Covering Multiple Submodular Constraints

Special Cases and Applications

Chandra Chekuri
Univ. of Illinois, Urbana-Champaign

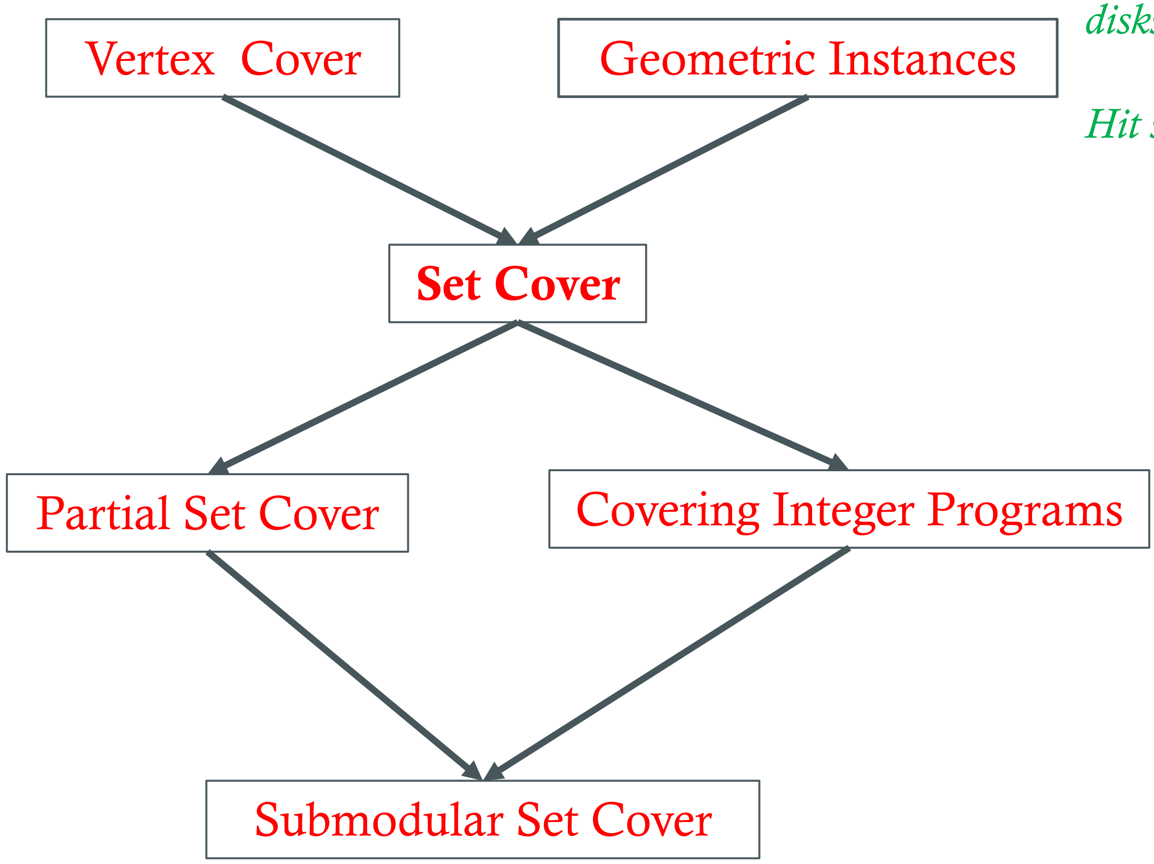
Joint work with
Tanmay Inamdar, Kent Quanrud, Kasturi Varadarajan, Zhao Zhang

CANADAM 2021, Approximation Algorithms Session

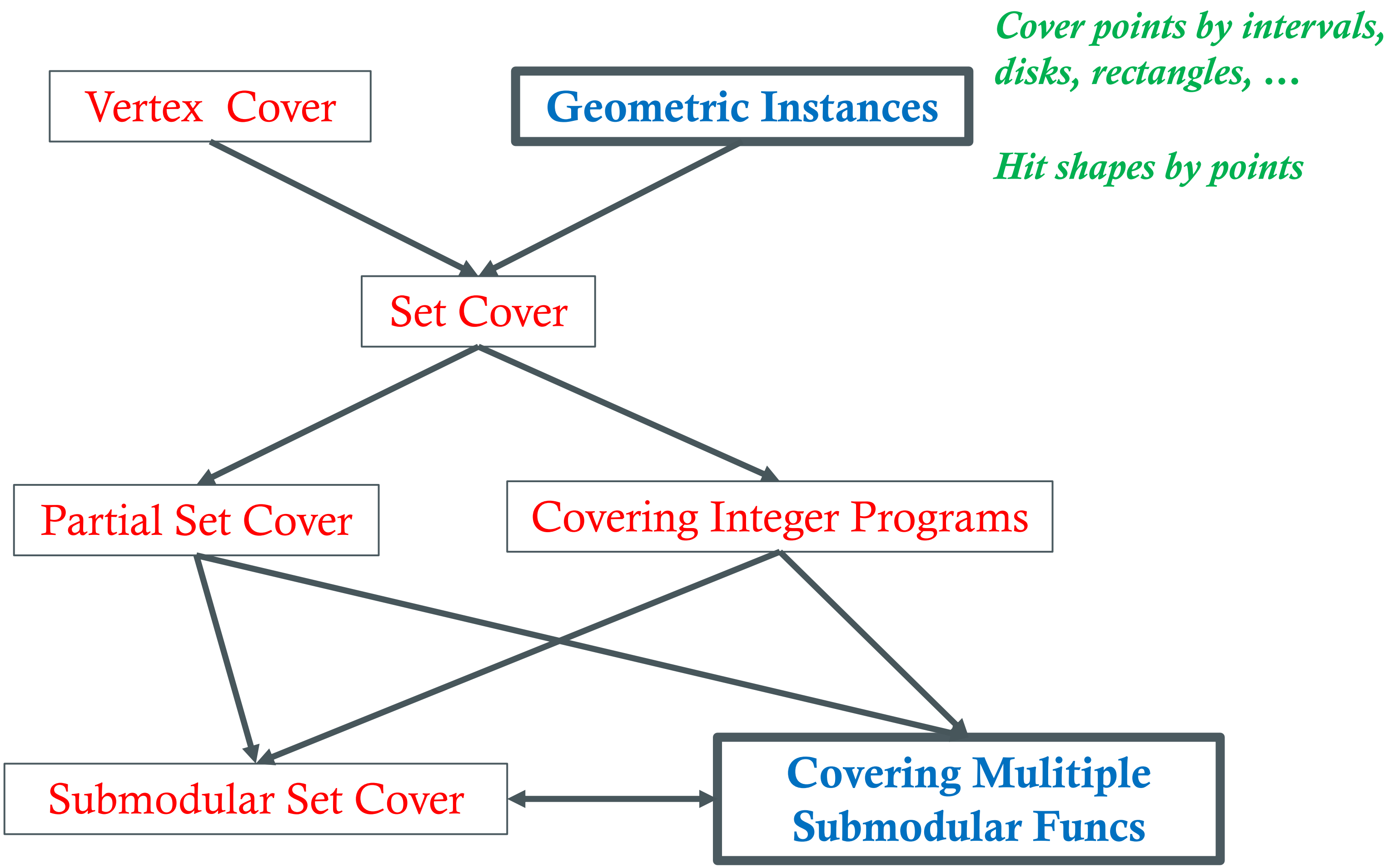
Set Cover and Variants

Cover points by intervals, disks, rectangles, …

Hit shapes by points



New Flavors


Cover points by intervals, disks, rectangles, …
Hit shapes by points
Vertex Cover
Geometric Instances
Set Cover
Partial Set Cover
Covering Integer Programs
Submodular Set Cover
Covering Mulitiple Submodular Funcs

Set Cover

- $U = \{1, 2, \ldots, n\}$
- Sets $S_1, S_2, \ldots, S_m$ each a subset of U
- c_i : non-negative cost of S_i

Goal: Find min-cost sub-collection of $S_1, S_2, \ldots, S_m$ whose union is U

d : max set size, f : max frequency over elements

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_{S_i \ni j} x_i \geq 1 \quad j \in [n]$$

$$x_i \in \{0,1\} \quad i \in [m]$$

Covering Integer Programs

Set Cover

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_{S_i \ni j} x_i \geq 1 \quad j \in [n]$$

$$x_i \in \{0,1\} \quad i \in [m]$$

CIP

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_i A_{i,j}\, x_i \geq b_j \quad j \in [n]$$

$$x_i \in \{0,1\} \quad i \in [m]$$

A, b non-negative

Submodular Set Cover

[Wolsey '82]

- Finite ground set $V = [m]$, c_i non-negative cost of i
- Monotone *submodular* set func: $g: 2^{[m]} \rightarrow R_+$

Goal: $min_{S \subseteq V}\, c(S)$ such that $g(S) \geq b$

Set Cover, CIP, Submod Cover

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_{S_i \ni j} x_i \geq 1 \quad j \in [n]$$

$$x_i \in \{0,1\} \quad i \in [m]$$

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_i A_{i,j}\, x_i \geq b_j \quad j \in [n]$$

$$x_i \in \{0,1\} i \in [m]$$

A, b non-negative

$$\min_S c(S)$$

$$g(S) \geq b$$

Approximation Algorithms

Techniques: *Greedy* and *LP Rounding*

- **Submodular Set Cover:** $(1 + \ln(\max_i g(i)))$
- **Set Cover:** $(1 + \ln d)$ or f (max freq of elements)
- **CIP:** $(\ln d + \ln \ln d + O(1))$ where d is column sparsity. Need Knapsack Cover inequalities

Results are essentially tight in various regimes

Covering Multiple Submodular Functions

[HarPeled-Jones'19] motivated by geom. application

- Finite ground set $V = [m]$, c_i non-negative cost of i
- r monotone *submodular* set funcs: $f_j: 2^{[m]} \rightarrow R_+$

Goal: $min_{S \subseteq V}\, c(S)$ such that $g_j(S) \geq k_j$ for $j = 1\ to\ r$

Set Cover

$$\min \; \sum_{i=1}^{m} c_i x_i$$

$$\sum_{S_i \ni j} x_i \geq 1 \quad j \in [n]$$

$$x_i \in \{0,1\} \; i \in [m]$$

CIP

$$\min \; \sum_{i=1}^{m} c_i x_i$$

$$\sum_i A_{i,j} \, x_i \geq b_j \quad j \in [n]$$

$$x_i \in \{0,1\} i \in [m]$$

A, b non-negative

Submod Cover

$$\min_S c(S)$$

$$g(S) \geq b$$

Multi Submod Cover

$$\min_S c(S)$$

$$g_j(S) \geq k_j \quad j \in [r]$$

Covering Multiple Submodular Functions

Goal: $\min c(S)$ such that $g_j(S) \geq k_j$ for $j = 1\ to\ r$

Can reduce to standard submodular cover problem

$$g(S) = \sum_j \min\left(g_j(S), k_j\right)$$

Greedy gives $O\left(\ln r + \ln\left(\sum_j k_j\right)\right)$ approx.

Covering Multiple Submodular Functions

Goal: $\min c(S)$ such that $g_j(S) \geq k_j$ for $j = 1\ to\ r$

Can reduce to standard submodular cover problem

$$g(S) = \sum_j \min\left(g_j(S), k_j\right)$$

Greedy gives $O(\ln r + \ln\left(\sum_j k_j\right))$ approx.

Main question: Can we avoid $\ln\left(\sum_j k_j\right)$ term?

Why? Applications including CIP

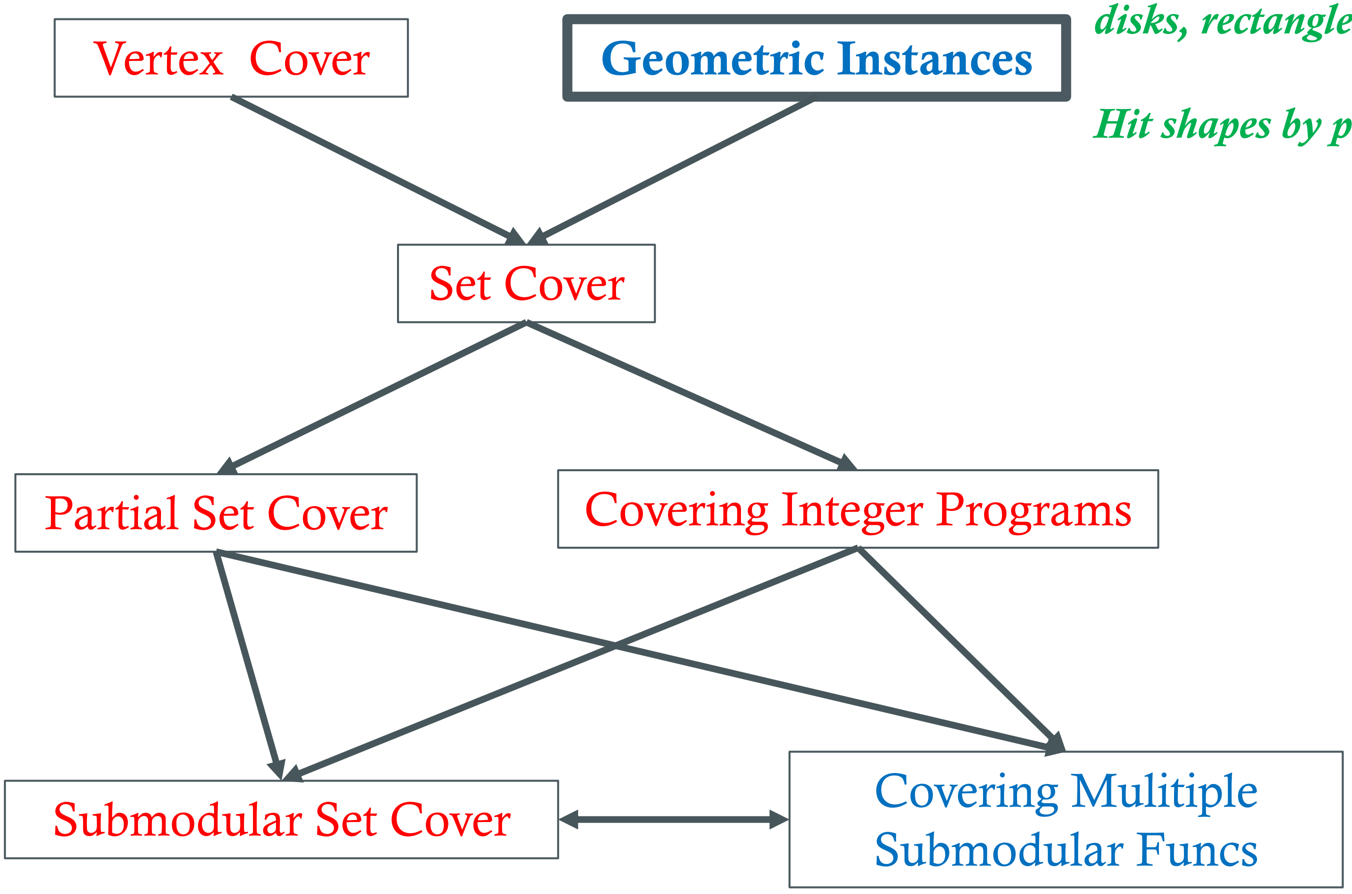
CIP and Submodular Cover

- Reduction to Submodular Cover/Greedy only yields $\Omega(m)$ approximation in worst case [Dobson'80, Wolsey'82]
- Can obtain $O(\log d)$ using LP and KC inequalities [Kolliopoulos-Young'01, Chen-Harris-Srinivasan'16, C-Quanrud'18]

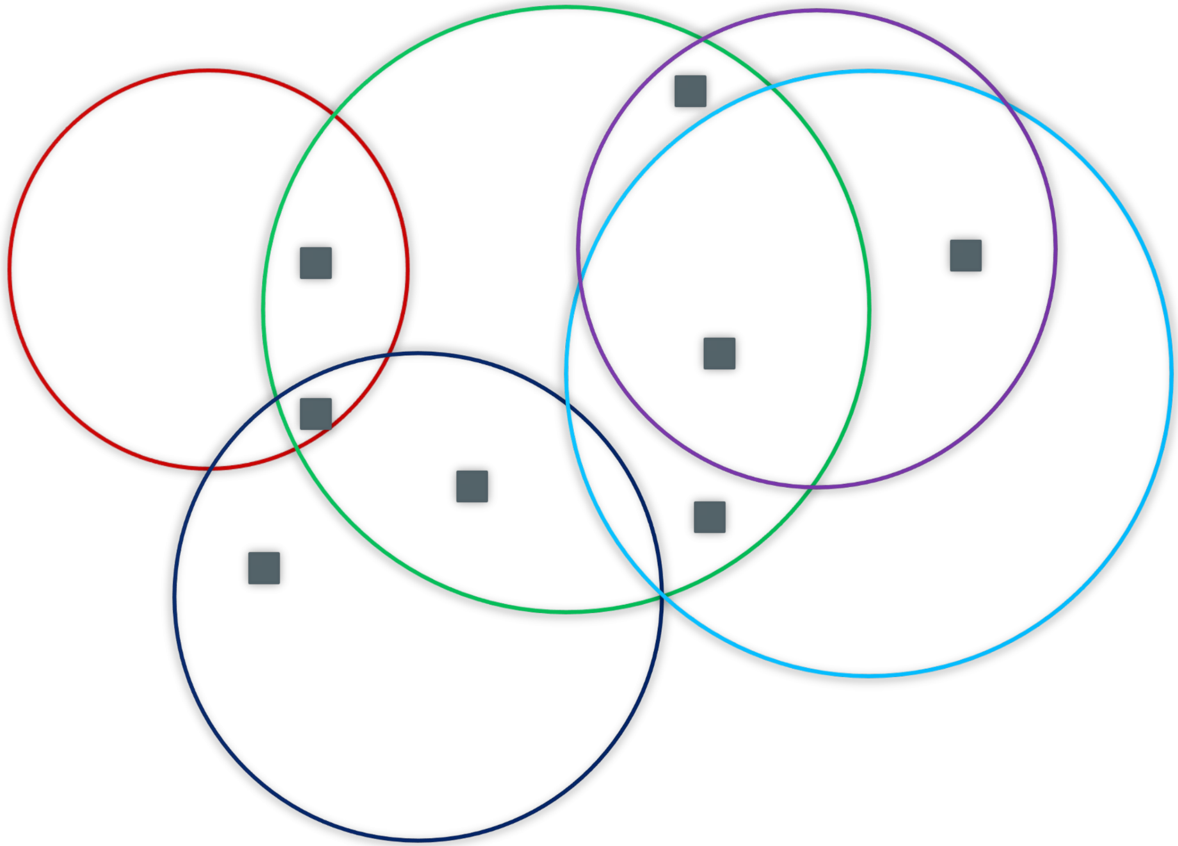
New Player

Cover points by intervals, disks, rectangles, …

Hit shapes by points

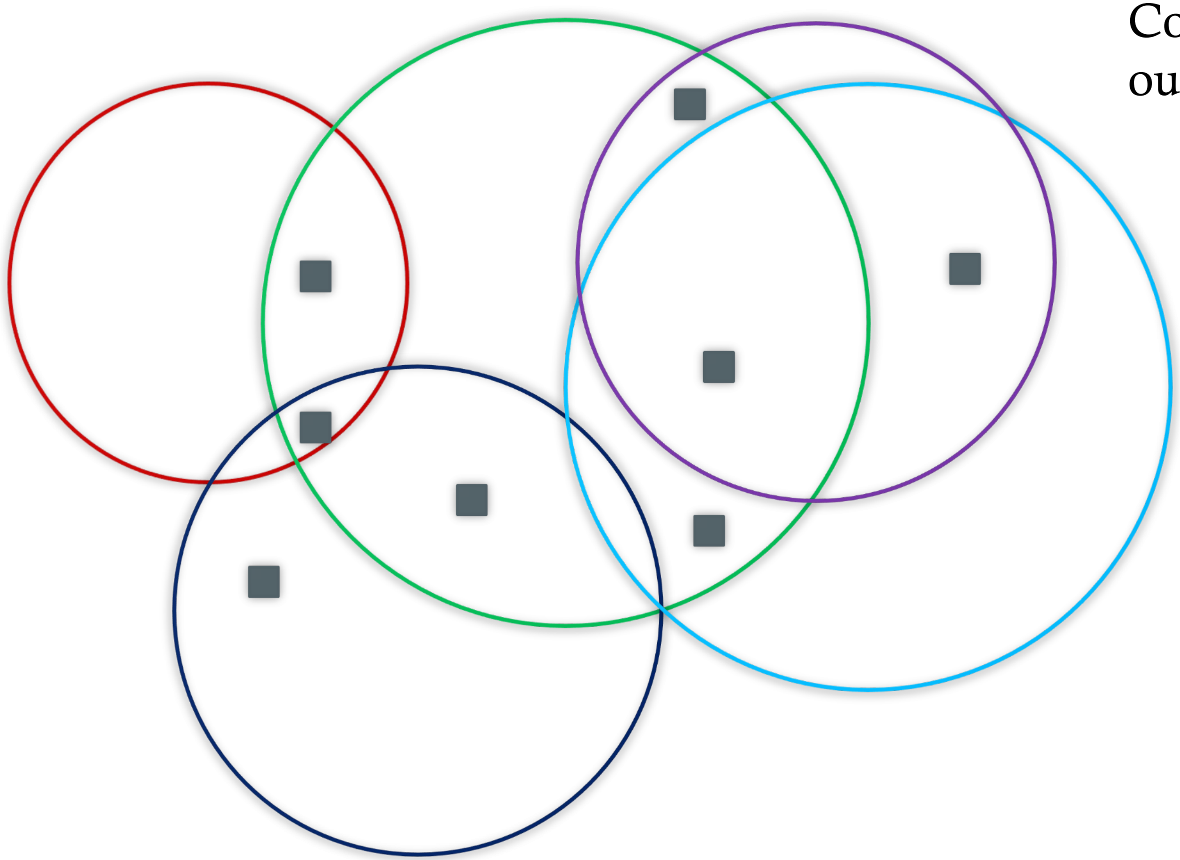


Cover points by disks



O(1) approx via **LP relaxation**: *union complexity bounds and quasi-uniform sampling* [Varadarajan'09][Chan et al'12]

Partial Cover points by disks

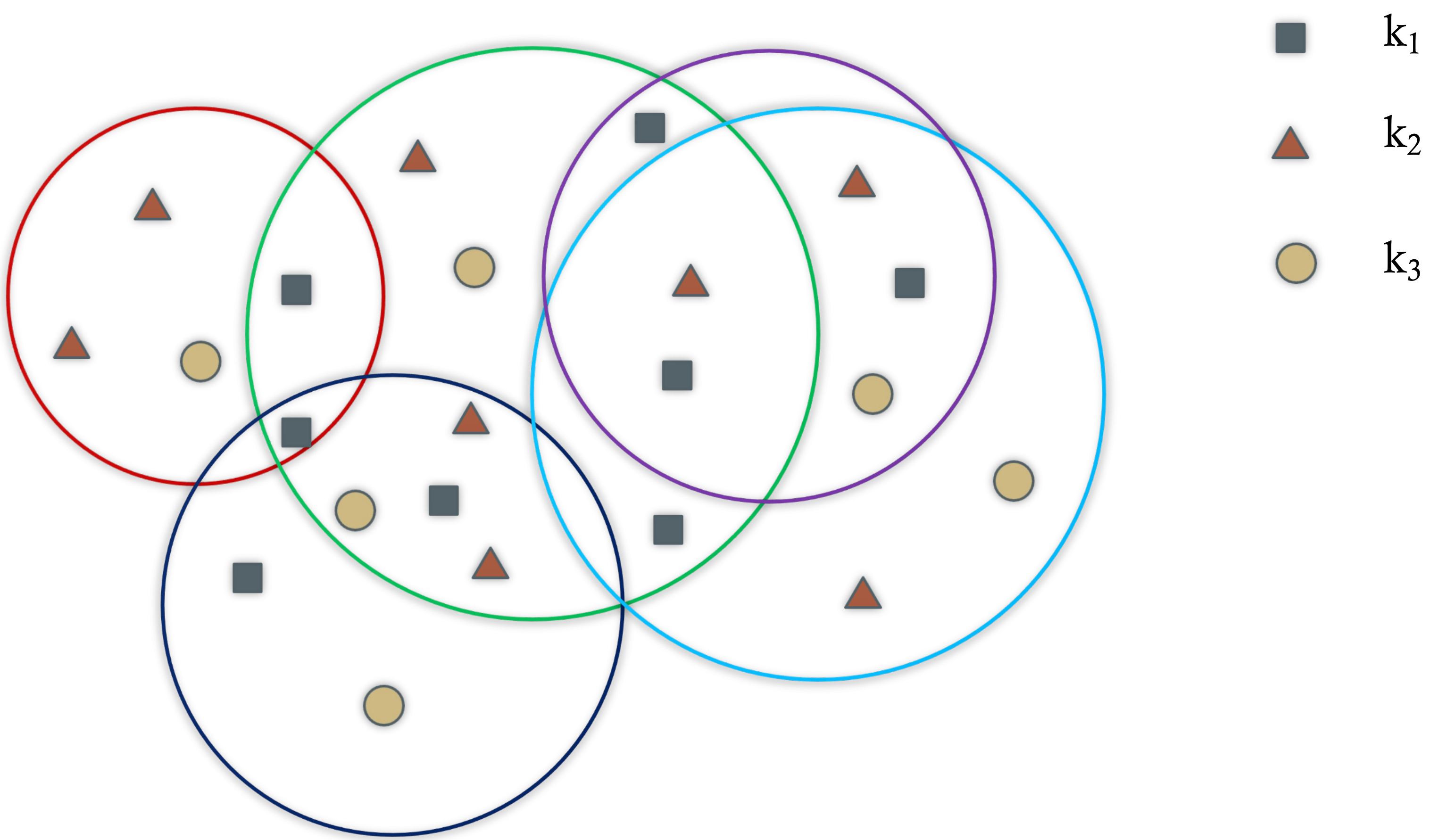


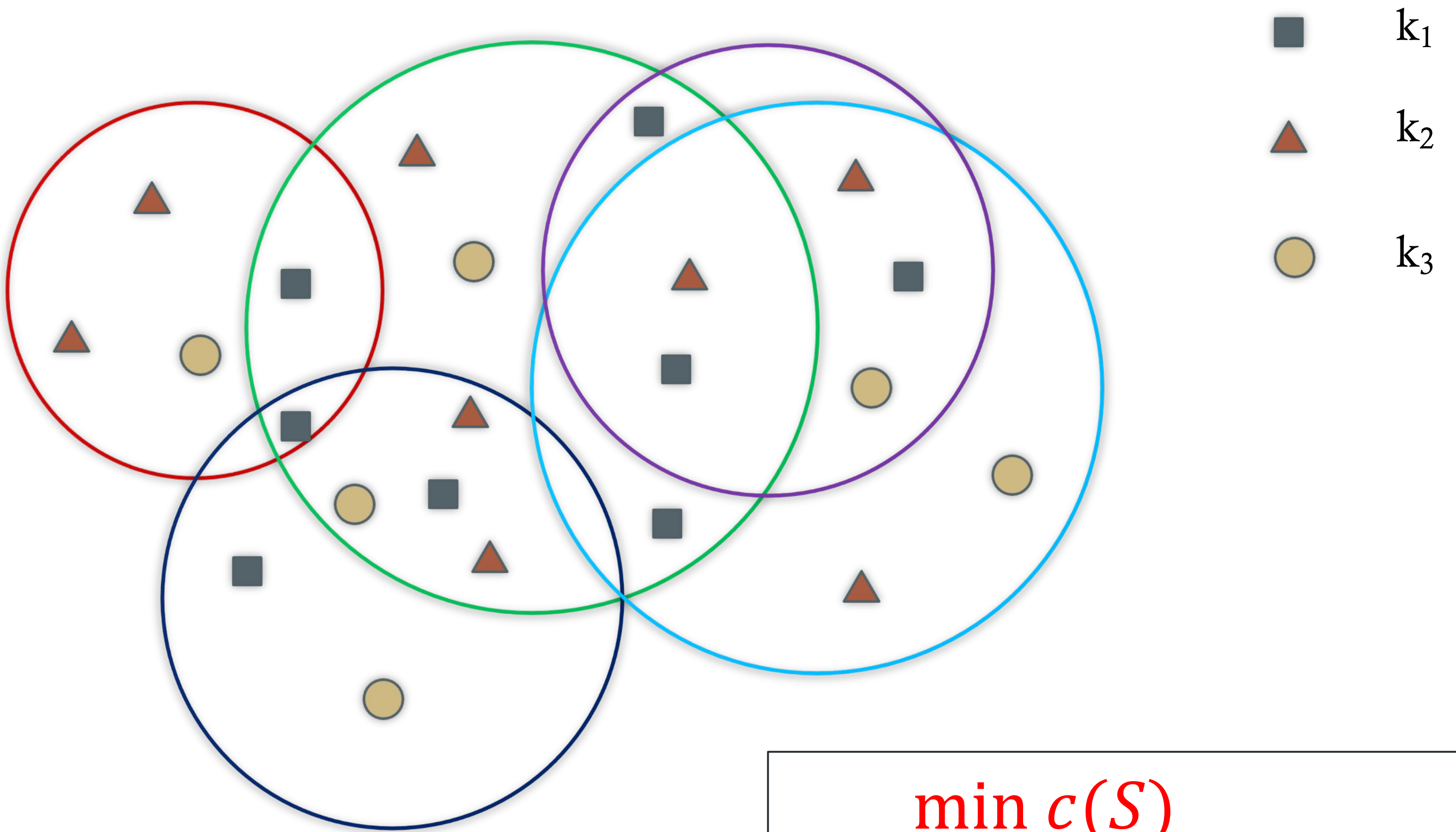
Cover k points out of n

[Inamdar-Varadarajan'18]

O(1) approx via **LP relaxation.** Reduce to covering all points

Partial Set Cover with Multiple Types





$$\min_{S} c(S)$$

$$g_j(S) \geq k_j \quad j \in [r]$$

$g_j(S)$: # of points covered by disks in S

Result 1

For Geometric Multiple Partial Cover

β approx. for Geometric Set Cover via natural LP implies $O(\beta + \ln r)$ approx.

Hence $\beta = O(1)$ implies $O(\ln r)$ approximation

Subsumes known result for CIPs as special case

Result 2

For Covering Multiple Submodular Covering Constraints

Bicriteria approximation

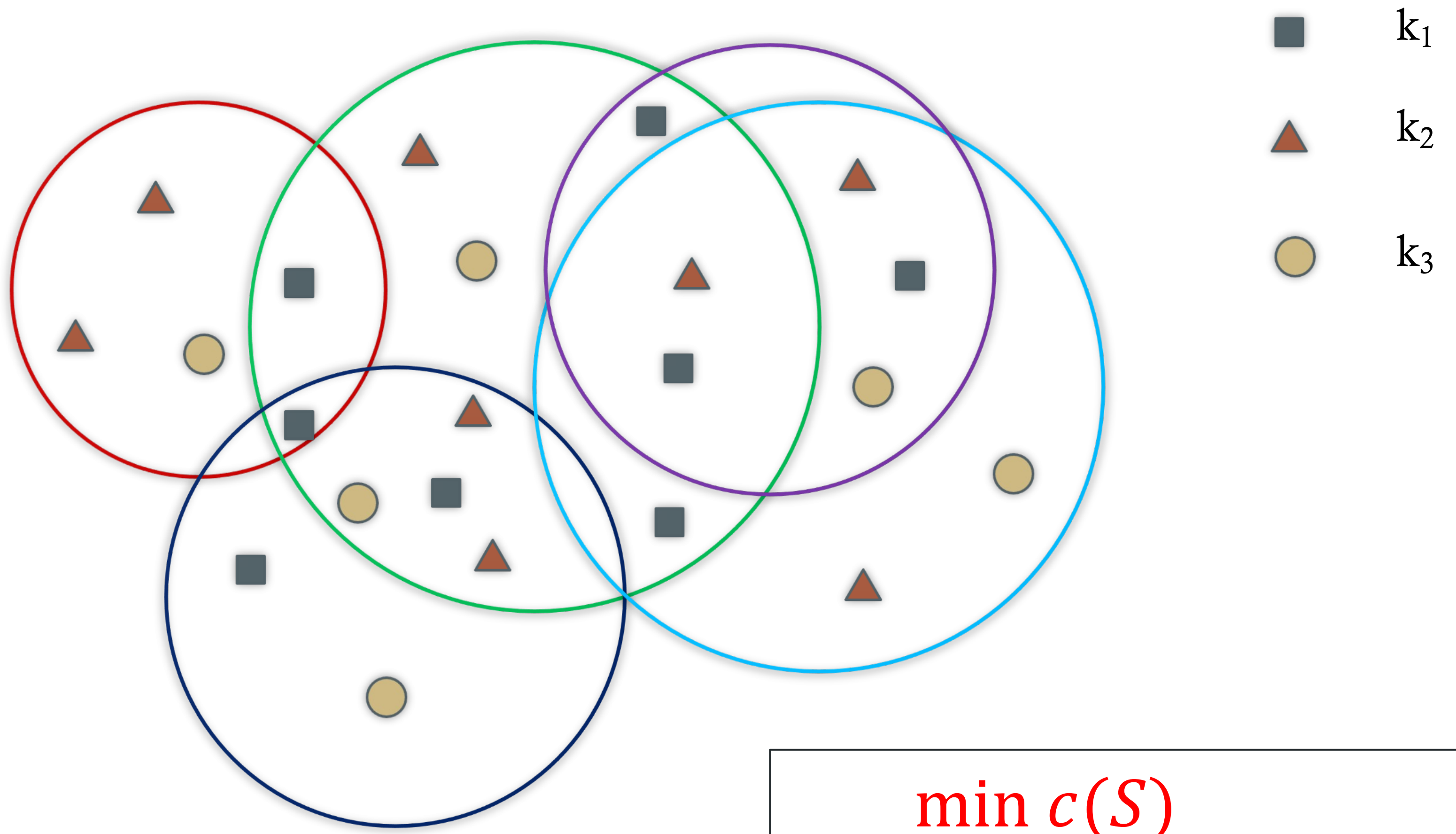
- Cost of solution S is $O(\frac{1}{\epsilon}\ln r)\ OPT$
- Each constraint satisfied approximately
  - $f_j(S) \geq \left(1 - \frac{1}{e} - \epsilon\right) k_j$ for each j
- Approx ratio depends only on r. Bicriteria sufficient for several applications including motivating one from [HarPeled-Jones'19]

Other results

- Improve and *simplify* [Inamdar-Varadarajan'18] reduction for Partial Set Cover: β approx. for Set Cover via natural LP implies $\frac{e}{e-1}(\beta + 1)$ approx. for Partial Cover
- Related results for clustering problems with outliers
  - $O(\ln r)$ approx for facility location with r client types and outliers
  - $O(\ln r)$ approx for minimizing sum of radii with r client types and outliers
  - Not black box reduction but very similar ideas

Ideas from

- [Inamdar-Varadarajan'18] A generic reduction via LP. β approx. for Set Cover via *natural LP* implies $2(\beta+1)$ approx. for Partial Cover version
- Multiple Vertex Cover constraints [Bera-Gupta-Kumar-Roy'14]
- Round and Fix algorithm for CIP [C-Quanrud'18]
- Continuous extension and concentration inequalities for submod functions [CCPV'07, Vondrak'10]



$$\min_S c(S)$$
$$g_j(S) \geq k_j \quad j \in [r]$$

$g_j(S)$: # of points covered by disks in S

LP for Set Cover

x_i: fraction of set S_i taken in solution

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_{S_i \ni j} x_i \geq 1 \quad j \in [n]$$

$$x_i \geq 0 \quad i \in [m]$$

Will assume β approx. for covering points by disks

LP for Partial Set Cover

x_i: fraction of set S_i taken in solution
z_j: fraction of elemt j covered

$$\min \sum_{i=1}^{m} c_i x_i$$

$$z_j = \min\{1, \sum_{S_i \ni j} x_i\} \quad j \in [n]$$

$$\sum_j z_j \geq k$$

$$x_i \geq 0 \quad i \in [m]$$

LP for Partial Set Cover

x_i: fraction of set S_i taken in solution
z_j: fraction of elemt j covered

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_{j \in S_i} x_i \geq z_j \quad j \in [n]$$

$$\sum_j z_j \geq k$$

$$z_j \in [0,1] \quad j \in [n]$$

$$x_i \geq 0 \quad i \in [m]$$

Rounding for Partial Set Cover

- $H = \{ j : z_j \geq \frac{1}{2} \}$ highly covered elements, L rest
- Setting $x'_i = 2\, x_i$ for each set S_i gives feasible fractional soln to cover all of H. Use β approx.
- Residual covering requirement is $k' = k - |H|$. Use simple Greedy algorithm on L until k' covered

Variant/simplification of [Inamdar-Varadarajan'18]

r Partial Covering Constraints

x_i: fraction of set S_i taken in solution
z_j: fraction of elemt j covered

$$\min \sum_{i=1}^{m} c_i x_i$$

$$\sum_{j \in S_i} x_i \geq z_j \quad j \in [n]$$

$$\sum_{j \in C_t} z_j \geq k_t \qquad t = 1\ to\ r$$

$$z_j \in [0,1] \quad j \in [n]$$

$$x_i \geq 0 \quad i \in [m]$$

Rounding

- $H = \{ j : z_j \geq \frac{1}{2} \}$ highly covered elements, L rest
- Setting $x_i' = 2\, x_i$ for each set S_i gives feasible fractional soln to cover all of H. Use β approx.
- Residual covering requirement for t'th constraint is $k_t' = k_t - |H|$.
- How to simultaneously satisfy r residual constraints?

Rounding

- $H = \{ j : z_j \geq \frac{1}{2} \}$ highly covered elements, L rest
- Setting $x_i' = 2\, x_i$ for each set S_i gives feasible fractional soln to cover all of H. Use β approx.
- Residual covering requirement for t'th constraint is $k_t' = k_t - |H|$.
- **Randomly round:** pick each S_i independently with probability $\Theta(\ln r)\ \ x_i$

Randomized Rounding Analysis

- Residual covering requirement for t'th constraint is $k'_t = k_t - |H|$.
- **Randomly round:** pick each S_i independently with probability $\Theta(\ln r)\ x_i$

Question: will constraints be satisfied?

Intuition: from standard Set Cover but constraints in terms of z_j variables while rounding wrt to x_i variables

Difficulty: Each x_i can influence many z_js

Overcoming difficulty

- Need *Knapsack Cover* inequalities (used already by [Bera et al] for Multiple Partial Vertex Cover)
- Cannot separate KC inequalities in the submodular setting. Use round and cut framework as in [Bera et al]

Analysis

With KC inequality

Can use concentration of submodular functions under independent rounding [Vondrak'10]

Proof via connection between two continuous extensions of submodular functions (concave closure and multilinear extension)

Rounding again

- $H = \{ j : z_j \geq \frac{1}{2} \}$ highly covered elements, L rest
- Cover H using β approx.
- Residual requirement for t'th const is $k_t - |H|$.
- **Randomly round:** pick each S_i independently with probability $\Theta(\ln r)\ \ x_i$
- With KC ineq: constraint covered with prob $\left(1 - \frac{1}{r}\right)$
- **Fix** unsatisfied constraints separately with Greedy alg.

Conclusions

- Covering Multiple Submodular Constraints is a *useful* model to keep in mind
- Several non-obvious applications/connections
- Simple in retrospect but nice interplay of techniques
  - Round and fix framework
  - KC inequalities
  - Concentration inequalities

Thank You!