**class** Data:

**def** extract\_training\_set(self, user\_num):

1. 选择用户集；
2. 提取用户之间的关系；
3. 提取用户微博；

预处理：

1. 分词；
2. 去除stopwords；