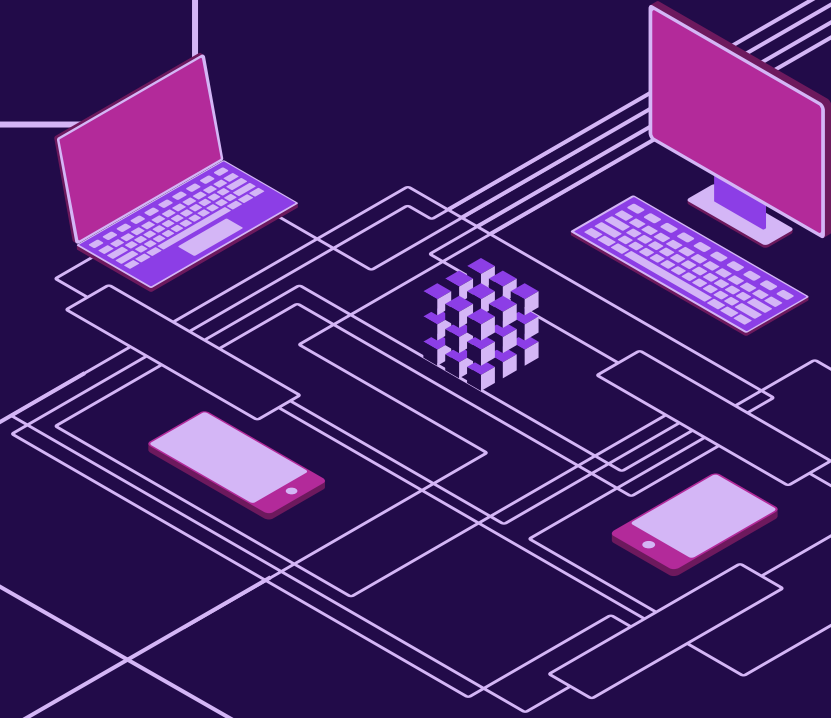


Proyecto 3

Neo4J

Ana Maria Garcia
Andres Delgado
Katherine Camacho
Leonardo Saez Wilches



CONTENIDOS

01

Descripción de la Colección

Grafo creado en Neo4J con sus relaciones

02

Medidas Topológicas y Comparación

Implmentación de 5 medidas y su desempeño

03

Mejor Medida Para la Predicción

Medida más útil para la predicción

04

Comparación Con y Sin Medidas Topológicas

Tabla Comparativa

05

Conclusiones

Conclusiones de los 3 proyectos del curso

DATOS AVILA

Los registros del dataset representan las páginas de la biblia Avila

Atributos



10 Atributos que representan las características de las páginas(F1-F10)



Un atributo clasificador "Authornum"

0



:



11





01

Descripción de la Colección

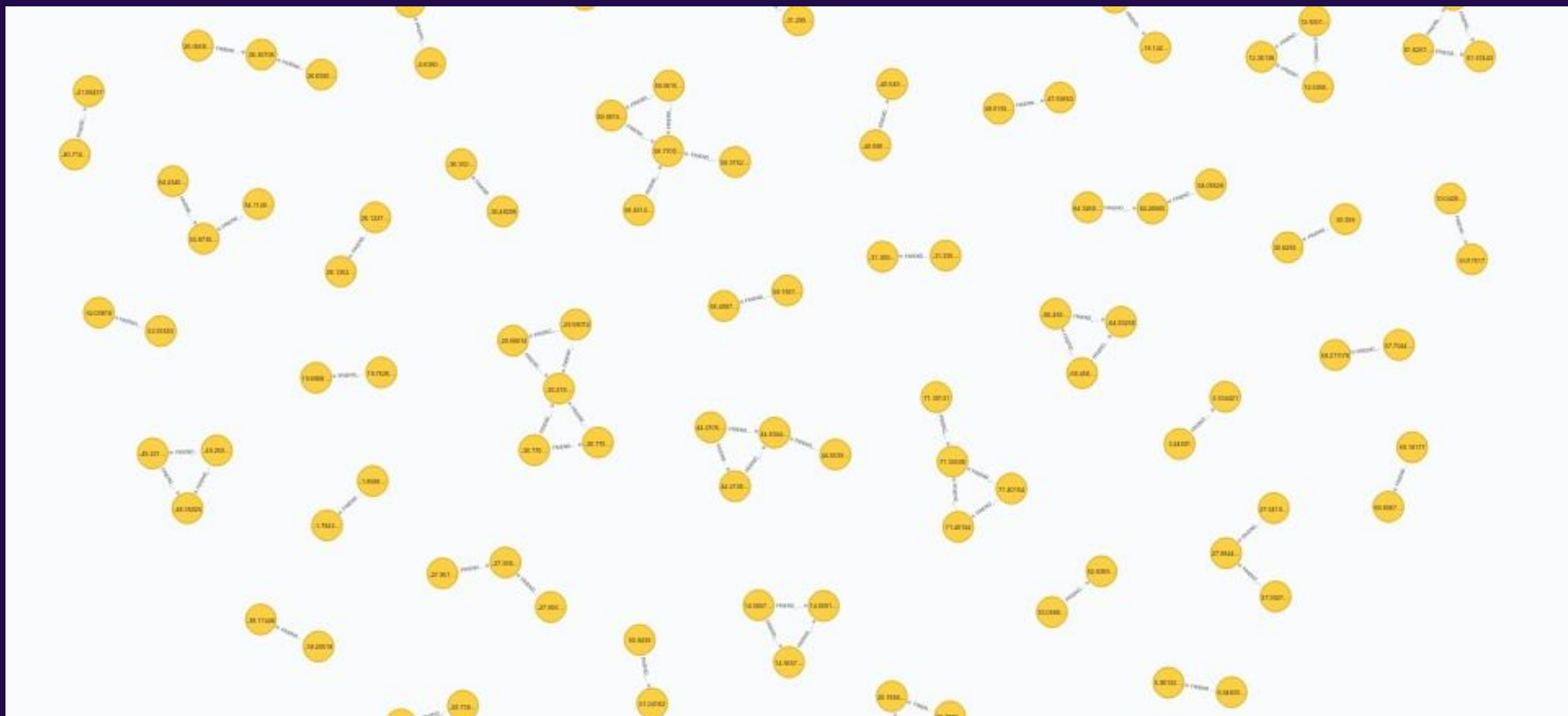
PROCESO

1. Preprocesamiento del conjunto
2. Reduccion de dimensionalidades(Sklearn.manifold.TSNE)
3. KDTree(Sklearn.neighbors.KDTree)
- 4 relaciones
5. Creación del csv de nodos y relaciones
6. Importacion a Neo4j

Grafo Final

Grafo de 25578 nodos y 53402 relaciones. Cada nodo está compuesto por dos atributos

Grafo Neo4j



Grafo Neo4j

Csv Nodos

idNodo	caracteristica1	caracteristica2
0	8.562754	50.605396
1	-10.000735	-3.0800116
2	76.77591	32.53666
3	-17.947903	-23.827896
4	-58.989773	50.16818

Csv Relaciones

idRelacion	idNodo1	idNodo2
0	0	9782
1	0	16827
2	1	9783
3	1	13017
4	2	11454



02

Medidas Topológicas y Comparación

Medidas Topológicas

01

Triangle Counting / Clustering Coefficient

Accuracy Random Forest:
0.638953040801

02

PageRank

Accuracy Random Forest:
0.643956889915

03

Betweenness Centrality

Accuracy Random Forest:
0.652040030793

04

Louvain

Accuracy Random Forest:
0.858352578907

05

Strongly Connected Components

Accuracy Random Forest:
0.957467282525



03

**Mejor Medida Para
la Predicción**

Mejor Medida Topológica

La medida más útil para la predicción de los autores fue **Strongly Connected Components**, pues esta medida fue la que tuvo el mejor desempeño al implementar el modelo de Random Forest, escogido por ser el modelo con mejor desempeño del proyecto 1.



04

**Comparación Con y Sin
Medidas Topológicas**

```
graph TD; A[ ] --- B[SIN MEDIDAS]; A --- C[CON MEDIDAS];
```

SIN MEDIDAS

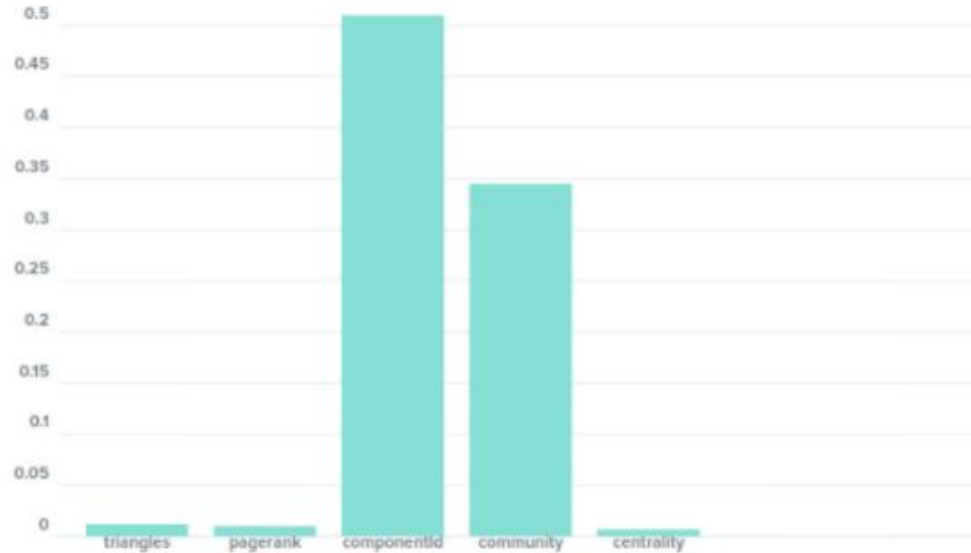
Accuracy
Random Forest:
0.8914

CON MEDIDAS

Accuracy Random
Forest:
0.9706

Medidas Topológicas

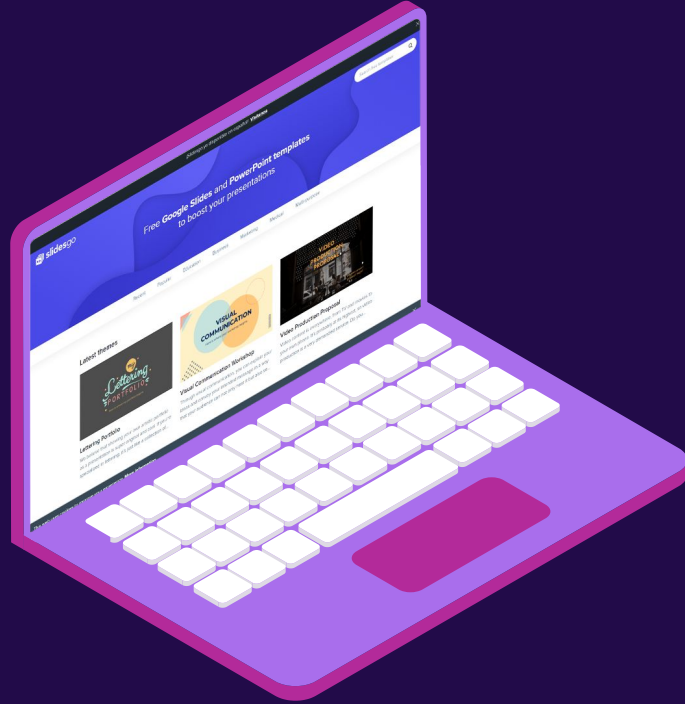
Aporte Medidas Topologicas





05

Conclusiones



Conclusiones

- Conclusiones Proyecto 1
- Conclusiones Proyecto 2
- Conclusiones Proyecto 3

MUCHAS GRACIAS

