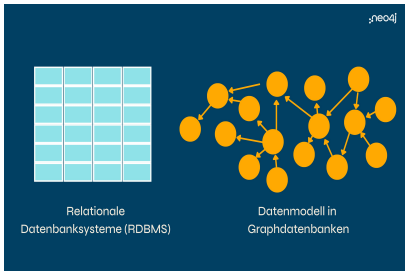
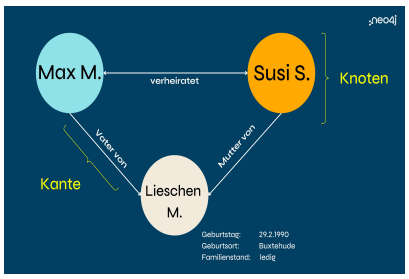
Im Einsatz – im Thema.
POLIZEI PRAXIS

DATENANALYSE MIT GRAPHDATENBANKEN





Die Polizei steht vor einem Datenproblem: Sie muss auf der einen Seite Unmengen an Informationen in kürzester Zeit sichten und verarbeiten. Auf der anderen Seite gilt es, strenge Datenschutz-Vorgaben zu erfüllen. Graphdatenbanken bieten hier neue Wege beim Umgang mit Daten.

Die polizeiliche Datenanalyse hat es in Deutschland nicht gerade einfach. Erst 2023 schränkt das Bundesverfassungsgericht die automatisierte Datenauswertung (Data-Mining) zur Vorbeugung von Straftaten ein. Das Urteil ist allerdings keine grundsätzliche Absage an intelligente Datensysteme. Vielmehr legt es detailreich die gesetzlichen Anforderungen dar, die mit Blick auf KI weiter an Bedeutung gewinnen dürften.

Fakt ist: Die Datenanalyse ist ein Muss, will die Polizei im digitalen Zeitalter einen Wissensvorsprung behalten. In den Asservatenkammern türmen sich Speichermedien, Festplatten, Smartphones, Notebooks und Tablets. Kriminalitätsfelder verlagern sich mehr und mehr ins Netz, wo sie eine lange Spur an digitalen Fußabdrücken hinterlassen. Die Menge an zu verarbeitenden Informationen wächst ohne Ende.

Manuell lässt sich diese Informationsflut nicht bewältigen und in der analogen Welt erprobte Prozesse und

Strukturen stoßen längst an ihre Grenzen. Das betrifft auch Datenbanken. Als Datenfundament für intelligente Anwendungen müssen sie technische Voraussetzungen rund um Echtzeit-Performance, Skalierbarkeit und Flexibilität erfüllen. Ermittler brauchen zudem einen einfachen Zugang zu Informationen. Gleichzeitig heißt es, die Daten unter Vorgabe des Datenschutzes zu verwalten und zu schützen. Und schließlich stellen Transparenz und Nachvollziehbarkeit sicher, dass die Datenanalyse der Polizei nicht zur Black-Box mutiert.

Beziehungen im Fokus

Graphdatenbanken erfüllen diese Kriterien. Sie kommen dort zum Einsatz, wo viele und heterogene Daten miteinander verknüpft sind. Die Verknüpfungen bzw. Beziehungen stehen dabei im Zentrum des Datenmodells (siehe Infobox). Das ermöglicht es den Graphen, Netzwerke in ihrer ganzen Komplexität abzubilden, Zusammenhänge und Muster offenzulegen und Informationen in einen Kontext zu stellen. Es entsteht ein realitätsnahes Bild der Welt, das nachvollziehbar und erklärbar bleibt.

Graphdatenbanken brechen mit der langen Tradition von relationalen Datenbanken. Diese modellieren Daten in Tabellen und Spalten und müssen für Verknüpfungen zeit- und rechenaufwändige Operationen (Joins) durchführen. Abfragen im Graphen hingegen folgen lediglich den inhärenten Verbindungen innerhalb der Daten und sind mit mehreren Millionen Sprüngen (Hops) pro Sekunde im Durchschnitt 1.000-mal schneller. Sie sind zudem hoch skalierbar. Das Knoten-Kanten-Prinzip ermöglicht es, neue Informationen einfach in den Graphen zu integrieren. Gemeinsam mit Graph Analytics und Graph Data Science gehören Graphdatenbanken zudem zu den Basistechnologien für Machine Learning (ML) und künstlicher Intelligenz (KI).

Graphen in der Praxis
Graphen erlauben es Anwendern ohne große technisches Vorwissen, tief in Datensätze einzusteigen und neue Verbindungen aufzuspüren, die so vorher nicht bekannt waren. Das Navigieren im Graphen gleicht dabei der Arbeit von Ermittlern: Man geht Hinweisen nach, wirft einen Blick auf den Umkreis von Personen, Dingen, Orten oder Ereignisse (Kontextinformationen) und kommt so zu neuen Erkenntnissen. Für die Polizeiarbeit ergeben sich damit unterschiedlichste Anwendungsszenarien:

1. Kontext für die Betrugsaufdeckung

Der Versicherer Zurich Schweiz nutzt Graphdatenbanken, um gemeldete Schadenfälle genauer unter die Lupe zu nehmen und im Kontext von allen verfügbaren Informationen zu prüfen. Der Versicherer prüfte jeden Versicherungsschaden automatisiert nach definierten Kriterien. Das Problem: Die Field Investigatoren kamen mit der manuellen Nachprüfung der Fälle nicht mehr hinterher. Warum es zu einem Alert kam oder wie sich der Risk-Score zusammensetzte, blieb zudem oft im Dunkeln.

Um Kontext zu schaffen, kombinierte Zurich Schweiz daher sein regelbasiertes System mit der Graphdatenbank Neo4j. Neue Informationen sowie Querreferenzen zu Bankkonten, Adressen, Kundendaten und Policen werden nun automatisch in einer Ansicht verknüpft. So kann das Ermittlerteam eine Triage vornehmen, die Zusammensetzung des Risk-Score einsehen und Betrugsversuche schnell und gezielt aufdecken.

2. Echtzeit-Analysen bei Banken

Bei der Betrugsaufdeckung geht es nicht nur um Kontextinformationen, sondern auch um Schnelligkeit. Banken bleibt oft nur ein kleines Zeitfenster, um Betrugsversuche zu erkennen und z. B. Kreditkarten und Konten zu sperren. Der IT-Dienstleister TODO1 entwickelte für Banken in Südamerika mit Neo4j eine graphbasierte Lösung, die solche Echtzeit-Analysen ermöglicht.

Der Kontinent ist eine der am schnellsten wachsenden Regionen für digitale Bankgeschäfte. Zahlungen werden in der Regel in Echtzeit verarbeitet. Das heißt, auch Betrugsversuche müssen in Echtzeit erkannt werden. Die Graph-Lösung analysiert pro Sekunde Hunderte von Transaktionen auf verdächtige Hinweise, was vor allem während der Stoßzeiten (z. B. Feiertage, Zahltage) entscheidend zur Betrugsprävention beiträgt. Unterm Schnitt verbesserte sich die Aufdeckungsrate um 200%.

3. Geldwäsche: Follow the Money

Große globale Banken verarbeiten täglich Millionen von Transaktionen in verschiedenen Sprachen und Transaktionsformaten. Kriminelle Banden nutzen diese Komplexität zur Geldwäsche aus und verschieben illegale Gelder über ein weit verzweigtes Netzwerk aus echten und falschen Identitäten, Konten und Briefkastenfirmen. Die UN schätzt, dass weltweit jährlich über 2 Billionen US-Dollar gewaschen werden – das sind 5% des weltweiten BIP.

Als Risikoindustrien gelten u. a. der Immobiliensektor und die Glücksspielindustrie. Um den dortigen Dokumentations- und Prüfungsplichten nachzukommen, entwickelte der Systemanbieter Kerberos auf Basis der Neo4j-Graphdatenbank ein Compliance-Management-System. Die Dokumentation erfolgt auch auf Knopfdruck und kann über eine Schnittstelle an die Meldestelle des FIU weitergeleitet werden. Jede Momentaufnahme des Graphen lässt sich speichern. So ist nachvollziehbar belegbar, welche Information zu einem bestimmten Zeitpunkt vorlag.

Smarte Lösungen für Anti-Money-Laundering (AML) sind dringend notwendig. Das International Consortium of Investigative Journalists (ICIJ) deckte in den letzten Jahren immer wieder Offshore-Steueroasen, Sanktionsverstöße und die Geheimgeschäfte zahlreicher Funktionäre, Politiker, Milliardäre und Berühmtheiten auf (u. a. Paradise Papers, FinCEN Files). Allein durch die Panama Papers konnten Behörden weltweit mehr als eine Milliarde Euro an Steuernachforderungen einnehmen. In Deutschland waren es mehr als 160 Millionen Euro. Um die diversen Datenleaks zu analysieren und der Spur des Geldes zu folgen, arbeiteten die Journalisten auch hier mit der Graphdatenbank von Neo4j.

4. Grenzschutz und -sicherheit

Datenanalysen helfen nicht nur im Finanzsektor. Auch Einwanderungsbehörden nutzen Graphdatenbanken weltweit bei Grenz- und Sicherheitskontrollen und bei der Echtzeit-Überprüfung von Dokumenten und Pässen. Ziel ist es, Einreisende ohne Arbeits- oder Aufenthaltserlaubnis zu erkennen und wiederholte Antragsteller von Erstanträgen zu unterscheiden. Dazu nutzen Graphdatenbanken die sog. Fuzzy-Logik. Diese erkennt z. B. verschiedene Schreibweisen eines Namens und ordnet diese eindeutig einer Person zu. Grenzbeamte können so sekundenschnell feststellen, ob es sich tatsächlich um die gleiche Person handelt.

In einem anderen Fall nutzt die Einwanderungsbehörde eines G8-Staates die Graph-Visualisierung, um Personendaten bei der Einreise mit Daten aus Strafregistern in Echtzeit abzugleichen. Liegt z. B. ein Haftbefehl gegen den Einreisenden vor? Oder steht er in Verbindung mit behördlich bekannten Schleusern oder Schmugglern? Erfolgt eine Anfrage weiß der zuständige Grenzposten innerhalb weniger Augenblicke, ob er der Person die Einreise gestatten oder verwehren kann.

5. Infrastrukturen überwachen: Kommunikation & Verkehr

Graphdatenbanken ermöglichen es, sicher und schnell in Netzwerken zu navigieren – egal ob es sich dabei um IT-Netzwerke, soziale Netzwerke oder um die öffentliche Infrastruktur handelt. Die Management- und Technologieberatung Sopra Steria entwickelte beispielsweise basierend auf Neo4j ein Intelligentes Netzwerk-Analysetool (INA), über das Netzwerkbetreiber eine Live-Ansicht auf das deutschlandweite Telekommunikationsnetzwerk erhalten. Die detailgetreue Landkarte zeigt die schnellste Route zwischen zwei Stationen, deckt Störungen auf und kann Peaks im Datenverkehr vorhersagen.

Für den Polizeibetrieb ergeben sich daraus ähnliche Anwendungsszenarien – von der computergestützten Disposition (CAD) bis hin zur Überwachung des öffentlichen Nahverkehrs und Verkehrsleitsystemen. So koordiniert die Dachorganisation Transport for London (TfL) seit 2001 das Verkehrssystem der britischen Hauptstadt. Um Störungen im Straßenverkehr schneller zu erkennen und lange und kostspielige Staus zu reduzieren, entwickelte die Behörde mit Hilfe der Graphtechnologie von Neo4j einen Digitalen Zwilling des gesamten Londoner Verkehrsnetzes – sowohl U-Bahn, Bahn als auch Oberflächenverkehr.

6. Im Netz gegen Cyberkriminelle

Cyberkriminelle sind oft die ersten, die neue Technologien missbrauchen. Das gilt auch für generative KI und

Large Language Models (LLMs). Laut BSI bergen die Sprachmodelle aber auch Chancen, um Spam oder Phishing-Mails sowie Fake News oder Hate Speech auf Social-Media-Plattformen zu identifizieren. In Graphdatenbanken lassen sich entsprechende Informationen zusammenführen und mittels Graph-Algorithmen analysieren. Überhaupt kann die Verknüpfung unterschiedlicher Datensätze (z. B. Telekommunikations- und Internetdaten, Webservern und IP-Adressen) der Polizei und dem Staatsschutz wichtige Informationen liefern, um Plattformen für Kinderpornografie, Marktplätze im Darknet oder terroristische Absprachen im Netz aufzuspüren.

Um IT-Netzwerke vor Cyberangriffen zu schützen, können Sicherheitsteam die IT-Infrastruktur virtuell in einem Graph nachmodellieren und simulieren. Der digitale Zwilling gibt eine ganzheitliche Sicht auf alle Risiken im Kontext aller Informationen. So lässt sich z. B. der Modus Operandi (MO) einer cyberkriminelle Gruppe als Threat Intelligence im Graphen hinterlegen. Graph-Algorithmen suchen nach entsprechenden Mustern oder erkennen Auffälligkeiten (wie ein unerklärlicher Anstieg des Netzwerkverkehrs), um eventuelle Attacken abzuwehren.

Die Digitalisierung stellt Polizeibehörden vor erhebliche Herausforderungen. Sie bietet jedoch auch erhebliches Potential, ihre Leistungsfähigkeit zu steigern und in Zeiten von Fachkräftemangel und Budgetknappheit schneller und schlagkräftiger zu agieren. Egal welche intelligenten Anwendungen letztendlich in der Polizeiarbeit zum Einsatz kommen, ohne saubere, relevante und sichere Daten geht es nicht. Es lohnt sich daher, am Datenfundament zu arbeiten und mit Hilfe von Graphtechnologie das Datenproblem langfristig anzugehen.

Das Graph-Datenmodell: Knoten und Kanten
Ein Graph bezeichnet eine abstrakte Struktur, die eine Menge von Objekten sowie die bestehenden Verbindungen zwischen diesen Objekten repräsentiert. In der Mathematik werden Objekte als Knoten und Verbindungen als Kanten dargestellt. Sowohl Knoten als auch Kanten können Eigenschaften, sogenannte Properties besitzen (Labeled-Property-Graph).

Was abstrakt klingt, ist tatsächlich sehr intuitiv und findet sich in zahlreichen Strukturen wieder. In einem Familienstammbaums beispielsweise werden Personen als Kreise (Knoten) dargestellt, die über Linien (Kanten) miteinander verbunden sind. Jeder Kreis kann mit einem Namen, jede Linie mit einem Verwandtschaftsgrad (z. B. „verheiratet“) versehen werden.

Text: Heiko Schönfelder, Country Manager Germany, Neo4j, Grafiken/Bilder: Neo4j

Alle Artikel dieser Kategorie

Folgen Sie uns!