queryMed package: how to display drug interactions

Y. Rivault, O.Dameron, and N. Le Meur

version 0.1 - 17/04/2018

Context

In the area of massive open-data access and big data in general, pharmaco-epidemiology and public health sciences are in need for computational tools. Researchers can now query large medical information systems such as medico-administrative databases and claim databases. Although those information systems are often well structured, their contents are highly codified with several medical terminologies and ontologies, which might be difficult to analyze by non-expert. Pharmaco-epidemiologists and public health scientists thus need batch translation of diverse medical codes like diagnostic codes (based on ICD9 or ICD10), medical procedures or drugs codes (based on ATC nomenclature). In addition to the translation of those codes, their annotation can help making the most of medico-administrative and claim databases analysis by public health researchers. For instance, in a patient care trajectory it might help identifying critical drug interactions that might impair the patient safety. It might also be used to predict evitable hospitalization due to inappropriated care trajectory.

Using ontologies in R has proven to be efficient and useful[Kurbatova et al., 2011], notably in *omic fields of research. For example, many genomes (including the Human genome) are available for downlaod through the BioConductor repository. Those genomes are annotated and statistical analyses of enrichment of standardised terms (commun or closely related) within part of living organisms have help discovering new or impaired functions.

Simmilarly, knowledges in pharmacoepidemiology, drug interactions, indications and contraindications have been combined for clinical data analyzes [Pathak et al., 2013]. However, while works like the Drug Indication Database (DID) and Drug Interaction Knowledge Base (DIKB) have pooled different sources of medical knowledge from the Web of Data [Ayvaz et al., 2015, Sharp, 2017], the use and merging of knowledge is still non trivial to perform due to the multitude of medical classifications and sources of knowledge.

Today the Linked Open Data and the Semantic Web provide technical solutions for the integration of distributed data, their interrogation and their interpretation. For instance, the Ressource Description Framework (RDF) standardise the representation of data and knowledge, and thus allows their sharing and reuse. Furthermore, SPARQL, an other standard from the Semantic Web, provides a way to query these structured data. So these technologies allow reasoning through shared knowledge in the form of ontologies, classifications, and medical thesauri. Although, querying Linked Data in R with SPARQL is feasible, it requiers expertise in the SPARQL language and data schema.

To facilitate this step, we propose a tool to integrate medical ontologies programmatically in the R environment. The queryMed package provides functions and algorithms to query the different sources of medical knowledge representations from the Web of data and to link them to the main medical classifications, for the enrichment and the analysis of medical data. The proposed functions are of two sort: for expert and non-expert. In this vignette, we illustrate the interest of our package to display drug interactions as graphs and explore ATC structure and information.

Applications

To illustrate how to visualize drug interactions with queryMed we explore the drugs that are listed to interact with the platelet antiaggregant CLOPIDOGREL.

First we load queryMed and the DIKB dataset that lists drug interactions gathered from various knowledge databases.

```
library(queryMed)
```

```
##
## Attaching package: 'queryMed'
```

```
## The following object is masked from 'package:base':
##
##     search
```

```
data(DIKB)
head(DIKB, n=3)
```

```
##     drug2   drug1     object precipitant contraindication ddiPkMechanism
## 1 DB00001 DB00636 CLOFIBRATE   LEPIRUDIN            FALSE           <NA>
## 2 DB00001 DB06605   APIXABAN   LEPIRUDIN            FALSE           <NA>
## 3 DB00001 DB00822 DISULFIRAM   LEPIRUDIN            FALSE           <NA>
##   effectConcept label precaution severity                              uri
## 1          <NA>  <NA>       TRUE     <NA> http://rest.kegg.jp/ddi/D00279
## 2          <NA>  <NA>       TRUE     <NA> http://rest.kegg.jp/ddi/D03213
## 3          <NA>  <NA>       TRUE     <NA> http://rest.kegg.jp/ddi/D00131
##   source evidenceStatement    atc1    atc2
## 1   Kegg              <NA> C10AB01 B01AE02
## 2   Kegg              <NA> B01AF02 B01AE02
## 3   Kegg              <NA> N07BB01 B01AE02
```

We subset the dataset and select only the rows where the object (drug of interest) is the CLOPIDOGREL.

```
clopData <- DIKB[DIKB$object=="CLOPIDOGREL",]
dim(clopData)
```

```
## [1] 5096   15
```

```
# unique interactions
dim(unique(clopData[,c("object","precipitant")]))
```

```
## [1] 130   2
```

Overall 5096 rows are selected. Although, some interactions are reported by more than one source, thus it corresponds to 130 unique interactions. The sources are:

```
 table(clopData[,"source"])
```

```
##
##       CredibleMeds    DDI-Corpus-2011    DDI-Corpus-2013
##                  1                  0                  1
##           Drugbank           FrenchDB                HEP
##                 13                  6                  0
##                HIV             NDF-RT         NLM-Corpus
##                  0                  0                  6
##   ONC-HighPriority ONC-NonInteruptive              OSCAR
##                  0                  2                  0
##          PK-Corpus        World-Vista              Kegg
##                  0                 12                119
```

To visualize the interactions we propose to use network representation. Currently, one can look at one drug (object) and its precipitants at a time. The circle layout positionned the object of interest at the center (see Figure below). The thickness of the edge represents the fact that some interactions are reported by more than one source. The graphs below display the interactions listed by at least 2 sources (nbinteractions=2). Some drugs belong to the same ATC family and are represented by the same vertex color. When considering more

general classes of the ATC nomenclature, the interactions concern 11 and 15 classes, respectively (Fig1B, Fig1C). Again some classes are characterized by a high number of interactions (size of the vertex) and sources (thickness of the edge), notably the anti-thrombotic and gastro-esophageal family. In the context of pharmacovigilance, one could then make assumptions about the value of expanding the study of interactions between clopidogrel and the 26 known drugs to those between clopidogrel and the families they belong to.

```
par(mfrow=c(2,2), mar=c(2,2,2,2))
pddi_plot("CLOPIDOGREL", level=5, nbinteractions=2)
```

```
##     atc4               label5 n mypalette
## 1  A01AD acetylsalicylic acid 3   #9E0142
## 2  A02BC           omeprazole 3   #D53E4F
## 3  A02BC         pantoprazole 2   #D53E4F
## 4  A02BC         lansoprazole 2   #D53E4F
## 5  A02BC         esomeprazole 5   #D53E4F
## 6  A10BX          repaglinide 2   #F46D43
## 7  B01AA        acenocoumarol 2   #FDAE61
## 8  B01AD            alteplase 2   #FEE08B
## 9  B01AE dabigatran etexilate 3   #FFFFBF
## 10 B01AF          rivaroxaban 4   #E6F598
## 11 B01AF             apixaban 3   #E6F598
## 12 J05AG           etravirine 2   #ABDDA4
## 13 L01XX           anagrelide 2   #66C2A5
## 14 M03BX             pridinol 3   #3288BD
## 15 N03AB            phenytoin 2   #5E4FA2
```

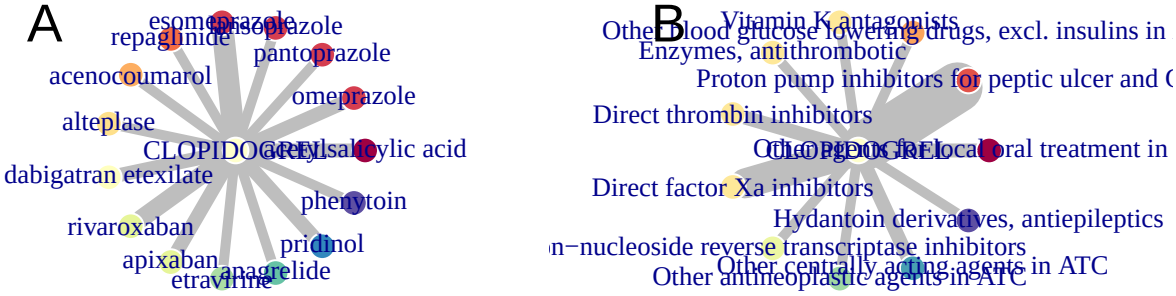
```
text(-1.5,1 ,"A", cex=2)
pddi_plot("CLOPIDOGREL", level=4 , nbinteractions=2)
```

```
##    atc3                                                  label4  n
## 1  A01A              Other agents for local oral treatment in ATC  3
## 2  A02B          Proton pump inhibitors for peptic ulcer and GORD 12
## 3  A10B Other blood glucose lowering drugs, excl. insulins in ATC  2
## 4  B01A                                    Vitamin K antagonists  2
## 5  B01A                                 Enzymes, antithrombotic  2
## 6  B01A                               Direct thrombin inhibitors  3
## 7  B01A                              Direct factor Xa inhibitors  7
## 8  J05A           Non-nucleoside reverse transcriptase inhibitors  2
## 9  L01X                        Other antineoplastic agents in ATC  2
## 10 M03B                     Other centrally acting agents in ATC  3
## 11 N03A                    Hydantoin derivatives, antiepileptics  2
##    mypalette
## 1    #9E0142
## 2    #E25249
## 3    #FBA45C
## 4    #FEE899
## 5    #FEE899
## 6    #FEE899
## 7    #FEE899
## 8    #EDF7A3
## 9    #A1D9A4
## 10   #48A0B2
## 11   #5E4FA2
```

```
text(-1.5,1 ,"B", cex=2)
pddi_plot("CLOPIDOGREL", level=3, nbinteractions=2)
```

```
##   atc2                                                          label3
## 1  A01                                                STOMATOLOGICAL PREPARATIONS
## 2  A02 DRUGS FOR PEPTIC ULCER AND GASTRO-OESOPHAGEAL REFLUX DISEASE (GORD)
## 3  A10                          BLOOD GLUCOSE LOWERING DRUGS, EXCL. INSULINS
## 4  B01                                                ANTITHROMBOTIC AGENTS
## 5  J05                                             DIRECT ACTING ANTIVIRALS
## 6  L01                                  OTHER ANTINEOPLASTIC AGENTS in ATC
## 7  M03                            MUSCLE RELAXANTS, CENTRALLY ACTING AGENTS
## 8  N03                                                       ANTIEPILEPTICS
##    n mypalette
## 1  3   #9E0142
## 2 12   #E25249
## 3  2   #FBA45C
## 4 14   #FEE899
## 5  2   #EDF7A3
## 6  2   #A1D9A4
## 7  3   #48A0B2
## 8  2   #5E4FA2
```

```
text(-2,1 ,"C", cex=2)
```




```
sessionInfo()
```

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
```

```
## [1] fr_FR.UTF-8/fr_FR.UTF-8/fr_FR.UTF-8/C/fr_FR.UTF-8/fr_FR.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] queryMed_0.1
##
## loaded via a namespace (and not attached):
##  [1] zip_1.0.0          Rcpp_0.12.17       compiler_3.5.0
##  [4] pillar_1.2.3       RColorBrewer_1.1-2 plyr_1.8.4
##  [7] bindr_0.1.1        bitops_1.0-6       tools_3.5.0
## [10] digest_0.6.15      jsonlite_1.5       evaluate_0.10.1
## [13] tibble_1.4.2       pkgconfig_2.0.1    rlang_0.2.1
## [16] openxlsx_4.1.0     igraph_1.2.1       yaml_2.1.19
## [19] bindrcpp_0.2.2     rJava_0.9-10       httr_1.3.1
## [22] stringr_1.3.1      dplyr_0.7.6        knitr_1.20
## [25] rprojroot_1.3-2    tidyselect_0.2.4   glue_1.2.0
## [28] data.table_1.11.4  rrdf_2.2.0         R6_2.2.2
## [31] tcltk_3.5.0        rmarkdown_1.10     purrr_0.2.5
## [34] tidyr_0.8.1        magrittr_1.5       rrdflibs_1.4.1
## [37] backports_1.1.2    htmltools_0.3.6    assertthat_0.2.0
## [40] stringi_1.2.3      RCurl_1.95-4.11
```

References

Serkan Ayvaz, John Horn, Oktie Hassanzadeh, Qian Zhu, Johann Stan, Nicholas P. Tatonetti, Santiago Vilar, Mathias Brochhausen, Matthias Samwald, Majid Rastegar-Mojarad, Michel Dumontier, and Richard D. Boyce. Toward a complete dataset of drug–drug interaction information from publicly available sources. *Journal of Biomedical Informatics*, 55:206–217, June 2015. ISSN 15320464. doi: 10.1016/j.jbi.2015.04.006. URL http://linkinghub.elsevier.com/retrieve/pii/S1532046415000738.

N. Kurbatova, T. Adamusiak, P. Kurnosov, M. A. Swertz, and M. Kapushesky. ontoCAT: an R package for ontology traversal and search. *Bioinformatics*, 27(17):2468–2470, September 2011. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btr375. URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr375.

Jyotishman Pathak, Richard C. Kiefer, and Christopher G. Chute. Using linked data for mining drug-drug interactions in electronic health records. *Studies in Health Technology and Informatics*, 192:682–686, 2013. ISSN 0926-9630.

Mark E Sharp. Toward a comprehensive drug ontology: extraction of drug-indication relations from diverse information sources. *Journal of Biomedical Semantics*, 8(1), December 2017. ISSN 2041-1480. doi: 10.1186/s13326-016-0110-0. URL http://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-016-0110-0.