Improved Algorithms for Overlapping and Robust Clustering of Edge-Colored Hypergraphs: An LP-Based Combinatorial Approach

NEURAL INFORMATION PROCESSING SYSTEMS
NeurIPS 2025: Dec 2-7
Paper URL

Changyeol Lee[1*] Yongho Shin[2*] Hyung-Chan An[1]

연세대학교 YONSEI UNIVERSITY
Uniwersytet Wrocławski

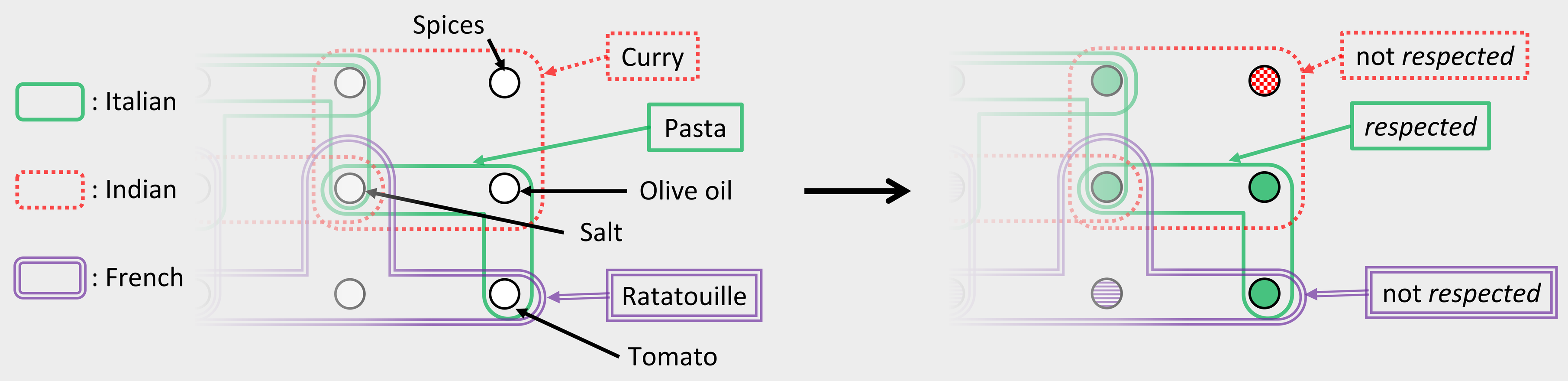
[1] Yonsei University, South Korea
[2] University of Wrocław, Poland
[*] Equal contributions

Local, Global, and Robust ECC

- **Overlapping** and **robust** clustering of edge-colored hypergraphs

Edge-Colored Clustering (ECC) Problems

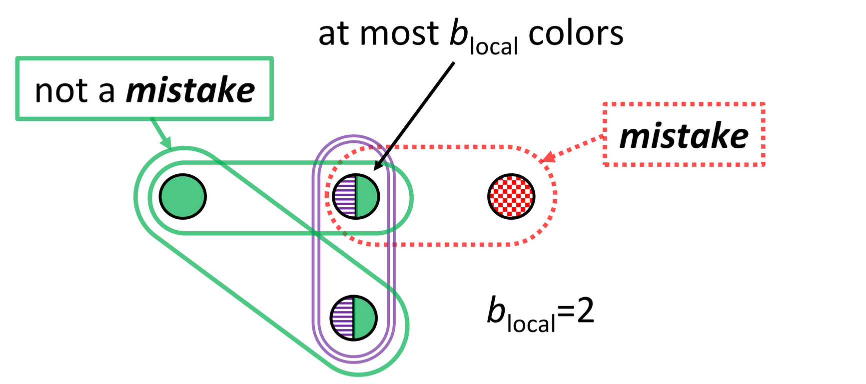
- Data with *categorical* *higher-order* interactions; cluster/color nodes while *respecting* interactions



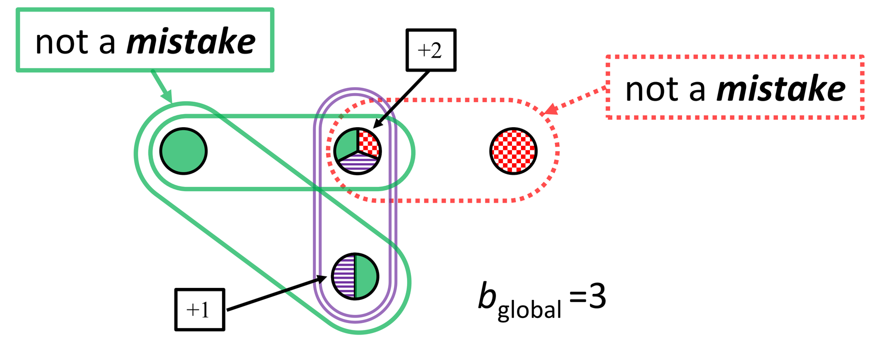
- Amburg et al. (WWW'20; SDM'22), Veldt (ICML'23), Crane et al. (ICML'25)
- **Limitation:** traditional ECC enforces *nonoverlapping* and *exhaustive* clustering (i.e., must assign exactly *one color* to *every node*)

- Three generalizations of traditional ECC (Crane et al., WSDM'24)
  - **Local ECC** and **Global ECC** – overlapping clustering
  - **Robust ECC** – non-exhaustive clustering

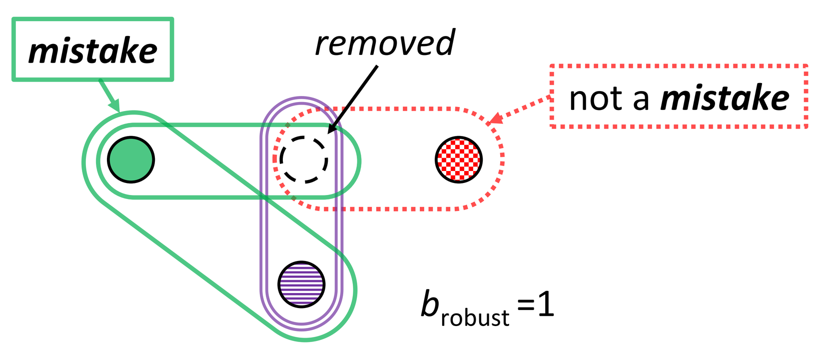
Local ECC
- A *local budget* $b_{\text{local}} \geq 1$
- May assign at most b_{local} colors to each node



Global ECC
- A *global budget* $b_{\text{global}} \geq 0$
- May assign at most b_{global} extra colors throughout the hypergraph



Robust ECC
- A *node-removal budget* $b_{\text{robust}} \geq 0$
- May remove b_{robust} nodes before the color assignment



- ***Mistake***: contains a (non-removed) node whose assigned color(s) $\not\ni$ edge color
- **Goal**: minimize the number of **mistakes**

Our Contributions

- Previous algorithms (Crane et al., 2024)
  - Greedy (combinatorial) algorithms
  - LP-rounding algorithms
- **Proposed algorithms: primal-dual**
  - **LP-based *and* combinatorial *at the same time***

	Crane et al. (2024)		**Proposed**
Local ECC			
	Greedy (combinatorial)	LP-rounding	**Primal-dual (LP-based combinatorial)**
	r-approx.	(b_{local}+1)-approx.	**(b_{local}+1)-approx.**
Global ECC			
	Greedy (combinatorial)	LP-rounding	**Primal-dual (LP-based combinatorial)**
	r-approx.	*bicriteria*	**($2b_{\text{global}}$+2)-approx.**
Robust ECC			
	Greedy (combinatorial)	LP-rounding	**Primal-dual (LP-based combinatorial)**
	r-approx.	*bicriteria*	**($2b_{\text{robust}}$+2)-approx.**

- (b_{local}+1)-approx. is *essentially optimal* for Local ECC (UGC-hard)
  – **Answers an open question posed by Crane et al. (2024)**
- Also works with *non-uniform local budgets* & in the *online* setting

- *bicriteria*: outputs an approx. soln. violating budget constraints by a given factor
- $r := \max |e|$

- Integrality gap lower bounds that (almost) match the approximation ratios
- All proposed algorithms can be analyzed as a bicriteria ($O(1)$,$O(1)$)-approx.
  – **Answers an open question posed by Crane et al. (2024)**

Proposed Algorithm for Local ECC

- **Primal LP**
  - $y_e = 1 \Leftrightarrow e$ is a mistake
  - $x_{v,c} = 1 \Leftrightarrow v$ is colored with c

$$\begin{aligned} \min\ & \textstyle\sum_{e\in E} y_e \\ \text{s.t.}\ & \textstyle\sum_{c\in C} x_{v,c} \leq b_{\text{local}}, && \forall v \in V, \\ & x_{v,c_e} + y_e \geq 1, && \forall e \in E, v \in e, \\ & x_{v,c} \geq 0, && \forall v \in V, c \in C, \\ & y_e \geq 0, && \forall e \in E. \end{aligned}$$

- **Dual LP**
  - $\delta_c(v) := \{e \in \delta(v) : c_e = c\}$

$$\begin{aligned} \max\ & \textstyle\sum_{e\in E, v\in e} \beta_{e,v} - \sum_{v\in V} b_{\text{local}}\alpha_v \\ \text{s.t.}\ & \textstyle\sum_{e\in\delta_c(v)} \beta_{e,v} \leq \alpha_v, && \forall v \in V, c \in C, \\ & \textstyle\sum_{v\in e} \beta_{e,v} \leq 1, && \forall e \in E, \\ & \alpha_v \geq 0, && \forall v \in V, \\ & \beta_{e,v} \geq 0, && \forall e \in E, v \in e. \end{aligned}$$

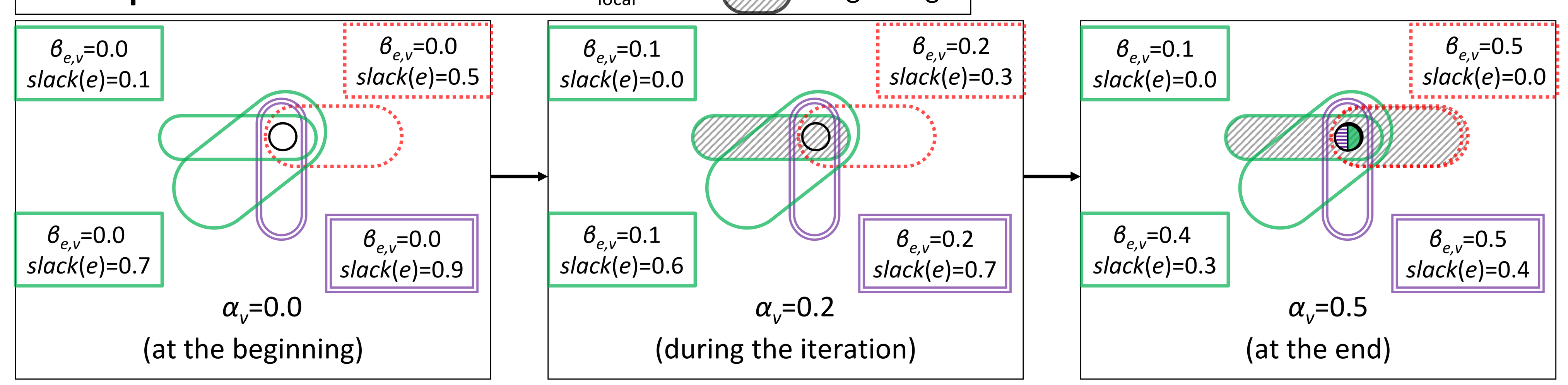
Definition
- *slack*(e) $:= 1 - \sum_{v\in e} \beta_{e,v}$; an edge e is *tight* if *slack*(e)=0
- A color c is *loose for v* if $\exists\, e \ni v$ s.t. c_e=c and *slack*(e)>0

Algorithm Proposed algorithm for LOCAL ECC

Runs in linear time

1: $(\alpha, \beta) \leftarrow (\mathbf{0}, \mathbf{0})$; $L \leftarrow E$
2: **for** $v \in V$ **do**
3: **while** there are more than b_{local} loose colors for v **do**
4: increase α_v and $\sum_{e\in\delta_c(v)\cap L} \beta_{e,v}$ for each loose color c for v at unit rate
5: **if** some edge becomes tight, **then** remove all such edges from L
6: assign loose colors for v to v

A sample run of one iteration b_{local}=2 : tight edge



Analysis #mistakes $\leq \sum_e \sum_{v\in e} \beta_{e,v} \leq (b_{\text{local}} + 1)(\sum_e \sum_{v\in e} \beta_{e,v} - \sum_v b_{\text{local}} \alpha_v) \leq (b_{\text{local}} + 1)\, \mathrm{OPT}_{\text{Primal-LP}}$

- every mistake is tight
- at any moment, $\sum_{v\in e} \beta_{e,v}$ increases by $\geq (b_{\text{local}} + 1)\, \alpha_v$
- weak duality

Computational Evaluation for Local ECC