stemmatology: An **R** Stemmatology Package

Jean-Baptiste Camps, Florian Cafiero

Abstract

Stemmatology is the name of the field dedicated to studying text genealogies and establishing genealogical tree-like graphs known as stemma codicum.

This package includes various functions for stemmatological analysis. It particularly implements functions following the Poole-Camps-Cafiero method, as well as functions to import data.

Keywords: stemmatology, philology, network, graphs.

---

1. Input

Most of the functions take, as input a *numeric matrix*, with witnesses in columns, variant locations in rows, and readings coded by a number, e.g.

	A	B	C	D	E	H	I	J	K	O
1	0	1	1	1	NA	1	1	NA	1	1
2	1	1	1	1	NA	1	1	NA	1	1
3	1	1	1	1	NA	1	1	NA	1	1
4	1	1	1	2	NA	1	1	NA	1	1
5	1	1	1	2	NA	1	1	NA	1	1
6	1	1	1	1	NA	1	1	NA	1	1

where $A, B, \ldots, O$ are the various witnesses in columns, $1 \ldots 6$ the various variant locations, in rows, and the differents readings are coded either 0 (omission), $1, 2, \ldots, n$. `NA` is used for the lack of information (physical lacuna, absence of observation, variant location not applicable to a given witness, etc.).

Alternatively , if `alternateReadings = TRUE`, the input can be a *character matrix*, with witnesses in columns, variant locations in rows, and, in each cell, one or several readings, coded by numbers and separated by a comma (e.g. '1,2,3', if the witness has three different readings), e.g.

	A	D	F	T	P
1	"1"	"2"	"2"	"2"	"1,2"
2	"1"	"2"	"1,2"	"2"	"1"
3	"1"	"1"	"1"	"1"	"2"
4	"1,3"	"1,2"	"1"	"2"	"3"

Notice how a witness can bear several readings (e.g., P at VL 1).

1.1. Create or import data

Data can be created inside R or imported. They can be imported by reading a csv file, for instance (e.g. with `read.csv`). They can also be imported from a TEI encoded apparatus in parallel-segmentation, either by using an XSL stylesheet, or the built-in function `import.TEIApparatus`.

The function `import.TEIApparatus` allows to import a TEI P5 encoded apparatus into a stemmatological matrix usable with other functions. It has some parameters to refine the import (variant types, . . .), and can read either from disk or from an URL.

2. PCC Method

Functions are made available for the PCC method (See Camps and Cafiero 2014 or `PCC` for more details). The most important are

`PCC` global shell for the PCC functions;

`PCC.Exploratory` global function for exploratory methods of the PCC family;

`PCC.Stemma` Building the Stemma Codicum.

3. Other functions

The package contains also various other functions, particularly aimed at detecting contamination. It contains for instance the function `PCC.contam`.

The package aims at making available various other stemmatological methods, including further functions for contamination detection, or for theoretical stemmatology.

References

Camps, Jean-Baptiste, and Florian Cafiero. 'Stemmatology: An R Package for the Computer-Assisted Analysis of Textual Traditions'. *Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CRH-2)*, edited by Andrew U. Frank et al., 2018, pp. 65–74, https://halshs.archives-ouvertes.fr/hal-01695903v1.

Camps, Jean-Baptiste. 'Detecting Contaminations in Textual Traditions Computer Assisted and Traditional Methods'. Leeds, International Medieval Congress, 2013, unpublished paper, https://www.academia.edu/3825633/Detecting_Contaminations_in_Textual_Traditions_Computer_Assisted_and_Traditional_Methods.

Camps, Jean-Baptiste, and Florian Cafiero. 'Genealogical Variant Locations and Simplified Stemma: A Test Case'. *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, edited by Tara Andrews and Caroline Macé, Brepols, 2015, pp. 69–93, https://halshs.archives-ouvertes.fr/halshs-01435633, DOI: 10.1484/M.LECTIO-EB.5.102565.

Poole, Eric. 'L'analyse stemmatique des textes documentaires'. *La pratique des ordinateurs dans la critique des textes*, Paris, 1979, p. 151-161.

Poole, Eric, 'The Computer in Determining Stemmatic Relationships'. *Computers and the Humanities*, 8-4 (1974), p. 207-16.

Bugs and Issues

Please report issues with this package to https://github.com/Jean-Baptiste-Camps/stemmatology.

Example of use

```
> # Interactive mode
> # Load data
> data(fournival)
> # or alternatively, import it
> fournival = import.TEIApparatus(file = "myFournival.xml",
+     appTypes = c("substantive"))
> # Analyse it with the PCC functions
> PCC(fournival)
> # Complete step-by-step non interactive use
> data("fournival")
> # look for conflicts
> myConflicts = PCC.conflicts(fournival)
> # remove conflicting VL
> myConflicts = PCC.overconflicting(myConflicts, ask = FALSE, threshold = 0.06)
> myNewData = PCC.elimination(myConflicts)
> # look for competing genealogies
> myConflicts = PCC.conflicts(myNewData)
> myNewData = PCC.equipollent(myConflicts, ask = FALSE, scope = "W", wits = "D")
> # build a stemma
> PCC.Stemma(myNewData$databases[[3]], ask = FALSE)
```

Affiliation:

JB Camps
École nationale des chartes
jbcamps@hotmail.com
URL: www.chartes.psl.eu/jean-baptiste-camps