HOW FAR WOULD YOU GO? COMPARING URBAN ACCESS IN 10 GLOBAL CITIES

Imperial College London

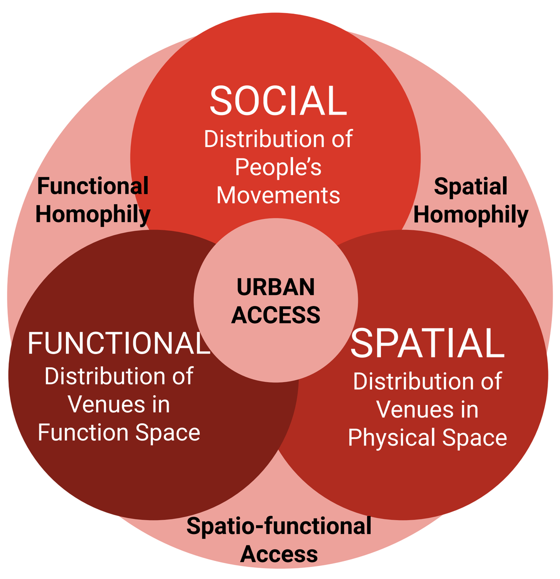
Sahil Loomba and Matthew Garrod

Department of Mathematics, Faculty of Natural Sciences, Imperial College London

ACCESS STATISTICS FOR URBAN MOBILITY

Research on urban growth is increasingly focused on mining patterns of human movement from location-based network data, spread across different axes of spatial, temporal and social dimensions. Some statistical analyses have captured dynamic patterns (Preoţiuc-Pietro and Cohn, 2013), while others have looked at contribution of friendships (Cho, Myers, and Leskovec, 2011) and social influence (Zhang and Pelechrinis, 2014). Furthermore, Bawa-Cavia, 2011 perform an inter-city comparison of polycentricity and fragmentation, but a general, interpretive statistical model is still lacking.

We provide a simple probabilistic framework that captures 3 inter-related yet independently varying views of urban mobility: spatial, functional and social. The spatial view captures number of public venues across physical space (say a city), and thus the opportunities of access. The functional view captures venues across functional categories (say transport, restaurants), and thus the type of access. The social view is the sum of all behavioural considerations (people tend to go from offices to metro stations), ultimately evidenced in number of trips taken between venues.



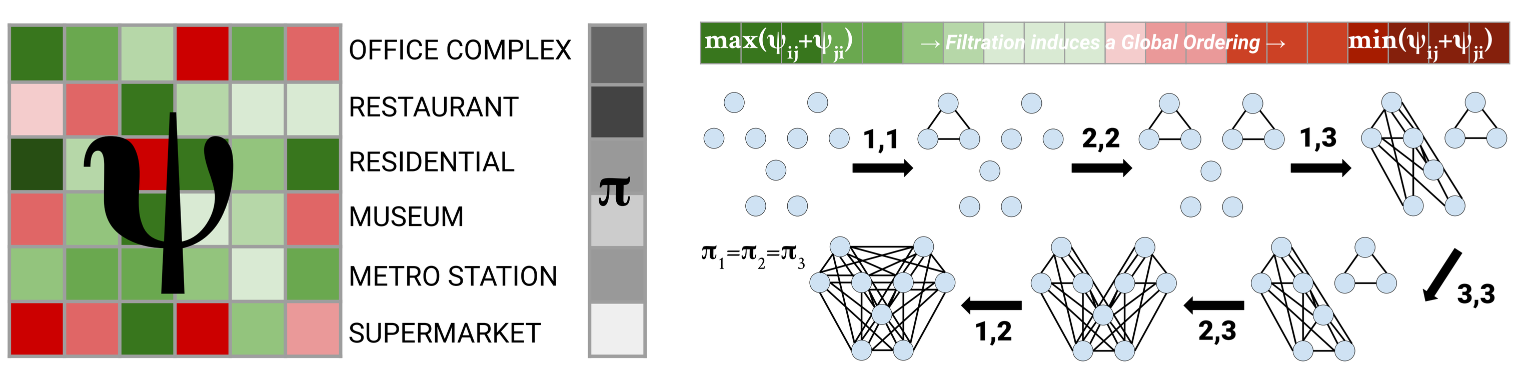
A 3-pronged view of Urban Access along social, spatial and functional dimensions. Social distribution refers to commuter movements, and the median trip distance is a measure for that. Conditioned on venue function it provides category-affinity matrix Ψ, and conditioned on physical space it provides region-affinity matrix $\overline{\Psi}$. Functional distribution refers to distribution of venues in categorical space: π. Spatial distribution refers to distribution of venues in space: $\overline{\pi}$.

Consider check-in data corresponding to venues across m categories according to $\pi \in [0,1]^m$. Let C_{ij} be the trip-count matrix between categories i, j. Trips between two categories can be high for two reasons: either there are many venues of those categories, or people indeed preferentially travel between venues of these two categories. We define this via an affinity matrix Ψ:

$$\begin{aligned} C_{ij} &\propto \pi_i \pi_j \Psi_{ij} \\ C_{ij} &= \mathrm{tr}(C^T C)\pi_i \pi_j \Psi_{ij} \\ \implies \Psi_{ij} &= \frac{C_{ij}}{\mathrm{tr}(C^T C)\pi_i \pi_j} \end{aligned} \tag{1}$$

A network realisation with n venues can be sampled from a Stochastic Block Model: $z_q \sim \mathrm{Multinomial}(\pi) \forall q \in \{0, 1, \ldots n\}, A \sim \mathrm{Bernoulli}(Z \Psi Z^T)$, where $Z \in \{0,1\}^{n \times m}$ is an assignment matrix. Similarly, upon considering distribution of venues across r regions according to $\overline{\pi} \in [0,1]^r$, we can define inter-region affinity matrix $\overline{\Psi}$. Let $\pi(i)$ refer to the functional distribution within region i. Then we define the following statistics for each i:

- **Social Distribution:** median trip distance for trips from region i
- **Functional Distribution:** dispersion in the distribution of venues across categories: $\sum_{k=1}^{m} \pi_k(i)^2$; smaller value indicates higher venue diversity
- **Spatial Distribution:** size of distribution of venues across space: $\overline{\pi}_i$
- **Functional Homophily:** area under the Betti-1 curve given by Ψ and $\pi(i)$; larger value implies propensity to travel between venues of same category
- **Spatial Homophily:** negative-log-ratio of affinities to venues in other regions relative to venues within the region: $-\sum_{j=1}^{r} \overline{\pi}_j \log \frac{\overline{\Psi}_{ij}}{\overline{\Psi}_{ii}}$
- **Spatio-functional access:** number of categories within region of interest (either predefined region i, or any arbitrary region centered around some point with radius R) relative to total number of categories in the city



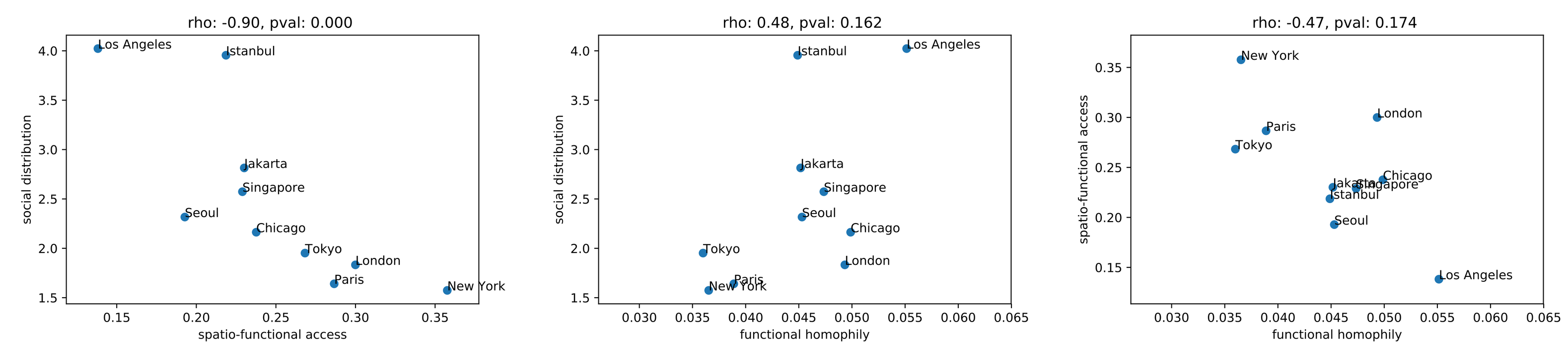
Schematic of Statistics derived using the given probabilistic generative model: (a) Negative-log-ratio is a pairwise measure of homophily, larger values indicating higher homophily (b) area under Betti-1 curve is a global one—it imposes an ordering of community pairs from highest to lowest affinities, asymptotically measuring edge density of networks sampled from these affinities.

ACKNOWLEDGEMENTS

We would like to thank Nick Jones for helpful discussions and the anonymous reviewers for their feedback.

ANALYSIS OF 10 GLOBAL CITIES

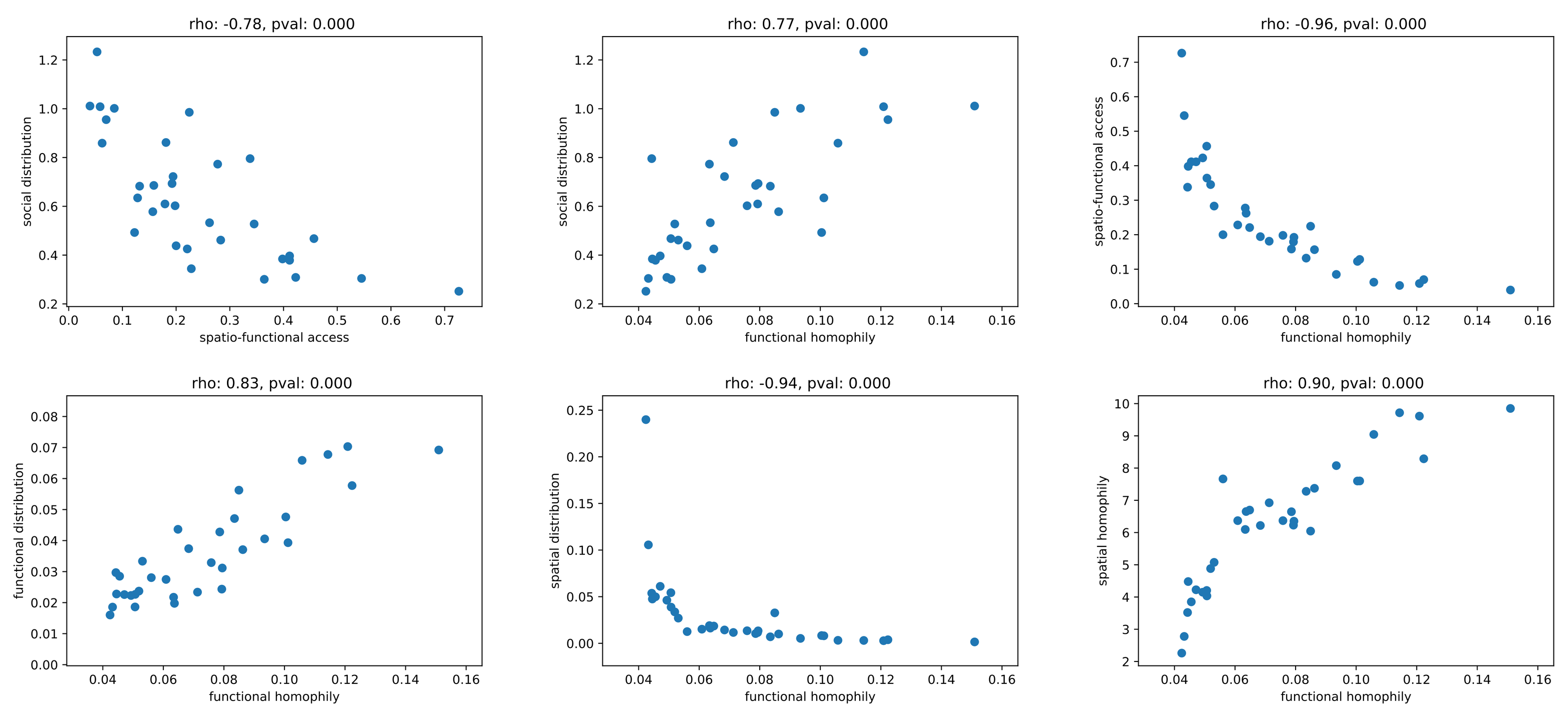
For the 10 cities in the Foursquare dataset, we derive select statistics (spatial homophily cannot be estimated since we do not have information about inter-city travel). Spatio-functional access was estimated by sampling 100 random points in a given city, taking $R = 1km$, and averaging over. Note that every city has its own Ψ, π.



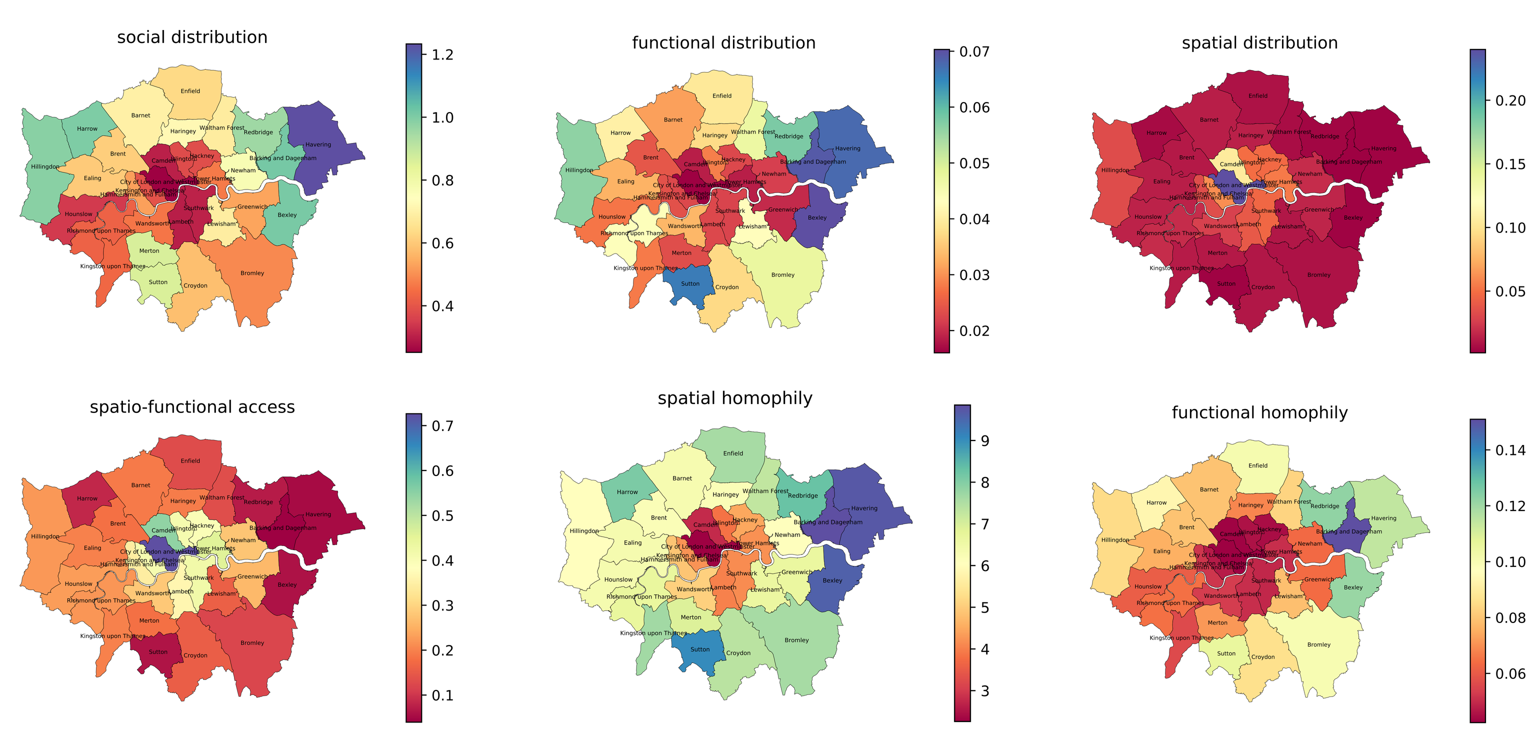
Scatter plots of select urban access statistics for 10 cities in the Foursquares Dataset. We note a significant negative correlation between social distribution and spatio-functional access, but insignificant ones between them and functional homophily.

ANALYSIS OF 33 LONDON BOROUGHS

To obtain more statistical power and comparability, we focus on analysing only one city: London. The 33 boroughs serve as pre-defined regions, for which we can estimate all the statistics mentioned above. $\overline{\Psi}$ would represent inter-borough affinities, and $\overline{\pi}$ indicates distribution of venues across boroughs. Every borough has the same inter-category affinities Ψ corresponding to London, but different $\pi(i)$'s—distribution of venues across categories, given the borough.



Scatter plots of select urban access statistics for 33 boroughs of London. Social distribution represents median trip distance (adjusted for borough size), which correlates negatively to spatio-functional access and positively to functional homophily, suggesting that people travel further to seek diverse venue types. Functional and spatial homophily positively correlate, indicating that areas that encourage venue mixing also encourage spatial mixing. This could be done by simply providing more venues, but the plot between spatial distribution and functional homophily shows that lower functional homophily can be achieved with a low number of venues—as long as they are of the "right" kind.



Maps of urban access statistics for 33 boroughs of London. Areas in central London tend to provide more number, diversity, and "mixing" of venues across space and function, thus requiring people within them to travel over smaller distances.

REFERENCES

Bawa-Cavia, Anil (2011). "Sensing the urban: using location-based social network data in urban analysis". In: *Pervasive PURBA Workshop*, pp. 1–7.

Cho, Eunjoon, Seth A Myers, and Jure Leskovec (2011). "Friendship and mobility: user movement in location-based social networks". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 1082–1090.

Preoţiuc-Pietro, Daniel and Trevor Cohn (2013). "Mining user behaviours: a study of check-in patterns in location based social networks". In: *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, pp. 306–315.

Zhang, Ke and Konstantinos Pelechrinis (2014). "Understanding spatial homophily: the case of peer influence and social selection". In: *Proceedings of the 23rd international conference on World wide web*. ACM, pp. 271–282.