

BÁO CÁO BÀI TẬP LỚN

LẬP TRÌNH HƯỚNG ĐỐI TƯỢNG

Tên đề tài : Tích hợp GNN vào app Cytoscape

GVHD : PGS.TS Lê Đức Hậu

Tên thành viên : Nguyễn Huy Hoàng – 20226107

Vũ Thị Khánh Linh - 20226113

1. Giới thiệu đề tài

1.1 Tổng quan

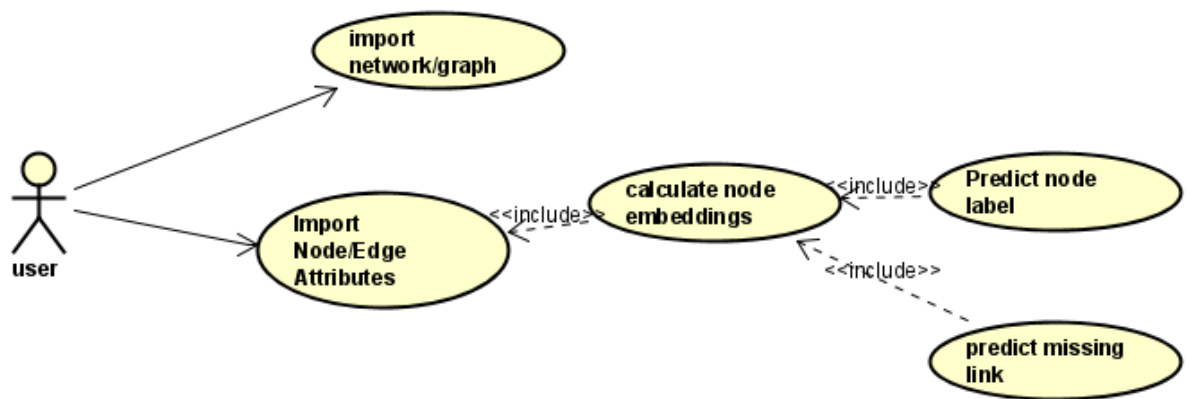
Cytoscape là một nền tảng phần mềm mã nguồn mở được sử dụng rộng rãi để trực quan hóa và phân tích mạng lưới phân tử phức tạp. Việc tích hợp Graph Neural Networks (GNN) vào Cytoscape sẽ mở rộng đáng kể khả năng phân tích và dự đoán của nền tảng này.

1.2 Mục tiêu tích hợp

Tích hợp và Sử dụng các mô hình GNN (sử dụng các thư viện trong Python hoặc Java) phù hợp (với loại graph đã import, e.g., đồng nhất (homogeneous, chỉ gồm một loại node và edge) và hỗn hợp (heterogeneous, gồm nhiều loại node và edge khác nhau)) để:

- Tính các véc tơ thuộc tính ẩn/nhúng (embedding vector) cho các node (véc tơ này là đặc trưng được học từ cấu trúc đồ thị và thuộc tính của node --> Mang toàn bộ đặc trưng của node)
- Từ các node embedding vector này, dự đoán nhãn cho các node (node classification) và các link còn thiếu (missing) (link prediction) trên đồ thị.
 - Đây là 2 lớp bài toán (đại diện cho rất nhiều bài toán) quan trọng trong hầu hết các lĩnh vực (Ví dụ: Dự đoán gen gây bệnh (node classification), dự đoán tác dụng mới của thuốc (link prediction giữa thuốc và bệnh), dự đoán khách hàng mua sản phẩm gì (link prediction) trong các Hệ khuyến nghị (Recsys), v...v...

2. Use case diagram



1. Tác nhân chính ở đây là user , người dùng sẽ import các thuộc tính của đồ thị , sau đó dựa vào các thuật toán sẽ tính ra được node embedding để có thể dự đoán nhãn của node và link còn thiếu

3. Ý tưởng triển khai

a. Phần Cytoscape App (Java)

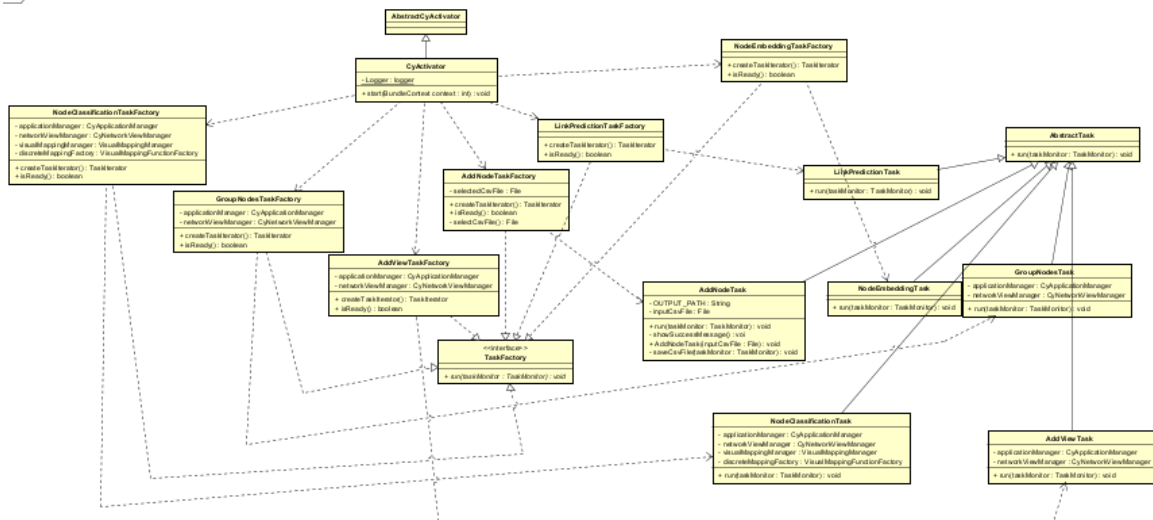
- CyActivator: Điểm khởi đầu của ứng dụng, đăng ký các services và tasks
- TaskFactories: Tạo và quản lý các task xử lý
 - NodeClassificationTaskFactory
 - NodeEmbeddingTaskFactory
 - LinkPredictionTaskFactory
 - AddNodeTaskFactory
 - AddViewTaskFactory
 - GroupNodesTaskFactory
- Tasks: Thực hiện các chức năng cụ thể
 - NodeClassificationTask: Phân loại nodes
 - NodeEmbeddingTask: Tạo embedding cho nodes
 - LinkPredictionTask: Dự đoán liên kết
 - AddNodeTask: Thêm node mới

- AddViewTask: Cập nhật giao diện
- GroupNodesTask: Nhóm các nodes

b) Python Backend Service

- Flask Server: Cung cấp RESTful API endpoints
- GNN Models:
 - Node Classification Model
 - Node Embedding Model
 - Link Prediction Model

4. Class diagram



5. Các tính chất của OOP

a) Tính đóng gói

- Thể hiện rõ qua việc sử dụng access modifiers và các biến được khai báo private

```

private final CyApplicationManager applicationManager;
private final CyNetworkViewManager networkViewManager;
private final VisualMappingManager visualMappingManager;
private final VisualMappingFunctionFactory discreteMappingFactory;
  
```

b) Tính kế thừa

- CyActivator kế thừa từ AbstractCyActivator:

```
public class CyActivator extends AbstractCyActivator {
    @Override
```

- Các Task classes kế thừa từ AbstractTask:

```
public class NodeClassificationTask extends AbstractTask {
    private final String name;
    private final String description;
```

c) Tính đa hình

- Thể hiện qua việc ghi đè (override) các phương thức:

```
@Override
public void run(TaskMonitor taskMonitor)

@Override
public void start(BundleContext context)

@Override
public boolean isReady()
```

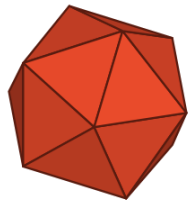
d) Tính trừu tượng

Sử dụng interfaces như TaskFactory để định nghĩa các hành vi chung

```
public class NodeClassificationTaskFactory implements TaskFactory
public class GroupNodesTaskFactory implements TaskFactory
```

6. Các công nghệ sử dụng

- Thư viện pytorch geometric
- Ngôn ngữ Java
- App Cytoscape



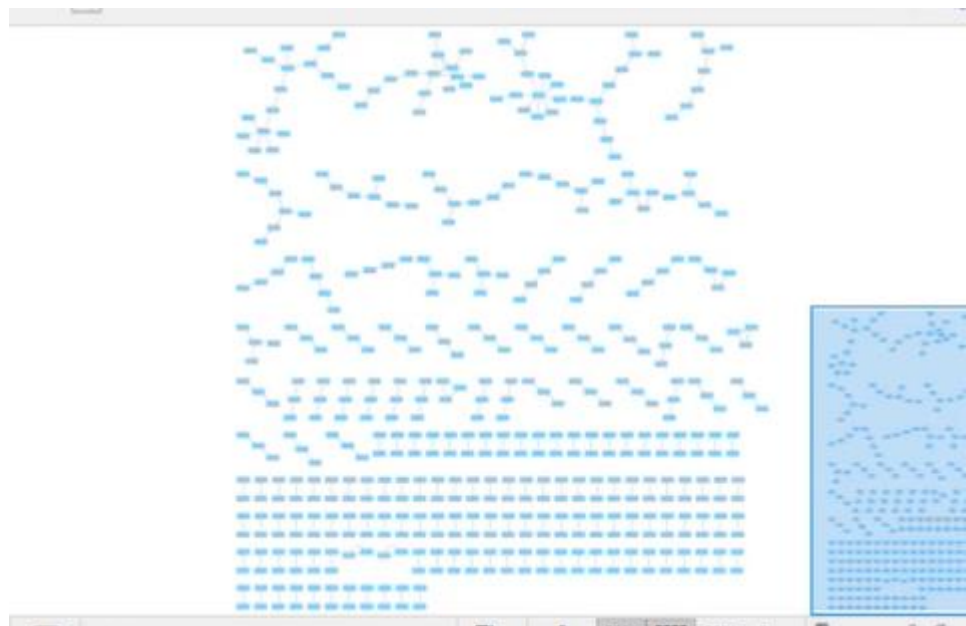
PyTorch

geometric

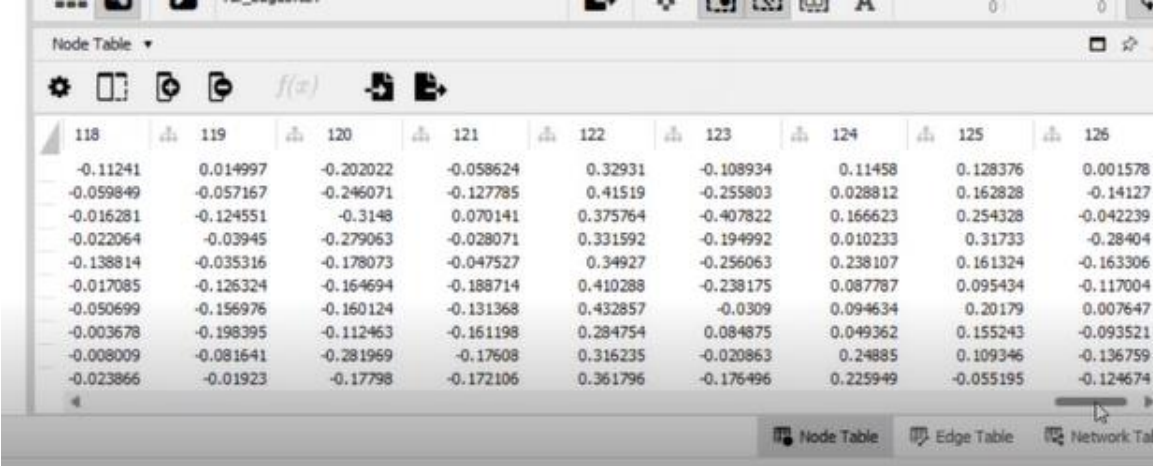


7. Demo

1. Import đồ thị

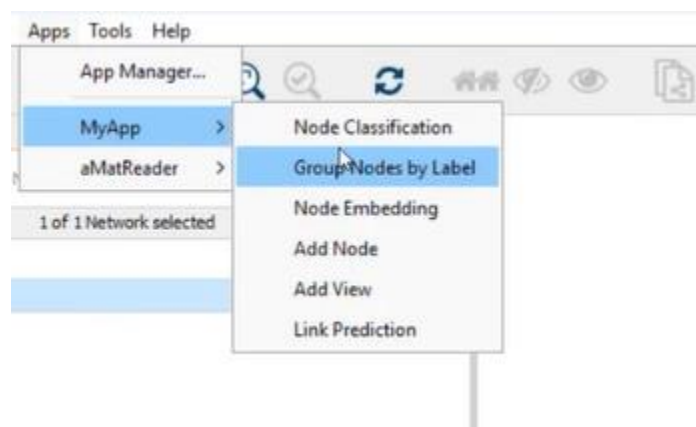


2. Import các thuộc tính của đồ thị



| 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 |
|-----------|-----------|-----------|-----------|----------|-----------|----------|-----------|-----------|
| -0.11241 | 0.014997 | -0.202022 | -0.058624 | 0.32931 | -0.108934 | 0.11458 | 0.128376 | 0.001578 |
| -0.059849 | -0.057167 | -0.246071 | -0.127785 | 0.41519 | -0.255803 | 0.028812 | 0.162828 | -0.14127 |
| -0.016281 | -0.124551 | -0.3148 | 0.070141 | 0.375764 | -0.407822 | 0.166623 | 0.254328 | -0.042239 |
| -0.022064 | -0.03945 | -0.279063 | -0.028071 | 0.331592 | -0.194992 | 0.010233 | 0.31733 | -0.28404 |
| -0.138814 | -0.035316 | -0.178073 | -0.047527 | 0.34927 | -0.256063 | 0.238107 | 0.161324 | -0.163306 |
| -0.017085 | -0.126324 | -0.164694 | -0.188714 | 0.410288 | -0.238175 | 0.087787 | 0.095434 | -0.117004 |
| -0.050699 | -0.156976 | -0.160124 | -0.131368 | 0.432857 | -0.0309 | 0.094634 | 0.20179 | 0.007647 |
| -0.003678 | -0.198395 | -0.112463 | -0.161198 | 0.284754 | 0.084875 | 0.049362 | 0.155243 | -0.093521 |
| -0.008009 | -0.081641 | -0.281969 | -0.17608 | 0.316235 | -0.020863 | 0.24885 | 0.109346 | -0.136759 |
| -0.023866 | -0.01923 | -0.17798 | -0.172106 | 0.361796 | -0.176496 | 0.225949 | -0.055195 | -0.124674 |

3. Menu sẽ bao gồm như sau



- Tính toán node embedding

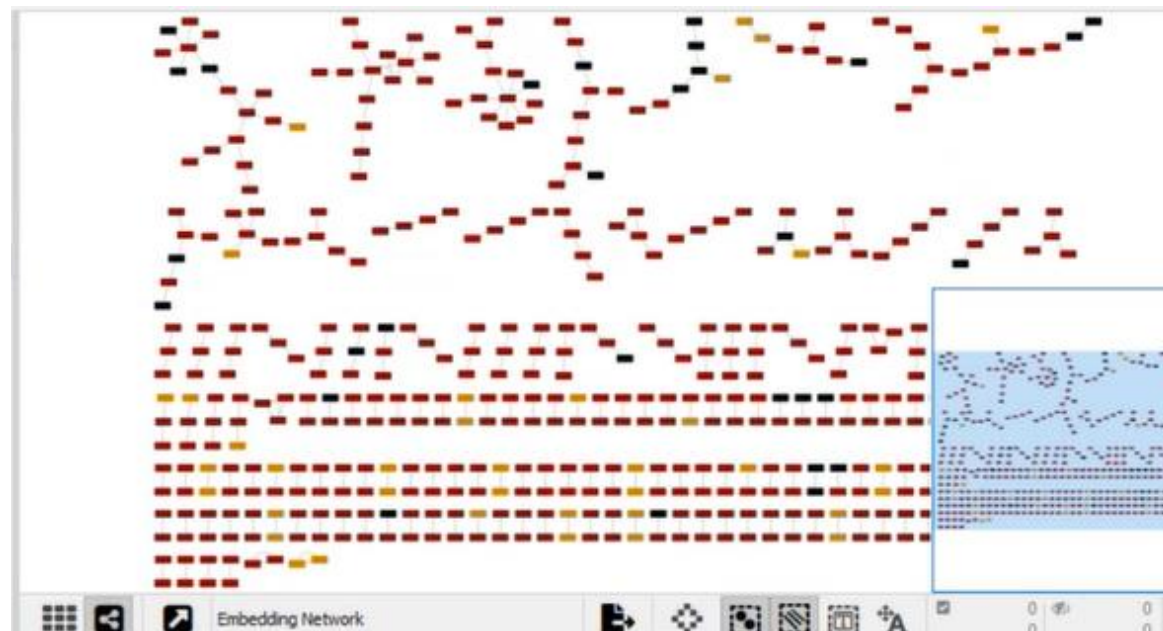
Embedding Network

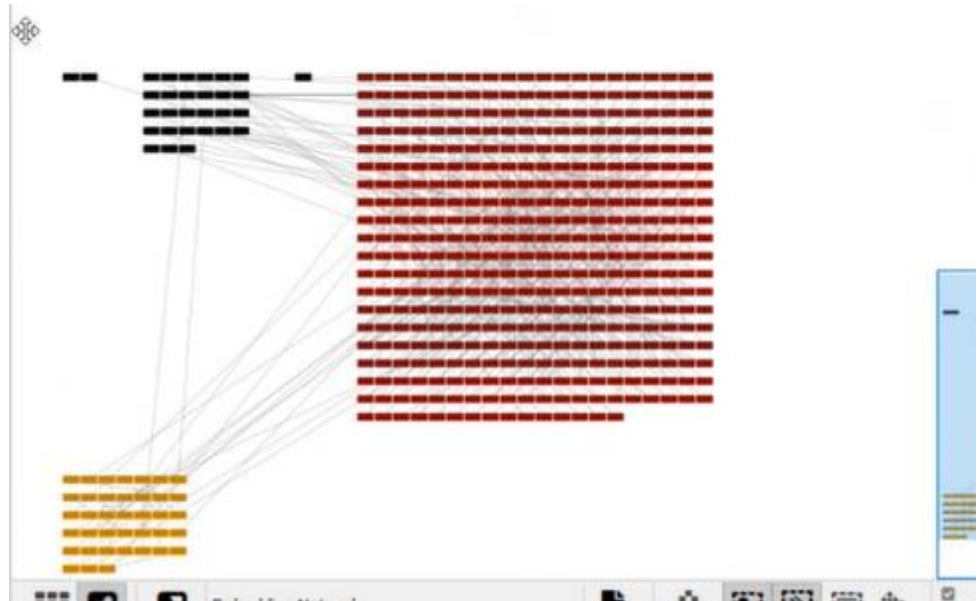
Node Table

| e | name | id | label | 0 | 1 | 2 | 3 | 4 | 5 |
|-----|------|-----|-------|------------|------------|------------|------------|------------|------------|
| 339 | | 339 | 28 | -5.4402604 | -4.5034094 | -6.063806 | -5.678567 | -5.7076917 | -6.1019096 |
| 356 | | 356 | 28 | -5.6257825 | -5.178976 | -5.724745 | -6.2030177 | -4.690882 | -5.3066106 |
| 190 | | 190 | 28 | -5.432766 | -5.2251444 | -5.0959764 | -5.522592 | -5.057819 | -3.8721564 |
| 373 | | 373 | 30 | -5.4196486 | -4.550557 | -4.7460747 | -4.4776945 | -5.934487 | -5.043645 |
| 340 | | 340 | 28 | -5.064489 | -4.676414 | -5.0392623 | -6.0025535 | -5.3079085 | -4.607872 |
| 439 | | 439 | 28 | -5.598569 | -5.4423075 | -5.4607973 | -6.5426855 | -5.28273 | -5.1459494 |
| 286 | | 286 | 28 | -5.879451 | -6.835737 | -4.5758686 | -5.577499 | -5.058521 | -4.8247194 |
| 60 | | 60 | 28 | -5.593797 | -6.165942 | -5.0358315 | -5.1781263 | -4.978565 | -5.5039268 |
| 357 | | 357 | 28 | -5.424221 | -4.566608 | -5.3067236 | -4.8065977 | -4.790107 | -4.7276564 |
| 12 | | 12 | 28 | -5.8735247 | -5.141758 | -5.2532654 | -5.123343 | -5.6308146 | -5.291598 |

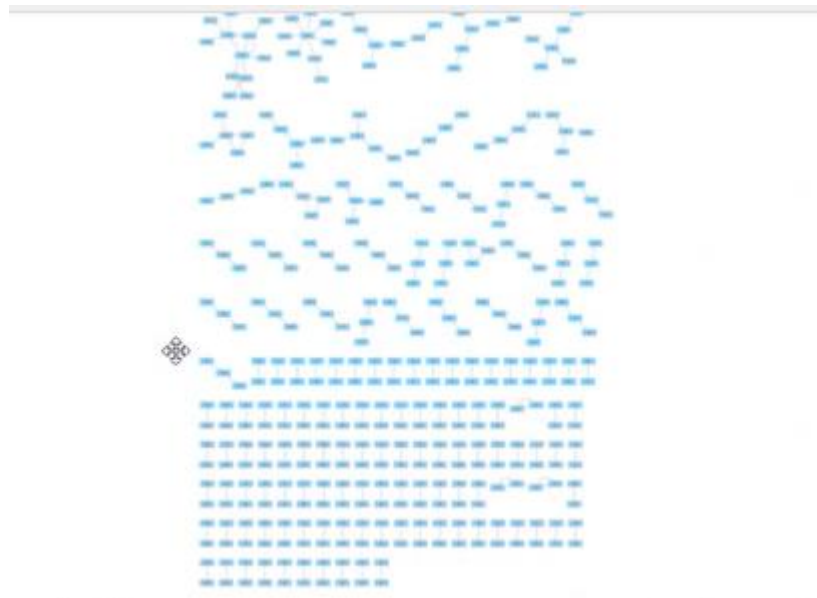
Node Table Edge Table Network Table

4. Phân loại nút và gom chúng lại thành nhóm





5. Thực hiện thêm các node mà không thêm link để có thể dự đoán link



Sau đó tiến hành dự đoán link

