INFLUENTIAL NODES DETECTION ON COMPLEX NETWORKS VIA NETWORK TOPOLOGY

Sidra Jawaid

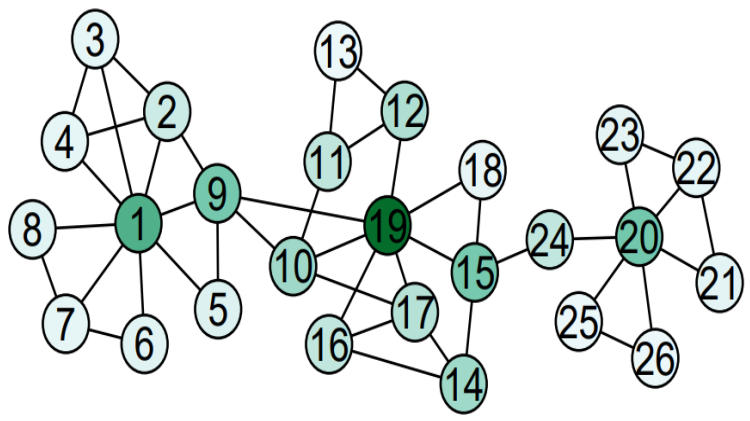
University of Central Oklahoma

In collaboration with

Dr. Mehmet Aktas
Dr. Esra Akbas
Ebony Harrington

INTRO: IDENTIFYING INFLUENTIAL NODES

- What is a network?
- Objects in a network are nodes or vertices.
- Connections between nodes are called edges.
- Influential node detection is an important graph mining problem.
- It has many important practical applications.



Facebook Network

Darker Node = More Influential

OUR METHOD: DIFFUSION FRÉCHET FUNCTION (DFF)

- DFF takes both network topology and attribute information into consideration.
- Heat source on each vertex or node.
- Finds similarity of vertices through differences in heat diffusion distances.
- Sums up heat diffusion distances for each node and finds weighted average.
- Varying parameter values.

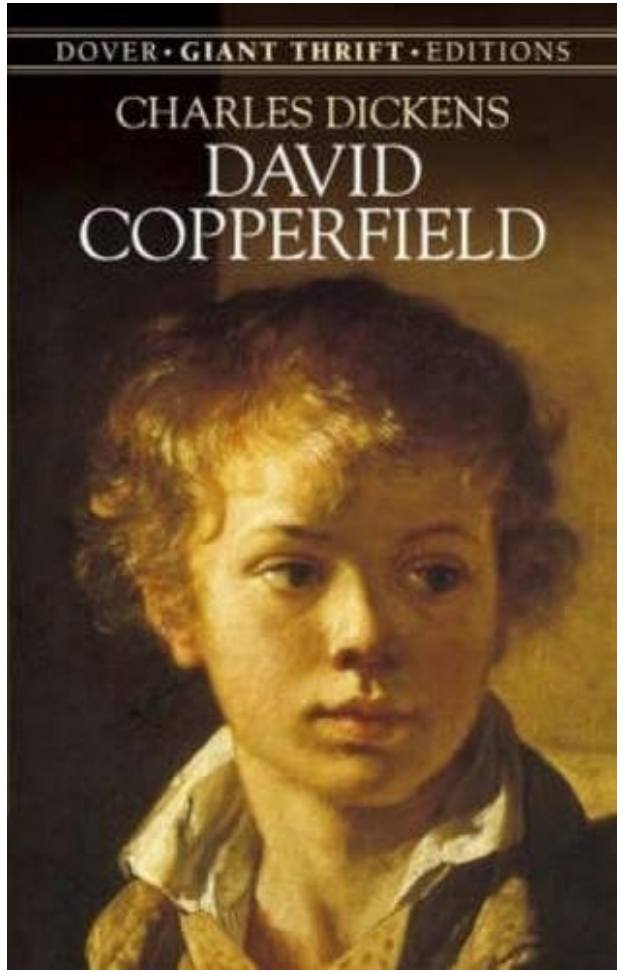
$$F_{\xi,t}(i) = \sum_{j=1}^{n} d_t^2(i,j)\xi_j$$

OVERVIEW

1. Apply proposed method Diffusion Frechet Function (DFF) with varying parameter values to real world data sets.
2. Apply centrality measures to real world data sets.
3. Rank nodes found by DFF and centrality measures from most to least influential.
4. Apply SIR model to influential nodes found by DFF as evaluation method to find most effective parameter value.
5. Apply SIR model to influential nodes found by DFF and centrality measures as evaluation method to compare performances.

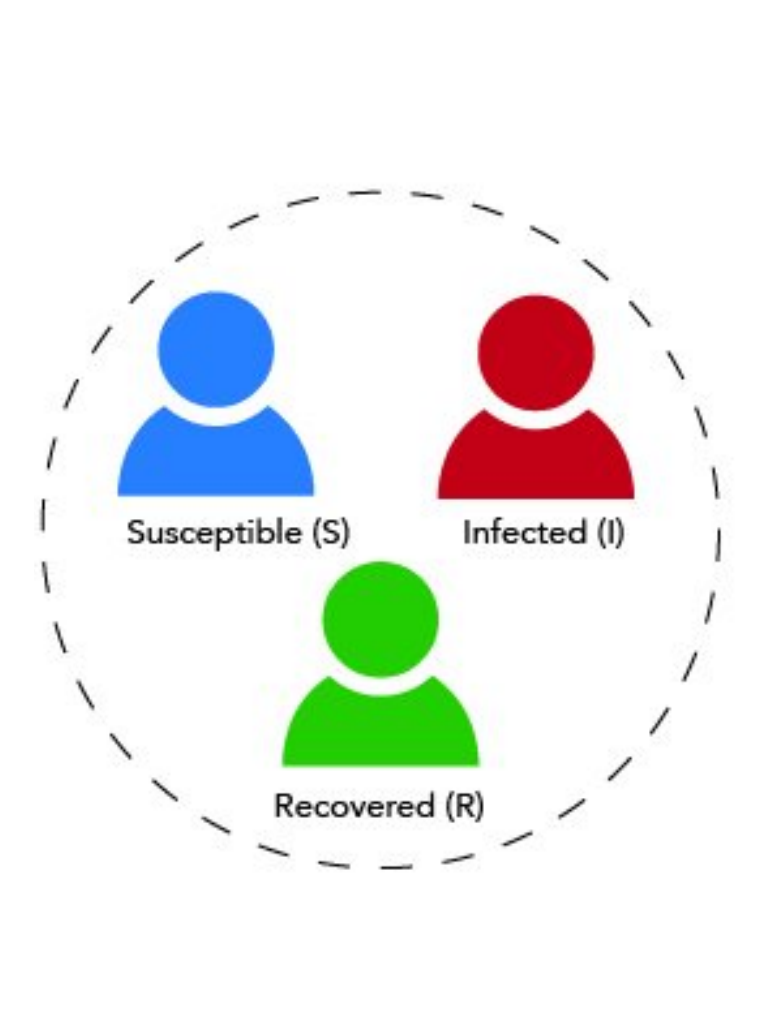
RESULTS: DATASETS

- **High School** - a network that consists of boys from a highschool in Illinois with an edge representing friendship between two boys as reported by one of them. Consists of 70 nodes and 366 edges.
- **Les Miserables** - a network of characters in chapters of *Les Misérables* written by Victor Hugo. An edge represents the appearance of two characters in the same chapter. Consists of 77 nodes and 254 edges.
- **Copper** - network of adjectives and nouns in the novel *David Copperfield* by Charles Dickens. Consists of 112 nodes and 425 edges.
- **Netscience** – a coauthorship network of scientists working on network theory and experiments. Consists of 1536 nodes and 2742 edges.

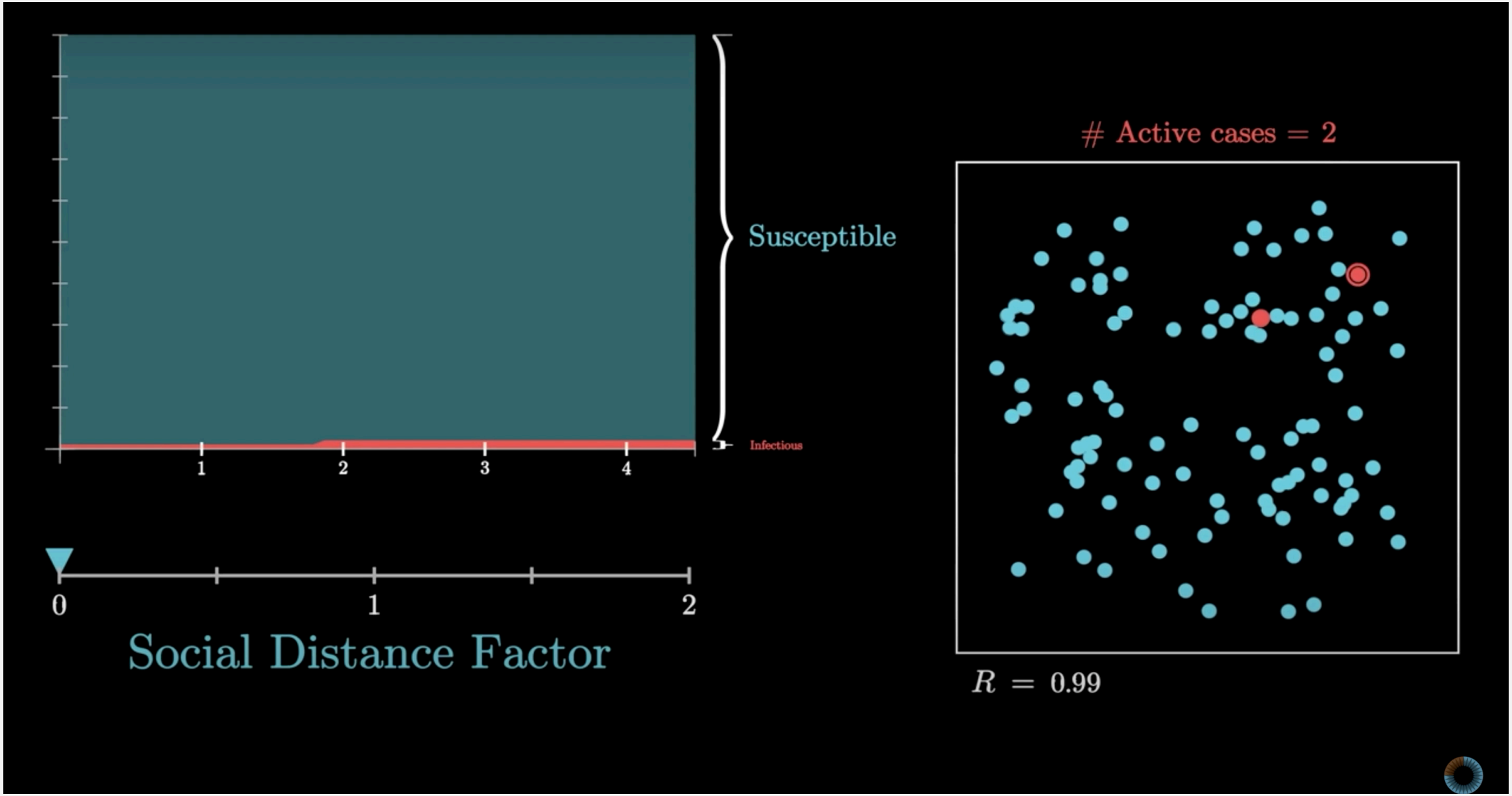


EVALUATION METHOD: SIR MODEL

- SIR model determines spreading ability of nodes.
- Each classified into one of three states: Susceptible nodes (S), Infected nodes (I), and Recovered nodes (R).
- Node(s) are selected as Infected (I) and all other nodes in network are Susceptible (S).
- In each propagation iteration, each Infected (I) node infects its direct neighbors with probability μ.
- Simultaneously, each infected node will be Recovered (R) with probability β.

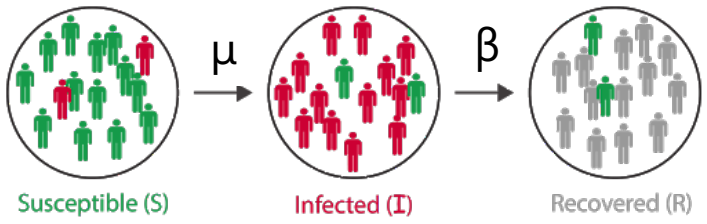


Simulating An Epidemic by 3Blue1Brown



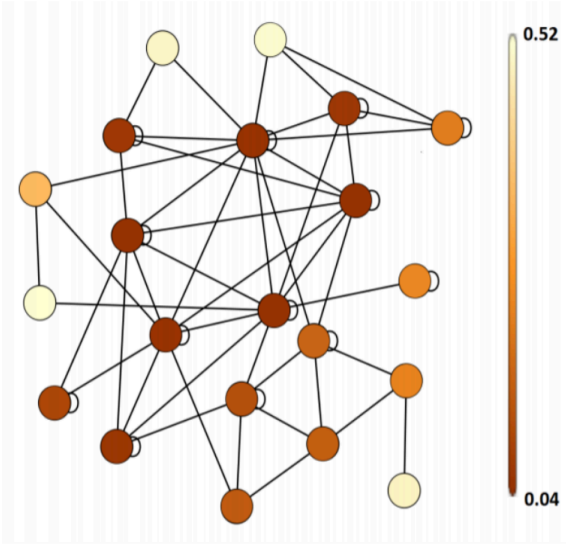
EXPERIMENTS
SIR MODEL

- We set the infection rate as $\mu=1.5\frac{\langle k\rangle}{\langle k^2\rangle-\langle k\rangle}$.
- The recovery rate was set as $\beta=1$ for simplicity.
- The experiment was run 100 times and the average was taken.
- Two types of results: number and time.

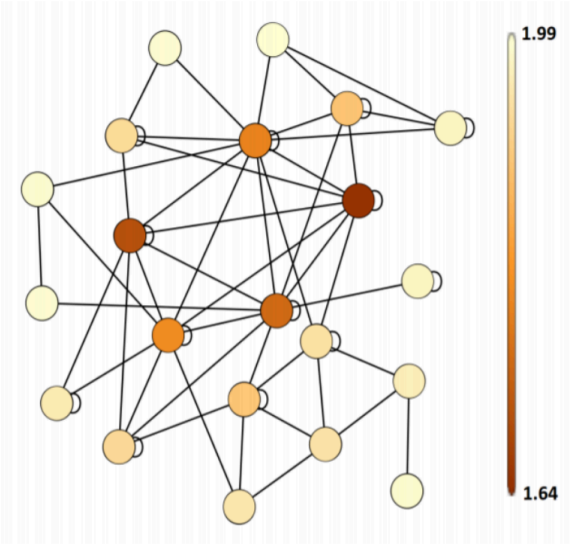


RESULTS: VARYING DELTA VALUES

Delta value in DFF determines locality of information



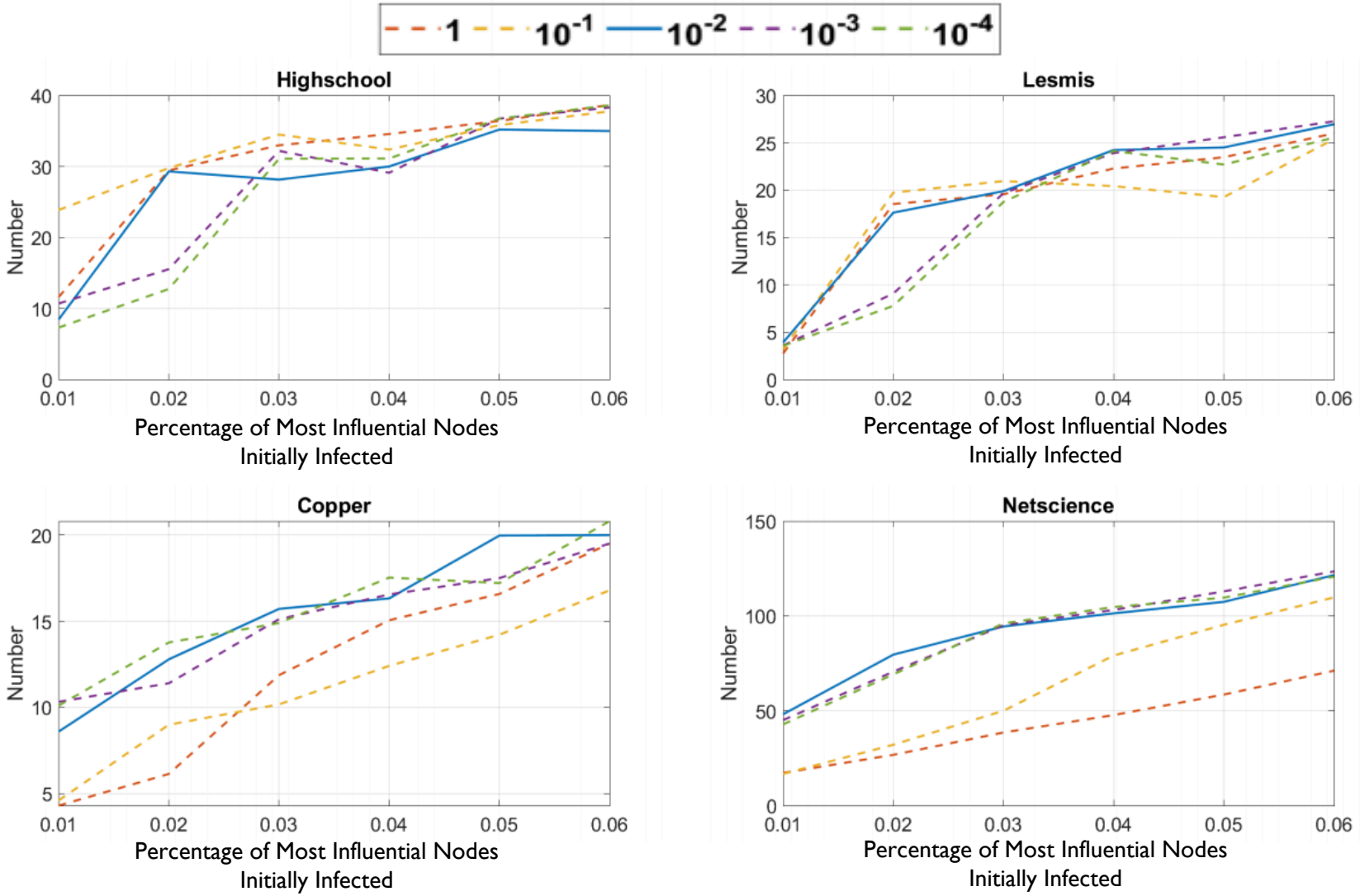
(a) $t = 10^{-1}$



(b) $t = 10^{-7}$

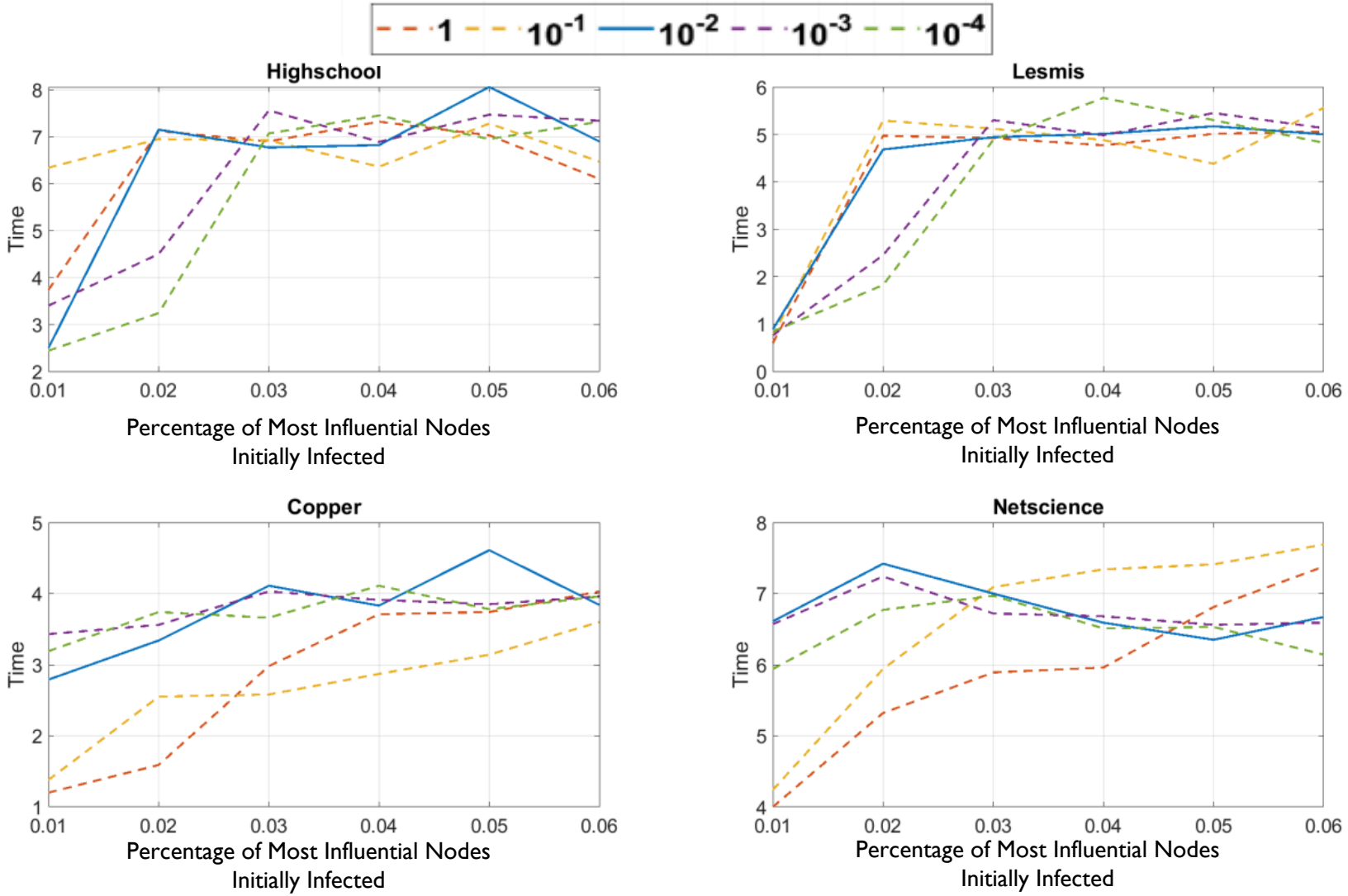
Darker color = More Influential

Greater Number Value = More Effective



RESULTS: VARYING DELTA VALUES

Greater
Time Value
=
More Effective



RESULTS: VARYING DELTA VALUES

RESULTS: CENTRALITY MEASURES

Betweenness Centrality - measures how often each graph node appears on a shortest path between two nodes in the graph.

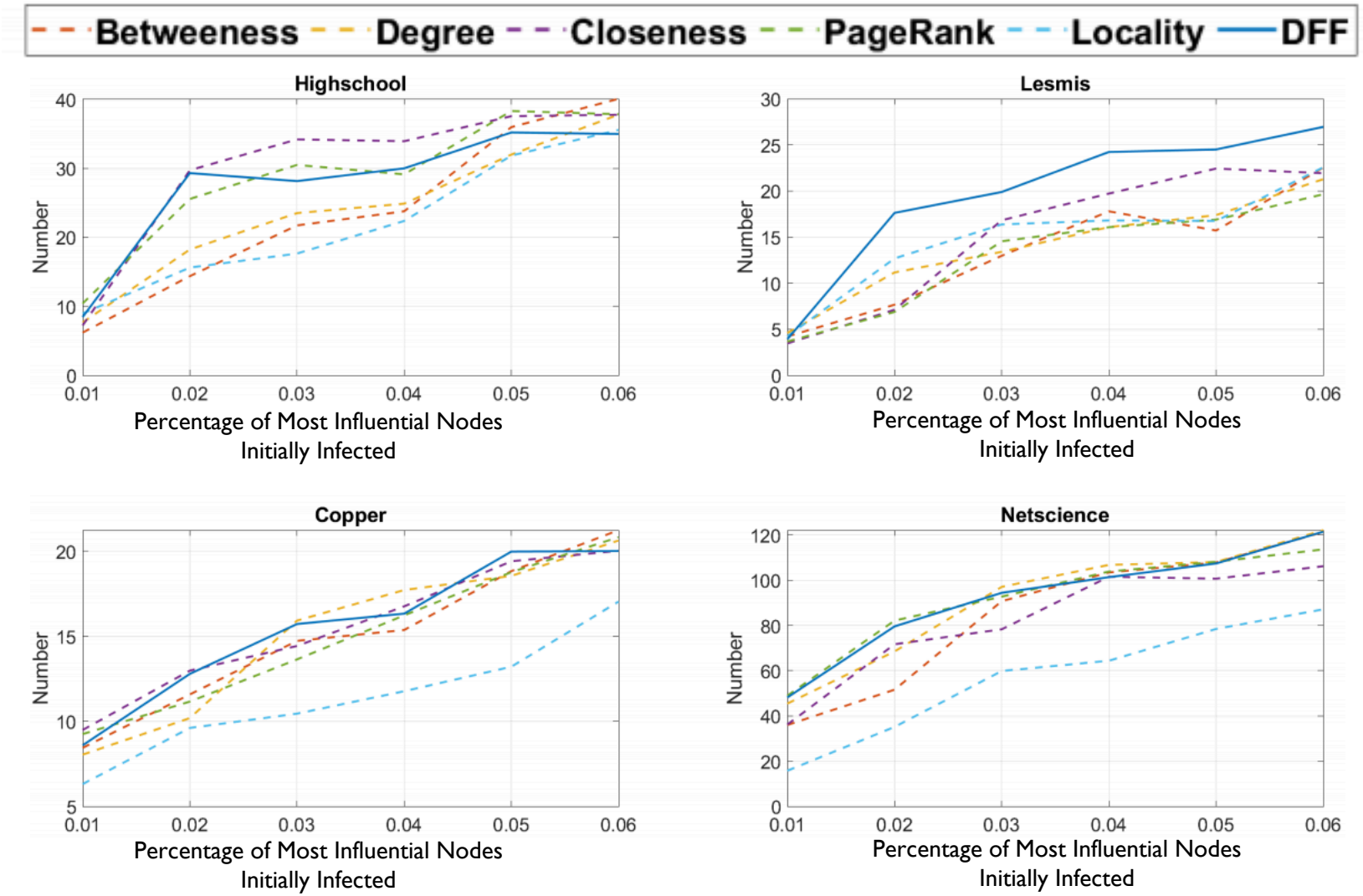
Degree – how many neighbors or connections a node has.

Closeness – uses the inverse sum of the distance from a node to all other nodes in the graph.

PageRank - At each node in the graph, the next node is chosen with the probability from the successors of the current node. The centrality score is the average time spent at each node during the random walk.

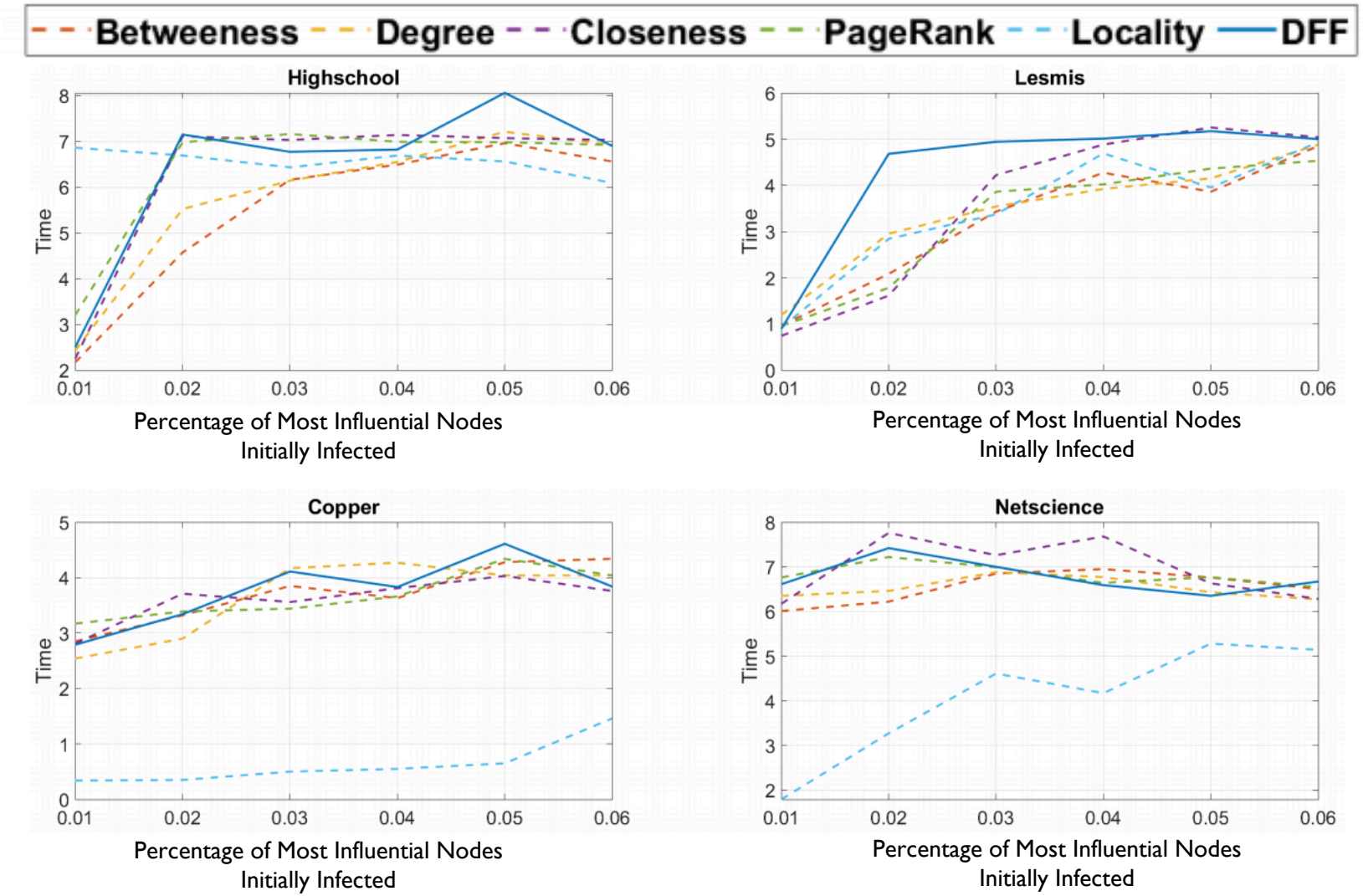
Locality – centrality based on neighbors and successive neighbors of a node.

Greater
Number Value
=
More Effective



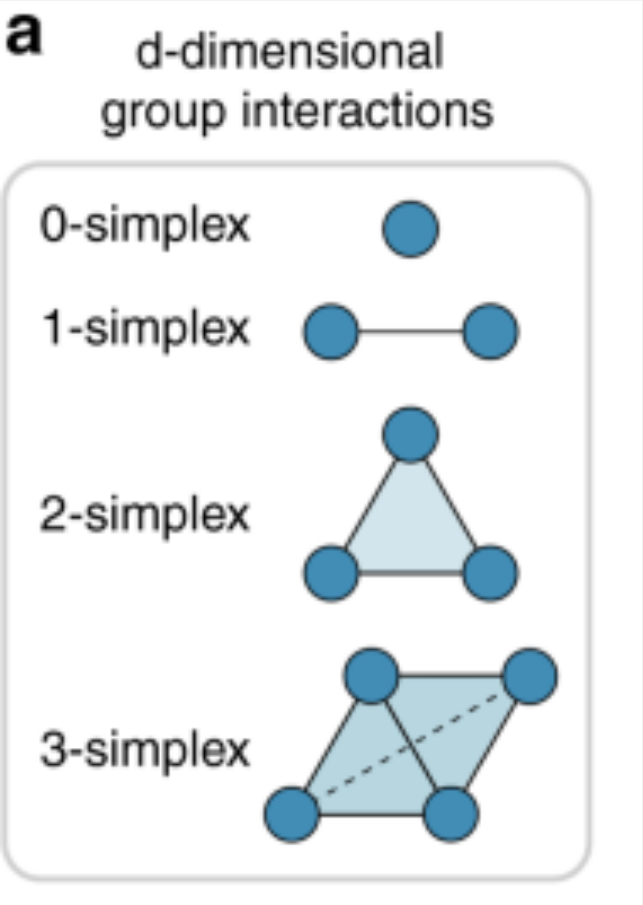
RESULTS:
CENTRALITY MEASURES

Greater
Time Value
=
More Effective



RESULTS:
CENTRALITY MEASURES

FUTURE WORK


a
d-dimensional
group interactions
0-simplex
1-simplex
2-simplex
3-simplex

SPECIAL THANKS TO DR. AKTAS, DR. AKBAS, EBONY HARRINGTON, AND DR. BANNISH AS WELL AS NCUWM

THANK YOU!