

Korišćenje alata Gephi na predmetu Društvene mreže

Ana Miletić, Vesna Petković, Slobodanka Đenić, Verica Vasiljević, Mirko Stupar

Visoka škola elektrotehnike i računarstva strukovnih studija

Beograd, Srbija

anam@viser.edu.rs, vesnap@viser.edu.rs, sdjenic@viser.edu.rs, vvasiljevic@viser.edu.rs, mircan@viser.edu.rs

Sadržaj— Web 2.0 tehnologije i društvene mreže su omogućile svojim korisnicima da samostalno kreiraju i uređuju sadržaje. Na ovaj način dogodila se prava eksplozija različitih sadržaja, koji su postali otvoreno dostupni za korišćenje. Analiza društvenih mreža koristi teoriju mreža i predstavlja društvenu mrežu kao graf koji se sastoji od čvorova i veza. Kao alat za analizu društvenih mreža koristi se Gephi, koji omogućava vizuelizaciju i analizu velikih grafova. Upoznavanje studenata sa različitim tehnikama za analizu društvenih mreža uči ih primeni različitih tipova društvenih medija i tome kako se društveni mreže mogu koristiti kao alat za interakciju.

Ključne riječi— društvene mreže; analiza društvenih mreža; Gephi

I. UVOD

Nove Web tehnologije, kao što su Web 2.0, Web 3.0, društvene mreže i mashupi, kao i standardi koji ih podržavaju, bi trebalo da omoguće realizovanje ideje Weba, kao univerzalnog medijuma za razmenu podataka, informacija i znanja. Web 2.0 predstavlja verziju Weba čija je osnovna karakteristika aktivno učešće korisnika u kreiranju sadržaja na Web sajtu. Ovo učešće ogleda se u kreiranju i uređivanju sadržaja, ocenjivanju, komentaranju i dodavanju tagova. Na taj način svi korisnici raspolazu novim informacijama, za koje kreatori sadržaja smatraju da su od značaja.

Društveni softver (eng. social software) omogućava kreiranje i održavanje postojećih društvenih veza između ljudi, korišćenjem računarskih sistema. Istraživanja u ovoj oblasti obuhvataju različite discipline uključujući matematičke alate i tehnike, teoriju igara i računarstvo. Društveni mediji (eng. social media) predstavljaju ljudske interakcije korišćenjem virtuelnih zajednica i mreža, kreiranje, diskusiju i modifikaciju korisnički generisanog sadržaja.

Činjenica da je korisnicima omogućeno da sami dodaju i menjaju sadržaje dovela je do prave eksplozije sadržaja. Web 2.0 tehnologije su još uvek daleko od vizije kreatora Web-a Tim Berners Lee-a, čija je ideja da Web predstavlja mrežu povezanih podataka [1]. Trenutno stanje je takvo da se podaci nalaze zarobljeni u okviru Web 2.0 sajtova, pa se oni nazivaju i ograđeni vrtovi (eng. walled gardens). Međutim, pojavom API-a (eng. application programming interface) i uvođenjem semantičkih opisa Web se konstantno otvara za javnost.

Kako bi išli u korak sa vremenom i naše studente obučili za korišćenje novih tehnologija, počevši od ove školske godine uvedena je nastava iz predmeta Društvene mreže. Jedna od

tema koja se obrađuje je i analiza društvenih mreža, gde će studenti naučiti o značaju analize društvenih mreža i kako da izvode analizu društvene mreže korišćenjem alata Gephi. Pojavom društvenih medija analiza društvenih mreža je kao nauka izašla iz "zapečka", upravo zato što se korišćenjem ovih sajtova svakodnevno na Webu pojavljuju velike količine podataka, i što su ti podaci javno dostupni preko API-a. Zbog toga se korišćenjem tehnika analize društvenih mreža dolazi do interesantnih novih informacija, koje mogu biti od značaja za različite oblasti, kao što su ekonomija, marketing, politika, sociologija itd.

II. DRUŠTVENE MREŽE

Sam termin Web 2.0 prvi je upotrebio Tim O'Reilly, osnivač i direktor O'Reilly Media Inc. Nakon O'Reilly Media Web 2.0 konferencije 2004. godine naglo se proširilo njegovo korišćenje. Njegovo značenje može se objasniti analizom razlika između tehnologija koje se koriste u Web 2.0 i Web 1.0 [2].

Web 2.0 poseduje novu paradigmu u kojoj je fokus na isporuci servisa. Servisi Web 2.0 mogu se koristiti i kombinovati na nove načine, čime je korisnicima omogućeno da odaberu sadržaje koji su im od značaja, ili koji za njih imaju najveću vrednost. Treba pomenuti i čitaj/piši (eng. read/write) interfejs Web sajtova, čija se celokupna vrednost stvara kroz doprinos velikog broja korisnika. S druge strane, Web 1.0 omogućava samo čitanje postojećeg sadržaja [3].

Društvene mreže podrazumevaju formiranje virtuelnih zajednica, u kojima korisnici mogu da razmenjuju mišljenja, diskutuju, sarađuju i raspravljaju na teme vezane za njihova zajednička interesovanja [3].

A. Pojam i karakteristike društvenih mreža

Društvene mreže se mogu definisati kao servisi na Web-u, koji omogućavaju pojedincima da:

- kreiraju javni ili polujavni profil u okviru ograđenog sistema,
- kreiraju listu ostalih korisnika sa kojima su povezani,
- vide i istražuju sopstvenu listu veza sa ostalim korisnicima i veza koje su ostali kreirali u okviru sistema.

Priroda i nazivi pomenutih veza se razlikuju među različitim sistemima [4].

Osnovne karakteristike društvenih mreža su:

- formiranje on-line zajednice korisnika koji dele zajednička interesovanja,
- kolektivna inteligencija,
- korisnički generisan sadržaj.

Otvorenost je još jedna karakteristika svojstvena za Web 2.0, i odnosi se na distribuciju sadržaja i na jedinstvene formate podataka. Podržana je s Open Data API, a jedan od načina distribucije sadržaja je i putem RSS/Atom formata.

B. Kolektivna inteligencija

Kolektivna inteligencija, fenomen koji se naziva i „mudrost gomile“ predstavlja, takođe jednu od osnovnih karakteristika društvenih mreža. Ovaj termin podrazumeva učešće velikog broja ljudi, obično različitog porekla, navika i načina razmišljanja, u procesu kreiranja digitalnog sadržaja. Na ovaj način uglavnom se dolazi do boljih odgovora na pitanja i rezultata pretraga, nasuprot procesu kada se mišljenja pojedinaca razmatraju odvojeno [5].

Primer kolektivne inteligencije je i kreiranje folksonomija, kolaborativnog klasifikovanja sadržaja. Ovo klasifikovanje postiže se kreiranjem i upravljanjem tagovima.

Kreiranje tagova ili „tagovanje“, predstavlja vrlo popularan način klasifikovanja sadržaja na Web 2.0 sajtovima. Tag je ključna reč čija je uloga označavanje predmeta ili kategorije. Korisniku je dozvoljeno da pridruži bilo koju ključnu reč kakobi identifikovao sadržaj, kao što je link u slučaju social bookmarking-a ili pesma, video ili fotografije u slučaju sajtova za razmenu digitalnog sadržaja. Bitno obeležje je da se tagovi mogu razmenjivati, koristiti u pretragama ili preporučiti na osnovu izbora drugih korisnika istog sadržaja.

C. Korisnički generisan sadržaj

Na sajtovima društvenih mreža korisnici su odgovorni za dodavanje i ažuriranje informacija koje su dostupne svima. Ove virtuelne zajednice omogućavaju njenim korisnicima da razmenjuju mišljenja, diskutuju, sarađuju i raspravljaju na teme vezane za njihova zajednička interesovanja. Korisnički generisan sadržaj i razmena poslatih informacija glavna su pokretačka snaga kod formiranja i održavanja on-line zajednica [3].

Korisnički generisan sadržaj jeste sadržaj veb sajta koji kreiraju i ažuriraju korisnici, a ne konvencionalni mediji. Tipičan primer bio bi Flickr, on-line zajednica za razmenu fotografija, na kojoj se može naći preko 37 miliona slika, i gde 1,2 miliona korisnika postavlja na sajt i do 200 000 slika dnevno. Korisnički generisan sadržaj promenio je način korišćenja Interneta, pa se umesto jednosmernog prenosa medijskog sadržaja od strane određenog broja kompanija sadržaji kreiraju od strane korisnika i razmenjuju između njih [6].

Wiki predstavljaju primer web sajtova čiji je celokupan sadržaj korisnički generisan. Članci u wiki-u kreiraju se učešćem većeg broja ljudi, korišćenjem jednostavnog markup jezika u okviru web browsera. Korisnički generisan sadržaj i otvoren pristup uređivanju predstavljaju najznačajniju prednost

wiki-ja, ali i osnovni nedostatak. Ostali nedostaci su nemogućnost provere validnosti sadržaja i nedostatak kontrole. Bilo je slučajeva, kao u incidentu Seigenthaler [7], kada se skepticizam vezan za tačnost sadržaja na Wikipediji pokazao opravdanim. Bez obzira na sve nedostatke, ovaj otvoreni model u praksi dobro funkcioniše. Mnogi analitičari, a i korisnici, Wikipediju smatraju ozbiljnim konkurentom enciklopediji Britannica [8].

Primer zajedničkog procesa kreiranja sadržaja je i <http://delicious.com>, sajt za on-line skladištenje, dodavanje i tagovanje bookmarka. Nedostatak bookmark-a koji se nalaze na korisnikovom računaru je da se ne mogu koristiti na drugoj lokaciji ili razmenjivati s drugim korisnicima. Sajtovi za on-line bookmarking, kao što je delicious, omogućavaju skladištenje i organizovanje bookmarka na centralnom mestu na Web-u, tako da korisnici mogu da kreiraju on-line društvene mreže koje ih povezuju s prijateljima, porodicom i kolegama, ili s potpunim strancima sa kojima dele slična interesovanja.

Upotrebom međusobnih veza u ovoj (ali i bilo kojoj drugoj) društvenoj mreži korisnici mogu da praćenjem ponašanja ostalih korisnika dolaze do novih informacija, lakše i efikasnije nego npr. korišćenjem e-maila.

Kroz analizu Web sajta delicious.com vidi se da su korisnici društvenih mreža u centru pažnje. Oni su glavni učesnici u procesu kreiranja i razmene sadržaja. Činjenica da su prijava i korišćenje ovih sajtova besplatni i jednostavni je ono što privlači veliki broj korisnika na sajtove društvenih mreža.

III. UVOD U ANALIZU DRUŠTVENIH MREŽA

Analiza društvenih mreža se ukratko može opisati kao izučavanje ljudskih veza korišćenjem teorije grafova [9]. Ona predstavlja društvene odnose korišćenjem teorije mreža, koja se sastoji od čvorova (koji predstavljaju pojedinačne aktore u okviru mreže) i veze (koji predstavljaju odnose između pojedinaca).

Društvena mreža može se opisati skupom iskaza koji opisuju relacije između entiteta mreže. Jednostavan iskaz odnosa dva entiteta predstavlja osnovnu jedinicu analize društvene mreže i označava se kao dijada ili par. Društvene mreže mogu se opisati grafom $G = (N, E)$, gde je N oznaka čvorova mreže (eng. nodes), dok E označava granu mreže (eng. edge). Svaka dijada označava pojedinačni odnos koji se u tradicionalnoj teoriji grafova naziva grana ili veza. Subjekti u iskazu predstavljaju članove uključene u relaciju koji se nazivaju temena ili čvorovi u matematičkoj literaturi [10].

Čvorovi individua su povezani jednom ili više zavisnosti uključujući:

- zajedničke vrednosti, vizije ili ideje,
- društvene kontakte,
- konflikte,
- finansijske razmene.

Veze između čvorova obezbeđuju različite tokove, kao npr.:

- poruke,

- novac,
- bolesti.

Podela na tipove čvorova u analizi društvenih mreža vrši se na osnovu toga kakva je reprezentacija čvora unutar mreže. Grafovi koji sadrže čvor jednog tipa označavaju se kao jednomodalni, dok se grafovi sa relacijama dva različita tipa čvorova nazivaju bimodalni. Grafovi sa relacijama više različitih tipova nazivaju se multimodalni [10].

Kao osnovni elementi svake mreže, čvorovi i grane se opisuju dodatnim parametrima koji se dobijaju primenom odgovarajućih algoritama za analizu. Kada govorimo o granama u mreži između čvorova, one mogu biti usmerene (eng. directed) i neusmerene (eng. undirected). Usmerene veze se još nazivaju lukovi ili linkovi i označavaju se kao: $A \rightarrow B$. Neusmerene veze se označavaju sa: $A \leftrightarrow B$ ili $A - B$. Atributi koji bliže opisuju grane su: težina, rangiranje, tip, kao i svojstva u zavisnosti od ostatka strukture mreže [11].

Kod analize grana čvorova jedna od osnovnih karakteristika koja se preračunava na osnovu povezanosti svakog čvora u mreži je metrika čvora. Povezanost čvora u mreži određuje njegov stepen (eng. degree) koji predstavlja broj veza ka datom čvoru. Što je povezanost čvora veća, raste i vrednost stepena čvora.

Kod čvorova u usmerenoj mreži razlikuju se:

- Ulazni stepen (eng. indegree) predstavlja vrednost usmerenih veza koje idu ka čvoru.
- Izlazni stepen (eng. outdegree) pokazuje koliko veza potiče od čvora koji se posmatra.
- Kod neusmerenih mreža stepen je jednak ukupnom broju grana koje se stiču u čvoru.

Jedno od bitnih svojstava unutrašnje strukture grafa društvene mreže je centralnost (eng. centrality).

Centralnost se deli na dva osnovna tipa:

- centralnost međupoloženosti (eng. betweenness centrality) i
- centralnost blizine (eng. closeness centrality) [11].

Centralnost međupoloženosti je jednaka broju najkraćih putanja koje prolaze kroz posmatrani čvor ka ostalim čvorovima u mreži.

Centralnost blizine je prosečan broj koraka potrebnih od vrha jednog grafa do svih ostalih čvorova u dometu grafa mreže [12].

IV. MOGUĆNOSTI ANALIZE DRUŠTVENIH MREŽA KORIŠĆENJEM GEPHI ALATA

Gephi je softverski alat otvorenog kôda koji se koristi za vizuelizaciju i analizu velikih grafova. Gephi koristi 3D alate za renderovanje da prikaže grafove u realnom vremenu i ubrza istraživanje. Primena ovog alata obuhvata istraživanje, analizu, prostorni prikaz, filtriranje, klasterovanje, manipulaciju i izvoz različitih vrsta grafova. Podržani formati u Gephi okruženju su:

GEXF, GML, GraphML, Tulip, Pajek NET, GDF, CSV i Compressed ZIP [11].

Kada je otvoren radni fajl, izveštaj sumira podatke i svojstva i to na sledeći način:

- broj čvorova,
- broj grana,
- tip grafa.

Nakon uvoza podataka, pozicija čvorova je prozvoljna, tako da se može uočiti nešto drugačija reprezentacija podataka.

Modul za rangiranje (eng. ranking module) omogućava podešavanje boje i veličine čvorova. Pomoću ovog modula bira se parametar "Degree" za rangiranje.

Izračunavanje srednje dužine putanje (eng. average path length) u mreži za sve moguće parove čvorova daje informaciju o tome koliko su čvorovi blizu jedan drugome.

Metrike koje su dostupne:

- dijametar,
- centralnost međupoloženosti,
- centralnost blizine,
- srednja dužina putanje,
- koeficijent klastera,
- ekscentricitet,
- rang strane i
- detekcija zajednice (modularnost) [11].

Značajan korak u analizi podataka društvene mreže je filtriranje, gde se aktiviraju filteri koji mogu da sakriju čvorove i veze unutar mreže koja se analizira. Može se kreirati filter koji uklanja tzv. "listove", tj čvorove sa jednom vezom. Filteri predstavljaju specijalizovane upite za izdvajanje podataka od interesa u posmatranoj mreži.

V. MOGUĆNOST ANALIZE DRUŠTVENE MREŽE POMOĆU NETVIZZ I GEPHI ALATA

Korisnički podaci dostupni na Facebook mreži sadrže implementirane kontrole privatnosti, što za rezultat ima ograničenje dostupnosti većeg dela podataka za korisnike van facebook mreže. Dva tipa podataka do kojih se može lako doći su privatna "ego-mreža" i javna stranica podataka. Uz podešena ograničenja privatnosti podataka, većina korisnika može pratiti informacije svojih prijatelja ili prijatelja prijatelja (eng. friend of a friend, foaf). Ovako se kreira ego mreža sa radiusom od 2 skoka [12]. U cilju upoznavanja sa osnovnim parametrima za analizu Facebook društvene mreže, planirano je da se za dobijanje podataka koristi već postojeći dodatak za Facebook Netvizz koji se može instalirati preko sledeće adrese:

<http://apps.facebook.com/netvizz/>

Kao dodatak za izdvajanje podataka korišćena je trenutno aktuelna verzija Netvizz v1.0. Ovaj plugin omogućava korisniku da izdvoji i izveze podatke ego mreže sa svog

Facebook profila u .gdf format pogodan za analizu podataka. Tekuća verzija Netvizz v1.0 dodatka omogućava izdvajanje podataka ličnog profila, grupa, kao i strana u istraživačke svrhe.

Moduli koji su na raspolaganju primenom ovog plugin-a su:

- lična mreža – izdvaja prijatelje i međusobne relacije
- mreža ličnih like-ova – kreira mrežu koja kombinuje prijatelje i njihove lajkove
- podaci grupe – kreira mrežu i tabularne fajlove i za prijateljstva i za interakcije u grupama
- mreža like-ovanih strana – kreira mrežu strana čije su veze međusobni lajkovi
- podaci strane – kreira mrežu i tabularne fajlove za strane.

Tipovi čvorova koji se mogu izdvojiti u analizi Facebook ontologije su: korisnik i like.

Od kategorija čvorova definisani su za korisnika:

- pol korisnika (user_sex),
- jezik korisničkog interfejsa (user_locale) i
- broj objekata koje je korisnik like-ovao (user_likecount).

Definisane kategorije za like-ove:

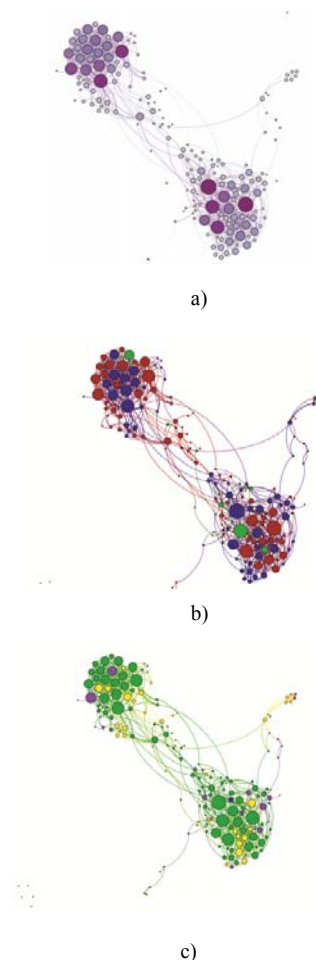
- broj like-ova i stavka koju je korisnik like-ovao (like_locallikes) i
- broj like-ova i stavka koja je like-ovana od strane svih korisnika na Facebooku (like_fulllikes).

Kako bi se studenti upoznali sa osnovnim konceptima analize društvenih mreža, planirano je da se na praktičnoj nastavi uradi analiza podataka ličnog Facebook profila. U primeru analizirane mreže korišćen je Force Atlas 2 layout algoritam u Gephi okruženju. Osnovne karakteristike mreže ličnog Facebook profila autora koja je analizirana prikazana na Sl. 1:

- broj čvorova za analizu: 185
- broj grana: 826
- tip grafa: undirected.

Ograničenje prilikom izdvajanja ovih podataka jeste da dostupnost podataka zavisi od postavljenih pravila privatnosti svakog korisnika na Facebooku. Na prvom grafu predstavljena je vizuelizacija lične Facebook mreže autora prema stepenu čvora. Vizuelizacijom podataka generisana je struktura sastavljena od čvorova koji predstavljaju kontakte na Facebook mreži autora. Uključeni su i parametri za izdvajanje čvorova po stepenu, gde je veličina čvora proporcionalna stepenu čvora. Čvorovi sa najvećim stepenom su najznačajniji i predstavljaju osobe koje su unutar grupe najpovezanije sa ostalim članovima grupe. Prosečan stepen čvora u primeru posmatrane mreže iznosi 4.465. U mreži se mogu vizuelno izdvojiti i komponente

međusobno povezanih čvorova. U ovom primeru se komponenta koja sadrži najviše čvorova, tzv. gigantska komponenta sastoji od dve grupacije međusobno povezane zajedničkim čvorovima. Primenom filtera Giant Component može se odrediti veličina gigantske komponente koja u posmatranom primeru iznosi 176 čvorova (95,14% ukupne veličine posmatrane mreže).



Slika 1. Prikaz Facebook mreže konekcija u Gephi okruženju - a) Force Atlas 2 layout prikaz čvorova sa granama konekcije b) Izdvajanje čvorova prema polu c) Izdvajanje čvorova prema jeziku korisničkog interfejsa.

Detaljnija vizuelizacija mreže iz Gephi okruženja može se izvršiti i primenom filtera za dodatno dostupne podatke mreže. Na osnovu podataka dobijenih sa autorovog profila, javno dostupni parametri dobijeni pomoću netvizz plugin-a omogućavaju pregled polne strukture mreže, kao i korišćenje jezika korisničkog interfejsa. Na osnovu postojećih kategorija čvorova koji predstavljaju korisnike izvršeno je partitionisanje po kategoriji pola korisnika. Za vizualizaciju korišćen je filter Partition. Plavom bojom označene su osobe muškog a crvenom osobe ženskog pola. Zelenom bojom označeni su korisnici koji u svojim profilima nisu popunili ovu stavku ili sama stavka zbog podešavanja privatnosti nije bila dostupna. Kod jezičkog interfejsa na posmatranom uzorku, najviše korisnika (70%) ima engleski korisnički interfejs (označen zelenom bojom).

VI. ZAKLJUČAK

Pojava sajтова društvenih mreža omogućilo je da Web postane najveći javni izvor podataka na svetu. Pristup i analiza ovim javnim podacima o ljudima i njihovim međusobnim vezama otvara nove primene tradicionalnih tehnika analize društvenih mreža. One omogućavaju identifikovanje najvažnijih ili najuticajnijih ljudi u mreži, kako se stvari šire kroz mrežu i prirodu ljudskih veza. Izučavanjem analize društvenih mreža i alata Gephi studente upoznaje sa tim tehnikama, pokazuje primenu različitih tipova društvenih medija i uči ih tome kako se društveni mediji mogu koristiti kao alat za interakciju sa online javnošću.

LITERATURA

- [1] T. Berners-Lee, J. Hendle and O. Lassila, "The Semantic Web," *Scientific American*, 2001.
- [2] T. O'Reilly, "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," 30 September 2005. [Online]. Available: <http://oreilly.com/web2/archive/what-is-web-20.html>. [Accessed 09 February 2014].
- [3] J. Bernal, Web 2.0 and Social Networking for the Enterprise: guidelines and examples for implementation and management within your organization, 2009, IBM Press.
- [4] D. Boyd and N. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, 2007.
- [5] J. Surowiecki, The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations, New York: Anchor Books, 2004..
- [6] K. Seongwoon, L. Inseong, L. Kih and J. Seungk, "Mobile Web 2.0 With Multi-Display Buttons," *Communications of the ACM*, 2010.
- [7] J. Seigentaler, "A false Wikipedia 'biography'," 29 November 2005. [Online]. Available:

http://www.usatoday.com/news/opinion/editorials/2005-11-29-wikipedia-edit_x.htm. [Accessed 09 February 2014].

- [8] "Wikipedia survives research test," 15 December 2005. [Online]. Available: <http://news.bbc.co.uk/2/hi/technology/4530930.stm>. [Accessed 09 February 2014].
- [9] M. Tsvetovat and S. N. A. Kouznetsov, Social Network Analysis for Startups, O'Reilly Media, 2011.
- [10] M. Čangalović, V. Manojlović and V. Baltić, Diskretne matematičke strukture, Beograd: FON, 2009.
- [11] "Gephi," [Online]. Available: <https://gephi.org/>.
- [12] "Social Network Analysis, Coursera," [Online]. Available: <https://www.coursera.org/course/sna>.

ABSTRACT

Web 2.0 technologies and social networks facilitate users to individually design and edit contents. Hence, real outbreak of diverse contents emerged, and its use became overtly accessible. Social network analysis uses Network theory and introduces social network as a graph consisting of nodes and ties. Social network analysis uses Gephi as a tool allowing visualization and analysis of large graphs. Presentation of various techniques for analyzing social networks to students introduces the application of distinctive types of social media and how social media can be implemented as interaction tool.

Application of Gephi tool for social network analysis in the Course Social networks

Ana Miletic, Vesna Petkovic, Slobodanka Djenic, Verica Vasiljevic, Mirko Stupar