Supplementary material

Testing Heaps' law for cities using administrative and gridded population data sets

Filippo Simini, Charlotte James

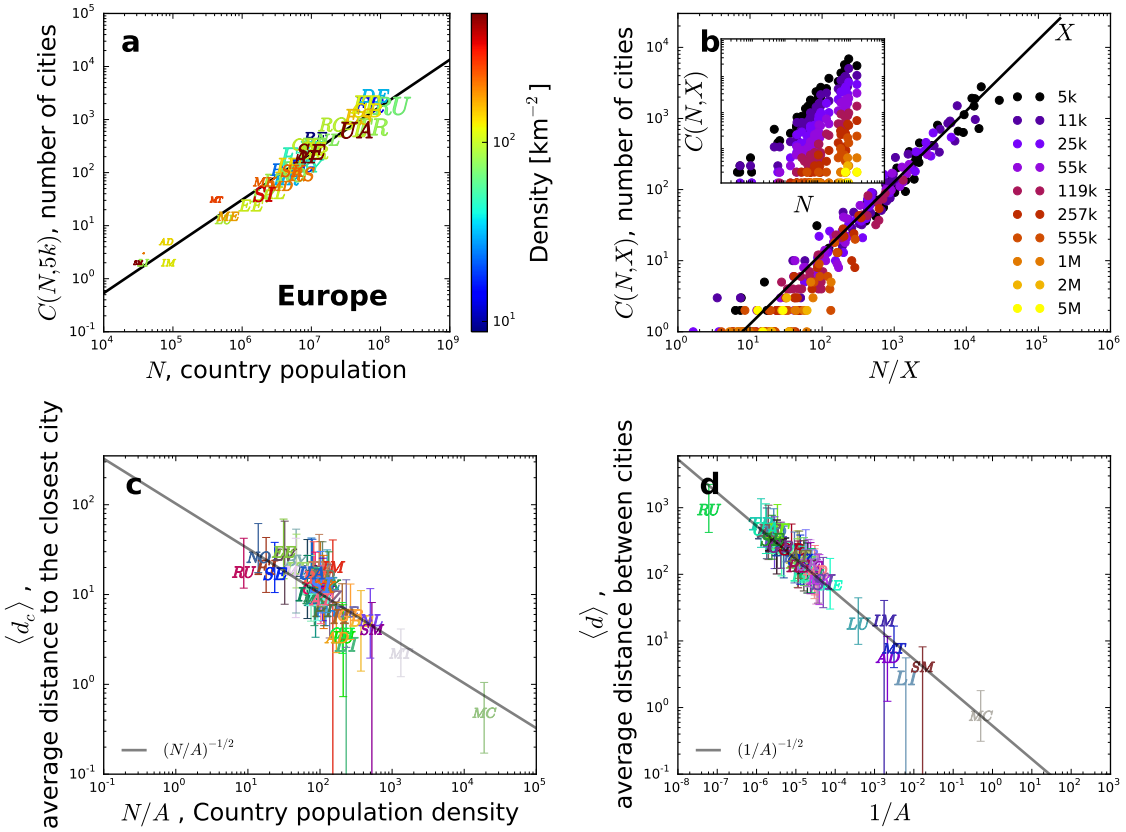

Figure S1: **a** Heaps' law for Europe. Population (x-axis), number of cities with more than 5,000 inhabitants (y-axis), 2-letter country code (marker), logarithm of the area (marker size) and population density (color). The black line is a power law fit of the scaling relationship between the number of cities and the total population. **b** The number of cities with more than X inhabitants for countries in Europe with population N; $C(N,X)$ (y-axis) versus the ratio N/X (x-axis). The scattered cloud of points resulting when plotting $C(N,X)$ against N for various X's in the range $5 \cdot 10^3 - 5 \cdot 10^6$ (inset) collapses on a straight line when $C(N,X)$ is plotted against the ratio N/X. **c** The average distance to the closest city for countries in Europe (y-axis) scales as the inverse of the square root of the country's population density (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average. **d** The average distance between cities for countries in Europe (y-axis) scales as the inverse of the country's area (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average.

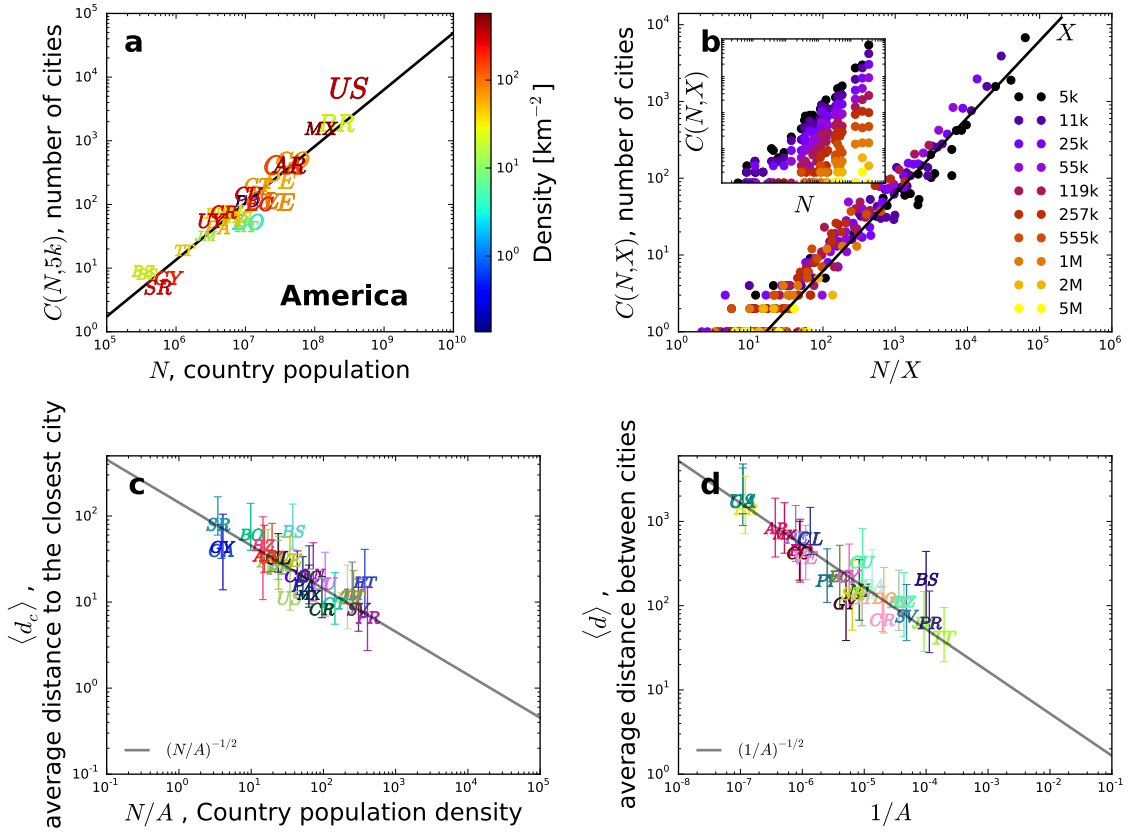

Figure S2: **a** Heaps' law for America. Population (x-axis), number of cities with more than 5,000 inhabitants (y-axis), 2-letter country code (marker), logarithm of the area (marker size) and population density (color). The black line is a power law fit of the scaling relationship between the number of cities and the total population. **b** The number of cities with more than X inhabitants for countries in America with population N; $C(N, X)$ (y-axis) versus the ratio N/X (x-axis). The scattered cloud of points resulting when plotting $C(N, X)$ against N for various X's in the range $5 \cdot 10^3 - 5 \cdot 10^6$ (inset) collapses on a straight line when $C(N, X)$ is plotted against the ratio N/X. **c** The average distance to the closest city for countries in America (y-axis) scales as the inverse of the square root of the country's population density (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average. **d** The average distance between cities for countries in America (y-axis) scales as the inverse of the country's area (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average.

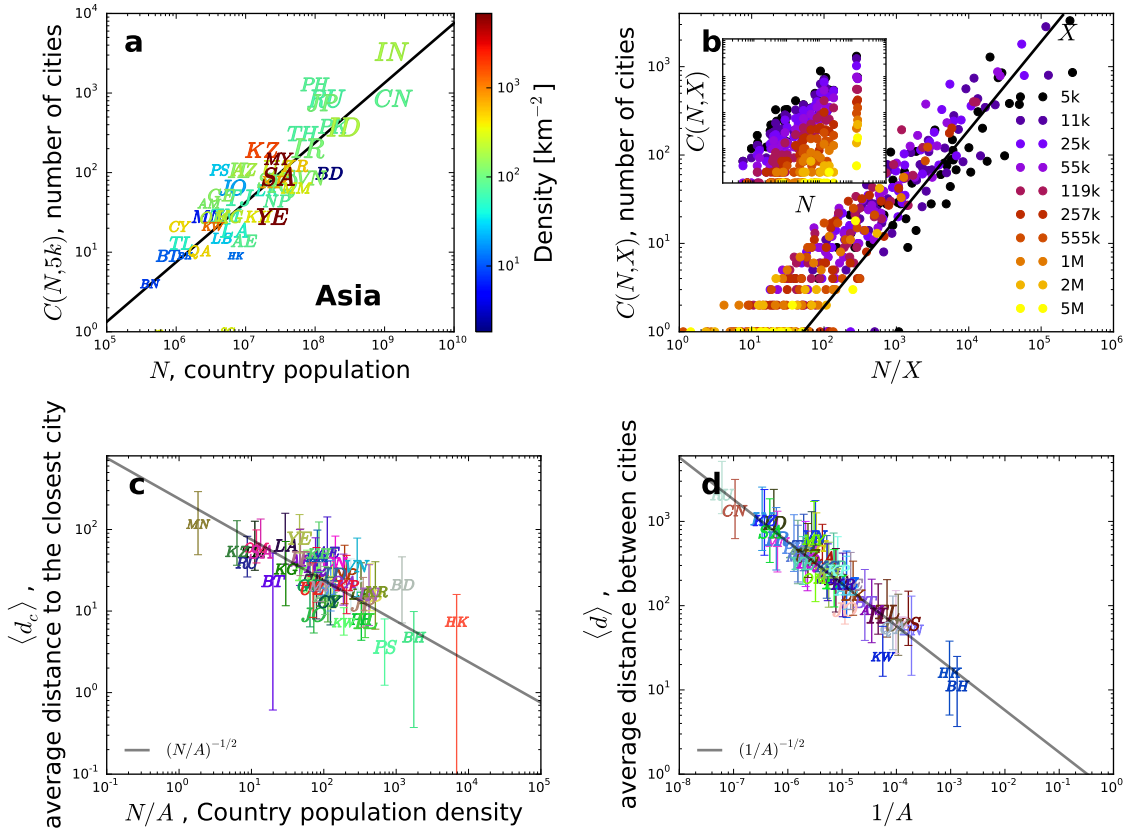

Figure S3: **a** Heaps' law for Asia. Population (x-axis), number of cities with more than 5,000 inhabitants (y-axis), 2-letter country code (marker), logarithm of the area (marker size) and population density (color). The black line is a power law fit of the scaling relationship between the number of cities and the total population. **b** The number of cities with more than X inhabitants for countries in Asia with population N; $C(N,X)$ (y-axis) versus the ratio N/X (x-axis). The scattered cloud of points resulting when plotting $C(N,X)$ against N for various X's in the range $5 \cdot 10^3 - 5 \cdot 10^6$ (inset) collapses on a straight line when $C(N,X)$ is plotted against the ratio N/X. **c** The average distance to the closest city for countries in Asia (y-axis) scales as the inverse of the square root of the country's population density (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average. **d** The average distance between cities for countries in Asia (y-axis) scales as the inverse of the country's area (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average.

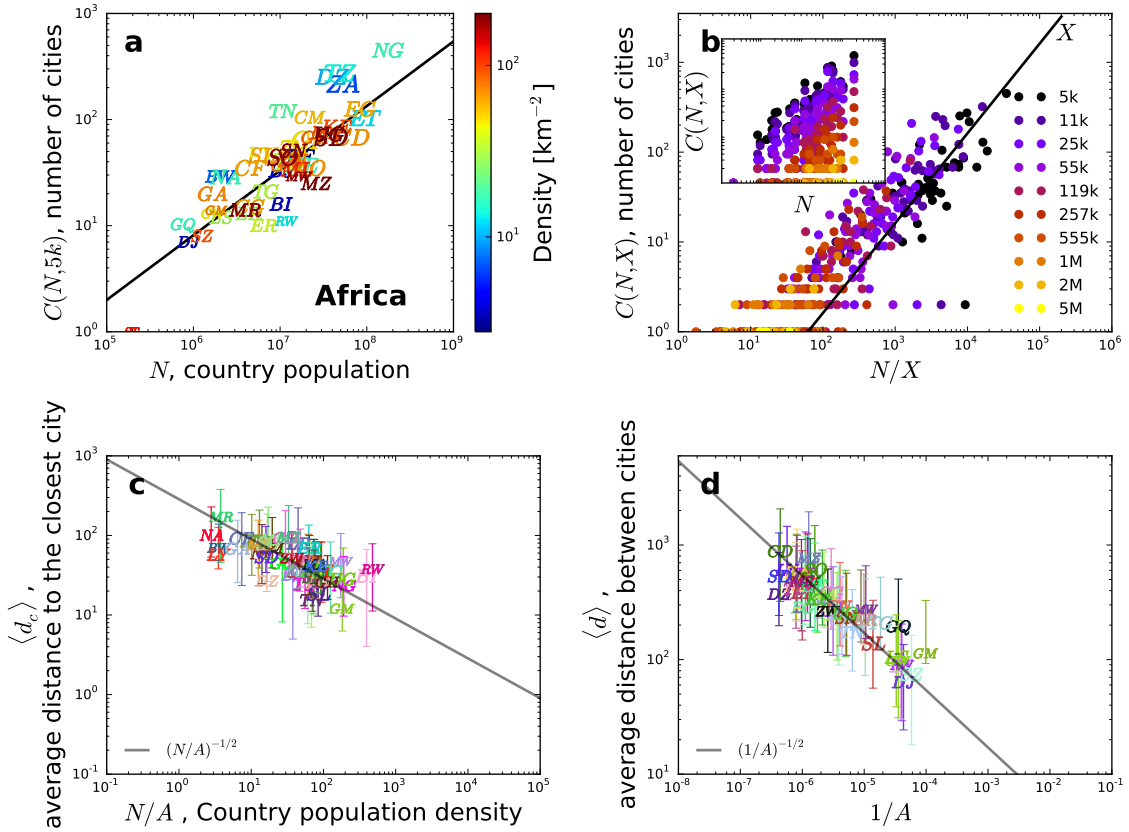

Figure S4: **a** Heaps' law for Africa. Population (x-axis), number of cities with more than 5,000 inhabitants (y-axis), 2-letter country code (marker), logarithm of the area (marker size) and population density (color). The black line is a power law fit of the scaling relationship between the number of cities and the total population. **b** The number of cities with more than X inhabitants for countries in Africa with population N; $C(N,X)$ (y-axis) versus the ratio N/X (x-axis). The scattered cloud of points resulting when plotting $C(N,X)$ against N for various X's in the range $5 \cdot 10^3 - 5 \cdot 10^6$ (inset) collapses on a straight line when $C(N,X)$ is plotted against the ratio N/X. **c** The average distance to the closest city for countries in Africa (y-axis) scales as the inverse of the square root of the country's population density (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average. **d** The average distance between cities for countries in Africa (y-axis) scales as the inverse of the country's area (x-axis). All cities with more than 5,000 inhabitants are considered. The asymmetric error bars denote the standard deviations above and below the average.