



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

AYATORI: IMPLEMENTACIÓN EFICIENTE DE CONNECTION SCAN EN GTFS
CON CASO DE ESTUDIO DE MOVILIDAD EN SANTIAGO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN COMPUTACIÓN

FELIPE IGNACIO LEAL CERRO

PROFESOR GUÍA:
EDUARDO GRAELLS GARRIDO

MIEMBROS DE LA COMISIÓN:
NELSON BALOIAN TATARYAN
HERNÁN SARMIENTO ALBORNOZ

SANTIAGO DE CHILE
2023

Resumen

El presente informe detalla la creación de una implementación en Python de Connection Scan Algorithm, algoritmo que utiliza la información del transporte público disponible en la ciudad para calcular la mejor ruta para ir desde un punto A hasta un punto B, en el contexto del desarrollo de una Memoria para optar al título de Ingeniero Civil en Computación. Actualmente, la mayor parte de las herramientas utilizadas en la planificación vial son lentas de procesar y tediosas de ejecutar. Por este motivo, la motivación principal es crear una herramienta que permita realizar simulaciones de trayectos viales con la información actualizada del transporte público, y que así pueda ser utilizada a futuro para aportar en la planificación urbana y de transporte. Para evaluar la utilidad y correcta implementación de esta solución, se realizó un caso de estudio para Santiago de Chile, del que se concluye que la herramienta creada funciona como se esperaba, cumpliendo su objetivo exitosamente.

Dijiste que tenías un sueño, y ahora... ¡se cumplirá! ¡Los sueños y los ideales tienen poder para cambiar el mundo!

-N.

Agradecimientos

A mi mamá, por enseñarme a estudiar, preocuparte por mi futuro, y darme una razón para salir adelante a pesar de todo. A mi papá, por todo tu amor, apoyo y respeto, por enseñarme a priorizar mi vida, por todos tus años de servicio como padre viudo, por nunca dejar de cuidarme, y ser mi ejemplo a seguir. Al amor de mi vida, por ser mi apoyo y compañía principal, por ayudarme a extender las fronteras de mis sueños y esperanzas, por dejarme estar en tu vida, darme alegría y luz en mis peores momentos, reírte de mis chistes, y todo tu amor incondicional.

Gracias Pauli, por amar a mi padre y darle la oportunidad de estar de nuevo felizmente casado, por tu amor y preocupación, y extender lo que entiendo por 'familia'. Gracias Ita, Renata y Felipe, por nuestra mutua adopción familiar y convertirse en mi abuela y hermanos. A mis tatas, Daniel y Pechita, por sentar las bases familiares que inspiraron mis valores y moral, por consentirme y preocuparse de mí aunque la distancia nos separe. A mi tía Helen, por su amor, apoyo, y preocupación constantes por mi bienestar. A mi tío Leo, por todas las risas y anécdotas que me ayudaron a apreciar la sobremesa en familia. A mis primos: Amalia, Cristobal, Benja, Naty y Bastian, por su amistad, amor, apoyo, y alegrías varias. Gracias especiales a Nico, mi hermano del alma. Gracias a todo el resto de mi familia extendida.

Gracias, tío Alvaro y tía Katy, por aceptarme como su yerno, por su preocupación y cariño, por darme una segunda familia, y por otorgarme la posibilidad de seguir trabajando en mi sueño cuando más lo necesité. Gracias, Florencia y Julieta, por enseñarme a ser un hermano mayor, todo su aceptación y amor. A toda la familia Luna, por aceptarme como uno más entre los suyos, y por todo el cariño, consejos, y buenas vibras.

Gracias a mi curso, 12°B, por acompañarme en la primera parte de mi vida y por los amigos que encontré entre sus filas. Gracias a Anime no Seishin Doukoukai, por darme la oportunidad de adquirir responsabilidades incluso con mis hobbies, y por todos los grandes amigos que me permitió hallar. Gracias a Ivancito, Kurisu, Gus y Chelo, por ser mi apoyo en los llantos y mi compañía en las celebraciones. Gracias Gabi, Julio, Gabo, Sofi, Naise, por su amistad. Gracias a Basti, Lucho y Seba, por ser mis primeros amigos en el mundo exterior y mantenerse a mi lado hasta el día de hoy. Gracias a todos aquellos compañeros de carrera con los que he podido compartir y colaborar al ir educándome, destacando especialmente a mi amigo Matías Vergara, y a todos los miembros de Team Michil.

Gracias a todas mis profesoras y profesores, por darme las herramientas para llegar a donde estoy hoy.

Tabla de Contenido

1. Introducción	2
1.1. Objetivos	4
2. Estado del Arte	5
2.1. Connection Scan Algorithm	5
2.1.1. Implementaciones existentes	7
2.2. OpenStreetMap	8
2.2.1. OSM integrado en CSA	8
2.3. GTFS	9
2.3.1. GTFS integrado en CSA	11
2.4. Datos: estructura y manejo de la información	12
3. Diseño	13
3.1. Stack tecnológico	13
3.2. Funcionamiento lógico	14
3.3. Criterio de Evaluación	15
4. Implementación	16
4.1. Obtención de datos	16
4.1.1. Procesamiento de OpenStreetMap	16
4.1.2. Procesamiento de GTFS	19
4.2. Generación del output	22

4.2.1. Texto	22
4.2.2. Mapa	26
5. Resultados	29
5.1. Ejemplos de uso del programa	29
5.2. Caso de estudio	32
5.3. Evaluación de resultados	32
6. Discusión	33
6.1. Implicancias	33
6.2. Limitaciones	34
6.3. Trabajo Futuro	34
7. Conclusión	36
Bibliografía	38
Anexos	39
Apéndice A. Implementaciones existentes	39
A.1. Aalto University	39
A.1.1. ULTRA: headers	44
Apéndice B. Código de la solución	48
B.1. Obtención de la información	48
B.1.1. OSM	48
B.1.2. GTFS	52
B.2. Creación del mapa	56
B.3. Operación del algoritmo	62

Índice de Ilustraciones

2.1.	Diagrama explicativo del funcionamiento de Connection Scan Algorithm. Fuente: "Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys" (R. Kujala et al., 2017), vía sciencedirect.com	5
2.2.	Posibles caminos para llegar desde FCFM hasta Derecho en transporte público. Fuente: Google Maps.	6
2.3.	Mapa de Santiago en OpenStreetMap. Fuente: openstreetmap.cl	8
2.4.	Visualización de archivos del feed GTFS para Santiago.	9
2.5.	Diagrama de uso de datos en tiempo real en formato GTFS para una aplicación. Fuente: watrifeed.ml	11
4.1.	Ejemplo de salida del algoritmo (texto). Ruta entre FCFM y Puente Cal y Canto.	25
4.2.	Mapa de Santiago, generado con folium.	26
4.3.	Ejemplo de salida del algoritmo (mapa). Ruta entre FCFM y Puente Cal y Canto.	28
5.1.	Resultado de la implementación del algoritmo: viaje entre Conchalí y Santiago un jueves de madrugada.	30
5.2.	Resultado de la implementación del algoritmo: error al encontrar una dirección.	31
5.3.	Resultado de la implementación del algoritmo: error al no existir rutas para la fecha u hora especificadas.	31
5.4.	Resultado de la implementación del algoritmo: error al encontrar rutas.	31

Estructura del Documento

Este informe presenta las distintas etapas del desarrollo de una implementación de Connection Scan Algorithm en Python, que corresponde a la Memoria para optar al título de Ingeniero Civil en Computación. El documento está dividido en 7 capítulos distintos, listados a continuación con su respectiva temática:

- **Capítulo 1: Introducción.** Entrega la base contextual y los objetivos del proyecto.
- **Capítulo 2: Estado del Arte.** Presenta los antecedentes del proyecto que componen su Marco Teórico, junto a los conceptos y herramientas a utilizar, para comprender la base teórica del mismo.
- **Capítulo 3: Diseño.** Explica el diseño de la solución propuesta, incluyendo el stack tecnológico a usar, el funcionamiento lógico de la solución, y el criterio de evaluación a considerar, explicitando el caso de estudio a realizarse.
- **Capítulo 4: Implementación.** Expone el desarrollo de las distintas fases de la implementación del algoritmo.
- **Capítulo 5: Resultados.** Muestra los resultados finales obtenidos al terminar la implementación, a través de ejemplos de uso del programa, la ejecución del caso de estudio establecido, y la posterior evaluación.
- **Capítulo 6: Discusión.** Desarrolla las discusiones posteriores a la evaluación de resultados, considerando las implicancias y limitaciones de la solución, además de posibles líneas de trabajo futuro.
- **Capítulo 7: Conclusión.** Sintetiza el trabajo realizado y concluye el desarrollo del Trabajo de Título.

Posteriormente, se presenta la bibliografía utilizada y referenciada a lo largo del informe. Además, un anexo que muestra partes de implementaciones existentes del algoritmo, y otro que incluye gran parte del código de la solución desarrollada.

Capítulo 1

Introducción

A día de hoy, es normal que las grandes ciudades experimenten cambios constantemente que las hagan crecer. Este fenómeno, común a nivel mundial, está presente también en Chile. Estudiando la situación local, existen múltiples causas asociadas, algunas de estas siendo más globales (como el cambio climático), y otras más específicas, como el importante aumento de la migración tanto interna como externa al país durante los últimos años, y la construcción de nueva infraestructura urbana. Si bien han existido ciertas condiciones que afecten negativamente el florecimiento de las ciudades, como la pandemia del COVID-19, la tendencia general de crecimiento se mantiene. Así es como, en un mundo donde las grandes urbes tienden a crecer exponencialmente, la planificación y buena gestión de las ciudades se ha visto afectada por el auge de estos fenómenos.

Es necesario, entonces, hallar maneras novedosas para comprender y caracterizar, correc-tamente, la vida de los habitantes de las grandes ciudades, tal como la capital de nuestro país, Santiago. En este mismo contexto, una arista muy importante a considerar es la movilidad vial, o el cómo las personas son capaces de movilizarse a través de las calles y avenidas de una ciudad, la cual es un factor determinante de la calidad de vida de sus habitantes. Las grandes ciudades suelen ser el hogar de una gran cantidad de personas, las cuales necesitan transportarse cada día para realizar sus jornadas de trabajo, de estudio, entre otras.

Existen múltiples registros de información que pueden ser utilizados para estudiar la movilidad urbana de ciudades como Santiago. Sin embargo, esto no implica que dicho estudio se pueda realizar sin inconvenientes notables. Por ejemplo, la *Encuesta Origen Destino* es una herramienta utilizada por los gobiernos para estudiar patrones de viajes de los habitantes de las ciudades, y el gobierno de Chile ha realizado esta encuesta en múltiples ciudades del país durante los últimos años. Esto, evidentemente, incluye también a Santiago, pero la última vez que se realizó fue en el año 2012, hace más de una década atrás [14]. Debido a esto, la información inferida gracias a la encuesta probablemente no represente, de forma correcta, la realidad actual del transporte en la capital, lo cual es una problemática común a esta clase de instrumentos de estudio. Se necesita, luego, una herramienta que permita hacer este trabajo más continuamente, y que represente al común de los habitantes de la ciudad.

Con respecto a los medios de movilización, la gente puede tener a su disposición múltiples tipos, tanto públicos como privados. Por ejemplo, se pueden mover a pie, en bicicleta, en auto, o utilizando el transporte público. Siguiendo la idea anterior, para poder caracterizar correctamente la movilidad urbana, sería útil verlo desde la perspectiva de un medio de transporte que esté disponible para toda la población, así que estudiar el uso del transporte público en Santiago resulta ser una buena opción para este fin. Dentro de la ciudad, esto incluye al Metro de Santiago y los buses de Red (antiguamente Transantiago), los cuales son usados por las personas en múltiples combinaciones, generando una cantidad enorme de rutas diferentes. Para almacenar y hacer pública la información del sistema de horarios de esta clase de medios de transporte, existe el formato GTFS (General Transit Feed Specification) [3] que utilizan las agencias de transporte en el mundo para estandarizar la información de, entre otras cosas, los diferentes servicios existentes, sus rutas y paradas respectivas.

Connection Scan [8] (CSA) es un algoritmo creado para responder de manera eficiente a consultas en los sistemas de información de horarios. Este algoritmo recibe como entrada una posición de origen y una posición de destino, generando una salida consistente en una secuencia de vehículos que el viajero debe tomar para recorrer una ruta entre ambos puntos. CSA se alimenta de la información del transporte público disponible, así que es útil enlazar la información cartográfica de la ciudad con datos en formato GTFS para que logre operar correctamente.

Este trabajo de título tiene por objetivo principal realizar una implementación de Connection Scan, alimentándolo con la información del transporte en Santiago, para poder analizar los patrones de movilidad disponibles para los habitantes de la ciudad. La idea es que el producto generado permita, mediante este análisis, realizar estudios de movilidad vial de forma más actualizada y directa. La visión a futuro es que puedan realizarse casos de estudio, que visibilicen el impacto de la ampliación del transporte público disponible sobre los patrones de movilidad de las personas.

1.1. Objetivos

Objetivo General

El objetivo general de este trabajo de título es crear una implementación de Connection Scan Algorithm (CSA) en Python, con el fin de generar un programa que permita trazar rutas entre dos puntos cualesquiera de Santiago de Chile. Para ello, se utilizarán los datos cartográficos de la ciudad provenientes de OpenStreetMap, un proyecto colaborativo de creación de mapas comunitarios [2], y la información del transporte público provista por la Red Metropolitana de Movilidad.

Objetivos Específicos

1. Obtener la información cartográfica de Santiago, además de la información del transporte público (en formato GTFS), y almacenarla en estructuras de datos pertinentes.
2. Enlazar la información de ambas fuentes de datos para ubicar las rutas de transporte en el mapa de Santiago.
3. Desarrollar una implementación de CSA en Python, que permita ubicar dos puntos cualesquiera de la ciudad y entregar la mejor ruta para llegar del origen al destino.
4. Utilizar el algoritmo para realizar un caso de estudio que ejemplifique la utilidad del trabajo realizado.

Capítulo 2

Estado del Arte

2.1. Connection Scan Algorithm

Como fue mencionado anteriormente, Connection Scan Algorithm (CSA) es un algoritmo desarrollado para responder, de manera eficiente, consultas en sistemas de información de horarios [8]. Este algoritmo es capaz de optimizar los tiempos de viaje entre dos puntos determinados de origen y destino, siendo alimentado por distintas fuentes de información de transporte. Como salida, entrega una secuencia de vehículos (como trenes o buses) que un viajero debería tomar para llegar al destino desde el origen establecido. La base teórica tras el algoritmo hace que este analice las opciones disponibles y optimice el número de transbordos, tal que sea Pareto-eficiente, es decir, llegando al punto en el cual no es posible disminuir el tiempo de viaje en un medio de transporte sin tener que aumentar el de otro. En la figura 2.1, se grafica el funcionamiento antes descrito:

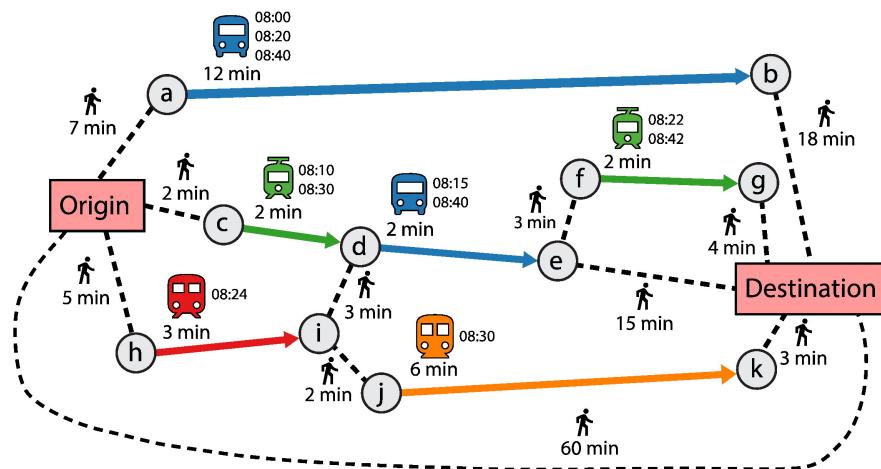


Figura 2.1: Diagrama explicativo del funcionamiento de Connection Scan Algorithm. Fuente: "Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys" (R. Kujala et al., 2017), vía sciencedirect.com .

Por ejemplo, si el punto de origen fuera la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile (Beauchef 850, Santiago), y el destino fuera la Facultad de Derecho de la Universidad de Chile (Pio Nono 1, Providencia), se debieran evaluar los medios de transportes que pueden ser utilizados para ir desde las coordenadas del punto de origen hasta las del punto de destino, y los transbordos necesarios. Posibles rutas podrían abarcar:

1. Una ruta con uso exclusivo del Metro de Santiago (subiendo en estación Parque O'Higgins de Línea 2, combinando en Los Héroes a Línea 1 y bajando en Baquedano).
2. Una ruta con uso exclusivo de buses de Red (tomar el recorrido 121 y luego el recorrido 502).
3. Una ruta que realice transbordos entre ambos medios de transporte (subir al metro en estación Parque O'Higgins y bajar en Puente Cal y Canto, para luego tomar el recorrido 502).

Los recorridos del ejemplo se muestran en la figura 2.2, generada utilizando el portal de Google Maps, el servidor web de visualización de mapas de Google [11], por su simplicidad de uso. Las rutas aparecen enumeradas según la lista previa.

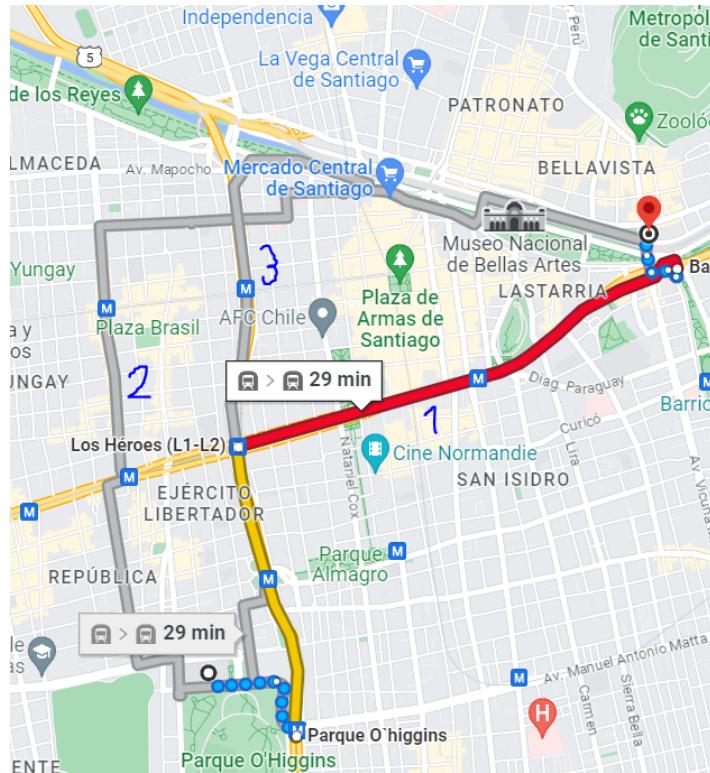


Figura 2.2: Posibles caminos para llegar desde FCFM hasta Derecho en transporte público.
Fuente: Google Maps.

Ejemplificando el caso anterior para resolverlo mediante CSA, el algoritmo recibe como entrada las coordenadas del punto de origen y el punto de destino. Luego, revisando la información del transporte público, calcula las rutas posibles (como las descritas anteriormente). CSA entonces buscaría el punto óptimo de Pareto con respecto a los transbordos, y entregaría

la ruta recomendada para llegar al destino deseado. En este caso, al trabajar con información estática, se toman ciertos supuestos, como una velocidad de caminata fija entre transbordos y la continuidad operativa del servicio en todo momento.

Connection Scan Algorithm precisa, en primera instancia, ser capaz de obtener y almacenar la información del transporte público disponible, para así ser capaz de calcular la ruta óptima. Sin embargo, dado que el algoritmo trabaja con las coordenadas de los puntos de origen y destino, es bueno contar también con los datos cartográficos del sector o ciudad en cuestión donde se desea realizar el viaje, con el fin de obtener mejores visualizaciones de los resultados. Trabajando con ambos flujos de información, es posible crear una sólida implementación del algoritmo.

2.1.1. Implementaciones existentes

Desde que Connection Scan Algorithm fue publicado, en marzo de 2017, ha sido implementado en varios formatos y lenguajes de programación. En el sitio web de Papers with Code, un portal que recopila códigos desarrollados sobre la idea central de diferentes papers, se muestran varias de estas implementaciones, enlazadas con su respectiva fuente de origen.

Destaca, entre estos, el repositorio de *ULTRA: UnLimited TRAnsfers for Multimodal Route Planning* [21], un framework desarrollado por el *Karlsruher Institut für Technologie* (KIT) en C++, para realizar planificaciones de viajes que incluyen diferentes medios de transporte. Este framework considera CSA, junto con otros algoritmos, para entregar posibles rutas entre dos puntos de una ciudad. Otra implementación disponible existe en el repositorio creado por Linus Norton, que creó una implementación del algoritmo en TypeScript [18]. Parte de estas implementaciones se pueden analizar en el código provisto en el anexo A.

Analizando el terreno actual, existe más de un ejemplo existente de implementación del algoritmo, en varios lenguajes de programación. Sin embargo, principalmente por el lenguaje en el que están programadas, la mayoría de estas implementaciones terminan siendo algo complejas de entender para alguien no especializado en ese *stack* tecnológico. Asimismo, existen programas o frameworks que dicen incluir a CSA en su procesamiento interno, pero no en todos estos casos el algoritmo se encuentra expuesto al público. Estas ideas motivan la decisión de que, en este trabajo de título, el algoritmo sea implementado en Python, por ser un lenguaje de alto nivel y de propósito general, lo que implica tener una fácil legibilidad de código y permite programar de forma más sencilla. La elección del lenguaje de programación motiva las decisiones posteriores de herramientas y librerías, que se mencionan en las secciones 2.2.1, 2.3.1, y 2.4, explicitadas y resumidas en la sección 3.1 del siguiente capítulo.

2.2. OpenStreetMap

OpenStreetMap (OSM) es un proyecto colaborativo cuyo propósito es crear mapas editables y de uso libre [10]. Los mapas, generados mediante la recopilación de información geográfica a través de dispositivos GPS móviles, incluyen detalles sobre las vías públicas (como pasajes, calles y carreteras), paradas de autobuses y diversos puntos de interés. Al ser un proyecto *Open-Source*, el desarrollo de los mapas locales es gestionado por organizaciones voluntarias de contribuyentes; en nuestro país, existe la Fundación OpenStreetMap Chile [2] cumpliendo ese papel. Se presenta la figura 2.3 a modo de ejemplo, donde se observa el mapa del Gran Santiago, en el que se pueden notar, entre otras vías, las carreteras más destacadas de la ciudad, como la Circunvalación Américo Vespucio y la Autopista Central.

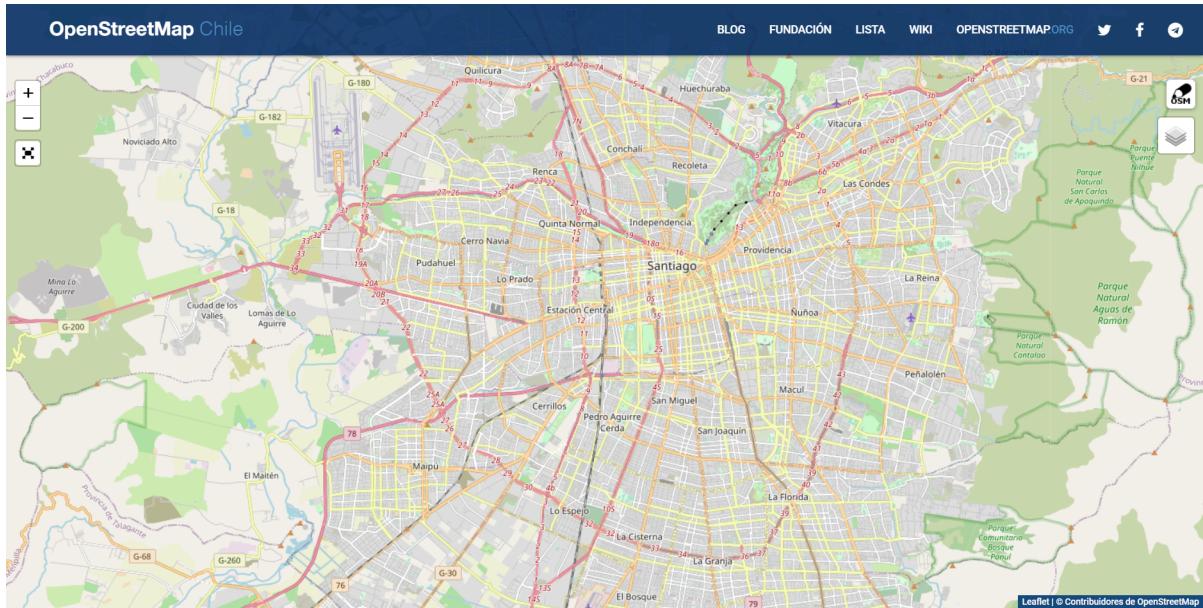


Figura 2.3: Mapa de Santiago en OpenStreetMap. Fuente: openstreetmap.cl .

2.2.1. OSM integrado en CSA

Connection Scan Algorithm (CSA) es capaz de alimentarse de los datos provenientes de OpenStreetMap para calcular la salida del algoritmo. Desde OpenStreetMap (OSM) se puede obtener la información cartográfica de una ciudad, la cual es utilizada para ubicar las coordenadas de los puntos de origen y destino en un mapa, y de esta forma obtener propiedades como la distancia entre los puntos, además de identificar detalles como las calles aledañas a las ubicaciones buscadas. Aquí también se obtienen las coordenadas de las paradas de los servicios de transporte público, tales como los paraderos de bus y las estaciones del Metro.

La librería **pyrosm** [22] funciona como un parser de la información de OSM en Python. **pyrosm** permite descargar la información más actualizada de la ciudad, almacenándola en un grafo dirigido donde cada nodo representan una intersección entre vías o algún lugar de interés, y las aristas entre los nodos representan las vías en sí; el que el grafo sea dirigido

responde al sentido de las vías (es diferente una calle que es *doble vía* a una que va en un solo sentido). Trabajar con grafos permite otorgar propiedades a los componentes, como las coordenadas a los nodos (su latitud y longitud) o el largo a las aristas (representando la distancia entre ambos nodos que conecta).

2.3. GTFS

Las Especificaciones Generales del Suministro de datos para el Transporte público, o en inglés, General Transit Feed Specification (GTFS), son un tipo de especificaciones ampliamente utilizado para definir y trabajar sobre datos de transporte público en las grandes ciudades. Este instrumento consiste en una serie de archivos de texto, recopilados en un archivo ZIP, de manera tal que cada archivo modela un aspecto específico de la información del transporte público, como paradas, rutas, viajes y horarios.

En la figura 2.4, se muestra el cómo se ven los archivos en GTFS, mostrando, como ejemplo, la información de los paraderos disponibles.

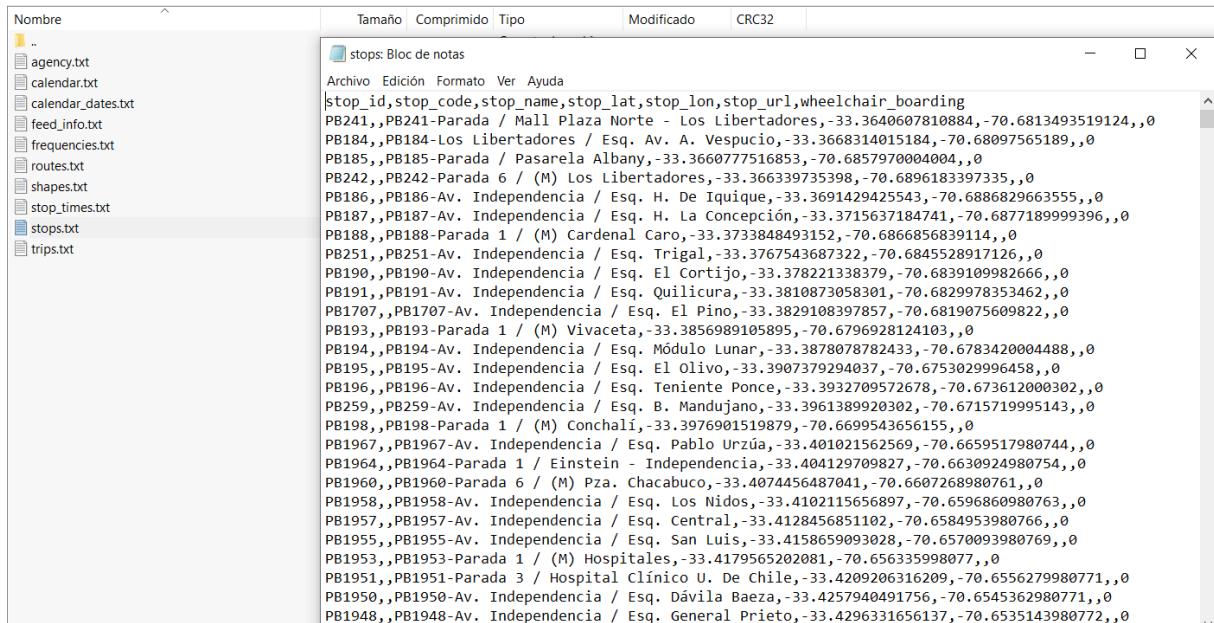


Figura 2.4: Visualización de archivos del feed GTFS para Santiago.

A nivel global, las organizaciones encargadas de la gestión administrativa del transporte público suelen utilizar este formato para compartir la información. En Santiago, el Directorio de Transporte Público Metropolitano (DTPM) es la entidad encargada de esta tarea. Este organismo, dependiente del Ministerio de Transportes y Telecomunicaciones, y cuya misión es mejorar la calidad del sistema de transporte público en la ciudad, tiene disponible públicamente esta información, y la actualiza periódicamente [6]. Al momento de la entrega de este informe, la última versión fue configurada para implementarse desde el 8 de julio de 2023.

La información en GTFS está contenida en diferentes archivos de texto, con sus valores separados por comas (similar a un CSV). Cada archivo concentra un área específica de los datos, las cuales se describen a continuación:

- **Agency**: entrega la información de las diferentes agencias de transporte que alimentan el GTFS. En este caso, se encuentra la Red Metropolitana de Movilidad (que engloba a todos los buses Red, antiguamente Transantiago), el Metro de Santiago, y EFE Trenes de Chile.
- **Calendar Dates**: especifica fechas especiales que alteran el funcionamiento habitual de los recorridos que varían por día. Para la última versión, este archivo contiene todas las fechas de feriados que caen entre lunes y sábado.
- **Calendar**: especifica los diferentes recorridos que varían por día, con su tiempo de validez. Acá se especifican los recorridos de Red para tres formatos diferentes: el cronograma para los días laborales (lunes a viernes), el cronograma para los sábados, y el cronograma para los domingos.
- **Feed Info**: información de la entidad que publica el GTFS.
- **Frequencies**: listado que, para todos los viajes de los recorridos disponibles, incluye sus tiempos de inicio y de término, y el *headway* o tiempo de espera estimado entre vehículos.
- **Routes**: contiene el identificador de cada ruta existente, su agencia, ubicación de origen y destino.
- **Shapes**: lista las diferentes 'formas' de los viajes de cada recorrido. Esto incluye el identificador de cada viaje (el recorrido y si acaso es de ida o retorno), y las latitudes y longitudes para cada secuencia posible.
- **Stop Times**: incluye las horas estimadas de llegada para que cada recorrido incluido en el GTFS llegue a cada parada incluida en su trayecto.
- **Stops**: contiene los identificadores, nombres, latitud y longitud de cada parada de transporte.
- **Trips**: contiene todos los viajes diferentes que realiza cada recorrido, señalando el nombre del recorrido, sus días de funcionamiento, si es de ida o retorno, y su dirección de destino.

Siguiendo este formato, los operadores de transporte pueden almacenar y publicar la información pertinente a sus sistemas, para que esta sea utilizada por las personas o entidades que lo estimen conveniente. Por ejemplo, los desarrolladores de aplicaciones que permitan a sus usuarios revisar el estado actual de los servicios de transporte público, con el fin de planificar sus viajes. Un ejemplo de flujo de información en el que estos datos pueden ser utilizados se detalla en el diagrama mostrado en la figura 2.5.

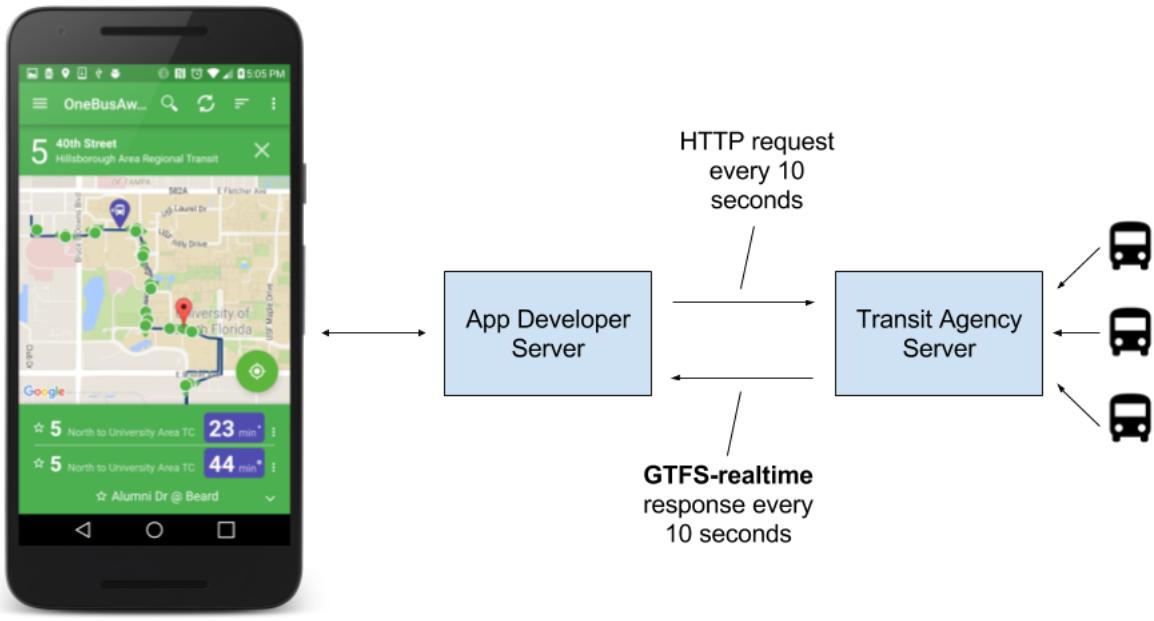


Figura 2.5: Diagrama de uso de datos en tiempo real en formato GTFS para una aplicación.
Fuente: watrifeed.ml.

En este ejemplo, se muestra cómo una aplicación móvil se conecta al servidor que almacena sus datos, el cual hace consultas periódicas al servidor de la agencia de tránsito. Este, al contener la información de los recorridos (en este caso, de buses), responde con información en tiempo real en formato GTFS, que finalmente el servidor de la aplicación interpreta para mostrarle al usuario la ruta en un mapa. Si bien este ejemplo muestra una aplicación con información en vivo, también pueden realizarse aplicaciones con la información programada de los recorridos.

2.3.1. GTFS integrado en CSA

Connection Scan necesita tener a su disposición la información del transporte público para ser capaz de calcular las rutas solicitadas, por lo que es útil entregarle dicha información en formato GTFS. De esta manera, el algoritmo tendrá información de los distintos recorridos disponibles, incluyendo sus rutas, paradas, horarios, y cualquier otra información que se estime necesaria para poder obtener la mejor ruta a seguir.

Similar al caso de OSM, existe una librería de Python llamada **pygtfs** [4], la que modela información en formato GTFS en una interfaz de Python. La librería almacena la información del transporte público en una base de datos relacional, tal que pueda ser usada en proyectos programados en este lenguaje.

2.4. Datos: estructura y manejo de la información

Implementar un algoritmo como CSA requiere trabajar con un gran volumen de datos. Sin ir más lejos, considerando el cómo se definen los nodos y aristas de OSM en **pyrosm** (como se mencionó en la sección 2.2.1), se deduce que, para una ciudad como Santiago, existe un volumen importante de información que debe almacenarse para poder operar con ella. Es por esta razón que es crucial saber elegir una buena herramienta para la creación de las estructuras de datos correspondientes. En esta misma línea, el tipo de estructura de datos a utilizar viene dado, precisamente, por la forma en la que se almacena la información de OSM: grafos. Dicho esto, y dado que existen múltiples librerías que manejan este tipo de estructura en Python, se debe elegir una que se adecúe mejor a las necesidades de este proyecto.

Una alternativa muy utilizada en conjunto a **pyrosm** es **networkx**, un paquete de Python para la creación, manipulación, y estudio de la estructura, dinámica, y funciones de redes complejas [7]. Esta librería está disponible para sistemas operativos Windows mediante **pip**, el sistema de gestión de paquetes de Python. La razón por la cual es ampliamente utilizada es por su simplicidad en el manejo y operación de la información. Sin embargo, su principal problema recae en su rendimiento, pues al estar programada completamente en Python, su desempeño es lento en comparación a otras opciones. Si, además, se toma en cuenta el gran volumen de datos que se requiere almacenar, se infiere que el uso de **networkx** terminará generando un importante *bottleneck* o cuello de botella en el desempeño del algoritmo.

Por los motivos antes mencionados, se decide utilizar una librería diferente para este fin. La opción seleccionada es **graph-tool**, un módulo de Python creado para la manipulación y análisis de grafos [5]. A diferencia de otras herramientas, **graph-tool** posee la ventaja de tener una base algorítmica implementada en C++, un lenguaje de programación basado en compilación, por lo que su desempeño es mucho más eficiente. Esto permite trabajar con grandes volúmenes de información de mejor manera, por lo que demuestra ser una excelente librería para utilizar en este proyecto. Cabe destacar, eso sí, que **graph-tool** solo se encuentra disponible para sistemas operativos GNU/Linux y MacOS. Como consecuencia, la programación del algoritmo se realiza en Ubuntu, una distribución de GNU/Linux, mediante WSL2 (Windows Subsystem for Linux 2) [17].

Capítulo 3

Diseño

3.1. Stack tecnológico

Basado en lo obtenido del capítulo anterior, se genera un formato para diseñar la solución propuesta. Para poder simular de manera correcta el funcionamiento del algoritmo, se debe procesar correctamente la información, tanto los datos cartográficos de Santiago, como la información del transporte público. Para desarrollar el algoritmo, se define lo siguiente:

- El algoritmo será programado en Python.
- La información cartográfica se obtendrá desde OpenStreetMap, procesada mediante la librería **pyrosm** [22].
- La información del transporte público se obtendrá en el formato GTFS, procesada mediante la librería **pygtfs** [4]. Los datos serán obtenidos previamente, descargando la última versión del sitio web del Directorio de Transporte Público Metropolitano [3].
- Para almacenar y trabajar con la información obtenida, se utilizará la librería **graph-tool** [5] para trabajar con grafos.

Otras librerías que son utilizadas para cumplir de mejor forma los objetivos especificados en la sección 1.1 responden a la necesidad de procesar, tanto la entrada como la salida del algoritmo, de una forma tal que facilite su funcionamiento para el usuario final. En primer lugar, la librería **Nominatim** [13] permite que el usuario pueda ingresar como entrada una dirección en palabras, en vez de coordenadas numéricas, lo que facilita el uso del algoritmo y resta la necesidad de obtener las coordenadas de los puntos deseados por otro medio; **No-minatim** permite realizar la geocodificación de estas direcciones, buscando sus coordenadas en los datos de OpenStreetMap. En segundo lugar, la librería **folium** [9] permite visualizar datos cartográficos en un mapa de fácil uso. **folium** está basado en la librería **Leaflet.js** de JavaScript, por lo que aprovecha todas sus características para generar un mapa interactivo.

3.2. Funcionamiento lógico

Con el fin de llegar al resultado esperado, Santiago estará representada como una red de capas. La primera vendrá entregada por OpenStreetMap mediante **pyrosm**, conformada por la infraestructura urbana de la ciudad (calles, edificios, etc.) La segunda provendrá de la información de transporte, proveniendo del GTFS del Directorio de Transporte Público Metropolitano, mediante **pygtfs**. De esta manera, el algoritmo entregará un conjunto de caminos en esa red, desde un punto de la infraestructura urbana hasta otro punto dentro de la misma, que pueden ser recorridos usando la red de transporte público.

Connection Scan Algorithm requiere de un punto de inicio y uno de destino como entrada para funcionar, y mediante estos calcula la ruta a seguir para efectuar el viaje. Sin embargo, luego de estudiar la estructura de GTFS y la información disponible para Santiago, se concluye que, además de considerar los puntos antes mencionados, el algoritmo debería obtener además una fecha y una hora de inicio del viaje, para entregar una respuesta que represente mejor la realidad del servicio disponible. Dado que no todos los recorridos funcionan en los mismos horarios, y algunos solo funcionan ciertos días de la semana, es provechoso tomar en consideración estas variables temporales. Los archivos que componen el GTFS provisto para Santiago incluyen un listado de días feriados, por lo que también se consideran como excepciones a los recorridos regulares.

La motivación del algoritmo, tal como fue discutida anteriormente, es obtener la *mejor* ruta para llegar al destino. En este caso, se prioriza la ruta que posea menos transbordos, y dentro de estas, la ruta cuyo recorrido demore menos en llegar desde la hora de inicio. Por lo tanto, el algoritmo debe procesar las paradas cercanas a los puntos de inicio y destino, obtener los recorridos que pasan por ellas, y teniendo la información de sus rutas y tiempos de espera, generar la mejor ruta y entregarla como salida al usuario.

Para poder realizar esto, el funcionamiento del algoritmo sería:

- El usuario ejecuta el programa.
- El programa descarga la información de OpenStreetMap y la procesa, en conjunto con la información de GTFS, y almacena todo en grafos.
- Al correr el algoritmo, el usuario ingresa una fecha y hora para el inicio del viaje, una dirección de origen y una dirección de destino.
- El programa busca las direcciones para obtener sus coordenadas, en la capa de OpenStreetMap.
- Con las coordenadas obtenidas y la información de los paraderos, se buscan los más cercanos y los recorridos que se detienen en ellos.
- Se busca la existencia de un recorrido que conecte ambos puntos. De no existir, se buscan los paraderos en los que se detienen los recorridos que llegan al destino, para revisar si alguno de los recorridos que parten del origen terminan allí y es posible hacer un transbordo. De esta forma, se consigue la secuencia de vehículos a tomar.
- Con los recorridos necesarios obtenidos, se calculan los tiempos de viaje. Se obtiene el tiempo de espera para el primer bus y se calcula cuánto tiempo demora la ruta hasta llegar al destino.

- Finalmente, se entrega la salida del algoritmo. Con las coordenadas de los puntos de origen y destino, además de las coordenadas de los paraderos necesarios para efectuar la ruta, se grafican los pasos a seguir en un mapa. Por otro lado, se entrega el tiempo de viaje como texto.

Como nota aparte, la implicancia de la estructuración propuesta es que, de cambiar la información por la de una ciudad diferente, el algoritmo también sería capaz de entregar una respuesta válida. Esto quiere decir que el algoritmo deberá estar abierto a recibir datos de GTFS y de OpenStreetMap provenientes de cualquier lugar. Evidentemente, como el enfoque de este trabajo de título es estudiar la movilidad vial en Santiago de Chile, el resto de posibilidades escapa de su alcance, y no serán consideradas. Sin embargo, pueden sentar bases para realizar trabajo futuro.

3.3. Criterio de Evaluación

El usuario final del proyecto será cualquier individuo u organización que requiera información sobre cómo movilizarse en Santiago. Por ejemplo, una persona que desee saber la mejor forma de movilizarse desde Padre Hurtado hasta Vitacura, o una entidad gubernamental (como la DTPM) que requiera estudiar las rutas más utilizadas por la gente para repartirles los recursos correspondientes. Por esto, para asegurar el éxito del proyecto, el criterio que se usará para evaluar que el resultado del trabajo cumpla su objetivo, es:

- El usuario final deberá ser capaz de, haciendo uso del algoritmo propuesto como solución, identificar patrones de movilidad a través de Santiago de Chile.

Posterior a la implementación, se realiza un caso de estudio para analizar la movilidad en Santiago. La finalidad es probar la efectividad y eficiencia de la solución desarrollada, y utilizar sus resultados para comparar la situación en la ciudad bajo diferentes condiciones. El caso de estudio definido consiste en estudiar las rutas disponibles en Santiago, antes y después de la implementación de la Línea 3 del Metro (la última en inaugurarse al momento del desarrollo de este trabajo de título). Para efectuarlo, se definen una serie de puntos de origen y destino que actualmente puedan conectarse realizando viajes en la Línea 3, y se obtienen tanto rutas que incluyan a la L3, como otras que la saquen de las posibilidades. Luego, se comparan los tiempos estimados de viaje, para así analizar qué tanto efecto tiene esta línea de Metro sobre los trayectos de las personas. El desarrollo del Caso de Estudio está documentado en la sección 5.2 del informe.

Capítulo 4

Implementación

En el presente capítulo, se detalla la implementación realizada de Connection Scan Algorithm y todo el trabajo que corresponde a su desarrollo. El código fuente de la implementación ha sido almacenado en un repositorio de GitHub creado para este fin [16].

4.1. Obtención de datos

4.1.1. Procesamiento de OpenStreetMap

La información almacenada en OpenStreetMap puede ser descargada en formato PBF (Protocolbuffer Binary Format), para luego ser filtrada y procesada según lo necesitado. Geofabrik, un portal comunitario para proyectos relacionados con OpenStreetMap [23], tiene disponible para descarga la información de los distintos países del mundo, incluído Chile [24]. Con esto, es posible obtener la información geoespacial de Santiago y trabajar con ella, para lo cual es necesario procesarla correctamente. En un principio, se pretendía realizar este proceso manualmente, pero se descubrió una mejor alternativa, que permite automatizarlo.

pyrosm [22], la librería utilizada para procesar la información, permite leer datos de OpenStreetMap en formato PBF e interpretarla en estructuras de GeoPandas [15], librería de Python de código abierto para trabajar con datos geoespaciales. Además de esto, **pyrosm** también permite directamente descargar la información de una ciudad y actualizarla en caso de existir una versión anterior en el directorio, permitiendo automatizar este proceso. De esta forma, una vez descargada la información de Santiago, se pueden crear gráficos según se necesite para su representación.

Los datos de OSM se obtienen mediante el método `get_network` de `pyrosm`, obteniendo así nodos y aristas. Esto se logra ejecutando el código 4.1:

```
1 fp = get_data(  
2     "Santiago", # Nombre de la ciudad  
3     update=True,  
4     directory=OSM_PATH  
5 )  
6  
7 osm = OSM(fp)  
8 # Se obtienen los datos en nodos y aristas  
9 nodes, edges = osm.get_network(nodes=True)  
10 osm_graph = Graph()
```

Código 4.1: Fragmento de `get_osm_data()` que se encarga de obtener los datos de OSM.

Como fue mencionado anteriormente, los nodos representan tanto puntos de interés de una ciudad, así como intersecciones entre calles. Por su parte, las aristas representan las calles en sí.

Estos datos se almacenan en un grafo de `graph-tool` (llamado `osm_graph` en el código anterior), siguiendo la misma estructura de nodos y aristas. A ambos grupos se les otorgan propiedades en el grafo para permitir acceder a su información de manera consistente. Los nodos almacenan su latitud y longitud en propiedades separadas, además de almacenar sus identificadores (un id interno del grafo, y su id otorgado en la estructura de OpenStreetMap). Por otro lado, las aristas tienen propiedades para marcar los nodos que conectan (una para el nodo de origen y otra para el nodo de destino), y una extra para representar la distancia que mide la arista. Iterando por todos los nodos y aristas existentes, se almacenan con sus propiedades respectivas, y termina la obtención de datos de OpenStreetMap.

Estudiando el grafo obtenido, se observó que es un grafo dirigido, es decir, sus aristas poseen una dirección fija, ocasionando que las conexiones entre nodos no sean, necesariamente, en ambas direcciones, pues puede existir el caso en el que un nodo u esté conectado a un nodo v y sea posible moverse en dicha dirección $u \rightarrow v$, pero sea imposible realizar el movimiento en la dirección opuesta, $u \leftarrow v$. En el trabajo posterior, se observaron problemas para trabajar con el grafo dirigido, puesto que la búsqueda de paraderos cercanos a los puntos de origen y destino se dificultaba por esta razón. Al tener que moverse por las calles para llegar a la parada, las caminatas desde y hacia una parada se vieron afectadas por la dirección de las aristas que conectaban los nodos, generando un comportamiento innecesario y complejo de sortear. Por este motivo, se decidió trabajar con una copia no dirigida del grafo original, poseyendo la misma estructura y propiedades, con la única diferencia de perder la dirección de las aristas. Así, se genera una versión no dirigida tomando como base el grafo inicial, recorriéndolo y copiando su estructura.

Naturalmente, sobre este grafo se gestionan las búsquedas de direcciones para los puntos de origen y destino. Para realizar esta gestión, se crea la función `address_locator`, presentada en el código 4.2:

```

1 def address_locator(graph, address):
2     # Suscripcion al agente de geocodificacion
3     geolocator = Nominatim(user_agent="ayatori")
4     while True: # Probar estado del servicio
5         try:
6             location = geolocator.geocode(address)
7             break
8         except GeocoderServiceError:
9             i = 0
10            if i < 15:
11                print("Geocoding service error. Retrying in 5 seconds...")
12                tm.sleep(5)
13                i+=1
14            else:
15                msg = "Error: Too many retries. Geocoding service may be
down. Please try again later."
16                print(msg)
17                return
18        if location is not None:
19            # Se obtienen las coordenadas
20            lat, lon = location.latitude, location.longitude
21            nearest = find_nearest_node(graph, lat, lon)
22            return nearest
23    msg = "Error: Address couldn't be found."
24    print(msg)

```

Código 4.2: Definición de `address_locator()`

Esta función recibe el grafo creado con la información de OpenStreetMap (graph) y una dirección (address), suscribe un agente de geocodificación con un nombre (en este caso, *ayatori*), e intenta buscar las coordenadas de la dirección mediante la librería **Nominatim**. Evidentemente, el funcionamiento del algoritmo depende directamente del estado del servicio de **Nominatim**, por lo que si ese servicio se encuentra caído en algún momento, el algoritmo no funcionará. Por esta razón, la función previene este caso, intentando acceder al servicio 3 veces; si no está disponible, muestra el error correspondiente.

Por otro lado, también puede existir el caso en el que la dirección no se encuentre (por ejemplo, por estar mal escrita), para cuyo caso también se gestiona el mensaje de error correspondiente. Si la dirección se logra encontrar, la función obtiene sus coordenadas, y luego entrega el nodo más cercano. La razón de esto es que la dirección buscada puede no coincidir exactamente con un nodo del grafo de OpenStreetMap, así que lo más correcto sería entregar el nodo más cercano. Esto se realiza con la función `find_nearest_node`, mostrada en el código 4.3:

```

1 def find_nearest_node(graph, latitude, longitude):
2
3     query_point = np.array([longitude, latitude])
4
5     # Obtiene las propiedades de vertice (coordenadas)
6     lon_prop = graph.vertex_properties['lon']
7     lat_prop = graph.vertex_properties['lat']
8
9     # Calcula la distancia entre el nodo y la direccion buscada
10    distances = np.linalg.norm(np.vstack((lon_prop.a, lat_prop.a)).T -

```

```

query_point, axis=1)

11
12 # Halla el indice del nodo mas cercano
13 nearest_node_index = np.argmin(distances)
14 nearest_node = graph.vertex(nearest_node_index)
15
16 return nearest_node

```

Código 4.3: Definición de `find_nearest_node` para hallar el nodo más cercano.

Este procedimiento se realiza para ambas direcciones entregadas por el usuario. Así, se obtienen las coordenadas de los puntos de origen y destino, y se puede comenzar a calcular la mejor ruta para efectuar el viaje.

4.1.2. Procesamiento de GTFS

El formato GTFS incluye múltiples archivos de texto que almacenan la información del transporte público, organizada según diversos criterios, tal y como se especificó en la sección 2.3. Toda la información viene en un archivo comprimido ZIP, descargado desde la web del DPTM [3]. Este archivo debe descargarse manualmente, y las versiones nuevas salen cada uno o dos meses. Sin embargo, las diferencias entre versiones suelen tener relativamente pocas diferencias.

`pygtfs` [4], la librería utilizada para procesar los datos de GTFS, posee un módulo llamado `Schedule`, encargado de gestionar toda la información. Instanciando el método al crear una nueva variable, permite obtener la información de GTFS y enlazarla a ella. Esto se muestra en el código 4.4, mediante la variable `sched`:

```

1 # Create a new schedule object using a GTFS file
2 sched = pygtfs.Schedule(":memory:")
3 pygtfs.append_feed(sched, "gtfs.zip")

```

Código 4.4: Fragmento de `get_gtfs_data()` que procesa los datos en GTFS.

En este fragmento, se solicita la memoria necesaria para generar la instancia del `scheduler`, y se procesa la información descargada previamente (el archivo `gtfs.zip`). Luego, se puede acceder a la información de cada archivo de texto como si fuese un método de `sched`. Por ejemplo, para obtener la información de las paradas, basta con llamar a `sched.stops`, y para obtener los servicios o rutas, se llama a `sched.routes`.

Al igual que en el caso anterior, estos datos se almacenan, inicialmente, en un grafo de `graph-tool`. En este caso, se guardan en los nodos los identificadores de las paradas, y las aristas conectan las paradas entre sí, generando la red de transporte. Sin embargo, durante el desarrollo de la implementación, se llegó a la conclusión de que usar un grafo para esta parte del código no era la mejor opción. Esto es debido a que existen algunos datos que dependen no solamente de una parada o de una ruta, sino que de la unión entre ambos, es decir, hay información que tiene sentido solamente al solapar los datos de una parada con los de una ruta. En específico, estos son:

- Orientación: refiere al sentido de la ruta cuando se detiene en una parada en específico. Cada ruta tiene un recorrido de ida y uno de vuelta, y por lo general, solo se detiene en un determinado paradero en uno de los sentidos.
- Número de Secuencia: al realizar una ruta en una orientación dada, el número de secuencia es el valor ordinal de una parada para esa ruta. En palabras simples, representa el orden en el que la ruta pasa por las paradas (la primera parada, la segunda, la tercera, etc.)
- Tiempos de llegada: representa la hora aproximada en la que una ruta llega a una parada.

La orientación y el número de secuencia deben utilizarse para filtrar las rutas que sirven para viajar entre dos puntos del mapa, mientras que los tiempos de llegada son cruciales para elegir la mejor ruta y entregar el resultado. Sin embargo, ninguno de estos datos son inherentes a una parada o a una ruta, pues, por ejemplo, no se puede decir que una ruta *posee* una orientación, sino que pasa por una parada al ir en cierta orientación. Por estos motivos, finalmente, se opta por usar un diccionario anidado para almacenar la información, el cual se denomina **route_stops**.

Este diccionario anidado, o diccionario de diccionarios, tiene como primera llave el identificador de la ruta, y luego posee un diccionario para cada parada por la que pasa dicha ruta. Cada diccionario interno, además de poseer los identificadores respectivos y las coordenadas correspondientes, tiene los tres datos mencionados anteriormente. En el código 4.5, se muestra cómo se define esta estructura:

```

1 route_stops[route.route_id][stop_id] = {
2     "route_id": route.route_id,
3     "stop_id": stop_id,
4     "coordinates": stop_coords[route.route_id][stop_id],
5     "orientation": "round" if orientation == "I" else "return",
6     "sequence": sequence,
7     "arrival_times": []
8 }
9 arrival_time = (datetime.min + stop_times[i].arrival_time).time()

```

Código 4.5: Fragmento de `get_gtfs_data()` que almacena la información de rutas y paradas.

Utilizando el diccionario **route_stops**, es posible crear múltiples funciones que sean de utilidad para generar la ruta del viaje. Por ejemplo, `get_near_stop_ids`, para obtener los identificadores de las paradas cercanas a un punto del mapa. Esta función recibe como entrada el diccionario en cuestión, una tupla de coordenadas, y un margen numérico. Iterando sobre los elementos del diccionario, obtiene las coordenadas de cada parada, y revisa si está a una distancia cercana de las coordenadas entregadas, cercanía dada por el margen entregado. La existencia de este margen permite, en la práctica, modificar la distancia máxima hasta la cual una parada se considera *cercana* a los puntos del mapa. En el código 4.6, se puede observar la definición de la función:

```

1 def get_near_stop_ids(route_stops, coords, margin):
2     stop_ids = []
3     orientations = []
4     for route_id, stops in route_stops.items():

```

```

5     for stop_info in stops.values():
6         stop_coords = stop_info["coordinates"]
7         distance = haversine(coords[1], coords[0], stop_coords[1],
8                               stop_coords[0])
9         if distance <= margin:
10            orientation = stop_info["orientation"]
11            stop_ids.append(stop_info["stop_id"])
12            orientations.append((stop_info["stop_id"], orientation))
13
14    return stop_ids, orientations

```

Código 4.6: Definición de `get_near_stop_ids()` para obtener las paradas cercanas a un punto.

Como se ve en la definición de la función, para calcular la distancia entre el punto y las paradas, se usa la función `haversine`, definida en el código 4.7:

```

1 def haversine(lon1, lat1, lon2, lat2):
2     R = 6372.8 # Radio de la Tierra en kilometros
3     dLat = radians(lat2 - lat1)
4     dLon = radians(lon2 - lon1)
5     lat1 = radians(lat1)
6     lat2 = radians(lat2)
7     a = sin(dLat / 2)**2 + cos(lat1) * cos(lat2) * sin(dLon / 2)**2
8     c = 2 * asin(sqrt(a))
9     return R * c

```

Código 4.7: Definición de `haversine()` para calcular la distancia entre dos puntos.

La fórmula Haversine, o fórmula del semiverseno en español, es una ecuación que calcula la distancia entre dos puntos de una esfera, en base a su longitud y latitud. Esta fórmula es ampliamente utilizada en la navegación astronómica, pues permite calcular de forma fidedigna la distancia entre dos puntos del planeta.

Otra función importante que ha sido creada utilizando `route_stops` es `connection_finder`. Esta obtiene el diccionario y los identificadores de dos paradas como entrada, y luego de revisar todos los recorridos, entrega el listado de aquellos que se detienen en ambas paradas, es decir, los recorridos que pueden tomarse para ir de la primera parada a la segunda. La implementación de `connection_finder` se puede observar en el código 4.8:

```

1 def connection_finder(route_stops, stop_id_1, stop_id_2):
2     connected_routes = []
3     for route_id, stops in route_stops.items():
4         stop_ids = [stop_info["stop_id"] for stop_info in stops.values()]
5         if stop_id_1 in stop_ids and stop_id_2 in stop_ids:
6             connected_routes.append(route_id)
7
7

```

Código 4.8: Definición de `connection_finder()` para obtener las rutas que visitan dos paradas específicas.

Con esto, se puede realizar un uso correcto de los datos provistos por ambas capas de información para ejecutar el algoritmo. Si bien, en esta sección se omite el código completo, gran parte de este puede revisarse en el anexo B. Luego, queda analizar cómo utilizar estas herramientas para generar la salida del programa.

4.2. Generación del output

Tal como se mencionó en la sección 3.2 del capítulo de Diseño, el *output* del algoritmo está conformado por dos partes, explicadas a continuación.

4.2.1. Texto

La primera parte del output está en formato de texto, que consiste en las instrucciones que el usuario debe seguir para realizar la mejor ruta entre su punto de origen y su punto de destino. Estas instrucciones incluyen los identificadores de los recorridos a tomar, los identificadores de los paraderos de subida y bajada, y el tiempo aproximado del viaje. El tiempo se calcula con la diferencia entre la hora a la que pasa el siguiente bus del recorrido deseado y la hora actual, a lo que se le agrega el tiempo aproximado que demorará en llegar al paradero de destino. Esto se puede realizar al leer y cruzar la información de dos archivos de GTFS: *stop_times.txt* y *frequencies.txt*, lo que permite saber los rangos de tiempo en los que opera cada servicio y la frecuencia de salida de nuevos buses dentro de estos, además del tiempo que demora un bus desde que inicia el recorrido y hasta que llega a una parada en específico.

Para poder obtener y trabajar con estos tiempos estimados, se crean un par de funciones. Primero, la función `get_arrival_times()`, que permite obtener los tiempos de llegada de una ruta específica a una parada específica, considerando las frecuencias de la ruta y la fecha del viaje. En el código 4.9, se puede observar su implementación:

```
1 def get_arrival_times(route_id, stop_id, source_date):
2     # Lee las frecuencias desde frequencies.txt
3     frequencies = pd.read_csv("frequencies.txt")
4
5     # Filtra las frecuencias de la ruta especificada
6     route_frequencies = frequencies[frequencies["trip_id"].str.startswith(
7         route_id)]
8
9     # Obtiene el sufijo del dia
10    day_suffix = get_trip_day_suffix(source_date)
11
12    # Obtiene los tiempos de viaje para cada trayecto de la ruta
13    stop_route_times = []
14    bus_orientation = ""
15    for _, row in route_frequencies.iterrows():
16        start_time = pd.Timestamp(row["start_time"])
17        # Nota: el DTPM almacena el ultimo tiempo de llegada como
18        # "24:00:00", por lo que se considera como "23:59:59" para que no genere
19        # conflictos tecnicos con las librerias utilizadas
20        if row["end_time"] == "24:00:00":
21            end_time = pd.Timestamp("23:59:59")
22        else:
23            end_time = pd.Timestamp(row["end_time"])
24
25        headway_secs = row["headway_secs"] # Tiempo entre salidas
26        round_trip_id = f"{route_id}-{I}-{day_suffix}" # ID del trayecto de
```

```

1  ida
24     return_trip_id = f"{route_id}-R-{day_suffix}" # ID del trayecto de
      vuelta
25
26     # Se obtienen los tiempos para ida y vuelta
27     round_stop_times = pd.read_csv("stop_times.txt").query
28     (f"trip_id.str.startswith('{round_trip_id}') and stop_id == '{{
      stop_id}'")
29     return_stop_times = pd.read_csv("stop_times.txt").query
30     (f"trip_id.str.startswith('{return_trip_id}') and stop_id == '{{
      stop_id}'")
31
32     if len(round_stop_times) == 0 and len(return_stop_times) == 0:
33         return
34     elif len(round_stop_times) > 0:
35         bus_orientation = "round"
36         stop_time = pd.Timestamp(round_stop_times.iloc[0] ["arrival_time"])
37     elif len(return_stop_times) > 0:
38         bus_orientation = "return"
39         stop_time = pd.Timestamp(return_stop_times.iloc[0] ["arrival_time"])
40     for freq_time in pd.date_range(start_time, end_time, freq=f"{{headway_secs}}"):
41         freq_time_str = freq_time.strftime("%H:%M:%S")
42         freq_time = datetime.strptime(freq_time_str, "%H:%M:%S")
43         stop_route_time = datetime.combine(datetime.min, stop_time.
44 time()) + timedelta(seconds=(freq_time - datetime.min).seconds)
45         if stop_route_time not in stop_route_times:
46             stop_route_times.append(stop_route_time)
47             stop_time += pd.Timedelta(seconds=headway_secs)
48
49     return bus_orientation, stop_route_times

```

Código 4.9: Definición de `get_arrival_times()` para obtener los tiempos de llegada de una ruta a una parada.

La función `get_arrival_times()` es crucial para elegir la mejor ruta existente, puesto que sienta la base para considerar los tiempos de espera de los buses.

Por otro lado, la función `get_time_until_next_bus()` permite filtrar estos tiempos de llegada, considerando la hora y fecha de inicio del viaje. Esta toma todos los tiempos de llegada, y a partir de la hora, obtiene los tiempos de llegada para los siguientes 3 buses. La implementación de esta función puede observarse en el código 4.10:

```

1 def get_time_until_next_bus(arrival_times, source_hour, source_date):
2     arrival_times_remaining = []
3     for a_time in arrival_times:
4         # Filtra los tiempos de llegada posteriores a la hora de inicio
5         if a_time.time() >= source_hour:
6             arrival_times_remaining.append(a_time)
7     if len(arrival_times_remaining) == 0:
8         return None
9     else:
10        # Ordena los tiempos filtrados en orden ascendente
11        arrival_times_remaining.sort()

```

```

12
13     # Para los siguientes 3 buses, obtiene sus tiempos de llegada,
14     # considerando la fecha del viaje
15     next_buses = []
16     for i in range(min(3, len(arrival_times_remaining))):
17         next_arrival_time = arrival_times_remaining[i]
18         next_bus = datetime.combine(next_arrival_time.date(),
19                                     next_arrival_time.time())
20         next_buses.append(next_bus)
21
22     if next_buses is None:
23         print("No buses remaining for the specified date.")
24     else:
25         # Calcula el tiempo de espera para los siguientes buses
26         time_until_next_buses = []
27         for next_bus in next_buses:
28             time_until_next_bus = (next_bus - datetime.combine(
29             next_bus.date(), source_hour)).total_seconds()
30             minutes, seconds = divmod(time_until_next_bus, 60)
31             time_until_next_buses.append((int(minutes), int(seconds)))
32
33     return time_until_next_buses

```

Código 4.10: Definición de `get_time_until_next_bus()` para obtener el tiempo de espera para los siguientes buses de una ruta en una parada.

La razón por la cual se decidió que `get_time_until_next_bus()` filtrara los tiempos para los siguientes 3 buses, responde a la necesidad de informarle al usuario más de una opción para tomar dicho bus. Dado que se trabaja con cronogramas y tiempos estimados, pueden existir diferencias con los tiempos reales de llegada de los buses. Por este motivo, se consideró que entregar el tiempo del siguiente bus y dos buses adicionales consistiría en una buena decisión de diseño.

Además, se define la función `get_travel_time()`, que permite obtener el tiempo aproximado que le toma a un bus, de una ruta determinada, viajar entre dos paradas de su trayecto. Su utilidad es permitir calcular el tiempo completo del viaje, para entregarle esta información al usuario en el output de texto. La definición de esta función se puede apreciar en el código 4.11:

```

1 def get_travel_time(trip_id, stop_ids):
2     stop_times = pd.read_csv("stop_times.txt").query(f"trip_id.str.
3     startswith('{trip_id}') and stop_id in {stop_ids}")
4     if len(stop_times) < 2:
5         return None
6     arrival_times = [datetime.strptime(arrival_time, "%H:%M:%S") for
7     arrival_time in stop_times["arrival_time"]]
8     travel_time = arrival_times[1] - arrival_times[0]
9     return travel_time

```

Código 4.11: Definición de `get_travel_time()` para obtener el tiempo aproximado de viaje para una ruta entre dos de sus paradas.

Uno de los detalles importantes a tener en cuenta, a la hora de calcular los tiempos, es el sentido u orientación del recorrido, pues de no considerar esto, se podría obtener un tiempo de llegada que no se apegue a la realidad. Cada recorrido del transporte público de Santiago posee un viaje de ida y un viaje de regreso, los cuales difieren en paradas y tiempos relativos para todos los buses (solo exceptuándose el Metro), por lo que no basta con buscar una ruta que sirva para realizar un viaje determinado, sino que su orientación sea concordante. Desafortunadamente, no existe una columna en algún archivo que posea esta información directamente. Sin embargo, los identificadores del archivo *frequencies.txt* son *strings* compuestos que incluyen, además del identificador del recorrido y del tipo de viaje según el día de la semana, un identificador que señala la orientación del bus, pudiendo filtrar esta información correctamente.

Así, se puede procesar lo necesario para entregar esta parte de la salida del algoritmo: el usuario lee en pantalla el recorrido del bus que debe tomar, en qué parada subirse y en cuál bajarse. Adicionalmente, se le dice a qué hora pasa el próximo bus del recorrido, además de los dos posteriores a ese, junto al tiempo de viaje sobre el bus y el tiempo total que demorará el trayecto. A modo de ejemplo se presenta, en la figura 4.1, la salida en texto del algoritmo, cuando se busca una ruta para ir desde la Facultad de Ciencias Físicas y Matemáticas hasta Puente Cal y Canto, a las 10:00 durante un día de semana:

```
To go from: Universidad de Chile (Facultad de Ciencias Físicas y Matemáticas), 850, Avenida Almirante Blanco Encalada, Barrio Ejército, Santiago, Provincia de Santiago, Región Metropolitana de Santiago, 8370456, Chile
To: Puente Cal y Canto, Avenida Cardenal José María Caro, Centro Histórico, Santiago, Provincia de Santiago, Región Metropolitana de Santiago, 8320012, Chile

The best option is to walk for 1 minutes and 0 seconds to stop PA433, and take the route 509.
The next bus arrives at 10:01.
The other two next buses arrives in:
2 minutes, 32 seconds (10:02)
9 minutes, 12 seconds (10:09)

You will get off the bus on stop PA300 after 13 minutes and 22 seconds.
After that, you need to walk for 1 minutes and 6 seconds to arrive at the target spot.
Total travel time: 15 minutes, 30 seconds. You will arrive your destination at 10:15.
```

Figura 4.1: Ejemplo de salida del algoritmo (texto). Ruta entre FCFM y Puente Cal y Canto.

Se puede apreciar cómo se señala que la mejor opción es subirse al recorrido 509 en la parada PA433 hasta bajarse en la parada PA300. Además, se muestran los tiempos de caminata hasta y desde las paradas, las horas de llegada de los siguientes buses, y el tiempo estimado para estar arriba del bus. Finalmente, se entrega el tiempo completo de viaje, el cual es de 15 minutos y 30 segundos, y se menciona la hora de llegada, 10:15.

4.2.2. Mapa

La segunda parte de la salida consiste en el mapa interactivo que grafica los puntos relevantes de la ruta calculada. Este mapa debe ser un complemento para el texto, destacando los puntos de origen y destino, y las paradas pertinentes. De esta forma, el usuario final puede entender mejor la información otorgada.

Como fue mencionado anteriormente, los mapas son generados usando la librería **folium**. En la figura 4.2, se muestra un mapa de Santiago generado con esta librería:

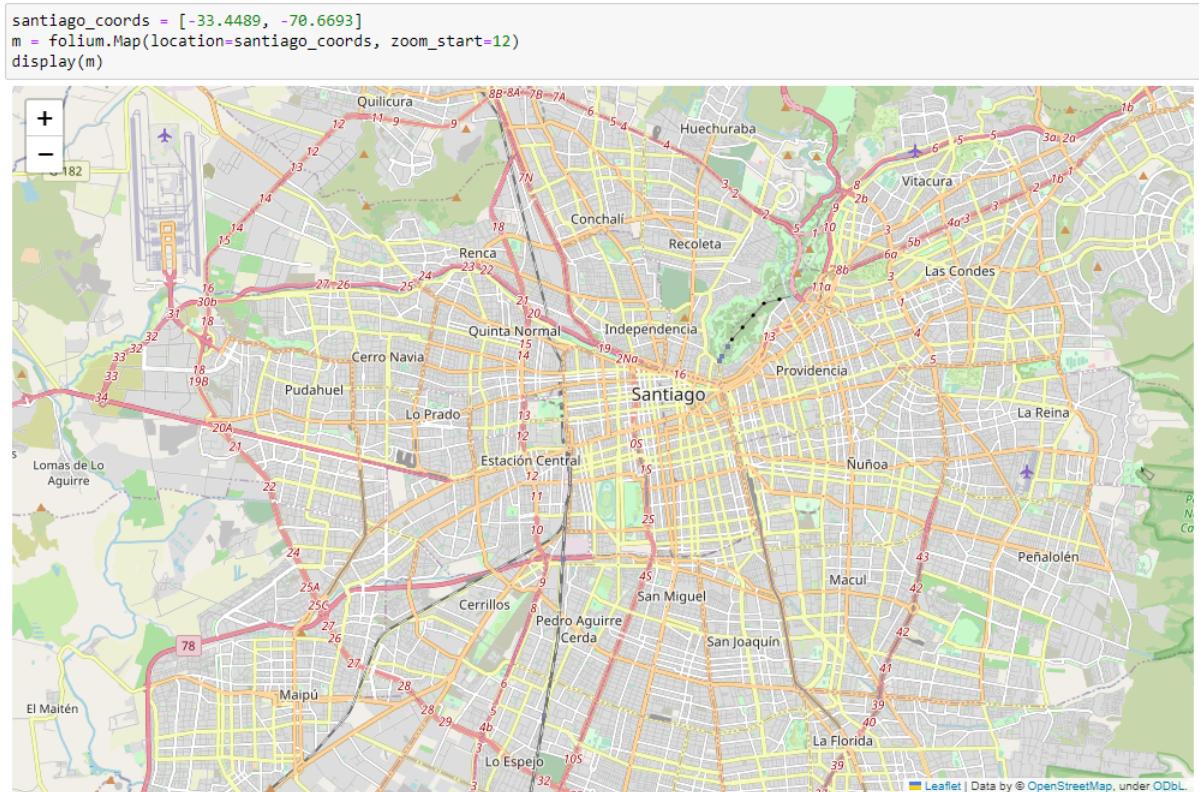


Figura 4.2: Mapa de Santiago, generado con folium.

El método **folium.Map** crea el mapa, centrado en una locación dada por una tupla de coordenadas (en este caso, Santiago), y con un nivel configurable de zoom en la imagen. Al ser interactivo, este nivel de zoom es modifiable utilizando los botones de la esquina superior izquierda o la rueda del mouse, mientras que se puede cambiar el foco al clickear y arrastrar.

Para destacar los puntos importantes, **folium** posee métodos como **folium.Marker**, que agrega un marcador a una ubicación específica. Este es configurable estéticamente con respecto al color e ícono, y además se le puede dar un mensaje para mostrar al clickear el marcador, como **popup**. En este caso, lo que se muestra para los puntos de inicio y destino corresponde a la información completa de la dirección buscada, proveniente de **Nominatim**. Por otro lado, para las paradas, se señalan los identificadores del recorrido y la parada en cuestión.

Usando estas herramientas en conjunto, se puede generar el mapa correctamente para entregarlo como salida del algoritmo. La función creada para gestionar la salida, tanto el texto como el mapa, se llama **create_transport_map**. El código completo se puede revisar en el anexo B, pero en el fragmento de código 4.12 se señalan las partes correspondientes a la generación del mapa:

```

1 def create_transport_map(route_stops, selected_path, source_date,
2     source_hour, margin):
3     m = folium.Map(location=[selected_path[0][0], selected_path[0][1]],
4         zoom_start=13)
5
6     folium.Marker(location=[selected_path[0][0], selected_path[0][1]],
7         popup="Origen: {}".format(source), icon=folium.Icon(color='green')).add_to(m)
8     folium.Marker(location=[selected_path[-1][0], selected_path[-1][1]],
9         popup="Destino: {}".format(target), icon=folium.Icon(color='red')).add_to(m)
10    ...
11    for stop_id in near_source_stops:
12        if stop_id in valid_source_stops:
13            ...
14            for service in valid_stop_services:
15                if service == best_option[0] and stop_id == best_option[1]:
16                    folium.Marker(location=[stop_coords[1], stop_coords[0]],
17                        popup="Mejor opcion: subirse al recorrido {} en el paradero {}".format(best_option[0], best_option[1]),
18                        icon=folium.Icon(color='cadetblue', icon='plus')).add_to(m)
19                    initial_distance = [(selected_path[0][0],
20                        selected_path[0][1]),(stop_coords[1], stop_coords[0])]
21                    folium.PolyLine(initial_distance,color='black',
22                        dash_array='10').add_to(m)
23                    ...
24                    for service in valid_target:
25                        if service == best_option[0]:
26                            ...
27                            folium.Marker(location=[selected_stop_coords[1],
28                                selected_stop_coords[0]],
29                                popup="Mejor opcion: bajarse del recorrido {} en el paradero {}".format(best_option[0], selected_stop),
30                                icon=folium.Icon(color='cadetblue', icon='plus')).add_to(m)
31                            ending_distance = [(selected_path[-1][0], selected_path[-1][1]),
32                                (selected_stop_coords[1], selected_stop_coords[0])]
33                            folium.PolyLine(ending_distance,color='black',
34                                dash_array='10').add_to(m)
35
36                            folium.PolyLine(visited_stops_sorted_coords, color='red').
37                                add_to(m)
38
39    return m

```

Código 4.12: Fragmento de **create_transport_map()** que se encarga de generar el mapa y sus elementos.

De esta forma, se procesa la información para generar el mapa, señalando los puntos relevantes de la ruta. A modo de ejemplo, continuando lo mostrando en la subsección anterior, en la figura 4.3 se muestra la salida del algoritmo, en la parte del mapa, al realizar la misma ruta entre la FCFM y Puente Cal y Canto:

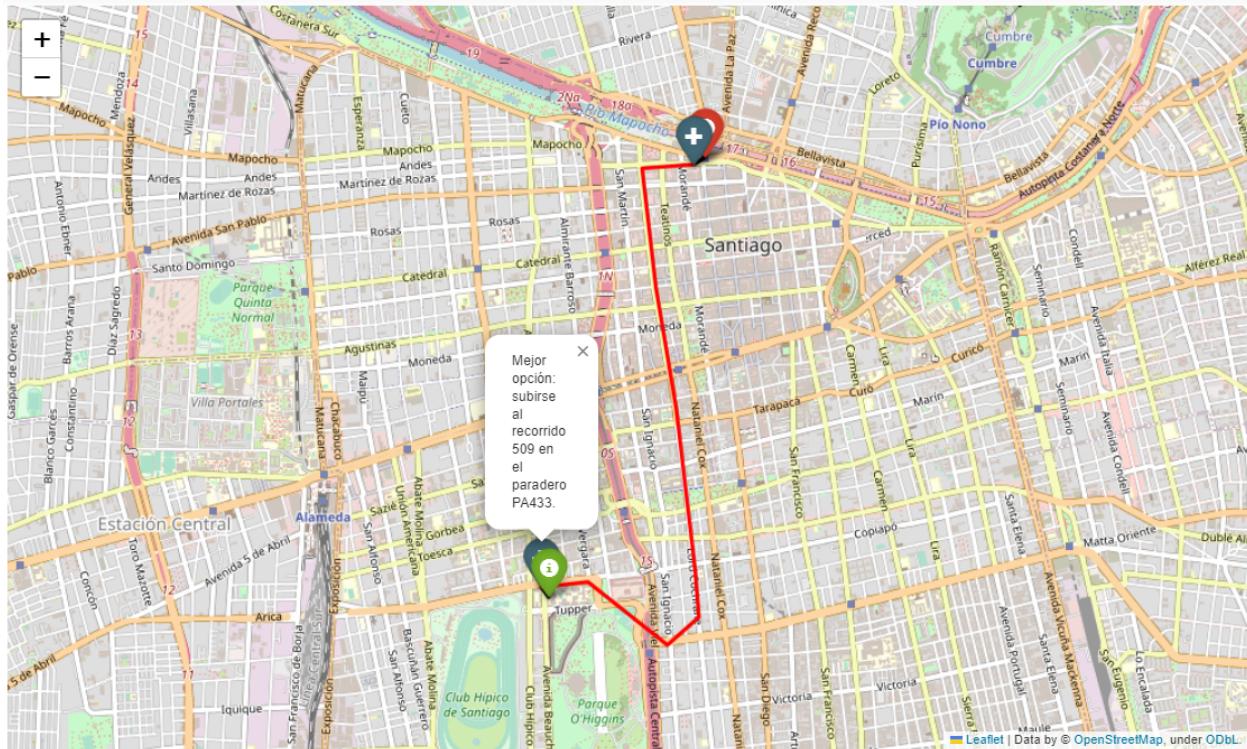


Figura 4.3: Ejemplo de salida del algoritmo (mapa). Ruta entre FCFM y Puente Cal y Canto.

Se pueden apreciar los diversos marcadores presentes en el mapa. El de color verde corresponde al punto de origen, el de color rojo corresponde al punto de destino, y los de color gris corresponden a las paradas de subida y bajada de la ruta, que está delimitada por una línea de color rojo. Además, la figura muestra el mensaje popup del marcador correspondiente a la parada de subida, que destaca que la mejor opción es subir al recorrido 509 en el paradero PA433, tal y como se indica en la salida de texto mostrada en la figura 4.1.

Capítulo 5

Resultados

En el presente capítulo, se describen los resultados obtenidos al terminar la implementación del algoritmo.

5.1. Ejemplos de uso del programa

Al ejecutar el programa, el algoritmo le solicita al usuario que ingrese los inputs para las 4 entradas del algoritmo; en orden, la fecha del viaje, la hora de inicio del viaje, la dirección de origen, y la dirección de destino. El programa permite ingresar valores por defecto, que consisten en la fecha y hora actuales del sistema, el Campus Beauchef de la Universidad de Chile como punto de origen, y el Campus Antumapu de la Universidad de Chile como punto de destino. La ventana de salida mostrará el estado actual del programa, imprimiendo mensajes cuando se encuentren las direcciones, cuando se esté calculando la mejor ruta, y el resultado final, incluyendo las paradas de origen, de destino, los recorridos a tomar, y el tiempo final del viaje. Además, se mostrará el mapa interactivo que graficará la ruta.

En la figura 5.1, se muestra como ejemplo la salida del algoritmo al buscar la mejor ruta para realizar un viaje desde una dirección en la comuna de Conchalí, hasta una dirección en la comuna de Santiago Centro. En específico, se busca llegar desde Avenida Diagonal Cardenal José María Caro #1553 hasta Avenida Libertador Bernardo O'Higgins #706. La fecha del viaje es 20/07/2023, y la hora de inicio es a las 01:00.

```

Enter the travel's date, in DD/MM/YYYY format (press Enter to use today's date) : 20/07/2023
20/07/2023
Enter the travel's start time, in HH:MM:SS format (press Enter to start now) : 01:00:00
01:00:00
Enter the starting point's address, in 'Street #No, Province' format (Ex: 'Beauchef 850, Santiago'):Av. Diagonal Cardenal José
María Caro 1553, Conchali
Enter the ending point's address, in 'Street #No, Province' format (Ex: 'Campus Antumapu Universidad de Chile, Santiago'):Aveni
da Libertador Bernardo O'Higgins 706, Santiago
Both addresses have been found.
Processing...

```

Routes have been found.
Calculating the best route and getting the arrival times for the next buses...

To go from: 1551, Avenida Diagonal Cardenal José María Caro, Conchali, Provincia de Santiago, Región Metropolitana de Santiago, 8710022, Chile
To: 706, Avenida Libertador Bernardo O'Higgins, Barrio París-Londres, Santiago, Provincia de Santiago, Región Metropolitana de Santiago, 8330069, Chile

The best option is to walk for 1 minutes and 7 seconds to stop PB968, and take the route B31N.

The next bus arrives at 01:28.

The other two next buses arrives in:

87 minutes, 52 seconds (02:27)

147 minutes, 52 seconds (03:27)

You will get off the bus on stop PA91 after 24 minutes and 38 seconds.

After that, you need to walk for 2 minutes and 11 seconds to arrive at the target spot.

Total travel time: 55 minutes, 49 seconds. You will arrive your destination at 01:55.

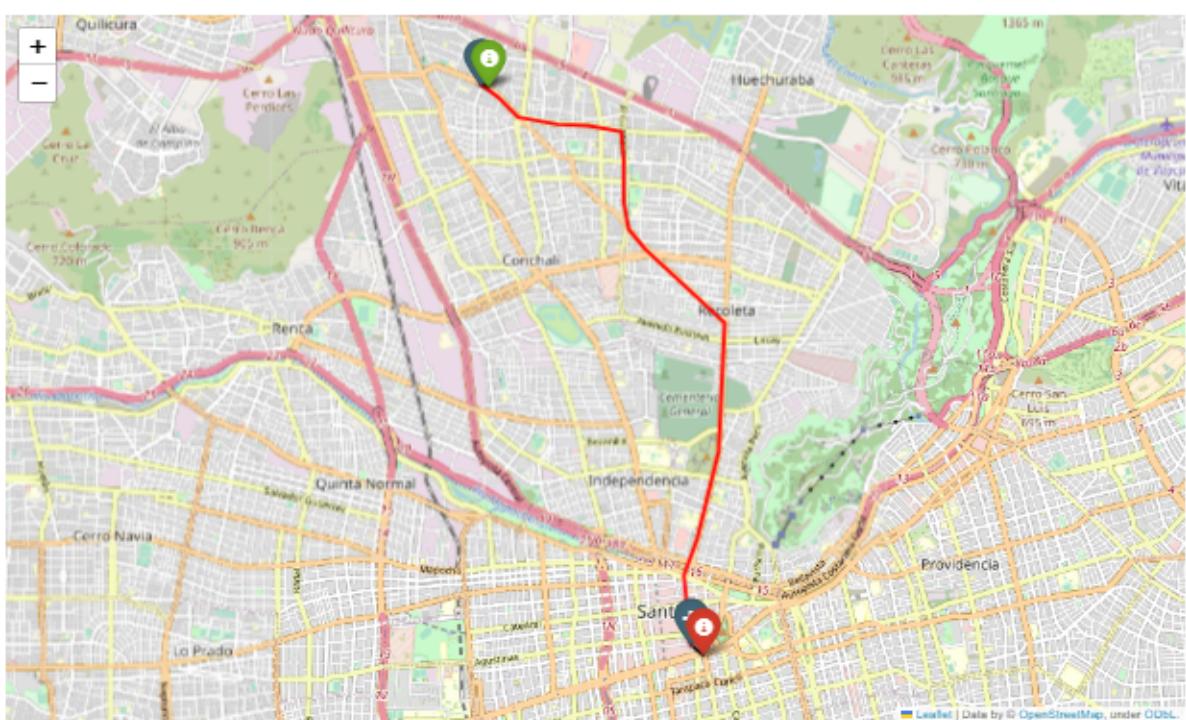


Figura 5.1: Resultado de la implementación del algoritmo: viaje entre Conchalí y Santiago un jueves de madrugada.

Como se puede apreciar arriba, en la parte superior de la figura se muestra la salida de texto del algoritmo. Se incluye, además, la sección inicial donde se procesan los inputs del usuario. Para este ejemplo, la mejor opción es caminar 2 minutos y 39 segundos para tomar el recorrido B31N en la parada PB968, cuyo siguiente bus llega a las 01:28 del día ingresado. Al bajar en la parada PA91, y luego de caminar por 2 minutos con 11 segundos adicionales, se llega al destino seleccionado a las 01:55. Posteriormente, se muestra el mapa interactivo que grafica el recorrido descrito.

Para ejemplificar qué pasa en aquellos casos en los que algo falla en el proceso de generar la ruta, se presentan las siguientes figuras. En la figura 5.2, se muestra el error que aparece cuando el programa es incapaz de hallar una de las direcciones entregadas por el usuario. En este caso, la dirección que no se encuentra es la dirección de destino: Avenida Siempreviva #742, Springfield (que corresponde a la dirección ficticia de la casa de *Los Simpson*).

```
Enter the travel's date, in DD/MM/YYYY format (press Enter to use today's date) :
22/07/2023
Enter the travel's start time, in HH:MM:SS format (press Enter to start now) :
15:35:24
Enter the starting point's address, in 'Street #No, Province' format (Ex: 'Beauchef 850, Santiago'):Beauchef 850, Santiago
Enter the ending point's address, in 'Street #No, Province' format (Ex: 'Campus Antumapu Universidad de Chile, Santiago'):Aveni
da Siempreviva 742, Springfield
Error: Address couldn't be found.

Something went wrong. Please try again later.
```

Figura 5.2: Resultado de la implementación del algoritmo: error al encontrar una dirección.

Por otro lado, en la figura 5.3 se muestra el error que aparece cuando sí existen rutas válidas para ir desde el origen al destino, pero estas rutas no tienen buses funcionando para la fecha u hora de inicio entregadas por el usuario. Una de las causas posibles es que sean rutas que sólo andan de día o de noche, como sugiere el mensaje de la figura.

```
Enter the travel's date, in DD/MM/YYYY format (press Enter to use today's date) : 20/07/2023
20/07/2023
Enter the travel's start time, in HH:MM:SS format (press Enter to start now) :
15:56:55
Enter the starting point's address, in 'Street #No, Province' format (Ex: 'Beauchef 850, Santiago'):Av. diagonal Cardenal José
María Caro 1553, conchali
Enter the ending point's address, in 'Street #No, Province' format (Ex: 'Campus Antumapu Universidad de Chile, Santiago'):aveni
da libertador bernardo o'higgins 706, santiago
Both addresses have been found.
Processing...

Routes have been found.
Calculating the best route and getting the arrival times for the next buses...
Error: There are no available services right now to go to the desired destination.
Possible reasons: the valid routes are not available at the specified date or starting time.
Please take into account that some routes have trips only during or after nighttime, which goes between 00:00:00 and 05:30:00
```

Figura 5.3: Resultado de la implementación del algoritmo: error al no existir rutas para la fecha u hora especificadas.

Finalmente, en la figura 5.4 se muestra el error que aparece cuando, por algún motivo, el programa determina que no existen rutas que permitan llegar desde el punto de origen hasta el punto de destino. En este caso, incluso habiendo hallado ambas direcciones, se concluye que es imposible realizar la ruta. Una causa posible es que el margen que se entrega para hallar las paradas cercanas sea muy pequeño, y como menciona el mensaje de la imagen, se recomienda probar con uno más grande.

```
Enter the travel's date, in DD/MM/YYYY format (press Enter to use today's date) :
22/07/2023
Enter the travel's start time, in HH:MM:SS format (press Enter to start now) :
15:44:18
Enter the starting point's address, in 'Street #No, Province' format (Ex: 'Beauchef 850, Santiago'):Beauchef 850, Santiago
Enter the ending point's address, in 'Street #No, Province' format (Ex: 'Campus Antumapu Universidad de Chile, Santiago'):Aveni
da Santa Rosa 1502, La Pintana
Both addresses have been found.
Processing...
Error: There are no available services right now to go to the desired destination.
Possible reasons: no routes that have stops near the source and target addresses.
You can try changing the search margin and try again.
```

Figura 5.4: Resultado de la implementación del algoritmo: error al encontrar rutas.

5.2. Caso de estudio

Para efectuar el caso de estudio definido, se define buscar rutas entre múltiples pares de puntos para los que, hoy en día, viajar usando la Línea 3 del Metro sería la mejor opción. La idea es comparar los resultados obtenidos en el caso regular donde la L3 forma parte de los recorridos permitidos, con un caso artificial donde se omite completamente. Esto forzará al algoritmo a buscar rutas en los buses Red, realizando un filtro en los vehículos disponibles. Finalmente, se comparan los tiempos finales de viaje para ambos casos, para inferir qué tanto hace la diferencia la existencia de esta línea de Metro.

Se definen cuatro escenarios diferentes, con respecto al contexto temporal:

- Iniciar el viaje a las 07:00 un lunes cualquiera.
- Iniciar el viaje a las 01:30 un sábado cualquiera.
- Iniciar el viaje a las 14:30 un domingo cualquiera.
- Iniciar el viaje a las 22:00 un día feriado.

La idea es estudiar la situación en distintos escenarios, para estudiar el viaje de forma más completa e inferir patrones de viajeros con diferentes necesidades.

Sección a completar...

5.3. Evaluación de resultados

El algoritmo es capaz de funcionar correctamente para generar rutas en Santiago. Se buscan correctamente las direcciones entregadas, y en caso de hallarlas, se toma en consideración el contexto temporal entregado por el usuario (fecha y hora) para entregar una ruta acorde a este. Además, cuando el algoritmo no es capaz de calcular una ruta por algún motivo, se muestra correctamente el mensaje correspondiente a la razón tras ello (dirección no encontrada, inexistencia de recorridos a la hora estipulada, etc.) En general, esto indica que se cumple correctamente el objetivo del proyecto.

Por otro lado, como fue mencionado en la sección 3.3 del capítulo de Diseño, el criterio específico que fue elegido para evaluar el cumplimiento de los objetivos del proyecto consiste en que el usuario sea capaz de *identificar patrones de movilidad* en Santiago. Esto implica que el usuario debe ser capaz de utilizar el algoritmo para obtener rutas, y a través de estas, poder inferir información sobre el cómo se movilizan las personas en la ciudad. Para este caso, como mostró el caso de estudio... (Sección a completar...)

Capítulo 6

Discusión

En este capítulo, habiendo ya presentado los resultados obtenidos, se procede a realizar una discusión sobre las implicancias y limitaciones del proyecto. Además, se discuten ideas para continuar con el desarrollo del proyecto a modo de Trabajo Futuro.

6.1. Implicancias

La creación de esta herramienta permitirá a potenciales usuarios realizar simulaciones de viajes en el sistema de transporte público disponible. Como se vio en el capítulo anterior, esta implementación del algoritmo logra su objetivo y es capaz de calcular la mejor ruta entre dos puntos de la ciudad, habiendo sido alimentado por la información cartográfica y del transporte de Santiago. Dado que el repositorio del proyecto estará disponible de forma pública para cualquiera que lo desee usar, esto permitirá que futuros potenciales usuarios puedan realizar estas simulaciones sin necesidad de utilizar una plataforma externa ni exponer su información a terceros. Además, como el algoritmo se halla expuesto en el código, el usuario es libre de modificarlo según sus necesidades.

Una gran implicancia derivada del último punto es que el algoritmo podría ser modificado para estudiar el transporte público en otras ciudades de Chile, o incluso, del extranjero, cambiando la ciudad al llamar a la función `get_network` de `pyrosm` (como se muestra en la sección 4.1.1) para obtener la información cartográfica , y utilizando un archivo en formato GTFS con los datos del transporte público de la misma al usar el módulo `Schedule` de `pygtfs` (como se muestra en la sección 4.1.2). La razón de esto es que la lógica tras en funcionamiento del código no está ligada directamente a los datos de Santiago, sino que necesita, simplemente, que la información desde OSM y desde GTFS sea provista correctamente y corresponda a la misma ciudad. Por ello, se abre un mundo de posibilidades para extender lo discutido en el desarrollo de este trabajo de título.

6.2. Limitaciones

Si bien la implementación del algoritmo logró sus objetivos al cumplir el criterio de evaluación estipulado, esto no le resta de tener ciertas limitaciones en su estado actual. Por ejemplo, una de sus limitantes principales es que la búsqueda de direcciones se realiza utilizando la geocodificación de Nominatim, librería dependiente de su propia API externa, por lo que el algoritmo será incapaz de calcular una ruta en caso de que el servicio de **Nominatim** se encuentre caído. La decisión de tener que depender de un servicio externo se realizó teniendo esta limitante en mente, puesto que se priorizó la comodidad del usuario final al permitirle entregar una dirección en vez de ingresar directamente las coordenadas, con el fin de mejorar la usabilidad del programa. Habiendo concluido el trabajo, se considera que fue la decisión correcta, mas no deja de ser una limitación.

Otra limitante de vital importancia recae en el hecho de que los datos ingresados para el transporte público son todos de naturaleza *estática*, es decir, información previamente ingresada que, en teoría, debiera ser representativa de la realidad, pero en la práctica, existe más de un inconveniente para que esto no ocurra tan así. El caso más directo de ver recae en los tiempos de espera de los buses, utilizados tanto para calcular el tiempo de viaje completo, como para decidir cuál es la mejor ruta (el fin último del algoritmo). Estos tiempos son aproximados, pues responden al cronograma mandatorio que los buses deben de seguir, pero dejan de lado variables que pueden existir al momento de realizar los viajes: los retrasos, que pueden ocurrir por múltiples razones, como fallas humanas, embotellamientos, malfuncionamiento del bus, entre otras. Al no poseer información en tiempo real, el algoritmo funciona bajo un supuesto de *continuidad operativa perfecta*, y al entregar una respuesta basada en este supuesto, puede presentar un sesgo importante.

Por último, dado que el algoritmo requiere de información de transporte público entregada en el formato GTFS, una limitación directa es que la responsabilidad de utilizar la última versión disponible recae en el usuario, puesto que él es quien debe preocuparse de ingresar manualmente la última versión disponible cada vez que aparece una nueva actualización. El uso de una versión desactualizada puede llevar a otro sesgo, pudiendo llevar a, por ejemplo, omitir un recorrido nuevo que llegue a ser la solución óptima para obtener la mejor ruta. Además, en el caso de extender el funcionamiento para otra ciudad, basta que el organismo encargado del transporte no entregue la información en formato GTFS para que la simulación no pueda realizarse.

6.3. Trabajo Futuro

Inspirado directamente por las limitaciones señaladas anteriormente, se pueden derivar múltiples aristas a desarrollar para un posible trabajo a futuro sobre esta implementación. Además, durante el desarrollo del proyecto, existieron varias ideas de desarrollo e implementación del algoritmo que quedaron fuera del enfoque del Trabajo de Título, pero que pueden motivar una línea de trabajo adicional para el futuro. Las ideas presentes son:

- Desarrollar una versión que obtenga constantemente la información del transporte público en tiempo real, proveniente de bases como los GPS integrados en los buses o datos provistos por los mismos usuarios. Esto ya existe en otras plataformas que poseen objetivos similares, por lo que en teoría debiera ser posible agregarlo a esta implementación. Esto permitiría que el algoritmo entregue una respuesta mucho más cercana a la realidad, y que pierda el sesgo de *continuidad operativa perfecta*, al menos de forma parcial.
- Añadir nuevas fuentes de datos cartográficos adicionales a OpenStreetMap, para adecuarse a otros medios de transporte fuera del transporte público. Por ejemplo, obtener información de las ciclovías disponibles para un usuario que desee movilizarse en bicicleta (pudiendo provenir también de proyectos comunitarios, como Bicineta Chile [1]). Esto puede aumentar las aristas estudiadas y desarrollar estudios de movilidad mucho más enriquecedores.
- Desarrollar una versión completamente *offline* de esta implementación, que no dependa de servicios externos ni conexión a Internet. Esto implica forzar al usuario a descargar previamente la información cartográfica de la ciudad (al igual que con *gtfs.zip*) y a que deba ingresar las coordenadas de los puntos deseados en vez de buscar por dirección, con el fin de que se pueda acceder a la información en todo momento. Esto aseguraría la disponibilidad operativa del algoritmo.
- Desarrollar un método de actualización de datos del transporte público, permitiendo automatizar la obtención de la información. Esto permitirá al usuario librarse de una responsabilidad para el funcionamiento del programa, mejorando la usabilidad del mismo. Esta idea puede extenderse a desarrollar una versión que permita entregar la información del transporte público en un formato distinto a GTFS, y/o que permita su traducción para usarlo como tal.
- Realizar una revisión completa a la implementación para mejorarla u optimizarla. Si bien el proyecto fue realizado con el cuidado de generar una implementación lo más limpia y eficiente posible, siempre puede existir lugar a mejoras. Además, dado que el funcionamiento del algoritmo está garantizado con el stack tecnológico discutido, en sus versiones disponibles a julio de 2023, a futuro será necesario actualizar las dependencias y librerías a las últimas versiones y realizar un chequeo general para verificar que todo siga funcionando correctamente.

Se espera que este proyecto pueda continuar su desarrollo en el futuro, ya que es un buen aporte como herramienta para el estudio de movilidad urbana, y posee un gran potencial explotable.

Capítulo 7

Conclusión

El trabajo de título tuvo por objetivo generar una implementación de Connection Scan Algorithm en Python, con la finalidad de ser utilizado para obtener rutas entre dos puntos de Santiago, siguiendo una secuencia de transbordos de servicios del transporte público. Su motivación fue el problema de estudiar la movilidad vial en una ciudad, algo bastante más crítico de lo que parece a primera vista. Si bien existen múltiples herramientas para cumplir este objetivo, una de sus carencias comunes es la rápida obsolescencia de la información entregada, dado que al momento de terminar de procesar y analizar los datos obtenidos, suele ocurrir que se hayan generado cambios en el entorno que generen un sesgo negativo, como un cambio en la ruta de un servicio de autobús o la apertura de nuevas estaciones de Metro.

El resultado obtenido fue una implementación adecuada de CSA, cumpliendo así el objetivo principal del proyecto. La ruta obtenida como salida del algoritmo optimiza la cantidad de trasbordos a realizar y los tiempos de viaje, entregando así la mejor secuencia disponible. Dado que la implementación considera la información cartográfica actual de Santiago (proveniente de OpenStreetMap), así como las especificaciones vigentes del transporte público (en formato GTFS y provistas por el Directorio de Transporte Público Metropolitano), se necesita estar constantemente actualizando la alimentación del algoritmo para que entregue una ruta válida y actualizada; sin embargo, esta acción se puede hacer de forma sencilla, y permite la continuidad operativa de esta implementación. Aun así, dado que el algoritmo se alimenta de información estática (porque no recibe datos en vivo), al momento de ocurrir una eventualidad o emergencia, la implementación podría quedar igualmente sesgada bajo un supuesto de normalidad permanente en el estado del servicio. Esta línea de pensamiento puede ser utilizada para motivar trabajo futuro en esta área, mejorando la implementación para obtener información en directo del transporte público disponible.

En definitiva, esta implementación de Connection Scan Algorithm en Python representa un paso significativo hacia la mejora de la planificación de rutas de transporte público en Santiago. La información generada por el algoritmo puede aportar datos concretos a las organizaciones encargadas de administrar los servicios de transporte, abriendo nuevas posibilidades para el estudio y análisis de la movilidad urbana. Finalmente, la capacidad de adaptación y actualización sencilla del algoritmo sienta las bases para futuras investigaciones y desarrollos en busca de una movilidad más eficiente y resiliente en la ciudad.

Bibliografía

- [1] Bicineta Chile. Mapa de Ciclovías de la Región Metropolitana. Disponible en <https://www.bicineta.cl/ciclovias>. Revisado el 2023/03/07.
- [2] Fundación OpenStreetMap Chile. Mapa de OpenStreetMap Chile. Información disponible en <https://www.openstreetmap.cl>. Revisado el 2023/03/07.
- [3] GTFS Community. General Transit Feed Specification. Disponible en <https://gtfs.org>. Revisado el 2023/06/28.
- [4] Yaron de Leeuw. pygtfs. Repositorio disponible en <https://github.com/jarondl/pygtfs>. Revisado el 2023/06/29.
- [5] Tiago de Paula Peixoto. graph-tool: Efficient network analysis with python. Documentación disponible en <https://graph-tool.skewed.de>. Revisado el 2023/07/17.
- [6] Directorio de Transporte Público Metropolitano. GTFS Vigente. Disponible en <https://www.dtpm.cl/index.php/gtfs-vigente>. Revisado el 2023/07/22. Última versión: 2023/07/08.
- [7] NetworkX developers. Networkx - network analysis in python. Documentación disponible en <https://networkx.org>. Revisado el 2023/07/22. Última versión: 2023/04/04.
- [8] Julian Dibbelt, Thomas Pajor, Ben Strasser, and Dorothea Wagner. Connection Scan Algorithm. *ACM Journal of Experimental Algorithms*, 23(1.7):1–56, 2018.
- [9] Filipe Fernandes. Folium. Repositorio disponible en <https://github.com/python-visualization/folium>. Revisado el 2023/07/17.
- [10] OpenStreetMap (Global). Mapa de OpenStreetMap. Información disponible en <https://www.openstreetmap.org>. Revisado el 2023/03/07.
- [11] Google. Google Maps. Disponible en <https://www.google.com/maps/>.
- [12] Eduardo Graells-Garrido. Aves: Análisis y Visualización, Educación y Soporte. Repositorio disponible en <https://github.com/zorزالerrante/aves>. Revisado el 2023/03/07.
- [13] Sarah Hoffmann. Nominatim. Disponible en <https://nominatim.org>. Revisado el 2023/07/17.
- [14] Universidad Alberto Hurtado. Actualización y recolección de información del sistema de transporte urbano, IX Etapa: Encuesta Origen Destino Santiago 2012. Encuesta origen destino de viajes 2012. Disponible en <http://www.sectra.gob.cl/biblioteca/detalle1.asp?mfn=3253> (2012). Revisado el 2023/03/07. Última versión lanzada el 2014.

- [15] Kelsey Jordahl. Geopandas. Documentación disponible en <https://geopandas.org>. Revisado el 2023/03/07. Última versión: 2022/12/10.
- [16] Felipe Leal. CC6909-Ayatori (Repositorio del Trabajo de Título). Repositorio disponible en <https://github.com/Lysorek/CC6909-Ayatori>. Revisado el 2023/03/07.
- [17] Microsoft. Windows subsystem for linux. Documentación disponible en <https://learn.microsoft.com/en-us/windows/ws1/>. Revisado el 2023/07/17.
- [18] Linus Norton. Connection Scan Algorithm (implementación en TypeScript). Repositorio disponible en <https://github.com/planarnetwork/connection-scan-algorithm>. Revisado el 2023/06/29.
- [19] Data Reportal. Digital 2021 Report for Chile. Disponible en <https://datareportal.com/reports/digital-2021-chile> (2021/02/11). Revisado el 2023/03/07.
- [20] Audrey Roy and Cookiecutter community. Cookiecutter: Better project templates. Documentación disponible en <https://cookiecutter.readthedocs.io/en/stable/>. Revisado el 2023/03/07.
- [21] Jonas Sauer. ULTRA: UnLimited TRAnsfers for Multimodal Route Planning. Repositorio disponible en <https://github.com/kit-algo/ULTRA>. Revisado el 2023/03/07.
- [22] Henrikki Tenkanen. Pyrosm: Read OpenStreetMap data from Protobuf files into GeoDataFrame with Python, faster. Repositorio disponible en <https://github.com/HTenkanen/pyrosm>. Revisado el 2023/03/07.
- [23] Jochen Topf and Frederik Ramm. Geofabrik. Disponible en <https://www.geofabrik.de>. Revisado el 2023/03/07.
- [24] Jochen Topf and Frederik Ramm. Geofabrik download server - chile. Disponible en <https://download.geofabrik.de/south-america/chile.html>. Revisado el 2023/03/07. Última versión: 2023/03/06.
- [25] Papers with Code. Connection Scan Algorithm implementations. Disponible en <https://cs.paperswithcode.com/paper/connection-scan-algorithm>. Revisado el 2023/03/07.

ANEXOS

Apéndice A

Implementaciones existentes

A.1. Aalto University

```
1 class ConnectionScan(AbstractRoutingAlgorithm):
2     """
3         A simple implementation of the Connection Scan Algorithm (CSA) solving
4             the first arrival problem
5             for public transport networks.
6
7             http://i11www.iti.uni-karlsruhe.de/extra/publications/dpsw-isftr-13.
8             pdf
9             """
10
11
12     def __init__(self, transit_events, seed_stop, start_time,
13                  end_time, transfer_margin, walk_network, walk_speed):
14         """
15             Parameters
16             -----
17             transit_events: list[Connection]
18             seed_stop: int
19                 index of the seed node
20             start_time : int
21                 start time in unixtime seconds
22             end_time: int
23                 end time in unixtime seconds (no new connections will be
24             scanned after this time)
25             transfer_margin: int
26                 required extra margin required for transfers in seconds
27             walk_speed: float
28                 walking speed between stops in meters / second
29             walk_network: networkx.Graph
```

```

26     each edge should have the walking distance as a data attribute
27     ("d_walk") expressed in meters
28     """
29
30     AbstractRoutingAlgorithm.__init__(self)
31     self._seed = seed_stop
32     self._connections = transit_events
33     self._start_time = start_time
34     self._end_time = end_time
35     self._transfer_margin = transfer_margin
36     self._walk_network = walk_network
37     self._walk_speed = walk_speed
38
39     # algorithm internals
40     self.__stop_labels = defaultdict(lambda: float('inf'))
41     self.__stop_labels[seed_stop] = start_time
42
43     # trip flags:
44     self.__trip_reachable = defaultdict(lambda: False)
45
46     def get_arrival_times(self):
47         """
48             Returns
49             -----
50             arrival_times: dict[int, float]
51                 maps integer stop_ids to floats
52
53         assert self._has_run
54         return self.__stop_labels
55
56     def _run(self):
57         self._scan_footpaths(self._seed, self._start_time)
58         for connection in self._connections:
59             departure_time = connection.departure_time
60             if departure_time > self._end_time:
61                 return
62             from_stop = connection.departure_stop
63             to_stop = connection.arrival_stop
64             arrival_time = connection.arrival_time
65             trip_id = connection.trip_id
66             reachable = False
67             if self.__trip_reachable[trip_id]:
68                 reachable = True
69             else:
70                 dep_stop_reached = self.__stop_labels[from_stop]
71                 if dep_stop_reached + self._transfer_margin <=
72                     departure_time:
73                     self.__trip_reachable[trip_id] = True
74                     reachable = True
75                     if reachable:
76                         self._update_stop_label(to_stop, arrival_time)
77                         self._scan_footpaths(to_stop, arrival_time)
78
79     def _update_stop_label(self, stop, arrival_time):
80         current_stop_label = self.__stop_labels[stop]
81         if current_stop_label > arrival_time:
82             self.__stop_labels[stop] = arrival_time

```

```

80
81     def _scan_footpaths(self, stop_id, walk_departure_time):
82         """
83             Scan the footpaths originating from stop_id
84
85         Parameters
86         -----
87         stop_id: int
88         """
89         for _, neighbor, data in self._walk_network.edges(nbunch=[stop_id],
90 ], data=True):
91             d_walk = data["d_walk"]
92             arrival_time = walk_departure_time + d_walk / self._walk_speed
93             self._update_stop_label(neighbor, arrival_time)
94
95 class ConnectionScanProfiler(AbstractRoutingAlgorithm):
96     """
97         Implementation of the profile connection scan algorithm presented in
98
99         http://i11www.iti.uni-karlsruhe.de/extr/publications/dpsw-isftr-13.
100        pdf
101        """
102
103     def __init__(self,
104                 transit_events,
105                 target_stop,
106                 start_time=None,
107                 end_time=None,
108                 transfer_margin=0,
109                 walk_network=None,
110                 walk_speed=1.5,
111                 verbose=False):
112         """
113         Parameters
114         -----
115         transit_events: list[Connection]
116             events are assumed to be ordered in DECREASING departure_time
117             (!)
118         target_stop: int
119             index of the target stop
120         start_time : int, optional
121             start time in unixtime seconds
122         end_time: int, optional
123             end time in unixtime seconds (no connections will be scanned
124             after this time)
125         transfer_margin: int, optional
126             required extra margin required for transfers in seconds
127         walk_speed: float, optional
128             walking speed between stops in meters / second.
129         walk_network: networkx.Graph, optional
130             each edge should have the walking distance as a data attribute
("distance_shape") expressed in meters
131         verbose: boolean, optional
132             whether to print out progress
133         """

```

```

131     AbstractRoutingAlgorithm.__init__(self)
132
133     self._target = target_stop
134     self._connections = transit_events
135     if start_time is None:
136         start_time = transit_events[-1].departure_time
137     if end_time is None:
138         end_time = transit_events[0].departure_time
139     self._start_time = start_time
140     self._end_time = end_time
141     self._transfer_margin = transfer_margin
142     if walk_network is None:
143         walk_network = networkx.Graph()
144     self._walk_network = walk_network
145     self._walk_speed = float(walk_speed)
146     self._verbose = verbose
147
148     # algorithm internals
149
150     # trip flags:
151     self._trip_min_arrival_time = defaultdict(lambda: float("inf"))
152
153     # initialize stop_profiles
154     self._stop_profiles = defaultdict(NodeProfileSimple())
155     # initialize stop_profiles for target stop, and its neighbors
156     self._stop_profiles[self._target] = NodeProfileSimple(0)
157     if graph_has_node(walk_network, target_stop):
158         for target_neighbor in walk_network.neighbors(target_stop):
159             edge_data = walk_network.get_edge_data(target_neighbor,
160                                         target_stop)
161             walk_duration = edge_data["d_walk"] / self._walk_speed
162             self._stop_profiles[target_neighbor] = NodeProfileSimple(
163                                         walk_duration)
164
165     def _run(self):
166         # if source node in s1:
167         previous_departure_time = float("inf")
168         connections = self._connections # list[Connection]
169         n_connections = len(connections)
170         for i, connection in enumerate(connections):
171             # basic checking + printing progress:
172             if self._verbose and i % 1000 == 0:
173                 print(i, "/", n_connections)
174             assert (isinstance(connection, Connection))
175             assert (connection.departure_time <= previous_departure_time)
176             previous_departure_time = connection.departure_time
177
178             # get all different "accessible" / arrival times (Pareto-
179             # optimal sets)
180             arrival_profile = self._stop_profiles[connection.arrival_stop]
181             # NodeProfileSimple
182
183             # Three possibilities:
184
185             # 1. earliest arrival time (Profiles) via transfer
186             earliest_arrival_time_via_transfer = arrival_profile.

```

```

183     evaluate_earliest_arrival_time_at_target(
184         connection.arrival_time, self._transfer_margin
185     )
186
187     # 2. earliest arrival time within same trip (equals float('inf')
188     # if not reachable)
189     earliest_arrival_time_via_same_trip = self.
190     _trip_min_arrival_time[connection.trip_id]
191
192     # then, take the minimum (or the Pareto-optimal set) of these
193     # three alternatives.
194     min_arrival_time = min(earliest_arrival_time_via_same_trip,
195                           earliest_arrival_time_via_transfer)
196
197     # If there are no 'labels' to progress, nothing needs to be
198     # done.
199     if min_arrival_time == float("inf"):
200         continue
201
202     # Update information for the trip
203     if earliest_arrival_time_via_same_trip > min_arrival_time:
204         self._trip_min_arrival_time[connection.trip_id] =
205         earliest_arrival_time_via_transfer
206
207     # Compute the new "best" pareto_tuple possible (later: merge
208     # the sets of pareto-optimal labels)
209     pareto_tuple = LabelTimeSimple(connection.departure_time,
210                                     min_arrival_time)
211
212     # update departure stop profile (later: with the sets of
213     # pareto-optimal labels)
214     dep_stop_profile = self._stop_profiles[connection.
215     departure_stop]
216     updated_dep_stop = dep_stop_profile.
217     update_pareto_optimal_tuples(pareto_tuple)
218
219     # if the departure stop is updated, one also needs to scan the
220     # footpaths from the departure stop
221     if updated_dep_stop:
222         self._scan_footpaths_to_departure_stop(connection.
223         departure_stop,
224                                         connection.
225                                         departure_time,
226                                         min_arrival_time)
227
228     def _scan_footpaths_to_departure_stop(self, connection_dep_stop,
229                                         connection_dep_time, arrival_time_target):
230         """ A helper method for scanning the footpaths. Updates self.
231         _stop_profiles accordingly"""
232         for _, neighbor, data in self._walk_network.edges(nbunch=[
233             connection_dep_stop],
234                                         data=True):
235             d_walk = data['d_walk']
236             neighbor_dep_time = connection_dep_time - d_walk / self.
237             _walk_speed
238             pt = LabelTimeSimple(departure_time=neighbor_dep_time,
239                                 arrival_time_target=arrival_time_target)

```

```

220             self._stop_profiles[neighbor].update_pareto_optimal_tuples(pt)
221
222     @property
223     def stop_profiles(self):
224         """
225             Returns
226             -----
227             _stop_profiles : dict[int, NodeProfileSimple]
228                 The pareto tuples necessary.
229             """
230
231     assert self._has_run
232     return self._stop_profiles

```

A.1.1. ULTRA: headers

```

1  namespace CSA {
2
3  template<bool PATH_RETRIEVAL = true, typename PROFILER = NoProfiler>
4  class CSA {
5
6  public:
7      constexpr static bool PathRetrieval = PATH_RETRIEVAL;
8      using Profiler = PROFILER;
9      using Type = CSA<PathRetrieval, Profiler>;
10     using TripFlag = Meta::IF<PathRetrieval, ConnectionId, bool>;
11
12 private:
13     struct ParentLabel {
14         ParentLabel(const StopId parent = noStop, const bool
15         reachedByTransfer = false, const TripId tripId = noTripId) :
16             parent(parent),
17             reachedByTransfer(reachedByTransfer),
18             tripId(tripId) {
19         }
20
21         StopId parent;
22         bool reachedByTransfer;
23         union {
24             TripId tripId;
25             Edge transferId;
26         };
27     };
28
29 public:
30     CSA(const Data& data, const Profiler& profilerTemplate = Profiler() :
31         data(data),
32         sourceStop(noStop),
33         targetStop(noStop),
34         tripReached(data.numberOfTrips(), TripFlag()),
35         arrivalTime(data.numberOfStops(), never),
36         parentLabel(PathRetrieval ? data.numberOfStops() : 0),
37         profiler(profilerTemplate) {
38         AssertMsg(Vector::isSorted(data.connections), "Connections must be
39             sorted in ascending order!");

```

```

38     profiler.registerPhases({PHASE_CLEAR, PHASE_INITIALIZATION,
39     PHASE_CONNECTION_SCAN});
40     profiler.registerMetrics({METRIC_CONNECTIONS, METRIC_EDGES,
41     METRIC_STOPS_BY_TRIP, METRIC_STOPS_BY_TRANSFER});
42     profiler.initialize();
43 }
44
45 inline void run(const StopId source, const int departureTime, const
46 StopId target = noStop) noexcept {
47     profiler.start();
48
49     profiler.startPhase();
50     AssertMsg(data.isStop(source), "Source stop " << source << " is
51     not a valid stop!");
52     clear();
53     profiler.donePhase(PHASE_CLEAR);
54
55     profiler.startPhase();
56     sourceStop = source;
57     targetStop = target;
58     arrivalTime[sourceStop] = departureTime;
59     relaxEdges(sourceStop, departureTime);
60     const ConnectionId firstConnection = firstReachableConnection(
61     departureTime);
62     profiler.donePhase(PHASE_INITIALIZATION);
63
64     profiler.startPhase();
65     scanConnections(firstConnection, ConnectionId(data.connections.
66     size()));
67     profiler.donePhase(PHASE_CONNECTION_SCAN);
68
69     profiler.done();
70 }
71
72 inline bool reachable(const StopId stop) const noexcept {
73     return arrivalTime[stop] < never;
74 }
75
76 inline int getEarliestArrivalTime(const StopId stop) const noexcept {
77     return arrivalTime[stop];
78 }
79
80 template<bool T = PathRetrieval, typename = std::enable_if_t<T ==
81 PathRetrieval && T>>
82     inline Journey getJourney() const noexcept {
83         return getJourney(targetStop);
84     }
85
86 template<bool T = PathRetrieval, typename = std::enable_if_t<T ==
87 PathRetrieval && T>>
88     inline Journey getJourney(StopId stop) const noexcept {
89         Journey journey;
90         if (!reachable(stop)) return journey;
91         while (stop != sourceStop) {
92             const ParentLabel& label = parentLabel[stop];
93             if (label.reachedByTransfer) {

```

```

86         const int travelTime = data.transferGraph.get(TravelTime,
87 label.transferId);
88         journey.emplace_back(label.parent, stop, arrivalTime[stop]
89 - travelTime, arrivalTime[stop], label.transferId);
89     } else {
90         journey.emplace_back(label.parent, stop, data.connections[
91 tripReached[label.tripId]].departureTime, arrivalTime[stop], label.
92 tripId);
93     }
94     stop = label.parent;
95 }
96 Vector::reverse(journey);
97 return journey;
98 }

99 inline std::vector<Vertex> getPath(const StopId stop) const noexcept {
100     return journeyToPath(getJourney(stop));
101 }

102 inline std::vector<std::string> getRouteDescription(const StopId stop)
103 const noexcept {
104     return data.journeyToText(getJourney(stop));
105 }

106 inline const Profiler& getProfiler() const noexcept {
107     return profiler;
108 }

109 private:
110     inline void clear() {
111         sourceStop = noStop;
112         targetStop = noStop;
113         Vector::fill(arrivalTime, never);
114         Vector::fill(tripReached, TripFlag());
115         if constexpr (PathRetrieval) {
116             Vector::fill(parentLabel, ParentLabel());
117         }
118     }

119     inline ConnectionId firstReachableConnection(const int departureTime)
120 const noexcept {
121     return ConnectionId(Vector::lowerBound(data.connections,
122 departureTime, [] (const Connection& connection, const int time) {
123         return connection.departureTime < time;
124     }));
125 }

126     inline void scanConnections(const ConnectionId begin, const
127 ConnectionId end) noexcept {
128         for (ConnectionId i = begin; i < end; i++) {
129             const Connection& connection = data.connections[i];
130             if (targetStop != noStop && connection.departureTime >
131 arrivalTime[targetStop]) break;
132             if (connectionIsReachable(connection, i)) {
133                 profiler.countMetric(METRIC_CONNECTIONS);
134                 arrivalByTrip(connection.arrivalStopId, connection.

```

```

        arrivalTime, connection.tripId);
    }
}
}

137 inline bool connectionIsReachableFromStop(const Connection& connection)
138 ) const noexcept {
139     return arrivalTime[connection.departureStopId] <= connection.
140 departureTime - data.minTransferTime(connection.departureStopId);
141 }

142 inline bool connectionIsReachableFromTrip(const Connection& connection)
143 ) const noexcept {
144     return tripReached[connection.tripId] != TripFlag();
145 }

146 inline bool connectionIsReachable(const Connection& connection, const
147 ConnectionId id) noexcept {
148     if (connectionIsReachableFromTrip(connection)) return true;
149     if (connectionIsReachableFromStop(connection)) {
150         if constexpr (PathRetrieval) {
151             tripReached[connection.tripId] = id;
152         } else {
153             suppressUnusedParameterWarning(id);
154             tripReached[connection.tripId] = true;
155         }
156         return true;
157     }
158     return false;
159 }

160 inline void arrivalByTrip(const StopId stop, const int time, const
161 TripId trip) noexcept {
162     if (arrivalTime[stop] <= time) return;
163     profiler.countMetric(METRIC_STOPS_BY_TRIP);
164     arrivalTime[stop] = time;
165     if constexpr (PathRetrieval) {
166         parentLabel[stop].parent = data.connections[tripReached[trip]
167 ].departureStopId;
168         parentLabel[stop].reachedByTransfer = false;
169         parentLabel[stop].tripId = trip;
170     }
171     relaxEdges(stop, time);
172 }

173 inline void relaxEdges(const StopId stop, const int time) noexcept {
174     for (const Edge edge : data.transferGraph.edgesFrom(stop)) {
175         profiler.countMetric(METRIC_EDGES);
176         const StopId toStop = StopId(data.transferGraph.get(ToVertex,
177 edge));
178         const int newArrivalTime = time + data.transferGraph.get(
179 TravelTime, edge);
180         arrivalByTransfer(toStop, newArrivalTime, stop, edge);
181     }
182 }

```

```

180     inline void arrivalByTransfer(const StopId stop, const int time, const
181         StopId parent, const Edge edge) noexcept {
182         if (arrivalTime[stop] <= time) return;
183         profiler.countMetric(METRIC_STOPS_BY_TRANSFER);
184         arrivalTime[stop] = time;
185         if constexpr (PathRetrieval) {
186             parentLabel[stop].parent = parent;
187             parentLabel[stop].reachedByTransfer = true;
188             parentLabel[stop].transferId = edge;
189         }
190     }
191
192     private:
193         const Data& data;
194
195         StopId sourceStop;
196         StopId targetStop;
197
198         std::vector<TripFlag> tripReached;
199         std::vector<int> arrivalTime;
200         std::vector<ParentLabel> parentLabel;
201
202         Profiler profiler;
203     };

```

Apéndice B

Código de la solución

B.1. Obtención de la información

B.1.1. OSM

```

1 def get_osm_data():
2     """
3     Obtains the required OpenStreetMap data using the 'pyrosm' library.
4     This gives the map info of Santiago.
5
6     Returns:
7         graph: osm data converted to a graph
8     """

```

```

8 # Download latest OSM data
9 fp = get_data(
10     "Santiago",
11     update=True,
12     directory=OSM_PATH
13 )
14
15 osm = OSM(fp)
16
17 nodes, edges = osm.get_network(nodes=True)
18
19 graph = Graph()
20
21 # Create vertex properties for lon and lat
22 lon_prop = graph.new_vertex_property("float")
23 lat_prop = graph.new_vertex_property("float")
24
25 # Create properties for the ids
26 # Every OSM node has its unique id, different from the one given in
27 # the graph
28 node_id_prop = graph.new_vertex_property("long")
29 graph_id_prop = graph.new_vertex_property("long")
30
31 # Create edge properties
32 u_prop = graph.new_edge_property("long")
33 v_prop = graph.new_edge_property("long")
34 length_prop = graph.new_edge_property("double")
35 weight_prop = graph.new_edge_property("double")
36
37 vertex_map = {}
38
39 print("GETTING OSM NODES...")
40 for index, row in nodes.iterrows():
41     lon = row['lon']
42     lat = row['lat']
43     node_id = row['id']
44     graph_id = index
45     node_coords[node_id] = (lat, lon)
46
47     vertex = graph.add_vertex()
48     vertex_map[node_id] = vertex
49
50     # Assigning node properties
51     lon_prop[vertex] = lon
52     lat_prop[vertex] = lat
53     node_id_prop[vertex] = node_id
54     graph_id_prop[vertex] = graph_id
55
56     # Assign the properties to the graph
57     graph.vertex_properties["lon"] = lon_prop
58     graph.vertex_properties["lat"] = lat_prop
59     graph.vertex_properties["node_id"] = node_id_prop
60     graph.vertex_properties["graph_id"] = graph_id_prop
61
62     print("DONE")
63     print("GETTING OSM EDGES...")

```

```

63
64     for index, row in edges.iterrows():
65         source_node = row['u']
66         target_node = row['v']
67
68         if row["length"] < 2 or source_node == "" or target_node == "":
69             continue # Skip edges with empty or missing nodes
70
71         if source_node not in vertex_map or target_node not in vertex_map:
72             print(f"Skipping edge with missing nodes: {source_node} -> {target_node}")
73             continue # Skip edges with missing nodes
74
75         source_vertex = vertex_map[source_node]
76         target_vertex = vertex_map[target_node]
77
78         if not graph.vertex(source_vertex) or not graph.vertex(
79             target_vertex):
80             print(f"Skipping edge with non-existent vertices: {source_vertex} -> {target_vertex}")
81             continue # Skip edges with non-existent vertices
82
83         # Calculate the distance between the nodes and use it as the
84         # weight of the edge
85         source_coords = node_coords[source_node]
86         target_coords = node_coords[target_node]
87         distance = abs(source_coords[0] - target_coords[0]) + abs(
88             source_coords[1] - target_coords[1])
89
90         e = graph.add_edge(source_vertex, target_vertex)
91         u_prop[e] = source_node
92         v_prop[e] = target_node
93         length_prop[e] = row["length"]
94         weight_prop[e] = distance
95
96         graph.edge_properties["u"] = u_prop
97         graph.edge_properties["v"] = v_prop
98         graph.edge_properties["length"] = length_prop
99         graph.edge_properties["weight"] = weight_prop
100
101    print("OSM DATA HAS BEEN SUCCESSFULLY RECEIVED")
102    return graph
103
104 # OSM Graph
105 osm_graph = get_osm_data()
106
107 def make_undirected(graph):
108     """
109     Given a directed graph, returns an undirected version of the graph.
110
111     Parameters:
112     graph (Graph): A directed graph. In this specific case, the osm graph.
113
114     Returns:
115     Graph: An undirected version of the graph.
116     """

```

```

114     undirected_graph = Graph(directed=False)
115     vprop_map = graph.new_vertex_property("object")
116
117     # Create vertex properties for lon and lat
118     lon_prop = undirected_graph.new_vertex_property("float")
119     lat_prop = undirected_graph.new_vertex_property("float")
120     node_id_prop = undirected_graph.new_vertex_property("long")
121     graph_id_prop = undirected_graph.new_vertex_property("long")
122
123     # Create edge properties
124     u_prop = undirected_graph.new_edge_property("long")
125     v_prop = undirected_graph.new_edge_property("long")
126     length_prop = undirected_graph.new_edge_property("double")
127     weight_prop = undirected_graph.new_edge_property("double")
128
129     undirected_vertex_map = {}
130
131     for v in graph.vertices():
132         new_v = undirected_graph.add_vertex()
133         vprop_map[new_v] = v
134         lon = graph.vertex_properties["lon"][v]
135         lat = graph.vertex_properties["lat"][v]
136         node_id = graph.vertex_properties["node_id"][v]
137         graph_id = graph.vertex_properties["graph_id"][v]
138
139         undirected_vertex_map[node_id] = new_v
140         #print("NODO {} EN GRAFO {}".format(node_id, graph_id))
141
142         # Assigning node properties
143         lon_prop[new_v] = lon
144         lat_prop[new_v] = lat
145         node_id_prop[new_v] = node_id
146         graph_id_prop[new_v] = graph_id
147
148     # Assign the properties to the graph
149     undirected_graph.vertex_properties["lon"] = lon_prop
150     undirected_graph.vertex_properties["lat"] = lat_prop
151     undirected_graph.vertex_properties["node_id"] = node_id_prop
152     undirected_graph.vertex_properties["graph_id"] = graph_id_prop
153
154
155     for e in graph.edges():
156         source, target = e.source(), e.target()
157         source_node = graph.edge_properties["u"][e]
158         target_node = graph.edge_properties["v"][e]
159         lgt = graph.edge_properties["length"][e]
160         wt = graph.edge_properties["weight"][e]
161
162         if lgt < 2 or source_node == "" or target_node == "":
163             continue # Skip edges with empty or missing nodes
164
165         if source_node not in undirected_vertex_map or target_node not in
undirected_vertex_map:
166             print(f"Skipping edge with missing nodes: {source_node} -> {
target_node}")
167             continue # Skip edges with missing nodes

```

```

168     source_vertex = undirected_vertex_map[source_node]
169     target_vertex = undirected_vertex_map[target_node]
170
171     if not undirected_graph.vertex(source_vertex) or not
172         undirected_graph.vertex(target_vertex):
173         print(f"Skipping edge with non-existent vertices: {
174             source_vertex} -> {target_vertex}")
175         continue # Skip edges with non-existent vertices
176
177     e = undirected_graph.add_edge(source_vertex, target_vertex)
178     u_prop[e] = source_node
179     v_prop[e] = target_node
180     length_prop[e] = lgt
181     weight_prop[e] = wt
182
183     undirected_graph.edge_properties["u"] = u_prop
184     undirected_graph.edge_properties["v"] = v_prop
185     undirected_graph.edge_properties["length"] = length_prop
186     undirected_graph.edge_properties["weight"] = weight_prop
187
188     return undirected_graph
189
190 # Convertir el grafo en no dirigido
191 undirected_graph = make_undirected(osm_graph)

```

B.1.2. GTFS

```

1 def get_gtfs_data():
2 """
3     Reads the GTFS data from a file and creates a directed graph with its
4     info, using the 'pygtfs' library. This gives
5     the transit feed data of Santiago's public transport, including "Red
6     Metropolitana de Movilidad" (previously known
7     as Transantiago), "Metro de Santiago", "EFE Trenes de Chile", and "
8     Buses de Acercamiento Aeropuerto".
9
10    Returns:
11        graphs: GTFS data converted to a dictionary of graphs, one per
12        route.
13        route_stops: Dictionary containing the stops for each route.
14        special_dates: List of special calendar dates.
15    """
16
17    # Create a new schedule object using a GTFS file
18    sched = pygtfs.Schedule(":memory:")
19    pygtfs.append_feed(sched, "gtfs.zip")
20
21    # Get special calendar dates
22    special_dates = []
23    for cal_date in sched.service_exceptions: # Calendar_dates is renamed
24        in pygtfs
25            special_dates.append(cal_date.date.strftime("%d/%m/%Y"))
26
27    # Create a graph per route

```

```

22     graphs = {}
23     stop_id_map = {} # To assign unique ids to every stop
24     stop_coords = {}
25     route_stops = {}
26     for route in sched.routes:
27         graph = Graph(directed=True)
28         stop_ids = set()
29         trips = [trip for trip in sched.trips if trip.route_id == route.
route_id]
30
31         weight_prop = graph.new_edge_property("int") # Propiedad para
almacenar los pesos de las aristas
32
33         for trip in trips:
34             stop_times = trip.stop_times
35
36             # Get the orientation of the trip
37             orientation = trip.trip_id.split("-")[1]
38
39             for i in range(len(stop_times)):
40                 stop_id = stop_times[i].stop_id
41                 sequence = stop_times[i].stop_sequence
42
43                 if stop_id not in stop_id_map:
44                     vertex = graph.add_vertex() # Add empty vertex
45                     stop_id_map[stop_id] = vertex
46                 else:
47                     vertex = stop_id_map[stop_id] # Obtain existing
vertex
48
49                     stop_ids.add(vertex)
50
51                     if i < len(stop_times) - 1:
52                         next_stop_id = stop_times[i + 1].stop_id
53
54                         if next_stop_id not in stop_id_map:
55                             next_vertex = graph.add_vertex() # Add an empty
vertex for the next stop
56                             stop_id_map[next_stop_id] = next_vertex # Assign
the vertex
57                         else:
58                             next_vertex = stop_id_map[next_stop_id] # Obtain
the vertex
59
60                         e = graph.add_edge(vertex, next_vertex) # Add an edge
between the vertexes
61                         weight_prop[e] = 1
62
63                         # Store the coordinates of each stop for this route
64                         if route.route_id not in stop_coords:
65                             stop_coords[route.route_id] = {}
66                         if stop_id not in stop_coords[route.route_id]:
67                             stop = sched.stops_by_id(stop_id)[0]
68                             stop_coords[route.route_id][stop_id] = (stop.
stop_lon, stop.stop_lat)
69

```

```

70             # Store the sequence of each stop for this route
71             if route.route_id not in route_stops:
72                 route_stops[route.route_id] = {}
73                 route_stops[route.route_id][stop_id] = {
74                     "route_id": route.route_id,
75                     "stop_id": stop_id,
76                     "coordinates": stop_coords[route.route_id][
77                         stop_id],
78                     "orientation": "round" if orientation == "I"
79                     "else "return",
80                     "sequence": sequence,
81                     "arrival_times": []
82                 }
83
84             # Get the arrival time for the current stop
85             arrival_time = (datetime.min + stop_times[i].arrival_time)
86             .time()
87
88             # Check if the stop ID is already in the dictionary
89             if stop_id in route_stops[route.route_id]:
90                 # If the stop ID is already in the dictionary, append
91                 # the arrival time
92                 route_stops[route.route_id][stop_id]["arrival_times"].append(arrival_time)
93
94             # Else:
95             # If the stop ID is not in the dictionary, create a
96             # new entry with the arrival time
97             route_stops[route.route_id][stop_id] = {
98                 "# "route_id": route.route_id,
99                 "# "stop_id": stop_id,
100                "# "coordinates": stop_coords[route.route_id][
101                    stop_id],
102                "# "visited_on_round_trip": True if orientation == "I"
103                "# "else False,
104                "# "visited_on_return_trip": True if orientation == "R"
105                "# "else False,
106                "# "sequence": sequence,
107                "# "arrival_times": [arrival_time]
108                "#     }
109
110             graphs[route.route_id] = graph
111             # Group the stops by direction to get the stops visited on the
112             # round trip and the return trip
113             stops_by_direction = {"round_trip": [], "return_trip": []}
114             for trip in trips:
115                 stop_times = trip.stop_times
116                 stops = [stop_times[i].stop_id for i in range(len(stop_times))
117             ]
118
119             # Determine the direction of the trip
120             if trip.direction_id == 0:
121                 stops_by_direction["round_trip"].extend(stops)
122             else:
123                 stops_by_direction["return_trip"].extend(stops)

```

```

115     # Get the unique stops visited on the round trip and the return
116     trip
117     round_trip_stops = set(stops_by_direction["round_trip"])
118     return_trip_stops = set(stops_by_direction["return_trip"])
119
120     for stop_id in round_trip_stops:
121         if stop_id in stop_coords[route.route_id]:
122             if stop_id in route_stops[route.route_id]:
123                 route_stops[route.route_id][stop_id]["orientation"] =
124                     "round"
125             else:
126                 route_stops[route.route_id][stop_id] = {
127                     "route_id": route.route_id,
128                     "stop_id": stop_id,
129                     "coordinates": stop_coords[route.route_id][stop_id]
130                 },
131                     "orientation": "round",
132                     "sequence": sequence,
133                     "arrival_times": []
134                 }
135             for stop_id in return_trip_stops:
136                 if stop_id in stop_coords[route.route_id]:
137                     if stop_id in route_stops[route.route_id]:
138                         route_stops[route.route_id][stop_id]["orientation"] =
139                             "return"
140                     else:
141                         route_stops[route.route_id][stop_id] = {
142                             "route_id": route.route_id,
143                             "stop_id": stop_id,
144                             "coordinates": stop_coords[route.route_id][stop_id]
145                         },
146                             "orientation": "return",
147                             "sequence": sequence,
148                             "arrival_times": []
149
150     print("DONE")
151     print("STORING ROUTE GRAPHS...")
152
153     # Store graphs into a file
154     for route_id, graph in graphs.items():
155         weight_prop = graph.new_edge_property("int") # Crear una nueva
156         propiedad de peso de arista
157
158         for e in graph.edges(): # Iterar sobre las aristas del grafo
159             weight_prop[e] = 1 # Asignar el peso 1 a cada arista
160
161         graph.edge_properties["weight"] = weight_prop # Asignar la
162         propiedad de peso al grafo
163
164         data_dir = "gtfs_routes"
165         if not os.path.exists(data_dir):
166             os.makedirs(data_dir)
167         graph.save(f"{data_dir}/{route_id}.gt")
168
169     print("GTFS DATA RECEIVED SUCCESSFULLY")

```

```

164     return graphs, route_stops, special_dates
165
166 # GTFS Graph
167 gtfs_graph, route_stops, special_dates = get_gtfs_data()

```

B.2. Creación del mapa

```

1 # Define the function to set the optimal zoom level for the map
2 def fit_bounds(points, m):
3     """
4         Fits the map bounds to a given set of points.
5
6     Parameters:
7         points (list): A list of points in the format [(lat1, lon1), (lat2,
8             lon2), ...].
9         m (folium.Map): A folium map object.
10    """
11    df = pd.DataFrame(points).rename(columns={0:'Lat', 1:'Lon'})[['Lat', 'Lon']]
12    sw = df[['Lat', 'Lon']].min().values.tolist()
13    ne = df[['Lat', 'Lon']].max().values.tolist()
14    m.fit_bounds([sw, ne])
15
16 # Define the function to create a map that shows the correct public
17 # transport services to take from a source to a target
18 def create_transport_map(route_stops, selected_path, source_date,
19     source_hour, margin):
20     """
21         Creates a map that shows the correct public transport services to take
22         from a source to a target.
23
24     Parameters:
25         route_stops (dict): A dictionary that maps route IDs to a list of stop
26             IDs.
27         selected_path (list): A list of points in the format [(lat1, lon1), (lat2,
28             lon2), ...].
29         nighttime_flag (bool): A flag indicating whether to include nighttime
30             routes.
31         rush_hour_flag (bool): A flag indicating whether to include express
32             routes during rush hour.
33         margin (float): The margin in kilometers around the given addresses.
34
35     Returns:
36         folium.Map: A folium map object.
37     """
38
39     # Note: The margin represents the kilometers around the given
40     # addresses.
41     # Example: a margin of 0.1 represents 0.1 km, or 100 meters.
42
43     geolocator = Nominatim(user_agent="ayatori")
44     source_lat = selected_path[0][0]
45     source_lon = selected_path[0][1]
46     target_lat = selected_path[-1][0]

```

```

37     target_lon = selected_path[-1][1]
38     source = geolocator.reverse((source_lat,source_lon))
39     target = geolocator.reverse((target_lat,target_lon))
40
41     # Create a map that shows the correct public transport services to
42     # take from the source to the target
43     m = folium.Map(location=[selected_path[0][0], selected_path[0][1]],
44                     zoom_start=13)
45
46     # Add markers for the source and target points
47     folium.Marker(location=[selected_path[0][0], selected_path[0][1]],
48                   popup="Origen: {}".format(source), icon=folium.Icon(color='green')).add_to(m)
49     folium.Marker(location=[selected_path[-1][0], selected_path[-1][1]],
50                   popup="Destino: {}".format(target), icon=folium.Icon(color='red')).add_to(m)
51
52     # Add markers for the nearest stop from the source and target points
53     source_coords = (selected_path[0][1], selected_path[0][0])
54     near_source_stops, source_orientations = find_nearest_stops(source,
55     margin)
56
57     target_coords = (selected_path[-1][1], selected_path[-1][0])
58     near_target_stops, target_orientations = find_nearest_stops(target,
59     margin)
60
61     fixed_orientation = None
62     valid_services = set()
63     for source_stop_id in near_source_stops:
64         for target_stop_id in near_target_stops:
65             services = connection_finder(route_stops, source_stop_id,
66             target_stop_id)
67             for service in services:
68                 source_orientation = get_bus_orientation(service,
69                 source_stop_id)
70                 target_orientation = get_bus_orientation(service,
71                 target_stop_id)
72                 source_sequence = int(get_trip_sequence(route_stops,
73                 service, source_stop_id))
74                 target_sequence = int(get_trip_sequence(route_stops,
75                 service, target_stop_id))
76                 if source_sequence > target_sequence:
77                     continue
78                 if isinstance(source_orientation, list) and isinstance(
79                 target_orientation, list):
80                     # If both source and target orientations are lists,
81                     # check if any of the values match
82                     valid_orientation = any(x in target_orientation for x
83                     in source_orientation) or any(x in source_orientation for x in
84                     target_orientation)
85                     if valid_orientation and service not in valid_services
86                     :
87                         valid_services.add(service)
88                         fixed_orientation = [x for x in source_orientation
89                         if x in target_orientation][0] if [x for x in source_orientation if x
90                         in target_orientation] else source_orientation[0]

```

```

73             elif source_orientation == target_orientation and service
74                 not in valid_services: # Check if both stops are visited in the same
75                     orientation
76                         valid_services.add(service)
77                         fixed_orientation = target_orientation
78                         elif isinstance(source_orientation, list) and
79                             target_orientation in source_orientation and service not in
80                             valid_services:
81                                 valid_services.add(service)
82                                 fixed_orientation = target_orientation
83                                 elif isinstance(target_orientation, list) and
84                                     source_orientation in target_orientation and service not in
85                                     valid_services:
86                                         valid_services.add(service)
87                                         fixed_orientation = source_orientation
88
89             #print("ORIENTATION FIXED: {}".format(fixed_orientation))
90
91             if len(valid_services) == 0:
92                 print("Error: There are no available services right now to go to
93                     the desired destination.")
94                 print("Possible reasons: no routes that have stops near the source
95                     and target addresses.")
96                 print("You can try changing the search margin and try again.")
97                 return
98
99             nighttime_flag = is_nighttime(source_hour)
100            rush_hour_flag = is_rush_hour(source_hour)
101            holiday_flag = is_holiday(source_date)
102            if holiday_flag:
103                rush_hour_flag = 0
104
105            # Nighttime check
106            daily_time_services = check_night_routes(valid_services,
107                nighttime_flag)
108
109            if daily_time_services is None:
110                print("Error: There are no available services right now to go to
111                     the desired destination.")
112                print("Possible reasons: Source hour is during nighttime.")
113                print("Please take into account that nighttime goes between
114                     00:00:00 and 05:30:00.")
115                return
116
117            # Rush hour check
118            valid_services = check_express_routes(daily_time_services,
119                rush_hour_flag)
120
121            valid_services = list(set(valid_services))
122
123            valid_source_stops = [stop_id for stop_id in near_source_stops if any(
124                route_id in valid_services for route_id in route_stops.keys() if
125                stop_id in route_stops[route_id])]
126            valid_source_stops = list(set(valid_source_stops))
127            valid_target_stops = [stop_id for stop_id in near_target_stops if any(
128                route_id in valid_services for route_id in route_stops.keys() if

```

```

114     stop_id in route_stops[route_id]])
115     valid_target_stops = list(set(valid_target_stops))
116
117     # Give info
118     print("")
119     print("Routes have been found.")
120     print("Calculating the best route and getting the arrival times for")
121     print("the next buses...")
122
123     best_option = None
124     best_option_times = None
125     source_time = timedelta(hours=source_hour.hour, minutes=source_hour.
126     minute, seconds=source_hour.second)
127
128     valid_orientations = set(source_orientations)
129
130     best_option_orientation = None
131
132     valid_target = []
133     for target_stop in valid_target_stops:
134         target_routes = get_routes_at_stop(route_stops, target_stop)
135         valid_target.extend(target_routes)
136     valid_target = list(dict.fromkeys(valid_target))
137
138     for stop_id in valid_source_stops:
139         routes_at_stop = get_routes_at_stop(route_stops, stop_id)
140         valid_stop_services = [stop_id for stop_id in valid_services if
141         stop_id in routes_at_stop]
142         for valid_service in valid_stop_services:
143             a_t = get_arrival_times(valid_service, stop_id, source_date)
144             if a_t is not None and a_t[0] == fixed_orientation:
145                 orientation = a_t[0]
146                 flag = False
147                 for target_stop_id in valid_target_stops:
148                     flag = False
149                     target_stop_routes = get_routes_at_stop(route_stops,
150                     target_stop_id)
151                     if valid_service in target_stop_routes:
152                         target_orientation = get_bus_orientation(
153                         valid_service, target_stop_id)
154                         if a_t[0] != target_orientation:
155                             flag = True
156                             continue
157
158                         if flag:
159                             continue
160
161                         if valid_service not in valid_target:
162                             continue
163
164                         arrival_times = a_t[1]
165                         #print(arrival_times)
166                         time_until_next_buses = get_time_until_next_bus(
167                         arrival_times, source_hour, source_date)
168
169                         if not time_until_next_buses:

```

```

163         print("Error: There are no available services right
now to go to the desired destination.")
164         print("Possible reasons: There are no buses left today
. Maybe the source hour is too close to the ending time for the service
.")
165         return
166
167
168     # Print the time until the next three buses in the desired
format
169     for i in range(len(time_until_next_buses)):
170         minutes, seconds = time_until_next_buses[i]
171         waiting_time = timedelta(minutes=minutes, seconds=
seconds)
172         arrival_time = source_time + waiting_time
173         time_string = timedelta_to_hhmm(arrival_time)
174
175         target_orientation = get_bus_orientation(valid_service
, target_stop_id)
176
177         # Update the best option
178         if (best_option is None or (arrival_time < best_option
[2])) and orientation == fixed_orientation:
179             best_option = (valid_service, stop_id,
arrival_time, waiting_time)
180             best_option_times = time_until_next_buses
181             best_option_orientation = orientation
182
183     if best_option is None:
184         print("Error: There are no available services right now to go to
the desired destination.")
185         print("Possible reasons: maybe a small margin. You can try using a
bigger one")
186         return
187
188     # Print the best option
189     arrival_time = None
190
191     print("")
192     print("To go from: {}".format(source))
193     print("To: {}".format(target))
194     best_arrival_time_str = timedelta_to_hhmm(best_option[2])
195     print("The best option is to take the route {} on stop {}. The next
bus arrives at {}.".format(best_option[0], best_option[1],
best_arrival_time_str))
196     print("The other two next buses arrives in:")
197     for i in range(len(best_option_times)):
198         if i == 0:
199             continue
200         minutes, seconds = best_option_times[i]
201         waiting_time = timedelta(minutes=minutes, seconds=seconds)
202         arrival_time = source_time + waiting_time
203         time_string = timedelta_to_hhmm(arrival_time)
204         print(f"{minutes} minutes, {seconds} seconds ({time_string})")
205
206

```

```

207     # Map the options
208     for stop_id in near_source_stops:
209         if stop_id in valid_source_stops:
210             stop_coords = get_stop_coords(route_stops, str(stop_id))
211             routes_at_stop = get_routes_at_stop(route_stops, stop_id)
212             valid_stop_services = [stop_id for stop_id in valid_services
213             if stop_id in routes_at_stop]
214
215             for service in valid_stop_services:
216                 if service == best_option[0] and stop_id == best_option
217                 [1]:
218                     folium.Marker(location=[stop_coords[1], stop_coords
219                     [0]],
220                     popup="Mejor opcion: subirse al recorrido {} en
221                     el paradero {}".format(best_option[0], best_option[1]),
222                     icon=folium.Icon(color='cadetblue', icon='plus')
223                     ).add_to(m)
224
225                     initial_distance = [(selected_path[0][0],
226                     selected_path[0][1]), (stop_coords[1], stop_coords[0])]
227                     folium.PolyLine(initial_distance, color='black',
228                     dash_array='10').add_to(m)
229
230             for stop_id in near_target_stops:
231                 if stop_id in valid_target_stops:
232                     stop_coords = get_stop_coords(route_stops, str(stop_id))
233                     routes_at_stop = get_routes_at_stop(route_stops, stop_id)
234                     valid_stop_services = [stop_id for stop_id in valid_services
235                     if stop_id in routes_at_stop]
236
237                     target_orientation = None
238                     for service in valid_target:
239                         if service == best_option[0]:
240                             if fixed_orientation == "round":
241                                 trip_id = service + "-I-" + get_trip_day_suffix(
242                     source_date)
243                             else:
244                                 trip_id = service + "-R-" + get_trip_day_suffix(
245                     source_date)
246
247                             best_travel_time = None
248                             selected_stop = None
249                             for stop_id in valid_target_stops:
250                                 bus_time = get_travel_time(trip_id, [best_option[1],
251                     stop_id])
252                                 target_stop_routes = get_routes_at_stop(route_stops,
253                     stop_id)
254                                 target_orientation = get_bus_orientation(best_option[0],
255                     stop_id)
256                                 if service in target_stop_routes and bus_time > timedelta
257                                 () and (best_travel_time is None or bus_time < best_travel_time):
258                                     if fixed_orientation == target_orientation:
259                                         best_travel_time = bus_time
260                                         selected_stop = stop_id
261
262                                         selected_stop_coords = get_stop_coords(route_stops,
263                                         selected_stop)

```

```

248     minutes, seconds = timedelta_separator(best_travel_time)
249
250     print("DEBUG")
251     seq_1 = route_stops[best_option[0]][best_option[1]]["sequence"]
252 ]
253     seq_2 = route_stops[best_option[0]][selected_stop]["sequence"]
254     print(seq_1, seq_2)
255     print("")
256     print("You will get off the bus on stop {} after {} minutes"
257           "and {} seconds.".format(selected_stop, minutes, seconds))
258
259         folium.Marker(location=[selected_stop_coords[1],
260                           selected_stop_coords[0]],
261                           popup="Mejor opcion: bajarse del recorrido {} en el"
262                           "paradero {}.".format(best_option[0], selected_stop),
263                           icon=folium.Icon(color='cadetblue', icon='plus')).add_to
264                           (m)
265
266         ending_distance = [(selected_path[-1][0], selected_path
267                           [-1][1]), (selected_stop_coords[1], selected_stop_coords[0])]
268         folium.PolyLine(ending_distance, color='black', dash_array='10')
269         .add_to(m)
270
271         total_time = best_option[3] + best_travel_time
272         minutes, seconds = timedelta_separator(total_time)
273
274         destination_time = source_time + total_time
275         time_string = timedelta_to_hhmm(destination_time)
276         print(f"Total travel time: {minutes} minutes, {seconds}"
277               "seconds. You will arrive your destination at {time_string}.")
278
279 # Set the optimal zoom level for the map
280 fit_bounds(selected_path, m)
281
282 return m

```

B.3. Operación del algoritmo

```

1 def connection_scan(graph, source_address, target_address, departure_time,
2                     departure_date):
3     """
4     The Connection Scan algorithm is applied to search for travel routes
5     from the source to the destination,
6     given a departure time and date. By default, the algorithm uses the
7     current date and time of the system.
8     However, you can specify a different date or time if needed.
9
10    Args:
11        graph (graph): the graph used to visualize the travel routes.
12        source_address (string): the source address of the travel.
13        target_address (string): the destination address of the travel.
14        departure_time (time): the time at which the travel should start.
15        departure_date (date): the date on which the travel should be done
16

```

```

13
14     Returns:
15         list: the list of coordinates of the travel connections needed to
16             arrive at the destination.
17     """
18
19     node_id_mapping = create_node_id_mapping(graph)
20
21
22     if source_node is not None and target_node is not None:
23         # Convert source and target node IDs to integers
24         source_node_graph_id = graph.vertex_properties["graph_id"][
25             source_node]
26         target_node_graph_id = graph.vertex_properties["graph_id"][
27             target_node]
28
29         print("Both addresses have been found.")
30         print("Processing...")
31         #print("SOURCE NODE: {}. TARGET NODE: {}".format(
32             source_node_graph_id, target_node_graph_id))
33
34         path = [source_node_graph_id, target_node_graph_id]
35         path_coords = []
36         for node in path:
37             lon, lat = graph.vertex_properties["lon"][node], graph.
38             vertex_properties["lat"][node]
39             path_coords.append((lat, lon))
40             #print(path)
41
42
43     def csa_commands():
44         """
45             Process the inputs given by the user to run the Connection Scan
46             Algorithm.
47         """
48
49         # System's date and time
50         now = datetime.now()
51         dt_string = now.strftime("%d/%m/%Y %H:%M:%S")
52         #print("Fecha y hora actuales =", dt_string)
53
54         # Date formatting
55         today = date.today()
56         today_format = today.strftime("%d/%m/%Y")
57
58         # Time formatting
59         moment = now.strftime("%H:%M:%S")
60         used_time = datetime.strptime(moment, "%H:%M:%S").time()
61
62         # User inputs
63         # Date and time

```

```

63     source_date = input(
64         "Enter the travel's date, in DD/MM/YYYY format (press Enter to use
65 today's date) : ") or today_format
66     print(source_date)
67     source_hour = input(
68         "Enter the travel's start time, in HH:MM:SS format (press Enter to
69 start now) : ") or used_time
70     if source_hour != used_time:
71         source_hour = datetime.strptime(source_hour, "%H:%M:%S").time()
72     print(source_hour)
73
74 # Source address
75 source_example = "Beauchef 850, Santiago"
76 while True:
77     source_address = input(
78         "Enter the starting point's address, in 'Street #No, Province'
79 format (Ex: 'Beauchef 850, Santiago'):")
80     if source_address.strip() != '':
81         #print("Direccion de Destino ingresada: " + target_address)
82         break
83
84 # Destination address
85 destination_example = "Campus Antumapu Universidad de Chile, Santiago"
86 while True:
87     target_address = input(
88         "Enter the ending point's address, in 'Street #No, Province'
89 format (Ex: 'Campus Antumapu Universidad de Chile, Santiago'):")
90     if target_address.strip() != '':
91         #print("Direccion de Destino ingresada: " + target_address)
92         break
93
94     start = tm.time()
95
96     path_coords = connection_scan(undirected_graph, source_address,
97 target_address, source_hour, source_date)
98
99     if path_coords:
100         map = create_transport_map(route_stops, path_coords, source_date,
101 source_hour, 0.2)
102         if map:
103             display(map)
104
105             end = tm.time()
106             exec_time = round((end-start) / 60,3)
107             print("MAP IS READY. EXECUTION TIME: {} MINUTES".format(exec_time))
108         )
109         return path_coords
110
111     else:
112         print("")
113         print("Something went wrong. Please try again later.")
114         return
115
116 selected_path = csa_commands()

```