
POLÍTICAS PÚBLICAS, GRANDES DATOS, TEORÍA DE REDES Y COVID-19

Emiliano Isaza Villamizar

Servinformación

emiliano.isaza@servinformacion.com

Felipe González-Casabianca

Universidad de Los Andes

f.gonzalez1899@uniandes.edu.co

Santiago Herrera

Universidad de Los Andes

s.herrera@uniandes.edu.co

Tomás Rodríguez-Barraquer

Universidad de Los Andes

t.rodriguezb@uniandes.edu.co

Andrés Ángel

Universidad de Los Andes

ja.angel1908@uniandes.edu.co

Vladimir Corredor

Universidad Nacional

vcorredore@unal.edu.co

Alejandro Feged-Rivadeneira

Universidad del Rosario

alejandro.feged@urosario.edu.co

3 de julio de 2020

1. Introducción

La crisis mundial desatada por un coronavirus de murciélagos, rastreada hasta un mercado de comida marina en Wuhan, China, ha puesto de cabeza el mundo de las políticas públicas [Zhu et al., 2020, Zhou et al., 2020]. Todos los países, más allá del espectro político y la aprobación del gobierno, de su momento económico, y de cualquier otra caracterización posible, han tenido que atender una emergencia humanitaria en medio de una crisis global cuyo par más cercano ocurrió hace más de 100 años. La economía global, una tecnología intangible y planetaria, en jaque por un agente infeccioso tan pequeño que no es considerado vivo en un sentido estricto.

La respuesta de países alrededor del mundo ha sido tan variada como los sistemas de gobierno, lenguas y credos en los cuales el virus encuentra terreno fértil. Al igual que en Egipto durante la primera guerra mundial, cuando los mosquitos y la malaria brotaron en medio de profundas transformaciones a los sistemas de producción y gobierno [Mitchell, 2002], la pandemia de Covid-19 conlleva una parafernalia de transformaciones sobre las cuales, ante la emergencia, se ha tenido poco tiempo para reflexionar. En apenas unos meses, los gobiernos han improvisado (o se han rehusado a hacerlo) las mejores respuestas que sus funcionarios de salud pública y grupos de expertos han podido diseñar. Comparado con el ritmo paquidérmico en el que operan las políticas públicas en términos de innovación, estas transformaciones se han operado con celeridad. Los países más preparados, han sido sin duda, aquellos afectados por la última amenaza de pandemia, también por coronavirus, en 2003. En los albores de la epidemia, Japón, Corea del Sur, Singapur, Taiwán, Hong Kong (los países con más casos de SARS), han sido ejemplo para controlar la dispersión del SARS-CoV-2. La tecnología ha cumplido un papel fundamental para el exitoso control de la enfermedad, y quizás otro tanto hayan influido factores como el acceso a salud, la adopción y acatamiento de medidas preventivas por parte de la población civil, y la desigualdad [Parodi and Liu, 2020, Lee et al., 2020, Wang et al., 2020, Ting et al., 2020, Wang and Tang, 2020, Chin et al., 2020, Ainslie et al., 2020, Kraemer et al., 2020, Lai et al., 2020]. Países como Estados Unidos, Rusia, Brasil, y Reino Unido, encabezan la segunda oleada de casos confirmados, después

de que brotes vistos Italia y España, o China en sus inicios, parecieran un límite superior. En el momento de redactar este manuscrito, Perú supera a China en casos confirmados acumulados, y la Organización Mundial de la Salud (OMS) declara que Latinoamérica es el nuevo epicentro de la pandemia. Uno de los misterios de las primeras etapas del fenómeno global está marcado por paradojas que nadie puede explicar bien: Perú parecía haber respondido rápido y con medidas preventivas efectivas, pero ha visto un crecimiento de casos que lo pone como segundo en una región donde no hay muchos ejemplos de éxito en la contención del virus. Países ricos (Suecia) y pobres (Nicaragua) han visto rebozados sus sistemas de salud. Populismo de izquierda (Nicaragua, México) y derecha (Estados Unidos, Brasil), con alto índice democrático (Italia, Francia, Alemania) y países que gozan de menos libertades civiles (Rusia), todos por igual han sufrido brotes que se han salido de control. En este contexto, Colombia no ha sido (hasta ahora) ni un país con brotes que desborden el sistema de salud, ni tampoco uno donde se haya tenido controlado el brote. El país entró en cuarentena de manera relativamente temprana (comparado con la respuesta mundial y gracias al ejemplo de países como USA, España e Italia), pero a pesar de eso la transmisión ha mantenido una tendencia de crecimiento de casos y muertes. La pandemia ha significado retos para la división político administrativa (diferencias entre alcaldes y gobierno nacional, retos de ordenamiento territorial en zonas metropolitanas), para la comunidad académica (representado pero no limitado por debate entre economía y cuarentena), entre muchos otros.

La respuesta más efectiva ante el Covid-19 parece haber estado relacionada con el uso de tecnología digitales para vigilancia epidemiológica: Singapur, Corea del Sur, Taiwan y China, todos con ejemplos de innovaciones tecnológicas para rastreo de contactos y uso de grandes datos para toma de decisiones en salud pública relacionada con la pandemia [Parodi and Liu, 2020, Lee et al., 2020, Wang et al., 2020, Ting et al., 2020]. En general, hay tres aspectos críticos que han sido resaltados para entender tanto la heterogeneidad de la epidemia por regiones, como la eficacia de las acciones llevadas a cabo por gobiernos (y la respectiva adherencia de sus poblaciones): 1. La heterogeneidad del contagio, o importancia de los superdispersores del virus. 2. Congregaciones que facilitan la dispersión (eventos atípicos en contexto de cuarentena), como celebraciones religiosas, funerales, carnavales, entre otros. 3. Conectividad regional e internacional [Hébert-Dufresne et al., 2020, Shim et al., 2020, MacIntyre, 2020].

2. Objetivo

Este texto pretende hacer una reflexión sobre algunos temas críticos (heterogeneidad del contagio, eventos de contacto atípicos, y movilidad) en el marco de la respuesta de Colombia a la pandemia y el proceso de toma de decisión en salud pública. Para ello, se presentan resultados de análisis de datos utilizados en el marco de las diferentes intervenciones implementadas en Colombia, y se discuten su alcance y pertinencia. Esto incluye:

- caracterizar el análisis de grandes datos en términos de superdispersión, eventos atípicos de contacto, y movilidad intra e inter focos epidémicos, con el fin de ilustrar los insumos que tienen las políticas de salud pública relacionadas con Covid-19.
- analizar la toma de decisión basada en los insumos discutidos anteriormente, con el fin de discutir la efectividad de las políticas basadas en datos en contexto de Covid-19.
- discutir y reflexionar sobre el papel de las políticas en salud pública (vigilancia epidemiológica basada en grandes datos) en el contexto Colombiano.

3. Aspectos importantes para tener en cuenta al pensar en Covid-19 desde los grandes datos

Dos tipos de datos son de particular interés, los datos de movilidad en unidades espaciales definidas que permiten monitorear los cambios en los flujos de movilidad y los datos de contactos de infecciones activas para poder tomar medidas de aislamiento de contactos. Ambos tipos de datos pueden obtenerse a partir del rastreo de teléfonos celulares y proporcionan información invaluable para la toma de decisiones en política pública.

Los datos agregados y anonimizados de movilidad proporcionados por operadores privados permiten obtener información casi en tiempo real de los patrones de movilidad humana [Buckee et al., 2020]. De esta manera es posible: 1) evaluar el nivel de acatamiento de las medidas asociadas a la cuarentena y el distanciamiento social a través del monitoreo cuantitativo de cambios (reducciones o aumentos) de movilidad al interior de unidades espaciales o entre unidades espaciales. 2) Identificar mecanismos y consecuencias de las medidas adoptadas. 3) Predecir la tendencia, en períodos cortos de tiempo, de cómo progresa la infección en relación a los cambios de movilidad comparados con situaciones similares ocurridas previamente en otros lugares del mundo, una aproximación propensa a la automatización a través del entrenamiento de algoritmos o "machine learning". 4) Delimitar de manera cuantitativa unidades espaciales de conectividad (o aislamiento) de acuerdo a los niveles de conectividad reflejados en los flujos de movilidad.

La integración de este tipo de datos con información epidemiológica (e.g. número casos, utilización hospitalaria) y demográfica (e.g. ingreso per cápita; pirámide poblacional) permite la elaboración de políticas dirigidas a distintos sectores de la población.

Los datos anonimizados de contactos de una persona contagiada o de una persona con síntomas y su red de contactos identificados a través de datos de cercanía de celulares en una determinada unidad de tiempo permiten identificar y hacer el rastreo y aislamiento de casos sospechosos [Kishore et al., 2020, Oliver et al., 2020]. Estos datos son particularmente útiles en la medida en que permiten el aislamiento de casos, y el rastreo de la cadena de contactos. Este tipo de información (información de redes) puede además ser eventualmente utilizada para determinar parámetros de importancia en la progresión de la infección que pueden ser utilizados en modelos mecanísticos de la infección [Chinazzi et al., 2020].

Un problema con respecto a los modelos que tratan de simular las condiciones en que progresa la infección y predecir el tamaño de la misma es que muchos de los parámetros sobre los cuales depende el evento son inciertos o mal comprendidos [Holmdahl and Buckee, 2020]. Uno de ellos es precisamente la movilidad humana, vista como la probabilidad de introducción de un evento infeccioso en una población y el carácter estocástico del evento infeccioso, en el sentido en que un individuo puede contagiar un número muy grande de personas mientras que otros por su parte pueden no contagiar a nadie [Hébert-Dufresne et al., 2020]. Este efecto refleja por lo tanto la heterogeneidad de la distribución de las infecciones secundarias. Los datos de rastreo de la cadena de contactos permiten obtener datos acerca de esta distribución, incorporarlos eventualmente en los modelos y así disminuir la incertidumbre en la estimación del tamaño de la infección [Hébert-Dufresne et al., 2020].

4. Métodos

El análisis de datos presentado acá es una muestra de los informes y productos que se han desarrollado para diferentes entes territoriales y de salud pública en Colombia. Hay dos fuentes de datos principales: datos de movilidad, agregados y dispuestos en forma de Infraestructura como Servicio (IaaS, por su sigla en inglés) en la plataforma GeoInsights de Facebook. Estos datos están compuestos por diferentes vectores que denotan movimiento entre dos puntos georreferenciados en intervalos de ocho horas, con información correspondiente a la línea base de personas, el valor observado para ese intervalo de tiempo, y algunos cálculos estadísticos basados en estos valores y la distribución (cambio porcentual, valor-z, significancia, diferencia, entre otros). Adicionalmente está disponible un conjunto de datos que caracteriza el movimiento para cada municipio con base en dos variables: cambio porcentual en número de cuadrantes visitados por usuario, y proporción de usuarios que visitaron un solo cuadrante. Con los datos agregados de movilidad entre puntos geográficos se construyó una red pesada en la cual los nodos son municipios y las aristas el número de personas que se movilizan entre municipios. A esta representación en red de la movilidad se le hizo un análisis de estadística descriptiva, y así mismo se realizó un ejercicio de detección de comunidades. Finalmente, para algunas comunidades se implementó un algoritmo de predicción a corto plazo (*nowcasting*) basado en datos para evaluar la evolución de la epidemia en los siguientes 9 días. Dado que la epidemia en Colombia llevaba un retraso con respecto a otros lugares del mundo, para evaluar las dinámicas con más información, se implementó un algoritmo de búsqueda de dinámicas similares (*dynamic time warping*, DTW por su sigla en inglés) para todo el mundo, con la cual se entrenó un algoritmo de inteligencia artificial que para un conjunto de datos de movimiento interno y externo, y una distribución espacial de casos, proyecta en el tiempo los casos para 9 días.

Por otro lado, gracias a la colaboración con la empresa ServiInformación, está disponible un conjunto de datos de movilidad de teléfonos celulares que provienen de un servicio de agregación de aplicaciones para minería de datos sobre usuarios con fines de mercadeo digital disponibles a nivel mundial. Estos datos son anónimos y están disponibles comercialmente, y han sido utilizados en el pasado con fines de investigación para caracterizar la movilidad de las personas. El conjunto de datos cuenta con un ID anonimizado para cada usuario, y coordenadas donde interactuó con diferentes aplicaciones. Con base en estos datos, se construyó una matriz de distancia entre usuarios que compartieron el mismo tiempo y espacio, la cual puede ser representada como una red de contactos, donde A está conectado con B si se encontraron en el mismo lugar y en tiempos similares (con una diferencia entre las dos observaciones no mayor a 10 minutos, definido arbitrariamente). Para esta red, construida para intervalos de días y semanas, se implementaron métricas de red básicas, como centralidad de grado, de valor propio y Pagerank, así como modularidad y otros algoritmos para detección de comunidades.

Ambos datos pueden contener sesgos importantes por acceso a tecnología en dispositivos móviles. Sin embargo, también han sido utilizados exitosamente para caracterizar movilidad desde un interés académico y comercial, y en menor medida, para formulación de políticas públicas.

Los datos de morbilidad y mortalidad por Covid-19 están disponibles a nivel municipal para Colombia para el público general y se pueden descargar directamente y en su versión más actualizada desde la página del Instituto Nacional de Salud (INS) de Colombia.

5. Resultados del análisis de grandes datos, insumos para políticas públicas

Los reportes producidos para tomadores de decisión en políticas públicas contienen información similar a la presentada en esta sección. Esta información hace parte de una amplia gama de soportes, muchos de los cuales no están contenidos en este manuscrito pues desbordan su objetivo, con los cuales se toman decisiones a escalas sin precedentes en una historia reciente: cerrar fronteras y vuelos internacionales, decretar cuarentenas nacionales, cerrar sectores de una ciudad o de la economía, relajar dichas restricciones, y desplegar equipos y recursos como epidemiología de campo, todas son decisiones que se toman con base en, entre otras, esta información.

5.1. Movilidad

El gobierno nacional decretó cuarentena obligatoria el 24 de marzo, 5 días después de que varios alcaldes de algunas de las ciudades más importantes del país iniciaran un simulacro. La movilidad del país se redujo significativamente, como lo muestran las diferentes representaciones de la figura 1, como consecuencia de las medidas preventivas implementadas por el gobierno (primero el simulacro a nivel subnacional, y luego el decreto por el cual se reglamenta la restricción de movilidad). Los datos muestran, además, que la cuarentena, lentamente, ha perdido su adherencia en la población. En general, vemos que las ciudades principales (en colores púrpura y rosa) han reducido mucho más la movilidad que otros territorios con ciudades más pequeñas o de vocación rural (azule, verde, amarillo, naranja y rojo).

La misma información pero agregada a nivel de unidades funcionales (generadas con el algoritmo de detección de comunidades en redes de movimiento)[[Amekudzi et al., 2012](#), [Nelson and Rae, 2016](#)] permite caracterizar el riesgo territorialmente. En la gráfica se ve cómo la movilidad interna y externa caracteriza el riesgo en las regiones en términos de importar casos, contagio interno, o la combinación.

Con esta información se pueden hacer estimaciones para los casos en un futuro cercano, entrenando el algoritmo con las dinámicas de regiones con proyecciones similares en otros lugares del mundo. La figura 3 muestra las series vecinas a la progresión de casos y movilidad de Bogotá, y la predicción que se hace para la capital en el contexto global de la pandemia, pero hecho específicamente para sus características.

5.2. Redes

En Colombia, el trabajo de campo epidemiológico y el rastreo digital de contactos ha revelado aspectos importantes de la epidemia. En Bogotá, uno de los brotes más fuertes hasta la fecha ha sido el de la localidad Kennedy, pero los primeros casos de la ciudad (y el país, dicho sea de paso) tuvieron lugar en las localidades de Usaquén y Chapinero. ¿Cómo y por qué el virus migró a Kennedy (también Suba y Engativá han sido afectadas por la segunda oleada) y se dispersó por toda la ciudad? El evento de dispersión a otras zonas probablemente tuvo lugar alrededor de la semana 10, por lo cual decidimos explorar las dinámicas de contactos alrededor de esas fechas. La gráfica 5 muestra el cambio en la red de contactos en el tiempo, y en particular para la semana en cuestión (se incluye la gráfica de la semana 9 también). En esta gráfica podemos ver cómo un evento durante esa semana cambió drásticamente la red de contactos observada para todo el periodo general. Al explorar la naturaleza de este fenómeno, encontramos que dicho incremento en las tasas de contacto (eventos asociados a la transmisión del virus en múltiples contextos) tuvo lugar en proximidad a la plaza de Lourdes, en la calle 63 con carrera 9, para ser más precisos. Esta es una zona de bastante tráfico, pero donde además hay varios establecimientos nocturnos. Sobre todo, este método muestra que los datos de proveedores para mercadeo digital pueden ser utilizadas para detectar contextos sociales asociados a la transmisión del virus en tiempo casi real, y para responder preguntas retrospectivas con el fin de entender procesos que facilitan o previenen la dispersión del virus.

6. Reflexiones sobre datos y políticas públicas

Uno de los temas más visibles como medida de salud pública en Colombia ante la epidemia de Covid-19, sin duda, ha sido el rastreo digital de contactos, que ha seguido los métodos acá descritos. Hasta ahora, desde nuestro conocimiento, tres entidades territoriales han implementado estas medidas: Bogotá, Valle y Antioquia, con resultados bastante heterogéneos. ¿Por qué en un lugar (Antioquia) sí ha funcionado el empoderamiento tecnológico de la salud pública, mientras que en los otros no parece haber dado resultados tan impactantes? En diversas comunicaciones, tanto la gobernación como la alcaldía de Medellín han dado parte de victoria sobre el uso de tecnología, mientras Cali y Bogotá han visto la epidemia crecer, probablemente a una tasa más baja que sin medidas.

Es importante hacer dos aclaraciones importantes. La primera, es que Bogotá ha implementado un modelo de confinamiento diferencial basado, entre otras cosas, en datos de rastreo digital. Mientras tanto, Antioquia ha fortalecido su sistema de seguimiento epidemiológico de campo con la tecnología como insumo, con gran interés en la epidemiología de campo. Esta diferencia de enfoques en políticas públicas no hace de un método de rastreo digital de contactos

eficiente o no, sino que por el contrario, ilustra que los datos en sí mismos no son una política pública. Más allá de comparar la evolución del brote en Medellín y Bogotá, algo que consideramos a todas luces inapropiado por múltiples razones, este ejemplo ilustra que el análisis de datos es solo un eslabón más en una larga cadena de procesos que componen una política pública. El éxito y su pertinencia está dada no solamente por las características de los datos y la calidad del análisis, sino también por el contexto cultural y político en el cual están inmersos. En otras palabras, los grandes datos en políticas públicas no pueden ser entendidos como un atajo para superar dificultades en un entorno específico, sino por el contrario, son producto de dicho contexto y nutren las prácticas de políticas públicas previas.

Por otro lado, los datos de movilidad que componen tanto las unidades funcionales para el control de Covid-19 en Colombia y su respectivo perfil de movilidad y casos presenta desafíos de otra índole. Si bien es cierto que han sido utilizados como herramienta para evaluar el grado de cumplimiento de medidas de distanciamiento social, en términos de políticas públicas es poco más lo que se ha podido aprovechar este conjunto de datos. En cualquier caso, esta información ha sido un insumo clave para alimentar modelos de simulación con los que posteriormente se toman decisiones, así como han alimentado el análisis que derivó en Bogotá en las zonas de cuidado especial (aunque el insumo principal en términos de grandes datos ha sido la red de contactos). Esto no significa que la información no sea útil, sino que uno de los retos que existe es entender la manera como se pueden incorporar divisiones político-administrativas de manera más fluida en el ordenamiento territorial y del gasto, así como el potencial que tienen para identificar condiciones de vulnerabilidad (se ha reportado en todo el mundo que los sectores más vulnerables de una ciudad son los que menos posibilidad tienen de resguardarse y cumplir la cuarentena). Sin duda, atender este reto será de vital importancia de cara a la siguiente etapa de la epidemia: el relajamiento de las medidas de distanciamiento social para aliviar la crisis económica derivada de las medidas de salud pública.

Referencias

- [Ainslie et al., 2020] Ainslie, K. E., Walters, C. E., Fu, H., Bhatia, S., Wang, H., Xi, X., Baguelin, M., Bhatt, S., Boonyasiri, A., Boyd, O., et al. (2020). Evidence of initial success for china exiting covid-19 social distancing policy after achieving containment. *Wellcome Open Research*, 5(81):81.
- [Amekudzi et al., 2012] Amekudzi, A. A., Banerjee, T., Barringer, J., Cmapbell, S., Contant, C. K., Doyle, J. L. H., Ankner, W., Fainstein, N., Fainstein, S. S., Faludi, A. K., et al. (2012). *Megaregions: Planning for global competitiveness*. Island Press.
- [Buckee et al., 2020] Buckee, C. O., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., Grad, Y. H., Grenfell, B., Halloran, M. E., Kraemer, M. U., et al. (2020). Aggregated mobility data could help fight covid-19. *Science (New York, NY)*, 368(6487):145.
- [Chin et al., 2020] Chin, T., Kahn, R., Li, R., Chen, J. T., Krieger, N., Buckee, C. O., Balsari, S., and Kiang, M. V. (2020). Us county-level characteristics to inform equitable covid-19 response. *medRxiv*.
- [Chinazzi et al., 2020] Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A. P., Mu, K., Rossi, L., Sun, K., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400.
- [Hébert-Dufresne et al., 2020] Hébert-Dufresne, L., Althouse, B. M., Scarpino, S. V., and Allard, A. (2020). Beyond r0: Heterogeneity in secondary infections and probabilistic epidemic forecasting. *medRxiv*.
- [Holmdahl and Buckee, 2020] Holmdahl, I. and Buckee, C. (2020). Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*.
- [Kishore et al., 2020] Kishore, N., Kiang, M. V., Engø-Monsen, K., Vembar, N., Balsari, S., and Buckee, C. (2020). Mobile phone data analysis guidelines: applications to monitoring physical distancing and modeling covid-19.
- [Kraemer et al., 2020] Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Du Plessis, L., Faria, N. R., Li, R., Hanage, W. P., et al. (2020). The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497.
- [Lai et al., 2020] Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., et al. (2020). Effect of non-pharmaceutical interventions to contain covid-19 in china.
- [Lee et al., 2020] Lee, V. J., Chiew, C. J., and Khong, W. X. (2020). Interrupting transmission of covid-19: lessons from containment efforts in singapore. *Journal of Travel Medicine*, 27(3):taaa039.
- [MacIntyre, 2020] MacIntyre, C. R. (2020). Global spread of covid-19 and pandemic potential. *Global Biosecurity*, 1(3).
- [Mitchell, 2002] Mitchell, T. (2002). Can the mosquito speak? *Rule of experts: Egypt, techno-politics, modernity*, pages 19–53.

- [Nelson and Rae, 2016] Nelson, G. D. and Rae, A. (2016). An economic geography of the united states: From commutes to megaregions. *PloS one*, 11(11).
- [Oliver et al., 2020] Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Delataille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., et al. (2020). Mobile phone data for informing public health actions across the covid-19 pandemic life cycle.
- [Parodi and Liu, 2020] Parodi, S. M. and Liu, V. X. (2020). From containment to mitigation of covid-19 in the us. *Jama*, 323(15):1441–1442.
- [Shim et al., 2020] Shim, E., Tariq, A., Choi, W., Lee, Y., and Chowell, G. (2020). Transmission potential and severity of covid-19 in south korea. *International Journal of Infectious Diseases*.
- [Ting et al., 2020] Ting, D. S. W., Carin, L., Dzau, V., and Wong, T. Y. (2020). Digital technology and covid-19. *Nature medicine*, 26(4):459–461.
- [Wang et al., 2020] Wang, C. J., Ng, C. Y., and Brook, R. H. (2020). Response to covid-19 in taiwan: big data analytics, new technology, and proactive testing. *Jama*, 323(14):1341–1342.
- [Wang and Tang, 2020] Wang, Z. and Tang, K. (2020). Combating covid-19: health equity matters. *Nature Medicine*, 26(4):458–458.
- [Zhou et al., 2020] Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798):270–273.
- [Zhu et al., 2020] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in china, 2019. *New England Journal of Medicine*.

7. Agradecimientos

Ana María Bermúdez y Manuel Riaño facilitaron el espacio de trabajo y gestionaron los datos y el equipo. Mónica Patiño y ServiInformación facilitaron datos y recursos para que esta investigación fuera posible. El equipo de Salud Pública de la Secretaría Distrital de Salud de Bogotá participó en discusiones en las cuales se discutió la importancia de este tipo de análisis.

8. Figuras

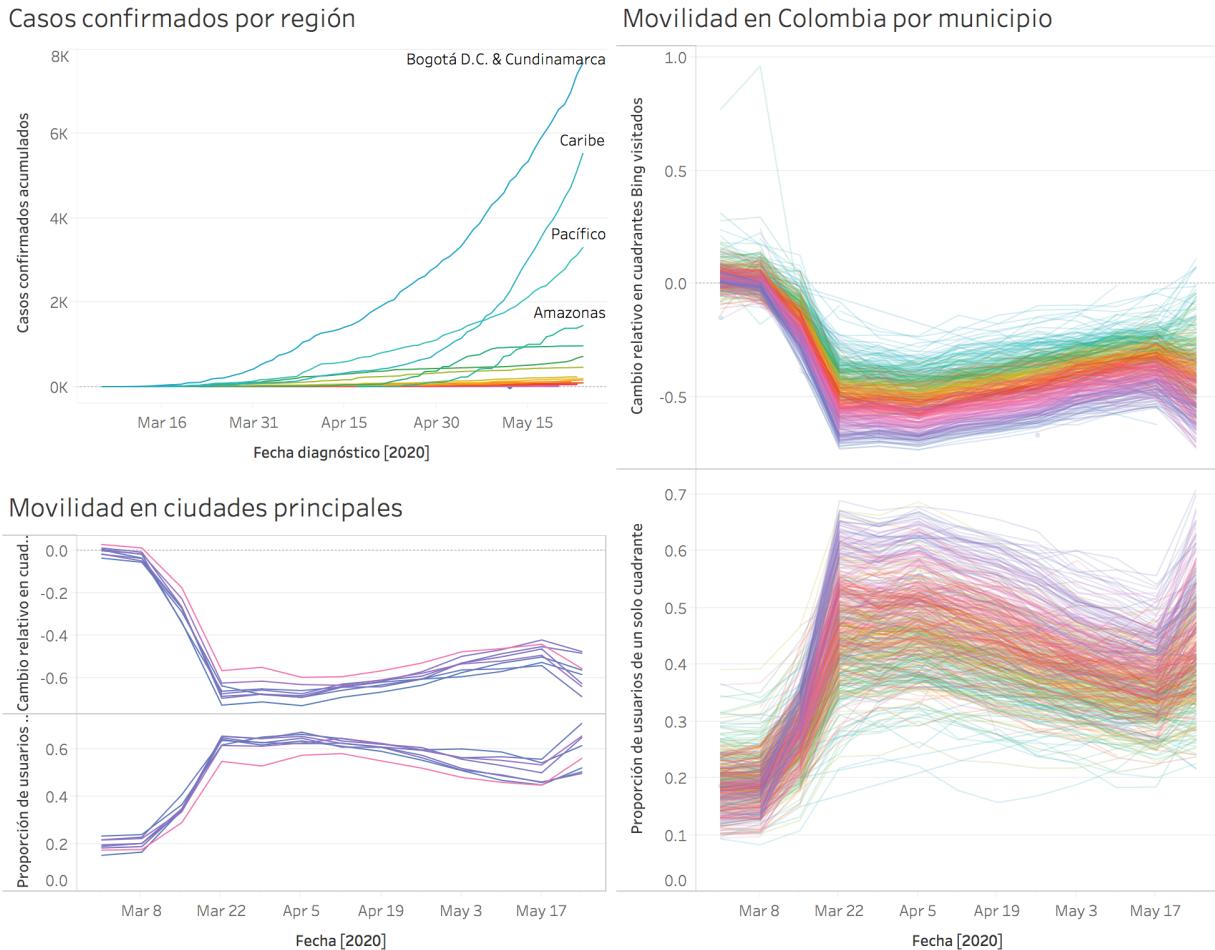


Figura 1: Movilidad y casos confirmados de Covid-19 en Colombia por región y ciudades principales. Izquierda, arriba: Casos confirmados de Covid-19 para Colombia, por región. Izquierda, abajo: Movilidad para algunas ciudades de Colombia (en orden decreciente: Barrancabermeja, Santa Marta, Medellín, Bucaramanga, Cali, Cartagena, Barranquilla y Bogotá). La parte superior muestra el cambio relativo en número de cuadrantes visitados. La parte inferior corresponde a la fracción de usuarios que visitaron solamente un cuadrante. Derecha: Movilidad para Colombia por municipio. Cada línea representa un municipio diferente. Se puede ver que la movilidad se redujo en más del 50% con el inicio de la cuarentena, y desde entonces ha subido progresivamente en todo el país. La parte superior muestra el cambio relativo en número de cuadrantes visitados. La parte inferior corresponde a la fracción de usuarios que visitaron solamente un cuadrante. La fuente de los datos de movilidad es GeoInsights de Facebook, mientras los datos de casos confirmados son publicados diariamente por el Instituto Nacional de Salud.

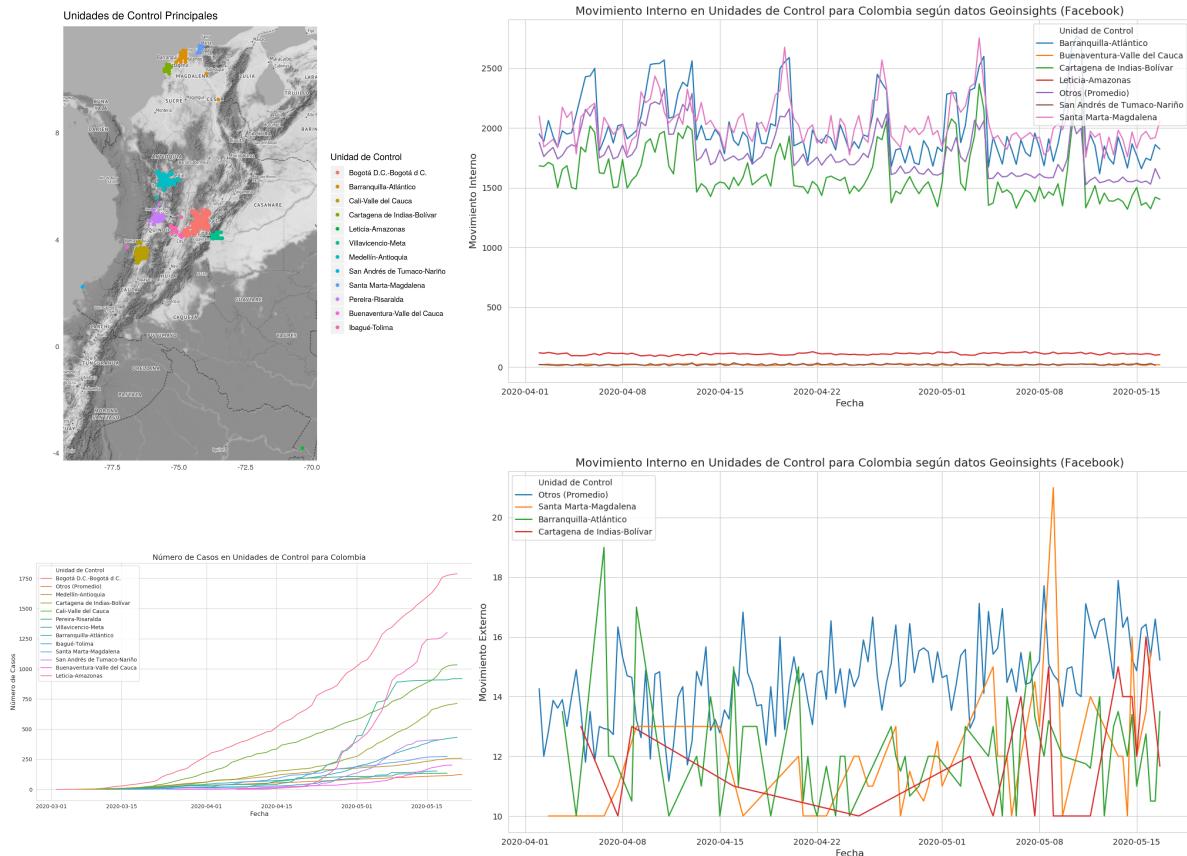


Figura 2: **Movilidad y casos confirmados de Covid-19 en Colombia por unidades funcionales** Izquierda, arriba: Unidades funcionales detectadas por el algoritmo de detección de comunidades con base en los datos de movilidad de GeoInsights de Facebook. Izquierda abajo: Casos confirmados de Covid-19 por unidad funcional. Derecha, arriba: movimiento interno para cada una de las unidades funcionales. Derecha, abajo: Movimiento externo entre unidades funcionales.

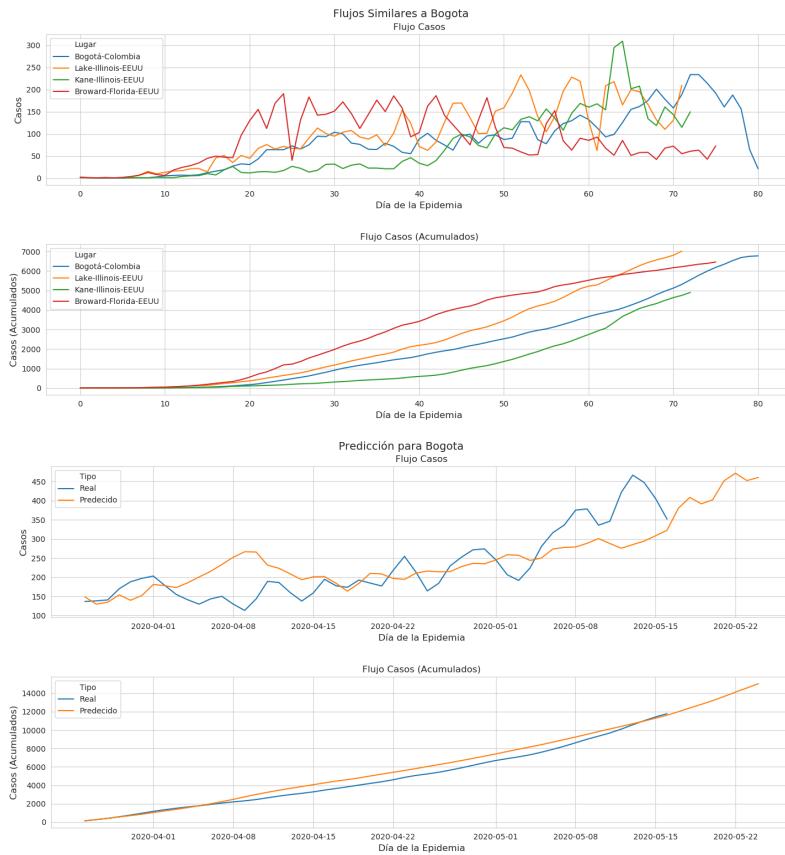


Figura 3: **Predictión de casos de Covid-19 para Bogotá basada en datos.** Con base en el algoritmo DTW se identifican regiones de otros países con tendencias similares a las presentadas por Bogotá (se toma como ejemplo pues es la región con más casos de Colombia, y por lo tanto la que produce resultados con mejor ajuste).

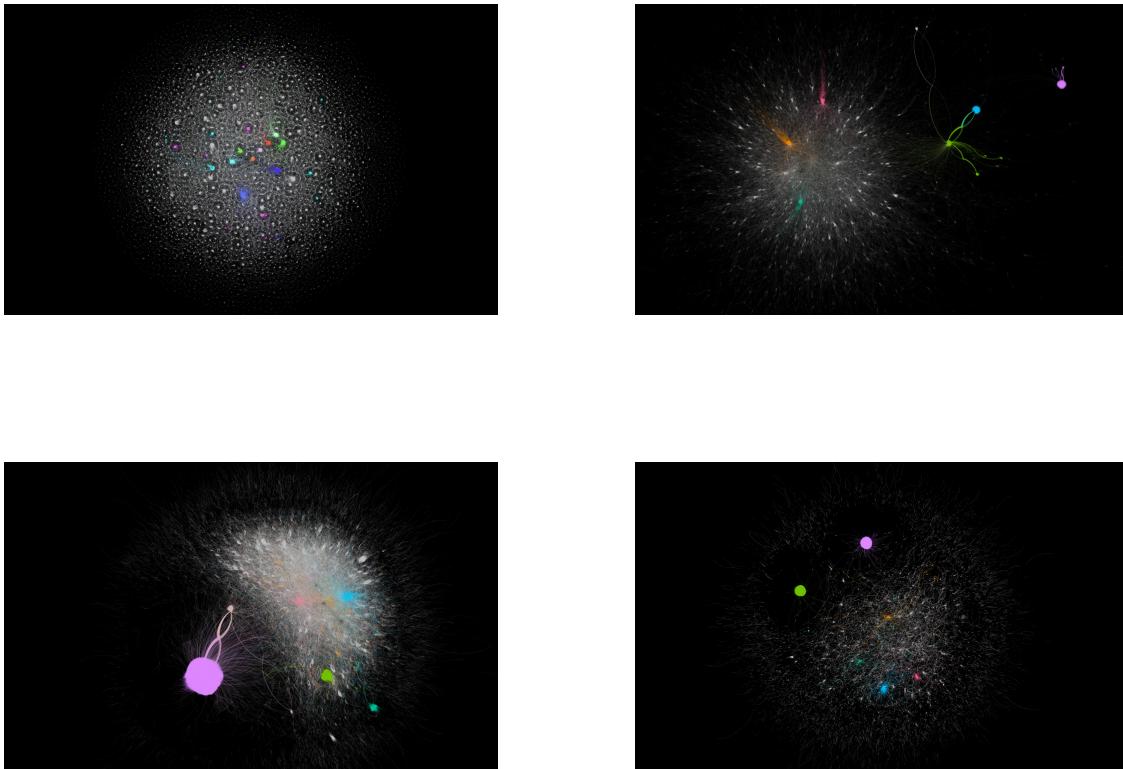


Figura 4: Red de contactos para Bogotá en el tiempo. Arriba, izquierda: Esta visualización permite ver todos los contactos observados para todos los dispositivos de los cuales se tiene registro en una red cuyos contactos están distribuidos en toda la red (no hay un solo individuo que tenga un número mucho más alto de contactos que el resto de la red, a diferencia, por ejemplo, de la red creada por vínculos de páginas web). Se pueden observar algunos *clusters*, resaltados en colores, que representan pequeñas comunidades en las cuales los individuos tienen más contactos entre sí que con personas fuera de la comunidad. Arriba, derecha: red de contactos para la semana 9, con los mismos métodos usados para construir la red completa. En esta se puede observar que hay varios contactos en una red donde no hay comunidades grandes, al igual que la red completa. Abajo, izquierda: red para la semana 10. Se observa un evento atípico, marcado por una comunidad de personas que tuvieron contacto con un dispositivo. Este evento tuvo lugar los días tres y cuatro de marzo (antes de que fuera instaurada la cuarentena). Abajo, derecha: Red de contactos para Bogotá en la semana 17, en la cual se presentó un brote en Corabastos. Se pueden observar dos clusters de contactos atípicos en Bogotá durante esta semana. Examinamos las semanas anteriores, que presentaban un fenómeno similar.

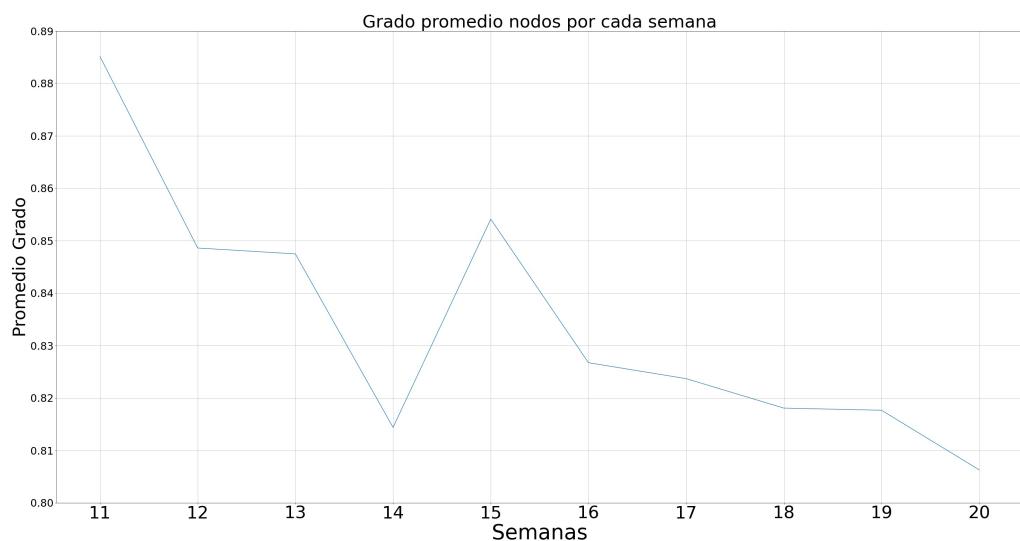


Figura 5: **Serie de tiempo de los grados promedio de las redes por semana** Esta serie de tiempo permite mostrar, en promedio, cuantos contactos tienen las personas cada semana. Las semanas se codifican de acuerdo al estándar ISO-8601. En la gráfica se puede observar una tendencia decreciente entre la semana 11 y la semana 20.