Identifying exceptional (dis)agreement between groups

Adnene Belfodil, Sylvie Cazalens, Philippe Lamarre, Marc Plantevit

Identifying exceptional (dis)agreement between groups

**Adnene Belfodil · Sylvie Cazalens ·
Philippe Lamarre · Marc Plantevit**

Abstract Under the term behavioral data, we consider any type of data featuring individuals performing observable actions on entities. For instance, voting data depict parliamentarians who express their votes w.r.t. legislative procedures. In this work, we address the problem of discovering exceptional (dis)agreement patterns in such data, i.e., groups of individuals that exhibit an unexpected (dis)agreement under specific contexts compared to what is observed in overall terms. To tackle this problem, we design a generic approach, rooted in the Subgroup Discovery/Exceptional Model Mining framework, which enables the discovery of such patterns in two different ways. A branch-and-bound algorithm ensures an efficient exhaustive search of the underlying search space by leveraging closure operators and optimistic estimates on the interestingness measures. A second algorithm abandons the completeness by using a sampling paradigm which provides an alternative when an exhaustive search approach becomes unfeasible. To illustrate the usefulness of discovering exceptional (dis)agreement patterns, we report a comprehensive experimental study on four real-world datasets relevant to three different application domains: political analysis, rating data analysis and healthcare surveillance.

Keywords Supervised Pattern Mining, Subgroup Discovery, Exceptional Model Mining, Behavioral Data Analysis.

Adnene Belfodil, Sylvie Cazalens, Philippe Lamarre
INSA Lyon, CNRS, LIRIS UMR5205, F-69621 France

Marc Plantevit
Université Lyon 1, CNRS, LIRIS UMR5205, F-69622 France

E-mail: firstname.surname@liris.cnrs.fr

1 Introduction

Consider data describing the organization and votes of a parliamentary institution (e.g., European Parliament[1], US Congress[2]). Such datasets record the activity of each member in voting sessions held in the parliament, as well as information on the parliamentarians and the sessions. Table 1 provides an example. It reports the votes of Members of the European Parliament (MEPs) on legislative procedures. These procedures are described by attributes such as themes and dates. MEPs are characterized by their country, parliamentary group and age. The general trends are well known, and easy to check on these data with basic queries. For instance, the Franco-German axis is reflected by consensual votes between parliamentarians of both countries as well as the usual opposition between right wing and left wing. An analyst (e.g., a data journalist) is aware of these political positions and expects deeper insights. To this end, it is of major interest to discover groups of individuals that exhibit an unexpected mutual agreement (or disagreement) under specific conditions (contexts). For example, from Table 1, an exceptional (dis)agreement pattern is $p = (c = \langle \text{themes} = 7.30 \text{ Judicial Coop} \rangle, u_1 = \langle \text{country} = \text{France} \rangle, u_2 = \langle \text{country} = \text{Germany} \rangle)$, which reads: "in overall terms, while German and French parliamentarians are in agreement (comparing majority votes leads to 66%[3] of equal votes), an unexpected strong disagreement between the two groups is observed for *Judicial Cooperation related* legislative procedures (the respective majorities voted the same way only 33% of the time in the corresponding voting sessions, i.e. $\{e_3, e_5, e_6\}$)".

In this work, we aim to discover such exceptional (dis)agreement patterns not only in voting data but also in more generic data which involves individuals, entities and outcomes. In fact, voting data are particular instances of a

ide	themes	date
e_1	1.20 Citizen's rights	20/04/16
e_2	2.10 Free Movement of goods	16/05/16
e_3	1.20 Citizen's rights; 7.30 Judicial Coop	04/06/16
e_4	7 Security and Justice	11/06/16
e_5	7.30 Judicial Coop	03/07/16
e_6	7.30 Judicial Coop	29/07/16

(a) Entities (Voting sessions)

idi	country	group	age
i_1	France	S&D	26
i_2	France	PPE	30
i_3	Germany	S&D	40
i_4	Germany	ALDE	45

(b) Individuals (Parliamentarians)

idi	ide	outcome
i_1	e_1	For
i_1	e_2	Against
i_1	e_5	For
i_1	e_6	Against
i_2	e_1	For
i_2	e_3	Against
i_2	e_4	For
i_2	e_5	For
i_3	e_1	For
i_3	e_2	Against
i_3	e_3	For
i_3	e_5	Against
i_4	e_1	For
i_4	e_4	For
i_4	e_6	Against

(c) Outcomes

Table 1: Example of behavioral dataset - European Parliament Voting dataset

---

[1] http://parltrack.euwiki.org/, last access on October 15, 2019.

[2] https://voteview.com/data, last access on October 25, 2019.

[3] Since the majorities of $\langle u_1, u_2 \rangle$ voted respectively on $\{e_1, e_2, \underline{e_3}, e_4, \underline{e_5}, \underline{e_6}\}$ as follows: $\langle \text{For}, \text{For} \rangle, \langle \text{Against}, \text{Against} \rangle, \underline{\langle \text{Against}, \text{For} \rangle}, \langle \text{For}, \text{For} \rangle, \underline{\langle \text{For}, \text{Against} \rangle}, \underline{\langle \text{Against}, \text{Against} \rangle}$.

more general type of data that we call *behavioral data*. It regroups any type of data featuring individuals (e.g., parliamentarians, users in social networks, patients) performing observable actions (e.g., votes, ratings, intakes) on entities (e.g., procedures, movies, restaurants, medicines). From such datasets, we aim to discover exceptional (dis)agreement between groups of individuals on specific contexts. That is to say, an important difference between the groups' behaviors is observed compared to the usual context (i.e., the whole data). This could answer a large variety of questions. For instance, considering political data, an analyst may ask: *what are the controversial subjects in the European parliament in which groups or parliamentarians have divergent points of view?* In collaborative rating analysis, one may ask *what are the controversial items? And which groups are opposed?* In Healthcare surveillance, the analyst may want to know if some medicines are prescribed much more often for one group of individuals than another one.

The discovery of regions within the data that stand out with respect to a given target has been widely studied in data mining and machine learning communities under several names (Novak et al, 2009) (subgroup discovery (Wrobel, 1997), emerging patterns (Dong and Li, 1999), contrast sets (Bay and Pazzani, 2001)). Subgroup Discovery (SD) is known as the most generic one as it is agnostic of the data and the pattern domain. For instance, subgroups can be defined by a conjunction of conditions on symbolic (Lavrac et al, 2004) or numeric attributes (Grosskreutz and Rüping, 2009; Atzmüller and Puppe, 2006) as well as sequences (Grosskreutz et al, 2013). Furthermore, the single target can be discrete or numeric (Lemmerich et al, 2016). Exceptional Model Mining (EMM) (Leman et al, 2008) extends SD by offering the possibility to handle complex targets (Duivesteijn et al, 2016). However, no model in the EMM/SD framework makes it possible to investigate exceptional contextual (dis)agreement between groups. We made a first attempt to discover exceptional (dis)agreement patterns in (Belfodil et al, 2017). However, the model proposed required the specification of many non-intuitive parameters that may be the source of misleading interpretations. In this paper, we strive to provide a simpler and more generic framework to analyze behavioral data.

Figure 1 gives an overview of the approach we devise to discover exceptional (dis)agreement between groups. At a high level of description, five steps are necessary to discover interesting (dis)agreement patterns. First, two groups of individuals (u_1, u_2) are selected (1) . Then, their usual agreement on all their expressed outcomes is computed in step (2). All characterizable subsets of entities (contexts (c)) are then enumerated (3). For each selected subset, the agreement between the two groups is measured (4) and compared to their usual agreement (5) to evaluate to what extent the mutual agreement changes conditioned at a (dis)agreement pattern (c, u_1, u_2). Eventually, all pairs of groups (at least conceptually) are confronted. The discovery of exceptional (dis)agreement patterns requires to explore (simultaneously) both the search space associated to the individuals and the search space related to the entities. Moreover, behavioral datasets may contain several types of attributes (e.g., numerical and/or categorical attributes potentially organized by a hierarchy),
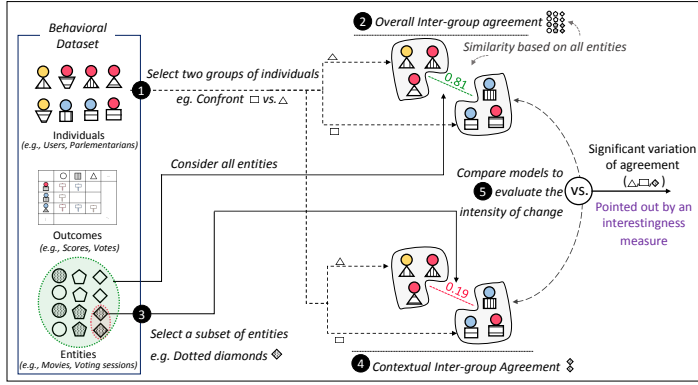
Fig. 1: Overview of the task of discovering exceptional (dis)agreement between groups

and outcomes. This requires efficient enumeration strategies. Last but not least, different measures to capture agreement may be considered depending on the application domain. Accordingly, the proposed method must be generic.

The main contributions of this paper are threefold:

Problem formulation. We define the novel problem of discovering exceptional (dis)agreement between groups of individuals when considering a particular subset of outcomes compared to the whole set of outcomes. Particular attention is paid to ensure a generic framework for behavioral data analysis.

Algorithms. We propose two algorithms to tackle the problem of discovering exceptional (dis)agreement patterns. DEBuNk[4] is a branch-and-bound algorithm that efficiently returns the complete set of patterns. It exploits both closure operators and optimistic estimates. Quick-DEBuNk is an algorithm that samples the space of (dis)agreement patterns in order to support instant discovery of patterns.

Evaluation. We report an extensive empirical study on four real-world datasets to demonstrate the efficiency of our algorithms as well as the interest of the discovered patterns. In particular, we report three case-studies from different application domains: political analysis, rating data analysis and healthcare surveillance to demonstrate that our approach is generic. The analysis of political votes, which is done in collaboration with journalists, is also available online[5].

The rest of this paper is organized as follows: the problem formulation is given in Section 2. We present the *agreement* measure and how it is integrated into an interestingness measure to capture changes of inter-group agreement in Section 3. Algorithm DEBuNk is presented in Section 4 while a pattern space sampling version, Quick-DEBuNk, is defined in Section 5. We report an empirical study in Section 6. Section 7 reviews the literature. We conclude and discuss future directions of research in Section 8.

---

[4] DEBuNk stands for Discovering Exceptional inter-group Behavior patterNs
[5] http://contentcheck.liris.cnrs.fr, last access on October 25, 2019.

2 Problem Definition

We are interested in discovering exceptional (dis)agreements among groups in *Behavioral Datasets*, which are defined as follows.

Definition 1 (*Behavioral Dataset*) A behavioral dataset $\langle G_I, G_E, O, o \rangle$ consists of (i) a collection of Individuals G_I, (ii) a collection of Entities G_E, (iii) a domain of possible Outcomes O, and (iv) a mapping function $o : G_I \times G_E \to O$ that gives the outcome of an individual i over an entity e.

In order to define appropriately the form of the sought patterns, we need first to characterize subgroups of data records in G_I and G_E. These two sets are collections of records defined over a set of descriptive attributes (Schema). We denote such *collection of records* by G, reintroducing the subscripts only in case of possible confusion. We assume $\mathcal{A} = (a_1, ..., a_m)$ to be the ordered list of attributes constituting the schema of G. Each attribute a_j has a domain of interpretation, noted $\mathrm{dom}(a_j)$, which corresponds to all its possible values. Attributes may be numerical or categorical potentially augmented with a taxonomy referred to by *Hierarchical Multi-Tag* (HMT) attributes (see section 4.2). For instance, in Table 1, parliamentarians, described by their country (categorical), their political group (categorical) and their age (numerical), decide on some voting sessions outlined by a date (numerical) and themes (HMT attribute). The attributes' domains define a description domain $\mathcal{D}$ which corresponds to the set of all possible descriptions characterizing *subgroups* in G.

Definition 2 (Description) Let G be a collection defined over the schema $\mathcal{A} = \{a_1, ..., a_m\}$, a **description** $d \in \mathcal{D}$ is a conjunction of **conditions** of the form $d = \langle r_1, ..., r_m \rangle$ where r_j depends on the type of the attribute a_j:

– If a_j is a categorical attribute then **condition** r_j is an equality test of the form $a_j = v$ with $v \in \mathrm{dom}(a_j) \cup \{*_j\}$. Note that $*_j$ is a symbol representing any value in $\mathrm{dom}(a_j)$. (i.e. every record in G satisfies $a_j = *_j$).
– If a_j is a numerical attribute then **condition** r_j is a membership test of the form $a_j \in [v..w]$ with $v, w \in \mathrm{dom}(a_j)$.

A description d characterizes a subgroup of records, also called *support*, in G denoted $G^d = \{g \in G \text{ s.t. } g \text{ satisfies } d\}$.

Descriptions are partially ordered in $\mathcal{D}$ by a *specialization relationship* defined as follows.

Definition 3 (Specialization between descriptions $\sqsubseteq$) Let d and d' be two descriptions from $\mathcal{D}$. d' is said to be a *specialization* of d, denoted $d \sqsubseteq d'$, iff $d' \Rightarrow d$, i.e. each condition r'_j in d' implies its corresponding condition r_j in d.

As a consequence of the previous definition, if $d \sqsubseteq d'$ then $G^{d'} \subseteq G^d$, since each record satisfying d' satisfies d.

Formally, the concept of *description* is used to describe both sets of individuals and sets of entities. Yet, for the ease of interpretation, we use two

different terms to name them: *group description* and *context*. For example, in Table 1, the *context* $c = \langle$ date $\in [05/06/16..30/07/16] \rangle$ identifies the set of entities $G_E^c = \{e_4, e_5, e_6\}$. Similarly, the *group description* $u = \langle$ group='S&D' $\rangle$ selects the set of individuals $G_I^u = \{i_1, i_3\}$.

Since we are interested in patterns highlighting exceptional (dis)agreement between two groups of individuals described by u_1 and u_2, in a context c compared to the overall context, the sought patterns are defined as follows:

Definition 4 ((Dis)Agreement Pattern) A *(dis)agreement pattern* is a triplet $p = (c, u_1, u_2)$ where $c \in \mathcal{D}_E$ is a *context* and $(u_1, u_2) \in \mathcal{D}_I^2$ is a pair of *group descriptions*.

The *extent* of a (dis)agreement pattern p is $ext(p) = (G_E^c, G_I^{u_1}, G_I^{u_2})$ with G_E^c the set of entities satisfying the conditions of context c, and $G_I^{u_1}$ (resp. $G_I^{u_2}$) the set of individuals supporting the description u_1 (resp. u_2). The set of all possible patterns is denoted as $\mathcal{P} = \mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$. Furthermore, as $\mathcal{P} = \mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$ is the product of three partially ordered collections, patterns of $\mathcal{P}$ are also partially ordered.

Definition 5 (Specialization between patterns $\sqsubseteq$) Let $p = (c, u_1, u_2)$ and $p' = (c', u_1', u_2')$ be two patterns from $\mathcal{P}$, p' is a *specialization of a pattern* p, denoted $p \sqsubseteq p'$, iff $c \sqsubseteq c'$, $u_1 \sqsubseteq u_1'$ and $u_2 \sqsubseteq u_2'$.

Notice that, if $p \sqsubseteq p'$ then $ext(p') \subseteq ext(p)$, that is $G_E^{c'} \subseteq G_E^c$ and $G_I^{u_1'} \subseteq G_I^{u_1}$ and $G_I^{u_2'} \subseteq G_I^{u_2}$. Moreover, two descriptions $d_1, d_2 \in \mathcal{D}$ are said to be equivalent if they characterize the same subset $S \subseteq G$, i.e. $G^{d_1} = G^{d_2}$. Similarly, two patterns $p, p' \in \mathcal{P}$ are equivalent if $ext(p) = ext(p')$.

To objectively evaluate how interesting a (dis)agreement pattern is, a *quality measure* φ is required (Duivesteijn et al, 2016).

Definition 6 (Quality measure) A quality measure is a function $\varphi : \mathcal{P} \to \mathbb{R}$ which assigns to each pattern $p = (c, u_1, u_2) \in \mathcal{P}$ a real number $\varphi(p) \in \mathbb{R}$.

A quality measure is designed to compare patterns: the quality of one will be compared to the quality of the others, most of the time to choose the best one. Consequently, it must be carefully designed with respect to what the algorithm is expected to produce. Our first objective is to identify particular parts of the data. This naturally leads to quality evaluation functions focusing only on the extent of the pattern. In particular, this means that, in this paper, syntax is not considered to assess the quality of a pattern.

Consequently, the quality measures we propose[6] are of the form: $\varphi(p) = \varphi'(ext(p))$. It follows that two patterns characterizing the same data, i.e. with the same extent, share the same quality measurement: $\forall p, p' \in \mathcal{P}$, if $ext(p) = ext(p')$ then $\varphi(p) = \varphi(p')$.

The user will be provided with a collection of patterns that captures exceptional (dis)agreements in a given behavioral dataset. A first intuitive idea

---

[6]Different quality measures are proposed in Sec. 3.

is to provide all patterns of high quality, i.e. with a quality greater than a user-defined threshold σ_φ. However, by construction of the quality measures, different patterns sharing the same extent will reach the same quality level, leading to multiple descriptions of the same parts of the data. Assuming that the user can be quickly bothered by such duplication, we propose to expose each interesting part of the data only once. More interestingly, the system should provide the user with the best generalizations only, i.e., patterns whose extent is not included in some other found ones. Additionally, some cardinality constraints can be added to avoid patterns of too small extent. Given two minimum support thresholds σ_E and σ_I, these constraints ensure, for a pattern $p = (c, u_1, u_2)$, that the size of the context extent (i.e. $|G_E^c| \geq \sigma_E$) and the size of both groups (i.e. $|G_I^{u_1}| \geq \sigma_I$ and $|G_I^{u_2}| \geq \sigma_I$) are large enough. Now, we introduce formally the core problem we tackle in this paper.

Problem Def. (*Discovering Exceptional (Dis)Agreement between Groups*).
   Given a behavioral dataset $\langle G_I, G_E, O, o \rangle$, a quality measure φ, a quality threshold σ_φ and a set of cardinality constraints $\mathcal{C}$, the problem is to find the pattern set $P \subseteq \mathcal{P}$ such that the following conditions hold:

1. (*Validity*) $\forall p \in P$, p is *valid* ; that is p *satisfies* $\mathcal{C}$ and $\varphi(p) \geq \sigma_\varphi$.
2. (*Maximality*) $\forall p \in P \; \forall q \in \mathcal{P}, \mathrm{ext}(q) = \mathrm{ext}(p) \Rightarrow q \sqsubseteq p$.
3. (*Completeness*) $\forall q \in \mathcal{P} \setminus P$, q is *valid* $\Rightarrow \exists p \in P$, $\mathrm{ext}(q) \subseteq \mathrm{ext}(p)$.
4. (*Generality*) $\forall (p, q) \in P^2$, $p \neq q \Rightarrow ext(p) \nsubseteq ext(q)$.

   Condition (1) ensures that the patterns in P are of high quality and satisfy the cardinality constraints. Condition (2) retains only one unique representative among patterns sharing the same extent: the maximal one w.r.t. $\sqsubseteq$. Such a pattern exists only if the specialization relation $\sqsubseteq$ over the pattern space induces a lattice structure (Ganter and Kuznetsov, 2001). It is commonly referred to as the *closed pattern* (Pasquier et al, 1999). We confine ourselves to such pattern spaces. Condition (3) ensures that each valid pattern in $\mathcal{P}$ has a representative in P covering it, while condition (4) ensures that only the most general patterns w.r.t. their extents are in P. In other words, the combination of conditions (3) and (4) guarantees that the solution P is minimal in terms of the number of patterns while having each valid pattern represented in the solution. Considering the generic definition of the quality measure discussed here, this problem extends the top-k problem addressed in (Belfodil et al, 2017) by introducing conditions (3) and (4). That is, for a sufficiently large k, the method formerly provided in (Belfodil et al, 2017) solves this problem only limited to the two first conditions providing, hence, a solution with a much larger number of redundant patterns.

3 Quality Measures and Inter-group Agreement Measurement

The previous section has already hinted at the fact that pattern interestingness is assessed with a quality measure φ (cf. Definition 6). Here, we propose such measures to capture the deviation between the *contextual inter-group agreement* and the *usual inter-group agreement*.

3.1 Quality Measures

For any pattern $p = (c, u_1, u_2) \in \mathcal{P}$, we denote by p^* the pattern $(*, u_1, u_2)$ which involves the same groups but all the entities. IAS (p^*) (resp. IAS (p)), which stands for *Inter-group Agreement Similarity*, represents the usual (resp. contextual) inter-group agreement observed between the two groups u_1, u_2. In order to discover interpretable patterns, we define two quality measures using IAS (p^*) and IAS (p).

- φ_{consent} assesses the strengthening of inter-group agreement in a context c:

$$\varphi_{\text{consent}}(p) = \max\left(\text{IAS}(p) - \text{IAS}(p^*), 0\right). \tag{1}$$

- φ_{dissent} assesses the weakening of inter-group agreement in a context c:

$$\varphi_{\text{dissent}}(p) = \max\left(\text{IAS}(p^*) - \text{IAS}(p), 0\right). \tag{2}$$

For instance, one can use φ_{consent} to answer: "*What are the contexts for which we observe more consensus between groups of individuals than usual?*".

3.2 Inter-group Agreement Similarity (IAS)

Several IAS measures can be designed according to the domain in which the data was measured (e.g., votes, ratings) and the user objectives. The evaluation of an IAS measure between two groups of individuals over a context requires the definition of two main operators: the *outcome aggregation operator* (θ) which computes an aggregated outcome of a group of individuals for a given entity, and a *similarity operator* (sim) which captures the similarity between two groups based on their aggregated outcomes over a single entity. These operators are defined in a generic way as follows.

Definition 7 (Outcome Aggregation Operator θ) An *aggregation operator* is a function $\theta : 2^{G_I} \times G_E \to \mathbb{D}$.

$\theta(u, e)$ simply transforms the outcomes of individuals u over an entity e into a value belonging to a domain $\mathbb{D}$ (e.g., $\mathbb{R}$, *categorical values*). From this, two elements of $\mathbb{D}$ have to be compared.

Definition 8 (Similarity between aggregated outcomes sim**)** A *similarity betweed aggregarted outcomes* is a function sim : $\mathbb{D} \times \mathbb{D} \to \mathbb{R}^+$.

sim(x, y) assigns a real positive value to any couple of aggregated outcomes (x,y).

Based on these operators, we properly define IAS which assigns to each pattern $p = (c, u_1, u_2) \in \mathcal{P}$ a value IAS$(p) \in \mathbb{R}^+$. This similarity evaluates how the two groups of individuals (u_1, u_2) behave similarly given their outcomes w.r.t. the context c. In the scope of this work, we confine ourselves to IAS measures that can be expressed as weighted averages. The next definition, though limiting, is generic enough to handle a wide range of behavioral data.

Definition 9 (Inter-group Agreement Similarity Measure IAS) Let w be a function associating a weight to each triple from $(G_E \times 2^{G_I} \times 2^{G_I})$. IAS: $\mathcal{P} \to \mathbb{R}^+$ associates to each pattern p the weighted average of the similarities of the aggregated outcomes for each entity e supporting the context c.

$$\text{IAS}\,(p = (c, u_1, u_2)) = \frac{\sum_{e \in G_E^c} w(e, G_I^{u_1}, G_I^{u_2}) \times \text{sim}(\theta(G_I^{u_1}, e), \theta(G_I^{u_2}, e))}{\sum_{e \in G_E^c} w(e, G_I^{u_1}, G_I^{u_2})}. \tag{3}$$

3.3 Examples of IAS Measures

By simply defining sim and θ, we present two instances of IAS measures that address two types of behavioral data with specific aims.

Behavioral Data With Numerical Outcomes: Rating datasets are a classic example of behavioral data with numerical outcomes. Such datasets describe users who express numerical ratings belonging to some interval $O = [\min, \max]$ (e.g., 1 to 5 stars) over reviewees (*e.g., movies, places*). A simple and adapted measure for aggregating individual ratings over one entity is the weighted mean $\theta_{\text{wavg}} : 2^{G_I} \times G_E \to [\min, \max]$.

$$\theta_{\text{wavg}}(G_I^u, e) = \frac{1}{\sum_{i \in G_I^u} w(i)} \sum_{i \in G_I^u} w(i) \times o(i, e). \tag{4}$$

Weight $w(i)$ corresponds to the importance of ratings given by each individual $i \in G_I$. Such weight may depend on the confidence of the individual or the size of the sample population if fine granularity ratings (*rating of each individual*) are missing. If no weights are given, θ_{wavg} computes a simple average over ratings, denoted θ_{avg}. To measure agreement between two aggregated ratings over a single entity, we define $\text{sim}_{\text{rating}} : [\min, \max] \times [\min, \max] \to [0, 1]$.

$$\text{sim}_{\text{rating}}(x, y) = 1 - \left(\frac{|x - y|}{\max - \min} \right). \tag{5}$$

Behavioral Data with Categorical Outcomes: A typical example of such datasets are *Roll Call Votes*[7] datasets where voting members cast categorical votes. The outcome domain (e.g., $O = \{\text{For}, \text{Against}, \text{Abstain}\}$) regroups all possible votes. To aggregate categorical outcomes we adapt majority vote to handle potential ties (non unique majority vote). Hence, $\theta_{\text{majority}} : 2^{G_I} \times G_E \to 2^O$ returns all the outcomes that received the majority of votes. Given $\#\text{votes}(z, G_I^u, e) = |\{(i, e) \text{ s.t. } i \in G_I^u \wedge o(i, e) = z\}|,$[8] we have:

---

[7] Roll-Call vote is a voting system where the vote of each member is recorded, such as http://www.europarl.europa.eu (EU parliament, last access on October 25, 2019) or https://voteview.com (US Congresses, last access on October 25, 2019).

[8] $o(i, e)$ returns the outcome (e.g., vote, rating) expressed by an individual i (e.g., parliamentarian, user) to an entity e (e.g., legislative procedure, movie) if given.

$$\theta_{\mathrm{majority}}(G_I^u, e) = \{v \in O \text{ s.t. } v = \underset{z \in O}{\mathrm{argmax}} \ \#\mathrm{votes}(z, G_I^u, e)\}. \tag{6}$$

We use a Jaccard index to assess the similarity between two majority votes x and y. Hence, $\mathrm{sim}_{\mathrm{voting}} : 2^O \times 2^O \to [0,1]$ is defined as follows.

$$\mathrm{sim}_{\mathrm{voting}}(x, y) = \frac{|x \cap y|}{|x \cup y|}. \tag{7}$$

3.4 Discussion

Above we introduced simple outcome aggregating functions and similarities that can be used to build an IAS measure to assess how similar two groups of individuals are. More sophisticated measures can be considered. For instance, in behavioral datasets with categorical outcomes, one can define an outcome aggregation measure which takes the empirical distribution of votes into account and then a similarity measure which builds up on a statistical distance (e.g., Kullback-Leibler divergence (Csisz et al, 1967; Johnson and Sinanovic, 2001)). Such measures can also be used on behavioral datasets which involves numerical outcomes, for instance the *Earth Mover Distance* measure was investigated in similarly structured datasets (rating datasets) in (Amer-Yahia et al, 2017). Several other measures can be relevant to analyze behavioral data with numerical outcomes depending on the aim of the study. In Section 6, "empirical study", we investigate another IAS measure which relies on a ratio to highlight discrepancies between the medicine consumption rates of two subpopulations.

4 A Branch-and-Bound Algorithm for Mining Relevant (Dis)Agreement Patterns

In this section, we present DEBuNk, a branch-and-bound algorithm that solves the problem outlined in Section 2. To this end, DEBuNk uses an exhaustive search strategy that exploits properties of closure operators to enumerate groups and contexts in a systematic manner and without redundancy. This allows the exploration of the complete search space, at least conceptually, by confronting all pair of groups $(u_1, u_2) \in \mathcal{D}_I \times \mathcal{D}_I$ in every possible context $c \in \mathcal{D}_E$. To speed up this exploration process while ensuring completeness, it is necessary to identify areas of the search space that cannot lead to the discovery of interesting patterns (i.e. patterns whose qualities are above the minimum quality threshold σ_φ). This is where optimistic estimates for the used quality measures are needed, since they provide an upper bound on the maximum possible quality measurement that a pattern can observe in each sub-search space. Algorithm 1 outlines the different ingredients of DEBuNk

---

Algorithm 1: PseudoDEBuNk($\langle G_I, G_E, O, o \rangle, \sigma_E, \sigma_I, \varphi, \sigma_\varphi$)

Inputs : $\langle G_I, G_E, O, o \rangle$ a behavioral dataset;
σ_E (resp. σ_I) minimum support threshold of a context (resp. group);
φ the quality measure; σ_φ quality threshold on the quality.
Output: P the set of exceptional (dis)agreement patterns.

1 $P \leftarrow \emptyset$
2 **foreach** *pair of groups* $(u_1, u_2) \in \mathcal{D}_I \times \mathcal{D}_I$ **do**
3   **foreach** *context* $c \in \mathcal{D}_E$ **do**
4   Evaluate the quality of the pattern $\varphi(c, u_1, u_2)$
5   **if** *the pattern is interesting (i.e.* $\varphi(c, u_1, u_2) \geq \sigma_\varphi$*)* **then**
6   Update P with (c, u_1, u_2)
7   **else**
8   Evaluate the optimistic estimate $oe(c, u_1, u_2)$
9   **if** $oe(c, u_1, u_2) < \sigma_\varphi$ **then**
10   Prune the sub-search space (backtrack).
11 *return P*

---

and explains how they are combined to compute the complete set of exceptional (dis)agreement patterns.

From Algorithm 1, we see that the main ingredients of DEBuNk basically consists of:

1. **Enumerating candidate patterns** (lines $2-3$): Efficiently enumerating candidate patterns in the whole search space $\mathcal{P} = \mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$ is a crucial part of DEBuNk. We give in Section 4.1 an overview of how the search space is structured and its main properties that one can leverage for an efficient enumeration of candidate patterns. We then pay particular attention to hierarchical multi-tag attributes in Section 4.2.
2. **Evaluating the quality of candidate patterns (lines** $4-6$**)** : Section 3 already discussed how the quality of patterns is assessed by capturing the deviation between the usual inter-group agreement and the contextual one using IAS measures (see Definition 9).
3. **Pruning unpromising areas of the search space (lines** $7-10$**)**: For a more efficient exploration, one need to safely discard, as early as possible, unpromising parts of the search space. To this end, we define optimistic estimates for the proposed quality measures in Section 4.3.

Once these ingredients are properly defined, we introduce, in Section 4.4, Algorithm DEBuNk which efficiently computes the complete set of exceptional (dis)agreement patterns.

4.1 Enumerating Candidate Subgroups

Exploring the space of (dis)agreement patterns from $\mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$ is equivalent to enumerating descriptions in $\mathcal{D}_E$ and $\mathcal{D}_I$ concurrently. Therefore, we explain how to enumerate descriptions for an arbitrary collection of records G associated to an arbitrary descriptions space $\mathcal{D}$.

Recall that several descriptions are considered to be equivalent when they characterize the same subset $S \subseteq G$. This may cause redundancy which we want to avoid. Focusing on *closed descriptions* enables to only retain in the results set the unique maximal element w.r.t. $\sqsubseteq$ – identified by $\text{clo}(d)$ – of each equivalence class $\dot{c} = \{d \in \mathcal{D} \text{ s.t. } G^c = G^d\}$ (Ganter and Wille, 1999).

We need to introduce some basic notions to define the operator $\text{clo}(d)$. Similarly to (Ganter and Kuznetsov, 2001), we define a mapping function $\delta_j(s)$ which provides a condition describing the value v of attribute a_j. This clearly depends on the type of a_j. $\delta_j(s)$ is of the form $(a_j = v)$ (resp. $(a_j \in [v..v])$) if a_j is a categorical (resp. numerical) attribute. Applying such mappings to all attributes leads to the tightest (maximal) description of a record $g \in G$: $\delta(s) = \langle \delta_1(g), ..., \delta_m(g) \rangle$. Given these mappings, the subgroup associated to a description d can be defined formally as such $G^d = \{g \in G \text{ s.t. } d \sqsubseteq \delta(g)\}$.

Extending the definition of δ to subsets $S \subseteq G$ characterizes a subset with the maximum common description of its elements:

$$\delta(S) = \bigwedge_{g \in S} \delta(g) = \langle \bigwedge_{g \in S} (\delta_1(g)), ..., \bigwedge_{g \in S} (\delta_m(g)) \rangle. \qquad (8)$$

$\delta(S)$ requires to compute a conjunction[9] $\bigwedge_j$ of conditions related to each attribute a_j. Using the aforementioned concepts, the closure operator can be defined as $\text{clo}(d) = \delta(G^d)$. This results from the fact that $\square^d$ and $\delta(\square)$ form a pair of Galois derivation operators between 2^G and $(\mathcal{D}, \sqsubseteq)$ providing as a consequence a closure operator $\delta(\square^d)$(Ganter and Kuznetsov, 2001).

Algorithm 2, called EnumCC (*Enumerate Closed Candidates*), first introduced in (Belfodil et al, 2017), enumerates once and only once all the *closed descriptions* that fulfill the support constraint $|G^c| \geq \sigma_G$ with σ_G a user defined support threshold similarly to the *ClosebyOne* (Kuznetsov and Obiedkov, 2002) algorithm and the *divide-and-conquer* (Boley et al, 2010b) algorithm.

Algorithm 2: EnumCC(G, d, σ_G, f, cnt)

Inputs : G is the collection of records depicted each by m descriptor attributes
$\qquad\qquad d$ a description from $\mathcal{D}$, σ_G a support threshold,
$\qquad\qquad f \in [1, m]$ a refinement flag, cnt a boolean
Output: yields all closed descriptions, i.e. $\text{clo}[\mathcal{D}] = \{\text{clo}(d) \text{ s.t. } d \in \mathcal{D}\}$

1 **if** $|G^d| \geq \sigma$ **then**
2 $\quad$ $\mathit{closure\_d} \leftarrow \delta(G^d)$
3 $\quad$ **if** $d \lessdot_f \mathit{closure\_d}$ **then**
4 $\quad\quad$ $\mathit{cnt\_c} \leftarrow \text{copy}(\mathit{cnt})$ $\qquad$ ▷ $\mathit{cnt\_c}$ value can be modified by a caller algorithm
5 $\quad\quad$ $\mathbf{\mathit{yield}}$ $(\mathit{closure\_d}, G^{\mathit{closure\_d}}, \mathit{cnt\_c})$ $\qquad$ ▷ yield the results and wait for next call
6 $\quad\quad$ **if** $\mathit{cnt\_c}$ **then**
7 $\quad\quad\quad$ **foreach** $j \in [f, m]$ **do**
8 $\quad\quad\quad\quad$ **foreach** $d' \in \eta_j(\mathit{closure\_d})$ **do**
9 $\quad\quad\quad\quad\quad$ **foreach** $(nc, G^{nc}, \mathit{cnt\_nc}) \in \text{EnumCC}(G, d', \sigma_G, j, \mathit{cnt\_c})$ **do**
10 $\quad\quad\quad\quad\quad\quad$ $\mathbf{\mathit{yield}}$ $(nc, G^{nc}, \mathit{cnt\_nc})$

---

[9]$(a_j \in [v_1..w_1]) \wedge_j (a_j \in [v_2..w_2]) = a_j \in [min(v_1, v_2)..max(w_1, w_2)]$ for numerical attributes. For categorical attributes $(a_j = v_1) \wedge_j (a_j = v_2) = v_1$ if $v_1 = v_2$ else $true_j$.

EnumCC, initially called with the following inputs $\text{EnumCC}(G, *, \sigma, 1, true)$, explores the search space $\mathcal{D}$ in a depth-first search manner starting from the most general description $\langle \text{true}_{a_1}, \ldots, \text{true}_{a_m} \rangle$, denoted $*$. It proceeds by *atomic refinements* to progress, step by step, toward more specific descriptions. Intuitively, an *atomic refinement* of a description d produces a more specific description d' by reinforcing the condition of one attribute only. Such descriptions are provided by a refinement operator denoted $\eta : \mathcal{D} \to 2^{\mathcal{D}}$ which returns $\eta(d)$ the collection of *neighbor* descriptions d', i.e. $\eta(d) = \{d' \in \mathcal{D} \text{ s.t. } d \sqsubset d' \wedge \nexists e \in \mathcal{D}, d \sqsubset e \sqsubset d'\}$. η can be built up by defining a refinement operator η_j for conditions related to each attribute $a_j \in \mathcal{A}$. For numerical attributes, an atomic refinement corresponds to a left-minimal (resp. right-minimal) change w.r.t. existing values of the attribute a_j in G^d, that is $a_j \in [\text{next}_{G^d}(\inf)..\sup]$ (resp. $a_j \in [\inf..\text{pred}_{G^d}(\sup)]$) on the interval bounds of the condition (Kaytoue et al, 2011). An empty resulting interval means that there is no possible refinement. Considering a categorical attribute, the atomic refinement of a condition true_{a_j} gives a condition of the form $a_j = v \in \text{dom}(a_j)$. Otherwise, a condition of the form $a_j = v$ does not admit any refinement.

Given a description d, EnumCC first computes G^d. If the support constraint is fulfilled (line 1), the closure of d is computed (line 2). Subsequently, a *canonicity test* between $\text{clo}(d)$ and d is assessed (line 3). It enables to determine whether a description after closure was already generated and to discard it, if appropriate. The canonicity test relies on an arbitrary order between attributes in $\mathcal{A}_G = \{a_1, a_2, ..., a_m\}$ indicating that, in the enumeration process, attributes' conditions are refined following this arbitrary order. Let $d = \langle r_1, ..., r_f, ..., r_m \rangle$ be a description resulting from the refinement of the fth condition of some preceding description, and $d' = \langle r'_1, ..., r'_f, ..., r'_m \rangle = \text{clo}(d)$ the closure of d. Following the arbitrary order, we expect for d', if it is the first time that it is encountered, that no condition before r'_f (i.e. $r'_1, ..., r'_{f-1}$) is refined; otherwise, $\text{clo}(d)$ was already generated after a refinement of preceding conditions and has thus to be discarded. This canonicity test is based on lectic order (cf. (Ganter and Wille, 1999, p.66-68)) between d and its closure d' denoted $d \lessdot_f d'$ which is defined as follows: $d \lessdot_f d' \leftrightarrow \forall i \in [1..f-1], r_i = r'_i \wedge r_f \lessdot r'_f$. The latter condition, $r_f \lessdot r'_f$, corresponds to an analogous canonicity test between conditions and makes sense for multi-valued attributes types only (e.g., HMT in Section 4.2). It does not need to be calculated for simple attributes (numerical, categorical). If the canonicity test is successful (line 3), $\text{clo}(d)$ is returned as a valid closed candidate (line 5). The algorithm then generates the neighbors by refining the attributes $\{a_f, ..., a_n\}$ from d on the condition that $\mathtt{cnt\_c}$ is not switched to False (lines 6-8). Flag f determines the index of the last attribute that was refined in the description d. Boolean $\mathtt{cnt\_c}$ can be modified externally by some caller algorithm to prune the search space, for instance, when using optimistic estimates. Eventually, a recursive call is done to explore the sub search space related to d (lines 9-10).

4.2 Hierarchical Multi-Tag Attribute (HMT)

Vote and review datasets often contain multi-tagged records whose tags are part of a hierarchical structure. For instance, voting sessions in the EU parliament can have multiple tags. For example, procedure *Gender mainstreaming in the work of the EU Parliament* is tagged by *4.10.04-Gender equality* and *8.40.01-EU Parliament*. Tag *4.10.04* is identified in a a hierarchy as a specialization of tag *4.10* that depicts *Social policy* and which is itself a specialization of tag *4* that covers all the sessions related to *Economic, social and territorial cohesion*. We formally define this type of attribute named HMT.

Definition 10 (HMT Attribute) Let $T = \{t_1, t_2 \ldots t_k\} \cup \{*\}$ be a set of values (also called tags), $<$ be a partial order over T inducing a tree structure $(T, <)$ whose root is '$*$'. $t_i < t_j$ denotes the fact that t_j is a descendant of t_i in T. In addition, the ascendants (resp. descendants) of a tag $t \in T$ is $\uparrow t = \{t' \in T \text{ s.t. } t' \leq t\}$ (resp. $\downarrow t = \{t' \in T \text{ s.t. } t' \geq t\}$).
A HMT attribute a_j takes its values in $dom(a_j) = 2^T$.

As an example, Fig. 2b describes G, a set of tag records defined by a unique attribute *tags*. Elements of *tags* are organized through the tree from Fig. 2a. We have $* < 1 < 1.20$ and $\uparrow 1.20 = \{1.20, 1, *\}$.

For a HMT attribute a_j, each record $g \in G$ is mapped by $\delta_j(g)$ to its corresponding tightest set of tags in $\text{dom}(a_j)$. If $\delta_j(g) = \{t_1, ..., t_n\}$, the record g is tagged *explicitly* by all the tags t_k for $k \in [1, n]$ and also *implicitly* by all their generalizations $\uparrow t_k$. Figure 2c illustrates this by reporting the flat representation of the collection of tagged records depicted in Figure 2b. It follows that a condition over a HMT attribute is defined as follows:

Definition 11 (Condition on a HMT attribute) (extends Definition 2) Let G be a collection defined over the schema $\mathcal{A} = \{a_1, ..., a_m\}$.

- If a_j is a HMT attribute then **condition** r_j is a superset test of the form $a_j \supseteq \chi$ with $\chi \in \text{dom}(a_j)$.

Accordingly, a HMT condition can be depicted by a rooted sub-tree of T and a record supports such a condition if it contains at least all tags of the sub-tree. It follows that, the partial order between two HMT conditions r, r' denoted $r \sqsubseteq r'$ (r' *is a specialization* r) is valid if the sub-tree r covers the sub-tree r'. i.e. , $r \sqsubseteq r'$ means $\forall t \in r, \exists t' \in r', t' \in \downarrow t$.

	$\ast$	1	1.10	1.20	2	2.10	3
g^1	$\times$	$\times$		$\times$	$\times$	$\times$	
g^2	$\times$	$\times$					$\times$
g^3	$\times$	$\times$	$\times$		$\times$		$\times$
g^4	$\times$				$\times$	$\times$	
g^5	$\times$	$\times$		$\times$			

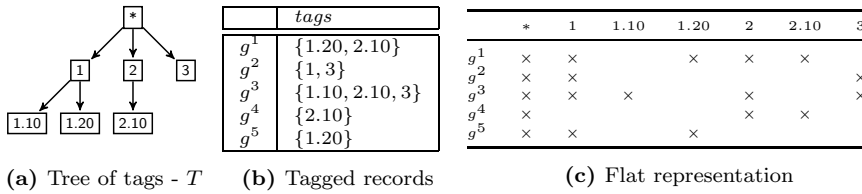**(a)** Tree of tags - T **(b)** Tagged records   **(c)** Flat representation

Fig. 2: A collection of records labeled each by a set of tags and its flat representation.
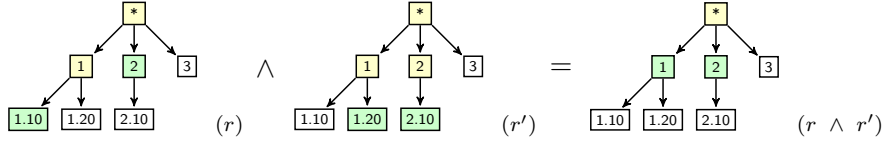
Fig. 3: Illustration of the conjunction operator $\wedge$ between two HMT descriptions.

Two ways are possible to take this attribute into account in the enumeration of descriptions from the complex search space aforementioned. One straightforward solution is to consider HMT attribute values as *itemsets* as depicted in the vector representation in Fig. 2c. However, such a solution ignores the taxonomy T implying the enumeration of *chain descriptions*. For instance, a chain description $\{1, 1.20.01\}$ is regarded as a different description than $\{1.20.01\}$. This stems from the fact that items are unrelated from the viewpoint of itemsets solution. As a consequence, a larger search space is explored while determining the same number of closed descriptions. To tackle this issue, we define a HMT description language.

Similarly to the aforementioned attributes, we define the conjunction operator $\wedge$ between two conditions which computes the *maximum common sub-tree* covering a set of conditions. Let $r = \{t_1, ..., t_n\}$ and $r' = \{t'_1, ..., t'_m\}$ be two HMT conditions, $r \wedge r' = \max(\cup_{t \in r} \uparrow t \cap \cup_{t' \in r'} \uparrow t')$ where $\max : 2^T \to 2^T$ maps a subset of tags $s \subseteq T$ to the leafs of the sub-tree induced by s: $\max(s) = \{t \in s \text{ s.t. } (\downarrow t \setminus \{t\}) \cap s = \emptyset\}$. Intuitively, $r \wedge r'$ is the set of the maximum explicit and implicit tags shared by the two descriptions (cf. Fig. 3).

Moreover, we define an atomic refinement operation which enables calculating neighbors of a HMT condition r. A condition r' is said to be a neighbor of r if: either only one tag of r is refined in r' or a new tag is added in r' that shares a parent with a tag in r or with one of its ascendants.

Finally, we define the lectic order between two conjunctions of tags $r = \{t_1, ..., t_n\}$ and its closure $r' = \{t'_1, ..., t'_n, ..., t'_m\}$ for the *canonicity test* to avoid the enumeration of already visited descriptions. Let r be generated after a refinement of the fth tag, the lectic order is defined as: $r \prec_f r' \Leftrightarrow \forall i \in [1..f-1] : t_i = t'_i \wedge t_f \prec t'_f$. The linear order $\prec$ between tags can be provided by a depth first search order on T. These concepts being defined, the mapping function δ can be extended easily to handle HMT among other attributes. Note that HMT supports itemsets. This can be done simply by considering a flat tree T with all the items as leaves. Hence, HMT can be seen as a generalization of itemsets, where implications between items are known.

4.3 Optimistic Estimates on Quality Measures

The enumeration of closed patterns enables a non-redundant traversal of the search space. For a more efficient exploration, we follow a branch and bound

scheme to discard unpromising parts of the search space. This requires the definition of safe pruning properties leveraging optimistic estimates on φ.

Let u_1, u_2 be two descriptions from $\mathcal{D}_I$ that respectively cover the two groups $G_I^{u_1}$, $G_I^{u_2}$. We consider optimistic estimates only with regards to the search space $\mathcal{D}_E$. We assume that u_1 and u_2 are instantiated a priori. Moreover, without loss of generality, we assume that the input domains of oe and φ are defined over both the pattern space $\mathcal{P}$ and $2^{G_E} \times 2^{G_I} \times 2^{G_I}$. This is possible, since the quality measure only depends on extents (cf. Section 2). Below, we give the definition of an optimistic estimate (Grosskreutz et al, 2008).

Definition 12 (Optimistic Estimate) An optimistic estimate oe for a given quality measure φ is a function such that:

$$\forall \text{ contexts } c, d \in \mathcal{D}_E . c \sqsubseteq d \Rightarrow \varphi(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{oe}(G_E^c, G_I^{u_1}, G_I^{u_2}).$$

Tight optimistic estimates, defined in (Grosskreutz et al, 2008), offer more pruning abilities than simple optimistic estimates.

Definition 13 (Tight Optimistic Estimate) An optimistic estimate oe is tight iff: $\forall c \in \mathcal{D}_E. \exists S \subseteq G_E^c : \text{oe}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \varphi(S, G_I^{u_1}, G_I^{u_2}).$[10]

4.3.1 Lower Bound and Upper Bound for the IAS *Measure*

The two quality measures φ_{consent} and φ_{dissent} rely on the IAS measure. Since u_1 and u_2 are considered to be instantiated for optimistic estimates, we can rewrite the IAS measure for a context $c \in \mathcal{D}_E$ and its extent G_E^c:

$$\text{IAS}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \frac{\sum_{e \in G_E^c} w_e \times \alpha(e)}{\sum_{e \in G_E^c} w_e} \; with \begin{cases} \alpha(e) = & \text{sim}(\theta(G_I^{u_1}, e), \theta(G_I^{u_2}, e)) \\ w_e = & w(e, G_I^{u_1}, G_I^{u_2}) \end{cases}.$$

$$(9)$$

We can now define a lower bound LB and an upper bound UB for the IAS measure based on the following operators that are defined for any context $c \in \mathcal{D}_E$ and for $n \in \mathbb{N}$:

– $\text{m}(G_E^c, n) = \text{Lowest}_{e \in G_E^c}(\{w_e \times \alpha(e) \text{ s. t. } e \in G_E^c\}, n)$ returns the set of the n distinct records e from G_E^c having the lowest values of $w_e \times \alpha(e)$.

– $\text{M}(G_E^c, n) = \text{Highest}_{e \in G_E^c}(\{w_e \times \alpha(e) \text{ s. t. } e \in G_E^c\}, n)$ returns the set of the n distinct records e from G_E^c having the highest values of $w_e \times \alpha(e)$.

– $\text{mw}(G_E^c, n) = \text{Lowest}_{e \in G_E^c}(\{w_e \text{ s. t. } e \in G_E^c\}, n)$ returns the set of the n distinct records e from G_E^c having the lowest values of w_e.

– $\text{Mw}(G_E^c, n) = \text{Highest}_{e \in G_E^c}(\{w_e \text{ s. t. } e \in G_E^c\}, n)$ returns the set of the n distinct records e from G_E^c having the highest values of w_e.

---

[10]Note that this does not require S to have a corresponding description in $\mathcal{D}_E$.

Proposition 1 (Lower bound LB **for** IAS**)** *we define function* LB *as:*

$$\mathrm{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \frac{\sum_{e \in \mathrm{m}(G_E^c, \sigma_E)} w_e \times \alpha(e)}{\sum_{e \in \mathrm{Mw}(G_E^c, \sigma_E)} w_e}. \tag{10}$$

For any context c (corresponding to a subgroup G_E^c), LB provides a lower bound for IAS w.r.t. contexts with σ_E a minimum context support threshold:

$$\forall c, d \in \mathcal{D}_E. \quad c \sqsubseteq d \Rightarrow \mathrm{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) \leq \mathrm{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}).$$

Proposition 2 (Upper bound UB **for** IAS**)** *we define function* UB *as:*

$$\mathrm{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \frac{\sum_{e \in \mathrm{M}(G_E^c, \sigma_E)} w_e \times \alpha(e)}{\sum_{e \in \mathrm{mw}(G_E^c, \sigma_E)} w_e}. \tag{11}$$

For any context c, UB provides an upper bound for IAS w.r.t. contexts. i.e.

$$\forall c, d \in \mathcal{D}_E. \quad c \sqsubseteq d \Rightarrow \mathrm{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \mathrm{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}).$$

Now that both the lower bound and the upper bound of IAS are defined w.r.t. contexts, we define the optimistic estimates corresponding to $\varphi_{\mathrm{consent}}$ and $\varphi_{\mathrm{dissent}}$. The proofs of the propositions are given in Appendix A.

4.3.2 Optimistic Estimates for Quality Measures

Proposition 3 (Optimistic estimate for $\varphi_{\mathrm{consent}}$ and $\varphi_{\mathrm{dissent}}$) oe$_{\mathrm{consent}}$ *(resp.* oe$_{\mathrm{dissent}}$*) is an **optimistic estimate** for $\varphi_{\mathrm{consent}}$ (resp. $\varphi_{\mathrm{dissent}}$) with:*

$$\mathrm{oe}_{\mathrm{consent}}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \max(\mathrm{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) - \mathrm{IAS}(G_E, G_I^{u_1}, G_I^{u_2}), 0). \tag{12}$$

$$\mathrm{oe}_{\mathrm{dissent}}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \max(\mathrm{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) - \mathrm{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}), 0). \tag{13}$$

The two defined optimistic estimates are tights if IAS is a simple average.

Proposition 4 *If $\forall(\{e\}, G_I^{u_1}, G_I^{u_2}) \subseteq G_E \times G_I \times G_I \ : \ w(e, G_I^{u_1}, G_I^{u_2}) = 1$,* oe$_{\mathrm{consent}}$ *(*oe$_{\mathrm{dissent}}$*) is a **tight optimistic estimate** for $\varphi_{\mathrm{consent}}$ ($\varphi_{\mathrm{dissent}}$).*

4.4 Algorithm DEBuNk

DEBuNk is a Branch-and-Bound algorithm which returns the complete set of patterns as specified in the problem definition (Section 2). DEBuNk follows the same guidelines given in Algorithm PseudoDEBuNk (see Algorithm 1). For an efficient exploration, it exploits the defined closure operator and optimistic estimates. Relying on algorithm EnumCC, DEBuNk starts by generating the couples of confronted groups of individuals that are large enough w.r.t. σ_I (lines 2-3). Then it computes the usual agreement observed between these two groups of individuals when considering all entities in G_E (line 4). Next, the

context search space is explored to generate valid contexts c (line 5). Subsequently, the optimistic estimate oe is evaluated and the context sub search space is pruned if oe is lower than σ_φ (lines 7-8). Otherwise, the contextual inter-group agreement is computed and the quality measure is calculated (lines 9-10). If the pattern quality exceeds σ_φ, two scenarios are possible. Either the current pattern set P already contains a more general pattern, or it does not. In the former case, the pattern is discarded. In the latter, the newly generated pattern is added to pattern set P while removing all previously generated patterns whose extents are included in the extent of p (lines 11-14). Since the current pattern quality exceeds the threshold and all the remaining patterns in the current context sub search space have their extents included in the extent of the current one, the sub search space is pruned (line 15). Eventually, in the case that the quality measure is symmetric w.r.t. u_1 and u_2 (i.e. $\forall u_1, u_2 \in \mathcal{D}_I^2, \; \varphi(c, u_1, u_2) = \varphi(c, u_2, u_1)$) there is no need to evaluate both qualities. It is therefore possible to prune the sub search space of the couple descriptions (u_1, u_2) whenever $u_1 = u_2$ (lines 16-17).

DEBuNk and DSC algorithm (Belfodil et al, 2017) differ on several levels. First, DEBuNk overcomes the lack of diversity of results provided by DSC which was designed to discover the top-k solutions. The proposed algorithm discards all patterns for which a generalization w.r.t. extents is already a solution. Second, DEBuNk handles a wider range of bounded quality measures (i.e. weighted mean IAS) in contrast to DSC algorithm. Finally, DSC requires the prior definition of an aggregation level which makes it difficult to use and interpret. DEBuNk overcomes this issue by reducing the number of input parameters and integrating relevancy checks between patterns. Hence, it requires less effort from the end-user both in terms of setting the parameters, and in terms of interpreting the quality of the patterns.

Algorithm 3: DEBuNk($\langle G_I, G_E, O, o \rangle, \sigma_E, \sigma_I, \varphi, \sigma_\varphi$)

Inputs : $\langle G_I, G_E, O, o \rangle$ a behavioral dataset;
$\qquad$ σ_E (resp. σ_I) minimum support threshold of a context (resp. group);
$\qquad$ φ the quality measure; σ_φ quality threshold on the quality.
Output: P the set of exceptional (dis)agreement patterns.

1 $P \leftarrow \{\}$
2 **foreach** $(u_1, G_I^{u_1}, cont\_u_1) \in \text{EnumCC}(G_I, *, \sigma_I, 0, \text{True})$ **do**
3 $\quad$ **foreach** $(u_2, G_I^{u_2}, cont\_u_2) \in \text{EnumCC}(G_I, *, \sigma_I, 0, \text{True})$ **do**
4 $\quad\quad$ $\text{IAS}_{\text{ref}} \leftarrow \text{IAS}(*, u_1, u_2)$
5 $\quad\quad$ **foreach** $(c, G_E, cont\_c) \in \text{EnumCC}(G_E^c, *, \sigma_E, 0, \text{True})$ **do**
6 $\quad\quad\quad$ **if** $\text{oe}_\varphi(c, u_1, u_2) < \sigma_\varphi$ **then**
7 $\quad\quad\quad\quad$ $cont\_c \leftarrow \text{False}$ $\qquad\qquad$ ▷ Prune the unpromising sub-search space under c
8 $\quad\quad\quad$ **else**
9 $\quad\quad\quad\quad$ $\text{IAS}_{\text{context}} \leftarrow \text{IAS}(c, u_1, u_2)$
10 $\quad\quad\quad\quad$ quality $\leftarrow \varphi(c, u_1, u_2)$ $\qquad\qquad$ ▷ computed using IAS_{ref} and $\text{IAS}_{\text{context}}$
11 $\quad\quad\quad\quad$ **if** quality $\geq \sigma_\varphi$ **then**
12 $\quad\quad\quad\quad\quad$ $p_{\text{new}} \leftarrow (c, u_1, u_2)$
13 $\quad\quad\quad\quad\quad$ **if** $\nexists p_{\text{old}} \in P \mid \text{ext}(p_{\text{new}}) \subseteq \text{ext}(p_{\text{old}})$ **then**
14 $\quad\quad\quad\quad\quad\quad$ $P \leftarrow (P \cup p_{\text{new}}) \setminus \{p_{\text{old}} \in P \mid \text{ext}(p_{\text{old}}) \subseteq \text{ext}(p_{\text{new}})\}$
15 $\quad\quad\quad\quad\quad$ $cont\_c \leftarrow \text{False}$ $\qquad\qquad$ ▷ Prune the sub search space
16 $\quad\quad$ **if** φ *is symmetric* **and** $u_1 = u_2$ **then**
17 $\quad\quad\quad$ *break* $\qquad\qquad$ ▷ Prune the sub search space
18 **return** P

5 Sampling (Dis)Agreement Patterns

DEBuNk returns the exact solutions according to the problem definition. Although it relies on the enumeration of closed descriptions and pruning techniques, such an exploration may take a considerable time depending on the size and the complexity (i.e. attributes types) of the behavioral data. To address this concern, we devise a (dis)agreement pattern sampling approach called Quick-DEBuNk. It enables instant mining of (dis)agreement patterns by yielding approximate solutions that improve over time. Quick-DEBuNk follows almost the same guidelines given in Algorithm PseudoDEBuNk (see Algorithm 1), while Quick-DEBuNk adopt a sampling strategy instead of enumerating candidate patterns (Ingredient 1), it uses the same ingredients to evaluate the quality of patterns (Ingredient 2) and to prune unpromising areas of the search space (Ingredient 3). To sample the candidate patterns, Quick-DEBuNk relies on two major steps as depicted in Fig. 4:

Frequency-Based Sampling (Step 1). A (dis)agreement pattern $p \in \mathcal{P}$ is drawn with a probability proportional to the size of its extent (i.e. $|\text{ext}(p = (c, u_1, u_2))| = |G_E^c| \times |G_I^{u_1}| \times |G_I^{u_2}|$). The key insight is to provide a greater chance to patterns supported by larger groups and contexts which are less likely to be discarded by more general ones generated by future iterations. This technique is inspired by the direct frequency-based sampling algorithm proposed in (Boley et al, 2011) which considers only Boolean attributed datasets. Here, this method is extended to handle more complex data with HMT, categorical and numerical attributes.

Random Walk on Contexts (Step 2). Starting from a context obtained in step 1, a random walk traverses the search tree corresponding to the contexts description space $\mathcal{D}_E$. We introduce some bias to fully take ad-
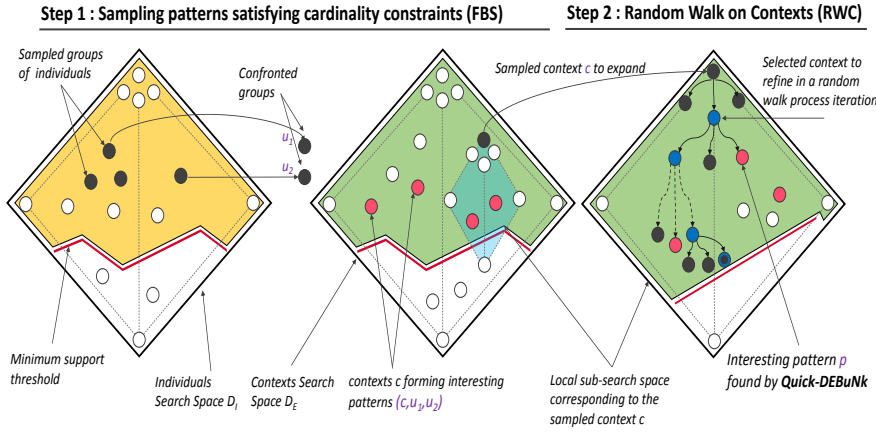

Fig. 4: Quick-DEBuNk approach in a nutshell

vantage of the devised quality measures and the optimistic estimates , this being done to reward high quality patterns by giving them more chance to be sampled by the algorithm.

5.1 Frequency-Based Sampling (Step 1)

To sample patterns of the form $p = (c, u_1, u_2)$, we aim to draw description c, respectively u_1 and u_2, from description space $\mathcal{D}_E$, respectively $\mathcal{D}_I$, with a probability proportional to their respective support size. To this end, we devise the algorithm FBS (Frequency-Based Sampling, cf. Algorithm 4). FBS generates a description d from an arbitrary description space $\mathcal{D}$ (i.e. $\mathcal{D}_E$, $\mathcal{D}_I$ in practice) with a probability proportional to the size of its support.

Proposition 5 *A description $d \in \mathcal{D}$ has a probability of being generated by* FBS *equal to* $\mathbb{P}(d) = \frac{G^d}{\sum_{d' \in \mathcal{D}} |G^{d'}|}$ *(see Appendix A for proofs).*

In the following, for any $d \in \mathcal{D}$, $\downarrow d$ denotes the set of all descriptions subsuming d, i.e: $\downarrow d = \{d' \in \mathcal{D} \text{ s.t. } d' \sqsubseteq d\}$. Since $\mathcal{D}^{11} = \mathcal{D}_1 \times \mathcal{D}_2 \times ... \times \mathcal{D}_m$, it follows that: $\downarrow d = \downarrow \langle r_1, r_2, ..., r_m \rangle = \downarrow r_1 \times \downarrow r_2 \times ... \times \downarrow r_m$, where $\downarrow r_j$ is the set of conditions less specific than (implied by) r_j in the conditions space $\mathcal{D}_j$.

To achieve the property defined in Proposition 5, FBS performs two steps following the reasoning of Boley et al (2011):

Step 1.1 (line 1): FBS starts by drawing a record g from G with a probability proportional to the number of descriptions $d \in \mathcal{D}$ covering g (i.e: $\downarrow \delta(g)$). To enable this, each record g is weighted by $w_g = | \downarrow \delta(g)|$. Knowing that $\delta(g) = \langle r_1^g, ..., r_m^g \rangle$, the weight $w_g = | \downarrow \delta(g)| = \prod_{j \in [1,m]} | \downarrow r_j^g|$ is the product of the numbers of conditions subsuming each r_j^g. The size of $| \downarrow r_j^g|$ depends on the type of the related attribute a_j[12].

Step 1.2 (line 2): Given g, the record returned from the first step and its corresponding description $\delta(g) = \langle r_1^g, ..., r_m^g \rangle$, FBS uniformly generates a description d from the set of descriptions covering g, that is $\downarrow \delta(g)$. This can be done by uniformly drawing[12] conditions r_j from $\downarrow r_j^g$, hence returning a description $d = \langle r_1, ..., r_m \rangle$. This comes from the fact that $\forall j \in [1, m]$: $\mathbb{P}(r_j) = \frac{1}{|\downarrow r_j^g|}$:

Algorithm 4: FBS(G)

Input: G a collection of records which may be G_E or G_I
Output: a description d from $\mathcal{D}$ which may be $\mathcal{D}_E$ or $\mathcal{D}_I$
1   ***draw*** $g \curvearrowleft (w_g = | \downarrow \delta(g)|)$ ***from*** G ▷ Step 1.1
2   ***draw*** $d \curvearrowleft \text{uniform}(\downarrow \delta(g))$ ▷ Step 1.2
3   ***return*** d

---

[11]Cartesian product of the m lattices related to the attributes' conditions spaces forms a lattice (Roman, 2008)

$$\mathbb{P}(d|g) = \prod_{j \in [1,m]} \mathbb{P}(r_j) = \frac{1}{\prod_{j \in [1,m]} |\downarrow r_j^g|} = \frac{1}{|\prod_{j \in [1,m]} \downarrow r_j^g|} = \frac{1}{|\downarrow \delta(g)|}. \qquad (14)$$

FBS algorithm makes it possible to generate valid patterns $p = (c, u_1, u_2)$ from the pattern space $\mathcal{P} = \mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$. This is achieved in the first step of Quick-DEBuNk (lines 3-6 in Algorithm 6) by sampling two group descriptions u_1, u_2 from $\mathcal{D}_I$ and a context c from $\mathcal{D}_E$ followed by assessing if the three descriptions satisfy the cardinalities constraints $\mathcal{C}$ (min. support thresholds).

Proposition 6 *Given $\mathcal{C}$, every valid pattern p is reachable by the first step of* Quick-DEBuNk: $\forall p \in \mathcal{P}$, *p satistifies $\mathcal{C} \Rightarrow \mathbb{P}(p) > 0$ (Proofs in Appendix A)*

Step 1 of Quick-DEBuNk does not favor the sampling of high quality patterns as it does not involve an exploitation phase. The random walk process on contexts used in Step 2 enables a smarter traversal of the search space while taking into account the devised quality measures and optimistic estimates.

5.2 RWC - Random Walk on Contexts (Step 2)

RWC (Algorithm 5) enumerates contexts of the search space $\mathcal{D}_E$ while considering closure and optimistic estimates. RWC takes as input two confronted groups u_1, u_2 for which it looks for relevant contexts following a random walk process starting from a context c. Mainly, RWC has two steps that are recursively executed until a terminal node is reached. RWC starts by generating all neighbors d of the current context c (line 2). Next, RWC assesses whether the size of the corresponding support G_E^c and the optimistic estimates respectively exceed the support threshold σ_E and the quality threshold σ_φ (line 3). If

---

Algorithm 5: RWC($\langle G_I, G_E, O, o \rangle, c, u_1, u_2, \sigma_E, \varphi, \sigma_\varphi$)

Inputs : $\langle G_I, G_E, O, o \rangle$ a behavioral dataset; c the current context;
(u_1, u_2) couple of confronted group descriptions;
σ_E threshold on support; φ the quality measure; σ_φ quality threshold.
Output: yield valid patterns (c, u_1, u_2)

1 NtE $\leftarrow \{\}$
2 **foreach** $d \in \eta(c)$ **do**
3   **if** $|G_E^d| \geq \sigma_E$ *and* $\text{oe}_\varphi(d, u_1, u_2) \geq \sigma_\varphi$ **then**
4        $closure\_d \leftarrow \delta(G_E^d)$
5   **if** $\varphi(d, u_1, u_2) \geq \sigma_\varphi$ **then**
6   *yield* $closure\_d$
7   **else**
8   NtE $\leftarrow$ NtE $\cup \{d\}$
9 **if** NtE $\neq \emptyset$ **then**
10   *draw* next $\curvearrowright \varphi(\text{next}, u_1, u_2)$ *from* NtE
11   **foreach** $c_{\text{next}} \in$ RWC($\langle G_I, G_E, O, o \rangle, \text{next}, \sigma_E, \varphi, \sigma_\varphi, u_1, u_2$) **do**
12   *yield* c_{next}

---

[13]For the sake of brevity and clarity, the reader is referred to (Belfodil et al, 2019) for technical details about the computation of $|\downarrow r_j^g|$ and the uniform sampling of conditions r_j from $\downarrow r_j^g$ for each attribute a_j.

appropriate, the closed description d is computed (line 4). The algorithm proceeds by evaluating the quality of the pattern (line 5). If the quality exceeds the threshold σ_φ, the pattern is valid and is hence returned (line 6). Otherwise, the pattern is added to NtE (*Neighbors to be Explored*) (line 8) as its related sub search space may contain interesting patterns (*i.e* $oe_\varphi(d, u_1, u_2) \geq \sigma_\varphi$). The second step of RWC consists of selecting a neighbor from NtE to be explored with a probability proportional to its quality (lines $10 - 12$). This process is recursively repeated until a terminal node is reached (i.e. $NtE = \emptyset$).

5.3 Quick-DEBuNk

Quick-DEBuNk (Algorithm 6) samples patterns from the full search space $\mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$. It is based on FBS and RWC. It takes as input the same parameters as DEBuNk in addition to a *timebudget*. It starts by generating a couple of closed group descriptions of individuals u_1, u_2 that fulfill the cardinality constraints (lines $3-5$) using FBS. Next, Quick-DEBuNk generates a context while only considering entities having a quality greater than the threshold σ_φ (line 6). The reason behind considering only $G_E^{\geq \sigma_\varphi}$ is clear: we have $\forall p \in \mathcal{P}$ p *satisfies* $\mathcal{C}$ and $\varphi(p) \geq \sigma_\varphi \Rightarrow \exists e \in G_E^c$ s.t. $\varphi(\{e\}, G_I^{u_1}, G_I^{u_2}) \geq \sigma_\varphi$ (since the quality measure is a weighted mean). If the context fulfills the cardinality constraint and its evaluated optimistic estimate is greater than the quality threshold (line 7), the algorithm then evaluates the quality of the sampled pattern (line 8). If this quality is greater than the threshold σ_φ, the pattern

---

Algorithm 6: Quick-DEBuNk($\langle G_I, G_E, O, o\rangle, \sigma_E, \sigma_I, \varphi, \sigma_\varphi, timebudget$)

> **Inputs** : $\langle G_I, G_E, O, o\rangle$ a behavioral dataset;
> σ_E (resp. σ_I) minimum support threshold of a context (resp. group);
> φ the quality measure; σ_φ threshold on the quality;
> *timebudget* the maximum amount of time given to the algorithm.
> **Output:** P the set of sampled exceptional (dis)agreement patterns

1   $P \leftarrow \{\}$
2   **while** *executionTime* < *timebudget* **do**
3     $u_1 \leftarrow clo(FBS(G_I))$
4     $u_2 \leftarrow clo(FBS(G_I))$
5   **if** $|G_I^{u_1}| \geq \sigma_I \wedge |G_I^{u_2}| \geq \sigma_I$ **then**
6      $c \leftarrow clo(FBS(G_E^{\geq \sigma_\varphi}))$ $\triangleright G_E^{\geq \sigma_\varphi} = \{e \in G_E$ s.t. $\varphi(\{e\}, G_I^{u_1}, G_I^{u_2}) \geq \sigma_\varphi\}$
7   **if** $|G_E^c| \geq \sigma_E \wedge oe_\varphi(c, u_1, u_2) \geq \sigma_\varphi$ **then**
8   **if** $\varphi(c, u_1, u_2) \geq \sigma_\varphi$ **then**
9        $p_{new} \leftarrow (c, u_1, u_2)$
10   **if** $\nexists p_{old} \in P$ s. t. $ext(p_{new}) \subseteq ext(p_{old})$ **then**
11         $P \leftarrow (P \cup p_{old}) \setminus \{p_{old} \in P$ s. t. $ext(p_{old}) \subseteq ext(p_{new})\}$
12   **else**
13   **foreach** $d \in RWC(\langle G_I, G_E, O, o\rangle, c, u_1, u_2, \sigma_E, \varphi, \sigma_\varphi)$ **do**
14         $p_{new} \leftarrow (d, u_1, u_2)$
15   **if** $\nexists p_{old} \in P$ s. t. $ext(p_{new}) \subseteq ext(p_{old})$ **then**
16          $P \leftarrow (P \cup p_{new}) \setminus \{p_{old} \in P$ s. t. $ext(p_{old}) \subseteq ext(p_{new})\}$
17   **else**
18   *break*
19   **if** *executionTime* $\geq$ *timebudget* **then**
20   *return* P
21   *return* P

is added to the resulting pattern set if and only if its extent is not included in an extent of an already found pattern (lines $9 - 11$). Otherwise, a random walk is launched starting from context c (line 13). This is done by relying on RWC. The algorithm continues by updating the resulting pattern set P with each pattern p_{new} returned by RWC, as long as there is no pattern p_{old} in P whose extent includes the extent of p_{new} (lines $14 - 16$). Otherwise, RWC is interrupted (line 18). The process is repeated as long as the time budget allows.

6 Empirical Study

In this section, we report on both quantitative and qualitative experiments over the implemented algorithms. For reproducibility purposes, source code (in Python) and data are made available on a companion page[13].

6.1 Aims and Datasets

The experiments aim to answer the following questions:

- Do the algorithms provide interpretable patterns?
- How effective is DEBuNk compared to state-of-the-art algorithms?
- Are the closure operators and optimistic estimate based pruning, efficient?
- How effective is HMT closed description enumeration?
- Does DEBuNk scale w.r.t. different parameters?
- How effective is Quick-DEBuNk at sampling patterns?

Most of the experiments were carried out on four real-world behavioral datasets whose main characteristics are given in Table 2. Each dataset involves entities and individuals described by an HMT (H) attribute together with categorical(C) and numerical(N) ones.

EPD8[14] features voting information of the eighth European Parliament about the 958 members who were elected in 2014 or after. The dataset records $2.7M$ tuples indicating the outcome (For, Against, Abstain) of a member voting during one of the 4161 sessions. Each session is described by its themes (H), a voting date (N) and the organizing committee (C). Individuals are described by a national party (C), a political group (C), an age group (C), a country(C) and additional information about countries (date of accession to the European Union (N) and currency (C)). To analyze (dis)agreement patterns in this dataset, we consider $\mathrm{IAS}_{\mathrm{voting}}$ which is defined by using $\theta_{\mathrm{majority}}$ and $\mathrm{sim}_{\mathrm{voting}}$.

Movielens[15] is a movie review dataset (Harper and Konstan, 2016) consisting of $100K$ ratings (ranging from 1 to 5) expressed by 943 users on

---

1681 movies. A movie is characterized by its genres (H) and a release date (N), while individuals are described with demographic information such as age group (C), gender (C) and occupation (C). To handle the numerical outcomes, we use the measure IAS_{rating} which relies on θ_{wavg} and sim_{rating}.

Yelp[16] is a social network dataset featuring individuals who rate (scores ranging from 1 to 5) places (stores, restaurants, clinics) characterized by some categories (H) and a state (C). The dataset originally contains 1M users. We preprocessed the dataset to constitute 18 groups of individuals based on the size of their friends network (C), their seniority (C) in the platform and their account type (e.g., elites or not) (C). We also use IAS_{rating} measure in this dataset.

Openmedic[17] is a drug consumption monitoring dataset that has been recently made available by *Ameli*[18]. This dataset inventories the number of drug boxes (described by their Anatomical Therapeutic Chemical (ATC) Classification[19](H)) yearly administered to individuals (from 2014 to 2016). Individuals are described with demographic information such as age (C), gender (C) and region (C). We further discuss an adapted IAS measure.

Comparing the size and the complexity of these datasets is difficult because of the heterogeneity of the attributes. In particular, the hierarchies of the HMT

		Entities	Individuals	Outcomes
EPD8	Size (Nb. records)	4161	958	2.7M
	attribute types	$1H + 1N + 1C$	$1N + 5C$	
	size after scaling	$347 + 26 + 40 = 413$	$16 + 285 = 301$	
	avg scaling per record	20.44	14	
Movielens	Size (Nb. records)	1681	943	100K
	attribute types	$1H + 1N$	$3C$	
	size after scaling	$20 + 144 = 164$	$4 + 2 + 21 = 27$	
	avg scaling per record	75.72	3	
Yelp	Size (Nb. records)	127 000	18	750K
	attribute types	$1H + 1C$	$3C$	
	size after scaling	$1175 + 29 = 1204$	$3 + 2 + 3 = 8$	
	avg scaling per record	5.77	3	
Openmedic	Size (Nb. records)	12 221	78	500K
	attribute types	$1H$	$3C$	
	size after scaling	14 094	$2 + 13 + 3 = 18$	
	avg scaling per record	7	3	

Table 2: Behavioral datasets characteristics before and after scaling.

---

[16]https://www.yelp.com/dataset/challenge, last access on April 25, 2017.

[17]http://open-data-assurance-maladie.ameli.fr/, last access on November 16, 2017.

[18]*Ameli* - France National Health Insurance and Social Security Organization

[19]The Anatomical Therapeutic Chemical classification system classifies therapeutic drugs according to the organ or system on which they act and their chemical, pharmaco- logical and therapeutic properties – https://www.whocc.no/atc/structure_and_principles/, last access on October 18, 2019.

attributes are very different, as well as the range of the numerical ones. To enable a fair comparison, we employ a conceptual scaling (Ganter and Wille, 1999). The attributes are "projected" on a set of items by transforming each one to a Boolean representation. Each possible value of a categorical attribute provides a single item (e.g., *gender* gives *male*, *female* and *unknown*). The items corresponding to an HMT attribute are all the nodes of the tag tree (T). Each numerical attribute is transformed to an itemset via *interordinal scaling* (Kaytoue et al, 2011). To a given set of values $[v_1, v_2, ...v_n]$, we associate $2n$ items $\{\leq v_1, \leq v_2, ... \leq v_n, \geq v_1, \geq v_2, ... \geq v_n\}$. Table 2 illustrates this step, while Table 3 shows the obtained comparable characteristics.

Dataset	Transactions	Items	AverageSize
EPD8	1 727 032 585	1 015	34.48
Movielens	16 807 109	218	79.37
Yelp	5 860 354	1 220	9.00
Openmedic	28 512 418	14 130	10.00

Table 3: Characteristics of the datasets considered as plain collections of itemsets records - the plain collections correspond to $G_E \times G_I \times G_I$ while considering only pairable individuals (i.e., the cartesian product contains a record (e, i_1, i_2) only if both individuals expressed an outcome on the entity e, that is $o(i_1, e)$ and $o(i_2, e)$ are given).

6.2 Qualitative Results

First, we focus on illustrating patterns discovered by DEBuNk. To this end, we report three real world case studies: (i) In collaborative rating platforms (Yelp, Movielens), we study the affinities between groups of users with regard to their expressed ratings. (ii) In a voting system (European Parliament Dataset), we show how the voting behavior of parlementarians can provide interesting insights about the cohesion and the polarization between groups of parliamentarians in different contexts. Such information can be valuable for journalists and political analysts. (iii) We give example patterns reporting substantial differences in medicine consumption behavior between groups. Such results can be leveraged by epidemiologists to study comparative prevalence of sicknesses among subpopulations.

6.2.1 Study of Collaborative Rating Data

Table 4 describes some patterns returned by DEBuNk on the Movielens dataset when looking for contexts that lead to a disagreement between groups of individuals labeled by their professional occupations. The first pattern describes that, while students and health professionals agree 74% of the time, they tend to disagree for horror and comedy-like movies released between 1986 and 1994

(e.g., *Evil Dead II*, *Braindead*). Figure 5 illustrates the usual and the contextual rating distribution of each groups. We observe from this rating distributions, that the students like the movies highlighted by the pattern, whereas the healthcare professionals dislike them.
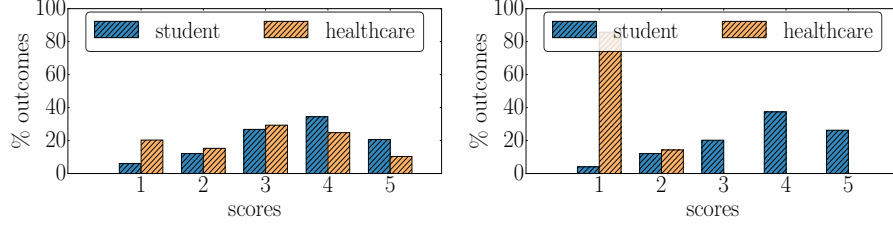

Fig. 5: Pattern 1 Illustration – distribution of ratings of individuals constituting the group of students versus distribution of ratings of individuals constituting the group of health professionals. Left figure corresponds to the usual distribution observed over all movies. Right figure corresponds to the contextual distribution observed over the context of pattern 1 from Table 4.

In Table 5, we present some results provided by DEBuNk over Yelp dataset. The groups of individuals are labeled by the size of their friend network and

	(c, u_1, u_2)	$|G_E^c|$	$|G_I^{u_1}|$	$|G_I^{u_2}|$	$o(i,e)$	φ_{dissent}
1	Student vs. Healthcare in ['11 Horror', '5 Comedy'] [1986, 1994]	6	196	16	106	$0.42 = 0.74 - 0.33$
2	Student vs. Healthcare in ['5 Comedy'] [1991, 1991]	5	196	16	40	$0.41 = 0.74 - 0.33$
3	Healthcare vs. Artist in ['5 Comedy', '8 Drama'] [1987, 1993]	5	16	28	28	$0.42 = 0.73 - 0.3$

Table 4: Top-3 w.r.t. number of expressed outcomes ($o(i,e)$ *column*) of disagreement patterns discovered on Movielens ($|\mathcal{A}_E| = 2$, $|\mathcal{A}_I| = 1$, $\sigma_E = 5$, $\sigma_I = 10$ and $\sigma_\varphi = 0.4$).

	(c, u_1, u_2)	$|G_E^c|$	$|G_I^{u_1}|$	$|G_I^{u_2}|$	$o(i,e)$	φ_{dissent}
1	(Newcomer,*) vs. (Middler,*) in ['03 Automotive', '14.22 Electronics Repair', '22.06 Battery Stores', '22.21 Electronics'] *	10	6	6	43	$0.4 = 0.8 - 0.4$
2	(Senior, Medium) vs. (Middler, Large) in ['10.55.21 Internal Medicine'] Nevada	15	2	2	39	$0.43 = 0.81 - 0.38$
3	(Newcomer, Medium) vs. (Middler, Large) ['11.59.01 Apartments', '11.59.18 University Housing'] Arizona	14	2	2	30	$0.4 = 0.78 - 0.38$

Table 5: Top-3 w.r.t. number of expressed outcomes ($o(i,e)$ *column*) of disagreement patterns discovered on the Yelp dataset ($|\mathcal{A}_E| = 2$, $|\mathcal{A}_I| = 2$, $\sigma_E = 10$, $\sigma_I = 1$ and $\sigma_\varphi = 0.4$).

their seniority in the Yelp platform. Notice that additional demographic data about users are missing. This prevents DEBuNk from obtaining concrete results similar to the ones obtained in Movielens. The resulting patterns highlight the places for which groups of individuals have divergent opinions. For example, pattern 2 states that Senior Yelp users (registered in Yelp before 2010) having a friend network of medium size (less than 100 friends) disagree with users registered in Yelp before 2015 having a large friend network (more than 100 friends) on Internal Medicines Clinics in Nevada (e.g., Las Vegas Urgent Care), contrary to the usual, where these two groups roughly share the same opinions about places (81% of the time).

6.2.2 Analysis of the Voting Behavior in the European Parliament Dataset

Table 6 reports patterns obtained by DEBuNk where the aim is to find contexts (subsets of voting sessions) that lead groups of parliamentarians (labeled by their countries and their corresponding date of accession to the European Union) to strong disagreement compared to the usual observed agreement. Note that we choose carefully $\sigma_E \geq 25$ to reach subgroups of the third level of the themes hierarchy which on average contain $\backsim 25$ voting sessions. Such analysis can be valuable to political analysts and journalists as it enables to uncover subjects/thematics of votes on which countries have divergent opinions. For instance, the second pattern in Table 6 illustrated in Figure 6, states that the voting sessions about theme 4.15.05 (Industrial Restructuring, job losses, EGF, e.g., Mobilization of the European Globalization Adjustment Fund: redundancies in aircraft repair and installation services in Ireland) lead to strong disagreements between parliamentarians from the United Kingdom and their peers. In Figure 6, we provide a visualization of this pattern through a similarity matrix where each cell represents the similarity between two countries. This can be seen as a post-processing step where the end-user chooses to augment the pattern with more related information (similarities between other countries). Such visualization brings more context to the pattern. While the second pattern conveys that British parliamentarians are in strong disagreement with

	(c, u_1, u_2)	$|G_E^c|$	$|G_I^{u_1}|$	$|G_I^{u_2}|$	$o(i,e)$	φ_{dissent}
1	([1973, 1973] United Kingdom) vs. (*,*) ['4 Economic, social & territorial cohesion', '8.70 Budget of the Union']	47	88	958	30255	0.54 = 0.68 − 0.14
2	([1973, 1973] United Kingdom) vs. (*,*) ['4.15.05 Industrial restructuring, job losses, Globalization Adjustment Fund']	47	88	958	30250	0.54 = 0.68 − 0.14
3	([1958, 1958] Italy) vs. ([1981, 2013] *) ['3.40 Industrial policy', '6.20.02 Export /import control, trade defence']	79	99	433	29501	0.51 = 0.87 − 0.35

Table 6: Top-3 w.r.t. number of expressed outcomes ($o(i,e)$ *column*) of Relevant (dis)agreement patterns discovered on EPD8 ($|\mathcal{A}_E| = 1$, $|\mathcal{A}_I| = 2$, $\sigma_E = 25$, $\sigma_I = 1$ and $\sigma_\varphi = 0.5$ using φ_{dissent}).

their peers, the visualization goes beyond by reporting that all other countries formed a coalition against the voting decision of British parlementarians. The algorithms elaborated in this work also allow to discover patterns exhibiting consensual subjects, thanks to the quality measure φ_{consent}. Additional experiments are reported in the technical report (Belfodil et al, 2019) illustrating such patterns. Moreover, as part of a collaboration with political journalists, we provide an online tool[20], dubbed ANCORE that makes it possible to analyze European parliament voting sessions.

Fig. 6: Illustration of pattern 2 reported in Table 6. The left matrix depicts the agreement observed in general between countries when considering all voting sessions. The right matrix corresponds to the inter country agreement for the context of pattern 2.

6.2.3 Illnesses Prevalence on the Basis of Medicine Consumption

One interesting analysis task to be carried out on the *Openmedic* dataset is to look for subgroups of drugs where the ratio of intakes between two groups of individuals is substantially different than the one usually observed. For instance, we find that while *Females* takes $1.32\times$ more drugs than *Males* in overall terms, this ratio increases up to $5\times$ when considering drugs prescribed for *Hyperthyroidism* (see Pattern 3 in Table 7). These results are similar to those reported in an epidemiology study by Wang et Al. in (Wang and Crapo, 1997). Such task can provide some insight regarding illness prevalence for particular groups of individuals. In the behavioral dataset *Openmedic*, the outcomes expressed by individuals are depicted by numerical values reporting the count of medicine boxes. As we are interested in characterizing the agreement by the consumption ratio, we instantiate IAS as follows:

$$\text{IAS}_{ratio}(c, u_1, u_2) = \frac{\sum_{e \in G_E^c} \theta_{avg}(G_I^{u_1}, e)}{\sum_{e \in G_E^c} \theta_{avg}(G_I^{u_2}, e)}. \tag{15}$$

This ratio falls under the definition of IAS considered in Definition 9 as it can be expressed as a weighted average:

---

[20] http://contentcheck.liris.cnrs.fr, last access October 25, 2019.

$$\text{IAS}_{ratio}(c, u_1, u_2) = \frac{\sum\limits_{e \in G_E^c} w(e, G_I^{u_1}, G_I^{u_2}) \times sim_{ratio}(\theta_{avg}(G_I^{u_1}, e), \theta_{avg}(G_I^{u_2}, e))}{\sum_{e \in G_E^c} w(e, G_I^{u_1}, G_I^{u_2})} \quad (16)$$

$$\text{with } w(e, G_I^{u_1}, G_I^{u_2}) = \theta_{avg}(G_I^{u_2}, e) \text{ and } sim_{ratio}(x, y) = \frac{x}{y} \ .$$

In order to provide interpretable patterns according to the aim of the study, we define an adapted quality measure φ_{ratio} as:

$$\varphi_{ratio}(p) = \frac{\text{IAS}_{ratio}(p)}{\text{IAS}_{ratio}(p^*)} \quad \text{with } p = (c, u_1, u_2) \in \mathcal{P} \text{ and } p^* = (*, u_1, u_2). \quad (17)$$

Drug boxes are labeled by tags in the ATC classification system. We aim at leveraging the medical consumption differences between groups of individuals to investigate the comparative prevalence[21] of illnesses between gender groups. Table 7 shows some patterns discovered by DEBuNk on Openmedic. Note that we carefully choose $\sigma_E \geq 10$ to reach subgroups of drugs of the fifth level of ATC tree which on average contains $\backsim$ 10 drugs. Pattern 4 states that, for drugs prescribed for $Gout$[22], men consume $3\times$ more drugs than women, whereas in overall terms, men consume $0.76\times$ less drugs than women. Similar results were reported by an epidemiology study of $Gout$ in (Roddy and Doherty, 2010) giving an incidence of gout per 1,000 person-years of 1.4 in women and 4.0 in men. Patterns 3 and 4, depicted in Figure 7, report details on the differences between the two gender groups in terms of population size and number of drugs consumed both in overall and in the context highlighted by the pattern.

	(c, u_1, u_2)	$|G_E^c|$	$|G_I^{u_1}|$	$|G_I^{u_2}|$	$o(i, e)$	φ_{ratio}
1	Men vs. Women in N07B - Drugs used in addictive disorders	138	39	39	4195	$4.59 = \frac{3.48}{0.76}$
2	Women vs. Men in A12A - Calcium	54	39	39	3174	$3.96 = \frac{5.21}{1.32}$
3	Women vs. Men in H03 - Thyroid Therapy	31	39	39	1981	$3.89 = \frac{5.13}{1.32}$
4	Men vs. Women in M04A - Antigout preparations	42	39	39	1940	$3.91 = \frac{2.97}{0.76}$

Table 7: Top-4 w.r.t. the number of expressed outcomes on Openmedic considering by default the full dataset, $|\mathcal{A}_E| = 1$, $|\mathcal{A}_I| = 1$, $\sigma_E = 10$, $\sigma_I = 1$ and $\sigma_\varphi = 3.5$ using φ_{ratio}.

---

[21]   https://www.idf.org/component/attachments/attachments.html?id=811&task=download [Page=97], last access on October 18, 2019.

[22]https://bit.ly/2zuorQK, last access on October 18, 2019.

6.3 Comparison to state-of-the-art techniques

We have investigated the ability of classical SD/EMM techniques to tackle the problem of discovering exceptional (dis)agreement among groups of individuals. To this end, we have considered three appropriate SD/EMM adaptations[23] and tested them on synthetic datasets with ground truth. No existing quality measure (in classical SD) or model (in classical EMM) makes it possible to uncover exactly the (dis)agreement patterns, and these experiments obviously supported this observation (for more details, please refer to the technical report (Belfodil et al, 2019)). This is due to the fact that SD and EMM techniques are usually tailored to tackle a specific mining task. Therefore and for the interest of brevity, we report here only comparative experiments against our first attempt (Belfodil et al, 2017) implemented by DSC.

DSC aims at discovering top-k patterns that elucidate exceptional (dis)agreement between groups of individuals. In addition, for a sufficiently large k, DSC solves the core problem tackled in this paper limited to the two first conditions (i.e., validity and maximality). Note that, we disable the aggregation dimension parameter for DSC to obtain comparable pattern sets. To compare between DEBuNk and DSC, we designed experiments to answer to the two following questions:

Q1. How concise is the patterns set provided by DEBuNk compared to the one provided by DSC?
Q2. How diverse is the pattern set, limited to k patterns, provided by DEBuNk compared to the one provided by DSC?

In order to answer (Q1), we evaluate the number of patterns returned by DEBuNk and DSC when looking for complete pattern set P (i.e., k sufficiently
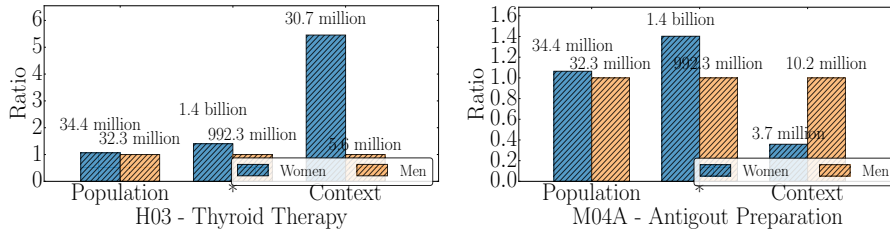
Fig. 7: Drugs consumption behavior of gender groups in Patterns 3 (left) and 4 (right).

---

[23]Since common SD techniques require flat representations of the underlying dataset augmented with a target attribute, we have proposed two adaptations: SD-Majority for discovering (dis)agreement with the majority and SD-Cartesian for discovering (dis)agreement between two groups on the cartesian product $G_E \times G_I \times G_I$. In both of the aformentioned adaptations, the target is equal to 1 if there is an agreement, 0 else. Experiments are performed using PySubgroup(Lemmerich and Becker, 2018) while utilizing the precision gain (Fürnkranz et al, 2012) as a quality measure. Moreover, to take into account the usual agreement between groups, we adapt Exceptional Subgraph Mining (Kaytoue et al, 2017) to discover contextual (dis)agreement in subgraphs representing individuals group pairs.

large for DSC). For this, we run both methods on EPD8 with various[24] quality thresholds σ_φ and descriptive attributes $\mathcal{A}_E$, $\mathcal{A}_I$. Figure 8 reports the results of these experiments. Results demonstrate that DEBuNk considerably reduces the desired pattern set while ensuring that each pattern returned by DSC is represented by a pattern returned by DEBuNk (according to the problem definition). On average, DSC returns 38 times more patterns than DEBuNk. Moreover, DEBuNk achieves better performance than DSC in terms of run time . This is explained by (i) the model simplification which reduces the complexity of computing the interestingness measure and (ii) the pruning property implemented by DEBuNk supported by condition (3) of the problem definition.

So far, we compared DEBuNk against DSC when looking for the complete pattern set. Experiments (Q1) demonstrated the fact that in such a setting DSC returns an overwhelmingly large result set. To tackle this problem, DSC implements a top-k algorithm to control the size of the returned pattern set. Of course, the main drawback of using a top-k algorithm is the lack of diversity even when redundancy is avoided by closure operators. This lack of diversity is induced by the fact that, most likely, the patterns observing the highest qualities are condensed in small region of the dataset.

To fairly evaluate the diversity of patterns returned by both DSC and DE-BuNk (Q2), we run both algorithms for several parameters[25] and compare the size of the datasets regions covered by both returned pattern sets. This quantity can be captured by the number of outcomes covered by a result set, that is $|o[P^k]| = |\{(i, e) \in G_I \times G_E \text{ s.t. } o(i, e) \text{ is expressed}\}|$ with P^k an ar-
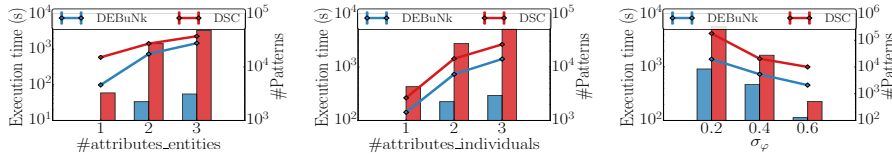

Fig. 8: Comparison between DEBuNk and DSC for the task of discovering the complete set of the desired patterns on EPD8 dataset (default parameters are: $|\mathcal{A}_E| = 2$, $|\mathcal{A}_I| = 2$, $\sigma_\varphi = 0.4$, $\sigma_E = 40$, $\sigma_I = 10$ and φ_{dissent}). Lines correspond to the execution time and bars correspond to the number of returned patterns.
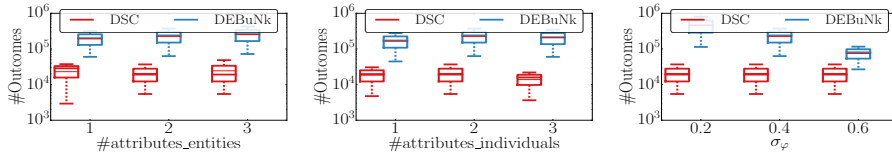

Fig. 9: Comparison between DEBuNk and DSC (top-k) for the task of discovering k-sized pattern sets on EPD8 Dataset (default parameters: $|\mathcal{A}_E| = 2$, $|\mathcal{A}_I| = 2$, $\sigma_\varphi = 0.4$, $\sigma_E = 40$, $\sigma_I = 10$ and φ_{dissent}). Box plots correspond to the size of $O[P^k]$ when varying k in $[10, 50, 100]$.

---

[24] 27 runs for each method by varying $(|\mathcal{A}_E|, |\mathcal{A}_I|, \sigma_\varphi) \in [[1, 2, 3], [1, 2, 3], [0.2, 0.4, 0.6]]$
[25] 81 runs by varying $(k, |\mathcal{A}_E|, |\mathcal{A}_I|, \sigma_\varphi) \in [[10, 50, 100], [1, 2, 3], [1, 2, 3], [0.2, 0.4, 0.6]]$

bitrary pattern set containing k patterns. For a fair comparison, we compare $|o[P^k_{\text{DSC}}]|$ (top-k patterns) against $|o[P^k_{\text{DEBuNk}}]|$. To obtain the latter quantity, we run DEBuNk so as to obtain the complete pattern set P_{DEBuNk}. Next, we draw 100 k-sized samples drawn uniformly from the obtained P_{DEBuNk} and then compute the average $|o[P^k_{\text{DEBuNk}}]|$. It is worth mentioning that comparison can be made also by taking the top-k patterns P_{DEBuNk} rather than an arbitrary k-sized sample. We decided to study the latter scenario, since the philosophy of DEBuNk is to retrieve the complete patterns set summarizing exceptional (dis)agreement in an underlying behavioral dataset.

Results are reported in Figure 9. Clearly, DEBuNk's k-sized pattern set covers larger (and different) parts of the dataset compared to DSC's top-k pattern set. We observe that DEBuNk surpasses DSC by one order of magnitude ($\times 12.5$ in average) when comparing the portions of the dataset covered by their respective k-sized pattern set. Simply put, when the pattern set related to DEBuNk covers 10% of the dataset, DSC patterns cover less than 1% of the underlying dataset records.

6.4 Performance study

6.4.1 Efficiency of closure operators and optimistic estimates

To evaluate the efficiency of closure operators and optimistic estimates, we compare DEBuNk against two baseline algorithms. The first baseline, named *Baseline*, is obtained by disabling both closure operators and the pruning properties supported by the defined optimistic estimates. Thus, *Baseline* only
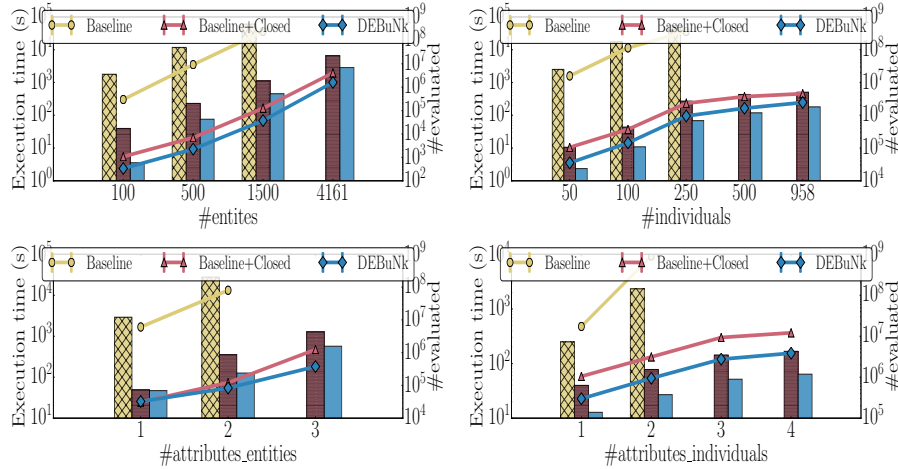

Fig. 10: Effectiveness of *DEBuNk* considering EPD8 Dataset with $|G_E| = 2000$, $|G_I| = 500$, $|Outcomes| = 750k$, $|\mathcal{A}_E| = 3$, $|\mathcal{A}_I| = 4$, $\sigma_E = 40$, $\sigma_I = 10$, $\sigma_\varphi = 0.5$ and φ_{dissent}. Lines correspond to the execution time and bars correspond to the number of evaluated patterns.

pushes the anti-monotonic constraints (i.e., the set of cardinality constraints $\mathcal{C}$). The second baseline, *Baseline+Closed*, is proposed to study more precisely the efficiency of the optimistic estimates. Thus, it is obtained by disabling the optimistic estimate based pruning. In these experiments, we interrupt a method if its execution time exceeds 10 hours. Figure 10 reports the execution time and the number of candidate patterns processed by each of the three methods when varying the size of the dataset w.r.t. both the number of records and the size of the descriptions space.

Experiments give evidence that the closure operator and the canonicity tests performed by EnumCC are effective as they drastically reduce the number of evaluated patterns. Additionally, DEBuNk is about one order of magnitude faster than the *Baseline+Closed* algorithm, thanks to the optimistic estimate-based pruning. This especially happens when the IAS measure is a simple average, which is the case of the IAS measure used for EPD8, Yelp and Movielens. This is explained by the fact that the corresponding optimistic estimate is tight. Similar observations hold for the experiments carried on Yelp and Movielens datasets, in contrast, experiments run on Openmedic exhibit slightly different performance profiles. For more details, please refer to the technical report (Belfodil et al, 2019).

6.4.2 Efficiency of HMT closed descriptions vs. closed itemsets enumeration

In order to evaluate the performance of the closed descriptions enumeration in the presence of a taxonomy linking the tags (items), we study the behavior of DEBuNk (i.e. execution time and the number of explored patterns) both with and without leveraging the hierarchy between items. The latter can be done by scaling the HMT values (as illustrated in Fig. 2) using a vector representation for each tagged record. Experiments are carried out on EPD8 and Yelp datasets whose entities are characterized by a hierarchy of 347 tags and 1175 tags respectively. To vary the number of items/tags constituting the hierarchy, we remove tags from the tree in a bottom-up fashion until the desired number of
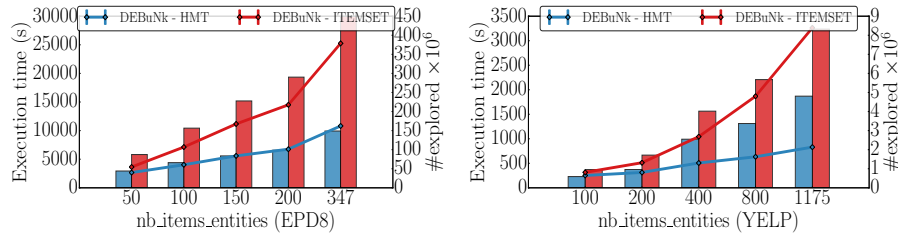

Fig. 11: Efficiency of HMT against itemsets closed descriptions enumeration according to the number of items/tags constituting the hierarchy for the two datasets EPD8 (left) and Yelp (right). For both datasets we only consider the HMT attribute for entities $|\mathcal{A}_E| = 1$. The used parameters for EPD8 are: $|\mathcal{A}_I| = 6$, $\sigma_E = 1$, $\sigma_I = 10$, $\sigma_\varphi = 0.5$ and φ_{dissent}. The used parameters for Yelp are: $|\mathcal{A}_I| = 3$, $\sigma_E = 5$, $\sigma_I = 1$, $\sigma_\varphi = 0.5$ and φ_{dissent}. Lines correspond to the execution time and bars correspond to the number of visited patterns.

tags/items is reached, followed by replacing the HMT values of each entity by the set of ascendants tags remaining in the obtained tree.

Experiments reported in Figure 11 demonstrate that taking into account the hierarchy of tags significantly improves the performance of DEBuNk ($5\times$ faster). This results from the fact that, in contrast to itemsets enumeration, HMT descriptions enumeration exploits the structure of the hierarchy and therefore avoids considering chain descriptions (e.g., {1, 1.10.40}). Note that the bars depict the number of patterns that are visited by EnumCC used in DEBuNk to generate the closed patterns. Obviously, the HMT and Itemset closed description enumeration return the same number of closed patterns. We choose to represent the number of visited patterns rather than the number of closed patterns to illustrate the differences between the HMT and Itemset enumeration in terms of the size of the explored search space.

6.4.3 Performance study of DEBuNk

We now focus on the study of DEBuNk according to the size of the description spaces ($\mathcal{D}_E$, $\mathcal{D}_I$), the support thresholds, the quality threshold and the quality measures. To study the behavior of DEBuNk according to the size of the description spaces, we choose to vary the number of items resulting from projecting the attributes values of each record (entity/individual) on their corresponding vector representation. To this end, we select values from each attribute according to the size of its corresponding domain so as to obtain the required number of items. We follow the same approach as in the experiments reported in Figure 11 to select the required number of tags for an HMT attribute. Numerical attributes domains are discretized according to the required number of items. Subsets of values of categorical attributes are regrouped under single categories in order to obtain the desired number of values.
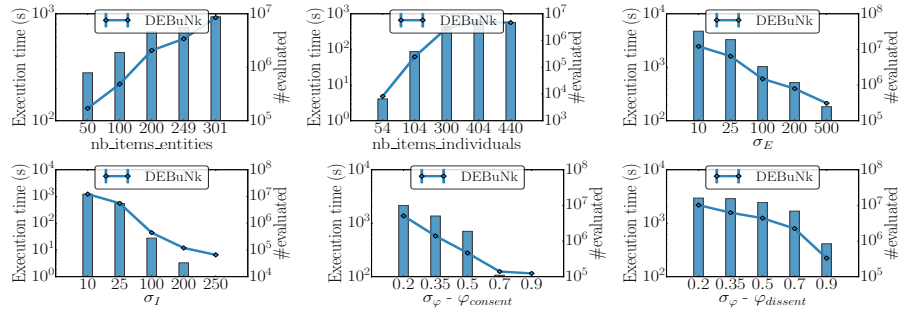

Fig. 12: Effectiveness of DEBuNk on EPD8 according to the sizes of G_E, G_I, $\mathcal{D}_E$, $\mathcal{D}_I$, the supports and quality measures thresholds. Considering by default $|G_E| = 4161$, $|G_I| = 958$, $|Outcomes| = 2.7M$, $|\mathcal{A}_E| = 3$, $|\mathcal{A}_I| = 6$. $\sigma_E = 40$, $\sigma_I = 10$, $\sigma_\varphi = 0.5$ and φ_{dissent}. Lines correspond to the execution time and bars correspond to the number of evaluated patterns.

Figure 12 reports the behavior of DEBuNk on EPD8. Clearly, the number of evaluated candidates and the execution time increase with regards to the size of description spaces $\mathcal{D}_I$ and $\mathcal{D}_E$. These experiments confirm that pushing monotonic constraints (i.e. supports threshold σ_E, σ_I) drastically improves the efficiency of DEBuNk. Finally, a higher threshold on the quality σ_φ leads to an important reduction of the number of visited patterns and therefore to a better execution time. This demonstrates the effectiveness of the pruning properties enabled by the use of optimistic estimates. We also notice that φ_{consent} performs slightly better than φ_{dissent}. This effect arises mainly from the fact that, in the EU Parliament dataset, the overall observed agreement between groups of individuals is rather consensual. The same conclusions can be drawn as well from experiments performed on Movielens, Yelp and Openmedic. For more details, see the technical report (Belfodil et al, 2019).

6.4.4 Quick-DEBuNk vs. DEBuNk

To evaluate the efficiency of Quick-DEBuNk, we compare it against the exhaustive search algorithm DEBuNk over different time budgets. To objectively measure how well Quick-DEBuNk results approximates DEBuNk results, let us first define a similarity measure $\text{sim}_{\mathcal{P}}$ between two patterns $p = (c, u_1, u_2)$ and $p' = (c', u_1', u_2')$ from $\mathcal{P}$. It captures to what extent two patterns covers the same context and groups and relies on a Jaccard Index (J in what follows):

$$
\text{sim}_{\mathcal{P}}(p, p') = \sqrt{J(G_E^c, G_E^{c'}) \times \frac{1}{2} \cdot \left(J(G_I^{u_1}, G_I^{u_1'}) + J(G_I^{u_2}, G_I^{u_2'}) \right)}
$$
$$
\text{with } J(G, G') = \frac{|G \cap G'|}{|G \cup G'|}. \tag{18}
$$

Note that, the quantity $(J(G_I^{u_1}, G_I^{u_1'}) + J(G_I^{u_2}, G_I^{u_2'}))$ is replaced by the following measure if the quality measure φ is symmetric:

$$
\max(J(G_I^{u_1}, G_I^{u_1'}) + J(G_I^{u_2}, G_I^{u_2'}), J(G_I^{u_1}, G_I^{u_2'}) + J(G_I^{u_2}, G_I^{u_1'})) \tag{19}
$$

For comparing two pattern sets P, P' returned by respectively DEBuNk and Quick-DEBuNk, we use precision and recall defined as follows.

$$
\text{precision}(P, P') = \frac{\sum_{p \in P} \max(\{\text{sim}_{\mathcal{P}}(p, p') \ s.t. \ p' \in P'\})}{|P|} \ ,
$$
$$
\text{recall}(P, P') = \frac{\sum_{p' \in P'} \max(\{\text{sim}_{\mathcal{P}}(p', p) \ s.t. \ p \in P\})}{|P'|} \ . \tag{20}
$$

A similar measure to recall has been proposed by Bosc et al (2018) to evaluate the ability of their algorithm to retrieve ground-truth patterns. We extend this measure with the precision to evaluate not only that all the patterns returned by DEBuNk have been retrieved by Quick-DEBuNk (i.e. recall=1) but

also the conciseness of the returned set (i.e. precision=1 if and only if all returned patterns by Quick-DEBuNk are actually present in the ground-truth result set, namely the returned patterns by DEBuNk).

Figures 13a, 14a and 15a report the comparative study between DEBuNk and Quick-DEBuNk carried out on respectively EPD8, Movielens and Yelp. We notice that in all situations, Quick-DEBuNk is able to promptly return high quality patterns. Interestingly, some differences can be observed between datasets. Quick-DEBuNk is less efficient on the Yelp dataset. We argue that this is due to the fact that the corresponding context search space is much larger than the three other behavioral datasets (see Table 2) which might impede random walk step RWC for finding high quality patterns.

We also investigate the empirical distribution from which the patterns are sampled when using Quick-DEBuNk. This requires the true distribution of the qualities of valid patterns in the corresponding datasets. To this end, we run DEBuNk by disabling the generality condition (see Problem definition). This makes it possible to identify all interesting (dis)agreement patterns in the dataset. In these experiments, we choose an arbitrary threshold set to $\sigma_\varphi = 0.1$. Similarly, we run Quick-DEBuNk so as to obtain a sufficiently large pattern set, and calculate the sampling distribution from the retrieved patterns' qualities. We observe from the empirical distributions depicted in
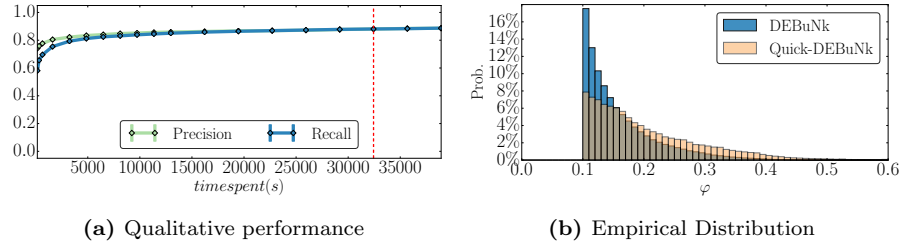

(a) Qualitative performance   **(b)** Empirical Distribution

Fig. 13: Efficiency of Quick-DEBuNk compared to DEBuNk on EPD8. Parameters used are $\sigma_E = 40$, $\sigma_I = 10$, $\sigma_\varphi = 0.5$ and φ_{dissent}. The red line corresponds to the required time by DEBuNk to perform an exhaustive search.

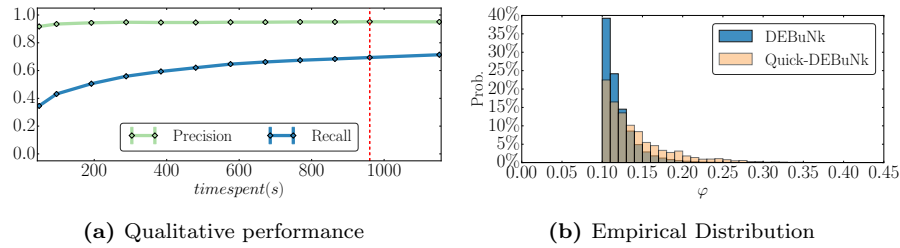**(a)** Qualitative performance   **(b)** Empirical Distribution

Fig. 14: Efficiency of Quick-DEBuNk compared to DEBuNk on Movielens. Parameters used are $\sigma_E = 5$, $\sigma_I = 10$, $\sigma_\varphi = 0.25$ and φ_{dissent}. The red line corresponds to the required time by DEBuNk to perform an exhaustive search.

Figures 13b, 14b and 15b that Quick-DEBuNk rewards high quality patterns by giving them a better chance to be sampled.

Finally, to evaluate the importance of the RWC (Random Walk on Contexts) step in Quick-DEBuNk, we perform the same experiments with the same time budgets with the RWC step disabled. This configuration, Quick-DEBuNk without RWC returned only 3 472, 389 and 120 valid patterns compared to 408 610, 64 198 and 75 398 valid patterns when carried out on, respectively, EPD8, Movielens and Yelp. In average, Quick-DEBuNk without RWC retrieved 20× fewer valid patterns than the original Quick-DEBuNk. This clearly indicates that RWC improves the performance of Quick-DEBuNk. This stems from the fact that when the first step (FBS step) generates a pattern, most of the time the pattern is not of a sufficient quality. RWC tackles this issue by locally searching for interesting patterns, starting from the generated pattern.

6.5 Discussion

DEBuNk scales well w.r.t. the size of the search space corresponding to the entities collection thanks to the defined optimistic estimates which enable to prune unpromising parts of the search space. However, DEBuNk does not scale according to the size of the description spaces related to the individuals. This limits its application when behavioral datasets have a large number of individuals described with many attributes. This is due to the need of taking into account the usual inter-group agreement in the interestingness measures. As a consequence, it is notoriously difficult to define an optimistic estimate which not only works on the entities related search space, but also on the one corresponding to the confronted couples of groups of individuals. This should be the scope of future research, starting with definition of bounds on the usual agreement quantity. Algorithm Quick-DEBuNk partially addresses this scalability issue by sampling the couples of groups directly from the patterns space rather than starting from the search tree root. Interestingly, the experiments demonstrated that Quick-DEBuNk makes it possible to retrieve most of the interesting patterns in a relatively small amount of time (i.e. compared to
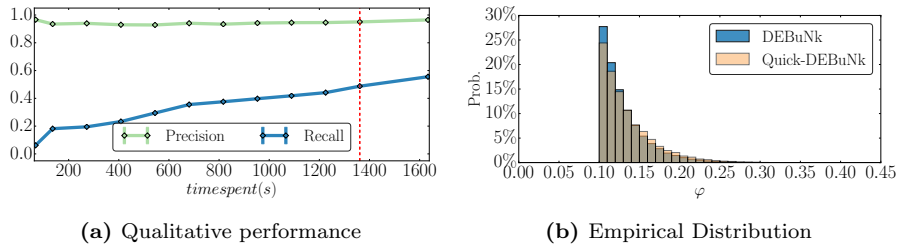

(a) Qualitative performance

(b) Empirical Distribution

Fig. 15: Efficiency of Quick-DEBuNk compared to DEBuNk over Yelp. Parameters used are $\sigma_E = 15$, $\sigma_I = 1$, $\sigma_\varphi = 0.1$ and φ_{dissent}. The red line corresponds to the required time by DEBuNk to perform an exhaustive search.

what returns the exhaustive search algorithm DEBuNk and the ground truth in artificial data). This is particularly observed for EPD8 dataset involving the largest descriptions space $\mathcal{D}_I \times \mathcal{D}_I$, hence empirically demonstrating its interest. Nevertheless, Quick-DEBuNk does not have theoretical guarantees on the distribution of the sampled patterns (we only proved that all valid patterns are reachable and are generated proportionally to their size). This shortcoming is due to two reasons. On the one hand, the three-set format of the patterns makes them challenging to be sampled proportionally to their interestingness measure since the value is computed only when the context is known (no information is available before the instantiation of the two groups). On the other hand, quality measures that are expressed as average functions are complex to apprehend under direct pattern sampling framework. Dealing with this two issues is required to obtain theoretical guarantees.

To avoid misleading interpretations, it is important to be aware of the data sparsity. To this end, the proposed approaches enable to discard some patterns that involve too small subset of entities on which the two confronted groups haven't expressed enough outcomes. Moreover, the strength of the claim related to the pattern should be assessed according not only to the data sparsity but also to the representativeness of the two subpopulation of interest (e.g., the claims drawn from the EU parliament votes are usually consistent even though the data are fairly sparse).

7 Related Work

Subgroup Discovery and Exceptional Model Mining. Scientists have always seen Exploratory Data Analysis (EDA) as an important research area since its introduction (Tukey, 1977). Among the various EDA techniques that aim to maximize insight into datasets and uncover underlying structures, Subgroup Discovery (SD) (Klösgen, 1996; Wrobel, 1997; Atzmueller, 2015; Herrera et al, 2011) is a generic data mining task concerned with finding regions in the data that stand out with respect to a given target. Many other data mining tasks have similar goals as SD, e.g., emerging patterns (Dong and Li, 1999), significant rules (Terada et al, 2013), contrast sets (Bay and Pazzani, 2001) or classification association rules (Liu et al, 1998). However, among these different tasks, SD is known as the most generic one, especially SD is agnostic of the data and the pattern domain. For instance, subgroups can be defined with a conjunction of conditions on symbolic (Lavrac et al, 2004) or numeric attributes (Grosskreutz and Rüping, 2009; Atzmüller and Puppe, 2006) as well as sequences (Grosskreutz et al, 2013). Furthermore, the single target can be discrete or numeric (Lemmerich et al, 2016). Exceptional Model Mining (EMM) (Leman et al, 2008), while sharing exactly the same exploration space (i.e., the description space), extends SD by offering the possibility to handle complex targets, e.g., several discrete attributes (van Leeuwen and Knobbe, 2012; Duivesteijn et al, 2010, 2016), graphs (Kaytoue et al, 2017; Bendimerad et al, 2016, 2017), two numeric targets (Downar and Duivesteijn, 2017) and

preferences (de Sá et al, 2016, 2018). Our method is rooted in the SD/EMM framework. Nevertheless, the problem we tackle cannot be directly addressed with an instance of SD/EMM because a target space is provided instead of explicit targets. As a consequence, the discovery of (dis)agreement patterns with a SD/EMM instance would consist in performing a SD discovery algorithm per pair of confronted groups of individuals, which is not feasible in practice due to the exponential number of possible pairs of groups. Dynamic EMM (i.e., EMM with a non-fixed model) has been recently investigated for different aims. Bosc et al (2016) propose a method to handle multi-label data where the number of labels per record is much lower than the total number of labels which prevents the use of usual EMM model. Other dynamic EMM approaches aim to discover exceptional attributed sub-graphs (Kaytoue et al, 2017; Bendimerad et al, 2016, 2017). The present paper, which is an extension of (Belfodil et al, 2017), is the first attempt to discover (dis)agreement patterns with a method rooted in dynamic EMM. Our preliminary work (Belfodil et al, 2017) is extended on many levels: (1) we provide an easier to use framework to discover exceptional (dis)agreement between groups which requires less parameter setting and interpretation effort by the end-user. (2) Our proposal enables to use a wider spectrum of interestingness measures that can be enriched by relying on the building blocks discussed in Section 3. (3) This work provides a more elaborate exhaustive search algorithm compared to the former one as discussed in Section 4 and (4) An alternative sampling approach Quick-DEBuNk is proposed. It enables instant mining of (dis)agreement patterns, which sets the ground for interactive pattern mining tools.

Rating Data Analysis. Behavioral data analysis has received a wide interest in the last two decades allowing the development of new services for consumers, citizens, companies and organizations. In (Das et al, 2011), the problem of discovering meaningful ratings interpretation is introduced. It can be formalized as a SD problem, the authors' aim is to identify groups of users that substantially agree or disagree w.r.t. some given subsets of entities while using a mono-objective measure (ratings average). Extensions have been proposed to enable multi-objective groups identification thanks to more complex statistical measure (rating distribution) (Amer-Yahia et al, 2017; Omidvar-Tehrani et al, 2016). This makes it possible to discover intra-group behavior patterns such as polarized and homogeneous opinions. The main differences with our work are: (i) these methods consider intra-group agreement only (no inter-group agreement) without taking into account the usual agreement observed between the individuals; (ii) the set of reviewees on which the study is performed is given in prior, in contrast to our SD/EMM based approach, which discovers relevant contexts by leveraging the reviewees dimension.

Voting Data Analysis. Similarly, the past two decades have witnessed an increasing emergence of Open Government Data[26] (OGD) promoting transparency and accountability in public institutions. Consequently, many re-

---

[26]http://www.oecd.org/gov/digital-government/open-government-data.htm, last access on October 25, 2019.

searchers from different fields (e.g., information science, political and social sciences, data mining and machine learning) have studied such data (Charalabidis et al, 2016). For instance, (Jakulin, 2004) uses hierarchical clustering and PCA to identify cohesion blocs and dissimilarity blocs of voters within the US Senate. Similar work was done on the Finnish (Pajala et al, 2004), the Italian (Amelio and Pizzuti, 2012) and the Swiss (Etter et al, 2014) parliaments to study the polarization and cohesion between parliamentarians. Similarly, Grosskreutz et al (2010) investigate the voting behavior of citizens instead of politicians relying on SD. Our work goes further and supports the discovery of new insights in such data.

Outpatient Data Analysis. Monitoring the disease prevalence is an important task. Many researchers dedicated much effort to analyze the prevalence of diseases considering different sources of data. Orueta et al (2012) highlight the importance of considering outpatient data (e.g., medical prescriptions) in such epidemiology studies. This motivates the analysis task proposed over Openmedic data. It provides an interesting additional tool in epidemiology surveillance applications by revealing substantial differences in medicine consumption between subpopulations.

Pattern Sampling Approaches. The discovery of the complete set of interesting patterns (e.g., frequent, discriminant) has two disadvantages that limit the use of such methods in practice. It is time consuming to compute the complete set of solutions. Furthermore, this set can be huge and non-manageable for a human expert. To overcome these limitations, many approaches that can effectively sample the pattern space for interesting patterns have been proposed for a decade. These methods address frequent or discriminant itemset mining tasks (Boley et al, 2011; Giacometti and Soulet, 2016; Li and Zaki, 2016; Moens and Goethals, 2013) offering some theoretical guarantees on the sampling quality or more generic ones (Dzyuba et al, 2017; Boley et al, 2010a; Hasan and Zaki, 2009). Dzyuba et al (2017) define the problem of sampling pattern sets and propose a method based on a SAT solver sampling solution. However, this approach only supports pattern languages that can be compactly represented by binary variables such as itemsets. It requires the discretization of numerical attributes. Authors in (Boley et al, 2010a; Hasan and Zaki, 2009) use a MCMC (Monte-Carlo Markov-Chain) based algorithm to achieve sampling with guarantees according to a desired probability distribution. Despite the generic nature and the interesting guarantees that MCMC algorithms provide, it requires a number of steps that grows exponentially in the input size to generate a single pattern (Boley et al, 2010a). This may prevent the user to obtain instant results. The problem we are interested in has several specificities. First, the search space involves attributes of different types (i.e., numerical, symbolical, HMT attributes) which prevents us to use sampling techniques based on itemset language. Second, the quality measure is not considered in the state-of-the-art methods that mainly support frequency and discriminative measures (Boley et al, 2011, 2012). Finally, the method proposed in (Moens and Boley, 2014) for EMM is not suited to our problem since we have to

simultaneously consider both description space and target space. Algorithm Quick-DEBuNk handles the specificity of the problem by combining an exploration step (i.e. direct sampling from the pattern space) and an exploitation step while taking profit of the quality measures properties (i.e. random walk step on contexts search space). However, we have no theoretical guarantee on the quality of the sampled (dis)agreement patterns.

8 Conclusion

In this paper, we have defined the problem of discovering exceptional (dis)-agreement in behavioral data. The generic definition of behavioral data enables to encompass datasets featuring individuals and their outcomes on some entities whatever the application domain (e.g., ratings, votes, medicines consumption). The discovery of exceptional (dis)agreement patterns is rooted in SD/EMM with a novel pattern domain and associated quality measures. However, the targets are not specified and have to be enumerated in our framework. We have defined DEBuNk, a branch-and-bound algorithm which takes benefit from closure operators, properties of the underlying descriptions space (as for HMT attributes) and (tight) optimistic estimates to efficiently enumerate the patterns. Alternatively, we devised Quick-DEBuNk that samples the space of patterns instead of returning the complete set of (dis)agreement patterns. We have investigated several quality measures to assess inter-group agreement. The extensive experimental study demonstrates the efficiency of our algorithms as well as their ability to provide new insights in three case-studies: (i) the investigation of contexts that impact the inter-group agreement between parliamentarians, (ii) the characterization of affinities and contrasted opinions between reviewers in rating platforms and (iii) the study of prevalence of certain sicknesses that can be pointed out by high discrepancies between the medicine consumption rates of two subpopulations.

We believe that this work opens new directions for future research. This generic framework can be extended by paying a particular attention to the analysis of intra-group agreement within a group of individuals. It may support the discovery of contexts that divide a political group. This requires the definition and the integration of suited similarity measures into the IAS (Intergroup Agreement Similarity) measure. For instance, the cohesion of a political group can be assessed by the "agreement index" (Hix et al, 2005), which is an application-specific measure to the study the European parliament. More generic measures, such as Krippendorff's alpha coefficient (Hayes and Krippendorff, 2007), could also be investigated. While our method is able to analyze behavioral datasets with large collections of entities (e.g., Yelp), tackling large collections of individuals still remains challenging to ensure the scalability of both DEBuNk and Quick-DEBuNk. Indeed, the search space related to individuals does not have, according to our problem definition, properties that can be leveraged to prune unpromising parts of this search space. Another interesting future direction is to take into account the temporal dimension into the

analysis of behavioral data. This can offer the opportunity to investigate how the relationship (e.g., inter-group agreement) between groups of individuals evolves through time.

References

Amelio A, Pizzuti C (2012) Analyzing voting behavior in italian parliament: Group cohesion and evolution. In: International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012, Istanbul, Turkey, 26-29 August 2012, pp 140–146, DOI 10.1109/ASONAM.2012.33

Amer-Yahia S, Kleisarchaki S, Kolloju NK, Lakshmanan LVS, Zamar RH (2017) Exploring rated datasets with rating maps. In: Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017, pp 1411–1419, DOI 10.1145/3038912.3052623

Atzmueller M (2015) Subgroup discovery. Wiley Interdiscip Rev Data Min Knowl Discov 5(1):35–49, DOI 10.1002/widm.1144

Atzmüller M, Puppe F (2006) Sd-map - A fast algorithm for exhaustive subgroup discovery. In: Knowledge Discovery in Databases: PKDD 2006, 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, Berlin, Germany, September 18-22, 2006, Proceedings, pp 6–17, DOI 10.1007/11871637_6

Bay SD, Pazzani MJ (2001) Detecting group differences: Mining contrast sets. Data Min Knowl Discov 5(3):213–246, DOI 10.1023/A:1011429418057

Belfodil A, Cazalens S, Lamarre P, Plantevit M (2017) Flash points: Discovering exceptional pairwise behaviors in vote or rating data. In: Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Proceedings, Part II, pp 442–458, DOI 10.1007/978-3-319-71246-8_27

Belfodil A, Cazalens S, Lamarre P, Plantevit M (2019) Identifying exceptional (dis)agreement between groups (Technical Report). Tech. rep., LIRIS UMR CNRS 5205, URL https://contentcheck.liris.cnrs.fr/public/technical_report_2019_02.pdf

Bendimerad AA, Plantevit M, Robardet C (2016) Unsupervised exceptional attributed sub-graph mining in urban data. In: IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain, pp 21–30, DOI 10.1109/ICDM.2016.0013

Bendimerad AA, Cazabet R, Plantevit M, Robardet C (2017) Contextual subgraph discovery with mobility models. In: Complex Networks & Their Applications VI - Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications), COM-

PLEX NETWORKS 2017, Lyon, France, November 29 - December 1, 2017., pp 477–489, DOI 10.1007/978-3-319-72150-7_39

Boley M, Gärtner T, Grosskreutz H (2010a) Formal concept sampling for counting and threshold-free local pattern mining. In: Proceedings of the SIAM International Conference on Data Mining, SDM 2010, April 29 - May 1, 2010, Columbus, Ohio, USA, pp 177–188, DOI 10.1137/1.9781611972801. 16

Boley M, Horváth T, Poigné A, Wrobel S (2010b) Listing closed sets of strongly accessible set systems with applications to data mining. Theor Comput Sci 411(3):691–700, DOI 10.1016/j.tcs.2009.10.024

Boley M, Lucchese C, Paurat D, Gärtner T (2011) Direct local pattern sampling by efficient two-step random procedures. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011, pp 582–590, DOI 10.1145/2020408.2020500

Boley M, Moens S, Gärtner T (2012) Linear space direct pattern sampling using coupling from the past. In: The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Beijing, China, August 12-16, 2012, pp 69–77, DOI 10.1145/2339530.2339545

Bosc G, Golebiowski J, Bensafi M, Robardet C, Plantevit M, Boulicaut J, Kaytoue M (2016) Local subgroup discovery for eliciting and understanding new structure-odor relationships. In: Discovery Science - 19th International Conference, DS 2016, Bari, Italy, October 19-21, 2016, Proceedings, pp 19–34, DOI 10.1007/978-3-319-46307-0_2

Bosc G, Boulicaut J, Raïssi C, Kaytoue M (2018) Anytime discovery of a diverse set of patterns with monte carlo tree search. Data Min Knowl Discov 32(3):604–650, DOI 10.1007/s10618-017-0547-5

Charalabidis Y, Alexopoulos C, Loukis E (2016) A taxonomy of open government data research areas and topics. Journal of Organizational Computing and Electronic Commerce 26(1-2):41–63

Csisz I, et al (1967) Information-type measures of difference of probability distributions and indirect observations. Studia Sci Math Hungar 2:299–318

Das M, Amer-Yahia S, Das G, Yu C (2011) MRI: meaningful interpretations of collaborative ratings. PVLDB 4(11):1063–1074, URL http://www.vldb.org/pvldb/vol4/p1063-das.pdf

Dong G, Li J (1999) Efficient mining of emerging patterns: Discovering trends and differences. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 15-18, 1999, pp 43–52, DOI 10.1145/312129.312191

Downar L, Duivesteijn W (2017) Exceptionally monotone models—the rank correlation model class for exceptional model mining. Knowledge and Information Systems 51(2):369–394, DOI 10.1007/s10115-016-0979-z

Duivesteijn W, Knobbe AJ, Feelders A, van Leeuwen M (2010) Subgroup discovery meets bayesian networks – an exceptional model mining approach. In: ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010, pp 158–167, DOI 10.1109/ICDM.

2010.53

Duivesteijn W, Feelders A, Knobbe AJ (2016) Exceptional model mining - supervised descriptive local pattern mining with complex target concepts. Data Min Knowl Discov 30(1):47–98, DOI 10.1007/s10618-015-0403-4

Dzyuba V, van Leeuwen M, Raedt LD (2017) Flexible constrained sampling with guarantees for pattern mining. Data Min Knowl Discov 31(5):1266–1293, DOI 10.1007/s10618-017-0501-6

Etter V, Herzen J, Grossglauser M, Thiran P (2014) Mining democracy. In: Proceedings of the second ACM conference on Online social networks, COSN 2014, Dublin, Ireland, October 1-2, 2014, pp 1–12, DOI 10.1145/2660460.2660476

Fürnkranz J, Gamberger D, Lavrac N (2012) Foundations of Rule Learning. Cognitive Technologies, Springer, DOI 10.1007/978-3-540-75197-7

Ganter B, Kuznetsov SO (2001) Pattern structures and their projections. In: Delugach HS, Stumme G (eds) Conceptual Structures: Broadening the Base, 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, USA, July 30-August 3, 2001, Proceedings, Springer, Lecture Notes in Computer Science, vol 2120, pp 129–142, DOI 10.1007/3-540-44583-8_10

Ganter B, Wille R (1999) Formal Concept Analysis - Mathematical Foundations. Springer, DOI 10.1007/978-3-642-59830-2

Giacometti A, Soulet A (2016) Frequent pattern outlier detection without exhaustive mining. In: Advances in Knowledge Discovery and Data Mining - 20th Pacific-Asia Conference, PAKDD 2016, Auckland, New Zealand, April 19-22, 2016, Proceedings, Part II, pp 196–207, DOI 10.1007/978-3-319-31750-2_16

Grosskreutz H, Rüping S (2009) On subgroup discovery in numerical domains. Data Min Knowl Discov 19(2):210–226, DOI doi:10.1007/s10618-009-0136-3

Grosskreutz H, Rüping S, Wrobel S (2008) Tight optimistic estimates for fast subgroup discovery. In: Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I, pp 440–456, DOI 10.1007/978-3-540-87479-9_47

Grosskreutz H, Boley M, Krause-Traudes M (2010) Subgroup discovery for election analysis: A case study in descriptive data mining. In: Discovery Science - 13th International Conference, DS 2010, Canberra, Australia, October 6-8, 2010. Proceedings, pp 57–71, DOI 10.1007/978-3-642-16184-1_5

Grosskreutz H, Lang B, Trabold D (2013) A relevance criterion for sequential patterns. In: Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part I, pp 369–384, DOI 10.1007/978-3-642-40988-2_24

Harper FM, Konstan JA (2016) The movielens datasets: History and context. TiiS 5(4):19:1–19:19, DOI 10.1145/2827872

Hasan MA, Zaki MJ (2009) Output space sampling for graph patterns. PVLDB 2(1):730–741, DOI 10.14778/1687627.1687710, URL http://www.

vldb.org/pvldb/2/vldb09-713.pdf

Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. Communication methods and measures 1(1):77–89

Herrera F, Carmona CJ, González P, del Jesús MJ (2011) An overview on subgroup discovery: foundations and applications. Knowl Inf Syst 29(3):495–525, DOI 10.1007/s10115-010-0356-2

Hix S, Noury A, Roland G (2005) Power to the parties: cohesion and competition in the european parliament, 1979–2001. British Journal of Political Science 35(2):209–234

Jakulin A (2004) Analyzing the us senate in 2003 : Similarities , networks , clusters and blocs. URL http://kt.ijs.si/aleks/Politics/us_senate.pdf

Johnson D, Sinanovic S (2001) Symmetrizing the kullback-leibler distance. IEEE Transactions on Information Theory

Kaytoue M, Kuznetsov SO, Napoli A, Duplessis S (2011) Mining gene expression data with pattern structures in formal concept analysis. Inf Sci 181(10):1989–2001, DOI 10.1016/j.ins.2010.07.007

Kaytoue M, Plantevit M, Zimmermann A, Bendimerad AA, Robardet C (2017) Exceptional contextual subgraph mining. Machine Learning 106(8):1171–1211, DOI 10.1007/s10994-016-5598-0

Klösgen W (1996) Explora: A multipattern and multistrategy discovery assistant. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in Knowledge Discovery and Data Mining., AAAI/MIT Press, pp 249–271

Kuznetsov SO, Obiedkov SA (2002) Comparing performance of algorithms for generating concept lattices. J Exp Theor Artif Intell 14(2-3):189–216, DOI 10.1080/09528130210164170

Lavrac N, Kavsek B, Flach PA, Todorovski L (2004) Subgroup discovery with CN2-SD. J Mach Learn Res 5:153–188, URL http://jmlr.org/papers/volume5/lavrac04a/lavrac04a.pdf

van Leeuwen M, Knobbe AJ (2012) Diverse subgroup set discovery. Data Min Knowl Discov 25(2):208–242, DOI doi:10.1007/s10618-012-0273-y

Leman D, Feelders A, Knobbe AJ (2008) Exceptional model mining. In: Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II, pp 1–16, DOI 10.1007/978-3-540-87481-2_1

Lemmerich F, Becker M (2018) pysubgroup: Easy-to-use subgroup discovery in python. In: Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2018, Dublin, Ireland, September 10-14, 2018, Proceedings, Part III, pp 658–662, DOI 10.1007/978-3-030-10997-4_46

Lemmerich F, Atzmueller M, Puppe F (2016) Fast exhaustive subgroup discovery with numerical target concepts. Data Min Knowl Discov 30(3):711–762, DOI 10.1007/s10618-015-0436-8

Li G, Zaki MJ (2016) Sampling frequent and minimal boolean patterns: theory and application in classification. Data Min Knowl Discov 30(1):181–225,

DOI 10.1007/s10618-015-0409-y

Liu B, Hsu W, Ma Y (1998) Integrating classification and association rule mining. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York City, New York, USA, August 27-31, 1998, pp 80–86, URL http://www.aaai.org/Library/KDD/1998/kdd98-012.php

Moens S, Boley M (2014) Instant exceptional model mining using weighted controlled pattern sampling. In: Advances in Intelligent Data Analysis XIII - 13th International Symposium, IDA 2014, Leuven, Belgium, October 30 - November 1, 2014. Proceedings, pp 203–214, DOI 10.1007/978-3-319-12571-8_18

Moens S, Goethals B (2013) Randomly sampling maximal itemsets. In: Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics, IDEA@KDD 2013, Chicago, Illinois, USA, August 11, 2013, pp 79–86, DOI 10.1145/2501511.2501523

Novak PK, Lavrac N, Webb GI (2009) Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. J Mach Learn Res 10:377–403

Omidvar-Tehrani B, Amer-Yahia S, Dutot P, Trystram D (2016) Multi-objective group discovery on the social web. In: Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I, pp 296–312, DOI 10.1007/978-3-319-46128-1_19

Orueta JF, Nuño-Solinis R, Mateos M, Vergara I, Grandes G, Esnaola S (2012) Monitoring the prevalence of chronic conditions: which data should we use? BMC health services research 12(1):365

Pajala A, Jakulin A, Buntine W (2004) Parliamentary group and individual voting behavior in finnish parliament in year 2003 : A group cohesion and voting similarity analysis. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.2295&rep=rep1&type=pdf

Pasquier N, Bastide Y, Taouil R, Lakhal L (1999) Discovering frequent closed itemsets for association rules. In: Database Theory - ICDT '99, 7th International Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings., pp 398–416, DOI 10.1007/3-540-49257-7_25

Roddy E, Doherty M (2010) Epidemiology of gout. Arthritis research & therapy 12(6):223

Roman S (2008) Lattices and ordered sets. Springer Science & Business Media, Berlin/Heidelberg, Germany

de Sá CR, Duivesteijn W, Soares C, Knobbe AJ (2016) Exceptional preferences mining. In: Discovery Science - 19th International Conference, DS 2016, Bari, Italy, October 19-21, 2016, Proceedings, pp 3–18, DOI 10.1007/978-3-319-46307-0_1

de Sá CR, Duivesteijn W, Azevedo PJ, Jorge AM, Soares C, Knobbe AJ (2018) Discovering a taste for the unusual: exceptional models for preference mining. Machine Learning 107(11):1775–1807, DOI 10.1007/s10994-018-5743-z

Terada A, Okada-Hatakeyama M, Tsuda K, Sese J (2013) Statistical significance of combinatorial regulations. Proceedings of the National Academy of Sciences 110(32):12,996–13,001

Tukey JW (1977) Exploratory data analysis. Addison-Wesley series in behavioral science : quantitative methods, Addison-Wesley, URL http://www.worldcat.org/oclc/03058187

Wang C, Crapo LM (1997) The epidemiology of thyroid disease and implications for screening. Endocrinology and Metabolism Clinics 26(1):189–218

Wrobel S (1997) An algorithm for multi-relational discovery of subgroups. In: Principles of Data Mining and Knowledge Discovery, First European Symposium, PKDD '97, Trondheim, Norway, June 24-27, 1997, Proceedings, pp 78–87, DOI 10.1007/3-540-63223-9_108

Symbol	Definition
G_E	A finite collection of records depicting entities
G_I	A finite collection of records depicting individuals
O	the domain of possible outcomes
o	function returning the outcome of an individual over an entity
$\langle G_I, G_E, O, o \rangle$	A behavioral dataset
$\mathcal{A}_E$	Descriptive attributes of entities
$\mathcal{A}_I$	Descriptive attributes of individuals
θ	An outcome aggregation measure
sim	a similarity function between two aggregated outcomes
IAS	Inter-group Agreement Similarity Measure
φ	An interestingness measure
$\mathcal{D}_E$	The description domain of entities containing all possible contexts
$\mathcal{D}_I$	The description domain of individuals
G_E^d	A subgroup of entities supporting a description $d \in \mathcal{D}_E$
G_I^u	A subgroup of individuals supporting a description $u \in \mathcal{D}_I$
δ	A mapping function that binds an entity from G to a description in $\mathcal{D}$
$\mathcal{P}$	$= \mathcal{D}_E \times \mathcal{D}_I \times \mathcal{D}_I$ and denotes the pattern space
p	$= (c, u_1, u_2) \in \mathcal{P}$ depicts a (dis)agreement pattern
p^*	$= (*, u_1, u_2) \in \mathcal{P}$ depicts the referential (dis)agreement pattern
	related to some pattern $p = (c, u_1, u_2)$
P	$\subseteq \mathcal{P}$ denotes a pattern set
$\sqsubseteq$	read "less restrictive than" is a partial order between descriptions (resp.
	patterns) in some descriptions space $\mathcal{D}$ (resp. patterns space $\mathcal{P}$)
$\wedge$	the conjunction operator between descriptions in $\mathcal{D}$
η	the refinement operator which applies atomic refinements of a given
	description $d \in \mathcal{D}$, thereby yielding neighbor descriptions of d w.r.t. $\sqsubseteq$

Table 8: Symbol table

A Appendix: Proofs of Theorems and Propositions

Before giving the proof of the proposition 1 we present the following lemma.

Lemma 1 *Let $n \in \mathbb{N}^*$, $A = \{a_i\}_{1 \leq i \leq n}$ and $B = \{b_i\}_{1 \leq i \leq n}$ such that:*

$$\forall i \in 1..n-1, \ 0 \leq a_i \leq a_{i+1},$$
$$\forall i \in 1..n-1, \ 0 < b_{i+1} \leq b_i.$$

We have:

$$\forall k \in 1..n, \ \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i} \leq \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq \frac{\sum_{i=n-k+1}^{n} a_i}{\sum_{i=n-k+1}^{n} b_i}.$$

Proof (Lemma 1) Using the same notations of the lemma, we know that:

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} - \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i}$$

is of the same sign of:

$$\left(\sum_{i=1}^{n} a_i \right) \times \left(\sum_{i=1}^{k} b_i \right) - \left(\sum_{i=1}^{k} a_i \right) \times \left(\sum_{i=1}^{n} b_i \right).$$

This above quantity is equal to:

$$\left(\sum_{i=1}^{k} a_i + \sum_{i=k+1}^{n} a_i \right) \times \left(\sum_{i=1}^{k} b_i \right) - \left(\sum_{i=1}^{k} a_i \right) \times \left(\sum_{i=1}^{k} b_i + \sum_{i=k+1}^{n} b_i \right).$$

Which is equal to

$$\left(\sum_{i=k+1}^{n} a_i \right) \times \left(\sum_{i=1}^{k} b_i \right) - \left(\sum_{i=1}^{k} a_i \right) \times \left(\sum_{i=k+1}^{n} b_i \right).$$

Using the lemma hypotheses (orders between a_i's and b_i's), we have:

$$\sum_{i=k+1}^{n} a_i \geq (n-k) \times a_k,$$

$$\sum_{i=1}^{k} b_i \geq k \times b_k,$$

$$\sum_{i=1}^{k} a_i \leq k \times a_k,$$

$$\sum_{i=k+1}^{n} b_i \leq (n-k) \times b_k.$$

Thus:

$$\left(\sum_{i=k+1}^{n} a_i \right) \times \left(\sum_{i=1}^{k} b_i \right) \geq (n-k) \times k \times a_k \times b_k,$$

$$\left(\sum_{i=1}^{k} a_i \right) \times \left(\sum_{i=k+1}^{n} b_i \right) \leq (n-k) \times k \times a_k \times b_k.$$

We conclude that

$$\left(\sum_{i=k+1}^{n} a_i \right) \times \left(\sum_{i=1}^{k} b_i \right) - \left(\sum_{i=1}^{k} a_i \right) \times \left(\sum_{i=k+1}^{n} b_i \right) \geq 0.$$

Hence, we have:

$$\forall k \in 1..n, \; \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} b_i} \leq \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.$$

Similarly the inequality $\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq \frac{\sum_{i=n-k+1}^{n} a_i}{\sum_{i=n-k+1}^{n} b_i}$ can be easily proved following the same line of reasoning of the proof of the first part of the inequality.

$\square$

Proof (Proposition 1) By a straightforward application of Lemma 1 we obtain for any d s.t. $|G_E^d| \geq \sigma_E$ the following inequality.

$$LB(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}). \tag{21}$$

This stems from the fact that $LB(G_E^d, G_I^{u_1}, G_I^{u_2})$ takes the sum of the lowest σ_E quantities constituting the numerator of $\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2})$ and divides them by the sum of the greatest σ_E quantities forming the denominator of $\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2})$.

Moreover, we have that LB is monotonic w.r.t. $\sqsubseteq$ of $\mathcal{D}_E$. i.e.

$$c \sqsubseteq d \Rightarrow LB(G_E^c, G_I^{u_1}, G_I^{u_2}) \leq LB(G_E^d, G_I^{u_1}, G_I^{u_2}). \tag{22}$$

This results from $c \sqsubseteq d \Rightarrow G_E^d \subseteq G_E^c$. Hence, if we reorder values of G_E^c and G_E^d where $G_E^c = \{e_1^c, ..., e_{|G_E^c|}^c\}$ and $G_E^d = \{e_1^d, ..., e_{|G_E^d|}^d\}$ as such:

$$\begin{cases} w_{e_1^c}.\alpha(e_1^c) \leq w_{e_2^c}.\alpha(e_2^c) \leq ... \leq w_{e_{\sigma_E}^c}.\alpha(e_{\sigma_E}^c) \leq ... \leq w_{e_{|E_c|}^c}.\alpha(e_{|G_E^c|}^c) \\ w_{e_1^d}.\alpha(e_1^d) \leq w_{e_2^d}.\alpha(e_2^d) \leq ... \leq w_{e_{\sigma_E}^d}.\alpha(e_{\sigma_E}^d) \leq ... \leq w_{e_{|G_E^d|}^d}.\alpha(e_{|G_E^d|}^d) \end{cases} .$$

Given that $G_E^d \subseteq G_E^c$, it is clear that: $\forall i \leq \sigma_E, \; w_{e_i^c}.\alpha(e_i^c) \leq w_{e_i^d}.\alpha(e_i^d)$. Having that $m(G_E^c, \sigma_E) = \{e_1^c, ..., e_{\sigma_E}^c\}$ and $m(G_E^d, \sigma_E) = \{e_1^d, ..., e_{\sigma_E}^d\}$, it follows that:

$$\sum_{e \in m(G_E^c, \sigma_E)} w_e \times \alpha(e) \leq \sum_{e \in m(G_E^d, \sigma_E)} w_e \times \alpha(e). \tag{23}$$

Similarly, if we reorder entities e in descending order w.r.t the weights w_e we have $\forall j \leq \sigma_E \mid w_{e_j^d} \leq w_{e_j^c}$. Resulting in:

$$\sum_{e \in Mw(G_E^c, \sigma_E)} w_e \geq \sum_{e \in Mw(G_E^d, \sigma_E)} w_e. \tag{24}$$

Hence, from (23) and (24) we have $\text{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) \leq \text{LB}(G_E^d, G_I^{u_1}, G_I^{u_2})$ and provided that $LB(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2})$ from (21), we have: $\forall c, d \in \mathcal{D}_E, \; c \sqsubseteq d \Rightarrow \text{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) \leq \text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2})$.

$\square$

Proof (Proposition 2) This proof is similar to the proof of *Proposition 1*. For the sake of brevity, we give a *proof sketch*. By a direct application of Lemma 1, it is clear that for any d s.t. $|G_E^d| \geq \sigma_E$,

$$\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{UB}(G_E^d, G_I^{u_1}, G_I^{u_2}). \tag{25}$$

We have that UB is anti-monotonic w.r.t. $\sqsubseteq$ of $\mathcal{D}_E$. i.e.

$$c \sqsubseteq d \Rightarrow \text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) \geq \text{UB}(G_E^d, G_I^{u_1}, G_I^{u_2}). \tag{26}$$

This results from $c \sqsubseteq d \Rightarrow G_E^d \subseteq G_E^c$. Thus,

$$\sum_{e \in M(G_E^c, \sigma_E)} w_e \times \alpha(e) \geq \sum_{e \in M(G_E^d, \sigma_E)} w_e \times \alpha(e) \ and \ \sum_{e \in mw(G_E^c, \sigma_E)} w_e \leq \sum_{e \in mw(G_E^d, \sigma_E)} w_e.$$

Hence, given (25) and (26) it follows that:

$$\forall c, d \in \mathcal{D}_E. \ c \sqsubseteq d \Rightarrow \text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}).$$

$\square$

Proof (Proposition 3) given $c, d \in \mathcal{D}_E$ such that $c \sqsubseteq d$, using *proposition 1* we have:

$$\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}),$$

$$\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) \leq \text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}).$$

Since $\varphi_{\text{consent}}(G_E^d, G_I^{u_1}, G_I^{u_2}) = \max(\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}), 0)$ thus $\varphi_{\text{consent}}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{oe}_{\text{consent}}(G_E^c, G_I^{u_1}, G_I^{u_2})$.

Similarly we have:

$$\text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) \geq \text{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}),$$

$$\text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq \text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) - \text{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}).$$

Since $\varphi_{\text{dissent}}(G_E^d, G_I^{u_1}, G_I^{u_2}) = \max(\text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E^d, G_I^{u_1}, G_I^{u_2}), 0)$ we get $\varphi_{\text{dissent}}(G_E^d, G_I^{u_1}, G_I^{u_2}) \leq oe_{dissent}(G_E^c, G_I^{u_1}, G_I^{u_2})$.

$\square$

Proof (Proposition 4) Given that $\forall (e, G_I^{u_1}, G_I^{u_2}) \in E \times 2^I \times 2^I : w(e, G_I^{u_1}, G_I^{u_2}) = 1$, we have for any $c \in \mathcal{D}_E$ s.t. $|G_E^c| \geq \sigma_E$,

$$\text{IAS}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \frac{\sum\limits_{e \in G_E^c} \alpha(e)}{|G_E^c|} \ \ and \ \ \text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \frac{\sum\limits_{e \in M(G_E^c, \sigma_E)} \alpha(e)}{\sigma_E}.$$

It follows from the fact that $M(G_E^c, \sigma_E) \subseteq G_E^c$:

$$\exists S \subseteq G_E^c, \ \text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \text{IAS}(S, G_I^{u_1}, G_I^{u_2}),$$
$$\text{UB}(G_E^c, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) =$$
$$\text{IAS}(S, G_I^{u_1}, G_I^{u_2}) - \text{IAS}(G_E, G_I^{u_1}, G_I^{u_2}),$$
$$\text{oe}_{\text{consent}}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \varphi_{\text{consent}}(S, G_I^{u_1}, G_I^{u_2}).$$

The subset S being for example the set $M(G_E^c, \sigma_E)$ itself. The same reasoning applies when considering $\text{oe}_{\text{dissent}}$. In this case we consider the lower bound LB. We have:

$$\text{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \frac{\sum_{e \in m(G_E^c, \sigma_E)} \alpha(e)}{\sigma_E}.$$

Given that $m(G_E^c, \sigma_E) \subseteq E$, we have:

$$\exists S \subseteq G_E^c, \mathrm{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \mathrm{IAS}(S, G_I^{u_1}, G_I^{u_2}),$$
$$\mathrm{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) - \mathrm{LB}(G_E^c, G_I^{u_1}, G_I^{u_2}) =$$
$$\mathrm{IAS}(G_E, G_I^{u_1}, G_I^{u_2}) - \mathrm{IAS}(S, G_I^{u_1}, G_I^{u_2}),$$
$$\mathrm{oe}_{\mathrm{dissent}}(G_E^c, G_I^{u_1}, G_I^{u_2}) = \varphi_{\mathrm{dissent}}(S, G_I^{u_1}, G_I^{u_2}).$$

This proves that, if IAS is a simple mean, for any $c \in \mathcal{D}_E$ s.t. $|G_E^c| \geq \sigma_E$:

$$\exists S, S' \subseteq G_E^c \begin{cases} \mathrm{oe}_{\mathrm{consent}}(G_E^c, G_I^{u_1}, G_I^{u_2}) &= \varphi_{\mathrm{consent}}(S, G_I^{u_1}, G_I^{u_2}) \\ \mathrm{oe}_{dissent}(G_E^c, G_I^{u_1}, G_I^{u_2}) &= \varphi_{\mathrm{dissent}}(S', G_I^{u_1}, G_I^{u_2}) \end{cases}.$$

Hence $\mathrm{oe}_{\mathrm{consent}}$ and $oe_{dissent}$ are tight optimistic estimates for respectively $\varphi_{\mathrm{consent}}$ and $\varphi_{\mathrm{dissent}}$ if the underlying IAS is a simple average. □

Before giving the proof of the proposition 5 we present the following lemma.

Lemma 2 *The sums of the number of all descriptions covering each record in G is equal to the sum of the supports of all descriptions in $\mathcal{D}$. That is:*

$$\sum_{g \in G} |\downarrow \delta(g)| = \sum_{d \in \mathcal{D}} |G^d|.$$

Proof (Lemma 2) For $g \in G$, we have $\downarrow \delta(g) = \{d \in \mathcal{D} \text{ s. t. } d \sqsubseteq \delta(g)\}$ and for $d \in \mathcal{D}$, we have $G^d = \{g \in G \text{ s. t. } d \sqsubseteq \delta(g)\}$. Let us define the indicator function on $\mathcal{D} \times G$:

$$\mathbb{1}_{\sqsubseteq}(d, g) = \begin{cases} 1 & \text{if } d \sqsubseteq \delta(g) \\ 0 & \text{else} \end{cases}.$$

Hence, we have $|\downarrow \delta(g)| = \sum_{d \in \mathcal{D}} \mathbb{1}_{\sqsubseteq}(d, g)$ and $|G^d| = \sum_{g \in G} \mathbb{1}_{\sqsubseteq}(d, g)$ thus:

$$\sum_{g \in G} |\downarrow \delta(g)| = \sum_{g \in G} \sum_{d \in \mathcal{D}} \mathbb{1}_{\sqsubseteq}(d, g) = \sum_{d \in \mathcal{D}} \sum_{g \in G} \mathbb{1}_{\sqsubseteq}(d, g) = \sum_{d \in \mathcal{D}} |G^d|.$$

□

Proof (Proposition 5) We denote by **gs** the random record drawn in line 1 and by **ds** the random description drawn in line 2 of FBS.

$$\mathbb{P}(\mathbf{ds} = d) = \sum_{g \in G} \mathbb{P}((\mathbf{gs} = g)(\mathbf{ds} = d|g))$$
$$= \sum_{g \in G^d} \frac{1}{|\downarrow \delta(g)|} \times \underbrace{\frac{|\downarrow \delta(g)|}{\sum_{i \in G} |\downarrow \delta(i)|}}_{\text{weight } w_g \text{ normalized}} = \frac{|G^d|}{\sum_{g \in G} |\downarrow \delta(g)|}.$$

It follows that from *Lemma 2* that $\mathbb{P}(\mathbf{ds} = d) = \dfrac{|G^d|}{\sum_{d' \in \mathcal{D}} |G^{d'}|}$. □

Proof (Proposition 6) Given *Proposition 5*, it is clear that $\forall p \in \mathcal{P} : p = (c, u_1, u_2)$ *satisfies* $\mathcal{C} \Rightarrow \mathbb{P}(p) = \frac{|\mathrm{ext}(p)|}{Z} > 0.$ with $|\mathrm{ext}(p)| = |G_E^c| \times |G_I^{u_1}| \times |G_I^{u_2}|$ and $Z = \sum_{p' \in \mathcal{P}} |\mathrm{ext}(p')|$ a normalizing factor. □