Documentation for Social Random Walk

Christian Geldermann
on behalf of CISPA

August 20, 2018

1 Program Options

1.1 Mandatory options

1.1.1 $--if$

Defines the input file. The path can be relative or absolute, as long as *fopen* accepts it.

1.1.2 $--of$

Defines the output file, The path can be relative or absolute, as long as *fopen* accepts it.

1.1.3 $--walks$

Defines the number of walks.

1.1.4 $--length$

Defines the number of nodes that a single walk should consist of.

1.2 Optional options

1.2.1 $-p$ or $--print$

Prints the transitions of the graph. One line for each transition.

1.2.2 $-w$ or $--weighted$

Indicates that the input file contains a weighted graph.

1.2.3 $-o$ or $--optiTransOrder$

Indicates that the order of transitions should be optimized for better performance. This optimization is only available for weighted graphs and has no effect on non-weighted graphs.

For weighted graphs the option sorts the transitions by weight (highest weight first). If all weights are mostly equal or sorted with highest weight first anyway, you can skip the option.

1.2.4 $-u$ or $--undirected$

Converts a graph into an undirected graph. If the graph is weighted, the reverse of a transition is only added if it does not already exist. If it exist, nothing changes for those two associated transitions.

1.2.5 $-r$ or $--reflect$

Reflects the whole graph. The result is an undirected graph. For un-weighted graphs the effect is the same as $-u$. For weighted graphs, the reverse of a transition is added even if it already exists. This way the resulting graph contains transitions in both directions for every transition of the original graph. The weight of a transition from node n_1 to n_2 is the sum of the transition from n_1 to n_2 and that from n_2 to n_1 in the original graph.

2 Input format

2.1 Non-weighted graphs

The input file has to contain one transition per line (determined by '\n'). Every line has to consist of 2 numbers (the source node followed by the destination node) separated by any non-digits (any ASCII character except '0' to '9'). To see an example, have a look at the *Examples* section.

2.2 Weighted graphs

The input file has to be structured as that for non-weighted graphs except that there hast to be a third number per line that specifies the weight. The weight is interpreted as the probability to calculate the random walks.

Assume a node has 3 outgoing transitions with weights 2, 3 and 5. The probability that the transition with weight 2 is taken is $\frac{2}{2+3+5} = \frac{2}{10} = 20\%$.

3 Examples

3.1 Original graphs

3.1.1 Non-weighted

Lets assume the following graph to be defined in the file *graph.txt*:

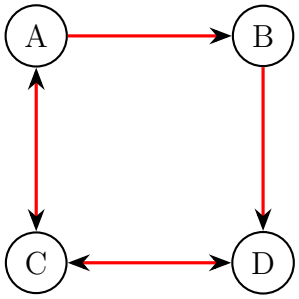

Figure 1: Basic non-weighted graph

The *graph.txt* file content could be:

0,1
0,2
1,3
2,0
2,3
3,2

The node numbers may also start with 1 or any other number. But the maximum node number should be kept as small as possible. The transitions don't need to be sorted in any way.

3.1.2 Weighted

Lets assume the following weighted graph to be defined in the file *graph_weighted.txt*:

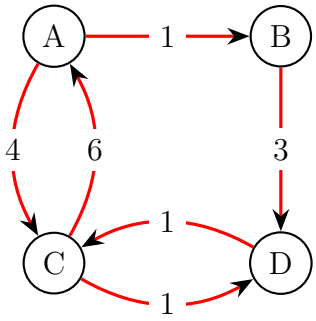

Figure 2: Basic weighted graph

The *graph.txt* file content could be:
0,1,1
0,2,4
1,3,3
2,0,6
2,3,1
3,2,1

3.2 Make undirected ($-u$)

3.2.1 Non-weighted

Calling *./walk -u –if=./graph.txt –of=./out.txt –walks=20 –length=10* would output 20 random walks of length 10 to *out.txt*. The graph from *graph.txt* (see figure 1) is converted to the graph in figure 3 before performing the walks:

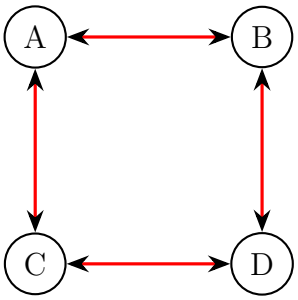

Figure 3: Undirected non-weighted graph

Using $-r$ or $--ru$ instead of $-u$ would do exactly the same.

3.3 Weighted

Calling *./walk -uow –if=./graph_weighted.txt –of=./out.txt –walks=20 –length=10* would output 20 random walks of length 10 to *out.txt*. The graph from *graph_weighted.txt* (see figure 2) is converted to the following before performing the walks:

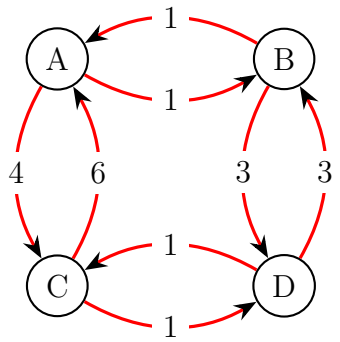

Figure 4: Undirected weighted graph

Calling *./walk -row –if=./graph_weighted.txt –of=./out.txt –walks=20 –length=10* instead would perform the walks on the graph:

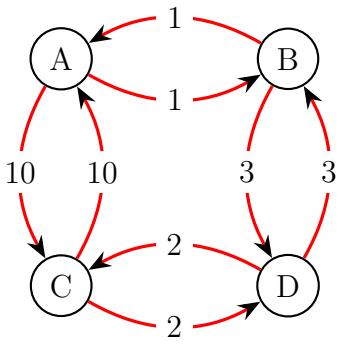

Figure 5: Reflected weighted graph

The $-o$ option is only for performance reasons.