VELIKA OMREŽJA IZ REALNEGA SVETA

Lovro Šubelj, Slavko Žitnik, Marko Janković, Bojan Klemenc, Aleš Kumer, Aljaž Zrnec, Marko Bajec

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Tržaška 25, 1000 Ljubljana

[ime.priimek]@fri.uni-lj.si

Povzetek

V zadnjih petnajstih letih so postala omrežja izjemno popularna, saj predstavljajo močno orodje za predstavitev in analizo kompleksnih sistemov, sestavljenih iz velikega števila med seboj povezanih komponent. Takšne sisteme najdemo tako v realnem kot navideznem svetu, v naravi in tehnologiji, in so trenutno predmet številnih analiz in študij. Teorija omrežij je v preteklosti ponudila že številna pomembna spoznanja kot na primer pojav malega sveta, odpornost tehnoloških omrežij in obstoj skupnosti v socialnih omrežjih. Poleg slednjega pa je omogočila razlago različnih dinamičnih procesov kot na primer gibanje ljudi v družbi, navigacija po svetovnem spletu in pa nadzor v bioloških sistemih. Primere analize omrežij tako najdemo na številnih področjih, kljub temu pa je uporaba teorije velikih omrežij še vedno v veliki meri omejena zgolj na akademske kroge. V prispevku zato predstavimo osnovne lastnosti velikih realnih omrežij ter podamo nekatere primere uporabe.

Ključne besede

Analiza omrežij, lastnosti omrežij, realna omrežja, primeri uporabe.

Abstract

**LARGE REAL-WORLD NETWORKS**

Recently, networks have become extremely popular, since they represent a powerful tool for the representation and analysis of large complex systems of interacting parts. Such systems can be found in real and virtual world, nature and society, while real-world networks are currently part of numerous studies and analyses. Network theory has already resulted in important discoveries in the past like small-world phenomena, network robustness and community structure. Moreover, the analysis of networks also gave further comprehension about different dynamical processes occurring on networks like spreading of diseases and network controllability. Examples of network analysis can thus be found in various fields, nevertheless, theory of large networks is still mainly used only in academic circles. Therefore, we her present basic properties of large real-world networks and describe some interesting applications of network analysis.

Key words

Network analysis, network properties, real-world networks, applications.

# UVOD

Omrežja predstavljajo izjemno močno orodje za predstavitev in analizo kompleksnih sistemov, sestavljenih iz velikega števila med seboj povezanih komponent. Takšne sisteme najdemo tako v realnem kot navideznem svetu, v naravi in tehnologiji, in so trenutno predmet številnih analiz in študij. Teorija omrežij je v preteklosti ponudila že številna pomembna spoznanja in odkritja (Watts & Strogatz, 1998; Barabasi & Albert, 1999), ter omogočila razlago različnih procesov, kot je na primer navigacija po svetovnem spletu (Kleinberg, 2000) ali pa nadzor nad lastniško shemo podjetij (Vitali, Glattfelder & Battiston, 2011). Omenimo, da moč analize omrežij ne temelji na kompleksnosti posameznih povezav med komponentami nekega sistema, temveč na kompleksnih vzorcih medsebojnega povezovanja večjega števila komponent. Enostavnost predstavitve poljubnega sistema povezanih komponent pa je sicer eden od ključnih razlogov za izjemno popularnost omrežij v zadnjih letih (Newman, 2010; Newman, 2003).

Potrebno je poudariti, da sodobna analiza omrežij dejansko temelji na matematični teoriji grafov z začetki v 18. stoletju. Dočim so bila prva dela na področju večinoma zabavne ali teoretske narave, pa so že stoletje kasneje raziskovalci nakazali možnosti uporabe v elektrotehniki in molekularni kemiji. V 20. stoletju se je analiza omrežij izkazala kot učinkovito orodje v družbenih znanostih in bibliometriki, ter postala podpodročje operacijskih raziskav. Različne primere uporabe najdemo tudi v ekonomiji, biologiji, organski kemiji, ter še na številnih drugih področjih. Poleg tega je analiza omrežij tudi pomemben del raziskav kompleksnih sistemov v fiziki, predvsem z razmahom računalniške tehnologije ter pojavom velikih komunikacijskih omrežij konec prejšnjega stoletja pa so vidno zanimanje za omrežja pokazali tudi raziskovalci v računalništvu in informatiki.

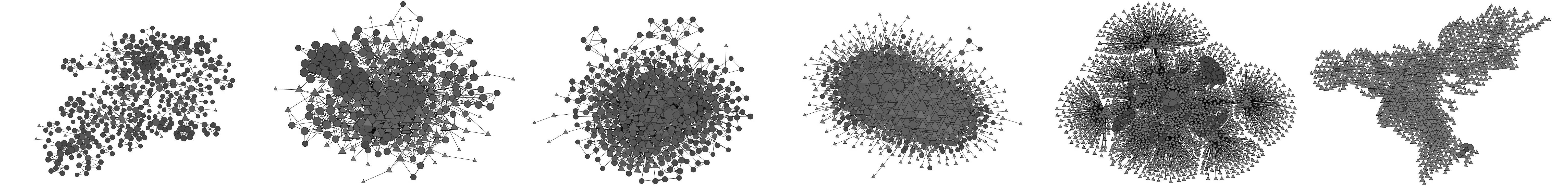
Analiza omrežij je tako danes izjemno interdisciplinarno področje na katerem delujejo tudi številni slovenski raziskovalci (poleg avtorjev prispevka). Med njimi Jure Leskovec iz Univerze Stanford; Vladimir Batagelj, Andrej Mrvar, Anuška Ferligoj, Valentina Hlebec, Tina Kogovšek in Gregor Petrič iz Univerze v Ljubljani; Matjaž Perc iz Univerze v Mariboru, Bosiljka Tadić in Marko Grobelnik iz Instituta Jožef Stefan; ter še nekateri drugi. Med najvidnejšemi prispevki lahko sicer omenimo svetovno uveljavljen program za analizo velikih omrežij *Pajek* (Batagelj & Mrvar, 1998; de Nooy, Mrvar & Batagelj, 2012), ki ga avtorji razvijajo že več kot petnajst let.

Kljub zgornjemu pa je uporaba teorije velikih omrežij še vedno v veliki meri omejena na akademske kroge, dočim so primeri uporabe v praksi razmeroma redki. V prispevku zato predstavimo skupne lastnosti realnih omrežij ter izpostavimo nekatere zanimivejše možnosti uporabe.

V nadaljevanju najprej podamo osnovne pojme analize omrežij ter predstavimo različne vrste velikih realnih omrežij (razdelek 2). V razdelku 3 nato izpostavimo skupne lastnosti realnih omrežij, vključujoč nekatera vidnejša odkritja v zadnjih petnajstih letih. V razdelku 4 predstavimo še izbrane možnosti uporabe analize omrežij v praksi, s poudarkom na socialnih (družbenih) omrežjih. Sledita sklep in zaključek v razdelku 5.

# REALNA OMREŽJA

Zgradbo omrežja navadno predstavimo z diskretnim matematičnim objektom, ki mu pravimo graf. Graf  je sestavljen iz množice vozlišč , , ter množice povezav med njimi , . V primeru storitev za socialno mreženje (npr. *Facebook*), vozlišča predstavljajo različne uporabnike storitve, povezave pa ustrezajo relaciji prijateljstva. Povezave so tako neusmerjene, dočim nekatere vrste omrežij zahtevajo usmerjene povezave. V primeru svetovnega spleta, vozlišča predstavljajo množico spletnih strani, ki so med seboj povezane preko usmerjenih spletnih povezav. Torej, . Zgornji definiciji  lahko še razširimo tako, da dopuščata več vzporednih povezav med vozlišči ter zanke (povezave vozlišča s samim seboj). Poleg omenjenega pa omrežje vključuje tudi poljubno dodatno znanje o vozliščih in povezavah grafa. Tako so le-ta lahko različnih tipov, vsebujejo časovne značke in druge podatke.



Slika 1: Od leve proti desni: omrežje sodelovanj med znanstveniki (Newman, 2006); prijateljstva v družabna omrežju *Facebook* (Blagus, Šubelj & Bajec, 2012); metabolično omrežje (Jeong, Tombor, Albert, Oltvai & Barabasi, 2000); spletni dnevniki o ameriški politiki (Adamic & Glance, 2004); spletne strani v domeni *amazon.com* (Šubelj & Bajec, 2012a) in evropsko avtocestno omrežje (Šubelj & Bajec, 2011). Oblika vozlišč ustreza gostoti omrežja v neposredni okolici posameznega vozlišča (Soffer & Vazquez, 2005).

Poleg družabnih omrežij in svetovnega spleta obstajajo še številne druge vrste realnih omrežij (Slika 1). V grobem jih delimo v štiri kategorije (Newman, 2003):

* socialna (družbena) omrežja, kjer vozlišča predstavljajo ljudi ali ustanove, povezave pa ustrezajo neki obliki interakcije med njimi. V preteklosti so bila najpogosteje preučevana klasična omrežja prijateljstev (angl. offline social), spletna socialna ali družabna omrežja (angl. online social), omrežja sodelovanj med znanstveniki in filmskimi igralci (angl. author, actor collaboration), ter nekatera druga;
* informacijska omrežja, kjer povezave v omrežju ustrezajo toku informacij (ali podatkov) skozi analiziran sistem. Sem sodijo svetovni splet (angl. web graph), omrežje citiranj med znanstvenimi prispevki (angl. citation), ter mobilna in druga komunikacijska omrežja (angl. communication);
* biološka omrežja, ki ponazarjajo interakcije med geni, celicami, beljakovinami, ipd. v živih organizmih (angl. gene regulatory, metabolic, protein-protein interaction);
* tehnološka omrežja, ki navadno predstavljajo neko umetno infrastrukturo, ki je podvržena tehnološkim (ali drugim) omejitvam. Sem sodi internetno omrežje (angl. Internet map), cestno omrežje (angl. road), električno omrežje (angl. power grid) ter omrežja zgrajena na podlagi programske kode (angl. software).

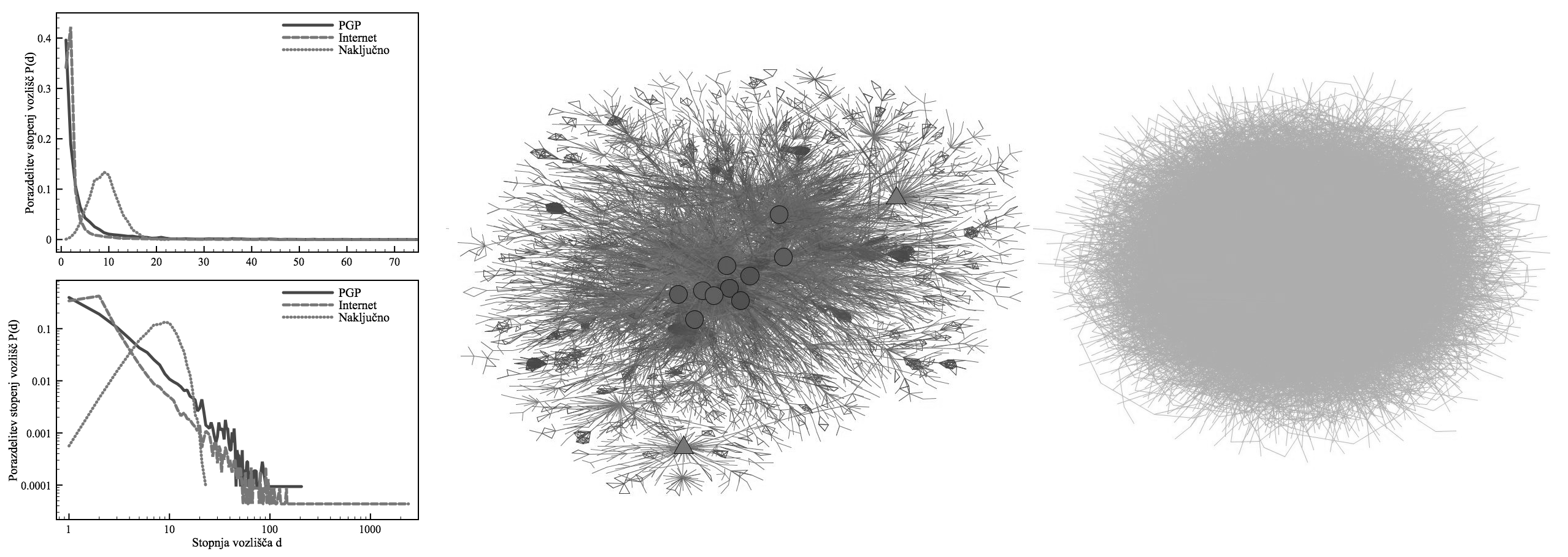
Delitev ni stroga, saj so na primer številna komunikacijska omrežja prav tako tudi socialna. Poleg tega se v zadnjih letih vse pogosteje preučuje omrežja, ki ne sodijo v nobeno od zgornjih kategorij (npr. različna transportna ali ekonomska omrežja).

# LASTNOSTI OMREŽIJ

Kljub svoji raznolikosti pa imajo realna omrežja številne skupne lastnosti. Nekatere izmed teh predstavimo v nadaljevanju (zaradi enostavnosti predpostavimo neusmerjene povezave).

## Porazdelitev stopenj vozlišč ter potenčni zakon

Stopnja vozlišča je definirana kot število povezav z enim krajiščem v obravnavanem vozlišču. Naj bo  povprečna stopnja vozlišča, . V primeru socialnih omrežij,  predstavlja povprečno število prijateljev posamezne osebe. Izkaže se, da je porazdelitev stopenj vozlišč v večini realnih omrežij značilno drugačna kot na primer v naključnih omrežjih (Erdős & Rényi, 1959). Pri slednjih predpostavimo, da osebe izbirajo prijatelje popolnoma naključno (verjetnost povezave med poljubnim parom različnih vozlišč je ). V naključnih omrežjih so stopnje vozlišč porazdeljene tesno okrog povprečja  (podobno kot pri normalni porazdelitvi), dočim je porazdelitev stopenj v realnih omrežjih močno razpotegnjena v desno (angl. right skewed). Natančneje, porazdelitev pogosto sledi potenčnemu zakonu ,  (angl. power law), omrežja z omenjeno lastnostjo pa označujemo z izrazom brezlestvična omrežja (Barabasi & Albert, 1999) (angl. scale-free). Omenimo, da se potenčna porazdelitev odraža kot premica v dvojno logaritemskem prikazu, kjer je  naklon premice (Slika 2).



Slika 2: (levo) Porazdelitev stopenj vozlišč v družabnem omrežju *Pretty Good Privacy* (*PGP*) (Boguná, Pastor-Satorras, Díaz-Guilera & Arenas, 2004), v internetnem omrežju na nivoju avtonomnih sistemov (http://www-personal.umich.edu/~mejn/netdata/) ter v naključnem omrežju (Erdős & Rényi, 1959). (na sredini) Družabno omrežje *PGP*. Stopnja prikazanih vozlišč je vsaj 75 (povprečje znaša le 4.6), dočim oblika ustreza gostoti v okolici (Soffer & Vázquez, 2005). (desno) Naključno omrežje z enakim številom vozlišč in povezav.

Ena od ključnih lastnosti brezlestvičnih omrežij je obstoj vozlišč z izjemno visoko stopnjo (Barabási & Albert, 1999; Kleinberg, 2000) (angl. hub). Na primer, največja stopnja v družabnem omrežju na Slika 2 znaša 206, dočim je povprečna stopnja  zgolj 4.6. Podobno, internetno omrežje zgoraj ima , medtem ko je največja stopnja kar 2390. Število vozlišč v omenjenih omrežjih je 10680 in 22963.

Obstaja več razlag o nastanku brezlestvičnih omrežij. Eno od njih je načelo prednostnega povezovanja (Barabási & Albert, 1999) (angl. preferential attachment), ki pravi, da bogati bogatijo (angl. rich get richer). V primeru družabnih omrežij slednje pomeni, da osebe z velikim številom občudovalcev lažje pridobivajo nove občudovalce, kar povzroči obstoj oseb oziroma vozlišč z izjemno visoko stopnjo v omrežju.

Omenimo še, da imajo vozlišča z visoko stopnjo prav poseben vpliv na različne dinamične procese, ki se odvijajo nad omrežji (npr. širjenje novic po socialnem omrežju ali virusov preko svetovnega spleta). Natančneje, pri  lahko že manjše število okuženih vozlišč (vozlišča z visoko stopnjo) povzroči razširjenost po celotnem omrežju (Pastor-Satorras & Vespignani, 2001; Sinha, 2011). Za družabno omrežje *PGP*  znaša 2.2, dočim  za internetno omrežje zgoraj. Podobno imajo omenjena vozlišča močan vpliv na odpornost in ranljivost realnih omrežij (glej spodaj).

## Razdalje med vozlišči ter pojav malega sveta

Stanley Milgram je v 60. letih prejšnjega stoletja izvedel naslednji eksperiment. Posame- znikom je zadal nalogo naj preko zaporedja pisem dosežejo neko izbrano osebo. Večina pisem se je tekom eksperimenta sicer izgubila, ostala pisma pa so prišla na cilj v presenetljivo majhnem številu korakov. V objavljenih primerih je slednje v povprečju znašalo le šest korakov, od koder izhaja izraz angl. six degrees of separation (Milgram, 1967). Eksperiment velja za prvi prikaz pojava malega sveta (angl. small-world effect), ki pravi, da sta v realnih omrežjih poljubni vozlišči povezani preko zelo kratke poti.

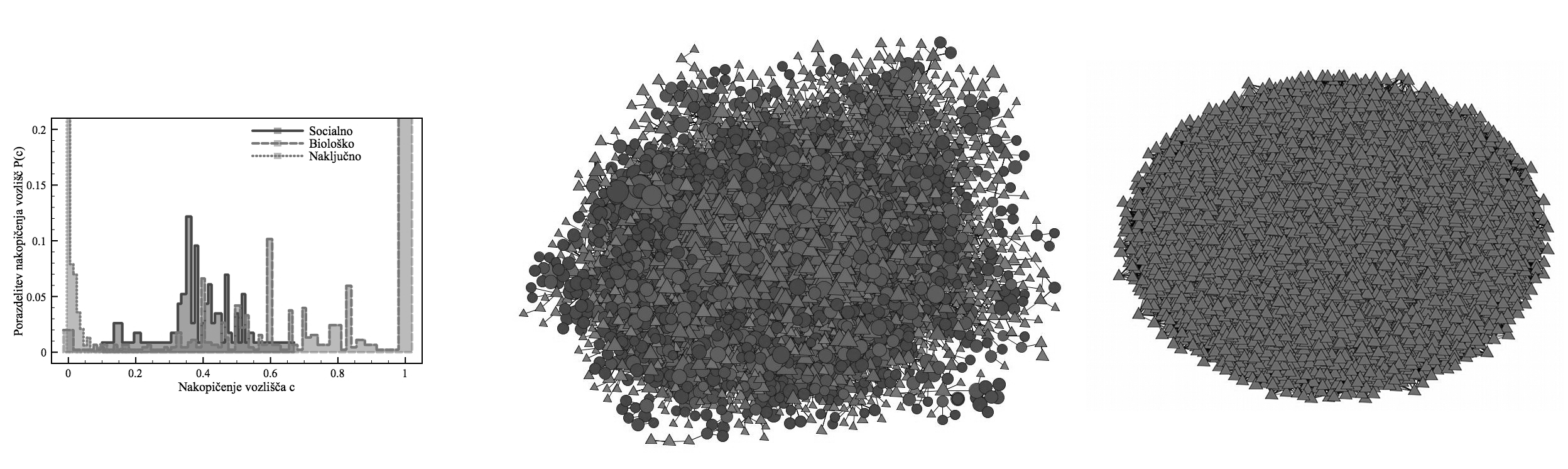
Naj bo  razdalja med vozliščema  in , ki je definirana kot število povezav v najkrajši poti med  in . Pojav malega sveta najpogosteje opazujemo preko povprečne razdalje  (Newman, 2003).

Med omrežja malega sveta (Watts & Strogatz, 1998) navadno štejemo tista, pri katerih povprečna razdalja  ne narašča s številom vozlišč  (oziroma narašča zelo počasi kot npr. ). Kot primer povejmo, da je povprečna razdalja med 700 milijoni uporabniki storitve *Facebook* v letu 2011 znašala zgolj  (Backstrom, Boldi, Rosa, Ugander & Vigna, 2012), dočim je največja razdalja med preko milijardo spletnimi stranmi *Yahoo!* manjša od 8 (Kang, Tsourakakis, Appel, Faloutsos & Leskovec, 2010). Podobno velja za mnoga druga velika omrežja, kjer je .

Potrebno je omeniti, da kratka razdalja med vozlišči ni presenetljiva sama po sebi, saj imajo podobno lastnost tudi naključna omrežja (Erdős & Rényi, 1959). Vendar pa slednja niso tako (lokalno) gosta ali nakopičena kot primerljiva realna omrežja. Natančneje, naj bo  nakopičenost (Watts & Strogatz, 1998; Newman & Park, 2003) (angl. clustering coefficient), ki meri tranzitivnost v omrežju. V primeru socialnih omrežij  dejansko ustreza verjetnosti, da je prijatelj prijatelja prav tako prijatelj. Večina realnih omrežjih ima  med 0.3 in 0.6, dočim je  v primerljivih naključnih omrežjih blizu nič (Slika 3).

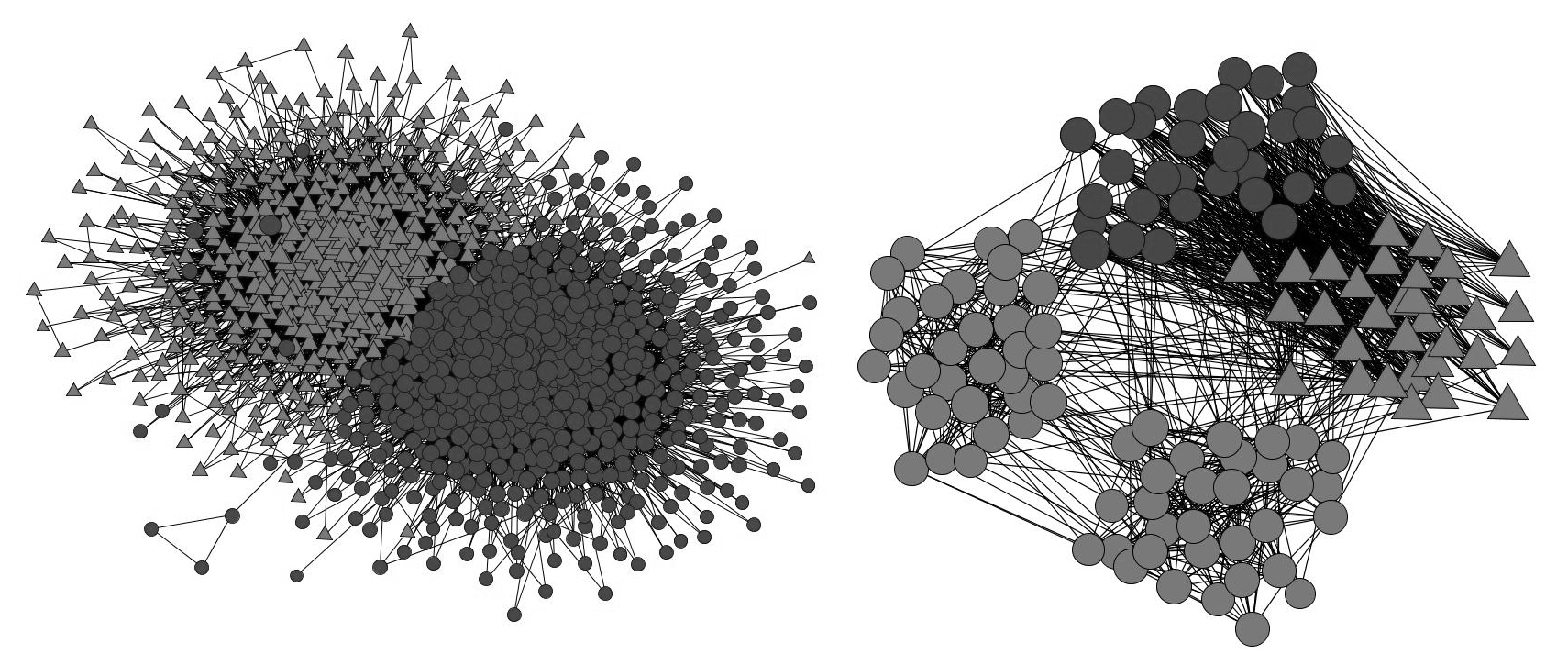
## Kopičenje vozlišč ter strukturni moduli v omrežjih

Realna omrežja pogosto vsebujejo skupnosti ali komune (Girvan & Newman, 2002) (angl. communities), s čimer označujemo množice tesno povezanih vozlišč, ki pa so med seboj le šibko povezane. Skupnosti v socialnih omrežjih ustrezajo osebam s podobnimi interesi, medtem ko skupnosti v spletnih omrežjih navadno vežejo spletne strani s podobno vsebino (Slika 4). Omenimo, da je odkrivanje skupnosti trenutno eno najbolj »vročih« področij v teoriji omrežij (Porter, Onnela & Mucha, 2009; Fortunato, 2010).



Slika 3: (levo) Porazdelitev nakopičenja vozlišč (Watts & Strogatz, 1998) v socialnem omrežju (Girvan & Newman, 2002), biološkem omrežju (Jeong, Tombor, Albert, Oltvai & Barabási, 2000) ter naključnem omrežju (Erdős & Rényi, 1959). (na sredini) Omrežje sodelovanj med slovenskimi znanstveniki do leta 2011 (nakopičenje znaša 0.47). (desno) Naključno omrežje z enakim številom vozlišč in povezav (nakopičenje je manjše od 0.01). Oblika vozlišč ustreza gostoti omrežja v okolici vozlišča (Soffer & Vázquez, 2005).

Odkrivanje skupnosti dejansko ustreza razvrščanju (angl. data clustering) vozlišč omrežja glede na razdalje med vozlišči (glej zgoraj). Poleg skupnosti pa realna omrežja običajno vsebujejo tudi druge značilne skupine vozlišč ali strukturne module (angl. modules). Zadnje raziskave kažejo, da je v mnogih omrežjih moč najti šibko povezane množice vozlišč, ki pa so podobno povezane z ostalim omrežjem (Newman & Leicht, 2007; Šubelj & Bajec, 2012a) (Slika 4). Slednje označujemo z izrazom funkcijski moduli (Šubelj & Bajec, 2012b) (angl. functional modules), saj navadno vežejo komponente nekega sistema, ki imajo podobno funkcijo ali vlogo. Na primer, funkcijski moduli v programskih omrežjih ustrezajo množicam programskih konstruktov, ki podpirajo podobno funkcionalnost v okviru obravnavanega projekta (Šubelj & Bajec, 2012b; Šubelj & Bajec, 2012a) (npr. vhod/izhod). Podobno funkcijski moduli v bioloških omrežjih ustrezajo npr. beljakovinam, ki opravljajo podobno nalogo v človeškem telesu (Pinkert, Schultz & Reichardt, 2010). Značilne funkcijske module je sicer moč najti še v informacijskih, različnih tehnoloških ter tudi socialnih omrežjih.

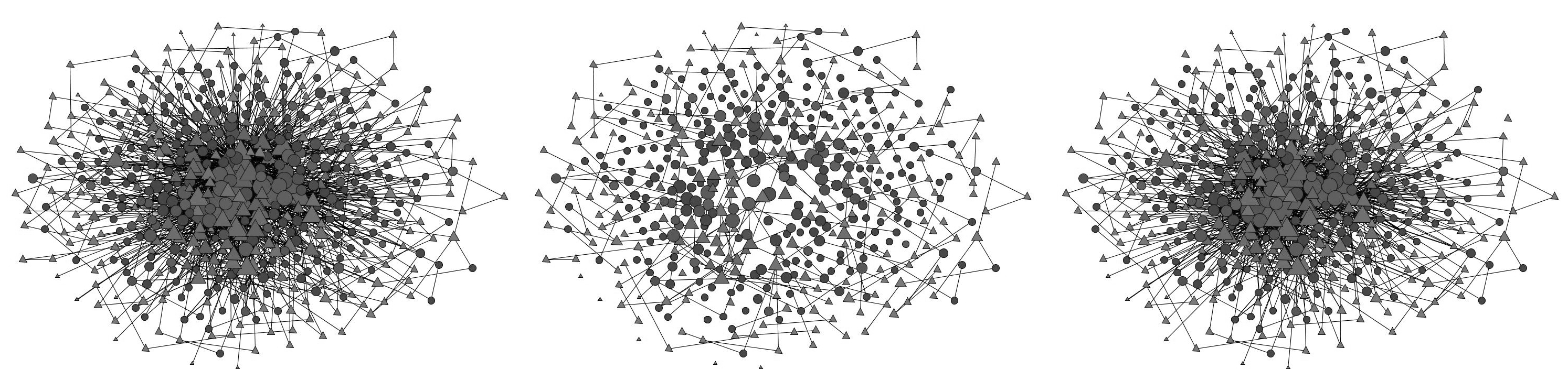


Slika 4: (levo) Omrežje povezav med spletnimi dnevniki o ameriški politiki (Adamic & Glance, 2005). Omrežje vsebuje dve tesno povezani skupnosti, ki pa sovpadata z delitvijo na levo in desno usmerjene dnevnike (Adamic & Glance, 2005). (desno) Sintetično omrežje, ki vsebuje dve skupnosti in dva funkcijska modula (dve množici podobno povezanih vozlišč). Oblika vozlišč ustreza razvrstitvi v skupine.

## Odpornost, ranljivost in nadzor realnih omrežij

Kot že omenjeno zgoraj, v večini realnih (brezlestvičnih) omrežij že majhno število izbranih vplivnih vozlišč povzroči razširjenost po celotnem omrežju (Pastor-Satorras & Vespignani, 2001; Sinha, 2011) (npr. širjenje bolezni med preko socialnega omrežja, novic po komunikacijskem omrežju ali pa napak po tehnoloških omrežjih). Pomembno vlogo pri tem igrajo vozlišča z visoko stopnjo, v splošnem pa taka vplivna vozlišča označujemo z izrazom središčna vozlišča (Freeman, 1977) (angl. centrality). Obstaja več mer središčnosti (Freeman, 1977; Freeman, 1979), med njimi tudi , kjer je  stopnja obravnavanega vozlišča. Za podrobnejši pregled glej (Newman, 2010; Newman, 2006).

Vozlišča z visoko stopnjo imajo pomemben vpliv tudi na odpornost in ranljivost omrežij. Brezlestvična omrežja so izjemno odporna na naključno odstranjevanje vozlišč, dočim odstranitev že manjšega števila vozlišč z visoko stopnjo lahko povzroči, da omrežje razpade na več ločenih komponent (Albert, Jeong & Barabási, 2000). Na primer, internetno omrežje je relativno odporno na naključne izpade strojne opreme, a izjemno ranljivo na namerne napade nad ključnimi deli omrežja (Cohen, Erez, Ben-Avraham & Havlin, 2000) (Slika 5).



Slika 5: (levo) Internetno omrežje na nivoju avtonomnih sistemov zbrano v letu 2003 (Leskovec, Lang, Dasgupta & Mahoney, 2009). (na sredini) Omrežje po odstranitvi 10% vozlišč z največjo stopnjo. (desno) Omrežje po odstranitvi 10% naključno izbranih vozlišč. Oblika vozlišč ustreza gostoti omrežja v neposredni okolici vozlišča (Soffer & Vázquez, 2005).

V letu 2011 so številni avtorji preučevali nadzor v realnih omrežjih (Liu, Slotine & Barabási, 2011; Vitali, Glattfelder & Battiston, 2011). Slednje v primeru spletnih socialnih storitev pomeni, da je moč preko nekaterih vozlišč (uporabnikov storitve) nadzirati mnenje vseh uporabnikov v omrežju (pri določenih predpostavkah). Omenjena vozlišča lahko označimo z izrazom upravljalska vozlišča (angl. driver). Zanimivo, upravljalska vozlišča navadno ne sovpadajo z vozlišči z visoko stopnjo (glej zgoraj), dočim je število upravljalskih vozlišč  potrebnih za nadzor v brezlestvičnih omrežjih odvisno le od povprečne stopnje  in eksponenta  (Liu, Slotine & Barabási, 2011). Natančneje, pri , .

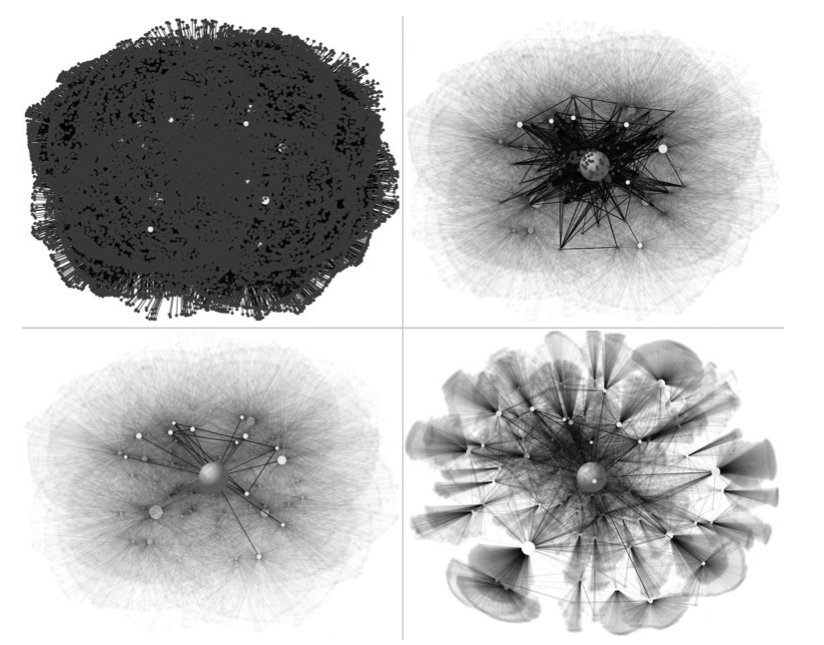
Na primer, družabno omrežje *PGP* na Slika 2 ni moč nadzirati preko manj kot 64% uporabnikov spletne storitve *PGP*. Za primerjavo,  v nekaterih drugih spletnih socialnih omrežjih znaša okrog 30%, dočim 80% v bioloških omrežjih (angl. regulatory) ter okrog 50% v primeru interneta (Liu, Slotine & Barabási, 2011). Na drugi strani pa je  v omrežjih lastništev med podjetji manj kot 3% (Vitali, Glattfelder & Battiston, 2011). Podobno je programski jezik *Java* moč nadzirati preko zgolj 17% programskih konstruktov, ki sestavljajo jedro jezika (Šubelj & Bajec, 2012c).

# PRIMERI UPORABE

V nadaljevanju so predstavljeni nekateri praktični primeri uporabe analize omrežij.

## Družabno omrežje Facebook

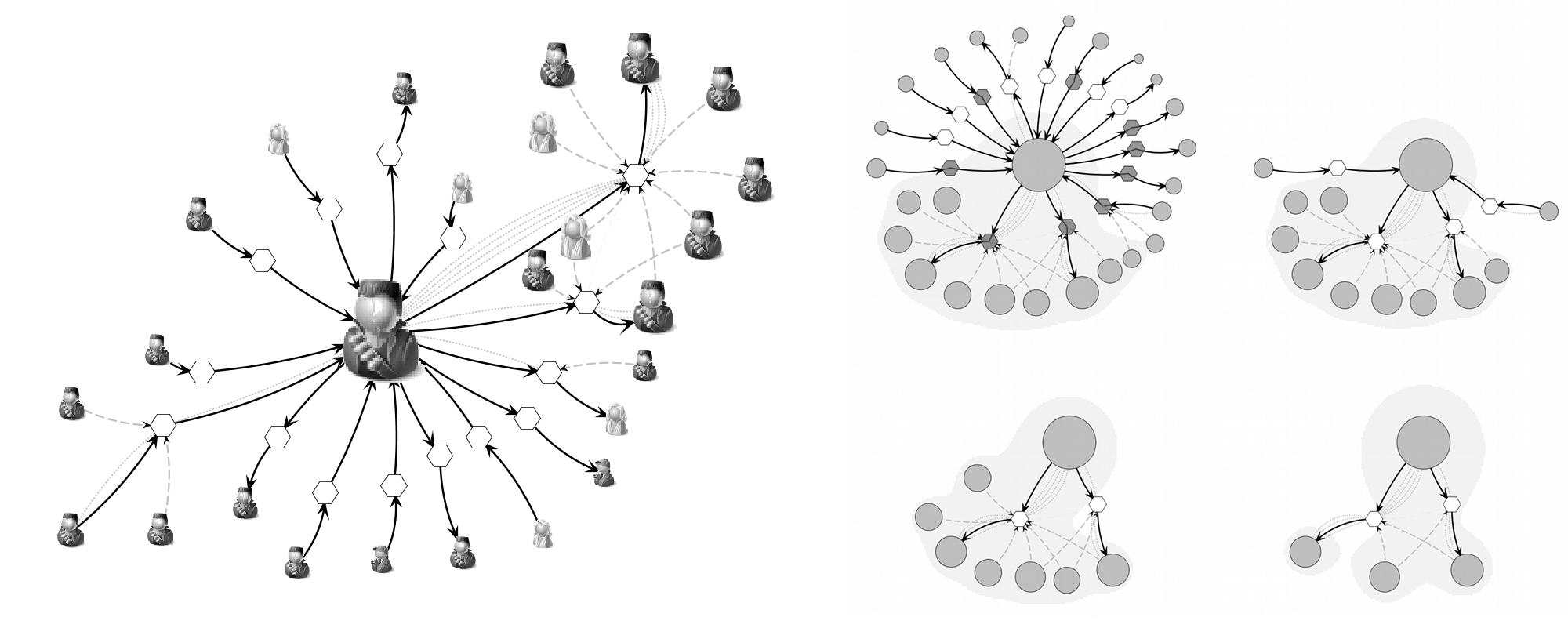
Slika 6 prikazuje 25 tisoč uporabnikov družabnega omrežja *Facebook*, ki so povezani glede na relacijo prijateljstva (Catanese, De Meo, Ferrara & Fiumara, 2010). Vozlišča z visoko središčnostjo, ki imajo izjemno močan vpliv na celotno omrežje (razdelek 3), so posebej izpostavljena.



Slika 6: Različni prikazi družabnega omrežja *Facebook* (glej tekst).  
(Vir: Catanese, De Meo, Ferrara & Fiumara, 2010.)

## Odkrivanje avtomobilskih goljufij

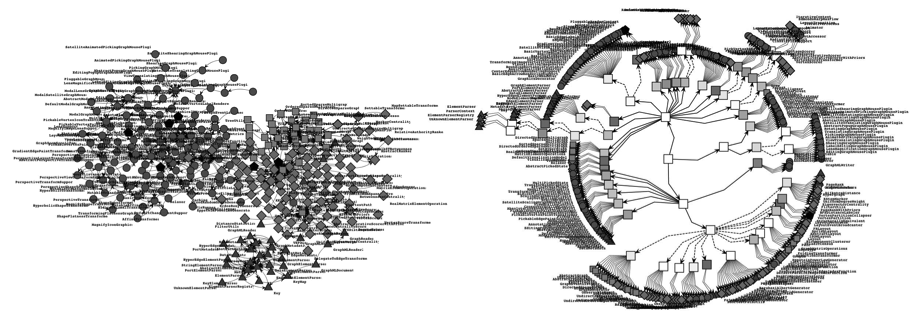
Analiza omrežij se pogosto uporablja tudi za odkrivanje anomalij (odstopanj) v podatkih. Na primer, z analizo omrežij prometnih nesreč lahko razkrijemo goljufive skupine posameznikov, ki uprizarjajo prometne nesreče ter se tako okoristijo preko svojega zavarovanja (Šubelj, Furlan & Bajec, 2009; Šubelj, Furlan & Bajec, 2011). S pomočjo razširjanja sumljivosti po omrežju lahko tako odkrijemo goljufive osebe, nesreče in vozila, ter izpostavimo ključne povezave med njimi (Slika 7). Uporabljeni pristopi so podobni algoritmu *PageRank* (Brin & Page, 1998), ki ga v svojem brskalniku uporablja *Google*, ter pa tudi nekaterim meram središčnosti vozlišč (razdelek 3).



Slika 7: (levo) Socialno omrežje udeležencev povezanih prometnih nesreč. (desno) Rezultat odkrivanja avtomobilskih goljufov v omrežju na levi. Okrogla (oglata) vozlišča predstavljajo udeležence (prometne nesreče), velikost vozlišč pa je sorazmerna njihovi sumljivosti.   
(Vir: Šubelj, Furlan & Bajec, 2011)

## Reorganizacija programske opreme

Z uporabo primernih algoritmov lahko v omrežju razkrijemo celotno hierarhijo različnih strukturnih modulov (razdelek 3). V primeru omrežij, ki predstavljajo kompleksno programsko kodo, omenjena hierarhija dejansko sovpada s programskimi paketi ter tako razkriva odvisnosti med različnimi deli programske opreme (Šubelj & Bajec, 2012b; Šubelj & Bajec, 2012c) (Slika 8). Slednje je moč uporabiti za reorganizacijo paketov (npr. glede na načelo modularnosti, funkcionalnosti) ter napovedovanje odvisnosti med različnimi programskimi konstrukti (Šubelj & Bajec, 2012c; Lavbič, Lajovic & Krisper, 2010).



Slika 8: (levo) Omrežje odvisnosti med programskimi razredi knjižnice *JUNG* (O’Madadhain, Fisher, White, Smyth & Boey, 2005)*.* (desno) Hierarhija strukturnih modulov razkrita z algoritmom v (Šubelj & Bajec, 2012b). Oblika vozlišč ustreza programskim paketom.   
(Vir: Šubelj & Bajec, 2012c.)

## Analiza znanstvenih prispevkov

V preteklosti so bila pogosto preučevana tudi omrežja citiranj med znanstvenimi prispevki iz različnih področij. Izkaže se, da je struktura znanosti podobna obliki črke »U« (Rosvall & Bergstrom, 2008). Natančneje, naravoslovne in družboslovne znanosti so razmeroma dobro ločene med seboj (levo in desno zgoraj), medtem ko so povezane preko interdisciplinarnih znanosti kot sta medicina in biologija (spodaj).

# SKLEP

V prispevku podamo osnovne pojme analize omrežij ter predstavimo različne vrste realnih omrežij. V nadaljevanju nato izpostavimo skupne lastnosti velikih realnih omrežij, vključujoč nekatera vidnejša odkritja v zadnjih petnajstih letih. Na koncu predstavimo še izbrane primere uporabe analize omrežij v praksi, s poudarkom na socialnih (družbenih) omrežjih.

Zaključimo, da teorija omrežij predstavlja močno orodje za analizo kompleksnih sistemov sestavljenih iz velikega števila povezanih komponent.

# VIRI IN LITERATURA

1. Adamic, Lada A., Glance, Natalie: The political blogosphere and the 2004 U.S. election, Proceedings of the KDD Workshop on Link Discovery, Chicago, IL, USA, 2005, str. 36 - 43.
2. Albert, Reka, Jeong, Hawoong, Barabási, Albert L.: Error and attack tolerance of complex networks, Nature, 2000, št. 406, str. 378 - 382.
3. Backstrom, Lars, Boldi, Paolo, Rosa, Marco, Ugander, Johan, Vigna, Sebastiano: Four degrees of separation, e-print arXiv:1111.4570v3, 2012.
4. Barabási, Albert L., Albert, Reka: Emergence of scaling in random networks, Science, 1999, št. 286, str. 509 - 512.
5. BATAGELJ, Vladimir, MRVAR, Andrej: Pajek: Program for large network analysis, Connections, 1998, št. 21, str. 47 - 57.
6. Blagus, Neli, Šubelj, Lovro, Bajec, Marko: Self-similar scaling of density in complex real-world networks, Physica A, 2012, št. 391, str. 2794 - 2802.
7. Boguná, Marian, Pastor-Satorras, Romualdo, Díaz-Guilera, Albert, Arenas, Alex: Models of social networks based on social distance attachment, Physical Review E, 2004, št. 70, str. 056122.
8. BRIN, Sergey, PAGE, Lawrence: The anatomy of a large-scale hypertextual web search engine, Computer Networks and ISDN Systems, 1998, št. 30, str. 107 - 117.
9. CATANESE, Salvatore, DE MEO, Pasquale, FERRARA, Emilio, FIUMARA, Giacomo: Analyzing Facebook friendship graph, Proceedings of the International Workshop on Mining the Future Internet, Berlin, Germany, 2010, str. 6.
10. COHEN, Reuven, EREZ, Keren, Ben-Avraham, Daniel, Havlin, Shlomo, Resilience of the Internet to random breakdowns, Physical Review Letters, 2000, št. 85, str. 4626.
11. DE NOOY, Wouter, MRVAR, Andrej, BATAGELJ, Vladimir: Exploratory social network analysis with Pajek, (Revised and Expanded 2nd Edition) 2012, Cambridge University Press.
12. Egerstedt, Magnus, Complex networks: Degrees of control, Nature, 2011, št. 473, str. 158 - 159.
13. Erdős, Paul, Rényi, Alfréd: On random graphs I, Publicationes Mathematicae Debrecen, 1959, št. 6, str. 290–297.
14. FORTUNATO, Santo: Community detection in graphs, Physics Reports, 2010, št. 486, str. 75 - 174.
15. FREEMAN, Linton: A set of measures of centrality based upon betweenness, Sociometry, 1977, št. 40, str. 35 - 41.
16. FREEMAN, Linton: Centrality in social networks: Conceptual clarification, Social Networks, 1979, št. 1, str. 215 - 239.
17. GIRVAN, Michele, Newman, Mark E. J.: Community structure in social and biological networks, Proceedings of the National Academy of Sciences of the USA, 2002, št. 99, str. 7821 - 7826.
18. Jeong, Hawoong, Tombor, Bwalint, Albert, Reka, Oltvai, Zoltán N., Barabási, Albert L.: The large-scale organization of metabolic networks, 2000, Nature, št. 407, str. 651 - 654.
19. Kang, U, Tsourakakis, Charalampos E., Appel, Ana P., Faloutsos, Christos, Leskovec, Jure: Radius plots for mining tera-byte scale graphs: Algorithms, patterns, and observations. Proceedings of the SIAM International Conference on Data Mining, Columbus, OH, USA, 2010, str. 548 - 558.
20. Kleinberg, Jon M.: Authoritative sources in a hyperlinked environment, Journal of the ACM, 1999, št. 46, str. 604 - 632.
21. Kleinberg, Jon M.: Navigation in a small world, Nature, 2000, št. 406, str. 845.
22. LAVBIČ, Dejan, LAJOVIC, Iztok, KRISPER, Marjan: Facilitating information system development with Panoramic view on data, Computer Science and Information Systems, 2010, št. 7, str. 737 - 768.
23. LESKOVEC, Jure, LANG, Kevin J., DASGUPTA, Anirban, MAHONEY, Michael W.: Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, Internet Mathematics, 2009, št. 6, str. 29 - 123.
24. LIU, Yang-Yu, Slotine, Jean-Jacques, Barabási, Albert L.: Controllability of complex networks, Nature, 2011, št. 473, str. 167 - 173.
25. Milgram, Stanley: The small world problem, Psychology Today, 1967, št. 1, str. 60 - 67.
26. Newman, Mark E. J.: The structure and function of complex networks, SIAM Review, 2003, št. 45, str. 167 - 256.
27. Newman, Mark E. J.: Finding community structure in networks using the eigenvectors of matrices, Physical Review E, 2006, št. 74, str. 036104.
28. Newman, Mark E. J.: Networks: An introduction, 2010, Oxford University Press.
29. Newman, Mark E. J., LEICHT, Elizabeth A.: Mixture models and exploratory analysis in networks, Proceedings of the National Academy of Sciences of the USA, 2007, št. 104, str. 9564 - 9569.
30. Newman, Mark E. J., PARK, Juyong: Why social networks are different from other types of networks, Physical Review E, 2003, št. 68, str. 036122.
31. O’Madadhain, Joshua, Fisher, Danyel, White, Scott, Smyth, Padhraic, Boey, Yan-Biao: Analysis and visualization of network data using JUNG, Journal of Statistical Software, 2005, št. 10, str. 35.
32. Pastor-Satorras, Romualdo, VESPIGNANI, Alessandro: Epidemic spreading in scale-free networks, Physical Review Letters, 2001, št. 86, str. 3200 - 3203.
33. Pinkert, Stefan, Schultz, Jörg, Reichardt, Jörg: Protein interaction networks: More than mere modules, PLoS Computational Biology, 2010, št. 6, str.. e1000659.
34. Porter, Mason A., Onnela, Jukka-Pekka, Mucha, Peter J.: Communities in networks, Notices of the American Mathematical Society, 2009, št. 56, str. 1082 - 1097.
35. ROSVALL, Martin, BERGSTROM, Carl: Maps of random walks on complex networks reveal community structure, Proceedings of the National Academy of Sciences of the USA, 2008, št. 105, str. 1118 - 1123.
36. SINHA, Sitabhra: Few and far between, Physics, 2011, št. 4, str. 81.
37. Soffer, Sara N., Vázquez, Alexei: Network clustering coefficient without degree-correlation biases, Physical Review E, 2005, št. 71, str. 057101.
38. Šubelj, Lovro, FURLAN, Štefan, Bajec, Marko: Odkrivanje goljufij na osnovi analize socialnih omrežij, Zbornik Prispevkov Konference Dnevi Slovenske Informatike, Portorož, Slovenija, 2009, str. 10.
39. Šubelj, Lovro, Bajec, Marko: Robust network community detection using balanced propagation, European Physical Journal B, 2011, št. 81, str. 353 - 362.
40. Šubelj, Lovro, FURLAN, Štefan, Bajec, Marko: An expert system for detecting automobile insurance fraud using social network analysis, Expert Systems with Applications, 2011, št. 38, str. 1039 - 1052.
41. Šubelj, Lovro, Bajec, Marko: Ubiquitousness of link-density and link-pattern communities in real-world networks, European Physical Journal B, 2012a, št. 85, str. 32.
42. Šubelj, Lovro, Bajec, Marko: Clustering assortativity, communities and functional modules in real-world networks, e-print arXiv:1202.3188v1, 2012b.
43. Šubelj, Lovro, Bajec, Marko: Software engineering through network science: Review, analysis and applications, Proceedings of the KDD Workshop on Software Mining, Beijing, China, 2012c, str. 8.
44. Vitali, Stefania, Glattfelder, James B., Battiston, Stefano: The network of global corporate control. PLoS ONE, 2011, št. 6, str. e25995.
45. WATTS, Duncan J., Strogatz, Steven H.: Collective dynamics of ‘small-world’ networks, Nature, 1998, št. 393, str. 440 - 442.
46. Pridobljeno iz <http://www-personal.umich.edu/~mejn/netdata/>.



Lovro Šubelj je mladi raziskovalec in asistent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Poučuje predvsem predmete s področja podatkovnih baz. Raziskovalno se ukvarja z analizo realnih omrežij, natančneje z odkrivanjem značilnih skupin vozlišč v velikih kompleksnih omrežjih. Je avtor ali soavtor številnih prispevkov v strokovnih in znanstvenih publikacijah.



Slavko Žitnik je doktorski študent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani in hkrati zaposlen kot mladi raziskovalec iz gospodarstva v podjetju Optilab d.o.o.. Raziskovalno se ukvarja predvsem s procesiranjem besedil, bolj natančno z razpoznavanjem entitet in povezav med njimi z uporabo metod strojnega učenja in semantičnih tehnologij.



Marko Janković je mladi raziskovalec v Laboratoriju za podatkovne tehnologije na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Njegova glavna raziskovalna področja obsegajo dogodkovno vodene arhitekture, procesiranje in rudarjenje po podatkovnih tokovih ter internet stvari.



Bojan Klemenc je podiplomski študent in asistent na Fakulteti za računalništvo in informatiko  Univerze v Ljubljani. Diplomiral je leta 2008. Poučuje predmete s področja podatkovnih baz, operacijskih sistemov, interaktivne vizualizacije podatkov in navidezne resničnosti. Raziskovalno se ukvarja z vizualizacijo podatkov in obogateno resničnostjo.



Aleš Kumer je asistent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani.



Aljaž Zrnec magistriral leta 2002 na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Leta 2006 je doktoriral s področja konstruiranja metodologij. Zaposlen je v Laboratoriju za podatkovne tehnologije kot asistent za področje podatkovnih baz. Na raziskovalnem področju se ukvarja s konstruiranjem metodologij, podatkovnimi bazami NoSQL in računalništvom v oblaku. Je avtor ali soavtor številnih prispevkov v strokovnih in znanstvenih publikacijah.



Marko Bajec je izredni profesor na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, kjer poučuje dodiplomske in podiplomske predmete s področja razvoja informacijskih sistemov in podatkovnih baz. Raziskovalno se ukvarja z metodami in pristopi k snovanju in razvoju informacijskih sistemov, obvladovanjem informatike ter v zadnjih letih predvsem s podatkovnimi tehnologijami za predstavitev, analizo in vizualizacijo podatkov. Leta 2009 je ustanovil Laboratorij za podatkovne tehnologije ter prevzel njegovo vodenje. Je član številnih domačih in tujih združenj, komisij in odborov. V okviru fakultete je vodil več aplikativnih in raziskovalnih projektov. Svoje raziskovalne rezultate in dosežke iz prakse redno objavlja v domačih in mednarodnih znanstvenih in strokovnih krogih.