DrugDiseaseNet: A package for system-level drug repurposing

Azam Peyvandipour and Sorin Draghici
Department of Computer Science, Wayne State University, Detroit MI 48202

August 5, 2017

Abstract

This package generates a drug-disease network (DDN) composed of the genes that are most likely perturbed by a drug. By performing a system-level analysis on this network using disease gene expression signatures and drug-exposure gene expression signatures, the amount of perturbation caused by a drug on the genes can be estimated. This can help to identify the associations to a disease of interest. Although such network is studied in the context of drug repurposing, it also can be used to identify novel targets for FDA-approved drugs and understanding their mechanism of action.

1 DrugDiseaseNet

This document provides an example code that describes the usage of the package DrugDiseaseNet. This package uses two different sources of data: one is databases of pathways and the other is drug targets and disease related genes. It obtains signaling pathways from Kyoto Encyclopedia of Genes Genomics (KEGG) [2]. A signaling pathway in KEGG is modeled by a graph in which nodes represent genes or proteins, and directed edges between them represent signals between genes or proteins. The edges are weighted based on the various types of signals, such as activation, inhibition, etc. It constructs a global network (GN) by performing the union of all nodes and edges of KEGG human signaling pathways.

Drug targets and disease-related genes (genes associated with the disease of interest) are retrieved from the Comparative Toxicogenomics Database (CTD) [3] and Drugbank [5]. CTD is a database that provides curated data describing cross-species chemical-gene/protein interactions and gene-disease associations. Drugs with no known targets are removed from the study. Such drugs are mostly not FDA-approved.

Next, given the two sets of disease-related genes as $Disease_t = \{x_1, x_2, ..., x_n\}$, and drug targets as $Drug_t = \{y_1, y_2, ..., y_n\}$, we extract a subgraph of GN that consists of all the shortest paths connecting genes belonging to these sets. It means that a gene from either $Disease_t$ or $Drug_t$ can be a source or destination of the shortest path extracted from GN. This subgraph called Drug-disease network (DDN) represents all the interactions between drug targets and genes related to the given disease, through all the interactions described in KEGG signaling pathways.

The impact analysis method [1] can be applied on DDN using the drug and disease gene expressions signatures to generate gene perturbation signatures. The gene perturbation signature is represented by the amount of perturbation estimated upon genes belonging to the drug-disease network (DDN) for all drug-disease pairs.

1.1 KEGG global graph construction

We design the function `keggGlobalGraph` to generate the KEGG global graph. This function uses a copy of KEGG human signaling pathways. We obtained the KEGG pathways and their names using the *ROntoTools* package [4]. The following code will construct a network using the available cached data for the human KEGG signaling pathways.

```
> library(DrugDiseaseNet)
> gg<-keggGlobalGraph()
```

The parameter *updateKEGG* will allow the user to download the latest KEGG signaling pathways as follow:

```
> library(DrugDiseaseNet)
> gg<-keggGlobalGraph(updateKEGG=TRUE)
> gg
```

At this point, gg is a directed graph of class *graphNEL* with weighted edges. The edge weights are 1 for activation/expression signals and -1 for inhibition/repression signals. The nodes and edges labels can be accessed as follow:

```
> library(graph)
> allEdges<-edges(gg)
> allEdges$"hsa:2065"

 [1] "hsa:2549"   "hsa:25759"  "hsa:2885"   "hsa:3091"   "hsa:3716"
 [6] "hsa:3717"   "hsa:399694" "hsa:4609"   "hsa:5290"   "hsa:5291"
[11] "hsa:5293"   "hsa:5295"   "hsa:5296"   "hsa:5335"   "hsa:53358"
[16] "hsa:5336"   "hsa:6464"   "hsa:6714"   "hsa:8503"

> head(nodes(gg))

[1] "hsa:2065" "hsa:2064" "hsa:1956" "hsa:2932" "hsa:1978" "hsa:6198"
```

The code above shows the edges starting from node "hsa:2065". The weight of each edge is obtained using the following code:

```
> library(graph)
> edgeData(gg,from="hsa:2065",to = "hsa:2549" )

$`hsa:2065|hsa:2549`
$`hsa:2065|hsa:2549`$weight
[1] 1

$`hsa:2065|hsa:2549`$subtype
[1] "activation"
```

Another way to obtain the weight of an edge is shown as follow:

```
> library(graph)
> adjmatrix<-as(gg,"matrix")
> adjmatrix [1:4,1:4]
```

```
         hsa:2065 hsa:2064 hsa:1956 hsa:2932
hsa:2065        0        0        0        0
hsa:2064        0        0        0        0
hsa:1956        0        0        0        0
hsa:2932        0        0        0        0
```

```
> adjmatrix ["hsa:2065","hsa:2549"]
```

```
[1] 1
```

1.2 Drug-disease network construction

In this analysis, the inputs are drug targets, disease-related genes, and the KEGG global graph. The output is the directed graph of class *graphNEL* with weighted edges. This graph is a subgraph of KEGG global graph connecting drug-targets and disease-related genes through KEGG signaling pathways.

```
> gg<-keggGlobalGraph()
> #drug target genes can be obtained from The CTD
> drug_targets<-c("hsa:9368","hsa:2322", "hsa:3932", "hsa:4067", "hsa:6714")
> #disease-related genes can be obtained from CTD
> disease_genes<-c("hsa:1832" ,"hsa:5073" ,"hsa:5328")
> drugDiseaseNetwork<-shortestPathsGraph(drug_targets,disease_genes,gg)
> drugDiseaseNetwork
```

```
A graphNEL graph with directed edges
Number of Nodes = 48
Number of Edges = 169
```

2 Funding

This package was supported in part by the following grants: NIH R01 DK089167, NSF 213 DBI-0965741, and R42 GM087013, and by the Robert J. Sokol Endowment in Systems Biology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the funding agencies.

3 Citing

More detail about the proposed approach is discussed in the manuscript which is in the publication process.

References

[1] S. Drăghici, P. Khatri, A. L. Tarca, K. Amin, A. Done, C. Voichiţa, C. Georgescu, and R. Romero. A systems biology approach for pathway level analysis. *Genome Research*, 17(10):1537–1545, 2007.

[2] KEGG. Kyoto Encyclopedia of Genes and Genomes. Technical report, Kyoto University, 2001.

[3] C. Mattingly, M. Rosenstein, G. Colby, J. Forrest Jr, and J. Boyer. The comparative toxicogenomics database (CTD): a resource for comparative toxicological studies. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*, 305(9):689–692, 2006.

[4] C. Voichiţa and S. Drăghici. *ROntoTools: R Onto-Tools suite*, 2013. R package.

[5] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(suppl 1):D668–D672, 2006.