Leveraging Graphical Diversity Under Constraints to Predict Structure

Nicole Woytarowicz, Brennan McConnell, Jonathan Lavington

University of Colorado, Boulder: CSCI 5352

What is Graph Diversity?

Graph diversity considers the number of possible realizations a graph can take on under specific constraints (e.g., degree sequence, diameter, global clustering coefficient, etc.).

What is Graph Diversity?

Graph diversity considers the number of possible realizations a graph can take on under specific constraints (e.g., degree sequence, diameter, global clustering coefficient, etc.).

Our project focuses on the investigation of how the imposition of certain constraints can implicitly affect the range of other metrics and how we can leverage underlying structure in a predictive manner.

Graph Diversity and the Random Graph Models

The configuration model generates graphs with a specific degree sequence D under maximum entropy.

Graph Diversity and the Random Graph Models

The configuration model generates graphs with a specific degree sequence D under maximum entropy.

Alderson & Li used this to empirically investigate how the s metric (a measure of the extent to which a graph has a hublike core) varies for graphs with different degree sequences.

Graph Diversity and the Random Graph Models

The configuration model generates graphs with a specific degree sequence D under maximum entropy.

Alderson & Li used this to empirically investigate how the s metric (a measure of the extent to which a graph has a hublike core) varies for graphs with different degree sequences.

Ultimately they showed that graphs with the same degree sequence can have considerable structural differences.

Graph Diversity and the Random Graph Models

For our project, we try to extend this idea to investigate the following questions...

- What do distributions of statistics resemble for various random graph models (i.e., how does the underlying structure represent itself)?

Graph Diversity and the Random Graph Models

For our project, we try to extend this idea to investigate the following questions...

- What do distributions of statistics resemble for various random graph models (i.e., how does the underlying structure represent itself)?
- Under imposed constraints, how are these statistics affected?

Graph Diversity and the Random Graph Models

For our project, we try to extend this idea to investigate the following questions...

- What do distributions of statistics resemble for various random graph models (i.e., how does the underlying structure represent itself)?
- Under imposed constraints, how are these statistics affected?
- Can we utilize restrictive measures (such as the joint degree distribution) to compare graph structure in order to predict which random graph model best represents a given graph?

Graph Diversity and the Random Graph Models

For our project, we try to extend this idea to investigate the following questions...

- What do distributions of statistics resemble for various random graph models (i.e., how does the underlying structure represent itself)?
- Under imposed constraints, how are these statistics affected?
- Can we utilize restrictive measures (such as the joint degree distribution) to compare graph structure in order to predict which random graph model best represents a given graph?
- Can we recover missing edges in such a way that the original structure is maintained/recreated by leveraging observed constraints alongside known random graph structure?

Statistical Distributions for Random Graph Models

1. Do random graphs have structural *fingerprints* that distinguish them from one another?
2. That is, does a graph's structure produce varying statistical differences across various metrics?

Graph Diversity - Preferential Attachment Model

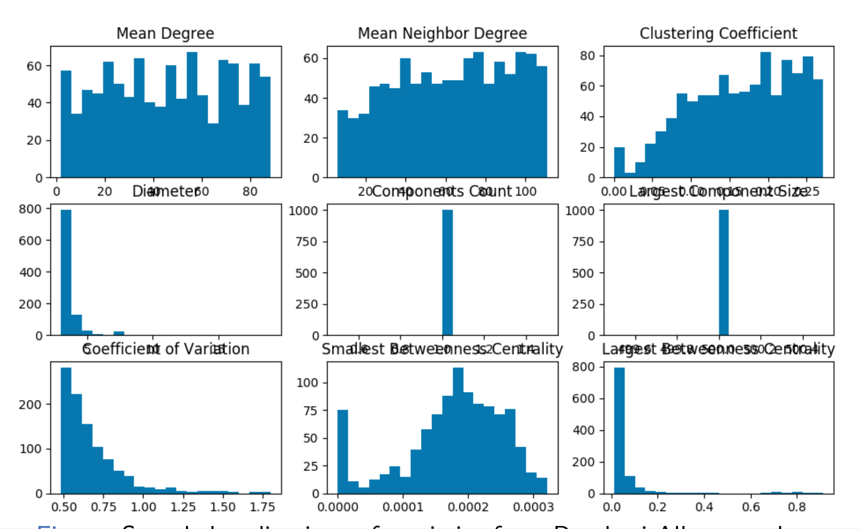


Figure: Sampled realizations of statistics for a Barabasi-Albert random graph

Graph Diversity - Small World Effect Model

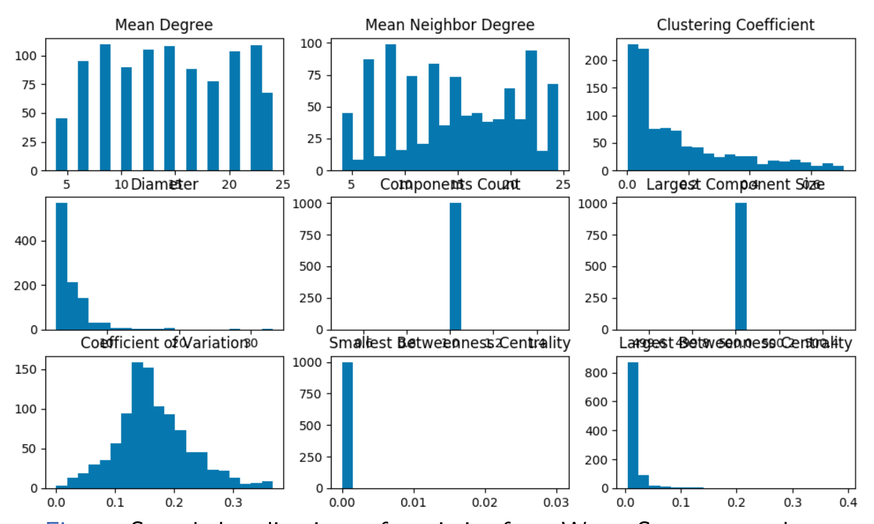


Figure: Sampled realizations of statistics for a Watts-Strogatz random graph

Given that we have observed realizations across random graphs that allow for large variation in random graph parameters and metric outcomes, we can impose constraints on specific metric distributions and/or model parameters in order to make the fingerprint more uniquely identifiable.

Furthermore, we can iteratively add more constraints to further restrict the realization space of a type of random graph.

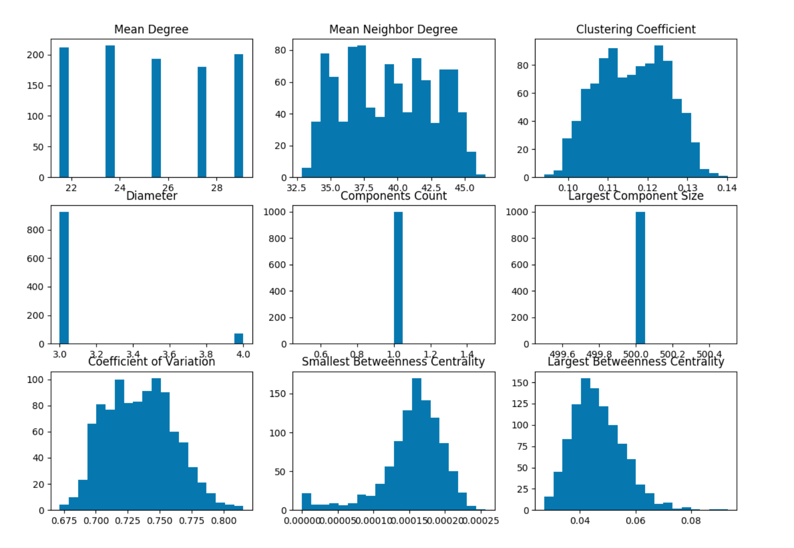


Figure: Restricting mean degree to an interval alters distributions of certain metrics.

Leveraging Graph Structure

We hypothesize the following...

- We can utilize the *average* structure across realizations of different types of random graphs in order to predict the type of structure a newly introduced graph contains within
- Given a graph with missing edges and/or corrupted structure - by observing graph characteristics, and generating random graph realizations with constraints based on observed characteristics, we can see which random graph best fits the underlying structure (and can predict missing edges in this manner).
- We can perform edge prediction while maintaining and recovering the proper statistical distributions and underlying structure

How can we compare the structure of two graphs? And how can we add edges in such a manner that we push the structure of one graph in a desired direction of a known realization?

Joint Degree Distribution

Degree distribution is the probability that a randomly chosen node in a graph will have degree k.

Joint degree distribution (JDD) is the probability that a randomly chosen edge will be between nodes of degree k_1 and k_2.

Joint Degree Distribution

Degree distribution is the probability that a randomly chosen node in a graph will have degree k.

Joint degree distribution (JDD) is the probability that a randomly chosen edge will be between nodes of degree k_1 and k_2.

The joint degree distribution JDD as we use it is a matrix defined as

$$m(k_1, k_2) = \text{number of edges connecting nodes of degrees } k_1 \text{ and } k_2$$

Joint Degree Distribution

Degree distribution is the probability that a randomly chosen node in a graph will have degree k.

Joint degree distribution (JDD) is the probability that a randomly chosen edge will be between nodes of degree k_1 and k_2.

The joint degree distribution JDD as we use it is a matrix defined as

$$m(k_1, k_2) = \text{number of edges connecting nodes of degrees } k_1 \text{ and } k_2$$

JDD is more restrictive than degree distribution and thus contains more information about the connectivity in a graph.

Why is this relevant?

Imposing specific statistical restrictions affects the types of models that can be realized.

Based on a paper by Mahadevan, we would like to infer structure within an incomplete graph by sampling JDD from the space of graphs associated with restricted distributions.

In this way we can infer the most likely true underlying structure.

JDD as a measurement of graph structure similarity

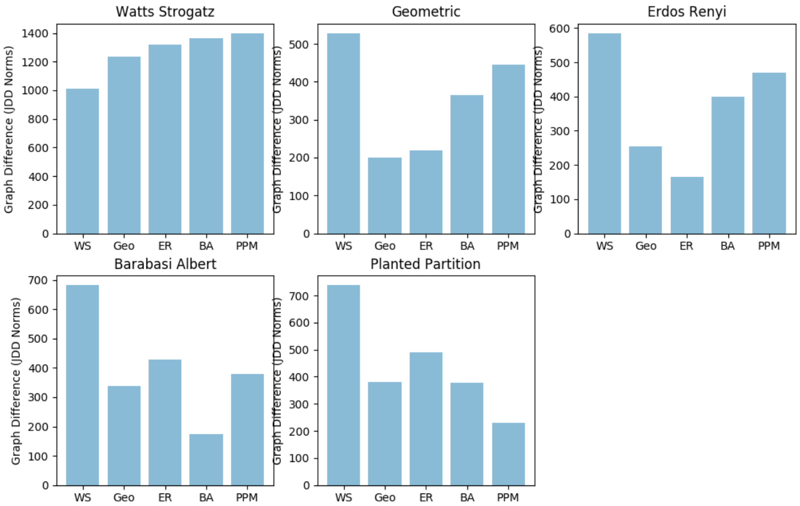


Figure: We show that the norm of the difference between the JDDs is a useful comparison for underlying structure.

Forced Structure Via Optimization

To achieve this, we assume a host of underlying graphical models, and then force our incomplete empirical graph to be as similar as possible. In the context of this project we measured the similarity of two graphs by,

$$\text{Similarity} = ||JDD(G_1) - JDD(G_2)||$$

- Assume n possible random graph models
- under each of the n models, minimize the function described above via the addition of edges.
- stop after a specific number of iterations, or after no better minimum can be reached.
- return the random graph that was closest to its target graph

Future Investigation

Edge Imputation:
To implement the above method, we first considered the iterative addition of edges of variable batch size using a greedy approach.

This gave mediocre results, as the algorithm tended to cease execution before a true minimum was reached. Furthermore, the metric for closeness often tended to favor graphs that "grew" structure such as the barabasi-albert.

To avoid this, we will consider a wider range of proposal functions, thereby allowing for net of possible regimes to consider. We also hope to include better optimization metric, and implement something to make more intelligent decisions about the final proposal distribution.

Future Investigation

Reconstruction under the restricted configuration model:
Regardless of edge imputation accuracy, can we recreate the structure (i.e., is structure independent of specific edges and can we recreate a similar graph but with different edges entirely)?

Future Investigation

How does it all work on real world data?
Do predictions translate over from real-world data sets? Can we predict structure of real world data sets by utilizing the realizations of the random graph models under constraints.

References

Mahadevan, Priya and Krioukov, Dmitri and Fomenkov, Marina and Dimitropoulos, Xenofontas and Vahdat, Amin and others

The Internet AS-level topology: three data sources and one definitive metric

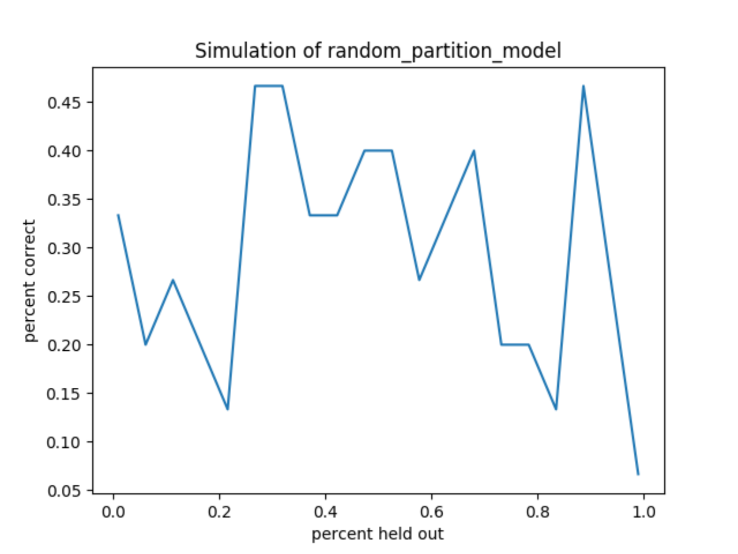
ACM SIGCOMM Computer Communication Review 36(1), 17–26.

Alderson, David L and Li, Lun

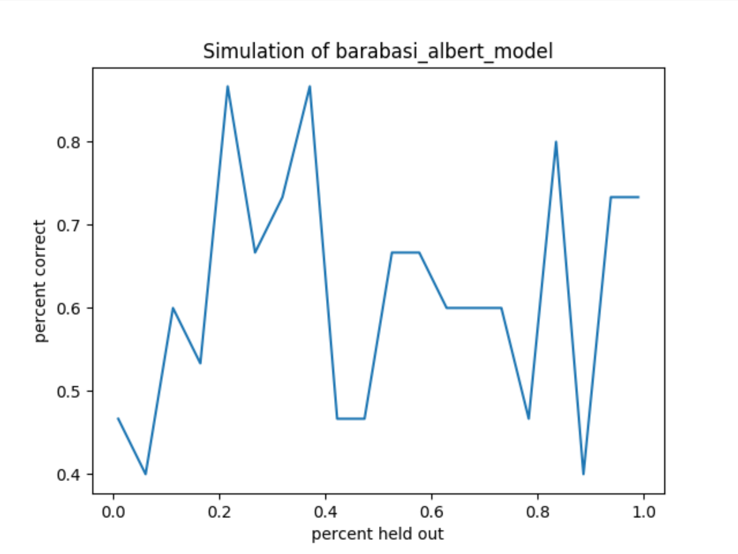
Diversity of graphs with highly variable connectivity

Physical Review E 75(4), 046102.

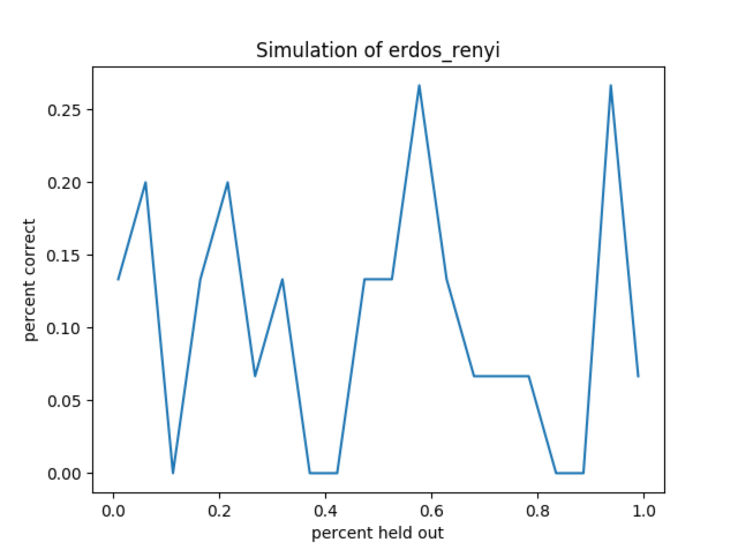
Appendix:


Simulation of random_partition_model
percent correct
percent held out
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.0
0.2
0.4
0.6
0.8
1.0

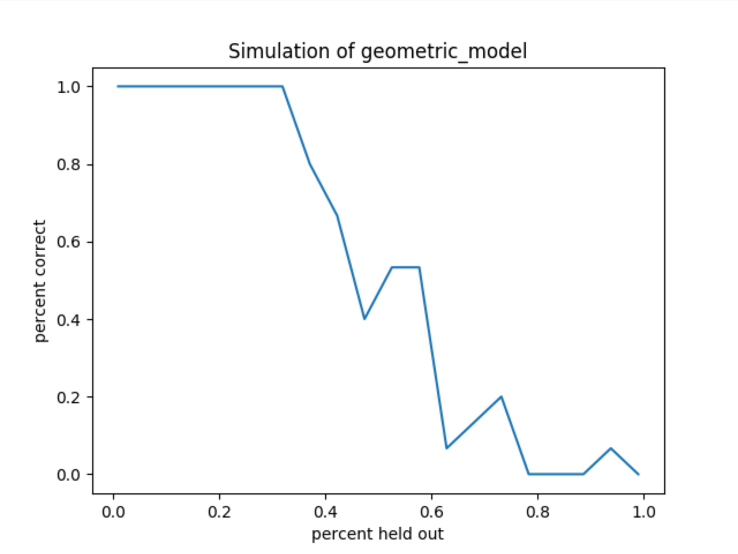
Appendix:


Simulation of barabasi_albert_model
percent correct
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
percent held out

Appendix:


Simulation of erdos_renyi
0.25
0.20
0.15
0.10
0.05
0.00
percent correct
0.0
0.2
0.4
0.6
0.8
1.0
percent held out

Appendix:


Simulation of geometric_model
1.0
0.8
0.6
0.4
0.2
0.0
percent correct
0.0
0.2
0.4
0.6
0.8
1.0
percent held out