# K-partitioning with imprecise probabilistic edges

Tom Davot, Sébastien Destercke, David Savourey

# K-partitioning with imprecise probabilistic edges

Tom Davot, Sébastien Destercke, David Savourey

**Abstract** Partitioning a set of elements into disjoint subsets is a common problem in unsupervised learning (clustering) as well as in networks (e.g., social, ecological) where one wants to find heterogeneous subgroups such that the elements within each subgroup are homogeneous. In this paper, we are concerned with the case where we imprecisely know the probability that two elements should belong to the same partition, and where we want to search the set of most probable partitions. We study the corresponding algorithmic problem on graphs, showing that it is difficult, and propose heuristic procedures that we test on data sets.

## 1 Introduction

Partitioning a set of elements into heterogeneous groups such that elements within each group are as homogeneous as possible is a common task. It is at the very core of unsupervised learning and clustering problems, as well as when one considers networks of different kinds (e.g., social, voting, . . . ).

A natural way to encode the relations existing between elements is through graphs, where the presence of an edge indicates that elements should be grouped together. However, the existence of such a link may be subject to various uncertainties. For instance, if one thinks of grouping persons (e.g., in parliament) voting in the same way, it may be that we rarely observe two persons voting at the same time, or that two persons do not always have the same behaviour (sometimes voting in the same way, sometimes not). Imprecise probabilities offer a rich and natural model to describe this uncertainty.

However, once one has modelled link uncertainty by imprecise probabilities, it remains to infer what are the more likely clusters. In this paper, we study the problem of extracting possibly optimal clusters from imprecise

Université de Technologie de Compiègne, CNRS, Heudiasyc (Heuristics and Diagnosis of Complex Systems), CS 60319 - 60203 Compiègne Cedex,
e-mail: surname.name@hds.utc.fr

probabilistic graphs. We show that solving this problem exactly is NP-hard, and propose some heuristics. While there are existing approaches trying to extract partial clusters from imprecise probabilistic knowledge (Denoeux and Kanjanatarakul, 2016; Masson et al., 2020), to our knowledge this is the first paper to view the problem as a robust decision-making one.

## 2 Notations and problem definition

Let $G$ be a simple loopless graph. We denote $V(G)$ and $E(G)$ the set of vertices and edges of $G$, respectively (or simply $V$ or $E$ if no ambiguity occurs). The *complement graph* of $G$, denoted $\bar{G}$ is the graph defined by $V(\bar{G}) = V(G)$ and $E(\bar{G}) = \{uv \mid uv \notin E(G)\}$. A *cluster graph* is a disjoint union of complete graphs [1], called *cliques*. A $k$-cluster graph is a cluster graph that contains $k$ non-empty connected components. Let $G$ be a complete graph, a $k$*-partition* of $G$ is a $k$-cluster subgraph $G'$ of $G$ such that $V(G) = V(G')$.

In this paper, an *imprecise probability* $p$ is an interval $[\underline{p}, \overline{p}] \subseteq [0,1]$ of probabilities, and $p$ is called *precise* if $\overline{p} = \underline{p}$. An *imprecise probabilistic graph* $(G, \mathcal{P})$ is a graph with a function $\mathcal{P}$ that associates to each edge in the graph an imprecise probability. If $uv$ is an edge, we denote $\underline{p}_{uv}$ and $\overline{p}_{uv}$ the lower and upper bounds of $\mathcal{P}(uv)$, respectively. The probability bounds of an absence of an edge can be deduced by duality (*i.e.* $[1-\overline{p}_{uv}, 1-\underline{p}_{uv}]$). $p_{uv}$ being the marginal probability $uv$ is an edge (and $1-p_{uv}$ that it is not), we only assume $[\underline{p}_{uv}, \overline{p}_{uv}] \subseteq [0,1]$.

A *probability realisation* $R : E(G) \mapsto [0,1]$ of $\mathcal{P}$ is a function that associates to each edge $uv$ a probability within $[\underline{p}_{uv}, \overline{p}_{uv}]$. We denote $\mathcal{R}_{\mathcal{P}}$ the set of probability realisations of $\mathcal{P}$. Let $G'$ be a subgraph of $G$ and $R \in \mathcal{R}_{\mathcal{P}}$ be a probability realisation. The probability of $G'$ under $R$, denoted $R(G')$ corresponds to

$$R(G') = \prod_{uv \in E(G')} R(uv) \prod_{uv \notin E(G')} 1 - R(uv).$$

Let $G_1$ and $G_2$ be two vertices $k$-partitions of $G$. We say that $G_1$ is *certainly more probable* than $G_2$, denoted by $G_1 \succ_p G_2$, if

$$\forall R \in \mathcal{R}_{\mathcal{P}}, R(G_1) - R(G_2) > 0.$$

Let ${}^{G_1}/_{G_2}$ denote the following value

$${}^{G_1}/_{G_2} = \prod_{uv \in E(G_1) \setminus E(G_2)} \frac{\underline{p}_{uv}}{1-\overline{p}_{uv}} \prod_{uv \in E(G_2) \setminus E(G_1)} \frac{1-\overline{p}_{uv}}{\underline{p}_{uv}}.$$

[1] A complete graph is a simple graph where every pair of vertices is connected.

Notice that if an edge $uv$ belongs (resp. does not belong) to both $G_1$ and $G_2$, then the factor $R(uv)$ (resp. $1 - R(uv)$) is present on both sides of the substraction. Thus, to verify if $G_1$ is certainly more probable than $G_2$, we only need to consider edges in $E(G) \setminus (E(G1) \cap E(G_2))$. Moreover, by duality $R(G_1) - R(G_2)$ is minimum if for every edge $uv \in E(G_1) \setminus E(G_2)$ (resp. $uv \in E(G_2) \setminus E(G_1)$), we have $R(uv) = \underline{p_{uv}}$ (resp. $R(uv) = \overline{p_{uv}}$). Hence, we have the following property.

*Property 1* Given two $k$-partitions $G_1$ and $G_2$ of the imprecise probabilistic graph $(G, \mathcal{F})$, we have

$$G_1 \succ_p G_2 \Leftrightarrow {}^{G_1}/_{G_2} > 1.$$

Notice that the order given by $\succ_p$ is partial since we may have $G_1 \not\succ_p G_2$ and $G_2 \not\succ_p G_1$. Given a constant $k$, we are then interested in finding the most probable $k$-partitions of $G$. Let $P_k(V(G))$ be the set of $k$-partitions of $G$. We define $\mathcal{M}_{G,k} = \{G \in P_k(V(G)) \mid \nexists G' \in P_k(V(G)), G' \succ_p G\}$ the set of *non-dominated* $k$-partitions under $\succ_p$. In the following, we are interested in enumerating every partition of $\mathcal{M}_{G,k}$. Hence, we define the following problem.

Most probable $k$-partitions ($k$-MPP)
**Input** A complete graph $G$ and an integer $k$.
**Output** Enumeration of $\mathcal{M}_{G,k}$.

# 3 Analysis

## 3.1 Computational Complexity

We first show that finding one element of $\mathcal{M}_{G,k}$ is NP-hard, even if $k = 2$ and $\mathcal{R}_{\mathcal{P}}$ has one element. To do so, we construct in the following way a reduction from the Max Cut (Karp, 1972) problem (that aims at finding a spanning bipartite subgraph with a maximum number of edges in a graph $H$).

**Construction 1** *Given $H$ an instance of* Max Cut*, we construct an imprecise probabilistic graph $(G, \mathcal{P})$ such that:*

- $V(G) = V(H)$,
- *for each pair of vertices $u$ and $v$, $\underline{p}_{uv} = \overline{p}_{uv} = 0.1$ (*red edges*) if $uv \in E(H)$ and, $\underline{p}_{uv} = \overline{p}_{uv} = 0.5$ (*blue edges*), otherwise.*

The proof idea is that a 2-partition is non-dominated if and only if it contains a minimum number of red edges, and thus its complement graph is a bipartite graph with a maximum number of edges. Hence, we can show the following.

**Theorem 1** *Let $(G, \mathcal{P})$ be an imprecise probabilitstic graph. Computing any element of $\mathcal{M}_{G,k}$ is NP-hard, even if $k = 2$ and $\mathcal{P}$ is a singleton.*

***Proof*** Let $H$ be an instance of Max Cut and let $G$ the graph resulting from Construction 1. First, let $G'$ be a two 2-partition of $\mathcal{M}_{G,2}$. Let $r =$

$|E(H) \cap E(G')|$ be the number of red edges in $G'$. The probability of $G'$, under any realisation, is equal to

$$P(G') = 0.5^{|E(\bar{H})|} \cdot 0.1^{r} \cdot 0.9^{|E(H)|-r}.$$

Thus, a 2-partition belongs to $\mathcal{M}_{G,2}$ if and only if it contains a minimum number of red edges. Later, we now show that $G'$ contains a 2-partition with $k$ red edges if and only if there is a bipartite subgraph of $H$ with $|E(H)| - r$ edges.

- Let $G'$ be a 2-partition of $G$ containing $k$ red edges. Notice that $\bar{G}'$ is a bipartite graph and that by duality its contains a $|E(H)| - r$ red edges. Hence, since the red edges correspond to the edges of $H$, the graph $H'$ defined by $V(H') = V(H)$ and $E(H') = E(\bar{G}) \cap E(H)$ contains $|E(H) - r|$ edges.
- Let $H'$ be a bipartite subgraph of $H$ containing $|E(H)| - r$ edges. Let $X$ and $Y$ be the bipartition of $H'$. Let $G'$ be the 2-partition of $G$ such that $G' = G[X] \cup G[Y]$. The number of red edges in $G'$ is equal to $r$.

Hence, computing a non-dominated 2-partition of $G$ is equivalent to compute an optimal solution for Max Cut. □

### 3.2 Easy cases

In this section we present three easy cases in which some element of $k$-MPP can be polynomially computed in the size of graph. These easy cases appear when one value appears in every probabilistic interval of $\mathcal{P}$. The first case is when 0.5 is contained in every probabilistic interval which implies that any $k$-partition is non-dominated.

**Theorem 2** *Let $(G, \mathcal{P})$ be an imprecise probabilistic graph such that $\forall uv \in E(G), 0.5 \in \mathcal{P}(uv)$. We have $\mathcal{M}_{G,k} = P_k(V)$.*

***Proof*** Let $R$ be the probability realisation of $\mathcal{P}$ such that $R(uv) = 0.5$ for any edge $uv$. Notice that for any $k$-partition $G'$ we have $R(G') = 0.5^{|E(G)|}$. Hence, there is a probabilistic realisation for which all $k$-partitions have the same probability and therefore, any $k$-partition is non-dominated. □

The second case is when a value inferior to 0.5 is contained in every probabilistic interval. In that case, every $k$-partition that contains a minimum number of edges (i.e., is balanced) is non-dominated.

**Theorem 3** *Let $x < 0.5$ and $(G, \mathcal{P})$ be an imprecise probabilistic graph such that $\forall uv \in E(G), x \in \mathcal{P}(uv)$. Let $G'$ be a $k$-partition of $G$ with connected components $\{V_1, \ldots, V_k\}$ of respective orders $n_1, \ldots, n_k$. If we have*

$$\forall i, j, |n_i - n_j| \leq 1$$

*then, $G' \in \mathcal{M}_{G,k}$.*

***Proof*** Let $R$ be the probability realisation of $\mathcal{P}$ such that $R(uv) = x$ for any edge $uv$. First, note that for any $k$-partition $G'$ of $G$ we have $R(G') = x^{|E(G')|} \cdot (1-x)^{E(G)-|E(G')|}$. Hence, since $x < 0.5$, $G'$ is more probable under $R$ if $|E(G')|$ is minimum. Toward a contradiction, suppose there is a $k$-partition $G'$ with a minimum number of edges and such that $G'$ has two cliques $V_i$ and $V_j$ such that $|V_i| - |V_j| > 1$ (we assume without loss of generality that $|V_i| < |V_j|$). Let $v_j$ be a vertex of $V_i$. Let $H$ be the $k$-partition obtained by replacing $V_i$ and $V_j$ by $V_i \cup \{v_j\}$ and $V_j \setminus \{v_j\}$. We have

$$|E(G')| - |E(H)| = |E(V_j)| - |E(V_j \setminus \{v_j\})| + |E(V_i)| - |E(V_i \cup \{v_j\})|$$
$$|E(G')| - |E(H)| = |V_j| - 1 - |V_i|$$
$$|E(G')| - |E(H)| > 1.$$

Thus, $G'$ is not a $k$-partition with a minimum number of edges.

Further, let $G'$ be a $k$-partition of $G$ with connected components $\{V_1, \ldots, V_k\}$ of respective orders $n_1, \ldots, n_k$ and such that $\forall i, j, |n_i - n_j| \leq 1$. Since $G'$ has a minimum number of edges, $G'$ is non-dominated under $R$ and thus, it belongs $\mathcal{M}_{G,k}$ . □

Finally, the last case is when a value greater to 0.5 is contained in every probabilistic interval. In that case, every $k$-partition that contains a maximum number of edges (i.e., is unbalanced) is non-dominated.

**Theorem 4** *Let $x > 0.5$ and $(G, \mathcal{P})$ be an imprecise probabilistic graph such that $\forall uv \in E(G), x \in \mathcal{P}(uv)$. Let $G'$ be a $k$-partition of $G$ with connected components $\{V_1, \ldots, V_k\}$ such that $\forall i < j, |V_i| \leq |V_j|$. If we have*

*1.* $|V_i| = 1, \forall i < k$, *and*
*2.* $|V_k| = |V(G)| - k + 1$.

*then,* $G' \in \mathcal{M}_{G,k}$.

***Proof*** Let $R$ be the probability realisation of $\mathcal{P}$ such that $R(uv) = x$ for any edge $uv$. First, note that for any $k$-partition $G'$ of $G$ we have $R(G') = x^{|E(G')|} \cdot (1-x)^{E(G)-|E(G')|}$. Hence, since $x < 0.5$, $G'$ is more probable under $R$ if $|E(G')|$ is maximum. A $k$-partition has a maximum number of edges if and only if every clique but one is constitued of one vertex, that is, if it respects $(a)$ and $(b)$. Let $G'$ be such $k$-partition. Since $G'$ is non-dominated under $R$, it belongs to $\mathcal{M}_{G,k}$ . □

## 4 Heuristic

In this section, we describe some heuristic method used to approach $\mathcal{M}_{G,k}$ or to improve the computation time. This method relies on the use of a pattern and some associated reductions rules.

### 4.1 Pattern and Reduction Rules

Let $(G, \mathcal{F})$ be an imprecise probabilistic graph. A *pattern* $X$ of $G$ is a subset of edges. We say that a $k$-partition $G'$ respects a pattern $X$ if $G'$ contains every edge of $X$ (*i.e.* $X \subset E(G')$). We denote $P_k(V(G), X)$ the set of $k$-partitions that respects $X$. Let $\mathcal{M}_{G,k}(X) = \{G \in P_k(V(G), X) \mid \nexists G' \in P_k(V(G), X), G' \succ_p G\}$ the set of non-dominated $k$-partitions respecting $X$. We give some reduction rules that reduce the size of $G$ without altering the computation of $\mathcal{M}_{G,k}(X)$.

Let $uv$ be an edge of $X$ and $x$ be any vertex. Note that for any $k$-partition $G' \in P_k(V(G), X)$, $u$ and $v$ are contained in the same clique and $G'$ contains either both $xu$ and $xv$ or none of them. Hence, $u$ and $v$ are acting like a single vertex and thus, we can contract $uv$ into a single vertex and merge $xu$ and $xv$ together. Formally, the *contraction* of $uv$, denoted $f_{uv}$, is the application which given any graph $G$, constructs the graph $H$ where:

- $V(H) = (V(G) \setminus \{uv\}) \cup \{w\}$,
- $E(H) = (E(G) \setminus \{xy \mid \forall xy \in E(G), xy \cap uv \neq \emptyset\}) \cup \{xw \mid \forall x \in V(G) \setminus \{uv\}\}$.

The contraction rule uses $f_{uv}$ and adapts the imprecise probability set and the pattern so that the sets of non-dominated $k$-partitions respecting the pattern are equivalent in the original graph and the newly created graph.

**Rule 1 (Contraction rule)**

*Let $(G, \mathcal{F})$ be an imprecise probabilistic graph and let $X = (C, A)$ be a pattern. Let $uv$ be an edge of $C$. We reduce $G$ to the following imprecise probabilistic graph $(H, \mathcal{G})$.*

- $H = f_{uv}(G)$,
- *for any edge $xy$ of $E(G)$ such that $xy \cap uv = \emptyset$, we set $\mathcal{G}(xy) = \mathcal{F}(xy)$, and*
- *for any vertex $x \notin uv$ of $V(G)$, we set*

$$\mathcal{G}(xw) = [\frac{\underline{p}_{xu} \cdot \underline{p}_{xv}}{\underline{p}_{xu} \cdot \underline{p}_{xv} + (1 - \underline{p}_{xu}) \cdot (1 - \underline{p}_{xv})}, \frac{\bar{p}_{xu} \cdot \bar{p}_{xv}}{\bar{p}_{xu} \cdot \bar{p}_{xv} + (1 - \bar{p}_{xu}) \cdot (1 - \bar{p}_{xv})}].$$

*This corresponds to compute bounds over the conditional probability $P(xu \wedge xv | (xu \wedge xv) \vee (\neg xu \wedge \neg xv))$. We construct a new pattern $X'$ for $H$ as follows.*

- *Let $T_X = \{xy \mid xy \in X, xy \cap uv \neq \emptyset\}$ and $R_X = \{xw \mid \exists xy \in X, y \in uv\}$. We set $X' = (X \setminus T_X) \cup R_X$.*

An example of the application of Rule 1 is depicted in Figure 1. The two next properties show that Rule 1 is safe, that is, the computation of $\mathcal{M}_{G,k}(X)$ is equivalent to the computation of $\mathcal{M}_{H,k}(Y)$.

*Property 2* $f_{uv}$ is a bijection from $P_k(V(G), X)$ to $P_k(V(H), Y)$.

***Proof*** First, we show that for any $G' \in P_k(V(G), X)$, we have $f_{uv}(G') \in P_k(V(H), Y)$. Let $H' = f_{uv}(G')$. Let $V_{uv}$ be the clique containing $uv$ (which exists since $uv \in C$). Since contracting an edge in a clique leads to another

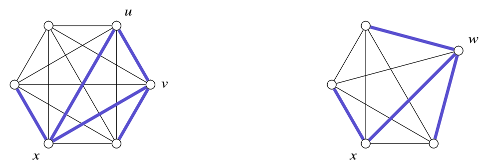


**Fig. 1** Example of an application of Rule 1 on the edge $uv$ in an imprecise probabilistic graph with a motif $X$ (blue edges).

clique, $f_{uv}(V_{uv})$ is a clique. Moreover, since any other connected component remains unchanged by $f_{uv}(G')$, $H'$ is a $k$-partition of $H$. For any edge $xw$ in $R_c$ (resp. $R_A$), since $xu$ or $xv$ belongs (resp. does not belong) to $C$ (resp. $A$) then the vertex $x$ belongs (resp. does not belong) to $V_{uv}$ and thus, the edge $xu$ belongs (resp. does not belong) to $E(H')$. Moreover, since $C' \setminus R_c \subseteq C$ (resp. $A' \setminus R_A \subseteq A$), any edge $e \in C' \setminus R_c$ (resp. $e \in A' \setminus R_A$) belongs (resp. does not belong) to $G'$ and since $e$ is not altered by $f_{uv}$, we have $e \in E(H')$ (resp. $e \notin E(H')$). Hence, $C' \subseteq E(H')$ and $A' \cap E(H') = \emptyset$, that is, $H'$ is a $k$-partition that respects $Y$.

Further, we show that $f_{uv}$ is surjective. Let $H'$ be any $k$-partition of $P_k(V(H), X')$. Let $V_w$ be the clique containing $w$. Since splitting a vertex in two in a clique leads to another clique, then $f_{uv}^{-1}(V_w)$ is also a clique. Moreover, since any other connected component remains unchanged by $f_{uv}^{-1}$, then $G' = f_{uv}^{-1}(H')$ is a $k$-partition of $(G, \mathcal{F})$ that contains $uv$. For any vertex $y \in uv$ and for any edge $xy$, in $T_c \setminus \{uv\}$ (resp. $T_A$), since $xw \in C'$ (resp. $xw \in A'$) then the vertex $x$ belongs (resp. does not belong) to $f_{uv}^{-1}(V_w)$ and thus $xy \in E(G')$ (resp. $xy \notin E(G')$). Moreover, since $C \setminus T_c \subseteq C'$ (resp. $A \setminus T_A \subseteq A'$), any edge $e \in C \setminus T_c$ (resp. $e \in A \setminus T_A$) belongs to (resp. does not belong to) $E(H')$ and since $e$ is not altered by $f^{-1}$, we have $e \in E(G')$ (resp. $e \notin E(G')$). Hence, $C \subseteq E(G')$ and $A \cap E(G') = \emptyset$, that is, $G'$ is a $k$-partition that respects $X$.

Finally, we show that $f_{uv}$ is injective. Let $G_1$ and $G_2$ be two $k$-partitions of $P_k(V(G), X)$. Let $xy$ be an edge such that $xy \in E(G_1)$ and $xy \notin E(G_2)$. If $xy \cap uv = \emptyset$, then $xy$ is not altered by $f_{uv}$ and thus, $xy \in E(f_{uv}(G_1))$ and $xy \notin E(f_{uv}(G_2))$. Otherwise, without loss of generality, suppose that $x \notin uv$. Since $uv \in C$, $u, v$ and $x$ belong to the same clique in $G_1$ and there is a clique in $f(G_1)$ that contains $x$ and $w$. Moreover, $u, v$ are not in the same clique as $x$ in $G_2$, and thus, $w$ is not in the same clique as $x$ in $f(G_2)$. Hence, we have $f(G_1) \neq f(G_2)$. □

*Property 3* Given two $k$-partitions $G_1$ and $G_2$ in $P_k(V(G), X)$ we have

$$G_1 \succ_p G_2 \Leftrightarrow f_{uv}(G_1) \succ_g f_{uv}(G_2).$$

***Proof*** Let $X = (C, A)$ be a pattern of $(G, \mathcal{F})$ and let $G_1$ and $G_2$ be two k-partitions in $P_k(V(G), X)$. Let $V_{uv}^1$ and $V_{uv}^2$ be the cliques containing $uv$ in $G_1$

and $G_2$, respectively (which exist since $uv \in C$). If $V_{uv}^1 = V_{uv}^2$ then, ${V_{uv}^1}/{V_{uv}^2} = 1$. Since any edge that does not belong to $V_{uv}^1$ or $V_{uv}^2$ are not altered by $f_{uv}$, we have ${G_1}/{G_2} = {f_{uv}(G_1)}/{f_{uv}(G_2)}$. Thus, $G_1 \succ_p G_2 \Leftrightarrow f_{uv}(G_1) \succ_p f_{uv}(G_2)$.

Now, suppose that $V_{uv}^1 \neq V_{uv}^2$. For any edge $xw$ such that $xw \in f_{uv}(V_{uv}^1)$ and $xw \notin f_{uv}(V_{uv}^2)$, we have $\{xu, xv\} \subseteq E(G_1)$ and $\{xu, xv\} \cap E(G_2) = \emptyset$. We have

$$\begin{aligned} \frac{\underline{g}_{xw}}{1 - \underline{g}_{xw}} &= \frac{\frac{\underline{p}_{xu} \cdot \underline{p}_{xv}}{\underline{p}_{xu} \cdot \underline{p}_{xv} + (1 - \underline{p}_{xu}) \cdot (1 - \underline{p}_{xv})}}{1 - \frac{\underline{p}_{xu} \cdot \underline{p}_{xv}}{\underline{p}_{xu} \cdot \underline{p}_{xv} + (1 - \underline{p}_{xu}) \cdot (1 - \underline{p}_{xv})}} \\ &= \frac{\underline{p}_{xu} \cdot \underline{p}_{xv}}{(1 - \underline{p}_{xv}) \cdot (1 - \underline{p}_{xu})}. \end{aligned}$$

Samewise, for any edge $xw$ such that $xw \notin f_{uv}(V_{uv}^1)$ and $xw \in f_{uv}(V_{uv}^2)$, we have $\{xu, xv\} \cap E(G_1) = \emptyset$ and $\{xu, xv\} \subseteq E(G_2)$. We have

$$\begin{aligned} \frac{1 - \bar{g}_{xw}}{\bar{g}_{xw}} &= \frac{1 - \frac{\bar{p}_{xv} \cdot \bar{p}_{xu}}{\bar{p}_{xu} \cdot \bar{p}_{xv} + (1 - \bar{p}_{xu}) \cdot (1 - \bar{p}_{xv})}}{\frac{\bar{p}_{xu} \cdot \bar{p}_{xv}}{\bar{p}_{xu} \cdot \bar{p}_{xv} + (1 - \bar{p}_{xu}) \cdot (1 - \bar{p}_{xv})}} \\ &= \frac{(1 - \bar{p}_{xu}) \cdot (1 - \bar{p}_{xv})}{\bar{p}_{xu} \cdot \bar{p}_{xv}}. \end{aligned}$$

Hence, since any other edge in $V_{uv}^1$ or $V_{uv}^2$ is not altered by $f_{uv}$, we have ${f_{uv}(G_1)}/{f_{uv}(G_2)} = {G_1}/{G_2}$. Hence, $G_1 \succ_p G_2 \Leftrightarrow f_{uv}(G_1) \succ_p f_{uv}(G_2)$. □

## 4.2 Algorithm Description

Our method to approach $\mathcal{M}_{G,k}$ is described by Algorithm 1. The basic idea is to reduce the size of the graph by computing a pattern $X$ and applying the reduction rules described above. Once the size of the reduced graph is small enough, it seems possible to enumerate every $k$-partitions and thus compute $\mathcal{M}_{G,k}(X)$ in a relatively small amount of time. The difficulty of this method is to find a pattern $X$ such that $|\mathcal{M}_{G,k}(X) \Delta \mathcal{M}_{G,k}|$ is minimum. In the next section, we take the pattern $X_i = \{uv \mid \underline{p}_{uv} \geq 0.9\}$.

**Algorithm 1:** Heuristic method

**Data:** An imprecise probabilistic graph $(G, \mathcal{F})$ and two integers $t$ and $k$.
**Result:** A set of $k$-partitions for $G$.

**forall** $i \leq t$ **do**
  Compute a feasible pattern $X_i$ for $(G, \mathcal{F})$;
  Apply the contraction rule until $X_i$ is empty;
**end**
$X \leftarrow \bigcup_{i \leq k} X_i$;
Enumerate every possible $k$-partitions of $G$ to compute $M \leftarrow \mathcal{M}_{G,k}(X)$;
**return** $M$;

# 5 Numerical Experiments

In the following, we provide some tests for the special case where $k = 2$.

## 5.1 Dataset

For our tests, we use two types of dataset, one using some real data and one using some generated instances.

**Real dataset:** we use the dataset used by Arinik et al. (2017) that contains vote information for French and Italian members of the european parlement. Each member is represented by a vertex in the graph. To generate the imprecise probability sets, we use the following formula. Let $u$ and $v$ be two members. Let $k$ be the number of sessions in which both $u$ and $v$ participated and let $t$ be the number of sessions in which $u$ and $v$ voted the same. We set $\mathcal{P}(uv) = [\frac{t}{k+s}, \frac{t+s}{k+s}]$ where $s$ is a parameter settling the speed at which the intervals converge to a precise value. In the following, we set $s = 5$. Since the generated graph contains 870 vertices, it is not possible to compute exactly the set of non-dominated 2-partitions, we create twenty subinstances by randomly drawing 15 vertices of the original graph.

**Randomly Generated Instances:** we proceed as follows. First, we define $k$ groups of vertices with a given size and an application $f : \{1, \ldots, k\}^2 \mapsto [0, 1]^4$ which associate to each pair of integers $\{i, j\}$ a tuple $\{\underline{m}_{i,j}, \bar{m}_{i,j}, \underline{\ell}_{i,j}, \bar{\ell}_{i,j}\}$. Then for each pair of vertices $u$ and $v$ such that $u$ belongs to the group $i$ and $v$ belongs to the group $j$ ($i$ can be equal to $j$), we draw two real numbers $m \in [\underline{m}_{i,j}, \bar{m}_i, j]$ and $\ell \in [\underline{\ell}_{i,j}, \bar{\ell}_{i,j}]$. Finally, we introduce the edge $uv$ with the imprecise probabilistic interval $\mathcal{F}(uv) = [max(0, m - \ell), min(1, m + \ell)]$.

We test two differents groups configurations.

- *Configuration A*. The graph contains two groups of 7 vertices, and we set $f(1, 1) = f(2, 2) = \{0.9, 0.95, 0, 0.3\}$ and $f(1, 2) = \{0.1, 0.2, 0, 0.3\}$.
- *Configuration B*. The graph contains three groups of 6 vertices, and we set $f(1, 1) = f(2, 2) = f(3, 3) = \{0.9, 0.95, 0, 0.1\}$ and $f(1, 2) = \{0.1, 0.2, 0, 0.1\}$. For the values of $f(1, 3)$ and $f(2, 3)$, we test three differents variations.
  - B1: $f(1, 3) = f(2, 3) = \{0.45, 0.55, 0, 0.1\}$,
  - B2: $f(1, 3) = f(2, 3) = \{0.45, 0.55, 0.2, 0.0.35\}$,
  - B3: $f(1, 3) = f(2, 3) = \{0.45, 0.55, 0.3, 0.7\}$.

For each configuration and each variation, we generate twenty instances.

## 5.2 Results

The tests were run on a personal laptop with 16Go of RAM and with an Intel Core 7 processor 2.5GHz. Results are displayed in Table 1. In our tests, we compare two exacts algorithms with our heuristic with two different values of $t$. The "brute" version enumerate every 2-partitions to construct the solution set $S$. For each enumerated 2-partition $G'$, if $G'$ is not dominated by another 2-partition of $S$, then $G'$ is added to $S$ and every 2-partition of $S$ that is dominated by $G'$ is removed from $S$. The "init" version does the same thing but $S$ is initialized by a set of 2-partitions. To initialize $S$, we take the value returned by Algorithm 1 with $t = 1$. The idea behind the “init” version is that initializing $S$ with a set of possibly non-dominated solutions can reduce computational time (similar ideas can be found in Nakharutai et al. (2019)).For every configuration, the results correspond to the average of the twenty instances.

**Table 1** Results on real and generated data for the 2-MPP problem. The columns “Brute” and “Init” correspond to exact algorithms. The columns “t=1” and “t=2” correspond to Algorihm 1 with two different values of $t$. “Sol” is the size of the enumerated 2-partitions, “$n$” is the size of the graph after the application of the reduction rules. “ER1” is the percentage of enumerated 2-partitions that do not belong to $\mathcal{M}_{G,k}$. “ER2” is the percentage of 2-partitions that belong to $\mathcal{M}_{G,k}$ and that are not enumerated. Time is given in seconds.

| Config | Brute | | Init | t=1 | | | | | t=2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Time | Sol | Time | Time | n | Sol | ER1 | ER2 | Time | n | Sol | ER1 | ER2 |
| A | 15.5 | 1 | 0.36 | $10^{-3}$ | 9.55 | 1 | 0% | 0% | $10^{-5}$ | 4.3 | 2 | 0% | 0% |
| B1 | 1.8 | 2 | 3 | $10^{-3}$ | 6.25 | 2 | 0% | 0% | $10^{-5}$ | 3.05 | 2 | 0% | 0% |
| B2 | 296 | 2.7 | 3.05 | $10^{-3}$ | 6.4 | 2 | 0% | 17.2% | $10^{-5}$ | 3 | 2 | 0% | 17.2% |
| B3 | 5017 | 71.35 | 34.9 | $10^{-4}$ | 6.55 | 2.7 | 0% | 97% | $10^{-5}$ | 3 | 2 | 0% | 97% |
| Real | 29 | 827.35 | 28.5 | $10^{-3}$ | 6.9 | 9.15 | 30% | 99.25% | $10^{-5}$ | 3.7 | 2.15 | 37.5% | 99.5% |

We can see for the generated instances every 2-partition enumerated by the heuristic belongs to the exact solution. However, the number of 2-partitions returned by the heuristic can be relatively small compared to the size of $\mathcal{M}_{G,k}$. For example, for the B3 configuration, 97% of $\mathcal{M}_{G,k}$ is not enumerated by the heuristic. Nevertheless, the results of 1 help to drastically reduce the computation time of the exact algorithm. For instances from real data, the results are more mixed: almost all of $\mathcal{M}_{G,k}$ is not enumerated by the heuristic and at least 30% of the 2-partitions returned by the heuristic does not belong to $\mathcal{M}_{G,k}$. Moreover, since the results of the heuristic are not good enough, the computation time is not significantly reduced for the “init” version. We can explain this bad performance by the fact that drawing randomly 15 vertices in a real instance can lead to a subinstance that is not really representative since the 15 vertices can belong to the same group.

## 6 Conclusion

In this paper, we adressed the problem of the most probable $k$-partition with imprecise probabilistic edges. After some theoritical results, we developed a heuristic to tackle this problem. We show that this heuristic can have good results in practice but becomes less performant if the probability intervals are to large. A natural perspective of our work can be to find another way to compute some $k$-partitions to make the initialisation for the exact version, since we show that it can significantly reduce the computation time. It can be interesting to find another method since our heuristic can not perform well for some instances.

**Acknowledgements** This work was funded in the framework of the Labex MS2T. It was supported by the French Government, through the program "Investments for the future" managed by the ANR (Reference ANR-11-IDEX-0004-02).

## References

Arinik N, Figueiredo R, Labatut V (2017) Signed graph analysis for the interpretation of voting behavior. CoRR abs/1712.10157, URL http://arxiv.org/abs/1712.10157, 1712.10157

Denoeux T, Kanjanatarakul O (2016) Evidential clustering: a review. In: International symposium on integrated uncertainty in knowledge modelling and decision making, Springer, pp 24–35

Karp RM (1972) Reducibility among combinatorial problems. In: Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA, pp 85–103

Masson MH, Quost B, Destercke S (2020) Cautious relational clustering: A thresholding approach. Expert Systems with Applications 139:112837

Nakharutai N, Troffaes MC, Caiado CC (2019) Improving and benchmarking of algorithms for decision making with lower previsions. International Journal of Approximate Reasoning 113:91–105