



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría en Ciencias (Física)

Instituto de Física

Detección de patrones en la actividad del sistema Metrobús
en la Ciudad de México

Tesis

QUE PARA OPTAR AL GRADO DE:

Maestro en Ciencias (Física)

PRESENTA:

Fis. Jaspe Uriel Martínez González

DIRECTOR DE TESIS:

Dr. Alejandro Pérez Riascos

Instituto de Física

COMITÉ TUTOR:

Dr. José Luis Mateos Trigos

Instituto de Física

Dr. Ricardo Atahualpa Solórzano Kraemer

Facultad de Ciencias

CIUDAD UNIVERSITARIA, CDMX, JUNIO 2021



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología, instancia que me otorgó una beca de posgrado para cursar mi maestría sin distracciones económicas.

A la beca PAPIIT IN116220 por su apoyo económico para continuar con la investigación que se aborda en este trabajo.

A mi Comité Tutor, conformado por el Dr. José Luis Mateos Trigos y Dr. Ricardo Atahualpa Solórzano Kraemer por su orientación y consejos a lo largo del proceso de maestría. En particular, agradezco a mi Tutor Principal, el Dr. Alejandro Pérez Riascos, por su invaluable labor como maestro y director de esta investigación, a él pertenece gran parte del mérito en este trabajo.

A mi Sínodo, por haber dedicado parte de su tiempo para hacer las correcciones necesarias para que este trabajo llegara a buen puerto, en especial, menciono la profunda revisión del Dr. Víctor Dossetti Romero.

A la UNAM y al Instituto de Física por ser las casas de estudio que me brindaron todas las facilidades para continuar con mi formación académica a nivel de posgrado.

A mi tío, Ing. Salatiel Trejo Martínez, por el invaluable apoyo que no agradecí en mi trabajo anterior, enmiendo así mi falta.

Dedicatorias

A mi padre[†] por seguir estando aquí para ser mi guía.

Sólo rocío
es el mundo, rocío
y sin embargo...

A mi madre y a mi hermana, por su apoyo imperecedero.

Pone al niño a dormir
y ahora lava la ropa
luna de verano.

A Julieta, por otra campaña con su amor y compañía.

De no estar tú
demasiado grande
sería el bosque.

Al Dr. Adolfo Cordero Borboa, por ser el mejor maestro.

Dijo: “antaño, el mar
llegaba hasta aquí”,
y puso más leña en el fuego.

Resumen

En este trabajo de investigación se analiza el movimiento de vehículos del sistema Metrobús en la Ciudad de México por medio del análisis de datos recolectados durante un período de seis meses. Estos datos dan la actividad en tiempo real del sistema y son suministrados de manera abierta por la administración del sistema Metrobús. A esta información se le agregaron los datos geográficos de una segmentación de este sistema de transporte. Con base en ello, se encontraron los patrones diarios de velocidad, así como los patrones de cada uno de los segmentos. Posteriormente, se usó la distancia de Kullback-Leibler para asignar un valor objetivo a la similitud entre cada una de las distribuciones de velocidad anteriormente descritas; el conjunto de valores obtenido sirvió para encontrar la medida en que determinados días o segmentos ven afectadas sus tendencias de velocidad por la incidencia de factores internos o externos al sistema. Por otro lado, se organizaron en un arreglo matricial las distancias de Kullback-Leibler entre cada uno de los segmentos para obtener una red equivalente que describe similitudes entre la actividad de cada segmento y a la que se le aplicó un algoritmo para detectar la formación de comunidades en la red que describe la actividad del sistema Metrobús. Finalmente, estos últimos resultados se interpretaron en el mapa segmentado de la red para asociar tendencias de velocidad y de formación de comunidades con la ubicación geográfica de cada uno de los segmentos.

Índice general

1. Transporte en zonas urbanas	12
1.1. Introducción	12
1.2. Desarrollo del transporte	13
1.3. Sistemas de transporte público	14
1.4. Descripción de sistemas mediante redes	18
1.4.1. Redes y grafos	18
1.4.2. Propiedades de las redes	20
1.4.3. Comunidades	21
1.5. Divergencia de Kullback-Leibler	25
2. Sistema Metrobús CDMX	27
2.1. Introducción	27
2.2. Autobús de tránsito rápido	28
2.3. Comparación con otros sistemas de transporte	29
2.4. Características del sistema Metrobús	31
2.4.1. Líneas del sistema Metrobús	32
2.4.2. Estaciones	35
2.4.3. Parque vehicular	36

2.4.4. Infraestructura de ingreso al sistema	37
2.5. Problemas actuales y perspectivas a futuro	38
3. Actividad del metrobús: caracterización por segmentos	39
3.1. Introducción	39
3.2. Software utilizado para el análisis	40
3.2.1. Formato <i>General Transit Feed Specification</i>	40
3.2.2. Formato JSON	40
3.2.3. Python	41
3.2.4. PANDAS	41
3.2.5. QGIS	41
3.2.6. Mathematica	42
3.3. Descarga y tratamiento de los datos	42
3.4. Histogramas de velocidad	48
3.5. Actividad vehicular en el sistema metrobús	53
3.6. Divergencia de Kullback-Leibler	57
4. Patrones en histogramas de velocidad	65
4.1. Introducción	65
4.2. Matriz de distancias de Kullback–Leibler	66
4.3. Redes de similitud entre segmentos	70
4.4. Comunidades en la red de similitud	73
4.5. Comunidades en el sistema Metrobús	74

Índice de figuras

1.1. Medios de transporte urbano de gran capacidad.	15
1.2. Medios de transporte sin gran capacidad de movilización.	16
1.3. Esquematización de varias redes.	19
1.4. Ejemplo de la formación de comunidades en una red.	22
2.1. Comparación de sistemas de transporte masivo.	31
2.2. Esquema de las líneas del Metrobús.	33
2.3. Fotografías de los cuatro tipos de autobuses usados en la red de metrobús de la Ciudad de México.	36
2.4. Sistema de cobro del metrobús.	37
3.1. Registros del sistema metrobús del día 2 de marzo del 2020.	43
3.2. Datos del Metrobús en el software QGIS.	46
3.3. Histograma de las velocidades registradas el 16 de febrero.	47
3.4. Histogramas de velocidad del 3 de marzo, 8 de marzo, 24 de abril y 11 de junio.	49
3.5. Histogramas de velocidad de todo el sistema durante los días lunes de las 18 semanas de estudio.	51
3.6. Distribuciones de velocidad en cada una de las horas de servicio.	52

3.7. Histograma del número de vehículos en circulación en toda la red de Metrobús durante cada día del tiempo de estudio.	53
3.8. Histogramas del número de unidades en circulación en toda la red para cada uno de los días de la semana.	54
3.9. Promedios del número de unidades en circulación.	56
3.10. Histogramas <i>total</i> y diarios, para toda la red.	58
3.11. Histograma de las distancias de Kullback-Leibler.	59
3.12. Comparación entre el <i>histograma total</i> y el histograma para dos segmentos.	61
3.13. Densidad de probabilidad de las distancias de Kullback-Leibler \mathcal{D}_{KL} para cada uno de los segmentos.	62
3.14. Valores promedio de las distancias de Kullback-Leibler $\langle \mathcal{D}_{KL} \rangle$ para los segmentos del sistema metrobús.	62
3.15. Densidad de probabilidad $\rho(v)$ para las velocidades v en cada uno de los segmentos.	63
4.1. Matriz de distancias de Kullback-Leibler simétrica \mathcal{D}_{KLS}	67
4.2. Comparación entre pares particulares de segmentos.	68
4.3. Densidad de probabilidad $\rho(\mathcal{D}_{KLS})$ de todas las distancias \mathcal{D}_{KLS}	69
4.4. Matrices adyacencia graficadas según la distancia mínima h	71
4.5. Red de similitud generada con $h = 0.05$	72
4.6. Comunidades detectadas en la componente gigante en redes de similitud.	74
4.7. Detección de comunidades en el sistema Metrobús con $h = 0.035$	76
4.8. Detección de comunidades en el sistema Metrobús con $h = 0.05$	78

Índice de tablas

2.1. Sistemas de autobús de tránsito rápido en México, enlistados por orden de inauguración.	28
3.1. Ejemplo de los datos ordenados en un <i>Data Frame</i>	44
3.2. Número de unidades en circulación por cada una de las semanas de estudio.	55
4.1. Propiedades de las redes generadas a partir de matrices de distancias KL .	73
4.2. Propiedades comunidades encontradas con $h = 0.05$	77

Introducción

Actualmente, los seres humanos nos trasladamos de un lugar a otro en una amplia variedad de escalas: desde el viaje al trabajo dentro de una misma ciudad hasta éxodos que cruzan continentes enteros; esta extensa gama de variantes hacen que la movilidad humana se analice desde enfoques igual de numerosos. Aún así, popularmente se considera que tales estudios competen sólo al ámbito de las ciencias sociales, en ese tenor, autores como Juval Portugali [1] consideran que existen *dos culturas* de estudio de las ciudades, una que se basa en las ciencias sociales y otra que lo hace en las llamadas *ciencias duras*. Aquí se aborda esa segunda cultura para estudiar un problema que las grandes ciudades tienen en común: el transporte público eficiente [2]; mismo que incide en la *salud* de la ciudad y, sobre todo, en la de sus habitantes [3]. La solución no se limita a un aumento de la infraestructura, también requiere eficiencia [4]. Para ello, la combinación de la física estadística y el análisis del *big data* amplía los horizontes para afrontar tal desafío [5], [6], [7]. Estos antecedentes inspiran la presente tesis, que pretende ser un primer acercamiento al estudio matemático del sistema de autobús de tránsito rápido de la Ciudad de México, no obstante, el espíritu aquí depositado aspira a sentar las bases de nuevos estudios en el área.

La investigación que aquí se ofrece consta de 4 capítulos. El primero contiene el marco teórico necesario para entender el objetivo de la investigación y la estrategia que se utilizará. Este marco teórico se divide en dos: un breve resumen sobre el transporte público y la exposición de las herramientas matemáticas que se usarán para encontrar las propiedades de nuestro objeto de estudio. La primera parte es tema de las dos secciones con las que abre el capítulo, y la segunda se aborda en las 2 secciones restantes, una versa sobre la *teoría de redes* y la otra sobre la *divergencia de Kullback-Leibler*.

En el segundo capítulo, el texto se centra en el objeto de la investigación: el sistema de autobús de tránsito rápido de la Ciudad de México. El capítulo empieza explicando a detalle el concepto de autobús de tránsito rápido y ofrece una breve reseña histórica de su

implementación en el mundo y en México. En la siguiente sección se comparan los costos y la capacidad de movilización de los modelos de transporte masivo. En la sección 2.4 se abunda en la descripción del *Metrobús* y se cierra el capítulo con la exposición de los problemas a los que se enfrenta este medio de transporte, así como algunas perspectivas a futuro que de él se tienen.

Al inicio del capítulo 3 se presentan las herramientas de cómputo usadas para analizar los datos que la administración del Metrobús libera por medio de un servidor dinámico. Posteriormente, se explica la manera en que se descargaron los datos y el tiempo total de recolección; también se detalla cómo estos se manipularon para optimizar su estudio. Las secciones 3.4 y 3.5 contienen los primeros resultados de este trabajo. El capítulo cierra con los resultados de comparar las bases de datos mediante el uso de la *distancia de Kullback-Leibler*.

El último capítulo consta de 4 secciones, en la primera de ellas se ordenan en un arreglo matricial las distancias de Kullback-Leibler entre los segmentos de la red. De la matriz anterior, se obtienen 6 redes por medio de la variación de un criterio de cercanía. En seguida, se usa un algoritmo para detectar la formación de comunidades dentro de cada una de las redes obtenidas. Finalmente, se relacionan cada una de estas comunidades de segmentos con su ubicación geográfica en el mapa y dilucidar el origen de las propiedades de cada comunidad.

Al final de este texto, se presentan las conclusiones obtenidas a lo largo de la investigación, y a estas se agregan las perspectivas para trabajos futuros.

Capítulo 1

Transporte en zonas urbanas

1.1. Introducción

A menudo se considera que el estudio de las ciudades es un campo que sólo compete a quienes se forman en las ciencias sociales, tales como la sociología o la administración pública; la fama de los aportes que hacen la física o las matemáticas para ampliar nuestro entendimiento de las ciudades son relativamente menores. Autores como Joval Portugali [1] emplean el término *dos culturas* para abordar el estudio de las ciudades. Una de ellas se basa en las ciencias sociales y la otra en las *ciencias duras*; división que sintetiza adecuadamente ambas visiones. El capítulo inicial de este trabajo tiene como finalidad mostrar al lector las ventajas que resultan del enfoque de esta segunda cultura y, en lo particular, se mostrará la utilidad de aplicar las técnicas de esta segunda cultura de manera directa al problema de interés en esta tesis: las propiedades subyacentes del sistema de autobús de tránsito rápido de la Ciudad de México.

Con estas ideas se explicarán la historia y el desarrollo del transporte, mencionando el papel que juega el factor económico en ambos desarrollos. En la primera parte se hace énfasis en el tema que atañe a esta tesis, el transporte público; iniciando con la diferenciación entre el transporte *foráneo* y el transporte *urbano*, describiendo en detalle estos últimos. También se presentan algunas razones para implementar sistemas de transporte público que sean eficientes. Esta primera parte cierra con una exposición breve de las ciudades como un sistema complejo.

En la segunda parte, se presenta la segunda cultura de estudio de las ciudades repasando

los elementos matemáticos que se usarán a lo largo de este trabajo de tesis. Estos conceptos provienen de dos ramas de las matemáticas: la *ciencia de redes* y la estadística. En cuanto a la teoría de redes, se hace mención a propiedades como *grado de la red* y *coeficiente de agrupamiento*, así como *red de mundo pequeño* y *red libre de escala*, complementando lo anterior con el concepto de *comunidad* dentro de una red. Por último, se presenta la *distancia de Kullback-Leibler*, una técnica matemática que sirve para obtener la similitud entre dos distribuciones de datos; dado que la información descargada puede verse como una distribución de datos, este concepto será muy útil para el desarrollo del presente trabajo.

1.2. Desarrollo del transporte

La historia del transporte ha sido parte fundamental de la historia del ser humano: desde los orígenes mismos de nuestra especie ha existido la necesidad de cruzar largas distancias en búsqueda de nuevas fuentes de alimentación. Con nuestra conversión en una especie sedentaria, el transporte quedó ligado a la actividad económica que surge como consecuencia del desarrollo de la agricultura [8]. Ejemplo de ello es el uso de animales domesticados para el transporte, como los equinos o los camélidos. Por otra parte, la necesidad de mover mercancías impulsa la creación de medios de transporte, como el barco, cuya importancia es tal que en la actualidad el 80 % de las mercancías del mundo se transportan por este medio [9].

A partir de la revolución industrial se desarrollan los motores, tanto de combustión externa como interna [8]. El primero posibilitó la construcción de locomotoras al aprovechar la fuerza de una máquina de vapor para tirar de vagones; este invento da origen al tren y a los transportes que de él derivan, como el metro y el tren ligero, y hoy en día el tren sigue usándose para el transporte masivo terrestre de personas y mercancías. El motor de combustión interna posibilitó la creación de máquinas como el automóvil o el avión; estos motores siguen desarrollándose para obtener el máximo rendimiento del combustible utilizado, lo que impulsa el desarrollo de la ciencia aplicada, pues, entre otras cosas, la propulsión del transporte también utiliza combustible nuclear y electricidad generada a partir de fuentes renovables como el sol cuya tecnología sigue perfeccionándose. Lo que es más, el estudio de los primeros ciclos de los motores dio origen a la Termodinámica moderna.

Hoy en día, las tecnologías de la información han reducido la necesidad de traslados

debidos al trabajo [2], pero sigue siendo importante el transporte de personas a causa del turismo, cuyo impacto económico continúa siendo muy grande. Además, el transporte es un componente fundamental de cualquier ciudad del mundo, y por tanto, una fuente de complejidad en la misma. En vista de ello, pueden retomarse las ideas de Michael Batty [10] para dejar de ver a la ciudad como un conjunto de edificios, lugares y personas; y verlas como un conjunto de acciones, interacciones y transacciones, cuya complejidad puede abordarse desde un punto de vista matemático.

1.3. Sistemas de transporte público

El primer sistema de transporte masivo en el que las ciudades basan la movilidad pública de sus habitantes es el de las flotillas de minibuses, organizadas en rutas fijas, que suelen concentrarse en torno a paraderos conocidos popularmente en México como bases. Hasta antes de la aparición de los sistemas de autobuses de tránsito rápido, las ciudades que no se podían permitir un metro basaban el transporte masivo de pasajeros en este formato de transporte público. Normalmente operan como una concesión dada por el gobierno local a particulares, aunque en algunos casos se da una administración mixta y en otros casos es el gobierno local quien toma en sus manos la provisión del servicio, como el de la extinta *Ruta 100*, que puede observarse en la Figura 1.1 (d). La capacidad de las unidades varía, desde unos 15 pasajeros para vagonetas, hasta capacidades de 75 pasajeros para minibuses [11].

El metro es un medio de transporte que deriva del tren, con la diferencia de que el primero es exclusivamente urbano [11]. Consiste en un convoy de entre 2 y 10 vagones tirados por una locomotora, estos pueden ser de rodadura férrea o neumática. La implementación de un metro precisa de infraestructura especial, tales como vías exclusivas e instalaciones para las estaciones. Es el sistema de transporte público con mayor capacidad y su implementación suele ser indicativo del poder económico de una ciudad, dado su alto costo [11], cuestión que se abordará en detalle en el siguiente capítulo. Este sistema de transporte nace en Londres y desde entonces ha sido implementado en más de 150 ciudades en todo el mundo, acumulando más de 8000 estaciones y más de 10000 km de vías [18]. Una de las consecuencias de la implantación del metro como sistema de transporte es la creación de *ciudades dormitorio* [19], ubicadas a unas decenas de kilómetros de la ciudad principal. Los viajes entre la ciudad principal y las ciudades dormitorio son posibles debido a la reducción de tiempo de viaje asociada con este sistema de transporte. En México existen

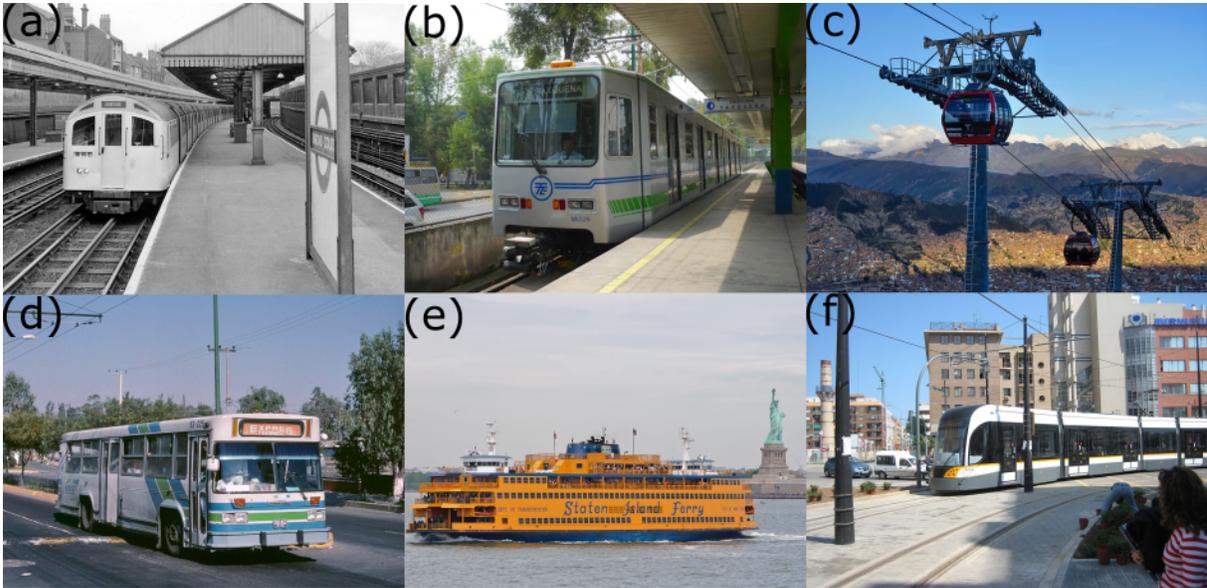


Figura 1.1: Existen diversos medios de transporte urbano con capacidad para movilizar grandes cantidades de pasajeros, como: (a) el metro, cuya invención se dió en Londres [12]. (b) El tren ligero, como el de la Ciudad de México [13]. (c) El teleférico para zonas urbanas ubicadas en terrenos montañosos, como el que comunica a La Paz con El Alto, en Bolivia [14]. (d) Flotillas de minibuses, como el de la extinta *Ruta 100* en la Ciudad de México [15]. (e) El transbordador, o ferry que sirve para transportar pasajeros a través de grandes cuerpos de agua, como el que comunica a Staten Island con Nueva York [16]. (f) El Tranvía, como el que se ubica en Sevilla, España [17].

3 sistemas de metro: el de la Ciudad de México con 195 estaciones repartidas en 12 líneas a lo largo de 225 kilómetros; el de Guadalajara, que tiene 48 estaciones repartidas en tres líneas sobre 47 kilómetros de vías; y el Monterrey, cuyo sistema de metro tiene 31 estaciones en 2 líneas que totalizan 31 kilómetros de vías.

La distinción entre metro, tren eléctrico urbano y tren ligero no siempre es muy clara, y suele basarse en el tamaño y capacidad de los vagones usados [11], así como de una infraestructura menos robusta para operar el sistema, que se ejemplifica en la Figura 1.1 (b), además del hecho de ser de rodadura totalmente férrea y no asegurar la exclusividad de sus vías [11], pues existen algunos cruces con calles y avenidas de las ciudades. Dentro de este tipo de transportes están los sistemas modernos de *tranvías*, cuyas vías yacen totalmente dentro del arroyo vehicular de las ciudades como se advierte en el ejemplo ilustrado de la Figura 1.1 (f).

Para aquellas ciudades en que una parte importante de su extensión y de su población se ubique en zonas montañosas existe la opción de operar un teleférico para comunicarla eficientemente con el resto de la ciudad. Dado que en estas áreas los caminos son estrechos y sinuosos, por las condiciones del terreno, es complicado introducir sistemas como el



Figura 1.2: Existen otros medios de transporte públicos, pero que no tienen la capacidad de transportar a muchos pasajeros en un sólo viaje, como es el caso del taxi; uno de los sistemas más famosos es el de los taxis de Nueva York (a) [21]. Un sistema de creación reciente para transporte individual es el de las bicicletas compartidas, como en el sistema público de la Ciudad de México (b) [22].

metro o el autobús de tránsito rápido [20]. Este es un medio de transporte en el que una serie de cabinas son transportadas vía aérea por medio de cables sostenidos entre enormes postes que se elevan decenas de metros sobre el suelo, como puede observarse en la Figura 1.1 (c). Este mecanismo ya es utilizado como sistema de transporte masivo en ciudades como La Paz, Bolivia; Medellín, Colombia; y Ecatepec, México; y en la Ciudad de México está por implementarse este servicio, que plantea convertirse en el más grande de su tipo en el mundo. Por otro lado, el *transbordador* es una embarcación que comunica zonas de una ciudad que están separadas por un grandes cuerpos de agua, como ríos caudalosos, lagos o estrechos marítimos, un ejemplo de este sistema se ilustra en la Figura 1.1 (e). Y puede encontrarse en ciudades como Estambul, Turquía; Nueva York, EUA y Hong Kong.

Por otra parte, el Taxi es un automóvil, usualmente tipo *sedán*, véase Figura 1.2 (a), que es contratado por un pasajero o, como mucho, por la máxima capacidad de la unidad. A diferencia del minibús, no tienen una ruta preestablecida por lo que el pasajero fija el destino. En ciudades grandes, los taxis cuentan con un dispositivo que calcula el monto total del viaje que no es menor que un monto mínimo. Normalmente, los taxis tienen una base, que es donde empiezan su jornada laboral. En los últimos años, han surgido empresas que ofrecen una plataforma digital para que el dueño de un automóvil de uso particular registre una cuenta y ofrezca servicio de taxi; la inserción de este tipo de negocio no ha estado exento de problemas con las asociaciones de taxis en las ciudades donde se implementan estos formatos de viaje.

En un esfuerzo para disminuir los índices de contaminación, gobiernos de ciudades de todo el mundo fomentan el uso de la bicicleta para trayectos cortos dentro de ciudades

[3], para lo cual han implementado la creación de vías exclusivas para bicicletas. También, en muchas de esas ciudades se han creado sistemas públicos de bicicletas compartidas. La Ciudad de México opera desde hace 11 años el sistema *EcoBici* que tiene una infraestructura de 6800 bicicletas y 480 cicloestaciones [23]. Esta infraestructura se observa en la Figura 1.2 (b). Para acceder al servicio, el usuario registra una cuenta bancaria para obtener una tarjeta con la que podrá hacer uso de las bicicletas del sistema al colocar dicha tarjeta en el sensor de cada una de las estaciones. Al terminar de usarla, el usuario debe estacionar la bicicleta en alguna otra estación. El tiempo máximo de uso es de 45 minutos, antes de que venza el plazo debe cambiar de bicicleta o se hará acreedor a una sanción económica cuyo monto dependerá del tiempo extra en que tenga posesión de la unidad. Además de esta empresa pública, otras compañías privadas hacen lo mismo. Por último, el autobús de tránsito rápido es otro sistema de transporte público de adopción relativamente reciente, en comparación de aquellos que ya han sido mencionados en párrafos anteriores. En esta parte no se explicará su definición, pues el siguiente capítulo versa sobre este sistema, en el cual se detallarán todas sus características.

Dado el porcentaje actual de población urbana en el mundo, mayor al 50 %, [10], es sencillo imaginarse la cantidad de viajes realizados dentro de las ciudades; y aunque las telecomunicaciones redujeron la necesidad de los traslados para el transporte de información y personas [2], sigue siendo clara la importancia que tiene un medio de transporte eficiente para el correcto funcionamiento de las ciudades. Por un lado, está el factor económico y, por otro, el factor humano. Del primero puede decirse que, si el transporte es ineficiente, entonces habrán pérdidas económicas asociadas a los retrasos en entregas de mercancías [2], por ejemplo en una empresa ubicada en el centro de una ciudad, la entrada y salida de mercancías se verá entorpecida si la ciudad padece de un sistema ineficiente de transporte. La concentración de población y actividad económica en corredores bien definidos lleva a la justificación económica de un sistema masivo de transporte de alta calidad que no podría ser viable en un área más dispersa [1]. Esto se observa en la ciudad de México, cuyas áreas de mayor actividad económica tienen sistemas de transporte masivo como el metro, metrobús y rutas de minibuses; caso contrario a zonas periféricas, que carecen de transporte público eficiente.

En el caso del factor humano, los efectos se reflejan en la psicología de sus usuarios [3]: habiendo pasajeros que viven a decenas de kilómetros de sus centros de estudio y trabajo, es fundamental acortar los tiempos de viaje y hacerlos lo más cómodos posibles para evitar factores como el cansancio y la ansiedad aparejados con esta situación. Dicha situación es especialmente grave en los estratos sociales más desfavorecidos, ya que suelen encontrarse

en lugares más inaccesibles. Por ejemplo, en la Ciudad de México y su área metropolitana, los asentamientos ubicados en la Sierra de Guadalupe o en la zona de Chalco dan cuenta de ello, pues los usuarios tienen que tomar hasta 3 sistemas de transporte para llegar a sus centros de estudio o trabajo, o realizar recorridos de hasta dos horas y media por viaje, por lo que gastan 5 horas del día sólo en el transporte [24]. Esto es más del 25% de las horas del día; si se consideran jornadas laborales de, en promedio 9 horas. Así, hay personas que sólo disponen de 10 horas para dormir y realizar otras actividades que les permitan un desarrollo pleno como personas.

La solución de estos problemas precisa de un mantenimiento continuo de la infraestructura de los diversos sistemas de transporte, pero esta solución no se limita solo a eso. Se ha constatado que aumentar las vialidades no soluciona a largo plazo los embotellamientos y el aumento de unidades de transporte tampoco lo hace, tal como Carlos Gershenson ejemplifica en sus investigaciones recientes [4]. Por ello, es importante complementar la búsqueda de soluciones con las propiedades del sistema que puedan encontrarse con enfoques como el de la ciencia de redes.

1.4. Descripción de sistemas mediante redes

Una gran cantidad de fenómenos naturales y artificiales surgen en sistemas de los que emergen vastas redes complejas [25]. Esta *estructura de red* es consecuencia de que todas estas manifestaciones se encuentran gobernados por los mismos principios organizativos [26] que rigen otros procesos, como sucede en los fenómenos sociales y los tecnológicos. A continuación, se presentan las herramientas que se usarán en este trabajo para determinar las propiedades de nuestro objeto de estudio.

1.4.1. Redes y grafos

Una *red* se puede entender como un sistema cuyos elementos, o *nodos*, tienen una relación entre sí, o *conexiones* [26]. En el mundo hay abundancia de ejemplos: las relaciones sociales, los lenguajes, Internet, las conexiones aeroportuarias o la comunidad mundial de países. Así, en Internet, los servidores son los nodos y las conexiones son las conexiones físicas entre ellos; este ejemplo y algunos otros han sido esquematizadas en la Figura 1.3. A la formalización de las redes en el ámbito matemático se le denomina *Ciencia de Redes*, que encuentra propiedades generales para aplicarlas a cualquier ejemplo en particular.

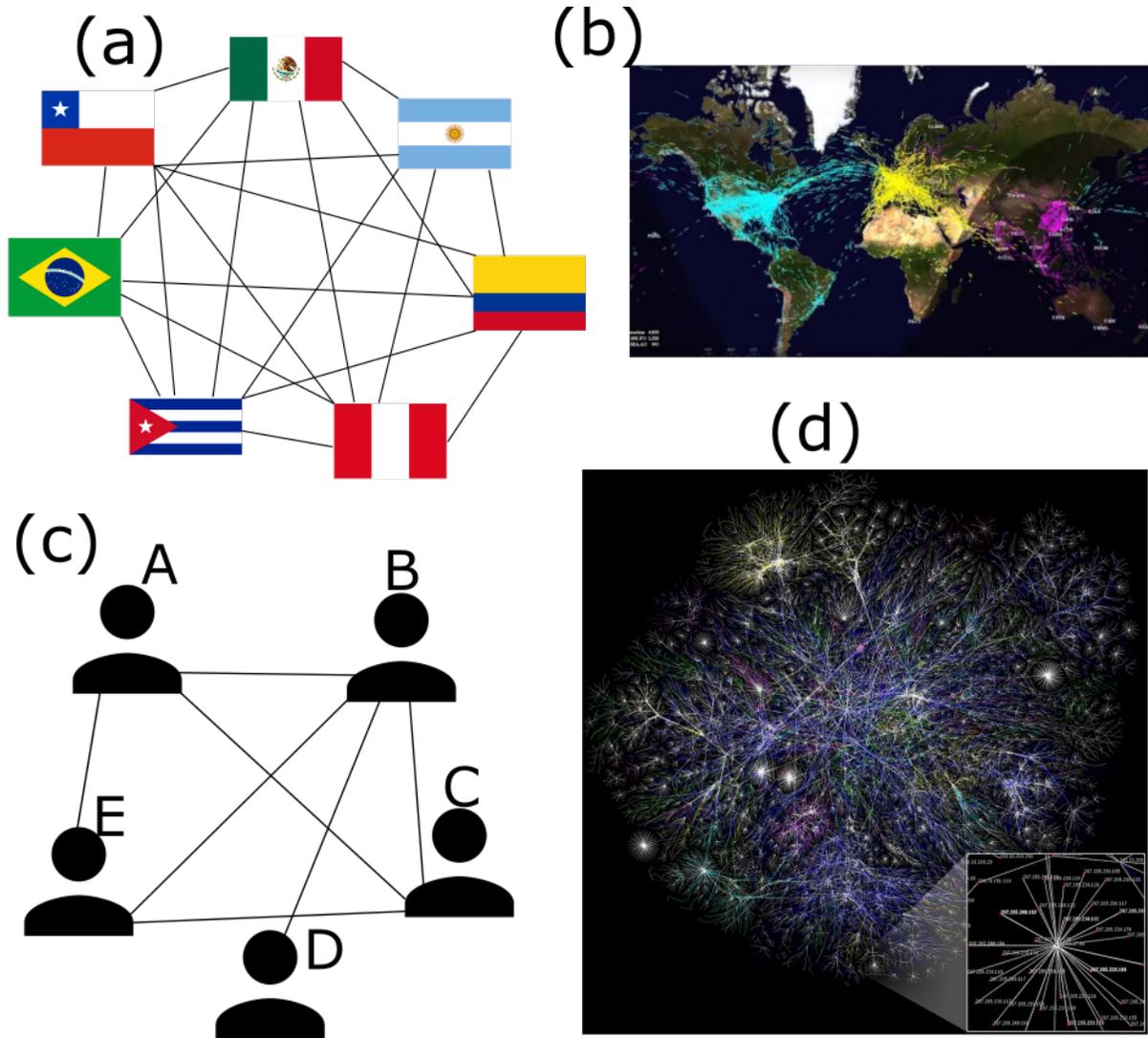


Figura 1.3: Esquematzación de varias redes. (a) Misiones diplomáticas entre algunos países de latinoamérica, obsérvese que en este ejemplo hay reciprocidad total. (b) Conexiones aéreas entre algunos aeropuertos del mundo [27]. (c) Red de amistades entre 5 personas, la persona B tiene relación con las otras 4, en cambio, la persona D sólo tienen relación con una. (d) Mapa de Internet [28].

El andamiaje matemático para el estudio de las redes se basa en dos ramas de las matemáticas: la estadística y la *teoría de grafos* [26]. Esta última se utiliza cuando la representación de la red se hace de manera *gráfica*, en cuyo caso, el grafo de la red se compondrá de *vértices* y *aristas*; que matemáticamente se expresa como un *par* $G = (E, V)$, donde E es el conjunto de aristas y V es el conjunto de vértices. En la actualidad, muchos problemas se resuelven total o parcialmente, con ayuda de la teoría de redes, especialmente aquellos que tienen que ver con las redes humanas o sociales y también en biología, economía e inteligencia artificial. Esto la hace de esta una rama de las matemáticas en pleno

ascenso; pues, allá donde hayan elementos de un grupo que tengan algún tipo de conexión entre sí, siempre habrá espacio para la teoría de redes.

1.4.2. Propiedades de las redes

Las redes pueden ser clasificadas en dos grandes grupos: *dirigidas* y *no-dirigidas*. Las redes no-dirigidas son aquellas en las que los nodos tienen relaciones recíprocas, por lo que las aristas pueden transitarse en ambos sentidos. Ejemplos de este tipo de redes son las amistades en *Facebook* o la red neuronal. En cuanto a las redes dirigidas, se dice que una red lo es si las interacciones entre cada par de nodos tiene dirección. Considérese, por ejemplo, las interacciones entre usuarios de *Twitter* y el de las citas entre autores de artículos científicos.

Una manera compacta de representar una red de N nodos es mediante una matriz de adyacencia, \mathbf{A} , de orden $N \times N$ con elementos 0 y 1. En esta representación, valores no nulos corresponden a una conexión entre nodos y los nulos corresponden al caso contrario. El número de conexiones puede calcularse contando los elementos no nulos de la matriz. para el caso de una red no-dirigida, el grado de un nodo se calcula como la suma de los valores correspondientes a su fila, o su columna si se tienen una red no dirigida, entonces $k_i = \sum_{j=1}^N A_{ij} = \sum_{i=1}^N A_{ij}$. También existen redes en las que las conexiones entre nodos no se restringen a una naturaleza booleana. En este caso, las entradas de \mathbf{A} pueden tener todo un espectro de valores; estas son las *redes pesadas* y las *redes múltiples*. En las primeras, la conexión puede tener asociado un valor de entre 0 y 1, mientras que en el segundo caso, puede darse el caso de más de una conexión entre un par de nodos y, más aún, cualquiera de ellas puede tener un *peso*. A su vez, estos dos casos pueden ser dirigidos o no-dirigidos.

Al número de nodos conectados con un nodo dado i se le denomina simplemente *grado*, k_i . En redes no-dirigidas, se define el *grado promedio de la red*, $\langle k \rangle$, como:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2L}{N}, \quad (1.1)$$

siendo L el número total de líneas en la red.

Se puede definir otra propiedad de los nodos de una red, llamada *coeficiente de agrupamiento*, que cuantifica la conectividad local de una red en torno a un nodo i , que se calcula como:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}, \quad (1.2)$$

donde L_i es el número de conexiones entre los vecinos del nodo i . Si se considera a los k_i vecinos de i como una subred aislada, vemos que el coeficiente es la razón entre el número de enlaces de esta subred y el número máximo de enlaces, con lo que queda sustentada la afirmación anterior. Este concepto también puede extenderse a toda la red definiendo el *coeficiente de agrupamiento promedio* $\langle C \rangle$, que se calcula como:

$$C = \frac{1}{N} \sum_{i=1}^N C_i, \quad (1.3)$$

Este coeficiente indica, por extensión del coeficiente para cada nodo, la probabilidad de que dos nodos vecinos escogidos al azar estén conectados. En toda red existe el concepto de *camino* que se refiere a la distancia, d , que se recorre para ir de un nodo i a un nodo j . Si d es finita para todo par de nodos i y j entonces se dice que la red es conexa, si $d = \infty$ entonces se tendrá una red inconexa. Normalmente, existe más de un camino para ir de un nodo a otro. Al camino más corto se le denomina *camino más corto* y a d_{max} se le denomina *diámetro de la red*. También se tiene la *distancia promedio*, $\langle d \rangle$, que se calcula como:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i=1, i \neq j}^N d_{ij}. \quad (1.4)$$

Finalmente, es importante mencionar que las redes pueden tener la propiedad de ser de *mundo pequeño* y *libres de escala*. Se dice que una red es de mundo pequeño si al escoger dos nodos al azar de la red, la distancia entre estos es mucho menor que el diámetro de la misma. Así, suele decirse que las interacciones sociales son de mundo pequeño, con el famoso dicho de los *seis grados de separación*, pues se dice que entre dos personas cualesquiera el número promedio de intermediarios es 5 [26]. Esta conclusión está basada en el experimento de Stanley Milgram [29], aunque este experimento famoso ha tenido algunos detractores.

1.4.3. Comunidades

Una red, o un grafo, se denomina *completa* si el número de sus conexiones, o vértices, es igual $L_{max} = \frac{N(N-1)}{2}$, por lo que su correspondiente matriz adjunta estará llena de valores no nulos, salvo aquellos de la diagonal, siempre que los elementos no interactúen consigo mismos. En caso de que no sea completa, puede suceder que haya subconjuntos de elementos que interactúan mucho entre sí con respecto a aquellos elementos que no están en tales subconjuntos. En este caso, se dirá que se ha formado una *comunidad*, pues

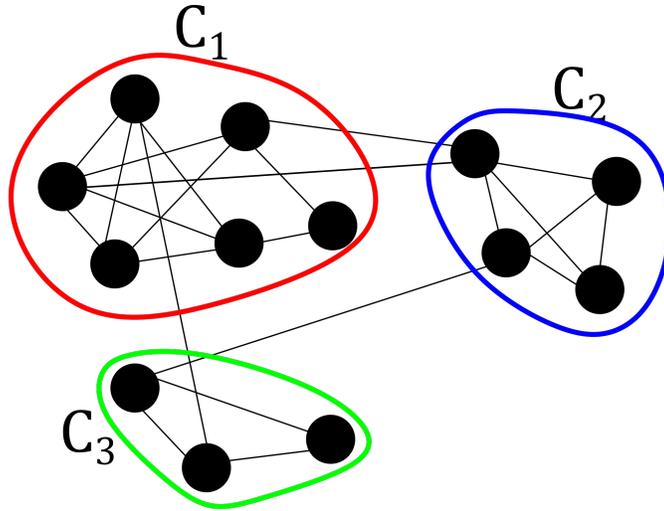


Figura 1.4: Ejemplo de la formación de comunidades en una red, las cuales han sido encerradas con curvas de distinto color. En ellas puede observarse que sólo algunos elementos de cada comunidad tienen comunicación con los elementos de otras comunidades, mientras que las conexiones dentro de cada comunidad son abundantes.

esta situación coincide con la definición intuitiva de comunidad: un grupo de elementos o individuos que tienen características o tendencias similares. Como recurso ilustrativo, en la Figura 1.4 se observa una red con las comunidades que hay en ella. Encontrar comunidades haciendo uso de la matriz de adyacencia correspondiente puede ser sencillo, si esta tiene una estructura de bloques definida, pero en el caso en que esto no sea tan claro, es muy útil tener criterios que eviten la determinación de comunidades de manera arbitraria o discrecional. La detección computacional de comunidades en redes es importante porque, usando criterios objetivos previamente establecidos, encuentra similitudes o tendencias dentro de conjuntos con un número muy grande de elementos, como los habitantes de un país, la cadena alimenticia de una zona geográfica, las llamadas telefónicas o las unidades de transporte en una ciudad, tarea que de otra manera sería imposible de llevar a la práctica.

Ahora bien, considérese un subgrafo C de N_C nodos dentro de una red. El *grado externo*, k_i^{ext} , es el número de conexiones de ese nodo i con el resto de la red, mientras que *grado interno*, k_i^{int} , es el número de conexiones de i con otros nodos contenidos en C . Si $k_i^{ext} = 0$, entonces todos los vecinos de i están dentro de C y la comunidad es adecuada para i . Si $k_i^{int} = 0$, quiere decir que todos los vecinos de i están fuera de C , e i debe ser asignado a otra comunidad. Lo anterior, ofrece un parámetro para determinar si una comunidad es *fuerte* o *débil*. Así, se tendrá que una comunidad es fuerte sí [26]:

$$\forall i \in C k_i^{int}(C) > k_i^{ext}(C) \quad (1.5)$$

y una comunidad débil si:

$$\sum_{i \in C} k_i^{int}(C) > \sum_{i \in C} k_i^{ext}(C), \quad (1.6)$$

esto es, si el grado interno total es mayor que el grado externo total. Con el fin de buscar de manera eficiente comunidades dentro de una red, hay que definir los conceptos de *bisección del grafo* y de *tamaño mínimo de corte* [26]. El primer término se refiere, simplemente, a partir en dos comunidades a un grafo de manera que no tengan elementos en común. En cuanto al tamaño mínimo de corte, la bisección debe hacerse de manera tal que el número de conexiones entre las dos comunidades sea mínimo. Ahora bien, el número de combinaciones posibles que cumplen esta condición es [26]:

$$B \simeq \frac{N^{N+\frac{1}{2}}}{N_1^{N_1+\frac{1}{2}} N_2^{N_2+\frac{1}{2}}}, \quad (1.7)$$

donde N es el número de nodos de la red y N_1 y N_2 son los nodos en la comunidad 1 y 2, respectivamente. Si consideramos el caso en que $N_1 = N_2$, entonces, la ecuación 1.7 se reduce a

$$B \simeq \frac{2^{N+1}}{\sqrt{N}} = \exp \left\{ (N+1) \log 2 - \frac{1}{2} \log N \right\}, \quad (1.8)$$

con lo que se ve que el número de combinaciones crece de forma exponencial con cada aumento de nodos. Así, para una red de 10 nodos en las que las comunidades 1 y 2 tengan el mismo tamaño y con un tiempo de cálculo de 1 milisegundo por inspección, tardaríamos 10^{16} años en encontrar las comunidades correspondientes. Este ejemplo muy sencillo es claro para entender que son necesarios algoritmos eficaces para encontrar comunidades usando el poder de las computadoras. A continuación se describen algunos de los algoritmos usados para la detección de comunidades.

- **Algoritmo de agrupamiento jerárquico.** Este procedimiento ayuda a conseguir tiempos de ejecución algorítmica que crecen polinomialmente con cada nodo N . El método inicia con una *matriz de similaridad* X cuyos elementos x_{ij} representan la similitud entre los nodos i y j [26]. Esta distancia puede calcularse por la *similitud coseno*, el *índice Jaccard* y la *distancia Hamming* [30]. Una vez que las distancias para cada par de nodos ha sido calculada, el algoritmo agrupa aquellos nodos con alta similitud; este último paso puede hacerse de diferentes maneras: por medio de *algoritmos aglomerativos* o con *algoritmos divisivos*. Entre los algoritmos aglomerativos está el *Algoritmo de Ravasz* y entre los divisivos, el *Algoritmo de Girvan-Newman*.

- **Algoritmo de Ravasz.** Se lleva a cabo en cuatro pasos. En el primero se calcula la

matriz de traslapamiento topológico, cuyos coeficientes x_{ij}^o se calculan a partir de

$$x_{ij}^o = \frac{J(i, j)}{\min(k_i, k_j) + 1 - \Theta(A_{ij})}, \quad (1.9)$$

donde $\Theta(x)$ es la función escalón de Heaviside, $J(i, j)$ es el número de vecinos en común de los nodos i y j ; a este valor se le suma un uno si existe una conexión entre los nodos; $\min(k_i, k_j)$ es el más pequeño de los grados de dichos nodos. Entonces, se tendrá que: $x_{ij}^o = 1$ si los nodos están conectados y tienen los mismos vecinos y $x_{ij}^o = 0$ en caso contrario. El siguiente paso es determinar qué tan similares son las comunidades que se han obtenido, ello se hace calculando el promedio de x_{ij} sobre todos los pares de nodos que pertenecen a comunidades distintas. El tercer paso es aplicar el agrupamiento jerárquico. Por último, se usa un *dendograma* para extraer la organización de comunidad que subyace en este conjunto.

• **Algoritmo de Girvan-Newmann.** Inventado en 2001 por M. Girvan y M. E. J. Newman, el algoritmo de Girvan-Newmann se lleva a cabo en cuatro pasos [31]:

- 1. Calcular la *betweenness* para todos los vértices en la red; este es el término con el que se le denomina a los elementos x_{ij} .
- 2. Eliminar el vértice con la *betweenness* más pequeña.
- 3. Recalcular la *betweenness* para todos los vértices afectados.
- 4. Repetir desde el paso 2 hasta que no queden sobren vértices.

En el artículo original [31], la *betweenness* se calcula con el algoritmo de Newman, aunque esto no es obligatorio. Este algoritmo toma un tiempo de orden $O(m^2n)$. Este método demuestra una eficacia superior a métodos como el de *clustering jerárquico*, aunque su efectividad se limita a redes con menos de unos cuantos miles de nodos.

La maximización de la *modularidad* es uno de los métodos más usados [32]. En este método, la modularidad se refiere a qué tan buena es la división de una red en comunidades. Para ello, se toma una red de N nodos y L enlaces, en la que cada comunidad tiene N_c nodos. Entonces, se mide la diferencia M_c entre el diagrama real de enlaces de la red y el número esperado de enlaces entre los nodos i y j si los enlaces de la red se construyen aleatoriamente. Así, la modularidad se calcula como [26]:

$$M_c = \frac{1}{2L} \sum_{(i,j) \in C_c} (A_{ij} - p_{ij}), \quad (1.10)$$

donde p_{ij} suele expresarse como $p_{ij} = \frac{k_i k_j}{2L}$. Los valores de la modularidad llevan a 3 posibles casos: 1. si $M_c > 0$ entonces el subgrafo C_c tiene más conexiones que las esperadas por azar; 2. para $M_c = 0$ la conectividad entre los N_c nodos es aleatoria; 3. si $M_c < 0$ entonces los nodos de C_c no forman una comunidad.

El *método de Louvain*, llamado así por el lugar en el que se creó, fue ideado por Vincent Blondel y sus estudiantes en 2007 [33]. Este método se usa en redes pesadas y propone la optimización de la modularidad, que se calcula como:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j). \quad (1.11)$$

Se observa que la ecuación 1.11 es muy similar a la ecuación 1.10, en donde se reemplaza a L por m , que es la suma de todos los pesos de las conexiones de la red y se calcula como $m = \frac{1}{2} \sum_{i,j} A_{ij}$, y k_i es el grado de i pero considerando los pesos de las conexiones. El método en sí, consta de dos fases que se repiten de manera iterativa. Primero, se asigna una comunidad diferente a cada nodo de la red, por lo que en este paso hay tantas comunidades como nodos en la red. Entonces, se consideran los vecinos j del nodo i y se evalúa la ganancia en modularidad ΔQ que habría si i se extrajera de su comunidad y se colocara en la de j . De esta manera el nodo se queda en la comunidad donde la ganancia de modularidad sea mayor siempre que esta ganancia sea positiva. Si no existe ganancia positiva i se queda en su comunidad, el cálculo se hace como sigue:

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right]. \quad (1.12)$$

En la segunda fase del método, se construye una nueva red, donde los nodos son las comunidades obtenidas en el paso anterior. Las conexiones entre nodos de la misma comunidad se convierten en *auto-conexiones* y las conexiones a otras comunidades se condensan en una sola conexión cuyo peso se determina con el número de conexiones anteriores. En esta fase se crea una nueva red, y el proceso se repite tantas veces como se considere necesario.

1.5. Divergencia de Kullback-Leibler

También referida como *distancia* de Kullback-Leibler, fue introducida por Solomon Kullback y Richard Leibler en 1951; y suele interpretarse como la diferencia entre una distribución P de probabilidad y otra Q [34]. Hoy en día, esta herramienta es ampliamente usada

para la comparación de bases de datos [35]. Su expresión para distribuciones discretas es

$$D_{KL}(P||Q) = \sum_{i=1}^N P(x_i) \log \frac{P(x_i)}{Q(x_i)}. \quad (1.13)$$

Si estas distribuciones de valores fueran *continuas*, la suma cambia por una integral, y entonces:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} P(x) \log \frac{P(x)}{Q(x)} dx. \quad (1.14)$$

Es importante hacer énfasis en que esta no es una distancia en sentido de una *métrica*, ya que la distancia entre P y Q no necesariamente es la misma que entre Q y P , sin embargo, es común encontrar en la literatura la referencia a esta medida como *distancia de Kullback-Leibler*. También, de estas ecuaciones, es claro ver que D_{KL} nunca es negativa, y es igual a cero cuando $P = Q$. A pesar de la brevedad de esta sección, se resalta aquí su importancia en la presente tesis, pues esta herramienta, en apariencia sencilla, será usada para encontrar una estructura de red que subyace al sistema del metrobús, analizando la distribución de velocidades, en un enfoque novedoso para comprender las características de este sistema de transporte.

Capítulo 2

Sistema Metrobús CDMX

2.1. Introducción

La Ciudad de México tiene una población estimada de 8.9 millones de residentes [36] pero, considerando a la *población flotante*, la cifra aumenta a 10.6 millones de personas [37]. Esta enorme población flotante proviene en su mayoría de la Zona Metropolitana del Valle de México, de la cual, el 48.8% carece de transporte privado por lo que tiene que hacer uso del transporte público [38]. Parte del transporte público de la Ciudad de México se realiza a través del transporte masivo que, como su nombre lo indica, tiene la capacidad de movilizar a cientos o miles de pasajeros por viaje; el transporte público de la zona metropolitana se compone del Metro, el tren ligero, el tren suburbano y el autobús de tránsito rápido.

Como se ha mencionado anteriormente, esta tesis se ocupa únicamente del estudio del autobús de tránsito rápido. Para ello se incluye aquí una descripción detallada del sistema de autobús de tránsito rápido de la Ciudad de México, llamado Metrobús, con la secuencia que a continuación se escribe. Primero, se explica el concepto de autobús de tránsito rápido como sistema de transporte, así como sus orígenes y algunas de sus primeras implementaciones en el mundo. Posteriormente, se hace una comparación con otros sistemas de transporte masivo, a saber: sistemas tipo metro y trenes ligeros. Esta comparación incluye los costos de construcción, operación y mantenimiento, la capacidad de movilización y la relación costo-beneficio. Se prosigue con la descripción detallada de su aplicación en la Ciudad de México, así como de su infraestructura. El capítulo concluye con una breve exposición de los problemas a los que se enfrenta el metrobús, así como de

sus perspectivas a futuro.

2.2. Autobús de tránsito rápido

El sistema tipo Autobús de Tránsito Rápido, frecuentemente encontrado en la literatura como *BRT* (por sus siglas en inglés *Bus Rapid Transit*), es una forma de transporte masivo que se implementa en grandes ciudades como alternativa a sistemas tipo metro. Este sistema de transporte consiste en autobuses urbanos que transitan únicamente por carriles exclusivos en calles y avenidas ya existentes [39], por tal motivo, su infraestructura sólo necesita la construcción de estaciones de ascenso y descenso de pasaje y la instalación de un sistema de cobro, que consiste en un sistema de lectores electrónicos de tarjetas creadas para tal propósito. Este sistema se usó por primera vez en Ottawa en la década de 1970 [40] bajo el nombre de *Transitway* y en Curitiba, Brasil, bajo el nombre de *Red Integrada de Transporte*, donde su primera línea de autobús entró en operación en 1973 y se estableció propiamente como un sistema completo en 1982 [41]. En el caso de México, se implementó por primera vez en León, Guanajuato en 2003 [42] por medio del sistema *Optibús*. Hoy en día, 12 estados de la república han adoptado este medio de transporte, información que está resumida en la Tabla 2.1.

Tabla 2.1: Sistemas de autobús de tránsito rápido en México, enlistados por orden de inauguración [43].

Ciudad	Nombre del sistema	Año de inauguración	Líneas
León, Guanajuato	Optibús	2003	10
Ciudad de México	Metrobús	2005	7
Villahermosa, Tabasco	Transbús	2008	1
Guadalajara, Jalisco	Macrobús	2009	1
Estado de México*	Mexibús	2010	3
Puebla, Puebla	UTA	2013	3
Chihuahua, Chihuahua	Vivebús	2013	1
Ciudad Juárez, Chihuahua	Vivebús	2013	1
Monterrey, Nuevo León	Ecovía	2014	1
Pachuca, Hidalgo	Tuzobús	2015	1
Acapulco, Guerrero	Acabús	2016	1
Tijuana, Baja California	SITT	2016	1
Querétaro, Querétaro	Qrobús	2017	2

*A lo largo de la zona metropolitana del valle de México.

Hoy en día se pueden encontrar estudios sobre los impactos positivos asociados a la construcción un sistema de autobús de tránsito rápido en las zonas circundantes a este. Por ejemplo: propicia un ambiente amigable para el peatón, revitaliza las zonas próximas a las estaciones y eleva el precio de los inmuebles que las rodean [44]. Aunado a lo anterior, las políticas actuales de implementación de este sistema contemplan la reducción de emisión de gases de efecto invernadero [45], dada la tecnología que aplican en la construcción de sus unidades, a lo que se agrega que un solo autobús biarticulado llega a equivaler a 100 automóviles. Esto implica no sólo una reducción de esas emisiones, sino que contribuye al desahogo de los embotellamientos vehiculares [46], un problema muy serio en todas las grandes ciudades del mundo. Impulsar diferentes alternativas de transporte público coadyuva a la eliminación de la llamada *cultura del coche*: el hecho de destinar espacio público a avenidas y estacionamientos para los automóviles. Si una ciudad cuenta con muchas alternativas de transporte, puede pensarse incluso en una forma sana de movilidad, al combinar transporte en autobús urbano más bicicleta o caminar las cortas distancias que separen a una estación del lugar de destino de un pasajero, esto es, articular el transporte masivo con el individual en la cuestión de la *última milla* [47].

2.3. Comparación con otros sistemas de transporte

Cuando una ciudad sobrepasa un cierto umbral de habitantes y extensión, así como de actividad económica, es normal que se plantee la adopción de un sistema de transporte que movilice a una gran cantidad de pasajeros. Esto suele manifestarse cuando el transporte basado en rutas de minibuses se ve rebasado por una demanda que crece con rapidez [2]. Entre los sistemas que se plantean están: el *metro*, el *tren ligero* y el *autobús de tránsito rápido* [47]. Los primeros necesitan la construcción de vías exclusivas para el tránsito de sus vehículos e instalaciones eléctricas de alta potencia para movilizar a un número relativamente grande de pasajeros [11]. Con respecto a este último punto, la alimentación eléctrica de los vehículos puede ser aérea, por medio de un mecanismo denominado *pantógrafo*, o estar integrada en las vías mismas, como ocurre en casi todas las líneas del *metro de la Ciudad de México* [48]. De estos tres sistemas, el que tiene mayor capacidad de transporte es el metro, seguido por el tren ligero, por tanto, el autobús de tránsito rápido es el de menor capacidad [11].

En la decisión de implementar alguno de estos sistemas tiene un peso decisivo la relación *costo-beneficio*, ya que se busca optimizar el número de pasajeros que son movilizados en

relación al costo de la infraestructura necesaria en tales sistemas, dado el presupuesto con el que se cuenta para ello. Para el caso del metro, en un primer acercamiento cuantitativo, la experiencia nos muestra que en este sistema es común la construcción subterránea tanto de las vías como de las estaciones, razón por la cual, en algunos países se le denomina *subte* o *underground*; esta primera consideración generalmente eleva los costos de construcción con respecto a una alternativa en la superficie. También se observa que, en ocasiones, tramos de este sistema son de *viaducto elevado*, como se observa en la Figura 2.1(a), así que, similarmente al caso anterior, se intuye un costo elevado en tal opción. En segundo lugar, se observan diferencias entre los vehículos usados en estos tres sistemas de transporte. El metro utiliza carros con una longitud promedio de 21 metros [11], de rodadura férrea o neumática, los cuales transitan unidos formando largos convoyes de hasta 10 carros. El tamaño y la constitución más robusta de sus unidades nos dan una idea de un costo superior por carro que el de los autobuses que se usan en los sistemas de autobús de tránsito rápido. Un razonamiento similar se haría en el caso del tren ligero, con la diferencia de la longitud del convoy, que es menor que en un metro. Por último, se puede contemplar el costo de operación y mantenimiento; en ambos casos, se hace mantenimiento de las unidades y de las estaciones, pero en el caso del metro y el tren ligero se agrega el mantenimiento y costo de operación de las vías exclusivas y de la instalación eléctrica.

En números concretos, se puede encontrar que, en 1990, el costo promedio por kilómetro en Estados Unidos era de 128.2 millones de dólares para el metro, 26.4 millones de dólares para un tren ligero y 10.2 millones para un autobús de tránsito rápido [11]. En el caso europeo [49], el costo promedio por kilómetro en 2004 fue de 130 millones de euros para un metro y 15 millones de euros para un tren ligero y un promedio de 10 millones de euros para el autobús de tránsito rápido en la década de 2010 [50]. En el caso mexicano se puede citar la construcción de la línea 12 del metro, que tuvo una inversión inicial de casi 18,000 millones de pesos [51], aunque al final terminó costando casi 25,000 millones de pesos [52], un aproximado de 1,000 millones de pesos por kilómetro o, aproximadamente, 77 millones de dólares por kilómetro. En contraste, el metrobús tuvo un costo promedio [42] de 2.8 millones de dólares por kilómetro. En una última revisión, se enlistan los costos que algunas ciudades de Latinoamérica hicieron para tener un sistema de autobús de tránsito rápido [42]: Quito, 1 millón de dólares por kilómetro; Guayaquil, 2 millones por kilómetro; Guadalajara: 3.5 millones por kilómetro y Bogotá: 12.5 millones por kilómetro. Resumiendo, en Estados Unidos, construir un sistema de autobús de tránsito rápido es casi 13 veces más barato que un metro y casi 3 veces más barato que un tren ligero; en Europa es 13 veces más barato que un metro y un 33% más barato que un tren ligero; y



Figura 2.1: Comparación de sistemas de transporte masivo. Obsérvese la infraestructura de ambos sistemas (a) Metro de la Ciudad de México, Fuente: [55]; (b) Metrobús de la Ciudad de México, Fuente: [56]

en México es 30 veces más barato.

A continuación, se considera la movilidad de pasajeros tomando como ejemplo la Ciudad de México. Las siete líneas del autobús de tránsito rápido en la Ciudad de México atienden a 1.2 millones de pasajeros al día [53], en unidades y estaciones como las mostradas en la Figura 2.1(b), mientras que las doce líneas del metro transportan a 4.9 millones de pasajeros diariamente [54]. De esta manera, el metro transporta un número 4 veces mayor de personas además de tener una extensión casi dos veces mayor (140 km para el metrobús en comparación con 226 km del metro). Esto permite calcular que movilizar a 4.8 millones de pasajeros costaría 11.2 millones de dólares por kilómetro, un costo 7 veces menor que el requerido para movilizar a la misma cantidad de personas en el metro. Con este ejemplo sencillo, se puede ver claramente la relación costo-beneficio, y la ventaja a la hora de considerar la implementación de un sistema de autobús de tránsito rápido.

2.4. Características del sistema Metrobús

La red de Metrobús se inauguró el 19 de junio de 2005 [57], como una estrategia para diversificar la oferta de transporte público en la capital del país. El conteo en un mapa actualizado de la red del sistema, permite saber que la red consta de 7 líneas, cada una con un color distintivo, en las cuales se distribuyen 238 estaciones, cada una con un ícono

distintivo similar al usado en el metro. Toda esta información se encuentra ilustrada y condensada en la Figura 2.2. En cuanto a su distribución geográfica, el metrobús se extiende íntegramente en el territorio capitalino, a diferencia del metro que tiene algunas estaciones en el Estado de México. En este sentido, vale la pena agregar que sólo la estación Indios Verdes, tanto de la línea 1 como de la línea 7, tiene conexión con el sistema de autobús de tránsito rápido del Estado de México, en su línea 3.

El metrobús opera bajo el esquema de licitación a empresas privadas [53], encargadas de operar tanto las unidades que brindan el servicio como el cobro del pasaje, mientras que el gobierno de la ciudad se encarga de administración, planeación y control de sistema por medio de un organismo público. Esto último lo diferencia del Metro, el Tren Ligero y el sistema de Trolebuses, sistemas totalmente operados por el estado. Por último, es importante destacar el beneficio ambiental, pues tan sólo en 2016 el sistema Metrobús evitó la emisión de 143,952 toneladas de dióxido de carbono [58], una mejoría con respecto a las 15,804 toneladas que dejaron de emitirse en 2005 cuando el Metrobús comenzó a operar.

2.4.1. Líneas del sistema Metrobús

Las 7 líneas atienden a 10 de las 16 alcaldías de la Ciudad de México, a saber: Gustavo A. Madero, Venustiano Carranza, Cuauhtémoc, Azcapotzalco, Miguel Hidalgo, Iztacalco, Iztapalapa, Benito Juárez, Coyoacán y Tlalpan. Una de las razones por las que el resto de las demarcaciones no tienen el servicio es debido a las dificultades orográficas para implementar el servicio. Por ejemplo, la alcaldía Magdalena Contreras tiene un terreno altamente montañoso que dificulta la introducción del Metrobús en sus calles estrechas; este es un patrón que se repite con los otros dos sistemas de transporte masivo, el Metro y el Tren Ligero. Condiciones similares se presentan en las delegaciones Álvaro Obregón, Cuajimalpa y Milpa Alta y, de hecho, en estas demarcaciones se plantea la introducción de un teleférico [59]

Con respecto a las características de las líneas, la línea 1 tiene 46 estaciones; 37 de paso, 7 de transbordo y 2 terminales (Indios Verdes y El Caminero). Su longitud es de 30 kilómetros, la cual discurre completamente en la Av. de los Insurgentes, en dirección norte-sur. Sus trabajos de construcción iniciaron en 2005 y terminaron en Marzo de 2008. La demanda promedio es de 480 mil pasajeros por día. Su color distintivo es el rojo. Tiene las rutas Indios Verdes - El Caminero, Indios Verdes - Insurgentes, Indios Verdes - Dr.

en septiembre de 2007 y finalizaron en diciembre de 2008. La demanda estimada promedio es de 180 mil pasajeros por día. Su color distintivo es el morado. Tiene los itinerarios Tepalcates - Tacubaya, Tepalcates - Etiopía, Tepalcates - Colonia del Valle (Línea 1).

La línea 3 tiene 33 estaciones; 25 de paso, 7 de transbordo y 2 terminales (Tenayuca y Etiopía), siendo Etiopía estación terminal y de transbordo. Esta línea tiene una longitud de 17 kilómetros que yacen a lo largo del eje poniente 1 y de las vialidades: Calzada Vallejo, Calle Guerrero y Balderas. Su construcción se desarrolló entre marzo de 2010 y febrero de 2011 y la afluencia estimada es de 155 mil pasajeros por día. Su color distintivo es el verde. Cuenta con las rutas Tenayuca - Etiopía, Tenayuca - Balderas, Tenayuca - La Raza y Tenayuca - Buenavista.

La línea 4 tiene 35 estaciones; 32 de paso y 3 terminales (Buenavista, Terminal 1 y Terminal 2). Esta es la línea más compleja del sistema, pues además de estar dividida en 2 ramales, denominadas “ruta norte” y “ruta sur”, en pocas estaciones coinciden las direcciones de los vehículos. Aunado a lo anterior, es la que más transbordos tiene. Su longitud total es de 28 kilómetros que yacen a lo largo de las siguientes vialidades: Eje 1 Norte (Mosqueta), Juan Cuamatzin, Eje 1 Oriente (Anillo de Circunvalación), Eje 2 Oriente (Congreso de la Unión), Eje 3 Oriente (Ing. Eduardo Mallida), para la Ruta sur; y Puente de Alvarado, República de Venezuela, Héroes de Nacozari, General Miguel Alemán, Eje 3 Oriente, Fuerza Aérea Mexicana y Circuito Interno del Aeropuerto; para la Ruta Norte. El período de construcción estuvo comprendido entre julio de 2011 y abril de 2012. Es usada por, aproximadamente, 65 mil pasajeros por día. Su color distintivo es el naranja. Tiene las rutas San Lázaro - T1 y T2, Buenavista - San Lázaro (Ruta Sur), Buenavista - San Lázaro (Ruta Norte).

La línea 5 tiene 18 estaciones, 16 de paso, 2 transbordo y 2 terminales; Río de los Remedios y San Lázaro, siendo esta última también una estación de transbordo. Río de los Remedios, 314 / Memorial New's Divine, 5 de Mayo, Vasco de Quiroga, El Coyol, Preparatoria 3, San Juan de Aragón, Río Guadalupe, Talismán, Victoria, Oriente 101, Río Santa Coleta, Consulado, Canal del Norte, Deportivo Eduardo Molina, Mercado Morelos, Archivo General de la Nación y San Lázaro. Tiene una longitud total de 10 kilómetros contenidos completamente en el Eje 3 Oriente “Ing. Eduardo Molina”, en dirección nortesur. Su construcción inició en marzo de 2013 y terminó en noviembre del mismo año. Tiene una demanda estimada de 70 mil pasajeros por día. Su color distintivo es el azul. Río de los Remedios - San Lázaro es su única ruta.

La línea 6 del metrobús tiene 38 estaciones, 30 de paso, 6 de transbordo y 2 terminales,

El Rosario y Villa de Aragón. Su longitud es de 20 kilómetros, cuyo trazado yace sobre el Eje 5 Norte (Av. Montevideo), en dirección poniente-oriente. Su construcción empezó en agosto de 2014 y terminó en octubre del mismo año. La demanda estimada de esta línea es de 150 mil pasajeros por día. Su color distintivo es el rosa. Sus rutas son: Villa de Aragón - El Rosario, IPN - El Rosario y 18 de Marzo - El Rosario.

La línea 7 del metrobús tiene 31 estaciones, 19 estaciones de paso, 10 de transbordo y 2 terminales: Campo Marte e Indios Verdes. Tiene una longitud de 15 kilómetros que discurren completamente sobre La Avenida Paseo de la Reforma. Su construcción inició en enero de 2017 y finalizó en marzo de 2018. Aún no se han dado a conocer datos oficiales sobre su afluencia. Su color distintivo es el verde bandera. Sus dos rutas son: Indios Verdes - Campo Marte e Indios Verdes - Hospital Infantil La Villa.

2.4.2. Estaciones

Las estaciones son los puntos por los cuales se accede al sistema. Puede hacerse una clasificación de las estaciones en 4 tipos: primero, aquellas en las que los autobuses en ambas direcciones cargan y descargan pasaje en el mismo recinto de la estación y son construidas dentro del camellón de las avenidas por donde transitan los autobuses; este formato de estación es el predominante en las estaciones de las líneas 1, 2, 3, 5 y 6. Luego, están las estaciones en las que las instalaciones de ascenso y descenso se han construido al centro de las avenidas, pero que cuentan con instalaciones individuales para cada dirección. Un ejemplo es la estación Deportivo 18 de Marzo (Línea 1), cuyas instalaciones están separadas por la línea 3 del Metro.

Por otra parte, también están aquellas estaciones en las que los puntos de acceso se ubican sobre las banquetas de las vialidades, por lo que las instalaciones para el ascenso y descenso están al aire libre, esto es, no tienen torniquetes de acceso ni paredes de ningún tipo y sólo cuentan con una *sombra* para proteger a los pasajeros que esperan el autobús; en tales estaciones suele pasar que los puntos de acceso para una dirección y otra ni siquiera coincidan geográficamente, pudiendo estar separadas por varios metros. Este es el tipo de estación que predomina en las líneas 4 y 7. Por último, aquellas estaciones que son sólo de una dirección; esto pasa porque que la línea se bifurca y las direcciones transitan por vialidades distintas para unirse en un punto posterior. Por poner un ejemplo, después de la estación “Constitución de Apatzingán” de la línea 2, dirección Tepalcates, el flujo vehicular se divide en dos: con el flujo de ida discurriendo por Anillo Periférico y la Calzada

Ignacio Zaragoza y el flujo de vuelta yace sobre la Avenida Constitución de Apatzingán. Ambos flujos se vuelven a unificar en la estación terminal Tepalcates. En el primer caso, hay dos estaciones intermedias y en el segundo sólo una. Comportamientos similares se encuentran en las líneas 4, 6 y 7.

Todas las estaciones cuentan con elementos que facilitan la accesibilidad para las personas con capacidades diferentes y todas ellas cuentan con información escrita en sistema *Braille* y con líneas guía en el piso. También se cuenta con un botón que avisa el tiempo de llegada del siguiente autobús a la estación. Para personas con capacidades motoras diferentes, cada estación tiene rampas de acceso.

2.4.3. Parque vehicular

El parque vehicular está compuesto por 660 unidades de transporte, repartidas en los siguientes tipos: 303 autobuses articulados que tienen 18 metros de largo y una capacidad de 160 pasajeros, 124 autobuses biarticulados que tienen 24 metros de largo y una capaci-



Figura 2.3: Fotografías de los cuatro tipos de autobuses usados en la red de metrobús de la Ciudad de México. Fuente: [53]

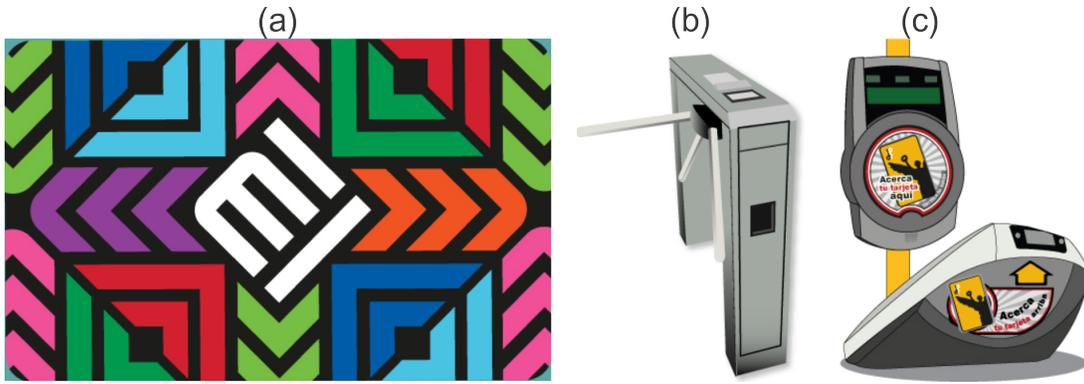


Figura 2.4: Sistema de cobro del metrobús. (a) Ejemplo de una tarjeta de acceso al metrobús. (b) Torniquetes de acceso que se activan por medio de una tarjeta válida. (c) Sensores de cobro usados únicamente en las unidades de las líneas 4 y 7. Fuente: [61]

dad para 240 personas y 70 autobuses de 12 metros de los cuales 61 funcionan con Diésel y 9 tienen un sistema híbrido; en ambos caben 100 pasajeros. También se cuenta con 90 autobuses de doble piso, con una capacidad de 130 pasajeros por unidad; estos sólo se usan en la línea 7. Las imágenes de estos tipos de unidades se pueden ver en la Figura 2.3. Todas las unidades cuentan con rampas de acceso y espacios designados para sillas de ruedas, también cuentan con asientos reservados para personas gestantes, personas de edad avanzada y personas con dificultades motrices. También cuentan con espacios de uso exclusivo para mujeres y niños menores de 12 años. Dicha exclusividad es permanente y la violación de esta medida se castiga con la expulsión del pasajero del autobús e incluso con una sanción administrativa.

2.4.4. Infraestructura de ingreso al sistema

Para ingresar al sistema es necesario colocar una tarjeta válida, como la mostrada en la Figura 2.4 (a), sobre alguno de los sensores electrónicos que están en los accesos a la estación, ilustrado en la Figura 2.4 (b); o en la entrada de los autobuses, ejemplificado en la Figura 2.4 (c), según la estación de la que se trate. Estos dispositivos descontarán el pasaje del saldo de dicha tarjeta si esta es válida o si tiene el crédito suficiente. Hay dos tarifas en el sistema [61]. La tarifa normal, de 6 pesos, da derecho al usuario de permanecer dentro del sistema el tiempo que desee, dentro del horario de servicio y una tarifa especial de 30 pesos, con la cual, el usuario puede arriivar a las estaciones Terminal 1 y Terminal 2 de la línea 4. Esta tarifa también permite el transbordo gratuito en cualquiera de las intersecciones del sistema. Para las personas de la tercera edad y con capacidades diferentes el acceso es gratuito. Otra facilidad con la que cuenta es que los domingos los

usuarios pueden abordar los autobuses con sus bicicletas.

Todas las estaciones, a excepción de aquellas de la línea 4, cuentan con puntos de recarga de saldo para la tarjeta del Metrobús. Estos puntos cuentan con máquinas expendedoras en las que se pueden vender tarjetas con un costo de 16 pesos, que incluyen un saldo de 6 pesos y se puede recargar saldo hasta un máximo de 120 pesos; la información e instrucciones de estas máquinas se ofrecen al usuario en Español, Inglés o Francés. También se cuenta con una *Red de Recarga Externa*, que consiste en establecimientos comerciales donde puede realizarse la recarga y venta de las tarjetas [61].

2.5. Problemas actuales y perspectivas a futuro

Existen diversos retos a los que se enfrenta el Metrobús. Uno de ellos es la saturación del sistema, pues el número de unidades en servicio es insuficiente ante una demanda cada vez mayor, por lo que es frecuente observar aglomeraciones en las estaciones, especialmente en horas pico, como los horarios laborales y escolares. Otra deficiencia es la formación de convoyes que ralentizan el tránsito de las unidades [4], provocados por períodos relativamente largos en los que dejan de pasar unidades, y que, una vez que pasan se amontonan dos o tres vehículos al mismo tiempo, este no es un problema que se resuelva únicamente con la adquisición de más vehículos, una ampliación de la infraestructura o una mejora de las mismas. Es el servicio mismo el que debe eficientarse para evitar estas molestias al usuario que pueden desalentarlo a usar el transporte público y retornar al uso del automóvil [2].

El sistema metrobús, actualmente, se encuentra en fase de expansión, llegando a zonas en que aún no existe el transporte masivo de pasajeros, por lo que se espera que abarque zonas como Xochimilco [62]. Por ejemplo, el sistema de transporte público de la Ciudad de México no está articulado [63], por lo que es primordial que haya una coordinación efectiva entre este y los demás sistemas de transporte. Para ello es necesario un estudio profundo de la movilidad pública en la Ciudad, en el que además de las técnicas que se mostrarán en capítulos posteriores, se plantee el enfoque de *multicapas*, cuya utilidad ha sido demostrada en otras zonas del mundo [64].

Capítulo 3

Actividad del metrobús: caracterización por segmentos

3.1. Introducción

En este capítulo se presentan los primeros resultados que arrojó el análisis de la información obtenida del servidor de *datos dinámicos* que Metrobús pone a disposición de quien solicite los datos que este sistema genera a lo largo del día. Primeramente, se ofrece una explicación somera de las herramientas informáticas usadas, tanto en la recolección de la información como en su tratamiento y en su análisis posterior. Después, se detalla al lector cómo los datos fueron descargados; con el mismo nivel de detalle, se expone la forma en que estos fueron tratados para dejarlos listos de manera que su análisis se realizase de forma adecuada.

Los resultados aquí mostrados, se enfocan en el análisis de las velocidades de las unidades, aunque también se estudian brevemente otros parámetros de los datos, tales como el número de unidades y la ubicación de éstas. El análisis empieza con la presentación de histogramas de velocidad, se aumenta gradualmente la profundidad y nivel de abstracción del análisis hasta desembocar en el uso de la *divergencia de Kullback-Leibler*, concepto utilizado para realizar una comparación de las tendencias que muestran los datos. Al final del capítulo, se presentan resultados de cuyo análisis más exhaustivo emana el tema de estudio que se presentará en el siguiente capítulo.

3.2. Software utilizado para el análisis

A continuación, se describen algunas de las herramientas computacionales para la adquisición y tratamiento de datos que describen la actividad de los vehículos en el sistema Metrobús de la Ciudad de México.

3.2.1. Formato *General Transit Feed Specification*

El formato General Transit Feed Specification, cuyas siglas son GTFS, es un proyecto que fue desarrollado por Chris Harrelson, en colaboración con Bibiana McHugh, siendo su primer lanzamiento el 7 de diciembre de 2005, usando como base los datos de área circundante al metro de Portland, Estados Unidos [65]. Este servicio permite a las agencias de transporte público, como el metrobús de la CDMX, publicar los datos que estos generan. Es tal su alcance que hoy en día 261 agencias hacen públicos sus datos por medio de este formato [66]. La publicación de los datos puede hacerse de dos formas, por medio de un *GTFS estático* o de un *GTFS en tiempo real*. El primero define un formato común para itinerarios del transporte público y para información geográfica asociada (estaciones, líneas, tráfico, transbordos, etc), mientras que el segundo añade la publicación en tiempo real de las ubicaciones de las unidades de transporte público así como de su velocidad [67]. Los *feeds* GTFS dejan a las agencias de transporte público sus datos de tránsito y los desarrolladores hacen aplicaciones que consumen esos datos de una forma que puedan ser aprovechados para fines como consulta de itinerarios o consulta sobre el tiempo de arribo de una unidad de transporte público. Cualquier persona, con un usuario y clave, puede acceder a uno de estos *feeds* o puede crear uno. Entre sus aplicaciones están: el planeo de viajes, la comparación de niveles de servicio y la accesibilidad de la investigación [66].

3.2.2. Formato JSON

JSON corresponde a las siglas de JavaScript Object Notation [68] (Notación de Objeto de JavaScript), que es un subconjunto del lenguaje de programación JavaScript y, como tal, no posee ninguna característica que JavaScript no posea; más que un lenguaje de programación, es un formato de texto sencillo para el intercambio de datos. Su facilidad de uso proviene de la sencillez para escribirlo (sintaxis) y la facilidad que tiene la máquina para leerlo e interpretarlo, lo que ha convertido a este formato en un estándar dentro del intercambio de datos [68]. JSON está constituido por dos estructuras: una colección de

pares de nombres/valor y una lista ordenada de valores. Estas son estructuras universales y, virtualmente, todos los lenguajes de programación las soportan de una forma u otra.

3.2.3. Python

Python es un lenguaje de programación desarrollado en 1991 por Guido van Rossum en el Centro para las Matemáticas y la Informática (*Centrum Wiskunde & Informatica, CWI*) en los Países Bajos. Es un lenguaje de programación orientado a objetos que está disponible para la mayoría de los sistemas operativos. Python puede verse como un lenguaje emergente pues sigue siendo desarrollado y refinado. En su estado actual, es un lenguaje de programación que es tan bueno como Fortran o MATLAB [69]. A lo anterior hay que agregar que tiene disponibles una gran cantidad de bibliotecas que ofrecen miles de operaciones matemáticas en diversas ramas de esta ciencia, por ejemplo: estadística, álgebra lineal o cálculo.

3.2.4. PANDAS

Es una biblioteca de Python escrita como una extensión de NumPy, se usa para la manipulación y el análisis de datos. Su nombre se deriva de "PANel DATA". Esto describe de manera apropiada la función de la biblioteca: sirve para el análisis de datos, toma datos (CSV, TSV, SQL, etc.) y los convierte en un objeto de Python con columnas y filas llamada *Data Frame* [70]. Fue desarrollada por Wes McKinney en el 2008 mientras trabajaba en AQR Capital, por la necesidad que tenía de una herramienta flexible de alto rendimiento para realizar análisis cuantitativo de datos financieros.

3.2.5. QGIS

Según su sitio web [71], QGIS es un Sistema de Información Geográfica (en inglés *Geographic Information System*) de código abierto, licenciado bajo GNU. Es un proyecto oficial de Open Source Geospatial Foundation, que puede ser usado en Linux, Unix, Mac OSX, Windows y Android. Hasta 2013 era conocido como *Quantum GIS* y posteriormente la palabra *Quantum* pasó a formar parte de su acrónimo. Este software permite a los usuarios importar información geográfica para trabajar mapas, analizar y editar esta información, componer y exportar mapas. Soporta numerosos formatos y funcionalidades de

datos vector, datos ráster y bases de datos; estos vectores son almacenados como puntos, líneas o polígonos; también es capaz de georeferenciar imágenes.

3.2.6. Mathematica

Proyecto fundado por Stephen Wolfram en 1987. Es un paquete de software muy utilizado en ciencia e ingeniería por su capacidad de resolver muchos problemas de la Física directamente en la computadora [72], dado el catálogo de librerías disponibles. Que en este trabajo fueron usadas para ahorrar tiempo de escritura y de ejecución, pues algunas instrucciones precargadas en este software tienen algoritmos optimizados para realizar operaciones en un menor tiempo. También se aprovechó la facilidad de admitir una gran cantidad de extensiones para trabajar con los archivos descargados y aquellos que fueron usados en programas especializados en información geográfica; que sin duda alguna hubieran tornado parte del análisis en una tarea larga y tediosa de hacerse en lenguajes menos sofisticados.

3.3. Descarga y tratamiento de los datos

Los datos analizados en este trabajo son aquellos que la administración del metrobús libera para quien los requiera. Estos datos pueden solicitarse en su página oficial, en la sección *Datos Libres*. Son dos los tipos de datos disponibles: *estáticos* y *dinámicos*; en esta página web se encuentra un enlace a un formulario que debe llenarse para solicitar los datos. Si dicha solicitud es aprobada, las autoridades correspondientes responden con un correo en el que se incluye un nombre de usuario y una contraseña para acceder a un servidor *General Transit Feed System* de datos dinámicos, al que las unidades del metrobús mandan datos de su ubicación, velocidad y número de identificación cada 30 segundos, desde las 4:59:00 a.m. hasta las 11:59:59 p.m, que es el horario de servicio del metrobús. Mediante un código escrito en Python se colectan estos datos cada vez que el servidor actualiza los mismos y los escribe en una base de datos que se exporta como un archivo de extensión *.JSON*; deben descargarse de esta manera porque metrobús no los guarda debido a la gran cantidad de información que se acumula a lo largo del tiempo, un tamaño suficientemente grande como para considerarlos dentro de la categoría de *Big Data*. Además, al finalizar la jornada diaria, este mismo código comprime todos los archivos descargados y borra los archivos originales para evitar la saturación de la unidad

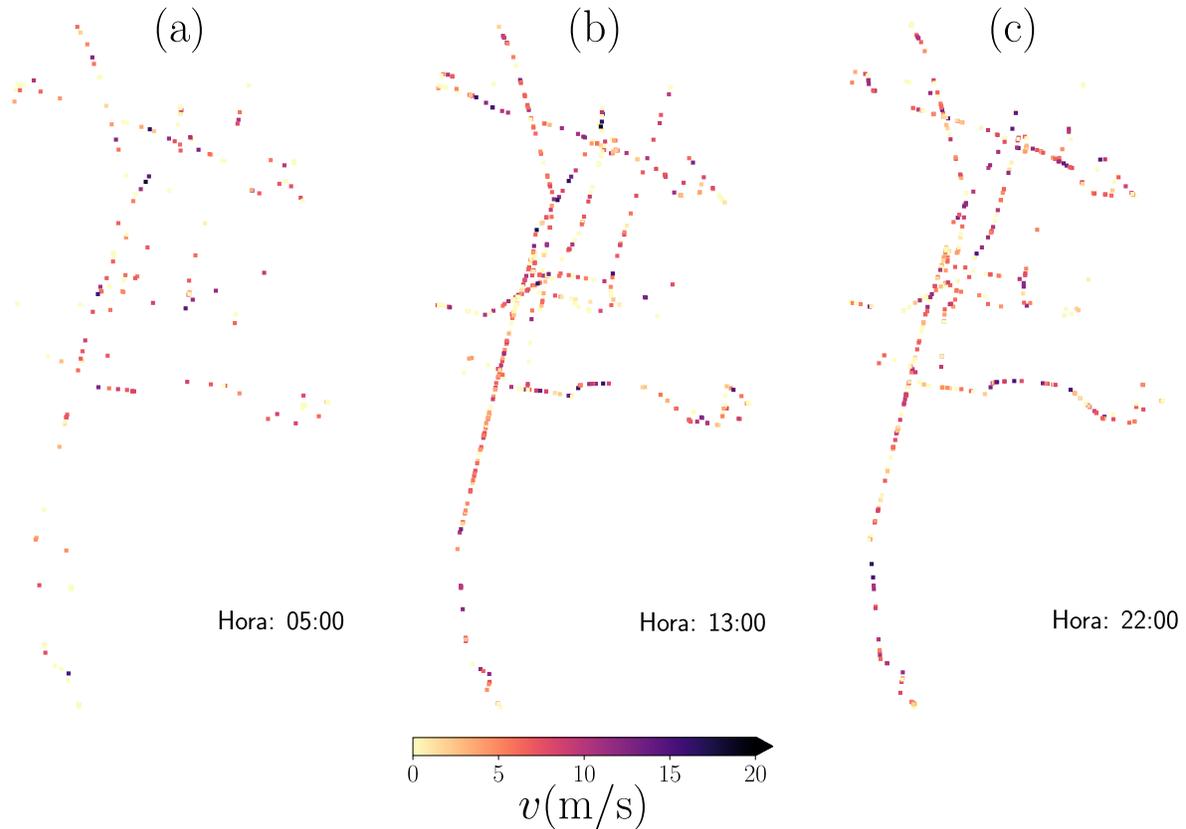


Figura 3.1: Visualización de datos del sistema metrobús con la actividad registrada el día 2 de marzo del 2020 en tres momentos del día: (a) 5:00 AM, (b) 1:00 PM y (c) 10:00 PM. Cada vehículo activo en el sistema se representa por un punto en el que su color indica la velocidad v codificada en la barra de color.

de almacenamiento que esté en uso.

Los datos se descargaron desde dos equipos: el primero en el Instituto de Física, a través de un equipo que forma parte del inventario de dicho Instituto; el segundo, usado a manera de respaldo, obtenía los datos a partir de una conexión externa a la Universidad. Es conveniente mencionar que los datos con los que se ha trabajado se recolectaron dentro del período 16 de febrero de 2020 - 23 de junio de 2020, con excepción de la semana del 12 al 18 de junio; los datos obtenidos se organizaron en 18 semanas y 5 meses. La primera semana empieza el 16 de febrero del año en curso y la última empieza el 18 de junio del mismo año.

A partir de los datos descargados, es posible consultar la velocidad, medida en m/s, y el número de identificación de las unidades que transitan en el sistema, información útil cuando se quiera determinar el estado de la red en un momento determinado. De esta manera puede reconstruir la actividad de toda la red en cualquier período de tiempo, tal

Tabla 3.1: Primeros 11 datos del día 16 de febrero de 2020, ordenados en un *DataFrame*. De esta manera, los datos pueden analizarse más fácilmente, pues pueden crearse filtros por columna para escudriñar únicamente los datos de interés.

id_vehículo	latitud	longitud	velocidad (m/s)	día	hora	id_segmento
3	-99.159203	19.510201	0.0000	6	4	34
18	-99.063599	19.383101	8.8889	6	4	231
19	-99.104897	19.397301	8.3333	6	4	220
20	-99.087402	19.391899	3.8889	6	4	225
3	-99.058899	19.389900	0.0000	6	4	233
4	-99.159203	19.510201	0.0000	6	4	34
5	-99.144096	19.463499	7.7778	6	4	142
6	-99.159103	19.508200	4.7222	6	4	35
20	-99.063599	19.383101	8.8889	6	4	231
21	-99.104897	19.397301	8.3333	6	4	220
22	-99.089699	19.393400	0.2778	6	4	224

como se muestra en la Figura 3.1 en la que se recrea la actividad de la red durante el día 2 de marzo para los tres momentos indicados; se observa que a las 5:00 a.m. la actividad muestra poca densidad en toda la red, en comparación con la actividad registrada a la 1:00 p.m. y a las 10:00 p.m. Esta imagen da una visión general de las tendencias de comportamiento de la red, tal como horarios de máxima afluencia, líneas más demandadas y tramos donde la velocidad de las unidades es mayor que el promedio.

Los datos se procesaron con el *software* QGIS, versión 3.14 Pi. Posteriormente, los datos que resultan de este paso se procesan con otro código escrito en Python, del que se obtienen los resultados que se mostrarán en la siguiente sección. Los archivos reprocesados en Python tienen una extensión .csv, es decir, son archivos separados por comas. En un *Notebook de Jupyter* se trabaja con ellos transformándolos en objetos tipo *DataFrame*, de esta forma los datos son organizados en columnas que tienen información acerca de la unidad que emitió los datos, la longitud y latitud, la hora, el día y la velocidad de la unidad.

Por su importancia en el análisis de los datos realizado en este trabajo, es necesario explicar la última columna de la Tabla 3.1 y la manera en la que se generan los datos contenidos en ella. Un segmento o *polígono* (palabras que en la siguiente parte se usan indistintamente a modo de sinónimos) es cada uno de los polígonos construidos a lo largo de las 7 líneas del sistema, tal como se muestra en la Figura 3.2 (a). Con la herramienta *trazar polígono*, de QGIS, se trazaron paralelogramos con la siguiente regla general: dos de

sus lados son paralelos a la calle donde transitan los autobuses y los dos restantes la cruzan perpendicularmente, de manera que corten por la mitad cada punto que represente una estación; esta regla tiene sus excepciones en los puntos donde convergen 3 o más segmentos. El ancho de los polígonos intenta encerrar la mayor cantidad posible de puntos dentro de ellos.

Para ilustrar la forma en que son acomodados los datos, en la Tabla 3.1 se presentan las primeras 11 filas de datos de los datos correspondientes al domingo 16 de febrero de 2020. Un conteo revela que ese día se registraron un total de 315,205 entradas de datos. Tal como se puede apreciar en los registros presentados en Tabla 3.1, la base de datos contiene información sobre el vehículo que emite los datos (*id_vehículo*), sobre sus coordenadas (longitud y latitud), la velocidad, el día de la semana (con una numeración del 0 al 6 que le asigna el 0 al lunes y 6 al domingo), la hora (0 a 23 horas) y el segmento (*id_segmento*) por el que transitaba el vehículo al momento de transmitir los datos. La manera en que están organizados los datos en este formato permite filtrar la columna de interés dependiendo del enfoque que se desee dar al análisis. Por ejemplo, en la Tabla 3.1, se puede ver que las unidades 18, 19 y 20 empezaron a enviar datos cuando estaban en los polígonos 231, 220 y 225; también se puede ver cuáles eran sus velocidades en ese momento. De esta manera es relativamente sencillo dar un tratamiento adecuado a los más de 533 millones de datos que se colectaron durante el tiempo de estudio, tamaño que hace entrar a este trabajo dentro de la categoría del *big data*.

En cuanto a la notación para nombrar a los polígonos, esta se compone de dos o tres dígitos. El primero de ellos denota la línea en la que se trazó el polígono y el siguiente, o siguientes dos dígitos corresponde a su número ordinal dentro de su línea correspondiente. Por ejemplo, el segmento 19 corresponde al segmento número 9 de la línea 1, y el segmento 622 es el polígono número 22 de la línea 6. Sólo en el caso de algunos polígonos de la línea 4 su número de identificación se compone de 4 dígitos. En esta línea, los primeros dos polígonos se enumeran como los de las otras líneas y a los polígonos siguientes se les agrega un cero después del número que identifica a la línea a la que pertenecen. Esto se debe a que a partir del segundo polígono, el flujo vehicular de la línea se divide en *Ruta Norte* y *Ruta Sur*. De esta manera se podía tener un mayor control sobre el censo de los segmentos.

Por su sencillez, se han clasificado a los polígonos en dos tipos: *Sencillos* y *Fusionados*, mostrados en la Figura 3.2 (b). Los polígonos sencillos son aquellos que se trazaron con el procedimiento descrito anteriormente. Los polígonos fusionados resultan, como su nom-

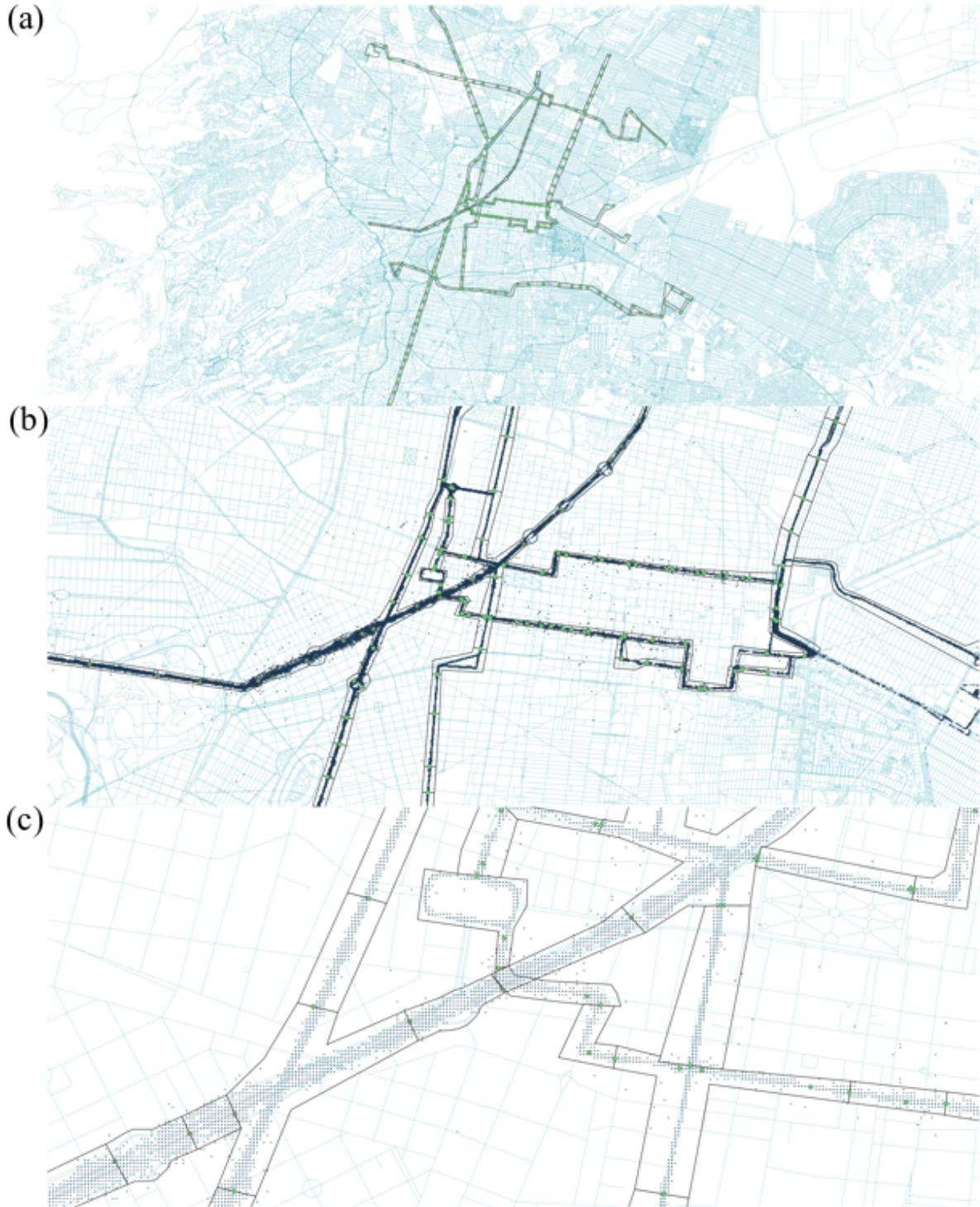


Figura 3.2: Panel de trabajo en QGIS. (a) Red de estaciones del Metrobús (puntos verdes) y los polígonos en que fueron divididas las líneas (polígonos de borde negro). (b) Detalle del mapa anterior sobre la *zona centro de la ciudad*, obsérvese cómo los polígonos contienen a la mayoría de los puntos negros (posiciones de las unidades de metrobús). (c) Detalle a mayor escala del inciso (b)

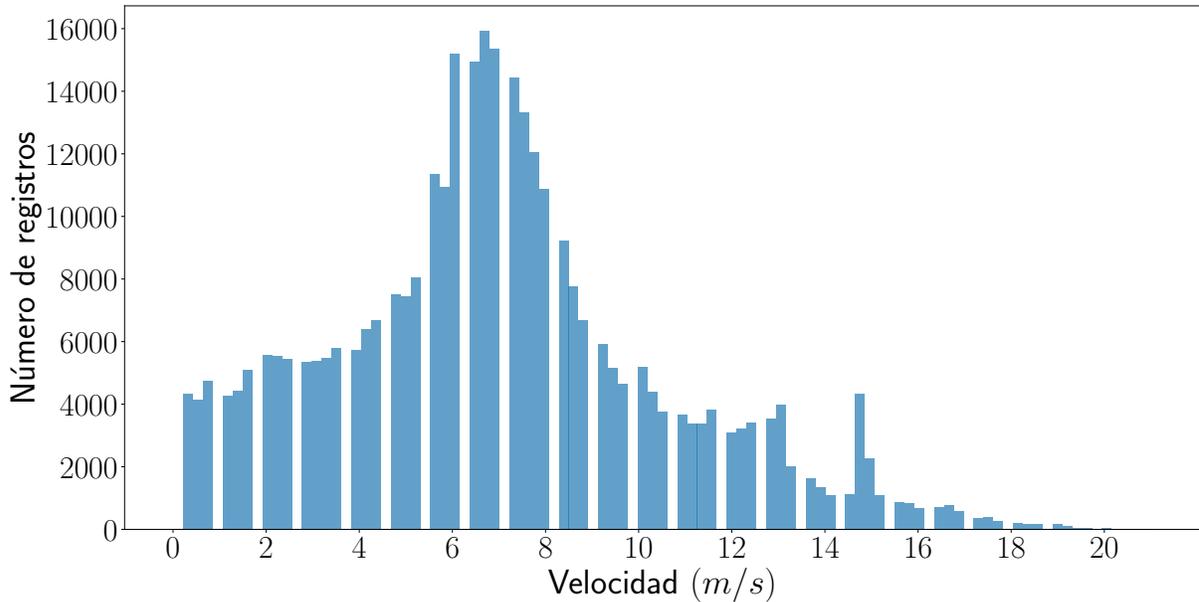


Figura 3.3: Histograma de las velocidades registradas el 16 de febrero. Se han omitido los registros de velocidad mayores a 20 m/s dado el bajo porcentaje de estos.

bre lo indica, de la fusión de dos o más polígonos sencillos que se traslapan parcial o totalmente; a los polígonos fusionados se los asignó dentro de una hipotética Línea 8, ello debido a las limitaciones del software de manejo geográfico para nombrar a los polígonos, que sólo acepta números sin espacios. La inclusión de esta línea extra fue necesaria debido a la dificultad de clasificar a las estaciones que pertenecen a más de una línea. Esta complejidad radica en que contienen el flujo de más de una línea, lo cual puede ocasionar análisis erróneos. Por ejemplo, en un intersección de polígonos habrá un mayor número de vehículos que en los polígonos sencillos adyacentes, caso diferente a los polígonos sencillos, en los que se puede afirmar que los datos extraídos de ellos son fiables.

Por otro lado, es preciso mencionar que en los datos importados solamente se incluyen las localizaciones de las estaciones de las primeras 6 líneas. Las estaciones de la línea 7 tuvieron que introducirse “a mano”, buscando su localización en Google Maps, haciendo una tabla con estos datos y guardando los datos de la misma en un archivo separado por comas, es por ello que su simbología tiene un color diferente al de las estaciones de las otras líneas. Dentro del mismo archivo de QGIS, también se muestran las ubicaciones de las unidades de Metrobús, simbolizados con puntos negros. Los puntos, las estaciones y las calles de la ciudad están cada una en una capa del proyecto. Por lo que pueden ocultarse o mostrarse, según se convenga.

3.4. Histogramas de velocidad

En esta sección se presenta el análisis estadístico de las velocidades a diferentes escalas, tanto espaciales como temporales. Primero se estudia la distribución de las velocidades en toda la red. Esta distribución puede presentarse como un histograma de barras en los que la altura de cada barra corresponde al número de *registros* asociados con esa velocidad. De ahora en más, registro se entenderá como el componente mínimo del paquete de datos enviado por una unidad en un lapso de 30 segundos. Por ejemplo, en la figura 3.3, se muestra la distribución de velocidades, graficada en un histograma, de toda la red durante toda la jornada del 16 de febrero del año en curso. En dicha Figura, se observa que el máximo de registros está en torno a los 7.5 m/s (27 km/h), cuyo número es aproximadamente 16,000; esto significa que en un total aproximado de 16000 lapsos de 30 s, el conjunto de todas las unidades en circulación circulaban a 7.5 m/s. Así, en ese día, las unidades totalizaron 133 horas avanzando a esa velocidad. En cuanto a las características del perfil del histograma se puede decir que el aumento de registros aumenta de forma monótona desde los 0 m/s, cuyo número es ligeramente mayor a los 4000 registros, hasta alcanzar el máximo total. Después disminuye de forma monótona, aunque a un ritmo menor. También, se observa la existencia de un máximo local en torno a los 15 m/s, a partir del cual el número de registros con velocidades mayores va descendiendo hasta que es cero más allá de los 20 m/s (72 km/h).

Analizando de esta misma manera los datos para el resto de los días, se encuentra que este comportamiento se mantiene constante a lo largo del tiempo; en todos los días el máximo absoluto y el máximo secundario están en lugares similares del histograma, así como los ritmos de aumento y descenso, antes y después del máximo absoluto, también son similares. Por otro lado, la diferencia más notoria recae en el número de registros; cantidad que varía según se analice un día entre semana o uno de fin de semana. Por ejemplo, mientras que en un domingo las cifras suelen ser menores a 20000, en los días entre semana los registros son mayores a 30000. En general, se observa una relación cercana a 2:1. Esta primera conclusión concuerda con nuestra experiencia diaria, dado que el número de registros es directamente proporcional al número de unidades en circulación. Es lógico pensar que el número será menor en fines de semana debido a la disminución de actividades laborales y educativas.

En este mismo tenor comparativo, en la Figura 3.4 se presentan los histogramas para algunos días de actividad *normal*, contrastados con eventos que tuvieron efectos importantes en la movilidad de la ciudad: 3 de marzo (martes) un día entre semana; 8 de marzo

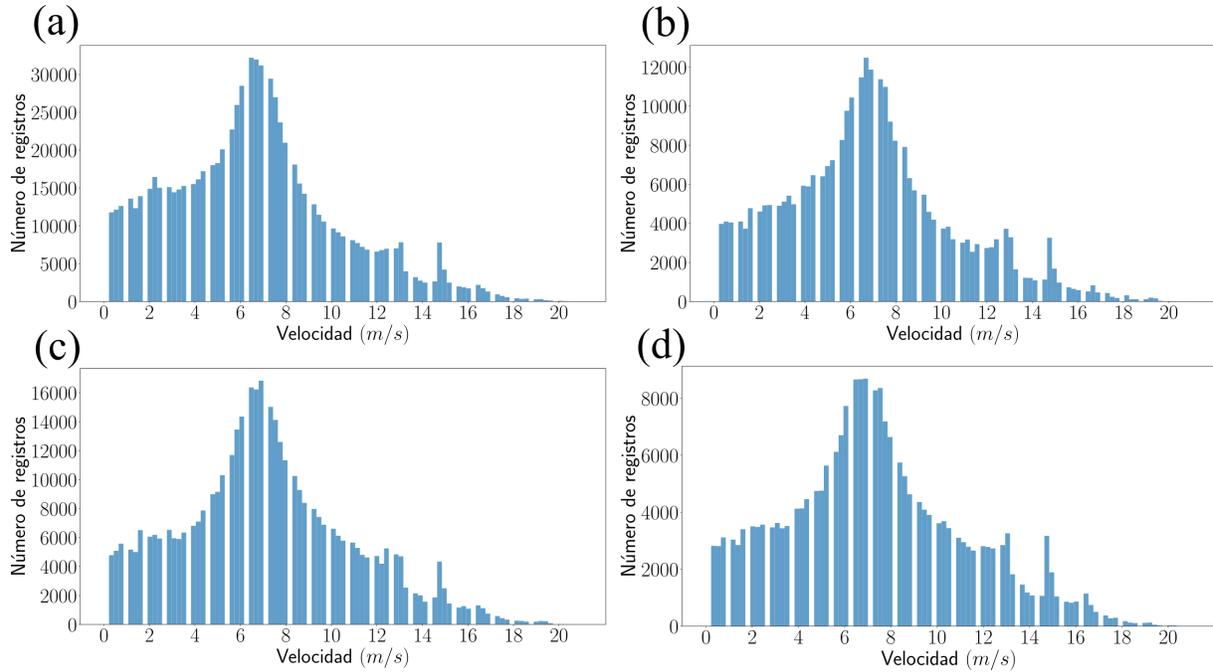


Figura 3.4: Histogramas de velocidad para 4 días diferentes. (a) 3 de marzo, (b) 8 de marzo, (c) 24 de abril y (d) 11 de junio. De la misma manera que en la Figura 3.3, los registros se han limitado a velocidades menores a 20 m/s .

(domingo), día de la mujer; 24 de abril (viernes), inicio de la restricción de movilidad por la pandemia de COVID-19; y 11 de junio (jueves), días después del inicio de restricción de movilidad. Se ha escogido el 3 de marzo 3.4 (a) porque hasta entonces la movilidad en la ciudad no había sido afectada por medidas oficiales. En esa fecha, se puede observar que el máximo absoluto tiene 30000 registros, por lo que es posible calificar su actividad como *normal*; el 8 de marzo se registró una actividad más baja de lo habitual para ser domingo, pues esta descendió a casi un tercio del 3 de marzo, descenso notorio si se considera que ambos días son de la misma semana. El día que inició formalmente la restricción de la movilidad, se observó un descenso pronunciado de casi el 50% de la actividad normal; y ya en el histograma correspondiente al 11 de junio se puede observar el efecto no sólo de medidas oficiales. También se manifestó un confinamiento autoimpuesto, ya que muchos sectores de la población evitaron salir en aquellos días.

La familiarización con el análisis de estos datos lleva a preguntarse si la red tiene comportamientos característicos en cada día de la semana, durante la semana inglesa o los fines de semana. Como punto de partida para responder esta interrogante, en la Figura 3.5, se presentan los histogramas de velocidad para todos los días lunes, tanto en el número de registros (a), como en la densidad de estos (b). Es posible ver que todos los días se comportan de manera muy similar, pero se observa que para los lunes correspondientes a

la segunda mitad del período de estudio, el número de registros desciende sensiblemente. Sin embargo, a pesar de la diferencia de magnitudes, persisten las mismas características principales. Por ejemplo, el máximo absoluto en torno a los 7.5 m/s, un máximo secundario en torno a los 15 m/s, y las mismas tendencias de aumento y descenso en el número de registros. La diferencia en magnitudes se atribuye al período de restricción de movilidad por la pandemia de COVID-19. Fuera de ello, hay ligeras diferencias que son difícilmente medibles en términos *cualitativos*, asunto que se tratará más adelante.

Observando el histograma en términos de la densidad de velocidades, los cambios en el tiempo disminuyen sensiblemente; de acuerdo con los resultados, se advierte que en todos los lunes las densidades de velocidad son prácticamente las mismas. Por ejemplo, el 15 % de los registros tienen una velocidad de 7.5 m/s y, aproximadamente, el 5 % lo hace a 15 m/s. También, es de notarse que la fluctuación en las densidades de velocidades es más ancha en velocidades bajas que en velocidades altas, fluctuaciones análogas a las del inciso (a) de la Figura 3.5. De la misma manera, se puede hacer un análisis más fino analizando estas distribuciones por hora en vez de hacerlo por día. Así es posible saber cuál es la dinámica de la red a distintas horas de una jornada de servicio.

Como ya se ha mencionado, los perfiles de los histogramas de velocidad por día son similares entre sí, por lo que puede tomarse un día cualquiera para empezar el análisis; aquí se ha tomado el 24 de febrero (lunes) como ejemplo. Por otro lado, en la Figura 3.6 (a) se han graficado los histogramas de velocidades para cada una de las horas en que el servicio de Metrobús está activo. Sobre lo que se ha observado, se hace mención destacada de lo siguiente: todos los histogramas tienen un perfil muy similar y su diferencia más notoria radica en el número de registros. Se puede decir, con un buen grado de exactitud, que un histograma cualquiera es un escalamiento de otro, salvo en el caso de aquel correspondiente a las 5:00 horas, pues todos los espectros comparten los mismos rasgos característicos, como el máximo absoluto en torno a los 7.5 m/s, un máximo secundario en torno a los 15 m/s y las mismas tendencias ascendentes y descendentes. En cuanto a la actividad, se observa que registra sus números más altos en dos períodos: entre las 7 y las 11 horas (región naranja-amarilla del espectro) y entre las 17 y las 21 horas (región azul claro del espectro). Tales comportamientos son fácilmente atribuibles a la dinámica de la ciudad, ya que esas son las horas de inicio y término de las jornadas escolares y laborales. Si también se analiza la actividad, pero ahora en términos relativos, se observa que los histogramas son muy parecidos entre sí. Esto significa que sin importar la hora o el nivel de actividad en el sistema, la distribución de velocidades es, cualitativamente, la misma.

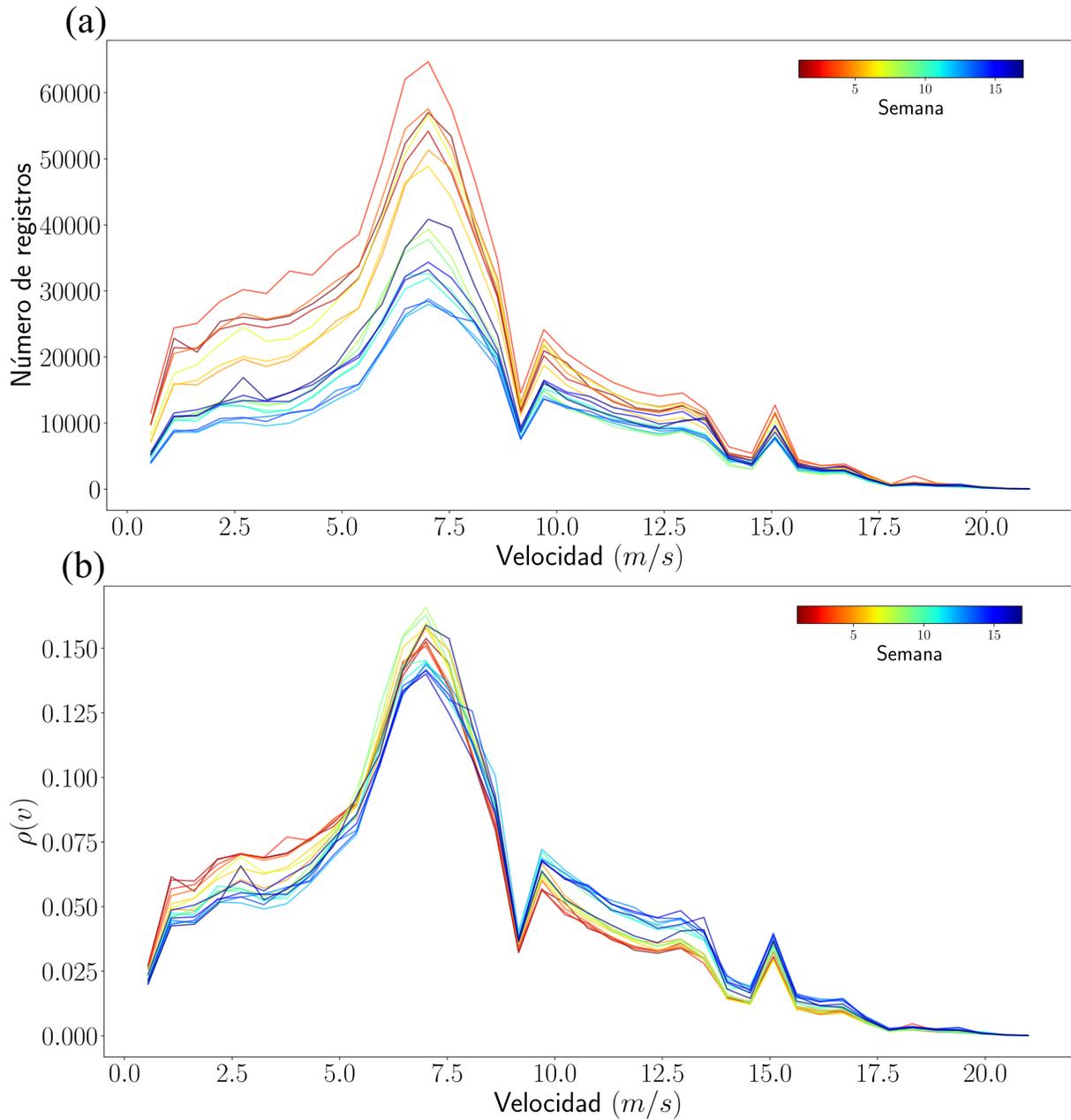


Figura 3.5: Histogramas de velocidad de todo el sistema durante los días lunes de las 18 semanas de estudio, donde cada histograma se colorea según la barra de color mostrada en ambas gráficas. En (a) se han graficado estos histogramas en términos absolutos, obsérvese cómo es que los colores del rojo al amarillo tienen los registros más altos, en contraparte de aquellos que van del verde al azul; en (b) se muestran estos histogramas pero en cantidades normalizadas. En ambos casos el período temporal es presentado con una barra de colores.

A continuación, se abre un paréntesis para centrar este análisis en el estudio del número de vehículos que transitan por cada una de las líneas, a fin de caracterizarlas individualmente. De esta manera, se muestra al lector que el número de unidades en circulación también muestran patrones de comportamiento que permitan predecir, con un buen nivel

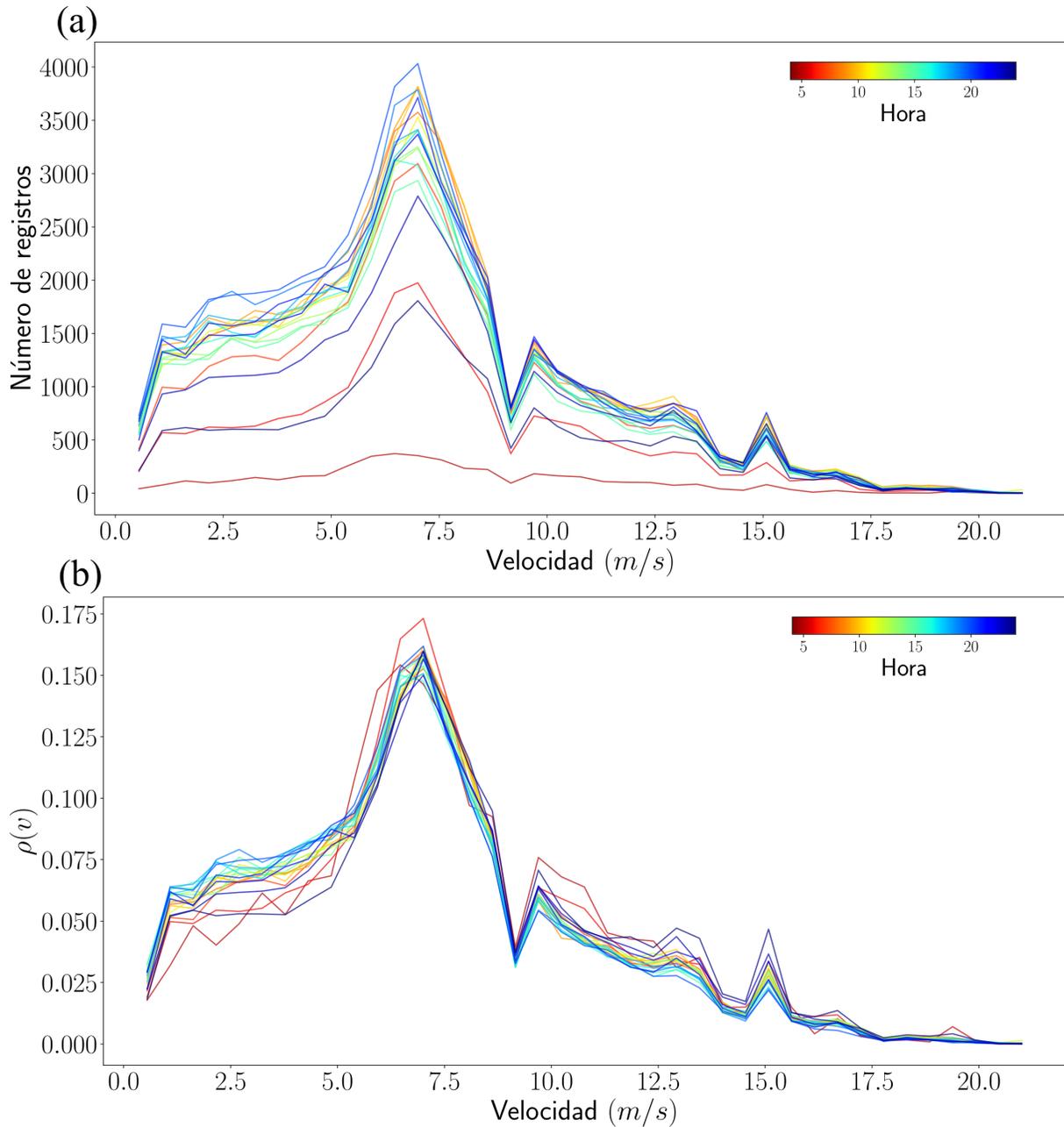


Figura 3.6: Distribuciones de velocidad en cada una de las horas de servicio. En (a) se presentan las distribuciones medidas en términos absolutos, obsérvese cómo estas líneas se aglutinan en torno a los horarios de máxima demanda del servicio. En (b) se presentan las distribuciones normalizadas, en este gráfico se observa que a todas horas se conserva la misma distribución.

de certeza, el comportamiento de cada una de las líneas del sistema; para posteriormente profundizar el análisis de los datos concernientes a la velocidad de las unidades del sistema.

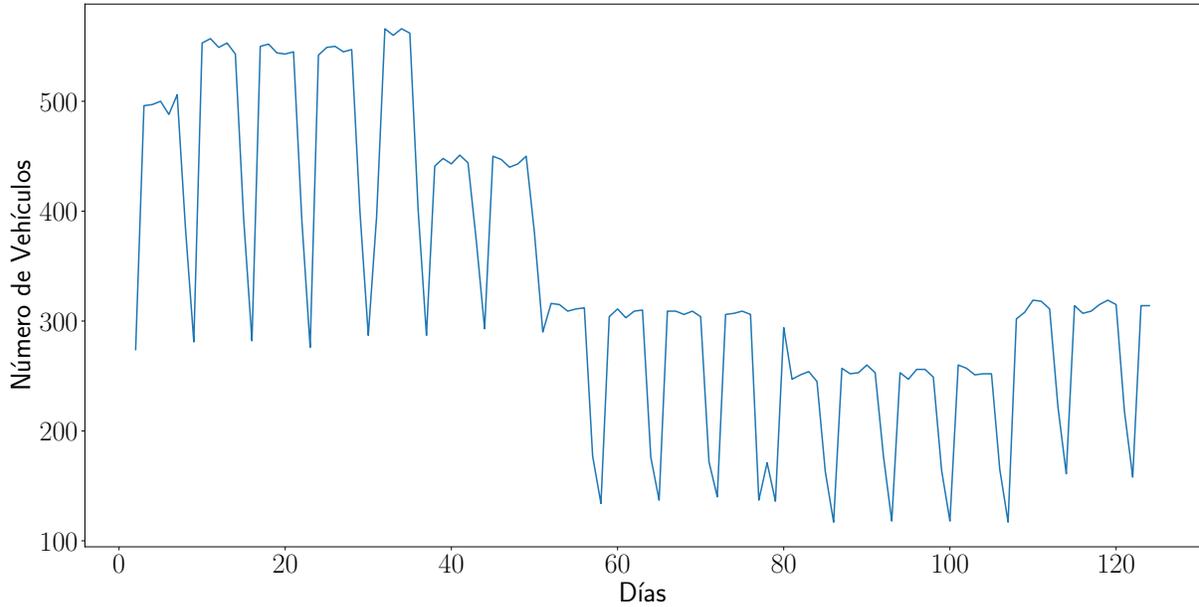


Figura 3.7: Histograma del número de vehículos en circulación en toda la red de Metrobús durante cada día del tiempo de estudio. Obsérvese la estructura *en bloques* que conforman al histograma. Dichos bloques corresponden a los días entre semana, mientras que los descensos abruptos corresponden a los fines de semana

3.5. Actividad vehicular en el sistema metrobús

En esta sección se hace una breve pausa para estudiar la estadística sobre el número de unidades en circulación, a diferencia de la sección anterior en donde el análisis se basó en la velocidad de los vehículos. El número de unidades en circulación muestra un patrón de comportamiento en el que es posible una identificación rápida de los días entre semana y de los días de fin de semana. Al observar los datos en un archivo de texto plano, se observan secuencias de 5 datos en que la magnitud de los números son similares, y, posteriormente, un par de datos en los que se registra una caída abrupta en el número de unidades en circulación. En conjunto con esta observación y aquella sobre el número de registros que se hizo en la sección anterior, es posible inferir que la primer secuencia a la que se hace referencia son los días lunes, martes, miércoles, jueves y viernes; y la segunda corresponde al sábado y al domingo.

Esta cadena de datos se presenta gráficamente en la Figura 3.7, donde se presenta el histograma de actividad de toda la red del Metrobús. Ahí se observa una estructura conformada por *bloques*, que corresponden a una semana de actividad. Con esta presentación gráfica de los datos se puede hablar de tres períodos bien diferenciados en cuanto a la magnitud de la actividad vehicular: *período normal*, *período de transición* y *período pandémico*. En

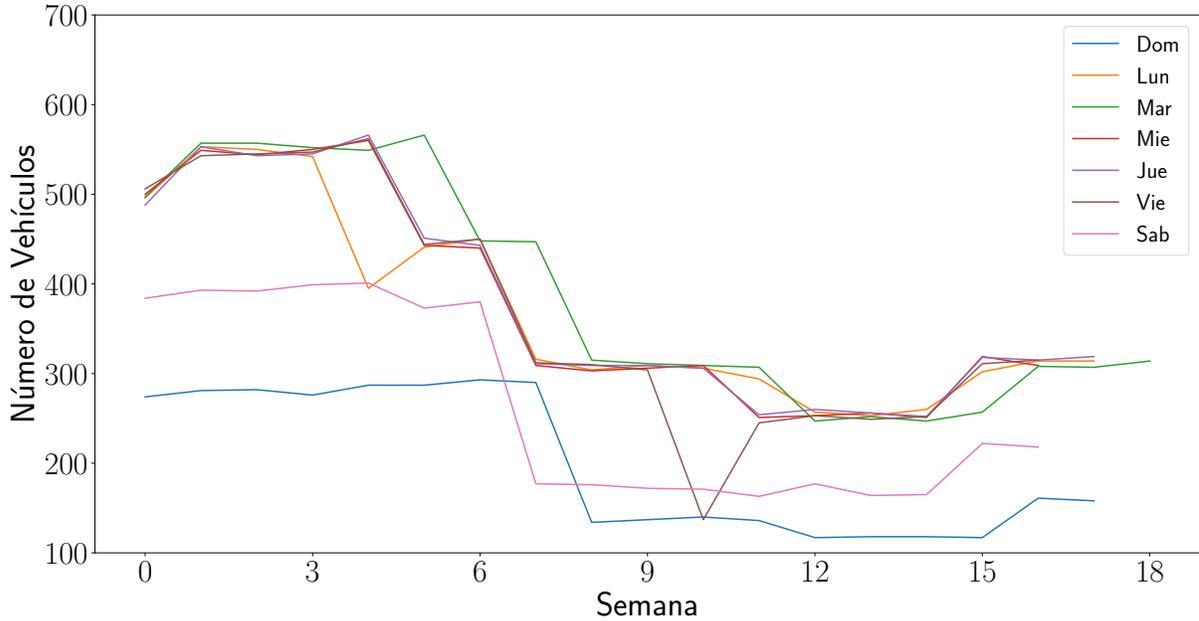


Figura 3.8: Histogramas del número de unidades en circulación en toda la red para cada uno de los días de la semana. Obsérvese la predominancia de actividad en los días de la *semana inglesa* y la regularidad en los números a pesar de la restricción de movilidad que empezó en la semana 6.

el primer período, conformado por 5 semanas, se observa que entre semana el número de vehículos es mayor a 500 y en los fines de semana esta actividad cae hasta un promedio de 280 vehículos. En el período de transición (que empieza el 22 de abril) se observa que la actividad entre semana cae por debajo de los 450 vehículos pero la actividad en fin de semana permanece igual al primer período. Finalmente, en el período pandémico se observa una nueva caída en la actividad entre semana hasta ser similar a la actividad de fin de semana de los dos períodos anteriores, mientras que en fin de semana la actividad disminuye hasta 100 unidades en circulación.

Se puede hacer un análisis más detallado de la actividad del sistema si lo que se estudia es la actividad en cada uno de los días de la semana y de esta manera, saber cuáles de ellos son más activos que los otros. En la Figura 3.8 se han superpuesto los histogramas de actividad de los siete días de la semana; donde se observa que hay una ligera predominancia de actividad en los días martes. Los demás días de la *semana inglesa* tienen números muy similares, salvo el lunes de la semana 4, en el que se observa una caída en la actividad incluso por debajo de la del sábado de esa misma semana. Lo mismo que el viernes, que registra un marcado descenso en la semana 10, cayendo incluso por debajo del domingo.

El promedio de vehículos en circulación también muestra constancia a lo largo del tiempo, aunque se pueden encontrar dos períodos bien diferenciados con base en el número de

unidades en circulación. Según se observa en la Figura 3.9 (a), entre la semana 0 y la número 6 se registra que el número de unidades en circulación oscila entre 450 y 500 vehículos diarios. Pero a partir de la semana 7, se aprecia un marcado descenso en su número, menor a 300. Es útil hacer notar que, a partir de la semana 13 se registra una leve recuperación en el promedio de vehículos en circulación. Con ayuda de estos números, se observa con facilidad el efecto que tuvo la contingencia del COVID-19 en la movilidad de la ciudad. Las restricciones derivadas de este fenómeno provocaron un cese que llegó a superar el 50% de la movilidad en este sistema de transporte. Estas cifras pueden consultarse en detalle en la Tabla 3.2, donde también se han agregado los intervalos de cada una de las semanas tomadas en consideración en este estudio.

Por otro lado, en la Figura 3.9 (b) se aprecia un resultado que concuerda con nuestro sentido común, los días más activos son aquellos entre semana, cuyas cifras son: Lunes: 413 , Martes: 425 , Miércoles: 422, Jueves: 428, Viernes: 396, Sábado: 298, Domingo: 235. En cuanto a las líneas, se observa que cada una tiene un perfil de distribuciones de actividad vehicular muy constante a lo largo del tiempo y que además es característico

Tabla 3.2: Número de unidades en circulación por cada una de las semanas del estudio. También se especifican las fechas correspondientes a cada semana. *semanas incompletas por falta de datos

Semana	Unidades en circulación (prom)	Período
0	449	16 - 22 de febrero
1	490	23 - 29 de febrero
2	487	1 - 7 de marzo
3	487	8 - 14 de marzo
4	477	15 - 21 de marzo
5	412	22 - 28 de marzo
6	415	29 de marzo - 4 de abril
7	290	5 - 11 de abril
8	264	12 - 18 de abril
9	263	19 - 25 de abril
10	239	26 de abril - 2 de mayo
11	271	3 - 9 de mayo
12	241	10 - 16 de mayo
13	220	17 - 23 de mayo
14	222	24 - 30 de mayo
15	271	31 de mayo - 6 de junio
16	281	7 - 11 de junio*
17	273	18 - 23 de junio*

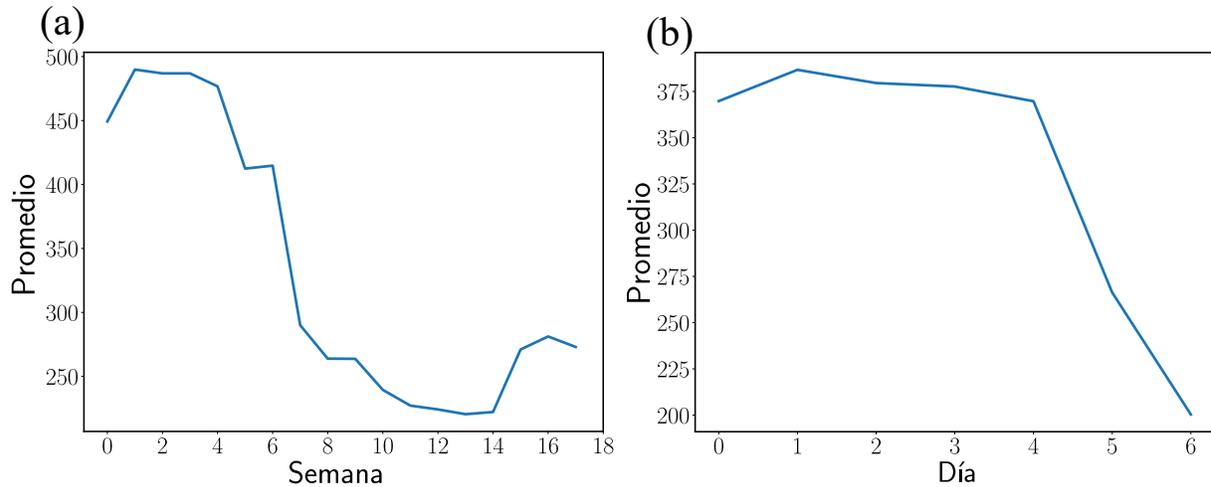


Figura 3.9: Promedio del número de unidades en circulación, en cada una de las semanas que abarca el período de estudio (a), donde la semana 0 corresponde a la primera semana de estudio; y para cada uno de los días de la semana (b), donde el día 0 corresponde al lunes.

para cada línea.

,

Línea 1: muestra un comportamiento fluctuante, pero con una tendencia general “recta”, salvo un aumento inusual entre las estaciones Colonia del Valle y Nuevo León. Línea 2: de la misma manera que en la línea 1, se observa un comportamiento fluctuante. Las primeras estaciones tienen menos unidades en servicio, cantidad que aumenta a partir de la octava estación, y cuya tendencia permanece más o menos horizontal. Línea 3: muestra un comportamiento muy constante en casi todas sus estaciones, con excepción de los segmentos 6-7 y 19-21, que corresponden a los polígonos comprendidos entre las estaciones Poniente 46 y Poniente 134, para el primer caso y Mina y Balderas para el segundo caso. Por último, también se observa un aumento de la actividad en la estación Etiopía. Línea 4: en la estación Glorieta de Colón se observa un descenso muy fuerte en el número de unidades en circulación, para recuperar sus valores anteriores en la siguiente estación y mantener una tendencia horizontal. Se observa un aumento notable en la actividad a partir de la estación San Lázaro, cantidad que se mantiene en la siguiente estación, para después disminuir, pero no a niveles iguales a los anteriores, manteniendo una tendencia uniforme hasta la estación Morelos en la que existe otro marcado descenso para, posteriormente, volver a sus valores anteriores en la última estación. Línea 5: como en el caso de la línea 3, muestra un comportamiento muy regular en todas sus estaciones, menos en la estación San Juan de Aragón. Línea 6: es una línea que muestra un comportamiento menos caótico que las líneas 1, 2 y 3; pero no tan uniforme como las líneas 3 y 5. La tendencia describe

una curva cóncava que empieza su ascenso en la estación Montevideo, y termina en la estación La Pradera, donde vuelve a aumentar su actividad progresivamente hasta tener su máximo en la estación Villa de Aragón. Línea 7: esta línea exhibe una caoticidad similar a la línea 6, pero con una tendencia horizontal. Su rasgo más distintivo es que tiene un abrupto descenso en el tramo entre la estación Garrido y Gustavo A. Madero. Las líneas más concurridas son la 1 y la 2, que son, a su vez, las más largas y las más antiguas en operación. La línea 3 es la que tiene menos afluencia

Promediando la actividad de cada línea se puede observar la evolución que han tenido a lo largo del tiempo de estudio. Lo que se nota es que el comportamiento promedio de una semana comparado al de otra no muestra cambios importantes en su forma, sino únicamente en el número de vehículos en circulación. La diferencia más notoria es la disminución en el número de vehículos, agregando que en las gráficas correspondientes a las últimas 4 semanas en la Línea 1, desaparece el aumento de actividad entre las estaciones 20 y 25, mientras que el aumento sólo se registra en la estación 25. Las primeras siete semanas muestran números similares, mientras que las últimas 3 muestran un decremento de actividad, tanto así, que la diferencia entre los números de las primeras semanas y los de la última tienen una relación de aproximadamente 2 a 1 en las líneas 1, 2, 5 y 6 y de 3 a 1 en las líneas 3, 4 y 7.

3.6. Divergencia de Kullback-Leibler

Los análisis de datos realizados en las secciones anteriores permiten establecer comparaciones cuantitativas que se limitan al comportamiento de los histogramas asociados a la velocidad en cada uno de los segmentos, así como determinar su evolución temporal. En ciertos casos, esta comparación basta para establecer diferencias determinantes entre un cierto grupo de datos y otros. Por ejemplo, la diferencia entre el perfil de actividad promedio de las diferentes líneas que componen a toda la red, o la diferencia de actividad entre las primeras y las últimas semanas que comprenden el período de estudio aquí presentado. Un análisis más detallado exige establecer una escala que permita comparar diferencias, o semejanzas, entre tales bases de datos; en tal caso, la comparación realizada hasta ahora es insuficiente.

Entonces, es necesario implementar un método que permita establecer una medida de las diferencias o semejanzas en las distribuciones de velocidad que se han estudiado hasta ahora. Por ejemplo, en la Figura 3.10 se han superpuesto los histogramas de velocidad de

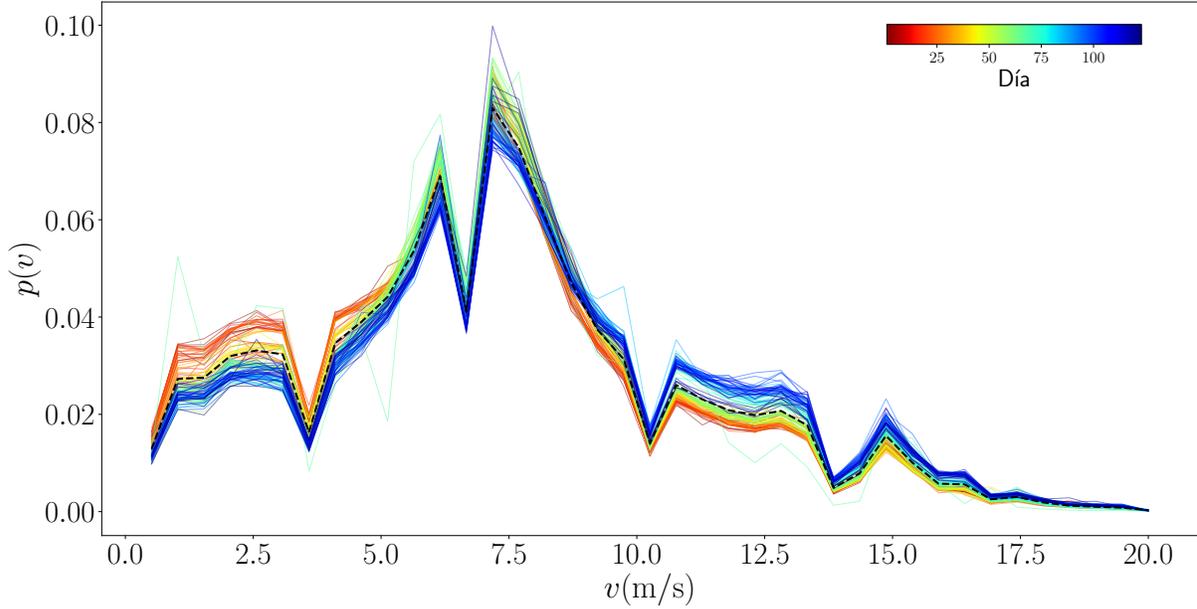


Figura 3.10: Comparación entre el histograma *total* (línea punteada) de velocidades y cada uno de los histogramas diarios durante el tiempo de estudio (líneas continuas), para toda la red de metrobús. Los valores graficados están normalizados.

toda la red (con sus valores normalizados) para cada uno de los días del tiempo de estudio, junto con el histograma *total* de velocidades para toda la red. En esta imagen se aprecia que una gran cantidad de los histogramas diarios caen dentro de una tendencia marcada por la línea punteada. Hasta ahí se puede inducir una tendencia general. El siguiente paso en el análisis es saber cuáles días se parecen más y cuáles menos, a una referencia que se haya convenido previamente. Como se ha visto que el comportamiento de toda la red es poco variable con respecto al histograma total, entonces este último puede ser un referente conveniente. Así, en aquellos días que no se parezcan demasiado al histograma total debió presentarse alguna eventualidad que afectó esa *normalidad* y entre menos parecido sea el histograma de un día en particular, mayor tendría que ser dicha perturbación. De esta manera, se muestra la utilidad que puede tener la identificación de días que se parecen más o menos al histograma total.

Para cuantificar la similitud entre histogramas, es necesario establecer un criterio que asigne una magnitud a dicha semejanza. Para ello se utiliza la *divergencia de Kullback-Leibler*, dada por

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}, \quad (3.1)$$

que permite comparar la distribución de valores P con la distribución de valores Q . En el uso que se le ha dado en esta investigación, se toman los datos de un histograma y se los

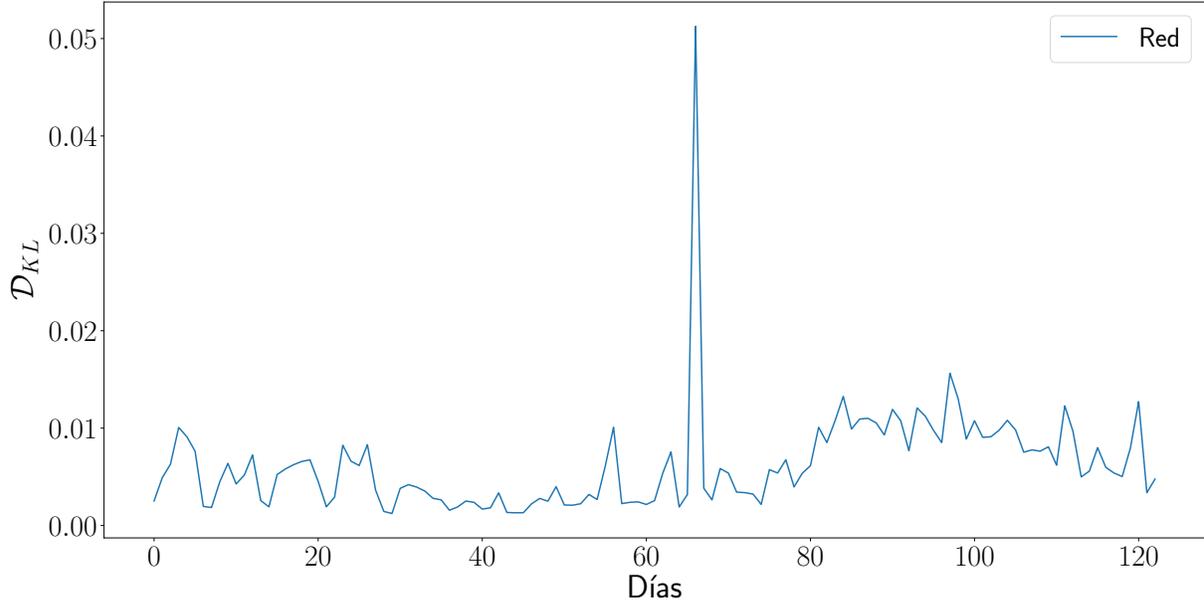


Figura 3.11: Histograma de las distancias de Kullback-Leibler calculadas para las distribuciones graficadas en la Figura 3.10, esto es, la distancia entre la distribución de cada uno de los días con la *distribución total de velocidades*.

considera como la *distribución P*. Se hace lo mismo con los datos de otro histograma, que juegan el papel de la *distribución Q*. La medida presentada en la ecuación 3.1 permite, por ejemplo, identificar cuáles histogramas, de los que se encuentran en torno a la línea punteada de la Figura 3.10, tienen mayor similitud con este último histograma. En la Figura 3.11 se han graficado las distancias de Kullback-Leibler entre la distribución de velocidades de cada uno de los días, cuyos datos juegan el papel de la distribución *P* y la distribución de velocidades que da origen al histograma total que toma el pape de *Q*; esta aclaración es importante ya que la distancia de Kullback-Leibler no es simétrica.

Por ejemplo, y según estas mediciones, el día 4 desde que se tomaron los datos, muestra una distancia de Kullback-Leibler de 0.001 con respecto a la *distribución total de velocidades*. Otro ejemplo es la distancia igual a 0.0156, registrada en el día 98. Así, en la Figura 3.11 se observa que hasta antes del día 67 todas las distancias de Kullback-Leibler son menores a 0.01 y justo en ese día esta distancia crece hasta 0.05 para, posteriormente, regresar a la tendencia anterior pero con un ligero aumento de escala. Este análisis revela que las diferencias entre cada par de datos son menores al 5%, por lo que el histograma de velocidades de cualquier día es un buen representante del comportamiento que tiene la red en cualquier momento del tiempo de estudio.

Esta es una consecuencia importante que muestra que la operación del sistema metrobús

es muy estable, manteniendo condiciones de operación similares durante todos los días de la semana e incluso en situaciones tan variadas como las generadas por las restricciones de movilidad en la Ciudad de México durante los meses de abril y mayo de 2020. Se hace hincapié en que lo anterior se cumple sólo cuando se estudia la totalidad de la red, puede ser que esto no siempre sea verdad al estudiar una línea, un segmento o una hora en particular. Aunque, de acuerdo con los diferentes análisis realizados, se encuentra que para el caso de los segmentos sólo un número pequeño de estos muestran comportamientos diferentes al caso general.

Para apreciar claramente esta última afirmación, en la Figura 3.12 se presentan los histogramas diarios superpuestos de dos segmentos: el 8 de la línea 8 (Figura 3.12(a)), comprendido entre las estaciones: Garrido, de la línea 7, La Villa, de la línea 6, y Delegación GAM, Hospital Infantil la Villa y Misterios, de las líneas 6 y 7; y el segmento 9 de la línea 1 (Figura 3.12(b)), comprendido entre las estaciones Centro Cultural Universitario y Ciudad Universitaria de la línea 1; además se presentan sus correspondientes histogramas de distancias de Kullback-Leibler comparando el histograma total del segmento (histograma Q) y los histogramas de velocidad generados para cada uno de los días de estudio (histograma P). En el segmento 8 de la línea 8 se observa que la mayoría de los histogramas de velocidad caen dentro de una tendencia en torno al *histograma total del segmento*, y sólo unos cuantos se salen de dicha tendencia, alcanzando su único valor notable el día 67, que corresponde al 23 de abril, fecha en que la distancia alcanza un valor de 1.27, casi 5 veces mayor que la segunda distancia más alta, 0.276, registrada el día 98. Dados los resultados, se concluye que este es un ejemplo de segmento que en buena medida la actividad registrada en cada día se mantiene constante, en tanto que los histogramas diarios son similares al histograma total del segmento.

En el caso del segmento 9 de la línea 1, los histogramas de los primeros 67 días son muy parecidos al histograma total de este segmento y por lo tanto, tienen distancias pequeñas. Después de ese día, sucede algo que salta a la vista. Los histogramas cambian notablemente y los histogramas de los días siguientes muestran una alta similitud entre sí. Ahondando, en el día 67 se registra la distancia más alta, de 1.33 y por otro lado, a partir de ese día las distancias son siempre superiores a 0.45, pero entre estos días la diferencia suele ser menor a 0.2. Esto es, la actividad es diferente pero se estabiliza en torno a un comportamiento nuevo que puede ser denominado como una *nueva normalidad*. Este es un ejemplo de segmento que no mantiene los mismos histogramas de velocidad durante todo el tiempo de estudio. Se reitera en que estos casos constituyen un número pequeño comparado con el número total de segmentos.

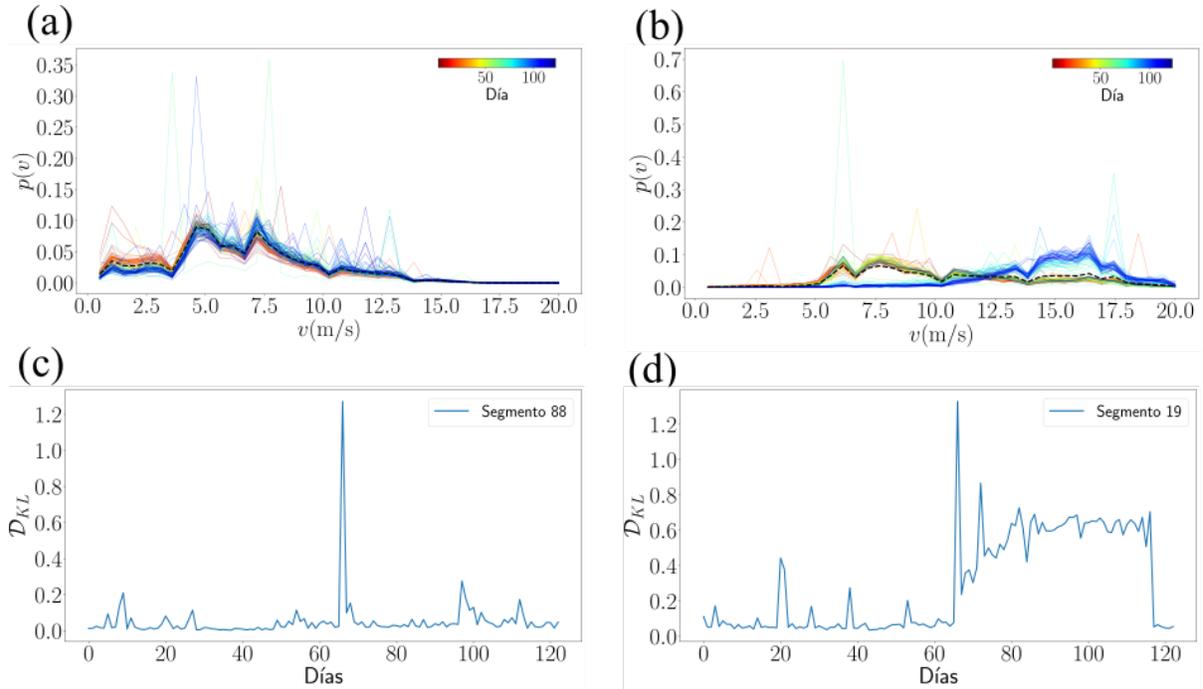


Figura 3.12: Comparación entre el *histograma total* y el histograma diario de velocidades para cada uno de los días del estudio, para dos segmentos en particular, (a) segmento 88 y (b) segmento 19. En (c) y (d), respectivamente, se presentan las distancias de Kullback-Leibler obtenidas a partir de la comparación del histograma diario con el histograma total de cada segmento (presentados con líneas punteadas en (a) y (b)).

Al extenderse este análisis a todos los segmentos del sistema metrobús, se concluye que, en general, todos los segmentos muestran comportamientos diarios en los que las diferencias entre las distribuciones diarias de velocidad y la distribución de todos los días tienen distancias de Kullback-Leibler que son menores a 0.5. Este dato aporta indicios para pensar que en un segmento dado, el histograma de velocidades de cualquier día, e incluso el de todos los días, es un buen representante de ese segmento.

En la Figura 3.13 se grafica el análisis estadístico de las distancias de Kullback-Leibler \mathcal{D}_{KL} obtenidas al comparar los histogramas diarios con su respectivo histograma total. Se presenta la densidad de probabilidad $\rho(\mathcal{D}_{KL})$ para cada uno de los segmentos del sistema metrobús, con la finalidad de obtener una mejor comprensión sobre la magnitud de esta distancia en toda la red. Al ser vistas estas densidades de probabilidad para todos los segmentos, es posible advertir que para la mayoría de los segmentos no hay distancias mayores a 0.5, y que el máximo de apariciones se encuentra en alrededor de 0.1. También se puede apreciar que para aquellos segmentos coloreados en tonalidades rojizas llegan a haber distancias de hasta 1.0, pero estos ejemplos son escasos. Un conteo permite ver que son menos de 10 segmentos los que registran distancias mayores a 0.6. Además, es útil

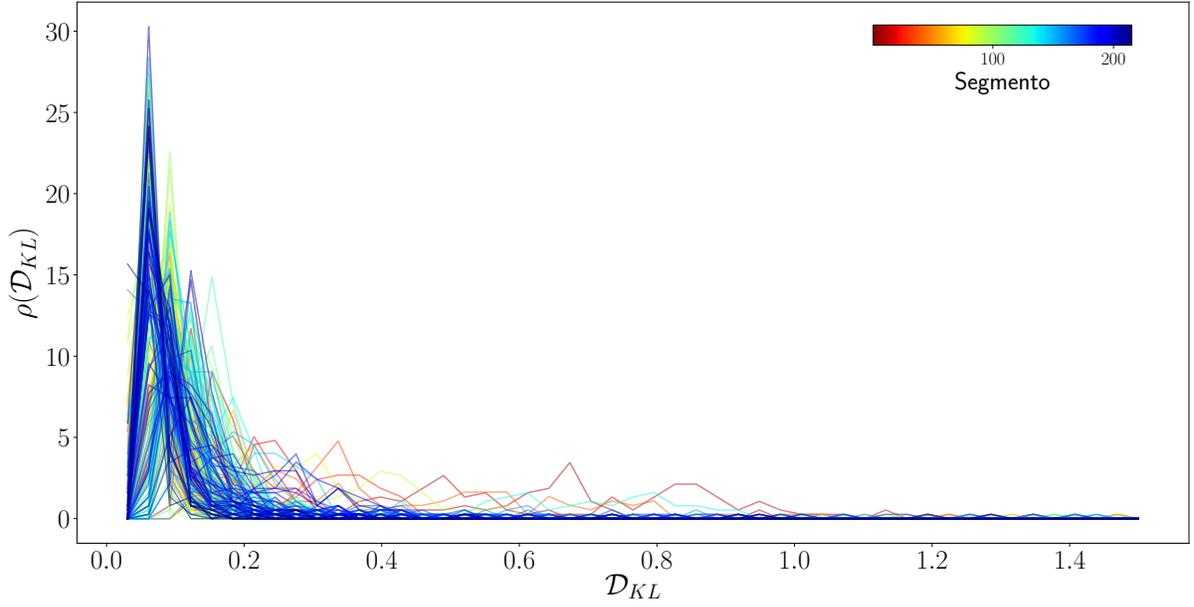


Figura 3.13: Densidad de probabilidad de las distancias de Kullback-Leibler \mathcal{D}_{KL} para cada uno de los segmentos, coloreados según la barra de color, se observa que el máximo de probabilidad está comprendido entre 0 y 0.2; y sólo para algunos segmentos existe probabilidad de encontrar distancias mayores a 0.6.

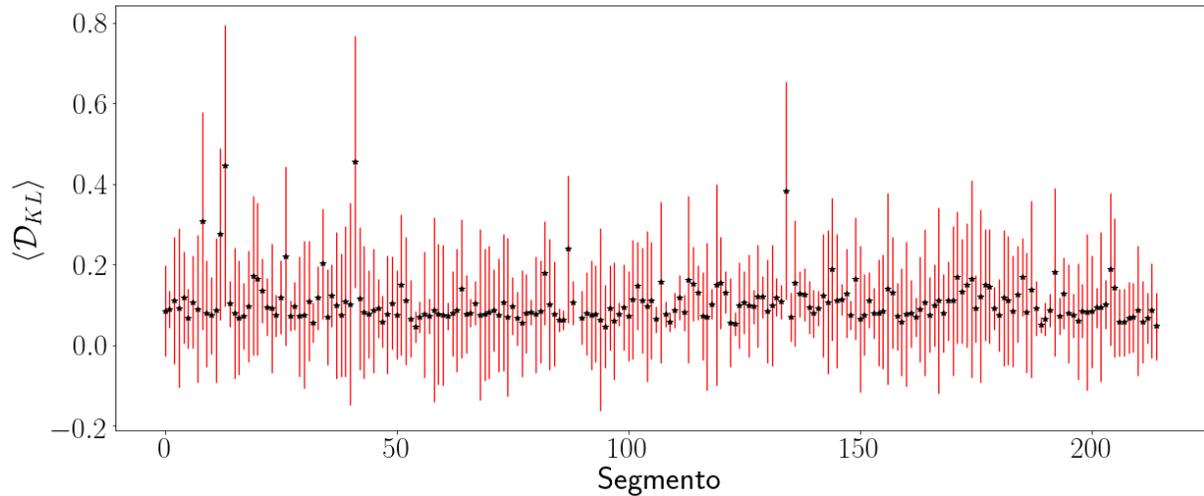


Figura 3.14: Valores promedio de las distancias de Kullback-Leibler $\langle \mathcal{D}_{KL} \rangle$ obtenidas al comparar el histograma diario con el histograma total de cada segmento. Las *barras de error* muestran las desviaciones estándar en la distribución de valores de las distancias \mathcal{D}_{KL} en cada uno de los segmentos.

mencionar que este número de segmentos representa un porcentaje menor al 5% de todo el sistema por lo que su incidencia en el comportamiento de la red puede considerarse marginal.

El análisis de los datos muestra que las distancias de Kullback-Leibler promedio en cada segmento son menores a 0.3, salvo en cuatro segmentos, a saber: el 9 de la línea 1, 14

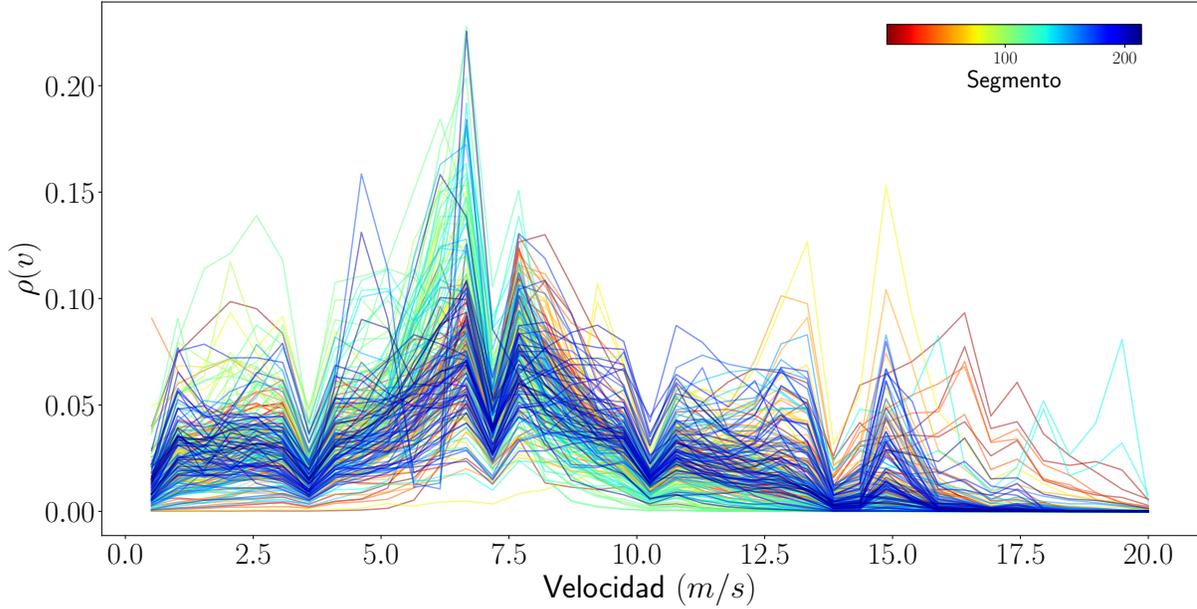


Figura 3.15: Densidad de probabilidad $\rho(v)$ para las velocidades v en cada uno de los segmentos. Cada histograma se construye utilizando los datos del segmento con toda de la base de datos cubriendo el periodo entre 16 de febrero de 2020 y 23 de junio de 2020.

de la línea 1, 1 de la línea 2 y 36 de la línea 4 Ruta Sur, cuyas distancias promedio respectivas son: 0.307, 0.445, 0.454 y 0.383. En la Figura 3.14 se presenta esta información gráficamente, en donde también se incluye las desviaciones estándar en cada segmento, representadas con barras rojas. Se puede observar que para casi todos los segmentos las desviaciones estándar son pequeñas, lo que indica que la distribución de esos datos es estrecha, y en el caso de los segmentos con distancias promedio mayores a 0.3 se observan distribuciones más anchas que el resto. Así, entre menor sea la barra roja, más constante será el comportamiento del segmento correspondiente.

Con lo presentado hasta ahora, se puede concluir que el histograma de velocidad de cualquier segmento es un buen representante de ese segmento; y este mismo razonamiento se puede aplicar si se analiza la distribución de velocidades de todo el sistema donde, de hecho, esta afirmación es aún más fuerte.

Para finalizar este capítulo, en la Figura 3.15 se grafica la densidad de probabilidad de las velocidades para cada uno de los segmentos de la red durante el tiempo de estudio, con el fin de detectar conjuntos de segmentos en los que se registren comportamientos similares. La diferencia con las Figuras 3.10 y 3.12 es que la búsqueda de similitudes se hacía entre los días de estudio. De esta manera, se buscará una conexión entre los comportamientos de los polígonos en que ha sido dividida la red, por ejemplo, saber si el

cambio de velocidad en un segmento repercute en el cambio de velocidad de otro segmento. A simple vista se observa que los segmentos representados en tonalidades azul fuerte muestran tendencias similares, que difieren de las tendencias registradas en los segmentos coloreados en tonalidades rojizas. Como se ha mencionado anteriormente, este análisis visual presenta limitaciones para otros enfoques de estudio, por lo que, nuevamente, se hará uso de la distancia de Kullback-Leibler, pero ya no se graficará en un histograma, sino en un arreglo bidimensional que facilitará la comparación entre las tendencias de varios segmentos a la vez. Este análisis se aborda en detalle en el siguiente capítulo.

Capítulo 4

Patrones en histogramas de velocidad

4.1. Introducción

En este último capítulo se presentan los resultados de un análisis de los datos apoyado en la teoría de las matrices y un análisis gráfico estos mismos datos, que sirve como método auxiliar al primero que se ha mencionado; este método gráfico hace uso de la teoría de redes. En la siguiente sección, se organizan en un arreglo matricial las distancias de Kullback-Leibler, mostradas en la Figura 3.15. Con esta forma de ordenar los datos se profundiza el análisis de las semejanzas y diferencias entre los segmentos realizado hasta ahora. En la sección 4.3, se define un *criterio de cercanía* con el que se construirán diferentes matrices de adyacencia al variar este valor y con ello termina el uso de la teoría de matrices.

En la sección 4.4, se convierten las matrices de adyacencia en grafos, en las que los nodos simbolizan los segmentos en los que se ha dividido la red y las aristas indicarán que esos nodos son cercanos entre sí. Teniendo las redes, se buscarán propiedades del sistema encontrando las de la red, tales como el grado promedio o el coeficiente de agrupamiento. Después, se hará uso de ciertos algoritmos para detectar la formación de comunidades dentro de cada una de las redes y se registrará la dinámica de las mismas conforme varía el criterio de cercanía.

Finalmente, se presentan estas comunidades en un mapa del sistema metrobús, en el que

se identifican por comunidades los segmentos implementados en este trabajo, de manera que se puedan apreciar fácilmente las presencias de las comunidades encontradas en el análisis de la red. Esto con la finalidad de enriquecer el análisis al establecer una relación entre las comunidades y su ubicación geográfica.

4.2. Matriz de distancias de Kullback–Leibler

Una forma de presentar las distancias de Kullback-Leibler entre las distribuciones de velocidad entre un segmento i y uno j es por medio de una matriz, pero en vez de valores numéricos, cada uno de los elementos se colorea según una escala de valores asociada. Así, se puede hacer un análisis visual que resulta más práctico que la sola comparación numérica. Esta *simetrización* se puede realizar promediando la distancia de Kullback-Leibler entre los segmentos i y j y la distancia j e i . Esta operación se reduce a sumar la matriz de distancias entre i y j con su matriz transpuesta y multiplicar el resultado por $1/2$, entonces:

$$\mathcal{D}_{KLS}(i, j) = \frac{1}{2} (\mathcal{D}_{KL}(i, j) + \mathcal{D}_{KL}(j, i)). \quad (4.1)$$

Con la ecuación anterior, se obtiene una matriz de $214 \times 214 = 45796$ que se muestra en la Figura ???. Dado que la matriz es simétrica, sólo se tiene que poner atención a los valores del triángulo superior o inferior con lo que se tienen únicamente 22791 valores útiles. A partir de la escala cromática a un costado de la matriz, se puede apreciar que gran parte de las distancias son menores a 2, dada la predominancia visual de los tonos oscuros. También se puede advertir la formación de *bloques* de tonos oscuros y *líneas* de tonos brillantes. La primera observación indica que la mayoría de los segmentos tienen distribuciones de velocidad cuyo comportamiento es muy similar a los demás y son pocos aquellos segmentos cuya distribución de velocidades difiere mucho, relativamente hablando, del resto de la red. De la segunda observación se puede inferir que existen segmentos consecutivos que muestran patrones de comportamiento similares, así como segmentos solitarios que destacan por tener una actividad diferente a la del resto de los tramos de la red. Esta segunda observación se tratará en detalle en las secciones siguientes.

El valor más grande de la matriz mostrada en la Figura 4.1 es 4.699, que corresponde al elemento $\mathcal{D}_{KLS}(75, 0)$; la fila 75 se refiere al segmento 37 de la línea 2, y la columna 0 al segmento 1 de la línea 1. El primer tramo está comprendido entre las estaciones Canal de San Juan y Nicolás Bravo y el segundo está comprendido entre las estaciones El Caminero y La Joya. De esta manera, se encuentra que durante el tiempo de estudio los histogramas

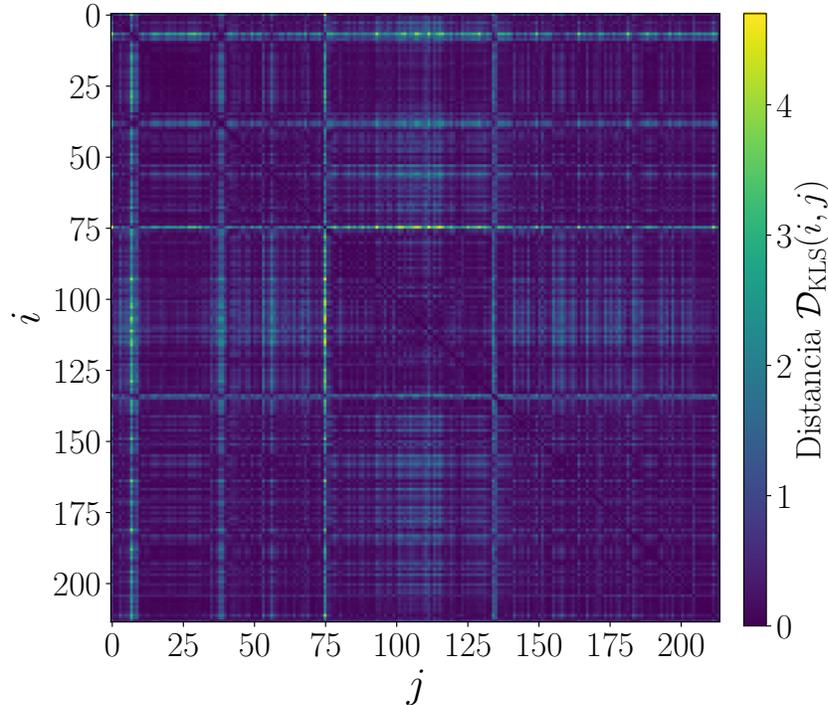


Figura 4.1: Matriz de distancias de Kullback–Leibler simétrica \mathcal{D}_{KLS} de las densidades de probabilidad correspondientes a todos los segmentos en la Figura ???. Los valores obtenidos se han sustituido por una escala cromática cuya referencia está a la derecha de la matriz.

de velocidad de estos dos segmentos, son los que registran la mayor diferencia de todo el sistema. De esto se concluye que lo que pasa en un segmento influye en lo mínimo a lo que pasa en el otro y viceversa.

En cuanto a la distancia más pequeña, esta tiene un valor de 0.00713567 y corresponde al elemento $\mathcal{D}_{KLS}(15, 24)$, esto es, la distancia entre el segmento 16 de la línea 1 y el segmento 25 de la misma línea. El primero, ubicado entre las estaciones José María Velasco y Teatro de los Insurgentes y el segundo ubicado entre las estaciones La Piedad y Nuevo León. Entonces, se tiene que este par de segmentos son los que muestran histogramas de velocidad con los patrones de comportamiento más similares de entre todos aquellos que conforman la división de polígonos implementada en esta investigación. Así, se deduce que lo que pasa en un segmento tiene una influencia apreciable en el otro, conclusión plausible considerando la cercanía geográfica entre ambos tramos de la red. Los valores máximos y mínimos adquieren una mejor perspectiva cuando se considera el valor promedio de las distancias de la red, que es 0.4225.

En la Figura 4.2 (a) se presenta gráficamente la similitud entre los segmentos 16 y 25 de la línea 1. Nótese cómo a cada incremento o disminución de velocidad en un segmento, le corresponde uno similar en el otro. Geográficamente hablando, ambos segmentos son

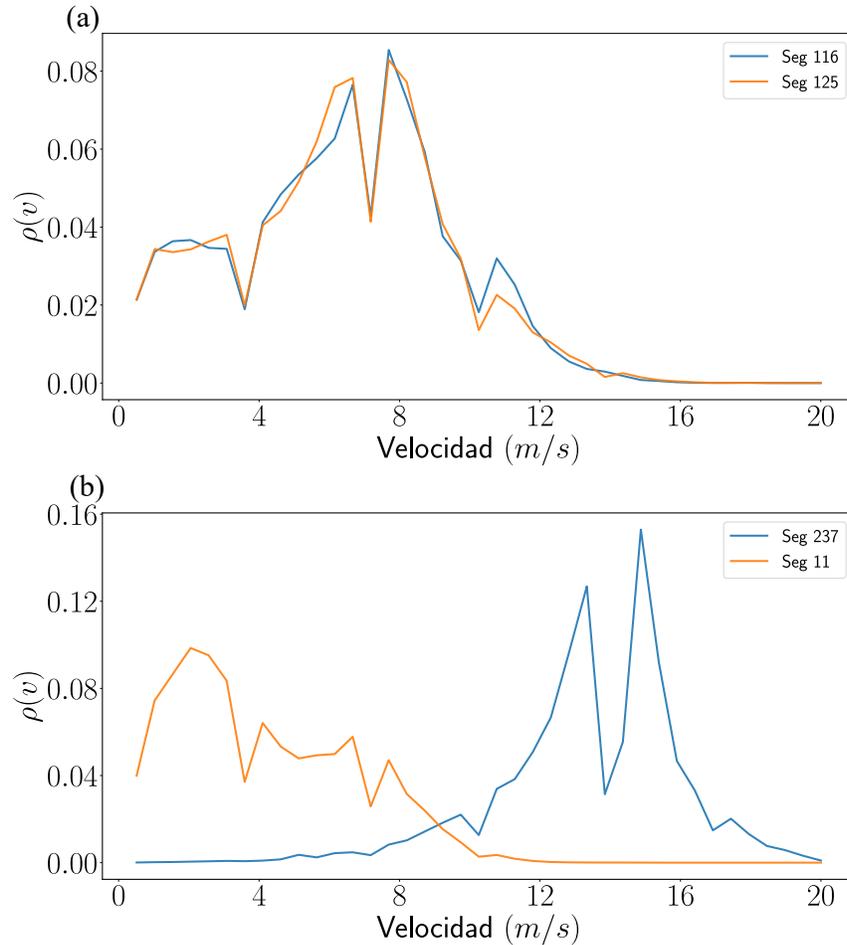


Figura 4.2: Comparación entre el par de segmentos que más se parecen en sus histogramas de velocidades (a) y entre el par con la menor distancia de Kullback–Leibler (b). Obsérvese cómo en (a), a cada aumento del segmento 116 le corresponde uno en 125 que es muy parecido y cómo el comportamiento en 237 y 11 no muestra similitudes.

muy cercanos, a una distancia de 4 kilómetros, ubicados dentro de lo que se conoce como *Insurgentes Sur*. Una zona conocida por ser sede de muchas oficinas y comercios, por lo que ambos segmentos son frecuentados en horarios similares por perfiles similares de usuarios. En cuanto a la comparación gráfica de segmentos 11 y 237, esta se muestra en la Figura 4.2 (b).

Retomando la observación hecha en el segundo párrafo de esta sección, en la fila número 75 (segmento entre las estaciones Canal de San Juan y Nicolás Bravo), se observa que los tonos predominantes son claros, por contraposición a sus filas vecinas con tonalidades oscuras. Ello indica que ahí las distancias de Kullback–Leibler simetrizadas son, mayormente, superiores a 2.0 y de hecho, es así, pues 124 de sus 214 valores lo son. Este conjunto de distancias cobra notoriedad si se considera que sólo 320 valores, dentro del conjunto relevante de datos, son mayores que esta cota. Esto es, este tramo de la red muestra una

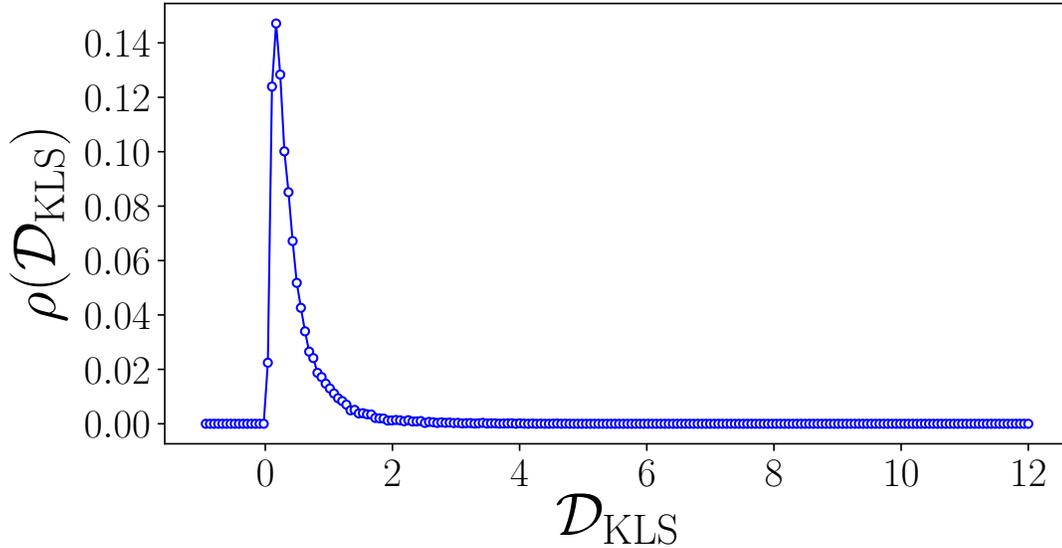


Figura 4.3: Densidad de probabilidad $\rho(\mathcal{D}_{KLS})$ de todas las distancias \mathcal{D}_{KLS} en la matriz presentada en la Figura 4.1.

gran diferencia en su patrón de velocidades con respecto a todos los demás segmentos en los que la red de metrobús ha sido dividida en este estudio. Algo similar ocurre con el segmento 18 de la línea 1, comprendido entre las estaciones Perisur y Centro Cultural Universitario; en la que 128 de sus valores son mayores a 2.0. Entonces, se observa que estos dos segmentos se ven poco afectados por lo que pase en gran parte de la red. En la Figura 4.3 se muestra la distribución de distancias de Kullback-Leibler en toda la matriz, donde se puede apreciar que el máximo de valores se sitúa en torno a 0.5, siendo este el 14% de todos los valores y casi la totalidad de las distancias están entre 0 y 2. Considerando este dato, la moda de esta distribución y la consideración anterior sobre los segmentos 18 de la línea 1 y 37 de la línea 2, se puede concluir que existen diferencias importantes entre algunos de los histogramas de los segmentos.

Ahora bien, si lo que se plantea es la distancia entre un segmento en particular y el resto de la red, es buena estrategia calcular el promedio de las distancias en la fila correspondiente a ese segmento. De esta manera, se encuentra que la distancia promedio más grande es 2.176, correspondiente al segmento 38 de la línea 2, entre las estaciones Nicolás Bravo y Tepalcates, por lo que este es el segmento cuya actividad promedio tiene la mayor diferencia con el resto de la red. Por otro lado, la distancia promedio más pequeña es 0.2185, que corresponde al segmento 5 de la línea 8, formado por la intersección de la línea 3 con la 6. Por lo tanto, se encuentra que este es el segmento cuya actividad se parece más a la de los demás segmentos.

Dado que las distancias entre histogramas asociados a segmentos se han calculado tomando en cuenta todo el tiempo de estudio, se puede pensar que entre aquellos segmentos en los que la distancia sea menor puede existir una conexión que explique tal similitud. Al igual que el propósito de la matriz en la Figura 4.1, presentar estas conexiones de forma gráfica ayudaría a entender más fácilmente lo que está pasando. Por lo que es preciso buscar las comunidades que se forman en el sistema metrobús haciendo uso de la teoría de redes. Antes de pasar a ello, es necesario crear una matriz de adyacencia a partir de la matriz simetrizada.

4.3. Redes de similitud entre segmentos

Dado que una matriz de adyacencia equivale a una red y el propósito de este trabajo es obtener una red gráfica a partir de las matrices ya obtenidas, es necesario establecer un criterio para saber cuáles elementos de \mathcal{D}_{KLS} se convertirán en valores nulos y cuáles en no nulos. En este trabajo se considerará que 1 representa cercanía y 0 lejanía, entre cada par de segmentos. Para saber qué valor de la matriz simetrizada representa cercanía o lejanía, se propone una *distancia mínima* h . Así, elementos menores a h se tomarán como cercanos y se les asignará valor 1, mientras que para valores mayores se les considerará como lejanos y su valor será 0. En la Figura 4.4 se presentan 6 matrices de adyacencia obtenidas a partir distintos valores de h , tomados del conjunto $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.35\}$. Estos valores fueron elegidos considerando la distribución de valores mostrada en la Figura 4.3, tomando en cuenta que el promedio de las distancias entre pares de segmentos es 0.4225.

En las matrices de adyacencia de la Figura 4.4 se observa cómo a medida que h aumenta, aumenta el número de entradas con valor no nulo. Esto es, a medida que se aumenta h se considera como *cercanos* a un número cada vez mayor de elementos.

Siguiendo el procedimiento descrito en la Sección 1.4.3, se construyen las redes asociadas a las distintas matrices de adyacencia obtenidas. Con esta nueva representación gráfica se obtiene una perspectiva que complementará al estudio realizado hasta ahora. En la Figura 4.5 se muestra la red generada por $h = 0.05$. En ella se aprecia la existencia de 42 vértices aislados del resto de la red. Para estos se infiere que son segmentos cuyo comportamiento es diferente al del resto de la red y corresponden a filas compuestas únicamente de valores nulos en la matriz de adyacencia. También, se observan *cadena*s de vértices que sólo tienen aristas con otros dos vértices sin que estos últimos tengan conexión entre sí. La diferencia entre este tipo de segmentos y aquellos que están aislados es que tienen una relación

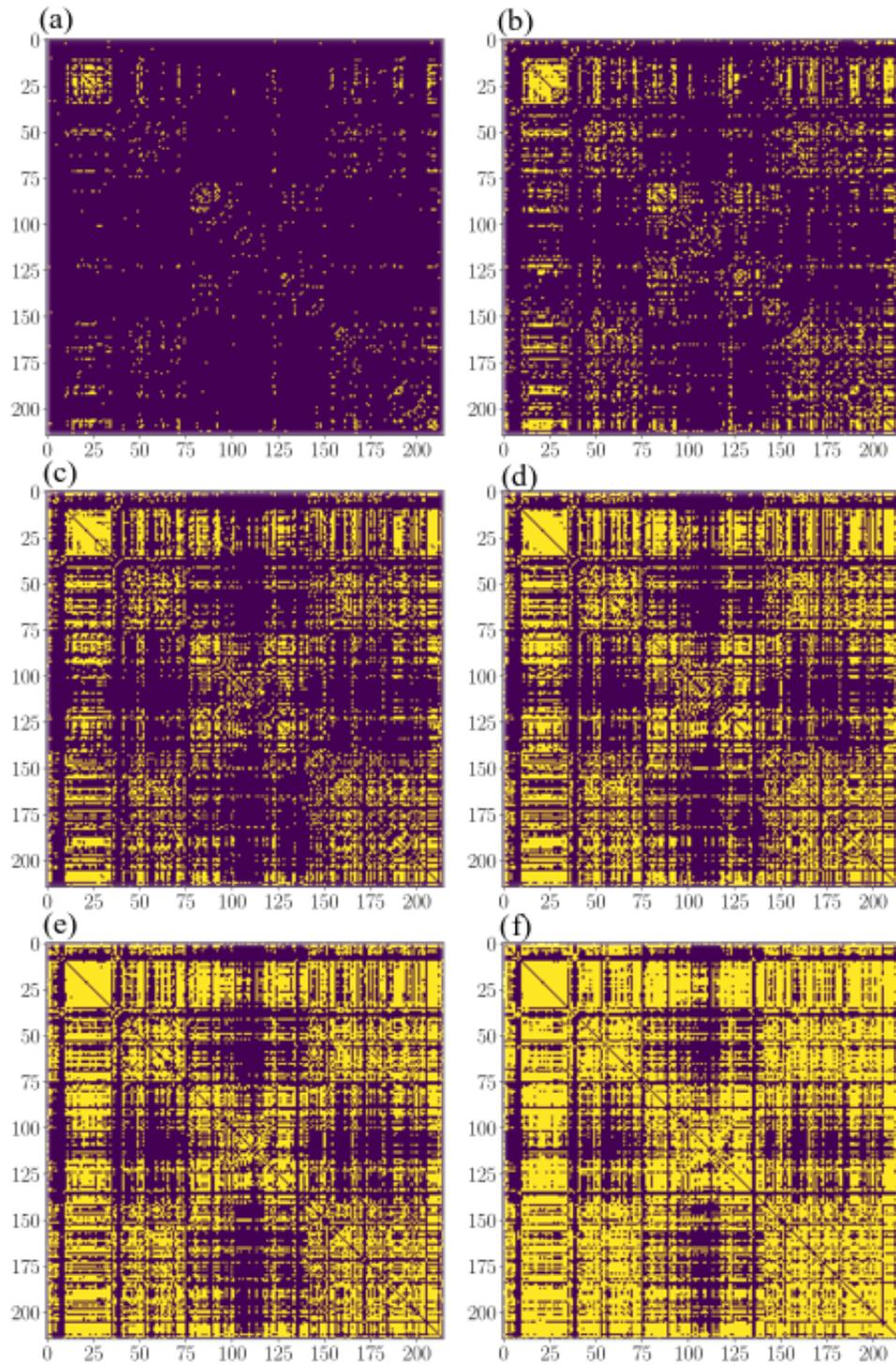


Figura 4.4: Matrices de adyacencia graficadas según la distancia mínima h : (a) $h=0.05$, (b) $h=0.1$, (c) $h=0.15$, (d) $h=0.2$, (e) $h=0.25$, y (f) $h=0.35$. En estas matrices, el color violeta corresponde a elementos de la matriz cuyo valor es nulo, y el color amarillo a aquellos elementos con valor no nulo.

exclusiva con otros segmentos. Como se advierte, la información que este tipo de vértices

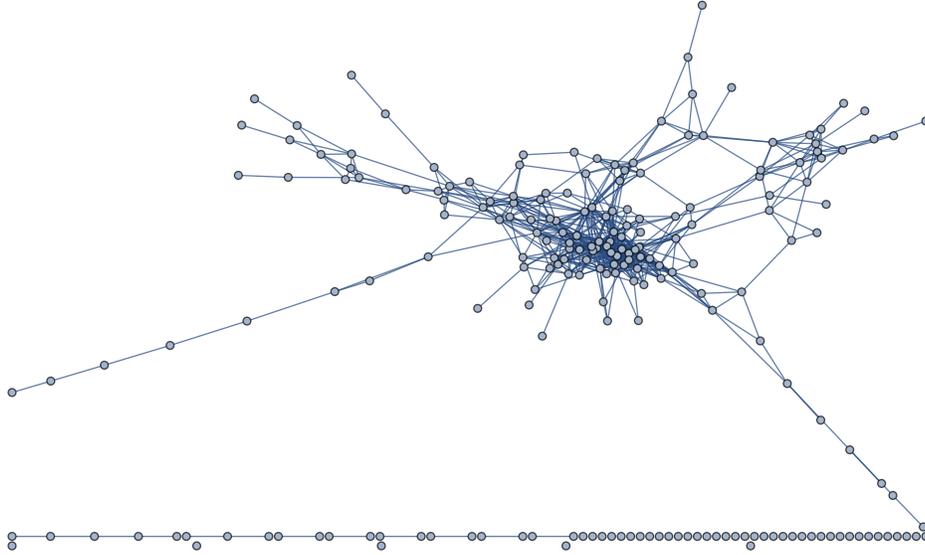


Figura 4.5: Red de similitud generada con $h = 0.05$.

aportan al estudio no es relevante, en comparación con aquella que ofrece el conjunto de vértices que está densamente conectada. Este conjunto se compone de 152 vértices y entre estos se encuentra el grado máximo de esta red, cuyo valor es $k_{max} = 33$, correspondiente al segmento comprendido entre las estaciones Río Mayo y Rojo Gómez de la línea 2. Esto es, dicho segmento es el que guarda mayor similitud con el mayor número de segmentos de la red.

En la siguiente parte, el enfoque del estudio se hace tomando aquella porción de la red que está densamente conectada y a ella se le denomina *componente gigante*. En la Tabla 4.1 se enlista el número de nodos que hay en la componente gigante. Nótese el incremento de nodos entre $h = 0.05$ y $h = 0.1$ entonces, en tales circunstancias pequeños cambios en h significa un aumento grande en el número de segmentos que guardan parecido con otros. También se observa que para $h < 0.25$ hay dos segmentos que tienen distancias de Kullback-Leibler lo suficientemente grandes como para ser considerados *distantes* en términos de sus patrones de velocidad. Estos dos segmentos son el 37 de la línea 2 y el 18 de la línea 1.

En la Tabla 4.1 también se enlistan los grados promedio de las distintas componentes gigantes. Esto dice que, en promedio, los segmentos cuando $h = 0.05$ son cercanos a otros 6 y segmentos. Esta cantidad sube a casi 28 para $h = 0.1$ y a casi 53 cuando $h = 0.15$. En este punto es interesante señalar que, en promedio, cada segmento es cercano la cuarta parte de todos los segmentos en la componente gigante. En este mismo tenor, si $h = 0.2$, un

segmento cualquiera será cercano a otros segmentos que constituyen la tercera parte de la componente gigante. Así, al final se observa que para $h = 0.35$, un segmento será cercano a otros 125, que es algo más de la mitad de los 214 segmentos totales. Esta observación es importante y su naturaleza se dilucida con claridad en la siguiente sección.

En la última columna de la Tabla 4.1 se listan los coeficientes de agrupamiento promedio de toda la red. Para concluir esta sección, se menciona que las observaciones sobre las submatrices y líneas en $\mathcal{D}_{\mathcal{KLS}}$, así como la mención sobre el porcentaje del grado promedio en la componente gigante, obedecen a la formación de comunidades dentro de la red; tema que finalmente será abordado en la siguiente sección.

4.4. Comunidades en la red de similitud

Usando el algoritmo *modularidad* en las componentes gigantes del conjunto de redes obtenidas para diferentes valores de h , se encuentran las comunidades que se forman dentro de estas subredes. En la Figura 4.6 se muestran las comunidades obtenidas.

Para $h = 0.05$ la componente gigante de la red de similitud tiene 7 comunidades que se describen en la Tabla 4.2. Para $h = 0.1$, la componente gigante aumenta a 200 elementos, lo que significa que a esta se han incorporado nodos de aquellas comunidades previamente descartadas. Además, el número de comunidades se reduce a 5, siendo \mathcal{C}_6 y \mathcal{C}_7 quienes fueron absorbidas por \mathcal{C}_1 , \mathcal{C}_2 y \mathcal{C}_3 , cuyos elementos aumentaron a 92, 68 y 33, respectivamente. En tanto que algunos elementos de \mathcal{C}_5 y \mathcal{C}_4 se integraron a las tres primeras comunidades centrales, reduciéndose de 12 a 2 elementos en el caso de \mathcal{C}_5 y de 15 a 11 en el de \mathcal{C}_4 . En lo concerniente a los puentes entre comunidades, las tres primeras se conectan

Tabla 4.1: Propiedades de las redes generadas a partir de las matrices de la Figura 4.4, en las que se presenta el número de comunidades detectadas, el número de nodos en la componente gigante, el grado promedio y el coeficiente de agrupamiento de la red.

h	Nodos	Grado Promedio	Coficiente de Agrupamiento
0.05	152	6.579	0.3078
0.1	200	27.654	0.5647
0.15	209	52.729	0.6628
0.2	211	75.430	0.7232
0.25	212	94.561	0.7584
0.35	214	124.953	0.819

entre sí; \mathcal{C}_4 sólo tiene conexión con \mathcal{C}_2 y \mathcal{C}_5 sólo lo hace con \mathcal{C}_1 .

Si $h = 0.15$, las comunidades se reducen a 3: \mathcal{C}_1 , \mathcal{C}_2 y \mathcal{C}_3 , que reúnen un total de 209 nodos. El número de elementos es 105, 99 y 5, respectivamente. Se observa que \mathcal{C}_1 y \mathcal{C}_2 aumentan su tamaño, absorbiendo a \mathcal{C}_4 y \mathcal{C}_5 e integrando 28 nodos de \mathcal{C}_3 . Finalmente, cuando $h > 0.2$, sólo hay 2 comunidades en toda la red: \mathcal{C}_1 y \mathcal{C}_2 . En particular, para $h = 0.35$, todos los nodos de la red están en una u otra comunidad, siendo 108 de estos nodos elementos de \mathcal{C}_1 y 106 nodos de \mathcal{C}_2 , aproximadamente la mitad en cada una.

4.5. Comunidades en el sistema Metrobús

Ahora que se han identificado los segmentos que pertenecen a las distintas comunidades encontradas en cada variación h es conveniente identificar estas comunidades en el mapa procesado en QGIS. Para ello se colorean los segmentos según la comunidad a la que pertenecen y finalmente se identifican por medio de sus histogramas de actividad, cuáles son las características en común entre cada segmento y cuáles son las diferencias entre cada comunidad de la componente gigante.

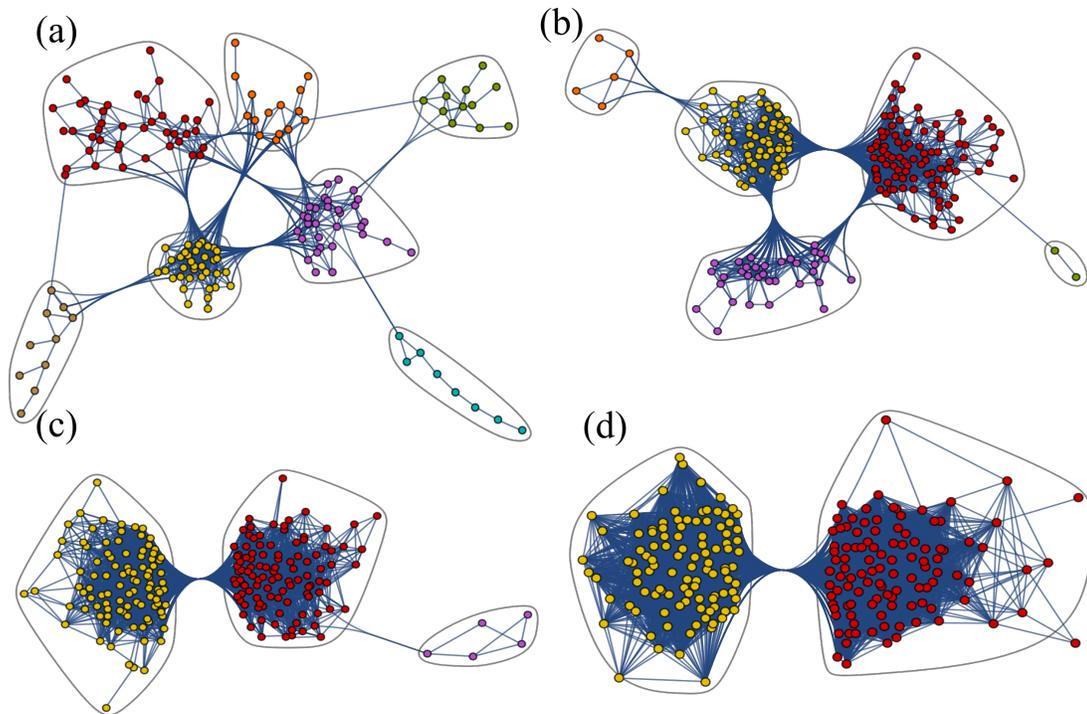


Figura 4.6: Comunidades detectadas en la componente gigante de la red cuando $h = 0.05$ (a), $h = 0.1$ (b), $h = 0.15$ (c) y $h = 0.35$ (d).

En la Figura 4.7 (a) se muestran nuevamente a \mathcal{C}_1 y \mathcal{C}_2 cuando $h = 0.35$. En (b) se muestra la división por segmentos del metrobús, en el que se ha indicado a qué comunidad pertenece cada uno de los tramos en que se ha dividido a la red. Sobre esto último, se observa que hay segmentos de ambas comunidades en todas las líneas, pero hay predominancia de una u otra comunidad en cada una de ellas. Por ejemplo, \mathcal{C}_1 tiene predominancia en las líneas 1, 2, 5 y 7, mientras que \mathcal{C}_2 la tiene en las líneas 3 y 6, mientras que en las líneas 4 y 8 hay un equilibrio entre ambas. De ambas comunidades, también se puede decir que tienen una cierta tendencia a formar largos tramos. Ejemplo de ello son las partes sur y norte de la línea 1, la *ruta sur* de la línea 4 y la parte centro-norte de la línea 7, además del tramo entre la estación Archivo General de la Nación y Terminal 2, de la línea 4. \mathcal{C}_2 lo hace en la parte sur de la línea 3 y en los extremos de la línea 6.

En las Figura 4.7 (c) y (d) se presentan los histogramas de de velocidad para los segmentos que componen a \mathcal{C}_1 y \mathcal{C}_2 , respectivamente. En cada uno de ellos se ha indicado con línea gruesa y punteada el *histograma promedio* respectivo. La actividad de \mathcal{C}_1 muestra que la distribución de velocidades es ancha, esto es, tiene un porcentaje significativo para todas las velocidades en el rango de estudio. Se observan tres máximos destacables: un primer máximo en torno a los 7 m/s, un segundo en torno a los 8 m/s y un tercer máximo cercano a los 15 m/s; este tercer máximo se destaca porque aparece en medio de una tendencia a la baja en la densidad de velocidades justo después del segundo máximo. Los dos primeros máximos mencionados tienen una distribución similar ya que, en promedio, estas dos velocidades ocurren un 7% para cada una y para el tercer máximo, su velocidad asociada ocupa en promedio el 3% de las velocidades registradas en los segmentos. Sobre la actividad de \mathcal{C}_2 se menciona que su distribución de velocidades en todos los segmentos no tiene cantidades significativas más allá de los 9 m/s siendo siempre menor al 5% y con una tendencia monótona decreciente. Además, en los dos máximos que se observan, hay segmentos que registran hasta el 25% de sus registros de velocidad en a los 6 m/s y hasta el 15% en velocidades cercanas a los 7.5 m/s. El promedio muestra una distribución más estrecha y más allá de los 12 m/s, es menor al 2.5%. En este caso, el histograma promedio exhibe dos máximos notables: el primero situado en torno a los 6.5 m/s que concentra cerca del 10% de los registros de velocidad; después se observa una tendencia a la baja, salvo en el segundo máximo, que se registra en torno a los 7.5 m/s. De esta manera, puede afirmarse que una de las características de \mathcal{C}_2 , es que es una comunidad de *baja velocidad*. Por otra parte, teniendo en cuenta las características de los histogramas en \mathcal{C}_1 , puede decirse que es una comunidad de *alta velocidad*.

De manera análoga a la Figura 4.7, en la Figura 4.8 (a) se presentan las comunidades de la

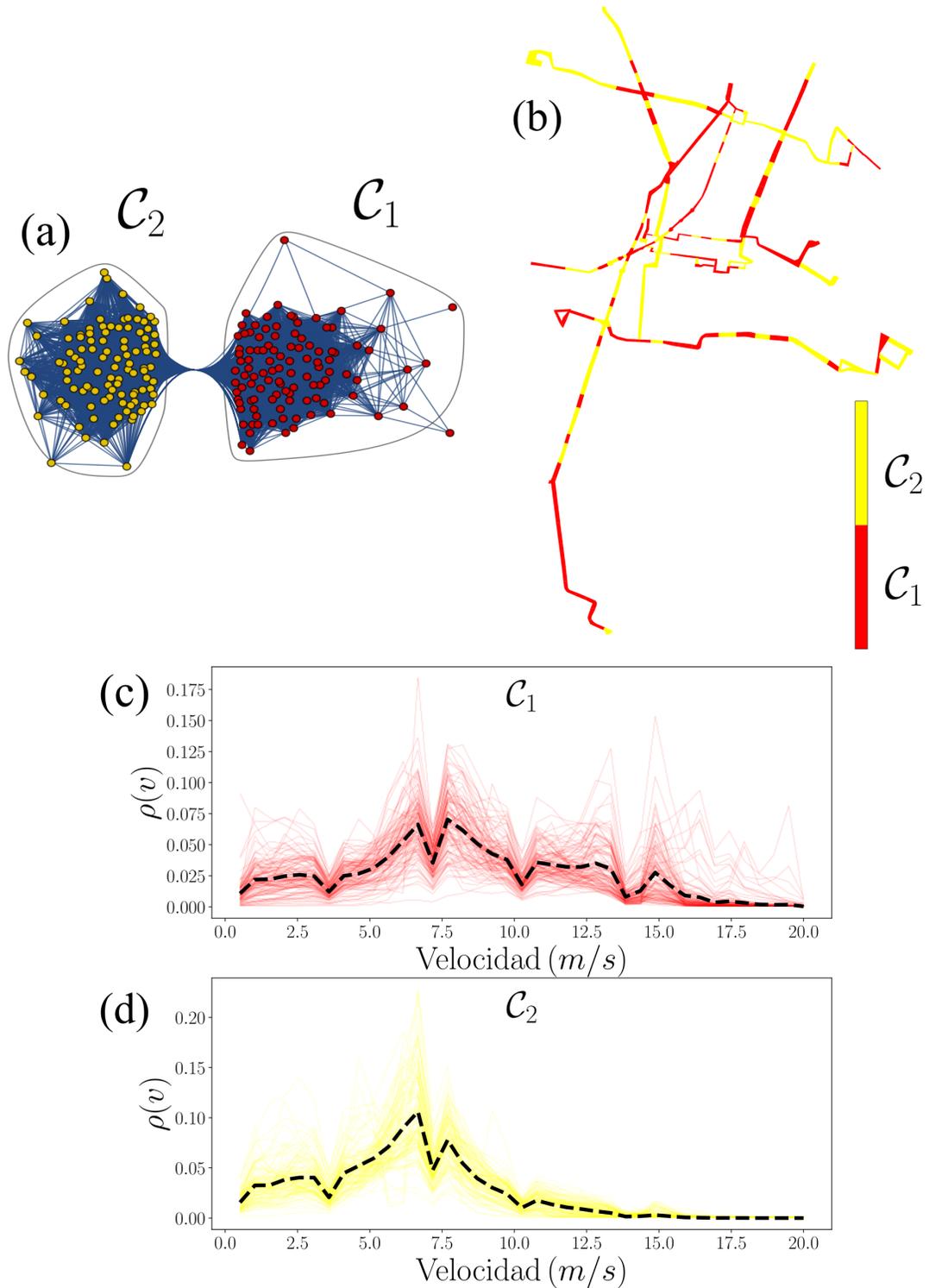


Figura 4.7: Detección de comunidades en el sistema Metrobús con $h = 0.35$: (a) Componente gigante de la red cuando $h = 0.35$, (b) visualización en el mapa de las comunidades en (a), se conserva la simbología de color en este mapa; (c) *histogramas totales* de las comunidades de la componente gigante.

Tabla 4.2: Propiedades de las comunidades encontradas con $h = 0.05$.

Comunidad	Nodos	$\langle v \rangle$ (m/s)
\mathcal{C}_1	40	5.56
\mathcal{C}_2	36	4.65
\mathcal{C}_3	31	6.04
\mathcal{C}_4	15	5.41
\mathcal{C}_5	12	7.52
\mathcal{C}_6	10	3.67
\mathcal{C}_7	8	6.94

componente gigante cuando $h = 0.05$. En este caso se forman 7 comunidades, etiquetadas según su tamaño como: $\{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5, \mathcal{C}_6, \mathcal{C}_7\}$, cuyo número de nodos integrantes se muestra en la Tabla 4.2. También se enlistan las velocidades promedio de cada una de las comunidades. \mathcal{C}_5 tiene la velocidad promedio más alta y \mathcal{C}_6 la más baja.

En la Figura 4.8(b) se muestra el mapa del sistema Metrobús con la simbología de colores de (a) para identificar los segmentos que pertenecen a cada comunidad. En este mapa, aquellos segmentos sin color pertenecen a comunidades fuera de la componente gigante. Este tipo de segmentos aparecen en todas las líneas y en particular, se observa una predominancia en la línea 1 donde se advierte la formación de largos tramos de segmentos *blancos*, en las zonas sur y norte de esta línea, con lo que se empieza a apreciar con claridad que gran parte de los elementos de estas comunidades son de baja velocidad. Se observa que, en general, todas las líneas tienen una composición variada de colores. Así, \mathcal{C}_7 concentra la mayoría de sus componentes en la línea 4, en donde forma un tramo continuo en su zona oriente, desde las estaciones Archivo General de la Nación hasta la Terminal 2 y sólo aparece en unos pocos segmentos de las líneas 2 y 7. En cuanto a \mathcal{C}_1 , esta muestra una predominancia en la zona oriente de la línea 2, de manera opuesta a la línea 1 en donde sólo aparece en un segmento y a la línea 5, en donde no hay ningún segmento que pertenezca a esta comunidad. En las demás líneas aparece en proporción similar a otras comunidades, sin destacar en alguna en especial. Aunado a lo anterior, observando los histogramas de velocidad para comunidad, se observa que de \mathcal{C}_1 a \mathcal{C}_4 y \mathcal{C}_6 tienen distribuciones estrechas y cargadas a velocidades bajas, en cambio, \mathcal{C}_5 y \mathcal{C}_7 muestran distribuciones más anchas. De estas dos últimas observaciones se deduce que conforme van surgiendo sólo dos comunidades, \mathcal{C}_1 , \mathcal{C}_5 y \mathcal{C}_7 conforman la comunidad de alta velocidad mientras que \mathcal{C}_2 , \mathcal{C}_3 , \mathcal{C}_4 y \mathcal{C}_6 conforman la comunidad de baja velocidad.

Finalmente, es importante comentar que los segmentos más largos tienden a ser de alta

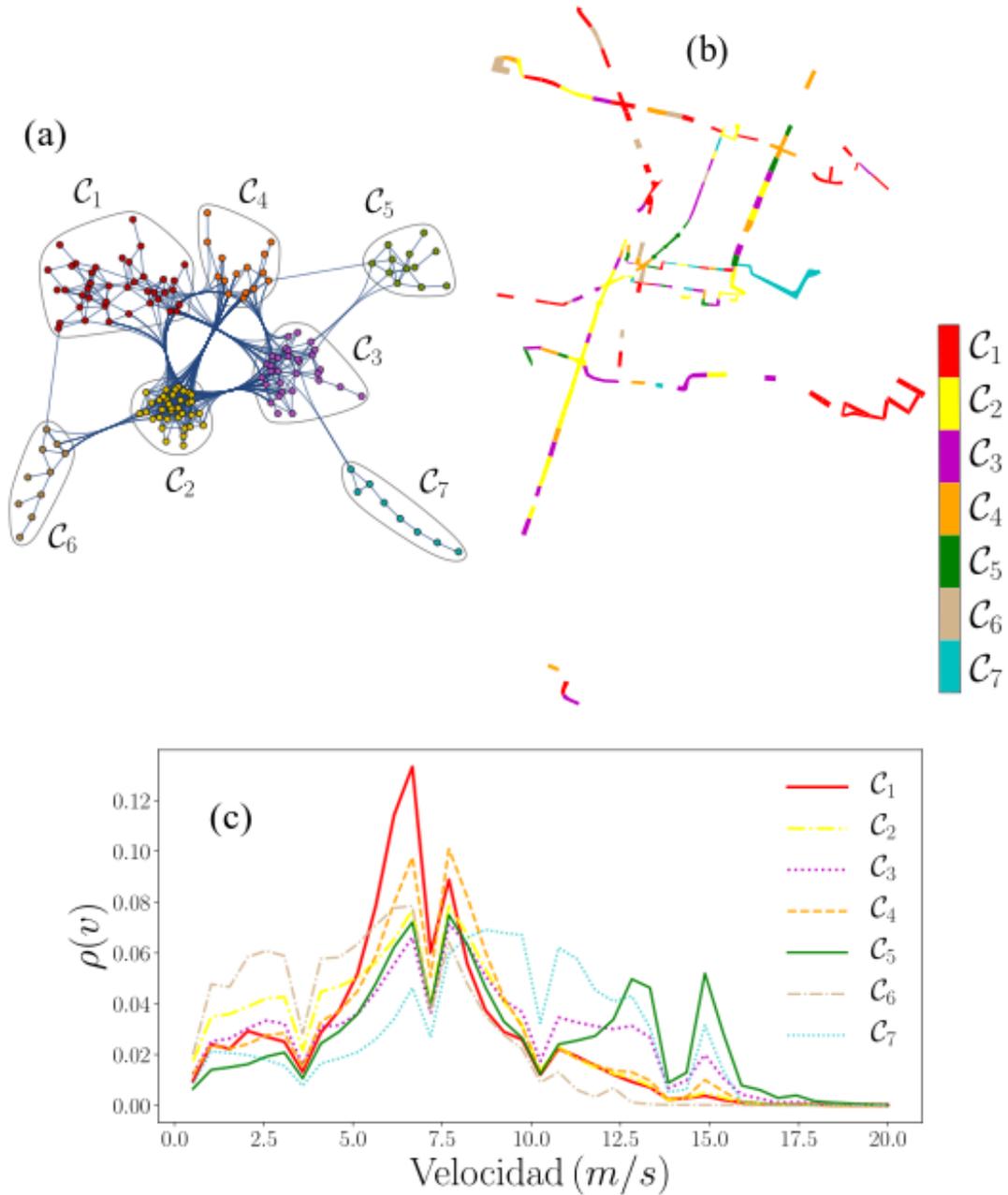


Figura 4.8: Detección de comunidades en el sistema Metrobús con $h = 0.05$: (a) Componente gigante de la red cuando $h = 0.05$, (b) visualización en el mapa de las comunidades en (a), se conserva la simbología de color en este mapa; (c) *histogramas totales* de cada una de las comunidades de la componente gigante durante el tiempo de estudio.

velocidad. Por ejemplo, en el caso de la línea 5, 11 de sus 15 segmentos también son de alta velocidad. De esta línea, se destaca el que sea la menos transitada del sistema. Por otro lado, los segmentos más céntricos del sistema Metrobús tienen una propensión a ser parte de la comunidad de baja velocidad. Ambas observaciones forman parte de una ten-

dencia que conviene tomar como punto de partida para explicar la distribución geográfica de las comunidades y las características de estas. De tal manera, se puede deducir que los vehículos alcanzan altas velocidades cuando la distancia entre las estaciones es grande y al desplazarse por zonas con menor densidad de usuarios, en el caso contrario, cuando los vehículos se desplazan por el centro de la ciudad. Por tanto, con una alta densidad de usuarios, los operadores se ven obligados a reducir su velocidad. Este comportamiento se mantiene casi sin cambios incluso cuando el tránsito general en toda la ciudad disminuyó a causa de las restricciones de movilidad por el COVID-19, por lo que estos resultados son otra manifestación del comportamiento estable del sistema.

Conclusiones

A partir del estudio de los datos de movilidad recolectados durante seis meses, los análisis desarrollados en este trabajo de investigación muestran que el sistema Metrobús es un sistema que opera con regularidad en cada uno de sus segmentos, esto es, se observa que los límites de velocidad de operación y las tendencias de velocidad tanto en todo el sistema como en cada uno de los segmentos se conservan, presentado pequeñas variaciones dependiendo del día o la hora en que se realice el análisis. Más aún, la aplicación de la distancia de Kullback-Leibler en las distribuciones de velocidad anteriormente mencionadas indica que el patrón de velocidades de todo el sistema Metrobús, tomado en cualquier día, y el patrón de velocidades de un solo segmento son prácticamente indistinguibles, salvo las excepciones expuestas en el texto, las cuales merecen ser estudiadas más exhaustivamente para determinar qué las hace diferentes. Esta regularidad puede estar relacionada con que el sistema Metrobús cuenta con carriles exclusivos para su funcionamiento, de esta manera, los histogramas de cada segmento son un buen referente del comportamiento específico de los vehículos en estas partes del sistema.

Por otro lado, se reporta la aparición de comunidades compuestas por los segmentos en que fue dividida la red del sistema. El número de comunidades varía según los criterios de cercanía que se impongan. Las características de las comunidades están estrechamente relacionadas con la ubicación geográfica y las estaciones que conforman a los segmentos. Ello permite relacionar la velocidad de los vehículos con la distancia entre las estaciones y las zonas por las que estos transitan.

Como conclusión general, se observa que el sistema Metrobús es una red robusta que se mantiene estable incluso con las restricciones de movilidad implementadas para contrarrestar la pandemia de COVID-19 en Ciudad de México. Parece plausible asociar este comportamiento estable con las reglas en el servicio que deben cumplir los operadores de las unidades del sistema Metrobús.

Perspectivas a futuro

Como se mencionó al principio del presente trabajo, la intención primordial fue hacer un estudio del Metrobús caracterizando la actividad de los vehículos en diferentes partes, utilizando métodos de la ciencia de redes. Este estudio puede ser profundizado para conocer con mayor detalle las características aquí esbozadas. A futuro, podría buscarse el por qué de la existencia de sectores singulares que muestran un patrón de actividad distinto al del resto del sistema. Por otro lado, para extender el propósito de esta investigación, podrían conseguirse datos sobre el número de pasajeros que usan el sistema y así complementar el estudio que aquí se ha presentado por medio del análisis de matrices de origen-destino, que son comunes en estudios de movilidad.

También es importante explorar los métodos desarrollados en esta investigación al análisis de otros modos de transporte masivo, por ejemplo, taxis, buses, servicios tipo Uber, entre otros. En estos casos, mediante un enfoque de redes multicapas, se pueden diseñar estrategias para una integración funcional de todo el transporte público de la ciudad. Este tipo de análisis podrían ser aplicados en sistemas de transporte de cualquier ciudad del mundo.

Bibliografía

- [1] Juval Portugali. *Complexity, cognition and the city*. Springer, 2011.
- [2] Juan de Dios Ortúzar and Luis Willumsen. *Modelling Transport*. Wiley, fourth edition, 2011.
- [3] Billie Giles-Corti, Anne Vernez-Moudon, Rodrigo Reis, Gavin Turrell, Andrew L. Dannenberg, Hannah Badland, Sarah Foster, Melanie Lowe, James F. Sallis, Mark Stevenson, and Neville Owen. City planning and population health: a global challenge. *The Lancet*, 388(10062):2912–2924, 2016.
- [4] Carlos Gershenson and Luis A. Pineda. Why does public transport not arrive on time? The pervasiveness of equal headway instability. *PLoS ONE*, 4(10), 2009.
- [5] A. P. Riascos and José L. Mateos. Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City. *Scientific Reports*, 10(1):1–14, 2020.
- [6] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE*, 7(5), 2012.
- [7] Samiul Hasan, Christian M. Schneider, Satish V. Ukkusuri, and Marta C. González. Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics*, 151(1-2):304–318, 2013.
- [8] Jean-Paul Rodrigue. *The Geography of Transport Systems*. Routledge, New York, fifth edition, 2020.
- [9] Organización Marítima Mundial. El transporte marítimo: indispensable para el mundo. <https://www.imo.org/es/MediaCentre/PressBriefings/Paginas/47-WMD-theme-2016-.aspx>, 2016.

-
- [10] Michael Batty. *The new science of cities*. The MIT Press, 2012.
- [11] Ming Zhang. Bus versus rail: Meta-analysis of cost characteristics, carrying capacities, and land use impacts. *Transportation Research Record*, (2110):87–95, 2009.
- [12] Ben Brooksbank. Metro Londres. https://commons.wikimedia.org/wiki/File:Barons_Court_Station_-_geograph.org.uk_-_1762897.jpg, 1962.
- [13] Daniel Osorio. Tren Ligero CDMX. https://commons.wikimedia.org/wiki/File:Tren_ligero_mexico_DF.JPG, 2008.
- [14] Gobierno de La Paz. Teleférico. https://commons.wikimedia.org/wiki/File:Línea_Roja_de_Mi_Teleférico_en_La_Paz,_Bolivia.jpg, 2016.
- [15] Steve Morgan. Ruta 100. https://commons.wikimedia.org/wiki/File:Ruta_100_bus_in_Tláhuac_1995.jpg, 1995.
- [16] Staten Island Ferry. Ferry. https://commons.wikimedia.org/wiki/File:Spirit_of_America_-_Staten_Island_Ferry.jpg, 2012.
- [17] Wikimedia Commons. Tranvía. https://commons.wikimedia.org/wiki/File:Bombardier_Flexity_Outlook_Cityrunner_1.jpg, 2007.
- [18] Marco Gonzalez-Navarro and Matthew A. Turner. Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108:85–106, 2018.
- [19] Garry Glazebrook and Peter Newman. The city of the future. *Urban Planning*, 3(2):1–20, 2018.
- [20] Juan Pablo Bocarejo, Ingrid Joanna Portilla, Juan Miguel Velásquez, Mónica Natalia Cruz, Andrés Peña, and Daniel Ricardo Oviedo. An innovative transit system and its impact on low income users: the case of the metrocable in medellín. *Journal of Transport Geography*, 39:49–61, 2014.
- [21] Wikimedia Commons. Taxi. https://commons.wikimedia.org/wiki/File:NYC_Taxi-Ford_Crown_Victoria.jpg, 2021.
- [22] Mario Durán. Ecobici. https://commons.wikimedia.org/wiki/File:Ecobici_03_2014-Mex_8163.JPG, 2014.
- [23] Gobierno de la Ciudad de México. Ecobici. <https://www.ecobici.cdmx.gob.mx/>.

-
- [24] Instituto de Investigaciones Parlamentarias. Diagnóstico de movilidad en la Ciudad en México: El impacto del crecimiento vehicular (problemas, estadísticas y evaluación de políticas). <http://aldf.gob.mx/archivo-9f6f5328e0f0853d4453d481cbffa2b6.pdf>.
- [25] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, New York, 2008.
- [26] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.
- [27] Wikimedia Commons. Tráfico aéreo. <https://informadorrd.com/los-fascinantes-mapas-que-muestran-los-miles-de-vuelos-que-cruzan-los-cielos-cada-dia-y-los-planes-de-la-nasa-para-mejorar-el-trafico-aereo/>.
- [28] Wikimedia Commons. Internet. https://commons.wikimedia.org/wiki/File:Internet_map_1024.jpg, 2006.
- [29] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.
- [30] Alejandro J. Alvarez, Carlos E. Sanz-Rodríguez, and Juan Luis Cabrera. Weighting dissimilarities to detect communities in networks. *Phil. Trans. R. Soc. A*, 373(2056), 2015.
- [31] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.
- [32] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
- [33] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.
- [34] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [35] Dmitry I. Belov and Ronald D. Armstrong. Distributions of the kullback–leibler divergence with applications. *British Journal of Mathematical and Statistical Psychology*, 64(2):291–309, 2011.
- [36] INEGI. Delimitación de Zonas Metropolitanas de México, 2015. http://internet.contenidos.inegi.org.mx/contenidos/Productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825006792.pdf, 2018.

-
- [37] María Fernanda Navarro. Población flotante: DF, centro laboral y escolar del Edo-mex. <https://www.excelsior.com.mx/comunidad/2015/12/24/1065140>, 2015.
- [38] INEGI. Encuesta Origen -Destino en Hogares de la Zona Metropolitana del Valle de México (EOD) 2017. Technical report, 2017.
- [39] H. S. Levinson; S. Zimmerman; J. Clinge; J. Gast. BUS RAPID TRANSIT Synthesis of Case Studies. Technical report, 2003.
- [40] Sami Al-Dubikhi and Paul Mees. Bus rapid transit in ottawa, 1978 to 2008: Assessing the results. *The Town Planning Review*, 81(4):407–424, 2010.
- [41] Luis Antonio Lindau, Dario Hidalgo, and Daniela Facchini. Curitiba, the cradle of bus rapid transit. *Built Environment*, 36(3):274–282, 2010.
- [42] Dario Hidalgo and Aileen Carrigan. BRT in latin America - high capacity and performance, rapid implementation and low cost. *Built Environment*, 36(3):283–297, 2010.
- [43] Wikipedia. Sistemas de autobús de tránsito rápido en México. https://es.wikipedia.org/wiki/Anexo:Sistemas_de_autobús_de_tránsito_rápido_en_México.
- [44] Robert Cervero and Danielle Dai. BRT TOD: Leveraging transit oriented development with bus rapid transit investments. *Transport Policy*, 36:127–138, 2014.
- [45] International Council on Clean Transportation. Cost and benefits of clean technologies for bus rapid transit (BRT) Summary of results for Nairobi. Technical report, 2012.
- [46] Dante Pérez-Méndez, Carlos Gershenson, María Elena Lárraga, and José L. Mateos. Modeling adaptive reversible lanes: A cellular automata approach. *PloS one*, 16(1):e0244326, 2021.
- [47] S. C. Wirasinghe, L. Kattan, M. M. Rahman, J. Hubbell, R. Thilakaratne, and S. Anowar. Bus rapid transit - a review. *International Journal of Urban Sciences*, 17(1):1–31, 2013.
- [48] Sistema de Transporte Colectivo. Conformación de un tren del metro. <https://www.metro.cdmx.gob.mx/operacion/conformacion-de-un-tren>.

- [49] The European Rail Research Advisory Council. Metro, light rail and tram systems in Europe. Technical report, 2009.
- [50] Brendan Finn, O Heddebaut, and S Rabuel. Bus with a high level of service (bhls): the european brt concept. In *Transportation Research Board 89th Annual Meeting, Washington, DC*, 2010.
- [51] Sistema de Transporte Colectivo. Análisis Costo y Beneficio a Nivel Perfil del Proyecto de Prestación de Servicios de Largo Plazo para poner a Disposición del Sistema de Transporte Colectivo un Lote de 30 Trenes de Rodadura Férrea. Technical report, 2009.
- [52] Editorial. 3 candidatos a la CDMX prometen construir más Metro... ¿Pero qué tan viable es? <https://politica.expansion.mx/cdmx/2018/04/19/3-candidatos-a-la-cdmx-prometen-construir-mas-metro-pero-que-tan-viable-es>, 2018.
- [53] Metrobús. Acerca de. <https://www.metrobus.cdmx.gob.mx/dependencia/acerca-de>.
- [54] INEGI. Transporte de pasajeros. <https://www.inegi.org.mx/temas/transporteurb/>.
- [55] Wikimedia Commons. Viaducto. https://commons.wikimedia.org/wiki/File:Estacion_Viaducto_02.jpg, 2006.
- [56] Wikimedia Commons. Metro. <https://upload.wikimedia.org/wikipedia/commons/9/-90/Metrofrayservando.JPG>.
- [57] Metrobús. Preguntas Frecuentes. <https://www.metrobus.cdmx.gob.mx/portal-ciudadano/preguntas-frecuentes>.
- [58] Metrobús. Reducción de emisiones. <https://www.metrobus.cdmx.gob.mx/dependencia/acerca-de/reduccionemisiones>.
- [59] Arturo Páramo. Sheinbaum reitera que habrá teleférico en Magdalena Contreras. <https://www.excelsior.com.mx/comunidad/sheinbaum-reitera-que-habra-teleferico-en-magdalena-contreras/1367103>, 2021.
- [60] Metrobús. Mapa del sistema. <https://www.metrobus.cdmx.gob.mx/mapas-de-sistema/mapa-completo-del-sistema>.
- [61] Metrobús. Tarifa, formas de pago y red externa. <https://www.metrobus.cdmx.gob.mx/portal-ciudadano/tarifas-y-formas-de-pago>.

-
- [62] Metrobús. Ampliación Línea 5. <https://www.metrobus.cdmx.gob.mx/linea-5>.
- [63] SEMOVI. MOVILIDAD + COVID 19 EN LA CIUDAD DE MÉXICO. <https://semovi.cdmx.gob.mx/storage/app/media/movilidad-covid-19-en-la-cdmx.pdf>, 2020.
- [64] Alberto Aleta, Sandro Meloni, and Yamir Moreno. A Multilayer perspective for the analysis of urban transportation systems. *Scientific Reports*, 7(March):1–9, 2017.
- [65] Goldstein Brett, Dyson Lauren, et al. *Beyond transparency: open data and the future of civic innovation*, volume 256. 2013.
- [66] Aaron Antrim, Sean J Barbeau, et al. The many uses of gtfs data—opening the door to transit and multimodal applications. *Location-Aware Information Systems Laboratory at the University of South Florida*, 4, 2013.
- [67] Google. Cómo hacer que los datos de transporte público sean accesibles para todos. <https://developers.google.com/transit?hl=es>.
- [68] Ben Smith. Introducing JSON. In *Beginning JSON*, chapter 4. Apress, 2015.
- [69] J. Kiusalaas. *Numerical Methods in Engineering with Python 3*. Cambridge University Press, 2013.
- [70] W. McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O’Reilly Media, 2012.
- [71] QGIS. Proyecto QGIS. <https://qgis.org/es/site/>.
- [72] G. Baumann. *Mathematica for Theoretical Physics: Classical Mechanics and Nonlinear Dynamics*. Mathematica for Theoretical Physics. Springer New York, 2006.