

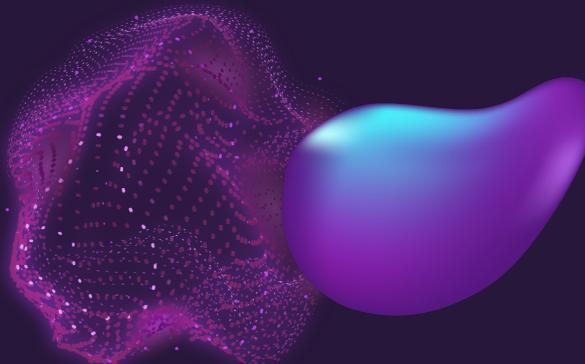
PageRank in Python

NAML Project

March 13, 2023,
Politecnico of Milan



x

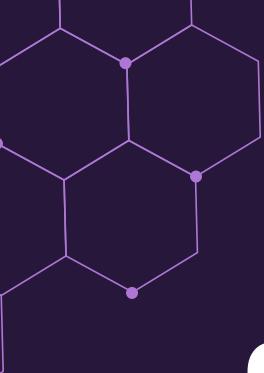


x



x





x

TABLE OF CONTENTS



x

01 THEORY

Graphs, History, Method

02 GRAPH ADTs

O.O., Data structures

03 ALGORITHM

Code, optimization

04 PERFORMANCE

Dataset, time execution

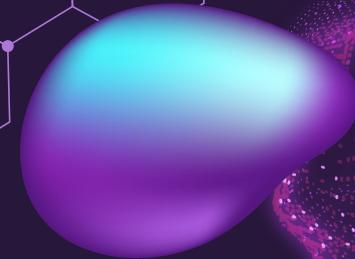
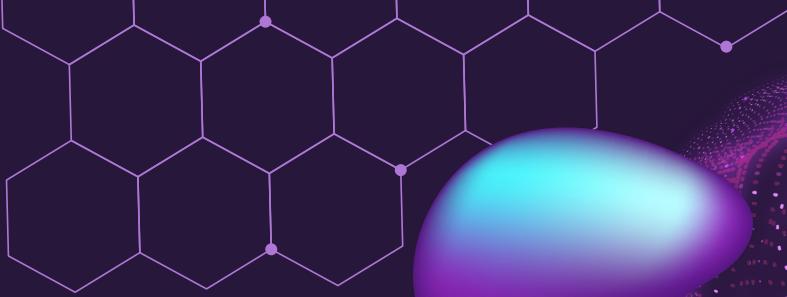
01

THEORY

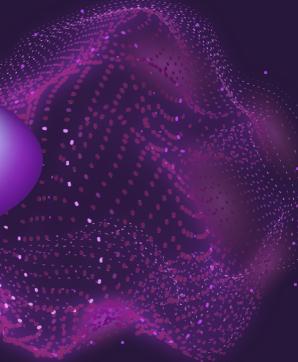
Graphs, History, Method



*



*



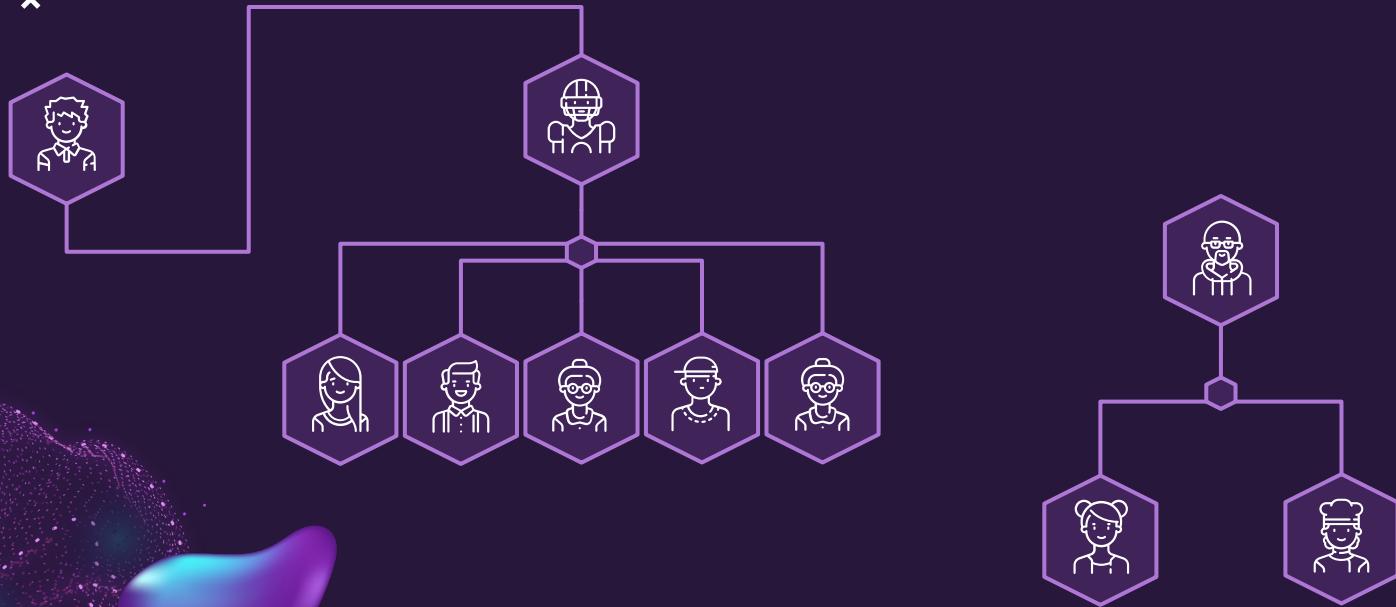
*



THEORY

In 1998, Larry Page and Sergey Brin introduced PageRank.
At the bases of Google search engine.
It can approximate the “importance” of any given node in a graph structure.

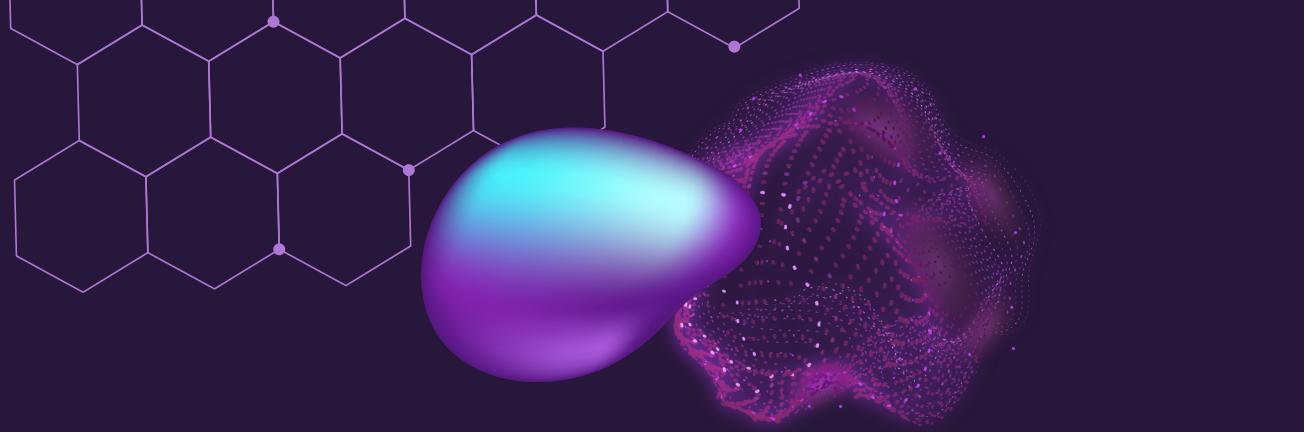
Let's consider an example



02

GRAPH ADTs

O.O., Data structures



ADTs

CLASS	DESCRIPTION
GraphError	Error handling
Node	Base element of the graph
Edge	Relationship between two nodes
BaseGraph	General purpose graph implementation
DirectedGraph	Graph with directed edges
UndirectedGraph	Graph with undirected edges

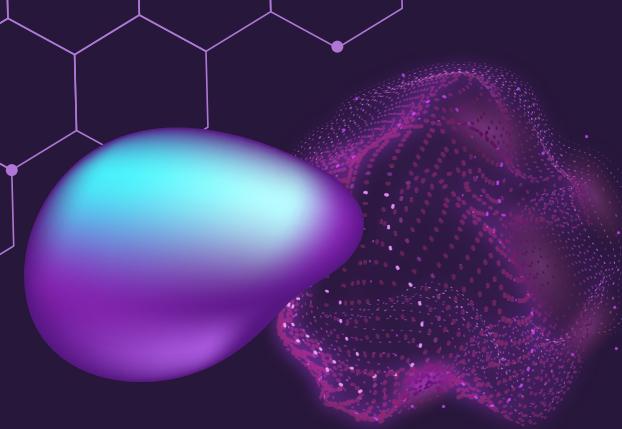
03

ALGORITHM

Code, optimization



x



x

+



UPDATE RULE

It is applied until convergence.
The pagerank score is an out-of-place computation.

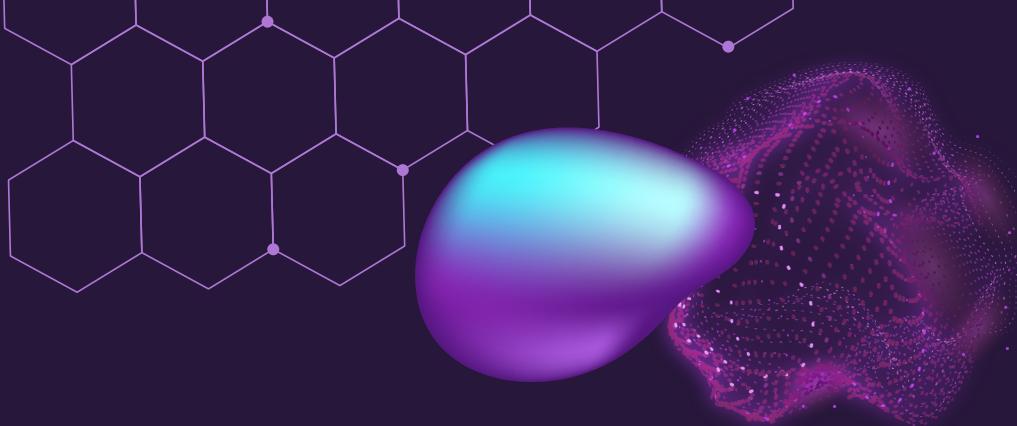
$$PR_{k+1}(u) = \frac{1-d}{N} + d \left(\sum_{v \in BL(u)} \frac{PR_k(v)}{v^+} + \sum_{w \text{ is sink}} \frac{PR_k(w)}{N} \right)$$



04

PERFORMANCE

Code, Dataset, time execution



x

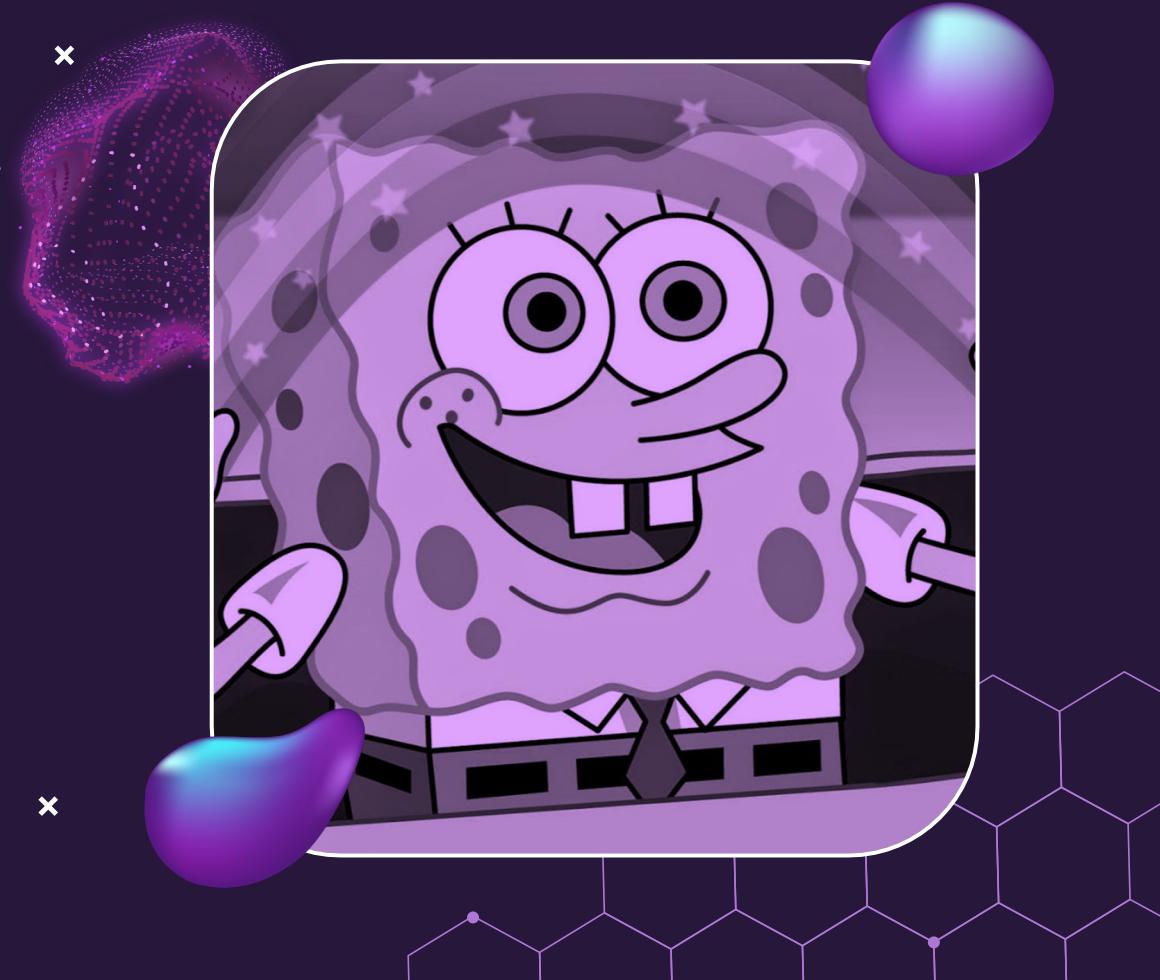


+



SMALL FIRST DATASET

It's a toy use case to check the correctness of the algorithm.
There are 9 nodes and 8 edges.

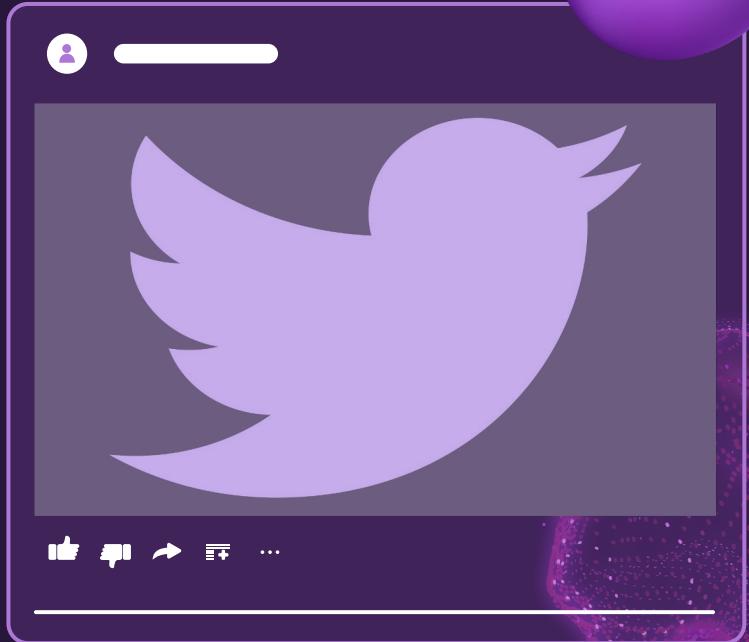


EMAIL DATASET



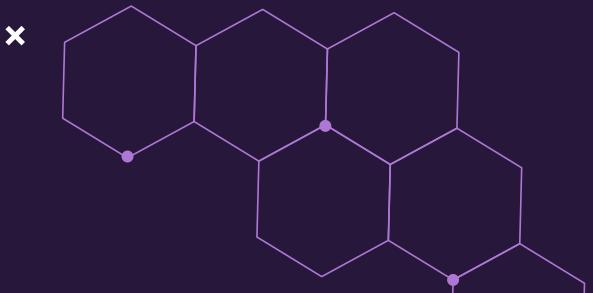
MEDIUM DATASET

This dataset contains 1006 nodes and 25572 edges. Each sample represents the connection between 2 users.



HUGE DATASET

This dataset contains 81307 nodes and 1768150 edges.
The alg. takes 2'30" to evaluate it.





THANKS!

DO YOU HAVE ANY QUESTIONS?

matteo3.negro@mail.polimi.it