Using GRAND database's API for reproducible network reconstruction

Marouen Ben Guebila

Introduction

GRAND database https://grand.networkmedicine.org/ hosts a collection of gene regulatory networks reconstructed using netZoo tools. The networks encompass a variety of human tissues, cell lines, and small molecule regulatory effects.

For example, the cell line view in https://grand.networkmedicine.org/cell , we can retrieve cell line gene regulatory network reconstruction available for download in the network column. For example, the LCL cell line gene regulatory network [1], has been reconstructed using PANDA as specified in the tool column, as implemented in netZooM version 0.1.

This information is necessary if we would like to reproduce the network provided in GRAND database. Additionally, all the aruguments necessary to run a reconstruction of LCL cell line are provided in PPI, Motif, and Expression columns.

Fetching netZoo version

To make sure that the reconstructed network is exactly the same, we have to use the specific implmentation of PANDA that was used to generate the network, in this case netZooM 0.1. Although, in the vast majority of cases, the code is expected to have the same behaviour between versions, unless there is a major bug fix.

To check all the version of netZooM

```
git clone https://github.com/netZoo/netZooM.git
cd netZooM
```

Then

```
git tag -l
```

Let's switch to version 0.1

```
git checkout tags/0.1 -b v0.1
```

At any time, we can retrieve the current version through

```
git checkout master
```

Downloading priors using programmatic access

Let's start by making a new folder for LC cell line

```
mkdir lcl
cd lcl
```

then let's fetch information about cell lines programmatically, without having to navigate the website. This is particulalrly useful for large queries. To do this, we use the GET api method. In MATLAB, it is called as follows:

```matlab
url = 'https://grand.networkmedicine.org/api/v1/cellapi/';
cells = webread(url);
cells.results(1)
```

The first entry is LCL cell line. We are intersted in downloading the regulatory network, and all the data that was generated to reconstruct the network. The goal is to rebuild the network from scratch and comapre it to the one that we downloaded. The files are hosted in Amazon Web Servies (AWS), so we will use their client to do the download. Alternatively, we can use the function webread to get the file directly in our working space.

```matlab
system(['aws s3 cp ' cells.results(1).network ' .']);
%networkPair = webread(gb.results(2).network ) %Alternative

% Download PPI
system(['aws s3 cp ' cells.results(1).ppi ' .']);
% Download expression data
system(['aws s3 cp ' cells.results(1).expression ' .']);
% Download TF binding motif data
system(['aws s3 cp ' cells.results(1).motif ' .']);
```

Reproducing GRAND database cell line network

Before calling PANDA, we need to set the parameters

```matlab
% Set Program Parameters
exp_file   = 'lcl/LCL_exp.txt';
motif_file = 'lcl/LCL_motif.txt';
ppi_file   = 'lcl/LCL_ppi.txt';
panda_out  = '';  % optional, leave empty if file output is not required
lib_path   = '../netZooM';  % path to the folder of PANDA source code
save_temp  = '';
alpha      = 0.1;%learning rate, equals to 0.1 unless specified
save_pairs = 0; %save in pairs format
```

Now, we can call PANDA

```matlab
AgNet = panda_run(lib_path,exp_file, motif_file, ppi_file, panda_out, save_temp, alpha,
```

The network we downloaded from GRAND database is in `.pairs` format, which is in the form of TF-gene-edge weight in each line. Let's transform it to a TF-by-gene matrix format to be able to compare to our computed network, using a netZooM tool function `Pairs2Mat`

```matlab
networkPair = 'LCL.pairs';
matNet = Pairs2Mat(networkPair,size(AgNet,2),0);
```

Finally, we compare to the output network and we find that indeed they are identical up to a 6-digit precision.

```matlab
tolMat=1e-6;
deltaMat=max(max(abs(AgNet-matNet')));
```

```
assertTrue( deltaMat < tolMat );
```

References

[1] Lopes-Ramos, Camila M., et al. "Regulatory network changes between cell lines and their tissues of origin." *BMC genomics* 18.1 (2017): 723.