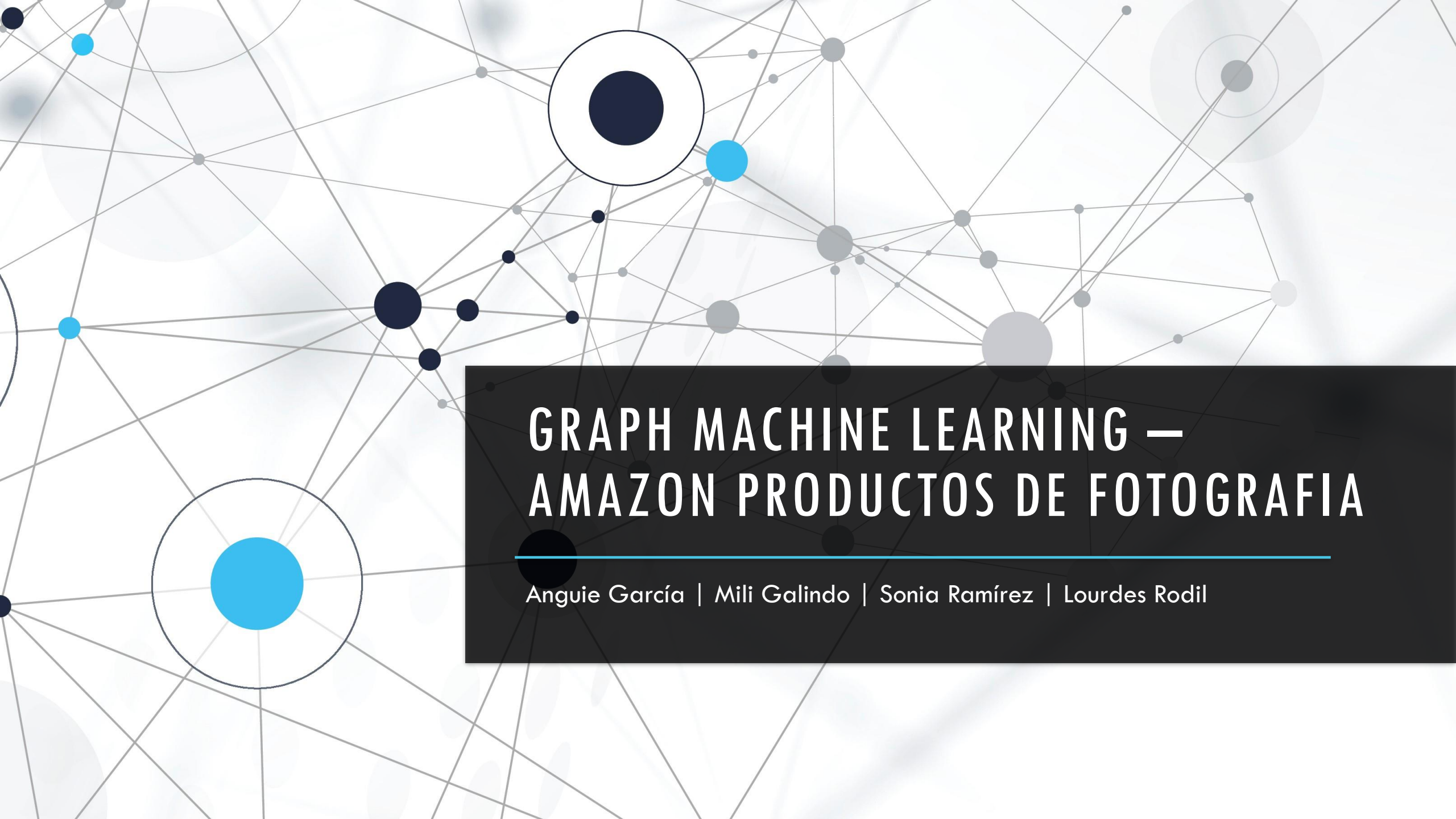


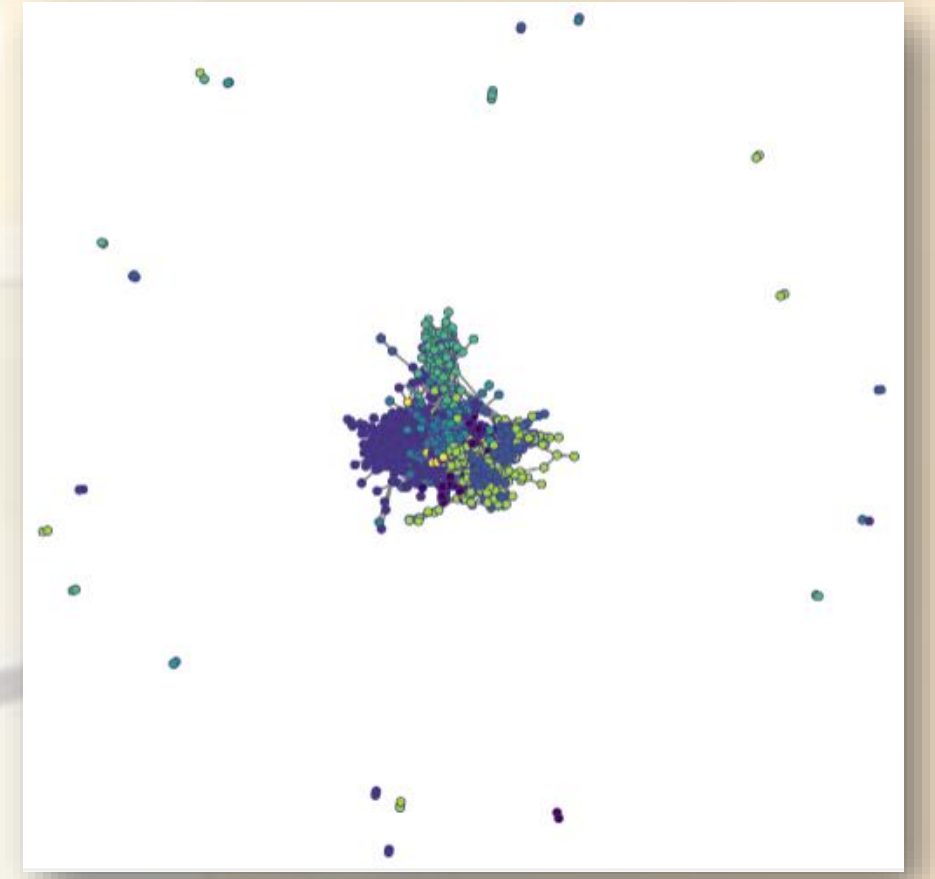
GRAPH MACHINE LEARNING – AMAZON PRODUCTOS DE FOTOGRAFIA

Anguie García | Mili Galindo | Sonia Ramírez | Lourdes Rodil

GRAPH MACHINE LEARNING - AMAZON PRODUCTOS DE FOTOGRAFÍA

Objetivo Clasificar los productos de Amazon en una de las ocho (8) categorías en el segmento de fotografía a través de modelos de grafos.

Dataset Amazon: Las redes de Amazon Computers provienen del artículo "Pitfalls of Graph Neural Network Evaluation" y forma parte de la libreria **Pythorch**. Los nodos representan productos y los enlaces indican que dos productos se compran frecuentemente juntos.



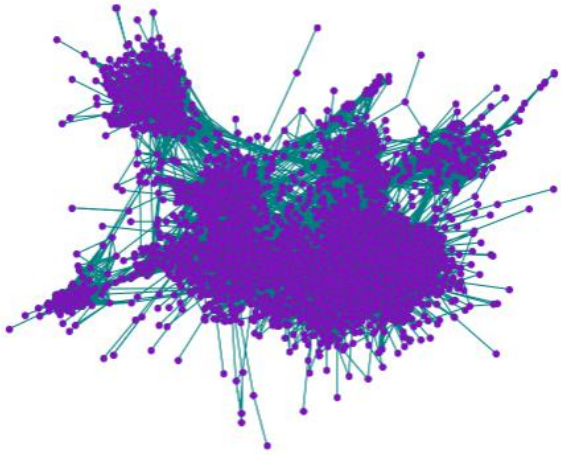
Fuente: https://pytorch-geometric.readthedocs.io/en/2.5.0/generated/torch_geometric.datasets.Amazon.html
Oleksandr Shchur, Pitfalls of Graph Neural Network Evaluation. Technical University of Munich, Germany.

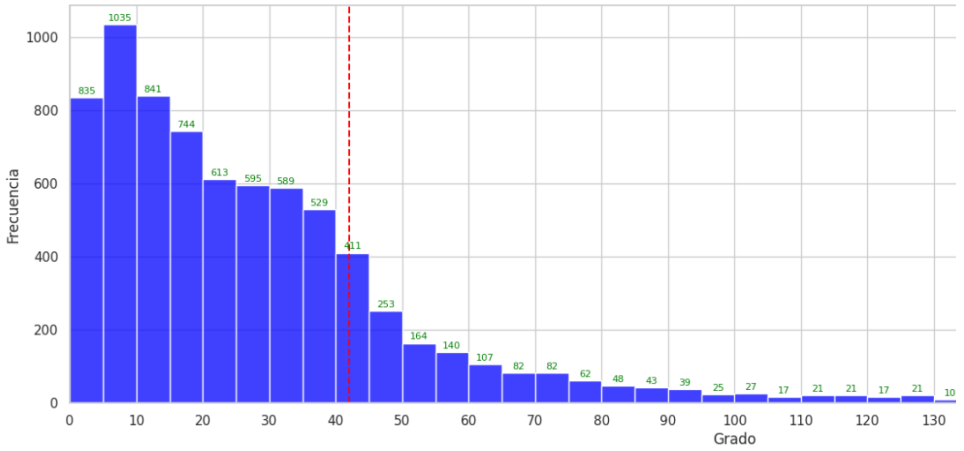
MÉTRICAS DEL GRAFO

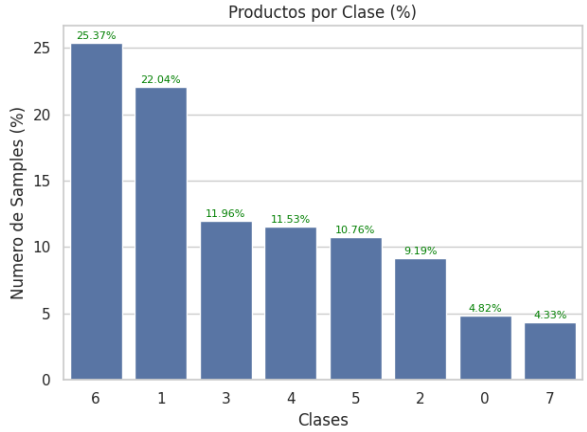
Número de Nodos: 7.535
Número de edges: 238.162
Número de clases: 8
Número de features: 745

Media de conexiones de los nodos: 31,6
Densidad: 0,004
Transitividad: 0,177

Clase 6: 25,37%
Clase 7: 4,33%







MODELOS

1 → Perceptrón Multicapa (MLP) → MLP con tunning

2 → GCN (Graph Convolutional Network) → GCN con tunning

3 → GATv2Conv → GATv2Conv con tunning

4 → GATvConv → GATvConv con tunning

5 → GraphSAGE → GraphSAGE con tunning & Early Stopping

PERCEPTRÓN MULTICAPA (MLP)

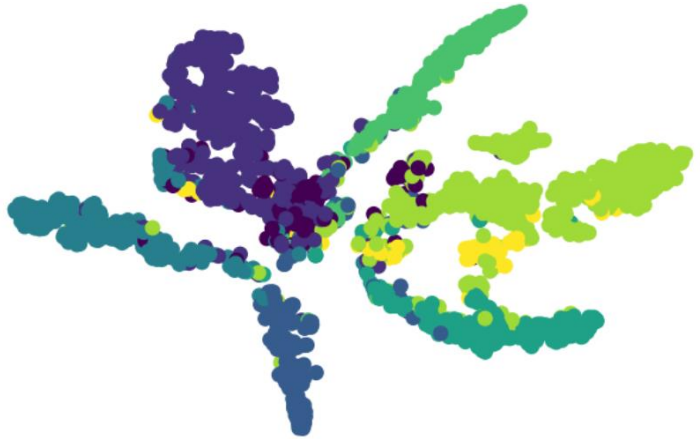
- **Setup:**

2 # capas	0.01 LR	5e-4 Weight decay
100 Épocas	16 Hidden channels	0,5 Drop out

- **Training:**

```
Epoch   0 | Train Loss: 2.243 | Train Acc:   6.69% | Val Loss: 2.23 | Val Acc: 7.21%
Epoch  10 | Train Loss: 1.626 | Train Acc:  24.29% | Val Loss: 1.65 | Val Acc: 24.67%
Epoch  20 | Train Loss: 1.455 | Train Acc:  25.45% | Val Loss: 1.50 | Val Acc: 24.67%
Epoch  30 | Train Loss: 1.368 | Train Acc:  25.45% | Val Loss: 1.45 | Val Acc: 24.67%
Epoch  40 | Train Loss: 1.333 | Train Acc:  33.45% | Val Loss: 1.42 | Val Acc: 30.17%
Epoch  50 | Train Loss: 1.312 | Train Acc:  34.24% | Val Loss: 1.37 | Val Acc: 31.31%
Epoch  60 | Train Loss: 1.284 | Train Acc:  35.27% | Val Loss: 1.36 | Val Acc: 33.02%
Epoch  70 | Train Loss: 1.290 | Train Acc:  35.38% | Val Loss: 1.32 | Val Acc: 33.78%
Epoch  80 | Train Loss: 1.276 | Train Acc:  35.67% | Val Loss: 1.35 | Val Acc: 33.40%
Epoch  90 | Train Loss: 1.233 | Train Acc:  38.34% | Val Loss: 1.28 | Val Acc: 36.81%
Epoch 100 | Train Loss: 1.212 | Train Acc:  39.21% | Val Loss: 1.26 | Val Acc: 38.52%
```

- **Results:**



43,5%

Grid Search

- Hidden channels: [8, 16, **32**]
- LR: [0.001, **0.01**, 0.1]
- Weight decay: [1e-4, 5e-4, **1e-3**]

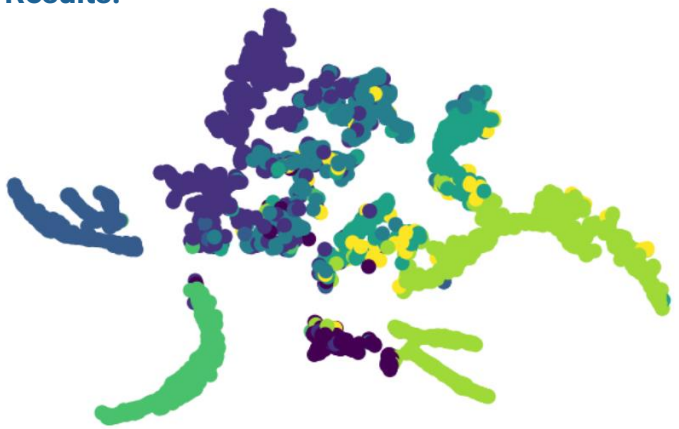
85,8%

GCN (GRAPH CONVOLUTIONAL NETWORK)

- **Setup:**

5 # capas	ReLU Función Activación	0.01 LR
100 Épocas	5e-4 Weight decay	

- **Results:**



- **Training:**

```
Epoch   0 | Train Loss: 2.199 | Train Acc:   8.91% | Val Loss: 2.20 | Val Acc: 10.44%
Epoch  10 | Train Loss: 1.765 | Train Acc:  25.45% | Val Loss: 1.74 | Val Acc: 24.67%
Epoch  20 | Train Loss: 1.439 | Train Acc:  25.45% | Val Loss: 1.39 | Val Acc: 24.67%
Epoch  30 | Train Loss: 1.200 | Train Acc:  35.93% | Val Loss: 1.17 | Val Acc: 33.78%
Epoch  40 | Train Loss: 1.040 | Train Acc:  49.92% | Val Loss: 1.03 | Val Acc: 48.77%
Epoch  50 | Train Loss: 1.000 | Train Acc:  57.15% | Val Loss: 0.99 | Val Acc: 58.63%
Epoch  60 | Train Loss: 0.838 | Train Acc:  64.79% | Val Loss: 0.85 | Val Acc: 64.90%
Epoch  70 | Train Loss: 0.788 | Train Acc:  65.76% | Val Loss: 0.80 | Val Acc: 66.60%
Epoch  80 | Train Loss: 0.616 | Train Acc:  74.27% | Val Loss: 0.62 | Val Acc: 77.04%
Epoch  90 | Train Loss: 0.515 | Train Acc:  82.40% | Val Loss: 0.53 | Val Acc: 82.54%
Epoch 100 | Train Loss: 0.431 | Train Acc:  86.41% | Val Loss: 0.45 | Val Acc: 86.53%
```

Grid Search

- LR: [**0,01**, 0,001]
- Weight decay: [**5e-4**, 1e-4]

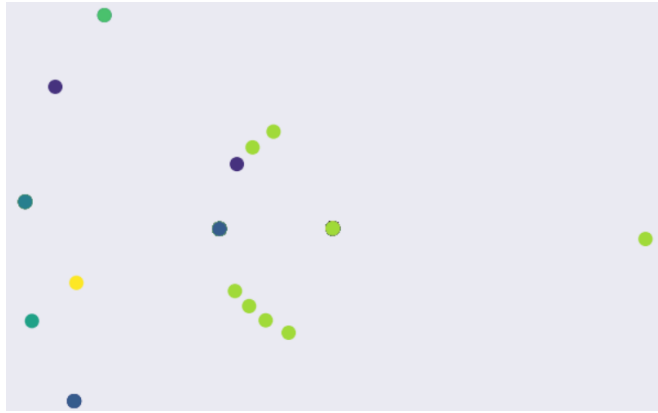
87%

MODELO GATVCONV

- **Setup:**

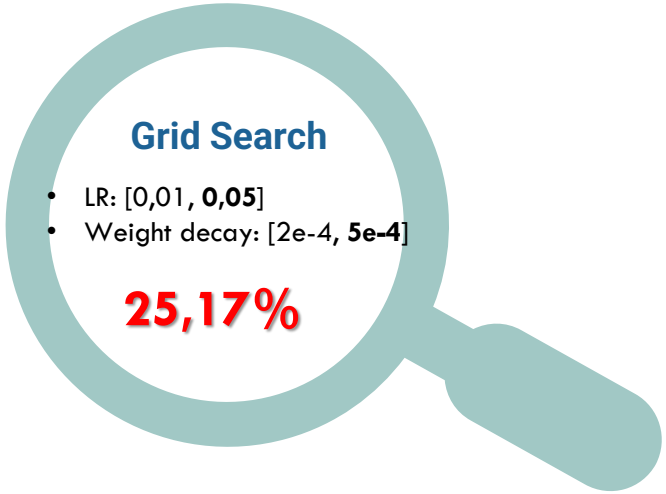
5 # capas	0.01 LR	5e-4 Weight decay	60 Épocas
100 Épocas			

- **Results:**



- **Training:**

```
Epoch   0 | Train Loss: 2.143 | Train Acc:  11.74% | Val Loss: 2.14 | Val Acc: 11.20%
Epoch  10 | Train Loss: 2.007 | Train Acc:   8.97% | Val Loss: 2.00 | Val Acc: 7.02%
Epoch  20 | Train Loss: 2.009 | Train Acc:  25.45% | Val Loss: 2.01 | Val Acc: 24.67%
Epoch  30 | Train Loss: 1.964 | Train Acc:  25.45% | Val Loss: 1.97 | Val Acc: 24.67%
Epoch  40 | Train Loss: 1.921 | Train Acc:  25.45% | Val Loss: 1.92 | Val Acc: 24.67%
Epoch  50 | Train Loss: 1.882 | Train Acc:  25.45% | Val Loss: 1.88 | Val Acc: 24.67%
Epoch  60 | Train Loss: 1.849 | Train Acc:  25.45% | Val Loss: 1.85 | Val Acc: 24.67%
Epoch  70 | Train Loss: 1.820 | Train Acc:  25.45% | Val Loss: 1.82 | Val Acc: 24.67%
Epoch  80 | Train Loss: 1.797 | Train Acc:  25.45% | Val Loss: 1.80 | Val Acc: 24.67%
Epoch  90 | Train Loss: 1.777 | Train Acc:  25.45% | Val Loss: 1.78 | Val Acc: 24.67%
Epoch 100 | Train Loss: 1.762 | Train Acc:  25.45% | Val Loss: 1.76 | Val Acc: 24.67%
```



MODELO GATV2CONV

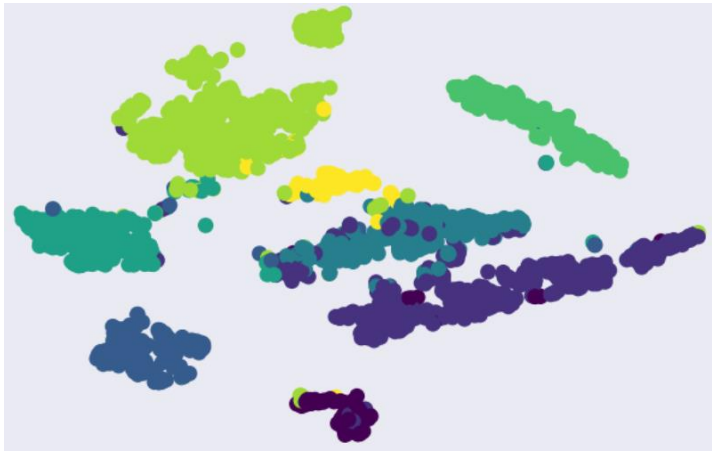
- **Setup:**

2 # capas	0.01 LR	5e-4 Weight decay
100 Épocas	8 heads	0,5 Drop out
	16 Hidden channels	

- **Training:**

```
Epoch   0 | Train Loss: 2.342 | Train Acc:  11.87% | Val Loss: 2.33 | Val Acc: 12.52%
Epoch  10 | Train Loss: 1.495 | Train Acc:  31.55% | Val Loss: 1.48 | Val Acc: 31.12%
Epoch  20 | Train Loss: 0.826 | Train Acc:  67.37% | Val Loss: 0.88 | Val Acc: 65.28%
Epoch  30 | Train Loss: 0.477 | Train Acc:  82.46% | Val Loss: 0.50 | Val Acc: 83.68%
Epoch  40 | Train Loss: 0.331 | Train Acc:  86.61% | Val Loss: 0.37 | Val Acc: 87.29%
Epoch  50 | Train Loss: 0.271 | Train Acc:  90.71% | Val Loss: 0.35 | Val Acc: 89.37%
Epoch  60 | Train Loss: 0.206 | Train Acc:  93.48% | Val Loss: 0.29 | Val Acc: 92.79%
Epoch  70 | Train Loss: 0.187 | Train Acc:  94.05% | Val Loss: 0.28 | Val Acc: 93.17%
Epoch  80 | Train Loss: 0.158 | Train Acc:  94.50% | Val Loss: 0.26 | Val Acc: 92.41%
Epoch  90 | Train Loss: 0.134 | Train Acc:  95.68% | Val Loss: 0.25 | Val Acc: 94.31%
Epoch 100 | Train Loss: 0.135 | Train Acc:  95.68% | Val Loss: 0.20 | Val Acc: 93.74%
```

- **Results:**



94,6%

Grid Search

- Dim_hidden: [8, **16**]
- Heads: [**4**, 8]
- LR: [**0,01**, 0,05]
- Weight decay: [2e-4, **5e-4**]

94,96%

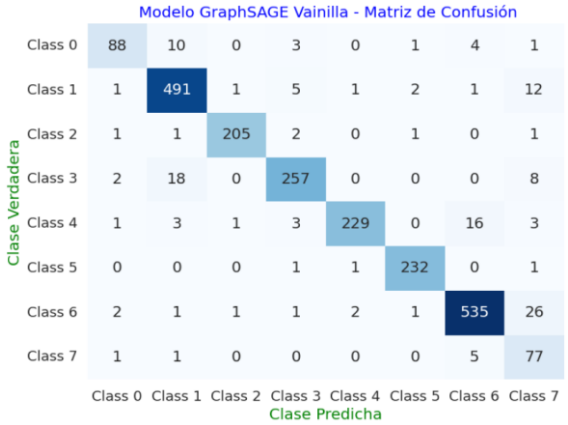
MODELO GRAPHSAGE

- **Setup:**

2 # capas	0.01 LR	5e-4 Weight decay
4 # Subgrafos	0,1 Drop out	

- **Results:**

- **Matriz de confusión**



95,22%

- **Training:**

```
GraphSAGE(
  (sage1): SAGEConv(745, 64, aggr=mean)
  (sage2): SAGEConv(64, 8, aggr=mean)
)
Epoch   0 | Train Loss: 3.199 | Train Acc:  28.31% | Val Loss: 2.92 | Val Acc: 31.26%
Epoch  10 | Train Loss: 0.527 | Train Acc:  80.44% | Val Loss: 0.63 | Val Acc: 79.12%
Epoch  20 | Train Loss: 0.305 | Train Acc:  88.09% | Val Loss: 0.49 | Val Acc: 85.94%
Epoch  30 | Train Loss: 0.172 | Train Acc:  94.37% | Val Loss: 0.41 | Val Acc: 90.26%
Epoch  40 | Train Loss: 0.112 | Train Acc:  96.82% | Val Loss: 0.41 | Val Acc: 91.45%
Epoch  50 | Train Loss: 0.083 | Train Acc:  97.59% | Val Loss: 0.42 | Val Acc: 91.61%
Epoch  60 | Train Loss: 0.071 | Train Acc:  97.94% | Val Loss: 0.44 | Val Acc: 91.57%

GraphSAGE test accuracy: 95.22%
```

Grid Search

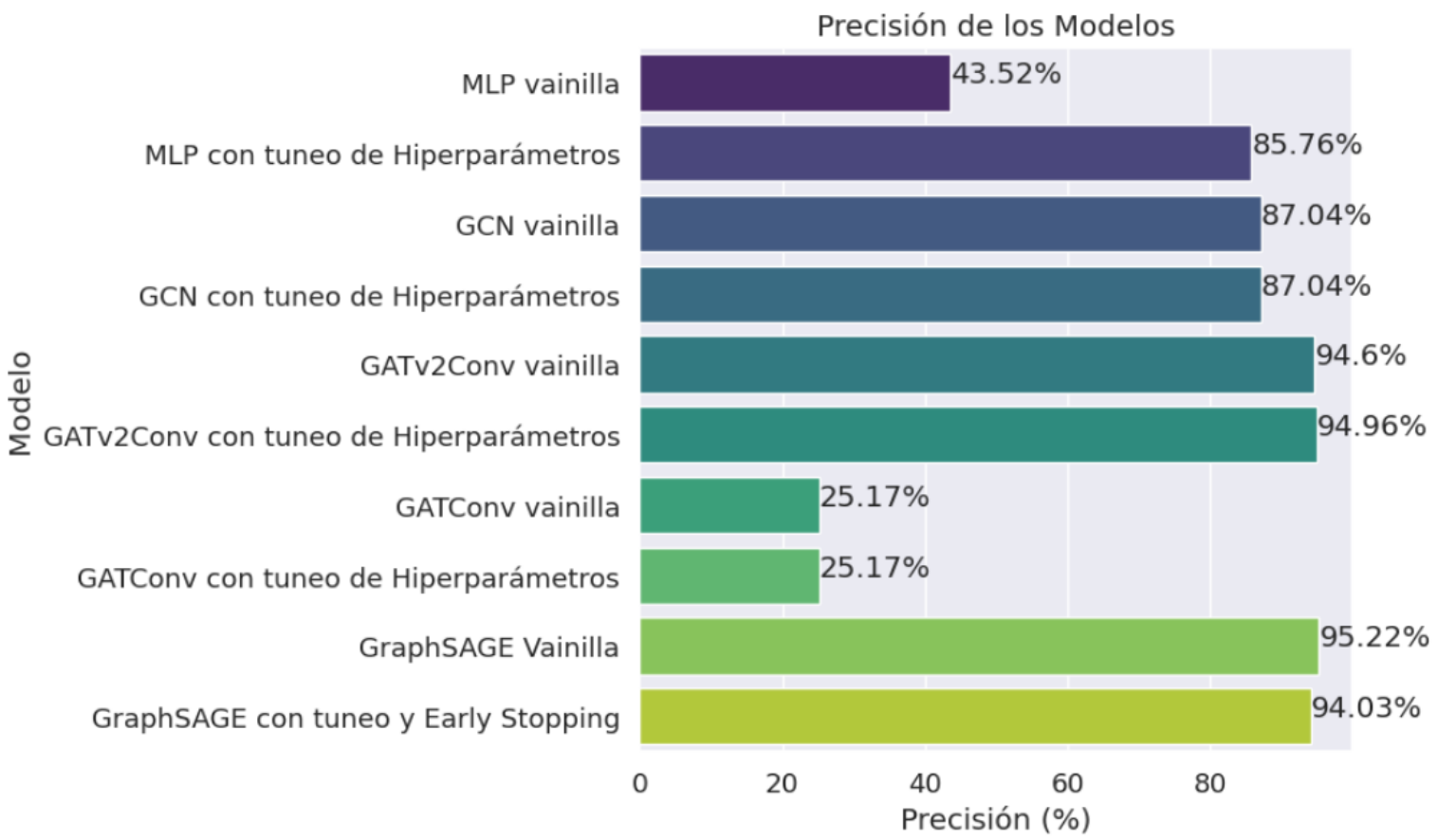
- LR: [**0,01**, 0,05]
- Weight decay: [**2e-4**, 5e-4]
- Hidden: [**8**, 16, 32, 64]

Early Stopping

- Paciencia: 5

94,03%

EVALUACIÓN Y CONCLUSIONES



- Los modelos en los que se aplicó optimización de hiperparámetros tuvieron en general un mejor desempeño comparativamente con aquellos en los que no se realizó optimización.
- Los dos modelos de mejor desempeño (GATv2Conv y GraphSAGE), tienen la capacidad de capturar y procesar patrones complejos de los grafos y sus propiedades.