Identifying Industry Clusters in Colombia Based on Graph Theory

Juan Carlos Duque[1] Dairo Andrés Gómez [2]

October, 2008

[1]Research in Spatial Economics (RISE-group), Department of Economics, EAFIT University, jduquec1@eafit.edu.co
[2]Research in Spatial Economics (RISE-group), Department of Economics, EAFIT University, dgomezca@eafit.edu.co

Identifying Industry Clusters in Colombia Based on Graph Theory

ABSTRACT: This paper presents a new way to identify and understand the industry clusters in the Colombian economy. The analysis relies on a recent methodology proposed by Duque and Rey (2008) where intrincated input-output linkages between industries are simplified using network analysis. In addition to other techniques for cluster identification available in the literature, this novel methodology allows not only to classify each industry in a given cluster, but also to understand how industries are related to each other within their clusters. This methodology offers a conciliatory approach to two radically different views about the economic unit from which policy makers should design their strategies for resource allocation: Porter's cluster strategy versus Hausmann's industrial targeting.

JEL classification: C67, D57, L22.

Palabras Clave: industry clusters, graph theory, input-output, impact analysis.

RESUMEN: Este paper presenta una nueva forma de identificar y entender los clusters industriales en la economía colombiana. Este análisis se basa en una metodología propuesta recientemente por Duque and Rey (2008) en la cual se aplica teoría de redes para simplificar los complejos vínculos comerciales entre industrias presentes en una matríz insumo-producto. En comparación con otras técnicas existentes en la literatura, esta novedosa técnica permite no solo clasificar cada industria dentro de un cluster, sino también entender como las industrias están relacionadas dentro de su cluster. Esta metodología ofrece una aproximación conciliadora entre dos puntos de vista radicalmente diferentes con respecto a la unidad económica sobre la cual se deben diseñar las estrategias para la asignación de recursos: La visión de Porter basada en el apoyo a los clusters versus la visión de Hausmann basada en el apoyo a industrias estratégicamente seleccionadas.

Clasificación JEL: C67, D57, L22.

Key Words: clusters industriales, teoría de redes, insumo-producto, análisis de impacto.

1 Introduction

Industry clusters have been a matter of investigation since the beginning of twentieth century. One of the first references dates back to 1890 when Marshall (1890) introduced the concept of “agglomeration economies” as the benefits derived from the synergy that industries can generate by locating near each other. While the concept of “agglomeration economies” focuses on the spatial allocation of industries, the term “industry cluster”, introduced by Porter (1990), define the benefits derived from the vertical or horizontal relationships between the industries of a given economy.

> “Vertical clusters are those that gather industries characterised by buyer-supplier relations. While horizontal clusters include industries that share a common market for final goods, or use same technology or employees, or need a similar natural resource”
>
> (Porter, 2003, p. 205, translated)

Kaufman et al. (1994) stress that cluster analysis offers guidance to policy makers in the identification of a state’s competitive advantage. In a same way, Doeringer and Terkla (1995) state that by widening the focus of development policies, cluster analysis offers the possibility of integrating non-export as well as export-based industries into regional growth strategies.

Nowadays, industry cluster identification is still an important research topic in public, private and academic sectors. On one hand, a great deal of cluster observatories have been created worldwide as an important resource for policy makers and planners who are concerned with the strategic and tactical deployment of resources (Europe INNOVA, 2007). On the other hand, cluster identification is also beneficial for industries since they provide insights about potential clients and suppliers, alternative markets, and as a way to clarify the roll of the industry in the economy (High level advisory group on clusters, 2008).

The benefits of liberalized trade have increased the search for trade agreement between Colombia and other countries around the world. For example, the recently approval of a free trade agreement with the European Free Trade Association (EFTA), or the posibility of a trade agreement with the US, has increased the need for a deeper undestanding of Colombia’s productive structure. The identification of industry clusters in this country would help to detect strengths and weaknesses when facing the arrival of products and services from external competitors, and to implement strategies to improve the competitiveness and innovation level of local industries.

Colombia's central government is aware of this situation. Proof of this is the fact that the development of *world-class clusters* is one of the five pillars of the national policy of competitiveness and productivity (Sistema Nacional de Competitividad, 2008).

Nevertheless, there is a clear lack of research on the identification of industry clusters at national level. Two studies on this topic have been commissioned by the government. The first study dates back to 1993 when the central government[1] commissioned a study to assess the competitiveness level of Colombia and to identify the most representative clusters to establish the bases of the Colombia's competitive policy (Monitor Company, Inc, 1993). The second study is an advisory service contracted by the National Planning Department in 2007 to identify the key industries at national and regional level that could lead growth in Colombia Hausmann and Klinger (2007).

This paper presents the results of the identification of industry clusters and interindustry networks based on 2005 input-output tables for Colombia. The methodology applied in this paper has been recently developed by Duque and Rey (2008) in which an algorithm, based on network theory, is proposed to identify the most representative vertical clusters in a given economy.[2]

The outline of this paper is as follows. Section 2 describes the methodology that is applied to identify industry clusters. Section 3 offers the main results of applying this methodology using information from 2005 input-output tables for Colombia. Finally, Section 4 summarizes the main findings and provides recommendations for future work.

2 Methodology

The algorithm applied in this paper is known as the network-based industry clusters (NBIC). This algorithm was recently developed by Duque and Rey (2008) and it is designed to identify vertical industry clusters using I/O tables.[3]

---

[1] Ministry of Development, the Instituto de Fomento Industrial, Confecámaras and Bancoldex.

[2] For overviews on regional cluster research see Steiner (1998); Bergman and Feser (1999).

[3] More thorough treatments of regional input-output modeling can be found in Roepke et al. (1974); Miller and Blair (1985); Hewings and Jensen (1986); Lahr (1993); Isard et al. (1998).

Czamanski and Ablas (1979) provide a very clear definition of vertical industrial clusters:

> "…'cluster' means a subset of industries of the economy connected by flows of goods and services stronger than those linking them to the other sectors of the national economy."
>
> *Czamanski and Ablas (1979, p. 62)*

Figure 1 shows six graphs illustrating the main steps of Duque and Rey's methodology. It starts by representing the transactions between industries as a dense directed network in which each node represents an industry and the links joining each pair of nodes represent a transaction between two sectors (see figure 1 graph I).

From this initial representation, the NBIC algorithm starts a simplification process in order to transform the initial network into a network type known as a "tree network" in which each pair of nodes is connected by a unique path. This simplification process is possible given a set of assumptions associated to interindustry transactions that allow the network reduction without losing valuable information.

The first assumption declares that the transactions between two industries flow mainly in one direction.[4] This assumption allows for the first reduction in the network which consist of deleting the link representing the smallest transaction between each pair of industries (see figure 1 graph II). Thus, having $Z_{i,j}$ representing sales from industry i to j, and $Z_{j,i}$ representing sales from industry j to i, the NBIC algorithm only keeps the largest of those two values, i.e.:

$$Z_{i,j} = max(Z_{i,j}, Z_{j,i}) \tag{1}$$

Up to this point each link represents a transaction between industries. The next step transforms the directed network into a undirected network (see figure 1 graph III) by applying the following expression:

$$MRW_{i,j} = max(RW_{i,j}^{out}, RW_{i,j}^{inp}) \tag{2}$$

Where, $RW_{i,j}^{out} = \frac{z_{i,j}}{\sum_{j=1}^{n} z_{i,j}}$ and $RW_{i,j}^{inp} = \frac{z_{i,j}}{\sum_{i=1}^{n} z_{i,j}}$.

[4]Empirical evidence from San Diego California shows that 72.2% of the relative differences between opposite flows are greater than 90%.

With this transformation, the link between industries i and j has not longer a direction. This link is now a weight (MRW - Maximum Relative Weight-) that represents how important is this link for either industry i or industry j.

In the next step, Duque and Rey apply an algorithm known as the Kruskal's algorithm (Kruskal, 1956) to reduce the network from a dense network to a sparse network type known as a "tree network" in which each pair of nodes is connected by a unique path. The links that remain in the network are the ones that better summarize the most important relationships between the industries in the economy (see figure 1 graph IV).

One important characteristic in the tree network is that the removal of a link "breaks" the tree network into two disconnected subnetworks. More generally, the removal on k links breaks the tree network into $k-1$ subnetworks, with each subnetwork representing a cluster.

Thus, the next step is to break the network into subnetworks or clusters. At this point, the logical question is: How many subnetworks?

In order to decide the number of clusters, the NBIC algorithm scores each industry according to three different criteria:

- Transaction volume (TV_i): Measures the share of I-O transactions accounted for industry i, with respect the total I-O transactions in the economy.

$$TV_i = \frac{\sum_{j=1}^{n} z_{i,j} + \sum_{j=1}^{n} z_{j,i}}{\sum_{i,j} z_{i,j}} \tag{3}$$

- Adjusted transaction volume (TVa_i): This criteria seeks to give a higher score to those industries which are related to many other industries in the economy.

$$TVa_i = \frac{2 \cdot TV_i}{\left(GINI_i^{out} + GINI_i^{inp}\right)} \tag{4}$$

  Where $GINI_i^{out}$ measures how dispersed the outputs of industry i are; and $GINI_i^{inp}$ measures how dispersed the inputs of industry i are.

- Market power (MP): Consist of an iterative procedure that measures how important each industry is for its direct buyers and suppliers. Thus, those industries whose outputs (inputs) represent a high percentage of its client's inputs (supplier's outputs) will receive a high MP score.

Finally, the NBIC algorithm uses factor analysis to merge the three criteria into one single vector value by extracting the first factor (referenced to as *factor score*). This factor score makes it possible to sort the industries by their level of importance for the economy.[5]

The last step in NBIC algorithm incorporates an iterative procedure that breaks the tree network into different number of clusters (subnetworks) such that each clusters contains one core industry (see figure 1 graph V and VI). Thus, if the number of clusters is set to four, then the core industries will be the first four industries with the highest factor scores. The optimal number of clusters k is the highest value of k such that the proportion of "weak" clusters in the economy do not exceed 50%.

The proportion of weak clusters is calculated in two different ways: i) based on internal linkages, and ii) based on external linkages:

- Internal linkages: It classifies each cluster according to how important the cluster is to the industries belonging to it. Two coefficients are calculated in order to carry out this classification. First, the intra-clusters Purchase Share coefficient (PS) that measures the share of interindustry purchases made by the cluster industries that are supplied by other industries within the cluster. The second measure, intra-cluster Sales Share coefficient (SS) that measures the share of interindustry sales made by the cluster industries that are purchased by other industries within the cluster:

$$PS_c = \frac{\sum_{i \in c}^{n} \sum_{j \in c}^{n} Z_{i,j}}{\sum_{i=1}^{n} \sum_{j \in c}^{n} Z_{i,j}}, \text{ with } Z_{i,i} = 0 \tag{5}$$

$$SS_c = \frac{\sum_{i \in c}^{n} \sum_{j \in c}^{n} Z_{i,j}}{\sum_{i \in c}^{n} \sum_{j=1}^{n} Z_{i,j}}, \text{ with } Z_{i,i} = 0 \tag{6}$$

  Where $Z_{i,j}$ represents the interindustry deliveries from industry i to industry j. $i \in c$ indicates that industry i is a member of cluster c. Thus,

  - if $PS_c > \overline{PS}$ and $SS_c < \overline{SS}$, $\Rightarrow c$ is a *purchase oriented cluster*;
  - if $PS_c < \overline{PS}$ and $SS_c > \overline{SS}$, $\Rightarrow c$ is a *sale oriented cluster*;
  - if $PS_c > \overline{PS}$ and $SS_c > \overline{SS}$, $\Rightarrow c$ is a *strong cluster*;

[5] The factor analysis technique has been also applied by Czamanski (1971) within the context of industry clusters based on I-O tables.

– if $PS_c < \overline{PS}$ and $SS_c < \overline{SS}$, $\Rightarrow$ c is a *weak cluster*;

- External linkages: It classifies each cluster according to how important the cluster is to the larger economy. Two coefficients are calculated in order to carry out this classification, backward linkages (BL), and forward linkages (FL):

$$BL_c = \sum_{i \in c} \frac{\frac{1}{n}\sum_{j=1}^{n} l_{ji}}{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} l_{ij}}{n^2}} \frac{X_i}{\sum_{j \in c}^{n} X_j} \tag{7}$$

$$FL_j = \frac{\frac{1}{n}\sum_{i=1}^{n} l_{ji}}{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} l_{ij}}{n^2}} \tag{8}$$

  Where l_{ij} is each element of the the Leontief inverse for the regional input-output matrix.

  – if $BL_c > 1$ and $FL_c > 1$, $\Rightarrow$ c is a *key cluster*;
  – if $BL_c > 1$ and $FL_c < 1$, $\Rightarrow$ c is a *driver cluster*;
  – if $BL_c < 1$ and $FL_c > 1$, $\Rightarrow$ c is a *enabler cluster*;
  – if $BL_c < 1$ and $FL_c < 1$, $\Rightarrow$ c is a *weak cluster*;

The NBIC algorithm is currently being used by the San Diego East County Economic Development Council as a part of its project Connectory.com whose primary objective is to "link California businesses to each other and to provide information about the industrial/technology base of the nation's largest economy." (Connectory, 2008). It is also being used by the San Diego Association of Governments (SANDAG) which serves as the "forum for regional decision-making." (Sandag, 2008).[6]

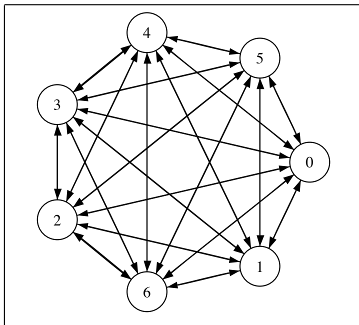
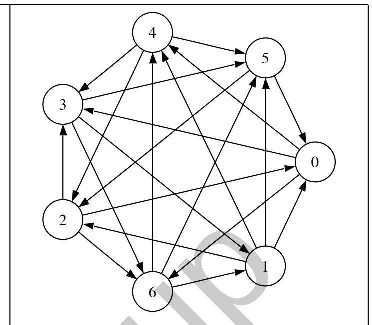
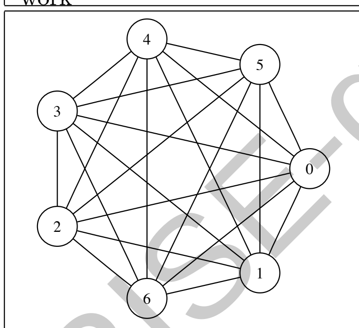
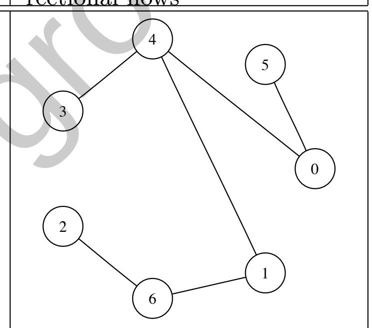
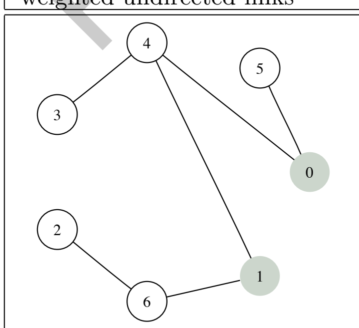
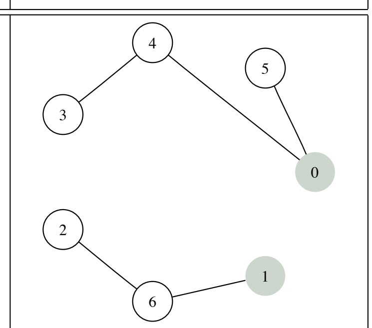
3 Empirical Application

3.1 Data

The product utilization matrix used in this application was obtained from the National Administrative Department of Statistic (DANE). Since 1951, this institution has been responsible for planning, collecting, processing,

[6]For additional information in this project see Rey et al. (2007) and Regal (2008)

Figure 1: Main steps of Duque and Rey's network-based industry clusters algorithm

	
I. I-O tables as a directed network	II. From bidirectional to unidirectional flows
	
III. From unidirectional to weighted undirected links	IV. From dense to tree network
	
V. Number of clusters and core industries	VI. Network partitioning

analysing, and disseminating the official statistics of Colombia. In this particular case, we used the projected 2005 product utilization matrix. This matrix contains the input-output transactions (in millions of COPs) between 60 industries.[7] Table 1 presents a brief description of 2005p I-O table for Colombia.

Table 1: Total interindutry output and input per industry

Code	Industry	Output	Input
1	No roasted Coffee	572.760	962.872
2	Other farming products	7.117.830	3.674.472
3	Animals and products derived from animals	12.440.416	5.854.613
4	Forestry products and wood extraction	620.008	85.699
5	Fish and other products from fishing	1.298.434	181.906
6	Lignite and Peat	314.705	1.488.147
7	Oil, natural gas, uranium and thorium minerals	4.858.510	3.525.375
8	Metallic minerals	131.385	501.656
9	Other no metallic minerals	2.308.560	275.007
10	Electricity and city gas	15.103.078	9.933.208
11	Water and sewerage services	1.014.610	1.097.957
12	Fish and Meat	1.932.724	9.950.050
13	Animal and vegetable oil, fur and cakes	1.618.824	2.077.625
14	Dairy products	722.773	3.084.200
15	Mill products, starch and its products	4.643.820	7.048.511
16	Sugar	1.073.156	1.339.273
17	Processed Coffee	291.699	578.128
18	Cacao, chocolate and sugar products	196.035	1.230.154
19	Other food products	768.981	2.006.509
20	Beverages	2.330.148	2.754.415
21	Tobacco products	8.769	198.090
22	Threads and textile fibre weaved	3.176.414	2.196.738
23	Textile items (except clothes)	-	-
24	Clothes	194.365	3.742.095
25	Leather and shoes	591.121	1.149.032
26	Wood products, cork, straw and plait materials	1.358.315	633.666
27	Cardboard and paper	5.290.013	3.395.399
28	Printing and similar goods	1.772.532	2.031.022
29	Refined oil products, nuclear combustibles and coke furnace products	11.104.862	6.548.943
30	Basic and elaborated chemical products (except plastic and rubber products)	18.947.064	9.517.910
31	Plastic and rubber products	7.468.413	3.559.427
32	Glass, glass products and other no metallic products	6.662.406	3.877.940
33	Furniture and other transportable goods	634.985	2.283.662
34	Waste products	520.799	-
35	Common metals and metallic products, except machines and equipment	14.273.878	6.775.892
36	Special and general use Machines	3.629.288	1.869.796
37	Other machines and electric supply	5.260.476	1.797.704
38	Transport equipment	5.305.930	5.236.637
39	Construction	2.298.579	13.080.363
40	Civil engineering works	2.008.051	5.797.583
41	Commerce	-	12.993.976
42	Repair services of engines and domestic stuff	4.493.140	3.146.614
43	Hotel and restaurant services	1.803.366	6.667.451
44	Road transport services	5.682.332	11.782.006
45	Fluvial transport services	288.055	573.974
46	Aerial transport services	2.294.848	2.388.190
47	Complementary road transport services	3.665.573	1.485.860
48	Post office services and telecommunications	7.206.901	3.488.016
49	Financial intermediation services, computer and related services	19.324.778	6.226.284
50	Real-state services and house renting	1.943.976	808.499
51	Services to enterprises, except financial and real-state services	11.501.014	2.072.623
52	Domestic services	-	-
53	No market education services	605.174	574.856
54	Social service and market health services	3.970.503	1.222.560
55	Leisure services and other market services	1.623.293	1.103.878
56	Government administration services and other services for the community	-	9.076.643
57	Market education services	-	915.259

continued on next page

[7]This data is available at: http://www.dane.gov.co/files/investigaciones/pib/anuales/anuales.zip. (accessed October, 2008)

Table 1: *continued*

Code	Industry	Output	Input
58	Social service and no market health services	-	5.524.693
59	Leisure services and other no market services	-	493.315
60	Financial intermediation services indirectly measured	-	12.381.296

According to Duque and Rey (2008), before applying the NBIC algorithm it is necessary to assess whether or not a network reduction can lead to a considerable loss of information. In this context, table 2 describes the distribution of transactions across deciles of links. As can be seen, 10% of the I-O links account for 76.11% of all the interindustry transactions in Colombia's economy.[8] This level of concentration in interindustry transactions is a first step to guarantee that the network reduction will not be too damaging for the analysis. The second step in assessing the impact of this reduction is to calculate the relative difference between opposite flows. The results in table 3 show that in 74.18% of the pairwise industry relationships, the difference between opposite flows is greater than 90%.[9] This difference also guarantee that the reduction described in equation 1 is valid for the Colombian economy.

Table 2: Analysis by Deciles of Transaction volume

Decile	Transaction Volume
10th	76.11%
20th	87.93%
30th	93.68%
40th	96.71%
50th	98.37%
60th	99.27%
70th	99.72%
80th	99.93%
90th	99.99%
100th	100.00%

[8]This results are similar to the ones obtained by Duque and Rey for San Diego's economy.

[9]Duque and Rey reports 72.2% of relative difference between opposite flows for San Diego's economy.

Table 3: Relative difference between pairwise opposite flows

Relative difference	% of transaction
0-10%	2.11%
10-20%	2.26%
20-30%	2.50%
30-40%	2.03%
40-50%	1.79%
50-60%	2.89%
60-70%	3.59%
70-80%	3.74%
80-90%	4.91%
90-100%	74.18%

3.2 Evaluation Measures

In this section we report on the results of applying the NBIC algorithm using the 2005p I-O table for Colombia. Following the procedure for determining the optimal number of clusters, the industries can be aggregated into twelve clusters. At this level of aggregation the proportion of weak clusters do not exceed 50% for neither the classification based on internal linkages nor the classification based on external linkages.

Table 4 presents several characteristics of the twelve industry clusters.

Table 4: Summary of the clusters

Name	Core*	Size	Total Output	Output Share	BL	FL	Type	PS	SS	Type
Construction	35	13	106380839.00	18.20	1.27	1.06	key	0.45	0.37	strong
Petrochemical	30	6	75244833.00	12.87	1.12	1.10	key	0.41	0.31	purchase
Transport	44	6	61860666.00	10.58	1.23	0.99	driver	0.27	0.35	sales
Food	3	7	56094322.00	9.59	1.30	1.02	key	0.52	0.65	strong
Commerce	41	5	51931533.00	8.88	1.11	0.85	driver	0.19	0.33	weak
Educational Services	56	3	47953590.00	8.20	1.31	0.66	driver	0.03	0.47	sales
Restaurants	43	6	39058335.00	6.68	1.09	0.77	driver	0.21	0.49	sales
Energy for industries	29	3	36727719.00	6.28	1.03	1.17	key	0.34	0.23	purchase
Banking	49	2	35045780.00	5.99	1.17	1.08	key	0.67	0.64	strong
Farming	2	3	29587632.00	5.06	0.89	0.92	weak	0.12	0.08	weak
Energy for resid. and comm. use	10	2	26369770.00	4.51	1.20	1.41	key	0.02	0.02	weak
Public Utilities	51	2	18390011.00	3.15	0.80	1.36	enabler	0.11	0.03	weak
Total		58	584645030.00	100.00				0.28	0.33	← avg

* *Id of the core industry in each cluster*

The size of the clusters, measured as the number of industry members, varies from 2 to 13 industries. *Construction* is the largest cluster in the economy with 13 industries representing a 22.41% of the industries included in this study. *Energy for residential and commercial use*, *Banking*, and *Public Utilities* are the smallest clusters in the economy. Each of these clusters contain 2 industry members.

The total output (production) of a given industry corresponds to the sum of all interindustry sales and its sales to final demand, which includes purchases by consumers, governments, and sales to other activities of investment goods. The total output of a cluster is then the sum of the total output of each industry assigned to the cluster; and the total output share is estimated as the ratio between the total output of each cluster and the total output generated by all the industries in the economy. The average total output share per cluster is 8.33%, with values ranging from 3.15% to 18.20%. The largest cluster in terms of output share is *Construction*, producing over 106.380.839 million COP, representing 18.20% of the total output. The second largest cluster is *Petrochemical* producing over 75.244.833 million (12.87 percent of the total output).

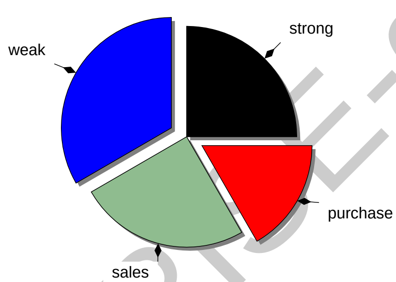
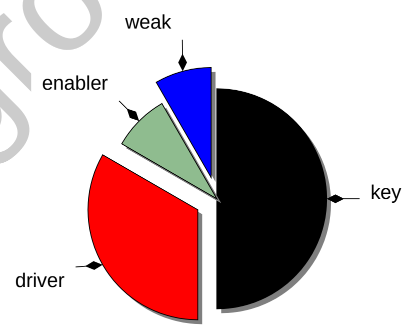
Table 5 shows the distribution of the 12 clusters based on internal and external linkages. The classification based on internal linkages (right figure) reflects the importance of the cluster as an internal sales and/or purchase market for the industries assigned to the same cluster (industry members). Internal linkages can also be understood as a measure of interactivity between the industries within the cluster.

In Colombia, the average purchase share ($\overline{PS}$) and the average sales share ($\overline{SS}$) are 0.28 and 0.33 respectively. This means that, on average, the clusters purchased 28% of their intermediate products from industries within the cluster, and 33% of the intermediate sales occur within the clusters. According to the results 33.33% of the clusters are important for industry members as both sales and purchase market (strong clusters); 25.00% are an important sales market (sales oriented clusters); 16.67% are an important purchase market for the industry members (purchased oriented clusters); and 33.33% of the clusters are classified as weak clusters.

The figure on the left (table 5) presents the distribution of the clusters according to the external linkages classification. External linkages seeks to measure how important each cluster is to the larger economy. On one hand, clusters with backward linkages greater than one ($BL_c > 1$), indicate that the cluster creates an above average increase in activity for the regional economy when the cluster experiences a marginal increase in its final demand. These clusters can be viewed as output drivers for the regional economy due to their reliance on locally produced inputs. On the other hand, A cluster with a forward linkage coefficient greater than one ($FL_c > 1$) shows an above average increase to marginal increase in other industries' final demand. This is indicative of the sector playing a strategic enabling role as a core supplier of inputs to other industries (Rey and Mattheis, 2000).

In Colombia, 50.00% of the clusters were classified as "key" clusters ($BL_c > 1$ and $FL_c > 1$). 8.33% as "enabler" clusters ($BL_c < 1$ and $FL_c > 1$); 33.33% as "driver" clusters ($BL_c > 1$ and $FL_c < 1$); and 8.33% as "weak" clusters ($BL_c < 1$ and $FL_c < 1$).

Table 5: Distribution of the clusters based on internal and external linkages

Internal linkages	External linkages
	

3.3 Composition of the clusters and impact analysis

This section provides an in depth analysis of each cluster. This analysis includes information about the roll of each cluster within the larger economy, as well as the internal composition and dynamics of each cluster.

Table 6 shows a graphical representation of each cluster. This representation is a useful tool to get an initial idea about the dynamics within each cluster. For example, through these graphics it is possible to recognize the role of each core industry (represented as squared shaped nodes). Thus, there are some cluster where the core industry is an important supplier for the other industry members (e.g. *Farming*, *Petrochemical*). In some other clusters, the core industry stands out as an important customer for the industry members (e.g. *Restaurants*, *Transport*, and *Commerce*). And finally, there are clusters where the core industries is a "connector" that purchases from some industry members and sales its intermediate products to other industry members (e.g. *Construction*, *Food*, and *Energy for industry*). According to Duque and Rey (2008) very few algorithms for industry cluster identification provide a clear way to understand the relationships between the industries assigned to the same cluster.

Table 7 describes the internal composition of each cluster and provides information to assess how important the cluster is for its industry members (Output and Input) and for the largest economy (Multiplier). This assessment is carry out with three measures:

- Output: Percentage of the industry output that goes to the other industries in the cluster. This value is a ratio between the industry output flowing within the cluster, and the total output of the industry. On average, 20% of the industries' output flows within the clusters. This value varies from cluster to cluster, and it strongly depends on the main activity of the cluster. Thus, there are clusters, like *farming* and *public utilities*, whose products and services are purchased by a wide range of industries, leading in low values of this index, 1.54% and 3.68% respectively. On the other hand, clusters like *food* and *banking* have an important portion of their sales flowing within the cluster, 31.7% and 30.84% respectively.

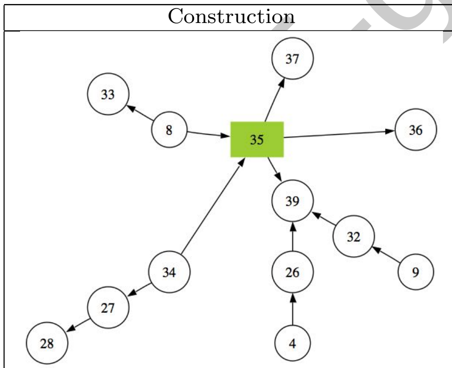
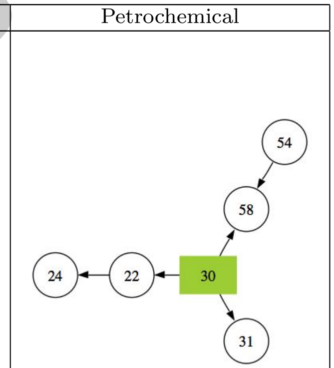
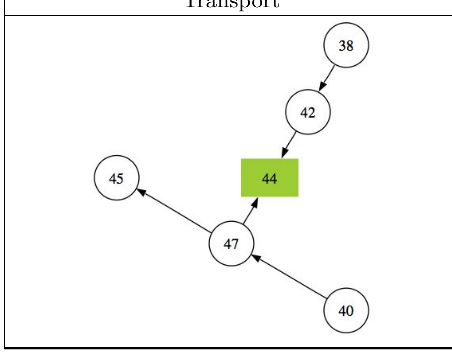
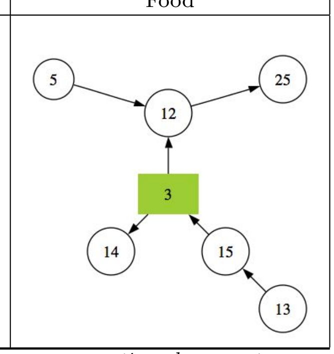
- Input: Percentage of the industry purchases that come from other industries in the cluster. This value is a ratio between the industry's input flowing within the cluster, and the total input of the industry. On average, 17.13% of the industries' purchases are supplied by other industries within the clusters. The clusters *farming* and *public utilities* present the smallest values of this index, 7.64% and 6.81% respectively; and *banking* and *food* have the highest values, 53.53% and 27.76% respectively.

- Multiplier: The multipliers estimated in this study are known in the literature as simple output multipliers, which seek to measure the change in the gross output of the local economy when there is a COP's worth change of final demand for a given industry, cluster or local economy (Leontief, 1953).

An output multiplier for a given industry i is defined as the total value of production in all the industries of the local economy that is necessary in order to satisfy an additional dollar's worth of final demand for industry i's output. The average output multiplier for Colombia is 1.55. Thus, a change of one COP's worth of final demand for Colombia will generate a change of 1.55 COP in the gross output of Colombia. At industry level *construction* is the industry with the highest output multiplier (3.61) and *banking* is the cluster with the highest output multiplier (1.84).

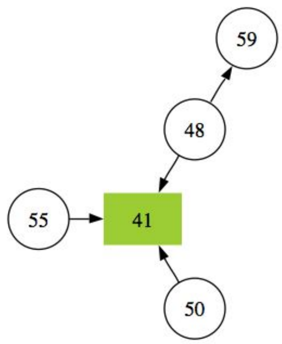
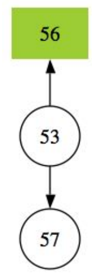
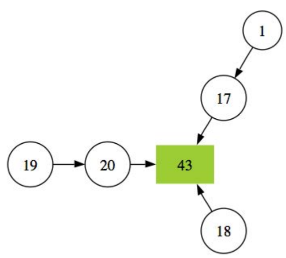
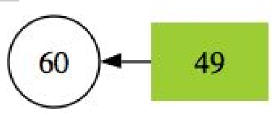
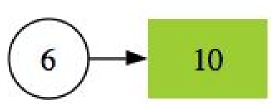
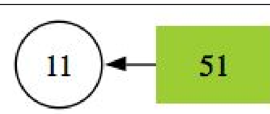
Two important results are derived from the figures in the table 7: i) seven out of twelve industries selected as "core industries", using the NBIC algorithm, have output multipliers ranked in the top ten; ii) only two clusters contain more than one industry ranked in the top ten highest output multipliers. This result may suggest an easier way to define core industries within the NBIC algorithm.

Table 6: Graphical representation of each cluster

Construction	Petrochemical
	
Transport	**Food**
	

continued on next page

Table 6: *continued*

Commerce	Educational Services
	
Restaurants	Energy for Industry
	
Banking	Farming
	
Energy for residential and commercial use	Public Utilities
	

3.3.1 Construction

Construction is the biggest cluster in Colombia with 13 industries that account for a 18.20% of the gross output of Colombia. This cluster includes four complementary activities (branches) related to the construction industry: industries 9 and 32 representing the branch of glass; industries 26 and 4 representing the branch of wood; industries 8, 33 and 35 representing the branch of iron and steel; industries 27 and 28 representing the branch of recycled products.

According to the backward and forward linkages, this cluster is classified as a "key" cluster, meaning that it plays an important roll as both driver and enabler of the economy. It also represents a wide market of resources and clients: 45% of the purchases of the cluster come from its industry members, and 37% of the sales of the cluster are directed toward industries in the cluster. At industry level, four industry members sale more than 50% of their output to industries within the cluster. It is also important to note the key roll played by the industry *Construction* which purchases 44.47% of its input from other industry members, and reports the highest output multiplier in the Colombian economy (3.61).

3.3.2 Petrochemical

This is a medium sized cluster with 6 industries belonging to it. However, this cluster contributes a significant share of the Colombia's gross output, 12.87%, ranking as the second highest cluster in output contribution.

The core industry *Basic and elaborated chemical products* plays an important role in the cluster. This industry supply chemical products for three different usages: i) textil and clothing, ii) plastic and rubber products, and iii) health services related products. From the arrange of the industries within the cluster and their relationships, it is easy to predict that this is a purchase oriented cluster. In fact, 41% of the purchases of the industry members remain within the cluster.

3.3.3 Transport

Six industries compose the third biggest cluster in term of output share, representing a 10.58% of colombian's gross output. This cluster covers road and fluvial transportation and other related services such as transport equipment, road transport and repair services, and civil engineering works.

The transport cluster is classified as a driver cluster of the economy. With backward linkages greater that one (1.23), this cluster creates an above average increase in activity for the regional economy when the cluster experiences a unit increase in its final demand. It is also an important sales market for its industry members. In fact, industries such as *Complementary* *road transport services*, *Repair services of engines and domestic stuff*, and *Transport equipment* have an important portion of their sales market in the transport cluster.

The transport cluster is also one of the two clusters that contain two industries with output multipliers ranked within the top ten highest multipliers, they are *Road transport services* (2.65), and *Civil engineering works* (1.97).

3.3.4 Food

The seven industries belonging to this cluster represent the production of food from animal and its derivatives. This cluster produces the 9.59% of the total output of the Colombian economy and it is considered to be a key cluster with forward and backward linkages greater than one. The industries within the cluster are strongly related to each other with purchase and sale share above the average, 0.52 and 0.65 respectively. The cluster presents the second highest multiplier in the economy with 1.69.

This cluster is the main market for the core industry *Animals and products derived from animals*, representing the 80% of its intermediate sales. Moreover, the cluster is the main supplier of the industry *Fish and Meat* providing a 62.61% of its inputs, mainly coming from *Fish and other products from fishing* and *Animals and products derived from animals. Fish and Meat*, has also the highest industry multiplier (2.72) of this cluster.

3.3.5 Commerce

Commerce is a medium sized cluster in term of industries. The five industries of this cluster represent a 8,88% of Colombia's gross output. This cluster is classified as a driver cluster in terms of its external linkages, meaning that its reliance on locally produced inputs creates an above average increase in activity for the regional economy when the cluster experiences a unit increase in its final demand.

Although the cluster has not a particularly high output multiplier, its core industry *Commerce* has the second highest output multiplier in Colombia (3.10).

This cluster is an important source of inputs for the industries *Commerce* and *Leisure services and other no market services*, which obtain 23.35% and 20.97% of their inputs from other industry members.

3.3.6 Educational Services

The educational services industry includes a variety of institutions that offer vocational, career or technical instruction, and other educational and training services to millions of students each year. Three industries compose this cluster. They produce 8.20% of the total output of the economy. The multiplier for the cluster is 1.58, slightly higher than the Colombian one, however the core industry, *Government administration services and other services for the community*, has a relatively high multiplier ranking among the top 5 highest multipliers.

Educational services is one of the four driver clusters in the economy. It has the highest backward linkage 1.31. This is mainly due to the fact that its core industry is highly associated to the government sector, and government expenses are a great engine for the economy.

3.3.7 Restaurants

Three branches converge to the core industry *Hotels and restaurants services*: i) coffee production, ii) cacao products, and iii) beverages and other food products. Its core industry has the sixth highest output multiplier (2.29).

This cluster accounts for 6.68% of Colombia's gross output, and it is classified as a driver cluster in terms of its external linkages, which indicates that the backward linkages reflect purchases of intermediate goods and services by this cluster that are necessary to meet the demand. Regarding the internal linkages, the cluster is classified as sales oriented, suggesting that the cluster is an important sales market for its industry members.

No Roasted Coffee is one of the main industries in the Colombian economy. in 2005, this industry exported 88.7% of its total production, having as main customers the USA, Germany and Japan.

3.3.8 Energy for Industries

The three industries included in this cluster represent the production of energy for powering industries and air transportation. It contributes with a 6,28% of Colombia's gross output. With forward and backward linkages greater that one, this cluster is classified as a key cluster for the Colombian economy.

The industries in this cluster purchase 34% of their intermediate products from other industries within the cluster. This figure places the cluster as one of the two purchase oriented clusters in this analyisis.

When analysing the graphical representation in table 6 a natural relation can be observed. The industry *Oil, natural gas, uranium and thorium minerals* extracts raw oil and sells 26.33% of its intermediate products to the other industry members. Next, the industry *Refined oil products, nuclear combustibles and coke furnace products* provides 15.52% of the inputs required by *Aerial transport services*.

3.3.9 Banking

Although the banking cluster does not have important output share in the economy, 5.99% of gross output, it is classified as a key/strong cluster in terms of its external/internal linkages. The two industries belonging to this cluster are responsible for moving the monetary resources through all the economic sectors. The relevance of this cluster in the economy is also reflected in its output multiplier, 1.84, the highest multiplier at cluster level.

3.3.10 Farming

Farming is composed of three industries of the agricultural sector: *Other farming products*, *Sugar* and *Tobacco Products*. This cluster generates 5.06% of the gross output of the Colombian economy, and it is classified as a weak cluster in terms of both, internal and external linkages.

The figures in table 7 suggest that this cluster is not an important market or source of inputs for its industry members. This may be consequence of the export orientation of industries like *Sugar* and *Tobacco*.

3.3.11 Energy for residential and commercial use

The cluster has two industry members, *Lignite and Peat* and *Electricity and city gas*. This small cluster represents the production of energy for the households. Although the contribution of this cluster to the Colombian gross output is not outstanding (4.51%), this cluster is classified as a key cluster in terms of its external linkages.

According to the internal linkages, it is classified as a weak cluster. This is mainly due to the fact that the *Lignite and Peat* industry exports almost the totality of its production to Europe and the USA.

3.3.12 Public Utilities

This cluster is compose by two service industries. Their contribution to the economy's gross output is 3.15%, the lowest among the clusters.

As expected, this cluster is an enabler cluster in the economy (the only enabler cluster). A cluster with forward linkages coefficient greater than one is considered to have an above average sensitivity to unit changes in all sectors' final demands. This is indicative of the sector playing a strategic enabling role as a core supplier of inputs to other industries. Given the enabler role of this cluster, it is not surprising that its internal linkages are not very important for its industry members.

Table 7: Industry clusters in the Colombian economy

Code	Sector	Output	Input	Multiplier
	Construction			**1.54**
35	Common metals and metallic products, except machines and equipment	53.19	23.64	2.04*
9	Other no metallic minerals	68.85	1.73	1.03
34	Waste products	46.23	0.00	1.00
8	Metallic minerals	3.75	4.93	1.06
27	Cardboard and paper	26.41	22.51	1.73
36	Special and general use Machines	7.48	9.86	1.23
4	Forestry products and wood extraction	45.48	1.49	1.03
26	Wood products, cork, straw and plait materials	63.42	33.38	1.41
32	Glass, glass products and other no metallic products	50.28	21.40	1.76
39	Construction	3.29	44.47	3.61*
37	Other machines and electric supply	7.84	5.71	1.22
28	Printing and similar goods	9.80	22.85	1.30
33	Furniture and other transportable goods	2.16	13.45	1.58
	Petrochemical			**1.56**
30	Basic and elaborated chemical products (except plastic and rubber products)	30.72	20.10	2.02*
31	Plastic and rubber products	12.44	21.76	1.36
22	Threads and textile fibre weaved	41.11	18.72	1.27
24	Clothes	0.65	22.03	1.65
54	Social service and market health services	61.98	8.60	1.18
58	Social service and no market health services	0.00	40.53	1.89
	Transport			**1.62**
44	Road transport services	1.75	23.07	2.65*
38	Transport equipment	25.26	10.46	1.53
42	Repair services of engines and domestic stuff	29.20	32.44	1.29
40	Civil engineering works	4.96	1.46	1.97*
47	Complementary road transport services	41.04	14.52	1.20
45	Fluvial transport services	0.60	43.03	1.10
	Food			**1.69**
3	Animals and products derived from animals	80.01	27.61	1.94
13	Animal and vegetable oil, fur and cakes	34.20	18.24	1.40

continued on next page

Table 7: *continued*

Code	Sector	Output	Input	Multiplier
15	Mill products, starch and its products	30.22	17.56	1.96
5	Fish and other products from fishing	51.80	3.71	1.03
12	Fish and Meat	6.15	62.61	2.72*
14	Dairy products	3.14	31.66	1.46
25	Leather and shoes	16.71	32.95	1.33
	Commerce			**1.59**
41	Commerce	0.00	23.35	3.10*
50	Real-state services and house renting	2.96	0.27	1.11
48	Post office services and telecommunications	18.98	1.66	1.50
59	Leisure services and other no market services	0.00	20.97	1.09
55	Leisure services and other market services	14.50	4.34	1.16
	Educational Services			**1.58**
56	Government administration services and other services for the community	0.00	0.65	2.51*
53	No market education services	3.90	0.02	1.09
57	Market education services	0.00	0.89	1.15
	Restaurants			**1.43**
43	Hotel and restaurant services	0.49	15.94	2.29*
19	Other food products	10.16	5.59	1.36
18	Cacao, chocolate and sugar products	4.79	1.79	1.19
20	Beverages	15.61	4.38	1.50
1	No roasted Coffee	12.52	0.35	1.10
17	Processed Coffee	20.42	38.29	1.14
	Energy for Industries			**1.61**
29	Refined oil products, nuclear combustibles and coke furnace products	7.41	23.38	1.76
7	Oil, natural gas, uranium and thorium minerals	26.33	1.90	1.46
46	Aerial transport services	1.88	15.52	1.38
	Banking			**1.84**
49	Financial intermediation services, computer and related services	61.69	7.06	1.74
60	Financial intermediation services indirectly measured	0.00	100.00	1.95
	Farming			**1.19**
2	Other farming products	3.71	1.19	1.43

continued on next page

Table 7: *continued*

Code	Sector	Output	Input	Multiplier
16	Sugar	0.50	19.19	1.13
21	Tobacco products	0.43	2.54	1.03
	Energy for residential and commercial use			**1.62**
10	Electricity and city gas	32.12	32.25	2.06*
6	Lignite and Peat	2.32	1.91	1.17
	Public Utilities			**1.19**
51	Services to enterprises, except financial and real-state services	5.34	2.97	1.26
11	Water and sewerage services	2.02	10.64	1.12

** Belongs to the top ten industries with higher output multiplier*

4 Conclusions

This paper presented an empirical application of a recent methodology proposed by Duque and Rey (2008) to identify industry clusters based on network analysis. Such application utilizes the projected 2005 Colombia's product utilization matrix.

This algorithm shows up as a novel methodology which conciliates Porter's approach, who emphasizes the importance of creating industry clusters to enhance countries development, and Hausmann's research which offers guidance to policy makers in the identification and support of key industries. On one hand, the NBIC algorithm sorts the industries by their level of importance for the economy identifying the core industries; on the other hand, it sets the most representative vertical clusters in a given economy. Thus, the methodology allows policy makers design specific policy initiatives for both clusters and key industries.

The NBIC algorithm identified twelve industry clusters in Colombia's economy. According to the cluster's internal linkages, 66.67% of the cluster were classified as either purchase oriented, sales oriented or strong clusters. In addition, based on the external linkages analysis, 91.67% of the cluster were classified as either key, enabler or driver clusters.

Among the outstanding clusters are: the clusters of *Construction* (the largest cluster in terms of industry members), *Petrochemical* and *Transport* represent together a 41.65% of 2005 Colombia's total interindustry output; The cluster of *Banking* and *Food* are the largest clusters in terms of output multiplier, with 1.84 and 1.69 respectively.

The possibility of visualize the way the industries are related to each other within the cluster is one of the main strengths of the NBIC algorithm. This visualization tool has been very appreciated by policy makers, since it offers useful information about the structure of the supply chain; it allows to easily identify the roll that each industry plays in the cluster; and speeds up the interpretation process as well.

An interesting finding is the strong relationship between being a core industry and having a high simple output multiplier. This will be matter of further research since it can be useful for simplifying the process of core industries identification, which is the most complex step in the NBIC algorithm.

Among the weaknesses of this algorithm stands out. First, its sensitivity to the Modifiable Areal Unit Problem (Openshaw, 1984), since it only takes into account interindustry transactions within a predefined geographical area, excluding those flows that cross its boundaries. This situation may become an important issue when the analysis is carried out at small geographical scale. Second, the NBIC does not account for those flows toward final demand (consumption, investment, government and exports), which in some industries represent an important share of their outputs.

The possibility of including information from more than one period; relaxing the assumption of exclusivity, where an industry is forced to belong to one and only one cluster; and the inclusion of flows toward final demand, seem to be very fruitful areas for future research.

References

Bergman, E. and Feser, E. (1999). Industrial and regional clusters: Concepts and comparative applications. The Web Book of Regional Science.

Connectory (2008). Connectory home page. http://www.connectory.com (accessed Dicember, 2008).

Czamanski, S. (1971). Some empirical evidence of the strengths of linkages between groups of related industries in urban-regional complexes. *Papers, Regional Science Association*, 27:137–150.

Czamanski, S. and Ablas, L. (1979). Identification of industrial clusters and complexes: a comparison of methods and findings. *Urban Studies*, 16:61–80.

Doeringer, P. and Terkla, D. (1995). Business strategy and cross-industry clusters. *Economic Development Quarterly*, 9:225237.

Duque, J. and Rey, S. (2008). *The economics of regional clusters: Networks, technology and policy*, chapter A Network Based Approach Towards Industry Clustering. Edwar Elgar Publishing.

Europe INNOVA (2007). Europe INNOVA. Annual report 2007. Technical report, European Commission, http://www.europe-innova.org/servlet/Doc?cid=10392&lg=EN (accessed October, 2008).

Hausmann, R. and Klinger, B. (2007). Achieving export-led growth in colombia. Technical report, Quantum Advisory Group.

Hewings, G. J. and Jensen, R. (1986). Regional, interregional and multiregional input-output analysis. In Nijkamp, P., editor, *Handbook of Regional and Urban Economics: Volume 1 Regional Economics*, pages 295–355. North-Holland, Amsterdam.

High level advisory group on clusters (2008). The European cluster memorandum. Promoting European innovation through Clusters: An agenda for policy actions. Technical report, Center for strategy and competitiveness, Stockholm School of Economics., http://www.europe-innova.org/servlet/Doc?cid=8857&lg=EN (accessed October, 2008).

Isard, W., Azis, I. J., Drennan, M. P., Miller, R. E., Saltzman, S., and Thorbecke, E. (1998). *Methods of Interregional and Regional Analysis*. Ashgate, Brookfield.

Kaufman, A., Gittlell, R., Merenda, M., and Wood, W. N. C. (1994). Porter's model for geographic competitive advantage: The case of new hampshire. *Economic Development Quarterly*, 8:4366.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50.

Lahr, M. L. (1993). A review of the literature supporting the hybrid approach to constructing regional input-output models. *Economic Systems Research*, 5:277–293.

Leontief, W. W. (1953). *Studies in the Structure of the American Economy*, chapter Interregional theory. Harvard University Press, Cambridge, MA.

Marshall, A. (1890). *Principles of Economics*. Prometheus Books.

Miller, R. and Blair, P. (1985). *Input-Output Analysis: Foundations and Extensions*. Prentice-Hall, englewood cliffs edition.

Monitor Company, Inc (1993). Creating the competitive advantage of colombia. Technical report, Ministry of Development, Instituto de Fomento Industrial, Confecámaras and Bancoldex.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Porter, M. (1990). *The Competitive Advantage of Nations*. Free Press.

Porter, M. (2003). *Ser competitivo: nuevas aportaciones y conclusiones*. Ediciones Deusto.

Regal (2008). Regal home page. http://regionalanalysislab.org/?n=Connectory (accessed Dicember, 2008).

Rey, S. J., Duque, J., Schmidt, C., and Li, X. (2007). Industrial clusters in San Diego County. Technical report, Regional Analysis Laboratory (San Diego State University).

Rey, S. J. and Mattheis, D. J. (2000). Identifying regional industrial clusters in California: Volume I conceptual design. Technical report, Californian Employment Development Department, Sacramento.

Roepke, H., Adams, D., and Wiseman, R. (1974). A new approach to the identification of industrial complexes using input-output data. *Journal of Regional Science*, 14:15–29.

Sandag (2008). Sandag home page. http://www.sandag.org (accessed Dicember, 2008).

Sistema Nacional de Competitividad (2008). Política nacional de competitividad y productividad. Technical Report 3527, Departamento Nacional de Planeación.

Steiner, M. (1998). *Clusters and regional specialization: on geography, technology, and networks.* Pion, London.