Concise representation of hypergraph minimal transversals : approach and application on the dependency inference problem

Sadok Ben Yahia [1]

Nidhal Jelassi [1,2], Christine Largeron [2]

[1] University of Tunis El Manar, Tunis, Tunisia
[2] Laboratoire Hubert Curien, Saint-Étienne, France

May 2015

Motivations

- Hypergraph minimal transversals (MTs) are useful in many computer science applications [Hagen, 2008]:
  - Data mining, AI, Databases, Semantic web, etc.
- Several algorithms have been proposed in the literature.
  - Berge[1], Fredman et *al.*[2], Kavvadias et *al.*[3], MTMINER[4], etc.
  - Number of MTs may be **exponential** relative to hypergraph size.
- Our approach : An **information lossless concise** representation of MTs.
  - Notion of irredundant hypergraph.
  - Proposition of a new algorithm, IRRED-ENGINE, for the extraction of all the MTs.

[1] [Berge, 1989]
[2] [Fredman, 1996]
[3] [Kavvadias et Stavropoulos, 2005]
[4] [Hébert, 2007]

Outline

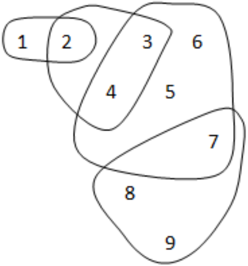
1. Key notions
2. Irredundant MTs
3. Approach and algorithm
4. Experimentations
5. Application : dependency inference problem
6. Conclusion

Outline

Hypergraph and minimal transversal

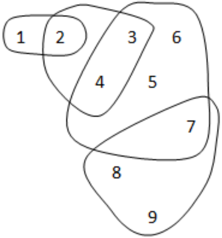
- Let $H = (\mathcal{X}, \xi)$ such as :
  - $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$: a finite set of vertices.
  - $\xi = \{e_1, e_2, \ldots, e_m\}$: a set of subsets of $\mathcal{X}$.
- $H = (\mathcal{X}, \xi)$ is a hypergraph if [Berge, 1989] :
  - $e_i \neq \emptyset$, i $\in\{1, \ldots, m\}$
  - $\bigcup_{i=1,\ldots,m} e_i = \mathcal{X}$
- $T \subset \mathcal{X}$ is a transversal of H if $T \bigcap e_i \neq \emptyset \; \forall i = 1, \ldots, m$.
- T is called minimal if $\nexists T_1 \subset T$ $s.t.$ T_1 is a transversal.
  The set of MTs is denoted $\mathcal{M}_H$.

- Let $H = (\mathcal{X}, \xi)$ such as :
  - $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$
  - $\xi = \{e_1,\ e_2,\ e_3,\ e_4\}$



$$e_1 = \{1,\ 2\}$$
$$e_2 = \{2,\ 3,\ 4\}$$
$$e_3 = \{3,\ 4,\ 5,\ 6,\ 7\}$$
$$e_4 = \{7,\ 8,\ 9\}$$

- A hypergraph :



- 15 MTs :
  - $\{2,7\}$ - $\{1,3,7\}$
  - $\{1,4,7\}$ - $\{1,3,8\}$
  - $\{1,3,9\}$ - $\{1,4,8\}$
  - $\{1,4,9\}$ - $\{2,3,8\}$
  - $\{2,4,8\}$ - $\{2,4,9\}$
  - $\{2,3,9\}$ - $\{2,5,8\}$
  - $\{2,6,8\}$ - $\{2,5,9\}$
  - $\{2,6,9\}$

Outline

Definitions and notations

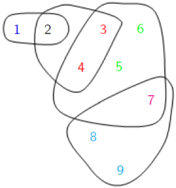
- EXTENT (x): extension of a vertex $x \in \mathcal{X}$.
  - The set of hyperedges to which belongs x.
- GENERALIZED NODE : set of vertices which have the same extension.
- REPRESENTATIVE : the first element of a generalized node, whenever its elements are sorted w.r.t lexicographic order.
  - Each GENERALIZED NODE has only one REPRESENTATIVE.

Notion of irredundance

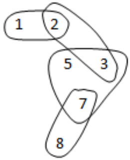
- Irredundant hypergraph : $H' = (\mathcal{X}', \xi')$ of $H = (\mathcal{X}, \xi)$.
  - $\mathcal{X}' \subseteq \mathcal{X}$: the set of the representatives of the generalized nodes associated to the vertices of H.
  - ξ': the set of hyperedges of H deprived of the elements of $\mathcal{X}$-$\mathcal{X}'$
- Irredundant minimal transversals ($\mathcal{M}_{H'}$)
  - $TM_{H'} \Leftrightarrow TMirr_H$

Study case

- Initial hypergraph H



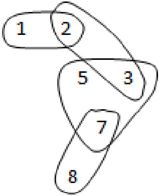
- Irredundant hypergraph H'



Vertex	Extension	G.N (Rep)
1	$\{e_1\}$	$S_1(1)$
2	$\{e_1, e_2\}$	$S_2(2)$
3	$\{e_2, e_3\}$	$S_3(3)$
4	$\{e_2, e_3\}$	$S_3(3)$
5	$\{e_3\}$	$S_4(5)$
6	$\{e_3\}$	$S_4(5)$
7	$\{e_3, e_4\}$	$S_5(7)$
8	$\{e_4\}$	$S_6(8)$
9	$\{e_4\}$	$S_6(8)$

Study case

- Hypergraph H'



- $\mathcal{M}_{H'}$:
  - $\{2, 7\}$
  - $\{1, 3, 7\}$
  - $\{1, 3, 8\}$
  - $\{2, 3, 8\}$
  - $\{2, 5, 8\}$
- $\mid \mathcal{M}_{H'} \mid = 5$
- $\mid \mathcal{M}_{H} \mid = 15$

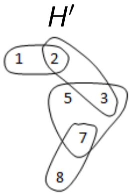
Outline

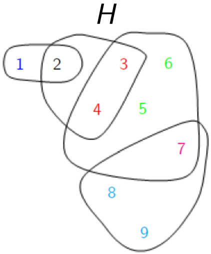
1. Key notions
2. Irredundant MTs
3. **Approach and algorithm**
4. Experimentations
5. Application : dependency inference problem
6. Conclusion

Steps

- Input : an hypergraph H
- Output : The set of MTs of H

1. Computation of all generalized nodes of H.
2. Generation of H'.
3. Extraction of MTs of H' ($\mathcal{M}_{H'}$) using the representative set.
4. Derive faithfully all the MTs of H from $\mathcal{M}_{H'}$ using a simple substitution process.





$\mathcal{M}_{H'}$	$\mathcal{M}_{H}$
$\{2,7\}$	$\{2,7\}$
$\{1,\mathbf{3},7\}$	$\{1,\mathbf{3},7\}$, $\{1,\mathbf{4},7\}$
$\{1,\mathbf{3},\underline{8}\}$	$\{1,\mathbf{3},\underline{8}\}$, $\{1,\mathbf{4},\underline{8}\}$, $\{1,\mathbf{4},\underline{9}\}$, $\{1,\mathbf{3},\underline{9}\}$
$\{2,\mathbf{3},\underline{8}\}$	$\{2,\mathbf{3},\underline{8}\}$, $\{2,\mathbf{4},\underline{8}\}$, $\{2,\mathbf{4},\underline{9}\}$, $\{2,\mathbf{3},\underline{9}\}$
$\{2,\mathbf{5},\underline{8}\}$	$\{2,\mathbf{5},\underline{8}\}$, $\{2,\mathbf{5},\underline{9}\}$, $\{2,\mathbf{6},\underline{8}\}$, $\{2,\mathbf{6},\underline{9}\}$

MTs of H' and MTs of H

Outline

Datasets

Experiments were carried out on several types of hypergraphs :

- The first one are generated from the datasets "*Accidents*" and "*Connect-4*"
- The second one are a randomly hypergraphs generated using the "*random hypergraph generator*" (Boros et *al.*,2003).

Algorithms considered

- IRRED-ENGINE
- MMCS (Murakami et *al.*, 2013)
- KS (Kavvadias et *al.*, 2005)

Experimentations

- Compactness rate θ = 1 - $|\mathcal{M}_{H'}|$ / $|\mathcal{M}_H|$.
  - The percentage of MTs that can be built from the irredundant MTs, without loss of information.

	$\lvert\mathcal{M}_H\rvert$	$\lvert\mathcal{M}_{H'}\rvert$	θ
Accidents1	1.961	1.866	4.84%
Accidents2	17.486	17.199	1.64%
Connect-Win	4.587.967	4.423.837	3.57%

Accidents and Connect Hypergraphs statistics

	Accidents1	Accidents2	Connect-Win
MMCS(H)	0.301	2.787	**88.491**
KS(H)	8.620	-	-
MMCS(H')	0.001	2.366	84.015
$+$			
GET-ALL-MT($\mathcal{M}_{H'}$)	0.035	0.010	7.291
$=$			
IRRED-ENGINE (H)	**0.036**	**2.376**	91.306

Computation time on Accidents and Connect Hypergraphs (in seconds)

	$\|\mathcal{X}\|$	$\|\xi\|$	$\|Gen\_nodes\|$	$\tau(H)$
$H1$	95	51	65	8
$H2$	99	101	64	9
$H3$	117	20005	68	4
$H4$	60	20	20	20

	$\|\mathcal{M}_H\|$	$\|\mathcal{M}_{H'}\|$	θ
$H1$	832.564.740	358.392	99.95%
$H2$	265.765.380	189.444	99.92%
$H3$	1693	1250	26.17%
$H4$	3^{20}	1	99.99%

Matching and randomly generated hypergraphs characteristics

	$H1$	$H2$	$H3$	$H4$
MMCS(H)	2083.792	766.19	124.692	245.610
KS(H)	-	-	1041.29	**18.550**
MMCS(H')	2.740	0.37	87.54	0.001
+				
GET-ALL-MT($\mathcal{M}_{H'}$)	11.696	1.53	0.031	19.806
=				
IRRED-ENGINE (H)	**14.436**	**1.90**	**87.571**	19.807

Computation time on matching and randomly generated hypergraphs (in seconds)

Outline

Main challenge

- Ilustrate the relevance of our concise representation of hypergraph MTs
- Improve the dependecy inference problem :
  - Computing a compact cover of FDs that faithfully represents the set of all FD holding in a relation r.

Functional dependency

Given a relation schema R, a FD overs R is an expression $X \rightarrow Y$, where $X, Y \subseteq R$. A FD holds in r, denoted $r \models X \rightarrow Y$, where r is a relation over R, if all tuples $u, v \in R$ with $u[X] = v[X]$ satisfy also $u[Y] = v[Y]$.

Related work

- Several approach proposed in the literature classified into two categories : top-down algorithms and bottom-up algorithms.
- Bottom-up algorithms compute a subset of FD, called the cover of the fd and denoted $\textsc{Cover}(\textsc{dep}(r))$.

- Notions of agree sets, maximal sets, disagree sets, necessary sets, etc.
- The minimal cover problem can be solved by computing the hypergraph MTs.
- **Approach** : Computing an irredundant minimal cover$\textsc{Cover}'(D_r)$
  - To illustrate our proposition, we consider DEP-MINER [Lopes et al., 2000]

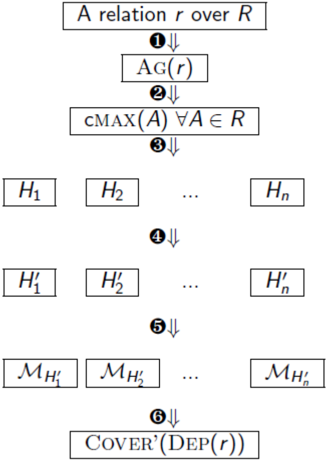


Figure: Methodology description

Study case

	A	B	C	D	E
t_1	1	100	1	2	50
t_2	4	101	1	2	50
t_3	1	102	2	2	70
t_4	1	200	1	2	50
t_5	2	101	3	3	100
t_6	2	200	1	3	70
t_7	1	100	3	2	50

A relation r

$BE \to A$	$A \to D$	$AB \to E$
$BD \to A$	$CE \to D$	$BD \to E$
	$BE \to D$	$AC \to E$
		$CD \to E$

$\textsc{Cover}(D_r)$

Study case

	A	B	C	D	E
t_1	1	100	1	2	50
t_2	4	101	1	2	50
t_3	1	102	2	2	70
t_4	1	200	1	2	50
t_5	2	101	3	3	100
t_6	2	200	1	3	70
t_7	1	100	3	2	50

A relation r

$BE \rightarrow A$	$A \rightarrow D$	$AB \rightarrow E$
$BD \rightarrow A$	$CE \rightarrow D$	$BD \rightarrow E$
	$BE \rightarrow D$	$AC \rightarrow E$
		$CD \rightarrow E$

$\textsc{Cover}(D_r)$

$BD \rightarrow A$
$A \rightarrow D$
$BE \rightarrow D$
$AB \rightarrow E$

$\textsc{Cover'}(D_r)$

Outline

Conclusion

- We propose an approach to improve hypergraph minimal tranversals computation.
  - Concise representation of the set of MTs.
  - Notions of irredundant hypergraph and irredundant MT.
- We introduce a new algorithm, called IRRED-ENGINE which extracts all the MTs from a subset.
- Experiments, carried out on several types of hypergraphs, showed very interesting results evaluated through a compactness measure.
- Application on the dependency inference problem.

Thanks for your attention...