

Software for network metrics

Short description

Purpose

Software module is designed to calculate PCHI-C network metrics. Results obtained using this software are described in the paper “Characteristic topological features of promoter capture Hi-C interaction networks”. This software serves as a supplementary material for the paper and allows to reconstruct completed experiments using the same data source or provide similar experiments using other data sets. Software package contains software module together with the corresponding data files in JSON format.

Requirements

Python version 3.6 or above is required.

Module

Module calculates network metrics for all chromosomes.

Contents

Python script `HiCNetworkMetrics.py` together with supplementary script `CReader.py` in `Utils` folder, data files in `Data` folder.

Usage

In the `Data` folder there should be 24 JSON files:

1. `tissueDistance.json`, describing distances between tissue types,
2. `graphChr<chromosome_id>.json` (where `chromosome_id` = 1...22, X), specifying network structure for the particular chromosome.

After executing “`HiCNetworkMetrics.py`” from some framework or command line, there should be the following output on the console:

```
DONE load ./Data/tissueDistance.json
process chr 1
DONE load ./Data/graphChr1.json
process chr 2
DONE load ./Data/graphChr2.json
process chr 3
DONE load ./Data/graphChr3.json
process chr 4
DONE load ./Data/graphChr4.json
process chr 5
DONE load ./Data/graphChr5.json
process chr 6
DONE load ./Data/graphChr6.json
process chr 7
DONE load ./Data/graphChr7.json
process chr 8
DONE load ./Data/graphChr8.json
process chr 9
DONE load ./Data/graphChr9.json
process chr 10
```

```

DONE load ./Data/graphChr10.json
process chr 11
DONE load ./Data/graphChr11.json
process chr 12
DONE load ./Data/graphChr12.json
process chr 13
DONE load ./Data/graphChr13.json
process chr 14
DONE load ./Data/graphChr14.json
process chr 15
DONE load ./Data/graphChr15.json
process chr 16
DONE load ./Data/graphChr16.json
process chr 17
DONE load ./Data/graphChr17.json
process chr 18
DONE load ./Data/graphChr18.json
process chr 19
DONE load ./Data/graphChr19.json
process chr 20
DONE load ./Data/graphChr20.json
process chr 21
DONE load ./Data/graphChr21.json
process chr 22
DONE load ./Data/graphChr22.json
process chr X
DONE load ./Data/graphChrX.json
file saved ./Results/stats-all-17-norm.csv

```

Results are saved in the comma-separated file `stats-all-17-norm.csv` in the `Results` folder.

This file is in the following format:

Chr	tissue1	tissue2	clasif	dist1	dist2	trans all	nodes	links	CC_V_3	CC_V_4	CC_V_5	CC_V_6	CC_V_8	CC_V_+	CC_V_max	C
9	nCD8	nB	CLP	488	11	24137	0.639388	1.007283	0.0	0.001161	0.001161	0.000871	0.001161	0.010453	0.135892	0
9	nCD8	tB	CLP	488	11	24137	0.61049	1.021048	0.0	0.001097	0.001371	0.000274	0.001371	0.009325	0.146177	0
9	tCD8	nB	CLP	488	11	24137	0.643388	1.015421	0.0	0.001787	0.000894	0.000894	0.000596	0.010425	0.127784	0
9	tCD8	tB	CLP	488	11	24137	0.613874	1.022092	0.0	0.001125	0.001407	0.001125	0.000281	0.007877	0.275428	0
9	nB	tB	CLP	359	1	24137	0.784427	1.511662	0.0	0.001121	0.001682	0.000841	0.000561	0.010934	0.196527	0
10	Mon	Mac0	CFU-GM	546	12	31173	0.704941	0.947549	0.0	0.002214	0.001476	0.002707	0.001476	0.014763	0.072586	0
10	Mon	Mac1	CFU-GM	546	12	31173	0.687121	0.904784	0.0	0.001979	0.001484	0.001237	0.001979	0.015088	0.057631	0
10	Mon	Mac2	CFU-GM	546	12	31173	0.681624	0.903152	0.0	0.00216	0.0012	0.00168	0.00216	0.014641	0.060482	0
10	Mon	Neu	CFU-GM	450	8	31173	0.605859	0.845271	0.0	0.001086	0.001629	0.000814	0.000814	0.011401	0.046959	0
10	Mon	MK	CMP	546	12	31173	0.61556	0.787199	0.0	0.003264	0.000816	0.001904	0.00272	0.016593	0.04053	0
10	Mon	EP	CFU-GM	546	12	31173	0.537925	0.654644	0.0	0.002404	0.00187	0.002671	0.003472	0.013088	0.043269	0
10	Mon	Ery	CMP	546	12	31173	0.618705	0.805545	0.0	0.002458	0.002732	0.000819	0.002458	0.014751	0.063373	0
10	Mon	FoeT	HSC	575	13	31173	0.576066	0.785637	0.0	0.002045	0.001533	0.001533	0.002811	0.009201	0.067983	0
10	Mon	nCD4	HSC	575	13	31173	0.603899	0.821566	0.0	0.002199	0.00171	0.000733	0.001466	0.012948	0.073533	0

Each row contains chromosome number, two tissue type identifiers, distances in the Haematopoietic tree hierarchy, number of links in the graph for the particular chromosome, normalized number of nodes, links and various network metrics.