

# Implementació d'un servei de MapMatching online per a traces de camins a Catalunya

Guillem Herrera Masabeu  
Màster en Sistemes de Informació Geogràfica (UPC)  
2017-2018

## INTRODUCCIÓ

La geolocalització integrada als dispositius mòbils ha permès una explosió en la capacitat de emmagatzemar els trajectes personals i/o professionals en format digital. Tot i la gran qualitat dels satèl·lits actuals, i a gratuïtat dels serveis GPS, els tracks elaborats son sempre una aproximació a la trajectòria real que els usuaris han transcorregut. En primer lloc degut a la precisió dels sistemes GPS, que sempre tenen un error associat, i en segon lloc per la metodologia d'obtenció de dades, en que es prioritza l'eficiència energètica dels dispositius obtenint les dades de la posició en rangs de temps determinats (segons o minuts). Els tracks elaborats per les aplicacions GPS sempre presenten una geometria lineal, que va resseguint aproximadament el camí o carretera que el dispositiu va transcórrer.

El MapMatching és la tècnica, o conjunt de tècniques, que permet determinar quina carretera o camí s'estava recorrent quan es va emmagatzemar el track GPS, per posteriorment generar una nova geometria lineal que representi el recorregut real que es va recórrer. La tècnica busca la ruta més probable per a una seqüència de mesuraments de GPS.

En el context del nostre treball, en que es coneix la seqüència completa dels punts GPS, podem definir el MapMatching com el procés d'emparellar una trajectòria GPS sencera amb un graf de carreteres.

## OBJECTIUS

L'objectiu final és la creació d'un servei de MapMatching online, per a la conversió de tracks GPS generats específicament en activitats d'excursionisme i ciclisme (hiking i cicling). Aquest servei ([www.realpath.online](http://www.realpath.online)), disposa també d'una plataforma de registre d'usuaris, on es podrà accedir a tots els tracks processats per cada un dels usuaris i disposar-ne en format GeoJson, així com obtenir una adreça d'internet permanent on poder visualitzar el mapmatching final.

Per a la simplificació dels serveis web necessaris ens limitarem crear el servei per a traces GPS generades en l'àmbit de la península ibèrica.

En el procés de desenvolupament de l'eina, ens hem adonat que tots els algorismes consultats no tenien en compte trajectòries d'un origen que no fos de trànsit rodat, i per tant, no es contemplava que el trajecte real no transcorregués per una carretera o camí cartografiat. En el conjunt d'activitats que nosaltres volem avaluar, es donà freqüentment el cas en que els generadors de traces utilitzen petits corriols o van camp a través. Hem hagut d'implementar la capacitat de detectar aquestes circumstàncies.

Com esmentàvem, la transformació del track GPX a Trajectòria de PostGis és un dels objectius fonamentals. Aquesta tipologia de format vectorial està definida per ser un vector LinesStringM amb un valor M que creix des de el primer vèrtex fins a l'últim. El valor M conté el timestamp associat a la trajectòria. D'aquesta manera es poden realitzar anàlisis molt eficients sobre grans conjunts de dades. Per exemple: «Quines trajectòries van transcórrer en aquest període de temps

determinat?», «Quines trajectòries circulen a menys de x metres d'aquest equipament?», «Quines trajectòries van realitzar-se per aquesta carretera en aquest període de temps?». Sens dubte aquest format vectorial presenta un gran potencial d'anàlisi en els temps del BigData.

Un altre objectiu primordial del projecte és usar software lliure i recursos lliures. La cerca en l'ús d'aquestes tecnologies respon a la inquietud personal de l'autor, i la convicció de la potencialitat de les eines desenvolupades en aquest context. D'aquesta manera, la cartografia base per a la realització del MapMatching és obtinguda del projecte Open Street Maps (OSM), el desenvolupament i execució de l'algoritme serà en base de dades Postgres/Postgis, i el backend que dona suport al servei web està implementat en Django.

El codi de bona conducta de les eines que pertanyen al context del Free and Open-Sources Software (FOSS), recomana retornar l'ús que se'n fa, lliure i gratuït, a la comunitat d'usuaris. En aquest sentit hem fet molts esforços en poder localitzar els passos i senders no cartografiats que hem comentat amb anterioritat. D'aquesta manera, amb un volum suficientment gran de trajectòries emmagatzemades, serem capaços de detectar tots aquells camins que encara no estan cartografiats per la comunitat OSM, i contribuir a millorar la cartografia lliure existent.

## METODOLOGIA

El projecte presenta dos línies de treball principals. Paral·leles, tot i que no independents.

Per una banda el desenvolupament, testeig i millora del propi algoritme de Mapmatching. Basat en traces GPS i enfocat a recorreguts obtinguts en activitats de ciclisme i excursionisme.

Per altra banda el desenvolupament del backend necessari per a la gestió i processament dels arxius pujats pels usuaris a la plataforma realpath.online.

### Base teòrica

Actualment existeixen multituds de algorismes per a la realització del MapMatching (MM). Cal tenir en compte que la tendència actual és dissenyar algorismes per processar dades GPS en viu. Gràcies a la capacitat de processament dels dispositius mòbils actuals això ja és més que possible, i intenta respondre a la voluntat de generar serveis de routing online i en viu per als usuaris d'aplicacions mòbils. Evidentment un bon algoritme MM ( gran fiabilitat en determinar la teva posició real), es fonamental per poder calcular correctament rutes optimes de transport.

En el nostre cas hem desenvolupat un algoritme per a processar les dades generades a posteriori.

Això té certes avantatges, tal com mostrarem més endavant.

Actualment els algorismes de MM es divideixen en diferents categories :

- Geomètrics (gMM): Necessiten de poca informació, però de poca fiabilitat en casos de conjuncions de carreteres o existència de carreteres paral·leles.

- Topològics (tMM): Incorporen criteris històrics (carretera assignada als punts anteriors, ...), velocitat del vehicle o criteris topològics del segments del graf en ús (connectivitat entre segments).
- Probabilístics (pMM): Ús de l'estudi de la estadística per identificar els segments candidats per als punts.
- Avançats (aMM): Ús d'algoritmes avançats per a la identificació dels segments del graf, com ara Xarxes Neurons Artificials (ANN).

L'algoritme desenvolupat en el present treball es basa en els treballs VELAGA et al. i KOLLER et al. El primer d'ells és un algoritme de base topològica (tMM), i el segon probabilístic (pMM), basat un model Hidden Markov Model (HMM) simplificat. El treball substitueix la clàssica resolució del HMM amb l'algoritme de Viterbi per un model basat en un algoritme Dijkstra Bidireccional.

Podem considerar tanmateix que el nostre algoritme es de categoria topològica, ja que si bé hem utilitzat alguna eina publicada per KOLLER et al. per a identificar els segments candidats més probables, es realitza només per a l'avaluació seqüencial de cada un dels punts i no per la totalitat dels punts de la trajectòria. En aquest sentit, l'algoritme que presentem realitza un seqüència de processament i detecció de segments dels graf més probable per a cada un dels punts ( com ho fa VELAGA et al.), i no un anàlisi complet d'estimació de la millor trajectòria possible ( com KOLLER et al.).

## Algoritme

### **Descripció**

En el nostre cas, el core de l'algoritme està desenvolupat íntegrament en llenguatge PL/pgSQL de la Base de Dades PostgreSQL. Hem creat un conjunt de funcions que executen tota la seqüència d'encadenaments lògics que conformen l'algoritme en si mateix. En el cas de la plataforma realpath.online que hem desenvolupat, l'execució de l'algoritme està gestionat en Django responnent a la interacció dels usuaris de la plataforma (com expliquem més endavant).

A nivell esquemàtic, el servei rep un arxiu GPX de l'usuari, el processa, i genera i emmagatzema una trajectòria nova, que «serveix» als usuaris en forma d'arxius GeoJson i enllaç a un visor de mapes interactiu. Tantmateix, l'algoritme es pot executar independentment i ser usat per al tractament d'arxius a nivell local, o de forma dedicada en servidor per a altres tipus d'estudis.

Sempre serà necessari, però, adequar les dades contingudes als arxius GPX, com també les dades provinents del projecte OSM.

## Preprocessament

En primer lloc, degut a la naturalesa erràtica i els inherents errors associats a la tecnologia GPS, és necessari realitzar un preprocessament de les dades per tal de descartar els outliers (mostres no significatives degut a un grau massa gran d'error), punts duplicats, i parades (stops) en el transcurs de la trajectòria del track GPS. En aquest sentit, és de nou imprescindible que els tracks GPX processats tinguin associats un timestamp en cada un dels trackpoints que el componen.

Aquest preprocessament és absolutament necessari. Si bé la presència de outliers no és un fenomen molt freqüent (  $\approx 0,05\%$  de les mostres analitzades), si que és significatiu. Un sol punt de la trajectòria desviat uns centenars de metres de la seva posició real, altera en gran mesura el traçat final de la trajectòria en un segment important d'aquesta.

Un cop la informació bàsica del GPX, està carregada a la base de dades (geolocalització, i timestamp), processem les dades en els següents passos:

1. Detecció i eliminació de punts duplicats.
2. Càlcul de distàncies i velocitats (km/h) per cada un dels punts respecte el seu precedent
3. Detecció de parades en la trajectòria (Stops).
4. Detecció i eliminació de outliers.

Tenim en compte tant les velocitats i distàncies relatives de cada un dels punts (amb el seu precedent), com les mitjanes i desviacions estàndard absolutes del conjunt de dades. Considerem que un punt té massa error associat si s'allunya massa de la distribució normal de la mostra.

## Funcionament

En termes generals, l'algoritme avalua per cada un dels punts de la trajectòria, un cert numero de segments (edges) candidats, i selecciona aquell que més probabilitats té de ser el correcte. Pera cada emparellament (match) s'obté el tram de graf entre aquest i el match anterior, i se'n assigna el timestamp corresponent.

El numero de possibles candidats (treshold de recerca) a emparellar l'hem establert en 5 per al nostre algoritme, ja que empíricament hem trobat que ens produïa molt bons resultats.

Seguint el treball de VELAGA et al., l'algoritme s'inicia localitzant l'origen del primer punt de la seqüència GPX, basant-se en la **direcció (heading)** entre els dos primers punts de la trajectòria, la direcció dels segments de grafs més propers, i la **distància ortogonal (D)** entre els el punt i els segments de graf. Els segments de grafs que tinguin una orientació més similar a la direcció de la traça GPS, obtindran un **pes per direccionalitat** major ( $W_h$ ). De la mateixa manera, aquells segments de graf més propers al punt GPX tindran un **pes per distància** major ( $W_d$ ).

Establim el paràmetre **TWS (Total Weight Score)** definit per:

$$TWS = W_h + W_d \rightarrow TWS = (H_w * f(\theta)) + (D_w * f(D))$$

on:

$H_w$  es el coeficient del pes per direccionalitat

$D_w$  és el coeficient del pes per distancia

$$f(D) = \frac{(80-D)}{80}$$

$$f(\theta) = \cos \theta$$

$\theta$  angle (radians) entre la direcció del segment del graf i la direcció del traçat GPX.

Un cop seleccionat el millor segment candidat (TWS més elevat), es calcula el punt més proper de la recta al trackpoint del GPX, i se li assigna el valor timestamp del punt GPX associat. Aquest serà el primer punt de la nova trajectòria generada.

Cal tenir en compte, per això, diversos escenaris:

- El trackpoint és una parada/Stop:

Derivat del preprocessament, si un trackpoint és considerat una parada, se li assigna el seu timestamp, però la projecció del punt anterior.

- El trackpoint anterior s'ha considerat fora del graf (breakpoint).

En aquest cas es valora, en funció a la distancia al segment de graf més proper, si considerem que el trajecte torna a transcórrer per un camí present al graf. Per al present estudi el llindar es fixa als 5 metres. Si està sota el llindar, s'agafa el punt més proper del segment com a vàlid, amb el seu corresponent timestamp. En cas contrari, es pren la posició del trackpoint en «brut», i es segueix considerant fora del graf (breakpoint).

- Trackpoint apte per al mapmatch:

Si el punt trackpoint a avaluar ha passat els dos filtres anteriors, es realitza primer un **anàlisi de proximitat topològica** als segments mes propers del graf.

Per a cada un dels possibles candidats, s'analitza si són el mateix segment seleccionat que el trackpoint anterior, si estan directament connectats per un node (1 hop) o si la seva distancia topològica és més gran.

En funció d'aquesta distància s'assignen diferents paràmetres als càlcul del TWS. Si no estan directament relacionats el TWS és calculat de la mateixa manera que per el MapMatch inicial.

En els altres casos el TWS es veu modificat afegint un paràmetre de probabilitat ( $W_p$ ). Aquest està inspirat en KOLLER et all, i té en compte la relació entre la distància del trackpoint avaluat i el immediatament anterior ( $\delta(Z_{i-1}, Z_i)$ ) amb la distancia del trajecte entre el punt calculat anterior i el punt mes proper al segment candidat ( $\varphi(X_{i-1}, X_i)$ ). Hem adaptat aquest concepte a les nostres necessitats, definint  $W_p$  com:

$$W_p = P_w * f(\delta, \varphi)$$

on :

$$f(\delta, \varphi) = 1 - \left| 1 - \frac{\delta(Z_{i-1}, Z_i)}{\varphi(X_{i-1}, X_i)} \right|$$

Donant com a resultat:

$$TWS = W_h + W_d + W_p \rightarrow TWS = (H_w * f(\theta)) + (D_w * f(D)) + (P_w * f(\delta, \varphi))$$

En aquest punt es realitza un primer anàlisi de detecció de breakpoints. Si el el candidat seleccionat està directament connectat, i presenta una distancia major de 30 metres del segment seleccionat, es considera que és un breakpoint. Es pren la geometria en brut del GPX com a vàlida i es passa a analitzar el següent trackpoint. En cas contrari, es projecta sobre el segment candidat.

Ara bé, el cas més complex d'anàlisi en l'algoritme proposat és aquell en que el candidat amb un TWS més elevat està a una considerable distància topològica amb l'anterior punt processat.

En aquest cas, per tal de determinar la ruta transcorreguda entre els dos punts, es necessari realitzar un anàlisi de camí més curt entre els punts.

Malauradament la base cartogràfica de OSM no està té una topologia adequada per a l'anàlisi de routing. Seria necessària un conjunt de arcs i nodes en cada una de les interseccions reals entre ells, així com una relació definida de node d'origen i node final per a cada un dels segments.

Així doncs, hem creat una funció pròpia, sostinguda sobre PgRouting, en que creem un graf local i acotat per a tots els casos necessaris. Es seleccionen un conjunt adequat de segments entre el trackpoint a avaluar i el punt processat anterior de forma dinàmica, ja que la distancia entre ells pot variar molt en cada cas avaluat.

Cal destacar que aquest procés de generació de graf local és la part de l'algoritme que més temps de processament necessita. És el nostre coll d'ampolla (bottleneck).

Ara de nou, basant-nos en el treball de KOLLER et all, tindrem en compte  $\delta (Z_{i-1}, Z_i)$  i  $\varphi (X_{i-1}, X_i)$ . En canvi utilitzarem aquest paràmetres ara per valorar el cost dels trajectes possibles (de cada un dels candidats). Tal com expliquen al article:

$$w(u, v) = e(u, v) + n(v)$$

$$e(u, v) = \beta * \frac{\varphi(X_{i-1}, X_i)}{\delta(Z_{i-1}, Z_i)}$$

on:

u: node anterior (trackpoint anterior processat)

v: node candidat (Projecció del trackpoint al segment candidat)

$\beta$ : Paràmetre de calibratge.

w(u, v): cost total de trajecte entre els nodes de graf u i v

e(u, v): cost de realitzar el trajecte per el segment de graf comprés entre u i v

n(v): cost de realitzar el trajecte a través de v

$\delta (Z_i, x_i)$ : Distància entre el trackpoint i la seva projecció sobre el segment candidat

El candidat escollit ara serà aquell que presenti un cost de trajecte menor ( $w(u, v)$ ). Hem establert empíricament el valor de  $\beta$  en 30.

## Construcció de la trajectòria

En primer lloc es calcula el segment de graf que transcorre entre els dos punts consecutius, i s'interpolava linealment el valor de timestamp a tots els vèrtexs de la geometria, de forma lineal.

Es a dir, per a un suposat segment, amb temps inicial i final 1s i 2s respectivament, que presenta 3 nodes (inicial, intermedi a un 40% del la seva longitud, i un node final), interpolem el valor temporal al node 2 i hi assignaríem un valor de 1,4s.

Aquest procés es pot realitzar sense que importi el numero de nodes que presenti la geometria lineal. Finalment, realitzant un merge entre tots segments de graf extrets ( i LinesString en «brut» als segments fora de graf), obtenim la geometria LinesStringM final.

## Servei online realpath.online

Paral·lelament al desenvolupament de l'algoritme, hem creat un servei online de tractament dels arxius GPX dels usuaris. Aquest servei, encara en procés de desenvolupament i millora, permet als usuaris pujar un arxiu GPX que serà processat per el nostre algoritme en el backend, i es crearà un enllaç perenne i públic on es pot visualitzar el recorregut generat de forma dinàmica. Si es tracta



d'un usuari registrat, podrà accedir posteriorment a tots els recorreguts que hagi processat, i disposarà d'un enllaç de baixada del resultat del processament en format GeoJson. L'arxiu GeoJson disponible (LineString) conté tres variables per a cada punt que conformen la línia: latitud, longitud i timestamp en format epoch, projectats al CRS WGS84. Hem hagut de generar la geometria en aquest format degut a que no està definit el format LinesStringM en les especificacions oficials que defineixen aquest format (RFC 7946). De totes maneres el format personalitzat és acceptat de forma correcta pel programari GIS d'escriptori, i se'n pot extreure amb posterioritat el valor temporal associat.

## CONCLUSIONS

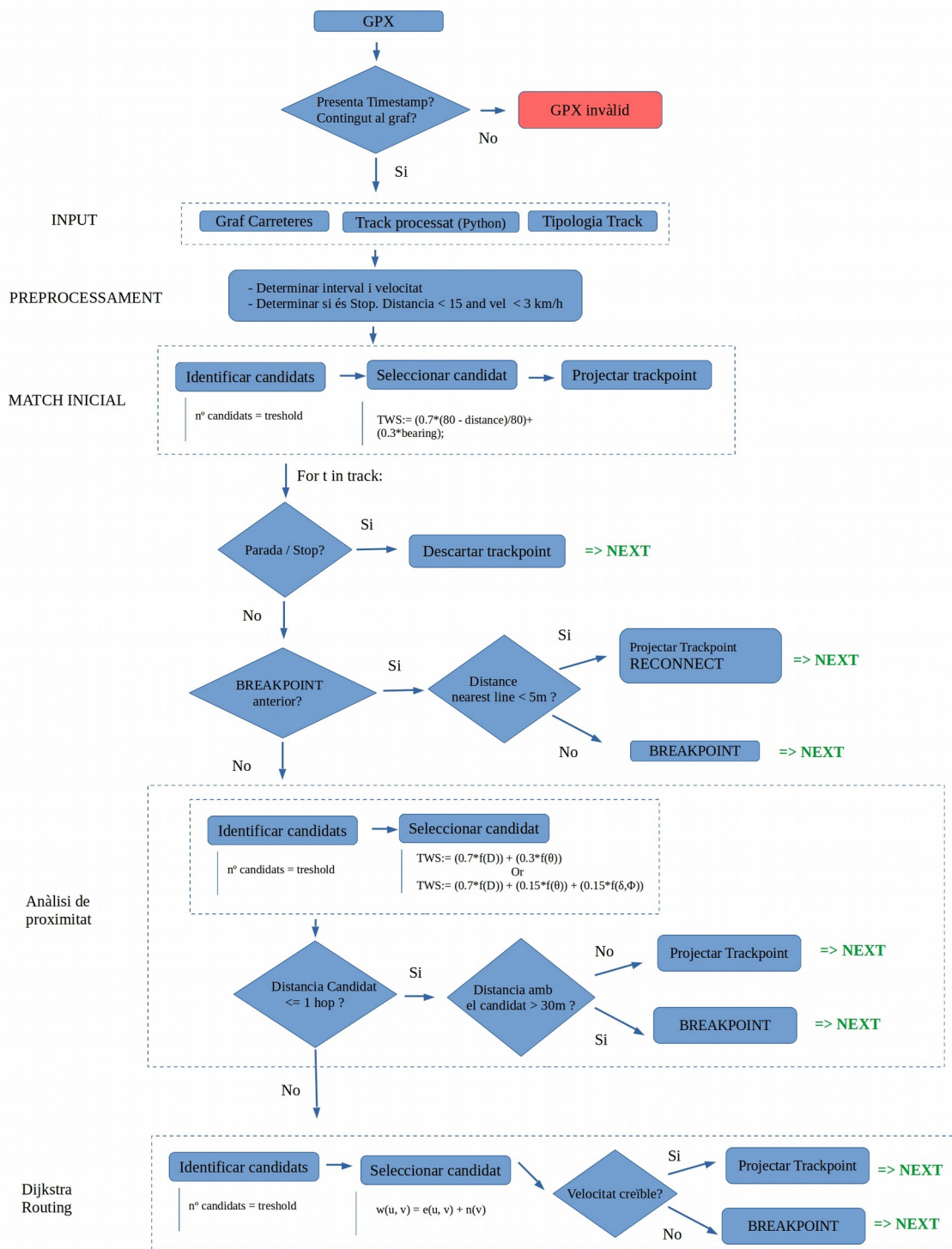
Trencant el tòpic de major eficàcia que pesa sobre el software privatiu, el software lliure és la punta de llança dels avanços tecnològics actuals, sobretot en l'àmbit del Big Data, com ho demostra el gran conjunt d'eines lliures en base R i Python que s'utilitzen avui en dia en els projectes més avançats. En el context GIS, també la BBDD PostgreSQL/PostGIS és una eina inigualable en la gestió de grans volums de dades geogràfiques. Aquest treball demostra com les eines FOSS no han de patir l'estigma de tenir baixa funcionalitat, és més, presenten una major versatilitat i flexibilitat, que les fan ideals per qualsevol tipus de projecte.

L'algoritme desenvolupat és una eina que ens permet millorar la cartografia lliure present a OSM, gràcies a la col·laboració de tots els possibles usuaris de la plataforma realpath.online. D'aquesta manera l'ús de dades lliures es retroalimentarà i es contribuirà a la millora de la cartografia existent. El propi usuari és converteix d'aquesta manera en generador de dades, que seran usades per a la millora de la comunitat, i podrà gaudir i compartir una visualització online moderna de les seves activitats esportives.

En qualsevol projecte que hagi de tractar un gran volum de dades, sempre és necessari un primer preprocessament i neteja de les dades. Així doncs, aquest algoritme també és una eina per a la gestió del Big Data geogràfic en el context de les dades preses amb tecnologia GPS.

## BIBLIOGRAFIA

- GRASER A. 2018. Evaluating Spatio-temporal DataModels forTrajectories in PostGISDatabases. DOI:10.1553/giscience2018\_01\_s16
- KOLLER H., WIDHALM, P., DRAGASCHNIG, M., GRASER A. 2015. Fast Hidden Markov Model Map-Matching for Sparse and Noisy Trajectories
- VELAGA, N.R., QUDDUS, M.A. and BRISTOW, A.L., 2009. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems. Transportation Research Part C:Emerging Technologies, 17 (6), pp.672-683.
- Zheng, Y., Xie, X., and Ma, W.-Y. (2010). GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory. IEEE Data Eng. Bull., 33(2):32\_39.



Il·lustració 1: Diagrama de fluxos de l'algorithm proposat