The Head/tail Division Rule for Characterizing the Scaling of Geographic Space

(Extended abstract from full paper http://arxiv.org/abs/1009.3635)

Bin Jiang and Xintao Liu

Department of Technology and Built Environment, Division of Geomatics
University of Gävle, SE-801 76 Gävle, Sweden
Email: bin.jiang@hig.se, xintao.liu@hig.se

Scaling of geographic space refers to the fact that for a large geographic area its small constituents or units are much more common than the large ones. This paper develops a novel perspective to the scaling of geographic space using large street networks involving both cities and countryside. This paper is motivated by the belief that geographic space essentially exhibits a heavy tailed distribution rather than a normal distribution. We attempt to investigate the scaling of geographic space from the perspective of city and field blocks. This paper is further motivated by another intriguing issue, i.e., how to delineate city boundaries. Delineating city boundaries objectively is essential for many urban studies and urban administrations. Researchers and practitioners alike usually rely on the boundaries provided by census or statistical bureaus. These imposed boundaries are considered to be subjective or even arbitrary.

Three largest European countries France, Germany and UK were adopted in our study. Given a street network of an entire country, we decompose the street network into individual blocks, each of which forms a minimum ring or cycle such as city blocks and field blocks. We adopt the street networks of three largest European countries for the computation and experiments. Before the extraction of individual blocks for scaling analysis, we need to build up topological relationships, which lead to an arc-based graphs or networks.

We compute the arc-based networks to extract individual blocks in order to investigate some scaling properties. To introduce the computation, we adopt a fictive street network shown in Figure 1, which includes forty blocks and several dangling arcs that do not constitute any part of the blocks. To extract the individual blocks, we first need to set a minimum bounding box for the network in order to select an outmost arc to start traversal processes. There are two kinds of traversal processes: left traversal process and right traversal process. The left traversal process means that when comes to a node with two or more arcs, it always chooses the most left arc. On the other hand, it always chooses the most right arc for the right traversal process. Once the traversal process (starting from the outmost arc) is over, it ends up with one cycle: either a minimum cycle (which is a block) or a maximum cycle which is the outmost border. If the maximum cycle is not generated, then the program chooses a reverse direction for the traversal process until the maximum cycle is detected, and the corresponding arcs are marked with the traversal direction (left or right).

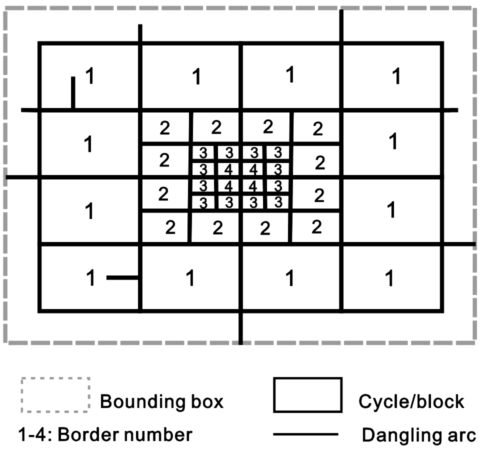


Figure 1: Illustration of the minimum cycles (or blocks) and the maximum cycle

The next step is to choose an arc on the border, and begin the traversal process along the opposite direction as previously marked, until all arcs on the border are processed. This way all the blocks on the border are detected and are assigned to border number 1. This process goes recursively for the blocks that are adjacent to the blocks with border number 1. We will get all the blocks with border number 2. The above process continues until all the blocks are exhausted and are assigned to an appropriate border number; refer to Appendix and Figure A1 for details on the algorithmic procedures. As a note on computation, it takes many hours for the server-like machine to have the process done: France and UK each about 5 hours, and Germany 63 hours. Eventually those dangling arcs are dropped out in the process of extracting the blocks. The border number is a de facto topological distance of a block far from the outmost border (Note: the border is not necessarily a country border). Every block has a border number, showing how far it is from the outmost border. The higher the border number, the farther the block is from the border, or reversely the lower the border number, the closer the block is to the border.

The block sizes demonstrate the scaling property, i.e., far more small blocks than large ones. Interestingly, we find that the mean of all the block sizes can easily separate between small and large blocks- a high percentage (e.g., 90%) of smaller ones and a low percentage (e.g., 10%) of larger ones. Statistically, the block sizes demonstrate one of the heavy tailed distributions, lognormal distribution (Figure 2). This regularity is termed as the head/tail division rule, i.e., *given a variable X, if its values x follow a heavy tailed distribution, then the mean (m) of the values can divide all the values into two parts: a high percentage in the tail, and a low percentage in the head.*

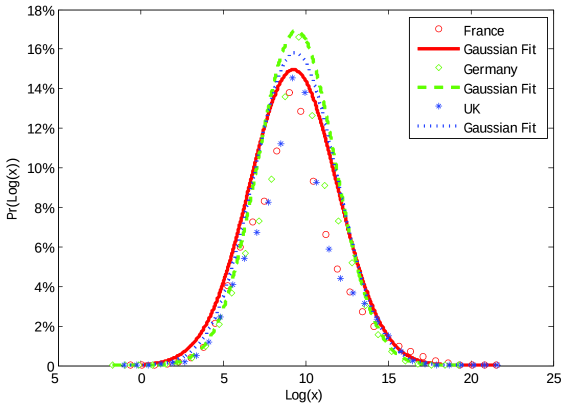
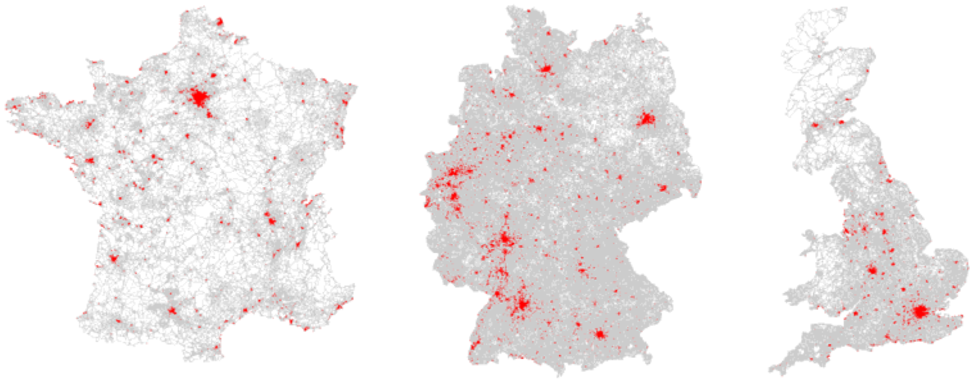


Figure 2: (Color online) Lognormal distribution of the block sizes for the three street networks

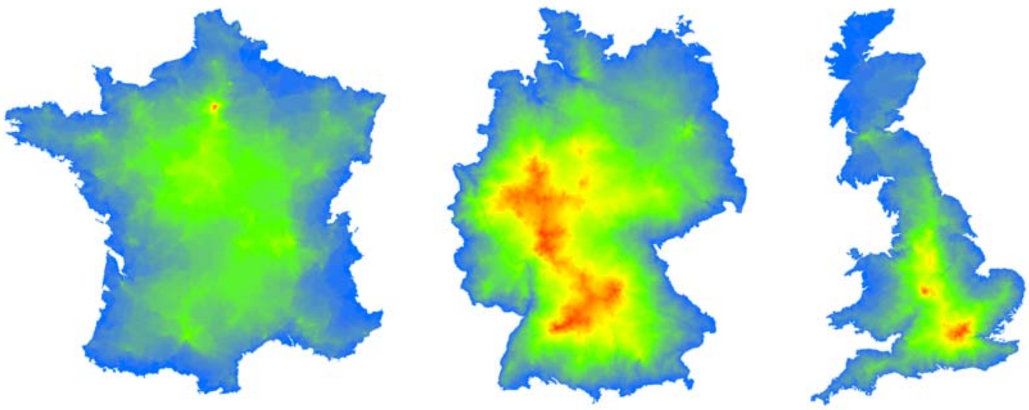
Because all blocks in one country exhibit a heavy tailed distribution (i.e., lognormal distribution), we can use the mean to divide all the blocks into smaller ones (smaller than the mean) and larger ones (larger than the mean). We then cluster the smaller blocks into individual groups. This clustering process goes like this. Starting from any smaller block whose neighboring blocks are also smaller ones, we design a program to traverse its adjacent blocks, and cluster those smaller blocks whose adjacent blocks are also smaller ones. This processing continues recursively until all the smaller ones are exhausted. We find that the sizes of the clustered groups demonstrate a heavy tailed distribution. Because of this, we then rely on the head/tail division rule to divide the groups into smaller ones (smaller than the mean) and larger ones (larger than the mean). The larger groups are de facto cities or natural cities as shown in Figure 3.



(a) (b) (c)

Figure 3: (Color online) All natural cities in red identified for the three networks of (a) France, (b) Germany, and (c) UK (Note: the gray background shows the extracted blocks)

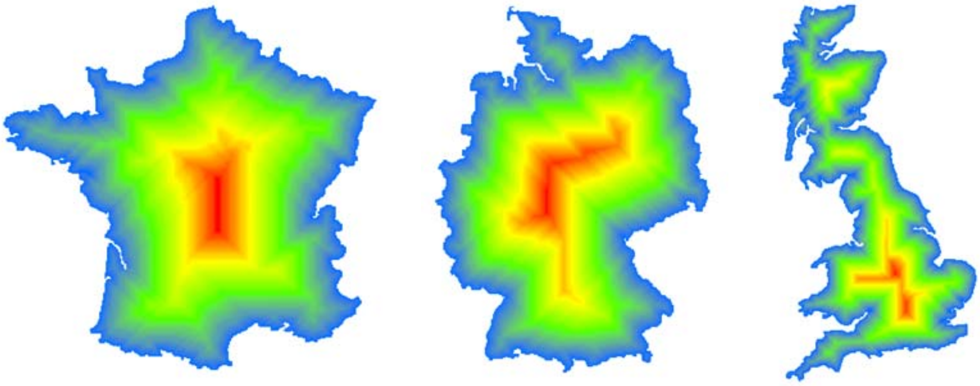
We further define the concept of border number as a topological distance of a block far from the outmost border to map the center(s) of the country and the city or to characterize the scaling of geographic space (Figure 4).



(a) (b) (c)

Figure 4: (Color online) Mapping the border number using a spectral color legend (Note: the higher the border number the warmer the color; red and blue represent respectively the highest and lowest border numbers)

The patterns shown in Figure 4 are illustrated from a topological perspective, which is very different from a geometric one. For example, given any country border or shape, we can partition the shape into equal sized rectangular cells (at a very fine scale, e.g., 1000 x 1000), and then compute the border number for the individual cells. Eventually, we obtain the patterns shown in Figure 5. As we can see, the centers of the countries are geometric or gravity centers that are equal distances to the corresponding edges of the borders. Essentially the country forms or shapes are viewed symmetrically. This is a distorted image of the countries, since the geometric centers are not true centers that the human minds perceive. This geometric view is the fundamental idea behind the concept of medial axis (Blum 1967), which has found a variety of applications in the real world in describing the shape of virtually all kinds of objects from the infinitely large to the infinitely small including biological entities (Leymarie and Kimia 2008). While medial axis is powerful enough in capturing a symmetric structure of a shape, it presents a distorted image of a shape as seen from Figure 5. This distortion is particularly true for France, since the true center Paris is far from the geometric or gravity center.



(a) (b) (c)

Figure 5: (Color online) Distorted images of the country based on the geometric distance far from the outmost boundaries

We add some implications to understanding the morphology of organisms. The city and field blocks can be compared to the cells of complex organisms. We believe that this kind of scaling analysis of geographic space can be applied to complex organisms and we consequently conjecture that a similar scaling structure is appeared in complex organisms like human bodies or human brains. This would reinforce our belief that cities, or geographic space in general, can be compared to a biological entity in terms of their structure and their self-organized nature in their evolution. Our future work will concentrate on the further verification of the findings and applications of the head/tail division rule.

For more details about the study, the reader is encouraged to refer to Jiang and Liu (2011) and references therein.

Jiang B. and Liu X. (2011), Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information, Preprint, arxiv.org/abs/1009.3635.