

# COMP599 Network Science Assignment 3

Ann Huang, Jenny Cao

October 19, 2021

## 1 Introduction

In the first part of this project, we compared the performance of two node classification algorithms: the Harmonic functions and the local and global consistency. In the second part of this project, we compared the AUC score of two link prediction algorithms: the Jaccard coefficient and the Preferential Attachment coefficient.

## 2 Node Classification

### 2.1 Real-classic Datasets

#### 2.1.1 Strike Dataset

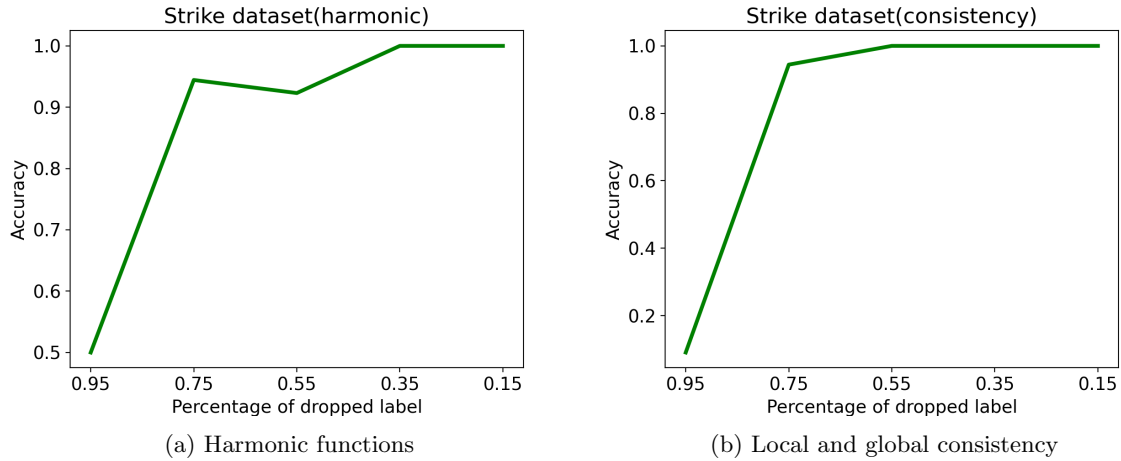


Figure 1: Node classification on the strike dataset using harmonic functions and local and global consistency on the Strike dataset.

Algorithm/Percentage dropped	0.95	0.75	0.55	0.35	0.15
Harmonic functions	0.5	0.9444	1.0	1.0	1.0
Local and global consistency	0.0909	0.944	1.0	1.0	1.0

Table 1: Performance of node classification algorithms as a function of percentage of dropped nodes on the Strike dataset

### 2.1.2 Karate Dataset

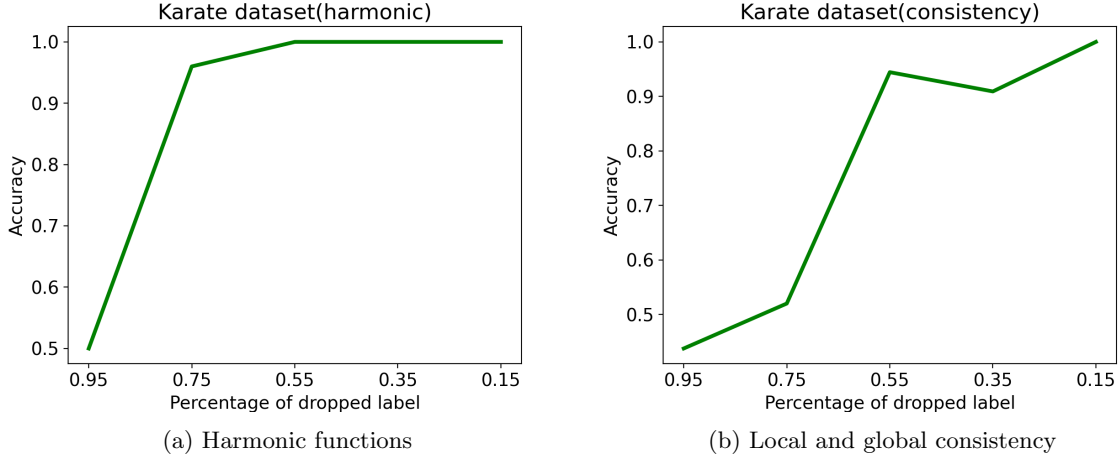


Figure 2: Node classification on the strike dataset using harmonic functions and local and global consistency on the Karate dataset.

Algorithm/Percentage dropped	0.95	0.75	0.55	0.35	0.15
Harmonic functions	0.5	0.96	1.0	1.0	1.0
Local and global consistency	0.4375	0.52	0.9444	0.9091	1.0

Table 2: Performance of node classification algorithms as a function of percentage of dropped nodes on the Karate dataset

### 2.1.3 Polbooks Dataset

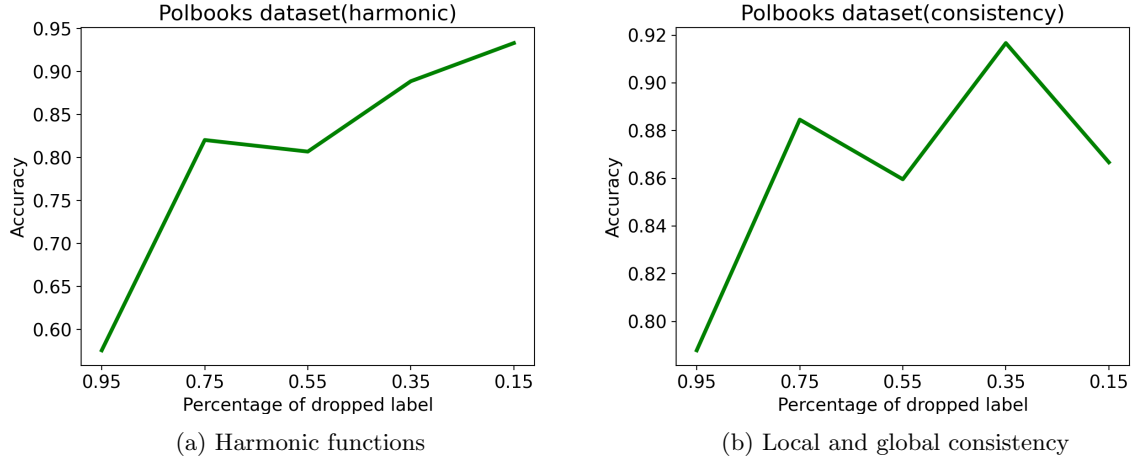


Figure 3: Node classification on the strike dataset using harmonic functions and local and global consistency on the Polbooks dataset.

Algorithm/Percentage dropped	0.95	0.75	0.55	0.35	0.15
Harmonic functions	0.9011	0.8854	0.9048	0.9002	0.9103
Local and global consistency	0.8961	0.8872	0.9048	0.8964	0.8969

Table 3: Performance of node classification algorithms as a function of percentage of dropped nodes on the Polbooks dataset

#### 2.1.4 Polblogs Dataset

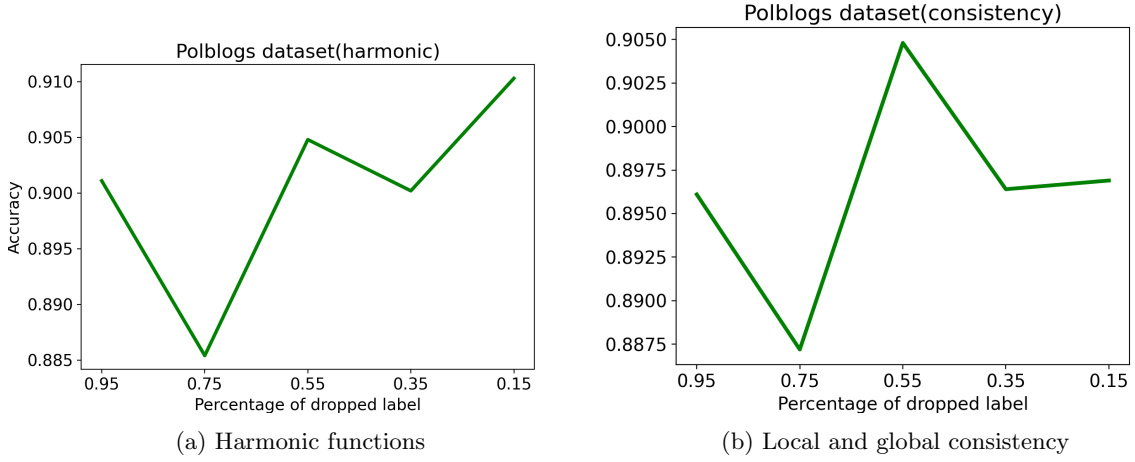


Figure 4: Node classification on the strike dataset using harmonic functions and local and global consistency on the Polblogs dataset.

Algorithm/Percentage dropped	0.95	0.75	0.55	0.35	0.15
Harmonic functions	0.9011	0.8854	0.9048	0.9002	0.9103
Local and global consistency	0.8961	0.8872	0.9048	0.8964	0.8969

Table 4: Performance of node classification algorithms as a function of percentage of dropped nodes on the Polblogs dataset

### 2.1.5 Football Dataset

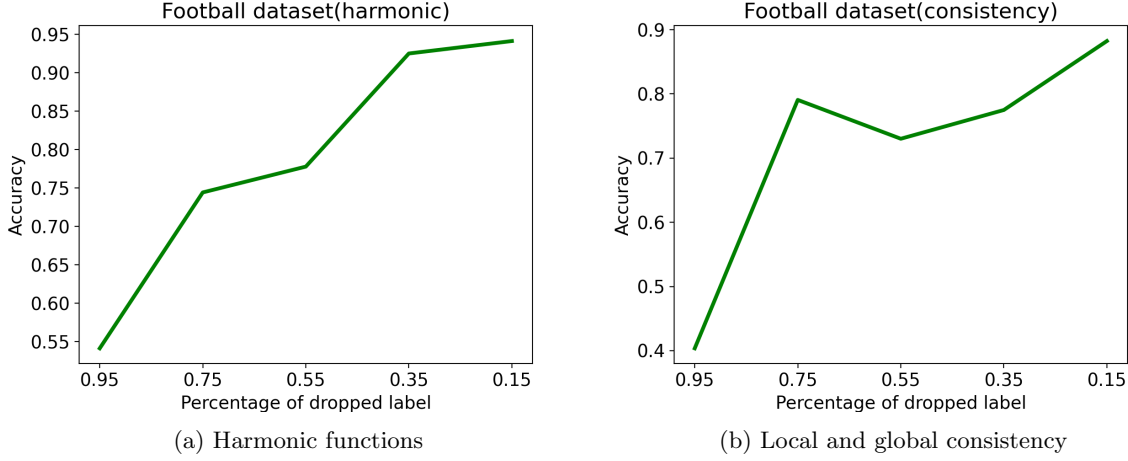


Figure 5: Node classification on the strike dataset using harmonic functions and local and global consistency on the Football dataset.

Algorithm/Percentage dropped	0.95	0.75	0.55	0.35	0.15
Harmonic functions	0.9011	0.8854	0.9048	0.9002	0.9103
Local and global consistency	0.8961	0.8872	0.9048	0.8964	0.8969

Table 5: Performance of node classification algorithms as a function of percentage of dropped nodes on the Football dataset

## 2.2 Real-node-label Datasets

We used the same split for train and test set as given by the dataset. We first assign labels to nodes in the training set, use these labels to carry out node classification on every node in a graph. After that, we evaluate the node classification accuracy only on the nodes in the test set.

### 2.2.1 Cora Dataset

The average test accuracy over 10 run for the algorithm using harmonic functions is 0.844.

The average test accuracy over 10 run for the algorithm using local and global consistency is 0.855.

### 2.2.2 Citeseer Dataset

The average test accuracy over 10 run for the algorithm using harmonic functions is 0.736.

The average test accuracy over 10 run for the algorithm using local and global consistency is 0.746.

### 2.2.3 Pubmed Dataset

The average test accuracy over 10 run for the algorithm using harmonic functions is 0.802.

The average test accuracy over 10 run for the algorithm using local and global consistency is 0.838.

## 3 Link Prediction

### 3.1 Real-classic Datasets

#### 3.1.1 Strike Dataset

For the Jaccard Coefficient, the average AUC over 10 runs is 0.015451895043731779.

For the Preferential Attachment, the average AUC over 10 runs is 0.08571428571428569.

### **3.1.2 Karate Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 0.016599732262382854.

For the Preferential Attachment, the average AUC over 10 runs is 0.046050870147255674.

### **3.1.3 Polbooks Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 0.009186366306495528.

For the Preferential Attachment, the average AUC over 10 runs is 0.010109786923474017.

### **3.1.4 Polblogs Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 0.0013543795995350771.

For the Preferential Attachment, the average AUC over 10 runs is 0.010921470925551641.

### **3.1.5 Football Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 0.006584816601064064.

For the Preferential Attachment, the average AUC over 10 runs is 0.0010354037804403589.

## **3.2 Real-node-label Datasets**

### **3.2.1 Cora Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 0.00017353387523852404.

For the Preferential Attachment, the average AUC over 10 runs is 0.019587162872101223.

### **3.2.2 Citeseer Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 4.761453370226686e-05.

For the Preferential Attachment, the average AUC over 10 runs is 0.01784481792807561.

### **3.2.3 Pubmed Dataset**

For the Jaccord Coefficient, the average AUC over 10 runs is 0.0004377949291660789.

For the Preferential Attachment, the average AUC over 10 runs is 0.0150141128077486.