Hypergraph Communities

Dimitrios Vogiatzis

Dec 2020

Table of contents

Definitions

Hypergraph

$G = (V, \mathbf{E})$, where V is the set of vertices and $\mathbf{E} = \bigcup_i V_i$, and $V_i \subset V$

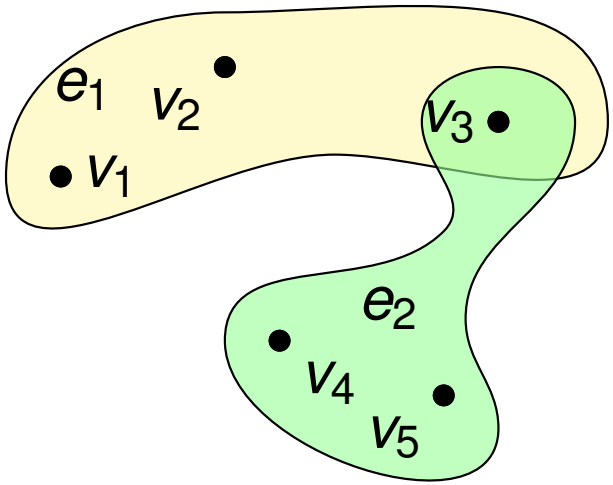


Figure: Uniform Hypergraph: equal number of vertices per edge

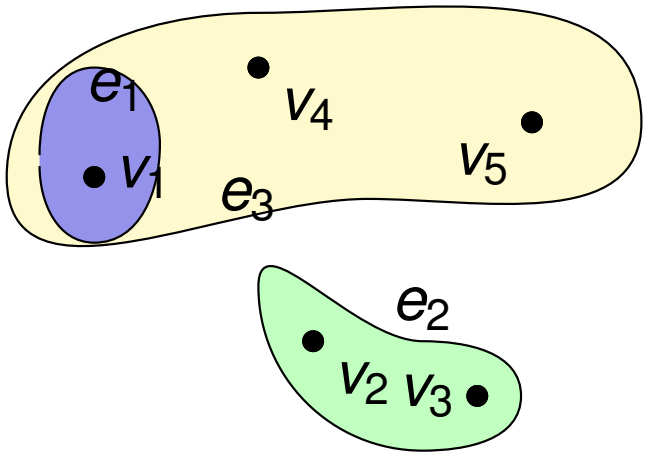


Figure: Non Uniform Hypergraph: different number of vertices per edge

Motivation or Applications I

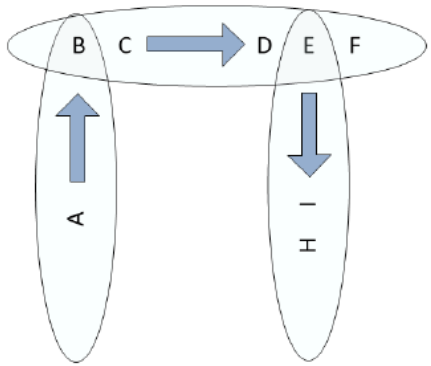


Figure: Chemical Reaction Network

- Social Tagging: e.g, user, seeing a movie, and taggs
- Chemical Reaction networks

Reducing a hypergraph to a simpler graph

- Reduction to a simple graph
- Reduction to a bi-partite graph

Reduction of hypergraph to bipartie

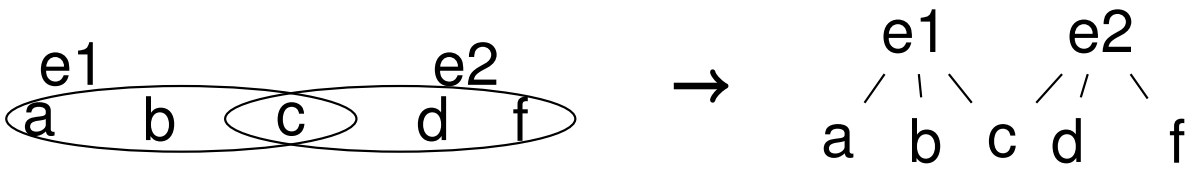
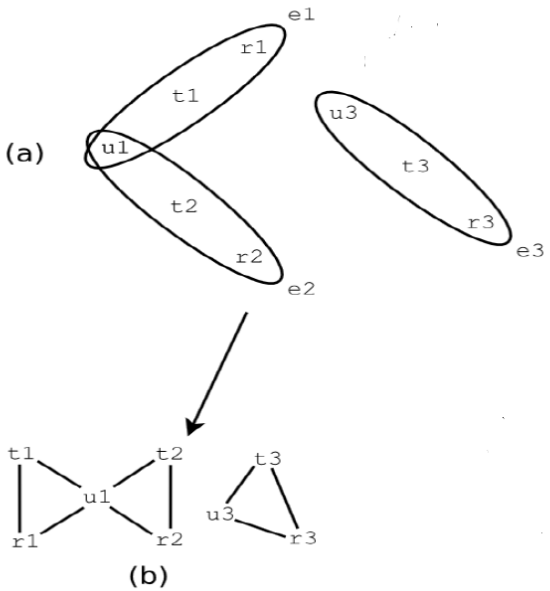


Figure: Hypergraph and its Star expansion

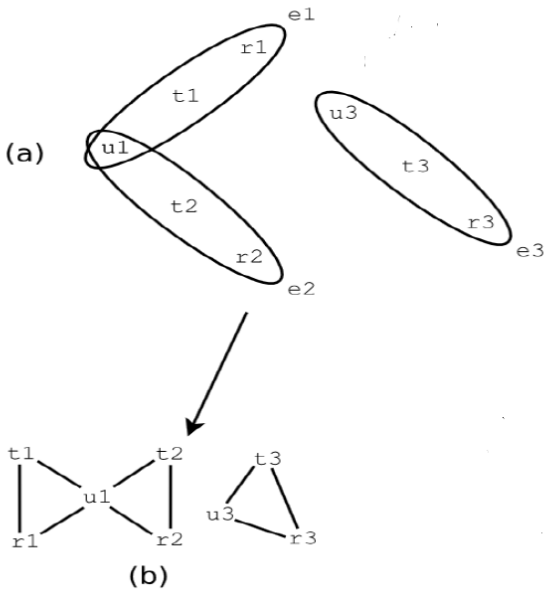
- $E^* = \{(u, v) : u \in V, v \in E\}$
- $w(e*) = w(e)/\delta_e$

Reduction of hypergraph to a simple graph: cliques



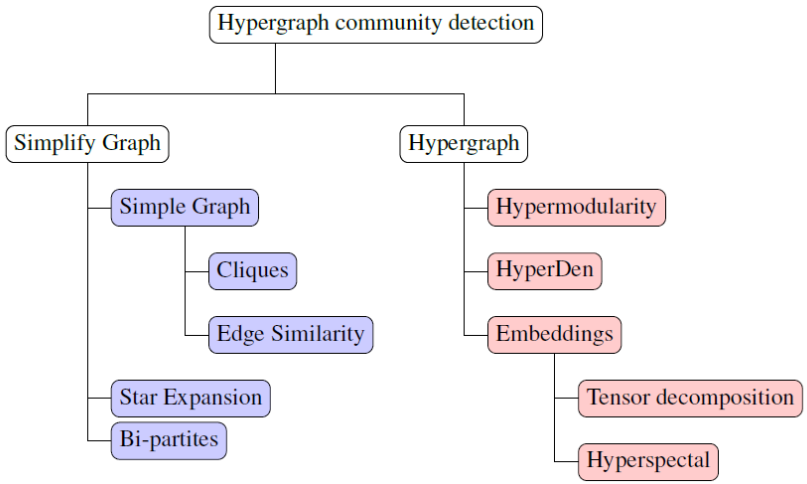
- each hyperedge can be represented as a clique
- works also for hyperedges of variable size (i.e. non-uniform)

Reduction of hypergraph to a simple graph: cliques



- each hyperedge can be represented as a clique
- works also for hyperedges of variable size (i.e. non-uniform)
- but a complete weighted hypergraph will have $\binom{n}{k}$ weights
- the resulting graph will have only $\binom{n}{2}$ weights
- Thus a weighted graph cannot accurately represent the original graph

Clustering Methods



- Hyper-Modularity Maximization
- Hyper-Spectral clustering
- Hyper-Density

Hyper-spectral clustering

- in spectral clustering, we define a graph Laplacian
- in simple graphs (synnetric normalized): $L = I - D^{-1/2} A D^{-1/2}$
- in hyper graphs: $L_{HYP} = I - D_v^{-1/2} H W D_e H^T D_v^{-1/2}$

Tensor Factorization

- A hypergraph represented as a Tensor T. For simplicity, assume a 3-order tensor $\mathbf{T} \in \mathbb{R}^{I \times P \times Q}$
- The PARAFAC tensor decomposition is,

$$\mathbf{T} = [[A, B, C]] + E = \sum_{f=1}^{F} \mathbf{a}_f \circ \mathbf{b}_f \circ \mathbf{c}_f + E$$

$A \in R,^{I \times J}$ $B \in R,^{P \times J}$ $C \in R,^{Q \times J}$: factor matrices, F: #latent factors interpreted as communities, $\circ$: outer product of vectors $\mathbf{a}_f \in R^I$, $\mathbf{b}_f \in R^P$, $\mathbf{c}_f \in R^Q$, E is the error

- Is tensor decomposition same as hyperspectral clustering?

Hyper-Density

An extension of DBScan for hyper-graphs

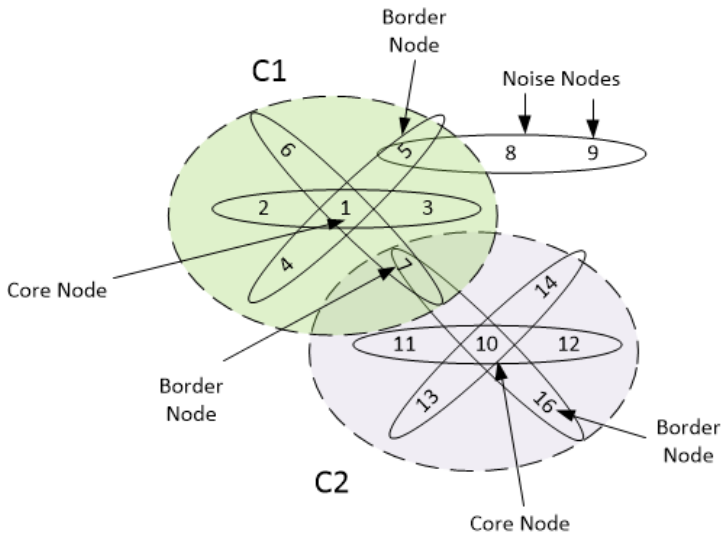


Figure: HyperDen algorithm: Community C1 build with $\eta = 6$, comprising vertices 1,2,3,4,5,6 and 7. vertices 2,3,4,5,6 and 7 are directly density reachable from vertex 1. Vertices 8 and 9 are noise vertices. The degree of vertex 1 is 6; vertices 2,3,4,6,7,8 and 9 have degree 2; and vertex 5 has degree 4. Likewise, community C2 is formed around vertex 10. vertex 7 belongs to both communities

Experiments: Cliques, Hyper-Spectral, Hyper-Density

Comparing three clustering algorithms

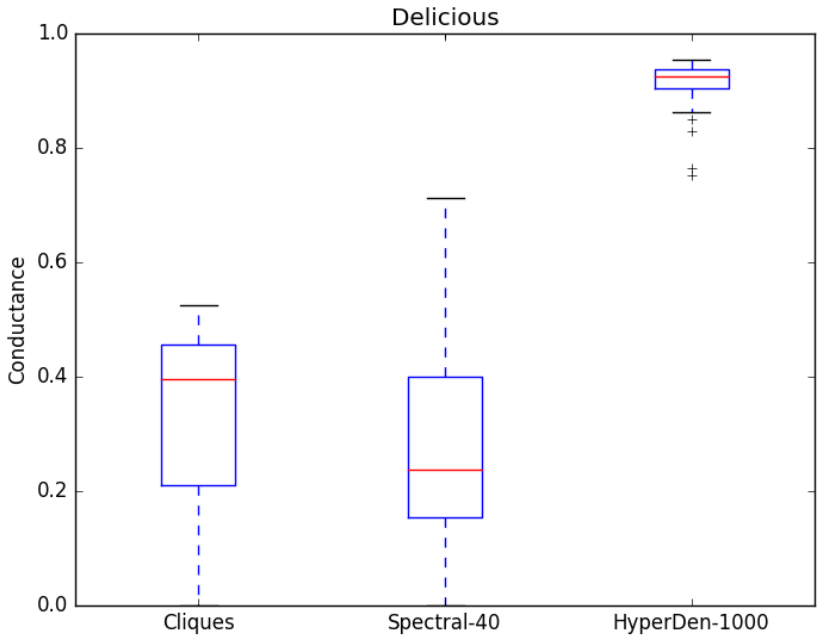


Figure: Conductance measures the connectivity of a community to other communities. Lower values are better

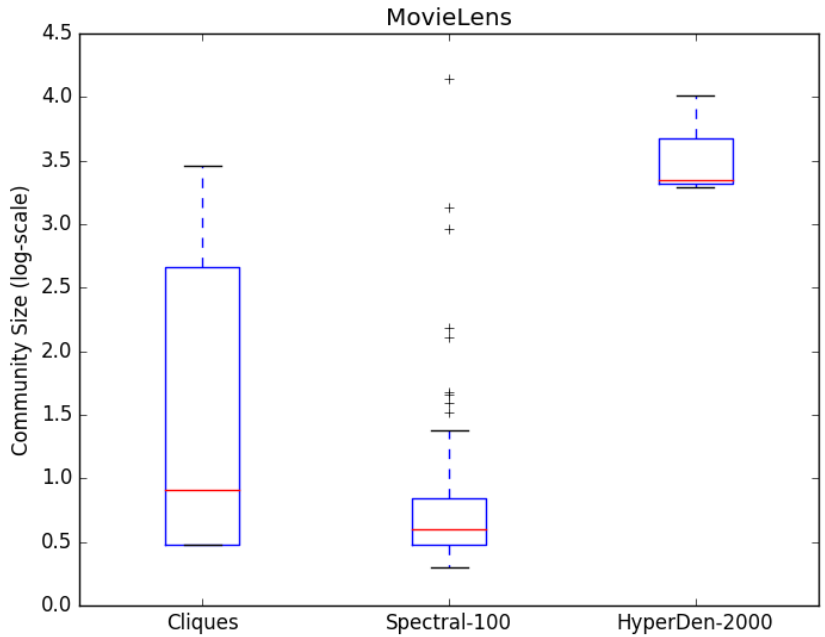


Figure: Size of communities in log-scale