

# Določanje pragovne funkcije za zapis kompleksne mreže kompeticije dreves z evolucijskim algoritmom

David Jesenko<sup>a</sup>, Urša Vilhar<sup>b</sup>, Mitja Skudnik<sup>b</sup>, Borut Žalik<sup>a</sup>, Domen Mongus<sup>a</sup>

<sup>a</sup>Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Koroška cesta 46, 2000 Maribor

<sup>b</sup>Gozdarski inštitut Slovenije, Večna pot 2, 1000 Ljubljana

E-pošta: david.jesenko@um.si

## Determination of the threshold function for defining a complex network of tree competition with an evolutionary algorithm

**Abstract.** *Complex networks have become a powerful mathematical tool for modelling natural phenomena. Most of studies target primarily on topological properties of nodes, while the non-topological properties remain unnoticed. This paper presents an evolutionary algorithm for discovering the relations of nodes' non-topological properties. To provide an user-friendly interpretation of results, they are expressed with a threshold function. In this way, not only those non-topological properties of nodes that have the largest influences on the connectivity in complex network are revealed, but also the intensities of their influences are estimated by the threshold function coefficients. Applicability of the presented method is demonstrated on a real-world study of tree competition complex networks.*

## 1 Uvod

Kompleksne mreže so postale pomembno orodje raziskovanja na različnih področjih znanosti, kot so biološke in kemijske reakcije [1], evolucijska dinamika [2], interakcije živih bitij [3] in socialna omrežja [4]. Skupna značilnost teh, vsebinsko zelo raznolikih področij je, da so posamezni elementi v interakciji [5]. Omenjene pojave je formalno najlažje predstaviti z grafi, kjer opazovane entitete predstavljajo vozlišča, medtem ko njihove interakcije opišemo s povezavami. Grafom, ki opisujejo interakcije z velikim številom entitet, opisujejo naravne pojave [6], ter imajo pogosto dinamične lastnosti [7], pravimo kompleksne mreže [8].

Do danes se je mnogo raziskav osredotočalo na razumevanje razvoja mrež [9], njihovo strukturo in dinamiko [10] ter druge topološke značilnosti [11]. Kljub temu da omenjene analize in topološke lastnosti omogočajo odkrivanje različnih vplivov na obnašanje kompleksnih mrež, ostajajo kompleksne mreže aktualno raziskovalno področje. Eno izmed njih je zagotovo študija netopoloških vplivov (lastnosti vozlišč) na interakcije v mreži. Netopološke lastnosti vozlišč v kompleksnih mrežah predstavljajo lastnosti entitet in niso povezane s topologijo mreže. Mednje lahko uvrščamo odzive vozlišč v različnih časovnih obdobjih, geometrijske lastnosti vozlišč (npr. višina, prostornina, površina elementa, ki je predstavljen z vozliš-

ščem), položaje posameznih vozlišč in druge od različnih pojavov odvisne lastnosti elementov. Mreže, pri katerih je položaj vozlišč v prostoru znan in pomemben, imenujemo prostorska omrežja [12]. Primeri takšnih mrež so mreže letalskih letov [13], mest s cestnimi povezavami [14], ladijskega prometa [15] in celo kompeticije med drevesi v gozdu [16, 17]. Prav kompeticiji med drevesi, kjer poleg razdalje med drevesi (vozlišči), tudi druge, predvsem geometrijske lastnosti vozlišč (npr. višina, premer, temeljnica, socialni položaj drevesa, drevesna vrsta) odločilno vplivajo na povezljivost v kompleksno mrežo [18], bomo v tem članku posvetili pozornost.

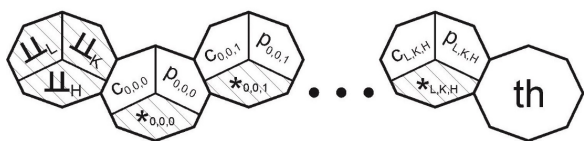
Rezultate naše metode bomo predstavili v obliki pragovne funkcije. Za razliko od večine metod strojnega učenja, takšen zapis omogoča preprosto in intuitivno interpretacijo rezultatov. Zapis v obliki pragovne funkcije ne samo da razkriva tiste netopološke lastnosti vozlišč, ki najbolj vplivajo na povezljivost v kompleksni mreži, ampak s koeficienti določa tudi intenzivnost njihovega vpliva.

Algoritem bomo predstavili v drugem poglavju. V poglavje tri bomo predstavili rezultate in interpretacijo pragovnih funkcij. S poglavjem štiri bomo zaključili članek.

## 2 Algoritem

V tem poglavju bomo predstavili dvonivojski evolucijski algoritem za iskanje pragovne funkcije, ki določa povezljivost v kompleksni mreži na podlagi netopoloških lastnosti vozlišč [19]. Naloga evolucijskega algoritma je poiskati pragovno funkcijo, ki bo na podlagi referenčne kompleksne mreže generirala mrežo, ki bo karseda podobna iskani (referenčni) mreži. Iskanje pragovne funkcije je izvedeno na dveh nivojih, kjer se na prvem nivoju išče oblika funkcije, na drugem pa optimalni koeficienti in prag funkcije. Za predstavitev pragovne funkcije v algoritmu uporabljamo posameznike (kromosome), ki so sposobni predstaviti polinomske, racionalne, iracionalne, eksponentne in preproste geometrijske funkcije. Posamezni členi funkcije so v kromosomu zakodirani v genih, razen prvega in zadnjega. Prvi predstavlja seštevanje ali množenje ( $\prod_L$ ) vseh členov funkcije in stopnjo funkcije ( $\prod_K$  in  $\prod_H$ ), zadnji pa prag pragovne funkcije  $th$ . Ostali geni, ki predstavljajo posamezne člene funkcije, so sestavljeni iz koeficientov  $c$ , potenc  $p$  in generičnega sim-

bola  $*$ , ki zajema eno izmed možnih operacij ( $+$ ,  $-$ ,  $\times$  ali  $/$ ) med enakima lastnostima obeh obravnavanih vozlišč (npr. razlika med prostornino prvega in drugega vozlišča). Grafično predstavitev kromosoma lahko vidimo na sliki 1.



Slika 1: Grafična predstavitev posameznika, kjer šrafirane dele optimiziramo na prvem nivoju, ostale pa na drugem.

Najbolj primerno obliko funkcije za dano kompleksno mrežo poiščemo na prvem nivoju, kjer optimiziramo zgolj parametre  $\square$  in  $*$ . Na tem nivoju inicializiramo parametre algoritma (velikost populacije, verjetnost mutacije in križanja, največje dovoljeno število funkcijskih evaluacij, interval za naključno generiranje koeficientov in praga ter delež naključno generiranih posameznikov na drugem nivoju). Začetna populacija se generira na prvem nivoju ter evaluira na drugem pred vstopom v glavno zanko mutacij, križanj in selekcij. Te operacije so definirane kot:

- **Križanje** dosežemo z izmenjavo posameznih genov med dvema staršema, kjer je prvi izbran zaporedno iz generacije, drugi pa naključno (s tem zagotovimo, da je vsak kromosom za križanje izbran vsaj enkrat). Implementirali smo dvotočkovno križanje, kjer sta začetni in končni točki križanja izbrani naključno. Zaradi determinističnega načina izvedbe algoritma, je vsak potomec ocenjen takoj po križanju in v primeru boljšega rezultata zamenja zaporedno izbranega starša.
- **Mutacija** je implementirana z naključno zamenjavo operatorjev (šrafirani deli na sliki 1). Število mutiranih posameznikov je enako verjetnosti mutacije. Operacije so izbrane in mutirane naključno, vsak mutant pa zamenja starša v primeru boljšega rezultata.
- **Selekcijo** dosežemo s primerjavo dveh posameznikov iz generacije in samo posameznik z boljšim rezultatom napreduje v naslednjo generacijo. Izvedena je turnirska selekcija velikosti dva.

Zanka omenjenih operacij se zaključí, ko dosežemo največje dovoljeno število funkcijskih evaluacij ali pa, ko dosežemo največjo možno uspešnost posameznika [20]. Ocenitev posameznika izvedemo na drugem nivoju, kjer za obliko pragovne funkcije iz prvega nivoja poiščemo najbolj optimalne parametre.

Na drugem nivoju določimo optimalne vrednosti koeficientov  $c$ , potenc  $p$  in pragu funkcije  $th$ . Ob prvi ocenitvi vsakega posameznika iz prvega nivoja vse parametre na drugem nivoju generiramo naključno. V vsaki naslednji ocenitvi delež posameznikov (vhodni parameter

algoritma) generiramo naključno, ostale pa generiramo iz najboljših parametrov prejšnjega zagona drugega nivoja za obravnavanega posameznika. Tudi na drugem nivoju inicializaciji sledi zanka operacij evolucionjskega algoritma, ki so definirani:

- **Križanje** je zelo podobno križanju iz prvega nivoja s to razliko, da med staršema izmenjamo samo koeficiente  $c$ , potence  $p$  in prag  $th$ , saj so ostali deli vseh posameznikov na drugem nivoju enaki.
- **Mutacija** je podobna križanju, saj mutiramo samo koeficiente  $c$ , potence  $p$  in prag  $th$ . Največja dovoljena sprememba  $c$  in  $th$  je določena z vhodnim parametrom algoritma, potenco  $p$  pa lahko spreminjamo samo za 0, 5.
- **Selekcija** je enaka kot na prvem nivoju, turnirska, velikosti dva.
- **Ocenitvena funkcija** je realizirana s konstrukcijo kompleksne mreže s pragovno funkcijo, ki je predstavljena v posamezniku. Podobnost referenčne in generirane kompleksne mreže dosežemo s primerjavo povezav, saj so vozlišča in netopološke lastnosti vozlišč v obeh mrežah enaki. Kvantitativno ocenitve podobnosti dosežemo s štejetjem pravilno povezanih, nepravilno povezanih, pravilno nepovezanih in nepravilno nepovezanih povezav. S temi meritvami lahko izračunamo harmonično razmerje FI med metrikama celovitosti in pravinosti [21].

Drugi nivo se zaključí, ko najdemo optimalne vrednosti za parametre  $c$ ,  $p$  in  $th$  ali pa v primeru, ko se najboljši posameznik ne spremeni pet zaporednih generacij. Po končanem drugem nivoju se najboljši posameznik prenese nazaj na prvi nivo, kjer evolucionjski algoritem nadaljuje izvajanje.

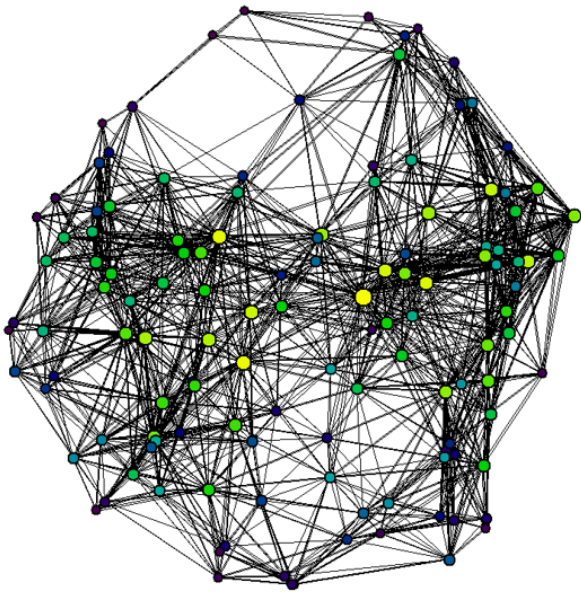
## 3 Rezultati

### 3.1 Podatki

Raziskave so potekale na desetih ploskvah intenzivnega monitoringa gozdov (IMGE) v Sloveniji, ki poteka v okviru mednarodnega programa sodelovanja za oceno in sledenje učinkov onesnaženega zraka na gozdove (Program intenzivnega monitoringa gozdnih ekosistemov, ang. ICP Forests, dosegljivo na <http://www.icp-forests.org/>) [22]. Ploskve velikosti  $50m \times 50m$  se nahajajo v gozdovih različnih gozdnih združb in v različnih ekoloških regijah: smrekov gozd na Pokljuki (alpska ekološka regija) in Pohorju, bukovi gozdovi v dinarski in predalpski ekološki regiji ter hrastovi gozdovi v predpanonski ekološki regiji. Za vsako drevo na vsaki ploskvi s premerom večjim od  $7cm$  so bile izmerjene oziroma določene naslednje lastnosti: drevesna vrsta  $t$ , status (mrtvo ali živo)  $l$ , lokacija (koordinati  $x$  in  $y$ ), premer  $d$  na višini  $1.3m$ , socialni status (5 razredov)  $s$ , višina  $h$  ter lesna zaloga  $v$  drevesa.

Mreže kompeticije (glej sliko 2) na ploskvah IMGE so bile določene z Bitterlichovo metodo [23]. Omenjeni pristop za razpoznavo kompetitorjev dreves (sosednjih dreves v gozdu) določí linearno razmerje med drevesnim

premerom in območjem kompeticije drevesa. Glede na splošno stanje opazovanih gozdov je bil za izračun območja kompeticije izbran faktor (ang. basal area factor)  $A = 5$ , saj je na ploskvah IMGE zaradi majhne gostote dreves visoka vidljivost. Prednost Bitterlichovega pristopa je relativno preprosto računanje območja kompeticije in pa dejstvo, da sta za računanje območja kompeticije potrebna le podatka o lokaciji in premeru drevesa, ki je v gozdarstvu največkrat merjena lastnost dreves in omogoča uporabo Bitterlichovega pristopa tudi v drugih gozdovih.



Slika 2: Kompleksna mreža kompeticije med drevesi, kjer večja in svetlejša vozlišča predstavljajo dominantnejša drevesa.

### 3.2 Interpretacija pragovnih funkcij

Primeri pragovnih funkcij za kompleksne mreže kompeticije med drevesi in povprečna uspešnost  $F1$  po dvajsetih zagonih algoritma za vseh deset ploskev IMGE so prikazani v tabeli 1, iz katere lahko izluščimo različna spoznanja in intenzivnost vpliva posamezne lastnosti. Prva in najbolj izstopajoča lastnost je razdalja  $D$  med posameznima drevesoma (vozliščema). Pri vseh pragovnih funkcijah ima koeficient razdalje  $D$  visoko negativno vrednost, kar pomeni, da večja kot je razdalja med drevesoma, manjša je kompeticija med njima. To je pričakovano, saj drevesa rastejo na enem mestu, med seboj pa konkurirajo za svetlobo, hranila in vodo v tleh. Bolj oddaljena drevesa v gozdu v manjši meri vplivajo na razpoložljivost teh dejavnikov za izbrano drevo. Druga prav tako pomembna lastnost, ki se pojavi v vseh pragovnih funkcijah, je višina dreves  $h$ . Višja drevesa imajo praviloma večje premere debel, večje krošnje in večje koreninske sisteme, ter posledično bolj konkurirajo za svetlobo, hranila in vodo v tleh. Območje kompeticije je torej sorazmerno z višino posameznega drevesa, na kar kažejo tudi samo pozitivni koeficienti pri višinah  $h$  v pragovnih funkcijah. V primeru višine je vidna tudi povezava

s pragom pragovne funkcije, saj je v primeru manjšega koeficienta praviloma manjši tudi prag. Vplivi drugih lastnosti se razlikujejo od vzorčne ploskve do vzorčne ploskve. Poglejmo si primer za drevesno vrsto  $t$ . V smrekovem gozdu na Pokljuki je vpliv te lastnosti skoraj zanemarljiv, saj je vrednost koeficienta zelo majhna (ista drevesna vrsta), na drugi strani pa je v gozdu na Kostanjevici zelo pomembna, saj gre za mešan gozd listavcev (visoka vrednost koeficienta). Tudi za ostale lastnosti lahko poiščemo podobne razlage in povezave z dejanskim stanjem v gozdu.

## 4 Zaključek

Predstavili smo algoritem za iskanje pragovne funkcije, ki določa povezave v kompleksni mreži. Algoritem se osredotoča in povezuje vozlišča na podlagi netopoloških lastnosti vozlišč. S primeri kompleksnih mrež kompeticije med drevesi smo pokazali uporabnost predstavljenega dvonivojskega evlucijskega algoritma. Zaradi zapisa rezultata v obliki pragovne funkcije, kjer so vidne tako netopološke lastnosti kot tudi njihove intenzivnosti (koeficienti), je interpretacija preprosta (ob predpostavki, da ima uporabnik specifično domensko znanje). Interpretacija rezultatov (pragovnih funkcij) je pokazala oziroma potrdila že znana dejstva (velik vpliv razdalje med drevesi in višine dreves na kompeticijo v gozdu), kar kaže na uporabnost metode. Metodo želimo v prihodnje preizkusiti še na drugih podatkih.

## Zahvala

Delo je podprla Javna agencija za raziskovalno dejavnost Republike Slovenije s projektom J2-6764 in programskima skupinama P2-0041 ter P4-0107.

## Literatura

- [1] L. Santana, H. Gonzalez-Diaz, E. Quezada, E. Uriarte, M. Yanez, D. Vina, F. Orallo, Quantitative structure-activity relationship and complex network approach to monoamine oxidase a and b inhibitors, *Journal of medicinal chemistry* 51 (21) (2008) 6740–6751.
- [2] I. Zelinka, D. Davendra, S. Roman, J. Roman, Do evolutionary algorithms dynamics create complex network structures?, *Complex Systems* 20 (2) (2011) 127–140.
- [3] J. Mourier, J. Vercelloni, S. Planes, Evidence of social communities in a spatially structured network of a free-ranging shark species, *Animal Behaviour* 83 (2) (2012) 389–401.
- [4] K. R. Žalik, Maximal neighbor similarity reveals real communities in networks, *Scientific reports* 5 (2015) 18374.
- [5] B. J. O’Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D.

| Ploskev       | Pragovna funkcija  | F1     |
|---------------|--|--------|
| Pokljuka      | $-13.9D(n_i, n_j) + 8.5(h_i + h_j) + 0.2(t_i + t_j) > 112.3$                                 | 0.9686 |
| Trnovski gozd | $-14.6D(n_i, n_j) - 2.2(s_i + s_j) + 4.6(h_i + h_j) > 51.1$                                  | 0.9625 |
| Sežana        | $-11.1D(n_i, n_j) + 3.9(s_i - s_j) + 1.8(h_i + h_j) + 9.1(t_i + t_j) > 2.2$                  | 0.9378 |
| Kranj         | $-13.2D(n_i, n_j) - 14.4(t_i/t_j) + 1.9(s_i + s_j) + 5.7(h_i + h_j) - 2.3(d_i + d_j) > 26.6$ | 0.9677 |
| Kočevska reka | $-7.1D(n_i, n_j) + 2.8(h_i + h_j) + 2.9(v_i + v_j) > 9.1$                                    | 0.9287 |
| Zasavje       | $-12.3D(n_i, n_j) + 7.2(h_i + h_j) + 10.4(s_i + s_j) > 130.7$                                | 0.9528 |
| Loški potok   | $-9.5D(n_i, n_j) + 1.9(s_i - s_j) + 9.4(d_i/d_j) - 1.4(v_i + v_j) + 2.3(h_i + h_j) > 15.4$   | 0.9441 |
| Kostanjevica  | $-13.7D(n_i, n_j) + 2.7(h_i/h_j) + 8.5(v_i + v_j) + 3.2(s_i + s_j) - 11.8(t_i + t_j) > 10.6$ | 0.9359 |
| Lendava       | $-12.7D(n_i, n_j) + 5.1(h_i + h_j) + 0.8(t_i + t_j) > 116.3$                                 | 0.9384 |
| Pohorje       | $-8.6D(n_i, n_j) + 0.4(v_i \times v_j) + 5.5(h_i + h_j) - 3.6(d_i + d_j) > 87.5$             | 0.9534 |

Tabela 1: Primeri pragovnih funkcij za kompleksne mreže kompeticije med drevesi, kjer  $D$  predstavlja funkcijo razdalje med vozliščema  $n_i$  in  $n_j$ .

- Smith, et al., Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations, *Nature* 485 (7397) (2012) 246–250.
- [6] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, *Science* 298 (5594) (2002) 824–827.
- [7] R. J. Prill, P. A. Iglesias, A. Levchenko, Dynamic properties of network motifs contribute to biological network organization, *PLoS biology* 3 (11) (2005) e343.
- [8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, *Physics reports* 424 (4) (2006) 175–308.
- [9] S. N. Dorogovtsev, J. F. Mendes, Evolution of networks, *Advances in physics* 51 (4) (2002) 1079–1187.
- [10] K. R. Žalik, B. Žalik, A local multiresolution algorithm for detecting communities of unbalanced structures, *Physica A: Statistical Mechanics and its Applications* 407 (2014) 380–393.
- [11] D. Jesenko, D. Mongus, B. Žalik, Spatially embedded complex network estimation using fractal dimension, in: *Computer Science Research Conference, 2014*, pp. 5–8.
- [12] L. Daqing, K. Kosmidis, A. Bunde, S. Havlin, Dimension of spatially embedded networks, *Nature Physics* 7 (6) (2011) 481–484.
- [13] Z. Wu, L. A. Braunstein, V. Colizza, R. Cohen, S. Havlin, H. E. Stanley, Optimal paths in complex networks with correlated weights: The worldwide airport network, *Physical Review E* 74 (5) (2006) 056104.
- [14] R. Guimera, M. Sales-Pardo, L. A. N. Amaral, Modularity from fluctuations in random graphs and complex networks, *Physical Review E* 70 (2) (2004) 025101.
- [15] T. Notteboom, J.-P. Rodrigue, Containerisation, box logistics and global supply chains: The integration of ports and liner shipping networks, *Maritime economics & logistics* 10 (1-1) (2008) 152–174.
- [16] G. Kunstler, D. Falster, D. A. Coomes, F. Hui, R. M. Kooyman, D. C. Laughlin, L. Poorter, M. Vanderwel, G. Vieilledent, S. J. Wright, et al., Plant functional traits have globally consistent effects on competition, *Nature* 529 (7585) (2016) 204–207.
- [17] D. Mongus, B. Žalik, An efficient approach to 3d single tree-crown delineation in lidar data, *ISPRS Journal of Photogrammetry and Remote Sensing* 108 (2015) 219–233.
- [18] M. Van Noordwijk, G. Lawson, K. Hairiah, J. Wilson, Root distribution of trees and crops: competition and/or complementarity, *Tree-Crop Interactions: Agroforestry in a Changing Climate*. CABI, Wallingford, UK (2015) 221–257.
- [19] D. Jesenko, M. Mernik, B. Žalik, D. Mongus, Two-level evolutionary algorithm for discovering relations between nodes’ features in a complex network, *Applied Soft Computing* 56 (2017) 82–93.
- [20] M. Črepinšek, S.-H. Liu, M. Mernik, Replication and comparison of computational experiments in applied evolutionary computing: common pitfalls and guidelines to avoid them, *Applied Soft Computing* 19 (2014) 161–170.
- [21] D. M. W. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, *Bioinfo Publications* 2 (1) (2011) 37–63.
- [22] M. Marenče, I. Sinjur, M. Skudnik, U. Vilhar, D. Žlindra, 30 let spremljanja stanja gozdov v Sloveniji, *Gozdarski inštitut Slovenije, Založba Silva Slovenica*, 2017.
- [23] H. E. Burkhart, M. Tomé, Indices of individual-tree competition, *Modeling Forest Trees and Stands* (2012) 201–232.