

# COMP 9102 Assignment 3

Wang Cheng

## 1 Implementation

### 1.1 Building the matrix

Since we treat retweet data as an undirected graph, the matrix should be symmetric.

---

```
for row in data:
    matrix[row[0] - 1, row[1] - 1] = 1
    matrix[row[1] - 1, row[0] - 1] = 1
```

---

### 1.2 Personal page-rank

PPR-based proximity vector for node  $u$  is defined as follows:  $p_u = (1 - \alpha)Ap_u + \alpha e_u$ . By setting  $p_u^{(0)} = \mathbf{0}$ ,  $p_u$  is computed iteratively.

Note: The value of  $p_u$  is independent of  $p_u^{(0)}$ .

## 2 Experiment

To run the code: `python assignment3.py -k 5`. Here, `-k` is the number of clusters to form.

	Purity	Entropy	Normalized mutual information (NMI)
k = 5	11	11	11
k = 10	11	11	11
k = 15	11	11	11