SOCIAL NETWORK ASSIGNMENT 2

MARIAM MOSTAFA AMIN

2205084

DATASET DESCRIPTION

we used the facebook combined dataset since bot labels are not provided we assume the 10% of the nodes are synthetic bots

```
Nodes: 4039
Edges: 88234
```

```
bot_ratio = 0.10   # 10% bots
num_bots = int(bot_ratio * num_nodes)
```

BASELINE BOT DETECTION MODEL

we used random forest classifier

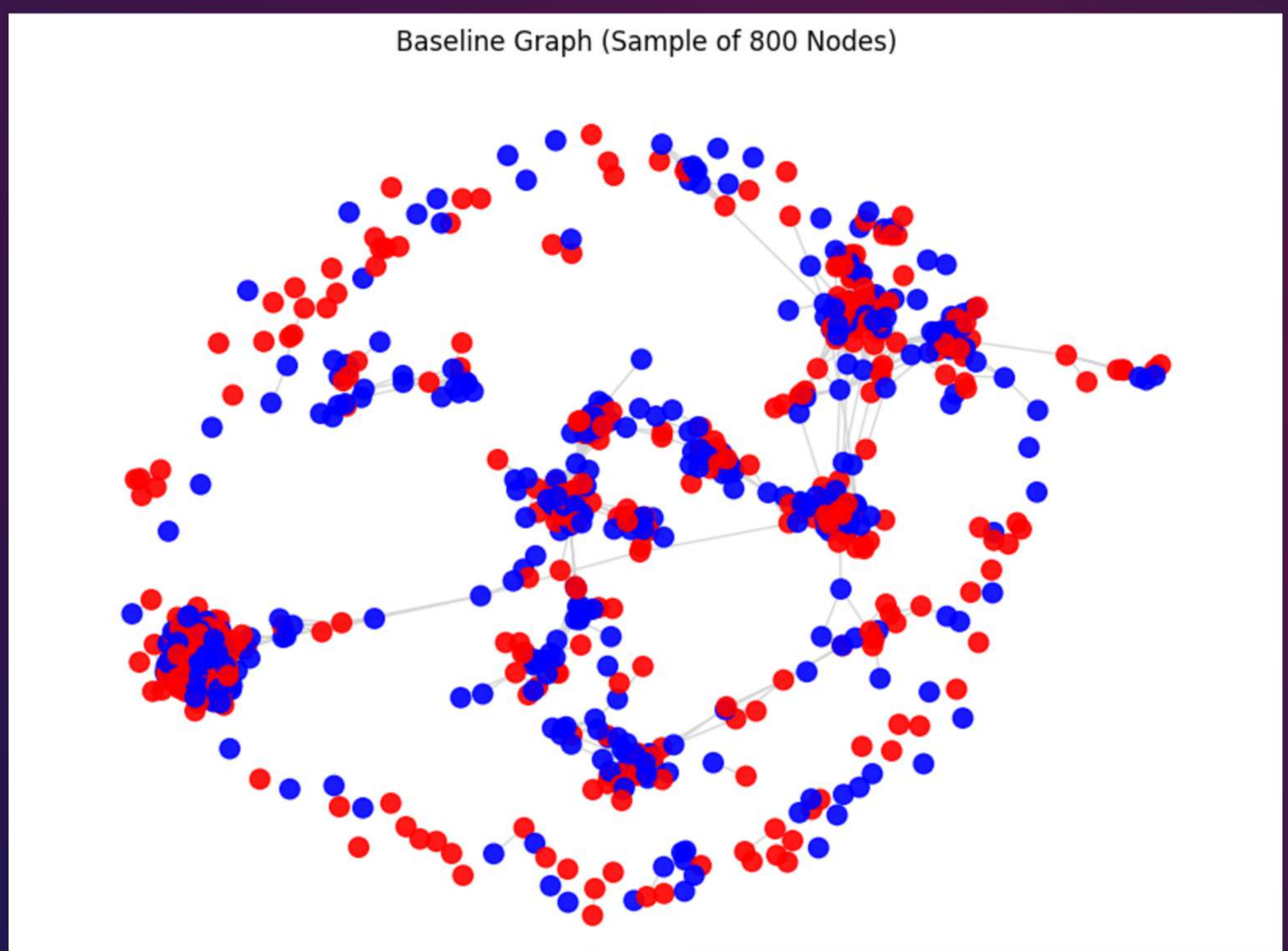
```
=== Baseline (no attack) ===
              precision    recall  f1-score   support

           0       0.90      0.99      0.94      1091
           1       0.12      0.02      0.03       121

    accuracy                           0.89      1212
   macro avg       0.51      0.50      0.49      1212
weighted avg       0.82      0.89      0.85      1212
```

this model performs poorly in bots as they are limited and synthetic but identify human nodes accurately

BASELINE BOT DETECTION MODEL



Baseline visualization of the original Facebook graph (sample of 800 nodes)
Blue = humans, Red = synthetic bots. No clear separation or pattern is visible before any attack."

STRUCTURAL EVASION ATTACK

in this attack synthetic bots try to appear as human nodes

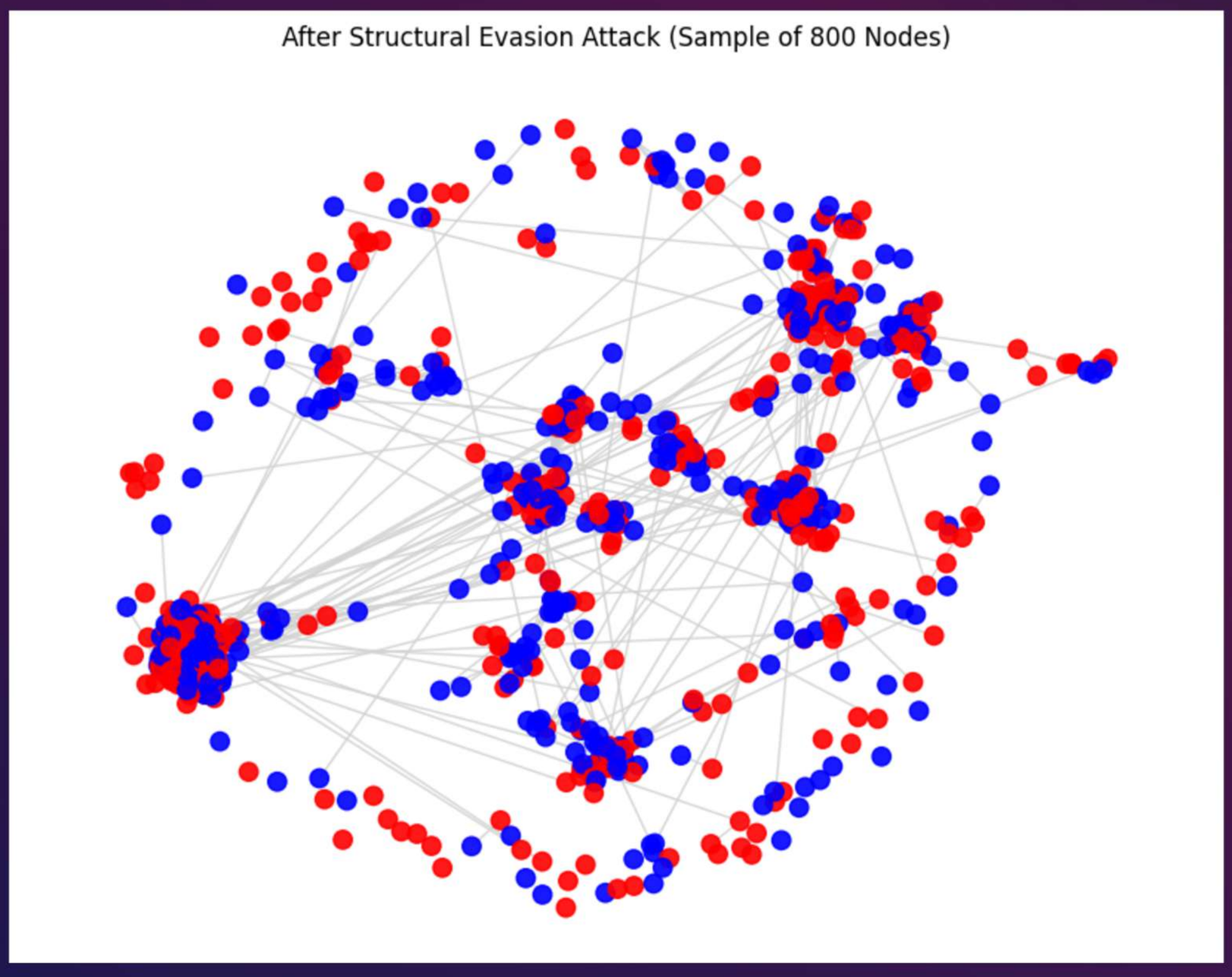
```
=== After Structural Evasion Attack ===
              precision    recall  f1-score   support

           0       0.93      0.98      0.96      1091
           1       0.69      0.36      0.47       121

    accuracy                           0.92      1212
   macro avg       0.81      0.67      0.71      1212
weighted avg       0.91      0.92      0.91      1212
```

Structural attack improved performance because the synthetic bots were originally random and did not follow any structural pattern

STRUCTURAL EVASION ATTACK



Bots (red) form more connections with humans (blue), blending deeper into communities."

GRAPH POISONING ATTACK

This attack manipulates the training data directly

```
=== After Graph Poisoning Attack ===
              precision    recall  f1-score   support

           0       0.90      0.99      0.94      1091
           1       0.56      0.15      0.23       136

    accuracy                           0.89      1227
   macro avg       0.73      0.57      0.59      1227
weighted avg       0.86      0.89      0.86      1227
```

bot detection becomes more difficult

POISONED MODEL TESTED ON CLEAN DATA

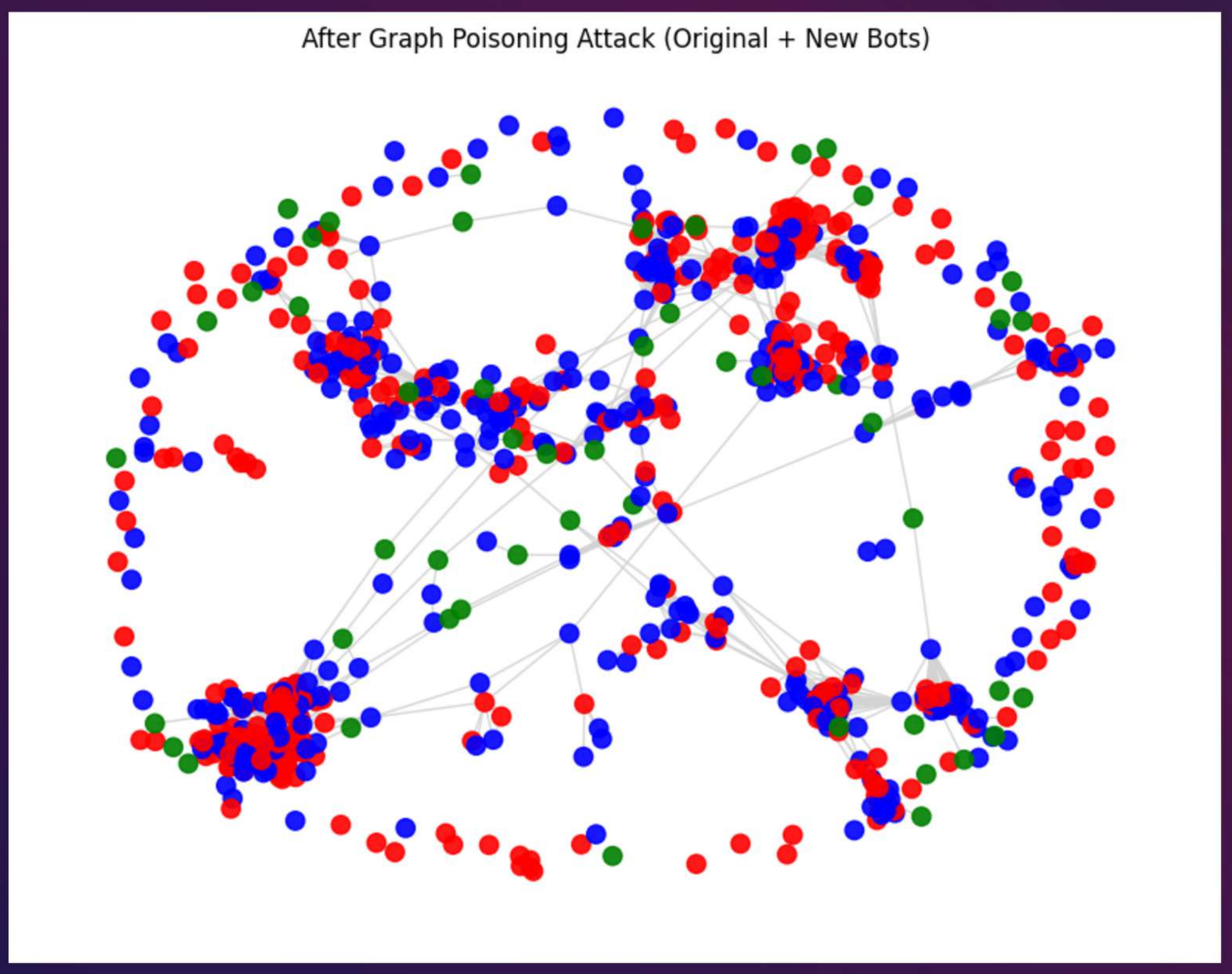
To assess how poisoning affects generalization the poisoned model was tested on the original clean graph

```
=== Poisoned model tested on CLEAN data ===
              precision    recall  f1-score   support

           0       0.91      1.00      0.95      3636
           1       0.62      0.06      0.11       403

    accuracy                           0.90      4039
   macro avg       0.76      0.53      0.53      4039
weighted avg       0.88      0.90      0.86      4039
```

GRAPH POISIONING



where new bot nodes (green) are injected
and connected to humans (blue)

The added bots blend into the network
structure, increasing density and
confusing the classifier

PERFORMANCE COMPARISON

	accuracy	bot f1	recall	percision
baseline	0.89	0.03	0.02	0.12
structural attack	0.92	0.47	0.36	0.69
poisioning attack	0.89	0.23	0.15	0.56
poisioning after clean data	0.90	0.11	0.06	0.62