



**HAL**  
open science

## Exploration de données temporelles floues : application en archéologie

Cyril de Runz, Éric Desjardin

► **To cite this version:**

Cyril de Runz, Éric Desjardin. Exploration de données temporelles floues : application en archéologie. Atelier Représentation et Raisonnement sur le Temps et l'Espace, Plateforme AFIA, 2011, Chambéry, France, France. pp.26-35. hal-00669222

**HAL Id: hal-00669222**

**<https://hal.science/hal-00669222>**

Submitted on 12 Feb 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exploration de données temporelles floues : application en archéologie

Cyril de Runz, Eric Desjardin

CReSTIC,  
IUT de Reims Châlons Charleville,  
Université de Reims Champagne-Ardenne,  
Rue des Crayères, BP1035, 51687 Reims Cedex 2  
cyril.de-runz@univ-reims.fr et eric.desjardin@univ-reims.fr

**Résumé** : Dans cet article, nous proposons une approche d'analyse temporelle de l'information basée sur l'exploration des périodes d'activités d'un ensemble d'objets archéologiques. En raison de l'imprécision de leurs limites, elles sont représentées par des nombres flous. Notre démarche exploite un indice d'antériorité entre deux nombres flous, afin de déterminer un indice temporel de chaque objet dans l'échantillon. Cet indice est alors utilisé pour extraire les positions temporelles de chaque objet ainsi que des représentants et des regroupements. L'information émergente met en lumière les rapports temporels entre objets. Nous avons appliqué cette démarche aux données portant sur les tronçons de rues de l'époque Romaine.

**Mots-clés** : Relation d'antériorité, rangement, groupement, représentant, intervalle flou, archéologie

## 1 Introduction

Dans un Système d'Information Géographique (SIG) dédié à l'archéologie, l'analyse exploratoire de données cherche à dégager des relations et des corrélations afin d'extraire de nouvelles connaissances entre les objets d'une base de données spatio-temporelles. Dans cet article, nous explorons la base de données afin de déterminer le positionnement temporel de chaque objet. Cette exploration est riche d'information pour l'archéologue.

Le comportement temporel des objets dans leur ensemble fournit de nouvelles connaissances sur l'information portée par ces objets. Ces nouvelles connaissances facilitent la compréhension des données archéologiques et donc l'expertise. Évaluer ce comportement lorsque les périodes étudiées sont imprécisément définies nécessite de mettre en oeuvre des méthodes non classiques. Pour ce faire, nous utilisons une représentation temporelle par des nombres flous et se pose alors le problème complexe de leur comparaison.

En effet, les comparaisons de deux nombres flous sont souvent non transitives (voir (Wang *et al.*, 1995)) ; celles-ci ne sont pas directement utilisées pour le classement de nombres flous dans un ensemble. Cependant, une alternative consiste à positionner chaque nombre flou individuellement dans l'ensemble. Celle-ci a conduit à la définition de nombreuses techniques permettant de positionner un nombre flou relativement à un ensemble de nombres flous et non vis-à-vis de chaque nombre.

Parmi ces méthodes, on trouve l'approche proposée par (Kerre, 1982). Selon lui, après avoir déterminé le maximum de l'ensemble selon le principe d'extension défini dans (Zadeh, 1965), le positionnement d'un élément correspond à sa distance au maximum des éléments. Une autre approche est celle de Jain qui considère les nombres selon leur intersection à un ensemble flou appelé « maximisant » construit sur l'union des supports des nombres flous à positionner (Jain, 1977). Dans la démarche proposée dans cet article, nous proposons d'étudier/explore le comportement temporel de chaque objet de la base de données via sa capacité à être antérieur ou/et postérieur à chaque objet de la base.

Pour cela, nous utilisons l'indice d'antériorité proposé dans (de Runz *et al.*, 2010). A l'aide de cet indice, nous définissons, au niveau de l'ensemble des objets, l'indice temporel de chaque objet par comparaison entre d'une part son aptitude à être antérieur aux autres et d'autre part son aptitude à avoir des objets qui lui sont antérieurs.

Comme l'exploration de données nécessite généralement une métrique, et que définir une métrique sur des informations floues est complexe, nous proposons d'exploiter l'indice temporel en entrée de cette étude. A partir de cet indice, nous pouvons définir le positionnement temporel (le rang) de chaque objet dans un ensemble, puis étudier de possibles groupes dans les objets ; ce que nous ferons en exploitant une démarche proposée dans (Blanchard *et al.*, 2010) pour des données multidimensionnelles, étendue dans Blanchard *et al.* (2011) aux données relationnelles.

Dans le cadre du projet ArchéoChamp (Desjardin & de Runz, 2009), nous proposons de construire une telle approche sur les données portant sur les tronçons de rues romaines découverts lors de fouilles dont les périodes d'activité ont préalablement été modélisées par des nombres flous (afin de représenter les données et leur imprécision). Ces données sont stockées dans la base *BDFRues*.

Ainsi, après avoir présenté dans la section 2 les données étudiées, nous définirons l'indice temporel de chaque objet dans leur ensemble (section 3). Nous exposerons ensuite des premières structures qui apparaissent à l'aide de l'indice (section 4). Nous appliquerons ensuite une approche permettant l'extraction de représentants et de groupements temporels (section 5). Enfin, nous présenterons une conclusion.

## 2 Modélisation des données

Au-delà de l'élaboration de la cartographie archéologique de la ville de Reims et de ses environs à l'époque romaine, le projet ArchéoChamp soutenu par la

région Champagne Ardenne et cadre applicatif de ce travail, porte sur la mise en place d'un Système d'Information Géographique (SIG) pluridisciplinaire.

La base de données spatiotemporelles *BDFRues* est, dans ce cadre, dédiée aux éléments de rues romaines à Reims, *BDFRues*. Dans cette base de données, les tronçons de rues sont caractérisés notamment par une période d'activité.

La datation de la période d'activité des objets est généralement issue d'interprétations ou d'estimations dépendantes de l'environnement de la découverte (lieux de fouilles, stratigraphie, comparaison aux objets se situant dans la même pièce...). De plus, la codification linguistique de périodes temporelles n'a pas toujours la même représentation. Par exemple l'estimation du début du Bas Empire varie selon les experts entre 193 et 284 après J.C. Elle est donc largement imprécise.

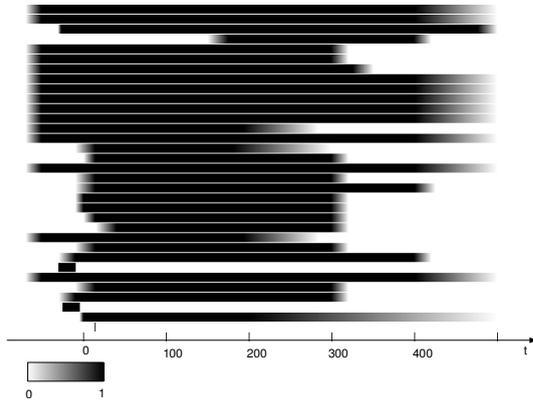


FIGURE 1 – Périodes floues d'activité des objets de *BDFRues* (Chaque nombre flou -chaque période- est représenté par une "bande". Le niveau de gris correspond au degré d'appartenance et l'abscisse au temps).

Nous représentons les périodes d'activité par des ensembles flous convexes et normalisés (généralement des intervalles flous). On peut ainsi prendre en compte cette imprécision. Une représentation visuelle de ces ensembles est proposée dans la figure 1. Les intervalles flous sont définis à l'aide d'intervalles de début et de fin donnés par les experts.

Dans cet article, chaque objet archéologique  $A_i$  aura pour composante temporelle un nombre flou  $A_i.FDate$  ayant pour fonction d'appartenance  $A_i.fdate$ . L'ensemble de tous les objets sera appelé  $\Omega$ .

### 3 Indice temporel

L'analyse exploratoire proposée se base sur la construction d'un indice temporel à partir de l'indice d'antériorité *Ant* défini dans (de Runz *et al.*, 2010).

Ce dernier quantifie une relation binaire, définie sur les nombres flous et interprétée comme antériorité, entre deux nombres flous  $A_i.FDate$  et  $A_j.FDate$  quelconques. Cet indice est construit de la manière suivante dans le cadre de deux objets archéologiques  $A_i$  et  $A_j$  :

$$Ant(A_i, A_j) = \begin{cases} \frac{K_\omega(A_i.FDate)}{K_\omega(A_i.FDate) + K_\omega(A_j.FDate)} & \text{si } K_\omega(A_i.FDate) + K_\omega(A_j.FDate) > 0, \\ 1 & \text{si } K_\omega(A_i.FDate) + K_\omega(A_j.FDate) = 0; \end{cases}$$

avec  $\omega = \{A_i.FDate, A_j.FDate\}$ .  $K$  correspond dans cette formule à l'indice de Kerre (Kerre, 1982) défini comme suit :

$$K_\omega(A_i.FDate) = D_H(A_i.FDate, \widetilde{max}(\omega))$$

avec

$$D_H(A_i.FDate, A_j.FDate) = \int |A_i.fdate(x) - A_j.fdate(x)| dx.$$

et  $\widetilde{max}$  est le maximum selon le principe d'extension de Zadeh.

L'indice temporel d'un objet  $A_i$  doit permettre de définir sa position temporelle. Son calcul doit donc prendre en considération à la fois son aptitude à être antérieur ainsi que celle à ce qu'on lui soit antérieur. C'est pourquoi, l'indice temporel  $IndTemp_\Omega(A_i)$  de  $A_i$  est déterminé comme suit :

$$IndTemp_\Omega(A_i) = \sum_{A_j \in \Omega} Ant(A_j, A_i) - Ant(A_i, A_j).$$

Les valeurs de cet indice appartiennent à l'intervalle  $[-card(\Omega), card(\Omega)]$ .

## 4 Positions temporelles et premiers regroupements

### 4.1 Classes temporelles

D'après la définition de notre indice temporel, si l'appétence d'un objet à ce que l'on lui soit antérieur est supérieure (resp. inférieure) à son appétence à être antérieur aux autres, son indice temporel sera positif (resp. négatif). On peut donc considérer qu'un objet ayant un indice temporel négatif peut être qualifié, par rapport à l'ensemble des objets, comme un objet « plutôt antérieur » tandis qu'un objet ayant un indice temporel positif correspond à un objet « plutôt postérieur ». Nous proposons ainsi une classification temporelle simple des objets. La figure 2 présente les objets de *BDFRues* selon ces deux classes.

Pour les données de *BDFRues*, la classe des objets « plutôt postérieurs » regroupe deux fois plus d'objets que celle des objets « plutôt antérieurs ». On peut

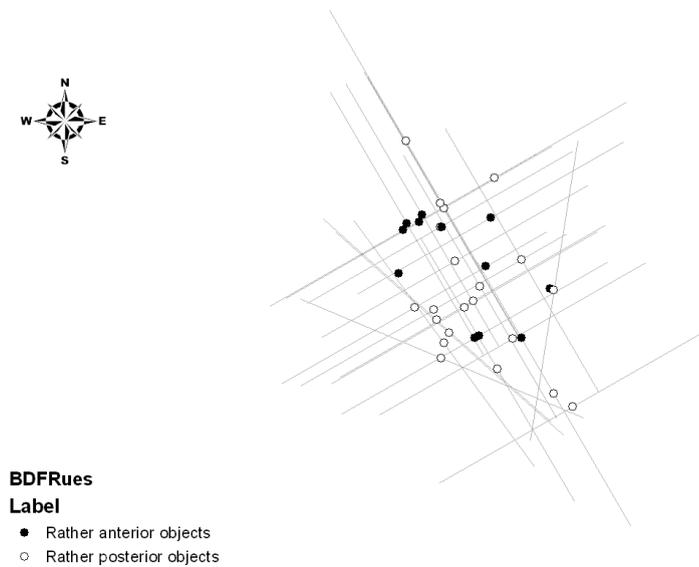


FIGURE 2 – Carte des objets de *BDFRues* classés en « plutôt postérieurs », « plutôt antérieurs »

supposer que les objets « plutôt antérieurs » ont des valeurs de l'indice d'antériorité avec les autres objets proches de 1. Cela expliquerait le fort différentiel entre le cardinal de la classe des objets « plutôt postérieurs » avec celui de la classe des objets « plutôt antérieurs ». Cette hypothèse est d'ailleurs vérifiée puisque, sachant qu'il y a trente trois objets :

- la valeur de l'indice temporel de l'élément le plus antérieur est de  $-29.8$  ;
- la plus grande valeur d'indice temporel des éléments « plutôt antérieurs » est  $-5.4$  ;
- plus de 72% des objets « plutôt postérieurs » ont une valeur d'indice temporel supérieure à 6.

La bipolarisation des valeurs est donc pertinente. Cependant, on peut remarquer que trois objets ont une valeur de l'indice temporel supérieur à 0 mais proche de 0. Ces trois éléments ont une période d'activité singulière débutant au début de la période d'étude mais finissant à un peu plus de sa moitié. Ils ont donc un comportement temporel très particulier en étant presque tout autant antérieur que postérieur.

## 4.2 Positions temporelles

À l'aide de l'indice temporel, nous pouvons ranger simplement les données. Si la valeur de l'indice d'un objet  $A_i$  est plus grande que celui de  $A_j$  alors son rang sera plus grand. Nous appelons ce rang la position temporelle. Pour l'objet  $A_i$ , il sera noté  $PosTemp_{\Omega}(A_i)$ . Le principe de rangement est le suivant :

Si  $IndTemp_{\Omega}(A_i) > IndTemp_{\Omega}(A_j)$ , alors  $PosTemp_{\Omega}(A_i) > PosTemp_{\Omega}(A_j)$ .

La position temporelle de  $A_i$  correspond donc au rang de  $A_i$  dans la liste des objets de  $\Omega$  ordonnée selon les valeurs de l'indice temporel.

Bien que les positions temporelles obtenues par notre approche peuvent différer des rangs obtenus par les méthodes classiques (Kerre, Jain, etc.), les positions temporelles ont une interprétabilité plus forte que les rangs issus de Kerre ou de Jain. En effet, la position temporelle d'un objet archéologique est obtenue à l'aide de l'indice temporel de l'objet. Ainsi, elle reflète donc la relation temporelle de l'objet à l'ensemble des autres. La figure 3 fournit les positions temporelles des objets de *BDFRues*. On peut noter que les positions temporelles fournissent un pré-ordre total d'un ensemble de nombres flous.

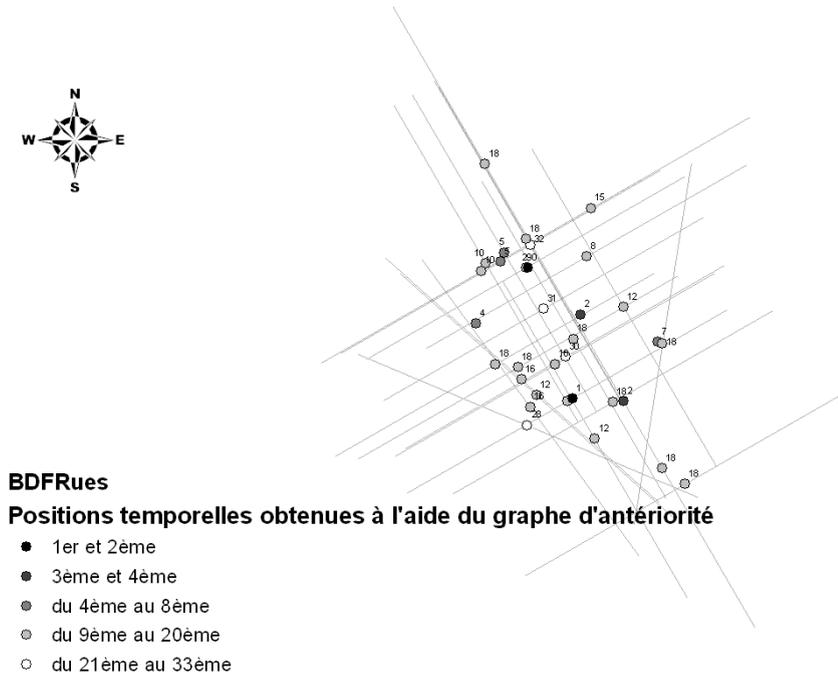


FIGURE 3 – Carte des positions temporelles des 33 objets de *BDFRues* selon leurs périodes d'activité

On peut de plus extraire trois objets particuliers : l'objet le plus antérieur, l'objet le plus postérieur et l'objet temporellement médian. L'objet le plus ancien dans le cadre applicatif, c'est-à-dire le plus antérieur, noté  $PA$ , est celui dont la valeur de l'indice temporel est la plus petite dans l'ensemble des valeurs de l'indice temporel des objets de  $\Omega$ . La position temporelle de l'objet le plus antérieur est la position temporelle minimale des objets de  $\Omega$  :

$$PosTemp(PA) = \min_{A_i \in \Omega} (PosTemp(A_i))$$

L'objet le plus récent dans le cadre applicatif, c'est-à-dire le plus postérieur, noté  $PP$ , est celui dont la valeur de l'indice temporel est la plus grande dans l'ensemble des valeurs de l'indice temporel des objets de  $\Omega$ . La position temporelle de l'objet le plus antérieur est la position temporelle maximale des objets de  $\Omega$  :

$$PosTemp(PP) = \max_{A_i \in \Omega} (PosTemp(A_i))$$

Grâce à l'approche par rang (position temporelle), il est trivial de définir l'objet temporellement médian. Il est celui dont la valeur de l'indice temporel est médiane à l'ensemble des valeurs de l'indice temporel pour les objets de  $\Omega$ . Le médian étant en statistique un estimateur robuste de position dans un échantillon, l'objet temporellement médian est donc un estimateur de position de l'ensemble des objets étudiés. Ces objets particuliers sont identifiables dans la figure 4.

Les classes et les positions extraites à l'aide de l'indice temporel ne fournissent pas directement des informations sur de possibles groupes de données proches temporellement ni sur les représentants de ces groupes.

## 5 Représentants et groupes temporels

Basée sur la théorie du vote de Borda, l'approche exploratoire proposée dans (Blanchard *et al.*, 2010, 2011) a pour objectif de faire émerger des représentants et des groupes de l'ensemble étudié. Pour cela, elle se base sur la construction d'un graphe de représentativité à l'aide de l'établissement du degré de représentativité de chaque élément dans l'ensemble.

### 5.1 Degré et graphe de représentativité

Le calcul du degré de représentativité se décompose en trois étapes :

- Chaque objet de  $\Omega$  évalue tous les objets de  $\Omega$  afin de définir des relations valuées entre les objets, deux à deux.
- A partir de ces relations, chaque objet range les autres par ordre de préférence. Ces rangs sont ensuite transformés en scores.
- Le degré de représentativité d'un objet  $A_i$  résulte de l'agrégation des scores de rangs obtenus par  $A_i$  pour l'ensemble des objets de  $\Omega$ .

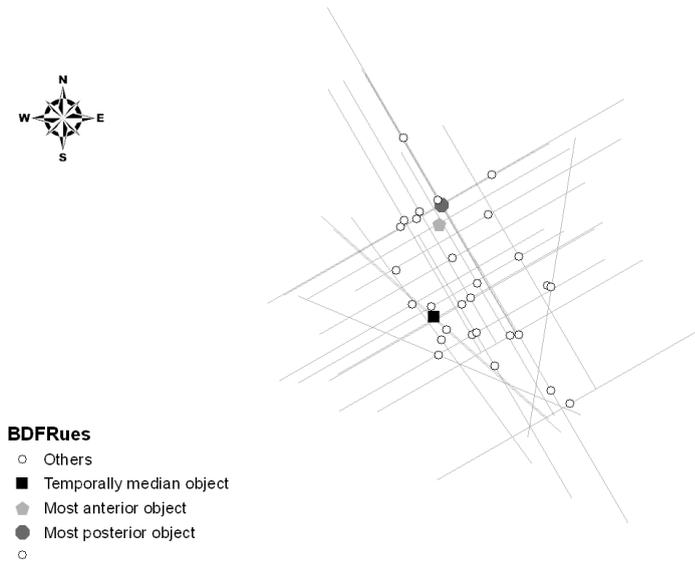


FIGURE 4 – Carte des objets de *BDFRues* surlignant l’objet le plus antérieur, le plus postérieur et le temporellement médian.

Le rang de  $A_i$  selon  $A_j$  noté  $RG_{\Omega}^{A_j}(A_i)$  est, dans le cadre de notre travail, défini à l’aide des valeurs de l’indice temporel. Considérons trois objets  $A_i, A_j, A_k$  et la distance euclidienne  $D$  entre les valeurs de l’indice temporel. Si

$$D(IndTemp_{\Omega}(A_i), IndTemp_{\Omega}(A_j)) < D(IndTemp_{\Omega}(A_k), IndTemp_{\Omega}(A_j))$$

alors

$$RG_{\Omega}^{A_j}(A_i) < RG_{\Omega}^{A_j}(A_k).$$

$RG_{\Omega}^{A_j}$  prend ces valeurs dans l’intervalle  $[1, card(\Omega)]$  et  $RG_{\Omega}^{A_j}(A_j) = 1$ .

Chaque objet classe les autres objets selon ce principe. Le degré de représentativité d’un objet  $A_i$  dans  $\Omega$  est défini comme suit :

$$DR_{\Omega}(A_i) = \sum_{A_j \in \Omega} (card(\Omega) - RG_{\Omega}^{A_j}(A_i) + 1)$$

On le calcule pour l’ensemble des objets  $\Omega$ .

## 5.2 Représentants et groupements

La deuxième partie de l’approche consiste à faire émerger une structuration des objets étudiés. Un graphe des représentants, appelé  $GR_{\Omega}$  défini par  $\langle X, U \rangle$

( $X$  est l'ensemble des sommets,  $U$  l'ensemble des arcs), est alors construit à l'aide d'un voisinage au sens des  $k$  plus proches voisins (selon les proximités des valeurs des indices temporels). À chaque objet  $A_i$  de  $\Omega$  est associé un sommet  $i$  et ce dernier est relié au sommet représentant l'objet dans son voisinage dont le degré de représentativité est le plus élevé, c'est à dire :

$$i \in X \Leftrightarrow A_i \in \Omega$$

$$\forall i, j \in X, (i, j) \in u \Leftrightarrow A_j = \underset{A_s \in V_k(A_i)}{\operatorname{argmax}} (DR_\Omega(A_s))$$

Dans ce graphe, chaque composante connexe forme un groupement de données et un arbre. Le nombre de groupements (de composantes connexes) est dépendant du voisinage choisi. Les éléments à la racine des arbres qui forment les composantes connexes sont les meilleurs représentants des groupes. Ils ont aussi le degré de représentativité le plus élevé dans le groupe.

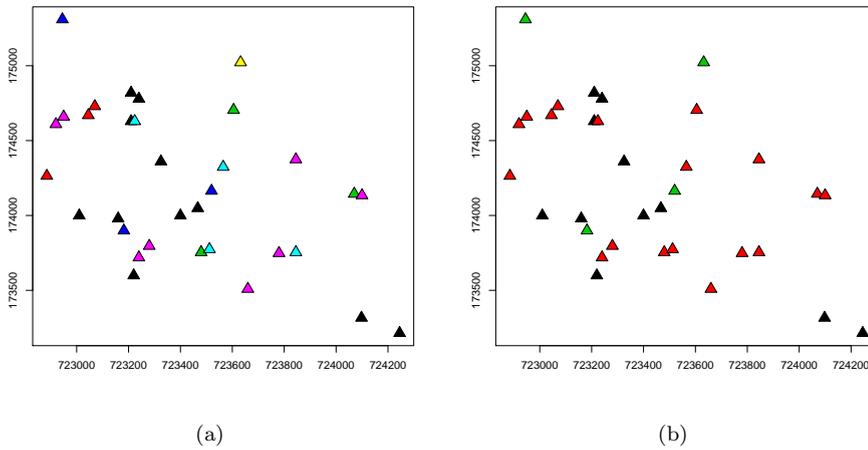


FIGURE 5 – Objets de *BDFRues* regroupés selon leur proximité d'indice temporel par le graphe de représentativité : (a) avec un *3-ppv* et (b) avec un *4-ppv*.

En utilisant cette approche sur les données de *BDFRues* nous obtenons, à l'aide d'un *3-ppv* et d'un *4-ppv*, les cartes présentées dans la figure 5. Nous passons de 8 groupements trouvés dans la figure 5a à 3 dans la figure 5b. Ces structures et leur évolution donnent une information complémentaire aux informations précédemment extraites. Ainsi, par exemple, dans l'approche *3-ppv* les groupements de données singulières peuvent être étudiés à l'instar des données d'indice temporel proche de 0, tandis que dans la recherche à l'aide du *4-ppv*, nous travaillerons plus sur les grandes tendances (*e.g.* objets présents sur toute la période Gallo-Romaine) et sur le ralliement des objets singuliers à ces dernières.

## 6 Conclusion

Dans cet article, nous avons proposé un processus exploratoire ayant pour but l'analyse des relations entre objets selon la temporalité de l'information archéologique disponible. Pour cela, nous avons construit un indice temporel de chaque objet dans l'ensemble des objets. A partir de cet indice, nous avons étudié des structures émergentes construites à partir d'une interprétation de l'indice, d'un rangement des objets, et d'une méthode de fouille de données.

Dans leurs expertises, les archéologues évaluent les objets qu'ils sont susceptibles de trouver dans un site tant d'un point de vue fonctionnel que temporel. Ils ont notamment besoin d'étudier les relations temporelles entre les objets stockés dans les bases de données afin de (i) regarder si la logique temporelle est respectée, et (ii) dégager une évolution temporelle de la cité. Dans cet objectif, ils peuvent utiliser les approches proposées dans cet article.

## Références

- BLANCHARD F., DE RUNZ C., HERBIN M. & AKDAG H. (2011). Représentativité et graphe de représentants : une approche inspirée de la théorie du choix social pour la fouille de données relationnelles. In *Atelier Fouille de Données Complexes, Conférence Extraction et Gestion des Connaissances 2011*, p. 73–83, Brest, France.
- BLANCHARD F., VAUTROT P., AKDAG H. & HERBIN M. (2010). Data representativeness based on fuzzy set theory. *Journal of Uncertain Systems*, **4**(3), 216–228.
- DE RUNZ C., DESJARDIN É., PIANTONI F. & HERBIN M. (2010). Anteriority index for managing fuzzy dates in archaeological gis. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, **14**(4), 339–344.
- DESJARDIN É. & DE RUNZ C. (2009). Gissar : de la saisie de fouilles à l'analyse spatiotemporelle en archéologie. In *Spatial Analysis and GEomatics*, Paris, France.
- JAIN R. (1977). A procedure for multiple-aspect decision making using fuzzy set. *Internat. J. Systems Sci.*, **8**, 1–7.
- KERRE E. E. (1982). The use of fuzzy set theory in electrocardiological diagnostics. In M. GUPTA & E. SANCHEZ, Eds., *Approximate Reasoning in Decision-Analysis*, p. 277–282. North-Holland Publishing Company.
- WANG X., KERRE E. E., CAPPELLE B. & RUAN D. (1995). Transitivity of Fuzzy Orderings Based on Pairwise Comparis. *The Journal of Fuzzy Mathematics*, **3**(2), 455–463.
- ZADEH L. A. (1965). Fuzzy Sets. *Information Control*, **8**, 338–353.