

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
lime.priimekl@fri.uni-lj.si

Izvleček

V zadnjih petnajstih letih so postala omrežja izjemno popularna, saj so 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 so npr. pojav malega sveta, odpornost tehnoloških omrežij in obstoj skupnosti v socialnih omrežjih. Poleg tega pa je omogočila razlago različnih dinamičnih procesov, kot npr. gibanje ljudi v družbi, navigacija po svetovnem spletu in 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 le na akademske kroge. V prispevku zato predstavimo glavne 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 the real and the virtual world, both in 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 provided a deeper understanding of different dynamical processes occurring on networks, such as the spreading of diseases and network controllability. Examples of network analysis can thus be found in various fields, and yet the theory of large networks is still mainly used in academic circles. Here we present the 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.

1 UVOD

Omrežja so 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 npr. 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. Eno-stavnost predstavitve poljubnega sistema povezanih komponent pa je sicer eden od ključnih razlogov za izjemno popularnost omrežij v zadnjih letih (Newman, 2010; Newman, 2003).

Poudariti je treba, da sodobna analiza omrežij dejansko temelji na matematični teoriji grafov z začetki

v 18. stoletju. Medtem ko so bila prva dela na tem področju večinoma zabavne ali teoretske narave, pa so že stoletje kasneje raziskovalci nakazali možnosti uporabe v elektrotehniko 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 pomemben del raziskav kompleksnih sistemov v fiziki, predvsem z razmahom računalniške tehnologije ter s 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 z Univerze Stanford, Vladi-

mir Batagelj, Andrej Mrvar, Anuška Ferligoj, Valentina Hlebec, Tina Kogovšek in Gregor Petrič z Univerze v Ljubljani, Matjaž Perc z Univerze v Mariboru, Bosiljka Tadić in Marko Grobelnik z Instituta Jožef Stefan ter še nekateri drugi. Med najvidnejšimi prispevki lahko sicer omenimo svetovno uveljavljeni program za analizo velikih omrežij *Pajek* (Batagelj & Mrvar, 1998; de Nooy, Mrvar & Batagelj, 2012), ki ga avtorji razvijajo že več kot petnajst let.

Kljub navedenemu pa je uporaba teorije velikih omrežij še vedno v veliki meri omejena na akademske kroge in 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 temeljne 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. Sledi sklep v razdelku 5.

2 REALNA OMREŽJA

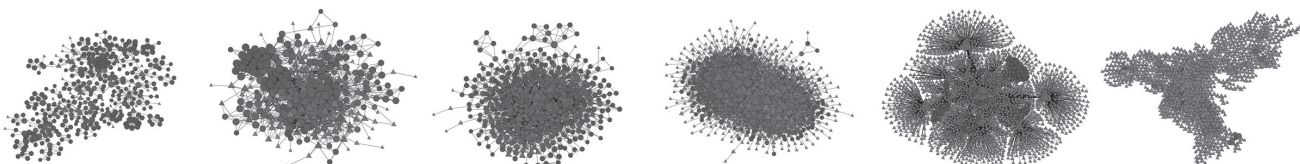
Zgradbo omrežja navadno predstavimo z diskretnim matematičnim objektom, ki mu pravimo graf. Graf $G(V, E)$ je sestavljen iz množice vozlišč $V = \{v_1, v_2, \dots, v_n\}$, $n = |V|$ ter množice povezav med njimi $E \subseteq \{\{v_i, v_j\} \mid v_i, v_j \in V\}$, $m = |E|$. V primeru storitev za socialno mreženje (npr. *facebook*), vozlišča predstavljajo različne uporabnike storitve, povezave pa ustrezajo relaciji prijateljstva. Povezave so tako neusmerjene, medtem ko nekatere vrste omrežij zahtevajo usmerjene povezave. V primeru svetovnega spleta vozlišča predstavljajo množico spletnih strani, ki so med seboj povezane prek usmerjenih spletnih povezav; torej $E \subseteq \{\{v_i, v_j\} \mid v_i, v_j \in V\}$. Definiciji E lahko še razširimo, tako da dopuščata več vzporednih

povezav med vozlišči ter zanke (povezave vozlišča s samim seboj). Poleg omenjenega omrežja vključuje tudi poljubno dodatno znanje o vozliščih in povezavah grafa. Tako so ta lahko različnih tipov, vsebujejo časovne značke in druge podatke.

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, v katerih 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, v katerih povezave v omrežju ustrezajo toku informacij (ali podatkov) skozi analizirani sistem. Sem spadajo 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 spada 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 npr. številna komunikacijska omrežja prav tako tudi socialna. Poleg tega v zadnjih letih vse pogosteje preučujejo omrežja, ki ne spadajo v nobeno od navedenih kategorij (npr. različna transportna ali ekonomska omrežja).



Slika 1: Od leve proti desni: omrežje sodelovanj med znanstveniki (Newman, 2006); prijateljstva v družabnem omrežju *facebook* (Blagus, Šubelj & Bajec, 2012); metabolično omrežje (Jeong, Tombor, Albert, Oltvai & Barabasi, 2000); spletni dnevnik 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).

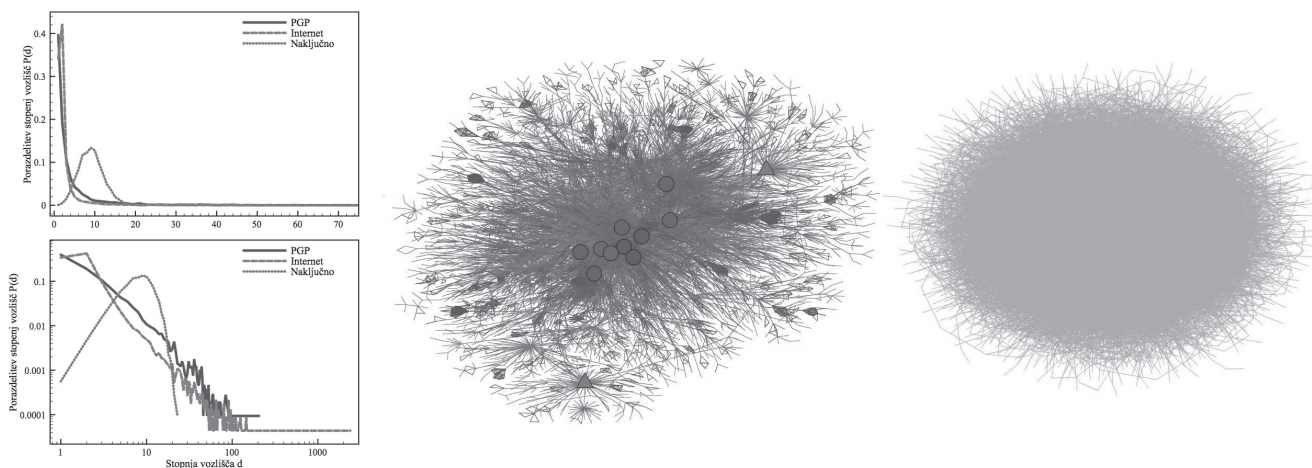
3 LASTNOSTI OMREŽIJ

Kljub svoji raznolikosti 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 \bar{d} povprečna stopnja vozlišča $-\bar{d} = 2m/n$. V primeru socialnih omrežij \bar{d} predstavlja povprečno število prijateljev posamezne osebe. Izkaže se, da je porazdelitev stopenj vozlišč v večini realnih omrežij značilno drugačna, kot npr. v naključnih omrežjih (Erdős & Rényi, 1959). Pri teh predpostavimo, da osebe izbi-

rajo prijatelje popolnoma naključno (verjetnost povezave med poljubnim parom različnih vozlišč je $\bar{d}/(n-1)$). V naključnih omrežjih so stopnje vozlišč porazdeljene tesno okrog povprečja d (podobno kot pri normalni porazdelitvi), medtem ko je porazdelitev stopenj v realnih omrežjih močno razpotegnjena v desno (angl. right skewed). Natančneje, porazdelitev pogosto sledi potenčnemu zakonu $P(d) \propto d^{-\gamma}$, $\gamma > 1$ (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 logaritmskem prikazu, pri čemer je $-\gamma$ 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 ravni 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), mestem ko 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 sliki 2 znaša 206, medtem ko je povprečna stopnja \bar{d} le 4,6. Podobno ima internetno omrežje zgoraj $\bar{d} = 4.2$, 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 to pomeni, da osebe z velikim številom občudovalcev lažje pridobi-

vajo nove občudovalce, kar povzroči obstoj oseb oz. vozlišč z izjemno visoko stopnjo v omrežju.

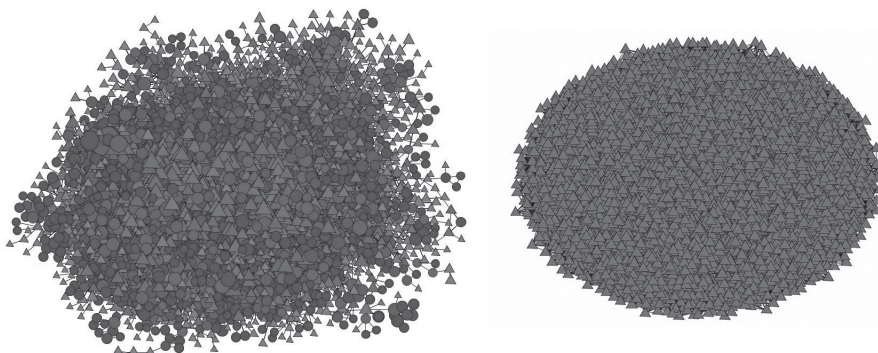
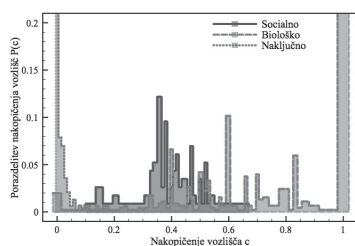
Omenimo še, da imajo vozlišča z visoko stopnjo prav poseben vpliv na različne dinamične procese, ki se odvijajo nad omrežji (npr. širjenje novic po socialnem omrežju ali virusov prek svetovnega spleta). Natančneje, pri $\gamma \in (2,3)$ lahko že manjše število okuženih vozlišč (vozlišča z visoko stopnjo) povzroči razširjenost po vsem omrežju (Pastor-Satorras & Vespignani, 2001; Sinha, 2011). Za družabno omrežje PGP γ znaša 2,2, medtem ko za internetno omrežje zgoraj velja $\gamma = 2,1$. Podobno imajo omenjena vozlišča močan vpliv na odpornost in ranljivost realnih omrežij (gl. spodaj).

Razdalje med vozlišči ter pojav malega sveta

Stanley Milgram je v šestdesetih letih prejšnjega stoletja izvedel eksperiment, v katerem je posameznikom zadal nalogo naj prek zaporedja pisem dosežejo neko izbrano osebo. Večina pisem se je med eksperimentom sicer izgubila, ostala pisma pa so prišla na cilj v presenetljivo majhnem številu korakov. V objavljenih primerih je to 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 prek zelo kratke poti.

Naj bo I_{ij} razdalja med vozliščema i in j , ki je definirana kot število povezav v najkrajši poti med i in j . Pojav malega sveta najpogosteje opazujemo prek povprečne razdalje $I = \frac{2}{n(n-1)} \sum_{ij} I_{ij}$ (Newman, 2003).

Med omrežja malega sveta (Watts & Strogatz, 1998) navadno štejemo tista, pri katerih povprečna razdalja I ne narašča s številom vozlišč n (oziroma narašča zelo počasi, kot npr. $I \propto \log(n)$). Kot primer povejmo, da je povprečna razdalja med 700 milijoni uporabnikov storitve *facebook* leta 2011 znašala le $I = 4,7$ (Backstrom, Boldi, Rosa, Ugander & Vigna, 2012), mestem ko je največja razdalja med več kot milijardo spletnimi stranmi *Yahoo!* manjša od 8 (Kang, Tsourakakis, Appel, Faloutsos & Leskovec, 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 (gl. zgoraj). Poleg skupnosti pa realna omrežja običajno vsebujejo tudi druge značilne skupine vozlišč ali strukturne module (angl.

Podobno velja za mnoga druga velika omrežja, pri katerih je $I \in (4,8)$.

Omeniti je treba, da kratka razdalja med vozlišči ni presenetljiva sama po sebi, saj imajo podobno lastnost tudi naključna omrežja (Erdős & Rényi, 1959). Vendar ta niso tako (lokalno) gosta ali nakopičena kot primerljiva realna omrežja. Natančneje, naj bo C nakopičenost (Watts & Strogatz, 1998; Newman & Park, 2003) (angl. *clustering coefficient*), ki meri tranzitivnost v omrežju. V primeru socialnih omrežij C dejansko ustreza verjetnosti, da je prijatelj prijatelja prav tako prijatelj. Večina realnih omrežjih ima C med 0,3 in 0,6, medtem ko je C v primerljivih naključnih omrežjih blizu nič (slika 3).

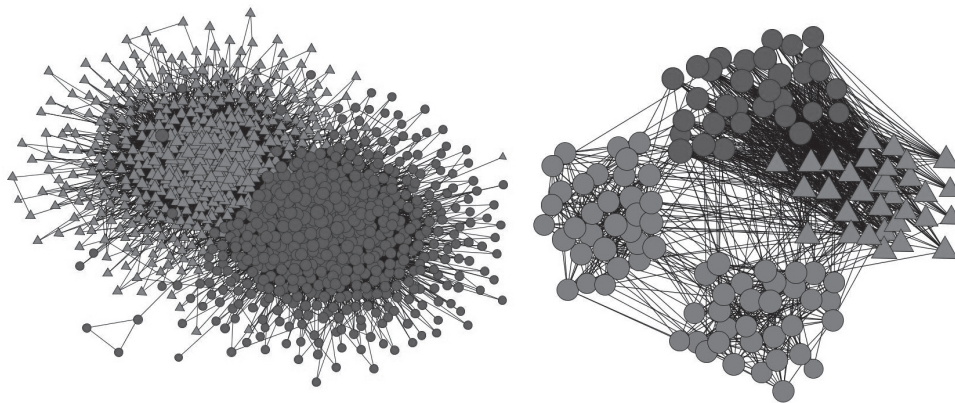
Kopičenje vozlišč ter strukturni moduli v omrežjih

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).

modules). Zadnje raziskave kažejo, da je v mnogih omrežjih mogoče najti šibko povezane množice vozlišč, ki pa so podobno povezane z ostalim omrežjem (Newman & Leicht, 2007; Šubelj & Bajec, 2012a) (slika 4). To 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. beljakovinom, ki opravljajo podobno nalogo v človeškem telesu (Pinkert, Schultz & Reichardt, 2010). Značilne funkcijske module je sicer mogoče najti še v informacijskih, različnih tehnoloških, pa tudi v socialnih omrežjih.



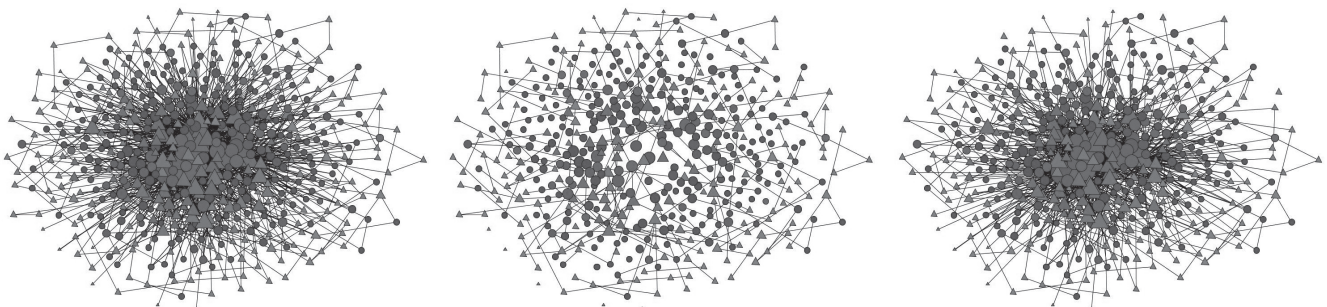
Slika 4: **Levo: Omrežje povezav med spletnimi dnevniki o ameriški politiki (Adamic & Glance, 2005). Omrežje vsebuje dve tesno povezani skupnosti, ki sovpadata z delitvijo na levo in desno usmerjene dnevnike (Adamic & Glance, 2005). Desno: Sintetično omrežje, ki vsebuje dve skupnosti in dva funkcijska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

Odpornost, ranljivost in nadzor realnih omrežij

Kot smo že omenili, 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 prek socialnega omrežja, novic po komunikacijskem omrežju ali 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 $d/(n-1)$, pri čemer je d stopnja obravnavanega voz-

lišča. (Za podrobnejši pregled gl. 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, medtem ko 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 ravni avtonomnih sistemov, zbrano leta 2003 (Leskovec, Lang, Dasgupta & Mahoney, 2009). Na sredini: Omrežje po odstranitvi 10 odstotkov vozlišč z največjo stopnjo. Desno: Omrežje po odstranitvi 10 odstotkov naključno izbranih vozlišč. Oblika vozlišč ustreza gostoti omrežja v neposredni okolici vozlišča (Soffer & Vázquez, 2005).**

Leta 2011 so številni avtorji preučevali nadzor v realnih omrežjih (Liu, Slotine & Barabási, 2011; Vitali, Glattfelder & Battiston, 2011). To v primeru spletnih socialnih storitev pomeni, da je mogoče prek nekaterih vozlišč (uporabnikov storitve) nadzirati mnenje vseh uporabnikov v omrežju (pri določenih predpostavkah). Omenjena vozlišča lahko označimo z izrazom upravljavska vozlišča (angl. driver). Zanimivo, upravljavska vozlišča navadno ne sovpadajo z vozlišči z visoko stopnjo (gl. zgoraj), medtem ko je število upravljavskih vozlišč n_D , potrebnih za nadzor v brezlestvičnih omrežjih, odvisno le od povprečne stopnje \bar{d} in eksponenta γ (Liu, Slotine & Barabási, 2011); natančneje, pri $\gamma > 2$, $n_D/n \approx e^{\bar{d}(\gamma-2)/(2-2\gamma)}$.

Na primer, družabnega omrežja PGP na sliki 2 ni mogoče nadzirati prek manj kot 64 odstotkov uporabnikov spletne storitve PGP. Za primerjavo: n_D/n v nekaterih drugih spletnih socialnih omrežjih znaša okrog 30 odstotkov, medtem ko v bioloških omrežjih (angl. regulatory) znaša 80 odstotkov ter v primeru

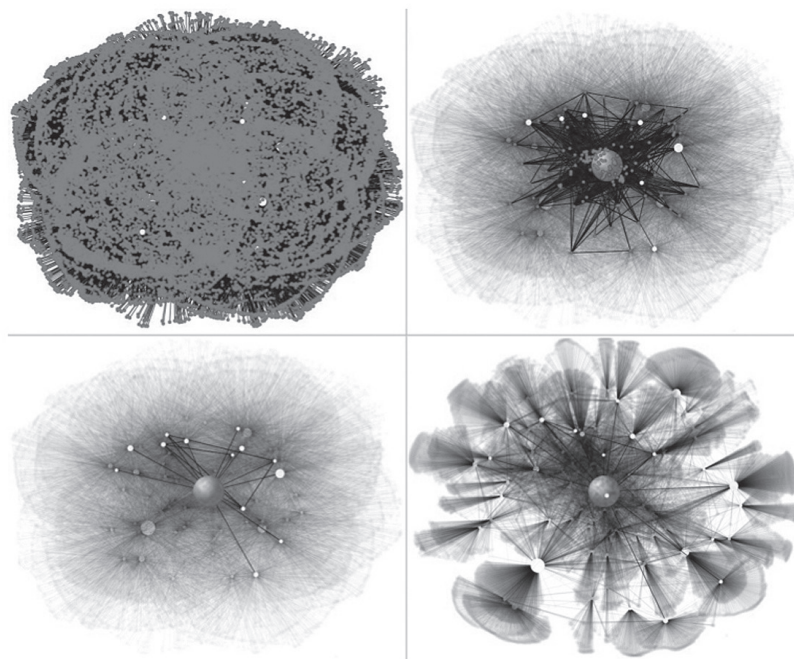
interneta okrog 50 odstotkov (Liu, Slotine & Barabási, 2011). Na drugi strani pa je n_D/n v omrežjih lastništev med podjetji manj kot 3 odstotke (Vitali, Glattfelder & Battiston, 2011). Podobno je programski jezik *java* mogoče nadzirati prek samo 17 odstotkov programskih konstruktorjev, ki sestavljajo jedro jezika (Šubelj & Bajec, 2012c).

4 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 (Vir: Catanese, De Meo, Ferrara & Fiumara, 2010)

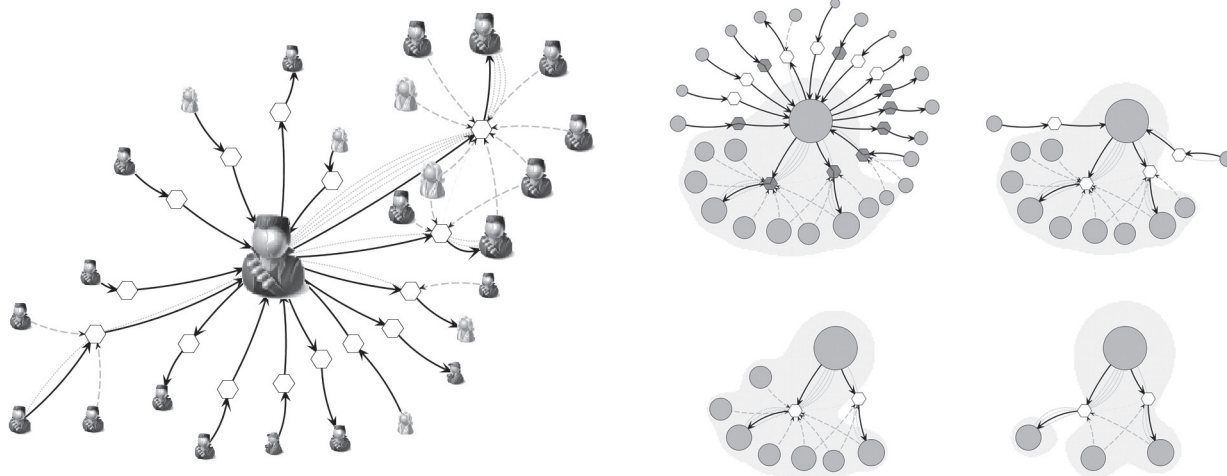
Odkrivanje avtomobilskih goljufij

Analizo omrežij pogosto uporabljajo tudi za odkrivanje anomalij (odstopanj) v podatkih. Na primer z analizo omrežij prometnih nesreč lahko razkrijemo goljufive skupine posameznikov, ki uprizarjajo pro-

metne nesreče ter se tako okoristijo prek 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 svo-

jem brskalniku uporablja *Google*, ter tudi nekaterim meram središčnosti vozlišč (razdelek 3).

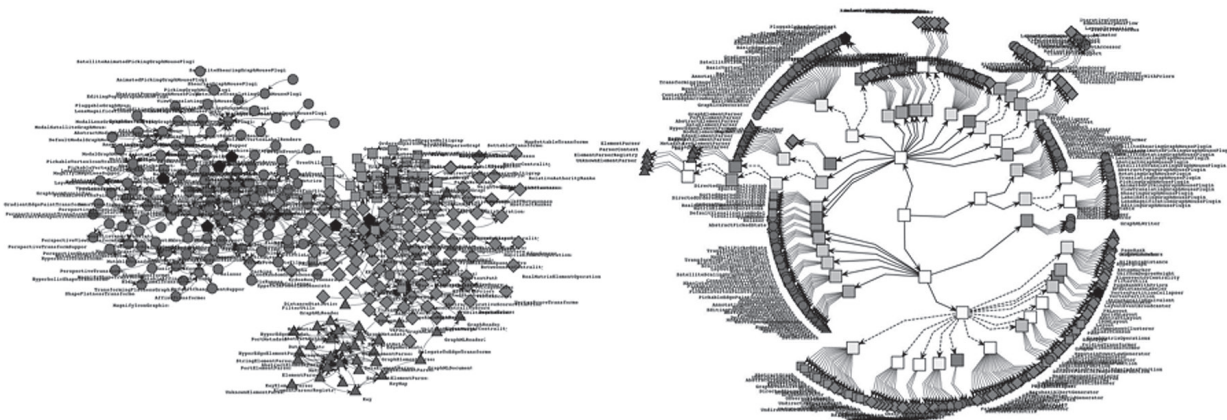


Slika 7: **Levo: Socialno omrežje udeležencev povezanih prometnih nesreč. Desno: Rezultat odkrivanja avtomobilskih goljufov v omrežju na levi. Okrogla (oglata) vozlišča predstavljajo udeležence (prometne nesreče), velikost vozlišč pa je sorazmerna njihovi sumljivosti. (Vir: Šubelj, Furlan & Bajec, 2011)**

Reorganizacija programske opreme

Z uporabo primernih algoritmov lahko v omrežju razkrijemo vso hierarhijo različnih strukturnih modulov (razdelek 3). V primeru omrežij, ki predstavljajo kompleksno programsko kodo, omenjena hierarhija dejansko sovпада s programskimi paketi ter tako razkriva odvisnosti med različnimi deli pro-

gramske opreme (Šubelj & Bajec, 2012b; Šubelj & Bajec, 2012c) (slika 8). To je mogoče 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 pogosto preučevali tudi omrežja citiranj med znanstvenimi prispevki z različnih področij. Izkaže se, da je struktura znanosti podobna obliki črke U (Rosvall & Bergstrom, 2008). Natančne-

je, naravoslovne in družboslovne znanosti so razmeroma dobro ločene med seboj (levo in desno zgoraj), medtem ko so povezane prek interdisciplinarnih znanosti, kot sta medicina in biologija (spodaj).

5 SKLEP

V prispevku so podani temeljni pojmi analize omrežij ter predstavljene različne vrste realnih omrežij. V nadaljevanju smo nato izpostavili skupne lastnosti velikih realnih omrežij, vključujoč nekatera vidnejša odkritja v zadnjih petnajstih letih. Na koncu predstavljamo še izbrane primere uporabe analize omrežij v praksi s poudarkom na socialnih (družbenih) omrežjih.

Ugotavljamo, da je teorija omrežij 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, HALVIN, 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 Zrnc 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 z 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.

Iz Islovarja

Islovar je spletni terminološki slovar informatike, ki je prosto dostopen na naslovu <http://www.islovar.org>. Varovanje podatkov je v informatiki zelo aktualno področje, pojavljajo se novi pojmi in tudi novi izrazi, zato se mu pri urejanju Islovarja posebej posvečamo. V tej številki revije objavljamo zbirko, ki smo jo sestavili na temelju izraza **napad**.

Izraze lahko komentirate tako, da se prijavite v poglavju *Nov uporabnik*, poiščete izraz, ki ga želite komentirati, in zapišete svoj komentar ter predlog spremembe.

APT APT-ja [apətə] m krat. (*angl. advanced persistent threat*)

gl. organizirana trajna grožnja

beléženje tipkanja -a -- s (*angl. keystroke logging, keylogging*)

uporaba beležnika tipkanja za sledenje uporabnikovih aktivnosti; prim. beležnik tipkanja

beléžnik tipkanja -a -- m (*angl. keylogger*)

program za beleženje tipkanja, ki se navadno uporablja zlonamerno, npr. za prisvajanje gesel; prim. zlonamerno programje, beleženje tipkanja, zlonamerna programska oprema

bótnet -a m (*angl. botnet, robot network*)

več v omrežje povezanih računalnikov, nad katerimi napadalec prevzame nadzor, da bi jih uporabil za izvajanje zlonamernih dejanj; sin. omrežje robotskih računalnikov

CSRF CSRF-ja [cəsərəfə] m krat. (*angl. cross site request forgery*)

gl. ponarejanje spletnih zahtev

DDoS DDoS-a [dədəós] m krat. (*angl. distributed denial-of-service*)

gl. porazdeljena ohromitev storitve

internétna prevára -e -e ž (*angl. internet fraud*)

prevara z uporabo internetne storitve

krája gésla -e -- s (*angl. password stealing*)

napad, pri katerem napadalec nepooblaščno pridobi gesla pooblaščenih uporabnikov

krája zasébnega kljúča -e -- -- ž (*angl. private key stealing*)

napad, pri katerem napadalec nepooblaščno pridobi zasebni ključ uporabnika; prim. kraja identitete

maskíranje -a s (*angl. 1. masquerade, masking, 2. masking*)

1. napad, pri katerem se napadalec z identiteto pooblaščenega uporabnika nepooblaščno prijavi v sistem; prim. kraja identitete
2. uporaba nadomestnega znaka pri poizvedovanju

MitB MitB-ja [mitəbə] m krat. (*angl. man in the browser*)

gl. prestrezanje v brskalniku

napàd -áda m (*angl. attack*)

zlonamerno dejanje, ki škodljivo vpliva na delovanje informacijskega sistema

napadálec -lca m (*angl. perpetrator*)

kdor zagreši zlonamerno, kriminalno dejanje v kibernetnem prostoru; prim. vdiralec, kreker

obrámni gostítelj -ega -a m (*angl. bastion host*)

strežnik, ki je dostopna točka in je zasnovan tako, da vzdrži napade

omréžje robótskih račúnálnikov -a -- -- s (*angl. botnet, robot network*)

več v omrežje povezanih računalnikov, nad katerimi napadalec prevzame nadzor, da bi jih uporabil za izvajanje zlonamernih dejanj; sin. botnet

organizírana trájna gróžnja -e -e -e ž (*angl. advanced persistent threat, krat. APT*)

trajno ogrožanje varnosti informacijskega sistema, ki ga izvaja organizirana skupina, organizacija

podtáknjena sêja -e -e ž (*angl. session fixation*)

napad z namenom pridobitve uporabnikovih pravic, pri katerem napadalec zvabi uporabnika, da se prijavi

Izbor pripravlja in ureja Katarina Puc s sodelavci Islovarja

Koledar prireditev

The 4 th Annual European Data Protection and Privacy Conference	17. september 2013	Bruselj, Belgija	www.dataprotection2013.eu
The 12 th International Symposium on Operational Research in Slovenia, SOR'13	25.-27. september 2013	Dolenjske Toplice, Slovenija	http://sor13.fis.unm.si
The 2013 M2M & Internet of Things Global Summit – Smart technologies for an interconnected and intelligent world	1.-2. oktober 2013	Washington, ZDA	www.iotsummit2013.com
8. mednarodna poslovna konferenca Management poslovnih procesov	16.-17. oktober 2013	Ljubljana, Slovenija	www.process-conference.org
The 3 rd Annual Americas Spectrum Management Conference	6.-7. november 2013	Washington, ZDA	

Pomembni spletni naslovi

- IFIP News: <http://www.ifip.org/images/stories/ifip/public/Newsletter/news> ali www.ifip.org → Newsletter
- IT Star Newsletter: www.itstar.eu
- ECDL: www.ecdl.com
- CEPIS: www.cepis.com

Dostop do dveh tujih strokovnih revij

- Revija **Upgrade** (CEPIS) v angleščini (ISSN 1684-5285) je dostopna na spletnem naslovu: <http://www.upgrade-cepis.org/issues/2008/4/upgrade-vol-IX-4.html>.
- Revija **Novática** (CEPIS) v španščini (ISSN 0211-2124) je dostopna na spletnem naslovu: <http://www.ati.es/novatica/>.