



Generalitat de Catalunya
Institut d'Estadística de Catalunya

Creació d'una infraestructura de càlcul de rutes per a l'anàlisi espacial de resultats estadístics. Un assaig per a l'estudi de la distribució dels desplaçaments obligats a Catalunya.

Gabriel Garcies Ramon
Treball final de màster
Barcelona, juliol de 2016

Supervisió a càrrec de:

Dr Joan Nunes

Eduard Suñé

Sumari

Aquest treball és un encàrrec de l'Institut d'Estadística de Catalunya (IDESCAT), institució que té la voluntat d'obtenir la infraestructura d'un SIG que permeti l'anàlisi espacial de resultats estadístics propis. El desenvolupament del projecte es basa en el procés de creació d'un graf per a diferents modes de transport, el càlcul de rutes massiu i l'anàlisi estadística dels resultats sobre desplaçaments obligats a Catalunya. De totes maneres, serà un assaig per al futur desenvolupament de la infraestructura de càlcul de rutes que es pretén obtenir. Es farà ús de programari lliure i de dades obertes per dur a terme el projecte.

Paraules clau: *Graf, càlcul de rutes, bases de dades, mobilitat, SIG.*

Abstract

This research is an assignment from the Statistics Institute of Catalunya (IDESCAT), an institution with the will to obtain a GIS infrastructure that provides a spatial analysis of their own statistical results. This project is based on the creation of a routing network (including different means of transport), the mass routing calculations and the statistical analysis of this data. Nevertheless, this will be a test for a future development of the expected infrastructure. Free software and open data will be used during the project.

Key words: *Network, routing, database, mobility.*

Índex

I. Introducció

- 1. Antecedents..... pàg 4
- 2. Objectius..... pàg 5

II. Desenvolupament del projecte

- 1. Plantejament del problema..... pàg 6
- 2. Pla de treball..... pàg 7
- 3. Entorn de treball..... pàg 7
- 4. Fonaments teòrics..... pàg 8
- 5. Metodologia..... pàg 11

III. Implementació

- 1. Preparació i càrrega de dades..... pàg 18
 - o Dades sobre desplaçaments
 - o Dades sobre la xarxa de transport de Catalunya
- 2. Creació del graf..... pàg 20
 - o Modelització del transport a peu
 - o Modelització del transport ferroviari
- 3. Càlcul de rutes..... pàg 34

IV. Anàlisi dels resultats

- 1. Anàlisi estadístic..... pàg 42
- 2. Diagnosi de casos irregulars: els *outliers* de la xarxa..... pàg 45
- 3. Correcció de la xarxa..... pàg 49

V. Conclusions..... pàg 52

VI. Bibliografia..... pàg 55

VII. Annexos..... pàg 56

I. Introducció

1. Antecedents:

En el marc de la creixent utilització de les tecnologies de la informació i la seva implementació en múltiples àmbits professionals, aquest treball és un exemple més de la rellevància que té avui en dia el fet de dotar la informació de dimensió espacial. Bé sigui en matèria empresarial com en la gestió pública, el valor afegit que aporta conèixer la localització i el comportament sobre el territori d'un fenomen concret, ha assolit una importància cabdal en els últims anys. Conseqüència d'això, n'és l'interès que mostren tant les empreses com les institucions públiques per adquirir professionals que aportin aquest tipus de coneixement, que combina tecnologia i coneixement geogràfic a la vegada.

En el cas particular d'aquest treball, tot parteix d'un encàrrec de l'Institut d'Estadística de Catalunya davant la voluntat d'incloure una vessant d'anàlisi espacial sobre els seus treballs estadístics susceptibles de ser analitzats no només en el sentit més estricte de l'Estadística, sinó també a partir de la seva distribució sobre el territori. Això suposa un plus qualitatiu molt important per a qualsevol institució o empresa que disposi de la capacitat per incloure aquest tipus d'anàlisi.

L'Institut disposa d'un ampli recull de dades que són aptes per dur a terme anàlisis d'aquest tipus. I, en el seu cas, significa un pas diferencial en la seva tasca perquè (a part de poder estudiar la distribució en l'espai dels seus resultats estadístics) servirà –també– per comprovar la qualitat i el grau de confiança que es pot donar a les dades que obtenen amb els seus mètodes de recollida i manipulació d'informació estadística. A més (com es veurà en aquest treball), permetrà valorar el grau d'eficiència d'aquests mètodes i, en qualsevol cas, estudiar la fiabilitat de les dades a l'hora de ser explotades.

Existeix un alt nombre d'institucions i entitats públiques que s'han fet ressò de la importància dels Sistemes d'Informació Geogràfica i que, per tant, disposen ja de departaments especialitzats en aquest sector que treballen i produeixen eines, serveis i productes que aporten un alt valor afegit a la seva tasca que, d'una o altra manera, servirà per contribuir en la millora de l'atenció i el benestar de la societat per a la qual treballen. Per anomenar-ne alguns exemples propers, és sabuda la rellevància dels SIG en organismes com els Bombers de Barcelona o el Port de Barcelona; així com també són presents en la gestió municipal de molts ajuntaments del territori català.

2. Objectius:

- Produir un graf que representi la xarxa de comunicacions en diferents mitjans de transport de l'àrea d'estudi i que sigui capaç de suportar el càlcul de rutes.
- Establir un sistema de càlcul de rutes òptimes per a dades massives.
- Analitzar els resultats estadístics obtinguts.

Objectius específics:

- Obtenció, preparació i manipulació de dades per adaptar-les al sistema i a l'estructura de dades amb la qual treballa el programa de càlcul de rutes emprat.
- Gestionar una base de dades a nivell professional.
- Aprenentatge de funcions de PG *Routing*.
- Aprenentatge de llenguatge *SQL*.
- Producció de resultats estadístics.

II. Desenvolupament del projecte

1. Plantejament del problema

El projecte que es presenta a continuació, es desenvolupa en el marc d'un conveni de pràctiques entre el món acadèmic i l'administració pública. Davant la necessitat de formar nous professionals recent sortits d'un entorn universitari, l'Institut d'Estadística de Catalunya veié oportú incorporar un estudiant en pràctiques per a l'elaboració d'un projecte amb l'objectiu de crear una infraestructura de rutes pròpia per tal de poder analitzar material estadístic que conté una important component espacial.

Inicialment, l'Institut sol·licita la creació de tal infraestructura per l'anàlisi de la distribució dels desplaçaments obligats a Catalunya per motius laborals o d'estudis. Aquest recull estadístic és propietat de l'IDESCAT, i és amb el qual es posarà en funcionament un sistema de càlcul de rutes amb el graf de transport que es desenvoluparà durant aquest projecte.

Per motius pràctics i tenint en compte la complexitat del projecte i el temps disponible durant el conveni de durada de les pràctiques, es veu oportú crear la xarxa de comunicacions (incloent diferents modes de transport) a partir de programari lliure i fonts de dades *open data*. Aquesta decisió es pren per tal d'evitar retards en l'espera de les dades procedents d'altres entitats, la qual cosa limitarà la construcció del graf als mitjans ferroviaris (Metro, FGC, ADIF i TRAM) i per anar a peu.

Des del principi es confia i es fa ús dels continguts oberts al web, una tipologia de continguts que va més enllà de permetre l'ús i difusió lliure d'informació a Internet, sinó que –a més a més– es basa en principis de creació i participació col·lectiva entre usuaris que aporten coneixement i material a una iniciativa cooperativa sense ànim de lucre.

2. Pla de treball

Tasques & Cronograma	S1 Dl - Dv 18 - 22 abril	S2 Dl - Dv 25 - 29 abril	S3 Dl - Dv 2 - 6 maig	S4 Dl - Dv 9 - 13 maig	S5 Dl - Dv 16 - 20 maig	S6 Dl - Dv 23 - 27 maig	S7 Dl - Dv 30 - 3 juny	S8 Dl - Dv 6 - 10 juny	S9 Dl - Dv 13 - 17 juny	Resultat
Creació BD espacial										Definició de la base de dades espacial
Càrrega de dades (origen/destinació + desplaçaments)										BD amb taules definides i dades sobre desplaçaments carregades
Obtenció, exploració i alternatives de <i>Graf viari</i>										Documentació i descarregada d'arxius necessaris per crear un <i>graf</i> des d' <i>open sources</i>
Modelització del transport a peu <i>OpenStreetMaps</i>										Graf de transport a peu
Modelització del transport públic (ferro + metro) <i>Gencat i OSM</i>										Graf de transport públic
Càlcul de rutes pels desplaçaments d'IDESCAT										<i>PgRouting</i> + resultats estadístics
Explotació estadística dels desplaçaments i anàlisi del graf										Anàlisi estadístic, detecció i correcció d'errors
Memòria del treball										Document escrit i presentació

3. Entorn de treball

Programari

- Postgre SQL (extensió espacial: PostGIS i càlcul de rutes: PgRouting)
- Quantum GIS
- Weka Data Mining (The University of Waikato)

Font de les dades:

- Open Street Maps
- Generalitat de Catalunya – www.gencat.cat



El software escollit com a programa gestor de bases de dades és PostgreSQL. Concretament, es treballarà amb PostGIS: l'extensió de PostgreSQL amb capacitats espacials. PgRouting (extensió dins de PostgreSQL) permetrà el càlcul de rutes a partir de funcions i algorismes complexos.

Quantum GIS serà el software client de GIS que s'utilitzarà per visualitzar els avanços del treball mitjançant la càrrega directa des de la base de dades de Postgre SQL. Servirà –també– per documentar el procés gràficament.

Weka Data Mining serà important de cara a la fase d'anàlisi estadística dels resultats (una vegada hagi finalitzat el càlcul de rutes).

4. Fonaments teòrics

El graf

El projecte desenvolupat a l'Institut d'Estadística de Catalunya es fonamenta – principalment– en la teoria de grafs. Un graf (dibuix o imatge, en etimologia grega), és la representació d'una xarxa interconnectada a partir d'un conjunt d'objectes que es classifiquen en arcs i nodes. Els primers, s'encarreguen de connectar els nodes (que equivaldrien a les interseccions de la xarxa). Els dos conjunts –normalment– reben aquests noms per definició. Tot i així, en algunes ocasions és possible veure'ls definits com vèrtexs i arestes, respectivament. De fet, el software amb el qual s'ha treballat per la creació del graf, utilitza tant el mot vèrtex com el de node en diferents fases del procés de creació (tot i que és més usual utilitzar el terme node).

La teoria de grafs

La teoria de grafs té el seu origen en un problema que es plantejà a la ciutat prussiana de Königsberg durant el segle XVIII. *El problema dels set ponts de Königsberg* és el primer precedent del qual es té constància de la resolució d'un problema mitjançant la teoria de grafs, a partir de l'anàlisi de les relacions –espacials o no– que representa un graf. Leonard Euler fou qui resolgué el problema transformant la realitat en un graf l'any 1736.

La formulació del problema a partir d'una abstracció de la realitat, on el territori es converteix en arcs (ponts) i nodes (nuclis de la ciutat) simplifica el plantejament del problema de manera gràfica. I, això, en facilita el seu estudi; sobretot per a problemes majors (on s'estableixen un nombre de relacions molt superior al del problema inicial dels set ponts de Königsberg). Aquesta abstracció permet descriure les connexions que s'analitzen, així com la interacció entre els objectes representats o fins i tot permet la identificació de patrons gràcies a la representació del problema en forma de graf.

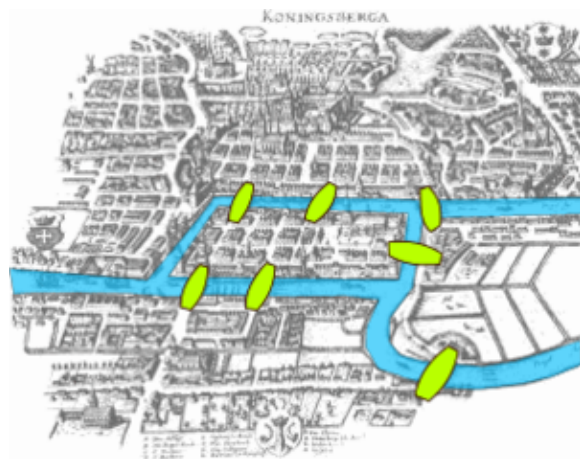


Figura 1. *El problema dels set ponts de Königsberg.*
Font: Viquipèdia.

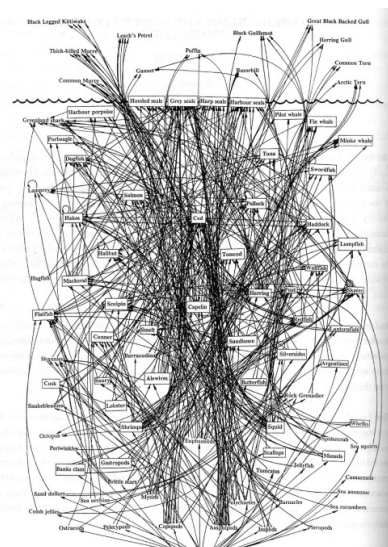


Figura 2. *La complexitat d'una cadena tròfica representada mitjançant un graf.*
Font: *Redes: una breu introducció.*
G. Caldarelli et M. Catanzaro.

La topologia geoespacial

La topologia geoespacial, essent una de les característiques diferencials dels Sistemes d'Informació Geogràfica, permet que un *software* sigui capaç d'estudiar les relacions espacials entre els elements que representen un entorn real gràcies a la classificació dels elements de la realitat en forma de nodes, arcs i polígons. El software en qüestió, estableix aquestes relacions a partir d'un llenguatge i unes regles basades en geometria matemàtica.

Bases de dades espacials

En el cas particular d'aquest estudi, el software que s'utilitzarà com a Sistema de Gestió de Bases de Dades espacials és PostGIS (l'extensió espacial de PostgreSQL). PostgreSQL és un sistema de gestió de bases de dades orientat a objectes. La principal aportació de PostGIS és que permet l'emmagatzematge de la informació geogràfica dels objectes gràcies a l'atribut *Geometria*, compost per codis binaris. Sobre aquests sistemes, cal destacar que –com qualsevol SIG– treballa amb la georeferenciació per tal que l'estructuració de les dades i els càlculs que es realitzin es facin sota un mateix sistema de referència espacial.

La gestió de les bases de dades i la informació emmagatzemada es basa en el llenguatge SQL (*Structured Query Language*), amb el qual es té accés directe a la base de dades mitjançant un mecanisme de consultes. A més, el llenguatge SQL es recolza d'operadors, sentències i clàusules molt diversos que permeten a l'usuari engrandir el ventall de possibilitats per generar consultes SQL complexes. A més, les consultes poden anar acompanyades de funcions que automatitzen processos llargs (compilats en forma de funcions creades a partir de llenguatges de programació), la qual cosa converteix les bases de dades en un atractiu destacat en l'àmbit de les tecnologies de la informació.

Taules i vistes

Quan es treballa amb un elevat volum de dades, és convenient que el conjunt de dades estigui adequadament estructurat. Per facilitar aquesta tasca, les bases de dades permeten l'emmagatzematge de les dades ordenades per taules. També existeixen les vistes. Les primeres, són l'espai principal per guardar la informació amb la qual es pretén treballar. Cada taula està definida per camps i registres. Normalment, una taula farà referència a una entitat del sistema. Les vistes, en canvi, són consultes que generen quelcom semblant a una taula virtual a través de la qual es pot accedir a informació de la base de dades. La seva definició es basa en una consulta concreta (es mostra aquella informació que la consulta ordena sobre una taula preexistent). Les vistes són una eina realment útil per desenvolupar certes tasques d'un treball, que d'altra manera resultarien més costoses o feixugues. Treballant amb vistes, la manipulació i realització de consultes se simplifica. Posteriorment, la informació resultant d'un procés en una vista pot ser transferida de nou a una taula.

Optimització de la cerca

Un altre factor a destacar quan es parla de bases de dades és el de l'optimització dels processos de cerca. Quan es treballa amb bases que superen els milers o milions de registres, l'optimització esdevé una prioritat. I és que, pel bon funcionament del software cal establir –de vegades– uns patrons de cerca que facilitin les tasques de recerca entre volums d'informació enormes.

El motor de cerca de les bases de dades acostuma a basar-se en els *B-tree* (*Binary search tree*) és un mètode d'estructuració de les dades creat per optimitzar el motor de cerca d'una base de dades. Establint una jerarquia, permet fer cerques concretes (obviant aquells espais de la base de dades on –de partida– se sap que no s'ha de buscar). Del contrari, per a cada consulta s'hauria de fer una inspecció al total de la base de dades, la qual cosa resultaria insostenible.

El llenguatge de consulta explícita què vol buscar, però no la manera en què s'ha de fer la cerca. Per tant, una manera de millorar l'eficiència de les cerques és la creació d'índexs que estructurin els registres de les taules de tal manera que s'accelera el procés de cerca. Que una consulta s'executi mitjançant índexs o no, es pot saber afegint una sentència d'*EXPLAIN*, resultat de la qual indicarà com està funcionant el motor de cerca del sistema.

5. Metodologia

5.1 Obtenció i preparació de les dades

5.1.1 Creació de la base de dades espacial

El programa gestor de bases de dades PostgreSQL, d'inici necessita activar l'extensió que li dóna caràcter espacial:

```
create extension postgis
select postgis_full_version()
```

A continuació, es crea la base de dades que allotjarà tant les taules com les vistes amb les que es treballarà durant el projecte.

5.1.2 Càrrega de les dades

En aquest espai s'hi emmagatzemarà tant el material estadístic com geogràfic necessari per dur a terme el projecte. PostGIS disposa d'un *wizard* específic per importar i exportar *shapefiles* i taules de dades, els quals s'inclouen –directament– a la base de dades en forma de taules on s'hi emmagatzema la informació de cada objecte.

[PostGIS 2.0 Shapefile and DBF Loader Exporter]



Per una banda, s'han d'importar les dades sobre desplaçaments de l'IDESCAT a la base de dades. Serà la part corresponent al material estadístic d'aquest treball.

D'altra banda, s'ha d'importar –també– el material que contextualitza el treball geogràficament i permet la creació del graf.

5.1.3 Preparació de les dades

- **Dades sobre desplaçaments obligats a Catalunya**

El material estadístic amb el qual s'ha de treballar està recollit en un full de càlcul amb 378.564 registres de partida que corresponen a **desplaçaments obligats**¹ a Catalunya, amb la informació referent a l'origen, el destí, el tipus de mitjà de transport, el temps mitjà de desplaçament, etc. Com a punt d'origen es prenen les coordenades **GPS**² del domicili de cada individu. Com a destinació es disposa del codi postal del punt de destinació, així que tan sols es pot fer una aproximació estructurant el territori en funció dels codis postals i extraient el centroides per a cada polígon resultant. Així, el centroides del codi postal serà el punt de destí per a cada desplaçament. Dit això, es preveu que els resultats del procés de càlcul de rutes no podrà ser més que una aproximació a causa de la manca d'exactitud que aporten les dades; per tant, en

la fase d'anàlisi dels resultats, es farà difícil estudiar empíricament com es comporta la distribució dels desplaçaments com a tal.

- **Dades sobre la xarxa de transports de Catalunya (*OpenStreetMap* i dades obertes de la Generalitat de Catalunya)**

Per a aquest cas d'estudi, s'opta per *OpenStreetMap* com a font de les dades que permetran la creació del graf. Emparat per l'*OpenStreetMap Foundation*, aquesta iniciativa permet la visualització i descàrrega de material cartogràfic a nivell mundial. Precisament, és d'aquí d'on s'extreu la informació de base amb la qual es treballarà durant el projecte. Des del seu lloc web (<https://www.openstreetmap.org>), el servidor de mapes permet la descàrrega de cartografia digital en diferents formats. En aquest cas, s'ha descarregat un *dataset* comprimit que correspon a l'**Àrea Metropolitana de Barcelona**³. L'AMB servirà com a àmbit d'estudi en aquesta primera fase del projecte. Al descomprimir-se, el *shapefile* amb un conjunt de dades vectorial s'estructura en punts, línies i polígons. A continuació, caldrà fer la importació d'aquest conjunt de dades a la base de dades amb la qual es treballarà durant tot el projecte. Essencialment, es treballarà amb els conjunts de línies i punts. Les línies equivalen als segments de la xarxa de comunicacions i els punts a les estacions de la xarxa de transport públic.

A més a més, es recorre al web de la Generalitat de Catalunya (*Dades obertes*) d'on s'obté un altre *dataset* comprimit que correspon a la xarxa ferroviària de Catalunya. Dins de l'arxiu comprimit (4,8 MB), s'hi troben principalment *shapefiles* individualitzats per tipus de mitjà de transport públic (línies de Metro, FGC, TRAM, ADIF...) i les respectives estacions en format de punts.

¹**Desplaçaments obligats:** Són aquells desplaçaments que es realitzen en el dia a dia sobre el territori per motius laborals o d'estudis. Les dades provenen del cens de població de 2011 (INE amb la col·laboració d'IDESCAT).

²**Coordenades GPS:** En l'elaboració de la mostra (a partir d'enquestes en domicilis de tot Catalunya), es prengueren les coordenades GPS dels domicilis enquestats.

³**Àrea Metropolitana de Barcelona:** Es pren com a referència l'entitat supramunicipal que engloba un conjunt de 36 municipis, amb una superfície de 636 km² i més de tres milions d'habitants (vuitena aglomeració metropolitana d'Europa pel que fa a població). Es calcula que es realitzen més de deu milions de desplaçaments a l'any a l'AMB, amb una mitjana de 3,9 desplaçaments per persona i dia segons dades de l'AMB (2014).

Mapa de localització de l'Àrea Metropolitana de Barcelona



Figura 3. Mapa de localització de l'Àrea Metropolitana de Barcelona. Font: Elaboració pròpia.

Concretament, s'ha escollit un retall per a l'àrea d'estudi que sobrepassa lleugerament els límits definits anteriorment per l'AMB:

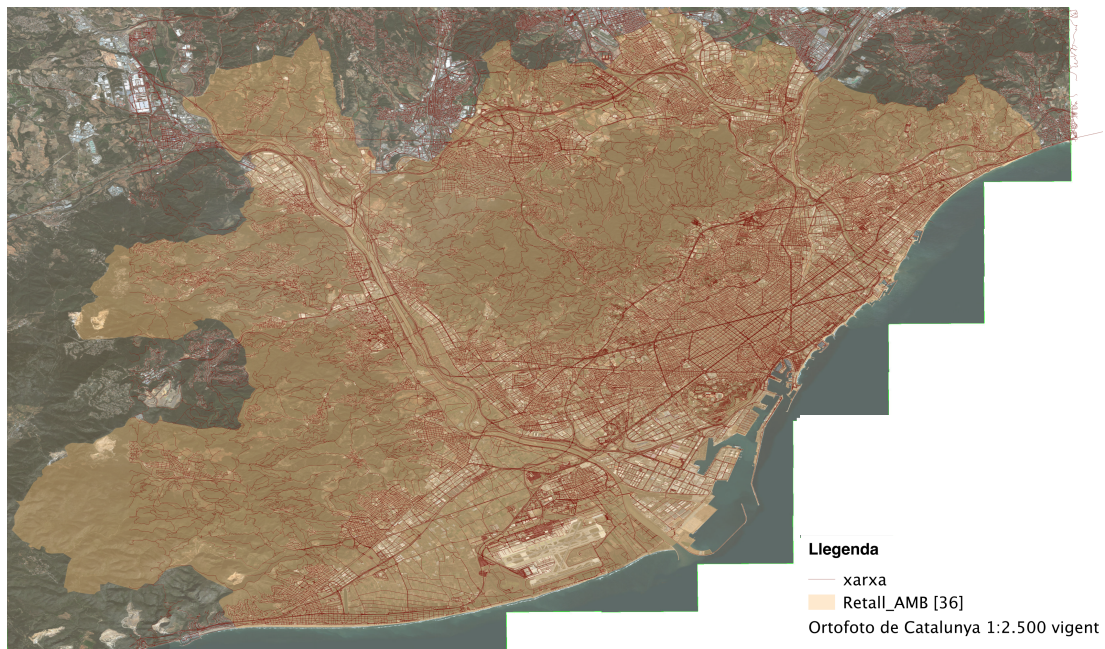


Figura 4. Delimitació de l'àrea d'estudi. Font: Elaboració pròpia.

5.2 Creació del graf

Una vegada es disposa del conjunt de dades necessàries per crear la xarxa de transport que es planteja a l'inici del projecte, caldrà dur a terme una exploració exhaustiva i un tractament de les dades meticulós per tal d'adaptar-les a l'entorn de treball de què es disposa. Aquest procés és el que es presenta a continuació: una modelització del graf dividida segons modes de transport i finalment unificada per poder-hi dur a terme el càlcul de rutes.

A grans trets, el procés que es durà a terme es pot definir per la creació de diversos grafos (cada un serà apte per ser utilitzat amb un determinat tipus de transport) que –finalment– seran unificats per formar una xarxa que representi, tant com sigui possible, la xarxa real de carrers i les línies del transport ferroviari existents a l'AMB.

Els procediments, en cada cas, s'estructuren en base al següent ordre:

- 1- Inspecció de la xarxa inicial.
- 2- Operacions a realitzar:
 - Existeix informació redundant que cal excloure.
 - Existeixen excepcions que necessiten un tractament especial.
 - Creació d'enllaços/connexions que no vénen donades d'inici.
 - Càlcul d'atributs que les funcions de *Routing* necessiten per ser executades.
- 3- Segmentació de la xarxa.
- 4- Creació d'una topologia única de referència.

5.3 Càlcul de rutes

D'entre les funcions de *Pg_Routing* més utilitzades per a càlcul de rutes trobem la *Pgr_Dijkstra* i *Pgr_Astar*.

Pgr_Dijkstra: L'algoritme amb què es basa la funció *pgr_Dijkstra* fou descrit l'any 1959 per primera vegada pel físic i científic de la computació holandès Edsger Dijkstra. L'algoritme es basa en la recerca del camí més curt donat un vèrtex d'origen, des del qual s'avança node a node buscant els vèrtexs més propers (la qual cosa ve determinada pel cost de les arestes que uneixen cada node). Per tant, el funcionament consisteix en una exploració de tots els camins possibles i s'acaba escollint el de menor cost (o més curt).

```
SELECT pgr_dijkstra('SELECT gid::int as id, source::int,
target::int,cost::float8 as cost FROM routing_simple',1, 4,
false, false);
```

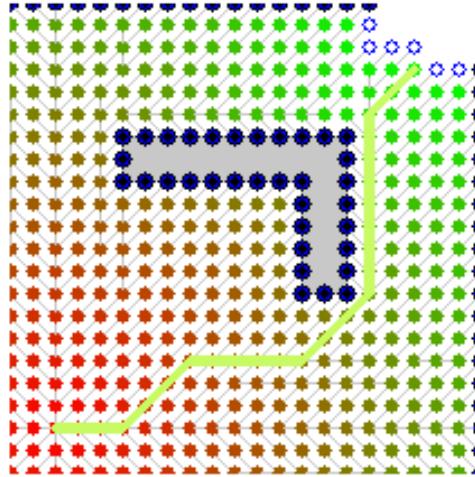


Figura 5. Representació gràfica del funcionament de la funció de Dijkstra.
Font: GitHub.com.

Pgr_Astar: L'algoritme en què es basa la funció *pgr_Astar* té el seu origen a l'any 1968 en un intent per donar solució a aquells algoritmes del mateix tipus que no eren suficientment eficients en el càlcul del camí més curt en un graf. L'element diferencial d'aquest algoritme es basa en el fet que durant el procés d'identificació del node més proper (donat un punt d'origen), la funció selecciona el vèrtex més proper a la destinació. Aquest fet és possible gràcies a la incorporació de variables com les coordenades geogràfiques de l'àrea d'estudi. Per tant, cal introduir una sèrie de camps addicionals que permetran optimitzar el temps d'execució de la funció, la qual cosa resulta realment interessant en casos en que es treballi amb un gran volum de dades.

```
SELECT pgr_astar('SELECT gid::int as id, source::int,
target::int,cost::float8 as cost, x1, x2, y1, y2 FROM
routing_simple',1, 4, false, false);
```

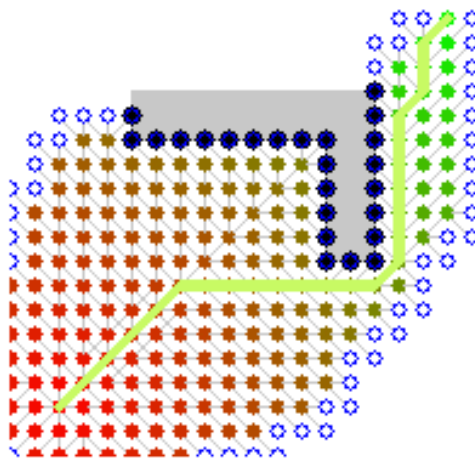


Figura 6. Funcionament de la funció Astar.
Font: GitHub.com.

Per a aquest cas d'estudi, s'opta per la segona opció ja que (com s'acaba d'explicar) optimitza el procés de cerca de la ruta més curta mitjançant la incorporació de les coordenades geogràfiques.

Les funcions de càlcul de rutes es basen en la recerca del camí òptim entre un punt d'origen i un de destí. Aquests dos punts (sovint anomenats *source* i *target*) poden ser qualsevol punt de la xarxa que es genera a partir de la segmentació i creació de la topologia. Aquests punts equivalen als nodes d'enllaç dels arcs que representen les vies (PostGIS, en el procés de creació de la topologia, els anomena *vèrtexs* i els emmagatzema en una taula amb identificadors únics per poder executar les funcions de *Routing*).

En relació als punts d'origen i destí, és important anticipar que (en vista que els punts d'origen i destí de les dades sobre desplaçaments no es corresponen amb punts existents del graf) s'ha optat per realitzar una operació que segueixi un procés automatitzat a la vegada que s'executa la funció de *Routing* per tal que la funció pugui prendre un node o vèrtex d'origen i de destí de referència. En l'apartat de càlcul de rutes s'explicarà en detall aquest procés.

Finalment, s'adjunta un esquema de la metodologia que seguirà el procés de càlcul de rutes massiu:

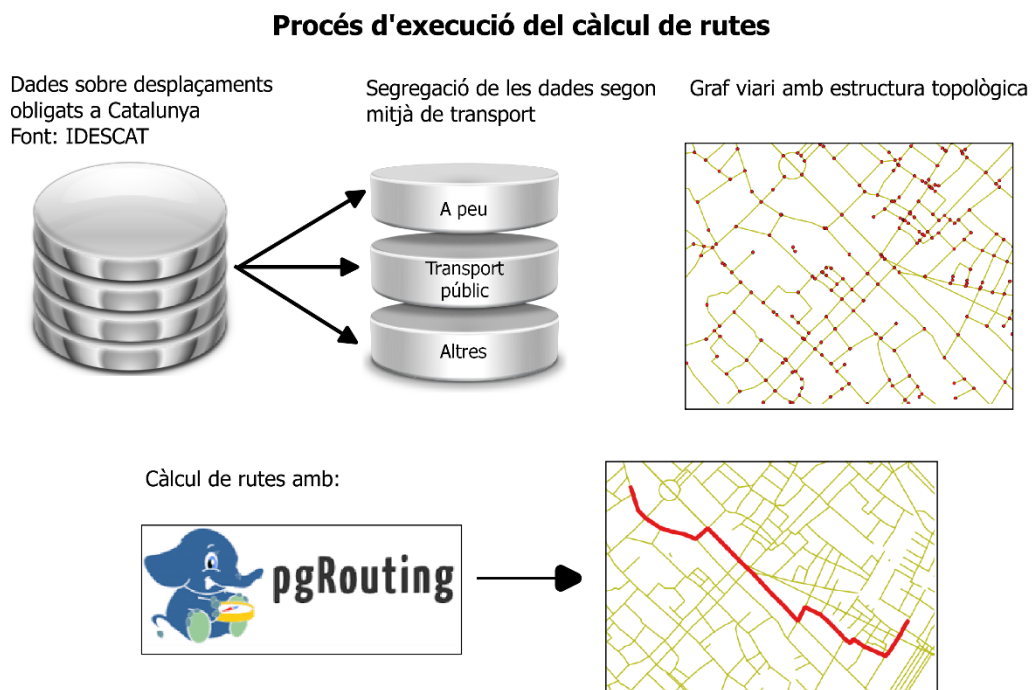


Figura 7. Procés d'execució del càlcul de rutes. Font: Elaboració pròpia.

Com s'observa, es farà una distinció per grups (segons el mode de transport, o la combinació de modes utilitzats en cada cas). Aquesta segregació es farà a partir de la creació de vistes dins la base de dades on s'hi agruparan els individus segons el mode de transport utilitzat. Llavors, s'executarà el càlcul de rutes per a cada grup i –posteriorment– se'n faran les modificacions necessàries per ordenar els resultats amb els atributs i agregats que es desitgin per procedir amb l'anàlisi dels resultats.

III. Implementació

1. Preparació i càrrega de dades

- **Dades sobre desplaçaments obligats a Catalunya**

Per tal d'importar els desplaçaments a la base de dades espacial, és necessari un procés de tractament de les dades per adequar-les al cas d'estudi. D'entrada, cal definir el marc geogràfic en el qual s'han de localitzar les dades. Per obtenir-lo, s'escull la base del cadastre de parcel·les agrícoles (SIGPAC), d'on s'extreuen les àrees urbanes de Catalunya (estructurades per codi de província i municipi), per creuar-les amb la base de l'Institut Geogràfic Nacional (IGN) d'on s'extreu –en aquest cas– la informació referent als codis postals de cada municipi. Amb el creuament d'aquestes dues bases (mitjançant una operació que intersecta la geometria i la informació dels polígons d'ambdues bases) s'obté la base territorial d'aquest treball: la que divideix les àrees urbanes de Catalunya segons el codi postal per a cada àrea urbana. A la imatge, s'observa seleccionat de color vermell el municipi de Barcelona (definit com una única àrea urbana) que, a la vegada es troba subdividit en polígons definits pel codi postal que representen. En altres casos, coincideix un sol codi postal amb una entitat urbana que equival a un municipi sencer.

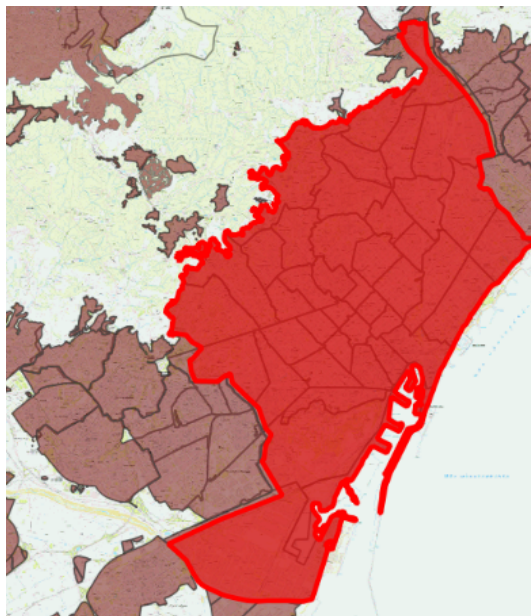


Figura 8. Base cartogràfica del treball. Creuament d'àrees urbanes (SIGPAC) i polígons segons els codis postals (IGN). Font: Elaboració pròpia.

A continuació, es pot procedir amb la importació dels desplaçaments, una vegada es disposa del context territorial definit. La importació s'efectua mitjançant la creació d'una nova taula dins la base de dades espacial. Com que l'estudi es limita a l'Àrea Metropolitana de Barcelona, primer

cal destriar aquells codis postals que pertanyen a les altres províncies de Catalunya. Per tant, es trien –només– aquells que comencen amb el codi de província “08” (que equival a la província de Barcelona) i es definirà el sistema de referència internacional corresponent a l'àrea geogràfica de Catalunya (SRID 25831). Una vegada s'ha creat la taula, es genera la capa de punts d'origen a partir de les coordenades GPS amb la següent sentència SQL:

```
update punts_origen set
geom=st_transform(ST_SetSRID(ST_MakePoint(gps_x,
gps_y),25831),25831)
```

El resultat de destriar aquells desplaçaments per a l'àrea d'estudi redueix la mostra total a 130.641 registres. Arribats a aquest punt, el sistema disposa dels punts d'origen i destí – més o menys concrets– per a l'anàlisi dels desplaçaments.

A continuació, vindrà el torn de crear la infraestructura que definirà la xarxa. Se'n farà una modelització per crear un graf que suporti el càlcul de rutes massiu i permeti la unió entre qualsevol punt d'origen amb el respectiu punt de destí, segons estableixin les dades sobre desplaçaments d'IDESCAT.

- **Dades sobre la xarxa de transports de Catalunya**

En aquest cas, la inspecció de les dades es duu a terme una vegada s'han importat a la base de dades mitjançant *PostGIS 2.0 Shapefile and DBF Loader Exporter*.

Una vegada importades, es procedeix a treballar dins de PostGIS. Primer, caldrà fer una inspecció detinguda de les dades procedents d'*OpenStreetMap* per veure si es corresponen amb la realitat i si són aptes per ser utilitzades en aquest projecte. En qualsevol cas, es veu imprescindible fer un tractament per adequar-les a les necessitats d'ús d'aquest treball, ja que – originalment– en el cas de les línies apareixen com una xarxa de línies interconnectades sense una estructura que permeti començar amb la modelització del graf. Aleshores, la preparació i tractament de les dades equival al procés de modelització del graf, el qual s'ha estructurat de manera que es pugui diferenciar un procés de modelització diferent per a cada tipus de transport amb el qual podrà ser explotat el graf. Tanmateix, el resultat final serà una única xarxa, tot i que haurà hagut de passar per processos diferents de modelització abans de ser unificada.

2. Creació del graf

2.1 Modelització del transport a peu

D'inici, s'escull adequar la xarxa per a la creació del graf que servirà per al càlcul de rutes de transport a peu. En aquest cas, la complexitat del procés és menor donades les poques restriccions que comporten els usuaris d'aquest graf: els vianants (també es consideren els ciclistes com a usuaris d'aquest graf) no estan subjectes a tantes restriccions com ho estan els vehicles de motor. Els carrers no necessiten tenir un sentit de circulació definit i les interseccions són – generalment– transitables en qualsevol direcció per aquest tipus d'usuari. A continuació s'exposa el procés que s'ha dut a terme en aquest primer cas de modelització.

Les dades inicials procedents d'*OpenStreetMap* (OSM) són bàsicament un conjunt de línies interconnectades. Un dels primers problemes a resoldre que s'identifiquen és que l'arxiu original està definit com una col·lecció de geometries *MultilineString*. Caldrà transformar-les a *LineString* a l'hora de crear el graf (si es pretén que el software executi les funcions correctament); contràriament, ho faria entenent que les línies són una cadena de trams de línia unificats amb aquest format “*multi*”. És a dir, que un carrer qualsevol estaria definit com un únic segment, sense tenir en compte els trams d'aquest carrer que es defineixen per les interseccions que suposen el creuament d'aquest carrer amb d'altres vies. Per tant, és necessària una transformació del tipus de línies per tal de tenir una xarxa adequadament segmentada en trams (que vindran donats segons les interseccions de la xarxa real de carrers):

```
ALTER TABLE my_table
ALTER COLUMN geom TYPE geometry(linestring,3857) USING
ST_GeometryN(geom, 1);
```

A continuació, arrenca el procés per crear l'estructura topològica del graf. És sabut que no totes les línies han de convergir en un node concret, tot i que la majoria sí que ho han de fer. Hi ha excepcions que deriven de l'existència de casos, com els túnels o els ponts, els quals no poden ser considerats de la mateixa manera que les vies ubicades a la calçada. Per tant, s'escull l'eina *Pgr_NodeNetwork*, una funció de *PgRouting* que serveix per segmentar la xarxa a partir de les interseccions de la mateixa. La funció descompon les línies en segments únics amb identificador propi, determinats pels nodes inicials i finals de cada tram. Aquest pas és fonamental si es pretén tenir una xarxa de transport que actuï com la xarxa real d'un territori concret.

Abans d'executar dita funció, cal fer un pas previ per excloure aquells registres que s'han considerat invàlids. El motiu d'això rau en l'existència d'excepcions que s'entenen a partir de la classificació de la xarxa en diferents alçades. A les dades d'origen procedents d'*OSM*, aquesta

informació es troba en el camp anomenat “z-order”. En aquest cas, s’han d’excloure els camps “*parking_aisle*”, “*tunnel*”, “*bridge*” i “*emergency_access*”:

```
select * from graf_bcn_25831 where
(trim(service) <> 'parking_aisle' or service is null ) and
(trim(tunnel) <> 'yes' or tunnel is null) and
(trim(bridge) <> 'yes' or bridge is null) and
(trim(service) <> 'emergency_access' or service is null)
```

I, a continuació, s’executa la funció *Pgr_NodeNetwork* amb els camps vàlids:

```
pgr_nodenetwork(text edge_table, float8 tolerance, text
id='id', text the_geom='the_geom',text table_ending='noded')
```

A la imatge s’observa el resultat del procés. Les línies representades només en groc corresponen a aquelles línies excloses en la creació dels nodes per tal d’evitar que aquestes creuessin amb la resta de la xarxa (se sap que no ho han de fer perquè corresponen a vies soterrades). A la imatge, els túnels de la Ronda litoral en el seu pas per la Barceloneta i l’Estació de França. A més, els vianants tenen restringit l’accés a aquest tipus de via.

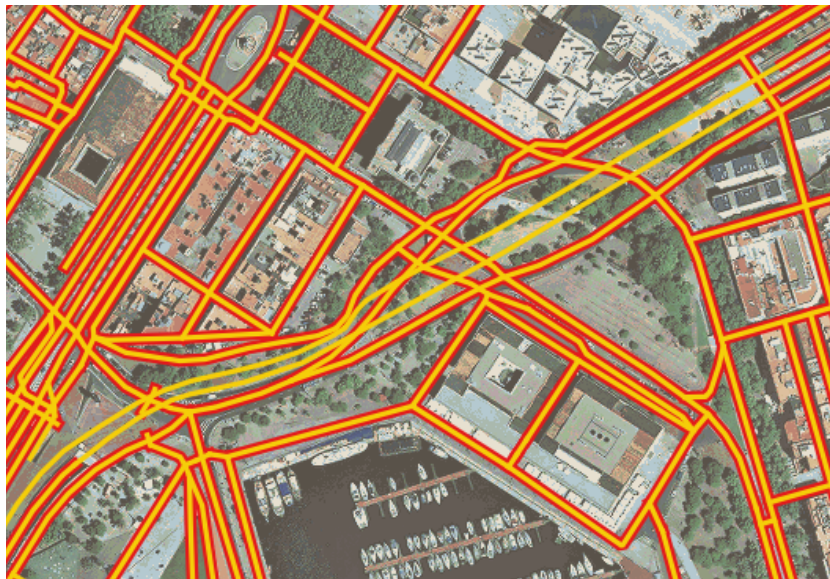


Figura 9. Resultat de segmentar la xarxa exclouent els casos invàlids. Font: Elaboració pròpia.

Tot seguit, ve el moment de crear la topologia de la xarxa viària. Per crear-la, es farà a partir d’una taula nova, on s’introduiran les dades resultants del procés de segmentació de la xarxa (*Pgr_NodeNetwork*) a més d’aquells registres exclosos anteriorment (ara que ja no són un inconvenient en el procés de segmentació).

Per establir la xarxa completa (unint els casos exclosos anteriorment), cal fer el pas d’unió entre taules, que en el llenguatge SQL que s’utilitza en PostGIS, equival a una sentència d’ *insert into*:

```

insert into routing_network
select
gid,gid,0,source,target,geom,z_order,highway,name,junction,oneway,ref,service,surface,tunnel,width,tags      from  graf_bcn_25831
where
trim(service) = 'parking_aisle' or
trim(tunnel) = 'yes' or
trim(bridge) = 'yes' or
trim(service) = 'emergency_access'

```

Arribats a aquest punt, es procedeix a la creació de la topologia que permetrà treballar amb la geometria dels elements, fer càlculs i executar funcions específiques de rutes:

```

select  pgr_createTopology ('routing_network', 0.001, 'geom',
'best_id', 'source', 'target')

```

La funció retorna els registres corresponents als segments de la xarxa amb la geometria de cada segment i uns camps (*source* i *target*) que corresponen als nodes ubicats en les interseccions. S'estableix una classificació automàtica amb identificador únic per a tots els *source* i *target* existents a la xarxa.

A més, la funció crea una taula de vèrtexs, els que s'observen en forma de punts a la següent imatge. Aquests vèrtexs equivalen als punts d'inici i destí per a cada tram; per tant, són idèntics que els nodes (esmentats anteriorment). Senzillament, són creats perquè *PgRouting* funcioni així. Un cop fet aquest pas, la xarxa per a transport a peu està quasi llesta per executar funcions de rutes, ja que ha estat adequadament tractada per que, en el moment d'executar el càlcul de rutes, la funció tingui un comportament com el que tindria una persona en decidir la seva ruta en la realitat.

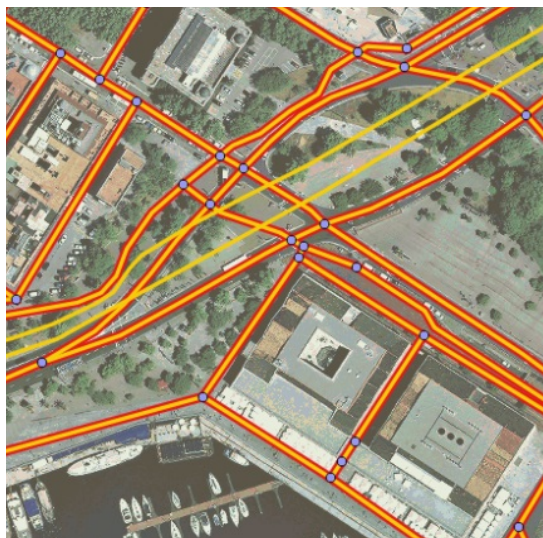


Figura 10. Resultat del procés de creació de la topologia. Font: Elaboració pròpia.

Manca incloure aquells camps que les funcions de ruta exigeixen per ser executades segons especifica la documentació dels seus creadors. Per a les funcions de ruta més utilitzades, és necessari incloure camps com les coordenades x1, x2, y1, y2 i la longitud (o *shape length*) de cada segment de la xarxa.

Càlcul de les coordenades:

```
alter table routing_network add column x1 double precision;
alter table routing_network add column y1 double precision;
alter table routing_network add column x2 double precision;
alter table routing_network add column y2 double precision;
update routing_network set x1 = ST_x(ST_PointN(geom,1));
update routing_network set y1 = ST_y(ST_PointN(geom,1));
update routing_network set x2 = ST_x(ST_PointN(geom,
ST_NumPoints(geom)));
update routing_network set y2 = ST_y(ST_PointN(geom,
ST_NumPoints(geom)));
```

Càlcul de la longitud o *Shape Length*:

```
alter table routing_network add column shape_length double
precision;
update routing_network set shape_length = st_length(geom)
```

El *shape length* representa el cost que, inicialment, es correspon únicament a la distància a recórrer per l'usuari. Més endavant, es pot modificar el cost perquè hi entrin en joc altres variables com pot ser la velocitat. Llavors, la variable cost esdevindrà la variable temps si s'estableix la relació $\text{Temps} = \text{Distància} / \text{Velocitat}$.

A continuació es mostra el resultat d'una primera comprovació del graf per a transport a peu. Amunt, la representació gràfica i, a baix, la representació del procés de càlcul dins la base de dades de PostGIS. El resultat en forma de taula s'expressa de manera seqüencial, on cada registre correspon a l'enllaç d'un vèrtex amb el següent fins a arribar al vèrtex de destinació final. El cost correspon a la distància (en metres) per a cada tram.

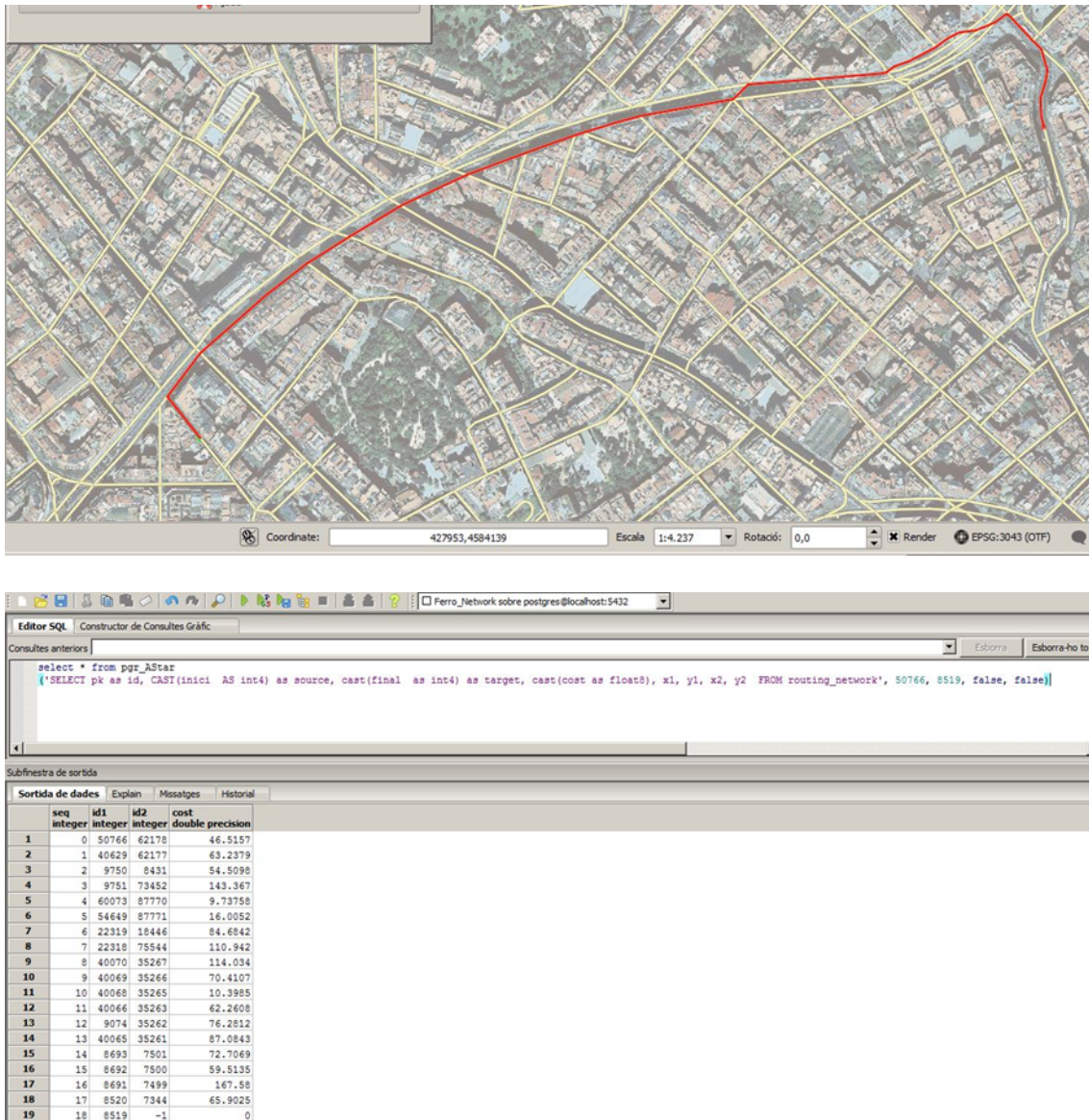


Figura 9. Resultat de la primera comprovació de càlcul de rutes per a desplaçaments a peu entre dos punts de la xarxa. Font: Elaboració pròpia.

2.2 Modelització del transport ferroviari

Per confeccionar la xarxa del graf corresponent als desplaçaments efectuats amb transport públic es decideix utilitzar una font que prové del web de dades obertes de la Generalitat – Departament de Territori i Sostenibilitat (en format *Shapefile*). Hi ha un *dataset* disponible que correspon a la xarxa del transport ferroviari. Conté les diferents línies de ferrocarril, de metro i de tramvia (segmentades per trams entre parades del servei) i les estacions en forma de punts i estructurades en base a uns identificadors únics. Amb aquest conjunt de dades, caldrà fer un tractament de modelització particular abans de poder unir aquest graf amb el qual ja s'ha creat per a desplaçaments a peu o en bicicleta.

En primer lloc, interessa establir nexes entre els diferents tipus de transport ferroviari que s'inclouen en aquest nou graf. Amb altres paraules, s'han de crear segments que equivaldran als enllaços entre diferents mitjans de transport públic (que majoritàriament corresponen a vies soterrades). Més endavant, es podran establir les connexions entre la xarxa de transport ferroviari amb la xarxa de carrers exterior per tal de poder fer el càlcul de rutes d'aquells usuaris que utilitzin la combinació de diversos mitjans de transport per desplaçar-se.

2.3 Creació dels enllaços de Metro

Inicialment, s'estableixen les connexions entre diferents línies de Metro. El *shapefile* de la xarxa de Metro s'estructura en segments de la xarxa (un total de 176). Per una banda, conté informació relativa a la línia i al nom de les estacions d'origen i destí donat un tram concret de la xarxa. D'altra banda, conté una capa de punts que representa les estacions, però no tenen cap clau primària. D'aquesta manera, la solució passa per buscar la geometria dels enllaços en altres fonts d'informació o bé per establir la clau primària d'una estació a partir del seu nom textual.

Gràcies a la informació dels noms de les estacions, es pot establir una relació que indiqui aquells casos on es dona un enllaç entre estacions (suposant que aquells trams de línia que comparteixen nom d'origen o destí corresponen a enllaços de la xarxa). D'aquí, es crea una taula que s'anomena "transbords" on romandran tots els enllaços possibles entre les diferents línies de Metro (22 enllaços, en total). Cada enllaç estarà definit per la unió a partir de segments que connecten l'estació d'origen amb la de destí i –aquesta línia– estarà definida a la vegada per una geometria única creada mitjançant l'operació que proporciona PostGIS: *ST_MakeLine*.

La taula de transbords s'ha creat gràcies a una operador de concatenació que té en compte la repetició dels noms de les estacions on s'estableixen enllaços reals de la xarxa de Metro.

102	599	Universitat-Passeig de Gràcia	Metro L2
26	523	Universitat-Catalunya	Metro L1

Figura 11. Detall de com ve donada la informació sobre les línies de Metro. Font: www.gencat.cat.

A continuació, es procedeix a introduir els transbords a la xarxa de Metro completa amb la següent sentència SQL (que inclou, també, la operació de concatenació utilitzada). *Srv1* i *Srv2* són dos camps que contenen el nom de les estacions de Metro implicades en el servei d'un segment donat de la xarxa:

```
insert into metro_sline (gid,nom,supra_serv,shape_length,geom)
select gid,nom,(srv1 || ' ' || srv2) ,st_length(geom),geom from
transbord
```



Figura 12. Resultat de la creació dels enllaços de Metro. Font: Elaboració pròpia.

2.4 Creació dels enllaços ferroviaris

El procés explicat anteriorment es farà extensiu a tota la xarxa de transport ferroviari. Fins ara només s'han creat els enllaços entre diferents línies de Metro; per tant, toca fer el mateix amb les connexions entre Ferrocarrils de la Generalitat (FGC), xarxa de trens ADIF i TRAM. En aquest cas, les dades no vénen preparades per fer-ho de la mateixa manera que en el cas del Metro (ja que no existeix cap camp que indiqui on es donen aquests enllaços), sinó que s'ha de fer manualment a partir de la recerca dels enllaços que s'estableixen entre aquests tres tipus de mitjans de transport i la xarxa de Metro. Per aconseguir-ho, s'opta per utilitzar *QGIS* per visualitzar les estacions i els codis que les identifiquen (ni les dades dels *shapefile* ni els webs de cada operador faciliten la informació pel conjunt dels modes de transport inclosos en aquest treball). Finalment, es crearà una taula amb *Excel* amb els enllaços corresponents a la xarxa de transport públic per a l'Àrea Metropolitana de Barcelona.

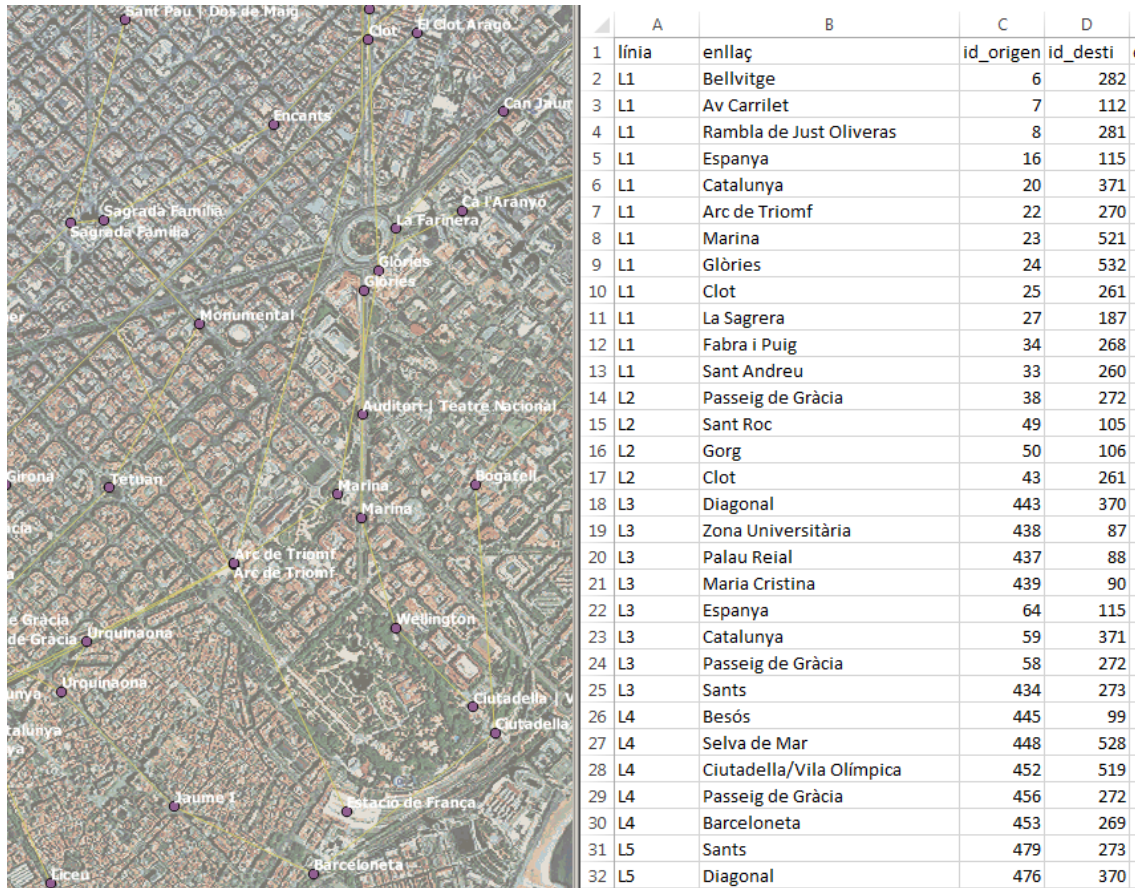


Figura 13. Classificació de la resta d'estacions de la xarxa ferroviària. Font: Elaboració pròpia.

Un cop s'ha elaborat la taula, s'importa a la base de dades com a arxiu .csv i es crearan les línies (*ST_MakeLine*, a partir dels identificadors d'estacions) que definiran els enllaços de la xarxa de transport públic real. Tot seguit, es procedeix a unir aquesta taula amb la de transbords de Metro, que ja existia.

2.5 Creació dels enllaços amb la xarxa de carrers

A part de la creació dels enllaços entre els diferents operadors del transport ferroviari, cal enllaçar-les –també– amb la xarxa de transport a peu; és a dir, amb les vies corresponents als carrers. Aquest pas és imprescindible ja que qualsevol individu que parteixi del seu domicili de residència no pot iniciar el seu recorregut des de cap línia de transport públic (directament) sense abans haver pres un carrer per accedir-hi. Per tant, s'hauran de crear enllaços entre les dues grans xarxes que confeccionen el graf complet per aquest cas d'estudi. Per fer-ho, es segueix el mateix mètode que el que s'ha utilitzat per crear els enllaços entre estacions de la xarxa ferroviària. Mitjançant la creació de línies que busquen el punt més proper de la xarxa de transport a peu (PostGIS: *ST_ShortestLine*) –donat un punt d'origen corresponent a una estació de Metro, ferrocarril o Tram– s'obté la interconnexió dels dos grafs. Exemple del query utilitzat:

```

SELECT *
      st_shortestline(estacions.geom, routing_network.geom) AS
segment
FROM estacions,
      routing_network
WHERE estacions.id_esttren = 443::numeric
ORDER BY (st_distance(estacions.geom, routing_network.geom));

```

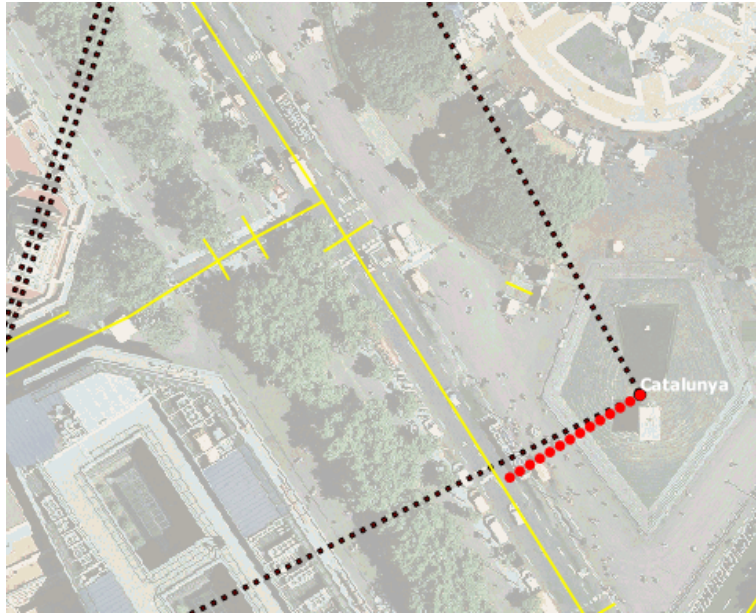


Figura 14. Resultat de la creació dels enllaços entre la xarxa de transport a peu i la xarxa ferroviària. Font: Elaboració pròpia.

2.6 Delimitació de l'àrea d'estudi

Un altre procediment a realitzar es correspon amb la definició dels límits de l'àrea d'estudi (PostGIS: *ST_MakeEnvelope* amb les coordenades $x1,y1, x2, y2$) per tal d'evitar que la xarxa estableixi connexions amb estacions i vies perifèriques que es troben a fora d'aquesta. Es prenen les coordenades que delimiten l'Àrea Metropolitana de Barcelona com a marc de referència:

```

Select * from Routing_Network where geom &&
ST_MakeEnvelope(407497.31, 4566938.00, 442755.97, 4598654.00,
25831);

```

2.7 Assignació de tipologia i velocitats

A continuació caldrà definir el camp *Velocitat*, per fer el càlcul de rutes el més real possible.

Per definir-les, primer serà necessari distingir entre els diferents tipus de transport que s'inclouen en l'estudi; quedaran definits prèviament en un camp que s'anomena *Tipologia* (on es poden distingir sis tipus diferents de transport). El resultat de la tipologia i les respectives velocitats queda estructurat de la següent manera:

Mitjà de transport	Tipologia	Velocitat
A peu	1	1,1 m/s
Metro	2	7,2 m/s
ADIF	3	11,2 m/s
Tramvia	4	5 m/s
FGC	5	9,8 m/s
Bus	6	3,2 m/s

Figura 15. Taula de velocitats mitjanes segons tipologia de transport. Font de les dades: TMB, TRAM i elaboració pròpia.

S'introduiran aquests atributs a la xarxa completa del graf a partir d'operacions d'*update* per a cada cas. En el cas dels Ferrocarrils de la Generalitat de Catalunya:

```
Update routing_network set velocitat = 9.8 where tipologia = 5
```

2.8 Unió dels grafes

Finalment, es procedeix a la unió de tots els elements que participen en la creació del graf. Fins ara, estan separats segons el tipus de transport que representen i caldrà unir-los en un taula definitiva que s'anomenarà *Xarxa*. La xarxa amb els diferents mitjans de transport ferroviari s'unirà amb la taula corresponent als enllaços entre estacions, així com els enllaços entre estacions i carrers i –finalment– amb la xarxa de carrers. Tot i la unió, cada conjunt d'elements mantindrà les característiques pròpies amb les velocitats i la tipologia per a cada cas; gràcies a la prèvia assignació dels camps que així ho designen. Exemple de la unió:

SELECT v_enllasos_altres.nom,		SELECT <u>v_enllasos_metro_carrer.nom,</u>
v_enllasos_altres.srv1,		v_enllasos_metro_carrer.srv1,
v_enllasos_altres.srv2,	UNION ALL	v_enllasos_metro_carrer.srv2,
v_enllasos_altres.geom		<u>v_enllasos_metro_carrer.geom</u>
FROM v_enllasos_altres		FROM <u>v_enllasos_metro_carrer;</u>

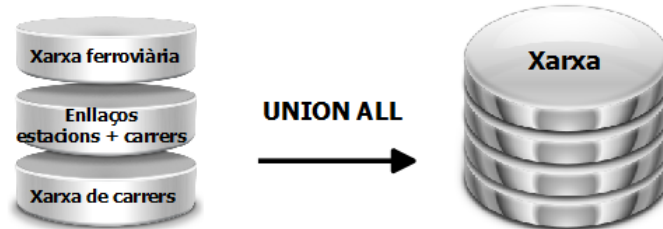


Figura 16. Representació gràfica de la unió de les diferents xarxes.
Font: Elaboració pròpia.

2.9 Creació de la topologia

Seguint els mateixos passos que en el cas de la modelització del transport a peu, és procedeix amb la creació de la topologia per la xarxa completa, ja que és un pas fonamental en la creació de qualsevol graf. S'haurà de refer el procés que inclou les operacions de *PgRouting*: *Pgr_NodeNetwork* + *Pgr_CreateTopology*. Els detalls i el funcionament d'aquesta operació estan explicats en l'apartat de Modelització del transport a peu.

2.10 Problemes detectats en la modelització del graf

Tot projecte professional, per molts mitjans de què disposi i per molt organitzat que estigui, passa per fases inesperades que no s'havien contemplat d'inici. Sovint, sobre aquests impasses no se'n deixa constància en els resultats d'un treball o una memòria. En aquest cas –però– es creu oportú documentar una sèrie de problemes amb què s'ha topat al llarg del desenvolupament del projecte, per tal que se sàpiga que problemes com els que s'expliquen a continuació són part del procés i és quelcom positiu que se'n faci menció, ja que poden resultar útils per a futurs treballs relacionats amb la temàtica d'aquest projecte.

Problema 1 – Pas per enllaçar la xarxa de carrers amb les estacions de transport públic.

El problema que sorgí en aquest pas, es deu a l'existència de segments aïllats en el graf original d'*OpenStreetMap*, els quals representen objectes com escales, pàrquings, etc. Com que el procediment per enllaçar els carrers amb les estacions es basava en buscar el punt més proper de la xarxa, en alguns casos aquests punts corresponien a segments aïllats de la xarxa que

provocaren la no connectivitat d'algunes estacions amb la xarxa real. Per resoldre-ho, s'hagué de definir un *query* per excloure els esmentats segments.

Per seleccionar els segments aïllats s'utilitzà l'operació de PostGIS: *ST_Touches*. Quan s'indica el contrari (és a dir, els casos en què “no toca / no troba connectivitat”), es fa possible discriminar els segments aïllats:

```
WHERE NOT (EXISTS ( SELECT 1
                    FROM routing_network b
                    WHERE st_touches(a.geom, b.geom) = true))
```

La següent imatge constata l'error que deriva de l'existència de segments aïllats:

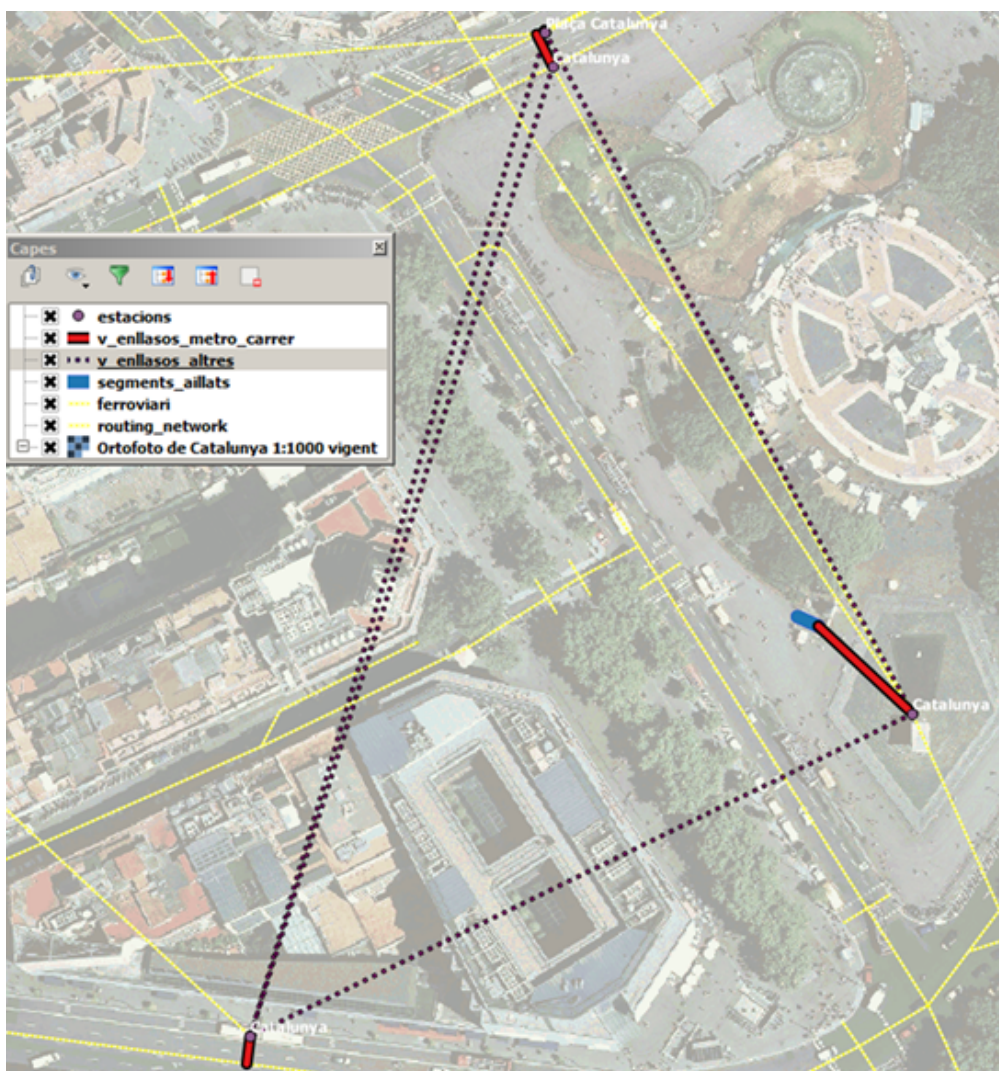


Figura 17. Resultat d'un cas erroni per enllaçar la xarxa originat per l'existència d'un segment aïllat. Font: Elaboració pròpia.

El *query* resultant d'aquest procés és el següent:

```
SELECT estacions.nom,
       estacions.supra_serv AS srv1,
       (SELECT routing_network.name AS srv2
        FROM routing_network where not exists (select 1 from
        segments_aillats where segments_aillats.gid=routing_network.gid)
        ORDER BY (estacions.geom <-> routing_network.geom)
        LIMIT 1) AS srv2,
       (SELECT st_shortestline(estacions.geom, routing_network.geom) AS
        geom
        FROM routing_network where not exists (select 1 from
        segments_aillats where segments_aillats.gid=routing_network.gid)
        ORDER BY (estacions.geom <-> routing_network.geom)
        LIMIT 1) AS geom
FROM estacions
```

Problema 2 – Error en la segmentació de la xarxa.

En el pas de *Pgr_NodeNetwork*, quan es pretén segmentar la xarxa establint nodes en les interseccions, s'ha detectat que –en alguns casos– l'operació per crear segments que connectin estacions amb carrers (*ST_MakeLine*) no s'efectua correctament amb la versió de PostGIS que s'utilitza (la v 2.2.0). A causa d'això, en el pas de creació de nodes, l'operació no estableix nodes en aquests trams conflictius ja que no hi detecta cap intersecció, per tant, el graf resulta mal connectat en certs trams i –més endavant– pot derivar en errors greus en el càlcul de rutes.

La causa del problema és que no existeix cap punt del segment (creat amb l'operació *ST_MakeLine*) que intersecti físicament amb l'altre segment. És a dir, que el punt final d'un segment no es troba realment a sobre de l'altre segment, com hauria de succeir. Aquest problema és un efecte de la no aplicació de la tolerància indicada en el procés de creació del segment que ha d'establir l'enllaç entre una línia i l'altra.

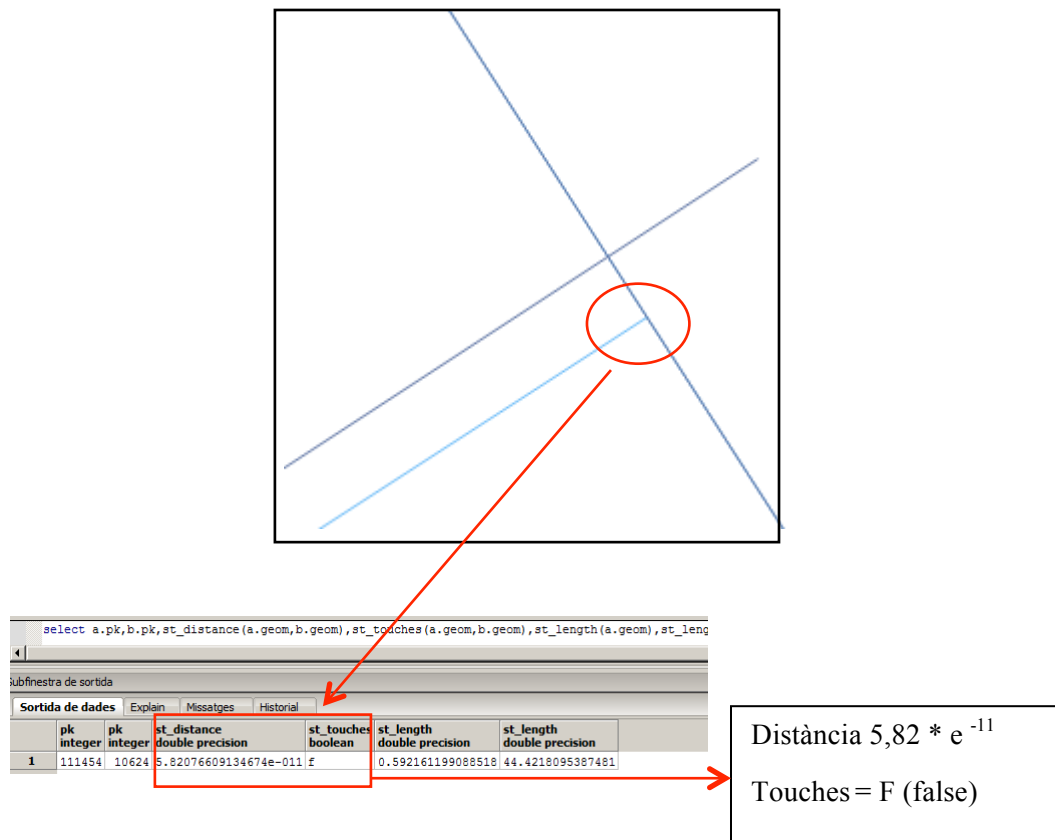


Figura 18. Representació de l'error.

En ser una distància tan reduïda, resulta difícil la identificació dels casos conflictius perquè visualment es fa impossible veure-ho. Per fortuna, amb la versió de PostGIS anterior (v 2.0.), aquest problema no passa i es procedeix amb l'execució del procés de segmentació de la xarxa.

3 Càlcul de rutes

3.1 Prova inicial per veure el funcionament d'Astar

És convenient provar d'executar la funció amb un cas senzill per tal de veure'n el funcionament i la resposta que té davant d'un càlcul determinat entre dos punts. Primer, s'han de definir els punts d'inici i de destí de la ruta, o *source* i *target* (62539 i 62740). Aquests punts estan inclosos a la taula *Routing_Network* (la xarxa completa), que equivalen a dos punts també presents a la taula de vèrtexs que genera automàticament *PgRouting* al crear la topologia.

```
select * from pgr_AStar
('SELECT id, CAST(source AS int4), cast(target as int4),
cast(cost as float8), x1, y1, x2, y2 FROM routing_network',
62539, 62740, false, false)
```

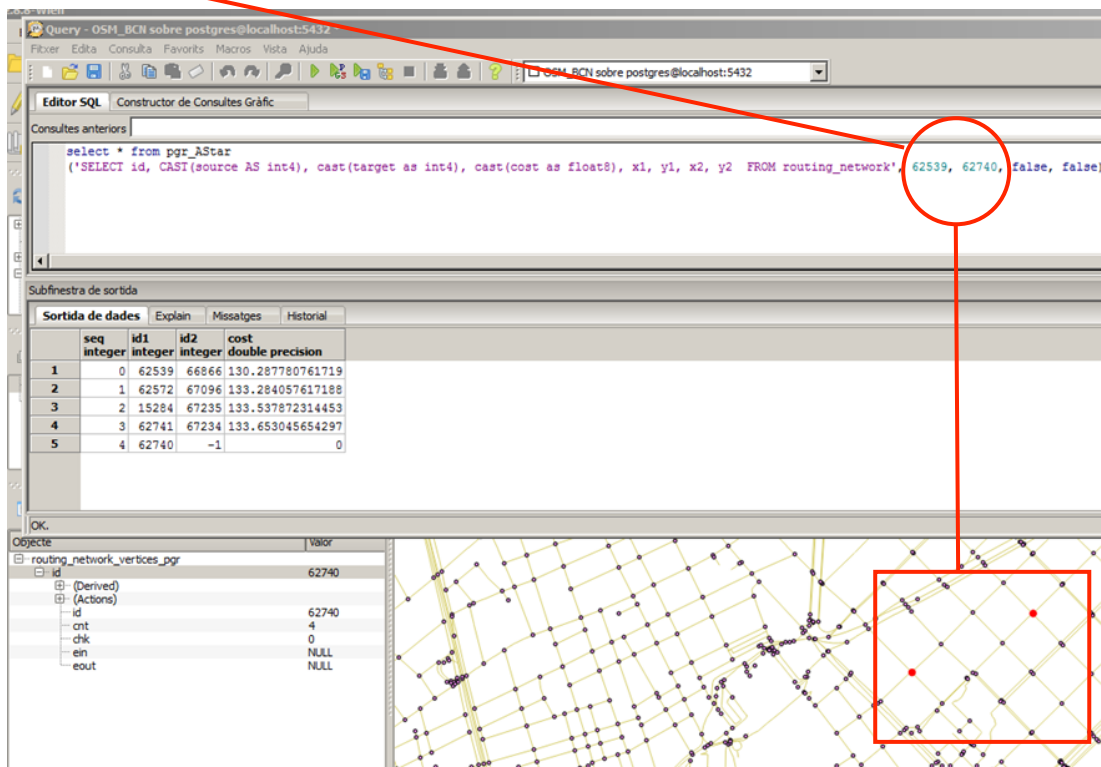


Figura 19. Comprovació del càlcul de rutes amb la funció Astar. Font: Elaboració pròpia.

3.2 Càlcul del vèrtex més proper

Fins aquí, la funció executa la cerca de ruta entre dos punts donats d'entre el total de vèrtexs existents en el graf. Per tal que la funció s'adeqüi a la realitat que representa el graf i –sobretot– a la disponibilitat de dades com a punt d'origen i destí (domicili i lloc de treball), s'ha hagut d'idear la manera perquè els vèrtexs d'origen i destí de la funció equivalguin a un punt proper

(el més proper possible) als quals indiquen les dades sobre desplaçaments. Per tant, s'ha creat una funció nova (que parteix de l'original *Astar*) de tal manera que identifica el vèrtex més proper respecte a l'origen i el destí i calcula la ruta en base a aquests dos vèrtexs. D'altra manera, seria impossible aproximar-se al punt d'origen o de destí que indiquen les dades, ja que només es podria indicar el càlcul de la ruta a través de la xarxa de vèrtexs que genera PostGIS en crear la topologia.

```
select *
pgr_astar('SELECT id, CAST (source AS int4), CAST (target as
int4), cast (longitud/velocitat as float8) as cost, x1, y1, x2,
y2 FROM xarxa where tipologia=1',
vertex_mes_proper(geom), vertex_mes_proper(desti), false, false)
from orig_dest_utils where idpers='05363016310004'
```

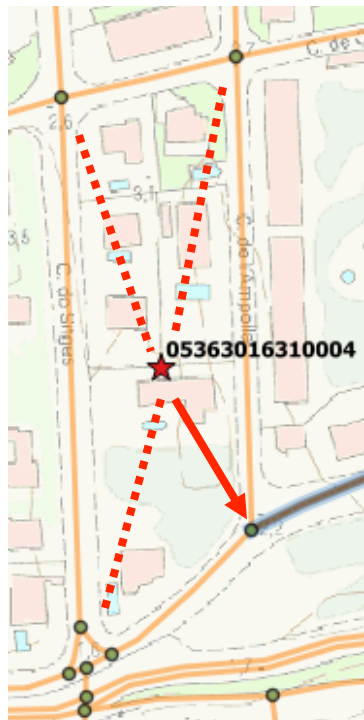


Figura 20. Representació gràfica del funcionament de la funció `vertex_mes_proper`. Font: Elaboració pròpia.

3.3 Càlcul de la ruta amb la variable *Velocitat*

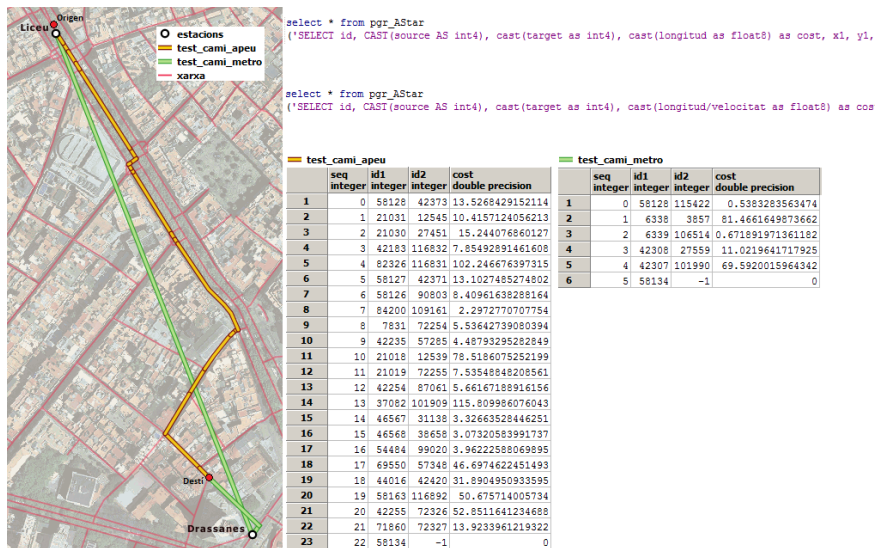


Figura 21. Representació del càlcul de rutes incorporant la variable *Velocitat*. Font: Elaboració pròpia.

Aquí s'observen dos casos, el primer dels quals executa la funció Astar sense contemplar les velocitats de cada segment de la xarxa. Per tant, no discrimina les vies segons el tipus o la velocitat, sinó que les considera totes igual (escull una ruta per fer a peu, enlloc d'utilitzar el Metro). En el segon cas, amb la variable sobre velocitats definida, sí que opta per la ruta que utilitza el Metro i torna sortir al carrer per arribar al punt de destinació.

A continuació, l'SQL d'ambdós casos. La diferència entre un i l'altre (subratllada de color vermell) es basa en atribuir una impedància que tingui en compte la velocitat de cada segment. És a dir, en el primer cas, el *cost* només ve donat per la longitud del segment. I, en el segon cas, el *cost* ve donat per la relació longitud/velocitat.

1)

```
Select * from pgr_AStar
('SELECT id, CAST (source AS int4), cast (target as int4), cast
(longitud as float8) as cost, x1, y1, x2, y2 FROM xarxa', 58128,
58134, false, false)
```

2)

```
Select * from pgr_AStar
('SELECT id, CAST (source AS int4), cast (target as int4), cast
(longitud/velocitat as float8) as cost, x1, y1, x2, y2 FROM
xarxa', 58128, 58134, false, false)
```

3.4 Preparació de les dades sobre desplaçaments per poder executar la funció Astar

3.4.1 Preparació dels tipus de mitjà de transport

Finalment, per poder utilitzar la funció Astar amb la càrrega de les dades sobre desplaçaments és necessari crear unes vistes a la base de dades on estiguin separats els diferents mitjans de transport i les combinacions possibles entre tipus de transport que recullen les dades d'IDESCAT. Aquests són els tipus de mitjans de transport que existeixen:

1	En cotxe o furgoneta com a conductor
2	En cotxe o furgoneta com a passatger
3	En autobus, autocar, minibus
4	En metro
5	En moto
6	Caminant
7	En tren
8	En bicicleta
9	Altres mitjans

En aquest cas d'estudi, només hi ha combinacions possibles entre els desplaçaments en metro, en tren, caminant o en bicicleta.

Es crearan 10 vistes, tantes com combinacions hi ha recollides a les dades sobre desplaçaments. S'utilitza la següent sentència (que s'haurà d'adaptar en cada cas):

```
SELECT orig_dest_utils.idpers,
pgr_astar('SELECT id, CAST (source AS int4), CAST (target as
int4), cast (longitud/velocitat as float8) as cost, x1, y1, x2,
y2 FROM xarxa where tipologia=1'::text,
persones_vertex_origen_desti.id_vertex_origen,
persones_vertex_origen_desti.id_vertex_desti, false, false) AS
pgr_astar
FROM orig_dest_utils,
persones_vertex_origen_desti WHERE orig_dest_utils.mdesp1 =
'6'::text AND orig_dest_utils.mdesp2 = ' '::text AND
orig_dest_utils.idpers = persones_vertex_origen_desti.idpers;
```

f	mdesp1	mdesp2	
bigint	text	text	
31547	6		+ [] astar_bici_8_0
15307			+ [] astar_camina_6_0
12446	4		+ [] astar_camina_6_7
2966	7		+ [] astar_camina_6_8
2152	8		+ [] astar_metro_4_0
813	4	7	+ [] astar_metro_4_6
478	4	6	+ [] astar_metro_4_7
179	4	8	+ [] astar_metro_4_8
134	6	8	+ [] astar_tren_7_0
87	6	7	+ [] astar_tren_7_8
20	7	8	

Figura 22. Estructuració de les dades en vistes (segons els modes de transport utilitzats).

Durant el procés de tractament de dades i preparació del graf per al càlcul de rutes, s'han anat destriant aquelles dades sobre desplaçaments que es podien utilitzar i les que no. Arribats a aquest punt, queden prop de 31.000 desplaçaments per analitzar (dels 378.000 registres inicials). Aquesta reducció tan important es deu –principalment- a la impossibilitat d'incloure aquells desplaçaments catalogats com a transport de mitjans motoritzats privats o autobús per falta de dades sobre els sentits dels carrers i les línies de bus que operen dins de l'àmbit d'estudi.

3.4.2 Preparació pel càlcul de ruta individualitzat

Per establir els punts d'origen i de destí per a cada individu implicat en el desplaçament que apareix en les dades d'IDESCAT, és necessari crear una taula que relacioni l'identificador de persona i les seves coordenades d'origen amb el vèrtex més proper de la xarxa que s'ha creat del graf. De la mateixa manera, també cal que s'indiqui el punt més proper al punt de destinació (que en aquest cas correspon al centroide del codi postal).

```
insert into persones_vertex_origen_desti
select orig_dest_utils.idpers,
       (SELECT      xarxa_vertices_pgr.id      as      d      FROM
xarxa_vertices_pgr,xarxa                                where
(xarxa_vertices_pgr.id=xarxa.source                    or
xarxa_vertices_pgr.id=xarxa.target)                    and
xarxa.tipologia=1

ORDER BY  xarxa_vertices_pgr.the_geom <-> orig_dest_utils.geom
limit 1) as id_vertex_origen,

(SELECT xarxa_vertices_pgr.id as d FROM xarxa_vertices_pgr,xarxa
where      (xarxa_vertices_pgr.id=xarxa.source          or
xarxa_vertices_pgr.id=xarxa.target) and xarxa.tipologia=1

ORDER BY  xarxa_vertices_pgr.the_geom <-> orig_dest_utils.desti
limit 1) as id_vertex_desti

from orig_dest_utils
```

A continuació, i sobre la taula recentment creada *persones_vertex_origen_desti*, s'executa el càlcul de rutes amb *pgr_Astar*. Com s'observa a la següent taula, el resultat s'estructura en registres que representen en cada cas el recorregut per un únic segment de la xarxa. És a dir, que cada vegada que el traçat canvia de segment (o el que és igual, passa d'un vèrtex a un altre) es genera un nou registre. I, així, successivament, fins que l'individu acaba el seu recorregut. El camp *pgr_costresult* indica els identificadors dels vèrtexs implicats en un determinat segment i el temps utilitzat per recorre'l. De color taronja, és el recorregut d'un mateix individu fins a la seva destinació. Quan el color passa a ser blau, comença el registre d'un altre individu.

idpers text	pgr_astar pgr_costresult
4021012040004	(5, 8098, 97921, 5.12761442007374)
4021012040004	(6, 10332, 117043, 2.5261965579403)
4021012040004	(7, 32504, 102109, 3.10368963836257)
4021012040004	(8, 14111, 53181, 1.57588048180671)
4021012040004	(9, 66673, 53180, 4.684446428975)
4021012040004	(20, 66672, 53179, 0.247542092130638)
4021012040004	(21, 26937, 38276, 9.54329616686832)
4021012040004	(22, 8099, 53182, 6.2676406278112)
4021012040004	(23, 26976, 58880, 5.77326697892641)
4021012040004	(24, 8123, 4923, 13.1467300556498)
4021012040004	(25, 8124, 4926, 2.10468879608372)
4021012040004	(26, 8125, 4924, 1.36872096174351)
4021012040004	(27, 8126, 77237, 25.4668441930015)
4021012040004	(28, 28972, 17578, 17.7132511843602)
4021012040004	(29, 14146, 92106, 26.4536596778386)
4021012040004	(30, 10353, 12836, 17.4736925325081)
4021012040004	(31, 21476, -1, 0)
4023000990003	(0, 32483, 79475, 2.79649697714169)
4023000990003	(1, 32484, 19970, 4.59761010327461)
4023000990003	(2, 32485, 109260, 20.292813016054)
4023000990003	(3, 78411, 73742, 2.00982368769803)
4023000990003	(4, 64069, 64677, 1.3279771446485)
4023000990003	(5, 55792, 39870, 2.77252495641437)
4023000990003	(6, 28933, 17551, 19.7063204350738)
4023000990003	(7, 28934, 43872, 1.52665599599018)
4023000990003	(8, 10258, 6089, 2.98730219736639)

→ Temps de desplaçament
→ ID vèrtexs

Figura 23. Taula dels registres que emmagatzema la base de dades amb identificador de persona, vèrtex i temps utilitzat per recórrer cada segment de la xarxa.

3.4.3 Preparació pel càlcul de resultats agrupats

Es fa evident que –de cara a l’anàlisi dels resultats– serà més pràctic presentar el recorregut total que realitza l’individu agrupant el temps i la distància que necessita per desplaçar-se. Per aconseguir-ho, es fa el següent *join* sumant els registres de les variables per a un mateix identificador de persona:

```
create table desplaçaments_persona as
SELECT
idpers,
sum((pgr_astar).cost) as temps,
sum(xarxa.longitud) as distancia
FROM desplaçaments
JOIN xarxa ON (pgr_astar).id2 = xarxa.id group by idpers
```

	idpers text	temps double precision	distancia numeric
1	03775000830001	350.649353575581	385.7142889331391132
2	03775002200003	591.896805775868	651.0864863534553132
3	03775002690001	596.050534857287	655.6555883430154532
4	03775002690002	596.050534857287	655.6555883430154532
5	03775002690003	5902.60201185809	35381.51008132646182
6	03775003360003	475.931113743654	523.5242251180197132
7	03775003870001	472.721064078395	519.9931704862343932
8	03775003870002	155.686577989891	519.9931704862343932
9	03775003870003	472.721064078395	519.9931704862343932
10	03775005150001	850.730677686386	935.8037454550241132
11	03775005500002	522.491759985332	574.7409359838648932
12	03775006210003	628.519752704209	691.3717279746295132

Figura 24. Taula amb els resultats del càlcul de ruta agrupats per individu.

3.5 Problemes detectats en el càlcul de rutes

Problema 1 – Utilització d’enllaços subterranis per a desplaçaments a peu

En el càlcul de rutes s’ha detectat que els individus que utilitzen la xarxa de carrers per anar a peu o amb bicicleta, sovint utilitzen els enllaços de metro com a mètode de desplaçament. És un error que es deu a no haver restringit l’accés a aquestes línies subterrànies mitjançant una tipologia especial que només pogués combinar-se amb l’ús de la xarxa de metro o ferrocarril. Com que els enllaços s’han creat manualment traçant línies rectes entre estacions, seran presumiblement força utilitzats pels casos d’usuaris que es desplacen a peu o en bicicleta, la qual cosa no és desitjable perquè no es correspon amb la realitat.

Problema 2 – Pèrdua d'un 2% dels registres útils sobre desplaçaments a causa de l'existència de segments aïllats.

Els segments aïllats de la xarxa *OSM* han provocat que es perdessin alguns registres de desplaçaments. Aquests segments no tenen connexió amb la resta de la xarxa i per això estan impeditos de ser connectats amb la resta del graf (tant si és com a punt d'origen o punt de destí). En el moment de buscar el punt més proper de la xarxa donat un punt d'origen per un individu que ha d'iniciar el seu recorregut a través de la xarxa, si es donava la coincidència que la connexió s'establiria amb un segment aïllat, provocava la no execució del càlcul de ruta per aquest cas.

S'haurien d'haver exclòs els segments aïllats permanentment per evitar tal pèrdua de registres. Així mateix, la funció *Astar* hauria d'alertar d'això però, en canvi, el que fa és executar la funció i –directament– obviar aquests casos.

IV. Anàlisi dels resultats

1. Anàlisi estadística

Entrant en matèria d'anàlisi, una vegada finalitzats els processos de creació del graf i de càlcul de rutes, el material que servirà per encarar la fase actual està emmagatzemat en forma de taula dins la base de dades. Les dades que resulten del càlcul de rutes per a cada desplaçament es presenten de manera agrupada (amb un identificador únic per a cada persona) i els camps d'interès per a l'anàlisi estadística que ve a continuació són: la distància euclidiana entre el punt d'origen i el de destí, la distància real a través dels grafs entre els mateixos punts i el temps total que necessita l'usuari en cada recorregut (donades unes velocitats específiques per a cada tipus de mitjà de transport). La taula que exemplifica el resultat final amb la informació sobre el càlcul de rutes pot ser consultada en l'apartat d'annexos del treball.

Tot seguit, els resultats estadístics s'analitzen amb el programa *Weka Data Mining* de la University of Waikato, New Zealand. Aquest programa de mineria de dades (d'*Open Source*), funciona com a maquinària de processament, classificació, visualització i aplicació d'algoritmes sobre les dades extretes de l'estudi. En aquest cas, ha servit per analitzar la distribució dels diferents casos de desplaçaments que es tenien per executar el càlcul de rutes. Veure el comportament de les funcions de càlcul de ruta sobre el graf que s'ha creat és quelcom interessant per incloure en aquest estudi. Les dades que s'extreuen de l'anàlisi amb *Weka* són:

		Distancia				
mdesp1	mdesp2	Casos	max	min	Mitjana	Desv. típica
6		30787	41708.96	37.34	2458.70	2796.34
4		12127	45045.85	81.98	6146.36	5551.54
7		2904	62438.54	130.96	10701.13	9393.16
8		2100	40686.25	121.99	3868.54	4543.12
4	7	793	43818.75	124.72	15323.21	8119.08
4	6	472	26490.70	207.05	4707.41	3865.17
4	8	177	32050.82	293.49	5172.22	4615.83
6	8	130	10790.60	245.45	2434.75	1911.05
6	7	84	30261.97	291.55	11573.65	8933.73
7	8	20	33682.37	3871.75	14217.28	7873.18
Total		49594				
Coef. correlació		abs(dist. Euclídea-distancia)/distancia				
		minim	màxim	mitjana	desv. Típica	
0.988		0.001	0.917	0.248	0.125	
0.964		0.004	0.865	0.332	0.148	
0.936		0.011	0.813	0.329	0.144	
0.992		0.010	0.875	0.225	0.120	
0.924		0.051	0.819	0.325	0.140	
0.936		0.022	0.750	0.350	0.160	
0.990		0.061	0.532	0.239	0.089	
0.986		0.029	0.652	0.220	0.109	
0.967		0.096	0.646	0.294	0.111	
0.948		0.095	0.554	0.288	0.127	

Figura 25. Taules amb els resultats de l'anàlisi estadística. Font: Elaboració pròpia.

Aquest material estadístic ha permès estudiar la relació entre la distribució de la distància euclidiana entre el punt d'origen i de destí respecte a la distància real recorreguda a través del graf. A més, s'ha calculat l'error relatiu (ε) definit com:

$$\varepsilon = \frac{|d_e - d_r|}{d_r} (*)$$

d_e distància euclidiana entre els punts d'origen i destinació

d_r distància *Astar* corregida entre els punts d'origen i destinació

S'observa que la distribució de ε és molt semblant a una **log** normal i que la correlació entre d_e i d_r és molt forta (en cap cas el coeficient de correlació és inferior a 0.9).

S'ha calculat el **ln(ε)** i la seva distribució és molt semblant a una normal en tots els casos. S'ha calculat també el z-score relatiu al **ln(ε)** i s'han localitzat els outliers de la distribució.

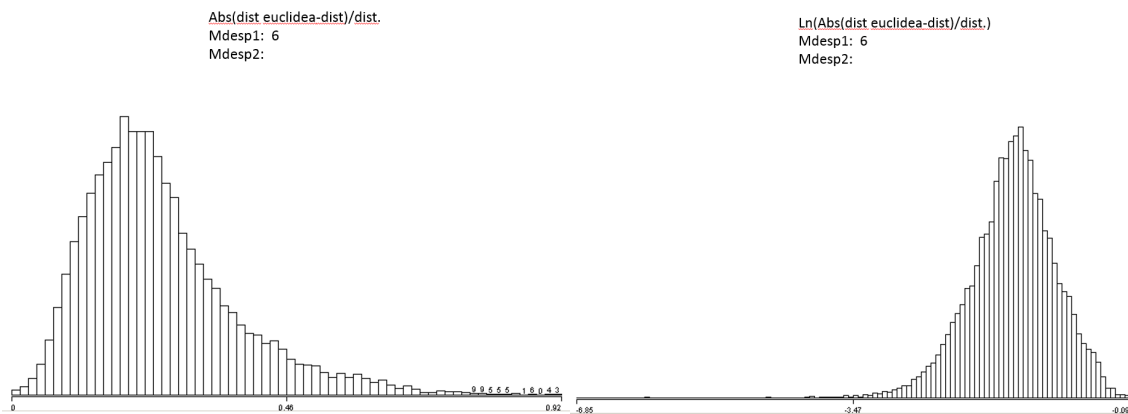


Figura 26. A l'esquerra, distribució de l'error relatiu (ε). A la dreta, la distribució del logaritme de l'error relatiu **ln(ε)**.

Tenint la distribució amb un comportament (quasi) normal, es poden identificar els *Standard Score* (o Z-Score), que són aquells valors que s'allunyen de la distribució mitjana de la mostra i equivalen als *outliers* que es pretén analitzar (valors anormalment alts del Z-Score).

Per indentificar els casos d'*outliers*, s'analitza la distribució per a cada grup de desplaçaments, segons el mitjà de transport utilitzat. S'ha vist oportú incloure aquest càlcul perquè suposa un mètode per avaluar l'eficiència del graf o –si més no– per detectar aquells casos poc coherents respecte a la resta. Aquests casos són fàcils d'identificar amb aquesta metodologia. S'interpreta que els *outliers* es deuen a potencials errors en la construcció del graf, al mètode de selecció dels vèrtex d'origen i destí o a ineficiències de la xarxa real.

				test normalitat		Casos amb zscore per sobre de					
				1.281	1.645	1.960	2.326	2.576			
mdesp1	mdesp2			20%	10%	5%	2%	1%			
6		30787	no	2799	1239	435	78	13			
4		12127	no	1042	257	17	0	0			
7		2904	no	281	83	9	0	0			
8		2100	no	186	88	41	9	3			
4	7	793	no	69	19	4	0	0			
4	6	472	no	37	2	0	0	0			
4	8	177	si	14	8	4	0	0			
6	8	130	si	13	5	4	2	0			
6	7	84	si	5	1	1	0	0			
7	8	20	si	2	0	0	0	0			
Total		49594									

Figura 27. Taula amb els resultats de Z-Score. Font: Elaboració pròpia.

Sembren remarcables els valors de Z-Score per als desplaçaments que utilitzen el graf per desplaçar-se a peu o amb bicicleta. També destaquen alguns casos en què s'utilitza el tren com a mitjà de transport.

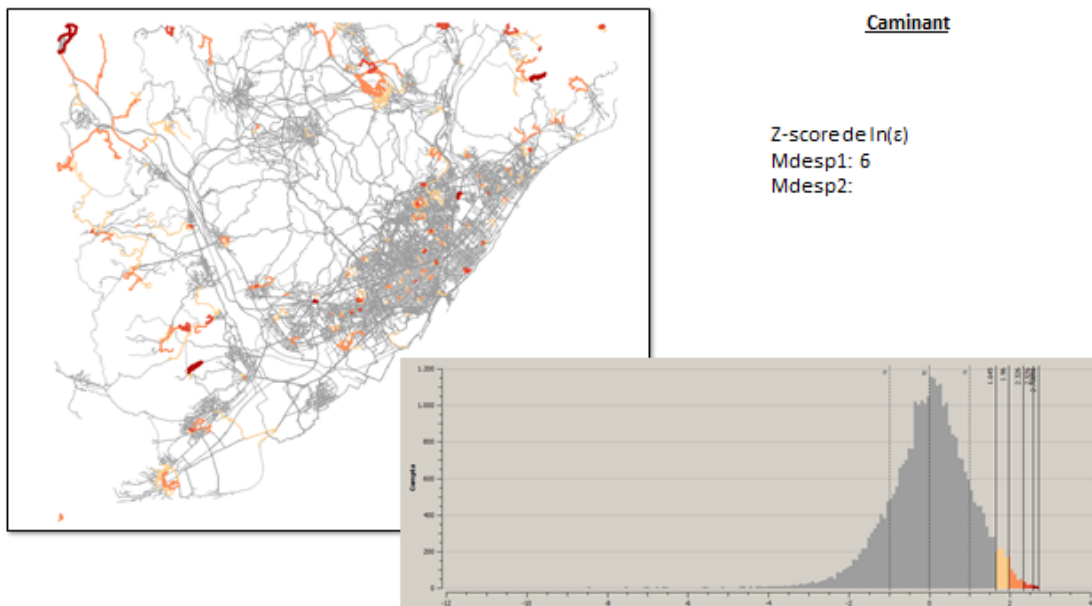


Figura 28. Distribució dels casos de Z-Score per a desplaçaments a peu. Font: Elaboració pròpia

Amb l'estudi individualitzat per a cada mitjà de transport i la representació gràfica amb QuantumGIS, es durà a terme un diagnòstic dels outliers.

2. Diagnosi de casos irregulars: els *outliers* de la xarxa

2.1 Casos de desplaçaments amb ferrocarril com a mitjà de transport:

En alguns casos, es veu una desviació de la norma, especialment per a casos amb desplaçaments amb tren com a mitjà de transport. Aquests casos coincideixen amb els quals més distància s'ha de recórrer per assolir la destinació. L'anàlisi d'aquests casos conclou en que la desviació pot ser donada per diversos factors, dels quals se'n plantegen les següents hipòtesis:

- **Que el recorregut sigui erroni degut a la falta d'enllaços entre estacions.**

Com que –d'inici– es van introduir a la base de dades els enllaços de manera manual i a partir de les fonts d'informació disponibles en aquell moment, és possible que –principalment– s'establissin els enllaços de la zona central de l'àrea d'estudi i que s'obviessin enllaços de zones perifèriques dels quals no es tenia constància.

És un problema derivat de la falta d'informació, tot i que de fàcil solució. Tal problema no succeeix amb la xarxa de Metro, ja que –en aquell cas– sí que hi havia disponibilitat de dades que permetien establir els enllaços amb certesa.

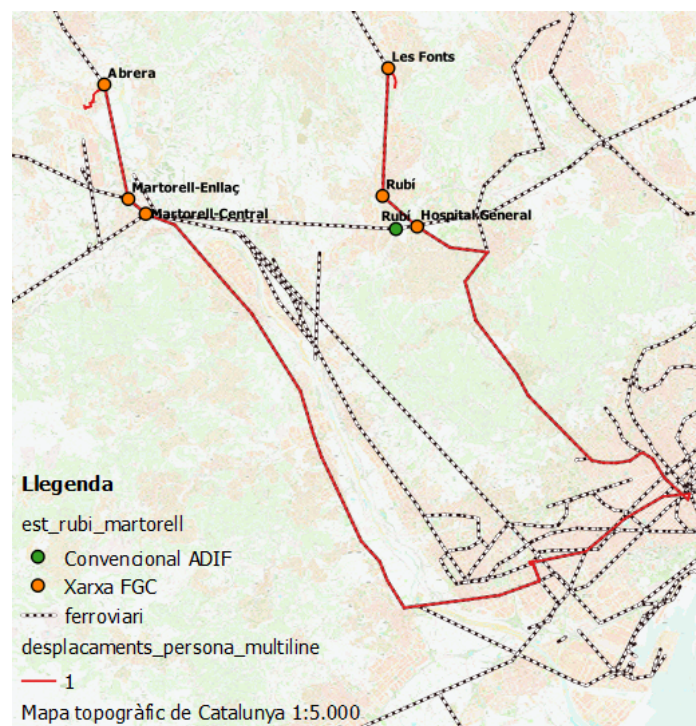


Figura 29. Representació gràfica dels possibles errors derivats de la falta d'enllaços entre estacions de la xarxa ferroviària. Font: Elaboració pròpia.

- **Que el retall de l'àrea d'estudi hagi impossibilitat la connexió amb altres vies (existents a la realitat però que no han cabut a l'àrea d'estudi).**

Aquesta hipòtesi es pot donar –especialment– en aquells casos on l'origen i el destí d'un desplaçament concret estiguin situats en la zona perifèrica de l'àrea d'estudi. És a dir, prop del límit de la xarxa que representa el graf. S'utilitza el mateix cas de desplaçament en ferrocarril que en la hipòtesi anterior. Ambdues es consideren vàlides i seria possible comprovar-ho definitivament creant el graf per a tot Catalunya.

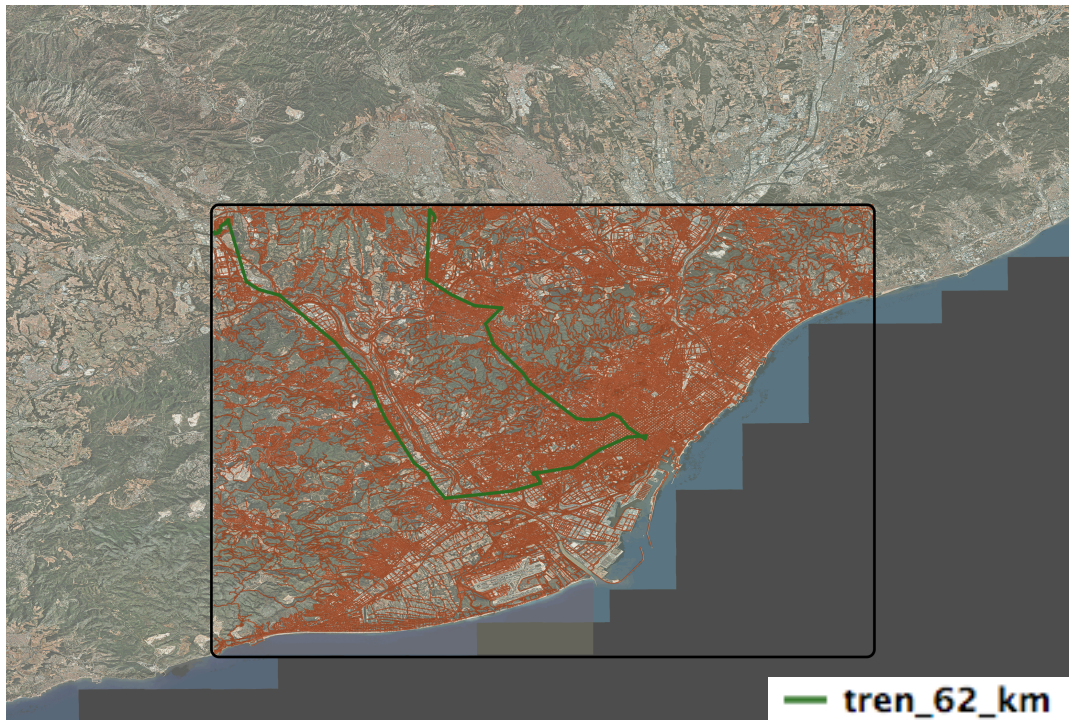


Figura 30. Representació gràfica dels possibles errors derivats de la delimitació de l'àrea d'estudi. Font: Elaboració pròpia.

2.1 Casos de desplaçaments a peu com a mitjà de transport:

D'entrada, se sabia que les dades disponibles per al desenvolupament del treball no eren suficientment precises per dur a terme el càlcul de rutes amb l'exactitud que es desitjaria (a causa de les fonts de les dades amb què s'ha basat la creació del graf i a causa de la qualitat de les dades sobre desplaçaments, que estableixen un punt de destinació aproximat a partir del centroide del codi postal del lloc de treball). Per tant, tot i obviar aquest fet, i establint la connexió entre els punts d'origen i destí amb la xarxa mitjançant operacions de buscar el node més proper en cada cas, es pensava que s'obtidrien resultats prou satisfactoris. Ara bé, gràcies a l'anàlisi dels *outliers* s'ha pogut constatar que s'han donat casos en què el procés de connectar els punts d'origen i de destí al node més proper de la xarxa conduïa a errors destacables. I és que es pensava que els problemes només es correspondrien amb el moment de buscar el punt més proper a la destinació (ja que era una aproximació que s'havia solucionat amb el centroide del

codi postal). Però el que s'ha detectat és que el problema també succeeix amb el punt d'origen. Per tant, no és un problema de falta d'exactitud en el *source* i *target*, sinó que el problema rau en el mètode utilitzat per traçar la ruta donat el node més proper a l'origen i el destí.

S'observa que, tot i que la xarxa de carrers per desplaçaments a peu sigui la més densa i la més segmentada, no ho és suficientment. I és que, en alguns casos, buscant el node més proper pel punt d'origen i de destí, el traçat de la ruta és completament erroni a causa de com s'havia fet la segmentació del graf. En algunes zones la xarxa estarà suficientment segmentada per executar les funcions de ruta i en d'altres no.

A continuació es mostra el cas d'un dels *outliers* per a desplaçaments a peu:

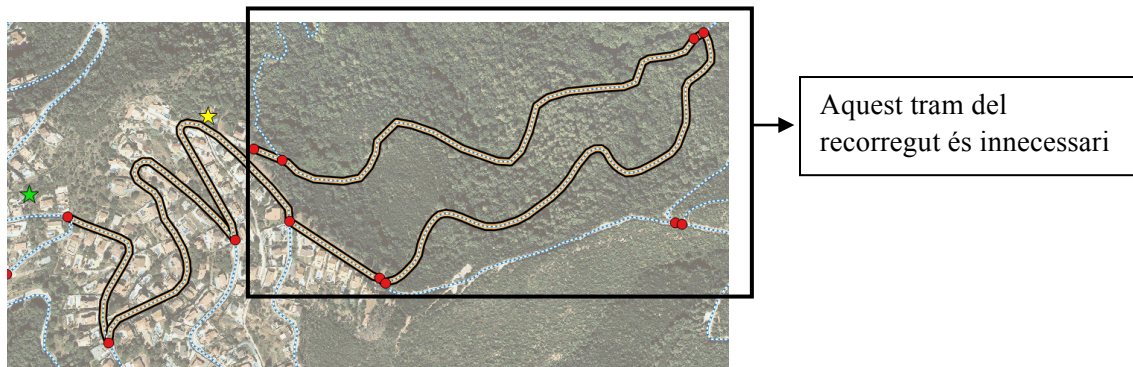


Figura 31. Representació gràfica de possibles errors derivats del procés de cerca del node més proper. Font: Elaboració pròpia.

Executant la funció (buscant el vèrtex més proper de la xarxa), pot passar que en alguns casos trobi un vèrtex que condiciona la ruta de tal manera que fa un traçat erroni.

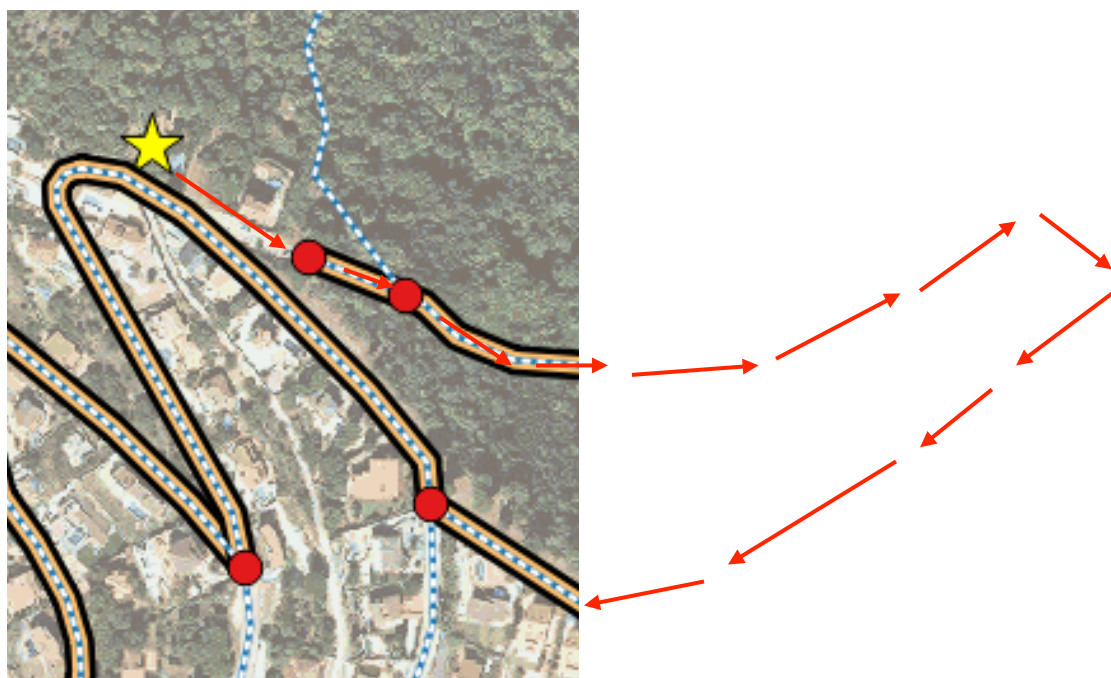


Figura 32. Representació gràfica de possibles errors derivats del procés de cerca del node més proper (2). Font: Elaboració pròpia.

Com que en el tram de la xarxa més proper a l'origen no hi ha interseccions properes, pren un vèrtex que provoca l'error i, enlloc d'anar cap el destí, fa que traci una volta innecessària.

Aleshores, es demostra que, per obtenir uns resultats més satisfactoris, s'hauria de dur a terme un procés que segmenti la xarxa de nou (partint de la hipòtesi que s'acaba d'argumentar). Es planteja una alternativa que es basa a segmentar els trams de carrer segons tantes adreces de portal hi hagi geocodificades a l'àrea d'estudi. Així, el resultat de buscar el node més proper del punt d'origen i de destí resultaria en un procés molt més acurat de com és ara mateix (pràcticament no es donarien inexactituds derivades de buscar el node més proper perquè la xarxa estaria segmentada fins a tal punt que el node més proper seria molt pròxim i no es traçarien rutes errònies).

Una alternativa més simple es basaria en la creació de segments que uneixin el punt d'origen o de destí mitjançant el traçat d'una línia que busqui el punt més proper d'un segment de la xarxa en el seu pas prop del punt d'origen o destí. Seria una solució costosa i només es podria fer cas a cas perquè es basa en la modificació de la xarxa donades unes dades concretes de desplaçaments per traçar-hi les rutes. Per tant, no seria una solució que pogués establir un mètode d'aplicació automatitzat per a qualsevol consulta (només serviria per a casos on es conegui amb antelació l'origen i el destí del desplaçament).

Conseqüentment, es decideix dur a terme la primera alternativa: una modificació de la xarxa per tal d'augmentar la densitat de segments i tractar de corregir els errors que s'identifiquen a partir del diagnòstic dut a terme sobre els *outliers*.

3. Correcció de la xarxa

A continuació es duu a terme una modificació de la xarxa per tal de corregir els errors detectats en el càlcul de rutes. D'inici, s'intueix que aquests errors són més propensos d'ocórrer en zones de baixa densitat d'interseccions, com poden ser casos d'urbanitzacions o zones perifèriques de la xarxa.

3.1 Procés de correcció:



Figura 33. Xarxa inicial

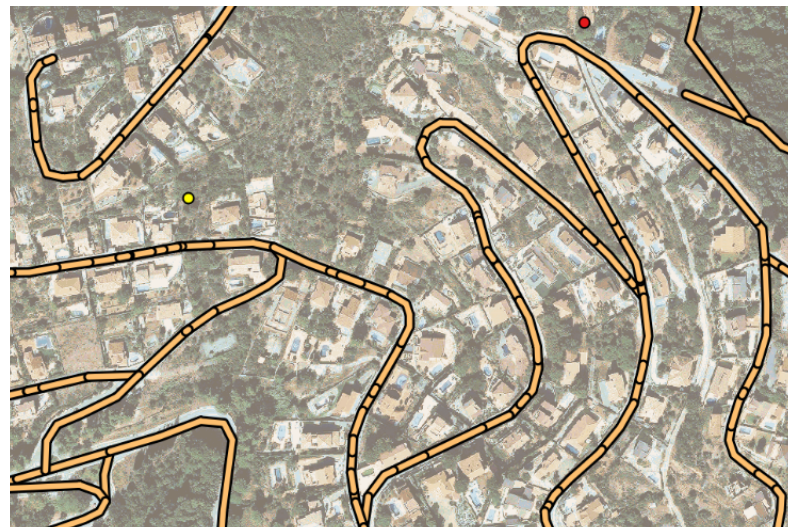


Figura 34. Xarxa segmentada de nou

L'increment de segments que s'observa d'una imatge a l'altra, s'aconsegueix a partir de repetir el procés de creació de nodes i creació de topologia mitjançant la incorporació de les adreces de portal a la base de dades. Cada adreça de portal estableix una connexió amb la xarxa. El punt on connecta la xarxa amb la línia imaginària procedent del punt d'adreça segmentarà la xarxa en N segments. Una vegada feta la segmentació, les connexions desapareixeran.

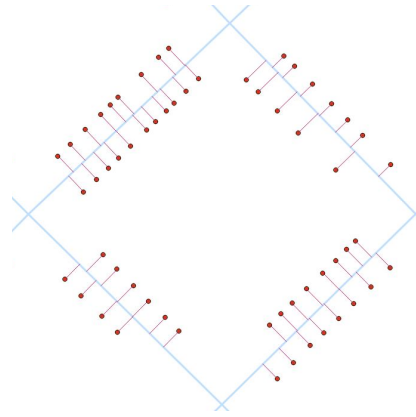


Figura 35. Representació gràfica de la incorporació de les adreces de portal en el procés

3.2 Càlcul de la ruta amb la xarxa segmentada de nou:

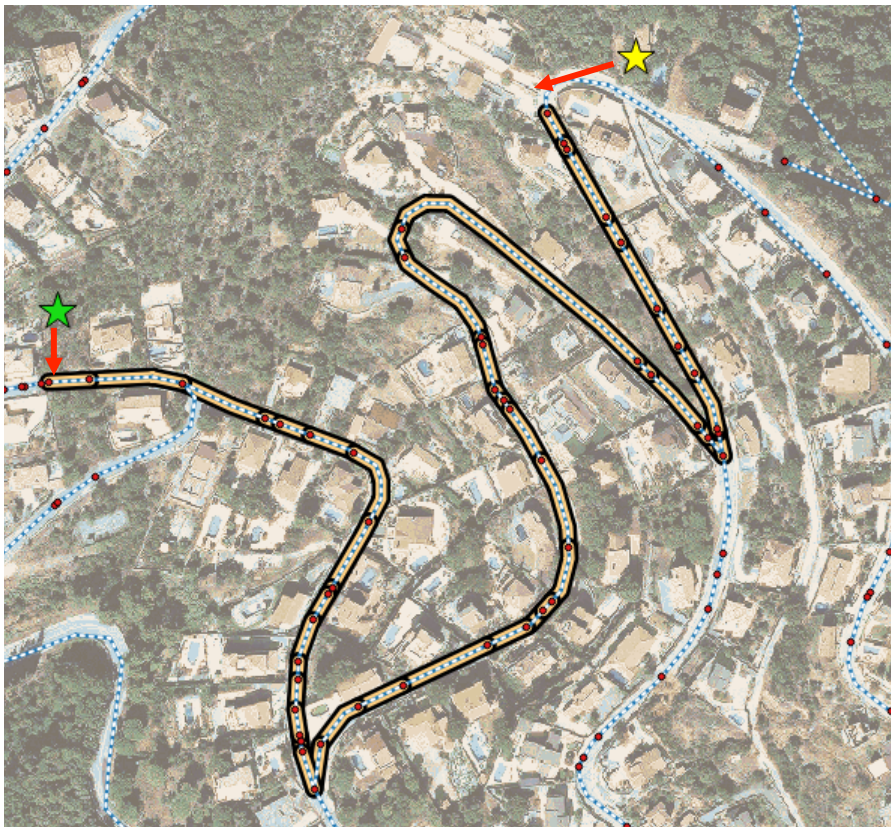


Figura 36. Nou traçat de la ruta després del procés de segmentació.

Els punts vermells equivalen als nous vèrtexs (n'hi ha molts més que en el cas de la xarxa inicial).

S'observa que executant la funció de nou sobre la xarxa més segmentada, el vèrtex que pren com a referència des de l'origen és un altre i, en conseqüència, es modifica el traçat de la ruta (partint dels mateixos punts d'origen i destí). La correcció en xifres parla per sí sola:

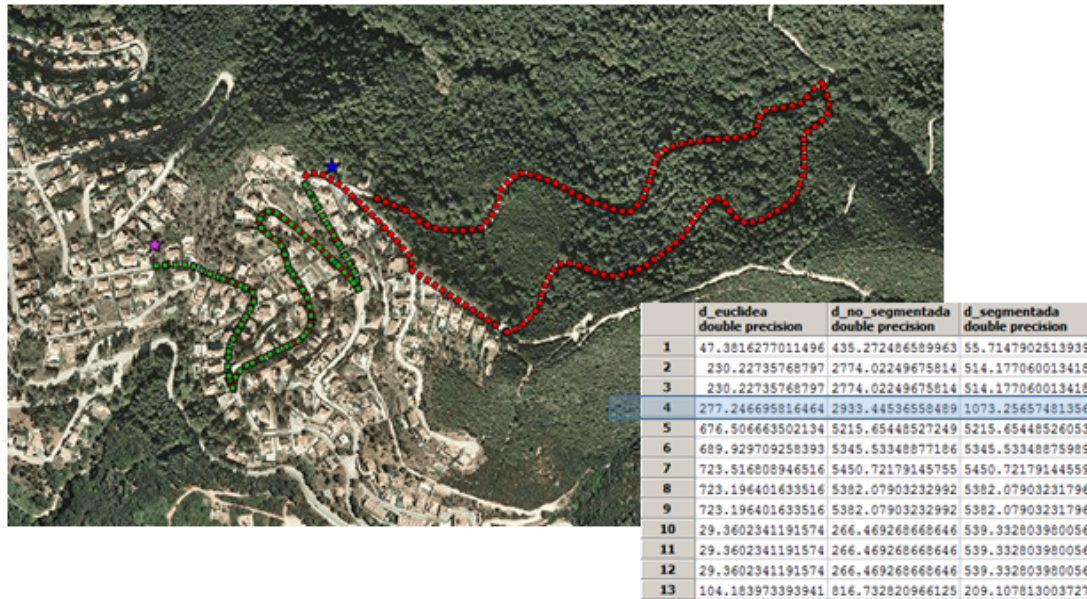
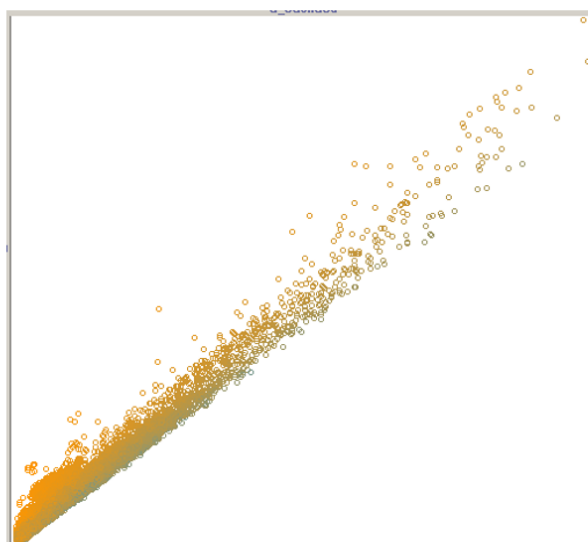


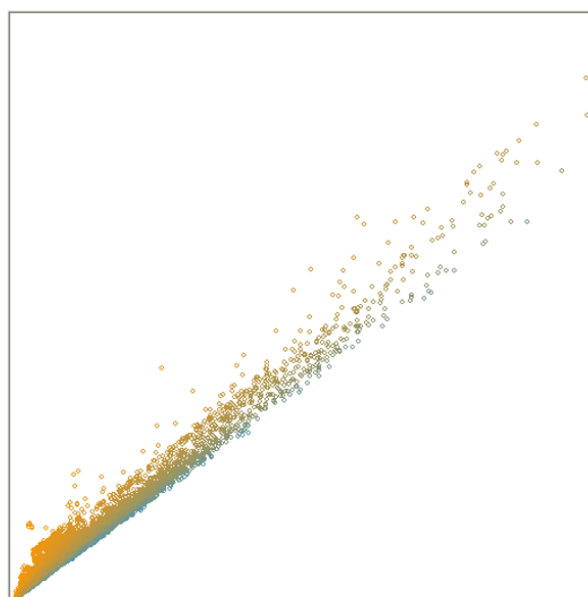
Figura 37. Comparativa de les distàncies segons distància euclidiana, distància abans i després de segmentar la xarxa de nou. Font: Elaboració pròpia.

Es demostra que el mètode funciona, la qual cosa s'evidencia tornant a executar la funció de càlcul de rutes pels casos que van a peu com a mitjà de transport. El resultat torna a passar pel programa *Weka* de mineria de dades i amb els següents gràfics, que representen la correlació entre la distància euclidiana i la distància real a través del graf, s'observa com la correlació s'ha normalitzat en 0.00017 punts i han disminuït els casos d'*outliers*. Per tant, el procés no s'ha emprat només per solucionar un cas d'estudi concret, sinó que s'ha fet extensiu a tota la xarxa.



R=0.98788

Sense segmentar



R=0.98805

segmentada

Figura 38. Gràfiques de la correlació entre la distància euclidiana i la distància real recorreguda a través del graf. A dalt, abans de segmentar. A baix, segmentada de nou. Font: Elaboració pròpia.

V. Conclusions

Donades per concloure les diferents fases que componien aquest treball, desenvolupat durant el període de pràctiques professionals a l'Institut d'Estadística de Catalunya, és moment de fer balanç i extreure les valoracions de tot el procés i la feina feta durant aquests mesos d'intens treball.

En primer lloc, es considera satisfactòriament finalitzada la tasca de creació del graf que representa la xarxa de transport a peu i transport ferroviari per a l'Àrea Metropolitana de Barcelona. Segons la disponibilitat de dades amb què s'ha pogut treballar, el graf ha estat testat i analitzat de tal manera que es considera un producte acabat i preparat per ser utilitzat. Això sí, existeix un marge de millora important amb el qual es podria treballar de cara al futur. L'aportació de noves dades i la realització d'algunes modificacions específiques del producte actual suposarien un plus qualitatiu respecte al resultat que s'ha aconseguit en tres mesos i amb els mitjans disponibles a l'inici. La disponibilitat de dades referents als sentits dels carrers o dades sobre les línies d'autobusos, permetrien fer extensiu el càlcul de rutes per aquells usuaris que utilitzen el vehicle privat o l'autobús com a mitjà de transport. Malauradament, en aquest cas, han hagut de ser exclosos.

En segon lloc, el càlcul massiu de rutes ha estat -en sí mateix- el millor testador possible del graf. Tot i la impossibilitat de dur a terme el càlcul per al 100% dels desplaçaments, els objectius en aquest cas també s'han complert satisfactòriament. I és que, a més d'haver suportat el càlcul de rutes i d'haver obtingut uns resultats estadístics amb propietats espacials, la infraestructura que s'ha creat a partir d'un sistema gestor de bases de dades té més possibilitats de les que s'han explotat durant aquest període. L'adequació del graf per incloure més mitjans de transport, la combinació d'impedàncies diverses com a variable *cost* en el càlcul de rutes o la creació d'un servidor de mapes per fer consultes sobre desplaçaments dins de l'àrea d'estudi, són exemples que complementarien el treball fins ara realitzat. Tanmateix, l'alt grau d'automatització dels processos i les operacions necessàries per dur a terme el càlcul de rutes massiu és una fita que s'ha assolit i creiem que -aquest- és el veritable valor afegit que aporta aquest assaig.

En tercer lloc, es valora molt positivament la possibilitat d'haver establert una metodologia estadística per l'anàlisi del graf que s'ha creat. D'una banda, aquesta metodologia ha servit per detectar errors generats durant el procés de creació. D'altra banda, serveix també per comprovar l'eficiència de la xarxa de transport existent, la qual cosa resulta especialment interessant de cara a la planificació estratègica del territori en matèria de mobilitat. La possibilitat d'haver treballat en un entorn on l'Estadística és la ciència predominant, m'ha permès l'aprenentatge i la inclusió d'una anàlisi estadística d'alt valor analític per concloure aquest treball.

Finalment, m'agradaria fer èmfasi en la manera i els mitjans amb què s'ha dut a terme el projecte. Partint d'una petició ambiciosa de l'IDESCAT, la idea inicial del projecte era un tant difusa. Com en la majoria dels començaments, la incertesa del principi plantejava dubtes sobre com arrencar amb tot plegat. Val a dir que, tot i la successió dels dubtes i entrebancs que anaven sorgint al llarg del desenvolupament del projecte, el coneixement i la implicació d'Eduard Suñé (el meu tutor per part de l'entitat col·laboradora) van permetre que en cap moment es posés en dubte la consecució dels objectius proposats inicialment. A més, la seva fidel aposta pels *software* lliures m'ha ensenyat que és totalment factible dur a terme encàrrecs professionals amb programari lliure i que moltes vegades pequem de prejudicis a l'hora de considerar la utilització d'aquests programes, quan el que realment importa és la qualitat de les dades amb què es treballa.

A manera de reflexió, voldria recalcar la importància que té el concepte de *Dades Obertes* en l'àmbit professional dels Sistemes d'Informació Geogràfica. Més enllà de les iniciatives internacionals de col·lectius emparats per l'*Open Geospatial Consortium* o la voluntat de crear biblioteques de dades obertes als webs de les institucions públiques locals i estatals, cosa que suposaria realment aplicar el concepte de *Dades Obertes* en aquesta societat seria el fet de facilitar l'accés i la lliure interoperabilitat de la majoria de les dades que avui en dia romanen privatitzades o són de molt difícil accés per la comunitat (tot i que pertanyin a una entitat pública). Aquest projecte és un cas més de tots aquells treballs que, si s'hagués tingut accés a unes dades de qualitat, haurien conclòs amb uns resultats millors del que s'han pogut aconseguir amb el material i les dades que s'han hagut d'utilitzar davant la falta de temps per aconseguir la col·laboració de les entitats que gestionen la mobilitat i el transport a Catalunya.

VI. Bibliografia

- CALDARELLI, G et CATANZARO, M (2014). *Redes: una breve introducción*, Madrid, Alianza (184 p.)
- VALLCORBA, O (2010). *BCN – Routing: Càlcul web de rutes amb pgRouting, OpenStreetMap i OpenLayers*. Barcelona, treball final de carrera (UOC).
- BUDHRANI, R (2009). *Palma Routing: Càlculo web de rutas con pgRouting, OpenStreetMap y OpenLayers*. Palma, treball final de carrera (UOC).

Netgrafia

- Documentació de PostgreSQL:

<https://www.postgresql.org/docs/>

- Mapping GIS – Cómo configurar una base de datos PostGIS para pgRouting:

<http://mappinggis.com/2015/01/como-configurar-una-base-de-datos-postgis-para-pgrouting/>

- Boston Geographic Information Systems – pgRouting: Loading OpenStreetMap with Osm2Po and route querying:

http://www.bostongis.com/PrinterFriendly.aspx?content_name=pgrouting_osm2po_1

- Free and open source GIS ramblings – A beginner's guide to pgRouting:

<https://anitagraser.com/2011/02/07/a-beginners-guide-to-pgrouting/>

- GitHub.com – Càlculo de rutas con pgRouting:

<https://github.com/geomatico/postgis2.0-advanced/blob/master/tema4.rst>

- GitHub.com – Análisis de redes con pgRouting:

https://github.com/CentroGeo/practicas_sig/tree/master/practica_3

- Universitat de Girona – Teoria de Grafs:

<http://www3.udg.edu/dghha/cat/secciogeografia/prac/models/Apunts%20grafs%20%283%29.htm>

VII. Annexos

idpers	cpo	cmun	dis	secc	edat_q	ltraha	paistra	protrab	muntrab	nviatje	mdesp1	mdesp2	ltdesp	codtrab	motiu	desp	cod_min	tipus	trans	transport	gps_x	gps_y	factor	distancia_euclidea	temps	distancia_astar
text	text	text	text	text	integer	text	text	text	text	text	text	text	text	text	text	text	text	text	text	text	double precision	double precision	double precision	double precision	double precision	numeric
1	03976002220001	08	019	01	001	133	108	08	019	2	6	1	08001	2	08019	3	08019	3	3	3	932763.28	459352.56	15.076739057	623.30835428616	761.585816254478	837.744397
2	0397600303010001	08	019	01	001	93	108	08	019	2	6	1	08004	2	08019	3	08019	3	3	3	932672.31	4593415.2	39.078489541	1225.55634449104	1338.8540539701	1472.7394593
3	0397600303010002	08	019	01	001	84	108	08	200	2	7	4	08830	2	08200	2	08200	2	2	2	932672.31	4593415.2	39.078489541	12161.9748743955	4776.58362494942	14224.3601964
4	039760030320001	08	019	01	001	94	108	08	073	2	4	5	08940	2	08073	2	08073	2	2	2	932669.3	4593404.35	12.693577645	8187.66294017185	1573.38576939154	9158.9387006
5	039760030320002	08	019	01	001	43	108	08	019	2	6	2	08010	1	08019	3	08019	3	3	3	932669.3	4593404.35	12.693577645	1783.72607954253	1989.3878992376	2187.2261289
6	039760030330003	08	019	01	001	73	108	08	019	2	6	3	08013	2	08019	3	08019	3	3	3	932781.06	4593624	61.233763095	2520.94227584662	2807.91183423093	3088.70301765
7	03976003920001	08	019	01	001	113	108	08	019	3	4	2	08040	2	08019	2	08019	2	2	2	932655.21	4593660.39	28.420478989	6409.73131766268	4730.34621424058	12537.13121396
8	03976003920002	08	019	01	001	83	108	08	019	2	6	2	08028	2	08019	3	08019	3	3	3	932655.21	4593660.39	28.420478989	4494.97630289145	4827.14491760767	5309.8594093
9	03976004410001	08	019	01	001	83	108	08	019	3	4	1	08001	2	08019	2	08019	2	2	2	932715.71	4593590.8	26.988569019	613.991980658593	681.179226649901	1246.2407557
10	03976004650001	08	019	01	001	73	108	08	019	2	4	2	08001	2	08019	2	08019	2	2	2	932784.44	4593592.6	33.727852885	625.61417891942	263.152752923091	1183.3895916
11	03976006520001	08	019	01	001	83	108	08	019	2	6	2	08002	2	08019	3	08019	3	3	3	932766.83	4593423.84	45.037992097	689.53924743338	817.663591462066	899.4299506
12	03976006730003	08	019	01	001	13	108	08	019	3	4	4	08001	1	08019	2	08019	2	2	2	932765.88	4593420.66	16.800346951	720.95149311787	634.787149652458	1195.2094710
13	03976006790001	08	019	01	001	93	108	08	019	2	4	3	08003	2	08019	2	08019	2	2	2	932765.88	4593420.66	9.4803748479	1357.02912080867	450.750953371833	1505.5081842
14	03976010340001	08	019	01	001	113	108	08	019	3	6	2	08001	2	08019	3	08019	3	3	3	932836.25	4593706.96	11.811429189	623.062444069471	769.23866661199	846.2230332
15	03977001590001	08	019	01	002	83	108	08	019	2	4	3	08014	2	08019	2	08019	2	2	2	932736.22	4593618.36	19.882017165	2745.54591732338	769.223057156252	3171.19416775
16	03977002080001	08	019	01	002	123	108	08	019	3	4	3	08004	2	08019	2	08019	2	2	2	932649.75	4593611.54	12.5724816	1322.12466039319	1098.62352338947	2324.70962113

Figura 39. Taula amb els resultats del càlcul de rutes agrupats per individu.

En referència a la correcció dels *outliers*, s'adjunten dos casos més en què (una vegada s'ha tornat a segmentar la xarxa) la distància recorreguda a través del graf s'aproxima molt al valor de la distància euclidiana.

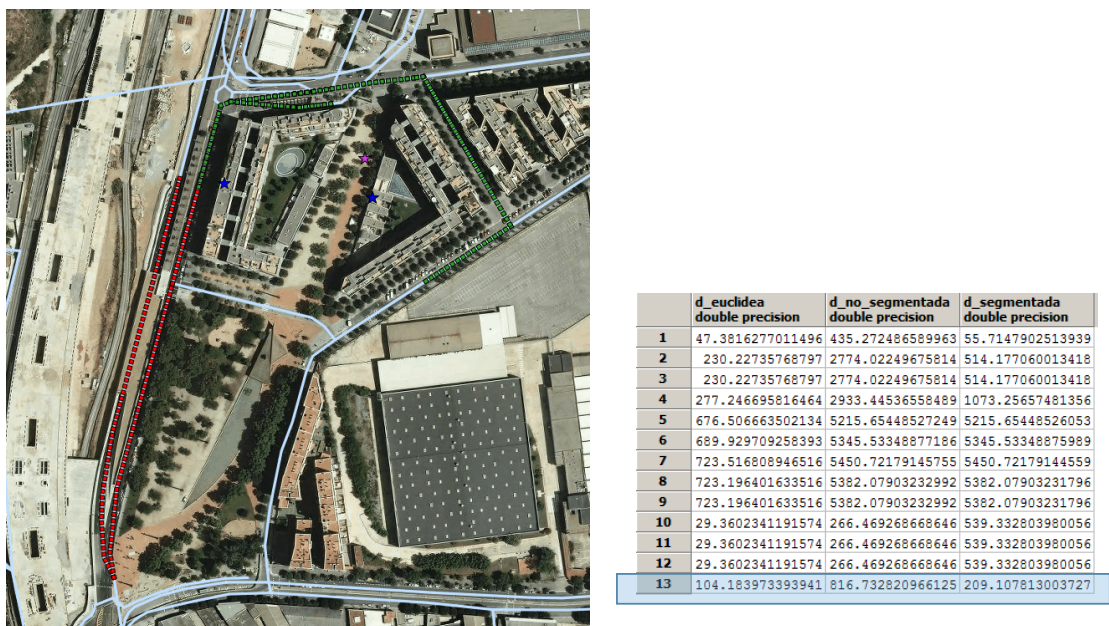
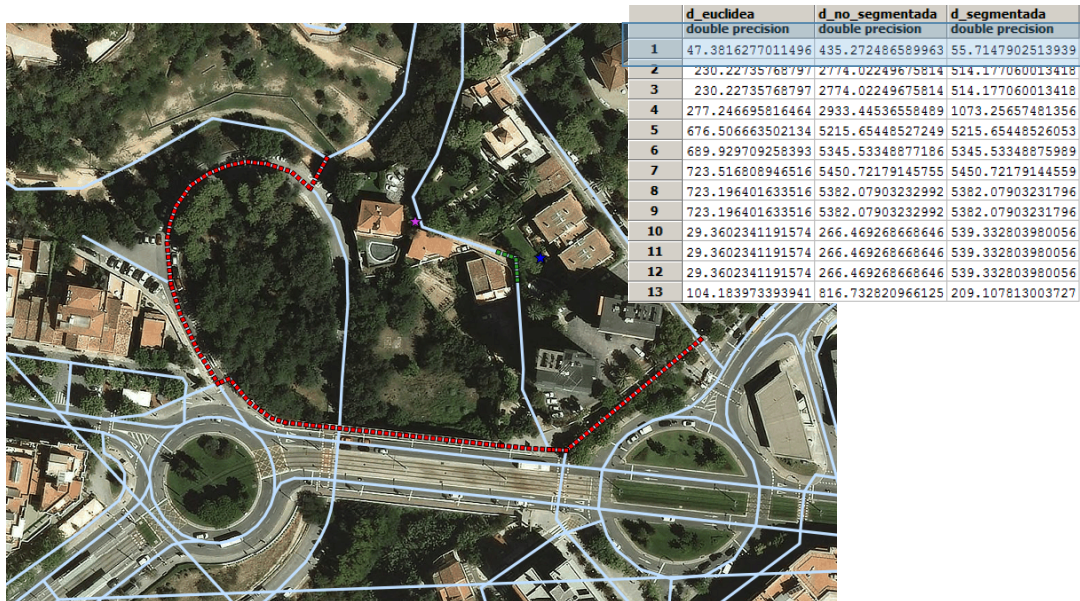


Figura 40. Comparativa de les distàncies: distància euclidiana, distància abans i després de segmentar la xarxa de nou. Font: Elaboració pròpia.