

Network motif identification in stochastic networks

Rui Jiang, Zhidong Tu, Ting Chen, and Fengzhu Sun

1. Running environment

The program provided in this web site need the support of *message passing interface* (mpi). You must have *mpich2* installed in order to run the provided program. Regarding how to install a copy of *mpich2*, please refer to <http://www-unix.mcs.anl.gov/mpi/>.

2. Configuration file

The program *emnm.exe* is the only executable. It first reads a configuration file to determine which task it should perform. A sample configuration file is given as follows,

```
1.  [Action]
2.      Command=EnumSubNumb
3.  [Common]
4.      DataPathWin32=c:\emnm\data\yeast\
5.      WorkPathWin32=c:\emnm\work\yeast\
6.      TempPathWin32=c:\emnm\temp\yeast\
7.      CommPathWin32=c:\emnm\comm\
8.      DataPathLinux=/home/username/emnm/data/yeast/
9.      WorkPathLinux=/home/username/emnm/work/yeast/
10.     TempPathLinux=/home/username/emnm/temp/yeast/
11.     CommPathLinux=/home/username/emnm/comm/
12. [Graph]
13.     NodeFile=node.txt
14.     DirEdgeFile=dedge.txt
15. [Subgraph]
16.     Node=3
17.     DirEdgeNorm=1
18.     DirEdgeSelf=0
19.     DirEdgeDble=0
20.     DirEdgeFull=0
21.     MutEdgeNorm=0
22.     MutEdgeSelf=0
23.     MutEdgeFull=0
24. [EnumSubNumb]
25.     Method=1
26.     NumbFileAscii=1
27.     NumbFileBinary=1
28.     NumbFilePrefix=subnumb
```

A configuration file is composed of several *sections* marked in a pair of “[]”. For example, [Action], [Common], [Graph], [Subgraph], and [EnumSubNumb] in the above example file mark the beginning of a

few sections. In each section, there are a few statements, normally in for format of `parameter=value`, where the `parameter` is a named variable in the program and the `value` is the value that a user wants to pass to the corresponding variable. All available parameters are already given in the included example for a yeast regulatory network, user can simply copy and modify that example file in order to get their own configuration file.

3. Running steps

In order to identify a stochastic network in a given network, a user needs to run the following 4 steps.

1. Enumerate the number of occurrence of each type of subgraph in the given network. The command for this step is

```
mpirun -np 1 emnm.exe step1.txt
```

2. Randomize the given network to obtain a set of randomized networks. The command for this step is

```
mpirun -np 2 emnm.exe step2.txt
```

Here, the user can choose a proper number (2 in the above example) for parallel running of the program. Typically, set the number as `#CPUs + 1`.

3. Enumerate the number of occurrence of each type of subgraph in each of the randomized networks and summarize the number to obtain the expected number and probability of observing each type of subgraph in the background. The command for this step is

```
mpirun -np 2 emnm.exe step3.txt
```

Also, set the number of parallel program as `#CPUs + 1`.

4. Invoke the EM algorithm to identify network motifs. The command for this step is

```
mpirun -np 2 emnm.exe step4.txt
```

Also, set the number of parallel program as `#CPUs + 1`.

4. Output file

The following is an example of the output file `emsimu.o3c.txt`.

I=	0	5	1.664669E-018
N=	100	100	1.00
M=	-2.293843E+006	-2.293792E+006	1.013611E+002
L=	9.999990E-001	1.012572E-006	
T=	0.000000E+000	1.000000E+000	1.000000E+000
	0.000000E+000	0.000000E+000	9.998834E-001
	0.000000E+000	0.000000E+000	0.000000E+000
	0.000000E+000	0.000000E+000	0.000000E+000
	0.000000E+000	0.000000E+000	0.000000E+000
	0.000000E+000	0.000000E+000	0.000000E+000

The detailed explanation of the meaning of each number is given as below

I=	Not used(ignore)	Internal status (ignore)	p-value according to χ^2 (ignore)
N=	#(the motif identified)	#(the EM repeated)	%(the motif identified)
M=	Log likelihood under the null	Log ikelihood under the alternative	(log likelihood)
L=	Estimated $1-\lambda$	Estimated λ	Not used (no value)
T=	The identified motif in matrix format (directed part)		
	The identified motif in matrix format (undirected part)		

References

Rui Jiang, Zhidong Tu, Ting Chen, and Fengzhu Sun. Network Identification in Stochastic Networks. *Proc. Natl. Acad. Sci. USA*, Vol.103, No.25, 9404~9409, 2006.