

“What Makes the Identity of a Scientific Method? A History of the “Structural and Analytical Typology” in the Growth of Evolutionary and Digital Archaeology in Southwestern Europe (1950s–2000s)”

Supplementary material

Sébastien PLUTNIAK^{1,2}

¹ Centre Émile Durkheim, Bordeaux, France

² TRACES laboratory, Toulouse, France

Supplementary material to: Sébastien Plutniak [2022], “What makes the Identity of a Scientific Method? A History of the ‘Structural and Analytical Typology’ in the Growth of Evolutionary and Digital Archaeology in Southwestern Europe (1950s–2000s),” *Journal of Paleolithic Archaeology: Cultural Taxonomies in the Palaeolithic. Old Questions, Novel Perspectives*, ed. by Felix Riede and Shumon T. Hussain, DOI: [10.1007/s41982-022-00119-7](https://doi.org/10.1007/s41982-022-00119-7).

Contents

- 1 Data and Packages
- 2 Table 1: Number of concepts by hierarchical levels in ten versions of the TA’s vocabulary
- 3 Figure 1: Proportion of concepts by hierarchical level by version of the TA taxonomy
- 4 Figure 2: Representation of the TA taxonomy by a directed graph
- 5 Figure 3: Degree centralisation value of the structure of the TA taxonomy by version

1 Data and Packages

The R packages used in the following analyses include: *igraph* (Csárdi and Nepusz 2006) for network analysis, *tidyr* (Wickham and Girlich 2022), *plyr* (Wickham 2011), and *dplyr* (Wickham, François, et al. 2021) for data manipulation and transformation.

```
library(igraph)
library(ggplot2)
library(tidyr)
library(plyr) # for dplyr()
library(dplyr)
```

Loading of the data (one file by version of the “Analytical Typology” taxonomy):

```
v1954 <- read.csv2("data/grille_1954.csv")
v1955 <- read.csv2("data/grille_1955.csv")
v1956 <- read.csv2("data/grille_1956.csv")
v1957 <- read.csv2("data/grille_1957.csv")
v1960 <- read.csv2("data/grille_1960.csv")
```

```

v1964 <- read.csv2("data/grille_1964.csv")
v1968 <- read.csv2("data/grille_1968.csv")
v1972 <- read.csv2("data/grille_1972.csv")
v1974 <- read.csv2("data/grille_1974.csv")
v1986 <- read.csv2("data/grille_1986.csv")
v2000 <- read.csv2("data/grille_2000.csv")

all.versions.df <- rbind(v1954, v1955, v1956, v1957, v1960, v1964,
                        v1968, v1972, v1974, v1986, v2000)
all.versions.df[ all.versions.df==""] <- NA

```

2 Table 1: Number of concepts by hierarchical levels in ten versions of the TA's vocabulary

```

all.versions.df.long <- pivot_longer(all.versions.df,
                                     cols = 2:ncol(all.versions.df),
                                     names_to = "level")
all.versions.df.long <- all.versions.df.long[!is.na(all.versions.df.long$value),]
n.by.lvl.df <- all.versions.df.long %>% group_by(version, level) %>%
  summarise(n = length(unique(value)))
n.by.lvl.df <- pivot_wider(n.by.lvl.df, names_from = version, values_from = n)
n.by.lvl.df <- n.by.lvl.df[order(n.by.lvl.df$level), -1]
n.by.lvl.df <- as.data.frame(n.by.lvl.df)
n.by.lvl.df.out <- rbind(n.by.lvl.df,
                        "Sum" = apply(n.by.lvl.df, 2, sum, na.rm = T) )
n.by.lvl.df.out

```

##	1954	1956	1957	1960	1964	1968	1972	1974	1986	2000
## 1	NA	NA	5	NA	5	4	6	6	5	5
## 2	10	11	12	14	14	15	16	16	16	21
## 3	NA	22	NA	NA	NA	37	51	52	50	56
## 4	56	49	60	80	82	85	105	104	59	85
## 5	NA	NA	NA	NA	NA	NA	NA	NA	NA	13
## 6	66	82	77	94	101	141	178	178	130	180

3 Figure 1: Proportion of concepts by hierarchical level by version of the TA taxonomy

Figure 1.

```

perc.by.lvl.df <- t(n.by.lvl.df) / apply(n.by.lvl.df, 2, sum, na.rm=T)
perc.by.lvl.df <- round(perc.by.lvl.df, 2)
perc.by.lvl.df <- as.data.frame(t(perc.by.lvl.df))
perc.by.lvl.df$level <- as.character(1:nrow(perc.by.lvl.df))
perc.by.lvl.df[is.na(perc.by.lvl.df)] <- 0
perc.by.lvl.df <- pivot_longer(perc.by.lvl.df, cols=1:10,
                              values_to = "proportion",
                              names_to = "version")
perc.by.lvl.df <- perc.by.lvl.df[order(perc.by.lvl.df$proportion),]
perc.by.lvl.df$level <- factor(perc.by.lvl.df$level, levels= 1:5)

```

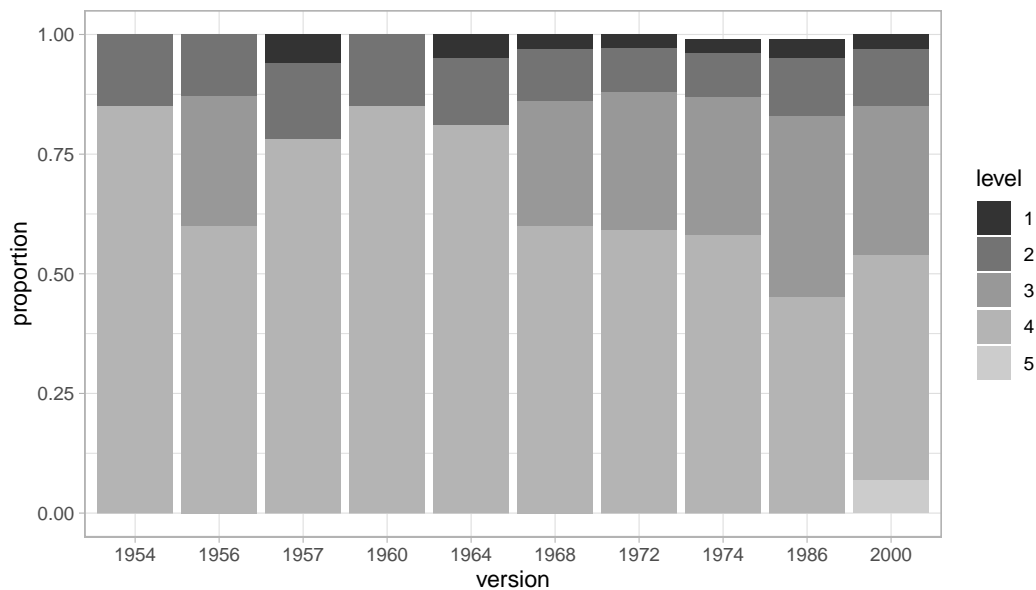


Figure 1: Proportion of concepts by hierarchical level by version of the TA taxonomy.

```
ggplot(perc.by.lvl.df,
  aes(group=level, fill=level, x=version, y=proportion) ) +
  theme_light(base_size=11) +
  geom_bar(position="stack", stat="identity") +
  scale_fill_grey("level")
```

4 Figure 2: Representation of the TA taxonomy by a directed graph

Figure 2.

```
simple1974 <- v1974[v1974$niveau.1 == "simples",]
classification <- rbind(
  cbind(simple1974$niveau.1, simple1974$niveau.2),
  cbind(simple1974$niveau.2, simple1974$niveau.3),
  cbind(simple1974$niveau.3, simple1974$niveau.4)
)
classification <- unique(classification)

g <- graph_from_data_frame(classification)
g <- simplify(g)

V(g)[V(g)$name %in% simple1974$niveau.1]$level <- 1
V(g)[V(g)$name %in% simple1974$niveau.2]$level <- 2
V(g)[V(g)$name %in% simple1974$niveau.3]$level <- 3
V(g)[V(g)$name %in% simple1974$niveau.4]$level <- 4

par(mar=c(0, 1, 0, 1))
plot(g, vertex.size = .01,
  vertex.label.cex = as.numeric(as.character(factor(V(g)$level,
    labels=c(2.5, 1.5, 1.1, 1)))),
  vertex.label.color = as.character(factor(V(g)$level,
```

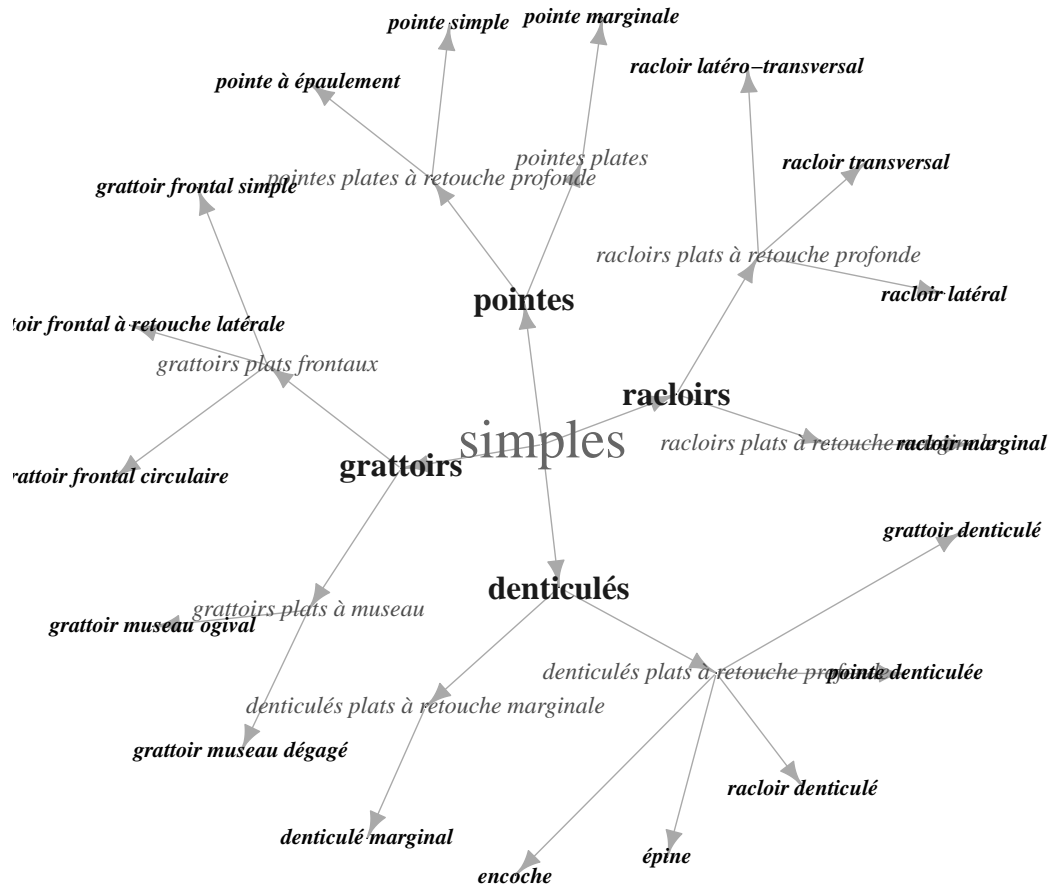


Figure 2: Representation of the TA taxonomy by a directed graph. Partial example from the 1974 version: only the sub-graph corresponding to the “simple” order is represented, with its classes, famille, and primary types.

```

labels=c("gray40", "gray10", "gray30", "black") ),
vertex.label.font = as.numeric(as.character(factor(V(g)$level))),
edge.arrow.size=1, vertex.label.dist=0,
layout = layout_as_tree(g, mode="out",
                        root = which(V(g)$level == "1"),
                        circular=T, flip.y=F)
)

```

5 Figure 3: Degree centralisation value of the structure of the TA taxonomy by version

```

# tag each concept with its hierarchical level:
all.versions.df[ ! is.na(all.versions.df$niveau.1),]$niveau.1 <- paste(
  all.versions.df[ ! is.na(all.versions.df$niveau.1),]$niveau.1, "1", sep="_")
all.versions.df[ ! is.na(all.versions.df$niveau.2),]$niveau.2 <- paste(
  all.versions.df[ ! is.na(all.versions.df$niveau.2),]$niveau.2, "2", sep="_")
all.versions.df[ ! is.na(all.versions.df$niveau.3),]$niveau.3 <- paste(
  all.versions.df[ ! is.na(all.versions.df$niveau.3),]$niveau.3, "3", sep="_")
all.versions.df[ ! is.na(all.versions.df$niveau.4),]$niveau.4 <- paste(

```

```

all.versions.df[ ! is.na(all.versions.df$niveau.4),]$niveau.4, "4", sep="_")
all.versions.df[ ! is.na(all.versions.df$niveau.5),]$niveau.5 <- paste(
  all.versions.df[ ! is.na(all.versions.df$niveau.5),]$niveau.5, "5", sep="_")
# create an edge list
all.versions.list <- plyr::dply(all.versions.df, .(version), .drop = F)
all.versions.list <- lapply(all.versions.list, function(x)
  Filter(function(xx) ! all(is.na(xx)), x) )
# function:
get.hierarchical.edgelist <- function(df){
  times <- ncol(df) - 1
  incr <- 1
  colnames(df) <- NA
  res <- data.frame()
  while( incr < times ){
    res <- rbind(res, df[ , c(1 + incr, 2 + incr, 1) ] )
    incr <- incr + 1
  }
  colnames(res) <- c("sup", "inf", "version")
  res[complete.cases(res), ]
}
# apply function:
all.versions.list2 <- lapply(all.versions.list, get.hierarchical.edgelist)

# make graphs:
g.list <- lapply(all.versions.list2, graph_from_data_frame, directed=T)
g.list <- lapply(g.list, simplify)
# compute centralisation:
deg.centr <- sapply(g.list, function(x)
  centr_degree(x, mode="total", normalize=T)$centralization)
deg.centr <- data.frame(version = names(deg.centr), deg.centr)

```

Figure 3.

```

ggplot(deg.centr, aes(x=version, y = deg.centr)) +
  theme_light(base_size=11) +
  geom_bar(stat="identity") +
  ylab("Degree centralisation value") + xlab("Version")

```

Session information

- R version 4.2.1 (2022-06-23), x86_64-pc-linux-gnu
- Running under: Arch Linux
- Matrix products: default
- BLAS: /usr/lib/libblas.so.3.10.1
- LAPACK: /usr/lib/liblapack.so.3.10.1
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: dplyr 1.0.9, ggplot2 3.3.6, igraph 1.3.2, knitr 1.39, plyr 1.8.7, tidyr 1.2.0
- Loaded via a namespace (and not attached): cli 3.3.0, colorspace 2.0-3, compiler 4.2.1, crayon 1.3.4, digest 0.6.25, ellipsis 0.3.2, evaluate 0.15, fansi 0.4.1, farver 2.1.0, generics 0.1.2, glue 1.6.2, grid 4.2.1, gtable 0.3.0, highr 0.9, labeling 0.4.2, lifecycle 1.0.1, magrittr 1.5, munsell 0.5.0, pillar 1.7.0, pkgconfig 2.0.3, purrr 0.3.4, R6 2.4.1, Rcpp 1.0.8.3, rlang 1.0.2,

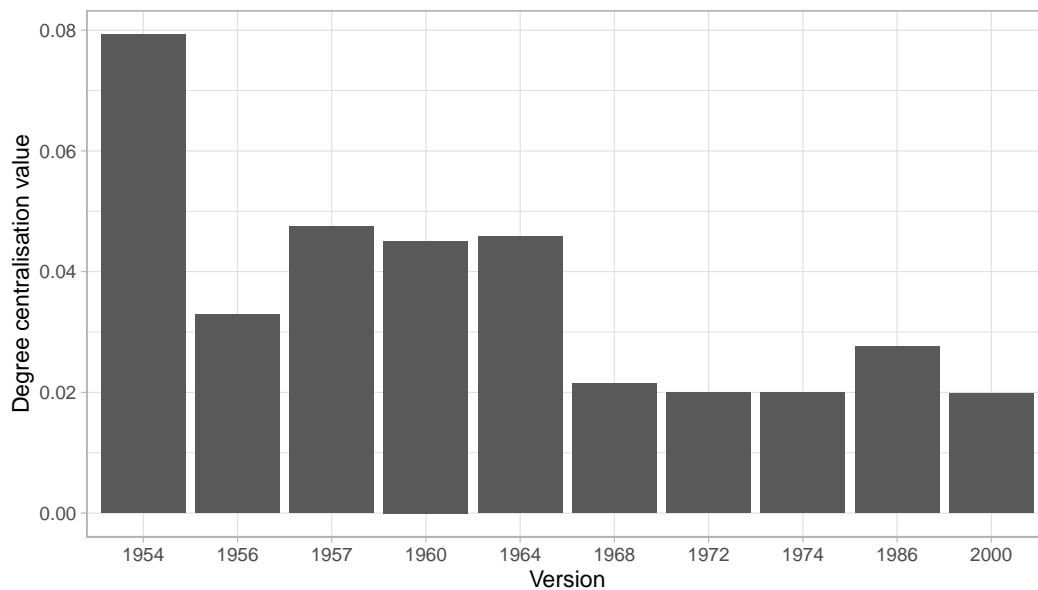


Figure 3: Degree centralisation value of the structure of the TA taxonomy by version.

rstudioapi 0.11, scales 1.2.0, stringi 1.7.6, stringr 1.4.0, tibble 3.1.7, tidyselect 1.1.2,
tools 4.2.1, utf8 1.2.2, vctrs 0.4.1, withr 2.5.0, xfun 0.31

References

- Csárdi, Gábor and Tamás Nepusz [2006], “The igraph Software Package for Complex Network Research,” *InterJournal*, 1695, 5, pp. 1-9, <http://igraph.org>.
- Plutniak, Sébastien [2022], “What makes the Identity of a Scientific Method? A History of the ‘Structural and Analytical Typology’ in the Growth of Evolutionary and Digital Archaeology in Southwestern Europe (1950s–2000s),” *Journal of Paleolithic Archaeology: Cultural Taxonomies in the Palaeolithic. Old Questions, Novel Perspectives*, ed. by Felix Riede and Shumon T. Hussain, doi: [10.1007/s41982-022-00119-7](https://doi.org/10.1007/s41982-022-00119-7).
- Wickham, Hadley [2011], “The Split-Apply-Combine Strategy for Data Analysis,” *Journal of Statistical Software*, 40, 1, pp. 1-29, <http://www.jstatsoft.org/v40/i01/>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller [2021], *dplyr: A Grammar of Data Manipulation*, R package version 1.0.5, <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley and Maximilian Girlich [2022], *tidyr: Tidy Messy Data*, R package version 1.2.0, <https://CRAN.R-project.org/package=tidyr>.