Africa Growth Initiative at BROOKINGS

AFRICAN DEVELOPMENT BANK GROUP

UNITED NATIONS UNIVERSITY

UNU-WIDER

World Institute for Development

LEARNING TO COMPETE

Working Paper No. 2

Are Spatial Networks of Firms Random?
Evidence from Vietnam

Emma Howard[1], Carol Newman[1] and Jacco Thijssen[2]

Abstract

We present a new approach for the empirical investigation of agglomeration patterns. We examine the clustering of manufacturing firms by identifying patterns of spatial network formation that deviate from randomly generated networks. Using firm-level panel data from Vietnam we calculate transitivity, a measure to determine the strength of clustering of manufacturing firms. We then test whether the observed clustering of firms is greater than that of a randomly generated network. Our findings suggest that the extent of clustering is over and above that which can be attributed to the legal and regulatory framework, economic zoning, or population patterns.

Keywords: D21, D22, D85, L14, O12, O25

JEL classification: Clustering, network formation, manufacturing firms, Vietnam

[1]Department of Economics, Trinity College Dublin, Ireland; [2] Department of Economics & Related Studies, University of York, UK.

The Brookings Institution is a private non-profit organization. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. Brookings recognizes that the value it provides is in its absolute commitment to quality, independence and impact. Activities supported by its donors reflect this commitment and the analysis and recommendations are not determined or influenced by any donation.

Learning to Compete (L2C) is a collaborative research program of the Africa Growth Initiative at Brookings (AGI), the African Development Bank, (AfDB), and the United Nations University World Institute for Development Economics Research (UNU-WIDER) on industrial development in Africa. Outputs in this Working Paper Series have been supported by all three institutions.

AGI-Brookings is grateful for the contribution of an anonymous donor for funding its work under the collaborative research program.

1 Introduction

The clustering of economic activities and the resulting agglomeration effects on firm outcomes has received considerable attention in the economics literature. It is well established that the clustering of firms not only facilitates specialisation leading to efficiency improvements, but the concentration of knowledge and the existence of a tight social network can also lead to productivity enhancing technology spillovers (Feldman 2001; Fujita et al. 1999; Gorg and Greenaway 2001; Kenney and von Burg 1999; Krugman 1991; Markusen and Venables 1999). Most of the literature focuses on why firms cluster and the existence and prevalence of clustering is usually taken as given. Two notable exceptions include Ellison and Glaeser (1997) and Duranton and Overman (2005) who provide measures of the extent of clustering in the USA and the UK, respectively.

In this study, we do not examine the reasons why firms cluster but take a step back and examine the overall pattern of clustering of manufacturing firms in Vietnam. Like Ellison and Glaeser (1997) and Duranton and Overman (2005) we believe that it is helpful to first determine that significant clustering exists before investigating the exact combination of natural endowments and external economies that lead to the particular pattern of clustering observed. Understanding to what extent the observed location patterns of firms in a country are actual departures from randomness is an important, often overlooked step in understanding whether agglomeration effects are present in an industry. While these questions are relevant in all contexts, they are of particular importance for developing countries at earlier stages of the industrialisation process where most evidence on the spatial distribution of firms is anecdotal.

In this study, we attempt to address the gap in the literature in this area in the following ways. First, we analyse the clustering patterns of manufacturing firms in Vietnam, a country that has undergone rapid industrialisation in recent years that has been characterised, anecdotally, by dynamic cluster formation. Second, we propose a new framework for understanding and measuring clustering that builds on the methodology proposed by Duranton and Overman (2005) and overcomes some of the shortcomings of their work. Third, we make use of a unique dataset of the population of registered firms in Vietnam that includes firm location.

We find a high degree of clustering in the manufacturing industry in Vietnam relative to that of a randomly generated network. We find that clustering is driven to a large extent by large enterprises but also by the location of state owned enterprises. The results of our analysis could be helpful in the design of industrial policy, in particular in designing industrial spatial strategies. To our knowledge no similar analysis has been conducted on firms in this country or in a developing country context.

The remaining sections below are organised as follows. Section 2 presents a conceptual framework for the formation of spatial clusters. The data and methodological approach are presented in Section 3. Section 4 presents the results while Section 5 describes a range of robustness checks. Section 6 concludes the paper.

2 Conceptual framework

The behaviour of economic agents, such as firms, countries, or individuals, is influenced by their relationships with other economic agents. In turn aggregate outcomes are influenced by relationships between agents. Relationships between and among agents can have a number of different dimensions. In order to meaningfully analyse the impact of an agent's relationships on their decisions, and on aggregate outcomes, we need a framework to systematically examine these relationships. Network analysis provides such a framework. A network describes a collection of nodes and the links between them. Nodes can represent individuals, firms, countries or even collections of such entities and the links between them can be defined in many different ways. For example, a network could consist of individuals as the nodes and define links as the friendships between them, or nodes could be defined as countries and links defined as trade between those countries.

A particular class of complex networks is composed of those embedded in real space, in other words networks whose nodes occupy a precise position in two/three dimensional Euclidean space and whose edges are real physical connections. When nodes and links are embedded in a physical space this induces a distance between nodes and such networks are designated 'spatial'. For example, manufacturing firms in Vietnam form a spatial network where the nodes are firms and the links are the physical distances between them. The position of a firm within the spatial network may impact on the costs and production decisions of the firm itself and on other firms within the network.

The 'symmetric connections model' (JW model) by Jackson and Wolinsky (1996) motivates our approach to analysing cluster formation. The JW model is a social network model where the nodes are agents/players and the links are the friendships between them. The model assumes that players benefit not only from direct links but also from indirect links or 'friends of friends'. This benefit is assumed to deteriorate in the distance between nodes. It is assumed that there is some cost associated with the formation of a direct link and therefore the network is not fully connected. Jackson and Rogers (2005) extend the JW model to include a simple geographic cost structure of link formation whereby agents are based on a number of 'islands' and the cost of forming a link with another agent depends on the geographic location of that agent. This cost structure, together with the deteriorating benefit structure of indirect links produces a network that exhibits 'small world' characteristics.

Jackson and Rogers (2005) (JR) apply the geographic cost structure extension of the JW model to the case of a social network. We argue that with one simple extension the model is a good fit for a spatial network of firms. Firms in clusters can experience agglomeration effects. For example, the costs of production may fall as firms have multiple competing suppliers or greater specialisation and division of labour can lead to efficiency gains. There can also be diseconomies of agglomeration. For example, a greater number of firms can increase competition and drive down prices. A manufacturing firm is affected by the production decisions of all firms in the network, but the magnitude of the impact decreases as the distance between the two firms increases. Agglomeration effects also decrease with increasing distance. In order to apply the JR model to a spatial network of firms we introduce the concept of a threshold distance and define firms to be linked if they are within a threshold distance, d^*. This implies that firms receive the 'benefit' if a firm is within the threshold distance. For the network of manufacturing firms this can be positive or negative. Firms cannot sever spatial links and they do not have any control over which firms, if any, locate near them, once they have entered the market and made their own location choice. In our extension of the JR model, the 'benefit' is simply the impact of the direct link between two firms i and j, or the effect that i locating within d^* of j has on the firms. Other firms in the spatial network are also effected by i's location choice but the 'benefit' deteriorates with increasing distance.

On entering the market each firm decides on a location. Their choice of location determines their position in the spatial network and also determines the spatial links they form with other firms. The cost associated with link formation is the cost of locating in a particular area. This cost clearly has a geographic structure as the cost of entering the market will differ depending on the area in which the firm decides to locate. Costs such as rent and land prices will vary from region to region and firms receive different incentives and levels of state investment depending on the area in which they choose to locate. What we aim to establish empirically in this paper is that this extension of the JR model gives a better explanation of the observed patterns of clustering than the alternative that firms simply form a random network.

3 Data and methodology

The data are taken from the Vietnamese Enterprise Survey for 2002-2007 inclusive, provided by the General Statistics Office of Vietnam (GSO, 2010). The dataset includes all registered manufacturing enterprises at the end of each year with more than 30 employees, plus a random sample of 15 per cent of small registered enterprises with less than 30 employees. The data includes information on the main economic activity of the firm, the type of ownership (state owned, privately owned or foreign owned), the number of employees, and the commune in which the firm is located. There are three levels of administrative areas in Vietnam; communes, districts and provinces. Communes are very small areas so commune level location information is sufficient for the analysis. The diameter of the largest commune is 66.37km. Table 1 presents the number of manufacturing firms and the percentage increase in the number of firms in each year.

Table 1: Number of manufacturing firms in Vietnam 2002-2007

Year	Number of firms	% Increase
2002	61,997	.....
2003	71,135	12.8
2004	91,499	28.6
2005	109,947	20.2
2006	128,637	17.0
2007	155,520	20.9

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

Our methodological approach to measuring clustering is similar to that proposed by Duranton and Overman (2005); however we use an alternative measure of clustering which is based on measures used in network analysis. We measure the distance between every set of firms, and define two firms to be linked if the distance between them is less than the threshold distance. In other words, $L_{ij} = 1$ if and only if $|l_i - l_j| \leq d^*$, where l_i is the location of firm *i*, l_j is the location of firm *j*, and d^* is the threshold distance. We define a *triad* as a subgraph of three nodes (*i, j, k*). A triad is called *transitive* if all three nodes are linked ($L_{ij} = 1, L_{ik} = 1$ and $L_{jk} = 1$) and *potentially transitive* if two of these three links exist. The *transitivity* of a network (τ) is defined as the ratio of the number of transitive triads to the number of potentially transitive triads, i.e.

$$\tau = \frac{\sum_{i,j,k}^{N} L_{ij} L_{ik} L_{jk}}{\sum_{i,j,k}^{N} L_{ij} L_{ik} + \sum_{i,j,k}^{N} L_{ij} L_{jk} + \sum_{i,j,k}^{N} L_{ik} L_{jk}} \quad (1)$$

The spatial network is assumed to be undirected in that $L_{ij} = L_{ji}$, $L_{ik} = L_{ki}$ and $L_{jk} = L_{kj}$. The transitivity has an upper bound of 1 and a lower bound of zero. If firm *i* is linked to firm *j* and firm *i* is also linked to firm *k*, the transitivity can be interpreted as the probability that firms *j* and *k* are also linked.

The threshold distance is chosen to be the diameter of the largest commune and firms are therefore defined to be linked if the distance between them is less than or equal to 66.34km. The choice of the diameter of the largest commune as the threshold distance is to ensure that if firms are located beside one another but in adjacent communes, they will still be linked. The use of a threshold distance rather than a geographic region, such as a commune, avoids problems associated with firms located on the border of regions (see for example Gorman and Kulkarni (2004)). A matrix of links is then created and the transitivity of the network is calculated. This procedure is repeated for all years.

The measure proposed here is similar to that of Duranton and Overman (2005). They plot the location of each firm and measure the distance between all pairs of firms. They then construct a measure of localisation called a K-density. For each four-digit sector they compute the density of bilateral distances between all pairs of establishments and compare this density to counterfactuals generated by randomly re-shuffling plants across sites. Their measure assesses departures from randomness in the location of firms in a four-digit sector, conditional on overall manufacturing agglomeration. Our approach departs from this in several ways that represent a number of unique methodological contributions to the literature on clustering and spatial networks.

First, unlike the K-density measure, the use of a threshold distance allows us to calculate transitivity when data on the location of firms is not exact, as is the case in our study.[1] In addition, as we are interested in the clustering of firms and agglomeration effects, it makes sense in the context of our conceptual framework to choose a threshold distance. Second, rather than analysing four-digit sub-sectors of the manufacturing industry, we look at the manufacturing industry as a whole. Duranton and Overman (2005) acknowledge that analysing four digit sectors can create problems, as two firms within the same sector do not necessarily produce the same thing[2], and suggest that one way to overcome this problem is to consider the whole population of manufacturing firms, which is what we do in this paper. Finally, this measure of clustering also has the advantage of being comparable not only across industries but also across countries and other networks (provided the same threshold distance is used).

4 Results

---

[1] Our data tell us which commune the firms are in but not the exact addresses of the firms. Consequently we could not do a full spatial statistical analysis as the exact distance between firms is not known. Given these data limitations the use of a threshold distance is appealing.

[2] Duranton and Overman (2005) use the example of naval constructions; large firms will mostly be located along the coast, but there may be smaller firms inland, in the same sector, but producing entirely different goods.

Before investigating the extent of clustering of manufacturing firms in Vietnam, we first plot the location of each firm to generate a visual representation of the location pattern. In order to plot the location of the firms, the midpoint of each commune is calculated. As the exact location of a firm within a commune is unknown, all firms are plotted in the midpoint of the commune in which they are located. As communes are very small in area this has a negligible impact on the plot of the firms.

Figures 1 to 6 in the Appendix show the plots of the locations of all manufacturing firms for the years 2002 to 2007. The firms are plotted semi-transparently so that the clusters appear darker the more firms are located in a particular commune. Of particular interest is the pattern of clustering of firms. It is easy to see that in 2002 there are two main large clusters, located in the north near Hanoi and in the south near Ho Chi Minh City (HCMC). We can also see that over the time period of the dataset new smaller clusters appear along the coast in the east most region in south central Vietnam.

Table 2 presents the transitivity of the spatial network of all manufacturing firms for each year of the dataset. Consistent with the pattern illustrated on the maps, the transitivity of the network is increasing over the time period of the data. Transitivity can be interpreted in the following way; in 2002, the reported value of 0.9482 indicates that if firm *i* is linked with firm *j* and firm *i* is also linked with firm *k*, then there is a 94.82 per cent probability that firms *j* and *k* are also linked. This is true for all firms *i*, *j* and *k* in the network. Note that because of the way we have defined the links, this means that if firm *i* is located within 66.34km of both firms *j* and *k*, there is a 94.9 per cent probability that firms *j* and *k* are located with 66.34km of each other.

Table 2: Transitivity of the network of all manufacturing firms over time

1	2	3	4	5	6	7
Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9486	0.6869	0.7974	1.381	1.190	0.0032
2003	0.9541	0.6862	0.8004	1.390	1.192	0.0025
2004	0.9667	0.6865	0.7956	1.408	1.215	0.0028
2005	0.9693	0.6857	0.7882	1.414	1.230	0.0028
2006	0.9716	0.6870	0.7797	1.414	1.246	0.0023
2007	0.9675	0.6883	0.7705	1.406	1.256	0.0023

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

We can say that there appears to be a high degree of clustering as the transitivities are all large numbers, however this may be purely because there are a large number of firms in a small area. In other words, the more firms in Vietnam the higher we would expect the transitivity to be and so the density of the network needs to be controlled for. We generate random networks for each year and compare the transitivity of the network of manufacturing firms to the transitivity of the randomly generated networks (RGNs) (Erdos and Renyi 1960). To generate the random networks for each year we take the same number of firms from that year and randomly allocate each of them to a commune. Using the same threshold distance as for the original networks we then calculate the transitivity. The results for each year are presented in column 3 of Table 2. For each year the transitivity of the network of firms is considerably higher than the transitivity of the randomly generated network.

If the data are consistent with our extension of the JR model for the location choice of firms, then we expect to observe a network with 'small world' characteristics, meaning a network that exhibits a high degree of clustering relative to that of a randomly generated network. To test if this is in fact the case, we use the benchmark suggested by Watts and Strogatz (1998). They propose that if the ratio of the clustering coefficient (or transitivity) of the actual network to the randomly generated network is greater than one, the network is a 'small world'. Column 5 of Table 2 presents this ratio for each year of our analysis. We find that the ratio is greater than 1 for all years and hence we can conclude that the network of manufacturing firms in Vietnam exhibits small world characteristics.

In the case of firms in Vietnam it could be argued that generating purely random networks to control for density is not an appropriate comparison. Randomly allocating firms to communes will of course result in firms being allocated to communes that do not in reality have any manufacturing firms, and it may not be possible for firms to locate in these communes due to legal or geographical restrictions. Therefore, to control for the regulatory framework, we follow Duranton and Overman (2005) and consider that the set of all existing sites for firms constitutes the set of all possible locations for any firm. We generate a second random network, denoted RGN*, for each year, where the firms are randomly allocated only to communes where at least one firm is located in that particular year. However comparing the observed transitivity in each year to a single RGN is somewhat problematic, as a single RGN is just one draw from the distribution of RGNs. To offer a more complete analysis we generate 30 RGN*s and consider the average transitivity. The average transitivities of these 30 random networks, for each year, are presented in column 4 of Table 2, and the ratio of the transitivities of the actual networks to the average transitivities of RGN*s is presented in the second last column of Table 2. Controlling for the regulatory framework we can still conclude that the networks of manufacturing firms exhibit a high degree of clustering as is evidenced by the fact that the ratio of transitivities is greater than one in all time periods.

In order to quantify how much higher the observed transitivities are than the mean of the distribution of RGN*s, the last column of table 2 shows the standard deviation (σ) of the distribution of RGN*s. If we take 2007 for example, the observed transitivity of 0.9675 is 85.6σ higher than the mean of the RGN* distribution. If we assume that the RGN*s are normally distributed[3] then we know that 99.7 per cent of RGN*s lie within 3σ of the mean. As the observed transitivity is 85.6σ greater than the mean we reject the null hypothesis that the spatial network of firms are drawn from the distribution of RGN*s and conclude that the network exhibits a high degree of clustering. Similar analyses of the other years leads to the same conclusion.

5 Robustness checks and further investigation

[3] The p-value for the Shapiro-Wilk test for normality is 0.31 so we fail to reject the null hypothesis of normality. Therefore our assumption that the RGN*s are normally distributed is justified.

5.1 Firm size and statistical significance

Our data include all registered enterprises with over 30 employees but only a sample of small enterprises (those firms with fewer than 30 employees). Therefore, as a robustness check, we remove all firms with fewer than 30 employees and repeat the analysis. These data represent the population of manufacturing firms with more than 30 employees and so the calculated transitivities are parameters rather than statistics. Table 3 shows the number of manufacturing firms remaining in our dataset each year after the small firms have been removed. These represent the population of 'large' manufacturing firms in Vietnam. The results of the analysis using only those firms with 30 employees or more are presented in Table 4. We find very similar measures of clustering to those presented in Table 2 when all firms are included in the analysis, revealing that the clustering of manufacturing firms is robust to the exclusion of small firms.

Table 3: Numbers of large and manufacturing firms

Year	Large Manufacturing Firms
2002	14,794
2003	16,915
2004	20,531
2005	24,044
2006	26,208
2007	31,057

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

Table 4: Transitivity of the network of large manufacturing firms over time

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9606	0.6896	0.8194	1.393	1.172	0.0049
2003	0.9696	0.6920	0.8488	1.401	1.142	0.0037
2004	0.9750	0.6914	0.8381	1.410	1.163	0.0042
2005	0.9720	0.6915	0.8311	1.010	1.170	0.0040
2006	0.9709	0.6904	0.8238	1.406	1.179	0.0043
2007	0.9644	0.6886	0.8101	1.401	1.190	0.0036

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

As Duranton and Overman (2005) note, firms with one or two workers may have different location patterns than firms with a large number of employees. In many industries small firms may be involved in very different activities to larger firms. It is also possible that the location behaviour of small and large plants simply differ. To further investigate the possibility that firms with different numbers of employees have different clustering patterns, the firms with 30 employees or more are divided into small, medium and large firms. The large firms are defined as those in the 95th percentile in terms of numbers of employees. These firms represent a large percentage of employment in the manufacturing industry in Vietnam, as revealed in Table 5. Small firms are defined as those below the 50th percentile in terms of employment while medium firms are defined as those within the 50th percentile number of employees and above, but below the 95th percentile threshold.

Table 5: Employment by size distribution of manufacturing firms in Vietnam

Year	Total Employment	95th Percentile Lower Bound	Employment in 95% Percentile	%Total Employment in 95th Percentile	50th Percentile Lower Bound
2002	4,221,624	1,152	868,066	20.6	107.5
2003	2,448,586	1,126	1,022,720	41.8	106
2004	2,757,753	1,095	1,182,508	42.9	100
2005	2,938,227	1,050	1,282,116	43.6	96
2006	3,136,891	1,069	1,402,208	44.7	97
2007	3,569,523	1,062	1,593,183	44.6	96

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

These three groups of manufacturing firms (small, medium and large) are plotted on three different maps for 2002 and 2007. Figures 7 to 12 in the Appendix show these plots based on firm size.[4] It is clear that the firms in the 95th percentile are at the very centre of the two largest clusters, near Hanoi in the north and HCMC in the south. In addition, the smaller clusters on the coast that have formed over the period of the dataset appear to be caused by large firms locating there and smaller firms clustering around them. Medium firms are located in the clusters but some are also in seemingly random locations around the country. In contrast, very few small firms are located away from clusters.

5.2 Ownership structure

We also explore the extent to which ownership structure impacts on the pattern of clustering in the manufacturing industry in Vietnam. We divide the manufacturing firms into three groups, this time defined by their ownership structure. We define state owned firms as those either wholly or partly owned by the state, private owned firms as those owned wholly by domestic non-state owners, and foreign owned firms as those either partly or wholly owned by foreigners. We plot the locations of the firms in 2002 and 2007 once again on three separate maps for the three ownership categories.[5] This allows us to see not only the pattern of clustering but also its progression over time. The resulting maps are illustrated in Figures 13 to 18 in the Appendix. Once again a clear pattern of clustering is revealed. Private enterprises are very clearly clustering around state owned enterprises in the two largest clusters while there are very few privately owned enterprises located outside of the clusters.

If we compare Figures 15 and 18 it is clear that the number of foreign owned enterprises increased over the period of the dataset. The raw data reveal a particularly pronounced increase in foreign owned firms between the end of 2006 and the end of 2007. This sharp increase is most likely due to changes in investment laws that came into effect in July 2006. Prior to July 2006 foreign and domestic investments were governed by two separate laws; the new law was intended to apply equally to domestic and foreign owned enterprises. The Investment Law introduced in 2006 was considered a significant watershed for improvement of the legal environment on investment activities and corporate governance (Foreign Investment Agency, 2009). The introduction of this new law in 2006 most likely assured and

[4] Colour maps provide a clearer picture of the clustering pattern and are available on request.

[5] Colour-coded maps available on request.

encouraged foreign investors in Vietnam, leading to an increase in the number of firms with some level of foreign ownership. In addition, although the purpose of the new investment law was to ensure both domestic and foreign-invested companies would be regulated in the same way, it has been argued that in some instances the regulatory burden to establish a business under the new law was actually greater for domestic investors (Freshfields Bruckhaus Deringer 2006). This effect may also have lead to further increases in the number of foreign owned firms relative to domestically owned firms between 2006 and 2007.

5.3 Location restriction

It could be argued that privately owned domestic firms have the greatest freedom in terms of choosing where to locate. The locations of state owned firms are determined by government rather than management of the firms and foreign owned firms are more restricted than domestically owned enterprises in terms of where they are allowed to locate. Therefore the high degree of clustering we observe in the network may be driven by the state owned enterprises and foreign owned firms and this clustering may be purely a result of government or planning policies. If this is the case then we may observe a less clustered network if we run our analysis for the network of private domestic owned firms only. The results of this analysis using only large domestic firms are presented in Table 6.

As we can see from the results presented in Table 6 when only private firms are used the resulting networks have higher transitivities than when all firms are used. So location restrictions on foreign and state owned firms are not driving the results. We can also say that the observed transitivities are significantly higher than the average transitivity of the distribution of transitivities from RGNs. If we take 2006 as an example, the observed transitivity is 7.4σ higher than the mean of the RGN* distribution. We can conclude then that the spatial network of private firms is not drawn from the RGN* distribution and exhibits a high degree of clustering.

Table 6: Transitivities of the networks of private firms only

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev RGN*
2002	0.9883	0.7698	0.9464	1.284	1.044	0.0068
2003	0.9888	0.7762	0.9386	1.274	1.053	0.0065
2004	0.9908	0.7657	0.9445	1.294	1.049	0.0064
2005	0.9901	0.7718	0.9432	1.283	1.050	0.0065
2006	0.9874	0.7746	0.9423	1.275	1.048	0.0061
2007	0.9844	0.7827	0.9260	1.258	1.063	0.0093

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

5.4 Economic/industrial zones

In 1991 the Vietnamese government introduced a policy to develop special economic zones (Foreign Investment Agency, 2009).[6] There are 96 different zones located across the country. The zones cover large areas and can span a number of communes. The zones have a high quality of infrastructure and zoned land is made available exclusively for industrial use. The government offers fiscal incentives to firms to locate within zones (Foreign Investment

6 www.business-in-asia.com/countries/vietnam_industrial_zones.html has a map of Vietnam's industrial zones.

Agency, 2009). It is therefore possible that economic zoning is driving the clustering observed in the spatial network of firms.[7] To determine if this is in fact the case we split the dataset into firms located in districts with economic zones and firms located in districts with no economic zones. We then run the analysis to determine if significant clustering occurs within economic zones and if significant clustering occurs outside economic zones. The results of the analysis are presented in Table 7. Our results reveal that there is significant clustering outside economic zones as well as within economic zones. In 2004 the within zone transitivity is 40σ greater than the mean of the RGN* distribution while outside zones the observed transitivity is 91.2σ greater than the mean of the RGN* distribution. This is evidence that it is not economic zoning that drives the clustering observed in the spatial network of firms in Vietnam. Significant clustering occurs both inside and outside of economic zones.

Table 7: Transitivity of the networks of firms within zones and outside zones

Year	Transitivity *Within Zones*	Transitivity RGN	Ratio: $c(g)/c(g^{rd})$	Std. Dev. RGN	Transitivity *Outside Zones*	Transitivity RGN	Ratio: $c(g)/c(g^{rd})$	Std. Dev. RGN
2002	0.9742	0.8220	1.185	0.004	0.9555	0.8266	1.156	0.007
2003	0.9572	0.7978	1.200	0.003	0.9653	0.8097	1.192	0.005
2004	0.9684	0.8080	1.199	0.004	0.9728	0.7903	1.231	0.002
2005	0.9680	0.8015	1.207	0.004	0.9767	0.7847	1.245	0.004
2006	0.9277	0.8024	1.156	0.004	0.9780	0.7747	1.262	0.003
2007	0.9646	0.7963	1.211	0.003	0.9753	0.7589	1.285	0.003

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

5.4 Choice of threshold distance

The results of our analysis may also be driven by the choice of threshold distance in the construction of the spatial network. As a final robustness check we repeat the analysis for all large firms using approximately half the original threshold distance (33km), and using approximately double the original threshold distance (132km). The results are presented in Tables 8 and 9, respectively. The choice of threshold distance does not have a significant impact on the results of the analysis. The networks still exhibit a high degree of clustering relative to the RGN.

Table 8: Transitivity of the networks of large firms over time, using small threshold

[7] It is also possible that other physical characteristics of provinces drive the observed pattern of clustering. We explore this using a linear regression of the density of firms within provinces as the dependent variable and a number of province characteristics such as the number of ports, train stations, airports and industrial zones as explanatory variables. The number of industrial zones is the only significant variable, and so we exclusively focus on this in our analysis. (Results available on request).

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9385	0.6645	0.8651	1.412	1.085	0.0045
2003	0.9303	0.6635	0.8842	1.402	1.052	0.0042
2004	0.9338	0.6625	0.8742	1.410	1.068	0.0047
2005	0.9273	0.6705	0.8606	1.383	1.078	0.0041
2006	0.9168	0.6610	0.8728	1.387	1.050	0.0040
2007	0.9674	0.6615	0.8376	1.462	1.155	0.0045

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

Table 9: Transitivity of the networks of large firms over time, using large threshold

Year	Transitivity	Transitivity RGN	Transitivity RGN*	Ratio: $c(g)/c(g^{rd})$	Ratio: $c(g)/c(g^{rd*})$	Std. Dev. RGN*
2002	0.9577	0.7722	0.8765	1.240	1.093	0.0025
2003	0.9731	0.7771	0.9017	1.252	1.078	0.0051
2004	0.9736	0.7779	0.9015	1.252	1.073	0.0055
2005	0.9730	0.7760	0.9056	1.254	1.074	0.0018
2006	0.9722	0.7773	0.8984	1.251	1.082	0.0027
2007	0.9708	0.7783	0.8949	1.247	1.085	0.0024

Sources: The Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations.

5.5 Population density

It is possible that the pattern of clustering we observe is driven by population density; firms could simply choose to locate where there are workers available to them and so clusters form in areas with high population density. To investigate if population density drives the observed high degree of clustering in the spatial network of manufacturing firms we use provincial-level population density data for 2008 to construct a set of province location probabilities.[8] We assume that the higher the population density is the more likely firms are to locate in a particular province. Using the same number of firms as in our dataset in 2007 we randomly allocate firms to communes based on the province location probabilities. As before, we define any commune in which a firm is actually located in 2007 to be a commune in which a firm may choose to locate and restrict the random allocation of firms to these communes. We generate 30 random networks based on the province location probabilities and calculate the transitivity of the random network for each. The resulting distribution of random networks has a mean transitivity of 0.7894 and a standard deviation (σ) of 0.0015. If we compare this to the observed transitivity for the spatial network of firms in 2007 of 0.9644 it is 116.7σ greater than the mean of the distribution of RGN*s based on population density. We can therefore conclude that the population density alone is not driving the high degree of clustering observed.

6 Conclusion

8 Data are provided by the General Statistics Office of Vietnam (2008). Data are not available for other years and so it is not possible to complete the analysis for all years of our data.

Even though we do not attempt to explain the reasons why firms cluster in this paper, we take a step forward in the literature by trying to determine if significant clustering actually occurs. We do this by comparing, statistically, whether the clustering of firms is better explained by a random graph network or by a network generated by our extension of the JR model. With the exceptions of Ellison and Glaeser (1997) and Duranton and Overman (2005) most papers take clustering as given and make no attempt to measure the extent and existence of clustering within an industry. Any analysis of why firms cluster should begin by determining that the observed pattern of clustering is a significant departure from randomness. Our paper uses tools of network analysis to measure the clustering of manufacturing firms in Vietnam. This paper is the first to apply the transitivity measure (a network analysis measure of clustering) to firms, and the first to attempt to measure clustering in a developing country context.

We find that there is significant clustering in the spatial network of manufacturing firms in Vietnam. The difference between the observed transitivity and the mean transitivity of the distribution of randomly generated networks is large enough (in terms of standard deviations) in all years to reject the null hypothesis that the manufacturing firms in Vietnam form a random network.

Marshall (1920) proposed three theoretical reasons why firms cluster; the clustering of economic activity reduces the transport costs for firms; where industry is concentrated a large pool of labour will also emerge facilitating better matching of workers to employers; and information and technology spillovers are more likely when firms are geographically located close together. Before we can establish which of these agglomerative forces are at work in a country, we must first establish that there is significant clustering, and that this clustering is not driven purely by the regulatory framework, population patterns or government industrial policy. We have established that this is the case for Vietnam. Manufacturing firms are highly clustered, and this clustering is not driven by economic zoning, population density or location restrictions. This paper provides valuable insight into the extent and pattern of clustering within Vietnam, and paves the way for similar investigations for other developing countries.

References

Duranton, Gilles and Henry G. Overman (2005). 'Testing for Localization Using Micro-Geographic Data'. *Review of Economic Studies*, 72(4): 1077-1106.

Ellison, Glenn and Edward L. Glaeser (1997). 'Geographic Concentration in US Manufacturing Industries: A Dartboard Approach'. *Journal of Political Economy*, 105(5): 889–927.

Erdos, P., and A. Rényi (1960). 'On the evolution of random graphs'. *Publications of the Institute of Mathematics, Hungarian Academy of Science*, 5: 17-60.

Feldman M.P. (2001). 'The entrepreneurial event revisited: firm formation in a regional context'. *Industrial and Corporate Change*, 10(4): 861–91.

Foreign Investment Agency (FIA), Ministry of Planning and Investment of Vietnam (2009). Vietnam Investment Guide. Downloaded on 20.1.2011 from: http://www.aseankorea.org/files/upload/board/120/6/2009%20Investment%20Guidebook%20Vietnam.pdf.

Freshfields Bruckhaus Deringer (2006). 'Vietnam-new investment law'. Briefing downloaded on 20.1.2011 from: http://www.freshfields.com/publications/pdfs/2006/14675.pdf.

Fujita, M., P.R. Krugman and A J. Venables (1999). *The Spatial Economy: Cities, Regions and International Trade*. Cambridge MA: MIT Press.

Gorg, Holger and David Greenaway (2001). 'Foreign direct investment and intra-industry spillovers: A review of the literature'. GEP Research Paper 37. Nottingham: University of Nottingham.

Gorman S.P., and R. Kulkarni (2004). 'Spatial small worlds: new geographic patterns for an information economy'. *Environment and Planning B: Planning and Design*, 31(2): 273-96.

Jackson, Matthew O., and Brian W. Rogers (2005). 'The Economics of Small Worlds'. *Journal of the European Economic Association*, 3(2-3): 617-27.

Jackson, Matthew O., and Asher Wolinsky (1996). 'A Strategic Model of Social and Economic Networks'. *Journal of Economic Theory*, 71:44-74.

Kenney, M., and U. von Burg (1999). 'Technology, Entrepreneurship and Path Dependence: Industrial Clustering in Silicon Valley and Route 128'. *Industrial and Corporate Change,* 8(1): 67–103.

Krugman, P. R. (1991). 'Increasing Returns and Economic Geography'. *Journal of Political Economy*, 99: 483-99.

Markusen, J., and A. Venables (1999). 'Foreign Direct Investment as a Catalyst for Industrial Development'. *European Economic Review*, 43: 335–56.

Marshall, Alfred (1920). *Principles of Economics*. London: Macmillan and Co. Ltd.

Watts, D. J., and S.H. Strogatz (1998). 'Collective dynamics of "small-world" networks'. *Nature*, 393: 440–442.

Appendix

Figure 1: Clustering 2002

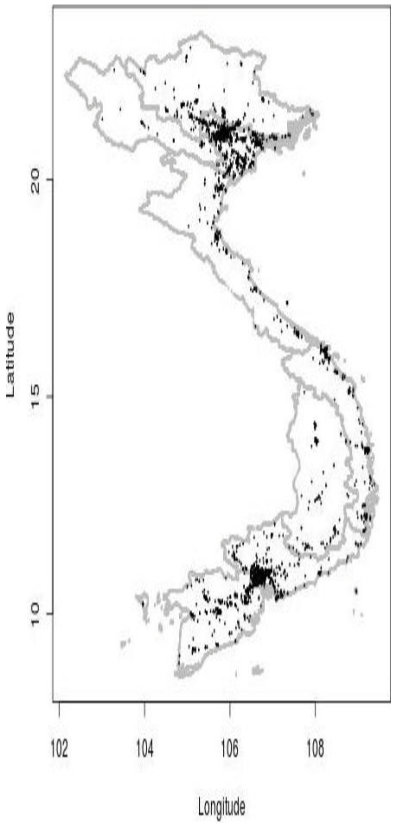


Figure 2: Clustering 2003

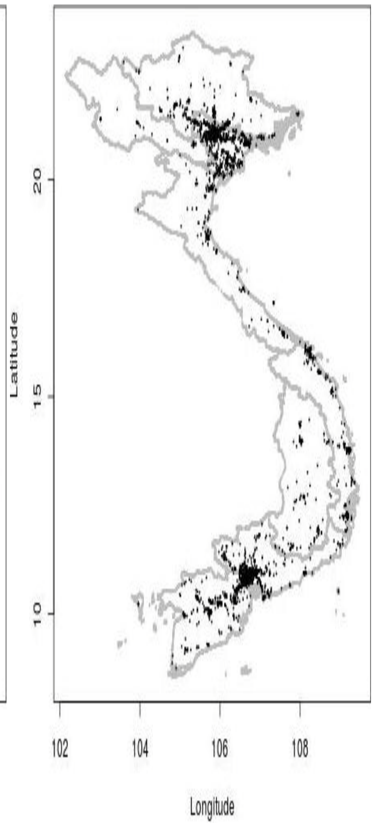


Figure 3: Clustering 2004

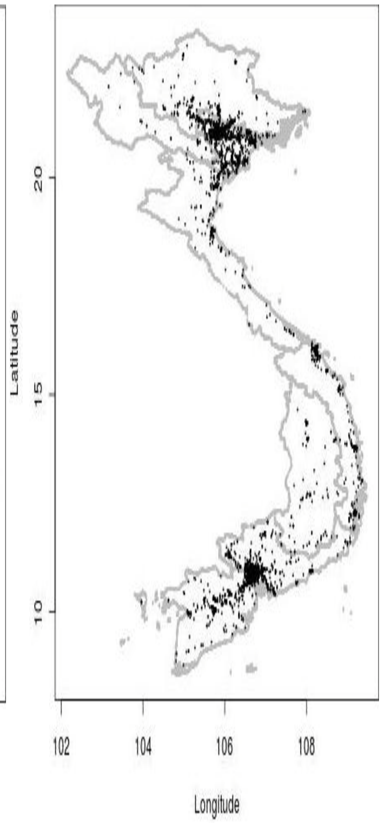


Figure 4: Clustering 2005

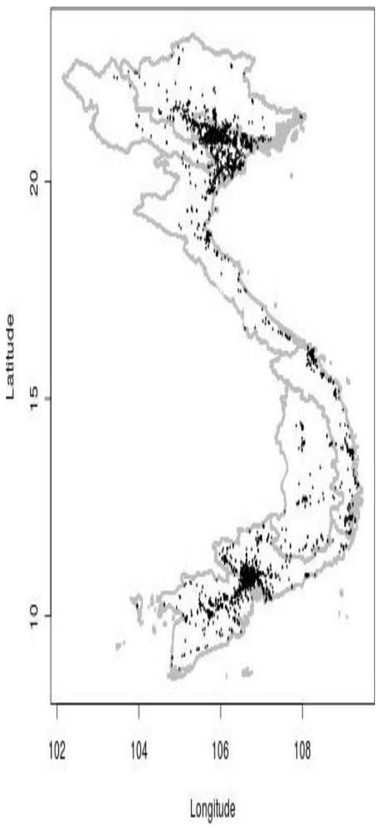


Figure 5: Clustering 2006

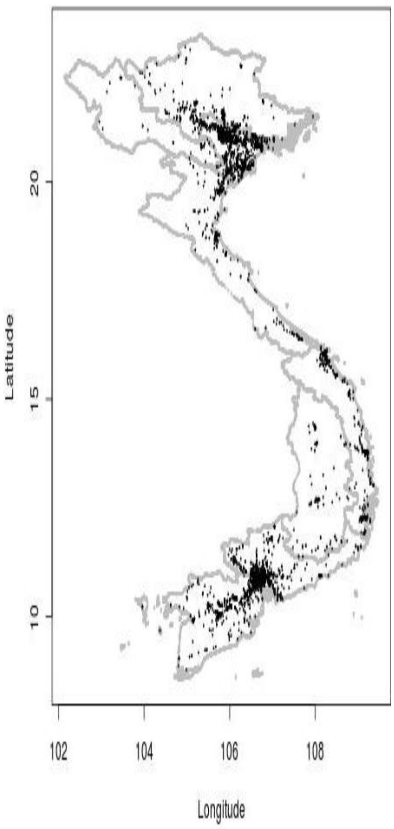


Figure 6: Clustering 2007

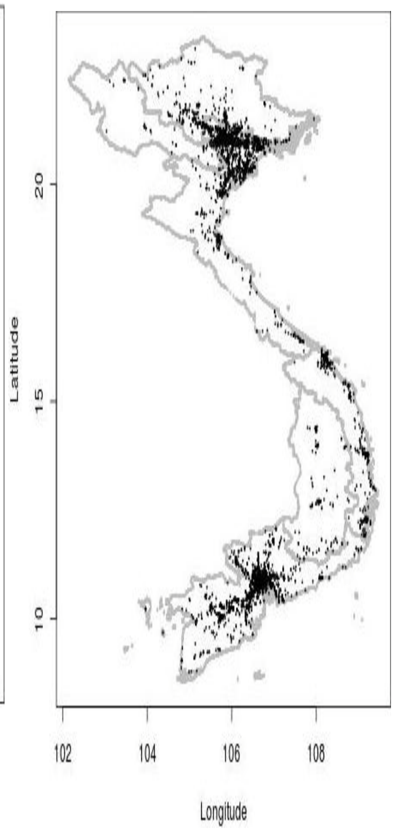
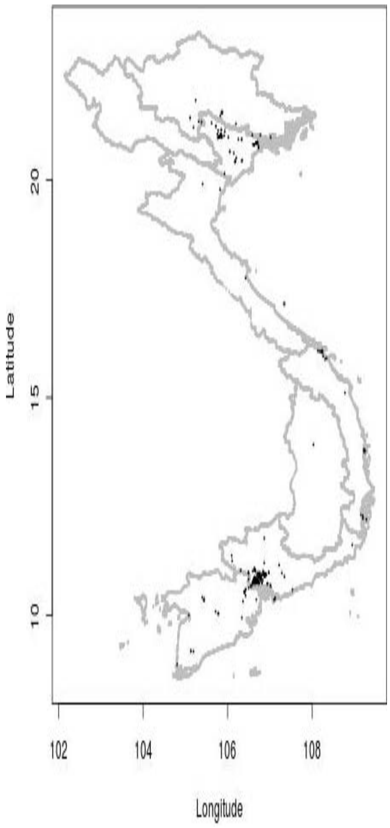
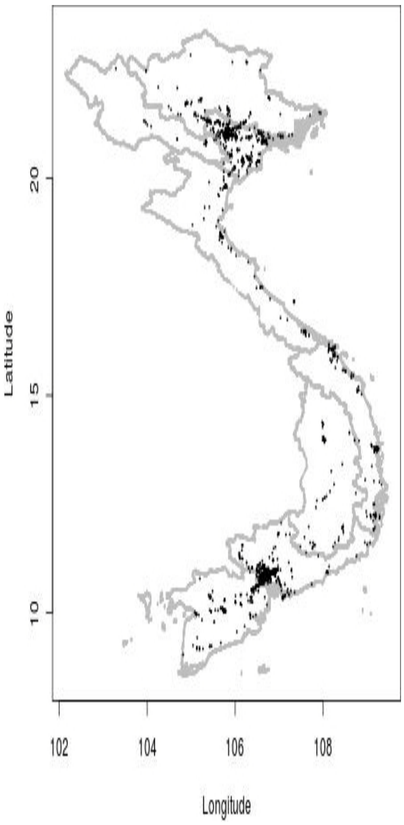


Figure 7: Large Firms 2002 Figure 8: Small Firms 2002 Figure 9: Medium
Firms 2002





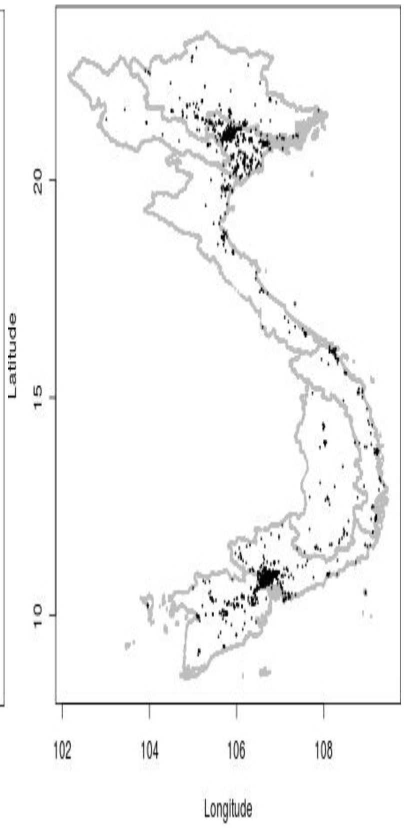
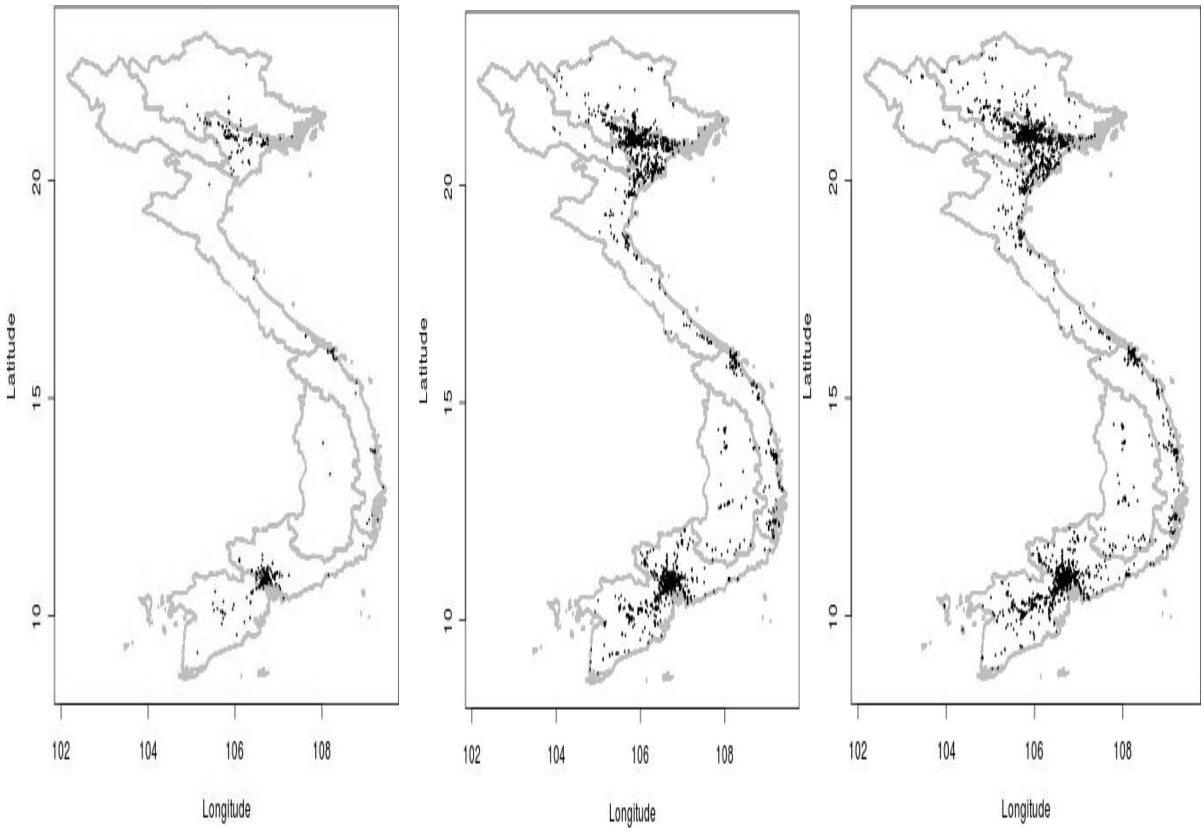


Figure 10: Large Firms 2007 Figure 11: Small Firms 2007 Figure 12: Medium
Firms 2007



Latitude
20
15
10
102
104
106
108
Longitude

Figure 13: State-owned 2002

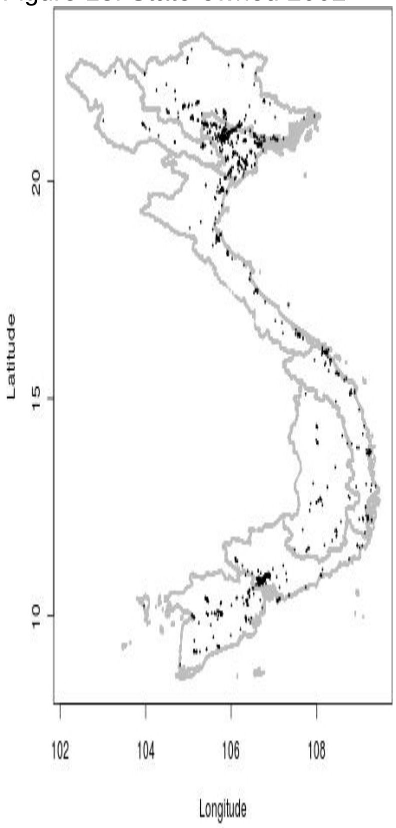


Figure 14: Private owned 2002

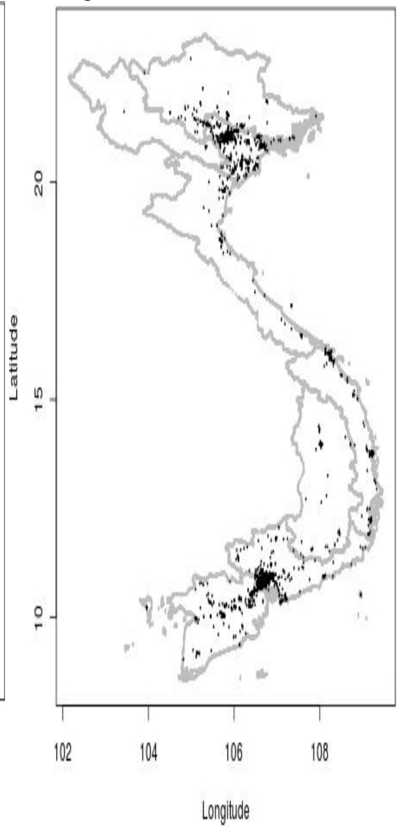


Figure 15: Foreign owned 2002

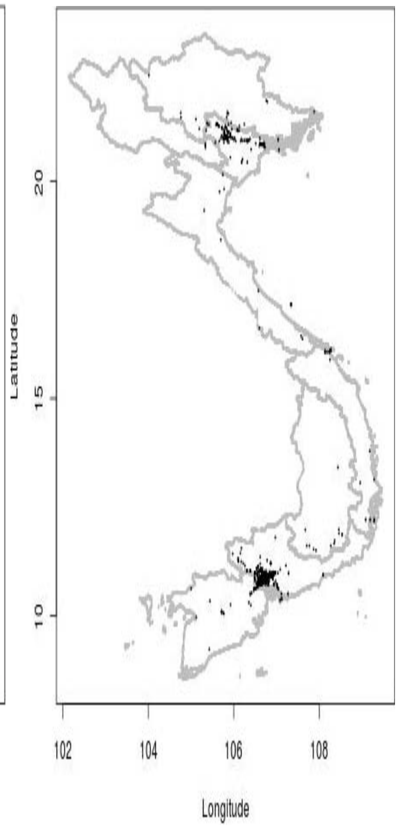


Figure 16: State owned 2007

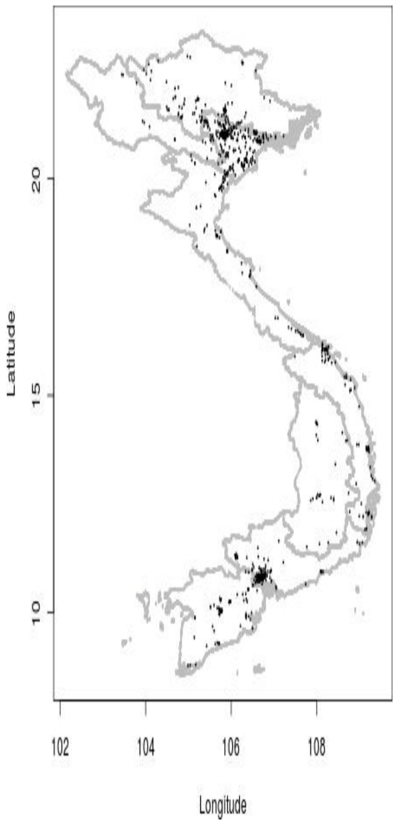


Figure 17: Private owned 2007

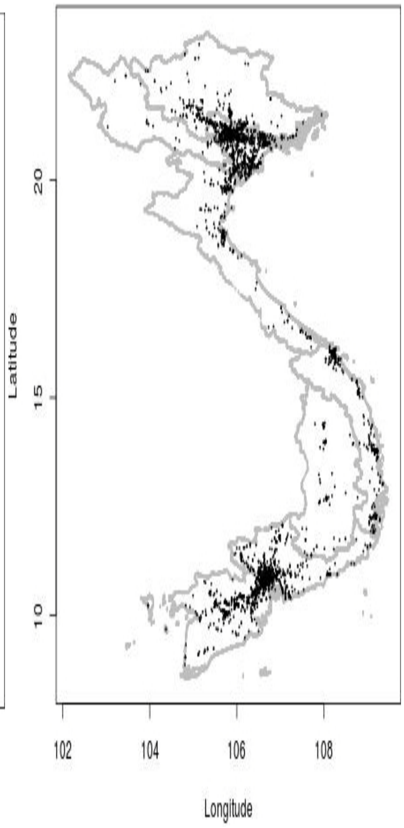
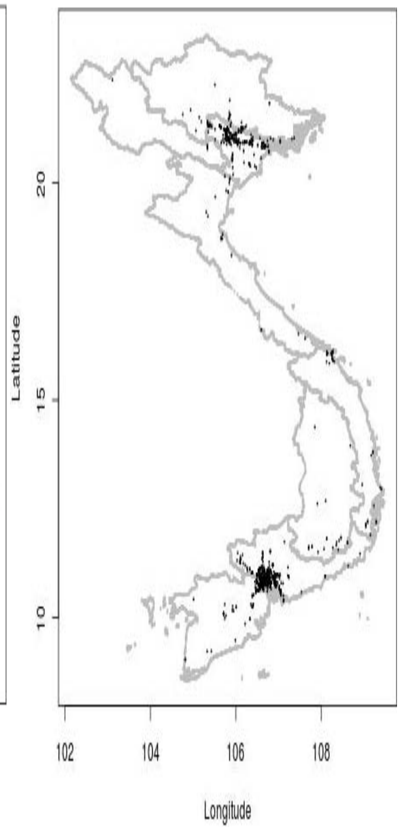


Figure 18: Foreign owned 2007



Sources for Figures 1-18: the Vietnamese Enterprise Survey 2002-07, provided by the General Statistics Office of Vietnam 2010; authors' calculations and illustrations.