



NetSlo '18

Univerza v Ljubljani

Fakulteta za družbene vede

**Srečanje raziskovalcev s področja
analize omrežij**

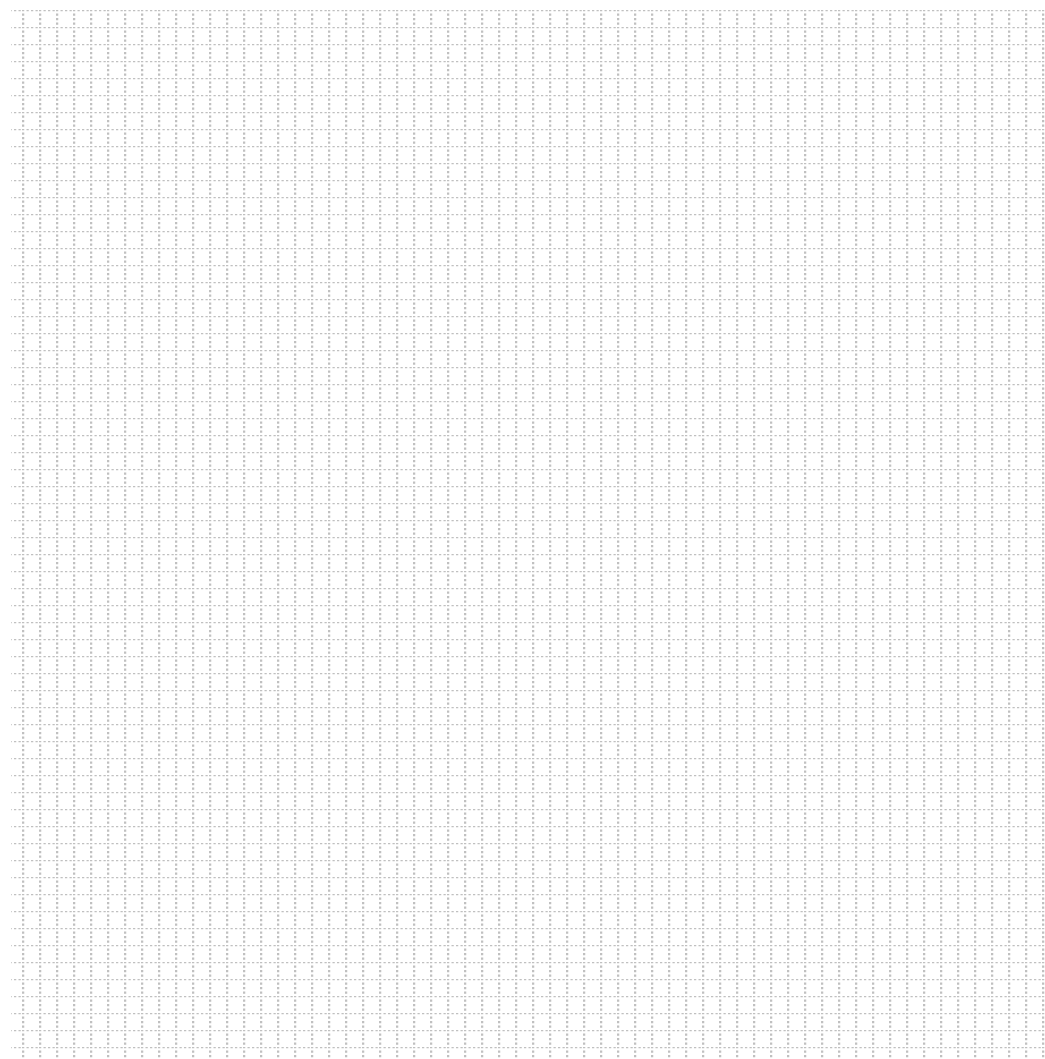
Ljubljana, 25. januar 2018

Organizatorji

Marjan Cugmas
Zoran Levnajić
Ljupčo Todorovski
Lovro Šubelj

Luka Kronegger
Anuška Ferligoj
Vlado Batagelj

Zapiski



Marjan Cugmas, mag. upor. stat.

Fakulteta za družbene vede, Univerza v Ljubljani

Generiranje globalnih struktur omrežij na podlagi različnih tipov triad

Predstavitel naslavlja vprašanje zmožnosti generiranja omrežij s točno določeno globalno strukturo zgolj z upoštevanjem lokalnih podstruktur. Različne globalne strukture omrežij so operacionalizirane z različnimi tipi bločnih modelov. Bločni model je omrežje, kjer so enote skupine enakovrednih enot opazovane ga omrežja. Pri tem se omejimo na strukturno enakovrednost (dve enoti sta strukturno enakovredni, če sta enako povezani z ostalimi enotami). V analizi preučujemo najbolj znane tipe bločnih modelov: koheziven, center-periferija, tranzitiven in hierarhičen tip bločnega modela. Lokalne strukture omrežij predstavljajo različni tipi triad. Za generiranje omrežij sta uporabljena algoritem predstavljanja povezav in Monte Carlo Multi Chain algoritem, implementiran v programski paket »ergm« programskega jezika R. Rezultati kažejo, da je, zgolj z upoštevanjem različnih tipov triad, mogoče generirati omrežja različnih globalnih struktur, a vključitev dodatnih lokalnih struktur (na primer poti dolžine tri) lahko znatno pripomore k nastanku želene globalne strukture. Pri generiranju tovrstnih omrežij ne upoštevamo lastnosti enot.

PROGRAM

13.00 - 13.10	doc. dr. Luka Kronegger	Uvodni nagovor
13.10 - 13.55	prof. dr. Vladimir Batagelj	Vabljen predavatelj: Polkolobarji in analiza omrežij
13.55 - 14.15	dr. Jan Kralj	NetSDM - analiza omrežij v semantičnem rudarjenju podatkov
14.15 - 14.35	asist. Tilen Marc	Vložitve časovnih omrežij v vektorski prostor za odkrivanje vandalov in goljufov
14.35 - 15.35	Predstavitel plakatov in čas za kavo	
15.35 - 16.20	doc. dr. Vinko Zlatič	Vabljen predavatelj: Unexpected properties of color avoiding percolation
16.20 - 16.40	Tadej Matek	Uporaba odvisnosti višjih redov za zaznavanje karakteristik prometa v računalniških omrežjih
16.40 - 16.55	Čas za kavo	
16.55 - 17.40	prof. dr. Tina Kogovšek prof. dr. Valentina Hlebec	Vabljen predavatelj: Merjenje osebnih omrežij z anketami
17.40 - 18.00	asist. Marjan Cugmas	Generiranje globalnih struktur omrežij na podlagi različnih tipov triad
18.00 - 18.45	Filip Dobranić	Danes je nov dan: Predstavitel podatkovnega izziva NetSlo '19
18.45	Pogostitev v prostorih FDV	
20.00	Večerja v gostilni Čad (Cesta na Rožnik 18, 1000 Ljubljana)	



Polkolobarji in analiza omrežij

V omrežjih so povezave pogosto utežene. Pri analizi sprehodov po omrežju/jih pride zelo prav operacija množenja omrežij. Polkolobar je algebrska struktura, ki omogoča definicijo produkta (usklajenih) omrežij.

Spoznali bomo nekaj, za analizo omrežij pomembnih, polkolobarjev in nekaj uporabnih izrekov o izpeljanih omrežjih oziroma njihovih matrikah.

Viri:

Batagelj, V: Semirings for social networks analysis. J Math Sociol 19 (1): 53-68 1994

Batagelj, V., Praprotnik, S.: An algebraic approach to temporal network analysis based on temporal quantities. Social Network Analysis and Mining, 6(2016) 1, 1-22. arXiv:1505.01569

Batagelj, V, Cerinšek, M: On bibliographic networks. Scientometrics 96(2013) 3, 845-864. arXiv:1301.4655

Cerinšek, M, Batagelj, V: Semirings and Matrix Analysis of Networks. in Alhajj, R, Rokne, J (Eds.) Encyclopedia of Social Network Analysis and Mining. Springer, October 2017, 1-8.

Merjenje osebnih omrežij z anketami

Podobno kot na drugih področjih v družboslovju, se tudi pri raziskovanju socialnih omrežij za merjenje najpogosteje uporablja anketa. Merjenje omrežij je za razliko od drugih objektov merjenja, kot so na primer stališča, mnenja ali vrednote kompleksnejše in zato v nekaterih vidikih še zahtevnejše. Merjenje pogosto poteka v dveh osnovnih fazah, saj običajno najprej izmerimo (generiramo) enote omrežja, nato pa še relacije med njimi in druge značilnosti enot (npr. demografske značilnosti ipd.). Posebnosti so tudi glede na to, ali merimo popolna ali osebna (egocentrična) omrežja. V tem prispevku se bomo omejili na merjenje osebnih omrežij in sicer predvsem na primerjavo različnih metod generiranja omrežij. Poznamo štiri osnovne načine generiranja omrežij: generator imen, generator vlog, generator omrežja z dogodki ter pozicijski generator. Vsak od teh načinov ima glede na različne raziskovalne cilje ter stroške (finančne, časovne, obremenitev anketirancev) svoje prednosti in slabosti. Poleg tega je treba zaradi kompleksnosti in specifik generiranja omrežij upoštevati tudi značilnosti uporabljenih metod merjenja (npr. osebna, telefonska ali spletna anketa). V tem kontekstu bomo predstavili prednosti in slabosti različnih načinov generiranja osebnih omrežij in jih ilustrirali s tipičnimi, pogosto uporabljenimi primeri.



Uporaba odvisnosti višjih redov za zaznavanje karakteristik prometa v računalniških omrežjih

Pri analizi omrežij se, poleg strukturne analize, že dolgo časa preučuje dinamiko, ki zajema razvoj omrežja in spreminjanje interakcij skozi čas. Pred kratkim so se raziskave usmerile v preučevanje t.i. odvisnosti višjih redov, ki omogočajo natančnejše modeliranje prehodnih verjetnosti naključnega sprehajalca. V raziskavi se posvetimo učinkovitemu odkrivanju vzorcev v sekvenčnih podatkih, pri čemer je velikost vzorcev lahko večja od običajnih pod-poti dolžine ena. Za modeliranje podatkov uporabimo Markovsko verigo z variabilno velikostjo spomina. Pomembnejše vzorce višjega reda zaznamo s statističnimi testi. Metode preizkusimo nad podatki zaporedne uporabe računalniških komunikacijskih protokolov. Rezultati kažejo, da se v podatkih nahajajo odvisnosti višjega reda, tako takšne ki bi jih pričakovali kot tudi zanimivi vzorci.

NetSDM - analiza omrežij v semantičnem rudarjenju podatkov

Področje Semantičnega rudarjenja podatkov (Semantic Data Mining - SDM) se ukvarja z analizo podatkov, v katerih so primeri anotirani z ontološkimi termi. Za razliko od standardnih metod rudarjenja podatkov z metodami SDM lahko v podatkih najdemo kompleksne vzorce (v obliki pravil, ki opisujejo zanimive dele vhodnih podatkov), ki imajo za končnega uporabnika veliko uporabno vrednost. Težava trenutnih metod je, da končnemu uporabniku poleg uporabnih rezultatov izstavijo tudi visok račun - iskanje je računsko zahtevno in dolgotrajno. Ta račun lahko občutno zmanjšamo tako, da na vhodne podatke pogledamo iz stališča analize omrežij. Ontološki termi ter relacije med njimi tvorijo omrežje, v katerem metode SDM iščejo "zanimiva" vozlišča ki tvorijo kakovostna pravila. Za razliko od teh metod v analizi omrežij obstajajo metode (na primer dobro znani PageRank), s katerimi lahko zelo hitro ocenimo pomembnost vozlišča v danem omrežju. Z metodo NetSDM smo prvič uporabili metode analize omrežij, da smo hitro in učinkovito poiskali zanimiva vozlišča in odstranili nezanimiva vozlišča - Iskanje kvalitetnih pravil smo tako uspeli močno pohitriti brez izgube kakovosti.



Vložitve časovnih omrežij v vektorski prostor za odkrivanje vandalov in goljufov

Časovna omrežja predstavljajo enega boljših modelov za zajemanje časovno pogojenih interakcij v družbenih omrežjih. Eden izmed pogostih pojavov v takih omrežjih je obstoj vandalov ali goljufov, ki se s pomočjo interakcij želijo okoristiti na škodo drugih. Preprost način odkrivanja takega početja je identifikacija različnih lastnosti vozlišč iz časovnih in omrežnih podatkov, ki razlikujejo običajna vozlišča od zlonamernih. Te lastnosti se nato uporabijo kot vhodni podatki modela strojnega učenja za klasifikacijo vozlišč. Vendar je identifikacija dobrih lastnost težak in časovno zahteven proces, vezan na dane podatke. Predstavili bomo nov model za napovedovanje oznak vozlišč v časovnih omrežjih, ki se s pomočjo znanih podatkov nauči vložiti vozlišča omrežja v vektorski prostor tako, da optimizira klasifikacijo vozlišč. Model se izkaže kot uspešen za odkrivanje vandalov na družbenem omrežju pridobljenem iz podatkov Wikipedije in drugih podobnih nalogah.

Unexpected properties of color avoiding percolation

Many real world networks have groups of similar nodes which are vulnerable to the same failure or adversary. Nodes can be colored in such a way that colors encode the shared vulnerabilities. Using multiple paths to avoid these vulnerabilities can greatly improve network robustness. Color-avoiding percolation provides a theoretical framework for analyzing this scenario, focusing on the maximal set of nodes which can be connected via multiple color-avoiding paths. We explicitly account for the fact that the same particular link can be part of different paths avoiding different colors. This fact was previously accounted for with a heuristic approximation. We compare this approximation with a new, more exact theory and show that the new theory is substantially more accurate for many avoided colors. Further, we formulate our new theory with differentiated node functions, as senders/receivers or as transmitters. In both functions, nodes can be explicitly trusted or avoided. With only one avoided color we obtain standard percolation. With one by one avoiding additional colors, we can understand the critical behavior of color avoiding percolation. For heterogeneous color frequencies, we find that the colors with the largest frequencies control the critical threshold and exponent. Colors of small frequencies have only a minor influence on color avoiding connectivity, thus allowing for approximations. We show that modularity of network provides an interesting analogue in the study of networks to the Ising model in the presence of a magnetic field. Lastly, we find from the perspective of statistical physics, that found exponents define new universality classes characterizing higher-order transitions with corresponding higher-order derivative analogues of γ . To the best of our knowledge the present study represents the first time that higher-order transitions are observed in percolation type systems with their higher-order scaling exponents defined and measured.

