

# GRÁFBEÁGYAZÓ (GRAPH EMBEDDING) ALGORITMUSOK IDŐHATÉKONYSÁGA

SZAKÁCS LILI KATA

TÉMAVEZETŐ: BÉRES FERENC, BENCZÚR ANDRÁS  
SZTAKI, INFORMATIKAI KUTATÓLABORATÓRIUM

ÖNÁLLÓ PROJEKT I., 2021/22 I. FÉLÉV

BUDAPEST, 2021. 12. 16.



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

- strukturálisan hasonló gráfokhoz közeli vektorok
- csúcsok permutációjára invariáns

Két megközelítés:

- Laplace-mátrix spektrális tulajdonságai
- Természetes nyelvfeldolgozással (NLP) analóg módon, csúcsok környezetével

- *Definíció:*  $G$  gráf Laplace-mátrixa  $L(G) = D(G) - A(G)$ , ahol  $D(G) = \text{diag}(d_i)$  és  $A(G)$  az adjacenciamátrix.
- *SF, IGE:* normalizált Laplace-mátrix sajátértékei közül a legkisebb pozitívokat használja vektorként
- *FGSD:* Laplace-mátrix sajátértékeivel definiál egy metrikát, amivel izometrikusan beágyaz
- *NetLSD:* Hőmennyiség terjedésének modellezése, aminek a megoldása Laplace-mátrix sajátértékeivel felírható

Szavak és környezetük egy szövegben  $\longleftrightarrow$  Csúcsok és környezetük egy gráfban

Csúcsok környezete  $\approx$  Random sétákkal mintavételezett szomszédság

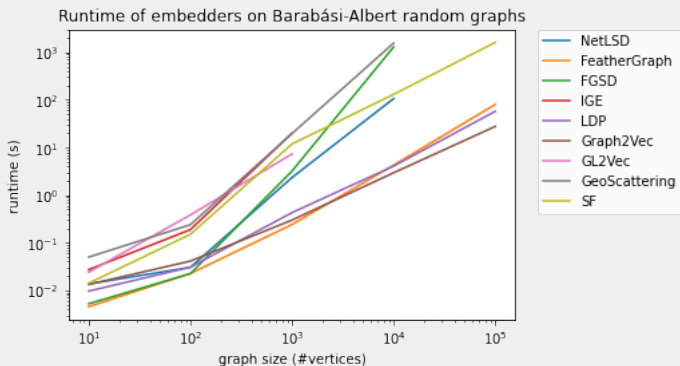
Használhatók a *NLP* eredményei (*word2vec*, *doc2vec*), ha a gráf tulajdonságaiból "környezeteket csináltunk"  
– ezek motorjai is szavak szövegbeli környezete alapján ágyaznak be szavakat/szövegrészleteket/dokumentumokat

- *Graph2Vec*, *GL2Vec*: csúcsok környezeteit random sétával mintavételezi, majd *doc2vec* beágyazás (*GL2Vec*-nél élgráfból is)
- *GeoScattering*: Csúcsokon vett jelfüggvény Wavelet transzformálása környezetek alapján
- *FeatherGraph*: Környezetek alapján minden csúcshoz karakterisztikus függvény, ezek kiértékelése, majd mean pooling

*Barabási-Albert-modell* által generált gráfok (1-4 új él, 50-50db nagyságrendenként) – valósághű a skálafüggetlenség miatt

$10^1$  –  $10^5$ -ig nagyságrendek

*KarateClub* csomagban lévő embedding eljárások (9db)



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

Nagyobb nagyságrendekre is próba futtatások

Performanciavizsgálat klaszterezési feladatoknál



KÖSZÖNÖM A FIGYELMET!

# REFERENCES I



NETWORKX PACKAGE.

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



**CHEN CAI AND YUSU WANG.**

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

*CoRR*, abs/1811.03508, 2018.



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 II

-  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.
-  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.

## REFERENCES III

-  **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.**
-  **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.