Vizualizacija kompleksnosti

Kinda, Jerolim

Undergraduate thesis / Završni rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:137:698453

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

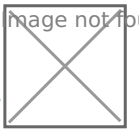
Download date / Datum preuzimanja: **2024-03-29**



Repository / Repozitorij:

Digital Repository Juraj Dobrila University of Pula





Sveučilište Jurja Dobrile u Puli
Odjel za informacijsko-komunikacijske tehnologije

JEROLIM KINDA

VIZUALIZACIJA KOMPLEKSNOSTI

Završni rad

Pula, lipanj, 2017 godine

Sveučilište Jurja Dobrile u Puli
Odjel za informacijsko-komunikacijske tehnologije

JEROLIM KINDA

VIZUALIZACIJA KOMPLEKSNOSTI

Završni rad

JMBAG: 0303046094, redoviti student
Studijski smjer: Informatika

Predmet: Teorija Sustava
Znanstveno područje: Društvene znanosti
Znanstveno polje: Informacijske i komunikacijske znanosti
Znanstvena grana: Informacijski sustavi i informatologija
Mentor: doc. dr. sc. Darko Etinger

Pula, lipanj, 2017 godine

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Jerolim Kinda, kandidat za prvostupnika informatike ovime izjavljujem da je ovaj Završni rad rezultat isključivo mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student
Jerolim Kinda

U Puli, lipanj, 2017 godine

IZJAVA
o korištenju autorskog djela

Ja, Jerolim Kinda dajem odobrenje Sveučilištu Jurja Dobrile
u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom Vizualizacija kompleksnosti koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.
Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, 06.07.2017

Potpis
Jerolim Kinda

Sadržaj

1. Uvod 1

2. Vizualizacija 2

2.1 Povijest vizualizacije 2

2.2. Današnje potrebe vizualizacije 5

2.2.1 Big data 5

2.2.2 Prednosti vizualizacije u poslovanju 6

2.2.3 Društvene mreže 6

2.2.4 Budućnost vizualizacije 7

3. Kompleksne mreže 8

3.1 Grafovi 8

3.1.1 Usmjerenost i povezanost grafa 8

3.1.2 Težinski graf 10

3.1.3 Stabla 11

3.2 Vrste i karakteristike kompleksnih mreža 12

3.2.1 Distribucija stupnja povezanosti 12

3.2.2 Koeficijent grupiranosti 12

3.2.3 Struktura zajednice 14

3.2.4 Nerazmjerne mreže 14

3.2.5 Mreže malog svijeta 15

4. Alati i tehnike za vizualizaciju kompleksnih mreža 16

4.1 Alati za vizualizaciju kompleksnih mreža 16

4.1.1 R i RStudio 16

4.1.2 Gephi 17

4.1.3 D3 18

4.1.4 Cytoscape 18

4.2 Tehnike vizualizacija kompleksnih mreža 19
4.2.1 Čvor-brid dijagram 19
4.2.2 Aluvijalni dijagram 22
4.2.3 Vizualizacija geografskih sustava. 23
4.3 Smjernice za vizualizaciju 24
5. Praktični primjer vizualizacije kompleksne mreže 27
5.1 Primjer vizualizacije pomoću lukova 27
6. Zaključak 30
Literatura 31
Popis slika 34
Popis programskih kodova 35

1. Uvod

U ovom radu ćemo opisati vizualizaciju kompleksnih mreža. Što su kompleksne mreže, što očekujemo od njih i kako ih prikazati su neka od ključnih pitanja na koje ćemo odgovoriti u ovom završnom radu. Sastoji se od uvoda, cjelina u kojima dio po dio objašnjavamo ideju rada i od zaključka. Cjeline su podijeljene tako da što jednostavnije objasne određene koncepte koji su bitni za vizualizaciju kompleksnih mreža.

Rad započinje uvodom u vizualizaciju, prolazeći kroz povijest prikazivanja znanja i informacija do analize današnjih potreba vizualizacije. Zbog učestale korištenosti prikazivanja informacija, analiza se neće samo zadržavati na informatičkim granama, nego će objasniti kakve su današnje potrebe poslovanja. Pošto bi direktni prelazak s vizualizacije na kompleksne mreže mogao biti nejasan, u početku treće cjeline objasnit ćemo osnovna svojstva mreža. Pomoću osnovnih definicija i jednostavnih grafova na početku cjeline i kompleksnih mreža pri kraju cjeline možemo objasniti sve o kompleksnim mrežama.

Te dvije cjeline mogle bi se svrstati u prvi dio završnog rada, gdje je objašnjeno što su mreže i što se traži od njih. Drugi dio rada odnosi se na izvršavanje tih vizualizacija. Četvrta cjelina se bazira na analizi popularnih alata za vizualizaciju i metodama koje se često koriste pri radu s kompleksnim mrežama. Osim toga, spominju se i određene smjernice koje se odnose na kvalitetu vizualnog prikaza. Zadnja cjelina je praktičan primjer vizualizacije kompleksne mreže, gdje će se izraditi primjer koristeći se alatima koji su obrađeni u prijašnjoj cjelini.

2. Vizualizacija

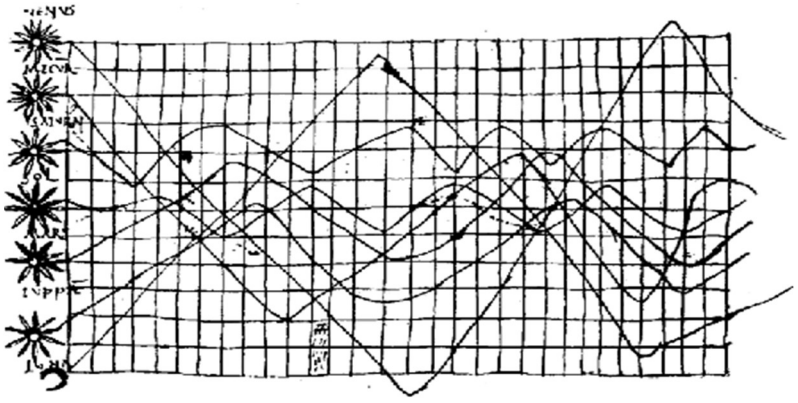
Ljudi su vizualna bića. Najlakši način da osobi prenesemo određenu informaciju jest vizualno. Bez obzira na to što smo napredniji od drugih stvorenja, naš mozak još uvijek najbolje procesira vizualne podatke. Najbolji primjer su stare civilizacije koje su zapisivale informacije na način da one budu vizualno razumljive.

Kroz godine razvili smo naprednije metode vizualizacije koje su nam dobro došle za razumijevanje informacija. Napretkom tehnologije, počinjemo sve više skladištiti informacije. Zbog viška informacija dolazi do problema razumijevanja istih jer postaje sve teže organizirati informacije u cjeline koje imaju smisla. Rješenje se nalazi u brzom razvoju tehnologija i metoda koje nam omogućuju jednostavan prikaz kompliciranih podataka. Današnji alati nam omogućavaju da u par minuta vidimo što bi ručnom obradom trajalo satima, možda i danima. Da bismo lakše razumjeli kako smo došli do današnjih metoda, trebali bismo barem ukratko proći vizualizaciju kroz ljudsku povijest.

2.1 Povijest vizualizacije

Među prve vizualizacije mogli bismo ubrojiti karte nebeskih tijela koje su radili drevni Egipćani. Te karte su se radile na temelju podataka o pozicijama zvijezda i drugih nebeskih tijela, dok su iste bile korištene kao pomagalo pri putovanju i istraživanju. Ideju o koordinatama kartografi su preuzeli od Egipatskih geodeta koji su na taj način planirali izgradnju[1].

Slika 1. Prikaz putanja nebeskih tijela



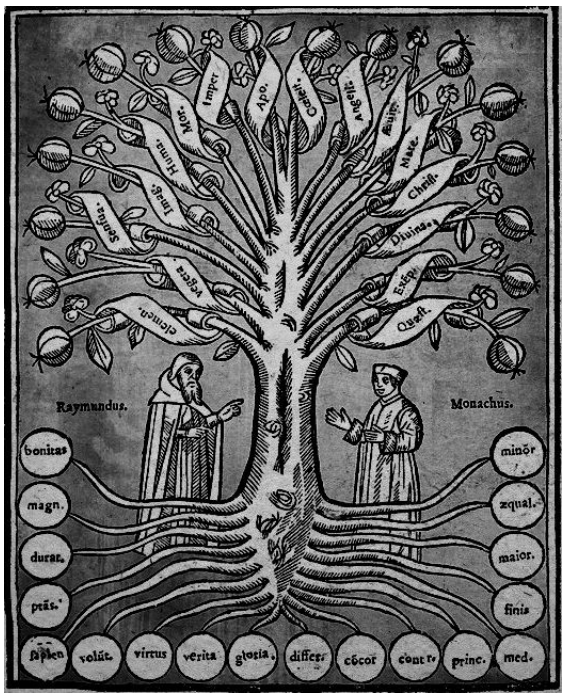
Izvor: Friendly M. (2006)

[1] http://www.datavis.ca/papers/hbook.pdf pristupno 10.05.2017

Jedan od prvih kvantitativnih prikaza informacije jest grafički prikaz vremenskog niza pozicija i putanja određenih planeta i zvijezda. Slika 1 prikazuje grafikon koji je napravljen u desetom stoljeću, iako koristi metode zapisivanja i prikazivanja podataka koje su se tek razvile u sedamnaestom stoljeću. Razni znanstvenici su vidjeli kakve prednosti donosi vizualni prikaz podataka pa su odlučili svoja istraživanja zapisati u takvom formatu.

Kartografija je jedna od glavnih područja koja je proizašla iz prije navedenih istraživanja[2]. Kartografi su osim geografskih položaja željeli pokazati podatke kao što su nadmorska visina, jakost morskih struja i slično. Statističari su raznim grafikonima počeli prikazivati podatke iz svojih istraživanja kako bi lakše pronašli rješenja. Iako se radi o grafikonima koju su nastali u osamnaestom stoljeću, danas su jedni od najkorištenijih u prikazivanju podataka. Kao primjer bismo mogli uzeti linijski i stupičasti grafikon koji se nalaze u svakom alatu za prikazivanje i analizu podataka. Osim striktno znanstvenog pristupa, razvijale su se metode vizualnog prikaza kao što su stabla znanja.

Slika 2. Ramon Llull – Arbor scientiae



Izvor: http://educationopinion.blogspot.hr/

Stablo je odabrano kao reprezentacija znanja zbog svojeg izgleda. Svojim korijenjem, deblom, granama i listovima može grupirati i prikazati određenu hijerarhiju podataka. Stablom se

[2] https://www.gislounge.com/mapping-through-the-ages/ pristupano 10.05.2017

najčešće prikazuje znanje u granama znanosti kao što su lingvistika, arheologija, biologija, filozofija i rodoslovlje[3]. Kao što možemo vidjeti na slici 2, španjolski filozof Ramon Llull je u trinaestom stoljeću pomoću stabala prikazivao stečeno znanje[4]. Obiteljska stabla su primjer prikaza znanja koja danas koristimo. Bez obzira na brzinu razvoja, znanstvenici su tek razvojem računala prepoznali potencijal vizualizacije. Za razliku od ručnog obrađivanja podataka koje ima mnogo fizičkih ograničenja, računalo je znanstvenicima omogućilo bržu i precizniju analizu podataka koja stavlja fokus na finalni produkt, tj. na prikaz podataka umjesto na točnost obrade.

Fokus razvoja vizualizacije se prebacio na razvoj računalnih aplikacija i programskih jezika. Osim softvera, potaknut je razvoj raznih ulaznih i izlaznih jedinica kojima bismo lakše unosili i prikazivali podatke. Takav ubrzan razvoj nas je doveo do današnje tehnologije koju svakodnevno koristimo. Danas rijetko koje računalo nije opremljeno alatima za obradu i prikaz podataka. Osim alata, svakodnevno smo izloženi vizualizacijama u medijima, bio to kartografski prikaz pobjede političkih stranaka na izborima po županijama ili stupičasti dijagram u kojem uspoređujemo prosječne plaće određenih zemalja. Vizualizacija je postala jedna od glavnih metoda prijenosa informacija.

[3] Lima M. „Visual complexity“, str 25.
[4] http://www.scottbot.net/HIAL/index.html@p=39166.html pristupano 12.05.2017

2.2. Današnje potrebe vizualizacije

Napretkom interneta svatko može pristupiti ogromnim količinama podataka jer veliku količinu njih možemo naći besplatno na internetu. Naravno, nekada ćemo imati pristup osiromašenim podacima jer vlasnici podataka žele sačuvati privatnost svojih korisnika.
Problem se nalazi u kompleksnosti i veličini podataka. Nekada jednostavno ne možemo na prvu ruku razumjeti što se događa.

2.2.1 Big data

Big data je izraz koji opisuje veliku količinu podataka, bili to strukturirani ili nestrukturirani podaci[5]. Koliko god se činilo da je količina bitna, iz tih podataka je najvažnije izvući informacije koje su nam potrebne.
Jedna od popularnijih definicija je podjela na varijable:

1. Volumen – Razne organizacije prikupljaju podatke iz raznih izvora, od društvenih mreža do poslovnih transakcija. Današnje tehnologije nam omogućavaju jednostavan zapis tih podataka.
2. Brzina – Podaci se koriste i zapisuju u stvarnom vremenu.
3. Raznolikost – Podaci dolaze u svim mogućim formatima. Mogu biti tradicionalne baze podataka, email poruke, transakcije, itd..

Jedan od načina na koji možemo bolje razumjeti veliku količinu podataka jest vizualizacijom. Svrha prikazivanja podataka jest uočavanje onog što se događa među podacima. Kompleksnost prikaza ovisi o publici kojoj je namijenjen. Znanstvenicima ili ekonomistima će trebati detaljnije informacije o određenim podacima dok će široj publici za informativne svrhe poslužiti pojednostavljena verzija. Jedan od popularnijih formata je infografikon koji zbog svoje jednostavnosti i jasnoće može vrlo efikasno prenijeti veliku količinu podataka publici koja nije toliko informirana o temi[6]. Osim popularnih formata prikazivanja podataka, postoje i detaljniji prikazi podataka kojima se više služe zahtjevniji korisnici koji pokušavaju izvući informacije za osobne ili poslovne svrhe.

[5] https://www.sas.com/en_us/insights/big-data/what-is-big-data.html pristupano 12.05.2017
[6] https://www.wired.com/insights/2014/01/big-data-look-like-visualization-key-humans/ pristupano 12.05.2017

2.2.2 Prednosti vizualizacije u poslovanju

Problem s tablicama podataka i izvješćima jest taj što formati za prikazivanje informacija nekada nisu dovoljno dobri. Za razliku od klasičnih formata, noviji načini prikazivanja informacija omogućavaju lakše razumijevanje podataka i primjećivanje sličnosti među njima. Postoji veća mogućnost da će korisnik naći ono sto mu je potrebno u grupiranim grafikonima ili povezanim podacima, te da će ta brzina interpretacije pozitivno utjecati na produkt[7]. Aplikacije u kojima se obavljaju sve ove radnje imaju i drugi zadatak na tržištu, njihova upotreba treba biti jednostavna i jasna. Smanjenjem potrebnog tehničkog znanja korištenja alata dozvoljava korisniku da se posveti onome što mu je bitno.

Praćenje poslovanja u stvarnom vremenu je luksuz koji si danas puno poduzeća može priuštiti. Najveća prednost takvog praćenja je mogućnost primjećivanja novih trendova koji se događaju. Takva tehnologija omogućava poduzećima da na vrijeme mijenjaju svoje poslovne strategije kako bi se prilagodili tržištu. Zahvaljujući internetu, današnji trendovi mogu postati popularni u svijetu u roku od par dana. Klasičnim metodama trend bi kasno bio primjećen.

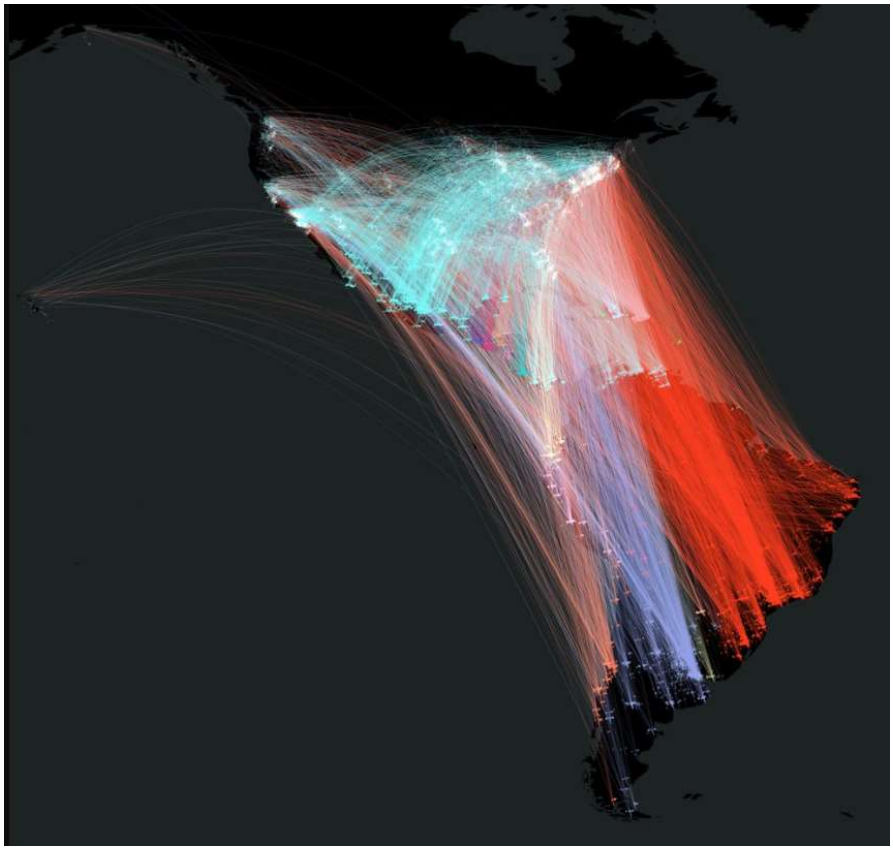
2.2.3 Društvene mreže

Jedan od boljih izvora podataka o korisnicima mogu biti društvene mreže koje uz jednostavne alate prikazuju aktivnosti na profilu proizvođača. Osim jednostavnih grafikona o demografiji, interakciji ili o korištenim uređajima za pristup sadržaju, neke stranice nude povratnu informaciju u realnom vremenu kako bi mogli vidjeti što se točno događa kada objavimo novi sadržaj. Osim trenutnih pretplatitelja, kroz analizu možemo naći korisnike društvenih mreža koji nisu pretplaćeni na sadržaj ali bi mogao zanimati.

Jedan od načina na koji možemo vidjeti koji se korisnici uključuju u razne trendove jest praćenjem objava društvenih stranica. Kao primjer možemo uzeti Twitter[8] koji nerijetko objavljuje vizualne prikaze svojih analiza. Kako bi si olakšali posao, uvijek analiziraju aktualna događaja kao što su politički izbori ili sportske utakmice.

[7] http://data-informed.com/top-5-business-benefits-using-data-visualization/ pristupano 12.05.2017
[8] www.twitter.com pristupano 15.05.2017

Slika 3. Povezanost Latinoamerikanaca



Izvor: https://twitter.com/TwitterAlas/status/712058816041345024

Na slici 3 prikazana je komunikaciju između Latinoamerikanaca u Sjedinjenim Američkim Državama s ostalim Latinoamerikancima u srednjoj i južnoj Americi.

2.2.4 Budućnost vizualizacije

Jedna od generalnih ideja jest ta da se u budućnosti automatiziraju svi procesi vizualizacije[9]. Iako je danas dosta procesa automatizirano, naprednije postupke još uvijek moramo ručno izvoditi. Problem današnjih alata jest taj da nisu dostigli razinu razumijevanja podataka koja bi mogla odlučiti što je za korisnika bitno a što ne. Zato računala i dalje gledamo kao pomagalo pri rješavanju problema, a ne kao rješenje.

[9] https://www.statsilk.com/blog/benefits-and-future-data-visualization pristupano 15.05.2017

3. Kompleksne mreže

Jedna od glavnih tema ovog rada su komplekse mreže. Vizualizacije tih mreža mogu se naći u sljedećem poglavlju, dok će u ovom biti razrađena teoretska strana.
Kompleksne mreže bi se mogle definirati kao grafovi koji nisu striktno regularni ni nasumični, ali imaju slične karakteristike grafova koji su simulacija stvarnih sustava. Da bismo mogli bolje razumjeti kompleksne mreže moramo razumjeti što su grafovi i kako funkcioniraju.

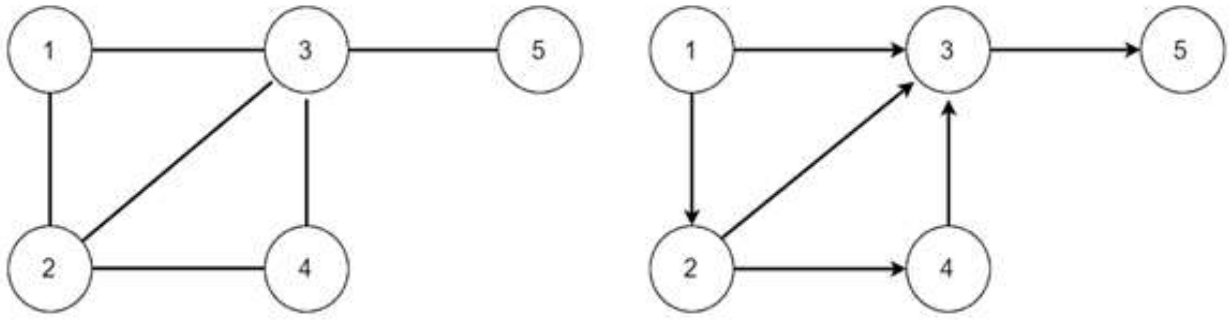
3.1 Grafovi

Najjednostavnija definicija grafa jest skup objekata u kojem objekti mogu biti povezani. U grafu objekte možemo prikazati kao vrhove dok veze možemo prikazati kao bridove[10]. Prvo ćemo obradit usmjerenost grafa.

3.1.1 Usmjerenost i povezanost grafa

Kao što vidimo na slici 4, postoje usmjereni i neusmjereni grafovi. U neusmjerenim grafovima bridovi ne sadrže strelice što znači da je veza između dva čvora obostrana. U usmjerenim grafovima, brid koji sadrži strelicu pokazuje jednostranu vezu. Smjer strelice obilježava iz kojeg čvora kreće veza. Neusmjereni grafovi ne moraju striktno biti prikazani bez strelica. U slučaju da se radi o neusmjerenom grafu možemo staviti strelice na oba vrha brida da bi prikazali obostranu vezu. U slučaju da graf sadrži usmjerene i neusmjerene veze, radi se o mješovitom grafu.

Slika 4. Neusmjereni i usmjereni graf

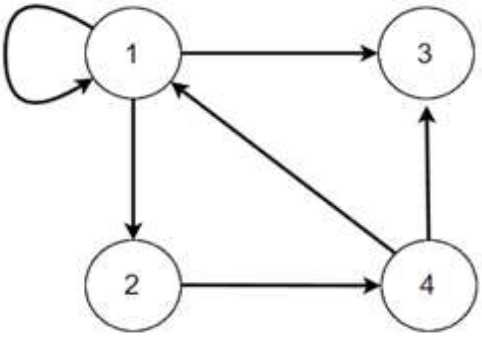


Izvor: Izrada autora

[10] http://www.math.fsu.edu/~pkirby/mad2104/SlideShow/s6_1.pdf pristupano 16.05.2017

Osim što čvor može biti povezan s drugim čvorovima, postoje slučajevi gdje brid kreće i završava u istom čvoru. Takva veza naziva se petljom.

Slika 5. Graf s petljom

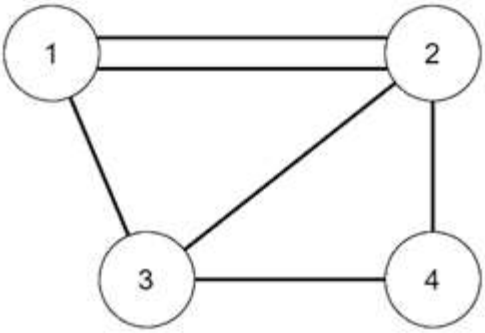


Izvor: Izrada autora

Za čvorove koji su povezani istim bridom kažemo da su susjedni čvorovi. U slučaju da čvor nema petlju, možemo definirati stupanj čvora za svaki čvor. Stupanj određuje broj bridova kojima je zadan čvor kranja točka.

Osim standardnih veza gdje jedan brid povezuje dva čvora, ako postoji više bridova koji povezuju dva čvora, za graf možemo reći da je multigraf. Jednostavan primjer multigrafa možemo vidjeti na slici 6.

Slika 6. Primjer multigrafa



Izvor: Izrada autora

Iako postoji par dodatnih pravila za multigraf, on ne smije biti usmjeren i ne smije imati petlje. U slučaju da želimo prikazati multigraf s petljama koristili bi pseudograf, koji dozvoljava korištenje petlja.

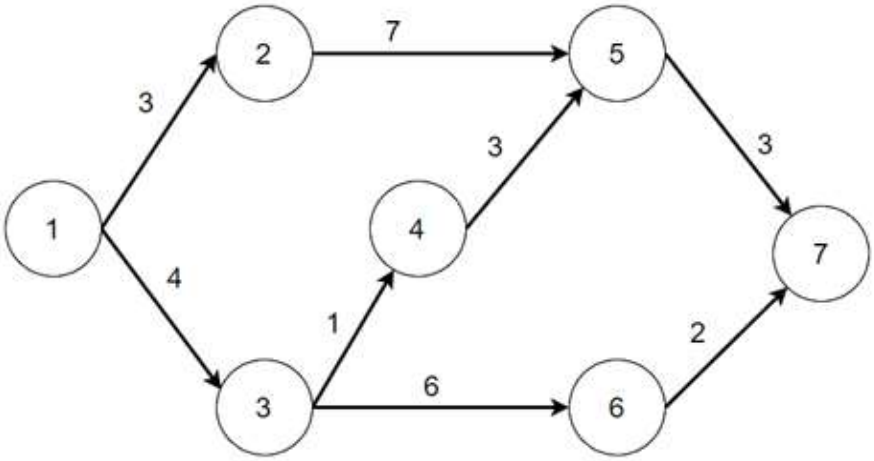
Vrijedi spomenuti druge vrste povezanosti kao što su:

1. Povezan graf – Ako se bilo koja dva čvora mogu povezati određenim putem. U slučaju da ne mogu, graf je nepovezan.
2. Potpuni graf – Ako su svi čvorovi u grafu međusobno povezani direktnom vezom.
3. Planiran graf – Ako se bridovi ne sijeku kada graf nacrtamo u ravnini.

3.1.2 Težinski graf

Jedina razlika između težinskog i prije navedenih grafova jest ta da težinski ima pridruženu vrijednost svakom bridu[11]. Za nadodanu vrijednost često se koristi izraz cijena, kako se ova vrsta grafova često koristi u zadacima kao što je odabir najkraće rute ili minimalizacija cijene. Na slici 7 možemo vidjeti jednostavan primjer težinskog grafa i kako su označene težine veza.

Slika 7. Primjer usmjerenog težinskog grafa



Izvor: Izrada autora

Jedan od načina korištenja ovog grafa jest da izračunamo put koji ima najmanju vrijednost tj. izračunamo najkraći put između dva čvora. U slučaju da se radi o usmjerenom grafu, smanjuje se broj raspoloživih puteva.

Postoji mnogo algoritama za rješavanje ovakvih problema. Grafom možemo simulirati probleme u stvarnom svijetu, ali ćemo ih izostaviti jer nisu bitni za ovaj rad.

[11] http://courses.cs.vt.edu/~cs3114/Fall10/Notes/T22.WeightedGraphs.pdf pristupano 16.05.2017

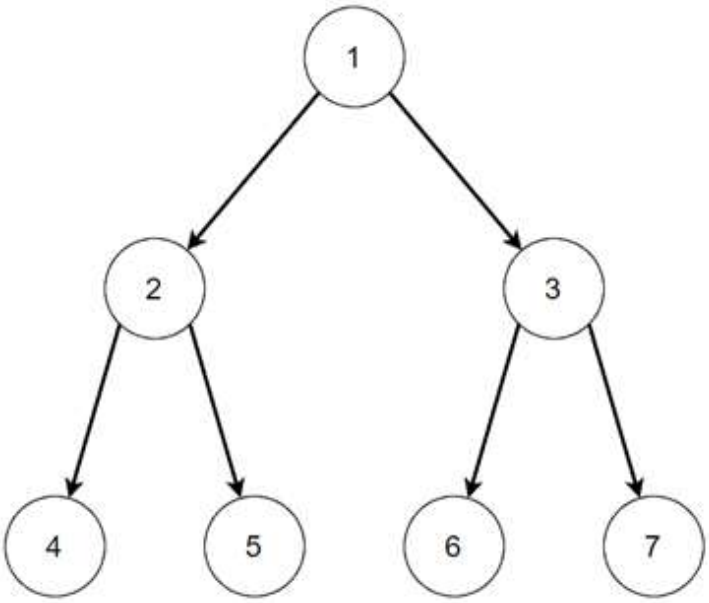
3.1.3 Stabla

Vrsta grafova koja se znatno razlikuje od drugih u ovom poglavlju su stabla. Rijetko kada su usmjerena jer bi takav graf imao slabu povezanost[12]. Većinu stabala možemo podijeliti na 4 dijela:

1. Korijen – Ako stablo ima korijen radi se o ukorijenjenom stablu. U tom stablu svi bridovi postaju usmjereni, i to u smjeru od čvora koji je bliži korijenu prema čvoru daljem od korijena.
2. Roditelj – Ako su dva čvora povezana, čvor bliži korijenu bit će roditelj.
3. Dijete – Ako su dva čvora povezana, čvor dalji od korijena bit će dijete.
4. List – Ako čvor nema niti jedno dijete, onda ga zovemo listom

U slučaju da čvor ima samo jedno dijete, za njega se kaže da je unutrašnji čvor.

Slika 8. Primjer stabla



Izvor: Izrada autora

Među korištenijim grafovima ove vrste jest binarno stablo. Takvo stablo je ukorijenjeno i ima maksimalno dvoje djece po roditelju. Obično te čvorove zovemo lijevim i desnim djetetom. Slika 8 prikazuje jednostavno binarno stablo gdje svaki roditelj ima po dvoje djece.

[12] https://www.fer.unizg.hr/_download/repository/Osnovni_pojmovi-teorija_grafova.pdf pristupano 17.05.2017

3.2 Vrste i karakteristike kompleksnih mreža

Nakon što smo pojasnili jednostavne mreže, možemo se vratiti na temu kompleksnih. Istraživanje kompleksnih mreža je relativno mlado znanstveno područje koje se bazira na mrežama stvarnog svijeta. Do nedavno takve mreže nisu postojale ili nismo imali tehnologiju za istražiti željena područja. Mreže poput društvenih su tek postale popularne u zadnjih desetak godina. Računalne mreže su malo starije ali su još relativno nove naprema drugim znanstvenim područjima. Kao što je navedeno u početku poglavlja, kompleksne mreže se razlikuju od standardnih mreža. Takve mreže znaju imati vrlo asimetričnu distribuciju stupnja povezanosti, mogu imati koeficijent grupiranja i često su strukturirane poput zajednica. Da bismo lakše razumjeli te pojmove, ukratko ćemo ih objasniti.

3.2.1 Distribucija stupnja povezanosti

Ova distribucija se temelji na stupnju povezanosti kojeg smo već objasnili u prošlom dijelu o grafovima. Nakon što izračunamo stupanj povezanosti za svaki čvor, cilj nam je napraviti distribuciju kojom možemo prikazati taj stupanj i izračunati prosječni stupanj. Kako kompleksne mreže znaju biti vrlo asimetrične, dobar dio čvorova biti će ispod prosječnog stupnja povezanosti. Zbog dugog repa distribucije par čvorova zna imati daleko veći stupanj povezanosti. Samim time mogu uzrokovati povećanje prosječne povezanosti[13]. U slučaju da se koriste usmjereni grafovi, možemo odvojeno računati distribuciju stupnja izlaznih i ulaznih bridova.

3.2.2 Koeficijent grupiranosti

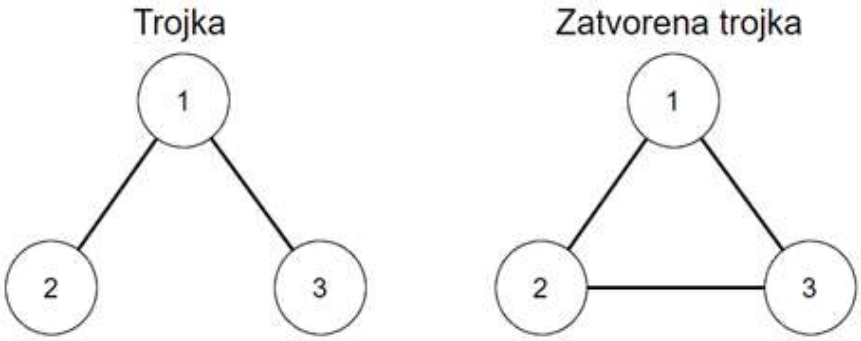
Ovaj koeficijent mjeri stupanj grupiranosti čvorova u grafu. Istraživanja su pokazala da u većini društvenih mreža i mreža stvarnog svijeta, čvorovi znaju formulirati snažno povezane grupe. Vjerojatnost tako jako povezanih čvorova je dosta veća od prosječne vjerojatnosti nasumične povezanosti između dva čvora[14].

Društvene mreže su izvrstan primjer ovakvog grupiranja zbog ponašanja njihovih korisnika. Kako postoji velika vjerojatnost da dva prijatelja imaju dosta zajedničkih prijatelja, u vrlo maloj skupini ljudi može doći do visokog stupnja grupiranosti. Kad bismo nasumično odabrali par osoba, vrlo je vjerojatno da bi povezanosti između njih bila mala do nikakva.

[13] http://mathinsight.org/degree_distribution pristupano 28.05 .2017
[14] https://networkscience.wordpress.com/2013/09/08/defining-the-clustering-coefficient/ pristupano 28.05.2017

Jedan od načina na koji možemo provjeriti grupiranost jest brojanje trokuta između određenih čvorova. Postoje dvije varijante mjerenja koje koriste drugačije metode. Prva metoda računa globalni koeficijent grupiranosti na način da gleda 3 međusobno povezana čvora kao trojku. Slika 9 prikazuje kako izgleda obična i zatvorena trojka.

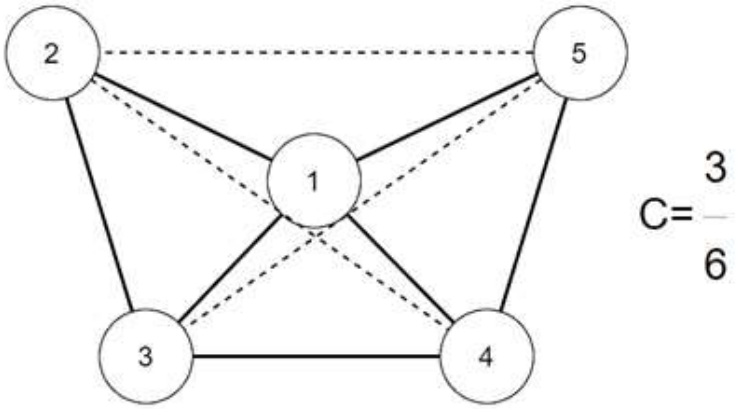
Slika 9. Primjer trojke i zatvorene trojke



Izvor: Izrada autora

Ako imamo tri zatvorene trojke onda se radi o trokutu, naravno ako gledamo da je središte svake trojke po jedan čvor u trokutu. Koeficijent računamo tako da podijelimo broj zatvorenih trojki sa sveukupnim brojem trojki.

Slika 10. Primjer lokalnog koeficijenta



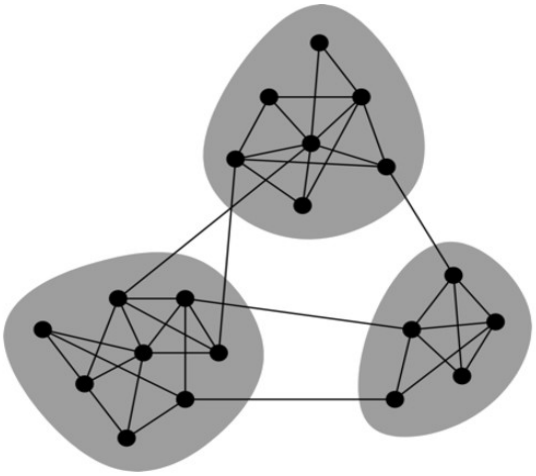
Izvor: Izrada autora

Druga metoda računa lokalni koeficijent grupiranosti grafa. U ovoj metodi imamo središnji čvor koji je povezan s drugim čvorovima. Trebamo prebrojati međusobne veze ostalih čvorova, ne brojeći veze sa središnjim čvorom i onda podijeliti sa svim mogućim vezama. Na slici 10 vidimo veze kao pune crte dok su moguće veze isprekidane. C je koeficijent grupiranosti.

3.2.3 Struktura zajednice

Za graf se kaže da ima strukturu zajednice kada ima grupe čvorova koje su unutarnje jako povezane, dok su grupe u mreži međusobno slabo povezane[15]. Takvu vrstu povezanosti možemo vidjeti na slici 11. Kod kompleksnih mreža postoji veća vjerojatnost da će čvorovi biti povezani ako pripadaju istoj zajednici tj. manje povezani ako ne pripadaju istoj zajednici.

Slika 11. Primjer zajednica u grafu



Izvor: http://www.pnas.org/content/103/23/8577.full

U stvarnom svijetu dosta često nailazimo na ovakve zajednice. Isto tako možemo reći za društvene mreže, u kojima se korisnici povezuju na temelju geografskog položaja i zajedničkih interesa.

Postoje dvije vrste kompleksnih mreža koje su znanstvenici istraživali, a to su nerazmjerne mreže i mreže malog svijeta. Jedna od zajedničkih karakteristika jest što im distribucija stupnja povezanosti poprima oblik Paretove distribucije.

3.2.4 Nerazmjerne mreže

U nerazmjernoj mreži gdje prevladava paretova distribucija stupnja povezanosti, neki čvorovi znaju imati stupanj povezanosti koji je daleko iznad prosjeka. U takvim slučajevima te čvorove zovemo koncentratorima (eng. Hub)[16]. Ne postoji definirana granična razina koju čvor mora proći da bi bio koncentrator, jer bi postojanjem takve razine mreža postala razmjerna.

[15] http://www.pnas.org/content/99/12/7821.full pristupano 29.05.2017
[16] http://barabasi.com/f/124.pdf pristupano 29.05.2017

Jedna od prednosti ovakve mreže jest otpornost na kvarove. Postoji hijerarhija gdje su na najveće hubove spojeni manji hubovi, a na te manje još manji hubovi.
Ako dođe do nasumičnog kvara, zbog oblika distribucije, velika je vjerojatnost da će pogoditi čvor koji ima manju povezanost. U slučaju da pogodi koncentrator, ne znači da će povezanost mreže pasti jer postoje drugi koncentratori. Problem leži u najvećim koncentratorima. Ako oni ispadnu iz mreže postoji mogućnost raspada i nastajanja manjih izoliranih grupa.

3.2.5 Mreže malog svijeta

Mreže malog svijeta predstavljaju mreže u kojima svi čvorovi nisu nužno povezani, iako se do svakog čvora može doći kroz par koraka. U ovakvim mrežama, prosječna udaljenost dvaju nasumično odabranih čvorova raste proporcionalno logaritmu količine čvorova u mreži[17].
Karakteristike poput velikog koeficijenta grupiranosti i postojanja koncentratora pokazuju sličnost s nerazmjernim mrežama. U mrežama malog svijeta imamo prekomjernu količinu koncentratora koji služe kao veze preko kojih čvor može naći najkraću putanju do drugog čvora. Zanimljiva teorija u mrežama malog svijeta jest šest stupnjeva odvojenosti[18]. Ideja te teorije jest da su sve stvari na svijetu 6 koraka ili manje udaljene jedne od drugih. Osim testiranja u stvarnom svijetu, mnogo društvenih mreža je testiralo ovu teoriju te su došli do zaključka da jedan korisnik stvarno može doći do bilo kojeg drugog korisnika u za 6 koraka ili manje. Rastom društvenih mreža, prosječan broj koraka opada jer postoji sve više koncentratora preko kojih mogu naći najkraći put.

[17] http://info-centre.jenage.de/assets/pdfs/library/newman_SIAM_REV_2003.pdf pristupano 29.05.2017
[18] http://whatis.techtarget.com/definition/six-degrees-of-separation pristupano 29.05.2017

4. Alati i tehnike za vizualizaciju kompleksnih mreža

Nakon što smo objasnili što su kompleksne mreže, kako funkcioniraju, kakve vrste postoje i što očekujemo od njihovih prikaza, možemo obraditi određene alate i tehnike za njihov prikaz. Zbor širokog izbora dostupnih podataka, nije praktično koristiti samo jedan alat ili samo jednu vrstu prikaza tih podataka. Nekada se podaci najbolje mogu prikazati jednom tehnikom ili u jednom alatu, dok drugi nemaju toliko ograničenja pa imamo slobodan izbor alata.

4.1 Alati za vizualizaciju kompleksnih mreža

Alata za prikaz kompleksnih mreža ima previše stoga ću odabrati par najkorištenijih u ovom području i one koje ću koristiti za praktičan dio rada.

4.1.1 R i RStudio

Jedan od najkorištenijih jezika u ovom području jest R[19] koji se većinom koristi za statističke obrade i vizualizaciju tih istih obrada. Radi se o besplatnom alatu koji je vrlo sličan S[20] po načinu rada. Vrlo brzo i efikasno pretvara ideje u softver. R nam pruža vrlo mnogo statističkih metoda kao što su:

1. Linearno i nelinearno modeliranje
2. Klasični statistički testovi
3. Analiza vremenskih serija
4. Klasifikacija
5. Grupiranje

Jedna od najboljih karakteristika ovog alata jest lakoća generiranja visokokvalitetnih grafičkih prikaza, uključujući ubacivanje matematičkih simbola i formula po potrebi. Jednostavan je za upotrebu zbog kvalitetnih zadanih postavki kojima možemo dobiti izvrsne rezultate. U slučaju da nam se ne sviđaju zadane postavke, korisnici imaju hrpu priručnika i literature o korištenju alata kako bi mogli prilagoditi grafičke prikaze svom ukusu.

[19] https://www.r-project.org/ pristupano 20.05.2017
[20] http://ect.bell-labs.com/sl/S/ pristupano 21.05.2017

Osim što sam alat nudi mnogo opcija, ako nam dodatno zatreba još opcija uvijek imamo mogućnost dodavanja paketa. Ti paketi sadržavaju nove funkcionalnosti kojima možemo lakše izračunati, analizirati ili prikazati neke podatke.
R sam po sebi nije toliko efikasan za rad jer dolazi bez grafičkog sučelja. Na svu sreću danas imamo par sučelja kojima možemo olakšati rad u R-u. Jedan od najpoznatijih je RStudio[21] koji je besplatan za osobnu upotrebu. U slučaju da ga želimo koristiti u komercijalne svrhe, postoji i plaćena inačica navedenog softvera koja nudi više mogućnosti i bolju korisničku podršku. Cilj ovog sučelja je da nam olakša dodavanje i manipulaciju podataka, korištenje paketa, razumijevanje koda i slično.

4.1.2 Gephi

Gephi[22] je besplatan alat koji je usko specijaliziran za analizu i vizualizaciju mreža. Koristi 3D metode za prikaze velikih podataka u stvarnom vremenu koje isto tako možemo u stvarnom vremenu istražiti. Takav način 3D obrade koristi grafičku karticu umjesto procesore, kao i današnje računalne igrice . Neke od mogućnosti su:

1. Istraživanje podataka
2. Analiza podataka
3. Filtriranje podataka
4. Grupiranje podataka
5. Manipulacija podataka

Gephi se bazira na vizualnom i manipulativnom principu rada koji omogućava korisnicima otkrivanje mreža i njihova svojstava. Svojim korisnicima dozvoljava visoku podesivost opcija mrežnih prikaza i rada algoritama. Osim što pruža interaktivan način rada i prikaza podataka, podatke isto tako može prikazati u standardnim statičkim prikazima.

[21] https://www.rstudio.com/ pristupano 23.05.2017
[22] https://gephi.org/ pristupano 23.05. 2017

4.1.3 D3

Data-Driven Documents ili skraćeno D3[23] je JavaScript biblioteka za manipulaciju dokumenata koji su bazirani na podacima. D3 se najčešće koristi za izradu dinamičnih i interaktivnih prikaza podataka kojima se može pristupiti preko web preglednika. Kako se radi o online načinu rada, ovaj alat koristi tehnologije kao što su HTML, CSS i SVG. Na tisuće stranica koristi D3 kako bi kreirale interaktivne prikaze podataka. Kao primjer bismo uzeli HTML tablicu generiranu od određenih podataka. Zatim podatke iz tablice možemo pretvoriti u interaktivan grafikon kojim bi lakše prikazali željene podatke.

D3 je ekstremno brz alat u svom području jer ne pokušava ponuditi sve moguće opcije nego se fokusira na bitne dijelove kao što je podržavanje velike količine podataka. Tim podacima u D3-u možemo vrlo brzo napraviti dinamičnu animaciju ili interaktivan prikaz.

Podešavanje najčešće spada u kategoriju stiliziranja kako se radi o CSS-u. S obzirom na veliku kolekciju radova korisnika, na službenoj D3 web stranici možemo naći razne primjere i njihov kod kojim se možemo poslužiti u unaprjeđenju vlastitih projekata.

4.1.4 Cytoscape

Kao i ostali navedeni alati u ovom radu, Cytoscape[24] je besplatan alat za vizualizaciju mreža. Cytoscape je usko specijalizirani alat za vizualizaciju molekularne interakcije i bioloških puteva. Osim prikaza možemo integrirati razne podatke kao što su razne anotacije ili genetski zapisi.

Iako je ovaj alat prvenstveno osmišljen za korištenje u uskom specijaliziranom području, korisnici su ga kroz godine počeli koristiti za sve moguće vizualizacije kompleksnih mreža.

Cytoscape dolazi s osnovnim mogućnostima za integraciju, analizu i vizualizaciju podataka. Dodatne mogućnost nalaze se u obliku aplikacija koje možemo naknadno instalirati. Tim aplikacijama možemo dodati nove funkcionalnosti kao što su:

1. Naprednije analize molekula i mreža
2. Nove vrste vizualizacija
3. Bolja podrška formata podataka
4. Povezanost s bazama podataka

[23] https://d3js.org/ pristupano 25.05.2017
[24] http://www.cytoscape.org/ pristupao 25.05.2017

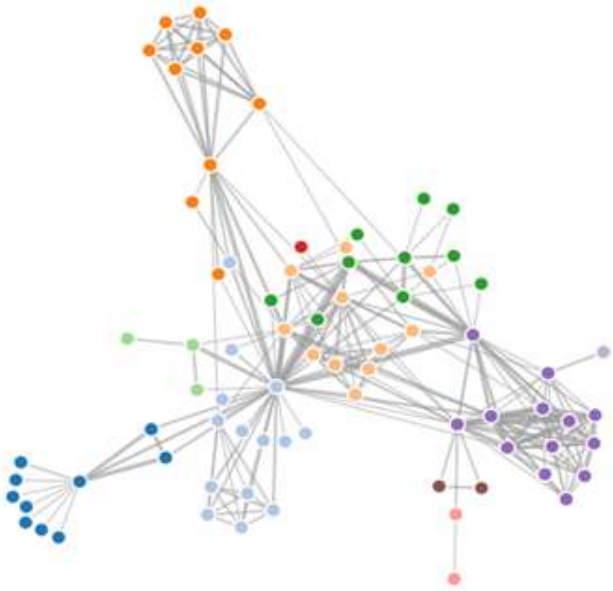
4.2 Tehnike vizualizacija kompleksnih mreža

Nakon što smo obradili alate, pojasnit ćemo par najčešćih tehnika za vizualizaciju koje se danas koriste. Cilj vizualizacije jest da vizualno možemo vidjeti što se događa unutar mreže. Kako bismo najbolje procjenu trebamo imati vizualizaciju koja je najprikladnija korištenim podacima. Za početak ćemo obraditi vrstu prikaza koju smo već spomenuli u prošlom poglavlju.

4.2.1 Čvor-brid dijagram

S ovakvim prikazom smo već upoznati, radi se o najjednostavnijoj tehnici prikaza mreža.
Unutar ovako jednostavne tehnike postoji mnogo načina na koje možemo prikazati mrežu.
Prvi problem na koji nailazimo jest kompleksnost mreže. Loše prikazana kompleksna mreža može izgubiti svo značenje. Da bismo urednije prikazali kompleksnu mrežu, služimo se silom usmjerenim crtanjem grafova. Takvi algoritmi vizualiziraju mreže na način da budu ugodnije oku.

Slika 12. Primjer silom usmjerenog grafa



Izvor: https://bl.ocks.org/mbostock/4062045

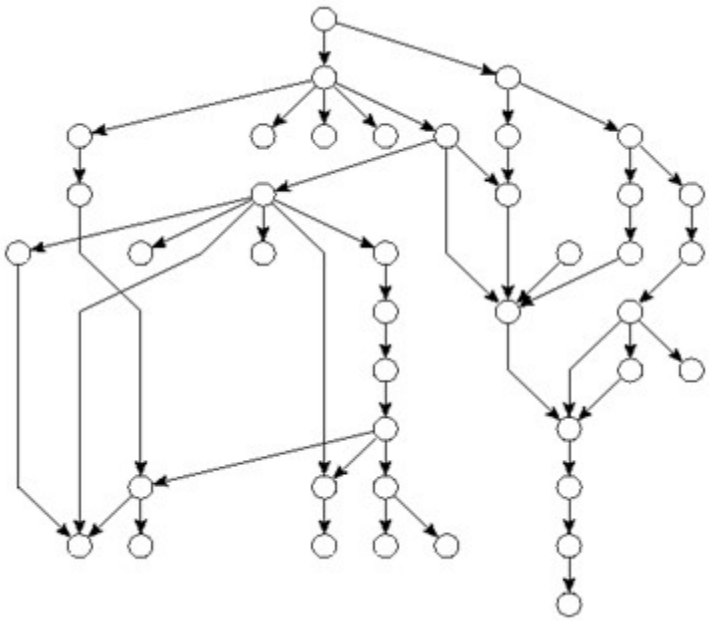
Algoritam izvršava zadatak tako da čvorove mreže postavlja u dvodimenzionalna ili trodimenzionalna polja gdje je udaljenost između svih čvorova više-manje jednaka, tj. gdje je duljina bridova više-manje jednaka[25]. Kao primjer možemo uzeti sliku 12 gdje su svi čvorovi relativno jednako udaljeni.

[25] https://cs.brown.edu/~rt/gdhandbook/chapters/force-directed.pdf pristupano 26.05.2017

Kako algoritmi ne mogu sve čvorove postaviti na optimalne pozicije, pokušavaju što manje bridova sjeći u mreži. Algoritam svim čvorovima i bridovima dodijeli određenu silu, zatim se elementi u mreži privlače i odguruju. Ovaj proces se ponavlja sve dok svi elementi ne dođu do mehaničke ravnoteže gdje algoritam zaključuje da im je to finalna pozicija.

Osim takvih algoritamskih prikaza postoje i drugi načini prikaza unutar ove tehnike. Jedna od tih vrsta jest hijerarhijski način crtanja kojim čvorove prikazujemo u horizontalnim redovima dok bridove obično crtamo vertikalno i to prema dolje[26]. Ovakav graf većinom počinjemo crtati od dna prema vrhu kako bi svi smjerovi bridova bili u istom smjeru i da se ne bi sjekli. Kako mreže znaju biti nasumične i mogu sadržavati petlje, ova metoda može vrlo neuspješna prikazivanju podataka ako mreža nema jaku hijerarhiju. Kao što vidimo na slici 13, dolazi do presječenih bridova iako se radi o jednostavnijoj mreži.

Slika 13. Primjer hijerarhijskog grafa



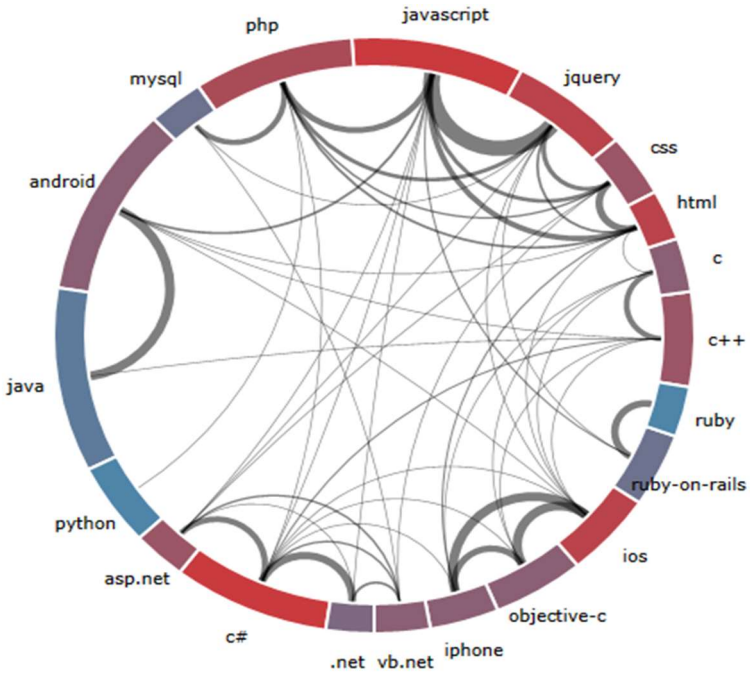
Izvor: http://goblin2.sourceforge.net/refman/pageGraphLayout.html

[26] http://goblin2.sourceforge.net/refman/pageGraphLayout.html pristupano 26.05.2017

Iako postoji još mnogo verzija ove tehnike prikazivanja, za sada ćemo objasniti još dvije koje se najčešće koriste:

1. Cirkularna metoda – Čvorovima kreira kružnicu dok se bridovi prikazuju kao veze unutar kružnice. Kao primjer će poslužiti slika 14 koja prikazuje rezultate 1000 pitanja sa stranice Stack Overflow[27]. Slika je preuzet sa stranice Codeproject[28]. Na grafu možemo vidjeti koje su se tehnologije najviše spominjale dok veze predstavljaju koliko su se često zajedno spominjale određene tehnologije.

Slika 14. Cirkularni prikaz

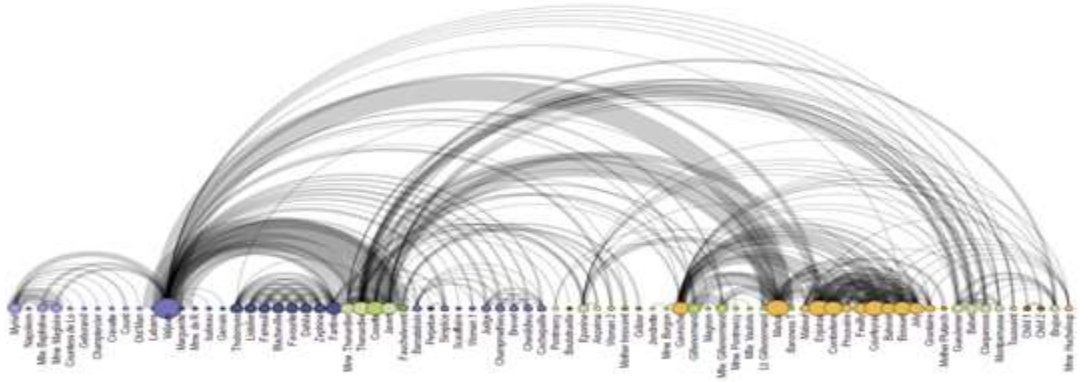


Izvor: https://www.codeproject.com/Articles/342715/Plotting-Circular-Relationship-Graphs-with-Silverl

2. Metoda prikazivanja pomoću lukova – Čvorove postavljamo u ravnini dok veze prikazujemo lukovima između zadanih čvorova. Slika 15 je preuzet sa stranice udruge za računalne znanosti sveučilišta u Washingtonu[29]. Na ovom grafu čvorovi predstavljaju likove knjige „Les Misérables“[30] a veze, tj. lukovi predstavljaju međusobnu interakciju.

[27] https://stackoverflow.com/ pristupano 26.05.2017
[28] https://www.codeproject.com pristupano 27.05.2017
[29] https://www.cs.washington.edu/ pristupano 27.05.2017
[30] http://www.goodreads.com/book/show/24280.Les_Mis_rables pristupano 27.05.2017

Slika 15. Prikaz pomoću lukova

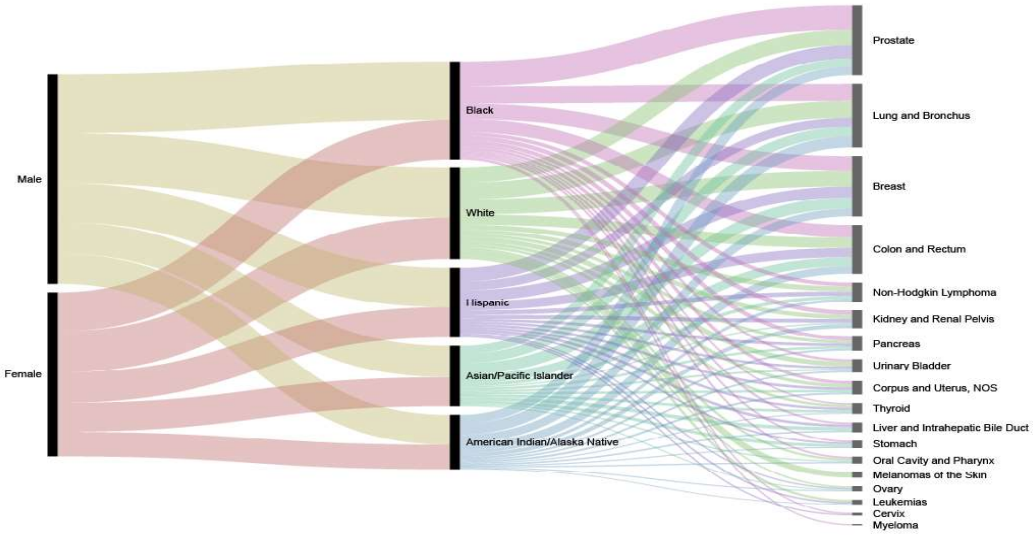


Izvor: https://homes.cs.washington.edu/~jheer/files/zoo/ex/networks/arc.html

4.2.2 Aluvijalni dijagram

Za aluvijalni dijagram moglo bi se reći da je vrsta dijagrama toka. Ovim dijagramom možemo prikazati tok podataka u određenom vremenskom periodu i količinu interakcije između entiteta tj. čvorova[31]. S lijeve i desne strane imamo entitete koji su povezani vezama. Te veze predstavljaju tok dok širina te veze prikazuje količinu protoka kroz vrijeme. Još jedna prednost ovog dijagrama jest jednostavan prikaz interakcije između entiteta, jer grananjem toka možemo vidjeti povezanost među njima. Na slici 16 imamo prikaz oboljelih od vrsta raka prema spolu i rasi.

Slika 16. Prikaz aluvijalnog dijagrama



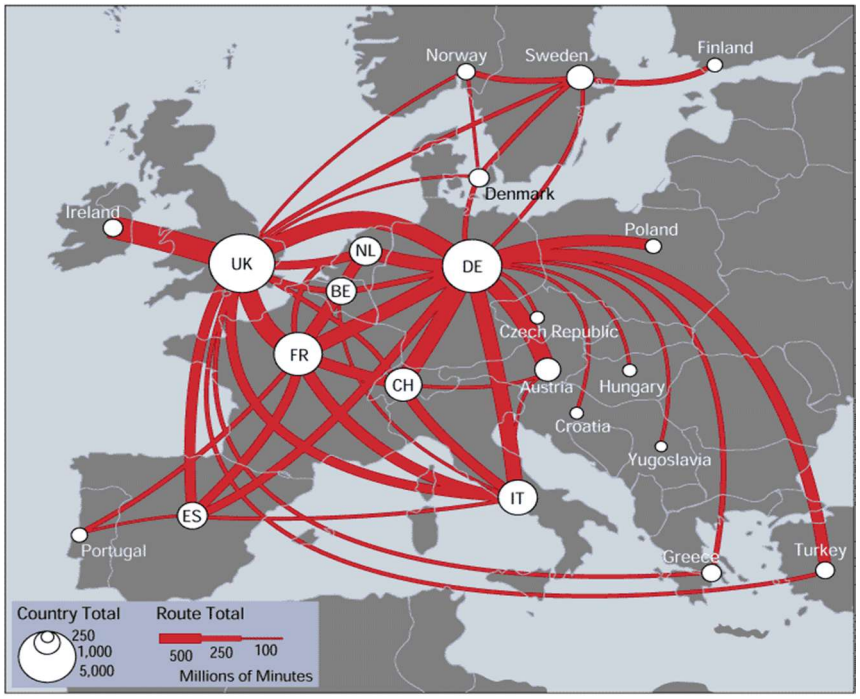
Izvor: http://digitalsplashmedia.com/2014/06/visualizing-categorical-data-as-flows-with-alluvial-diagrams/

[31] http://www.datavizcatalogue.com/methods/parallel_sets.html pristupano 27.05.2017

4.2.3 Vizualizacija geografskih sustava.

Ovo je jedan on najboljih primjera kako vizualizacija može pomoći pri razumijevanju kompleksnih mreža. Geografski informacijski sustavi se koriste u svakodnevnom životu, bilo da se radi o GPS navigaciji ili o praćenju pošiljke preko interneta. U navedenim primjerima još možemo proći bez dobrog vizualnog primjera, ali problem nastaje pri većoj količini podataka. U tako kaotičnom sustavu nemoguće je razumjeti podatke i primijetiti trendove bez detaljnijih analiza. Najlakša metoda prikaza jest običnim vezama gdje povezujemo lokacije i samim time možemo vidjeti koje sve rute postoje i koje su najkorištenije. Kako takva mreža nekada zna biti neuredna, veze možemo zamijeniti tokovima kojima širina označava korištenost određene rute[32]. Tokovi mogu pomoći pri prikazivanju letova aviona ili plovidba brodova jer se ti podaci znaju analizirati na mjesečnoj ili na godišnjoj bazi.

Slika 17. Prikaz telekomunikacijske mreže



Izvor: http://mappa.mundi.net/maps/maps_014/

Na slici 17 možemo vidjeti količinu komunikacije između čvorova tj. lokacija na karti. Ovako pojednostavljenim prikazom možemo lako zaključiti između kojih država postoji najviše komunikacije.

[32] http://www.datavizcatalogue.com/methods/flow_map.html pristupano 27.05.2017

4.3 Smjernice za vizualizaciju

Uz primjere koje smo prošli u ovom radu, proučavali smo kakve sve mreže postoje, kako funkcioniraju, kako ih možemo vizualizirati, ali ne kako bi ih trebali prikazati. Iako se kompleksne mreže mogu svrstati u razna znanstvena područja kao što prirodne ili tehničke znanosti, dobar prikaz mreže još uvijek treba površno razumijevanje prikaza detalja iz umjetničke perspektive. Koliko god vizualizacije bila detalja ili točna, ako ju loše prikažemo postoji mogućnost da ciljana publika neće razumjeti što je autor želio reći. Prije nego što počnemo projekt, trebao razmisliti o tome što želimo prikazati i kome će taj prikaz biti od koristi. Zbog popularnosti vizualizacije, postoji previše razloga zašto bi netko želio vizualizirati mrežu. Danas dosta korisnika ima umjetnički pristup vizualizaciji mreža, jer im nije cilj što detaljnije i preciznije objasniti što se događa nego na estetski ugodan način prikazati aktivnosti u mreži. Osim takvog pristupa, podijelio bih vizualizaciju na dva konkretna pristupa:

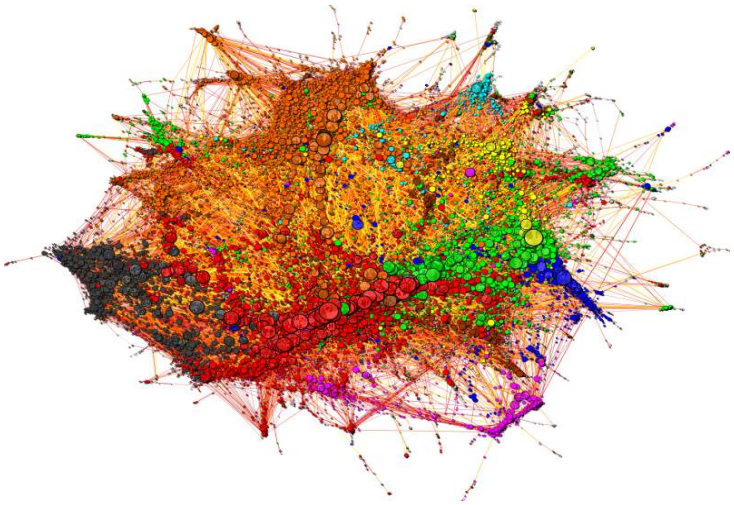
1. Obrazložavajući pristup – Pristup u kojem želimo što jednostavnije ciljanoj publici objasniti podatke koje imamo. U ovakvom pristupu želimo što efikasnije odgovoriti na postavljena pitanja, bez da korisnici moraju previše razmišljati. Ovaj pristup se većinom koristi u poslovnom okruženju gdje nemamo puno vremena za objašnjavanje. Većinom se prikazuju jednostavni grafovi ili grafikoni u kojima nema toliko detalja, minimalistički stil prikazivanja je ključni detalj lako razumljive prezentacije.

2. Istraživački pristup – U ovom pristupu imamo puno više detalja nego u prošlom. Cilj ovog pristupa nije da odmah razumijemo što se događa, nego da istražimo vizualizaciju, da si postavljamo pitanja i da eventualno nađemo odgovore na ta pitanja. U ovakvim okolnostima možemo koristiti kompleksnije prikaze, više izvora podataka ili veću količinu detalja. Ovakve vizualizacije možda nisu najprikladnije za poslovna okruženja jer korisniku treba vremena da bi došao do željene informacije. Iz tog razloga se više koriste u znanstvenim istraživanjima.

Bez obzira na pristup koji koristimo, postoji par pravila tj. smjernica koje ne bi bilo loše pratiti ako želimo što kvalitetnije vizualizirati mrežu. Iako u prikazivanju imamo dosta slobode što se tiče metoda i tehnika, također postoje razlozi zašto bi se za vizualizaciju reklo da je loša.

1. Upotrebljivost – Procijeniti radi li se o kvalitativnom ili o kvantitativnom prikazu. Ova stavka se može povezati s gore navedenim pristupima vizualizacije. Treba procijeniti ciljanu publiku i napraviti vizualizaciju koja odgovara situaciji. Koliko god ponosni bili na svoj radi, nepotrebno kompliciranje može dovesti do loših kritika.

Slika 18. Prikaz kompleksne mreže



Izvor: http://drewconway.com/zia/2013/3/29/the-value-of-edges-in-complex-network-visualization

Iz primjera kao što je slika 18, vrlo teško možemo izvući korisne informacije bez da potratimo previše vremena.

2. Integritet – Pri izradi vizualnog prikaza, trebamo biti konzistentni u korištenju određenih mjera ili pokazala. Umjesto da prikaz bude umjetnička interpretacija, želimo prikazati podatke onakvima kakvi jesu. Treba se držati istih dimenzija kako ne bi stvorili lažnu sliku o podacima. Pomicanjem određenih elemenata u grafu možemo primijeniti cijelu percepciju, što može uzrokovati krive zaključke o podacima.

3. Udio i teksta i slika – Iako želimo što više informacija prenijeti preko vizualnog prikaza, moramo paziti da ne dođe do zasićenja. Treba znati koji je najbolji omjer preko kojeg možemo prenijeti poruku. Iako je ova smjernica dosta objektivna jer za vizualizaciju ne treba puno teksta, isto bi trebalo pripaziti s nazivima čvorova ili veza ako radimo s kompleksnim podacima.

4. Estetika – Pod estetiku ne spada samo koliko lijepo izgleda vizualizacija, nego i jednostavnost načina na kojeg možemo prenijeti informaciju.

Osim prije navedenih kriterija isto treba pripaziti na neke manje detalje, kao što su :

1. Oblik elemenata – Procijeniti koji oblik čvorova ili veza najviše odgovara vrsti prikaza. Ako se radi o kompleksnim podacima, možda bismo mogli odabrati manje čvorove i tanje veze kako ne bi postalo nečitljivo.

2. Boje – Koliko god lijepo izgledalo kada iskoristimo cijelu paletu raspoloživih boja, nekada možemo efikasnije prikazati informacije ako koristimo manje boja. Ako se radi o prikazu grupiranja u mreži, bilo bi prikladno koristiti onoliko boja koliko ima grupa. U slučaju da ima previše grupa, može se mijenjati nijansa određene boje po grupi ako ne želimo da postane prešareno. Isto tako treba pripaziti na kontrast između odabranih boja, previše žarkih boja može napraviti prikaz vizualno odbojnim.

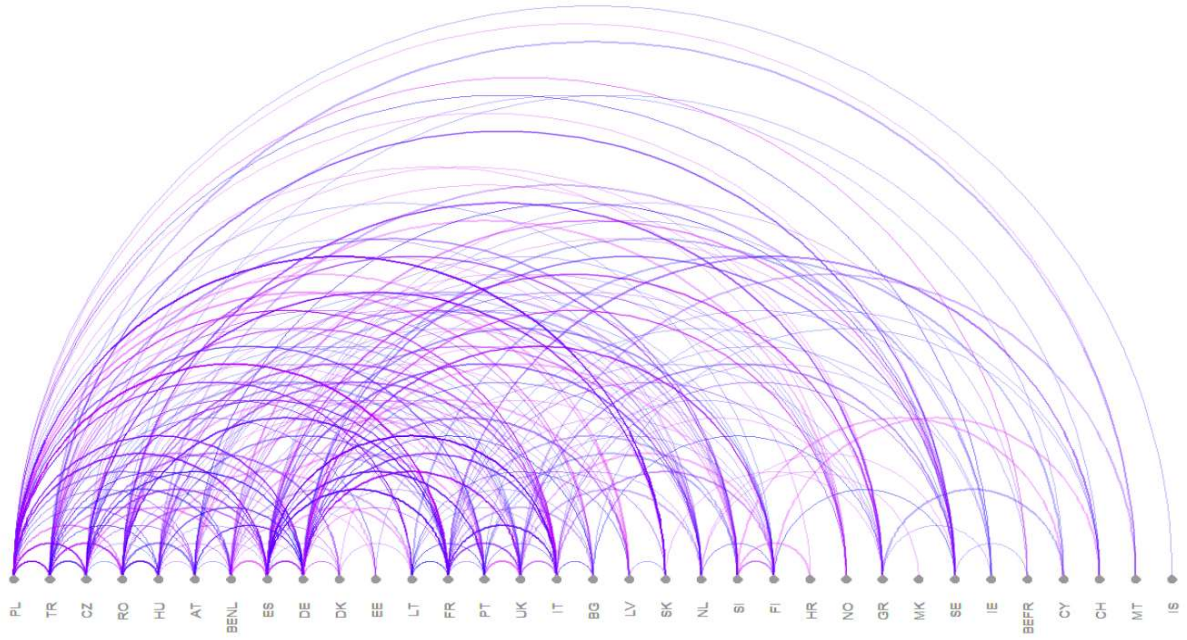
5. Praktičan primjer vizualizacije kompleksne mreže

Nakon što smo prošli bitne stavke za vizualizaciju, u ovom poglavlju ćemo demonstrirati te metode kroz praktičan primjer. Primjer je napravljen u programskom jeziku R, dok je RStudio korišten kao grafičko sučelje.

5.1 Primjer vizualizacije pomoću lukova

U ovom primjeru prikazana je mobilnost profesora preko Erasmus programa u akademskoj godini 2012/1013. Radi se o otvorenim podacima koji se mogu naći na portalu otvorenih podataka Europske unije[33]. Uz pomoć lukova možemo prikazati kako su profesori putovali i koje su zemlje najviše uključene u Erasmus programu.

Slika 19. Vizualizacija mobilnosti profesora



Izvor: Izrada autora

Kako je projekt napravljen u R-u, nije potrebno puno koda da bi došli do vizualnog prikaza kao što je slika 19. Bitno je odabrati prikladan paket koji će omogućiti prikaz podataka. Većinom ti paketi znaju bez ikakvih dodatnih parametara izraditi dovoljno dobar prikaz. Ako nam se ne sviđa uvijek možemo modificirati.

[33] https://data.europa.eu/euodp/en/data/dataset/erasmus-mobility-statistics-2012-13 pristupano 02.06.2017

U programskom kodu 1 imamo instalaciju i učitavanje paketa koji su potrebni za izradu prikaza.

Programski kod 1. Korišteni paketi

```
############### ucitavanje paketa ################
library(igraph)
library(githubinstall)

######## instalacija i ucitavanje arcdiagram ###############
githubinstall('arcdiagram', username='gastonstat')
library(arcdiagram)
```

Izvor: Izrada autora

U ovom slučaju pakete „igraph“ i „githubinstall“ možemo instalirati preko grafičkog sučelja tako da pod opciju „tools“ kliknemo na „Install Packages“. Paket „arcdiagram[34]“ ne možemo tako instalirati pa ćemo morati koristiti paket za instalaciju paketa s Github-a[35]. Nakon što ga instaliramo, možemo ga učitati i nastaviti s radom.

Programski kod 2. Čišćenje podataka

```
################# ciscenje podataka ##########################
pod <- prof[sample(1:nrow(prof),500, replace=FALSE),]
podaci <- pod[,c(3,13)
```

Izvor: Izrada autora

Sad možemo unijeti Erasmus podatke u RStudio kako bismo mogli nastavi rad s njim. Nakon toga u varijabli „pod“ radimo nasumično uzorkovanje od 500 uzoraka jer je rad s originalnom datotekom dosta zahtijevan. Uzimamo manji uzorak jer nam ova vizualizaciju služi kako demonstracija alata, a ne kao informacija za daljnju analizu. Nakon toga uzimamo stupce koji su nam potrebni u varijabli „podaci“, a to su zemlja institucije u kojoj profesor radi i zemlja institucije u koju ide.

[34] https://github.com/gastonstat/arcdiagram pristupano 05.06.2017

[35] https://github.com/ Pristupano 06.06.2017

Programski kod 3. Spremanje podataka

```
############## spremanje podataka u prikladnom formatu ###############
graf <-graph.data.frame(podaci, directed=FALSE)
write.graph(graph = graf, file="graf.gml", format = "gml")
tips_prof <-read.graph("/graf.gml", format = "gml")
prof_edge = get.edgelist(tips_prof)
```

Izvor: Izrada autora

Kako radimo s „igraph“ paketom, trebamo podatke modificirati da bi bili prikladni paketu za korištenje. Podatke spremamo u formatu koji je prikladniji za izradu grafova. U tom formatu su posebno deklarirani svi čvorovi i bridovi. Kada učitavamo podatke, potrebno je unijeti lokaciju tražene datoteke. U programskom kodu 3 imamo „read.graph“ koji zahtijeva da u navodnike unesemo gdje se nalazi.

Programski kod 4. Izrada prikaza

```
######################### izrada prikaza ############################
arcplot(prof_edge,
    lwd.arcs=runif(20,1,2.5), col.arcs=hsv(runif(10,0.65,0.82),alpha=0.3),
    show.nodes=TRUE, pch.nodes=10,
    col.nodes="gray60", lwd.nodes=4)
```

Izvor: Izrada autora

U programskom kodu 4 radimo na izradi vizualnog prikaza mreže. Prikaz se može obaviti samo s “arcplot“ funkcijom bez dodatnih varijabli. U ovom slučaju želio sam promijeniti neke stvari na prikazu, pa sam dodao par modifikacija:

1. ldw.arcs – Određujemo širinu luka, dodana je „runif“ funkcija koja nasumično generira brojeve da svi lukovi nebi bili iste širine.
2. col.arcs – Određujemo boju lukova, također je korišteno nasumično odabiranje boja. Osim toga imamo parametar „alpha“ koji određuje prozirnost lukova.
3. pch.nodes –Biramo oblik čvorova
4. col.nodes –Biramo boju čvorova
5. lwd.nodes –Biramo veličinu čvorova

6. Zaključak

Na kraju ovog rada možemo vidjeti koje su sve osnove potrebne za razumijevanje i vizualizaciju kompleksnih mreža. Najvažniji dio izrade vizualnog prikaza jest razumijevanje potreba projekta i izrada odgovarajućeg prikaza. Zbog jednostavnosti današnjih alata, možemo se više posvetiti odabiru i analizi mreže koju ćemo vizualizirati, a ne kako ćemo prikazati. Vizualizaciju ne bi trebalo gledati kao glavni dio posla, nego kao alat koji će nam omogućiti da jednostavnije obavimo traženi posao. Bilo to prezentiranje na fakultetu ili na poslu, ili možda istraživanje društvenih mreža s ciljem otkrijemo nešto novo u njima. Ako ne znamo što tražimo ili što želimo prikazati, ni najbolje vizualizacije nam neće pomoći.

Nakon što razumijemo što želimo iz podataka, možemo se posvetiti odabiru najprikladnijih tehnika za vizualizaciju mreža. U današnje vrijeme alati ne igraju veliku ulogu u izvedbi jer dosta njih ima slične mogućnosti. Osim ako se radi o nekom specijaliziranom projektu, alat koji odaberemo neće toliko utjecati na konačni proizvod. Zbog učestalog korištenja veće količine alata za jedan projekt, neke su kompanije razvile pakete kako bi se njihov stil vizualiziranja mogao dobiti u drugim alatima, što bi znatno olakšalo posao korisniku kada bi morao koristiti samo jedan alat. Zbog veće mogućnosti može doći do otežanih odluka vezanih za izradu vizualnog prikaza. Upravo zbog svih tih paketa korisnik sve teže odabire koji paket koristiti. Zato bi se uvijek trebao izraditi što jednostavniji prikaz, zatim naknadno dodati elementi koji su potrebni za poboljšanje vizualnog prikaza.

Literatura

1. Lima M. (2011) Visual Complexity – Mapping Patterns of Information, Princeton Architectural Press http://www.scottbot.net/HIAL/index.html@p=39166.html (pristupano 10.05.2017)
2. Friendly M. (2006) A Brief History of Data Visualization, York University Dostupno na: http://www.datavis.ca/papers/hbook.pdf (pristupano 10.05.2017)
3. Mapping Through the Ages: The History of Cartography, dostupno na: https://www.gislounge.com/mapping-through-the-ages/ (pristupano 10.05.2017)
4. Diagrams of knowledge, dostupno na: http://www.scottbot.net/HIAL/index.html@p=39166.html (pristupano 12.05.2017)
5. Big Data, What it is and why it matters, dostupno na: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html pristupano (12.05.2017)
6. What Does Big Data Look Like? Visualization Is Key for Humans, dostupno na: https://www.wired.com/insights/2014/01/big-data-look-like-visualization-key-humans/ pristupano (12.05.2017)
7. The Top 5 Business Benefits of Using Data Visualization, dostupno na: http://data-informed.com/7-great-visualizations-history/ (pristupano 12.05.2017)
8. Twitter www.twitter.com (pristupano 15.05.2017)
9. Twitter Alas, dostupno na: https://twitter.com/TwitterAlas/status/712058816041345024 (pristupano 15.05.2017)
10. The Benefits and Future of Data Visualization, dostupno na: https://www.statsilk.com/blog/benefits-and-future-data-visualization (pristupano 15.05.2017)
11. Few S. Edge P. (2007) Data Visualization Past, Present, And Future, dostupno na: https://www.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf (pristupano 15.05.2017)
12. Introduction to Graph Theory dostupno na: http://www.math.fsu.edu/~pkirby/mad2104/SlideShow/s6_1.pdf (pristupano 16.05.2017)
13. Weighted Graphs, dostupno na: http://courses.cs.vt.edu/~cs3114/Fall10/Notes/T22.WeightedGraphs.pdf (pristupano 16.05.2017)

14. Grafovi, dostupno na:
https://www.fer.unizg.hr/_download/repository/Osnovni_pojmovi-teorija_grafova.pdf
(pristupano 17.05.2017)
15. The degree distribution of a network, dostupno na:
http://mathinsight.org/degree_distribution (pristupano 28.05 .2017)
16. Defining the Clustering Coefficient, dostupno na:
https://networkscience.wordpress.com/2013/09/08/defining-the-clustering-coefficient/
(pristupano 28.05.2017)
17. M. Girvan, M. E. J. Newman (2002) Community structure in social and biological networks
dostupno na: http://www.pnas.org/content/99/12/7821.full (pristupano 29.05.2017)
18. Scale free networks, dostupno na: http://barabasi.com/f/124.pdf pristupano 29.05.2017
19. M. E. J. Newman (2003) The structure and function of complex networks, dostupno na:
http://info-centre.jenage.de/assets/pdfs/library/newman_SIAM_REV_2003.pdf
(pristupano 17.05.2017)
20. Six degrees of separation, dostupno na:
http://whatis.techtarget.com/definition/six-degrees-of-separation (pristupano 29.05.2017)
21. Remco van der Hofstad - Random Graphs And Complex Networks, dostupno na:
https://www.win.tue.nl/~rhofstad/NotesRGCN.pdf (pristupano17.05.2017)
22. R-project https://www.r-project.org/ (pristupano 20.05.2017)
23. The S System http://ect.bell-labs.com/sl/S/ pristupano (21.05.2017)
24. RStudio https://www.rstudio.com/ (pristupano 23.05.2017)
25. Gephi https://gephi.org/ (pristupano 23.05. 2017)
26. D3 https://d3js.org/ pristupano (25.05.2017)
27. Cytoscape http://www.cytoscape.org/ pristupao (25.05.2017)
28. Force-Directed Graph, dostupno na: https://bl.ocks.org/mbostock/4062045
pristupano (26.05.2017)
29. Stephen G. Kobourov; Force-Directed Drawing Algorithms, dostupno na:
https://cs.brown.edu/~rt/gdhandbook/chapters/force-directed.pdf (pristupano 26.05.2017)
30. Graph layout, dostupno na:
http://goblin2.sourceforge.net/refman/pageGraphLayout.html (pristupano 26.05.2017)
31. Stack Overflow, dostupno na: https://stackoverflow.com/ (pristupano 26.05.2017).

31. Codeproject https://www.codeproject.com (pristupano 27.05.2017)

32. Plotting Circular Relationship Graphs with Silverlight dostupno na: https://www.codeproject.com/Articles/342715/Plotting-Circular- Relationship-Graphs-with-Silverl pristupano (pristupano 27.05.2017)

33. Computer Science & Engineering university of Washington, dostupno na: https://www.cs.washington.edu/ (pristupano 27.05.2017)

34. Les Misérables, dostupno na: http://www.goodreads.com/book/show/24280.Les_Mis_rables (pristupano 27.05.2017)

35. A Tour Through the Visualization Zoo, dostupno na: https://homes.cs.washington.edu/~jheer/files/zoo/(pristupano 27.05.2017)

36. Parallel Sets, dostupno na: http://www.datavizcatalogue.com/methods/parallel_sets.html (pristupano 27.05.2017)

37. Digital Splash Media, dostupno na: http://digitalsplashmedia.com/2014/06/visualizing-categorical-data-as-flows-with-alluvial-diagrams/ (pristupano 27.05.2017)

38. TeleGeography's Traffic Flow Maps http://mappa.mundi.net/maps/maps_014/ (pristupano 27.05.2017)

39. Flow map, dostupno na: http://www.datavizcatalogue.com/methods/flow_map.html pristupano (27.05.2017)

40. Drewconway, dostupno na: http://drewconway.com/zia/2013/3/29/the-value-of-edges-in-complex- network-visualization (pristupano 01.06.2017)

41. Erasmus mobility statistics 2012-13, dostupno na: https://data.europa.eu/euodp/en/data/dataset/erasmus-mobility-statistics-2012-13 (pristupano 02.06.2017)

42. Arcdiagram, dostupno na: https://github.com/gastonstat/arcdiagram (pristupano 05.06.2017)

43. Github, dostupno na: https://github.com (pristupano 06.06.2017)

Popis slika

Slika 1. Prikaz putanja nebeskih tijela 2
Slika 2. Ramon Llull – Arbor scientiae 3
Slika 3. Povezanost Latinoamerikanaca 7
Slika 4. Neusmjereni i usmjereni graf 8
Slika 5. Graf s petljom 9
Slika 6. Primjer multigrafa 9
Slika 7. Primjer usmjerenog težinskog grafa 10
Slika 8. Primjer stabla 11
Slika 9. Primjer trojke i zatvorene trojke 13
Slika 10. Primjer lokalnog koeficijenta 13
Slika 11. Primjer zajednica u grafu 14
Slika 12. Primjer silom usmjerenog grafa 19
Slika 13. Primjer hijerarhijskog grafa 20
Slika 14. Cirkularni prikaz 21
Slika 15. Prikaz pomoću lukova 22
Slika 16. Prikaz aluvijalnog dijagrama 22
Slika 17. Prikaz telekomunikacijske mreže 23
Slika 18. Prikaz kompleksne mreže 25
Slika 19. Vizualizacija mobilnosti profesora 27

Popis programskih kodova

Programski kod 1. Korišteni paketi 28
Programski kod 2. Čišćenje podataka 28
Programski kod 3. Spremanje podataka 29
Programski kod 4. Izrada prikaza 29

SAŽETAK

Ovaj rad se temelji na razradi bitnih područja pri vizualizaciji kompleksnih mreža. Od analize potreba za vizualizacijom, obrada jednostavnih i kompleksnih mreža do alata i metoda za vizualizaciju. Obrada mreža kreće od jednostavnih grafova do kompleksnih mreža i njihovih karakteristika. Osim teoretskog i tehničkog dijela, rad se fokusira na što treba pripaziti pri izradi vizualnog prikaza. Nakon obrade svih važnih cjelina, izrađuje se praktičan primjer koristeći R.

Ključne riječi: vizualna kompleksnost, vizualizacija mreža, kompleksne mreže, grafovi, RStudio, R

SUMMARY

This paper is based on an elaboration of essential areas in visualization of complex networks. From the analysis of the need for visualization, explaining simple and complex networks to describing tools and methods for visualization. Explaining network ranges from simple graphs to complex networks and their properties. Apart from the theoretical and technical part, the work focuses on what to look for when creating a visualisation. After processing all important units, a practical example is made using R.

Keywords: visual complexity, network visualization, complex networks, RStudio, R