Multimodal Clustering for Community Detection

Dmitry I. Ignatov, Alexander Semenov, Daria Komissarova, and Dmitry V. Gnatyshak

1 Introduction

Online social networking services generate massive amounts of data, which can become a valuable source for guiding Internet advertisement efforts or provide sociological insights. Each registered user has a network of friends as well as specific profile features. These profile features describe the user's tastes, preferences, the groups he or she belongs to, etc. Social Network Analysis (SNA) is a popular research field in which methods are developed for analysing one-mode networks, like friend-to-friend,[1] two-mode or affiliation networks [57, 60, 69], three-mode [10, 20, 38, 46, 66], and even multi-mode dynamic networks [75, 76, 81, 89]. By multi-mode networks we mean namely such networks where actors can be related with other types of entities by edges like those between users and their interests in two-mode case or by hyperedges like those relating users, tags, and resources in three-mode case; sometimes such networks are called heterogeneous since different types of nodes are involved [48]. We focus on the subfield of bicommunity identification and its higher order extensions. Thus, in particular, we present tri- and tetracommunities examples extracted from real data. For one-mode case a reader may refer to an extensive survey on community detection [21].

---

[1]www.https://en.wikipedia.org/wiki/Friend-to-friend.

D.I. Ignatov (✉) • D. Komissarova • D.V. Gnatyshak
National Research University Higher School of Economics, Moscow, Russia
e-mail: dignatov@hse.ru; komissarovadaria93@gmail.com; dgnatyshak@hse.ru

A. Semenov
National Research University Higher School of Economics, Moscow, Russia

Mobile TeleSystems PJSC, Moscow, Russia
e-mail: SemenoffAlex@gmail.com

The notion of community in SNA and Complex Networks is closely related to the notion of cluster in Data Analysis [3, 21]. There is the main issue in both disciplines: what is a common definition of community and what is a common definition of cluster? On the one hand, it is clear that actors from the same community should be similar as well as objects in one cluster; on the other hand, these actors (or objects) should be less similar to actors (or objects) from another community (or cluster). This general idea allows a variety of definitions suitable for concrete purposes in both domains [3, 21, 63].

There is a large amount of network data that can be represented as bipartite or tripartite graphs. Standard techniques for community detection in two-mode networks like "maximal bicliques search" return a huge number of patterns (in the worst case exponential w.r.t. the input size) [56, 77]. Moreover, not all members of such bicommunities should be related to the same items, for example, exactly the same vocabulary used by each member in case of epistemic communities. Therefore we need some relaxation of the biclique notion as well as appropriate interestingness measures and constraints for mining and filtering such "relaxed" biclique communities.

Applied lattice theory provides us with the definition of formal concept [27], which is closely related to maximal biclique in a bipartite graph; formal concepts and concept lattices (or Galois lattices) are widely known in the social network analysis community (see, e.g., [19, 23, 24, 65, 77, 86]). However, these methods are overly rigid for analysing large amounts of data resulting in a huge number of concepts even if their computation is feasible.

A concept-based bicluster (or object-attribute bicluster) [37] is a scalable approximation of a formal concept (maximal biclique in a bipartite graph). The advantages of concept-based biclustering are

1. Less number of patterns to analyse (no more than the number of edges in the original network);
2. Less computational time (polynomial vs exponential);
3. Tolerance to missing (object, attribute) pairs;
4. Filtering of biclusters (communities) by density threshold.

In general, the method of biclustering dates back to the seminal work of Hartigan on the so-called direct clustering [31], where clusters of objects may appear sharing only a subset of attributes. The term biclustering was introduced later in the book of Mirkin [63]:

> The term biclustering refers to simultaneous clustering of both row and column sets in a data matrix. Biclustering addresses the problems of aggregate representation of the basic features of interrelation between rows and columns as expressed in the data.

Following this terminology, formal concepts can be considered as maximal inclusion biclusters of constant values in binary data [49], whereas their relaxations tolerant to missing object-attribute pairs can be called object-attribute biclusters [37, 39].

There are several successful attempts to mine two-mode [51, 78], three-mode [46], and even four-mode communities [47] by means of Formal Concept Analysis.

For analysing three-mode network data like folksonomies [83] we have also proposed a scalable triclustering technique [40, 45].

These studies for higher-mode cases were enabled by the previous introduction of the so-called triconcepts by Lehman and Wille [58, 87]; a formal triconcept consists of three components: extent (objects), intent (attributes), and modus (conditions under which an object has an attribute). It is a matter of curiosity, but such triconcepts had been used for analysing triadic data in social cognition studies [52] before their formal introduction. Later, a polyadic (or multimodal) extension of FCA was introduced in [85].

Previously, we have introduced a pseudo-triclustering technique for tagging groups of users by their common interests [28]. This approach differs from traditional triclustering methods because it relies on the extraction of biclusters from two separate object-attribute tables and rather belongs to methods for analysing multi-relational networks. Here we investigate applicability of biclustering and triclustering (as well as *n*-clustering, its higher-mode extension) to community detection in two-, three-, and higher-mode networks directly.

The remainder of the paper is organised as follows. In Sect. 2, we introduce basic notions of Formal Concept Analysis. Section 3 describes object-attribute biclustering and its direct generalisations to higher dimensions. Section 4 briefly discusses a variety of quality measures used in clustering, FCA, and SNA domains and their interrelation with multimodal clustering. In Sect. 5, we describe datasets which we have chosen to illustrate the performance of the approach. We present the results obtained during experiments on these datasets in Sect. 6. Related work is discussed in Sect. 7, while Sect. 8 concludes our paper and describes some interesting directions for future research.

2 Basic Definitions

2.1 Formal Concept Analysis

A formal context in FCA [27] is a triple $\mathbb{K} = (G, M, I)$, where G is a *set of objects*, M is a *set of attributes*, and the relation $I \subseteq G \times M$ shows which object possesses which attribute. For any $A \subseteq G$ and $B \subseteq M$ one can define *Galois operators*:

$$A' = \{m \in M \mid gIm \text{ for all } g \in A\},$$
$$B' = \{g \in G \mid gIm \text{ for all } m \in B\}. \tag{1}$$

The operator $''$ (applying the operator $'$ twice) is a *closure operator*: it is idempotent ($A'''' = A''$), monotone ($A \subseteq B$ implies $A'' \subseteq B''$), and extensive ($A \subseteq A''$). The set of objects $A \subseteq G$ such that $A'' = A$ is called closed. Similar properties are valid for closed attribute sets, subsets of a set M. A pair (A, B) such that $A \subseteq G, B \subseteq M, A' = B$, and $B' = A$, is called a *formal concept* of a context $\mathbb{K}$. The sets A and B are closed and called *extent* and *intent* of a formal concept

(A, B), respectively. For the set of objects A the set of their common attributes A' describes the similarity of objects of the set A, and the closed set A'' is a cluster of similar objects (with the set of common attributes A'). The relation "to be a more general concept" is defined as follows: $(A, B) \geq (C, D)$ iff $A \supseteq C$. The concepts of a formal context $\mathbb{K} = (G, M, I)$ ordered by extensions inclusion form a lattice, which is called *concept lattice*. For its visualisation *line diagrams* (Hasse diagrams) can be used, i.e. the cover graph of the relation "to be a more general concept". In the worst case (Boolean lattice) the number of concepts is equal to $2^{\{\min |G|, |M|\}}$, thus, for large contexts, to make application of FCA machinery tractable the data should be sparse. Moreover, one can use different ways of filtering formal concepts (for example, choosing concepts by their stability index or extent size).

Let us consider a formal context $\mathbb{K}$ that consists of four objects, persons (Alex, Mike, Kate, David), four attributes, books (Romeo and Juliet by William Shakespeare, The Puppet Masters by Robert A. Heinlein, Ubik by Philip K. Dick, and Ivanhoe by Walter Scott), and incidence relation showing which person which book read or liked.

| $\mathbb{K}$	*Romeo and Juliet*	*The Puppets Masters*	*Ubik*	*Ivanhoe*
Kate	×			×
Mike	×		×	
Alex		×	×	
David		×	×	×

There are nine concepts there. For example,
$C_1 = (\{Kate, Mike\}, \{Romeo\ and\ Juliet\})$
$C_2 = (\{Alex, David\}, \{The\ Puppet\ Masters, Ubik\})$
$C_3 = (\{Kate, David\}, \{Ivanhoe\})$.
Note that the pair of sets $(A, B) = (\{Alex, David\}, \{Ubik\})$ does not form a formal concept since we can enlarge its extent by one more object Mike to fulfill $(A \cup \{Mike\})' = B$ and $B' = A \cup \{Mike\}$. So, $C_4 = (\{MIke, Alex, David\}, \{Ubik\})$ is a formal concept. The corresponding bipartite graph is shown in Fig. 1 along with the biclique formed by elements of concept C_2.

Fig. 1 Two-mode network of
readers and its community of
Sci-Fi readers (*shaded*)
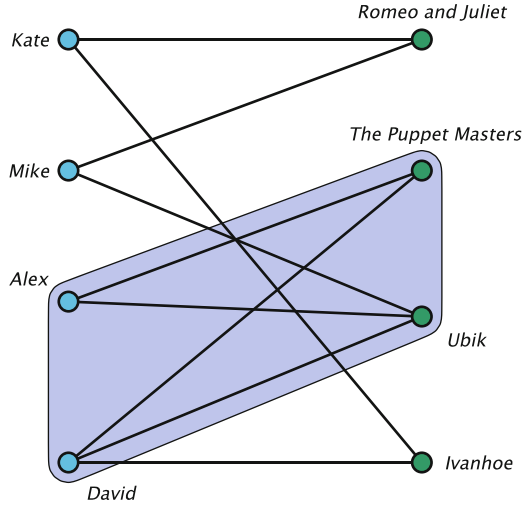
From SNA viewpoint, if we assume that an OA-bicluster (*event'*, *actor'*) is a found community, we are looking for a pair (*actor*, *event*) in an input network, where this actor participated in all of the events typical for the community, while the chosen event is typical for all the members of that community.

3 Higher-Order Extensions of FCA and Multimodal Clustering

3.1 Triadic and Polyadic FCA

For convenience, a *triadic context* is denoted by (X_1, X_2, X_3, Y). A triadic context $\mathbb{K} = (X_1, X_2, X_3, Y)$ gives rise to the following dyadic contexts:

$$\mathbb{K}^{(1)} = (X_1, X_2 \times X_3, Y^{(1)}), \quad \mathbb{K}^{(2)} = (X_2, X_1 \times X_3, Y^{(2)}), \quad \mathbb{K}^{(3)} = (X_3, X_1 \times X_2, Y^{(3)}),$$

where $g Y^{(1)}(m, b) :\Leftrightarrow m Y^{(2)}(g, b) :\Leftrightarrow b Y^{(3)}(g, m) :\Leftrightarrow (g, m, b) \in Y$. The *derivation operators* (primes or concept-forming operators) induced by $\mathbb{K}^{(i)}$ are denoted by $(.)^{(i)}$. For each induced dyadic context we have two kinds of such derivation operators. That is, for $\{i, j, k\} = \{1, 2, 3\}$ with $j < k$ and for $Z \subseteq X_i$ and $W \subseteq X_j \times X_k$, the (*i*)-derivation operators are defined by:

$$Z \mapsto Z^{(i)} = \{(x_j, x_k) \in X_j \times X_k | x_i, x_j, x_k \text{ are related by } Y \text{ for all } x_i \in Z\},$$

$$W \mapsto W^{(i)} = \{x_i \in X_i | x_i, x_j, x_k \text{ are related by } Y \text{ for all } (x_j, x_k) \in W\}.$$

Formally, a *triadic concept* of a triadic context $\mathbb{K} = (X_1, X_2, X_3, Y)$ is a triple (A_1, A_2, A_3) of $A_1 \subseteq X_1, A_2 \subseteq X_2, A_3 \subseteq X_3$, such that for every $\{i, j, k\} = \{1, 2, 3\}$ with $j < k$ we have $(A_j \times A_k)^{(i)} = A_i$. For a certain triadic concept (A_1, A_2, A_3), the components A_1, A_2, and A_3 are called the *extent*, the *intent*, and the *modus* of (A_1, A_2, A_3). Since a tricontext $\mathbb{K} = (X_1, X_2, X_3, Y)$ can be interpreted as a three-dimensional cross table, according to our definition, under suitable permutations of rows, columns, and layers of this cross table, the triadic concept (A_1, A_2, A_3) is interpreted as a maximal cuboid full of crosses. The set of all triadic concepts of $\mathbb{K} = (X_1, X_2, X_3, Y)$ is denoted by $\mathfrak{T}(X_1, X_2, X_3, Y)$.

To avoid additional technical description of *n*-ary concept-forming operators, we introduce *n*-adic formal concepts without their usage. The *n*-adic concepts of an *n*-adic context $(X_1, \ldots, X_n, Y)$ are exactly the maximal *n*-tuples $(A_1, \ldots, A_n)$ in $2^{X_1} \times \cdots \times 2^{X_n}$ with $A_1 \times \cdots \times A_n \subseteq Y$ with respect to component-wise set inclusion [85]. The notion of *n*-adic concept lattice can be introduced in the similar way to the triadic case [85]. For mining *n*-adic formal concepts one can use DATA-PEELER algorithm described in [12].

3.2   Biclustering

An alternative approach to define patterns in formal contexts can be realised via a relaxation of the definition of formal concept as a maximal rectangle full of crosses w.r.t. the input incidence relation. One of such relaxations is the notion of an object-attribute bicluster [37]. If $(g, m) \in I$, then (m', g') is called an *object-attribute bicluster*[2] (OA-bicluster or simply bicluster if there is no collision) with the density $\rho(m', g') = |I \cap (m' \times g')| / (|m'| \cdot |g'|)$.

The main features of OA-biclusters are listed below:

1. For any bicluster $(m', g') \subseteq 2^G \times 2^M$ it follows that $\frac{|m'| + |g'| - 1}{|g'||m'|} \leq \rho(A, B) \leq 1$.
2. OA-bicluster (m', g') is a formal concept iff $\rho = 1$.
3. If (m', g') is a bicluster, then $(g'', g') \leq (m', m'')$.

Let $(A, B) \subseteq 2^G \times 2^M$ be a bicluster and $\rho_{\min}$ be a non-negative real number such that $0 \leq \rho_{\min} \leq 1$, then (A, B) is called *dense*, if it fits the constraint $\rho(A, B) \geq \rho_{\min}$. The above-mentioned properties show that OA-biclusters differ from formal concepts by the fact that they do not necessarily have unit density. Graphically it means that not all the cells of a bicluster must be filled by crosses (see Fig. 2). The rectangle in Fig. 2 depicts a bicluster extracted from an object-attribute table. The horizontal grey line corresponds to object g and contains only non-empty cells. The vertical grey line corresponds to attribute m and also contains only non-empty cells. By applying the Galois operator, as explained in Sect. 2.1, one time to g we obtain all its attributes g'. By applying Galois operator $'$ twice to g we obtain all objects that

---

[2]We omit curly brackets here it what follows implying that $\{g\}' = g'$ and $\{m\}' = m'$.

Fig. 2 OA-bicluster
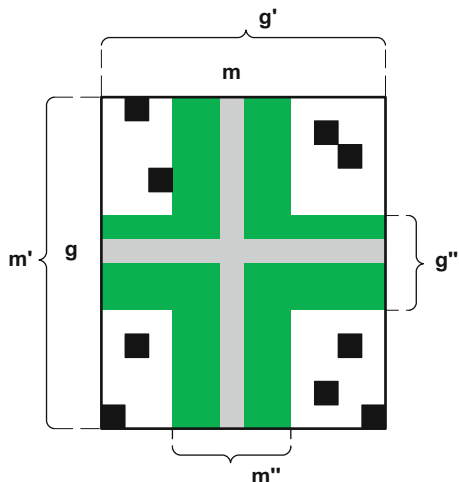
Algorithm 1: Add procedure for the online algorithm for OA-biclustering

---

   Input: I is an input set of object-attribute pairs;
      $\mathscr{B} = \{B = (*X, *Y)\}$ is a current set of OA-biclusters;
      PrimesOA, *PrimesAO*;
   Output: $\mathscr{B} = \{T = (*X, *Y)\}$;
      PrimesOA, *PrimesAO*;
  1: **for all** $(g, m) \in I$ **do**
  2:     *PrimesOA*[g] := *PrimesOA*[g] $\cup$ m
  3:     *PrimesAO*[m] := *PrimesAO*[m] $\cup$ g
  4:     $\mathscr{B}$:= $\mathscr{B}$ $\cup$ (&*PrimesAO*[m], &*PrimesOA*[g])
  5: **end for**

---

have the same attributes as g. This is depicted in Fig. 2 as g''. By applying Galois operator $'$ twice to m we obtain all attributes that belong to the same objects as m. This is depicted in Fig. 2 as m''. The white spaces indicate empty cells. The filled black boxes indicate non-empty cells. Whereas a traditional formal concept would cover only the green and grey area, the bicluster also covers the white and black cells. This gives to OA-biclusters fault-tolerance properties (see Proposition 1).

To generate biclusters fulfilling a minimal density requirement we can perform computations in two phases. The online phase, Add procedure (see Algorithm 1), allows to process pairs from incidence relation I and generate biclusters in one pass by means of pointer and reference variables for access to primes of objects and attributes even without knowing the number of objects and attributes in advance; see the version of this online algorithm for triadic case in [29]. Thus, the generation of all biclusters is realised within $O(|I|)$. Note that the algorithm can start with a non-empty collection of biclusters obtained previously. Then all biclusters can be enumerated in a sequential manner and only those fulfilling the minimal density constraint are retained.

For the context shown in Fig. 1 one can find two concepts,

$C_2 = (\{Alex, David\}, \{The\ Puppet\ Masters, Ubik\})$ and

$C_4 = (\{Alex, Mike, David\}, \{Ubik\})$, and one bicluster,

$B_1 = (Ubik', David') = (\{Alex, Mike, David\}, \{The\ Puppet\ Masters, Ubik\})$,

with density $\rho = 5/6 \approx 0.83$.

These two concepts can be interpreted as Sci-Fi readers and cyber punk readers (or P.K. Dick's readers at least), respectively. However, bicluster B_1 by allowing one missing pair (*Mike, The Puppet Masters*) can be considered as a community of Sci-Fi readers as well, which is larger than C_2.

3.3   OAC-Triclustering and Prime-Based n-Clustering

Guided by the idea of finding scalable and noise-tolerant alternatives to triconcepts, we have had a look at triclustering paradigm in general for a triadic binary data, i.e. for tricontexts as input datasets.

Definition 1 Suppose $\mathbb{K} = (G, M, B, I)$ is a triadic context and $Z \subseteq G, Y \subseteq M, Z \subseteq B$. A triple $T = (X, Y, Z)$ is called an *OAC-tricluster*. Traditionally, its components are called *extent, intent, and modus*, respectively.

The *density* of a tricluster $T = (X, Y, Z)$ is defined as the fraction of all triples of I in $X \times Y \times Z$:

$$\rho(T) = \frac{|I \cap (X \times Y \times Z)|}{|X||Y||Z|} \tag{2}$$

Definition 2 A tricluster T is called *dense* iff its density is not less than some predefined threshold, i.e. $\rho(T) \geq \rho_{\min}$.

The collection of all triclusters for a given tricontext $\mathbb{K}$ is denoted by $\mathscr{T}$.

Since we deal with all possible cuboids in Cartesian product $G \times M \times B$, it is evident that the number of all OAC-triclusters, $|\mathscr{T}|$, is equal to $2^{|G|+|M|+|B|}$. However, not all of them are supposed to be dense, especially for real data which are frequently quite sparse. Below we discuss one of the possible OAC-tricluster definitions, which gives us an efficient way to find, within polynomial time, a number of (dense) triclusters not greater than the number of triples in the initial data, $|I|$.

Here, let us define the prime operators and describe *prime OAC-triclustering*, which extends the biclustering method from [39] to the triadic case.

Derivation (prime) operators for elements of a triple $(\widetilde{g}, \widetilde{m}, \widetilde{b}) \in I$ from a triadic context $\mathbb{K}$ can be defined as follows:

$$\widetilde{g}' := \{ (m, b) \mid (\widetilde{g}, m, b) \in I \} \tag{3}$$

$$\widetilde{m}' := \{ (g, b) \mid (g, \widetilde{m}, b) \in I \} \tag{4}$$

$$\widetilde{b}' := \{ (g, m) \mid (g, m, \widetilde{b}) \in I \} \tag{5}$$

$(\widetilde{g}, \widetilde{m})'$, $(\widetilde{g}, \widetilde{b})'$, $(\widetilde{m}, \widetilde{b})'$ prime operators can be defined in the same way.

$$(\widetilde{g}, \widetilde{m})' := \{ b \mid (\widetilde{g}, \widetilde{m}, b) \in I \} \tag{6}$$

$$(\widetilde{g}, \widetilde{b})' := \{ m \mid (\widetilde{g}, m, \widetilde{b}) \in I \} \tag{7}$$

$$(\widetilde{m}, \widetilde{b})' := \{ g \mid (g, \widetilde{m}, \widetilde{b}) \in I \} \tag{8}$$

The following definition uses only prime operators [Eqs. (6)–(8)] to generate triclusters, however, other variants are possible. Thus, in [45], *OAC-triclusters based on box operator* have been studied; this type of tricluster relies on Eqs. (3)–(5).

Definition 3 Suppose $\mathbb{K} = (G, M, B, I)$ is a triadic context. For a triple $(g, m, b) \in I$ a triple $T = ((m, b)', (g, b)', (g, m)')$ is called a *prime operator based OAC-tricluster*. Its components are called *extent, intent, and modus*, respectively.

Prime-based OAC-triclusters are more dense than the ones based on box operator. Their structure is illustrated in Fig. 3: every element corresponding to the "grey" cell is an element of I. Thus, prime operator based OAC-triclusters in a three-dimensional matrix (tensor) form contain an absolutely dense cross-like structure of crosses (or ones).

The proposed OAC-tricluster definition has a fruitful property (see Proposition 1): for every triconcept in a given tricontext there exists a tricluster of the same
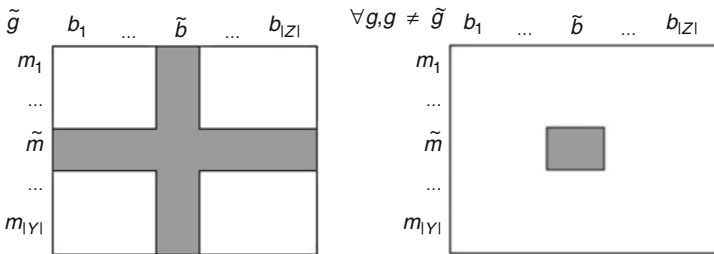

Fig. 3 Prime operator based tricluster structure

tricontext in which the triconcept is contained w.r.t. component-wise inclusion. It means that there is no information loss, we keep all the triconcepts in the resulting tricluster collection.

Proposition 1 *Let $\mathbb{K} = (G, M, B, I)$ be a triadic context and $\rho_{\min} = 0$. For every $T_c = (X_c, Y_c, Z_c) \in \mathfrak{T}(G, M, B, I)$ with non-empty X_c, Y_c, and Z_c there exists a prime OAC-tricluster $T = (X, Y, Z) \in \mathscr{T}'(G, M, B, Y)$ such that $X_c \subseteq X, Y_c \subseteq Y, Z_c \subseteq Z$. Here, $\mathscr{T}'(G, M, B, I)$ denotes the set of all OAC-prime triclusters fulfilling the chosen value of $\rho_{\min}$.*

Proof Let $(g, m, b) \in X_c \times Y_c \times Z_c$. By the definition of prime operators $(m, b)' := \{\widetilde{g} \mid (\widetilde{g}, m, b) \in I\}$. Since $m \in Y_c$ and $b \in Z_c$ then by the definition of formal triconcept (m, b) is related by Y to every $\widetilde{g} \in X_c$, therefore $(m, b)' \cap X_c = X_c$. Consequently for all $g_i \in X_c$ we have $g_i \in (m, b)'$. For $(g, b)'$ and $(g, m)'$ tricluster components the proof is similar. Finally, we have $X_c \subseteq X = (m, b)', Y_c \subseteq Y = (g, b)'$, and $Z_c \subseteq Z = (g, m)'$.

Prime-based n-clustering can be introduced similarly. Let $\mathbb{K} = (X_1, X_2, \ldots, X_n, Y)$ be an n-adic context and Y is binary relation between $X_1 \ldots X_n$.

Then for a tuple $(x_1, x_2, \ldots, x_n) \in Y$ we define n prime operators for each tuple $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ as follows:

$$(\{x_1\}, \ldots, \{x_{i-1}\}, x_{i+1}, \ldots, \{x_n\})' = \{z_i \mid (x_1, \ldots, x_{i-1}, z_i, x_{i+1}, \ldots, x_n) \in Y\}.$$

For a given tuple $(x_1, x_2, \ldots, x_n) \in Y$, a prime operator based n-cluster is defined as follows:

$$P = ((\{x_2\}, \ldots, \{x_n\})', \ldots, (\{x_1\}, \ldots, \{x_{i-1}\}, \{x_{i+1}\}, \ldots, \{x_n\})', \ldots,$$

$$(\{x_1\}, \ldots, \{x_{n-1}\})').$$

The density of n-cluster $P = (Z_1, Z_2, \ldots, Z_n)$ is $\rho(P) = \frac{|Y \cap Z_1 \times Z_2 \times \ldots \times Z_n|}{|Z_1 \times Z_2 \times \ldots \times Z_n|}$. To keep analogy of ρ with physical density we refer to its numerator as the mass of P, i.e. $\mathrm{mass}(P)$, while its denominator plays a role of the volume of P, i.e. $\mathrm{vol}(P)$.

The description of a one-pass algorithm for OAC-prime tricluster generation can be found in [29]. A Map-Reduce based prototype of OAC-prime triclustering and possible implementation variants are presented in [94].

4   Quality Measures for Multimodal Clustering

4.1   Connection Between ρ and Local Clustering Coefficient

Since we use density as a local measure of n-cluster quality, it is useful to find its connection to local clustering coefficients (we use $cc_\bullet(\cdot)$ notation from [57]). For

$(V, E \subseteq V \times V)$, the local clustering coefficient is $cc_\bullet(v) = \frac{|N(v) \times N(v) \cap E|}{N(v)(N(v)-1)/2}$, here $N(v)$ is the degree of $v \in V$.

If one considers a one-mode network $(V, E \subseteq V \times V)$ as a formal context $\mathbb{K} = (G, G, I \subseteq G \times G)$, where $V = G$, and for $g, m \in V$ $gEm \iff gIm$, then for bicluster (g', g') it follows that[3]

$$\rho(g', g') = \frac{|g' \times g' \cap I|}{|g'||g'|} = \frac{|N(g) \times N(g) \cap I|}{|N^2(g)|} = \frac{|N(g) \times N(g) \cap I|}{\frac{(|N(g)|-1)|N(g)|}{2}} \frac{1 - 1/|N(g)|}{2}$$

$$= cc_\bullet(g) \frac{1 - \frac{1}{|N(g)|}}{2}.$$

Note that $N(g) = \deg(g) = \{u | gEu\} = g'$.

Moreover, for large neighbourhoods $\rho(g', g') \approx \frac{cc_\bullet(g)}{2}$.

4.2 Connection Between ρ and Modularity

Since we do not optimise any modularity-like criterion in our study, multimodal clusters are supposed to be overlapping in general, and, moreover, to the best of our knowledge there is no widely accepted modularity criterion even for bipartite overlapping communities; the introduction and study of such criteria could be a subject of a separate research. However, we show the interconnection between average of values in the input modularity matrix for a particular bicluster and its density.

Let A_{gm} be the adjacency matrix of an input context $\mathbb{K} = (G, M, I \subseteq G \times M)$, i.e. $A_{gm} = [gIm]$[4] for $(g, m) \in G \times M$. For bipartite graphs an entry of modularity matrix is defined as follows:

$$B_{gm} = A_{gm} - \frac{\deg(g)\deg(m)}{|I|} = [gIm] - \frac{|g'||m'|}{|I|}.$$

For non-overlapping communities modularity in two-mode networks is defined as follows [4]:

$$Mod = \frac{1}{|I|} \sum_{(g,m) \in G \times M} \left([gIm] - \frac{|g'||m'|}{|I|} \right) [(g, m) \in C], \text{ where}$$

---

[3]Note that technically (g', g') is not an OA-bicluster since $(g, g) \notin I$.

[4]Here $[\cdot]$ means Iverson bracket defined as $[P] = \begin{cases} 1 & \text{if } P \text{ is true;} \\ 0 & \text{otherwise,} \end{cases}$.

$C \subseteq G \times M$ is a module (or community) from a set of non-overlapping communities $\mathscr{C}$ of the original network. Non-overlapping here is formally defined as follows: $\forall C, D \in \mathscr{C} \; C \cap D = \emptyset$.

Let (m', g') be a bicluster of $\mathbb{K}$, then the sum over all entries $(\widetilde{g}, \widetilde{m}) \in m' \times g'$ in B gives

$$|m' \times g' \cap I| - \frac{\displaystyle\sum_{\widetilde{(g,m)} \in m' \times g'} |\widetilde{g'}||\widetilde{m'}|}{|I|}.$$

Instead of normalising that sum by $|I|$ as in modularity definition, we can try to calculate (local) bicluster modularity, $\mathrm{Mod}_l(m', g')$, by normalising the sum by the bicluster volume $\mathrm{Vol}(m', g') = |g'||m'|$:

$$\mathrm{Mod}_l(m', g') = \frac{|m' \times g' \cap I|}{|g'||m'|} - \frac{\displaystyle\sum_{\widetilde{g} \in m'} |\widetilde{g'}| \sum_{\widetilde{m} \in g'} |\widetilde{m'}|}{|g'||m'||I|} = \rho(m', g') - \frac{\overline{\deg(\widetilde{g})}\,\overline{\deg(\widetilde{m})}}{|I|}, \text{ where}$$

$\overline{\deg(\widetilde{g})} = \frac{\sum_{\widetilde{g} \in m'}}{|g'|}$ is the average degree of $\widetilde{g}$ in the input bicluster and $\overline{\deg(\widetilde{m})}$ is the average degree of $\widetilde{m}$ and defined similarly.

It is clear that to maximise Mod_l criterion one needs to find a bicluster with high density and low average degrees of its elements.

However, the original modularity criterion for bipartite non-overlapping networks has intrinsic drawbacks. The first problem, low resolution, consists in the dependence between the size of detected communities and the size of an input graph [21]. Another one can be demonstrated by a model example.

Let $\mathbb{K} = (G, M, I)$ be a formal context, where for a certain pair $(g, m) \in I$ we have $g' = M, m' = G$, and $I = m' \times m \cup g \times g'$. Without loss of generality let $|G| = |M| = n$. Then

$$B_{gm} = [gIm] - \frac{|g'||m'|}{|I|} = 1 - \frac{n^2}{2n - 1}.$$

For large n, $B_{gm} \approx 1 - n/2$ and this value tends to $-\infty$ by implying $n \to \infty$. To keep the second term of an entry of the modularity matrix no greater than 1 (the maximal probability of incidence of g and m), one needs to require $|g'|, |m'| \leq \sqrt{|I|}$ (which is in fact should be normally fulfilled for large and sparse (real) networks).

4.3 Least Square Optimal n-Clusters

One of the important statistics in clustering is the data scatter of an input matrix, i.e. the sum of squares of all its entries [63]. In [64], least squares based maximisation criterion to generate n-cluster was proposed:

$$g(P) = \rho^2(P) \cdot \text{Vol}(P) = \rho(P) \cdot \text{mass}(P), \text{ where}$$

P is an n-cluster of a certain n-adic context. On the one hand, its direct interpretation implies that we care about dense n-clusters of large size instead of only dense (that may be small) or only large (that may be sparse); in other words such n-clusters tend to be massive (with low number of missing tuples in the input binary relation) and dense. On the other hand, this criterion measures the contribution of P to the data scatter of the input n-adic context.

In [45], one can find a theorem saying that by maximisation of $g(P)$ we require higher density within n cluster P than in the corresponding outside regions along its dimensions.

4.4 Weak Bicluster Communities and Graph Cuts

In network analysis, a community is called weak if its average internal degree is greater than its average out degree [3].

In two-mode case, for an input context $\mathbb{K} = (G, M, I)$ and its bicluster (m', g'), we have

$$\sum_{\widetilde{g} \in m'} |(\{\widetilde{g}\} \cup \{g\})'| + \sum_{\widetilde{m} \in g'} |(\{\widetilde{m}\} \cup \{m\})'| \geq \sum_{\widetilde{g} \in m'} |\widetilde{g}' \cap M \setminus g'| + \sum_{\widetilde{m} \in g'} |\widetilde{m}' \cap G \setminus m'|.$$

The left-hand side of the inequality is the doubled sum of the number of object-attribute pairs from (m', g'). The right-hand side shows how many pairs the objects from bicluster extent and the attributes from bicluster intent form with the remaining attributes and objects of the context, respectively. In network analysis this measure is known as *cut* [21], i.e. the number of edges one should delete to make the community disconnected from the remaining vertices in the input graph. Thus, the inequality can be rewritten as follows:

$$\rho(m', g') \geq \frac{\text{cut}(m', g')}{2|g'||m'|}.$$

This criterion can be used for selection of biclusters during their generation instead of fixed $\rho_{\min}$.

4.5 Stability of OA-Biclusters

Stability of formal concepts [53, 54] has been used as a means of concepts' filtering in studies on epistemic communities [56, 77, 78] and communities of website visitors [55].

Let $\mathbb{K} = (G, M, I)$ be a formal context and (A, B) be a formal concept of $\mathbb{K}$. The *(intensional) stability index*, σ, of (A, B) is defined as follows:

$$\sigma(A, B) = \frac{|\{C \subseteq A \mid C' = B\}|}{2^{|A|}}$$

As we know, not all of the OA-biclusters of a given formal context are formal concepts.

Only those OA-biclusters that fulfill condition $(m', g') = (g'', m'')$ are formal concepts. However, stability index can be technically computed for any OA-bicluster as follows:

$$\sigma(m', g') = \frac{|\{A \subseteq m' \mid A' = g'\}|}{2^{|m'|}}$$

Set $2^{m'}$ can be decomposed into three parts: $2^{g''} \cup 2^{m' \setminus g''} \cup \Delta$. The numerator is equal to $|\{A \in 2^{g''} \mid A' = g'\}| + |\{A \in 2^{m' \setminus g''} \mid A' = g'\} \setminus \emptyset| + |\{A \in \Delta \mid A' = g'\} \setminus \emptyset|$. Since every set of objects from $m' \setminus g''$ does not have all attributes from g', the second summand is 0, and the same applies to the third one due to each set from Δ contains at least one object $\widetilde{g}$ from $m' \setminus g''$ such that $\widetilde{g}' \neq g'$. Hence,

$$\sigma(m', g') = \frac{|\{A \in 2^{g''} \mid A' = g'\}|}{2^{|m'|}}.$$

Since the number of all A that contain g is $|2^{g'' \setminus g}|$, the tight lower bound of OA-bicluster's stability is $2^{|g'' \setminus g| - |m'|}$.

The stability index of a concept indicates how much the concept intent depends on particular objects of the extent.

4.6 Coverage and Diversity

Diversity is an important measure in Information Retrieval for diversified search results and in Machine Learning for ensemble construction [82].

To define diversity for multimodal clusters we use a binary function that equals to 1 if the intersection of triclusters T_i and T_j is not empty, and 0 otherwise.

$$\text{intersect}(T_i, T_j) = \left[G_{T_i} \cap G_{T_j} \neq \emptyset \wedge M_{T_i} \cap M_{T_j} \neq \emptyset \wedge B_{T_i} \cap B_{T_j} \neq \emptyset \right] \qquad (9)$$

It is also possible to define *intersect* for the sets of objects, attributes, and conditions. For instance, $\text{intersect}_G(T_i, T_j)$ is equal to 1 if triclusters T_i and T_j have non-empty intersection of their extents, and 0 otherwise.

Now we can define *diversity of the tricluster set $\mathscr{T}$*:

$$\text{diversity}(\mathscr{T}) = 1 - \frac{\sum_j \sum_{i<j} \text{intersect}(T_i, T_j)}{\frac{|\mathscr{T}|(|\mathscr{T}|-1)}{2}} \qquad (10)$$

The *diversity for the sets of objects (attributes or conditions)* is similarly defined:

$$\text{diversity}_G(T) = 1 - \frac{\sum_j \sum_{i<j} \text{intersect}_G(T_i, T_j)}{\frac{|\mathscr{T}|(|\mathscr{T}|-1)}{2}} \qquad (11)$$

Coverage is defined as a fraction of the triples of the context (alternatively, objects, attributes, or conditions) included in at least one of the triclusters of the resulting set.

More formally, let $\mathbb{K} = (G, M, B, I)$ be a tricontext and $\mathscr{T}$ be the associated triclustering set obtained by some triclustering method, then coverage of $\mathscr{T}$:

$$\text{coverage}(\mathscr{T}) = \sum_{(g,m,b)\in I} \left[(g, m, b) \in \bigcup_{(X,Y,Z)\in\mathscr{T}} X \times Y \times Z \right] / |I|. \qquad (12)$$

The *coverage of the object set G* by the tricluster collection $\mathscr{T}$ is defined as follows:

$$\text{coverage}_G(\mathscr{T}) = \sum_{g\in G} \left[g \in \bigcup_{(X,Y,Z)\in\mathscr{T}} X \right] / |G|. \qquad (13)$$

Coverage of attribute or condition sets can be defined analogously. These measures may have sense when one would like to know how many actors or items in the network do not belong to any found community.

We also use the *coverage of formal concepts by biclusters*, i.e. we count the number of concepts covered by at least one bicluster in the corresponding bicluster collection B. We say that bicluster $B = (X, Y)$ covers concept $C = (Z, W)$ w.r.t. component-wise inclusion of their extents and intents, namely $C \sqsubseteq B : \iff Z \subseteq X$ and $W \subseteq Y$.

$$\text{coverage}_{\mathscr{B}}(\mathfrak{B}(G, M, I)) = \frac{\{C \in \mathfrak{B}(G, M, I) \mid \exists B \in \mathscr{B} : C \sqsubseteq B\}}{|\mathfrak{B}(G, M, I)|}. \qquad (14)$$

5 Data

For our experiments we collected datasets from one-mode to four-mode networks.

In particular, we have analysed the following classic one-mode datasets:

- Karate club, 34×34, 78 edges;
- Florentine family 1, 16×16, 40 edges;
- Florentine family 2, 16×16, 30 edges;
- Hi-tech, 36×36, 147 edges;
- Mexican people, 35×35, 117 edges.

For two-mode datasets we have used Southern women of size 18×14 with 93 edges and four datasets studied in [57]:

- co-authoring, 19,885×16,400, and 45,904 edges;
- co-occurrence, 13,587×9,263, and 1,833,63 edges;
- actor, 127,823×383,640, and 1,470,418 edges;
- p2p, 1,986,588 peers×5,380,546 data, and 55,829,392 links (edges).

As for three-mode network, we have analysed Bibsonomy dataset[5] with $|U| = $ 2467 users, $|T| = $ 69,904 tags, $|R| = $ 268,692 resources that related by $|Y| = $ 816,197 triples.

Finally, MovieLens data[6] with 100,000 ratings (integers from 1 to 5) and 1300 tag applications applied to 9000 movies by 700 users is considered as a four-mode dataset. We have used only user, movie, rating, and time modes.

6 Experiments

We have tested our implementations for one- and two-mode networks in Python 2.7 and for higher modes in C# with our tool, Multimodal Clustering Toolbox, on a Mac Pro computer with 3.7 GHz and 16 GB RAM.

6.1 Two-Mode Networks

For each two-mode dataset we report the number of unique biclusters and the number of all generated biclusters; note that when all objects (and attributes) are pairwise different there are no duplicates by definition.

---

[5]http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/.

[6]http://grouplens.org/datasets/movielens/.

Table 1 Southern women: 18×14, 93 edges

ρ	Concept coverage	Unique biclusters	Biclusters	Fraction of covered concepts
0	65	83	93	1.00
0.05	65	83	93	1.00
0.1	65	83	93	1.00
0.15	65	83	93	1.00
0.2	65	83	93	1.00
0.25	65	83	93	1.00
0.3	65	83	93	1.00
0.35	65	82	92	1.00
0.4	65	81	91	1.00
0.45	65	77	87	1.00
0.5	65	71	81	1.00
0.55	65	63	73	1.00
0.6	65	60	7	1.00
0.65	64	51	59	0.98
0.7	63	40	47	0.97
0.75	57	33	4	0.88
0.8	51	22	28	0.78
0.85	35	13	19	0.54
0.9	20	7	9	0.31
0.95	0	0	0	0.00
1	0	0	0	0.00

For small and medium size classic two-mode and one-mode datasets we have reported the number of formal concepts covered by the generation bicluster collection for a specific ρ_{min} as well as their fraction, i.e. coverage$_{\mathscr{B}}(\mathfrak{B}(G, M, I))$.

In 1930s, a group of ethnographers collected data on the social activities of 18 women over a nine-month period [17]. Different subgroups of these women had met in 14 informal social events; the incidence of a woman to a particular event was established using "interviews, the records of participant observers, guest lists, and the newspapers" [17, p. 149]. Later on, this Southern Women dataset has become a benchmark for comparing communities detection methods in two-mode social network analysis, in particular, including concept lattices as a community detection approach [22, 23].[7]

The results of our experiments with Southern Women dataset are summarised in Table 1.

---

[7]There is a small inconsistency in the profiles of women w_{14} (Helen) and w_{15} (Dorothy), namely between their description in [22] and the downloaded dataset provided at https://networkdata.ics. uci.edu/netdata/html/davis.html, thus according to the latter $e_{12}, e_{13} \in w'_{14}$ and $e_{11}, e_9 \in w'_{15}$.

There are 66 formal concepts for the Southern woman network. Since OA-biclusters are tolerant to missing values, let us illustrate how rather dense biclusters include the largest concepts with non-empty extent and intent.

For example, with $\rho_{\min} = 0.8$ we show five bicluster-concept pairs $B_i = (e', w')$, $C_i = (W, E)$ related by component-wise inclusion of their extents and intents, respectively, namely $C_i \sqsubseteq B_i : \iff W \subseteq e'$ and $E \subseteq w'$:

1. $C_1 = (\{w_0, w_1, w_2, w_3, w_5, w_6, w_7\}, \{e_5, e_7\}) \sqsubseteq B_1 = (\{w_0, w_1, w_2, w_3, w_5, w_6, w_7, w_8\}, \{e_2, e_4, e_5, e_7\})$ with $\rho(B_1) = 0.84$;
2. $C_2 = (\{w_0, w_2, w_3\}, \{e_2, e_3, e_4, e_5, e_7\}) \sqsubseteq B_2 = (\{w_0, w_2, w_3, w_4\}, \{e_0, e_2, e_3, e_4, e_5, e_6, e_7\})$ with $\rho(B_2) = 0.82$;
3. $C_3 = (\{w_9, w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}\}, \{e_{11}\}) \sqsubseteq B_3 = (\{w_9, w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}\}, \{e_6, e_7, e_8, e_{11}\})$ with $\rho(B_3) = 0.82$;
4. $C_4 = (\{w_{10}, w_{11}, w_{12}, w_{15}\}, \{e_7, e_8, e_9, e_{11}\}) \sqsubseteq B_4 = (\{w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}\}, \{e_7, e_8, e_9, e_{11}\})$ with $\rho(B_4) = 0.92$;
5. $C_5 = (\{w_{16}, w_{17}, w_{13}\}, \{e_1, e_8\}) \sqsubseteq B_5 = (\{w_{16}, w_{17}, w_{13}, w_{14}\}, \{e_1, e_8\})$ with $\rho(B_5) = 0.88$.

The corresponding bipartite graph is shown in Fig. 4 along with the biclique formed by elements of concept C_1 and bicluster B_1, and concept C_3 and bicluster B_3. According to [18, 22] there is the "true structure" of the Southern women network: namely, there are two groups of women $\{w_0, \ldots, w_8\}$ and $\{w_1, \ldots, w_{17}\}$. The first group of women participated in events e_0 through e_4, while the second group was not. The second group participated in events e_3 through e_{13}, while the first group was not. Both groups participated e_6, e_7, and e_8.

Since the Southern women network is a well-studied case in SNA community and one of the first SNA datasets analysed by sociologists using concept lattices, an interested reader may refer to [22, 23] to find professional interpretation of several important communities of women found by means of formal concepts.

Even though that such networks as co-authoring, co-occurrence, actor, and p2p are two-mode and known to SNA community about a decade, even the number of concepts (maximal bicliques) for these datasets is not reported in the literature (Tables 2 and 3).

An interesting issue has appeared: At which $\rho_{\min}$ the generated biclusters do not cover all formal concepts with non-empty extent and intent? According to our experiments for two-mode (see also Appendix) and one-mode networks, it usually happens around $\rho_{\min} = 0.5$ or higher (containing intervals marked by two horizontal lines in the tables), so, we may hypothesise that one can normally set minimal density value equal to 0.5.
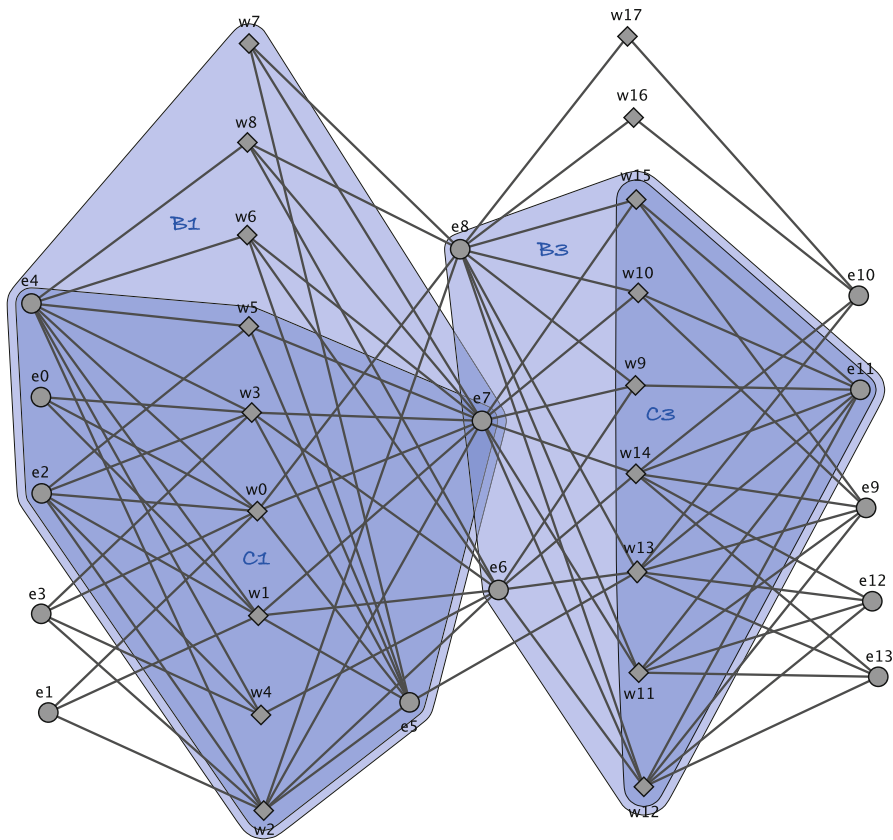
Fig. 4 The two-mode network for the Southern women dataset, bicluster B_1 and concept C_1, and bicluster B_3 and concept C_3

6.2 Folksonomies as Three-Mode Networks

Folksonomy is a typical example of a three-mode network, where a hyperedge connects a user, a tag, and an attribute. Thus each hyperedge is a set of size three with three vertices of different types; it is convenient to represent edges as tuples (*user*, *tag*, *resource*). Since we experiment with Bibsonomy, a Folksonomy-based resource sharing system for scientific bibliography, our users are scientists, resources are papers that they bookmarked or even authored; a tag is assigned by a scientist to a particular paper while bookmarking.

Table 2 The numbers of unique and all OA-biclusters for the four large two-mode networks

	Datasets							
	Co-authoring		Co-occurrence		Actor		p2p	
ρ	Unique biclusters	Biclusters	Unique biclusters	Biclusters	Unique biclusters	Biclusters	Unique biclusters	Biclusters
0	43,253	45,904	161,386	183,363	1,278,989	1,470,418	54,789,256	55,829,169
0.05	43,253	45,904	161,386	183,363	1,226,429	1,417,827	41,937,580	42,973,016
0.1	43,253	45,904	160,200	181,630	962,389	1,153,704	27,178,639	28,196,480
0.15	43,253	45,904	124,383	137,367	700,207	891,401	18,320,253	19,321,315
0.2	43,251	45,902	69,283	75,761	523,446	714,509	13,179,196	14,165,402
0.25	43,184	45,835	39,081	43,252	410,118	601,065	9,789,039	10,759,880
0.3	42,748	41,774	24,484	27,672	318,245	509,068	7,019,097	7,969,965
0.35	41,774	44,423	17,011	19,718	269,642	460,361	5,088,606	6,017,582
0.4	39,366	42,008	12,796	15,100	214,979	405,543	3,950,659	4,856,567
0.45	36,194	38,809	10,111	12,251	190,704	381,106	3,369,522	4,261,678
0.5	34,141	36,737	8539	10,515	182,906	373,191	3,056,597	3,938,536
0.55	29,404	31,960	6926	8699	110,464	299,895	1,156,887	1,918,111
0.6	23,150	25,615	5395	7036	84,459	272,894	764,584	1,483,586
0.65	20,604	23,007	4572	6127	77,904	265,699	614,743	1,308,939
0.7	16,391	18,707	3929	5386	72,651	259,877	50,981	1,182,631
0.75	15,951	18,234	3726	5129	71,663	258,550	472,869	1,126,702
0.8	12,989	15,137	3490	4846	69,449	255,904	419,533	1,046,786
0.85	11,533	13,530	3313	4568	68,555	254,703	39,189	986,811
0.9	11,053	12,976	3214	4437	68,186	254,138	377,377	949,637
0.95	10,875	12,756	3105	4290	67,871	253,623	369,401	929,765
1	10,874	12,756	3079	4250	67,798	253,390	367,946	926,380

Table 3 Elapsed time for online OA-biclustering

Dataset	$|I|$	$G|$	$|M|$	Time, s
Co-authoring	45,904	19,885	16,400	0.13
Co-occurrence	183,363	13,587	9264	0.25
Actor	1,470,418	127,823	383,640	3.55
p2p	55,829,392	19,86,588	5,380,546	260.13

Let us consider a toy imaginary example of Bibsonomy data; the input context is shown by three layers in Table 4. There are four users (u_1 = *Fortunato*, u_2 = *Freeman*, u_3 = *Newman*, and u_4 = *Roth*) and three tags (t_1 = *Galois Lattices*, t_2 = *SNA*, and t_3 = *Statistical Physics*). Three papers p_1, p_2, and p_3 are marked according to the research interests of those users. Thus Freeman and Roth marked paper 1 by tags "Galois Lattices" and "SNA",

(continued)

while Fortunato and Newnam tagged paper 3 by tags 'SNA" and "Statistical Physics". All the users assigned tag "SNA" to paper 2. Three corresponding communities can be easily captured by formal triconcepts:

$$C_1 = (\{u_2, u_4\}, \{t_1, t_2\}, \{p_1\})$$
$$C_2 = (\{u_1, u_3\}, \{t_2, t_3\}, \{p_3\})$$
$$C_3 = (\{u_1, u_2, u_3, u_4\}, \{t_2\}, \{p_2\}).$$

Concept C_3 is more general than C_1 and C_2 w.r.t. extent inclusion, and corresponds to SNA-interested users, while C_1 corresponds to those who are interested in concept lattices for SNA domain, and C_2 unites users interested in SNA by means of methods similar to their prototypes in Statistical Physics. The corresponding hypergraph with these triconcepts is shown in Fig. 5.

Table 4 A toy example with Bibsonomy data

	t_1	t_2	t_3
u_1			
u_2	×	×	
u_3			
u_4	×	×	
	p_1		

	t_1	t_2	t_3
u_1		×	
u_2		×	
u_3		×	
u_4		×	
	p_2		

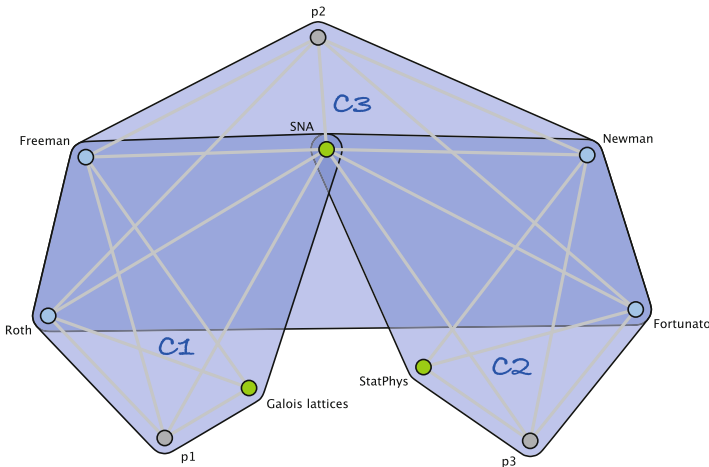|	t_1	t_2	t_3
u_1		×	×
u_2			
u_3		×	×
u_4			
	p_3		

Fig. 5 Three triconcepts C_1, C_2, C_3 for the Bibsonomy three-mode network

Table 5 Experimental results for k first triples of Bibsonomy dataset with $\rho_{\min} = 0$

k, number of first triples	$|U|$	$|T|$	$|R|$	$|\mathfrak{T}|$	$|\mathscr{T}_{OAC'}|$	TRIAS, s	OAC-Prime,s	
							Full time	Online phase
100	1	47	52	57	77	0.2	0.02	0.003
1000	1	248	482	368	656	1	0.043	0.001
10,000	1	444	5193	733	1461	2	273	0.031
100,000	59	5823	28,920	22,804	33,172	3386	24,185	0.542
200,000	340	14,982	61,568	–	105,571	> 24 h	25,446	1.268
500,000	1191	45,232	148,695	–	316,139	> 24 h	29,035	3.529
816,197	2467	69,904	268,692	–	484,349	> 24 h	241,341	5.186

Table 6 Density distribution of OAC-prime triclusters for 816,197 triples of Bibsonomy dataset with $\rho_{\min} = 0$

Lower bound of ρ	Upper bound of ρ	Number of triclusters
0	0.05	172
0.05	0.1	3070
0.1	0.2	36,878
0.2	0.3	77,170
0.3	0,4	90,005
0.4	0.5	67,659
0.5	0.6	66,711
0.6	0.7	41,507
0.7	0.8	22,225
0.8	0.9	11,662
0.9	1	67,290

To build all triconcepts of a certain context we have used a Java implementation of the TRIAS algorithm by R. Jäschke [46]. The last two columns in Table 5 mean time of execution of TRIAS and OAC-prime algorithms.

Note that here we have reported both the full execution time of OAC-prime algorithm, i.e. tricluster generation with density calculation, and the time of online phase for tricluster generation only. One may note a dramatical drop-off in time efficiency between the last and penultimate lines in Table 5 for the full execution time, while online phase took only about half a second more. The devil is in the hashing data structures used for duplicate elimination and we believe the timing can be improved, for example, by a specially designed Bloom filter. Note that a more general and efficient algorithm Data-Peeler [13] could be used suitable for mining n-concepts.

Distribution of density of triclusters for all the triples of Bibsonomy dataset is given in Table 6.

6.3 MovieLens Data as Four-Mode Network

We summarise the results of prime-based tetraclustering execution on Movielens data below:

Time:	13, 252 ms
Number of n-clusters:	89,931
Average volume, $\overline{Vol}$:	455.4
Average density, $\overline{\rho}$:	0.35
Average coverage:	0.1%
Average mass, $\overline{mass}$:	103.7
Average $\rho \cdot mass$:	28.1

In addition to average density we report average volume, average coverage (the number of covered original tuples by each tetracluster on average), average mass (the number of tuples inside each tetraclusters on average), and quite an interesting statistic, average $\rho \cdot mass$. If we maximise the latter criterion, then we require for our tetraclusters to be dense and large at the same time while criterion $\rho \cdot Vol$ could result in sparse patterns.

To provide concrete examples of tetraclusters, we have selected rather small-sized dense communities in Table 7.

Table 7 Tetraclusters for Movielens data

No.	Generating tuple	Volume	ρ	Coverage	$mass$	$\rho \cdot mass$
1	(483, Star Trek IV, 5, 1997/11)	27	0.93	0.03 %	25	23.1
2	(384, Evita, 5, 1998/03)	15	0.87	0.01 %	13	11.3
3	(872, Scream 2, 5, 1998/02)	15	0.87	0.01 %	13	11.3
4	(102, Face/Off, 3, 1997/10)	12	0.92	0.01 %	11	10.1
5	(750, Gang Related, 1, 1997/11)	9	1.00	0.01 %	9	9.0

No.	Users	Movies	Rating	Time
1	{109,307,374,483, 87,545,815,882,927}	{Star Trek: The Wrath of Khan (82), Star Trek IV: The Voyage Home (86), Star Wars (77) }	{5}	{97/11}
2	{378,384,392}	{Good Will Hunting (97), Evita (96), Titanic (97), L.A. Confidential (97), As Good As It Gets (97)}	{5}	{98/03}
3	{206,332,872}	{Time to Kill, A (96), Scream (96), Scream 2 (97), Air Force One (97), Titanic (97)}	{5}	{98/02}
4	{102,116,268,430}	{Grosse Pointe Blank (1997), Face/Off (1997) } Air Force One (1997)}	{3}	{97/10}
5	{181,451,750}	{Gang Related (1997), Rocket Man (1997) Leave It to Beaver (1997)}	{1}	{97/11}

For example, one can easily identify the community of modern space opera lovers in 4-cluster no. 1. Note that their third and fourth components are always sets containing a single element due to the chosen mode nature: the same people cannot rate the same movies by different marks simultaneously or within a different month.

6.4 One-Mode Networks as Two-Mode Ones

There are different techniques called projections to transform two-mode graphs to their one-mode versions [57, 67]. Sometimes, researchers even do transformations in backward direction to consider interactions between different subgroups of actors as they were from different modes of the corresponding two-mode network [18, 91].

An undirected one-mode network in the form $\Gamma = (G, E \subseteq G \times G)$ can be considered as the two-mode network by composing a context $\mathbb{K} = (G, G, I)$ where $gEh \iff gIh$ for any $g, h \in G$, with two options for I being a symmetric relation: a) reflexive and b) irreflexive.

In reflexive case, each concept (A, B) of such context $\mathbb{K}$ that fulfills $A = B$ corresponds to the maximal clique A in the original one-mode network.

We provide the reader with the results of OA-biclustering for one-mode networks in Tables 8, 9, 10, 11, and 12.

In addition to the fraction of covered concepts by component-wise set inclusion we have reported intervals $[\rho_\alpha, \rho_\beta]$, where the fraction of covered concepts decreases below 1 first time for each dataset (see two vertical lines in the tables).

In addition to the reported statistics, let us demonstrate found biclusters and concepts for Zachary's karate club dataset. Originally, the author of [90], an anthropologist, described social relationships between members of a karate club in the period of 1970–72; the network contains 34 active members of the karate club who interacted outside the club, including 78 pairwise links between them. The club was split into two parts after a conflict between its instructor and president. This dataset is usually used as a benchmark for demonstration and testing of community detection algorithms [3].

In Fig. 6, one can see three biclusters (B_1, B_2, and B_3) with density less than 1 but greater 0.8 each. Thus none of them is a concept; moreover, union of their intent and extent does not form a clique of the input one-mode network.

$$B_1 = (29', 29') = (\{32, 33, 26, 29, 23\}, \{32, 33, 26, 29, 23\}) \text{ with } \rho = 0.84$$

$$B_2 = (3', 12') = (\{0, 1, 2, 3, 7, 12, 13\}, \{0, 3, 12\}) \text{ with } \rho = 0.81$$

$$B_3 = (5', 4') = (\{0, 10, 4, 6\}, \{0, 10, 4, 5\}) \text{ with } \rho = 0.88$$

Among all generated concepts, each concept (X, Y) with $X = Y$ results in clique X.

Thus concept $(\{0, 1, 2, 3, 7\}, \{0, 1, 2, 3, 7\})$ forms clique $Q_1 = \{0, 1, 2, 3, 7\}$, while concepts $(\{0, 1, 2, 3, 13\}, \{0, 1, 2, 3, 13\})$ and $(\{32, 33, 29, 23\}, \{32, 33, 29, 23\})$ result in $Q_2 = \{0, 1, 2, 3, 13\}$ and $Q_3 = \{32, 33, 29, 23\}$, respectively. Those are cliques of maximal size 5 and 4 from two parts of the karate club after its fission. It is evident that for each of those cliques its set of vertices can be found in some OA-bicluster. One can check that the set of vertices of B_1 contains those of Q_3, and vertices of B_2 include those of Q_1 and Q_2. So, it is possible to conclude that even though the density of a bicluster may be less than 1, they can contain more vertices resulting in larger communities than cliques. Note that the club instructor, 0, belongs to extents of B_2 and B_3 being a "missing link" between two corresponding subcommunities, which lack in active interaction otherwise.

Table 8 Karate club: 34×34, 190 edges

| ρ	Covered concepts	Unique biclusters	Biclusters	Fraction of covered concepts
0	134	190	190	1.00
0.05	134	190	190	1.00
0.1	134	190	190	1.00
0.15	134	190	190	1.00
0.2	134	190	190	1.00
0.25	134	190	190	1.00
0.3	134	184	184	1.00
0.35	134	178	178	1.00
0.4	134	163	163	1.00
0.45	134	142	142	1.00
0.5	132	128	128	0.99
0.55	126	108	108	0.94
0.6	115	91	91	0.86
0.65	97	71	71	0.72
0.7	90	67	67	0.67
0.75	68	47	47	0.51
0.8	31	25	25	0.23
0.85	27	20	20	0.20
0.9	12	12	12	0.09
0.95	12	12	12	0.09
1	12	12	12	0.09
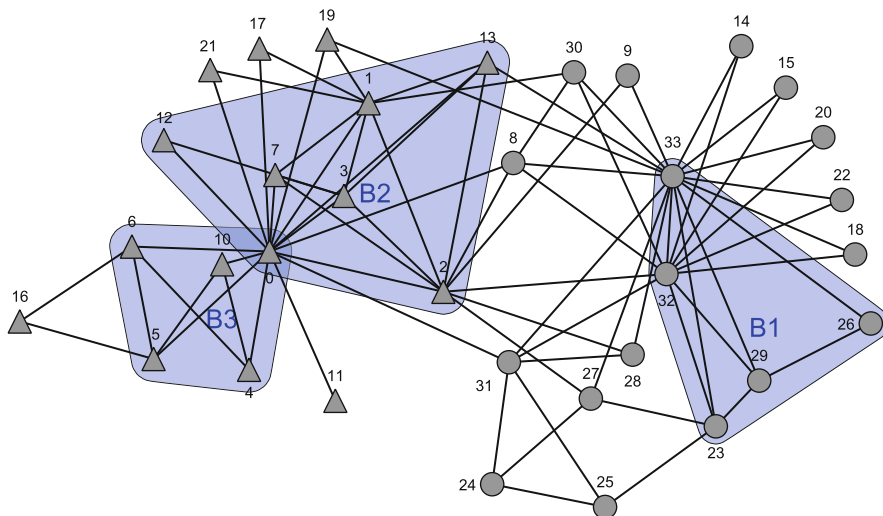
Fig. 6 Three dense biclusters B_1, B_2, B_2 found in Karate club network with $\rho_{\min} = 0.8$

In Fig. 7, one can see three found communities that are composed of vertices corresponding to three concepts C_1, C_2, and C_3.

$$C_1 = (\{32, 33\}, \{32, 33, 8, 14, 15, 18, 20, 22, 23, 29, 30, 31\})$$

$$C_2 = (\{0, 1\}, \{0, 1, 2, 3, 7, 13, 17, 19, 21\})$$

$$C_3 = (\{0, 10, 6\}, \{0, 4, 5\})$$

In this concrete example, the usage of formal concepts for representing communities seems to be even more beneficial than that of dense OA-biclusters since we have been able to cover almost both parts of the separated karate club by three concepts without sharing members between the counterparts; concepts C_1 and C_2 contain more vertices than biclusters B_1 and B_2 shown in Fig 6. Note that the semantic of C_1 lies in the interpretation of its intent as common contacts of 32 and 33, an active club member who is loyal to the club's president and the president, respectively. Intent of C_2 contains members mutually connected with the club instructor, 0, and member 1.
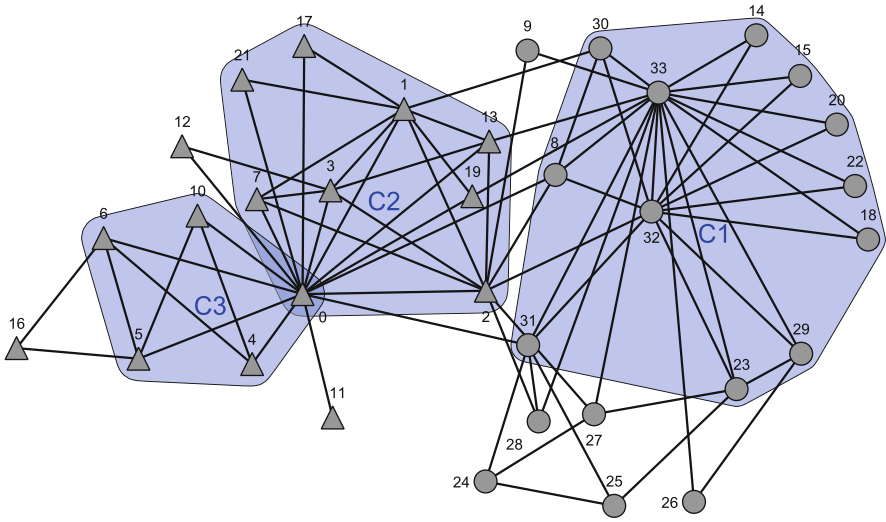
Fig. 7 Three formal concepts C_1, C_2, C_2 found in Karate club network

7 Related Work

There is a so-called subspace clustering [1] closely related to biclustering, where objects are considered as points in high dimensional space and clustered within multidimensional grid of a certain granularity. However, these methods cannot be directly applied to multidimensional relational data, i.e. multi-mode networks, since entities from different modes are often numbered arbitrarily and do not follow a pre-specified order like values along numerical axes. However, biclustering of numerical data, which may describe two-mode weighted networks, can be realised with Triadic Concept Analysis in case we consider attribute values as a mode of conditions under which an object has an attribute [50]. These results are also applicable to n-dimensional numerical datasets. Two other ways to deal with numeric data are to apply the so-called scaling, e.g., using a binary threshold, or Pattern Structures defined on vectors of numeric intervals [16, 25, 49]. Pattern Structures were also used to rethink collaborative filtering and find relevant taste communities for a particular user in terms of vectors of desirable rating intervals for good movies [34].

As for OA-biclustering, it has been used in several applications; for example, OA-biclustering has been applied for finding market segments in two-mode data on Internet advertising to recommend advertising terms to companies playing on

these segments [35, 39]. In crowdsourcing platforms, OA-biclustering helps to find similar ideas (proposals) to discuss potential collaborators [43, 44] as well as answer questions [14]; in case we consider opinions of users over a set of different ideas (proposals), it is possible to find antagonists, which may be prospective opponents in crowdsourcing teams [41].

In fact, biclustering is a well-established tool in Bioinformatics, especially for Gene Expression Analysis in genes-samples networks [49, 70]. A non-exhaustive concept lattice based taxonomy of biclustering techniques can be found in [36]. Methods for three-mode networks are applicable in this domain when in addition to genes and samples time mode comes [92].

Going back to networks, several researchers define other kinds of networks where the role of dimensions is played by different types of labels of multi-edges between actors [8, 9]; they call such networks multidimensional while others use the term multi-relational networks [88].

One more variation of networks is realised by n-partite networks where connection are edges between vertices of allowed types [80]. It is possible to mine maximal closed and connected subgraphs in them and interpret them as communities [59]; these patterns coincide with bicliques and formal concepts in two-mode case. However, for higher dimensions such n-partite graphs are not equivalent to n-adic contexts and may result in information loss or phantom hyperedges if we reduce the latter to the former or vice versa [33]. In [28], for analysing such tripartite network composed by two two-mode networks with one shared part, biclusters from these two networks have been used. Namely, those biclusters that are similar with respect to their extents are merged by taking the intersection of their extents. The intent of the first bicluster and the intent of the second bicluster become the intent and modus, respectively, of the resulting tricluster. In FCA domain, analysis of n-partite and multi-relational networks can be unified within Relational Concept Analysis where objects can be involved in different types of relations with attributes and each other [30].

Another related subject is tensor factorisation, which is of high importance in Data Mining [71] and Machine Learning [15] due to its ability to reduce data dimensionality, find the so-called hidden factors, and even perform information fusion. The closest approaches to ones in the presented study can be found in works on Boolean matrix [6, 7] and tensor factorisation [5, 62]. Thus in [7] it was shown that formal concepts may result in optimal factors in Boolean matrix decomposition; in [2, 42] these decompositions showed their competitive applicability to collaborative filtering by finding communities of similar tastes. Tensor clustering is another way to find dense patterns; this approach is very similar to multimodal clustering in n-ary relations, especially in case of Boolean tensors, which normally represent n-ary relations between entities [38, 61, 64, 79]. An interesting issue here, whether it is possible to obtain improvements in classification accuracy for tensors with labeled objects from one of their dimensions over conventional object-attribute representations [93].

Since the proposed multimodal clustering is an approach to find approximate patterns, not absolutely dense as closed *n*-sets or *n*-adic concepts, various similar ideas can be proposed. Thus, in [13] another type of fault-tolerant patterns was proposed, which is guided by the number of allowed non-missing tuples inside an *n*-cluster rather than by maximising their relative number. It seems that techniques searching for relaxed *n*-cliques maximal according a density-like criteria can be proposed for multi-mode networks as well [84]. The classic definition of *biplex* can be compared with the one of the OA-biclusters as many more similar relaxations for cliques and their possible *n*-adic generalisations [11].

Comparison of several existing triclustering techniques based on spectral clustering (SᴘᴇᴄTʀɪᴄ), least squares approximation (TʀɪBᴏx), OAC-prime and OAC-box operators, and formal triconcepts (TRIAS) can be found in [40, 45]. In [45], the complexity of the problem of optimal triclustering cover with respect to several quality criteria is discussed; it is shown that the problem belongs to NP-complete complexity class whereas the problem of the number of such covers belongs to #P.

Formal concepts and their lattices have been used in criminal studies to find communities of criminals operating together [72]. Many more successful applications based on FCA are known as well as related models and techniques [73, 74]. A comprehensive introduction to FCA can be found in the recent book [26] and application-oriented tutorial [32].

8   Conclusions

We have proposed a scalable technique for community detection in *n*-mode networks (where nodes are normally connected by hyperedges in case of $n > 2$). The approach welcomes improvements and may benefit from fine tuning and efficient filtering criteria in order to increase the scalability at the stage of density calculation and guarantee high-quality of the found communities. We consider several directions for such improvements: efficient hashing for elimination of duplicate patterns, strategies for approximate density calculation and selection of meaningful *n*-clusters as well as theoretical justification of choosing good thresholds for minimal density of *n*-clusters.

The proposed technique can also be compared with other existing approaches like fault-tolerant *n*-concepts [13] and with possible multimodal extensions of the existing ones like different techniques for relaxed cliques [84], variations of bicliques [68], or higher-order extensions of modularity-based criteria [66].

Since we have only showcased several relevant examples to community detection in multi-mode networks, validation of the method for analysing similar cases requires domain expert feedback, for example, by a sociologist practitioner.

Appendix: Experiments with One-Mode Networks

Table 9 Florentine family 1: 16×16, 58 edges

ρ	Covered concepts	Unique biclusters	Biclusters	Fraction of covered concepts
0	43	58	58	1.00
0.05	43	58	58	1.00
0.1	43	58	58	1.00
0.15	43	58	58	1.00
0.2	43	58	58	1.00
0.25	43	58	58	1.00
0.3	43	58	58	1.00
0.35	43	58	58	1.00
0.4	43	57	57	1.00
0.45	43	53	53	1.00
0.5	43	47	47	1.00
0.55	43	40	40	1.00
0.6	37	31	31	0.86
0.65	33	28	28	0.77
0.7	29	19	19	0.67
0.75	29	19	19	0.67
0.8	11	8	8	0.26
0.85	9	6	6	0.21
0.9	5	5	5	0.12
0.95	5	5	5	0.12
1	5	5	5	0.12

Table 10 Florentine family 2: 16×16, 46 edges

ρ	Covered concepts	Unique biclusters	Biclusters	Fraction of covered concepts
0	27	46	46	1.00
0.05	27	46	46	1.00
0.1	27	46	46	1.00
0.15	27	46	46	1.00
0.2	27	46	46	1.00
0.25	27	46	46	1.00
0.3	27	46	46	1.00
0.35	27	46	46	1.00
0.4	27	46	46	1.00
0.45	27	46	46	1.00
0.5	27	44	44	1.00
0.55	27	43	43	1.00
0.6	27	41	41	1.00
0.65	27	41	41	1.00
0.7	25	26	26	0.93
0.75	23	22	22	0.85
0.8	23	19	19	0.85
0.85	17	14	14	0.63
0.9	12	12	12	0.44
0.95	10	10	10	0.37
1	10	10	10	0.37

Table 11 Hi-tech: 36×36, 218 edges

ρ	Covered concepts	Unique biclusters	Biclusters	Fraction of covered concepts
0	191	218	218	1.00
0.05	191	218	218	1.00
0.1	191	218	218	1.00
0.15	191	218	218	1.00
0.2	191	218	218	1.00
0.25	191	218	218	1.00
0.3	191	218	218	1.00
0.35	191	213	213	1.00
0.4	191	198	198	1.00
0.45	191	174	174	1.00
0.5	189	134	134	0.99
0.55	163	99	99	0.85
0.6	126	78	78	0.66

Table 11 (continued)

ρ	Covered concepts	Unique biclusters	Biclusters	Fraction of covered concepts
0.65	86	49	49	0.45
0.7	65	31	31	0.34
0.75	47	22	22	0.25
0.8	28	16	16	0.15
0.85	16	13	13	0.08
0.9	16	13	13	0.08
0.95	12	12	12	0.06
1	12	12	12	0.06

Table 12 Mexican people: 35×35, 268 edges

ρ	Covered concepts	Unique biclusters	Biclusters	Fraction of covered concepts
0	373	268	268	1.00
0.05	373	268	268	1.00
0.1	373	268	268	1.00
0.15	373	268	268	1.00
0.2	373	268	268	1.00
0.25	373	266	266	1.00
0.3	373	260	260	1.00
0.35	373	247	247	1.00
0.4	373	225	225	1.00
0.45	371	189	189	0.99
0.5	360	151	151	0.97
0.55	348	119	119	0.93
0.6	298	69	69	0.80
0.65	211	45	45	0.57
0.7	141	24	24	0.38
0.75	86	15	15	0.23
0.8	17	5	5	0.05
0.85	13	4	4	0.03
0.9	1	1	1	0.00
0.95	1	1	1	0.00
1	1	1	1	0.00

References

1. Agrawal, R., Gehrke, J., Gunopulos, D., Raghavan, P.: Automatic subspace clustering of high dimensional data. Data Min. Knowl. Discov. **11**(1), 5–33 (2005). doi:10.1007/s10618-005-1396-1. http://dx.doi.org/10.1007/s10618-005-1396-1
2. Akhmatnurov, M., Ignatov, D.I.: Context-aware recommender system based on boolean matrix factorisation. In: Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications, pp. 99–110, Clermont-Ferrand, 13–16 October 2015. http://ceur-ws.org/Vol-1466/paper08.pdf

3. Barabási, A.: Network Science. Cambridge University Press, Cambridge (2016)

4. Barber, M.J.: Modularity and community detection in bipartite networks. Phys. Rev. E **76**, 066102 (2007). doi:10.1103/PhysRevE.76.066102. http://link.aps.org/doi/10.1103/PhysRevE.76.066102

5. Belohlávek, R., Glodeanu, C.V., Vychodil, V.: Optimal factorization of three-way binary data using triadic concepts. Order **30**(2), 437–454 (2013). doi:10.1007/s11083-012-9254-4. http://dx.doi.org/10.1007/s11083-012-9254-4

6. Belohlávek, R., Trnecka, M.: From-below approximations in boolean matrix factorization: Geometry and new algorithm. J. Comput. Syst. Sci. **81**(8), 1678 – 1697 (2015). doi:http://dx.doi.org/10.1016/j.jcss.2015.06.002. http://www.sciencedirect.com/science/article/pii/S002200001500063X

7. Belohlávek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Comput. Syst. Sci. **76**(1), 3–20 (2010). doi:10.1016/j.jcss.2009.05.002. http://dx.doi.org/10.1016/j.jcss.2009.05.002

8. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: Multidimensional networks: foundations of structural analysis. World Wide Web **16**(5), 567–593 (2013). doi:10.1007/s11280-012-0190-4. http://dx.doi.org/10.1007/s11280-012-0190-4

9. Berlingerio, M., Pinelli, F., Calabrese, F.: Abacus: frequent pattern mining-based community discovery in multidimensional networks. Data Min. Knowl. Discov. **27**(3), 294–320 (2013). doi:10.1007/s10618-013-0331-0. http://dx.doi.org/10.1007/s10618-013-0331-0

10. Bohman, L.: Bringing the owners back in: an analysis of a 3-mode interlock network. Soc. Netw. **34**(2), 275 – 287 (2012). doi:http://dx.doi.org/10.1016/j.socnet.2012.01.005. //www.sciencedirect.com/science/article/pii/S037887331200007X

11. Borgatti, S.P., Everett, M.G.: Network analysis of 2-mode data. Soc. Netw. **19**(3), 243 – 269 (1997). doi:http://dx.doi.org/10.1016/S0378-8733(96)00301-2. //www.sciencedirect.com/science/article/pii/S0378873396003012

12. Cerf, L., Besson, J., Robardet, C., Boulicaut, J.: Closed patterns meet *n*-ary relations. TKDD **3**(1), 3:1–3:36 (2009). doi:10.1145/1497577.1497580. http://doi.acm.org/10.1145/1497577.1497580

13. Cerf, L., Besson, J., Nguyen, K., Boulicaut, J.: Closed and noise-tolerant patterns in n-ary relations. Data Min. Knowl. Discov. **26**(3), 574–619 (2013). doi:10.1007/s10618-012-0284-8. http://dx.doi.org/10.1007/s10618-012-0284-8

14. Chatterjee, S., Bhattacharyya, M.: Judgment analysis of crowdsourced opinions using biclustering. Inf. Sci. **375**, 138–154 (2017). doi:10.1016/j.ins.2016.09.036. http://dx.doi.org/10.1016/j.ins.2016.09.036

15. Cichocki, A., Lee, N., Oseledets, I.V., Phan, A.H., Zhao, Q., Mandic, D.P.: Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. Found. Trends Mach. Learn. **9**(4–5), 249–429 (2016). doi:10.1561/2200000059. http://dx.doi.org/10.1561/2200000059

16. Codocedo, V., Napoli, A.: Lattice-based biclustering using partition pattern structures. In: ECAI 2014 - 21st European Conference on Artificial Intelligence, 18–22 August 2014, Prague - Including Prestigious Applications of Intelligent Systems (PAIS 2014), pp. 213–218 (2014). doi:10.3233/978-1-61499-419-0-213. http://dx.doi.org/10.3233/978-1-61499-419-0-213

17. Davis A., B.B.G., Gardner, M.R.: Deep South. The University of Chicago Press, Chicago (1941)

18. Doreian, P., Batagelj, V., Ferligoj, A.: Generalized blockmodeling of two-mode network data. Soc. Netw. **26**(1), 29–53 (2004). doi:http://dx.doi.org/10.1016/j.socnet.2004.01.002. //www.sciencedirect.com/science/article/pii/S0378873304000036

19. Duquenne, V.: Lattice analysis and the representation of handicap associations. Soc. Netw. **18**(3), 217–230 (1996). doi:10.1016/0378-8733(95)00274-X. http://www.sciencedirect.com/science/article/pii/037887339500274X

20. Fararo, T.J., Doreian, P.: Tripartite structural analysis: generalizing the Breiger-Wilson formalism. Soc. Netw. **6**(2), 141–175 (1984). doi:http://dx.doi.org/10.1016/0378-8733(84)90015-7. http://www.sciencedirect.com/science/article/pii/0378873384900157

21. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010). doi:http://dx.doi.org/10.1016/j.physrep.2009.11.002. http://www.sciencedirect.com/science/article/pii/S0370157309002841

22. Freeman, L.: Finding social groups: a meta-analysis of the southern women data. In: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, pp. 39–97. National Academy Press, Washington, DC (2003)

23. Freeman, L.C., White, D.R.: Using galois lattices to represent network data. Sociol. Methodol. **23**, 127–146 (1993)

24. Freeman, L.C.: Cliques, galois lattices, and the structure of human social groups. Soc. Netw. **18**, 173–187 (1996)

25. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In: Conceptual Structures: Broadening the Base, Proceedings of the 9th International Conference on Conceptual Structures, ICCS 2001, pp. 129–142, Stanford, CA, 30 July–3 August 2001. doi:10.1007/3-540-44583-8_10. http://dx.doi.org/10.1007/3-540-44583-8_10

26. Ganter, B., Obiedkov, S.A.: Conceptual Exploration. Springer, Heidelberg (2016). doi:10.1007/978-3-662-49291-8. http://dx.doi.org/10.1007/978-3-662-49291-8

27. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations, 1st edn. Springer, New York (1999)

28. Gnatyshak, D., Ignatov, D.I., Semenov, A., Poelmans, J.: Gaining insight in social networks with biclustering and triclustering. In: BIR, pp. 162–171 (2012)

29. Gnatyshak, D., Ignatov, D.I., Kuznetsov, S.O., Nourine, L.: A one-pass triclustering approach: Is there any room for big data? In: Proceedings of the Eleventh International Conference on Concept Lattices and Their Applications, pp. 231–242, Košice, 7–10 October 2014. http://ceur-ws.org/Vol-1252/cla2014_submission_26.pdf

30. Hacene, M.R., Huchard, M., Napoli, A., Valtchev, P.: Relational concept analysis: mining concept lattices from multi-relational data. Ann. Math. Artif. Intell. **67**(1), 81–108 (2013). doi:10.1007/s10472-012-9329-3. http://dx.doi.org/10.1007/s10472-012-9329-3

31. Hartigan, J.A.: Direct clustering of a data matrix. J. Am. Stat. Assoc. **67**(337), 123–129 (1972). doi:10.2307/2284710. http://dx.doi.org/10.2307/2284710

32. Ignatov, D.I.: Introduction to formal concept analysis and its applications in information retrieval and related fields. In: Information Retrieval - 8th Russian Summer School, RuSSIR 2014, pp. 42–141, Nizhniy, Novgorod, 18–22 August 2014. Revised Selected Papers (2014). doi:10.1007/978-3-319-25485-2_3. http://dx.doi.org/10.1007/978-3-319-25485-2_3

33. Ignatov, D.I.: Towards a closure operator for enumeration of maximal tricliques in tripartite hypergraphs. CoRR **abs/1602.07267** (2016). http://arxiv.org/abs/1602.07267

34. Ignatov, D.I., Kornilov, D.: RAPS: a recommender algorithm based on pattern structures. In: Proceedings of the 4th International Workshop "What Can FCA Do for Artificial Intelligence?", FCA4AI 2015, co-located with the International Joint Conference on Artificial Intelligence (IJCAI 2015), pp. 87–98, Buenos Aires, 25 July 2015. http://ceur-ws.org/Vol-1430/paper9.pdf

35. Ignatov, D.I., Kuznetsov, S.O.: Concept-based recommendations for internet advertisement. In: Belohlavek, R., Kuznetsov, S.O. (eds.) Proceedings of the CLA 2008, CEUR WS, vol. 433, pp. 157–166. Palacký University, Olomouc (2008)

36. Ignatov, D.I., Watson, B.W.: Towards a unified taxonomy of biclustering methods. In: Kuznetsov, S.O., Watson, B.W. (eds.) Proceedings of Russian and South African Workshop on Knowledge Discovery Techniques Based on Formal Concept Analysis (RuZA 2015). CEUR Workshop Proceedings, vol. 1552, pp. 23–39 (2015)

37. Ignatov, D., Kaminskaya, A., Kuznetsov, S., Magizov, R.: A concept-based biclustering algorithm. In: Proceedings of the Eight International Conference on Intelligent Information Processing (IIP-8), pp. 140–143. MAKS Press, Moscow (2010) [in Russian]

38. Ignatov, D.I., Kuznetsov, S.O., Magizov, R.A., Zhukov, L.E.: From triconcepts to triclusters. In: Rough Sets, Fuzzy Sets, Data Mining and Granular Computing - Proceedings of the 13th International Conference, RSFDGrC 2011, pp. 257–264, Moscow, 25–27 June 2011. doi:10.1007/978-3-642-21881-1_41. http://dx.doi.org/10.1007/978-3-642-21881-1_41

39. Ignatov, D.I., Kuznetsov, S.O., Poelmans, J.: Concept-based biclustering for internet advertise-ment. In: ICDM Workshops, pp. 123–130. IEEE Computer Society, Brussels (2012)

40. Ignatov, D.I., Kuznetsov, S.O., Poelmans, J., Zhukov, L.E.: Can triconcepts become triclusters? Int. J. Gen. Syst. **42**(6), 572–593 (2013). doi:10.1080/03081079.2013.798899. http://dx.doi.org/10.1080/03081079.2013.798899

41. Ignatov, D.I., Mikhailova, M., Zakirova, A.Y., Malioukov, A.: Recommendation of ideas and antagonists for crowdsourcing platform witology. In: Information Retrieval - 8th Russian Summer School, RuSSIR 2014, pp. 276–296, Nizhniy, Novgorod, 18–22 August 2014, Revised Selected Papers (2014). doi:10.1007/978-3-319-25485-2_9. http://dx.doi.org/10.1007/978-3-319-25485-2_9

42. Ignatov, D.I., Nenova, E., Konstantinova, N., Konstantinov, A.V.: Boolean matrix factorisation for collaborative filtering: an fca-based approach. In: Artificial Intelligence: Methodology, Systems, and Applications - Proceedings of the 16th International Conference, AIMSA 2014, pp. 47–58, Varna, 11–13 September 2014. doi:10.1007/978-3-319-10554-3_5. http://dx.doi.org/10.1007/978-3-319-10554-3_5

43. Ignatov, D.I., Kaminskaya, A.Y., Konstantinova, N., Konstantinov, A.V.: Recommender system for crowdsourcing platform witology. In: 2014 IEEE/WIC/ACM International Joint Confer-ences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), vol. II, pp. 327–335, Warsaw, 11–14 August 2014. doi:10.1109/WI-IAT.2014.52. http://dx.doi.org/10.1109/WI-IAT.2014.52

44. Ignatov, D.I., Kaminskaya, A.Y., Konstantinova, N., Malioukov, A., Poelmans, J.: Fca-based recommender models and data analysis for crowdsourcing platform witology. In: Graph-Based Representation and Reasoning - Proceedings of the 21st International Conference on Conceptual Structures, ICCS 2014, pp. 287–292, Iaşi, 27–30 July 2014. doi:10.1007/978-3-319-08389-6_24. http://dx.doi.org/10.1007/978-3-319-08389-6_24

45. Ignatov, D.I., Gnatyshak, D.V., Kuznetsov, S.O., Mirkin, B.G.: Triadic formal concept analysis and triclustering: searching for optimal patterns. Mach. Learn. **101**(1–3), 271–302 (2015). doi:10.1007/s10994-015-5487-y. http://dx.doi.org/10.1007/s10994-015-5487-y

46. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: TRIAS–an algorithm for mining iceberg tri-lattices. In: Proceedings of the Sixth International Conference on Data Mining, ICDM '06, pp. 907–911. IEEE Computer Society, Washington, DC (2006). doi:http://dx.doi.org/10.1109/ICDM.2006.162. http://dx.doi.org/10.1109/ICDM.2006.162

47. Jelassi, M.N., Yahia, S.B., Nguifo, E.M.: Towards more targeted recommendations in folk-sonomies. Soc. Netw. Anal. Min. **5**(1), 68:1–68:18 (2015). doi:10.1007/s13278-015-0307-8. http://dx.doi.org/10.1007/s13278-015-0307-8

48. Jones, I., Tang, L., Liu, H.: Community discovery in multi-mode networks. In: Paliouras, G., Papadopoulos, S., Vogiatzis, D., Kompatsiaris, Y. (eds.) User Community Discovery, pp. 55–74. Springer, Cham (2015). doi:10.1007/978-3-319-23835-7_3. http://dx.doi.org/10.1007/978-3-319-23835-7_3

49. Kaytoue, M., Kuznetsov, S.O., Napoli, A., Duplessis, S.: Mining gene expression data with pattern structures in formal concept analysis. Inf. Sci. **181**(10), 1989–2001 (2011)

50. Kaytoue, M., Kuznetsov, S.O., Macko, J., Napoli, A.: Biclustering meets triadic concept analysis. Ann. Math. Artif. Intell. **70**, 55–79 (2014). doi:10.1007/s10472-013-9379-1. http://liris.cnrs.fr/publis/?id=6292

51. Krasnov, F., Vlasova, E., Yavorskiy, R.: Connectivity analysis of computer science centers based on scientific publications data for major Russian cities. In: Proceedings of the Second International Conference on Information Technology and Quantitative Management, ITQM 2014, pp. 892–899, National Research University Higher School of Economics (HSE), Moscow, 3–5 June 2014. doi:10.1016/j.procs.2014.05.341. http://dx.doi.org/10.1016/j.procs.2014.05.341

52. Krolak-Schwerdt, S., Orlik, P., Ganter, B.: Tripat: a model for analyzing three-mode binary data. In: Bock, H.H., Lenski, W., Richter, M. (eds.) Information Systems and Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 298–307. Springer, Berlin/Heidelberg (1994). doi:10.1007/978-3-642-46808-7_27. http://dx.doi.org/10.1007/978-3-642-46808-7_27

53. Kuznetsov, S.O.: Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. Nauchn. Tekh. Inf., Ser.2 (Autom. Doc. Math. Ling.) **12**, 21 – 29 (1990)

54. Kuznetsov, S.O.: On stability of a formal concept. Ann. Math. Artif. Intell. **49**(1–4), 101–115 (2007)

55. Kuznetsov, S.O., Ignatov, D.: Concept stability for constructing taxonomies of web-site users,. In: Obiedkov, S., Roth, C. (eds.) Proceedings of ICFCA 2007 Satellite Workshop on Social Network Analysis and Conceptual Structures: Exploring Opportunities, pp. 19–24. Clermont-Ferrand (2007)

56. Kuznetsov, S.O., Obiedkov, S.A., Roth, C.: Reducing the representation complexity of lattice-based taxonomies. In: Conceptual Structures: Knowledge Architectures for Smart Applications, Proceedings of the 15th International Conference on Conceptual Structures, ICCS 2007, pp. 241–254, Sheffield, 22–27 July 2007. doi:10.1007/978-3-540-73681-3_18. http://dx.doi.org/10.1007/978-3-540-73681-3_18

57. Latapy, M., Magnien, C., Vecchio, N.D.: Basic notions for the analysis of large two-mode networks. Soc. Netw. **30**(1), 31 – 48 (2008). doi:10.1016/j.socnet.2007.04.006. http://www.sciencedirect.com/science/article/pii/S0378873307000494

58. Lehmann, F., Wille, R.: A triadic approach to formal concept analysis. In: Proceedings of the Third International Conference on Conceptual Structures: Applications, Implementation and Theory, pp. 32–43. Springer, London (1995). http://dl.acm.org/citation.cfm?id=645488.656867

59. Lijffijt, J., Spyropoulou, E., Kang, B., Bie, T.D.: P-n-rminer: a generic framework for mining interesting structured relational patterns. Int. J. Data Sci. Anal. **1**(1), 61–76 (2016). doi:10.1007/s41060-016-0004-3. http://dx.doi.org/10.1007/s41060-016-0004-3

60. Liu, X., Murata, T.: Evaluating community structure in bipartite networks. In: Elmagarmid, A.K., Agrawal, D. (eds.) SocialCom/PASSAT, pp. 576–581. IEEE Computer Society, Washington, DC (2010)

61. Metzler, S., Miettinen, P.: Clustering boolean tensors. Data Min. Knowl. Discov. **29**(5), 1343–1373 (2015). doi:10.1007/s10618-015-0420-3. http://dx.doi.org/10.1007/s10618-015-0420-3

62. Miettinen, P.: Boolean tensor factorizations. In: 11th IEEE International Conference on Data Mining, ICDM 2011, pp. 447–456, Vancouver, BC, 11–14 December 2011. doi:10.1109/ICDM.2011.28. http://dx.doi.org/10.1109/ICDM.2011.28

63. Mirkin, B.: Mathematical Classification and Clustering. Kluwer, Dordrecht (1996)

64. Mirkin, B.G., Kramarenko, A.V.: Approximate bicluster and tricluster boxes in the analysis of binary data. In: Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, RSFDGrC'11, pp. 248–256. Springer, Berlin/Heidelberg (2011). http://dl.acm.org/citation.cfm?id=2026782.2026831

65. Mohr, J.W., Duquenne, V.: The Duality of Culture and Practice: Poverty Relief in New York City, 1888–1917. Theory Soc. **26**(2/3), 305–356 (1997). Special Double Issue on New Directions in Formalization and Historical Analysis

66. Murata, T.: Detecting communities from tripartite networks. In: Rappa, M., Jones, P., Freire, J., Chakrabarti, S. (eds.) WWW, pp. 1159–1160. ACM, New York (2010)

67. Newman, M.E.J.: Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. Phys. Rev. E **64**, 016,132 (2001). doi:10.1103/PhysRevE.64.016132. http://link.aps.org/doi/10.1103/PhysRevE.64.016132

68. Nussbaum, D., Pu, S., Sack, J., Uno, T., Zarrabi-Zadeh, H.: Finding maximum edge bicliques in convex bipartite graphs. Algorithmica **64**(2), 311–325 (2012). doi:10.1007/s00453-010-9486-x. http://dx.doi.org/10.1007/s00453-010-9486-x

69. Opsahl, T.: Triadic closure in two-mode networks: redefining the global and local clustering coefficients. Soc. Netw. **34** (2011). doi:10.1016/j.socnet.2011.07.001. http://www.sciencedirect.com/science/article/pii/S0378873311000360 (in press)

70. Padilha, V.A., Campello, R.J.G.B.: A systematic comparative evaluation of biclustering techniques. BMC Bioinf. **18**(1), 55:1–55:25 (2017). doi:10.1186/s12859-017-1487-1. http://dx.doi.org/10.1186/s12859-017-1487-1

71. Papalexakis, E.E., Faloutsos, C., Sidiropoulos, N.D.: Tensors for data mining and data fusion: Models, applications, and scalable algorithms. ACM Trans. Intell. Syst. Technol. **8**(2), 16:1–16:44 (2016). doi:10.1145/2915921. http://doi.acm.org/10.1145/2915921

72. Poelmans, J., Elzinga, P., Ignatov, D.I., Kuznetsov, S.O.: Semi-automated knowledge discovery: identifying and profiling human trafficking. Int. J. Gen. Syst. **41**(8), 774–804 (2012). doi:10.1080/03081079.2012.721662. http://dx.doi.org/10.1080/03081079.2012.721662

73. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: a survey on applications. Expert Syst. Appl. **40**(16), 6538–6560 (2013). doi:10.1016/j.eswa.2013.05.009. http://dx.doi.org/10.1016/j.eswa.2013.05.009

74. Poelmans, J., Kuznetsov, S.O., Ignatov, D.I., Dedene, G.: Formal concept analysis in knowledge processing: a survey on models and techniques. Expert Syst. Appl. **40**(16), 6601–6623 (2013). doi:10.1016/j.eswa.2013.05.007. http://dx.doi.org/10.1016/j.eswa.2013.05.007

75. Roth, C.: Generalized preferential attachment: towards realistic socio-semantic network models. In: ISWC 4th Intl Semantic Web Conference, Workshop on Semantic Network Analysis, Galway, CEUR-WS Series (ISSN 1613-0073), vol. 171, pp. 29–42 (2005)

76. Roth, C., Cointet, J.P.: Social and semantic coevolution in knowledge networks. Soc. Netw. **32**, 16–29 (2010)

77. Roth, C., Obiedkov, S.A., Kourie, D.G.: Towards concise representation for taxonomies of epistemic communities. In: Yahia, S.B., Nguifo, E.M., Belohlávek, R. (eds.) CLA. Lecture Notes in Computer Science, vol. 4923, pp. 240–255. Springer, Heidelberg (2006)

78. Roth, C., Obiedkov, S.A., Kourie, D.G.: On succinct representation of knowledge community taxonomies with formal concept analysis. Int. J. Found. Comput. Sci. **19**(2), 383–404 (2008). doi:10.1142/S0129054108005735. http://dx.doi.org/10.1142/S0129054108005735

79. Shin, K., Hooi, B., Faloutsos, C.: M-zoom: Fast dense-block detection in tensors with quality guarantees. In: Machine Learning and Knowledge Discovery in Databases - Proceedings of the European Conference, ECML PKDD 2016, Part I, pp. 264–280, Riva del Garda, 19–23 September 2016. doi:10.1007/978-3-319-46128-1_17. http://dx.doi.org/10.1007/978-3-319-46128-1_17

80. Spyropoulou, E., Bie, T.D., Boley, M.: Interesting pattern mining in multi-relational data. Data Min. Knowl. Discov. **28**(3), 808–849 (2014). doi:10.1007/s10618-013-0319-9. http://dx.doi.org/10.1007/s10618-013-0319-9

81. Tang, L., Liu, H., Zhang, J., Nazeri, Z.: Community evolution in dynamic multi-mode networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 677–685, Las Vegas, NV, 24–27 August 2008. doi:10.1145/1401890.1401972. http://doi.acm.org/10.1145/1401890.1401972

82. Tsymbal, A., Pechenizkiy, M., Cunningham, P.: Diversity in search strategies for ensemble feature selection. Inf. Fusion **6**(1), 83–98 (2005)

83. Vander Wal, T.: Folksonomy coinage and definition. URL http://vanderwal.net/folksonomy.html. http://vanderwal.net/folksonomy.html (2007). Accessed 12 Mar 2012

84. Veremyev, A., Prokopyev, O.A., Butenko, S., Pasiliao, E.L.: Exact mip-based approaches for finding maximum quasi-cliques and dense subgraphs. Comput. Optim. Appl. **64**(1), 177–214 (2016). doi:10.1007/s10589-015-9804-y. http://dx.doi.org/10.1007/s10589-015-9804-y

85. Voutsadakis, G.: Polyadic concept analysis. Order **19**(3), 295–304 (2002)

86. White, D.R.: Statistical entailments and the galois lattice. Soc. Netw. **18**(3), 201–215 (1996). doi:10.1016/0378-8733(95)00273-1. http://www.sciencedirect.com/science/article/pii/0378873395002731

87. Wille, R.: The basic theorem of triadic concept analysis. Order **12**, 149–158 (1995)

88. Wu, Z., Bu, Z., Cao, J., Zhuang, Y.: Discovering communities in multi-relational networks. In: Paliouras, G., Papadopoulos, S., Vogiatzis, D., Kompatsiaris, Y. (eds.) User Community Discovery, pp. 75–95. Springer, Cham (2015). doi:10.1007/978-3-319-23835-7_4. http://dx.doi.org/10.1007/978-3-319-23835-7_4

89. Yavorsky, R.: Research challenges of dynamic socio-semantic networks. In: Ignatov, D., Poelmans, J., Kuznetsov, S. (eds.) CEUR Workshop Proceedings, CDUD'11 - Concept Discovery in Unstructured Data, vol. 757, pp. 119–122 (2011)

90. Zachary, W.W.: An information flow model for conflict and fission in small groups. J. Anthropol. Res. **33**(4), 452–473 (1977). http://www.jstor.org/stable/3629752

91. Zakhlebin, I., Semenov, A., Tolmach, A., Nikolenko, S.I.: Detecting opinion polarisation on twitter by constructing pseudo-bimodal networks of mentions and retweets. In: Information Retrieval - 9th Russian Summer School, RuSSIR 2015, pp. 169–178, Saint Petersburg, 24–28 August 2015, Revised Selected Papers (2015). doi:10.1007/978-3-319-41718-9_10. http://dx.doi.org/10.1007/978-3-319-41718-9_10

92. Zhao, L., Zaki, M.J.: Tricluster: An effective algorithm for mining coherent clusters in 3d microarray data. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 694–705, Baltimore, Maryland, 14–16 June 2005. doi:10.1145/1066157.1066236. http://doi.acm.org/10.1145/1066157.1066236

93. Zhuk, R., Ignatov, D.I., Konstantinova, N.: Concept learning from triadic data. In: Proceedings of the Second International Conference on Information Technology and Quantitative Management, ITQM 2014, pp. 928–938, National Research University Higher School of Economics (HSE), Moscow, 3–5 June 2014. doi:10.1016/j.procs.2014.05.345. http://dx.doi.org/10.1016/j.procs.2014.05.345

94. Zudin, S., Gnatyshak, D.V., Ignatov, D.I.: Putting oac-triclustering on mapreduce. In: Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications, pp. 47–58, Clermont-Ferrand, 13–16 October 2015. http://ceur-ws.org/Vol-1466/paper04.pdf