



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

ESTRUCTURA ESPACIAL URBANA DE MOVILIDAD DESDE DATOS MASIVOS DE  
TRANSPORTE PÚBLICO EN SANTIAGO DE CHILE

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS, MENCIÓN  
COMPUTACIÓN

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN

FELIPE ANDRÉS HERNÁNDEZ GODOY

PROFESORES GUÍA:  
BENJAMÍN BUSTOS CÁRDENAS  
MARCELA MUNIZAGA MUÑOZ

MIEMBROS DE LA COMISIÓN:  
BARBARA POBLETE LABRA  
NELSON BALOIAN TATARYAN  
RICARDO HURTUBIA GONZÁLEZ

SANTIAGO DE CHILE

2017

# RESUMEN

La estructura espacial urbana se refiere a la disposición del espacio en la ciudad producto de su forma actual junto con las relaciones subyacentes entre estas. Estas interacciones son generadas por el movimiento de personas, mercancías o información entre un origen y destino, y enmarcadas en un concepto de ciudad entendido como una colección de componentes interrelacionados donde se destacan: el sistema de actividades, sistema de transporte y las relaciones que se generan entre ellos.

En este trabajo se intenta caracterizar la estructura espacial para Santiago de Chile mediante tres indicadores: centros (zonas de la ciudad con la capacidad de atraer o concentrar personas), centros de pasada (zonas de la ciudad que sirven para conectar un par de zonas, funcionan como puentes espaciales) y comunidades (zonas de la ciudad que presentan un fuerte nivel de interacción interno). La metodología utiliza análisis de redes y análisis espacial sobre datos de tarjetas inteligentes de pago de transporte público.

Los datos son obtenidos a partir de las validaciones hechas con tarjetas inteligentes generadas en parte del sistema de transporte público de la ciudad de Santiago (Transantiago) entre el 14 al 21 de abril de 2013 (una semana). Con las validaciones y un proceso de estimación de bajada es posible estimar etapas y destino en cerca del 80 % de los viajes.

Con el origen, etapas y destino de cada viaje se construye una red dirigida con pesos en donde se relaciona cada indicador con una métrica de la red. Los centros son asociados al *PageRank*, los centros de pasada al *Betweenness* y las comunidades a *Infomap*.

Los resultados muestran que la ciudad continúa manteniendo su estructura con un distrito comercial central (CBD por sus siglas en inglés) en el sector centro-oriente, los centros de pasada son fuertemente influenciados por la red de metro y las comunidades presentan una fuerte unión espacial con una expresión de todas ellas en el CBD, mostrando que el centro de la ciudad es un territorio que les pertenece a todos. Lo anterior muestra una estructura de la ciudad más cercana al monocentrismo al compararla con Singapur.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Hipótesis . . . . .	3
1.2. Objetivos . . . . .	3
1.2.1. Objetivo general . . . . .	3
1.2.2. Objetivos específicos . . . . .	3
1.3. Alcance . . . . .	4
1.4. Importancia . . . . .	4
1.5. Estructura de la tesis . . . . .	5
<b>2. Marco Teórico</b>	<b>6</b>
2.1. Estructura Espacial Urbana . . . . .	6
2.2. Caracterización de redes complejas . . . . .	8
2.2.1. Grado . . . . .	9
2.2.2. Promedio de rutas más cortas . . . . .	9
2.2.3. Coeficiente de agrupamiento . . . . .	10
2.2.4. Componentes conexas . . . . .	11
2.2.5. Diámetro . . . . .	11
2.2.6. <i>PageRank</i> . . . . .	11
2.2.7. <i>Betweenness</i> . . . . .	12
2.2.8. Estructura de comunidad: <i>Infomap</i> . . . . .	13
2.3. Ciudades desde redes complejas . . . . .	14
<b>3. Datos</b>	<b>17</b>
3.1. Datos de estudio: viajes . . . . .	18
3.1.1. Análisis estadístico . . . . .	19
3.2. Datos referenciales . . . . .	22
3.2.1. Santiago de Chile . . . . .	22
3.2.2. Sistema de transporte . . . . .	23
3.2.3. Paraderos de Transantiago . . . . .	26
3.2.4. Caracterización espacial de la ciudad . . . . .	26
<b>4. Implementación</b>	<b>32</b>
4.1. Herramientas . . . . .	32
4.2. Procesamiento de datos . . . . .	33
4.2.1. Viajes . . . . .	33
4.2.2. Paradas . . . . .	33

4.2.3.	Geografía . . . . .	34
4.2.4.	Filtros . . . . .	35
4.2.5.	Normalización . . . . .	36
4.3.	Construcción de la red de movilidad . . . . .	37
4.4.	Cálculo de indicadores . . . . .	39
4.4.1.	PageRank . . . . .	39
4.4.2.	Betweenness . . . . .	40
4.4.3.	Infomap . . . . .	41
4.4.4.	Propiedades grafo . . . . .	42
4.4.4.1.	Grado de nodo . . . . .	42
4.4.4.2.	Intensidad ( <i>Strength</i> ) . . . . .	43
4.4.4.3.	Promedio de rutas más cortas ( <i>Shortest path length</i> ) . . . . .	43
4.4.4.4.	Coefficiente de agrupamiento ( <i>Clustering coefficient</i> ) . . . . .	44
4.4.4.5.	Diámetro . . . . .	44
4.4.4.6.	Componentes conexas . . . . .	44
4.5.	Asociación espacial . . . . .	45
4.5.1.	Interpolación espacial . . . . .	45
4.5.2.	Asignación de comunidad a zonificación . . . . .	46
4.6.	Visualización de <i>PageRank</i> y <i>Betweenness</i> . . . . .	47
<b>5.</b>	<b>Análisis</b> . . . . .	<b>49</b>
5.1.	Análisis de la red . . . . .	50
5.1.1.	Propiedades generales . . . . .	51
5.2.	Análisis espacial . . . . .	54
5.2.1.	Centros . . . . .	54
5.2.2.	Centros de pasada . . . . .	58
5.2.3.	Comunidades . . . . .	60
5.2.3.1.	Comunidades detectadas . . . . .	61
5.2.3.2.	Asociación espacial con la zonificación 777 . . . . .	65
5.2.3.3.	Comparación con Singapur . . . . .	66
5.2.3.4.	Estructuración de Transantiago el 2012 . . . . .	67
<b>6.</b>	<b>Conclusiones y trabajo futuro</b> . . . . .	<b>69</b>
6.1.	Trabajo futuro . . . . .	71
<b>7.</b>	<b>Anexos</b> . . . . .	<b>73</b>
A.	Descripción tabla de viajes . . . . .	73
B.	Descripción tabla de paradas y estaciones de metro . . . . .	76
C.	Conversión de archivo CSV a PAJEK . . . . .	77
D.	Generación de CSV de viajes . . . . .	78
	<b>Bibliografía</b> . . . . .	<b>80</b>

# Índice de Tablas

3.1. Datos utilizados. Fuente: elaboración propia . . . . .	17
3.2. Cantidad de viajes según su cantidad de etapas. Fuente: elaboración propia .	20
3.3. Antecedentes demográficos por sector separado por IFM (Ingreso Familiar Mensual). Fuente: EOD-2012 . . . . .	22
3.4. Distribución de los viajes según modo de transporte y propósito en día laboral temporada normal. Fuente: eod-2012 . . . . .	24
3.5. Caracterización de las líneas de metro. Fuente: www.metro.cl . . . . .	25
3.6. Comunas de la región metropolitana agrupadas por sector y provincia. Fuente: EOD-2012 . . . . .	30
4.1. Columnas utilizadas de la tabla de viajes. Fuente: elaboración propia . . . .	34
4.2. Columnas utilizadas de la tabla de redparadas. Fuente: elaboración propia .	34
4.3. Columnas utilizadas de la tabla de estaciones_metro. Fuente: elaboración propia	35
4.4. Atributos de archivo <i>shape</i> con la zonificación 777. Fuente: elaboración propia	35
4.5. Viajes iniciales y viajes eliminados en el filtro. Fuente: elaboración propia . .	36
4.6. Parámetros para generar visualización de <i>PageRank</i> y <i>Betweenness</i> . Fuente: elaboración propia . . . . .	47
5.1. Grafos construidos y las métricas a aplicar sobre cada uno. Fuente: elaboración propia . . . . .	51
5.2. Propiedades globales. Fuente: elaboración propia . . . . .	51
5.3. Comunidades generadas con la cantidad de nodos asignada por <i>Infomap</i> . . .	62
7.1. Columnas presentes en la tabla viajes . . . . .	73
7.2. Columnas presentes en la tabla estaciones_metro . . . . .	76
7.3. Columnas presentes en la tabla redparadas . . . . .	77

# Índice de Ilustraciones

2.1.	Ejemplo de coeficiente de agrupamiento. Fuente: wikipedia.org . . . . .	10
2.2.	<i>PageRank</i> en una red sin pesos. Fuente: Wikipedia.org . . . . .	12
2.3.	Ilustración del método ICDR. Fuente: Louail et al. (2015) . . . . .	16
3.1.	Cantidad de viajes por día. Fuente: elaboración propia . . . . .	20
3.2.	Cantidad de viajes por hora. Fuente: elaboración propia . . . . .	21
3.3.	Cantidad de viajes por comuna. Fuente: elaboración propia . . . . .	21
3.4.	líneas de metro. Fuente: elaboración propia . . . . .	25
3.5.	Estaciones de metro agrupadas por comuna de ubicación. Fuente: elaboración propia . . . . .	26
3.6.	Ubicación de paraderos de Transantiago. Fuente: elaboración propia . . . . .	27
3.7.	Cantidad de paraderos por comuna de Santiago. Fuente: elaboración propia .	28
3.8.	Sectores (borde rojo) y comunas (borde negro) de la ciudad de Santiago. Fuente: EOD-2012 . . . . .	29
3.9.	zonificación de Santiago, conocida como zonificación 777. Fuente: EOD-2012	31
4.1.	Proceso de construcción de indicadores. Fuente: elaboración propia . . . . .	32
4.2.	Proceso de filtrado y normalización de datos. Fuente: elaboración propia . .	37
4.3.	Red descrita usando el formato Pajek. Fuente: elaboración propia . . . . .	38
4.4.	Proceso de transformación desde la base de datos a formato pajek. Fuente: elaboración propia . . . . .	39
4.5.	Ejemplo de formato de archivo con extensión .tree. Fuente: elaboración propia	42
4.6.	Ventana de configuración para la interpolación en QGIS. Fuente: elaboración propia . . . . .	46
4.7.	Configuración de la capa que contiene el resultado de la interpolación. Fuente: elaboración propia . . . . .	48
5.1.	Fuentes de datos para el análisis. Fuente: elaboración propia . . . . .	49
5.2.	Distribución de probabilidad acumulada para grafos de etapas. Fuente: elaboración propia . . . . .	53
5.3.	Distribución de probabilidad acumulada para grafos de viaje. Fuente: elaboración propia . . . . .	54
5.4.	Grado vs strength en 2 variantes: entrada y salida sobre los grafos de mañana y tarde de viajes y etapas. Fuente: elaboración propia . . . . .	55
5.5.	Identificación de centros (zonas de color rojo). Corresponde al <i>PageRank</i> interpolado. Fuente: elaboración propia . . . . .	55

5.6.	Distribución de <i>PageRank</i> en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia . . . . .	56
5.7.	Visualización de los grados de entrada de un nodo. Fuente: elaboración propia	57
5.8.	Visualización del flujo de entrada de un nodo. Fuente: elaboración propia . .	57
5.9.	<i>Betweenness</i> con grafos de viaje en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia . . . . .	58
5.10.	Distribución de probabilidad acumulada de <i>Betweenness</i> con grafos de viaje en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia	59
5.11.	<i>Betweenness</i> usando datos de transbordo. Fuente: elaboración propia . . . .	59
5.12.	Distribución de probabilidad acumulada de <i>Betweenness</i> con grafos de etapa en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia	60
5.13.	<i>Real Betweenness</i> (usando datos de transbordo). Fuente: elaboración propia .	61
5.14.	Comunidad de Metro. (a) distribución espacial (b) distribución espacial con la proyección de las líneas de metro . Fuente: elaboración propia . . . . .	63
5.15.	Comunidades 12,13 y 14. Fuente: elaboración propia . . . . .	63
5.16.	Comunidades más significativas con los macro sectores de la ciudad (líneas negras). Fuente: elaboración propia . . . . .	64
5.17.	Asignación de comunidades a zonificación 777 usando las 41 comunidades encontradas con <i>Infomap</i> . Fuente: elaboración propia . . . . .	65
5.18.	Asignación de comunidades a zonificación 777 usando las 10 comunidades que dan forma a la ciudad. Fuente: elaboración propia . . . . .	66
5.19.	Comunidades presentes en Singapur con datos del 2012. Fuente: Zhong et al. (2014) . . . . .	66
5.20.	División operacional de Transantiago en el año 2012. Fuente: elaboración propia	67
5.21.	Comunidades más significativas contrastadas con la zonificación de Transantiago del año 2012. Fuente: elaboración propia . . . . .	68
7.1.	Algoritmo para transformar archivo con formato CSV a formato PAJEK. Fuente: elaboración propia . . . . .	78
7.2.	código que genera archivo CSV intermedio para la construcción del grafo en formato Pajek usando los datos de viaje. Fuente: elaboración propia . . . . .	79

# Capítulo 1

## Introducción

Muchas ciudades ya han alcanzado un gran tamaño y organismos internacionales como el departamento de asuntos económicos y sociales de la ONU (UNDESA por sus siglas en inglés) estiman que su crecimiento no se detendrá, lo que generará urbes cada vez más complejas al originar nuevas áreas de actividad e incrementar su volumen de viajes, permitiendo mayor variedad de estilos de vida (mayor número de combinaciones: hogar-trabajo-ocio).

Cualquier solución al problema descrito debe plantearse desde la base de un diagnóstico previo, donde uno de los aspectos es la estructura espacial urbana junto con los patrones de movilidad de las personas que viven o estudian en ella.

La estructura espacial urbana se refiere a la disposición del espacio en una ciudad producto de su forma urbana presente junto con las relaciones subyacentes entre éstas. Estas interacciones espaciales son generadas por el movimiento de personas, mercancías o información entre un origen y un destino (Rodrigue et al., 2013). Las motivaciones de estos flujos involucran una gran variedad de propósitos tales como trabajo, uso de servicios públicos o privados, transmisión de información o capital, actividad económica, etc.

Anas et al. (1998) plantean que la estructura espacial de las ciudades modernas es determinada en gran medida por los avances en transporte y telecomunicaciones. La complejidad de los movimientos humanos ha ido redefiniendo el uso del espacio urbano y la disposición de los recursos. Según lo anterior, las personas pueden ser vistas como transportadores físicos entre zonas de la ciudad y, en consecuencia, ser usados como representantes de interacciones espaciales para la identificación de elementos esenciales (espacios urbanos) de la estructura espacial urbana. Los espacios urbanos identificados en este trabajo corresponden a los definidos por Zhong et al. (2014):

- **Centros de pasada:** zonas que sirven para conectar lugares de la ciudad para transferencia de personas. Funcionan como puentes espaciales entre distintas áreas urbanas.
- **Centros:** zonas que concentran personas o actividades. Pueden diferir de los Centros de pasada, pero a menudo, coinciden.
- **Comunidades:** áreas compactas de la ciudad que presentan un fuerte nivel de interacción en su interior y a su vez, en menor medida con el resto de la ciudad. Este conjunto

de áreas particiona la ciudad en pequeños barrios llamados comunidades.

Por otro lado, los avances actuales en tecnologías de la información y redes sociales (e.g. GPS, 3G/4G, RFID, Twitter, Foursquare, smartphones, smartcards) han puesto a disposición de la comunidad fuentes de datos masivos con un nivel de detalle y profundidad a los que no se tenía acceso previamente y que han contribuido a conocer diversos aspectos de las ciudades, entre ellas, su estructura espacial subyacente (Louail et al., 2014). Una de estas fuentes de datos es la que se genera a partir del uso de tarjetas inteligentes para el pago del transporte público, sistema conocido como AFC por sus siglas en inglés (Automated Fare Collection). Este tipo de sistemas es utilizado en el sistema de transporte público en varias ciudades del mundo, y los usos que se han hecho de esa información son diversos (Pelletier et al., 2011).

La captura de datos de transporte por medio de tarjetas inteligentes ofrece la posibilidad de analizar el comportamiento de los usuarios por medio del conocimiento de sus patrones de desplazamiento. En particular, la estructura espacial de la ciudad puede ser analizada a partir de su estructura de viajes, tomando estos datos como indicador de la interacción espacial urbana. En el caso de Santiago de Chile, tomando los datos del sistema de monitoreo vehicular automatizado (AVL por sus siglas en inglés) basado en GPS instalados en los buses y transacciones de pago del sistema de transporte público, se ha construido matrices origen-destino de viajes de paradero a paradero. Cabe señalar que los viajes en transporte público representan casi el 50 % de los viajes motorizados en la ciudad (Munizaga and Palma, 2012; Munizaga et al., 2014).

El objetivo de este trabajo es identificar la estructura espacial de la ciudad de Santiago de Chile mediante una adaptación de la metodología propuesta por Zhong et al. (2014), quien hace uso de técnicas de análisis de redes y análisis espacial sobre la red de viajes (paradero a paradero) del sistema de transporte público de Singapur a través de los datos de sus tarjetas inteligentes.

El análisis de la estructura espacial urbana es realizado con tres métricas: el *PageRank*, que da información de la importancia de cada zona de la ciudad para atraer flujos (viajes), y por lo tanto se asocia con el concepto de espacio urbano “Centro”. La segunda métrica corresponde al *betweenness*, que en el ámbito de grafos sirve para identificar aquellos nodos que son relevantes para transmitir información, para este trabajo, la información corresponde a viajes realizados, y por lo tanto, su valor es asociado al concepto de espacio urbano de “Centro de pasada”. La tercera métrica es la detección de comunidades, donde en grafos tiene el objetivo de agrupar nodos por diversos criterios (Newman, 2003), para este caso el criterio corresponde al fuerte flujo de viajes que se produce entre las zonas de la ciudad, por lo que se usa el algoritmo *Infomap* y se asocia con la definición de espacio urbano “Comunidades”. Por otro lado, el análisis espacial corresponde a la utilización de la topología de la red dentro del estudio, ya que la red es una red espacial.

En esta tesis se utiliza la matriz origen-destino de viajes en transporte público para construir un grafo dirigido con pesos que luego, es usado para identificar los espacios urbanos definidos anteriormente por medio de métricas usadas en grafos. En el caso de Santiago, no se ha analizado la estructura urbana con otro tipo de datos desagregados, por lo cual su estructura es desconocida hasta el momento. También se busca contrastar los indicadores

evaluando otras métricas de grafos.

## 1.1. Hipótesis

La forma de una ciudad está influenciada por componentes históricos principalmente basados en avances en transporte y tecnología (Anas et al., 1998). Esta forma urbana complementada con las interacciones subyacentes definen la estructura espacial de la ciudad (Rodrigue et al., 2013). A raíz de lo anterior se originan la siguiente pregunta de investigación: ¿las fuentes de datos de transporte público actuales permiten identificar la estructura espacial de una ciudad?

Este trabajo contempla verificar si las métricas de centralidad en redes y el algoritmo de generación de comunidades elegidos son capaces de identificar la estructura espacial de la ciudad de Santiago de Chile. Específicamente se plantean las siguientes hipótesis:

1. El *PageRank* es capaz de identificar zonas que concentran personas.
2. El *Betweenness* es capaz de identificar zonas de transbordo.
3. Las comunidades se condicen con la división usada en análisis territoriales y de movilidad actuales. Un ejemplo de esto último es ver si las comunidades están dentro de los límites zonales o comunales de la ciudad.

## 1.2. Objetivos

A continuación se describen el objetivo general de este trabajo de tesis y sus respectivos objetivos específicos.

### 1.2.1. Objetivo general

Identificar la estructura espacial urbana de la ciudad de Santiago de Chile utilizando datos masivos de transporte público para entender el estado actual (desde el punto de vista de la movilidad) de la ciudad.

### 1.2.2. Objetivos específicos

- Construir un modelo de red para la ciudad de Santiago de Chile utilizando los viajes registrados con las tarjetas inteligentes de transporte público durante una semana.
- Analizar la estructura de la red mediante el cálculo de métricas en el grafo (grado, componentes conexas, promedio de rutas más cortas, intensidad, coeficiente de agrupamiento).

- Calcular centros, centros de pasada y comunidades usando las métricas de red *PageRank*, *Betweenness* y *Infomap* respectivamente.
- Visualizar espacialmente centros, centros de pasada y comunidades mediante herramienta GIS.

### 1.3. Alcance

Esta tesis desarrolla el problema desde un alcance de tipo descriptivo y exploratorio al analizar la estructura espacial de la ciudad de Santiago de Chile a través de los datos de transporte público por medio de la metodología propuesta por Zhong et al. (2014), la que genera métricas cuantitativas para la detección de zonas de aglomeración de personas, zonas utilizadas como intermediación en un viaje y comunidades.

El alcance exploratorio se justifica en que este tipo de trabajo (utilización de datos de transporte público para evaluar la estructura espacial de una ciudad) no ha sido realizado anteriormente en la ciudad de Santiago de Chile, por lo que se pretende generar una base para futuros trabajos relacionados en el área de movilidad. Desde el punto de vista descriptivo se busca evaluar los resultados con estudios que han usado otras metodologías para la identificación de la estructura espacial.

El enfoque a tratar en esta tesis es de carácter mixto, es decir, contiene tanto análisis cuantitativo como cualitativo. El primero está dado por los indicadores calculados en este trabajo, mientras que el segundo corresponde a la incorporación del contexto de la ciudad, como lo es la encuesta origen-destino de viajes realizada el año 2012 (Muñoz and Contreras, 2015).

### 1.4. Importancia

La importancia de este trabajo radica en los siguientes aspectos:

- El uso de *PageRank* para la identificación de zonas que concentran personas a partir de datos de movilidad en una ciudad es reciente y ha sido usado anteriormente en la ciudad de Singapur, en consecuencia, la aplicación y evaluación (positiva o negativa) de este en una ciudad distinta como Santiago es un aporte al área.
- El análisis de *PageRank*, *Betweenness* e *Infomap* permitirá conocer cuantitativamente el tipo de ciudad que es Santiago de Chile.
- La metodología construida en este trabajo puede ser usada para evaluar el efecto de políticas públicas en la movilidad de los habitantes (comparando dos o más cortes temporales).
- Puede contribuir a identificar posibles zonas de desarrollo local, permitiendo a los planificadores urbanos reconocerlas tempranamente.
- La metodología puede ser útil no solo para el estudio del desarrollo espacial urbano

sino también para ámbitos relacionados, como el entorno social, económico o ambiental de las ciudades.

## 1.5. Estructura de la tesis

Este trabajo está estructurado como sigue: una descripción del estado del arte e indicadores *PageRank*, *Betweenness* e *Infomap* se presentan en el capítulo 2. En el capítulo 3 se describen los datos utilizados separados en tres fuentes: datos geográficos de la ciudad de Santiago de Chile, datos geográficos del sistema de transporte público y los datos producidos por los usuarios del sistema de transporte público, Transantiago. El capítulo 4 explica la implementación de la metodología, abarcando desde los datos iniciales, su transformación y posterior técnica de visualización. En el capítulo 5 se lleva a cabo un análisis a de los resultados obtenidos de los indicadores *PageRank*, *Betweenness* e *Infomap*, además de su comparación con otras métricas. Por último, en el capítulo 6 se presentan las conclusiones de este estudio y se especifican los posibles pasos a seguir en el futuro a partir de este trabajo.

# Capítulo 2

## Marco Teórico

La investigación urbana se entiende como la aplicación del conocimiento multidisciplinario sobre problemas urbanos, y donde la palabra urbano se enmarca dentro de una ciudad conceptualizada como una colección de componentes interrelacionados, destacando principalmente tres: sistema de actividades, sistema de transporte y las relaciones que se generan entre ellos. Estos componentes generan procesos espaciales relevantes como: difusión espacial, intercambios y transferencias, interacción espacial, segmentación y cambios temporales (Páez and Scott, 2005). Es en este ámbito en donde se enmarca este trabajo, que busca evaluar la aplicabilidad de una metodología desarrollada por Zhong et al. (2014) para la identificación de la estructura espacial urbana con datos de Santiago de Chile.

A continuación se describen los aspectos relevantes del estado del arte en los que se enmarca este trabajo. El capítulo se divide en 3 partes, comenzando por una definición de la estructura espacial urbana y sus tipos, después se realiza una descripción de las redes complejas y las métricas existentes para su clasificación que serán usadas en este trabajo. Y por último, se muestran trabajos realizados sobre clasificación de ciudades o zonas geográficas a partir de métricas de redes complejas.

### 2.1. Estructura Espacial Urbana

La forma de las ciudades y los factores que la causan y modifican ha sido un problema abarcado en la literatura durante varios años. Por ejemplo, Anas et al. (1998) describe la estructura espacial urbana y su evolución desde el punto de vista económico, un punto de vista importante para este análisis debido a que los patrones de crecimiento de las ciudades están experimentando cambios cualitativos. Algunos de estos correspondieron a avances en actividades intensas en información, desregulación, globalización (que obligó a las empresas a ser más competitivas), y por último, los que el autor destaca como más relevantes, los avances en transporte y telecomunicaciones.

El autor señala que las primeras ciudades industriales fueron altamente centralizadas alrededor de zonas de negocios hasta la introducción del transporte motorizado y las teleco-

municaciones. En los últimos dos siglos al menos, el autor indica que las ciudades han estado expandiéndose, pero en décadas recientes este proceso de descentralización ha tomado una forma más policéntrica. Esta nueva forma es planteada clásicamente a través de la identificación de centros de empleo, los cuáles mayoritariamente, correspondieron a cúmulos de empresas dependientes de aquellas que formaron parte de la tradicional zona de negocios de las ciudades (*Central Business District (CBD)*), y por lo tanto, se les identifica con el nombre de sub-centro.

La justificación del cambio de una tendencia monocéntrica a policéntrica es justificada, en parte, a través de las economías de aglomeración, término que se refiere a la disminución del costo promedio de un producto como consecuencia del incremento de producción dentro de un área geográfica. Este fenómeno también fue la base para entender los CBD's.

La visión planteada por el autor radica en comprender la estructura espacial urbana como el resultado de procesos de mercado pero también intentando abarcar otros dos ámbitos que considera relevantes: el rol del gobierno, donde el autor destaca la política de uso de suelo y la infraestructura de transporte como ítems claves en la forma de las ciudades, conclusión que aparece del estudio del desarrollo urbano de América del Norte. El segundo ámbito corresponde a la importancia del espacio en la economía, donde la identificación de los recursos geográficos producen nuevos conocimientos al fenómeno económico que son analizados por medio de modelos espaciales.

Uno de los acercamientos de análisis propuestos para describir la forma urbana (descrito por Anas et al. (1998)), consiste en medir el grado de concentración espacial de la población y empleos, donde se definen dos tipos de concentración espacial. La primera a un nivel de ciudad, donde se habla de actividades **centralizadas** o **descentralizadas**, dependiendo de como esta concentración se relaciona (en cercanía) con el CBD de la ciudad. Como herramientas de análisis en este caso han sido utilizadas principalmente funciones de densidad espacial. El segundo tipo de concentración corresponde a un nivel local, donde las actividades podrían ser agrupadas en patrones policéntricos, o dispersos en otros patrones más regulares. La identificación de estos patrones no es sencilla, una de las propuestas de análisis contempla la evaluación por medio de visualizaciones 3D de las funciones de densidad, pero se señala que es difícil fijar una metodología no arbitraria para identificar sub-centros, destacando también que la escala espacial utilizada juega un rol relevante.

El estudio de Anas et al. (1998) permite comprender que la estructura espacial es un fenómeno macro (a nivel de ciudad) que se origina a partir de muchos factores, y que por lo tanto no tiene una única causa, ni hay una forma específica de analizarlo, pero si propuestas, entre ellas el autor destaca el transporte como un medidor del cambio que experimentan las ciudades.

Luego, Rodrigue et al. (2013) define la estructura espacial urbana como la disposición del espacio en una ciudad producto de su forma urbana presente junto con las relaciones subyacentes entre estas, destacando el hecho de que estas relaciones nacen de la interacción generada por el movimiento de personas, mercancías o información entre un origen y destino.

Por otro lado, Humphreys and Ahern (2013) resume desde la literatura cuatro tipos de ciudad: monocéntrica, mono-policéntrica, policéntrica, policéntrica con movimientos aleatorios.

Una ciudad monocéntrica tiene un claro gradiente de densidad que aumenta desde la periferia hacia el centro, lugar donde se concentra la actividad económica, además se caracteriza por fuertes movimientos radiales del transporte público y poco uso de transporte privado, se señala el casco histórico de Dublin como un ejemplo. Una ciudad mono-policéntrica es aquella en donde el centro se mantiene como principal área económica pero con un incremento en la descentralización de los empleos (lo que ha debilitado al CBD) y un transporte público que sigue presentando una tendencia fuertemente radial, en esta categoría se señala a Londres como ejemplo. Luego describe a las ciudades policéntricas como aquellas que presentan varias zonas con patrones de agrupamiento de población (viviendas) y actividad económica donde la mayoría vive cerca de su lugar de empleo, generando viajes en transporte público generalmente locales, como ejemplo de lo descrito se señala Seúl, Korea. Por último, una ciudad policéntrica con movimientos aleatorios se describe como aquella que presenta sub-centros pero sin un centro dominante, produciendo que la actividad económica esté distribuida uniformemente, además cada sub-centro genera viajes desde todas partes de la ciudad (alta dispersión en los orígenes y destinos de viaje).

Alineando las definiciones de Anas et al. (1998), Rodrigue et al. (2013) y Humphreys and Ahern (2013) podemos comprender la estructura espacial urbana como aquella forma física que tiene la ciudad más la interacción espacial entre sus distintas zonas, donde las personas son representantes de esta interacción y el transporte público funciona como un medio para la concretización de esta comunicación.

## 2.2. Caracterización de redes complejas

En los últimos treinta años ha emergido un nuevo movimiento de interés e investigación en el estudio de redes, las denominadas redes complejas (Boccaletti et al., 2006). Una red compleja es aquella que presenta una estructura irregular, evoluciona en el tiempo y presenta una diversidad de configuraciones tanto para nodos como para conexiones (Strogatz, Steven H., 2001), yendo desde pequeños grafos que contienen un par de nodos a miles o millones.

Los esfuerzos para estudiar estas redes complejas han generado una serie de modelos (*small world*, *scale free*, entre otros) que han buscado caracterizarlas (Boccaletti et al., 2006; Watts and Strogatz, 1998; Kleinberg, 2000). El resultado más relevante de esto ha sido la identificación de principios unificadores y propiedades estadísticas comunes en muchas redes de diverso origen (*diameter*, *shortest path*, *average degree*, *clustering*, *closeness*, *strength*, entre otros).

Una clase de red compleja son las redes espaciales. Boccaletti et al. (2006) realizan un resumen de este tipo de redes donde comienza definiéndolas como aquellas en que sus nodos presentan una posición en un espacio euclidiano de dos o tres dimensiones y sus arcos son conexiones físicas reales. Ejemplos de este tipo de redes son las redes neuronales, redes de comunicación, redes de transporte, red de calles, red de vuelos realizados entre aeropuertos, red de trenes o red de tren subterráneo. Una de las características de este tipo de redes es la presencia de restricciones, algunas de estas según Boccaletti et al. (2006) son:

1. limitaciones espaciales producto del lugar en el que se encuentra
2. grado de los nodos presenta una restricción puesto que la cantidad de conexiones que puede tener un nodo está limitado por el espacio físico que los conecta
3. nodos distantes tienen menos probabilidad de estar conectados debido a la dependencia que tienen los arcos sobre la distancia
4. muestran una correlación fuerte entre grado y coeficiente de agrupamiento

Es importante destacar estas restricciones puesto que la red que se construye en este trabajo contiene fuertes restricciones subyacentes que vienen dadas por los recorridos de los servicios del sistema de transporte, que a su vez, están limitados por la configuración de calles presente en la ciudad, lo que muestra una fuerte dependencia del espacio en donde habita.

A continuación se describen distintas métricas para caracterizar una red (grado promedio, promedio de rutas más cortas, coeficiente de agrupamiento, componentes conexas y diámetro) con el objetivo de entender su estructura y también medidas asociadas a sus nodos (grado de entrada, grado de salida, *PageRank*, *Betweenness*, *Infomap*), esto último para identificar qué función cumple cada nodo en la estructura de la red ya que es conocido que la posición de un nodo en un grafo repercute fuertemente en su función o importancia, según lo que se esté midiendo. Las definiciones que siguen fueron obtenidas de Boccaletti et al. (2006) y Newman (2003).

### 2.2.1. Grado

El grado o conectividad  $k_i$  de un nodo  $i$  es el número de arcos conectados al nodo, y es definido en términos de la matriz de adyacencia del grafo según 2.1:

$$k_i = \sum_{j \in N} a_{ij} \quad (2.1)$$

donde  $N$  es el número total de nodos en la red y  $a_{ij}$  toma valor 1 si hay un arco entre  $i$  y  $j$  o 0 en otro caso. Si el grafo es dirigido, el grado del nodo tiene dos componentes: el número de arcos de salida  $k_i^{out} = \sum_j a_{ij}$ , y el número de arcos de entrada  $k_i^{in} = \sum_j a_{ji}$ . El grado total es entonces definido como  $k_i = k_i^{out} + k_i^{in}$ .

La caracterización topológica más básica de un grafo  $G$  puede ser obtenida en términos de la distribución del grado  $P(k)$ , definido como la probabilidad de escoger un nodo uniformemente al azar que tenga grado  $k$ , o equivalentemente, como la fracción de nodos en  $G$  que tienen grado  $k$ . En el caso de redes dirigidas se requiere considerar dos distribuciones:  $P(k^{in})$  y  $P(k^{out})$ .

### 2.2.2. Promedio de rutas más cortas

El promedio de rutas más cortas (*average shortest path lengths* o *characteristic path length* en inglés) es una medida de separación entre dos nodos en un grafo  $G$  y se define como el

promedio de los caminos más cortos entre cada par de nodos. La ecuación 2.2 muestra su fórmula, donde  $N$  es la cantidad de nodos en  $G$  y  $d_{ij}$  corresponde a la cantidad de arcos mínimos requeridos para conectar el nodo  $i$  con el nodo  $j$ .

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij} \quad (2.2)$$

Un problema con esta definición es que  $L$  diverge si el grafo no es conexo. Una posibilidad para evitar la divergencia es limitar la sumatoria solo a los pares de nodos conectados y luego promediar el resultado de cada componente conexas.

### 2.2.3. Coeficiente de agrupamiento

El coeficiente de agrupamiento ( $C$ ), también conocido como transitividad, es una propiedad típica en redes sociales, donde dos individuos con un amigo en común son probablemente conocidos entre si. Desde la mirada de un grafo genérico  $G$  con  $N$  nodos, la transitividad significa la presencia de un alto número de triángulos (conjunto de tres nodos conectados a un arco de distancia entre ellos). Un ejemplo de lo anterior es mostrado en la figura 2.1, donde se observa la evaluación de  $C$  sobre el nodo morado: a la izquierda se observa que el nodo evaluado siempre puede formar triángulos con todos sus pares de vecinos, por lo que obtiene el valor máximo (1), para el segundo caso se eliminan dos arcos (arcos de color rojo), por lo que ahora solo es posible formar un triángulo de tres posibles, obteniendo un valor de  $C$  igual a  $1/3$ . Por último, se eliminan todos los arcos entre vecinos, resultando en un  $C$  igual a cero.

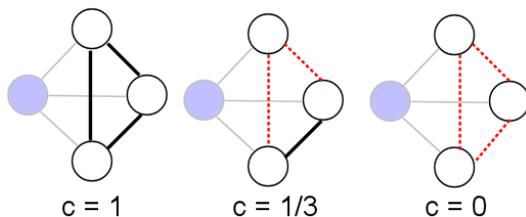


Figura 2.1: Ejemplo de coeficiente de agrupamiento. Fuente: wikipedia.org

Formalmente se define la transitividad  $T$  como:

$$T = \frac{3 \times \text{NúmeroDeTriángulosEnG}}{\text{NúmeroDeTripletes}} \quad (2.3)$$

el factor tres compensa que cada triángulo contribuye con tres tripletes conectados (uno centrado sobre cada uno de los tres nodos) y asegura que  $0 \leq T \leq 1$ . Notar que cuando  $T = 1$  se tiene el clique  $K_n$ . Para el caso particular de un nodo en la red, el coeficiente de agrupamiento está dado por:

$$T(i) = \frac{E_i}{k_i(k_i - 1)/2} \quad (2.4)$$

donde  $E_i$  son los arcos entre los vecinos de  $i$  y  $K_i$  el grado del nodo  $i$ , es decir, la proporción de arcos presente con respecto a todos los posibles arcos. La fórmula anterior entrega información

del clustering local (por nodo), para obtener el coeficiente de agrupamiento de la red se puede calcular el promedio (ecuación 2.5).

$$T = \frac{1}{N} \sum_{i \in N} T(i) \quad (2.5)$$

## 2.2.4. Componentes conexas

Una componente conexa en un grafo es aquella en que siempre existe un camino entre cualquier par de nodos asociados a ella, y un grafo siempre tiene una o más componentes conexas. Es importante mencionar (para efectos del análisis posterior) que el concepto puede ser ampliado cuando se trata de grafos dirigidos, incorporando dos definiciones: conexo fuerte y conexo débil. Una componente es conexas fuerte si existe un camino dirigido entre cada par de nodos de la red, mientras que se dice que es conexo débil si existe un camino entre cada par de nodos en la versión de la red sin dirección en los arcos.

## 2.2.5. Diámetro

Las rutas más cortas juegan un rol importante en el transporte y comunicación dentro de una red, por lo que resultan relevantes a la hora de caracterizar la estructura interna de un grafo. El diámetro se define como aquella ruta más larga (en número de arcos) entre todas las rutas más cortas para todos los pares de nodos y busca describir el tamaño de la red.

Un problema a considerar en su cálculo es cuando el grafo no es conexo; en esta situación su valor será el mayor diámetro obtenido de las componentes conexas.

## 2.2.6. *PageRank*

La idea detrás del *PageRank* es medir la capacidad de un nodo para atraer flujos desde todos los nodos de la red. Esta versión está basada en la utilizada por Rosvall and Bergstrom (2008) para medir la importancia de un nodo haciendo una analogía con el *PageRank* de *Google* (Brin and Page, 2012).

Según lo anterior, se define la probabilidad  $r_j$  de visitar el nodo  $j$  como:

$$r_j = [(1 - \rho)/N] + \rho \sum_i r_i p_{ij} \quad (2.6)$$

Donde  $1 - \rho$  puede ser visto como la probabilidad de que un caminante aleatorio que recorre la red decida cambiarse de nodo, y  $p_{ij}$  como la probabilidad de escoger ir desde nodo  $i$  al nodo  $j$ , cuyo valor es proporcional al peso del arco de  $i$  a  $j$ , es decir:

$$p_{ij} = w_{ij} / \sum_k w_{ik}, \text{ y } \sum_j p_{ij} = 1 \quad (2.7)$$

El parámetro  $\rho$  es conocido como factor de amortiguación y puede tomar valores entre 0 y 1. En estos experimentos usamos el valor 0,85, que es el usado por Zhong et al. (2014) en su análisis. La probabilidad del sistema estable<sup>1</sup>  $r_j$  puede ser calculada mediante diversos métodos (Zhong et al., 2014).

Un alto valor del *PageRank* para un nodo  $k$  en un grafo está justificado, según la definición del indicador, por alguno de los siguientes motivos:

- Nodo con alto PageRank apunta hacia el nodo  $k$  (ejemplo: nodo C en figura 2.2).
- Nodo presenta varios arcos que inciden en el nodo  $k$  (ejemplo: nodo E en figura 2.2).
- Una combinación de los dos puntos anteriores (ejemplo: nodo B en figura 2.2).

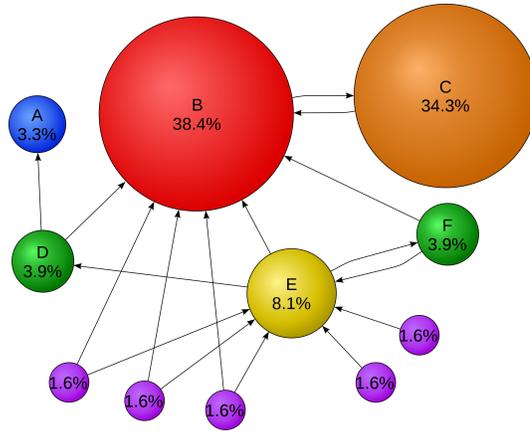


Figura 2.2: *PageRank* en una red sin pesos. Fuente: Wikipedia.org

### 2.2.7. *Betweenness*

El *Betweenness* es una medida de centralidad que caracteriza la importancia de un nodo en la organización del flujo en una red, la que está determinada por su habilidad para generar una ruta entre dos regiones (Newman, 2003). Para un nodo  $k$ , el *Betweenness* se define como el número de caminos más cortos que conecta otro par de nodos  $i$  y  $j$  en el grafo que pasa por el nodo  $k$ . Un nodo tiene un valor alto de *Betweenness* si un mayor número de rutas más cortas lo atraviesa. Según lo anterior se define como:

$$Betweenness(k) = \sum_{ij} \frac{\delta_{ij}(k)}{\delta_{ij}} \quad (2.8)$$

donde  $\delta_{ij}(k)$  es el número de caminos más cortos desde  $i$  a  $j$  que pasan por  $k$ , mientras que  $\delta_{ij}$  es el número de caminos más cortos desde  $i$  a  $j$ . Lo anterior se realiza para cada par  $i, j$  distintos de  $k$ .

<sup>1</sup>El *PageRank* puede ser modelado como un proceso de markov, donde la probabilidad asignada a cada nodo se obtiene cuando este alcanza su estado estable (o de equilibrio).

## 2.2.8. Estructura de comunidad: *Infomap*

Las comunidades en un grafo se definen como grupos de nodos que tienen una alta densidad de arcos entre sí y una baja densidad entre grupos. En la literatura se reportan variadas formas de detectar comunidades (Newman, 2003) y dada la naturaleza de los datos disponibles, en este trabajo se privilegian aquellas que consideran las variables de densidad (cantidad de arcos) y flujo de interacciones (comportamiento de un peatón aleatorio en la red). Concretamente se espera que dentro de una comunidad se observe una gran cantidad de arcos y que un peatón aleatorio se mueva mayoritariamente dentro de la comunidad. Asimismo, se espera que ambas variables (densidad y flujo) sean mayores al interior de la comunidad que entre comunidades. Otra característica importante que deben presentar las comunidades es ser excluyentes, dado que se busca identificar zonas de interacción en la ciudad bien definidas.

Dadas las consideraciones anteriores se hace uso de un método de detección de comunidades propuesto por Rosvall and Bergstrom (2008) llamado *Infomap*. Los análisis comparativos han mostrado un excelente funcionamiento en la generación de módulos con las características mencionadas anteriormente (Lancichinetti and Fortunato, 2009).

Este algoritmo consiste en generar recorridos aleatorios a lo largo de la red basado en la importancia de los nodos según el algoritmo *PageRank* (el mismo explicado en 2.2.6), para luego particionar el grafo en módulos tal que comprime una descripción del recorrido aleatorio. La compresión se logra minimizando la entropía que se define como:

$$\left. \begin{aligned} Lg(M) &= H(P) + \sum_{i=1}^m P_i H(p)_i \\ &= -p \sum_{i=1}^m P_i \log P_i - \sum_{i=1}^m P_i \sum_{k=1}^{M_i} \frac{P_k}{P_i} \log \frac{P_k}{P_i} \end{aligned} \right\}, P_i = \sum_k p_k \quad (2.9)$$

donde  $P_i$  es la probabilidad de ser visitada la comunidad  $i$ ,  $P_k/P_i$  es la probabilidad de visitar el nodo  $k$  dado que la comunidad  $i$  es visitada y  $M_i$  es la cantidad de nodos que contiene la comunidad  $i$ . Por último, la división por  $P_i$  que afecta a los  $P_k$  cumple la función de normalizar. Es importante decir que  $M_i$  es un módulo que contiene un conjunto de  $k \in M_i$  nodos y que llega a ser estable (sin alteraciones) una vez que se obtiene la entropía mínima.

La forma en que trabaja el algoritmo es primeramente asociando cada nodo a un módulo<sup>2</sup>, en cada paso, identificar que nodo se puede agregar a que módulo tal que la entropía general decrezca. Este proceso continúa hasta que no se pueda reducir más la entropía, asegurando que con esa configuración se obtiene la partición más estructurada.

Las métricas Grado, promedio de rutas más cortas, coeficiente de agrupamiento y diámetro son usadas para caracterizar los grafos usados en este trabajo, mientras que el *PageRank*, *Betweenness* e *Infomap* son usados como representación de los espacios urbanos descritos por Zhong et al. (2014) para la identificación de la estructura espacial urbana.

---

<sup>2</sup>Al inicio existen tantos módulos como nodos contiene la red

## 2.3. Ciudades desde redes complejas

Los avances actuales en tecnología (GPS, RFID, Twitter, Foursquare, teléfonos móviles, tarjetas inteligentes, entre otros) han puesto a disposición de la comunidad fuentes de datos masivos con un nivel de detalle y profundidad a los que no se tenía acceso previamente. Lo anterior ha permitido construir distintos tipos de redes espaciales sobre las ciudades, lo que ha permitido analizar distintos problemas desde otro punto de vista, el de las redes.

A continuación se resume algunos trabajos relacionados con la identificación de estructuras subyacentes en las ciudades, regiones o países (bordes geográficos, zonas de interés, zonas de actividad, etc.) a partir de redes espaciales creadas desde datos generados por la interacción de las personas o de la ciudad. Por ejemplo: llamados telefónicos, circulación espacial de billetes, movimientos en transporte público, rastreo de vehículos mediante GPS, calles.

Zhong et al. (2014) es la principal referencia de este trabajo, donde busca identificar cambios en la estructura espacial de la ciudad de Singapur analizando tres cortes temporales (2010, 2011 y 2012) sobre la red de viajes (nodos son paradas de bus o metro y sus arcos representan viajes realizados entre pares de paradas) en transporte público utilizando los datos capturados desde las tarjetas inteligentes de pago del sistema. En Singapur las personas validan en la subida y bajada de bus o metro dado que la tarifa es segmentada por distancia, esto permite generar las redes dirigidas con pesos necesarias para calcular los espacios urbanos que propone: zonas de concentración de personas con *PageRank*, zonas usadas para conectar otro par de zonas con *Betweenness* y la identificación de áreas con fuerte interacción interna (comunidades) con *Infomap*. Las conclusiones destacan la identificación clara (bordes geográficos) de comunidades que dividen el espacio de las actividades de la población, que además son distintos de los político-administrativos existentes. Además, con la evaluación temporal concluye que Singapur está cambiando su estructura espacial a una más policéntrica, esto basado en la presencia de nuevos subcentros (identificados con el *PageRank*) y nuevas comunidades que están emergiendo en línea con el plan de desarrollo de la ciudad.

Una característica importante del trabajo de Zhong et al. (2014) es la utilización de análisis espacial para el *PageRank* y *Betweenness*. Aquí plantea el uso de la interpolación espacial para obtener un mapa de calor de los indicadores sobre la ciudad, pasando de mediciones discretas (asociadas a los nodos de la red) a la asignación del indicador en cada parte de la ciudad. Para esto asume que las personas escogen la parada más cercana a su destino, por lo que cada nodo de la red espacial ejerce una influencia sobre su entorno que va disminuyendo con la distancia. Un método que refleja este comportamiento (y usado por Zhong et al. (2014)) es *Inverse Distance Weighted (IDW)*, el cual considera que cada punto genera una influencia que disminuye con la distancia y se define como:

$$W_i(x, y) = \frac{1}{d_{i(x,y)}^\lambda} \quad (2.10)$$

donde  $(x, y)$  es la coordenada geográfica de la cual queremos conocer la influencia que ejerce el punto de ubicación  $i$ , entonces  $W_i(x, y)$  es la contribución del punto  $i$  a la ubicación  $(x, y)$ , y  $d_{i(x,y)}$  es la distancia desde el punto  $i$  al punto  $(x, y)$ .  $\lambda$  es un parámetro establecido en 2, por lo que sigue una ley inverso al cuadrado. En este trabajo se decide mantener esta técnica

de interpolación para poder comparar los resultados de las métricas obtenidos en Santiago con los de Singapur.

El *PageRank* como ranking de zonas urbanas ha sido utilizado previamente. Agryzkov et al. (2012) planteó un método para medir la importancia relativa de nodos en redes urbanas considerando distintos atributos físicos como la presencia de monumentos, museos, restaurantes, entre otros. Para lograr este ranking Agryzkov et al. (2012) modifica la probabilidad de que el peatón aleatorio se cambie a un nodo según una matriz de atributos. Esto fue un intento de establecer una importancia relativa de las distintas zonas en la ciudad, donde a partir de atributos de las calles el autor intenta predecir el flujo o atractividad que podría tener. Al compararlo con Zhong et al. (2014) vemos que esta última plantea un punto de partida donde los flujos ya existen y se basa en ellos para medir la importancia del lugar, caso más acorde a la realidad de Santiago donde se tienen los datos de subida y una estimación de bajada de los flujos en transporte público.

Respecto a la generación de bordes geográficos, Zhong et al. (2014) se basa en el trabajo previo de Rinzivillo et al. (2012) que muestra como las interacciones sociales capturadas a través de movimientos de vehículos mediante GPS permiten extraer bordes geográficos relevantes por medio de la generación de comunidades. Trabajos previos realizan algo similar pero a un nivel macro, como lo son regiones o países. Por ejemplo, Thiemann et al. (2010) y Ratti et al. (2010) generan bordes geográficos a gran escala a partir de actividades de movimiento. Thiemann et al. (2010) analiza una red de movilidad humana extraída desde el proyecto *Where's George?*<sup>3</sup>, y usando un método estocástico extraen una partición de regiones por medio de una función basada en maximizar la modularidad. Los experimentos fueron realizados a gran escala espacial donde la granularidad estuvo dada por la identificación del código postal en los Estados Unidos. Por otro lado, el trabajo de Ratti et al. (2010) también adopta una función de modularidad para delinear bordes desde una red extraída desde registros telefónicos. Si bien ambos obtienen resultados interesantes, Rinzivillo et al. (2012) explica que otros autores han mostrado que la función de modularidad tiene un problema de resolución que causa que comunidades pequeñas sean ignoradas o unidas, por lo que opta por utilizar una técnica basada en peatón aleatorio y teoría de la información llamado *infomap*. Sus resultados muestran que a una escala más pequeña (municipios) *Infomap* logra asignar sectores de la misma municipalidad a la misma comunidad (generando comunidades espacialmente adyacentes), además muestra que ningún municipio es una comunidad en si misma, indicando que un municipio con flujos autocontenidos no existe.

Respecto a la identificación de la estructura espacial, se destacan dos trabajos interesantes: Louail et al. (2014) y Louail et al. (2015). El primero realiza un análisis de datos telefónicos de 31 ciudades de España capturados por 55 días. Primero define un índice de dilatación que mide como varía la distancia promedio entre los individuos durante el día, permitiendo destacar diferentes tipos de estructuras sobre la ciudad, en particular, se centran en las zonas de mayor concentración de señales telefónicas (*hotspot*), donde proponen un método libre de parámetros para detectarlos, descubriendo que el número de *hotspot* escala sublinealmente con el tamaño de la población. Además, analiza el comportamiento de los *hotspot* durante el día, descubriendo una jerarquía de permanencia, lo que les permite identificar lo que ellos

---

<sup>3</sup>Proyecto que consiste en que los usuarios (por medio de una página web) registran el número de serie de un billete y código postal de donde se encuentran, generando así un historial de donde estuvo el dinero.

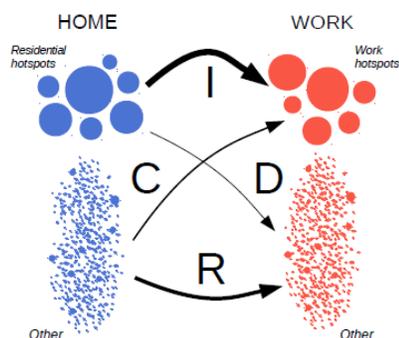


Figura 2.3: Ilustración del método ICDR. Fuente: Louail et al. (2015)

llaman el corazón de la ciudad, *hotspot* que nunca desaparecen a lo largo del día, resultado que está presente en todas las ciudades estudiadas. Por último, destaca la distribución espacial de *hotspot*, señalando que les permite distinguir distintas categorías de ciudades, desde monocéntricas a policéntricas. El segundo trabajo, y más reciente, utilizando los mismos datos del trabajo anterior (31 ciudades) destaca un método para capturar la estructura de flujos de viajes en una ciudad a partir de la separación de su flujo en cuatro categorías, agrupadas en una matriz de  $2 \times 2$  y generadas a partir de los datos de llamados telefónicos de las personas de una ciudad capturados por 55 días. En la figura 2.3 se puede observar las cuatro categorías de flujo, donde **I** corresponde al flujo integrado desde *hotspot* a *hotspot*, **C** es el flujo convergente que va desde lugares residenciales aleatorios a *hotspot* de trabajo, **D** es el flujo divergente que va desde *hotspot* residenciales a lugares de empleo aleatorios, y por último, **R** que representa los flujos aleatorios que ocurren en la ciudad, yendo desde y hacia lugares que no son *hotspots*. Parte de sus resultados muestra que las ciudades se diferencian esencialmente por la proporción de dos tipos de flujo: integrado (entre zonas residenciales y zonas de empleo) y aleatorio, donde la importancia de este último incrementa con el tamaño de la ciudad. Lo anterior les permite clasificar las ciudades a partir de su estructura de viajes.

Por último, otro trabajo interesante es el planteado por Soh et al. (2010), donde realiza un análisis sobre la red de viajes en Singapur (mismo sistema de transporte usado por Zhong et al. (2014)), donde el autor separa la red en dos: una de viajes hechos solo en bus y otra de viajes realizados solo en metro. Uno de los hechos rescatables de este trabajo es la descripción de los patrones de uso del sistema en transporte público, encontrando una similitud con los datos de Santiago: los días laborales (lunes a viernes) se comportan de manera similar mientras que sábados y domingos tienen comportamientos distintos. Esto muestra que una comparación como la que queremos hacer acá con los resultados de los espacios urbanos (centros, centros de pasada y comunidades) entre Santiago y Singapur son a partir de patrones de uso del sistema similares. Otros tipos de resultados están basados en propiedades calculadas desde las redes de bus y metro, pero escapan de este trabajo puesto que aquí se considera la red completa.

# Capítulo 3

## Datos

Los datos utilizados en este trabajo provienen de tres fuentes: viajes de usuarios en transporte público, sistema de transporte público Transantiago y la división geo-administrativa de la ciudad; en la tabla 3.1 se puede observar cada uno de ellos. El primero corresponde a los viajes de usuarios de transporte público con tarjeta bip!, datos de estudio de este trabajo y que representan una semana de viajes, específicamente entre el 14 y 21 de abril de 2013, de domingo a domingo. Es importante decir que la elección de una semana se fundamenta en el hecho de que estos días contemplan la mayoría de las variantes de movilidad que pueden realizar las personas en la ciudad<sup>1</sup>.

Las otras dos fuentes de datos corresponden a datos referenciales, el primero corresponde a datos espaciales del sistema de transporte público, Transantiago, y el segundo es la división territorial de la ciudad de Santiago de Chile. Del primero se tiene la distribución espacial de las paradas de buses a lo largo de la ciudad y la ubicación de las estaciones del tren subterráneo; el segundo entrega datos de los límites político-administrativos, aquí se encuentran tres configuraciones, yendo desde la zonificación agregada que divide la ciudad en siete grandes sectores, pasando por una división más detallada que corresponde a una escala de comunas, donde se divide la ciudad en 45 comunas, y por último, la zonificación más detallada, que divide la ciudad en 866 zonas, esta partición lleva por nombre 777 y fue creada por la oficina de planificación de transporte (SECTRA).

Dimensión \ Dato	Viajes de usuarios	Sist. de transporte	División territorial
Tipo de dato	estudio	referencial	referencial
Rango temporal	un semana	estático	estático
Descripción	viajes, etapas y fecha-hora de inicio	paradas de bus estaciones de metro	macro-sectores, comunas y zonificación 777

Tabla 3.1: Datos utilizados. Fuente: elaboración propia

<sup>1</sup>Una semana representa la máxima extensión temporal que el departamento de Transporte puede procesar al momento de realización de este trabajo

A continuación se describen los datos de viaje usados y de la ciudad de Santiago de Chile. Los datos demográficos fueron obtenidos de la encuesta origen-destino del año 2012, que corresponde a un estudio que se realiza cada diez años y es la base de datos fundamental para apoyar la formulación de planes y proyectos de infraestructura para mejorar las condiciones de movilidad de las personas y mercancías de Santiago, así como también de áreas estrechamente relacionadas, como el medio ambiente y desarrollo urbano (Muñoz and Contreras, 2015).

### 3.1. Datos de estudio: viajes

Los viajes realizados en el sistema de transporte público son obtenidos del sistema de pago de tarifa implementado en Transantiago, el que utiliza la tecnología AFC por medio de tarjetas inteligentes (conocidas comercialmente como tarjetas bip!) y el sistema AVL a través de GPS en los buses que operan en el sistema.

En términos simples, el proceso de identificación de viaje consta de dos etapas. El primero corresponde a la identificación del paradero de subida y el segundo a la identificación del paradero de bajada, siendo este último de mayor complejidad.

El proceso de identificación de paradero de subida considera tres casos: el primero corresponde a la situación en que un usuario ingresa a una estación de metro, en esta situación se conoce el lugar de subida automáticamente dado que los aparatos que descuentan el saldo de la tarjeta bip! son estáticos, por lo que se les puede asociar una posición que no cambia en el tiempo y obtener el lugar de subida. El segundo caso corresponde a un usuario abordando un bus. La tecnología implementada considera como parte del equipo de un bus dos aparatos: (1) El GPS que informa la posición del bus y hora de la medición cada 30 segundos y (2) un dispositivo que descuenta el saldo correspondiente de la tarjeta bip! cuando esta se acerca físicamente, dejando almacenado el número identificador de la tarjeta e instante en que se produjo la validación. En este caso el paradero de subida es obtenido al relacionar la hora en que se produjo la transacción y la hora de las posiciones del GPS, obteniendo una estimación de la posición del bus en el instante de la transacción. Con la posición del bus se obtiene el paradero más cercano y se asocia a la transacción de subida. El tercer caso corresponde a las validaciones hechas en las zonas pagas, que corresponde a espacios físicos donde los validadores se encuentran en el paradero y no en el bus, esto busca disminuir el tiempo de cada persona al abordar un bus. El proceso de asignación de zona paga es realizado previamente por lo que se realiza un procedimiento similar al primer caso, aunque no es posible asignar el bus que abordó (dado que no hay validación en este).

Por otro lado, el paradero de bajada presenta un desafío mayor toda vez que el usuario no deja registro de ningún tipo al momento de abandonar el sistema, lo que requiere implementar una serie de reglas (supuestos) que permitan identificar viaje y etapa. La definición de viaje y etapa usadas por Munizaga and Palma (2012) son:

- Viaje: traslado entre un origen y un destino donde el usuario realiza una actividad para satisfacer una necesidad (Ortúzar et al., 2011).
- Etapa: traslado en bus o metro necesario para completar el viaje.

Las reglas desarrolladas en Munizaga and Palma (2012) se resumen en las siguientes (algunas basadas en investigaciones previas):

- Cada tarjeta de prepago (bip!) corresponde a un usuario.
- Si la tarjeta de prepago (bip!) tiene solo un viaje, no es posible inferir con un solo día de información.
- Después de un viaje, los usuarios volverán al destino del viaje anterior (Barry et al., 2002).
- Al final del día, los usuarios volverán al origen del primer viaje del mismo día (Barry et al., 2002).
- Se asigna como lugar de bajada la parada más conveniente para acceder a la posición de la siguiente parada de subida dentro de un radio de 1000 metros.
- Se observan viajes de días siguientes para completar información perdida (Trépanier et al., 2007).

Lo anterior obtiene una estimación de bajada para más del 80 % de las transacciones usadas como muestra. A partir de los resultados anteriores es que se sabe que durante una semana se producen aproximadamente 35 millones de viajes con cerca de 3 millones de tarjetas.

La ventana temporal utilizada en este trabajo se establece entre el 15 y 21 de abril de 2013, es decir, una semana de datos, de lunes a domingo. La elección de una semana se fundamenta en el hecho de que estos días contemplan todas las variantes de movilidad que pueden realizar las personas. A continuación se caracterizan los datos a ser utilizados.

### 3.1.1. Análisis estadístico

Dado que el sistema de pago funciona sólo con validación en la subida al bus o metro, su bajada y destino es estimada utilizando las metodologías propuestas por Munizaga and Palma (2012) y Munizaga et al. (2014). Del resultado de la estimación se requieren aquellos datos de viaje que contienen bajada conocida en todas sus etapas, requisito necesario para generar los arcos de la red de movilidad. A partir de lo anterior se obtiene una muestra que corresponde a 18.535.153 viajes validados con 2.851.393 tarjetas inteligentes, siendo el 73,42 % de los viajes realizados durante la ventana de tiempo descrita.

Durante la semana de la muestra se observaron 25.244.791 viajes, que en su conjunto acumulan 35.983.995 etapas hechas con 3.419.038 tarjetas inteligentes. La tabla 3.2 muestra como se distribuyen los viajes basado en su cantidad de etapas, mostrando que gran parte de estos corresponden a viajes con una etapa (63,93 %), lo que indica que más de la mitad no realiza transbordo alguno para llegar a su destino. Los viajes que realizan dos etapas corresponden al 29,91 %, esto nos dice que los viajes están altamente concentrados en configuraciones de 1 o 2 etapas (93,84 %), el resto (6,16 %) está constituido por viajes de 3 o más etapas.

Un aspecto a destacar es la distribución temporal que tienen los datos, en la figura 3.1 se puede observar la cantidad de viajes realizados por cada día considerado en la muestra. Otro

Número de etapas	Número de viajes	% del total
1	16.139.875	63,93 %
2	7.549.652	29,91 %
3	1.487.254	5,89 %
4	59.176	0,23 %
5 o más	8.834	0,04 %

Tabla 3.2: Cantidad de viajes según su cantidad de etapas. Fuente: elaboración propia

aspecto importante en este ámbito es la similitud en volúmenes de viaje presentes durante los días laborales de la semana (lunes a viernes), los que fluctúan cercanos a los 4 millones de viajes diarios. Para los días del fin de semana, sábado y domingo se observa una diferencia considerable, indicando que para los ciudadanos representan días distintos entre ellos, y a su vez, con los días laborales.

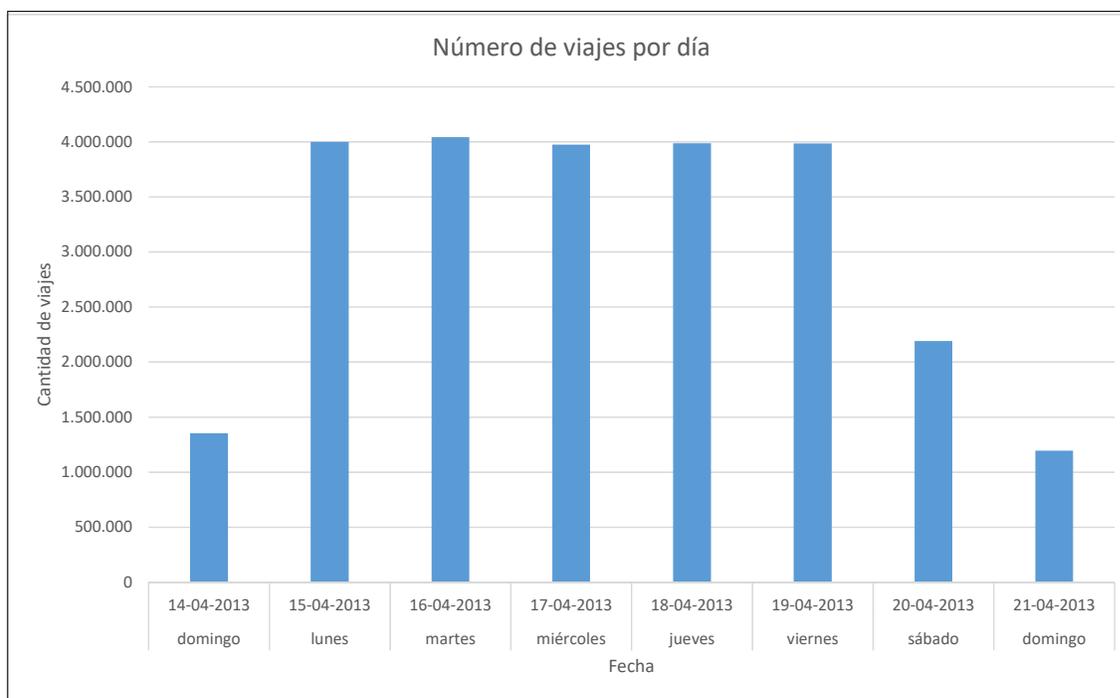


Figura 3.1: Cantidad de viajes por día. Fuente: elaboración propia

Desde el punto de vista de la hora del día en que estos viajes comienzan, se puede observar en la figura 3.2 que existen dos períodos temporales donde el volumen de viajes alcanza su máxima magnitud; el primero corresponde al período de la mañana, entre las 06:00 y 09:00 horas, lo que se relaciona con el inicio de la jornada laboral o educacional de gran parte de la población. El segundo máximo se alcanza en el período de la tarde, entre las 17:00 y 21:00 horas, esta vez comprendido como el efecto opuesto al pico anterior, gran parte de la población vuelve a su hogar. Otro suceso relevante es el que se produce al medio día, donde se alcanza un máximo local (13:00 horas) que irrumpe sobre las horas contiguas.

Otra característica importante es la distribución espacial presente en la muestra, en la figura 3.3 se puede observar el número de viajes iniciados por comuna de la ciudad, donde

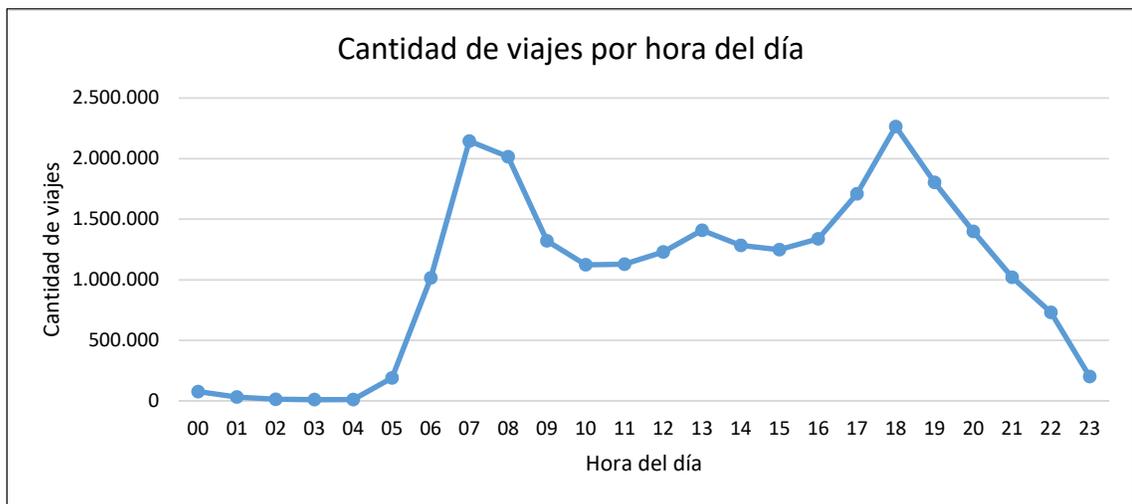


Figura 3.2: Cantidad de viajes por hora. Fuente: elaboración propia

la comuna de Santiago concentra por si solo casi el 20% de los viajes totales, seguido por Maipú y Puente Alto, que en su conjunto agrupan el 30% de los viajes. Lo anterior indica la importancia que tiene la comuna de Santiago a la hora de generar viajes, lo que puede ser entendido por la alta población flotante que presenta el sector al ser este utilizado en gran parte como zona de trabajo más que de vivienda. Además se puede observar que la comuna con menor cantidad de viajes corresponde a la comuna de San Ramón, que también corresponde a la comuna de menor tamaño donde Transantiago funciona.

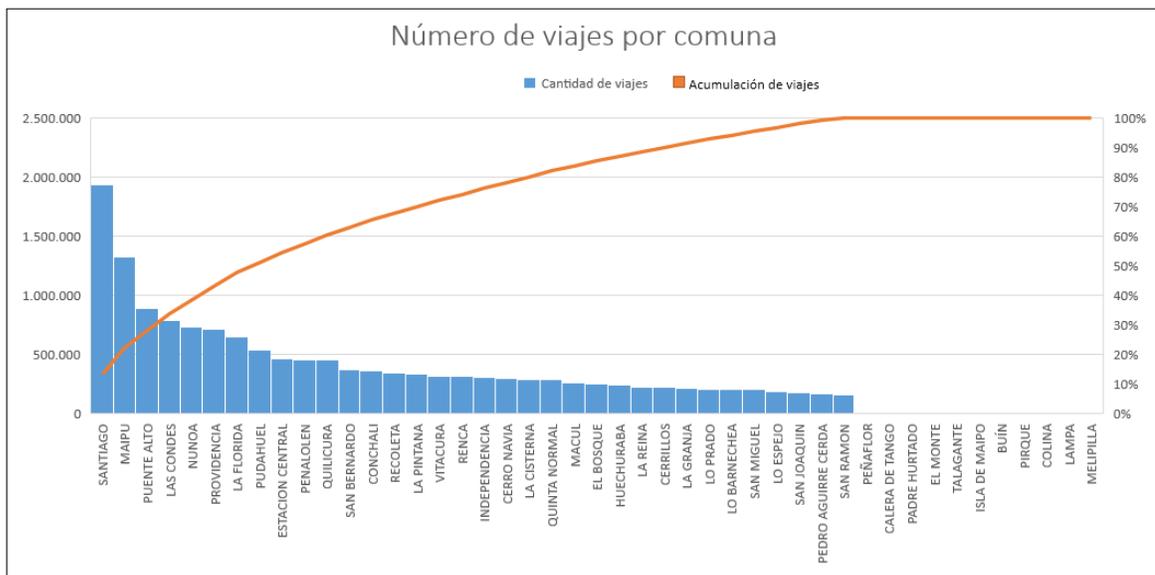


Figura 3.3: Cantidad de viajes por comuna. Fuente: elaboración propia

El análisis anterior permite rescatar observaciones que deben ser consideradas en los análisis de este trabajo puesto que son características relevantes de los datos, como lo son los días que presentan un comportamiento de flujo similar pudiendo unir estos días en un solo análisis, como lo son los días laborales (lunes a viernes). Además se debe considerar un análisis particular para aquellos tramos horarios que presentan una mayor demanda de viajes, como

lo son los picos de la figura 3.2. Por último se debe tener presente la distribución espacial de los viajes como un método para contrastar si las métricas generadas en este trabajo se condicen con los comportamientos presentes en la ciudad.

## 3.2. Datos referenciales

### 3.2.1. Santiago de Chile

La ciudad de Santiago es la capital de Chile y corresponde a la ciudad más poblada del país. En ella viven aproximadamente 6,5 millones de personas, representando a cerca del 35 % de la población nacional.

El centro geográfico ha sido el punto neurálgico de la ciudad, concentrando una gran cantidad de población flotante y una alta densidad de zonas de trabajo, mientras que la periferia se ha desarrollado como áreas de habitabilidad en su mayoría. Dado el desarrollo de la urbe es que presenta una forma circular con una gran proporción de viajes desde la periferia hacia el centro durante las mañanas y en sentido opuesto durante las tardes. En términos de género, los datos muestran una población ligeramente mayor de mujeres (51,02 %) y destacando que esta diferencia se observa en el rango de edad entre 51 a 65 años, donde los hombres concentran el 16,7 % de su población mientras que las mujeres el 17,8 %.

La tabla 3.3 muestra que la población de Santiago se concentra mayoritariamente en los sectores Sur, Suroriente y Poniente, confirmando el carácter mayoritariamente habitacional de la periferia. Otro aspecto relevante se observa en la relación entre habitantes por hogar, donde el sector centro y oriente presentan valores más bajos.

Sector	Habitantes (miles)	Hogares por IFM (miles)			Hab/Hogar
		Bajo	Medio	Alto	
Norte	976,0	103,4	174,8	9,7	3,4
Poniente	1.412,9	179,8	231,4	9,9	3,4
Oriente	966,2	46,0	150,2	131,1	3,0
Centro	344,8	40,2	95,1	11,5	2,3
Sur	1.236,7	160,2	190,9	9,5	3,4
Sur-Oriente	1.331,4	137,6	232,0	25,2	3,4
Extensión Surponiente	383,7	55,1	56,4	1,4	3,4
<b>TOTAL</b>	6.651,7	722,2	1.130,8	198,3	3,2

Tabla 3.3: Antecedentes demográficos por sector separado por IFM (Ingreso Familiar Mensual). Fuente: EOD-2012

Se sabe que Santiago es una ciudad segregada económicamente. Las personas de mayores ingresos habitan en el sector oriente de la ciudad, mientras que las personas de menores ingresos lo hacen en mayor medida en los sectores Poniente, Sur y Extensión Surponiente. La tabla 3.3 muestra esta realidad, donde un ingreso bajo se define como aquel ingreso familiar mensual entre \$0 y \$ 400.000 pesos, mientras que un ingreso medio se entiende como aquel

percibido entre \$ 400.001 y \$ 1.600.000, y por último, el ingreso alto agrupa a todas las familias que tienen un ingreso familiar mensual superior al \$ 1.600.000. Se puede ver que el sector oriente es el que presenta el mayor porcentaje de hogares de altos ingresos con un 40,1 %, mientras que el sector sur-poniente presenta el mayor porcentaje de hogares de ingreso bajo con un 48,8 %. Todos los montos nombrados corresponden a valores del año 2012.

### 3.2.2. Sistema de transporte

Transantiago es parte del sistema de transporte público de Santiago de Chile y existe como tal a partir del año 2007. Está presente en gran parte de la región metropolitana. Su estructura se definió inicialmente como troncal-alimentador, que quiere decir que existen servicios locales que alimentan servicios troncales que conectan con otros servicios locales ubicados en otras áreas de la ciudad. Esta estructura fue modificada posteriormente y hoy día es más bien mixta.

Organizacionalmente, el sistema lo componen el Directorio de Transporte Público Metropolitano (DTPM) que funciona como ente regulador, la Administración Financiera de Transantiago (AFT) que se encarga de recaudar, administrar y distribuir los ingresos entre los operadores, el tren subterráneo (Metro), cuyo dueño es el estado de Chile, y siete operadores de buses que corresponden a empresas privadas, donde cada una gestiona una flota que va desde los 400 a los 1.200 vehículos (Gschwender et al., 2016).

El sistema tarifario opera pagando al abordar un bus o metro, y permite realizar un viaje de hasta tres etapas, dentro de una ventana de dos horas, el sistema considera que los usuarios deben hacer transbordos. La tarifa en el metro es un poco más elevada durante las horas punta, mientras que los buses presentan una tarifa fija e igual a la del metro en horario valle.

Transantiago no es el único modo de transporte utilizado en la ciudad, la encuesta EOD del 2012 identifica 9 modos de viajes: Auto, Bip!, Bip!-Otros Público, Bip!-Otros Privado, Taxi Colectivo, Taxi Básico, Caminata, Bicicleta y otros. En la tabla 3.4 se muestra la partición modal por propósito de viaje. Dentro de este conjunto el modo de transporte “Bip!” corresponde a viajes que utilizan el sistema integrado de transporte Transantiago (buses y Metro) para llegar de un lugar a otro. Mientras el modo “Bip!-Otros Público” corresponde a una combinación entre Transantiago y taxi colectivo, taxi básico o transporte inter-urbano. El otro ítem que considera Transantiago es el modo “Bip!-Otros Privado”, que corresponde a aquellos viajes que utilizan Transantiago con una combinación de transporte privado como lo puede ser el auto, bicicleta o caminata. Si se consideran los 3 ítems nombrados anteriormente, se puede observar que Transantiago participa en el 25 % de los viajes de la población, cuando se consideran solo los viajes motorizados ese porcentaje sube a 38,2 %.

Otra información relevante que muestra la tabla 3.4 es que más de la mitad (51,9 %) de los viajes de un día son con el propósito de trabajo o estudio, indicando lo relevante de la ubicación espacial de estos sectores en la ciudad. Además, la mayoría de la gente que utiliza el sistema de transporte público (ítem Bip!, Bip!-Otros público y Bip!-Otros privado) lo hacen con el objetivo de ir a su lugar de trabajo.

Modo de transporte		Propósito			TOTAL
		Trabajo	Estudio	Otro	
Auto	Viajes (miles)	1.980,3	539,6	2.228,7	4.748,6
	%	41,7 %	11,4 %	46,9 %	
Bip!	Viajes (miles)	1.955,0	924,8	1.291,5	4.171,3
	%	46,9 %	22,2 %	31,0 %	
Bip! - Otros Público	Viajes (miles)	178,2	47,0	76,6	301,9
	%	59,0 %	15,6 %	25,4 %	
Bip! - Otros Privado	Viajes (miles)	92,2	27,6	28,4	148,2
	%	62,2 %	18,6 %	19,1 %	
Taxi Colectivo	Viajes (miles)	155,2	53,7	321,5	530,4
	%	29,3 %	10,1 %	60,6 %	
Taxi Básico	Viajes (miles)	73,4	16,0	225,6	315,0
	%	23,3 %	5,1 %	71,6 %	
Otros	Viajes (miles)	354,6	618,0	162,9	1.135,4
	%	31,2 %	54,4 %	14,3 %	
Caminata	Viajes (miles)	865,6	1.298,3	4.199,4	6.363,3
	%	13,6 %	20,4 %	66,0 %	
Bicicleta	Viajes (miles)	329,8	80,6	336,7	747,1
	%	44,1 %	10,8 %	45,1 %	
TOTAL	Viajes (miles)	5.984,2	3.605,6	8.871,3	18.461,1
	%	32,4 %	19,5 %	48,1 %	

Tabla 3.4: Distribución de los viajes según modo de transporte y propósito en día laboral temporada normal. Fuente: eod-2012

Respecto de los medios de transporte de la ciudad, según datos de la EOD del 2012, se estima que existen 1.160.000 vehículos particulares, 6.300 buses urbanos, 27.000 taxis básicos, 11.000 taxis colectivos urbanos y cinco líneas de metro con 104 kilómetros de vías que unen 100 estaciones.

Respecto a la estructura vial de la ciudad, esta cuenta con 68 kilómetros de vías reservadas, 150 kilómetros de vías segregadas, donde existe una división física en la calle que impide que vehículos no públicos puedan transitar por esas vías, y por último, metro aporta 104 kilómetros de rieles que recorren los puntos más demandados de la ciudad.

En relación a aspectos tecnológicos, Transantiago cuenta con un sistema de pago automático de tarifa (AFC) y uno de localización automática de vehículos (AVL). El primer sistema (AFC) es ejecutado a través de tarjetas inteligentes (conocidas comercialmente como tarjeta bip!) con modalidad de prepago que el usuario utiliza para cancelar la tarifa al abordar alguno de los medios de transporte disponibles (bus o metro). Respecto de la cobertura, es el único método de pago disponible en buses y en metro es el más utilizado, ya que la compra de un ticket individual no permite el beneficio del transbordo gratuito. La tecnología AVL consiste en dispositivos GPS (Global Positioning System) que monitorea cada 30 segundos la posición de los vehículos que operan en el sistema de transporte público. Transantiago se caracteriza por tener más de 300 servicios de buses, más de 6.000 buses disponibles y casi 11.200 paraderos de bus.

El sistema de metro funciona dentro de las zonas más demandadas de la ciudad. La figura 3.4 muestra las líneas existentes y su distribución en el espacio: Línea 1 (gris), Línea 2 (amarillo), Línea 4 (azul), línea 4a (celeste) y Línea 5 (verde).

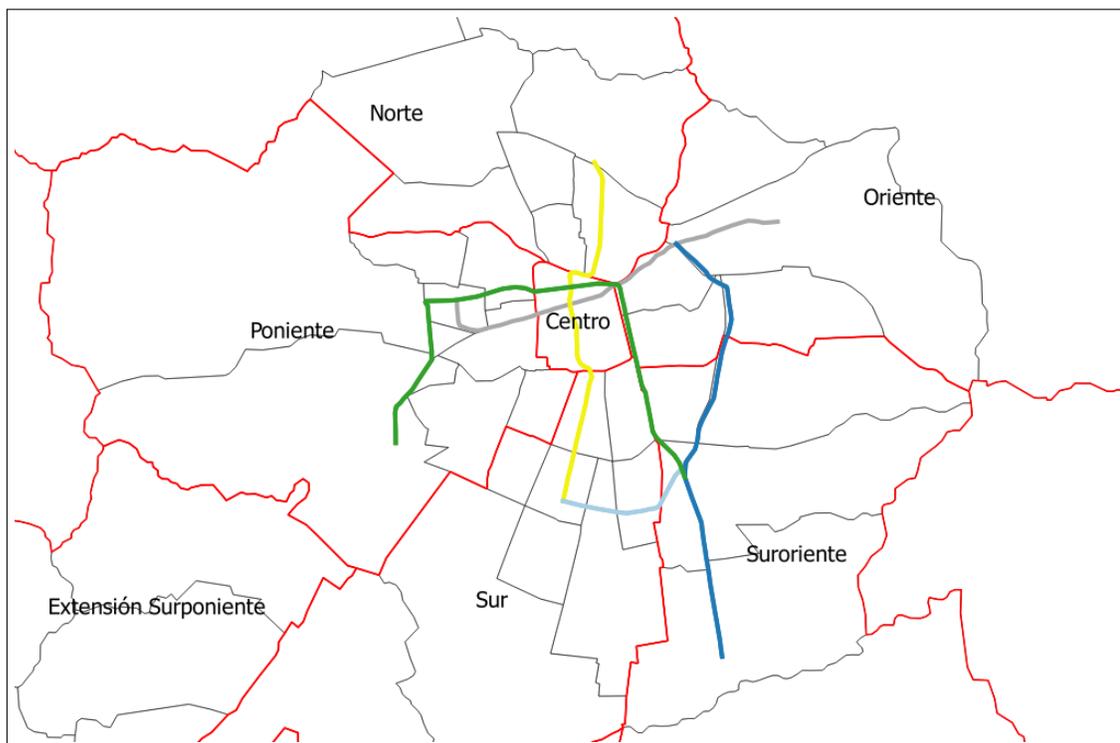


Figura 3.4: líneas de metro. Fuente: elaboración propia

Adicionalmente, en la tabla 3.5 se puede observar la cantidad de estaciones existentes en cada línea, la distancia cubierta, el sentido en el que se mueve en la ciudad y las líneas con las que se puede intercambiar. Respecto a la cobertura del metro, la Línea 5 es la más extensa, recorriendo una distancia total de 30 kilómetros entre el área surponiente hasta el suroriente pasando por el centro, aunque en términos de flujo, es la línea 1 la que presenta mayor demanda.

Línea	Número de estaciones	Distancia	Sentido	Combina con
L1	27	20 Km	Oriente-Poniente	L2, L4, L5
L2	22	20,6 Km	Norte-Sur	L1, L5, L4A
L4	23	24,7 Km	nororiente-suroriente	L1, L5
L4A	6	7,7 Km	Sur-Suroriente	L2, L4
L5	30	30 Km	Surponiente-Suroriente	L1, L2, L4

Tabla 3.5: Caracterización de las líneas de metro. Fuente: [www.metro.cl](http://www.metro.cl)

Si bien Metro es un ente muy relevante dentro de la movilidad de la ciudad, cubre un sector limitado debido a su alto costo. En la figura 3.5 se puede observar la distribución de estaciones en las comunas de la ciudad, donde se ve que la gran presencia de estas se encuentra en la comuna de Santiago. Lo anterior resulta relevante dado que parte de los datos utilizados en este trabajo corresponden a los de metro, dejando ver que estos forman

parte de una fracción de la ciudad y por lo tanto pueden tener un sesgo que se debe tener presente a la hora de analizar.

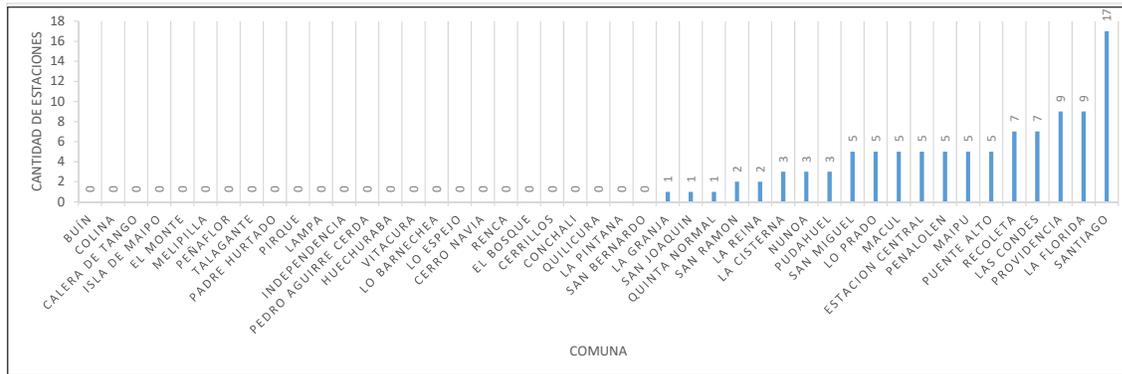


Figura 3.5: Estaciones de metro agrupadas por comuna de ubicación. Fuente: elaboración propia

### 3.2.3. Paraderos de Transantiago

Para abril de 2013, período analizado en esta tesis, existían 11.386 paraderos de buses ubicados en el sector más poblado de la ciudad. En la figura 3.6 se puede observar su distribución espacial, mostrando una alta densidad en el centro que luego va disminuyendo hacia la periferia. Además se puede observar que la cobertura no es completa ni uniforme a lo largo de la ciudad pero es superior a la entregada por metro. Lo anterior resulta relevante dado que los datos a utilizar en esta tesis están limitados al espacio que cubren los paraderos, por lo que el análisis de este trabajo se hará sobre esa superficie cubierta, que corresponde a 34 comunas de la ciudad (ver tabla 3.7).

Las comunas no cubiertas corresponden a localidades que se encuentran en la periferia de la ciudad y por el momento su servicio de transporte está cubierto por empresas de buses inter-urbanos, fuente de datos no disponible para este trabajo.

### 3.2.4. Caracterización espacial de la ciudad

Espacialmente, se puede describir a Santiago en una zonificación agregada de siete grandes sectores, los cuales han sido definidos por SECTRA según criterios socio-económicos y son ocupados para diversos análisis, entre ellos, la encuesta origen-destino del año 2012. Su denominación está dada por la zona geográfica que abarcan, es así que existen los sectores: Norte, Oriente, Suroriente, Centro, Poniente, Sur, extensión Surponiente.

Un nivel de zonificación con mayor detalle corresponde al de comuna. En la figura 3.8 se muestra la distribución espacial de las comunas y sectores. El borde rojo representa los límites que definen los sectores, mientras que las líneas negras definen los límites geográficos de cada comuna, que a su vez, corresponden a un límite político-administrativo. La comuna

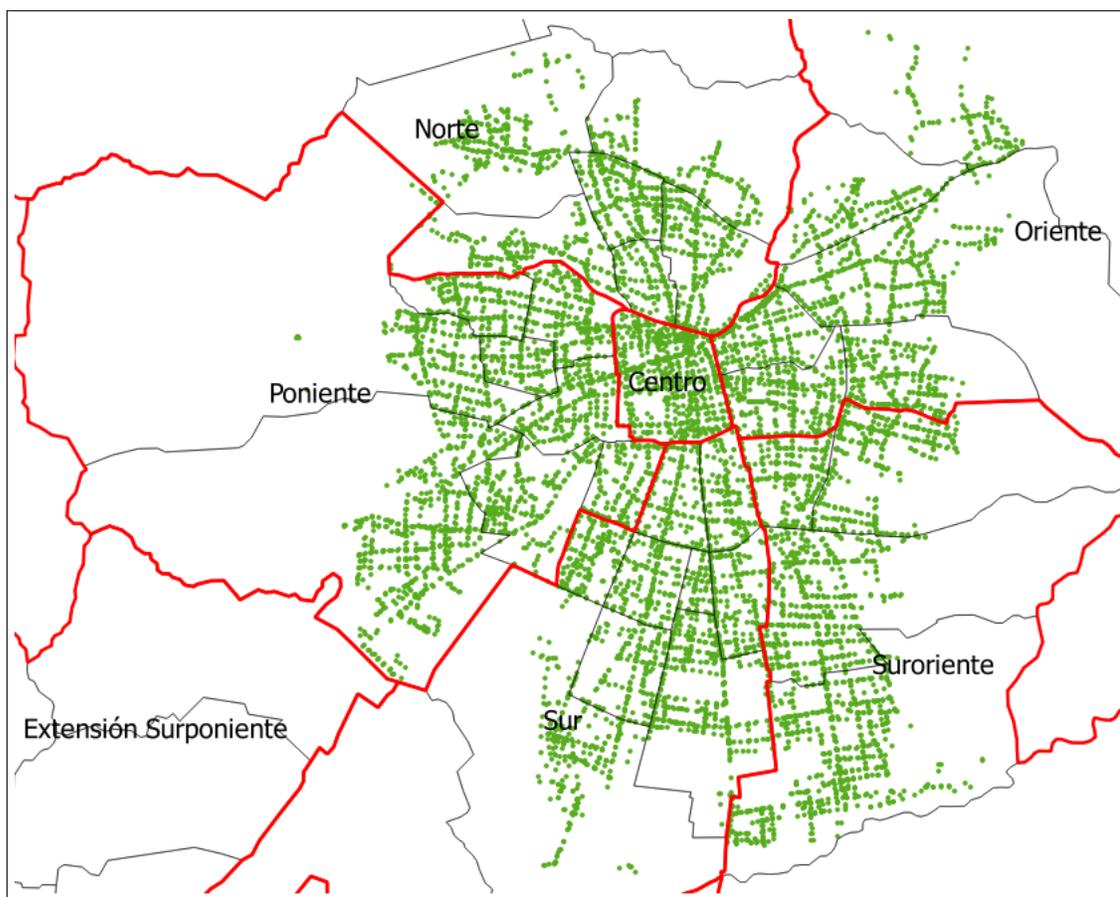


Figura 3.6: Ubicación de paraderos de Transantiago. Fuente: elaboración propia

de Santiago (ubicada en el centro de la ciudad), al presentar alta densidad en actividades, viajes y alta diversidad de población, es definida en si misma como un sector (Centro).

Desde el punto de vista de comunas, lo que se conoce como el gran Santiago, corresponde a 45 comunas de la región metropolitana: 32 comunas perteneciente a la provincia de Santiago, 5 comunas perteneciente a la provincia de Talagante y las comunas de San Bernardo, Calera de Tango, Puente Alto, Pirque, Colina, Lampa, Buin y Melipilla.

En la tabla 3.6 se muestra la estructura de comunas, provincias y sectores que forman parte de lo que se conoce como el Gran Santiago. Se destacan en azul aquellas donde hay presencia de paraderos de Transantiago y su cantidad entre paréntesis.

Una visión más desagregada que el esquema de comunas se puede hacer mediante la división por zonas. En Santiago suele utilizarse una división de 866 zonas cuya definición fue realizada por SECTRA (oficina de planificación de transporte) a través de criterios económicos y sociales. En la figura 3.9 se muestra esta zonificación llamada 777, donde la parte central presenta áreas más pequeñas que la periferia debido a que las densidades de las variables consideradas para la partición son más altas en el centro.

Una característica común de estos tres tipos de zonificación es su compatibilidad, dado que un sector está compuesto por un grupo de comunas, que a su vez, cada comuna está

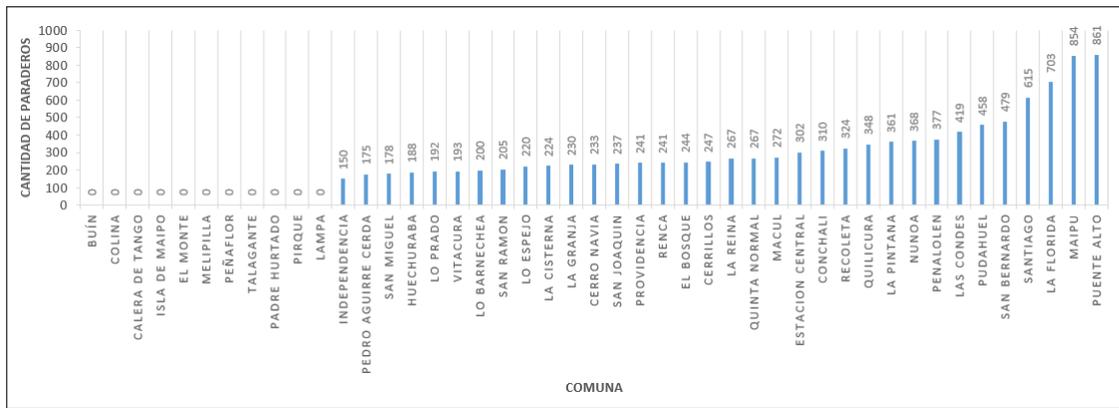


Figura 3.7: Cantidad de paraderos por comuna de Santiago. Fuente: elaboración propia

formada por un grupo de zonas, lo que permite agrupar o desagregar indicadores presentes en estos espacios geográficos.

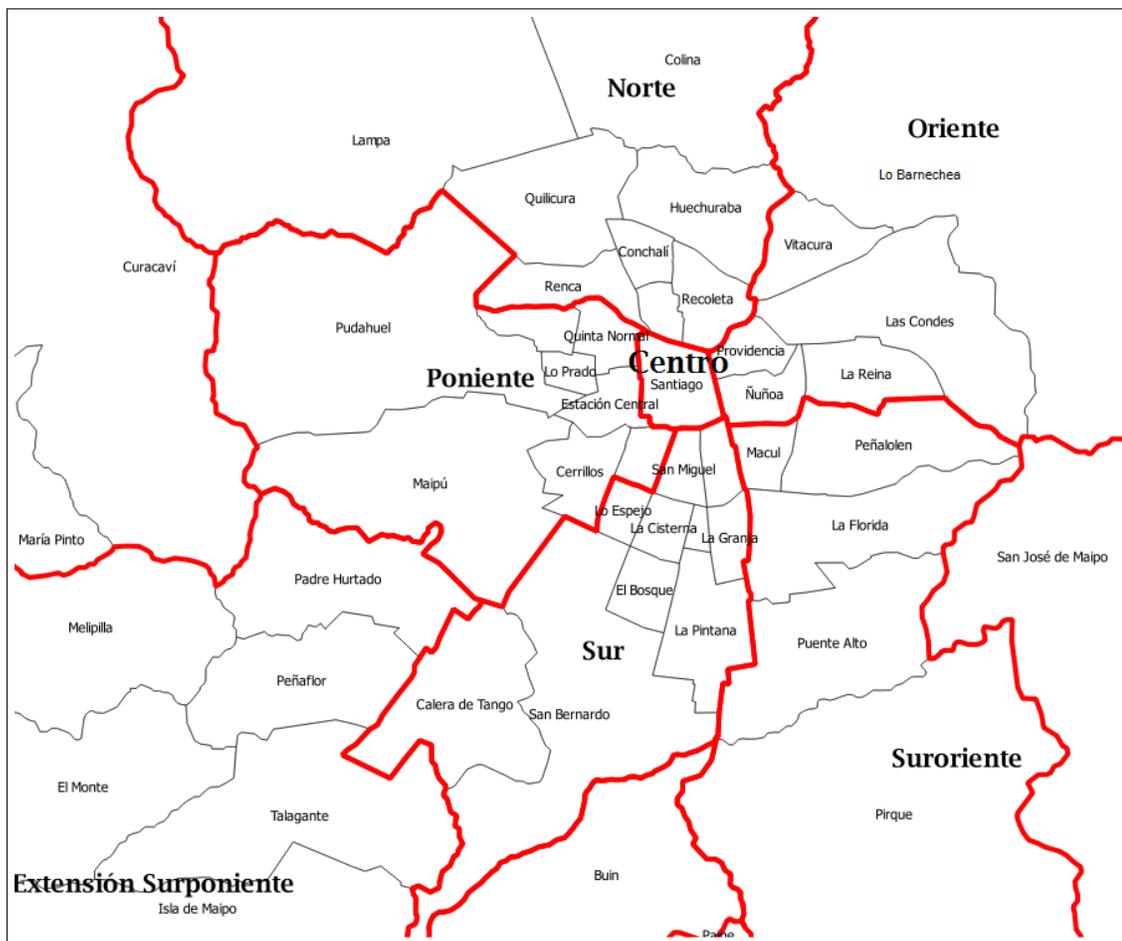


Figura 3.8: Sectores (borde rojo) y comunas (borde negro) de la ciudad de Santiago. Fuente: EOD-2012

Gran Santiago				
Sector	Provincia de Santiago		Provincia de Talagante	Otra provincia
Centro	Santiago (615)			
Norte	Independencia (150) Renca (241) Recoleta (324)	Huechuraba (188) Conchalí (310) Quilicura (348)		Lampa Colina
Oriente	Vitacura (193) Providencia (241) Ñuñoa (368)	Lo Barnechea (200) La Reina (267) Las Condes (419)		
Poniente	Pedro Aguirre Cerda (175) Cerro Navia (233) Quina Normal (267) Pudahuel (458)	Lo Prado (192) Cerrillos (247) Estación Central (302) Maipú (854)		
Sur	San Miguel (178) Lo Espejo (220) La Granja (230) El Bosque (244)	San Ramón (205) La Cisterna (224) San Joaquín (237) La Pintana (361)		San Bernardo (479) Calera de Tango
Suroriente	Macul (272) La Florida (703)	Peñalolén (377)		Puente Alto (861) Pirque
Extensión Surponiente			Isla de Maipo El Monte Padre Hurtado Peñaflor Talagante	Buín Melipilla

Tabla 3.6: Comunas de la región metropolitana agrupadas por sector y provincia. Fuente: EOD-2012

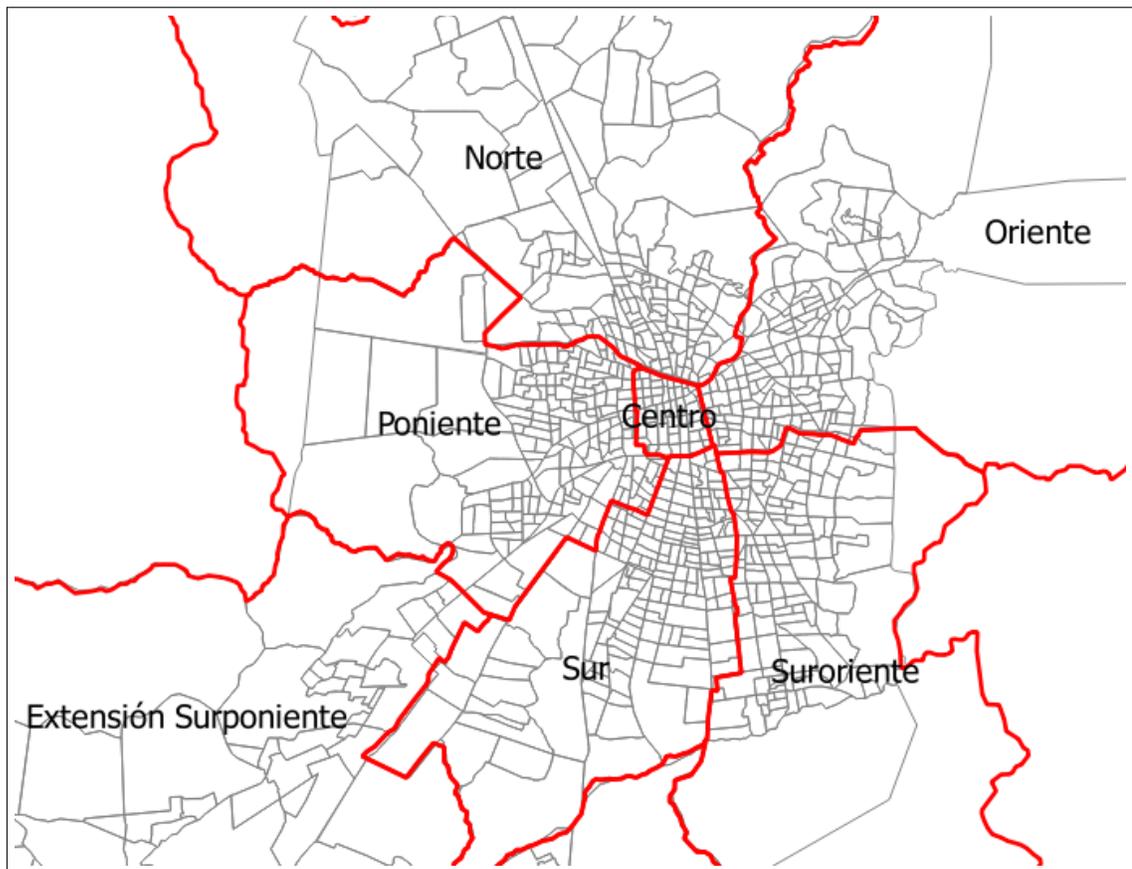


Figura 3.9: zonificación de Santiago, conocida como zonificación 777. Fuente: EOD-2012

# Capítulo 4

## Implementación

Este capítulo tiene por objetivo describir detalladamente el proceso de construcción de los indicadores *Page rank*, *Betweenness* e *Infomap* además de las propiedades generales de la red. La figura 4.1 muestra el proceso de manera general. Allí se puede observar que primero se hace uso de los datos de viajes para generar el grafo de movilidad, donde los nodos son los paraderos de buses y los arcos representan la posibilidad de un viaje entre un par de paradas. Con el grafo de movilidad se calculan las métricas requeridas que están asociadas a cada nodo, luego se utilizan los datos de paradas para conocer la posición de cada nodo en el espacio y se realiza una interpolación espacial para obtener una métrica continua a lo largo de toda la ciudad. por último se complementa los resultados con la información geográfica de Santiago de Chile (zonificaciones). A continuación se describe con mayor detalle cada parte del proceso.

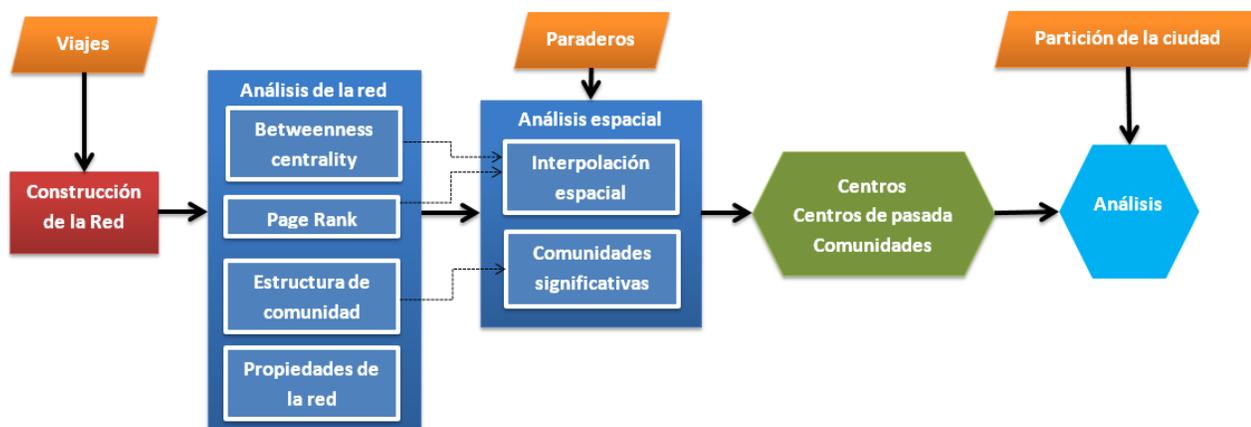


Figura 4.1: Proceso de construcción de indicadores. Fuente: elaboración propia

### 4.1. Herramientas

Las tareas descritas fueron realizadas con las siguientes herramientas:

- PostgreSQL<sup>1</sup>: Administrador de base de datos relacional para la manipulación de los datos
- PostGIS<sup>2</sup>: Módulo para PostgreSQL que añade soporte de objetos geográficos. Usado para asignar la métrica de comunidad a cada zona de la zonificación 777.
- Igraph<sup>3</sup>: Herramienta para análisis de redes. Usado para el cálculo de indicadores.
- MapEquation<sup>4</sup>: Framework para detección de comunidades basado en flujos, implementa el algoritmo *Infomap*, usado para la generación de comunidades.
- QGIS<sup>5</sup>: Sistema de información geográfica para llevar a cabo el análisis espacial, que involucra la interpolación espacial y la selección de las comunidades más significativas. La versión utilizada es la 2.8.2.
- Colorbrewer<sup>6</sup>: Página web que recomienda paleta de colores para visualización de mapas. Usado para la selección de colores usados en las vistas de *PageRank* y *Betweenness*.
- BASH: Lenguaje de programación utilizado para automatizar la construcción de los indicadores. Su elección se justifica por sus comandos para la manipulación de archivos en entornos Linux.

## 4.2. Procesamiento de datos

### 4.2.1. Viajes

Los datos de viajes son obtenidos a partir de una tabla alojada en una base de datos PostgreSQL, donde cada fila representa un viaje en transporte público. Las columnas presentes en ella pueden ser vistas en la tabla 7.1 del anexo A. Las columnas utilizadas de la tabla de viajes corresponden a las descritas en la tabla 4.1.

Con los datos de la tabla de viajes es posible rescatar los hitos que se requieren del viaje, que corresponden al momento en que inicia junto con los paraderos del que hace uso.

### 4.2.2. Paradas

La otra fuente de datos, que corresponde a las paradas, se obtiene de dos tablas, **red-paradas** y **estaciones\_metro**. La primera contiene las paradas presentes en el transporte público de superficie, mientras que la segunda contiene las estaciones subterráneas de metro. En la tabla 4.2 y 4.3 se pueden ver los datos utilizados.

Notar que en la tabla 4.3 se utiliza la columna *codigosinlinea* para dos fines distintos,

---

<sup>1</sup><https://www.postgresql.org/>

<sup>2</sup><https://www.postgis.net/>

<sup>3</sup><http://igraph.org/>

<sup>4</sup><http://www.mapequation.org/>

<sup>5</sup><http://www.qgis.org/es/site/>

<sup>6</sup><http://colorbrewer2.org/>

Nombre columna	Descripción
id	Identificador de tarjeta bip! encriptada
nviaje	Número correlativo de viaje para un id.
netapa	Número de etapas presente en el viaje.
par_subida_1er	Paradero de inicio del viaje.
par_bajada_1er	Paradero de fin del viaje.
tiemposubida_1er	Marca temporal en que el usuario subió a un transporte público al inicio del viaje.
par_subida_2da	Paradero inicial de la segunda etapa del viaje.
par_bajada_2da	Paradero final de la segunda etapa del viaje.
par_subida_3era	Paradero inicial de la tercera etapa del viaje.
par_bajada_3era	Paradero final de la tercera etapa del viaje.
par_subida_4ta	Paradero inicial de la cuarta etapa del viaje.
par_bajada_4ta	Paradero final de la cuarta etapa del viaje.
factor_expansion	cantidad de usuarios que representa el viaje.

Tabla 4.1: Columnas utilizadas de la tabla de viajes. Fuente: elaboración propia

Nombre columna	Descripción
codigo	Identificador de parada.
nombre	Nombre de la parada.
latitud	Latitud de la parada.
longitud	Longitud de la parada.
comuna	Comuna donde se encuentra la parada.
zona777	Zona donde se encuentra la parada.

Tabla 4.2: Columnas utilizadas de la tabla de redparadas. Fuente: elaboración propia

como identificador de estación y a su vez como nombre, esto con el fin de hacer coincidir la cantidad de columnas con la tabla 4.2 dado que luego los datos serán puestos en una sola tabla. El detalle completo de las tablas **redparadas** y **estaciones\_metro** puede ser visto en el anexo B.

### 4.2.3. Geografía

La información geográfica de la ciudad fue obtenida en el departamento de transporte de la universidad de Chile<sup>7</sup>. Los datos corresponden a la zonificación 777 en formato *shape*. Aparte de las geometrías de cada zona contiene atributos que permiten obtener los otros niveles de agrupación, como lo son las comunas y sectores. En la tabla 4.4 se muestran todos los atributos presentes.

Este archivo será utilizado para dar contexto a las vistas generadas de los indicadores calculados en esta tesis, por lo que el valor de esta fuente de datos radica en que permite

<sup>7</sup>Actualmente este archivo puede ser encontrado en el siguiente link: <https://github.com/felipinbombin/eod2012>

Nombre columna	Descripción
codigosinlinea	Identificador de estación.
codigosinlinea	Nombre de la estación.
latitud	Latitud de la parada.
longitud	Longitud de la parada.
comuna	Comuna donde se encuentra la parada.
zona777	Zona donde se encuentra la parada.

Tabla 4.3: Columnas utilizadas de la tabla de estaciones\_metro. Fuente: elaboración propia

columna	Descripción
id	Identificador de figura
área	Metros cuadrados que cubre la zona.
zona	Identificador de zona 777
comuna	Comuna donde se encuentra la zona

Tabla 4.4: Atributos de archivo *shape* con la zonificación 777. Fuente: elaboración propia

visualizar los límites físicos para los tres tipos de zonificación (sectores, comunas y zonas). Importante es notar que la asociación de las paradas y estaciones de metro a cada área (comuna y zona 777) ya se encuentra almacenada en la tabla **redparadas** y **estaciones\_metro**.

#### 4.2.4. Filtros

El proceso de filtros corresponde a la labor de seleccionar aquellos datos que cumplen con una serie de reglas que permiten llevar adelante sin problema los procesos posteriores. Cada filtro depende del tipo de dato a considerar y su importancia dentro del proceso. A continuación se describen las reglas impuestas para cada tabla de datos.

La tabla **viajes** es usada para la generación de grafos dirigidos con pesos donde cada viaje aporta 1 al peso de un arco entre el paradero inicial y final del trayecto (nodos en el grafo), por lo que una condición que debe cumplir es que tanto el paradero de inicio como de fin de viaje debe existir. Además, para los análisis se requiere tener la hora de inicio, por lo que este dato (al igual que los descritos anteriormente) debe ser distinto de vacío. Otro requisito es reconstruir el viaje completo, ya que se construirá otros grafos que toman en cuenta las etapas (no solo el origen y destino).

Para los datos de paradas y estaciones de metro se exige que todos los campos nombrados en las tablas 4.2 y 4.3 sean distintos de vacío.

La última condición exigida para los viajes, es que el tiempo de subida de una etapa debe ser superior al tiempo de bajada de la etapa anterior, esta inconsistencia en los datos se produce en el proceso de estimación de bajada (Munizaga and Palma, 2012) debido a que la asignación de bajada de cada etapa se realiza antes de la unión de estas para generar el viaje.

De lo descrito anteriormente se desprende la siguiente implementación para la etapa de filtrado<sup>8</sup>:

- Se genera una tabla para almacenar los viajes y otra para paradas (buses y metro) que cumplan con el filtro
- Se insertan los viajes que cumplen con las siguientes condiciones:
  - campos **id**, **nviaje**, **netapa**, **tiempo\_subida\_1er**, **par\_subida\_1er**, **par\_bajada\_1er**, **par\_subida\_2da**, **par\_bajada\_2da**, **par\_subida\_3era**, **par\_bajada\_3era**, **par\_subida\_4ta**, **par\_bajada\_4ta**, **factor\_expansion** distinto de vacío.
  - campo **tiempobajada\_1er** debe ser más antiguo que el campo **tiemposubida\_2da**
  - campo **tiempobajada\_2do** debe ser más antiguo que el campo **tiemposubida\_3era**
  - campo **tiempobajada\_3era** debe ser más antiguo que el campo **tiemposubida\_4ta**
- Se insertan las paradas de bus que cumplen con la siguiente condición: campos **codigo**, **nombre**, **latitud**, **longitud**, **comuna**, **zona777** distinto de vacío.
- Se insertan las estaciones de metro en la misma tabla que las paradas de buses y que cumplen la siguiente condición: campos **codigosinlinea**, **latitud**, **longitud**, **comuna**, **zona777** distinto de vacío.

En la tabla 4.5 se muestra el número de registros eliminados y los que quedaron para cada tabla de datos.

Fuente	Datos iniciales	Datos eliminados	Datos que cumplen con el filtro
Viajes	25.244.788	6.745.145	18.499.643
Estaciones de metro	112	0	112
Paraderos	11.183	0	11.183

Tabla 4.5: Viajes iniciales y viajes eliminados en el filtro. Fuente: elaboración propia

#### 4.2.5. Normalización

Esta etapa tiene la tarea de unificar convenciones usadas en las distintas fuentes de datos con el objetivo de usar un lenguaje común entre todas ellas.

La estructura inherente del sistema de metro introduce un problema, y es que contiene estaciones comunes para a lo más dos líneas (estación Baquedano, Los Heroes, Tobalaba, La cisterna, San Pablo, Santa Ana y Vicente Valdes), lo que permite que un usuario pueda realizar una combinación de una a otra sin tener que salir de la red. Para diferenciar por cual línea llegó o partió un usuario (proceso de generación de viajes) en la parada de inicio o fin de una etapa se agrega un sufijo que lo informa, por ejemplo “Baquedano\_L1” indica que el

<sup>8</sup>El significado de cada campo está descrito en la tabla del anexo A

usuario comenzó la etapa en la estación Baquedano por desde el acceso por la línea uno. Lo anterior es un problema puesto que para la generación del grafo de viajes “Baquedano.L1” y “Baquedano.L5” representan un solo nodo, por lo que se debe realizar la unión de todas aquellas situaciones. El criterio utilizado fue quitar los sufijos “L1” y “L5”.

Otro problema presente en los datos de metro tiene que ver con los códigos de las estaciones. En los datos de viaje y etapas se presentan códigos para una estación que no existen en la tabla **estaciones\_metro**, por lo que se debe realizar el match entre estos con el fin de poder asignar la localización espacial de esos nodos. El procedimiento consistió en revisar caso a caso las diferencias, lo que concluyó con 35 registros modificados de la tabla **estaciones\_metro** y la adición de tres nuevos registros que existen en la tabla de viajes y etapas pero no en la tabla **estaciones\_metro**.

Juntando el proceso de filtrado y normalización se obtiene el diagrama expuesto en la figura 4.2. El proceso inicia cuando se crea la base de datos (1) junto con las tablas que contienen los datos crudos (2), luego se insertan estos en las tablas (3) para luego ser filtrados (4) y normalizados (5). El resultado de lo anterior es almacenado en dos tablas: **viaje\_util** y **parada\_util** (6).

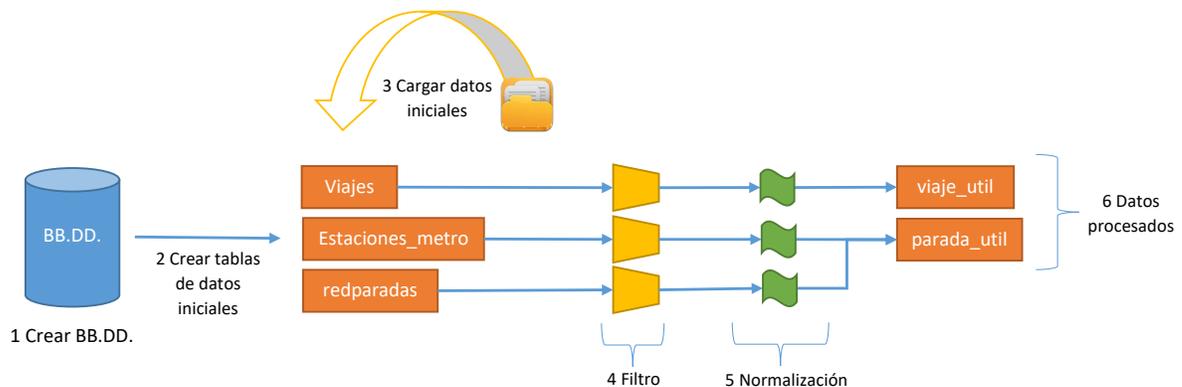


Figura 4.2: Proceso de filtrado y normalización de datos. Fuente: elaboración propia

### 4.3. Construcción de la red de movilidad

Una vez obtenidos los datos procesados es necesario construir la red de movilidad. Esta red corresponde a un grafo dirigido con pesos donde los nodos representan paradas de bus o estaciones metro, los arcos la posibilidad de viaje entre un par de paradas y su peso corresponde al volumen de viajes realizados en la ventana de tiempo usada.

Existen varias formas de representar un grafo<sup>9</sup>, el criterio que debe primar a la hora de elegir uno es su compatibilidad con las herramientas de procesamiento elegidas.

Las herramientas que recibirán como entrada el grafo de viaje serán dos: *Igraph*, herramienta para realizar el cálculo de métricas en grafos y *mapequation* para el cálculo de las

<sup>9</sup>Para mayor detalle visitar <http://igraph.org/python/doc/igraph.Graph-class.html#Read>

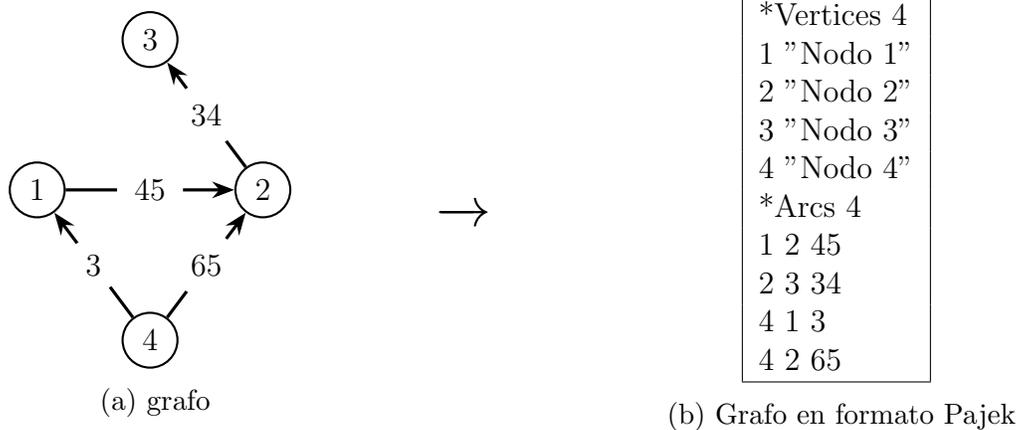


Figura 4.3: Red descrita usando el formato Pajek. Fuente: elaboración propia

comunidades con el algoritmo *Infomap*.

*Igraph* acepta los siguientes formatos: NCOL, LGL, graphdb, graphml, graphmlz, gml, net, pajek, dimacs, edgelist, edges o edge (lista de arcos), matriz de adyacencia, objetos serializables de python. Mientras que *mapequation* permite los siguientes: pajek, link-list, 3gram, multiplex or bipartite. Al observar las limitaciones se puede apreciar que el formato pajek cumple con las restricción de ser aceptado por ambas herramientas.

El formato pajek corresponde a un archivo de texto plano con extensión “.net” que primero define una lista de nodos, estableciendo un identificador y atributos adicionales (separados por un espacio), luego describe una lista de pares de identificadores de nodo para representar los arcos, atributos adicionales sobre los arcos pueden se agregados a la derecha de los pares de identificadores<sup>10</sup>. En la figura 4.3 se puede observar un grafo (izquierda) y su representación en formato pajek (derecha) para una red de 4 nodos y 4 arcos. En el archivo primero se define un encabezado que indica que las siguientes líneas describen los nodos del grafo y su cantidad, las siguientes cuatro líneas contienen la información de cada nodo, que corresponde a un identificador y una etiqueta, la línea 5 indica que las siguientes líneas expresan arcos con dirección (palabra “Arcs”) y el número de arcos. Por último, las 4 líneas siguientes definen los arcos señalando el nodo de origen (primer número), nodo de destino (segundo número) y el peso asociado (tercer número).

Una vez establecido el formato con que se expresará la red, el paso siguiente es convertir los datos presentes en la tabla **viaje\_util** de Postgres a la descripción Pajek. En la figura 4.4 se puede observar el proceso de transformación desde las tablas en Postgres al formato Pajek. Primero se realiza una consulta SQL con la que se construye un archivo intermedio con formato CSV, cuyo formato contiene 3 columnas: parada inicial de viaje, parada final de viaje y número de viajes realizados entre la parada inicial y final.

Una vez construido el archivo CSV se realiza el segundo procedimiento descrito en la figura 4.4, que consiste en generar el formato Pajek a partir del CSV de viajes. Para lo anterior se escribió un script en Python que recorre el CSV por fila, donde se extrae el paradero

<sup>10</sup>Una descripción más completa pueden ser encontrada en la documentación oficial: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf>

de subida y de bajada para asignarles un índice, si a estos ya se les asignó uno en alguna iteración previa, se rescata y se guarda la información del arco, que corresponde al índice del paradero de subida, índice de paradero de bajada y el volumen de viajes. Para más detalle consultar el anexo C.

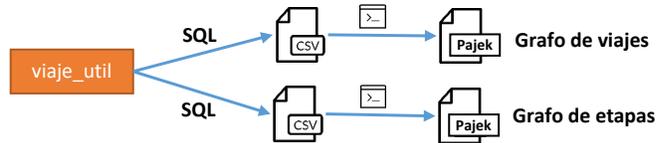


Figura 4.4: Proceso de transformación desde la base de datos a formato pajek. Fuente: elaboración propia

## 4.4. Cálculo de indicadores

Para el cálculo de los indicadores se creó un archivo *bash*, con el objetivo de ser manejados de manera independiente. Lo anterior tiene el beneficio de ser modular, permitiendo agregar alguna métrica nueva fácilmente a partir de alguna ya creada.

### 4.4.1. PageRank

El *PageRank* fue calculado utilizando la librería *Igraph*, que permite cargar un grafo en memoria a partir de un archivo con formato Pajek. El método utilizado es *igraph\_pagerank* y su definición es la siguiente:

```
int igraph_pagerank(const igraph_t *graph,
                  igraph_pagerank_algo_t algo,
                  igraph_vector_t *vector,
                  igraph_real_t *value,
                  const igraph_vs_t vids,
                  igraph_bool_t directed,
                  igraph_real_t damping,
                  const igraph_vector_t *weights,
                  void *options);
```

Los parámetros requeridos por el método `igraph_pagerank` son:

**graph:** grafo de viajes

**algo:** la implementación de PageRank a utilizar. Se usa la recomendada por *Igraph*, que corresponde a la opción `IGRAPH_PAGERANK_ALGO_PRPACK`. La diferencia entre uno y otro es la implementación, y en consecuencia, su rendimiento.

**vector:** Puntero a un vector que almacenará el resultado.

- value:** valor propio del vector que contiene el *PageRank*, debería ser siempre igual a 1.
- vids:** Identificadores de nodo para los cuales se debería retornar su *PageRank*. En este caso debe contener todos los nodos.
- directed:** boolean que indica si se considera la dirección de los arcos. Es ignorado para grafos sin dirección. Este valor es establecido como `TRUE`.
- damping:** damping factor, se establece en 0,85 para coincidir con la implementación de Zhong et al. (2014).
- weights:** Vector con los pesos de los arcos.
- options:** Las opciones disponibles dependen de la implementación usada. Para el cálculo que realizamos este parámetro es `NULL`.

Un hecho a tener en cuenta al momento de configurar la ejecución de la métrica con *Igraph* es la creación de un vector de pesos a partir de los pesos en el grafo ya que el método *igraph\_pagerank* lo recibe como un parámetro adicional. Por último, el archivo de salida generado por el procedimiento corresponde a un archivo CSV de dos columnas: la primera contiene el identificador de parada o estación de bus y la segunda el valor del *PageRank*.

#### 4.4.2. Betweennees

Esta métrica es construida en dos formatos: el primero considerando considerando el cálculo teórico, que está basada en las rutas más cortas entre cada par de nodos tomando en cuenta el peso de estos. El segundo cálculo se realiza reemplazando la concepción de ruta más corta por una ruta entre cada par de nodos presente en los datos.

El cálculo teórico del *betweenness* se construye con el método `igraph_betweenness` de la librería *Igraph*, que se define como sigue:

```
int igraph_betweenness(const igraph_t *graph,
                      igraph_vector_t *res,
                      const igraph_vs_t vids,
                      igraph_bool_t directed,
                      const igraph_vector_t* weights,
                      igraph_bool_t nobigint);
```

Los parámetros requeridos por el método `igraph_betweenness` son:

- graph:** Grafo de etapas.
- res:** vector de resultados.
- directed:** Boolean que indica si se considera la dirección de los arcos. Es ignorado para grafos sin dirección. Este valor es establecido como `TRUE`.

**weights:** Vector con los pesos de los arcos.

**nobigint:** Boolean, si es TRUE no usa enteros grandes para el cálculo.

Una observación importante sobre la ejecución de este método es que en *Igraph* los pesos son considerados como un costo, por lo que al momento de calcular la ruta más corta el algoritmo minimiza ese costo, lo que es un problema puesto que se quiere el efecto opuesto, es decir, maximizar el costo, ya que consideraremos una ruta más corta como aquella que contenga la mayor cantidad de viajes. Según lo anterior es que el vector de pesos debe ser invertido, lo que se realiza obteniendo el máximo peso del vector para luego restarlo con cada elemento del vector de pesos, de esta forma el máximo pasa a ser mínimo y viceversa.

El segundo cálculo del indicador se realiza con un código escrito en Python, donde la lógica es parecida al indicador teórico descrito anteriormente, pero en esta ocasión los viajes más cortos son reemplazados por los viajes presentes en los datos y para un nodo cualquiera, se considera la cantidad de veces que ese nodo participa en los viajes presentes en los datos.

Por último, el archivo generado por el procedimiento corresponde a un archivo CSV de dos columnas: la primera contiene el identificador de parada o estación de bus y la segunda el valor *Betweenness*.

### 4.4.3. Infomap

El cálculo de las comunidades presentes se realiza utilizando el framework *mapequation* que implementa el algoritmo *Infomap*. Para la ejecución del algoritmo se usan los siguientes parámetros:

- i ‘pajek’: indica que el formato del grafo recibido como entrada está en formato Pajek.
- two-level: Optimiza una partición de dos niveles en la red (comunidades y nodos).
- code-rate: el radio de codificación se refiere a la cantidad de pasos que realiza el peatón aleatorio antes de realizar la codificación. El valor usado acá (1.261595) fue obtenido por medio de pruebas de ensayo y error, hasta obtener la menor cantidad de comunidades.
- num-trials: Número de cálculos realizados antes de retornar. Se utiliza un parámetro de 1000, retornando la mejor optimización.
- directed: Indica que los arcos del grafo tienen dirección.
- undirdir: Considera la red sin dirección para definir el flujo y con dirección para minimizar el código de descripción de un recorrido aleatorio.

Adicionalmente a los parámetros descritos anteriormente es necesario indicar la ruta del archivo Pajek que describe el grafo de viajes junto con la ruta del archivo de salida.

El archivo de salida utiliza un formato “.tree”, que corresponde a una descripción de los

módulos asignados más la información del nodo (nombre e identificador de nodo dado en el archivo Pajek). En la figura 4.5 se muestra un ejemplo del formato `.tree`<sup>11</sup>.

Archivo <code>.tree</code>
...
1:1 0.0156089 "TOBALABA" 3125
1:2 0.0143866 "UNIVERSIDAD DE CHILE" 3130
1:3 0.0127237 "PLAZA DE ARMAS" 1916
...

Figura 4.5: Ejemplo de formato de archivo con extensión `.tree`. Fuente: elaboración propia

#### 4.4.4. Propiedades grafo

Las propiedades del grafo son calculadas con la librería *Igraph*. Se seleccionaron tres propiedades, de las cuales existen variantes que en total generan ocho métricas.

##### 4.4.4.1. Grado de nodo

El grado de un nodo se refiere a la cantidad de arcos en los que inciden en él; en un grafo dirigido existen 3 variantes de esta métrica: arcos de incidencia, arcos que nacen en el nodo o ambos. *Igraph* provee la función `igraph_degree` que ofrece una opción para calcular las variantes mencionadas.

```
int igraph_degree(const igraph_t *graph,
                 igraph_vector_t *res,
                 const igraph_vs_t vids,
                 igraph_neimode_t mode,
                 igraph_bool_t loops);}
```

Los parámetros que recibe la función son:

**graph** : grafo de viajes.

**res** : vector que almacenará los resultados.

**vids** : vector con los identificadores de nodo a ser calculados. En nuestro caso son todos.

**mod** : tipo de grado, `IGRAPH_IN` para considerar los arcos de incidencia, `IGRAPH_OUT` para arcos que se originan en el nodo y `IGRAPH_ALL` que considera todos los arcos.

**loops** : valor booleano, `TRUE` si se deben considerar arcos que tienen el mismo nodo de inicio y fin. La opción utilizada para los cálculos es `FALSE`.

---

<sup>11</sup>Para más información visitar: <http://www.mapequation.org/code.html#Tree-format>

#### 4.4.4.2. Intensidad (*Strength*)

El método que define la intensidad de un nodo está dado por la función `IGRAPH_STRENGTH`, la que realiza un cálculo similar a `IGRAPH_DEGREE` salvo que ahora los arcos son ponderados por su peso, lo que corresponde a sumar los pesos de los arcos. Al igual que `IGRAPH_DEGREE`, existen 3 formas de hacerlo y los parámetros requeridos son los siguientes:

```
int igraph_strength(const igraph_t *graph,
                   igraph_vector_t *res,
                   const igraph_vs_t vids,
                   igraph_neimode_t mode,
                   igraph_bool_t loops,
                   const igraph_vector_t *weights)
```

**graph:** grafo de viajes.

**res:** vector que almacenará los resultados.

**vids:** vector con los identificadores de nodo a ser calculados. En nuestro caso son todos.

**mode:** tipo de arco a considerar, `IGRAPH_IN` para considerar los arcos de incidencia, `IGRAPH_OUT` para arcos que se originan en el nodo y `IGRAPH_ALL` para considerar todos los arcos.

**loops:** valor booleano, `TRUE` si se deben considerar arcos que tienen el mismo nodo de inicio y fin. La opción utilizada para los cálculos es `FALSE`.

**weights:** vector con los pesos de cada arco.

#### 4.4.4.3. Promedio de rutas más cortas (*Shortest path length*)

El método que define el promedio de rutas más cortas está dado por la función `IGRAPH_AVERAGE_PATH_LENGTH`, los parámetros requeridos son los siguientes:

```
float igraph_average_path_length(const igraph_t *graph,
                                igraph_bool_t directed,
                                igraph_bool_t unconn)
```

**graph:** grafo de viajes.

**directed:** valor booleano, `TRUE` si se debe considerar la dirección de los arcos en caso de un grafo dirigido. La opción utilizada para los cálculos es `TRUE`.

**unconn:** valor booleano, indica que hacer cuando el grafo no es conexo. Si es `TRUE` el promedio de las rutas promedio de cada componente conexas es devuelto cuando el grafo no es conexo. La opción utilizada para los cálculos es `TRUE`.

#### 4.4.4.4. Coeficiente de agrupamiento (*Clustering coefficient*)

El coeficiente de agrupamiento es dado por la función `IGRAPH_TRANSITIVITY_UNDIRECTED`, este calcula la probabilidad que dos vecinos de un vértice estén conectados. Los parámetros requeridos son los siguientes:

```
float igraph_transitivity_undirected(const igraph_t *graph,
                                     igraph_bool_t mode)
```

**graph:** grafo de viajes.

**mode:** `TRANSITIVITY_ZERO` retorna cero si el grafo no tiene ningún triplete y `TRANSITIVITY_NAN` retorna `NULL` en el mismo caso. La opción utilizada para los cálculos es `TRANSITIVITY_ZERO`.

#### 4.4.4.5. Diámetro

El método que define el diámetro está dado por la función `IGRAPH_DIAMETER`, calcula la ruta más larga de los caminos más cortos entre todos los pares de nodos en la red. Los parámetros requeridos son los siguientes:

```
float igraph_diameter(const igraph_t *graph,
                      igraph_bool_t directed,
                      igraph_bool_t unconn,
                      igraph_bool_t weights)
```

**graph:** grafo de viajes.

**directed:** valor booleano, `TRUE` si se debe considerar la dirección de los arcos en caso de un grafo dirigido. La opción utilizada para los cálculos es `TRUE`.

**unconn:** valor booleano, indica que hacer cuando el grafo no es conexo. Si es `TRUE` el máximo de los diámetros de cada componente conexa es devuelto cuando el grafo no es conexo. La opción utilizada para los cálculos es `TRUE`.

**weights:** vector con los pesos de los arcos. El valor utilizado es `NULL`.

#### 4.4.4.6. Componentes conexas

El método que define el diámetro está dado por la función `IGRAPH_COMPONENTS`, calcula las componentes conexas de un grafo. Los parámetros requeridos son los siguientes:

```
float igraph_components(const igraph_t *graph,
                        igraph_bool_t mode)
```

**graph:** grafo de viajes.

**mode:** **STRONG** para calcular si es conexo considerando la dirección de los arcos, para realizar la verificación sin considerar la dirección se debe usar la opción **WEAK**. El cálculo de esta métrica se realiza con ambas opciones.

El cálculo de las métricas descritas se realiza tanto para el grafo de viajes como de etapas debido a que ambos forman parte del análisis de la estructura de la red.

## 4.5. Asociación espacial

Una vez obtenidas las métricas se procede a asociarlas al espacio geográfico; para eso se debe adjuntar a los datos de cada nodo la ubicación de este en la ciudad, dato que es obtenido de la tabla `parada_util`, además de incorporar datos geográficos como la zona 777 y la comuna a la que pertenece el paradero. A continuación se describe el proceso para realizar la interpolación espacial con la herramienta QGIS y luego, se explica la partición de la ciudad a partir de la generación de comunidades teniendo en consideración la partición zonal 777.

### 4.5.1. Interpolación espacial

La interpolación espacial es el proceso de utilizar puntos con valores conocidos para estimar valores desconocidos en otros puntos. Para este trabajo tiene el objetivo de convertir una métrica discreta en la ciudad, que en este caso corresponde al *PageRank* y *Betweenness*, en valores para cada parte de ella, generando una imagen continua en el espacio. Esta tarea fue realizada con el programa QGIS, sistema de información geográfico de código abierto.

QGIS contiene un módulo para calcular interpolaciones espaciales, que corresponde a la usada en este trabajo. El primer paso consiste en seleccionar la opción “*raster* → *interpolación* → *interpolación*”, desplegándose la ventana de configuración mostrada en la figura 4.6. Posteriormente se debe seleccionar la capa de trabajo que contienen los datos, que en este caso corresponde a la distribución de los paraderos en la ciudad (ver figura 3.6).

El siguiente paso es indicar el atributo de interpolación, que este caso fueron las siguientes: *PageRank*, *Betweenness*, grado e intensidad (*Strength*). Con la métrica elegida se debe seleccionar el método de interpolación IDW (*Inverse Distance Weight*), la elección de este método se debe a que se busca comparar los resultados con los obtenidos por Zhong et al. (2014).

Por último, queda indicar el tamaño de la celda en la interpolación, donde se escoge un nivel de detalle tal que genera celdas sobre la ciudad de 50 metros por 50 metros aproximadamente, específicamente se usa un valor de 1.000 tanto para el número de filas y columnas.

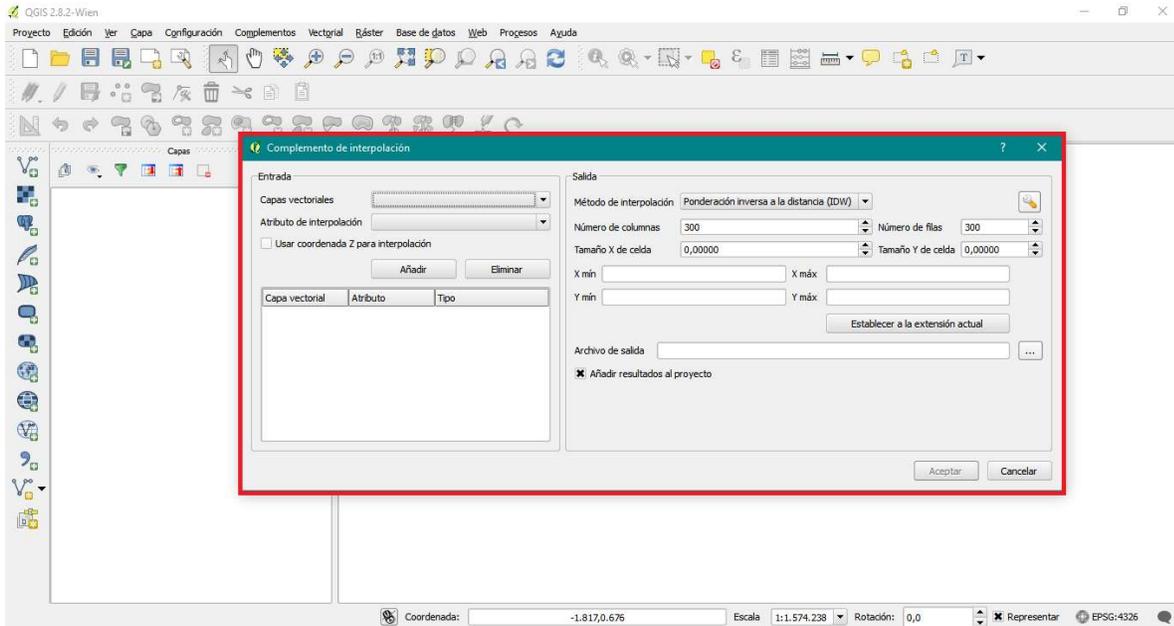


Figura 4.6: Ventana de configuración para la interpolación en QGIS. Fuente: elaboración propia

#### 4.5.2. Asignación de comunidad a zonificación

Los resultados de las comunidades están asociados a los nodos o paraderos, por lo que se tiene una imagen discreta de las áreas en donde una comunidad tiene presencia. Para generar bordes más claros (líneas divisorias) se utiliza la metodología de Zhong et al. (2014) para particionar la ciudad y asignar a cada zona la comunidad más relevante; esta relevancia se hace tomando todos los nodos que intersectan con la zona y asignándole la comunidad del nodo con mayor *PageRank*. Aquí se destaca que el *PageRank* es el entregado por *Infomap*.

La zonificación a utilizar será la zonificación 777 ya que presenta un alto detalle de la ciudad. El procedimiento para la asociación se realiza utilizando la extensión de PostgreSQL, PostGis, que permite realizar consultas espaciales. La que es de nuestro interés, es del estilo “Asigna una zona a un nodo si estos se intersectan”.

Para esta tarea se realizó el siguiente procedimiento:

- Se calcula el *PageRank* por comunidad para obtener la importancia relativa entre los nodos
- Para cada zona:
  - Tomar el conjunto de nodos que intersecta con la zona.
  - Obtener del conjunto el nodo con mayor *PageRank*.
  - La comunidad asociada al nodo con mayor *PageRank* es la comunidad de la zona.

## 4.6. Visualización de *PageRank* y *Betweenness*

La visualización de los datos se realiza tomando los valores obtenidos de la interpolación espacial y descartando los extremos (2 % superior y 2 % inferior). Luego, tomando el rango donde se encuentran los valores considerados ([2 %-98 %]) se procede a generar una clasificación de 10 tramos equidistantes. Cada tramo es mapeado a un color donde aquellos tramos con valores más altos son coloreados en rojo mientras que los más bajos en azul.

QGIS permite realizar lo descrito anteriormente en las opciones de la capa resultante del proceso de interpolación interpolación usando a la opción “estilo” y seleccionar los parámetros descritos en la tabla 4.6.

Opción	Parámetro
Tipo de renderizador	Unibanda pseudolor
Interpolación de color	Discreto
Cargar valores min/max	Corte del conteo acumulativo (ingresar valores 2,0 y 98,0)
Extensión	Completo
Precisión	estimar (más rápido)
Generar nuevo mapa de color	Paleta de colores RdBu
Invertir	Verdadero
Clases	10
Modo	Intervalo igual

Tabla 4.6: Parámetros para generar visualización de *PageRank* y *Betweenness*. Fuente: elaboración propia

Una vez establecido todos los parámetros se debe presionar el botón “clasificar”, que generará los rangos a colorear automáticamente basado en las opciones de la tabla 4.6. Una vista del resultado que se debería obtener es mostrado en la figura 4.7. En la imagen se puede observar el espacio de colores generado, asignando el color azul para los intervalos inferiores y rojo a los superiores, siendo estos últimos los que se analizan en el capítulo siguiente.

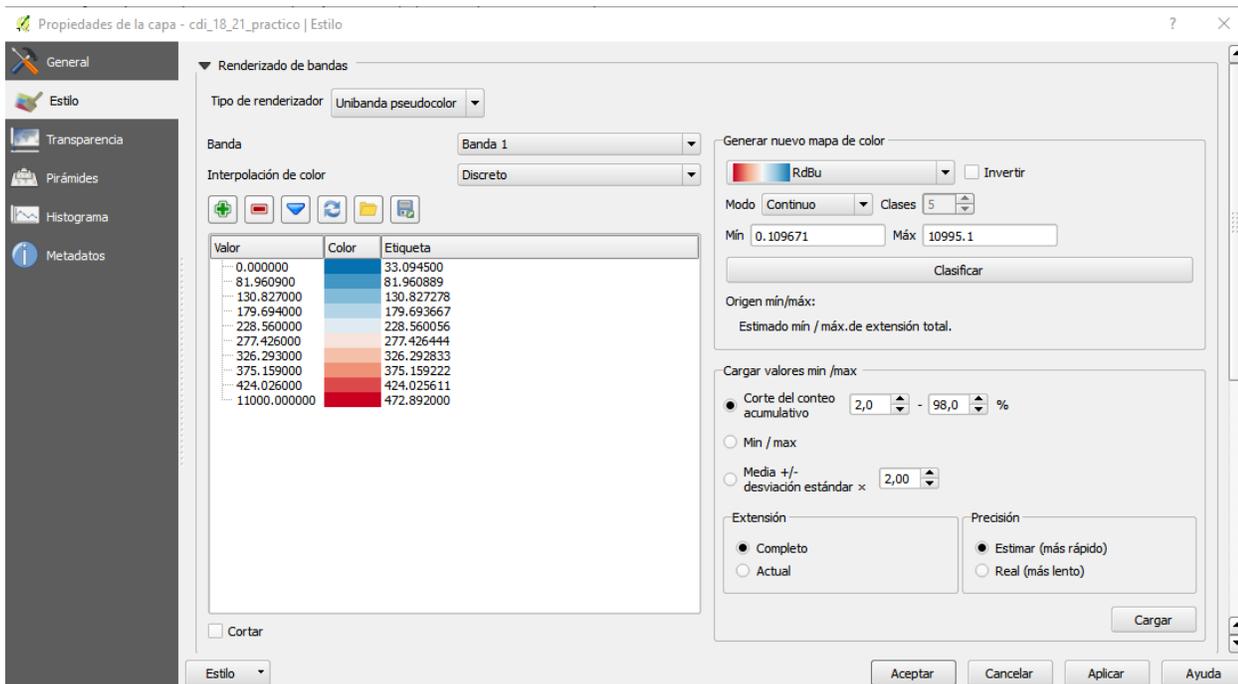


Figura 4.7: Configuración de la capa que contiene el resultado de la interpolación. Fuente: elaboración propia

# Capítulo 5

## Análisis

La identificación de la estructura espacial de movilidad se realiza por medio de la combinación de tres fuentes de datos: la red de viajes, la información del sistema de transporte, y la zonificación política-administrativa.

La red de viajes es un grafo dirigido con pesos que contiene información de interacciones entre distintos espacios físicos de la ciudad. La información del sistema de transporte provee la ubicación espacial de los nodos de la red, lo que permite identificar el área de impacto de cada métrica de la red. Por último, la zonificación político-administrativa permite conocer el contexto en el cuál está inmersa la ciudad. La combinación de las tres fuentes de datos permite tener una mirada completa de la estructura espacial de movilidad de la ciudad.

La figura 5.1 muestra las distintas fuentes utilizadas para el análisis. Un grafo construido a partir de los datos de tarjetas inteligentes (5.1.a, se muestran los nodos más relevantes), nodos de la red mapeados a su zona geográfica (5.1.b) y, por último, una segmentación político-administrativa de la ciudad definida por las autoridades con las líneas de tren subterráneo existentes (5.1.c).

Los datos de viajes consideran un período de una semana de abril del 2013, del 14 al 21, de domingo a domingo, lo cual permite capturar las zonas que son usadas para realizar

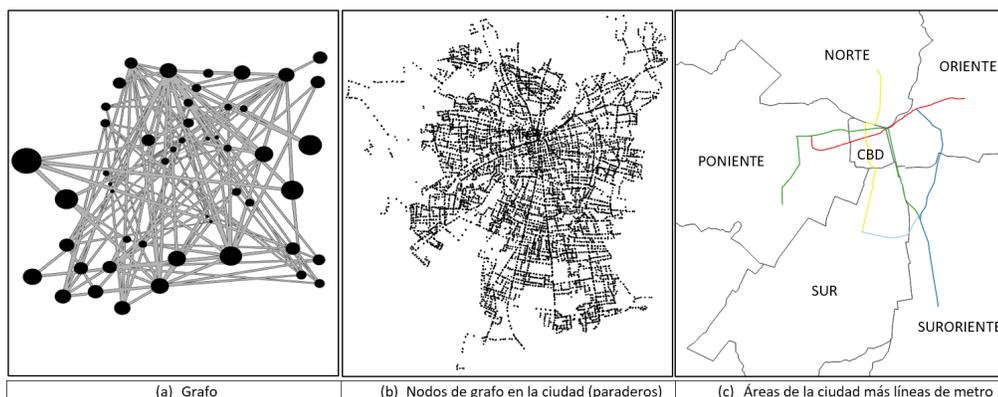


Figura 5.1: Fuentes de datos para el análisis. Fuente: elaboración propia

distintas actividades, incluyendo las zonas donde está localizados los hogares, lugares de trabajo o estudio, así como lugares donde se realiza otro tipo de actividades, sean estas en día laboral o durante el fin de semana. Además la distribución de validaciones realizada a lo largo de la semana<sup>1</sup> muestra un patrón regular a lo largo de los días, destacando dos aumentos considerables de intensidad en flujos, uno durante la mañana (06:00-09:00 hrs) y otro en la tarde (18:00-21:00 hrs), estos casos serán visualizados de manera independiente ya que proveen información general sobre intencionalidad de viaje.

El análisis de la red se realiza en dos grupos, el primero construido sobre el origen y destino de cada viaje (ignorando las etapas entre ellos) y dividido en tres ventanas temporales: Los viajes que comienzan entre las 06:00 y 09:00 hrs (punta mañana), los viajes que inician entre las 18:00 y 21:00 hrs (punta tarde) y por último uno que considera el día completo. El segundo grupo corresponde a un grafo construido sobre los viajes pero considerando las etapas dentro de este, es decir, los puntos de transbordo presentes en el viaje son considerados como nodos dentro de la red, y al igual que el grafo anterior, se construye sobre las tres ventanas temporales nombradas anteriormente. El por qué de esta separación se justifica en la interpretación que tiene cada métrica dentro de la ciudad (espacio urbano). La interpretación de espacio urbano para el *PageRank* tiene relación con aquellas zonas que concentran personas para la realización de una actividad, por lo que se opta por no involucrar zonas usadas como traslado. Por otro lado, la generación de comunidades tiene como objetivo la identificación de zonas de la ciudad que están fuertemente relacionadas entre sí pero no con el resto, para esta identificación se omiten las etapas intermedias de los viajes puesto que se busca identificar la relación entre el origen y destino de cada interacción espacial (viaje). Por último, el *Betweenness* como espacio urbano se relaciona con zonas de transbordo, áreas de la ciudad que las personas utilizan para llegar a su lugar de destino, es en esta métrica donde resulta relevante el conocer como fue el desplazamiento de las personas, por lo que uno de sus cálculos se realiza sobre la red con etapas.

Para construir la red se define un grafo dirigido con pesos como  $G = (N, L, W)$  que representa todos los viajes realizados entre cada par de paraderos y que comienzan (primera validación) dentro de la ventana de tiempo en consideración.  $N$  denota el número de zonas accesibles en la ciudad, representadas por los paraderos del transporte público de ésta. El conjunto  $L$  denota los traslados entre dos zonas de la ciudad, por lo que  $L$  corresponde a un conjunto de pares ordenados de  $N$ . Por último, el conjunto  $W$  denota el volumen (cantidad) de traslados entre dos zonas de la ciudad. Según lo anterior,  $N$  son los nodos del grafo,  $L$  representa los arcos y  $W$  denota los pesos de cada arco en  $L$ .

## 5.1. Análisis de la red

Como se mencionó en la introducción de este capítulo, se construyeron 6 grafos considerando tres ventanas temporales (todo el día, punta mañana y punta tarde) y dos configuraciones estructurales, un grafo que considera el origen y destino y otro que incorpora las etapas presente en los viajes. En la tabla 5.1 se puede ver un resumen de la información de los grafos.

---

<sup>1</sup>ver figura 3.2 del capítulo 3.

La columna ‘Id’ es un identificador del grafo que será usado a lo largo del capítulo, ‘Dato’ representa la fuente desde la cuál se construyó el grafo, ‘Rango horario’ corresponde a los distintos tramos horarios considerados, ‘Días’ indica los días considerados, que para nuestro caso corresponde a todos, desde el 14 al 21 de abril de 2013. La columna ‘Métrica a aplicar’ indica el uso que se le dará a cada grafo, mostrando que aquellos que consideran solo origen y destino de viaje serán usados para calcular el *betweenness* y *PageRank*, se destaca aquí el uso del grafo que utiliza los datos del día completo puesto que también será usado para calcular las comunidades con *Infomap*, mientras que el grafo de etapas será utilizado para el cálculo del *Betweenness*, donde se busca hacer una comparación en la utilización de ambos días.

Id	Dato	Rango horario	Días	Métrica a aplicar
viaje_día	Origen y destino	[00:00, 00:00)	Todos	<i>Betweenness, PageRank, Infomap</i>
viaje_mañana	Origen y destino	[06:00, 09:00)	Todos	<i>Betweenness, PageRank</i>
viaje_tarde	Origen y destino	[18:00, 21:00)	Todos	<i>Betweenness, PageRank</i>
etapa_día	Etapas	[00:00, 00:00)	Todos	<i>Betweenness</i>
etapa_mañana	Etapas	[06:00, 09:00)	Todos	<i>Betweenness</i>
etapa_tarde	Etapas	[18:00, 21:00)	Todos	<i>Betweenness</i>

Tabla 5.1: Grafos construidos y las métricas a aplicar sobre cada uno. Fuente: elaboración propia

### 5.1.1. Propiedades generales

Información importante puede ser obtenida a partir de las redes de movimientos humanos generadas, en particular, las distintas métricas para evaluar un grafo tienen una interpretación en la ciudad que permite caracterizarla. En la tabla 5.2 se muestran las propiedades globales de cada red. Las medidas son: número de nodos ( $N$ ), número de arcos, grado promedio ( $K$ ), largo promedio del camino más corto entre cada par de nodos ( $L$ ), coeficiente de agrupamiento ( $C$ ), existe un camino dirigido entre cada par de nodos ( $Conexo_{fuerte}$ ), existe un camino entre cada par de nodos ( $Conexo_{débil}$ ) y diámetro ( $D$ ).

Id	$N$	# Arcos	$K$	$L$	$C$	$Conexo_{fuerte}$	$Conexo_{débil}$	$D$
viaje_día	11.270	2.413.454	428,30	2.28	0.15	falso	verdadero	7
viaje_mañana	11.211	824.039	147	2.83	0.10	falso	verdadero	9
viaje_tarde	11.182	890.273	159,23	2.72	0.01	falso	verdadero	9
etapa_día	11.270	624.926	110,90	3.24	0.26	falso	falso	8
etapa_mañana	11.211	309.517	55,22	3.82	0.25	falso	falso	12
etapa_tarde	11.188	331.731	59,30	3.76	0.25	falso	falso	11

Tabla 5.2: Propiedades globales. Fuente: elaboración propia

El número de nodos representa cuantas zonas de la ciudad son accedidas, considerando el día completo para viajes y etapas se observa que el número de áreas accesibles de la ciudad disminuye durante la tarde. Por otro lado, el número de arcos presente en los grafos de viaje (viaje\_día, viaje\_mañana, viaje\_tarde) es considerablemente mayor que los presentes en los grafos de etapas (etapa\_día, etapa\_mañana, etapa\_tarde), lo que se puede explicar por el

hecho de que en el grafo de viajes existe una restricción más específica de asociación (origen y destino deben coincidir), impidiendo la reutilización de arcos en la red. Desde el punto de vista de la ciudad, los arcos indican cuantas zonas están conectadas directa (en el caso del grafo de etapas) o indirectamente (grafo de origen y destino) basado en el comportamiento de los viajes; en este punto se puede ver (grafos viaje\_mañana, viaje\_tarde y etapa\_mañana, etapa\_tarde) que la conectividad entre zonas de la ciudad cambia a lo largo del día, mostrando un valor mayor para la tarde, siendo consistente con el grado promedio ( $K$ ) de la red.

La evaluación de la presencia de componentes conexas en la red resulta de interés ya que las métricas con las que buscamos caracterizar la ciudad (*PageRank*, *Betweenness* e *Infomap*) dependen de ella. La variable  $Conexo_{fuerte}$  muestra que no existe un camino “dirigido” entre cada par de nodos para ninguno de los grafos generados (todos tienen el valor falso). Una situación distinta se observa en  $Conexo_{débil}$ , que evalúa la existencia de un camino sin tomar en cuenta el sentido de los arcos para cada par de nodos. Los grafos de viaje (viaje\_día, viaje\_mañana, viaje\_tarde) resultaron ser conexos pero no así los de etapas, los cuales presentan dos componentes conexas, situación que está relacionada por el modelado de los datos ya que una componente corresponde a los paraderos de bus y otra a las estaciones de metro. Cuando se describe una etapa se almacena el origen y destino (además de otros datos), si el origen es un paradero de bus, el destino será un paradero de bus, por el contrario, si el origen es una estación de metro su destino será una estación de metro. Los grafos de viaje no tienen este problema porque el origen y destino pueden pertenecer a etapas distintas. La componente inconexa de menor tamaño corresponde al conjunto de 100 estaciones de metro. Para efectos de las métricas, esto debe ser considerado en la evaluación de *Betweenness*, única métrica que utiliza la red de etapas.

Otras medidas calculadas fueron el largo de la ruta promedio (“average path length”) donde se observa que los grafos de viaje tienen un menor tamaño. Lo mismo ocurre con el coeficiente de Clustering, que para todos los grafos resultan ser bajos si se toma como referencia los obtenidos por Zhong et al. (2014) en la red de transporte público de Singapur.

El grado de cada nodo en la red muestra el nivel de interacción que tiene una zona de la ciudad con otras, cuando se observa el grado de entrada para un nodo  $k$ , se entiende como el número de zonas de la ciudad que están interesadas en la zona  $k$ , en el caso opuesto (grado de salida) denota el interés de la zona  $k$  por otras zonas de la ciudad, mientras que el grado representa cuantas áreas están conectadas. En la figura 5.2 se muestra la distribución de probabilidad acumulada del grado de entrada (5.2.a), grado de salida (5.2.b) y grado sin considerar el sentido (5.2.c) para los grafos etapa\_día, etapa\_mañana y etapa\_tarde, notar que el eje  $x$  presenta una escala logarítmica. Para el grado de entrada (figura 5.2.a) se ve que existe un mayor número de nodos con alto grado en el grafo etapa\_tarde con respecto al grafo etapa\_mañana, lo que muestra que durante la tarde hay un mayor número de nodos conectados con otros nodos en la red. Para el grado de salida (figura 5.2.b) existe una menor variación entre la mañana y la tarde, siendo más alto el grado de salida para el grafo etapa\_mañana. Lo anterior suena razonable ya que en una ciudad como Santiago la mayoría de las personas salen de su hogar a realizar alguna actividad en la mañana y luego vuelven a él durante la tarde. Para el grado sin sentido (figura 5.2.c), se observa que no existen diferencias considerables entre la mañana y tarde. Además, el grafo etapa\_día tiene un mayor volumen de arcos que el resto (su curva está desplazada hacia la derecha) en los tres gráficos, lo que es

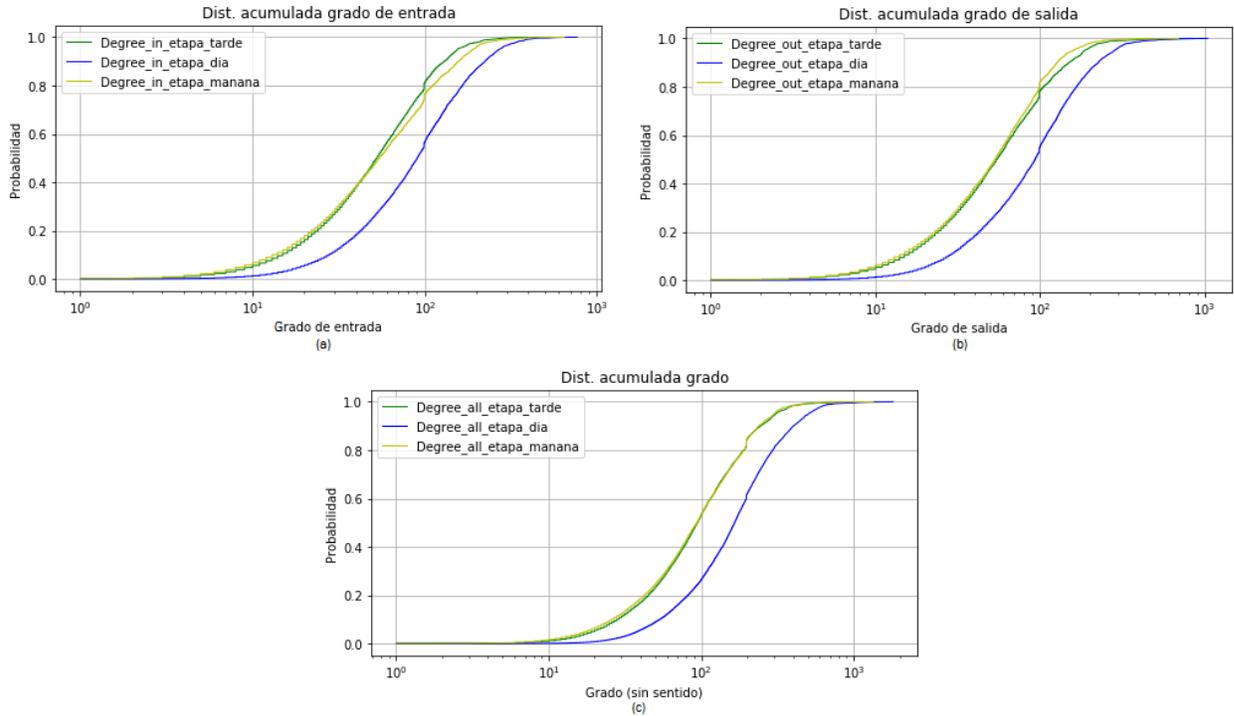


Figura 5.2: Distribución de probabilidad acumulada para grafos de etapas. Fuente: elaboración propia

entendible dado la ventana temporal que representa (las 24 horas del día), lo que a su vez le permite contener una mayor cantidad de etapas que contribuyen con nuevos arcos a la red.

Para el caso de los viajes ocurre un fenómeno más interesante. En la figura 5.3 se muestra la distribución de probabilidad acumulada para cada tipo de grado en los grafos viaje\_día, viaje\_mañana y viaje\_tarde. En el gráfico del grado de entrada (figura 5.3.a) se ve que el grafo viaje\_tarde tiene una mayor cantidad de nodos con alto grado que el grafo viaje\_mañana, mientras que en el gráfico del grado de salida (figura 5.3.b) ocurre el efecto opuesto, comportamiento también visto en los grafos de etapas (figura 5.2), lo que muestra consistencia entre los grafos de viajes y etapas. También se destaca la forma más plana de la distribución, lo que es indicador de una mayor uniformidad en esta, lo que se puede justificar por el hecho de que esta red considera solo el origen y destino de los viajes, mientras que en los otros grafos se consideran las etapas; criterio que aumenta el grado de un conjunto particular de nodos, como lo son aquellos usados para transbordo. Otro aspecto relevante es la similitud que presenta la distribución de los grafos viaje\_mañana y viaje\_día en el gráfico de la distribución de probabilidad del grado de entrada (figura 5.3.a), mostrando que el grado de accesibilidad alcanzado por el conjunto de nodos con mayor grado durante la mañana son similares a los alcanzados durante el día completo; el mismo efecto ocurre en el gráfico de la distribución del grado de salida (figura 5.3.b) entre el grafo viaje\_tarde y viaje\_día. Por último, el gráfico 5.3.c muestra la distribución del grado sin considerar su sentido; en este gráfico se puede ver que el nivel de conectividad general (sin considerar la dirección) entre el grafo viaje\_mañana y el grafo viaje\_tarde es bastante similar diferenciándose visiblemente en la parte final de la distribución, donde el grafo viaje\_mañana presenta nodos de alto grado con mayor conectividad.

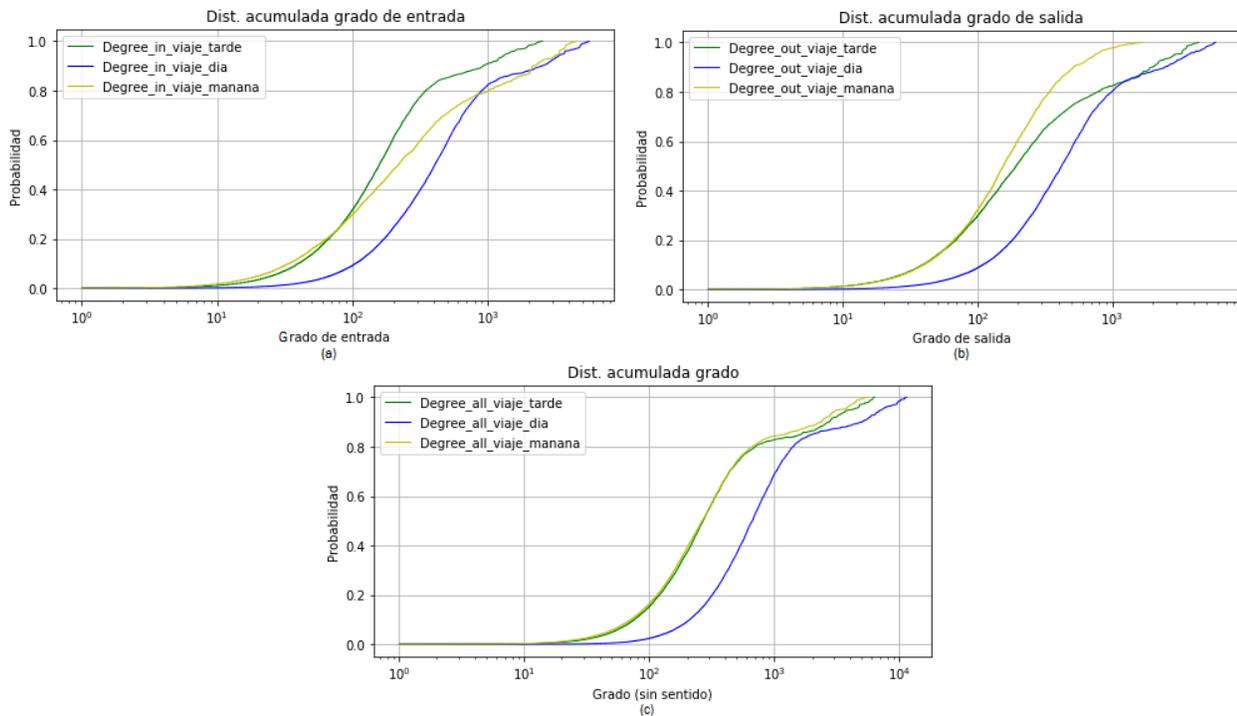


Figura 5.3: Distribución de probabilidad acumulada para grafos de viaje. Fuente: elaboración propia

Por otro lado, la métrica Intensidad (*Streength*) representa el volumen de viajes generado desde o hacia un área de la ciudad. En la figura 5.4 se muestra la relación entre el flujo de viajes (*Streength*) y el grado de cada nodo de la red. Se puede ver cómo en cada gráfico existe la presencia de pocos nodos que contienen un alto grado y un alto volumen de flujo (nodos ubicados en la parte superior derecha de cada gráfico), indicando las áreas de la ciudad que presentan una intensa (debido al flujo) conexión con otras. Este comportamiento junto con la presencia de un  $L$  pequeño es asociado a grafos que presentan comportamientos de *small-world*, que son redes a las que se les asocia la noción de que ganan eficiencia teniendo un número grande de conexiones locales y unas pocas conexiones globales que conectan los clusters locales (Boccaletti et al., 2006). Aunque tanto Zhong et al. (2014) como Boccaletti et al. (2006) son claros a la hora de indicar que dada las restricciones de las redes espaciales se deben emplear otros mecanismos para evaluar este fenómeno, en este caso las restricciones son las calles que sustentan la red de transporte público, que tienden a ser grafos planares, donde es poco probable encontrar un comportamiento de *small-world* (Cardillo et al., 2006).

## 5.2. Análisis espacial

### 5.2.1. Centros

Los centros son determinados sobre los grafos de viaje mediante el indicador de centralidad *PageRank*, que mide la importancia de un nodo para atraer flujo en la red. Con este indicador

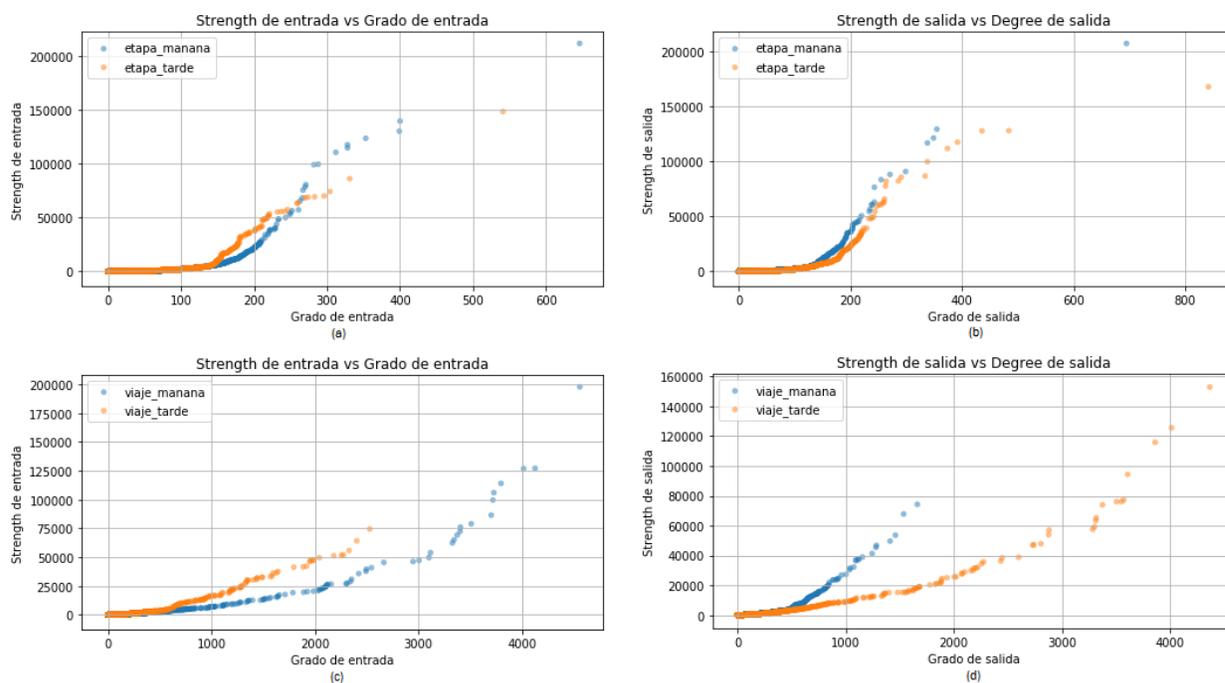


Figura 5.4: Grado vs strength en 2 variantes: entrada y salida sobre los grafos de mañana y tarde de viajes y etapas. Fuente: elaboración propia

se busca detectar aquellas zonas de la ciudad usadas por las personas para realizar actividades (trabajar, pernoctar, divertirse, etc.). Una visión general de las zonas que se identifican como centros puede ser encontrada analizando los datos de la semana completa (ver figura 5.5), donde en (a) se considera los datos de un día completo, mientras que (b) concentra los datos de mayor demanda durante la primera mitad del día y (c) concentra los datos de mayor demanda durante la segunda mitad del día.

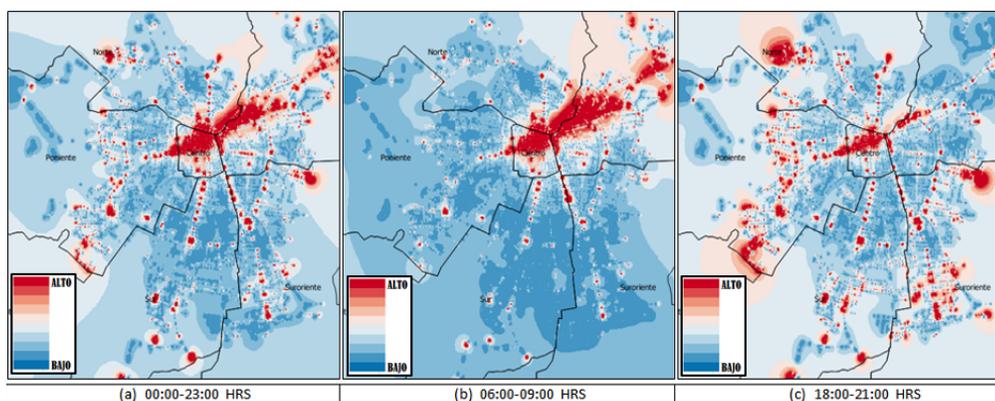


Figura 5.5: Identificación de centros (zonas de color rojo). Corresponde al *PageRank* interpolado. Fuente: elaboración propia

La figura 5.5.a muestra las zonas de destino más probables de ser preferidas por las personas para realizar actividades; a partir de esta se puede observar que gran parte de los destinos está fuertemente marcado por zonas donde se encuentra disponible el acceso al transporte subterráneo (metro), lo que es esperable dado que por una parte las inversiones en metro se

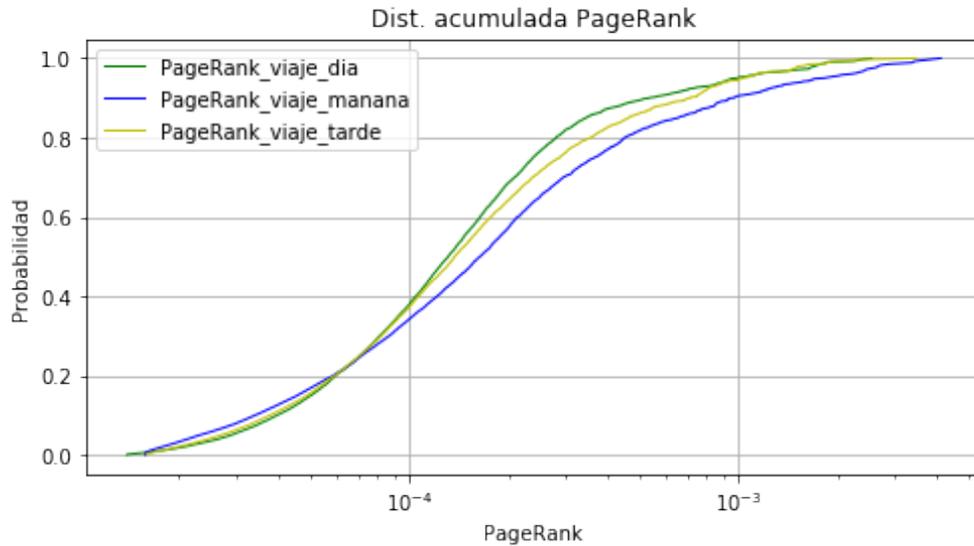


Figura 5.6: Distribución de *PageRank* en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia

definen de acuerdo a la demanda, privilegiando las zonas de mayor concentración de flujo. Por otra parte, el metro en sí es un modo de transporte masivo, con una alta concentración de flujo en sus estaciones. Este análisis permite detectar estas zonas, pero no es posible inferir la intencionalidad. Una estimación general del propósito de viaje puede ser inferido haciendo cortes temporales y considerando las dinámicas de flujo habituales usadas, por ejemplo, las descritas en Munizaga and Palma (2012), donde gran parte de los usuarios se mueven durante la mañana a su lugar de trabajo o estudio y en la tarde vuelven a su hogar, generando un patrón cíclico. Según lo anterior se puede hacer una distinción entre zonas de trabajo o actividad (figura 5.5.b) y zonas de vivienda (figura 5.5.c).

Otro aspecto es analizar la distribución que genera el indicador a lo largo de la red. En la figura 5.6 se muestra la distribución de probabilidad acumulada del *PageRank*. En ella se observan 3 ventanas temporales: día, mañana y tarde. La distribución de la mañana es la más uniforme, presentando una mayor cantidad de nodos con *PageRank* bajos en comparación con las otras dos distribuciones, luego durante la tarde se observa un crecimiento importante, indicando un mayor número de áreas de la ciudad atrayendo gente. La distribución del *PageRank* del día muestra la realidad actual de la ciudad mostrando una pendiente más elevada que las otras dos distribuciones, indicando que durante las horas no punta (06:00-09:00 y 18:00-21:00) los viajes realizados fortalecen aquellas zonas con mayor valor del indicador.

Si bien el *PageRank* por si solo aporta vasta información respecto a la ciudad, es posible obtener un poco más a partir de la adición del indicador del grado de entrada. En el análisis realizado aquí se busca determinar aquellas zonas que tienen un alto *PageRank* y además presentan un valor alto para el grado de entrada. En la Figura 5.7 es posible observar en mayor detalle aquellas zonas que presentan nodos con mayor grado de entrada, identificando áreas con buena accesibilidad y su influencia en el espacio, mostrando que éstas son muy apreciadas por individuos de distintas localidades (representado por un mayor número de grado de entrada).

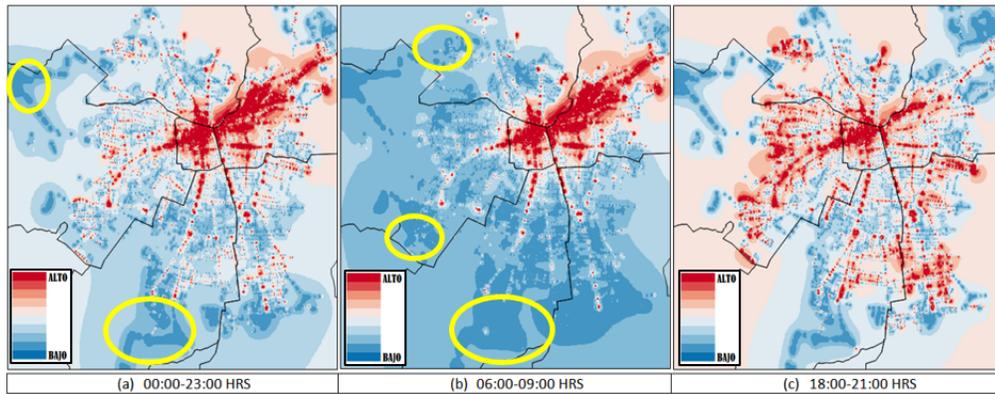


Figura 5.7: Visualización de los grados de entrada de un nodo. Fuente: elaboración propia

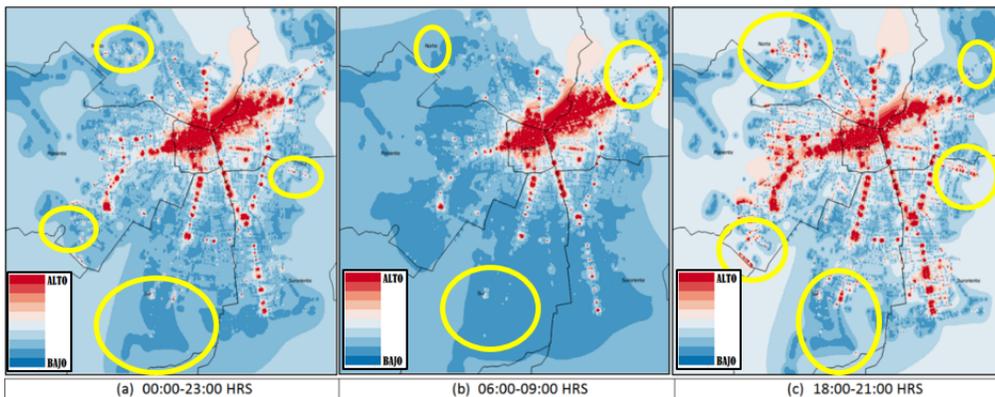


Figura 5.8: Visualización del flujo de entrada de un nodo. Fuente: elaboración propia

De lo anterior es posible diferenciar zonas con alto *PageRank* y que mantienen pocos arcos hacia él, mostrando que un fuerte flujo hacia aquellas zonas se mueve a través de pocas zonas de la ciudad (ver círculos demarcados en Figura 5.7).

Un aspecto relevante a destacar es la medición de accesibilidad presente en la Figura 5.7.c, la que es muy pareja a lo largo de la ciudad, lo que indica un fenómeno particular para el período de la tarde, donde gran parte de las zonas de la ciudad son accedidas desde muchos puntos de ésta.

Otro de los aspectos que se evalúa es la idoneidad del indicador para detectar zonas de actividad, para ello se considera la comparación contra la medición de la suma de los flujos de entrada (“*strength-in*”) sobre cada nodo para luego aplicar la misma técnica de visualización.

El flujo de entrada de un nodo  $i$ , es definido como  $f_i = \sum_j^N W_{ji}$ , que corresponde a la suma de los pesos de los arcos que llegan a él,  $W_{ji}$  es el peso del arco cuando hay un arco de  $j$  a  $i$ , 0 en otro caso.

Al comparar la Figura 5.5 con la figura 5.8 es posible observar que en esta última no se detectan zonas que sí detecta el algoritmo *PageRank* (ver círculos amarillos en figura 5.8).

También se debe ser cauto a la hora de considerar aquellas zonas de la periferia destacadas

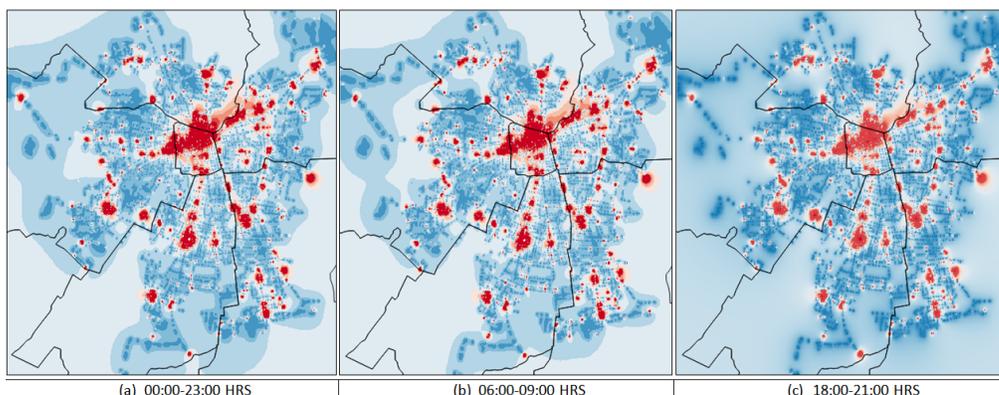


Figura 5.9: *Betweenness* con grafos de viaje en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia

como zonas centros dado que en esos lugares termina la cobertura de Transantiago y empiezan otros medios de transporte (buses interurbanos por ejemplo) de los cuales no tenemos datos, generando un sesgo que se debe tener en cuenta. Además, se destaca la capacidad del indicador de resaltar zonas no obvias (alejadas del centro y estaciones de metro). Por último, dado que la métrica es una probabilidad, se puede entender también como una expresión de la potencialidad de las áreas destacadas para atraer gente, lo que puede ser indicativo de un potencial de esas zonas para la generación de centros o sub-centros en el futuro.

### 5.2.2. Centros de pasada

Los centros de pasada corresponden a aquellas zonas de la ciudad que son usadas para realizar transbordos, y como espacio urbano son detectadas a través de la medida de centralidad *Betweenness*. De este modo se mide la importancia de un nodo en la organización del flujo presente en la red. A continuación se realizan tres cálculos del indicador: uno utilizando los grafos de viajes (figura 5.9), un segundo utilizando los grafos de etapas (5.11) y por último se define un *betweenness* a partir de los viajes considerando las etapas.

Las zonas identificadas como centros de pasada utilizando la definición teórica de *Betweenness* se muestran en la Figura 5.9 (en rojo). Para hacer este cálculo se ocupa solamente los datos de origen y destino de los viajes para construir el grafo. La estructura de las imágenes muestra una configuración que destaca en gran medida las líneas de metro, lo que muestra que este medio de transporte es muy relevante dentro de la red de transbordos.

Al analizar los períodos de la mañana y tarde por separado y el día completo, se puede ver que las imágenes son muy parecidas en sus zonas de transbordo, lo que indica que los flujos presentes en ambos períodos por cada punto son muy parecidos, dando indicios de un comportamiento similar de viaje en la mañana y tarde. Este hecho se puede corroborar mirando la distribución de probabilidad de los datos de viajes (ver figura 5.10).

Al aplicar el mismo algoritmo a los datos de transbordos realizados por las personas se obtienen los resultados de la figura 5.11. Al comparar la figura 5.9 con la figura 5.11 se puede

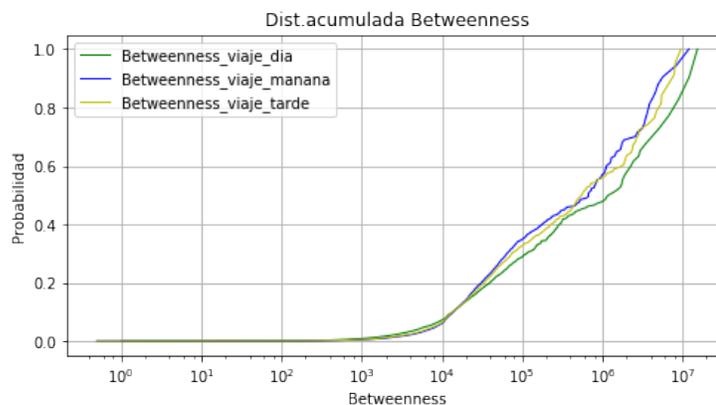


Figura 5.10: Distribución de probabilidad acumulada de *Betweenness* con grafos de viaje en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia

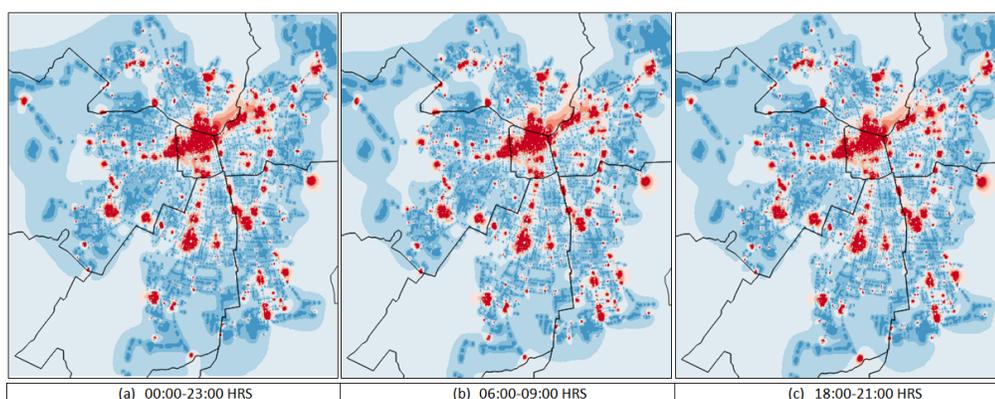


Figura 5.11: *Betweenness* usando datos de transbordo. Fuente: elaboración propia

apreciar que no existen diferencias sustanciales que permitan observar cambios relevantes dentro de los centros de pasada, al igual que el resultado obtenido anteriormente; el efecto se ve corroborado por la distribución de probabilidad acumulada de la métrica para cada grafo (ver figura 5.12). Los centros de pasada detectados tanto con los grafos de viaje como con los grafos de etapa son muy similares en los tres cortes temporales observados, esto puede tener sentido al considerar que usualmente las zonas de alta actividad (centros) presentan buena accesibilidad, por lo que también suelen ser usadas como centros de pasada. De lo anterior se puede desprender la idea de que los centros de pasada son detectables a partir de los datos de origen y destino solamente.

Es importante tener en cuenta que los grafos de etapa no son conexos, presentan dos componentes conexas, una construida con la red de paradas y otra con la red de metro. La forma en que se abordó este problema fue calcular la métrica *Betweenness* para cada nodo dentro de su red, por lo que las estaciones de metro ignoraron los nodos de paraderos de buses y viceversa.

Otra comparación interesante es el *Betweenness* que considera dentro de su cálculo sólo los pares origen-destino (con sus etapas) presentes en los datos, es decir, aquellos que haya realizado al menos una persona. Con esta comparación se busca observar si el hecho de

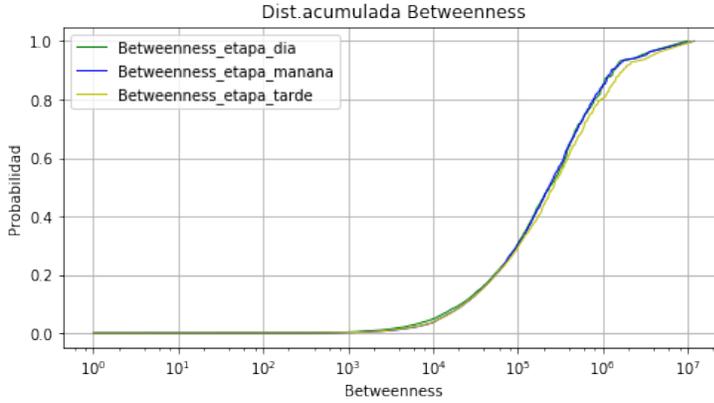


Figura 5.12: Distribución de probabilidad acumulada de *Betweenness* con grafos de etapa en tres ventanas temporales: día, mañana y tarde. Fuente: elaboración propia

considerar todos los pares de nodos dentro del grafo (como lo hace el *Betweenness* teórico) afecta la identificación de centros de pasada. Formalmente se define una adaptación del *Betweenness* que utiliza la información de los transbordos, es decir, considera aquellas rutas entre dos paraderos presente en los datos. Esta nueva versión que llamamos *Real Betweenness* aprovecha la información conocida de la ruta de los viajes y no sólo los extremos de estos. Según lo anterior, se define formalmente para un nodo  $k$  como:

$$RB(k) = \sum_{ij} \frac{\xi_{ij}(k)}{\xi_{ij}} \quad (5.1)$$

La Figura 5.13 muestra que al considerar solo los pares de origen y destino dentro de los datos como caminos válidos, desaparecen varios puntos considerados inicialmente como centros de pasada. Los centros de pasada detectados corresponden en gran medida a los sectores por donde viaja la red de metro. Esto puede significar que no todos los centros de pasada detectados por el *Betweenness* teórico que utiliza sólo el origen y destino tienen una existencia en la estructura espacial. Esencialmente desaparecieron varias zonas de la periferia que el *Betweenness* teórico reconoce, lo que se puede deber a nodos de la periferia que tienen pocas opciones de conexión y que por lo tanto eligen siempre los mismos nodos para conectarse a toda la red, potenciando aquellos nodos de la periferia.

### 5.2.3. Comunidades

Las comunidades corresponden a una partición de la ciudad que se ha ido formando a lo largo de los años como consecuencia de la interacción entre las actividades y la demografía, representada por intereses comunes (trabajo, recreación) según edad, ingreso y otras variables sociodemográficas (Newman, 2003). El algoritmo usado en este trabajo (*Infomap*) revela una partición de la ciudad al seleccionar las comunidades más significativas (más grandes y que dominan una zona definida de la ciudad). Es importante hacer notar que el algoritmo de detección de comunidades no contiene ningún tipo de información espacial por lo que su

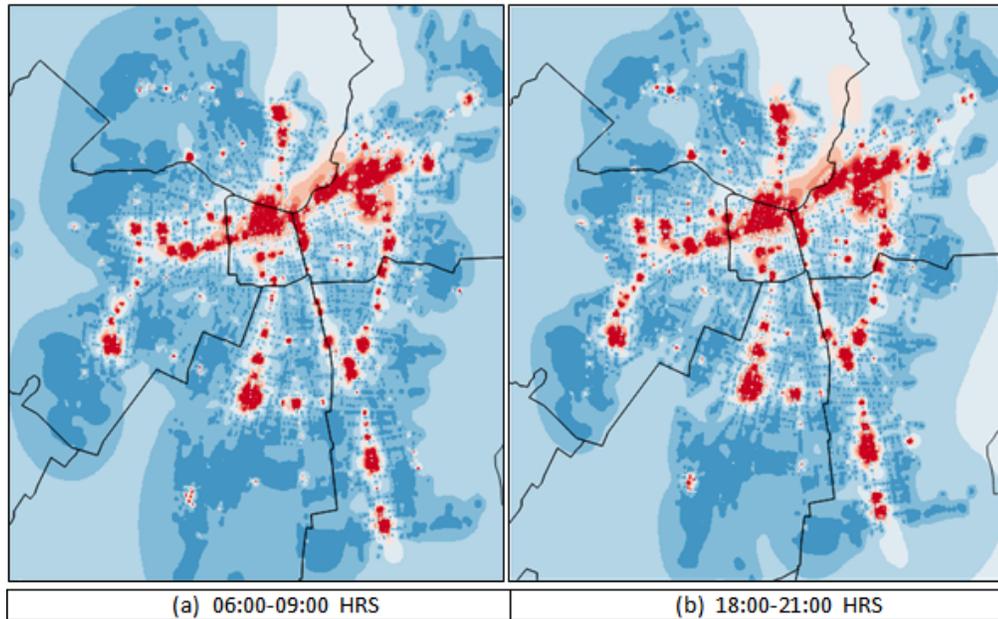


Figura 5.13: *Real Betweenness* (usando datos de transbordo). Fuente: elaboración propia

resultado depende solamente de las interacciones presentes en la ciudad generadas por las personas.

### 5.2.3.1. Comunidades detectadas

El algoritmo fue ejecutado varias veces hasta obtener un conjunto que representara de mejor manera la ciudad mediante observaciones visuales, llegando a una configuración de 41 comunidades. En la tabla 5.3 se puede observar la cantidad de nodos asignada a cada comunidad. Si bien son 41 comunidades, se muestra que muchas de ellas resultan ser un remanente del algoritmo y que no calzan con nuestro objetivo, que es encontrar aquellas que den forma a la ciudad. A partir de lo anterior se desechan desde la comunidad 15 en adelante, quedando como candidatos las primeras 14 comunidades (1 a 14). Es importante decir que el número de la comunidad fue asignado por *Infomap* y corresponde al orden (creciente) en que el algoritmo fue creando las comunidades.

Al inspeccionar cada comunidad vemos que la comunidad uno corresponde a aquella que concentra todas las estaciones de metro más un conjunto de paraderos de bus. En la figura 5.14 se muestra la distribución espacial de la comunidad 1. Es relevante esta comunidad puesto que concentra una componente esencial del sistema de transporte público como lo es el sistema de tren subterráneo. Además se ve que la comunidad de metro es una estructura diseminada a lo largo de la ciudad, mostrando que genera fuertes interacciones en gran parte de ella. También se muestra que es el sector sur-oriente el que tiene una mayor cantidad de nodos en la red, lo que podría indicar lo esencial del sistema para las personas que originan o finalizan su viaje en ese sector. Aunque resulta relevante, no cumple con nuestro objetivo dar forma a la ciudad por lo que se deja como un caso de estudio aparte. Lo mismo ocurre con las comunidades 12,13 y 14 (ver figura 5.15), que resultan ser de interés por su forma,

Comunidad	# nodos	Comunidad	# nodos	Comunidad	# nodos
1	581	15	61	29	1
2	1629	16	76	30	2
3	1969	17	86	31	1
4	1329	18	59	32	2
5	1352	19	20	33	1
6	671	20	32	34	1
7	1005	21	6	35	1
8	699	22	8	36	1
9	459	23	4	37	1
10	272	24	2	38	1
11	569	25	2	39	1
12	122	26	1	40	1
13	106	27	1	41	1
14	126	28	1		

Tabla 5.3: Comunidades generadas con la cantidad de nodos asignada por *Infomap*

dado que no se asocian a ninguna de las presentes en la administración geopolítica, ya sea a nivel de sectores, comunas o zonificación 777.

Las comunidades restantes han sido seleccionadas puesto que se asocian a la estructura de la ciudad. En la Figura 5.16 se pueden observar las comunidades que aparecen a partir de los flujos, las que están compuestas por un conjunto de paraderos que mantienen una fuerte interacción entre ellos pero no con el resto. Cada una muestra no tan solo una fuerte cohesión a nivel de flujo, sino que también física, ya que los nodos están geográficamente cercanos en su mayoría, situación no menor dado que su generación no utiliza en ningún momento información espacial (posición de los nodos). Otra característica es su expresividad dentro de la ciudad, existe un conjunto de comunidades que genera un anillo exterior (comunidades 2,3,5,6,8,9) para que luego aparezca otro conjunto que toma presencia al interior del anillo (comunidades 4,7,10,11).

Otro aspecto de estas comunidades es su tendencia a expresarse hacia el centro de la ciudad, si bien fuera del centro cada una puede ser asociada a un área geográfica en particular, todas ellas coexisten en esta gran zona central que se ha desarrollado a lo largo del tiempo y se ha ido expandiendo al oriente en las últimas décadas. Esto confirma el comportamiento fuertemente monocéntrico subyacente.

Otra característica importante a evaluar es la forma que dan estas comunidades a la ciudad y cómo se relacionan con los límites geo-administrativos. En el caso de Santiago, la división administrativa se realiza en torno a municipios o comunas, administradas por una institución llamada Municipalidad. El área metropolitana de Santiago contiene 38 comunas; la comuna asociada al *CBD* es la comuna de Santiago. Al comparar esta estructura con la estructura definida por las comunidades, se puede ver que ninguna comunidad detectada forma parte de una sola comuna, lo que muestra que la estructura de actividades de las personas trasciende la estructura político-administrativa. Se podría decir que las comunas tienen un grado de especialización tal que un grupo importante de personas debe salir de estas para encontrar

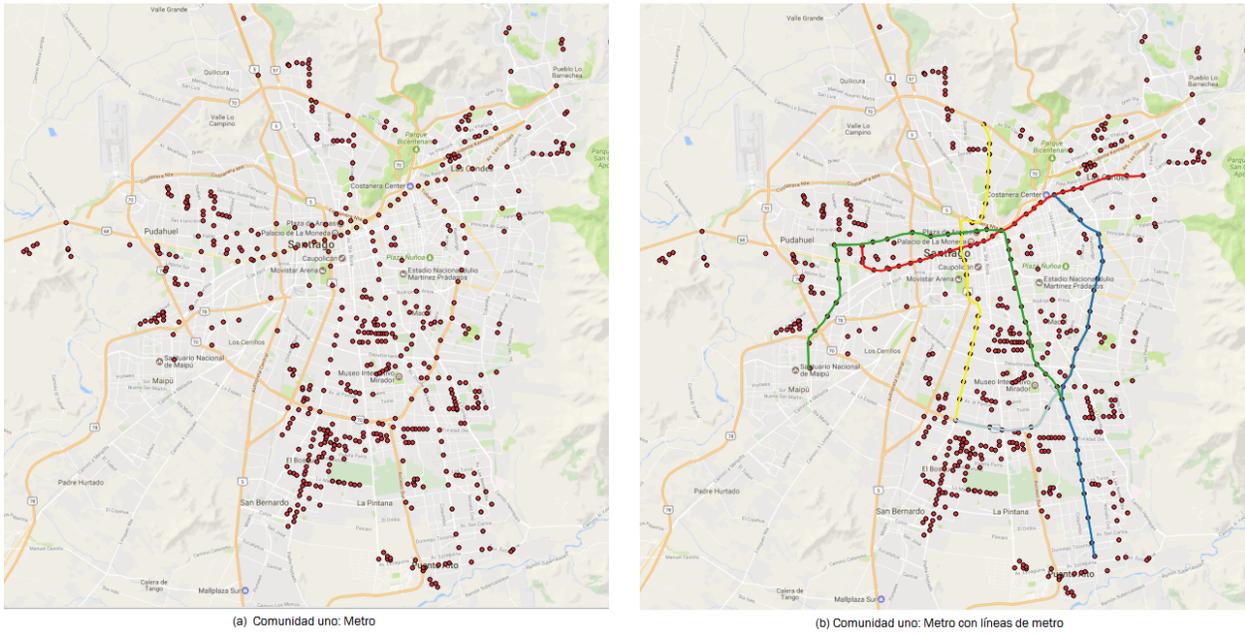


Figura 5.14: Comunidad de Metro. (a) distribución espacial (b) distribución espacial con la proyección de las líneas de metro . Fuente: elaboración propia

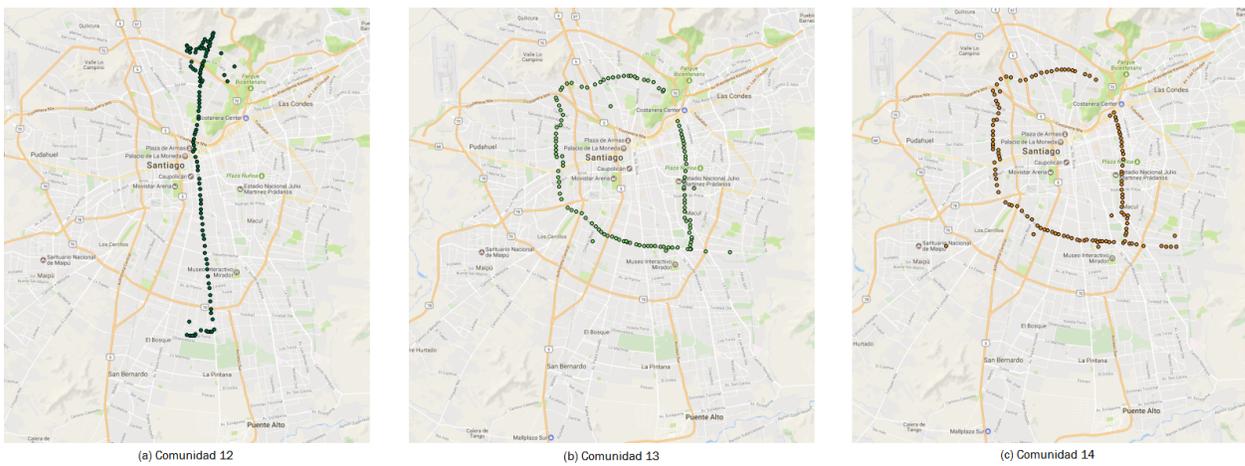


Figura 5.15: Comunidades 12,13 y 14. Fuente: elaboración propia

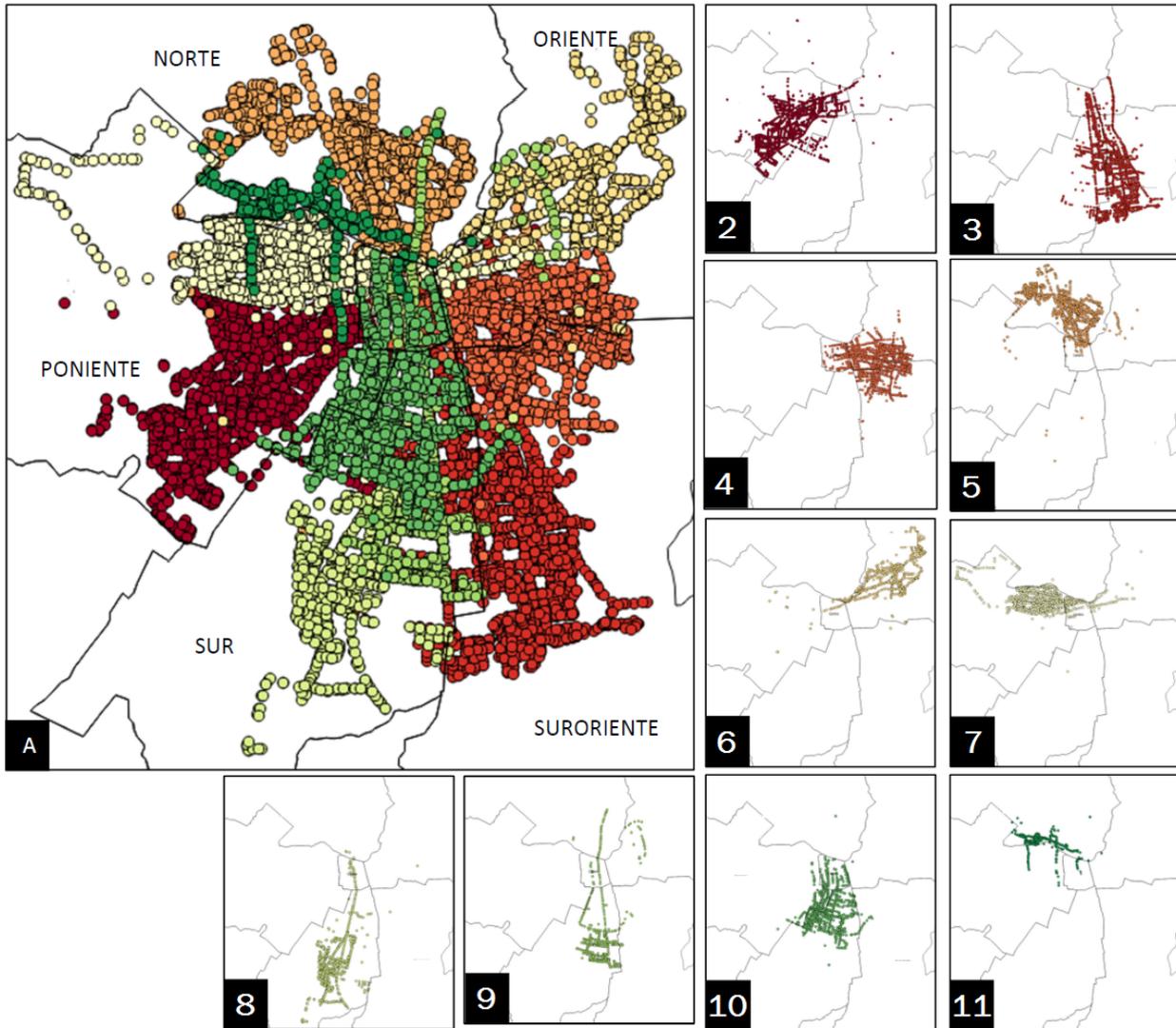


Figura 5.16: Comunidades más significativas con los macro sectores de la ciudad (líneas negras). Fuente: elaboración propia

las actividades que requiere.

### 5.2.3.2. Asociación espacial con la zonificación 777

Parte de la metodología propuesta por Zhong et al. (2014) sobre las comunidades consiste en la identificación de bordes geográficos que permitan identificar de manera clara la división espacial presente de las interacciones de movilidad en la ciudad. Zhong et al. (2014) Tomó una de las divisiones zonales de Singapur y construyó los bordes asignándole a cada zona la comunidad asignada al nodo presente en la zona con el *PageRank* más alto respecto de los otros nodos de la comunidad; este criterio asigna los territorios a aquellas comunidades para las cuales resulta más relevante, de esta manera se logra en Singapur el efecto mostrado en la figura 5.19. Para el caso de Santiago de Chile escogimos la zonificación 777, que es la utilizada por organismos estatales y consultoras para realizar mediciones en distintos ámbitos, uno de ellos es la encuesta origen-destino del año 2012 (Muñoz and Contreras, 2015). Se aplicó la misma metodología sobre dos casos: uno considerando las 41 comunidades y otro con las diez comunidades que dan forma a la ciudad.

En la figura 5.17 se muestra el resultado de la metodología de asignación de comunidad aplicada sobre las 41 comunidades encontradas con *Infomap*. Los bordes que aparecen son claros en la periferia pero se desordenan a medida que avanzan hacia el centro de la ciudad. Sin embargo en el sector sur de la ciudad se encuentra un desorden similar al del centro, mostrando el traslape presente de las comunidades que conviven en la misma zona y resultan ser más importantes en zonas aisladas. Otra característica importante es la dominación que ejercen aquellas comunidades con forma de anillo descritas en la figura 5.15, indicando que los valores de su *PageRank* resultan ser más altos que las comunidades que dominan la periferia.



Figura 5.17: Asignación de comunidades a zonificación 777 usando las 41 comunidades encontradas con *Infomap*. Fuente: elaboración propia

Al filtrar las comunidades usando el conjunto que da forma a la ciudad y aplicar la metodología se obtiene el resultado mostrado en la figura 5.18. Esta vez se obtienen bordes más ordenados, siendo más claros los bordes del sector sur y oriente de la ciudad en comparación con la fig 5.17, aunque igual persiste el desorden en el centro. Lo anterior indica que no es posible asignar comunidades al centro de la ciudad, todas participan en él y resulta relevante para todas, haciendo que nadie sea dominante en esa área. Además permiten complementar el análisis del *PageRank* que muestra al centro como un gran punto de atracción, con las comunidades podemos entender que esa atracción viene de todos lados de la ciudad, confirmando

el comportamiento monocéntrico de la ciudad.



Figura 5.18: Asignación de comunidades a zonificación 777 usando las 10 comunidades que dan forma a la ciudad. Fuente: elaboración propia

### 5.2.3.3. Comparación con Singapur

Otro aspecto relevante a ser presentado es la situación existente en las ciudades de Chile y Singapur, que es donde fue aplicada la metodología inicialmente. Mientras en la primera se observa una clara tendencia de todas las comunidades a generar presencia en el centro (ver figura 5.16), dado por la existencia de un gran *CBD* (*Central Business District*) que provee principalmente trabajo a diversas comunas, Singapur muestra una estructura más distribuida, con una partición de la ciudad que detalla una fuerte interacción en zonas acotadas de la ciudad, mostrando indicios de un desarrollo policéntrico, que va en línea con el plan impulsado por el país oriental (de 1991) centrado en el crecimiento económico sustentable que contempla la existencia 4 centros regionales Zhong et al. (2014). En la figura 5.19 se puede observar estructura de comunidades para Singapur.

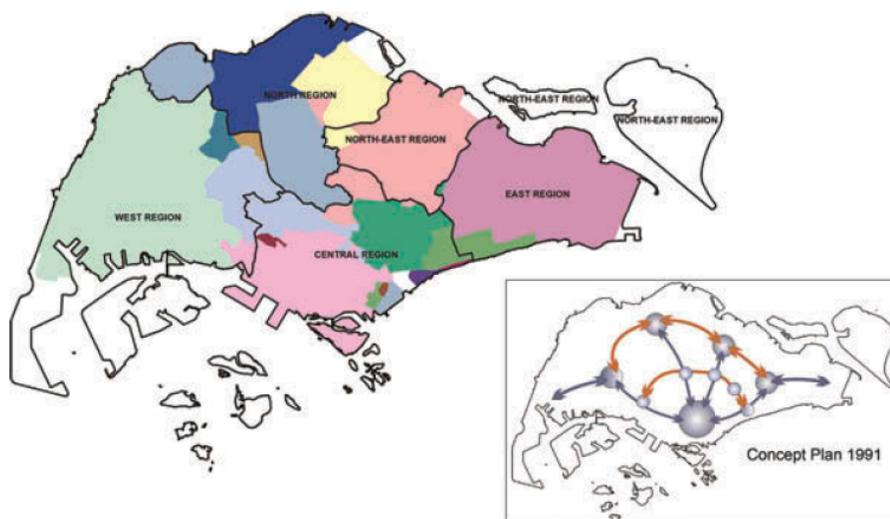


Figura 5.19: Comunidades presentes en Singapur con datos del 2012. Fuente: Zhong et al. (2014)

#### 5.2.3.4. Estructuración de Transantiago el 2012

Transantiago en un principio definió el sistema de transporte estableciendo servicios “locales” (o “alimentadores”) y “troncales”. Para lo anterior se dividió la ciudad en 10 zonas (identificadas con letras desde la A a la J más un color único) donde cada una era administrada por una empresa, con excepción de la zona A. La idea era simple: los buses locales realizan viajes cortos en zonas específicas de la ciudad y los buses troncales funcionarían como puentes entre alimentadores o locales y conectarían con el centro de la ciudad. En la figura 5.20 se muestra la zonificación explicada en este párrafo.

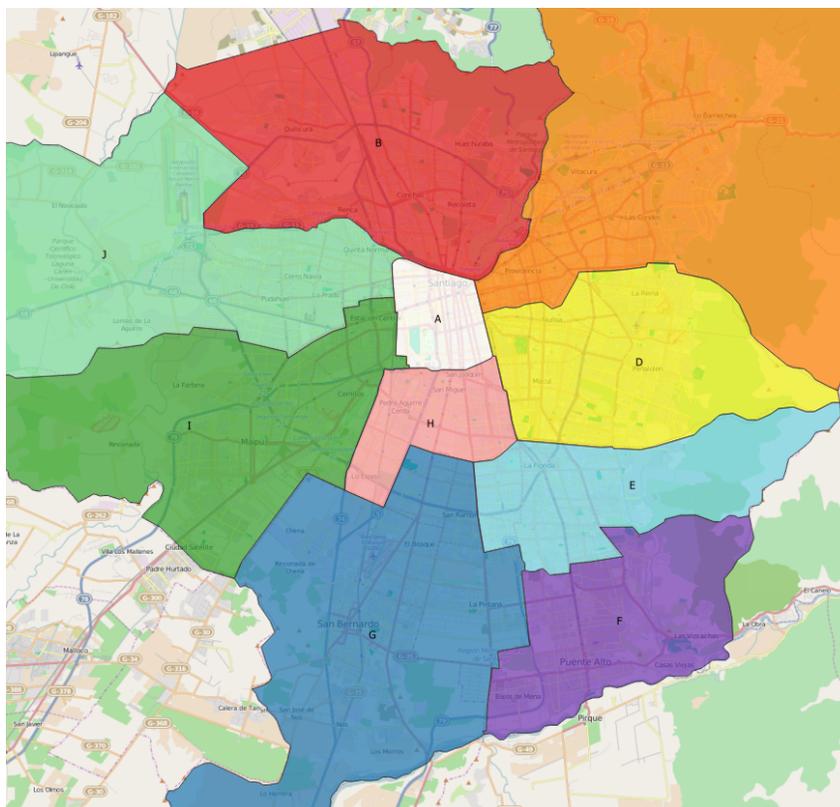


Figura 5.20: División operativa de Transantiago en el año 2012. Fuente: elaboración propia

En la figura 5.21 se muestran las comunidades junto con la zonificación establecida en los inicios del sistema, cada imagen corresponde a una comunidad y esta se identifica como el conjunto de puntos de color negro, donde cada punto es un paradero. Allí se muestra como una parte de las comunidades (1, 3, 4, 5, 6, 7, 8, 9, 10) concentran su núcleo (parte más densa) dentro de una zona, pero también es posible observar dos casos anómalos: el primero corresponde a la comunidad 2, que concentra su núcleo en dos zonas (celeste y violeta), lo que indica que un grupo importante de personas se mueve entre esos dos sectores. Lo anterior puede indicar un problema de asignación de zonas dado que una gran cantidad de gente deberá tomar buses troncales para moverse por un tramo corto (entre esas dos zonas) debido a que los buses alimentadores particionan la comunidad en dos; esto aconseja unir aquellas zonas y pensar recorridos locales que abarquen esta nueva y extensa sección de la ciudad. La segunda comunidad anómala corresponde a la número 11, donde no se observa un núcleo pero sí muestra una fuerte interacción entre dos extremos de la ciudad, dejando claro

que un grupo importante de gente se mueve constantemente de sur a norte (y viceversa). Al complementar los datos con el análisis realizado anteriormente del *PageRank* es posible concluir que parte importante del flujo se produce desde en el sector sur (vivienda) hacia el norte (lugar de actividad).

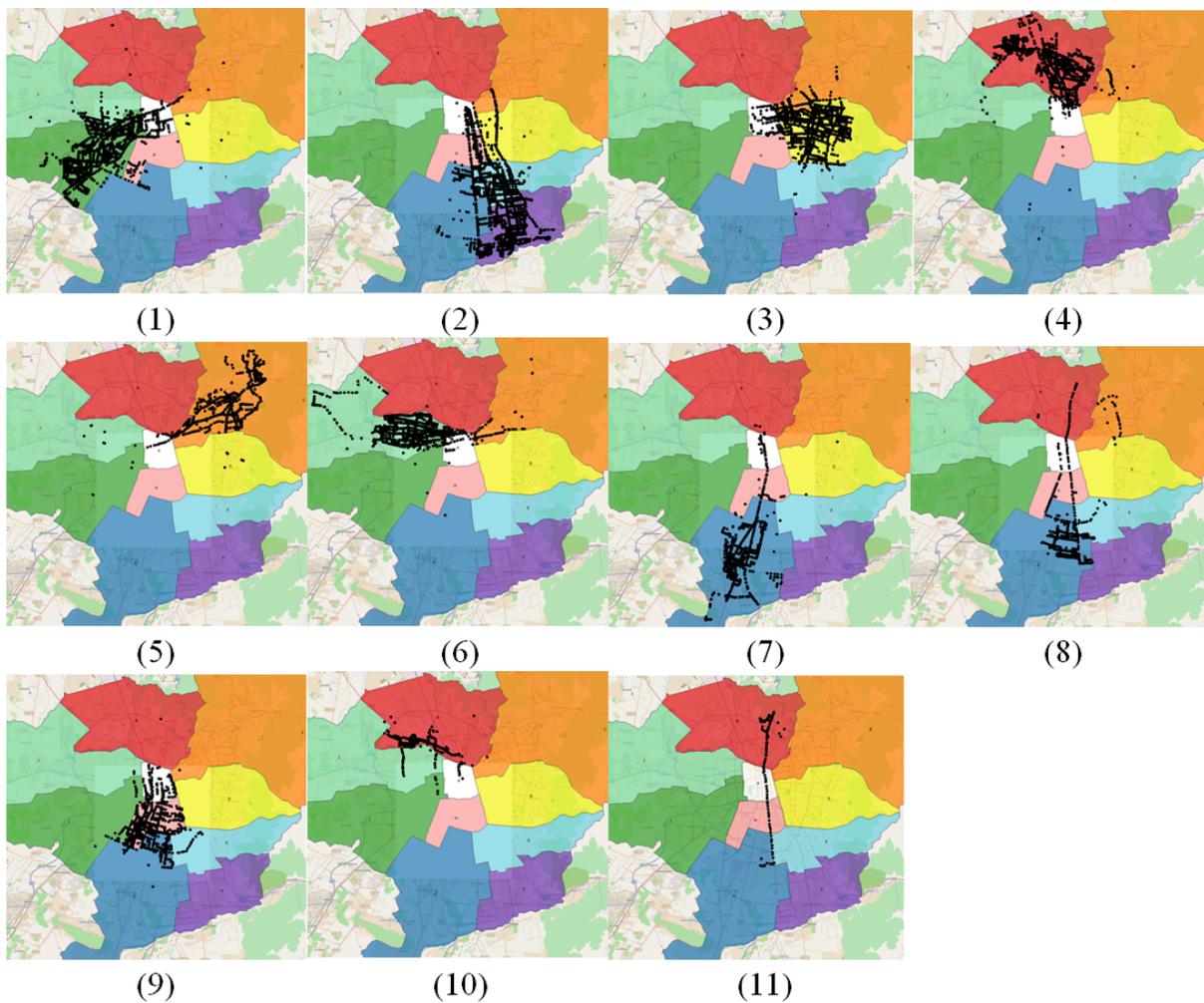


Figura 5.21: Comunidades más significativas contrastadas con la zonificación de Transantiago del año 2012. Fuente: elaboración propia

# Capítulo 6

## Conclusiones y trabajo futuro

La ciudad de Santiago de Chile por medio de su sistema integrado de transporte público (Transantiago) ha capturado datos de millones de viajes a lo largo de una semana, permitiendo generar información valiosa en distintos ámbitos (investigación, operación y planificación del sistema de transporte, entre otros). Este trabajo es un aporte al ámbito de la investigación sobre las estructuras espaciales urbanas, y en particular, sobre la estructura espacial de la ciudad de Santiago de Chile, aspecto que históricamente ha tenido una visión economicista pero que en los últimos años ha incorporado nuevas herramientas de análisis con resultados interesantes y prometedores.

En este trabajo se realizó un análisis espacial para descubrir la estructura espacial urbana a partir de los datos de transporte público por medio de la identificación cuantitativa de los espacios urbanos descritos por Zhong et al. (2014): centros, centros de pasada y comunidades.

Para los centros se hace uso de la medida de centralidad *PageRank*. Esta muestra que sigue existiendo un gran CBD (*Central Business District*) con una localización centro-oriente de la ciudad, lo que expresa un comportamiento monocéntrico, corroborado con la evaluación temporal de los datos, donde la ciudad se contrae por las mañanas para luego expandirse durante la tarde. Su presencia es muy fuerte cerca de las estaciones de metro subterráneo, mostrando la relevancia espacial que tiene la red de metro en la ciudad para definir su grado de atraktividad. Otro aspecto es su evolución durante el día, donde se observa el aumento de nodos con un *PageRank* bajo y durante la tarde ocurre el caso opuesto, hecho que corrobora el comportamiento monocéntrico. Otra característica de la métrica es la identificación de posibles nuevos centros ubicados fuera del CBD, lo que no se correlaciona necesariamente con los niveles actuales de viajes que atraen ya que el *PageRank* de un área depende del flujo y la cantidad de conexiones presentes con otras áreas. Esto podría ayudar a identificar potenciales sub-centros.

Respecto a los centros de pasada se hace uso de la medida de centralidad *Betweenness*, la cuál es calculada sobre los grafos de viajes, y grafos de etapas, no encontrando diferencias significativas, lo que se puede deber al alto grado de coincidencia entre las zonas de alta actividad y las que presentan buena accesibilidad. Inicialmente se podría pensar en omitir el grafo de etapas y realizar los cálculos solo con el grafo de viajes, pero no es aconsejable

puesto que áreas con poca actividad y alto grado de accesibilidad podrían ser ocultadas al ignorar las etapas.

Otro aspecto relevante son las preferencias de viaje mostradas durante el período de alta demanda de la mañana y período de alta demanda de la tarde, donde se observa que los flujos existentes en cada período utilizan los mismos centros de pasada, lo que puede indicar una regularidad en la gran mayoría de los viajes. A pesar de lo anterior, hay que ser cautos a la hora de asumir una detección de centro de pasada ya que el cálculo de éste a través del *Betweenness* teórico tiene la desventaja de considerar todos los pares de nodos como posibles viajes, lo cual no es necesariamente correcto para una ciudad por lo que un resultado más adecuado sería usar la métrica *Real Betweenness*, que tiene el beneficio de considerar solamente los pares de nodo presentes en los datos.

Los grafos de etapas utilizados no son conexos debido a la forma como es realizada la estimación de bajada para cada etapa, donde se supone un paradero de bus como destino cuando el origen fue un paradero de bus, y cuando el origen es una estación de metro se espera como destino una estación de metro impidiendo que se genere un arco entre un paradero de bus y estación de metro, los grafos de viaje no tienen este problema debido a que su origen y destino pueden pertenecer a etapas distintas. Cualquier métrica calculada sobre los grafos de etapa debe tener presente esta situación.

Para las comunidades se utiliza el algoritmo *Infomap* que agrupa nodos en función de los flujos presentes en la red. Se detectaron 41 comunidades donde se encontró que muchas de ellas son pequeñas (menor a 100 nodos), por lo que fueron ignoradas para el análisis. De las 41 se rescataron 14 comunidades, seleccionando diez para dar forma a la ciudad.

La primera comunidad descubierta agrupa todas las estaciones de metro, mostrando que el sistema subterráneo presenta un fuerte flujo de viajes entre estaciones. Además se observa la presencia de nodos dispersos a lo largo de la ciudad que tienen una mayor concentración en el sector sur-oriente, lo que podría indicar que ese sector de Santiago es más dependiente de él.

Las diez comunidades seleccionadas para dar forma a la ciudad parecen identificar sectores autocontenidos, pero todas presentan también un borde que se extiende al centro, mostrando que el CBD es alimentado por todas las comunidades. Además, cada comunidad en conjunto con el *PageRank* permite conocer sus zonas de alta interacción, lo que podría ser usado como identificación de zonas alimentadoras que realicen viajes dentro de ésta y que se conecten por medio de buses troncales.

Parte de la metodología para la identificación de comunidades corresponde a la identificación de sus bordes, esta parte arrojó buenos resultados en la periferia, logrando definir bordes claros, pero a medida que uno se acerca al CBD se observa un traslape en los bordes; lo que impide identificar una partición de la ciudad en ese sector, lo que lleva a concluir que el centro de la ciudad no es una zona perteneciente a una comunidad sino que a todas. Además muestra que el CBD es alimentado desde todas partes de la ciudad, mostrando un comportamiento más monocéntrico que el descrito en Singapur.

Otro resultado relevante resultó de estudiar la configuración del sistema de transporte

público del año 2012, esta dividió la ciudad en 10 zonas: los buses locales realizan viajes cortos en zonas específicas de la ciudad y los buses troncales funcionan como puentes entre alimentadores o locales. Al comparar estas zonas con las comunidades se observa un traslape entre comunidades que cubren más de una zona, lo que puede haber significado un aumento en el tiempo de viaje para los usuarios, puesto que las personas pertenecientes a la comunidad, verían siempre la necesidad de abordar buses troncales a la hora de realizar un viaje, aunque estos fueran de corta distancia. Lo importante de este punto radica en que análisis como el realizado en este trabajo hubiesen contribuido a la definición de zonas y haber evitado problemas como el descrito.

Según lo anterior, la metodología propuesta por Zhong et al. (2014) aplicada sobre la ciudad de Santiago de Chile genera resultados interesantes, al mostrar una estructura espacial urbana con carácter más monocéntrico revelada por los viajes hechos en transporte público. Es importante destacar que el uso de la misma metodología dio resultados distintos para ciudades que son distintas (Santiago de Chile y Singapur), lo que resulta ser positivo ya que muestra que las métricas son capaces de capturar la esencia de la ciudad.

## 6.1. Trabajo futuro

Este trabajo realizado es el primer paso para el estudio de la estructura espacial de la ciudad de Santiago de Chile a partir de datos de movilidad, un acercamiento que buscó adaptar una metodología aplicada en Singapur a una realidad distinta como lo es Santiago. El siguiente paso corresponde a mejorar o enriquecer el análisis con herramientas que se adapten mejor a la realidad local y al problema que se aborda, así como la generación de nuevas líneas de trabajo. A continuación se describen posibles mejoras al trabajo realizado e ideas que podrían derivar en nuevos estudios.

En este trabajo se buscó identificar la estructura espacial de la ciudad de Santiago de Chile presente en una semana del mes de abril de 2013 utilizando tres ventanas temporales: el día completo, mañana (06:00-09:00) y tarde (18:00-21:00). Adicional a lo anterior, resultaría interesante (dado el origen de los datos) agregar comparaciones de otros cortes temporales, como puede ser por el tipo de día (laboral, sábado y domingo) dado que se sabe que afecta considerablemente el comportamiento de los usuarios que utilizan el sistema (i.e. cambia el flujo a lo largo de la red). Un trabajo más profundo puede ser la evaluación de estas métricas a lo largo de los años, lo que puede ayudar a observar como la estructura de la ciudad evoluciona.

Respecto al *Betweenness*, la propuesta de Newman (2005) sobre la utilización de *random walks* para su cálculo podría ser una mejor opción que la utilizada en este trabajo. La definición convencional asume que el flujo se expande a través de las rutas más cortas, restricción que para el desplazamiento de personas no siempre es cierto (por ejemplo, tomar una ruta más larga para comprar alguna cosa de interés). Al utilizar *random walks* se relaja la restricción permitiendo que todas las rutas entre cada par de nodos contribuyan, pero manteniendo la importancia de las rutas más cortas.

Las visualizaciones planteadas aquí definen rangos estándar para la generación de escala de colores y, por ende, la identificación de zonas de actividades también es arbitraria. Una forma más apropiada podría ser la identificación de estas zonas mediante metodologías libres de parámetros como la propuesta por Louail et al. (2014) que permite hacer uso de métodos estadísticos para su definición.

Respecto al proceso de generación de comunidades, en este trabajo se abordó la evaluación de un algoritmo (*Infomap*), pero una línea de trabajo puede ser la evaluación de otros algoritmos de generación de comunidades (Newman, 2003) para luego analizar su rendimiento desde el punto de vista de la movilidad, incluso se podría evaluar si algunos son más adecuados para ciertos tipos de ciudad.

Al considerar solamente datos de transporte público se deja fuera otros medios de transporte, por lo que una mejora podría ser incorporar nuevas fuentes de datos de viajes en otros modos (e.g. taxi, auto, bicicleta) para tener una matriz de viajes más completa y complementar el análisis con otras fuentes de datos.

Respecto del *PageRank*, se habló de la presencia de potenciales centros o sub-centros fuera de los convencionales, esto es algo que aparece en el análisis pero que requiere validación, lo que corresponde a estudiar esas zonas de la ciudad con mayor profundidad.

Actualmente la interpolación utilizada para los paraderos contempla para cada uno una influencia que disminuye con la distancia sin establecer un límite superior, lo cuál es poco realista puesto que pasada una distancia la persona posiblemente decida optar por otro paradero para llegar a su destino. Según esto es que la influencia de la interpolación debería ser limitada, un primer paso podría ser la realización de la interpolación estableciendo como límite la distancia de caminata usada en el cálculo de estimación de bajada hecha por Munizaga and Palma (2012).

Otra posible línea a seguir corresponde a la generación de pruebas estadísticas para evaluar la significancia de los resultados encontrados, en particular, evaluar si la estructura encontrada resulta ser significativa, donde parte del proceso corresponde a definir cuales son esas pruebas.

Una información adicional que contiene la tarjeta bip! es su tipo, que puede ser estudiante o no, con esto se podría filtrar solamente aquellos viajes de estudiantes y calcular las métricas sobre este nuevo grafo, con el objetivo de establecer cuales son los centros, centros de pasada y comunidades establecidos por estos y luego compararlos con la población general.

Por último, un estudio adicional se puede realizar con el resultado de las comunidades al incorporar información demográfica en el análisis por medio de la construcción de estadígrafos por zona 777 para luego compararlo con las comunidades presentes en ella para ver si existe alguna correlación entre las comunidades y alguna de las variables, como puede ser el nivel socioeconómico. Lo anterior con el fin de caracterizar las comunidades.

# Capítulo 7

## Anexos

### A. Descripción tabla de viajes

A continuación se describe la tabla de viajes utilizada para este trabajo, contiene 104 columnas, la razón de tal cantidad de columnas se debe a que es una tabla que ha sido utilizada para diversos propósitos, tanto de análisis como de trabajos de tesis de otros estudiantes.

Tabla 7.1: Columnas presentes en la tabla viajes

N	Columna	Descripción
1	<b>id</b>	Identificador de tarjeta bip encriptado asociado al viaje
2	<b>nviaje</b>	Número de correlativo de viaje por id
3	<b>netapa</b>	Cantidad de etapas realizadas en el viaje
4	<b>netapasinbajada</b>	Cantidad de etapas de viaje que no fue posible estimar bajada
5	<b>ultimaetapaconbajada</b>	Número de etapa correlativa que tiene asignada una bajada
6	<b>tviaje_seg</b>	Tiempo total de viaje en segundos
7	<b>tviaje_min</b>	Tiempo total de viaje en minutos
8	<b>dviajeeuclidiana_mts</b>	Distancia euclidiana (en metros) recorrida entre la ubicación inicial y final del viaje
9	<b>dviajeenruta_mts</b>	Distancia (en metros) recorrida entre la ubicación inicial y final
10	<b>paraderosubida</b>	Código interno de paradero donde se estima que el usuario comenzó el viaje
11	<b>paraderobajada</b>	Código interno de paradero donde se estima que el usuario terminó el viaje
12	<b>comunsubida</b>	Comuna donde está ubicado el paradero de subida
13	<b>comunabajada</b>	Comuna donde está ubicado el paradero de bajada
14	<b>diseno777subida</b>	Zona de la partición 777 donde está ubicado el paradero de subida

N	Columna	Descripción
15	<b>diseno777bajada</b>	Zona de la partición 777 donde está ubicado el paradero de bajada
16	<b>tiemposubida</b>	Fecha y hora en que el usuario comenzó el viaje
17	<b>tiempobajada</b>	Fecha y hora en que el usuario terminó el viaje
18	<b>periodosubida</b>	Período Transantiago en que el usuario comenzó el viaje
19	<b>periodobajada</b>	Período Transantiago en que el usuario terminó el viaje
20	<b>tipodia</b>	Tipo de día en el que comenzó el viaje (Laboral, Sábado, Domingo)
21	<b>mediahora</b>	Período de media hora en que comenzó el viaje
22	<b>contrato</b>	Código para diferenciar los tipos de usuarios de la tarjeta
23	<b>factorexpancion</b>	Número de viajes que representa el registro
24	<b>tiempomediodeviaje</b>	media entre el tiempo de subida y tiempo de bajada
25	<b>periodomediodeviaje</b>	Período transantiago donde se encuentra la media entre el tiempo de subida y bajada
26	<b>mediahoramediodeviaje</b>	Período media hora donde se encuentra la media entre el tiempo de subida y bajada
27	<b>tipodiamediodeviaje</b>	tipo de día donde se encuentra la media entre el tiempo de subida y bajada
28	<b>t_1era_etapa</b>	Minutos de duración de la primera etapa del viaje
29	<b>d_1era_etapa</b>	Metros recorridos en la primera etapa del viaje
30	<b>tespera_1era_etapa</b>	tiempo esperado por el usuario para abordar un bus e iniciar la siguiente etapa
31	<b>ttransbordo_1era_etapa</b>	tiempo de movimiento entre la parada de bajada y subida de la siguiente etapa
32	<b>tcaminata_1era_etapa</b>	tiempo de movimiento que fue realizado a pie
33	<b>t_2da_etapa</b>	Minutos de duración de la segunda etapa del viaje
34	<b>d_2da_etapa</b>	Metros recorridos en la segunda etapa del viaje
35	<b>tespera_2da_etapa</b>	tiempo esperado por el usuario para abordar un bus e iniciar la siguiente etapa
36	<b>ttransbordo_2da_etapa</b>	tiempo de movimiento entre la parada de bajada y subida de la siguiente etapa
37	<b>tcaminata_2da_etapa</b>	tiempo de movimiento que fue realizado a pie
38	<b>t_3era_etapa</b>	Minutos de duración de la tercera etapa del viaje
39	<b>d_3era_etapa</b>	Metros recorridos en la tercera etapa del viaje
40	<b>tespera_3era_etapa</b>	tiempo esperado por el usuario para abordar un bus e iniciar la siguiente etapa
41	<b>ttrasbordo_3era_etapa</b>	tiempo de movimiento entre la parada de bajada y subida de la siguiente etapa
42	<b>tcaminata_3era_etapa</b>	tiempo de movimiento que fue realizado a pie
43	<b>t_4ta_etapa</b>	Minutos de duración de la cuarta etapa del viaje
44	<b>d_4ta_etapa</b>	Metros recorridos en la cuarta etapa del viaje

<b>N</b>	<b>Columna</b>	<b>Descripción</b>
45	<b>tespera_4ta_etapa</b>	tiempo esperado por el usuario para abordar un bus e iniciar la siguiente etapa
46	<b>ttransbordo_4ta_etapa</b>	tiempo de movimiento entre la parada de bajada y subida de la siguiente etapa
47	<b>tcaminata_4ta_etapa</b>	tiempo de movimiento que fue realizado a pie
48	<b>serv_1era_etapa</b>	servicio utilizado en la primera etapa
49	<b>serv_2da_etapa</b>	servicio utilizado en la segunda etapa
50	<b>serv_3era_etapa</b>	servicio utilizado en la tercera etapa
51	<b>serv_4ta_etapa</b>	servicio utilizado en la cuarta etapa
52	<b>linea_metro_subida_1</b>	Línea de metro utilizada en la subida de la primera etapa (cuando corresponde)
53	<b>linea_metro_subida_2</b>	Línea de metro utilizada en la subida de la segunda etapa (cuando corresponde)
54	<b>linea_metro_subida_3</b>	Línea de metro utilizada en la subida de la tercera etapa (cuando corresponde)
55	<b>linea_metro_subida_4</b>	Línea de metro utilizada en la subida de la cuarta etapa (cuando corresponde)
56	<b>linea_metro_bajada_1</b>	Línea de metro utilizada en la bajada de la primera etapa (cuando corresponde)
57	<b>linea_metro_bajada_2</b>	Línea de metro utilizada en la bajada de la segunda etapa (cuando corresponde)
58	<b>linea_metro_bajada_3</b>	Línea de metro utilizada en la bajada de la tercera etapa (cuando corresponde)
59	<b>linea_metro_bajada_4</b>	Línea de metro utilizada en la bajada de la cuarta etapa (cuando corresponde)
60	<b>paraderosubida_1era</b>	Código de paradero de subida para la primera etapa
61	<b>paraderosubida_2da</b>	Código de paradero de subida para la segunda etapa
62	<b>paraderosubida_3era</b>	Código de paradero de subida para la tercera etapa
63	<b>paraderosubida_4ta</b>	Código de paradero de subida para la cuarta etapa
64	<b>tiemposubida_1era</b>	Tiempo de subida para la primera etapa
65	<b>tiemposubida_2da</b>	Tiempo de subida para la segunda etapa
66	<b>tiemposubida_3era</b>	Tiempo de subida para la tercera etapa
67	<b>tiemposubida_4ta</b>	Tiempo de subida para la cuarta etapa
68	<b>zona777subida_1era</b>	zona 777 donde se realizó la subida para la primera etapa
69	<b>zona777subida_2da</b>	zona 777 donde se realizó la subida para la segunda etapa
70	<b>zona777subida_3era</b>	zona 777 donde se realizó la subida para la tercera etapa
71	<b>zona777subida_4ta</b>	zona 777 donde se realizó la subida para la cuarta etapa
72	<b>paraderobajada_1era</b>	Código de paradero de bajada para la primera etapa
73	<b>paraderobajada_2da</b>	Código de paradero de bajada para la segunda etapa
74	<b>paraderobajada_3era</b>	Código de paradero de bajada para la tercera etapa
75	<b>paraderobajada_4ta</b>	Código de paradero de bajada para la cuarta etapa

N	Columna	Descripción
76	<b>tiempobajada_1era</b>	Tiempo de bajada para la primera etapa
77	<b>tiempobajada_2da</b>	Tiempo de bajada para la segunda etapa
78	<b>tiempobajada_3era</b>	Tiempo de bajada para la tercera etapa
79	<b>tiempobajada_4ta</b>	Tiempo de bajada para la cuarta etapa
80	<b>zona777bajada_1era</b>	zona 777 donde se realizó la bajada para la primera etapa
81	<b>zona777bajada_2da</b>	zona 777 donde se realizó la bajada para la segunda etapa
82	<b>zona777bajada_3era</b>	zona 777 donde se realizó la bajada para la tercera etapa
83	<b>zona777bajada_4ta</b>	zona 777 donde se realizó la bajada para la cuarta etapa
84	<b>tipotransporte_1era</b>	Tipo de transporte usado en la primera etapa (ZONA PAGA, BUS o METRO)
85	<b>tipotransporte_2da</b>	Tipo de transporte usado en la segunda etapa (ZONA PAGA, BUS o METRO)
86	<b>tipotransporte_3era</b>	Tipo de transporte usado en la tercera etapa (ZONA PAGA, BUS o METRO)
87	<b>tipotransporte_4ta</b>	Tipo de transporte usado en la cuarta etapa (ZONA PAGA, BUS o METRO)
88	<b>escolar</b>	Boolean que indica si es una tarjeta de prepago estudiantil
89	<b>tviaje_en_vehiculo_min</b>	tiempo en minutos en que estubo viajando en bus o metro
90	<b>tipo_corte_etapa_viaje</b>	Criterio por el que el algoritmo de estimación de viaje cortó la etapa
91	<b>proposito</b>	Estimación de propósito de viaje (Actividad1Minuto, Estudio, sin_bajada, trabajo, hogar, otro)

## B. Descripción tabla de paradas y estaciones de metro

Tabla 7.2: Columnas presentes en la tabla estaciones\_metro

N	Columna	Descripción
1	<b>codigotrx</b>	Nombre de la estación de metro junto con la línea
2	<b>latitud</b>	posición
3	<b>longitud</b>	posición
4	<b>x</b>	Longitud en formato UTM
5	<b>y</b>	Latitud en formato UTM
6	<b>linea</b>	Línea de metro a la que pertenece la estación
7	<b>codigoestandar</b>	Nombre de estación de metro más línea a la que pertenece
8	<b>tipo</b>	Tipo de estación (Normal o Transbordo)

N	Columna	Descripción
9	<b>codigosinlinea</b>	código estándar sin línea a la que pertenece
11	<b>comunas</b>	Comuna donde está ubicada la estación de metro
10	<b>censal_1992</b>	formato de zonificación del Gran Santiago
12	<b>diseno_563</b>	formato de zonificación del Gran Santiago
13	<b>diseno_777</b>	formato de zonificación del Gran Santiago
14	<b>eod_2001</b>	formato de zonificación del Gran Santiago
15	<b>eod_2006</b>	formato de zonificación del Gran Santiago
16	<b>estraus_264</b>	formato de zonificación del Gran Santiago
17	<b>estraus_404</b>	formato de zonificación del Gran Santiago
18	<b>estraus_410</b>	formato de zonificación del Gran Santiago
19	<b>estraus_618</b>	formato de zonificación del Gran Santiago
20	<b>zonas_6</b>	formato de zonificación del Gran Santiago

Tabla 7.3: Columnas presentes en la tabla redparadas

N	Columna	Descripción
1	<b>codigo</b>	código de parada para transantiago
2	<b>codigousuario</b>	código de parada para usuarios
3	<b>comuna</b>	comuna donde está ubicado el paradero
4	<b>nombre</b>	nombre del paradero
5	<b>sentido</b>	orientación geográfica (O-P, N-S, S-N, P-O)
6	<b>fila_superior</b>	calle cercana considerando la orientación
7	<b>fila_inferior</b>	calle cercana considerando la orientación
8	<b>grupo_parada</b>	calle representativa
9	<b>x</b>	Longitud en formato UTM
10	<b>y</b>	Latitud en formato UTM
11	<b>latitud</b>	posición
12	<b>longitud</b>	posición
13	<b>censal_1992</b>	formato de zonificación del Gran Santiago
15	<b>diseno_563</b>	formato de zonificación del Gran Santiago
16	<b>diseno_777</b>	formato de zonificación del Gran Santiago
17	<b>eod_2001</b>	formato de zonificación del Gran Santiago
18	<b>eod_2006</b>	formato de zonificación del Gran Santiago
19	<b>estraus_264</b>	formato de zonificación del Gran Santiago
20	<b>estraus_404</b>	formato de zonificación del Gran Santiago
21	<b>estraus_410</b>	formato de zonificación del Gran Santiago
22	<b>estraus_618</b>	formato de zonificación del Gran Santiago
23	<b>zonas_6</b>	formato de zonificación del Gran Santiago

## C. Conversión de archivo CSV a PAJEK

La conversión de los archivos CSV (obtenidos desde la base de datos en PostgreSQL) al formato Pajek es realizado por medio de un script en Python. El procedimiento es descrito

en la figura 7.1.

```
arrParaderos = [] // paraderos encontrados
cantidadTotalArcos = 0
strArcos = ''
strVertices = ''
para cada linea en csv:
    si paraderoSubida existe en arrParaderos:
        indiceParaderoSubida = rescatar indice
    de lo contrario:
        agregar paraderoSubida en arrParaderos
        indiceParaderoSubida = rescatar indice
        strVertices += indiceParaderoSubida + nombreParadero + quiebreDeLinea

    si paraderoSubida existe en arrParaderos:
        indiceParaderoBajada = rescatar indice
    de lo contrario:
        agregar paraderoSubida en arrParaderos
        indiceParaderoBajada = rescatar indice
        strVertices += indiceParaderoBajada + nombreParadero + quiebreDeLinea

    strArcos += indiceParaderoSubida + indiceParaderoBajada + NumeroViajes +
        quiebreDeLinea
    cantidadTotalArcos++

f = crearArchivoPajek()
Escribir(f, "*Vertices" + cantidad(arrParaderos) + quiebreDeLinea)
Escribir(f, strVertices)
Escribir(f, "*Arcs" + cantidadTotalArcos)
Escribir(f, strArcos)
```

Figura 7.1: Algoritmo para transformar archivo con formato CSV a formato PAJEK. Fuente: elaboración propia

## D. Generación de CSV de viajes

La construcción de los CSV para la generación de los archivos Pajek se realiza por medio de consultas SQL directamente al gestor de base de datos PostgreSQL por medio del comando psql.

En la figura 7.2 se muestra la consulta ejecutada para los datos de viaje. El código define un arreglo de tramos horarios para generar cortes temporales en los datos.

```

# Define los tramos horarios que se van a procesar y generar los csv
# Estos deben tener la siguiente sintaxis XX-YY [XX-YY ...]
# donde XX e YY son números enteros de dos dígitos en el rango [00-23]
# y pueden ser iguales. Para concatenar varios tramos se usa el espacio. Ej:
  XX-YY ZZ-TT
TRAMOS=(06-09 18-21 00-23)

# CALCULO POR HORA DE LA SEMANA COMPLETA
for TRAMO in ${TRAMOS[@]}; do
  CONDICION=$(echo "$TRAMO" | sed -r 's/-/ AND /g')
  CONSULTA='copy (
    SELECT par_subida , par_bajada , SUM(peso)
    FROM (SELECT par_subida_1 AS par_subida ,
              par_bajada_1 AS par_bajada ,
              SUM(factor_expansion) AS peso
    FROM viaje_util
    WHERE netapa=1 AND
           extract(hour from tiempo_subida_1) BETWEEN $CONDICION
    GROUP BY par_subida_1 , par_bajada_1
    UNION
    SELECT par_subida_1 , par_bajada_2 , SUM(factor_expansion) AS peso
    FROM viaje_util
    WHERE netapa=2 AND
           extract(hour from tiempo_subida_1) BETWEEN $CONDICION
    GROUP BY par_subida_1 , par_bajada_2
    UNION
    SELECT par_subida_1 , par_bajada_3 , SUM(factor_expansion) AS peso
    FROM viaje_util
    WHERE netapa=3 AND
           extract(hour from tiempo_subida_1) BETWEEN $CONDICION
    GROUP BY par_subida_1 , par_bajada_3
    UNION
    SELECT par_subida_1 , par_bajada_4 , SUM(factor_expansion) AS peso
    FROM viaje_util
    WHERE netapa=4 AND
           extract(hour from tiempo_subida_1) BETWEEN $CONDICION
    GROUP BY par_subida_1 , par_bajada_4) AS viaje
  GROUP BY par_subida , par_bajada)
  To '$RUTA_DATOS_CSV/${TRAMO}_semana_viaje.csv' WITH DELIMITER ',' CSV;'

  sudo -u postgres -i psql -d memoria -c "$CONSULTA"
done

```

Figura 7.2: código que genera archivo CSV intermedio para la construcción del grafo en formato Pajek usando los datos de viaje. Fuente: elaboración propia

# Bibliografía

- Taras Agryzkov, Jose L. Oliver, Leandro Tortosa, and Jose F. Vicent. An algorithm for ranking the nodes of an urban network based on the concept of pagerank vector. *Applied Mathematics and Computation*, 219(4):2186 – 2193, 2012. ISSN 0096-3003. doi: <http://dx.doi.org/10.1016/j.amc.2012.08.064>. URL <http://www.sciencedirect.com/science/article/pii/S0096300312008570>.
- Alex Anas, Richard Arnott, and Kenneth A. Small. Urban Spatial Structure. *Journal of Economic Literature*, 36(3):1426–1464, September 1998. URL <https://ideas.repec.org/a/aea/jeclit/v36y1998i3p1426-1464.html>.
- Barry, Newhouser, R Rahbee, and A Sayeda. Origin and destination estimation in new york city with automated fare system data. *Transportation research record*, 1817:183–187, 2002.
- S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4–5):175 – 308, 2006. ISSN 0370-1573. doi: <http://dx.doi.org/10.1016/j.physrep.2005.10.009>. URL <http://www.sciencedirect.com/science/article/pii/S037015730500462X>.
- Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18):3825 – 3833, 2012. ISSN 1389-1286. doi: <http://dx.doi.org/10.1016/j.comnet.2012.10.007>. URL <http://www.sciencedirect.com/science/article/pii/S1389128612003611>. The {WEB} we live in.
- Alessio Cardillo, Salvatore Scellato, Vito Latora, and Sergio Porta. Structural properties of planar graphs of urban street patterns. *Phys. Rev. E*, 73:066107, Jun 2006. doi: 10.1103/PhysRevE.73.066107. URL <http://link.aps.org/doi/10.1103/PhysRevE.73.066107>.
- Antonio Gschwender, Marcela Munizaga, and Carolina Simonetti. Using smart card and gps data for policy and planning: The case of transantiago. *Research in Transportation Economics*, 59:242 – 249, 2016. ISSN 0739-8859. doi: <http://dx.doi.org/10.1016/j.retrec.2016.05.004>. URL <http://www.sciencedirect.com/science/article/pii/S0739885915300998>. Competition and Ownership in Land Passenger Transport (selected papers from the Thredbo 14 conference).
- Mr John Humphreys and Aoife Ahern. The influence of urban spatial structure on commuting and modal choice: evidence from the greater dublin area. In *Proc. Irish Transport Research Network*, page 11. ITRN2013, 2013.

- Jon M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, Aug 2000. ISSN 0028-0836. doi: 10.1038/35022643. URL <http://dx.doi.org/10.1038/35022643>.
- Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: A comparative analysis. *Phys. Rev. E*, 80:056117, Nov 2009. doi: 10.1103/PhysRevE.80.056117. URL <http://link.aps.org/doi/10.1103/PhysRevE.80.056117>.
- Thomas Louail, Maxime Lenormand, Oliva G. Cantu Ros, Miguel Picornell, Ricardo Herranz, Enrique Frias-Martinez, José J. Ramasco, and Marc Barthelemy. From mobile phone data to the spatial structure of cities. *Scientific Reports*, 4, June 2014. ISSN 2045-2322. doi: 10.1038/srep05276. URL <http://dx.doi.org/10.1038/srep05276>.
- Thomas Louail, Maxime Lenormand, Miguel Picornell, Oliva García Cantú, Ricardo Herranz, Enrique Frias-Martinez, José J. Ramasco, and Marc Barthelemy. Uncovering the spatial structure of mobility networks. *Nat Commun*, 6, Jan 2015. URL <http://dx.doi.org/10.1038/ncomms7007>. Supplementary information available for this article at [http://www.nature.com/ncomms/2015/150121/ncomms7007/supinfo/ncomms7007\\_S1.html](http://www.nature.com/ncomms/2015/150121/ncomms7007/supinfo/ncomms7007_S1.html).
- Marcela Munizaga, Flavio Devillaine, Claudio Navarrete, and Diego Silva. Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44:70–79, 2014.
- Marcela A. Munizaga and Carolina Palma. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile. *Transportation Research Part C: Emerging Technologies*, 24:9 – 18, 2012. ISSN 0968-090X. doi: <http://dx.doi.org/10.1016/j.trc.2012.01.007>. URL <http://www.sciencedirect.com/science/article/pii/S0968090X12000095>.
- Thomas A Navarrete C. Muñoz, V. and R. Contreras. Encuesta origen destino de santiago 2012: resultados y validaciones. *Ingeniería de Transporte*, 19(1):21–36, 2015. URL <http://ingenieriadetransporte.org/index.php/sochitran/article/view/141>.
- Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2): 167–256, 2003.
- Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.
- Juan Ortúzar, Luis G Willumsen, et al. *Modelling transport*. Wiley, Chichester, 2011.
- Antonio Páez and Darren M. Scott. Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, 61(1):53–67, 2005. ISSN 1572-9893. doi: 10.1007/s10708-005-0877-5. URL <http://dx.doi.org/10.1007/s10708-005-0877-5>.
- Marie-Pier Pelletier, Martin Trépanier, and Catherine Morency. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557 – 568, 2011. ISSN 0968-090X. doi: <http://dx.doi.org/10.1016/j.trc.2010.12.003>. URL <http://www.sciencedirect.com/science/article/pii/S0968090X1000166X>.

- Carlo Ratti, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H. Strogatz. Redrawing the map of great britain from a network of human interactions. *PLoS ONE*, 5(12):e14248, 12 2010. doi: 10.1371/journal.pone.0014248. URL <http://dx.doi.org/10.1371%2Fjournal.pone.0014248>.
- Salvatore Rinzivillo, Simone Mainardi, Fabio Pezzoni, Michele Coscia, Dino Pedreschi, and Fosca Giannotti. Discovering the geographical borders of human mobility. *KI - Künstliche Intelligenz*, 26(3):253–260, 2012. ISSN 0933-1875. doi: 10.1007/s13218-012-0181-8. URL <http://dx.doi.org/10.1007/s13218-012-0181-8>.
- J.-P. Rodrigue, C. Comtois, and B. Slack. *The Geography of Transport Systems*. Routledge, 3 edition, 2013. ISBN 978-0-415-82254-1.
- Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008. doi: 10.1073/pnas.0706851105. URL <http://www.pnas.org/content/105/4/1118.abstract>.
- Harold Soh, Sonja Lim, Tianyou Zhang, Xiuju Fu, Gary Kee Khoon Lee, Terence Gih Guang Hung, Pan Di, Silvester Prakasam, and Limsoon Wong. Weighted complex network analysis of travel routes on the singapore public transportation system. *Physica A: Statistical Mechanics and its Applications*, 389(24):5852 – 5863, 2010. ISSN 0378-4371. doi: <http://dx.doi.org/10.1016/j.physa.2010.08.015>. URL <http://www.sciencedirect.com/science/article/pii/S0378437110006977>.
- Strogatz, Steven H. Exploring complex networks. *Nature*, 410(6825):268–276, 2001. URL <http://dx.doi.org/10.1038/35065725>.
- Christian Thiemann, Fabian Theis, Daniel Grady, Rafael Brune, and Dirk Brockmann. The structure of borders in a small world. *PLoS ONE*, 5(11):e15422, 11 2010. doi: 10.1371/journal.pone.0015422. URL <http://dx.doi.org/10.1371%2Fjournal.pone.0015422>.
- Martin Trépanier, Nicolas Tranchant, and Robert Chapleau. Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*, 11(1):1–14, 2007.
- Duncan J. Watts and Steven H. Strogatz. Collective dynamics of /‘small-world/’ networks. *Nature*, 393(6684):440–442, Jun 1998. ISSN 0028-0836. doi: 10.1038/30918. URL <http://dx.doi.org/10.1038/30918>.
- Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11):2178–2199, 2014. doi: 10.1080/13658816.2014.914521. URL <http://dx.doi.org/10.1080/13658816.2014.914521>.