**Study Design**

**Data Generation**

1. Graph type selection: Random, Scale free, Small-world and SF & SW (Holme & Kim)
2. Graph sizes in 3 settings:
   1. 30, 35, … 50
   2. 100, 200, ..., 4500
   3. 5000, 6000, ..., 10000

**Attributes**

1. Network traits:
   1. Network size
   2. No. of Edges
   3. Edge-Degree Ratio
   4. Max. Degree
   5. Avg. Path Length
   6. Clustering Coefficient
   7. Diameter
   8. Density
   9. Distance
   10. No. of Triads
   11. Assortativity
   12. Graph Girth
2. Node traits
   1. Degree
   2. Closeness
   3. Betweenness
   4. Eigenvalue Centrality
   5. Eccentricity
   6. Page Rank
3. Resilience of each node:
   1. Optimal method for size up to 50 nodes
   2. Greedy method for size over 50 nodes
4. Time taken for each trail
5. Label top ‘k’ Nodes as *Influential*
   1. By Degree
   2. By Closeness
   3. By Betweenness
   4. By Eigenvalue
   5. By Page Rank

**Machine Learning Model**

1. Normalize traits
2. Training classification models
3. Linear Regression (glm)
4. Decision Trees (rpart)
5. Random Forest (randomForest)
6. Support Vector Machines (svm)
7. Neural Networks (avNNet)
8. Boosting (C5.0)
9. Deep Learning (mxnet)
10. Quantify model learning times

**Result Evaluation**

1. Generate test data (similar to training data with different seed)
2. Normalize test data
3. Create Confusion matrix for ML models
4. Compare results
   1. Accuracy
   2. Precision
   3. Recall
   4. F1-Score
5. Compare training time with Optimal/Greedy methods

**Evaluation on Scale**

1. Data sets:
   1. Twitter
   2. Epinions
   3. Actor Network
   4. Other recommended networks
2. Classify influential nodes via ML models
3. Classify influential nodes via Heuristics