

Informe lab 6

En este laboratorio se trabajó con un conjunto de datos de la cuenta de Twitter @TráficoGT, con el fin de analizar las interacciones entre usuarios y comprender cómo se organiza la información sobre movilidad en la ciudad de Guatemala. El análisis fue limpieza, preprocesamiento, construcción de grafos, detección de comunidades e interpretación de sentimientos.

El primer paso fue la limpieza y preprocesamiento de los datos. Se convirtió todo el texto a minúsculas, se eliminaron caracteres especiales como hashtags y menciones, se removieron enlaces, emoticones, signos de puntuación. Se extrajeron menciones, retweets y respuestas, lo cual permitió construir un dataframe estructurado con la información correcta. A partir de esto se generó una matriz de que eran las interacciones como un grafo dirigido, con nodos como usuarios y aristas como conexiones de interacción.

En el análisis exploratorio se encontró que la mayoría de tweets se concentraban en horarios de tráfico, principalmente en las mañanas y tardes. Las zonas más mencionadas fueron la Calzada Roosevelt y áreas como la zona 5, 9 y 13, que destacan por su alta congestión a estas horas. Entre los hashtags más frecuentes aparecieron términos relacionados con accidentes, lluvia, etc. Una nube de palabras confirmó que se hablaba demasiado de conceptos vinculados a tráfico, incidentes y clima.

Al construir la red de interacciones, se observó que el grafo presentaba baja densidad, lo que indica que la mayoría de usuarios interactúan con unos pocos nodos clave en lugar de entre sí. La aplicación del algoritmo Louvain permitió identificar comunidades bien definidas, tres de ellas con gran tamaño e influencia, donde cada una se concentró en temáticas o zonas específicas.

Los análisis de centralidad mostraron que la cuenta oficial de @TráficoGT es el nodo más influyente, pues recibe la mayor cantidad de menciones y retweets, seguida de usuarios que reportan tráfico de forma recurrente en sectores particulares. La centralidad mostró a usuarios que actúan como puentes entre comunidades.

También se identificaron pequeños grupos aislados que interactúan poco con el resto de la red. Estos grupos se enfocaron en incidentes locales o en áreas con menor cobertura. En el análisis de contenido y sentimiento, se observó que la mayoría de mensajes tenían un tono negativo, asociado a quejas, advertencias y reportes de accidentes. Algunos mensajes fueron neutrales, con un carácter más informativo, mientras que prácticamente no hubo mensajes positivos. El análisis de temas confirmó que los temas principales eran accidentes de tránsito, congestión en horas pico e impacto de la lluvia en la movilidad.

Los hallazgos muestran que la cuenta de @TráficoGT funciona como el eje central de la red, articulando los reportes ciudadanos y transformándolos en información útil para la gente. Las comunidades reflejan cómo distintas zonas de la ciudad concentran dinámicas específicas de tráfico, mientras que los influencers actúan como nodos que facilitan la difusión de datos en tiempo real. El predominio de sentimientos negativos refuerza la idea de que la conversación gira en torno a problemas y advertencias más que a experiencias positivas.

En conclusión, el análisis de esta red permitió comprender mejor cómo los usuarios organizan y comparten información en torno al tráfico de la ciudad. Se identificaron patrones de congestión consistentes en horas pico y en zonas críticas, lo que sugiere que estas condiciones se mantienen año tras año, especialmente durante la temporada de lluvia. Asimismo, se comprobó que los nodos influyentes y las comunidades detectadas tienen un papel clave en la formación de percepciones colectivas y en la manera en que la ciudadanía responde ante los problemas de movilidad urbana.