

Supplementary Information for  
A dominance tree approach to systems of cities

Thomas Louail,<sup>1</sup> Marc Barthelemy,<sup>2,3,\*</sup>

<sup>1</sup>CNRS/Géographie-cités, Campus Condorcet, 93322 Aubervilliers Cedex,  
France

<sup>2</sup>Institut de Physique Théorique, Université Paris Saclay, CEA, CNRS,  
F-91191 Gif-sur-Yvette, France

<sup>3</sup>Centre d'Analyse et de Mathématique Sociales, (CNRS/EHESS) 54,  
Boulevard Raspail, 75006 Paris, France

\*E-mail: marc.barthelemy@ipht.fr

October 11, 2022

# Supplementary text

## Basin sizes: Zipf plots

At each level  $h$ , the remaining cities are local centers. From these seeds, we construct the Voronoi tessellation and inside each cell  $i$  the total population  $P_{tot}(i)$  is given by

$$P_{tot}(i) = P_i + \phi_i \quad (S1)$$

where  $P_i$  is population of the seed city  $i$  and  $\phi_i = \sum_{j \in V(i)} P_j$  is the total population of the seed city  $i$ 's attraction basin, that is the sum of the population sizes of the cities  $j$  dominated by  $i$  and located in  $i$ 's Voronoi cell  $V(i)$ . For these quantities  $P_{tot}(i)$  and  $\phi_i$ , we construct the rank plots (also called Zipf's plots) that are shown in Fig. S1.

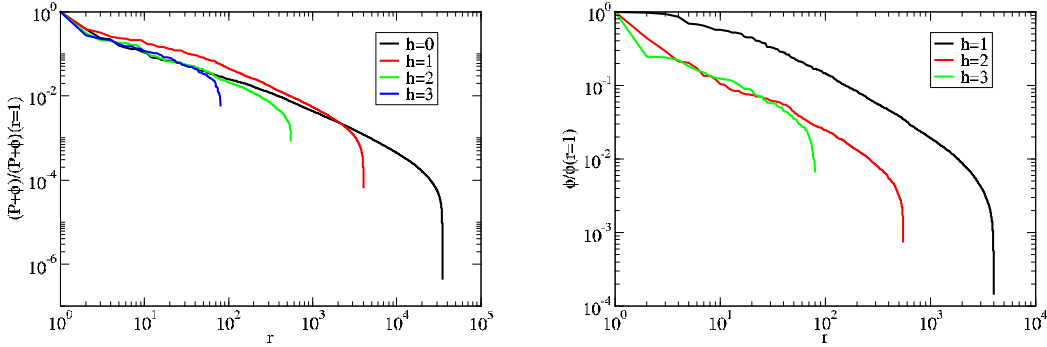


Figure S1: Zipf's plots for the (Left) total population inside Voronoi cells at a given height  $h$  and (Right) for the quantity  $\phi$  (defined as the total population inside Voronoi cells minus the population of the city seed). These plots are normalized by their initial value at rank  $r = 1$ . Power law fits give the exponent values shown in Table S1.

These plots can be fitted by a power law and the values of the exponent  $\alpha$  are summarized in Table S1. In this table, the exponent values are computed for variable ranges (from  $[100, 10000]$  for  $h = 0$  to  $[1, 100]$  for  $h = 3$ . For all these fits we have  $r^2 \approx 0.99$ . These values clearly show a decreasing trend for the exponent, indicating less fluctuations when  $h$  increases (the corresponding distribution varies a power law with exponent  $1 + 1/\alpha$ ).

Height $h$	Exponent $\alpha$ (case $P_{tot}$ )	Exponent $\alpha$ (case $\phi$ )
0	0.97	—
1	0.97	0.89
2	0.82	0.825
3	0.76	0.74

## Statistical definitions of cities

There are two main categories of definitions used by national statistics bureaus in France and in the US. The first one is based on the contiguously built-up area ('morphological' definition): starting from a dense core municipality, neighbor municipalities are aggregated to this core as long as there is no 'gap' (i.e. an empty area between buildings) that is larger than a given distance, which is a free parameter in this method (often taken as a few hundred of meters). The second category ('functional' definitions) consists of definitions that are based on the people's mobility, generally the journey-to-work commuting flows. In this case municipalities are aggregated if the proportion of their residents that commute for work to the previously aggregated municipalities is larger than a given threshold. The resulting objects are often referred to as 'functional urban areas'. These different statistical definitions capture different properties and serve different purposes. For a given city the municipalities that are included in its 'morphological' and 'functional' definitions may then be different. Consequently for a given city, the shape, the total area and the population size are different from one definition to another. For example, in 2017 according to the French INSEE, the municipality of Paris had approx. 2.2M inhabitants, while the Paris urban unit ('unité urbaine' definition) was about 10.7M, and the Paris urban area ('aire d'attraction des villes' definition) had 13M inhabitants.

## Supplementary Figures

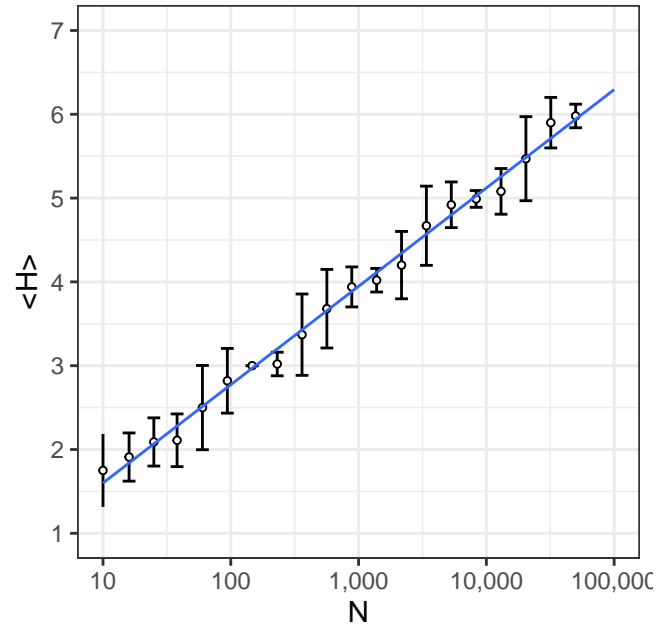


Figure S2: **Average tree height  $\langle H \rangle$  vs. number of points  $N$  in the box, for the ‘random’ toy model.** The blue line represents the best log-lin fit which gives  $\langle H \rangle = 0.5 \cdot \ln N + 0.42$ . Average values  $\langle H \rangle$  were computed over 500 random configurations for each value of  $N$ . Error bars represent  $\pm 1$  standard deviation.

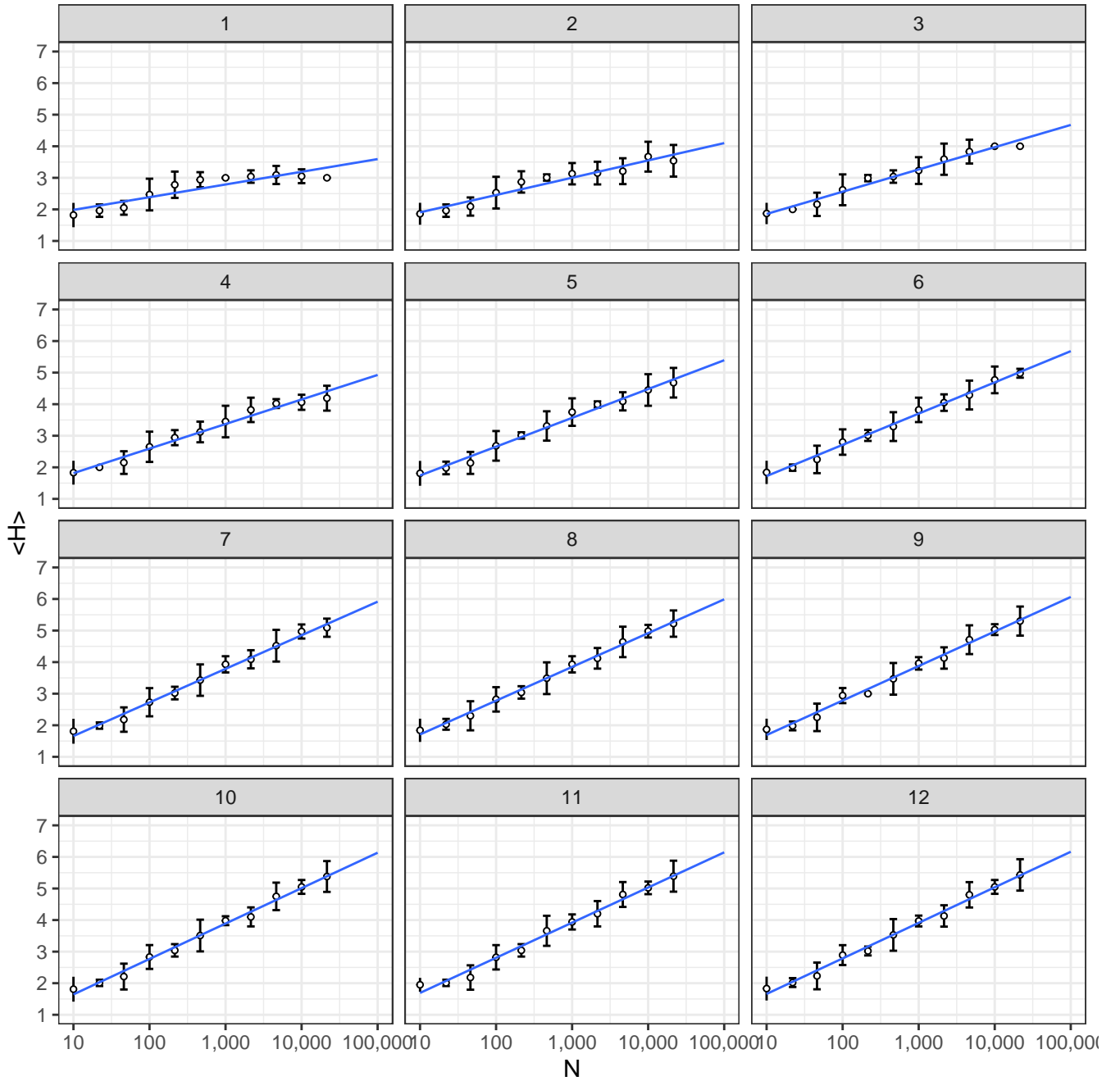


Figure S3: Average tree height vs. number of points in the box, for the ‘tunable’ toy model and for different values of parameter  $L$ . Panels numbers increase as the parameter  $L$  value increases.

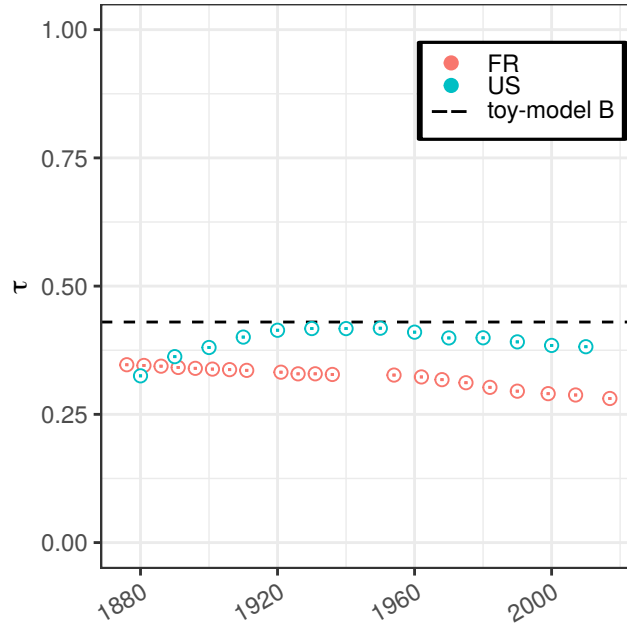


Figure S4: **Kendall  $\tau$  value measuring the correlation between the rank  $r$  and the height  $h$  of cities in the dominance tree, in the US and in France, at different periods for which city population data were available and for which the dominance tree was built.** The dashed line corresponds to the average value obtained for  $10^3$  configurations of the ‘random’ toy model, calculated for  $N = 10^4$  points randomly distributed in a square box. One sees that the two variables  $r$  and  $h$  are less correlated in the French case, which is farther from the total random pairing of ranks and locations of the ‘random’ toy model. This means that in France there are more municipalities with small ranks (large population) which are at small heights (dominated by a close, larger city). This is the case when large municipalities grow in the immediate neighbourhood of even larger ones (“metropolisation”).

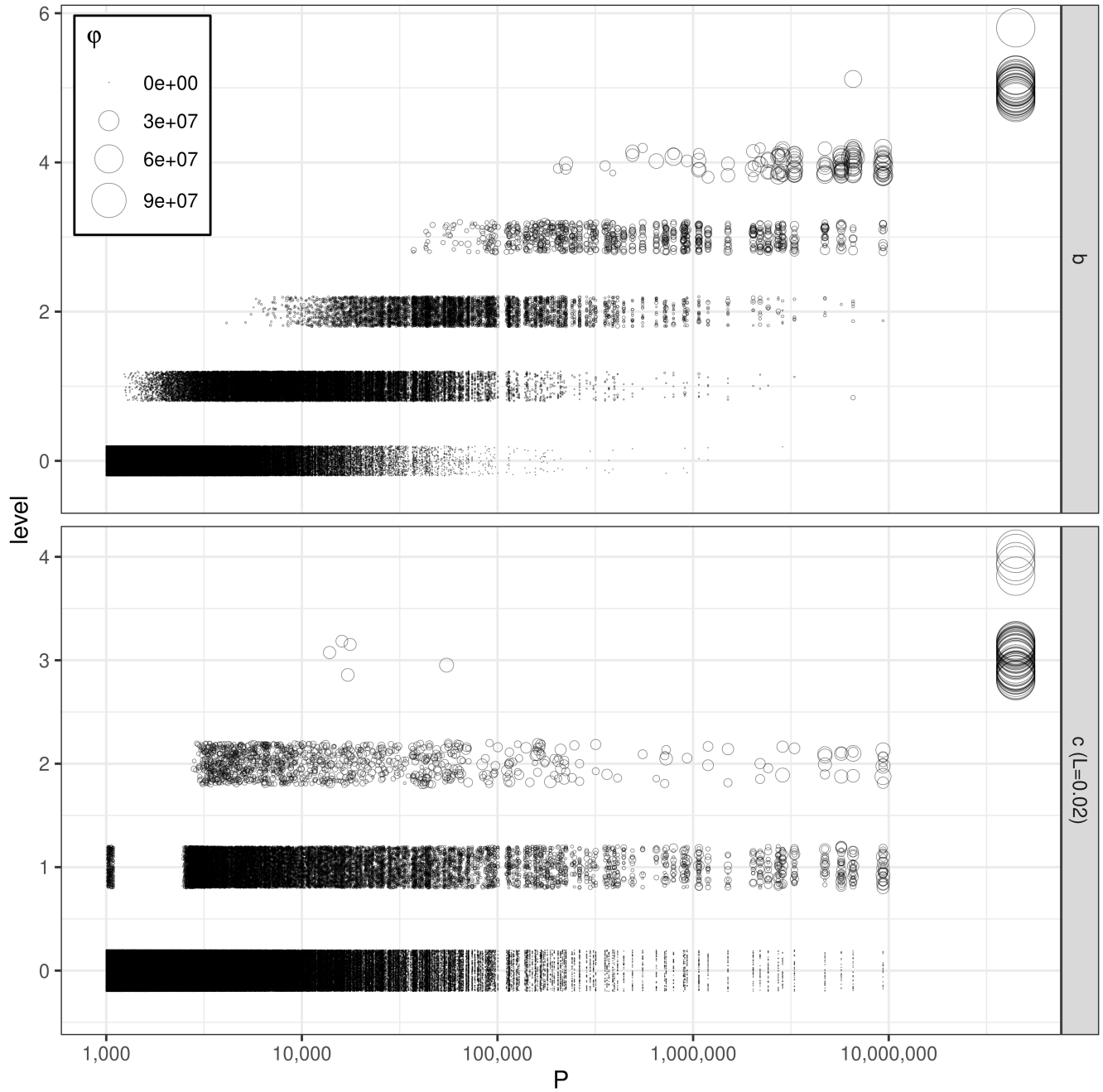


Figure S5: Height in the dominance tree vs. population size for multiple random configurations of the toy model ‘Random’ (top line) and multiple random configurations of the ‘tunable’ toy model, with  $L=0.02$  (bottom line). The overlap area of the horizontal stripes for two consecutive  $h$  values indicates, for that given model, to what extent the mark/rank of the point alone is a poor predictor of its height in the dominance tree, which depends not only on its mark but also on its location. As for each  $h$  value there are a large number of municipalities the dots are jittered along the  $y$  axis for a given  $h$  value to ease the visualization.

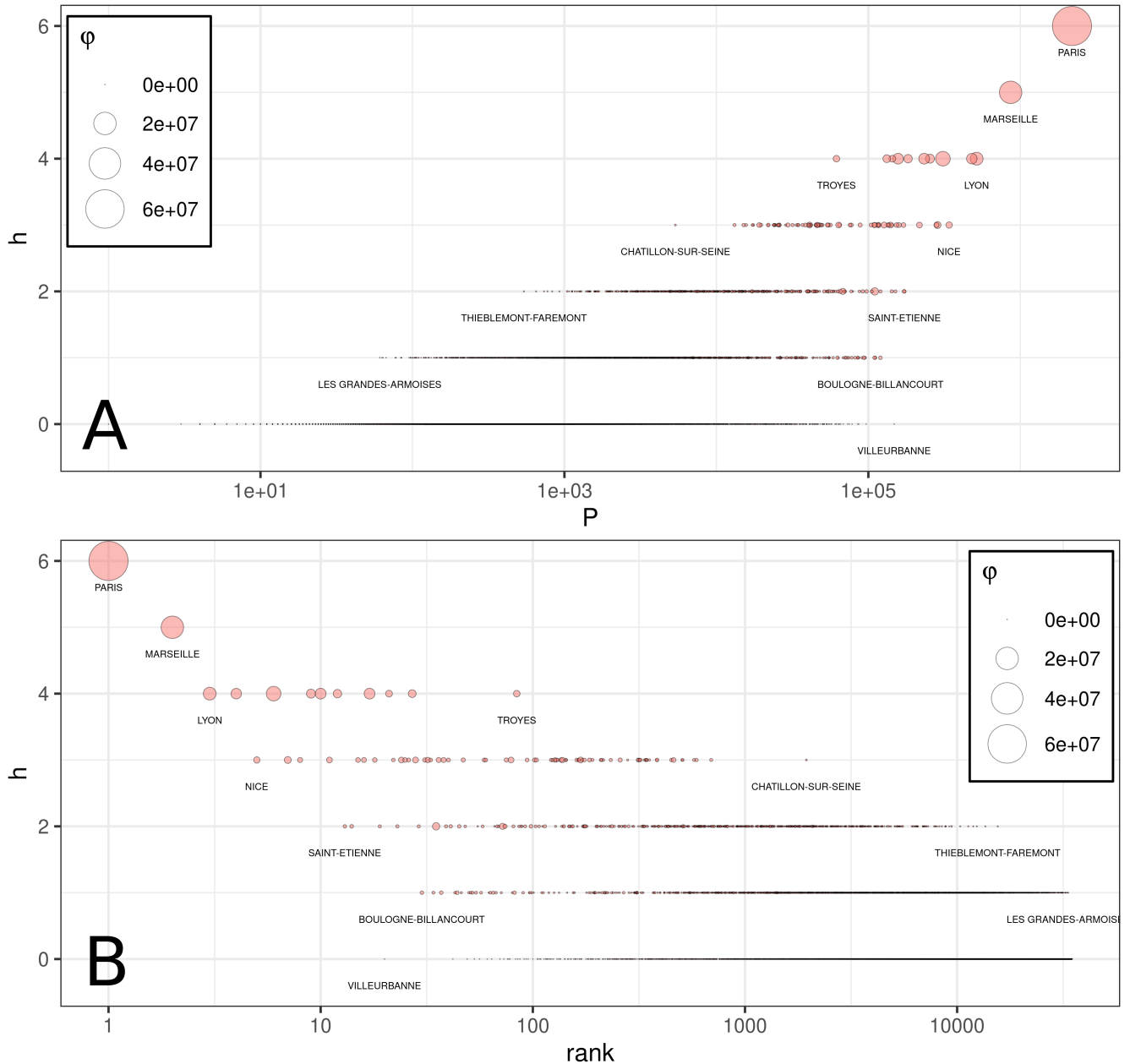


Figure S6: **Height in the dominance tree vs. population size (A) and vs. rank (B) of French municipalities, in 2017**). The size of the dots is proportional to the population size of the city's attraction basin  $\phi$ . Like for the toy models of Supplementary Figure 4, we see horizontal stripes that overlap, proving that the population size of a municipality alone is a poor predictor of its height in the dominance tree. Some large municipalities have a small height because they are absorbed by a larger municipality from which they depend – this is the case for example of Villeurbanne whose population is larger than Troyes, which is however four levels above in the dominance tree. On each line we report the names of the largest and the smallest municipality at this height.



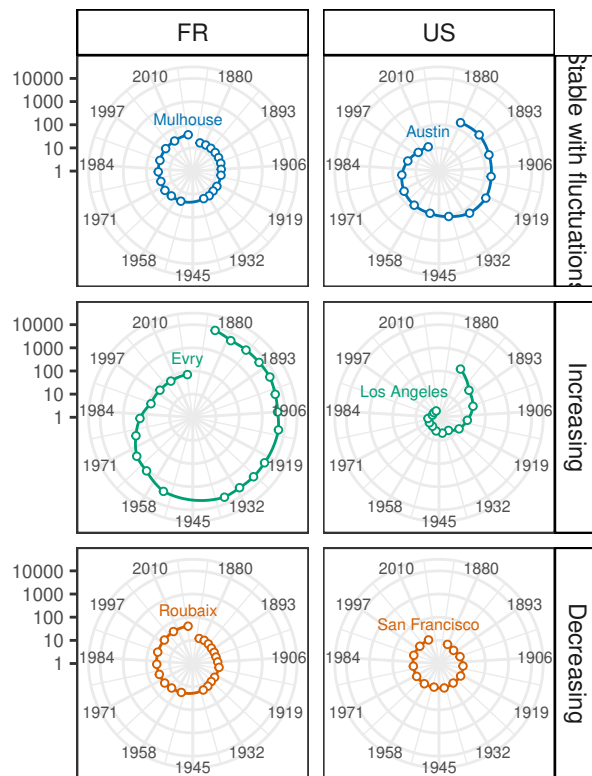
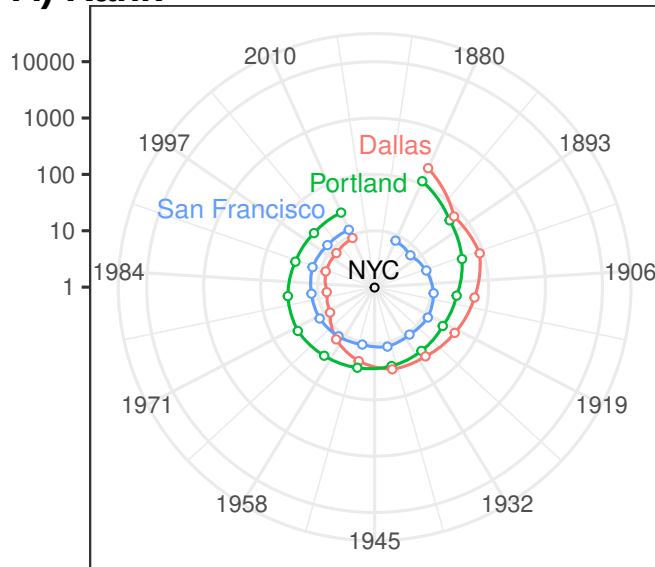


Figure S7: Rank clocks of different cities in the US and in France over the period 1870-2010.

### A) Rank



### B) Height

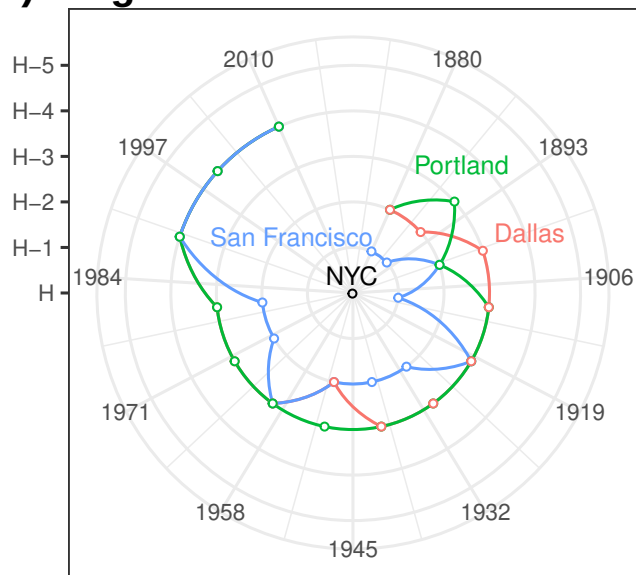


Figure S8: **Rank and height clocks of San Francisco, Portland and Dallas municipalities over the period 1880–2010.** Juxtaposing the rank and height clocks of the same cities highlights how the two variables provide different information, and can possibly evolve in opposite directions. Like other major metropolitan areas in the western part of the US, Dallas has grown very fast in the 20th century and consequently its rank in the urban hierarchy has decreased (nb: the largest city has rank 1, second largest has rank 2, etc.). But in the same time, the population of cities close to Dallas (including San Antonio and Houston) has also increased, with Houston becoming the largest city of the region. This explains that the height of Dallas in the spatial dominance tree did not increase. The same kind of dynamics occurred for Portland, whose height trajectory was limited by the growth of Seattle. For San Francisco the situation is different, its rank in the population hierarchy has remained almost constant, but its height has decreased due to the emergence of a larger city in California (Los Angeles) in the 20th century.