

Anexo A

A.1 Obtención de los datos

Gracias al auge del mercado de las aplicaciones móviles y al aumento del apoyo de las iniciativas “open data” algunos ayuntamientos han empezado a poner a disposición del público mucha información sobre el funcionamiento de las ciudades tanto para su estudio como para su uso en dichas aplicaciones.

Sin embargo, en el caso de la información sobre el transporte urbano no existe un formato estándar para los datos¹. Un ejemplo claro es que mientras que algunos ayuntamientos dan las coordenadas en unidades de longitud y latitud, es decir, en un sistema de referencia que sirve a escala mundial, otras lo dan en formato UTM ED50/ETRS89. El problema de este sistema es que posee 60 orígenes de coordenadas diferentes que dependen de la zona en la que te encuentres, y de hecho en España hay 4 zonas UTM diferentes. Por tanto, el primer paso es estandarizar la posición de todos los nodos de las redes en el formato latitud/longitud para lo cual se han empleado herramientas como los paquetes “raster” y “geosphere” de R así como la aplicación GeoRepository.

Además, en ocasiones los datos contenían errores o faltaba información (por ejemplo, estaban las paradas pero no las líneas de cada parada) por lo que muchas veces ha sido necesario consultar también las páginas oficiales de las distintas empresas de transporte y corregir los datos manualmente.

Por otra parte, los números de identificación de las paradas nunca se reutilizan. Es decir, si ahora se pone una parada nueva se le asigna el número 1000 y dentro de unos años se quita, ese número no volverá a usarse ninguna parada. No solo eso, sino que las empresas de autobuses, tranvía y metro suelen ser diferentes, así que cada uno emplea los números de identificación que quiere (unos empiezan a contar en 1, otros en 100000, etc). La consecuencia de todo esto es que es necesario reasignar nuevos números de identificación a todas y cada una de las paradas de las redes. Puesto que es una tarea imposible de realizar de forma manual para todo este tratamiento de “limpieza y estandarización” de los datos se ha empleado R.

A continuación se presenta la lista de los sitios web de los que se ha extraído la información:

1. Madrid

<http://opendata.emtmadrid.es>
<http://datos.madrid.es>
<http://www.emtmadrid.es>
<http://www.metromadrid.es>

2. Barcelona

<http://www.tmb.cat>
<http://www.metrobarcelona.es>
<http://opendata.bcn.cat>
<http://barcelonaapi.marcpous.com>

3. Valencia

<http://www.gvaoberta.gva.es>
<https://www.emtvalencia.es>

¹Existe una iniciativa de Google que busca estandarizar estos datos bajo un formato único, el GTFS (General Transit Feed Specification), pero todavía no hay muchas ciudades que se hayan unido a la iniciativa [<https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds>]

4. Zaragoza
<http://www.zaragoza.es/docs-api/>
<http://www.urbanosdezaragoza.es>
<http://www.tranviasdezaragoza.es>

5. Bilbao
<http://www.bilbao.net>
<ftp://ftp.geo.euskadi.net/cartografia/Transporte/Moveuskadi/>

6. Malaga
<http://datosabiertos.malaga.eu>
<http://www.emtmalaga.es>

7. Vitoria
www.vitoria-gasteiz.org/tuvisa
<ftp://ftp.geo.euskadi.net/cartografia/Transporte/Moveuskadi/>

Y en la tabla 1 se especifican los datos de cada una de las redes.

Ciudad	Bus		Tram		Metro		Total	
	Líneas	Paradas	Líneas	Paradas	Líneas	Paradas	Capas	Nodos
Madrid	177	4590	3	37	16	241	196	4703
Barcelona	97	2520	6	107	11	137	114	2512
Valencia	46	1136	3	41	6	94	55	1228
Zaragoza	35	902	1	50	-	-	36	915
Málaga	41	1027	-	-	2	17	43	1034
Bilbao	35	514	1	14	2	41	38	555
Vitoria	14	291	2	20	-	-	16	302

Table 1: Características de las redes estudiadas

A.2 Creación de la red

Dentro de cada capa se han considerado que los enlaces son dirigidos, es decir, que se puede ir de A a B pero no de B a A siguiendo un mismo enlace. Matemáticamente, esto significa que la matriz de adyacencia del sistema no es simétrica.

En cuanto al procedimiento para conectar las capas lo que se ha hecho es unir aquellos nodos de metro y de tranvía que estuviesen muy próximos a nodos de autobús, dado que estos últimos son los más numerosos. Para ello, lo primero es buscar el nodo de autobús más cercano a cada parada de tranvía. Si dicho nodo se encuentra a menos de 100m ambos nodos se consideran iguales y se sitúan en el punto medio geográfico que se calcula de la siguiente forma:

1. Se pasa la latitud y la longitud a radianes
2. Se calculan las coordenadas cartesianas mediante las fórmulas

$$\begin{aligned}
 x_i &= \cos(\text{lat}_i) \cos(\text{lon}_i) \\
 y_i &= \cos(\text{lat}_i) \sin(\text{lon}_i) \\
 z_i &= \sin(\text{lat}_i)
 \end{aligned}$$

3. Se calcula el punto medio

$$x = (x_1 + x_2)/2$$

$$y = (y_1 + y_2)/2$$

$$z = (z_1 + z_2)/2$$

4. Se pasan las nuevas coordenadas a latitud y longitud en radianes y, finalmente, a grados

$$\text{lon} = \text{atan2}(y, x)$$

$$\text{hyp} = \sqrt{x^2 + y^2}$$

$$\text{lat} = \text{atan2}(z, \text{hyp})$$

A continuación se repite el mismo proceso para los nodos de metro.

A.3 Determinación de las distancias

La fórmula del haversine es una de las más utilizadas para determinar la distancia entre dos puntos dadas su longitud y latitud. En nombre deriva de la función *versin*, que viene dada por $\text{versin}(\theta) = 2 \sin^2(\theta/2)$, puesto que en la fórmula aparece varias veces el término $\text{versin}(\theta)/2$ (*half-versine*).

Si denotamos por ψ a la latitud, por λ a la longitud y por R al radio de la tierra, la distancia entre los dos puntos, d , vendrá dada por:

$$a = \sin^2(\Delta\psi/2) + \cos\psi_1 \cos\psi_2 \sin^2(\Delta\lambda/2)$$

$$c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$d = R \cdot c$$