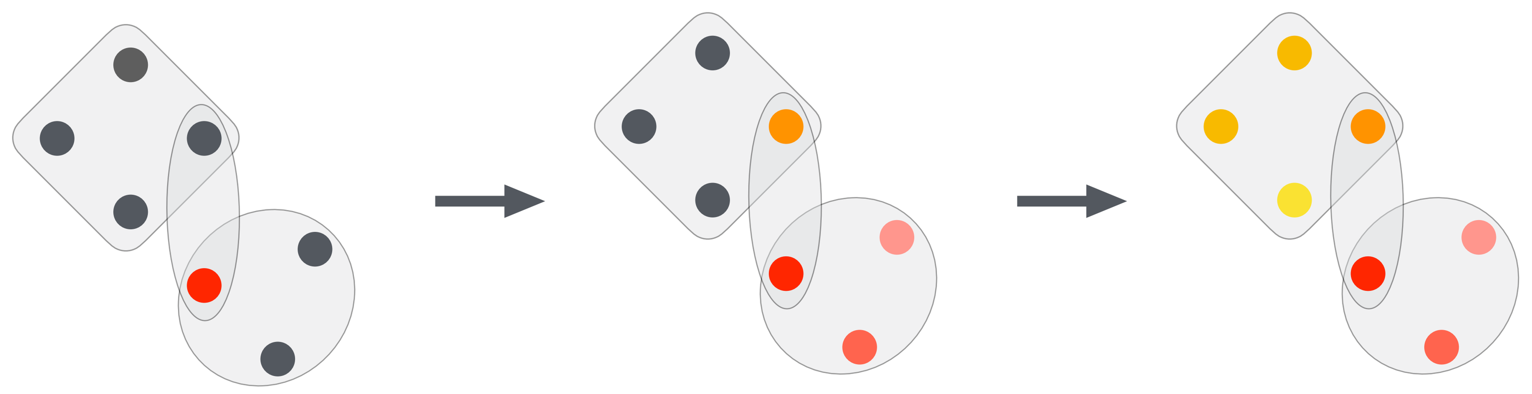
UNIVERSITY OF WATERLOO
PURDUE UNIVERSITY

Local Hyper-Flow Diffusion

Kimon Fountoulakis, Pan Li, Shenghao Yang

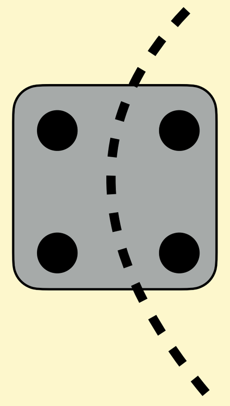
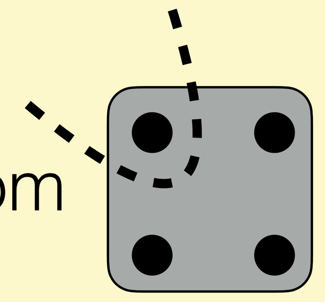


Motivation

- **Diffusion** in a graph is the process of spreading a given initial mass from some seed node(s) to nearby nodes along edges. A diffusion is **local** if it spreads mass to only a small fraction of the entire graph.
- **Hypergraphs** are capable of encoding complex higher-oder relations among entities. Many real-world systems are naturally modelled by hypergraphs.
- Existing local diffusion methods for hypergraphs do not model nontrivial higher-order relations.
- We propose **Hyper-Flow Diffusion (HFD)**, a generic local diffusion framework for hypergraphs. HFD applies to a rich class of higher-order relations which unify and extend many previously studied special cases.

Background: Hyperedge Cut

Observation:
There are distinct ways to cut a 4-node hyperedge.

Question:
How do we treat  differently from  ?

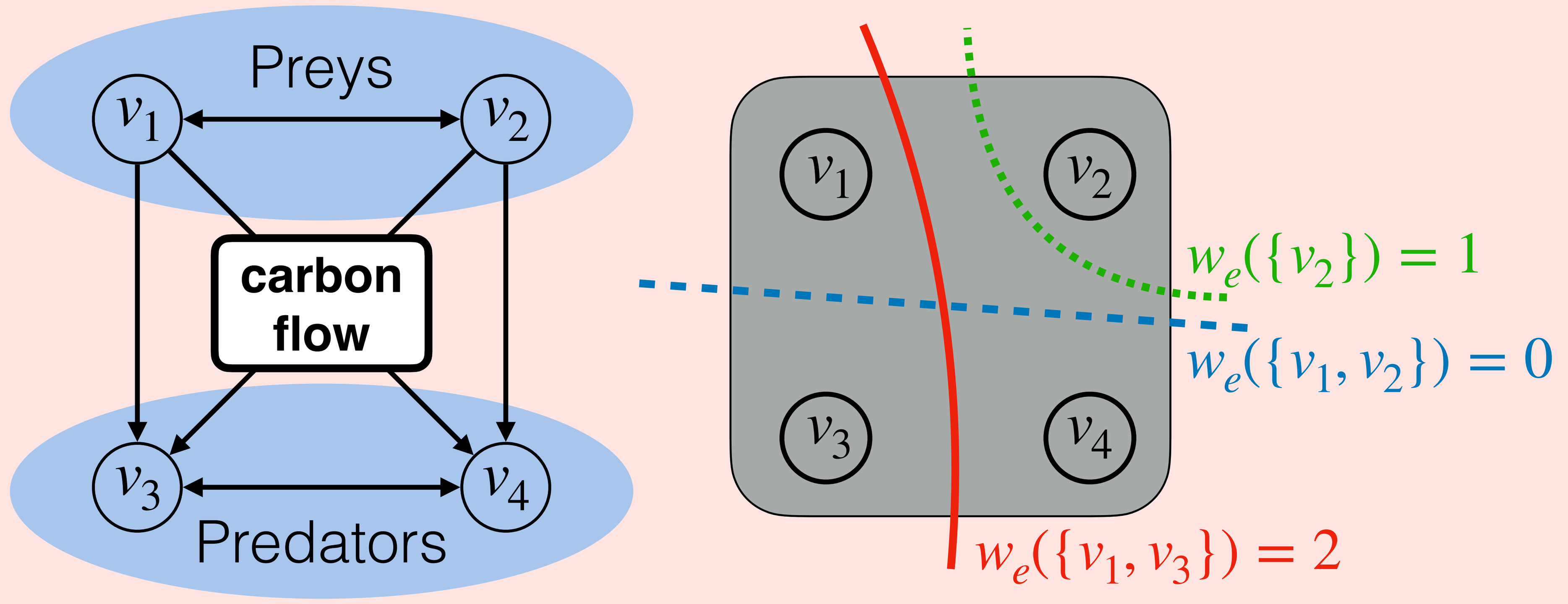
Given a hyperedge $e = \{v_1, v_2, \dots, v_k\}$, we define a **set function** $w_e : 2^e \rightarrow \mathbb{R}_+$, where $w_e(S)$ specifies the cost of splitting e into S and $e \backslash S$.

✓ The **cut-cost** w_e encodes nontrivial higher-order relations among nodes in e.

Hierarchy of Hyperedge Cut

- **Unit (i.e., all-or-nothing):** $w_e(S) = 1$ if $S \cap e \neq \emptyset$, 0 o/w.
- **Cardinality-based:** $w_e(S) = g(\min|S|, |e\backslash S|)$ for a concave function g.
- **Submodular:** w_e is a submodular set function.

Example: Why is submodular cut useful?



Cut-cost $w_e(\{v_1, v_2\}) = w_e(\{v_3, v_4\}) = 0$ encourages separation of predators and preys.

Cut-cost $w_e(\{v_1, v_3\}) = w_e(\{v_2, v_4\}) = 2$ discourages grouping of predators and preys.

Flow & Cut Primal-Dual Form

HFD is formulated as a **primal-dual convex optimization** problem. The primal problem has a physical interpretation as routing flows over hyperedges. The dual problem is

$$\min_{x \geq 0} \frac{1}{2} \boxed{\sum_{e \in E} \boxed{f_e(x)}^2} + \frac{\sigma}{2} x^T \boxed{D} x + (\boxed{d} - \boxed{\Delta})^T x$$

Generalized hypergraph Laplacian operator — *Lovasz extension of w_e* — *Degree matrix/vector* — *Initial mass vector*

Local Hypergraph Clustering

Given a set of seed node(s) S, the goal of local hypergraph clustering is to find a low conductance cluster C around S.

- The **conductance** of a set $A \subseteq V$ is

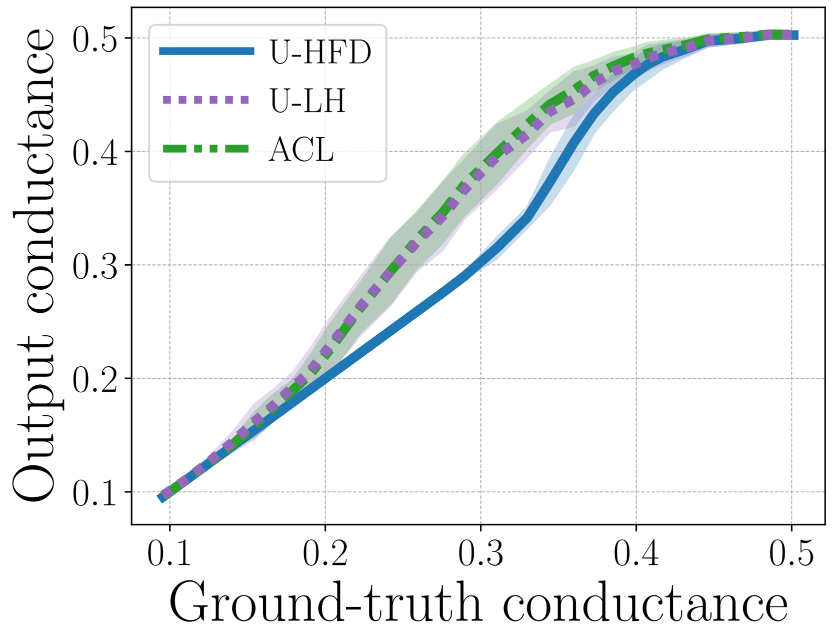
$$\Phi(A) = \sum_{e \in E} w_e(A)/\mathbf{vol}(A), \text{ where } \mathbf{vol}(A) = \sum_{v \in A} d(v).$$

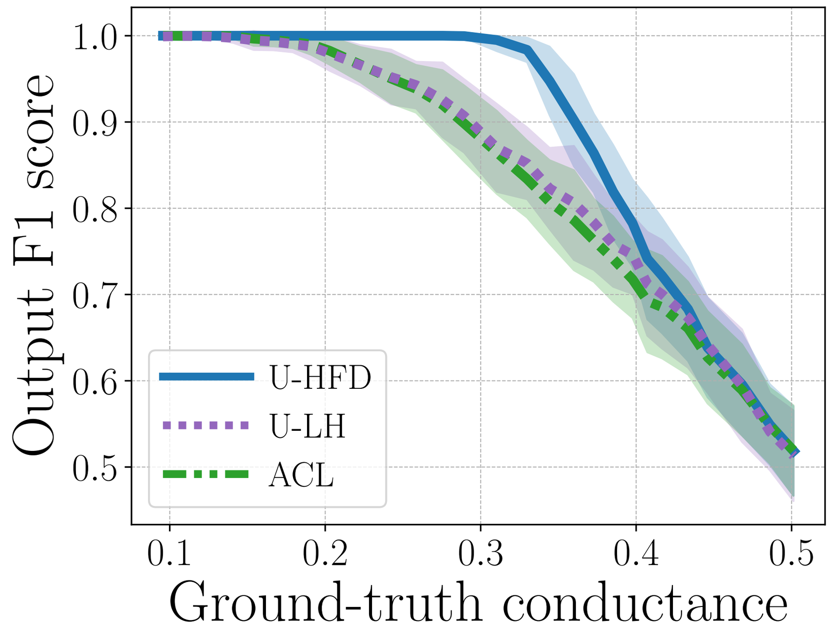
Main result:

If $\mathbf{vol}(S \cap C) \geq \beta \mathbf{vol}(S)$, $\mathbf{vol}(S \cap C) \geq \alpha \mathbf{vol}(C)$, $\alpha, \beta \geq \Omega(1/\log^t \mathbf{vol}(C))$ for some t, and $0 \leq \sigma \leq \tilde{\mathcal{O}}(\beta\Phi(C))$, then applying sweep-cut on the optimal solution x returns a cluster $\tilde{C}$ such that $\Phi(\tilde{C}) \leq \tilde{\mathcal{O}}(\sqrt{\Phi(C)})$.

Empirical Evaluations

- Synthetic experiments using hypergraph SBM:





- Local clustering in a real food network (F1 scores):

Method (cut-cost)	*Producers*	*Low-level consumers*	*High-level consumers*
ACL (unit)	**0.69**	0.44	0.57
LH (unit)	**0.69**	0.45	0.57
HFD (unit)	**0.69**	0.47	0.64
HFD (submodular)	**0.69**	**0.62**	**0.84**