

# Décors à la molette sur des vases en céramique sigillée d'Argonne et leurs dérivées - réflexions pour le traitement des décors répétitifs et cycliques – saison 3.

## Boîte à outils et discussion méthodologique

- type de projet : Activités (ludiques) collatérales d'un projet de publication du corpus
- porteur du projet : Pablo Ciezar (Inrap, UMR 7041 ArScAn, équipe GAMA)
- thématiques de l'atelier concernées : 3 - sériations et typologies des éléments matériels et 4 - exploration multidimensionnelle de sources croisées pour l'archéologie ;

## présentation

### Ciezar, Pablo (Inrap, UMR 7041 ArScAn, équipe GAMA)

Après avoir évoqué le regroupement des séquences à l'aide d'AFC sur les composantes d'une molette (Les molettes - saison 1, voir p. 74 du compte-rendu disponible à cette adresse [https://f-origin.hypotheses.org/wp-content/blogs.dir/4054/files/2018/04/2016\\_CR\\_SITRADA.pdf](https://f-origin.hypotheses.org/wp-content/blogs.dir/4054/files/2018/04/2016_CR_SITRADA.pdf) ), après avoir cherché à calculer des indices de similarités en employant les distances d'édition en considérant une molette comme une chaîne de caractères (Les molettes - saison 2, voir p. 58 du compte-rendu disponible à cette adresse : [https://f-origin.hypotheses.org/wp-content/blogs.dir/4054/files/2018/04/2017\\_CR\\_SITRADA.pdf](https://f-origin.hypotheses.org/wp-content/blogs.dir/4054/files/2018/04/2017_CR_SITRADA.pdf) ), pour cette troisième année je poursuis l'exploration des données, toujours dans la perspective d'une recherche d'indicateurs permettant d'aller vers une "classification" des séquences, ou en tout cas d'obtenir des indices de proximité.

La présentation porte cette fois-ci, sur l'aspect du voisinage physique entre deux motifs dans une même molette, en exploitant quelques paramètres de l'analyse des réseaux. Elle est suivie d'une composante "boîte à outils", en évoquant le logiciel open source "Gephi" (<https://gephi.org>).

Les données utilisées proviennent d'une extraction de la base de données en cours de constitution dans le but de publier un corpus des décors à la molette sur des vase en céramique sigillée d'Argonne et leurs dérivées. C'est un projet conduit depuis de nombreuses années par Paul Van Ossel (Univ. Paris Nanterre – UMR ArScAn, équipe GAMA), Lothar Bakker et Wim Dijkman. Ce travail a conduit à définir une codification simplifiée et schématique des motifs composant une molette afin de faciliter l'identification des séquences.

L'idée cette fois-ci était d'essayer de dégager à l'aide de l'analyse de graphes le rôle de certains motifs dans la composition d'une molette, et à plus grande échelle de vérifier si, à terme, des constantes ne pourraient pas être évoquées.

Chaque molette est ainsi décrite par une chaîne de caractères ou code alphanumérique qui peut ensuite être l'objet de plusieurs traitements comme une séparation en termes contigus.

En partant de cette codification des molettes par des séquences codées (séries numériques), j'ai produit une table associant deux motifs par le fait d'être l'un à côté de l'autre (à droite ou à gauche) dans une même séquence ou molette. À l'aide de quelques packages R pour le traitement des données ou de représentation des graphes, comme **igraph**, quelques premières représentations ont pu être initiées (fig.1).

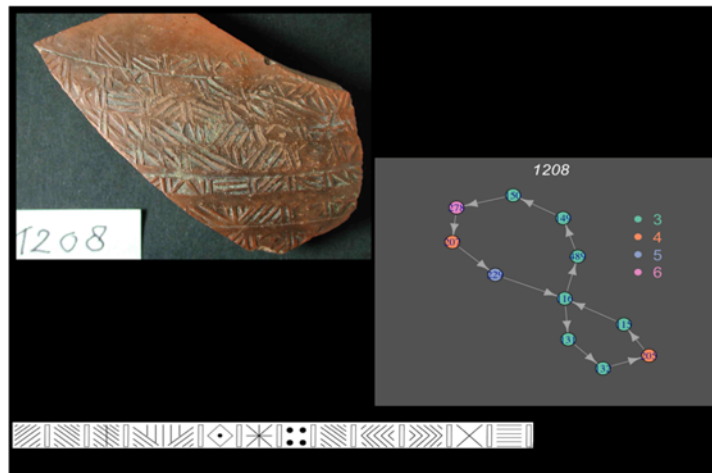


Figure 1 : exemple de chaînage des motifs dans un graphe<sup>1</sup>

Ensuite, la table « Source, Cible » a été étendue à l'ensemble des molettes et un graphe produit avec le même package. Immédiatement les molettes à motifs uniques, ou avec des associations de plusieurs motifs en « vase-clos » et sans liens avec les autres, ont été isolées dans les graphes (fig.2).

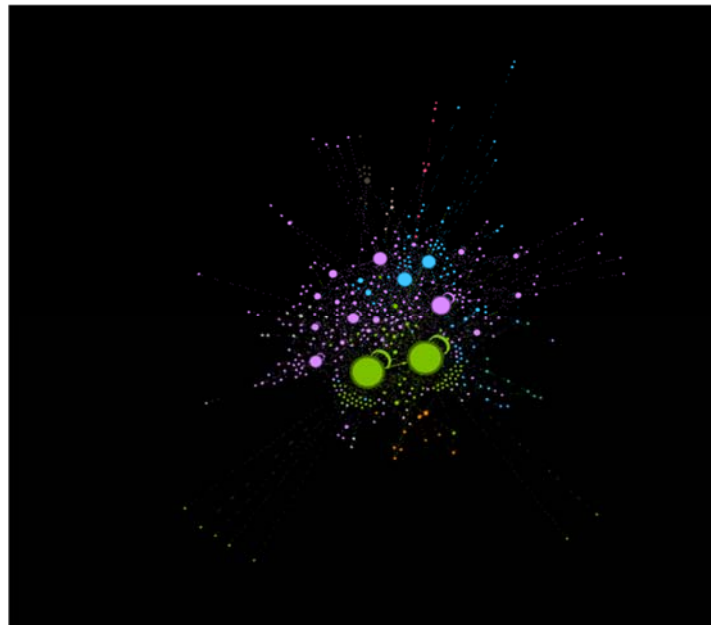


Figure 2 : représentation du graphe avec toutes les associations motif/motif<sup>2</sup>

Par la suite ce nuage a été traité à l'aide du logiciel Gephi.

Pour le reste, la très grande majorité des molettes partagent des motifs en commun, permettant un traitement d'un graphe unique. Malheureusement, le phénomène de répétition de certains motifs sur-représentés (particulièrement les motifs hachurés obliques à droit ou à gauche, dont la fréquence d'apparition est immense) font que tous les autres points gravitent autour (fig.3).

---

<sup>1</sup> Cliché et représentation schématique P. Van Ossel / graphe sous RStudio – package igraph

<sup>2</sup> Package igraph

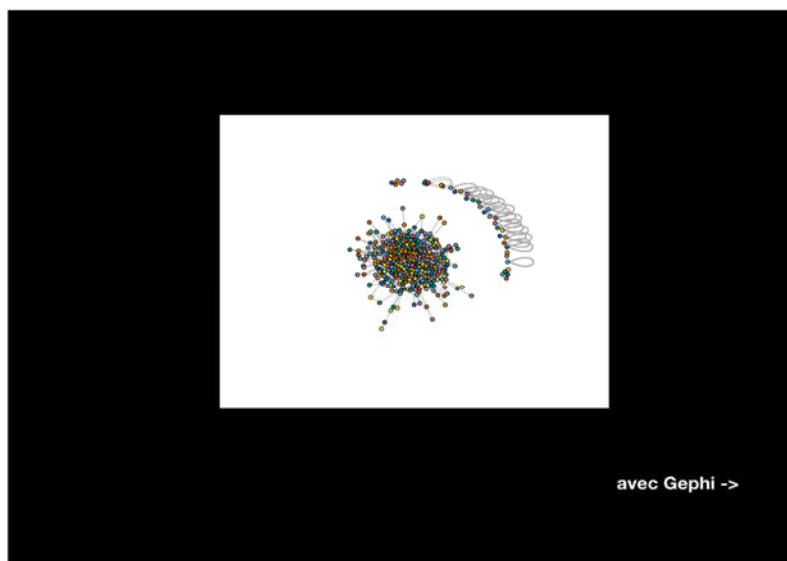


Figure 3 : exemple de l'application de l'algorithme de représentation Atlas-Force 2 sur la première composante connexe (avec Gephi)<sup>3</sup>

Cela reste trop balbutiant pour pouvoir être exposé en l'état et fera l'objet de retraitements qui seront peut-être évoqués dans un prochain atelier, si toutefois ils aboutissent.

## discussion

**BD : Bruno Desachy, JG : Julie Gravier, PC : Pablo Ciezar**

PC : l'ensemble des packages visualisation sur R sont disponibles ici <https://www.r-graph-gallery.com/Treemap>

JG (suite à la mention de l'utilisation de Gephi) : attention, Gephi est parfois critiqué pour utiliser des algorithmes dont la documentation n'est pas toujours précise et détaillée mathématiquement.

JG : pourquoi ne pas tester des graphes avec le nombre de sites en poids ?

PC : pour l'instant je suis dans une démarche totalement exploratoire ; une autre piste consisterait à "spatialiser" le graphe, en introduisant des contraintes géographiques.

BD : parmi les logiciels de fournissent des visualisations de graphe à partir d'un ensemble d'éléments (qui deviennent les sommets du graphe) et d'un ensemble de couples de ces éléments (qui deviennent les arcs ou arêtes) , on peut citer Tulip, un outil qui a été développé à Bordeaux. (Pour plus d'info sur cet outil, voir <http://tulip.labri.fr/TulipDrupal/>)

PC : j'aimerais souligner le côté ludique des multiples manipulations que je fais des données, en accompagnement ou parallèlement au projet de publication du Corpus Argonne conduit par les porteurs historiques du projet. J'évoque ces expérimentations dans les ateliers comme pistes sans avoir aucune certitude qu'elles puissent aboutir ou amener à des conclusions.

BD : C'est très bien comme ça : c'est aussi dans l'esprit de l'atelier de pouvoir venir parler d'essais exploratoires et libres, sans contraintes de « rentabilité » immédiate

Tous : que de pistes prometteuses ! Nous attendons tous avec impatience l'épisode 4... s'intitulera-t-il "Un nouvel espoir"!!!

---

<sup>3</sup>Copie d'écran d'une représentation Gephi.

## suites effectives ou envisagées du projet

Au-delà de l'état du travail exposé et discuté chaque année dans une approche personnelle et expérimentale, d'autres projets sont en cours sur ces données et mobiliseront l'analyse des données : projet de sériation des associations des molettes dans des contextes cohérents, préparation d'un module d'aide à l'identification schématique des séquences dans le cadre de la mise en ligne progressive du corpus.

## références et liens

Pour les traitements sous R (interface d'exploitation RStudio -<https://www.rstudio.com>)

packages utilisés :

tidyverse : <https://cran.r-project.org/package=tidyverse>

treemap : <https://cran.r-project.org/package=treemap>

igraph : Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. <http://igraph.org>

Pour les représentations graphiques nous attirons l'attention sur les pages réalisées par Yan Holtz (INRA) /[www.r-graph-gallery.com/](http://www.r-graph-gallery.com/) qui fournissent une grande aide à élaboration de nombreuses familles de graphiques avec des exemples variés et des fragments de code