



Universidad del Desarrollo
Facultad de Ingeniería

MOVILIDAD A TRAVÉS DE LA RED DE METRO DE SANTIAGO
EN TIEMPOS DE PANDEMIA

POR: ANDREA VERENA FAÚNDEZ QUEZADA

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dr. LEO FERRES

JUNIO 2021
SANTIAGO

Dedico este trabajo a mi esposo por su apoyo incondicional y a mi madre por abrirme las puertas al camino de la educación.

AGRADECIMIENTOS

Antes de presentar el trabajo con el que culmina este proceso de casi dos años, no puedo dejar de agradecer en primer lugar al Instituto de Data Science de la Universidad del Desarrollo y a Telefónica I+D Chile, por la confianza depositada en mí al permitirme analizar datos de telecomunicaciones en pos de aportar un granito de arena al estudio de relación entre movilidad y propagación del COVID-19 en la Región Metropolitana. Especialmente quiero agradecer al Dr. Leo Ferres, quien me guió en este sinuoso pero satisfactorio camino del entendimiento de lo que nos rodea.

También deseo expresar mi gratitud a todo el cuerpo docente del programa, quienes fueron indispensables en este proceso educativo, especialmente a Loreto Bravo por su disposición y capacidad de escucha, muchas veces fuera de horario.

Por último, pero no menos importante quiero agradecer a mi esposo por abastecerme de todos los pasteles que necesité para terminar este estudio, que espero sea de vuestro agrado.

ÍNDICE DE CONTENIDOS

RESUMEN	3
1. INTRODUCCIÓN	4
2. TRABAJO RELACIONADO	7
3. HIPÓTESIS Y OBJETIVOS	17
3.1. HIPÓTESIS.....	17
3.2. OBJETIVOS.....	17
4. DATOS Y METODOLOGÍA.....	19
4.1. DATOS	19
4.2. METODOLOGÍA.....	24
5. RESULTADOS.....	35
5.1. LIMPIEZA Y PREPROCESAMIENTO DE DATOS.....	35
5.2. ANÁLISIS EXPLORATORIO.....	37
5.3. GRAFOS PONDERADOS	45
5.4. CÁLCULO DE CORRELACIONES.....	55
5.5. CONTRASTE DE HIPÓTESIS	60
6. CONCLUSIONES.....	62
7. BIBLIOGRAFÍA	64
8. ANEXOS	69
ANEXO 1: ANTENAS POR ESTACIÓN	69
ANEXO 2: ESTACIONES CONSIDERADAS Y COMUNAS A LAS QUE SIRVEN	70
ANEXO 3: <i>BETWEENNESS CENTRALITY</i> DE LAS ESTACIONES CADA SEMANA	71
ANEXO 4: DISTRIBUCIÓN HORARIA METRO POR COMUNA, PREVIO AL 16/03/2020	72
ANEXO 5: DISTRIBUCIÓN HORARIA METRO POR COMUNA, POSTERIOR AL 16/03/2020	74

ÍNDICE DE FIGURAS

FIGURA 1. PROPORCIÓN DE VIAJES POR LÍNEA DE METRO.	6
FIGURA 2. REPRESENTACIÓN DE MODELO SIR.	8
FIGURA 3. CASOS NUEVOS CONFIRMADOS E ÍNDICE DE MOVILIDAD DIARIOS POR PROVINCIA.....	15
FIGURA 4. METODOLOGÍA PROPUESTA.....	24
FIGURA 5. EJEMPLO DE RECONSTRUCCIÓN DE VIAJE	28
FIGURA 6. UBICACIÓN DE ANTENAS Y COMUNAS ESTUDIADAS	36
FIGURA 7. DISPOSITIVOS ÚNICOS CONECTADOS POR DÍA	37
FIGURA 8. DISPOSITIVOS ÚNICOS CONECTADOS POR HORA EN DÍA LABORAL PROMEDIO SIN CUARENTENA	38
FIGURA 9. DISPOSITIVOS ÚNICOS CONECTADOS POR HORA EN DÍA LABORAL PROMEDIO CON CUARENTENA.....	40
FIGURA 10. COMPARACIÓN DE DISTRIBUCIÓN HORARIA DE DISPOSITIVOS EN EL METRO ANTES Y DESPUÉS DEL 16 DE MARZO DE 2020 EN LAS COMUNAS DE LAS CONDES, SANTIAGO Y SAN MIGUEL	41
FIGURA 11. CASOS CONFIRMADOS ACUMULADOS DE COVID-19 POR COMUNA	43
FIGURA 13. GRAFO RED DE METRO NO PONDERADO	45
FIGURA 14. GRAFO RED DE METRO PONDERADO POR CANTIDAD DE VIAJES ENTRE ESTACIONES, DESDE 02/03/2020 HASTA 05/07/2020.....	46
FIGURA 15. COMPARACIÓN DE CANTIDAD DE VIAJES ENTRE LAS DISTINTAS ESTACIONES DE METRO EN UNA SEMANA SIN CUARENTENA Y OTRA DURANTE LA CUARENTENA	49
FIGURA 16. GRAFOS SEMANALES DE ESTACIONES CON COLOR DE NODO SEGÚN CENTRALIDAD DE INTERMEDIACIÓN	51
FIGURA 17. GRAFOS SEMANALES DE COMUNAS CON COLOR DE NODO SEGÚN CENTRALIDAD DE INTERMEDIACIÓN NORMALIZADA	53

ÍNDICE DE TABLAS

TABLA 1. PERIODO DE ESTUDIO AGRUPADO EN SEMANAS.....	30
TABLA 2. CANTIDAD DE CASOS NUEVOS DE CONTAGIADOS CON COVID-19 CADA SEMANA.....	44
TABLA 3. TOP 10 NODOS CON MAYOR CENTRALIDAD DE INTERMEDIACIÓN DEL GRAFO NO PONDERADO Y PONDERADO POR CANTIDAD DE VIAJES.....	47
TABLA 4. TOP 10 NODOS CON MAYOR CENTRALIDAD DE PROXIMIDAD DE LA RED	48
TABLA 5. BETWEENNESS CENTRALITY NORMALIZADO DE COMUNAS POR SEMANA	53
TABLA 6. RESULTADOS CORRELACIÓN LINEAL ENTRE CENTRALIDAD DE INTERMEDIACIÓN Y CANTIDAD DE CASOS NUEVOS DE COVID-19	56
TABLA 7. RESULTADOS CORRELACIÓN DE SPEARMAN ENTRE BETWEENNESS CENTRALITY Y CANTIDAD DE CASOS NUEVOS DE COVID-19.	56
TABLA 8. RESULTADOS CORRELACIÓN DE PEARSON Y DE SPEARMAN ENTRE BETWEENNESS CENTRALITY Y CANTIDAD DE CASOS NUEVOS DE COVID-19 PARA CADA COMUNA	59
TABLA 9. RESULTADO PRUEBA DE HIPÓTESIS MUESTRAS DEPENDIENTES PARA DISMINUCIÓN DE CENTRALIDAD DE INTERMEDIACIÓN.....	61

Resumen

La enfermedad viral COVID-19 fue declarada como pandemia por la Organización Mundial (OMS) de la Salud el 11 de marzo de 2020, y ha recomendado diversas medidas a la población mundial y a los Gobiernos para frenar la propagación, dentro de las que se encuentra el distanciamiento social. En el caso de Chile se ha alternado entre cuarentenas dinámicas y totales, a fin de disminuir las interacciones en las comunas con mayor tasa de incidencia y evitar el desplazamiento desde, hacia y dentro de estas comunas. El presente estudio tiene por objetivo relacionar la movilidad en Metro y el aumento de contagios en las comunas de la Región Metropolitana a través de la centralidad de la red, para lo cual se utilizó el tráfico de datos entre dispositivos móviles y antenas que cubren la red de Metro de Santiago, y datos de contagios de COVID-19 semanales por comuna, en el periodo desde el 02 de marzo de 2020 hasta el 05 de julio del mismo año. Se construyeron redes direccionadas y ponderadas por la cantidad de viajes en Metro entre comunas, obteniendo como resultado que el *betweenness centrality* y la cantidad de casos nuevos de COVID-19 presentan una correlación positiva promedio de 0,1251 considerando 4 semanas de desfase entre los viajes y el aumento de casos. Las correlaciones más altas se obtuvieron para las comunas de La Reina, La Florida y Las Condes, con un coeficiente de correlación de Pearson promedio de 0,3421 para los distintos desfases considerados (1, 2, 3 y 4 semanas). Además, se obtuvo que la comuna de Santiago es la que posee mayor centralidad de intermediación en periodos sin restricciones y también durante cuarentenas, aun cuando la movilidad en Metro cayó un 80% aproximadamente.

1. Introducción

El día 11 de marzo de 2020, fue el día en que se declaró que el mundo estaba enfrentando una nueva amenaza: el COVID-19, enfermedad de carácter viral causada por el virus SARS-CoV-2. Si bien el virus se declaró como pandemia en esa fecha, la información oficial establece que el primer brote ocurrió en diciembre de 2019, en la ciudad de Wuhan, China[1]. El virus se expandió rápidamente por Asia, donde India acumula 29,3 millones de casos confirmados al 12 de junio de 2021. Luego se extendió hacia Europa, donde se esparció agresivamente en países como Francia con 5,8 millones de casos confirmados hasta la misma fecha. América no fue la excepción, donde Estados Unidos y Brasil cuentan con más de 33,5 y 17,3 millones de casos acumulados hasta el 12 de junio de 2021, respectivamente [2].

Con el fin de frenar los contagios en el mundo, la Organización Mundial de la Salud (OMS) entregó una serie de recomendaciones a la población mundial y a los Gobiernos, entre las que se encuentran lavado frecuente de manos con agua y jabón, uso de mascarilla y el distanciamiento social de al menos un metro[3].

Atendiendo a las recomendaciones emitidas por la OMS, el Gobierno de Chile, a través del Ministerio de Salud, optó inicialmente por una estrategia de cuarentenas dinámicas[4], para disminuir el desplazamiento desde, hacia y dentro de las comunas con mayores tasas de incidencia de COVID-19 en distintas ventanas de tiempo. Las decisiones de las comunas que entraban o salían de cuarentena se tomaban en base a los datos de tasa de contagios por comuna, los cuales se encuentran disponibles en el repositorio GitHub del

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación¹. Sin embargo, a partir del día viernes 15 de mayo de 2020 se optó por decretar cuarentena total en la mayoría de las comunas de la Región Metropolitana, medida que se renovó consecutivamente durante el periodo de estudio, desde el 02 de marzo de 2020 hasta el 5 de julio del 2020.

Si bien existen datos de contagios desagregados por comuna, no existen estudios acerca de la relación de transmisibilidad de la enfermedad entre las comunas. Desde esta perspectiva resulta relevante estudiar la transmisibilidad de la enfermedad a través de la red de Metro de Santiago, puesto que este es uno de los principales medios de transporte entre comunas de la Región Metropolitana. En efecto, los usuarios de transporte público utilizan el Metro en más del 60% de sus viajes, y la afluencia promedio en un día laboral es de 2,409 millones de viajes[5]. Considerando que la población de la provincia de Santiago asciende a 5,838 millones[6], la cifra de afluencia promedio de usuarios de Metro en días laborales indica que un 41% de la población utiliza este servicio de transporte diariamente. Cabe mencionar que los viajes en 2019 se redujeron en un 2,3% respecto de 2018, debido a que 19 estaciones de Metro quedaron fuera de servicio luego de daños causados en octubre de 2019[5], y que el Plan de Reconstrucción permitió volver a tener las estaciones operativas recién en septiembre de 2020, por lo que el periodo de estudio estuvo afecto a reducción de viajes por disponibilidad de estaciones[7]. En la Figura 1 se puede observar que la línea de Metro más utilizada es la Línea 1 con un 35,4% de los viajes, y le siguen las Líneas 2, 5 y 4 con 16,4% , 15,0% y 14.4%, respectivamente.

¹ <https://github.com/MinCiencia/Datos-COVID19>

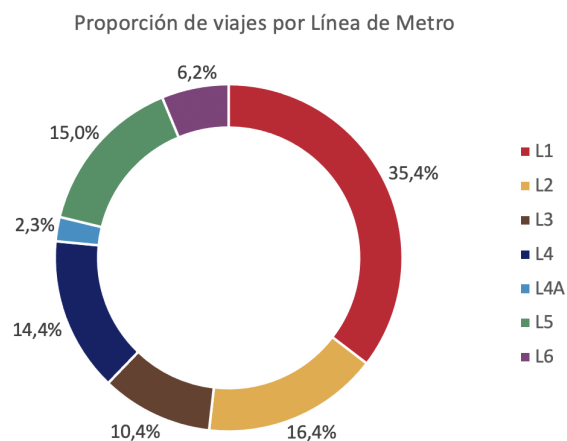


Figura 1. Proporción de viajes por Línea de Metro.
Fuente: Elaboración propia. Datos: Memoria anual 2020 de Metro de Santiago[7].

El presente estudio se limitó al análisis de viajes a través de las Líneas 1, 2, 4 y 5, donde se encuentran instaladas antenas que proporcionan cobertura en el Metro a clientes de telefonía móvil de la compañía Telefónica Movistar S.A., vale decir, las líneas de Metro estudiadas representan el 81,2% del uso total del servicio.

En el apartado de Hipótesis y Objetivos se plantea identificar una posible relación entre la movilidad en Metro y la cantidad de nuevos contagios de COVID-19 en las comunas de la Región Metropolitana. Para esto se utilizaron los datos que se detallan en la sección Datos y Metodología, que corresponden a conexiones entre dispositivos móviles y antenas ubicadas en las estaciones Metro, que al ser procesados permitieron cuantificar los viajes realizados entre las distintas estaciones. Con lo anterior, se construyeron grafos semanales donde los vértices correspondían a las estaciones, mientras que las aristas ponderadas representaron los viajes. Luego se calcularon las medidas de centralidad de intermediación y su relación con la cantidad de contagios, cuyos resultados se presentan en el capítulo de Resultados, para finalmente cerrar con las conclusiones y el trabajo futuro.

2. Trabajo Relacionado

El presente análisis tiene por objeto buscar relaciones entre la movilidad en Metro y contagios nuevos de la enfermedad COVID-19 por comuna, a través de métricas aportadas por la Teoría de Redes. Sin embargo, la enfermedad COVID-19 es una epidemia y, como tal, se vuelve necesario realizar una revisión de los modelos no sólo desde la óptica de Teoría de Redes, sino también de los modelos epidemiológicos que se han planteado hasta la fecha.

El primer modelo epidemiológico se remonta a 1927 con la investigación de Kendrick y McCormick, quienes proponen el modelo SIR[8], nombrado de esa forma por la abreviación de susceptibles, infectados y removidos, que son justamente las tres clasificaciones que realizan para contar a la población que: no está contagiada, la que está contagiada y puede seguir propagando la enfermedad, y la que ya no puede propagar la enfermedad ni contagiarse de nuevo, respectivamente. Este modelo se centra en el estudio de las derivadas parciales, es decir, las tasas de cambio de cada una de estas tres categorías, considerando que un individuo puede pasar de susceptible a infectado y posteriormente de infectado a removido; donde la suma de los tres estados en un determinado tiempo t corresponde a la población N . El modelo se resume en la Figura 2, donde $\langle k \rangle$ corresponde al número de contactos de un individuo típico, i es la fracción de individuos infectados, r es la fracción de individuos recuperados/removidos, β corresponde a la probabilidad de que se transmita la enfermedad de un individuo infectado a uno susceptible en un determinado momento t y μ es la tasa de recuperación de los individuos infectados.

Esta propuesta es bastante simple y no considera, por ejemplo, la posibilidad de que un individuo se contagie nuevamente o tenga una reinfección. Esto sí ha sido reportado en individuos que presentan COVID-19, aunque no existe claridad acerca de si se trataría de una reinfección o reactivación del virus [9][10]. No obstante, el modelo SIR es un modelo flexible en el sentido de su capacidad para extenderse o complementarse con otros modelos, razón por la cual sigue siendo una de las herramientas más utilizadas para modelar epidemias hasta la actualidad [11][12][13].



b.

$$\frac{ds}{dt} = -\beta \langle k \rangle i [1 - r - i]$$

$$\frac{di}{dt} = -\mu i + \beta \langle k \rangle i [1 - r - i]$$

$$\frac{dr}{dt} = \mu i$$

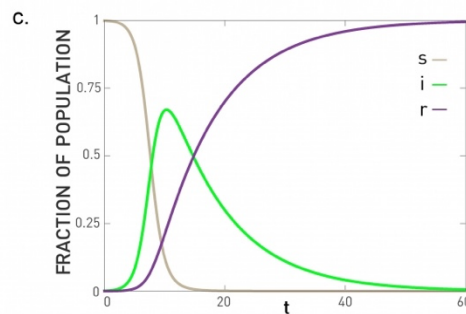


Figura 2. Representación de Modelo SIR.
Fuente: Barabási, A.L. Network Science (2013)[14]

Desde la óptica de la Teoría de Redes, existen estudios acerca de cómo se propaga un virus a través de los sistemas de transporte. Un ejemplo de lo anterior es la investigación de Colizza, Barrat, Barthélemy y Vespignani en 2006, donde se estudia la propagación de un virus a través de los vuelos[13]. En este estudio, se construyó una red global de transporte aéreo, donde los nodos correspondían a los aeropuertos del mundo; y las aristas, a los viajes entre los aeropuertos. Además, se complementó la data de nodos y aristas con data demográfica del número de habitantes de las áreas que son atendidas por cada aeropuerto, relacionando la población de un área metropolitana con los viajes que tienen por origen/destino el aeropuerto que las atiende. Los autores utilizan el enfoque del modelo SEIR, derivación del SIR que incorpora un estado de latencia o exposición E , donde el individuo se encuentra incubando la enfermedad y, por tanto, aún no tiene capacidad de contagiar a otros. Sin embargo, para no limitarse a un modelo determinístico, los autores aplicaron distintos métodos matemáticos estocásticos para modelar pasajeros que tienen una probabilidad determinada de estar en estado susceptible(S), latente (E), infectado (I) o recuperado (R) en un tiempo t dado, y que tienen también una probabilidad de viajar desde una ciudad a otra ciudad en un determinado momento. Para modelar esto, propusieron una variable aleatoria estocástica con distribución multinomial y así generaron las aristas, representando a los potenciales pasajeros que viajan de una ciudad a otra, considerando además las tasas de ocupación de vuelos y capacidad en términos de cantidad de asientos disponibles para cada conexión. También incluyeron variables relacionadas con la enfermedad viral: la tasa de transición de un estado a otro, como por ejemplo, de latente a infectado, considerando período de incubación de la enfermedad y

posibilidad de interacción entre los individuos, modelada con operadores de transporte y las ecuaciones epidémicas de Langevin[15]. Con este estudio los autores muestran que el comportamiento aparentemente errático de la propagación de una enfermedad por el mundo, en realidad sí tiene un patrón que se puede modelar a través de redes de transporte aéreo y modelos matemáticos complejos de gran escala. Este enfoque permitiría realizar predicciones acerca de los patrones epidémicos de una enfermedad, pero para aplicarlo con precisión se deben incluir variables relacionadas a la dinámica de la enfermedad, factores estacionales y capacidades de los sistemas de salud de cada país. Adicionalmente, se señala que el estudio podría ampliarse hacia otros sistemas de transporte como carreteras y vías férreas. Cabe señalar que en el presente estudio de viajes a través de la red de Metro, se propone utilizar la población de las comunas de manera análoga a lo propuesto por Colizza et al.[14] donde aeropuerto es a estación de Metro, como población del área metropolitana atendida por aeropuerto es a la población de la comuna atendida por la estación de Metro.

Más tarde, en el año 2007, Colizza, Barrat, Barthélemy, Valleron y Vespignani señalaron que el transporte aéreo juega un rol fundamental en los brotes de enfermedades virales globales, pero que la estrategia de contención de brotes debe basarse en protocolos terapéuticos antivirales, porque el implementar restricciones de viaje no es sólo difícil de llevar a la práctica y costoso económicamente, sino que además no tiene gran impacto en la propagación viral, pues sólo ralentiza unos días o semanas la evolución general de la pandemia[16].

Posteriormente, Balcan, Colizza, Gonçalves, Hud, Ramasco y Vespignani presentan un estudio en 2009, donde no sólo se considera el desplazamiento a gran escala mediante tráfico aéreo, sino también la movilidad a corta escala. Los autores señalan que el patrón de la epidemia a gran escala, es decir global, puede explicarse sólo con la movilidad de tráfico aéreo, siendo despreciable el impacto de la movilidad a corta escala en el patrón de propagación de la enfermedad a nivel mundial. Sin embargo, si se trata de explicar el comportamiento epidémico local, en la periferia de la infraestructura de transporte aéreo, entonces la movilidad de corto alcance sí se convierte en un factor clave[17]. Dentro de los métodos utilizados en este estudio se destaca la Tesselación similar a la de Voronoi, que permite dividir el mapa en áreas geográficas censales alrededor de los principales ejes de transporte dentro de un mismo país, a fin de estudiar el comportamiento de subpoblaciones. También se presentan componentes probabilísticos asociados a la posibilidad de que un individuo pueda ser asintomático o no (complementarias) o de que un individuo pueda viajar dado que no presenta síntomas. Adicionalmente se incorporan componentes estocásticas con variables aleatorias de distribución binomial y multinomial, entre las que destacan la tasa con la cual se contrae la enfermedad β , causando que los susceptibles pasen a estado de latencia; y la tasa μ con que los individuos pasan de estado infeccioso a recuperado/removido. Las variables anteriores permiten calcular el que es uno de los parámetros más relevantes y que ha sido mencionado en otras investigaciones tanto previas como posteriores [18][19][20]: el número básico de reproducción R_0 , que corresponde al promedio de nuevos infectados que un individuo con la enfermedad genera cuando se encuentra en una población completamente susceptible. Por otra parte, los

autores proponen un estimador para la fuerza de la infección, la cual se encuentra estrechamente relacionada con los desplazamientos y sus efectos en la propagación de la infección[20]. Cabe señalar que en el presente estudio de transmisibilidad del COVID-19 a través de la red de Metro se busca relacionar justamente el efecto que tiene el desplazamiento entre distintas comunas, con el aumento de casos en dichas comunas.

Existen también estudios que buscan explicar y predecir la propagación de un virus a nivel local, dentro de un país. Entre estos se destaca el de los autores Barrio, Varea, Govezensky y Marco, quienes en 2013 propusieron un modelo matemático de propagación local basándose en parámetros de movilidad de los individuos[21]. Para ellos generaron una cuadrícula bidimensional con celdas ponderadas por la densidad poblacional correspondiente. Dentro de los parámetros que dominan los efectos de transmisibilidad del modelo se encuentran nuevamente el coeficiente de transmisión β , el periodo de incubación, de infección, y a diferencia de otros modelos, considera un periodo de inmunidad ω luego del cual un individuo puede pasar de un estado de recuperado a susceptible nuevamente. Con los parámetros anteriores los autores establecen las ecuaciones de susceptibles, latentes, infectados y recuperados para el tiempo $t+1$, y además, proponen una fórmula para el cálculo del número básico de reproducción R_0 , que como se ha mencionado antes, es clave en la dinámica de propagación de la enfermedad, ya que corresponde el número promedio de individuos susceptibles a los que puede contagiar un individuo infectado. Por otra parte, los autores abordan la comunicación entre poblaciones de distintas cuadrículas como v_t , donde esta cantidad corresponde al promedio normalizado de viajes entre celdas por día; esta variable permite realizar

simulaciones de Monte Carlo para decidir si en un momento dado las personas infecciosas en una cuadrícula transmiten la enfermedad a las cuadrículas vecinas. Además, incorporan el parámetro β , que está relacionado con la infecciosidad de la enfermedad. El modelo fue ajustado y luego validado con los datos del virus de la influenza AH1N1 que afectó a México en 2009, logrando una concordancia entre lo modelado y los datos reales[21].

Respecto de estudios de propagación del COVID-19, Barrio, Kaski, Haraldsson, Aspelund y Govezensky han generado recientemente un modelo de propagación geográfica del virus a nivel local de acuerdo a patrones de movilidad de los individuos. Se hizo de manera similar al estudio mencionado anteriormente, pero esta vez, para el virus causante de la enfermedad COVID-19, aplicando el modelo a tres países con realidades culturales distintas: México, Islandia y Finlandia[22]. El estudio recoge el enfoque anterior de utilizar una cuadrícula bidimensional ponderada por la densidad de la población de cada celda para la movilidad a corta distancia, con su parámetro de movilidad terrestre v_t , donde cada celda tiene su sistema de ecuaciones diferenciales SEIRS que representan la dinámica local del virus. Considera también la dinámica entre celdas no adyacentes, cuya conexión es proporcional al número de pasajeros diarios que viajan de un aeropuerto a otro, localizando cada aeropuerto en la celda correspondiente y estableciendo un parámetro de movilidad área v_a . El planteamiento de este modelo permite realizar una distinción entre los parámetros biológicos (como tasas de latencia) y los parámetros socioculturales (como la movilidad), razón por la cual los parámetros biológicos fueron ajustados con los datos de México desde marzo a abril, y se aplicaron a los tres países (ya que la enfermedad era el COVID-19 para los tres); no obstante, los

parámetros de movilidad fueron calculados para cada país, según las decisiones restrictivas de movilidad implementadas por los respectivos gobiernos. Las predicciones realizadas en los tres países, considerando las distintas medidas de restricción de movilidad, muestran que el promedio de casos nuevos confirmados en los últimos 7 días se mantiene dentro o muy cercano al intervalo de confianza del 99% que el modelo predijo; esto para los tres países, aunque con mejores resultados para Islandia, donde la población es menor. El modelo propuesto, gracias a sus componentes estocásticas, permite realizar predicciones en distintos países, variando los parámetros acorde a la realidad local[22]. Cabe señalar que dentro de los hallazgos destacados se encuentra que las restricciones oportunas de movilidad y distanciamiento social son eficientes, especialmente en países homogéneos en población, donde la gente obedece las normas establecidas.

En relación a la efectividad de las restricciones de movilidad, Ferguson et al. han presentado recientemente un estudio donde se prueba que las medidas de distanciamiento social y confinamiento, así como las restricciones y recomendaciones públicas fueron efectivas para frenar la velocidad del aumento de casos nuevos en la población de China[23] hasta marzo del 2020. Para este estudio, los autores reunieron los reportes de casos nuevos confirmados diariamente desde el 1 de enero hasta el 17 de marzo de 2020, en las provincias chinas de Guangdong, Henan, Hubei, Hunan y Zhejiang, y en Hong Kong, data que les permitió calcular el número de reproducción del virus en función del tiempo R_t . También se utilizó datos de movilidad de las localidades considerando la misma ventana temporal, para relacionar la efectividad de restricciones con la disminución

de casos nuevos. Dentro de las dificultades del estudio que afectan al cálculo de R_t se encuentran: el desfase temporal que existe desde que un individuo contrae la enfermedad hasta que presenta los primeros síntomas y, por otra parte, el periodo que transcurre desde que una persona se toma el test hasta que éste finaliza, se registra y se convierte en un dato asociado a una fecha posterior a la de la toma de la muestra. Estos desfases fueron cuantificados a través de correlaciones cruzadas entre la movilidad diaria y el R_t diario. Los datos del estudio se encuentran en la Figura 3, donde se puede apreciar un patrón cíclico en Beijing y Zhejiang que se debe a la disminución de viajes los fines de semana.

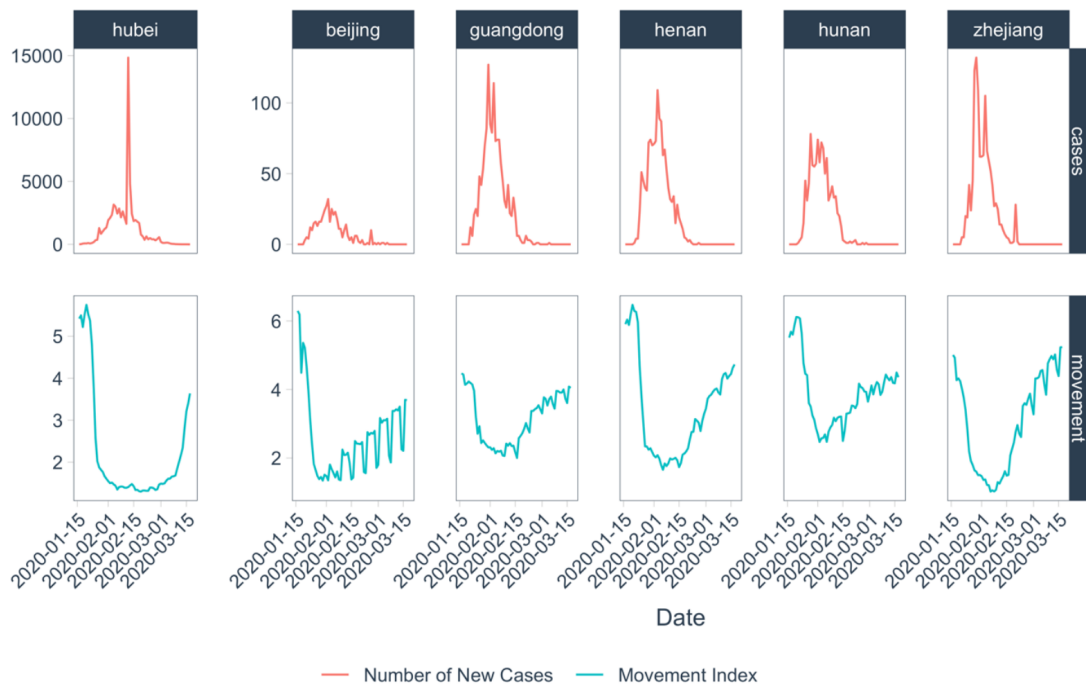


Figura 3. Casos nuevos confirmados e índice de movilidad diarios por provincia.
Fuente: Ferguson et al. Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment (2020)[23]

Dentro de los principales hallazgos de este estudio se encuentra que a medida que se implementaron restricciones de movimiento dentro de China continental (fines de enero a inicios de febrero de 2020), el movimiento en las ciudades y el R_t estaban altamente correlacionados en el mismo sentido. No obstante, con el aumento de la movilidad, la correlación entre el movimiento dentro de la ciudad y el R_t disminuyó abruptamente y se volvió fuertemente negativa. Estos resultados proporcionan evidencia de que las estrategias de contención de China continuaban siendo efectivas a medida que reiniciaban su economía, considerando el índice de movilidad como un indicador de la reactivación económica.

3. Hipótesis y Objetivos

3.1. Hipótesis

En el presente estudio se propone probar lo siguiente:

- 1) La cuarentena por COVID-19 están asociadas a un cambio en los patrones de viajes de la Red de Metro, haciendo que las estaciones, y por ende, las comunas, disminuyan en promedio su centralidad de intermediación con la cuarentena.
- 2) Las comunas donde se localizan las estaciones de Metro con mayor centralidad de intermediación dada por los viajes semanales, mostrarán un aumento mayor en los casos nuevos de COVID-19 semanales en la semana siguiente o las semanas subsiguientes.

3.2. Objetivos

Para probar las hipótesis planteadas se formulan los siguientes objetivos:

1. Cuantificar el número de viajes semanales por comuna, a través del seguimiento del tráfico de datos entre equipos móviles y antenas que dan cobertura a dispositivos localizados en las estaciones de Metro.
2. Generar grafos semanales de la red de Metro, ponderados por el número de viaje entre estaciones y calcular las métricas de centralidad de proximidad e intermediación de la red.

3. Cuantificar la relación entre la centralidad de intermediación y la cantidad de casos nuevos de COVID-19 para las comunas de manera agrupada y cada comuna por separado, considerando desfases de 1 a 4 semanas desde la fecha del viaje hasta la fecha de casos nuevos confirmados.

4. Datos y Metodología

4.1. Datos

En la presente investigación se utilizó un total de cinco datasets que corresponden a:

- Antenas: referenciado como archivo “metro_antennas.csv”
- Tráfico de conexiones entre dispositivos y antenas: referenciado como “tráfico.csv”
- Serie temporal de casos COVID-19 por comuna: Covid-19_T.csv”
- Cuarentenas históricas por comuna: “Cuarentenas-Historicas.csv”

La firma Telefónica S.A, que según el Análisis de Tráfico Internet Enero-Abril 2020 de la Subsecretaría de Telecomunicaciones, tiene el 24,6% de participación en el mercado de telefonía móvil[24], proporcionó archivos de carácter privado en formato “.csv”, es decir, de valores separados por comas. El primer archivo a analizar es el “metro_antennas.csv”, el cual contenía data no ordenada de las antenas, localización y descripciones asociadas, por lo que se pre-procesó a fin de obtener los datos organizados en las siguientes columnas:

- tower_id: corresponde al identificador de la torre, que consta de una combinación de cinco caracteres alfanuméricos. No permite valores nulos y es tipo string.
- lat: corresponde a la latitud en que se ubica la torre, es un dato de tipo float redondeado a la diezmilésima.
- lon: corresponde a la latitud en que se ubica la torre, es un dato de tipo float redondeado a la diezmilésima.

- line: es un dato de tipo string que indica la línea de Metro a la que pertenece la estación atendida por la torre.
- station: es un dato de tipo string que indica la estación de Metro a la que atiende la torre.
- comuna_id: dato de tipo numérico entero, de 5 caracteres, que corresponde al identificador de la comuna en la cual se sitúa la torre.
- comuna_desc: dato de tipo string que corresponde al nombre de la comuna en la cual se sitúa la torre.
- region_id: dato de tipo numérico que corresponde al identificador de la región en la cual se sitúa la torre.
- region_desc: dato de tipo string que corresponde al nombre de la región en la cual se sitúa la torre.

El archivo “tráfico.csv”, que también corresponde a data de carácter privado, contenía el tráfico entre equipos de telefonía móvil y las antenas ubicadas en torres. Los campos contenidos en este archivo son:

- device: dato de tipo string que corresponde al identificador del equipo móvil, codificado en una combinación alfanumérica de 64 caracteres. Cabe destacar que esta codificación resulta de un proceso de anonimización, por lo cual no es posible individualizar personas ni números telefónicos.
- cell_id: identificador de la antena ubicada en la torre, que corresponde a una combinación de cinco caracteres alfanuméricos, por tanto es de tipo string.

- `datetime`: dato de tipo `timestamp` que corresponde a la fecha y hora en que se realiza la transacción entre el dispositivo móvil y la torre.
- `bts_id`: dato de tipo `string` que corresponde al identificador de las torres, codificado en una combinación alfanumérica de 5 caracteres.
- `geo_area_id`: dato de tipo numérico entero, de 5 caracteres, que corresponde al identificador de la comuna donde se ubica la torre.

La segunda fuente de datos es de acceso público y corresponde al GitHub `MinCiencia/Datos-COVID19`² disponibilizado por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. En el repositorio se encuentra la carpeta “Data Product 1 - Casos totales por comuna incremental” con los datos de informes epidemiológicos emitidos por el Ministerio de Salud de Chile, en forma organizada y estructurada en archivos de valores separados por coma. Cabe mencionar que la data fue generada con una frecuencia de 2 a 3 días y es de carácter público. En este caso se utilizó el archivo “`Covid-19_T.csv`”, cuyo contenido es el siguiente:

- `region`: dato de tipo `string` que corresponde a la región a la que se hace referencia.
- `codigo_region`: dato de tipo entero que corresponde al código de la región a la que se hace referencia.
- `comuna`: dato de tipo `string` que corresponde a la comuna a la que se hace referencia.
- `codigo_comuna`: dato de tipo entero que corresponde al código de la comuna a la que se hace referencia.

² <https://github.com/MinCiencia/Datos-COVID19>

- poblacion: dato de tipo entero que corresponde a la cantidad de habitantes de la comuna.
- columnas aaaa-mm-dd: contienen datos de tipo numérico entero y corresponde a la cantidad de casos confirmados incremental (no acumulado) para el día aaaa-mm-dd (formato de fecha año-mes-día) informado por el Ministerio de Salud de Chile. Se encuentran datos desde el 2020-03-30 hasta el 2020-06-23 con una frecuencia de 2 a 3 días.
- tasa: dato de tipo numérico float que contiene el número de casos confirmados por cada 100 mil habitantes de la comuna a la que se hace referencia.

Adicionalmente, se incorporó data asociada a los decretos de cuarentenas, disponibles en el repositorio anteriormente mencionado, en la carpeta “Data Product 29” que contiene datos de las cuarentenas activas e históricas. Se utilizó el archivo “Cuarentenas-Historicas.csv”, que contiene las siguientes columnas:

- id: dato tipo numérico auto-incremental utilizado para identificar cada fila.
- nombre: dato de tipo string que corresponde al nombre de la comuna afecta a cuarentena.
- estado: dato de tipo string que corresponde al estado de la cuarentena, en este caso “Histórica”.
- alcance: dato de tipo string que corresponde a una breve descripción respecto del alcance geográfico de la cuarentena en esa comuna, pudiendo tomar valores como “Área Urbana Completa” o “Sector Específico”.

- fecha_inicio: dato de tipo datetime que corresponde a la fecha de inicio de la cuarentena decretada en esa comuna.
- fecha_fin: dato de tipo datetime que corresponde a la fecha de fin de la cuarentena decretada en esa comuna.
- codigo_CUT_comuna: dato tipo numérico entero que corresponde al identificador de la comuna.
- detalles: corresponde a un dato de tipo string que contiene una breve descripción asociada a la cuarentena en esa comuna.
- superficie: dato de tipo numérico float que corresponde al área afecta a cuarentena en metros cuadrados de la comuna.
- perímetro: dato de tipo numérico float que corresponde al perímetro que delimita la zona de la comuna afecta a cuarentena, en metros.

4.2. Metodología

A continuación se presenta la metodología que guió el presente estudio, la cual consta de ocho etapas, resumidas en la Figura 4. Cabe destacar que para el análisis se utiliza el lenguaje Python y sus diferentes librerías disponibles para análisis de redes, visualización de datos, y medidas estadísticas, tales como Pandas, Numpy, NetworkX, Netgraph, Matplotlib, Seaborn y SciPy, principalmente.



*Figura 4. Metodología propuesta.
Fuente: Elaboración Propia.*

Hipótesis

Como paso inicial de la metodología, se planteó la hipótesis que guió la investigación y la exploración de métricas de centralidad, a fin de dar respuesta a la conjetura planteada. Se establecieron también los objetivos necesarios que permiten probar o refutar la hipótesis. El detalle se encuentra en la sección anterior de “Hipótesis y Objetivos”.

Recopilación de datos

Respecto de la data pública disponible en el GitHub proporcionado por el Ministerio de Ciencia, se puede señalar que los datos ya se encontraban organizados en archivos de

valores separados por coma, lo cual facilitó el acceso a los datos necesarios respecto de contagios incrementales desagregados por comuna. De manera similar, los datos de torres y su localización, proporcionados por Telefónica S.A. a través del Instituto de Data Science de la Universidad del Desarrollo, estaban organizados en el archivo “metro_antennas.csv”. En cuanto a la data de tráfico, también se encontraba organizada en un archivo de valores separados por coma, llamado “tráfico.csv”.

Limpieza y preprocesamiento de datos

En esta etapa se revisaron los distintos conjuntos de datos señalados en la sección Datos. En el caso de la data correspondiente a torres, se buscaron sólo aquellas torres cuyas coordenadas (latitud y longitud) fuesen concordantes con la ubicación de una estación de Metro. Esto se realizó a fin de trabajar sólo con torres que representen las estaciones de la red de Metro de Santiago, que serían los nodos de la red. Con el conjunto de torres relevantes, se filtró el dataset de tráfico de datos entre dispositivos y antenas, con el objeto de rescatar sólo los datos correspondientes a conexiones en estaciones de Metro.

Los conjuntos de datos no contenían datos nulos, por lo que no fue necesario realizar imputación de valores nulos. Sí se verificó que el tipo de dato fuese consistente y que los valores pertenecieran al dominio correspondiente, por ejemplo, que la columna de longitud sólo estuviese poblada por valores numéricos tipo decimal (float) y cuyos valores fuesen desde -70.65 hasta -70.54, dado que son los meridianos entre los que se ubica la provincia de Santiago. De manera análoga, se verificaron los valores de latitud de modo

que los valores de esta columna fuesen numéricos tipo float entre -33.61 y -33.38, que son los paralelos entre los que se ubica la provincia de Santiago.

En cuanto al conteo de viajes, se procesaron los datos de tráfico entre dispositivos móviles y antenas de torres, agrupando los datos por los campos “device” y por la fecha del campo “datetime”. Esto permitió obtener las antenas con las que un dispositivo interactúa en un día determinado, y con ello, se reconstruyó el viaje que efectuó el individuo en posesión del dispositivo correspondiente. Cabe destacar que se trabajó con datos anonimizados, donde en el identificador del dispositivo se encontraba enmascarado.

Por otra parte, se filtró la data según el horario de funcionamiento del Metro, motivo por el cual sólo se consideraron dispositivos que interactuaron con antenas de Metro entre las 05:00hrs. y las 23:00hrs. Esto, ya que el horario operativo del Metro en un día laboral es de 06:00 a 22:00hrs. y se le agregó una hora de holgura al inicio, por si hay usuarios que llegan antes de que inicie el servicio, y una hora de holgura al final de la jornada, puesto que si bien el último viaje inicia a las 22:00hrs, las estaciones cierran una vez que todos los pasajeros llegan a destino y salen de las instalaciones. En fines de semana sólo varía el horario de inicio del servicio, desde las 07:00hrs. y 08:00hrs. los sábados y domingos, respectivamente. En cualquier caso, el ruido que podría generarse a las horas de inicio y fin de funcionamiento del servicio, fue eliminado al aplicar el filtro de que un dispositivo debe haber tenido conexión con al menos dos antenas distintas en un mismo día, de lo contrario, no se contabiliza pues no hubo desplazamiento en Metro.

Respecto de la metodología para contar viajes a partir de los dispositivos, ésta consistió en lo siguiente:

- 1) Agrupar los registros por dispositivo-día ordenados por hora, donde si el dispositivo-día presentó conexión a una única antena se borra el registro, debido a que esto indica que no hubo desplazamiento entre distintas estaciones de Metro.
- 2) Teniendo los registros ordenados por hora para cada dispositivo-día, se obtuvieron las estaciones de Metro donde el dispositivo tuvo conexión con la antena respectiva, con lo cual a cada registro le corresponde un listado de antenas ordenadas por hora de conexión en un día.
- 3) Se reconstruyó el viaje considerando la ruta más corta entre las estaciones ordenadas por hora donde el dispositivo tuvo conexión en el día. Por ejemplo, si el dispositivo x tuvo conexión en el día y con las antenas ordenadas $\{a, d, f\}$ que cubren las estaciones de Metro $\{A, D, F\}$, entonces se reconstruyó el viaje considerando las estaciones $\{B, C\}$ y $\{E\}$ que conforman la ruta más corta desde A a D y desde D a F , respectivamente; obteniéndose la ruta $A-B-C-D-E-F$.

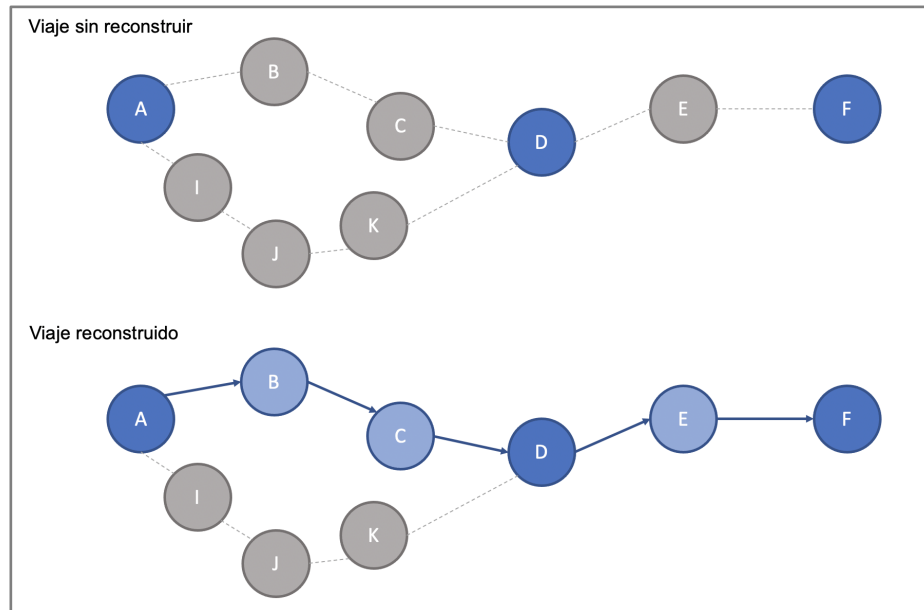


Figura 5. Ejemplo de reconstrucción de viaje.

Fuente: Elaboración Propia.

Cabe mencionar que si una estación se encuentra ubicada en el límite entre dos o más comunas, se consideró que atiende a cada una de ellas en partes iguales (Anexo 2).

Análisis exploratorio

Luego de la limpieza y preprocesamiento de los datos, se continuó con el análisis exploratorio. El objetivo de esta etapa fue identificar comportamientos de las variables, a través de tablas y visualizaciones como gráficos de barras, histogramas, gráficos de torta, entre otros. Este análisis permitió validar que el comportamiento de los datos hicieran sentido con los patrones de movilidad en el Metro, mostrando distribuciones de la cantidad de dispositivos en las estaciones de Metro a lo largo de las horas del día, así como también a lo largo de los días y semanas.

Con el número de conexiones, se procedió a calcular la representatividad de la muestra de estudio. Para ello se utilizó la fórmula para el cálculo de tamaño de muestra conocido el tamaño de la población[25]:

$$n = \frac{N \times Z_{\alpha}^2 \times p \times (1 - p)}{d^2 \times (N - 1) + Z_{\alpha}^2 \times p \times (1 - p)}$$

Donde n corresponde al tamaño de la muestra, N al tamaño de la población, Z_{α} representa el nivel de significancia, p es la proporción esperada o probabilidad de éxito y d es el error máximo admisible en términos de proporción. En este caso se utilizaron los parámetros: $p = 0.5$, ya que según recomienda bibliografía[25] en caso de desconocer la proporción se aplica este criterio conservador a fin de maximizar el tamaño de muestra n . Por convención, se exige un nivel de confianza del 95% (por lo que $\alpha = 0.05$), y desde la tabla de distribución Normal se obtiene $Z_{\alpha} = 1.96$; mientras que el error máximo admisible en este caso se establece en $d = 1\%$, por criterio conservador, a fin de maximizar el tamaño de muestra. Para el tamaño de la población N , se consideró la afluencia de un día laboral promedio, y de ella sólo la fracción correspondiente a las líneas de Metro estudiadas. Como se señaló en el apartado introductorio, Metro indica que en un día laboral promedio se realizan 2,409 millones de viajes. No se señalan cantidad de pasajeros, pero corresponden como máximo a 2,409 millones de personas, en el caso de que cada viaje se haya realizado por una persona diferente. Se estableció entonces la cota superior de 2,4 millones de usuarios del servicio en un día laboral, ponderado por 81%, que corresponde a la suma de las participaciones de las líneas 1, 2, 4 y 5 en el total de viajes, con lo que la población de estudio asciende a 1,94 millones de usuarios. De esta forma se calculó el

tamaño muestral requerido, el cual asciende a un mínimo de 10.000 dispositivos diarios, cifra que debe verificarse en los resultados.

En esta fase también se construyó un gráfico de línea para mostrar la evolución de los casos nuevos de COVID-19 confirmados por cada comuna a lo largo del tiempo, y la tabla de casos agrupados por comuna para su posterior análisis.

Con los resultados de esta etapa se pueden responder preguntas como ¿cuáles son los horarios en que hay más individuos en las estaciones de Metro? ¿existen comportamientos distintos en el periodo previo a la cuarentena y el periodo afecto a ella?. Las observaciones anteriores son de utilidad para el proceso que continúa, que es la generación de grafos.

Generación de Grafos ponderados

Luego del análisis exploratorio de los datos, se continuó con la generación de grafos ponderados, donde los nodos del grafo representan las estaciones de Metro de Santiago; los arcos representan la conexión entre estaciones dado por si existe o no ese servicio, y finalmente, el peso de cada arco está dado por el número de viajes asociados al traslado entre estaciones. Se generó un grafo por semana para un total de 18 semanas.

Semana	Periodo	Semana	Periodo
1	De 02/03/2020 a 08/03/2020	10	De 04/05/2020 a 10/05/2020
2	De 09/03/2020 a 15/03/2020	11	De 11/05/2020 a 17/05/2020
3	De 16/03/2020 a 22/03/2020	12	De 18/05/2020 a 24/05/2020
4	De 23/03/2020 a 29/03/2020	13	De 25/05/2020 a 31/05/2020
5	De 30/03/2020 a 05/04/2020	14	De 01/06/2020 a 07/06/2020
6	De 06/04/2020 a 12/04/2020	15	De 08/06/2020 a 14/06/2020
7	De 13/04/2020 a 19/04/2020	16	De 15/06/2020 a 21/06/2020
8	De 20/04/2020 a 26/04/2020	17	De 22/06/2020 a 28/06/2020
9	De 27/04/2020 a 03/05/2020	18	De 29/06/2020 a 05/07/2020

Tabla 1. Periodo de estudio agrupado en semanas.

Fuente: Elaboración propia.

Para probar la importancia de considerar la cantidad de viajes como ponderador de la red, se realizó una comparación entre el grafo de la red de Metro sin ponderación y con ponderación por cantidad de viajes. Se calculó el grado de centralidad de intermediación de los nodos (*betweenness centrality*) de ambas redes, a fin de comprobar que la capacidad conectiva de las estaciones de la red de Metro no está dada sólo por la estructura de la red, sino que por el patrón de movilidad debido al uso que las personas hacen de este servicio de transporte.

Exploración de métricas de centralidad

En esta etapa se calculan las métricas de centralidad de intermediación para cada grafo semanal, ponderado por la cantidad de viajes. La centralidad de intermediación o *betweenness centrality* (c_B) se puede calcular como:

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)}$$

Donde V corresponde al conjunto de nodos de la red, $\sigma(s, t)$ es la cantidad de caminos más cortos entre los nodos s y t , mientras que $\sigma(s, t|v)$ es el subconjunto de los caminos más cortos entre s y t , que pasan por v [26][27]. No obstante, para poder comparar entre distintas redes, se debe realizar una normalización aplicando el factor $2/((n - 1)(n - 2))$ para grafos no direccionados, y por $1/((n - 1)(n - 2))$ para grafos direccionados, donde n corresponde al número de nodos de la red. Cabe mencionar que un nodo con mayor centralidad de intermediación posee una mejor capacidad conectiva que otro nodo con menor centralidad de intermediación, puesto que acumula una mayor cantidad de rutas

más cortas para llegar de un nodo a otro. Análogamente, este estudio considera el supuesto de que si un usuario de Metro necesita llegar a B desde A, escogerá la ruta que le signifique menor cantidad de transbordos y menor cantidad de estaciones entre A y B.

Además, para explicar que la centralidad de intermediación del grafo ponderado calculada previamente no se debe necesariamente a la distancia entre nodos, se calculó la centralidad de proximidad (*closeness centrality*), que representa el recíproco de la distancia de la ruta más corta promedio al nodo v desde los demás $n-1$ nodos alcanzables:

$$c_c(v) = \frac{n-1}{\sum_{u=1}^{n-1} d(u, v)}$$

Donde $d(u, v)$ representa la distancia de la ruta más corta desde el nodo u hasta el nodo v , siendo n la cantidad de nodos que pueden llegar hasta v .

De esta manera, se propone la métrica de centralidad de intermediación como indicador que permite identificar los nodos más relevantes en términos de conexión de la red, para posteriormente evaluar su correlación con distintos desfases con la cantidad de nuevos contagios por comuna para el periodo estudiado.

Cálculo de correlaciones

Una vez que se calculó la centralidad de intermediación para las estaciones de Metro en cada semana, se procedió a calcular la correlación lineal de Pearson[28] entre la centralidad de intermediación y los casos nuevos con desfase de 1, 2, 3 y 4 semanas entre la semana en que ocurrieron los viajes y la semana en que se contabilizan los casos nuevos. Esto debido a que las investigaciones señalan que el virus posee un periodo de incubación

que tiene como mediana 7.76 días, con 14 días correspondientes al percentil 90, pero existe una probabilidad del 5 al 10% de que el virus posea un periodo de incubación superior a 14 días [29]. Para este proceso se utilizó la librería Scipy y se exigió un valor-p de a lo más 5% para considerar el coeficiente obtenido como significativo.

Paralelamente, con la misma librería se realizaron pruebas de correlación de Spearman[30] sobre la data, a fin de determinar si existía una relación monótona en las fluctuaciones de las variables centralidad de intermediación y cantidad de nuevos contagios. Al igual que en el cálculo del coeficiente de Pearson, se exigió un p-valor igual o menor al 5% para considerar el coeficiente de Spearman como significativo. Este análisis se realiza de manera agrupada y desagrupada, es decir, se calculó un coeficiente general y un coeficiente para cada comuna.

Contraste de hipótesis

En relación al primer objetivo, al tener la métrica de centralidad de intermediación para cada estación y semana, se verificó si la centralidad era o no mayor en las estaciones que sirven a comunas localizadas en el centro de la provincia (Estación Central, Santiago y Providencia), y luego se comparó si efectivamente la centralidad de intermediación en el periodo sin cuarentena es superior a la del periodo con cuarentena, a través de una prueba de hipótesis de comparación de medias para muestras dependientes[31], donde se tiene el antes y el durante cuarentena para cada comuna. Cabe señalar que en esta prueba la hipótesis nula correspondía a que ambas muestras tenían la misma media, por lo que se debía rechazar la hipótesis nula para probar la hipótesis planteada.

Respecto de la segunda hipótesis, para comprobar que las estaciones con mayor centralidad de intermediación conllevaban una mayor cantidad de contagios futuros, se debía encontrar una correlación positiva y significativa entre el *betweenness centrality* semanal de las estaciones agrupadas por comunas y la cantidad de contagios nuevos considerando los desfases descritos en la sección anterior.

5. Resultados

En la presente sección se presentan los hallazgos relevantes de cada etapa.

5.1. Limpieza y preprocesamiento de datos

Como se menciona en el apartado anterior, se aplicaron filtros a las antenas por latitud y longitud, y como resultado se obtuvo un total de 90 antenas que proporcionan cobertura a 79 estaciones de la red de Metro de Santiago. Sin embargo, al analizar las conexiones entre las antenas y los dispositivos, se encontraron datos para 84 antenas, correspondientes a 77 estaciones de Metro (Anexo 1), lo cual puede deberse a que tal como se explicó en el apartado introductorio, existen estaciones fuera de servicio debido a los daños ocurridos en octubre de 2019[7]. Las estaciones estudiadas prestan servicio a 18 comunas de la Provincia de Santiago, que corresponden a: Estación Central, Huechuraba, La Cisterna, La Florida, La Reina, Las Condes, Lo Prado, Macul, Maipú, Ñuñoa, Peñalolén, Providencia, Pudahuel, Puente Alto, Quinta Normal, Recoleta, San Miguel y Santiago (Anexo 2). Por tanto, estas fueron las comunas estudiadas en cuanto al conteo de viajes y análisis de correlaciones entre centralidad de intermediación, obtenida con los datos de movilidad, y nuevos contagiados COVID-19 semanales por comuna.

A continuación, se presenta una visualización producto del proceso de limpieza y procesamiento, donde se pueden observar las antenas, las comunas y sus localizaciones dentro de la provincia de Santiago.

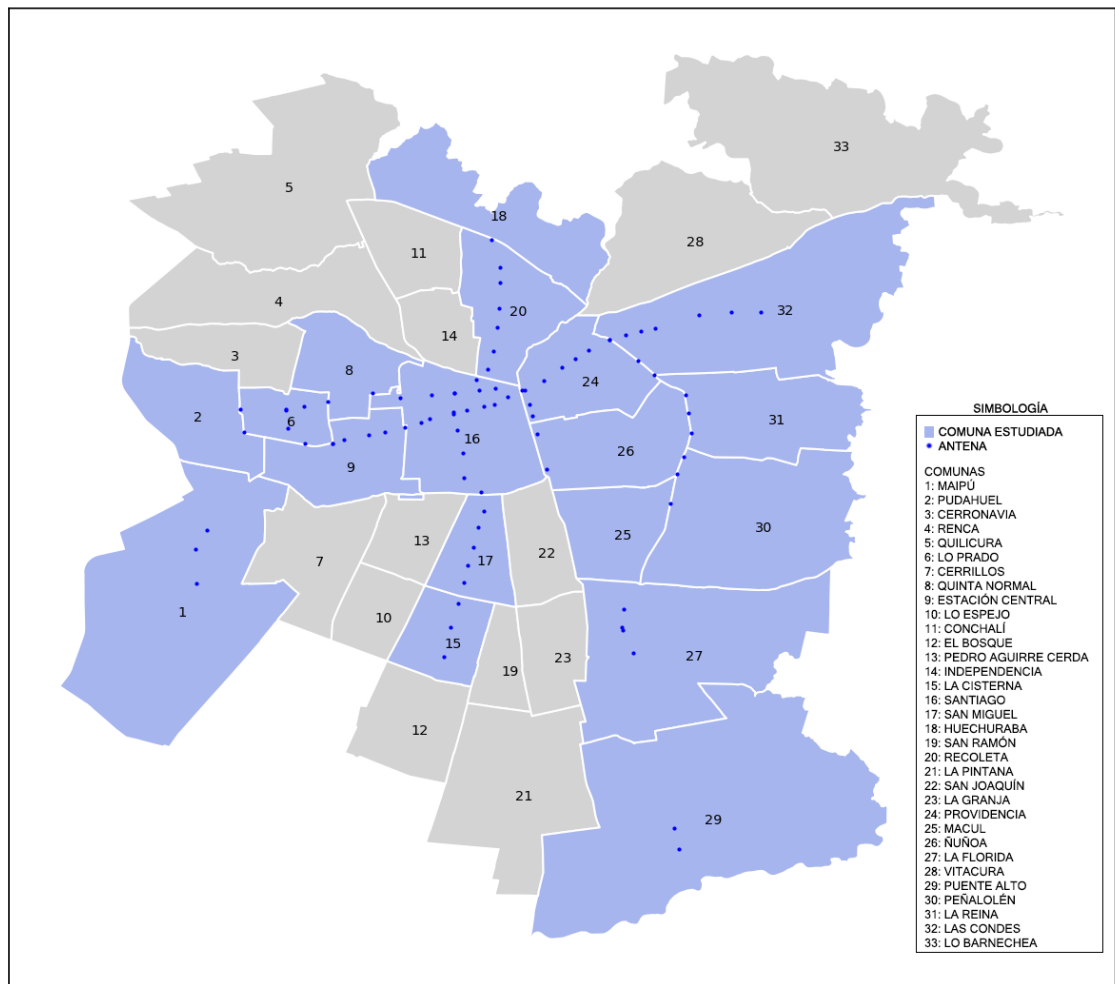


Figura 6. Ubicación de antenas y comunas estudiadas.

Fuente: Elaboración Propia.

5.2. Análisis exploratorio

Para comenzar con este análisis, se observó el comportamiento diario de las conexiones de dispositivos con las antenas que proporcionan cobertura en las estaciones de la Red de Metro de Santiago.

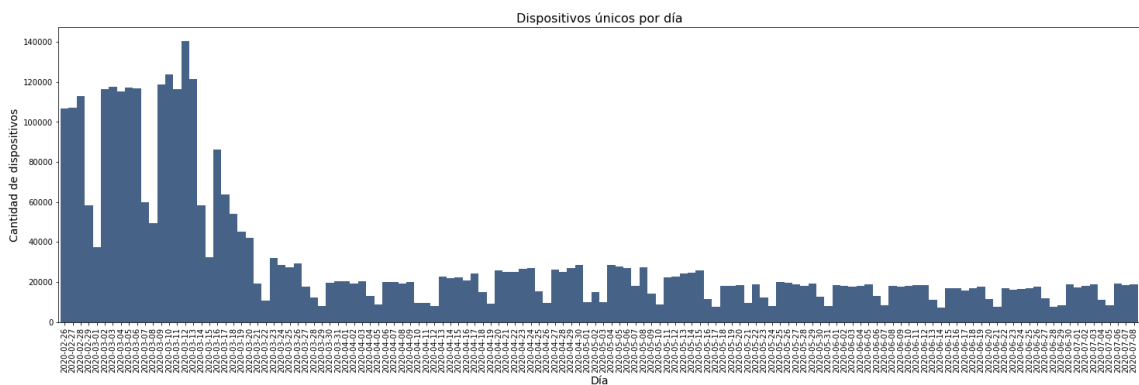


Figura 7. Dispositivos únicos conectados por día.

Fuente: Elaboración Propia.

En la Figura 7 se observa en primer lugar, una disminución de las conexiones diarias donde en el periodo previo al 16 de marzo de 2020, el orden de magnitud de las conexiones era de 100.000, mientras que después de esa fecha hasta el 17 de mayo se muestra una clara disminución hasta las 25.000 conexiones diarias, aproximadamente; y posterior al 17 de mayo las conexiones se mantienen en cerca de 20.000 dispositivos diarios. En otras palabras, se produjo una disminución del orden del 80% del uso del servicio de transporte proporcionado por Metro en la muestra de estudio. Esto es concordante con la realidad nacional, donde a partir del 16 de marzo de 2020 se suspendieron las clases presenciales en la Región Metropolitana[32], mientras que se decretó cuarentena total para la totalidad

de la Provincia de Santiago y las comunas de Puente Alto, Buin, San Bernardo, Padre Hurtado, Lampa y Colina a partir del 15 de mayo de 2020 [33].

Por otra parte, la Figura 7 muestra una señal semanal muy clara, respecto de la diferencia en el uso de servicio de Metro entre los días lunes a viernes y sábado-domingo. Cabe destacar que este patrón cíclico también se ha presentado en otros estudios de movilidad[23]. Tal como señala Metro en su Memoria Anual[5], el mayor uso del servicio se produce entre los días lunes y viernes, ambos inclusive, debido a que los usuarios se trasladan desde sus domicilios a sus lugares de trabajo y viceversa.

Respecto de la representatividad de los datos, se concluye que la muestra es representativa, ya que se dispone del orden de 100.000 dispositivos diarios para un día laboral promedio sin cuarentena, superando el mínimo requerido de 10.000 dispositivos, calculado previamente bajo criterio conservador explicado en el apartado de Metodología.

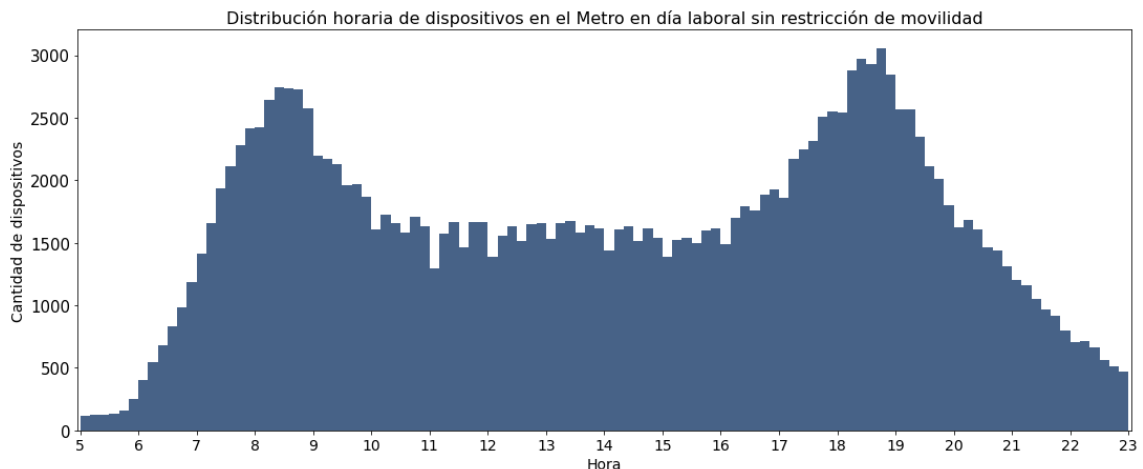
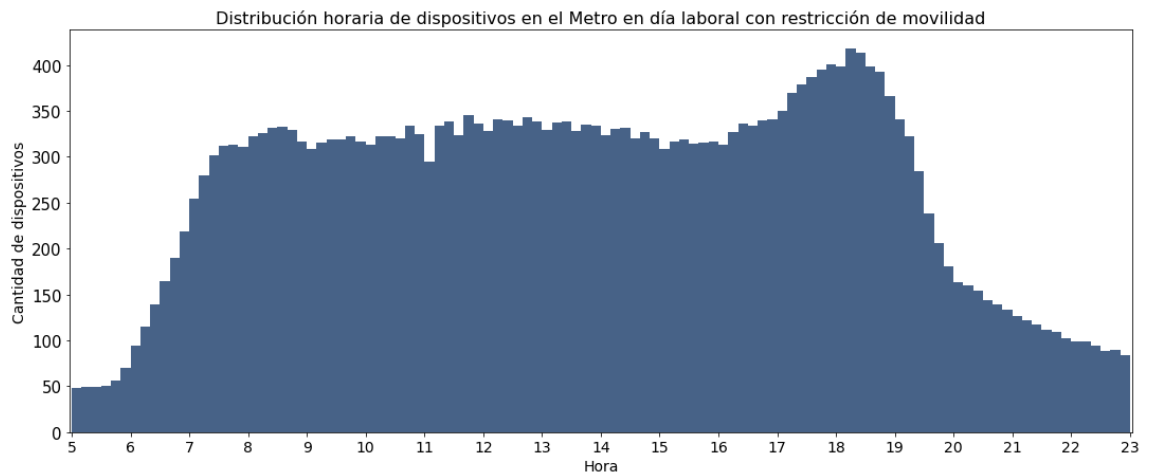


Figura 8. Dispositivos únicos conectados por hora en día laboral promedio sin cuarentena.
Fuente: Elaboración Propia.

Para continuar con el análisis exploratorio, se analizó el comportamiento de dispositivos, y por ende, de los individuos al interior de las estaciones de Metro, en un día laboral promedio sin restricciones de movilidad (días de lunes a viernes entre el 26/02/2020 y el 15/03/2020). En la Figura 8, se observa que a lo largo del día existen dos cimas pronunciadas, la primera alrededor de las 8:45hrs. y la segunda a las 18:45hrs. aproximadamente. Este comportamiento podría asociarse al inicio y término de la jornada laboral, ya que según indica el Informe de Encuestas Origen Destino de SECTRA en 2012[34], los viajes en días de semana, en esos horarios y utilizando transporte público se asocian a traslados entre el hogar y el lugar de trabajo. Además, se aprecia un leve aumento de individuos en el Metro alrededor de las 13:30hrs., que usualmente corresponde al horario de colación.

En contraste, en la Figura 9 se presenta el comportamiento en un día laboral promedio, pero para el periodo con restricciones de movilidad estudiado, que se extiende desde el 16 de marzo de 2020 hasta el 05 de julio del mismo año.



*Figura 9. Dispositivos únicos conectados por hora en día laboral promedio con cuarentena.
Fuente: Elaboración Propia.*

El primer aspecto a destacar corresponde a la disminución de la frecuencia: anteriormente las cimas contaban con alrededor de 2.550 dispositivos, en cambio, en esta figura se muestra que la cima más pronunciada cuenta con 400 dispositivos. En este sentido, se refleja la disminución significativa de usuarios de Metro que se había mencionado anteriormente en la Figura 7. Por otra parte, ahora sólo se presenta una única cima alrededor de las 18:15hrs., mostrando un mayor flujo de individuos en el Metro a esa hora. El flujo de individuos que sin restricciones de movilidad utilizaba el servicio en a las 8:45hrs. parece haber diversificado su horario de inicio de jornada laboral, distribuyéndose hacia las horas posteriores, pues el valle que se presentaba anteriormente entre las 10:00hrs. y las 17:00hrs. ya no existe, mostrando ahora una curva constante de 330 dispositivos entre esas horas.

Pero ¿se comportan de la misma manera todas las comunas estudiadas? A fin de responder la interrogante planteada se propuso visualizar el comportamiento promedio

por comuna en un día laboral, para lo cual se contabilizan los dispositivos en las estaciones de Metro que se ubican en dichas comunas.

A modo de ejemplo, se muestra el comportamiento en las comunas de Las Condes, Santiago y San Miguel. Como se puede observar en la Figura 10, el comportamiento entre las comunas es distinto (más comunas en Anexo 4 y Anexo 5).

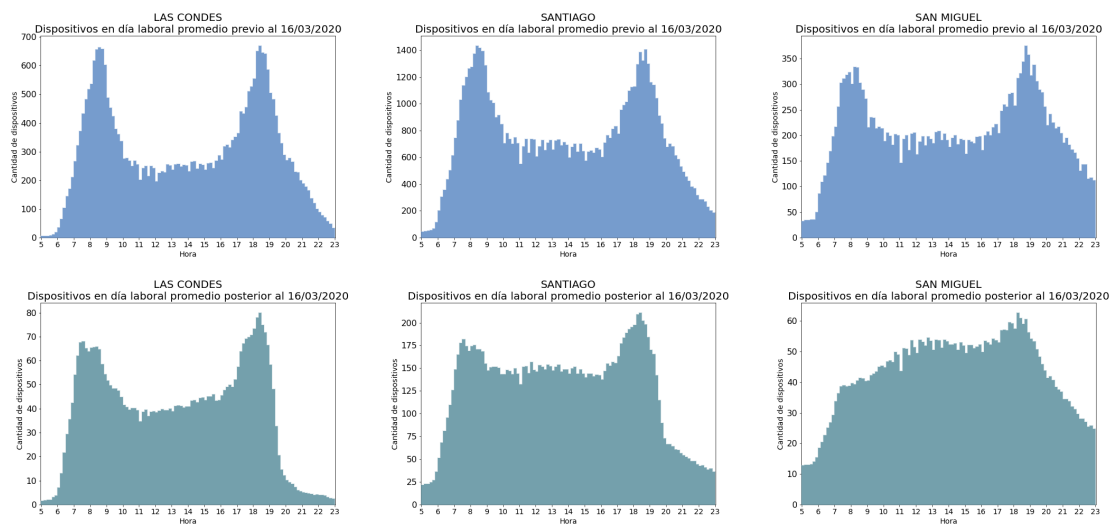


Figura 10. Comparación de distribución horaria de dispositivos en el Metro antes y después del 16 de marzo de 2020 en las comunas de Las Condes, Santiago y San Miguel.

Fuente: Elaboración Propia.

En primer lugar, al comparar la cantidad de dispositivos en las estaciones de Metro de cada comuna en un día laboral promedio antes del 16 de marzo, se puede apreciar que la comuna de Las Condes presenta un máximo de 650 dispositivos en un mismo momento del día, en cambio Santiago alcanza un máximo de 1400 dispositivos, mientras que San Miguel tiene un máximo de 350 dispositivos en el Metro en un determinado momento del

día. En otras palabras, el máximo de individuos que se encuentran simultáneamente en estaciones de San Miguel en un momento del día equivale sólo al 25% del máximo de individuos que se encuentran en estaciones de la comuna de Santiago en determinado momento del día. Por otra parte, en el periodo previo al 16 de marzo, se aprecia una distribución similar de los individuos en las estaciones de Metro a lo largo del día, asemejándose al perfil de distribución general presentado anteriormente (de toda las estaciones), con las dos cimas de las 8:45hrs. y las 18:45hrs.

Respecto del comportamiento desde el 16 de marzo en adelante, la distribución horaria del tráfico es distinta para cada comuna. En el caso de Las Condes, se aprecia una disminución del 88,5% de los dispositivos en las estaciones, en comparación al periodo previo al 16 de marzo. De manera similar, la disminución del número de dispositivos para las comunas de Santiago y San Miguel es de 86% y 83%, respectivamente. Sin embargo, la distribución de dispositivos a lo largo del día durante el periodo con restricciones de movilidad es distinto en las tres comunas presentadas. Mientras en Las Condes la cima que antes se producía a las 8:45hrs. ahora se aplana, distribuyéndose entre las 7:30hrs y las 8:45hrs., en San Miguel esta cima de la mañana desaparece por completo, ampliándose hacia la derecha, mostrando que en la medida que transcurre la mañana los viajes van en aumento hasta el mediodía. Santiago, en cambio, muestra un comportamiento en donde se aplana la cima de la mañana, extendiéndose desde 7:30 a 9:00hrs. pero a diferencia de Las Condes, esta cima es menos pronunciada respecto de la media diaria, que oscila en los 150 dispositivos. La distribución del *peak* de las 8:45hrs. podría explicarse debido a que con el surgimiento de la pandemia en la Región Metropolitana, la Dirección del Trabajo

implementó dictámenes que permitieron a los empleadores y empleados pactar horarios diferidos de inicio y término de la jornada laboral[35]. Sin embargo, el estudio en profundidad de esta medida escapa al alcance de la presente investigación.

Por otra parte, se visualizó la evolución de los casos confirmados de COVID-19 acumulados por comuna, para cada una de las 18 comunas estudiadas:

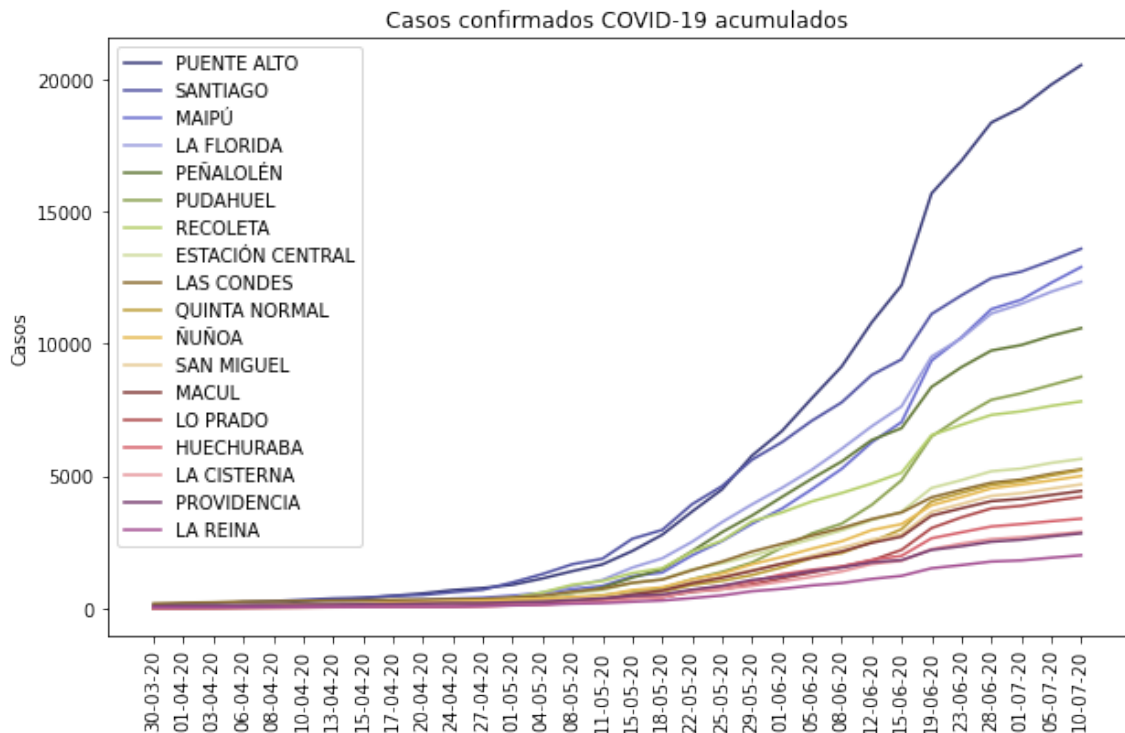


Figura 11. Casos confirmados acumulados de COVID-19 por comuna.

Fuente: Elaboración Propia. Datos: Ministerio de Ciencia, GitHub Datos COVID-19/output/producto1

Se calculó el incremento entre los distintos días y se agrupó por semana, a fin de obtener la cantidad de casos nuevos por comuna para cada semana en el periodo de estudio, resultado que se muestra en la Tabla 2.

Comuna	02/03-08/03	09/03-15/03	16/03-22/03	23/03-29/03	30/03-05/04	06/04-12/04	13/04-19/04	20/04-26/04	27/04-03/05	04/05-10/05	11/05-17/05	18/05-24/05	25/05-31/05	01/06-07/06	08/06-14/06	15/06-21/06	22/06-28/06	29/06-05/07
ESTACIÓN CENTRAL	0	0	0	0	9	36	40	61	179	237	376	512	520	631	716	1214	633	315
HUECHURABA	0	0	0	0	5	16	21	18	36	107	155	294	336	447	375	818	448	198
LA CISTERNA	0	0	0	0	2	22	13	26	49	83	143	298	199	340	486	565	379	170
LA FLORIDA	0	0	0	0	24	59	85	71	146	396	714	976	1413	1310	1642	2636	1628	823
LA REINA	0	0	0	0	6	22	7	6	43	65	74	137	260	220	246	403	253	146
LAS CONDES	0	0	0	0	43	57	39	26	63	217	335	478	702	606	623	822	555	336
LO PRADO	0	0	0	0	6	15	27	26	82	88	132	227	283	447	463	1222	742	281
MACUL	0	0	0	0	13	20	14	32	47	136	252	425	463	497	564	1028	549	234
MAIPÚ	0	0	0	0	21	103	117	80	116	238	494	788	1193	1306	1773	3089	1951	994
ÑUÑO A	0	0	0	0	26	62	62	52	60	106	257	399	571	588	705	939	638	300
PEÑALOLÉN	0	0	0	0	23	44	64	47	113	248	587	1003	1345	1394	1461	2008	1377	555
PROVIDENCIA	0	0	0	0	12	35	35	32	30	83	163	234	365	332	334	470	308	204
PUDAHUEL	0	0	0	0	15	35	36	34	96	139	250	473	633	1103	1075	2589	1385	563
PUENTE ALTO	0	0	0	0	31	163	233	216	201	515	761	1493	2112	2169	2865	4878	2684	1431
QUINTA NORMAL	0	0	0	0	14	34	56	47	90	115	183	319	376	642	614	1530	619	356
RECOLETA	0	0	0	0	15	29	46	67	224	491	441	805	1152	753	676	1827	768	341
SAN MIGUEL	0	0	0	0	14	30	42	41	61	137	213	452	429	544	609	1047	605	274
SANTIAGO	0	0	0	0	46	148	142	145	372	686	961	1308	1682	1472	1729	2305	1358	665

Tabla 2. Cantidad de casos nuevos de contagiados con COVID-19 cada semana.
Fuente: Elaboración Propia. Datos: Ministerio de Ciencia, GitHub Datos COVID-19/oupu/producto1

Se destaca que los primeros registros de casos nuevos COVID-19 disponibles en el repositorio corresponden al 30 de marzo del 2020, por lo cual, no existe incremental de casos en las primeras 4 semanas del presente estudio. De esta manera, para los análisis que necesitaban de una 1 semana de desfase entre la variable y los números de casos, se consideraron las semanas desde el 23 de marzo (semana 4) hasta la penúltima semana (del 22 de junio al 28 de junio de 2020). Para un desfase de 2 semanas la ventana de datos de la variable desplazó una semana más hacia la izquierda respecto de la cantidad de contagios, hasta la penúltima semana disponible de datos, y así sucesivamente.

5.3. Grafos ponderados

En primer lugar, se construyó el grafo de la red con las estaciones de Metro, ubicadas a escala en base a sus coordenadas geográficas de latitud y longitud, sin considerar ponderaciones ni dirección de aristas, tal como se muestra en la Figura 13.

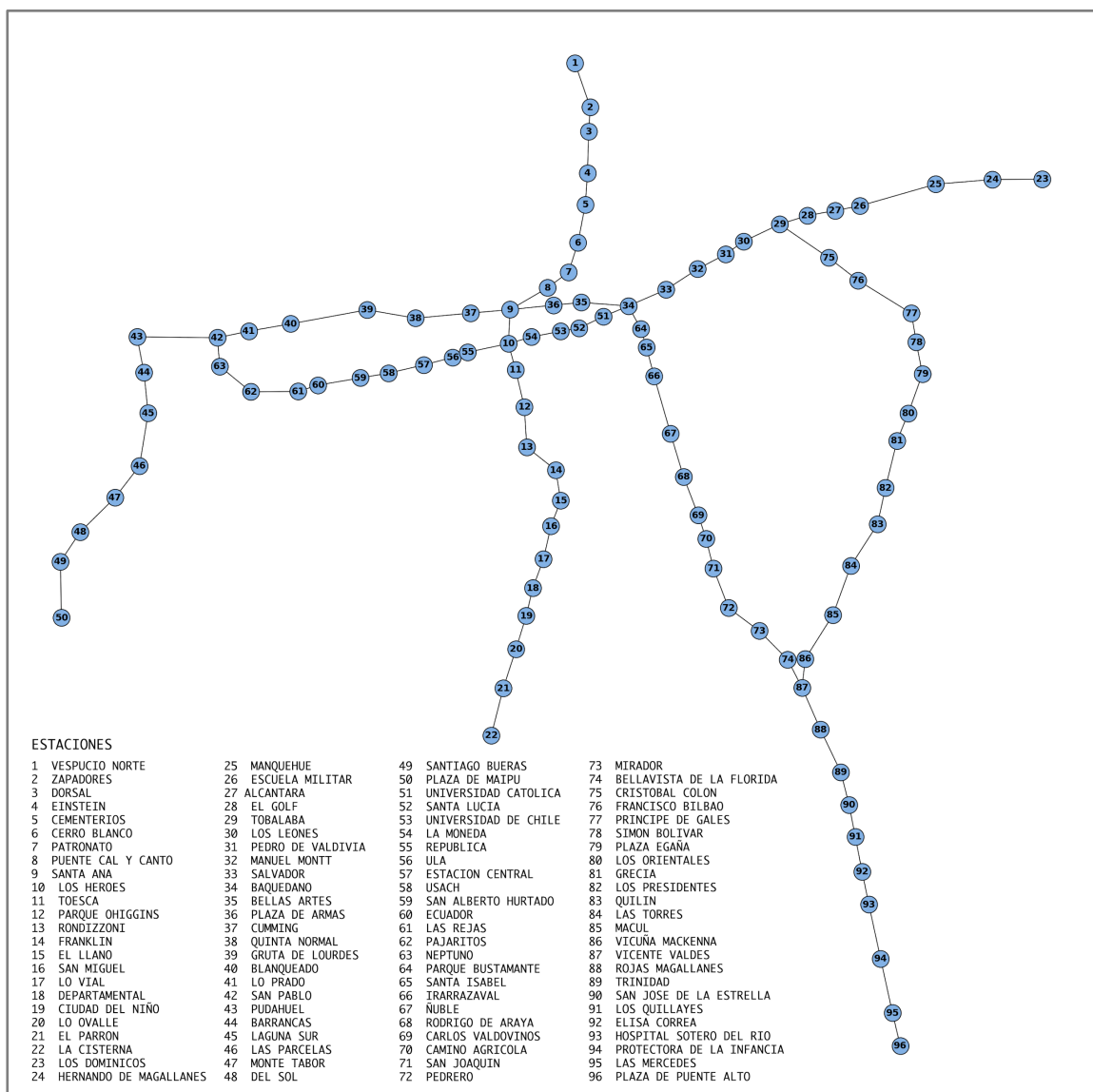


Figura 12. Grafo Red de Metro no ponderado.

Fuente: Elaboración Propia.

Paralelamente, se construyó el grafo ponderado por cantidad de viajes en el periodo comprendido desde el 02 de marzo de 2020 hasta el 05 de julio del mismo año, considerando la dirección en que se producen estos entre las estaciones de Metro, como se muestra en la Figura 14.

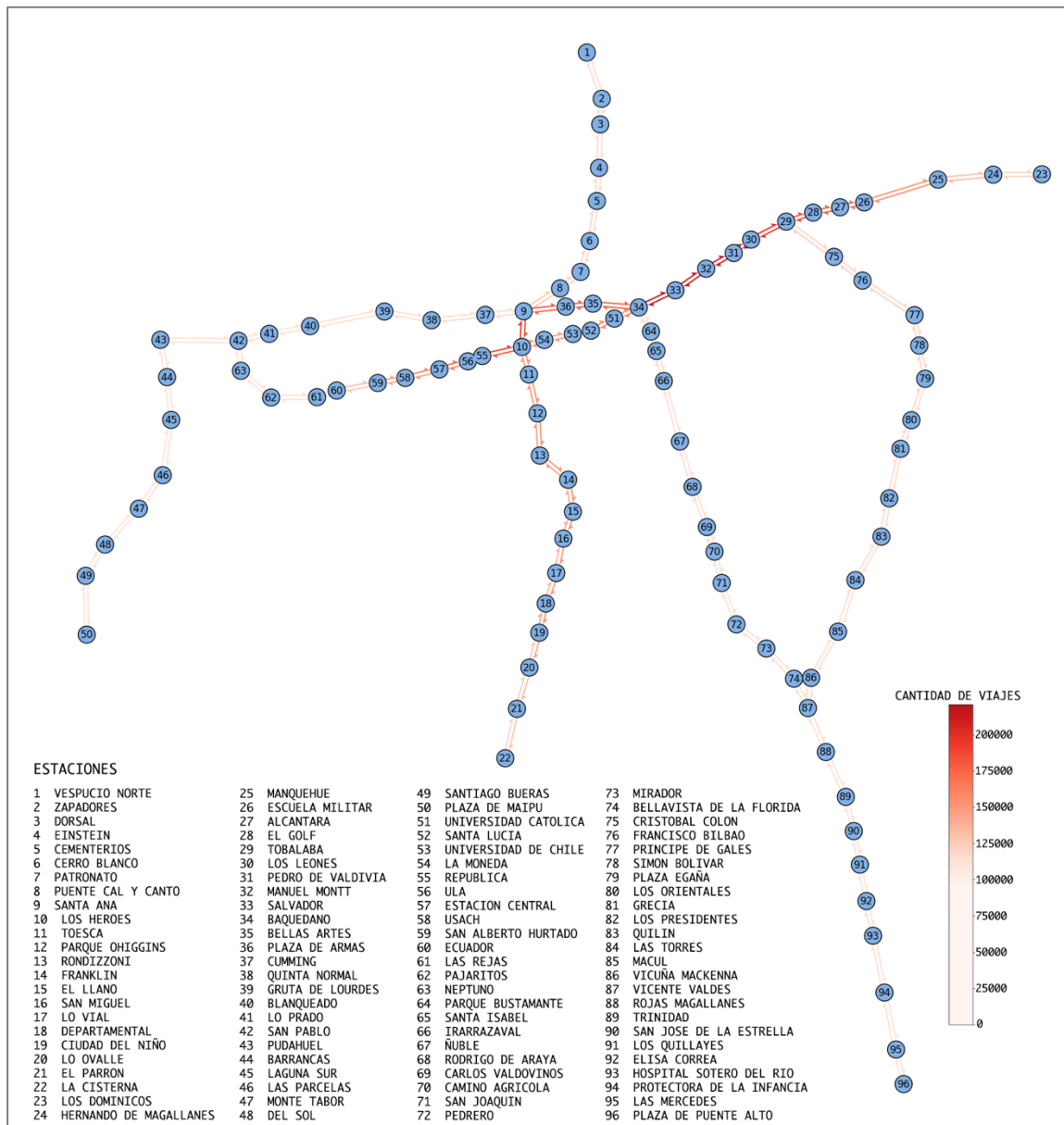


Figura 13. Grafo Red de Metro ponderado por cantidad de viajes entre estaciones, desde 02/03/2020 hasta 05/07/2020. Fuente: Elaboración Propia.

Luego, se calculó el grado de centralidad de intermediación normalizada (*betweenness centrality*) con la librería NetworkX considerando el recíproco de viajes, debido a que la librería interpreta los pesos de la red como resistencias y se requiere que los interprete como flujo. Se obtuvo que al considerar los viajes como ponderadores de la red, la centralidad de intermediación cambia. Esto expone que la capacidad conectiva de las estaciones de Metro no se debe sólo a su posición en la red, sino que también a los patrones de movilidad, dados por el uso que las personas hacen de este servicio de transporte, tal como se muestra en la Tabla 3.

RED NO PONDERADA			RED PONDERADA POR CANTIDAD DE VIAJES		
Rank	Estación	Betweenness centrality	Rank	Estación	Betweenness centrality
1	(9) SANTA ANA	0,5992	1	(9) SANTA ANA	0,6292
2	(34) BAQUEDANO	0,5489	2	(34) BAQUEDANO	0,5624
3	(35) BELLAS ARTES	0,4695	3	(35) BELLAS ARTES	0,4868
4	(36) PLAZA DE ARMAS	0,4693	4	(36) PLAZA DE ARMAS	0,4866
5	(10) LOS HÉROES	0,3412	5	(29) TOBALABA	0,4295
6	(64) PARQUE BUSTAMANTE	0,2974	6	(33) SALVADOR	0,4281
7	(65) SANTA ISABEL	0,2867	7	(32) MANUEL MONTT	0,4214
8	(66) IRARRÁZABAL	0,2759	8	(31) PEDRO DE VALDIVIA	0,4142
9	(67) ÑUBLE	0,2652	9	(30) LOS LEONES	0,4066
10	(33) SALVADOR	0,2618	10	(10) LOS HÉROES	0,3931

Tabla 3. Top 10 nodos con mayor centralidad de intermediación del grafo no ponderado y ponderado por cantidad de viajes

Fuente: Elaboración Propia.

Se observa que la centralidad de intermediación en la red ponderada presenta valores más altos que en la red no ponderada para el top 10 de estaciones con *mayor betweenness centrality* debido a la incorporación de la ponderación de los viajes. Además, las estaciones Tobalaba, Manuel Montt, Pedro de Valdivia y Los Leones entran en el ranking top 10 debido exclusivamente al uso que le dan las personas al Metro, ya que de no ser por la ponderación de cantidad de viajes, estas estaciones quedarían fuera del ranking. Se

destaca que la mayoría de las estaciones con mayor centralidad de intermediación prestan servicio a las comunas de Santiago y Providencia.

Entonces, las estaciones y comunas anteriormente mencionadas son las que tienen una mayor capacidad para conectar la provincia de Santiago a lo largo del periodo estudiado. Cabe mencionar que la capacidad conectiva no se debe necesariamente a la cercanía de un nodo respecto de los demás (*closeness centrality*), lo cual se comprueba al calcular la centralidad de proximidad para los nodos de la red, donde las estaciones Baquedano, Bellas Artes y Plaza de Armas son las que se encuentran más cercanas a todas las demás, como se muestra en Tabla 4.

Rank	Estación	Closeness centrality
1	(34) BAQUEDANO	0,1106
2	(35) BELLAS ARTES	0,1105
3	(36) PLAZA DE ARMAS	0,1103
4	(9) SANTA ANA	0,1102
5	(64) PARQUE BUSTAMANTE	0,1047
6	(10) LOS HÉROES	0,1044
7	(33) SALVADOR	0,1040
8	(37) CUMMING	0,1027
9	(51) UNIVERSIDAD CATÓLICA	0,1025
10	(8) PUENTE CAL Y CANTO	0,1008

Tabla 4. Top 10 nodos con mayor centralidad de proximidad de la red
Fuente: Elaboración Propia.

Ya probada la importancia de utilizar una red ponderada por cantidad de viajes y la métrica de *betweenness centrality*, se analizó la red de Metro ponderada en una semana sin cuarentena y una semana con cuarentena. Tal como se adelantó en el análisis exploratorio, la Figura 15 muestra una importante disminución de la cantidad de viajes en Metro en la provincia de Santiago, que puede observarse mediante el grado de los nodos

(viajes de entrada y salida de cada estación). Se presenta la red en ambas semanas, así como la probabilidad acumulada p_k de que al escoger un nodo al azar, éste tenga un determinado grado k , dado por la cantidad de viajes que ingresan o salen desde ese nodo. Es así como en la semana del 9 al 15 de marzo, la probabilidad acumulada $p_k = 1$ se alcanza en el grado $k = 250.430$ (representa viajes de entrada y salida del nodo Los Héroes), mientras que en la semana del 23 al 29 de marzo, $p_k = 1$ se alcanza tan sólo a los 34.919 viajes en la misma estación.

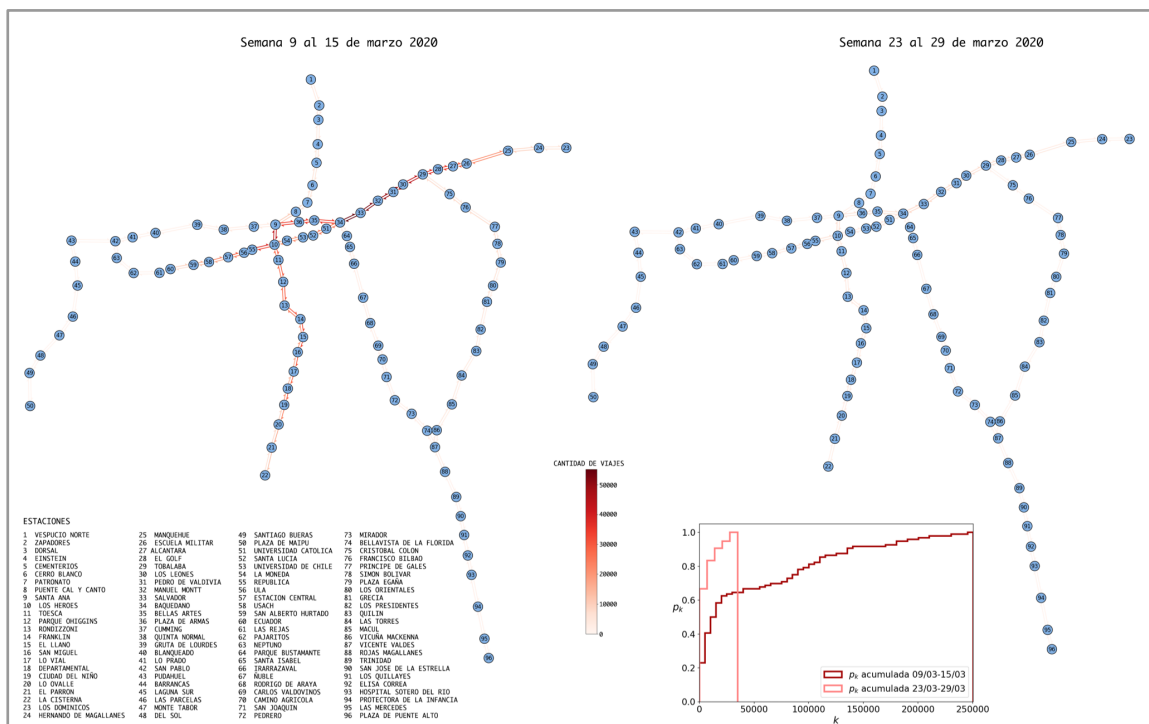
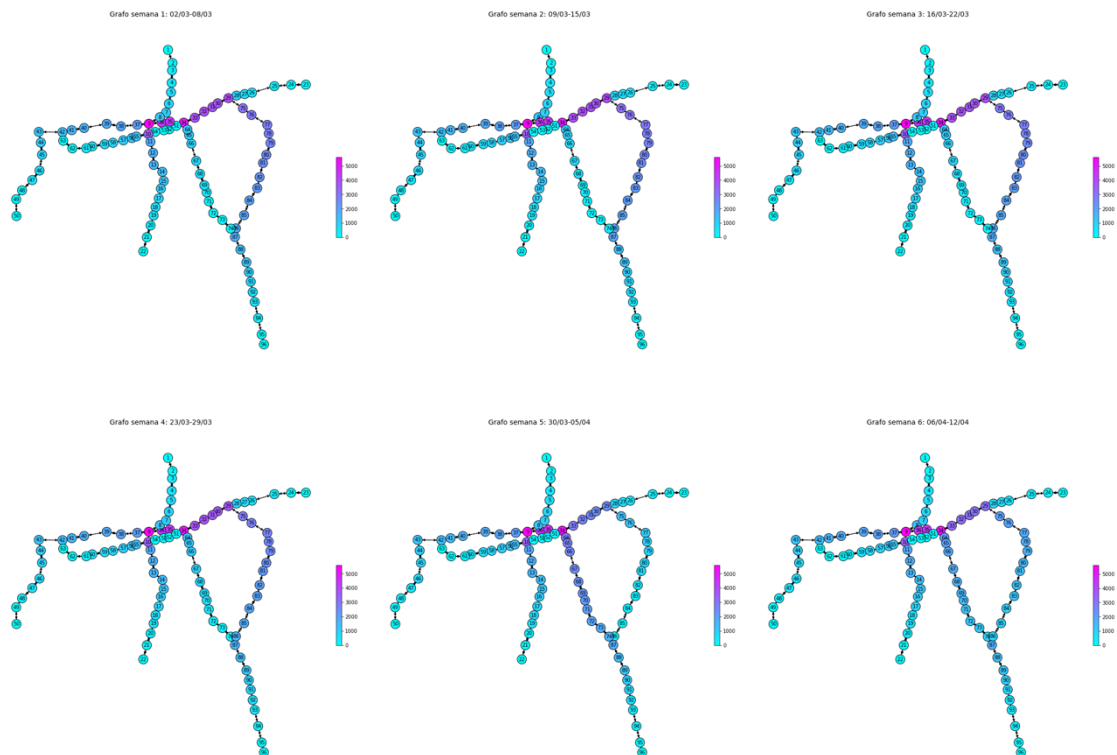


Figura 14. Comparación de cantidad de viajes entre las distintas estaciones de Metro en una semana sin cuarentena y otra durante la cuarentena.

Fuente: Elaboración Propia.

Cabe mencionar que las estaciones con mayor número de viajes tanto antes como durante la cuarentena corresponden a Baquedano y Salvador, con 109.944 viajes entre sí en la semana del 9 al 15 de marzo, disminuyendo a sólo 16.044 viajes entre sí en la semana del 23 al 29 del mismo mes.

Para analizar el comportamiento y cambios en los patrones de movilidad en la red de Metro, se construyó la red ponderada para 18 semanas: desde el 2 de marzo de 2020 hasta el 5 de julio de 2020, y se calculó el *betweenness centrality* de los nodos. En la Figura 16 se observan las redes semanales, donde el color de los nodos representa la centralidad de intermediación en cada grafo.



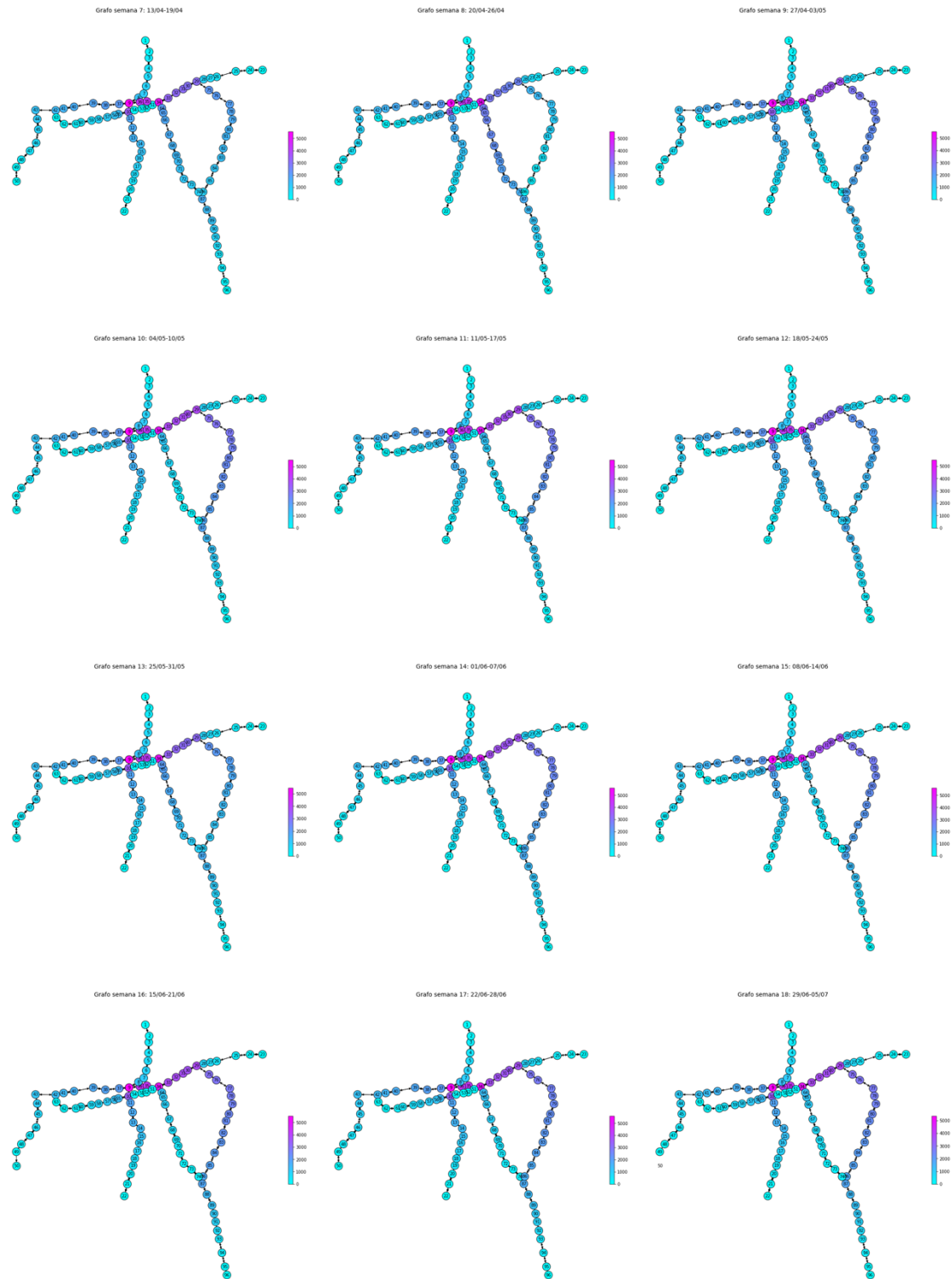
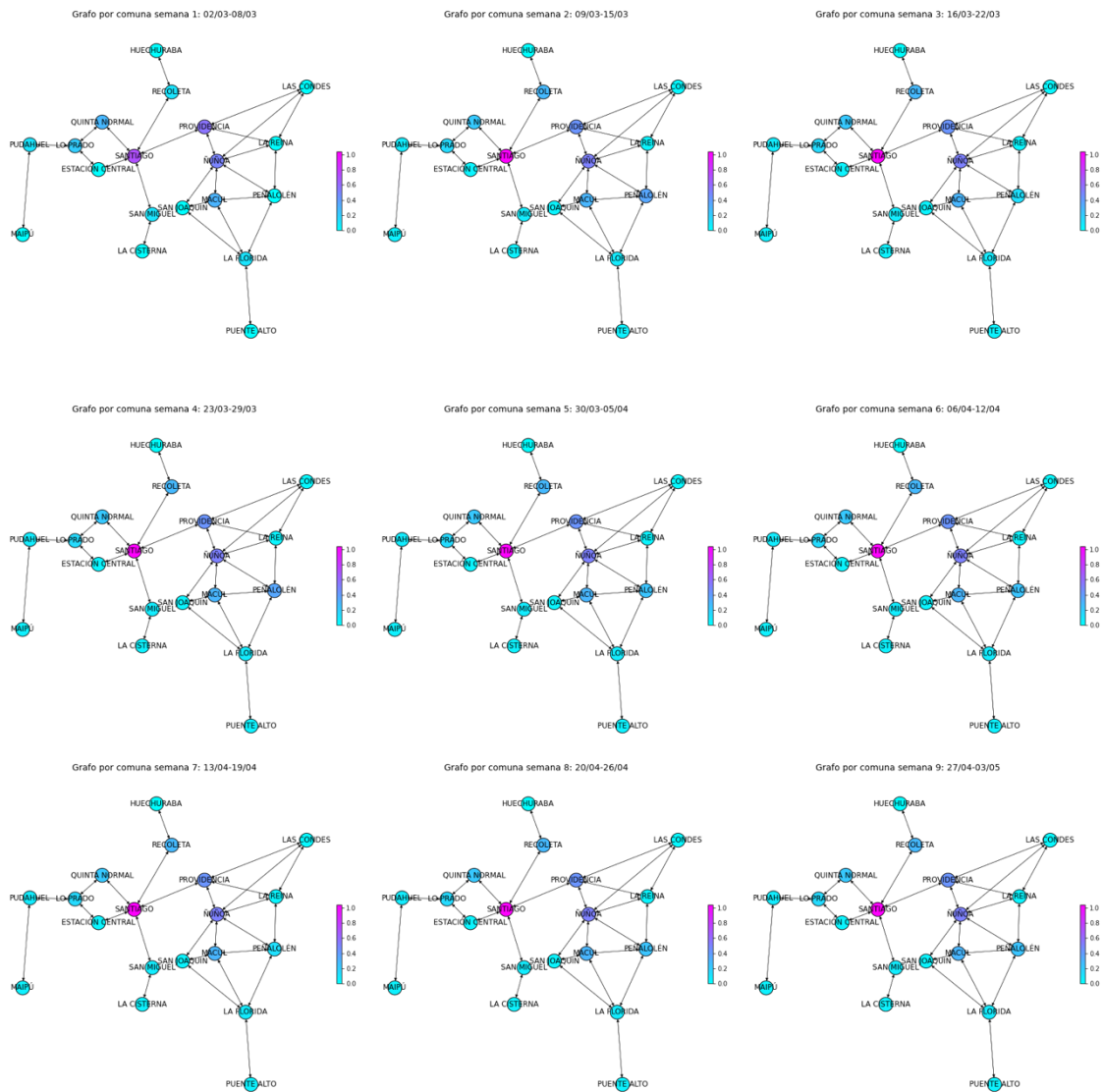


Figura 15. Grafos semanales de estaciones con color de nodo según centralidad de intermediación
Fuente: Elaboración Propia.

Tal como se observa en la Figura 16 y en el Anexo 3, la variación de la centralidad de intermediación entre las estaciones no varía significativamente en el tiempo, sin embargo, esto se prueba más adelante en las pruebas de hipótesis.

No obstante, dado que se planteó como objetivo el análisis por comuna, se realizó la agrupación de viajes según las comunas a las que sirve cada estación de Metro (Anexo 2), incluyendo los viajes intra-comuna, y se construyeron los respectivos grafos semanales.



Respecto de los grafos y las métricas calculadas, cabe señalar que el valor de *betweenness centrality* máximo normalizado es superior a 1 porque la librería NetworkX realiza la normalización considerando aristas sin *self-loops* (incluso con un objeto DiGraph como argumento), pero en este caso existen viajes al interior de un mismo nodo, que por simplicidad no se graficaron, pero se generan por viajes entre las estaciones al interior de una misma comuna, por ejemplo, los viaje desde Bellas Artes hasta La Moneda generan *self-loops* en el nodo Santiago.

Al realizar la agrupación de los viajes en comunas, se observa tanto en los grafos de la Figura 17 como en la Tabla 5, un notorio cambio en la centralidad de intermediación de la comuna de Santiago, que aumenta su *betweenness centrality* respecto de la primera semana de estudio, y se mantiene a lo largo de todos los periodos como la comuna con mayor capacidad de conectar flujos en lo que respecta a la red de transporte del Metro. Este cambio puede deberse a que otras comunas disminuyeron su cantidad de viajes en mayor proporción, motivo por el que Santiago aumenta su participación relativa dentro de la red, considerando que la métrica de centralidad de intermediación se encuentra normalizada para cada grafo semanal.

Por otra parte, Providencia es la comuna que más disminuyó su *betweenness centrality* entre la primera semana y las semanas siguientes, donde en la primera presentan con mayor capacidad conectiva junto a la comuna de Santiago, mientras que en las semanas subsiguientes esta capacidad conectiva es predominante sólo en la comuna de Santiago. De manera similar se aprecia que Quinta Normal y Lo Prado también disminuyen su métrica de manera relevante en cuarentena, por lo que la disminución absoluta de viajes fue proporcionalmente mayor en estas comunas, posiblemente porque son comunas mas bien residenciales.

Finalmente, como era de esperar, se encontró que las comunas como Maipú, La Cisterna, Puente Alto y Huechuraba presentan baja centralidad de intermediación dado que se encuentran en zonas extremas, por lo que las personas no pasan por estas comunas para llegar a otras, lo cual

implica una centralidad de intermediación cero o cercana a cero. No obstante, las comunas de Estación Central y San Joaquín también presentan centralidad de intermediación cero o cercanas a cero, siendo que su localización no es extrema. Este fenómeno puede deberse, en el caso de Estación Central, al supuesto de ruta más corta utilizado para la reconstrucción de viajes, ya que, por ejemplo, para ir desde Maipú hasta Santiago se recorren menos estaciones a través de Quinta Normal en comparación a pasar por Estación Central. En el caso de San Joaquín, el fenómeno de métrica cercana a cero se debe a estaciones fuera de servicio debido a los daños provocados en octubre de 2019, donde estaciones como San Joaquín(estación) fueron reparadas y puestas en marcha en noviembre de 2020, posterior a la fecha del periodo analizado[36].

5.4. Cálculo de correlaciones

1. Correlaciones de Pearson y de Spearman entre betweenness centrality y cantidad de casos nuevos de COVID-19 para todas las comunas agrupadas

Para calcular si existía correlación lineal de Pearson entre las variables, se concatenaron los datos de casos nuevos semanales por comuna (Tabla 2) para formar la variable Y de manera que $Y = [9, 5, 2, 24, 6, 43, 6, 13, 21, 26, 23, 12, 15, 31, 14, 15, 15, 14, 46, 36, 16, \dots, 274, 665]$. Análogamente, se concatenaron los datos de betweenness centrality semanal por comuna conformando un vector de 1×266 al igual que Y , considerando desde 1 hasta 4 semanas de desfase desde la fecha de la centralidad de intermediación hasta la fecha de casos nuevos. Los resultados obtenidos para la correlación lineal fueron los siguientes:

Desfase [semanas]	Coefficiente de Pearson	p-value	Observación
1	0,0765	0,2137	Resultado no significativo
2	0,0783	0,2029	Resultado no significativo
3	0,0742	0,2275	Resultado no significativo
4	0,0820	0,1824	Resultado no significativo

Tabla 6. Resultados correlación lineal entre centralidad de intermediación y cantidad de casos nuevos de COVID-19
Fuente: Elaboración Propia.

Por lo tanto, no es posible determinar si existe correlación lineal entre la métrica de betweenness centrality obtenida mediante la red ponderada por cantidad de viajes y la cantidad de casos nuevos de COVID-19 que se contabilizan 1, 2, 3 o 4 semanas después, debido a que los coeficientes obtenidos no son significativos.

A continuación, con los mismos vectores del proceso anterior, se analizó la existencia de correlación de Spearman entre la centralidad de intermediación y la cantidad de casos nuevos de COVID-19. El coeficiente de correlación de Spearman es no paramétrico (no asume distribución normal entre las muestras) y evalúa si dos variables se comportan de manera monótona, sin que la relación sea necesariamente lineal [30]. Sus valores fluctúan entre -1 y 1 al igual que el coeficiente de correlación de Pearson. Los resultados obtenidos para las comunas en forma agrupada se presentan a continuación.

Desfase [semanas]	Coefficiente de Spearman	p-value	Observación
1	0,0941	0,1258	Resultado no significativo
2	0,0956	0,1196	Resultado no significativo
3	0,0942	0,1261	Resultado no significativo
4	0,1251	0,0415	Resultado significativo

Tabla 7. Resultados correlación de Spearman entre betweenness centrality y cantidad de casos nuevos de COVID-19.
Fuente: Elaboración Propia.

Tal como se observa en la Tabla 7, el coeficiente de correlación de Spearman calculado no es de gran magnitud, pero sí es significativo para el caso de 4 semanas de desfase entre los viajes en Metro y los casos nuevos confirmados. El coeficiente de correlación de Spearman muestra que existe una correlación no lineal positiva de 0,1251 entre la centralidad de intermediación y cantidad de contagios nuevos de COVID-19 para todas las comunas en forma agrupada.

Respecto del desfase de 4 semanas, existen estudios que señalan una probabilidad del 5 al 10% de que el virus posea un periodo de incubación superior a 14 días [37]. Además del desfase entre la infección y la presentación de los primeros síntomas, existe el desfase desde ser consciente de los síntomas hasta buscar ayuda médica y realizarse el test PCR. Luego están las 24 a 72 horas de procesamiento del test PCR, y finalmente el desfase desde que se tiene el resultado de la prueba hasta que el caso es informado y contabilizado por las autoridades [23]. Considerando que la pandemia era una situación nueva y de la cual se tenía poca información, no es imposible que en un principio existiera gran desfase desde la infección de un individuo hasta que se contabilizara el caso.

2. Correlaciones de Pearson y de Spearman entre betweenness centrality y cantidad de casos nuevos de COVID-19 para cada una de las comunas

Luego de calcular las correlaciones de manera agrupada, se desagruparon los datos en vectores únicos según comuna, para luego calcular los coeficientes que se presentan en la siguiente tabla.

Comuna	Desfase [semanas]	Coefficiente de correlación Pearson	p-value	Coefficiente de correlación Spearman	p-value	Observación
ESTACIÓN CENTRAL	1	-	-	-	-	Centralidad constante
HUECHURABA	1	0,2688	0,1667	0,2786	0,1511	No significativo
LA CISTERNA	1	0,0561	0,7241	-0,0205	0,8977	No significativo
LA FLORIDA	1	0,3498	0,0082	0,2554	0,0574	Significativo para Pearson
LA REINA	1	0,3394	0,0040	0,2852	0,0167	Significativo
LAS CONDES	1	0,3223	0,0028	0,1959	0,0741	Significativo para Pearson
LO PRADO	1	0,1144	0,2620	0,1107	0,2777	No significativo
MACUL	1	0,0200	0,8341	0,0796	0,4044	No significativo
MAIPÚ	1	-0,0221	0,8060	0,0944	0,2929	No significativo
PEÑALOLÉN	1	0,0284	0,7390	0,1154	0,1747	No significativo
PROVIDENCIA	1	-0,0816	0,3144	0,0367	0,6517	No significativo
PUDAHUEL	1	-0,0854	0,2709	0,0417	0,5919	No significativo
PUENTE ALTO	1	-0,1184	0,1113	0,0297	0,6909	No significativo
QUINTA NORMAL	1	-0,1263	0,0777	0,0106	0,8825	No significativo
RECOLETA	1	-0,1117	0,1066	0,0204	0,7689	No significativo
SAN MIGUEL	1	-0,0878	0,1769	0,0356	0,5843	No significativo
SANTIAGO	1	0,0990	0,1168	0,0990	0,1168	No significativo
ÑUÑO A	1	0,0765	0,2137	0,0941	0,1258	No significativo
ESTACIÓN CENTRAL	2	-	-	-	-	Centralidad constante
HUECHURABA	2	0,2688	0,1667	0,2786	0,1511	No significativo
LA CISTERNA	2	0,0561	0,7241	-0,0205	0,8977	No significativo
LA FLORIDA	2	0,3498	0,0082	0,2554	0,0574	Significativo para Pearson
LA REINA	2	0,3394	0,0040	0,2852	0,0167	Significativo
LAS CONDES	2	0,3223	0,0028	0,1959	0,0741	Significativo para Pearson
LO PRADO	2	0,1144	0,2620	0,1107	0,2777	No significativo
MACUL	2	0,0200	0,8341	0,0796	0,4044	No significativo
MAIPÚ	2	-0,0221	0,8060	0,0944	0,2929	No significativo
PEÑALOLÉN	2	0,0383	0,6532	0,1179	0,1655	No significativo
PROVIDENCIA	2	-0,0770	0,3427	0,0389	0,6319	No significativo
PUDAHUEL	2	-0,0810	0,2966	0,0437	0,5735	No significativo
PUENTE ALTO	2	-0,1146	0,1235	0,0314	0,6740	No significativo
QUINTA NORMAL	2	-0,1229	0,0860	0,0133	0,8533	No significativo
RECOLETA	2	-0,1082	0,1178	0,0226	0,7444	No significativo
SAN MIGUEL	2	-0,0847	0,1929	0,0375	0,5651	No significativo
SANTIAGO	2	0,1011	0,1095	0,1008	0,1104	No significativo
ÑUÑO A	2	0,0783	0,2029	0,0956	0,1197	No significativo
ESTACIÓN CENTRAL	3	-	-	-	-	Centralidad constante
HUECHURABA	3	0,2688	0,1667	0,2786	0,1511	No significativo
LA CISTERNA	3	0,0561	0,7241	-0,0205	0,8977	No significativo
LA FLORIDA	3	0,3498	0,0082	0,2554	0,0574	Significativo para Pearson
LA REINA	3	0,3394	0,0040	0,2852	0,0167	Significativo
LAS CONDES	3	0,3223	0,0028	0,1959	0,0741	Significativo para Pearson
LO PRADO	3	0,1144	0,2620	0,1107	0,2777	No significativo
MACUL	3	0,0200	0,8341	0,0796	0,4044	No significativo
MAIPÚ	3	-0,0221	0,8060	0,0944	0,2929	No significativo
PEÑALOLÉN	3	0,0148	0,8622	0,1133	0,1824	No significativo
PROVIDENCIA	3	-0,0906	0,2636	0,0350	0,6666	No significativo
PUDAHUEL	3	-0,0937	0,2269	0,0403	0,6040	No significativo
PUENTE ALTO	3	-0,1246	0,0938	0,0284	0,7038	No significativo

QUINTA NORMAL	3	-0,1322	0,0647	0,0111	0,8777	No significativo
RECOLETA	3	-0,1172	0,0903	0,0191	0,7828	No significativo
SAN MIGUEL	3	-0,0931	0,1520	0,0348	0,5937	No significativo
SANTIAGO	3	0,0967	0,1259	0,0989	0,1175	No significativo
ÑUÑO A	3	0,0742	0,2275	0,0940	0,1261	No significativo
ESTACIÓN CENTRAL	4	0,3230	0,2600	0,4472	0,1089	No significativo
HUECHURABA	4	0,3437	0,0734	0,3901	0,0401	Significativo para Spearman
LA CISTERNA	4	0,1525	0,3351	0,1319	0,4052	No significativo
LA FLORIDA	4	0,3637	0,0059	0,3151	0,0180	Significativo
LA REINA	4	0,3590	0,0023	0,3605	0,0022	Significativo
LAS CONDES	4	0,3485	0,0012	0,2897	0,0075	Significativo
LO PRADO	4	0,1234	0,2261	0,1765	0,0822	No significativo
MACUL	4	0,0268	0,7790	0,1275	0,1804	No significativo
MAIPÚ	4	-0,0112	0,9007	0,1518	0,0898	No significativo
PEÑALOLÉN	4	0,0324	0,7043	0,1821	0,0313	Significativo para Spearman
PROVIDENCIA	4	-0,0820	0,3121	0,0928	0,2523	No significativo
PUDAHUEL	4	-0,0863	0,2659	0,0863	0,2662	No significativo
PUENTE ALTO	4	-0,1147	0,1232	0,0816	0,2737	No significativo
QUINTA NORMAL	4	-0,1246	0,0818	0,0551	0,4434	No significativo
RECOLETA	4	-0,1061	0,1253	0,0610	0,3788	No significativo
SAN MIGUEL	4	-0,0832	0,2009	0,0745	0,2520	No significativo
SANTIAGO	4	0,1052	0,0957	0,1341	0,0334	Significativo para Spearman
ÑUÑO A	4	0,0820	0,1824	0,1251	0,0415	Significativo para Spearman

Tabla 8. Resultados correlación de Pearson y de Spearman entre *betweenness centrality* y cantidad de casos nuevos de COVID-19 para cada comuna

Fuente: Elaboración Propia.

Como muestran los resultados, la comuna de La Reina presenta correlaciones significativas tanto para el coeficiente de Pearson como para el de Spearman, donde el primero varía entre 0,3394 y 0,3590 según el desfase que se considere para el análisis, mientras que el coeficiente de Spearman toma valores 0,2852 o 0,3605 según la cantidad de semanas de desfase que se aplique. El caso de esta comuna es consistente, ya que ambos coeficientes fueron positivos, indicando que cuando aumenta la centralidad de intermediación de la comuna La Reina debido a los viajes en Metro a través de sus estaciones, los casos nuevos de COVID-19 también aumentan 1, 2, 3 o 4 semanas después y lo hace de manera lineal. De modo similar, se obtuvo que las comunas de La Florida y Las Condes presentan coeficientes de Pearson positivos y significativos para todos los

desfases aplicados, lo cual indica que al aumentar el protagonismo de las estaciones que prestan servicios a estas comunas en términos de centralidad de intermediación, se produce un aumento en los casos nuevos confirmados de COVID-19 en las semanas posteriores. Cabe mencionar que correlación no necesariamente implica causalidad, sin embargo, existen estudios que relacionan la tasa de reproducción R del virus con la movilidad[37][38][39].

Respecto de otras comunas, Huechuraba, Peñalolén, Santiago y Ñuñoa, presentan coeficiente de Spearman positivos y significativos al aplicar 4 semanas de desfase entre los viajes y la cantidad de casos de nuevos contagiados COVID-19. Lo anterior indica que cuando aumenta la centralidad de intermediación, ocurre que después de 4 semanas la cantidad de casos de nuevos contagiados también lo hará; sin embargo, de manera no lineal.

Todas las correlaciones significativas encontradas fueron positivas, es decir, las variables aumentan en el mismo sentido; no obstante, son correlaciones “débiles” ya que su magnitud es inferior a 0.5 .

5.5. Contraste de hipótesis

¿Las comunas disminuyeron su centralidad de intermediación con la cuarentena?

Se aplicó la prueba estadística t-Student para comparación de medias de muestras dependientes, mediante la librería Scipy con la función `stats.ttest_rel` donde se consideró separar el análisis en 3 periodos: el primero corresponde las semanas 1 y 2 donde no existían restricciones de movilidad, el segundo periodo abarca desde la semana 3 hasta la

semana 11, mientras que el tercer periodo es desde la semana 12 hasta la semana 18. Para cada periodo se calculó el promedio de la centralidad de intermediación por comuna y luego se aplicó la prueba estadística, obteniéndose lo siguiente.

Comuna	BC Periodo 1	BC Periodo 2	BC Periodo 3	Resultado BC Periodo 1 y BC	Resultado BC Periodo 2 y BC
SANTIAGO	0,87361	1,04134	1,03901	Estadístico t = -1,3958 p-value = 0,1798	Estadístico t = 1,5491 p-value = 0,1388
PROVIDENCIA	0,50383	0,43904	0,43904		
ÑUÑO A	0,50825	0,51324	0,51324		
MACUL	0,30354	0,29989	0,29989		
QUINTA NORMAL	0,25704	0,22650	0,22417		
LO PRADO	0,18049	0,15183	0,15183		
PUDAHUEL	0,09533	0,07954	0,07954		
SAN MIGUEL	0,09472	0,07832	0,07832		
RECOLETA	0,19837	0,28562	0,28562		
LA FLORIDA	0,09701	0,08290	0,08290		
LAS CONDES	0,01585	0,03170	0,03170		
SAN JOAQUIN	0,02339	0,04678	0,04678		
PUENTE ALTO	0,02941	0,05882	0,05882		
LA REINA	0,02344	0,04687	0,04687		
PEÑALOLÉN	0,17917	0,21809	0,21081		
HUECHURABA	0,00000	0,00000	0,00000		
ESTACIÓN CENTRAL	0,02410	0,04820	0,04820		
LA CISTERNA	0,02941	0,05882	0,05882		
MAIPÚ	0,02941	0,05882	0,05882		

Tabla 9. Resultado prueba de hipótesis muestras dependientes para disminución de centralidad de intermediación
Fuente: Elaboración Propia.

Finalmente, no se rechaza la hipótesis nula, por lo que no es posible afirmar que la centralidad de intermediación disminuyó en promedio para las estaciones entre los periodos considerados. Este resultado es concordante con el resultado de correlaciones del apartado anterior, debido a que los casos de COVID-19 presentan un comportamiento monótonamente creciente en el periodo estudiado, por lo que si anteriormente se obtuvo que existe correlación positiva con la centralidad de intermediación, entonces no tendría sentido que en este apartado esa centralidad disminuyera.

6. Conclusiones

Las medidas de restricciones de movilidad impactaron en la cantidad de viajes en Metro, consiguiendo que estos disminuyeran en un 80% entre el 02 de marzo de 2020 y el 18 de mayo de 2020. No obstante, la disminución de la movilidad no está relacionada con la métrica de centralidad de intermediación que se construye mediante la red direccionada y ponderada por la cantidad de viajes. De esta manera, y mediante la aplicación de la prueba t-Student para muestras dependientes se obtuvo que el cambio en el patrón de viajes, en promedio, no causó una disminución significativa del betweenness centrality de las comunas, por lo que se rechaza la primera hipótesis planteada. En efecto, la comuna de Santiago es la que presenta mayor centralidad de intermediación tanto antes como durante las cuarentenas, siendo la que permite conectar más viajes dentro de la red incluso una vez disminuida la cantidad de viajes en un 80%.

Respecto de la centralidad de intermediación y su relación con la cantidad de casos nuevos de COVID-19, se comprobó la hipótesis planteada, ya que existe una correlación no lineal positiva significativa entre centralidad de intermediación y cantidad de nuevos casos a nivel agregado con desfase de 4 semanas. Adicionalmente, al desagregar la relación por comunas, se obtuvieron correlaciones significativas para 7 comunas a considerar desfase de 4 semanas.

Si bien se utilizaron antenas pertenecientes a la compañía Telefónica Movistar S.A., localizadas en las estaciones de Metro, la muestra tiene tamaño suficiente como para ser

representativa. Sin embargo, durante el periodo estudiado existían estaciones fuera de servicio por reparaciones, motivo por el cual se propone como trabajo futuro, poder realizar un análisis similar al presente, pero considerando la ventana de tiempo desde noviembre 2020 en adelante.

Por otra parte, si existiera la posibilidad de incorporar datos de movilidad en Transantiago, se podría ampliar el presente análisis a la red completa de transporte público y una hipótesis interesante a probar sería si la correlación entre centralidad de intermediación y casos de nuevos contagios se vuelve más fuerte en magnitud.

Finalmente, los resultados de este estudio permiten afirmar que efectivamente existe relación entre la movilidad y contagios, dejando abierta la posibilidad de que métricas de centralidad obtenidas desde la ciencia de redes, puedan ser incorporadas ya sea como predictores en modelos de *Machine Learning* o como parámetros en simulaciones estocásticas a fin de mejorar las herramientas que tenemos hoy para modelar enfermedades virales.

7. Bibliografía

1. Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19). World Health Organization. [Web; accedido el 03-09-2020]
<https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
2. COVID-19 Map. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). [Web; accedido el 12-06-2021]
<https://coronavirus.jhu.edu/map.html>
3. Orientaciones para el público. World Health Organization. [Web; accedido el 03-09-2020] <https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public>
4. 24horas.cl: Mañalich asegura que se está siguiendo una estrategia de “cuarentena dinámica”. 24horas.cl (2020). [Web; accedido el 03-09-2020]
<https://www.24horas.cl/coronavirus/ministro-manalich-explica-estrategia-de-cuarentena-dinamica-y-estrategica-del-gobierno-4094778>
5. Metro: Memoria Anual. Metro de Santiago (2019). [Web; accedido el 03-09-2020]
<https://www.metro.cl/documentos/memoria-anual-2019.pdf>
6. INE: Proyección base 2017. Estimaciones y proyecciones 2002-2035, comunas. Instituto Nacional de Estadística. [Web; accedido el 03-09-2020]
<https://www.ine.cl/estadisticas/sociales/demografia-y-vitales/proyecciones-de-poblacion>
7. Metro: Memoria Anual. Metro de Santiago (2020). [Web; accedido el 12-06-2021]
<https://www.metro.cl/documentos/memoria-anual-2020.pdf>

8. Kermack, W., McKendrick, A.: Contributions to the mathematical theory of epidemics. *Proc. Royal Soc. A* 115, 700–721 (1927)
9. Batisse, D., Benech, N., Botelho-Nevers, E., Bouiller, K., Collarino, R., Conrad, A., Gallay, L., Goehringer, F., Gousseff, M., Joseph, C., et al.: Clinical recurrences of covid-19 symptoms after recovery: viral relapse, reinfection or inflammatory rebound? *Journal of Infection* (2020)
10. To, K.K.-W., Hung, I.F.-N., Ip, J.D., Chu, A.W.-H., Chan, W.-M., Tam, A.R., Fong, C.H.-Y., Yuan, S., Tsoi, H.-W., Ng, A.C.-K., Lee, L.L.-Y., Wan, P., Tso, E., To, W.-K., Tsang, D., Chan, K.-H., Huang, J.-D., Kok, K.-H., Cheng, V.C.-C., Yuen, K.-Y.: COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. *Clinical Infectious Diseases* (2020). doi:10.1093/cid/ciaa1275. <https://academic.oup.com/cid/advance-article-pdf/doi/10.1093/cid/ciaa1275/33681713/ciaa1275.pdf>
11. Manrique-Abril, F.G., Agudelo-Calderon, C.A., González-Chordá, V.M., Gutiérrez-Lesmes, O., Téllez-Piñerez, C.F., Herrera-Amaya, G.: Sir model of the covid-19 pandemic in colombia. *Revista de Salud Pública* 22(1) (2020)
12. Daron Acemoglu, I.W.M.D.W. Victor Chernozhukov: Optimal targeted lockdowns in a multi-group sir model. *National Bureau of Economic Research* (2020)
13. Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A.: The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences* 103(7), 2015–2020 (2006)
14. Barabási, A.L.: *Network Science* (2013). doi:10.1098/rsta.2012.0375
15. Gardiner, C.W.: *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences*. Springer, Berlin (2004)

16. Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.J., Vespignani, A.: Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Medicine* 4 (1), 0095–0110 (2007). doi:10.1371/journal.pmed.0040013
17. Balcan, D., Colizza, V., Gonçalves, B., Hud, H., Ramasco, J.J., Vespignani, A.: Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America* 106 (51), 21484–21489 (2009). doi:10.1073/pnas.0906910106
18. Dietz, K.: The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research* (1993). doi:10.1177/096228029300200103
19. Delamater, P.L., Street, E.J., Leslie, T.F., Yang, Y.T., Jacobsen, K.H.: Complexity of the basic reproduction number (R_0). *Emerging Infectious Diseases* (2019). doi:10.3201/eid2501.171901
20. Wang, Y., You, X.Y., Wang, Y.J., Peng, L.P., Du, Z.C., Gilmour, S., Yoneoka, D., Gu, J., Hao, C., Hao, Y.T., Li, J.H.: Estimating the basic reproduction number of COVID-19 in Wuhan, China. *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi* (2020). doi:10.3760/cma.j.cn112338-20200210-00086
21. Barrio, R.A., Varea, C., Govezensky, T., Jose, M.V.: Modeling the geographical spread of influenza A(H1N1): the case of Mexico. *Applied Mathematical Sciences* 7 (44), 2143–2176 (2013). doi:10.12988/ams.2013.13193
22. Barrio, R.A., Kaski, K.K., Haraldsson, G.G., Aspelund, T., Govezensky, T.: Modelling covid-19 epidemic in mexico, finland and iceland. *arXiv preprint arXiv:2007.10806* (2020)
23. Ferguson, N. M., Ainslie, K. E. C., Walters, C. E., Fu, H., Bhatia, S., Wang, H., Xi, X., Baguelin, M., Bhatt, S., Boonyasiri, A., Boyd, O., Cattarino, L., Ciavarella, C.,

- Cucunuba, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., van Elsland, S. L., FitzJohn, R., Riley, S. (2020). Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment. *Wellcome Open Research*, 5(March), 1–8. <https://doi.org/10.12688/wellcomeopenres.15843.1>
24. Especial Análisis Tráfico Internet Enero-Abril 2020 (2020). Subsecretaría de Telecomunicaciones. [Web; accedido el 03-09-2020].
https://www.subtel.gob.cl/wp-content/uploads/2020/06/PPT_Series_MARZO_2020_V1.pdf
25. Torres, M., Paz, K., & Salazar, F. (2006). Tamaño de una muestra para una investigación de mercado. *Boletín electrónico*, 2, 1-13.
26. Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2), 136–145.
<https://doi.org/10.1016/j.socnet.2007.11.001>
27. Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.
28. Pearson, K. (1895). Correlation coefficient. In *Royal Society Proceedings* (Vol. 58, p. 214).
29. Qin, J., You, C., Lin, Q., Hu, T., Yu, S., & Zhou, X. H. (2020). Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. *Science advances*, 6(33), eabc1202.
30. Spearman, C. (1987). The proof and measurement of association between two things. *The American journal of psychology*, 100(3/4), 441-471.
31. Student. (1908). The Probable Error of a Mean. *Biometrika*, 6(1), 1-25.
[doi:10.2307/2331554](https://doi.org/10.2307/2331554)

32. Ministerio de Salud (2020). Presidente anuncia suspensión de clases y reduce actos públicos. <https://www.minsal.cl/presidente-anuncia-suspension-de-clases-y-reduce-actos-publicos/> [Web; accedido el 14-06-2021].
33. Ministerio de Salud (2020). Ministerio de salud decreta cuarentena total para ciudad de Santiago y seis comunas aledañas. <https://www.minsal.cl/ministerio-de-salud-decreta-cuarentena-total-para-la-ciudad-de-santiago-y-seis-comunas-aledanas/> [Web; accedido el 14-06-2021].
34. Informe Ejecutivo: Encuesta Origen Destino de Viajes 2012. Ministerio de Transportes y Telecomunicaciones Programa de Vialidad y Transporte Urbano: SECTRA (2012). <http://www.sectra.gob.cl/biblioteca/detalle1.asp?mfn=3253>
35. Dictamen 1239/005 Salud y Seguridad en el Trabajo. Dirección del Trabajo, Departamento Jurídico y Fiscalía, Santiago, Chile, 19 de marzo de 2020.
36. 24 Horas (2020). Siete estaciones de Metro volverán a operar y habrá cambios en el funcionamiento de buses para este lunes 18. <https://www.24horas.cl/tesirve/siete-estaciones-de-metro-volveran-a-operar-y-habra-cambios-en-el-funcionamiento-de-buses-para-este-lunes-18-3732336> [Web; accedido el 14-06-2021].
37. Kajitani, Y., & Hatayama, M. (2021). Explaining the effective reproduction number of COVID-19 through mobility and enterprise statistics: Evidence from the first wave in Japan. *Plos one*, 16(3), e0247186.
38. Nouvellet, P., Bhatia, S., Cori, A., Ainslie, K. E., Baguelin, M., Bhatt, S., ... & Donnelly, C. A. (2021). Reduction in mobility and COVID-19 transmission. *Nature communications*, 12(1), 1-9.
39. Noland, R. B. (2021). Mobility and the effective reproduction rate of COVID-19. *Journal of Transport & Health*, 20, 101016.

8. Anexos

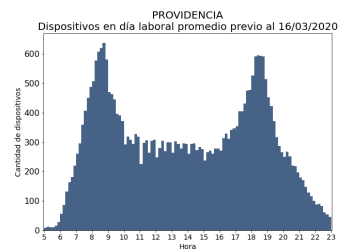
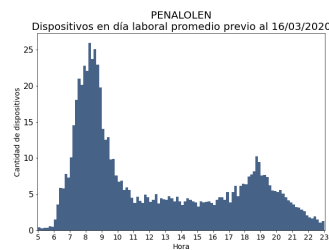
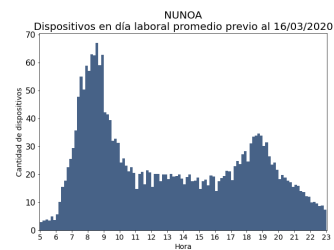
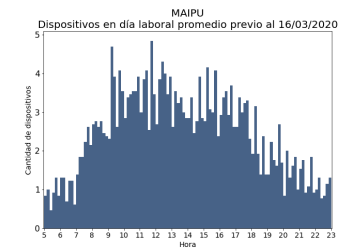
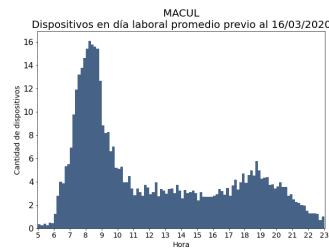
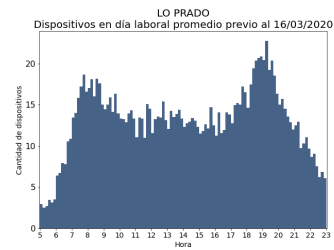
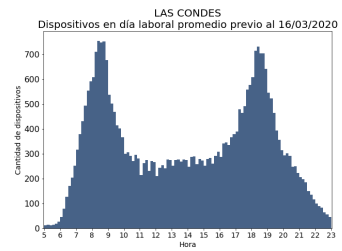
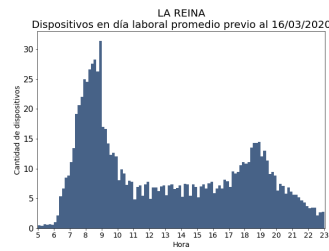
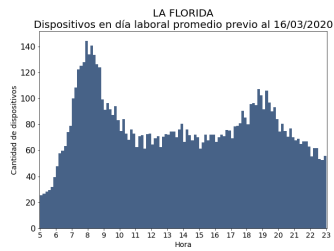
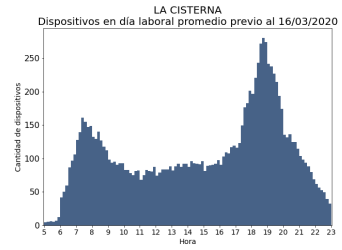
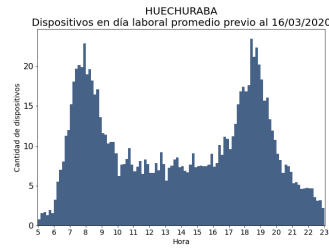
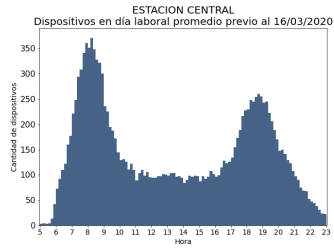
Anexo 1: Antenas por estación

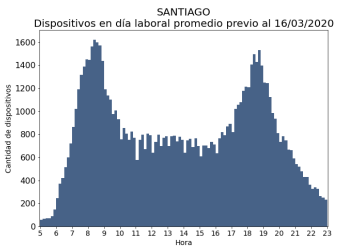
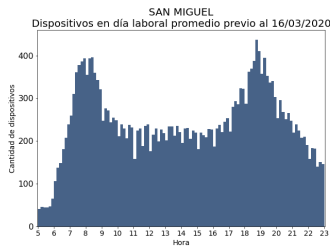
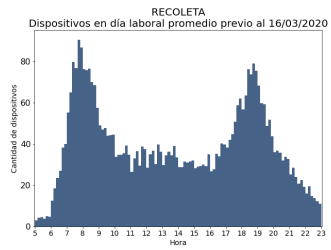
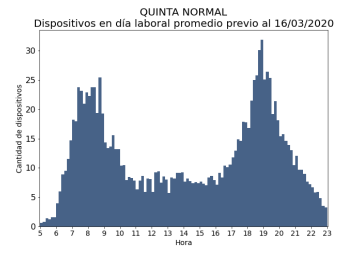
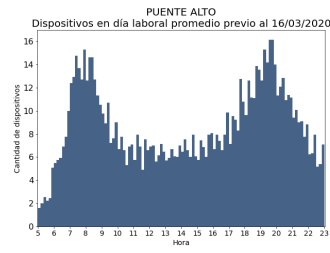
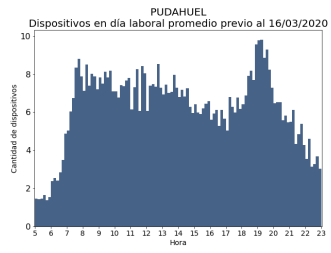
Nº	ESTACIÓN	ANTENAS	Nº	ESTACIÓN	ANTENAS
1	ALCÁNTARA	1	41	MANUEL MONTT	1
2	BAQUEDANO	2	42	NEPTUNO	1
3	BARRANCAS	1	43	ÑUBLE	1
4	BELLAS ARTES	1	44	PAJARITOS	1
5	BLANQUEADO	1	45	PARQUE BUSTAMANTE	1
6	CEMENTERIOS	1	46	PARQUE O'HIGGINS	1
7	CERRO BLANCO	1	47	PATRONATO	1
8	CIUDAD DEL NIÑO	1	48	PEDRO DE VALDIVIA	1
9	CRISTÓBAL COLÓN	1	49	PLAZA DE ARMAS	1
10	CUMMING	1	50	PLAZA DE MAIPÚ	1
11	DEL SOL	1	51	PLAZA DE PUENTE ALTO	1
12	DEPARTAMENTAL	1	52	PLAZA EGAÑA	1
13	DORSAL	1	53	PRÍNCIPE DE GALES	1
14	ECUADOR	1	54	PUDAHUEL	1
15	EINSTEIN	1	55	PUENTE CAL Y CANTO	1
16	EL GOLF	1	56	QUINTA NORMAL	1
17	EL LLANO	1	57	REPÚBLICA	1
18	EL PARRÓN	1	58	ROJAS MAGALLANES	1
19	ESCUELA MILITAR	1	59	RONDIZZONI	1
20	ESTACIÓN CENTRAL	1	60	SALVADOR	1
21	FRANCISCO BILBAO	1	61	SAN ALBERTO HURTADO	1
22	FRANKLIN	1	62	SAN JOAQUIN	1
23	GRECIA	1	63	SAN MIGUEL	1
24	GRUTA DE LOURDES	1	64	SAN PABLO	2
25	HERNANDO DE MAGALLANES	1	65	SANTA ANA	2
26	IRARRÁZVAL	1	66	SANTA ISABEL	1
27	LA CISTERNA	1	67	SANTA LUCÍA	1
28	LA MONEDA	1	68	SANTIAGO BUERAS	1
29	LAGUNA SUR	1	69	SIMÓN BOLÍVAR	2
30	LAS MERCEDES	1	70	TOBALABA	2
31	LAS REJAS	2	71	TOESCA	1
32	LO OVALLE	1	72	ULA	1
33	LO PRADO	1	73	UNIVERSIDAD CATÓLICA	1
34	LO VIAL	1	74	UNIVERSIDAD DE CHILE	4
35	LOS DOMINICOS	1	75	USACH	1
36	LOS HÉROES	2	76	VESPUCIO NORTE	1
37	LOS LEONES	1	77	VICENTE VALDES	2
38	LOS ORIENTALES	1	78	VICUÑA MACKENNA	1
39	LOS PRESIDENTES	1	79	ZAPADORES	1
40	MANQUEHUE	1			

Anexo 2: Estaciones consideradas y comunas a las que sirven

Nº	ESTACIÓN	COMUNAS A LAS QUE SIRVE	Nº	ESTACIÓN	COMUNAS A LAS QUE SIRVE
1	ALCÁNTARA	LAS CONDES	40	MANUEL MONTT	PROVIDENCIA
2	BAQUEDANO	PROVIDENCIA	41	NEPTUNO	LO PRADO
3	BARRANCAS	LO PRADO, PUDAHUEL	42	ÑUBLE	ÑUÑO A
4	BELLAS ARTES	SANTIAGO	43	PAJARITOS	LO PRADO, ESTACIÓN CENTRAL
5	BLANQUEADO	QUINTA NORMAL, LO PRADO	44	PARQUE BUSTAMANTE	PROVIDENCIA
6	CEMENTERIOS	RECOLETA	45	PARQUE O'HIGGINS	SANTIAGO
7	CERRO BLANCO	RECOLETA	46	PATRONATO	RECOLETA
8	CIUDAD DEL NIÑO	SAN MIGUEL	47	PEDRO DE VALDIVIA	PROVIDENCIA
9	CRISTÓBAL COLÓN	LAS CONDES, PROVIDENCIA	48	PLAZA DE ARMAS	SANTIAGO
10	CUMMING	SANTIAGO	49	PLAZA DE MAIPÚ	MAIPÚ
11	DEL SOL	MAIPÚ	50	PLAZA DE PUENTE ALTO	PUENTE ALTO
12	DEPARTAMENTAL	SAN MIGUEL	51	PLAZA EGAÑA	LA REINA, ÑUÑO A
13	DORSAL	RECOLETA	52	PRÍNCIPE DE GALES	LA REINA, ÑUÑO A
14	ECUADOR	ESTACIÓN CENTRAL	53	PUDAHUEL	LO PRADO, PUDAHUEL
15	EINSTEIN	RECOLETA	54	PUENTE CAL Y CANTO	SANTIAGO
16	EL GOLF	LAS CONDES	55	QUINTA NORMAL	QUINTA NORMAL, SANTIAGO
17	EL LLANO	SAN MIGUEL	56	REPÚBLICA	SANTIAGO
18	EL PARRÓN	LA CISTERNA	57	ROJAS MAGALLANES	LA FLORIDA
19	ESCUELA MILITAR	LAS CONDES	58	RONDIZZONI	SANTIAGO
20	ESTACIÓN CENTRAL	ESTACIÓN CENTRAL, SANTIAGO	59	SALVADOR	PROVIDENCIA
21	FRANCISCO BILBAO	LAS CONDES, PROVIDENCIA	60	SAN ALBERTO HURTADO	ESTACIÓN CENTRAL
22	FRANKLIN	SANTIAGO, SAN MIGUEL	61	SAN MIGUEL	SAN MIGUEL
23	GRECIA	PEÑALOLÉN, ÑUÑO A, MACUL	62	SAN PABLO	LO PRADO
24	GRUTA DE LOURDES	QUINTA NORMAL	63	SANTA ANA	SANTIAGO
25	HERNANDO DE MAGALLANES	LAS CONDES	64	SANTA ISABEL	PROVIDENCIA
26	IRARRÁZVAL	ÑUÑO A	65	SANTA LUCÍA	SANTIAGO
27	LA CISTERNA	LA CISTERNA	66	SANTIAGO BUERAS	MAIPÚ
28	LA MONEDA	SANTIAGO	67	SIMÓN BOLÍVAR	LA REINA, ÑUÑO A
29	LAS MERCEDES	PUENTE ALTO	68	TOBALABA	LAS CONDES, PROVIDENCIA
30	LAS REJAS	ESTACIÓN CENTRAL	69	TOESCA	SANTIAGO
31	LO OVALLE	LA CISTERNA	70	ULA	SANTIAGO
32	LO PRADO	LO PRADO	71	UNIVERSIDAD CATÓLICA	SANTIAGO
33	LO VIAL	SAN MIGUEL	72	UNIVERSIDAD DE CHILE	SANTIAGO
34	LOS DOMINICOS	LAS CONDES	73	USACH	ESTACIÓN CENTRAL
35	LOS HÉROES	SANTIAGO	74	VESPUCCIO NORTE	RECOLETA, HUECHURABA
36	LOS LEONES	PROVIDENCIA	75	VICENTE VALDÉS	LA FLORIDA
37	LOS ORIENTALES	PEÑALOLÉN, ÑUÑO A	76	VICUÑA MACKENNA	LA FLORIDA
38	LOS PRESIDENTES	PEÑALOLÉN, MACUL	77	ZAPADORES	RECOLETA
39	MANQUEHUE	LAS CONDES			

Anexo 4: Distribución horaria Metro por comuna, previo al 16/03/2020





Anexo 5: Distribución horaria Metro por comuna, posterior al 16/03/2020

