

Informe BDNR - Curso 2023

Grupo 27: Ignacio Corrales - Martin Katzenstein

Instituto de Computación

Facultad de Ingeniería, Universidad de la República

Montevideo, Uruguay

Resumen

El presente informe abarca el trabajo realizado en para la tarea final del curso 2023 de Bases de Datos no Relaciones. El objetivo del trabajo es generar un esquema de almacenamiento que permita organizar los datos disponibles del USGS <https://www.usgs.gov/about/about-us/who-we-are> (Servicio de Relevamiento Geológico de los Estados Unidos) y almacenarlos en una o más bases de datos no relacionales, con el fin de que puedan ser utilizados a futuro por usuarios que deseen hacer, a partir de dicho datos, análisis de datos, como por ejemplo, desarrollar un sistema de predicción de terremotos y tsunamis o alertas tempranas.

I. INFORMACIÓN CONTEXTUAL Y MOTIVACIÓN

Los terremotos son eventos naturales que ocurren cuando se libera energía acumulada en la Tierra. Si bien la mayoría de esos fenómenos geológicos son apenas perceptibles por la población en general y no tienen consecuencias en las infraestructuras instaladas hay casos que pueden tener efectos devastadores para la vida humana y dichas infraestructuras. Por lo tanto, el estudio de los terremotos es de vital importancia para comprender mejor estos eventos y tomar medidas adecuadas para reducir su impacto. Algunos ejemplos de estos eventos catastróficos ocurridos en los últimos años son el terremoto de Sumatra en el año 2004, donde se estima que fallecieron más de 200.000 personas, el terremoto ocurrido en Haití en 2010 con un número similar de víctimas o el ocurrido en Ciudad de México en 1985 con un estimado de 40.000 víctimas.

Más allá del conocimiento científico per se, el objetivo principal de estos estudios es comprender las dinámicas de estos eventos, entender mejor cómo se producen y cómo se propagan las ondas sísmicas identificando patrones y tendencias en su ocurrencia. Esto ayuda a predecir los lugares donde los terremotos son más probables y determinar la magnitud esperada de estos eventos. Esta información es crucial para la planificación urbana y la construcción de infraestructuras resistentes a los terremotos. El estudio de los terremotos contribuye a la mejora de los sistemas de alerta temprana. Los sistemas de monitoreo permiten detectar los primeros signos de un terremoto y enviar alertas a las áreas amenazadas. Esto brinda a las personas un tiempo valioso para ponerse a salvo y a las autoridades la oportunidad de tomar medidas preventivas. En definitiva los sistemas de alerta temprana salvan vidas y minimizan los daños materiales a la infraestructura al proporcionar información anticipada sobre los terremotos inminentes.

El estudio de los terremotos ayuda también a comprender las características geológicas de las regiones propensas a los sismos. Los terremotos están estrechamente relacionados con la dinámicas que rigen las placas tectónicas y mediante el análisis de los eventos pasados y presentes, los científicos pueden identificar y mapear las áreas donde la actividad sísmica es más intensa. Esta información es valiosa para evaluar los riesgos y tomar decisiones informadas sobre la ubicación de las distintas infraestructuras, como pueden ser las centrales nucleares.

Estos estudios contribuyen al avance de la ingeniería sísmica. Comprender cómo las estructuras se comportan durante un terremoto y cómo se pueden diseñar para resistir las fuerzas sísmicas es esencial para construir infraestructuras más seguras. Los técnicos y las autoridades utilizan los conocimientos obtenidos del estudio de los terremotos para desarrollar técnicas de diseño sísmico, códigos de construcción y estándares de seguridad.

En resumen, el estudio de los terremotos es de suma importancia debido a los riesgos que representan para la vida humana y las infraestructuras. Comprender mejor estos eventos nos permite predecirlos, desarrollar sistemas de alerta temprana, evaluar los riesgos, mejorar la planificación urbana y construir estructuras más seguras.

II. OBTENSIÓN DE LOS DATOS UTILIZANDO LA API DEL USGS

Los datos se obtienen utilizando la API del USGS¹. A modo de ejemplo se incluye una función en lenguaje Python que obtiene datos de eventos entre fechas dadas.

Listado 1 Solicitud de Datos de eventos a través de la API del USGS

```
def get_earthquake_data(start_date, end_date):
    url = "https://earthquake.usgs.gov/fdsnws/event/1/query"

    # Parametros para la consulta a traves de la API
```

¹<https://earthquake.usgs.gov/fdsnws/event/1/>

```

params = {
    "format": "geojson",
    "starttime": start_date,
    "endtime": end_date,
    "maxlatitude" : 44, #California
    "minlatitude" : 30, #California
    "minlongitude" : -125, #California
    "maxlongitude" : -114 #California
}

# Envio de la solicitud de datos
response = requests.get(url, params=params)

if response.status_code == 200:
    data = response.json()
    return data
else:
    print("Error al obtener los datos.")
    print(response)
    return None

```

III. ANALISIS INICIAL DE LOS DATOS DISPONIBLES

El objetivo de este trabajo es generar una herramienta que pueda ser utilizada en el análisis de los registros de eventos. Originalmente se planteó como hipótesis que una solución eficiente para el almacenamiento de la información disponible sea el de utilizar una Base de Datos documental con una posible segunda etapa en que se utiliza una Base de Datos basada en grafos como parte del producto final.

Durante el análisis inicial de los datos, resultó evidente que la complejidad de los fenómenos que se busca estudiar se encuentra principalmente representada en las relaciones entre los eventos. Estas relaciones son varias, de distinto tipo y de compleja caracterización. En función de este análisis inicial se llegó a la conclusión que la Base de Datos basada en grafos es, en principio, más adecuada para cumplir con todos los cometidos del trabajo. Tanto para realizar el almacenamiento propio de los datos como para representar eficientemente relaciones complejas entre los eventos estudiar y extraer información que pueda ser insumo para estudios posteriores.

Este trabajo se enfoca en dar herramientas orientadas a generar un set de datos, que pueda ser utilizado tanto para estudios supervisados de clasificación de eventos como para estudios no supervisados que ayuden a aportar nuevos conocimientos sobre los eventos sísmicos en general.

Para establecer el alcance de este trabajo vamos a representar las relaciones que surgen de los tiempos y ubicaciones de los mismos. La información obtenida del USGS y las relaciones que se establezcan serán almacenadas en una Base de Datos basada en grafos Neo4J.

IV. ESQUEMA NEO4J

Para la implementación se utilizará el servicio Aura de Neo4J y se complementará con herramientas de visualización utilizando la plataforma NeoDash².

Definición de los Nodos Como prueba de concepto proponemos un modelo simple en que existen 3 tipos de Nodos.

- El primer tipo de nodo representan los Eventos (N_E)
- El segundo tipo de nodos representan las Locaciones que interesa proteger (centros poblados, infraestructura crítica, etc) (N_L).
- El tercer tipo de nodo representan los Clusters que forman los nodos (N_E) y (N_L).

Propiedades de los nodos de Eventos (N_E).

- net: red que registra el evento
- code: código numérico que la red le asigna al evento (net + code forman el identificador)
- titulo: un titulo descriptivo del evento
- fecha: fecha y hora
- tipo: (earthquake, quarry blast, explosion, etc...)
- magnitud
- web: webpage de USGS donde se detalla el evento
- coordenadas (lat,long,prof)
- lugar: descripción textual del lugar del evento
- ids: (net+code, es una lista, ya que si varias redes lo identificaron, cada una le asignó un código distinto)

²<https://neodash.graphapp.io/>

- detalle: url que otorga mayor detalle del evento en formato geojson

Propiedades de los nodos de Ciudades (N_L).

- coordenadas (lat, long)
- nombre

Propiedades de los nodos de Clusters (N_C).

- inicio: fecha del evento asociado al Cluster con menor fecha.
- fin: fecha del evento asociado al Cluster con mayor fecha.
- coordenadas: (promedio latitud, promedio longitud)
- magnitud_avg: magnitud promedio
- cant_eventos: cantidad de eventos del Cluster

Definición de los vínculos entre los nodos Se propone un modelo en que los vínculos son sólo con nodos de eventos (N_E), es decir, $(N_E) - (N_E)$ y $(N_E) - (N_L)$. Estos se generarán de modo que si dos eventos están cerca en el espacio y el tiempo el vínculo es fuerte. Para medir esa fortaleza, proponemos utilizar dos parámetros.

- γ_1 para los vínculos entre eventos
- γ_2 para los vínculos entre eventos y locaciones relevantes (centros poblados, etc)

Para los vínculos entre eventos proponemos las variables $\gamma_1 = \frac{1}{\alpha\Delta T + \beta_1 d}$ con d la distancia física en kilómetros entre los eventos y ΔT el tiempo en segundos entre eventos.

Para los vínculos entre eventos y locaciones, como asumimos que los segundos son elementos estáticos, proponemos el parámetro solo dependa de la distancia $\gamma_2 = \frac{1}{\beta_2 d}$.

Para ambas variables γ_1 y γ_2 proponemos que existan valores mínimos en que si el valor es menor a este valor mínimo de referencia se asigna el valor cero y los nodos sobre los que se calculan la variable no estarán vinculados.

- $\frac{1}{\alpha\Delta T + \beta_1 d} \geq \epsilon_1 \rightarrow \gamma_1 = \frac{1}{\alpha\Delta T + \beta_1 d}$
- $\frac{1}{\alpha\Delta T + \beta_1 d} < \epsilon_1 \rightarrow \gamma_1 = 0$
- $\frac{1}{\beta_2 d} \geq \epsilon_2 \rightarrow \gamma_2 = \frac{1}{\beta_2 d}$
- $\frac{1}{\beta_2 d} < \epsilon_2 \rightarrow \gamma_2 = 0$

En el cálculo de los variables γ_1 y γ_2 se utilizan los parámetros α , β_1 y β_2 . Estos parámetros, juntos con los mínimos ϵ_1 y ϵ_2 podrán ser utilizados como hiperparámetros en análisis futuros ya que estos serán los que determinen la cantidad y estructura de los clusters de eventos y localidades.

A modo de ejemplo simplificado obtenemos los datos de los 1229 eventos ocurridos en el Estado de California entre el 20 y el 30 de mayo de 2023. Para determinar si los eventos están vinculados y la fortaleza del vínculo, utilizamos para este ejemplo los parámetros $\alpha = \beta_1 = 10^{-6}$ y $\epsilon_1 = 1$. Mostramos en la figura 1 un gráfico con la representación de los Clusters que se forman al utilizar estos parámetros. El gráfico fue exportado desde la plataforma Aura de Neo4J.

V. GENERACIÓN DE LA BASE DE DATOS

Para generar la Base de Datos se programó un Notebook en Colab³ del que se agrega una copia en la entrega de este informe. Para este trabajo, considerando las limitaciones del servicio ofrecido por la plataforma Aura, se tomaron períodos de tiempo y zonas geográficas acotadas. Se cargó la Base de Datos con eventos ocurridos en la semana del 24 al 30 de mayo de 2023 en el estado de California, EEUU. Esto se hizo con el Notebook de colab "Primeras Pruebas" que se adjunta a la entrega de este informe.

³<https://colab.research.google.com/drive/1bjyEDOpsT3Kp53fy6tbWXoFpMlp-8Uz6?usp=sharing>

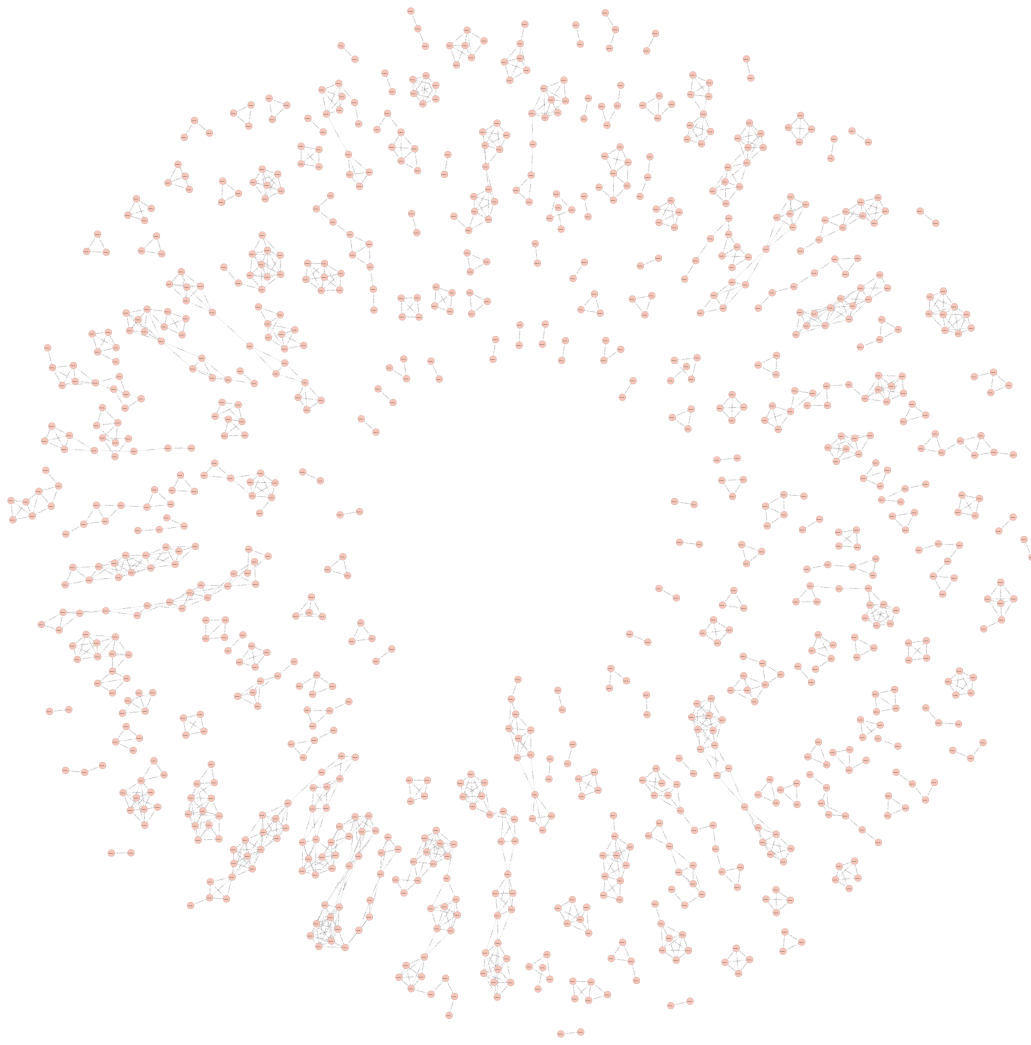


Figura 1: Ejemplo simplificado de como se forman los Clusters de Eventos ocurridos en el estado de California desde el 20 al 30 de Mayo de 2023

VI. CONSULTAS A LA BASE DE DATOS

Una vez que la información está almacenada en la Base de Datos se crea una lista de consultas con las que será posible generar conjuntos de datos que puedan ser insumo para futuros trabajos de análisis.

Las consultas que se incluyeron en la implementación del panel de visualización en NeoDash son las siguientes.

Listado 2 Consulta Cypher - Cantidad de Eventos

```
Match(e: Evento)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', ''))
return count(e)
```

Listado 3 Consulta Cypher - Consulta cantidad de ciudades afectadas

```
Match(e: Evento)-[r]-(c:Ciudad)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', ''))
return count(distinct c)
```

Listado 4 Consulta Cypher - Consulta magnitud mínima en el Cluster

```
Match(e: Evento)-[r]-(c:Ciudad)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
```

```

and e.magnitud >= toFloat(replace($neodash_magnitud, ',', '.'))
return min(e.magnitud)

```

Listado 5 Consulta Cypher - Consulta magnitud promedio en el Cluster

```

Match(e: Evento)-[r]-(c: Ciudad)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', '.'))

return avg(e.magnitud)

```

Listado 6 Consulta Cypher - Consulta Lista de Eventos en un período y zona determinada

```

Match(e: Evento)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', '.'))

return apoc.temporal.format(datetime(e.fecha), "dd/MM/yyyy_HH:mm") as Fecha,
e.tipo as Tipo,
e.magnitud as Magnitud,
e.coordenadas.y as lat,
e.coordenadas.x as long,
e.lugar as Lugar,
e.net+e.code as Id

```

Listado 7 Consulta Cypher - Consulta utilizada en la solapa Eventos del NeoDash

```

match(e:Evento)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', '.'))
return e

```

Listado 8 Consulta Cypher - Consulta utilizada en la solapa Ciudades del NeoDash

```

Mapa:
Match(c: Ciudad {nombre: $neodash_ciudad_nombre})-[r]-(e: Evento)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', '.'))
return e

Tabla:
Match(c: Ciudad {nombre: $neodash_ciudad_nombre})-[r]-(e: Evento)
where date(e.fecha) >= $neodash_desde
and date(e.fecha) <= $neodash_hasta
and e.magnitud >= toFloat(replace($neodash_magnitud, ',', '.'))
return apoc.temporal.format(datetime(e.fecha), "dd/MM/yyyy_HH:mm") as Fecha, r.distancia+'_km' as Distancia, r.direccion as Direc,

```

Listado 9 Consulta Cypher - Consulta utilizada en la solapa Clusters del NeoDash

```

tabla:
MATCH (c: Cluster)
WHERE date(c.inicio) >= $neodash_desde
and date(c.inicio) <= $neodash_hasta
and c.magnitud_avg >= toFloat(replace($neodash_magnitud, ',', '.'))
return ID(c) as ID, apoc.temporal.format(datetime(c.inicio), "dd/MM/yyyy_HH:mm") as Inicio,
apoc.temporal.format(datetime(c.fin), "dd/MM/yyyy_HH:mm") as Fin, c.magnitud_avg as Magn_Prom, c.cant_eventos as Cant_Eventos

```

Listado 10 Consulta Cypher - Consulta utilizada para generar el Mapa de correlación de eventos en el NeoDash

```

Mapa de correlacion entre eventos:
Match(e2:Evento)-[r]-(e:Evento)-[r2]-(c:Cluster)
where ID(c) = toInteger($neodash_id_de_cluster)
return r,e,e2
Mapa de Cluster y sus eventos:
Match(e:Evento)-[r]-(c:Cluster)
where ID(c) = toInteger($neodash_id_de_cluster)
return e, c

```

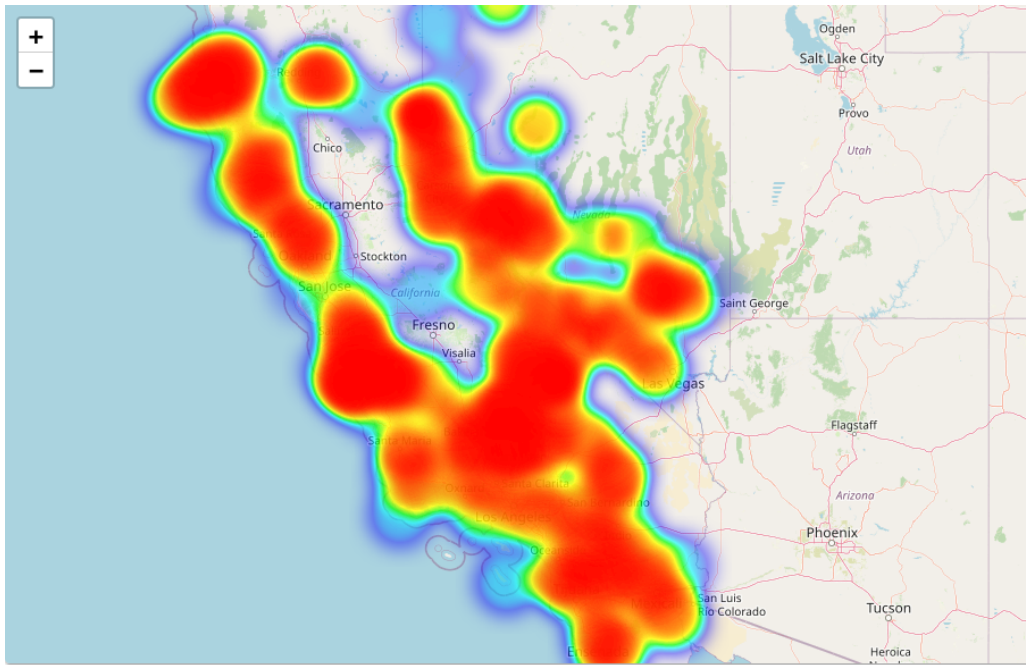


Figura 2: Mapa de Calor de los eventos en la costa oeste del estado de California desde el 24 de Mayo al 26 de Junio de 2023

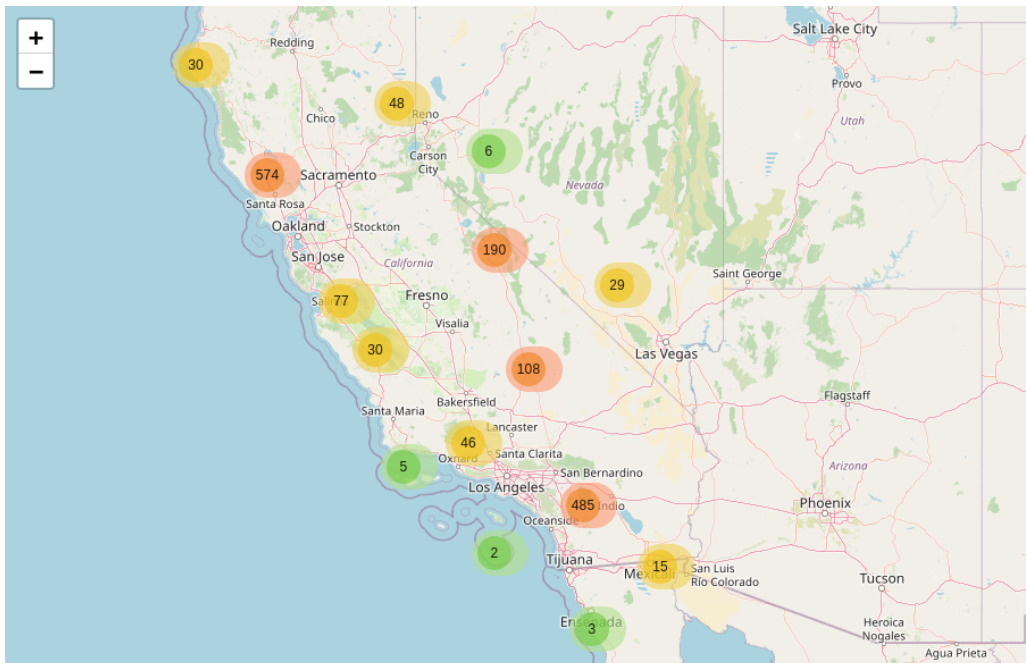


Figura 3: Eventos en la costa oeste del estado de California desde el 24 de Mayo al 26 de Junio de 2023

VII. VISUALIZACIONES

Utilizando las herramientas que nos provee NeoDash podemos generar visualizaciones que nos ayudan a comprender los datos con los que trabajamos y los vínculos que se almacenan en la Base de Datos. [Link a implementación en NeoDash](#)

En las figuras 2, 3 y 5, podemos ver ejemplos de eventos, clusters y las relaciones que hay entre los nodos que los forman.

Agregamos también la figura 4 donde se ve la ubicación geográfica de la falla de San Andrés que nos ayuda a contextualizar la info que se desprende de los otras imágenes.



Figura 4: Ubicación de la falla de San Andres

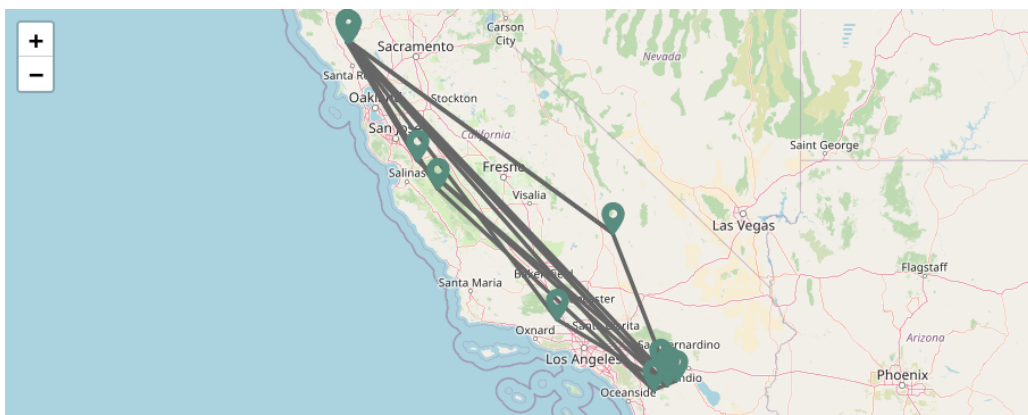


Figura 5: Cluster de eventos (cluster_id=3800)