pagerank_wikivote

PageRank is a popular algorithm in scoring nodes importance in a graph dataset. Here is an example of applying pagerank algorithm to wiki-administrators-vote data to find out: Who are important wiki administrators involved in elections?

First of all, let's analyze the dataset. data souce: https://snap.stanford.edu/data/wiki-Vote.txt.gz Source (citation)

```
J. Leskovec, D. Huttenlocher, J. Kleinberg. Signed Networks in Social Media. CHI 2010.
J. Leskovec, D. Huttenlocher, J. Kleinberg. Predicting Positive and Negative Links in Online Social Netw
```

Here displays some information in the data source:

```
wikivote = "/Users/lulu/BIGDATA/NetworkX/wiki-Vote.txt"
read.csv(wikivote,nrow = 3)
```

```
##                                                                                X..Directed.graph..each.u
## 1 # Wikipedia voting on promotion to administratorship (till January 2008). Directed edge A->B means
## 2
## 3
```

As you can see, the dataset is very similar to a graph type data structure, since there are only two elements in each line, showing a vote from A to B. In R, the 'igraph' library provides function make_graph to convert a vector to a graph object which is then suitable as the input of the function page.rank. The page.rank function gives the score of importance of each node using google pagerank algorithm.

```
dat = read.csv(wikivote,sep ='\t',skip = 3)
edges = as.character(as.vector(as.matrix(dat)))
library(igraph)
```

```
##
## Attaching package: 'igraph'
##
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
##
## The following object is masked from 'package:base':
##
##     union
```

```
graf = make_graph(edges,directed = TRUE)
prank = page.rank(graf,directed = TRUE,vids = V(graf))
```

Now we have IDs of the top 10 important/popular administrators as well as their scores.

```
##        2565        1549          15        1166        4310        2237
## 0.002282738 0.001589535 0.001235025 0.001231330 0.001165892 0.001147063
##        4037         737        5697        5524
## 0.001141931 0.001100773 0.001099142 0.001081451
```

The library igraph is also available in python. There is another similar library NetworkX in python which also does a very good job in dealing with graph data.