MSc Local Economic Development
Dissertation Proposal
To be submitted to the allocated LED Dissertation Tutor

Name
DYEVRE Arnaud

MSc Programme
MSc Local economic Development

Provisional Title of Dissertation
Regional specialisation patterns: evidence from the United States

Justification of research (Why is your topic interesting?)
My dissertation can contribute to our understanding of regional diversification for 4 main reasons:

1) With $1,575 trillion worth of exports in 2013 (est. CIA World Factbook), the United States are the second largest exporter after China. It is legitimate to ask what causes such omnipresence on world markets. I suggest here to explore the mechanisms that led America to reach high level of competitiveness, using fine export data and newly developed methods to analyse product relatedness and regional branching. My dissertation could contribute to understand the nexus that links global competitive pressures and regional development. The data, available from 2002 to 2012 will be fascinating to make sense of, as 2 important dynamics that could be of interest, are at work: the longstanding rise of China as the world's largest exporter (the country overtook USA in 2009) and the more punctual shock of the 2008 crisis.
2) While the case-study literature on product relatedness and related variety has been extremely fecund in the last decade, the American case has not been studied. The patterns of specialisation of the first economy in the world are of prime interest, and I suggest investigating how the USA developed comparative advantages at the local level. So far, only one study compared the importance of regional and national specialization using international trade data (Boschma, Minando and Navarro, 2012)
3) Former studies of the export specialization used rather coarse measures of exports, usually disaggregated at the 2-, 3- or 4-digit level. Here, I suggest to use export data disaggregated at the 6-digit level (HS6) and thus get a finer picture of the process of acquisition of comparative advantage.
4) With the data provided by the American customs, I will be able to perform a multi-layered analysis at 3 distinct levels: country-, State- and city-level. Academic papers on product relatedness and regional branching focus at most on a combination of two levels or one level at a time. A 3-layered focus will link the economic analysis to geography and will allow me to draw conclusions and potentially policy implications on the most relevant scale for capabilities acquisition and diversification.

Research questions/hypotheses

My dissertation aims at understanding the mechanisms of regional comparative advantage acquisition in the USA. To this end, I will apply a product relatedness approach to American regional and metropolitan export data.

The growing body of literature on product relatedness/related variety tends to confirm that regional and national economies alike grow using a set of capabilities, which slowly changes over time. This qualitative change of industrial diversification and specialization is amenable to quantitative analysis via different methodologies; my preference will go for Hausmann and Hidalgo's (2007) product space approach.

Using international trade data at the national and regional level over a long enough period to allow an economy to develop new comparative advantages (10 years), it would be possible to shed light on the regional and local specialization process in the United States.

While the processes through which capabilities are reallocated to create new comparative advantages are a fascinating question, I will mostly remain agnostic about their nature. My goal here is to let export data speak, and confirm or infirm the diversification of production in the United States via the product relatedness approach. I shall simply mention possible mediums of regional diversification. These are well explained in specialized streams of the economic development literature: labour mobility, firm spinoffs, forward and backward linkages, agglomeration economies, technological spillovers or university-industry linkages.

Research question:

(The question you will address in your dissertation)

I would like to test whether American States and cities economies diversify their economic output close to their current production fundamentals.

Hypothesis:

(Early a priori attempt to explain a given phenomena or process including an underpinning generic explanation)

How do new path of regional growth unfold is one of the most intriguing question in economic geography. The understanding of economic growth has benefitted from the recent contribution of product diversification theories, for until recently, relatively low attention was paid to the dynamics of industry branching.

From an evolutionary perspective, new economic output does not linearly emerges out of more inputs; it is created by reutilisation and recombination of existing capabilities.

Schumpeter first, pointed out that what drives economic development is not simply a quantitative change but a qualitative reshuffling of an economy's productive fundamentals via a never-ending process of creative destruction (1939). Subsequently, the economic literature of the 1950s and 1960s built on this idea: development is about a structural transformation of a nation's productive structure via reallocation of resources from low productivity productions to more productive ones (Lewis, 1955; Rostow, 1959; Kuznets, 1966; Kaldor, 1967).

Recent case studies support the view that new industries do not spring independently of the productive base of an economy (Klepper 2007; Boschma and Wenting 2007; Buenstorf, Fritsch, and Medrano 2010; Bathelt, Munro, and Spigel 2011; Buerger and Cantner 2011; Tanner 2011). Technologies and capabilities are the result of an evolution of related competences and assets.

The recent advances in the theorisation of 'regional branching' (Boschma & Frenken, 2007) allow us to makes sense of this phenomenon. In particular, some American scholars presented an interesting measure of relatedness amongst products (Hidalgo, Klinger, Barabási, and Hausmann 2007; Hausmann and Klinger 2007; Hausmann and Hidalgo 2010).
While most of the literature in regional economics used derivatives of the entropy measure, researchers agree that they suffer from the limitations imposed by industry categorisation.
I suggest using Hidalgo & Haussmann's method of proximity identification to find a better categorization of industries and groups of industries. Boschma et al. (2012) showed that this method gives better results than older measures (for instance Porter's, 2003; or Frenken's, 2007). The revealed industry relatedness index developed by Neffke and Henning (2008) based on product co-occurrence probabilities at the plant level requires very detailed data. It is nevertheless conceptually similar to Hausmann and Hidalgo's.

Under Hausmann and Hidalgo's categorisation, industries are related if they share untradeable inputs or 'capabilities'. Capabilities can be as concrete as harbours, bridges and highways, or be more abstract and culturally-rooted like institutions, skills and the existence of particular networks, or a combination thereof.
This method allows us to represent industry relatedness in the form of a network of capabilities. In this network, industries, which share an analogous set of capabilities, are closely connected. Hausmann et al. (2011) use international trade data and considers a nation possesses a capability if it has a revealed comparative advantage in the concerned industry.

However, I think relying only on international trade data might not give us a full picture of how economies acquire, develop and lose capabilities at the regional level.

I would hence try to analyse the process of industrial branching at the subnational scale, as capabilities are mainly transferred at the regional and urban level (although not exclusively).

Data
Following Poncet and De Waldemar (2013), I will use an international trade dataset provided by the *Centre d'Etudes Propestctives et d'Information Internationales* (CEPII), a French research institute for international economics. The dataset, BACI, is a world trade database at a high level of product disaggregation (HS 6-digit). It is based on original data provided by COMTRADE and correct the difference in values reported by the exporter (reporting 'FOB': free on board) and the importer (reporting 'CIF': cost, insurance and freight). Moreover it assesses the reliability of country reporting and correct for unreliable trade flows (a full description of their methodology can be found in Gaulier and Zignago, 2010). The dataset contains export data for more than 200 countries and more than 5000 products over 17 years (1995-2012). However, I will use only the observations from 2002 to 2012 to match the American data.

I will use this dataset to build one world product space per year studied. To date, a world product space for 5000+ products has not been created; Hausmann et al. (2011) use the SITC4 classification with only 800 products for theirs.
The database is accessible via
http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=1.

The American data is provided by USA Trade Online, a portal powered by the United States Census Bureau. I have been able to access export data for the 50 American States and for the following 57 'ports' at the HS 6-digit level:

Anchorage	*El Paso*	*Nogales*	*Savannah*
Baltimore	*Great falls*	*Norfolk*	*Seattle*
Boston	*Honolulu*	*Ogdensburg*	*St. Albans*
Buffalo	*Houston-Gavelston*	*Pembia*	*St. Louis*
Charleston	*Laredo*	*Philadelphia*	*Tampa*
Chicago	*Los Angeles*	*Port Arthur*	*U.S. Virgin Islands*
Cleveland	*Miami*	*Portland*	*Washington DC*
Columbia-Snake	*Milwaukee*	*Providence*	*Willmington*
Dallas-Fort Worth	*Minneapolis*	*San Diego*	
Detroit	*New Orleans*	*San Francisco*	
Duluth	*New York City*	*San Juan*	

With the city-, State- and Country-level export data, I will be able to test if diversification occurs through reallocation of capabilities close to the current production fundamentals at 3 different levels. My aim is to quantify the importance of local capabilities and I expect them to play a greater role in the acquisition of comparative advantage at a more local level.
The data can be accessed at: https://usatrade.census.gov.

Null Hypothesis [complete where appropriate]:
*(Is a plausible scenario which may explain a given phenomenon. A null hypothesis is tested to determine whether your data provide sufficient reason to pursue some alternative hypothesis. It cannot be rejected unless the evidence against it is sufficiently strong. Rejecting the null hypothesis then, suggests that the alternative hypothesis may be true
E.g. 'R&D investment lead to economic growth in all regions' or 'globalization is flattening the world economic geography' etc.)*

I want to test the following hypothesis:
H_0: 'American regional economies do not specialise in activities using the capabilities already used by existing industries'

Specific case studies in European countries and in China reject this hypothesis and thus back the regional branching theory. Regions seem to build up capabilities with what they already have.

To test this hypothesis, I plan to run 2 different regressions, each of them at the metropolitan and at the State level. The first will measure the influence of local product relatedness *vis-à-vis* the larger-scale density. The second will make this analysis slightly more complex by differentiating between persisting comparative advantages and newly created ones.

Methodology

(Describe the procedure or set of procedures that will be used in order to test your hypotheses and answer your research question)

I will proceed in 3 steps:
- Building a proximity index
- Based on the previously generated indices, represent the world and the American product spaces and find communities of capabilities
- Run 2 regressions to test the hypothesis and quantify the role of relatedness

Building the index

Many measures of relatedness exist, my choice to follow Hidalgo's method is motivated by:
- **Efficiency**: Hidalgo's method has proven to have more explanatory power in growth regressions than alternative measures based on entropy or the Herfindahl index
- **Feasibility**: This index is easier to compute with respect to the kind of data I have been able to get a hold on. Neffke, Boschma and Henning use their very own Revealed Relatedness method (2011, on Swedish regions) based on co-occurrence of products at the plant level. Obtaining this very detailed data at the firm level is daunting for this modest endeavour.
- Professor Iammarino's advices

Revealed Comparative advantages

Using trade data, one starts by identifying industries in which regions have a Revealed Comparative Advantage, using Belassa's formula (1965):

$$RCA_{c,i,t} = \frac{x_{c,i,t} / \sum_i x_{c,i,t}}{\sum_c x_{c,i,t} / \sum_c \sum_i x_{c,i,t}}$$

where $x_{c,i}$ is the total export value of good i exported by country c. We consider that a country c has a comparative advantage in good i at time t, if RCAc,i,t>1 i.e. if its share of export of product i is greater than the share of world exports for this product.

We compute RCA's:
- for countries (Including the United States) using the BACI database
- for US States, using the US Trade Online database for the numerator and the BACI's for the denominator
- for US cities in a similar fashion

Proximities

Then for each pair of goods (i,j) we calculate the probability for a country, a region or a city of having a revealed comparative advantage in i if it already has a RCA in j (we divide the number of regions exporting both i and j by the number of regions exporting j). Similarly, we get the conditional probability of exporting j with a comparative advantage when i is already exported with a comparative advantage. We then select the minimum of the two conditional probabilities in order to minimise false positive.

The proximity φ_{ijt} is formally defined as:

$$\varphi_{i,j,t} = \{Pr(RCA_{c,i,t} | RCA_{c,j,t}); Pr(RCA_{c,j,t} | RCA_{c,i,t})\}$$

φ_{ijt} is our measure of proximity, the greater, the more similar i and j are in their capabilities requirements.

Density
From the proximity measure φ_{ijt}, we can build a density index:

$$d_{i,c,t} = \frac{\sum_k \varphi_{i,k,t} x_{k,c,t}}{\sum_k \varphi_{i,k,t}}$$

where i is a product exported by country c. x_{kct} = 1 if country c has a comparative advantage in producing k at time t. d_{ict} is thus the ratio of the summation of 'distances' (measured as proximity probabilities) between product i and products in which the region has a comparative advantage, by the sum of the proximities to products in which country c has and has not a comparative advantage at t.
It gives us an idea of how densely embedded a product i is in the product space. If product i is well connected to products in which the region already has a comparative advantage, d_{ict} will be closer to 1 than if the product is isolated in the country export's basket.
The density measure captures the link between the intrinsic capabilities requirement of a specific production and the specialization pattern at the world or at the regional scale. It seems like the density of a good reflects its ability to benefit from export-enhancing spillovers from its 'neighbouring' goods. As stated by Hidalgo et al; (2007), Kali et al. (2013) and Poncet and Waldemar (2013), the density measure is a proxy for spillovers emanating from consistent specialization (like knowledge externalities, economies of scope and scale).
I will build these indices at the metropolitan, regional and country-level for each product.

Visualizing the product space
Based on the matrix of proximities between pairs of products, we can make the concept of product space more visual. I would like to represent the American product space as an undirected, weighted graph of good with softwares like Ucinet+Netdraw or Gephi. The greater the probability to co-export two goods amongst American states, the stronger the weight between these two products will be on the graph.

Ideally, I would get a result comparable to Neffke, Henning and Boschma's product space of Swedish regions

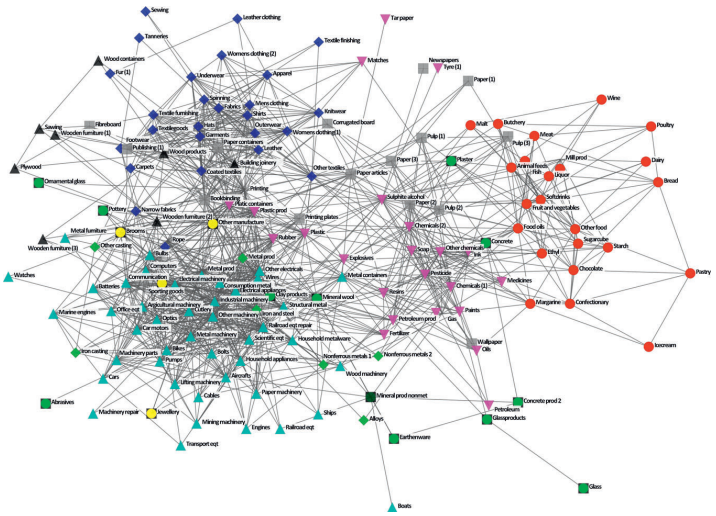
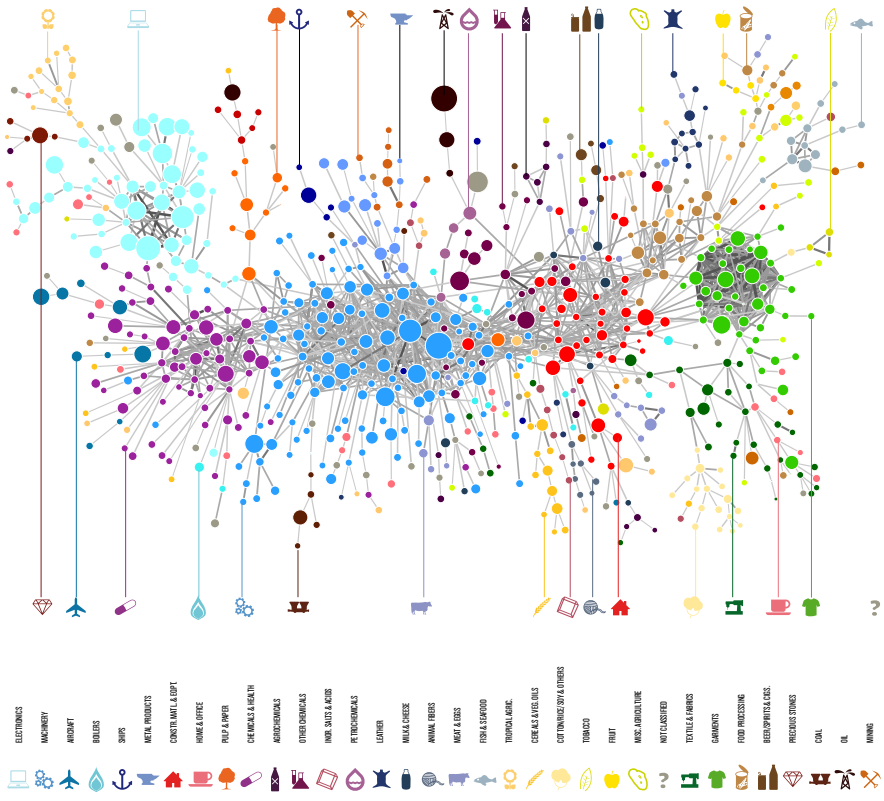


Figure 1. The average industry space in Swedish manufacturing, 1969–2002.

Or like Hausmann and Hidalgo's world product space:



On the graphs that I will generate, each node will represent one industry or product at a 6-digit classification in the dataset. Links represent connections between industries. For my dissertation, the proximities φ_{ijt} would weight the links.

Using functions of network analysis packages, we can display the nodes in a way such as the most related industries lie on the densest parts of the network on the two-dimensional plane. This operation can help getting a feeling of which industries share the same capabilities.

Following Hidalgo et al. (2011), we can use this networked representation of an economy output to find communities within the network. Communities in the product space are broad sets of industries requiring the same capabilities. This analysis can be performed with Ucinet and would allow us to find a more relevant classification of industries than the Standard Industrial Classification (SIC) based measures of relatedness.

To find communities in a product space network, the literature has proceeded using the Rosvall and Bergstrom (2008) algorithm: a collection of random walkers wandering on the network is generated, the process is repeated several times. Taking advantage of the fact that some walkers get 'stuck' in denser parts of the network –the communities– the algorithm is able to identify subgroups in the graph.
Alternatively, I can use methods of hierarchical clustering as proposed by Ruan and Zhang (2008) or the Girvan-Newman method based on deletion of edges with the highest betweenness centrality (2002). These algorithms can be implemented with UCINET and R. The choice of the method will depend on the shape of the network.

As a final step in exploratory data analysis, I think it would be interesting to identify the most densely embedded products and the most isolated products from a capability perspective. I would then try to make sense of the results and present them in a table. The general conclusion in the literature is that goods such as high-tech machines, tools and electronics are more densely embedded in a set of heterogeneous capabilities and thus rank higher on a 'complexity scale' than agricultural and extracting productions for instance.

First regression analysis: Do densities matter more at more localised levels?
Once 'communities' of industries are identified, I would like to use a logit model to test whether industries using a set of capabilities more densely embedded developed a comparative advantage in these industries, and if so, if it the specialisation process is more powerful at more localised geographies. In this respect, regression analysis could shed light on the magnitude of the regional export upgrading process.

I plan to slice the period of analysis in two 5-year intervals. According to Boschma et al. (2012), this interval length is long enough to let regions develop new industries and short enough to get a large number of intervals. I will then have 2 5-year intervals.

Density variables should preliminarily be normalised by subtracting the mean and dividing by the standard deviation to ease comparability.

To test my hypothesis, I will run a logit regression. The dependent variable, $x_{i,s,t+5}$, will take the value 1 if the State s has developed a comparative advantage in good i, at t+5. Independent variables will be observed at time t. For my purpose, one can define $\pi_{i,s,t+5} = Pr(x_{i,s,t+5}=1)$. I will thus use the following model to compare and quantify the impact of product density at the State and national level on the development of comparative advantage at the State level:

$$\log\left(\frac{\pi_{i,s,t+5}}{1-\pi_{i,s,t+5}}\right) = \beta_1 + \beta_2 x_{i,s,t} + \beta_c d^c_{i,s,t} + \beta_s d^s_{i,s,t} + \beta_{i,t} + \beta_{s,t} + \varepsilon_{i,s,t}$$

$x_{i,s,t+5}$ takes the value of 1 if State s has a comparative advantage in product i at time t+5, 0 otherwise.
$x_{i,s,t}$ takes the value of 1 if State s has a comparative advantage in product i at time t, 0 otherwise. Its inclusion in the model should reveal the persistence of a comparative advantage. I expect its coefficient to takes on a positive value *i.e.* States tend to keep a comparative advantage over a 5-year period.
$d^c_{i,s,t}$ is the density of product i at the country level at time t
$d^s_{i,s,t}$ is the density of product i at the State level at time t
$\beta_{i,t}$ and $\beta_{r,t}$ are products and States fixed effects.

Subsequently, I will run the same regression to compare the impact of metropolitan product density and State product density on specialization.
It might be possible that the density of a product exhibits diminishing returns to scale, as shown in Kali et al. (2013). That is, it is easier for a regional economy to conquer unexplored spaces of the product space if the density increases from a small value to an average one than from the average to a large value. If so, I should include a quadratic density term in the regression to account for this effect.

It is worth noting that the literature seems to face an incidental parameter problem as a lot of dummy variables (for products and regions) are included, leading to biased and inconsistent estimates. It might then be judicious to turn to linear probability OLS estimation.
One solution suggested by Boschma, Minondo and Navarro is to use a system-GMM model. The quality of the previous regressions will indicate if there is a need to use a GMM model or not.

Second regression: to acquire or to keep comparative advantages?
Subject to the same cautiousness regarding model selection, we can perform another analysis following Klinger & Hausmann (2007) and Boschma et al. (2012). We will try to assess how good US States and cities are at *keeping* and *acquiring* comparative advantage in a product. The models will take the form (here in linear probability-OLS):

$$x_{i,s,t+5} = \beta_1 + \beta_2 x_{i,s,t} + \beta_c^n (1 - x_{i,s,t}) d_{i,s,t}^c + \beta_c^o (x_{i,s,t}) d_{i,s,t}^c + \beta_s^n (1 - x_{i,s,t}) d_{i,s,t}^s + \beta_s^o (x_{i,s,t}) d_{i,s,t}^s + \beta_{i,t} + \beta_{s,t} + \varepsilon_{i,s,t}$$

β^n_c here is the estimate of the coefficient of *developing* a new comparative advantage in a new product i, at the country level
β^o_c is the coefficient of *keeping* an old comparative advantage in a product i at the country level.
β^n_s and β^o_s are similarly defined at the State level.
$d^c_{i,s,t}$ and $d^s_{i,s,t}$ are the country and State density levels for product i, exported by State s at t.

For instance, if we take a product in which neither the country nor the State had a comparative advantage at t, but acquired it at t+5, the regression equation becomes:

$$x_{i,s,t+5} = \beta_1 + \beta_c^n d_{i,s,t}^c + \beta_s^n d_{i,s,t}^s + \beta_{i,t} + \beta_{s,t} + \varepsilon_{i,s,t}$$

In other terms, I will be able to assess the efficiency of the American regional and national economies to conquer unexplored potentials of the product space. We expect β^{o}_{c} and β^{o}_{s} to be greater than β^{n}_{c} and β^{n}_{s} respectively, *i.e.* the United States as a whole and American States are better at keeping comparative advantages than developing new ones. It would be interesting to quantify the measure of this economic inertia.

Similarly to what will be done for the previous regression, the analysis will also be carried out at the metropolitan level, compared to the State level.

Check of robustness
I can check the robustness of results in 3 different ways:

1. Using the Revealed Comparative Advantage directly instead of using a dichotomous variable (revealed or not revealed)
2. Using a lower threshold for comparative advantage (e.g. if $RCA_{i,t}>0.5$, $x_{i,t} = 1$)
3. And using alternative time intervals.

Sources and Key References

Contextualisation, helpful for literature review:
BOSCHMA, R., & FRENKEN, K. (2011). The emerging empirics of evolutionary economic geography. Journal of Economic Geography, 11(2), 295-307.

Case-studies
KLEPPER, S. 2007. Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry. Management Science 53:616–31
BOSCHMA, R. A., and WENTING, R. 2007. The spatial evolution of the British automobile industry: Does location matter? Industrial and Corporate Change 16:213–38
BUENSTORF, G.; FRITSCH, M.; and MEDRANO, L. F. 2010. Regional knowledge and the emergence of an industry: Laser systems production in West Germany, 1975–2005. Papers in Evolutionary Economic Geography, No. 10.16.
BATHELT, H.; MUNRO, A. K.; and SPIGEL, B. 2011. Challenges of transformation: Innovation, re-bundling and traditional manufacturing in Canada's technology triangle. Papers in Evolutionary Economic Geography, No. 11.11.,
BATHELT, H., & BOGGS, J. S. (2003). Toward a reconceptualization of regional development paths: is Leipzig's media cluster a continuation of or a rupture with the past?. Economic Geography, 79(3), 265-293.
BUERGER, M., and CANTNER, U. 2011. The regional dimension of sectorial innovativeness: An empirical investigation of two specialized supplier and two science-based industries. Papers in Regional Science 90:373–93.
TANNER, A. N. 2011. The place of new industries. The case of fuel cell technology and its technological relatedness to regional knowledge bases. Papers in Evolutionary Economic Geography, No. 11.13.

1950s and 1960s papers mentioned in literature review:
LEWIS, W. A. (1955). The Theory of Economic Growth, Homewood, IL: Richard D.
ROSTOW, W.W., 1959. The stages of economic growth. Economic History Review 12 (1), 1–16.
KUZNETS, S., 1966. Modern Economic Growth. Yale University Press, New Haven, CT.

KALDOR, N., 1967. Strategic Factors in Economic Development. New York State School of Industrial and Labor Relations, Cornell University, Ithaca, NY.

Measure of relatedness:

HIDALGO, C. A.; KLINGER, B.; BARABÁsi, A. L.; and HAUSMANN, R. 2007. The product space conditions the development of nations. Science 317

HAUSMANN, R., HIDALGO, C., BUSTOS, S., COSCIA, M., SIMOES, A. & YILDIRIM, M. (2011). The atlas of economic complexity. Boston. USA.

HAUSMANN, R., and KLINGER, B. 2007. The structure of the product space and the evolution of comparative advantage. Working Paper No. 146. Cambridge, Mass.: Center for International Development, Harvard University.

HAUSMANN, R., and HIDALGO, C. A. 2010. Country diversification, product ubiquity, and economic divergence. Working Paper No. 201. Cambridge, Mass.: Center for International Development, Harvard University

HIDALGO, C. A. 2009. The dynamics of economic complexity and the product space over a 42 year period. Working Paper No. 189. Cambridge, Mass.: Center for International Development, Harvard University.

KALI, R., REYES, J., MCGEE, J., & SHIRRELL, S. (2013). Growth networks. *Journal of Development Economics*, *101*, 216-227.

BALASSA, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage1. The Manchester School, 33(2), 99-123.

Data explanations:

GAULIER, G., & ZIGNAGO, S. (2009). Baci: International trade database at the product-level.

Finding communities in the product space:

ROSVALL, M., & BERGSTROM, C. T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105(4), 1118-1123.

GIRVAN, M., & NEWMAN, M. E. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821-7826.

RUAN, J., ZHANG, W., 2008. Identifying network communities with a high resolution. Physical Review E 77 (016104), 1–12.

I would like to produce a piece of research in the vein of:

BOSCHMA, R., & IAMMARINO, S. (2009). Related variety, trade linkages, and regional growth in Italy. Economic Geography, 85(3), 289-311.

MAMELI, F., IAMMARINO, S., & BOSCHMA, R. (2012). Regional variety and employment growth in Italian labour market areas: services versus manufacturing industries. Papers in Evolutionary Economic Geography, 12.

BOSCHMA, R., MINONDO, A., & Navarro, M. (2012). Related variety and regional growth in Spain*. Papers in Regional Science, 91(2), 241-256.

NEFFKE, F., HENNING, M., & BOSCHMA, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Economic Geography, 87(3), 237-265.

NEFFKE, F. M., HENNING, M., & BOSCHMA, R. (2012). The impact of aging and technological relatedness on agglomeration externalities: a survival analysis. Journal of Economic Geography, 12(2), 485-517.
FRENKEN, K.; VAN OORT, F. G.; and VERBURG, T. 2007. Related variety, unrelated variety and regional economic growth. Regional Studies 41:685–97.
BRACHERT, M.; KUBIS, A.; and TITZE, M. 2011. Related variety, unrelated variety and regional functions. Identifying sources of regional employment growth in Germany from 2003 to 2008. IWH Discussion Papers No. 15. Halle, Germany: Halle Institute for Economic Research.

General:
SCHUMPETER, J. A. 1939. Business cycles: A theoretical, historical and statistical analysis of t he capitalist process. New York: McGraw-Hill.

Data

CEPII-BACI: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=1
USA Trade Online: https://usatrade.census.gov

Sections or Chapter Headings

Abstract
Acknowledgements

I. Introduction
Interest of the research question
Literature review
Novelty brought by the dissertation

II. Data
II.1 description of the BACI-CEPII dataset
II.2 description of the USA Trade Online dataset

III. Methodology
III.1 Generating the indices
Computing RCAs
Building proximity measures
Presenting the matrix of proximities
Representing the product space
Finding communities
Computing the density indexes at different levels
III.2 regression analysis
Presenting the logit model to be used (1st model)
Discuss problems of misspecification, suggest way forward (GMM)
Presenting the linear probability model to be used (2nd model)

IV. Results
V. Conclusions

Bibliography
Appendix

Outline of Research

- Cleaning the data
- Generating the indices
- Generating the proximity matrix and subsequently the product space of the United States
- Testing my hypotheses
- Drafting the paper

Do you see yourself encountering any elevated risk during the undertaking of your dissertation? No

Comments from Tutors:

Conceptual basis/focus
Empirical focus
Research questions/hypothesis
Methodological basis
Data sources

Proposed structure of report