Inhibiting Spread in Hypergraphs through Community-Aware and Process-Based Node Removal

A comparative study under the SICP model

TU Delft, EEMCS | **Quint Winthagen** | CSE3000

1. Introduction

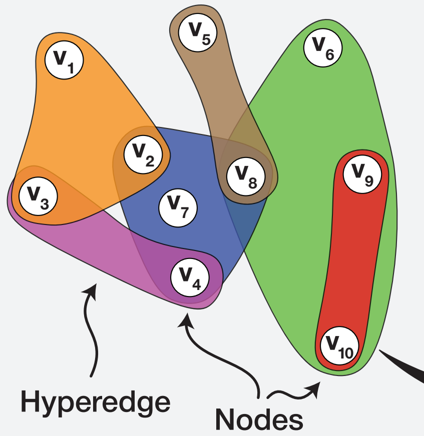
"Which nodes should we remove to effectively inhibit SICP spreading on a hypergraph?"

We investigate **14 strategies,** based on:

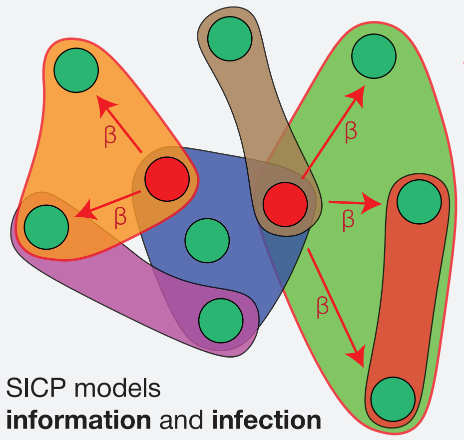
1. **Local Structure**
2. **Probability Model from spreading**
3. **Community Structure**

Hypergraphs

- A graph data structure, but **hyperedges** connect **2 vertices or more**
- Better representation of **complex networks** than regular graphs.
- Hyperedges are valuable in representing **group interactions**, e.g. social networks



Susceptible Infected Contact Process (SICP)



A node is either **Susceptible** OR **Infected**

Per timestep, for **each infected** node:

1. **One** incident hyperedge is chosen with **uniform probability**
2. **All nodes** in that hyperedges have a chance to be infected, with **probability β**

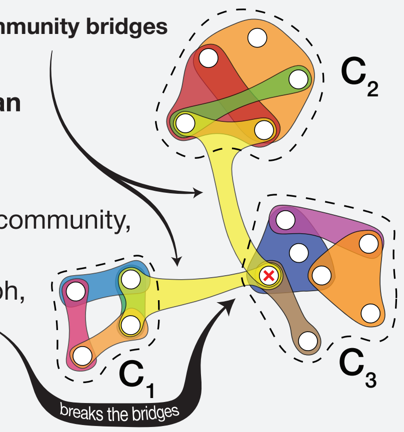
SICP models **information** and **infection** spreading **without recovery**

2. Auxiliary Structures

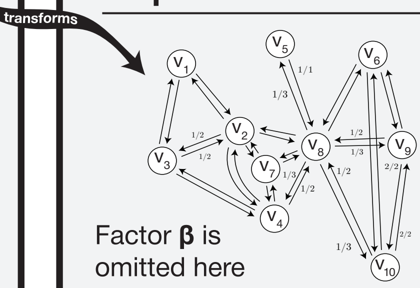
Communities

"A **group of nodes** that **appear together** in hyperedges **more often than expected** under a degree-preserving random hypergraph"

- Detected by the h-Louvain algorithm, which maximizes "**modularity**"
- Results in a partition of nodes where most connections occur inside a community, and **connections between communities are sparse.**
- Connections between communities **bridge** large parts of the hypergraph, **breaking those bridges** is interesting when **inhibiting spreading.**



Equivalent Probability Transformation (EPT)



Factor **β** is omitted here

- A **directed weighted graph** derived from the hypergraph, where edge weights give the **probability** that an infected node infects a neighbour, **in one SICP timestep.**
- Consequence of 2-step SICP process:
  1. **Choose one** incident hyperedge,
  2. Try infecting **all** nodes in that hyperedge

translates to → "what is the chance that node v chooses a hyperedge that also contains u?" (multiplied by β)

3. Strategies

A strategy is a function that **selects K nodes** to remove from the hypergraph, while **aiming to minimize spreading**

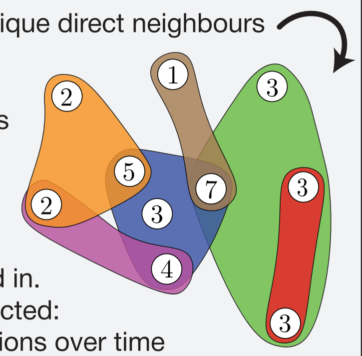
Local Structure -based

Degree (DEG): the number of unique direct neighbours

Hyperdegree (HDEG): the number of hyperedges that a node is contained in.

Average Hyperedge Size (AHS): the mean size of the hyperedges that a node is contained in.

- Shown that, when a node is infected: a high AHS leads to more infections over time



EPT -based (SICP-aware)

EPT In Strength (EPT-IS): sum of all probability **coming into** a node in the EPT. Targets nodes with a high change of being infected by their neighbours.

EPT Total Strength (EPT-TS): sum of probability **coming into** a node and **going out** of a node in the EPT. Sums the changes of being infected with the chances of infecting other nodes.

Community-aware (and SICP-aware)

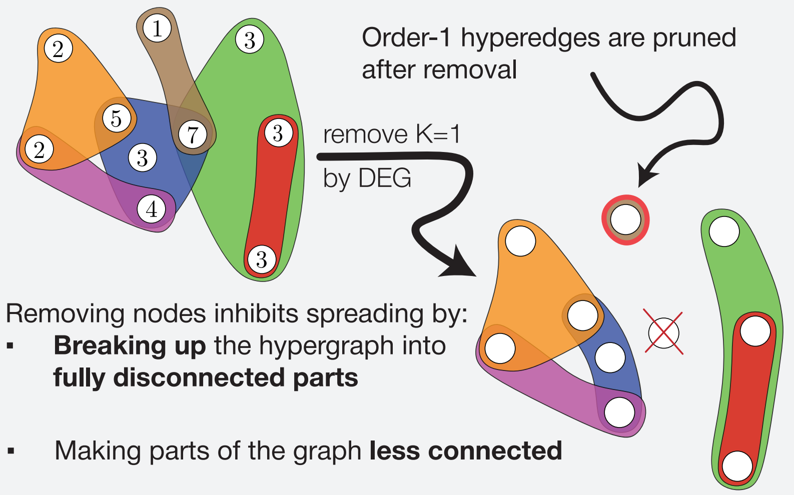
EPT Bridge In Strength (EPT-BIS) sum of all probability **coming into** a node in the EPT, **from** nodes **in other communities.** Targets nodes that are most likely to bring **infection into a community**

EPT Bridge Out Strength (EPT-BOS): sum of all probability **going out** of a node, **to** nodes in **other communities.** Targets nodes that are most likely to transmit the infection **to other communities.**

4. Node Removal

One-Shot Node Removal:
Calulate scores **once**, then rank and **remove top K** nodes

Adaptive Node Removal (strategies have -a prefix):
Remove the top ranked node, **recalculate scores**, **repeat K times**



Removing nodes inhibits spreading by:

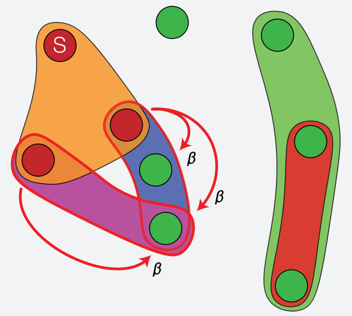
- **Breaking up** the hypergraph into **fully disconnected parts**
- Making parts of the graph **less connected**

5. Experimental Setup

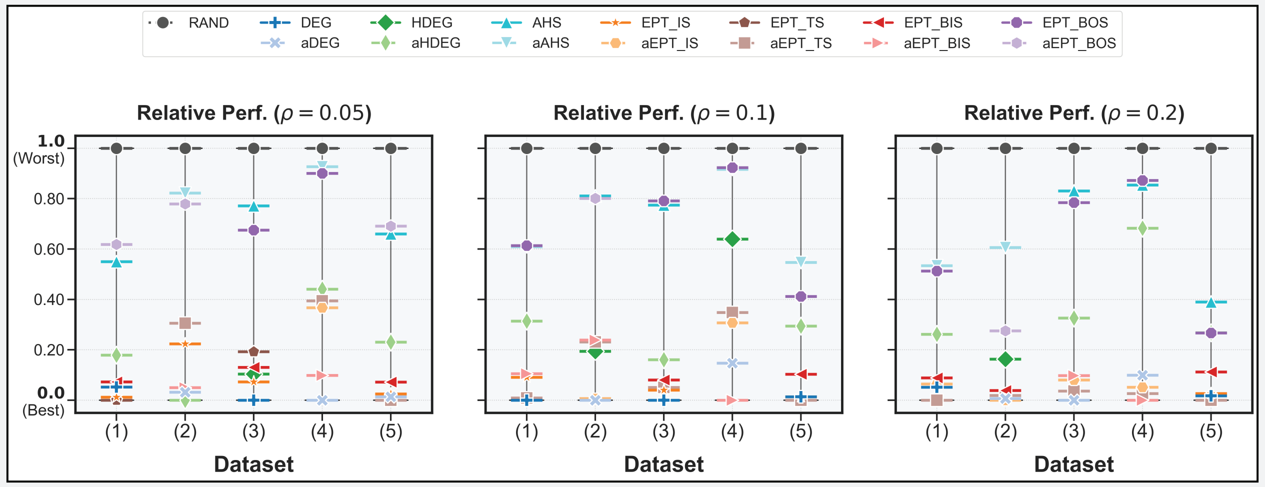
- We ran simulations on **6 real-world hypergraph datasets.**
- All datasets show evident **community structure**
- Over three **removal fractions:** 5%, 10%, 20% of all nodes in the hypergraph
- We use infection probability **β = 0.02**, and run the simulations until **T=25** timesteps

We ran **10 simulations for each seed node**

- **Random Baseline (RAND):** from all nodes, choose K nodes at random



6. Results and Conclusions



Conclusions

- **All** proposed strategies consistently **outperform random** removal
- Adaptive node degree (**aDEG**) is overall the **most robust strategy** for halting spread.
- **Adaptive** strategies typically **outperform one-shot** strategies.
- Complex **community** and **probability metrics** do **not** provide universal gains, though they do outperform aDEG in specific cases.
- **Influential seed nodes** make **poor targets** for hypergraph immunization.