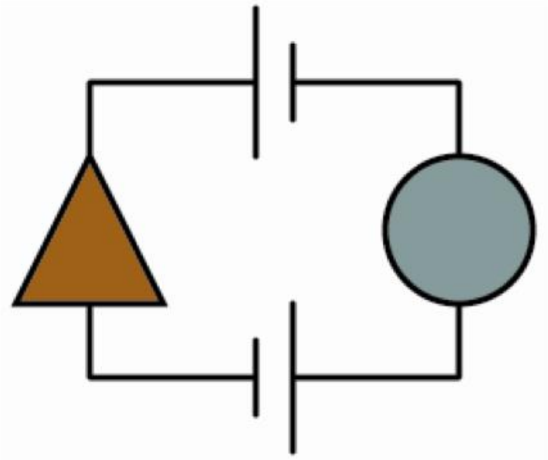


PUCK
Program for the Use and Computation of Kinship data

Written by Klaus Hamberger,
Christian Momon, Edoardo Savoia et Telmo Menezes (in Java 1.6)

Info and download : http://www.kintip.net/
Help : puck.aide@yahoo.fr

Summary

Introduction

Puck is a computer program for analyzing genealogical and other kinship related data and a continuation of the gradual introduction of computers in kinship studies since the 1970s ((Hamberger, Houseman, & Grange, 2009, p. 108). Puck is a product of the research group TIP and an outcome of the project « Informatical treatment of kinship phenomena » (« Traitement Informatique des phénomènes de parenté », 2006- 2009) financed by the French National Research Agency (ANR). The group's work revolved around a methodological and epistemological reflection on empirical data collection and analysis, based on the observation that "a paradox is at the heart of studies on kinship: the marriage choices, which as the foundation of kinship systems should be the primary object were neglected especially in their empirical dimension. The anthropology of kinship began as an analysis of terminology and developed into a science of rules and norms. Today it appears as a multifaceted research field dealing with representations and institutions, political strategies and symbolic operations. Still the actual practices that generate matrimonial networks continue to occupy a marginal position"(Hamberger et al., 2009, p. 107). Kinship cannot be treated as isolated elements, the use of a computer processing of matrimonial practices proved fundamental.

Starting from this premise, the software allows not only to make very fine and precise analyses of the matrimonial structures and configurations of a genealogical network, but also to consider the quality and biases: the data processing of kinship relationships is thus an opportunity to consider the theoretical issues and methodological choices, as it respects the individual way of the collection and organization of field data or records, to assess the quality of the dataset studied[1].

Puck accompanies the researcher throughout his work, from data input to final analysis to the final analysis. It is compatible with most commonly used formats (Excel, Gedcom, Pajek, etc.), and allows to import or export files in each of these formats. Puck has been written in Java 1.6 and is continuously updated[2].

Puck's main functionalities:

- Management and editing of kinship networks
- Navigation through kinship networks
- Diagnostics of kinship datasets
- Census and analysis of matrimonial and other circuits and chains in kinship networks
- Transformation, segmentation and fusion of kinship networks
- Simulation of kinship and alliance networks

[1] A presentation of the software and its issues can be found in a recent article in *l'Homme* (Hamberger, Houseman, & Grange, 2009).

[2] See the website www.kintip.net.

Installation

Download

The software is free to download on the website www.kintip.net, where you will find the latest version (PUCK 2.0) and earlier ones (PUCK 1.0). Puck 1.0 uses the external libraries JFreechart and JCommons for graphical visualization and Jxl for importation from and exportation to excel files (without them the functions involving diagrams will not work). Click to download the zip file « Lib » that contains them and put them into the same directory as the file containing the executable file « Puck.jar ».

Note for Puck 1.0 : the external jar files may be left inside or outside the « lib » directory. For Mac users however it is recommended to put them outside the « lib » directory. Puck is contained in an executable jar file and is started by double-clicking on the file icon.; For the treatment of large datasets (more than 10.000 individuals) download the batch file Puck512_bat or (for very large datasets with more than 30.000 individuals) Puck1024_bat to augment the memory of your Java virtual machine up to 512 MB or 1024 MB. Put it into the same directory as the file Puck.jar and double-click on its icon to start Puck.

System requirements

Make sure that your computer is equipped with an adequate version of Java. Java is compatible with most computers and operating systems, so Puck works on PC (Linux, Solaris, Windows) and Mac.

Languages

Puck can be run in English, French, German, Italian and Spanish. You can choose the language via the tab *Edit* → *Preferences* → *Language*

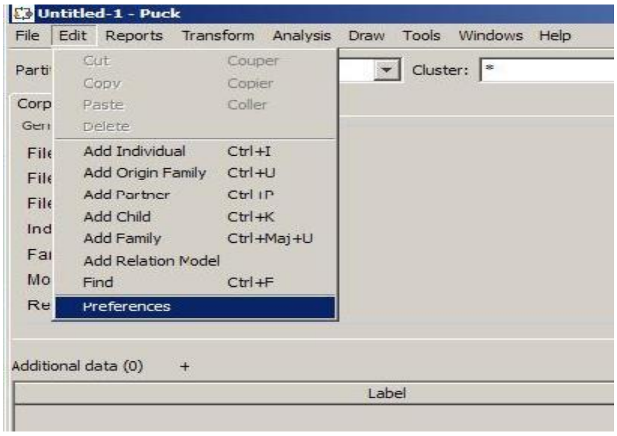


Fig. 1 : preferences

Codes and characters

By default, Puck assumes that the file is UTF-8. You can choose the encoding (*File* → *Open encoding*)

Contact

For questions regarding the software or to report a bug, please send an email to puck.aide@yahoo.fr.

Licence

Puck is a free software, distributed under CeCILL licence (a french variant of GPL). Puck has been deposited at the APP (*Agence pour la Protection des Programmes*, Program Protection Agency) and is protected by French law.

Disclaimer

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS « AS IS » AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

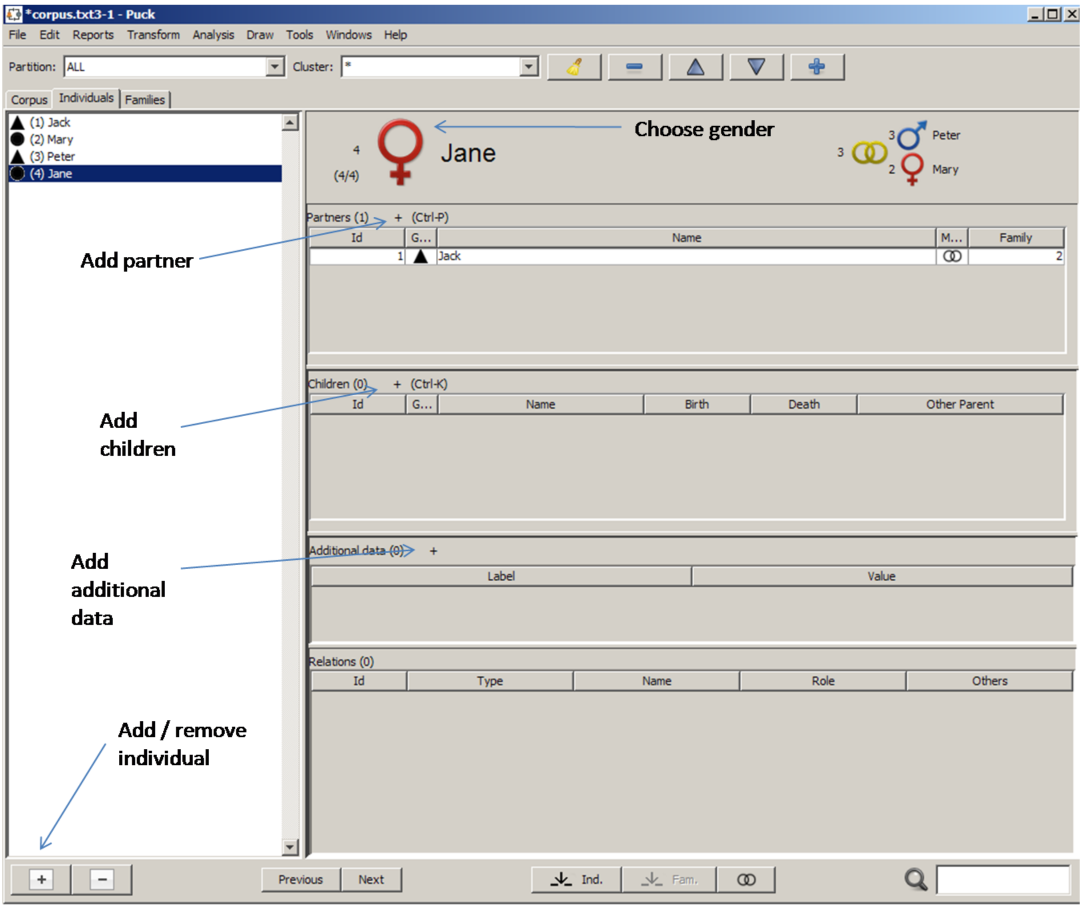
1. Start with Puck

a) Enter a new dataset

Note: before starting, see also How to establish a kinship dataset.

It is possible to enter data directly into Puck, but you can also use a software or a format of your choice, because the program is compatible with most popular formats. To do this, click on tab *File* → *new* → *empty network* and save it in the directory of your choice by giving it a name (*File* → *save as*). You can save the dataset in different formats (GED, txt, TIP, xls, etc.).

When the dataset is created, you can begin entering data: you create those as individuals, families (CTRL+MAJ+U), additional data (place of birth, occupation, religion, date of birth, etc.) and navigate in the dataset by clicking on individuals and families.



▲ (1) Jack
● (2) Mary
▲ (3) Peter
● (4) Jane

Id	G...	Name	M...	Family
1	▲	Jack	⚭	2

Choose gender

Add partner

Add children

Add additional data

Add / remove individual

Fig. 2 : enter data

Keyboard shortcuts :

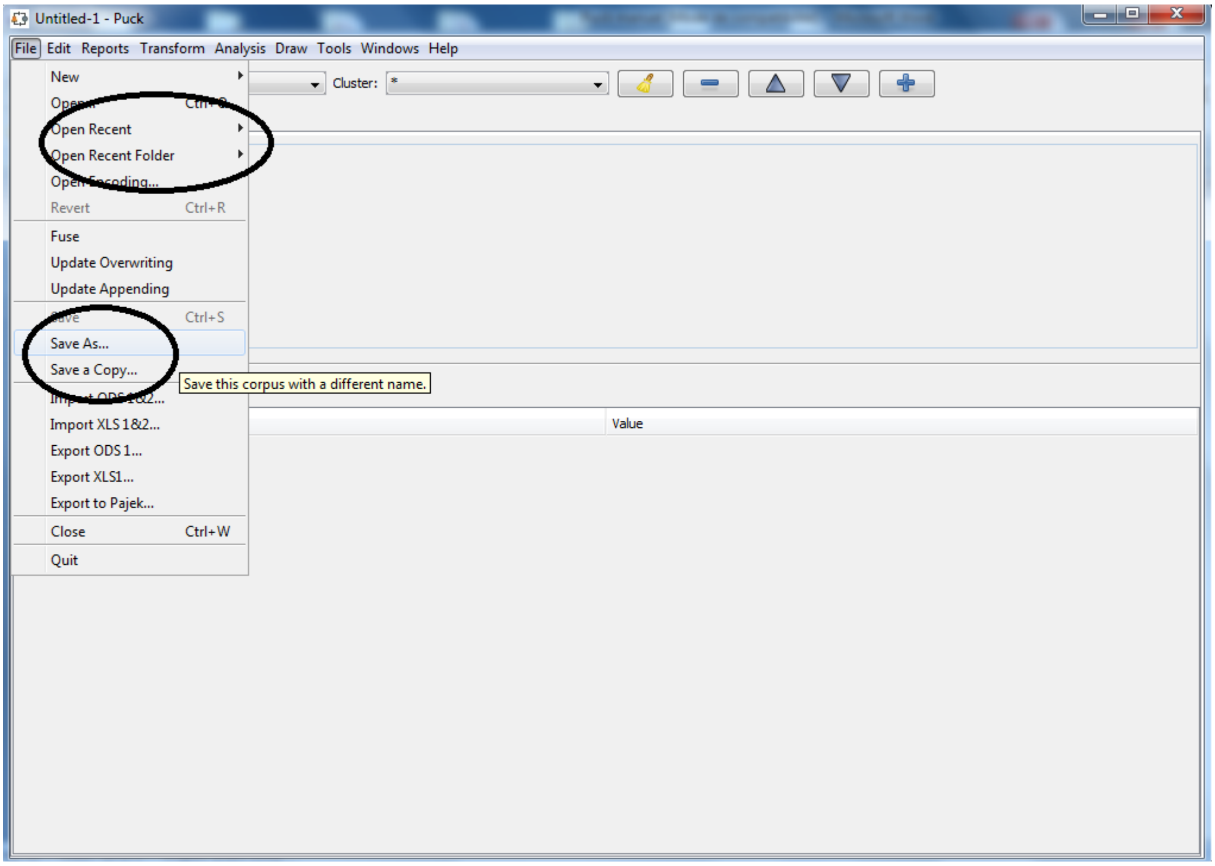
Add individual: Ctrl + I
Add Origin family Ctrl + U
Add partner Ctrl + P
Add child Ctrl + K
Add family Ctrl + Maj + U
Find: Ctrl + F

b) *Import dataset*

If you already have a family network, you can import it into Puck, that can read many files : text, excel, pajek, gedcom, tip, xml and prolog format. Import a dataset by clicking *File* → *Open* and choose your file. The data appears in the main window. It is possible to open a recently used file (*File* → *Open recent*).

c) *Save and export dataset*

You can save the dataset to the format of your choice: Gedcom, Pajek, Text (1 & 2-mode, Type 3, TIP, etc. (*File* → *Save as* / *Export*). Using *File* → *Save* you save as much as information exporting in .puc format.



Save this corpus with a different name.

Fig. 3 : import and save

d) *Update and merge dataset*

Update

Genealogical corpuses can be updated and new supplementary information added to them. To update a corpus and add data from another dataset, Puck provides a specific function : *Update*. For safety, Puck will generate a new dataset so that the original data are not erased by mistake, and you will store the new data in a new file in the same format as the main dataset. A corpus update requires a file which fulfills the same format requirements as the files used to load a corpus.

Note: if you use a file in text format containing only supplementary information on individuals' properties (so that the first block is empty), make sure that the file begins with two headlines (and not just one). The first headline (which may consist in a single letter) marks the presence of the first block, the second one the switch to the second block. Otherwise Puck will read your supplementary information as basic genealogical information, and the update will fail. These problems do not arise when you use files in tip format.

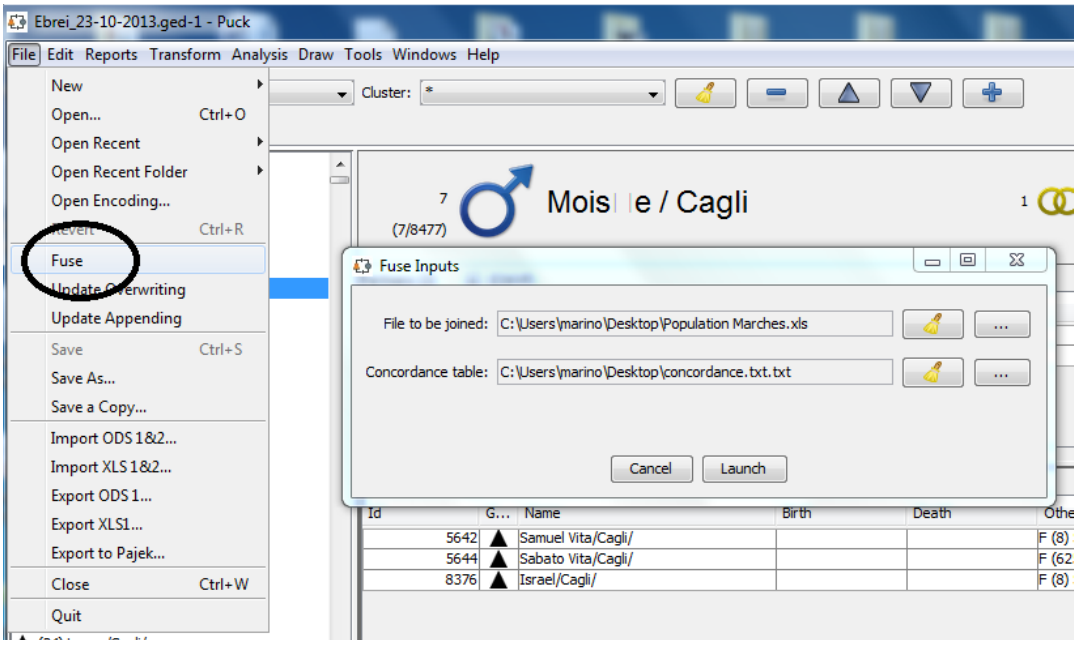
You have two choices: "happened" or "overwrite" (*File* → *Update overwriting / appending*). In the first case, data are simply appended without overwriting existing data (in this case, certain data may not be added, for instance a father that does not correspond to the actual father). In the second case, data are used to overwrite existing data by new ones (in this case, an individual which appears in the update file as having no father will lose its father in the corpus).

Fusion

You can merge two datasets using *File* → *Fuse* to produce a new corpus by fusing the current dataset with another dataset. This requires two files :

- a second dataset to be fused with the current corpus ("File to be joined")
- a concordance table which serves to identify the individuals that belong to both datasets ("Concordance table" : this is a simple text file with two columns containing the numbers of individuals in the second and the first corpus. Individuals that appear in only one corpus do not have to be listed in this table, and numbers need not to be ordered)

Warning: *merging two datasets implies an automatic renumbering of individuals: individuals of the second corpus who have no double in the first corpus obtain a new identity number by adding the number of the last individual of the first corpus to their old number. This renumbering should remain a transitory exception while establishing a definite dataset. As a rule, individuals should have one unique identity number and belong to one unique corpus.*



File to be joined: C:\Users\marino\Desktop\Population Marches.xls

Concordance table: C:\Users\marino\Desktop\concordance.txt.txt

Id	G...	Name	Birth	Death	Othe
5642	▲	Samuel Vita/Cagli/			F (8)
5644	▲	Sabato Vita/Cagli/			F (62
8376	▲	Israel/Cagli/			F (8)

Fig. 4 : merge a dataset

e) Transform a dataset

It is possible to perform a number of operations from the original dataset to duplicate and anonymize it (if working on a recent population and we do not want their identity appears) or to reduce it by eliminating individuals whose presence is not necessary (unmarried people, "structural children" that is to say individuals who have neither spouses nor children, etc.).

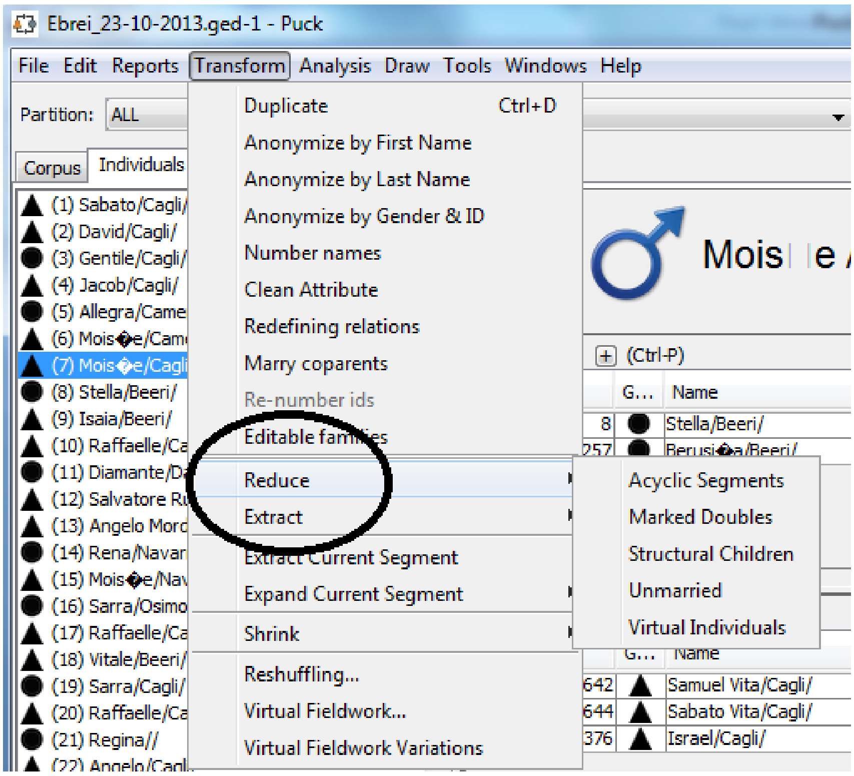


Fig. 5 : transform a dataset (reduce/extract)

Main window components (*Transform* → *reduce / Extract*)

- Anonymize by First / Last Name or Gender & ID
- Number names: adds the identity number between parentheses to the original name (e.g. "Pierre Dupont (5)").

Note: Numbered names are convenient if the corpus is exported in pajek format (pajek files require renumbering of individuals in order to assure continuous vertex numbers. Numbered names thus serve to keep the original numbers). If a pajek file with numbered names is imported to Puck, numbers between parentheses are automatically re-converted into identity numbers.

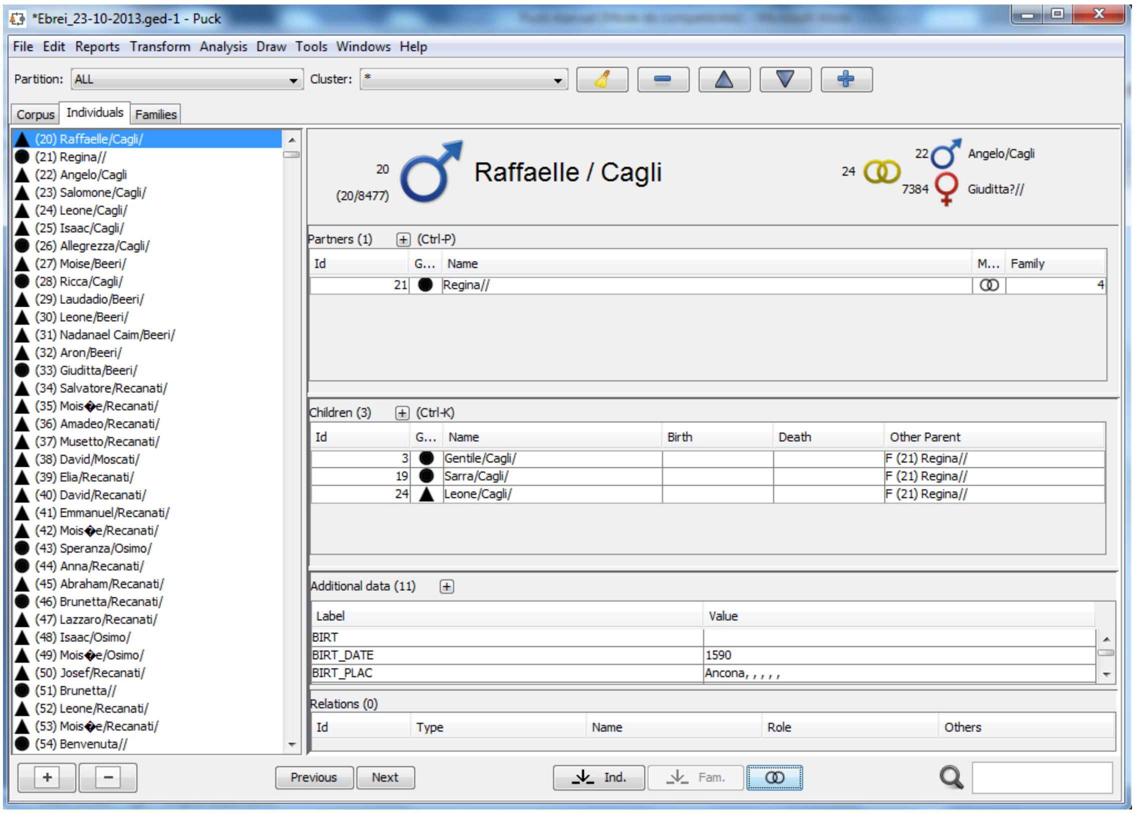
- *Reduce* → *Eliminates all virtual individuals* (individuals that only serve to represent the common parents of full siblings). Virtual individuals are individuals for whom no information is available except for their kinship relations (and perhaps their gender, if they are parents or spouses of existing individuals). This reduction is recommended for the exploratory analysis of a corpus containing fictive individuals (e.g Gender bias).
- *Reduce* → *Unmarried* : eliminates all individuals who are not married.
- *Reduce* → *Structural children* : eliminates all individuals who have neither spouses nor children (the structural children of the kinship network).
- *Reduce* → *Marked doubles* : doubles are individuals who are identical with another individual. The individual with the lower identity number is considered as the original, the one with the higher ID number as the double. Doubles can be marked by substituting the original's identity number for their name. They can be eliminated from the dataset by this function.

- *Extract* → *Kernel* : reduces the kinship network to its kernel

- *Extract* → *Max. bicomponent* : reduces the kinship network to its maximal bicomponent
- *Extract* → *Core* : reduces the kinship network to the sum of its matrimonial bicomponents, the core
- Extract → by cluster size / value : reduced the network to those vertices who have a positive cluster value in a chosen partition. For example, if "OCCU" (occupation) is chosen, then the network is reduced only to people who have an occupation.

f) Navigate in a dataset

The main window has three tabs : *corpus*, *individuals* and *family*. The *corpus* tab contains the general information of the dataset, the *Individuals* tab the list of all individuals, the information about them, and *family* the list of all families in the dataset. It is possible to add specific relations between them.



Partition: ALL

Cluster: *

▲ (20) Raffaelle/Cagli/
● (21) Regina//
▲ (22) Angelo/Cagli
▲ (23) Salomone/Cagli/
▲ (24) Leone/Cagli/
▲ (25) Isaac/Cagli/
● (26) Allegrezza/Cagli/
▲ (27) Moise/Beeri/
● (28) Ricca/Cagli/
▲ (29) Laudadio/Beeri/
▲ (30) Leone/Beeri/
▲ (31) Nadanael Caim/Beeri/
▲ (32) Aron/Beeri/
● (33) Giuditta/Beeri/
▲ (34) Salvatore/Recanati/
▲ (35) Mois�e/Recanati/
▲ (36) Amadeo/Recanati/
▲ (37) Musetto/Recanati/
▲ (38) David/Moscati/
▲ (39) Elia/Recanati/
▲ (40) David/Recanati/
▲ (41) Emmanuel/Recanati/
▲ (42) Mois�e/Recanati/
● (43) Speranza/Osimo/
● (44) Anna/Recanati/
▲ (45) Abraham/Recanati/
● (46) Brunetta/Recanati/
▲ (47) Lazzaro/Recanati/
▲ (48) Isaac/Osimo/
▲ (49) Mois�e/Osimo/
▲ (50) Josef/Recanati/
● (51) Brunetta//
▲ (52) Leone/Recanati/
▲ (53) Mois�e/Recanati/
● (54) Benvenuta//

20 (20/8477) ♂ Raffaelle / Cagli

24 22 ♂ Angelo/Cagli
7384 ♀ Giuditta?//

Partners (1) + (Ctrl-P)

Id	G...	Name	M...	Family
21	●	Regina//	⚭	4

Children (3) + (Ctrl-K)

Id	G...	Name	Birth	Death	Other Parent
3	●	Gentile/Cagli/			F (21) Regina//
19	●	Sarra/Cagli/			F (21) Regina//
24	▲	Leone/Cagli/			F (21) Regina//

Additional data (11) +

Label	Value
BIRT	
BIRT_DATE	1590
BIRT_PLAC	Ancona, , , , ,

Relations (0)

Id	Type	Name	Role	Others

Fig. 6 : dataset's main window

The identity section of individuals contains :

- The individual's identity number. Each individual has a unique identity number which serves to identify it in all contexts (such as matrimonial rings, or in the kinship section of another individual's page) and ideally should never be changed. By clicking on the identity number, a drop-down menu appears showing identity numbers and the corresponding names. By choosing an individual in this drop-down menu (or directly entering its identity number) one can jump to the chosen individual's page.
- The individuals gender : circle for women, triangles for men and squares for individuals whose gender is unknown or commons' symbols (♂ for men, ♀ for women and ⚲ for unknown). You can change the sex of individuals by clicking on the symbol.
- The individual's first name and last name. When importing or exporting kinship data, Puck identifies separate parts of names by a slash "/" separating them. The name part

before the first slash is identified as first name, all other parts are displayed in the "last name" field. However, any number of names can be distinguished by using the slash separator. The two name fields are auto-completing. Clicking on them opens a drop-down menu where all first and last names of the corpus are displayed, to facilitate data entry.

The window contains information on the immediate (first degree) relatives of Ego: one for parents, one for spouses, one for children. The headline bar above a kinship section indicates also the number of the respective relatives (number of spouses or children). By clicking on the name, one can jump to the relative's individual page. Jumping from one relative to another in this manner is a way to navigate through the corpus along kinship paths.

The attribute sections contain supplementary information on the individual, which can also be used to partition the corpus. All these attributes are exogenous properties (date and place of birth / marriage / death events, occupation, etc.).

Puck distinguishes three different categories of attributes and regroups them in separate sections. All attributes are characterized by a label (for which standard individual property codes should be used) and a value. For instance, a property with the label "OCCU" and the value "merchant" means that the individual's occupation is that of a merchant. When importing gedcom files, relational properties are automatically transformed into properties of the concerned individuals with reciprocal alter specification. For a list of standard property codes and guidelines for personalized property codes click here.

Note: even an ego's property with specification of alter and automatic attachment of a reciprocal property to alter with specification of ego still remains a combination of individual properties and is not a relation property. This is important when these properties are used to partition the corpus or to restrict a relational or matrimonial census to a subcorpus. For instance, search results for matrimonial rings among individuals whose marriages lie within a certain period may well contain marriages outside that period, if both partners have been married before or after.

The last number of a date is automatically interpreted as year (important for partitioning). This must be taken into account when specifying events as having happened before, after, or around a certain year. One and the same label cannot be used at the same time for a simple property and for an event property. For example, the code MARR cannot be used at the same time to indicate if a person is married (simple property) and when, where and whom he or she is married to (event property). "Notes" are simple properties with the only difference that they allow entering long texts and line breaks. The label of notes ("NOTE") is not displayed and cannot be chosen or changed.

All attribute fields (with the exception of notes) contain drop-down menus which allow choosing among existing label, value, place, and date data and existing individuals (for alter). The drop-down menu for alter contains identity numbers and names of all individuals in the corpus, just as for kinship relation entries.

Search dialog box

Searching individuals by name: this can be done by entering the name or name part of the searched individual(s) into the input field at the right below the dataset's main window (in the individuals / families section). If several individuals fit (for instance, when a family name has been entered), successive "enter" clicks permit to pass from one selected individual to another, and thus to navigate through the selected part of the corpus.

Data deletion

Use right click → *Delet* to remove an individual, a kinship relation or an individual's attribute from the dataset.

g) Explore a dataset

In addition to data entry and navigation through the corpus, it is possible to explore the kinship environment of individuals:

Ascendants / Progeniture (Analysis → Pedigree / Progeniture)

Getting an individual's pedigree / progeniture up to a certain degree: This is done by entering a single number that indicates the maximal generational distance of ascendants/descendants. The pedigree / progeniture is then listed in a tree structure. For instance, entering "3" produces the tree of known ascendants / descendants up to great-grandparents.

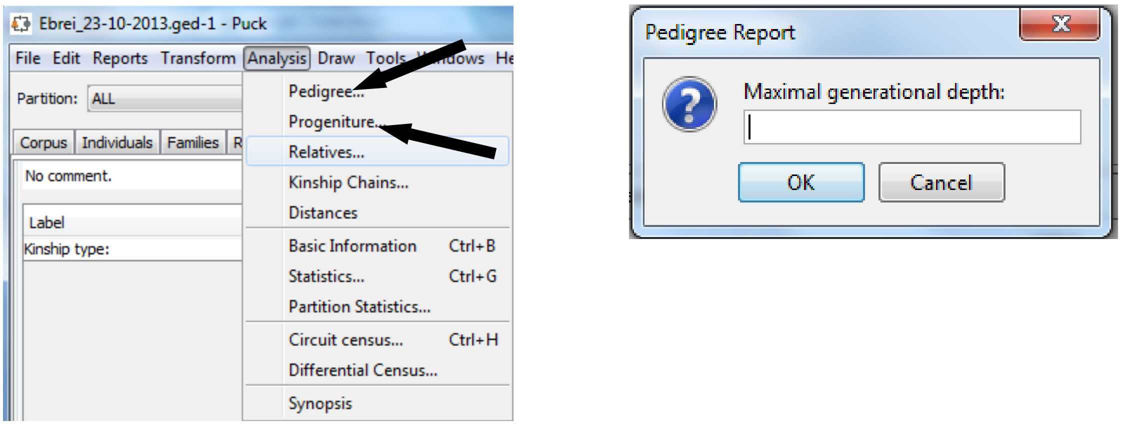
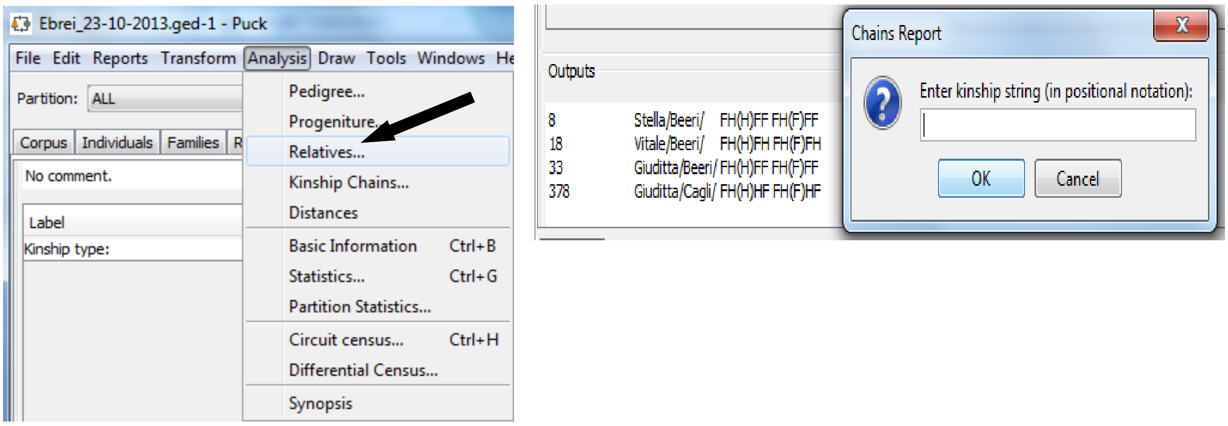


Fig. 7 : Ascendants / Progeniture

Relatives (Analysis → Relatives)

Getting an individual's relatives of a certain type: This is done by entering a structure formula in positional notation, just as in the case of a matrimonial or relational census. Note, however, that the present function is ego-centered: The first individual in the chain is the individual of the current page. For instance, entering "XX(X)XX" (*Enter kinship string* (in positional notation)) will produce the list of all cousins of the individual (with their names, identity numbers, and exact kinship relation types).

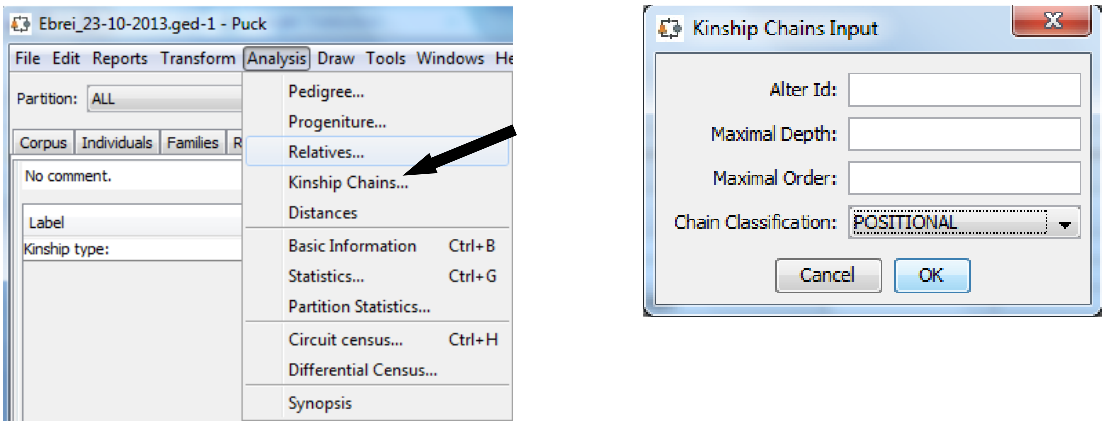


```
8      Stella/Beeri/   FH(H)FF FH(F)FF
18     Vitale/Beeri/   FH(H)FH FH(F)FH
33     Giuditta/Beeri/ FH(H)FF FH(F)FF
378    Giuditta/Cagli/ FH(H)HF FH(F)HF
```

Fig. 8 : relatives

Chains (*Analysis → Kinship Chains*)

Getting the kinship chains to another individual. This is done by entering two numbers, where the first one is alter's identity number, and the second the canonical degree. A third number can be entered which specifies the maximal width of the chains, that is, the maximal number of marriages they may contain. Puck lists all tracks between ego and alter within specified bounds in the classification of your choice.



Alter Id:

Maximal Depth:

Maximal Order:

Chain Classification: POSITIONAL

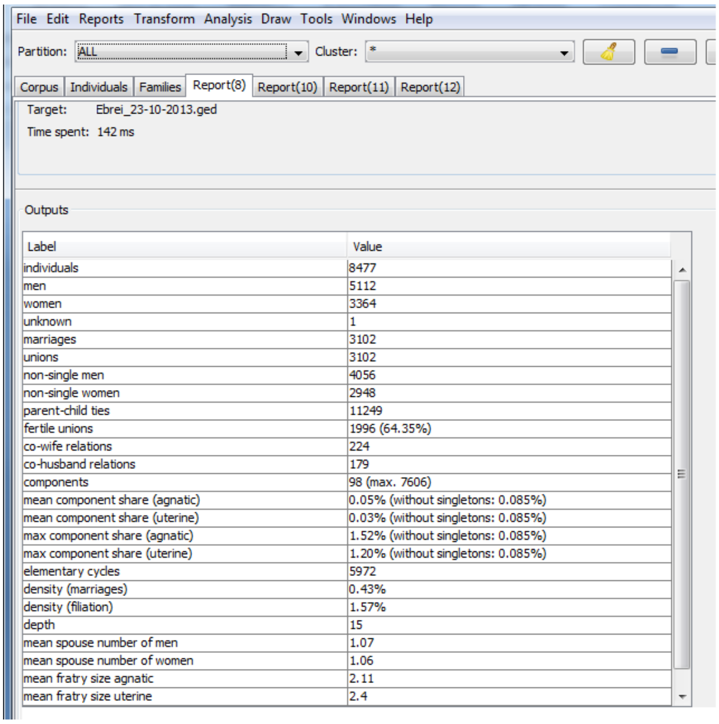
Fig. 9 : Kinship chains

2. Dataset diagnostic

Genealogical corpora produced by researchers are not neutral objects. The genealogies constructed from informants or sources are often incomplete and androcentered, implying an asymmetry of relations in the network. It is very important to detect errors of data collection or data entry to correct them, to know biases and relativize the raw results, (Barry & Gasperoni, 2008; Hamberger & Gargiulo, 2013; Hamberger, Houseman, & White, 2012, p. 546‑547). The first step before analyzing a family dataset is therefore to establish a profile of the network, from basic tools (count of the population and demographic composition, gender distribution, genealogical depth, density) to more complex tools (family completeness, gender bias), not only on the overall network, but also on specific parts. By partitioning the dataset, it is possible to refine the analysis focussing on parts of the network.

a) Basic statistics and potential errors

Knowing the basic information of a dataset is the starting point for an analysis of its structure (*Analysis* → *Basic information [CTRL+B]*). This produces many and basic information such as: the proportion of men and women or individuals of unknown sex, the number of individuals in the dataset, the number of marriages and married individuals, fertile unions, the number components of the network density, the genealogical depth, etc..



Partition: ALL
Cluster: *

Target: Ebrei_23-10-2013.ged
Time spent: 142 ms

Label	Value
individuals	8477
men	5112
women	3364
unknown	1
marriages	3102
unions	3102
non-single men	4056
non-single women	2948
parent-child ties	11249
fertile unions	1996 (64.35%)
co-wife relations	224
co-husband relations	179
components	98 (max. 7606)
mean component share (agnatic)	0.05% (without singletons: 0.085%)
mean component share (uterine)	0.03% (without singletons: 0.085%)
max component share (agnatic)	1.52% (without singletons: 0.085%)
max component share (uterine)	1.20% (without singletons: 0.085%)
elementary cycles	5972
density (marriages)	0.43%
density (filiation)	1.57%
depth	15
mean spouse number of men	1.07
mean spouse number of women	1.06
mean fratry size agnatic	2.11
mean fratry size uterine	2.4

Fig. 10 : basic report information

Main basic report information :

1) The number of individuals (differentiated by gender: men/women/unknown)
2) Relations:
   - The number of marriage relations (differentiated by gender)
   - The number of parent-child relations

- Fertile marriages: The number of couples with children (in absolute numbers and as percentage of total marriages)
- Co-spouse relations: The number of relations between co-wives and between co-husbands.

3) Components:
   - The number of components (maximal connected subnetworks), the size of the largest component, to know the dataset cohesion or disintegration.
   - The mean share (size divided by total network size) of agnatic/uterine components and the share of the largest agnatic/uterine component, and the percentage of marriages involving a member of the largest agnatic/uterine component.
4) Elementary cycles: The cyclomatic number (number of independent cycles) of the network.
5) Density: Basic concept in network analysis, the density of the network (number of marriage and filial relations divided by the total number of possible relations between two different individuals).
6) Maximal and mean depth: The mean depth of the network (computed as an average of the mean generational depth of each individual's pedigree, according to the formula of Cazes (Cazes & Cazes, 1996).
7) Mean number of spouses: The mean number of spouses (differentiated by gender).
8) Mean fratry size: The mean number of (cognatic, agnatic and uterine) siblings.

Common errors

To quickly find errors (including coding) that could be contained in the dataset, Puck offers to identify the most common of them (*Reports* → *Controls* to choose the type of error and *Reports* → *Controls* → *Special Features* for a full report of all errors). It will be easy to correct the dataset after having identified possible errors maybe due to errors during data entry. Some of these items (such as persons without name or gender, or marriages between same-sex spouses or between parents and children) may actually be correct and wanted (depending fields and sources), but they often are due to simple errors during data entry. Puck indicates them in order to facilitate error control for the researcher, but never automatically "corrects" possible errors. Even if they are true representations of the kinship network, some of these irregularities may hinder certain functions of Puck from working correctly or necessitate a reconsideration of analytical methods.

Some errors may cause failure of some functionalities (for instance, cyclic descent causes infinite loops and crash of the program) or will lead to erroneous results in others (for instance, the presence of male mothers or female fathers causes calculation errors in the matrimonial census).

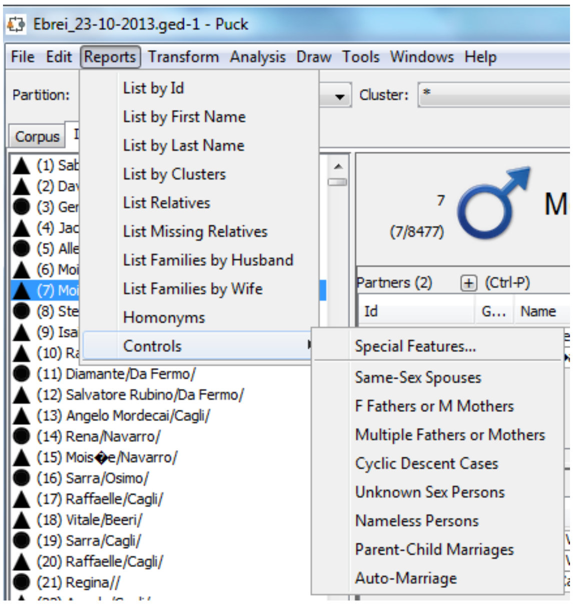


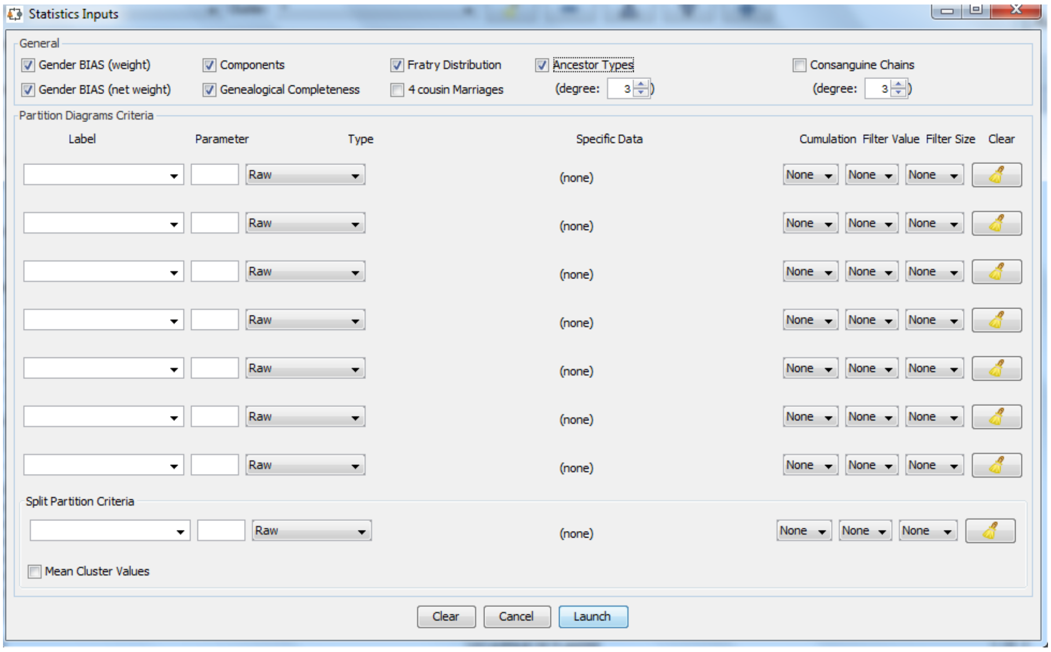
Fig. 11 : errors control

Main errors contained in the dataset:

1) Same-Sex spouses
2) Female fathers or male mothers
3) Multiple fathers or mothers
4) Nameless persons
5) Parent-child marriages
6) Auto-marriage
7) Unknown sex persons
8) Cyclic descent cases

b) Structure analyzis

After identifying errors (data entry / collection) in the dataset, the second step determines the limits and biases of the dataset, exploring the network morphology as a precondition for any analysis of matrimonial census. Puck offers a wide range of analysis, accessible from *Analysis* → *Statistics* (CTRL + G). The statistics' Inputs window allows five main operations : *Gender bias* (weight and net weight), *Components*, *Genealogical completeness*, *Ancestor chains* ("Ancestor types", choosing degree). A report window opens and displays the results in graphs which can be opened individually or saved by clicking on "save", choosing the destination folder. This window also provides information on the distribution of properties.



General

☑ Gender BIAS (weight) ☑ Components ☑ Fratry Distribution ☑ Ancestor Types ☐ Consanguine Chains

☑ Gender BIAS (net weight) ☑ Genealogical Completeness ☐ 4 cousin Marriages (degree: 3) (degree: 3)

Partition Diagrams Criteria

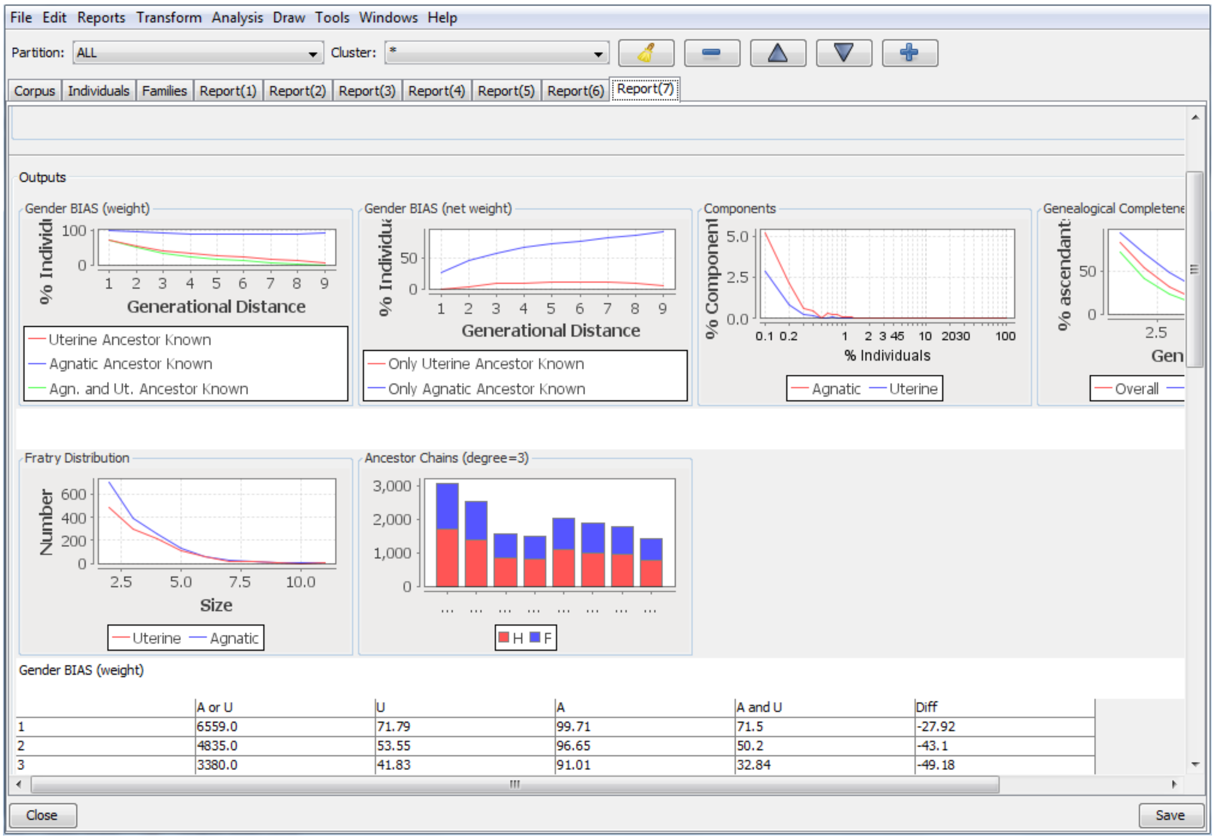
Label	Parameter	Type	Specific Data	Cumulation	Filter Value	Filter Size	Clear
		Raw	(none)	None	None	None	
		Raw	(none)	None	None	None	
		Raw	(none)	None	None	None	
		Raw	(none)	None	None	None	
		Raw	(none)	None	None	None	
		Raw	(none)	None	None	None	
		Raw	(none)	None	None	None	

Split Partition Criteria

		Raw	(none)	None	None	None	

☐ Mean Cluster Values

Fig. 12 : Statistics Inputs



Gender BIAS (weight)

	A or U	U	A	A and U	Diff
1	6559.0	71.79	99.71	71.5	-27.92
2	4835.0	53.55	96.65	50.2	-43.1
3	3380.0	41.83	91.01	32.84	-49.18

Fig. 13 : report window (statistics)

Gender bias:

There are two types of analysis for gender bias (Barry & Gasperoni, 2008; Hamberger & Daillant, 2008, p. 45):

- **Agnatic/uterine weight**: the number of individuals for whom the agnatic/uterine/the agnatic and the uterine linear ascendant of a given degree is known, as a percentage of individuals for whom the agnatic or the uterine ascendant of that degree is known. This is the first measure of the agnatic or uterine bias of the corpus.

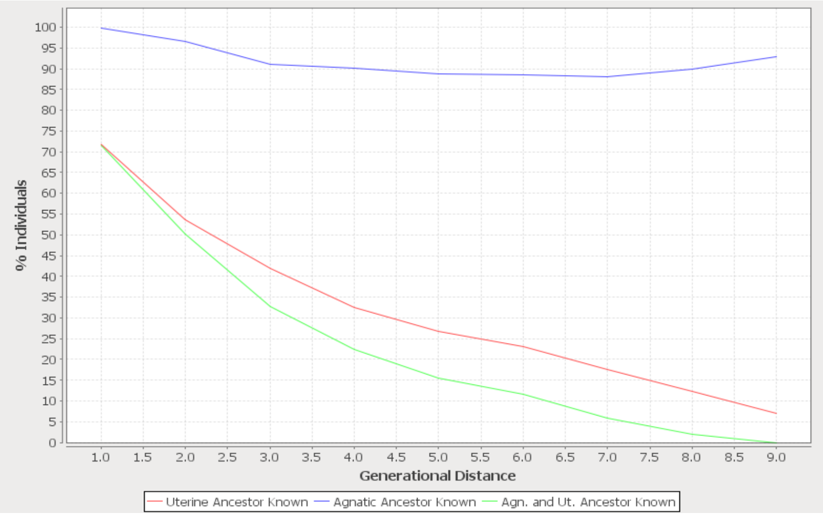


Fig. 14 : gender bias (weight)

- **Agnatic/uterine net weight**: the number of individuals for whom only the agnatic/uterine ascendant of a given degree is known, as a percentage of individuals for whom the agnatic/uterine ascendant of that degree is known. Another measure of the agnatic or uterine bias of the corpus are : this tool measures the interdependence and interconnection of genealogical knowledge : the more curves are close to one another the higher is the interdependence, the more they are apart the more they become autonomous, that is to say that we know the agnatic or uterine unisexual lines. If curves are low, it means that there is interconnection.

Note: it is important to analyze the gender bias on partitions (eg, depending on the generation or the age of birth of individuals). Partitioning is particularly useful for focussing on a part of networks .

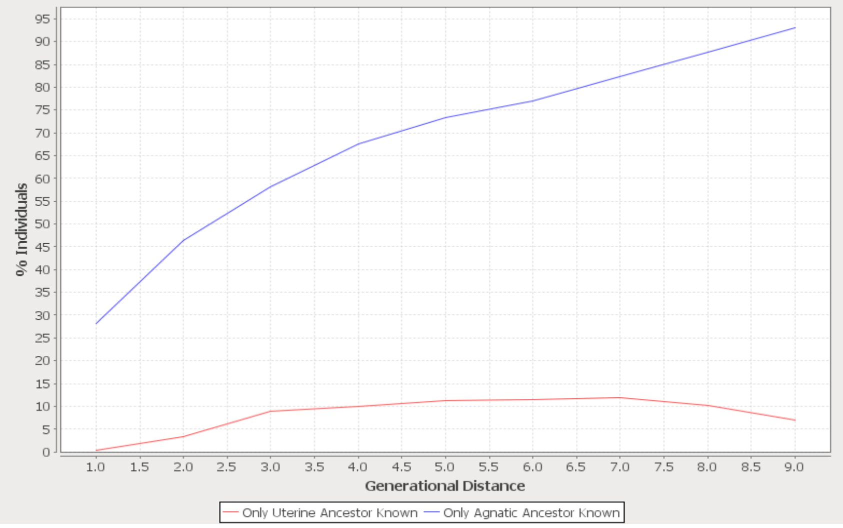


Fig. 15 : gender bias (net weigth)

Components:

It is the distribution of agnatic/uterine components (connected subnetworks made up entirely by paternal/maternal ties) according to their size: the abscissa of the diagram indicates the relative size of components (as a percentage of total network size, where size = number of individuals), the ordinate indicates the relative frequency of components of given size (as a percentage of the total number of components).

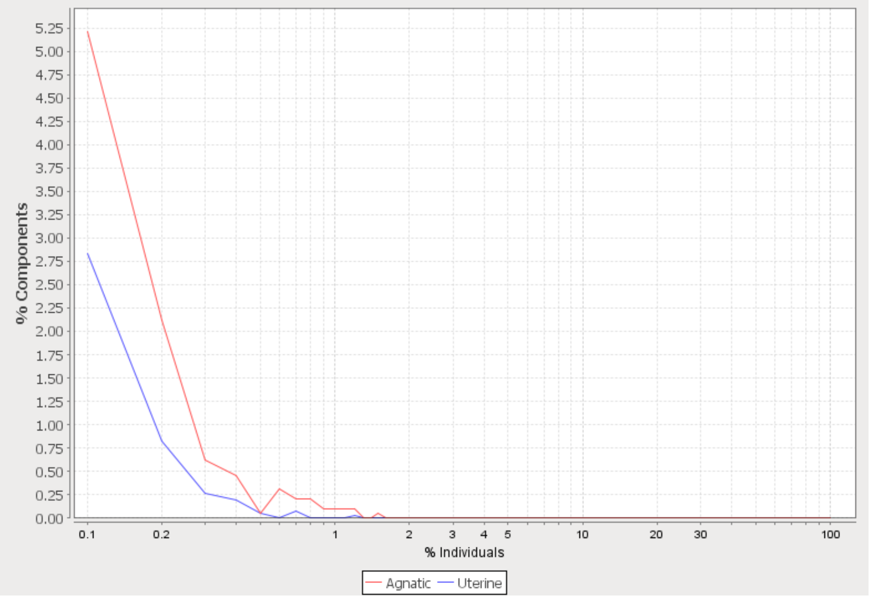


Fig. 16 : components

Genealogical completeness:

It is the percentage of known ascendants (agnatic, uterine and overall) by generation.

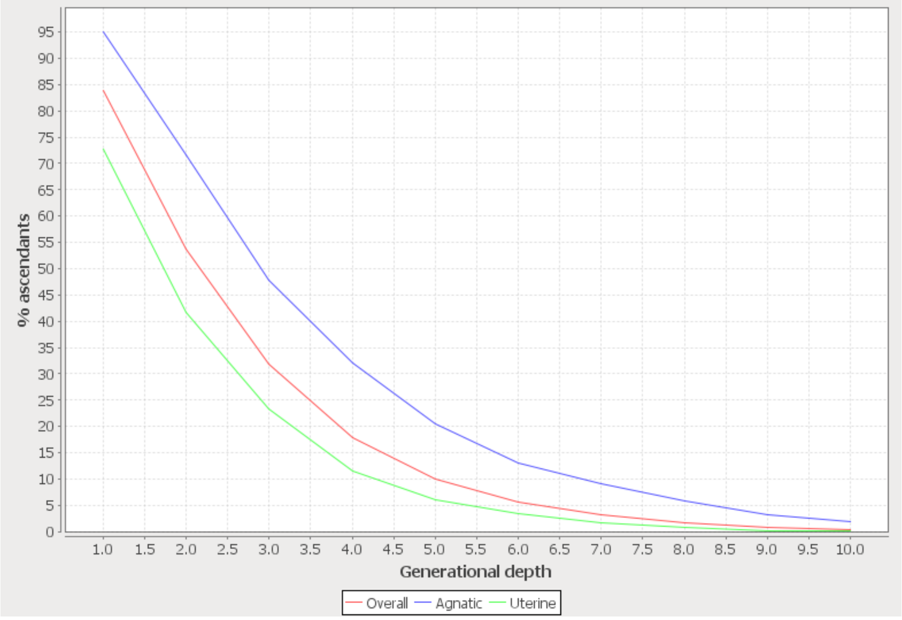


Fig. 17 : genealogical completeness

Fratry distribution:

It is the distribution of agnatic and uterine fratries (sibling groups) according to their size.

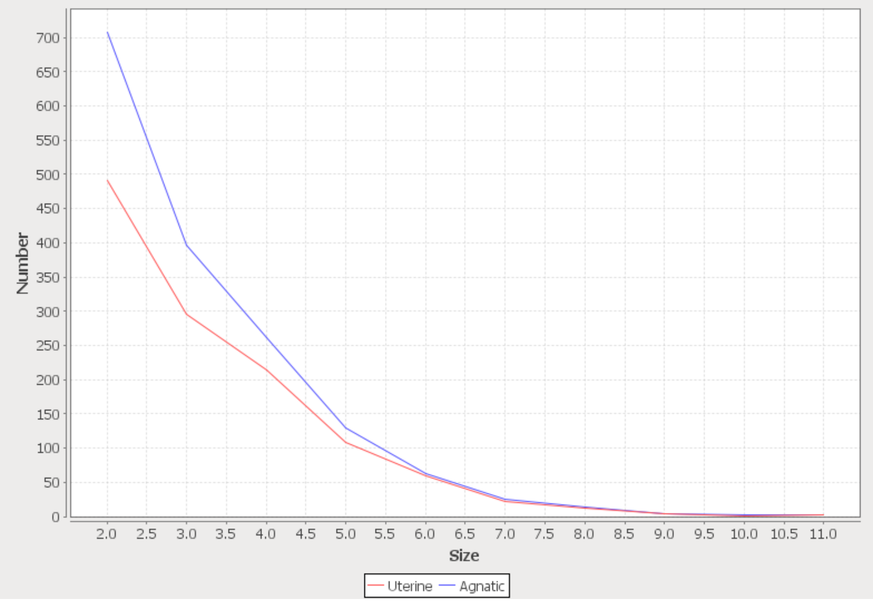


Fig. 18 : fratry distribution

Ancestor chains (choosing degree):

To know the nature of ancestor chains, depending on the degree (your choice) and gender, in positional notation. It is very important to know the distribution of consanguine and this is an additional measure of bias (Barry & Gasperoni, 2008, p. 71‐77).

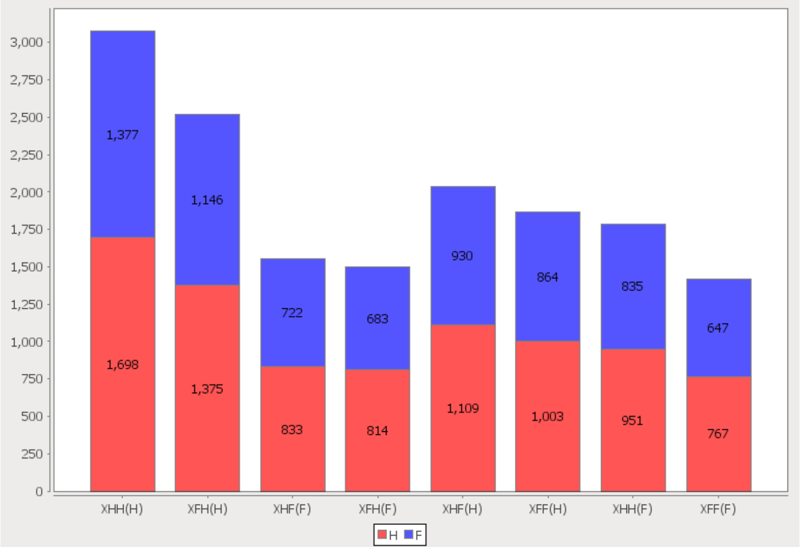


Fig. 19 : Ancestor chains

c) Distribution of properties

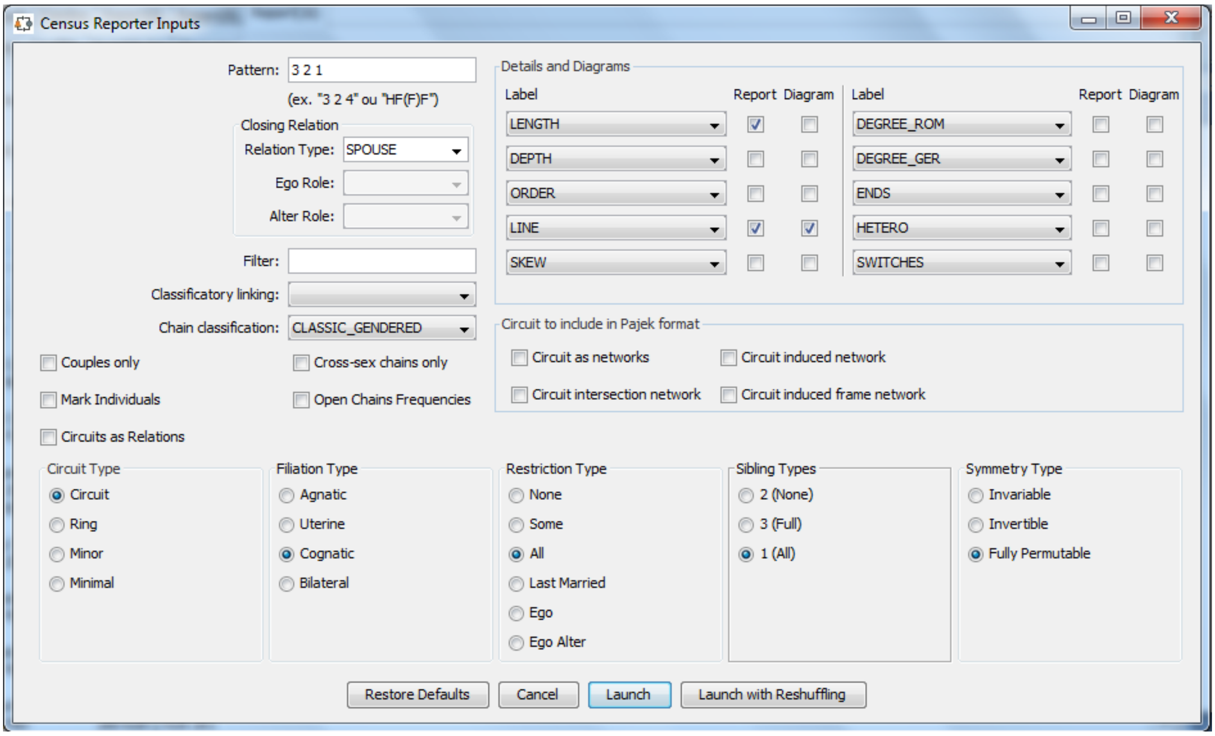
In the statistics window (CTRL + G), it is also possible to know precisely the distribution of endogenous and exogenous properties, combining queries and property codes and exporting the results as partitions, then using it with another software such as Pajek. This allows to know precisely the profile and composition of the dataset, not only for analysis according to certain criteria, but also to improve and complete it thereafter. Puck generates information and statistics on all genealogical (number and distribution of known ascendants/descendants etc.) or exogenous data (occupation, place and date of birth, etc.). The results appear in the form of tables, diagrams or partitions.

3. Matrimonial and relational census

One of the main interests of Puck is to identify all the circuits of a kinship network, but also all effective relations, to classify them and to analyze the topology of the set of matrimonial circuits (network of matrimonial components, circuit intersection network). The circuit intersection network allows to quickly identify the different combinations of marriages and to zoom on parts of the kinship network to check if an interleaving circuit is due to a special logic or shows simply the density of the network. The raw census must always be compared with the network quality: the matrimonial census is inseparable from consideration and a detailed analysis of his structure

a) *Circuit census*

The *circuit census* window has a number of parameters depending on the type of analysis desired (*Analysis* → *Circuit census* or CTRL+H). The "pattern" filed defines the type of census desired : consanguineous marriages, two-groups and three-groups relinking with two methods of circuits counting.



Pattern: 3 2 1
(ex. "3 2 4" ou "HF(F)F")

Closing Relation
Relation Type: SPOUSE
Ego Role:
Alter Role:

Filter:
Classificatory linking:
Chain classification: CLASSIC_GENDERED

□ Couples only
□ Cross-sex chains only
□ Mark Individuals
□ Open Chains Frequencies
□ Circuits as Relations

Details and Diagrams

Label	Report	Diagram	Label	Report	Diagram
LENGTH	☑	□	DEGREE_ROM	□	□
DEPTH	□	□	DEGREE_GER	□	□
ORDER	□	□	ENDS	□	□
LINE	☑	☑	HETERO	□	□
SKEW	□	□	SWITCHES	□	□

Circuit to include in Pajek format
□ Circuit as networks
□ Circuit induced network
□ Circuit intersection network
□ Circuit induced frame network

Circuit Type
● Circuit
○ Ring
○ Minor
○ Minimal

Filiation Type
○ Agnatic
○ Uterine
● Cognatic
○ Bilateral

Restriction Type
○ None
○ Some
● All
○ Last Married
○ Ego
○ Ego Alter

Sibling Types
○ 2 (None)
○ 3 (Full)
● 1 (All)

Symmetry Type
○ Invariable
○ Invertible
● Fully Permutable

Fig. 20 : circuit census window

- The first way is to use this field to specify the maximal dimension of the matrimonial rings of different widths to be searched for (that is, the maximal canonic degree of the consanguinity relations involved in the ring): the first number indicates the maximal canonic degree of rings with width 1 (i.e. incorporating 1 marriage arc, consanguineous unions), the second number the maximal canonic degree of rings with width 2 (i.e. incorporating 2 marriage arcs, two-group relinkings), the third number the maximal canonic degree of rings with width 3 (i.e. incorporating 3 marriage arcs, three-group relinkings). For instance, the code 3 2 1 sets the horizon of matrimonial ring search to blood marriages between 2nd cousins (degree 3), marriage redoublings between pairs of 1st cousins (degree 2) and marriage retriplings between pairs of siblings (degree 1). All rings/relations of lower dimensions (for instance, blood marriages between first cousins or marriage redoublings between pairs of siblings) are included in the search.

- The second way is to search by structural schema expressed in positional notation (i.e XX(X)XX will find all rings corresponding to marriages between cousins). Note that, as a rule, a structure formula is read by Puck in a socio-centered manner. That is, the formula X(H)HX and XH(H)X designate one and the same type of relation. This is important if the census is run on a subcorpus, where ego may be in the corpus but alter may not. Without further indication, a structure formula limits search to relations/rings that exactly fit this formula (contrary to the number code method which only sets an upper limit). We can, however, include all relations/rings which lie within the limits of the formula by letting it precede the sign "<". For instance, the formula "<XXX(X)XX" limits matrimonial ring search to all blood marriage rings within the limits of the 5rd civil degree (including marriages between first cousins, between uncles and nieces etc.). Juxtaposition of several structure formula is interpreted as a combination by logical "or". For instance, the formula X.X(X)X XX(X)XX limits ring search to marriages between siblings in-law or 1st cousins.

The census can be refined according to specific criteria, i.e *filiation, symmetry* or *sibling, circuit and restriction types*:

- The type of permitted consanguinity relations can be set in for possible ways:
  1. Cognatic: all consanguinity relations are permitted
  2. Agnatic: only agnatic relations are permitted (unilinear census)
  3. Uterine: only uterine relations are permitted (unilinear census)
  4. Bilateral: only bilateral relations are permitted

- The symmetry type of the relation between ego and alter (option symmetry in the search options window) decides on the permutability or non-permutability of ego and alter.

Example: if kinship chains between co-residents are searched, the option "symmetry" has to be activated, for co-residence is a symmetric relation. Accordingly, the chains "father-son" and "son-father" will be counted as one single category. By contrast, if kinship chains between persons and their heirs are searched, the option "symmetry" has to be deactivated, for inheritance is an asymmetric relation. Accordingly, the chains "father-son" and "son-father" will be counted as different categories.

Note: the symmetry type choice is only relevant for a non-matrimonial circuit census. If matrimonial circuits or open kinship chains are searched, ego and alter are always considered as permutable (in the first case, male or female ego will be chosen according to the chosen option, in the second case, there is no criterion for the selection of ego or alter)

- The type of sibling differentiation can be set in three possible ways:
  1. One single sibling type, "All" (all siblings assimilated). Full and half siblings are not distinguished. This method is recommended where half sibling relations rarely occur or do not matter for marriage rules.
  2. Two sibling types, "None" (no siblings assimilated. Only paternal and maternal siblings are distinguished, full siblings are counted twice (once as paternal and once as maternal siblings). This method is recommended where half-sibling relations are frequent (e.g. because of high rates of polygamy) the agnatic or uterine relationship is more important than the sibling relationship as such.
  3. Three sibling types, "Full" (full siblings are assimilated). True paternal and maternal half-siblings are distinguished from full siblings. This method provides the maximum of information.

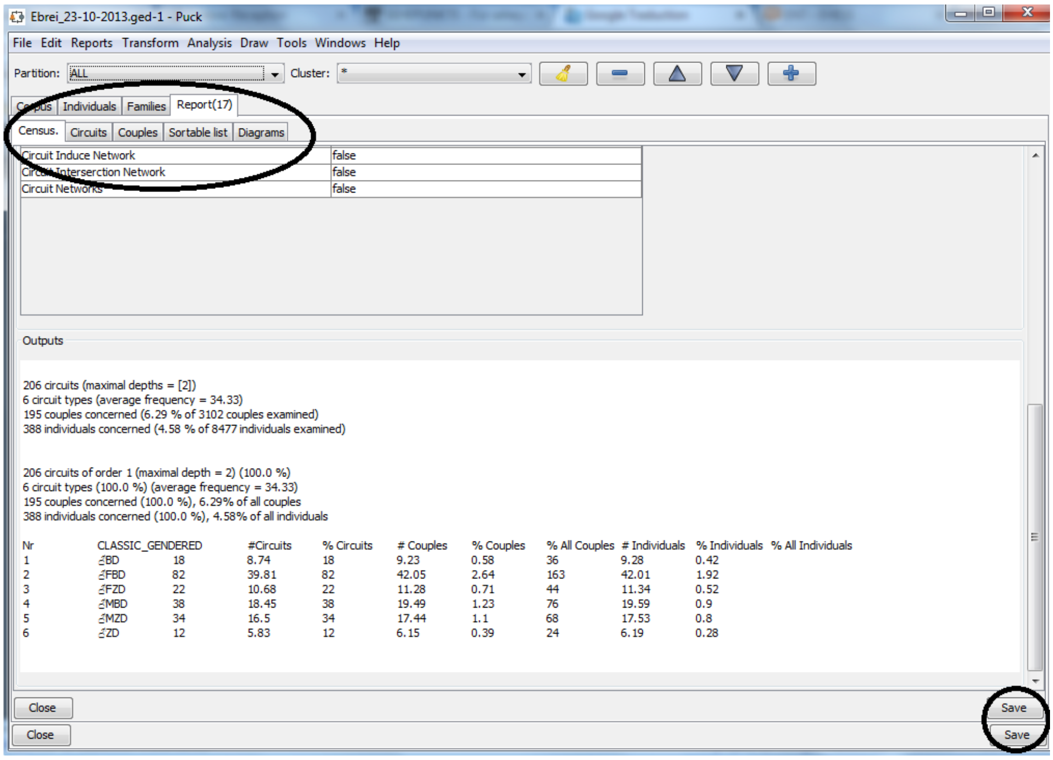
- The circuit types are:
    1. Circuit : counts matrimonial circuits.
    2. Ring: does not consider circuits/relations that completely include shorter ones (that is, matrimonial rings and not all matrimonial circuits).
    3. Minor: does not consider circuits/relations that intersect with shorter and narrower ones (those that only count minor matrimonial rings and not all matrimonial circuits).
    4. Minimal: Minimal rings only: does not consider circuits/relations that intersect with shorter ones (Those that only count minimal matrimonial rings and not all matrimonial circuits)

- Restriction types (censuses restricted to one cluster) are:
    1. All: all married individuals (matrimonial ring pivots) must belong to the chosen cluster
    2. Ego: ego (according to a chosen kinship schema) must belong to the chosen cluster (presupposes ring search by formula).
    3. Last married: the last married individual must belong to the chosen cluster (presupposes marriage dates, for the moment they were still treated as individual properties)

With *closing relation*, it is possible to choose different types of censuses (matrimonial, relational, absence of closing relations) from endogenous or exogenous properties (occupation, residence etc.), extending the search criteria and producing diagrams.

b) Results and tables

Puck produces a census report that indicates the precise kind of circuits, their classification as individuals and couples. The report can be saved in txt or xls.



Circuit Induce Network	false
Circuit Interserction Network	false
Circuit Networks	false

Outputs

```
206 circuits (maximal depths = [2])
6 circuit types (average frequency = 34.33)
195 couples concerned (6.29 % of 3102 couples examined)
388 individuals concerned (4.58 % of 8477 individuals examined)


206 circuits of order 1 (maximal depth = 2) (100.0 %)
6 circuit types (100.0 %) (average frequency = 34.33)
195 couples concerned (100.0 %), 6.29% of all couples
388 individuals concerned (100.0 %), 4.58% of all individuals

Nr      CLASSIC_GENDERED      #Circuits  % Circuits  # Couples  % Couples  % All Couples  # Individuals  % Individuals  % All Individuals
1       ⩽BD       18          8.74       18          9.23       0.58       36             9.28           0.42
2       ⩽FBD      82          39.81      82          42.05      2.64       163            42.01          1.92
3       ⩽FZD      22          10.68      22          11.28      0.71       44             11.34          0.52
4       ⩽MBD      38          18.45      38          19.49      1.23       76             19.59          0.9
5       ⩽MZD      34          16.5       34          17.44      1.1        68             17.53          0.8
6       ⩽ZD       12          5.83       12          6.15       0.39       24             6.19           0.28
```

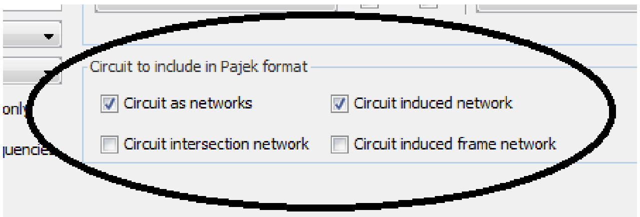
Fig. 21 : census report

Report window:

- Census: indicates the total number of circuits and the maximum height (the canonic degree), the number of different circuits, the number of individuals and couples involved, in absolute numbers and as percentages, both in total and on the circuits concerned.
- Circuits: list for each type of relation/ring (indicated in standard notation), all relations found in the corpus, with nominal indication of their pivots (and their relations: "=" for marriage, "-" for consanguinity) and the complete chain in positional notation (individuals being indicated by their identity numbers).
- Couples: list for each couple or pair of relatives concerned by the census, all the rings/relations that link them, both in standard notation of the type and in positional notation of the complete chain (individuals being indicated by their identity numbers).
- Sortable list: List all relations found in the census, with the index, standard and positional notation of the relation/ring type, nominal indication of their pivots and their relations ("=" for marriage, "-" for consanguinity) and the complete chain in positional notation (individuals being indicated by their identity numbers).

c) *Matrimonial networks*

Puck allows to generate matrimonial networks to be analyzed in Pajek. There are several types of operations:



Circuit to include in Pajek format

☑ Circuit as networks ☑ Circuit induced network

□ Circuit intersection network □ Circuit induced frame network

Fig. 22 : generating networks

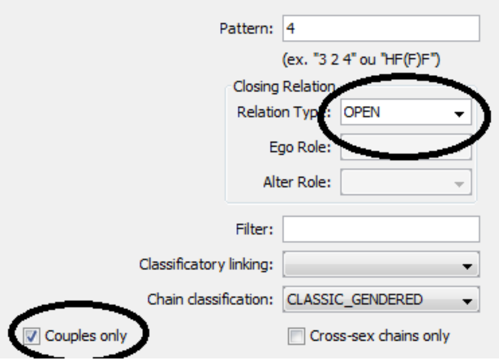
- Matrimonial network: produces the matrimonial network corresponding to the matrimonial census (that is, a network made up only of the links that form part of some matrimonial circuit), as well as a partition to distinguish vertices that occupy spouse positions in the matrimonial network.
- Frame of matrimonial network: produces the frame of the matrimonial network corresponding to the matrimonial census, embedded in the total kinship network.
- Circuit intersection network: produces the ring intersection network corresponding to the matrimonial census, as well as the circuit intersection matrix in text format (Nodes are circuit types and lines = marriages contained in circuits of both types). Using the concept of "circuit intersection network" (Hamberger & Daillant, 2008, p. 44) which can identify and relate marriages that fall within a particular type of circuit and another type in the same time, can be useful if one considers that some marriages must be understood not as the immediate expression of a relationship between spouses, but as unions that participate in much a much wider network of alliances.

d) Relational to complex matrimonial census :

If the matrimonial circuit census counts the matrimonial circuits in a kinship network, a non-matrimonial circuit census counts relational circuits in a kinship network, that is, kinship and marriage chains between people who are linked by previously defined relations (for instance, co-residence, friendship, etc.).

Open chains

To produce census of open chains for married people, use "closing relations" "open" and "couples only".



Pattern: 4

(ex. "3 2 4" ou "HF(F)F")

Closing Relation

Relation Type: OPEN

Ego Role:

Alter Role:

Filter:

Classificatory linking:

Chain classification: CLASSIC_GENDERED

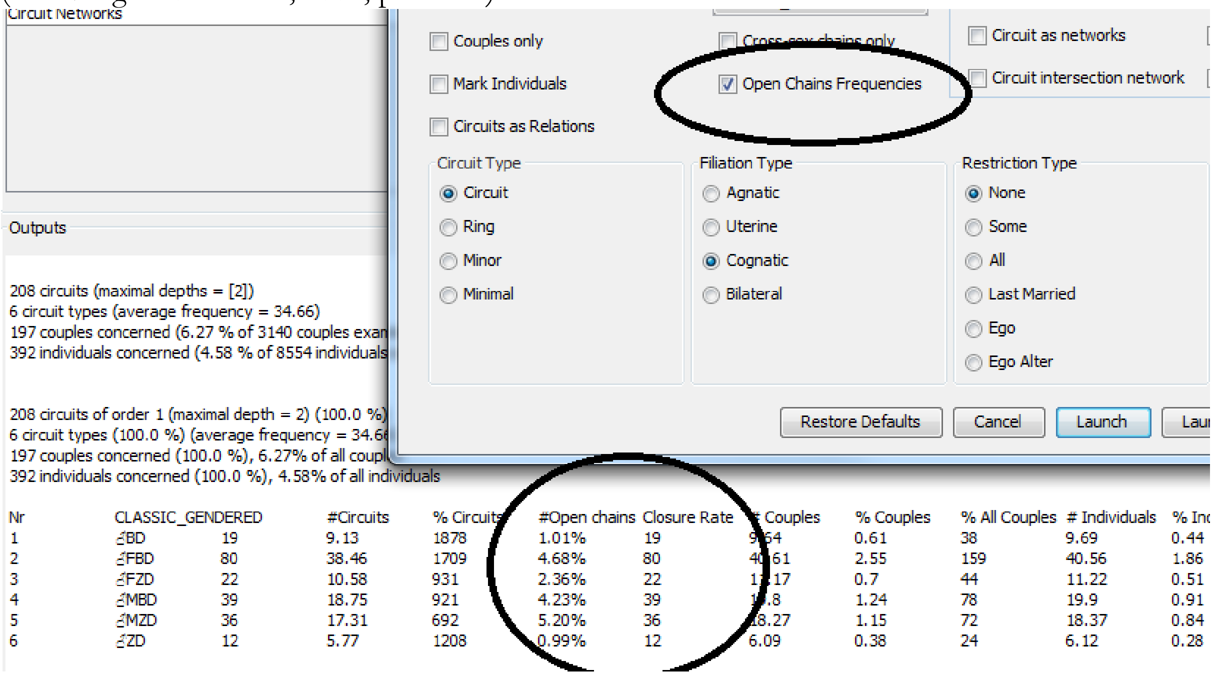
☑ Couples only

□ Cross-sex chains only

Fig. 23 : open chains

"Open chains frequencies" and "cloture rate"

The relational census can check the hypothesis and sometimes results in perspective. For example, "open chains frequencies" function counts the kinship chains in a kinship network, that is, all kinship and marriage chains, no matter how ego and alter are related otherwise. This allows in particular to know the "closure rate": indeed, the raw results of matrimonial census products must be reconsidered and according to the quality of the corpus. Thus, it is legitimate to ask whether these "preferential" marriages supposed to certain categories of parents are not the consequence of the imbalance of the genealogical information that is included in the network. It is therefore necessary to introduce other variables to refine the analysis and test hypotheses or trends expressed by the marital census. The "cloture rate" appears as an interesting indicator (Hamberger & Daillant, 2008, p. 27-28).



□ Couples only
□ Cross-sex chains only
□ Circuit as networks
□ Mark Individuals
☑ Open Chains Frequencies
□ Circuit intersection network
□ Circuits as Relations

Circuit Type: ● Circuit ○ Ring ○ Minor ○ Minimal

Filiation Type: ○ Agnatic ○ Uterine ● Cognatic ○ Bilateral

Restriction Type: ● None ○ Some ○ All ○ Last Married ○ Ego ○ Ego Alter

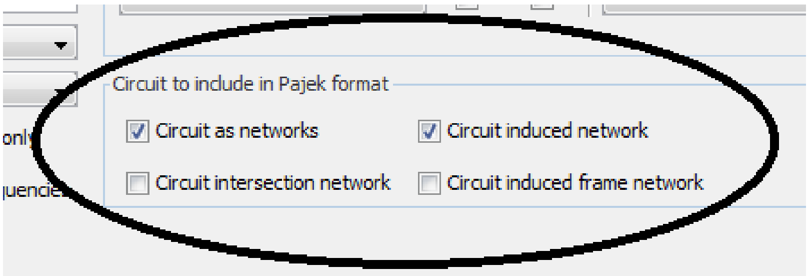
```
208 circuits (maximal depths = [2])
6 circuit types (average frequency = 34.66)
197 couples concerned (6.27 % of 3140 couples exam
392 individuals concerned (4.58 % of 8554 individuals


208 circuits of order 1 (maximal depth = 2) (100.0 %)
6 circuit types (100.0 %) (average frequency = 34.6
197 couples concerned (100.0 %), 6.27% of all coupl
392 individuals concerned (100.0 %), 4.58% of all individuals

Nr     CLASSIC_GENDERED    #Circuits   % Circuits  #Open chains  Closure Rate  # Couples  % Couples  % All Couples  # Individuals  % Inc
1      ♂BD      19         9.13        1878        1.01%         19            9.64       0.61       38             9.69           0.44
2      ♂FBD     80         38.46       1709        4.68%         80            40.61      2.55       159            40.56          1.86
3      ♂FZD     22         10.58       931         2.36%         22            11.17      0.7        44             11.22          0.51
4      ♂MBD     39         18.75       921         4.23%         39            19.8       1.24       78             19.9           0.91
5      ♂MZD     36         17.31       692         5.20%         36            18.27      1.15       72             18.37          0.84
6      ♂ZD      12         5.77        1208        0.99%         12            6.09       0.38       24             6.12           0.28
```

Fig. 24 : open chains frequencies and cloture rate

Matrimonial networks

Puck produces matrimonial networks exported in a *Pajek* format:

- *Circuit induced network* produces the matrimonial network corresponding to the matrimonial census (that is, a network made up only of the links that form part of some matrimonial circuit), as well as a partition to distinguish vertices that occupy spouse positions in the matrimonial network.
- *Circuit induced frame network* (consanguine chains reduced to lines) produces the frame of the matrimonial network corresponding to the matrimonial census, embedded in the total kinship network in which marriages are coded according to the ring(s) they are part of. This may provide a synthetic representation in which certain connective features of the matrimonial network can be identified.
- *Circuits as network* exports all matrimonial rings found by the census as separate networks (with nominal indication of individuals as vertex labels), as well as a partition to distinguish vertices that occupy spouse positions in the circuit.
- *Circuit intersection network* produces the ring intersection network (Hamberger & Daillant, 2008, p. 22-24) corresponding to the matrimonial census, as well as the circuit intersection matrix.



Circuit to include in Pajek format

☑ Circuit as networks

☑ Circuit induced network

□ Circuit intersection network

□ Circuit induced frame network

Fig. 25 : Networks

Mixed matrimonial & connubial circuits

Puck allows to regroup individuals according to a certain property (chosen from a drop-down menu by double clicking on the checkbox label) and effectuate a census of:

a) Mixted matrimonial circuits containing the relation "belonging to the same cluster". Puck distinguishes for the moment 9 types of mixed matrimonial rings, according as the H, HF, or HM belong to the same cluster as W, WF or WM.

b) Connubial circuits consisting of marriage alliance relations (each of which may represent several marriages) between groups of individuals (that is, rings in an alliance network, where arrows point from the wife's group to the husband's group): Puck distinguishes endogamous rings (1 group), redoubling circuits or exchange circuits (2 groups, arrows pointing in the same or in inverse directions) and cycle circuits (3 groups, arrows consistently directed).

For each ring, the census lists the number of connubial circuits, the number of distinct cycles that may be formed from the marriages that constitute the connubial circuit, the weight of the circuit (the geometric mean of marriages joining two groups in the ring) and the probability of the circuit, given the relative numbers of potential spouses in the constitutive groups.

4. Partitions

a) Segment a dataset

It is possible to make any kind of partitions based on needs analysis, choosing your criteria (individual of family properties). This help to refine and supplement the qualitative analysis (dataset diagnosis) and/or matrimonial census.

Any individual property (see individual property codes for details) can be used to split the corpus into subcorpuses. A subcorpus (whose title indicates the parent corpus as well as the property label and value) has the appearance of an autonomous corpus (with its own corpus window and all dependant windows), and every operation on a corpus can also be effectuated on a subcorpus. The important difference is that a subcorpus remains linked to the parent corpus, and all relation and ring search processes, while limiting results to individuals in the subcorpus, always run through the total corpus.

Warning: *a subcorpus which is saved and re-imported as a normal corpus loses this important subcorpus property. All links to individuals outside the subcorpus are cut, and external individuals can no longer act as intermediaries for chains between members of the subcorpus.*

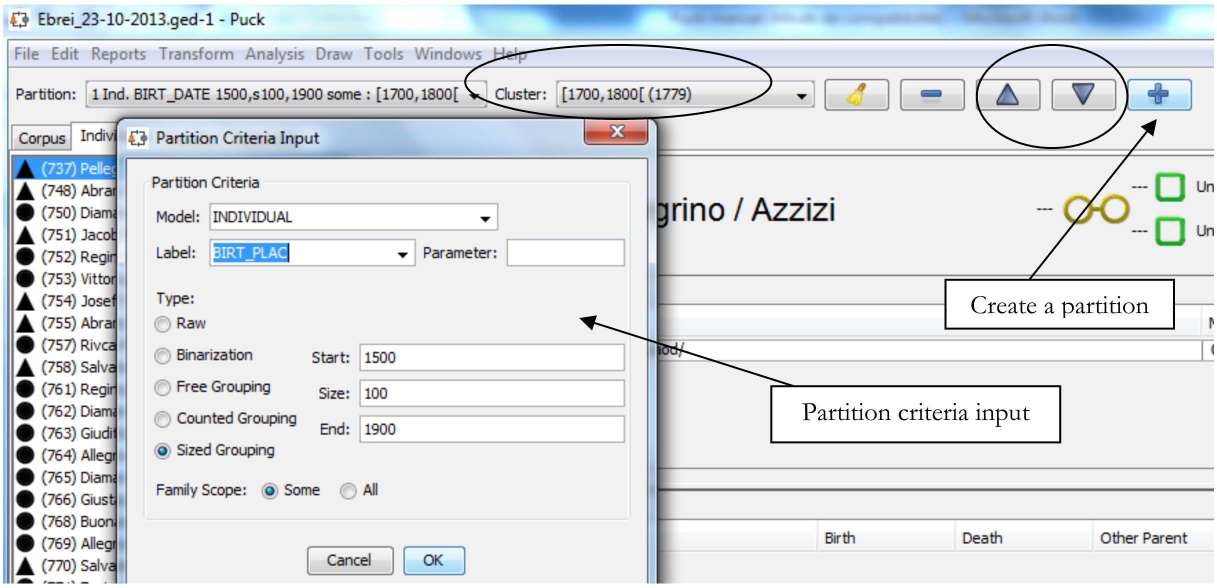
Examples (Fig. 26 & Fig. 27):

a) Genealogical corpora can be partitioned on a first level according to Gender (Model : INDIVIDUAL, Label :SEX) and then, one of the clusters, male or female, can be partitioned at a second level according to the Family Name (Model : INDIVIDUAL, Label LASTN) and one of these clusters is to be partitioned according to the surname (Model : INDIVIDUAL, Label FIRSTN). The arrow buttons are used to move up and down to the different levels of partition. For each item from the « Partition » list, the correspondent clusters appear in the « Cluster » list.

Different types can be chosen for the partition:

- **Raw** : produces a basic partition containing as much clusters as necessary e.g. Gender (Model : INDIVIDUAL, Label :SEX) produces 3 clusters : a Null, a Female and a Male. A Family-Name partition (Model : INDIVIDUAL, Label LASTN) produces much more clusters according to the number of names in the corpus and the presence of a null cluster or not.
- **Binarization** : always produces, as the name indicates, 2 clusters, one corresponding to the Pattern and another one regrouping everything but the Pattern.
- **Free grouping** : allows you to fix irregular (or regular) date intervals
- **Counted grouping** : allows you to fix the number of clusters in which you want to divide a given period (from start … to end).
- **Sized grouping** : allows you to fix the duration of clusters in which you want to divide a given period (from start … to end).

b) Partitions the corpus according to the date of birth of individuals: *Start* = 1500 ; *Size* = 100 and *End* = 1900 divides individuals born in 1500 to 1900 onwards into sub-sets corresponding to 100 year intervals, adding another sub-set containing individuals born before 1500 and a still further sub-set containing individuals whose birth date is unknown. (Sized grouping).



Ebrei_23-10-2013.ged-1 - Puck

Partition: 1 Ind. BIRT_DATE 1500,s100,1900 some : [1700,1800[

Cluster: [1700,1800[(1779)

Partition Criteria Input

Partition Criteria

Model: INDIVIDUAL

Label: BIRT_PLAC

Parameter:

Type:

○ Raw

○ Binarization

○ Free Grouping

○ Counted Grouping

● Sized Grouping

Start: 1500

Size: 100

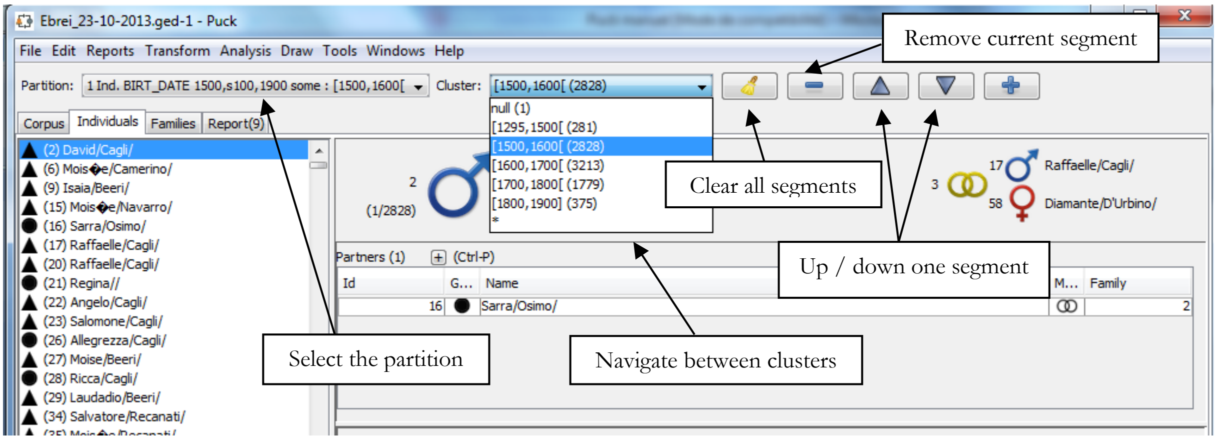
End: 1900

Family Scope: ● Some ○ All

(737) Pelleg
(748) Abrar
(750) Diama
(751) Jacob
(752) Regin
(753) Vittor
(754) Josef
(755) Abrar
(757) Rivca
(758) Salva
(761) Regin
(762) Diama
(763) Giudi
(764) Alleg
(765) Diama
(766) Giust
(768) Buon
(769) Alleg
(770) Salva

Fig. 26 : create a partition

Now, the dataset is reduced to the cluster to which the current individual belongs, and the other clusters show up in the drop-down menu. The dataset can be partitioned more than once by applying the current cluster (*Add segment*): successive application of different properties as partition criteria permits to refine partitioning.



Ebrei_23-10-2013.ged-1 - Puck

Partition: 1 Ind. BIRT_DATE 1500,s100,1900 some : [1500,1600[

Cluster: [1500,1600[(2828)

null (1)
[1295,1500[(281)
[1500,1600[(2828)
[1600,1700[(3213)
[1700,1800[(1779)
[1800,1900] (375)
*

(2) David/Cagli/
(6) Mois�e/Camerino/
(9) Isaia/Beeri/
(15) Mois�e/Navarro/
(16) Sarra/Osimo/
(17) Raffaelle/Cagli/
(20) Raffaelle/Cagli/
(21) Regina//
(22) Angelo/Cagli/
(23) Salomone/Cagli/
(26) Allegrezza/Cagli/
(27) Moise/Beeri/
(28) Ricca/Cagli/
(29) Laudadio/Beeri/
(34) Salvatore/Recanati/

Partners (1) (Ctrl-P)

Id	G...	Name
16	●	Sarra/Osimo/

M...	Family
ꝏ	2

Fig. 27 : browse and modify cluster

c) *Extract and expand a segment*

Puck allows to extract a particular segment a produce a truncated dataset (*Transform* → *Extract current segment*) and to expand it including the relatives (ascendants, descendants and affines) of segment members (*Transform* → *Expand current segment*).

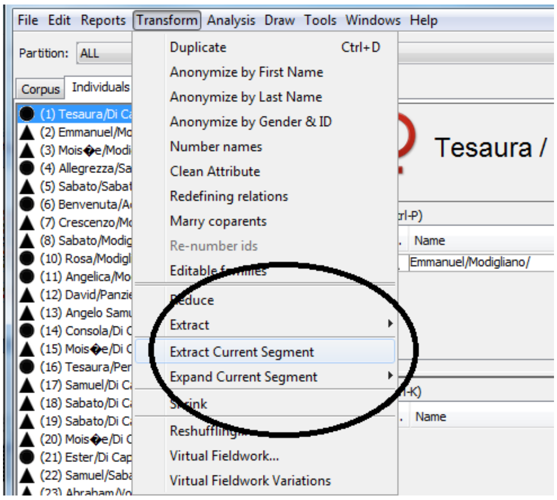


Fig. 28 : extract segment

d) Differential census

Conducting a segment-based census provides diverse comparative means, especially in concerns of relations that establish connections between individuals from identical segments. The significant advantage over a global census is the consideration of several segments. By applying a differential census (*Analysis* → *Differential census*) to all the dataset's segments configured by a certain segmentation statistic results can be achieved concerning for instance uterine-agnatic relations within a household or kinship relations within a profession. Differential census produce relational statistics for each cluster, global as well as mean percentages of relations and the distribution of relation percentages by cluster size.

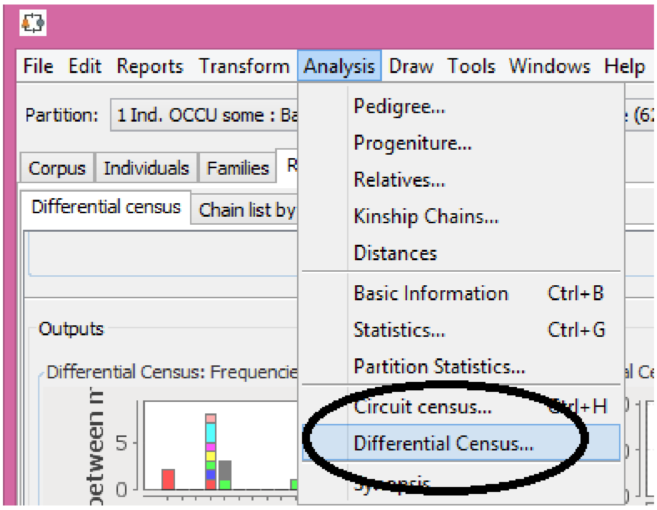


Fig. 29 : differential census

APPENDICES

How to establish a kinship dataset

A genealogical corpus is a set of individuals linked by relations of kinship and marriage with basic information and supplementary information for each individual that has been coded.

- Basic information for each individual:
    - A unique identity number (ID)
    - Name(s)
    - Gender: H (man), F (woman), X (gender unknown)
    - Father's ID number
    - Mother's ID number
    - Spouse(s) ID number

- Supplementary information for each individual:
    - Biographical information (birth, marriage, death dates and places, other properties)
    - Notes

Data are not only a result but also a means of data collection. They should be easily accessible in order to guide your research and to cross-check your informant's answers. When dealing with archives, this is often fairly simple: you can take a computer with you. But in many fieldwork situations this is not possible. However, noting kinship „by hand" can be extremely fast and efficient, if some basic principles are observed.

— Always use a compact medium, such as a notebook. Do not use filesheets or loose papers. You cannot use them during interviews, and there is a high risk of loosing some of them.

— Separate graphics and text. A good method is to use a notebook with the left page for drawing genealogies, the right page for listing the individuals and their properties, and numbers for identifying these individuals (if numbers get large, it is recommended to use, in addition, initial letters to prevent identification problems in case of numbering errors) — Attribute an identity number to each individual and never attribute that number to another individual. If you have „doubles", make a link to the original number but do not re-assign it. Holes in the series of numbers do not cause any damage, but ambiguities in identity numbers cause much damage, and are extremely difficult to detect.

Do not use identity numbers as codes. Identity numbers serve to identify individuals - and nothing else (except, perhaps, to recall the order in which you have entered them and to document the history of your corpus). If you want to convey information on individuals gender, clan affiliation, residence, etc., do not use identity numbers for that.

Never forget to make regularly copies and store them on different places. This holds for all data, but especially for kinship data, due to the network properties of kinship: one lost notebook may render twenty others useless.

Frequently asked questions

— Do I have to number individuals continuously?

No. Discontinuous numbering is no problem for Puck nor for most other genealogical programs. Pajek requires continuous numbering, but Puck can convert datasets into pajek file format including renumbering without loss of information on original numbers (by using the option "numbered" for exportation). However, you should avoid too large empty spaces between identity numbers, because some search methods may get more time intensive.

— Some individuals in my dataset are doubles, do not exist, or have become obsolete. Can I delete them ?

Yes, but do not reassign their identity numbers to other individuals! Just leave their positions empty. In the case of doubles, it can be useful to keep them in your dataset, so that you can easily find informations on the individuals in the different places in your notebooks. You can mark them as doubles by assigning them as a name the identity number of the original. If needed, you can always eliminate them by the eliminate doubles option.

— How do I code kinship relations between individuals when I ignore the exact genealogical chain ?

If you know the exact genealogical relation, you may introduce into your dataset virtual individuals - having « # » as a name - as intermediary links (for instance, if you know that A is B's paternal brother, you may introduce a virtual common father). Make sure, however, that the kinship term people give you really corresponds to the supposed genalogical relation (in many societies, kinship terms may designate large classes of relations, some of them may be without any genealogical foundation whatsoever!) If you are not 100% sure that your « brother » really is a brother in a genalogical sense, you should rather store the information in a note or as relational property of the concerned individuals.

— How do I code divorced spouses ?

Like all other spouses, living or dead, married or divorced. You can store the information on divorce among the individuals properties (see also File formats for kinship data).

File formats for kinship data

Kinship data can be stored in files of different formats: text, Pajek, Excel, TIP, Gedcom, Prolog. Each format has its own characteristics of encodings, and some limited treatment options exogenous data. With this in mind, Puck can use the best of datasets:

category	elements	txt/ods/xls v1	ged	txt/ods/xls v3	puck
corpus	attributes	no	yes	yes	yes
individual	id	yes	yes	yes	yes
	name	yes	yes	yes	yes
	gender	yes	yes	yes	yes
	attributes	yes	yes	yes	yes
family	id	no	yes	yes	yes
	unmarried status	yes	no	yes	yes
	married status	yes	yes	yes	yes
	divorced status	no	no	yes	yes
	father	yes	yes	yes	yes
	mother	yes	yes	yes	yes
	null parent	no	yes	yes	yes
	children	yes	yes	yes	yes
	attributes	no	yes	yes	yes
	unmarried sterile	no (lost)	yes (married)	yes	yes
	married sterile	yes	yes	yes	yes
	unmarried orphan	no (lost)	yes	yes	yes
	married orphan	no (lost)	yes (unmarried)	yes	yes
relation model		no	no	yes	yes
relation	id	no	no	yes	yes
	actors	no	no	yes	yes
	attributes	no	no	yes	yes

a) Text format (file extension .txt)

This is a tab delimited text file, composed of two blocks separated by an empty line.
The first block contains the basic information for each individual in separate columns:

- A unique identity number (ID)
- Name(s), where different name parts are separated by a slash (/)

- Gender: M or H (man), W or F(woman), X (gender unknown). Gender letters are not case sensitive
- Father's ID number
- Mother's ID number
- Spouse(s) ID number, where the ID numbers of different spouses appear in different columns

A headline may be convenient for data entry, but is not necessary for Puck.

Attention! If you use a headline, do not call the column of identity numbers " ID " (rather use " Id " or " Nr " or something similar), otherwise it will not be opend by Microsoft Excel (this has nothing to do with Puck, but is a general Microsoft Excel Bug).

In the case of multiple spouses, there are two possibilities (which may be combined):

- Either the ID numbers of an individual's spouses appear in one single line but different columns (from the sixth column on). This is the output produced by Puck and the most convenient solution if an individual's spouses are immediately known ;
- The ID numbers of an individual's spouses appear in the same column (the sixth) but in different lines, which means that the individual's ID number has to be entered several times. This solution may be more comfortable if information on an individual's spouses is dispersed.

In the second block, each line contains an individual's ID number and, in successive columns, items of supplementary information concerning the individual. These items are the values of individual properties, columns being labelled according to individual property codes, used to cluster individuals according to simple or more complex properties.

b) Pajek network format (file extension .paj)

Kinship data in pajek format can be used to transform, manipulate and analyze them with the computer program pajek. Puck exports data into a pajek project file (that is, a package of network, partition and vector files), which contains the basic information for each individual in a network, and the supplementary information in partitions and vectors.

➔ For an introduction to pajek see the pajek manual (pdf).
➔ For a series of macros that can be used to analyze kinship data with Pajek see Tip4Pajek .

c) Excel table format (file extension .xls)

The table format is the most simple format for entering data, specially by anthropologists (without using the Puck data entry form or another genealogy program). It is organized just as a file in text format, with the only difference that the various blocks are on different spreadsheets.
Excel table format is defined as the default format for the exportation of matrimonial census reports as well as for matrices and statistics.

d) Tip format (file extension .tip)

The tip format stores all information in three kind of lines, distinguished by the first number of the line. The different items of each entry are separated by tabs:

0 Identity line: contains

a. the individual's ID number
b. gender number (0 male, 1 female, 2 unknown gender)
c. name (different parts being separated by a slash (/)

1 Kinship line: contains

a. the individual's ID number
b. alter's ID number
c. the code of the kinsip relation (0 father, 1 mother, 2 spouse)

2 Property line: contains

a. the individual's ID number
b. the label of the property, using individual property codes
c. the property value
d. the place (in case of cv events)
e. the date (in case of cv events)
f. alter's ID number (in case of cv events)

There is no particular order prescribed. Although Puck exports data ordered by individuals, any information can be added manually at the end of the file.

e) Gedcom format (file extension .ged)

This is the format used by most genealogy programs (commercial and free). The individual property codes used by Puck correspond, as far as possible, to standard gedcom codes. For an introduction to gedcom formats click here.

f) Prolog format (file extension .pl)

Prolog is a general purpose logic programming language. The program logic is expressed in terms of relations. For more information on the prolog language click here. For an introduction to the use of prolog for representing kinship relations click here. To download a free prolog click here.

In prolog format, all relations and attributes are representing as pairs of the form r(a, b), where r is the name of the attribute or relation, a is the identity number of ego (preceded by the letter p), and b is either an attribute (between simple parentheses) or the identity number of alter (preceded by the letter p). For example :

➔ daughter(p1,p2) means that p1 is p2's daughter
➔ name(p1, "Mary") means that p1 has the name "Mary"

The prolog format for kinship data readable by Puck or the Kinship Editor uses the followig terminology:

➔ relations: father, mother, daughter, son, husband, wife
➔ attributes: gname, sex, info1, info2 etc.

Kinship Relation Notations

A kinship relation can be represented in several different notations. Puck basically uses two of them : the standart notation a the positional notation.

a) Standard notation

The conventional notation of kinship relations uses capital letters for indicating the type of 8 basic kinship relations. These letters are mostly abbreviations of the corresponding English kinship term. They contain information on the gender of Alter and of the direction of the basic kinship relation (ascendance, descendance, marriage, as well as siblingship).

	Male Alter	Female Alter
Ascendance :	F father	M mother
Descendance :	S son	D daughter
Marriage :	H husband	W wife
Siblingship :	B brother	Z sister

The gender of Ego must be indicated by additional signs such as ♂ [male Ego] or ♀ [female Ego] placed before the initial letter.

These basic relations are composed into more complex relations by simple juxtaposition of letters according to their position in the kinship chain, starting from ego (as in English, but contrary, for example, to French, where kinship terms have to be composed starting with alter!). The resulting combination of letters can be read as a direct abbreviation of an English kinship term: MBD (mother's brother's daughter, a matrilateral cross-cousin), ZH (sister's husband, a brother in-law), FWS(father's wife's son, a step-brother) are examples of this.

Half-sibling relations are distinguished from full sibling relations by using explicit combination of ascendance and descendance letters instead of sibling letters: for instance, FS (father's son, paternal half-brother). In addition to genealogical relations, relative age can be indicated by minor letters e (elder) and y (younger) placed before the kinship letter concerned: for instance, FeB (father's elder brother), MyZ (mother's younger sister). Standard kinship notation is highly intuitive and easy to read (at least for anglophones). However, it expresses the ethnocentric viewpoint of English kinship terminology and, by using simple abbreviations, tells us little or nothing about the structure of the kinship relation. It is therefore certainly not the best tool for analytical purposes.

b) Positional notation

In positional notation developed by Laurent Barry (Barry, 2004), a kinship relation is represented by a sequence of letters indicating gender (by abbreviations of the french terms H - homme - for male, and et F - femme - for female) and two diacritical signs:

- the point or full stop "." indicates marriage
- the parentheses () surround an apical position, that is, the position of an individual which is not descendant of any of its neighbors. If both neighbors are spouses, the parentheses may be dropped.
- Relations of ascendance and descendance are indicated by simple juxtaposition, where direction changes after every pair of parentheses and every marriage dot. By convention, the starting direction is ascendance.

By replacing gender letters with the variable X, more comprehensive classes of kinship relations can be represented in positional notation. For instance, X(H)X denotes paternal half-siblings, XX(X)F direct aunts, X(F)FH uterine nephews.

Note that the translation of kinship relations from standard notation (without using ♀ and ♂ signs for the gender of ego) into positional notation always implies the variable letter X in the first position.

Positional notation can be used not only to represent abstract kinship relations, but also concrete kinship chains. In this case, gender letters are replaced by identity numbers of the individuals in the respective positions.

The major advantages of this notation are :

- The clear representation of the structural properties of kinship relations, which is not radically unchanged by symmetry transformations (HF()HF becomes FH()FH, but MBD becomes FZS
- The integration of the sex of ego and not only of alter
- The applicability not only as a notation but as a classification tool (by use of gender variables
- The homogeneity of notations of kinship chains (with individual numbers), kinship relations (with gender letters) and kinship relation classes (with gender variables).

Examples :

Positional		**Standart**
H()F.(H)F	→	ZHD
H(F).H(H)HH	→	MHFSS
FF()F.H	→	MMZH

Property codes

Endogenous and exogenous properties are designated by standard codes. In addition to the standardized codes listed above, you are free to enter any other property label you want.
Warning: *only use single-word codes - Puck does not allow for empty spaces in property codes.*
Note: *property codes are fixed and language-independent. They do not change by switching from one language to another.*

Main endogenous properties

"Endogenous" criteria of classification are calculated by Puck from genealogical data and are derived automatically from the kinship network itself: sibling group size, number of known ascendants, number of spouses, etc. They need not and should not be explicitly specified, and their codes should not be used to enter properties or to load them from a file.

- ALL - a pseudo-property that serves to remove a partition and to restore the unity of the underlying corpus
- SEX - gender
- GEN - generation[3]
- FIRSTN - first name
- LASTN - last name
- FRATP - father, agnatic fratry
- FRATM - mother, uterine fratry
- PATRIC - agnatic apical ancestor, "patrilineage"
- MATRIC - uterine apical ancestress, "matrilineage"
- PATRID - distance to the agnatic apical ancestor, "agnatic generation"
- MATRID - distance to the agnatic apical ancestress, "uterine generation"
- DEPTH - distance to the most remote ancestor, maximal generational depth
- MDEPTH - mean distance to ancestors, mean generational depth (formula of (Cazes & Cazes, 1996))
- PEDG[4] x - number of ascendants (where x is a number specifying generational distance)
- PROG x - number of descendants (where x is a number specifying generational distance)
- SPOU - number of spouses

[3] Generation and generational distance are not unique concepts. Except in kinship networks that consist of trees, there are usually several alternative ways to arrange an individual on a generational level inferior to its ascendants and superior to its descendants. For instance, if a man has married his sister's daughter, his children will be at the same time grand-children and great-grand children of his father. One has to decide on the path along which generational distance shall be calculated.
The algorithm used by Puck is identical to that of Pajek. It consists in navigating through the
network along kinship paths and assigning to parents, spouses and children of each individual the generational level of that individual, augmented by 1, 0 or -1 according to the nature of the kinship tie.
***Note**: the identity of the algorithm does not necessarily imply the network of the results. The result depends on the navigation path which may be different for Pajek and Puck, since arcs are not necessarily stored in the same order.*

[4] The properties PEDG (pedigree) and PROG (progeny) require specification by a number that indicates generational distance. For instance, PEDG 2 is the number of grandparents, PROG 1 the number of children. Attention: there must be a blank between the label "PEDG" or "PROG" and the number that follows

Main exogenous properties

"Exogenous" criteria of classification are those that do not derive from the kinship network itself: dates of birth, death or marriage, profession, residence, religion, etc. Exogenous properties have to be specified explicitly for each individual in the file from which the corpus is loaded or by entering them in the data window. Puck uses the standard gedcom codes for exogenous properties.

- BIRT - birth (place/date/year)
- DEAT - death (place/date/year)
- MARR - marriage (place/date/year/alter)

Note: *Binarizing this property according to place, date or period and using this binarized property for redefining spouses in order to effect a second relational or matrimonial census permits a restricted matrimonial census*

- DIV - divorce (place/date/year/alter)
- BAP - baptism (place/date/year)
- BURI - burial (place/date/year)
- DECO - decoration (place/date/year)
- EDUC - education
- NATI - nationality
- OCCU - occupation
- RELI - religion
- RESI - residence
- TITL – title

The classification of kinship relations and property codes

Among the most basic criteria for classifying kinship relations are the following:

a) According to the arc and edge pattern of lines:
   - Length: the number of arcs and edges included (Roman degree in the case of consanguineous relations)
   - Height: the length of the longest linear chain included (German degree in the case of consanguineous relations)
   - Width: the number of marriage edges included (consanguineous relations have width 1, relinking marriages width 2 or more.)

b) According to the gender pattern of vertices:
   - Descent: agnatic, uterine or cognatic according to the gender of vertices in consanguineous chains
   - Crossness: cross or parallel according to the gender difference of intermediate pairs of vertices in consanguineous chains
   - Terminal crossness: cross or parallel according to the gender difference of terminal pairs of vertices in consanguineous chains

c) According to symmetry features:

   - Skewedness: horizontal, ascending or descending according to differences in the length of the linear chains composing a consanguineous chain
   - Automorphy: percentage of symmetry transformations that leave
   - the kinship relation unchanged

Property codes

Kinship relation property codes are used to cluster kinship relations and matrimonial rings.

Overall kinship relation properties

- SIMPLE - the relation or ring type as such (the "finest" classification: each relation is in a separate class)
- LENGTH - length : the number of links between ego and alter (in consanguineous relations this corresponds to civil or roman degree)
- HEIGTH - height : the maximal number of links to an apical ancestor (in consanguineous relations this corresponds to canonic or germanic degree)
- WIDTH - width: the number of consanguineous components implied in the relation
- SYM - symmetry: the number of automorphic transformations as a percentage of all possible transformations which leave gender and direction invariant
- HETERO - a binary property, true if all married couples as well as the pair ego/alter are heterosexual, false otherwise

Properties of consanguineous relations

For kinship relations implying marriages, the code yield the profile of properties of the implied consanguineous relations

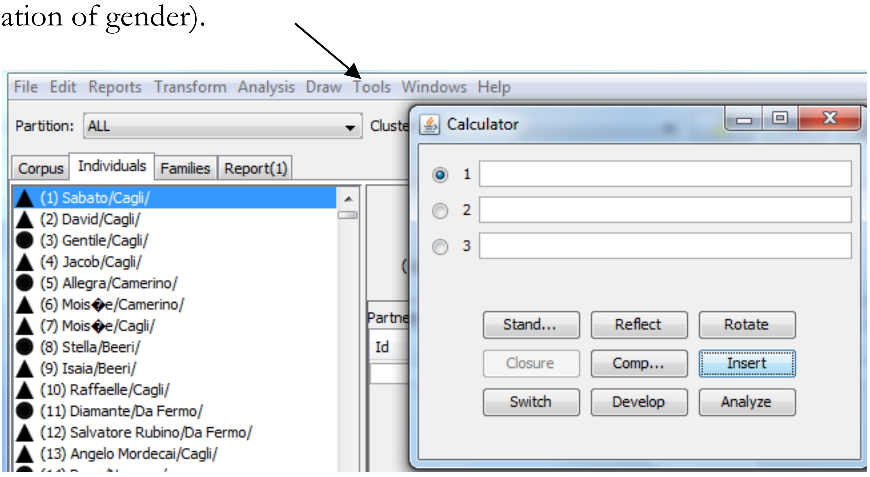
- DEGREE - civil degree (number of links between consanguines)
- ENDS - gender combination of ego/alter
- SKEW - skewedness (generational distance between ego and alter)
- SKEW+ - skewedness (in three classes: horizontal, oblique, alterne)
- LINE - unilinearity type (agnatic, uterine, cognatic, bilateral or identity)
- AGNA - agnatic coefficient (percentage of agnatic links)
- UTER - uterine coefficient (percentage of uterine links)
- DRAV - dravidian crossness
- SWITCHES - number of gender switches
- ARCH - gender combination of the apical siblings (children of the apical ancestor of the relation), not defined for linear relations

Status (allowed / not allowed / not defined) according to particular marriage systems

- DRAV-H - dravidian crossness (horizontal system, Chimane model)
- DRAV-O - dravidian crossness (oblique system, Parakana model)

Kinship calculator

The Kinship calculator serves to convert, transform and analyze kinship relations. The kinship calculator can be accessed from *Tools* → *Calculator*. Kinship relations can be entered in any notation. The calculator contains three lines so as to allow for unary and binary operations. These operations can act on fully specified relations or on a relational schema (without specification of gender).



Partition: ALL

▲ (1) Sabato/Cagli/
▲ (2) David/Cagli/
● (3) Gentile/Cagli/
▲ (4) Jacob/Cagli/
● (5) Allegra/Camerino/
▲ (6) Mois�e/Camerino/
▲ (7) Mois�e/Cagli/
● (8) Stella/Beeri/
▲ (9) Isaia/Beeri/
▲ (10) Raffaelle/Cagli/
● (11) Diamante/Da Fermo/
▲ (12) Salvatore Rubino/Da Fermo/
▲ (13) Angelo Mordecai/Cagli/

You can change the point of view by clicking *Reflection* (inverses ego and alter) and *Rotation* (replaces ego by the next married pivot (not married to ego). In consanguineous relations, equivalent to identity). You have some binary operations for composing and combine kinship relations with *Composition* (composes relation 3 by linking alter of relation 1 to ego of relation 2 by marriage in the heterosexual case, identity in the homosexual case), and *Insertion* (calculates the relation 3 implied between ego and alter of relation 1 if ego's parents are in relation 2). *Develop* all relations of a same type and *analyse* a relation producting the analytic profile of the relation in the Report window.

Tools and concepts

a) Matrimonial networks

Matrimonial networks are subgraphs induced by matrimonial rings. They are line-induced and not vertex-induced subgraphs. This means that for a line to be in the subgraph, it must be part of a ring (it is not enough that its endpoints are in a ring). The matrimonial network derived from a set of matrimonial rings found in a kinship network is thus simply the network composed of these rings. It consists, in other word, of the matrimonially "interesting" regions of the original kinship network.

The connected parts of the matrimonial network (the matrimonial components) represent continuous regions of densely interconnected rings, which may be studied from various perspectives.

- → On the one hand, we may suppose that the frequent occurrence of particular matrimonial patterns is correlated with other properties of the network region concerned (for instance social class, geographical region or historical period); we may then apply several partitions to the network in order to evaluate the degree to which partition clusters correspond to matrimonial components.
- → On the other hand, we may interpret the density of rings as an effect of self-reinforcing social mechanisms (behavior transmission, imitation or the presence of rules) or as a simple network effect (rings combining to compose other rings) which we did not consider when defining the criteria for our initial ring search.

The concept of a matrimonial network is also meaningful in and of itself, independent of any particular ring set. Even without being able to precisely identify all matrimonial rings (without limits of size) which may exist in a kinship network, it is possible to determine which part of the network is composed of matrimonial rings. The result is the absolute matrimonial network, the subgraph induced by all lines in the network which are in some ring whatsoever. This absolute matrimonial network is equivalent to the sum of all matrimonial bicomponents. It corresponds to what has been called the "core" in a P-graph context (Grange & Houseman, 2010; White & Houseman, 1996) .

Every matrimonial network constitutes a network without tails (every vertex must have a degree greater than one) and without structural children (every vertex must have an outdegree greater than zero). However, the reverse is not the case. There may be networks where all vertices fulfill these two degree criteria, but which nevertheless are not matrimonial, as they contain lines which do not form part of any matrimonial ring. Filial triads (father, mother and child) or marriage ties connecting disjoint matrimonial components are instances of this.

b) Matrimonial rings

A matrimonial ring is a chain of kinship (consanguineal and affinal) links closed by a marriage and which does not pass through childless an unmarried individuals (Hamberger, Houseman, Daillant, White, & Barry, 2004).

Pragmatically, matrimonial rings types correspond to types of consanguineal marriage (between consanguineal kin, such as between a man and his mother's brother's daughter) and types of affinal "relinkings" incorporating one, two or more intermediary marriage ties.

➔Consanguineal marriages, that incorporate a single marriage tie (and a single consanguineous kinship chain), form matrimonial rings of “width” 1, e.g. a man marries his mother‟s brother‟s daughter.

➔Relinkings incorporating two marriage ties (and two consanguineous kinship chains), form matrimonial rings of “width” 2, e.g. a man and his sister marry a sister and her brother, or a man marries his mother’s brother’s wife’s bother’s daughter.

➔Relinkings incorporating three marriage ties (and three consanguineous kinship chains) form matrimonial rings of “width” 3, e.g. a man marries his mother’s brother’s wife’s bother’s daughter’s husband’s sister.

c) Matrimonial bicomponent

A bicomponent (or bi-connected component) is a graph in which any two vertices can be linked to each other by two distinct paths (this can be stated equivalently by saying that a bicomponent contains no cut-point whose elimination would cut it into two disconnected components). As a consequence, any two vertices in a bicomponent form part of a cycle (Grange & Houseman, 2010; Hamberger & Daillant, 2008).

Restricting this condition from cycles to matrimonial rings gives rise to the concept of a matrimonial bicomponent, that is, a maximal subgraph in which every two vertices form part of a matrimonial ring. It can also be defined by the condition that any two vertices in a matrimonial bicomponent can be linked to each other by two distinct kinship chains that do not pass through and do not meet in “structural children”.

Matrimonial bicomponents are closely related (but not identical) to matrimonial components (the maximal connected parts of matrimonial networks): both are line-biconnected (two distinct line-series link each vertex to every other), but matrimonial bicomponents have the additional feature of vertex-biconnected as well (the two interconnecting line-series never run through the same vertex).

d) Matrimonial circuit

A matrimonial circuit is a circuit in a kinship network that does not contain a parental triad. As a consequence, it contains at least one (marriage) edge. Matrimonial circuits are indicators of sociological constraints of matrimonial choice (rules, preferences and avoidances, opportunities) and of the dynamics of self-organization of the network. They have to be studied as a whole. For the concept of the matrimonial circuit see (Hamberger et al., 2012, p. 539-540).

e) Some basic notions

- A **path** is an alternating sequence of vertices and lines (arcs or edges), where all vertices are distinct.
- A **cycle** is a path where the first and the last vertex are identical
- A **chain** is a (sub)graph whose vertices and arcs form a single path
- A **circuit** is a (sub)graph whose vertices and arcs form a single cycle
- A kinship chain is **linear** if it consists of uniformly oriented arcs
- A **parental triad** is a graph formed by three vertices and arcs pointing from two of them (the “parents”) to the third (the “child”)
- A **consanguineal** chain contains neither parental triads nor (marriage) edges

f) Kinship network representations

Ore graphs

Named after the scandinavian mathematician Oystein Ore (1970), developed by Vladimir Batagelj and Andrej Mrvar. In an Ore graph, vertices represent individuals, arcs filial ties and edges marriages. Gender is represented by vertex-labels, paternal and maternal ties are represented by two different types of lines.

Tip-graphs

Named after the research group TIP (Traitement Informatique de la Parenté), used by the macros of the Tip4Pajek series (2007). In a Tip-graph, filial and marriage ties are represented by arcs. All information on the type of tie and on gender is contained in line values. There are five types of lines:

- a marriage arc pointing from female to male,
- a filial arc pointing from female (mother) to female (daughter),
- a filial arc pointing from female to male (son),
- a filial arc pointing from male (father) to female,
- a filial arc pointing from male to male.

Because a Tip-Graph does not involve vertex labeling, it is a highly economical representation of a kinship network. Its major disadvantage is that it is not directed acyclic. Many analyses therefore require its being re-transformed into a conventional Ore-graph. To export a dataset in tip-graph format (as a pajek project file) the option "tip" has to be chosen.

P-graphs

Developed by Douglas White and Paul Jorion (1992), used by the homonymous computer program p-graph. In a P-graph couples or unmarried individuals are represented by vertices, married individuals by gender labeled lines running from the couple in which they are partner to the couple of which they are born.

P-graphs have the advantages of being directed acyclic and of incorporating fewer lines and vertices, allowing semi-cycles (that correspond to matrimonial rings in Ore-graphs) to be more easily detected. Note, however, that an individual who marries several times will be represented by several lines. Lines therefore have to be name-labeled in order to distinguish identity from siblingship.

Bibliography

Barry, L. (2004, juin). Historique et Spécificités techniques du programme Genos. Ecole « Collecte et traitement des données de terrains ». Consulté à l'adresse http://llacan.vjf.cnrs.fr/SousSites/EcoleDonnees/extras/Genos.pdf

Barry, L., & Gasperoni, M. (2008). L'oubli des origines. Amnésie et information généalogiques en histoire et en ethnologie. *Annales de démographie historique*, *n° 116*, 53-104.

Cazes, M.-H., & Cazes, P. (1996). Comment mesurer la profondeur généalogique d'une ascendance? *Population*, *51/1*, 117-140.

Grange, C., & Houseman, M. (2010). Objets d'analyse pour l'étude des réseaux de parenté: une application aux familles de la grande bourgeoisie juive parisienne XIXe-XXe siècles. *Annales de démographie historique*, *n° 116*(2), 105-144.

Hamberger, K., & Daillant, I. (2008). L'analyse de réseaux de parenté: concepts et outils. *Annales de démographie historique*, *n° 116*, 13-52.

Hamberger, K., & Gargiulo, F. (2013). Virtual Fieldwork. Modeling Observer Bias in Kinship and Alliance Networks. *Journal for Artificial Societies and Social Simulation*.

Hamberger, K., Houseman, M., Daillant, I., White, D. R., & Barry, L. (2004). Matrimonial ring structures. *Mathématiques et Sciences Humaines. Mathematics and Social Sciences*, (168). Hamberger, K., Houseman, M., & Grange, C. (2009). La parenté radiographiée. *L'Homme*, *n° 191*(3), 107-137.

Hamberger, K., Houseman, M., & White, D. R. (2012). Kinship Network Analysis. In P. Carrington & J. . Scotto (Éd.), *The Sage Handbook of Social Network Analysis* (p. 533-549). Sage Publications.

White, D. R., & Houseman, M. (1996). Structures réticulaires de la pratique matrimoniale. *L'Homme*, *36*(139), 59-85.

INDEX