

# GRÁFBEÁGYAZÓ (GRAPH EMBEDDING) ALGORITMUSOK PERFORMANCIÁJA

SZAKÁCS LILI KATA

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

ÖNÁLLÓ PROJEKT II., 2021/22 II. FÉLÉV

BUDAPEST, 2022. 05. 19.

*Cél:* Gráfok  $\rightarrow$  alacsonydimenziós euklideszi vektorok

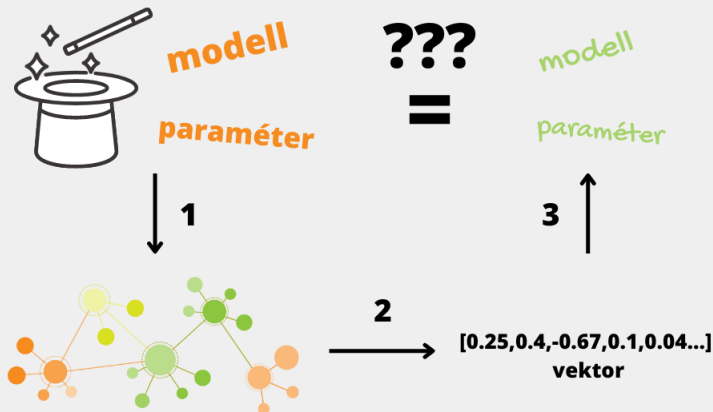
- strukturálisan hasonló gráfokhoz közeli vektorok
- csúcsok permutációjára invariáns

Két gyakori megközelítés:

- Spektrális tulajdonságok alapján
- Természetes nyelvfeldolgozással (NLP) analóg módon, csúcsok környezetével

# MÉRÉSI FELADAT

Klasszifikációs feladathoz előfeldolgozásként használhatjuk a beágyazott gráfokat



- 1) 5 különböző modellel generált random gráfok
- 2) 8 gráfbeágyazó algoritmus:
  - ▶ **NetLSD, FeatherGraph, FGSD, LDP, Graph2Vec, GL2Vec, GeoScattering, SF**
  - ▶ *KarateClub package*, alapértelmezett beállítások
- 3) 6 klasszifikációs feladat modellekre, paraméterekre
  - ▶ logisztikus regresszió *Newton CG* optimalizálóval
  - ▶ train-test: 0.75 - 0.25

Kb. 450 gráf 1000-1100 csúccsal

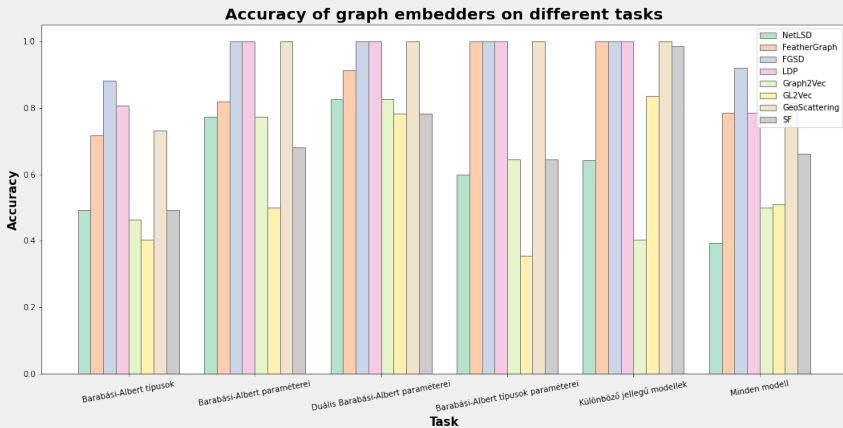
Ötféle modell:

- Barabási-Albert-modell
- Duális Barabási-Albert-modell
- Holme & Kim-algoritmus
- Erdős-Rényi-modell
- Random reguláris gráf

90-90 adott modellből → 10 vagy 20 adott paraméterrel

- **Barabási-Albert típusok:** a Barabási-Albert, duális Barabási-Albert és Holme-Kim modellek  
*3 kategória, bennük kb. 70 gráf a tanító halmazban és 20-22 gráf a teszt halmazban*
- **Különböző jellegű modellek:** Barabási-Albert, Erdős-Rényi és random reguláris gráf modellek  
*3 kategória, bennük kb. 70 gráf a tanító halmazban és 20-22 gráf a teszt halmazban*
- **Minden modell:** mind az öt használt modell  
*5 kategória, bennük kb. 70 gráf a tanító halmazban és 20-22 gráf a teszt halmazban*

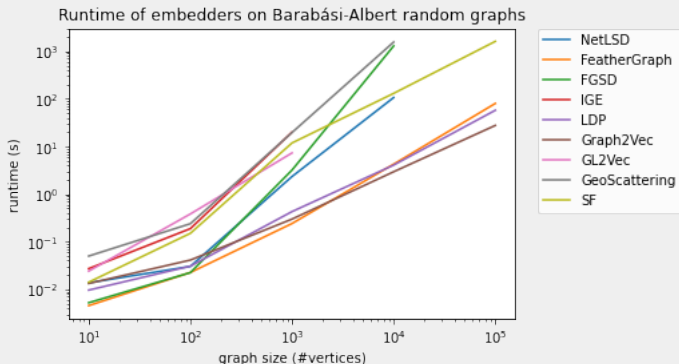
- **Barabási-Albert paramétere:** Barabási-Albert modell paramétere  
*6 kategória, bennük kb. 10-15 gráf a tanító halmazban és 3-5 gráf a teszt halmazban*
- **Duális Barabási-Albert-modell paramétere:** duális Barabási-Albert modell paramétere  
*5 kategória, bennük kb. 14-17 gráf a tanító halmazban és 4-5 gráf a teszt halmazban*
- **Barabási-Albert típusok paramétere:** a Barabási-Albert és duális Barabási-Albert modell paramétere  
*6 kategória, bennük kb. 25-35 gráf a tanító halmazban és 5-10 gráf a teszt halmazban*



Legjobbak: *FGSD, GeoScattering, LDP, FeatherGraph*







# EREDMÉNY ÉRTÉKELÉS



Sajnos az *FGSD* és a *GeoScattering* lassú...  
*LDP* egyszerű és gyors módszer, meglepően jó!




KÖSZÖNÖM A FIGYELMET!

# REFERENCES I





-  NETWORKX PACKAGE.  
<https://networkx.org/documentation/stable/reference/generators.html>.
-  SCIKIT-LEARN PACKAGE.  
<https://scikit-learn.org/stable/index.html>.
-  CHEN CAI AND YUSU WANG.  
**A SIMPLE YET EFFECTIVE BASELINE FOR NON-ATTRIBUTE GRAPH CLASSIFICATION.**  
*CoRR*, abs/1811.03508, 2018.
-  HONG CHEN AND HISASHI KOGA.  
**GL2VEC: GRAPH EMBEDDING ENRICHED BY LINE GRAPHS WITH EDGE FEATURES.**  
In Tom Gedeon, Kok Wai Wong, and Minho Lee, editors, *Neural Information Processing - 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12-15, 2019, Proceedings*,

## REFERENCES II




*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.
-  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.

## REFERENCES III

-  ADITYA GROVER AND JURE LESKOVEC.  
**NODE2VEC: SCALABLE FEATURE LEARNING FOR NETWORKS.**  
*CoRR*, abs/1607.00653, 2016.
-  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.

## REFERENCES IV

-  **BENEDEK ROZEMBERCZKI AND RIK SARKAR.**  
**CHARACTERISTIC FUNCTIONS ON GRAPHS: BIRDS OF A FEATHER, FROM STATISTICAL DESCRIPTORS TO PARAMETRIC MODELS, 2020.**
-  **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.