

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

# Condor

*Ivan Kvastek*

voditelj: *Mr. sc. Mile Šikić*

*Zagreb, 7. lipnja 2008.*

## Sadržaj

1. Uvod .....	2
2. Kompleksne mreže .....	2
2.1. Grupiranje.....	2
2.2. stupnjeva.....	3
2.3. Primjeri kompleksnih mreža .....	4
2.3.1. Mreže bez skale .....	4
2.3.2. Mali svjetovi.....	4
3. Condor .....	5
3.1. Uvod.....	5
3.2. Princip rada condora i njegove mogućnosti .....	6
3.2.1. E-mail database .....	6
3.2.2. Document database (web site database) .....	6
4. Zaključak .....	12
5. Sažetak .....	13
6. Literatura.....	14

## 1. Uvod

Kompleksne mreže su dugo bile predmet istraživanja matematike i fizike, te su danas zastupljene u mnogim znanostima i njihovim granama. Dakako, sam naziv kompleksne mreže odaje da je njihova složenost velika. U teoriji postoje dvije vrste slučajnih mreža: klasične sa Poissonovom distribucijom i kompleksne mreže sa . Kod klasičnih mreža su veze distribuirane nasumično između konačnog broja čvorova, a kompleksne mreže su prilično zamršenije od klasičnih mreža. Jedan od alata koji pojednostavljaju shvaćanje velikih kompleksnih mreža je TeCFlow. TeCFlow je alat kojim se koristi Condor koji vizualizira kompleksnu mrežu te uvelike olakšava njeno shvaćanje.

U ovome seminaru ću opisati aplikaciju Condor koja se koristi svojstvima kompleksnih mreža kao što su međupoloženost, grupiranje i centralitet stupnjeva, te ću opisati osnovne pojmove kompleksnih mreža.

## 2. Kompleksne mreže

Kompleksne mreže su mreže u kojima je organizacija veza kompleksna, npr. distribucija stupnjeva koja može biti kompleksnija od Poissonove distribucije zastupljene u klasičnim mrežama. Naravno tu su i druge karakteristike mreža i razne pojave, kao što je efekt „debelog repa“. Zbog velike složenosti stvarnih mreža klasične slučajne mreže se ne mogu uspješno koristiti kao modeli za objašnjavanje empirijski dobivenih podataka iz stvarnog svijeta, stoga i danas najrazvijeniji modeli rasta mreža, kao što je Barabási-Albert model, imaju ograničene mogućnosti predviđanja karakteristika stvarnih mreža. Neka od najvažnijih obilježja su grupiranje (*engl. Clustering*), te distribucija stupnjeva koja se može dogoditi u nekoliko na nekoliko načina.

### 2.1. Grupiranje

Sama riječ grupiranje govori da je riječ o nakupljanju više čvorova oko samo jednog, tj. jedan čvor ima sve više veza sa drugima. Tada nastaju petlje koje su specifične veze u mrežama.

Koeficijent grupiranja ( $C$ ) određuje gustoću veza u okolini bliskoj čvoru. Taj koeficijent je omjer svih  $x$  veza koje spajaju najbližih  $y$  susjeda čvora i svih mogućih veza koje bi ih mogle spajati:

$$C = \frac{2x}{z(z-1)}$$

Iako bi se mogla uvesti distribucija od  $C$  najčešće se koristi prosječna vrijednost grupiranja  $\bar{C}$ . koeficijent grupiranja je direktno vezan za prisutnost „trokuta“ u mreži, te je vrlo dobar pokazatelj toga postoje li korelacije između čvorova mreže.

No kod klasičnih slučajnih mreža vrijedi omjer  $\bar{C} = \frac{\bar{k}}{N}$  gdje je  $N$  broj čvorova mreže, a  $\bar{k}$  prosječni stupanj čvorova. Kako je  $\bar{k}$  fiksna vrijednost on se približava nuli kako veličina mreže raste. Empirijska istraživanja pokazuju da to nije slučaj s mnogim stvarnim mrežama.

## 2.2. Distribucija stupnjeva

Stupnjevi čvorova u slučajnim mrežama su statički raspodijeljeni, gdje je stupanj čvora broj veza koje čvor ima sa susjednim čvorovima. U neusmjerenoj mreži, ako se čvorovi međusobno razlikuju, svaki čvor ima stupanj distribucije  $p(k,s,N)$ , ili kažemo da je to vjerojatnost da čvor  $s$  u mreži veličine  $N$  čvorova ima  $k$  veza (susjeda). Poznavajući distribuciju svakog stupnja možemo naći ukupnu distribuciju stupnjeva mreže:

$$P(k, N) = \frac{1}{N} \sum_{s=1}^N p(k, s, N)$$

Ako su čvorovi u slučajnim mrežama statički jednaki, tada svaki od njih ima jednaku distribuciju stupnjeva  $P(k,N)$ . Prvi moment te raspodjele daje nam srednji stupanj mreže:

$$\bar{k} = \sum_k kP(k)$$

Neko od tipičnih distribucija stupnjeva su:

- Poissonova distribucija

$$P(k) = \frac{e^{-\bar{k}} \bar{k}^k}{k!}$$

Klasični slučajni grafovi imaju ovakvu distribuciju ako im se broj čvorova približava beskonačnosti uz ograničenje da je srednji stupanj grafa konstantan.

- Eksponencijalna distribucija

$$P(k) = e^{-\frac{k}{\bar{k}}}$$

- Zakon potencije (*polinomna distribucija*)

$$P(k) = k^{-\alpha}$$

Za razliku od prve dvije distribucije ona nema prirodni opseg te ju se još naziva distribucija bez skale (*engl. scale-free*), a mreže s takvom distribucijom su mreže bez skale.

### 2.3. Primjeri kompleksnih mreža

Najvažniji primjeri kompleksnih mreža su: mreže bez skale (*scale-free*) i mali svjetovi (*small-world*)

#### 2.3.1. Mreže bez skale

Mreže bez skale su važne zbog svoje polinomne distribucije stupnjeva, što je svojstvo koje je zamijećeno u brojnim realnim mrežama od Interneta i WorldWideWeba do proteinskih interakcija u metabolizmu. Karakteristično za mreže bez skale je da se u njima pojavljuju čvorovi (takozvani hubovi) koji imaju stupnjeve za red veličine veće od prosječnog stupnja svih čvorova. Pri tome karakteristike mreža bez skale ne divergiraju bitno bez obzira na veličinu mreže to jest broj čvorova. Najvažnija karakteristika takvih mreža je njihova polinomna distribucija stupnjeva

$$P(k) \sim k^{-\alpha}$$

$\alpha$  se naziva eksponent distribucije.

#### 2.3.2. Mali svjetovi

Fenomen "malog svijeta" poznat je već neko vrijeme iz stvarnog svijeta. Najpoznatiji primjer je slavni pokus Stanleya Milgrama iz 1960-tih i njegova demonstracija da su svaka dva čovjeka na svijetu povezana u prosjeku preko najviše šest drugih ljudi. Formalno, efekt malog svijeta zahtjeva preciznu definiciju prosječne najkraće udaljenosti između dva čvora mreže. Očito, prosječni broj  $n$  najbližih susjeda jednog čvora raste usporedivo s  $\langle k \rangle^n$ . Znači, prosječan najkraći put  $\bar{l}$  ugrubo se aproksimira relacijom  $\langle k \rangle^{\bar{l}} \sim N$ . Stoga vrijedi

$$\bar{l} = \frac{\ln N}{\ln \langle k \rangle}$$

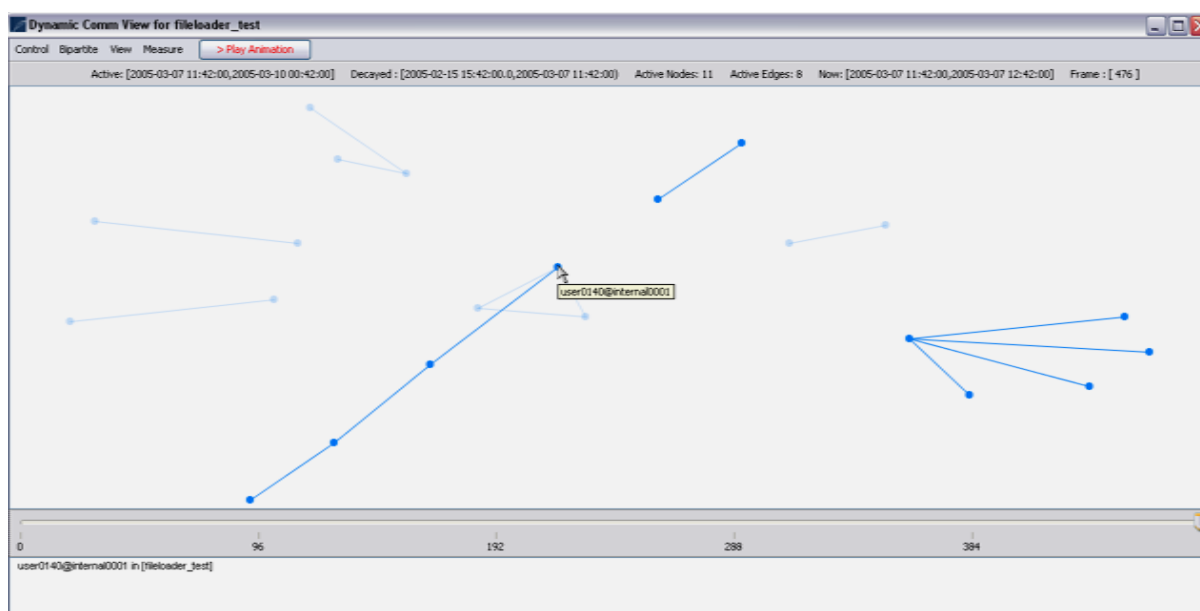
Uglavnom, ako  $\bar{l}(N)$  raste sporije nego ijedna pozitivna potencija od  $N$  za mrežu možemo reći da ima efekt malog svijeta.

### 3. Condor

Condor je aplikacija koja koristi TeCFlow, *Temporal Communication Flow Visualizer*. Taj alat je razvijen u znanstvenom projektu na sveučilištu MIT u Americi koji je vodio prof. Dr. sc Peter Gloor profesor na tom sveučilištu.

#### 3.1. Uvod

Jedan primjer koji poboljšava samo shvaćanje toga alata, i približava ga drugima, koji je naveden u Gloorovoj knjizi [3] je: „Recimo da ste ljubitelj klasične glazbe i da slušate jednu sonatu od Beethovena. Te što je duže slušate, to će te je dublje upoznati i znati će te što očekivati sljedeće.“, te takav pristup ima i TeCFlow. S tim se alatom može isto što i sa slušanjem muzike, samo to treba primijeniti na razvoj odnosa i povezanosti u socijalnoj mreži. Njegov cilj je bio da napravi vizualizaciju dinamičnosti ljudskih povezanosti u nekom određenom prostoru ili okolini. Tako na primjer Condor, koristeći se s tim alatom može predočiti postupni razvoj neke mreže prikupljajući podatke tijekom nekog unaprijed određenog vremena. Na osnovi tih podataka Condor napravi „film“ koji kada se pokrene prikazuje, za unaprijed određene točke vremena, razvoj mreža. Mjesta na kojima je to moguće su blogovi, tražilice, određene internet stranice ili preko elektoničke pošte. Naravno mogućnost Condora je da prikazuje cijeli sadržaj neke pošte ili nekog „posta“ na blogu. Pošto je to zadiranje u nečiju privatnost, a i zbog ograničenja koja sam imao, uzeo sam primjer samog profesora. On je organizirao grupu (tim) ljudi koji su dobrovoljno pristali biti dio testa, te je prikupljao njihove pošte. Na osnovi prometa pošte između članova toga tima on je mogao napraviti vizualni prikaz razvoja mreže (povezanosti) toga tima.



Slika 1. Prikaz komunikacije grupe ljudi putem e-maila

Na primjer svaka točka (čvor) predstavlja osobu, te linija između dvije točke je kao odnos između dvije osobe. Što se više te osobe zbližavaju, to je njihov odnos sve intenzivniji,

odnosno više poruka su si poslale. I tako s još dosta osoba i odnosa među njima dobijemo malu mrežu, tj. tim ljudi. Osobe koje su najaktivnije u tome timu, odnosno osobe koje dobiju i pošalju najviše poruka, su smještene u sredinu mreže.

Jedna od prednosti takvoga ispitivanja povezanosti i komunikacije je to što bi se mogli, pomoću takvog uspoređivanja dinamičkih prikaza i rada virtualnih timova, odrediti najbolji komunikacijski kanali i tako poboljšati rad virtualnog tima i donekle kontrolirati njegov razvoj.

### 3.2. Princip rada Condora i njegove mogućnosti

Kao glavni orijentir u snalaženju sam koristio „Condor Manual 2.2“, te već pripremljene baze podataka<sup>1</sup>.

Rad Condora se zasniva na prikupljanju podataka i njihovoj pohrani. Pohrana podataka se vrši preko MySQL servera na internetu. Pošto je želja bila da se Condor pojednostavni, nekakvo predznanje za programski jezik MySQL-a nije bilo potrebno jer je program napravljen tako da se jednostavnim pozivom funkcije „database→new“ stvara nova datoteka na serveru, te se upisi u nju ne rade ručno nego automatski preko programa. Interakcijski „film“ koji Condor stvara od tih podataka je dosta jednostavan i pun mogućnosti. Kao npr. vlastito bojanje radi bolje vizualizacije, mjerenje međupoloženosti (*engl. Betweenness*), prikazi aktivnosti određenih čvorova u određenom periodu vremena, i dr.

#### 3.2.1. E-mail database

U prvome primjeru sam koristio već gotove unose koji su bili spremljeni u tekst datoteku pod nazivom „fileloader\_sample.txt“ koju je g. Peter Gloor napravio kao primjer za vizualizaciju. Ta datoteka je bila sastavljena od djelomičnih e-mail poruka koje su si međusobno slali on i njegov tim. U njemu je bilo vidljivo krot šesnaest dana kako se poruke šalju i tko je koji dan bio aktivan.

*E-mail database* općenito predstavlja bazu podataka koja se stvara unosom cijelih „mailboxova“.

#### 3.2.2. Document database (web site database)

*Document database* je tip baze podataka koji služi za unos opširnijih podataka. Drugi primjer koji je g. Gloor priložio uz program je bila baza podataka svih unosa sa jednog foruma (Hipforum) koju je trebalo samo učitati u program. Ta baza podataka je prikupljala unose svih članova foruma tijekom jedne godine, te je čisti primjer neetičnosti kod takvih primjena jer je narušila privatnost forumaša (*slika 2*).

---

<sup>1</sup> Kao razlog tome je to što je Google® prestao izdavati klasične API ključeve sredinom 2007. godine, a prilagođavanje drugim vrstama ključeva i razvoj novije verzije programa je još u procesu.

datasetId	comm_id	comm_from	comm_to	comm_time	comm_subject	comm_content	tag
http://hipforums.com/foru...	1554	lodui	jono	2004-09-15 10:13:00.0	Spacer's post[thread 2399...	Well first off I'd like to say....	To
http://hipforums.com/foru...	1563	geckopelli	jono	2004-10-05 06:22:00.0	Spacer's post[thread 2399...	Smoked it many times. It u...	To
http://hipforums.com/foru...	1606	geckopelli	jono	2004-10-05 06:24:00.0	Foppful's post[thread 3545...	It's not real.If it was, you ...	To
http://hipforums.com/foru...	1605	Foppful	jono	2004-10-04 02:08:00.0	Foppful's post[thread 3545...	I recently aquired an 1/8 o...	Cc
http://hipforums.com/foru...	1604	jono	Foppful	2004-10-04 02:54:00.0	Foppful's post[thread 3545...	could be some incense. i d...	To
http://hipforums.com/foru...	1696	gottssunfire	jono	2004-09-13 06:20:00.0	gottssunfire's post[thread ...	I want to try opiumboth ea...	Cc
http://hipforums.com/foru...	1695	jono	gottssunfire	2004-09-15 02:32:00.0	gottssunfire's post[thread ...	growing opium poppy requi...	To
http://hipforums.com/foru...	1562	jono	therefore...	2004-10-04 02:56:00.0	Spacer's post[thread 2399...	there is no such thing as w...	To
http://hipforums.com/foru...	1552	jono	sofaraway	2004-09-15 02:35:00.0	Spacer's post[thread 2399...	i sincerely doubt that any ...	To
http://hipforums.com/foru...	1553	spacer	jono	2004-08-17 09:36:00.0	Spacer's post[thread 2399...	I'm gonna be heading off t...	Cc

Search

Search Words :

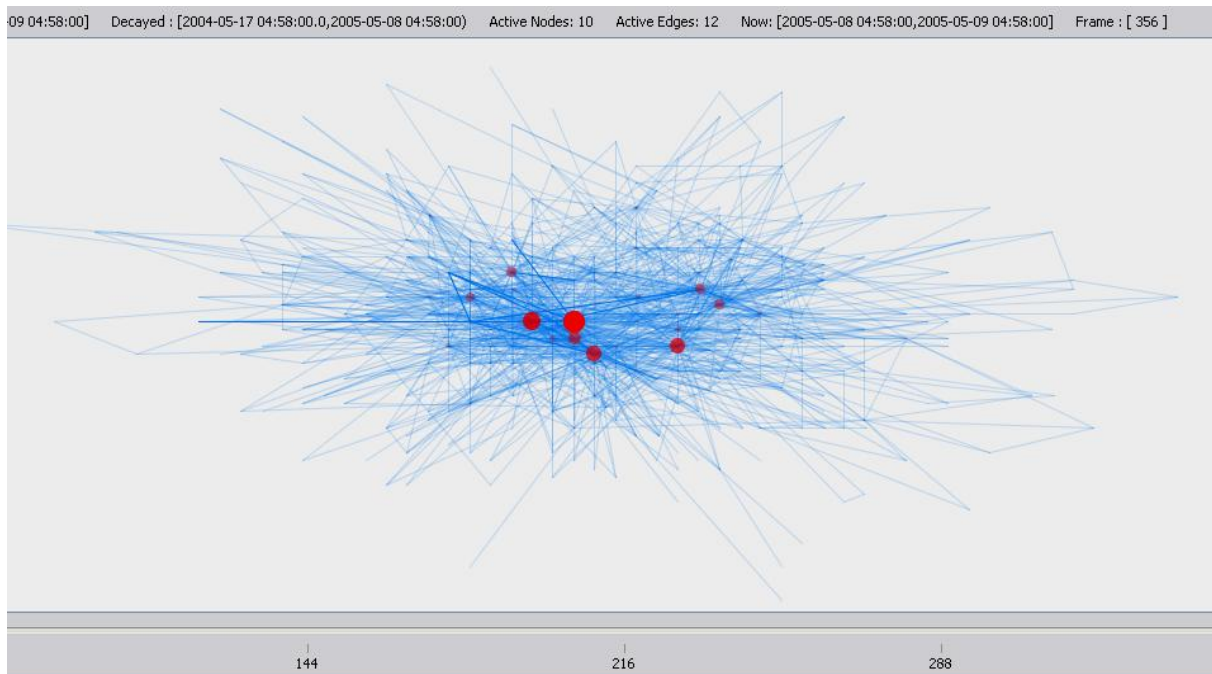
---

there is no such thing as white rock opium. sorry dude you may have seen that shit at bonnaroo or somewhere else, its white and it turns to black oil when you burn it right? its crap. i've grown poppies and know that opium is hard to come by and what i

Slika 2. Ovdje se jasno može vidjeti kako se svaka poruka može evidentirati i prikazati njezin sadržaj



No, usprkos tome, taj primjer je jako dobro poslužio za proučavanje postupnog razvoja veza (mreža) među članovima foruma, te je moglo jasno pokazati tko je najaktivniji član foruma.



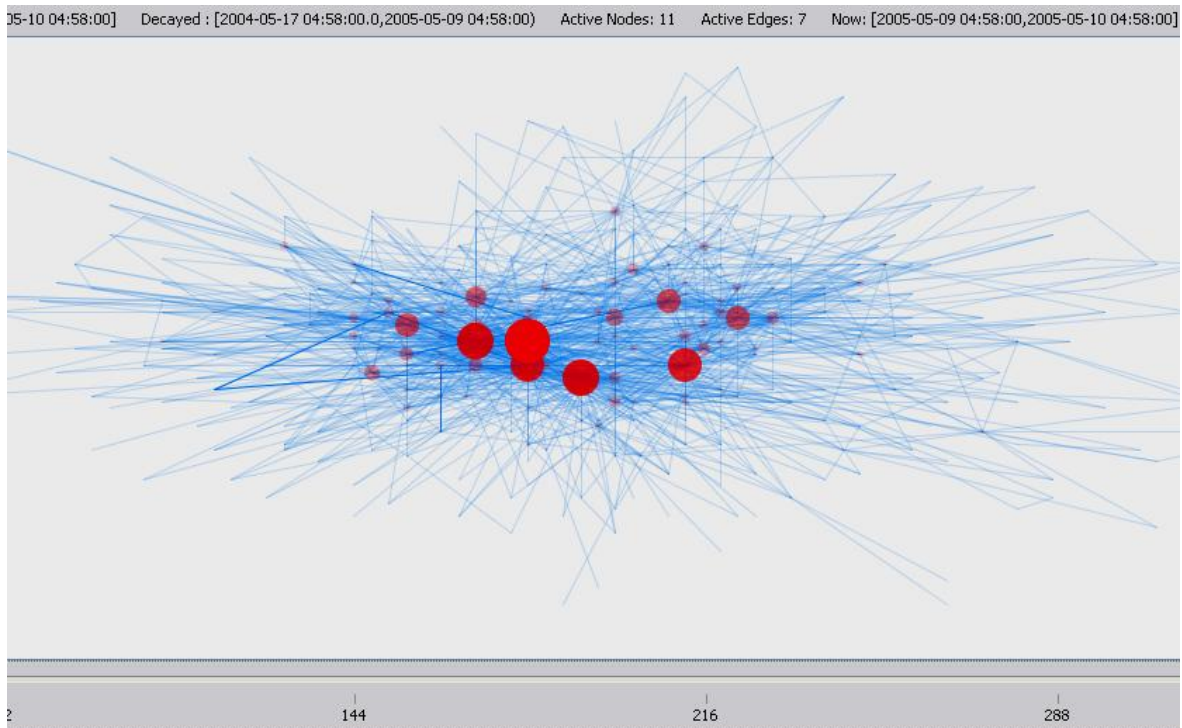
Slika 3. Prikaz međupoloženosti pomoću crvenih točki

Na *slici 3.* crvene točke prikazuju vrijednost centraliteta međupoloženosti među članovima. Što je mreža veća to je međupoloženost manja, ali što je aktivnost između članova veća to je međupoloženost veća, tako ove crvene točke svojom veličinom i intenzitetom jasno pokazuju međupoloženost. Centralitet međupoloženosti je svojstvo mreža koje se može definirati kao izračun centraliteta čvora u mreži. To znači da čvorovi koji se nalaze na mnogo kratkih „puteva“ do drugih čvorova imaju veću vrijednost nego on koji se ne nalaze. međupoloženost se može dobiti preko formule

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

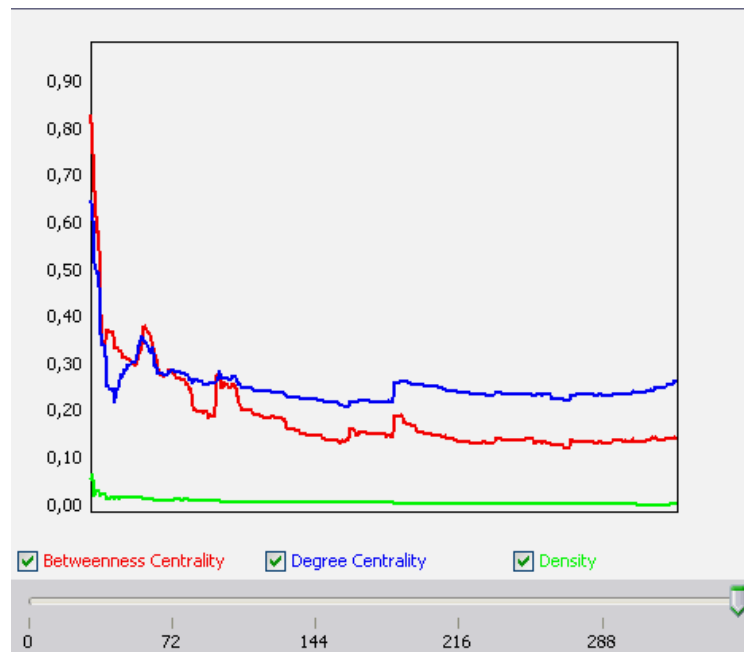
gdje je  $\sigma_{st}$  broj najkraćih puteva od  $s$  do  $t$ , a  $\sigma_{st}(v)$  je broj najkraćih puteva od  $s$  do  $t$  kroz neki čvor  $v$ .

Drugo svojstvo koje se može vidjeti je centralitet stupnjeva (*engl. degree centrality*).



Slika 4. Centralitet stupnjeva

Centralitet stupnjeva određuje broj veza po jednom čvoru. Pošto se radi o vezama sa susjednim čvorovima, usko je povezan sa međupoloženošću, no preko toga se ne može izračunati najkraći put između određenog čvora i susjeda.



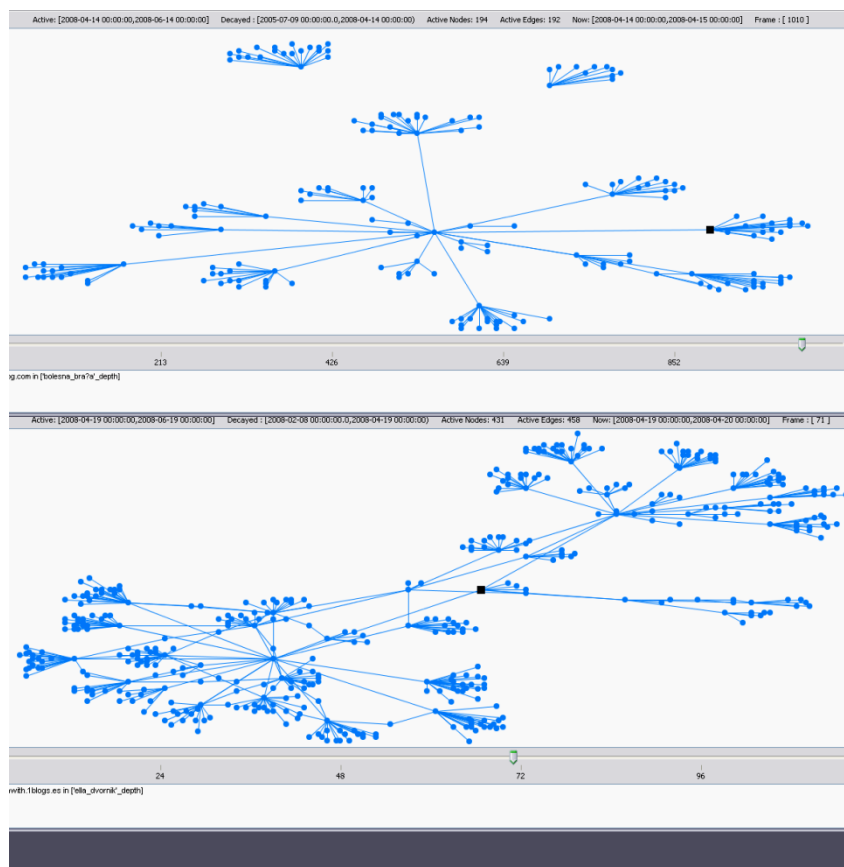
Slika 5. Graf centraliteta stupnjeva i međupoloženosti

Centralitet stupnjeva i međupoloženosti imaju zajedničko svojstvo, a to je da se njihov graf ponaša kao i graf eksponencijalne distribucije stupnjeva čvorova. No kod centraliteta stupnjeva postoje dvije vrste veza, *indegree* i *outdegree*. *Indegree* predstavlja dolazne veze, odnosno poruke, a *outdegree* odlazne. Pošto se centralnost odnosi na one najaktivnije i najpopularnije onda se uzima *indegree*. Znači, što je veći broj članova to je manja međupoloženost, ali i centralitet stupnjeva.

Kod Condora uz forume se može još pratiti i blogove, internet stranice i RSSFeed pomoću istoimenih kolektora.

Blog Collector pruža način na koji se mogu prikupiti podaci o nekim određenim stvarima za koje bi htjeli nešto predvidjeti ili vidjeti njihovu popularnost među blogerima ili na internetu općenito. To je moguće tako da se upiše željena riječ ili url neke internet stranice. Takav kolektor radi na principu da preko jedne od tražilica blogova od kojih su dostupne: *Google Blog Search*, *MSN Live WebSearch*, *Google News Search* ili *Google Scholar Search*, traži upisani parametar.

Svaka od tih tražilica pregledava različite stranice. Uzmimo na primjer Google blog Search i MSN Live WebSearch. Google blog Search je tražilica koja pretražuje blogove i njom se služi Condor da bi nam prikazao koliko je zapravo to što smo mi tražili popularno u zadnjih par dana, odnosno koliko ljudi je spomenulo barem jednom naš traženi parametar. Jedna od mogućnosti Blog Collectora je da se izaberu neke detaljnije opcije kao što su stupanj separacije, proizvoljan broj nađenih blogova tražilice, datum koji hoćemo ili interval od nekoliko dana. Stupanj separacije je stupanj koji određuje da li ćemo tražiti našu riječ ili internet stranicu u blogovima koji su nađeni od tražilice, ili hoćemo da nam se prikažu blogovi ili Internetske stranice koje imaju linkove na tim nađenim blogovima, a da se spominje naš pojam. Naravno što je viši broj stupnja separacije to kolektor „dublje gleda“. Primjera radi, sa estrade sam uzeo poznatu grupu „Bolesna Braća“ i „Ellu Dvornik“ te sam usporedio tko je od njih u zadnjih nekoliko mjeseci popularniji kod blogera. Bolsena Braća su se u opširnijem pretraživanju našla u jednom trenutku u 194 bloga, dok je Ella u skoro mjesec do dva bila glavna tema blogera, te se u jednom trenutku pojavila na čak 431 blogu (*slika 6*).



Slika 6. Gornji prikaz je za grupu „Bolesna Braća“, a donji za Ellu Dvornik

MSN Live WebSearch funkcioniра na istome principu, jedino što je razlika je u tome što se ne orijentira samo na blogove, pa su traženja opširnija, i više se orijentiraju na postojeće Internet stranice.

#### 4. Zaključak

Mreže su zastupljene u svakakvim prirodnim procesima i u društvima, u digitalnom svijetu i u prirodnim znanostima. Svaka od njih je kompleksna i funkcionira na slične načine.

Condor se zasniva na kompleksnim (stvarnim) mrežama i donekle može pojednostavniti njihovo shvaćanje u stvarnome svijetu. On je uspio vizualizirati i pojasniti ono što već dugo postoji. Mreže, kao što su Internet i WWW, je za sad nemoguće prikazati u cijeloj svojoj veličini, ali nije niti potrebno, jer kako smo vidjeli svi fenomeni i sve karakteristike mreža se mogu dokazati i vidjeti u puno manjim mrežama ili isječcima velikih mreža.

Condor, kako smo vidjeli, može započet lavinu pitanja o etičnosti i moralu u takvim postupcima, jer pruža veliku količinu privatnih podataka na jednome mjestu, te koliko god oni bili izloženi nisu u tolikoj opasnosti od zloupotrebe koliko su korištenjem Condora.

No, kao što sam rekao prije, jedna od prednosti Condora je stvaranje virtualnih timova (timova na daljinu) i njihovo stalno praćenje razvoja i interakcije, ili praćenje odnosa radnika u nekoj tvrtci koja ne mora biti nužno sastavljena od virtualnih timova. Takvo praćenje bi moglo osigurati da se ne dogode nekakvi propusti oko važnih informacija ili raspršenje odgovornosti, jer bi se komunikacija između svih članova tvrtke učvrstila i pratila, te bi se tako mogle spriječiti neke greške. Što se tiče privatnosti u tvrtci ona se ne bi narušavala, te takav način praćenja može biti posve siguran i profesionalan. Tu je i mogućnost predviđanja budućnosti, što se tiče „kolektivne svijesti“, kao npr. glasanja, ili na principu dijeljenja određenih podataka, tj. umreživanja određenih izvora informacija, što bi pomoglo pri prognoziranju vremena ili nekakvih opasnosti.

## 5. Sažetak

Kompleksne mreže su svugdje oko nas. Ima ih u prirodi, u biljkama i životinjama, pojavljuje u raznim oblicima i objašnjava se na razne načine, ali samo je jedan mali dio koji je čovjek napravio, ali to djelo je veliko i fenomenalno. Internet je jedna velika mreža u kojoj su ljudi spojeni sa cijelim svijetom samo sa klikom miša.

Ovim seminarom sam opisao osnovna svojstva kompleksnih mreža, način rada i mogućnosti programa „Condor“. Objasnio sam kako se njime koristiti i čemu služi. Na kojim principima radi i čime se služi za pohranu podataka, za obradu podataka, vizualizaciju i lakše shvaćanje rješenja.

Zapravo nije Condor približio mrežu nama, nego nas mreži, njenom shvaćanju i dočaravanju njenih fenomena i karakteristika.

## 6. Literatura

- [1] Đani Glavinić: Osnovna svojstva kompleksnih mreža i njihova promjena (seminar), 2007.
- [2] Matija Piškorec: Modeli rasta kompleksnih mreža(seminar), 2007.
- [3] Peter A. Gloor: Swarm creativity: Competitive Advantage through Collaborative Innovation Networks, 28. Studenoga 2007.
- [4] Peter Gloor: Condor Manual 2.2, 20. veljače 2007.
- [5] Internet, <http://tiger.soberit.hut.fi/virtualbrownbag/tiki-index.php>
- [6] Internet, <http://ccs.mit.edu/pgloor.html>
- [7] Internet, <http://en.wikipedia.org/wiki/Centrality>