Étude de cas

NBC News

NBC News analyse des centaines de milliers de trolls russes sur Twitter grâce à Neo4j

SECTEUR
Médias & édition

CAS D'USAGE
Graphes de médias et de réseaux sociaux

OBJECTIF
Dévoiler les réseaux de trolls derrière l'ingérence russe dans les élections de 2016 aux États-Unis

DÉFI
Restaurer et analyser plus de 200 000 tweets

SOLUTION
Réseau de trolls mis à jour en utilisant la base de données de graphes Neo4j

RÉSULTATS
- Mise en lumière de trolls russes se faisant passer pour des citoyens américains, des médias et des groupes politiques locaux
- Identification de modèles de retweets, de hashtags et de pics d'activité pendant les heures de bureau en Russie

Au cours de l'élection de 2016 aux États-Unis, des trolls russes ont infiltré les conversations en ligne. NBC News a voulu mené l'enquête, mais a rencontré deux problèmes : récupérer les tweets supprimés et analyser les données pour identifier des modèles. Ses journalistes ont utilisé Neo4j pour passer au crible des centaines de milliers de tweets et mettre à jour la tactique de ces réseaux de trolls russes.

L'entreprise

NBC News est l'un des principaux médias journalistiques au monde. La chaîne diffuse des émissions emblématiques telles que «NBC Nightly News», «Today» et «Meet the Press». Elle anime la chaîne d'information 24 heures sur 24 MSNBC, plus diverses plateformes numériques d'information. La tentative d'ingérence de la Russie dans la politique américaine a été confirmée en 2016 et une équipe de NBC a cherché à comprendre comment les trolls appuyés par le Kremlin exploitaient Twitter.

Le défi

C'est un fait : des trolls russes sur Twitter ont interféré dans l'élection présidentielle américaine de 2016. Toutefois, il est difficile de déterminer exactement comment ils s'y sont pris à cause de l'opacité de la guerre en ligne, de l'anonymat sur Internet, de la facilité avec laquelle il est possible de se dissimuler derrière des identités fictives et du vaste volume de données des réseaux sociaux.

En novembre 2017, la Commission permanente de la Chambre des représentants des États-Unis sur le renseignement* a publié une liste de 2 752 comptes Twitter en lien avec l'Internet Research Agency, la « ferme de trolls » associée au Kremlin (Twitter a ensuite élargi cette liste à 3 814 comptes). Les agents russes se faisaient passer pour des citoyens américains, des organismes de presse et des groupes politiques afin de diffuser de la désinformation et susciter des divisions.

Au moment de la parution de la liste, Twitter avait suspendu les comptes en question, supprimé les tweets et les profils des utilisateurs. Les journalistes de NBC devaient donc remettre la main sur les tweets effacés des trolls.

Comment ces données pouvaient-elles être retrouvées et analysées ? Comment fonctionnaient les réseaux ? Comment les trolls infiltraient-ils les conversations en ligne d'Américains ordinaires et tentaient-ils d'influencer l'opinion publique ? Ces questions avaient un immense intérêt public, mais sans outils pour récupérer et analyser les données, les réponses restaient évasives.

La stratégie

NBC News s'est adressé à Neo4j, une plateforme de graphes adaptée à l'identification de relations au sein de grands ensembles de données. La première tâche a consisté à récupérer autant de données que possible.

Les enquêteurs ont récupéré des tweets auprès d'Internet Archive – avec leur Wayback Machine (machine à revenir en arrière) - et de groupes indépendants qui ont supervisé Twitter pendant les élections. Ces sources ont produit une base de données de 202 973 tweets provenant de 454 comptes.

* U.S. House of Representatives Permanent Select Committee on Intelligence

Étude de cas

« Grâce à une approche donnant la priorité aux relations pour analyser ce type d'ensemble de données, les gouvernements et les plateformes de médias peuvent détecter et contrer de manière plus proactive ce genre de comportement d'ingérence avant qu'il ne risque de faire dérailler la démocratie ou d'empoisonner les conversations citoyennes. »

– Will Lyon,
Ingénieur en charge des relations avec les développeurs chez Neo4j

Cette base de données ne représentait qu'une portion de toute l'activité, mais s'avérait suffisante pour permettre aux journalistes et analystes de commencer à répondre à cette question majeure : que faisaient ces trolls ?

La solution

Le graphe a montré des relations entre entités telles que tweets, utilisateurs (pour certains, des trolls connus), hashtags, applications sources et liens.

Les algorithmes de graphes ont mesuré la centralité des nœuds en fonction de leurs relations avec les autres entités. Les algorithmes de détection de communautés ont révélé des réseaux d'utilisateurs en interaction fréquente et ont identifié quels étaient les trolls influenceurs, et ceux qui se contentaient de leur faire écho. Dans chaque cluster, PageRank a identifié les comptes les plus influents.

Les journalistes ont commencé à voir le réseau de trolls à l'œuvre. Chaque communauté disposait d'un petit noyau de générateurs de contenu et d'un spectre plus large relayant les tweets. Seuls 25 pourcents des tweets étaient des originaux, le reste consistait en retweets. Les trolls ont mis à profit des hashtags partagés et ont répondu à des comptes populaires afin d'accumuler les followers et de se bâtir une influence.

Les trolls ont laissé beaucoup de traces. Les utilisateurs légitimes de Twitter envoient souvent des tweets à partir de leur téléphone, mais les enquêteurs ont découvert un nombre disproportionné de tweets partants du client Web Twitter. Une grille de lecture horaire y montre des pics de tweets aux heures de bureau en Russie.

Les résultats

Les enquêteurs ont identifié plusieurs sortes de comptes de trolls russes. Certains étaient conçus pour ressembler à ceux d'Américains moyens, comme @LeroyLovesUSA. D'autres imitaient des sites d'actualité, par exemple @OnlineCleveland. Une troisième catégorie se faisait passer pour des organisations politiques, notamment @TEN_ GOP qui se présentait comme le parti républicain du Tennessee.

En fait, tous étaient pilotés par l'Internet Research Agency (Centre de recherche d'internet) en Russie. Neo4j a mis en lumière la façon dont des centaines de faux comptes se coordonnaient en réseaux.

Quelques semaines après la parution de la liste des trolls, NBC et Neo4j ont généré une base de données de plus de 200 000 tweets. NBC a sorti un dossier basé sur les analyses de Neo4j. Il a été mis en lumière comment les trolls russes se faisaient passer pour des Américains, attiraient des centaines de millions de followers et diffusaient de la propagande dans la vie politique américaine.

En réfléchissant à ce que les plateformes de médias et les gouvernements pourraient faire pour éviter de nouveaux abus à l'avenir, Will Lyon, Ingénieur en charge des relations avec les développeurs chez Neo4j, indique « Grâce à une approche donnant la priorité aux relations pour analyser ce type d'ensemble de données, les gouvernements et les plateformes de médias peuvent détecter et contrer de manière plus proactive ce genre de comportement d'ingérence avant qu'il ne risque de faire dérailler la démocratie ou d'empoisonner les conversations citoyennes. »

Ben Popken, Reporter senior de NBC sur les sujets Business a tweeté un message de reconnaissance à l'équipe de Neo4j : « Un grand merci à Neo4j pour nous avoir aidés à compiler et analyser les données effacées de Twitter, à dégager les tendances et à découvrir de nouveaux angles."

Neo4j est le leader de la technologie des bases de données de graphes. Avec le plus grand nombre de déploiements au monde, Neo4j aide des entreprises mondiales - comme Airbus, Michelin, NASA, Crédit Agricole et Volvo Cars – à prédire et identifier la façon dont les personnes, les processus, les lieux et les systèmes sont interconnectés. Grâce à cette approche par les relations, les applications mises au point en utilisant Neo4j relèvent les défis associés aux données connectées, tels que l'analytique et l'intelligence artificielle, la détection de fraude, les recommandations en temps réel et les graphes de connaissance. Pour en savoir plus, merci de consulter Neo4j.com et @Neo4jFr.

Des questions sur Neo4j ?

Contactez-nous :
info@neo4j.com
neo4j.com/contact-us