# Wave Top-k Random-d Family Search : comment guider un expert dans un espace structuré

Etienne Lehembre*, Bruno Cremilleux*, Bertrand Cuissart*, Abdelkader Ouali*
Albrecht Zimmermann*

* UNICAEN, ENSICAEN, CNRS - UMR GREYC, Normandie Univ 14000 Caen, France
{prenom.nom}@unicaen.fr

**Résumé.** Dans cet article, nous développons une méthode (WTRFS) incluant le retour utilisateur dans le but de le guider parmi les résultats d'une fouille de motifs. Ce travail vise à remplacer l'étape de déclaration des descripteurs utilisée dans la fouille interactive de motifs. Pour cela, la méthode s'appuie sur l'existence hypothétique d'un lien entre les différents motifs intéressants un expert. Nous montrons empiriquement que WTRFS renvoie rapidement les résultats les plus pertinents pour l'utilisateur. De plus, même si les retours de l'utilisateur sont imparfaits, le comportement de WTRFS n'en est pas altéré.

## 1 Introduction

Le but de la fouille de données est d'aider les experts de domaines applicatifs (ils ou elles) à analyser leurs données en leur montrant des associations d'intérêt. Lorsque ces résultats sont fournis sous la forme d'un ensemble de motifs saillants, un problème récurrent est la grande quantité de solutions fournies, souvent impossible à appréhender par un humain. Différentes approches traitent ce problème comme les représentations condensées de motifs qui synthétisent *l'espace des solutions* (Pasquier et al., 1999), les nombreuses *mesures de qualité* (Tan et al., 2004) et, plus récemment, les techniques de *fouille d'ensembles de motifs* (De Raedt et Zimmermann, 2007). Cependant, la combinaison de ces résultats reste insuffisante à rendre l'espace des solutions humainement abordable. Aussi, une proposition est d'intégrer l'expert au processus via une fouille qualifiée *d'interactive*.

Alors que plusieurs méthodes de fouille interactive de motifs traitent les données sous forme d'itemsets (Boley et al., 2013; Van Leeuwen, 2014), peu de travaux portent sur la recherche interactive de motifs à partir de données structurées, comme la fouille interactive de sous-graphes (Bhuiyan et Hasan, 2016; Bhuiyan et Al Hasan, 2016). De plus, même dans ces travaux, un sous-graphe est traité comme un itemset et les relations entre motifs sont peu exploitées. Les algorithmes considèrent les motifs comme un ensemble, sans exploiter la taille des sous-graphes pour induire leur degré de spécificité. Bien que certains travaux (van Leeuwen et al., 2016) travaillent à retranscrire l'intérêt subjectif dans la distribution de l'échantillonnage. Cette dernière est généralement impactée globalement et non localement. Pourtant, l'expert est sensible à ces paramètres locaux et son intérêt peut diverger lorsqu'il étudie deux régions distinctes de l'espace des solutions.