

GRÁFBEÁGYAZÓ (GRAPH EMBEDDING) ALGORITMUSOK PERFORMANCIÁJA

SZAKÁCS LILI KATA

TÉMAVEZETŐ: BÉRES FERENC, BENCZÚR ANDRÁS,
MOLNÁR ANDRÁS JÓZSEF
SZTAKI, INFORMATIKAI KUTATÓLABORATÓRIUM

ÖNÁLLÓ PROJEKT III., 2022/23 III. FÉLÉV

BUDAPEST, 2022. 12. 22.

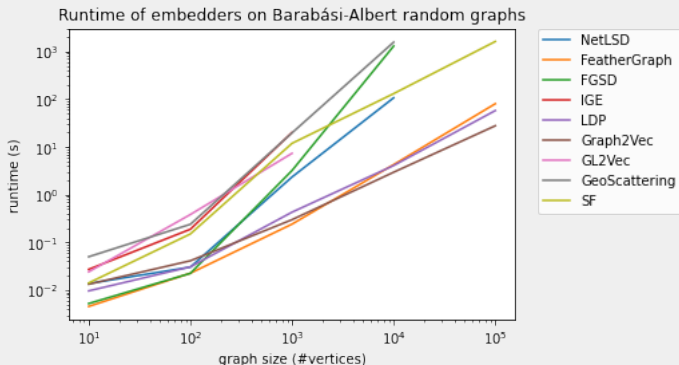
Cél: Gráfok \rightarrow alacsonydimenziós euklideszi vektorok

- strukturálisan hasonló gráfokhoz közeli vektorok
- csúcsok permutációjára invariáns
- hatékony előfeldolgozás pl. klasszifikációhoz

Két gyakori megközelítés:

- Spektrális tulajdonságok alapján
- Sétákkal, csúcsok környezetével

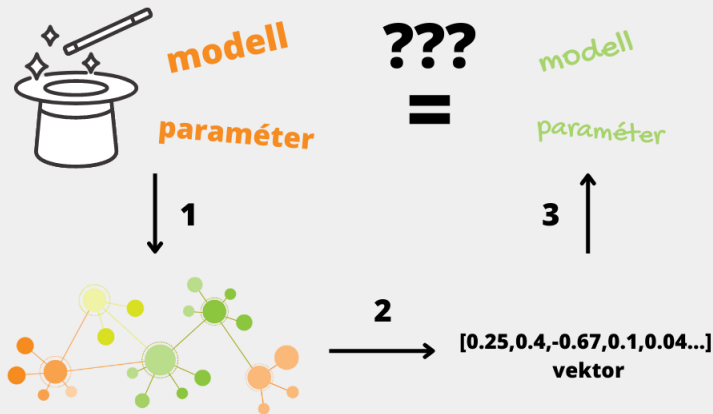
Futásidő vizsgálat a gráf csúcsainak növelésével

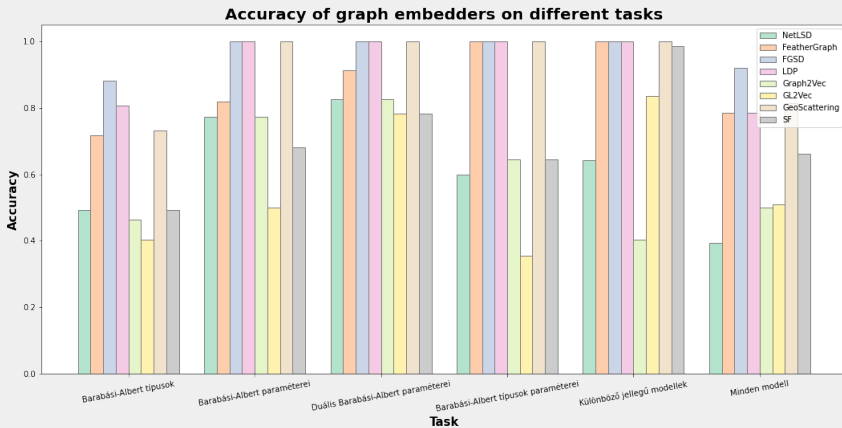


További elemzésre érdemes: *FeatherGraph*, *LDP*, *Graph2Vec* (*SF*)

PROJEKT II.

Klasszifikációs feladathoz előfeldolgozásként használhatjuk a beágyazott gráfokat

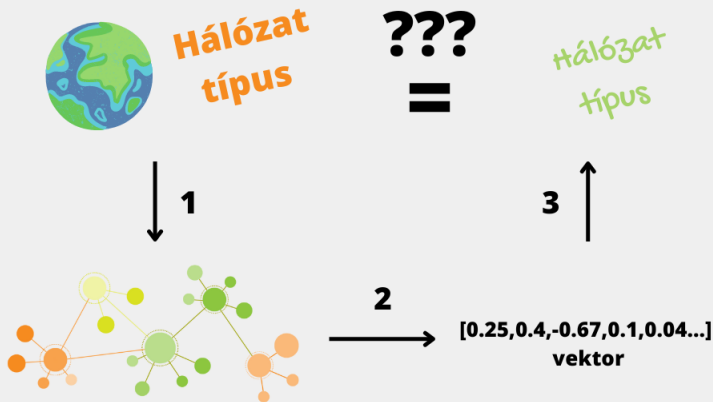




Legjobbak: *FGSD, GeoScattering, LDP, FeatherGraph*

MÉRÉSI FELADAT – PROJEKT III.

Klasszifikációs feladathoz előfeldolgozásként használhatjuk a beágyazott gráfokat



Változó méretű, valós hálózatok 6 területről

| Domain | Description | Range of network size | Number of networks |
|-----------------|--|---------------------------|--------------------|
| Brain | Human and animal connectomes | 50 - 2,995 (avg: 946) | 100 |
| Cheminformatics | Protein-protein (enzyme) interaction networks | 44 - 125 (avg: 55) | 100 |
| Food | What-eats-what, consumer-resource networks | 19 - 1,500 (avg: 118) | 100 |
| Infrastructural | Transportation (metro, bus, road, airline) and distribution networks (power and water) | 39 - 40K (avg: 4,562) | 68 |
| Social | Facebook, Twitter and collaboration networks | 85 - 34K (avg: 5,183) | 118 |
| Web | Pieces of the World Wide Web | 146 - 16K (avg: 4,488) | 14 |

Molontay Roland, Nagy Marcell: 91.2% accuracy 8 dimenziós vektorokkal

■ **SF (Spectral Features):**

Normalizált Laplace-mátrix k legkisebb pozitív sajátértéke

Paraméterek: k

Dimenzió: k

■ **NetLSD (Network Laplacian Spectral Descriptor):**

Hőmennyiség terjedésének modellezése, megoldása Laplace-mátrix sajátértékeivel felírható; a gráf nyomát különböző időpontokban kiértékeljük

Paraméterek: sajátérték közelítés lépésszáma, kiértékelési időpontok eleje, vége, száma

Dimenzió: időpontok száma

■ **LDP (Local Degree Profile):**

Csúcsok szomszédainak fokszámstatisztikái hisztogramban

Paraméterek: binek száma

Dimenzió: $5 \times$ binek száma

■ Graph2Vec:

Csúcsok környezetei random fákkal mintavételezve, majd *doc2vec* beágyazás

"Használt" paraméterek: dimenzió, epoch-szám, learning rate

Dimenzió: dimenzió

■ FeatherGraph:

Adott hosszú séták végpontjainak eloszlása alapján minden csúcshoz karakterisztikus függvény, ezek kiértékelése, majd pooling

Paraméterek: sétahossz, kiértékelési intervallum, pontok száma, pooling

Dimenzió: $4 \times$ sétahossz \times kiértékelési pontok száma

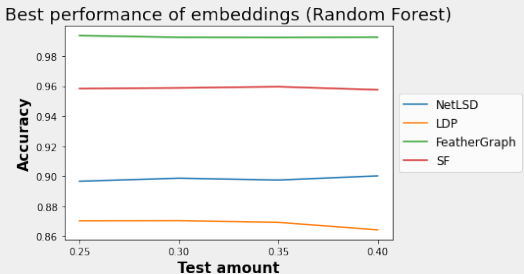
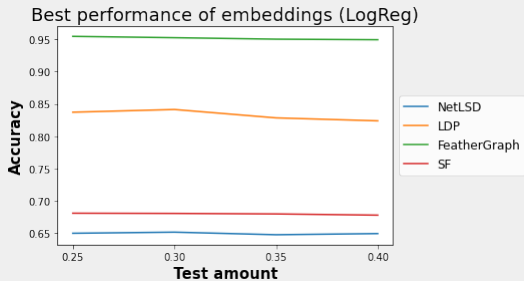
Logistic Regression *Newton CG* optimalizálással

- train-test: 0.25 – 0.4

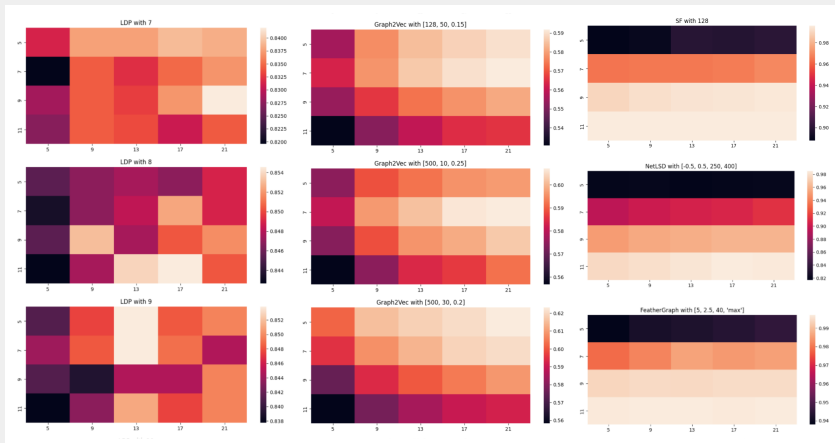
Random Forest Classifier

- 15 fa, max. 7 mélység, train-test: 0.25 – 0.4
- train-test: 0.3, (fa, mélység) $\in \{5, 9, 13, 17, 21\} \times \{5, 7, 9, 11\}$

EREDMÉNY – TRAIN-TEST VÁGÁS



EREDMÉNY – RANDOM FOREST PARAMÉTEREZÉS



KÖSZÖNÖM A FIGYELMET!

REFERENCES I



NETWORKX PACKAGE.

<https://networkx.org/documentation/stable/reference/generators.html>.



SCIKIT-LEARN PYTHON PACKAGE.

<https://scikit-learn.org/stable/index.html>.



**STRUCTURAL ANALYSIS OF REAL NETWORKS AND THEIR
MODEL-GENERATED COUNTERPARTS – SUPPLEMENTARY DATA.**

<https://github.com/marcessz/Complex-Networks>.






CHEN CAI AND YUSU WANG.




**A SIMPLE YET EFFECTIVE BASELINE FOR NON-ATTRIBUTE GRAPH
CLASSIFICATION.**

CoRR, abs/1811.03508, 2018.

REFERENCES II

-  HONG CHEN AND HISASHI KOGA.
GL2VEC: GRAPH EMBEDDING ENRICHED BY LINE GRAPHS WITH EDGE FEATURES.
In Tom Gedeon, Kok Wai Wong, and Minhoo Lee, editors, *Neural Information Processing - 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12-15, 2019, Proceedings, Part III*, volume 11955 of *Lecture Notes in Computer Science*, pages 3–14. Springer, 2019.
-  NATHAN DE LARA AND EDOUARD PINEAU.
A SIMPLE BASELINE ALGORITHM FOR GRAPH CLASSIFICATION.
CoRR, abs/1810.09155, 2018.
-  ALEXIS GALLAND AND MARC LELARGE.
INVARIANT EMBEDDING FOR GRAPH CLASSIFICATION.
In *ICML 2019 Workshop on Learning and Reasoning with Graph-Structured Data*, Long Beach, United States, June 2019.



REFERENCES III

-  FENG GAO, GUY WOLF, AND MATTHEW HIRN.
GEOMETRIC SCATTERING FOR GRAPH DATA ANALYSIS.
In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors,
Proceedings of the 36th International Conference on Machine Learning, volume 97 of *Proceedings of Machine Learning Research*, pages 2122–2131. PMLR, 09–15 Jun 2019.
-  ADITYA GROVER AND JURE LESKOVEC.
NODE2VEC: SCALABLE FEATURE LEARNING FOR NETWORKS.
CoRR, abs/1607.00653, 2016.
-  MARCELL NAGY AND ROLAND MOLONTAY.
NETWORK CLASSIFICATION-BASED STRUCTURAL ANALYSIS OF REAL NETWORKS AND THEIR MODEL-GENERATED COUNTERPARTS.
Network Science, pages 1–24, 2022.

REFERENCES IV

-  ANNAMALAI NARAYANAN, MAHINTHAN CHANDRAMOHAN, RAJASEKAR VENKATESAN, LIHUI CHEN, YANG LIU, AND SHANTANU JAISWAL.
GRAPH2VEC: LEARNING DISTRIBUTED REPRESENTATIONS OF GRAPHS.
CoRR, abs/1707.05005, 2017.
-  BRYAN PEROZZI, RAMI AL-RFOU, AND STEVEN SKIENA.
DEEPWALK: ONLINE LEARNING OF SOCIAL REPRESENTATIONS.
CoRR, abs/1403.6652, 2014.
-  BENEDEK ROZEMBERCZKI, OLIVER KISS, AND RIK SARKAR.
AN API ORIENTED OPEN-SOURCE PYTHON FRAMEWORK FOR UNSUPERVISED LEARNING ON GRAPHS.
CoRR, abs/2003.04819, 2020.
-  BENEDEK ROZEMBERCZKI AND RIK SARKAR.
CHARACTERISTIC FUNCTIONS ON GRAPHS: BIRDS OF A FEATHER, FROM STATISTICAL DESCRIPTORS TO PARAMETRIC MODELS, 2020.

REFERENCES V

-  **ANTON TSITSULIN, DAVIDE MOTTIN, PANAGIOTIS KARRAS, ALEXANDER M. BRONSTEIN, AND EMMANUEL MÜLLER.**
NETLSD: HEARING THE SHAPE OF A GRAPH.
CoRR, abs/1805.10712, 2018.
-  SAURABH VERMA AND ZHI-LI ZHANG.
HUNT FOR THE UNIQUE, STABLE, SPARSE AND FAST FEATURE LEARNING ON GRAPHS.
In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.