



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

MOVILIDAD HUMANA EN SISTEMAS DE
TRANSPORTE AÉREO

TESIS

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

PRESENTA:

JOSÉ ALBERTO RUIZ GAYOSSO

TUTOR:

DR. ALEJANDRO PÉREZ RIASCOS



CIUDAD UNIVERSITARIA, CD. MX., 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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

Primeramente quiero agradecer a mis padres pues me han mostrado su apoyo durante toda mi vida, les agradezco por su cariño y paciencia. También quiero agradecer a todos los profesores que he tenido durante mi vida, por todas sus enseñanzas y porque me guiaron a donde estoy ahora. Agradezco al Dr. Alejandro Pérez Riascos, mi tutor, que siempre estuvo atento durante la realización de este trabajo y porque siempre respondió con amabilidad a cualquier duda que tenía, gracias por su tiempo. Agradezco al Dr. Marcelo Del Castillo Mussot quien hizo posible que parte de los resultados de este trabajo se utilizaran en la realización de un artículo, proceso durante el cual aprendí muchas cosas. Agradezco a los sinodales por su tiempo y por sus valiosos comentarios que permitieron mejorar este trabajo. Finalmente, quiero agradecer el apoyo financiero otorgado por la Dirección General de Asuntos del Personal Académico de la UNAM a través del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) con No. IN105320 que permitió la realización de este trabajo.

Resumen

En esta investigación se analiza la red del sistema de transporte aéreo de los Estados Unidos utilizando los métodos de la teoría de redes y modelos de movilidad humana. En la primera parte se construye la matriz de origen destino asociada a este sistema de transporte utilizando la base de datos DB1B, obtenida de la página web de la Oficina de Transporte de los Estados Unidos, que contiene información detallada acerca de los vuelos nacionales entre aeropuertos dentro de los territorios estadounidenses. A partir de la matriz de origen destino se analizan las propiedades estadísticas de los grados de la matriz y se ajustan distribuciones de cola pesada por el método de máxima verosimilitud. Los resultados indican que este tipo de distribuciones describen correctamente el comportamiento estadístico de los grados. También se analizan las propiedades topológicas de la red y se comparan con las propiedades observadas en la red de transporte aéreo mundial la cual ha sido estudiada extensamente en el contexto de las redes complejas.

En la segunda parte se proponen dos modelos de movilidad para explicar los patrones de movimiento observados de los pasajeros. Estos modelos, basados en el modelo de tipo gravedad, se definen en términos de probabilidades de transición y distancias de vuelo. Los parámetros de cada modelo, de fácil interpretación, se obtienen minimizando una función error definida en el texto. Finalmente, la utilidad de los modelos para explicar los patrones de movimiento se confirma por medio de simulaciones Monte Carlo encontrando que estos modelos reproducen adecuadamente la dinámica de los pasajeros en la red de aeropuertos.

Índice general

1. Introducción	10
1.1. Motivación	10
1.2. Planteamiento del problema	10
1.3. Objetivos	11
2. Movilidad humana en diferentes medios de transporte	12
2.1. Introducción	12
2.2. Matriz de origen-destino	13
2.3. Redes	14
2.4. Movilidad humana a nivel de poblaciones	18
2.4.1. Modelos de tipo gravedad	19
2.4.2. Modelos de oportunidades	21
2.4.3. Modelo de radiación	22
2.5. Redes de transporte aéreo	23
2.5.1. Estudio del transporte aéreo mediante redes	24
2.5.2. Movilidad en el transporte aéreo	25
3. Matrices de Origen-Destino en transporte aéreo	29
3.1. Introducción	29
3.2. Descripción de los datos	30
3.3. Matrices de Origen-Destino	36

4. Modelamiento del transporte en una red de aeropuertos	53
4.1. Introducción	53
4.2. Matrices de probabilidad de transición	54
4.3. Modelos de movilidad	55
4.3.1. Modelamiento de la base de datos Market	56
4.3.2. Modelamiento de la base de datos Coupon	58
4.4. Estimación de los parámetros del modelo	59
4.4.1. Parámetros del modelo para Market	61
4.4.2. Parámetros del modelo para Coupon	61
4.5. Simulaciones Monte Carlo	62

Índice de figuras

2.1. Fuentes de datos para el estudio de la movilidad humana	13
2.2. Representación simplificada de una red social	15
2.3. Ejemplo de una gráfica simple.	16
2.4. Representación gráfica de la matriz de Origen-Destino	17
2.5. Esquema del modelo de oportunidades y del modelo de radiación.	22
2.6. Mapa mundial de tráfico aéreo.	24
2.7. Aeropuertos con mayor flujo de pasajeros en Estados Unidos.	26
3.1. Red de Aeropuertos en Estados Unidos	32
3.2. Estadísticas de los desplazamientos en Market y Coupon	36
3.3. Matrices de origen-destino y de distancias del año 2018	38
3.4. Gráficas y comunidades de la red de aeropuertos del año 2018	40
3.5. Distribución geográfica de las comunidades para el año 2018	41
3.6. Longitudes de camino mínimo de la red de aeropuertos del año 2018.	42
3.7. Longitud de camino mínimo promedio de la red de aeropuertos del año 2018.	43
3.8. Densidad de probabilidad de los grados de la matriz OD	45
3.9. Densidad de probabilidad de los grados de la matriz OD y distribuciones ajustadas	48
3.10. Densidad de probabilidad de la centralidad de los aeropuertos en Estados Unidos.	49
3.11. Rango de los grados de entrada y salida de la matriz OD.	50

3.12. Aeropuertos con mayor número de pasajeros en vuelos de salida y aeropuertos con mayor centralidad del año 2018.	52
4.1. Histograma bidimensional del conjunto de probabilidades de transición y distancias en Market	55
4.2. Histograma bidimensional del conjunto de probabilidades de transición y distancias en Coupon	56
4.5. Densidad de probabilidad de los desplazamientos en Market	63
4.6. Densidad de probabilidad de los desplazamientos en Coupon	64

Índice de tablas

3.1. Ejemplo extraído del conjunto de datos DB1B.	31
3.2. Variables seleccionadas de las tablas Market y Coupon para el análisis . . .	32
3.3. Veinte aeropuertos con mayor número de pasajeros en vuelos de salida . . .	34
3.4. Número de pasajeros y desplazamientos en Market y Coupon	35
3.5. Número de pasajeros totales que viajan entre aeropuertos activos	37
3.6. Número de aeropuertos por comunidad para el año 2018	39
3.7. Estadística de Kolmogorov-Smirnov de la distribución de los grados de la matriz OD	46
4.1. Características de la red de aeropuertos en los Estados Unidos	62

1. Introducción

1.1. Motivación

Los sistemas de transporte representan una parte importante de las sociedades debido al impacto en el crecimiento económico de las naciones. Estos sistemas, en los que existe una compleja interacción entre la infraestructura de transporte y los usuarios, pueden modelarse matemáticamente como redes complejas cuyo estudio permite identificar sus propiedades dinámicas y topológicas. Es posible también estudiar las propiedades estadísticas de estos sistemas mediante la introducción de las matrices de origen destino y a partir de estas, proponer modelos de movilidad para describir los patrones de movimiento de los usuarios. Aunque las redes de transporte aéreo han sido estudiadas desde el punto de vista de las redes complejas, en la mayoría de estos trabajos, se considera la matriz de adyacencia que ofrece información únicamente acerca de la topología de la red con menor énfasis en la dinámica del transporte en estas redes. En este trabajo, se estudia el transporte aéreo como un proceso dinámico en el caso particular de la red de aeropuertos de Estados Unidos para la cual se tiene una extensa base de datos bien documentada, esto permite explorar modelos que reproducen de manera adecuada el comportamiento de la red.

1.2. Planteamiento del problema

Diversos problemas relacionados con el sistema de transporte han surgido a través de los años debido a una inadecuada planificación de la infraestructura necesaria o al rápido incremento en su demanda. En este sentido, es necesario estudiar a los sistemas de transporte desde un punto de vista científico cuyo objetivo es el de entender los principios básicos que rigen el comportamiento de estos sistemas. De esta forma, se busca construir un modelo matemático, es decir, una representación simplificada de la realidad para describir las características principales y el comportamiento del sistema o los sistemas de transporte bajo estudio y al mismo tiempo obtener nuevas conclusiones que conduzcan a un mayor entendimiento del problema. Esto conducirá, en un futuro, a sistemas de trans-

porte más eficientes que minimicen los problemas de la actualidad como la congestión, la contaminación, los accidentes y el difícil acceso al transporte en ciertas regiones.

Esta caracterización de los sistemas de transporte por medio de modelos matemáticos luce cada vez más plausible. Los sistemas de transporte, por su constitución, pueden representarse mediante redes complejas las cuales exhiben propiedades que emergen de la interacción entre su gran número de componentes [1]. Por su naturaleza, el problema de estudiar los flujos humanos entre regiones depende en gran medida de la capacidad de cómputo de los ordenadores de la época. Actualmente, el incremento en el poder de cómputo de los ordenadores y la capacidad de almacenamiento han permitido estudiar infraestructuras de transporte desde un punto de vista nunca antes visto. La capacidad para procesar una cantidad enorme de datos ha permitido utilizar herramientas y modelos cada vez más sofisticados para la descripción más detallada de los sistemas de transporte.

Las redes complejas como una área de estudio interdisciplinaria incorpora a la física que hace uso de métodos propios para estudiar las propiedades de los sistemas complejos. En el contexto de los modelos de movilidad, la física también hace aportaciones importantes ya que varios de los modelos propuestos para estudiar los flujos de personas en diferentes escalas de espacio y tiempo, se derivan de modelos que originalmente se han utilizado para explicar fenómenos físicos; por ejemplo, la difusión anómala, los vuelos de Lévy, los procesos Markovianos, entre otros procesos estocásticos.

1.3. Objetivos

El objetivo general de este trabajo es caracterizar los patrones de movimiento de los pasajeros en la red de aeropuertos de Estados Unidos utilizando la matriz de origen destino y modelos de movilidad que dependen de cantidades que caracterizan a la red. De esta manera, se plantea realizar un procesamiento de los datos para la construcción de la matriz de origen destino del sistema de transporte aéreo y partir de estas matrices se propone buscar un modelo para caracterizar la movilidad en este sistema.

2. Movilidad humana en diferentes medios de transporte

2.1. Introducción

La aparición de dispositivos digitales móviles y la disponibilidad de una mayor cantidad de información proveniente de censos y encuestas, registros de teléfonos celulares, datos GPS, entre otros, ha permitido estudiar los patrones de movilidad humana con un mayor nivel de detalle y también la introducción de métodos y modelos matemáticos cada vez más sofisticados [2] (veáse Figura 2.1). Por ejemplo, diversos autores han estudiado la movilidad en el sistema de taxis de la ciudad de Nueva York. Cada uno de los vehículos en este sistema, está equipado con un dispositivo que permite conocer de forma precisa los puntos donde el pasajero comienza y termina su viaje, esto permite estudiar las características de los viajes en taxi en una de las ciudades más grandes e importantes del mundo [3, 4].

El término *movilidad humana* se refiere al movimiento de los seres humanos en diferentes escalas de espacio y tiempo [2]. La importancia de estudiar los patrones de movilidad humana radica en el impacto y en sus posibles aplicaciones en diferentes áreas como la migración, los sistemas de transporte, la planificación urbana y el modelamiento de epidemias [2]. La movilidad humana combina un conjunto de disciplinas entre las cuales destaca la geografía por su papel fundamental en los primeros estudios sobre migración humana. En el siglo pasado, diversos autores realizaron los primeros estudios cuantitativos en diferentes escalas espaciales; por ejemplo, sobre migración entre países, a una gran escala, y sobre el movimiento de personas que viajan desde sus hogares hacia sus respectivos lugares de trabajo, en una escala local. Algunos autores, entre los cuales destacan el sociólogo Samuel Stoufer [5] y el lingüista George Kingsley Zipf [6] introdujeron diversos modelos matemáticos para caracterizar el movimiento de los viajeros. Estos estudios continuaron y extendieron parte del trabajo realizado por el geógrafo Ernst Ravenstein en su obra *The laws of migration* [7] y significaron una importante contribución en el desarrollo del estudio de la movilidad humana [2].

En el estudio de la movilidad humana se distinguen dos enfoques distintos para modelar los



Figura 2.1: Fuentes de datos para el estudio de la movilidad humana. En los últimos años, la aparición de dispositivos móviles y tecnologías GPS ha permitido observar los movimientos humanos con un nivel de detalle nunca antes logrado. De la misma forma, los avances en informática en términos de procesamiento y memoria posibilitan el estudio de grandes cantidades de información en busca de los patrones que caracterizan a la movilidad humana.

patrones de movimiento [2]: uno a nivel individual, en el cual el objetivo es caracterizar los patrones de movilidad de cada individuo; y otro a nivel de poblaciones, en el cual el objetivo es caracterizar la movilidad de un agregado de individuos. En este capítulo se describen algunos de los modelos y métodos que se han propuesto para estudiar los patrones que emergen en movilidad humana a nivel de poblaciones.

2.2. Matriz de origen-destino

En el estudio de la movilidad a nivel individual, los modelos utilizados para caracterizar los patrones de movimiento se basan en conceptos y métodos de los caminantes aleatorios y el movimiento browniano [2]. En el caso de la movilidad a nivel de poblaciones, la herramienta principal para estudiar los patrones de movimiento humano es la matriz de origen-destino (OD), que contiene información sobre el número de personas que se mueven entre zonas de una región espacial en un intervalo específico de tiempo [2, 8].

La matriz de origen-destino \mathbf{T} es una matriz de tamaño $n \times m$ donde n corresponde al número de zonas “origen”, lugares donde se origina el viaje; y m corresponde al número de zonas “destino”, lugares donde termina el viaje, generalmente se considera $n = m$. Cada una de las entradas T_{ij} de la matriz OD es igual al número de personas que viajan

desde la zona i hacia la zona j [2, 9]. Las zonas consideradas están definidas por el sistema bajo estudio: en algunos casos, el conjunto de zonas se obtiene tomando una partición de la región espacial de interés y el tamaño de cada una de estas zonas depende del nivel de detalle de la información contenida en la base de datos [2]. En otros casos, el conjunto de zonas está predefinido por la naturaleza del sistema; por ejemplo, en varios sistemas de transporte, las diferentes zonas corresponden a la estaciones de la red de transporte: estaciones de autobús, estaciones de metro, aeropuertos, etc. En general, $T_{ij} \neq T_{ji}$, es decir, la matriz OD no es simétrica. Esto refleja el hecho de que el número de personas que viajan desde la zona i hacia la zona j no es igual al número de personas que viajan desde la zona j hacia la zona i . Por ejemplo, el número de personas que migran desde México hacia Estados Unidos no es igual al número de personas que migran desde Estados Unidos hacia México.

A partir de la definición de la matriz OD se pueden definir los grados de la matriz por medio de las siguientes ecuaciones. El grado de entrada $k_i^{(e)}$ de la matriz OD, donde (e) se refiere a entrada, está dado por [4]:

$$k_i^{(e)} = \sum_{j=1}^N T_{ji}, \quad (2.1)$$

donde N corresponde al número total de zonas. Esta cantidad representa el número total de viajes cuyo destino es la zona i . Análogamente, el grado de salida $k_i^{(s)}$ de la matriz OD, donde (s) se refiere a salida, está dado por [4]:

$$k_i^{(s)} = \sum_{j=1}^N T_{ij} \quad (2.2)$$

y denota el número total de viajes cuyo origen es la zona i .

Por otra parte, a partir de la definición de la matriz OD y la ecuación (2.2) se puede obtener la fracción $w_{i \rightarrow j}$ de viajes que inician en la zona i y terminan en la zona j por medio de la relación:

$$w_{i \rightarrow j} = \frac{T_{ij}}{k_i^{(s)}}. \quad (2.3)$$

La ecuación (2.3) define una nueva matriz; la matriz de probabilidad de transición \mathbf{W} (con elementos $w_{i \rightarrow j}$, $i, j = 1, 2, \dots, N$), que permite utilizar a la matriz OD para definir las probabilidades de transición entre dos zonas dentro de la región espacial de interés [8]. El valor $w_{i \rightarrow j}$ es la probabilidad de que un viaje que comienza en la zona i termine en la zona j . De esta forma, la matriz OD es útil para caracterizar los movimientos en redes espaciales.

2.3. Redes

Una red se define como cualquier sistema que puede representarse mediante una gráfica (también llamada grafo) cuyos nodos denotan los elementos de dicha red y las aristas

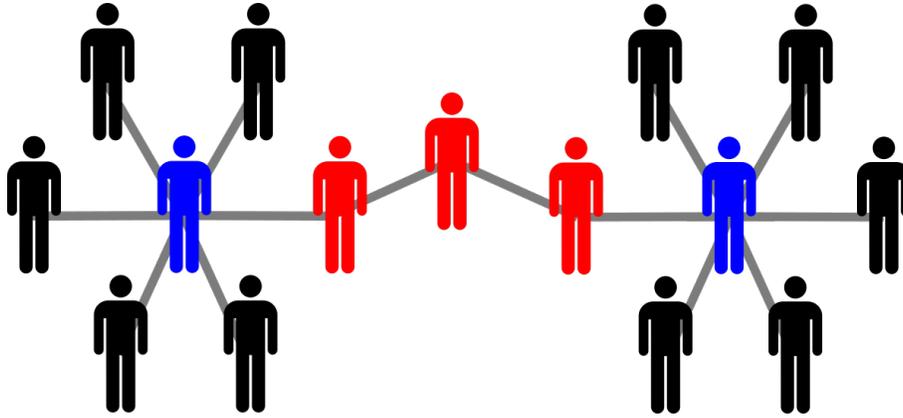


Figura 2.2: Representación simplificada de una red social. Las redes sociales son un caso especial de las redes espaciales, las cuales consisten de nodos o vértices localizados en el espacio. En el caso de las redes sociales, una métrica posible para describir la distancia entre sus elementos está dada por el número de personas que se necesitan para conectar a cualquier par de individuos. En la figura, la distancia entre los individuos en color azul es igual a cuatro.

corresponden a la presencia de interacciones o algún tipo de relación entre estos elementos [10]. En particular, una red espacial consiste de elementos localizados en un espacio, para el cual, se define una métrica que permite medir las distancias entre sus elementos; estas distancias, no necesariamente están asociadas magnitudes físicas; por ejemplo, en una red social, la distancia puede definirse como el número de personas necesarias para conectar a cualquier par de individuos [11]. Esto último se muestra en la Figura 2.2.

Las redes son estudiadas de forma matemática utilizando la teoría de gráficas, las gráficas son objetos matemáticos que se utilizan para describir a las redes y constan de un conjunto de vértices y un conjunto de aristas [10, 11]. En la Figura 2.3 se muestra una representación de una gráfica simple en la cual cada círculo rojo corresponde a un elemento de la red y la existencia de una línea que une a cualesquiera dos elementos implica una interacción o relación entre estos. Para estudiar una red se define la matriz de adyacencia A , en la cual, cada entrada A_{ij} asume dos valores posibles, 1 o 0. La entrada A_{ij} de la matriz de adyacencia es 1 si existe una conexión o interacción entre los nodos i , j y 0 en caso contrario [10]. Esta interacción puede ocurrir en una sola dirección o en ambas direcciones; en este último caso, la gráfica es dirigida. La matriz de adyacencia presenta un inconveniente: no cuantifica la magnitud de la interacción entre nodos. Para considerar la magnitud de las interacciones entre los elementos de la red se define la matriz de pesos Ω ; en la cual, cada entrada Ω_{ij} especifica la interacción entre los nodos i , j [10]. Nótese que, en general, $\Omega_{ij} \neq \Omega_{ji}$.

Muchas de las cantidades que se definen a partir de la matriz de adyacencia A y que permiten caracterizar estadísticamente a la red pueden generalizarse de forma directa para la matriz de pesos Ω [10]. Por ejemplo, una de las principales medidas de caracterización es la importancia de los nodos o aristas. Esta importancia se denomina centralidad; de esta manera, un nodo o arista es importante si este es “central” y permite conectar o interactuar a partes distantes de la red. Existen diversas formas de medir la centralidad

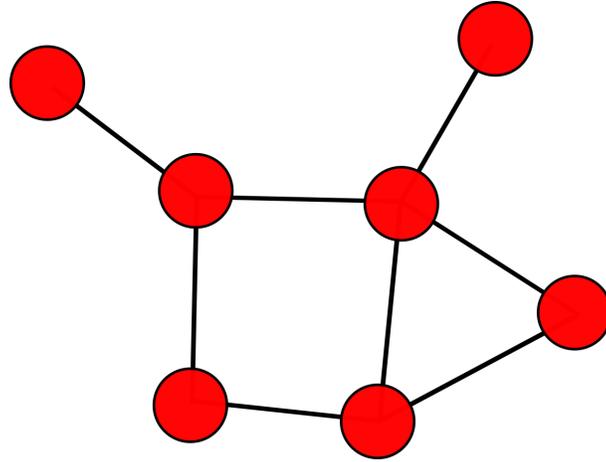


Figura 2.3: Ejemplo de una gráfica (o grafo) simple. En teoría de gráficas, las gráficas o grafos son los objetos matemáticos bajo estudio y permiten modelar relaciones o interacciones entre los elementos de un sistema. Dichos objetos están constituidos por un conjunto de vértices, o nodos, y un conjunto de aristas que representan a los elementos y sus interacciones, respectivamente. La teoría de gráficas es importante en el estudio de las redes complejas pues estas se describen matemáticamente mediante gráficas.

según las propiedades que se consideren interesantes [11, 10]. La medida más utilizada es el grado de la matriz de adyacencia o de la matriz de pesos. Si T representa cualquiera de estas dos matrices, el grado de entrada se define como $k_i^{(e)}$ en la ecuación (2.1) mientras que el grado de salida $k_i^{(s)}$ se define utilizando la ecuación (2.2). En efecto, la matriz OD corresponde a la matriz de pesos de una red de movilidad, tal como se ilustra en la Figura 2.4.

Los grados de la red cuantifican el número de interacciones o conexiones entre un elemento particular (nodo o arista) y el resto de elementos de la red. Otras medidas de centralidad están basadas en la métrica o distancia que se define en la red y permiten definir nodos o aristas que unen partes separadas de la red, funcionando como puentes entre grupos de elementos poco conectados, o elementos que conectan a cualquier par de elementos en el menor número de pasos [10]. Por ejemplo, la centralidad de intermediación (betweenness centrality) b_i del elemento i se define como el número de caminos de mínima longitud que atraviesan i ; donde un camino de longitud mínima l_{jk} es aquel que conecta a dos nodos j, k en el menor número de pasos. De esta forma, la centralidad de intermediación b_i queda definida mediante la ecuación

$$b_i = \sum_{i \neq j \neq k} \frac{\sigma^{(i)}_{jk}}{\sigma_{jk}} \quad (2.4)$$

donde σ_{jk} es el número de caminos de longitud mínima que conectan a los nodos j, k y $\sigma^{(i)}_{jk}$ el número de dichos caminos que pasan a través de i .

Si se calcula el promedio del camino de longitud mínima sobre el conjunto de todos los pares posibles (j, k) se obtiene una medida que indica el número de pasos promedio que

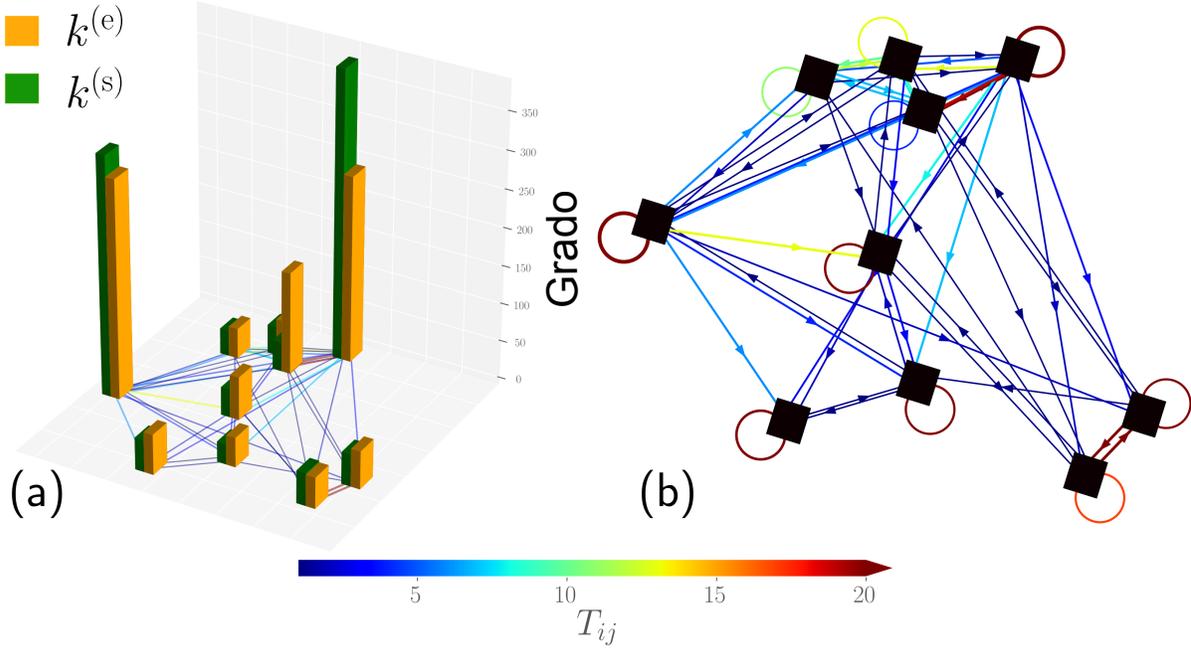


Figura 2.4: Representación gráfica de la matriz de Origen-Destino. (a) Representación tridimensional de la información en la matriz OD donde mediante barras verticales se muestran los grados de entrada y salida de cada una de las zonas. (b) Red dirigida asociada con la matriz OD, las flechas denotan la dirección de las interacciones y los colores describen su magnitud. Figura tomada de [3] con permiso del autor.

se necesitan para conectar cualquier par de nodos. Este promedio se denomina longitud de camino mínimo promedio y se define por medio de la relación

$$\langle l \rangle = \frac{1}{N(N-1)} \sum_{ij} l_{ij}, \quad (2.5)$$

donde N es igual al número de nodos de la red. Esta cantidad es importante para establecer si una red satisface la propiedad de mundo pequeño la cual se traduce en que cualquier par de nodos dentro de la red puede conectarse mediante un número relativamente pequeño de pasos [1].

Otra parte importante en el estudio de redes son las comunidades; formadas por grupos de nodos cuya principal característica es que presentan un mayor número de conexiones e interacciones entre ellos en comparación con el resto de los nodos. En redes que representan sistemas reales la detección de comunidades es importante; por ejemplo, la detección de comunidades en redes sociales para entender su estructura y organización. Existen diversas técnicas para la identificación de comunidades; sin embargo, cada una tiene sus ventajas y desventajas y no existe un algoritmo “universal” que pueda aplicarse en todas las situaciones [11]. Varios algoritmos para detectar comunidades se han desarrollado para ser aplicados en redes complejas que surgen en diferentes áreas como la biología, ciencias de la computación, economía, ingeniería, política, etc [12].

2.4. Movilidad humana a nivel de poblaciones

El estudio de la movilidad humana a nivel de poblaciones utiliza a la matriz OD como herramienta principal. Esta matriz se construye empíricamente mediante un agregado de información acerca de los patrones de movimiento de los individuos en una población [2]; por ejemplo, el flujo de personas entre dos lugares de una ciudad puede estimarse utilizando datos sobre el número de viajes en taxi que ocurren entre dichos lugares [3]. La matriz OD es particularmente útil cuando las diferentes zonas y sus interacciones se representan mediante una red espacial dirigida, en la cual, los nodos o vértices corresponden a los lugares y las aristas corresponden a las interacciones entre estos lugares, estas interacciones pueden describir flujos de personas o bienes. En la Figura 2.4 se presenta una ilustración de la matriz de Origen-Destino y los grados de entrada y de salida, los cuadrados son sitios de interés y las líneas en la red espacial dirigida conectando estos lugares representan el número de viajes.

Para predecir los flujos entre las diferentes zonas de una red espacial, se han propuesto algunos modelos que buscan describir sus interacciones en términos de variables locales; es decir, cantidades asociadas a las propiedades de dichas zonas. Por ejemplo, el tamaño de su población o ciertos indicadores económicos [2]. De esta forma, los diferentes modelos propuestos difieren en las variables que se consideran importantes para describir los flujos y en la forma específica en que se relacionan dichas variables [2]. Estos modelos, se pueden clasificar en dos clases generales: una primera clase que corresponde a los modelos de tipo gravedad, en analogía a la ley de gravitación universal y cuya principal característica, es la suposición, de que el número de viajes entre dos zonas es una función decreciente de la distancia entre estas zonas; y, una segunda clase que corresponde a los modelos de oportunidades y cuya principal característica es asumir que el flujo entre dos zonas está determinado por el número de oportunidades (las oportunidades son los posibles destinos entre la zona de origen y la zona de destino que el viajero rechaza en favor de continuar su camino) que se presentan entre estas dos zonas [2]. En cada una de estas clases existen diferentes modelos que se derivan de un primer modelo más simple.

El objetivo de estos modelos es predecir el número de viajes entre la zona i y la zona j para todos los valores posibles de i, j . Estos modelos buscan representar la distribución p_{ij} que corresponde a la probabilidad de que ocurra un viaje entre las zonas i, j [8]. Cada modelo asume que dicha distribución de probabilidad depende de ciertas variables que caracterizan a la zonas de la red y sus interacciones.

Utilizando la distribución p_{ij} se genera, a partir de una simulación, una nueva matriz de origen destino \hat{T} cuyos elementos representan los viajes entre las zonas consideradas en el estudio. Esta nueva matriz permite analizar si el modelo propuesto describe los patrones de movilidad que se observan en la realidad. Para generar esta matriz se pueden considerar diferentes tipos de restricciones de acuerdo a los objetivos del estudio y a la cantidad de información disponible [2, 8]:

- **Modelo sin restricción.** La única restricción consiste en mantener el número total

de viajes simulados igual al número total de viajes observados N .

- **Modelo de única restricción en producción.** Se añade una restricción adicional al modelo; el número total de viajes *simulados* que se originan en la zona i (para toda i) es igual al número de viajes *observados* que se originan en la zona i .
- **Modelo de única restricción en atracción.** Se añade una restricción adicional al modelo; el número total de viajes *simulados* que son atraídos a la zona i (para toda i) es igual al número de viajes *observados* hacia la zona i desde cualquier otra zona.
- **Modelo de doble restricción.** En este modelo, el número de viajes *simulados* que son originados y atraídos a la zona i (para toda i), debe ser igual al número de viajes *observados* que son originados y atraídos a la zona i .

Estas dos clases de modelos de movilidad (de tipo gravedad y de oportunidades) han sido consideradas en diferentes contextos y aplicaciones, siendo el modelo de tipo gravedad preferido por su simplicidad matemática [8]. En diversos estudios se realizaron comparaciones entre ambas clases de modelos en diferentes contextos y conjuntos de datos [8, 13, 14]. Esta comparación entre ambas clases sugiere que dependiendo el contexto en que sean aplicadas, una u otra clase arroja mejores resultados; sin embargo, en estos estudios, la comparación no se lleva a cabo de forma equitativa pues para cada modelo se utilizan variables distintas para simular los flujos y las restricciones no son siempre las mismas [8]. En [8], Lenormand et al. realizaron una comparación entre modelos con las mismas variables y restricciones, y los resultados mostraron que el modelo de tipo gravedad supera al modelo de oportunidades en la estimación de flujos y en la preservación estructural de la red, además de reproducir correctamente la distribución de desplazamientos de los viajeros $P(d)$. A continuación se presenta una descripción detallada sobre estas dos clases de modelos y se enuncian algunas de sus aplicaciones en el estudio de la movilidad humana.

2.4.1. Modelos de tipo gravedad

Como se señaló anteriormente, la clase de modelos de tipo gravedad tiene como principal característica la suposición de que la distancia es el factor de disuasión más importante en la decisión de un viajero sobre viajar de un lugar a otro. Este modelo fue uno de los primeros modelos cuantitativos que se utilizaron para describir los flujos migratorios entre dos lugares y en su forma más simple, debida a Zipf (1946), puede expresarse a través de la relación [6, 4]:

$$T_{ij} \propto \frac{P_i P_j}{d_{ij}} \quad (2.6)$$

donde T_{ij} es el número de viajes desde la zona i hacia la zona j y P_i, P_j son las poblaciones de las zonas i, j respectivamente; d_{ij} es la distancia entre estas dos zonas [2].

En la ecuación (2.6) se observa la analogía entre este modelo y la ley de gravitación universal, donde P_i y P_j corresponden a las masas m_i, m_j de dos cuerpos físicos. Extendiendo

este tipo de razonamiento para considerar diferentes tipos de “masas” e interacciones entre las zonas consideradas, a través de la distancia como factor de disuasión, se obtiene una formulación más general de este tipo de modelo definido a través de la ecuación [2]:

$$T_{ij} \propto m_i m_j f(d_{ij}) \quad (2.7)$$

donde m_i y m_j son funciones que codifican el número de viajes que salen desde la zona i y que llegan a la zona j respectivamente. Estas funciones pueden incluir a la población o alguna otra variable que permita caracterizar estas zonas; por ejemplo, $m_i = P^\alpha$ y $m_j = P^\beta$. La función $f(d_{ij})$ es una función decreciente de la distancia y su forma funcional depende de diversos factores como el propósito de viaje y el modo de transporte [11]. En otras aplicaciones, el factor de disuasión puede expresarse en términos de costos monetarios o tiempo de viaje.

Este modelo de tipo gravedad ha sido utilizado en diferentes contextos; por ejemplo, los patrones de movilidad en sistemas de transporte como el de bicicletas [15] en las ciudades de Nueva York y Chicago, y el de taxis [3] en la ciudad de Nueva York, han sido estudiados utilizando variaciones del modelo tipo gravedad. En estos estudios, la distancia sigue siendo uno de los principales factores de disuasión; sin embargo, únicamente a largas distancias, en escalas locales, existe una vecindad de radio R alrededor del punto de origen en la cual todos los posibles destinos tienen la misma probabilidad de ser alcanzados. Lo anterior se expresa matemáticamente de la siguiente forma:

$$p_{ij} \propto \begin{cases} 1 & \text{si } 0 \leq d_{ij} \leq R \\ \left(\frac{R}{d_{ij}}\right)^\alpha & \text{si } R < d_{ij} \end{cases}$$

donde p_{ij} es la probabilidad de que exista un viaje desde la zona i hacia la zona j . De esta forma, los individuos consideran la distancia como un factor importante a la hora de viajar, únicamente si esta distancia excede un límite R definido por el sistema. Por ejemplo, en [15], Loaiza y Riascos, encontraron para la ciudad de Chicago que $R = 1.194$ km y $\alpha = 2.11$ mientras que para la ciudad de Nueva York encontraron que $R = 1.094$ km y $\alpha = 2.12$. En [3], Riascos y Mateos, aplicaron un modelo similar para estudiar el sistema de taxis en la ciudad de Nueva York para el cual encontraron que $R = 2.0$ km y $\alpha = 2$.

A una escala mayor, a nivel de países, Jung et al. [16] estudiaron el sistema de carreteras de Corea del Sur que conecta a 30 ciudades con una población mayor a los 200,000 habitantes. En este caso, el modelo utilizado para caracterizar el flujo de individuos corresponde a la formulación original del modelo tipo gravedad:

$$T_{ij} \propto \frac{P_i P_j}{d_{ij}^2}$$

donde P_i , P_j son las poblaciones de las ciudades i , j respectivamente. Los resultados muestran una alta correlación entre el flujo de personas y la cantidad $\frac{P_i P_j}{d_{ij}^2}$, lo cual demuestra que el sistema de carreteras de Corea del Sur puede caracterizarse mediante un modelo de tipo gravedad.

En un estudio sobre la propagación de enfermedades infecciosas que involucra diferentes escalas, locales y globales, en 29 países alrededor del mundo, Balcan et al. [17], estudiaron la movilidad de individuos que se desplazan en transporte no aéreo y la movilidad en la red mundial de aeropuertos, utilizando un modelo con un mayor número de parámetros:

$$T_{ij} \propto p_i^\alpha p_j^\gamma e^{-\frac{d_{ij}}{r}}$$

donde r representa una distancia característica que refleja el decaimiento de los flujos de individuos a distancias cada vez más grandes. Balcan et al, encontraron que el modelo de tipo gravedad describe correctamente los patrones de movilidad a nivel local y aunque el número de viajes en escala local es de un orden mayor, la propagación de enfermedades está principalmente determinada por los flujos en la red mundial de aeropuertos.

2.4.2. Modelos de oportunidades

En contraste con el modelo de tipo gravedad, en el modelo de oportunidades, la distancia juega un papel indirecto como factor de disuasión. En el modelo de oportunidades, el principal factor de disuasión es el número de oportunidades entre dos zonas; es decir, el número de potenciales destinos que el viajero rechaza en favor de seguir hacia su destino original [2]. Este modelo fue propuesto por Stouffer en 1940 en su trabajo titulado, *Intervening opportunities: a theory relating mobility and distance* [5]. En este trabajo se enuncia la ley de oportunidades intermedias como: “El número de personas que se desplazan una cierta distancia es directamente proporcional al número de oportunidades a esa distancia e inversamente proporcional al número de oportunidades intermedias”. Matemáticamente esto se traduce en la siguiente relación:

$$\frac{dy}{dr} \propto \frac{1}{x} \frac{dx}{dr}$$

donde y es el número acumulado de personas que viajan una distancia r y x es el número acumulado de oportunidades entre el origen y el destino. Una formulación debida a Schneider [18] en términos de probabilidades, está dada por:

$$T_{ij} = K_i O_i (e^{-LV_{ij-1}} - e^{-LV_{ij}}). \quad (2.8)$$

Aquí, O_i es el número total de viajes que se originan desde la zona i y K_i representa un término de normalización. El término en paréntesis corresponde a la probabilidad de que un viaje que comienza en i finalice en j donde L representa una probabilidad constante de aceptar una oportunidad y V_{ij} es el número de oportunidades acumuladas desde la zona i hasta la zona j [2, 18].

Este modelo ha sido aplicado en diferentes contextos; sin embargo, su aplicación se ha limitado a escalas espaciales locales; como flujos de individuos en ciudades o flujos entre ciudades. Por ejemplo, en [19], Elffers et al. utilizaron el modelo de oportunidades para estudiar la criminalidad en la ciudad de La Haya, Países Bajos, bajo la hipótesis de que el número de oportunidades para cometer un crimen determina el número de viajes, que

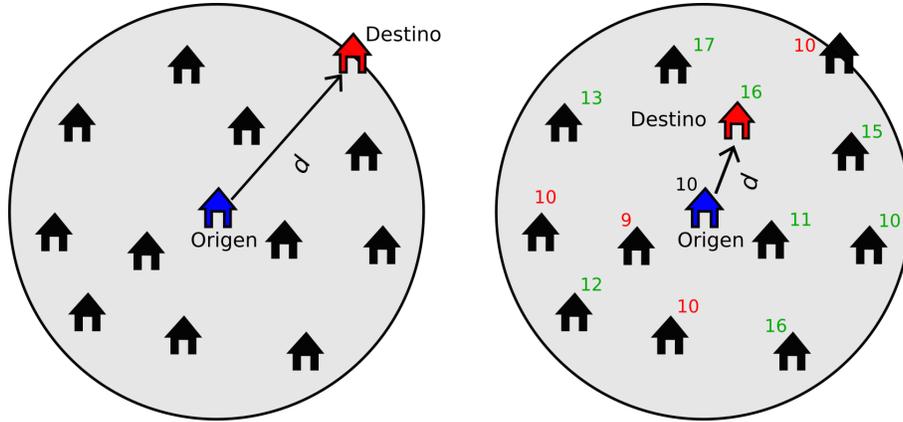


Figura 2.5: Esquema del modelo de oportunidades y del modelo de radiación. En (a), se ilustra el modelo de oportunidades, cada oportunidad (representada por un icono en forma de casa) dentro del círculo de radio d es un posible destino que el viajero rechaza en favor de continuar hacia su destino (casa en rojo). En (b), se muestra un esquema del modelo de radiación, aquí, cada oportunidad tiene asociado un valor que cuantifica su importancia. El viajero escoge la oportunidad más cercana y con la mayor importancia. En ambos modelos, y a diferencia del modelo de tipo gravedad, la distancia juega un papel indirecto.

realizan los criminales entre dos zonas. En este caso, los resultados indican que la distancia entre dos zonas explica mejor el número de viajes con propósitos criminales; por lo tanto, un modelo de tipo gravedad es más adecuado para estudiar los patrones de movimientos de los criminales.

En [20], Akwawua y Pooler, utilizaron el modelo de oportunidades para estudiar los flujos migratorios (de forma similar al artículo original de Stouffer) entre los 48 estados contiguos de Estados Unidos y el Distrito de Columbia. Los resultados mostraron que este modelo es capaz de reproducir razonablemente la distribución de la migración en Estados Unidos y los resultados son similares a los obtenidos con un modelo de tipo gravedad.

2.4.3. Modelo de radiación

El modelo de radiación es una reformulación del modelo de oportunidades y recibe su nombre de la analogía que se establece entre los procesos físicos de radiación y absorción, y el movimiento de un individuo que parte de una zona i y el cual es “absorbido” por otra zona j [8]. En este modelo, el concepto de oportunidades sigue siendo el factor principal de disuasión en la decisión de un viajero sobre viajar de un lugar a otro; sin embargo, ahora cada oportunidad (destino potencial) tiene asignado un número z obtenido de la distribución de probabilidad $p(z)$ que describe que tan buena es dicha oportunidad. La oportunidad en el lugar de origen también tiene asignada un valor z' . El destino escogido por el viajero es aquel que es más cercano al lugar de origen y que satisface $z > z'$. Puesto que los valores que representan la calidad de las oportunidades es generado de forma aleatoria, el número promedio de personas que se mueven entre las zonas i, j , está dado

por [2, 8]:

$$T_{ij} = O_i \frac{1}{1 - \frac{m_i}{M}} \frac{m_i m_j}{(m_i + s_{ij})(m_j + s_{ij})}, \quad (2.9)$$

donde O_i es el número de viajes que se originan desde la zona i , m_j es el número de oportunidades en la zona j y s_{ij} el número de oportunidades dentro del círculo de radio r_{ij} , siendo r igual a la distancia entre las zonas i, j , centrado en la zona i . Una ventaja de este modelo sobre otros modelos de movilidad es que no requiere de calibración pues no depende de parámetros a estimar; por lo que puede aplicarse de forma general en diversos contextos, además, resuelve ciertos problemas relacionados con el modelo de tipo gravedad como la falta de pautas para escoger la función $f(d_{ij})$ y la incapacidad para predecir la movilidad en regiones en las cuales se carece de datos sobre el movimiento de los individuos [13].

El modelo de radiación fue introducido por primera vez en [13]. En este artículo, Simini et al. utilizaron el modelo de radiación, un modelo estocástico, para estudiar flujos de migración ocurridos en periodos de tiempo largos y otros patrones de movilidad relacionados con el transporte. En particular, el modelo fue aplicado para predecir los movimientos de 10,000 individuos en el condado de New York. Los resultados mostraron una mejor aproximación a los patrones de movilidad observados en la realidad en comparación con los resultados del modelo de tipo gravedad.

En [21], Stefanouli y Polyzos, utilizaron el modelo de radiación y dos variantes del modelo de tipo gravedad para estudiar la movilidad de individuos que se desplazan a sus lugares de trabajo en Grecia. Los resultados obtenidos mostraron que el modelo de radiación describe correctamente los patrones de movilidad y aunque los modelos de gravedad mostraron mejores valores de las estadísticas utilizadas para evaluar la calidad de los ajustes, el gran número de parámetros requeridos por estos modelos significan su principal desventaja.

En la Figura 2.5 se muestra un esquema de los modelos de oportunidades y de radiación en la cual se hace énfasis en las diferencias de cada modelo.

2.5. Redes de transporte aéreo

En el año 2018, el número de pasajeros transportados en la red mundial de aeropuertos fue de 4.122 mil millones de acuerdo a datos del banco mundial [22]. Cada año, se incrementa el número de personas que utilizan el transporte aéreo con diferentes propósitos; por lo tanto, resulta de vital importancia estudiar el sistema de aeropuertos donde se llevan a cabo estos movimientos. En particular, el estudio de los patrones de movilidad en las redes de transporte aéreo es fundamental para comprender su impacto en la economía mundial y en las economías nacionales, el flujo de bienes y personas, así como la propagación de enfermedades como la influenza y el SARS [23, 24]. En la Figura 2.6 se muestra el mapa de tráfico aéreo del año 2009; el gran número de conexiones entre aeropuertos de diferentes partes del mundo permite viajar de un punto a otro en un número pequeño de vuelos.



Figura 2.6: Mapa mundial de tráfico aéreo [25]. La figura muestra 54,317 rutas aéreas alrededor del mundo para el año 2009. El sistema de transporte aéreo representa una de las infraestructuras de transporte más importantes en el mundo debido a su capacidad de transportar bienes y personas en periodos de tiempo cortos.

Varios autores han investigado distintas redes de aeropuertos alrededor del mundo utilizando la metodología de las redes complejas. Diversos aspectos relacionados con el sistema de transporte aéreo como la movilidad de pasajeros, la evolución temporal de la red de aeropuertos, y su adaptación a cambios en la demanda de los pasajeros o factores económicos externos han sido estudiados considerando las propiedades topológicas de la red [24]. Estos estudios son relevantes para el presente y futuro del transporte aéreo; esto, debido a los límites en la capacidad de los sistemas de transporte aéreo (por ejemplo, la red mundial de aeropuertos enfrenta cada año una mayor demanda) y a la integración en un futuro entre el sistema de transporte aéreo y otros sistemas de transporte [24] (veáse Figura 2.7).

2.5.1. Estudio del transporte aéreo mediante redes

Existen diversas representaciones posibles, en términos de redes, de los sistemas complejos que se observan en el mundo real. El sistema de transporte aéreo no es la excepción y es posible estudiarlo utilizando distintos enfoques que dan lugar a diferentes redes para analizar a este sistema de transporte. La representación más utilizada en el estudio del transporte aéreo considera a los aeropuertos como los nodos de la red y la existencia de un vuelo directo entre un par de aeropuertos está asociada a una interacción entre los nodos correspondientes; esta representación recibe el nombre de red de vuelos y permite estudiar las propiedades topológicas de la red. Debido a que no todos los vuelos son iguales o a que ciertas rutas tienen un mayor flujo de aeronaves y pasajeros, posibles variaciones de este

enfoque se obtienen al considerar a las interacciones en términos del número de vuelos totales entre cada par de aeropuertos, el número de pasajeros transportados, o el número de asientos disponibles en una aeronave [24]. Estas últimas representaciones ofrecen información acerca de la dinámica que ocurre en la red pues cuantifican la magnitud de la interacción entre cualesquiera dos nodos.

Otras redes complejas asociadas con el sistema de transporte aéreo están relacionadas con la estructura del espacio aéreo; por ejemplo, las redes que se obtienen a partir de las rutas que siguen las aeronaves, compuestas por marcadores de navegación (navaids), permiten estudiar las trayectorias con el objetivo de evitar desastres en el espacio aéreo [26]. También se pueden considerar las redes que consideran la dinámica de los aeropuertos en términos de los tiempos de retraso de cada vuelo, lo cual tiene repercusiones económicas importantes [27]. Estas redes complejas son importantes para entender el impacto económico de este tipo de factores; sin embargo, su objetivo principal no es el estudio de la movilidad humana.

2.5.2. Movilidad en el transporte aéreo

La red de transporte aéreo ha sido estudiada extensamente durante mucho tiempo, en particular, el crecimiento de la teoría de redes complejas en este último par de décadas ha permitido estudiar a este sistema de transporte con una variedad de métodos más sofisticados, y por lo tanto, entender su estructura topológica y dinámica con un mayor nivel de detalle [24].

En uno de los primeros estudios sobre el sistema de transporte aéreo en el contexto de las redes complejas, Guimerà et al. [28] investigaron las propiedades de la red de transporte aéreo mundial considerando los vuelos directos entre cada par de ciudades alrededor del mundo. En este trabajo, Guimerà et al. encontraron que esta red de transporte satisface la propiedad de mundo pequeño, es decir, que en promedio, el número de vuelos necesarios d para conectar cualesquiera dos ciudades es pequeño; globalmente $d = 4.4$ vuelos y regionalmente $d = 3.5$ vuelos. En general, encontraron que si S es el número de ciudades consideradas, entonces $d \approx \log S$. Guimerà et al. también encontraron que aquellas ciudades con un gran número de conexiones (vuelos directos) con otras ciudades no son necesariamente las más centrales; una ciudad es central si permite conectar diferentes regiones de la red.

Otro resultado importante de esta red mundial es que el grado k de una ciudad, dado por el número de conexiones con otras ciudades, tiene una distribución acumulada que decae como una ley de potencias truncada:

$$P(> k) \propto k^{-\alpha} f\left(\frac{k}{k_x}\right) \quad (2.10)$$

donde $\alpha = 1.0$ y k_x representa una constante que depende del tamaño de la red (número de ciudades). Otras redes de transporte aéreo también han sido investigadas y se ha

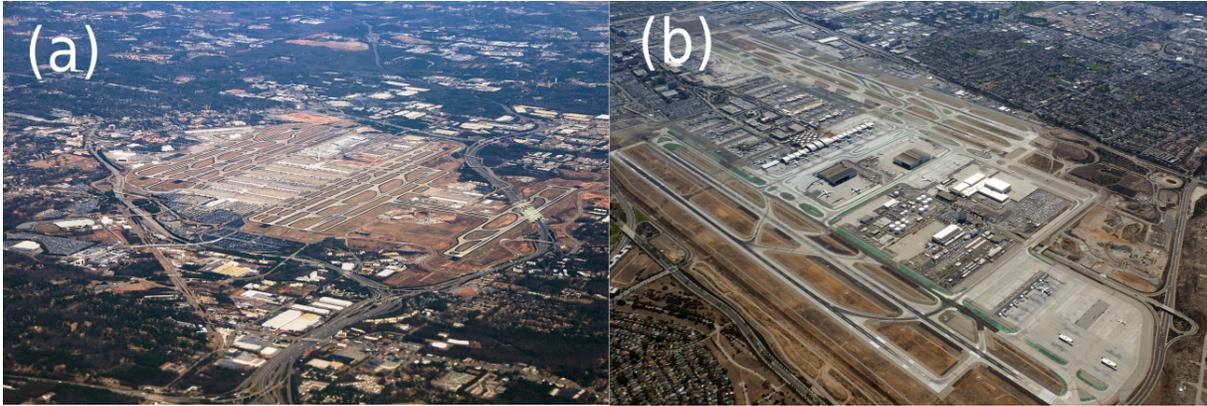


Figura 2.7: Aeropuertos con mayor flujo de pasajeros en Estados Unidos. En (a), se muestra el aeropuerto internacional Hartsfield Jackson en Atlanta [33], el aeropuerto de Estados Unidos con mayor flujo de pasajeros. En 2018, el flujo anual en este aeropuerto fue de 107,394,029 pasajeros. En (b), se muestra el aeropuerto internacional de Los Angeles [34], el segundo aeropuerto con mayor flujo de pasajeros en Estados Unidos. En 2018, su flujo anual fue de 87,534,384 pasajeros.

encontrado que siguen leyes de potencia truncadas similares; tal es el caso de la red de transporte aéreo en China con $\alpha \sim 1.7$, Estados Unidos con $\alpha \sim 2.6$ e India con $\alpha \sim 2.2$ [29, 30, 31]. En el caso de la red mundial de transporte aéreo, también se encontró que la centralidad intermedia b (betweenness centrality) tiene una distribución que obedece una ley de potencias truncada, similar a la distribución del grado de la matriz de adyacencia k .

En [32], Guimerà y Amaral, propusieron un modelo para explicar la generación y la evolución de la red mundial de transporte aéreo. A partir de este modelo que considera conexiones preferenciales, la distancia entre ciudades, y ciertas restricciones geopolíticas como factores importantes en la generación de la red; los autores lograron explicar las propiedades observadas en la realidad en esta red de transporte.

Aunque estos resultados permiten entender la estructura topológica de las redes de transporte aéreo a nivel nacional o a nivel global; debido a que estos son obtenidos a partir de la matriz de adyacencia \mathbf{A} , y en la mayoría de los casos, de la suposición de que esta matriz es simétrica, los resultados no reflejan la dinámica que ocurre dentro de la red en términos del número de aeronaves o pasajeros que se mueven en ella, es decir, no se considera la magnitud de la interacción entre las ciudades (dos ciudades pueden tener el mismo número de conexiones con otras ciudades pero con diferentes flujos de tráfico de aeronaves o pasajeros).

Sin embargo, utilizando redes con pesos (véase Sección 2.4), estos resultados pueden generalizarse para los casos en que las interacciones entre nodos están representadas por el número de vuelos o número de pasajeros transportados. En estos casos, los patrones de movilidad que se generan dentro de la red se vuelven importantes y los modelos presentados en la Sección 2.4 pueden ser utilizados para explicar dichos patrones. Aunque este tipo de generalizaciones no es común en la literatura sobre la red de transporte aéreo, algunos

autores han explorado las redes de transporte aéreo de países y territorios como China, India, Estados Unidos y el continente europeo utilizando la magnitud o la intensidad de la interacción entre los nodos [24].

Uno de los primeros trabajos en considerar la magnitud de las interacciones entre los nodos fue realizado por Barrat y colaboradores en [23]. En este trabajo, además de estudiar la red científica de colaboraciones, los investigadores analizaron a la red mundial de transporte aéreo (también estudiada en [32] y [28]) cuantificando la interacción entre aeropuertos mediante el número de asientos disponibles en cada vuelo directo. De esta forma, obtuvieron resultados importantes acerca de las propiedades estadísticas de la magnitud o intensidad de las interacciones. Específicamente, Barrat et al. definieron la importancia de un aeropuerto en términos de la cantidad de tráfico que dicho aeropuerto es capaz de manejar. Matemáticamente, si Ω_{ij} representa el número de asientos disponibles en un vuelo directo entre los aeropuertos i, j , entonces la importancia del aeropuerto i está dada por:

$$S_i = \sum_j^N \Omega_{ij}$$

Barrat et al. encontraron que esta cantidad S , que corresponde al grado de la matriz OD, tiene propiedades similares al grado k de la matriz de adyacencia. En el caso de S , sus valores siguen una distribución de cola pesada por lo que se observan valores en extremo grandes (asociados a aeropuertos con un flujo de tráfico elevado) en comparación con el valor promedio $\langle S \rangle$; una propiedad típica en redes que representan a sistemas reales complejos [23]. También encontraron la existencia de una relación no lineal entre S y el grado k que se expresa como $S(k) \sim k^\beta$ con $\beta = 1.5$. De este último resultado, los investigadores concluyeron que existe una alta correlación entre la intensidad de las interacciones entre aeropuertos y las propiedades topológicas de la red mundial de transporte aéreo; comportamiento que no se observa en redes aleatorias para las cuales $S(k) \propto k$, es decir, el grado de la matriz OD S no proporciona información adicional a la ya provista por el grado de la matriz de adyacencia k .

De forma similar, otras redes de transporte aéreo a nivel nacional, han sido estudiadas. Por ejemplo, Li y Cai, en [29]; y Bagler, en [30], utilizaron un enfoque similar al utilizado en [23] para estudiar las propiedades topológicas y la dinámica de los vuelos en la red de aeropuertos de China e India respectivamente. En ambos casos, los investigadores cuantificaron la magnitud de la interacción entre aeropuertos considerando el número de vuelos directos entre cualesquiera dos aeropuertos. En el caso de la red de transporte aéreo China, se construyó una red de 128 aeropuertos con un total de 1165 vuelos directos entre pares de aeropuertos; en el caso de la red de transporte aéreo India, se construyó una red de 79 aeropuertos con un total de 492 vuelos directos entre pares de aeropuertos. Ambos estudios encontraron que las redes presentan propiedades de mundo pequeño y que los grados de la matriz de adyacencia siguen distribuciones que decaen como una ley de potencias. Además, para la red de aeropuertos de India y de forma similar a la red mundial de transporte aéreo, se encontró una relación no lineal entre la intensidad de tráfico aéreo S y el grado k de la matriz de adyacencia: $S(k) \sim k^\beta$ con $\beta = 1.43$.

Sin embargo, de acuerdo a los resultados encontrados en [35], los resultados obtenidos para redes que describen un subconjunto de un sistema real podrían no extrapolarse para el caso global. En [29] y [30], ambas redes representan subconjuntos de la red de transporte real; por lo tanto, los resultados obtenidos podrían diferir una vez que se considere la red de transporte completa.

En los tres ejemplos presentados el flujo del tráfico entre aeropuertos es cuantificado en términos del número de asientos disponibles o el número de vuelos entre aeropuertos. Debido a que no todos los asientos son ocupados por pasajeros o no, todos los vuelos son iguales (los vuelos pueden tener un número diferente de pasajeros), estos estudios no toman en cuenta el número exacto de pasajeros en cada vuelo. En [31], Xu y Harriss exploraron la red de transporte aéreo de Estados Unidos y analizaron sus propiedades topológicas y la dinámica que ocurre en la red en términos del número de pasajeros diarios que se mueven entre cualquier par de aeropuertos. En este caso, se estudiaron cuatro años de datos, desde el 2002 hasta el 2005; periodo en el cual se consideraron hasta 272 ciudades y 6636 vuelos directos entre ciudades. Xu y Harriss encontraron que el grado k de la matriz de adyacencia y el grado S de la matriz OD siguen distribuciones que decaen como una ley de potencias truncada, es decir, $P(k) \sim k^{-\gamma}$ con $\gamma = 2.63$, y $P(S) \sim S^{-\gamma}$ con $\gamma = 2.13$. Además, satisfacen la relación no lineal $S(k) \sim k^{\beta}$ con $\beta = 1.7$ como ocurre también en la red mundial de aeropuertos y la red de transporte aéreo en India.

En estos trabajos se han estudiado las propiedades de las redes asociadas a los sistemas de transporte aéreo, a nivel nacional o global. Sin embargo, no se proponen modelos para explicar la movilidad de los pasajeros en estos sistemas. La literatura sobre movilidad humana en sistemas de transporte aéreo está principalmente enfocada en el estudio de estos sistemas en el contexto de las redes complejas. En este trabajo se pretende estudiar el sistema de transporte aéreo de Estados Unidos desde el punto de vista de las redes complejas; pero también tomando en consideración los modelos de movilidad presentados en la Sección 2.4. De esta forma, es posible caracterizar no sólo la red donde se lleva a cabo la movilidad sino también caracterizar los patrones de movilidad de los pasajeros que utilizan la infraestructura del sistema de transporte aéreo.

3. Matrices de Origen-Destino en transporte aéreo

3.1. Introducción

En este capítulo se exploran las matrices de origen-destino que se construyen a partir de información de los vuelos en la red aeroportuaria de los Estados Unidos. El enfoque de estudio utilizado en este capítulo es el de las redes complejas que permite realizar un análisis estadístico de cantidades que caracterizan la estructura de la red. En la primera parte se describe de forma detallada el conjunto de datos “The Airline Origin and Destination Survey (DB1B)” obtenido de la página web oficial de la Oficina de transporte de los Estados Unidos [36]. Dicho conjunto de datos provee información de los vuelos nacionales que ocurren entre aeropuertos de los diferentes estados y territorios de Estados Unidos. Las variables consideradas relevantes para este trabajo y que son útiles para el estudio de la movilidad humana son: el número de pasajeros en cada vuelo, la distancia recorrida y el número de identificación único de cada aeropuerto que permite realizar un seguimiento a través de los años.

A partir de la información contenida en la base de datos, se construyen dos matrices de origen-destino, utilizando la definición dada en la Sección 2.2, para cada uno de los años entre 2010 y 2018. Estas dos matrices representan dos redes distintas: (1) la red que permite estudiar la movilidad de los pasajeros que se denomina Market y (2) la red de aeropuertos donde se lleva a cabo dicha movilidad; es decir, el conjunto de aeropuertos intermedios por los que tiene que pasar cada pasajero para llegar a su destino final. Esta segunda red se denomina Coupon. Estas dos matrices proveen información importante acerca del sistema de transporte aéreo. Análogamente, se construyen las correspondientes matrices de distancias D que contienen información acerca de la distancia recorrida en cada vuelo entre cualesquiera dos aeropuertos.

Una vez que se ha hecho una descripción detallada de los datos y de las matrices de origen-destino, se analiza la red de aeropuertos. En términos del número de pasajeros, dos aeropuertos están conectados si el número de pasajeros viajando entre estos es mayor que un umbral previamente definido. Los resultados obtenidos muestran que la red de

aeropuertos satisface la propiedad de mundo pequeño la cual establece que cualesquiera dos aeropuertos pueden conectarse mediante un número pequeño de viajes. Finalmente, se analizan los grados de salida y entrada para cada una de las matrices OD de los años considerados. Los resultados obtenidos en este capítulo muestran que la red de aeropuertos de Estados Unidos presenta propiedades similares a las observadas en otras redes de transporte aéreo alrededor del mundo.

3.2. Descripción de los datos

En esta sección se describe el conjunto de datos “The Airline Origin and Destination Survey (DB1B)” recopilado por la Oficina de Estadísticas de Transporte del Departamento de Transporte de los Estados Unidos y que se encuentran disponible de forma abierta en su página web oficial [36]. Este conjunto consiste de una muestra aleatoria de un 10% del total de tickets de vuelo de pasajeros que han viajado en vuelos nacionales durante un año, estos datos han sido recopilados desde el año 1993 hasta la fecha y se dividen de forma trimestral. Los datos proveen información detallada acerca del itinerario de cada vuelo, por ejemplo: aeropuerto de origen y aeropuerto de destino, número de pasajeros, distancia recorrida, indicador de vuelo redondo y tarifas de clase [37].

Para cada año, el conjunto de datos se divide en tres tablas: “DB1BTicket”, “DB1BMarket” y “DB1BCoupon” a las cuales se hará referencia en lo que sigue como Ticket, Market y Coupon, respectivamente. La unidad básica en Coupon es lo que se denomina *coupon* o cupón y se refiere a una porción del vuelo completo en la que un pasajero es transportado entre dos aeropuertos de la red que forman parte de la trayectoria total, es decir, cada registro en Coupon contiene detalles acerca de cada vuelo de conexión dentro del vuelo completo. Un *ticket* es la unidad básica de Ticket, corresponde a un único itinerario de vuelo y puede contener uno o varios cupones. De acuerdo a la descripción de la base de datos en la página oficial [37], Ticket contiene un total de 303, 276, 607 registros (desde 1993 hasta 2019) y provee información básica acerca de cada itinerario de vuelo como la aerolínea operadora, número de pasajeros, costo del ticket y la distancia total de vuelo. La tabla Market contiene un total de 535, 639, 256 registros y provee información acerca de las características de cada segmento del itinerario como número de pasajeros, tarifas prorrateadas e información sobre el tipo de vuelo (nacional o internacional). Por su parte, Coupon contiene un total de 880, 384, 622 registros con información específica acerca de cada vuelo dentro del itinerario completo como número de pasajeros, aeropuertos de origen y destino, tarifas de clase, distancia recorrida, etc. En la Tabla 3.1 se muestra un ejemplo extraído de la base de datos correspondiente a un itinerario único que permite identificar la relación entre cada una de las tablas, obsérvese que la clave primaria en (a) Ticket es el número de itinerario (ItinID). De esta forma, cada registro en (a) está asociado con uno o más registros en las tablas (b) Market y (c) Coupon. Por ejemplo, si un pasajero adquiere un boleto para un vuelo redondo el cual consiste en ir de Los Ángeles, California a la ciudad de Nueva York y de regreso, la tabla Ticket contendrá un registro especificando el aeropuerto de origen en California, la tabla Market contendrá un

Tabla 3.1: Ejemplo extraído del conjunto de datos DB1B. Cada itinerario de vuelo se identifica mediante una clave única ItinID. En (a) se muestra el registro correspondiente al itinerario de vuelo con clave única 201811724603 que contiene información detallada como número de cupones en el itinerario (Coupons), año (Year) y trimestre (Quarter) en el que ocurrió el viaje, y código del aeropuerto de origen (Origin). Este registro tiene asociados uno o más registros en Market y Coupon. Cada fila en (b) corresponde a un segmento del vuelo completo y se identifica mediante la clave única MktID, en este caso particular, el primer segmento del vuelo consiste de tres vuelos mientras que el segundo consta de dos vuelos de acuerdo a lo que se observa en la columna MktCoupons. Por otra parte, cada fila en (c) representa una escala o vuelo de conexión dentro del vuelo completo, aquí se tienen cinco registros que corresponden al total de vuelos dentro del itinerario, la columna SeqNum establece el orden de cada uno de estos vuelos. En (a), (b) y (c), los aeropuertos de origen y destino se especifican mediante un número de identificación único que aparece en las columnas OriginAirportID y DestAirportID.

(a) Ticket					
ItinID	Coupons	Year	Quarter	Origin	OriginAirportID
201811724603	5	2018	1	GSO	11995

(b) Market					
ItinID	MktID	MktCoupons	Year	Quarter	OriginAirportID
201811724603	20181172460301	3	2018	1	11995
201811724603	20181172460304	2	2018	1	11423

(c) Coupon					
ItinID	MktID	SeqNum	Coupons	Year	OriginAirportID
201811724603	20181172460304	5	5	2018	11433
201811724603	20181172460301	1	5	2018	11995
201811724603	20181172460301	3	5	2018	13367
201811724603	20181172460301	2	5	2018	10397
201811724603	20181172460304	4	5	2018	11423

registro para el viaje de ida y otro para el de vuelta. Mientras que Coupon contendrá registros para cada vuelo de conexión que se tenga que realizar durante ambos trayectos, es decir, para cada vuelo de conexión existe un registro en esta tabla.

En el caso particular de la Tabla 3.1, el registro en (a) corresponde a un itinerario que consta de cinco vuelos entre diferentes aeropuertos de acuerdo a lo que se puede observar en la columna “Coupons”; por otra parte, en (b) existen dos registros con clave MktID distinta correspondientes a los dos segmentos en los que se divide el vuelo completo, para el primer segmento se especifican tres vuelos y para el segundo únicamente dos vuelos de acuerdo a lo observado en la columna “MktCoupons”. En cuanto a Coupon, se tienen cinco registros, uno para cada vuelo dentro del itinerario. Aunque cada una de las tablas está relacionada en la forma mencionada, cada una contiene información única que no está disponible en las restantes. Este conjunto de datos ha sido explorado por diversos autores en diferentes contextos como dinámica de redes [38], memoria en flujos de redes [39], dinámica en redes multicapa [40], el estudio del impacto de las condiciones meteorológicas en el transporte aéreo [41], entre otros [42, 43, 44].

Tabla 3.2: Variables seleccionadas de las tablas (a) Market y (b) Coupon para el análisis. A partir de estas variables se obtienen las matrices de Origen-Destino que permiten estudiar la dinámica de la red de aeropuertos de Estados Unidos. En ambos casos, las variables OriginAirportID y DestAirportID asociadas a un aeropuerto son únicas y no cambian en el tiempo lo cual permite analizar el comportamiento de la red a lo largo de los diferentes años.

(a) Market	
Variable	Descripción
ItinID	Id del Itinerario. Número único para cada registro en Coupon.
MktID	Market Id. Número único para cada registro en Market.
OriginAirportID	Clave única con la que se identifica al aeropuerto de origen.
DestAirportID	Clave única con la que se identifica al aeropuerto de destino.
Passengers	Número de pasajeros.
MktMilesFlown	Total de millas recorridas (en vuelo).

(b) Coupon	
Variable	Descripción
ItinID	Id del Itinerario. Número único para cada registro en Coupon.
MktID	Market Id. Número único para cada registro en Market.
OriginAirportID	Clave única con la que se identifica al aeropuerto de origen.
DestAirportID	Clave única con la que se identifica al aeropuerto de destino.
Passengers	Número de pasajeros.
Distance	Total de millas recorridas (en vuelo).

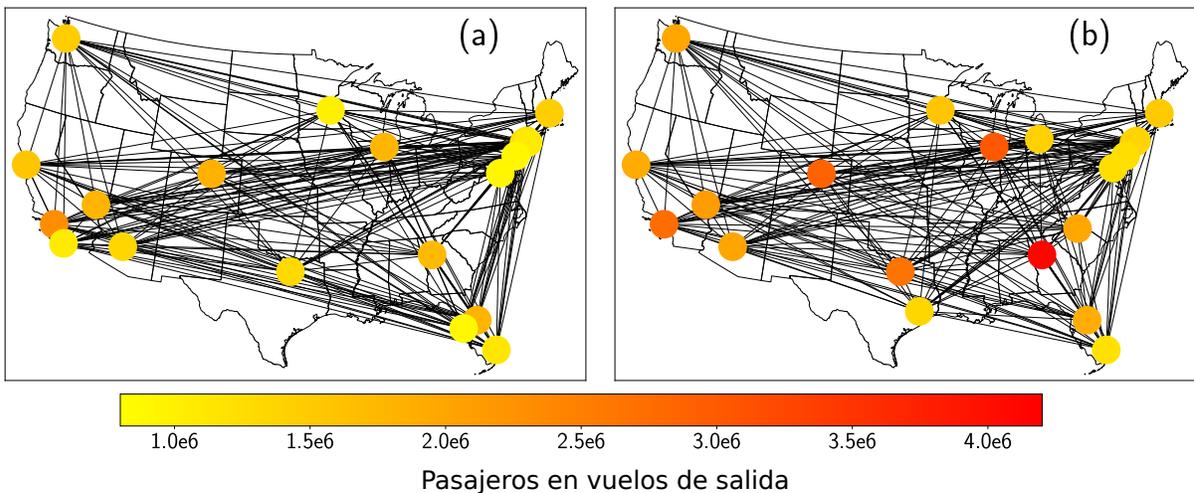


Figura 3.1: Red de Aeropuertos en Estados Unidos. En (a) se muestra la red de aeropuertos en el contexto de la movilidad humana utilizando el conjunto de datos Market, en (b) la red de aeropuertos donde se lleva a cabo dicha movilidad a partir de los datos en Coupon. Por claridad, en ambos casos se presentan únicamente los veinte aeropuertos con mayor número de pasajeros en vuelos de salida. El color de cada nodo indica el número de dichos pasajeros y la conexión entre aeropuertos representa el hecho de que existe por lo menos un vuelo entre estos. Observe que en general, los aeropuertos de las dos redes no son los mismos y aquellos que están presentes en ambas no tienen el mismo color.

Cabe mencionar que este conjunto de datos presenta algunas limitaciones; por ejemplo, la cantidad de información disponible representa únicamente un 10 % de los vuelos totales. Esto es particularmente importante en el caso de aerolíneas con pocos vuelos para las cuales se tienen un número menor de registros dentro del conjunto de datos. Tampoco es posible llevar a cabo un análisis detallado de los patrones temporales de movilidad ya que no se especifican fechas ni horas de salida o llegada. En el estudio realizado se consideran únicamente las tablas Market y Coupon ya que contienen información acerca de los aeropuertos de origen y destino de cada uno de los vuelos, estas permiten estudiar la movilidad humana (Market) y la red de aeropuertos en donde se lleva a cabo dicha movilidad (Coupon). Es decir, en el caso de Market, para cada registro, el aeropuerto de destino representa el lugar el cual tienen como objetivo los pasajeros del vuelo, por el contrario, en Coupon, el aeropuerto de destino representa una escala o vuelo de conexión y por lo tanto, un lugar por el cual tienen que transitar los pasajeros para llegar a su destino final. Para propósitos de este trabajo sólo se consideran los datos correspondientes a los años dentro del rango [2010, 2018], esto para mantener aproximadamente constante el número de vuelos totales por año y debido a que al momento de realización de este trabajo la información correspondiente al año 2019 no cuenta con los dos últimos trimestres.

Por otra parte, para cada una de las tablas solo se considera un subconjunto de variables que son relevantes para el análisis de la red de aeropuertos. Para estudiar la movilidad humana y la red de aeropuertos como una red espacial dirigida, es necesario conocer el flujo de pasajeros que viajan entre aeropuertos de la red así como la distancia que recorren dichos pasajeros entre cada par de aeropuertos. De esta forma, las variables relevantes que se consideran para el análisis son el número de pasajeros y distancia recorrida (sin considerar transporte por tierra) en cada vuelo, las claves únicas ItinID y MktID que permiten identificar los vuelos que pertenecen a un mismo itinerario y los números de identificación de los aeropuertos de origen (OriginAiportID) y destino (DestinationAiportID) que son utilizados para realizar un seguimiento de los aeropuertos a través de los años y que permanecen constantes. En la Tabla 3.2 se hace una descripción de cada una de las variables consideradas para el análisis.

Con el objetivo de representar gráficamente la red de aeropuertos en Estados Unidos, en la Figura 3.1 se muestra una representación de la red de aeropuertos donde se lleva a cabo la movilidad, esta nos permite observar las conexiones entre los veinte aeropuertos con mayor número de pasajeros en vuelos de salida. En esta figura, cada aeropuerto está representado por un nodo de la red y su color denota el número de pasajeros en vuelos de salida desde dicho aeropuerto. Por claridad, los aeropuertos restantes de la red no se muestran en la figura, sin embargo de la figura es posible observar que aquellos con un mayor número de pasajeros cubren gran parte del territorio de los Estados Unidos. La conexión entre dos nodos de la red representa el hecho de que existe por lo menos un vuelo entre estos dos aeropuertos. De la figura, también se puede observar que Market y Coupon difieren en la localización de los aeropuertos con mayor número de pasajeros. En el caso de Market, el aeropuerto con mayor número de pasajeros es el de Los Ángeles, California mientras que para Coupon el aeropuerto con mayor número de pasajeros es el de Atlanta, Georgia. De acuerdo a lo discutido en la Tabla 3.1 y los párrafos anteriores, esto

Tabla 3.3: Veinte aeropuertos de la red correspondiente a Coupon con mayor número de pasajeros en vuelos de salida. Para cada aeropuerto se presenta: nombre, número de pasajeros en vuelos de llegada, número de pasajeros en vuelos de salida y coordenadas geográficas (latitud y longitud). Nótese que los aeropuertos aparecen ordenados de acuerdo al número de pasajeros en vuelos de salida y que este número es aproximadamente igual al número de pasajeros en vuelos de llegada.

Nombre	Llegadas	Salidas	Latitud	Longitud
Atlanta Municipal	4109321	4121668	33.6408	-84.4272
Chicago O'Hare International	2996279	3019123	41.9780	-87.9061
Stapleton International	2876284	2881765	39.7744	-104.8797
Los Angeles International	2740622	2746288	33.9425	-118.4080
Dallas Fort Worth Regional	2634030	2656274	32.8944	-97.0297
McCarran International	2102290	2115118	36.0800	-115.1522
Phoenix Sky Harbor International	1999147	2007573	33.4361	-112.0094
Seattle International	1987097	1986723	47.4472	-122.3055
Douglas Municipal	1956472	1981425	35.2191	-80.9358
San Francisco International	1923334	1916681	37.6188	-122.3750
Orlando International	1894868	1893104	28.4316	-81.3247
Minneapolis-St Paul International	1578480	1579137	44.8861	-93.2177
Logan International	1517121	1510324	42.3636	-71.0061
Newark Liberty International	1431651	1424282	40.6958	-74.1716
Detroit Metro Wayne County	1418209	1421662	42.2166	-83.3444
LaGuardia	1375879	1375739	40.7794	-73.8758
Houston Intercontinental	1338819	1340130	29.9833	-95.3400
Philadelphia International	1237744	1241460	39.8680	-75.2486
Friendship International	1241665	1240244	39.1755	-76.6713
Fort Lauderdale-Hollywood International	1205981	1209226	26.0741	-80.1516

significa que el aeropuerto en Atlanta sirve como un punto de conexión entre diferentes aeropuertos, mientras que el aeropuerto en Los Ángeles representa un origen y destino frecuente para los viajeros. En otras palabras, la función de cada aeropuerto en cada una de las redes de la Figura 3.1 es distinta, los aeropuertos en la red correspondiente a Coupon representan puntos intermedios por los cuales deben transitar los pasajeros mientras que los aeropuertos en la red asociada a Market representan los verdaderos orígenes y destinos de los pasajeros.

En la Tabla 3.3 se presentan las características de los veinte aeropuertos de la red correspondiente a Coupon con mayor número de pasajeros en vuelos de salida. También se presenta el número de pasajeros en vuelos de llegada, nombre y coordenadas geográficas de cada aeropuerto. Nótese que para cada uno de los aeropuertos, el número de pasajeros en vuelos de llegada es aproximadamente igual al número de pasajeros en vuelos de salida lo cual señala cierta simetría en el comportamiento de un aeropuerto como punto de partida y punto de llegada. Esto último será explorado con más detalle en la siguiente sección. Para complementar la descripción de los datos, en la Tabla 3.4 se muestran algunas características para cada uno de los años considerados en el análisis; por ejemplo, número total de pasajeros, distancia máxima d_{max} y distancia promedio \bar{d} . Aquí se observa que el

Tabla 3.4: Número de pasajeros y desplazamientos en (a) Market y (b) Coupon. Para cada año desde 2010 hasta 2018, se presenta el número total de pasajeros, los desplazamientos mínimo d_{min} y máximo d_{max} en kilómetros así como el desplazamiento promedio \bar{d} . Los desplazamientos corresponden a la distancia de vuelo reportada en la base de datos y no a la distancia geográfica entre los puntos de origen y destino.

(a) Market				
Año	Total de pasajeros	d_{min} (km)	d_{max} (km)	\bar{d} (km)
2010	43092916	27.3	17973.1	2083.4
2011	43671977	33.8	19617.8	2104.8
2012	43900575	28.9	22485.7	2210.0
2013	44455899	17.7	17664.1	2093.5
2014	45884943	17.7	17416.2	2060.8
2015	48495394	49.9	17630.3	2072.5
2016	51225753	17.7	18060.0	2065.1
2017	53569257	17.7	18789.0	2055.9
2018	56784850	17.7	20177.9	2063.2

(b) Coupon				
Año	Total de pasajeros	d_{min} (km)	d_{max} (km)	\bar{d} (km)
2010	58995052	16.0	12409.6	1393.0
2011	59732478	16.0	12768.5	1400.2
2012	59887861	17.7	11983.1	1405.0
2013	60194281	3.2	13468.5	1420.1
2014	62009665	3.2	11983.1	1426.7
2015	65001626	3.2	9944.1	1443.1
2016	67086350	3.2	12818.3	1466.9
2017	68986721	3.2	9131.3	1478.2
2018	72560126	3.2	8199.5	1489.2

número total de pasajeros se ha incrementado a través de los años; sin embargo, es posible notar que la distancia promedio \bar{d} de cada año sigue una tendencia constante. Este no es el caso de la distancia mínima d_{min} o distancia máxima d_{max} en las cuales se observan variaciones más grandes en sus valores. Por otra parte, en la Figura 3.2 se muestran las distribuciones estadísticas de los desplazamientos, estos corresponden a la distancia total recorrida durante el vuelo reportada en la base de datos y no a la distancia geográfica entre los puntos de origen y destino.

De la Figura 3.2 se puede observar que en ambos casos, la mayoría de los desplazamientos se concentran en intervalos diferentes. En el caso de Market, la distribución tiene una mayor simetría respecto al valor máximo en $d = 1500$ km mientras que en Coupon, se observa una tendencia hacia los valores más bajos de los desplazamientos. Esto es consistente con la descripción que se hizo de estas dos tablas pues refleja el hecho de que la mayoría de los vuelos completos cuyas distancias van desde los 100 hasta los 10000 km, están conformados por vuelos entre aeropuertos intermedios con distancias más cortas. Este hecho también puede observarse en la columna de los desplazamientos promedio \bar{d}

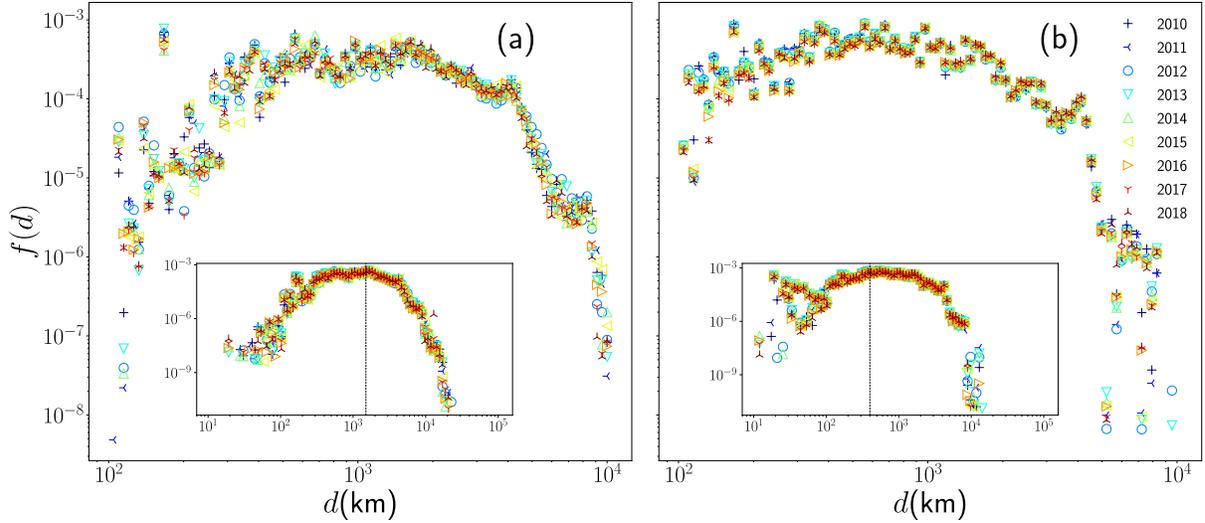


Figura 3.2: Estadísticas de los desplazamientos en Market y Coupon. En (a) se muestra la frecuencia de los desplazamientos (distancias de vuelo recorridas) en Market que van desde los 10^2 hasta los 10^4 km para cada año desde 2010 y hasta 2018. En (b) se muestra la misma gráfica pero para el conjunto de datos Coupon. En cada caso, se presenta una gráfica insertada en la cual se muestra un mayor intervalo de la distribución y una línea vertical que denota el máximo. Obsérvese la diferencia en el máximo en ambas gráficas, me Market, este se encuentra aproximadamente en $d = 1500$ km y en Coupon aproximadamente en $d = 400$ km.

en la Tabla 3.4.

3.3. Matrices de Origen-Destino

En esta sección se definen las matrices de origen-destino (OD) de la red de aeropuertos en Estados Unidos. De acuerdo a la definición de la matriz de Origen-Destino dada en la Sección 2.2, para Market y Coupon se construye la matriz de Origen-Destino \mathbf{T} como aquella cuyo entrada T_{ij} representa el número de pasajeros que parten desde el aeropuerto i y que llegan al aeropuerto j donde $i, j = 1, 2, \dots, N$. Aquí, N es el número de aeropuertos considerados y es igual al número de aeropuertos activos en la red. En este caso, se considera que un aeropuerto es activo si el número de pasajeros que salen desde y que llegan a dicho aeropuerto es mayor o igual que 1000. Este valor resulta razonable considerando el hecho que para cada año se tienen más de 40 millones de registros y al mismo tiempo se evita tomar en cuenta aeropuertos que no están bien representados en el conjunto de datos. Para estudiar el flujo de pasajeros a través de la red de aeropuertos, se considera que cada persona que viaja entre dos aeropuertos i, j contribuye con una unidad a la entrada T_{ij} . De esta forma, el conjunto de aeropuertos puede considerarse como una red espacial dirigida y ponderada, en la que cada enlace entre dos aeropuertos tiene un peso que corresponde al número de pasajeros que viajan entre dichos aeropuertos.

En la Tabla 3.5 se muestra para cada año considerado, el número de aeropuertos activos así

Tabla 3.5: Número de pasajeros totales que viajan entre aeropuertos activos. En (a) y en (b) se presenta para Market y Coupon respectivamente y para cada años desde 2010 hasta 2018, el número de pasajeros totales que viajan entre aeropuertos activos, el número de aeropuertos activos N y la fracción de pasajeros resultantes respecto al total de pasajeros en la base de datos original. Un aeropuerto i se considera activo si el número de pasajeros que llegan y salen de este es mayor o igual a 1000.

(a) Market			
Año	\mathcal{T}	N	Fracción de los datos originales (%)
2010	43020628	332	99.8
2011	43600643	339	99.8
2012	43818899	331	99.8
2013	44386869	340	99.8
2014	45827090	342	99.8
2015	48446289	338	99.8
2016	51183778	339	99.9
2017	53528222	340	99.9
2018	56738321	345	99.9

(b) Coupon			
Año	\mathcal{T}	N	Fracción de los datos originales (%)
2010	58918111	333	99.8
2011	59663048	343	99.8
2012	59806954	333	99.8
2013	60124641	341	99.8
2014	61949408	342	99.9
2015	64956487	341	99.9
2016	67046661	341	99.9
2017	68948084	342	99.9
2018	72517675	348	99.9

como el total de pasajeros que viajan entre estos aeropuertos. Con el propósito de estudiar la actividad espacial de la red, también se define la matriz de distancias \mathbf{D} cuya entrada d_{ij} corresponden a la distancia recorrida en un vuelo que parte desde el aeropuerto i y que llega al aeropuerto j . Estas matrices se obtienen para todos los años desde 2010 hasta 2018 y permiten analizar la dinámica espacial de la red de aeropuertos en Estados Unidos. En la Figura 3.3 se muestran la matriz de Origen-Destino \mathbf{T} y la matriz de distancias \mathbf{D} de los conjuntos de datos Market y Coupon que corresponden al año 2018. El valor de cada entrada de estas matrices está codificado mediante un conjunto de colores, de tal forma que a cada valor le corresponde un único color. De esta figura es posible observar la diferencia entre los conjuntos de datos Market y Coupon; particularmente, en el caso de Market, la matriz de Origen-Destino tiene un mayor número de entradas distintas de cero. De acuerdo a la relación entre las tablas Market y Coupon explicada en la sección anterior, esto último se traduce en que, en la mayoría de los casos, dados dos aeropuertos i, j , es posible encontrar un registro en Market (un vuelo) de tal forma que el aeropuerto de origen es i y el aeropuerto de destino es j o viceversa. Por el contrario, puesto que cada registro en Coupon corresponde a un vuelo de conexión o escala dentro del vuelo completo, gran

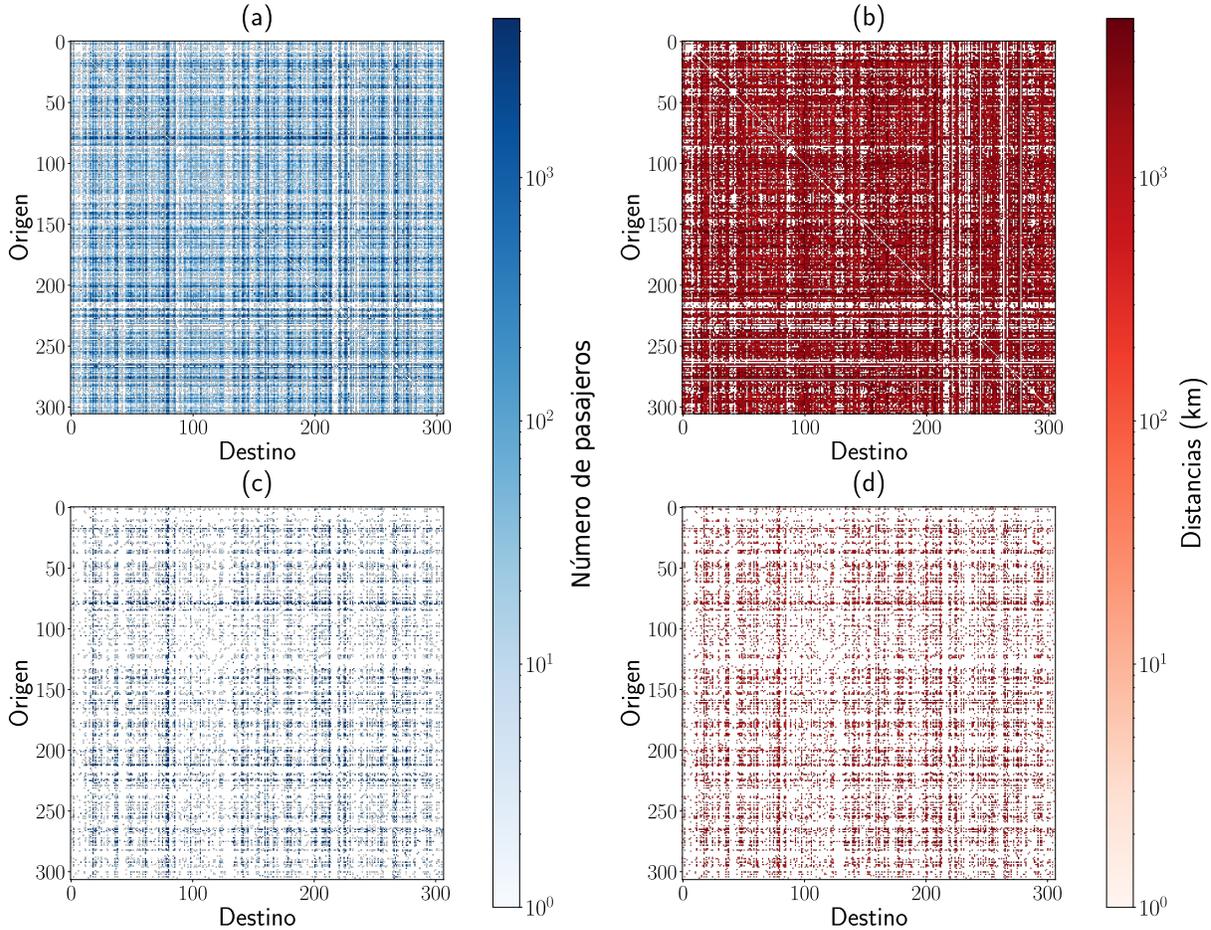


Figura 3.3: Matrices de origen-destino y de distancias del año 2018. En (a) y (b) se muestran para Market las matrices de origen-destino \mathbf{T} y distancias \mathbf{D} respectivamente. En (c) y (d) se muestran las respectivas matrices para Coupon. El valor de cada entrada es codificado mediante una barra de colores que representa el número de vuelos T_{ij} entre dos aeropuertos en el caso de la matriz de origen-destino y la distancia total recorrida d_{ij} (en vuelo) entre dos aeropuertos en el caso de la matriz de Distancias. De (a) y (c) podemos observar la diferencia entre los conjuntos de datos Market y Coupon: la matriz de origen-destino en (a) en comparación con (b) tiene un mayor número de entradas distintas de cero, esto puede verse en la diferencia de las regiones coloreadas entre las dos matrices. Lo anterior se traduce en que, en su mayoría, los vuelos de pasajeros no son directos y dados cualesquiera dos aeropuertos, se puede viajar entre estos en vuelos que involucran pasar por aeropuertos intermedios.

parte de las entradas en la matriz Origen-Destino son iguales a cero, en otras palabras, la mayoría de los vuelos involucran pasar a través de aeropuertos intermedios entre el punto de partida y el destino final. Lo anterior muestra indicios de que la red de aeropuertos está altamente conectada, es decir, es posible viajar entre cualesquiera dos aeropuertos de la red en un número de pasos mayor o igual que uno, donde cada paso corresponde a un movimiento entre dos aeropuertos. También se puede observar el alto grado de simetría de las matrices lo cual significa que para cada aeropuerto, el número de pasajeros en vuelos de salida es aproximadamente igual a el número de pasajeros en vuelos de llegada. Este comportamiento también se observa en las matrices de Origen-Destino de los años anteriores.

Tabla 3.6: Número de aeropuertos por comunidad para el año 2018. Para cada gráfica se presenta el número de nodos o aeropuertos de cada una de las comunidades de los conjuntos de datos Market y Coupon.

Comunidad	Número de nodos (Market)	Número de nodos (Coupon)
1	49	78
2	157	59
3	119	103
4	20	108

A partir de la representación de esta red de aeropuertos como una red espacial dirigida y ponderada, es posible estudiar y calcular ciertas medidas estadísticas como las definidas en la Sección 2.3 que caracterizan la estructura y la dinámica de cualquier red asociada a un sistema con un gran número de elementos e interacciones. La información contenida en estas medidas permite obtener un mayor entendimiento de los procesos que ocurren en dicho sistema. En primer lugar, se puede explorar las diferentes comunidades de aeropuertos que conforman la red. Esto permite identificar grupos de aeropuertos cuyos miembros están altamente conectados entre sí, es decir, existen una gran cantidad de vuelos de entrada o salida entre cada par de aeropuertos pertenecientes a una comunidad determinada, además de observar cómo se relacionan estas comunidades con la distribución geográfica de los aeropuertos.

Para detectar comunidades en redes, existen diversos algoritmos que minimizan la *modularidad*, esta cantidad representa el número de conexiones entre elementos de una comunidad en comparación con el número de conexiones o enlaces entre comunidades diferentes [45]. En particular, el algoritmo de *Louvain* permite obtener las comunidades de una red ponderada con un gran número de nodos en intervalos de tiempo relativamente cortos [46]. Como ya se ha mencionado, la red de aeropuertos en Estados Unidos es una red ponderada y dirigida. Existen diversos métodos para detectar comunidades en redes dirigidas, sin embargo la interpretación del significado de estas comunidades es más complicada que en el caso de las redes no dirigidas [47, 48]. Por lo tanto, para obtener las comunidades en este caso particular, se parte de la suposición de que la red de aeropuertos es no dirigida. Esta suposición resulta justificada cuando se toma en cuenta que las matrices de Origen-Destino son aproximadamente simétricas como puede observarse en la Figura 3.3. El significado de lo anterior es que en general el número de vuelos que salen y el número de vuelos que llegan a un aeropuerto determinado es aproximadamente el mismo. En la Figura 3.4 se presentan para Market y Coupon, las comunidades obtenidas utilizando el algoritmo de *Louvain* para el año 2018 sin considerar particiones iniciales del conjunto de aeropuertos ni parámetros adicionales. El color de cada nodo representa la comunidad a la que pertenece y el número de enlaces que salen o llegan a este, representa el número de conexiones entre el aeropuerto correspondiente y otros aeropuertos de la red.

De la figura se observa que los nodos o aeropuertos altamente conectados se concentran en el centro de cada comunidad, mientras que aquellos que no presentan muchas conexiones se sitúan en la periferia. Aquí se observa la diferencia entre Market y Coupon que se mencionó

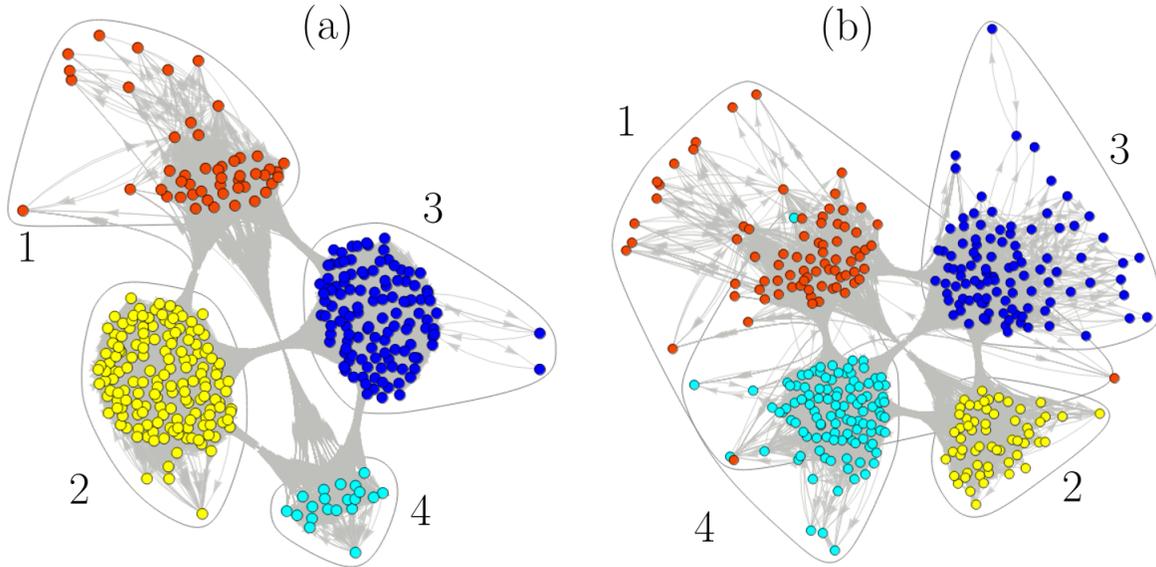


Figura 3.4: Gráficas y comunidades de la red de aeropuertos del año 2018 obtenidas mediante el algoritmo de Louvain. En (a) y en (b) se muestran las gráficas y comunidades de aeropuertos para los conjuntos de datos Market y Coupon respectivamente. Para cada nodo, el número de enlaces que se originan del mismo representa el número de conexiones con otros aeropuertos mientras que el color del nodo representa la comunidad a la que pertenece. En ambos casos, la dirección de las flechas representa la dirección de los vuelos entre nodos.

anteriormente; en el caso de Market la mayoría de aeropuertos están conectados entre sí, es decir, dados dos aeropuertos siempre existe un vuelo (que involucra pasar por aeropuertos intermedios) que los conecta. Esto explica el hecho de que la mayoría de nodos se concentra en el centro de cada comunidad. En el caso de Coupon una porción significativa de nodos se encuentra en la periferia de su comunidad respectiva. Esto refleja el hecho de que la mayoría de aeropuertos no se pueden conectar por medio de vuelos directos sino por vuelos que involucran escalas en aeropuertos intermedios. En ambos casos se obtuvo que el número de comunidades es igual a 4 y de la figura puede observarse que una de estas presenta un número de nodos pequeño en comparación con el resto de las comunidades. Por otra parte, en la Tabla 3.6 se muestra para Market y Coupon, el número de nodos en cada comunidad. En la Figura 3.5 se muestra para Market y Coupon, la distribución geográfica de las comunidades utilizando las coordenadas geográficas (longitud y latitud) de cada aeropuerto de la red. Esta representación permite observar la organización de los aeropuertos de cada comunidad y cómo se relaciona cada una de estas con el resto de las comunidades. Por ejemplo, en el caso de Coupon (Figura 3.5(b)), se muestra que las comunidades dividen al territorio de Estados Unidos en cuatro zonas bien definidas que podrían identificarse como zonas Oeste, Norte, Este y Sur. Por el contrario, en el caso de Market, no se observa una división tan clara de las comunidades las cuales se traslapan geográficamente (Figura 3.5(a)). Sin embargo, en ambos casos, es posible notar que ciertos elementos pertenecientes a una comunidad permiten conectar partes alejadas del territorio. Esto se observa con mayor claridad en la comunidad 3 (color azul) que abarca la parte occidental del territorio estadounidense y que contiene elementos que la

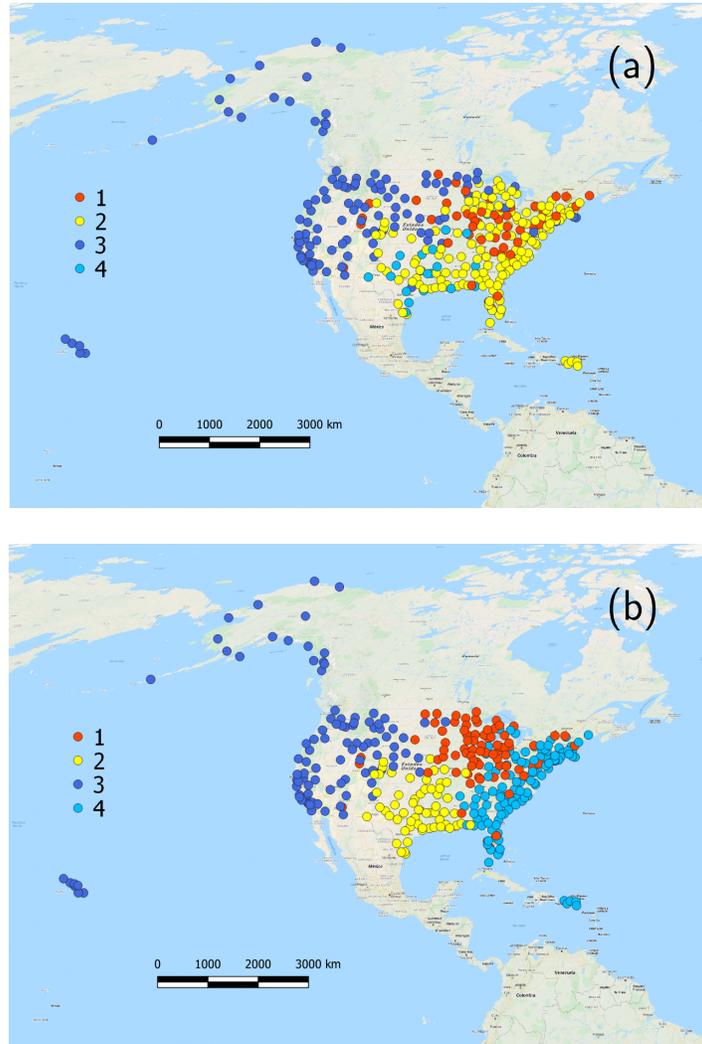


Figura 3.5: Distribución geográfica de las comunidades para el año 2018. En (a) y (b) se muestra la disposición geográfica de las comunidades de Market y Coupon respectivamente. El color de cada nodo o aeropuerto representa la comunidad a la que pertenece. En (a) se observa que no hay regiones geográficas bien definidas por las comunidades. Por el contrario, en (b) se observa que las comunidades dividen al territorio en cuatro regiones bien definidas (Norte, Sur, Este y Oeste). Nótese que en (a) y (b), la comunidad número 3 contiene nodos o aeropuertos que permiten conectar con otras comunidades y en particular con parte alejadas del territorio.

conectan con la parte oriental del territorio, Alaska y con islas como Hawaii y Puerto Rico. Estos nodos o aeropuertos son particularmente importantes para conectar regiones distantes del territorio.

Otra medida importante que permite caracterizar la red de aeropuertos donde se lleva a cabo la movilidad es la longitud de camino mínimo promedio $\langle l \rangle$ definida por la ecuación (2.5). Aún cuando la red de aeropuertos tiene asociada una medida de distancia que corresponde al número de kilómetros que se recorren durante un vuelo para llegar de un aeropuerto a otro, es interesante considerar una definición alternativa de la distancia

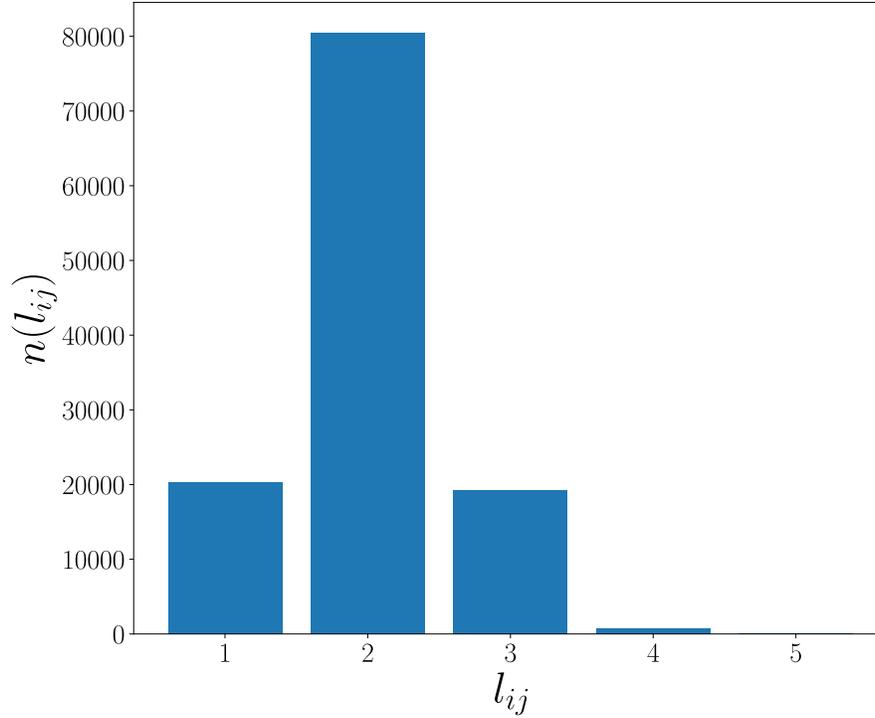


Figura 3.6: Longitudes de camino mínimo de la red de aeropuertos del año 2018. Para el año 2018 se muestra el número de pares de aeropuertos n cuya longitud de camino mínimo, es decir, el número de enlaces mínimo para conectarlos es igual a l_{ij} . Los resultados indican que cada par de aeropuertos de la red puede conectarse con un número no mayor de 5 enlaces entre aeropuertos intermedios. Esto es indicativo de que la red de aeropuertos en Estados Unidos presenta la propiedad de mundo pequeño.

entre dos nodos de la red como el número mínimo de enlaces en la red que se necesitan para conectar dichos nodos. Esta distancia se denomina longitud de camino mínimo y no depende de la distancia geográfica entre un par de aeropuertos. Dados dos aeropuertos i, j esta distancia se denota como l_{ij} y la longitud de camino mínimo promedio es el promedio sobre todos los pares de aeropuertos posibles. Para calcular la longitud de camino mínimo promedio de la red de aeropuertos en Estados Unidos, se construye la matriz de adyacencia \mathbf{A} (véase sección 2.3) como aquella cuyo elemento A_{ij} está dado por la relación

$$A_{ij} = \begin{cases} 1 & \text{si } T_{ij} \neq 0 \\ 0 & \text{si } T_{ij} = 0 \end{cases} \quad (3.1)$$

donde \mathbf{T} es igual a la matriz de Origen-Destino.

La matriz de adyacencia permite conocer si dos aeropuertos están conectados entre sí por al menos un vuelo. Puesto que los datos en Market corresponden a los destinos finales de los pasajeros y no contiene información acerca de los aeropuertos intermedios que son transitados durante el vuelo completo, únicamente se calcula l_{ij} para el conjunto de datos Coupon. Utilizando el algoritmo de *Lloyd* es posible calcular las longitudes de camino mínimo para cada par de aeropuertos.

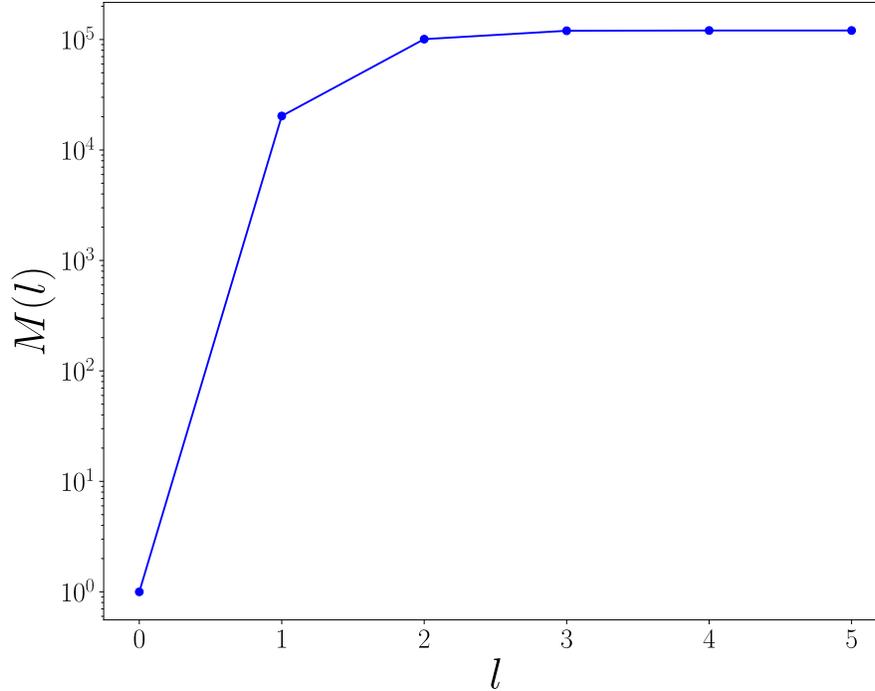


Figura 3.7: Longitud de camino mínimo promedio de la red de aeropuertos del año 2018. Para el año 2018 se muestra la función $M(l)$ cuyo valor representa el número de nodos promedio que se encuentran a una distancia menor o igual a l de cualquier vértice de la red. Esta función permite conocer si una red presenta la propiedad de mundo pequeño cuando no se tienen datos de la misma red para diferentes escalas. En el caso de redes con esta propiedad, la función sigue un crecimiento al menos de forma exponencial. En el caso particular de la red de aeropuertos en Estados Unidos, se observa que esta función crece exponencialmente para los primeros dos valores de l , posteriormente la función llega a un valor de saturación debido al tamaño finito de la red. Nótese que la escala de gráfica es lineal en el eje de las abscisas y logarítmica en el eje de las ordenadas.

En la Figura 3.6 se muestra el número de pares de aeropuertos con diferentes longitudes de camino mínimo, en esta se puede observar que la red de aeropuertos tiene la siguiente propiedad: cada par de aeropuertos puede conectarse mediante un número pequeño de desplazamientos entre aeropuertos intermedios, específicamente, entre 1 y 5 desplazamientos. En efecto, si se calcula la longitud de camino mínimo promedio dada por la ecuación (2.5) se obtiene que $\langle l \rangle = 2.0$. En la teoría de redes esto se conoce como la propiedad de mundo pequeño (*small world property*) y corresponde al hecho de que se puede viajar desde un nodo hacia cualquier otro, ambos pertenecientes a la misma red, mediante un número pequeño de saltos entre nodos intermedios [1]. En el caso particular de la red de aeropuertos de Estados Unidos, se necesitan en promedio únicamente 2 enlaces para conectar cualquier par de aeropuertos.

Formalmente, la propiedad de mundo pequeño se refiere a redes en las cuales se satisface que $\langle l \rangle = O(\log N)$ conforme $N \rightarrow \infty$, es decir, la longitud de camino mínimo promedio crece a lo sumo de forma logarítmica con el número de nodos N de la red. Puesto que no siempre se tiene información sobre diferentes tamaños de la red, otra forma de medir la propiedad de mundo pequeño es a través de la función $M(l)$ definida como el número

de nodos promedio que se encuentran a una distancia menor o igual a l de cualquier otro vértice [1]. Para redes con la propiedad de mundo pequeño esta función crece al menos de forma exponencial. En la Figura 3.7 se muestra la gráfica de la función M para la red de aeropuertos en Estados Unidos en la cual se observa que dicha función crece exponencialmente hasta llegar a un valor de saturación debido al tamaño finito de la red [1].

Es interesante el hecho de que en la red de aeropuertos mundial el número de enlaces necesarios para conectar cualquier par de aeropuertos es menor o igual a cinco siendo la longitud de camino mínimo promedio $\langle l \rangle$ igual a 4 [49, 28]. Este comportamiento es similar al observado en la red de aeropuertos de Estados Unidos que requiere el mismo número de enlaces para conectar cualquier par de aeropuertos y para la cual $\langle l \rangle = 2$. Lo anterior muestra indicios de que las propiedades de la red de aeropuertos se mantienen aproximadamente constantes cuando se cambia la escala en la que se estudia el sistema; es decir, cuando el número de aeropuertos se incrementa.

Adicionalmente, se introducen los grados de la matriz OD que fueron definidos en la Sección 2.2 por medio de las ecuaciones (2.1) y (2.2), y que permiten caracterizar la importancia de cada uno de los aeropuertos. Aquí, se definen de nuevo los grados de la matriz OD en el contexto de la red de aeropuertos. El grado de entrada $\mathbf{K}^{(e)}$ es el vector cuya entrada i -ésima está dada por la relación

$$k_i^{(e)} = \sum_{j=1}^N T_{ji} \quad (3.2)$$

y representa el número total de pasajeros que llegan al aeropuerto i de la red. De la misma forma, el grado de salida $\mathbf{K}^{(s)}$ es el vector cuya entrada i -ésima se define a través de la relación

$$k_i^{(s)} = \sum_{j=1}^N T_{ij} \quad (3.3)$$

que representa el número total de pasajeros cuyo origen es el aeropuerto i . En ambos casos, N representa el número de aeropuertos activos.

De las ecuaciones (3.2) y (3.3), se observa que los grados de la matriz OD satisfacen la siguiente relación

$$\sum_{i=1}^N k_i^{(e)} = \sum_{i=1}^N k_i^{(s)} = \sum_{i=1}^N \sum_{j=1}^N T_{ij} = \mathcal{T} \quad (3.4)$$

donde \mathcal{T} es el número de vuelos totales cada año.

A partir de las matrices de origen-destino de Market y Coupon representadas en la Figura 3.3 es posible obtener los grados de entrada y salida utilizando las ecuaciones (3.2) y (3.3). En la Figura 3.8 se presenta la densidad de probabilidad de los grados de las matrices OD para Market y Coupon. En cada gráfica, ambos ejes coordenados presentan una escala logarítmica lo cual permite observar que los datos siguen una tendencia lineal, en la figura esto se representa mediante una línea recta (a puntos) de pendiente -1. En particular, una

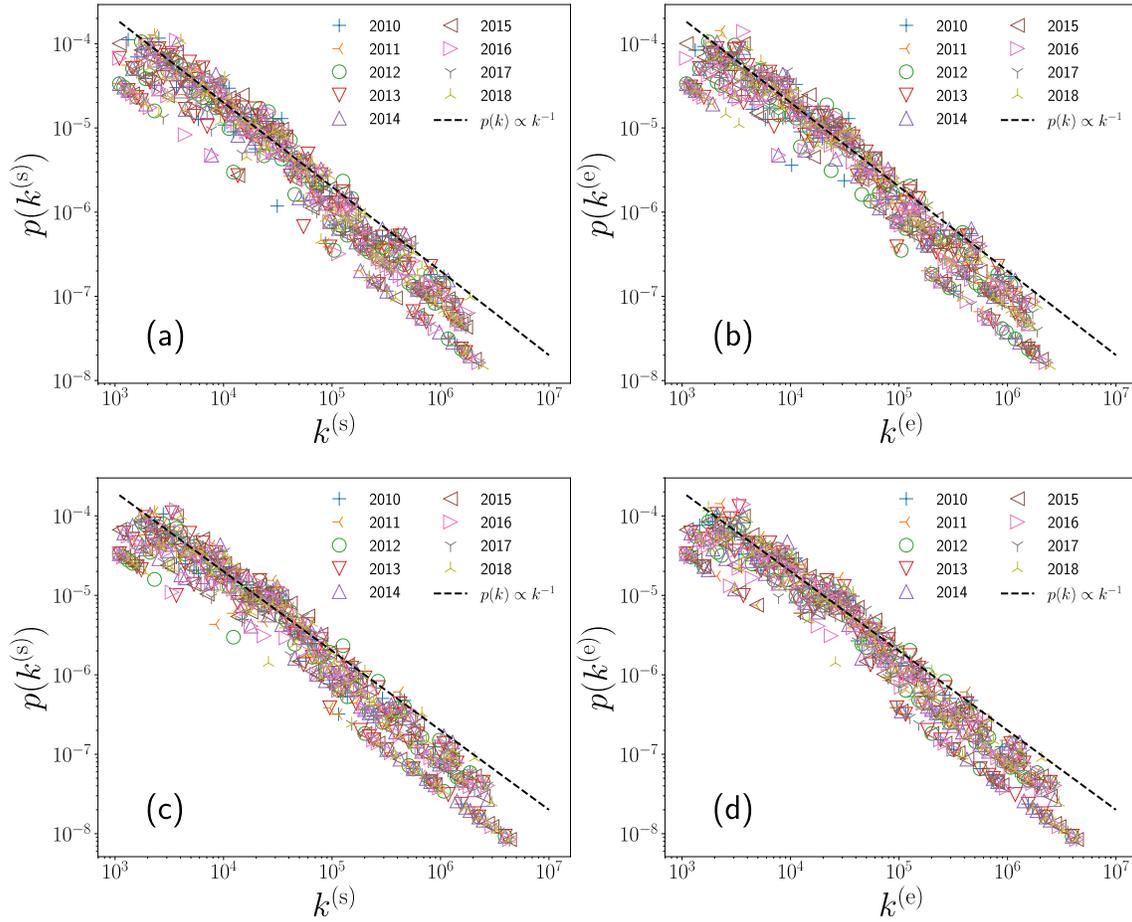


Figura 3.8: Densidad de probabilidad de los grados de la matriz OD. En (a) y (b) se muestra las densidades de distribución de los grados de salida $k_i^{(s)}$ y entrada $k_i^{(e)}$ respectivamente de la matriz OD correspondiente a Market. En (c) y (d) se muestran las densidades de probabilidad correspondientes a Coupon. En todos los casos, los ejes coordenados siguen una escala logarítmica y una línea punteada con pendiente -1 sirve como guía para resaltar la tendencia lineal de los datos. Para obtener la densidad de probabilidad (pdf) se graficó el histograma de los grados utilizando intervalos cuyo tamaño aumenta de forma logarítmica, esto permite una mejor visualización de los datos.

de las posibles distribuciones que pueden ajustarse a estos datos es una ley de potencias $p(k) \propto k^{-\alpha}$ ya que al aplicar la función logaritmo en ambos lados de la relación se obtiene la ecuación de una recta con pendiente negativa α . Sin embargo, otras distribuciones de cola pesada como la ley de potencias truncada, la distribución log-normal o la distribución de estiramiento exponencial (stretched exponential) podrían describir mejor los datos que se presentan en la Figura 3.8.

Actualmente, existen diversos métodos estadísticos que permiten ajustar y medir la calidad del ajuste de este tipo de distribuciones de forma precisa. Por ejemplo, en el caso de la distribución de ley de potencias, se ha demostrado que el método de máxima verosimilitud (MLE, por sus siglas en inglés) es un método más confiable para el ajuste de este

Tabla 3.7: Estadística de Kolmogorov-Smirnov de la distribución de los grados de la matriz OD. Los valores D_e y D_s representan la estadística KS de los grados $k_i^{(e)}$ y $k_i^{(s)}$ respectivamente. En (a) y (b) se presenta, utilizando los datos en Market y en Coupon respectivamente, la estadística KS de cada una de las distribuciones consideradas. En ambos casos se considera $x_{min} = 1000$, en acuerdo con el criterio establecido para verificar si un aeropuerto es activo.

(a) Market		
Distribución	D_e	D_s
Exponencial	0.377	0.377
Ley de potencias	0.1894	0.1891
Ley de potencias truncada	0.0776	0.0771
Exponencial estirada	0.0465	0.0460
Log-normal	0.0388	0.0392
(b) Coupon		
Distribución	D_e	D_s
Exponencial	0.421	0.421
Ley de potencias	0.186	0.185
Ley de potencias truncada	0.070	0.070
Exponencial estirada	0.0477	0.0477
Log-normal	0.0332	0.0328

tipo de distribución en comparación con el método usual de regresión lineal por mínimos cuadrados. Estos métodos también permiten comparar entre dos distribuciones y decidir cuál de ellas se ajusta mejor a los datos [50].

De esta forma, para caracterizar la distribución de los grados de la matriz OD se lleva a cabo un ajuste de distribuciones de cola pesada sobre los datos que se muestran en la Figura 3.8. Utilizando el paquete de software en lenguaje Python *powerlaw*, es posible ajustar diversas distribuciones de cola pesada. En particular, en este trabajo se consideran las siguientes distribuciones

$$p(x) = e^{-\lambda x} \quad (\text{Exponencial}) \quad (3.5)$$

$$p(x) = x^{-\alpha} \quad (\text{Ley de potencias}) \quad (3.6)$$

$$p(x) = x^{-\alpha} e^{-\lambda x} \quad (\text{Ley de potencias truncada}) \quad (3.7)$$

$$p(x) = x^{\beta-1} e^{-\lambda x^\beta} \quad (\text{Exponencial estirada}) \quad (3.8)$$

$$p(x) = \frac{1}{x} \exp \left[\frac{-(\ln x - \mu)^2}{2\sigma^2} \right] \quad (\text{Log-normal}) \quad (3.9)$$

En primer lugar, es necesario verificar la plausibilidad de suponer que los datos obedecen una distribución de cola pesada. Para esto se tiene que comprobar que alguna de las distribuciones consideradas se ajusta mejor a los datos en comparación con la distribución exponencial de la ecuación (3.5), esto debido a que por definición una distribución de cola pesada es aquella que no está acotada por una distribución exponencial para valores

suficientemente grandes. Si el resultado indica que la distribución elegida no ajusta mejor a los datos entonces no existen argumentos para suponer que la distribución que genera los datos es en efecto de cola pesada.

Lo anterior puede verificarse mediante una prueba de razón de probabilidad, en inglés, *Likelihood-ratio test* (LR) en la cual se calcula el logaritmo de la razón de las probabilidades de dos modelos que posiblemente ajustan a los datos [50]. En el caso en que el nivel de significancia p asociado al resultado satisface que $p < 0.05$ se puede concluir dependiendo del signo de la razón LR que alguna de las distribuciones es un mejor ajuste. Sin embargo esta prueba no garantiza que en efecto los datos sigan esta distribución, únicamente permite elegir entre dos posibles ajustes. Por otra parte, para medir la calidad de un ajuste se utiliza la estadística de Kolmogorov-Smirnov (KS) definida como [50]

$$D = \max_{x \geq x_{min}} |S(x) - P(x)|$$

donde $S(x)$ y $P(x)$ son respectivamente, la función de distribución acumulada (FDA) obtenida a partir de los datos y la correspondiente FDA del modelo seleccionado para el ajuste, x_{min} representa una cota inferior a partir de la cual se observa que los datos siguen una distribución de cola pesada. Un mejor ajuste implica un menor valor de D . La estadística de Kolmogorov-Smirnov D , también puede ser utilizada para estimar el valor de x_{min} considerando diversos valores de x y seleccionando aquel que minimice D [50].

Al comparar la distribución de ley de potencias con la distribución exponencial para cada uno de los conjuntos de datos presentados en la Figura 3.8, se obtiene mediante la prueba LR que la ley de potencias es un mejor modelo para describir la distribución de los datos. En todos los casos, el valor de significancia satisface $p \ll 0.05$ lo cual permite tener un alto grado de confianza en la afirmación anterior. Sin embargo, como ya se mencionó, otras distribuciones de cola larga pueden resultar en mejores ajustes a los datos. En la Tabla 3.7 se muestra la estadística D calculada para cada una de las distribuciones en las ecuaciones (3.5) a (3.9) ajustadas a los datos en Market y Coupon. Para $k_i^{(s)}$ y $k_i^{(e)}$ se consideró $x_{min} = 1000$ de acuerdo al criterio establecido para verificar si un aeropuerto es activo. De la Tabla 3.7 es posible observar que en el caso de Market y Coupon las distribuciones de las ecuaciones (3.7), (3.8) y (3.9) tienen un valor menor de la estadística D en comparación con la distribución de ley de potencias (3.6), esto último sugiere que estas tres distribuciones ajustan mejor a los datos. Para evaluar y comparar cual de estas describe mejor la distribución de los grados de la matriz OD se utiliza el test LR . Los resultados indican que en efecto, la distribución de ley de potencias truncada es un mejor ajuste que la ley de potencias; sin embargo, tanto la distribución exponencial estirada como la distribución log-normal superan a la ley de potencias truncada. En todos estos casos el nivel de significancia satisface $p < 0.05$. Finalmente, comparando entre las distribuciones log-normal y exponencial estirada se obtiene en el caso de Market, que estadísticamente ninguna de las dos es superior a la otra ya que $p \approx 0.90$, mientras que en Coupon la distribución log-normal parece un mejor ajuste pues $p = 0.02$. Sin embargo, como ya se dijo, esto no es indicativo de que la distribución de los grados de la matriz OD sea en efecto una distribución de tipo log-normal, solamente se puede concluir que es un mejor ajuste.

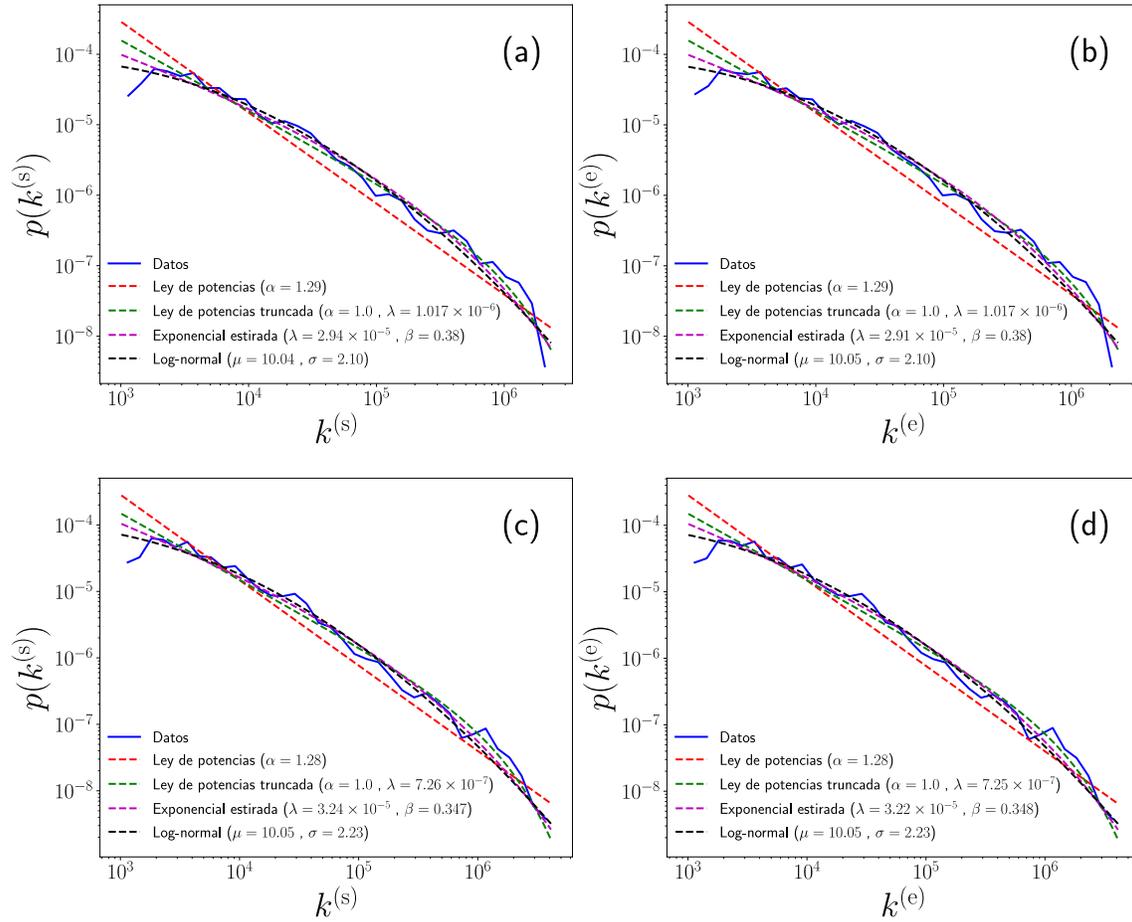


Figura 3.9: Densidad de probabilidad de los grados de la matriz OD y distribuciones ajustadas. En (a) y (b) se muestra las densidades de probabilidad de los grados de salida ($k_i^{(s)}$) y entrada ($k_i^{(e)}$) respectivamente de la matriz OD correspondiente a Market. En (c) y (d) se muestran las densidades de probabilidad correspondientes a Coupon. En todos los casos, también se muestra la gráfica de las distribuciones ajustadas y se especifican los parámetros estimados mediante el método de máxima verosimilitud.

En la Figura 3.9 se muestran para Market y Coupon, la función de densidad de probabilidad de los grados de la matriz OD así como las distribuciones ajustadas utilizando los parámetros estimados mediante el método de máxima verosimilitud.

Una pregunta natural que surge al observar la distribución de los grados de la matriz OD y sus respectivos ajustes es el por qué dicha distribución no sigue una ley de potencias en todo su intervalo, es decir, por qué la distribución se trunca a partir de un cierto valor de k . Esto se debe a las limitaciones físicas de cada aeropuerto, en particular, los aeropuertos con un mayor número de conexiones (aquellos con un mayor número de vuelos de salida o llegada) no tienen una capacidad ilimitada para manejar un número infinito de vuelos de llegada o salida. Por lo tanto, la capacidad de un aeropuerto limita el número de posibles conexiones con otros elementos de la red [49].

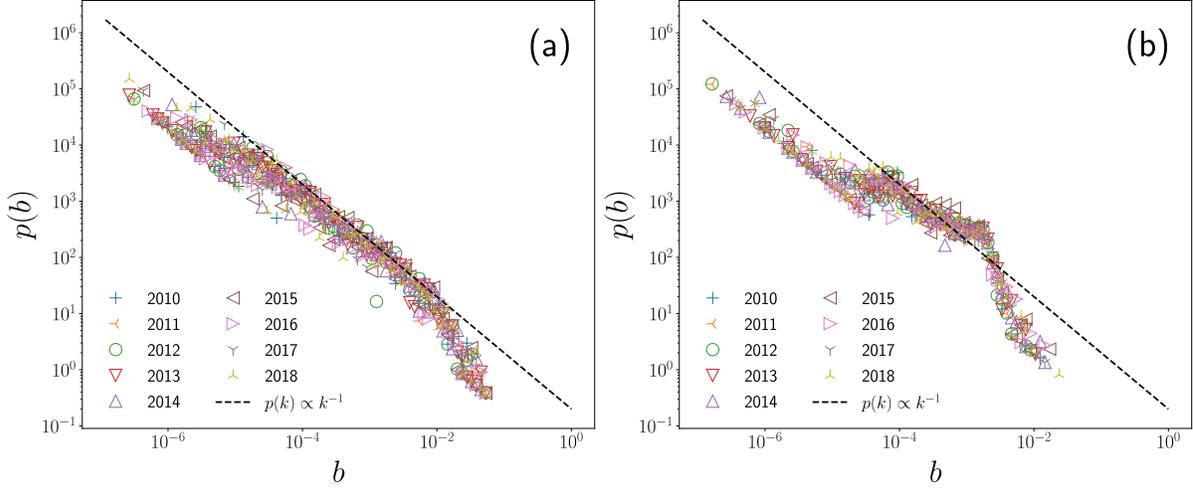


Figura 3.10: Densidad de probabilidad de la centralidad de los aeropuertos en Estados Unidos. En (a) y en (b) se muestra respectivamente, la densidad de probabilidad de la centralidad (*Betweenness centrality*) de los conjuntos de datos Market y Coupon para cada año desde 2010 hasta 2018. Los ejes coordenados presentan una escala logarítmica y como guía se ha utilizado una línea punteada con pendiente -1 que permite observar que ambas distribuciones presentan características de una distribución de cola pesada y que a partir de un valor de b dichas distribuciones se truncan debido al tamaño finito de la red.

Por otra parte, es posible estudiar la distribución del rango de los grados de la matriz OD. El rango permite ordenar a cada uno de los aeropuertos en términos del número de pasajeros que llegan o salen de estos. De esta forma, el rango número uno corresponde al aeropuerto con un mayor número de pasajeros. En diversos contextos, este tipo de ordenamiento da lugar a distribuciones que siguen leyes de potencias en un intervalo del rango; por ejemplo, la ley de Zipf, que describe el comportamiento de diversas cantidades relacionadas con actividades humanas en diferentes contextos como la escritura y la economía [51], establece que el rango r de la variable de interés sigue una distribución dada por:

$$f(r) \propto \frac{1}{r^a}. \quad (3.10)$$

En gran parte de los casos; sin embargo, la ley de potencias es válida solamente para una parte del rango. En la referencia [52], los autores proponen una distribución de dos parámetros que permite describir el comportamiento de una gran diversidad de fenómenos que surgen en las artes y en las ciencias sociales y naturales. Esta distribución, corresponde a la versión discreta de la función beta generalizada y está dada por:

$$f(r) = A \frac{(N - 1 + r)^b}{r^a}, \quad (3.11)$$

donde N es el rango máximo y A es una constante de normalización. Nótese que esta distribución se reduce a una ley de potencias en el caso cuando $b = 0$. La distribución 3.11

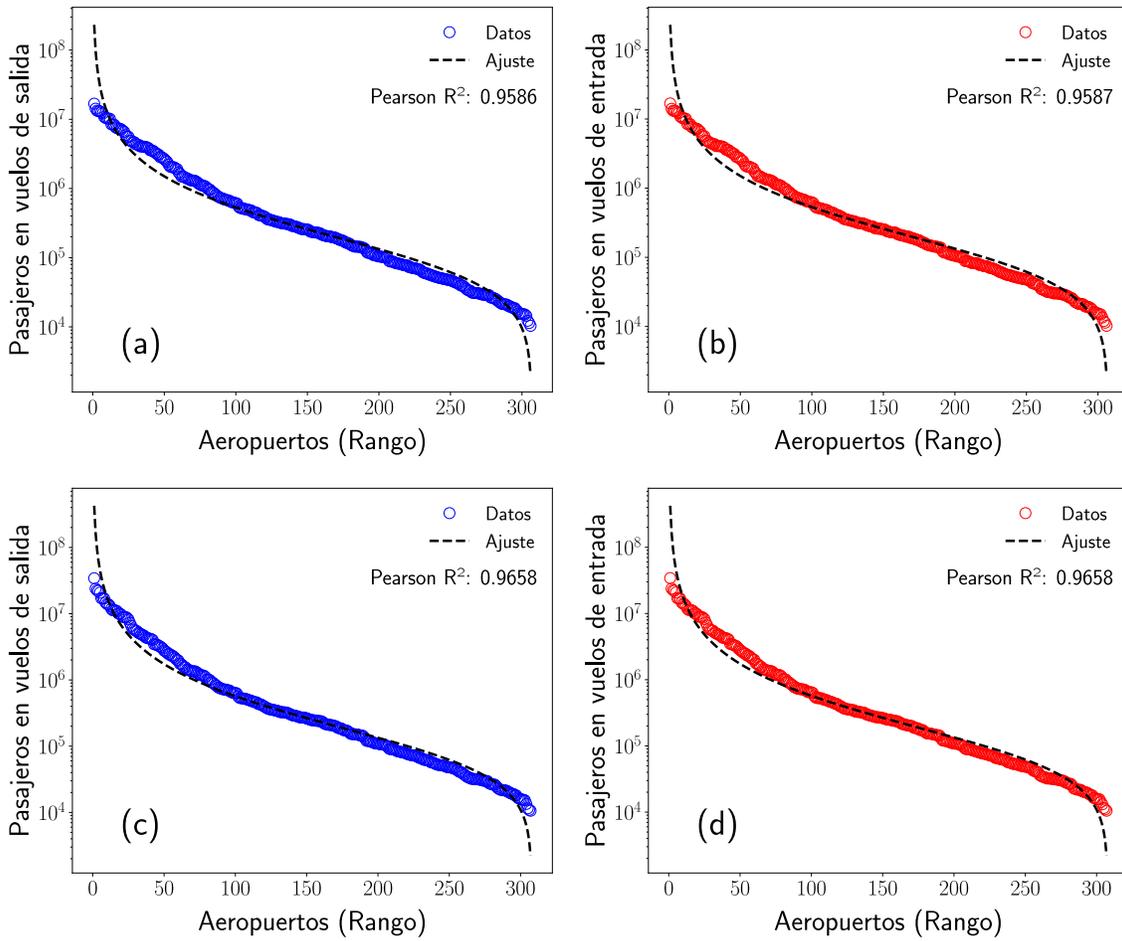


Figura 3.11: Rango de los grados de entrada y salida de la matriz OD. En (a) y (b) se muestra, respectivamente, el rango de los grados de salida ($k_i^{(s)}$) y entrada ($k_i^{(e)}$) de la matriz OD correspondiente a Market. En (c) y (d) se muestran los rangos correspondientes a Coupon. En todos los casos, se muestra (como una línea punteada) el ajuste obtenido por regresión lineal múltiple.

es particularmente útil para estudiar correcciones a la ley de potencias debido a efectos finitos o restricciones en la red y ha sido ampliamente utilizada en diversos contextos. Para aplicar esta distribución al caso particular de los aeropuertos en Estados Unidos, se aplica la función logaritmo en ambos lados de la ecuación 3.11 y se realiza una regresión lineal múltiple para obtener los parámetros a , b , A . Los resultados obtenidos de ajustar esta distribución al caso de la red de aeropuertos, ordenados de acuerdo al número de pasajeros, se muestran en la Figura 3.11 donde una línea a puntos representa el ajuste obtenido.

De la Figura 3.11 se observa que el coeficiente de correlación de Pearson R^2 es alto y permite obtener un ajuste de la distribución beta generalizada discreta que describe bien los resultados para rangos mayores a 100. Para el caso de valores con rango menor a 10, los resultados aunque cualitativamente son similares, no describen correctamente el rango

de estos aeropuertos de Estados Unidos.

Aunque los grados de la matriz OD permiten obtener una medida de la importancia de un aeropuerto en términos del número de pasajeros que llegan o salen del mismo, no proveen información acerca de la centralidad de un aeropuerto, es decir, qué tan importante es este aeropuerto para conectar cualquier par de aeropuertos o diferentes zonas de la red de la red. Para evaluar esta propiedad, se calcula la centralidad de intermediación (*Betweenness Centrality*) de un aeropuerto i definida en la ecuación (2.4). De esta forma, en el caso de la red de aeropuertos, se obtiene la siguiente ecuación

$$b_i = \sum_{i \neq j \neq k} \frac{\sigma(i)_{jk}}{\sigma_{jk}} \quad (3.12)$$

donde σ_{jk} es el número de caminos de longitud mínima que conectan a los aeropuertos j, k y $\sigma(i)_{jk}$ el número de dichos caminos que pasan a través de i . Esta cantidad representa la fracción de caminos de longitud mínima que atraviesan un aeropuerto en particular. Entre mayor sea esta cantidad, mayor es la importancia de un aeropuerto para conectar diferentes regiones de la red. Utilizando la matriz de adyacencia definida en la ecuación (3.1) de los conjuntos de datos Market y Coupon, se puede calcular la centralidad de los aeropuertos mediante el algoritmo de *Brandes* [53].

En la Figura 3.10 se muestra la densidad de probabilidad de la centralidad de los aeropuertos en la red aeroportuaria de los Estados Unidos para el conjunto de datos Market y Coupon. En ambos casos se puede observar que la distribución es de tipo cola pesada y que esta se trunca a partir de un valor lo suficientemente grande de b como sucede en el caso de los grados de la matriz OD. En este trabajo no se realizan ajustes a estas distribuciones; sin embargo, es interesante notar que los aeropuertos con mayor número de pasajeros no son necesariamente los aeropuertos con mayor centralidad como puede observarse en la Figura 3.12. En esta figura se han graficado para el año 2018, los diez aeropuertos con mayor número de pasajeros en vuelos de salida ($k_i^{(s)}$) y los diez aeropuertos con el valor de centralidad más grande dentro de la red para el año 2018.

Lo anterior resulta relevante al estudiar la movilidad de las personas en la red de aeropuertos ya que los grados de la matriz OD, o bien, la centralidad asociada a cada uno de los aeropuertos representan medidas diferentes de la importancia de un aeropuerto o del papel que juega dentro de la red. En el siguiente capítulo se estudiarán modelos de movilidad que consideran a los grados de la matriz OD como elementos importantes en la caracterización de la red de aeropuertos.

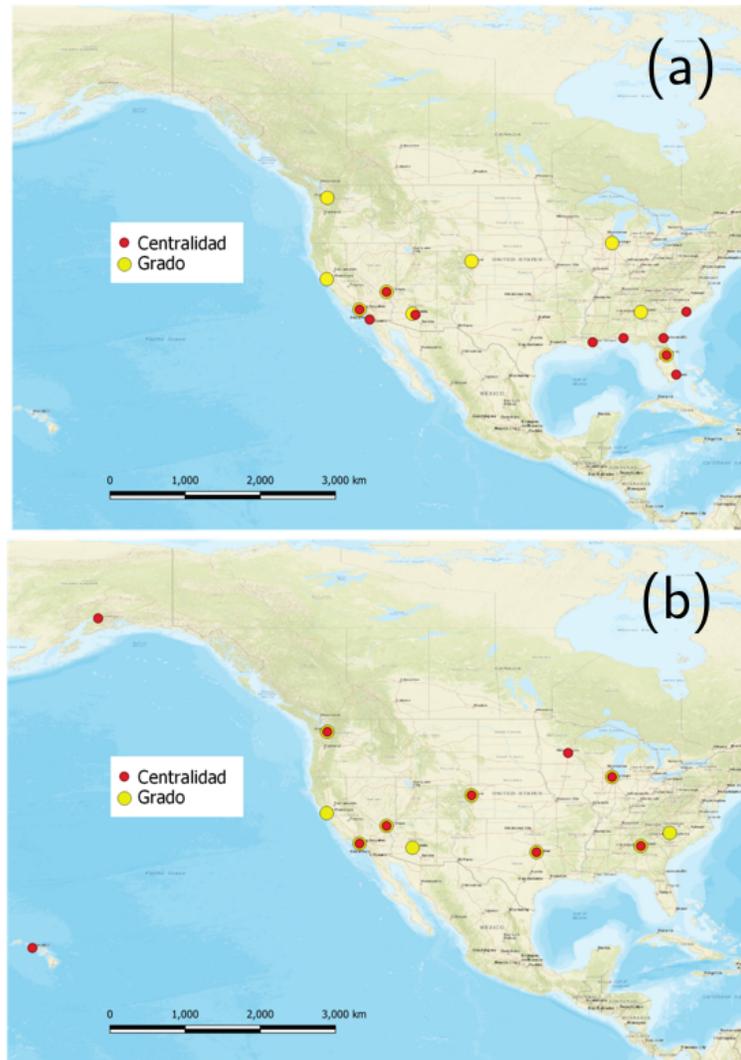


Figura 3.12: Aeropuertos con mayor número de pasajeros en vuelos de salida y aeropuertos con mayor centralidad del año 2018. En (a) y en (b) se muestran respectivamente para Market y Coupon, los diez aeropuertos con mayor número de pasajeros en vuelos de salida (en color amarillo) y los diez aeropuertos con los valores más grandes de centralidad (en color rojo). Se observa que en general, los aeropuertos con un mayor grado de salida no coinciden con aquellos que presentan una mayor centralidad.

4. Modelamiento del transporte en una red de aeropuertos

4.1. Introducción

En este capítulo se analizan las matrices de Origen-Destino desde un enfoque dinámico que va más allá de un tratamiento utilizando redes. Mediante estos métodos, se busca explicar el comportamiento de los pasajeros en las redes asociadas a las matrices OD construidas en el Capítulo 3 utilizando modelos de movilidad como aquellos que se presentaron en la Sección 2.4. El objetivo es caracterizar los patrones de movimiento que se observan en la matriz de Origen-Destino \mathbf{T} y en la matriz de distancias \mathbf{D} por medio de un número pequeño de parámetros.

En la Sección 4.2 se construyen las matrices de probabilidad de transición \mathbf{W} para cada una de las matrices OD consideradas. Estas matrices contienen información acerca de la probabilidad de que un pasajero realice un viaje entre cualesquiera dos aeropuertos de la red. A partir de las matrices de probabilidad y las matrices de distancias \mathbf{D} se analizan patrones de movimiento de los pasajeros. Con base a los resultados se propone, para cada una de estas redes, un modelo de movilidad basado en el modelo de tipo gravedad definido en la Sección 2.4, con un número relativamente pequeño de parámetros, para explicar los patrones observados. Se comparan los resultados obtenidos a partir de dichos modelos y se establece que los modelos reproducen de forma adecuada el comportamiento observado de los pasajeros.

Los resultados obtenidos en este capítulo demuestran que la movilidad de los pasajeros en la red de aeropuertos de Estados Unidos puede caracterizarse mediante modelos con un conjunto relativamente pequeño de parámetros. Además, estos parámetros tienen una interpretación simple en términos de cantidades conocidas; por ejemplo, la distancia de vuelo recorrida o el grado de salida de la matriz OD que denota una medida de la importancia de los aeropuertos de la red.

4.2. Matrices de probabilidad de transición

La matriz de origen-destino \mathbf{T} definida en la Sección 2.2 y que se construyó para los conjuntos de datos Market y Coupon contiene información que permite estudiar de forma estadística a la red de aeropuertos de Estados Unidos en términos del grado de salida $\mathbf{K}^{(s)}$ y del grado de entrada $\mathbf{K}^{(e)}$, estos últimos representados en las Figuras 3.8 y 3.9 con sus respectivos ajustes. Para relacionar la información contenida en la matriz OD con la información en la matriz de distancias \mathbf{D} , es decir, el número de pasajeros entre aeropuertos y las distancias de vuelo recorridas entre cada par de aeropuertos, se construye la matriz de probabilidad de transición \mathbf{W} definida en la Sección 2.2 por medio de la ecuación (2.3). En este caso particular, cada elemento $w_{i \rightarrow j}$ de la matriz de probabilidad de transición representa la probabilidad de que un pasajero viaje desde el aeropuerto de origen i hacia el aeropuerto de destino j y está dado por la siguiente relación

$$w_{i \rightarrow j}^{(OD)} = \frac{T_{ij}}{k_i^{(s)}}. \quad (4.1)$$

De la definición anterior se observa que los elementos de la matriz \mathbf{W} satisfacen la condición de normalización

$$\sum_{j=1}^N w_{i \rightarrow j}^{(OD)} = 1, \quad (4.2)$$

de tal forma que estas probabilidades están bien definidas. Esta condición implica que necesariamente ocurre un vuelo entre el aeropuerto i y cualquier otro aeropuerto de la red. A partir de la relación (4.1), la información en la matriz OD y la matriz de distancias \mathbf{D} se obtiene la relación entre los elementos $w_{i \rightarrow j}$ y d_{ij} por medio de la distribución de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ donde $d_0 = 1$ km representa una longitud de referencia.

En la Figura 4.1 y Figura 4.2 se muestran los análisis para las bases de datos Market y Coupon respectivamente a través del histograma bidimensional de dichos puntos para cada uno de los años considerados desde 2010 hasta 2018. En cada uno de los años, la frecuencia de cada punto está codificada mediante un conjunto continuo de colores. De la información en estos histogramas, se puede analizar, en el caso de Coupon, la dinámica de los vuelos que se llevan a cabo en los aeropuertos de la red, y en el caso de Market, el comportamiento de los pasajeros que viajan entre diferentes nodos de la red, es decir, la movilidad.

Es importante notar que en ambos casos, la dinámica de la red permanece constante a lo largo de los años que se consideran para el análisis tal y como puede observarse en cada uno de los recuadros de las Figuras 4.1 y 4.2. Esto sugiere que es posible construir un modelo que explique la dinámica de los vuelos de todos los años considerados en el estudio.

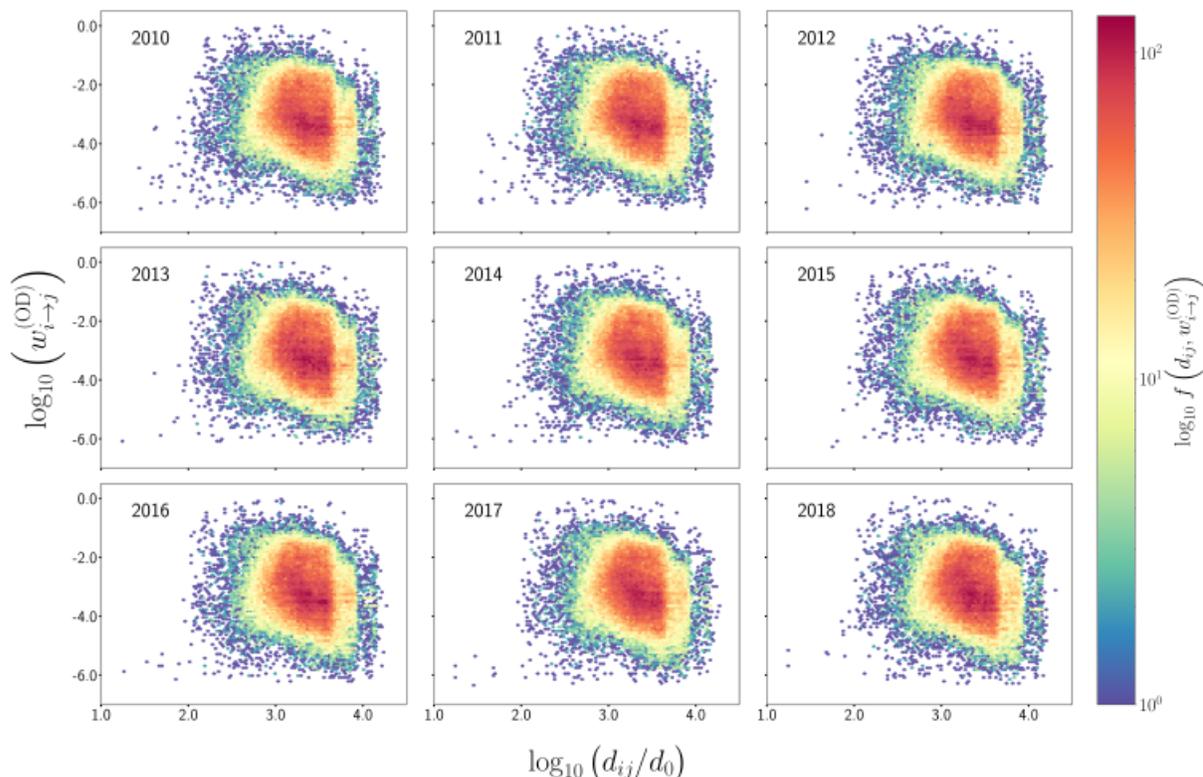


Figura 4.1: Histograma bidimensional del conjunto de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ en la base de datos Market. Para cada año desde 2010 hasta 2018 se muestra la frecuencia $f \left(d_{ij}, w_{i \rightarrow j}^{(OD)} \right)$ de los puntos $\left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} w_{i \rightarrow j}^{(OD)} \right)$ en Market mediante un histograma hexagonal bidimensional. El valor de la frecuencia se codifica en la barra de color.

4.3. Modelos de movilidad

En esta sección se proponen dos modelos que explican de forma aproximada la dinámica de los vuelos en la red de aeropuertos y que se observa en los patrones de las Figuras 4.1 y Figura 4.2. El objetivo de estos modelos es describir mediante un número reducido de parámetros los procesos de movilidad que ocurren en la red de aeropuertos y obtener predicciones de los desplazamientos entre aeropuertos utilizando la información contenida en la matriz de probabilidad de transición y en los grados de la matriz OD. Los modelos propuestos en esta sección se basan en aquellos que se discutieron en la Sección 2.4; en particular, en el modelo de tipo gravedad definido por medio de la ecuación (2.7). Este modelo tiene la ventaja de tener una forma matemática simple y de fácil interpretación, esto debido a la analogía establecida con la ley de gravitación universal.

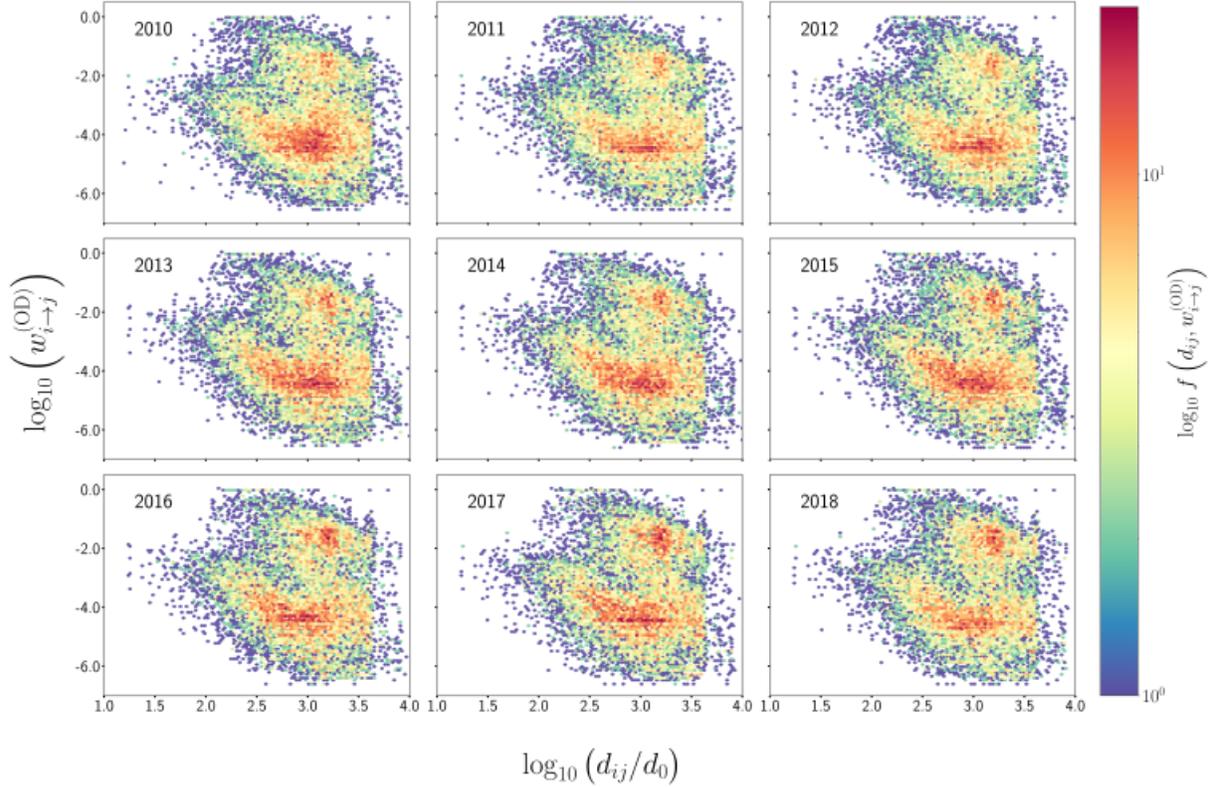


Figura 4.2: Histograma bidimensional del conjunto de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ en la base de datos Coupon. Para cada año desde 2010 hasta 2018 se muestra la frecuencia $f \left(d_{ij}, w_{i \rightarrow j}^{(OD)} \right)$ de los puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ en Coupon mediante un histograma hexagonal bidimensional. El valor de la frecuencia se codifica mediante un conjunto continuo de colores.

4.3.1. Modelamiento de la base de datos Market

En el caso de Market, se observa en la Figura 4.1 que la dinámica espacial de los pasajeros puede describirse de forma aproximada mediante un modelo que considera transiciones hacia zonas dentro de una localidad de radio R centrada en el aeropuerto de origen y que son independientes de la distancia d_{ij} . Estas transiciones son de tipo local pues ocurren entre aeropuertos en los que la distancia de vuelo satisface $d_{ij} < R$. Para distancias mayores que R , las transiciones están descritas por probabilidades que dependen de la distancia de vuelo entre dos aeropuertos d_{ij} .

Específicamente, se considera un modelo que clasifica los vuelos de pasajeros en dos tipos: los vuelos locales cuya distancia recorrida es menor que una distancia R , para los cuales la probabilidad de viajar entre los aeropuertos i, j es proporcional únicamente al grado de salida del aeropuerto j elevado a una potencia β y vuelos de largo alcance para los cuales la distancia recorrida es mayor que R y cuya probabilidad de transición es inversamente proporcional a cierta potencia γ de la distancia d_{ij} , similar a los modelos de movilidad tipo-gravedad 2.7. De esta forma, el modelo considera tres cantidades o parámetros, R, β y

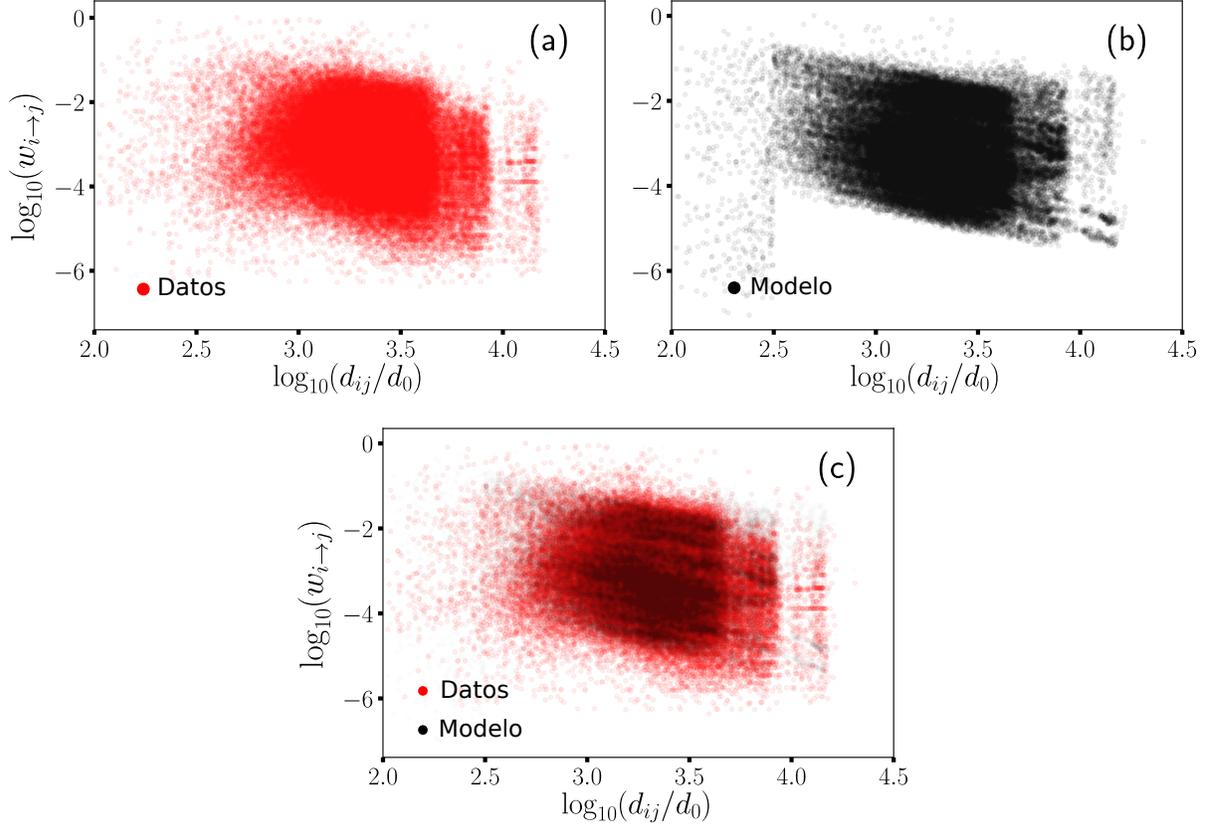


Figura 4.3: Comparación entre datos y modelo del conjunto de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ correspondiente a Market. En (a) se muestra el conjunto de puntos obtenido a partir de los datos empíricos. En (b) se muestra el conjunto de puntos obtenido a partir del modelo (4.3). En (c) se muestran, superpuestos, ambos conjuntos de puntos para facilitar su comparación. Los parámetros utilizados para generar la información correspondiente al modelo son presentados en la Tabla 4.1.

γ que deben determinarse a partir del conjunto de datos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$. Matemáticamente, lo anterior se expresa a través del número de vuelos T_{ij}^{modelo} que corresponde a la matriz de Origen-Destino del modelo y se define como:

$$T_{ij}^{modelo} = \begin{cases} \left(k_j^{(s)} \right)^\beta & \text{si } d_{ij} \leq R, \\ \frac{k_i^{(s)} k_j^{(s)}}{d_{ij}^\gamma} & \text{si } d_{ij} > R. \end{cases} \quad (4.3)$$

La interpretación de los parámetros del modelo (4.3) es simple: para cualquier aeropuerto i las transiciones hacia aeropuertos dentro de una vecindad de radio R están caracterizadas por la importancia del aeropuerto de destino j , dada por el número de vuelos de salida. El exponente β cuantifica el impacto del grado de salida $k_j^{(s)}$ en la probabilidad de que ocurra un viaje entre los aeropuertos i y j . En efecto, si $\beta \rightarrow 0$ entonces $\left(k_j^{(s)} \right)^\beta \rightarrow 1$ para todo j tal que $d_{ij} < R$ y por lo tanto las probabilidades de transición son las mismas para todos los aeropuertos dentro de esta localidad. Por otra parte, R determina el tamaño o

radio de dicha localidad y expresa el hecho de que la distancia juega un papel importante en la decisión de un pasajero sobre realizar un viaje o no siempre que $d_{ij} > R$. Mientras mayor sea la distancia de vuelo entre dos aeropuertos, menor es la probabilidad de que ocurra un viaje entre ellos.

Para demostrar visualmente que el modelo propuesto reproduce correctamente el comportamiento de la red de aeropuertos (al menos de forma aproximada), en la Figura 4.3 se muestra el conjunto de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ correspondiente al conjunto de datos Market, en los tres casos siguientes: En (a) se muestra dicho conjunto de puntos para la matriz de probabilidad de transición \mathbf{W} correspondiente a la matriz OD obtenida a partir los registros en la base de datos; en (b) se muestra este conjunto de puntos para la matriz de probabilidad de transición \mathbf{W}^{modelo} que se obtiene de la definición del modelo (4.3) y la ecuación (4.1); en (c) se superponen los conjuntos de puntos (a) y (b) para facilitar la comparación entre ambos. De esta figura es posible observar que en efecto, el modelo reproduce de forma aproximada el comportamiento de la red.

Una vez determinados los parámetros R y β , es posible obtener predicciones a partir del modelo propuesto acerca de el número de vuelos entre cada par de aeropuertos i y j . Para tal fin, se obtiene la matriz de probabilidades de transición del modelo, $w_{i \rightarrow j}^{modelo}$ mediante la definición dada en la ecuación (4.1).

4.3.2. Modelamiento de la base de datos Coupon

En el análisis de la base de datos Coupon, se presenta una situación diferente. Aquí, no se distinguen los vuelos en términos de la distancia de vuelo recorrida, en este caso se observan dos tipos de vuelos con una dinámica similar pero que ocurren con diferentes probabilidades. Dicho de otra forma, para cualquier par de aeropuertos i, j cuya distancia de vuelo es específicamente d_{ij} , existen dos tipos de viajes entre este par aeropuertos, uno de estos con menor probabilidad que el otro. Esto último puede notarse en la Figura 4.2 en la que para cada recuadro (año), se identifican dos regiones que se caracterizan, cada una, por la presencia de una mayor frecuencia de puntos (regiones en color rojo).

Lo anterior puede expresarse mediante la construcción de dos matrices de Origen-Destino, una para cada tipo de vuelo. Además, se propone como en el caso de Market, que el número de vuelos entre los aeropuertos i, j es proporcional a los grados de salida de la matriz OD e inversamente proporcional a cierta potencia de la distancia de vuelo entre dichos aeropuertos, es decir, se están considerando modelos tipo-gravedad. De forma matemática, esto se expresa mediante la relación

$$T_{ij}^{modelo} = \begin{cases} T_{ij}^1 = \frac{k_i^{(s)} k_j^{(s)}}{d_{ij}^\gamma} & \text{si } a > \log_{10} (w_{i \rightarrow j}) + \log_{10} (d_{ij}^\gamma), \\ T_{ij}^2 = \frac{k_i^{(s)} k_j^{(s)}}{d_{ij}^\gamma} & \text{si } a \leq \log_{10} (w_{i \rightarrow j}) + \log_{10} (d_{ij}^\gamma). \end{cases} \quad (4.4)$$

El valor a corresponde al término constante de la recta $\log_{10} (w_{i \rightarrow j}^{modelo}) = \gamma \log_{10} (d_{ij}) + a$ con pendiente γ y permite separar los datos de los dos tipos de vuelo mediante las

condiciones establecidas en el modelo (4.4).

Para obtener la matriz de probabilidades de transición $w_{i \rightarrow j}^{modelo}$ que describe ambos tipos de vuelo, se propone que cada tipo de vuelo contribuye a la probabilidad de transición entre dos aeropuertos en un factor que debe determinarse utilizando los datos. Matemáticamente esto se escribe como:

$$w_{i \rightarrow j}^{modelo} = \alpha w_{i \rightarrow j}^{1\ model} + (1 - \alpha) w_{i \rightarrow j}^{2\ model} \quad (4.5)$$

donde $w_{i \rightarrow j}^{\kappa\ model}$ es la matriz de probabilidad de transición del modelo obtenida a partir de la ecuación (4.1) y correspondiente a la matriz OD, T_{ij}^{κ} para $\kappa = 1, 2$. El parámetro α debe tener un valor entre 0 y 1 de tal forma que $w_{i \rightarrow j}^{modelo}$ satisfaga la condición de normalización (4.2).

Por lo tanto, en el modelo para Coupon, las cantidades a , α y γ son los parámetros del modelo. La interpretación de los dos últimos parámetros es como en el caso de Market, sencilla: en primer lugar, γ representa la importancia de la distancia entre dos aeropuertos en la probabilidad de que ocurra un vuelo entre estos. Para una distancia d_{ij} fija, un incremento en el valor de γ implica un decremento en el valor de la probabilidad de transición. En segundo lugar, α cuantifica la contribución de cada tipo de vuelo a la probabilidad de transición entre aeropuertos con una distancia de vuelo d_{ij} . Por otra parte, el parámetro a caracteriza a la recta $\log_{10}(w_{i \rightarrow j}) = \gamma \log_{10}(d_{ij}) + a$ una vez que γ tiene un valor fijo. De esta forma, a permite dividir al conjunto de datos en dos partes que corresponden a cada tipo de vuelo. El valor de a también determina la fracción de vuelos que son de un tipo u otro.

Como en el caso de Market, en la Figura 4.4 se muestra una comparación entre el conjunto de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ generado a partir de la base de datos y el conjunto de puntos obtenido a partir del modelo (4.4), la ecuación (4.1) y el grado de salida de la matriz OD. En la siguiente sección, se discutirá con detalle, los algoritmos para obtener los valores adecuados de los parámetros de los modelos correspondientes a Market y Coupon.

4.4. Estimación de los parámetros del modelo

En la sección anterior se definieron los modelos para describir la movilidad y dinámica en la red de aeropuertos de los Estados Unidos. Cada modelo está caracterizado por un conjunto de parámetros cuya interpretación en términos de las probabilidades de transición y las distancias de vuelo es relativamente simple. Para calcular los valores óptimos de dichos parámetros que logran un mejor ajuste a los datos se define una función error, en este caso, el error cuadrático, que cuantifica la discrepancia entre las probabilidades de transición obtenidas empíricamente y las probabilidades de transición que se obtienen de los modelos propuestos. El error cuadrático S^2 es una función de los parámetros del modelo y se define

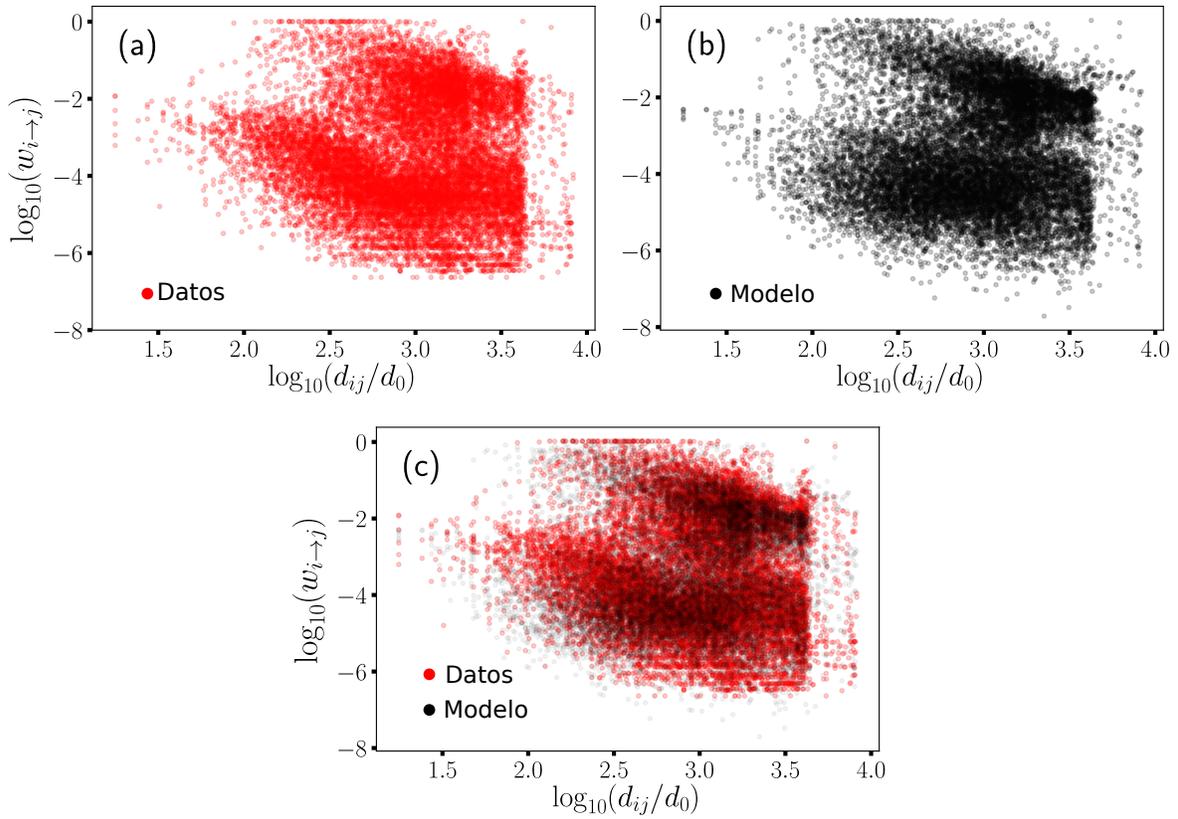


Figura 4.4: Comparación entre datos y modelo del conjunto de puntos $\left\{ \left(\log_{10} \left(\frac{d_{ij}}{d_0} \right), \log_{10} \left(w_{i \rightarrow j}^{(OD)} \right) \right) \right\}$ correspondiente a Coupon. En (a) se muestra el conjunto de puntos obtenido a partir de los datos empíricos. En (b) se muestra el conjunto de puntos obtenido a partir del modelo (4.3). En (c) se muestran, superpuestos, ambos conjuntos de puntos para facilitar su comparación. Los parámetros utilizados para generar la información correspondiente al modelo son presentados en la Tabla 4.1.

como

$$S^2(\theta) = \sum_{i=1}^N \sum_{j=1}^N \left(\log_{10}(w_{i \rightarrow j}) - \log_{10}(w_{i \rightarrow j}^{modelo}) \right)^2 \quad (4.6)$$

donde θ representa el conjunto de parámetros de cada modelo. El objetivo es obtener los valores de los parámetros para los cuales S^2 alcanza un valor mínimo.

Para obtener el conjunto de parámetros óptimo de cada modelo se utilizan los datos correspondientes al año 2018. El modelo obtenido se utilizará posteriormente para simular transiciones entre los diferentes aeropuertos de la red y comparar la distribución de desplazamientos resultante con la distribución empírica obtenida a partir de las matrices de origen-destino \mathbf{T} y de distancias \mathbf{D} de cada uno de los años considerados. De esta forma, se pretende explicar la movilidad y la dinámica de la red de aeropuertos de cada uno de los años considerando el conjunto de parámetros correspondiente al año 2018.

4.4.1. Parámetros del modelo para Market

Primero, se considera el conjunto de datos Market. De acuerdo a la ecuación (4.1), la probabilidad de transición entre los aeropuertos i, j está dada por

$$w_{i \rightarrow j}^{modelo} = \frac{T_{ij}^{modelo}}{\sum_{l=1}^N T_{il}^{modelo}} \quad (4.7)$$

donde T_{ij}^{modelo} está especificado de acuerdo al modelo (4.3). Para obtener los valores óptimos de los parámetros, se construye una cuadrícula de puntos de la forma (γ, R, β) en los intervalos de valores $0.0 < \gamma < 2.0$, $10^{2.5} \text{ km} < R < 10^{3.5} \text{ km}$, $0.5 < \beta < 1.5$. El error cuadrático es evaluado en una cuadrícula de $200 \times 100 \times 100$ puntos y en donde los extremos de dichos intervalos se escogieron con base en la Figura 4.1. La razón por la cual se ha escogido este método de minimización sobre métodos que involucran calcular la derivada del error cuadrático, es meramente por simplicidad, ya que la derivada con respecto a β resulta en una expresión complicada.

Los resultados obtenidos para los parámetros se muestran en la Tabla 4.1 e indican que un 1.4% del total de los vuelos satisfacen que $d_{ij} < R$, es decir, son vuelos locales. El resto de los vuelos son de largo alcance y corresponden al 98.6% del total de vuelos. Esto implica que una gran cantidad de vuelos ocurren entre aeropuertos distantes. Por otra parte, el radio de la vecindad que define a los vuelos locales tiene un valor de $R = 10^{2.5} \text{ km} = 316.2 \text{ km}$, este valor es pequeño en comparación con la distancia de vuelo promedio de cada uno de los años considerados cuyo valor es aproximadamente 2000 km (véase Tabla 3.4).

4.4.2. Parámetros del modelo para Coupon

Para Coupon, se sigue el mismo procedimiento anterior. Se construye una cuadrícula, en este caso, de puntos de la forma (γ, a, α) . De la ecuación (4.4) se puede observar que el valor del parámetro a permite separar los datos en las dos regiones correspondientes a ambos tipos de vuelo. Para calcular este parámetro, se requiere del conocimiento previo de γ , pues diferentes valores de este último dan lugar a diferentes divisiones de los datos. Los intervalos utilizados para γ y α son $0.0 < \gamma < 2.0$ y $0.0 < \alpha < 1.0$. Puesto que el valor de a depende de γ , no se especifica un intervalo. El algoritmo para determinar el valor de a es como sigue: Se establece un valor fijo de γ de tal forma que para cada par de puntos $(w_{i \rightarrow j}^{modelo}, d_{ij})$ se calcula la cantidad $a_{ij} = \log_{10}(w_{i \rightarrow j}) + \log_{10}(d_{ij}^{\gamma})$ donde $w_{i \rightarrow j}$ se calcula a partir de la ecuación (4.1). El valor del parámetro a , se calcula como el promedio de este conjunto de valores $\{a_{ij}\}$.

Con los valores de γ y a establecidos, se calcula la matriz de origen-destino T_{ij}^{modelo} utilizando el modelo (4.4) y a partir de esta se obtiene la matriz de probabilidades de transición $w_{i \rightarrow j}^{modelo}$ definida en la ecuación (4.5) para diferentes valores de α . Los parámetros óptimos del modelo para el conjunto de datos Market, se muestran en la Tabla 4.1 e indican que sólo un 0.6% del total de vuelos corresponden a vuelos del tipo I, mientras que el 99.4% restante corresponde a vuelos del tipo II.

Tabla 4.1: Características de la red de aeropuertos en los Estados Unidos. Se presenta para el año 2018, el número de aeropuertos activos N , el número total de desplazamientos para los conjuntos de datos Market y Coupon, los parámetros 1, 2 y 3 de cada modelo cuyos valores minimizan la función error definida en la relación (4.6) y que fueron obtenidos utilizando los datos del año 2018. También se muestra para Market, la fracción del total de vuelos que son locales y de largo alcance y para Coupon, la fracción de vuelos de tipo I y tipo II. El valor obtenido del parámetro R del modelo asociado a Market corresponde a una vecindad de radio $R = 10^{2.5}$ km = 316 km alrededor de cada aeropuerto y define si un vuelo es local o de largo alcance.

Valores	Market	Coupon
Aeropuertos activos N	345	348
Número total de desplazamientos	56738321	72517675
Parámetro 1	(γ) 1.0	(γ) 1.1
Parámetro 2	(R) $10^{2.5}$ km	(a) - 0.2
Parámetro 3	(β) 1.2	(α) 0.005
Fracción vuelos locales (%)	1.4	–
Fracción vuelos largo alcance (%)	98.6	–
Fracción vuelos tipo I (%)	–	0.6
Fracción vuelos tipo II (%)	–	99.4

4.5. Simulaciones Monte Carlo

Una vez obtenidos los parámetros óptimos de cada modelo, se procede a evaluar su capacidad para reproducir las transiciones entre los diferentes aeropuertos de la red. Para tal fin, se realizan simulaciones numéricas utilizando el método de Monte Carlo. Considerando que cada aeropuerto dentro de la red tiene una importancia diferente que se ha definido en términos de los grados de salida de la matriz OD, se construye para cada aeropuerto i , una probabilidad p_i proporcional a su grado de salida $k_i^{(s)}$ que representa la frecuencia con la cual un vuelo se origina desde dicho aeropuerto. Para cada año y a partir del conjunto de probabilidades $\{p_i\}_{i=1}^N$ donde N representa el número de aeropuertos activos, se generan el mismo número de aeropuertos de origen que el número de vuelos totales en ese año en particular. Posteriormente, para cada ocurrencia de un aeropuerto de origen i , se simula una transición hacia cualquier otro aeropuerto de la red utilizando las probabilidades de transición $\{w_{i \rightarrow j}^{modelo}\}_{j=1}^N$ obtenidas de cada uno de los modelos. Cada transición se registra como un par (i, j) donde i es el aeropuerto de origen y j el aeropuerto de destino. De esta forma, se obtienen el mismo número de vuelos simulados que el número de vuelos en la base de datos original.

Una vez generadas las transiciones por medio de la simulación, se obtiene la distribución de distancias utilizando la matriz de distancias \mathbf{D} . Es decir, para cada distancia de vuelo d_{ij} entre los aeropuertos i, j , se cuenta el número de transiciones entre este par de aeropuertos que resultan de la simulación y de esta forma se construye una distribución de distancias como la que se muestra en la Figura 3.2. Comparando los resultados de esta simulación con la información contenida en los conjuntos de datos Market y Coupon, se puede establecer la utilidad de los modelos para explicar la movilidad de los viajeros y la

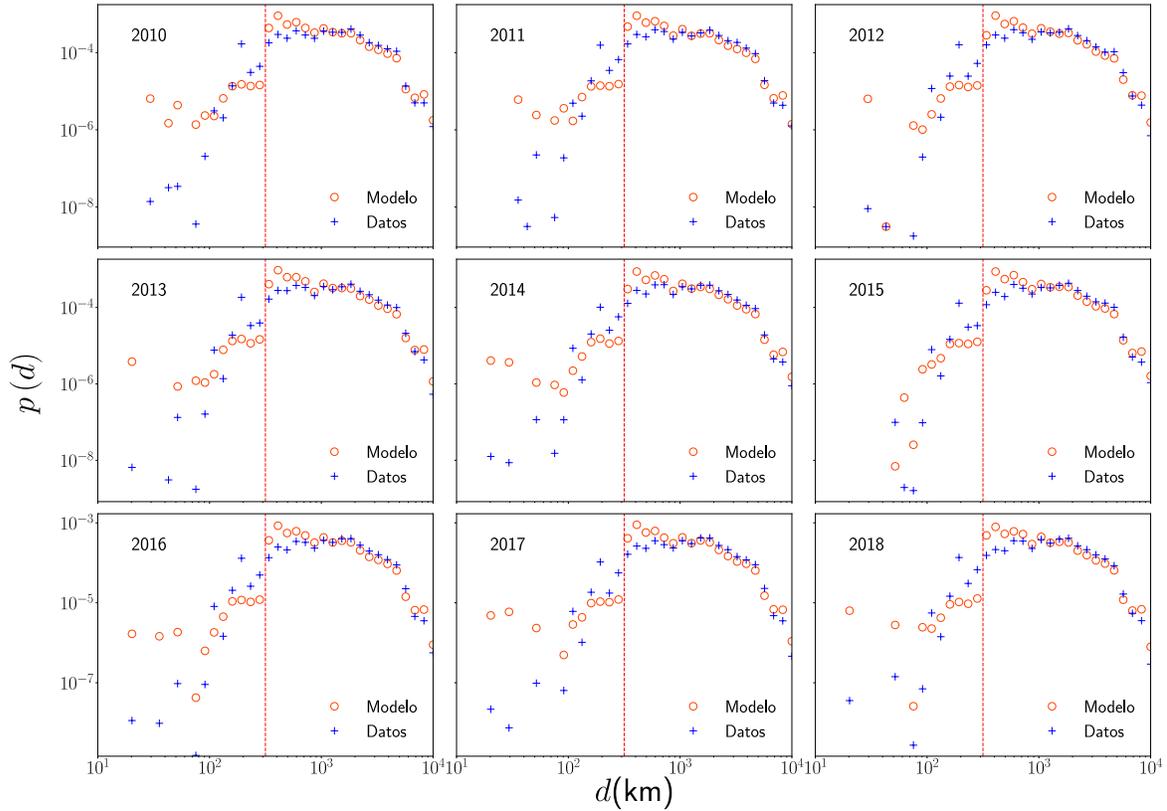


Figura 4.5: Densidad de probabilidad de los desplazamientos en Market. Para cada año seleccionado para el análisis, se muestra la densidad de probabilidad $p(d)$ de la distancia de vuelo recorrida d obtenida a partir de las transiciones entre aeropuertos generadas de forma aleatoria con probabilidad $w_{i \rightarrow j}^{modelo}$ y la correspondiente densidad de probabilidad obtenida a partir de los datos en la matriz OD y la matriz de distancias \mathbf{D} . En cada caso una línea vertical punteada denota el radio R de la localidad la cual define si un vuelo es local o de larga distancia. Ambos ejes coordenados presentan una escala logarítmica.

dinámica espacial de la red de aeropuertos.

En la Figura 4.5 se presenta el análisis estadístico de los desplazamientos para el caso de Market, en esta figura se muestra la distribución de los desplazamientos obtenidos de forma aleatoria por medio de la simulación Monte Carlo y la distribución obtenida de la base de datos original. En la Figura 4.6 se muestra el mismo par de distribuciones pero para el conjunto de datos Coupon.

De la Figura 4.5 se puede observar que el modelo (4.3) describe correctamente la dinámica de los vuelos de larga distancia ($R > 10^{2.5}$ km). En el caso de los vuelos locales, el modelo no captura de forma exacta el comportamiento de los pasajeros; sin embargo, es importante mencionar que si se considera un modelo exclusivamente de tipo-gravedad para el cual $R = 0$, la probabilidad de que ocurran vuelos entre aeropuertos cercanos (con distancias de vuelo menores a 315 km) es sobreestimada. Esto es un indicativo de que en efecto, existen vuelos locales con una dinámica diferente al de los vuelos que ocurren

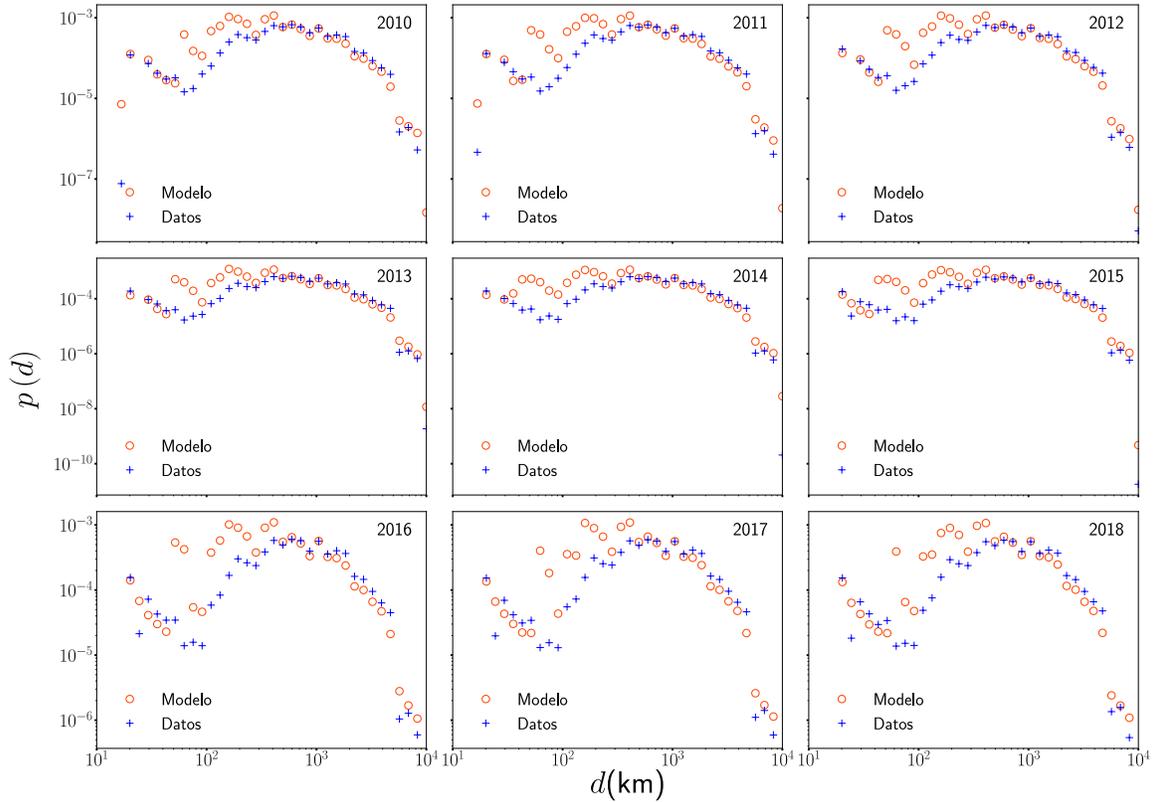


Figura 4.6: Densidad de probabilidad de los desplazamientos en Coupon. Para cada año seleccionado para el análisis, se muestra la densidad de probabilidad $p(d)$ de la distancia de vuelo recorrida d obtenida a partir de las transiciones entre aeropuertos generadas de forma aleatoria con probabilidad $w_{i \rightarrow j}^{modelo}$, esta última obtenida a partir del modelo definido en (4.4) y (4.5) así como la correspondiente densidad de probabilidad obtenida a partir de los datos en la matriz OD y la matriz de distancias \mathbf{D} . Ambos ejes coordenados presentan una escala logarítmica.

entre aeropuertos con una mayor de distancia de vuelo.

En el caso de la Figura 4.6 también se observa que el modelo (4.4) reproduce correctamente la dinámica de los vuelos en gran parte del intervalo de distancias de vuelo d . En particular, se observa que en los primeros años [2010-2015], la discrepancia entre la densidad probabilidad real y aquella obtenida a partir de las simulaciones es menor que para el resto de los años aún cuando los parámetros de modelo se obtuvieron a partir de los datos correspondientes al año 2018.

Es importante notar que para los modelos asociados a los conjuntos de datos Market y Coupon, el conjunto de parámetros óptimo obtenido a partir de los datos del año 2018 generaliza de forma adecuada para el resto de los años. Esto permite concluir que tanto la movilidad como la dinámica espacial en la red de aeropuertos puede explicarse en términos de un par de modelos con parámetros únicos.

Del trabajo realizado en este capítulo se puede concluir que la dinámica de las redes

definidas por los conjuntos de datos Market y Coupon puede describirse adecuadamente mediante modelos de movilidad con un número pequeño de parámetros. Estos parámetros, de fácil interpretación, permiten entender la movilidad de una forma intuitiva en términos de la distancia de vuelo como un factor de disuasión. Es decir, la distancia juega un papel importante en la decisión de un viajero sobre realizar un viaje o no. Por otra parte, se encontró que en ambas redes existen diferentes tipos de vuelos; por ejemplo, en el caso de Market, vuelos locales y vuelos de larga distancia. Esto último resulta importante si el objetivo es diseñar sistemas de transporte más eficientes pues permite considerar los diferentes tipos de pasajeros y sus necesidades.

Conclusiones

En este trabajo de grado se estudió la red de aeropuertos de los Estados Unidos en el contexto de las redes complejas y la movilidad humana. Por su importancia para el transporte de personas y bienes, las redes de aeropuertos alrededor del mundo tienen un impacto considerable en las economías nacionales y en la economía global. Por lo tanto, es importante entender la dinámica que ocurre en estas redes así como los patrones de movimiento que siguen las personas que viajan entre los aeropuertos que las conforman. Estas redes han sido extensamente estudiadas desde el punto de vista de las redes complejas y se ha encontrado que las redes de aeropuertos de varios países como las de China e India, tienen propiedades en común que caracterizan a este sistema de transporte. Sin embargo, el enfoque principal de estos trabajos es el de las redes complejas, con particular énfasis en el estudio de la matriz de adyacencia que ofrece información únicamente acerca de la topología de la red y no de la dinámica que ocurre en términos de los pasajeros de la red. Estudiar la movilidad de los pasajeros en la red de aeropuertos de los Estados Unidos fué el objetivo principal de este trabajo.

Se analizó la red aeroportuaria de Estados Unidos desde el punto de vista de las redes complejas, pero también, desde el punto de vista de la movilidad humana y de diferentes modelos que se han propuesto a lo largo de los años para estudiar los patrones de movimiento de las personas en diferentes escalas de espacio y tiempo; por ejemplo, el modelo de tipo gravedad que ha sido utilizado exitosamente para explicar los flujos de personas en diversos contextos. La simplicidad de este modelo ha permitido estudiar varios sistemas de transporte a diferentes escalas.

También se analizaron los movimientos de alrededor de 360 millones de pasajeros en la red de aeropuertos repartidos en el periodo 2010-2019. Algunas propiedades observadas en la red mundial de aeropuertos como la propiedad de mundo pequeño o la caracterización de los grados de salida por medio de leyes de potencias se verifican también para la red de aeropuertos de Estados Unidos lo cual implica que las propiedades de este sistema de transporte permanecen aproximadamente constantes cuando se cambia la escala espacial en la que se analiza el sistema. Primeramente, se construyeron las matrices de origen-destino que contienen la información completa de la red de aeropuertos y representan la herramienta principal de este trabajo. Posteriormente, se ajustaron distribuciones de cola pesada a los grados de entrada y salida de la matriz de origen-destino utilizando la metodología introducida en [50] y se encontró que los grados son mejor aproximados por

la distribución log-normal. Esto último sugiere que los grados siguen, en efecto, una ley de potencias en una parte del intervalo, como puede observarse de la ecuación (3.9). El truncamiento de la distribución para valores grandes de los grados de la matriz origen-destino se explica en términos de la capacidad limitada de los aeropuertos para manejar el tráfico aéreo. De forma similar, se analizaron las propiedades estadísticas del rango de los grados de la matriz de origen-destino mediante la distribución beta generalizada discreta, utilizada ampliamente para estudiar fenómenos relacionados con sistemas complejos.

Por otra parte, el análisis estadístico de cantidades asociadas a una red compleja como la longitud de camino mínimo promedio y la centralidad de intermediación se estudió la topología de esta red de aeropuertos y se encontraron resultados relevantes acerca de su organización en términos de comunidades. En particular, se encontró que el número promedio de vuelos $\langle l \rangle$ que se necesitan para conectar cualesquiera dos aeropuertos de la red es igual a 2; resultado que puede interpretarse como una alta eficiencia de la red de aeropuertos en Estados Unidos para transportar a los pasajeros. Utilizando la matriz de origen-destino, se aplicó el algoritmo de Louvain para la detección de comunidades en redes pesadas y se encontró para las dos redes construidas, Market y Coupon, cuatro comunidades que separan el territorio estadounidense en cuatro áreas geográficas bien definidas. Sin embargo, dentro de estas comunidades existen aeropuertos que permiten conectar partes distantes del territorio lo cual permite que un pasajero viaje largas distancias en un número pequeño de vuelos.

Para el modelamiento de la movilidad, se propusieron dos modelos para explicar el movimiento de los pasajeros en dos redes distintas asociadas con este sistema de aeropuertos: una red para estudiar la movilidad de los pasajeros y la red donde se lleva a cabo dicha movilidad. Se definieron las matrices de probabilidad de transición (a partir de las matrices de origen-destino) que ofrecen información acerca de la probabilidad con la que ocurre un viaje entre cualesquiera dos aeropuertos. Los modelos de movilidad se construyeron con base en el modelo de tipo gravedad donde los grados de salida de la matriz origen-destino cumplen la función de “masas”, es decir, los grados miden la importancia de cada aeropuerto dentro de la red. Los parámetros asociados con cada modelo son de fácil interpretación y permiten cuantificar el impacto de los grados de salida y de la distancia de vuelo en la probabilidad de que un pasajero realice un viaje entre dos aeropuertos.

Finalmente, se aplicaron simulaciones Monte Carlo para verificar la utilidad de estos modelos para generar el comportamiento observado de los pasajeros. Utilizando el grado de salida como una medida proporcional a la probabilidad con la que un viaje comienza en un aeropuerto particular, se generaron el mismo número de vuelos que el reportado en las bases de datos por medio de las matrices de probabilidad de transición. Con los resultados obtenidos en la simulación se encontró que, en efecto, los modelos propuestos reproducen de forma aproximada la dinámica observada en la red. En este último sentido, el trabajo de esta tesis es completamente nuevo, ya que no existe literatura que aborde el tema de las redes de aeropuertos desde el punto de vista de los modelos de movilidad. De esta forma, este trabajo es relevante en un mundo globalizado en el que el sistema de transporte aéreo juega un papel protagónico.

Trabajo futuro

En este trabajo se propusieron modelos de movilidad para explicar los patrones de movimiento de los pasajeros en la red de aeropuertos de Estados Unidos. Las variables relevantes de estos modelos son el número de pasajeros que viajan entre cualquier par de aeropuertos y las distancias de vuelo recorridas. Sin embargo, otras variables importantes para entender los movimientos de los pasajeros en la red de aeropuertos, y en general cualquier sistema de transporte, son variables de tipo económicas o temporales; por ejemplo, el costo de viaje puede estar determinado por diversos factores como la localización de las ciudades donde se localizan los aeropuertos de origen y destino o la temporada del año en que se realiza el viaje. Otro enfoque interesante para explorar este tipo de sistemas es el del aprendizaje automático o *machine learning* que permite identificar patrones de movimiento que son difíciles de identificar y por lo tanto, generar modelos de movilidad que puedan caracterizar mejor la dinámica de los pasajeros a diferentes escalas espaciales y temporales.

Bibliografía

- [1] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.
- [2] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R. James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J. Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, Mar 2018.
- [3] 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):4022, mar 2020.
- [4] Juan de Dios Ortúzar and Luis G Willumsen. *Modelling transport*. John Wiley & Sons, 2011.
- [5] Samuel A. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.
- [6] George Kingsley Zipf. The $p_1 p_2/d$ hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.
- [7] E. G. Ravenstein. The laws of migration. *Journal of the Statistical Society of London*, 48(2):167–235, 1885.
- [8] Maxime Lenormand, Aleix Bassolas, and José J. Ramasco. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography*, 51:158–169, Feb 2016.
- [9] Marc Barthelemy. *The Structure and Dynamics of Cities: Urban Data Analysis and Theoretical Modeling*. Cambridge University Press, 2016.
- [10] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.
- [11] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1-3):1–101, Feb 2011.
- [12] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, Feb 2010.

-
- [13] Filippo Simini, Marta C. González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, April 2012.
- [14] Kevin E Heanue and Clyde E Pyers. A comparative evaluation of trip distribution procedures. In *Origin and destination: Methods and Evaluation*, pages 20–50. Highway Research Board, 1966.
- [15] D. Loaiza-Monsalve and A. P. Riascos. Human mobility in bike-sharing systems: Structure of local and non-local dynamics. *PLOS ONE*, 14(3):e0213106, 03 2019.
- [16] Woo-Sung Jung, Fengzhong Wang, and H. Eugene Stanley. Gravity model in the korean highway. *EPL (Europhysics Letters)*, 81(4):48005, Jan 2008.
- [17] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [18] Morton Schneider. Gravity models and trip distribution theory. *Papers in Regional Science*, 5(1):51–56, 1959.
- [19] Henk Elffers, Danielle Reynald, Margit Averdijk, Wim Bernasco, and Richard Block. Modelling crime flow between neighbourhoods in terms of distance and of intervening opportunities. *Crime Prevention and Community Safety: An International Journal*, 10:85–96, 04 2008.
- [20] S. Akwawua and J.A. Pooler. An intervening opportunities model of u.s. interstate migration flows. *Geography Research Forum*, 20:33–51, 01 2000.
- [21] Maria Stefanouli and Serafeim Polyzos. Gravity vs radiation model: two approaches on commuting in greece. *Transportation Research Procedia*, 24:65–72, 12 2017.
- [22] World bank. Air transport, passengers carried. <https://data.worldbank.org/indicator/IS.AIR.PSGR?end=2018&start=1970>.
- [23] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004.
- [24] Massimiliano Zanin and Fabrizio Lillo. Modelling the air transport with complex networks: A short review. *The European Physical Journal Special Topics*, 215(1):5–21, Jan 2013.
- [25] Wikimedia Commons. World-airline-routemap-2009, 2009.
- [26] Kai-Quan Cai, Jun Zhang, Wen-Bo Du, and Xian-Bin Cao. Analysis of the chinese air route network as a complex network. *Chinese Physics B*, 21(2):028903, feb 2012.

-
- [27] Nikolas Pyrgiotis, Kerry Malone, and Amedeo Odoni. Modelling delay propagation within an airport network. *Transportation Research Part C: Emerging Technologies*, 27:–, 01 2011.
- [28] R. Guimera, S. Mossa, A. Turttschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, May 2005.
- [29] W. Li and X. Cai. Statistical analysis of airport network of china. *Physical Review E*, 69:046106, Apr 2004.
- [30] Ganesh Bagler. Analysis of the airport network of india as a complex weighted network. *Physica A: Statistical Mechanics and its Applications*, 387:2972–2980, 05 2008.
- [31] Zengwang Xu and Robert Harriss. Exploring the structure of the u.s. intercity passenger air transportation network: A weighted complex network approach. *GeoJournal*, 73:87–102, 10 2008.
- [32] Roger Guimerà and Luís Amaral. Modeling the world-wide airport network. *The European Physical Journal B*, 387540:381–385, 03 2004.
- [33] Wikimedia Commons. Atlanta hartsfield-jackson, 2010.
- [34] Wikimedia Commons. Los angeles international airport aerial photo, 2014.
- [35] Michael Stumpf, Carsten Wiuf, and Robert May. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102:4221–4, 04 2005.
- [36] Bureau of transportation statistics. <https://www.bts.gov/>.
- [37] Descripción de la base de datos db1b. https://www.transtats.bts.gov/tables.asp?DB_ID=125&DB_Name=&DB_Short_Name=.
- [38] Marius Agasse-Duval and Steve Lawford. Subgraphs and Motifs in a Dynamic Airline Network. Working Papers hal-02017122, HAL, February 2019.
- [39] Martin Rosvall, Alcides V. Esquivel, Andrea Lancichinetti, Jevin D. West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5(1):4630, August 2014.
- [40] Hector Rodriguez-Deniz, M. Villani, and Augusto Voltes-Dorta. A multilayered block network model to forecast large dynamic transportation graphs, with an application to us air transport. *arXiv: Machine Learning*, 2019.
- [41] Pablo Fleurquin, José J. Ramasco, and Victor M. Eguiluz. Systemic delay propagation in the us airport network. *Scientific Reports*, 3(1), Jan 2013.

-
- [42] Steve Lawford and Yll Mehmeti. Cliques and a new measure of clustering: With application to U.S. domestic airlines. *Physica A: Statistical Mechanics and its Applications*, 560:125158, 2020.
- [43] Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio J. Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature Communications*, 5(1), Sep 2014.
- [44] Martin Rosvall and Carl T. Bergstrom. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *Plos One*, 6(4):e18209, Apr 2011.
- [45] M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5), Nov 2004.
- [46] 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.
- [47] V Nicosia, G Mangioni, V Carchiolo, and M Malgeri. Extending the definition of modularity to directed graphs with overlapping communities. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03024, Mar 2009.
- [48] Fragkiskos D. Malliaros and Michalis Vazirgiannis. Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4):95–142, Dec 2013.
- [49] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, Sep 2000.
- [50] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, Nov 2009.
- [51] P. Sargant Florence. Human behavior and the principle of least effort. *The Economic Journal*, 60(240):808–810, 1950.
- [52] Gustavo Martínez-Mekler, Roberto Alvarez Martínez, Manuel Beltrán del Río, Ricardo Mansilla, Pedro Miramontes, and Germinal Cocho. Universality of rank-ordering distributions in the arts and sciences. *Plos One*, 4(3):e4791, 2009.
- [53] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.