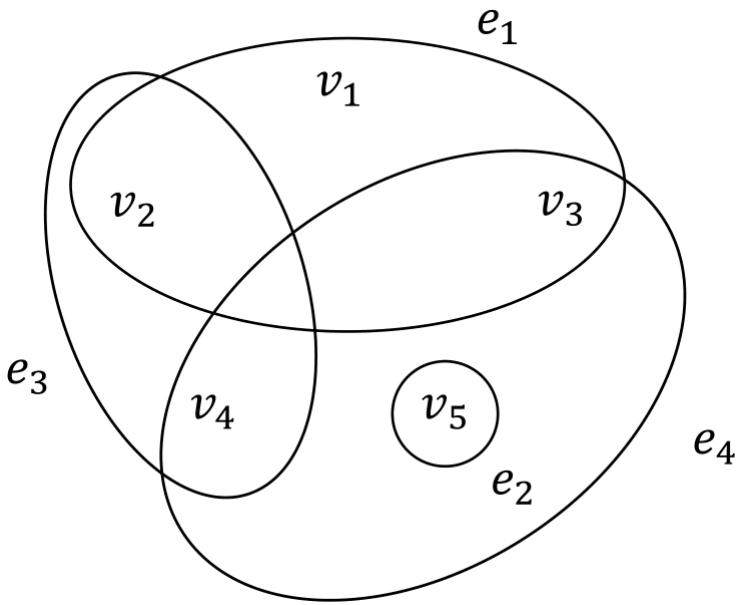
The research is financed by NAWA - The Polish National Agency for Academic Exchange

Modelling hypergraphs in Julia with the Hypergraphs.jl package

Przemysław Szufel & Bogumił Kamiński

https://github.com/pszufe/SimpleHypergraphs.jl

Hypergraph



Hypergraph = a generalization of a graph where a **single edge** can connect **more than two** vertices

- Applications
  - social data analysis
  - sending a single email to several recipients
  - a customer giving reviews to several restaurants
  - security vulnerabilities in information networks

- Benefits - better capture and analyze dependencies within the network than a standard graph

LightGraphs.jl library - base for the analysis

- Pure Julia
- Extensible
- Full array of standard methods for graph analysis
- Built for performance

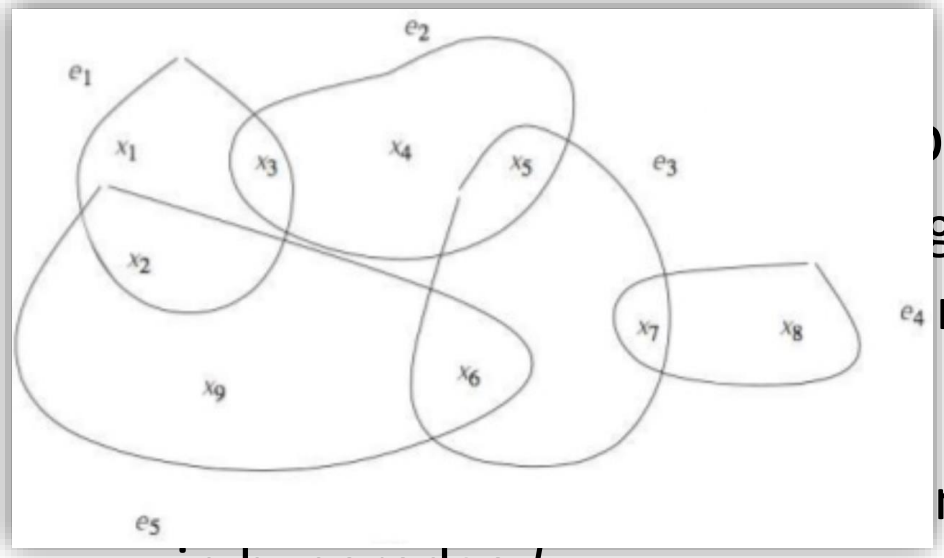
Test	LightGraphs	NetworkX		graph-tool
G1 = Erdos-Renyi (10k, 0.1) (s)	7.13		2.65	19.3
G2 = Barabassi-Albert (10k, 400) (s)			3.6	10.1
Betweenness (G2[1:3000]) (s)		DNF	6.77	3.34
Closeness (G2, s)	35.79	DNF	82	44.2
PageR	28.20	5 130	75.8	30.2
lustering Coefficient (G2, ms)	255.53	37 400	167	270

Source: James Fairbanks, Seth Bromberger (2017), Light graphs: Our Network, Our Story

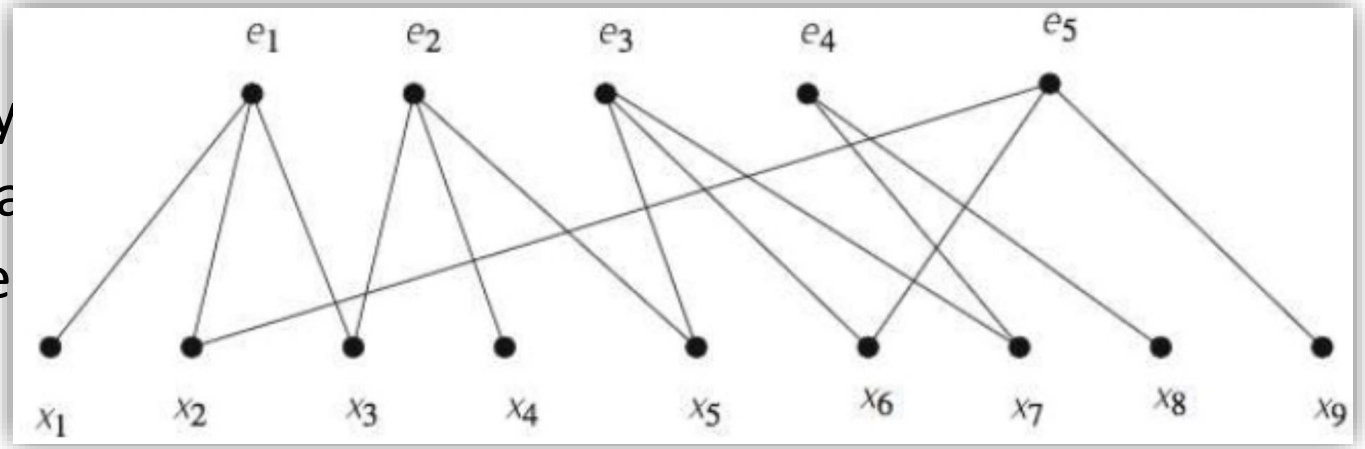
SimpleHypergraphs.jl

- Data representation of a hypergraph (redundancy)
  - A collection of vertices belonging to hyperedges
  - A collection of hyperedges containing vertices
- API
  - hypergraph represented as a matrix where $A[v, h]$ indicates weight of vertex v in hyperedge h
- Algorithms
  - Modularity and community detection
  - Bipartite view of a hypergraph (zero overhead) → LightGraphs
  - Two-section view of a hypergraph (zero overhead) → LightGraphs

SimpleHypergraphs.jl



- bergraph (redundancy)
  - ging to hyperedges
  - ntaining vertices
  - matrix where $A[v, h]$ indicates weight of vertex v in hyperedge h
- Algorithm
  - Modula y and community
  - Bipartite iew of a hyp rgra
  - Two-section view of a hype



Yelp dataset use case

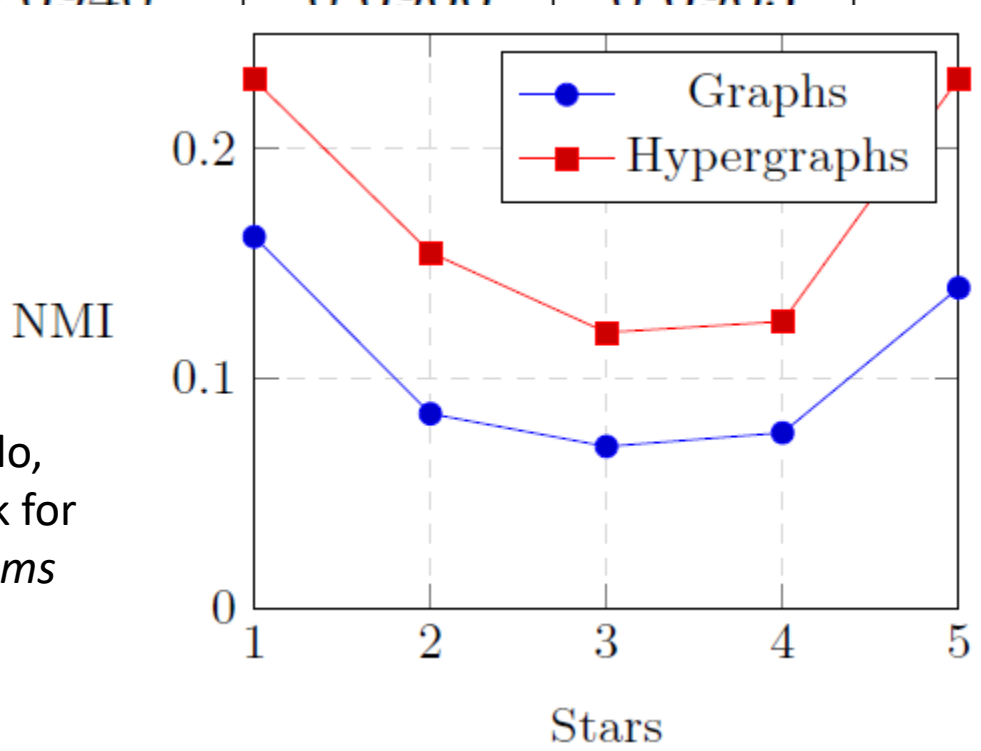
- Vertices - business
- Edges - users
- Hyperedge - "a single user has reviewed several restaurants"

- Total:
  - ~35k restaurants and ~1M reviews

source: Yelp Dataset Challenge - https://www.yelp.com/dataset/challenge

Yelp dataset use case

Stars	H_i $(\|V\|; \|E\|)$	City	State	Alcohol	Noise Level	Take Out	Category
1	(29479; 244671)	0.8833	0.9562	0.8166	0.8104	0.8176	0.8163
2	(28055; 173140)	0.8582	0.9462	0.7744	0.7651	0.7731	0.7702
3	(30369; 177792)	0.8132	0.9226	0.7075	0.6940	0.6966	0.6965
4	(32987; 301578)	0.7812	0.9081	0.6573	(		
5	(32558; 590320)	0.8027	0.9145	0.6963	(		
ALL	(35856; 950488)	0.7500	0.8985	0.6162	(		



Source: Antelmi, A., Cordasco, G., Kamiński, B., Prałat, P., Scarano, V., Spagnuolo, C., & Szufel, P. (2019, July). SimpleHypergraphs. jl—Novel Software Framework for Modelling and Analysis of Hypergraphs. In *International Workshop on Algorithms and Models for the Web-Graph* (pp. 115-129). Springer, Cham.