

Exploration temporelle de données archéologiques imprécises : graphe d'antériorité.

Cyril de Runz, Eric Desjardin

CRéSTIC-SIC

IUT de Reims Châlons Charleville

Rue des Crayères, BP 1035, 51687 Reims Cedex 2

{cyril.de-runz,eric.desjardin}@univ-reims.fr,

<http://crestic.univ-reims.fr>

Résumé. Dans cet article, nous proposons une nouvelle technique d'exploration temporelle d'un ensemble d'objets archéologiques dont les périodes d'activité sont représentées par des nombres flous. Pour cela, en se basant sur la définition d'un indice d'antériorité entre deux nombres flous, on construit un graphe orienté pondéré dont les sommets seront les objets archéologiques. À l'aide de ce graphe d'antériorité, nous déterminons le potentiel d'antériorité, de postériorité ainsi que la position temporelle de l'objet associé au sommet dans l'ensemble des objets de la base. L'information dégagée par ce graphe met en lumière les rapports temporels entre objets. Nous avons appliqué cette démarche aux données portant sur les tronçons de rues datant de l'époque Romaine et trouvés à l'époque romaine.

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 spatiotemporelles. Par exemple, dans de Runz et al. (2008), nous avons proposé une méthode pour l'extraction de représentants d'un ensemble d'objets archéologiques dans un SIG. 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. C'est d'ailleurs une démarche largement utilisée pour l'étude stratigraphique d'un chantier archéologique par le diagramme de Harris¹.

Le positionnement temporel des objets fournit de nouvelles connaissances sur l'information portée par ces objets. Ces nouvelles connaissances facilitent la compréhension des données archéologiques et donc l'expertise. Dans cet article, nous positionnerons temporellement, dans une base de données spatiotemporelles dédiée à l'archéologie, chaque objet archéologique par rapport aux autres objets, dont les périodes d'activité sont représentées par des nombres flous.

Comme 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 retrouve l'approche proposée par Kerre (1982). Pour Kerre, 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 de regarder le positionnement 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. (2006, 2008)). En utilisant cet indice sur un ensemble d'objets archéologiques dont les périodes d'activité sont représentées par des nombres flous, nous proposons de définir un graphe orienté pondéré dont les sommets représentent les tronçons de rues et dont les arcs ont pour coût l'indice d'antériorité de l'origine de l'arc par rapport à la destination.

¹Diagramme de Harris : méthode de représentation qui à l'aide de traits et de cases modélise les rapports qu'entretiennent les couches stratigraphiques entre elles.

La capacité d'antériorité d'un objet est déterminée par la somme des coûts des arcs sortants du sommet associé à l'objet. La capacité de postériorité d'un objet est définie par la somme des coûts des arcs entrants dans le sommet. L'indice temporel d'un objet est le différentiel entre sa capacité de postériorité et sa capacité d'antériorité.

Le positionnement temporel de chaque objet dans un ensemble correspond au rang obtenu par la valeur de son indice temporel dans l'ensemble des valeurs de l'indice temporel de tous les objets. Ce graphe permet donc une représentation synthétique et formelle des structures temporelles de l'échantillon.

Dans le cadre du programme SIGRem (Pargny et Piantoni (2005)), nous proposons de construire un tel graphe sur les données portant sur les tronçons de rues romaines dont les périodes d'activité ont préalablement été modélisées par des nombres flous². Ces données sont stockées dans la base *BDFRues*. Par la construction de ce graphe sur les objets de *BDFRues*, nous proposons donc de déterminer le positionnement temporel de chaque objet. À partir de ces positions temporelles, nous extrayons de *BDFRues* l'objet le plus antérieur ou l'objet le plus postérieur.

Afin d'illustrer cet article, le cas de trois objets spatiotemporels A_1 , A_2 et A_3 , dont les composantes temporelles sont respectivement représentées par les nombres flous $A_1.fDate = (0, 48, 68)$, $A_2.fDate = (22, 22, 57, 57)$ et $A_3.fDate = (11, 33, 81)$, dont les fonctions d'appartenance sont présentées figure 1, sera considéré.

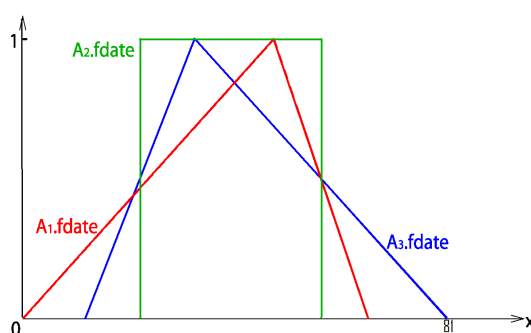


FIG. 1 – Fonctions d'appartenance $A_1.fdate$, $A_2.fdate$ et $A_3.fdate$ de respectivement $A_1.fDate$, $A_2.fDate$ et $A_3.fDate$

Ainsi, après avoir présenté la démarche classique pour le positionnement d'un nombre flou dans son ensemble, nous étudierons via la construction du graphe, la position temporelle de chaque objet par l'intermédiaire de sa capacité d'antériorité et de sa capacité de postériorité dans l'ensemble des objets. Nous terminerons cet article par l'application de notre démarche sur les données archéologiques contenues dans *BDFRues*.

2 Rangement de nombres flous dans un ensemble

Considérons un ensemble Ω de n nombres flous $\{F_1, \dots, F_n\}$. Les méthodes usuelles de classement les rangent par comparaison des valeurs associées à chacun d'entre eux et calculées selon un indice. Celles-ci sont obtenues à l'aide d'une ou plusieurs valeurs de référence (val_{ref}) définies sur l'ensemble des nombres flous (Wang et Kerre (2001a)). Dans Ω , la position d'un nombre flou F_i correspond au rang que confère la valeur de l'indice pour F_i au regard des valeurs de l'indice pour tous les autres nombres flous.

Classiquement on retrouve deux grands types de méthodes de rangement. Dans le premier type, les techniques (*c.f.* Jain (1977) et ses dérivées) se basent sur un ensemble flou de référence, appelé ensemble flou maximisant Ω . Dans le second type (*c.f.* Kerre (1982) et ses dérivées), les méthodes utilisent l'ensemble flou maximum sur Ω , au sens de l'extension de Zadeh.

2.1 Approche de Jain et ses dérivées

Dans l'approche de Jain (1977), considérant une valeur $k > 0$ donnée, on définit un ensemble flou F_{max}^k maximisant Ω dont la fonction d'appartenance f_{max}^k est :

$$f_{max}^k(x) = \left(\frac{x}{x_{max}} \right)^k,$$

²La modélisation floue des objets archéologiques a été présentée dans de Runz et al. (2007).

avec $k \in \mathbb{R}^{+*}$, $x \in \bigcup_{i=1}^n \text{Support}(F_i)$, $x \geq 0$, $x_{max} = \sup(\bigcup_{i=1}^n \text{Support}(F_i))$.

Cela est valable si : $\min(\bigcup_{i=1}^n \text{Support}(F_i)) \geq 0$. Si ce n'est pas le cas on procède à un changement de variable.

L'indice de Jain d'un nombre flou F_i ($J_{\Omega}^k(F_i)$) est alors la hauteur de la t -norme proposée par Zadeh entre F_i et F_{max}^k :

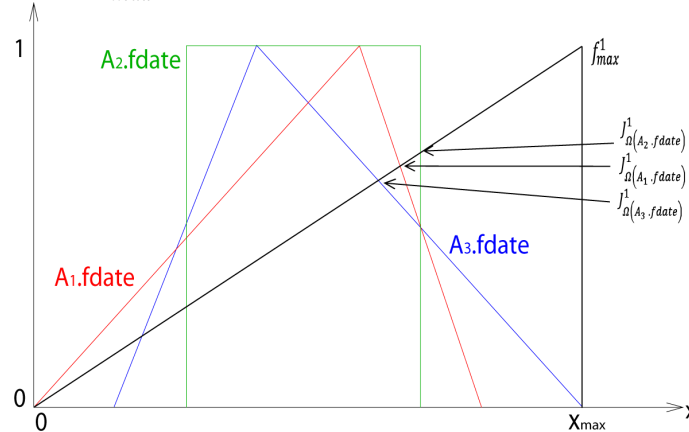
$$J_{\Omega}^k(F_i) = \text{Hauteur}(F_i \wedge F_{max}^k)$$

Soient F_i et F_j appartenant à Ω , si $J_{\Omega}^k(F_i)$ est plus grand que $J_{\Omega}^k(F_j)$ alors F_i aura un plus haut rang que F_j .

Différentes approches dérivées de celles de Jain ont été proposées dans la littérature. Par exemple, l'approche de Chen (1985) complète celle de Jain par l'ajout d'une valeur de référence : l'ensemble flou minimisant Ω . A l'aide des ensembles flous maximisant et minimisant Ω , il détermine une utilité gauche et une utilité droite d'un nombre flou sur lesquelles il base le calcul de l'indice pour le nombre flou considéré.

Pour les objets A_1 , A_2 et A_3 dont les périodes d'activité sont illustrées dans la Figure 1, Ω est composé de $A_1.fDate$, $A_2.fDate$ et $A_3.fDate$. La fonction d'appartenance f_{max}^1 de l'ensemble flou F_{max}^1 maximisant Ω pour $k = 1$ est illustrée dans la figure 2.

FIG. 2 - f_{max}^1 pour $\{A_1.fdate, A_2.fdate, A_3.fdate\}$ avec $k = 1$



Les valeurs de l'indice de Jain pour les représentations des périodes d'activité des objets A_1 , A_2 et A_3 , pour $\Omega = \{A_1, A_2, A_3\}$ et pour $k = 1$ sont :

$$J_{\Omega}^1(A_1.fDate) = 0.67, \quad J_{\Omega}^1(A_2.fDate) = 0.71, \quad J_{\Omega}^1(A_3.fDate) = 0.63.$$

Donc $J_{\Omega}^1(A_3.fDate) < J_{\Omega}^1(A_1.fDate) < J_{\Omega}^1(A_2.fDate)$. Le rangement selon Jain pour $k = 1$ sera donc dans l'ordre croissant des rangs $A_3.fDate$, $A_1.fDate$, et $A_2.fDate$. Les objets sont donc positionnés dans l'ordre croissant de la manière suivante : A_3 , A_1 et A_2 .

2.2 Approche de Kerre et ses dérivées

La proposition de Kerre est la suivante : si on prend l'ensemble des nombres flous à classer, et que l'on compare les distances séparant chaque nombre flou du maximum de l'ensemble, alors on peut ranger les nombres flous par ordre de distance décroissante ; le plus petit nombre aura la distance au maximum la plus grande.

Pour un ensemble Ω de n nombres flous $\{A_1, A_2, \dots, A_n\}$, Kerre propose de ranger ces nombres flous par la comparaison de leur distance de Hamming au maximum de l'ensemble Ω selon le principe d'extension Kerre (1982).

La distance de Hamming entre deux nombres flous F et G de fonctions d'appartenance f et g est définie dans Wang et Kerre (2001b) comme suit :

$$D_H(F, G) = \int |f(x) - g(x)| dx.$$

La distance de Hamming entre A_i , avec $i \in [1, n]$, et $\widetilde{max}(A_1, A_2, \dots, A_n)$ est appelée indice de Kerre $K_\Omega(A_i)$. Ainsi, l'indice de Kerre de A_i dans $\{A_1, A_2, \dots, A_n\}$ est obtenu de la manière suivante :

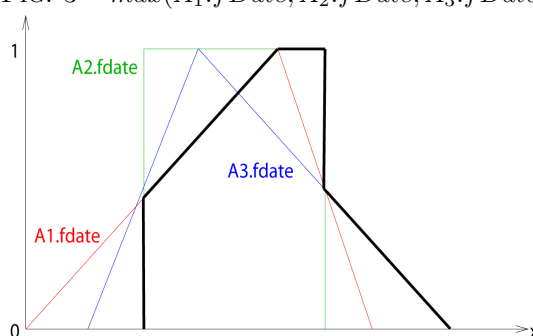
$$K_\Omega(A_i) = D_H(A_i, \widetilde{max}(A_1, A_2, \dots, A_n)) \tag{1}$$

Pour Kerre, un nombre flou est plus petit qu'un autre dans un ensemble donné si et seulement si son indice de Kerre est supérieur à celui de l'autre. C'est à dire $A_i \preceq A_j$ par rapport à $\Omega = \{A_1, A_2, \dots, A_n\}$, avec $(i, j) \in [1, n]$, si et seulement si $K_\Omega(A_i) \geq K_\Omega(A_j)$.

Depuis, différentes approches se sont basées sur l'évaluation de la proximité de chaque nombre flou F_i au maximum ou au minimum de Ω . Ainsi, par exemple, en 1987, Wang (voir dans Wang et Kerre (2001a)) proposa d'utiliser d'autres méthodes de quantification de la proximité au maximum.

Pour les objets A_1, A_2 et A_3 dont les représentations des périodes d'activité sont illustrées dans la Figure 1, Ω est composé de $A_1.fDate, A_2.fDate$ et $A_3.fDate$. La fonction d'appartenance du maximum sur Ω selon le principe d'extension de Zadeh est représentée Figure 3.

FIG. 3 – $\widetilde{max}(A_1.fDate, A_2.fDate, A_3.fDate)$



Les valeurs de l'indice de Kerre pour les représentations des périodes d'activité des objets A_1, A_2 et A_3 et pour $\Omega = \{A_1.fDate, A_2.fDate, A_3.fDate\}$ sont :

$$K_\Omega(A_1.fDate) = 11.1, \quad K_\Omega(A_2.fDate) = 13.1, \quad K_\Omega(A_3.fDate) = 10.7.$$

Ainsi $K_\Omega(A_1.fDate) < K_\Omega(A_2.fDate)$ et $K_\Omega(A_1.fDate) > K_\Omega(A_3.fDate)$. À partir de ces comparaisons, le rangement, selon Kerre, des objets par l'intermédiaire de leurs périodes d'activité est, dans l'ordre croissant des rangs $A2.fDate, A1.fDate, A3.fDate$.

Comme on peut le voir, dans ces méthodes, le rangement dans l'ensemble ne se fait pas par comparaison deux à deux des nombres flous mais par le calcul d'un indice reposant sur la définition de valeurs de référence sur l'ensemble des nombres à comparer. Une des causes à cela est que les méthodes de comparaison deux à deux de nombres flous sont le plus souvent non transitives (Wang et al. (1995)).

Dans l'approche proposée ci-après, l'idée est d'utiliser l'indice d'antériorité *Ant* (voir de Runz et al. (2006, 2008)) plutôt qu'une méthode de comparaison deux à deux donnant une décision binaire. Ainsi, au regard des objets archéologiques, dont les périodes d'activité sont représentées par un nombre flou, la quantification de l'antériorité permet d'étudier le positionnement temporel de chaque objet à l'aide de la construction d'un graphe orienté pondéré (le graphe d'antériorité). Ce graphe permet l'exploration schématique des données selon l'information temporelle.

3 Graphe d'antériorité

L'analyse exploratoire proposée dans ce chapitre se base sur la construction d'un graphe orienté pondéré, appelé graphe d'antériorité, à partir de l'indice d'antériorité *Ant*. Ce dernier quantifie une relation binaire, définie sur les nombres flous et interprétée comme antériorité, entre deux nombres flous F et G quelconques de la manière suivante :

$$Ant(F, G) = \begin{cases} \frac{K_{\{F,G\}}(F)}{K_{\{F,G\}}(F) + K_{\{F,G\}}(G)} & \text{si } K_{\{F,G\}}(F) + K_{\{F,G\}}(G) > 0, \\ 1 & \text{si } K_{\{F,G\}}(F) + K_{\{F,G\}}(G) = 0; \end{cases}$$

avec $\Omega = \{F, G\}$.

L'ensemble des nombres flous à comparer, la relation d'antériorité et les valeurs de l'indice d'antériorité fournissent les sommets, les arcs et les coûts du graphe d'antériorité.

La construction du graphe d'antériorité ayant pour objectif l'observation du positionnement temporel de chaque objet d'une base de données archéologiques par rapport aux autres, les nombres flous mis en relation sont les représentations des périodes d'activité des objets archéologiques. Ainsi on apparie temporellement les objets archéologiques les uns aux autres et on quantifie les paires obtenues à l'aide d'une représentation schématique.

3.1 Construction du graphe

Considérons un ensemble d'éléments E , une relation binaire \mathcal{R} sur E (i.e. un sous-ensemble de $E \times E$) et une application App sur cette relation binaire prenant valeur dans \mathbb{R} .

Un graphe orienté pondéré $G_{\mathcal{R}}(L_S, L_A, L_C)$ est une représentation schématique constituée d'un ensemble L_S de sommets, d'un ensemble L_A d'arcs reliant les sommets deux à deux, et d'un ensemble L_C de coûts associés aux arcs.

Soit deux éléments A et B de E reliés par la relation binaire \mathcal{R} , on associe, dans $G_{\mathcal{R}}(L_S, L_A, L_C)$, à A le sommet S_A et à B le sommet S_B :

$$A \in E \Leftrightarrow S_A \in L_S.$$

L'arc (S_A, S_B) représente alors le fait que ARB :

$$ARB \Leftrightarrow (S_A, S_B) \in L_A.$$

Le coût $C(S_A, S_B)$ d'un arc (S_A, S_B) est égal à $App(A, B)$:

$$(S_a, S_b) \in L_A \Leftrightarrow (C(S_A, S_B) = App(A, B) \text{ et } C(S_A, S_B) \in L_C).$$

L'analyse exploratoire de cette partie se base sur la construction d'un graphe orienté pondéré à partir de l'indice d'antériorité et de l'ensemble Ω d'objets d'une base de données archéologiques dont les périodes d'activité sont définies et représentées par des nombres flous. Ce graphe sera noté $G_{Ant}(L_S, L_A, L_C)$ et sera appelé graphe d'antériorité.

La construction de ce graphe suit la démarche suivante. La relation \mathcal{R} est une relation qui relie deux à deux tous les objets de Ω :

$$\forall A_i, A_j \in \Omega, A_i \mathcal{R} A_j \text{ et } A_j \mathcal{R} A_i.$$

À chaque objet A_i de Ω de période d'activité $A_i.fDate$, nous associons un sommet S_{A_i} du graphe $G_{Ant}(L_S, L_A, L_C)$. Ainsi, le cardinal de L_S est égal à celui de Ω . Soit deux objets A_i et A_j de Ω représentés respectivement par les sommets S_{A_i} et S_{A_j} du $G_{Ant}(L_S, L_A, L_C)$, l'arc (S_{A_i}, S_{A_j}) représente l'antériorité possible de A_i à A_j au regard de leurs périodes d'activité. Le coût $C(S_{A_i}, S_{A_j})$ de l'arc (S_{A_i}, S_{A_j}) est égal à $Ant(A_i.fDate, A_j.fDate)$ et représente la quantification de l'antériorité de A_i à A_j .

Pour l'exemple des objets dont les périodes d'activité sont représentées dans la figure 1, $\Omega = \{A_1, A_2, A_3\}$ et les valeurs de l'indice d'antériorité pour l'ensemble des couples de Ω sont :

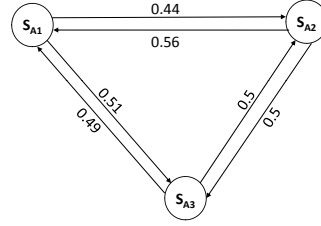
$$\begin{aligned} Ant(A_1.fDate, A_2.fDate) &= 0.44, & Ant(A_2.fDate, A_1.fDate) &= 0.56, \\ Ant(A_1.fDate, A_3.fDate) &= 0.51, & Ant(A_3.fDate, A_1.fDate) &= 0.49, \\ Ant(A_2.fDate, A_3.fDate) &= 0.50, & Ant(A_3.fDate, A_2.fDate) &= 0.50. \end{aligned}$$

Le graphe d'antériorité $G_{Ant}(L_S, L_A, L_C)$ est alors représenté dans la figure 4, dans lequel un sommet S_{A_i} correspond à l'objet A_i ayant pour période d'activité $A_i.fDate$.

Pour chaque sommet du graphe $G_{Ant}(L_S, L_A, L_C)$ ainsi construit, la somme des coûts des arcs sortants, celle des coûts des arcs entrants et leur différence sont des valeurs particulières et ont une signification importante.

3.2 Capacité d'antériorité, capacité de postériorité et positionnement d'un objet

Soient S_{A_i} et S_{A_j} deux sommets de $G_{Ant}(L_S, L_A, L_C)$ correspondant respectivement aux objets A_i et A_j , le coût $C(S_{A_i}, S_{A_j})$ correspond à la valeur de l'indice d'antériorité $Ant(A_i.fDate, A_j.fDate)$. Ainsi, la somme

FIG. 4 – Graphe d'antériorité pour $\Omega = \{A_1, A_2, A_3\}$


des coûts des arcs sortants de S_{A_i} correspond à la capacité d'antériorité $CapAnt_{\Omega}(A_i)$ de A_i relativement à l'ensemble des nombres flous Ω :

$$CapAnt_{\Omega}(A_i) = \sum_{\substack{S_{A_j} \in L_S \\ (S_{A_i}, S_{A_j}) \in L_A}} C(S_{A_i}, S_{A_j}).$$

Pour l'exemple où $\Omega = \{A_1, A_2, A_3\}$, d'après le graphe d'antériorité présenté Figure 4, les capacités d'antériorité des objets de Ω sont les suivants :

$$CapAnt_{\Omega}(A_1) = 0.95, \quad CapAnt_{\Omega}(A_2) = 1.06, \quad CapAnt_{\Omega}(A_3) = 0.99.$$

L'indice $Ant(A_j.fDate, A_i.fDate)$ entre deux nombres flous $A_j.fDate$ et $A_i.fDate$ a été présenté comme l'indice d'antériorité de A_j à A_i . On peut aussi le voir comme l'indice de postériorité de A_i à A_j . Ainsi, la somme des coûts des arcs entrants sur le sommet S_{A_i} de $G_{Ant}(L_S, L_A, L_C)$ associé à l'objet A_i de Ω peut être assimilée à la capacité de postériorité $CapPost_{\Omega}$ de A_j relativement à l'ensemble des nombres flous Ω :

$$CapPost_{\Omega}(A_i) = \sum_{\substack{S_{A_j} \in L_S \\ (S_{A_j}, S_{A_i}) \in L_A}} C(S_{A_j}, S_{A_i}).$$

Pour l'exemple où $\Omega = \{A_1, A_2, A_3\}$, d'après le graphe présenté Figure 4, les capacités de postériorité des objets de Ω sont les suivants :

$$CapPost_{\Omega}(A_1) = 1.05, \quad CapPost_{\Omega}(A_2) = 0.94, \quad CapPost_{\Omega}(A_3) = 1.01.$$

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 sa capacité à être postérieur et sa capacité d'antériorité. C'est pourquoi, l'indice temporel $IndTemp_{\Omega}(A_i)$ de A_i est déterminé comme suit :

$$IndTemp_{\Omega}(A_i) = CapPost_{\Omega}(A_i) - CapAnt_{\Omega}(A_i).$$

Les valeurs de l'indice temporel des objets A_1 , A_2 et A_3 , dont les fonctions d'appartenance des périodes d'activité sont représentées dans la figure 1 et pour le graphe d'antériorité présenté dans la figure 4 (avec $\Omega = A_1, A_2, A_3$), sont :

$$IndTemp_{\Omega}(A_1) = 0.10, \quad IndTemp_{\Omega}(A_2) = -0.12, \quad IndTemp_{\Omega}(A_3) = 0.02.$$

À l'aide de cet indice, nous déterminons la position temporelle $PosTemp(A_i)$ de A_i et $PosTemp(A_j)$ de A_j dans l'ensemble Ω en suivant le principe suivant :

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

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

Pour l'exemple de la figure 1, avec $\Omega = \{A_1, A_2, A_3\}$, d'après le graphe présenté figure 4, les positions temporelles des objets de Ω sont dans l'ordre croissant : A_2, A_3, A_1 . Cet ordre n'est donc pas le même que celui obtenu par l'approche de Kerre ni le même que celui issu de la démarche de Jain. Toutefois, grâce aux valeurs de l'indice d'antériorité, la méthode proposée définit un indice de position temporelle d'un individu vis-à-vis des autres dans un ensemble. On obtient une vision pondérée de la structure temporelle de l'ensemble.

Les valeurs de l'indice temporel et les positions temporelles des objets donnent des informations importantes sur les relations temporelles entre objets.

3.3 Analyse des objets selon le graphe d'antériorité

À l'aide du graphe d'antériorité, on peut extraire trois objets particuliers : l'objet le plus antérieur, l'objet le plus postérieur et l'objet temporellement médian.

L'objet le plus vieux 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 Ω . La position temporelle de l'objet le plus antérieur est la position temporelle minimale des objets de Ω :

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

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 Ω . La position temporelle de l'objet le plus antérieur est la position temporelle maximale des objets de Ω :

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

Pour les objets de la figure 1, l'objet A_2 est le plus antérieur.

Grâce à l'approche par rang (position temporelle), il est trivial de définir l'objet temporellement médian, noté TM . 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 Ω .

Pour les objets de l'exemple, on a :

$$PA = A_2, \quad PP = A_1, \quad TM = A_3.$$

De plus, on peut considérer qu'un objet ayant un indice temporel négatif peut être considéré, 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 ». Ceux ayant un indice d'antériorité nul ont une capacité d'antériorité égal à celle de postériorité. Ils occupent une position intermédiaire spécifique dans l'ensemble : nous les nommerons « anté-postérieurs ». Nous proposons ainsi une classification des objets selon qu'ils soient « plutôt antérieurs », « plutôt postérieurs » ou « anté-postérieurs ».

Dans l'exemple de la figure 1, $\Omega = \{A_1, A_2, A_3\}$, il n'y a pas d'élément anté-postérieur, A_2 est « plutôt antérieur » dans Ω et $\{A_1, A_3\}$ forme l'ensemble des objets « plutôt postérieurs » dans Ω .

La construction du graphe d'antériorité est une approche originale pour le rangement d'objets archéologiques dont les périodes d'activité sont imparfaites et sont représentées par des nombres flous. Ce graphe donne une vision globale des relations chronologiques entre les objets archéologiques. Il donne de nombreuses indications en vue de la classification des objets archéologiques, de l'analyse à l'échelle locale (le chantier de fouille) et globale (la ville) et donc de la généralisation. Dans le cadre du projet SIGRem, les périodes d'activité des objets représentant les tronçons de rues trouvés à Reims et datant de l'époque romaine et stockées dans *BDFRues*, étant représentées par des nombres flous, ces objets sont dans la suite explorés afin d'en dégager leurs positions temporelles.

4 Application aux données archéologiques

Dans la problématique de la valorisation et de la gestion du patrimoine archéologique, la démarche développée par l'Université de Reims Champagne Ardenne, l'Institut National de Recherches Archéologiques Préventives et Ministère de la Culture et de la Communication dans le Centre Interinstitutionnel de Recherches Archéologiques de Reims peut être considérée comme novatrice par l'intégration de la géomatique au cœur de l'analyse urbaine et régionale.

Au-delà de l'élaboration de la cartographie archéologique de la cité des Rèmes³, le projet *SIGRem* Pargny et Piantoni (2005), soutenu par la région Champagne Ardenne, l'état et la ville de Reims, et cadre applicatif de ce travail, porte sur la mise en place d'un Système d'Information Géographique (SIG) pluridisciplinaire Piantoni (2005). Il relève d'une ambition scientifique puisant ses outils conceptuels dans la recherche fondamentale, ses méthodes opérationnelles dans les technologies informatiques en matière d'analyse spatiale et son application pratique dans la mise en valeur des données archéologiques recueillies durant les trentes dernières années.

Dans cette partie, nous proposons d'appliquer le processus exploratoire proposé dans cet article sur la base de données *BDFRues*, partie intégrante du projet SIGRem. Cette base est dédiée aux éléments de rues romaines

³Cité des Rèmes : Reims et ses environs à l'époque romaine

à Reims, *BDFRues*. Les objets, qui y sont stockés, ont chacun une période d'activité représentée par un ensemble flou convexe et normalisé.

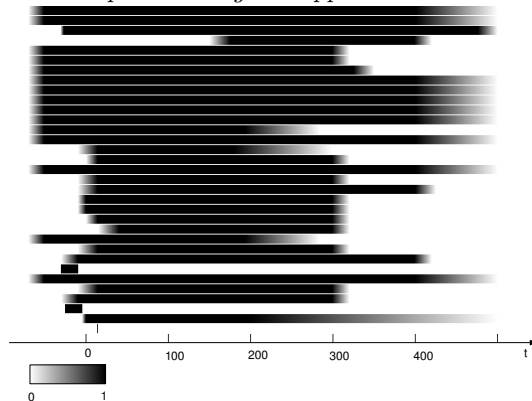
Ainsi dans un premier temps, nous présenterons la base de données *BDFRues*. Ensuite, afin d'obtenir une grille de lecture des résultats plus globale, nous présenterons les rangs des objets obtenus par le classement des représentations de leurs périodes d'activité selon les valeurs des indices de Jain et de Kerre. Enfin, nous exposerons les résultats issus de la construction du graphe d'antériorité.

4.1 A propos de *BDFRues*

Les données archéologiques sont des données spatio-temporelles, ce qui diffère des cas classiques des données géographiques. Quelques études, telles que Dragicevic et Marceau (2000), s'approchent conceptuellement de notre cadre de travail. Dans la base de données sur les rues de Durocortorum⁴, 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.

FIG. 5 – Périodes floues d'activité des objets de *BDFRues* (Chaque ensemble 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 5.

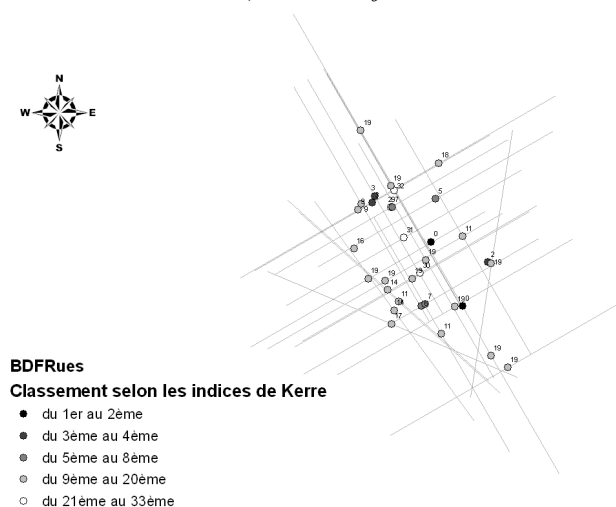
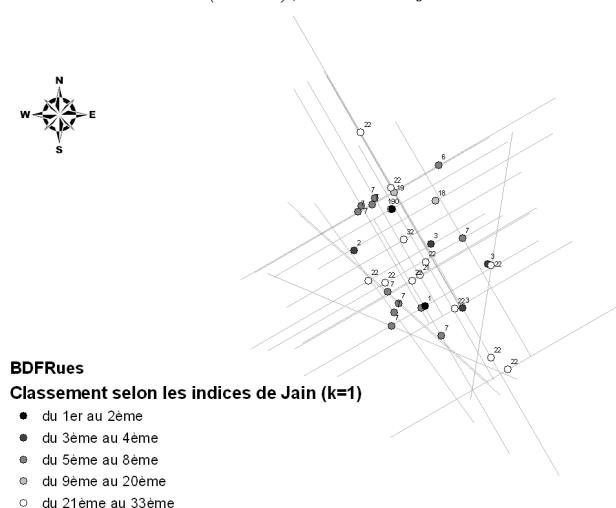
4.2 Rangs des objets selon leurs périodes d'activités pour Kerre et Jain

En utilisant Kerre pour classer les représentations des périodes d'activité des objets de *BDFRues* on obtient la figure 6. Dans cette figure, relativement à l'ensemble des nombres flous associés aux périodes d'activité des objets de *BDFRues*, plus le rang d'un objet est grand plus le nombre flou associé à la période d'activité de l'objet est grand selon Kerre (1982).

En utilisant l'indice de Jain avec $k = 1$ pour classer les nombres flous associés aux périodes d'activité des objets de *BDFRues* on obtient la figure 7. Dans cette figure plus le rang d'un objet est grand plus le nombre flou associé à sa période d'activité est grand selon le classement des indices de Jain (1977).

On peut remarquer que certains rangs attribués aux objets selon l'indice de Jain diffèrent de ceux attribués selon l'indice de Kerre. L'interprétation de ces rangs est difficile, car il n'y a pas de sémantique associée à ces indices. A contrario, les positions temporelles obtenues à l'aide de la construction du graphe d'antériorité ont une forte interprétabilité qui sera illustrée dans la suite.

FIG. 6 – Rangs, déterminés à l'aide de Kerre, des 33 objets de BDFRues selon leurs périodes d'activité

FIG. 7 – Rangs, déterminés à l'aide de Jain ($k = 1$), des 33 objets de BDFRues selon leurs périodes d'activité

4.3 Positionnement temporel des objets selon le graphe d'antériorité

En construisant le graphe d'antériorité sur les données de *BDFRues*, on obtient la figure 8. Dans cette figure, plus le rang d'un objet est grand, plus la position temporelle est élevée, c'est-à-dire plus sa capacité de postériorité à tous les autres est grande.

On peut remarquer que les positions temporelles obtenues à l'aide du graphe d'antériorité, et les rangs obtenus à l'aide des approches de Kerre ou de Jain présentent dans certains cas des divergences. Cependant, les positions temporelles ont une interprétabilité plus forte que les rangs issus de Kerre ou de Jain, ce qui permet de définir une classification temporelle des objets selon leur positionnement temporel.

La position temporelle d'un objet archéologique est obtenue à l'aide de l'indice temporel de l'objet lui-même issu des capacités d'antériorité et de postériorité. Ces capacités de l'objet sont calculées à l'aide de l'indice d'antériorité de l'objet à chacun des autres objets archéologiques. Cet indice quantifie l'antériorité d'un objet à un autre. Ainsi, la position temporelle d'un objet reflète la relation temporelle de l'objet à l'ensemble des autres.

Par exemple, de ces positions temporelles, on peut extraire les trois objets particuliers suivants : le plus antérieur, le plus postérieur et le médian. Ces objets particuliers sont identifiables dans la figure 9.

⁴Durocortorum : Reims à l'époque romaine

FIG. 8 – Positions temporelles des 33 objets de *BDFRues* selon leurs périodes d'activité

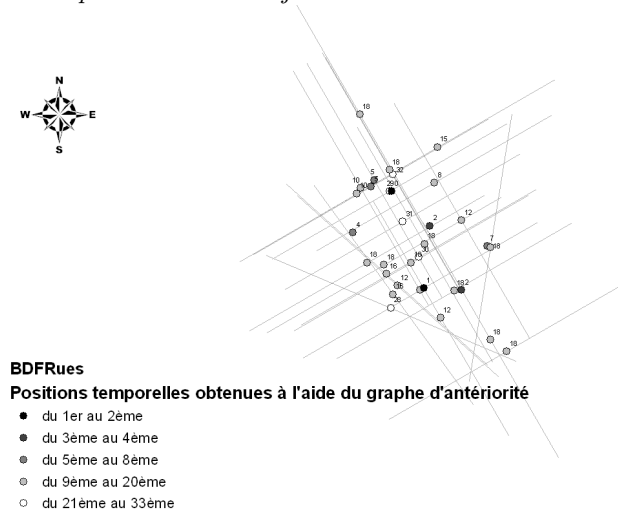
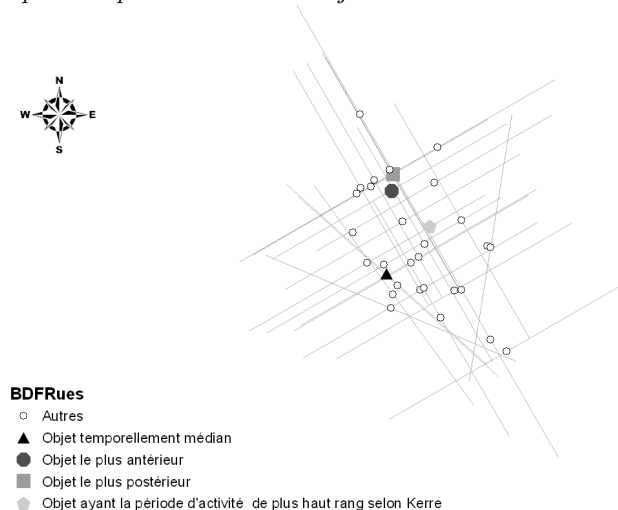


FIG. 9 – Positions temporelles particulières des objets de *BDFRues* selon leurs périodes d'activité



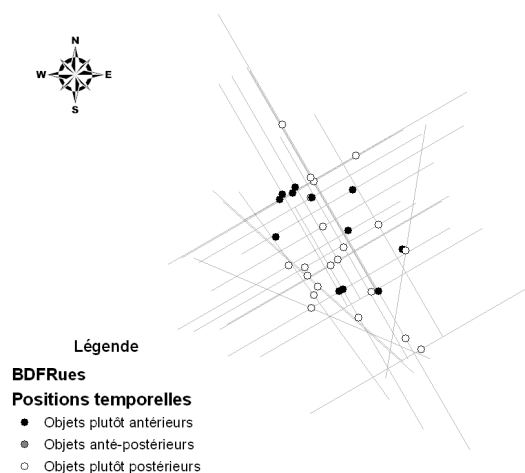
La figure 10 présente les objets de *BDFRues* selon les trois classes : « plutôt postérieurs », « plutôt antérieurs » et « anté-postérieurs ».

Pour les données de *BDFRues*, la classe des objets « plutôt postérieurs » regroupe deux fois plus d'objets que celle des objets « plutôt antérieurs ». La classe des objets « anté-postérieurs » est vide.

On peut 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 . Une définition plus vague de la notion *anté-postérieure*

FIG. 10 – Objets de *BDFRues* classés en « plutôt postérieurs », « plutôt antérieurs » et « anté-postérieurs »

aurait sûrement permis de détecter ces éléments. Mais cette nouvelle définition demande un paramétrage fortement dépendant des données. Ce paramétrage n'entre pas dans la démarche exploratoire proposée dans cette partie et donc dans cet article.

De plus, par le recouplement de la figure 10 avec la figure 9, on remarque que l'objet temporellement médian appartient aux objets « plutôt postérieurs » à l'instar de l'objet le plus postérieur. L'objet le plus antérieur appartient à la classe des « plutôt antérieurs ».

Ainsi, les valeurs de l'indice temporel des objets de *BDFRues* obtenues à partir du graphe d'antériorité permettent de classer les objets archéologiques selon l'antériorité aux objets de la base. Ces positions permettent de définir des classes d'objets et d'extraire des objets particuliers. Ce graphe permet donc, par une organisation schématique structurée, de dégager de nouvelles connaissances temporelles sur les informations contenues dans *BDFRues*. Ces nouvelles connaissances sont riches d'information pour l'expertise archéologique.

5 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, on construit un graphe orienté pondéré à partir des objets de la base et des valeurs de l'indice d'antériorité les reliant deux à deux. De ce graphe, nous dégageons l'indice temporel de chaque objet afin d'en déterminer sa position temporelle dans l'ensemble et l'analyse associée (« plutôt antérieur », « plutôt postérieur », « anté-postérieur »). Les positions, relativement au temps, des objets archéologiques dans une base de données facilitent la compréhension des relations liant les objets de la base dans le SIG.

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 le graphe d'antériorité.

En perspective, nous avons l'intention d'étudier différents modes de visualisation 3D des informations portées à la fois par le graphe d'antériorité (positions temporelles des objets) et par les localisations 2D des objets afin de pouvoir les croiser pour l'analyse spatiotemporelle des données. De plus, nous souhaitons détecter grâce à une réduction du graphe les possibles incohérences temporelles dues au changement d'échelle : les périodes d'activité sont estimées à l'échelle du site, le processus exploratoire se fait à l'échelle de la ville.

Remerciements

Nous tenons à remercier le Service Régional d'Archéologie de Champagne-Ardenne et le centre rémois de l'Institut National de Recherche en Archéologie Préventive pour nous avoir permis d'accéder à leurs données.

Nous tenons de même à souligner la contribution de Dominique Pargny, ingénieur d'études au laboratoire GE-GENA, de Frédéric Piantoni, Maître de Conférences au laboratoire HABITER, et de Michel Herbin, Professeur au CReSTIC, au projet SIGRem, porté par l'Université de Reims Champagne-Ardenne, dans le contexte duquel c'est fait ce travail.

Références

- Chen, S.-H. (1985). Ranking fuzzy numbers with maximizing set and minimizing set. *Fuzzy Sets and Systems* 17, 113–129.
- de Runz, C., F. Blanchard, E. Desjardin, et M. Herbin (2008). Fouilles archéologiques : à la recherche d'éléments représentatifs. In *Atelier Fouilles de Données Complexes - Conférence Extraction et Gestion des Connaissances - EGC'08*, Sophia Antipolis, France, pp. 95–103.
- de Runz, C., E. Desjardin, M. Herbin, et F. Piantoni (2006). A new Method for the Comparison of two fuzzy numbers extending Fuzzy Max Order. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems - IPMU'06*, Paris, France, pp. 127–133. Editions EDK.
- de Runz, C., E. Desjardin, F. Piantoni, et M. Herbin (2007). Using fuzzy logic to manage uncertain multi-modal data in an archaeological GIS. In *International Symposium on Spatial Data Quality - ISSDQ'07*, Enschede, Pays-Bas.
- de Runz, C., E. Desjardin, F. Piantoni, et M. Herbin (2008). Anteriority index for managing fuzzy dates in archaeological GIS. *Soft Computing*. Soumis le 30-10-2007, en révision majeure depuis le 09-06-2008.
- Dragicevic, S. et D. J. Marceau (2000). An application of fuzzy logic reasoning for GIS temporal modeling of dynamic processes. *Fuzzy Sets and Systems* 113, 69–80.
- 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 et E. Sanchez (Eds.), *Approximate Reasoning in Decision-Analysis*, pp. 277–282. North-Holland Publishing Company.
- Pargny, D. et F. Piantoni (2005). SIGRem : un Système d'Information Géographique pour l'archéologie en Champagne-Ardenne. In *colloque Archéologie en Champagne-Ardenne*. INRAP.
- Piantoni, F. (2005). Le SIGRem. Problématique et méthodologie. In *séminaire de recherche du CIRAR*.
- Wang, X. et E. E. Kerre (2001a). Reasonable properties for the ordering of fuzzy quantities (I). *Fuzzy Sets and Systems* 118, 375–385.
- Wang, X. et E. E. Kerre (2001b). Reasonable properties for the ordering of fuzzy quantities (II). *Fuzzy Sets and Systems* 118, 387–405.
- Wang, X., E. E. Kerre, B. Cappelle, et D. Ruan (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.

Summary

In this paper, we propose a new temporal data mining method considering a set of archaeological objects which are temporally represented with fuzzy numbers. Our method uses an index which quantifies the anteriority between two fuzzy numbers for the construction of a weighted oriented graph. The vertices of the graph correspond to the archaeological objects. Using this anteriority graph, we determine the potential of anteriority, of posteriority and the temporal position of each object. The information obtained designs the temporal relations between objects and allows us to extract the most anterior object and the most posterior object. We apply this approach to the roman streets discovered in Reims.