Heterogeneity in knowledge flows across regions: Investigating patterns and mechanisms

Abstract

Using patent citations as measures of knowledge flows, we explore if the different types of knowledge flows in a region affects the quality of the inventions originating in that region. We leverage a database of worldwide urban centers obtained from remote sensing data and find that local knowledge flows do not seem to impact inventive quality of regions.

Introduction

Scholars of economics and strategy have for long recognized that clusters and agglomeration economies play an important role in fostering innovation (Marshall, 2009; Porter, 1990). Agglomeration economies arise due to labor pooling advantages, economies of specialization of local suppliers, or knowledge spillovers (Krugman, 1991; Porter, 1990). In this study, we examine how the nature of knowledge flows in a region may affect the quality of the inventions generated in the region.

Patent citations have traditionally been used as a measure of knowledge flows (Jaffe et al., 1993) although recent work raises questions about the efficacy of using patent citations as a measure of knowledge flows (Arora et al., 2017). Several studies have, however used patent citations to demonstrate that knowledge flows are localized (Alcácer and Gittelman, 2006; Almeida and Kogut, 1999; Jaffe et al., 1993). Regions, however vary in their inventive output (Agrawal et al., 2014) and the nature of knowledge flows in a region may be one source of this variation. In this study we use patent citation data for patents applied for between 2001 and 2012 to empirically estimate the relationship between knowledge flows within a region and quality of inventions from that region.

The rest of the paper is organized as follows. The next section defines a framework to classify knowledge flows in a region and motivates our work by demonstrating how different regions fare differently. We then describe our data and methods in the following section. Our preliminary results are then presented, followed by a discussion of the results. We conclude with next steps and open questions for further research.

Effects of knowledge flows on quality of inventions

We categorize all knowledge flows incorporated into an invention along two dimensions: first, whether the knowledge flows among inventors are local to a geographical region or not, and second, whether knowledge flows are within the boundary of the firm or not. This classification allows us to analyze knowledge flows in four mutually exclusive, but collectively exhaustive categories as illustrated in Figure 1. We next describe each quadrant and discuss how the category of knowledge flow in that quadrant can affect invention quality. Within the context of this framework, we ask what is the net effect of each category of knowledge flow on invention quality, and which category of flow will have the largest effect.

The top left quadrant, labelled an "Independent Research Center" captures those knowledge flows that reflect competence building. Since these knowledge flows are both local and within the firm, this may be seen to represent local search on two dimensions (within firm and within region). Thus, while the competence that is being built up the Independent Research Center can be expected to have a positive effect on invention quality, the localness of the search on both dimensions may have a negative effect on the invention quality. Figure 2 depicts the normalized knowledge flows

	Same Region	Different Region
Same Assignee	Independent Research Center	Geographic Diversification
Different Assignee	Cluster	Diffusion

Figure 1: Categorizing inventions by matching inventor regions and assignees

for this category across time for five regionsBangalore, Beijing, Tel-Aviv-Yafo, Boston and San Jose (core of "Silicon Valley"). While our empirical analysis covers all the major regions, we chose these five as illustrative examples. We note that San Jose and Boston both report a higher proportion of knowledge flows within the same firm in the same region, while Bangalore and Tel Aviv-Yafo have the lowest proportion (fewer than 1%) of their citations to the same firm within the same region.

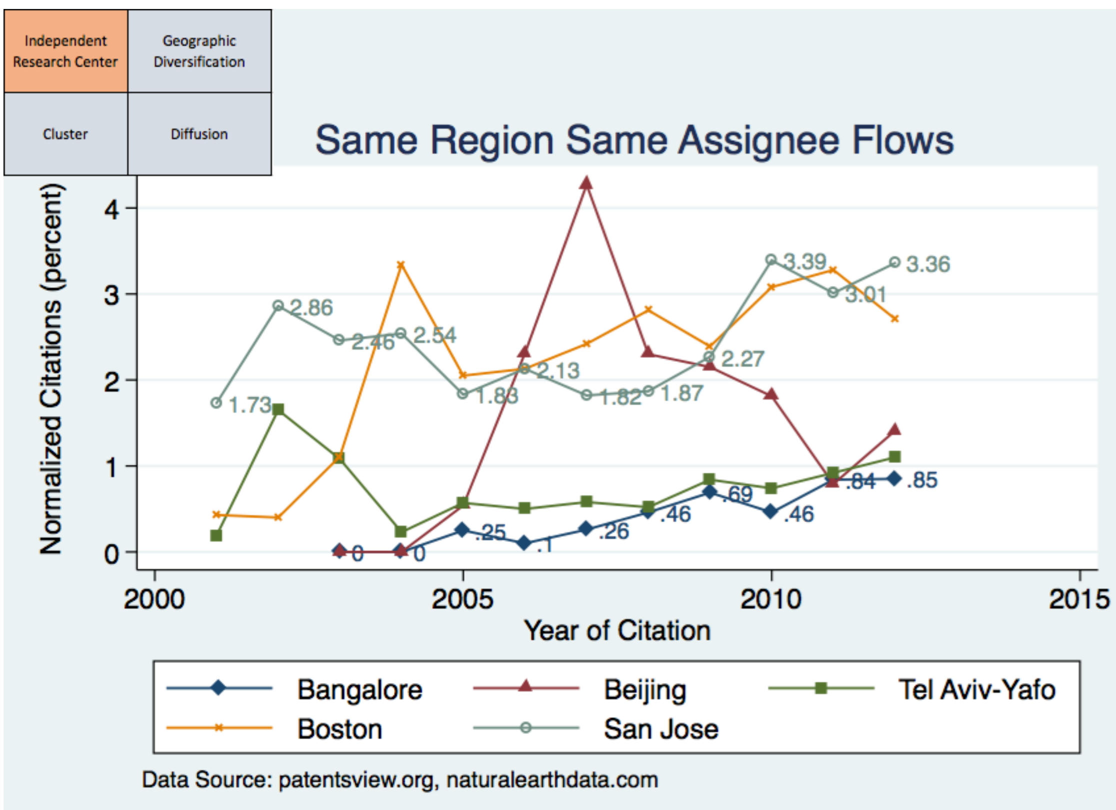

Figure 2: Flows within regions and within assignees

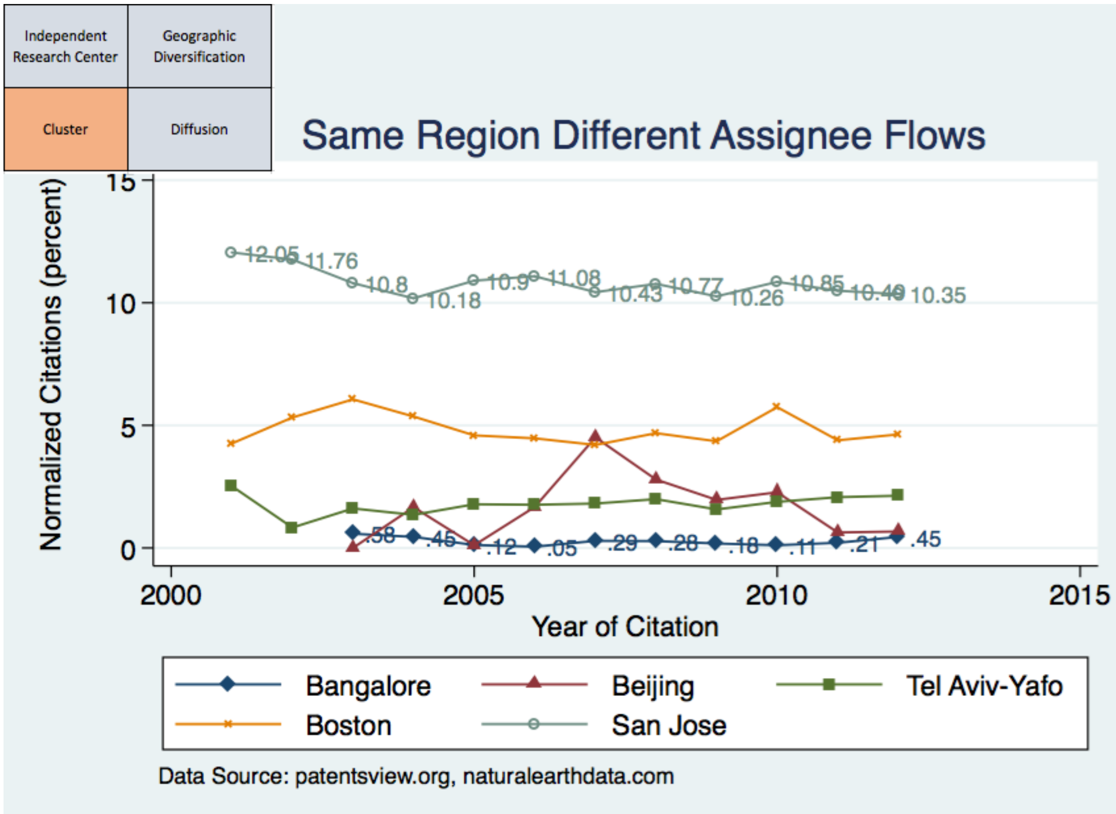

Figure 3: Flows within regions and across assignees

The quadrant on the bottom left, labelled "Cluster" captures knowledge spillovers within a region. Here firms may be seen as performing local search on one dimension (within regions) but not the other (within firms). Bangalore, Beijing, Tel-Aviv-Yafo, Boston and San Jose (core of ?Silicon Valley?). While our empirical analysis covers all the major regions, we chose these five as illustrative examples. Figure 3 depicts the normalized knowledge flows for this category across time for the same five regions. San Jose clearly stands out from the rest, suggesting a higher amount of across firm flows of knowledge in Silicon Valley, a result consistent with several prior studies.

The quadrant on the top right, labelled as "Geographic Diversification" captures local search on the dimension of the firm (across geographies) but not across regions. Innovations that are built on knowledge from several regions can be expected to benefit from the diversity of knowledge across regions. Yet, as in the previous quadrant, there is localness along the dimension of firm and such localness can have a negative effect on invention quality. Figure 4 depicts the normalized knowledge flows for this category across time for the five regions. We note that Bangalore and Beijing have a relatively higher proportion of knowledge flows to same assignees in different locations, thus confirming the role of these regions as R&D outposts of multinational firms.

Finally, the bottom right quadrant labelled "Diffusion" captures high exploration along both dimensions, indicating the development of a global pipeline. Figure 5 depicts the normalized knowledge flows for this category across time for five regions. We note that Bangalore, Beijing and Tel Avis-Yafo have a higher level of knowledge flows from other firms in other regions compared to Boston and San Jose, which is to be expected given that the absolute level of innovative activity in these emerging hotspots is still lower compared to that in Boston and San Jose.

As can be seen from the preceding discussion, prior theory suggests both positive and negative effects for each of these four categories of knowledge and it is not clear what the net effect will be. It also means that prior theory does not provide guidance on which category of knowledge flows will have the highest effect on invention quality. Since

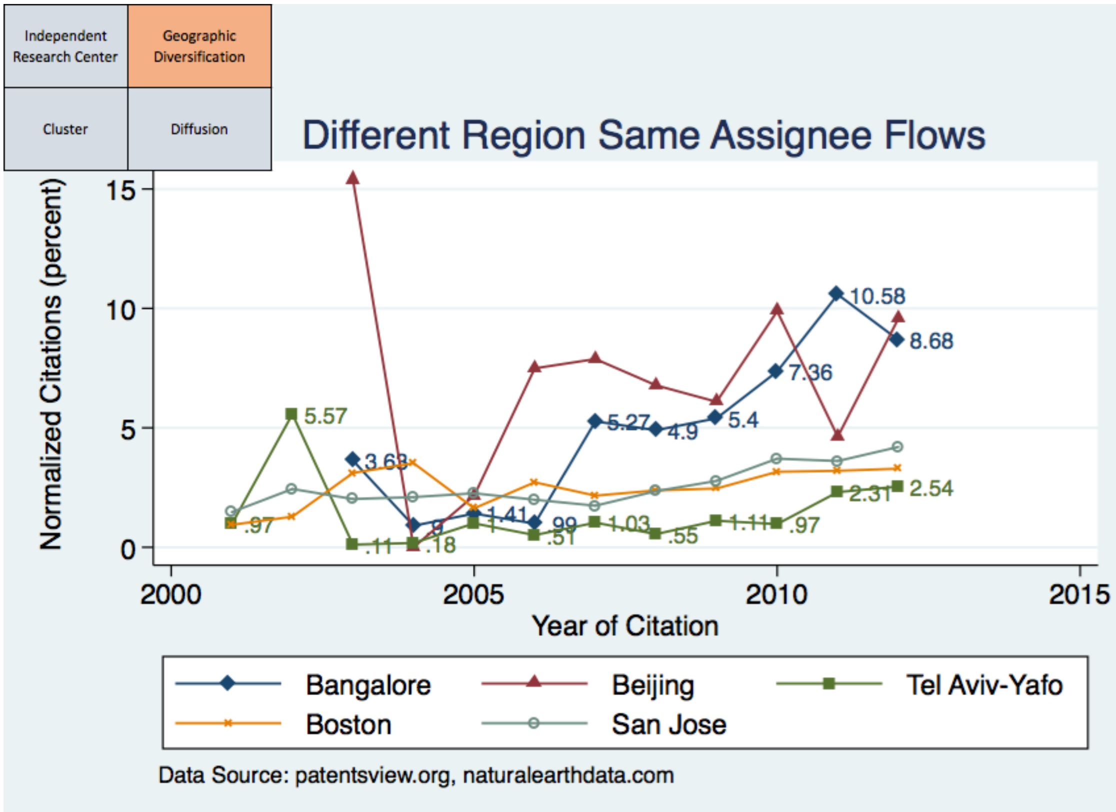

Figure 4: Flows across regions and within assignees

theory does not provide us with an answer, we rely on empirical analysis to inform us on the net effect of each category of knowledge flow and which category has the highest effect.

Data and Measures

We use patent citations data from the U.S. Patent Office (USPTO) as provided by patentsview.org. Additionally to map inventors to regions, we use urban centers data for world wide locations from Natural Earth Data that uses remote sensing data to determine urban agglomerations developed by Schneider et al. (2003). While it has been common to use Metropolitan Statistical Areas (MSA) for analyses related to economic geography in the U.S., an equivalent measure is unavailable for the rest of the world. For comparability and consistency, we choose to use the urban centers definitions from Natural Earth Data for all regions both within U.S. and outside U.S.

Our unit of analysis is the region-year. To be consistent with our objective of measuring knowledge flows, we restrict ourselves to those citations categorized as 'cited by applicant'and leave out those categorized as 'cited by examiner 'because the latter may not represent knowledge flows. This decision has the additional effect of limiting our period of analysis to citing patents applied for after the year 2000 since the data on which citations were added by examiners is available for patents from only 2000 onwards. We restrict our sample to patents applied for between 2001 - 2012, but citations received till 2015.

Our primary dependent variable is the total count of citations received (till 2015) by patents belonging to a region-year. For robustness, we also report results for total non-self citations received as the dependent variable. Self citations are patent citations where the assignee id on both the citing patent and the cited patent are identical. Our independent variables are the share of citations made to each of the four categories in our defined framework: those to a) same region, same assignee, b) same region, different assignee, c) different region, same assignee, and d) different region,

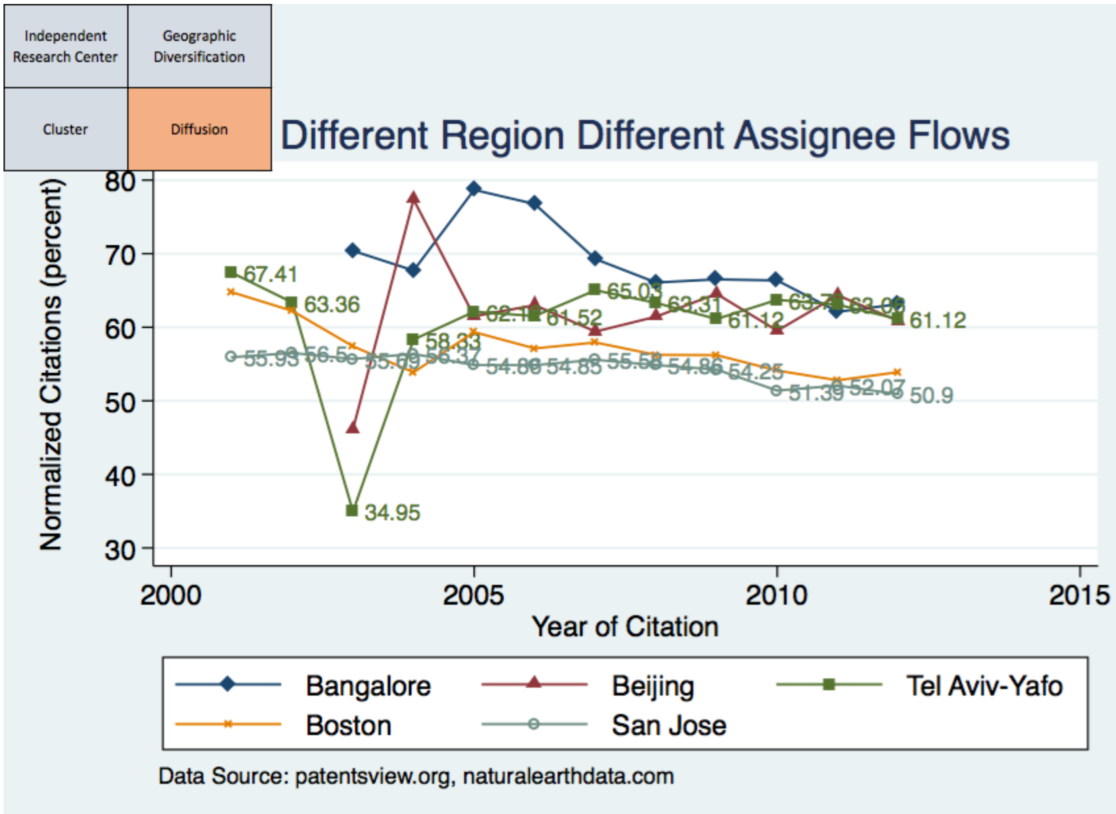

Figure 5: Flows across regions and across assignees

same assignee.

We control for the total number of citations made in the region-year, the total number of patents in the region-year), the size of the patent pool in the region-year, as well as the percentage of patents in region-year in each technology subcategory as defined by Hall et al. (2001). We include region fixed effects and year dummies in all regression models so as to control for region level and year specific effects, if any. Since our dependent variable is a count variable, we used negative binomial regression analysis with fixed effects.

Results

The preliminary results from our analysis are presented in Table 1. Model 1 reports the results for all regions worldwide, while Model 2 and Model 3 report results for U.S. locations and non-U.S. locations respectively. We find across the three models, that local knowledge flows within the firm do not seem to have a significant impact on the quality of inventions as measured by the number of citations received. This result is robust to using non-self citations as the dependent variable as presented in Table **??**.

Our preliminary findings in Table 1 suggest no evidence that local knowledge spillovers lead to higher quality inventions. On the other hand, building on knowledge from outside the firm and outside the region seems to improves invention quality. One possible explanation for the lack of evidence for local knowledge spillovers may be to suggest U.S. specific effects. However as the results in model 2 of Table 1 suggest, the trend is confirmed in a U.S.-only sample. It seems therefore that the implication for the effect of localized knowledge flows on invention quality is robust to regional or country sampling.

The effect of non-local knowledge flows however is less clear. While geographic diversification is seen to benefit all regions in Table 1, the effect is not statistically significant for non-U.S. locations in Table 2. This may suggest that there are regional, and potentially country effects that may be able to explain the phenomenon better.

Limitations

While the use of patent citations as a measure of knowledge flows has been popular in the literature, as Arora et al. (2017) point out, this may be subject to error. Our definition of regions is dependent on the latitude/longitude assignment in the patentsview.org data and on the urban centers definition in the Natural Earth Data. Any systematic biases in the definition of regions can create biases in measures of within and outside region knowledge flows.

Conclusion

While still at a preliminary stage, our analysis seem to suggest that local knowledge flows as measured by patent citations may not have a significant effect on the quality of invention produced in a region. This casts a doubt on the widely accepted idea that local knowledge spillovers are an important source of agglomeration economies. A potential extension of the study could be to analyze at level of firm-year rather than region-year. We could additionally look at the additional dimension of technology (within and outside technological domain) in addition to those of within/outside region and within/outside firm. This may provide us a more nuanced understanding of the factors affecting invention quality. Future studies could potentially examine other measures of invention outcomes such as breakthrough inventions. Finally, while our work suggests that local knowledge spillovers do not effect invention quality, it is not quite as clear why this may be the case and the mechanisms that underlie such an effect. We hope that our current work spurs further research in this direction.

References

Agrawal, A., Cockburn, I., Galasso, A., and Oettl, A. (2014). Why are some regions more innovative than others? the role of small firms in the presence of large labs. *Journal of Urban Economics*, 81:149 – 165.

Alcácer, J. and Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics*, 88(4):774–779.

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

Arora, A., Belenzon, S., and Lee, H. (2017). Reversed citations and the localization of knowledge spillovers. Working Paper 23036, National Bureau of Economic Research.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The nber patent citation data file: Lessons, insights and methodological tools. Working Paper 8498, National Bureau of Economic Research.

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.

Krugman, P. R. (1991). *Geography and trade*. MIT press.

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

Porter, M. E. (1990). *The Competitve Advantage of Nations*. Free Press New York.

Schneider, A., Friedl, M. A., McIver, D. K., and Woodcock, C. E. (2003). Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 69(12):1377–1386.

Table 1: Distribution of Citations Made on Total Citations Received

	(1) Citations Received	(2) Citations Received	(3) Citations Received
Share Citations Made[Same Region, Same Assignee]	-0.169	-0.104	-0.134
	(0.218)	(0.623)	(0.454)
Share Citations Made[Same Region, Different Assignee]	-0.149	-0.191	-0.0919
	(0.215)	(0.458)	(0.468)
Share Citations Made[Different Region, Same Assignee]	0.217	0.266	0.348
	(0.012)	(0.037)	(0.004)
Share Citations Made[Different Region, Different Assignee]	0.00828	0.0315	0.0179
	(0.789)	(0.499)	(0.661)
Log(Total Citations Made)	0.0180	0.0198	0.0134
	(0.000)	(0.001)	(0.014)
Log (Num Patents)	0.785	0.811	0.830
	(0.000)	(0.000)	(0.000)
Log (Patent Pool Size)	-0.111	-0.227	-0.0944
	(0.000)	(0.000)	(0.000)
Observations	9241	3885	5356
Groups	1314	529	785
Sample	All Locations	U.S. Locations	Non-U.S. Locations

p-values in parentheses
All models include region fixed effects, year dummies and technology subcategory controls

Table 2: Distribution of Citations Made on Non-Self Citations Received

	(1) Non-Self Citations Received	(2) Non-Self Citations Received	(3) Non-Self Citations Received
Share Citations Made[Same Region, Same Assignee]	-0.232	-0.183	-0.316
	(0.105)	(0.392)	(0.123)
Share Citations Made[Same Region, Different Assignee]	0.0391	0.0945	0.0505
	(0.733)	(0.721)	(0.699)
Share Citations Made[Different Region, Same Assignee]	0.195	0.263	0.174
	(0.035)	(0.041)	(0.212)
Share Citations Made[Different Region, Different Assignee]	0.00989	0.0203	0.00983
	(0.762)	(0.675)	(0.826)
Log(Total Citations Made)	0.0150	0.0139	0.0107
	(0.000)	(0.023)	(0.075)
Log (Num Patents)	0.772	0.795	0.828
	(0.000)	(0.000)	(0.000)
Log (Patent Pool Size)	-0.0453	-0.108	-0.0685
	(0.039)	(0.006)	(0.017)
Observations	8879	3732	5147
Groups	1199	478	721
Sample	All Locations	U.S. Locations	Non-U.S. Locations

p-values in parentheses
All models include region fixed effects, year dummies and technology subcategory controls