

# Analiza znanstvene literature pomoću vektorskih ugradnji i mreža suautorstva

20 godina istraživanja u Hrvatskoj

Nikola Balić, doktorski student

Mentor: Prof. dr. sc. Andrina Granić

Fakultet elektrotehnike, strojarstva i brodogradnje, Sveučilište u Splitu

# Podatci o temi doktorskog rada

- Naslov: "Digitalna transformacija sveučilišta s naglaskom na istraživanje, razvoj i inovacije"
- Povezanost seminara: Ovaj seminar doprinosi doktorskom istraživanju razvijanjem metoda za analizu znanstvene produkcije i suradnje. Takav pristup pruža korisne uvide za razumijevanje digitalne transformacije sveučilišta i unapređenje istraživačkih procesa.

# Objavljeni radovi

1. Balić, D.E., Balić, N. (2024). "Navigating through ocean literacy gaps: an analysis of elementary school textbooks in Croatian education". Mediterranean Marine Science.
2. Balić, N., Grubišić, A., Granić, A. (2023). "Perceptions of Digital Learning and Teaching: The Case of a Croatian University Transition to an Emergency Digital Environment". Technology, Knowledge and Learning.
3. Balić, N., Vuka, I. (2023). "State of Sustainable Entrepreneurship in SMEs: Development of a Cross-Country Sustainability Pulse Check". Distributed Computing and Artificial Intelligence, Special Sessions I, 20th International Conference.
4. Vuka, I., Balić, N., Mifsud, M. (2022). "Technology transfer offices as a facilitator of knowledge triangle integration in the knowledge valorisation era: focus group discourse analysis". 15th International Technology Transfer Conference.

# Primjenjene metode u istraživačkom radu

## 1. Prikupljanje podataka

- 225,298 znanstvenih radova iz CroRIS-a
- Razdoblje: 2004. - 2023.

## 2. Predobrada teksta:

- Tokenizacija, uklanjanje stop riječi, stemming

## 3. Konstrukcija grafa suautorstva:

- Korištenje NetworkX i nxCuGraph (GPU akceleracija) biblioteka

## 4. Ugrađivanje autora:

- Primjena Jina ColBERTv2 modela

$$a = \frac{1}{n} \sum_{i=1}^n E(\text{concat}(\text{title}_i, \text{abstract}_i))$$

## 1. Analiza mreže:

- Izračun centralnosti (stupanj, međupoloženost, svojstveni vektor)
- Analiza strukturnih rupa
- Detekcija zajednica pomoću Louvain algoritma

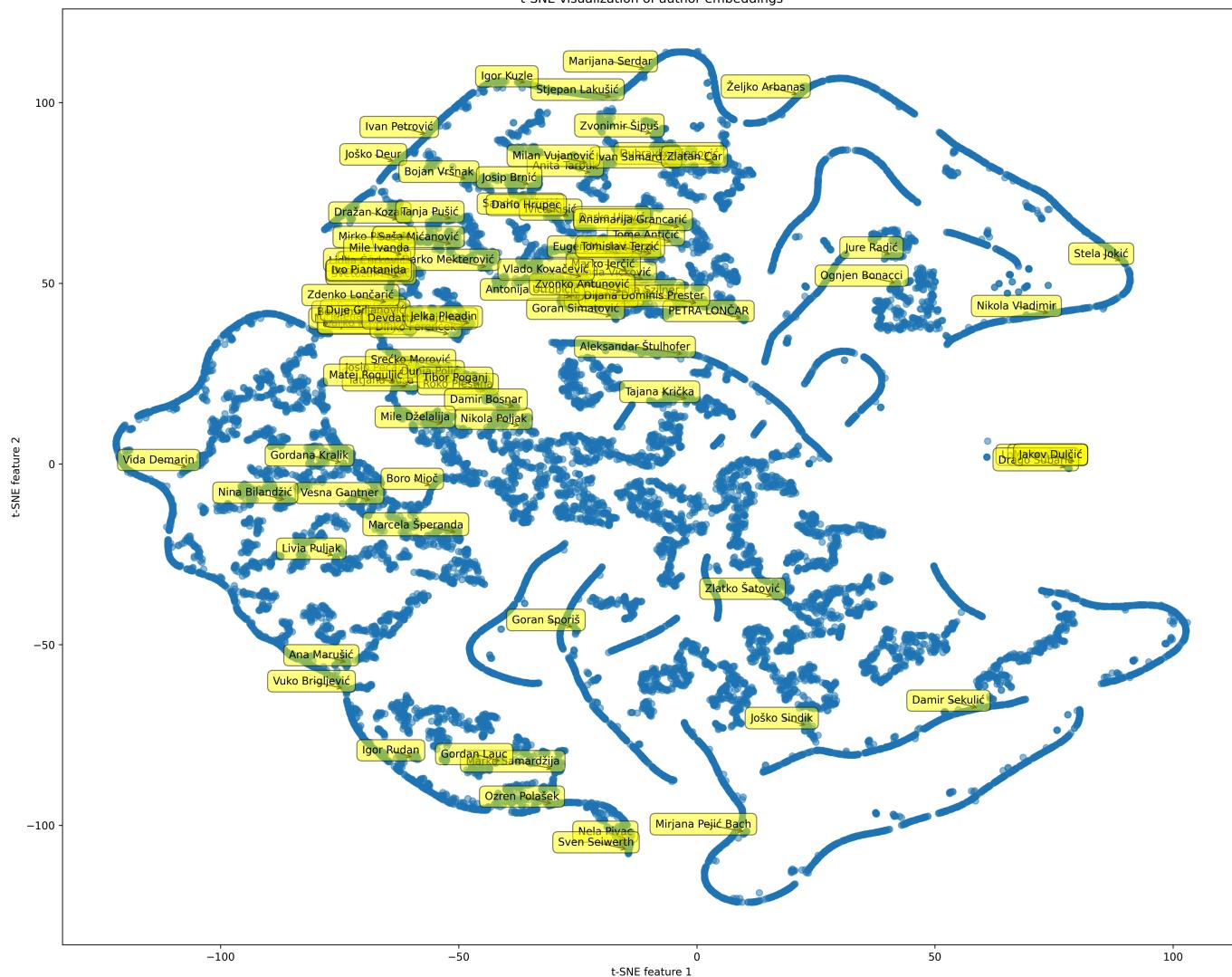
## 2. Grafovske neuronske mreže:

- Implementacija GCN slojeva za učenje ugrađivanja čvorova

# Ostvareni rezultati

- Konstruiran graf suautorstva s 25,582 čvora i 223,419 veza
- Identificirani ključni autori prema različitim mjerama centralnosti
- Otkriveni obrasci suradnje i evolucija kroz vrijeme
- Generirane visokokvalitetne reprezentacije autora
- Vizualizirane zajednice autora pomoću t-SNE

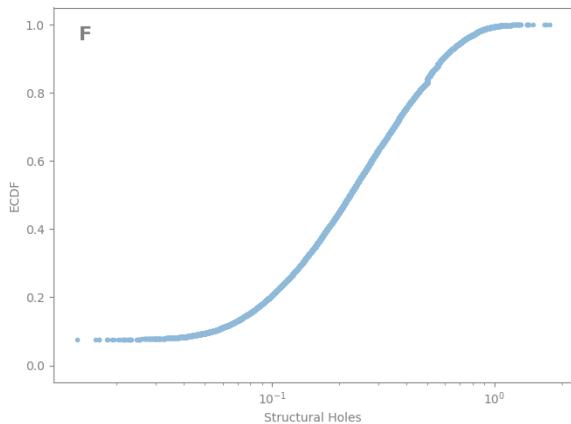
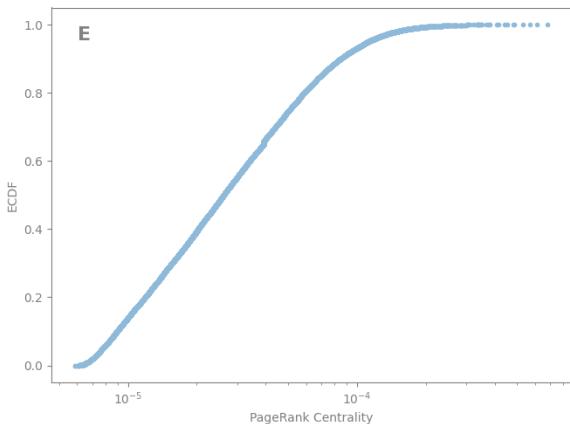
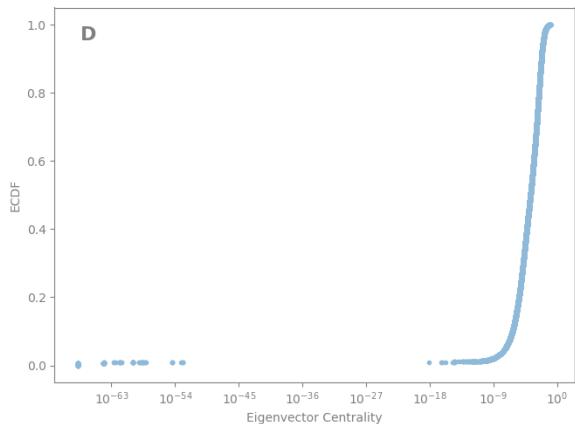
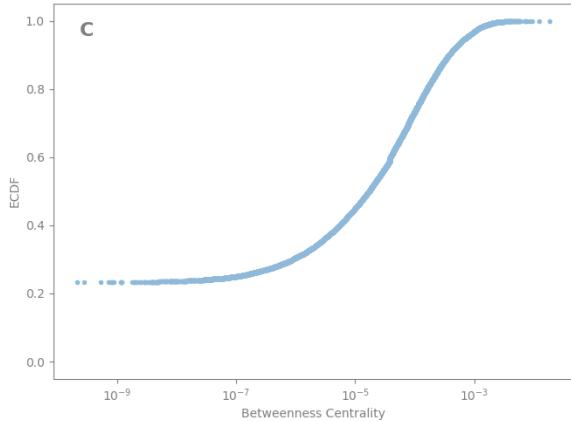
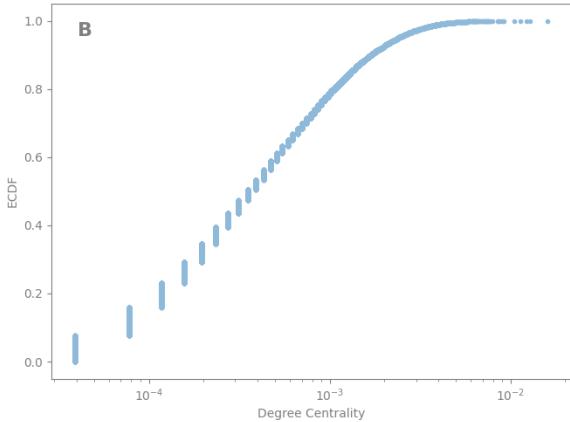
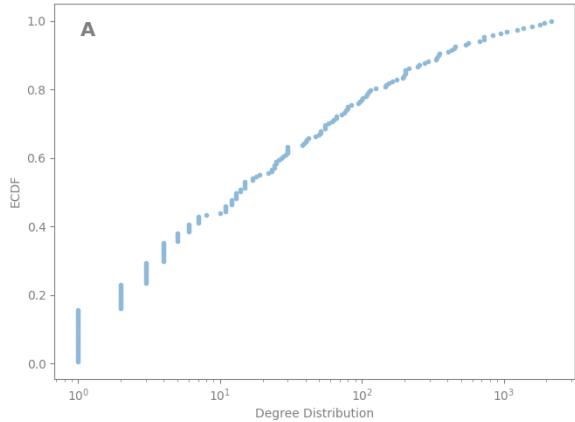
t-SNE visualization of author embeddings



# Ključni nalazi istraživanja

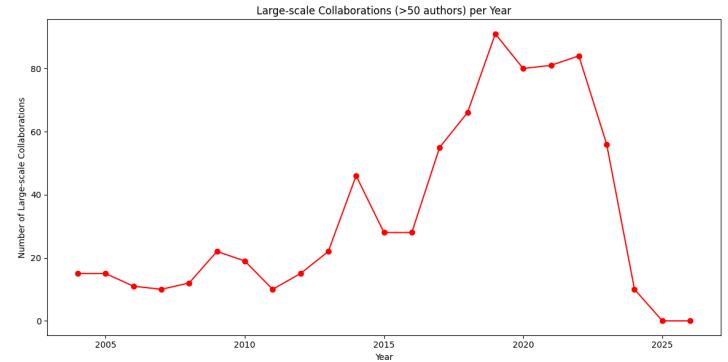
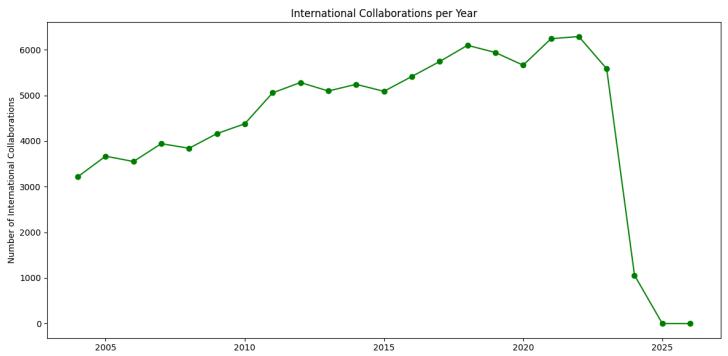
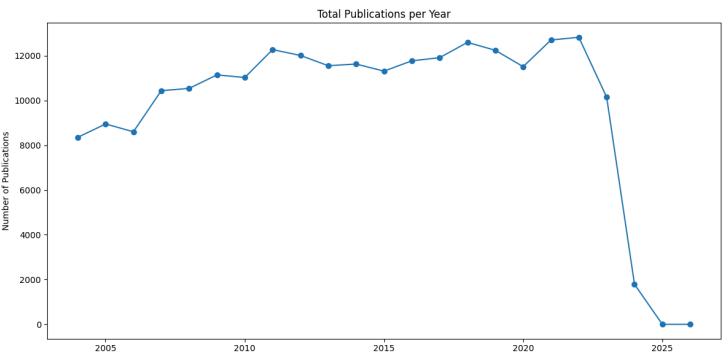
- Prosječan broj autora po publikaciji porastao je s 3,06 u 2004. na 5,00 u 2023.
- Međunarodne suradnje porasle su za 73,57% tijekom promatranog razdoblja
- Identificirano je 143 znanstvene zajednice unutar mreže
- Otkrivena je struktura mreže bez skale (scale-free network)
- Uočen je trend prema većim kolaborativnim naporima i interdisciplinarnim istraživanjima
- Gustoća mreže suautorstva povećala se s 0,0003 u 2004. na 0,0007 u 2023. (omjer stvarnog broja veza u mreži i maksimalnog mogućeg broja veza)

# Analiza centralnosti autora



# Trendovi u znanstvenim publikacijama i suradnjama

- Ukupan broj publikacija porastao je za 21,45% od 2004. do 2023.
- Udio međunarodnih suradnji povećao se s 38,50% u 2004. na 55,02% u 2023.
- Velike suradnje (>50 autora) porasle su za 273,33%, s vrhuncem u 2019.
- Uočen je trend prema većim kolaborativnim skupinama
- Pandemija COVID-19 nije značajno utjecala na broj publikacija i međunarodne suradnje



# Implikacije i budući rad

## Implikacije za znanstvenu politiku:

- Poticanje interdisciplinarnih istraživanja
- Ulaganje u međunarodnu suradnju
- Podrška ključnim autorima koji povezuju različite zajednice

## Potencijalni budući rad:

- Uključivanje podataka o citiranosti za cjelovitiju sliku znanstvenog utjecaja
- Proširenje analize na međunarodne baze podataka
- Presjecanje projekata i publikacija
- Razvoj sofisticiranih temporalnih modela za praćenje evolucije istraživačkih tema
- Razvoj mjere interdisciplinarnosti
- Istraživanje odnosa između položaja u mreži, interdisciplinarnosti i znanstvenog utjecaja

# Korištena literatura

1. Jawahar, G., et al. (2016). "Author2Vec: Learning author representations by combining content and link information". WWW '16.
2. Makarov, I., et al. (2018). "Predicting publication success for authors and co-authorship link prediction using multi-output neural networks". AIST.
3. Jha, S., et al. (2024). "Jina ColBERTv2: A Multilingual Late Interaction Retriever". arXiv preprint.
4. Molontay, R., Nagy, M. (2021). "Two decades of network science: As seen through the co-authorship network of network scientists". Scientometrics.
5. Wang, S., et al. (2017). "Attributed Signed Network Embedding". CIKM '17.

# Zaključak

- Razvijen je okvir za analizu literature koji kombinira NLP tehnike, analizu mreža i duboko učenje
- Identificirani su visoko kolaborativni i utjecajni autori u hrvatskoj znanstvenoj zajednici
- Otkriveni su trendovi prema većoj međunarodnoj suradnji i interdisciplinarnim istraživanjima
- Demonstrirana je učinkovitost pristupa ugrađivanja autora u hvatanju semantičkih i strukturnih odnosa
- Rezultati pružaju uvide za znanstvenu politiku, evaluaciju istraživanja i analizu trendova
- Metodologija ima primjenjivost izvan nacionalnog konteksta i doprinosi polju računalne scientometrije

Table I  
COMPARISON OF TOP 10 AUTHORS ACROSS DIFFERENT CENTRALITY MEASURES

Rank	Degree Centrality	Betweenness Centrality	Eigenvector Centrality	Structural Holes	PageRank Centrality
1	Milan Milošević (0.016028)	Milan Milošević (0.018467)	Marko Samardžija (0.120992)	Ajla Lejlić (1.768000)	Josip Pečarić (0.000682)
2	Sven Seiwerth (0.012822)	Aleksandar Včev (0.012373)	Miroslav Benić (0.115295)	Stjepan Božičević (1.704142)	Sven Seiwerth (0.000611)
3	Gordan Lauc (0.012236)	Ozren Polašek (0.009384)	Maja Popović (0.106211)	Melinda Kostelac (1.680555)	Ozren Polašek (0.000569)
4	Ozren Polašek (0.011297)	Gordan Lauc (0.008402)	Silvijo Vince (0.101444)	Darovan Tušek (1.486400)	Gordan Lauc (0.000532)
5	Aleksandar Včev (0.010516)	Aleksandar Racz (0.007728)	Nino Maćešić (0.097367)	Borislav Klarin (1.430280)	Milan Milošević (0.000486)
6	Marijana Čorić (0.009187)	Maja Miškulin (0.007256)	Hrvoje Valpotić (0.094220)	Mario Duras (1.430280)	Predrag Sikirić (0.000482)
7	Zdenko Lončarić (0.009030)	Sandra Kraljević Pavelić (0.007237)	Željko Cvjetnić (0.085849)	Sandra Jelić (1.411295)	Aleksandar Včev (0.000453)
8	Dunja Rogić (0.008913)	Sven Seiwerth (0.007204)	Ljiljana Bedrica (0.085787)	Maja Ferenc Kuća (1.405625)	Vida Demarin (0.000441)
9	Sandra Kraljević Pavelić (0.008600)	Igor Jerković (0.006011)	Dražen Đuričić (0.085142)	Sanja Zagrajski (1.393811)	Neven Duić (0.000417)
10	Božo Krušlin (0.008600)	Anamarija Jazbec (0.005790)	Blanka Beer Ljubić (0.082339)	Aleksandra Perković (1.393811)	Nela Pivac (0.000407)