Let me Explain! Exploring Large-Scale Networks with Mechanism Characteristics

Jasmine DeHart and Christan Grant, Ph.D.

School of Computer Science, Gallogly College of Engineering, University of Oklahoma

oudatalab.com

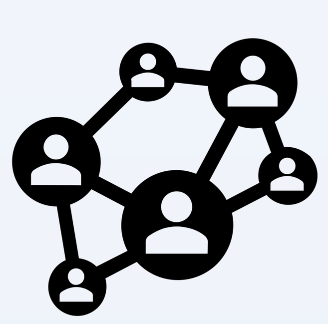
BACKGROUND

Understanding and explaining the intricacies of a large and complex network can be difficult. Large and complex networks can include protein-protein networks, food webs, social networks and genetic structures. Researchers have begun to describe large networks using models and measurements. One method for describing graphs is the use of grammars. Grammars describe operations to summarize these large graphs in an expressive way. Given the massive amounts of possible grammars, researchers cannot use grammars to describe graphs [1]. Another way to understand networks is with the use of centrality measures. Researchers have suggested centrality measurements to give meaning to their findings [2].

Does **characterizing** complex networks using mechanism **generation** and **verification** improve explainability?

MOTIVATION

There exists a graph, $G=(V,E)$, with nodes ($v \in V$) and edges ($e \in E$). Researchers use graphs to represent different ecosystems or domains. The structure represents the relationships in that domain or ecosystem.

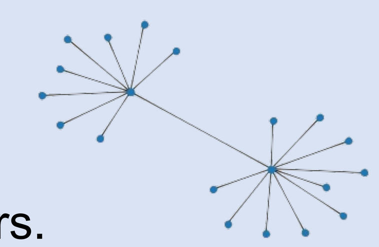


As these graphs are created and evolved, they can become quite large and complex. The complexity of these graphs is increased with the inclusion of multiple properties to represent its' features. Which leads to the question, how can we describe large and complex graphs?

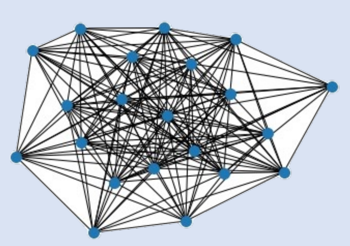
WHAT IS A MECHANISM?

A network structure or developmental process which relies on a set of constraints and features.

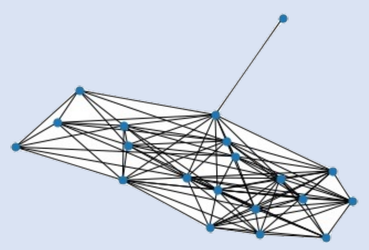
Divergence and Duplication
A node in the network is duplicated and given a probability of divergence. The duplicate is attached to its neighbors.



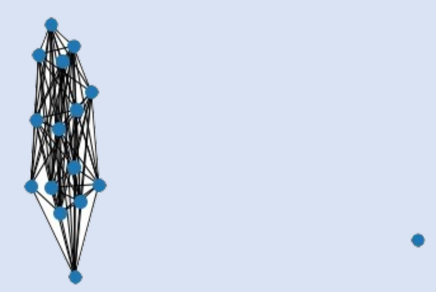
Erdős–Rényi model
All nodes are connected using a probability threshold. Each node is connected based on this probability.



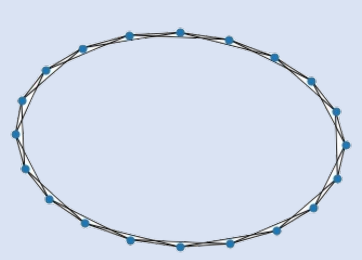
Niche model
This mechanism represents a hierarchy or food chain.



Preferential Attachment
Nodes prefer to connect with popular nodes in the network.



Small World
A small world network asserts that there are no more the 6 degrees between two nodes.



EXPLORING MECHANISM BREEDING ALGORITHMS

We develop mechanism generation algorithms to simulate the intentional creation of realistic graphs with mechanisms. The architecture for each mechanism is displayed in Figure 1.

Minimal Inter-Mechanism Breeding Algorithm
MIMBA: A → C → E
Breeds each mechanisms **individually** and combines the mechanisms using a **random sample** of nodes from each after growth.

Organic Inter-Mechanism Breeding Algorithm
OIMBA: B → C → D
Breeds each mechanisms **interdependently** and selects nodes **before** each mechanism's growth process.

Stochastic Inter-Mechanism Breeding Algorithm
SIMBA: A → C → F
Breeds each mechanisms **individually** and combines and **rewires** a **random sample** of nodes and edges from each mechanism **after** growth.

VERIFYING NETWORK STRUCTURES

Large graphs contain multiple mechanism types. To identify mechanisms in graphs, researchers must be able to break down these graphs into smaller components or summarize graphs as embeddings.

Dissecting Communities (Figure 2a)
We use community detection to identify substructures. We assume that these communities contain a mechanism.

1. Identify communities in graph
2. Create null mechanisms for each community
3. Compare community to template mechanisms

Finding the closest matching mechanism
Identifying closest mechanism may require several techniques. We explore a combination of:

1. Image Comparison for large graphs
2. Centrality Measures for any size graphs
3. Graph Edit Distance for small graphs

Graph Neural Networks
Using GNNs to dissect graph features to explore the existence of mechanisms in a graph. Graph features can be extracted using embeddings.

- Preserves the graph's structure
- Compresses representations of a graph
- Captures the topology of a graph

Node and Graph Embeddings (Figures 2b & 3)
Node embeddings are a way of representing graph nodes as vectors. Graph embeddings are vectors that represent all the nodes and edges in a graph. Those embeddings can be used to compare similarity between graphs or make predictions for mechanism types.

FUTURE WORK

Further studies should investigate understanding privacy vulnerabilities across mechanisms, demonstrating critical applications using mechanisms, and compressed graph evaluation.

REFERENCES

[1] Hibshman et al. Towards Interpretable Graph Modeling with Vertex Replacement Grammars. 2019

[2] Freeman, L. Centrality in Social Networks Conceptual Clarification. 1978

FIGURE 1. Mechanism Breeding Flow Chart

We visualize the architecture for the Mechanism Breeding Algorithms.

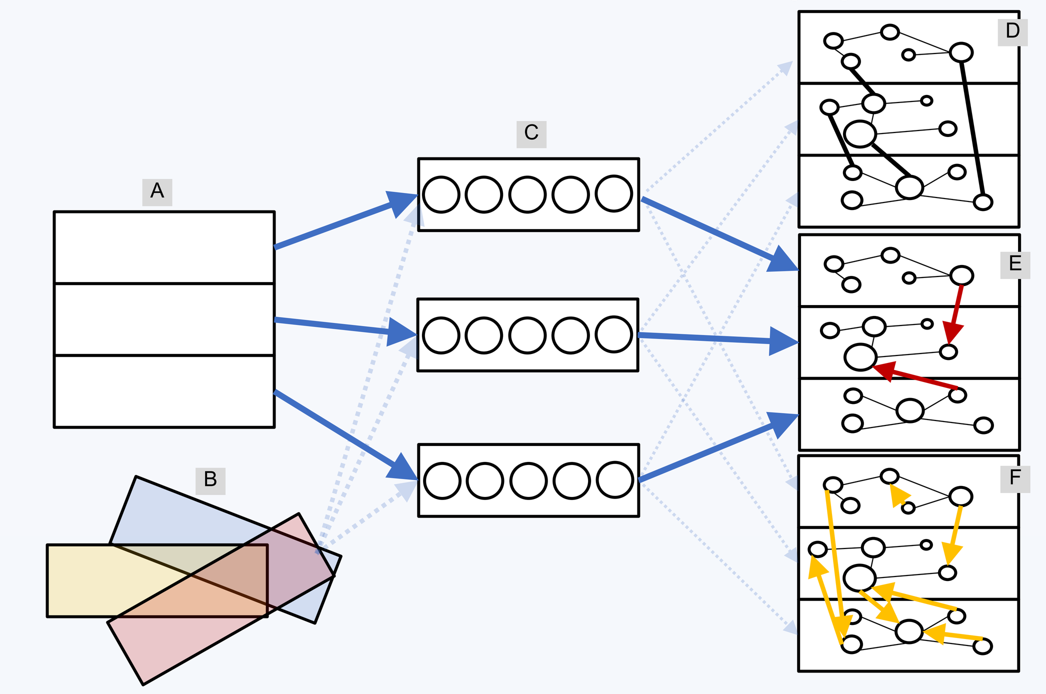
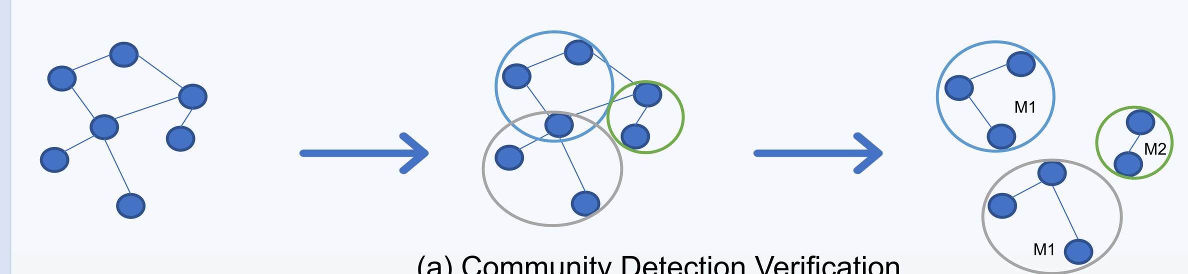
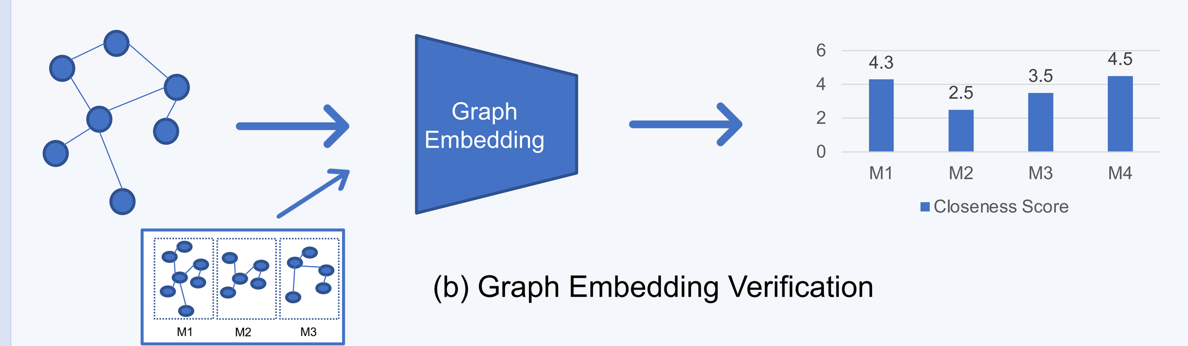


FIGURE 2. Verification Methods



(a) Community Detection Verification

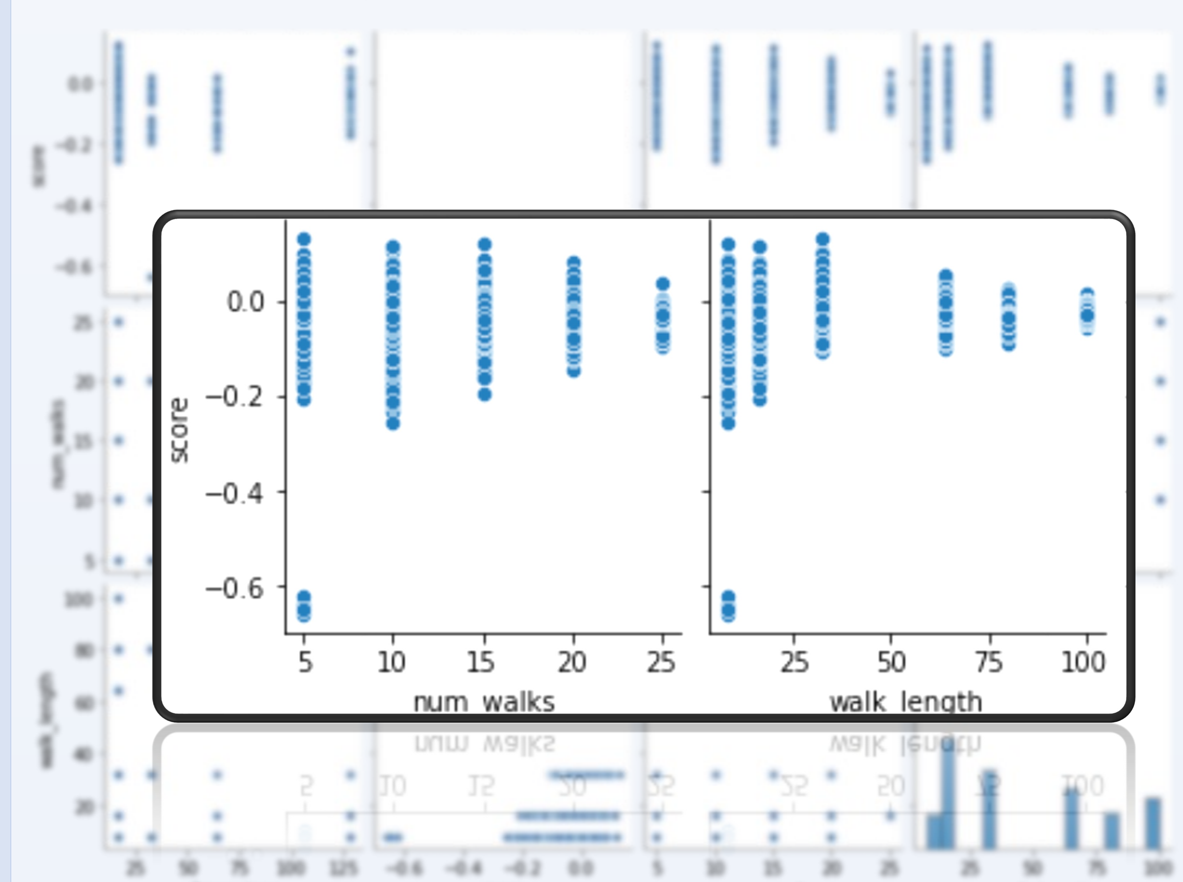


(b) Graph Embedding Verification

FIGURE 3. Exploring Embedding Parameters

We explore possible parameter configurations for node2vec algorithm on a small graph; generating 36,000 parameter configurations. We find num_walk @ 5 and walk_length @ 8 parameters can produce the best scores.

Since there is no ground truth dataset for mechanisms, we generate several different pairs of mechanisms graphs. We extract the embeddings from the mechanism graphs and score the embedding over of the all mechanisms.



Acknowledgements

The researchers are partially supported by an award from the National Science Foundation under Grant No. #1952181.

Contact us!

Jasmine DeHart
dehart.jasmine@ou.edu

Christan Grant, Ph.D.
cgrant@ou.edu

SCAN ME