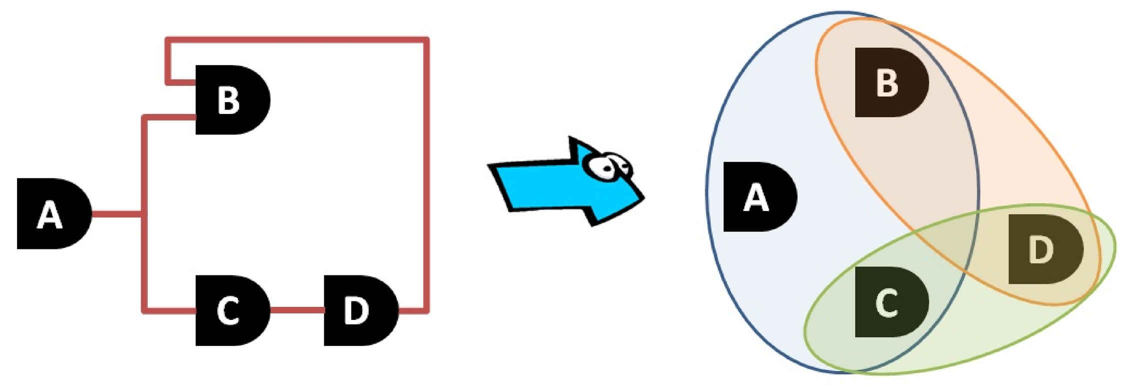
Consistency of Spectral Algorithms for Hypergraphs under Planted Partition Model

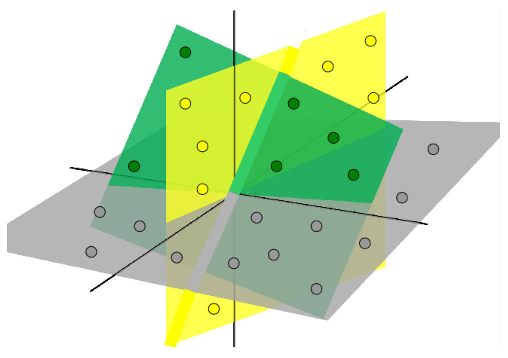
Debarghya Ghoshdastidar (Ph.D. - 4^{th} year)
Advisor: Prof. Ambedkar Dukkipati
Department of Computer Science & Automation, Indian Institute of Science

Hypergraphs and Clustering

Circuit partitioning for VLSI design:



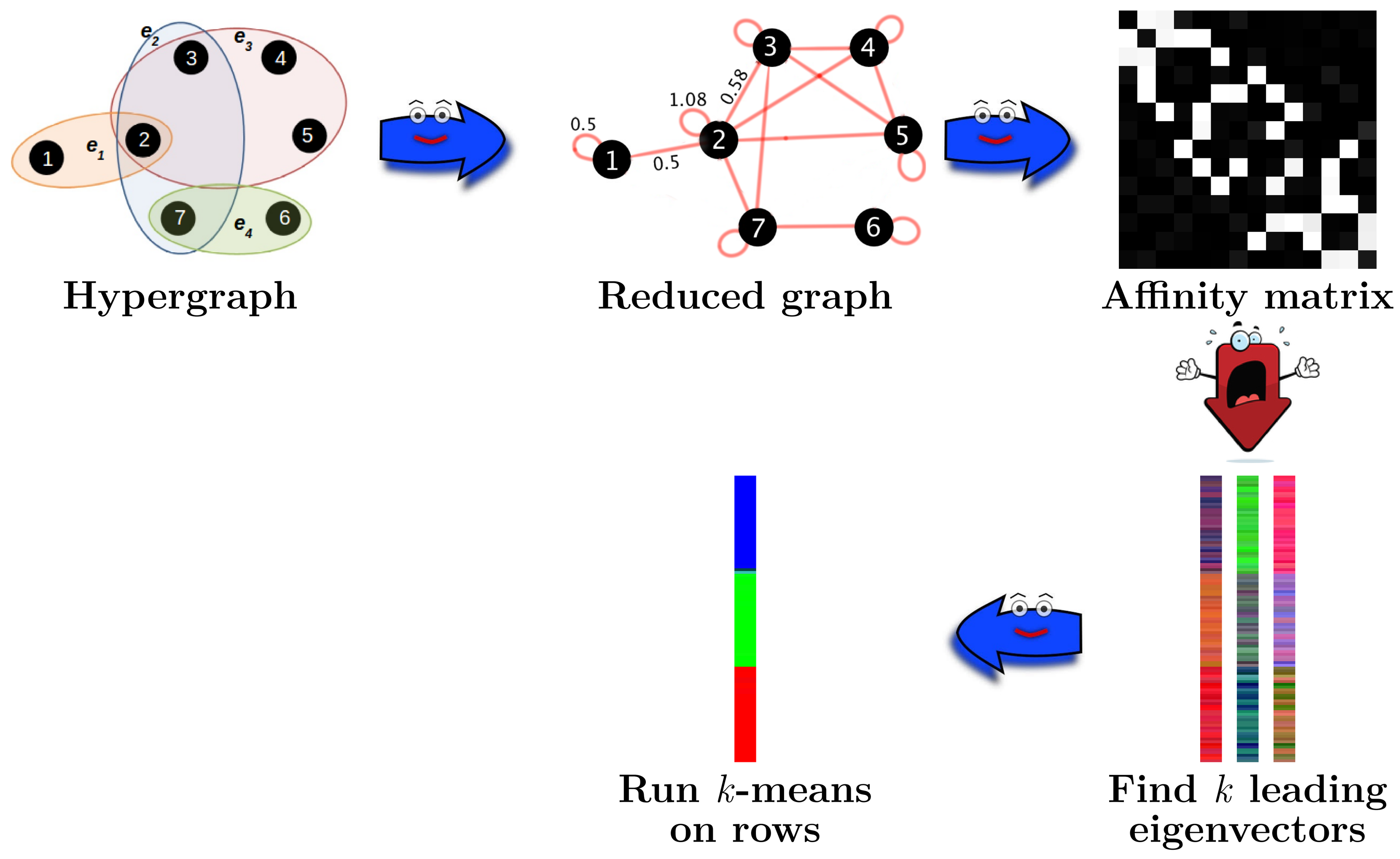
Subspace clustering and applications:



Group coplanar points

Motion segmentation

Spectral Hypergraph Partitioning



Consistency of Hypergraph Partitioning

Planted partition model (special case)

- m-uniform hypergraph on n nodes
- k unknown classes of equal size, $k = O\left(\frac{n^{1/4}}{\log n}\right)$
- Edge prob. within class $= p$ (unknown)
- Edge prob. across classes $= q < p$ (unknown)

Graphs ($m = 2$) – Spectral clustering

- Extensively studied [Rohe et al '11; Lei & Rinaldo'15]
- %error $\to 0$ as $n \to \infty$ (weak consistency)

No. of misclustered nodes $= O\left(\frac{n^{1/2}}{(\log n)^2}\right)$

Algorithms studied in our works

HOSVD [Govindu '05]

- Based on tensor decomposition
- Applicable only for uniform hypergraphs

No. of misclustered vertices $= O\left(\frac{n^{(4-m)/2}}{(\log n)^{2m-1}}\right)$

NH-Cut [Bolla '93; Zhou et al '07]

- Minimizes hypergraph cut
- Can tackle non-uniform hypergraphs

No. of misclustered vertices $= O\left(\frac{n^{(3-m)/2}}{(\log n)^{2m-3}}\right)$

TTM / TTM-ext [proposed]

- Maximizes hypergraph associativity
- Partitions uniform / non-uniform hypergraphs
- Unifies several higher order learning methods

No. of misclustered vertices $= O\left(\frac{n^{(3-m)/2}}{(\log n)^{2m-3}}\right)$

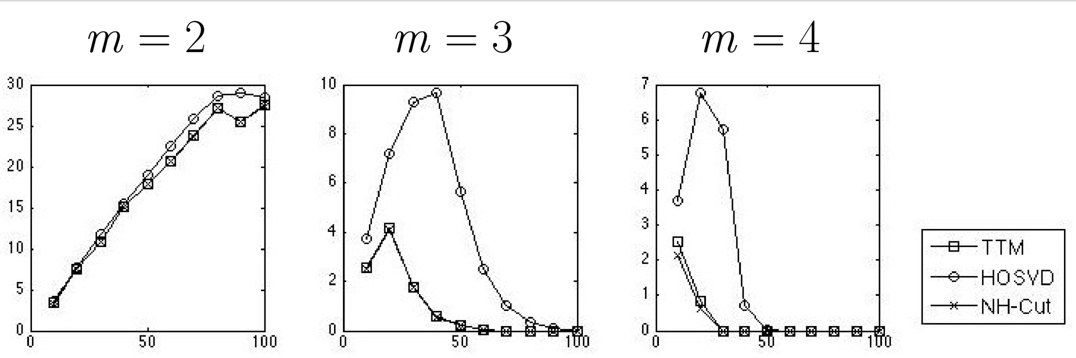
Tetris [proposed]

- Computationally efficient variant of TTM
- Applicable only for weighted hypergraphs
- Partitions using only few sampled edges

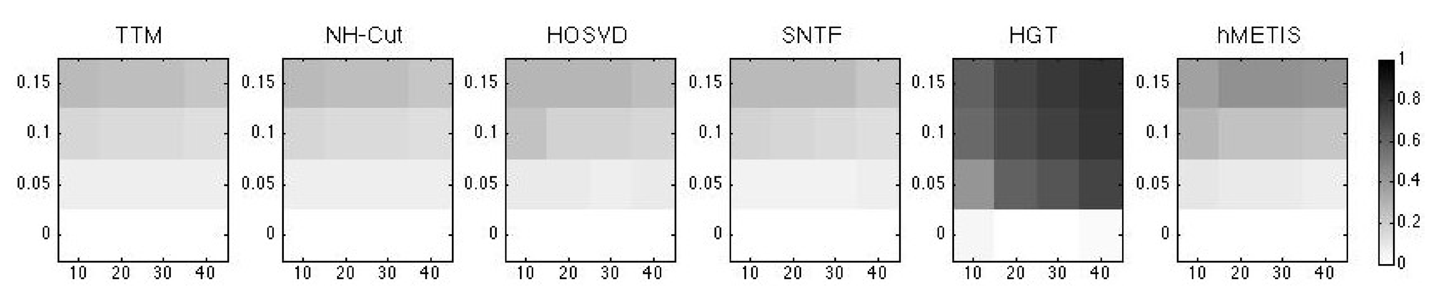
Consistency for edge sampling ratio $= \Omega\left(\frac{n^{(1.5-m)/2}}{(\log n)^{2m-3}}\right)$

- Retrieves graph results for $m = 2$; but for $m > 2$, error $\to 0$ as $n \to \infty$ (strong consistency)

Empirical Studies



Variation of error for HOSVD, TTM and NH-Cut with increase in n under planted partition model



Fractional error incurred by hypergraph partitioning algorithms in clustering noisy points from three intersecting lines when the cluster size n/k (horizontal), and the noise level (vertical) are varied

Algorithm	Mean eror (%)	Time (sec)
k-means	19.58	0.03
k-flats	13.19	0.38
SSC	1.53	0.80
LRR	2.13	0.94
SSC-OMP	16.93	0.72
TSC	18.44	0.19
NSN+Spec	3.62	0.08
SCC	2.53	0.45
SGC	3.50	0.54
Tetris	1.31	0.50

Performance on 2 motion videos in Hopkins dataset

References

- M. Bolla. *Discrete Mathematics* 117(3). 1993.
- V. M. Govindu. In *CVPR*. 2005.
- D. Zhou, J. Huang & B. Schölkopf. In *NIPS*. 2007.
- K. Rohe, S. Chatterjee & B. Yu. *Annals of Statistics* 39(4). 2011.
- D. Ghoshdastidar and A. Dukkipati. In *NIPS*. 2014.
- J. Lei & A. Rinaldo. *Annals of Statistics* 43(1). 2015.
- D. Ghoshdastidar and A. Dukkipati. In *ICML*. 2015.
- D. Ghoshdastidar and A. Dukkipati. *Annals of Statistics* (in press). arXiv: 1505.01582.
- D. Ghoshdastidar and A. Dukkipati. *Manuscript submitted*. arXiv: 1602.06516.