

A Machine Learning Perspective

ECE 18-755 M4 Poster Session Dec 7, 2017 || Team Member: Jiangtian Qian(jiangtiq) & Rui Wu(rwu2)

Introduction

The world's platform for change

212,279,858 people taking action. [Victories every day.](#)

Start a petition

Project aim:

Analyze emerge and spread phenomena of dynamic sustainable pro-social behavior in social network.

Data sets:

Collected from Change.org, one of the most active petition platform.

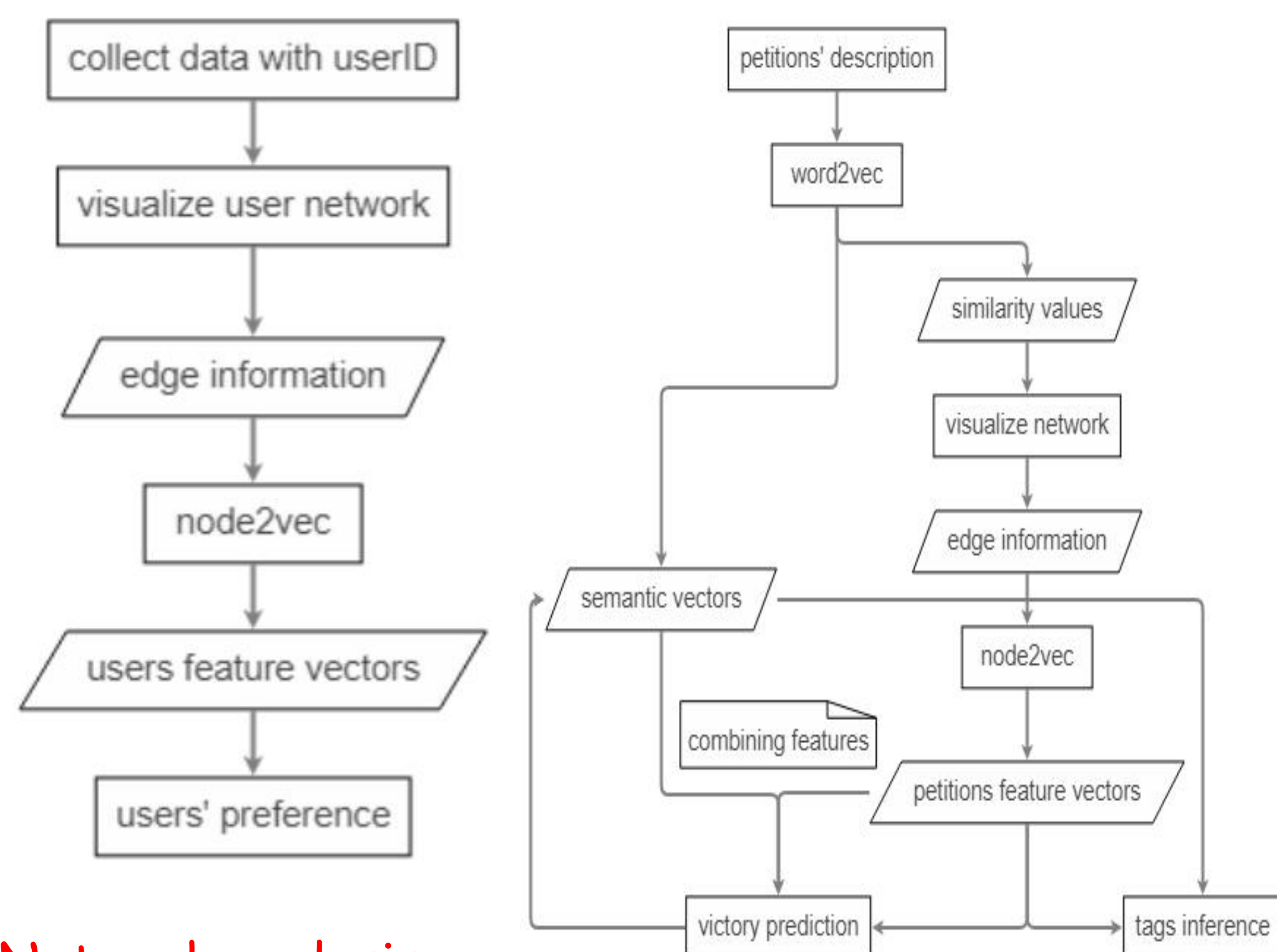
Methods:

Node2vec, word2vec & combination of both.

Approach

Users research:
node2vec & linear regression

Petition research:
word2vec & multi-label regression



Network analysis:

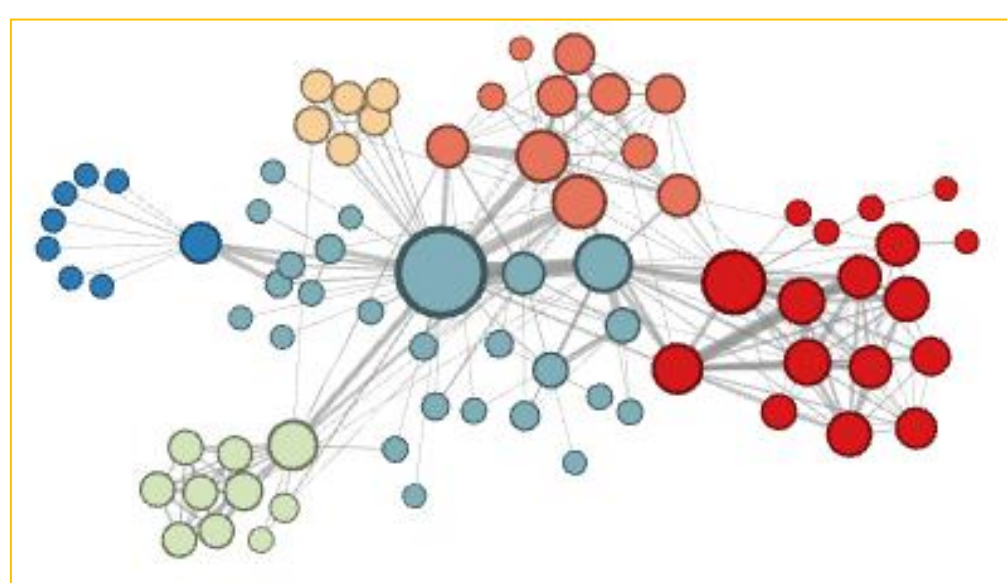
Community detection
Degree distribution

Set up:

Most frequent 20 tags

Evaluation metric:

Precision and recall in binary task
Macro and Micro in multilabel task



Problem description

Problem 1: How participants influence each other.

Problem 2: How large online social network brings single behavior others and influence them.

Problem 3: How much influential users will determine the number of signatures a petition could get since all followers of an influential user will also sign a petition that this influential user signs.

Problem 4: Is there any possibility that certain content may lead to the victory of the petition.

Results

User Research:

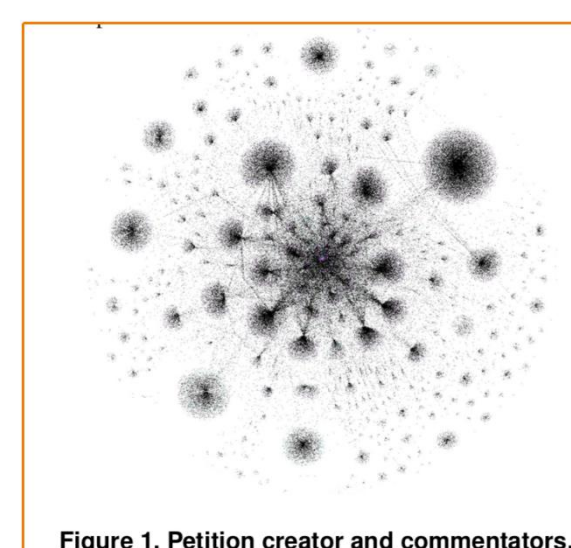


Figure 1. Petition creator and commentators.

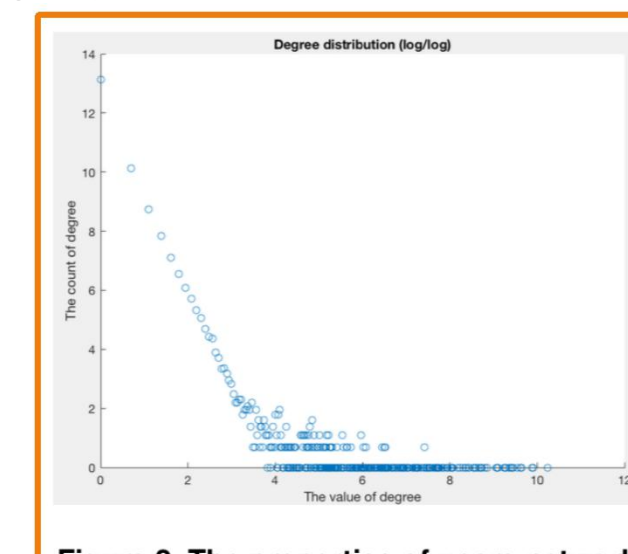


Figure 2. The properties of users network

For user network, we have 650000 nodes and 3400000 edges, and degree distribution follows power law, so user network is a scale free network.

Petition network:

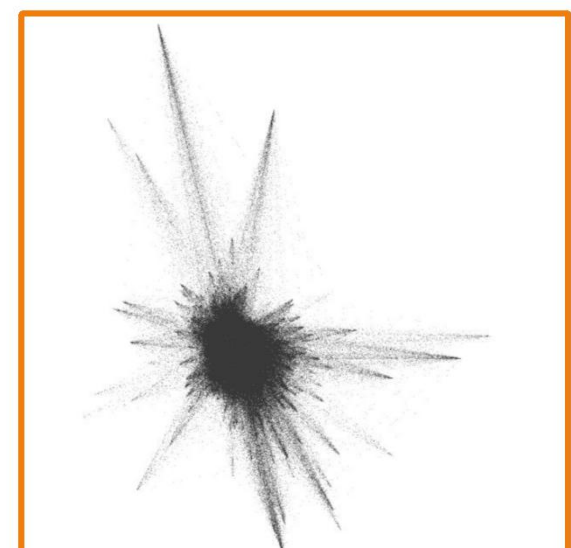


Figure 3. Petitions Network from 21 Topics

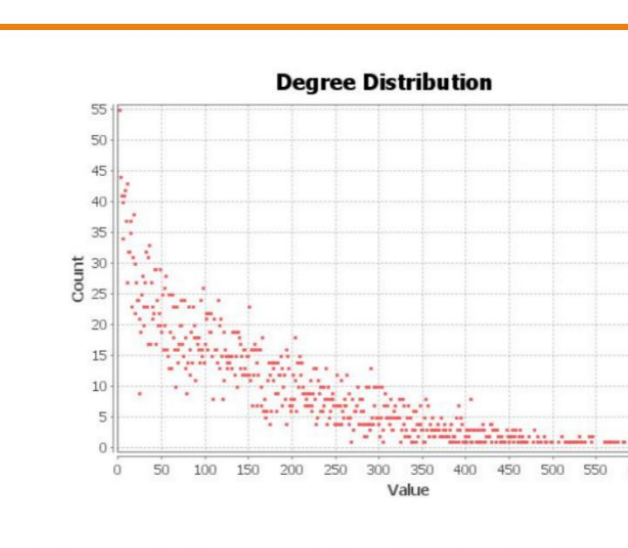


Figure 4. Degree Distribution in Petition Network

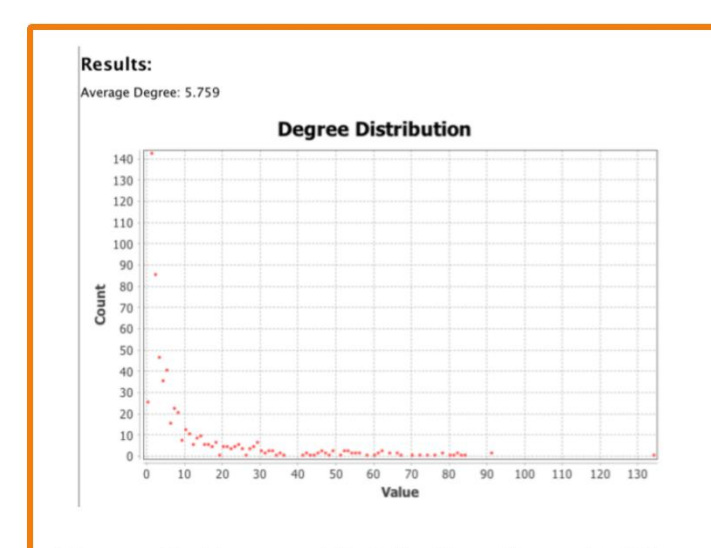


Figure 5. Degree Distribution from Petition Network with Victory Status

For petition network, we have 4545 nodes and 341352 edges, and degree distribution is power law, so petitions also forms into a scale free network.

Petition multi-label predict:

| | word2vec | node2vec | word2vec & node2vec | feature all 0 |
|-------|----------|----------|---------------------|---------------|
| macro | 0.479 | 0.480 | 0.481 | 0.481 |
| micro | 0.923 | 0.924 | 0.922 | 0.481 |

Macro and micro in combination reduce.

Petition victory predict:

| word | Prediction | | | node | Prediction | | | combination | Prediction | | | all zeros | Prediction | | |
|----------------|------------|---------|-------|----------------|------------|---------|-------|----------------|------------|---------|-------|----------------|------------|---------|-------|
| | Failed | Victory | Total | | Failed | Victory | Total | | Failed | Victory | Total | | Failed | Victory | Total |
| Actual Failed | 231 | 121 | 352 | Actual Failed | 265 | 87 | 352 | Actual Failed | 258 | 94 | 352 | Actual Failed | 352 | 0 | 352 |
| Actual Victory | 154 | 89 | 243 | Actual Victory | 167 | 76 | 243 | Actual Victory | 142 | 101 | 243 | Actual Victory | 243 | 0 | 243 |

Precision: 0.42

Precision: 0.47

Precision: 0.52

Precision: 0

We combine features for binary and multilabel, and find that simply combining was not good for multilabel classification while could improve the performance for victory prediction.