Innovation & Economic Development

- *Gang Zhao & Hassan Mohanna*

Introduction

One of the main challenges of urban science is how to model and measure the concepts we talk about, in terms of the choice of the selected metric and its suitability for measuring the concept. For example, these concepts:

- **City:** a conventional understanding of a city is that it is a place where people come together, live in proximity, and share activities. As such, it is impossible to define the boundaries of a city. In developed countries, metropolitan and regional areas span well beyond the boundaries of a city unless it is bound by restrictive geographies, such as rivers or mountains. In developing countries, drawing a land border for a city mandates determining how far the slums stretch, which then requires defining what a slum is, and raises the controversy of incorporating slum dwellers into urban population. This is one of the drivers of studying economic entities (Metropolitan Statistical Areas, or MSAs), rather than cities, which are political entities.
- **Human capital:** another name for "the amount of knowledge", this is a common phrase in political discourse and one of the first default answers when a city is looking for economic growth. But how is it measured or captured? We can assume a city's human capital is the aggregation or summation of all the human capital in this city. It is usually measured by "educational attainment", which is a reductionist view, but still correlated. This includes years of formal schooling, percent of labor force that achieves a degree of certain level or in certain domain (e.g. percent of labor force who have **graduate** degrees in **science/engineering**). One weakness for this criterion is that it does not account for quality of education.

- **Innovation/invention:** is it measured by patents, startups or another form? There is a school of thought in urban economics that is centered in the study of patents. The state of Iowa is one of the highest producers of bio-tech patents, but has very few bio-tech start-ups. Does this make it more or less innovative?
- **Economic Growth:** GDP data is available for urban areas in the US, starting in 2001. For other parts of the world, particularly developing countries, it will take a long time to measure GDP by statistics offices, and we can look for some other indicator: night-time light data has a good correlation. Its main weakness is exclusion of areas that lack electricity, traditionally slums, which brings us back to the aforementioned controversy of defining where a city ends.

Establishing correlations of either of these concepts and others with economic development is of great interest to urban economics. Muneepeerakul et al. (2013) use occupational data to illustrate how “the network of interdependencies among occupational specializations affects the ease with which urban economies can transform themselves”, which is the core of economic development.

On the other hand, Hidalgo et al. (2007) use an outcome-oriented approach focusing on product data to illustrate how economies move across the “product space”, a network of relatedness between products, and finds that economies achieve this by “developing goods close to those they currently produce”.

For our discussion, we will consider the linkage of patents to economic development. We selected this suggestion because patents are considered one form of recording knowledge creation, another concept which can also be viewed an output of human capital and innovation in a city or urban area.

Data Analysis – Part I: Patents and Economic Development

In this part, we aim to discuss the relationship between patenting performance and economic performance. Our unit of analysis is the MSA, where we analyze three datasets:

- Patenting Performance: MSA patents from the U.S. Patent & Trademark Office (USPTO),
- Economic Performance: MSA GDPs from the Bureau of Economic Analysis (BEA) at the U.S. Department of Commerce,
- MSA Population: this was not available from the Bureau of Economic Analysis (BEA) at the U.S. Department of Commerce, so we downloaded it from previous coursework (Prof Bettencourt).

We use the MSA Population dataset to normalize the data to a per-capita scale, which leads to two variables:

- **GDP Per Capita**: as a measure of economic performance, productivity and wealth.
- **Patent Intensity**: (patents per 100,000 residents) as a measure of "inventive productivity". This is calculated by using (MSA patents / MSA population) x 100,000, since the numbers tend to be small for some locations. It is important to note that patents are assigned to an MSA by the place of residence of an inventor; for patents with multiple inventors residing in different MSAs, each of the metropolitan areas represented in the authors' places of residence gets it patent count increased by one. It is important to note this link to the *Inventor* (e.g. faculty at a research facility) rather than the actual facility, referred to as the *Assignee*. We will come back to this point later.

Figure 1 (LN GDP per Capita vs LN Patent Intensity) plots the annual GDP data averaged over 2008 – 2012 period, against patent intensity annual data averaged over 2001 – 2005 period.

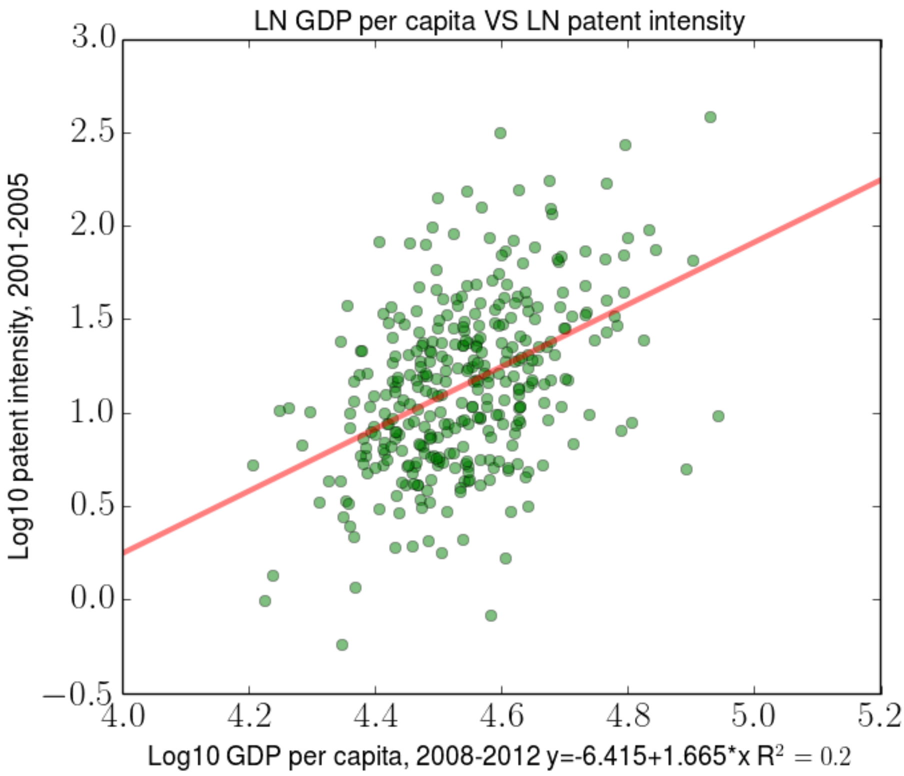

Figure 1. LN GDP per Capita vs LN Patent Intensity

The first observation is the low-quality fit between the two datasets. The *Coefficient of Determination* R^2 is equal to 1 in the case of a perfect fit, but is only 0.2 in this case. This demonstrates that there is weak statistical connection. A possible reason for this is that patent revenue and economic drive can be divided into three sources:

- Royalties: these go to the Inventor and/or Assignee. Sometimes they are the same person. However, a geographic scope bias is present, and can be attributed to the patent registration procedure of registering patents in the MSA of residence of inventors. We note that the work of any skilled professional is correlated to a certain extent with economic growth in their area of residence, but they might be working in an MSA other than the one they reside in, for example living in Connecticut and working in New York.

- Production revenues: this is subject to production costs and rarely takes place in the knowledge-centered U.S. cities where patents are registered, though it may happen that an industrialized suburb will produce the inventions patented in the same MSA.
- Contribution of the product/invention to improving the economy: this is not limited to any city or MSA since one of the main goals would be to sell wherever it is possible to obtain the desired product price.

Of additional significance are the timescale and lag for these datasets are significant. First, the adoption of a 5-year average allows buffering any exceptional incidents that could bias the data in a particular year. It can also be influenced by the timescale of research lifecycles, particularly for long-duration research programs. Second, the differing time windows accounts for the time it takes patents to impact the economy. Patents are treated as potential leading indicators because of this delayed impact, which occurs due to many reasons, such as time to market (TTM, or how long it takes a product from conception to market availability). Rothwell et al. (2013) note that inventions - the tangible descendants of patents - drive long-term economic performance, rather than making instantaneous impact.

As for the slope of the fitted line being 1.65, it is an indication of super-linear scaling, meaning that, in the best case scenario of a perfect statistical fit, an increase in patents would be accompanied by a less significant increase in GDP. While this can be attributed to the disparities in the quality of patents, meaning that only “useful” patents contribute to economic growth, it could also be due to a bias in patent data scale, in addition to the previously discussed bias in geographic scope.

The scale bias can be caused by many motivations and drivers for patenting that extend beyond the primary goal of patenting for direct profit or revenue. These include institutional goals (one of the criteria for success for the Cornell Technion NYC campus is how many patents come out of that in ten years), university/peer pressure (some engineering departments in top schools require candidates and faculty to have patents), and indirect market forces (start-up founders' having a patent portfolio is increasingly a criterion that investors consider in funding).

Data Analysis – Part II: Technological Profiles of Metropolitan Areas

In this part, we depict the technological profile of the New York MSA, and contrast/compare it with that of a few other MSAs. For this purpose, we selected, rather than three MSAs, the five metropolitan areas with the highest number of patents per capita, namely San Jose; Burlington, VT; Rochester, MN; Corvallis, OR; and Boulder, CO. (Rothwell et al., 2013).

We break down the count of patents in each MSA by technology class. Our unit of analysis is the MSA, where we analyze two datasets:

- Patenting counts by MSA and technology class: from USPTO,
- Patenting technology classes: from USPTO.

Figure 2 (NY MSA Top 10 Technology Classes) depicts the dominant patent technology classes in the New York MSA and hence its technology profile. Figures 3 to 7 depict the technology profiles of the MSAs selected for this comparison. Figure 8 (Comparison of Top 10 Technology Classes in Six MSAs) aggregates the top ten Technology Classes of the six MSAs and is the main basis for comparative analysis.

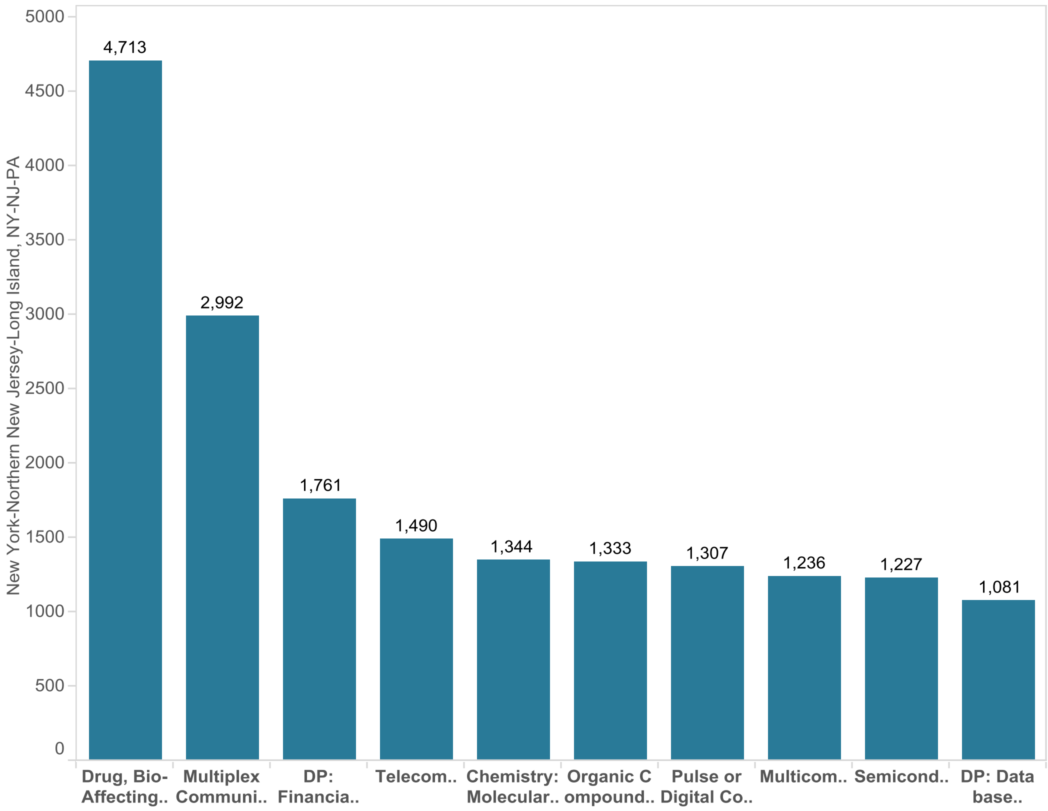

Figure 2. NY MSA Top 10 Technology Classes

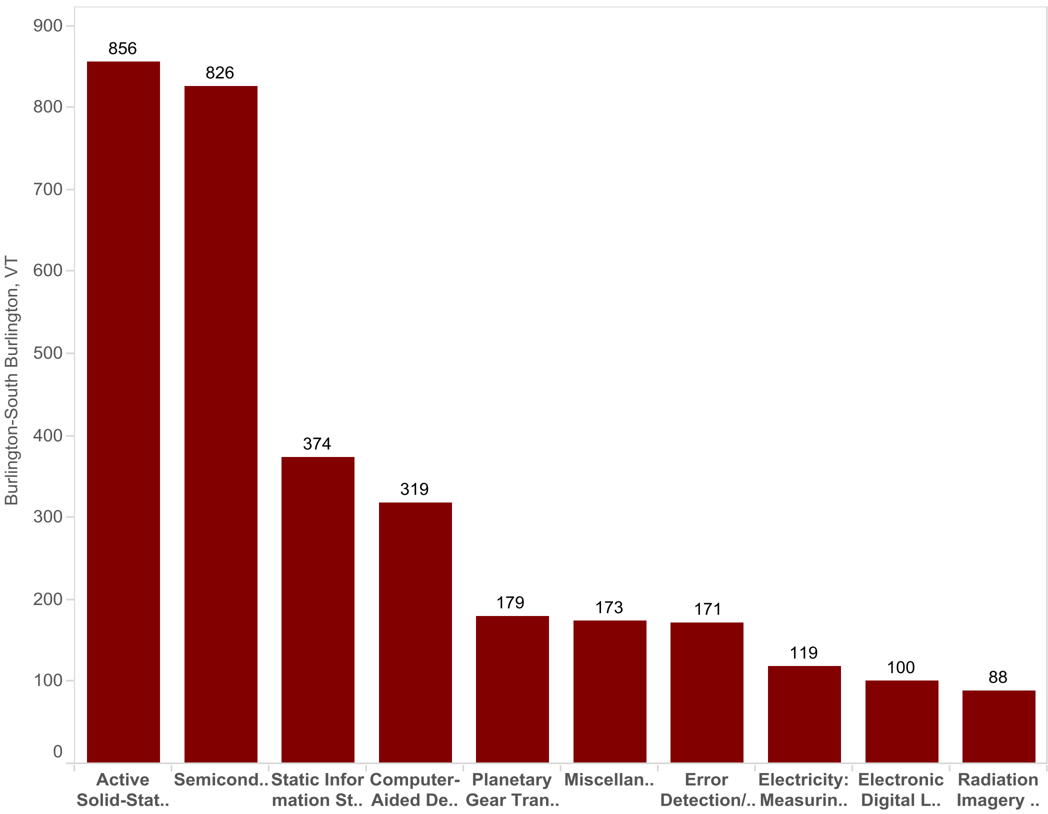

Figure 3. Burlington, VT MSA Top 10 Technology Classes

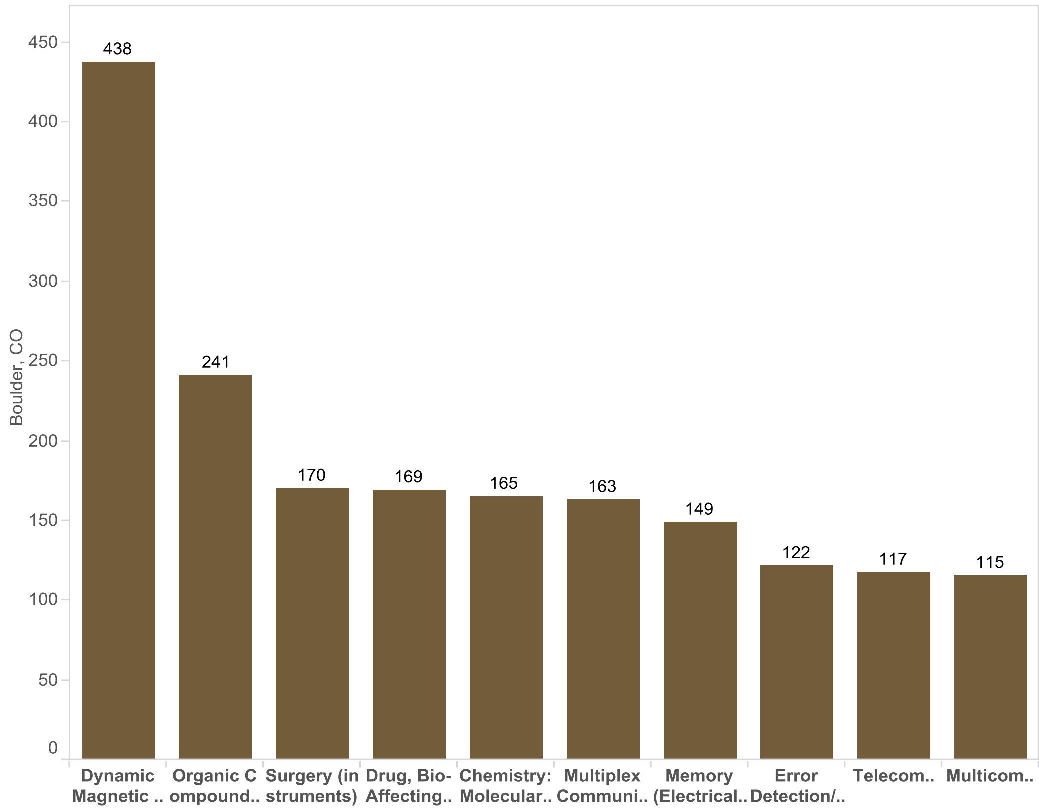

Figure 4. Boulder, CO MSA Top 10 Technology Classes

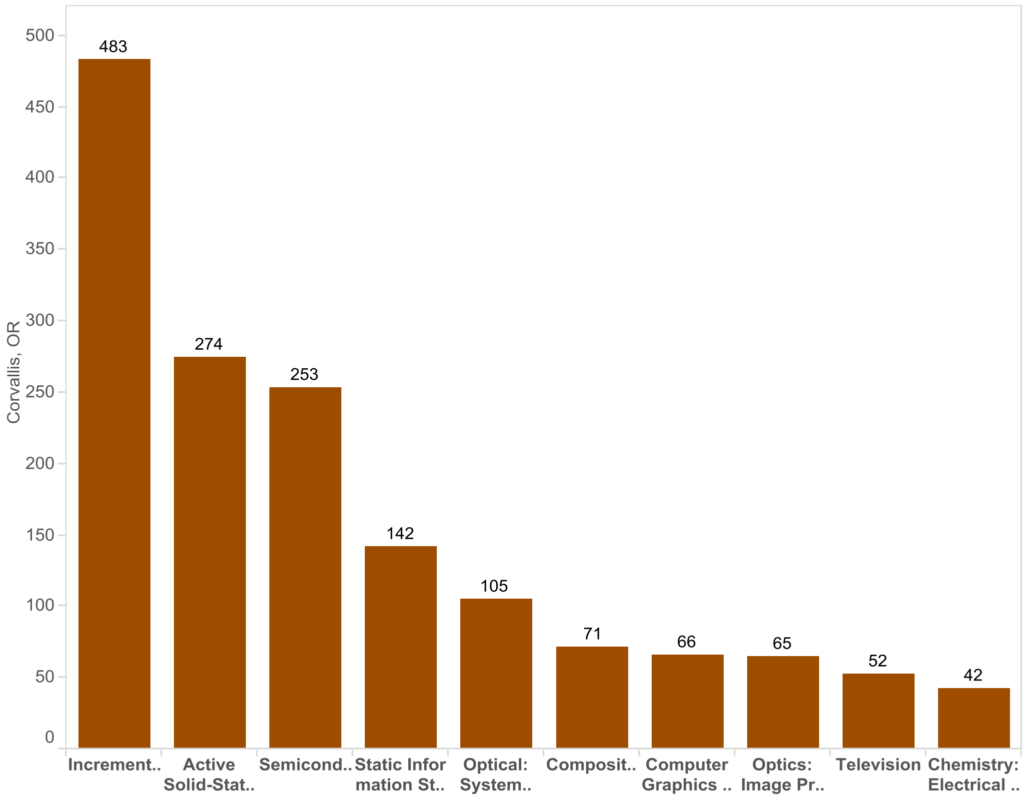

Figure 5. Corvallis, OR MSA Top 10 Technology Classes

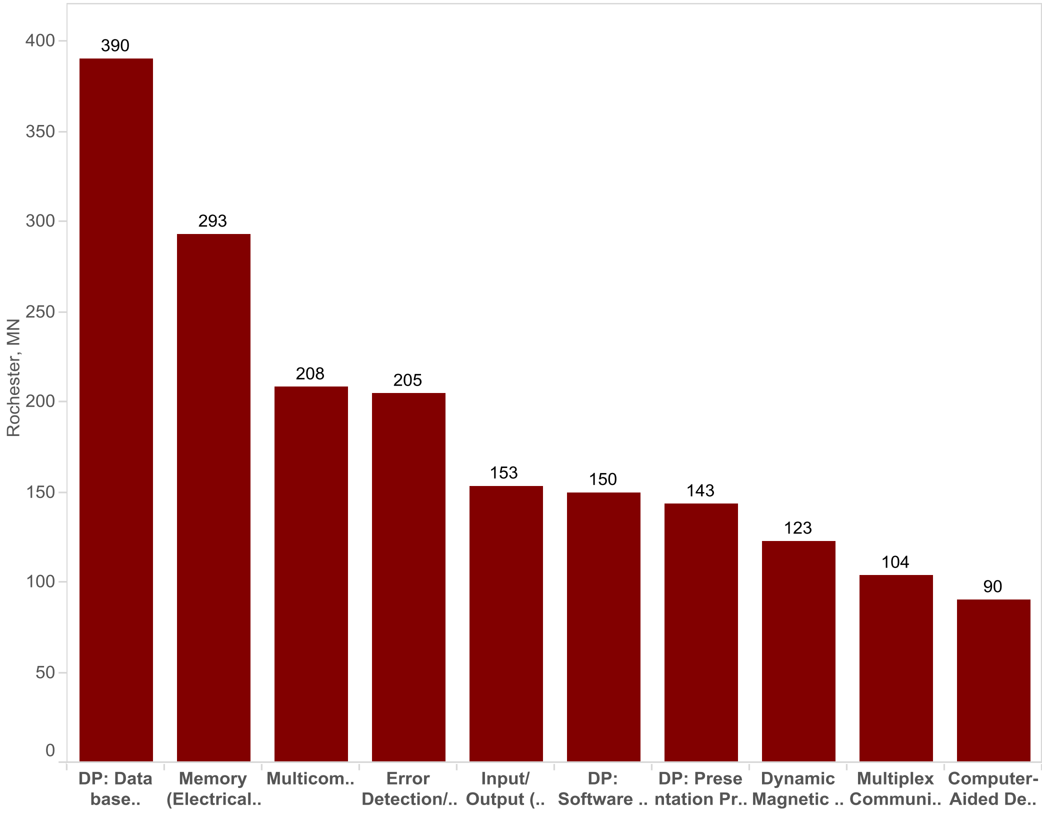

Figure 6. Rochester, MN MSA Top 10 Technology Classes

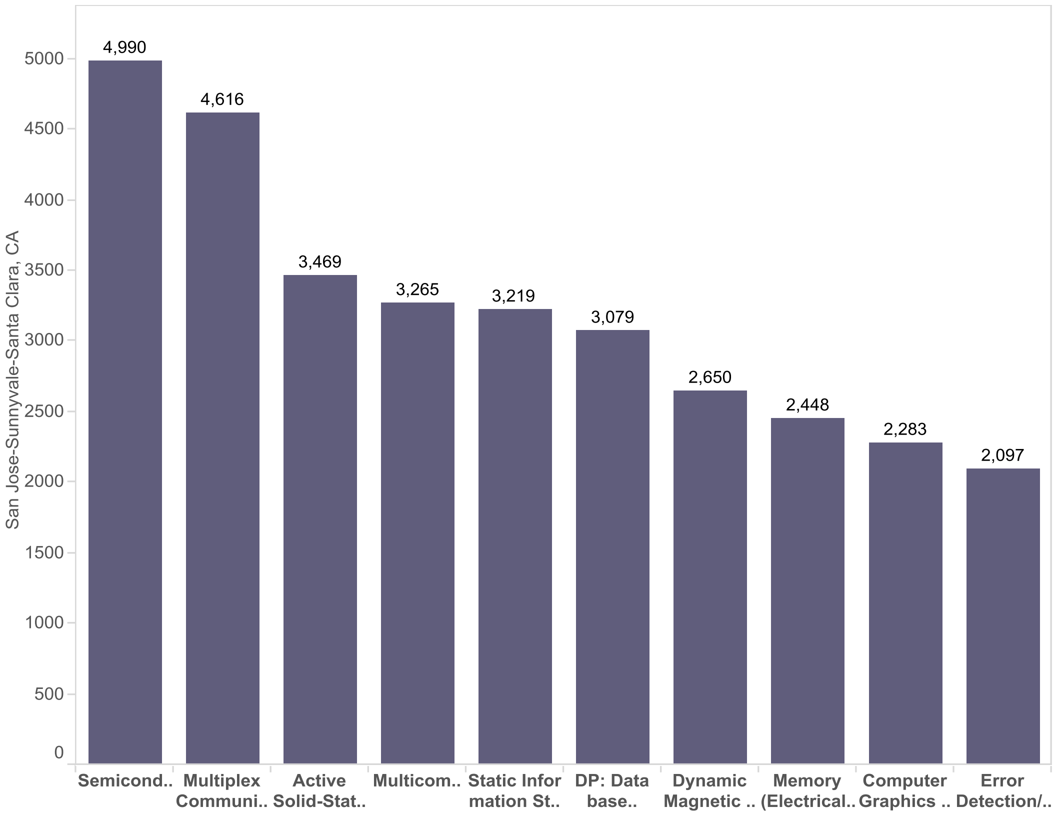

Figure 7. San Jose, CA MSA Top 10 Technology Classes

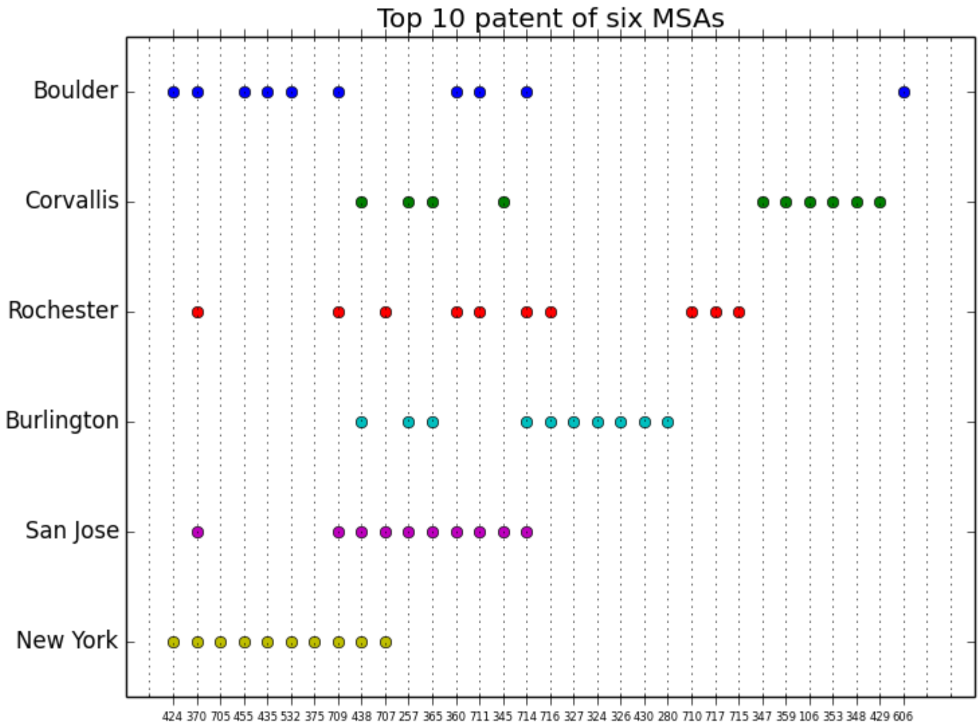

Figure 8. Comparison of Top 10 Technology Classes in Six MSAs

Comparing the six lists of top areas of inventive activities, we note that, of the 60 Technology Class records in the six MSA Technology Profiles, 32 are unique Technology Class records. (An attached/emailed XLS file will contain the source data and pivot tables if required). The remaining 32 are duplicates and cause the following commonalities to emerge:

- Class 438 (Semiconductor Device Manufacturing: Process) is common to **four** MSAs: **Burlington**, **Corvallis**, **New York**, and **San Jose**
- Class 370 (Multiplex Communications) & Class 709 (Multicomputer Data Transferring (Electrical Computers and Digital Processing Systems)) are common to **four** MSAs: **Boulder**, **New York**, **Rochester**, and **San Jose**
- Class 714 (Error Detection/Correction and Fault Detection/Recovery) is common to **four** MSAs: **Boulder**, **Burlington, Rochester**, and **San Jose**

- Class 257 (Active Solid-State Devices) & Class 365 (Static Information Storage and Retrieval) are common to **three** MSAs: **Burlington**, **Corvallis**, and **San Jose**
- Class 360 (Dynamic Magnetic Information Storage or Retrieval) & Class 711 (Memory (Electrical Computers and Digital Processing Systems)) are common to **three** MSAs: **Boulder**, **Rochester**, and **San Jose**
- Class 707 (DP: Database and File Management or Data Structures) is common to **three** MSAs: **New York**, **Rochester**, and **San Jose**
- Some other Classes are common to two MSAs…

This leads to the following conclusions:

- The top five MSAs in terms of patent intensity have many common patent trends, with ICT innovation accounting for most of their top-ranking Technology Classes.
- The duration 2001 – 2011 is simultaneous with the boom of ICT solutions and the rise of the Internet.
- These Technology Classes also have a high number of innovations related to computer hardware, not necessarily software and Internet-specific software solutions as one might imagine from media coverage.
- New York MSA, despite not being a significant player in terms of per capita number of patents, is aiming to replicate these models by tackling those domains where innovation (assuming we can measure it by patent generation) is rife. Judging from Part I, it is not a guaranteed route to more economic prosperity, unless accompanied by conscious contribution to the economy and assessing success by more relevant measures than patent counts.
- Other Technology Classes, not in the top ten of each of the six MSAs, will probably be common and can be explored for more comprehensive analysis

- In addition to linking each of the top five MSAs to ICT, we can point to possible reasons for the high concentration of patents. These can be further investigated in a more comprehensive analysis.
    - Corvallis, OR: in addition to Hewlett Packard, pharmaceuticals companies such as Sarepta Therapeutics are based there.
    - San Jose, CA: the largest city and “Capital of Silicon Valley”.
    - Burlington, VT: its high rank on the patent intensity scale is primarily due to its small population of around 42,000, coupled with its hosting the University of Vermont and General Electric, among others.
    - Boulder, CO: another example of university-centred (University of Colorado) or lab-centred metropolitan area (Rothwell et al., 2013)
    - Rochester, MN: it hosts Mayo Clinic and an IBM Campus, origin of IBM System i and home to the Blue Gene project.

A common effect of the technology concentration is the presence of a more specific set of skilled professionals, with the skills needed in such organizations. Rothwell et al. (2013) note that "Just a five percentage point increase in the share of workers with a STEM bachelor’s degree predicts an increase of 176 patents per million residents".

Data Analysis – Part III: Technological Heat Maps of Metropolitan Areas

In this part, we explore the diversity of the metropolitan patenting portfolios. This is achieved by developing a matrix of 481 technology classes by 367 MSAs. Figure 9 (MSA Technology Heat Map) is a binary representation of this distribution. Cells which contain patent count greater than zero, i.e. the MSA has generated patents of that respective Technology Class, are green, while cells with zero are yellow. The map is sorted from left to right in decreasing number of Technology Classes that an MSA has patents in. (The attached/emailed XLS file will contain the heat-map source data).

Figure 9. MSA Technology Heat Map (MSAs on horizontal axis, Technology Classes on vertical axis)

The five most diverse MSAs are:

1. New York-Northern New Jersey-Long Island NY-NJ-PA
2. Chicago-Joliet-Naperville IL-IN-WI
3. Los Angeles-Long Beach-Santa Ana CA
4. Boston-Cambridge-Quincy MA-NH
5. Minneapolis-St. Paul-Bloomington MN-WI

Many observations stand out in this heat-map:

- Some Technology Classes are common across the board, almost in all MSAs. This is visible in the horizontal streaks of green.
- The most "diverse" MSAs (highest number of Technology Classes, i.e. left quarter of the map) have patents across many classes, rather than being restricted. This

means innovation in one domain can be correlated with innovation in others. The significance of this can be elaborated on if Technology Classes are arranged in groupings of similar Technology Classes.

- Less diverse MSAs might be home to niche research and/or industries.
- The New York MSA does not stand out in number of patents per million residents (366, while the leader, San Jose has 9,237), but is still the most diverse because it has a high volume of patents.

The above heat-map is a good start for understanding diversity of Technological Profiles of MSAs, and, as indicated in the conclusion, it needs further elaboration to optimize this analysis.

Conclusion

While patent data is considered a log or register of human knowledge creation, it is weakly correlated with economic development at the level of the Metropolitan Statistical Area. This can be attributed to a variety of reasons and circumstances. To mitigate their effects, a revision of the work done in this analysis can be conducted using countries as the spatial unit of analysis. While this will significantly decrease the sample size, it might provide more insights into the relation between the two. Such analysis must account for the lag of impact of innovation on economic development, which means retaining the aforementioned lead of patent sample over GDP or other indicator of economic growth.

Additionally, and if/where the data is available, segmenting the patents and economic growth into relevant sectors will enable better identification of any correlational, maybe even causational, relation.

Another option is to expand the way we view patent data and qualify the sample further. This will not be straightforward because patent data is not designed with the requirement of being a researched dataset, because there is no mandate to make the data available to the public (e.g. as in the case for census). Changing that can facilitate researching the data. For example, it would be beneficial to examine the utility of at least a sample of patents, in terms of actual utilization, their respective outcome in terms of inventions, and the contribution of patents and inventions to economic growth. This also allows sifting through the clutter of less useful patents and working on quality records. Imposing a utility validation can be a part of the actual data cleaning before digging into the data. Tracking a "lifecycle" of the implementation of hundreds or thousands of patents would shed some light on issues such as whether innovation really drives economic growth.

As for diversity of Technology Classes within MSAs, we note, a clearly-defined scaling or normalization of patent numbers per Technology Class per MSA would allow analysis and comparison to establish trends of diverse MSAs. It would also allow assessing less diverse MSAs to determine if they have some niche strength that makes up for the low diversity. Additionally, a better understanding of the distribution of Technology Class "families", or what Rothwell et al. (2013) refers to as patent subcategories, would allow us to further analyze the clusters of Technology Classes visible as green beads on the heat-map.

References

Class notes, mostly from Prof Lobo's last two sessions

Hidalgo CA, Klinger B, Barabasi A-L, Hausmann R (2007) The product space conditions the development of nations. *Science* 317: 482–487.

Muneepeerakul R, Lobo J, Shutters ST, Gomez-Lievano A, Qubbaj MR (2013) Urban Economies and Occupation Space: Can They Get ''There'' from ''Here''? *PLoS ONE* 8(9): e73676. doi:10.1371/journal.pone.0073676

Rothwell, J., Lobo, J., Strumsky, D., & Muro, M. (2013, February). Patenting Prosperity: Invention and Economic Performance in the United States and its Metropolitan Areas. *Metropolitan Policy Program*.