Introducere

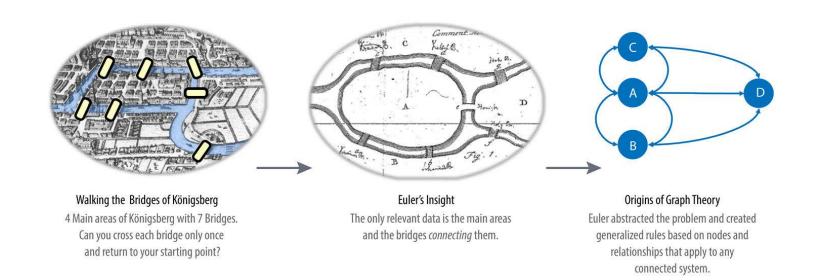
- Ce sunt grafurile?
- Ce este analiza si algoritmica grafurilor?
- Network Science Ştiinta Reţelelor
- Algoritmica grafurilor
- Studiu de caz
- Procesarea grafurilor, baze de date, interogari si algoritmi
- OLTP şi OLAP
- Hybrid Transactional and Analytical Processing (HTAP)
- Algoritmica grafurilor. Motivație
- Power Law (Lege de scalare)
- Explicaţii
- Cazuri de utilizare. Graph Analytics
- Tipuri de probleme în care este utilizată algoritmica grafurilor
- Concluzie

Introducere în analiza grafurilor si algoritmica grafurilor

- Cele mai noi provocari de date de astazi se concentreaza în jurul relatiilor, nu doar în tabelarea datelor discrete ("Algorithm Design Manual", Skiena, Ed. Springer)
- Tehnologiile bazate pe grafuri si analizele ofera instrumente puternice pentru datele conectate, care sunt utilizate în cercetare, initiative sociale si solutii de afaceri, cum ar fi:
 - Modelarea mediilor dinamice de la pietele financiare la serviciile IT
 - Prognozarea raspândirii epidemiilor, precum si a întârzierilor si întreruperilor serviciilor
 - Gasirea caracteristicilor predictive pentru machine learning în vederea combaterii infractiunilor financiare
 - Descoperirea tiparelor pentru experiente si recomandari personalizate
- Pe masura ce datele devin din ce în ce mai interconectate si sistemele din ce în ce mai sofisticate, este esential sa folosim relatiile bogate si în evolutie din cadrul datelor noastre.
- Urmeaza o scurta recapitulare despre originea grafurilor înainte de a introduce algoritmica grafurilor.
- Se va prezenta natura datelor moderne în sine si modul în care informatiile continute în conexiuni sunt mult mai sofisticate decât ceea ce putem descoperi cu metode statistice de baza.
- Cateva cazurilor de utilizare în care pot fi utilizati algoritmii grafurilor vor fi prezentate.

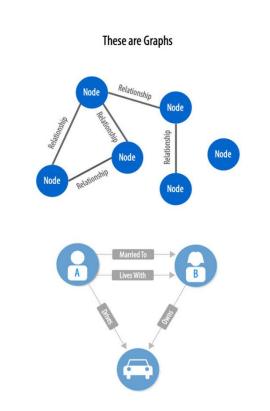
Ce sunt grafurile?

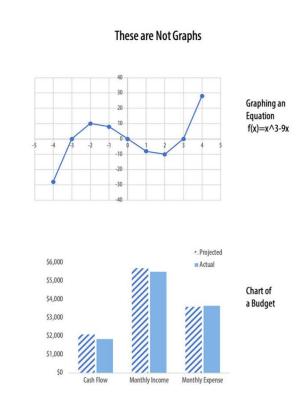
- Grafurile au o istorie care dateaza din 1736, când Leonhard Euler a rezolvat problema "celor sapte poduri din Königsberg".
- Problema a fost daca este posibil sa se viziteze toate cele patru zone ale unui oras conectate prin 7 poduri, în timp ce traverseaza fiecare pod o singura data.
- Folosind faptul ca numai conexiunile în sine erau relevante, Euler a pus bazele teoriei grafurilor si matematicii sale.
- Figura prezinta etapele lui Euler cu una dintre schitele sale originale, din lucrarea "Solutio problematis ad geome Triam situs pertinentis".



Ce sunt grafurile?

- În timp ce grafurilele îsi au originea în matematica, ele sunt, de asemenea, o modalitate pragmatica si de înalta fidelitate de modelare si analiza a datelor.
- Obiectele care alcatuiesc un graf se numesc noduri sau noduri, iar legaturile dintre ele sunt cunoscute sub numele de relatii, legaturi sau muchii.
- Folosim termenii noduri si relatii. Va gânditi la noduri ca substantive în propozitii si la relatii ca verbe care dau context nodurilor.
- Grafurile nu au nimic de-a face cu graficele bazate pe ecuatii sau diagrame (Figura).

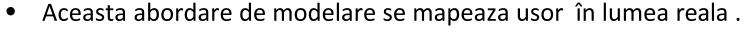




Ce sunt grafurile?

 Privind graful persoanei din Figura 1-2, se pot construi mai multe propozitii care o descriu.

 De exemplu, persoana A locuieste cu persoana B care detine o masina, iar persoana A conduce o masina pe care o detine persoana B.



- Acest lucru ajuta la modelarea si analiza datelor.
- Dar modelarea grafurilor este doar o parte din activitate.
- De asemenea, ar fi nevoie sa le procesam pentru a dezvalui o perspectiva care nu este imediat evidenta.
- Acesta este domeniul algoritmica grafurilor.

Ce este analiza si algoritmica grafurilor?

- Algoritmii grafurilor sunt un subset de instrumente pentru analiza grafurilor.
- Analiza grafurilor este utilizarea oricarei abordari bazate pe grafuri pentru a analiza datele conectate.
- Exista diverse metode pe care le-am putea folosi:
 - sa interogam datele grafurilor,
 - sa utilizam statistici de baza,
 - sa exploram vizual grafurile
 - sa încorporam grafurile în sarcinile de învatare automata.
- Interogarea bazata pe modelul grafurilor este adesea utilizata pentru analiza datelor locale, în timp ce algoritmii computationali bazati pe grafuri se refera de obicei la analize mai globale si iterative.
- Desi exista o suprapunere în modul în care aceste tipuri de analize pot fi utilizate, folosim termenul **algoritmica grafurilor** pentru a ne referi la acestea din urma.

Ce este analiza si algoritmica grafurilor?

- Algoritmica grafurilor ofera una dintre cele mai puternice abordari pentru analiza datelor conectate, deoarece calculele lor matematice sunt construite special pentru a opera asupra relatiilor.
- Acestea descriu pasii care trebuie urmati pentru a procesa un graf pentru a-i descoperi calitatile generale sau cantitatile specifice.
- Bazându-se pe matematica teoriei grafurilor, algoritmii grafurilor folosesc relațiile dintre noduri pentru a deduce organizarea si dinamica sistemelor complexe.
- Oamenii de stiinta inclusiv din Network Science, folosesc acesti algoritmi pentru
 - a descoperi informatii ascunse,
 - a testa ipoteze si
 - a face predictii despre comportament.

(ex. aplicatii in astronomie, comert, marketing, medicina)

Network Science – Știinta Rețelelor

- Network Science este un domeniu academic puternic înradacinat în teoria grafurilor, care se ocupa cu modele matematice ale relatiilor dintre obiecte.
- Oamenii de stinta din Network Science se bazeaza pe algoritmica grafurilor si sistemelor de gestionare a bazelor de date din cauza dimensiunii, conectivitații și complexitații datelor lor.
- Exista multe resurse pentru complexitate și știința rețelelor. Referințe pe care sa le explorați.
- <u>Network Science</u>, by A.L. Barabasi, ebook introductiv

http://networksciencebook.com/

- 'Complexity Explorer' ofera de exemplu cursuri online
- 'The New England Complex Systems Institute 'include resurse si lucrari
- Algorithms: R. Sedgewick, K. Wayne, MIT press (http://algs4.cs.princeton.edu/)
- Menczer et al. (2020) A First Course in Network Science. Cambridge University Press https://github.com/CambridgeUniversityPress/FirstCourseNetworkScience

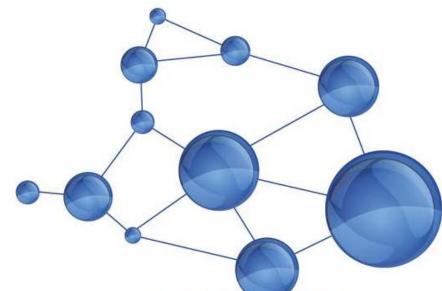
Free services to run Jupyter notebooks in the cloud:

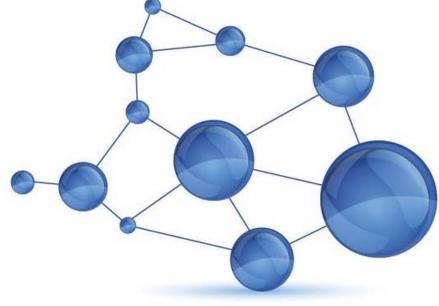
Google Colab;
 <u>Binder</u>;
 <u>Kaggle Kernels</u>;
 <u>Azure Notebooks</u>;
 <u>Datalore</u>;
 <u>Gryd</u>

Free online: Samples, (seturi de date) Tutorials:

NOUA ȘTIINȚĂ A REȚELELOR

Despre cum orice lucru este conectat cu oricare altul si ce reprezintă asta pentru afaceri, știință și viața cotidiană





barabasi albert graph

🏫 > Reference > Graph generators > barabasi albert graph

barabasi albert graph{n, m, seed=None, init Returns a random graph using Barabasi-Albert preferent

- Fizica si inginerie Universitatea Bucuresti
- Doctorat Boston University

Details:

Section Navigation

Introduction Graph types

Algorithms

Functions

"Linked" - Noua stiinta a retelelor

http://www.barabasilab.com/

Network Science, A. Barabasi,

http://barabasi.com/networksciencebook/



Section Navigation

introduction

Graph types

Algorithms

extended_barabasi_albert_graph

extended_barabasi_albert_graph(n, m, p, q, seed=None)

Returns an extended Barabási-Albert model graph.

A > Reference > Graph generators > extended basebasi albert graph

Barabási's latest book is The Formula (Little Brown, 2018). He is the author of "Network Science" (Cambridge, 2016).

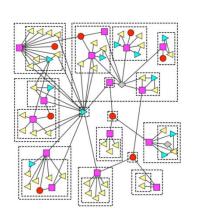
₹

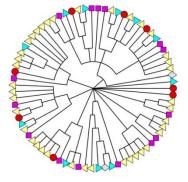
[source]

"Linked" (Penguin, 2002), and "Bursts:" (Dutton, 2010) He co-edited Network Medicine (Harvard, 2017) and "The Structure and Dynamics of Networks" (Princeton, 2005). His books have been translated in over twenty languages.

Algoritmica grafurilor

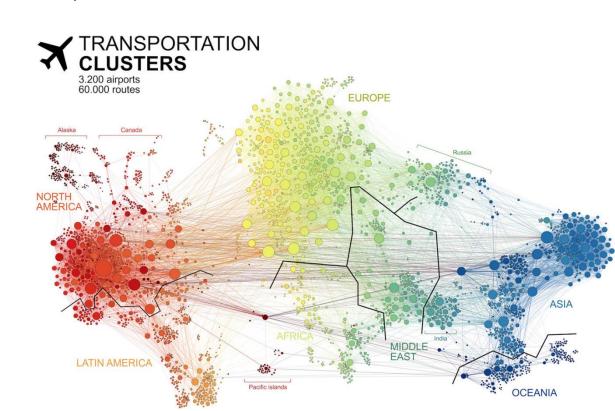
- Algoritmica grafurilor are un potential larg raspândit, de la prevenirea fraudei si optimizarea rutarii apelurilor pâna la prezicerea raspândirii gripei.
- Exemplu: sa punctam anumite noduri care ar putea corespunde conditiilor de suprasarcina dintr-un sistem de alimentare, sau sa descoperim grupari în graf care corespund congestiei într-un sistem de transport.
- Grafurile ajuta si la descoperirea modului în care interactiunile si dinamicile foarte mici duc la mutatii globale.
- Ele leaga împreuna scarile micro si macro, reprezentând exact ce lucruri interactioneaza în cadrul structurilor globale.
- Aceste asociatii sunt folosite pentru a prognoza comportamentul si pentru a determina legaturile lipsa.
- Figura este o retea trofica de interactiuni între speciile care folosesc pașunatul de verdeața care a folosit analiza grafului pentru a evalua organizarea ierarhica si interactiunile dintre specii si apoi pentru a prezice relatiile lipsa (Clauset et al. "Hierarchical Structure and the Prediction of Missing Links in Network")





Studiu de caz

- In 2010, sistemele de transport aerian din SUA au experimentat 2 evenimente grave care implica mai multe aeroporturi congestionate, ulterior studiate folosind analiza grafurilor.
- S-a folosit algoritmica grafurilor pentru a confirma evenimentele ca parte a întârzierilor sistematice în cascada si pentru a folosi aceste informatii pentru sfaturi corective (Eguíluz et. al. "Systemic Delay Propagation in the US Airport Network".)
- Figura ilustreaza reteaua care sta la baza transportului aerian (Grandjean "Connected World: Untangling the Air Traffic Net".)
- Aceasta ilustratie arata în mod clar structura extrem de conectata a clusterelor de transport aerian.
- Multe sisteme de transport prezinta o distributie concentrata a legaturilor cu modele clare (hub-and-spoke) care influenteaza întârzierile.



Procesarea grafurilor, baze de date, interogari si algoritmi

- Procesarea grafurilor include metodele de efectuare a sarcinilor grafului.
- Majoritatea interogarilor grafului iau în considerare anumite parti ale grafului (ex. un nod de pornire), iar lucrarea este de obicei concentrata în subgraful înconjurator.
- Numim acest tip de **graf de lucru local** si implica interogarea declarativa a structurii unui graf, asa cum se explica în cartea **Graph Databases** (Robinson et al. Editura O'Reilly).
- Acest tip de procesare locala a grafului este adesea utilizat pentru tranzactii în timp real si interogari bazate pe modele
- Când vorbim despre algoritmica grafurilor, cautam de obicei modele si structuri globale.
- Intrarea algoritmului este de obicei întregul graf, iar iesirea poate fi un graf îmbogatit sau o valoare agregata, cum ar fi un scor.
- Clasificam o astfel de procesare ca graf global si implica procesarea structurii unui graf folosind algoritmi computationali (adesea iterativ).
- Aceasta abordare arata natura generala a unei rețele prin conexiunile sale.
- Organizațiile tind sa utilizeze algoritmi grafurilor pentru a modela sistemele și a prezice comportamentul pe baza modului în care lucrurile se disemineaza, a componentelor importante, a identificarii grupului și a robusteții generale a sistemului.

Procesarea grafurilor, baze de date, interogari si algoritmi

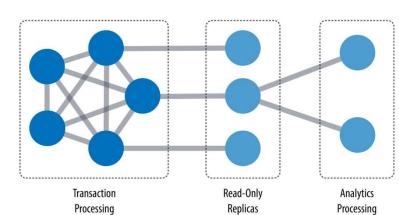
- Este posibil sa existe unele suprapuneri ale definitilor uneori se poate folosi procesarea unui algoritm pentru a raspunde la o riterogare locala sau invers -dar operatiile intregului graf sunt procesate de algoritmi computationali, iar operațiile subgrafului sunt interogate în baze de date.
- Tradițional, procesarea și analiza tranzacțiilor au fost separate
- Aceasta a fost o divizare nefireasca bazata pe limitarile tehnologice.
- Analiza grafurilor duce la tranzacții mai inteligente, ceea ce creeaza noi date și
 oportunitați pentru analize suplimentare.
- Mai recent, a existat o tendința de integrare a acestor "silos" (silozuri) pentru luarea mai multor decizii în timp real.

OLTP și **OLAP**

- Operațiunile de procesare a tranzacțiilor online (OLTP) sunt de obicei activitați
 scurte, cum ar fi rezervarea unui bilet, creditarea unui cont, rezervarea unei
 vânzari și așa mai departe.
- OLTP implica procesarea voluminoasa a interogarilor cu latența scazuta și integritate ridicata a datelor.
- Deși OLTP poate implica doar un numar mic de înregistrari per tranzacție, sistemele proceseaza multe tranzacții simultan.
- **Procesarea analitica online (OLAP)** faciliteaza interogari și analize mai complexe asupra datelor istorice.
- Aceste analize pot include mai multe surse de date, formate și tipuri.
- Detectarea tendințelor, efectuarea scenariilor "ce-ar fi daca", realizarea de predicții și descoperirea modelelor structurale sunt cazuri tipice de utilizare OLAP.
- În comparație cu OLTP, sistemele OLAP proceseaza tranzacții mai puține, dar mai îndelungate, în multe înregistrari.
- Sistemele OLAP sunt orientate spre o citire mai rapida, fara a se aștepta actualizari tranzacționale gasite în OLTP, iar operarea orientata pe loturi este obținuita.

OLTP și **OLAP**

- Recent, linia dintre OLTP şi OLAP a început sa se estompeze.
- Aplicațiile moderne cu volum mare de date combina acum operațiunile tranzacționale în timp real cu analizele.
- Aceasta fuziune a procesarii a fost stimulata de mai multe progrese în software, cum ar fi gestionarea mai scalabila a tranzacțiilor si procesarea incrementala a fluxului, precum si de hardware-ul cu costuri mai mici, cu memorie mare.
- Reunirea analizelor si tranzacţiilor permite analiza continua ca parte naturala a operaţiunilor obișnuite.
- Pe masura ce sunt colectate date de la mașini POS (POS), sisteme de producție sau dispozitive Internet of Things (IoT) analizele accepta acum capacitatea de a face recomandari și decizii în timp real în timpul procesarii.
- Aceasta tendința a fost observata în urma cu câțiva ani, iar termenii care descriu aceasta fuziune includ translitica și procesarea tranzactionala si analitica hibrida (HTAP).
- Figura ilustreaza modul în care replicile doar în citire pot fi utilizate pentru a reuni aceste tipuri diferite de procesare.



Hybrid Transactional and Analytical Processing (HTAP)

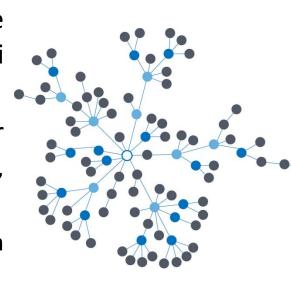
- HTAP ar putea redefini modul în care sunt executate unele procese de afaceri, deoarece analiza avansata în timp real (ex. planificarea, prognoza si analiza de tip "ce-ar fi daca") devine o parte integranta a procesului în sine, mai degraba decât o activitate separata efectuata ulterior.
- Acest lucru ar permite noi forme de proces decizional în timp real, bazat pe afaceri.
- În cele din urma, HTAP va deveni o arhitectura cheie pentru operațiunile inteligente de afaceri.
- Pe masura ce OLTP si OLAP devin mai integrate si încep sa accepte funcționalitațile oferite anterior într-un singur siloz, nu mai este necesar sa utilizam produse sau sisteme de date diferite pentru aceste fluxuri de lucru ne putem simplifica arhitectura utilizând aceeași platforma pentru ambele.
- Aceasta înseamna ca interogarile noastre analitice pot profita de date în timp real și putem eficientiza procesul iterativ de analiza.

Algoritmica grafurilor. Motivație

- Algoritmica grafurilor este utilizată pentru a ajuta la înțelegerea datelor conectate.
- Vedem relații în cadrul sistemelor din lumea reala, de la interacțiunile proteice la rețelele sociale, de la sistemele de comunicații la rețelele electrice și de la vânzarea cu amanuntul la planificarea misiunii pe Marte.
- Înțelegerea rețelelor și a conexiunilor din cadrul acestora ofera un potențial incredibil de înțelegere și inovare.
- Algoritmica grafurilor este potrivită în mod unic pentru înțelegerea structurilor și dezvaluirea modelelor în seturile de date care sunt foarte conectate.
- In Big Data conectivitatea și interactivitatea sunt cel mai evidente.
- Cantitatea de informații care a fost reunita, amestecata și actualizata dinamic este impresionanta.
- Aici algoritmica grafurilor poat ajuta la înțelegerea volumelor noastre de date, cu analize mai sofisticate care valorifica relațiile și îmbunatațesc informațiile contextuale ale inteligenței artificiale.

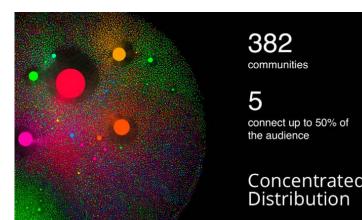
Algoritmica grafurilor. Motivație

- Pe masura ce datele devin mai conectate, este din ce în ce mai important sa înțelegem relațiile și interdependențele acestora.
- Oamenii de știință care studiaza creșterea rețelelor au observat faptul că conectivitatea crește în timp, dar nu uniform.
- Ataşamentul preferenţial este o teorie despre modul în care dinamica impactului creşterii se structurează.
- Figura descrie tendința unui nod de a se lega de alte noduri care au deja o mulțime de conexiuni.
- "Sync: How Order Emerges from Chaos in the Universe, Nature, and Daily Life" (Strogatz, editura Hachette) ofera exemple si explica diferite moduri în care sistemele din viața reala se auto-organizeaza.
- Indiferent de cauzele care stau la baza, mulți cercetatori cred ca modul în care rețelele cresc este inseparabil de formele și ierarhiile rezultate.
- Grupurile foarte dense şi reţelele de date aglomerate tind sa se dezvolte, complexitatea crescând odata cu dimensiunea datelor.



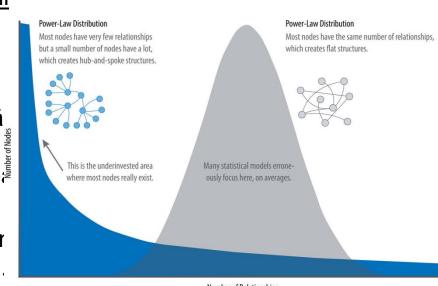
Algoritmica grafurilor. Motivație

- Vedem aceasta grupare de relații în majoritatea rețelelor din lumea reala astazi, de la internet la rețele sociale, cum ar fi comunitatea de jocuri (Figura).
- Analiza rețelei a fost creata de D'Orazio (Pulsar) pentru a ajuta la prezicerea viralitații conținutului și la informarea strategiilor de distribuție.
- D'Orazio a gasit o corelație între concentrarea distribuției unei comunitați și viteza de difuzare a unei bucați de conținut.
- Acest lucru este semnificativ diferit de ceea ce ar prezice un model mediu de distribuție, unde majoritatea nodurilor ar avea același numar de conexiuni.
- De exemplu, daca World Wide Web ar avea o distribuție medie de conexiuni, toate paginile ar avea aproximativ același numar de linkuri care intra și ies.
- Modelele medii de distribuție afirma ca majoritatea nodurilor sunt conectate în mod egal, dar multe tipuri de grafuri și multe ețrele eale rezinta concentrații.
 - Web-ul, are în comun cu grafurile precum calatoriile și rețelele sociale, o distribuție a legii puterii, cu câteva noduri fiind foarte conectate și majoritatea nodurilor fiind conectate modest.



Power Law

- Power Law (numita și lege de scalare) descrie relația dintre doua cantitați în care o cantitate variaza ca putere a alteia.
- Exemplu, aria unui cub este legata de lungimea laturilor sale printr-o putere de 3.
- Exemplu este distribuția Pareto sau "regula 80/20", folosita inițial pentru a descrie situația în care 20% dintr-o populație controla 80% din bogație.
- Vedem diferite legi ale puterii în lumea naturala și în rețele.
- Încercarea de a "media" o rețea, în general, nu va funcționa bine pentru investigarea relațiilor sau prognoza, deoarece rețelele din lumea reala au distribuții inegale de noduri și relații.
- Figura ilustreaza faptul că utilizarea unei medii a caracteristicilor datelor care este inegala ar duce la rezultate incorecte.
- Deoarece datele foarte conectate nu aderă la distribuţie medie, oamenii de ştiinţă utilizează analiza grafurilor pentru a căuta şi interpreta structuri şi distribuţii de relaţii în date din lume, reală.
- Nu există nicio rețea în natură despre care știr că ar fi descrisă de modelul de rețea aleatoriu.



Explicații

A.L. Barabási, director, Centrul pentru cercetarea rețelelor complexe, Universitatea Northeastern, autor numeroase cărți științifice despre rețele, evidențiază următoarele:

- Provocarea pentru majoritatea utilizatorilor este că datele conectate dens și inegal sunt dificil de analizat cu instrumente analitice tradiționale.
- S-ar putea să existe o structură specifică, dar este greu de găsit.
- Este tentant să adoptăm o abordare medie a datelor dezordonate, dar acest lucru va ascunde modele și va asigura că rezultatele noastre nu reprezintă grupuri reale.
- De exemplu, dacă faceți o medie a informațiilor demografice ale tuturor clienților dvs. și oferiți o experiență bazată exclusiv pe medii, veți pierde garantat majoritatea comunităților: comunitățile tind să se grupeze în jurul unor factori asociați, cum ar fi vârsta și ocupația sau starea civilă și locația.
- În plus, comportamentul dinamic, în special în jurul evenimentelor bruște și al izbucnirilor, nu poate fi văzut cu un instantaneu.
- Pentru a ilustra un grup social cu relaţii în creştere, v-aţi aştepta, de asemenea, la mai multe comunicări.
- Acest lucru ar duce la un punct critic de coordonare și la o coaliție ulterioară sau, alternativ, la formarea și polarizarea subgrupurilor (de exemplu in diverse alegeri).
- Sunt necesare metode sofisticate pentru a prognoza evoluția unei rețele în timp, dar putem deduce comportamentul dacă înțelegem structurile și interacțiunile din datele noastre.
- Analiza grafurilor este utilizată pentru a prezice rezistența grupului datorită concentrării asupra relațiilor.

Cazuri de utilizare. Graph Analytics

- La nivel abstract, analiza grafurilor este aplicată comportamentului de prognoză și
 prescrie acțiuni pentru grupurile dinamice.
- Acest lucru necesită înțelegerea relațiilor și structurii din cadrul grupului.
- Algoritmica grafurilor realizează acest lucru examinând natura generală a rețelelor prin conexiunile lor.
- Cu această abordare, se poate înțelege topologia sistemelor conectate și se pot modela procesele acestora.
- Există trei categorii generale de întrebări care indică dacă analiza grafurilor şi algoritmii sunt justificați, așa cum se arată în figura urmatoare

Propagation Pathways Influence Interactions & Resiliency

How do things spread?

What are the capacities, costs, and control points?

How do things interact and will that change?

Tipuri de probleme în care este utilizată algoritmica grafurilor

- Investigați traseul unei boli sau al unui eșec de transport în cascadă.
- Descoperiți componentele cele mai vulnerabile sau dăunătoare într-un atac de rețea.
- Identificați cel mai puțin costisitor sau cel mai rapid mod de a direcționa informații sau resurse.
- Preziceți linkurile lipsă din date
- Localizați influența directă și indirectă într-un sistem complex.
- Descoperiți ierarhii și dependențe nevăzute.
- Previzionați dacă grupurile se vor uni sau se vor separa.
- Găsiţi blocaje sau cine are puterea de a refuza / furniza mai multe resurse.
- Dezvăluiți comunități pe baza comportamentului pentru recomandări personalizate.
- Reduceți rezultatele fals pozitive în detectarea fraudelor și anomaliilor.
- Extrageți mai multe caracteristici predictive pentru învățarea automată.

Concluzie

- Azi datele sunt extrem de conectate şi acest lucru are implicaţii.
- Există practici științifice robuste pentru analiza dinamicii și relațiilor de grup, dar aceste instrumente nu sunt întotdeauna obișnuite de exemplu in afaceri.
- Pe măsură ce evaluăm tehnici avansate de analiză, ar trebui să luăm în considerare natura datelor și dacă trebuie să înțelegem atributele comunității sau să prezicem comportamentul complex.
- Dacă datele reprezintă o rețea, Network, se evită reducerea factorilor la o medie.
- În schimb, se vor folosi instrumente care se potrivesc cu datele și cu informațiile pe care le căutăm.