

# Social and Information Network Analysis:

## Homework 2 - Network Generator

CHENG-EN SUNG, National Cheng Kung University

### ACM Reference format:

Cheng-en Sung. 2018. Social and Information Network Analysis: Homework 2 - Network Generator. 1, 1, Article 1 (May 2018), 4 pages.

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

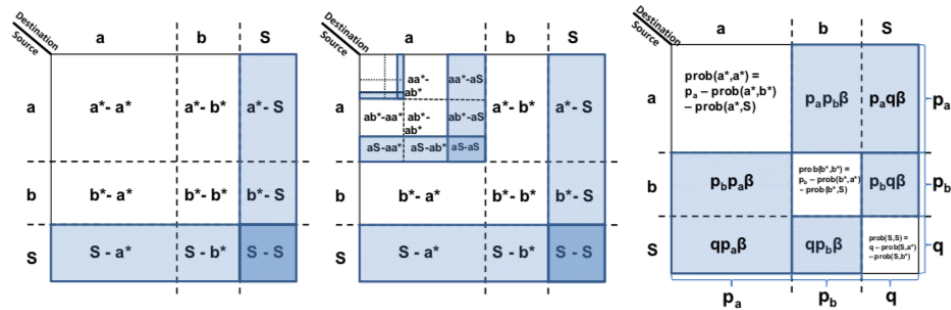
## 1 DESCRIPTION

In homework 2, I use RTG[1] algorithm to generate syntactic network, which satisfied following three properties:

- Power-law degree distribution
- High cluster coefficient
- Short average path length

RTG is short for Random Typing Generator, by type randomly on keyboard with  $k$  characters and a space bar. A space is hit with probability  $q$ ; all other characters are hit independent in-equivalence with the remaining probabilities. Space bar is used to separate sequence characters into words. author treat each separate word as node in network. To form links between nodes, mark the sequence of words as 'source' and 'target' nodes. That is, divide the sequence of words into group of word pairs, in each word pair, first element is source node and second element is target node. If two nodes are already linked, add 1 to the weight of the link. Therefore, if there are totally  $W$  words,  $W/2$  links will form into this network.

The author proposed a two-dimensional keyboard that generate source and destination labels in one shot (see following pictures)



Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 Association for Computing Machinery.

XXXX-XXXX/2018/5-ART1 \$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

In the above example, we assume only three keys('a', 'b', and 'space') in keyboard, we can regard this process as selecting one of the nine grids on left-most picture with probability corresponding to row character ( $p_r$ ) times column character ( $p_c$ ) in that grid. After the grid is selected, append the row character to source word and append the column to target word. This selecting process execute recursively as above picture in the middle part until the space bar(e.g., character 'S') is selected. In order to model 'homophily' and 'coomunities' author add an imbalance factor  $\beta$ , which will decrease the chance from select a node to b node (so called a-b/b-a form) and increase a-a and b-b form. In the above setting will formulate nine probabilities for 9 grids. (above right-most picture)

- $prob(a, b) = prob(b, a) = p_a p_b \beta$
- $prob(S, a) = prob(a, S) = q p_a \beta$
- $prob(S, b) = prob(b, S) = q p_b \beta$
- $prob(a, a) = p_a - (prob(a, b) + prob(a, S))$
- $prob(b, b) = p_b - (prob(b, a) + prob(b, S))$
- $prob(S, S) = q - (prob(S, a) + prob(S, b))$

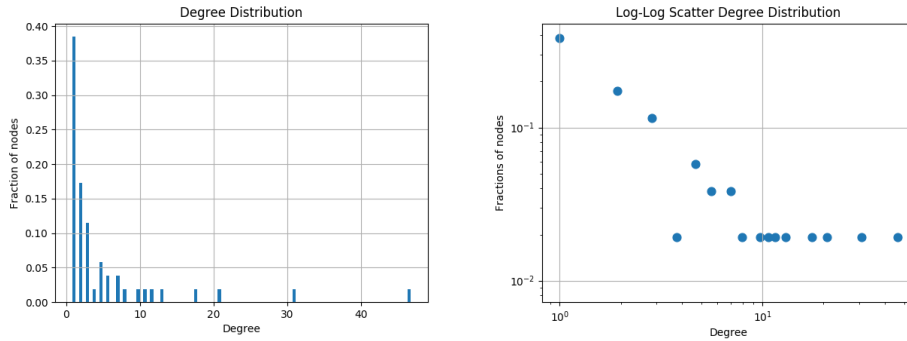
## 2 SIMULATION COMPARISON

In this section, we examine graph generate by RTG satisfied three properties mentioned in DESCRIPTION section before and execution time on different number of edges (i.e.,W) We fix

- Number of word pairs (i.e., W):  $W = 600000$
- Number of keys in keyboard (i.e., k):  $k = 2$
- Imbalance factor:  $\beta = 0.95$
- Terminated (i.e., space bar) probability:  $q = 0.8$

### 2.1 Power-law degree distribution

Visualize degree distribution in bar chart and log-log scatter plot showing below:



We can see that node's degree is satisfied power-law distribution.

### 2.2 High cluster coefficient

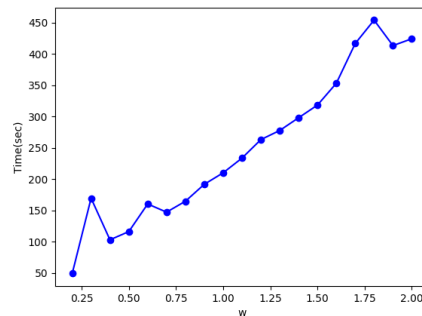
We got **Cluster Coefficient** using networkx build-in function(i.e., average\_clustering) equal to 0.547180. Though not as high as the value mention in class (0.7..), in average, 0.5 on cluster coefficient is larger than other network generator.

### 2.3 Short average path length

We got **Average Path Length** using networkx build-in function(i.e., `average_shortest_path_length`) equal to `2.009400`, which satisfied real network properties: average path length usually short

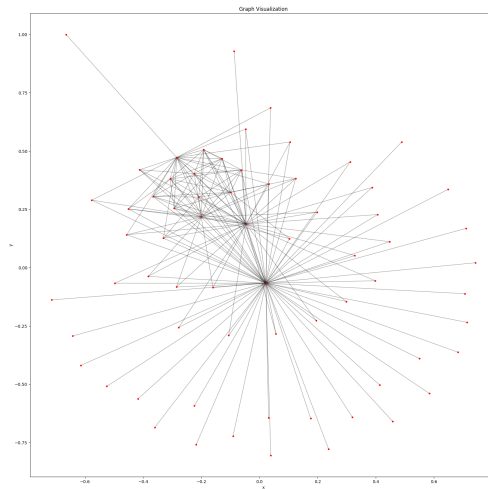
### 2.4 Execution Time Analysis

In this subsection, given different  $W$  from 0.1 2.0 (Note: x-axis scale of  $10^6$ ) and  $y$  is the corresponding execution time when other parameter setting are same as above setting. It's easy to aware that execution time increase when number of edges get larger and is linearly increasing.



## 3 VISUALIZATION

Visualize graph generate by RTG algorithm.



## REFERENCES

- [1] Leman Akoglu and Christos Faloutsos. 2009. RTG: A Recursive Realistic Graph Generator Using Radom Typing. *LNAI 5871* 19, 2 (Dec. 2009), 13–28. <http://dblp.uni-trier.de/db/journals/datamine/datamine19.html#AkogluF09>