HYPERHARMONIC CENTRALITY

SAMUEL RODRIGUEZ, MONIQUE RAMPERSAUD, & ONYINYECHI EMEROLE
Cornell University
ILYA AMBURG - FACULTY RESEARCH MENTOR
Cornell University Center for Applied Mathematics

INRODUCTION

Harmonic Centrality is a global measure of importance of a node within a network based on its distance from other nodes. It is given by the following formula:

$$c_H(x_i) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{\text{dist}(x_i, x_j)}$$

where n is the total number of nodes and i and j are any arbitrary nodes.

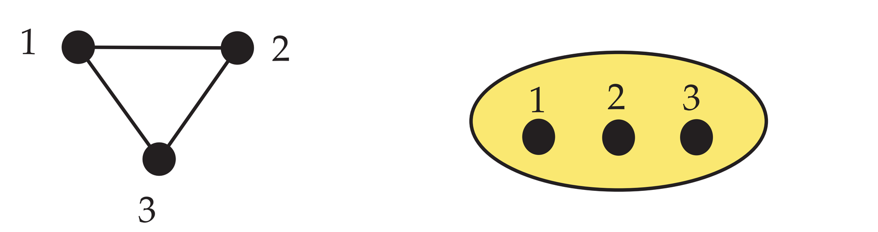
- Sums the reciprocal of the shortest path distances from all other nodes to node x_i
- Finds the harmonic mean of the shortest path among all nodes in a graph
- Gives preference to nodes with shorter paths
- Allows us to compute shortest-path-based centrality for unconnected networks

In this project, we create a way to apply the harmonic centrality measure on higher-order graphs.

Many real-world relationships reflect connections that require higher-order graphs (e.g., email networks, friendship networks, co-authorship).

- **Hypergraphs** allow modelling of simultaneous connections between 3+ entities whereas simple graphs are limited to pairwise connections

For example:



only pairwise connections | all nodes simultaneously connected

Practical applications include:

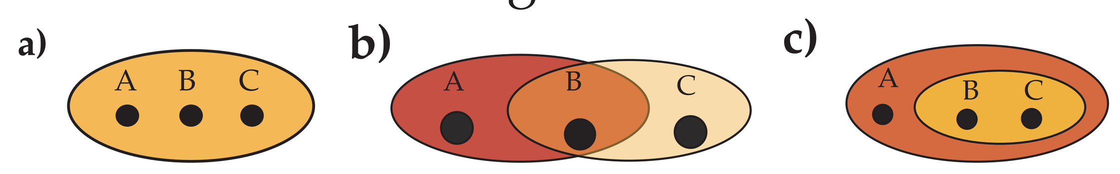
- Choosing a path that will spread a message to the most people in the shortest amount of time
- Understanding which geographic locations are most vulnerable to a disease
- Choosing where to place a facility in a network

THEORY

We measure *walks* in a hypergraph as follows:

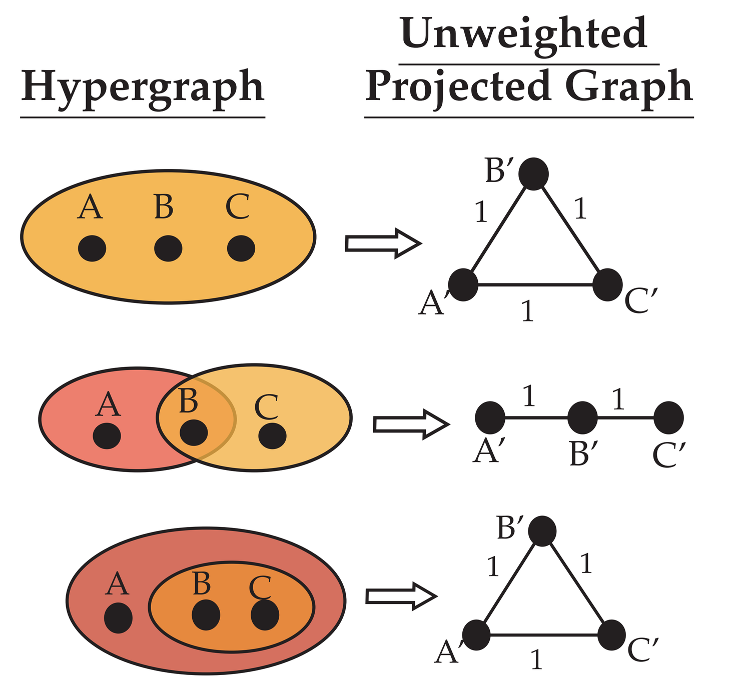
- Travelling from one node to another in the same hyperedge has a cost of 1.
- Travelling to a node that appears only in the adjacent hyperedge has a cost of 2.

Figure 1



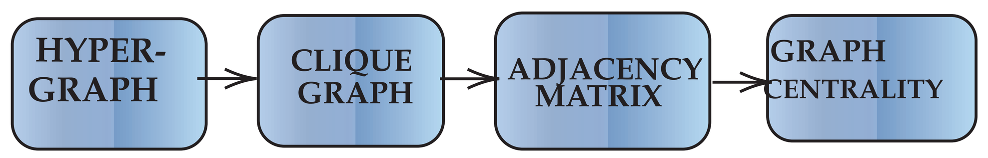
Let $l_s(u,v)$ be the length of the shortest path from node u to node v. Then:
In figure 1a: $l_s(A,B) = l_s(B,C) = l_s(A,C) = 1$
In figure 1b: $l_s(A,B) = l_s(B,C) = 1$ and $l_s(A,C) = 2$
In figure 1c: $l_s(A,B) = l_s(B,C) = l_s(A,C) = 1$

THEOREM: For any nodes u and v in hypergraph H and corresponding nodes u' and v' in the unweighted projected graph G, we have $l_s(u,v) = l_s(u',v')$



This unweighted graph approach could be used to generalize any path-based graph centrality

WORKFLOW



DATA STATISTICS

We use a collection of emails from Enron to generate a network where we have nodes as emails and hyperedges link the sender and recipient(s) of an email. Its statistics are as follows:

NODES	HYPEREDGES	PROJECTED GRAPH EDGES
143	10551	1800

RESULTS

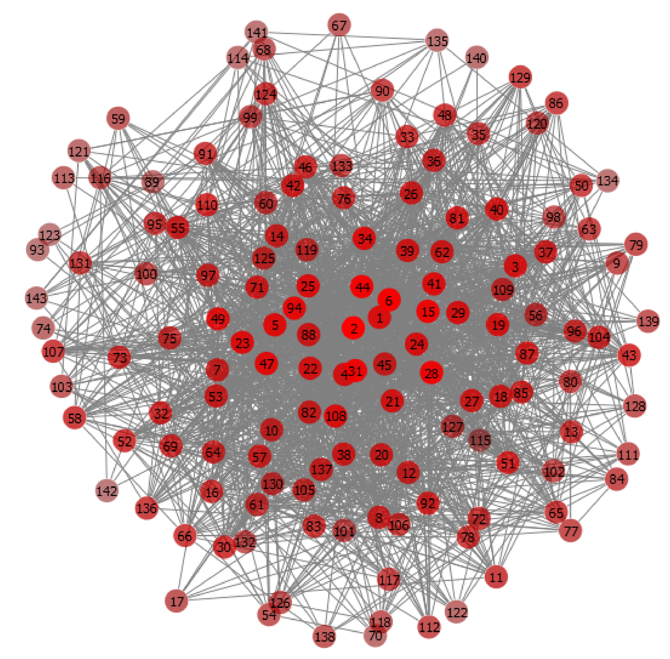


Figure 2: Harmonic centrality applied to unweighted projected graph

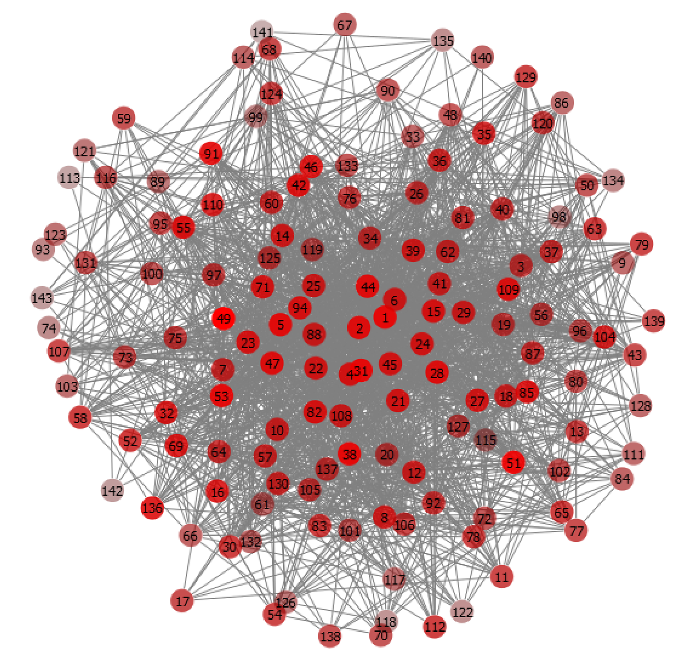


Figure 3: Harmonic centrality applied to weighted projected graph

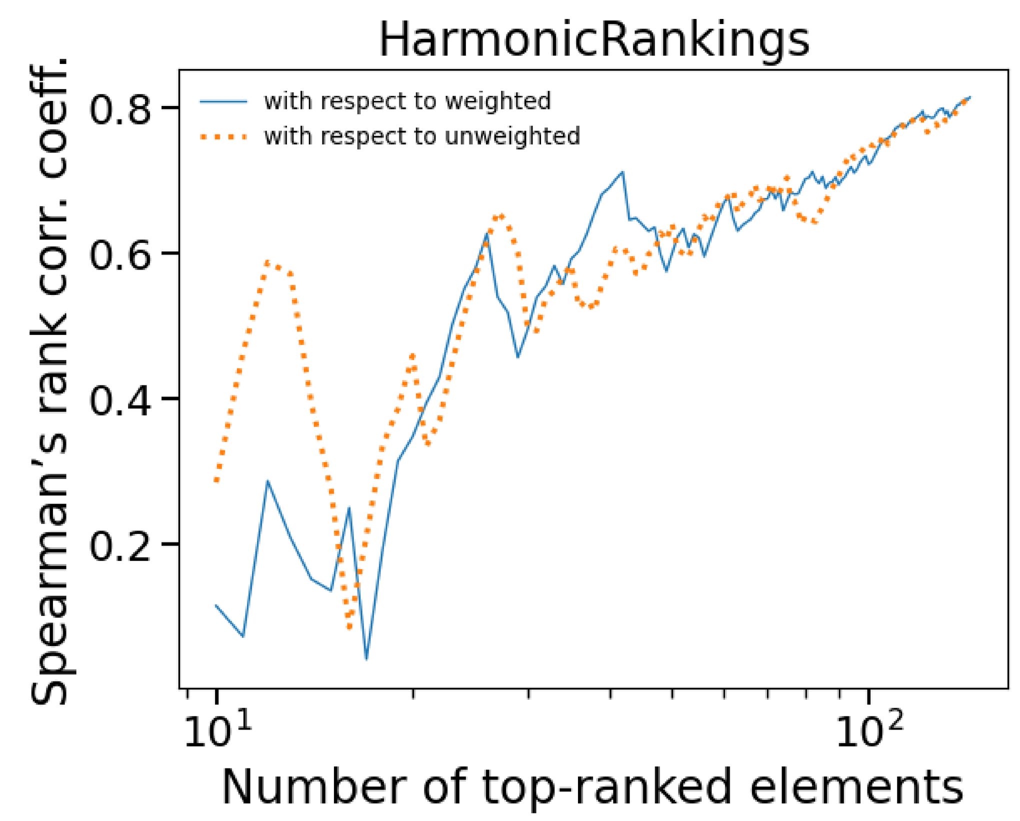


Figure 4: Harmonic centrality node-by-node comparison

- For *Figure 2* and *Figure 3*, color intensity = higher centrality
- There is a noticeable difference in the centrality obtained between weighted and unweighted
- We use L_2 normalization (0-1 scale) to measure the "distance" between the two centrality vectors
- L_2 difference = 0.11
- *Figure 4* shows approx. the first 20 nodes show the greatest disparity and the difference in centrality induced rank decreases as the node number increases

FUTURE RESEARCH

- Include weighted edges in framework
- Extend the measure to attributed and annotated hypergraphs as well

CODE

https://github.com/SamRod33/HyperharmonicCentrality

FUNDING

This research is supported by Engineering Summer Math Institute (ESMI), Diversity Programs In Engineering (DPE), Cornell University.

REFERENCES

[1] Yannick Rochat. Closeness centrality extended to unconnected graphs: The harmonic centrality index. Technical report, 2009.

[2] J. Gao et al. Dynamic shortest path algorithms for hypergraphs. *IEEE/ACM Transactions on Networking*, 23(6):1805–1817, 2015.