

Reporte de situación: movilidad y covid-19 en Colombia

Un informe de la Universidad del Rosario

Autores Principales

Felipe González-Casabianca
Alejandro Feged-Rivadeneira

Colaboradores

Andrés Ángel
Vladimir Corredor

REPORTE DE SITUACIÓN: MOVILIDAD Y COVID-19 EN COLOMBIA

Un informe de la Universidad del Rosario

Autores Principales

Felipe González-Casabianca
Universidad del Rosario

Alejandro Feged-Rivadeneira
Universidad del Rosario
alejandro.feged@urosario.edu.co

Colaboradores

Andrés Ángel
Universidad de Los Andes
ja.angel1908@uniandes.edu.co

Vladimir Corredor
Universidad Nacional
vcorredore@unal.edu.co

10 de julio de 2020

1. Destacado

- Se definen unidades funcionales con base en movimiento humano en tiempo real. Para cada unidad se reporta en gráficas que muestran tendencias en el tiempo y mapas.
- Las regiones Magdalena medio, Cúcuta, Choachí, Montelíbano, e Ituango han disminuido menos que otras regiones en sus movimientos (o han aumentado). Las zonas de tránsito pueden ser un foco vulnerable durante la cuarentena pues tienen con menor posibilidad de hacer aislamiento social, y no cuentan con servicios de salud robustos.
- Se observa un aumento en la movilidad urbana en diferentes ciudades del país. Se observa mayor movimiento en los sectores de menor ingreso de las ciudades.

2. Importante

El objetivo de este reporte es proveer información diaria y semanal a entidades gubernamentales sobre el efecto de las distintas medidas de restricción de la movilidad y distanciamiento social a nivel nacional y regional como herramienta para evaluar su efectividad en el control de la pandemia COVID-19.

3. Introducción

Este informe hace parte de una serie de documentos elaborados para aportar en la toma de decisiones en salud pública relacionada con la epidemia de SARS-COV-2 en Colombia. En este documento se presentan cambios en patrones de movilidad, a la fecha y acumulados, a partir de datos agregados. Los datos utilizados provienen del servicio GeoInsights, de Facebook.

Con base en los datos analizados, se busca identificar patrones de movilidad y establecer magnitudes de flujo de población entre y al interior de unidades espaciales definidas administrativamente o a través de esta herramienta. Para

ello se utilizó teoría de redes sociales, y se calcularon las comunidades con base en los flujos observados. Este trabajo está inspirado en la idea de mega-regiones[1], en la cual se utilizan viajes diarios para identificar grupos de unidades espaciales que componen unidades funcionales y/o administrativas. Otros ejemplos de trabajo teórico y empírico han abordado el tema de mega-regiones para distintos contextos[2, 3, 4]. El reporte genera análisis al nivel de las unidades detectadas por el algoritmo de detección de comunidades.

Para cada unidad se caracterizan las dinámicas de movilidad y casos por medio de estadística descriptiva, junto con análisis topológico de datos, y un ejercicio de aprendizaje de máquina.

Se discuten brevemente los resultados (semanalmente o con eventos atípicos), y las gráficas son actualizadas de manera diaria.

4. Objetivo

Proveer información para la toma de decisiones en políticas públicas asociadas a la pandemia de SARS-COV-2. Dada la escala de los datos de movilidad disponibles, y su relación con las diferentes medidas adoptadas de manera progresiva con respecto a la restricción/flexibilización de los contactos sociales, es crucial medir los flujos entre poblaciones para hacer seguimiento y controlar la diseminación de la infección en áreas con capacidades de respuesta en atención en salud diferentes. Este documento es un insumo para evaluar la efectividad de las medidas de aislamiento entre poblaciones estructuradas y establecer la magnitud de flujos poblacionales entre y al interior de unidades espaciales. Esta iniciativa persigue los siguientes objetivos:

- Identificar cómo se conectan por medio de la movilidad humana unidades espaciales definidas como municipios, localidades u otras unidades administrativas a través del tiempo.
- Identificar el riesgo de importar casos entre unidades espaciales dado un contexto epidemiológico y proveer una radiografía de los cambios en la movilidad humana a distintos tiempos.

5. Justificación

La movilidad humana constituye el principal riesgo de importación y diseminación de la infección por SARS-COV-2. Por consiguiente el efecto de las medidas adoptadas para reducir la movilidad encaminadas a disminuir la probabilidad de contactos, junto con información epidemiológica confiable, permiten hacer inferencias acerca del crecimiento de la infección y su efecto sobre la capacidad de respuesta del sistema de salud a nivel nacional y regional. Con base en esta información, es posible evaluar el riesgo de decisiones de aislamiento/distanciamiento social a nivel regional, diseñar políticas de reducción de la movilidad, de aislamientos parciales espacialmente segregados que permitan la movilidad de sectores y regiones dependiendo del riesgo, y focalizar esfuerzos en unidades espaciales donde hay distintos tipos de amenaza. Asimismo, constituye una herramienta para la aplicación de manera informada de pruebas serológicas que permitan establecer la magnitud de la población expuesta en regiones con diferentes conectividades. La utilidad de esta herramienta para las políticas de salud encaminadas a la contención de la epidemia por SARS-COV2 dependen en gran medida de la confiabilidad y disponibilidad de datos epidemiológicos sobre la infección.

6. Descripción de los datos

Las figuras de este reporte son generadas usando datos de movimiento recolectados por Facebook. Estos datos representan usuarios de Facebook en Colombia (y otros lugares) que tienen los servicios de ubicación habilitados. Los datos son representados de manera agregada en intervalos de tiempo de ocho horas (son datos anonimizados y agregados). Los datos de referencia corresponden a una recopilación de estos movimientos desde abril 4 de este año. Para este informe, los datos están agregados diariamente a nivel de municipio.

7. Resultados

Con base en los datos disponibles, aquí presentamos la situación de movilidad y casos de COVID-19 en Colombia.

7.1. Cambios Movilidad a Nivel Nacional

Las siguientes figuras (1-2) muestran el cambio de movilidad a nivel nacional y en lugares específicos desde principios de marzo hasta la última actualización de datos. Se toma como base el movimiento de promedio del mes de febrero. Se puede observar que la movilidad comienza a reducirse significativamente desde el inicio de la epidemia (marzo 6) en el territorio nacional, llegando a principios de abril con la menor cantidad de movimiento. Desde entonces se ve el crecimiento constante del movimiento a nivel nacional.

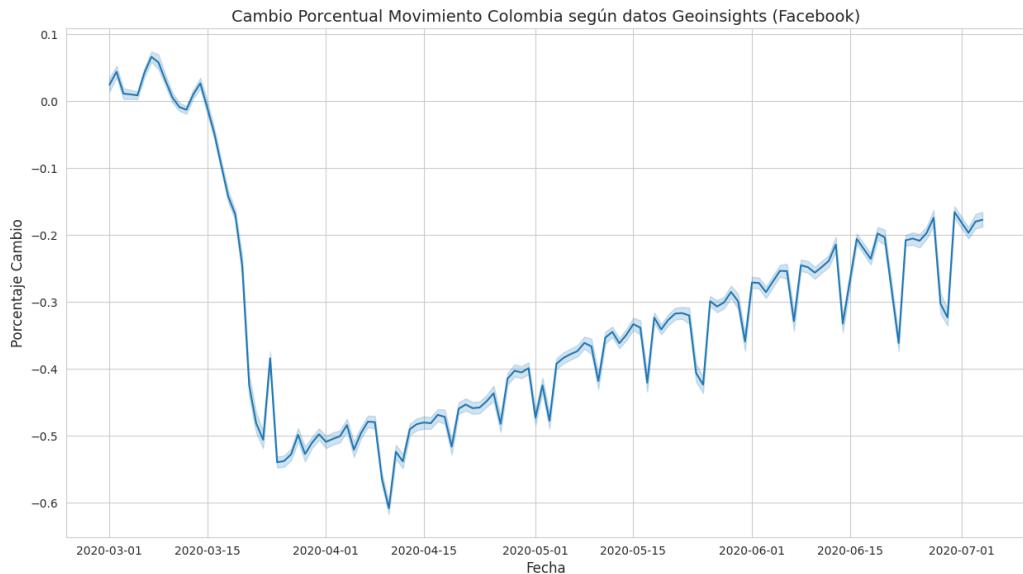


Figura 1: **Cambio de movilidad en Colombia.**

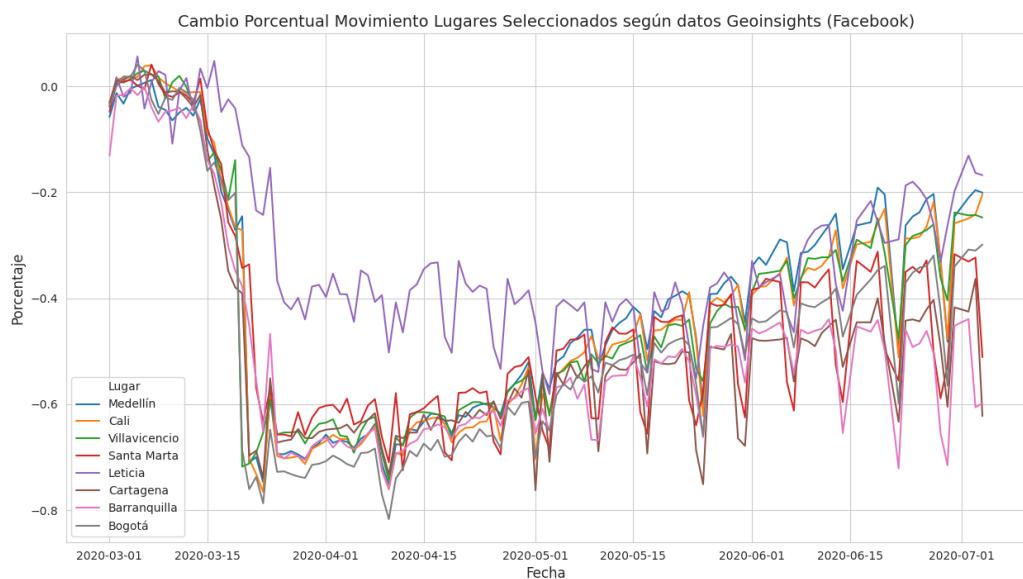
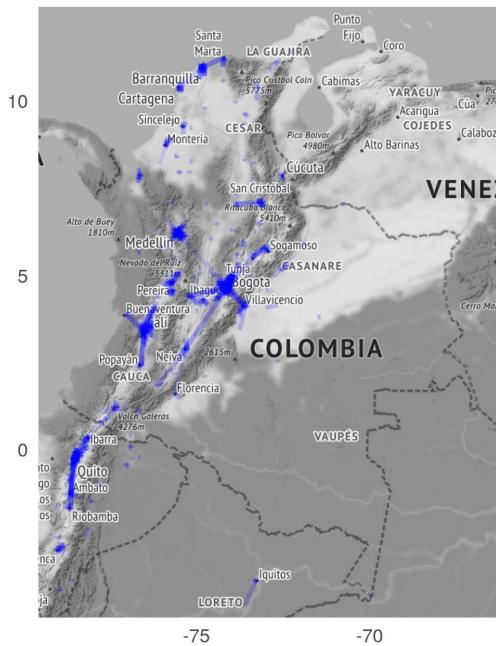


Figura 2: **Cambio de movilidad en lugares seleccionados.**

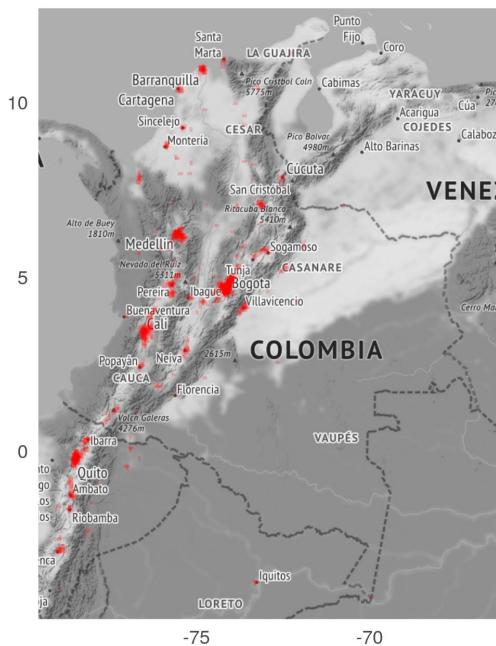
7.2. Cambios en la Movilidad (Visualización Geográfica)

Las siguientes figuras (3-4) muestran el movimiento de hace un mes (arriba izquierda), la semana anterior (abajo izquierda) y su cambio (derecha). En la figura del cambio, zonas resaltadas en azul muestran una reducción y las rojas muestran un aumento respecto al mes anterior.

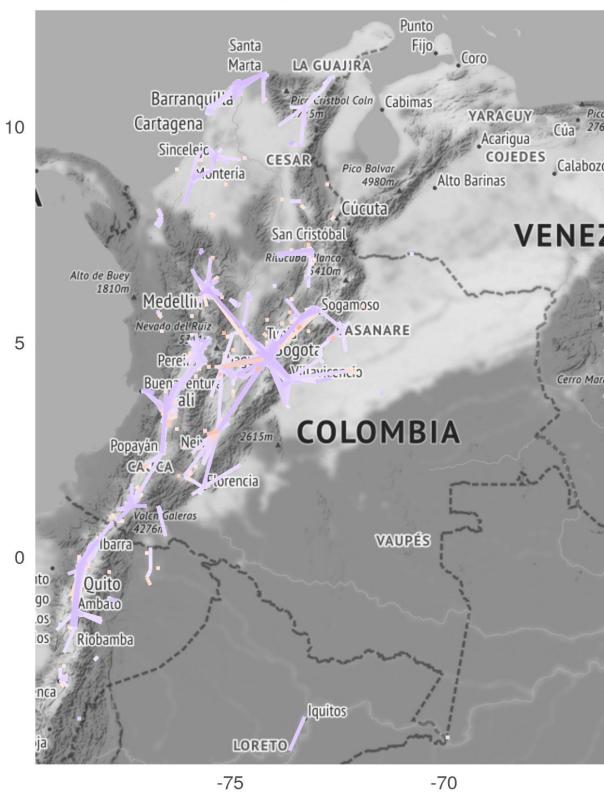
Un Mes Atrás



Durante Últimos 7 Días



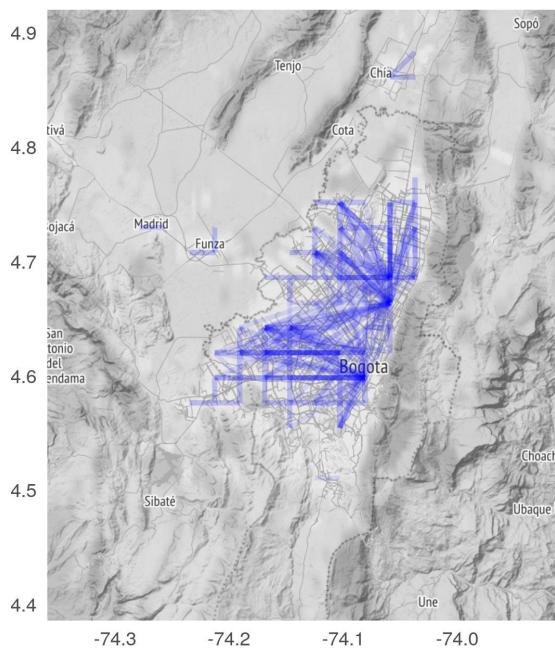
Cambio Promedio Últimos 7 Días
Cambios Netos Mayores a 25 %



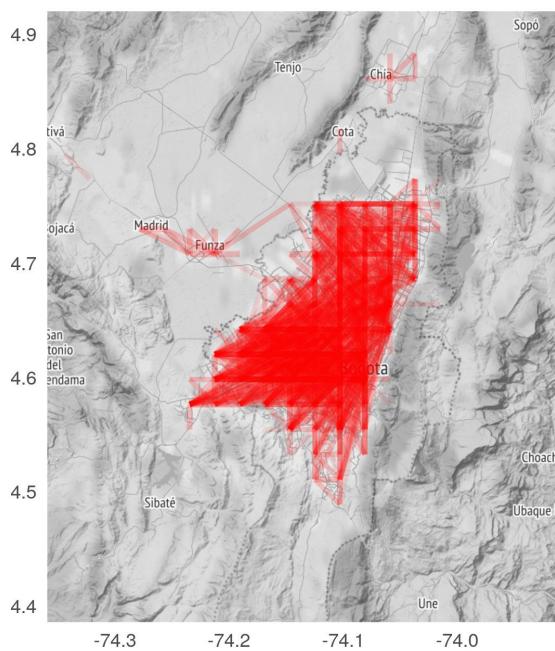
Cambio % Movimiento
-200-100 0 100 200

Figura 3: Cambio de movilidad en Colombia a fecha de 10 de julio de 2020.

Un Mes Atrás



Durante Últimos 7 Días



Cambio Promedio Últimos 7 Días
Cambios Netos Mayores a 150 %

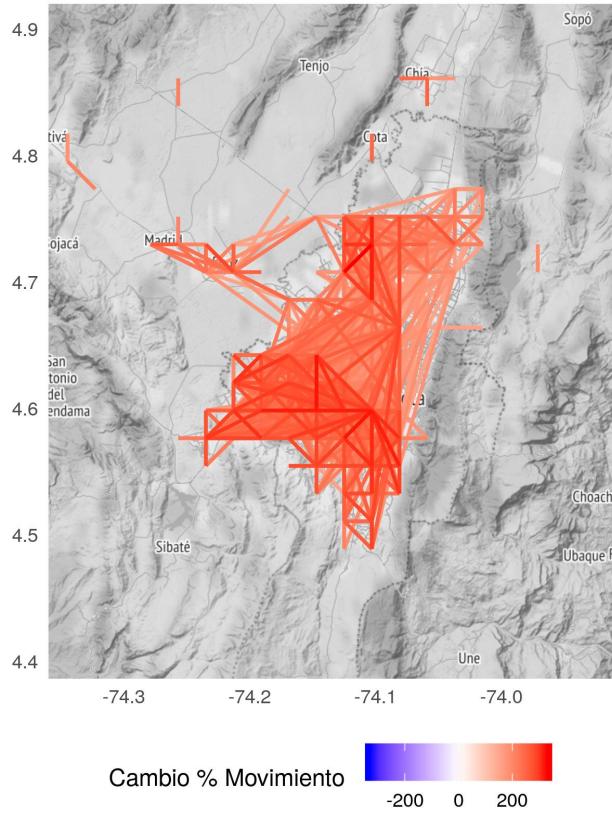


Figura 4: Cambio de movilidad en Bogotá a fecha de 10 de julio de 2020.

7.3. Cambios en la Movilidad por Unidad Espacial

Las siguientes figuras (5-6) muestran el cambio de movilidad interna y externa para las unidades espaciales encontradas (ver sección 11.1). Movimiento externo se refiere al movimiento entre unidades y el interno hace referencia al movimiento al interior de la unidad. Se toma como base el movimiento promedio de la primera semana de los datos utilizados: 4-10 de abril.

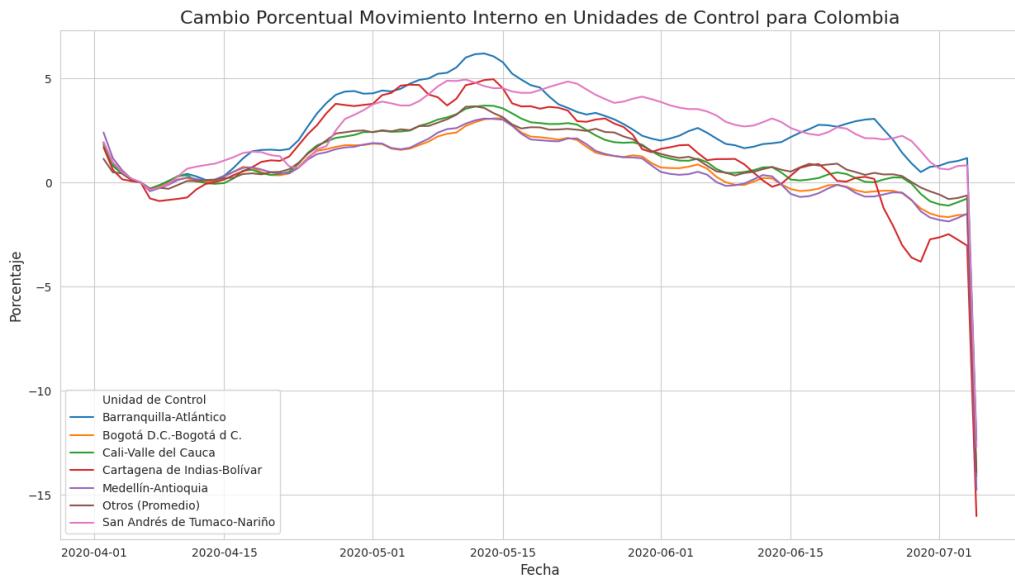


Figura 5: Movimiento interno por unidad espacial.

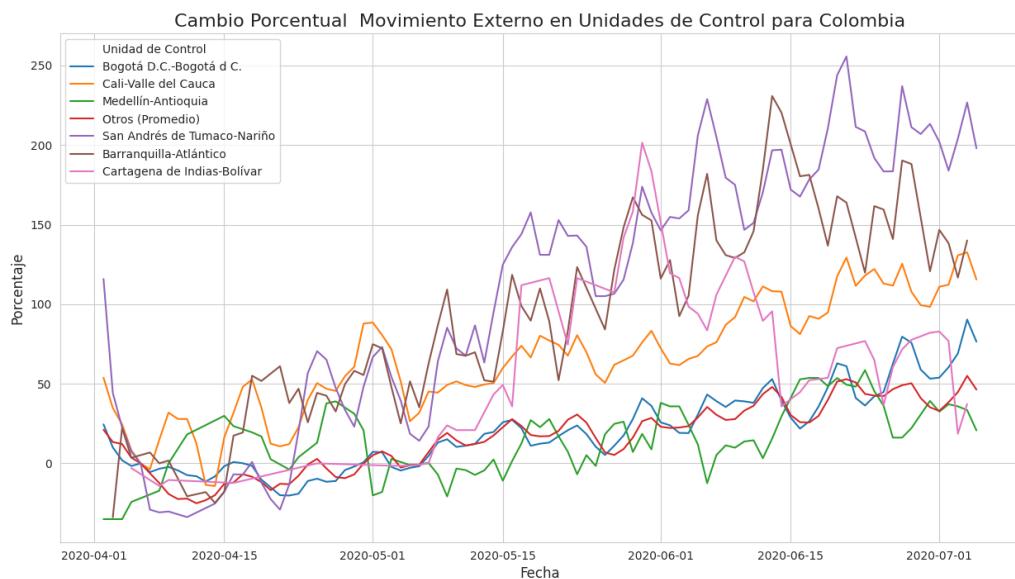


Figura 6: Movimiento externo por unidad espacial.

7.4. Casos por Unidad Espacial

Las siguientes figuras (7-8) muestran la progresión de casos para las unidades espaciales encontradas (ver sección 11.1).

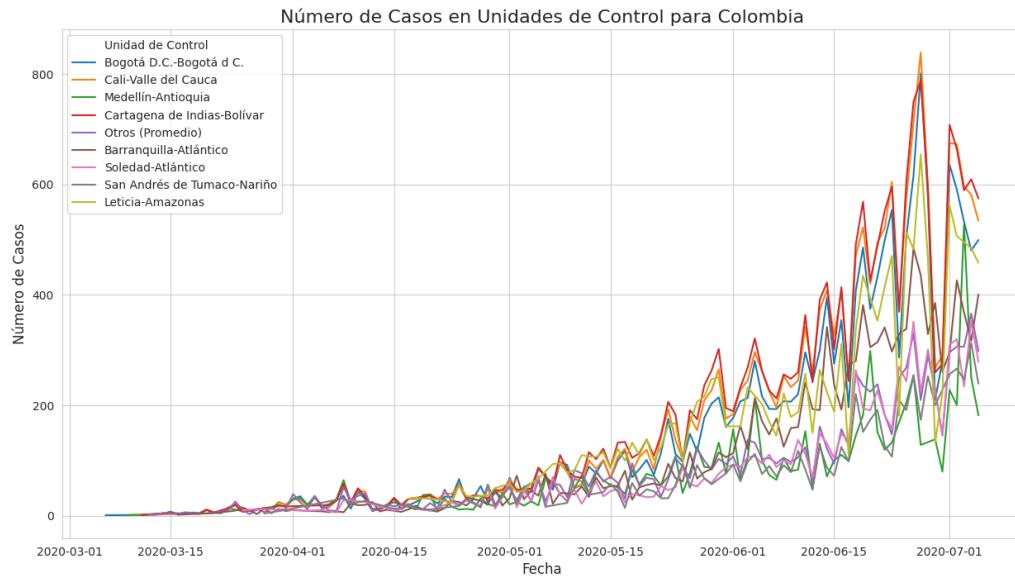


Figura 7: Número de casos diarios reportados en las unidades espaciales.

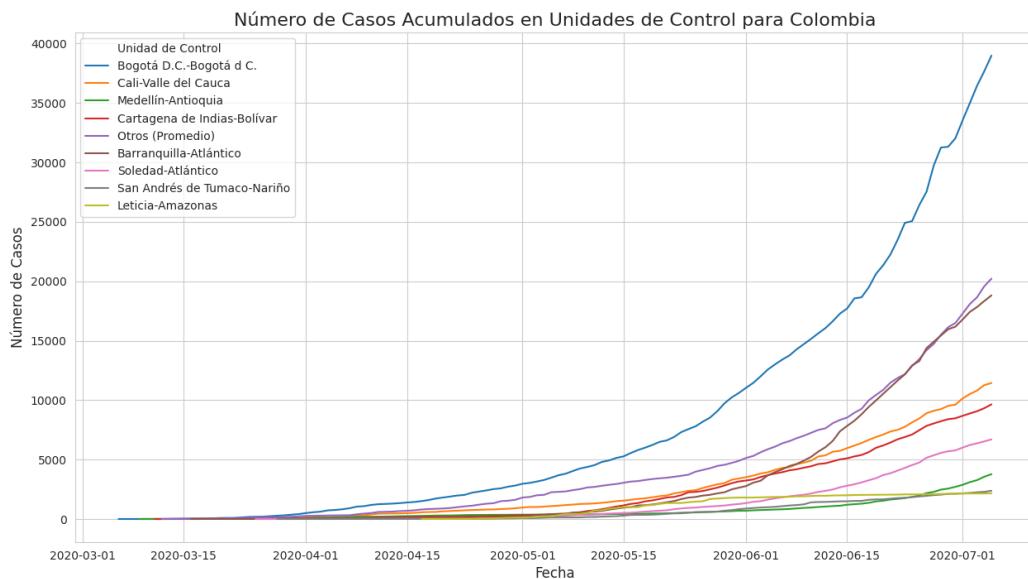


Figura 8: Número acumulado de casos reportados en las unidades espaciales.

7.5. Estado de Unidades Espaciales

Las siguientes figuras (9- 10) muestran el estado en distintas fechas para las distintas unidades espaciales (ver sección 11.1). Se puede ver el movimiento interno de la unidad (tamaño del círculo), movimiento entre distintas unidades (líneas amarillas) y la acumulación de casos (color rojo).

COVID-19 Dinámicas al Día: 27 (Casos Acumulados)

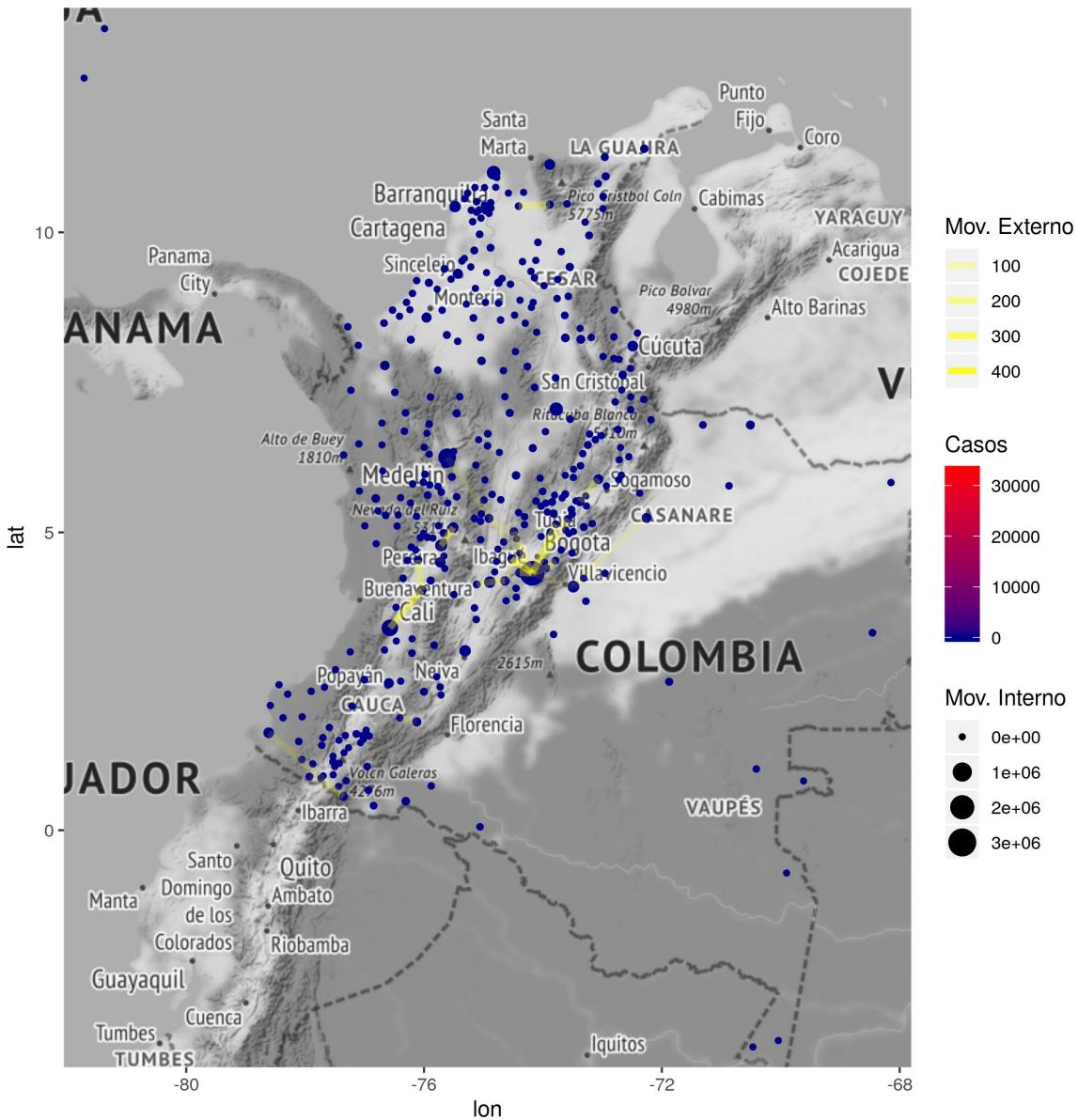


Figura 9: Mapa del estado de las unidades espaciales en el día 27 de la epidemia (abril 4).

COVID-19 Dinámicas al Día: 117 (Casos Acumulados)

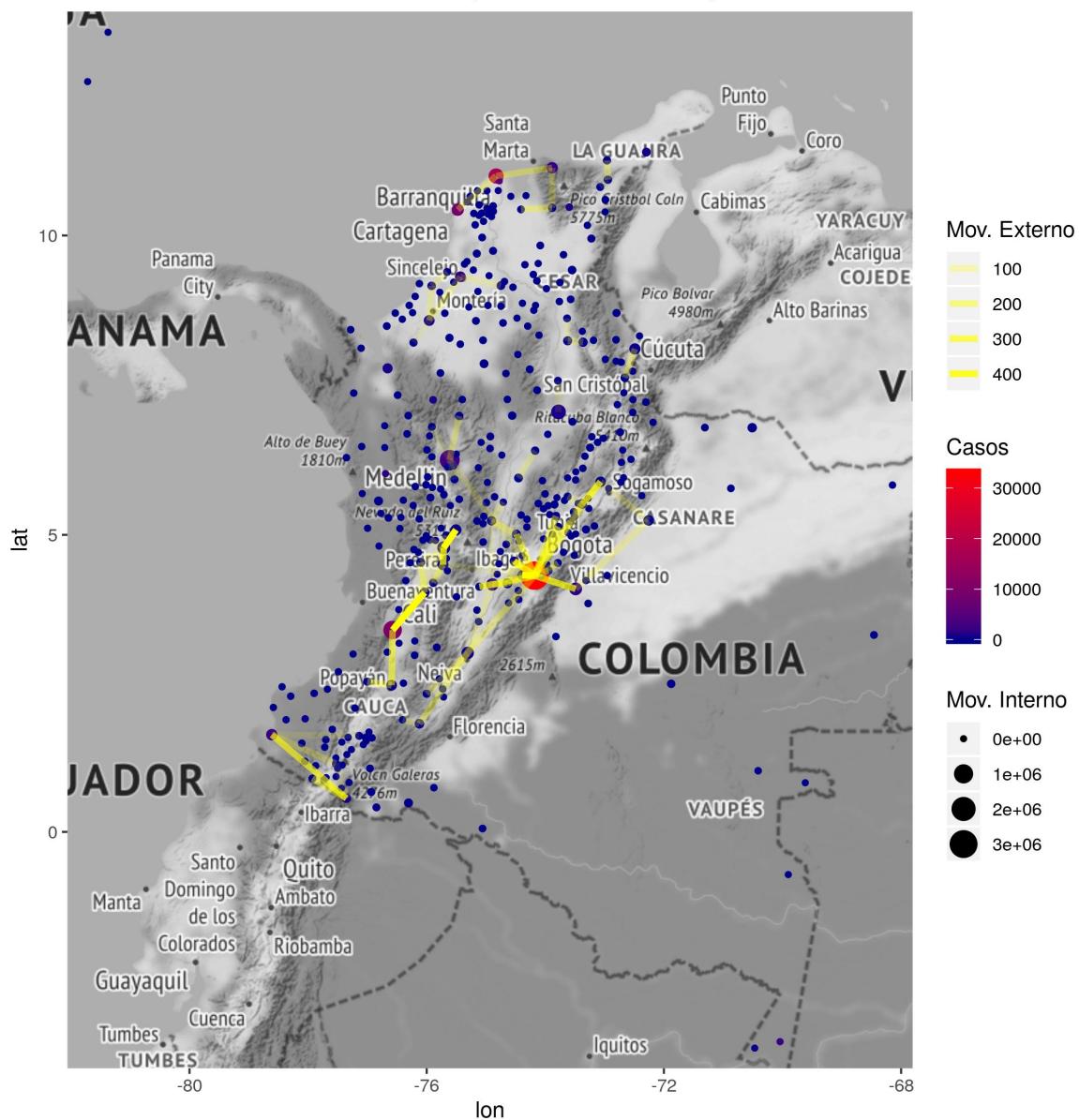


Figura 10: Mapa del estado de las unidades espaciales para el día más reciente.

7.6. Proyección a Futuro

Las siguientes figuras (11- 18) muestran las series de casos a través del tiempo mas similares y la proyección a futuro para las unidades espaciales con más casos del país. Los detalles de construcción de unidades espaciales se encuentran en la sección 11.1 y la explicación de los modelos de predicción se encuentran en la sección 12.1.

Este es un ejercicio que será actualizado constantemente, y no debe ser considerado una predicción epidemiológica sino una herramienta que busca generar alertas espacio-temporales con base en datos de movilidad en tiempo casi real.

7.6.1. Bogotá

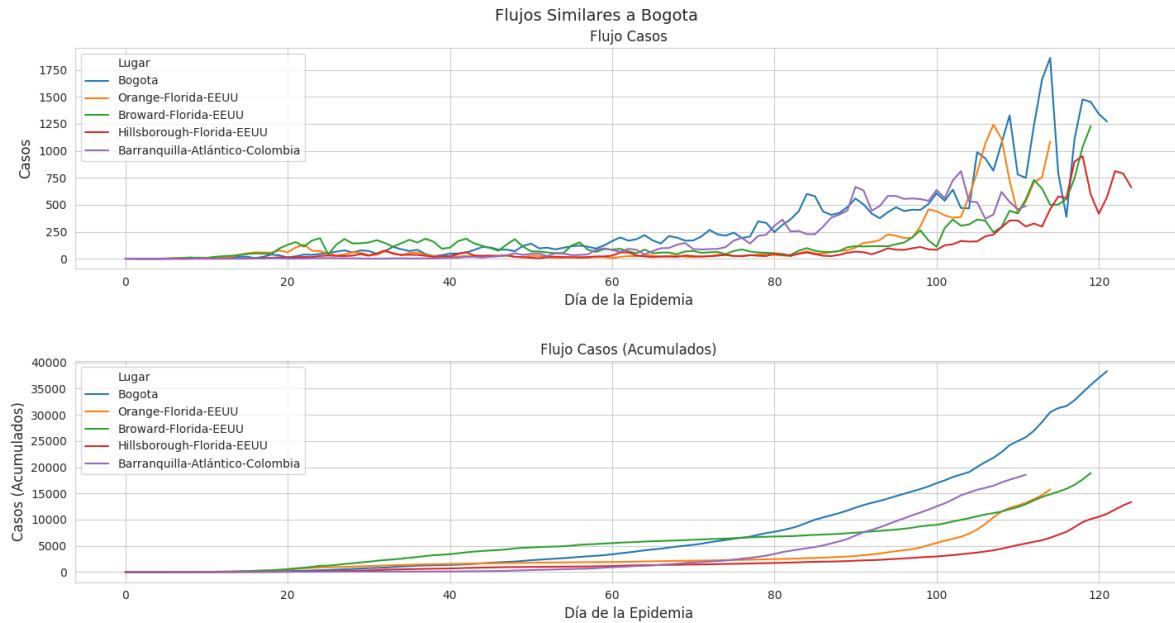


Figura 11: Ciudades más similares a Bogotá (según su curva epidemiológica).

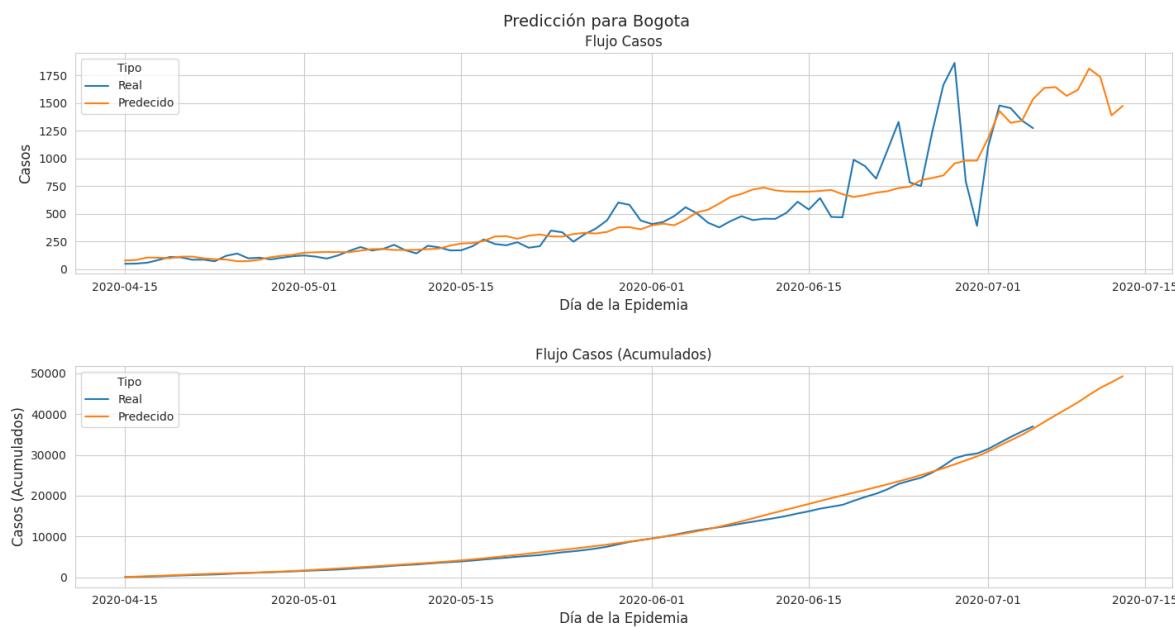


Figura 12: Proyección de la cantidad de casos diarios según el número de casos pasados, movimiento interno y movimiento externo histórico.

7.6.2. Cali

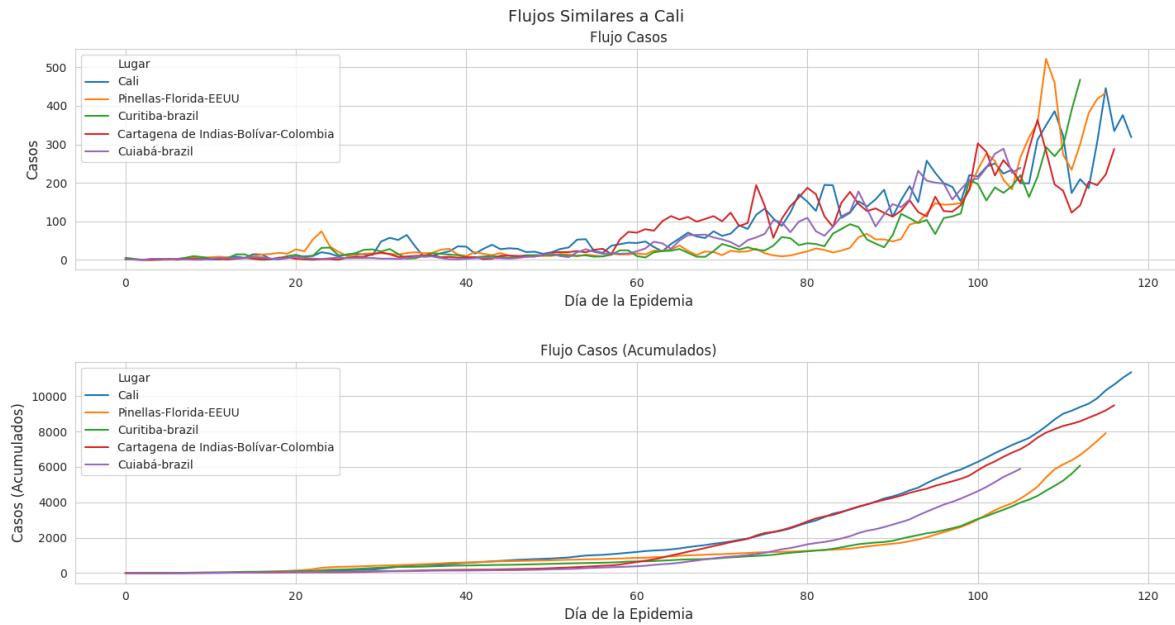


Figura 13: Ciudades más similares a Cali (según su curva epidemiológica).

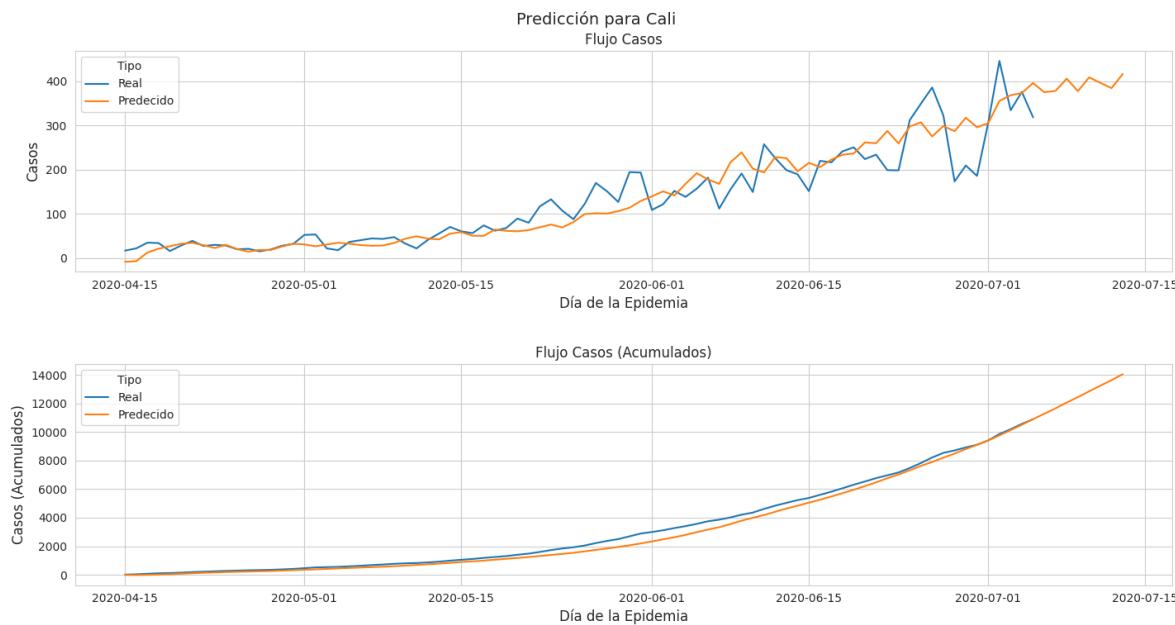


Figura 14: Proyección de la cantidad de casos diarios según el número de casos pasados, movimiento interno y movimiento externo histórico.

7.6.3. Barranquilla

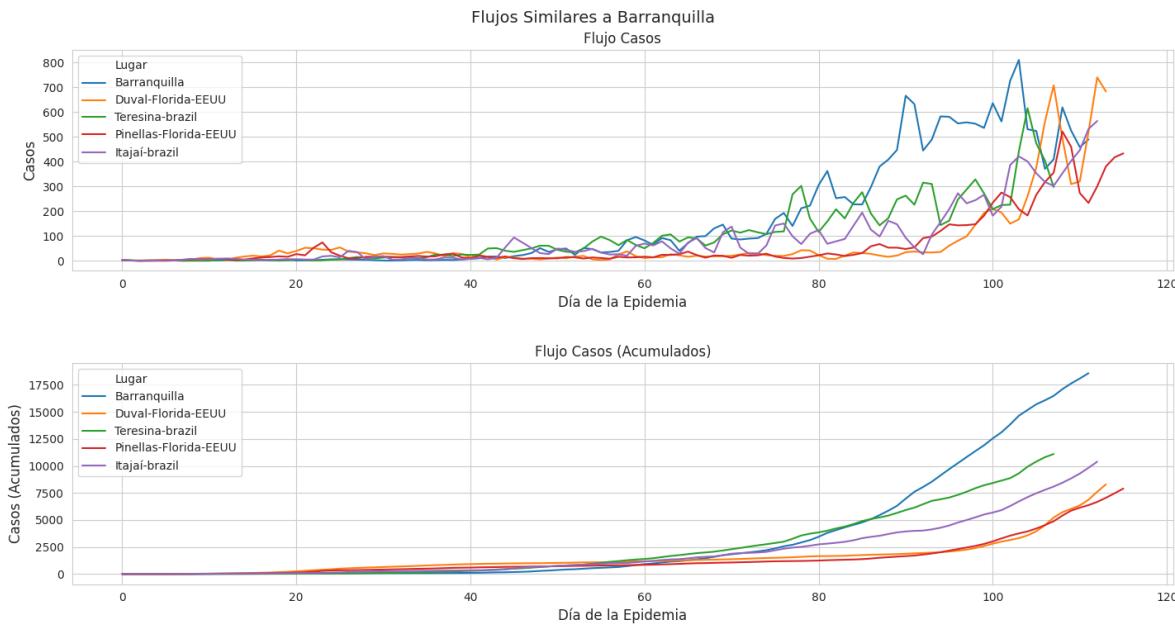


Figura 15: Ciudades más similares a Barranquilla (según su curva epidemiológica).

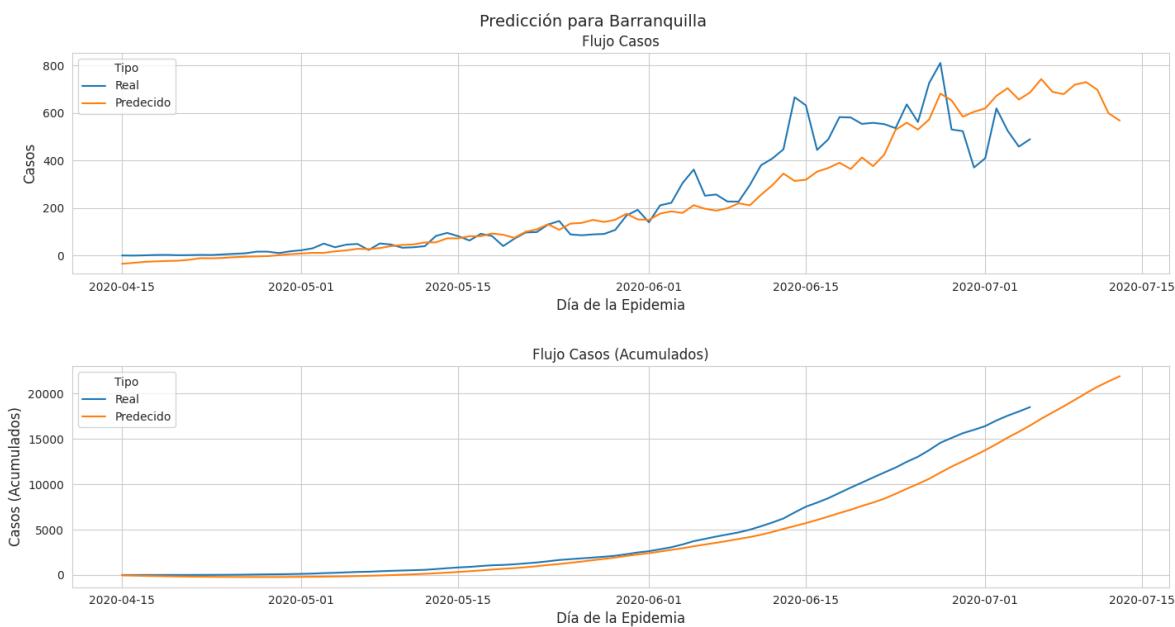


Figura 16: Proyección de la cantidad de casos diarios según el número de casos pasados, movimiento interno y movimiento externo histórico.

7.6.4. Cartagena

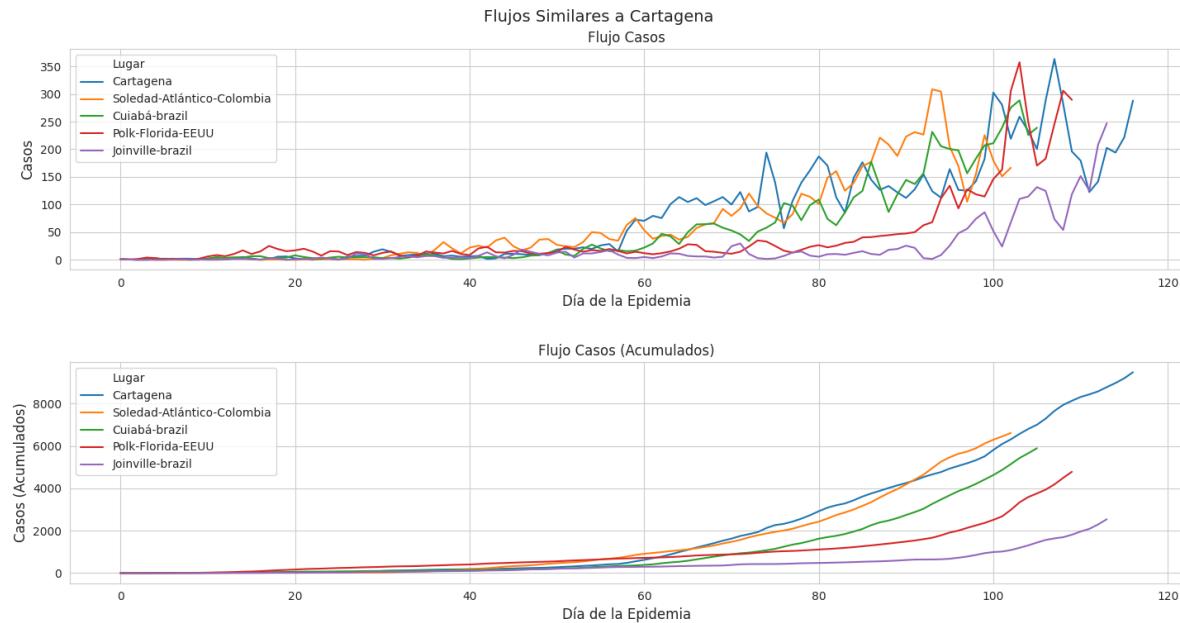


Figura 17: Ciudades más similares a Cartagena (según su curva epidemiológica).

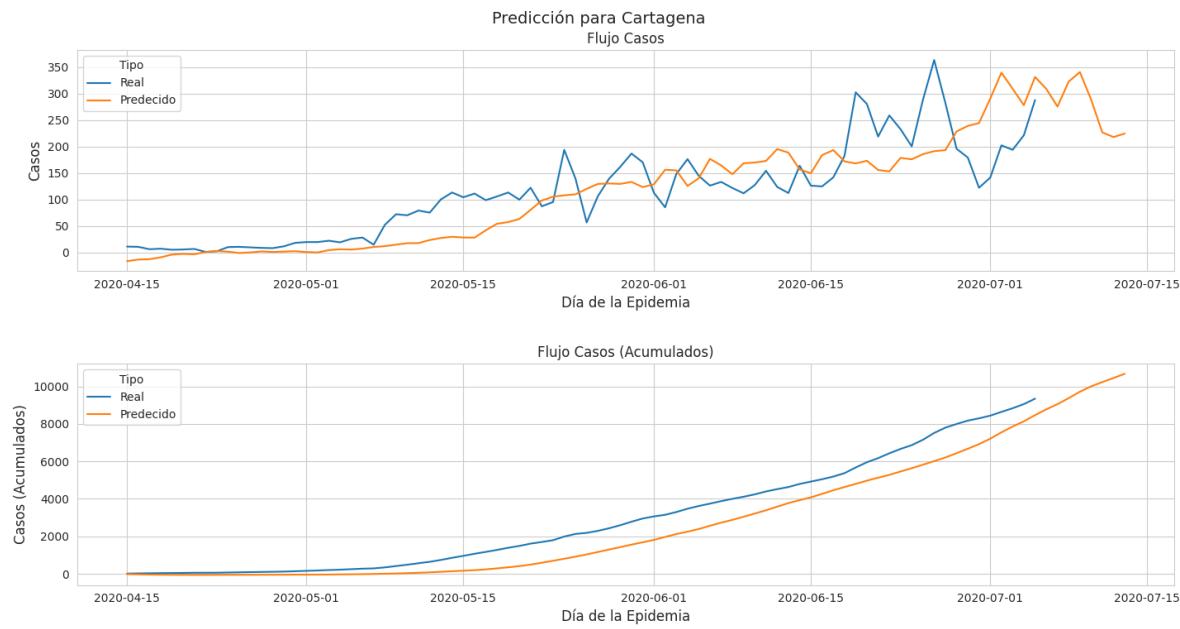


Figura 18: Proyección de la cantidad de casos diarios según el número de casos pasados, movimiento interno y movimiento externo histórico.

7.7. Análisis topológico de datos

Las siguientes figuras (19- 20) muestran los resultados del espacio de proyección de los grafos diarios y el 1-esqueleto obtenido al ejecutar el algoritmo *Mapper* (ver sección 12.2).

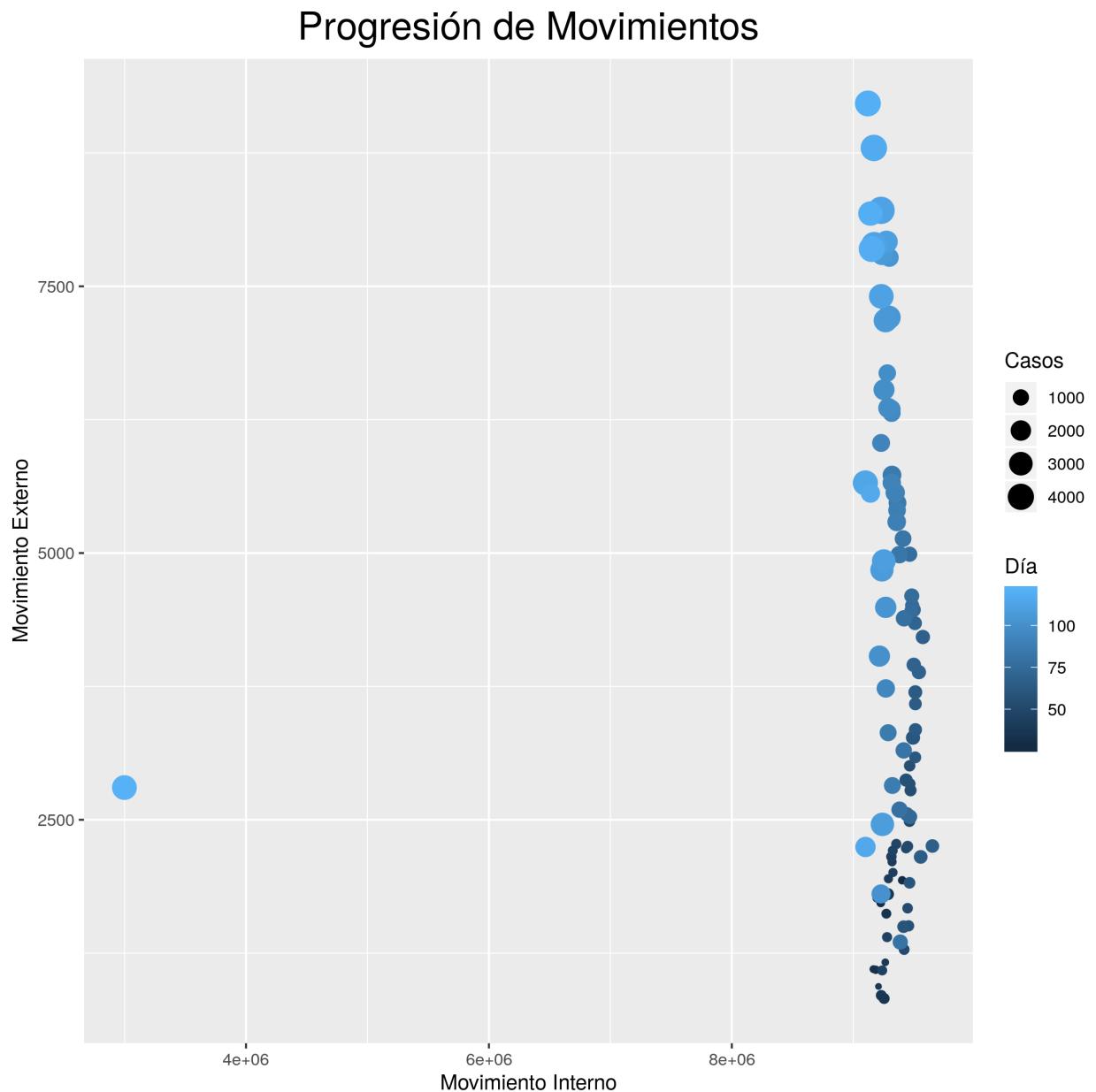


Figura 19: **Diagrama de dispersión de la movilidad interna/externa de las regiones.** Cada círculo representa un mapa como las figuras 9-10; el tamaño representa el número de casos reportados para ese día, el color representa el día. Los ejes representan la movilidad interior (eje-x) y exterior (eje-y). La primera muestra corresponde a abril 4.

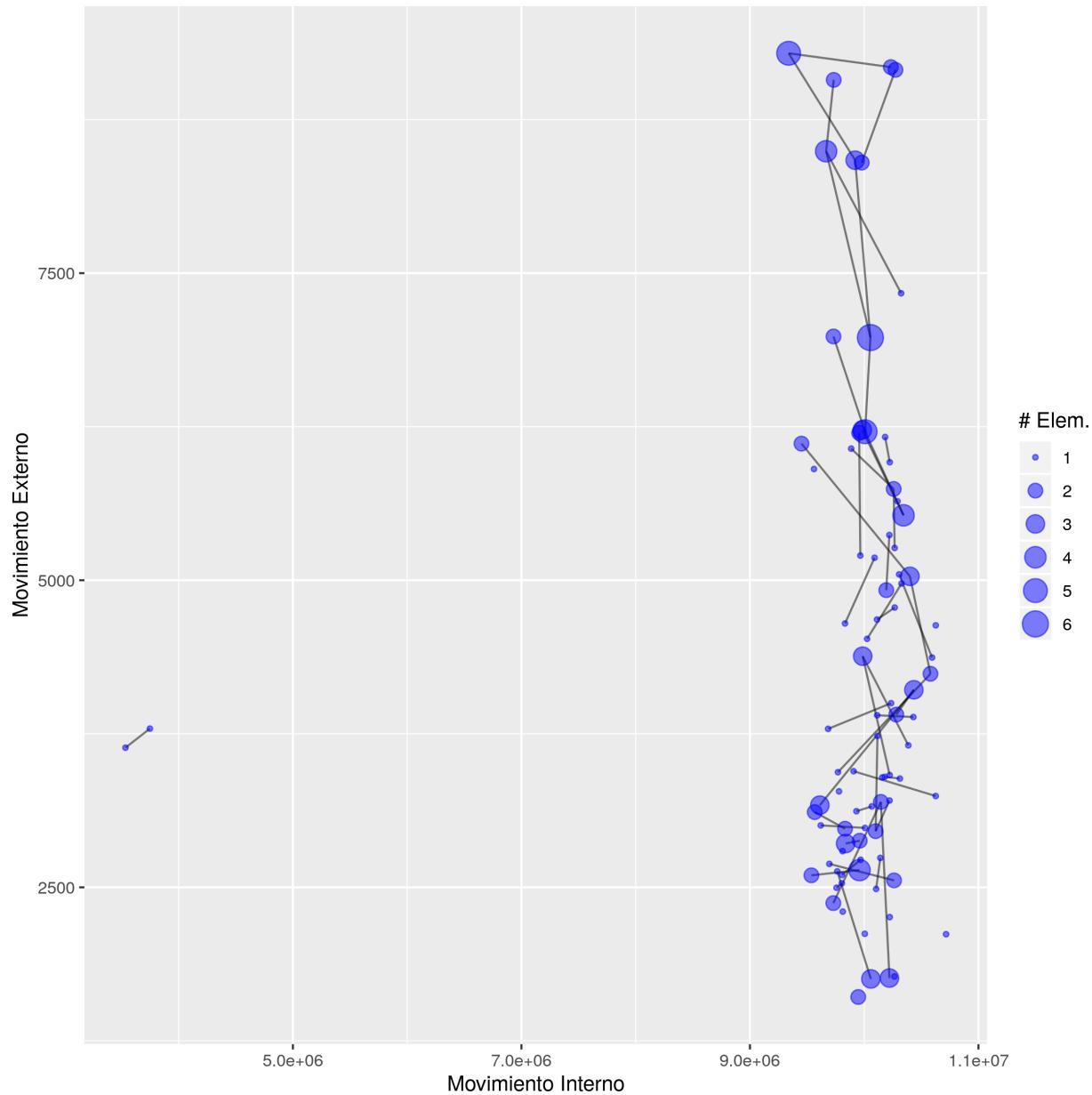


Figura 20: **Análisis Topológico de Datos (ATD)**. 1-esqueleto que representa agrupaciones (*clustering*) sobre estas dos dimensiones (movimiento interno y externo), con la distancia Wasserstein entre distribuciones espaciales de casos. Podemos observar que, a pesar de que la epidemia está en sus primeras etapas en Colombia, se aprecian ya grupos distintos de perfiles de movilidad y casos confirmados. Más interesante todavía, estos grupos pueden tener valores similares de movimiento interno y externo, pero la distribución espacial de casos es diferente y por eso se identifican diferentes grupos con valores similares. Esto quiere decir que perfiles de movilidad similares pueden conllevar distribuciones espaciales de casos diferentes.

8. Discusión (actualizado semanalmente)

Los datos muestran que el país, en general, aplicó medidas de distanciamiento y aislamiento social preventivo desde la fecha en que fue decretado el simulacro, el 21 de marzo. Si bien es cierto que el simulacro fue liderado por el gobierno local de Bogotá, el efecto en la movilidad se puede observar en las principales ciudades de Colombia.

Se observa un pico de movilidad a partir del 13 de abril, fecha en la cual la movilidad aumenta con respecto a lo observado los días anteriores. Si bien el movimiento a partir de esa fecha todavía está por debajo del promedio acumulado para cada lugar, en ciudades como Cúcuta se ha limitado significativamente el efecto, que ya está por debajo del 40 % con respecto a los valores de las semanas anteriores a la entrada del virus.

Es interesante ver cómo la movilidad se redujo de manera muy heterogénea en diferentes regiones del país. Si bien al hacer el análisis por centros poblados el resultado es bastante similar (todos redujeron de manera comparable la movilidad desde que empezó la cuarentena), el panorama a nivel regional es bastante diferente. Como es de esperarse, regiones compuestas por las principales vías terrestres del país han tenido una mayor dificultad para reducir su movilidad (probablemente dado su carácter de flujos de abastecimiento y movimiento entre regiones). Es importante anotar que estas regiones probablemente cuentan con servicios de salud mucho más precarios que otros lugares del país, con lo cual están en alto grado de vulnerabilidad.

El análisis de ATD detecta que si bien el país ha reducido su movilidad de manera similar en todo el territorio, las distribuciones de casos observadas son clasificadas en *clusters* diferentes. Esto significa que la respuesta de la epidemia al movimiento humano es heterogénea, y probablemente no impacte de la misma manera los casos observados en diferentes momentos en el tiempo (o diferentes distribuciones de casos).

Desde aproximadamente el 22 de abril, la movilidad en toda Colombia aumentó. La cuarentena lentamente ha ido perdiendo efectividad como medida de aislamiento social desde la perspectiva de los datos de movilidad.

9. Limitaciones

Este informe tiene como objetivo ser parte de un ecosistema de diagnósticos de la situación. El análisis presentado debe ser comparado con los resultados de diferentes métodos, el objetivo no es remplazarlos. La precisión de los datos de reportes diarios, por el cuello de botella que supone el diagnóstico, hace que los resultados estén sujetos al sesgo que muestran los datos. Debido a los datos a los que se tiene acceso, la geolocalización se hace de forma automática, sacrificando precisión por eficiencia. En Colombia, la epidemia apenas está en sus primeras etapas, hay pocos datos y las series de tiempo son relativamente cortas. A medida que aumente la información o se extienda a otros países, la precisión del análisis debe aumentar. En la medida en que ocurran diferentes cambios en la epidemiología, también el análisis se volverá más robusto. Es de esperar que tenga errores, pero es una herramienta más que puede brindar información en tiempo real.

Los datos de cambio de movilidad entre y al interior de unidades espaciales, aparte de ser un indicador de la eficacia de las medidas de distanciamiento social, permiten establecer cual es su efecto en la progresión de la infección y su posible carga para el sistema de salud a nivel regional si están acoplados a datos epidemiológicos mínimos disponibles y confiables (e.g. número de casos confirmados, número de individuos expuestos por unidad espacial, número de fallecidos por causa de la infección, número de camas ocupadas de los servicios de hospitalización general, unidades de cuidado intermedio (UCIM) y unidades de cuidado intensivo (UCI) en Instituciones Prestadoras de Servicios de Salud – IPS de naturaleza privada y pública frente al total disponible y habilitado). Por ejemplo, los datos de casos confirmados dependen del número de pruebas diarias realizadas por unidad espacial, del momento en que se realiza la prueba y del tiempo que tardan los laboratorios de diagnóstico en proporcionar el resultado. Estos datos no necesariamente están disponibles.

10. Definiciones

Reducción significativa en movilidad: Valor-z debajo de 2 desviaciones estándar del observado para las semanas anteriores (similar a un canal endémico).

Megaregión, comunidad, unidad funcional: La modularidad es definida como una función del total de conexiones del subconjunto de una red entre nodos de la misma comunidad (y no conexiones con nodos externos a una comunidad)[1]. Dado que no puede ser evaluada como la simple fracción de aristas (pues derivaría en una comunidad compuesta por todos los nodos), se evalúa en comparación con redes aleatorias[1].

11. Métodos

11.1. Construcción de Unidades Espaciales de SARS-COV-2 en Colombia

Con base en el algoritmo de clustering *walktrap*[5] (cuyo objetivo es identificar grupos de nodos cuyas conexiones entre sí sean mucho mayores que las conexiones con otros nodos), implementado con los datos de movilidad donde los municipios son los nodos y las aristas tienen como peso el número de viajes que se hacen entre estos, se identifican los grupos o unidades espaciales. Estas unidades son de vital importancia en el contexto de cuarentena flexible, pues son las unidades desde las cuales hay más riesgo de contagio entre sí.

Estas unidades se pueden pensar como las unidades regionales funcionales, de manera que los esfuerzos de control deben obedecer a estas dinámicas y no a las divisiones administrativas, pues en ocasiones municipios diferentes pueden hacer parte del mismo "grupo metropolitano", como Bogotá y Soacha, por mencionar solo un ejemplo.

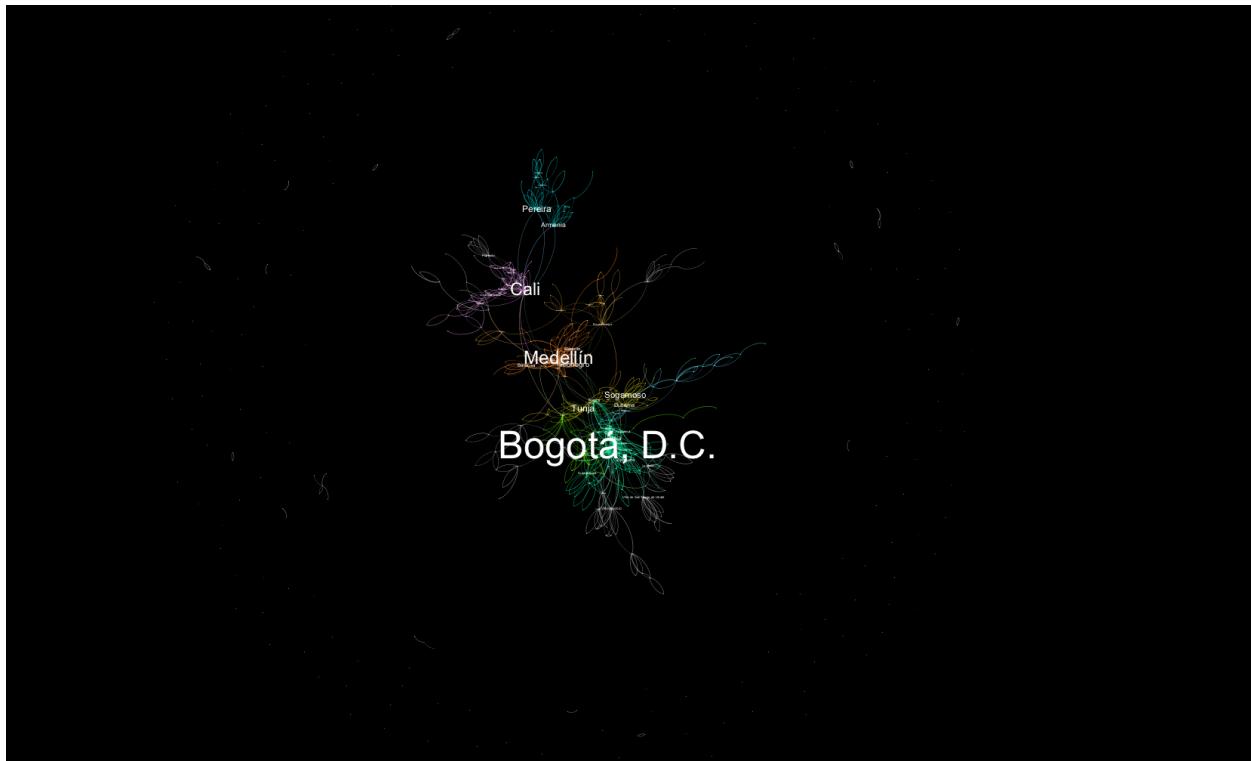


Figura 21: **Unidades espaciales de SARS-COV-2 en Colombia.** Con base en los datos, se definen las regiones de control, identificadas con diferentes colores en esta gráfica. Las unidades espaciales están pensadas de tal manera que son las comunidades regionales observadas.

Unidades de Control Principales

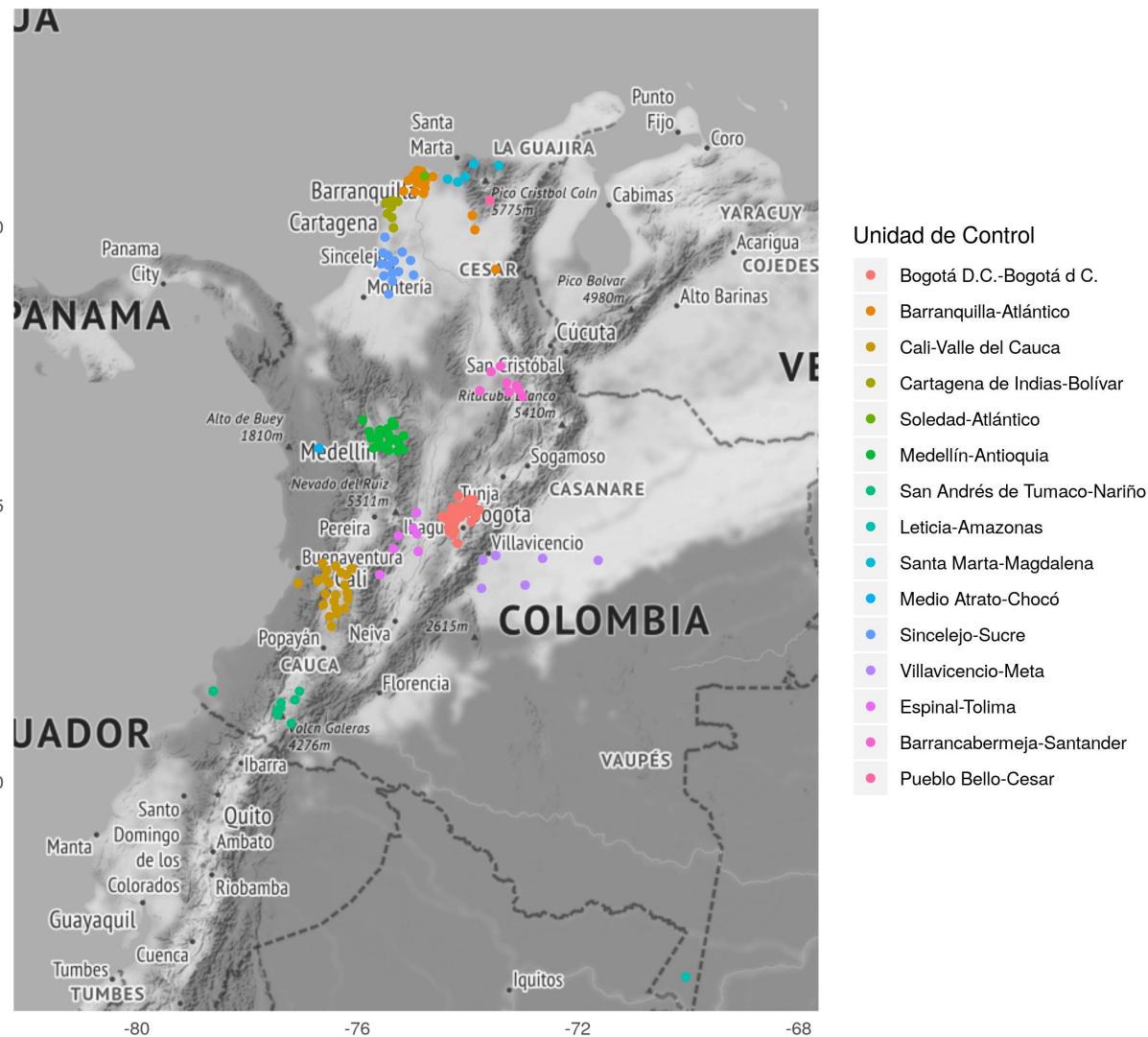


Figura 22: **Unidades espaciales de SARS-COV-2 en Colombia con más casos.** Representación geográfica.

12. Extracción de Datos de Movilidad y Casos

Uno de los principales beneficios de utilizar los datos de movilidad de Facebook GeoInsights, es que se encuentran disponibles para distintos lugares alrededor del mundo, muchos de los cuales también ofrecen repositorios públicos con datos estructurados de los casos confirmados en el lugar. Por lo tanto, los resultados expuestos anteriormente se pueden hacer para cualquier unidad espacial de Colombia y potencialmente para cualquier otro país del cual se tengan datos de casos confirmados.

Esto permite construir un conjunto estructurado de datos de movilidad y casos geolocalizados para distintos lugares del mundo. Actualmente se cuenta con los siguientes lugares integrados dentro del esquema de extracción de datos.

País	Lugar	Movimiento	Casos	Fecha Inicio Movimientos	# Unidades de Control
Colombia	País Entero	Si	Si	2020-04-02	246
Colombia	Bogotá	Si	Si	2020-04-12	494
USA	Indiana	Si	Si	2020-04-04	91
USA	Illinois	Si	Si	2020-04-05	94
USA	Kentucky	Si	Si	2020-04-02	96
USA	Florida	Si	Si	2020-04-09	64
USA	Vermont	Si	Si	2020-04-06	14
USA	Oregon	Si	Si	2020-04-28	34
Italia	País Entero	Si	Si	2020-03-23	29
Perú	País Entero	Si	Si	2020-03-25	160
Australia	New South Wales	Si	Si	2020-05-03	56
Alemania	País Entero	Si	Si	2020-03-25	223
Brasil	País Entero	Si	Si	2020-03-31	2318
Chile	País Entero	Si	Si	2020-03-23	100

Cuadro 1: Distintos lugares integrados en el esquema de análisis.

La mayoría de estos repositorios ya tienen los casos geocodificados, pero existen algunos que únicamente tienen asociado a cada caso el lugar específico (municipio, región, ciudad) en el que ocurrió, como sucede para Colombia, Alemania y Australia. Para estos lugares se hace una geocodificación automática, asignando las coordenadas geográficas del lugar que permitirá cruzar esta información con la información de movilidad de Facebook GeoInsights

La extracción de estos conjuntos y construcción de los datos unificados se hace a diario, descargando automáticamente los datos de movilidad de Facebook, los reportes de casos confirmados de los repositorios públicos de cada lugar y geocodificando los sitios de ocurrencia para los lugares que no se encuentran precomputados. A medida que se encuentran disponibles más lugares, estos se integrarán en el esquema de extracción.

12.1. Modelos de Proyección

Estos modelos permiten proyectar los casos de diferentes unidades hasta una semana en el futuro, utilizando el movimiento interno, movimiento externo de unidades de control adyacentes y el numero de casos confirmados por unidad. A continuación se presentan los detalles en la construcción de los datos y el esquema de entrenamiento para una unidad espacial.

12.1.1. Estructura de los Datos

A partir de los datos de movilidad de Facebook GeoInsights y los reportes de casos históricos, se construye el siguiente conjunto para cada día d en la unidad de control i (una unidad de control cualquiera):

- **Número de Casos Interno** (C_i^d): el número de casos confirmados en el día.
- **Movimiento Interno** (M_{ii}^d): la suma de todas las observaciones de movimiento que comienzan y terminan en dentro de la unidad espacial.
- **Número de Casos Externo** (CE_i^d): el número de casos confirmados en el día para el resto de unidades espaciales.

$$CE_i^d = \sum_{j \neq i} C_j^d$$

- **Movimiento Externo** (ME_i^d): el movimiento del resto de unidades espaciales del país hacia la unidad en cuestión.

$$ME_i^d = \sum_{j \neq i} M_{ij}^d$$

- **Día de la Epidemia** (δ_i^d) el número actual de la epidemia, tomando como el día 0 la fecha en que se presentó el primer caso en la unidad espacial analizada. Por ejemplo, si se tiene una unidad espacial A y B con primer caso reportado: 27 de Marzo y el 2 de Abril respectivamente, si d es el día 3 de Abril entonces: $\delta_A^d = 7$ y $\delta_B^d = 1$.

A partir de este conjunto de datos diarios, se construyen dos valores adicionales para el modelo de proyección. Estos valores también se encuentran asociados a la unidad espacial en cuestión para cada uno de los días de la epidemia:

- **Flujo Interno (FI_i^d):** La multiplicación de la cantidad de casos internos en el día por el movimiento interno de la unidad. Esta variable pretende capturar la “fuerza” interna de la epidemia en la unidad de control para un día en particular.

$$FI_i^d = C_i^d M_{ii}^d$$

- **Flujo Externo (FE_i^d):** La suma de la multiplicación de la cantidad de casos y el movimiento hacia la unidad de control en cuestión, del resto de unidades del país. Esta variable pretende capturar la “fuerza” entrante de la epidemia hacia la unidad de control para un día en particular.

$$FE_i^d = \sum_{j \neq i} C_j^d M_{ji}^d$$

El modelo implementado utilizará estas variables para los anteriores 5 días, para encontrar el número de casos 8 días en el futuro.

$$X_i^d = \{C_i^d, M_{ii}^d, CE_i^d, ME_i^d, FI_i^d, FE_i^d, \delta_i^d\}$$

$$(X_i^d, X_i^{d-1}, \dots, X_i^{d-4}) \mapsto C_i^{d+8}$$

Los valores de 5 y 8 fueron ajustados según los datos.

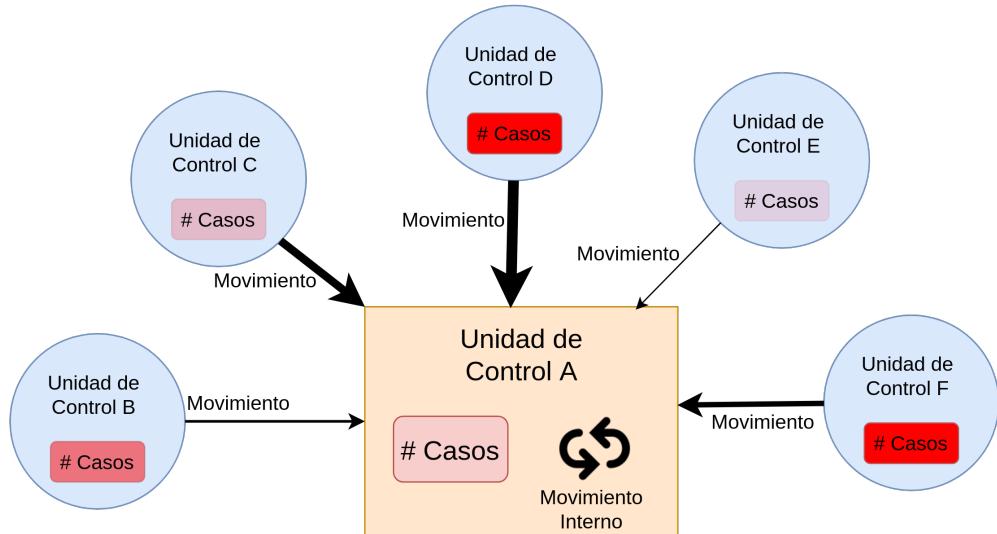


Figura 23: Funcionamiento modelo de proyección. A partir del movimiento y los casos para las distintas unidades espaciales alrededor, se construyen los datos de entrada al modelo para una unidad en particular.

12.1.2. Entrenamiento

A partir del conjunto de unidades de control para los distintos lugares de los que se tiene información (ver tabla 1), se seleccionan las 20 unidades con curvas epidémicas más similares a la unidad espacial que se quiere predecir. La similitud entre curvas epidémicas se hace utilizando una técnica llamada Dynamic Time Warping [6], únicamente comparando las partes de las curvas que coinciden en los días de la epidemia, esto quiere decir que lugares donde la epidemia está menos avanzada que la de la unidad serán descartadas y las curvas de los lugares donde la epidemia está más avanzada serán recortadas.

El conjunto de series más similares a la unidad de control se actualiza cada vez que se ejecuta la predicción, permitiendo adaptarse en parte a los cambios que presenten.

Los datos construidos para la unidad espacial serán unidos con los de sus unidades más similares para tener el conjunto final de datos para la proyección.

12.1.3. Modelo

El modelo de predicción utiliza una regresión lineal de mínimos cuadrados con regularización L2. Se usa un esquema de validación cruzada para encontrar el mejor parámetro del peso de la regularización.

12.2. Análisis Topológico de Datos (ATD)

El análisis topológico de datos, es un conjunto de técnicas y modelos que permiten inferir información sobre conjuntos de datos utilizando su estructura subyacente. Para este reporte, se utilizó nuestra propia implementación del algoritmo *Mapper*^[7], disponible en etapa beta en [github](#). Una de las virtudes de este método es que captura los fenómenos locales y globales mejor que métodos similares, como algoritmos que analizan grupos o que reducen las dimensiones de los datos (puede interpretarse como una combinación de ambos). Ejemplos exitosos del uso de *Mapper* incluyen la identificación de subgrupos de pacientes que padecen enfermedades crónicas y transmisibles, grupos de patógenos [8, 9, 10, 11, 12, 13, 14, 15], patrones de votación [8], configuraciones de equipos deportivos [8], entre otros. Para obtener más detalles sobre los métodos, consulte [8, 16].

Puntualmente, los datos analizados son grafos que se pueden generar a partir de las unidades espaciales y su estado en cada día de la epidemia. El esquema es el siguiente:

- Para un día en particular, se filtran únicamente los casos confirmados y el movimiento en esa fecha.
- A cada unidad espacial se le asocian sus casos correspondientes creando un nodo en el grafo.
- Los nodos son unidos por aristas con el peso correspondiente al movimiento entre cada par de nodos. Adicionalmente, cada nodo se une a sí mismo con el valor de su movimiento interno para la fecha.

Por lo tanto, se obtiene un grafo no dirigido pesado para cada día de la epidemia.

Para la distancia entre cada uno de estos grafos se calculó la métrica de Wasserstein sobre la distribución de casos a nivel nacional, donde la distancia entre cada unidad de control corresponde a la distancia espacial desde su punto central. Por último, se proyecta cada grafo al espacio bidimensional de la suma de su movimientos internos y la suma de sus movimientos externos.

Utilizando *Mapper*, se identifican agrupaciones (clusters) de distribuciones espaciales de casos confirmados en diferentes intervalos de movilidad. Dado que el objetivo es caracterizar la relación entre movimiento interno-externo y casos confirmados, este método identifica distribuciones de casos espaciales similares por intervalos de movilidad. La intuición de este análisis es que se identifican perfiles de casos a nivel espacial (diarios y acumulados), y se asocian con perfiles de movilidad. Eso nos da una idea de perfiles de movimiento observados y sus respectivos perfiles de casos confirmados, lo cual es el objeto de este informe. Por ejemplo, se puede observar si patrones de movilidad relativamente similares corresponden a distribuciones de casos similares, o si por el contrario, se observan patrones de casos confirmados diferentes y variados con perfiles de movilidad relativamente similares.

Referencias

- [1] Garrett Dash Nelson and Alasdair Rae. An economic geography of the united states: From commutes to megaregions. *PloS one*, 11(11), 2016.
- [2] Stephen Wheeler. Regions, megaregions, and sustainability. In *The Futures of the City Region*, pages 123–136. Routledge, 2013.
- [3] Adjo A Amekudzi, Tridib Banerjee, Jason Barringer, Scott Campbell, Cheryl K Contant, Jessica L Harbour Doyle, William Ankner, Norman Fainstein, Susan S Fainstein, Andreas KF Faludi, et al. *Megaregions: Planning for global competitiveness*. Island Press, 2012.
- [4] Margaret Dewar and David Epstein. Planning for “megaregions” in the united states. *Journal of Planning Literature*, 22(2):108–124, 2007.
- [5] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In *International symposium on computer and information sciences*, pages 284–293. Springer, 2005.
- [6] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

- [7] Gurjeet Singh, Facundo Mémoli, and Gunnar E Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *SPBG*, pages 91–100, 2007.
- [8] Pek Y Lum, Gurjeet Singh, Alan Lehman, Tigran Ishkanov, Mikael Vejdemo-Johansson, Muthu Alagappan, John Carlsson, and Gunnar Carlsson. Extracting insights from the shape of complex data using topology. *Scientific reports*, 3:1236, 2013.
- [9] Monica Nicolau, Arnold J Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.
- [10] Li Li, Wei-Yi Cheng, Benjamin S Glicksberg, Omri Gottesman, Ronald Tamler, Rong Chen, Erwin P Bottinger, and Joel T Dudley. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science translational medicine*, 7(311):311ra174–311ra174, 2015.
- [11] Timothy SC Hinks, Xiaoying Zhou, Karl J Staples, Borislav D Dimitrov, Alexander Manta, Tanya Petrossian, Pek Y Lum, Caroline G Smith, Jon A Ward, Peter H Howarth, et al. Innate and adaptive t cells in asthmatic patients: relationship to severity and disease mechanisms. *Journal of Allergy and Clinical Immunology*, 136(2):323–333, 2015.
- [12] Timothy SC Hinks, Tom Brown, Laurie CK Lau, Hitasha Rupani, Clair Barber, Scott Elliott, Jon A Ward, Junya Ono, Shoichiro Ohta, Kenji Izuhara, et al. Multidimensional endotyping in patients with severe asthma reveals inflammatory heterogeneity in matrix metalloproteinases and chitinase 3-like protein 1. *Journal of Allergy and Clinical Immunology*, 138(1):61–75, 2016.
- [13] Brenda Y Torres, Jose Henrique M Oliveira, Ann Thomas Tate, Poonam Rath, Katherine Cumnock, and David S Schneider. Tracking resilience to infections by mapping disease space. *PLoS biology*, 14(4):e1002436, 2016.
- [14] Alejandro Feged-Rivadeneira, Andrés Ángel, Felipe González-Casabianca, and Camilo Rivera. Malaria intensity in colombia by regions and populations. *PloS one*, 13(9), 2018.
- [15] A Knudson, F González-Casabianca, A Feged-Rivadeneira, MF Pedreros, S Aponte, A Olaya, CF Castillo, E Mancilla, A Piamba-Dorado, R Sanchez-Pedraza, et al. Spatio-temporal dynamics of plasmodium falciparum transmission within a spatial unit on the colombian pacific coast. *Scientific reports*, 10(1):3756, 2020.
- [16] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

13. Agradecimientos

Recibimos comentarios y valiosa retroalimentación de Mauricio Santos-Vega (Universidad de Los Andes), Jaime Urrego (Universidad del Rosario) Diego Echeverri (Universidad Icesi-Univialle), Pablo Cárdenas (MIT), Angélica Knudson Ospina (Universidad Nacional) y Manuela Carrasquilla (Harvard University).