

Ejercicio - Caso de estudio de Neo4j

Integrantes

- Abby Donis - 22440
- Mathew Cordero - 22982
- Josué Say - 220801

Glosario

- **NBC News(National Broadcasting Company):** es una división de noticias de la cadena de televisión estadounidense.
- **Trolls:** en este contexto se refiere a cuentas falsas o personas en internet que publican contenido engañoso, provocador o dañino con la intención de manipular la opinión pública, generar controversia o desinformar.
- **Kremlin:** es un complejo de edificios civiles y religiosos en el centro de Moscú, Rusia.
- **Wayback Machine:** es un servicio de archivo digital que permite acceder a versiones antiguas de páginas web.

Contexto

Tomando como referencia el caso de éxito de **NBC News Analyzes Russian Troll Tweets with Neo4j**, el artículo describe cómo se confirmó la participación de actores rusos en la política estadounidense en 2016, respaldados por el Kremlin y utilizando Twitter como medio de difusión.

En 2017, el Comité Permanente de Inteligencia de la Cámara de Representantes de EE.UU. publicó una lista de 2,752 cuentas de Twitter vinculadas a la injerencia rusa, denominadas granja de trolls. Posteriormente, Twitter amplió esta lista a 3,814 cuentas.

El objetivo de los trolls era hacerse pasar por ciudadanos estadounidenses, organizaciones de noticias y grupos políticos, creando cuentas falsas para difundir desinformación y generar división.

NBC se propuso investigar este fenómeno, enfrentándose al desafío de detectarlos en un entorno de anonimato en internet y un enorme volumen de datos.

Para cuando se publicó la lista, Twitter ya había suspendido las cuentas y eliminado los tweets y perfiles de usuario, lo que obligó a los reporteros de NBC a encontrar una manera de recuperar los tweets eliminados de los trolls.

Ante este desafío, los reporteros recurrieron a Neo4j para recuperar la mayor cantidad posible de datos eliminados.

Para ello, utilizaron Wayback Machine de Twitter para recuperar información adicional de las elecciones, logrando compilar una base de datos de 202,973 tweets de 454 cuentas.

El objetivo principal era exponer las redes de trolls detrás de la injerencia rusa en las elecciones de EE.UU. en 2016. El desafío consistía en restaurar y analizar más de 200,000 tweets, y la solución se basó en el uso de la base de datos en grafos Neo4j para exponer dichas redes.

El análisis arrojó resultados significativos, como la identificación de trolls rusos haciéndose pasar por ciudadanos estadounidenses, medios locales y grupos políticos locales.

Se descubrieron patrones de retweets, hashtags y picos de actividad durante el horario laboral en Moscú.

Los algoritmos de detección de comunidades revelaron redes de usuarios que interactuaban con frecuencia, permitiendo identificar cuáles trolls eran influyentes y cuáles solo amplificaban el contenido de otros trolls. PageRank ayudó a identificar las cuentas más influyentes dentro de cada grupo.

A través de este análisis se pudo determinar que cada comunidad tenía un pequeño núcleo de generadores de contenido y un grupo más grande de retuiteadores. Aproximadamente el 25% de los tweets de los trolls eran originales; el resto eran retweets. Para ganar seguidores e influenciar a más personas, los trolls utilizaban hashtags comunes y respondían a cuentas populares.

Otro hallazgo relevante fue que, al analizar la actividad por horario, los tweets de los trolls aumentaban significativamente durante el horario laboral en Rusia.

Además, algunas cuentas no solo simulaban ser ciudadanos comunes, sino que también imitaban canales de noticias y organizaciones políticas, haciéndose pasar por partidos políticos específicos para favorecer o perjudicar a determinados grupos.

Preguntas

- **¿Cómo utilizó NBC News la tecnología de Neo4j para identificar patrones específicos en la actividad de los trolls rusos en Twitter durante las horas laborales en Rusia?**

- Primero limpio los datos de los tweets de cuentas asociadas a la IRA , obtenidos a través de webscraping y por la wayback machine.

- Utilizo patrones para identificar trolls rusos durante las horas laborales en Rusia, esto para saber si existia una estructura de NetCenter. De hecho utilizo Ne4j para identificar las relaciones entre entidades como tweets, usuarios, hashtags, aplicaciones de origen y enlaces y como cada uno se relacionaba entre si.
- Entre los patrones que recabaron es que la mayoria de los twweets fueron durante el periodo de trabajo en Rusia, que se enviaban desde un cliente web y sospechosamente que venian casi del mismo lugar.
- **¿Qué métodos específicos emplearon los trolls rusos para interactuar con usuarios legítimos en Twitter y ganar seguidores de manera efectiva?**
 - El 25% de los tweets de los trolls eran contenido original
 - Cuentas falsas que imitaban a ciudadanos estadounidenses comunes, o instituciones.
 - Coordinacion mediante una cuenta que generaba contenido y el resto se encargaba de retuitearlo y asi tomar impulso.
- **¿Cómo logró Neo4j ayudar a NBC News a descubrir la verdadera naturaleza de cientos de cuentas falsas operadas por la Internet Research Agency en Rusia?**
 - Neo4j les permitió revelar las relaciones entre las diversas entidades involucradas como los usuarios, los tweets realizados, hashtags, entre otros. Por medio de algoritmos de detección de comunidad se reveló que estos usuarios interactuaban de forma frecuente, siendo muy pocos los que creaban el contenido, de los cuales los más influyentes fueron identificados, y la mayoría de los trolls se dedicaban a esparcir este contenido por medio de retweets y respuestas a cuentas populares.

Ademas de que se descubrió que los tweets provenían de la aplicación wen y no de la app móvil como los usuarios comunes, además de que la actividad de las cuentas correspondían a las horas laborales rusas.
- **¿Qué reveló el análisis de Neo4j sobre la influencia de diferentes cuentas de trolls rusos en las redes y cómo se identificaron los trolls más influyentes dentro de cada clúster?**
 - El análisis reflejó que el modus operandi de los trolls rusos consistía en dividirse en pequeñas comunidades con líderes y seguidores. Los líderes representaban aproximadamente el 25% de las cuentas rusas y eran responsables de publicar tweets originales, que luego eran replicados por los seguidores para amplificar su alcance. Además, los trolls generaban interacción masiva al responderse entre sí, logrando así llegar a más personas.

También empleaban la táctica de utilizar hashtags comunes, permitiendo que sus publicaciones aparecieran en diversas búsquedas, no solo dentro de su propio nicho, sino también en las interacciones de usuarios legítimos.

- **¿Cuál fue el impacto de la exposición de los trolls rusos por parte de NBC News basada en el análisis de Neo4j en la conversación política en Estados Unidos y qué medidas sugirió Will Lyon para prevenir futuros abusos en las plataformas de redes sociales?**

- La investigación de NBC News reveló la existencia de cuentas falsas que atraían seguidores y difundían propaganda en la política de EE.UU. Gracias a este descubrimiento, Will Lyon propuso un enfoque basado en conexiones para detectar y detener estas tácticas. Sugirió que tanto los gobiernos como las plataformas de redes sociales implementaran este método para identificar y prevenir la manipulación en línea antes de que pudiera afectar la democracia o la conversación pública.