X

A Gephi használata a közösségi hálózati adatok elemzésére kezdőknek



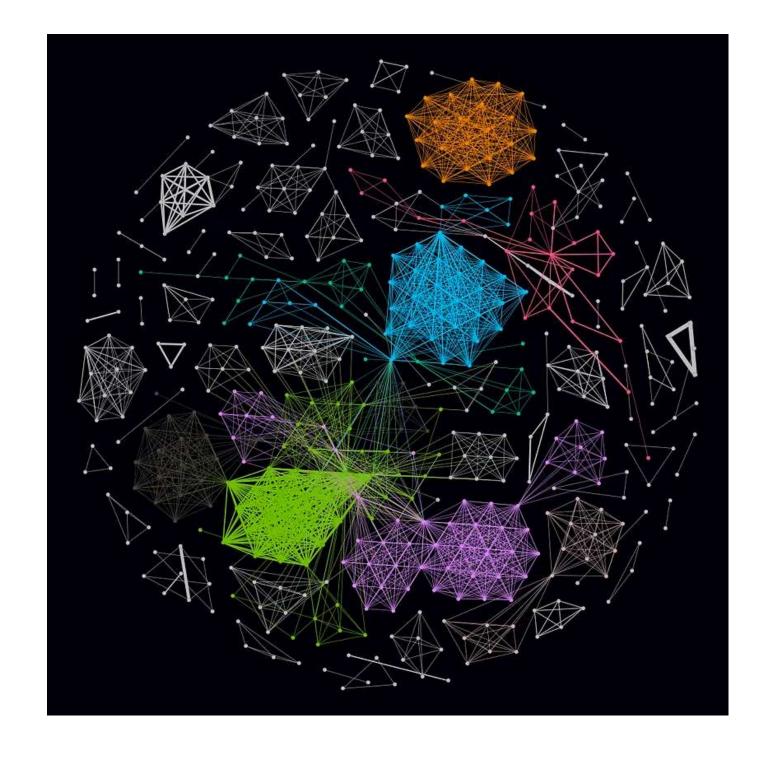
Ifeanyi Idiaye - Alábbi

Olvasási idő: 9 perc - 2023. augusztus 18.









Ebben a bejegyzésben bemutatom a <u>Gephi-t</u>, és megmutatom, hogyan használhatja a közösségi hálózati adatok elemzésére.

Ez a bejegyzés feltételezi, hogy nincs ismerete Gephiről.

Mi az a Gephi

A Gephi egy hatékony, nyílt forráskódú szoftver a hálózatok feltárására és manipulálására. Nyílt forráskódú szoftverként a kódja elágaztatható a GitHub adattárból, és más fejlesztők bővíthetik. Számos független fejlesztő bővítette a Gephi funkcionalitását harmadik féltől származó bővítmények létrehozásával, amelyek lehetővé teszik a Gephi számára, hogy számos feladatot végezzen, például a Twitter-adatok streamelését, a Neo4j-vel való interakciót és az adatok Oracle adatbázisból történő lekérését, hogy csak néhányat említsünk.

A Gephi egyszerű használata miatt népszerű szoftverválasztás a társadalomtudósok, kutatók, oktatók és adattudósok körében. A point-and-click felülettel könnyedén futtathat hatékony algoritmusokat az adatok elemzéséhez, és különféle elrendezési algoritmusok segítségével manipulálhatja a hálózati grafikon megjelenését is. Ezeket működés közben fogjuk látni, ahogy folytatjuk ezt a bejegyzést.

A Gephi használatához telepíteni kell a számítógépére. Tehát, ha még nincs telepítve a Gephi a számítógépére, látogasson el gephi.org oldalra a <u>Gephi</u> legújabb verziójának letöltéséhez és telepítéséhez a számítógépére.

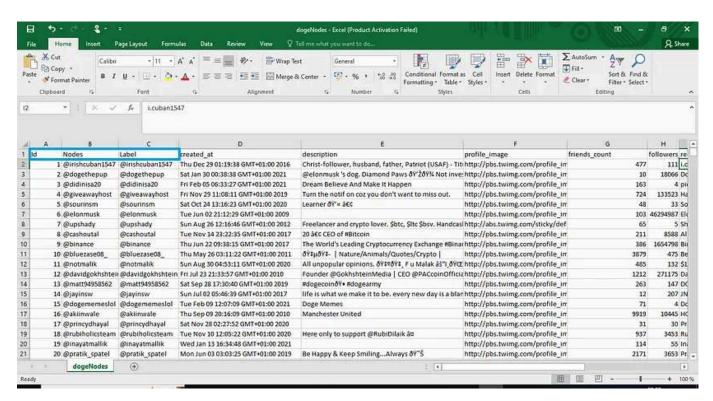
Gephi adatkészlet-formátumok

A Gephi különböző adatkészlet-formátumokat fogad el, például <u>CSV</u>, <u>GraphML</u> és <u>GEXF</u>. Most, ha CSV-fájlként tölti be az adatokat, két különböző CSV-fájlra lesz szüksége, amelyek közül az egyik tartalmazza a csomópontokat, a másik pedig az éleket.

Ebben a példában a CSV fájlformátumot fogjuk használni az adatok Gephibe való betöltéséhez. Az általunk használt adatkészlet a Doge érméről szóló beszélgetések a Twitteren az Elon Musk-korszak előtt. A csomópontokat és a peremhálózati CSV-fájlokat a <u>GephiDatasets GitHub adattárból</u> töltheti le, amely egy függetlenül kezelt adattár, és nem kapcsolódik közvetlenül a Gephihez.

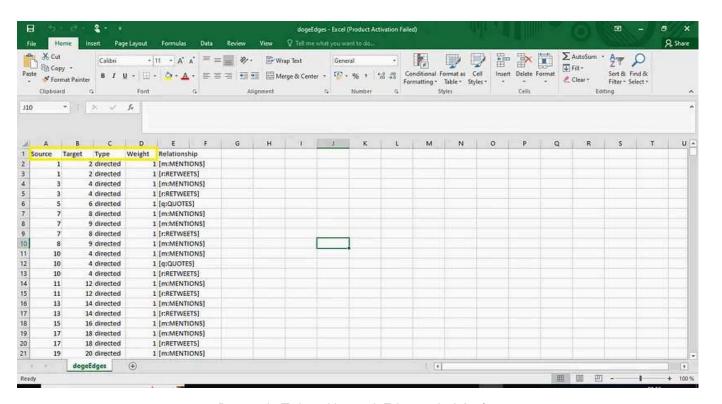
A fájlok "dogeNodes" és "dogeEdges" címkével vannak ellátva. Töltse le őket, hogy követni tudja.

Most nagyon fontos, hogy figyelnie kell az adatkészlet oszlopfejléceire. A csomópont CSV-fájljainak a következő alapértelmezett oszlopokkal kell rendelkezniük az "Id", "Nodes", "Label" fejlécekkel. Más oszlopokat is felvehet, például csomópont-attribútumokat vagy időelemeket a dinamikus hálózatelemzéshez, ahogy az alábbi képen látható.



Dogecoin Twitter hálózati csomópontok adatkészlete

A peremhálózati CSV-fájlok viszont általában a "Forrás", "Cél", "Típus" és "Súly" oszlopokat tartalmazzák, az alábbi ábrán látható módon



Dogecoin Twitter Network Edges adatkészlet

Észreveheti, hogy a Forrás és a Cél adatok karakterek helyett számok. Ennek az az oka, hogy a hálózat minden csomópontjának nevét az azonosítója váltotta fel. A Gephi így képes feltérképezni a csomópontok közötti kapcsolatokat vagy kapcsolatokat. Néhány más hálózatelemző eszköz mátrixot használ a kapcsolatok leképezésére.

Ha nem ehhez hasonló előre formázott adatkészletet használ, le kell cserélnie az egyes csomópontok nevét az azonosítójukra a peremhálózati adatkészletben. Egy egyszerű $\underline{\mathbf{R}}$ vagy $\underline{\mathbf{Python}}$ programmal ez könnyen elvégezhető.

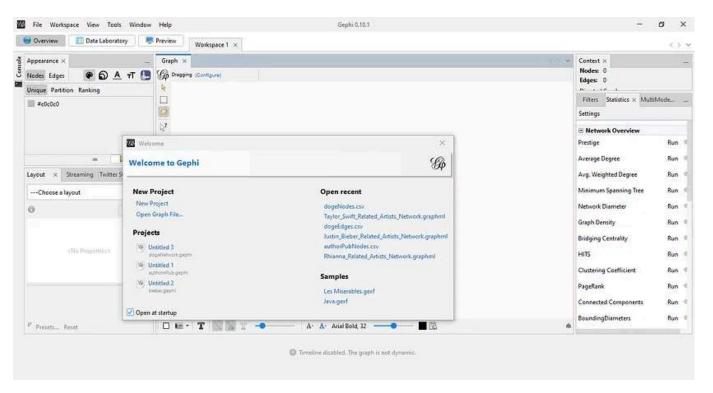
A "Típus" oszlop megmondja Gefinek, hogy milyen hálózati gráfot akarunk rajzolni. Az "irányított" azt jelenti, hogy egyirányú kapcsolatot képezünk le, azaz a kapcsolat az egyik csomópontról a másikra érkezik. Ennek ellentéte az "irányítatlan", ami kétirányú kapcsolat.

A "súly" a két csomópont közötti kölcsönhatás gyakoriságának meghatározására szolgáló mérték. A hálózati gráfban a vastag peremkapcsolatok azt jelentik, hogy nagy súlyuk van, ami azt jelenti, hogy a két csomópont sokat kölcsönhatásba lépett egymással a hálózatban.

Most töltsük be az adatkészleteket a Gephibe.

Indítsa el a Gephi-t

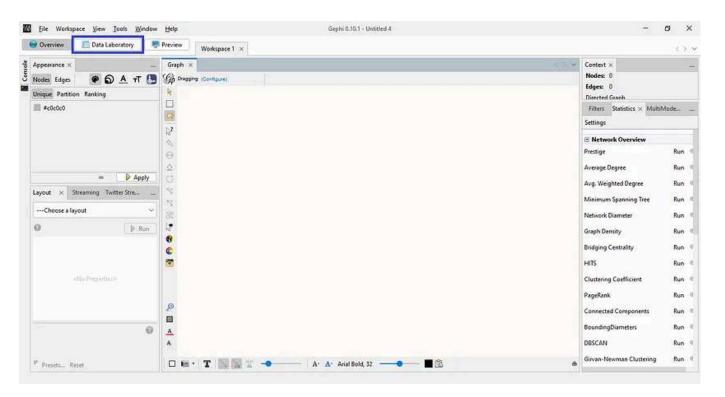
Indítsa el a Gephi-t az asztalon lévő ikonra duplán kattintva. Egy ehhez hasonló felületnek kell megjelennie



Gephi indítási felület

Kattintson az "Új projekt" gombra. Ekkor megjelenik egy üres vászon. A következő dolog, amit meg szeretne tenni, a csomópontok és a peremhálózati adatkészletek importálása.

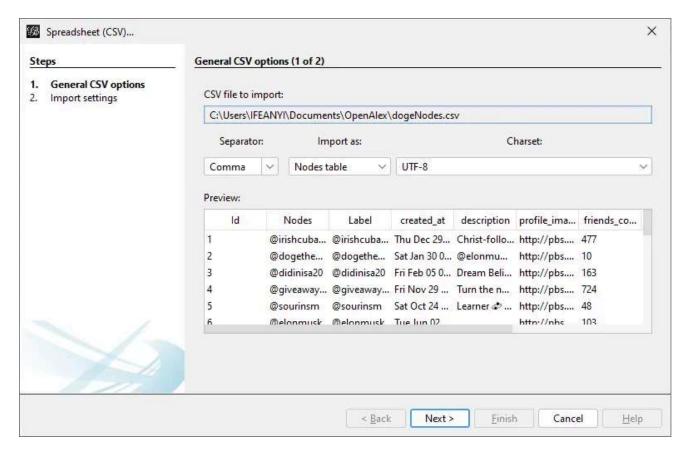
Ehhez lépjen az "Adatlaboratórium" fülre, és kattintson rá. Az alábbi képen a kék téglalap kiemeli



Gephi Data Laboratory lap

Amikor belép az Adatlaboratóriumba, üres Csomópontok és élek mezőt talál. Alapértelmezés szerint a "Csomópontok" van kiválasztva. Az élek importálása előtt mindenképpen importálja a csomópontok adatkészletét; Végül is az emberek a kapcsolatok előtt léteznek.

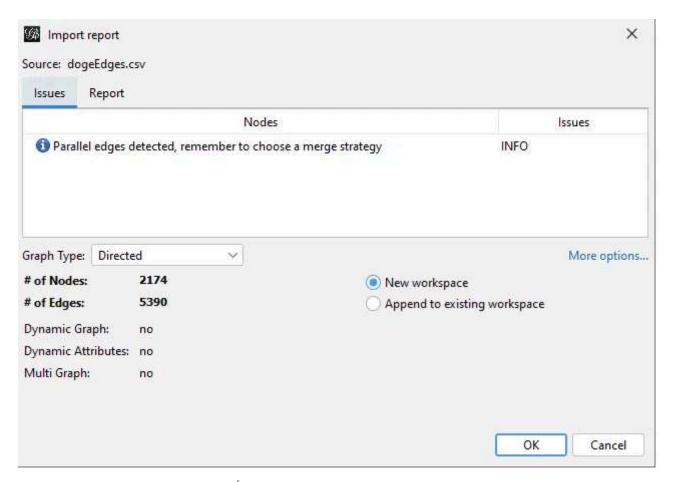
Ezután kattintson a "Táblázat importálása" gombra, és válassza ki a csomópontok CSV fájlját. Az alábbi ablakot kell látnia



Csomópontok adatkészlet-betöltési felülete

Kattintson a "Tovább", majd a "Befejezés" gombra. Ha minden jól megy, látnia kell a hálózati adatkészletben található csomópontok számát. Hagyjon mindent úgy, ahogy van, majd kattintson az "OK" gombra. Ezután látni fogja, hogy az adatkészlet betöltődik az Adatlaboratóriumba.

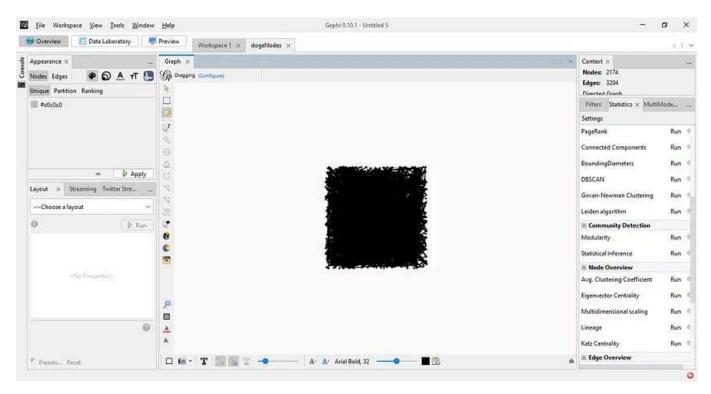
Most ugyanazokat a lépéseket követjük az élek importálásához. Először válassza az Adatlaboratórium élek lapját az üres élmező megjelenítéséhez. Ezután kattintson a "Táblázat importálása" elemre, és válassza ki az élek CSV fájlját. Kattintson a "Tovább", majd a "Befejezés" gombra. Az alábbi képet kell látnia



Élek adatkészletének importálása

Most nagyon fontos, hogy győződjön meg róla, hogy az "OK" gombra kattintás előtt válassza a "Hozzáfűzés meglévő munkaterülethez" lehetőséget, különben az élek nem lesznek hozzáfűzve a csomópontokhoz, és haszontalan grafikont kap.

Ismét, ha minden jól megy, most látnia kell a grafikont az alábbi képen látható módon



Betöltött grafikon

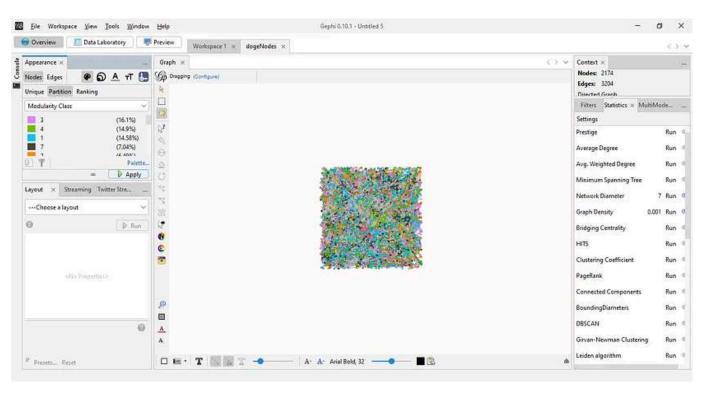
A grafikon elemzése

Most kezdjük el elemezni ezt a grafikont néhány algoritmus futtatásával. A munkaterület jobb oldalán kattintson a "Statisztika" fülre. Látni fogja a futtatható metrikákat, amelyek betekintést nyújtanak a diagramba. Az egyik mutató, amelyet először szeretek futtatni, a "Modularity", amely azonosítja a hálózat különböző csoportjait vagy közösségeit. A futtatás megmutatja, hogy hány közösséget észlelt a gráf.

Azt is szeretném tudni, hogy mennyire sűrű vagy ritka egy hálózati gráf a "Gráfsűrűség" metrika futtatásával, amely kiszámítja a hálózat összekapcsolt éleinek szintjét osztva a lehetséges kapcsolatok teljes számával, ahol minden egyes csomópont csatlakozik az összes többi csomóponthoz. Az 1-hez közelebbi sűrűségi pontszám azt jelenti, hogy a hálózati gráf sűrű, de a 0-hoz közelebbi pontszám azt jelenti, hogy a hálózat ritka.

Futtathatja a "Hálózati átmérő" metrikát is. Ez megmondja a hálózat két legtávolabbi csomópontja közötti távolságot. Egy másik jó metrika a futtatáshoz az "Átlagos útvonalhossz". Ez megmondja, hogy mennyi ideig tart a hálózat áthaladása. Általában, ha a hálózat ritka (a sűrűségi pontszám közelebb van a 0-hoz), akkor magas lesz az átlagos elérési út pontszáma, ami azt jelenti, hogy hosszabb ideig tart a teljes hálózat bejárása.

A Modularitás metrika futtatása után most már színezhetjük a gráfot az észlelt közösségek szerint. Ezért lépjen a munkaterület bal oldalára. A "Csomópontok" alatt 3 fül látható, nevezetesen "Egyedi", "Partíció" és "Rangsorolás". Kattintson a "Partíció" gombra, és a legördülő menüből válassza a "Modularitás osztály" lehetőséget, majd kattintson az "Alkalmaz" gombra. Azonnal észre fogja venni, hogy a grafikon színes

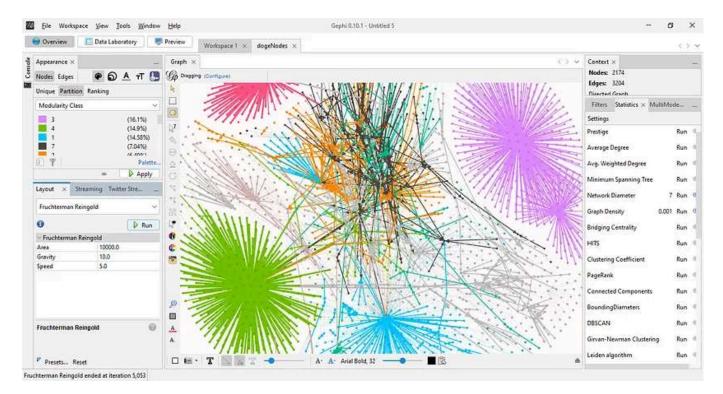


A gráfra alkalmazott modularitási osztály

Szép! A grafikon azonban még mindig nem mond el semmilyen történetet, mert még mindig nem tudjuk kitalálni a hálózat közösségeit. Ennek kijavításához elrendezési algoritmusokat futtatunk, hogy grafikonunk könnyebben érthető legyen. Az én algoritmusom a <u>Fruchterman Reingold algoritmus</u>, amely nagyon láthatóvá teszi a közösségeket. A futtatásához nézze meg az "Elrendezés" fület. Kattintson rá, és a megjelenő legördülő menüből válassza a Fruchterman Reingold lehetőséget. Hacsak nem tudja, mit csinál, hagyhatja a paramétereket úgy, ahogy vannak, és futtathatja.

A futtatás befejezése eltarthat egy ideig, mert a csomópontok a megfelelő fürtökbe vagy közösségekbe lesznek lehúzva. Amikor a folyamat befejeződött, az algoritmus leáll. Alternatív megoldásként manuálisan is leállíthatja, ha elégedett a kimenet minőségével.

A grafikon valahogy így fog kinézni az algoritmus futtatása után



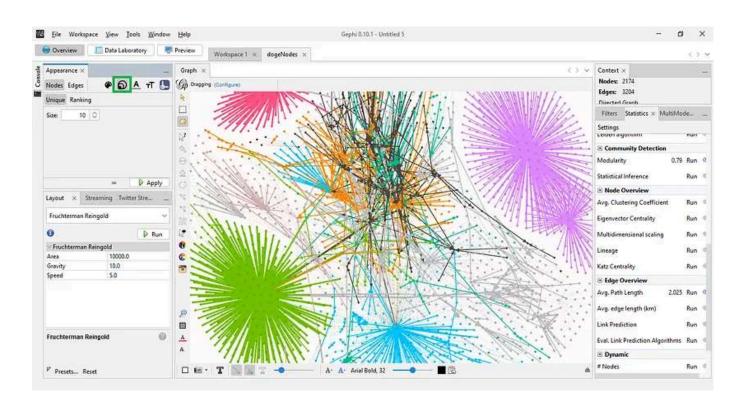
Fruchterman Reingold algoritmus hatás

Most a hálózat közösségei könnyebben láthatók. Futtathatja a "Noverlap" elrendezési algoritmust is, hogy eltávolítsa az átfedéseket a hálózatban.

Most tanuljunk meg néhány információt a hálózat csomópontjairól. Például szeretnénk tudni, hogy mely csomópontok a legbefolyásosabbak a beszélgetésekben. A befolyást itt az alapján mérik, hogy ki a beszélgetés legfontosabb személye, azaz ki a leginkább retweetelt, említett, idézett vagy válaszolt a hálózatban. Szeretnénk a befolyásos csomópont méretét

nagyobbá tenni, mint a többi, hogy könnyen láthatóak legyenek influencerként.

Ezért kattintson az alábbi képen a zöld négyzetben kiemelt ikonra



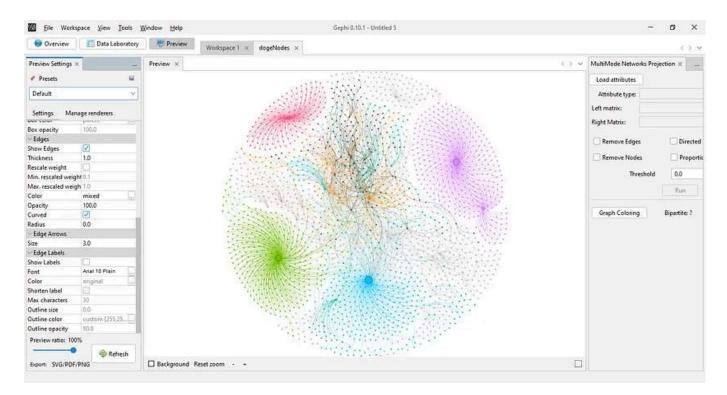
Kattintson a "Rangsor" gombra. A megjelenő legördülő menüben válassza az "In-Degree" lehetőséget. Adja meg a minimális méretet 10-nek és a maximális méretet 60-nak, majd kattintson a "Futtatás" gombra. Abban a pillanatban, amikor ezt megteszi, észre fogja venni, hogy a központi

csomópontok most jobban láthatók. Ezekkel a paraméterekkel játszhat a gráf csomópontjainak méretezéséhez. Vannak más típusú információk is, amelyeket megtudhat a hálózat csomópontjairól, például arról, hogy ki retweetelt, említette, idézett vagy válaszolt másoknak a legtöbbet. Ezt az "Out-Degree" kiválasztásával és a csomópontok ennek megfelelő méretezésével tudhatja meg.

Megtudhatja, hogy kik az egyes csomópontok, ha jobb gombbal kattint rájuk a grafikonon, és kiválasztja a "kiválasztás az adatlaboratóriumban" opciót. Ezután menjen az Adatlaboratóriumba, és látni fogja a csomópontot kiemelve a táblázatban.

Diagramkép előnézete és exportálása

A grafikon előnézetének megtekintéséhez kattintson a fenti "Előnézet" fülre. Egy üres vászon jelenik meg néhány beállítással a bal oldali ablaktáblában. Egyszerűen menjen lefelé, és nyomja meg a "Frissítés" gombot, és látni fogja a grafikonkép szép megjelenését.



A grafikon előnézete

Az exportálás előtt módosíthatja a képet, például az éleket ívelt helyett egyenessé teheti, teljesen eltávolíthatja az éleket, vastagabbá teheti a csomópontokat és éleket, és megjelenítheti a csomópontfeliratokat a diagram tetején. A diagram háttérszínét is módosíthatja.

Végül exportálhatja a grafikont SVG, PDF vagy PNG fájlként a bal oldali ablaktábla alján található "Exportálás" gombra kattintva. Inkább PNG

fájlként exportálom, hogy könnyen megoszthassam a képet másokkal a közösségi médiában.

És így kell használni a Gephi-t az adatok elemzésére. Ez csak a felszínét karcolta meg annak, hogyan kell használni a Gefit. Ennek a cikknek az a célja, hogy elkezdje használni a Gephi-t a közösségi hálózatok elemzési projektjeihez.

Sok más hasznos online forrás is található, amelyeket felhasználhat Gephi tanulási útján. Azt is fontolóra veheti, hogy csatlakozik a Gephi Facebook-csoporthoz, ahol az emberek kérdéseket tesznek fel, véleményeket, projekteket osztanak meg, és segítséget nyújtanak a tanulóknak.

Remélem, tetszett olvasni ezt a bejegyzést!

Kövess a Mediumon: @ifeanyidiaye

Kövess a Twitteren: @Ifeanyidiaye

A Mediumról szóló többi cikkemet itt olvashatod:

<u>Szöveg-kép alkalmazás létrehozása Pythonnal és Open Journey néhány</u> <u>sornyi kóddal | szerző: Ifeanyi Idiaye | 2023. augusztus | Közepes</u>

Könnyen konvertálhat PDF fájlt Word formátumba a {Convert2Docx} R csomag segítségével | szerző: Ifeanyi Idiaye | 2023. július | Közepes

R Shiny alkalmazás üzembe helyezése ölelő arctérként | szerző: Ifeanyi Idiaye | 2023. június | Közepes

<u>A számítógép helyi mappájának csatlakoztatása egy GitHub-adattárhoz | szerző: Ifeanyi Idiaye | 2023. július | Közepes</u>

Gefi Csicsergés GitHub



Írta: Ifeanyi Idiaye



56 követői

Adattudós vagyok, aki szenvedélyesen foglalkozik az Al-val és az Al-vezérelt alkalmazások építésével. A Dev Genius szoftverfejlesztője és írója is vagyok.

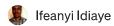
További információk lfeanyi Idiaye-től





Gyors adatfeltárás Python &; R rendszerben a Graphic-Walker...





Szöveg-kép alkalmazás létrehozása Pythonnal és az Ope... Data exploration is a critical step in building any analytical or machine learning pipeline,... In this post, I will show you how you can easily build your own text-to-image application wit...

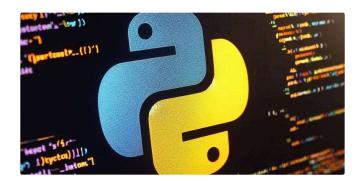
May 29 👋 268

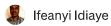


Aug 3, 2023











In this post, I will show you how you can easily build and run a Python Gradio application...

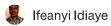
Nov 9, 2023











How to Deploy Flowise Application As A Hugging Face Space

In this post, I will show you how to deploy Flowise application as a Space on Hugging...

Sep 21, 2023













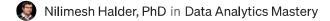




See all from Ifeanyi Idiaye

Recommended from Medium





Mastering Scatterplots in Data Science and Statistics: A...

Article Outline:





Abhay Parashar in The Pythoneers

17 Mindblowing Python Automation Scripts I Use Everyday

Scripts That Increased My Productivity and Performance

→ 5h ago 3 8K 4 76

(+ ···



Icon Design

36 stories 423 saves



Generative Al Recommended Reading

52 stories1403 saves •



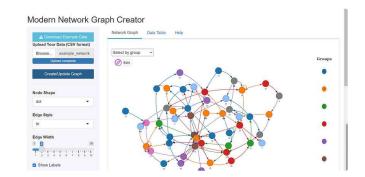
Staff Picks

742 stories1335 saves •



My Kind Of Medium (All-Time Faves)

96 stories 503 saves •





Create Stunning Network Graphs Without Code: Leveraging Python...

In today's data-driven world, visualizing complex relationships is crucial. Whether...

Software Development Engineer

Mar. 2020 - May 2021

- Developed Amazon checkout and payment services to handle traffic of 10 Million daily global transactions
 Integrated Iframes for credit cards and bank accounts to secure 80% of all consumer traffic and prevent CSRF
- Integrated Iframes for credit cards and bank accounts to secure 80% of all consumer traffic and prevent CSRI
 cross-site scripting, and cookie-jacking
- Led Your Transactions implementation for JavaScript front-end framework to showcase consumer transactions and reduce call center costs by \$25 Million
- · Recovered Saudi Arabia checkout failure impacting 4000+ customers due to incorrect GET form redirection

Projects

NinjaPrep.io (React)

- · Platform to offer coding problem practice with built in code editor and written + video solutions in React
- Utilized Nginx to reverse proxy IP address on Digital Ocean hosts
 Developed using Styled-Components for 95% CSS styling to ensure proper CSS scoping
- Implemented Docker with Seccomp to safely run user submitted code with < 2.2s runtime

HeatMap (JavaScript)

- Visualized Google Takeout location data of location history using Google Maps API and Google Maps
 - heatmap code with React
- Included local file system storage to reliably handle 5mb of location history data
- Implemented Express to include routing between pages and jQuery to parse Google Map and implement heatmen quadru



Alexander Nguyen in Level Up Coding

The resume that got a software engineer a \$300,000 job at Google.

1-page. Well-formatted.



Sep 20



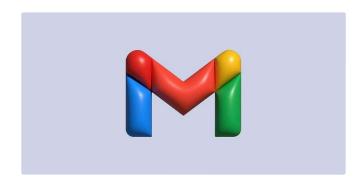
Ct.

•

Jun 1 👋

23K

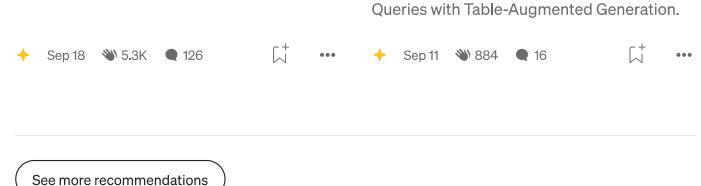
M





Goodbye Gmail: The Hard Truth About Why It's Time for a Change

The end of an era.



Shang-Chi

3

2257.8

Query Execution

2257.8 2257.8

Answer Generation

Pavan Ema... in Artificial Intelligence in Plain Engli...

Goodbye, Text2SQL: Why Table-

Exploring the Future of Natural Language

Augmented Generation (TAG) is...