

# **IDENTIFICACIÓN DE LA ESTRUCTURA ESPACIAL URBANA A PARTIR DE DATOS PASIVOS DE TRANSPORTE PÚBLICO EN SANTIAGO DE CHILE**

Felipe Hernández, Universidad de Chile, [fhernand@dcc.uchile.cl](mailto:fhernand@dcc.uchile.cl)

Ricardo Hurtubia, Pontificia Universidad Católica de Chile, [rhg@ing.puc.cl](mailto:rhg@ing.puc.cl)

Marcela A. Munizaga, Universidad de Chile, [mamuniza@ing.uchile.cl](mailto:mamuniza@ing.uchile.cl)

## **RESUMEN**

La estructura espacial urbana es la disposición del espacio urbano generado producto de la forma urbana presente junto con las relaciones subyacentes entre éstas. Estas interacciones son generadas por el movimiento de personas entre un origen y destino. Aquí intentamos identificar esta estructura para Santiago de Chile mediante tres indicadores: centros, centros de pasada y comunidades usando análisis de redes y espacial sobre datos de tarjetas inteligentes. Los resultados muestran que la ciudad continua manteniendo su estructura con un CBD al centro, la existencia de potenciales nuevos centros y regiones con interacciones fuertemente locales, todas con presencia en el CBD.

*Palabras clave: estructura de viajes, comunidades, datos masivos*

## **ABSTRACT**

The urban spatial structure is the disposition of the urban space generated by of the urban form and subyacent interaction. This interactions are generated by movement of people between origin and destination. Here we try to identify the urban structure of Santiago of Chile throught three indicators: centers, hubs and communities using network science and spatial analysis over smart card data. The results show that the city keeps his CBD in the middle of the city, the existency of new potential centers and regions with strong interactions locally, all of them with presence in the CBD.

*Keywords: travel structure, communities, massive data*

## 1. INTRODUCCIÓN

La estructura espacial urbana se refiere a la disposición del espacio urbano generado a partir de la forma urbana presente en conjunto 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). El propósito de estos flujos involucra una gran variedad de objetivos tales como trabajo, uso de servicios públicos o privados, transmisión de información o capital, actividad económica, etc.

La estructura espacial de las ciudades ha sido formada principalmente por avances en transporte y telecomunicaciones (Anas *et al.*, 1998). La complejidad de los movimientos humanos ha redefinido el uso del espacio urbano y la disposición de los recursos. Según lo anterior, las personas pueden ser vistas como transportadores físicos entre áreas presentes en el espacio urbano y en consecuencia, ser usados como representantes de interacciones espaciales para la identificación de elementos esenciales de la estructura espacial urbana (Zhong *et al.*, 2014). Estos elementos son definidos por Zhong y corresponden a los siguientes:

- **Centros de pasada:** Se refiere a las áreas que sirven para conectar otro par de áreas para transferencia de personas. Funcionan como puentes espaciales entre distintas áreas.
- **Centros:** Se refiere a áreas que concentran personas o actividades. Pueden diferir de los Centros de pasada, pero a menudo, son lo mismo.
- **Comunidades:** Se refiere a zonas compactas de la ciudad que ofrecen un conjunto suficientemente diverso de actividades, oportunidades de trabajo y una población relativamente heterogénea en términos socioeconómicos que tiende a realizar sus actividades dentro de la misma zona, generando así una especie de “*ciudad dentro de la ciudad*”.

Por otro lado, los avances actuales en tecnología (GPS, RFID, Twitter, Foursquare, teléfonos móviles y tarjetas inteligentes) 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 en particular 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 son las tarjetas inteligentes, que corresponde a una implementación de AFC (Automated Fare Collection) usada en el sistema de transporte público y que está presente en varias ciudades del mundo (Pelletier. *et al.*, 2010).

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.

Utilizando la matriz origen-destino de viajes paradero a paradero, construimos un grafo dirigido con pesos a partir de los datos pasivos del transporte público de Santiago de Chile. Cada nodo en el grafo denota un área urbana indicada por un paradero de bus, los arcos indican la posibilidad de viajar entre dos áreas, y el peso de los arcos denota el volumen de viajes, que representa el número de viajes hecho. Hacemos uso de medidas de centralidad en grafos (“*Page Rank*” y “*Betweenness centrality*”) para detectar centros y centros de pasada, y estructuras de comunidad

en grafos para descubrir las comunidades presentes. Finalmente los resultados de este análisis son desplegados sobre un sistema de información geográfica (SIG) para visualizar la estructura espacial.

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) que hace uso de técnicas de análisis de redes y análisis espacial sobre los datos de tarjetas inteligentes del sistema de transporte público de Santiago de Chile, estructura desconocida hasta el momento. También se busca contrastar los indicadores evaluando otras medidas de centralidad.

Las modificaciones a la metodología propuesta por Zhong consistieron en la utilización de los datos de transbordo para el cálculo de *Betweenness Centrality* y al reemplazo de la detección de bordes<sup>1</sup> por el de comunidades dado que los resultados muestran que para Santiago (ciudad monocéntrica) se pierde información relevante. Otra diferencia es que se limita el análisis a una semana de un año específico y además se calculan otros indicadores para enriquecer el análisis (“*degree-in*” y “*stregth-in*”).

El resto de este trabajo está organizado como sigue. Una descripción de la metodología, indicadores y herramientas utilizadas es presentada en el capítulo 2. En el capítulo 3 se describen los datos utilizados. En el capítulo 4 se muestran los resultados obtenidos a partir de la implementación de la metodología descrita en el capítulo 2 usando los datos del capítulo 3, separados por indicador, además se analiza parte de los resultados con los obtenidos en Singapur. En el capítulo 5 se presentan las conclusiones de este estudio y en el capítulo 6 se especifican los pasos a seguir en el futuro a partir de este trabajo.

## 2. METODOLOGÍA

La metodología utilizada corresponde a la definida por Zhong *et al.* (2014) donde a partir de los datos recolectados por los sistemas AFC y AVL se construye un grafo dirigido con pesos desde el cuál se realizan los cálculos de las medidas de centralidad (*Page Rank* y *Betweenness centrality*) para comprender las posiciones relativas de cada zona de la ciudad. La generación de comunidades busca entender la organización espacial de los patrones de interacción generados dentro de la ciudad. Luego se realiza un análisis espacial usando la información geográfica asociada a los nodos. Por un lado, se busca asociar los valores de centralidad a la zona geográfica de influencia asociada a cada nodo (paradero), y por otra, identificar las comunidades más significativas que generan una partición de la ciudad.

---

<sup>1</sup> Límite geográfico que divide la ciudad de acuerdo las comunidades predominantes de cada zona (Zhong et al., 2014).

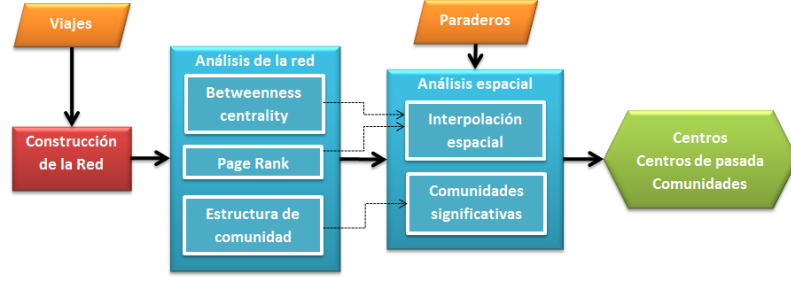


Ilustración 1: Metodología basada en Zhong et al. (2014). Fuente: elaboración propia.

## 2.1.Construcción de la red

Se define un grafo dirigido con pesos como  $G = (N, L, W)$  que representa todos los viajes realizados sobre cada par de paraderos.  $N$  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  $L$ .

## 2.2.Medidas de centralidad

Las medidas de centralidad son usadas para identificar posiciones relativas entre zonas de la ciudad, caracterizando los **centros de pasada** por medio de la medida *Betweenness centrality* y los **centros** a través de un análisis equivalente con el algoritmo *Page Rank* considerando los atributos de peso y sentido de todos los arcos presentes en la red.

### 2.2.1. Page Rank

La idea general de este algoritmo es medir el rol de un nodo (o zona dentro de la ciudad en nuestro caso) para atraer flujos desde todos los nodos de la red (otras zonas de la ciudad). Esta versión está basada en la utilizada por Rosvall y Bergstrom (2008) para medir la importancia de un nodo haciendo una analogía con el *Page Rank* de Google (Brin and Page, 1998).

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

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

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  $j$  ( $w_{ij}$ ), es decir:

$$p_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}, y \sum_j p_{ij} = 1 \quad (2)$$

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  $\{r_j\}$  puede ser calculada mediante diversos métodos (Zhong *et al.* 2014).

### 2.2.2. *Betweenness Centrality*

Esta medida de centralidad 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). El *Betweenness Centrality* de un nodo  $k$  se define como el número de caminos más corto que conecta otro par de áreas (nodos)  $i$  y  $j$  en el grafo que pasa por el nodo  $k$ . Un nodo tiene un valor alto de *Betweenness Centrality* si un mayor número de rutas más corta lo atraviesa.

$$Betweenness\ Centrality(k) = \sum_{ij} \frac{\delta_{ij}(k)}{\delta_{ij}} \quad (3)$$

Donde  $\delta_{ij}(k)$  es el número de rutas más cortas desde el nodo  $i$  al nodo  $j$  que pasan por el nodo  $k$ , mientras que  $\delta_{ij}$  corresponde al número de rutas más cortas desde nodo  $i$  a  $j$ . Lo anterior se realiza para cada par  $i, j$  distintos de  $k$ .

Dado que los datos utilizados (explicados en el capítulo 3) muestran los viajes hechos junto con los paraderos de transbordo, definimos una adaptación del *Betweenness Centrality* que los usa, descartando aquellas rutas entre dos paraderos que no estén presentes en los datos para luego compararlo con la definición anterior que usa el origen y destino. Esta nueva versión que llamamos *Betweenness Centrality Real* aprovecha la información conocida de los viajes y no solo los extremos. Según esto, lo definimos formalmente para un nodo  $k$  como:

$$Betweenness\ Centrality\ Real(k) = \sum_{ij} \frac{\delta_{ij}(k)}{\delta_{ij}} \quad (4)$$

Donde  $\delta_{ij}(k)$  es el número de alternativas de viajes entre los nodos  $i$  y  $j$  que pasan por el nodo  $k$ , mientras que  $\delta_{ij}$  corresponde al número de alternativas de viajes que hay entre el nodo  $i$  y  $j$ . Lo anterior se realiza para cada par  $i, j$  distintos de  $k$ .

## 2.3. Generación de comunidades

Existen varias formas para detectar comunidades en la literatura (Newman, 2003) pero una condición necesaria para este trabajo dada la naturaleza de los datos es la consideración de las

variables de densidad (cantidad de arcos) y flujo de interacciones (un peatón aleatorio en la red debiese moverse mayoritariamente dentro de la comunidad) al momento de crear éstas. Lo anterior basado en que estas dos variables deben ser mayor dentro de una comunidad y en términos de volumen, más grande que la presente entre éstas. Otra característica importante que deben presentar las comunidades es ser excluyentes dado que se busca identificar zonas de fuerte interacción bien definidas.

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

La idea detrás de este algoritmo consiste en generar recorridos aleatorios a lo largo de la red basado en la importancia de los nodos según el algoritmo *Page Rank* (el mismo explicado en 2.2.1), 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:

$$L(M) = q_{\sim}H(\mathcal{G}) + \sum_{i=1}^m p_{\cup}^i H(\mathcal{P}^i) \quad (5)$$

La ecuación (5) comprime dos términos: el primero es la entropía del movimiento entre módulos, y el segundo es la entropía de movimientos dentro de los módulos.  $q_{\sim}$  es la probabilidad de que el recorrido aleatorio cambie de módulo.  $H(\mathcal{G})$  es la entropía del nombre del módulo.  $H(\mathcal{P}^i)$  es la entropía de los movimientos al interior del módulo incluyendo su código de salida de éste.  $p_{\cup}^i$  es la fracción de movimientos que ocurren en el módulo  $i$  más la probabilidad de salir del módulo  $i$  tal que  $\sum_{i=1}^m p_{\cup}^i = 1 + q_{\sim}$ . El cómputo de este procedimiento se realiza mediante un algoritmo *Greedy* de búsqueda determinista.

## 2.4. Interpolación espacial

Por medio de la interpolación espacial se busca generar una imagen de movimientos sobre la ciudad que identifique las estructuras indicadas previamente. La idea es expresar las propiedades de cada área a partir del conjunto discreto de puntos geográficos que se tiene (paraderos de bus), para esto se asume que las personas escogen la parada más cercana a su destino. Si bien, existen varios tipos de interpolaciones que pueden reflejar este supuesto, se decide mantener la metodología de Zhong *et al.* (2014) y usar *Inverse Distance Weighted* (IDW), ésta considera que cada punto genera una influencia que disminuye con la distancia y se define como:

$$W_i(x, y) = 1/d_{ij}(x, y)^{\lambda} \quad (6)$$

Donde  $(x, y)$  es la coordenada geográfica de un punto de ubicación  $i$ ,  $W_i(x, y)$  es la contribución del punto  $i$  a su vecino  $j$ , y  $d_{ij}(x, y)$  es la distancia desde el punto  $i$  al punto  $j$ . El  $\lambda$  es un parámetro establecido en 2, por lo que sigue una ley inverso al cuadrado.

## 2.5.Herramientas utilizadas

Lo anterior fue realizado utilizando las siguientes herramientas:

- Igraph<sup>2</sup>: Herramienta para análisis de redes.
- MapEquation<sup>3</sup>: Framework para detección de comunidades basado en flujos.
- QGIS<sup>4</sup>: Sistema de información geográfica para llevar a cabo el análisis espacial.
- Colorbrewer<sup>5</sup>: Página web que recomienda paleta de colores para visualización de mapas.

## 3. DATOS

A continuación se detalla la forma en que funciona el sistema de transporte público (Transantiago) que origina los datos, como se estima el proceso de bajada y por último el detalle de los datos utilizados.

En el sistema de transporte público de Santiago de Chile se implementan los sistemas AFC (*Automated Fare Collection*) y AVL (*Automatic Vehicle Location*). El primero a través de tarjetas inteligentes, donde en buses es el único método de pago disponible y en metro es el más utilizado (sobre 97%), y el segundo por medio de la tecnología GPS (*Global Positioning System*) que monitorea los vehículos pertenecientes al sistema.

El sistema se caracteriza por tener cerca de 300 rutas de buses, 6000 buses disponibles agrupados en 6 operadores, aproximadamente 11.000 paraderos y una cifra que bordea los 150 kilómetros de rieles para el metro.

Hoy en día se sabe que se realizan aproximadamente 6.000.000 de validaciones durante un día laboral (por medio del sistema AFC), lo que genera una cifra cercana a los 35.000.000 de transacciones a la semana (incluyendo sábado y domingo) con aproximadamente 3.000.000 de tarjetas de pago. Por otro lado, hay 80.000.000 de emisiones proveniente de la tecnología AVL del sistema. Al procesar estos datos en conjunto es posible identificar el paradero de origen, recorrido utilizado para desplazarse (paraderos de transbordo) y el paradero de destino, este último requiere un procesamiento adicional basado en una metodología desarrollada por Munizaga y Palma (2012) que logra una identificación en el 80 % de las validaciones.

La muestra a utilizar corresponde al 80% de los datos procesados de la semana comprendida entre el 15/04/2013 al 21/04/2013 donde se logra una identificación de destino. Específicamente se hace uso de los siguientes datos:

---

<sup>2</sup> <http://igraph.org/>

<sup>3</sup> <http://www.mapequation.org/>

<sup>4</sup> <http://www.qgis.org/es/site/>

<sup>5</sup> <http://colorbrewer2.org/>

Nombre	Descripción
Viaje	Descripción de los viajes entre un punto de origen a uno de destino. Incluye la información de transbordos y horas de cada subida y bajada de bus/metro a lo largo de la ruta.
Red de paraderos	Contiene información geográfica asociada a cada paradero.
Red de metro	Contiene información geográfica asociada a cada estación de metro.

Tabla 1: Datos utilizados

#### 4. APLICACIÓN

Los datos consideran un período de una semana, de lunes a domingo, esto con la idea de capturar las zonas que son usadas para realizar una actividad, donde esta no está limitada a ser un hogar u trabajo sino también pueden ser identificadas como otras que realizan las personas durante el fin de semana. Adicionalmente revelan un comportamiento regular a lo largo de los días de la semana, 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.

Con el origen y destino de las tarjetas inteligentes se construye una red dirigida con pesos que contiene 11263 nodos y 2.421.076 arcos entre ellos que luego es asociada al espacio geográfico en que conviven, en éste cada nodo es representante de una zona de la ciudad con un paradero de bus en su interior.

La red construida es una parte para nuestro análisis, que por un lado, tiene un grafo que contiene información de interacciones entre distintos nodos desde el cuál se pueden calcular métricas que ayuden a entender su estructura, mientras que por otro se tiene información espacial de cada nodo, la combinación de ambos enfoques permite obtener una mirada completa de la estructura espacial de la ciudad (Ilustración 2.a y Ilustración 2.b).

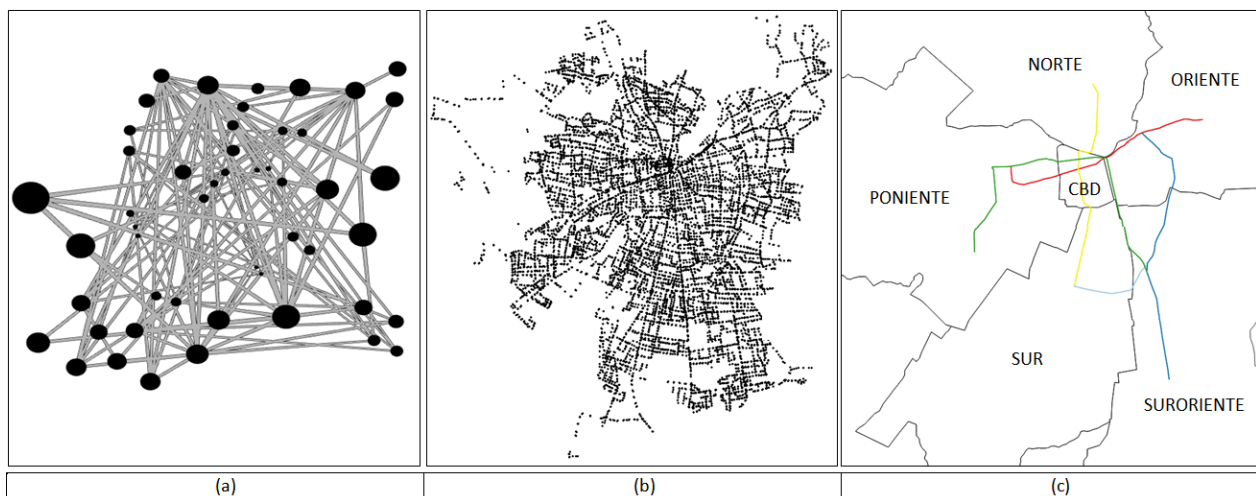


Ilustración 2: Distintos aspectos del análisis que en conjunto completan los datos faltantes. (a) grafo construido a partir de los datos de tarjetas inteligentes (se muestran nodos más relevantes). (b) Nodos de la red mapeados a su zona geográfica. (c) segmentación de la ciudad en regiones (líneas negras son los bordes) centro, oriente, suroriente, sur, poniente y norte.



Además se muestra el sistema de tren subterráneo presente en la ciudad donde cada color es una línea de metro. Fuente: elaboración propia.

#### 4.1.Centros

Los centros son determinados mediante el indicador de centralidad *Page Rank*, que mide la importancia de un nodo para atraer flujo en la red (sección 2.2.1). Con este indicador 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 Ilustración 3).

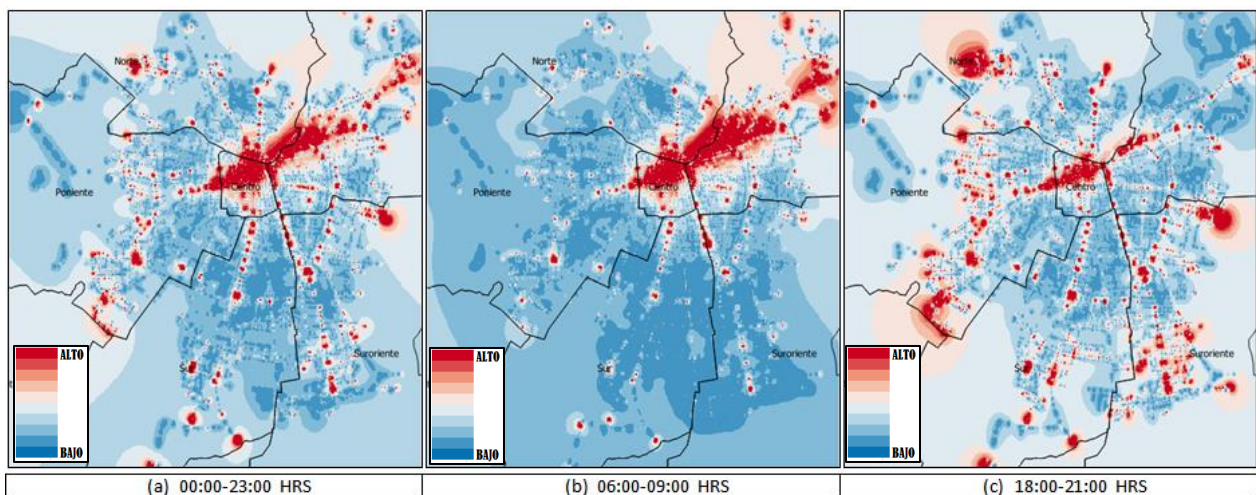


Ilustración 3: Identificación de centros (zonas de color rojo). La imagen (a) considera los datos de un día completo, mientras que (b) concentra los datos de mayor demanda durante la primera mitad del día y (c) concentra los datos de mayor demanda durante la segunda mitad del día. Fuente: elaboración propia.

La Ilustración 3.a muestra las zonas de destino preferida por las personas para realizar actividades, a partir de ésta se puede observar que gran parte de los destinos está fuertemente marcado por las zonas donde se encuentra disponible el acceso al transporte subterráneo (ver Ilustración 2.c), esto es algo esperable dado que estos proyectos de transporte son implementados en zonas de alta demanda, si bien se logra detectar éstas zonas, no es posible inferir la intencionalidad subyacente. Un propósito general puede ser inferido haciendo cortes temporales y considerando las dinámicas de flujo usadas por Munizaga y Palma (2012) y presente en otras ciudades donde gran parte de éste se mueve durante la mañana a su lugar de trabajo o actividad y en la tarde vuelve a su hogar, generando un patrón cíclico. Según lo anterior se puede hacer una distinción entre zonas de trabajo (Ilustración 3.b) y zonas de vivienda (Ilustración 3.c).

#### 4.1.1. Motivo de la detección de centro

Si bien el *Page Rank* por si solo aporta basta información respecto a la ciudad, es posible obtener un poco más a partir de la adición de otro indicador. El *Page Rank* de un nodo en la red, por la forma en que está definido, puede tener un alto valor por tres motivos:

- Nodo con alto *Page Rank* apunta hacia él (nodo C en Ilustración 4).
- Nodo presenta varios arcos que inciden en el (nodo E en Ilustración 4).
- Una combinación de los dos puntos anteriores (nodo B en Ilustración 4).

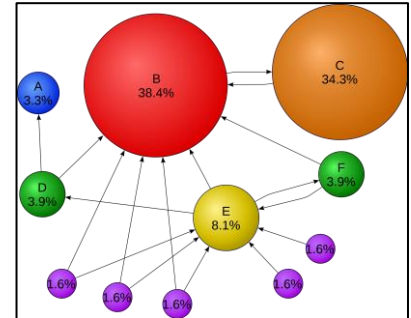


Ilustración 4: Page Rank en una red sin pesos. Fuente: Wikipedia.org

En el análisis realizado aquí se busca determinar aquellas zonas correspondientes a las alternativas b y c (sin diferenciación) de la enumeración anterior.

Las propiedades topológicas de la red pueden aportar información relevante sobre la interacción espacial. Desde el punto de vista de la conectividad, el grado de un nodo es una medición de su nivel de accesibilidad. Para un nodo  $i$ , su grado de entrada es definido como  $k_i = \sum_j^N a_{ji}$ , que corresponde al número de arcos que llegan a el,  $a_{ji}$  es 1 cuando hay un arco de  $j$  a  $i$ , 0 en otro caso. En la Ilustración 5 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.

De lo anterior es posible diferenciar zonas con alto *Page Rank* con pocos vínculos hacia el, mostrando que un fuerte flujo se mueve a el a través de pocas alternativas de viaje (Ver círculos demarcados en Ilustración 5).

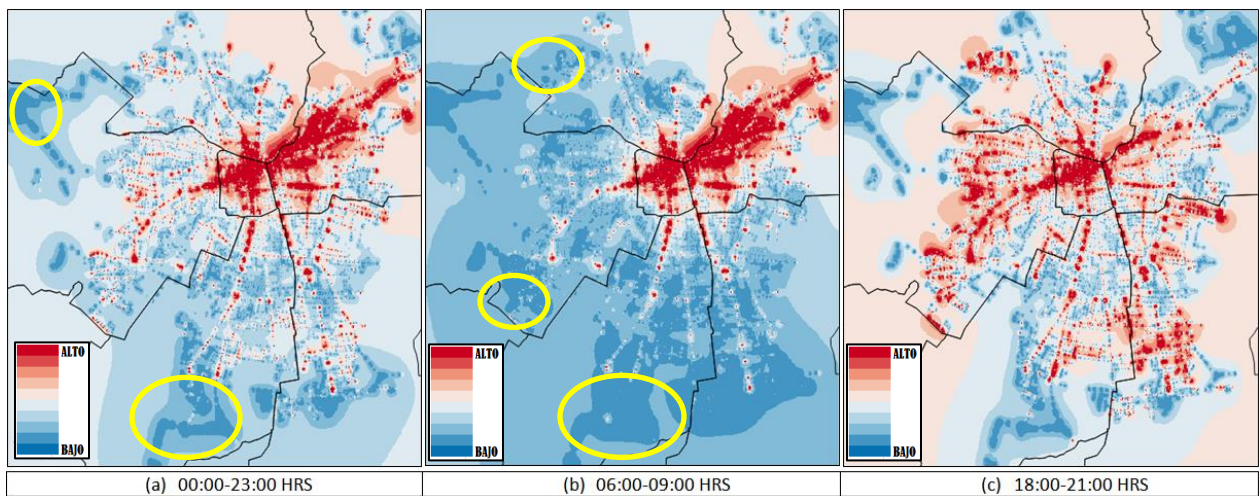


Ilustración 5: Visualización de los grados de entrada de un nodo. La imagen (a) considera los datos de todo un día. La imagen (b) contempla el período de alta demanda de la mañana y la imagen (c) el período de alta demanda de la tarde. Fuente: elaboración propia.

Un aspecto relevante a destacar es la medición de accesibilidad presente en la Ilustración 5.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.

#### 4.1.2. Suma de flujos de entrada sobre cada nodo

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 el,  $w_{ji}$  es el peso del arco cuando hay un arco de  $j$  a  $i$ , 0 en otro caso.

Al comparar la Ilustración 3 con la Ilustración 6 es posible observar en esta última la no detección de zonas (ver círculos amarillos en Ilustración 6) que si detecta el algoritmo *Page Rank*.

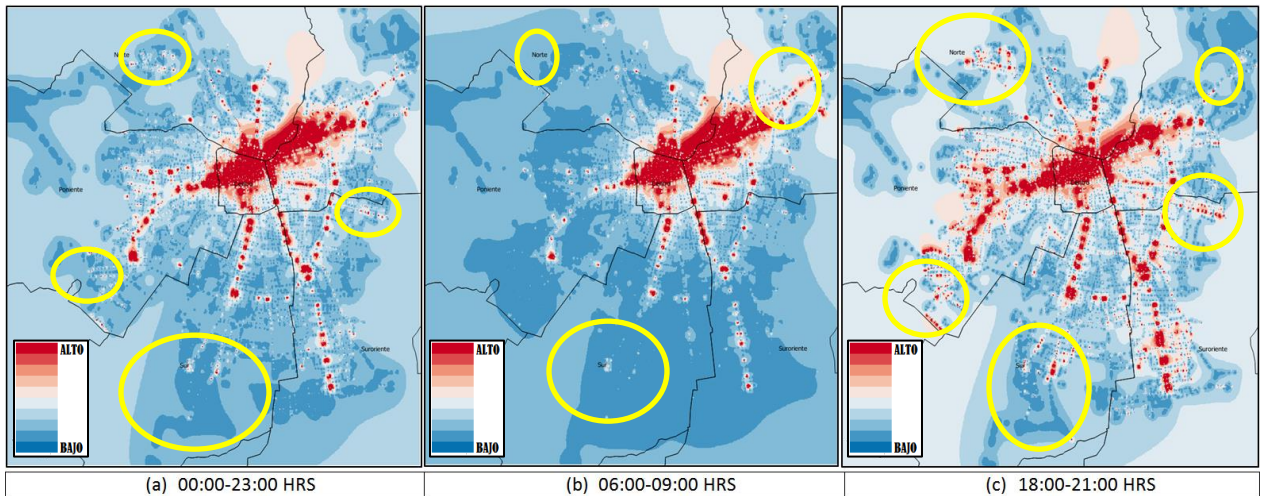


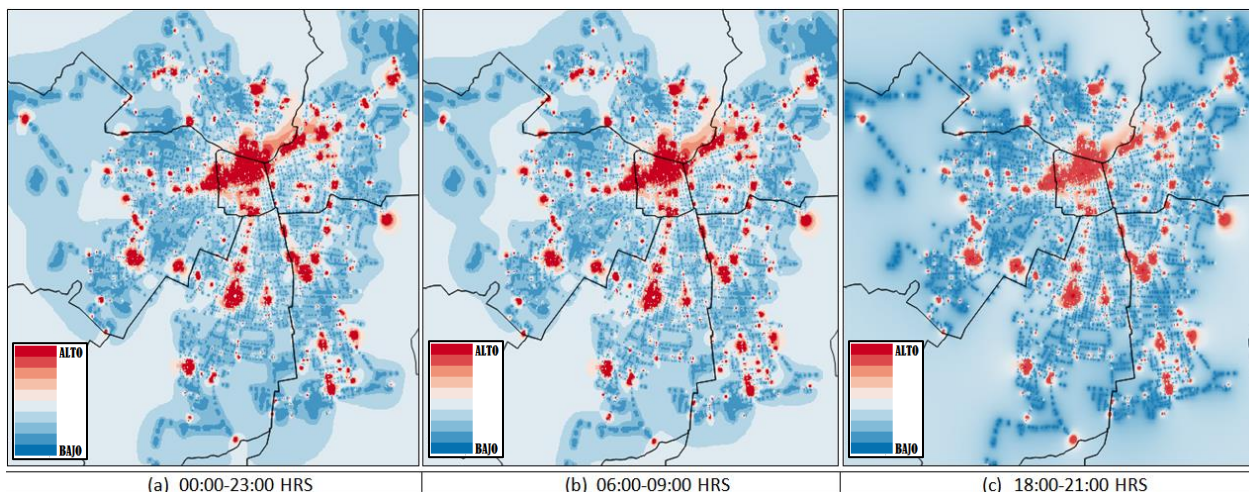
Ilustración 6: Visualización del flujo de entrada de un nodo. La imagen (a) contempla los datos de todo el día, (b) el período de alta demanda de la mañana y (c) el período de alta demanda de la tarde. Fuente: elaboración propia.

#### 4.2. Centros de pasada

Los centros de pasada corresponden a aquellas zonas de la ciudad que son usadas para realizar transbordos y son detectadas a través de la medida de centralidad *Betweenness Centrality*, que como se dijo antes (sección 2.2.2), mide la importancia de un nodo en la organización del flujo presente en la red. Las zonas identificadas pueden ser vistas en la Ilustración 7, la que ocupa los datos de origen y destino para construir el grafo. La estructura de las imágenes muestra una configuración que coincide en la estructura de las líneas de metro, lo que ubica a este medio de transporte como muy relevante dentro de la red de transbordos.



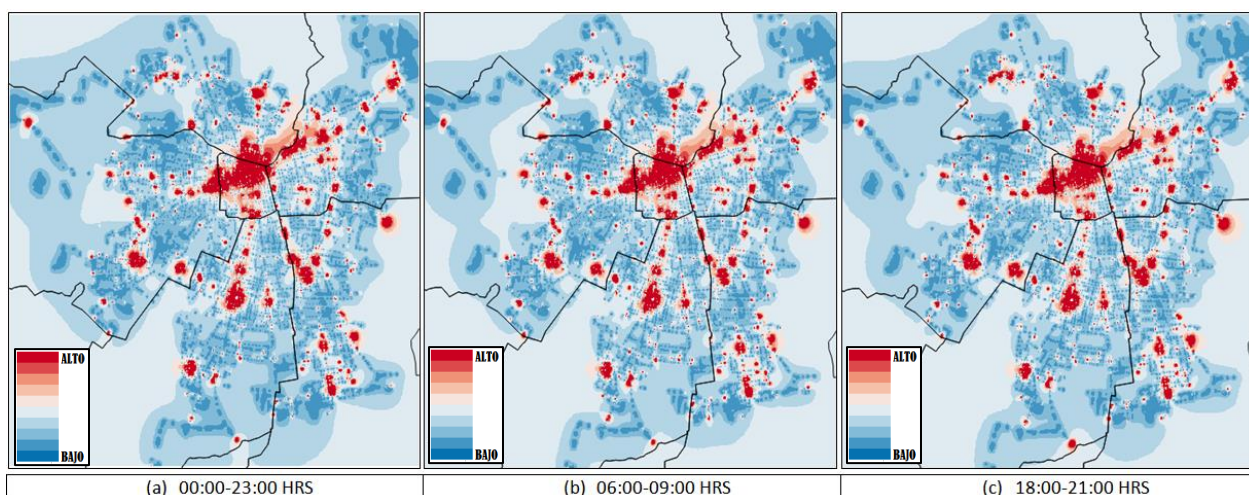
Otro fenómeno aparece al analizar los períodos de la mañana y tarde, donde las imágenes son casi idénticas en sus zonas de transbordo, lo que habla que los flujos presentes en ambos períodos por cada punto son muy parecidos, lo que indica que las personas no cambian en gran medida sus preferencias de viaje.



**Ilustración 7: Betweenness Centrality usando datos de origen y destino. Las zonas en rojo son las identificadas como zonas de transbordo. (a) datos de 00:00-23:00 HRS. (b) datos de 06:00-09:00 HRS. (c) datos de 18:00-21:00 HRS. Fuente: elaboración propia.**

#### 4.2.1. *Betweenness Centrality* usando datos de transbordo

Una forma de validar las visualizaciones anteriores es aplicando el mismo algoritmo sobre los datos de transbordo, la que es información real de los transbordos realizados por las personas.



**Ilustración 8: Betweenness Centrality usando datos de transbordo. Las zonas en rojo son las identificadas como zonas de transbordo. (a) datos de 00:00-23:00 HRS. (b) datos de 06:00-09:00 HRS. (c) datos de 18:00-21:00 HRS. Fuente: elaboración propia.**

Al comparar la Ilustración 7 con la Ilustración 8 se puede apreciar que no existen diferencias sustanciales que permitan observar cambios relevantes dentro de los centros de pasada. Los

centros de pasada detectados son idénticos 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.

#### 4.2.2. *Betweenness Centrality Real*

Esta variante del *Betweenness Centrality* considera dentro de su cálculo solo los pares origen-destino presente en los datos, es decir, aquellos que haya realizado al menos una persona. Con esta comparación se busca observar si el hecho de considerar todos los pares de nodos dentro del grafo (como lo hace el *Betweenness Centrality*) afecta la identificación de centros de pasada.

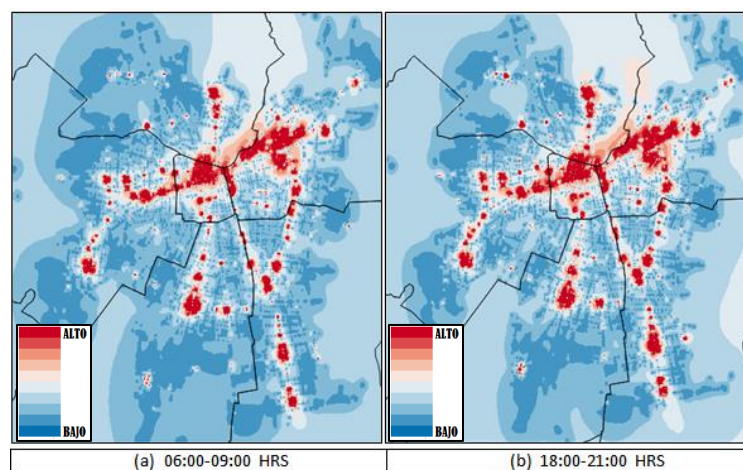


Ilustración 9: *Betweenness Centrality Real* usando datos de transbordo. Las zonas en rojo son las identificadas como zonas de transbordo. (a) datos de 06:00-09:00 HRS. (b) datos de 18:00-21:00 HRS. Fuente: elaboración propia.

La Ilustración 9 muestra que al considerar solo los pares de origen y destino dentro de los datos 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 pueden tener una existencia en la estructura espacial. La Ilustración 9 omite la imagen del día completo debido a errores presentados en su procesamiento.

### 4.3. Comunidades

Las comunidades generadas corresponden a una partición de la ciudad debido a intereses comunes como trabajo, recreación, edad, etc (newman 2003) que se ha ido formando a lo largo de los años. El algoritmo fue ajustado varias veces hasta obtener una estructura que 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 resultado es obra única de las interacciones presentes en la ciudad.



### 4.3.1. Comunidades detectadas

En la Ilustración 10 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. Otra característica es su expresividad dentro de la ciudad, existe un conjunto de comunidades que genera el anillo exterior de la ciudad (comunidades B,C,E,F,H,K) para que luego aparezca otro conjunto que toma presencia al interior del anillo (comunidades D,G,I,J).

Otro aspecto de estas comunidades es su tendencia a expresarse hacia el centro de la ciudad, todas de alguna manera se las arreglan para coexistir 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.

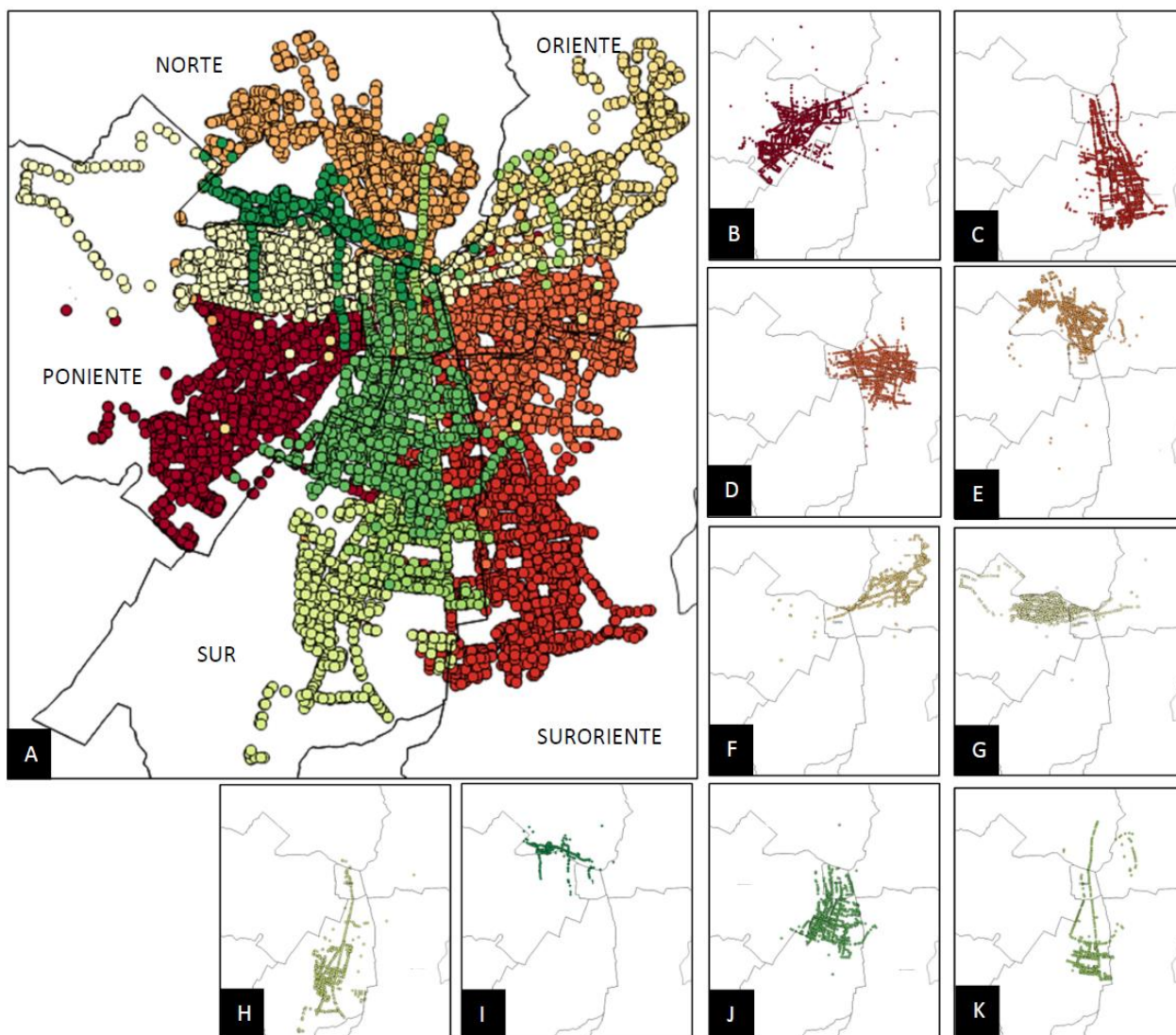


Ilustración 10: Comunidades más significativas detectadas con algoritmo *Infomap*. Cada cuadro muestra un conjunto de paraderos (puntos en la imagen) dentro de la ciudad asociados a una misma comunidad. El fondo de cada imagen

corresponde a las regiones centro (CBD), norte, oriente, suroriente, sur, extensión surponiente y poniente de la ciudad de Santiago de Chile. Fuente: elaboración propia.

#### **4.3.2. Comunidades y administración geo-política**

Otro aspecto importante a evaluar es la forma que dan estas comunidades a la ciudad y como se relacionan con los límites geo-administrativos. En la Ilustración 10 se puede observar que muchas de ellas mantienen gran parte de sus nodos en una sola región, identificando aquella zona como su área de predominancia. Cada región es subdividida en varias comunas que es administrada por una institución llamada Municipalidad, una comuna representa la unidad administrativa territorial mínima dentro de la ciudad. Desde el punto de vista de las comunas, ninguna comunidad detectada forma parte de una sola, lo que puede ser indicio de que las comunas tienen un grado de especialización tal que un grupo importante de personas debe salir de estas para encontrar otras actividades.

#### **4.3.3. 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 original. Mientras en la primera se observa una clara tendencia de todas las comunidades a desplazarse hacia el centro (ver Ilustración 10), dado por la existencia de una 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).

### **5. CONCLUSIONES**

La metodología propuesta por Zhong *et al.* (2014) aplicada sobre la ciudad de Santiago de Chile genera resultados que se ven prometedores pero se requiere validación. Es importante destacar del estudio que el uso de la misma metodología dio resultados distintos para ciudades que son distintas, lo que también resulta ser positivo.

Los datos pasivos de transporte público más las nuevas técnicas aplicadas en el ámbito de las ciudades han permitido conocer otro aspecto de la ciudad de Santiago de Chile que no se tenía precedente. Los indicadores muestran que sigue existiendo un gran CBD con una expresión centro-oriente de la ciudad, lo que destaca su comportamiento monocéntrico.

Los nuevos centros dados a conocer por el *Page Rank* tienen un mayor potencial de ser visitable, lo que no se correlaciona necesariamente con los niveles actuales de viajes que atraen. Esto se

podría interpretar como un potencial de atraktividad, que los planificadores podrían usar para decidir donde incentivar el desarrollo de nuevos sub-centros.

Los centros de pasada al parecer pueden ser detectados a partir del conocimiento del origen y destino de los viajes dado que las zonas de alta actividad también son las que presentan buena accesibilidad. 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, mostrando 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 Centrality tiene la desventaja que este considera todos los pares de nodos, lo cual no es necesariamente correcto para una ciudad por lo que debe ser validado.

Las comunidades parecen identificar sectores autocontenidos dentro de la ciudad generando grupos de éstas como anillo exterior y otro grupo de éstas como conjunto interior, aunque todas presentan una borde que se extiende al centro, mostrando que el CBD es alimentado por todas éstas. Además permite conocer 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.

El considerar solamente datos de transporte público se deja fuera otros medios de transporte existentes que podrían enriquecer los resultados.

## **6. Trabajo futuro**

En este trabajo se buscó identificar la estructura espacial de la ciudad de Santiago de Chile, dejando fuera un análisis temporal de éste, el que dado por la fuente de datos es posible abordar y puede dar indicios de otros aspectos que aquí se trataron tenuemente (corte temporal de mayor demanda en la mañana y tarde). Un trabajo más profundo puede ser la evaluación de estas métricas a lo largo de los años.

Validar si los nuevos centros que aparecen con el *page Rank* son efectivamente zonas que concentran actividades o son potenciales sub-centros es una tarea también pendiente.

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 (aquí se usaron deciles para agrupar los datos a ser visualizados). 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.

Otro aspecto pendiente es utilizar estos indicadores y asociarlos a unidades geográficas de manera tal que puedan ser comparables con otros estudios que permitan comunicar más información, por ejemplo, usando la zonificación EOD 2012, la que ya tiene otros indicadores.



Por último queda pendiente la tarea de incorporar nuevas fuentes de datos de viajes en otros modos (taxi, auto, bicicleta, etc) para tener una matriz de viajes más completas o hacer el mismo análisis con otras fuentes de datos.

## REFERENCIAS

Anas, A., Arnott, R., & Small, K. A. (1998) Urban spatial structure. **Journal of economic literature**, 36 (3), 1426-1464.

Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. **Computer Networks and ISDN Systems**, 30 (1–7), 107–117.

Lancichinetti, A. and Fortunato, S. (2009) Community detection algorithms: a comparative analysis. **Physical Review E**, 80 (5), 056117.

Louail, T., Lenormand, M., García, O., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. and Barthelemy, M. (2014) From mobile phone data to the spatial structure of cities. **Scientific reports**, 4:5276

Munizaga, M. and Palma, C. (2012) Estimation of a disaggregate multimodal public transport origin, destination matrix from passive smart card data from Santiago, Chile. **Transportation Research Part C**, 9-18.

Newman, M.E. (2003) The structure and function of complex networks. **SIAM Review**, 45 (2), 167-256.

Pelletier, M-P., Trépanier, M. and Morency, C. (2011) Smart card data use in public transit: A literature review. **Transportation Research Part C: Emerging Technologies**, 19, 4, 557-568.

Rodrigue, J.-P., Comtois, C. and Slack, B. (2013) **The geography of transport systems**. 3th Ed. New York: Routledge.

Rosvall, M. and Bergstrom, C.T. (2008) Maps of random walks on complex networks reveal community structure. **Proceedings of the National Academy of Sciences**, 105 (4), 1118–1123.

Zhong, C., Müller, S., Huang, X., Batty, M., Schmitt, G. (2014) Detecting the dynamics of urban structure through spatial network analysis. **International Journal of Geographical Information Science** 28(11).