The detection of "hot regions" in the geography of science – A visualization approach by using density maps

Ludo Waltman[1] and Lutz Bornmann[2]

[1] *waltmanlr@cwts.leidenuniv.nl*
Centre for Science and Technology Studies, Leiden University, Leiden (The Netherlands)

[2] *bornmann@gv.mpg.de*
Max Planck Society, Hofgartenstr. 8, D-80539 Munich (Germany)

Introduction

In a recent paper, Bornmann, Leydesdorff, Walch-Solimena, and Ettl (in preparation) introduced new methods to map centers of excellence around the world using freely available computer programs. By colorizing cities in dependency on the numbers of excellent papers, their maps provide visualizations where cities with a high (or low) output of excellent papers can be found. To demonstrate their approach, Bornmann et al. presented maps for excellent papers in neuroscience, physics & astronomy and social sciences as well as for *Nature* and *Science* papers. The present study is intended to follow their approach in general, but to change the focus from mapping of single data points (i.e., cities) to a more "sliding" visualization of regions. In our opinion this visualization, which is based on density maps, can serve as a useful complement to the maps proposed by Bornmann et al. The advantage of our new visualization is that the few broader regions characterized by an excellent paper output can be identified (worldwide or for a specific continent).

The material needed to produce our proposed density maps can be found online at www.ludowaltman.nl/density_map/. On this web page we have also made available detailed technical instructions on how to produce the maps.

Methods

Our proposed visualization is inspired by the density maps offered by the VOSviewer software for bibliometric mapping (Van Eck & Waltman, 2010). The main difference with the density maps produced by the VOSviewer software is that in the present study we work with geographical maps rather than with maps of some abstract space of, for instance, documents or keywords. Visualizations similar to those proposed in this study are also being used in other scientific fields (see here Shi, 2010).

We use the following methodology to make our density maps. We first construct a raster that exactly covers the area of which we want to make a map. The center of the cell in the ith row and the jth column of the raster has latitude $\varphi(i)$ and longitude $\lambda(j)$.

Suppose we have a large number of publications for which we know the geographical coordinates of the authors' cities (for a discussion of how these coordinates can be obtained, see Bornmann et al., in preparation). Suppose there are m different pairs of coordinates. The kth pair of coordinates is given by a latitude φ_k and a longitude λ_k. Each pair of coordinates has a weight w_k indicating the number of publications associated with that pair of coordinates. Based on these data, we calculate for each cell in our raster the density of publications. The way in which this is done is similar to the statistical technique of kernel density estimation (e.g., Scott, 1992). For the cell in the ith row and the jth column of our raster, we calculate the density of publications as

$$D(i,j) = \frac{1}{h^2} \sum_{k=1}^{m} w_k K\left(\frac{d(\varphi_k, \lambda_k, \varphi(i), \lambda(j))}{h}\right),$$

where $d(\varphi_k, \lambda_k, \varphi(i), \lambda(j))$ denotes the great-circle distance from φ_k and λ_k to $\varphi(i)$ and $\lambda(j)$, h denotes the kernel width parameter, and K denotes the kernel function. The kernel width parameter h determines the smoothness of the density map. Choosing an appropriate value for this parameter usually requires some trial and error. The kernel function K that we use is given by

$$K(u) = 0.5\exp(-u).$$

After calculating the density of publications for each cell in our raster, we determine the color of each cell. We use a color scheme consisting of the colors white, green, yellow, and red. White indicates the lowest publication density: no publications at all. Yellow means twice as many publications as green, and red means five (or more) times as many publications as green.

As a final step, we use a geographic information system (GIS) to display the density map. The GIS shows the raster of colors that we have constructed. On top of the raster, it shows the borders of countries and the locations of publications. The GIS that we use is Quantum GIS (www.qgis.org).

Results

To demonstrate our method for creating density maps, we produced a field-specific and a field-overlapping map focusing on the European region.

Figure 1 shows the locations of authors in Europe having published highly cited articles in biochemistry, genetics, & molecular biology. The map is based on the top 1% of articles published in 2007 in a journal of the Scopus journal set biochemistry, genetics, & molecular biology. On the map the individual locations of the authors having published the top 1% articles (the small circles on the map) and the geographic regions with the different density colors (red, yellow, green, and white) are visible. We decided to also draw densities on sea and ocean areas, as the map looks better when densities are also allowed to cover these areas. If we focus on geographic regions with a higher concentration of excellent papers (yellow colored regions), these regions are: (1) around Paris, (2) broad areas in Belgium and the Netherlands, as well as (3) some parts of Germany. Clearly, the highest concentration of excellent papers can be found in the region London – Cambridge – Oxford.

We now consider the results of a field-overlapping analysis of excellent papers. We downloaded from Web of Science the bibliographic data of all articles published in 2007 in the high-impact journals *Nature* and *Science*. The map based on this data is presented in Figure 2. In agreement to the field-specific map presented above, the highest density of papers is visible around London – Cambridge – Oxford. However, when comparing this map with a similar map for 2009 (see www.ludowaltman.nl/density_map/), it can be observed that the overall contribution of the UK to *Nature* and *Science* (relative to the total publication output of the journals) has decreased between 2007 and 2009.

We note that our density maps can also be viewed using Google Earth. Please visit our webpage mentioned above for examples of this possibility.

Discussion

The methods presented in this paper allow for an analysis revealing regions of excellence around the world using freely available computer programs. The methods as described above are especially oriented on broader excellence regions and do not allow for the identification of research institutions or single cities on the map where the authors of excellent papers are located.

Our proposed approach has a number of limitations:

1. A significant amount of manual parameter tuning is necessary in order to obtain insightful density maps.
2. The maps are not useful for small data sets.
3. The correct interpretation of a density map may not always be obvious to everyone. Some people may be confused because areas without a university (and even sea and ocean areas) may still have a green, yellow, or red color.

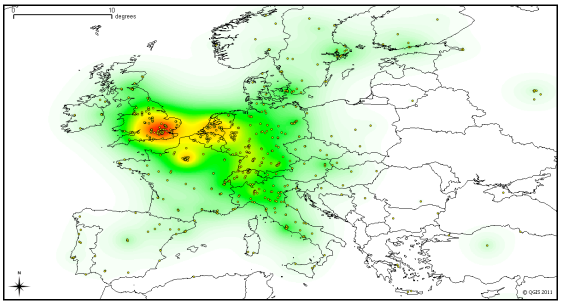

Figure 1. Density map of authors in Europe having published highly cited biochemistry, genetics, & molecular biology articles in 2007 (searched in Scopus).

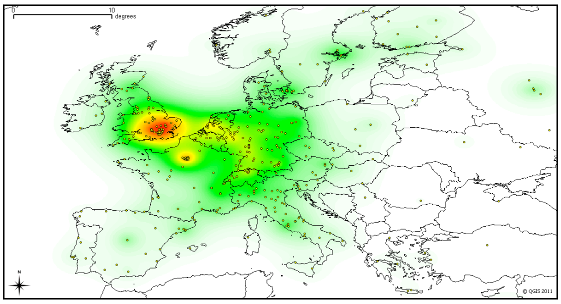

Figure 2. Density map of authors in Europe having published *Nature* or *Science* articles in 2007 (searched in Web of Science).

References

Bornmann, L., Leydesdorff, L., Walch-Solimena, C., & Ettl, C. (in preparation). How to map excellence in the sciences? A mapping approach made possible by using Scopus data. *Journal of Informetrics*.

Scott, D. W. (1992). *Multivariate density estimation: theory, practice, and visualization*. Hoboken, NJ, USA: Wiley.

Shi, X. (2010). Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. *International Journal of Geographical Information Science, 24*(5), 643-660.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538.