

NAPREDNA VIZUALNA ANALITIKA GRAFOVSKIH PODATKOVNIH BAZ

Lea Roj¹, Štefan Kohek¹, David Jesenko¹, Aleksander Pur², Domen Mongus¹, Niko Lukač¹

¹ Fakulteta za elektrotehniko, računalništvo in informatiko, Koroška cesta 46

² Ministrstvo za notranje zadeve, Policija

l.roj@um.si, stefan.kohek@um.si, david.jesenko@um.si, pur.aleksander@gmail.com, domen.mongus@um.si, niko.lukac@um.si

V zadnjem desetletju smo soočeni z eksponentno rastjo heterogenih digitalnih podatkov, kjer se lahko hitro izgubimo v preveliki količini informacij. Zaradi tega se spopadamo z izzivi pri razumevanju skritih informacij, ki so vidne šele z dobro vizualno analizo. Posledično narašča potreba po učinkovitih analitičnih orodjih, ki se izkazujejo kot učinkovit pristop k identifikaciji povezav med entitetami, kar omogoča vpogled, ki ga je težko doseči z uporabo tradicionalnih metod.

V analitični platformi STALITA smo razvili več inovativnih funkcij z namenom izboljšanja analitičnih procesov in uporabniške izkušnje pri delu z obsežnimi grafovskimi podatkovnimi bazami, ki temeljijo na tehnologiji Neo4j. Grafi nazorno prikazujejo povezave med podatki, pri čemer se je uporaba grafovskih podatkovne baze Neo4j izkazala za še posebej primerno. To velja zlasti, ko nas zanimajo povezave med podatki in iskanje skritih vzorcev in ne zgolj njihova agregacija. Vendar, so pri tem vizualizacije velikih količin podatkov lahko hitro nepregledne. Zato smo razvili raznovrstne pristope, ki izboljšajo iskanje koristnih informacij v velikih grafih, kar je predstavljeno v nadaljevanju.

Razporeditev vozlišč v grafih je ključnega pomena tako za razumevanje njihove širše slike kot iskanje skritih povezav. V STALITI smo za njihovo optimalno postavitev uporabili algoritem za izris grafov na podlagi sil (angl. force-directed graphs), ki vozlišča in povezave med njimi upodobi na način, ki posnema medsebojno delovanje fizikalnih sil. V okviru STALITE podpiramo tudi inteligentno agregacijo povezav, ki omogoča pregledno predstavitev kompleksnejših grafov z večjim številom povezav med enakimi vozlišči, tako da ustvari izhodni graf na podlagi uporabniško nastavljenih pravil agregacije.

V veliki količini informacij je tudi težko najti določene atribute, ki jih potrebujemo pri razumevanju podatkov. Zato smo v okviru STALITE podprli napredno poudarjanje vozlišč in povezav glede na atribute, ki jih le-ta vsebujejo. To omogoča uporabnikom, da hitro filtrirajo in zožijo obseg podatkov, s čimer se osredotočijo le na tiste dele grafa, ki so relevantni za trenutno analizo.

Vizualizacije podatkov v predstavitvi grafov omogočajo odkrivanje skritih in zapletenih vzorcev, ne omogočajo pa naprednega prikaza statistične porazdelitve, trendov, primerjave med izračunanimi vrednostmi v podatkih ipd. Zato smo v STALITI grafe nadgradili tudi s sekundarno vizualizacijo podatkov na podlagi grafikonov. Uporabljamo jih, kadar želimo kvantitativno oceniti podatke in jih razvrstiti glede na različne kategorije ali časovna obdobja. Na primer, radialni grafikoni lahko učinkovito prikažejo medsebojna razmerja merjenih entitet, časovni grafikoni pa omogočajo sledenje spremembam skozi čas. Histogrami so ključni za oceno porazdelitve in frekvence pojavljanja določenih vrednosti, medtem ko grafikoni raztrosa razkrivajo korelacije in vzorce razpršenosti med dvema ali več spremenljivkami. Tako smo v STALITI vizualizacijo grafov obogatili z raznovrstnimi grafikoni. Pri tem smo omogočili še prilagoditve vizualizacij specifičnim analitičnim potrebam, kjer lahko uporabnik filtrira po numeričnih vrednostih ali kategorijah.

S temi naprednimi funkcionalnostmi STALITA podpira ustvarjanje vizualizacij, ki so prilagojene specifičnim analitičnim potrebam in preferencam uporabnika, kar olajša interpretacijo kompleksnih podatkov, boljše razumevanje skritih vzorcev ter korelacij med podatki. Uporaba metod in orodij je omogočila bolj informirane odločitve in prispevala k širjenju znanja v različnih aplikacijskih domenah.

Analitična platforma je izjemno učinkovita in prilagodljiva, kar omogoča njeno široko uporabo tako v zasebnem kot javnem sektorju. V zasebnem sektorju bi jo lahko uporabili v finančnih institucijah. Te bi z napredno analizo finančnih transakcij učinkovito odkrile in preprečile vzorce finančnih prevar ter optimizirale strategije za obvladovanje tveganj. Tako bi izboljšali varnost in posledično povečali zaupanje strank. V javnem sektorju pa bi lahko platformo uporabili za zagotavljanje nacionalne varnosti in učinkovitejše sledenje kriminalnim aktivnostim. Služila bi kot orodje za izdelavo in analizo kompleksnih omrežij med osumljenci, kar bi omogočilo oblastem boljše razumevanje mreže kriminalnih dejavnosti. Ta pristop ne bi samo pospešil preiskovanja, temveč bi tudi pripomogel k preprečevanju prihodnjih kriminalnih dejanj z izboljšanjem predvidevanja in prepoznavanja groženj. Analitično platformo bi lahko uporabili tudi v različnih praktičnih aplikacijah, kjer so podatki prikazani z grafi. Primer tega sta analiza citiranja v revijah ter prikaz trendov v povezavi z onesnaževanjem okolja.

Nadaljnji razvoj na področju vizualne analitike nad grafi, predvsem z rešitvami umetne inteligence ter generativnimi modeli, bo prinesel še večjo učinkovitost in uporabnost v obravnavi grafovskih podatkov, spodbujal inovacije ter omogočil nove načine raziskovanja in razložljivosti podatkov.

Ključne besede: Grafovske podatkovne baze; Neo4j; Vizualna analitika; Interaktivna vizualizacija; Grafikoni

ADVANCED VISUAL ANALYTICS OF GRAPH DATABASES

Due to the numerous resources available, we are occupied with a flood of information. Given the abundance of digital data, we face challenges in understanding and analysing less obvious patterns, which become obvious only with advanced visual analytics. This is particularly evident in graph databases (such as Neo4j), where there are many complex relationships present. Visual analytics is proving to be a powerful approach to identifying these connections in datasets, providing insights that are difficult to gain using traditional methods. Using a combination of interactive visualisations and advanced analytics techniques, visual analytics enables users to explore, interpret, and extract insights from graph data, ultimately improving understanding and insight in various fields. In STALITA, we implemented a series of functions that are convenient for the visualization of graphs, highlighting crucial information, and providing a user-friendly experience while working on large graph databases.

Keywords: Graph databases; Neo4j; Visual analytics; Interactive visualization; Charts