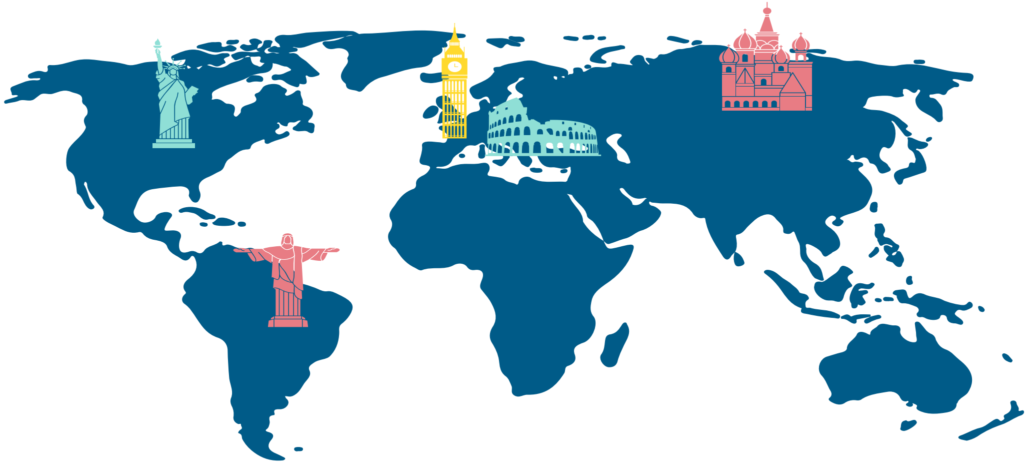
Estudio de métodos de detección de patrones de movimiento para sistemas de recomendación turísticos.

Sergio Torrijos López de la Manzanara



Escuela Politécnica Superior

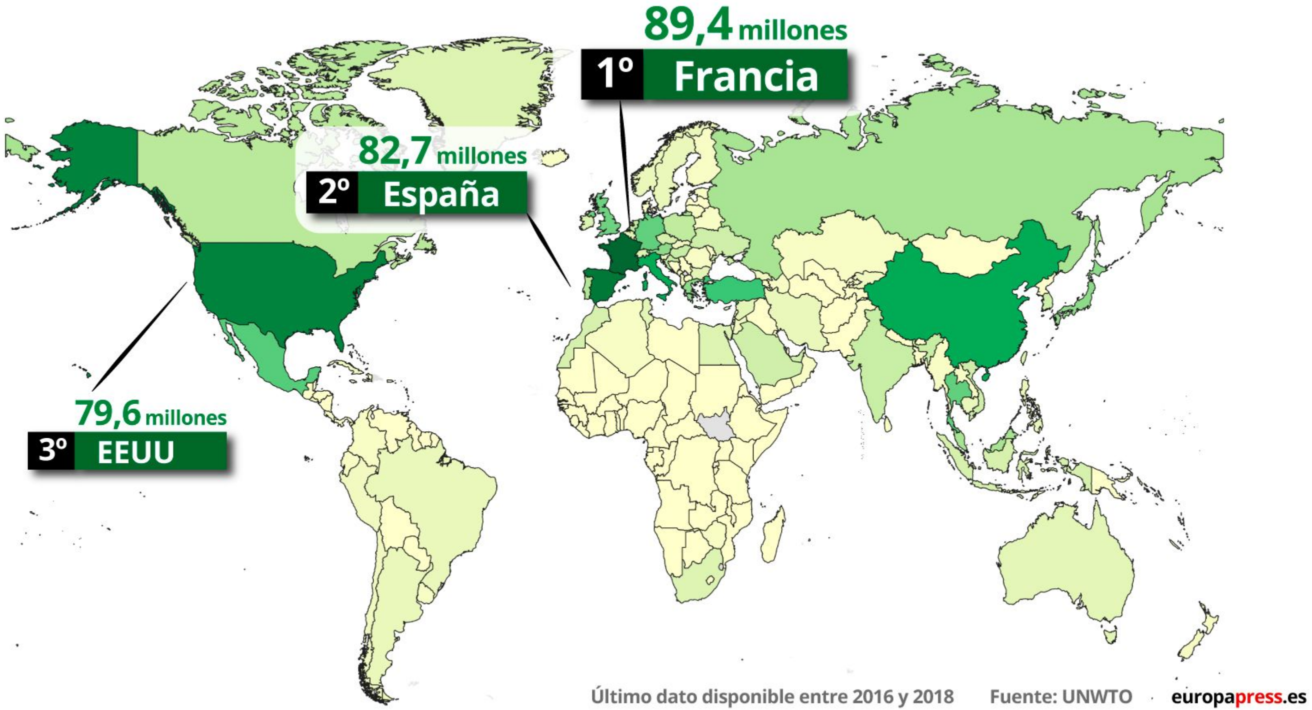
Universidad Autónoma de Madrid

TABLA DE CONTENIDOS

1. INTRODUCCIÓN

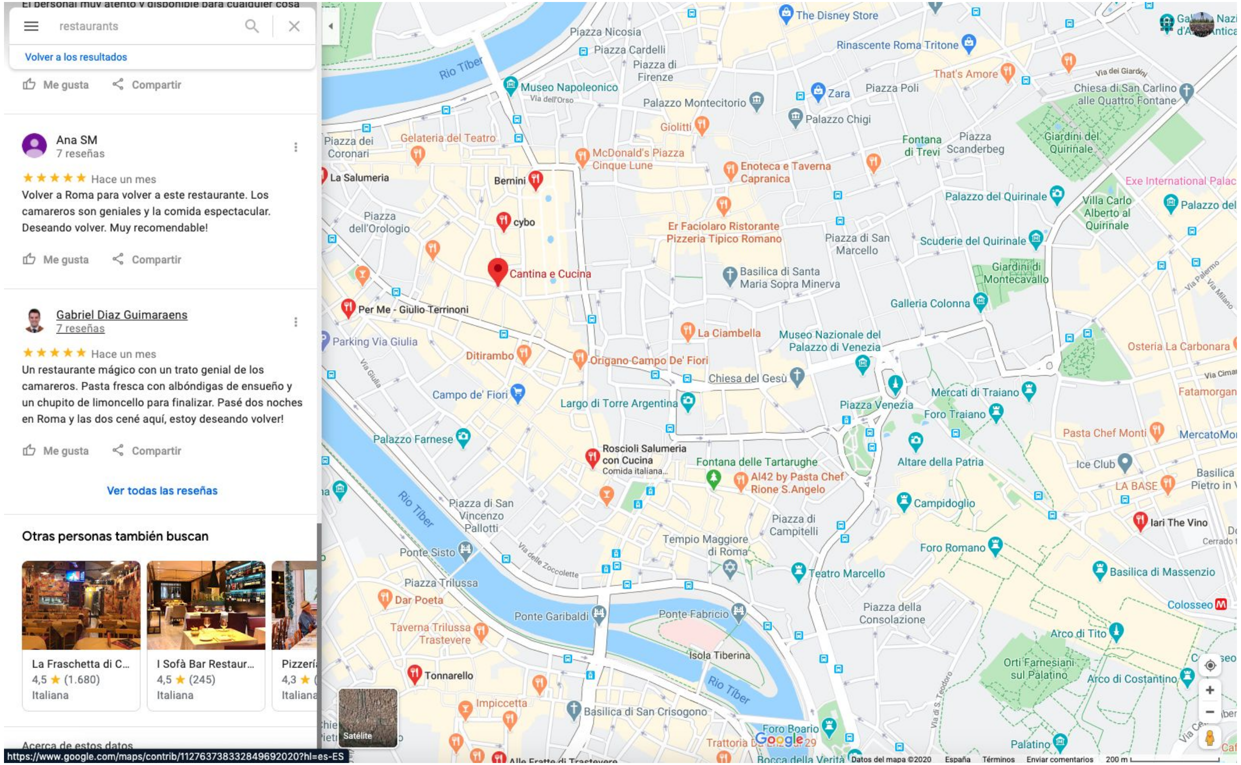
1. INTRODUCCIÓN

Datos turismo mundial



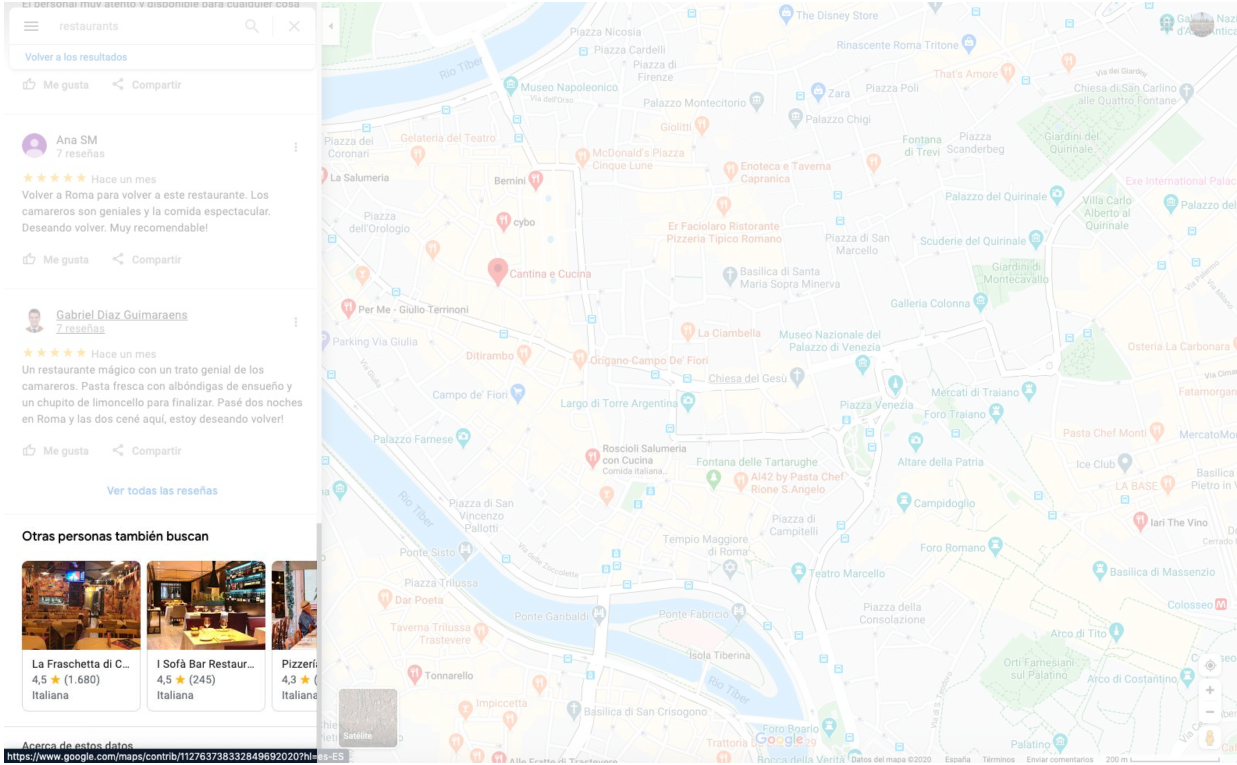
1. INTRODUCCIÓN

Sistemas de recomendación turísticos

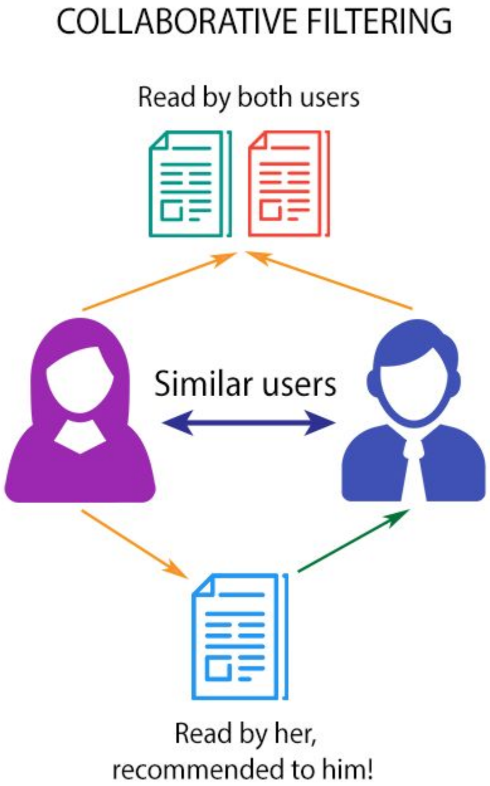


1. INTRODUCCIÓN

Sistemas de recomendación turísticos

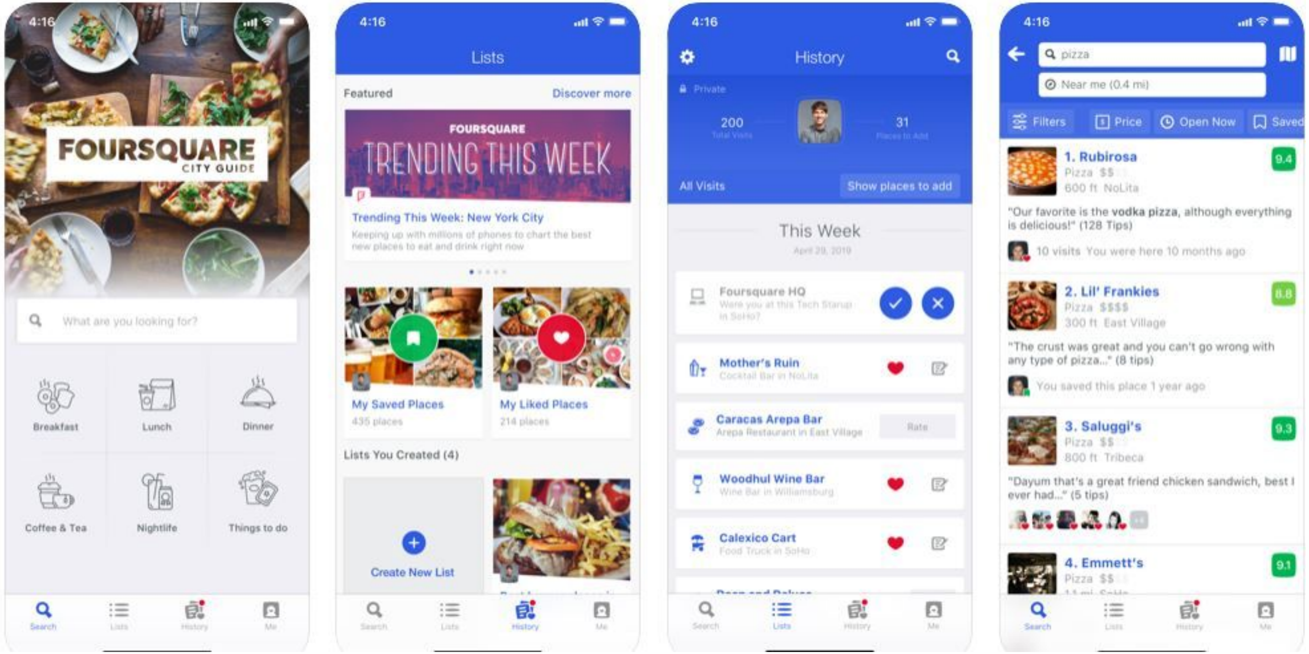


Sistemas de recomendación



1. INTRODUCCIÓN

Sistemas de recomendación turísticos



1. INTRODUCCIÓN

¿Bandada de pájaros = Patrones de movimiento?

1. INTRODUCCIÓN

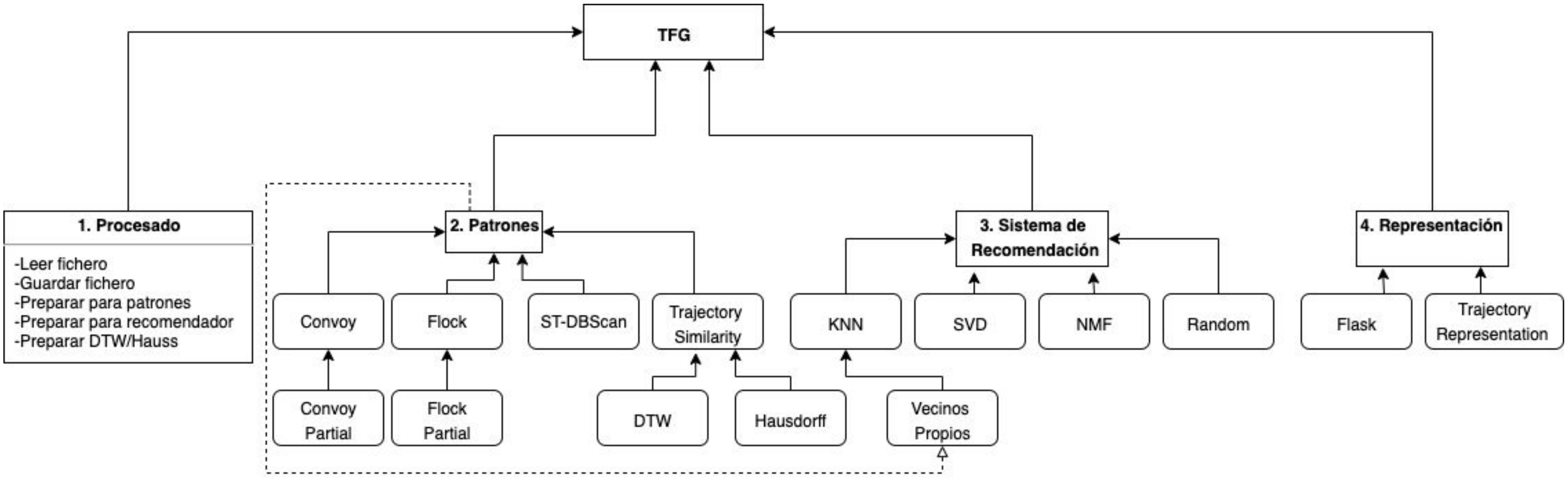
Tema de estudio

Estudiar, aplicar y adaptar los patrones de movimiento para los sistemas de recomendación.

2. DISEÑO E IMPLEMENTACIÓN

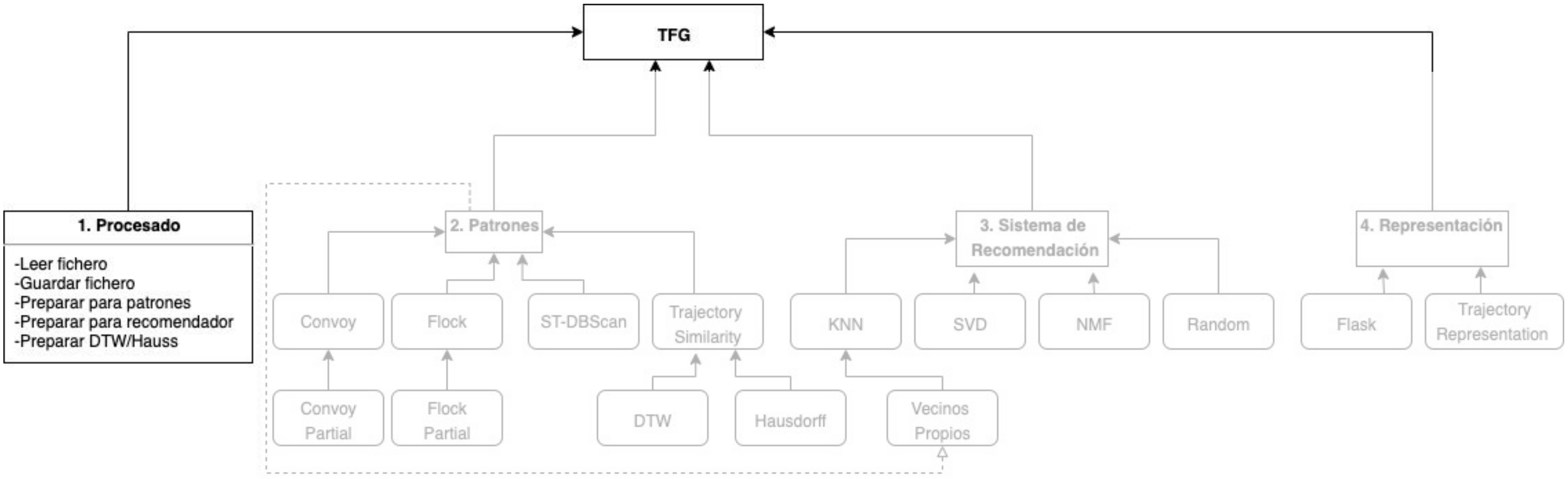
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



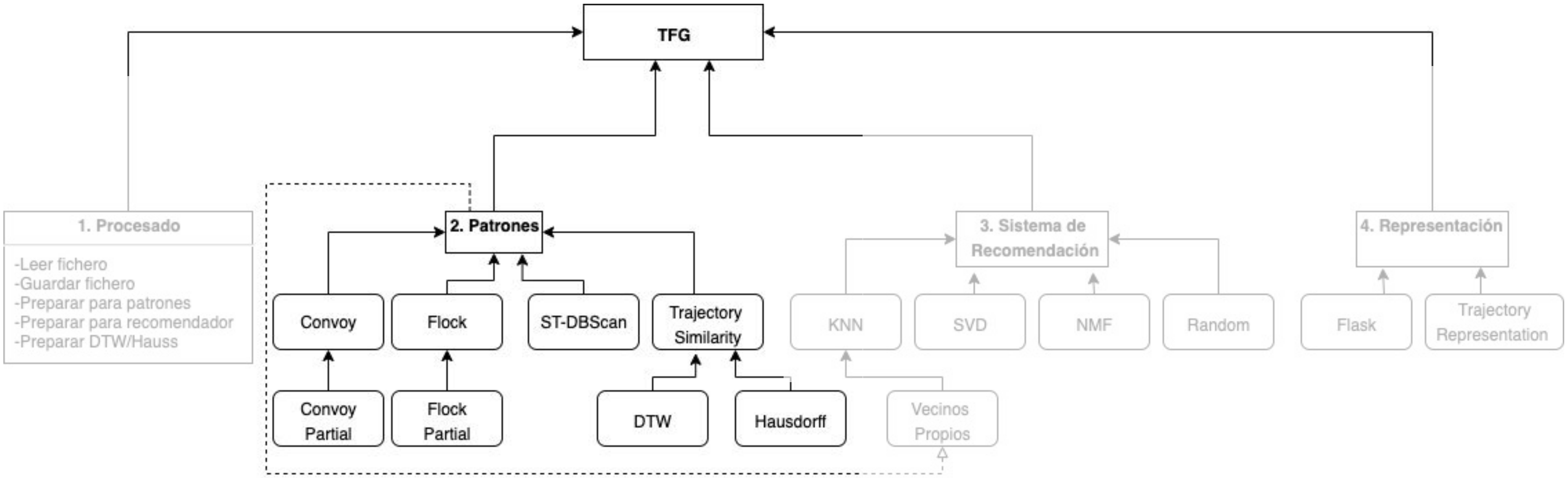
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



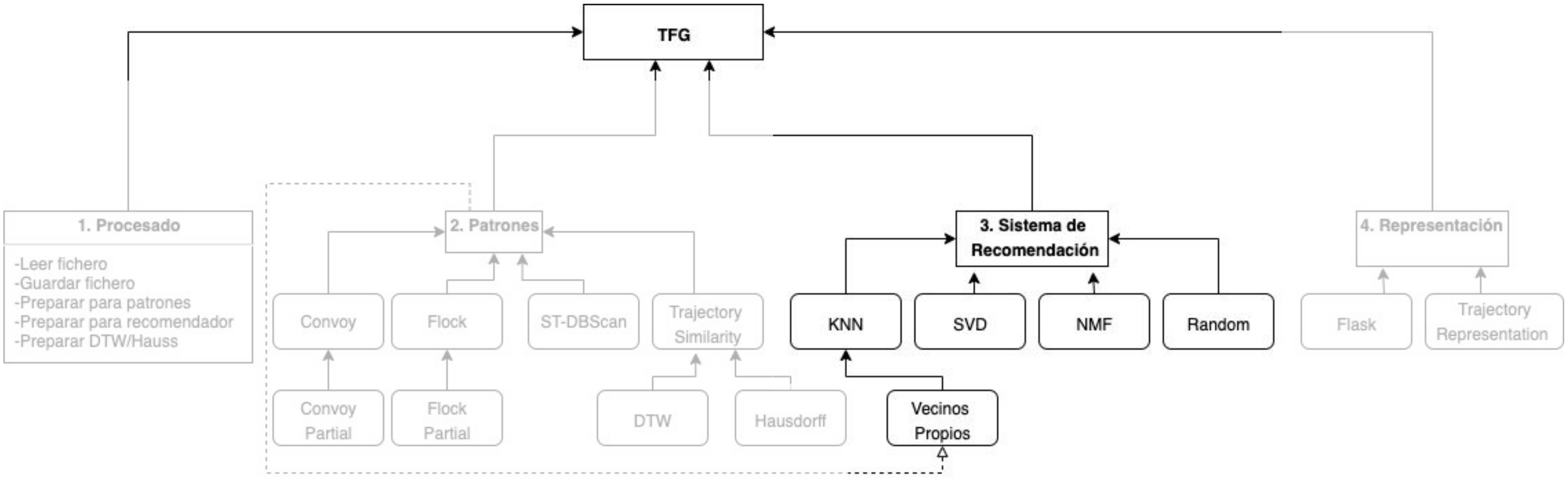
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



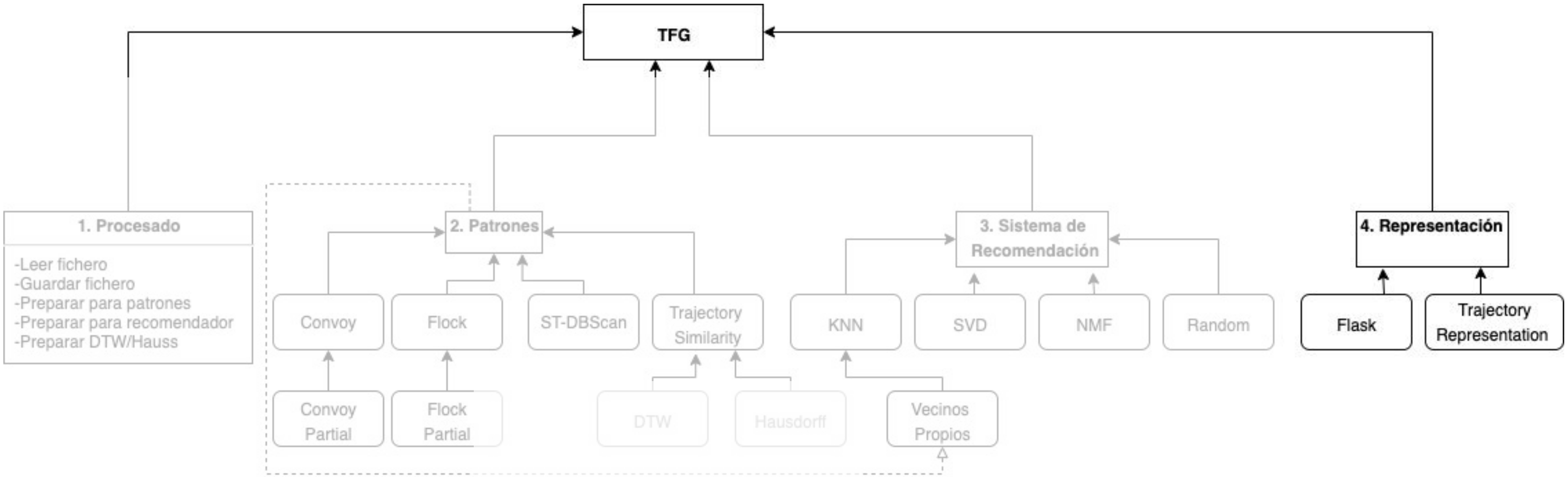
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



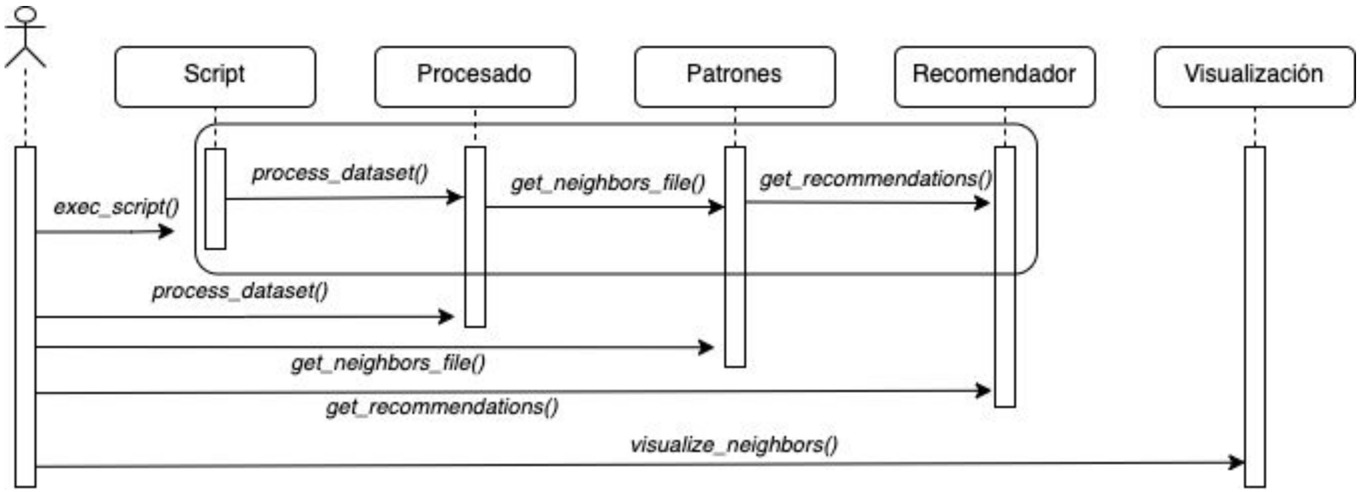
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



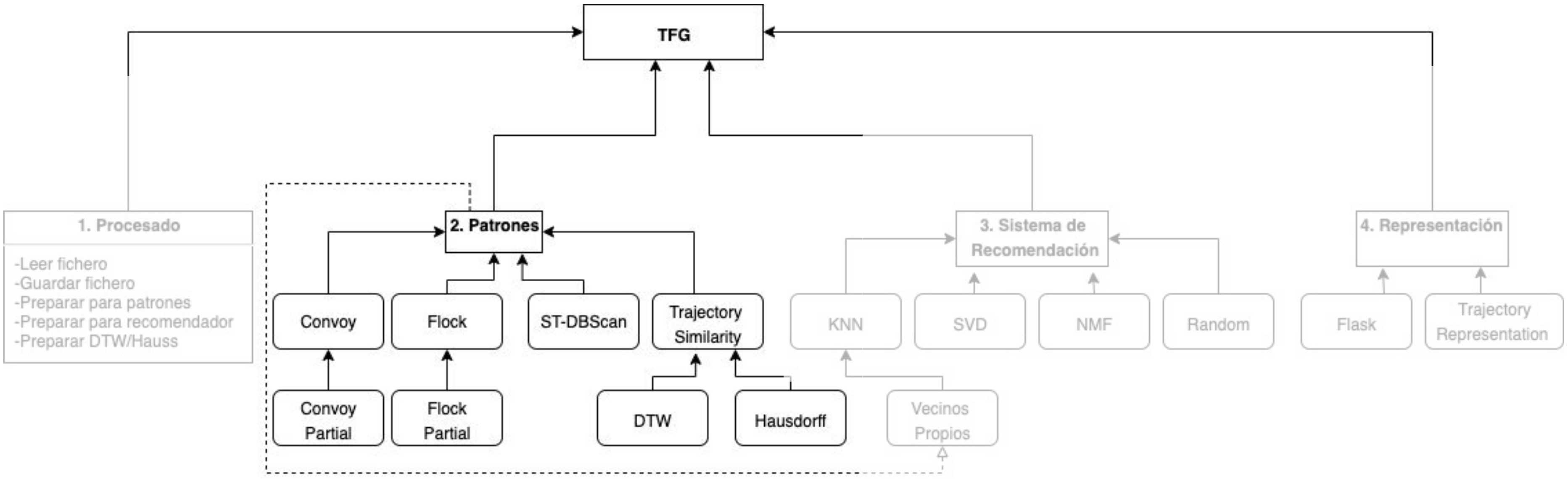
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Diagrama de secuencia



2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



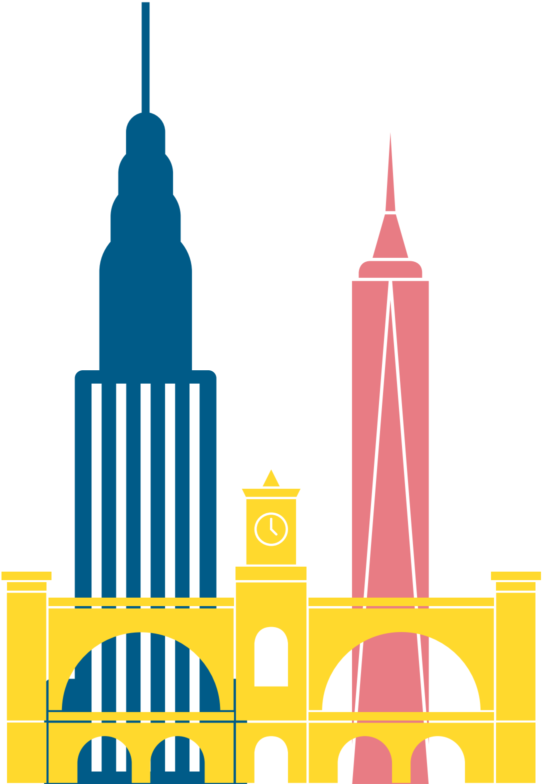
2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento

Flock

Convoy

ST-DBSCAN



Hausdorff

Dynamic Time Warping

2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento - Dynamic Time Warping

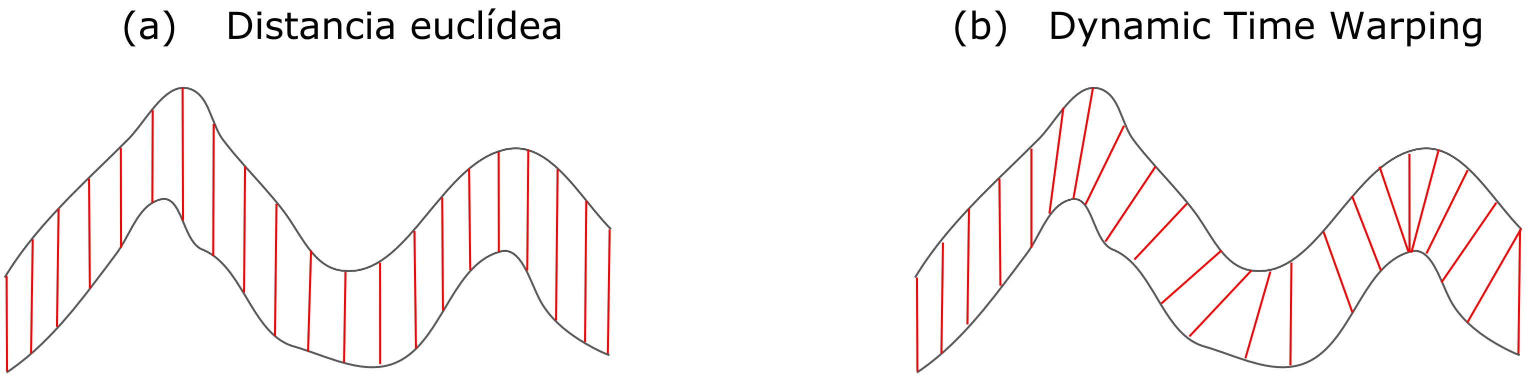
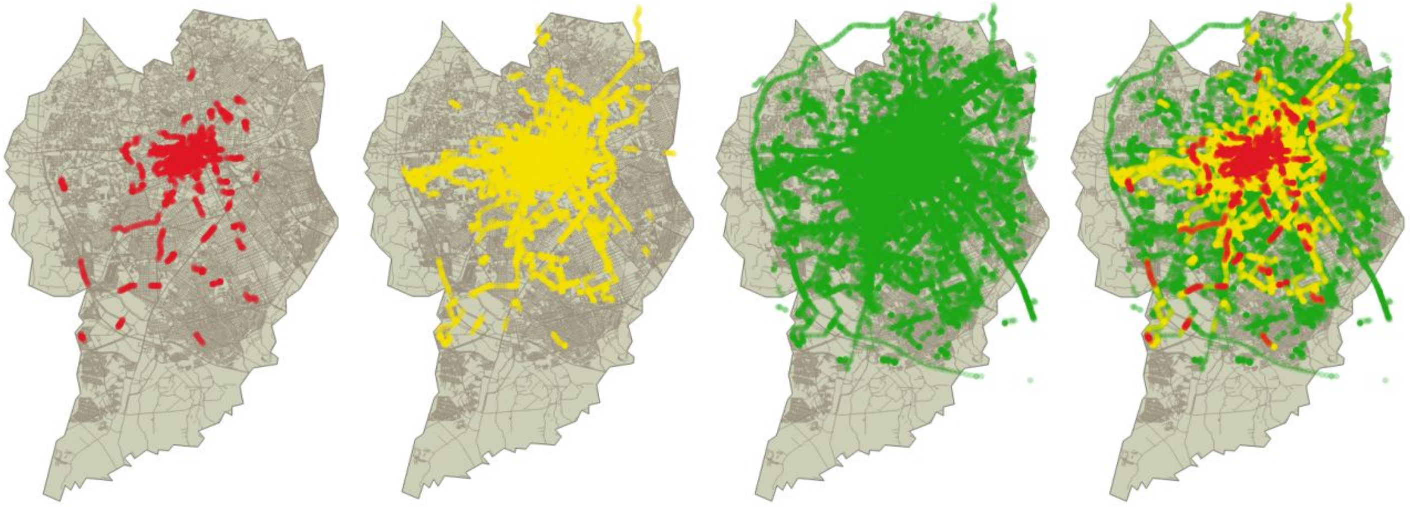


Figura 2.5: Comparación entre dos secuencias: (a) mientras la distancia euclídea es rígida en el tiempo, (b) Dynamic Time Warping (DTW) es flexible en el tiempo para tratar la posible distorsión de tiempo entre las secuencias [10].

2. DISEÑO E IMPLEMENTACIÓN

Implementación - ST-DBSCAN

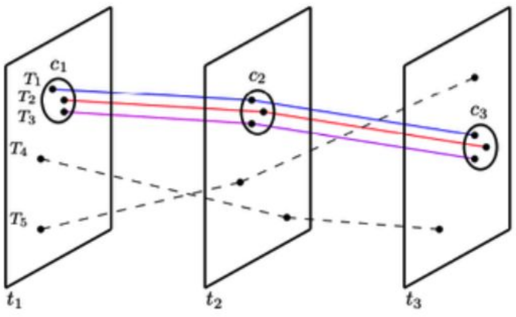


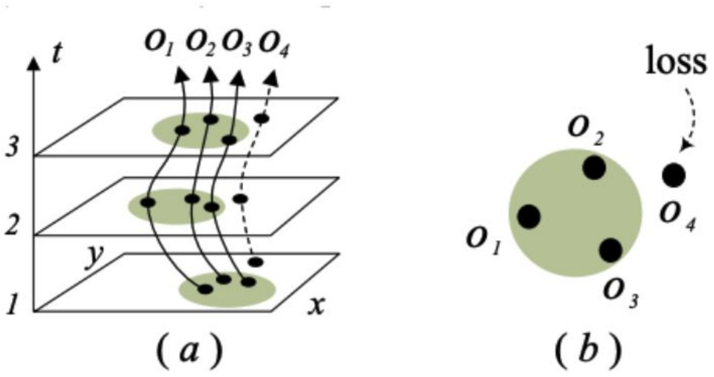
(a) Eps1 = 100m (b) Eps1 = 166m (c) Eps1 = 330m (d) Eps1 = 330, 166, 100m

Birant, D. and Kut, A. (2007). St-dbscan: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering, 60(1):208 –221. Intelligent Data Mining.

2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento - Flock





(a) Trayectoria Flock [7] donde se muestran cómo se escogen los puntos para la trayectoria en función de si cumplen el criterio del radio del disco.

(b) Demostración de la pérdida de elementos ocasionada por Flock [13], donde se aprecia que el punto *O4* no entra en las restricciones de Flock pese a seguir una trayectoria similar.

2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento - Convoy

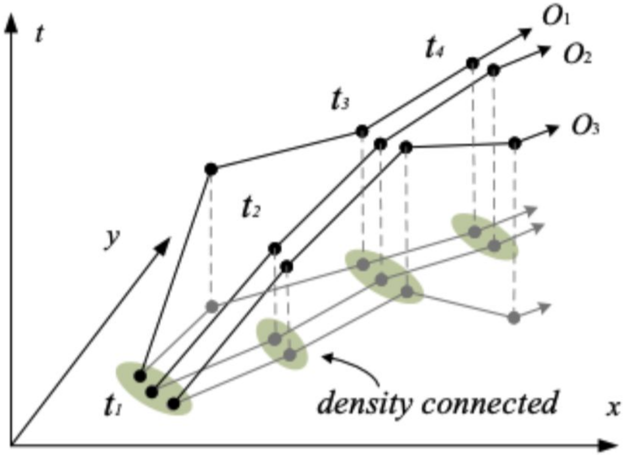


Figura 2.2: Ejemplo de Trayectoria Convoy [13], marcando en verde los discos que agrupan los puntos y las líneas marcando la trayectoria.

2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento - Librerías utilizadas

Pandas

Surprise

Flask

LCM

Coherent Moving Cluster

py-st-dbscan

FPFlock

2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento - Librerías utilizadas

Librerías implementadas y adaptadas a la caracterización de recomendación turística.

Implementación optimizada de Flock: Partials

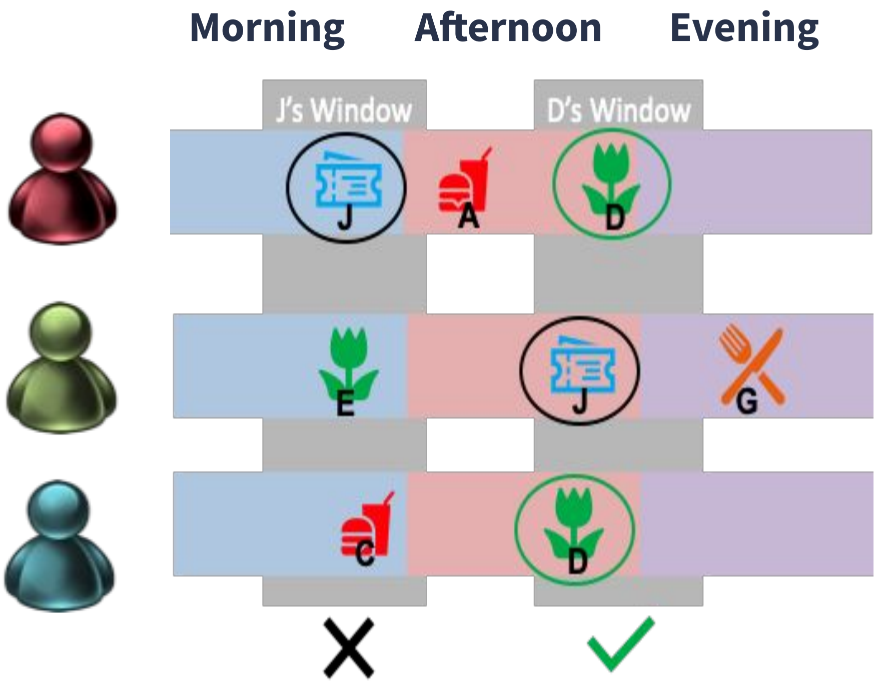
Ejecución en 2 fases:

1.– Calcular la similitud de todos los usuarios mediante un método rápido (DTW o ST-DBSCAN).

2.– Sobre los resultados obtenidos, coger los k usuarios más similares a nuestro usuario, y ejecutar Flock iterativamente la lista reducida.

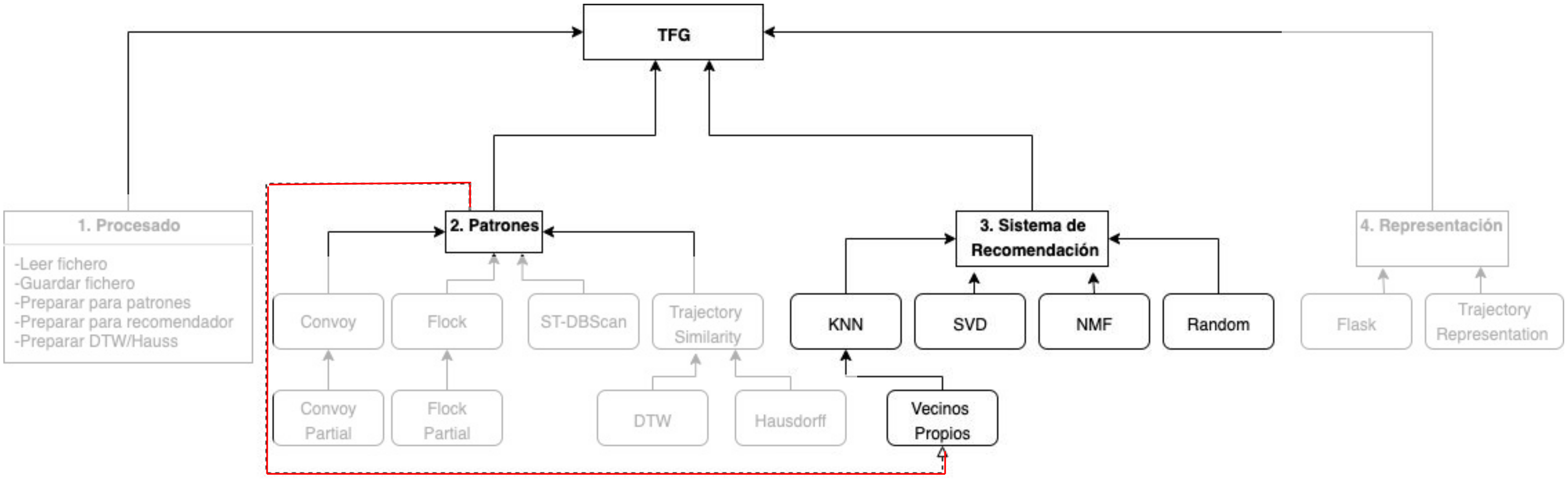
2. DISEÑO E IMPLEMENTACIÓN

Nuevo patrón de movimiento: Ad-Hoc



2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



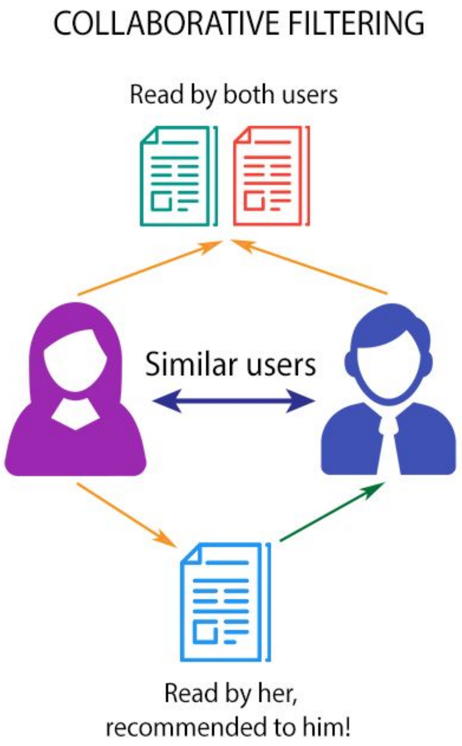
2. DISEÑO E IMPLEMENTACIÓN

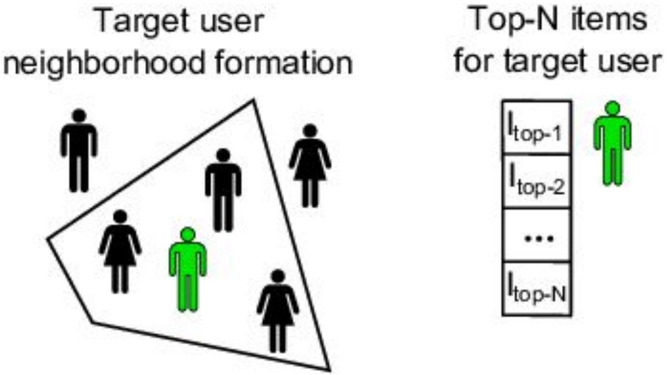
Implementación - Entrada al Sistema de Recomendación

```
user1_id user2_id occurrences
2        21       1.0
21       39       2.0
```

2. DISEÑO E IMPLEMENTACIÓN

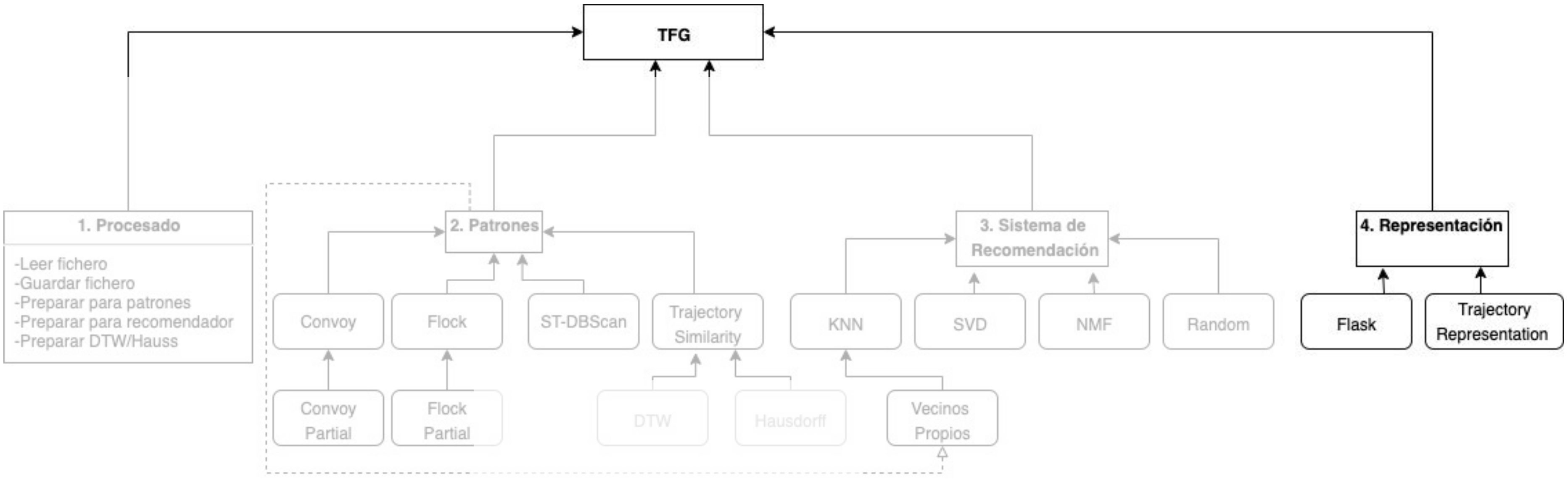
Implementación - Sistema de Recomendación





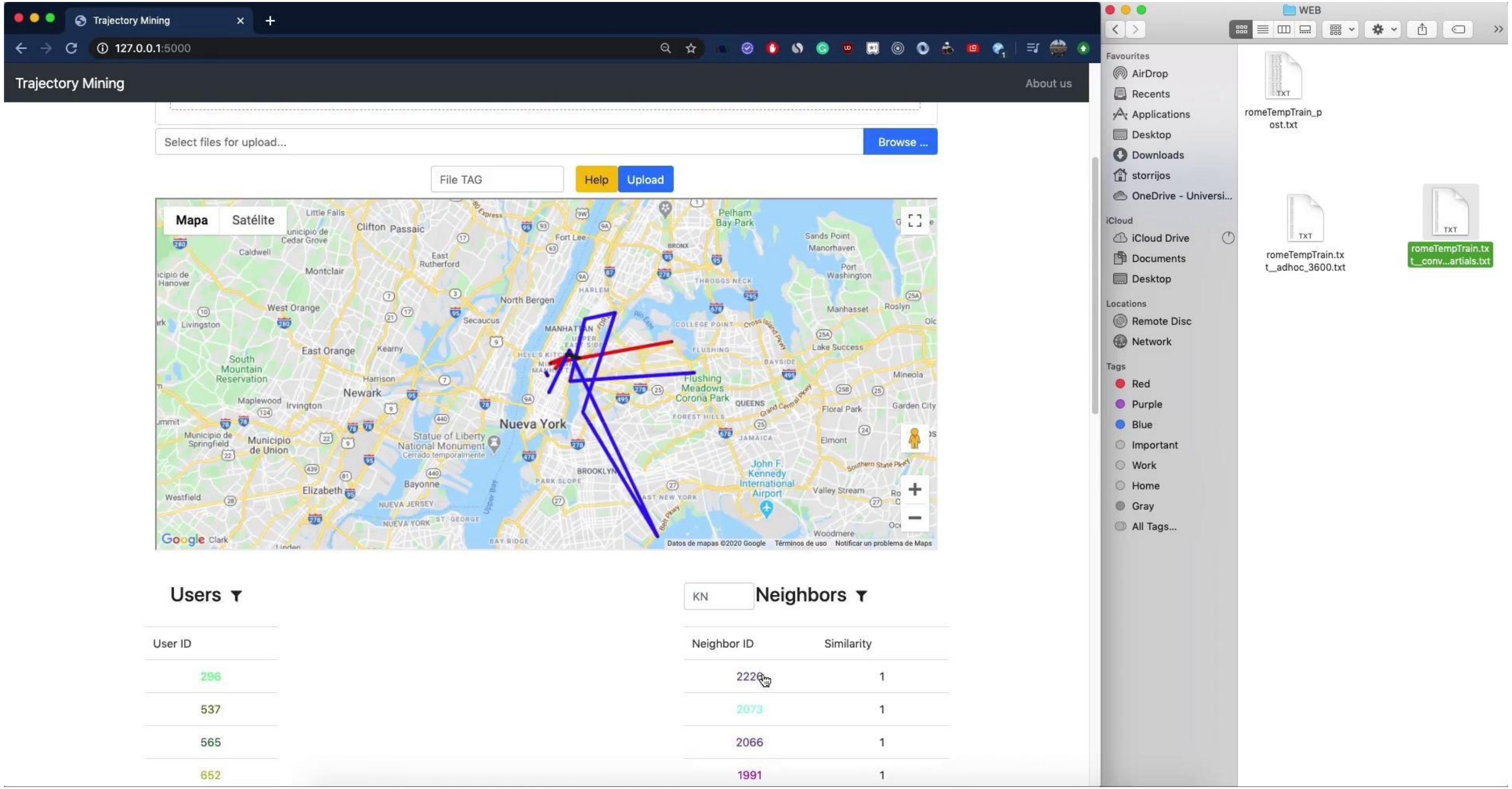
2. DISEÑO E IMPLEMENTACIÓN

Diseño - Estructura general



2. DISEÑO E IMPLEMENTACIÓN

Implementación - Visualización



3. PRUEBAS Y RESULTADOS

3. PRUEBAS Y RESULTADOS

Esquema de pruebas realizadas

01 Tiempos de ejecución

02 Número promedio de vecinos

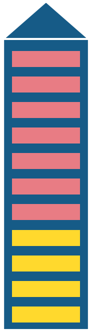
03 Efecto en la recomendación

04 Visualización y validación de trayectorias

3. PRUEBAS Y RESULTADOS

Dataset: Roma

user_id	item_id	latitude	longitude	timestamp
15	14627	40.757564	-73.989238	1354881399



319 POIs

1208 Usuarios

3. PRUEBAS Y RESULTADOS

1. Tiempos de ejecución

Algorithm	1K	10K	20K	60K
Ad-Hoc	0m0.477s	0m1.952s	0m5.253s	0m35.758s
Sim. tray. Hausdorff	0m10.568s	22m32.164s	88m17.315s	1157m31.251s
Sim. tray. DTW	0m3.386s	6m16.519s	26m9.038s	232m25.592s
ST-DBSCAN	0m0.817s	0m3.173s	0m9.487s	1m29.065s
Convoy	0m5.786s	0m6.086s	0m6.266s	0m6.098s
Convoy partials	0m2.786s	0m3.086s	0m3.266s	0m4.098s
Flock	0m5.199s	24m23.087s	49m34.065s	†
Flock partials	0m5.500s	0m30.545s	0m14.880s	3m11.811s

Dataset	1K	10K	20K	60K
Número POIs	163	319	346	394
Número usuarios	129	1208	2618	7954

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^{2}	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^2	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^2	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^{2}	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^{2}	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^{2}	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

2. Número promedio de vecinos

Algoritmo	Parámetros	Número vecinos promedio
Ad-Hoc	δ=3600	1.3
	δ=10800	1.8
Sim. tray.	Hausdorff	1206
	DTW	1206
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	1.7
	ε=5000, *th*=6000, *minEps*=5	3.1
	ε=4000, *th*=2000, *minEps*=1	2.2
	ε=1000, *th*=2000, *minEps*=1	1.5
Convoy	*min points*=20, *lifetime*=2, *dist max*=10^{-2}	22.8
	min points=2, *lifetime*=2, *dist max*=10^{-3}	2.1
	min points=2, *lifetime*=5, *dist max*=10^{-2}	1.23
	min points=5, *lifetime*=2, *dist max*=10^{-2}	5.92
	min points=5, *lifetime*=2, *dist max*=10^{-15}	6
	min points=3, *lifetime*=5, *dist max*=10^{2}	0
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-4}	1.4
	min points=2, *lifetime*=2, *dist max*=10^{-3}	1.7
Flock	ε=10, μ=2, δ=2	4
	ε=0.1, μ=2, δ=10	2
	ε=10, μ=2, δ=10^2	0
	ε=0.1, μ=2, δ=2	4
Flock partials	ε=10, μ=2, δ=2	5.2
	ε=0.1, μ=2, δ=10	5.09
	ε=0.1, μ=2, δ=2	8.04

3. PRUEBAS Y RESULTADOS

3. Efecto en la recomendación

Algoritmo	Parámetros	Precisión	MAE
Ad-Hoc	δ=3600	0.0361754	$\boxed{\textbf{0.3487827}}$
	δ=10800	0.0349337	0.3637682
Sim. tray.	Hausdorff	0.0175497	0.5784518
	DTW	0.0206126	1.2903128
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	**0.0369205**	0.4251326
	ε=5000, *th*=6000, *minEps*=10	0.0316225	0.3530357
	ε=5000, *th*=6000, *minEps*=10	**0.0369205**	0.4000756
	ε=1000, *th*=2000, *minEps*=1	0.0336921	0.3530357
	ε=4000, *th*=2000, *minEps*=1	0.0277318	0.3525041
Convoy	*min points*=2, *lifetime*=10, *dist max*=10^{-2}	0.0362583	0.3566276
	min points=20, *lifetime*=2, *dist max*=10^{-2}	0.0332781	1.0634231
	min points=2, *lifetime*=2, *dist max*=10^{-3}	0.0243377	0.5973882
	min points=2, *lifetime*=5, *dist max*=10^{-2}	0.0352649	0.4018430
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-3}	0.0360927	0.3925327
	min points=2, *lifetime*=2, *dist max*=10^{-1}	0.0368377	0.3525041
	min points=3, *lifetime*=5, *dist max*=10^{-1}	0.0276490	0.7646296
Flock	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
	ε=0.1, μ 2, δ=10	**0.0369205**	0.3704636
	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
Flock partials	ε=10, μ 2, δ=2	0.0368377	0.3525041
	ε=0.1, μ 2, δ=10	0.0355132	0.3530357
	ε=0.1, μ 2, δ=2	0.0368377	0.3670011

3. PRUEBAS Y RESULTADOS

3. Efecto en la recomendación

Algoritmo	Parámetros	Precisión	MAE
Ad-Hoc	δ=3600	0.0361754	**0.3487827**
	δ=10800	0.0349337	0.3637682
Sim. tray.	Hausdorff	0.0175497	0.5784518
	DTW	0.0206126	1.2903128
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	**0.0369205**	0.4251326
	ε=5000, *th*=6000, *minEps*=10	0.0316225	0.3530357
	ε=5000, *th*=6000, *minEps*=10	**0.0369205**	0.4000756
	ε=1000, *th*=2000, *minEps*=1	0.0336921	0.3530357
	ε=4000, *th*=2000, *minEps*=1	0.0277318	0.3525041
Convoy	*min points*=2, *lifetime*=10, *dist max*=10^{-2}	0.0362583	0.3566276
	min points=20, *lifetime*=2, *dist max*=10^{-2}	0.0332781	1.0634231
	min points=2, *lifetime*=2, *dist max*=10^{-3}	0.0243377	0.5973882
	min points=2, *lifetime*=5, *dist max*=10^{-2}	0.0352649	0.4018430
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-3}	0.0360927	0.3925327
	min points=2, *lifetime*=2, *dist max*=10^{-1}	0.0368377	0.3525041
	min points=3, *lifetime*=5, *dist max*=10^{-1}	0.0276490	0.7646296
Flock	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
	ε=0.1, μ 2, δ=10	**0.0369205**	0.3704636
	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
Flock partials	ε=10, μ 2, δ=2	0.0368377	0.3525041
	ε=0.1, μ 2, δ=10	0.0355132	0.3530357
	ε=0.1, μ 2, δ=2	0.0368377	0.3670011

3. PRUEBAS Y RESULTADOS

3. Efecto en la recomendación

Algoritmo	Parámetros	Precisión	MAE
Ad-Hoc	δ=3600	0.0361754	**0.3487827**
	δ=10800	0.0349337	0.3637682
Sim. tray.	Hausdorff	0.0175497	0.5784518
	DTW	0.0206126	1.2903128
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	**0.0369205**	0.4251326
	ε=5000, *th*=6000, *minEps*=10	0.0316225	0.3530357
	ε=5000, *th*=6000, *minEps*=10	**0.0369205**	0.4000756
	ε=1000, *th*=2000, *minEps*=1	0.0336921	0.3530357
	ε=4000, *th*=2000, *minEps*=1	0.0277318	0.3525041
Convoy	*min points*=2, *lifetime*=10, *dist max*=10^{-2}	0.0362583	0.3566276
	min points=20, *lifetime*=2, *dist max*=10^{-2}	0.0332781	1.0634231
	min points=2, *lifetime*=2, *dist max*=10^{-3}	0.0243377	0.5973882
	min points=2, *lifetime*=5, *dist max*=10^{-2}	0.0352649	0.4018430
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-3}	0.0360927	0.3925327
	min points=2, *lifetime*=2, *dist max*=10^{-1}	0.0368377	0.3525041
	min points=3, *lifetime*=5, *dist max*=10^{-1}	0.0276490	0.7646296
Flock	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
	ε=0.1, μ 2, δ=10	**0.0369205**	0.3704636
	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
Flock partials	ε=10, μ 2, δ=2	0.0368377	0.3525041
	ε=0.1, μ 2, δ=10	0.0355132	0.3530357
	ε=0.1, μ 2, δ=2	0.0368377	0.3670011

3. PRUEBAS Y RESULTADOS

3. Efecto en la recomendación

Algoritmo	Parámetros	Precisión	MAE
Ad-Hoc	δ=3600	0.0361754	**0.3487827**
	δ=10800	0.0349337	0.3637682
Sim. tray.	Hausdorff	0.0175497	0.5784518
	DTW	0.0206126	1.2903128
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	**0.0369205**	0.4251326
	ε=5000, *th*=6000, *minEps*=10	0.0316225	0.3530357
	ε=5000, *th*=6000, *minEps*=10	**0.0369205**	0.4000756
	ε=1000, *th*=2000, *minEps*=1	0.0336921	0.3530357
	ε=4000, *th*=2000, *minEps*=1	0.0277318	0.3525041
Convoy	*min points*=2, *lifetime*=10, *dist max*=10^{-2}	0.0362583	0.3566276
	min points=20, *lifetime*=2, *dist max*=10^{-2}	0.0332781	1.0634231
	min points=2, *lifetime*=2, *dist max*=10^{-3}	0.0243377	0.5973882
	min points=2, *lifetime*=5, *dist max*=10^{-2}	0.0352649	0.4018430
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-3}	0.0360927	0.3925327
	min points=2, *lifetime*=2, *dist max*=10^{-1}	0.0368377	0.3525041
	min points=3, *lifetime*=5, *dist max*=10^{-1}	0.0276490	0.7646296
Flock	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
	ε=0.1, μ 2, δ=10	**0.0369205**	0.3704636
	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
Flock partials	ε=10, μ 2, δ=2	0.0368377	0.3525041
	ε=0.1, μ 2, δ=10	0.0355132	0.3530357
	ε=0.1, μ 2, δ=2	0.0368377	0.3670011

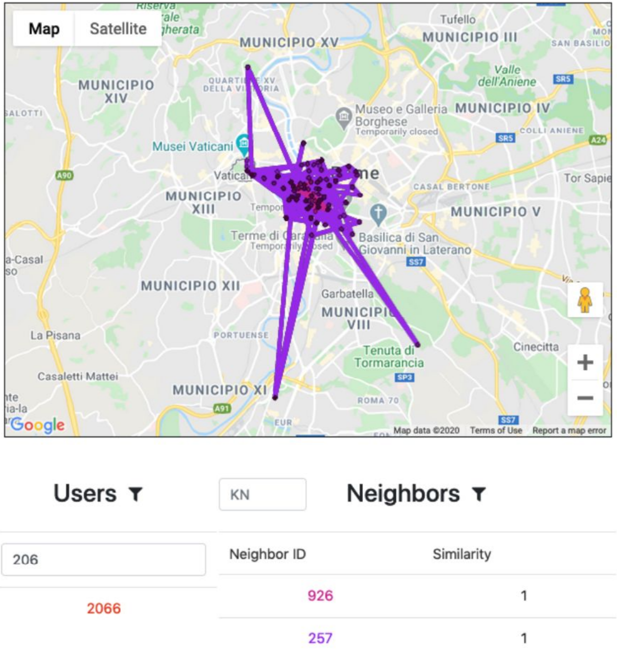
3. PRUEBAS Y RESULTADOS

3. Efecto en la recomendación

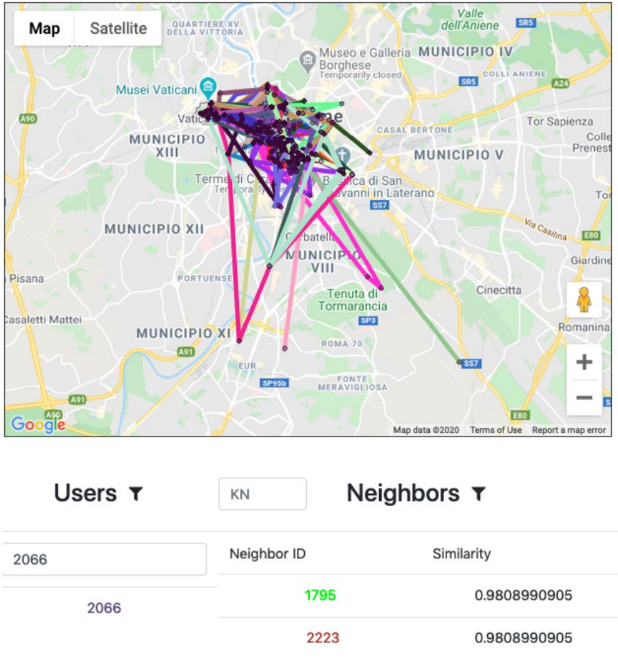
Algoritmo	Parámetros	Precisión	MAE
Ad-Hoc	δ=3600	0.0361754	**0.3487827**
	δ=10800	0.0349337	0.3637682
Sim. tray.	Hausdorff	0.0175497	0.5784518
	DTW	0.0206126	1.2903128
ST-DBSCAN	ε=5000, *th*=6000, *minEps*=1	**0.0369205**	0.4251326
	ε=5000, *th*=6000, *minEps*=10	0.0316225	0.3530357
	ε=5000, *th*=6000, *minEps*=10	**0.0369205**	0.4000756
	ε=1000, *th*=2000, *minEps*=1	0.0336921	0.3530357
	ε=4000, *th*=2000, *minEps*=1	0.0277318	0.3525041
Convoy	*min points*=2, *lifetime*=10, *dist max*=10^{-2}	0.0362583	0.3566276
	min points=20, *lifetime*=2, *dist max*=10^{-2}	0.0332781	1.0634231
	min points=2, *lifetime*=2, *dist max*=10^{-3}	0.0243377	0.5973882
	min points=2, *lifetime*=5, *dist max*=10^{-2}	0.0352649	0.4018430
Convoy partials	*min points*=2, *lifetime*=2, *dist max*=10^{-3}	0.0360927	0.3925327
	min points=2, *lifetime*=2, *dist max*=10^{-1}	0.0368377	0.3525041
	min points=3, *lifetime*=5, *dist max*=10^{-1}	0.0276490	0.7646296
Flock	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
	ε=0.1, μ 2, δ=10	**0.0369205**	0.3704636
	ε=10, μ 2, δ=2	**0.0369205**	0.3530357
Flock partials	ε=10, μ 2, δ=2	0.0368377	0.3525041
	ε=0.1, μ 2, δ=10	0.0355132	0.3530357
	ε=0.1, μ 2, δ=2	0.0368377	0.3670011

3. PRUEBAS Y RESULTADOS

4. Visualización y validación de trayectorias

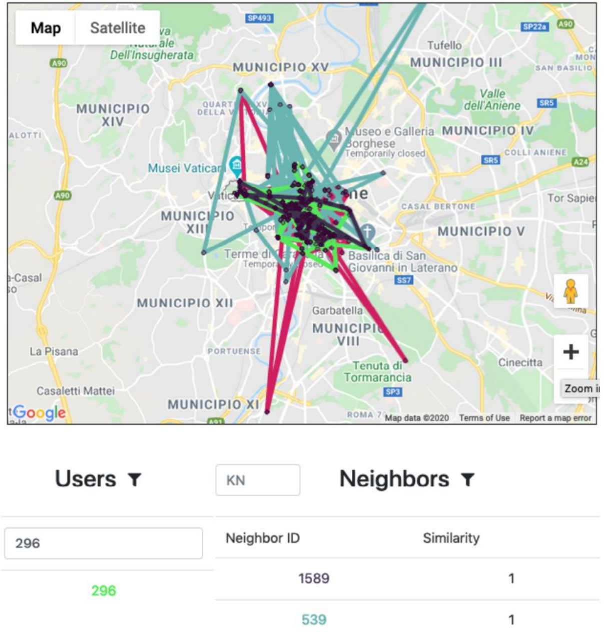


(c) Trayectoria Ad-Hoc δ=3600.

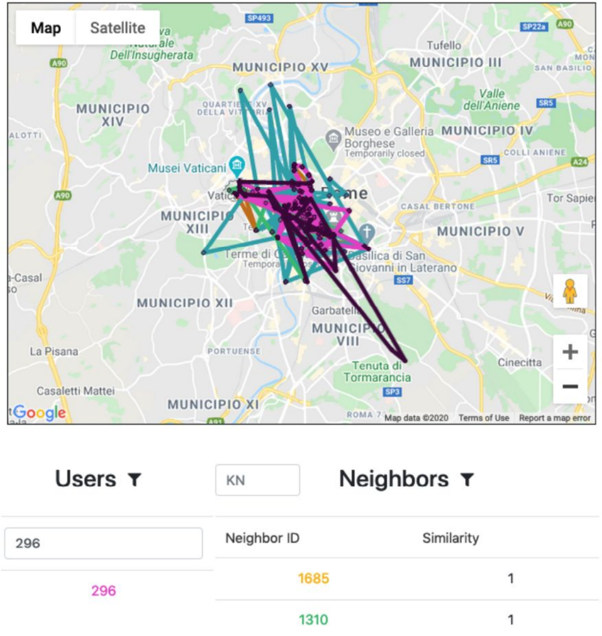


(d) Trayectoria DTW.

4. Visualización y validación de trayectorias - Convoy

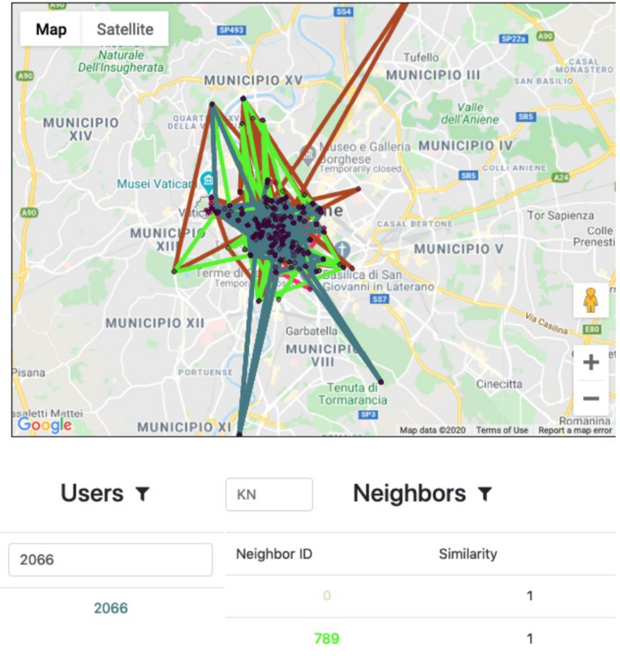


(a) Trayectoria Convoy *min points*=2, *lifetime*=2, *dist max*=0.01.

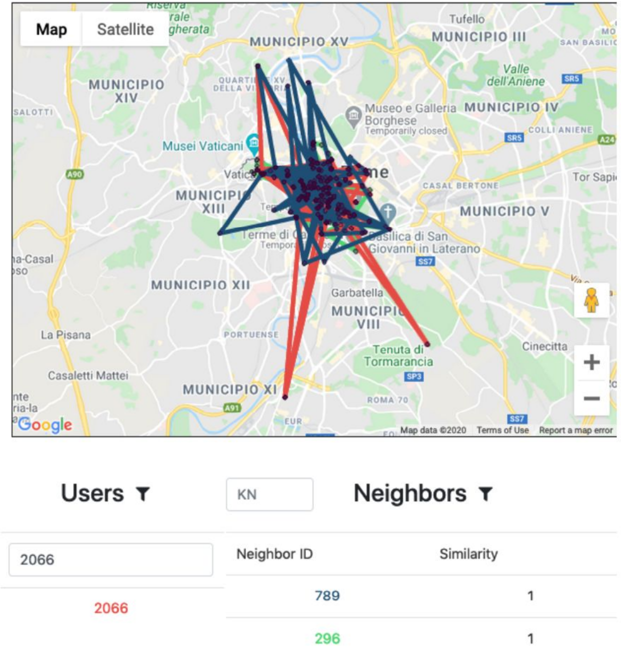


(b) Trayectoria Convoy *min points*=5, *lifetime*=2, *dist max*=0.01.

4. Visualización y validación de trayectorias - Flock



(a) Trayectoria Flock ε=0.1, μ=2, δ=2.



(b) Trayectoria Flock ε=0.1, μ=2, δ=10.

4. CONCLUSIONES Y TRABAJO FUTURO

4. CONCLUSIONES Y TRABAJO FUTURO

Conclusiones

1. Ventajas e inconvenientes

2. Utilidad estudios de trayectorias

3. Restricciones algoritmos

4. Publicación artículos

4. CONCLUSIONES Y TRABAJO FUTURO

Conclusiones - Publicaciones

Discovering Related Users in Location-Based Social Networks

Sergio Torrijos
Universidad Autónoma de Madrid
Madrid, Spain
sergio.torrijos@estudiante.uam.es

Alejandro Bellogín
Universidad Autónoma de Madrid
Madrid, Spain
alejandro.bellogin@uam.es

Pablo Sánchez
Universidad Autónoma de Madrid
Madrid, Spain
pablo.sanchezp@uam.es

ABSTRACT

Users from Location-Based Social Networks can be characterised by how and where they move. However, most of the works that exploit this type of information neglect either its sequential or its geographical properties. In this article, we focus on a specific family of recommender systems, those based on nearest neighbours; we define related users based on common check-ins and similar trajectories and analyse their effects on the recommendations. For this purpose, we use a real-world dataset and compare the performance on different dimensions against several state-of-the-art algorithms. The results show that better neighbours could be discovered with these approaches if we want to promote novel and diverse recommendations.

they visit, establish connections with other users, and check venue properties, such as their opening times, opinions, and pictures.

Because of the increasing number of users registered in LBSNs and similar systems, POI recommendation approaches have become particularly useful and several specific models have been proposed in recent years. In particular, such approaches tend to incorporate inherent properties of these systems, such as social, geographical, or temporal information [20, 21]. However, nearest neighbour techniques have been, in general, neglected in most of these studies, in favour of matrix factorisation or neural networks models [21, 24]. Nonetheless, we believe that algorithms based on similarities have a huge potential, since they may provide efficient computation, easy implementation, and explainable recommendations [23], but also because it has been demonstrated recently that these techniques

S. Torrijos, A. Bellogín, and P. Sánchez, "Discovering related users in location-based social networks," in User Modeling, Adaptation, and Personalization - 28th International Conference, UMAP 2020, Genoa, Italy, July 12-18, 2020. Proceedings, Lecture Notes in Computer Science, Springer, 2020.

Analysis of co-movement pattern mining methods for recommendation

Extended Abstract

Sergio Torrijos
Universidad Autónoma de Madrid
Madrid, Spain
sergio.torrijos@estudiante.uam.es

Alejandro Bellogín
Universidad Autónoma de Madrid
Madrid, Spain
alejandro.bellogin@uam.es

ABSTRACT

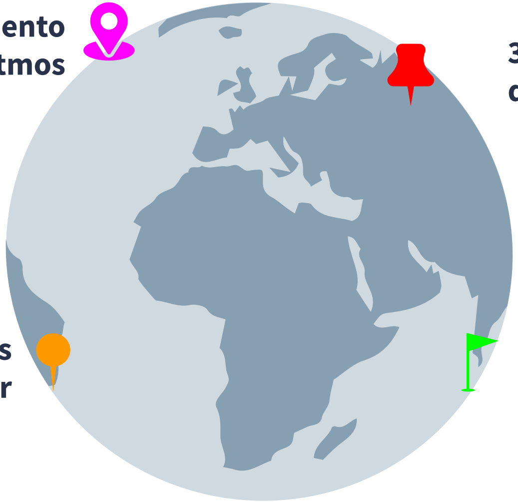
Location-Based Social Networks allow users to share the Points-of-Interest they visit, hence creating trajectories throughout their usual lives – even though they are also used by tourists to explore a city. There exist several algorithms in the trajectory pattern mining area able to discover and exploit interesting patterns from trajectory data, such as which *objects* tend to move together (co-movement), however, to the best of our knowledge, they have not been used with data coming from that type of systems. In this work, we analyse the extent to which these techniques can be applied to that type of data and under which circumstances they might be useful.

S. Torrijos and A. Bellogín, "Analysis of trajectory pattern mining methods for recommendation: Extended abstract," in Proceedings of the Joint Conference of the Information Retrieval Communities in Europe, CIRCLE 2020, Samatan, France, July 6-9, 2020, 2020.

4. CONCLUSIONES Y TRABAJO FUTURO

Trabajo futuro

1. Mejorar Rendimiento Algoritmos



2. Validación de las similitudes mediante Twitter

3. Solventar el problema de la dispersión en recomendación

4. Analizar resultados en función del tipo de usuario (local o turista)

Estudio de métodos de detección de patrones de movimiento para sistemas de recomendación turísticos.

Sergio Torrijos López de la Manzanara

Muchas gracias por su atención

Escuela Politécnica Superior

Universidad Autónoma de Madrid

2. DISEÑO E IMPLEMENTACIÓN

Implementación - Dynamic Time Warping y Hausdorff

Fórmula similitud:

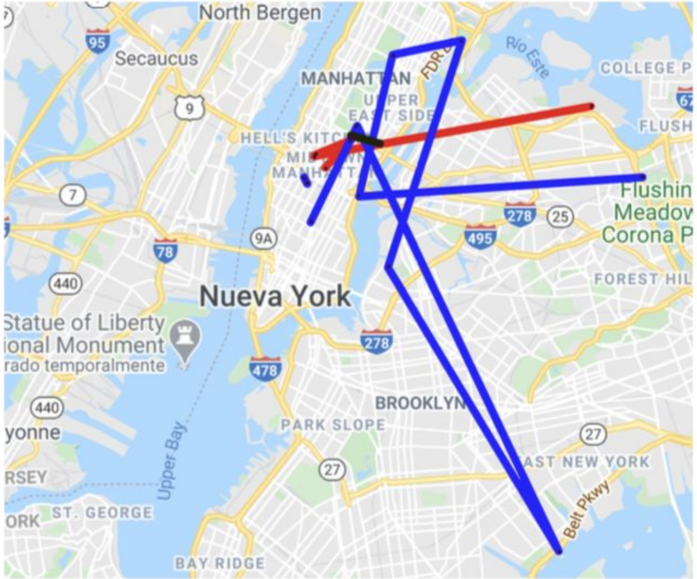
$$sim(u, v) = \frac{1}{n \cdot m} \sum_{j=1}^{n} \sum_{k=1}^{m} tsim(x_j^u, x_k^v)$$

Pseudocódigo:

```
D = [elementos_traj, elementos_traj]
for user in trajectorias_usuarios
    calcular similitud DTW/Hausdorff
Almacenar los k mas similares
```

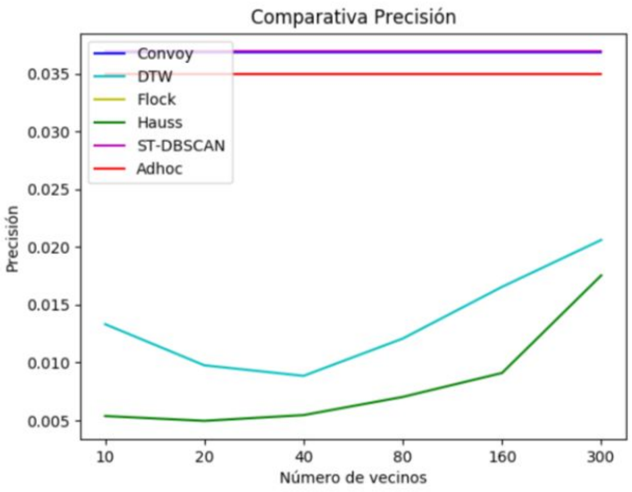
2. DISEÑO E IMPLEMENTACIÓN

Implementación - Dynamic Time Warping y Hausdorff

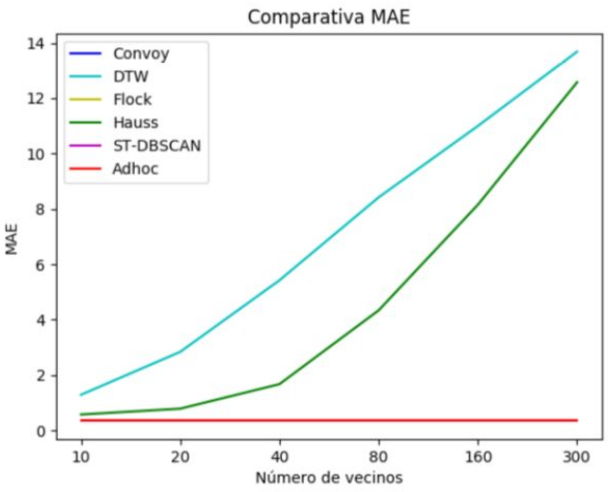


3. PRUEBAS Y RESULTADOS

Validación del Recomendador



(a) Evolución precisión recomendador.



(b) Evolución MAE recomendador.

3. PRUEBAS Y RESULTADOS

Comparativa entre recomendadores con similitud de trayectorias vs estándar

Recomendador	KNN Ad-Hoc	KNN Convoy	KNN DTW	KNN Flock	KNN Hausdorff	KNN ST-DBSCAN	SVD	Random	KNN Baseline	NMF
MAE	0.363768	0.352504	1.290313	0.352504	0.578452	0.353036	0.333831	0.475569	0.343762	**0.272056**

2. DISEÑO E IMPLEMENTACIÓN

Implementación - Recomendación

Formato fichero salida:

```
user_id    item_id    rating
   8          7       1.0964
   8          64      1.2289
   110        7       1.2235
```

3. PRUEBAS Y RESULTADOS

Obtención de Trayectorias - Cluster de K-Cores

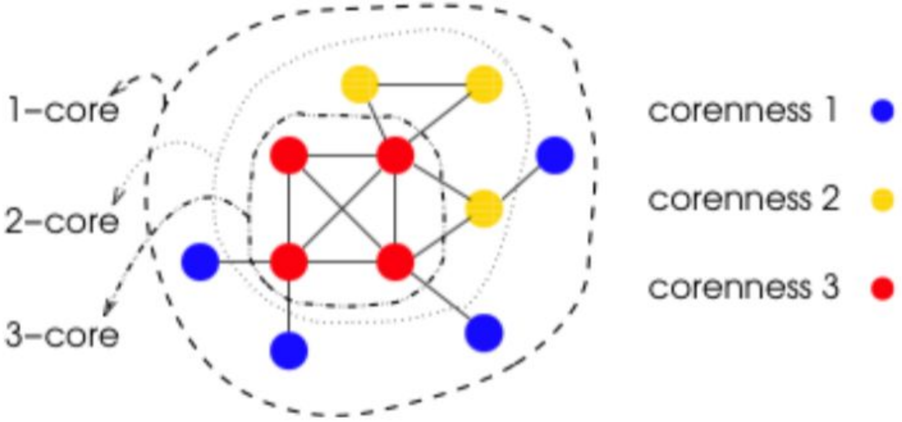


Figura 2.7: Procesado de K-Cores sobre un grafo [22].

3. PRUEBAS Y RESULTADOS

Relación entre tamaño del dataset y número de POIs/usuarios

Dataset	1K	10K	20K	60K
Número POIs	163	319	346	394
Número usuarios	129	1208	2618	7954

Implementación - Ad-Hoc

Fórmula similitud:

$$sim_\delta(u,v) = \| i \in I : |t(u,i) - t(v,i)| < \delta) \|$$

Ficheros necesarios:

- El fichero del dataset procesado con las columnas (user_id, item_id, lat, long, timestamp).
- El fichero de similitudes con las columnas (user1_id, user2_id, similitud).
- Opcional: El fichero de trayectorias con el formato (’user_id’: [’traj_counter’: [...]]).

3. PRUEBAS Y RESULTADOS

Entorno de pruebas

Recursos	Características
Versión Python	2.7
S.O.	macOS Mojave 10.14.6 (18G103)
CPU	2,7 GHz Intel Core i7 (I7-8559U)
GPU	Intel Iris Plus Graphics 655 1536 MB graphics
RAM	16 GB 2133 MHz LPDDR3

2. DISEÑO E IMPLEMENTACIÓN

Implementación - Salida común ejecución patrones de movimiento

```
user1_id user2_id occurrences
2        21       1.0
21       39       2.0
```

2. DISEÑO E IMPLEMENTACIÓN

Patrones de movimiento - Flock

Parámetros:

ε	μ	δ
Distancia entre los elementos móviles (radio).	Número mínimo de objetos móviles	Intervalo de tiempo definido mínimo entre elementos.

Implementación - Ejecución Flock

Ejecución en 2 fases:

1. Determinar los objetos móviles que se encuentran cerca según ε.
2. Combinación y agrupación de patrones que ocurran durante el parámetro δ.

```
KeyFlock: 30 Begin: 238 End 239 [0, 1025, 2437, 3720]
KeyFlock: 32 Begin: 231 End 239 [0, 1025, 2437]
KeyFlock: 34 Begin: 238 End 240 [0, 1025, 2437, 3720]
```

```
id      item_id   latitude    longitude    timestamp
796     140514    40.758328   -73.985457   0
1024    788744    40.730084   -73.989256   0
1024    788734    40.730085   -73.989257   1
1024    788784    40.730086   -73.989258   2
```

Implementación - Preparación del dataset

1. Selección atributos

```
         user_id   item_id   latitude     longitude    timestamp
          15       14627     40.757564    -73.989238   1354881399
```

1. 1. Selección atributos Hausdorff / DTW

```
{"user_id": [{"0": [[lat elements], [long_elements]], "1": ...}]}

{"15": [{"0": [[40.77383804321289, 40.75756454467734, 40.75837326049805,
40.76288986206055], [-73.87122344970703, -73.9892349243164, -73.98849487304688,
-73.97402954101562]], "1": [[40.753875732421875], [-73.98442840576172]],
"2": [[40.76355743408203], [-73.97286987304688]]}]}
```

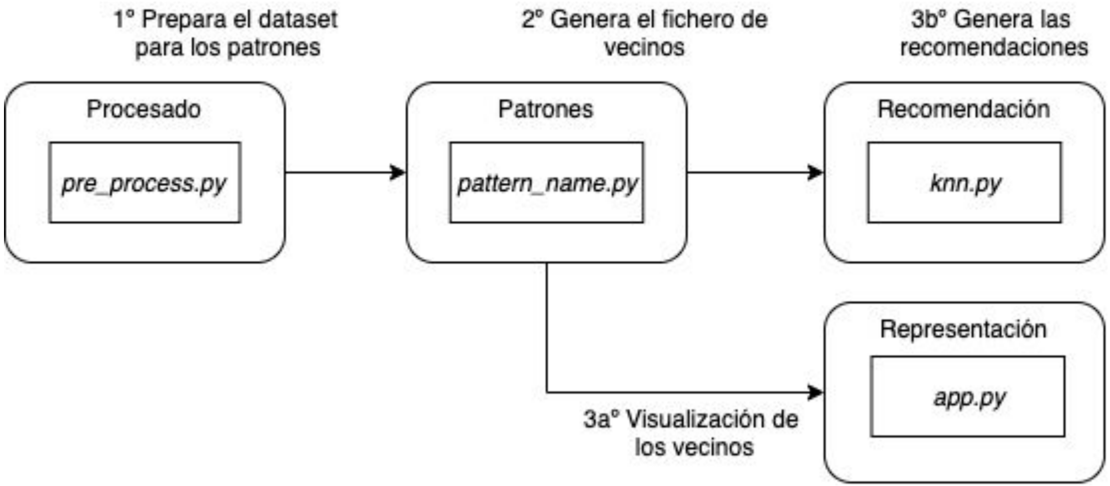
2. DISEÑO E IMPLEMENTACIÓN

Implementación - Flujo de la aplicación

```
1. python3 src/Processing/pre_process.py --input_file entradas/Rome10K/romeTempTrain.txt --
   coords_file entradas/POIS_rome__Coords.txt --output_file romePOISCompleto.txt
2. python3 src/Patterns/Convoy/ConvoyTrajectory.py --filename romePOISCompleto.txt --output
   similarity_output_convoy.txt --minpoints 3 --lifetime 2 --distance_max 0.1 --partials
   False
3. python3 src/Recommender/knn.py --train_file entradas/Rome10K/romeTempTrain.txt --
   test_file entradas/Rome10K/romeTest.txt --k 1 --neighbors_classified
   similarity_output_convoy.txt --output_file salida_knn_custom_rome.txt
4. python3 app.py
```

2. DISEÑO E IMPLEMENTACIÓN

Implementación - Flujo de la aplicación



2. ESTADO DEL ARTE

Parámetros:

ε

Radio espacial que delimita los puntos

minEps

Número mínimo de puntos para definir el cluster

temporal_threshold (th)

Ventana temporal que delimita los instantes de tiempo

2. ESTADO DEL ARTE

Patrones de movimiento - ST-DBSCAN

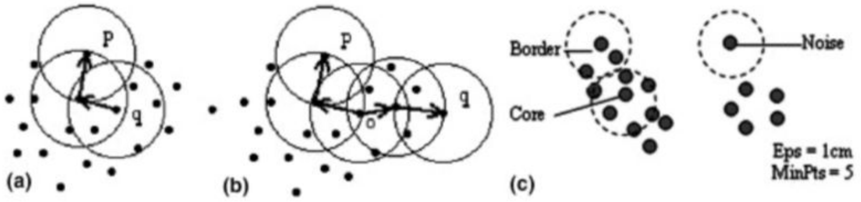
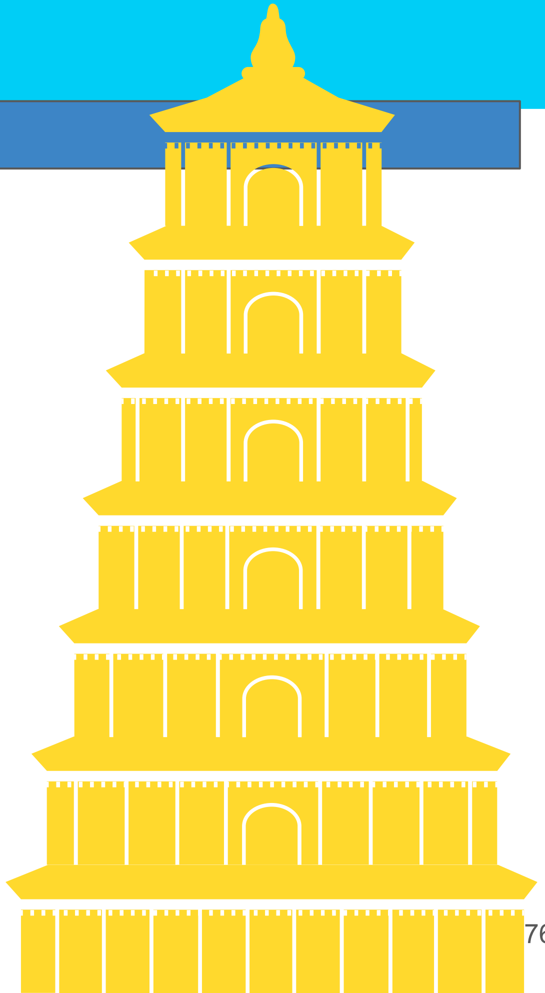


Figura 2.3: Ejemplo de cluster de puntos ST-DBSCAN [14], donde se aprecian los círculos delimitando el radio a buscar y cada uno de los puntos para definir cuáles entran en el criterio.

2. ESTADO DEL ARTE

Patrones de movimiento - Librerías utilizadas



LCM

FPFlock

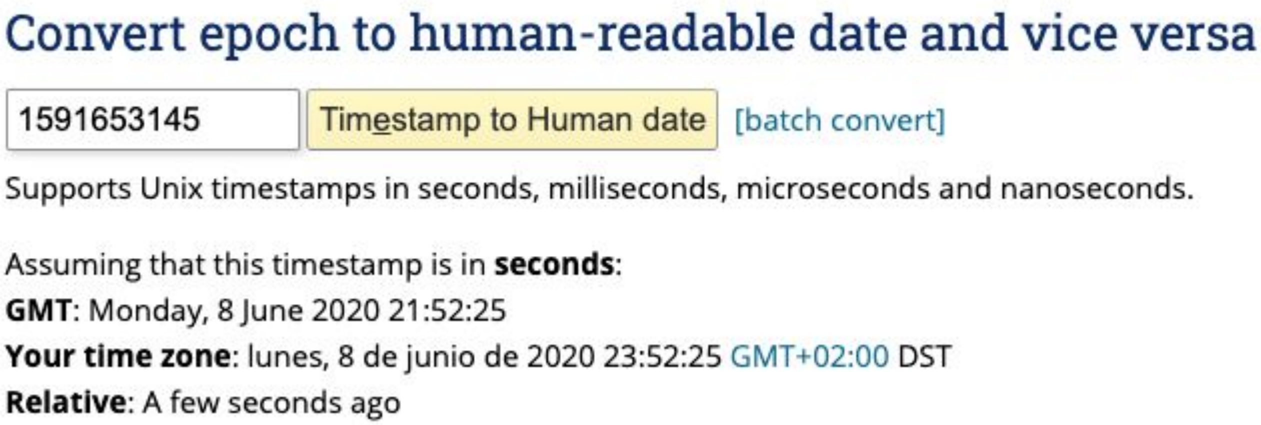
Pygmaps

py-st-dbscan

Coherent Moving Cluster Algorithm

Obtención de Trayectorias - Partición de trayectorias por instantes de tiempo

Timestamp:



2. ESTADO DEL ARTE

Sistemas de recomendación

Filtrado Colaborativo:

Métodos basados en modelos

1. Factorización de matrices

Métodos basados en memoria

1. Rating y Similitud basado en usuario
2. Rating y similitud basado en ítem

Factorización de matrices

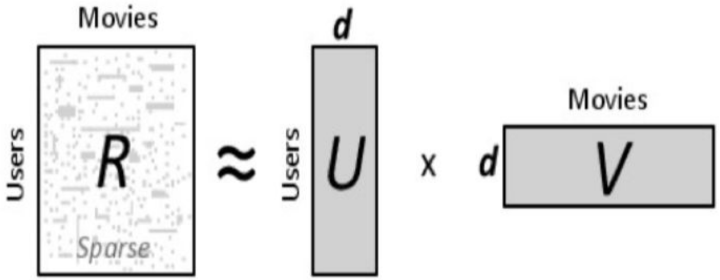


Figura 2.8: Representación de la factorización de matrices [24].

$$\boxed{\hat{r}(u,i) = x_u^T \cdot y_i}$$

Predicción de rating y similitud entre usuarios

$$\hat{r}(u,i) = C \times \sum_{v \epsilon N_k(u), r(v,i) \neq \varnothing} sim(u,v) r(v,i)$$

$$C = \frac{1}{\sum_{v \epsilon N_k(u), r(v,i) \neq \varnothing} |sim(u,v)|}$$

Predicción de rating y similitud entre ítems

$$\hat{r}(u,i) = C \times \sum_{r(u,j) \neq \varnothing} sim(i,j) r(u,j)$$

$$C = \frac{1}{\sum_{r(u,j) \neq \varnothing} |sim(i,j)|}$$

Similitud Coseno

$$sim(u,v) = \frac{\sum_{v \epsilon i: r(u,i) \neq \varnothing, r(v,i) \neq \varnothing} r(u,i) r(v,i)}{\sqrt{\sum_{i: r(u,i) \neq \varnothing} r(u,i)^2 \sum_{i: r(v,i) \neq \varnothing} r(v,i)^2}} \epsilon [0,1]$$

Correlación de Pearson

$$sim(u,v) = \frac{\sum_{v \epsilon i: r(u,i) \neq \varnothing, r(v,i) \neq \varnothing} (r(u,i) - \bar{r}_u)(r(v,i) - \bar{r}_v)}{\sqrt{\sum_{i: r(u,i) \neq \varnothing, r(v,i) \neq \varnothing} (r(u,i) - \bar{r}_u)^2 \sum_{i: r(u,i) \neq \varnothing, r(v,i) \neq \varnothing} (r(v,i) - \bar{r}_v)^2}} \epsilon [0,1]$$

Sistemas de recomendación - Métricas de evaluación

MAE

$$MAE = \frac{1}{|test|} \sum_{(u,i) \epsilon test} |\hat{r}(u,i) - r(u,i)|$$

RMSE

$$RMSE = \sqrt{\frac{1}{|test|} \sum_{(u,i) \epsilon test} (\hat{r}(u,i) - r(u,i))^2}$$

Precisión

$$P = \frac{TP}{TP + FP}$$

Recall

$$R = \frac{TP}{TP + FN}$$

2. DISEÑO E IMPLEMENTACIÓN

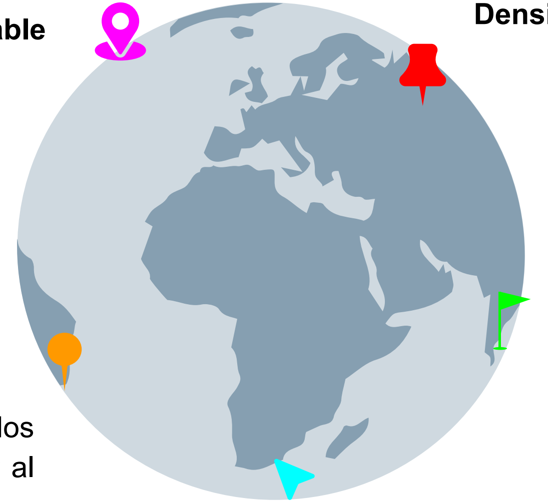
Patrones de movimiento - Convoy

Density-reachable

Un punto p es alcanzable desde un punto q si el punto p está a una distancia ε del punto q, y q tiene un número suficiente de puntos en sus vecinos a una distancia ε.

Core object

Un punto en cual sus vecinos conectados deben satisfacer la condición de contener al menos minPts.



Density-Connected

Un punto p ∈ S está densamente conectado a un punto q ∈ S con respecto a e y m si existe un punto x ∈ S tal que ambos p y q son alcanzables desde x.

Border-object

Un objeto p es un border-object si no es un core-object pero es density-reachable desde otro core-object.

Búsqueda Convoy

Convoy devuelve todos los grupos de objetos posibles tal que cada grupo consista en un grupo máximo de puntos densamente conectados con respecto a e y m durante al menos k puntos.