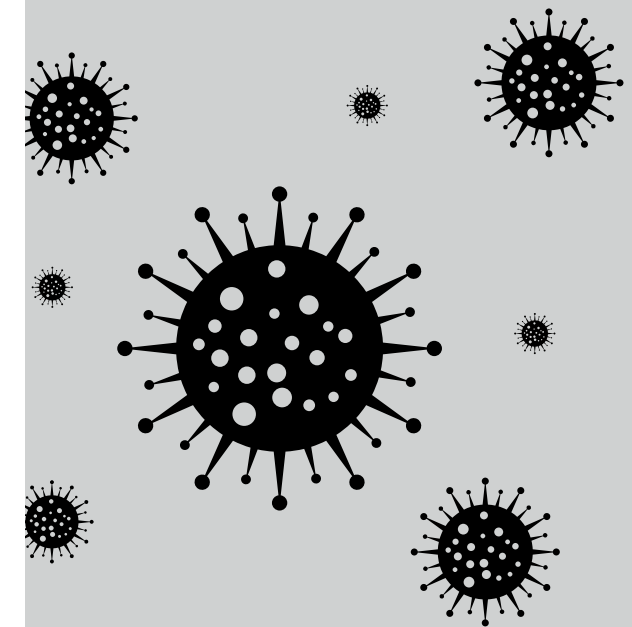




PAPER I

Applying Graph Databases and Human Mobility Data to Track Infectious Disease Spread in Brazil



PRESENTADO :
LUCIANA JULISSA HUAMAN COAQUIRA
DOCENTE :
ANA MARIA CUADROS VALDIVIA

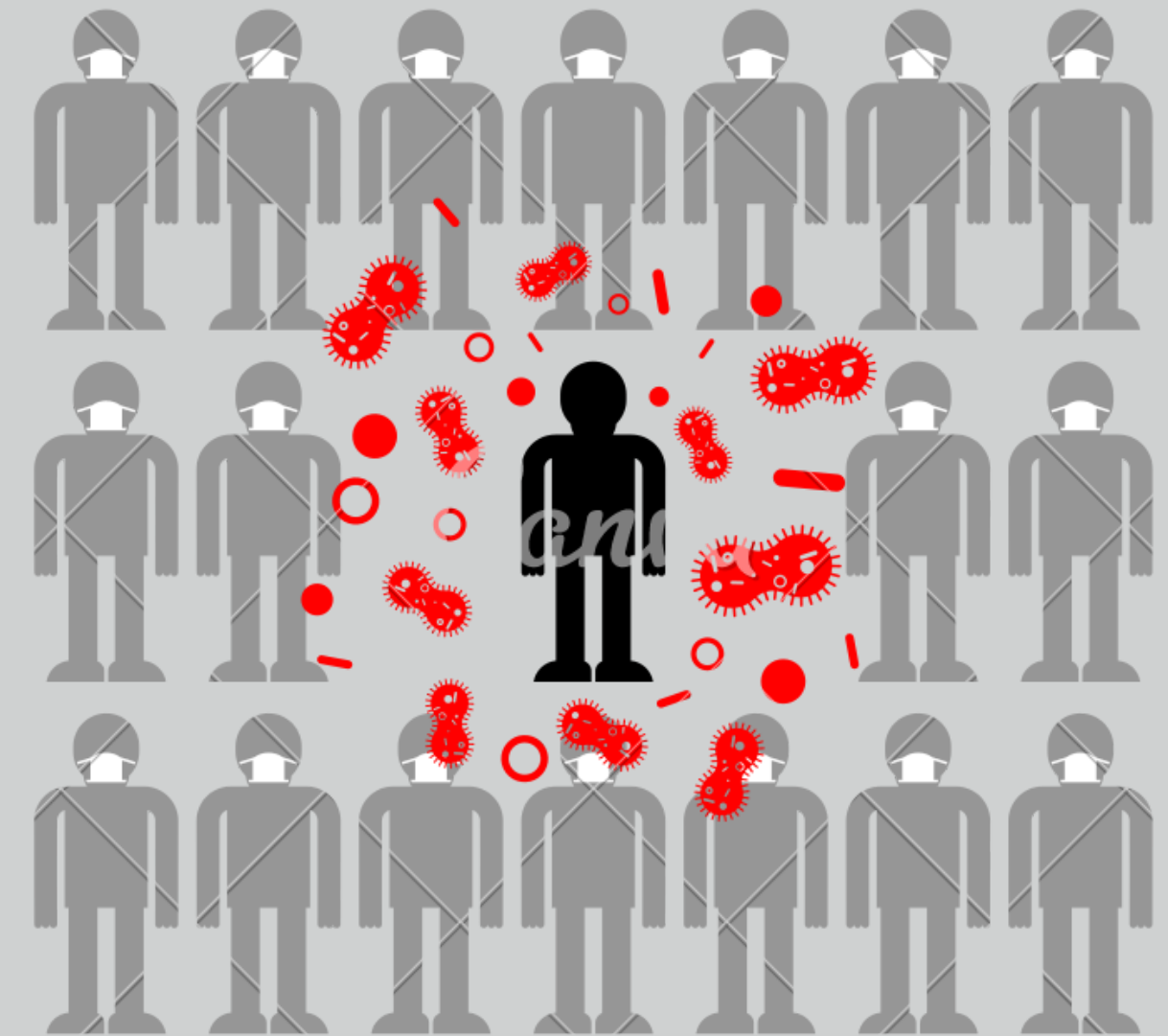


CONTENIDOS

- 01 **Introducción**
- 02 **Revisión Bibliográfica**
- 04 **Metodología**
- 06 **Herramienta**
- 08 **Resultados**
- 09 **Discusión**
- 10 **Conclusión**
- 11 **Bibliografía**

INTRODUCCIÓN

Las epidemias, como la COVID-19, tienen un impacto devastador en áreas clave como la salud y la economía. La propagación de enfermedades infecciosas está estrechamente relacionada con la movilidad humana, especialmente a través de viajes largos entre ciudades por transporte terrestre, aéreo o acuático. Según la Organización Mundial de la Salud (OMS), la COVID-19 demostró cómo los brotes pueden sobrepasar fronteras rápidamente, afectando a millones y causando enormes pérdidas económicas. Las epidemias, en general, no solo ponen en peligro la salud pública, sino que también alteran profundamente las estructuras sociales y económicas (OMS, 2020).



PROBLEMA

El principal desafío radica en la incapacidad de los modelos tradicionales para predecir de manera efectiva las rutas de propagación de enfermedades, especialmente cuando los datos de movilidad y transporte no están correctamente integrados. Aunque los datos de transporte existen, como los flujos de pasajeros entre ciudades, estos no siempre se utilizan de manera efectiva para prever la propagación de enfermedades. Esto limita la capacidad de las autoridades sanitarias para tomar decisiones informadas sobre la gestión de riesgos.

- **Retos:**

- Aunque existen grandes volúmenes de datos sobre movilidad humana (por ejemplo, transportes terrestre, aéreo), estos no se usan de manera efectiva en los modelos epidemiológicos.
- Los modelos existentes se enfocan más en el comportamiento biológico de las enfermedades y no en la influencia de los flujos de personas en la propagación de enfermedades.



REVISIÓN BIBLIOGRÁFICA

La movilidad humana es crucial en la propagación de enfermedades infecciosas (Mu et al., 2021; Bajardi et al., 2011). Modelos como metapoblación (Keeling y Rohani, 2008) y basados en agentes (Wei et al., 2021) han sido utilizados para simular la transmisión. Sin embargo, enfoques más simples, como el cálculo de probabilidades de flujo entre ciudades (Gilbert et al., 2020), son menos costosos computacionalmente. Este estudio usa probabilidades de viaje entre ciudades brasileñas, considerando transporte terrestre, aéreo y acuático, para predecir la propagación de enfermedades de manera más precisa.

METODOLOGIA Y DATSET

- **Datos de Transporte:**

- Fuentes: ANAC (aviación), ANTT (transporte terrestre).
- Incluyen flujos de pasajeros entre ciudades brasileñas, tanto para transporte aéreo, terrestre como acuático.

- **Datos de Salud Pública:**

- Fuentes: IBGE y SUS.
- Contienen información sobre el flujo de pacientes entre ciudades y la capacidad hospitalaria (número de camas, hospitales de referencia).

- **Datos Demográficos y Geográficos:**

- Fuente: IBGE.
- Datos sobre población, densidad y PIB de las ciudades brasileñas.

- **Datos de Propagación de Enfermedades:**

- Uso de datos de propagación del COVID-19 para validar el modelo, incluyendo variantes alfa y gamma.

- **Proceso de Transformación (ETL):**

- Se limpiaron y transformaron los datos usando técnicas como XGBoost para estimar flujos de pasajeros no reportados.

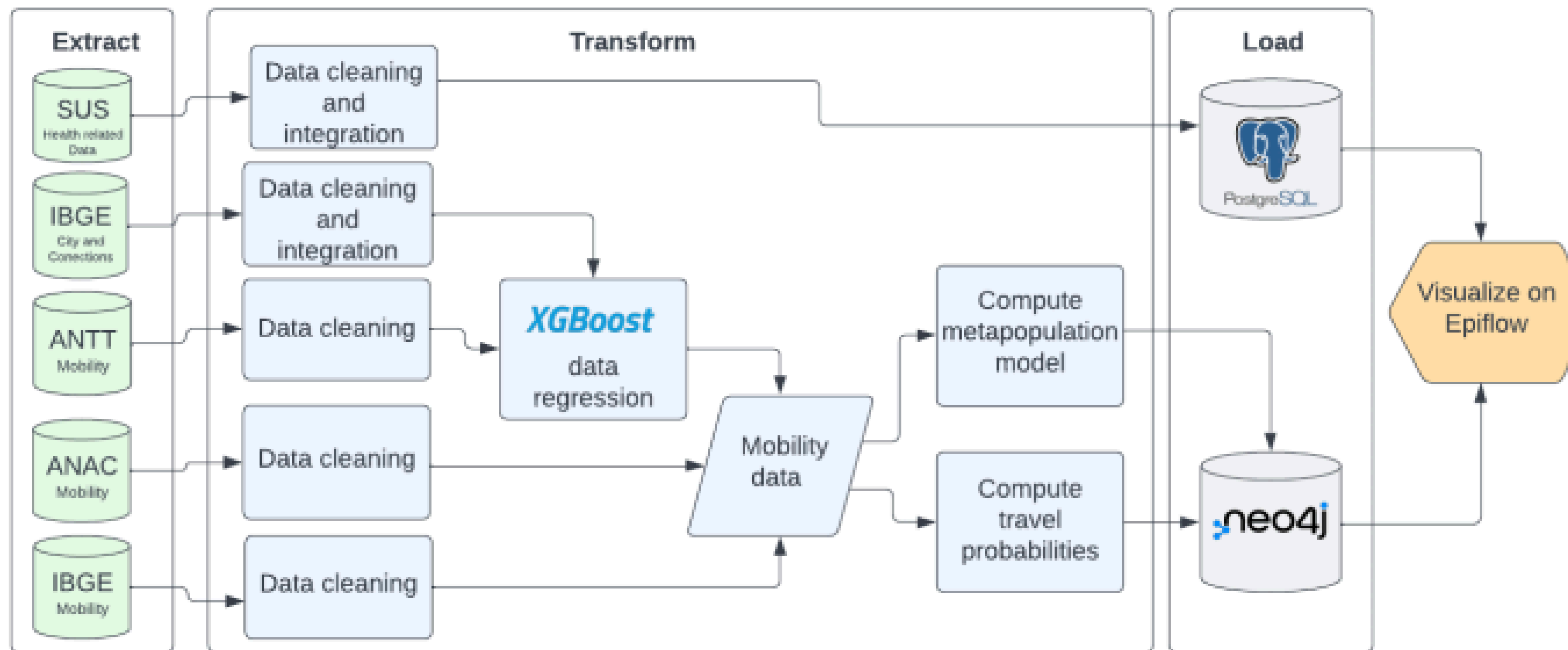
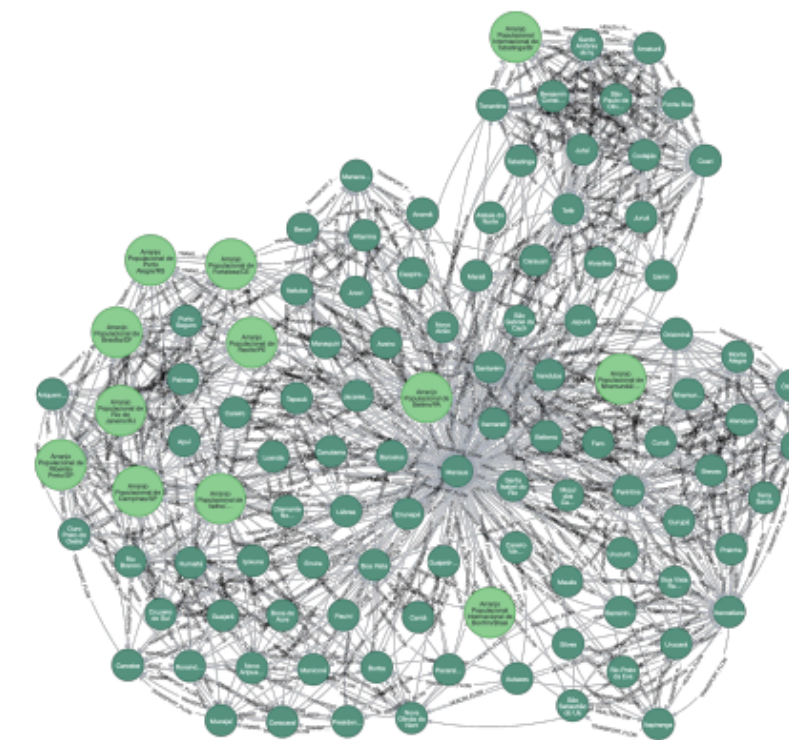


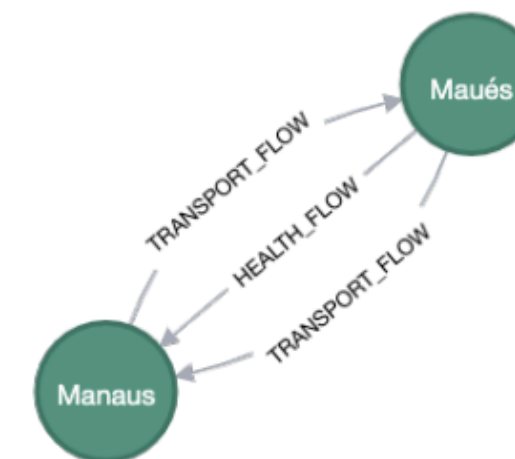
Figure 1. Flow diagram of the ETL methodology adopted.

METODOLOGÍA - ETL Y BASE DE DATOS GRÁFICA

- **ETL (Extract, Transform, Load):**
 - Extract: Recolección de datos de diversas agencias gubernamentales.
 - Transform: Limpieza y preparación de los datos utilizando herramientas como Pandas en Python.
 - Load: Carga de los datos transformados en las bases de datos.
- **Base de Datos Gráfica - Neo4j:**
 - Se utiliza para modelar las conexiones entre ciudades y las relaciones de transporte (bus, avión, barco) y salud.
 - Relaciones claves:
 - **TRANSPORT_FLOW:** Flujos de transporte entre ciudades.
 - **HEALTH_FLOW:** Flujo de pacientes entre ciudades.



(a)



(b)



METODOLOGÍA - IDENTIFICACIÓN DE CIUDADES EN RIESGO

- **Cálculo de Probabilidades:**

- Se calcula la probabilidad de que una ciudad reciba personas de una ciudad infectada, basada en el flujo de pasajeros.

- **Objetivo:**

- Identificar qué ciudades están en mayor riesgo de brotes, usando flujos de transporte y salud.

$$Pr(A \text{ to } B) = \frac{\text{No. of passengers from city A to city B}}{\text{Total no. of passengers from city A}} \quad (1)$$

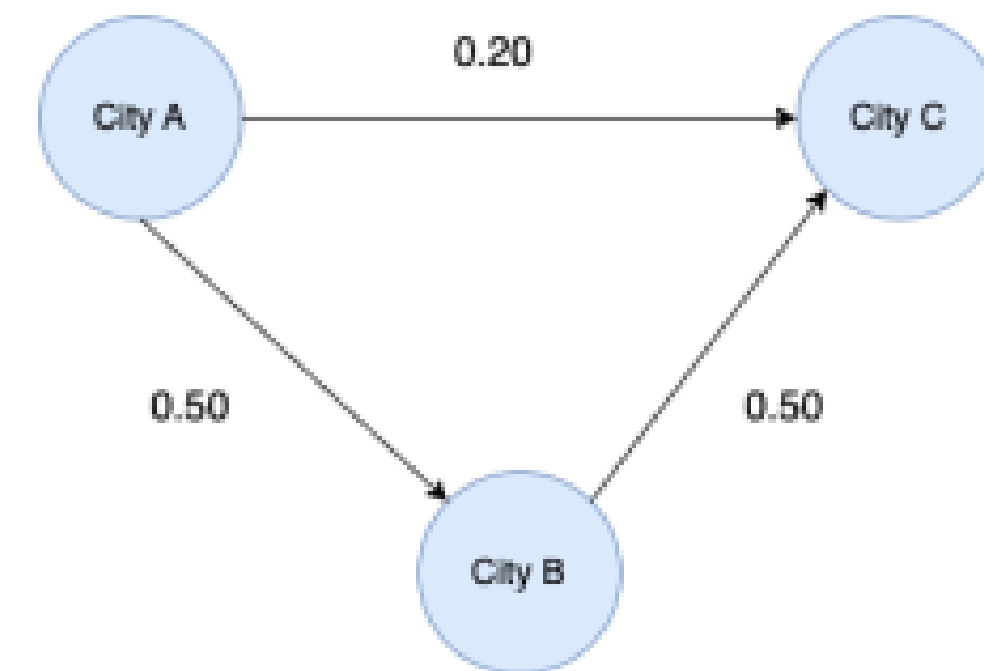
METODOLOGÍA - ALGORITMO DE PROPAGACIÓN - DIJKSTRA MODIFICADO

- **Dijkstra Modificado:**

- Se utiliza el algoritmo de Dijkstra para encontrar las rutas de propagación más probables entre ciudades, considerando las probabilidades de flujo como "pesos" de las aristas.

- **Modificaciones:**

- Multiplicación en lugar de suma para calcular la probabilidad total del camino.
 - Costo inicial de -1 para evitar que las probabilidades se anulen al multiplicarse.
- Esto permite identificar no solo la ruta más corta, sino la ruta de propagación más probable.



EPIFLOW - HERRAMIENTA DE VISUALIZACIÓN

- **Epiflow:**

- Herramienta interactiva para visualizar los flujos de transporte y las rutas de propagación de enfermedades.

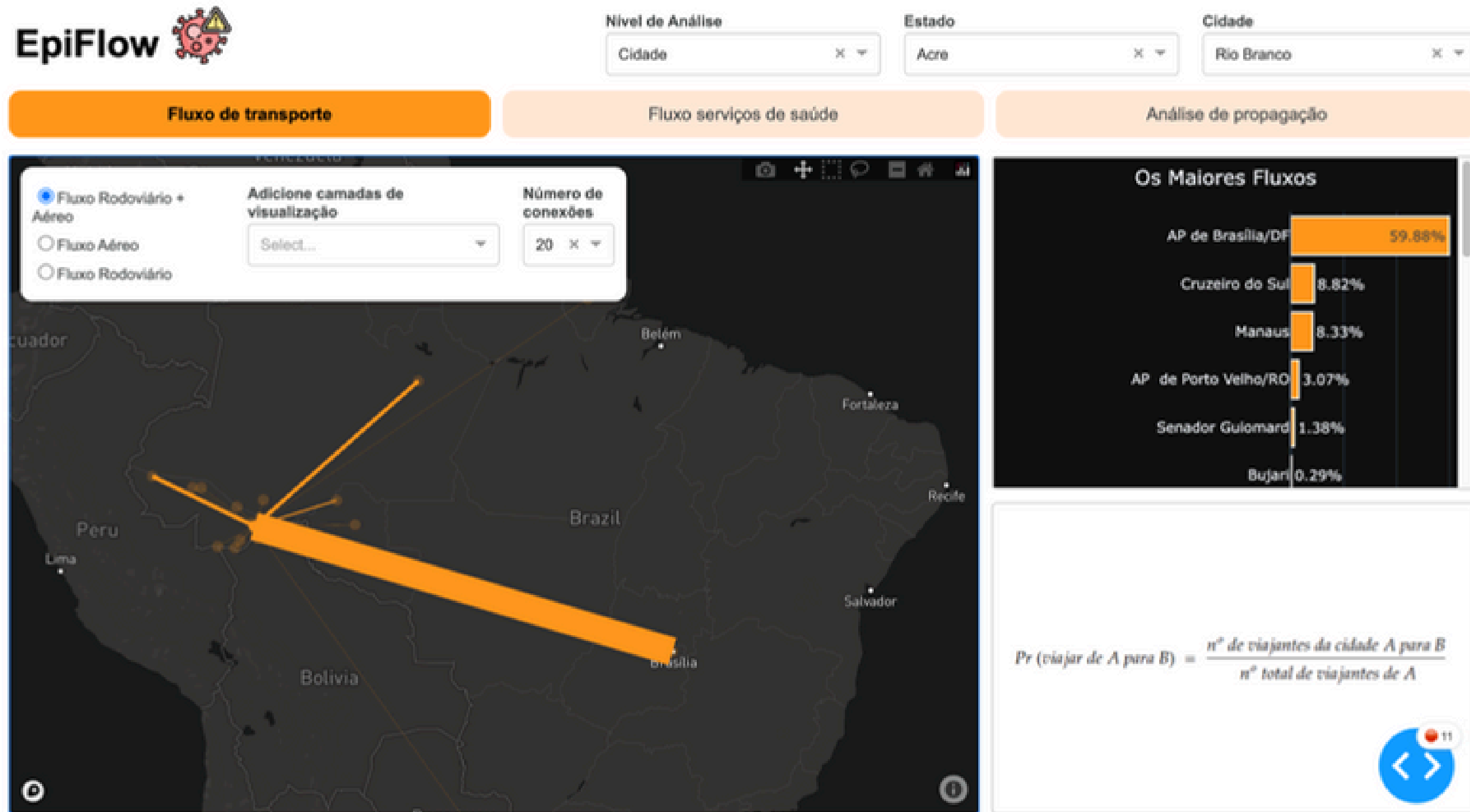
- **Funciones principales:**

- Visualizar flujos de transporte y salud desde una ciudad seleccionada.
- Explorar las rutas de propagación más probables hacia otras ciudades.

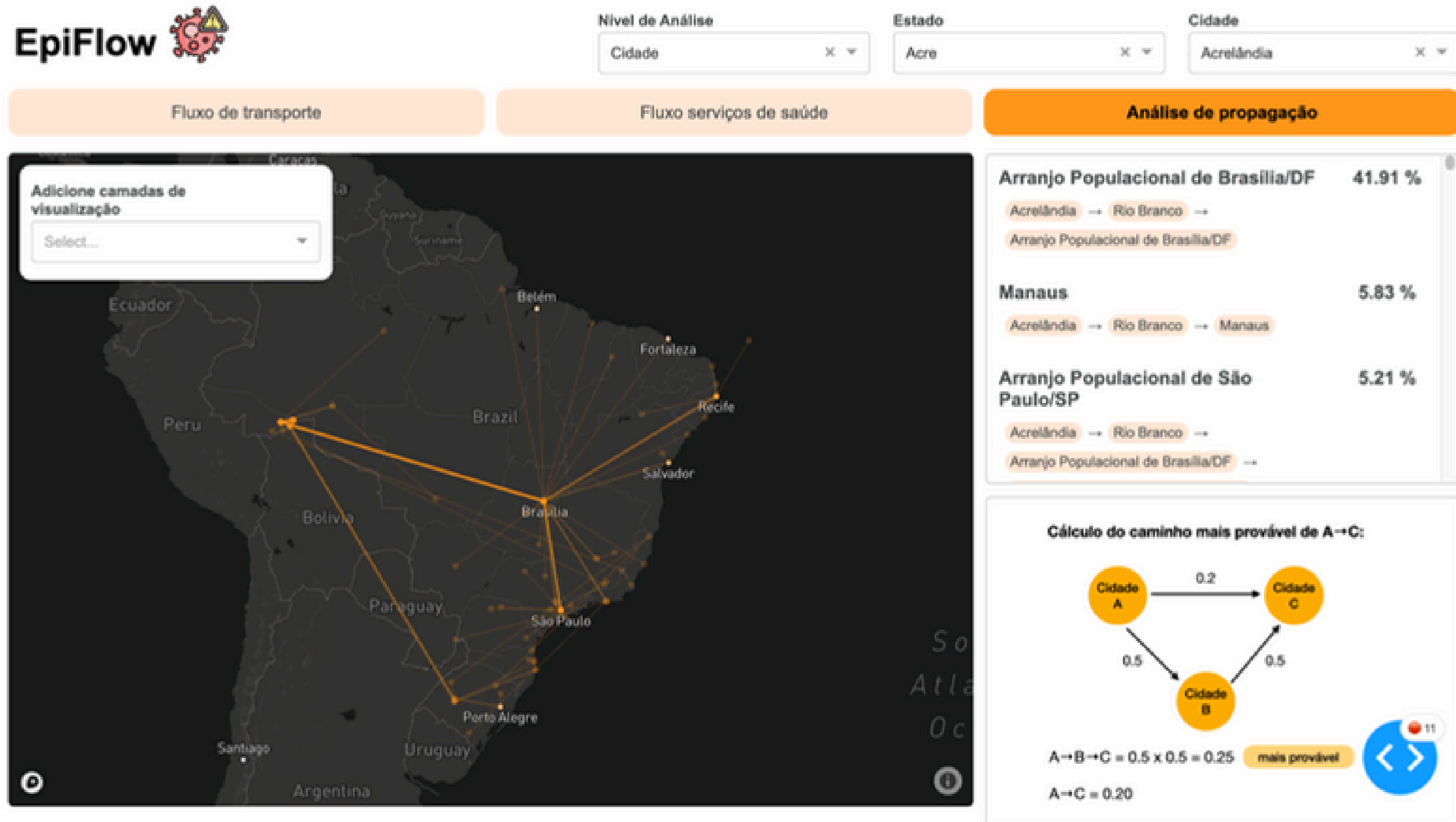
- **Interactividad:**

- El usuario puede seleccionar una ciudad y ver qué otras ciudades están en mayor riesgo.

EPIFLOW - HERRAMIENTA DE VISUALIZACIÓN



EPIFLOW - HERRAMIENTA DE VISUALIZACIÓN



RESULTADOS DE LA EVALUACIÓN

- **Validación con Datos de COVID-19:**

- Se validó el modelo con datos de las variantes alfa y gamma del COVID-19.

- **Resultados:**

- Variante alfa: Fuerte correlación entre las ciudades identificadas en riesgo y la propagación real.
- Variante gamma: Resultados inconclusos, probablemente debido a la falta de datos precisos de movilidad en ciertas áreas de Brasil.

- **Rutas de Propagación:**

- El modelo también identificó las rutas de propagación más probables, con una correlación moderada a fuerte para la variante alfa.

Table 2. Spearman's correlation between the system's probability of locations at risk vs. the actual spread of COVID-19

		Moving average threshold		
		0.5	1.0	1.5
Dataset 1	Alpha	$r(24) = -.87, p = .000$	$r(24) = -.82, p = .000$	$r(24) = -.84, p = .000$
	Gamma	$r(18) = -.38, p = .110$	$r(18) = -.20, p = .410$	$r(18) = -.19, p = .440$
Dataset 2	Alpha	$r(24) = -.86, p = .000$	$r(24) = -.82, p = .000$	$r(24) = -.84, p = .000$
	Gamma	$r(11) = -.40, p = .099$	$r(11) = -.27, p = .281$	$r(11) = -.27, p = .284$

Table 3. Spearman's correlation between the system's path spread probability vs. the actual spread of COVID-19

		Moving average threshold		
		0.5	1.0	1.5
Dataset 1	Alpha	$r(24) = -.82, p = .000$	$r(24) = -.83, p = .000$	$r(24) = -.81, p = .000$
	Gamma	$r(24) = -.40, p = .042$	$r(24) = -.14, p = .496$	$r(24) = -.11, p = .584$
Dataset 2	Alpha	$r(24) = -.83, p = .000$	$r(24) = -.84, p = .000$	$r(24) = -.77, p = .000$
	Gamma	$r(24) = -.28, p = .165$	$r(24) = -.13, p = .523$	$r(24) = -.09, p = .664$

DISCUSIÓN

- **Limitaciones:**

- Datos incompletos: Especialmente en áreas menos pobladas de Brasil y con bajo uso de transporte público.
- Modelo simplificado: No se consideraron viajes en automóvil, que son comunes en zonas rurales.

- **Áreas de Mejora:**

- Integrar datos más completos y precisos, como geolocalización móvil y más tipos de transporte.
- Evaluar el modelo con otros brotes epidémicos en Brasil (como Zika o Chikungunya).

CONCLUSIÓN

- El estudio confirma que el análisis de la movilidad humana mediante bases de datos gráficas puede predecir áreas de alto riesgo para brotes de enfermedades.
- La visualización interactiva es una herramienta poderosa para monitorear y mitigar la propagación de enfermedades.
- Se necesitan mejores modelos de flujo y datos más precisos para hacer predicciones más robustas.
- El enfoque propuesto puede ser replicado en otros contextos epidemiológicos con mejoras en los datos.

BIBLIOGRAFIA

- Cupkova, D., Azel, N., Mondor, C. (2015). EPIFLOW: Adaptive Analytical Design Framework for Resilient Urban Water Systems. In: Thomsen, M., Tamke, M., Gengnagel, C., Faircloth, B., Scheurer, F. (eds) Modelling Behaviour. Springer, Cham. https://doi.org/10.1007/978-3-319-24208-8_35

Gracias



METAPOPOPULATION GRAPH NEURAL NETWORKS: DEEP METAPOPOPULATION EPIDEMIC MODELING WITH HUMAN MOBILITY†

Luciana Julissa Huaman
Coaquira

La pandemia de COVID-19 generó una crisis global sin precedentes, afectando a millones de personas y desafiando la capacidad de los sistemas de salud pública para predecir y controlar su propagación. En este contexto, los modelos predictivos de epidemias jugaron un papel crucial en la planificación de intervenciones y la asignación de recursos. Sin embargo, los modelos tradicionales de predicción epidémica, como el modelo SIR, mostraron limitaciones al no poder capturar la complejidad de la propagación en múltiples regiones ni los cambios en los parámetros epidemiológicos con el tiempo. La movilidad humana, que juega un rol esencial en la propagación de enfermedades, era uno de los principales factores difíciles de modelar con precisión debido a la falta de datos detallados y actualizados sobre los desplazamientos entre regiones.



MOTIVACIÓN

La motivación principal de este estudio es superar las limitaciones de los modelos tradicionales de predicción epidémica, como el modelo SIR, que no capturan adecuadamente la propagación de enfermedades entre diferentes regiones ni la variabilidad temporal de los parámetros epidemiológicos. Además, la movilidad humana, un factor clave en la propagación de epidemias, a menudo no se integra bien en estos modelos debido a la falta de datos precisos y actualizados. Ante estos desafíos, el estudio propone un enfoque híbrido utilizando redes neuronales gráficas (GNNs) junto con el modelo SIR metapoblacional para mejorar las predicciones de la propagación de enfermedades, aprovechar los datos de movilidad y proporcionar resultados más interpretables y confiables.



PROBLEMA

El problema que aborda el artículo es la limitación de los modelos tradicionales de predicción epidémica, como el modelo SIR, que no son capaces de capturar la propagación de enfermedades entre múltiples regiones ni adaptarse a las variaciones temporales y espaciales de los parámetros epidemiológicos. Además, los datos de movilidad humana, esenciales para modelar esta propagación, son difíciles de obtener o no siempre están disponibles. El modelo MepoGNN busca resolver estos problemas combinando conocimientos epidemiológicos con redes neuronales gráficas para mejorar la precisión y la interpretabilidad de las predicciones.

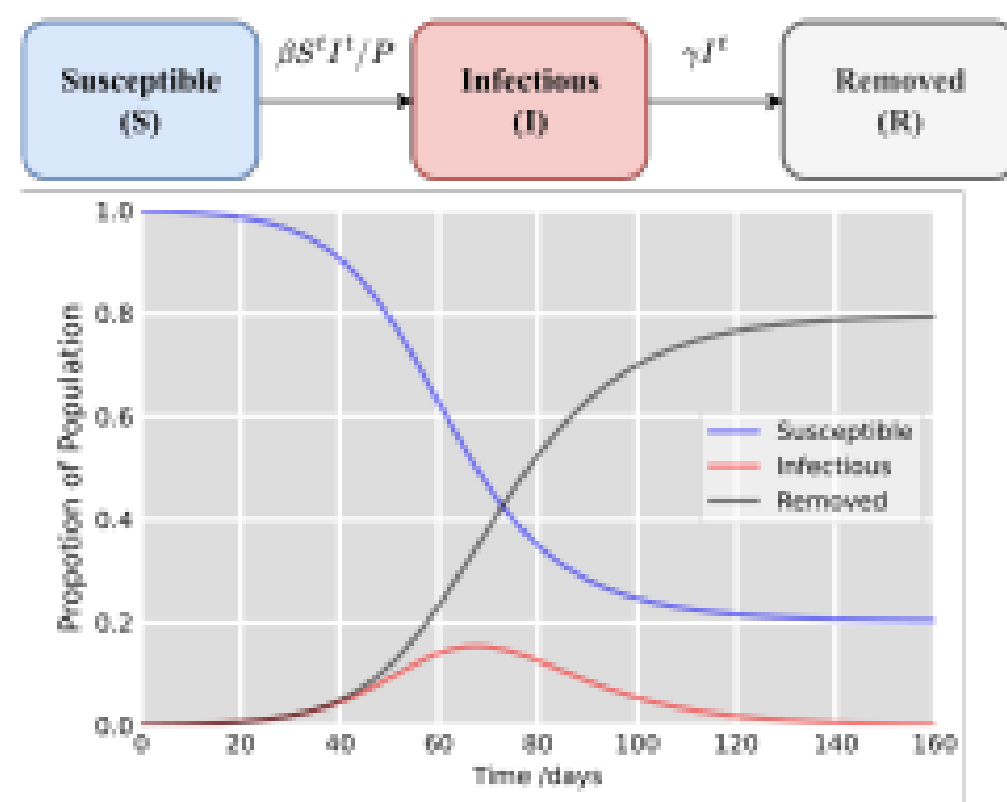


Fig. 2. Epidemic spread in classical SIR model.

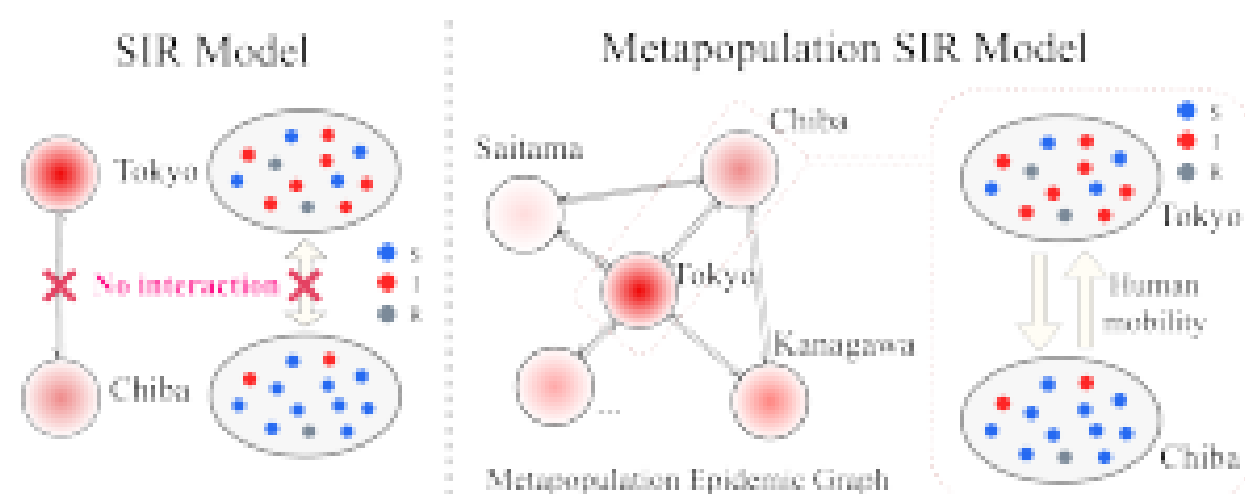


Fig. 3. Differences between SIR model and Metapopulation SIR model.

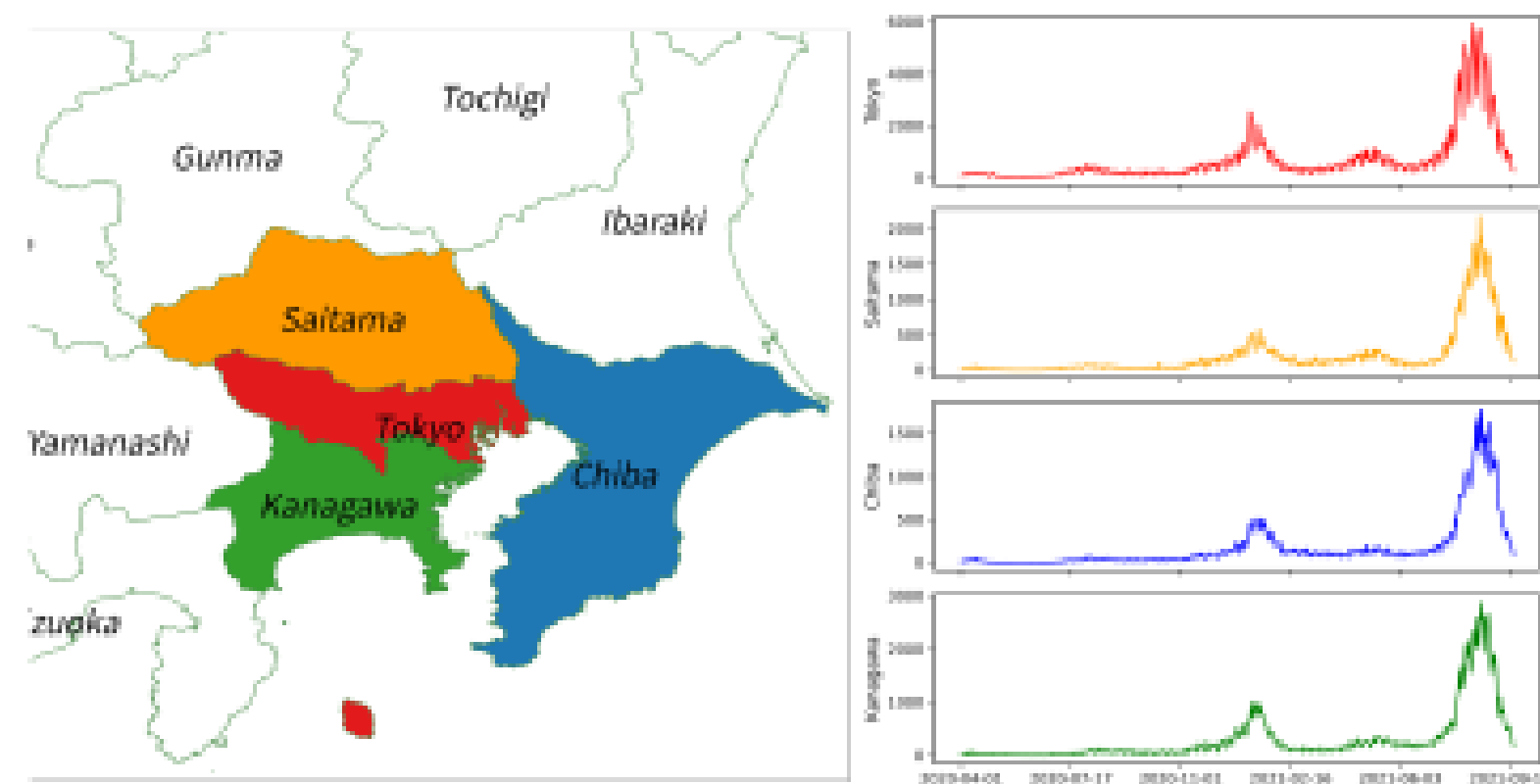


Fig. 4. Illustration of multi-regional epidemic forecasting.

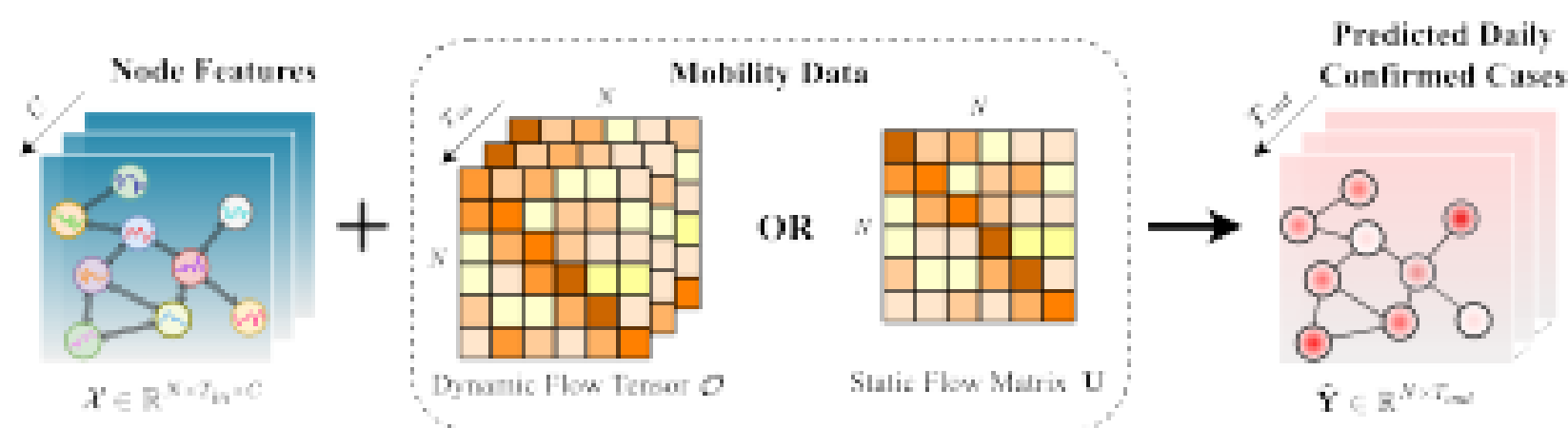


Fig. 5. Illustration of problem definition (model input and output).

ENFOQUE Y OBJETIVOS DEL ESTUDIO

- **Objetivo Principal:**
 - Desarrollar un modelo híbrido (MepoGNN) que combine el modelo SIR metapoblacional con redes neuronales gráficas (GNNs) para mejorar la predicción de epidemias a través de múltiples regiones, integrando datos de movilidad humana y otros factores externos.
- **Componentes del Modelo:**
 - Modelo SIR Metapoblacional: Utiliza la división de la población en subpoblaciones (por regiones) y la movilidad humana entre ellas para modelar la propagación epidémica.
- **Redes Neuronales Gráficas (GNNs):**
 - Capturan las dependencias espaciales y temporales, permitiendo aprender la estructura de propagación de epidemias a partir de los datos de movilidad.
- **Integración de Datos de Movilidad:**
 - Utiliza datos de movilidad humana (estáticos y dinámicos) para estimar las interacciones entre regiones y mejorar las predicciones.



METODOLOGÍA

Datos Utilizados:

- **Datos Epidémicos:**

- Incluyen el número de casos confirmados, recuperados y fallecimientos por región, obtenidos de la base de datos de COVID-19 en Japón.

- **Datos de Movilidad Humana:**


- Flujos Estáticos: Datos como las encuestas de movilidad (número de viajeros entre prefecturas).
- Flujos Dinámicos: Trayectorias GPS de las personas, que permiten modelar el movimiento en tiempo real entre regiones.

- **Datos Externos:** Incluyen factores como el cambio en los movimientos de las personas (en comparación con un período base) y las variaciones en las tasas de infectados activos.

METODOLOGÍA



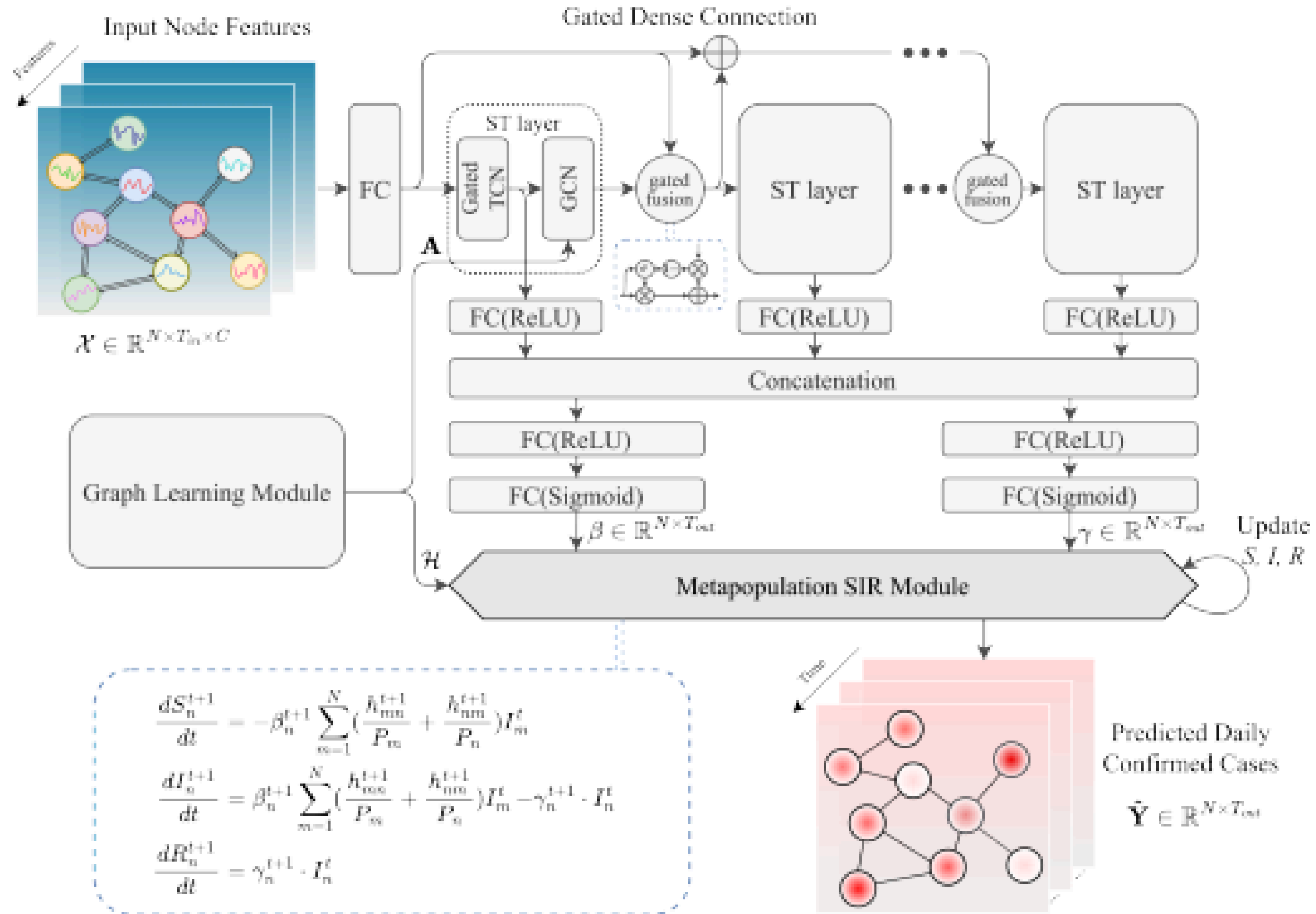
Procesamiento de Datos:

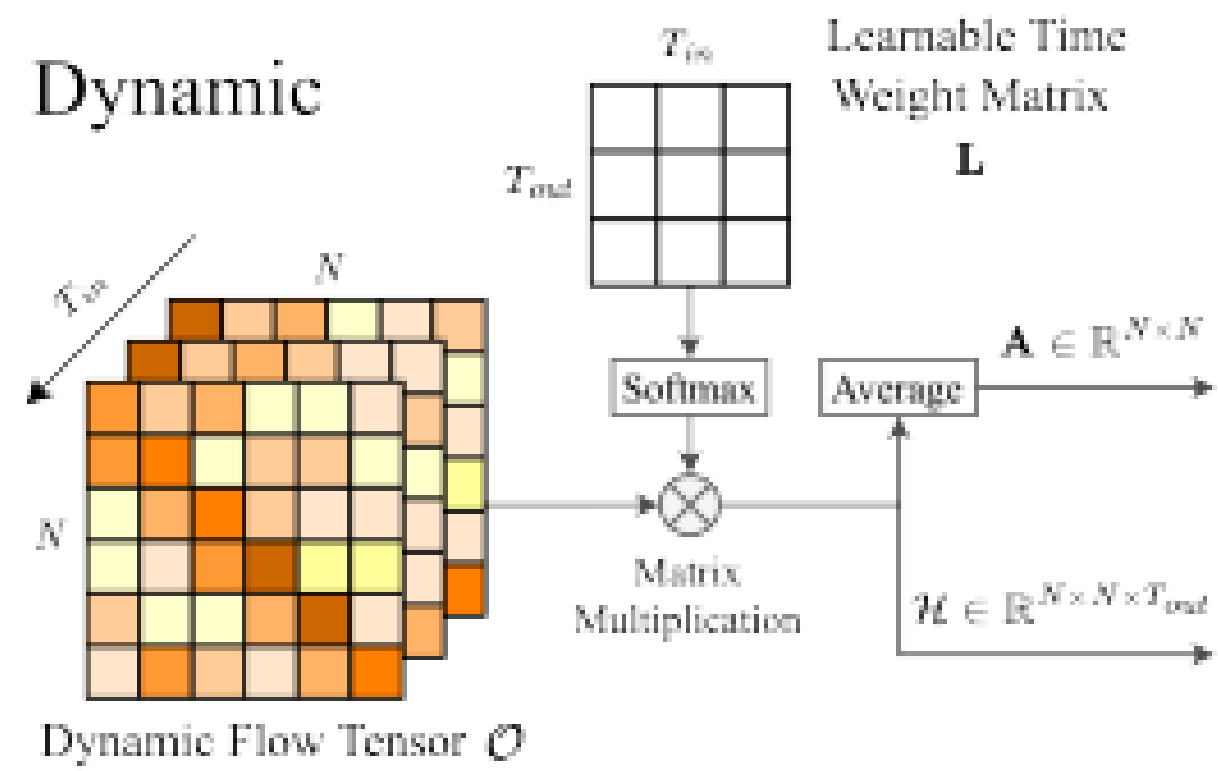
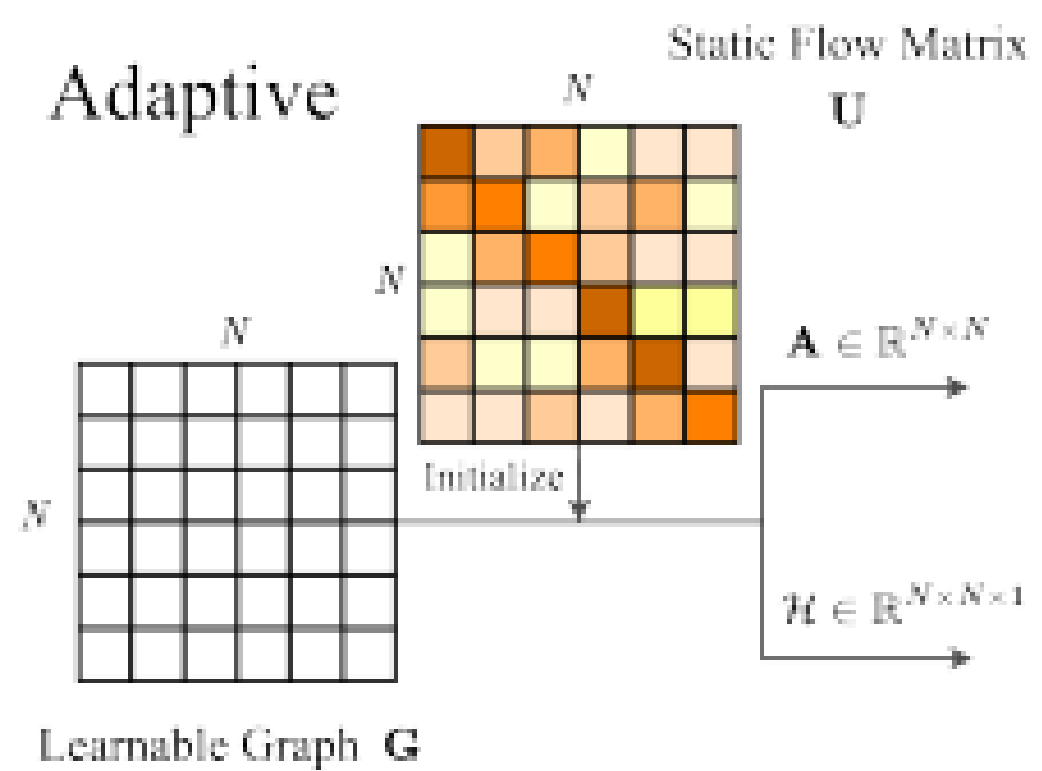
- Normalización de los Datos de Flujo: Se manejan las diferencias en la calidad de los datos de movilidad (por ejemplo, variabilidad en la tasa de usuarios de GPS) utilizando técnicas como la ponderación por población.
 - Generación de Datos de Movilidad: En ausencia de datos de movilidad detallados, se propone un método para generar movilidad relativa entre regiones utilizando solo datos de población y distancia entre las regiones.
- 

METODOLOGÍA

Construcción del Modelo MepoGNN:

- **Módulo SIR Metapoblacional:** Se modela la propagación epidémica entre subpoblaciones, con la movilidad humana como una matriz de interacción entre regiones.
- **Módulo Espaciotemporal:** Utiliza redes neuronales convolucionales para capturar la relación temporal (TCN) y espacial (GCN) entre las variables epidemiológicas.
- **Módulo de Aprendizaje de Gráficos:** Aprende y ajusta la estructura de movilidad entre regiones para mejorar las predicciones.





EVALUACIÓN

- **Entrenamiento del Modelo:** El modelo se entrena utilizando un conjunto de datos histórico de 47 prefecturas de Japón durante el período de abril de 2020 a septiembre de 2021.
- **Evaluación del Rendimiento:** Se comparan las predicciones del modelo MepoGNN con otros modelos base, utilizando métricas como RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), y RAE (Relative Absolute Error).
- **Evaluación de la Interpretabilidad:** Se analiza cómo el modelo MepoGNN es capaz de aprender parámetros interpretables, como el número reproductivo efectivo R_{tR_tRt} , y cómo estos parámetros se ajustan en función de los cambios en la movilidad y las intervenciones políticas.

CONCLUSIONES

- MepoGNN combina el modelo SIR metapoblacional con redes neuronales gráficas (GNNs), mejorando las predicciones epidémicas al capturar interacciones regionales y variaciones temporales.
- El modelo supera a los enfoques tradicionales y otros modelos de aprendizaje profundo, mostrando una mayor precisión en la predicción de casos y brotes inesperados.
- MepoGNN es altamente interpretable, permitiendo entender los parámetros clave de propagación y su relación con la movilidad humana y las intervenciones.
- Es aplicable a diversas epidemias y regiones, incluso cuando los datos de movilidad son limitados, gracias a su método de generación de movilidad.
- Futuras Mejoras: Aunque efectivo, el modelo aún necesita optimizarse para manejar brotes repentinos y extremos.



GRACIAS