1   RaTrav 13.09.16

RaTrav 13.09.16: mean first passage times and node occupancies with Hill's and Monte Carlo algorithms
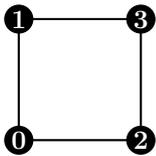
2   Input file

Figure 1: A simple network can take the form of a square with each corner representing a state, labelled from 0 to 3.

Let us start with a very simple network. For a square network, as illustrated in Fig. 1, the input file format takes the form:

```
WALK 1000000
SIMU 1
NODE 0 2 1
NODE 1 3 0
NODE 2 0 3
NODE 3 1 2
```

Keyword `NODE` is followed by the node number (the first numerical column, entries are counted from 0; it is important to maintain an increasing numeric order of node identities and gaps in numbering are not permitted). Subsequent numbers are the identities of neighbouring nodes and their order does not matter. In the case of performing a Monte Carlo simulation there is the requirement for `WALK` and `SIMU` keywords (number of walkers, number of simulations with `WALK` walkers). Entering `SIMU` > 1, the user will receive error estimates for each MFPT calculation (standard deviation of average MFPT over the number of simulations run). To establish waiting times on, for example the 9th node, the NODE 9 input line may be defined as:

```
NODE 9 3 9 9
```

i.e. the node number is repeated. This means that the node labelled as 9 in a graph connects directly to the node labelled as 3 and to itself twice, so the transition probability from node 9 to node 3 is 1/3, whereas the waiting probability on the node 9 is 2/3. If the user chooses to use Hill's algorithm, `WALK` and `SIMU` keywords won't be used even if present in the input file.

In advanced file format the user can define transition probabilities and weights of edges. To minimise rounding errors, in addition to the decimal format, it is possible to define the transition probabilities and weight of edges as common fractions. The same square as in Fig. 1 can be defined as follows (node number, followed by a triplet containing: identity of neighbouring node, transition probability, weight of edge):

```
NODE 0 2 1/2 1.0 1 1/2 1.0
NODE 1 3 1/2 1.0 0 1/2 1.0
NODE 2 0 1/2 1.0 3 1/2 1.0
NODE 3 1 1/2 1.0 2 1/2 1.0
```

With transition probabilities equal to 1/2 and weights of edges equal to 1.0, this format is equivalent to the basic file format. However, in advanced file format it is possible to use different combinations of probabilities and weights. Requirements are that the probabilities in each line have to sum up to 1.0, and in the case of Hill's method, at least one local timescale (weight of edge) must be different from 0. The last condition does not apply to the Monte Carlo method. Thus, on account of some specific problem, the advanced input file may take the following form:

```
NODE 0 2 1/2 2.0 1 1/2 1.0
NODE 1 3 1/3 0.5 0 2/3 0.1
NODE 2 0 3/5 3.0 3 2/5 0.2
NODE 3 1 3/4 5.0 2 1/4 4.0
```

In the above please note the different values of weights possible for two-way transitions between the same pair of nodes, for example, 1 and 3. We assume in general that a passage in either direction along a given edge does not need to correspond to the same weight or transition probability.

Use of the `MFPT` keyword allows definition of which MFPTs the user wants to compute, e.g.: `MFPT 0 1` means the MFPT between states 0 and 1 will be calculated, `MFPT 0 1 2` means the MFPT between initial state 0 and two final states will be calculated – either state 1 or 2 has to be reached (it is possible to define any number of final states).

The `OCCU` keyword works only for Monte Carlo method and returns node occupancies after given number of steps, e.g. after 100000:

```
OCCU 100000
```

If the user chooses this mode with Hill's method (`OCCU` not necessary), equilibrium or stationary occupancy probabilities for all nodes will be obtained.

The `INFO` keyword allows the user to pass any comments which will be copied to the output files, e.g.: `INFO This is a comment.`

3 Compilation

There is a Makefile attached so under Linux it is enough to call `make`. Two variables have to be set before doing so:

```
CXX=icc
BOOST=-I PATH_TO_BOOST/boost/include
```

First one is used to choose the compiler, e.g. `icc` for Intel C Compiler or `g++` for GNU C Compiler. Monte Carlo method uses Boost C++ Libraries (`http://www.boost.org/`) which have to downloaded and installed in your system. Second variable is used to set the path to the Boost C++ Libraries so the user has to change `PATH_TO_BOOST` to the local path.

Alternatively RaTrav tool may be compiled without Makefile, e.g. with ICC, as follows:

```
icc -o RaTrav -I PATH_TO_BOOST/boost/include MC.13.09.16.cpp HI.13.09.16.cpp
        RaTrav.13.09.16.cpp
```

4   Usage

```
RaTrav input output METHOD INPUT_FORMAT MODE
```

Five parameters have the following meaning: `input` is a name of the input text file which defines the network; `output` is a name of the output text file generated by the RaTrav; `METHOD` parameter equals `0` for Monte Carlo method or `1` for Hill's method; `INPUT_FORMAT` parameter equals `0` for basic input file or `1` for advanced input file; `MODE` parameter equals `0` when all MPFTs will be calculated, `1` when selected MFPTs will be calculated or `2` when occupancies will be calculated.

For example: for Monte Carlo method, basic input file and all MFPTs the user runs `RaTrav input.txt output.txt 0 0 0`; for Hill's method, advanced input file and occupancies the user runs `RaTrav input.txt output.txt 1 1 2`.

5   Functionalities

To better explain the functionalities of the RaTrav tool, we present in Fig. 2 a simple irregular network with nine nodes (black dots) and nine edges (links between nodes). It has one cycle (edges coloured in green); edges have different transition probabilities $u_{ll'}$ (probabilities of passing defined as common fractions; coloured in black, in case of a cycle in green, in case of waiting on a node in cyan), and local time scales $\tau_{ll'}$ (the cost of passing defined as decimal fractions with arrows; coloured in red). For some nodes there are waiting times (i.e. transitions to the same node; cyan circular arrows); sometimes the transition probability on an edge node is set equal to one, instead of a waiting time. Both variables $u_{ll'}$ and $\tau_{ll'}$ may be symmetric or asymmetric.

The network presented in Fig. 2 may be represented as follows (we are numbering the nodes from the left to the right, and from top to bottom, so S=0, E1=7, E2=8):

```
NODE 0 1 1.0 1.4
NODE 1 0 1/5 3.1 2 2/5 2.3 4 1/5 0.5 5 1/5 2.5
NODE 2 1 1/2 3.2 2 1/2 1.0
NODE 3 4 1/3 0.8 3 2/3 1.0
NODE 4 3 1/4 0.2 1 1/2 0.3 5 1/4 1.0
NODE 5 4 1/4 10.0 1 1/4 0.1 6 1/2 6.0
NODE 6 5 1/3 1.5 7 1/3 8.0 8 1/3 1.0
NODE 7 6 3/4 3.3 7 1/4 1.0
NODE 8 6 1.0 0.7
```

Please notice that node numbers must be written in increasing order, but the neighbours of each node may be written in any order. If we add the following lines to the input file:

Figure 2: A simple network with nine nodes and its properties (transition probabilities (black), local times scales (red), cycle (green), probability of waiting (cyan), initial (S) and final (E1, E2) states). See text for more details.

```
WALK 1000000
SIMU 1
MFPT 0 7
MFPT 0 8
MFPT 0 7 8
```

the output using Hill's method (WALK and SIMU not used) is:

```
===== OUTPUT =====
MFPT|Time in seconds
MFPT FROM 0 TO 7: 80.300 0.0
MFPT FROM 0 TO 8: 83.233 0.0
MFPT FROM 0 TO 7,8: 61.556 0.0
==================
```

and with the Monte Carlo method (WALK and SIMU used) is:

```
===== OUTPUT =====
MFPT|Time in seconds
MFPT FROM 0 TO 7: 80.437 0.9
MFPT FROM 0 TO 8: 83.121 0.9
MFPT FROM 0 TO 7,8: 61.581 0.7
==================
```

For the presented network, the MFPT between S and E1 is calculated to be 80.300, between S and E2 to be 83.233, and between S and E1 or E2 (multiple final state) to be 61.556. Two observations can be made: for the submitted network Hill's method is faster and gives exact results; using the Monte Carlo method, with 1000000 walkers, gives similar results. In case of reducing the number of walkers to 100000 but increasing the number of simulations to 10:

```
WALK 100000
SIMU 10
```

the following results with Monte Carlo can be obtained:

```
===== OUTPUT =====
MFPT|Error|Percentage error|Time in seconds
MFPT FROM 0 TO 7: 80.437 0.054 0.067% 0.8
MFPT FROM 0 TO 8: 83.121 0.074 0.089% 0.9
MFPT FROM 0 TO 7,8: 61.581 0.056 0.092% 0.6
=================
```

In this case the program also returns the standard deviation for each MFPT; comparing each MFPT, and its associated error, with the equivalent but exact result from Hill's method, we may notice that one standard deviation may be insufficient to denote each MFPT pair to match; however, within three standard deviations each MFPT pair should match. This may not always be the case for more complicated networks, since if the number of walkers in each simulation is too low, every node may not be visited by a walker from the ensemble (ergodicity issues).

Examples of input files for different topologies, in basic and advanced file format, may be found on the RaTrav website.

6 Important remarks

1. In Hill's algorithm the final state must be different from the initial state. However, when using the Monte Carlo method one can identify such states and obtain the MFPT, which is in fact a return time to the origin.

2. In Hill's algorithm the number of final states cannot exceed N-2, where N is the total number of states (nodes in a graph). However, for the Monte Carlo method one can define such a configuration of states, e.g. `MFPT 0 1 2 3` for a square.

3. By choosing the Monte Carlo method the user is able to define ultrafast transitions between all or selected states, for which the local passage time along an edge between such states is set equal to zero. In Hill's algorithm at least one local time scale must be different from zero.

4. A network does not have to be compact. If there is no path between states, RaTrav will return MFPT as 'Infinity (not accessible)'.

5. In order to calculate occupancy probabilities of network nodes in the case of various local time scales (various weights for edges) exactly for input network, please use Monte Carlo method. In order to use Hill's method in the case of various weights of edges, the network has to be transformed to equivalent network with the global time scale (all weights of edges equal to 1.0 and transformed transition probabilities do not have to sum up to 1.0). For the mean first passage times the values match between Monte Carlo and Hill's method (interplay of transformed transition probabilities and transformed occupancy probabilities). However, occupancy probabilities (only for the case of various time scales) will be different. The occupancy probabilities obtained from Monte Carlo mathod will match Hill's occupancy probabilities when the network after transformation (exit probabilities need not sum up to 1.0) is normalized by the highest sum of exit probabilities for one of nodes and waiting times are added to each node in order to sum up the transition probabilities to 1.0. The backward transformation, from Hill's to Monte Carlo occupancy probabilities, is not known yet and the study is still in progress.

7 References

1. M. Torchala*, P. Chelminiak*, M. Kurzynski and P. A. Bates, 'RaTrav: a tool for calculating mean first-passage times on biochemical networks', *BMC Syst. Biol.* **7**, 130 (2013). *Joint First Authors. http://dx.doi.org/10.1186/1752-0509-7-130

2. M. Torchala, P. Chelminiak and P. A. Bates, 'Mean first-passage time calculations: comparison of the deterministic Hill's algorithm with Monte Carlo simulations', *Eur. Phys. J. B* **85**, 116 (2012). http://dx.doi.org/10.1140/epjb/e2012-20760-8