Supplementary material

I- Directory Data

Each subdirectory (e.g. "08Barcelona") contains data about a **province**. There is a special directory ("**60Spain**") which collects all the information of Spain as a whole.

Contents

Each province subdirectory has:

1. A tab-separated **"txt" file** with all the information at the census-tract level about the province (Note: This data also includes Germany and Great Britain, which we remove afterwards in the paper). The columns are:
   - 1.1. Census tract ID.
   - 1.2. Census tract centroid (Longitude).
   - 1.3. Census tract centroid (Latitude).
   - 1.4. # Residens from Spain.
   - 1.5. # Residens from Germany.
   - 1.6. # Residens from Bulgaria.
   - 1.7. # Residens from France.
   - 1.8. # Residens from Italy.
   - 1.9. # Residens from Poland.
   - 1.10. # Residens from Portugal.
   - 1.11. # Residens from Great Britain.
   - 1.12. # Residens from Romania.
   - 1.13. # Residens from Russia.
   - 1.14. # Residens from Ukraine.
   - 1.15. # Residens from Algeria.
   - 1.16. # Residens from Morocco.
   - 1.17. # Residens from Nigeria.
   - 1.18. # Residens from Senegal.
   - 1.19. # Residens from Argentina.
   - 1.20. # Residens from Bolivia.
   - 1.21. # Residens from Brazil.
   - 1.22. # Residens from Colombia.
   - 1.23. # Residens from Cuba.
   - 1.24. # Residens from Chile.
   - 1.25. # Residens from Ecuador.
   - 1.26. # Residens from Paraguay.
   - 1.27. # Residens from Peru.
   - 1.28. # Residens from Dominican Republic.
   - 1.29. # Residens from Uruguay.
   - 1.30. # Residens from Venezuela.

1.31. # Residens from China.
1.32. # Residens from China.

2. A computed **".rsa" file** with the adjacency matrix **A** of the province, with a neighborhood radius of 400m. The matrix **A** is a 0-1 symmetric matrix that determines whether any two census tracts are connected geographically, i.e., whether the distance between their centroids is less tan 400 meters. The matrix comes in **Harwell-Boeing** format for sparse symmetric matrices with real numbers.
3. A computed **".rua" file** with the transition matrix **P** of the province, with a neighborhood radius of 400m. The matrix **P** is row-stochastic and it has been computed from **A**, where the transition probability to each node is proportional to its total population. The matrix comes in **Harwell-Boeing** format for sparse nonsymmetric matrices with real numbers.

Loading Sparse Matrices

Sparse matrices in Harwell-Boring format (".rsa" and ".rua" files) can be loaded into Mathematica with the *Import* command. For instance:

P = Import["dir/08BarcelonaFinalWeights(0.4).rua ","HarwellBoeing"];

Data sources

1. For the number of residents, we have manually downloaded population data at the census-tract level from the INE (National Institute of Statistics): http://www.ine.es/jaxi/menu.do?type=pcaxis&file=pcaxis&path=%2Ft20%2Fe245%2Fp07%2F%2Fa2009
   We incorporated this data into the ".txt" file.

2. We bought the centroids data at the census-tract level from "**Geograma S.L.**" and we also merged them into the ".txt" file.

II- Computation files

1. **Computation Example.nb**: A Mathematica file (with explanatory comments) with two simple algorithms for computing the segregation index. It explains two algorithms: the Pagerank-based algotirthm (Proposition 3) and the local-segregation-based (Proposition 1). This example consists of a 3x3 matrix.
2. **Example Province of Soria.nb**: This file contains the necessary code to reproduce our computations for each province or for Spain as a

whol. It allows to load/create the transition matrix and to compute the segregation indices in the province of Soria.