**Aspenia**
online

Pubblicata su *Aspenia online* (http://www.aspeninstitute.it/aspenia-online)

# Big data and economic complexity: opportunity and myth

Luciano Pietronero - Global Issues - 3/6/2015

The immense accumulation of data is a new phenomenon which induces many considerations, represents great potential and sometimes leads to mythical expectations. Here we show a specific example of "big data" applications, the case of Economic Complexity. This is a new perspective on fundamental economics, adopting a bottom-up approach which starts with a novel use of older data and then develops into its own streamline. It is currently being applied to a variety of problems by our team that comprises of ISC-CNR and Sapienza University of Rome (Ref. 1 – 3).

The approach confirms some expectations about big data but also disproves others.

For some practitioners the availability of big data is a problem of computer memory and access speed. Once the data set is large enough, it contains all the possible information one may need and it will speak for itself. This is actually not the case in general but it can basically work in some specific cases.

One such example is the analysis of economic inequality put forth by Thomas Piketty – which has attracted so much attention among economists and policymakers. In this case the problem is to compute a single ratio, i.e. the level of inequality. The work he undertook consists in the accumulation, cleaning and checking of the available data. Once these are completed, the calculation is indeed straightforward. The result is nevertheless remarkable vis-à-vis mainstream economics, and is based on a shift in attitude by the analyst: Piketty starts from the data and then discovers interesting correlations.

In general, when analyzing complex phenomena, things are less straightforward, as can be shown with the specific example of Economic Complexity (see footnotes 1-3). The standard analysis of the competitiveness of countries considers a number of elements like education, transportation, production, export, pollution etc., and through a suitable weighting of these elements leads to a global score for the country. In a way this is not specifically a problem with big data, but of a traditional nature. In the end this analysis needs the fixing of more than 100 parameters for all of these elements, each of which is characterized by a lot of data. Clearly this is a subjective task which cannot properly consider all the possible interactions between the elements involved. In addition, the analysis is made for each country individually and only at the end are different countries compared.

Economic Complexity entails a change of perspective and goes beyond the individual analysis. All countries are considered as nodes of an integrated network and the links are given by the products they produce. In practice one considers the bipartite network of countries and products. The first problem with this is that one needs coherent and homogeneous data which usually are not available or are only partly available. So the new vision immediately shows the limitations of the available data. It is a qualitative

problem, not dependent on how big the data are. One then starts working with what is available but would really like to have better raw materials.

In principle we have access to more information than just the products, but the problem is that these data are not independent of each other: thus, considering them all just leads to confusion or to an interpretation which is intrinsically subjective, as parameters and weights need to be assigned. We tried to go beyond this stage and make an analysis which is scientific, namely one that provides a unique result, not dependent on any subjective interpretation. This leads to a selection among the data and actually to a reduction of them. Only a selected subset is really useful (see footnotes 1-3); adding more data only leads to confusion.

This shows that a big data problem often starts with small data. One has to select a method of analysis and choose the questions and problems to consider. The data do not provide these things by themselves. In this case one needs an algorithm similar in spirit to the Google Page Rank. In economics, however, the Google algorithm is not appropriate and we had to look for a different one. This is a conceptual part of the work which, however, needs the appropriate data to be tested.

Once the conceptual challenge is overcome, the methodology begins to produce nice results and the approach demonstrates its practical value. Yet, as ambition grows, the limits of the original data set become apparent. This opens the search for much more data but in a specific direction which is identified by the method of analysis and the new algorithm. Now the problem can evolve at a real big data level by adding more and more information on the countries and the products but in the new perspective.

A natural evolution is then to move on to the analysis of individual companies (in addition to countries), and here a new situation appears. Companies are specialized in terms of products, so a matrix of companies and products would lead to a very limited set of information and not be particularly useful. One has to study which are the data suitable for companies and which are the new criterion and the new algorithm to extract useful information from these data. This is today's frontier.

Take home message 1: as soon as you have a new idea and a new algorithm you immediately realize that the data available (originally collected for different purposes) are not optimal and you want more data of a new type. There is no infinite dataset one may collect *a priori* which is good for all problems.

The step we have indicated corresponds to a shift from the individual country analysis with 100 parameters to a network algorithmic analysis with zero parameters. In practice we have the data of which country produces which product and the algorithm leads directly to the results. So one may think that the key to big data analysis is the study of Complex Networks. Indeed there is vast literature mostly on the statistical characterization of the properties of Complex Networks, but is this really what we need?

The example of Google Page Rank, rather, would point in a different direction. No matter what the specific characteristic and structure of the network are, this algorithm is successful in defining the correct hierarchy of websites. On the other hand, in the absence of such an algorithm, the classification of the specific properties of the network would not lead to much useful information.

The situation is actually similar for our algorithm for countries and products. With the standard Complex Network studies one can show, for instance, that in the past decade the economic cluster around China has become larger than that around Japan. Hardly a surprise and not a particularly interesting result. However, with the appropriate algorithm one can get a wealth of results like the ranking of countries and products, identification of hidden potential and forecasting of GDP growth, etc. (see footnotes 1-3).

Take home message 2: Big data science in the sense we have indicated can indeed produce a revolution in our knowledge in many fields. But for each area there should be a clear understanding of what the relevant information is and how to extract it from the data. This cannot be a single recipe for all fields of analysis: instead, it should be studied and tailored to each problem.

## References

[1] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli and L. Pietronero: *A New Metrics for Countries' Fitness and Products' Complexity*, Nature: Scientific Reports, 2-723 (2012)

[2] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli and L. Pietronero: *Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products*, PLOS One Vol. 8, e70726 (2013)

[3] M. Cristelli, A. Tacchella, L. Pietronero: *The Heterogeneous Dynamics of Economic Complexity,* PLOS One 10(2): e0117174 (2015) and Nature editorial 2015: http://www.nature.com/news/physicists-make-weather-forecasts-for-economies-1.16963

**URL di origine:** http://www.aspeninstitute.it/aspenia-online/article/big-data-and-economic-complexity-opportunity-and-myth

**Links:**
[1] http://www.aspeninstitute.it/aspenia-online/contributors/luciano-pietronero
[2] http://www.nature.com/news/physicists-make-weather-forecasts-for-economies-1.16963
[3] http://www.aspeninstitute.it/aspenia-online/article/i-media-nell'era-dei-big-data-innovazione-con-giudizio
[4] http://www.aspeninstitute.it/aspenia-online/article/intelligence-svelare-le-menzogne-nella-penombra-dei-big-data

**Links - other articles by this author:**
http://www.aspeninstitute.it/aspenia-online/article/big-data-and-economic-complexity-opportunity-and-myth