

# Découverte de biclusters avec présence ou absence de propriétés

Abdélilah Balamane\*, Rokia Missaoui\*  
Léonard Kwuida\*\*, Jean Vaillancourt\*\*\*

\*LARIM, Université du Québec en Outaouais, Québec, Canada  
bala04@uqo.ca, rokia.missaoui@uqo.ca

\*\*Haute École Spécialisée Bernoise (BFH), Suisse  
leonard.kwuida@bfh.ch

\*\*\*HEC Montréal, Québec, Canada  
jean.vaillancourt@hec.ca

**Résumé.** La plupart des algorithmes de biclustering existants tiennent compte uniquement de la présence de propriétés que possède un ensemble d'objets. Cependant, il pourrait être fort utile dans plusieurs domaines d'application tels que le crime organisé, la génétique ou le marketing numérique d'identifier des groupes homogènes d'objets présentant des similarités tant au niveau de la présence que de l'absence d'attributs. Dans cet article, nous présentons une méthode générique de biclustering qui exploite une matrice binaire pour produire au moins trois types de biclusters : (i) ceux où toutes les valeurs sont égales à 1, (ii) ceux où toutes les valeurs sont égales à 0, et (iii) ceux indiquant la présence de certains attributs et/ou l'absence d'autres attributs sans nécessité de tenir compte du complémentaire du contexte (matrice) binaire de départ. L'implémentation et la validation de la méthode sur des collections de données permettent d'illustrer son potentiel de découverte de motifs pertinents.

## 1 Introduction

La fouille et la gestion de motifs font référence à un ensemble d'activités d'extraction, de stockage et de manipulation de motifs à partir des données. Un motif (*pattern*) est un résultat d'un processus de fouille de données exprimé sous forme de connaissance. En analyse formelle de concepts, le motif prend deux principales formes : a) des concepts formels décrivant des objets/individus avec leurs attributs communs et représentant des nœuds d'un treillis de concepts (Galois), et b) des règles d'association, y compris des implications, entre des groupes d'attributs.

Dans le présent article, nous proposons une nouvelle procédure de biclustering (classification croisée ou bipartitionnement) générique appelée BiP (*Biclustering Procedure*) qui calcule des biclusters (blocs) sémantiquement significatifs à partir d'un graphe biparti issu d'une matrice d'adjacence à valeurs binaires. Différents types de biclusters peuvent alors être obtenus avec un contenu comportant : (i) des cellules remplies uniquement de 1 (type 1), (ii) des cellules remplies uniquement de 0 (type 2), et (iii) des colonnes complètement pleines de 1 et/ou