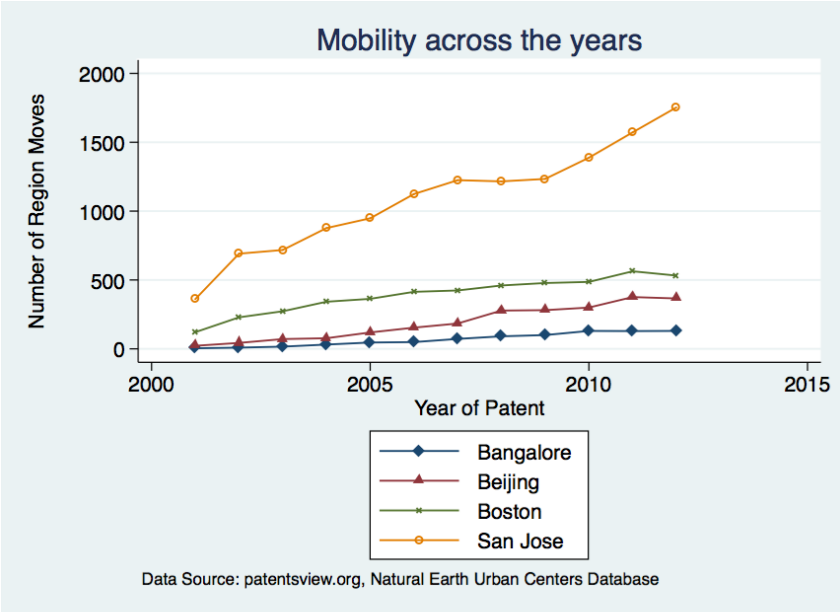

Figure 2: Region moves by year

3 Theory

Our approach toward theory building is assume the premise presented by prior scholars as valid to start, and then proceed step by step to nuance those arguments based on our arguments about the interplay of path dependence of the priors of the region, the level of complexity of the inventions produced and the relative strength of the intellectual property rights environment. Building off on Jaffe et al. (1993), I propose hypothesis 1 consistent with the priors supported in northern California.

Hypothesis 1: An increase in the average mobility of inventors in a region increases the average productivity of the inventor

I then build on top of Zhao (2006) and Singh (2007) to propose hypothesis 2.
Hypothesis 2: The effect in Hypothesis 1 is moderated positively by the strength of the prior pool of inventions by the inventor

Scholars (Baldwin and Henkel, 2015; Yayavaram and Ahuja, 2008), have argued that increased interaction with a larger number of components creates organizational impediments to an increase in reusability of prior work. In the presence of a stronger differential in the IPR environments between inventing locations, Zhao (2006) suggests that organizational mechanisms may stand to counter the treat posed by weaker property rights. In a similar vein, I argue that a differential in the IPR rights environment creates the organizational response to increase complexity of the inventions shared across country and IPR boundaries.

Hypothesis 3: The effect in Hypothesis 1 is moderated negatively by the strength of the prior pool of inventions by the inventing team

4 Research Design

4.1 Complexity

I construct my measure of complexity based interactions between the different patent sub-classes. Since each of the interactions between patent sub-classes may introduce a new interaction, I model interactions on a binomial function. Specifically, when `subclass` represents the number of distinct patent sub-classes, I define `interaction(subclass)` as follows:

$$interaction(subclass) = \begin{cases} 1 & : subclass \leq 2 \\ \binom{subclass}{2} & : subclass > 2 \end{cases}$$

I would expect, from a user perspective that the more number of contexts in which the patent is valuable, the lower should be the complexity. If `complexity` represents my measure of the complexity of the patent, and `usage contexts` represents the number of distinct contexts where the patent is found valuable, I should expect the following relationship to hold:

$$\texttt{Complexity} \propto \frac{1}{usage\ contexts}$$

Similarly, from an inventor perspective, the more the number of contexts that the patent is built on, the higher should be the complexity. A patent that is developed without citing any other patents is an extreme case of lowest complexity, while one that requires to be built upon several `source contexts` is properly understood as being more complex.

The relationship between `source contexts` and `complexity` is therefore a normal one as depicted below.

$$\texttt{complexity} \propto source\ contexts$$

Using the principles above, I therefore develop the following definition of complexity.

$$\texttt{complexity} = \frac{interaction(subclass_{\text{cited}})}{interaction(subclass_{\text{patent}})}$$

By the definition above, a patent that cites no patents (and hence has $subclass_{\text{cited}} = 0$) but is itself assigned to 4 sub-classes (and hence has $subclass_{\text{patent}} = 4$) will have a raw Complexity score of $\frac{1}{\binom{4}{2}} = 0.16$. If the patent itself had been assigned onto to 2 sub-classes, the raw complexity score would have been just 1. Therefore, the more the number of patent sub-classes a patent is assigned to, the lower its complexity score (by a square term). A similar but inverse relationship would hold for sub-classes arising out of cited patents. Here, I take a set union of patent sub-classes assigned to each cited patent, and use that count to determine the value of the `interaction` function.

4.2 IPR Classification

A review of the academic literature surrounding the construction of IPR indexes indicated that there were several, as was also evident in Zhao (2006) constructing a composite measure for the purposes of her article. Lesser (2010) provides an alternative, composite scoring system that includes the following components: protectable subject matter, membership in convention, enforcement, administration and duration of protection. I have therefore used the scores generated by Lesser (2010) for the purposes of this study. The extensive table of IPR scores has not been presented here to adhere to the page restriction, but can be made available on request. The listing has several countries for which scores have not been provided. However none of the top patenting nations were among them, and I therefore chose to go along with this scale.

4.3 Data Source

I derive all patents data for this study from patentsview.org. The dataset considered is for all USPTO patents filed in the period 1976 to 2015. For the IPR Scores, I rely on the scores generated by Lesser (2010). For country definitions, I use the resources provided by Thematic Mapping. To determine if spillovers are local, I use a composite data source as described in the following. For locations in the United States, it has been standard to use Metropolitan Statistical Areas (MSA) for analyses related to economic geography. Such standardized data is unavailable for non-US locations. Urban areas are a close substitute for economic centers, and I therefore determine to use one such definition for non-US locations. My data source for MSA of US locations is the US census and that for urban areas for world wide locations is Natural Earth Data.

This automatically raises conflicting definitions for locations in the United States. So that the MSA definitions take precedence, I eliminated all data pertaining to US locations from the Natural Earth urban centers data and integrated this with the MSA information. With this I generated a single database of location information for economic centers around the world. A sample region definition is depicted in Figure 3. Here all points in the highlighted region are considered to fall within the San Jose, CA Urban Center. In Figure 4, I present a non-MSA example of a geographic definition based on the urban centers data from Natural Earth Data. As will be noticed in Figure 4, the Bangalore urban center is seen to include parts of Hosur as well.

4.4 Unit of Analysis

The unit of analysis for this study is the `inventor-year`

4.5 Dependent Variable

My primary dependent variable is the productivity of an inventor in a year.

4.6 Explanatory Variables

Primary Explanatory Variable: Mobility of innovators (Between-Region Mobility, Between-Country Mobility) Moderating Variables: Prior patents of inventor (PPI), Prior patents of team (PPT) Control Variables: Technology subcategories, Year effects

4.6.1 Citations Made

My primary explanatory variables are the four counts of Citations Made along the two dimensions of same/different region and same/different assignee While most patents have multiple inventors, and some patents also have multiple assignees, my question requires us to associate a single location to the inventor of a patent, and a single location for the assignee of the patent. For the inventor location, I tabulate the count of each of the regions that each inventor is a resident of at the time of the filing of the patent application. In doing so, I treat all inventors equally and allocate the most frequently occurring location as the location of the inventor for that patent. In case of a tie, I assign the location of the first inventor (given by the sequence number of the inventor on the patent) as the location of the inventor of the patent.

For the assignee location, I treat multiple assignees as having been granted separate patents. I do this since the number of patents with multiple assignees is small, and so as to not lose potentially valuable information.

4.7 Control Variables

4.7.1 Size of the patent pool, and Number of patents generated in a region - year

Since the priors of the region may themselves explain the extent of citations received, I control for both the number of patents generated that year, as well as the aggregate pool of patents invented within that region. The reason both variables are logged is because of the exponential nature of the estimation method used.

5 Regression Results

6 Results

6.1 Spillover Effects

Table 6 presents the preliminary results from my regression. The first model uses year dummies but does not include region fixed effects, while the second includes both. As evident, the effects do not seem stable and much empirical work will be needed to determine the right identification method. A couple of approaches that are planned to be tried are to look at controlling for technology class, as regions may vary widely on this aspect. As indicated at the outset, the empirical aspect of this article remains work in progress.

Table 2: Regression Results

Country Mobility of Inventors and Productivity of Inventors

VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Country (MC)	1.7074***	0.9275***	0.5005***
	(0.4054)	(0.1655)	(0.1186)
Prior Patents of Inventor (PPI)		0.0875***	0.0775***
		(0.0066)	(0.0027)
Prior Patents of Team (PPT)		0.0004**	0.0005***
		(0.0002)	(0.0002)
MC x PPI			0.0381***
			(0.0062)
MC x PPT			-0.0009*
			(0.0005)
Constant	1.7096***	1.7260***	1.7337***
	(0.1609)	(0.0727)	(0.0810)
Observations	3,981,822	3,440,812	3,440,812
R^2	0.0176	0.2962	0.3063
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Country	Country	Country

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: Regression Results

Country Mobility of Inventors and Productivity of Inventors

VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Country (MC)	1.8086***	0.9959***	0.6032***
	(0.1945)	(0.0892)	(0.1072)
Prior Patents of Inventor (PPI)		0.0860***	0.0774***
		(0.0041)	(0.0036)
Prior Patents of Team (PPT)		0.0004***	0.0005***
		(0.0001)	(0.0001)
MC x PPI			0.0332***
			(0.0083)
MC x PPT			-0.0006
			(0.0006)
Constant	1.7019***	1.7212***	1.7294***
	(0.0871)	(0.0453)	(0.0483)
Observations	2,886,438	2,507,963	2,507,963
R^2	0.0173	0.2903	0.2980
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Region	Region	Region

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Regression Results

Country Mobility of Inventors and Productivity of Inventors

VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Country (MC)	1.7541***	0.9859***	0.5779***
	(0.1470)	(0.0872)	(0.1018)
Prior Patents of Inventor (PPI)		0.0812***	0.0719***
		(0.0139)	(0.0125)
Prior Patents of Team (PPT)		0.0004***	0.0005***
		(0.0000)	(0.0000)
MC x PPI			0.0351***
			(0.0075)
MC x PPT			-0.0009*
			(0.0005)
Constant	1.7081***	1.7267***	1.7340***
	(0.0213)	(0.0217)	(0.0213)
Observations	4,120,577	3,563,673	3,563,673
R^2	0.0179	0.2845	0.2940
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Technology	Technology	Technology

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Regression Results

Regional Mobility of Inventors and Productivity of Inventors			
	(1)	(2)	(3)
VARIABLES	Productivity	Productivity	Productivity
Moved Region (MR)	1.5328***	0.8679***	0.5668***
	(0.2967)	(0.1278)	(0.0851)
Prior Patents of Inventor (PPI)		0.0866***	0.0740***
		(0.0065)	(0.0019)
Prior Patents of Team (PPT)		0.0004*	0.0005***
		(0.0002)	(0.0002)
MR x PPI			0.0335***
			(0.0078)
MR x PPT			-0.0006***
			(0.0002)
Constant	1.7242***	1.7335***	1.7487***
	(0.1644)	(0.0746)	(0.0848)
Observations	3,981,822	3,440,812	3,440,812
R^2	0.0292	0.3001	0.3096
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Country	Country	Country

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Regression Results

Regional Mobility of Inventors and Productivity of Inventors

VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Region (MR)	1.3974***	0.7966***	0.5571***
	(0.0997)	(0.0531)	(0.0556)
Prior Patents of Inventor (PPI)		0.0852***	0.0741***
		(0.0041)	(0.0027)
Prior Patents of Team (PPT)		0.0004***	0.0005***
		(0.0001)	(0.0001)
MR x PPI			0.0282***
			(0.0054)
MR x PPT			-0.0005***
			(0.0002)
Constant	1.7123***	1.7258***	1.7404***
	(0.0894)	(0.0468)	(0.0504)
Observations	2,886,438	2,507,963	2,507,963
R^2	0.0282	0.2940	0.3010
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Region	Region	Region

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Regression Results

Regional Mobility of Inventors and Productivity of Inventors

VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Region (MR)	1.5681***	0.9204***	0.6239***
	(0.0926)	(0.0606)	(0.0654)
Prior Patents of Inventor (PPI)		0.0803***	0.0683***
		(0.0138)	(0.0124)
Prior Patents of Team (PPT)		0.0003***	0.0005***
		(0.0000)	(0.0000)
MR x PPI			0.0319***
			(0.0051)
MR x PPT			-0.0005***
			(0.0002)
Constant	1.7227***	1.7344***	1.7489***
	(0.0197)	(0.0218)	(0.0207)
Observations	4,120,577	3,563,673	3,563,673
R^2	0.0298	0.2888	0.2984
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Technology	Technology	Technology

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

7 Limitations and Looking Ahead

Direction of Causality Underestimation bias of mobility - effects Mechanism by which mobility affects productivity - other explanations Alternative measures of productivity I started this study attempting to understand if I could bring three mechanisms to bear together in explaining the heterogeneity in knowledge flows across regions. While there seems to be theoretical promise to exploring this question, this was a study too big to have been completed within the constraints of a term. Specifically, the endeavor has exposed me to the challenges to demonstrating empirically driven work with the objective of building theory. I intend to continue to pursue this further and integrate the IPR level data and complexity data to the flows. In addition, I plan to explore the prospect of controlling for technology classes, and see if that may lead to a strong result.

8 Acknowledgements

I am greatly indebted to Vidhya Soundarajan for having helped me with thinking deeply about my research design, level or analysis and causality. While I might have not done as much justice to the causality instruction, I cannot imagine having make as much empirical progress on this project if not for that guidance. I am also indebted to her for having encouraged me to see the theoretical relevance and contribution of empirical work.

I am also grateful to Sai Yayavaram for having introduced me to the literature on innovation, and for having hand held me with working on the patents data. Indeed many of the skills in understanding the data underlying this article owe their origin to him. All mistakes though, remain entirely mine.

References

Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45(7):905–917.

Baldwin, C. Y. and Henkel, J. (2015). Modularity and intellectual property protection. *Strategic Management Journal*, 36(11):1637–1655.

Ethiraj, S. K. and Levinthal, D. (2004). Modularity and innovation in complex systems. *Management Science*, 50(2):159–173.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598.

Lesser, W. (2010). Measuring intellectual property strength and effects: An assessment of patent scoring systems and casuality. *J. Bus. Entrepreneurship & L.*, 4:345.

Marshall, A. (2009). *Principles of Economics: Unabridged Eighth Edition*. Cosimo, Inc.

McAfee, A. and Brynjolfsson, E. (2014). The second machine age. *NY: WW Norton & Company*.

Singh, J. (2007). Asymmetry of knowledge spillovers between mncs and host country firms. *Journal of International Business Studies*, 38(5):764–786.

Singh, J. and Marx, M. (2013). Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity. *Management Science*, 59(9):2056–2078.

Yayavaram, S. and Ahuja, G. (2008). Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. *Administrative Science Quarterly*, 53(2):333–362.

Zhao, M. (2006). Conducting r&d in countries with weak intellectual property rights protection. *Management Science*, 52(8):1185–1199.

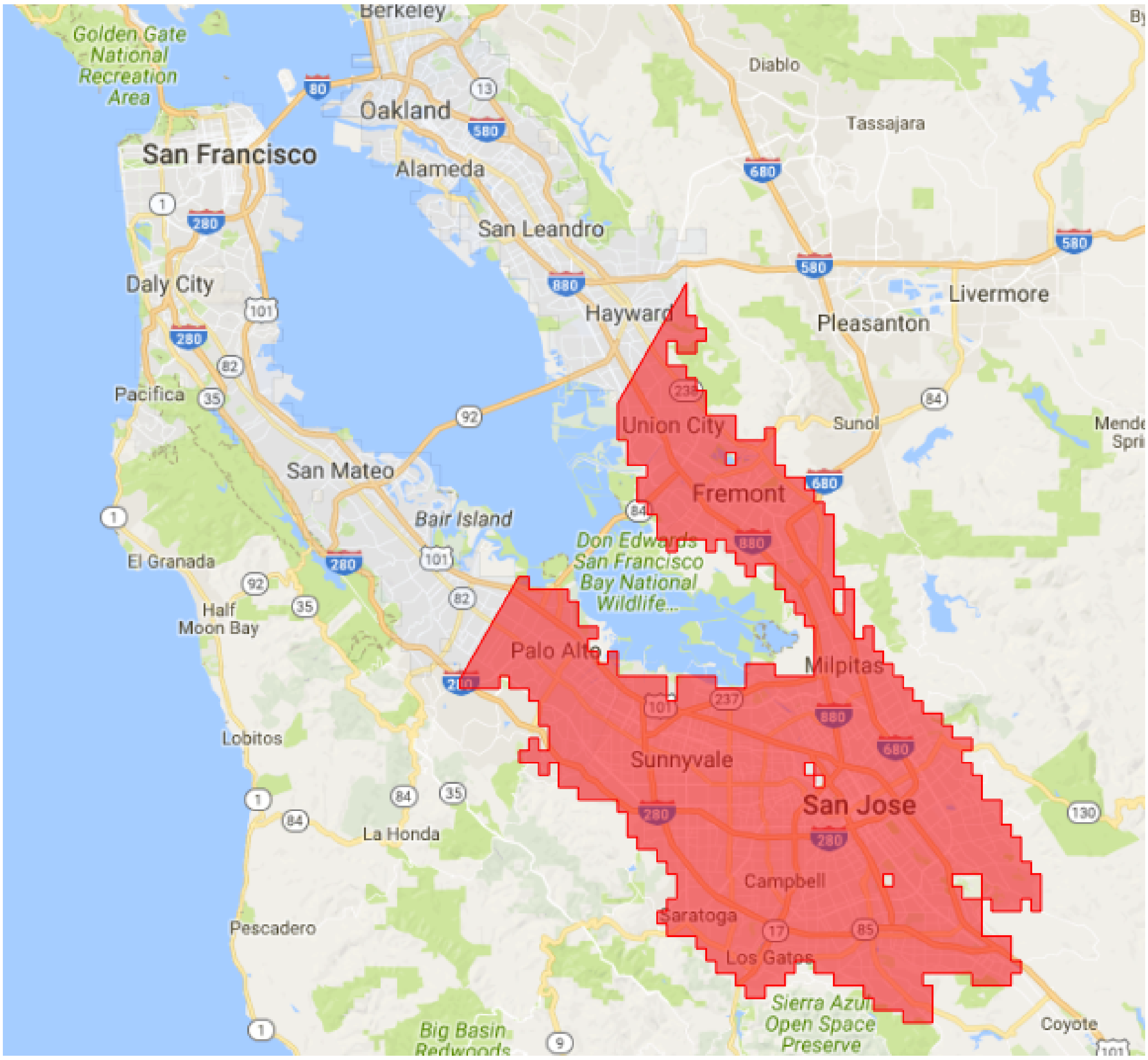

Figure 3: Geographic Definition of San Jose, CA

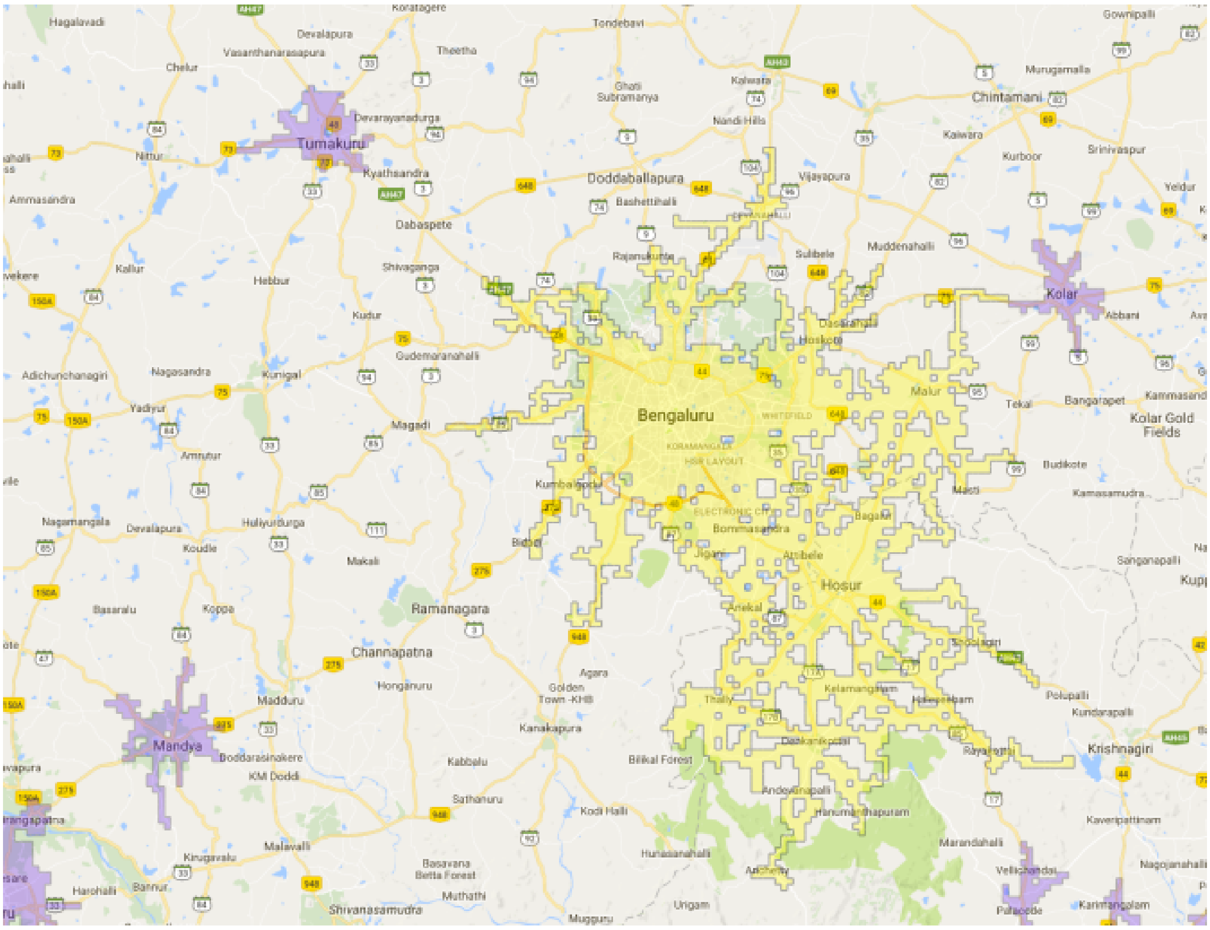

Figure 4: Geographic Definition of Bangalore