

Javni razpis za sofinanciranje raziskovalnih projektov za leto 2017

20. KRATKA PREDSTAVITEV RAZISKOVALNEGA PROJEKTA

1. Znanstvena izhodišča

V zadnjih dveh desetletjih so je na različnih področjih povečalo zanimanje za časovna omrežja. To je ustvarilo potrebo po razvoju metod za delo s časovnimi omrežji. Posebej pomembna je podpora za potovalne storitve in analiza zaporedij povezanih dogodkov, kakor so sporočila elektronske pošte, novice in telefonski klici. Ti pristopi in rezultati so povzeti v pregledih Holme in Saramäki (2012,2013) in Holme (2015). Poglejte še Nicosia (2013) in Kempe, Kleinberg, inKumar (2000). Še en pregled je mogoče najti v Casteigts, Flocchini, Quattrocioni, in Santoro (2012) in Casteigts in Flocchini (2013) ki uvaja njihovo formalizacijo časovnih omrežij kot časovno.spremenljivih grafov (time-varying graphs - TVGs).

V tem projektu bomo ubrali nekoliko drugačno pot. Oprli se bomo na naše rešitve in izkušnje s časovnimi omrežji v programu Pajek (Pajek 0.47, julij 1999) in jih razširili v splošnejši okvir za analizo časovnih omrežij.

Zamisel omrežjem dodati časovno razsežnost ni nova in je bila uporabljena na več področjih. Med zgodnejše poskuse spadajo transportna omrežja Bell in Lida (1997) in Correa in Stier-Moses (2011); upravljanje projektov (Critical Path Method (CPM), Pert) v operacijskih raziskavah Moder in Phillips (1970) in omrežja z omejitvami v umetni inteligenci Dechter (2003).

Obstajajo tudi kvalitativni pristopi k analizi časovnih omrežij, npr. Allen (1983) in Vilain, Kautz, in van Beek (1990). V tem projektu se bomo omejili na kvantitativne pristope. Več o statističnih pristopih najdete v Kolaczyk (2009) in na straneh programa Siena (Snijders in Ripley, Snijders, Boda, Vörös, in Preciado 2013).

Iz našega pregleda literature izhaja, da obstajata dva pogleda na časovna omrežja:

- Omrežje postavlja omejitve na dejavnosti vozlišč ali je z njimi določeno (npr. omrežje povezanosti letališč, kot je določeno z urniki letov); in
- Omrežje predstavlja dogodke, ki vključujejo pare vozlišč (npr. omrežja KEDS ali omrežja sklicevanj).

Koristno je razmišljati o procesih na omrežjih v smislu 'kaj se premika?' po povezavah. Na primer:

- Potovanja (premiki ljudi in stvari med mesti; mestni promet, pošta);
- Razširjanje (zamislj, slogi, informacij, tehnologije); in
- Tokovi (npr. vodovodno omrežje)

To razmišljanje prepletamo še nekaj novimi pojmi.

Večino raziskav uporablja neuskkljeno izrazje – običajno glede na področje uporabe – komunikacijska omrežja, transportna omrežja, računalniška omrežja, itd. Pogosto za iste ali podobne pojme uporabljajo različna imena in oznake. Na primer, časovna razdalja iz Xuan, Ferreira, and Jarry (2003), čas dosegljivosti iz Holme (2005), ali latenca informacije, itd. poimenujejo isti pojem na različnih področjih. Podobno je s pojmom potovanja (journey) iz Xuan et al. (2003), ki mu drugi avtorji pravijo časovna pot, pot, ki upošteva čas, rapored itd. Ne obstaja splošno sprejemljiv formalni opis časovnih omrežij.

Običajno izhodišče je prisotnost časovne razsežnosti in, da so spremembe v omrežju ena od ključnih informacij. Pogosto pristopi k analizi časovnih omrežij temeljijo na prirejenem zaporedju časovnih

Javni razpis za sofinanciranje raziskovalnih projektov za leto 2017

rezin Santoro, Quattrociochi, Flocchini, and Casteigts (2011). Časovno omrežje je predstavljeno kot zaporedje statičnih omrežij, ki predstavljajo stanje časovnega omrežja v izbranem časovnem trenutku ali intervalu. Predlagana sta dva pristopa, ki poskušata stvari poenotiti. Prvi je časovno-nakopičeni graf iz George in Kim (2013); drugi pa časovno-spremenljivi graf iz Casteigts in dr. (2012). Oba opisa omogočata na razmeroma zapleten način opisati prisotnost vozlišč in povezav. Mi pri obeh opisih pogrešamo možnost pripisati dodatne spremenljive podatke vozliščem in povezavam omrežja. V našem članku Batagelj in Praprotnik (2016) smo predlagali nov, splošnejši opis časovnih omrežij, ki odpravlja omenjene pomanjkljivosti.

2. Predstavitev problema

V časovnem omrežju se lahko prisotnost in dejavnost posameznih vozlišč in povezav spreminja skozi čas. V članku Batagelj in Praprotnik (2016) smo vpeljali pojem časovne količine, s katerimi opišemo obnašanje vozlišč in povezav v časovnem omrežju. Za časovne količine vpeljemo opraciji seštevanja in množenja na način, ki omogoča smiselno definicijo vsote in produkta omrežij. Ustrezne algebrske structure so polkolobarji. Običajni pristop k (podatkovni) analizi časovnih omrežij uporablja pretvorbo omrežja na zaporedje časovnih rezin - navadnih statičnih omrežij, ki ustrezajo izbranim časovnim intervalom. Z analizo (z običajnimi metodami) posameznih rezin dobimo ustrezno zaporedje rezultatov. Pristop, predlagan v tem članku, nam omogoča izračunati te rezultate naravnost. Razvili smo hitre algoritme za predlagane operacije. Dostopni so v odprto-kodni pythonski knjižnici TQ (Temporal Quantities). Predlagani pristop nam omogoča, da obravnavamo kot časovne količine tudi druge značilnosti omrežij, kot so: stopnje, komponente povezanosti, mere pomembnosti, ogrodje Pathfinder, itd. Glavni namen članka je bil podati teoretične temelje pristopa in pokazati "da gre". Pri tem je ostala še vrsta odprtih vprašanj. Tudi knjižnico TQ je mogoče z uporabo redke predstavitve omrežij za velika omrežja precej pohitriti. To so tudi naloge, ki jih nameravamo razrešiti v tem projektu.

3. Cilji raziskave s poudarkom na izvirnosti predlaganih raziskav ter njihovim potencialnim vplivom na razvoj novih raziskovalnih smeri

- Zbrati dodatne primerke časovnih omrežij. Razviti nameravamo novo obliko opisa (format) omrežnih podatkov netJSON, ki temelji na obliki opisa JSON, in jo uporabiti za opis vseh omrežij iz naše zbirke.
- Razširiti pristop temelječ na časovnih količinah na druge pojme iz analize omrežij kot so mere pomembnosti, sredice, otoki, itd.
- S konstrukcijo ustreznih polkolobarjev razširiti algebrajski pristop na probleme s potovanji (journeys) – časovni prehodi po povezavah z neničelnimi prehodnimi časi.
- Razviti učinkovite algoritme za analizo velikih časovnih omrežij opisanih s časovnimi količinami ob upoštevanju tega, da je večina dejanskih velikih omrežij redka (ima majhno povprečno stopnjo).
- Proučiti možnost uporabe knjižnic za vzporedno računanje, kot je na primer Solomonik in Hoefler 2015, za pohitritev izračunov.
- Razviti metodologije analize časovnih omrežij za izbrana področja uporabe, kot so npr scientometrika, analiza rodovnikov, analiza sporočil elektronske pošte, analiza telefonskih klicev, itd.
- Pomemben problem pri delu z večjimi časovnimi omrežji je obsežnost samih podatkov in prav tako obsežnost dobljenih rezultatov. Pri dosedanjem delu so se zelo dobro obnesli slikovni prikazi časovnih količin. V projektu nameravamo poiskati alternativne oblike pregledovanja časovnih omrežij in iskanja zanimivih sestavin v rezultatih. V javascriptu bomo razvili knjižnico

Javni razpis za sofinanciranje raziskovalnih projektov za leto 2017

netD3 za prikaze omrežij, ki temelji na knjižnici D3.js.

4. Izvirnost pričakovanih rezultatov

Pričakujemo, da bomo dobili več izvirnih teoretičnih rezultatov (izgradnja ustreznih polkolobarjev) in razvili več novih in izvirnih učinkovitih algoritmov za analizo velikih časovnih omrežij. Občestvu raziskovalcev, ki se ukvarjajo z analizo omrežij, bomo ponudili netJSON kot splošni "jezik" za opis omrežij.

5. Metode dela

Čeprav smo s tematiko časovnih omrežij že dobro seznanjeni se bomo morali poglobiti v literature za nekaj posebnih tem. Organizirali bomo poseben seminar, na katerem si bomo izmenjevali znanje in sledili dejavnostim na projektu. Nekatero nalogo (npr. zbiranje in kodiranje primerkov časovnih omrežij) so rutinske, a zahtevajo precej časa. Druge naloge zahtevajo najprej teoretične rešitve, ki jih kasneje prelijemo v računalniške programe. Te moramo še preizkusiti in uporabiti na dejanskih omrežjih. Dobljeni rezultati bodo predstavljeni na seminarjih, znanstvenih srečanjih in objavljeni kot članki in (prosto dostopne) programske knjižnice.

6. Relevantnost in potencialni vpliv rezultatov

Rezultati projekta bodo dali nova orodja in metodologije za analizo časovnih omrežij. Oblika zapisa netJSON bo ponujal splošen "jezik" za opis omrežij in bi ga lahko uporabljali kot posredno obliko za izmenjavo omrežnih podatkov med različnimi programi za analizo omrežij. Knjižnica netD3, ki bo temeljila na netJSON, bo podpirala najrazličnejše interaktivne spletne prikaze omrežij. Avtorji programov za analizo omrežij lahko pridejo do prikazov svojih rezultatov preprosto z izpisom the v obliki netJSON in uporabo ustreznih rutine iz knjižnice netD3 na tej datoteki. Našo zbirko časovnih omrežij bomo ponudili na spletu.

7. Izjemni družbeno-ekonomsko oziroma kulturno relevantni dosežki vodje projekta

Vodja projekta je leta 1996 začel z Andrejem Mrvarjem razvoj programa Pajek za analizo in prikaz velikih omrežij. Pajek je še vedno edini splošno namenski program za analizo omrežij z desetimi milijoni vozlišč. Leta 2013 je Pajek prejel William D. Richards Jr., Software Award združenja INSNA.

Vodja projekta je več desetletij sodeloval pri uvajanju IT v slovenske šole. Pomemben je tudi njegov prispevek pri uvajanju vrste programskih orodij v slovenski prostor: pascal, logo, TeX and LaTeX, Postscript, Python. Leta 2011 je za svoje prispevke prejel priznanje združenja Informacijska družba.

Leta 1991 je z Anuško Ferligoj in Patom Doreianom razvil nov pristop k bločnemu modeliranju omrežij. Rezultati več kot desetletnega raziskovanja so bili zbrani v knjigi "Generalized blockmodeling" (CUP 2004). Za to knjigo so avtorji leta 2007 prejeli priznanje združenja Harrison White Outstanding Book Award.

Več kot petnajst let predava analizo omrežij na delavnicah in kratkih tečajih širom po svetu na raznih univerzah in drugih ustanovah. Deset let je imel dvotedenski tečaj iz analize omrežij na mednarodni poletni šoli ECPR v Ljubljani. Analizo omrežij predava tudi na doktorskem študiju statistike na Univerzi v Ljubljani.

Javni razpis za sofinanciranje raziskovalnih projektov za leto 2017

8. Organiziranost in izvedljivost projekta.

Za vse naloge iz razdelka 20.3 že imamo zamisli, kako se jih lotiti. Najbolj "meglena" je naloga razširiti algebrski pristop na probleme s potovanji, toda tudi v tem primeru pričakujemo zanimive rezultate – vprašanje je le kako daleč se nam bo uspelo prigrusti.

Kot je bilo že omenjeno, bo project organiziran okrog posebnega seminarja, ki bo omogočal izmenjavo znanja in spremljanje dejavnosti na projektu. Vzpostavili bomo tudi projektni wiki, ki bo ponujal skupni prostor za projektno dokumentacijo in sodelovanja. Postavili bomo tudi spletno stran z omrežnimi podatki in programi.

V začetni fazi (prvo leto) projekta bomo razvili obliko zapisa netJSON in pretvorili našo zbirko časovnih omrežij vanj. Hkrati lahko steče teoretično delo na novih polkolobarjih in algoritmih.

Ko bo razvit nov algoritem, ga bomo sprogramirali in stestirali (od 7. meseca do konca projekta). Razvoj knjižnice netD3 bomo začeli v drugem letu.

Pričakujemo, da bo po letu in pol projekta učinkovita izvedba knjižnice TQ za velika redka omrežja že uporabna za resnejše raziskave. Takrat bomo začeli z razvojem metodologij za analizo izbranih vrst omrežij.

Rezultate projekta bomo predstavljali na raznih seminarjih, delavnicah in konferencah, na projektni spletni strani ter objavili v znanstvenih revijah. Rezultate bomo vključili tudi v svoja predavanja o analizi omrežij.

Če bodo na voljo, bomo v project vključili tudi posamezne doktorske študente.

9. Viri

Allen JF 1983 Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11), 832–843.

Batagelj V and Zaveršnik M 2011 Fast algorithms for determining (generalized) core groups in social networks. *Advances in Data Analysis and Classification* 5(2), 129–145.

Batagelj V and Praprotnik S 2016 An algebraic approach to temporal network analysis based on temporal quantities. <https://arxiv.org/abs/1505.01569>. In print in *Social Network Analysis and Mining*.

Bell MGH and Iida Y 1997 *Transportation Network Analysis*. Wiley, Chichester.

Casteigts A and Flocchini P 2013 *Deterministic Algorithms in Dynamic Networks: Formal Models and Metrics*. Technical Report DRDC Ottawa CR 2013-020, Defence R&D Canada.

Casteigts A, Flocchini P, Quattrociocchi W and Santoro N 2012 Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems* 27(5), 387–408.

Correa JR and Stier-Moses NE 2011 Wardrop equilibria In *Wiley Encyclopedia of Operations Research and Management Science* (ed. Cochran JJ, Cox LAJ, Keskinocak P, Kharoufeh JP and Smith JC) John Wiley & Sons, Inc.

Javni razpis za sofinanciranje raziskovalnih projektov za leto 2017

D3.js - Data-Driven Documents. <https://d3js.org/>

Dechter R 2003 Constraint Processing. Morgan Kaufmann, San Francisco.

George B, Kim S 2013 Spatio-temporal Networks; Modeling and Algorithms. Springer Briefs in Computer Science. Springer.

Holme, P. 2015 Modern temporal network theory: a colloquium. Eur. Phys. J. B 88, 234.

Holme P and Saramäki J 2012 Temporal networks. Physics Reports 519(3), 97–125.

Holme P and Saramäki J 2013 Temporal networks. Understanding Complex Systems. Springer.

Holme P, Kim BJ, Yoon CN and Han SK 2002 Attack vulnerability of complex networks. Physical Review E 65(5), 056109.

JSON. <http://www.json.org/>

Kempe D, Kleinberg J and Kumar A 2000 Connectivity and inference problems for temporal networks Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing, pp. 504–513. ACM, New York.

Kolaczyk ED 2009 Statistical Analysis of Network Data: Methods and Models. Springer, New York.

Moder JJ and Phillips CR 1970 Project Management with CPM and PERT second edn. Van Nostrand Reinhold Company, New York.

Nicosia V, Tang J, Mascolo C, Musolesi M, Russo G and Latora V 2013 Graph metrics for temporal networks In Temporal Networks (ed. Holme P and Saramäki J) Understanding Complex Systems Springer Heidelberg, New York, Dordrecht, London pp. 15–40.

Ripley RM, Snijders TAB, Boda Z, Vörös A and Preciado P 2013 Manual for SIENA version 4.0 (version November 1, 2013) University of Oxford, Department of Statistics; Nuffield College Oxford.

Santoro N, Quattrociocchi W, Flocchini P, Casteigts A 2011 Time-varying graphs and social network analysis: Temporal indicators and metrics. 3rd AISB Social Networks and Multiagent Systems Symposium (SNAMAS), pages 32–38.

Solomonik E, Hoefler T 2015 Sparse Tensor Algebra as a Parallel Programming Model. <http://arxiv.org/abs/1512.00066>

Vilain M, Kautz H and van Beek P 1990 Constraint propagation algorithms for temporal reasoning: A revised report In Readings in Qualitative Reasoning About Physical Systems (ed. Weld DS and de Kleer J) Morgan Kaufmann Publishers Inc. San Francisco, CA, USA pp. 373–381.

Xuan BB, Ferreira A, Jarry A 2003 Computing shortest, fastest, and foremost journeys in dynamic networks. International Journal of Foundations of Computer Science, 14(2), 267–285.