

Andrej Kastrin

NetSlo '23: poročilo s VII. srečanja raziskovalcev s področja analize omrežij

NetSlo '23: Report from the VII. Meeting of the Slovenian Network Analysis Researchers

O srečanju

Po dvoletnem premoru je 26. januarja na Medicinski fakulteti Univerze v Ljubljani potekalo tradicionalno, letos že sedmo, Srečanje raziskovalcev s področja analize omrežij. Analiza omrežij – pogosto označena kot znanost 21. stoletja – je vsebinsko razvejano in izrazito interdisciplinarno področje raziskovanja. V preteklem stoletju je analiza omrežij botrovala nekaterim paradigmatiskim miselnim preskokom v proučevanju kompleksnih socialnih sistemov. Kasneje je med drugim omogočila razumevanje zgradbe in mehanizmov delovanja svetovnega spleta. Danes analiza omrežij utira nova spoznanja v vedah o živem in pospešuje njihov razvoj.

Prvo srečanje, takrat še pod imenom Mreženje slovenskih networkkašev, se je odvijalo leta 2015 na ljubljanski Fakulteti za računalništvo in informatiko. Kasneje se je srečanje preimenovalo v NetSlo, po zgledu mednarodnih konferenc NetSci. Letošnji dogodek je potekal v organizaciji Inštituta za biostatistiko in medicinsko informatiko na ljubljanski Medicinski fakulteti z materialno in finančno pomočjo Fakultete za družbene vede ter novomeške Fakultete za informacijske študije. Srečanja se je udeležilo več kot 70 udeležencev, nastopilo pa je sedem predavateljev iz Slovenije in tujine. Pester program se je nadaljeval z neformalnim druženjem in zaključil s tradicionalno večerjo pod Rožnikom.

Z željo, da NetSlo približamo širši strokovni javnosti, v nadaljevanju predstavljamo povzetke prispevkov, več informacij pa je na voljo na spletni strani NetSlo '23 (<https://netslo.mf.uni-lj.si/>).

Povzetki

Sistemska medicina in bolezenska omrežja (Damjana Rozman)

Pojav visoko zmogljivih tehnologij ustvarja velike količine omskih podatkov, kar omogoča preučevanje organizmov na molekularni in celični ravni. Kljub spektakularnemu napredku v pogenomskem obdobju ostaja vrzel med podatki in medicinskim znanjem ter še večja vrzel med novim znanjem v smislu klinične uporabnosti in koristi za bolnike. Še vedno trpimo zaradi večfaktorskih bolezni, kjer ne poznamo vzrokov ali zdravil. Pogosto tudi ne vemo, kako motnje na molekularni ravni vplivajo na bolezenske fenotipe. Fenotipske značilnosti organizma so namreč določene s prepletenimi reakcijami presnovnih in signalnih poti, kar lahko opišemo kot biološka omrežja oziroma sisteme. Za višje razvite organizme, kot je človek, je izjemno težko preučiti vse njihove procese *in vivo* ali *in vitro*. Zato so se razvila številna računska orodja, kot so npr. modeli na ravni genoma, ki nam olajšajo vpogled v delovanje kompleksnih bioloških omrežij. Razumevanje, kako so posamezni geni in drugi molekularni dejavniki vključeni v patogenezo, je temeljni del posamezniku prilagojene medicine. Pri tem je nujno uporabiti celostne, sistemske pristope, saj geni in njihovi produkti ne delujejo vsak zase, temveč usklajeno. Govorimo lahko o pristopih sistemske medicine oziroma medicine omrežij, ki predstavlja novo trans-disciplinarno področje. Človeško telo obravnava kot sistem, sestavljen iz več delov, s kompleksnimi odnosi na več ravneh, ki je integrirano tudi v okolje. Raziskovanje s pristopi sistemske medicine pomeni razumevanje in združevanje konceptov, ki prihajajo iz diametralno različnih področij, vključno z medicino, biokemijo, biologijo, fiziko, statistiko, modeliranjem in simulacijo ter znanostjo o podatkih, vse več pa je tudi etičnih in pravnih dilem. Pri delu s človeškimi vzorci ali samo s

človeškimi podatki je treba vedno uveljavljati etična pravila in spoštovati zasebnost. V okviru predavanja bo predstavljenih nekaj primerov študija omrežij pri boleznih jeter, ki se lahko razlikujejo tudi med spoloma. Predstavljeni bodo pristopi, ki preko poznavanja bioloških omrežij lahko privedejo do opredelitve novih bolezenskih biooznačevalcev ali možnih tarč za zdravljenje.

Analiza omrežij časovnega razvoja seizmične aktivnosti (Romi Koželj)

Potres je naravni pojav, ki nastane kot posledica notranje Zemljine dinamike globoko pod površjem in ga z današnjim znanjem še ni mogoče napovedati. V raziskavi poizkušamo z uporabo omrežne znanosti pridobiti nova znanja o značilnostih in razvoju seizmične aktivnosti skozi čas. Implementiramo in med seboj primerjamo različne modele omrežij, ki temeljijo na interakciji med potresi v času in kraju ter na predpostavki o podobnostih potresne aktivnosti na izbranih geografskih območjih. Iz omrežij, ki jih konstruiramo v več manjših zaporednih časovnih oknih, izračunamo nabor značilk ter prikažemo njihovo spreminjanje skozi čas. Na koncu z uporabo modela ARIMA za napovedovanje časovnih vrst preverimo, ali je iz dobljenih vzorcev moč sklepati o značilnostih seizmičnega dogajanja v prihodnosti. Analiza dobljenih omrežij ter generiranih časovnih vrst pokaže, da preko večine obravnavanih omrežij dobimo zanesljiv in predvidljiv odziv vrednosti v časovnih vrstah. Iz rezultatov napovedovanja vrednosti časovnih vrst je razvidno, da oblike časovnih vrst, predvsem v intervalih, v katerih se zgodi močnejši potres, z uporabljenim modelom ARIMA ne moremo dobro napovedati.

Publicistična raziskovalna uspešnost slovenskih doktorjev znanosti (Katarina Rojko)

Prvi del predstavitve bo vključeval izsledke raziskave publicistične raziskovalne uspešnosti slovenskih doktorjev znanosti s poudarkom na vpliv Bolonjske reforme, drugi pa izsledke raziskave te uspešnosti s poudarkom na posamezne znanstvene discipline. Vzorec zajema vse, ki so od leta 2007 oz. 1991 doktorirali, a za vsakega le obdobje dve leti pred in dve oz. sedem let po letu doktoriranja. Njihova publicistična uspešnost je analizirana upoštevajoč produktivnost, sodelovanje in neodvisnost od mentorja. Rezultati prvega dela raziskave kažejo, da razlike med predbolonjskimi in bolonjskimi doktoranti niso vidne pri polovici indikatorjev, je pa vseeno opazen negativen trend. Drugi del raziskave, ki je med indikatorji iskal tudi korelacije, vključil podatke o citatih, zajel daljše časovno obdobje (28

generacij ter prvih 10 let kariere) in raziskovalce razdelil v 234 disciplin, pa je razkril v večini pričakovane razlike med disciplinami po vseh kazalnikih. Identificirani so bili tudi večinoma podobni trendi—zmanjševanje produktivnosti ter neodvisnosti, a vedno več sodelovanja in citatov.

Analiza slovenskega koščka Twitterja (Marko Plahuta)

Raziskovanje družabnega omrežja Twitter je zanimivo tako z vidika besedilnega rudarjenja po vsebinah objav kot tudi z vidika analize strukturnih lastnosti omrežja samega in proučevanja njegovega razvoja v času. V prispevku avtor obravnava podatkovne tehnologije, ki jih uporablja za zbiranje, shranjevanje, analizo in vizualizacijo omrežja Twitter. Omrežje trenutno sestavlja okrog 28.000 uporabniških imen in več kot 90 milijonov tвитov. V nadaljevanju avtor predstavi metapodatkovno shemo, ki jo uporablja za gradnjo omrežja in opiše algoritem za ugotavljanje spremembe uporabniškega imena uporabnika. Posebej obravnava dileme, na katere mora podatkovni analitik odgovoriti ob zajemu omrežja, npr. kolikšna je vrednost retvita v primerjavi z neposrednim odgovorom ali pa kako vrednotimo pripisan ključnik v odnosu do dodanega URL naslova. V tem okviru predstavi tudi shemo, ki jo uporablja za uteževanje povezav v svoji zbirki. Drugi del predstavitve avtor začne s kratkim uvodov v dve tehniki za vlaganje besed (word2vec) oz. vozlišč omrežja (node2vec) v vektorski prostor. Na izbranih primerih iz slovenske tвитosfere prednosti uporabe vektorskih vložitev tudi nazorno ilustrira; predstavi primere gručenja po uporabniških imenih, domenah, ter politični in tematski sorodnosti tвитov.

Struktura slovenske politične razprave na Twitterju (Luka Jesenšek)

Z metodami za analizo socialnih omrežij preučujem strukturo slovenske politične razprave na spletnem družbenem omrežju Twitter pred volitvami v Evropski parlament 2019. Opisujem omrežje 2867 uporabnikov in 40.670 objav (tвитov), središčne uporabnike, pretok informacij prek posredovanih objav (retvitov), skupnosti uporabnikov in širjenje medijskih vsebin na platformi. Med najaktivnejšimi uporabniki prevladujejo širši javnosti manj znani posamezniki, med najvplivnejšimi pa politične stranke, politiki in znane osebnosti. Polarizacija močno zaznamuje pretok informacij v omrežju: analiza retvitov, vzorci komuniciranja in analiza deljenih vsebin kažejo na obstoj jasno določenih, mnenjsko homogenih skupnosti. Uporabniki Twitterja pogosteje komunicirajo z osebami, s katerimi si delijo podobna prepričanja, pogosteje

delijo vsebine, ki prihajajo iz njihovih skupnosti ter pogosteje vključujejo povezave do vsebin, ki se skladajo z njihovimi pogledi. Medijske vsebine se pretežno razširjajo znotraj skupnosti uporabnikov in ne preko celotnega omrežja, znotraj različnih skupnosti pa se razširjajo medijske vsebine iz različnih virov. Politiki imajo velik vpliv na širjenje vsebin znotraj svojih skupnosti, a niso med središčnimi uporabniki, ki bi spodbujali komunikacijo med različnimi deli omrežja.

Stress knowledge map: from knowledge graph to Boolean modelling (Carissa Bleker)

With pressure on global food security set to increase due to a growing human population and the increasingly apparent effect of climate change on agriculture, our understanding of the complexity of plant responses to biotic and abiotic stressors is becoming ever more important. Knowledge on molecular processes occurring within the plant cell is currently scattered across various sources, and thus not easily accessible for analysis or modelling. Stress Knowledge Map (SKM, <https://skm.nib.si/>) is an attempt at integrating this dispersed information into a freely available resource. The types of entities (nodes) within SKM include genes and gene products, complexes, and metabolites, connected to one another via molecular interactions. Implemented in a Neo4j database, and translated to a number of standard formats, it supports the interactive exploration of its contents, and represents a basis for various modelling approaches. While the topology of Stress Knowledge Map is well described, many detailed kinetics are lacking, making classical dynamical modelling impractical. Instead, Boolean networks and Boolean modelling allows for the analysis of complex systems without the need for detailed mechanistic knowledge. In a Boolean network, nodes are assigned one of two states: »active« or »inactive«, defined from a set of regulator nodes. Realistically, the nodes have a degree of activity, and are not merely »active« or »inactive«. To model this, the functions in Boolean space can be translated to qualitative ODEs with the use of various transforms. We developed a Python package

(BoolDog) that can read in regulatory networks, as well as Boolean networks; run Boolean simulations and steady state analysis; transform the Boolean network to continuous ODEs; and run continuous simulations. The package is designed to be interoperable and easily extendable.

From causal networks to adverse outcome pathways (Vid Modic)

A decade ago, adverse outcome pathways (AOPs) and networks of AOPs have been put forward as a tool for organizing toxicological knowledge across different levels of biological organization, from the initial interaction of chemicals with the biological system (molecular initiating event) to the individual and population level effects relevant for environmental risk assessment (adverse outcome). There are currently only around 500 AOPs in the AOP-Wiki, which don't even closely cover all the different ways in which chemicals can adversely affect biological systems. Today we will present a semiautomatic approach for development of AOPs from causal toxicological networks (CTNs), which are available on the CausalBioNet database. The approach, demonstrated on a developmental neurotoxicity networks, includes additional curation of the CTN, removal of disconnected part of the networks and feedback loops, and separations into simple paths (candidate AOPs). In the second part we will show how to perform a weight of evidence analysis of an example pathway, which is necessary before such pathways are used in regulation of chemical safety.

Zahvala

Zapisal Andrej Kastrin (Univerza v Ljubljani), član organizacijskega odbora, zahvaljujoč in v imenu vseh ostalih članov: Gorazd Drevenšek (Univerza v Ljubljani, Univerza na Primorskem), Nataša Kejžar (Univerza v Ljubljani), Petra Kralj Novak (Institut Jožef Stefan, Central European University) in Zoran Levnjajić (Fakulteta za informacijske študije, Institut Jožef Stefan).

■ **Infor Med Slov** 2023; 28(1-2): 44-46