Software Manual

Institute of Bioinformatics, Johannes Kepler University Linz

Rchemcpp - Kernels for molecules

Michael Mahr and Günter Klambauer

Institute of Bioinformatics, Johannes Kepler University Linz
Altenberger Str. 69, 4040 Linz, Austria
michael.mahr@gmail.com

Version 1.0.2, November 19, 2012

Contents

1 Introduction

The Rchemcpp package is part of the CRAN project. The functionality of Chemcpp (http://chemcpp.sourceforge.net/html/index.html) is provided in R, that is the computation of similarities between molecules by kernel functions. The following kernels are implemented:

- the marginalized graph kernel between labeled graphs (Kashima *et~al.*, 2004).
- extensions of the marginalized kernel (Mahé *et~al.*, 2004).
- Tanimoto kernels (Ralaivola *et~al.*, 2005).
- graph kernels based on tree patterns (Mahé and Vert, 2009).
- kernels based on pharmacophores for 3D structure of molecules (Mahé *et~al.*, 2006).

2 Getting started and quick start

To load the package, enter the following in your R session:

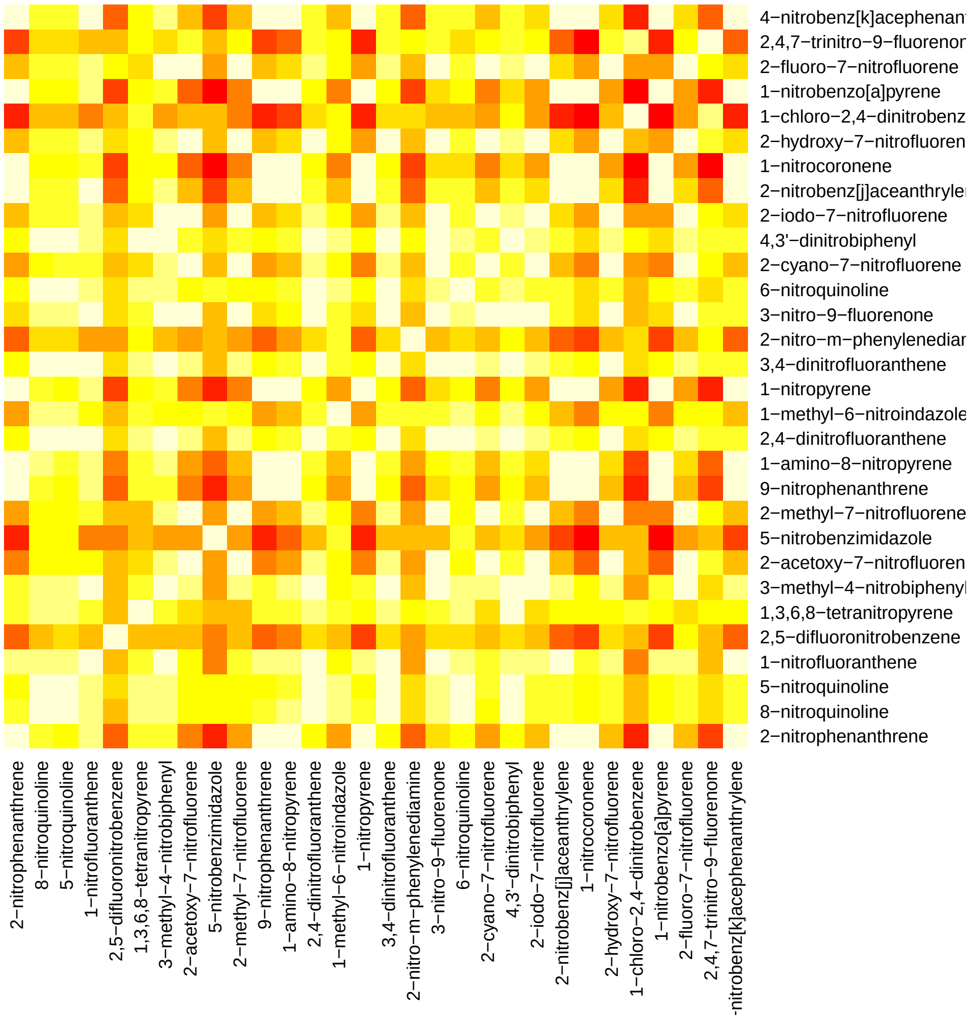
```
> library(Rchemcpp)
```

We enter the filename of and SDF file to the function sd2gram. This function computes the similarity of the molecules with the marginalized kernel (Kashima *et~al.*, 2004) approach.

```
> sdfolder <- system.file("sample_data",package="Rchemcpp")
> sdf <- list.files(sdfolder,full.names=TRUE,pattern="small")
> K <- sd2gram(sdf,returnNormalized=TRUE)
```

The similarity values are now stored in K. We visualize this matrix as a heatmap.

```
> heatmap(K,Rowv=NA,Colv=NA,scale="none")
```

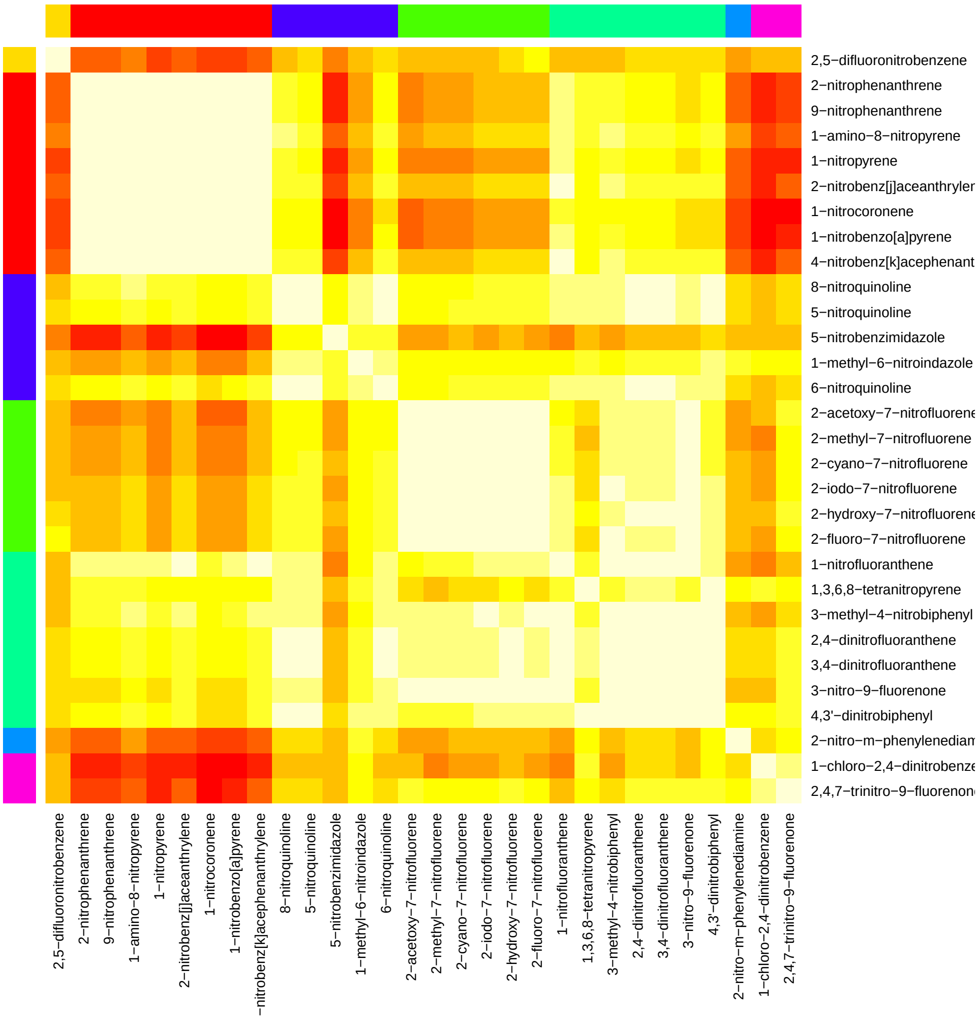

4−nitrobenz[k]acephenan
2,4,7−trinitro−9−fluorenor
2−fluoro−7−nitrofluorene
1−nitrobenzo[a]pyrene
1−chloro−2,4−dinitrobenz
2−hydroxy−7−nitrofluoren
1−nitrocoronene
2−nitrobenz[j]aceanthryle
2−iodo−7−nitrofluorene
4,3'−dinitrobiphenyl
2−cyano−7−nitrofluorene
6−nitroquinoline
3−nitro−9−fluorenone
2−nitro−m−phenylenediar
3,4−dinitrofluoranthene
1−nitropyrene
1−methyl−6−nitroindazole
2,4−dinitrofluoranthene
1−amino−8−nitropyrene
9−nitrophenanthrene
2−methyl−7−nitrofluorene
5−nitrobenzimidazole
2−acetoxy−7−nitrofluoren
3−methyl−4−nitrobiphenyl
1,3,6,8−tetranitropyrene
2,5−difluoronitrobenzene
1−nitrofluoranthene
5−nitroquinoline
8−nitroquinoline
2−nitrophenanthrene
2−nitrophenanthrene
8−nitroquinoline
5−nitroquinoline
1−nitrofluoranthene
2,5−difluoronitrobenzene
1,3,6,8−tetranitropyrene
3−methyl−4−nitrobiphenyl
2−acetoxy−7−nitrofluorene
5−nitrobenzimidazole
2−methyl−7−nitrofluorene
9−nitrophenanthrene
1−amino−8−nitropyrene
2,4−dinitrofluoranthene
1−methyl−6−nitroindazole
1−nitropyrene
3,4−dinitrofluoranthene
2−nitro−m−phenylenediamine
3−nitro−9−fluorenone
6−nitroquinoline
2−cyano−7−nitrofluorene
4,3'−dinitrobiphenyl
2−iodo−7−nitrofluorene
2−nitrobenz[j]aceanthrylene
1−nitrocoronene
2−hydroxy−7−nitrofluorene
1−chloro−2,4−dinitrobenzene
1−nitrobenzo[a]pyrene
2−fluoro−7−nitrofluorene
2,4,7−trinitro−9−fluorenone
-nitrobenz[k]acephenanthrylene

3 Molecular similarity for clustering

Based on the similarity measure we can run clustering algorithms on the data in order to find groups among the molecules. We use Affinity Propagation Clustering (Frey and Dueck, 2007) as implemented by Bodenhofer *et~al.* (2011) for this task, because the cluster centers are real molecules.

```
> library(apcluster)
> r <- apcluster(K)
```

```
> plot(r,K)
```



References

Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. *Bioinformatics*, **27**, 2463–2464.

Frey, B.~J. and Dueck, D. (2007). Clustering by passing messages between data points. *Science*, **315**, 972–977.

Kashima, H., Tsuda, K., and Inokuchi, A. (2004). Kernels for graphs. *Kernel methods in computational biology*, **39**(1), 101–113.

Mahé, P. and Vert, J.-P. (2009). Graph kernels based on tree patterns for molecules. *Mach. Learn.*, **75**(1), 3–35.

Mahé, P., Ueda, N., Akutsu, T., Perret, J.-L., and Vert, J.-P. (2004). Extensions of marginalized graph kernels. In R.~Greiner and D.~Schuurmans, editors, *Proc of the 21st ICML*, pages 552–559. ACM Press.

Mahé, P., Ralaivola, L., Stoven, V., and Vert, J.-P. (2006). The pharmacophore kernel for virtual screening with support vector machines. *J. Chem. Inf. Model.*, **46**(5), 2003–2014.

Ralaivola, L., Swamidass, S.~J., Saigo, H., and Baldi, P. (2005). Graph kernels for chemical informatics. *Neural Netw*, **18**(8), 1093–1110.